AD________________

Award Number: DAMD17-03-2-0038

TITLE: Structural Genomics of Bacterial Virulence Factors

PRINCIPAL INVESTIGATOR: Robert C. Liddington, Ph.D.
Adam Godzik, Ph.D.
Maurizio Pellecchia, Ph.D.

CONTRACTING ORGANIZATION: Burnham Institute
La Jolla, CA 92037

REPORT DATE: May 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE 01-05-2006	**2. REPORT TYPE** Final	**3. DATES COVERED** 1 May 2003 – 30 Apr 2006

4. TITLE AND SUBTITLE Structural Genomics of Bacterial Virulence Factors	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER DAMD17-03-2-0038
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Robert C. Liddington, Ph.D. Adam Godzik, Ph.D. Maurizio Pellecchia, Ph.D.	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Burnham Institute La Jolla, CA 92037	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

Original contains colored plates: ALL DTIC reproductions will be in black and white.

14. ABSTRACT

We applied a focused structural genomics approach to determine the atomic resolution crystal structures of key virulence factors from high priority pathogens. In our first year we focused on proteins encoded by the B. anthracis virulence plasmid, pX01, and the setting up of a virulence factor computational data base. In the second year we expanded our efforts to include genome-encoded proteins of B. anthracis, and structural studies on proteins encoded by Variola virus, the causative agent of smallpox. In year 3 we continued work on Variola proteins, including determining the structure of an important virulence factor, N1L. We also determined the structure of a SARS virus surface protein in complex with a neutralizing antibody. We have generated a large library of expression vectors for virulence factors, as well as research quantities of pure proteins, which could readily be adapted for vaccine design. We determined the structures of 12 novel virulence factors and more than 20 Lethal Factor complex, which can be utilized for rational drug design. The accumulated structural information will generate important and testable hypotheses that will increase our understanding of the molecular mechanisms of pathogenicity, putting us in a stronger position to anticipate and react to emerging pathogens.

15. SUBJECT TERMS

X-ray crystallography; structural genomics; bioinformatics; Biodefense; virulence factor; toxin.

16. SECURITY CLASSIFICATION OF:				**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	229	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

FINAL REPORT: Structural Genomics of Bacterial Virulence Factors

INTRODUCTION

We applied a comprehensive but focused structural genomics approach to determine the atomic resolution crystal structures of key virulence factors from high priority pathogens. The work in our first year focused on proteins encoded by the B. anthracis virulence plasmid, pX01, and the setting up of a virulence factor computational data base. In the second year we expanded our efforts to include genome-encoded proteins of B. anthracis, and structural studies on proteins encoded by Variola virus, the causative agent of smallpox. In year 3 we continued work on Variola proteins, including determining the structure of an important virulence factor, N1L. We also determined the structure of a SARS virus surface protein in complex with a neutralizing antibody. We have generated a large library of expression vectors for virulence factors, as well as research quantities of pure proteins, which could readily be adapted for vaccine design. In the broader and longer term, the accumulated structural information will generate important and testable hypotheses that will increase our understanding of the molecular mechanisms of pathogenicity, putting us in a stronger position to anticipate and react to emerging pathogens.

BODY

Task 1: Atomic resolution crystal structures of virulence factors:

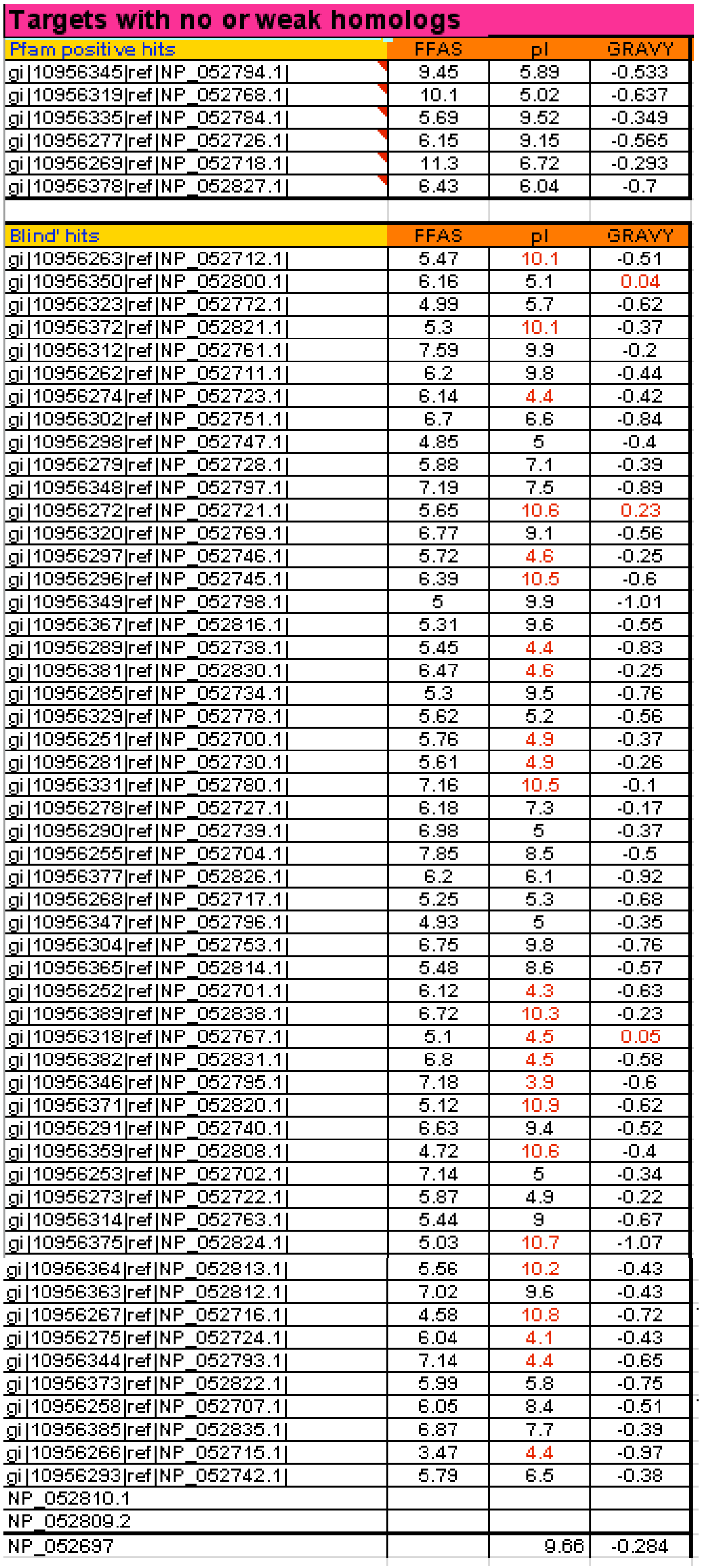
1.a Target Selection on B. anthracis pX01: We performed a detailed analysis of the Bacillus anthracis virulence plasmid (see also Task 3 and Appendix 1). Using a variety of bioinformatics tools we identified the possible function of about 40 proteins, and discovered several likely operons on the pXO1 plasmid. The most interesting discoveries include numerous DNA processing enzymes, several new regulatory proteins, and elements of a type IV secretion system. The results of the analysis of pXO1 are now being prepared for publication. A draft manuscript describing this work is provided in **Appendix 1**.

We identified a new domain in a broad range of bacterial, as well as single archaeal and plant proteins. Its presence in the virulence-related pXO1 plasmid of Bacillus anthracis (pX01-01) as well as in several other pathogens makes it a possible drug target. We term the new domain nuclease-related domain (NERD) because of its distant similarity to endonucleases. This work was published in Trends in Biochemical Sciences (Grynberg and Godzik, 2004) and is included as **Appendix 2.**

1.b Cloning and expression of novel B. anthracis proteins. Two target lists were generated from the bioinformatics approaches: proteins with distant homologues in the protein data base of structures, and a second list of proteins with no homologues. Research Associates from Dr. Liddington's laboratory each chose 5 targets from List 1 and 3 from list 2. The work is summarized below. For the most part, cloning was successful, and expression trials were performed, with several undergoing crystallization and NMR trials. Crystal structures of two novel pX01 proteins have been determined (described in **Appendix 3**). The work on pX01-118 and pX02-62 has led to a focus on the structure of the "master regulator" of the toxin genes, AtxA, and we have made a concerted effort to express full-length and domain fragments in different hosts and in a cell-free system. Our hit rate on soluble protein expression and crystallization has been disappointing when compared with our general success-rate for other bacterial and eukaryotic proteins. The reasons for this are unclear, although certainly several of the proteins appear to be toxic to the host.

Targets with distant sequence homologs in the PDB (Feb.2003) positive hits

Query	length	range	score	%id	covered by template(s)
gi|10956388|ref|NP_052837.1| pXO1-142 [Bacillus anthracis] (04/03/03)	887	1--634	-112	25	1i7d_A mol:protein length:659 DNA Topoisomerase III
gi|10956265|ref|NP_052714.1| pXO1-18 [Bacillus anthracis] (04/03/03)	315	1--315	-77.1	14	1kbu_A mol:protein length:349 Cre Recombinase
gi|10956379|ref|NP_052828.1| pXO1-132 [Bacillus anthracis] (04/03/03)	361	1--357	-75.4	13	1kbu_A mol:protein length:349 Cre Recombinase
gi|10956351|ref|NP_052799.1| pXO1-103 [Bacillus anthracis] (04/03/03)	317	12--310	-75.3	21	1a0p_ mol:protein length:290 Site-Specific Recombinase Xerd
gi|10956387|ref|NP_052836.1| pXO1-141 [Bacillus anthracis] (04/03/03)	214	57--201	-73.6	34	1ez6_A mol:protein length:149 Staphylococcal Nuclease
gi|10956362|ref|NP_052811.1| pXO1-115 [Bacillus anthracis] (04/03/03)	193	2--186	-64.6	31	1gdt_A mol:protein length:183 Gamma-Delta Resolvase
gi|10956342|ref|NP_052791.1| pXO1-95 [Bacillus anthracis] (04/03/03)	443	1--416	-64	25	1dli_A mol:protein length:402 Udp-Glucose Dehydrogenase
gi|10956341|ref|NP_052790.1| pXO1-94 [Bacillus anthracis] (04/03/03)	295	1--294	-63.9	29	1iim_A mol:protein length:292 Glucose-1-Phosphate Thymidylyltransferase
gi|10956343|ref|NP_052792.1| pXO1-96 [Bacillus anthracis] (04/03/03)	274	1--87	-12.6	10	1mdm_A mol:protein length:149 Paired Box Protein Pax-5
		38--270	-55	16	1k6y_A mol:protein length:288 Pol Polyprotein
		108--267	-62.6	19	1a5v_ mol:protein length:162 Integrase
gi|10956305|ref|NP_052754.1| pXO1-58 [Bacillus anthracis] (04/03/03)	272	4--265	-55.6	16	1ion_A mol:protein-het length:243 Probable Cell Division Inhibitor Mind
gi|10956284|ref|NP_052733.1| pXO1-37 [Bacillus anthracis] (04/03/03)	193	1--162	-55	10	1ghe_A mol:protein-het length:177 Acetyltransferase
gi|10956334|ref|NP_052783.1| pXO1-87 [Bacillus anthracis] (04/03/03)	160	2--160	-27.5	11	1jfu_A mol:protein length:186 Thiol: Disulfide Interchange Protein Tlpa
		53--157	-52.2	17	1quw_A mol:protein length:105 Thioredoxin
gi|10956338|ref|NP_052787.1| pXO1-91 [Bacillus anthracis] (04/03/03)	280	1--162	-51.6	14	1d2t_A mol:protein length:231 Acid Phosphatase
gi|10956358|ref|NP_052807.1| pXO1-111 [Bacillus anthracis] (04/03/03)	204	1--154	-51.5	37	1acc_ mol:protein length:735 Anthrax Protective Antigen
gi|10956383|ref|NP_052832.1| pXO1-137 [Bacillus anthracis] (04/03/03)	61	2--60	-51	40	1kq1_A mol:protein length:77 Host Factor For Q Beta
gi|10956292|ref|NP_052741.1| pXO1-45 [Bacillus anthracis] (04/03/03)	435	2--390	-49.5	17	1fsz_ mol:protein length:372 Ftsz
		17--427	-37.3	9	1ffx_A mol:protein length:451 Tubulin
gi|10956340|ref|NP_052789.1| pXO1-93 [Bacillus anthracis] (04/03/03)	366	2--254	-49.2	14	1qg8_A mol:protein length:255 Spore Coat Polysaccharide Biosynthesis Prote
gi|10956306|ref|NP_052755.1| pXO1-59 [Bacillus anthracis] (04/03/03)	477	3--477	-11.8	9	1e32_A mol:protein length:458 P97
		79--406	-47.2	19	1g6o_A mol:protein-het length:330 Cag-Alpha
		197--472	-14.5	12	1jj7_A mol:protein length:260 Peptide Transporter Tap1
gi|10956374|ref|NP_052823.1| pXO1-127 [Bacillus anthracis] (04/03/03)	214	1--87	-12.4	11	1mdm_A mol:protein length:149 Paired Box Protein Pax-5
		37--214	-37.1	14	1k6y_A mol:protein length:288 Pol Polyprotein
		111--214	-45.1	14	1vsd_ mol:protein length:152 Integrase
gi|10956294|ref|NP_052743.1| pXO1-47 [Bacillus anthracis] (04/03/03)	201	11--113	-42.3	13	1jbg_A mol:protein length:109 Transcription Activator Of Multidrug-Efflux
gi|10956257|ref|NP_052706.1| pXO1-10 [Bacillus anthracis] (04/03/03)	363	4--361	-41	14	2adm_A mol:protein length:421 Adenine-N6-DNA-Methyltransferase Taqi
gi|10956376|ref|NP_052825.1| pXO1-129 [Bacillus anthracis] (04/03/03)	137	7--137	-30.5	14	1k6y_A mol:protein length:288 Pol Polyprotein
		55--137	-39	16	1vsd_ mol:protein length:152 Integrase
gi|10956366|ref|NP_052815.1| pXO1-119 [Bacillus anthracis] (04/03/03)	475	8--139	-13.2	10	1j5y_A mol:protein-het length:187 Transcriptional Regulator, Biotin Repres
		161--387	-38.9	14	1h99_A mol:protein length:224 Transcription Antiterminator Lict
gi|10956287|ref|NP_052736.1| pXO1-40 [Bacillus anthracis] (04/03/03)	65	1--65	-33.2	16	1adr_ mol:protein length:76 P22 C2 Repressor (Amino-Terminal DNA-Binding
gi|10956337|ref|NP_052786.1| pXO1-90 [Bacillus anthracis] (04/03/03)	652	224--650	-20.6	11	1cii_ mol:protein length:602 Colicin Ia
		301--593	-32.8	12	2tma_A mol:protein length:284 Tropomyosin - Chain A
gi|10956356|ref|NP_052805.1| pXO1-109 [Bacillus anthracis] (04/03/03)	99	9--97	-29.2	22	1smt_A mol:protein length:122 Transcriptional Repressor Smtb
gi|10956384|ref|NP_052833.1| pXO1-138 [Bacillus anthracis] (04/03/03)	97	10--95	-29.1	18	1smt_A mol:protein length:122 Transcriptional Repressor Smtb
gi|10956326|ref|NP_052775.1| pXO1-79 [Bacillus anthracis] (04/03/03)	1222	3--977	-9.92	9	1i84_S mol:protein-het length:1184 Smooth Muscle Myosin Heavy Chain
		14--271	-10.2	14	1qle_C mol:protein length:273 Cytochrome C Oxidase Polypeptide III
		936--1169	-28.9	14	1qu7_A mol:protein length:227 Methyl-Accepting Chemotaxis Protein I
		987--1221	-14.3	8	1h6w_A mol:protein length:312 Bacteriophage T4 Short Tail Fibre
gi|10956286|ref|NP_052735.1| pXO1-39 [Bacillus anthracis] (04/03/03)	325	1--323	-28.1	11	1mm8_A mol:protein length:481 Tn5 Transposase
gi|10956283|ref|NP_052732.1| pXO1-36 [Bacillus anthracis] (04/03/03)	484	5--473	-27.6	12	1mm8_A mol:protein length:481 Tn5 Transposase
gi|10956328|ref|NP_052777.1| pXO1-81 [Bacillus anthracis] (04/03/03)	424	1--209	-27.3	13	1qsa_A mol:protein length:618 Soluble Lytic Transglycosylase Slt70
gi|10956282|ref|NP_052731.1| pXO1-35 [Bacillus anthracis] (04/03/03)	478	16--478	-25.3	11	1mm8_A mol:protein length:481 Tn5 Transposase
gi|10956380|ref|NP_052829.1| pXO1-133 [Bacillus anthracis] (04/03/03)	485	6--302	-10.1	12	1e32_A mol:protein length:458 P97
		183--485	-22.5	17	1pjr_ mol:protein length:724 Pcra
gi|10956270|ref|NP_052719.1| pXO1-23 [Bacillus anthracis] (04/03/03)	461	8--337	-20.7	9	1khv_A mol:protein length:516 RNA-Directed RNA Polymerase
		12--443	-16.1	9	1rdr_ mol:protein length:461 Poliovirus 3D Polymerase
gi|10956325|ref|NP_052774.1| pXO1-78 [Bacillus anthracis] (04/03/03)	405	125--398	-20.1	12	1gki_A mol:protein length:437 Conjugal Transfer Protein Trwb
gi|10956368|ref|NP_052817.1| pXO1-121 [Bacillus anthracis] (04/03/03)	57	1--56	-19.6	21	1qb7_A mol:protein length:236 Adenine Phosphoribosyltransferase
gi|10956254|ref|NP_052703.1| pXO1-07 [Bacillus anthracis] (04/03/03)	602	1--566	-12	8	1c2p_A mol:protein-het length:576 RNA-Dependent RNA Polymerase
		7--412	-19.1	11	1khv_A mol:protein length:516 RNA-Directed RNA Polymerase
		513--602	-11.5	17	1bvb_ mol:protein length:211 Cytochrome C-554
gi|10956271|ref|NP_052720.1| pXO1-24 [Bacillus anthracis] (04/03/03)	132	1--113	-16.1	16	1fa0_A mol:protein length:537 Poly(A)-Polymerase
gi|10956280|ref|NP_052729.1| pXO1-33 [Bacillus anthracis] (04/03/03)	259	1--257	-15.2	11	1kan_A mol:protein length:253 Kanamycin Nucleotidyltransferase (E.C. 2.7.7
gi|10956352|ref|NP_052801.1| pXO1-105 [Bacillus anthracis] (04/03/03)	67	6--28	-14.5	43	1ekt_A mol:protein length:53 Transcription State Regulatory Protein Abrb
gi|10956260|ref|NP_052709.1| pXO1-13 [Bacillus anthracis] (04/03/03)	1320	358--1310	-13.8	13	1k83_A mol:protein length:1733 DNA-Directed RNA Polymerase II Largest Subu
gi|10956317|ref|NP_052766.1| pXO1-70 [Bacillus anthracis] (04/03/03)	437	80--437	-13.7	15	1dd9_A mol:protein length:338 DNA Primase
gi|10956261|ref|NP_052710.1| pXO1-14 [Bacillus anthracis] (04/03/03)	564	503--556	-11.9	14	1b0n_A mol:protein length:111 Sinr Protein
gi|10956276|ref|NP_052725.1| pXO1-29 [Bacillus anthracis] (04/03/03)	274	1--61	-11.8	30	1ii8_A mol:protein length:195 Rad50 Abc-Atpase
gi|10956269|ref|NP_052718.1| pXO1-22 [Bacillus anthracis] (04/03/03)	91	9--91	-11.3	16	1ni5_A mol:protein-het length:433 Putative Cell Cycle Protein Mesj



Targets with no or weak homologs

Pfam positive hits	FFAS	pI	GRAVY
gi|10956345|ref|NP_052794.1|	9.45	5.89	-0.533
gi|10956319|ref|NP_052768.1|	10.1	5.02	-0.637
gi|10956335|ref|NP_052784.1|	5.69	9.52	-0.349
gi|10956277|ref|NP_052726.1|	6.15	9.15	-0.565
gi|10956269|ref|NP_052718.1|	11.3	6.72	-0.293
gi|10956378|ref|NP_052827.1|	6.43	6.04	-0.7
Blind' hits	FFAS	pI	GRAVY
gi|10956263|ref|NP_052712.1|	5.47	10.1	-0.51
gi|10956350|ref|NP_052800.1|	6.16	5.1	0.04
gi|10956323|ref|NP_052772.1|	4.99	5.7	-0.62
gi|10956372|ref|NP_052821.1|	5.3	10.1	-0.37
gi|10956312|ref|NP_052761.1|	7.59	9.9	-0.2
gi|10956262|ref|NP_052711.1|	6.2	9.8	-0.44
gi|10956274|ref|NP_052723.1|	6.14	4.4	-0.42
gi|10956302|ref|NP_052751.1|	6.7	6.6	-0.84
gi|10956298|ref|NP_052747.1|	4.85	5	-0.4
gi|10956279|ref|NP_052728.1|	5.88	7.1	-0.39
gi|10956348|ref|NP_052797.1|	7.19	7.5	-0.89
gi|10956272|ref|NP_052721.1|	5.65	10.6	0.23
gi|10956320|ref|NP_052769.1|	6.77	9.1	-0.56
gi|10956297|ref|NP_052746.1|	5.72	4.6	-0.25
gi|10956296|ref|NP_052745.1|	6.39	10.5	-0.6
gi|10956349|ref|NP_052798.1|	5	9.9	-1.01
gi|10956367|ref|NP_052816.1|	5.31	9.6	-0.55
gi|10956289|ref|NP_052738.1|	5.45	4.4	-0.83
gi|10956381|ref|NP_052830.1|	6.47	4.6	-0.25
gi|10956285|ref|NP_052734.1|	5.3	9.5	-0.76
gi|10956329|ref|NP_052778.1|	5.62	5.2	-0.56
gi|10956251|ref|NP_052700.1|	5.76	4.9	-0.37
gi|10956281|ref|NP_052730.1|	5.61	4.9	-0.26
gi|10956331|ref|NP_052780.1|	7.16	10.5	-0.1
gi|10956278|ref|NP_052727.1|	6.18	7.3	-0.17
gi|10956290|ref|NP_052739.1|	6.98	5	-0.37
gi|10956255|ref|NP_052704.1|	7.85	8.5	-0.5
gi|10956377|ref|NP_052826.1|	6.2	6.1	-0.92
gi|10956268|ref|NP_052717.1|	5.25	5.3	-0.68
gi|10956347|ref|NP_052796.1|	4.93	5	-0.35
gi|10956304|ref|NP_052753.1|	6.75	9.8	-0.76
gi|10956365|ref|NP_052814.1|	5.48	8.6	-0.57
gi|10956252|ref|NP_052701.1|	6.12	4.3	-0.63
gi|10956389|ref|NP_052838.1|	6.72	10.3	-0.23
gi|10956318|ref|NP_052767.1|	5.1	4.5	0.05
gi|10956382|ref|NP_052831.1|	6.8	4.5	-0.58
gi|10956346|ref|NP_052795.1|	7.18	3.9	-0.6
gi|10956371|ref|NP_052820.1|	5.12	10.9	-0.62
gi|10956291|ref|NP_052740.1|	6.63	9.4	-0.52
gi|10956359|ref|NP_052808.1|	4.72	10.6	-0.4
gi|10956253|ref|NP_052702.1|	7.14	5	-0.34
gi|10956273|ref|NP_052722.1|	5.87	4.9	-0.22
gi|10956314|ref|NP_052763.1|	5.44	9	-0.67
gi|10956375|ref|NP_052824.1|	5.03	10.7	-1.07
gi|10956364|ref|NP_052813.1|	5.56	10.2	-0.43
gi|10956363|ref|NP_052812.1|	7.02	9.6	-0.43
gi|10956267|ref|NP_052716.1|	4.58	10.8	-0.72
gi|10956275|ref|NP_052724.1|	6.04	4.1	-0.43
gi|10956344|ref|NP_052793.1|	7.14	4.4	-0.65
gi|10956373|ref|NP_052822.1|	5.99	5.8	-0.75
gi|10956258|ref|NP_052707.1|	6.05	8.4	-0.51
gi|10956385|ref|NP_052835.1|	6.87	7.7	-0.39
gi|10956266|ref|NP_052715.1|	3.47	4.4	-0.97
gi|10956293|ref|NP_052742.1|	5.79	6.5	-0.38
NP_052810.1			
NP_052809.2			
NP_052697		9.66	-0.284

1.3 Alternative Expression Systems for B. anthracis protein: We investigated Bacillus expression systems to see if these would provide a superior system for expressing B. anthracis proteins. Though *Bacillus* strains are broadly used for industrial expression of heterologous proteins, there was only one commercially available expression system. Furthermore, their shuttle plasmid was underdeveloped - it did not have purification tags and secretion peptides. There are numerous *Bacillus subtilis* strains and plasmids, but they have been used mostly for functional studies, where overexpression of a protein is not important. We tested two systems, *Bacillus subtilis* and *Bacillus megaterium*. Derivatives of *Bacillus subtilis* strain 168 (1A436, S53, 1A1) and the plasmid pDG148 were obtained from the Bacillus Genetic Stock Center (Ohio University). *Bacillus megaterium* strain WH320 and the plasmid pWH1520 were obtained from MoBiTec (Germany). *Bacillus subtilis* strain 168 has a natural ability for transformation (intake of plasmid DNA through the cell wall). The protein expression, however, is problematic, because this strain undergoes sporulation when the expressed protein is toxic or the growth conditions are not optimal. The value of this system for secreted expression is also limited, because *B. subtilis* produces too many proteases. *B. megaterium* strain WH320 does not sporulate, the shuttle plasmid is fairly stable there, and it not secrete many proteases. However, *B. megaterium* does not take plasmids by transformation. The alternative protocol, which requires removal of the cell wall by lysozyme, is unreliable.

We successfully adopted the two Bacillus expression systems and tested expression of following genes pX01-97, -99, -118, -119 and -125, which did not express well in *E. coli*. Gene pX01-118, which expressed well in E. coli, was used as a positive control. We found that the level of protein expression correlated closely with the level of expression in *E. coli*. The highest expression was obtained for pX01-118 using *B. megaterium*; nevertheless, the expression level per gram of cell mass was about 0.5-2 mg, which is 5 times lower then the expression from the pET plasmid in *E.coli*. The expression of other soluble proteins was detectable by Western blot against His-tag, but insufficient for crystallization. The expression of pX01-118 in *B. subtilis* strains was unstable. Often cells began to sporulate even before induction of protein expression (the IPTG promoter was very leaky). We tested the plasmid PDG148 with *B. megaterium* and the plasmid pWH1520 with *B. subtilis*. Contrary to the claims of MoBiTec, the plasmids did not perform well in foreign cells.

We concluded that intracellular expression in Bacillus species does not give an advantage over *E. coli* system, perhaps because the codon usage is similar and *E. coli* has a better developed chaperoning system. However, it is still possible that *B. megaterium* is beneficial for expression of secreted proteins.

1.4 Successful Structure determinations:

1.4.1 Structural Studies of inhibitor binding to Lethal Factor

We worked with Dr. Cantley from Harvard Medical School, and using an optimized peptidic substrate, defined the structural basis of substrate recognition and inhibition by peptidic mimetics at atomic resolution (Turk et al., 2004); **Appendix 4**. We also worked closely with in collaboration with Drs. Gussio and Bavari at USAMRIID/NCI. Compounds NSC 12155, NSC 357756, NSC 357777 had been identified as the top 3 hits in the NCI small molecules library high throughput screen for LF inhibition. We determined the crystal structure of LF-12155-Zn (LF wild-type bound to NSC 12155 in the presence of zinc), and this work has been published (Panchal et al., 2004); Figure 1; **Appendix 5**). It showed a compound that is able to bind and inhibit up to 95% of the native catalytic activity of LF. This compound does not require the presence of zinc to bind to the active site of LF, and appears to recognize the substrate-binding site immediately adjacent to the catalytic zinc site through hydrophobic interactions.

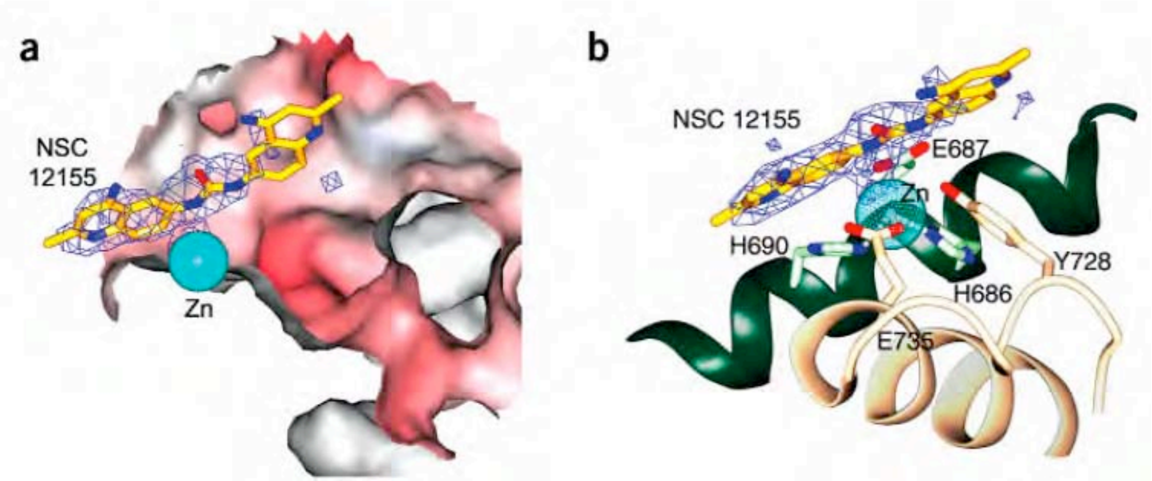


Figure 1: *X-ray crystal structure of the LF–NSC 12155–Zn complex. at 2.9-Å resolution. (a) Detailed view of the electron density trace and overall model fit of NSC 12155. Molecular surface of LF colored by charge (red, negative; blue, positive), with Zn2+ (cyan), and the model of the inhibitor molecule NSC 12155 (yellow) in stick representation. The difference map, 2Fo – Fc, is contoured at 1.1 σ level. (b) The inhibitor NSC 12155 bound in the active site of LF. The difference map, 2Fo – Fc, is contoured at 1.0 σ.*

Our work to determine LF-inhibitor complexes in collaboration with the Bavari and Gussio groups at USAMRIID and NCI continues. We have collected data sets from the following co-crystals crystals LF-357756-Zn and LF-357756-Zn (complex of LF wild-type bound to NSC 357756 or NSC 357777 in the presence of zinc), and model refinement is still in progress, and new data are being collected. So far, electron density maps indicate that compound NSC 357756 is bound in the immediate vicinity of the catalytic site, and may be coordinating the zinc atom. NSC 357777 however appears to be relying more on hydrophobic interactions in recognizing the substrate-binding site in LF, while still binding close to the zinc atom. Currently, the focus is on NSC 357756, which has been shown to have better cell permeability abilities than NSC 12155 and better inhibitory abilities than NSC. We also used the system developed here to test a distinct set of compounds, including an inhibitor that was successfully tested in a mouse model (Forino et al., 2005).

The crystal structure of full-length LF was grown under high salt conditions, and this may have hampered in several cases the determination of high quality inhibitor complexes. To try to overcome these problems we have cloned, expressed and crystallized a fragment of LF that lacks domain 1 (the PA-binding domain), but that contains the critical catalytic module (Domains 2-4). This protein expresses readily in E. coli, and crystallizes from low salt (PEG) conditions; it also diffracts X-rays to high resolution. We are now in the process of repeating our inhibitors soaks and co-crystallization experiments under these low salt conditions.

1.4.2 Crystal structure of an anthrax toxin-host cell receptor complex

Two closely related host cell receptor molecules, TEM8 and CMG2, bind to PA with high affinity and are required for toxicity. We determined the crystal structure of the PA-CMG2 complex at 2.5 Å resolution (Santelli et al., 2004); Figure 2; **Appendix 6**). The structure reveals an extensive receptor-pathogen interaction surface that mimics the non-pathogenic recognition of the extracellular matrix by integrins. The binding surface is closely conserved in the two receptors and across species, but quite different in the integrin domains, explaining the specificity of the interaction. CMG2 engages two domains of PA, and modeling of the receptor-bound PA63 heptamer suggests that the receptor acts as a pH-sensitive chaperone to ensure accurate and timely membrane insertion.

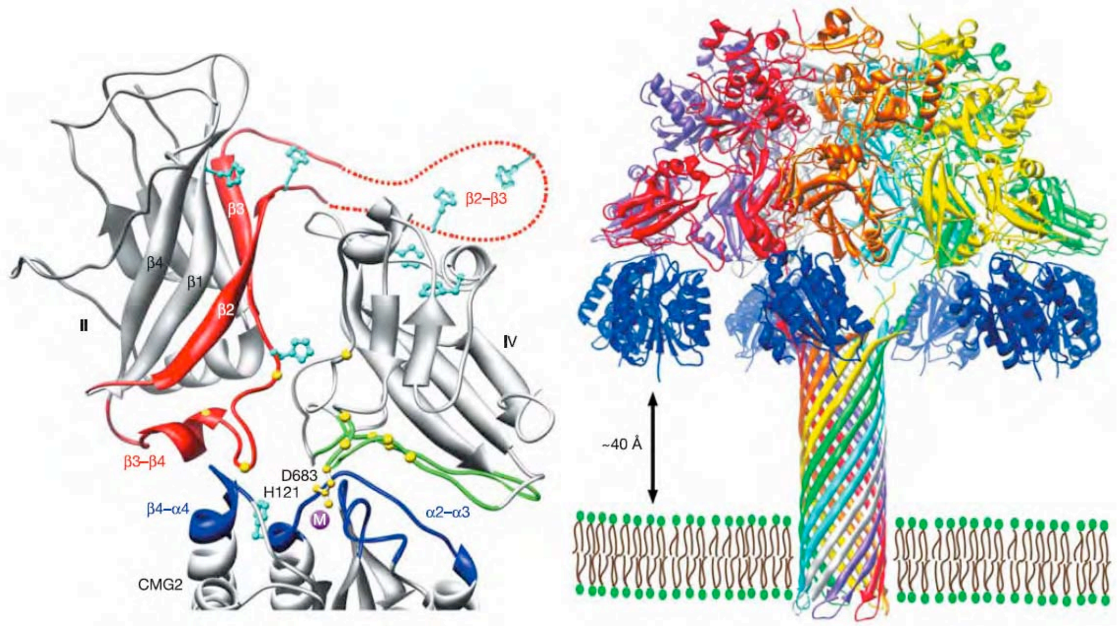


***Figure 2**: (Left Panel) Intermolecular contacts between PA domains II and IV and CMG2. Contacting regions are coloured blue and green for CMG2 and PA domain IV, respectively. The b2–b3 loop and flanking regions of PA domain II, which are implicated in pore formation, are highlighted in red.. Mutation sites that reduce binding by .100-fold are highlighted in gold. (Right Panel) Hypothetical model of the receptor-bound, membrane-inserted PA pore. The model is based on the pre-pore PA63 crystal structure, channel conductance studies, and the crystal structure of α-haemolysin.*

1.4.3 A new family of sensor histidine kinases involved in sporulation: Using bioinformatics approaches we discovered two plasmids encode proteins that are highly homologous to the signal sensor domain of a chromosomally encoded major sporulation sensor histidine kinase (BA2291). In collaboration with Dr. Marta Perego we showed that B. anthracis Sterne overexpressing the plasmid pXO2-61-encoded signal sensor domain exhibited a significant decrease in sporulation that was suppressed by the deletion of the BA2291 gene. Expression of the sensor domains from the pXO1-118 and pXO2-61 genes in Bacillus subtilis strains carrying the B. anthracis sporulation sensor kinase BA2291 gene resulted in BA2291-dependent inhibition of sporulation. These results indicate that sporulation sensor kinase BA2291 is converted from an activator to an inhibitor of sporulation in its native host by the virulence plasmid-encoded signal sensor domains. We speculate that activation of these signal sensor domains contributes to the initiation of B. anthracis sporulation in the bloodstream of its infected host, a salient characteristic in the virulence of this organism, and provides an additional role for the virulence plasmids in anthrax pathogenesis. This work has been published (White et al., 2006); Appendix 7. We have also determined the crystal structures of the two plasmid-encoded proteins, pXO1-118 and pXO2-61. The crystal structures suggest that competition with BA2291 for the binding of an unidentified signaling molecule provides a plausible mechanism for their inhibitory effect. This work is being prepapred for publication: Stranzl et al "Crystal structure of virulence plasmid-encoded sensor domains inhibiting sporulation in *Bacillus anthracis" (Figure 3:* Appendix 3*)*

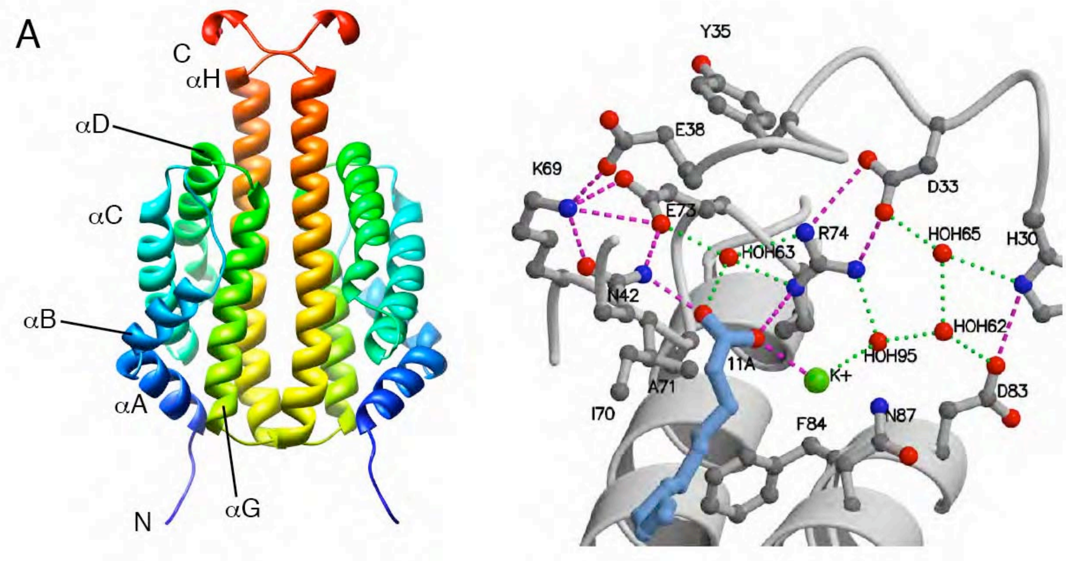


Figure 3: *(Left Panel) Ribbon representation of pXO1-118, colored blue to red from the N to the C terminus, showing the globin fold. (Right panel) The hydrogen bonding network at the top of the cavity containing the fatty acid (blue stick).*

1.4.4 Structure of the B. anthracis epimerase involved in lysine biosynthesis

Lysine biosynthesis in bacteria provides the essential components both for L-lysine for protein synthesis and meso-diaminopimelate for construction of the bacterial peptideglycan cell wall. Since it this process is unique to bacteria, the enzymes in the pathway may be useful for antibiotic design. Genome sequence analysis of *B. anthracis* revealed the complete sequences of enzymes involved in lysine biosynthesis. Moreover, humans are auxotrophic for lysine and therefore are unlikely to be affected by such compounds. The enzyme catalyzes the reversible conversion of *meso*-DAP to LL-DAP but not to DD-DAP; two cysteines constitute the active site and likely act as an acid/base couple. We determined the crystal structure of the DAP-epimerase of *B. anthracis* at the resolution of 2.4 Å (Figure 4) and its analysis suggests that it is in the reduced, active state; the activity of the enzyme still has to be confirmed, and we are currently screening for potential inhibitors.

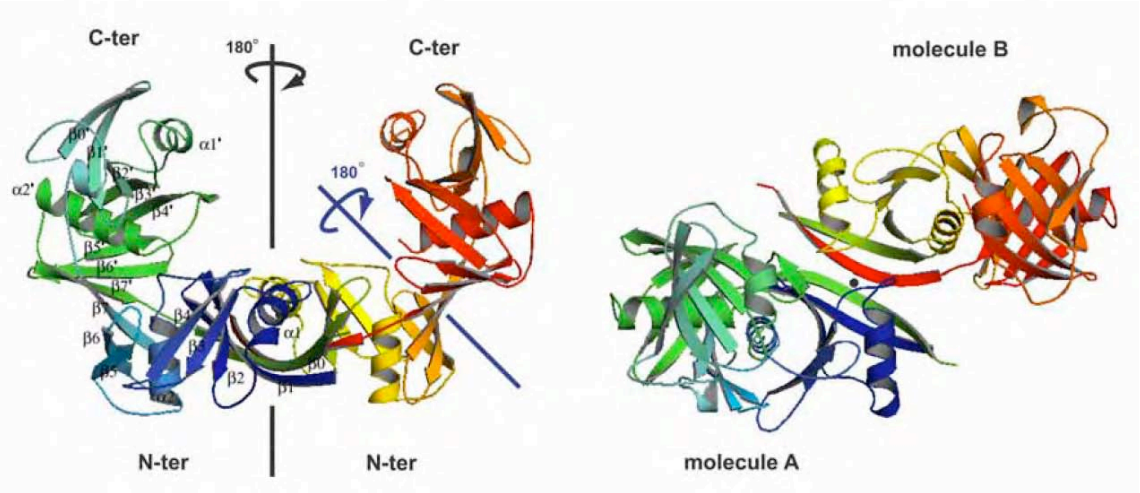


Figure 4: *Side view (left) and top view (right) ribbon representation of the B. anthracis DAP-EP dimer, corresponding to the asymmetric unit in the crystal; the monomers are related by a 2-fold pseudo rotation (black line). The N-terminal (N-ter) and C-terminal (C-ter) domains are also related by a 2-fold pseudo-rotation (blue line). The black dot in between molecules A and B in the right view indicates the pseudo-rotation axis for the dimer, placed in between both G280 of molecules A and B. No electron density could be seen from residues G219 to A225 for molecule B.*

1.4.5 B. anthracis endolysins studies (one paper published in J. Biol. Chem; three in preparation; see Appendices 8 and 9)

Endolysins are cell wall-dissolving enzymes used by bacteriophage to lyse its host to release its progeny, and are potential antibacterial agents. The aim of this study was to examine if the integrated copies of prophage endolysins within the *B. anthracis* Stern strain can be used as anti-bacterial agents for the treatment and prophylaxis of anthrax and other Gram positive bacterial infection when added as purified components exogenously. Two targets were selected, one prophage amidase and one prophage glycosidase, from the *B. anthracis* Stern strain. They are two-domain proteins, consisting of a N-terminal catalytic domain and a C-terminal 80 amino acid putative cell-wall binding domain. The amidase cleaves the bond between the N-Acetylmuramic acid and the L-Alanine, while the glycosidase cleaves the bond between N-acetylglucosamine and N-Acetylmuramic acid of the cell wall. The C-terminal cell wall binding domain of the two endolysin has very high sequence homology (68% identity). This minimum catalytic domains will be tested on other Gram positive bacteria strains in the near future, as soon as they become available.

We determined the structural carried out an in vitro functional analysis of the lambda prophage Ba02 endolysin (PlyL) encoded by the Bacillus anthracis genome (Low et al., 2005); Figure 4; **Appendix 8**. We showed that PlyL comprises two autonomously folded domains, an N-terminal catalytic domain and a C-terminal cell wall-binding domain. We determined the crystal structure of the catalytic domain; its three-dimensional fold is related to that of the cell wall amidase, T7 lysozyme, and contains a conserved zinc coordination site and other components of the catalytic machinery. We demonstrated that PlyL is an N-acetylmuramoyl-L-alanine amidase that cleaves the cell wall of several Bacillus species when applied exogenously. We show, unexpectedly, that the catalytic domain of PlyL cleaves more efficiently than the full-length protein, except in the case of Bacillus cereus, and using GFP-tagged cell wall-binding domain, we detected strong binding of the cell wall-binding domain to B. cereus but not to other species tested. We further show that a related endolysin (Ply21) from the B. cereus phage, TP21, shows a similar pattern of behavior. To explain these data, and the species specificity of PlyL, we propose that the C-terminal domain inhibits the activity of the catalytic domain through intramolecular interactions that are relieved upon binding of the C-terminal domain to the cell wall. Furthermore, our data show that (when applied exogenously) targeting of the enzyme to the cell wall is not a prerequisite of its lytic activity, which is inherently high. These results may have broad implications for the design of endolysins as therapeutic agents.

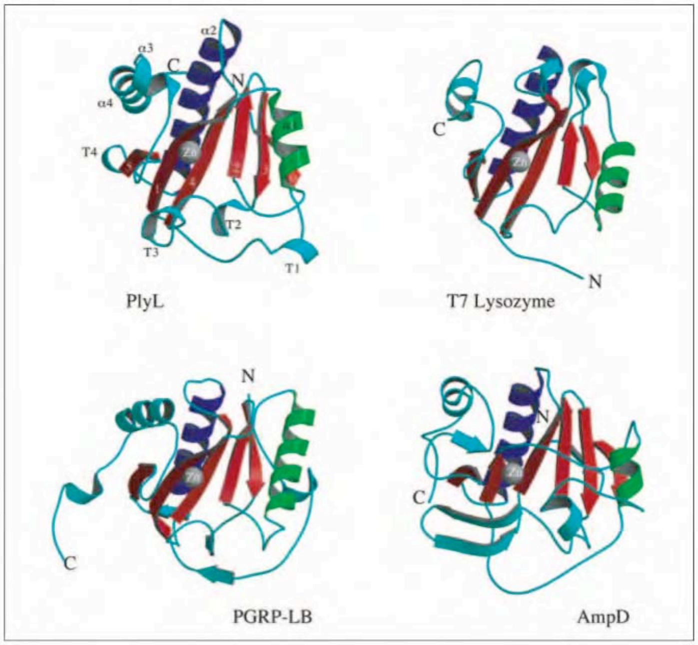


Figure 4: *3-dimensional structure of PlyL catalytic domain and related amidases: T7 lysozyme (PDB: 1LBA), PGRP-LB (PDB: 1OHT), and AmpD (PDB: 1J3G). The zinc ion is shown as a gray sphere. The colors represent the secondary structure arrangement. The backbone RMS differences with T7 lysozyme and PGRP-LB, are 1.8 Å (for 107 atoms) and 2.0 Å (for 106 atoms), respectively.*

Another endolysin, with a similar C-terminal cell wall binding (CWB) as the PlyL was found in the LambdaBa04 prophage region of the Bacillus anthracis str. Ames. We determined its structure at 1.4 Å resolution (manuscript in preparation; **Appendix 9**. The selectivity is similar to that of the PlyL, but kills bacilli four times faster than PlyL. We solved the structure of the N-terminal catalytic domain by single-isomorphous replacement to a resolution of 1.4 Å. Using calorimetry, we showed that the catalytic domain is more active and less selective than the full-length enzyme, which highlights the usefulness of using only the catalytic domain for deverloping therapeutic agents to treat bacillus infection.

1.4.6 Structure of the SARS S1 (spike protein) and its complex with a high affinity antibody.

The etiological agent of SARS is a novel coronavirus (SARS-CoV). The coronaviral surface spike protein S is a type I transmembrane glycoprotein that mediates initial host binding via the cell surface receptor angiotensin-converting enzyme 2 (ACE2), as well as the subsequent membrane fusion events required for cell entry. In collaboration with Dr, Wayne Marasco, Dana Farber Cancer Institute, Boston, we conducted a a structural study of the SARS S1 spike protein with a high affinity neutralizing antibody, "80R". Both the S1 protein and antibody were expressed and purified in milligram quantities. We crystallized the S1 receptor binding domain (RBD) at 2.2 Å resolution and its complex with the antibody at 2.3 Å resolution (Hwang et al., 2006); Figure 5; **Appendix 10).** This work showed that the 80R binding epitope on the S1 RBD overlaps very closely with the ACE2 binding site, providing a rationale for the antibody's strong binding and broad neutralizing ability. The work also provides a structural basis for the differential effects of certain mutations in the spike protein on 80R versus ACE2 binding, including escape mutants, which should facilitate the design of immunotherapeutics to treat a future SARS outbreak. We further showed that the RBD of S1 forms dimers via an extensive interface that is disrupted in receptor- and antibody-bound crystal structures, and we proposed a role for the dimer in virus stability and infectivity.

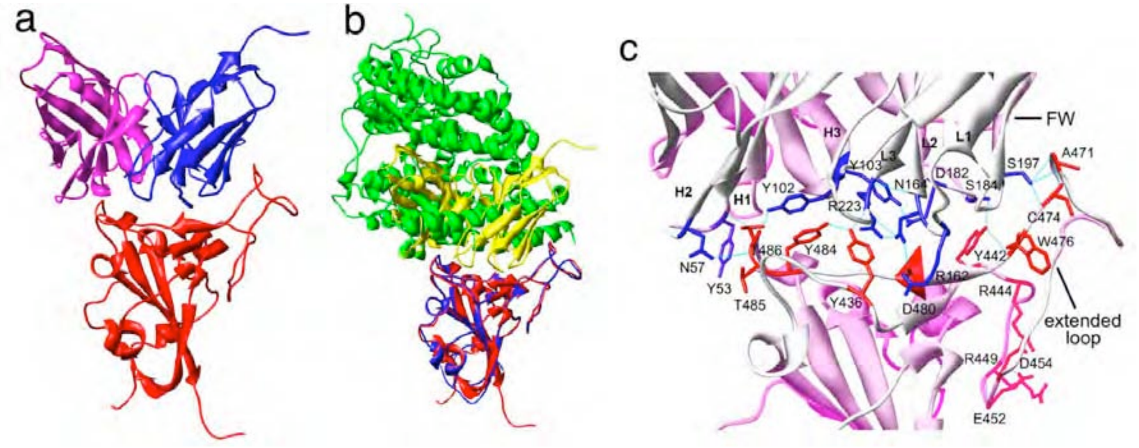


Figure 5: *Structure of the S1-RBD-80R complex. Panel a, Overall structure of the complex. Antibody variable region light chain is in blue and heavy chain is in magenta; S1-RBD is in red. Panel b, Comparison between the S1 RBD-80R complex (red and yellow) and the S1 RBDACE2 complex (blue and green) overlaid on the S1-RBD domain. Panel c, Close-up of the interface. Selected S1 side-chains are in red; 80R in blue; hydrogen bonds in cyan.*

1.4.7 Structural studies of Variola proteins

Despite its eradication from the world population in 1980 and the subsequent discontinuation of widespread vaccination, Variola virus, or smallpox, has emerged as a bioterrorist threat. Vaccinia virus is a close relative of Variola within the poxvirus family, being 99% identical at the amino acid level. Thus, our approach has been to structurally characterize Vaccinia viral virulence factors with the ultimate goal of identifying lead compounds that target these factors and serve as the foundation for the development of anti-poxviral therapeutics. Variola and Vaccinia have a large genome (192 kb), which encodes for 197 gene products. The poxviral “non-structural” (i.e., not part of the mature virus architecture) proteins allow poxviruses to be largely autonomous of host-encoded gene products. Our current focus has been directed at three virulence factors from Vaccinia.

The Vaccinia virus **H1**, also known as VH1, is a dual specificity phosphatase, dephosphorylating both serine-/threonine- and tyrosine-modified residues. VH1 is encapsidated within the virion, is essential for viral transcription (2), dephosphorylates the poxviral A14 protein, and blocks activation of Gamma Interferon. The last is an important mechanism for modulating the host immune system to promote viral viability. We have overexpressed VH1 in E. coli to >90% homogeneity as judged by SDS PAGE, with a yield of approximately 10 mg/L of cell culture and have also demonstrated phosphatase activity (Fig.). Crystallization trials are in progress.

The Vaccinia virus **F10** protein is one of two virally encoded protein kinases, and is an essential virulence factor. F10, like VH1, is encapsidated within the mature virus. In addition, similar to VH1, F10 is a dual specificity kinase. Together, these proteins constitute potentially complementary activities that regulate a variety of functions, indicative of a certain degree of viral economy. For example, like VH1, F10 has poxviral substrates, A14 and A17, as well as thus far ill-defined host substrates . We have over expressed F10 in baculovirus and purified to >90% homogeneity as judged by SDS-PAGE, with a yield of 2 mg/L of cell culture. The protein is active as judged by its ability to phosphorylate casein. Crystallization trials are in progress.

P28: The 28-kDa RING zinc finger-containing protein (p28) was first identified in ectromelia virus (EV) genome. It has been shown that p28 is not necessary for virus replication in cell cultures but is crucial for EV pathogenicity in mice. Consistently, p28 is highly conserved in variola virus while is disrupted in vaccinia virus, which is adapted to cell cultures. The molecular function of p28 in poxvirus virulence *in vivo* is still unclear. Studies of Shope Fibroma virus and EV showed that p28 is localized to viral replication factories and involved in the inhibition of viral infection-induced apoptosis. p28-knock out virus was unable to replicate in macrophage cultures. Recent studies have shown that the p28 RING domain possesses ubiquitin ligase activities in biochemical assays and in mammalian cell cultures. Given that the importance of p28 in EV virulence and the significant sequence conservation among the RING domains of EV, variola virus and other wild type orthopoxviruses, p28 is a potential antiviral drug target. As it has been shown that p28 is not soluble in bacterial cultures while could be expressed by insect cell cultures, we tried to express GST-p28 with baculoviral system. It turned out to be soluble. Now Joma is trying to purify the protein.

F1L: To evade immunity, poxviruses have developed numerous strategies to interfere host cell's signaling pathways. F1L is an anti-apoptotic protein which anchors at mitochondria via its C terminal hydrophobic domain. F1L-deficient virus has been shown to be more susceptible to apoptosis. However, over-expression of Bcl-2 could rescue cells infected with VAVC missing F1L. Although both F1L and Bcl-2 possess similar anti-apoptotic functions, sequence analysis shows insignificant homology between these two proteins. Therefore, comparing the structures of F1L and Bcl-2 would help us to understand the mechanism of interactions between Bcl-2 and other Bcl-2 family

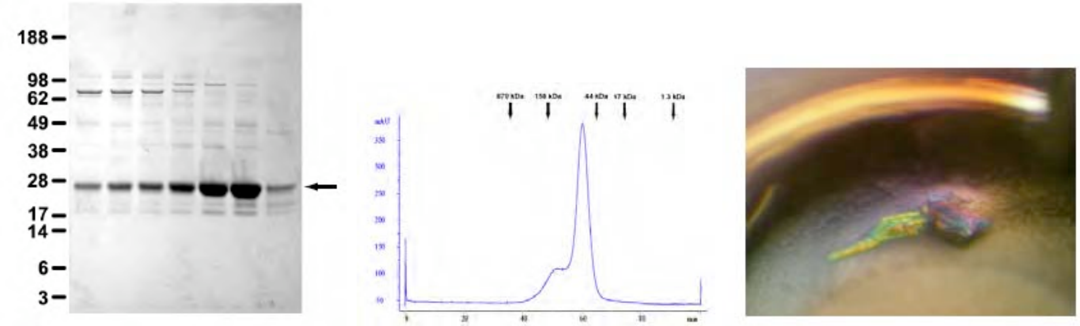


Figure 6: *GST-tagged F1L with a C-terminal truncation was expressed in E.coli strain BL21(DE3). The fusion protein was loaded onto glutathione sepharose beads and then cleaved by thrombin. Cleaved F1L was eluted and further purified using a Superdex 200 column. Crystals were grown from a factorial screen.*

1.4.8 Structure of the immunomodulatory protein, N1L: N1L is a small 14kDa protein, highly conserved among poxviruses, with 94% sequence identity between *Vaccinia* and *Variola* orthologs. N1L is considered one of the most potent virulent factors based on the attenuated phenotype of the recombinant mutant *Vaccinia* virus (Kotwal et al., 1989). The N1L gene was amplified from genomic DNA of *Vaccinia* Western Reserve and Cowpox Brighton Red (a gift from Dr. D.J. Pickup, Duke University). We determined the crystal structure of N1L (Aoyagi et al., 2006); in Press; Figure 7; **Appendix 11**), which reveals an unexpected but striking resemblance to host apoptotic regulators of the B cell lymphoma-2 (Bcl-2) family. Although N1L lacks detectable Bcl-2 homology (BH) motifs at the sequence level, we show that N1L binds with high affinity to the BH3 peptides of pro-apoptotic Bcl-2 family proteins in vitro, consistent with a role for N1L in modulating host antiviral defenses.

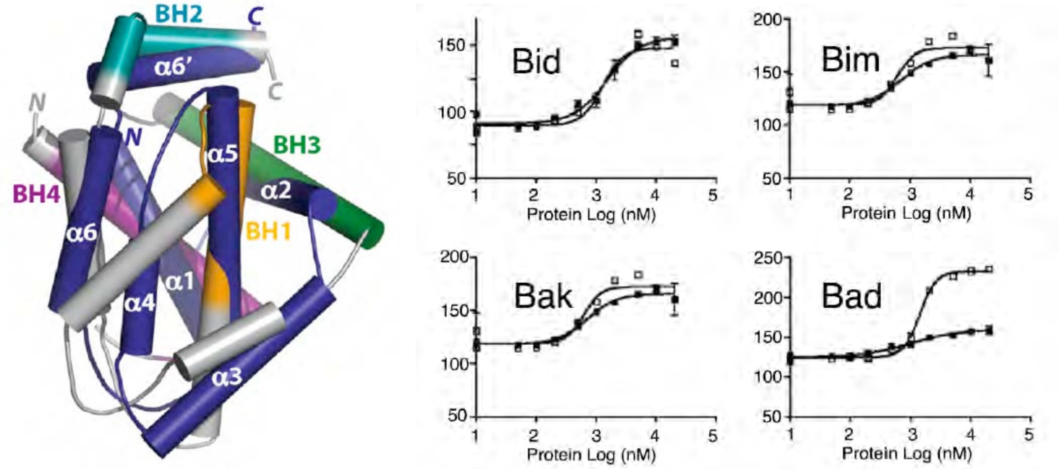


Figure 6: *(Left panel) Superposition of Vaccinia N1L (navy) and Bcl-X$_L$ (gray; 1MAZ. N1L helices are labeled. Functionally important BH regions of Bcl-X$_L$ are colored in magenta (BH4), green (BH3), orange (BH1) and cyan (BH2). (Right Panel) N1L binds to BH3 peptides. Fluorescence polarization plots of FITC-labeled BH3 domains (Bid, Bim, Bak and*

Bad) in the presence of varying concentrations of N1L (☐) or Bcl-XL (☐).

1.4.9 Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis The Chlamydia protein CADD (Chlamydia protein associating with death domains) has been implicated in the modulation of host cell apoptosis via binding to the death domains of tumor necrosis factor family receptors. Transfection of CADD into mammalian cells induces apoptosis. We determined the crystal structure of CADD (Schwarzenbacher et al., 2004b); Figure 7; **Appendix 14**, which reveals a dimer of seven-helix bundles. Each bundle contains a di-iron center adjacent to an internal cavity, forming an active site similar to that of methane mono-oxygenase hydrolase. We further showed that CADD mutants lacking critical metal-coordinating residues are substantially less effective in inducing apoptosis but retain their ability to bind to death domains. We concluded that CADD is a novel redox protein toxin unique to Chlamydia species and propose that both its redox activity and death domain binding ability are required for its biological activity.

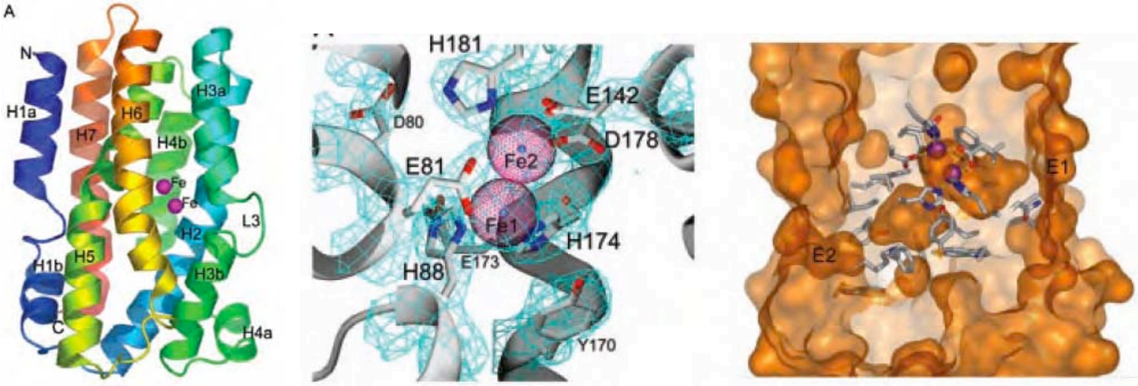


Figure 7: *(Left) The structure of CADD, rainbow color-coded from N terminus (blue) to C terminus (red), with helices H1–H7, the two iron ions, and loop L3 labeled. (Middle and Right) Active site analysis. (Middle) The di-iron site depicted in ball and stick format. The electron density map is contoured at 1.5σ. Right, close-up of the CADD molecule in a transparent surface representation (orange) showing the internal cavities, the di-metal site (purple spheres), and surrounding residues in ball and stick format.*

1.4.10 Crystal structure and functional analysis of PqqC from Klebsiella pneumoniae: The biosynthesis of pyrroloquinoline quinone (PQQ), a vitamin and redox cofactor of quinoprotein dehydrogenases, is facilitated by an unknown pathway that requires the expression of six genes, pqqA to -F. PqqC, the protein encoded by pqqC, catalyzes the final step in the pathway in a reaction that involves ring cyclization and eight-electron oxidation of 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-d icarboxylic-acid to PQQ. We determined the crystal structures of PqqC and its complex with PQQ, and determined the stoichiometry of H_2O_2 formation and O_2 uptake during the reaction (Magnusson et al., 2004; Schwarzenbacher et al., 2004a) (**Appendix 15**). The PqqC structure(s) reveals a compact seven-helix bundle that provides the scaffold for a positively charged active site cavity. Product binding induces a large conformational change, which results in the active site recruitment of amino acid side chains proposed to play key roles in the catalytic mechanism. PqqC is unusual in that it transfers redox equivalents to molecular oxygen without the assistance of a redox active metal or cofactor. The structure of the enzyme-product complex shows additional electron density next to R179 and C5 of PQQ, which can be modeled as O2 or H2O2, indicating a site for oxygen binding. We propose a reaction sequence that involves base-catalyzed cyclization and a series of quinone-quinol tautomerizations that are followed by cycles of O_2/H_2O -mediated oxidations.

1.5 NMR based structural characterization of virulence factors: The goal of Dr. Pellecchia's laboratory within this project was to provide support for the determination of the structures of key virulence factors using NMR spectroscopy. A group of bacterial genes homologous to the human Ubiquitin-like protease (Ulp) or SUMO-specific protease (SUMOylase) have been identified by bioinformatics methods in Dr. Godzik's laboratory. These proteins are also related to the *Yersinia* virulence factor YopP. Dr. Pellecchia focused his efforts on a particular protein construct from *Salmonella typhimurium* called Virulase ST. In unpublished work, Dr. Reed's laboratory has established that much like YopP, Virulase ST regulates apoptosis and inflammation in infected host cells, presumably via the NF-κB pathway. Recombinant Virulase ST (145-326) was produced from a pET-19b (Novagen) plasmid construct containing the nucleotide sequence for the catalytic domain fused to an *N*-terminal poly-His tag. Unlabeled protein was expressed in *E. coli* BL21 in LB media at 37°C, with an induction period of 3-4 hours with 1 mM IPTG. ^{15}N-labeled protein was similarly produced, with growth occurring in M9 media supplemented with 0.5 g/L $^{15}NH_4Cl$. Double $^{13}C/^{15}N$-labeled protein as well as triple labeled $^{2}H/^{15}N/^{13}C$ protein were similarly produced in M9 media supplemented with ^{13}C-glucose (2 g/L) and $^{2}H_2O$ (70%), respectively. Soluble protein was purified over a Hi-Trap chelating column (Amersham, Pharmacia).

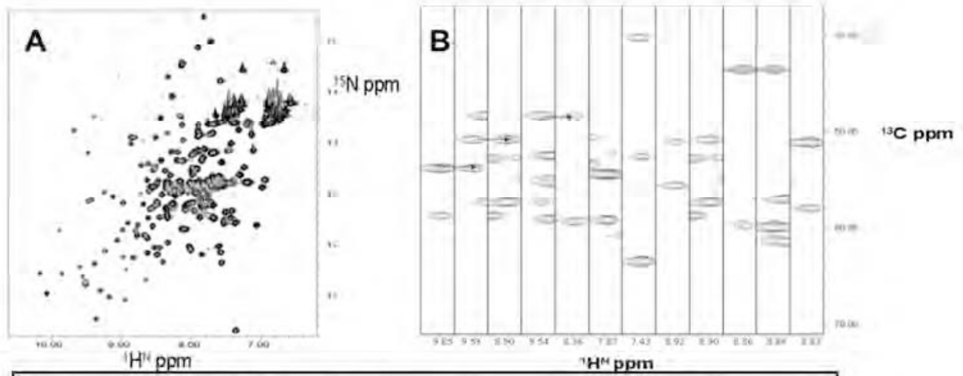


Figure 8 **A**) 2D $[^{1}H, ^{15}N]$ HSQC spectrum of the catalytic domain of Virulase ST acquired on a 1 mM sample in phosphate buffer, pH = 7.2, 50 mM each Arg/Glu, 100 mM NaCl. The spectrum was acquired with ns=16 at 20° C on a 600 MHz Avance Bruker instrument. **B**) Typical $^{13}C^{\alpha}/^{1}H^{N}$ strips taken at different ^{15}N chemical shifts from a 3D HNCA experiment measured with a triple $^{2}H/^{13}C/^{15}N$ labeled sample.

Unfortunately, the protein is not sufficiently long lived (2-3 weeks) for NMR assignments, and it tends to aggregate very rapidly (hours). In order to increase the stability of the domain, a number of different conditions were tested including temperature, pH, different detergents (TRITON and NP-40, both at 0.1%) and salts. Conditions that led to samples that are stable for ~ 3-7 days included a second step purification (ion-exchange purification with a MonoQ (Amersham, Pharmacia) column), pH = 7.2, 100 mM NaCl, and 50 mM each of arginine and glutamic acid. Because 3-7 days is still too short for a complete set of NMR experiments, several samples were prepared. 2D [^{1}H,^{15}N] HSQC and TROSY-type experiments were carried out on a 600 MHz spectrometer at 20°C and 30°C. A typical 2D [^{15}N, ^{1}H] HSQC spectrum of the is reported in Figure 8A. The number of peaks and the dispersion are indicative of a folded monomeric protein. Chemical shift dispersion in the $^{13}C^{\alpha}$ (and $^{13}C^{\beta}$) from initial triple resonance experiments (Figure 8B) suggests a mixed αβ secondary structure, although there is probably a flexible region as well. Therefore, while additional work is needed to complete the acquisition of a minimal data set for structural determination, samples that appear well behaved for high resolution studies have been obtained. The isotopically labeled samples and the preliminary NMR data collected lay the foundation for a detailed structure determination project.

Task 2: Collect expression vectors and purified proteins into a library suitable for use by other interested groups, and post the information on our website.

This task was performed for *B. anthracis* and variola/vaccinia virus; target selection and experimental updates were done on a monthly basis in the light of new cloning, expression and structural data. The final status for B. anthracis is summarized below. Expression vectors for vaccinia virus proteins are described in Task 1. We will make this information publicly available if it is deemed appropriate by USAMRMC.

Summary of cloning, expression and purification of novel pX01 proteins:

pXO1-1 has a single transmembrane region and could only be expressed as insoluble protein. Initial trials using high concentration of detergent TritonX-100 extraction failed to produce significant amount of soluble protein. Expression of the fragment excluding the predicted transmembrane also produce insoluble inclusion.

pXO1-37 (Acetyltransferase) His tagged full-length pXO1-37 (1-193) was solubly overexpressed by E. coli at 30°C. Previous instability problem upon concentrating to higher concentration is solved by adding 100 mM DTT to the protein solution after Ni-column purification. Crystallization setups have begun

pXO1-47 (Transcription Activator of multidrug-efflux) His tagged full-length pXO1-47 (1-201) was overexpressed in inclusion bodies . Varying expression conditions did not lead to soluble protein. pXO1-47 was purified under denatured condition by Ni-column and refolded as soluble protein. DSC experiment is underway to demonstrate correct folding.

pXO1-87 and pXO1-99 were expressed, but proved to be difficult to purify. Both proteins were co-purified with a 60 kDa protein, which is suspected to be a heat shock protein or chaperonin. High resolution columns, superdex200HR gel filtration, monoS and monoQ column could not separate the contaminants. Mg^{2+}-ATP has been shown to enhance dissociation of E. coli chaperonin from proteins with large hydrophobic surface area exposed. It will be used in the immediate future for the pXO1-99 and 87 protein purification.

pXO1-97 was cloned and gave soluble protein, and structural analysis by NMR is in progress.

pXO1-104 His tagged full-length pXO1-104 (1-61) was overexpressed as inclusion body. Other conditions have been tried to make it expressed solubly without success. Refolding experiments are underway.

pXO1-109/PagR Cloning and soluble expression; crystallization trials in progress.

pXO1-111 (homologous to PA domain 4). Cloning and soluble expression; crystallization trials in progress.

pXO1-116 Cloning unsuccessful so far.

pXO1-117 and 143 cloning successful but no expression in *E. coli*.

PX01-118 (and pX02-61) have been crystallized and their structures determined (see **Appendix 3**)

pXO1-121 His tagged full-length pXO1-121 (1-58) was overexpressed as inclusion body. Other conditions have been tried to express it solubly, without success. Refolding is underway.

pXO1-125 – cloning and expression successful – protein is insoluble and could not be refolded.

Cloning of all the following target genes as full-length proteins has been completed, and expression trials are in progress. All the genes are now subcloned into the bacterial expression vector, pET28a: **pXO1-96**, 274 residues, homologue to putative transposase; **pXO1-103**, 317 residues, homologue to site-specific recombinase; **pXO1-105**, 67 residues, homologue to regulators of stationary/sporulation gene expression; **pXO1-126**, 151 residues, homologue to uncharacterized ACR ML0644; **pXO1-130**, 237 residues, predicted periplasmic or secreted protein. **pXO1-04**, **pXO1-07**, **pXO1-10**, **pXO1-32**, **pXO1-90**, **pXO1-94**, **pXO1-98**, a truncated form of **pXO1-98**, **pXO1-117**, **pXO1-124**, **pXO1-127**, and **pXO1-132**.

pXO1-37 (Acetyltransferase) His tagged full-length pXO1-37 (1-193) was solubly overexpressed by E. coli at 30°C. Previous instability problem upon concentrating to higher concentration is solved by adding 100 mM DTT to the protein solution after Ni-column purification. Crystallization setups have begun

pXO1-47 (Transcription Activator of multidrug-efflux) His tagged full-length pXO1-47 (1-201) was overexpressed in inclusion bodies . Varying expression conditions did not lead to soluble protein. pXO1-47 was purified under denatured condition by Ni-column and refolded as soluble protein. DSC experiment is underway to demonstrate correct folding.

pXO1-87 and pXO1-99 were expressed, but proved to be difficult to purify. Both proteins were co-purified with a 60 kDa protein, which is suspected to be a heat shock protein or chaperonin. High resolution columns, superdex200HR gel filtration, monoS and monoQ column could not separate the contaminants. Mg^{2+}-ATP has been shown to enhance dissociation of E. coli chaperonin from proteins with large hydrophobic surface area exposed. It will be used in the immediate future for the pXO1-99 and 87 protein purification.

pXO1-97 was cloned and gave soluble protein, and structural analysis by NMR is in progress.

pXO1-104 His tagged full-length pXO1-104 (1-61) was overexpressed as inclusion body. Other conditions have been tried to make it expressed solubly without success. Refolding experiments are underway.

pXO1-109/PagR Cloning and soluble expression; crystallization trials in progress.

pXO1-111 (homologous to PA domain 4). Cloning and soluble expression; crystallization trials in progress.

pXO1-116 Cloning unsuccessful so far.

pXO1-117 and 143 cloning successful but no expression in *E. coli*.

PX01-118 (and pX02-61) have been crystallized and their structures determined (see **Appendix 3**)

pXO1-121 His tagged full-length pXO1-121 (1-58) was overexpressed as inclusion body. Other conditions have been tried to express it solubly, without success. Refolding is underway.

pXO1-125 – cloning and expression successful – protein is insoluble and could not be refolded.

Cloning of all the following target genes as full-length proteins has been completed, and expression trials are in progress. All the genes are now subcloned into the bacterial expression vector, pET28a:

pXO1-96, 274 residues, homologue to putative transposase;
pXO1-103, 317 residues, homologue to site-specific recombinase;

pXO1-105, 67 residues, homologue to regulators of stationary/sporulation gene expression;
pXO1-126, 151 residues, homologue to uncharacterized ACR ML0644;
pXO1-130, 237 residues, predicted periplasmic or secreted protein.

pXO1-109 (PagR) expressed in E. coli and purified.

pXO1-110 (PA) expressed in E. coli and purified;
 604-735 (domain IV) expressed, partially purified
 597-735 (domain IV) expressed, purified
 588-735 (domain IV) expressed, partially purified

pXO1-107 (LF) expressed in E. coli and purified; catalytic mutants E687C and E786A expressed and purified.
 263-776 (domains II-IV) expressed, purified and crystallized

pXO1-119 (AtxA) full-length and 1-393 expressed and purified;
 1-141 and 1-160 (putative DNA binding domain) expressed, insoluble;
 141-475, 162-475, 141-393, 162-393 (putative regulatory domains);
 388-475 expressed, soluble, precipitates during purification

pXO1-138 (PagR homolog) expressed, soluble

pXO2-53 (AcpB) expressed and purified
pXO2-64 (AcpA) expressed and purified (low yield << 1 mg/l)

The following gene products of unknown function have been cloned expressed and purified: **pXO1-04**, **pXO1-07**, **pXO1-10**, **pXO1-32**, **pXO1-90**, **pXO1-94**, **pXO1-98**, a truncated form of **pXO1-98**, **pXO1-117**, **pXO1-124**, **pXO1-127**, and **pXO1-132**.

pXO1-1, pXO1-15, pXO1-125, pXO1-117, pXO1-128 and pXO1-143 were expressed in *E. coli* as insoluble proteins. Refolding with arginine as refolding buffer solubilized the proteins but precipitations occurred during the removal. **pXO1-87 and pXO1-99** could be purified but as soluble aggregates, which precipitate at high concentration.

Expression and purification of AtxA and its homologs on pXO2, AcpA and AcpB

Full-length AtxA was expressed with or without a histidine-tag fusion and purified by Ni affinity, heparine-sepharose and/or anion-exchange, and gel filtration chromatography. Yields are around 2 mg/liter of cell culture. The presence of up to five species, partially separable by heparin-sepharose affinity chromatography, was evident. Native PAGE evidence at μM to mM concentration shows that AtxA interacts with DNA, since a band corresponding to DNA cannot be detected as the concentration of AtxA increases, but a stable specific complex could not yet be characterized, possibly due to the relatively high concentration of protein or the lack of a specific site on the DNA sequence used, a 300 bp stretch upstream of the transcriptional start site of the pag gene. Current work includes further separation of the above-mentioned AtxA species, determining whether they are stable or in slow exchange with each other, and whether this affects binding to DNA. Near-future plans are the characterization of the binding to DNA sequences from the promoters of other AtxA-regulated genes using radioactively labelled DNA, which will allow work at or near the protein-DNA dissociation constant, which is as yet undetermined but usually expected in the nM range.

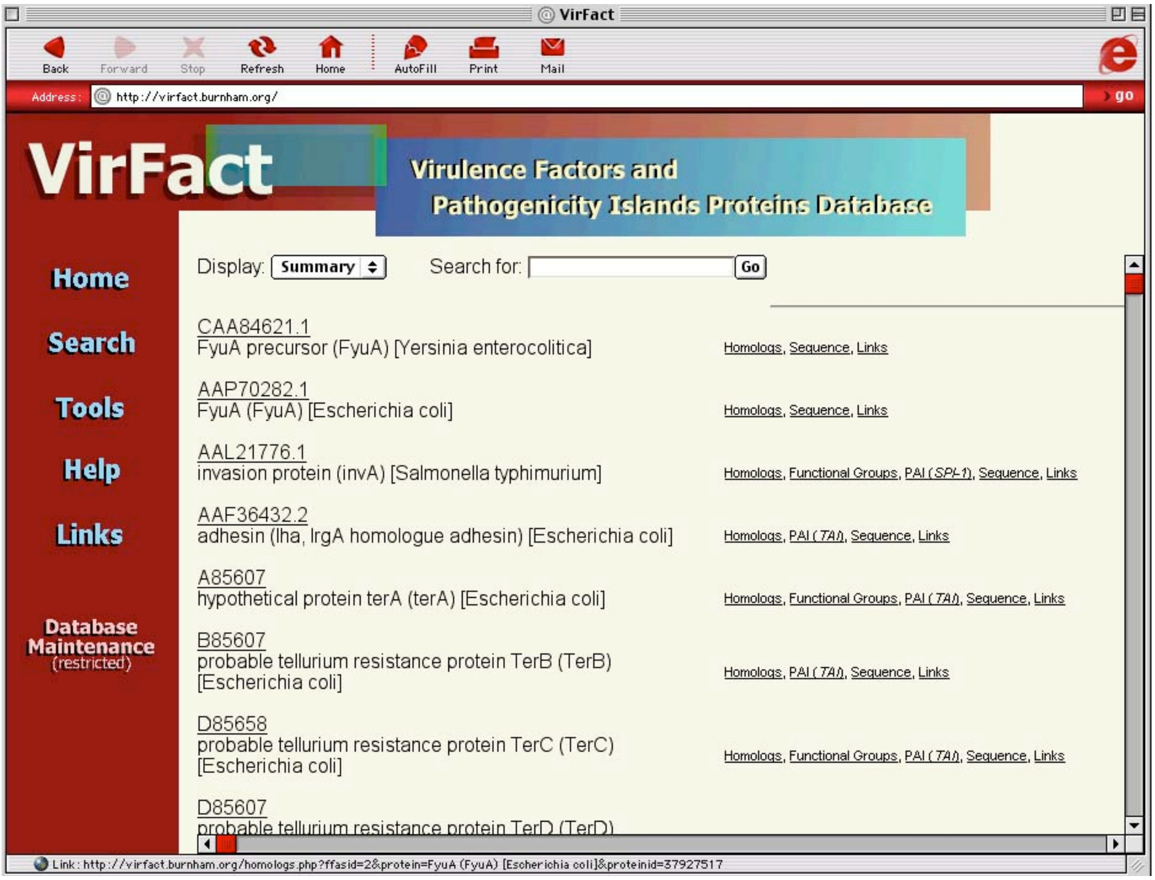
AcpB waas expressed as a histidine-tag fusion and purified with similar results. AcpA appears to be toxic to E.coli cells as their growth is significantly slowed down when transformed with a plasmid encoding the histidine-tagged protein, and yields were therefore an order of magnitude lower. Current work focuses on the cloning, expression and purification of native (untagged) AtxA and AtxB and future plans will include the characterization of their binding to DNA, similar to AtxA.

Collagen binding protein BA5258 of *B. anthracis*

B. anthracis, similar to other Gram positive bacteria, attaches to the host via cell-wall-anchoring proteins. Two of such protein from *B. anthracis* were characterized by Xu *et al* (2004), namely BA0871 and BA5258. These two proteins have sequence homology to CNA, a cell wall-anchored collagen adhesin of *S. aureus*. The full length BA5258, excluding the leader sequence, has been cloned into a *E. coli* expression vector. It can be expressed and purified to a final yield of 10 mg/L culture. The protein is extremely soluble and resistant to limited proteolysis with trypsin, elastase, and chymotrypsin. Crystallization trials of the protein by itself and with a collagen peptide are in progress, and small but promising protein crystals have been obtained.

Task 3: Develop a computational database of virulence-related genes

Bioinformatics and Target Selection. The main focus of the bioinformatics part of the grant was the development of an annotated collection of virulence factors. To this end we developed the VirFact database (http://virfact.burnham.org) (see **Appendix 12**), which contains information on microbial virulence factors and pathogenicity islands (PAIs) from major pathogens. The database initially contained information manually collected from literature, and then combined this with results obtained by genome context analysis and distant homology recognition. The database can be browsed by virulence factor, PAI or organism name. The annotations, including multiple alignments of proteins homologous to virulence factors, genomic context, models of three dimensional structures (if available) are presented using graphical web interface and standard visualization tools. The VirFact can also be used as a tool to recognize the presence of homologs of known virulence factors in the genome delivered by the user. For instance application of VirFact to *Francisella tularensis* genome allowed us to recognize over 50 known virulence factors in this genome.



http://virfact.burnham.org/

CAA84621.1
FyuA precursor (FyuA) [Yersinia enterocolitica] Homologs, Sequence, Links

AAP70282.1
FyuA (FyuA) [Escherichia coli] Homologs, Sequence, Links

AAL21776.1
invasion protein (invA) [Salmonella typhimurium] Homologs, Functional Groups, PAI (*SPI-1*), Sequence, Links

AAF36432.2
adhesin (Iha, IrgA homologue adhesin) [Escherichia coli] Homologs, PAI (*TAI*), Sequence, Links

A85607
hypothetical protein terA (terA) [Escherichia coli] Homologs, Functional Groups, PAI (*TAI*), Sequence, Links

B85607
probable tellurium resistance protein TerB (TerB) [Escherichia coli] Homologs, PAI (*TAI*), Sequence, Links

D85658
probable tellurium resistance protein TerC (TerC) [Escherichia coli] Homologs, Functional Groups, PAI (*TAI*), Sequence, Links

D85607
probable tellurium resistance protein TerD (TerD)

Link: http://virfact.burnham.org/homologs.php?ffasid=2&protein=FyuA (FyuA) [Escherichia coli]&proteinid=37927517

We used several of the annotation tools developed in Dr. Godzik's group for a detailed analysis of anthrax virulence plasmids. Using a combination of advanced bioinformatics tools, including context analysis, distant homology and fold recognition, we re-annotated the predicted open reading frames on the pXO1 plasmid, most of which were described as proteins of unknown function in previous analyses. Thanks to improved annotation tools we significantly enhanced the annotation of the pXO1 plasmid, bringing the total number of ORFs with some level of functional annotation from 48 to over 100. The new results also clearly show the mosaic nature of pXO1 and give tantalizing hints about the origin of anthrax virulence. The highlights of the new finding are two type IV secretion system-like clusters present on the pathogenicity island of the pXO1 plasmid, as well as at least three clusters related to DNA processing. This work is being prepared for publication **(Appendix 1).** Similar annotation of the pX02 plasmid as well as pathogenic islands of several bacteria from the Streptococcus group are now in preparation.

Other relevant work: The survival of human pathogens depends on their ability to modulate defence pathways in human host cells. This was thought to be attained mainly by pathogen specific "virulence factors". However, pathogens are increasingly being discovered that use distant homologs of the human regulatory proteins as virulence factors. We analyzed several cases of this approach, with a particular focus on virulence proteases. The analysis reveals clear cases of bacterial proteases mimicking the specificity of their human counterparts, such as strong similarities in their active and/or binding sites. With more sensitive tools for distant homology recognition, we could expect to discover many more such cases. This work has been published: (Sikora et al., 2005); **Appendix 13**.

Task 4: Form a consortium of groups with similar interests who are funded from other sources, developing a common website containing target selections and project status.

The issue of developing a web-site was suggested by Dr. Millard to be potentially politically sensitive. If the DoD would like us to create such a website we will be happy to oblige.

Key Research Accomplishments

- Development of the VirFact database (http://virfact.burnham.org) of virulence factors

- In-depth annotation of the anthrax virulence plasmid, and the identification of novel domains.

- Identification, crystal structure determination and in vivo characterization of B. anthracis sensor domains.

- Crystal structures and functional characterization of two B. anthracis endolysins.

- Crystal structure of two B. anthracis endolysins with potential as antibiotics.

- Crystal structure of anthrax Protective Antigen in complex with its host receptor

- 6 Crystal structures of anthrax Lethal Factor in complex with inhibitors; many more still being analyzed.

- Successful expression and/or cloning and of more than 50 proteins and domain fragments from the B. anthracis.

- Successful expression and/or cloning of 4 variola virus virulence factors; crystal structure of one of them (N1L) reveals a fold related to anti-apoptoticproteins.

- Crystal of the SARS virus spike protein in complex with a neutralizing antibody.

- Crystal structure and functional analysis of PqqC from Klebsiella .

- Structure and functional analysis of the Chlamydia protein CADD, revealing a redox enzyme that modulates host cell apoptosis

Reportable Outcomes

VirFact database (http://virfact.burnham.org) of virulence factors

Published manuscripts:

1. Aoyagi, M., Zhai, D., Jin, C., Aleshin, A., Stec, B., Reed, J. C., and Liddington, R. C. (2006). Vaccinia virus N1L protein resembles a B cell lymphoma-2 (Bcl-2) family protein. Protein Sci In press.

2. Grynberg, M., and Godzik, A. (2004). NERD: a DNA processing-related domain present in the anthrax virulence plasmid, pXO1. Trends Biochem Sci 29, 106-110.
Low, L. Y., Yang, C., Perego, M., Osterman, A., and Liddington, R. C. (2005). Structure and lytic activity of a Bacillus anthracis prophage endolysin. J Biol Chem 280, 35433-35439.

3. Panchal, R. G., Hermone, A. R., Nguyen, T. L., Wong, T. Y., Schwarzenbacher, R., Schmidt, J., Lane, D., McGrath, C., Turk, B. E., Burnett, J., et al. (2004). Identification of small molecule inhibitors of anthrax lethal factor. Nat Struct Mol Biol 11, 67-72.

4. Santelli, E., Bankston, L. A., Leppla, S. H., and Liddington, R. C. (2004). Crystal structure of a complex between anthrax toxin and its host cell receptor. Nature 430, 905-908.

5. Sikora, S., Strongin, A., and Godzik, A. (2005). Convergent evolution as a mechanism for pathogenic adaptation. Trends Microbiol 13, 522-527.

6. Turk, B. E., Wong, T. Y., Schwarzenbacher, R., Jarrell, E. T., Leppla, S. H., Collier, R. J., Liddington, R. C., and Cantley, L. C. (2004). The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. Nat Struct Mol Biol 11, 60-66.

7. White, A. K., Hoch, J. A., Grynberg, M., Godzik, A., and Perego, M. (2006). Sensor domains encoded in Bacillus anthracis virulence plasmids prevent sporulation by hijacking a sporulation sensor histidine kinase. J Bacteriol 188, 6354-6360.

8. Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004). Quinone biogenesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. Proc Natl Acad Sci U S A 101, 7913-7918.

9. Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Reed, J. C., and Liddington, R. C. (2004a). Crystal structure of PqqC from Klebsiella pneumoniae at 2.1 A resolution. Proteins 56, 401-403.

10. Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Robinson, H., Yuan, H., Bossy-Wetzel, E., Reed, J. C., and Liddington, R. C. (2004b). Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. J Biol Chem 279, 29320-29324.

11. Forino, M., Johnson, S., Wong, T. Y., Rozanov, D. V., Savinov, A. Y., Li, W., Fattorusso, R., Becattini, B., Orry, A. J., Jung, D., et al. (2005). Efficient synthetic inhibitors of anthrax lethal factor. Proc Natl Acad Sci U S A 102, 9499-9504.

Manuscripts in preparation:

1. Adrian Tkacz, Leszek Rychlewski and Adam Godzik "VirFact: a relational database of virulence factors and pathogenicity islands (PAIs)"

2. Marcin Grynberg, Iddo Friedberg, Marc Robinson-Rechavi, and Adam Godzik "Surprising connections: in-depth analysis of the *Bacillus anthracis* pXO1 Plasmid"

3. Gudrun R. Stranzl, Marcin Grynberg, Chandra La Clair, Dorinda Shoemaker, Robert Schwarzenbacher, Eugenio Santelli, Adam Godzik, Marta Perego, Robert C. Liddington "Crystal structure of virulence plasmid-encoded sensor domains inhibiting sporulation in *Bacillus anthracis"*

4. Lieh Yoon Low, Chen Yang, Andrea Osterman, and Robert Liddington. "Structure of a GH-25 N-acetylmuramidase from Bacillus anthracis prophage LambdaBa04 at 1.4 Å" (In preparation)

Reagents generated:

- Expression vectors and protocols for more than 60 virulence factors.

- Atomic coordinates and structure factors have been deposited in the Protein Data Bank for all of the structure described above.

Funding arising from these studies:

We developed the initial work on pX01-118, pX02-61 and AtxA funded by this grant into an in-depth structure-function study in a successful application for a Program Project grant from NIAID led by Dr. Liddington **(P01 AI55789) (2004-2009).**

Our work on the inhibitors of anthrax Lethal Factor played a large part in out successful application to NIAID to develop a novel class of inhibitors using in silico and NMR-based methods combined with crystallography **(U19 AI56385-01 to Dr. Alex Strongin, P.I. (2002-2006);** this has recently been renewed under new leadership of Dr. Pellecchia

(R01 AI059572 and U01 AI070494). Our general approach also led to the successful application for novel therapeutic treatments of smallpox **(U01 AI061139 - P.I. Dr. Alex Strongin)**

Conclusions

During the period of this grant, we (1) carried out cloning, expression and functional studies on B. anthracis plasmid-genome-encoded proteins; (2) extended this work to genome-encoded proteins; (3) extended our work to structural studies of virulence factors to other bacteria and viruses, including Variola virus and SARS CoV virus. We successfully cloned and expressed more than 60 new proteins, and determined more than 20 new LF-inhibitor complexes and 12 novel structures.

<u>**So what section**</u>: Post-exposure therapeutics do not exist for any of the major pathogens likely to be used in biowarfare or bioterrorism. Our work identifies and characterizes structurally and functionally key protein "virulence factors" from these organisms, allowing for the rational structure-based small molecule inhibitor design that can lead to the development of therapeutic drugs to treat anthrax, smallpox, plague and SARS.

References

Aoyagi, M., Zhai, D., Jin, C., Aleshin, A., Stec, B., Reed, J. C., and Liddington, R. C. (2006). Vaccinia virus N1L protein resembles a B cell lymphoma-2 (Bcl-2) family protein. Protein Sci *In press*.

Forino, M., Johnson, S., Wong, T. Y., Rozanov, D. V., Savinov, A. Y., Li, W., Fattorusso, R., Becattini, B., Orry, A. J., Jung, D., *et al.* (2005). Efficient synthetic inhibitors of anthrax lethal factor. Proc Natl Acad Sci U S A *102*, 9499-9504.

Grynberg, M., and Godzik, A. (2004). NERD: a DNA processing-related domain present in the anthrax virulence plasmid, pXO1. Trends Biochem Sci *29*, 106-110.

Hwang, W. C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W. A., and Liddington, R. C. (2006). Structural basis of neutralization by a human anti-SARS spike protein antibody, 80R. J Biol Chem.

Low, L. Y., Yang, C., Perego, M., Osterman, A., and Liddington, R. C. (2005). Structure and lytic activity of a Bacillus anthracis prophage endolysin. J Biol Chem *280*, 35433-35439.

Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004). Quinone biogenesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. Proc Natl Acad Sci U S A *101*, 7913-7918.

Panchal, R. G., Hermone, A. R., Nguyen, T. L., Wong, T. Y., Schwarzenbacher, R., Schmidt, J., Lane, D., McGrath, C., Turk, B. E., Burnett, J., *et al.* (2004). Identification of small molecule inhibitors of anthrax lethal factor. Nat Struct Mol Biol *11*, 67-72.

Santelli, E., Bankston, L. A., Leppla, S. H., and Liddington, R. C. (2004). Crystal structure of a complex between anthrax toxin and its host cell receptor. Nature *430*, 905-908.

Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Reed, J. C., and Liddington, R. C. (2004a). Crystal structure of PqqC from Klebsiella pneumoniae at 2.1 A resolution. Proteins *56*, 401-403.

Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Robinson, H., Yuan, H., Bossy-Wetzel, E., Reed, J. C., and Liddington, R. C. (2004b). Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. J Biol Chem *279*, 29320-29324.

Sikora, S., Strongin, A., and Godzik, A. (2005). Convergent evolution as a mechanism for pathogenic adaptation. Trends Microbiol *13*, 522-527.

Turk, B. E., Wong, T. Y., Schwarzenbacher, R., Jarrell, E. T., Leppla, S. H., Collier, R. J., Liddington, R. C., and Cantley, L. C. (2004). The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. Nat Struct Mol Biol *11*, 60-66.

White, A. K., Hoch, J. A., Grynberg, M., Godzik, A., and Perego, M. (2006). Sensor domains encoded in Bacillus anthracis virulence plasmids prevent sporulation by hijacking a sporulation sensor histidine kinase. J Bacteriol *188*, 6354-6360.

Appendices

Appendix 1: <u>Marcin Grynberg, Iddo Friedberg, Marc Robinson-Rechavi, and Adam Godzik</u> "Surprising connections: in-depth analysis of the *Bacillus anthracis* pXO1 Plasmid" (manuscript submitted)

Appendix 2: Grynberg, M., and Godzik, A. (2004). NERD: a DNA processing-related domain present in the anthrax virulence plasmid, pXO1. Trends Biochem Sci 29, 106-110.

Appendix 3: Gudrun R. Stranzl, Marcin Grynberg, Chandra La Clair, Dorinda Shoemaker, Robert Schwarzenbacher, Eugenio Santelli, Adam Godzik, Marta Perego, Robert C. Liddington "Crystal structure of virulence plasmid-encoded sensor domains inhibiting sporulation in *Bacillus anthracis"*

Appendix 4: Turk, B. E., Wong, T. Y., Schwarzenbacher, R., Jarrell, E. T., Leppla, S. H., Collier, R. J., Liddington, R. C., and Cantley, L. C. (2004). The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. Nat Struct Mol Biol 11, 60-66.

Appendix 5: Panchal, R. G., Hermone, A. R., Nguyen, T. L., Wong, T. Y., Schwarzenbacher, R., Schmidt, J., Lane, D., McGrath, C., Turk, B. E., Burnett, J., et al. (2004). Identification of small molecule inhibitors of anthrax lethal factor. Nat Struct Mol Biol 11, 67-72.

Appendix 6: Santelli E, Bankston LA, Leppla SH & Liddington RC. Crystal structure of a complex between anthrax toxin and its host cell receptor. ***Nature***. 2004 Aug 19;430(7002):905-8. Epub 2004 Jul 4.

Appendix 7: White, A. K., Hoch, J. A., Grynberg, M., Godzik, A., and Perego, M. (2006). Sensor domains encoded in Bacillus anthracis virulence plasmids prevent sporulation by hijacking a sporulation sensor histidine kinase. J Bacteriol 188, 6354-6360

Appendix 8: Low, L. Y., Yang, C., Perego, M., Osterman, A., and Liddington, R. C. (2005). Structure and lytic activity of a Bacillus anthracis prophage endolysin. J Biol Chem 280, 35433-35439.

Appendix 9: Lieh Yoon Low, Chen Yang, Andrea Osterman, and Robert Liddington. Structure of a GH-25 N-acetylmuramidase from Bacillus
anthracis prophage LambdaBa04 at 1.4 Å (in preparation)

Appendix 10: Hwang, W. C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W. A., and Liddington, R. C. (2006). Structural basis of neutralization by a human anti-SARS spike protein antibody, 80R. J Biol Chem. (published on-line 09/05/06).

Appendix 11: Aoyagi, M., Zhai, D., Jin, C., Aleshin, A., Stec, B., Reed, J. C., and Liddington, R. C. (2006). Vaccinia virus N1L protein resembles a B cell lymphoma-2 (Bcl-2) family protein. Protein Sci In press.

Appendix 12: Adrian Tkacz, Leszek Rychlewski and Adam Godzik "VirFact: a relational database of virulence factors and pathogenicity islands (PAIs)" (mansucript in preparation)

Appendix 13: Sikora, S., Strongin, A., and Godzik, A. (2005). Convergent evolution as a mechanism for pathogenic adaptation. Trends Microbiol 13, 522-527.

Appendix 14: Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Robinson, H., Yuan, H., Bossy-Wetzel, E., Reed, J. C., and Liddington, R. C. (2004b). Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. J Biol Chem *279*, 29320-29324.

Appendix 15: Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004). Quinone biogenesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. Proc Natl Acad Sci U S A *101*, 7913-7918; and Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Reed, J. C., and Liddington, R. C. (2004a). Crystal structure of PqqC from Klebsiella pneumoniae at 2.1 A resolution. Proteins *56*, 401-403.

Appendix 16: Forino, M., Johnson, S., Wong, T. Y., Rozanov, D. V., Savinov, A. Y., Li, W., Fattorusso, R., Becattini, B., Orry, A. J., Jung, D., et al. (2005). Efficient synthetic inhibitors of anthrax lethal factor. Proc Natl Acad Sci U S A 102, 9499-9504.

Surprising connections: in-depth analysis of the *Bacillus anthracis* pXO1 plasmid

Marcin Grynberg[1,2], Iddo Friedberg[1], Marc Robinson-Rechavi[3], and Adam Godzik[1,3,4]

[1]Bioinformatics and Systems Biology, The Burnham Institute, La Jolla, CA 92037, USA; [2]Institute of Biochemistry and Biophysics, PAS, Pawinskiego 5A, 02-106 Warsaw, Poland; [3]Joint Center for Structural Genomics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Keywords: anthrax, *Bacillus anthracis*, pathogenicity, virulence, pXO1, type IV secretion system, type IV pilus assembly system, context analysis, distant homology, ArsR, SmtB, regulators.

[3]Corresponding author.

E-mail: adam@burnham.org;

PHONE +1 (858) 646 3168

FAX: +1 (858) 713 9930.

ABSTRACT

Anthrax disease is caused by a bacterium *Bacillus anthracis*. Its virulence has been associated with two plasmids, pXO1 and pXO2. Using a combination of advanced bioinformatics tools, including context analysis, distant homology and fold recognition, we have re-annotated the predicted open reading frames on the pXO1 plasmid, most of which were described as proteins of unknown function in previous analyses. Thanks to improved annotation tools we significantly enhanced the annotation of the pXO1 plasmid, bringing the total number of ORFs with some level of functional annotation from 48 to over 100. The new results also clearly show the mosaic nature of pXO1 and give tantalizing hints about the origin of anthrax virulence. The highlights of the new finding are two type IV secretion system-like clusters present on the pathogenicity island of the pXO1 plasmid, as well as at least three clusters related to DNA processing.

Supplemental material available online at http://bioinformatics.burnham.org/pXO1.

INTRODUCTION

Anthrax is a disease primarily affecting herbivores but also sporadically attacking other mammals, including humans. Anthrax is known since antiquity and the quest for an effective treatment of anthrax is closely related to the birth of modern microbiology (Pasteur 1881). More recent work concentrated mostly on the anthrax toxin, leading to extensive structural and functional analysis of its components (for a review see (Turnbull 2002)). However, until recently the general level of interest in anthrax was limited, since it is not a major threat to human health. An era of more intensive work on *B.anthracis* has started since anthrax was adopted by military as a biological weapon, resulting in a threat of large scale anthrax outbreaks. These threats were kept alive by several large scale incidents, and more recently the threat of anthrax as a bioterror weapon.

At the same time, the origin and mechanism of *B.anthracis* virulence are very interesting on their own. Only very few *B.anthracis* virulence related proteins were studied in detail, among them the toxins (PagA, LEF, CyaA), cell envelope and germination genes (Cap, S-layer and Ger proteins), and the regulatory mechanisms triggering the virulence (Fouet and Mesnage 2002; Lacy and Collier 2002) and citations therein). The sequencing of the *B.anthracis* genome (Okinaka et al. 1999; Pannucci et al. 2002; Read et al. 2003) especially in the context of other *Bacilli* genome projects, highlighted the complex and little understood mechanism of its virulence (Koehler 2002). The *B.anthracis* genome consists of a single chromosome and two virulence associated megaplasmids, pXO1 and pXO2 (Okinaka et al. 1999; Read et al. 2003). The two plasmids together convey the pathogenic phenotype and are responsible for most of the difference between *B.anthracis* and its relatives with different pathogenicity profiles, such as *B.cereus* or *B.thuringiensis*. However, little is known about most proteins encoded by the two plasmids and only a few have been studied by experiment and shown to be directly involved in virulence. Most pXO1 and pXO2 proteins have no obvious sequence similarity to any other known genes. Therefore, the interest in *B.anthracis* pathogenicity transcends its immediate applications in bioterrorism and human health, and bears on fundamental questions of how novel and complex lifestyles, such as pathogenicity, can evolve.

Several earlier works focused on bioinformatic analysis of the anthrax genome and plasmids, often in the context of related organisms (Ariel et al. 2002 2002; Ariel et al. 2003; Rasko et al. 2004). These studies confirmed close relations between *B.anthracis*, *B.thuringiensis* and *B.cereus*, and identified previously unknown features of the virulence related plasmids, pXO1, pBtoxis and pBc10987, respectively. However, a vast majority of pXO1 genes remain uncharacterized, both in terms of their function and origin. A possible reason for this apparent novelty of pXO1 genes is that pathogenic plasmid encoded genes evolve rapidly and often bear little sequence similarity to their homologs from other species, hampering the detection of homology with most tools of sequence analysis. In this study we take advantage of recent improvements in super-sensitive tools for distant homology recognition. These include a profile based variant of the BLAST algorithm (Altschul et al. 1997), algorithms based on Hidden Markov Models (Bateman et al. 2002), and profile-profile based methods (Rychlewski et al. 2000). These algorithms are most often tested in the context of structural and fold predictions (Kinch et al. 2003), where predictions can be easily validated by comparing three dimensional structures. They are gaining acceptance also in function prediction and evolutionary analysis (Altschul and Koonin 1998; Sadreyev et al. 2003). In addition, context analysis, which takes advantage of the operon structure, has emerged as a powerful tool of annotation in prokaryotes (Overbeek, et al. 1999; Huynen, et al. 2000; Wolf, et al. 2001), and we have combined these results with those of distant homology to improve annotation of the pXO1 plasmid.

The origin of pathogenicity plasmids has often proved elusive, all the more that most of their ORFs were not annotated. Our annotation also allows us to put forward hypotheses on the evolutionary origin of the ORFs encoded in pXO1, which represent an interesting mix of vertical and horizontal transfer. Thus we are able to shed new light on the evolution of pathogenicity in the *Bacillus* genus.

RESULTS

Overview of the results

The results of our annotation effort are summarized in figure 1. All details are available as supplementary material tables on http://bioinformatics.burnham.org/pXO1. Despite previous reports, these results show that many pXO1 proteins do have recognizable homologues in other species. Overall, over 60 ORFs, previously described as unique, could be reliably identified as members of known protein families. Still, for many of them we are not able to confidently assign a molecular function. First, the full functional groups (operons, pathways) of many of the newly characterized proteins seem to be missing in pXO1. These groups may be completed by other proteins from anthrax plasmids or genome which are as yet uncharacterized, or the protein may have acquired a different functional context in anthrax. Second, many ORFs appear truncated and mutated to the point that it is unclear whether they have conserved the same function, or, in fact, whether they have any function at all (Supplementary data). This in turn might be related to the continuing evolution of the plasmid, where some genes are only partly degraded and still recognizable, like the region homologous to a part of the lethal factor (see: Particular cases section in Results) or a fragment of the NADH dehydrogenase (see: Supplementary data).

Despite these reservations, interesting tendencies emerge from our functional annotations: pXO1 contains many regulatory proteins, such as SinR (BXA0020, pXO1-14), AtxA (BXA0146, pXO1-119) or the MerR homologue (BXA0069, pXO1-47), with predicted DNA binding domains. Another interesting trend is that pXO1 has a significant number (15% of the whole plasmid) of proteins related to DNA metabolism (Supplementary data). We have also identified several probable operons, conserved among different groups of bacteria.

DNA level analysis

Several analyses were performed in order to analyze the DNA sequence of the pXO1 plasmid [Okinaka, 19999; Read, 2002; Pannucci, 2002]. The ORF prediction programs were used, the DNA motifs were discovered and a connection between promoter elements and ORFs was already done. Our analysis of the DNA sequence focused on two aspects. First, we were interested in the discovery of the origin of replication since no genes obviously involved in this process could be detected. Second, we searched for specific DNA regions related to pathogenicity.

Our goal was to find proteins directly involved in the plasmid replication. Unfortunately, we could not detect those. Therefore, we used the Oriloc program to predict the bacterial origin of replication [Frank, 2000]. In bacteria, the leading strands for replication are enriched in keto (G, T) basis while the lagging strand is enriched in amino bases (A, C) [Rocha, 1999]. This compositional assymetry allows the identification of probable origin and termination sites of replication. Oriloc analysis indicated a potential origin of replication between bases 66538 to 66558 which is quite close to the origin predicted earlier by Berry and colleagues (60955-62192 region)[Berry, 2002]. The origin is predicted in the neighbourhood of hypothetical proteins, with no recognizable homology to proteins from publicly available databases. It is located in between ORFs BXA0076 (pXO1-51) and BXA0077 (pXO1-52). The termination of replication may lie around the position 173914 on the pXO1 plasmid, between genes BXA0206 (pXO1-137) and BXA0207 (pXO1-138) which encode an RNA-binding Hfq (Host Factor I) protein and the transcription regulator from the ArsR family, respectively.

At the DNA level, we were interested in finding regions connected to the regulation of virulence. We focused on genes regulated by AtxA [Bourgogne, 2003]. Our goal was to characterize DNA regions involved in AtxA binding. For this purpose, we collected intergenic sequences preceding the AtxA-dependent genes (see Table 1 in Bourgogne, 2003] and analyzed it using the MEME [Bailey, 1994] and the MITRA [Eskin, 2002] programs. The only common motif that we could find was ANGGAG which was located in diversified distances (5-600 bp) from the putative ATG translation start codon. Large differences in the location of the ANGGAG motif can be attributed to unrecognized ORFs located upstream from some of the analyzed genes, in the same operon. Another possibility is that this signal is false. Deletion experiments of these *cis* elements should be performed to check our hypothesis.

Protein level analysis

Proposed operons: function and evolutionary conservation

A pathogenicity operon conserved in Bacilli

BXA0091 (pXO1-65) and BXA0094 are homologous to each other and to proteins from several other bacilli; *Enterococcus*, *Listeria*, *Lactococcus*, *Lactobacillus*, or other *Bacillus* species. Function of proteins from this family is unknown, but the proteins are hypothesized to be extracellular (Nakai and Horton 1999). Many members of this family have additional domains on the C-terminus, often repeats such as WD or LRR repeats, associated with protein-protein and receptor-like activities. Not only in anthrax, but also in *E.faecalis* and *B.thuringiensis*, this gene is represented by at least two copies in each operon. In *B. anthracis*, *B.thuringiensis*, *L.innocua* and *E.faecalis* the BXA0091 homologues colocalize with a surface layer domain protein. Interestingly, in species other than anthrax, these two proteins often colocalize with three proteins: a protein homologous (FFAS score: -10.100) to a protein containing the LysM domain (homology is not in the LysM region), a protein homologous to the RTX toxin and related Ca^{2+}-binding proteins family and a regulatory protein homologous to positive transcription regulators MGA. The LysM domain binds peptidoglycans and was first identified in bacterial lysins (Ponting et al. 1999). Several proteins, such as staphylococcal IgG binding proteins and *E.coli* intimins, contain LysM domains. RTX toxins are pore-forming, calcium-dependent cytotoxins encoded by various bacterial genomes (Braun and Cossart 2000), and MGA are important in streptococci virulence (McIver and Myles 2002). Other proteins from these operons in other organisms are also predicted to be extracellular and involved in pathogenesis, in *B. anthracis* this appears to be a minimal variant of this virulence related operon.

A DNA-modifying operon shared with Gram-positive bacteria

BXA0010 (pXO1-06), BXA0013 (pXO1-08) and BXA0015 (pXO1-10) form an operon that can also be found in two Gram-positive species, *Xanthomonas* and *Burkholderia* (Figure 1), and in the proteobacterial *Pseudomonas* group. BXA0010 and BXA0013 are homologues of the *Xanthomonas* orf8, of a *Burkholderia* protein and of a number of *Pseudomonas* proteins. Both BXA0010 and BXA0013 anthrax proteins belong to the superfamily II of DNA/RNA helicases, and BXA0010 seems to be a duplication of the middle part of the BXA0013 protein. In between these two proteins, in *B.anthracis*, there is an inserted reverse transcriptase (BXA0011, pXO1-07). One can hypothesize that this insertion occurred after the duplication and disrupted BXA0010. BXA0013 forms an operon with BXA0015, a protein with strong similarity to the N-terminal part of its homologues that encodes the coenzyme-binding domain of various DNA methyltransferases. The co-occurence of the DNA/RNA helicase and DNA methyltransferase is also conserved as an operon in other species mentioned above. *Xanthomonas*, *Burkholderia* and *Pseudomonas*, but not anthrax, preserve numerous other proteins in BXA0013-BXA0015 analogous operons. The function of these additional proteins is however unclear. From the functions of known members of this operon one can imply its DNA modifying function.

A nucleotide metabolism operon shared with Actinobacteria and Cyanobacteria

BXA0032 and BXA0033 (pXO1-22), if fused, would belong to the COG0175 family, members of the 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD synthetase group of enzymes which are linked to ATPase involved in DNA repair/chromosome segregation from *Anabaena* spp., *Nostoc* spp., *Bacillus stearothermophilus* and *Streptomyces avermitilis*. Functions of other proteins from this cluster are unknown. In *B.anthracis* however, it is located close to BXA0034. We described the members of this family as a new HEPN nucleotide-binding domain (Grynberg et al. 2003), and a connection with BXA0037 (pXO1-24), a nucleotidyltransferase domain protein, is obvious. As a complex they may catalyze the addition of a nucleotidyl group to unknown substrates, maybe to antibiotics or other poisonous substances, as their structural homolog kanamycin nucleotidyltransferase does (Matsumura et al. 1984). The specific function of the HEPN-nucleotidyltransferase operon in pXO1 is unknown.

Type IV secretion system machinery: two operons and missing links

Two operons in *B.anthracis* contain proteins strongly resembling elements of type IV secretion system proteins (Fig. 2). This specific secretion system is important in the delivery of effector molecules to the host cell (Christie 2001; Christie and Vogel 2000).

The first operon consists of four proteins (BXA0083/pXO1-57, BXA0085/pXO1-59, BXA0086/pXO1-60 and BXA0087/pXO1-61), of which the first is homologous to a protein involved in type IV pili biogenesis, CpaB/RcpC (COG3745). The next protein, BXA0085, belongs to the VirB11 family, and the remaining two are two paralogs belonging to the TadC family (COG2064), whose members are often found in the same operons with the VirB11. VirB11 family is well studied, (Christie 2001; Dang et al. 1999; Krause et al. 2000; Savvides et al. 2003; Yeo et al. 2000) and members of this family are ATPases that function as chaperones reminiscent of the GroEL family for translocating unfolded proteins across the cytoplasmic membrane (Christie 2001). Homologues of all four proteins from the pili biogenesis-like operon form operons in many Gram-negative bacterial species (Kachlany et al. 2000; Skerker and Shapiro 2000). To date, only in *Caulobacter crescentus* this operon was experimentally proven to be required for pilus assembly (Skerker and Shapiro 2000). Distant homologs of pilA and other pilin subunits necessary for pilus formation can be found scattered on pXO1 (for instance BXA0092) and on pXO2 (work in preparation).

The second operon contains the homologue of the VirB4 protein (BXA0107) and a fusion of the VirB6 homology region with a surface-located repetitive sequence, similar to coiled-coil proteins, with a methyl-accepting chemotaxis protein (MCP) signaling domain at the C terminus (BXA0108, pXO1-79). VirB4 family is one of the elements of the type IV secretion system. This system, ancestrally related to the conjugation machinery, is able to deliver DNA molecules as well as proteins. VirB4 is an ATPase that “might transduce information, possibly in the form of ATP-induced conformational changes, across the cytoplasmic membrane to extracytoplasmic subunits," according to Christie (Christie 2001) and Dang (Dang et al. 1999). It contains the Walker A motif responsible for ATP binding, which is well conserved in BXA0107 (200-207 fragment: GISGSGKS). The BXA0108 protein has at least 7 predicted N-terminal (55-281 aa) transmembrane motifs, similar to the central part of the VirB6 protein, and a surface-located repetitive sequence, most probably forming a coiled-coil structure. The C-terminal of this protein is homologous to a domain that is thought to transduce the external chemotaxis signal to the two-component histidine kinase CheA (for review see (Stock et al. 2002)). The next protein in this operon resembles the C-terminus of a *Bacillus firmus* integral membrane protein, which includes transmembrane domains in the N-terminal part. This region is homologous to the phosphatidate cytidylyltransferase (EC 2.7.7.41), an enzyme that catalyzes the synthesis of CDP-diglyceride, the source of phospholipids in all organisms (Icho et al. 1985; Sparrow and Raetz 1985). The function of the C-terminal part of the *B.firmus* protein is unknown.

The presence of three proteins with features characteristic of type IV secretion system and other ORFs related to type IV pilus formation strongly suggests that such a system may be active on the virulence plasmids in anthrax and may play a role in its virulence. It seems logical then to search for other elements of type IV secretion system in the anthrax plasmids or genome. We are able to detect some other distantly related elements of this machinery, but the system appears incomplete. Is it a fully functional, minimal type IV secretion system? Or are other parts of this system present in anthrax, but impossible to identify with available tools? The operons discussed here are good targets for experimental analysis, since they contain many as yet uncharacterized proteins. It is also not clear what molecules are secreted by this system, the anthrax toxin or other proteins. In any case, understanding of the function of this secretion system would be crucial for our understanding of diverse roles of pXO1 in virulence.

Putative pXO1 regulator proteins

The most important elements in the description of unknown biological systems are the regulatory proteins. They decide when, who and how is expressed in the cell. In pathogenic systems, frequently regulators of virulence genes are located in pathogenic regions. However, various permutations are known, where regulators regulate genes outside of the pathogenicity island, or regulators encoded outside of the pathogenicity island regulate genes located in the virulence regions (Hacker and Kaper 2000; Hentschel and Hacker 2001). Anthrax pX01 plasmid contains many uncharacterized regulatory proteins. We think that it is essential to describe the regulators on the anthrax pathogenicity vector in order to decipher the physiology of pXO1.

Specific duplications in the ArsR/SmtB family: BXA0166 and BXA0207

Both BXA0166 (pXO1-109) and BXA0207 (pXO1-138) are members of the ArsR/SmtB family of metalloregulatory transcriptional regulators. The vast majority of known family members are repressors. Indeed, BXA0166 has been characterized as the gene for repressor PagR (Hoffmaster and Koehler 1999). They act on operons linked to stress-inducing concentrations of diverse heavy metal ions. Derepression results from direct binding of metal ions by ArsR/SmtB transcription regulators. The founding members of the family are SmtB, the Zn(II)-responsive repressor from *Synecchococcus* PCC 7942 (Morby et al. 1993), and ArsR, that acts as the arsenic/antimony-responsive repressor of the *ars* operon in *Escherichia coli* (Wu and Rosen 1991). Another, less well studied, group in the ArsR/SmtB family are the transcriptional activators, with *Vibrio cholerae* HlyU as the founding member (Williams et al. 1993). HlyU is known to upregulate the expression of hemolysin and of two *hcp* genes, which are coregulated with hemolysin (Williams et al. 1996). We have conducted a phylogenetic analysis of this vast family, with a focus on the evolutionary history of ArsR/SmtB proteins in bacilli, notably in anthrax, and on the relation between phylogeny and function (i.e. repressor or activator).

In a phylogeny of representative members of the ArsR/SmtB family (Fig. 5A), the two pXO1 proteins are closely grouped with other *Bacillus* proteins. This group has very long branches in the tree, indicative of rapid evolution of the proteins. The only two known activators (HlyU and NolR) of the family appear closely related, in a clade with proteins of unknown function. These latter include clear orthologs of

HlyU or of NolR. It is thus reasonable to predict that these proteins form a clade of transcriptional activators. Interestingly, this "activator" clade appears closely related to the clade including both pXO1 proteins (clades boxed in Fig. 5A). PagR is known to act as a repressor, but in a weak manner (Hoffmaster and Koehler 1999) and is suspected of having an activation function as well (Mignot et al. 2003). A more detailed phylogeny of close homologues of the pXO1 proteins (Fig. 5B) shows that there has been a wave of gene duplications in the ancestor of *B.antracis* and *B.cereus* (full circles in Fig. 5B). All seven of the resulting paralogues were retained in *B.antracis*, including the two which were transferred to pXO1, while four were secondarily lost in *B.cereus*. There was an independent duplication in *B.thuringiensis* (open circle in Fig. 5B). Interestingly, these are the only bacilli represented in this clade of close homologues, all three have duplications of the gene, and all three are pathogens.

Overall, the phylogenetic analysis shows that both pXO1 ArsR/SmtB proteins are closely related members of a clade of fast evolving proteins, which have duplicated several times in pathogenic bacilli, and which are related to the only clade of transcriptional activators of the family.

Other putative regulators

BXA0020 (pXO1-14) is 564 amino acids long. The C-terminal 60-70 aa are homologous to DNA-binding domains of several repressor families (SCOP: a.35.1 superfamily of lambda repressor-like DNA-binding domains). The one that is the most similar is the SinR repressor domain (Gaur et al. 1986). In *Bacillus subtilis* the proteins of the *sin* (sporulation inhibition) region form a component of an elaborate molecular circuitry that regulates the commitment to sporulation. SinR is a tetrameric repressor protein that binds to the promoters of genes essential for entry into sporulation and prevents their transcription (Mandic-Mulec et al. 1995; Mandic-Mulec et al. 1992). In pXO1, BXA0020 does not form an operon with sin genes. Instead, it is located close to a protein (BXA0019, pXO1-13) that is characterized as similar to the middle fragment (417-1236 aa) of the 236 kDa rhoptry protein from *Plasmodium yoelii yoelii*, involved directly in the parasite attack of red blood cells (Khan et al. 2001). It is not certain whether they form one operon since both genes have putative independent ribosome binding sites. The N-terminal region of BXA0020 is not well described and has the strongest similarity to the α-helical part of the chromosome-associated kinesin, or the kinesin-like domain (KOG0244). Kinesins are microtubule-dependent molecular motors that play important roles in intracellular transport of organelles and in cell division (Mandelkow and Mandelkow 2002; Woehlke and Schliwa 2000).

The N-terminal part of BXA0048 (pXO1-34) is the DNA-binding helix-turn-helix motif that belongs to the TetR family (PF00440). Members of this family take part in the regulation of numerous pathways/operons, e.g. TetR is a tetracycline inducible repressor (Hillen and Berens 1994), BetI, a repressor of the osmoregulatory choline-glycine betaine pathway (Lamark et al. 1996), MtrR, a regulator of cell envelope permeability that acts as a repressor of *mtrCDE*-encoded and activator of *farAB*-encoded efflux pumps (Lee et al. 2003; Lee and Shafer 1999). We were unable to determine any reasonable homology to the distal part of BXA0048, therefore no functional hypothesis can be drawn. The only indication for the function of that regulator is the probable placement on one operon with a nucleotidyltransferase (BXA0047, pXO1-33). The presence on the same operon of the nucleotidyltransferase with a superfamily II DNA and RNA helicase family protein in *Streptomyces coelicolor* can be a suggestion that BXA0048 is involved in DNA metabolism.

BXA0060 (pXO1-40) belongs to a large superfamily of repressors (SCOP: a.35.1). It is composed of the DNA-binding domain only. Homologues of BXA0060 are present in numerous archaeal and eubacterial genomes, with no preservation of operon structure. It seems then that BXA0060 homologues are involved in very diverse functions/pathways.

BXA0069 (pXO1-47) belongs to the family of global transcription activators of membrane-bound multidrug transporters, responsible for bacterial multidrug resistance (MDR)(Paulsen et al. 1996). The closest homologue is the *B.subtilis* MtnA regulator that belongs to the MerR family (Summers 1992). It is known to activate two MDR transporters (*bmr* and *blt*), a transmembraneous protein-coding gene *ydfK* and its own gene (Baranova et al. 1999). It acts independently from two specific activators, BmrR and BltR, that are encoded by the *bmr* and *blt* operons (Ahmed et al. 1995). MtnA and other members of the MerR family are composed of three regions; N-terminal DNA-binding domain (winged helix-turn-helix motif), middle all-helical dimerization region and the C-terminal part specific for each protein that is probably involved in specific ligand binding (Godsey et al. 2001). BXA0069 perfectly fits this description, it possesses two quite conserved distal regions, and a 90 amino acid region of no homology that has an almost 80% probability of a coiled-coil structure (Lupas et al. 1991). Because of lack of resemblance of the C-terminus to any known regulatory domain, it is difficult to propose in what metabolism/gene(s) activation is the BXA0069 protein involved.

The FFAS analysis revealed low score similarity of BXA0122 (pXO1-89) to the MarR regulators of the multiple antibiotic resistance locus (Grkovic et al. 2002; Seoane and Levy 1995). This regulon consists of the *marRAB* operon and the *marC* gene. MarR acts as a repressor by binding as a dimer to promoter regions of the *mar* regulon (Martin and Rosner 1995). The repressive DNA-binding by MarR can be inhibited by several anionic compounds, e.g. salicylate (Alekshun and Levy 1999).

AtxA is a proven regulator of anthrax toxin genes (Dai et al. 1995; Koehler et al. 1994; Uchida et al. 1993). It is also known to influence the expression of other genes on pXO1, pXO2 plasmids and the anthrax genome (Bourgogne et al. 2003). AtxA is a member of a large, PTS (the phosphoenolpyruvate-dependent, sugar transporting phosphotransferase system) regulatory domain-containing family (Greenberg et al. 2002). Members of this family usually have a duplicated DNA/RNA binding domain and also duplicated PTS regulatory domain. Different variants of this structure are known, and additional domains are often present. Most probably, the presence of PTS EII homology domains is the necessity to act as an activator, since these domains are lacking in antiterminators (Greenberg et al. 2002). Because of its structure (Fig. 4), AtxA is believed to be a transcriptional activator. Knowing the architecture of this family, we searched the whole anthrax genome in order to find all similar regulators. Among the ones we found (Fig. 4), apart from the obvious AtxA and AcpA proteins, there is a very recent confirmation of the regulatory activity of the BXB0060 (pXO2-53), named AcpB (Drysdale et al. 2004). Diversity of domain composition and subtle structural differences in the group of evolutionary related anthrax regulators are certainly elements of a very fine regulation of stages of infection.

BXA0178 (pXO1-105) belongs to the AbrB family of "transition state regulators." AbrB was first described in *Bacillus subtilis* as an activator and repressor of numerous genes during transitions in growth phase (Phillips and Strauch 2002). Recently, Saile and Koehler (Saile and Koehler 2002) showed that the genomic copy of AbrB in *B.anthracis* regulates the expression of three toxin genes, whereas the truncated pXO1 version (BXA0178) of AbrB does not affect toxin gene expression. We can speculate then that the truncation could be crucial for BXA0178 function, or its influence on pXO1 function is not yet understood.

According to FFAS analysis, BXA0180 is an N-terminal part of the lambda repressor-like DNA-binding domain superfamily (a.35.1), as classified by the SCOP database (Andreeva et al. 2004). The ORF is truncated after the first half, and experiments are needed to check whether a shortened domain can exert any function.

BXA0206 (pXO1-137) belongs to a large family of Hfq proteins. Members of this family are known to be involved in various metabolic processes, like the regulation of iron metabolism (Masse and Gottesman 2002; Wachi et al. 1999), mRNA stability (Vytvytska et al. 1998), stabilization and degradation of RNAs (Takada et al. 1999; Tsui et al. 1997). Hfq proteins are similar to eukaryotic Sm proteins involved in RNA splicing (Moller et al. 2002). The function of the pXO1 version is not known and the RNA targeted by BXA0206 is not recognized. The question remains whether BXA0206 acts on an RNA encoded by the plasmid itself or has another function, e.g. acts on a chromosomal small RNA or disguises as the human Sm protein.

Interesting ORFs from the "pathogenic" region

The "pathogenic" region is defined as extending from BXA0057 to BXA0191 (Okinaka et al. 1999; Sirard et al. 2000), and is obviously of special interest.

BXA0139: an ORF implicated in Hemolysis?

The BXA0139 (pXO1-124) protein is located close to the oedema factor (CyaA) on the pXO1 sequence. It is 150 amino acids long, located on an operon with two unknown hypothetical proteins, BXA0138 (pXO1-125) and BXA0140 (pXO1-123). The only known fact about these proteins is the similarity of BXA0138 to BXA0149 (pXO1-117) (Supplementary data).

The most interesting finding is the homology of BXA0139 to the C-terminal end of the hemolysin II from *B.cereus* (Miles et al. 2002). This homology has already been described by Miles *et al.* (2002), but only as a similarity to a 46-amino acid segment of BXA0139. In reality, however, BXA0139 is a duplication of the same fragment, and C-end of hemolysin II is similar to both the N- and C-terminal parts of BXA0139 (Fig. 3). The significance of the C-terminus of the hemolysin II in *B.cereus* is unknown, and the functional studies suggest it has no influence on the hemolytic activity of the enzyme (Baida et al. 1999; Miles et al. 2002). Hemolysins form heptameric rings (Gouaux et al. 1997; Song et al. 1996), in which the C-terminal domain would reside in the outside part of each monomer (Miles et al. 2002). Miles and colleagues (2002) suggest three possible functions for this domain, however they do not exclude other possibilities. Either it is needed to form lattices or bind to surfaces, or has some catalytic activity. We also hypothesize an auxiliary function for the main monomer domain, maybe a regulatory function. Quite peculiar is the presence of a tandem tail-to-head repeat coded by the pXO1 plasmid. It is not fused to any catalytic domain and no overall function for the whole operon is known. The most attractive hypothesis would be the binding to surfaces. Maybe it serves as an anchor to the host cell membrane during the attack?

An interesting finding can maybe give a clue to a real function of BXA0139. We found a hemolysin II homolog in *B.anthracis* genome (gi: 21400399) that is almost identical to the *B.cereus* enzyme. However, in all anthrax strains sequenced, there is a nonsense mutation (TGG to TGA), instead of tryptophan 372 in *B.cereus*. In order to improve on the prediction of the encoded peptide, we ran the BLASTX program using the genomic sequence with large overhangs on both sides of the recognized ORF. The resulting sequence is given in the alignment in Figure 3. So, if the anthrax mutation is real (and its existence in all anthrax strains seems to reinforce this notion), we can hypothesize that BXA0139 is auxiliary to the hemolysin's function of the genomic copy of hemolysin.

Reverse homology of BXA0167

This hypothetical ORF (pXO1-108) has no identifiable homologs. Its function is also not known. It is a product of automatic translation. We could assume then that it is not an interesting target for analysis.

We performed a BLASTX analysis along its sequence and found an interesting homology coded by the opposite strand. Interspersed with nonsense mutations, we found a strong homology to the N-terminus of the lethal factor (corresponding to 9-176 amino acids of LEF)(data not shown). Noticeably, this homology region is encoded by the opposite strand from the LEF gene. Is it an example of a duplication event covered up by other events that happened later in the course of evolution? Was the part of the N-terminal LEF domain functional in the past?

DISCUSSION

In our work we described many novel features of the pXO1 plasmid that were not noticed previously. For instance, we show that parts of pXO1 are not only related to other bacilli plasmids, but also to proteins from more distant species. One of the most unexpected findings was the realization that pXO1 possesses two operons with homology to type IV secretion and pilus assembly systems. It is surprising because the type IV system is found mainly in Gram-negative bacteria (Burns 2003). Only some elements of the pilus are present in some Gram-positive bacteria (Grohmann et al. 2003; Wall and Kaiser 1999). It is even more surprising that the operons are not complete. A tempting hypothesis, which should be tested experimentally, is that the proteins present in pX01 constitute a minimal set indispensable for the formation and function of the secretion. Alternatively, these operons may have drifted from the original function. Cases both of minimal functional units, and of drift from original function, are known in pathogens and symbionts. The discovery of type IV secretion system has the potential for a significant impact on our understanding of anthrax virulence: a new pathogenic delivery pathway can be of major importance in the invasion process.

The similarity to other various bacteria and copying of parts of operons shows the phylogenetic kaleidoscope nature of this megaplasmid. Apparently, this killing agent has developed by collecting genomic pieces from a very broad range of bacteria, including pathogenicity agents as well as other organisms. Some of these pieces may be non-functional (at least in their original way) or not related to anthrax pathogenicity. It is worth noting that pXO1 shares similarity with other pathogenic bacteria also in regions *not* previously recognized as a part of the pXO1 pathogenicity island (see the operon preservation with *Burkholderia* and *Xanthomonas* in Results), whose status may have to be revised.

A detailed analysis of the pXO1 sequence by Okinaka *et al.* (Okinaka et al. 1999) focused mostly on the analysis of mobile elements, their number and possible implication for the evolution of the plasmid. Our findings not only suggest a thorough history of transposition but also allow us to hypothesize on the probable entities that were used to build pXO1. Interestingly, even if the type IV clusters are located inside the putative PAI, one can guess it was an indispensable part of the plasmid sequence, however the presence of the IS DD-E transposases suggest it is a new, independent insertion. Another option would be that we deal with a conjugative transposon, unusually equipped with a set of DD-E transposases instead of Tyr or Ser recombinases. The important question to understand pXO1 as a mobile entity is to localize the replication machinery. We were unable to find it, which makes this even more intriguing, however we identified the putative replication start and termination sites. The nature of replication should be informative on the nature and provenience of the pXO1 plasmid.

The discovery of previously unknown systems on pXO1 plasmid of course begs questions about their regulation. External signals, cell state or host-pathogen interaction certainly trigger bacterial response(s), and several of them are already known (for review see (Koehler 2002)). All these signals finally activate transcription of virulence-related genes. We have attempted to describe all possible regulators that we could find, using sensitive profile-profile alignment programs. Some of the regulatory proteins are known not to influence the toxin function (e.g. the homologue of AbrB), but others form priority targets for experimental studies of pathogenicity and *B.anthracis* biology. Notably, do the newly discovered factors regulate plasmid genes or chromosome genes?

We don't know how important is the presence of a common motif for AtxA-regulated genes. Its variable location throughout the putative promoter regions (closer or further to the ATG) poses questions. However, there may be ORFs not yet recognized 5' from the ones that are AtxA-dependent. In this case, the recognized ANGGAG sequence would directly precede the operon. Deletion experiments are needed to test whether these *cis* elements have any impact on the function of AtxA-regulated genes.

Another interesting finding is the diversity of ArsR homologs in *B.anthracis*. The majority of these, including those on pXO1, are related to the activator subfamily. The functions of MarR and TetR regulators are also intriguing.

There are two striking features of the whole plasmid that brought our special attention. First, the presence of so many DNA metabolism-related proteins (15%)(Supplementary data). It seems that DNA is a central point of the function of pXO1. Is this function related with the processing of pXO1, chromosomal DNA, transposons, or host DNA? None of these hypotheses can be excluded at the moment. The type IV delivery system could be an indication that some of them could have an external function. Second, when analyzing the DNA and proteome of pXO1 we realized how messy it is. pXO1 is full of incomplete and mutated ORFs (see Results and Supplementary data). There are many traces of ancient duplications, some still fresh (strong homology), but some almost completely faded away (homology barely recognizable), and often disrupted. It also consists of ORFs "borrowed" from other species. pXO1 seems to be the subject of constant evolutionary flux. The pXO1 plasmid should have a tag: "under construction."

METHODS

Gene names

The pXO1 plasmid was sequenced at least twice, by two independent research groups. Interestingly, the two sequences differ significantly, both on the DNA and on the (predicted) protein level. The second more recent sequencing identified almost 100 additional genes on pXO1. Several alternative naming conventions for *B.anthracis* plasmid proteins are used in literature. We use the names used by the pXO1 sequencing team (Read et al. 2003) (e.g. BXA007) as our primary names, but where appropriate we also provide the names used by the previous sequencing team (e.g. pXO1-04) or common gene names used in the literature (e.g. AtxA) when available.

DNA level analysis

The *Bacillus anthracis* strain A2012 pXO1 plasmid sequence was used for analysis (accession: NC_003980)(Read et al. 2003).

We used the Oriloc (Frank and Lobry 2000) program to detect pXO1 origin of replication, using the gene coordinates provided in pXO1 Genbank file.

For the analysis of common DNA features in promoter regions of AtxA-dependent genes (Bourgogne et al. 2003), we used the total DNA sequences between the end of a previous gene and the ATG neighbourhood of the AtxA-regulated gene. We used the 5' regions of the following genes from pXO1 and pXO2 plasmids: BXA0019 (pXO1-13), BXA0124 (pXO1-90), BXA0125 (pXO1-91), BXA0137 (pXO1-126), BXA0142 (cyaA), BXA0164 (pagA), BXA0172 (lef), BXB0045 (pXO1-31), BXB0060 (pXO1-40), BXB0066 (pXO1-58), BXB0074, BXB0084 (pXO1-124). We used MEME and MITRA programs to search for common motifs (Bailey and Elkan 1994; Eskin and Pevzner 2002).

Protein level analysis

For the analysis of the pXO1 proteome, we used proteins accessible with the BXAxxxx NCBI numbers, enforced with the BLASTX analysis (Altschul et al. 1990).

To analyze the protein sequences, we used the following programs: BLAST tools (Altschul et al. 1990; Altschul et al. 1997), SMART tool (Letunic et al. 2002), Pfam (Bateman et al. 2002), CDD (Marchler-Bauer et al. 2003), TMHMM2.0 (Sonnhammer et al. 1998), SEED (Read et al. 2003), Radar (Heger and Holm 2000), FFAS03 (Rychlewski et al. 2000), Metaserver.pl (Ginalski et al. 2003), Superfamily (Gough et al. 2001).

To align sequences we used: T-COFFEE (Notredame et al. 2000), AliBee (Nikolaev et al. 1997), MultAlin (Corpet 1988), BioEdit (Hall 1999).

Phylogenetic trees were estimated from amino acid alignments using PHYML (Guindon and Gascuel 2003), a fast and accurate Maximum Likelihood heuristic, under the JTT substitution model (Jones, Taylor et al. 1992), with a gamma distribution of rates between sites (eight categories, parameter alpha estimated by PHYML). Bootstrap support of branches was estimated using the programs SEQBOOT and CONSENSE of the PHYLIP package (Felsenstein 2002) with 1000 replicates; the parameter alpha was estimated independently for each repetition.

ACKNOWLEDGMENTS

We thank Dr. Andrei Osterman for help in the context analysis and Stephane Guindon for help with PHYML. This work was supported by a United States Army Medical Research and Materiel Command Grant DAMD17-03-2-0038.

REFERENCES

Ahmed, M., L. Lyass, P.N. Markham, S.S. Taylor, N. Vazquez-Laslop, and A.A. Neyfakh. 1995. Two highly similar multidrug transporters of Bacillus subtilis whose expression is differentially regulated. *J Bacteriol* **177:** 3904-3910.

Alekshun, M.N. and S.B. Levy. 1999. Alteration of the repressor activity of MarR, the negative regulator of the Escherichia coli marRAB locus, by multiple chemicals in vitro. *J Bacteriol* **181:** 4669-4672.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Altschul, S.F. and E.V. Koonin. 1998. Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem Sci* **23:** 444-447.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-3402.

Andreeva, A., D. Howorth, S.E. Brenner, T.J. Hubbard, C. Chothia, and A.G. Murzin. 2004. SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* **32 Database issue:** D226-229.

Ariel, N., A. Zvi, H. Grosfeld, O. Gat, Y. Inbar, B. Velan, S. Cohen, and A. Shafferman. 2002. Search for potential vaccine candidate open reading frames in the Bacillus anthracis virulence plasmid pXO1: in silico and in vitro screening. *Infect Immun* **70:** 6817-6827.

Ariel, N., A. Zvi, K.S. Makarova, T. Chitlaru, E. Elhanany, B. Velan, S. Cohen, A.M. Friedlander, and A. Shafferman. 2003. Genome-based bioinformatic selection of chromosomal Bacillus anthracis putative vaccine candidates coupled with proteomic identification of surface-associated antigens. *Infect Immun* **71:** 4563-4579.

Baida, G., Z.I. Budarina, N.P. Kuzmin, and A.S. Solonin. 1999. Complete nucleotide sequence and molecular characterization of hemolysin II gene from Bacillus cereus. *FEMS Microbiol Lett* **180:** 7-14.

Bailey, T.L. and C. Elkan. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2:** 28-36.

Baranova, N.N., A. Danchin, and A.A. Neyfakh. 1999. Mta, a global MerR-type regulator of the Bacillus subtilis multidrug-efflux transporters. *Mol Microbiol* **31:** 1549-1559.

Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy, S. Griffiths-Jones, K.L. Howe, M. Marshall, and E.L. Sonnhammer. 2002. The Pfam protein families database. *Nucleic Acids Res* **30:** 276-280.

Bourgogne, A., M. Drysdale, S.G. Hilsenbeck, S.N. Peterson, and T.M. Koehler. 2003. Global effects of virulence gene regulators in a Bacillus anthracis strain with both virulence plasmids. *Infect Immun* **71:** 2736-2743.

Braun, L. and P. Cossart. 2000. Interactions between Listeria monocytogenes and host mammalian cells. *Microbes Infect* **2:** 803-811.

Burns, D.L. 2003. Type IV transporters of pathogenic bacteria. *Curr. Opin. Microbiol.* **6:** 29-34.

Christie, P.J. 2001. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol Microbiol* **40:** 294-305.

Christie, P.J. and J.P. Vogel. 2000. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* **8:** 354-360.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16:** 10881-10890.

Dai, Z., J.C. Sirard, M. Mock, and T.M. Koehler. 1995. The atxA gene product activates transcription of the anthrax toxin genes and is essential for virulence. *Mol Microbiol* **16:** 1171-1181.

Dang, T.A., X.R. Zhou, B. Graf, and P.J. Christie. 1999. Dimerization of the Agrobacterium tumefaciens VirB4 ATPase and the effect of ATP-binding cassette mutations on the assembly and function of the T-DNA transporter. *Mol Microbiol* **32:** 1239-1253.

Drysdale, M., A. Bourgogne, S.G. Hilsenbeck, and T.M. Koehler. 2004. atxA Controls Bacillus anthracis Capsule Synthesis via acpA and a Newly Discovered Regulator, acpB. *J Bacteriol* **186:** 307-315.

Eskin, E. and P.A. Pevzner. 2002. Finding composite regulatory patterns in DNA sequences. *Bioinformatics* **18 Suppl 1:** S354-363.

Fouet, A. and S. Mesnage. 2002. Bacillus anthracis cell envelope components. *Curr Top Microbiol Immunol* **271:** 87-113.

Frank, A.C. and J.R. Lobry. 2000. Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics* **16:** 560-561.

Gaur, N.K., E. Dubnau, and I. Smith. 1986. Characterization of a cloned Bacillus subtilis gene that inhibits sporulation in multiple copies. *J Bacteriol* **168:** 860-869.

Ginalski, K., A. Elofsson, D. Fischer, and L. Rychlewski. 2003. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **19:** 1015-1018.

Godsey, M.H., N.N. Baranova, A.A. Neyfakh, and R.G. Brennan. 2001. Crystal Structure of MtaN, a Global Multidrug Transporter Gene Activator. *J. Biol. Chem.* **276:** 47178-47184.

Gouaux, E., M. Hobaugh, and L. Song. 1997. alpha-Hemolysin, gamma-hemolysin, and leukocidin from Staphylococcus aureus: distant in sequence but similar in structure. *Protein Sci* **6:** 2631-2635.

Gough, J., K. Karplus, R. Hughey, and C. Chothia. 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* **313:** 903-919.

Greenberg, D.B., J. Stulke, and M.H. Saier, Jr. 2002. Domain analysis of transcriptional regulators bearing PTS regulatory domains. *Res Microbiol* **153:** 519-526.

Grkovic, S., M.H. Brown, and R.A. Skurray. 2002. Regulation of bacterial drug export systems. *Microbiol Mol Biol Rev* **66:** 671-701, table of contents.

Grohmann, E., G. Muth, and M. Espinosa. 2003. Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev.* **67:** 277-301.

Grynberg, M., H. Erlandsen, and A. Godzik. 2003. HEPN: a common domain in bacterial drug resistance and human neurodegenerative proteins. *Trends Biochem Sci* **28:** 224-226.

Guindon, S. and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696-704.

Hacker, J. and J.B. Kaper. 2000. Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54:** 641-679.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41:** 95-98.

Heger, A. and L. Holm. 2000. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins* **41:** 224-237.

Hentschel, U. and J. Hacker. 2001. Pathogenicity islands: the tip of the iceberg. *Microbes Infect* **3:** 545-548.

Hillen, W. and C. Berens. 1994. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol* **48:** 345-369.

Hoffmaster, A.R. and T.M. Koehler. 1999. Autogenous regulation of the Bacillus anthracis pag operon. *J Bacteriol* **181:** 4485-4492.

Icho, T., C.P. Sparrow, and C.R. Raetz. 1985. Molecular cloning and sequencing of the gene for CDP-diglyceride synthetase of Escherichia coli. *J Biol Chem* **260:** 12078-12083.

Jeanmougin, F., J.D. Thompson, M. Gouy, D.G. Higgins, and T.J. Gibson. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23:** 403-405.

Kachlany, S.C., P.J. Planet, M.K. Bhattacharjee, E. Kollia, R. DeSalle, D.H. Fine, and D.H. Figurski. 2000. Nonspecific adherence by Actinobacillus actinomycetemcomitans requires genes widespread in bacteria and archaea. *J Bacteriol* **182:** 6169-6176.

Kachlany, S.C., P.J. Planet, R. DeSalle, D.H. Fine, and D.H. Figurski. 2001. Genes for tight adherence of Actinobacillus actinomycetemcomitans: from plaque to plague to pond scum. *Trends Microbiol* **9:** 429-437.

Khan, S., W. Jarra, and P. Preiser. 2001. The 235 kDa rhoptry protein of Plasmodium (yoelii) yoelii: function at the junction. *Mol Biochem Parasitol.* **117**.

Kinch, L.N., J.O. Wrabl, S.S. Krishna, I. Majumdar, R.I. Sadreyev, Y. Qi, J. Pei, H. Cheng, and N.V. Grishin. 2003. CASP5 assessment of fold recognition target predictions. *Proteins* **53 Suppl 6:** 395-409.

Koehler, T.M. 2002. Bacillus anthracis genetics and virulence gene regulation. *Curr Top Microbiol Immunol* **271:** 143-164.

Koehler, T.M., Z. Dai, and M. Kaufman-Yarbray. 1994. Regulation of the Bacillus anthracis protective antigen gene: CO2 and a trans-acting element activate transcription from one of two promoters. *J Bacteriol* **176:** 586-595.

Krause, S., W. Pansegrau, R. Lurz, F. de la Cruz, and E. Lanka. 2000. Enzymology of type IV macromolecule secretion systems: the conjugative transfer regions of plasmids RP4 and R388 and the cag pathogenicity island of Helicobacter pylori encode structurally and functionally related nucleoside triphosphate hydrolases. *J Bacteriol* **182:** 2761-2770.

Lacy, D.B. and R.J. Collier. 2002. Structure and function of anthrax toxin. *Curr Top Microbiol Immunol* **271:** 61-85.

Lamark, T., T.P. Rokenes, J. McDougall, and A.R. Strom. 1996. The complex bet promoters of Escherichia coli: regulation by oxygen (ArcA), choline (BetI), and osmotic stress. *J Bacteriol* **178:** 1655-1662.

Lee, E.H., C. Rouquette-Loughlin, J.P. Folster, and W.M. Shafer. 2003. FarR Regulates the farAB-Encoded Efflux Pump of Neisseria gonorrhoeae via an MtrR Regulatory Mechanism. *J Bacteriol* **185:** 7145-7152.

Lee, E.H. and W.M. Shafer. 1999. The farAB-encoded efflux pump mediates resistance of gonococci to long-chained antibacterial fatty acids. *Mol Microbiol* **33:** 839-845.

Letunic, I., L. Goodstadt, N.J. Dickens, T. Doerks, J. Schultz, R. Mott, F. Ciccarelli, R.R. Copley, C.P. Ponting, and P. Bork. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* **30:** 242-244.

Lupas, A., M. Van Dyke, and J. Stock. 1991. Predicting coiled coils from protein sequences. *Science* **252:** 1162-1164.

Mandelkow, E. and E.M. Mandelkow. 2002. Kinesin motors and disease. *Trends Cell Biol* **12:** 585-591.

Mandic-Mulec, I., L. Doukhan, and I. Smith. 1995. The Bacillus subtilis SinR protein is a repressor of the key sporulation gene spo0A. *J Bacteriol* **177:** 4619-4627.

Mandic-Mulec, I., N. Gaur, U. Bai, and I. Smith. 1992. Sin, a stage-specific repressor of cellular differentiation. *J Bacteriol* **174:** 3561-3569.

Marchler-Bauer, A., J.B. Anderson, C. DeWeese-Scott, N.D. Fedorova, L.Y. Geer, S. He, D.I. Hurwitz, J.D. Jackson, A.R. Jacobs, C.J. Lanczycki, C.A. Liebert, C. Liu, T. Madej, G.H. Marchler, R. Mazumder, A.N. Nikolskaya, A.R. Panchenko, B.S. Rao, B.A. Shoemaker, V. Simonyan, J.S. Song, P.A. Thiessen, S. Vasudevan, Y. Wang, R.A. Yamashita, J.J. Yin, and S.H. Bryant. 2003. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* **31:** 383-387.

Martin, R.G. and J.L. Rosner. 1995. Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences. *Proc Natl Acad Sci U S A* **92:** 5456-5460.

Masse, E. and S. Gottesman. 2002. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A* **99:** 4620-4625.

Matsumura, M., Y. Katakura, T. Imanaka, and S. Aiba. 1984. Enzymatic and nucleotide sequence studies of a kanamycin-inactivating enzyme encoded by a plasmid from thermophilic bacilli in comparison with that encoded by plasmid pUB110. *J Bacteriol* **160:** 413-420.

McIver, K.S. and R.L. Myles. 2002. Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. *Mol Microbiol* **43:** 1591-1601.

Mignot, T., M. Mock, and A. Fouet. 2003. A plasmid-encoded regulator couples the synthesis of toxins and surface structures in Bacillus anthracis. *Mol Microbiol* **47:** 917-927.

Miles, G., H. Bayley, and S. Cheley. 2002. Properties of Bacillus cereus hemolysin II: a heptameric transmembrane pore. *Protein Sci* **11:** 1813-1824.

Moller, T., T. Franch, P. Hojrup, D.R. Keene, H.P. Bachinger, R.G. Brennan, and P. Valentin-Hansen. 2002. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* **9:** 23-30.

Morby, A.P., J.S. Turner, J.W. Huckle, and N.J. Robinson. 1993. SmtB is a metal-dependent repressor of the cyanobacterial metallothionein gene smtA: identification of a Zn inhibited DNA-protein complex. *Nucleic Acids Res* **21:** 921-925.

Nakai, K. and P. Horton. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24:** 34-36.

Nikolaev, V.K., A.M. Leontovich, V.A. Drachev, and B.L. I. 1997. Building multiple alignment using iterative analyzing biopolymers structure dynamic improvement of the initial motif alignment. *Biochemistry* **62:** 578-582.

Notredame, C., D.G. Higgins, and J. Heringa. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302:** 205-217.

Okinaka, R., K. Cloud, O. Hampton, A. Hoffmaster, K. Hill, P. Keim, T. Koehler, G. Lamke, S. Kumano, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P. Jackson. 1999. Sequence, assembly and analysis of pXO1 and pXO2. *J Appl Microbiol* **87:** 261-262.

Okinaka, R.T., K. Cloud, O. Hampton, A.R. Hoffmaster, K.K. Hill, P. Keim, T.M. Koehler, G. Lamke, S. Kumano, J. Mahillon, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P.J. Jackson. 1999. Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the anthrax toxin genes. *J Bacteriol* **181:** 6509-6515.

Page, R.D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12:** 357-358.

Pannucci, J., R.T. Okinaka, R. Sabin, and C.R. Kuske. 2002. Bacillus anthracis pXO1 plasmid sequence conservation among closely related bacterial species. *J Bacteriol* **184:** 134-141.

Pasteur, L. 1881. Compte rendu sommaire des experiences faites a Pouilly-Le-Fort, pres de Meun, sur la vaccination charnonneuse (avec la collaboration de MM. Chamberland et Roux). *Compte Rendus Acad Sci* **XCII:** 1378-1383.

Paulsen, I.T., M.H. Brown, and R.A. Skurray. 1996. Proton-dependent multidrug efflux systems. *Microbiol Rev* **60:** 575-608.

Phillips, Z.E. and M.A. Strauch. 2002. Bacillus subtilis sporulation and stationary phase gene expression. *Cell Mol Life Sci* **59:** 392-402.

Ponting, C.P., L. Aravind, J. Schultz, P. Bork, and E.V. Koonin. 1999. Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol* **289:** 729-745.

Rasko, D.A., J. Ravel, O.K. OA, E. Helgason, R.Z. Cer, L. Jiang, K.A. Shores, D.E. Fouts, N.J. Tourasse, S.V. Angiuoli, J. Kolonay, W.C. Nelson, A.B. Kolsto, C.M. Fraser, and T.D. Read. 2004. The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis pXO1. *Nucleic Acids Res* **32:** 977-988.

Read, T.D., S.N. Peterson, N. Tourasse, L.W. Baillie, I.T. Paulsen, K.E. Nelson, H. Tettelin, D.E. Fouts, J.A. Eisen, S.R. Gill, E.K. Holtzapple, O.A. Okstad, E. Helgason, J. Rilstone, M. Wu, J.F. Kolonay, M.J. Beanan, R.J. Dodson, L.M. Brinkac, M. Gwinn, R.T. DeBoy, R. Madpu, S.C. Daugherty, A.S. Durkin, D.H. Haft, W.C. Nelson, J.D. Peterson, M. Pop, H.M. Khouri, D. Radune, J.L. Benton, Y. Mahamoud, L. Jiang, I.R. Hance, J.F. Weidman, K.J. Berry, R.D. Plaut, A.M. Wolf, K.L. Watkins, W.C. Nierman, A. Hazen, R. Cline, C. Redmond, J.E. Thwaite, O. White, S.L. Salzberg, B. Thomason, A.M. Friedlander, T.M. Koehler, P.C. Hanna, A.B. Kolsto, and C.M. Fraser. 2003. The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. *Nature* **423:** 81-86.

Rychlewski, L., L. Jaroszewski, W. Li, and A. Godzik. 2000. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* **9:** 232-241.

Sadreyev, R.I., D. Baker, and N.V. Grishin. 2003. Profile-profile comparisons by COMPASS predict intricate homologies between protein families. *Protein Sci* **12:** 2262-2272.

Saile, E. and T.M. Koehler. 2002. Control of anthrax toxin gene expression by the transition state regulator abrB. *J Bacteriol* **184:** 370-380.

Savvides, S.N., H.J. Yeo, M.R. Beck, F. Blaesing, R. Lurz, E. Lanka, R. Buhrdorf, W. Fischer, R. Haas, and G. Waksman. 2003. VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion. *Embo J* **22:** 1969-1980.

Seoane, A.S. and S.B. Levy. 1995. Characterization of MarR, the repressor of the multiple antibiotic resistance (mar) operon in Escherichia coli. *J Bacteriol* **177:** 3414-3419.

Sirard, J.C., C. Guidi-Rontani, A. Fouet, and M. Mock. 2000. Characterization of a plasmid region involved in Bacillus anthracis toxin production and pathogenesis. *Int J Med Microbiol* **290:** 313-316.

Skerker, J.M. and L. Shapiro. 2000. Identification and cell cycle control of a novel pilus system in Caulobacter crescentus. *Embo J* **19:** 3223-3234.

Song, L., M.R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J.E. Gouaux. 1996. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274:** 1859-1866.

Sonnhammer, E.L., G. von Heijne, and A. Krogh. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6:** 175-182.

Sparrow, C.P. and C.R. Raetz. 1985. Purification and properties of the membrane-bound CDP-diglyceride synthetase from Escherichia coli. *J Biol Chem* **260:** 12084-12091.

Stock, J.B., M.N. Levit, and P.M. Wolanin. 2002. Information processing in bacterial chemotaxis. *Sci STKE* **2002:** PE25.

Summers, A.O. 1992. Untwist and shout: a heavy metal-responsive transcriptional regulator. *J Bacteriol* **174:** 3097-3101.

Takada, A., M. Wachi, and K. Nagai. 1999. Negative regulatory role of the Escherichia coli hfq gene in cell division. *Biochem Biophys Res Commun* **266:** 579-583.

Tsui, H.C., G. Feng, and M.E. Winkler. 1997. Negative regulation of mutS and mutH repair gene expression by the Hfq and RpoS global regulators of Escherichia coli K-12. *J Bacteriol* **179:** 7476-7487.

Turnbull, P.C. 2002. Introduction: anthrax history, disease and ecology. *Curr Top Microbiol Immunol* **271:** 1-19.

Uchida, I., J.M. Hornung, C.B. Thorne, K.R. Klimpel, and S.H. Leppla. 1993. Cloning and characterization of a gene whose product is a trans-activator of anthrax toxin synthesis. *J Bacteriol* **175:** 5329-5338.

Vytvytska, O., J.S. Jakobsen, G. Balcunaite, J.S. Andersen, M. Baccarini, and A. von Gabain. 1998. Host factor I, Hfq, binds to Escherichia coli ompA mRNA in a growth rate-dependent fashion and regulates its stability. *Proc Natl Acad Sci U S A* **95:** 14118-14123.

Wachi, M., A. Takada, and K. Nagai. 1999. Overproduction of the outer-membrane proteins FepA and FhuE responsible for iron transport in Escherichia coli hfq:cat mutant. *Biochem Biophys Res Commun* **264:** 525-529.

Wall, D. and D. Kaiser. 1999. Type IV pili and cell motility. *Mol Microbiol.* **32:** 1-10.

Williams, S.G., S.R. Attridge, and P.A. Manning. 1993. The transcriptional activator HlyU of Vibrio cholerae: nucleotide sequence and role in virulence gene expression. *Mol Microbiol* **9:** 751-760.

Williams, S.G., L.T. Varcoe, S.R. Attridge, and P.A. Manning. 1996. Vibrio cholerae Hcp, a secreted protein coregulated with HlyA. *Infect Immun* **64:** 283-289.

Woehlke, G. and M. Schliwa. 2000. Walking on two heads: the many talents of kinesin. *Nat Rev Mol Cell Biol* **1:** 50-58.

Wu, J. and B.P. Rosen. 1991. The ArsR protein is a trans-acting regulatory protein. *Mol Microbiol* **5:** 1331-1336.

Yeo, H.J., S.N. Savvides, A.B. Herr, E. Lanka, and G. Waksman. 2000. Crystal structure of the hexameric traffic ATPase of the Helicobacter pylori type IV secretion system. *Mol Cell* **6:** 1461-1472.

FIGURE LEGENDS

Figure 1. A summary of the distribution of homologs of the predicted proteins (ORFs) encoded in pXO1 plasmid in a set of >100 diverse microbial genomes. Only relatively close homologues (with FASTA P-score above 10-3) were taken into account at this stage of the analysis. Relative size and polarity of ORFs (using the predictions and the nomenclature by TIGR) on the linearized map of pXO1 are illustrated by the heights (cutoff at 500 amino acids) and orientation of the bars along the X-axis (panel A, continued on panel

B). Open bars correspond to proteins for which no homologues have been detected in this analysis. Bars with matching colored borders correspond to "repeats" present in pXO1. Black and colored bars in correspond to proteins for which at least one homolog was detected in this analysis.

Panel C (and its continuation in panel D) mark the presence of respective homologues in at least one of the representative genomes in several groups (as indicated in respective boxes):

Group 1: *B. anthracis* (chromosome or pXO2), *B. thuringiensis* or *B. cereus*.

Group 2: *B. subtilis*, *B. halodurans* or *B. stearothermophilus*.

Group 3: *Staphylococci*, *Streptococci* or *Eneterococci* species.

Group 4: *Salmonella*, *Xanthomonas* or *Burkholderia* species.

Group 5: *Geobacter*, *Anabaena* or *Nostoc* species.

These genomes contain the largest number of homologues of pXO1-borne proteins, and jointly they provide a nearly complete coverage of the phylogenetic space of pXO1 homologues.

.

Figure 2. Type IV secretion and pilus systems representations with homologous genes in *B.anthracis* shown in red. It is worth noting that in the secretion operon representation, the anthrax VirB6 gene is fused to an adhesin-like long sequence, whereas in the pilus assembly operon the last homologue, TadC, has two representations in the anthrax operon. For more detailed comparison to known type IV secretion and pilus assembly systems, see (Christie 2001; Christie and Vogel 2000; Kachlany et al. 2000; Kachlany et al. 2001; Skerker and Shapiro 2000).

Figure 3. The multiple alignment of the *Bacillus cereus* terminal hemolysin II domain, two parts of the BXA0139/pXO1-124 protein, the *Streptococcus* phage Cp-1 orf16 and the *B.anthracis* hemolysin II copy with a truncated C terminus. The star represents the stop codon in the anthrax DNA sequence.

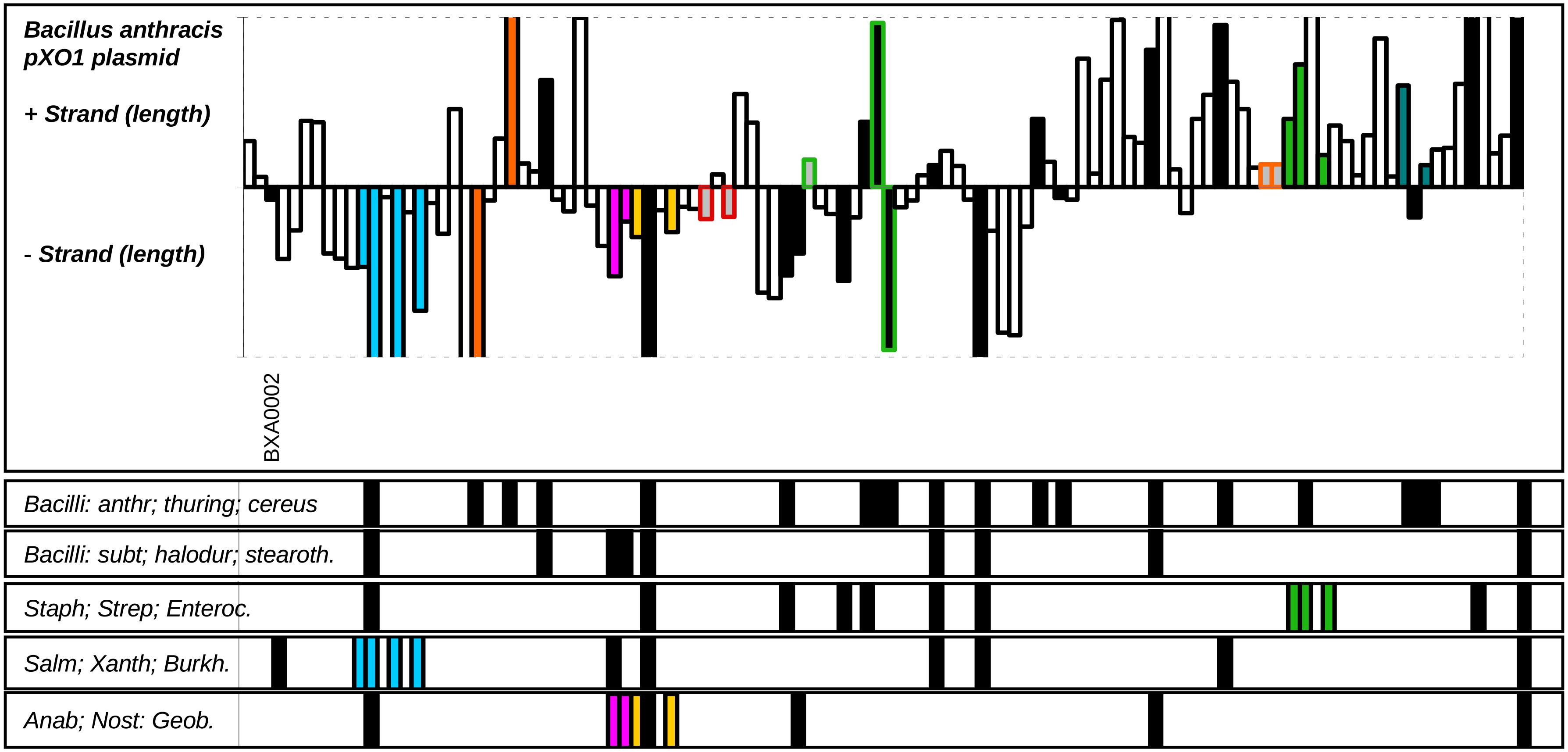
Figure 4. The domain structure of the AtxA family of protein from *B.anthracis*. Each colour depicts a family of most homologous sequences. Similar colours describe duplicated sequences.

Figure 5. Phylogenetic trees of ArsR/SmtB proteins.

Phylogenies estimated using PHYML (Guindon and Gascuel 2003). Figures at nodes are bootstrap support in % of 1000 replicates; bootstrap proportions under 50% are not reported. Branch length is proportional to the estimated number of substitutions per site. Proteins from pXO1 are boxed.

(A) Phylogeny of representative proteins sampling the diversity of the ArsR/SmtB family. Two *B.anthracis* proteins with short sequences are not included (Q81NE6 and Q81QQ6). Unrooted tree drawn using TreeView (Page 1996); the measure bar represents 0.1 substitutions/site. The boxes indicate clades (monophyletic groups) discussed in the text.

(B) Phylogeny of pXO1 ArsR/SmtB proteins and close homologues. This corresponds to the box "close homologs of pXO1 proteins" in (A), plus all closely related homologs as determined from a phylogeny of all available ArsR/SmtB sequences (487 sequences; tree not shown). Tree rooted according to the phylogeny of all ArsR/SmtB proteins, and drawn using NJplot (Jeanmougin et al. 1998); the measure bar represents 0.5 substitutions/site. Full circles indicate gene duplications in the common ancestor of *B.antracis* and *B.cereus*; the empty circle indicates a gene duplication in *B.thuringiensis*.


Bacillus anthracis pXO1 plasmid
+ Strand (length)
- Strand (length)
BXA0002
Bacilli: anthr; thuring; cereus
Bacilli: subt; halodur; stearoth.
Staph; Strep; Enteroc.
Salm; Xanth; Burkh.
Anab; Nost: Geob.

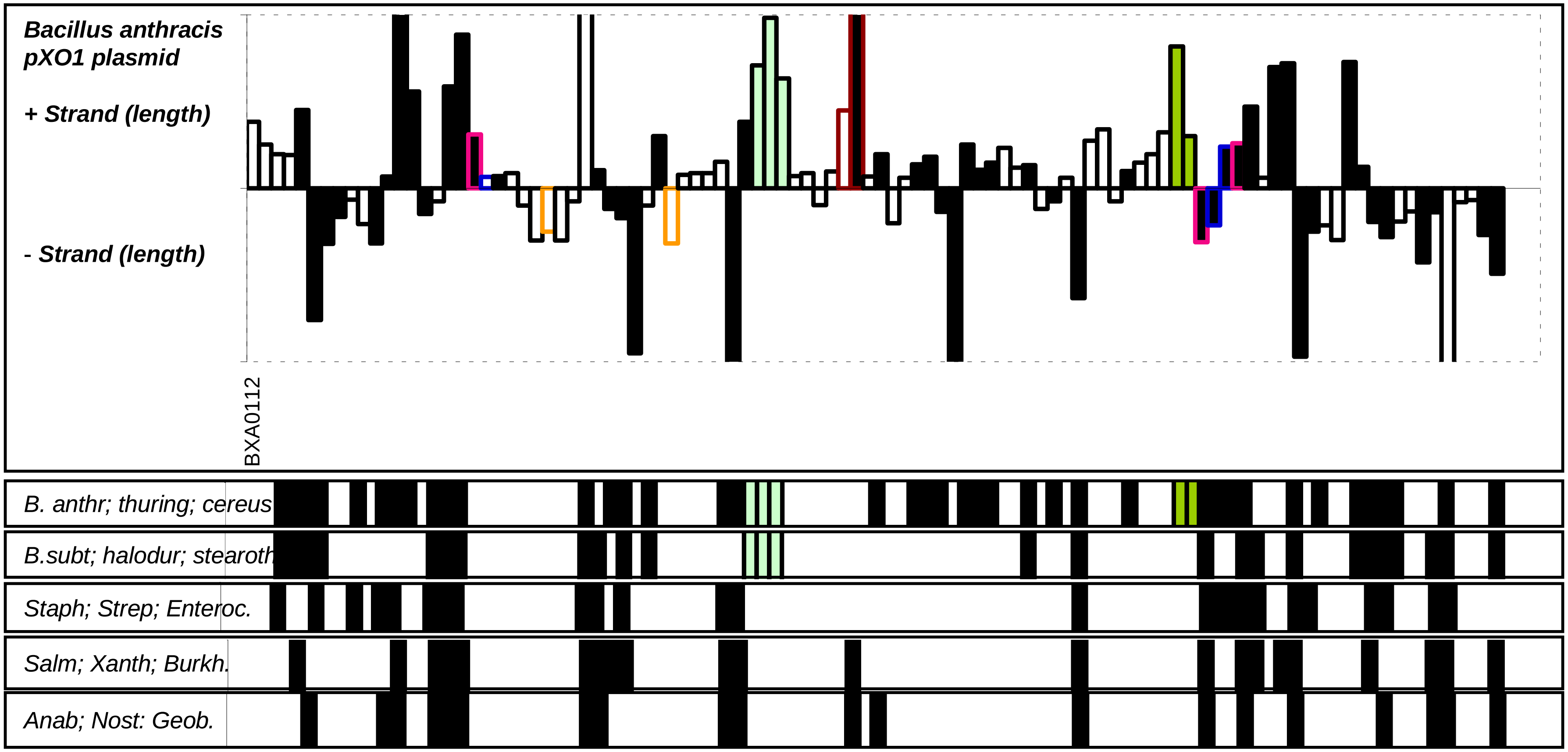

Bacillus anthracis pXO1 plasmid
+ Strand (length)
- Strand (length)
BXA0112
B. anthr; thuring; cereus
B.subt; halodur; stearoth
Staph; Strep; Enteroc.
Salm; Xanth; Burkh.
Anab; Nost: Geob.

Figure 2

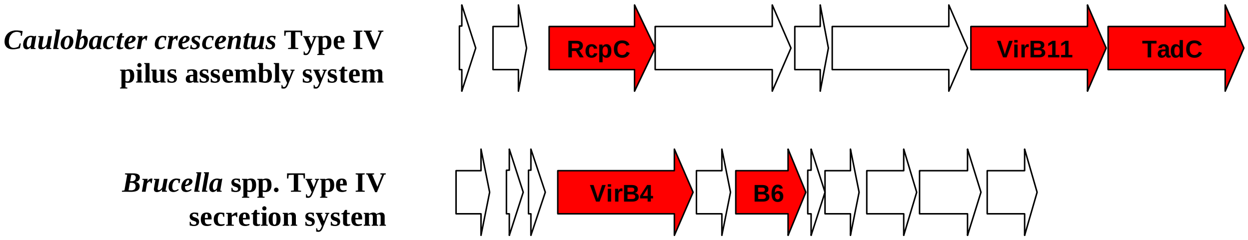
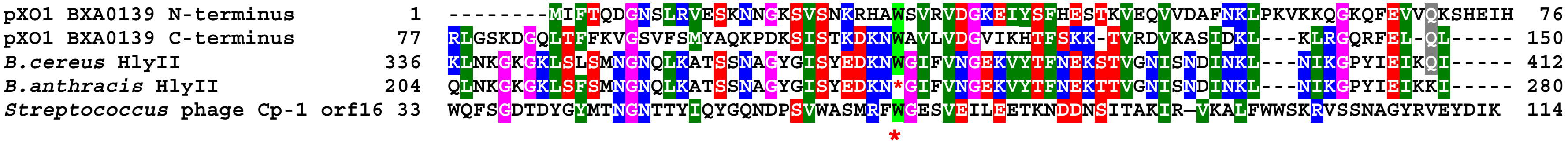
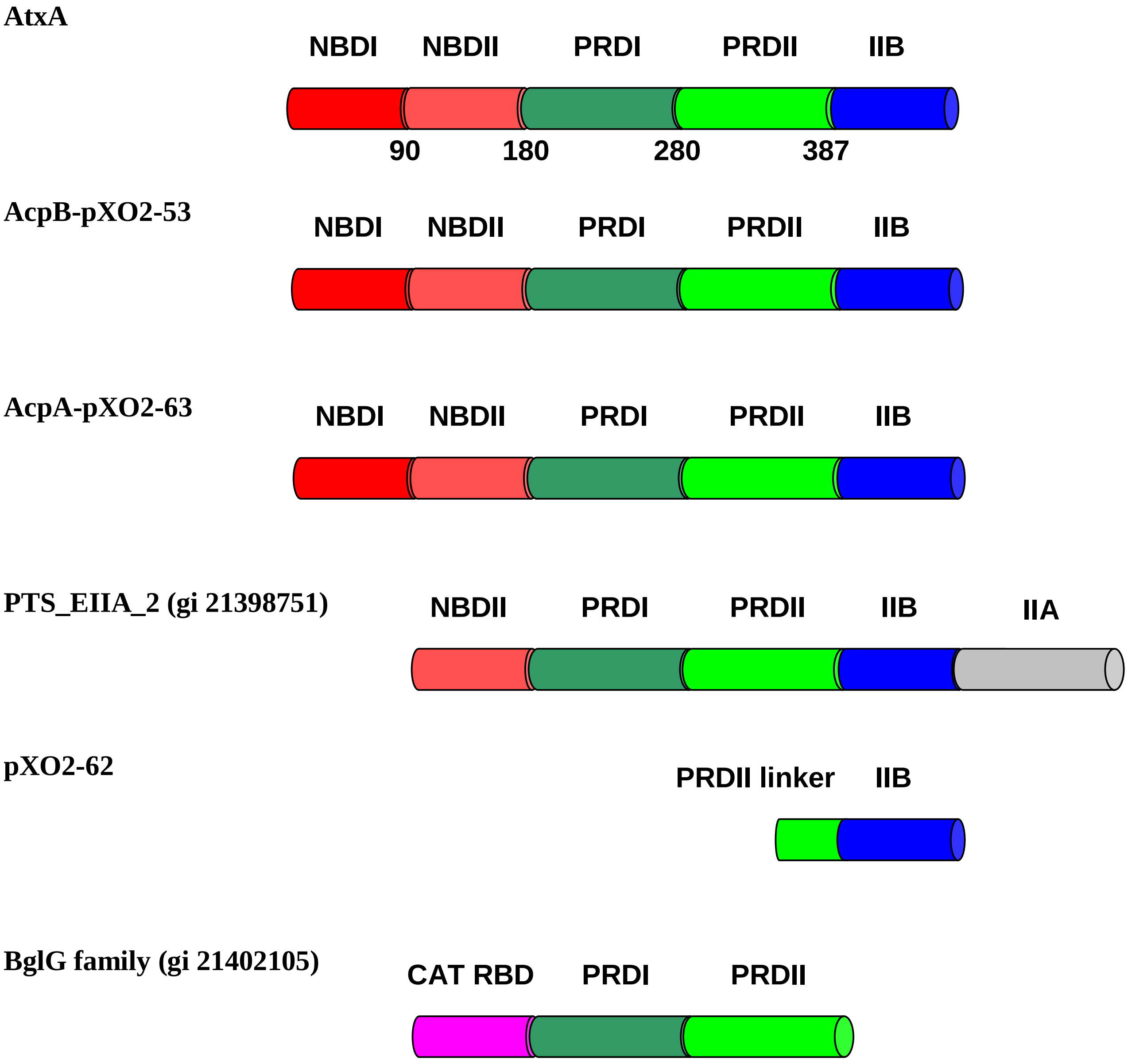

Caulobacter crescentus Type IV pilus assembly system
RcpC
VirB11
TadC
Brucella spp. Type IV secretion system
VirB4
B6

Figure 3



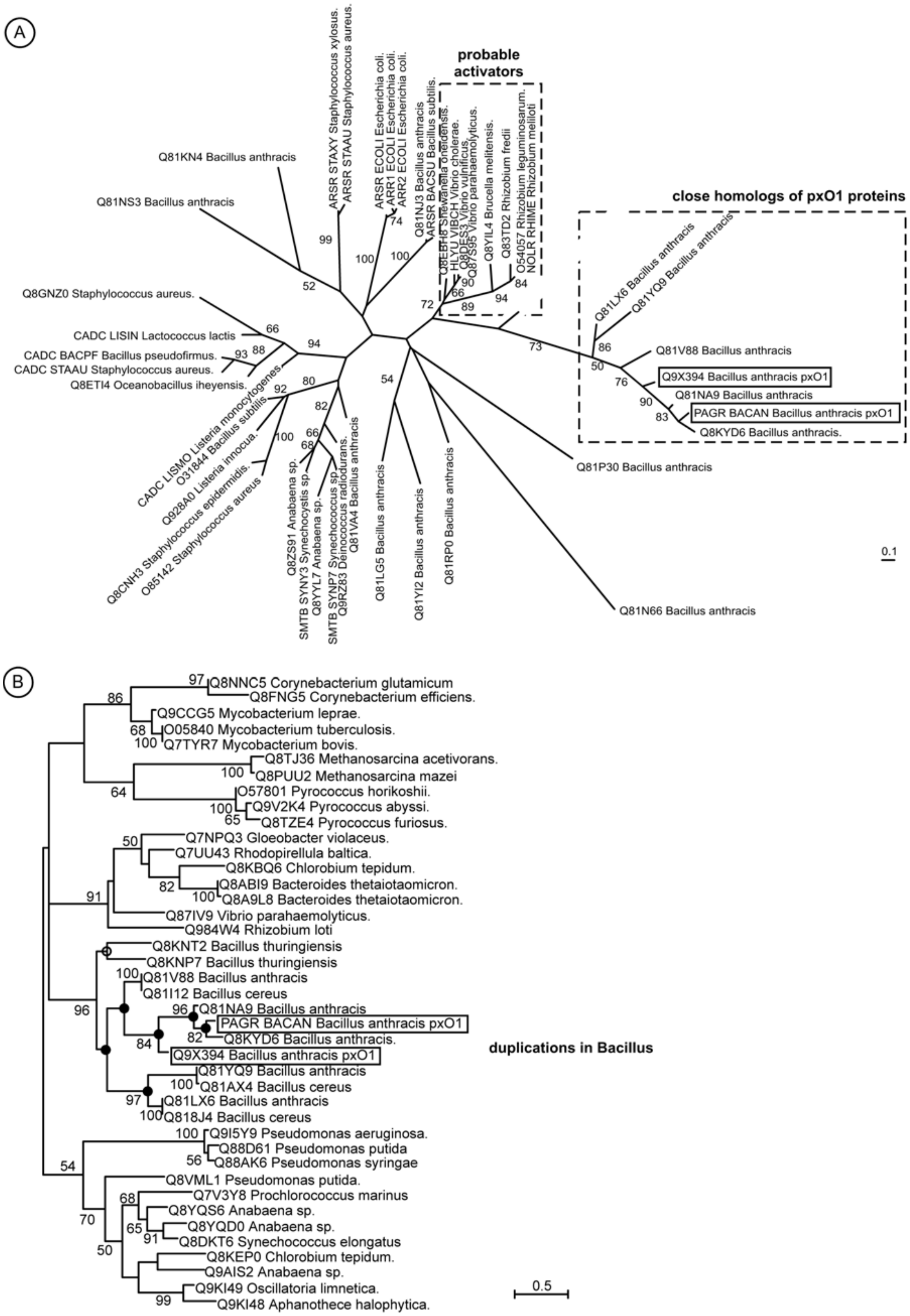
```
pXO1 BXA0139 N-terminus         1  --------MIFTQDGNSLRVESKNNGKSVSNKRHAWSVRVDGKEIYSFHESTKVEQVVDAFNKLPKVKKQGKQFEVVQKSHEIH  76
pXO1 BXA0139 C-terminus        77  RLGSKDGQLTFFKVGSVFSMYAQKPDKSISTKDKNWAVLVDGVIKHTFSKK-TVRDVKASIDKL---KLRGQRFEL-QL----- 150
B.cereus HlyII                336  KLNKGKGKLSLSMNGNQLKATSSNAGYGISYEDKNWGIFVNGEKVYTFNEKSTVGNISNDINKL---NIKGPYIEIKQI----- 412
B.anthracis HlyII              204  QLNKGKGKLSFSMNGNQLKATSSNAGYGISYEDKN*GIFVNGEKVYTFNEKTTVGNISNDINKL---NIKGPYIEIKKI----- 280
Streptococcus phage Cp-1 orf16 33  WQFSGDTDYGYMTNGNTTYIQYGQNDPSVWASMRFWGESVEILEETKNDDNSITAKIR—VKALFWWSKRVSSNAGYRVEYDIK  114
                                                                      *
```

Figure 4



NBDI - nucleic acid binding domain I
NBDII - nucleic acid binding domain II
PRDI - PTS regulatory domain I
PRDII - PTS regulatory domain II
IIB - PTS EIIB homology domain
IIA - PTS EIIA homology domain
CAT RBD - Co-AntiTerminator RNA binding domain

Figure 5


A
probable activators
close homologs of pxO1 proteins
Q81KN4 Bacillus anthracis
Q81NS3 Bacillus anthracis
ARSR STAXY Staphylococcus xylosus.
ARSR STAAU Staphylococcus aureus.
ARSR ECOLI Escherichia coli.
ARR1 ECOLI Escherichia coli.
ARR2 ECOLI Escherichia coli.
Q81NJ3 Bacillus anthracis
ARSR BACSU Bacillus subtilis.
Q8EBH8 Shewanella oneidensis.
HLYU VIBCH Vibrio cholerae.
Q8DES3 Vibrio vulnificus.
Q87S95 Vibrio parahaemolyticus.
Q8YIL4 Brucella melitensis.
Q83TD2 Rhizobium fredii
O54057 Rhizobium leguminosarum.
NOLR RHIME Rhizobium meliloti
Q81LX6 Bacillus anthracis
Q81YQ9 Bacillus anthracis
Q81V88 Bacillus anthracis
Q9X394 Bacillus anthracis pxO1
Q81NA9 Bacillus anthracis
PAGR BACAN Bacillus anthracis pxO1
Q8KYD6 Bacillus anthracis.
Q81P30 Bacillus anthracis
Q81N66 Bacillus anthracis
Q8GNZ0 Staphylococcus aureus.
CADC LISIN Lactococcus lactis
CADC BACPF Bacillus pseudofirmus.
CADC STAAU Staphylococcus aureus.
Q8ETI4 Oceanobacillus iheyensis.
CADC LISMO Listeria monocytogenes
O31844 Bacillus subtilis
Q928A0 Listeria innocua.
Q8CNH3 Staphylococcus epidermidis.
O85142 Staphylococcus aureus
Q8ZS91 Anabaena sp.
SMTB SYNY3 Synechocystis sp.
Q8YYL7 Anabaena sp.
SMTB SYNP7 Synechococcus sp.
Q9RZ83 Deinococcus radiodurans.
Q81VA4 Bacillus anthracis
Q81LG5 Bacillus anthracis
Q81YI2 Bacillus anthracis
Q81RP0 Bacillus anthracis
99
52
100
74
100
72
66
90
89
94
84
73
86
50
76
90
83
66
93
88
94
80
92
82
66
68
100
54
0.1
B
97 Q8NNC5 Corynebacterium glutamicum
Q8FNG5 Corynebacterium efficiens.
86
Q9CCG5 Mycobacterium leprae.
68 O05840 Mycobacterium tuberculosis.
100 Q7TYR7 Mycobacterium bovis.
Q8TJ36 Methanosarcina acetivorans.
100 Q8PUU2 Methanosarcina mazei
64
O57801 Pyrococcus horikoshii.
100 Q9V2K4 Pyrococcus abyssi.
65 Q8TZE4 Pyrococcus furiosus.
50 Q7NPQ3 Gloeobacter violaceus.
Q7UU43 Rhodopirellula baltica.
Q8KBQ6 Chlorobium tepidum.
82 Q8ABI9 Bacteroides thetaiotaomicron.
100 Q8A9L8 Bacteroides thetaiotaomicron.
91
Q87IV9 Vibrio parahaemolyticus.
Q984W4 Rhizobium loti
Q8KNT2 Bacillus thuringiensis
Q8KNP7 Bacillus thuringiensis
100 Q81V88 Bacillus anthracis
Q81I12 Bacillus cereus
96
96 Q81NA9 Bacillus anthracis
PAGR BACAN Bacillus anthracis pxO1
82 Q8KYD6 Bacillus anthracis.
84
Q9X394 Bacillus anthracis pxO1
duplications in Bacillus
Q81YQ9 Bacillus anthracis
100 Q81AX4 Bacillus cereus
97 Q81LX6 Bacillus anthracis
100 Q818J4 Bacillus cereus
100 Q9I5Y9 Pseudomonas aeruginosa.
Q88D61 Pseudomonas putida
56 Q88AK6 Pseudomonas syringae
54
Q8VML1 Pseudomonas putida.
68 Q7V3Y8 Prochlorococcus marinus
70
Q8YQS6 Anabaena sp.
65 Q8YQD0 Anabaena sp.
91 Q8DKT6 Synechococcus elongatus
50
Q8KEP0 Chlorobium tepidum.
Q9AIS2 Anabaena sp.
Q9KI49 Oscillatoria limnetica.
99 Q9KI48 Aphanothece halophytica.
0.5

NERD: a DNA processing-related domain present in the anthrax virulence plasmid, pXO1

Marcin Grynberg and Adam Godzik

Program in Bioinformatics and Systems Biology, The Burnham Institute, 10901 N. Torrey Pines Road, La Jolla, CA 92037, USA

Corresponding author: Adam Godzik (adam@burnham-inst.org).

We have identified a new domain in a broad range of bacterial, as well as single archaeal and plant proteins. Its presence in the virulence-related pXO1 plasmid of *Bacillus anthracis* as well as in several other pathogens makes it a possible drug target. We term the new domain nuclease-related domain (NERD) because of its distant similarity to endonucleases.

Anthrax, a disease of herbivores and primates (including humans), is caused by a gram-positive, spore-forming bacterium, *Bacillus anthracis*. The virulence of this bacterium is dependent on two megaplasmids: pXO1, which is required for the synthesis of the toxin protein [1]; and pXO2, which is required for the synthesis of an anti-phagocytic capsule [2–4]. Strains lacking either of the two megaplasmids are avirulent.

The pXO1 plasmid has been analyzed in several recent genome sequence studies [5–7], by using standard tools such as BLAST. Using sensitive homology-detection algorithms, we have found that a 117-amino acid fragment of the pXO1–01 protein, previously annotated as a hypothetical protein, defines a new domain that is shared by multiple proteins in other eubacteria and is also present in small numbers in archaea and plant proteins. We call it NERD for nuclease-related domain.

The NERD domain

Starting from the amino acid sequence of the *B. anthracis* pXO1–01 protein, a cascade of PSI–BLAST searches [8] identified >40 proteins with a region displaying statistically significant sequence similarity to the seed protein and to each other (Figure 1) and with varied domain combinations (Figure 2). The NERD domain partly overlaps two Pfam-B domains – Pfam-B_22501 and Pfam-B_26882 [9]. However, the Pfam-B families contain only a few sequences (5 and 4, respectively) with single domain context each. An alignment of NERD is presented in Figure 1 and covers 117 amino acids.

The NERD sequence is characterized by three conserved regions interspersed among weakly conserved or very diverse regions (Figure 1). Conserved hydrophobic, mainly aliphatic motifs (consisting of Leu, Ile and Val) and polar, mainly charged positions (e.g. Asp, His, Glu and Lys), alternate in the alignment. The first and most conserved region is formed by the N-terminal Glu followed by the [Gln/Glu]-[Ile/Val/Leu]-Asp motif, then a stretch of hydrophobic residues with two polar (Glu and Lys) and two hydrophobic (Gly and [Ile/Leu/Val]) residues at the end. The next 20 amino acids are not conserved, but the [Ser/Asn]-Pro-[Ile/Leu/Val/Met] motif with a neighboring Gln form a second conserved region. The third is at the C-terminal 25 amino acids, with mainly the hydrophobic amino acids conserved. An interesting feature of NERD is the existence of subgroups that have no conservation in motifs that are conserved in all other members of the family (e.g. two N-terminal glycine residues are missing in the plant domain) or with a charge difference (e.g. Glu instead of Gln in the most conserved [Gln/Glu]-[Ile/Val/Leu]-Asp motif). We can only hypothesize that these differences account for functional diversity within the NERD family.

The predicted α–β–β–β–β–(weak β/long loop)-α–β–β secondary structure of NERD domain helps rationalize the conservation of specific regions of the domain (Figure 1) because all the conserved residues coincide with secondary-structure elements, especially the third and fourth β strands. The only exception is the fifth β strand, which is likely to be a terminal strand or a long loop (Figure 1).

NERD-domain associations

The majority of NERD-containing proteins are single-domain, in several cases with additional (predicted) transmembrane helices. In only a few instances, proteins containing NERD have additional domains that, in 75% of these cases, are involved in DNA processing. In all cases in which NERD is present in multidomain proteins, it is found at the N terminus. There is also no evident operon conservation for NERD-containing proteins and no apparent connection between phyla and domain fusions.

Most NERD-containing proteins, including the group-defining *B. anthracis* pXO1–01 protein, consist entirely of the NERD domain, sometimes with short tails of several amino acids on both C and N termini. All proteins in this group are hypothetical open reading frames (ORFs). In addition, in several proteins the NERD domain is associated with one or two predicted transmembrane motifs, which could be located either at the N or C terminus (Figure 2).

In a hypothetical *Clostridium perfringens* protein (gi: 18309656), the NERD domain is followed by the helicase and RNaseD C-terminal (HDRC) domain (PF00570; Figure 2). HRDC is an 80-amino acid protein domain usually found at the C terminus of RecQ helicases and RNase D homologs from various organisms, including human [10]. An HRDC domain is present in genes linked to the human diseases Werner and Bloom syndromes [11,12]. The HRDC domain is involved in the binding of DNA to specific DNA structures (e.g. long-forked duplexes and Holliday junctions) that are formed during replication, recombination or transcription [13]. Interestingly, in the many HRDC-containing proteins, the N-terminal region in the 3′→5′ exonuclease domain (PF01612) that is responsible for the 3′→5′ exonuclease proofreading activity of the DNA polymerase I and other enzymes and catalyzes the hydrolysis of unpaired or mismatched nucleotides [14,15]. One can speculate that NERD, existing in analogous arrangement with the HRDC domain, has a related function.

In at least three proteins, including the hypothetical protein (gi: 22972752) from *Chloroflexus aurantiacus*, the NERD domain is found at the N terminus of the UvrD/Rep 3′→5′ DNA helicases (PF00580), which catalyze the ATP-dependent unwinding of double-stranded to single-stranded DNA (ssDNA) [16]. DNA helicases are essential for processes such as DNA replication, recombination and repair [17]. This domain co-occurs with the HRDC domain in several bacterial species (i.e. *Streptomyces coelicolor*, *Corynebacterium glutamicum*, *Mycobacterium leprae* and *Mycobacterium tuberculosis*).

In two proteins, in *Pseudomonas aeruginosa* (gi: 4406504) and the *Bacteroides* (gi: 8308027), NERD is followed by the DNA-binding C4 zinc finger (PF01396), which is a short motif present in two NERD proteins (Figure 2), usually a C-terminal region of prokaryotic topoisomerases I [18]. The role of topoisomerase in the bacterial cell is to remove excessive negative supercoils from DNA to maintain the optimal superhelical state [19]. The zinc motifs do not cleave or recognize the topoisomerase substrate, rather, they are believed to interact with ssDNA to relax negatively supercoiled DNA [20]. Apart from topoisomerases, there are a few proteins with proximally located restriction endonucleases (PF04471) or unknown N termini that possess the C4 zinc fingers. However, their role is unknown.

In five proteins, the NERD domain is followed by two STYKc domains (PF00069). STYKcs are protein kinases with possible dual serine, threonine and tyrosine kinase specificity [21]. For example, in the cases of *Thermomonospora fusca* and *Streptomyces coelicolor*, there are genomic associations with DNA polymerase III and transposase, and an adenine-specific methyltransferase, respectively, which can suggest a nucleotide-related function of these large proteins (ERGO database: http://ergo.integratedgenomics.com).

In most cases, only one copy of the NERD domain is present in a given organism. We found that in only three bacteria there are two copies of NERD per genome (in *Burkholderia fungorum*, *Oceanobacillus iheyensis* and *Desulfitobacterium hafniense*).

pXO1–01 function

None of the NERD-containing proteins have been studied by experiment, therefore, its exact function is not known. However, bioinformatics analyses offer some clues.

The closest homolog of pXO1–01 is the orf8 protein from *Bacteroides* spp. It is an ORF from the non-replicating *Bacteroides* unit 1 (NBU1) , a 10.3-kbp integrated element that can be excised and mobilized in *trans* by tetracycline-inducible *Bacteroides* conjugative transposons [22,23]. The elements responsible for integration and excision were recognized [24–26], but *orf8* is probably not involved in these processes. The large G+C content difference between *orf6*, *orf7* and *orf8* (35%), and other *Bacteroides* genes (42%) suggests a possible recent acquisition that is involved in a yet-undiscovered transposition process. The presence of NERD in a unique archaeal and only two plant species supports such a transposon-type transfer of the domain.

A more detailed prediction can be made based on the domain structure similarity between NERD proteins that contain the HRDC domains and the N-terminal region of exonuclease proteins that contain the HRDC domains **{Au: correction was not clear, is edit correct?}**. This is further supported by distant homology between NERD and the COG0792 family, a predicted endonuclease family distantly related to archaeal Holliday junction resolvase, members of which are involved in DNA replication and/or recombination, and/or repair. This homology is predicted by a profile–profile search algorithm FFAS (fold and function assignment system) [27], albeit with low statistical significance. Several fold-recognition algorithms (e.g. Superfamily and BASIC) [27–29] identify matches to the Holliday junction resolvase structure (PDB codes: 1gefA and 1hh1A) with statistically significant scores [30,31]. The alignment between the NERD and COG0792 families and the sequence of the Holiday junction resolvase (PDB code: 1gefA) is shown in Figure 1 (both alignments were obtained by the FFAS [27] algorithm). The alignment covers only the N-terminal half of NERD, and the 3D model of this is shown in Figure 3. Interestingly, all active-site residues of resolvase (black arrows in the alignment and residues shown in atomic detail in the Figure 3) are conserved in most NERD family members, which strongly supports the functional prediction. The common denominator of all these predictions suggests a nuclease function for NERD.

Concluding remarks

We have discovered a novel domain, NERD, with predicted connection to DNA processing. Genomic context analysis and distant homology analysis suggest a nuclease function.

The finding of this domain is important for the understanding of anthrax virulence. The location of pXO1–01 in the vicinity of other DNA processing-related ORFs, on the anthrax virulence plasmid, suggests an orchestrated function of the products of these genes. Is this machinery an anthrax DNA-remodeling system or is it involved in the eukaryotic cell attack? Maybe further advances in the studies of the NBU1 element will reveal its function.

The presence of NERD in only few non-bacterial species not only suggests that this domain might be involved in some mobility processes, but also that the species transfer must have happened quite recently.

Acknowledgements

We thank Marc Robinson-Rechavi for his critical reading of the article. This work was supported by the National Institutes of Health, grant GM60049.

References

1 Vodkin, M.H. and Leppla, S.H. (1983) Cloning of the protective antigen gene of *Bacillus anthracis*. *Cell* 34, 693–697

2 Uchida, I. *et al.* (1985) Association of the encapsulation of *Bacillus anthracis* with a 60 megadalton plasmid. *J. Gen. Microbiol.* 131, 363–367

3 Mikesell, P. *et al.* (1983) Evidence for plasmid-mediated toxin production in *Bacillus anthracis*. *Infect. Immun.* 39, 371–376

4 Green, B.D. *et al.* (1985) Demonstration of a capsule plasmid in *Bacillus anthracis*. *Infect. Immun.* 49, 291–297

5 Pannucci, J. *et al.* (2002) *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. *J. Bacteriol.* 184, 134–141

6 Okinaka, R.T. *et al.* (1999) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J. Bacteriol.* 181, 6509–6515

7 Read, T.D. *et al.* (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423, 81–86

8 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI–BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402

9 Bateman, A. *et al.* (2002) The Pfam protein families database. *Nucleic Acids Res.* 30, 276–280

10 Morozov, V. *et al.* (1997) A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases. *Trends Biochem. Sci.* 22, 417–418

11 Yu, C.E. *et al.* (1996) Positional cloning of the Werner's syndrome gene. *Science* 272, 258–262

12 Ellis, N.A. *et al.* (1995) The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* 83, 655–666

13 von Kobbe, C. *et al.* (2003) Werner syndrome protein contains three structure specific DNA binding domains. *J. Biol. Chem.* 278, 52997–53006

14 Moser, M.J. *et al.* (1997) The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic Acids Res.* 25, 5110–5118

15 Joyce, C.M. and Steitz, T.A. (1994) Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63, 777–822

16 Hickson, I.D. *et al.* (1983) The *E. coli uvrD* gene product is DNA helicase II. *Mol. Gen. Genet.* 190, 265–270

17 Matson, S.W. *et al.* (1994) DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *Bioessays* 16, 13–22

18 Tse-Dinh, Y.C. and Beran-Steed, R.K. (1988) *Escherichia coli* DNA topoisomerase I is a zinc metalloprotein with three repetitive zinc-binding domains. *J. Biol. Chem.* 263, 15857–15859

19 Drlica, K. (1990) Bacterial topoisomerases and the control of DNA supercoiling. *Trends Genet.* 6, 433–437

20 Ahumada, A. and Tse-Dinh, Y.C. (2002) The role of the Zn(II) binding domain in the mechanism of *E. coli* DNA topoisomerase I. *BMC Biochem.* 3, 13

21 Hanks, S.K. and Hunter, T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* 9, 576–596

22 Li, L.Y. *et al.* (1993) Characterization of the mobilization region of a *Bacteroides* insertion element (NBU1) that is excised and transferred by *Bacteroides* conjugative transposons. *J. Bacteriol.* 175, 6588–6598

23 Shoemaker, N.B. and Salyers, A.A. (1988) Tetracycline-dependent appearance of plasmidlike forms in *Bacteroides uniformis* 0061 mediated by conjugal *Bacteroides* tetracycline resistance elements. *J. Bacteriol.* 170, 1651–1657

24 Shoemaker, N.B. *et al.* (1996) NBU1, a mobilizable site-specific integrated element from *Bacteroides spp.*, can integrate nonspecifically in *Escherichia coli*. *J. Bacteriol.* 178, 3601–3607

25 Shoemaker, N.B. *et al.* (1996) The *Bacteroides* mobilizable insertion element, NBU1, integrates into the 3′ end of a Leu-tRNA gene and has an integrase that is a member of the λ integrase family. *J. Bacteriol.* 178, 3594–3600

26 Shoemaker, N.B. *et al.* (2000) Multiple gene products and sequences required for excision of the mobilizable integrated *Bacteroides* element NBU1. *J. Bacteriol.* 182, 928–936

27 Rychlewski, L. *et al.* (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci.* 9, 232–241

28 Ginalski, K. *et al.* (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 19, 1015–1018

29 Gough, J. *et al.* (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* 313, 903–919

30 Nishino, T. *et al.* (2001) Crystal structure of the archaeal Holliday junction resolvase Hjc and implications for DNA recognition. *Structure (Camb)* 9, 197–204

31 Bond, C.S. *et al.* (2001) Structure of Hjc, a Holliday junction resolvase, from *Sulfolobus solfataricus*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5509–5514

32 Nikolaev, V.K. *et al.* (1997) Building multiple alignment using iterative analyzing biopolymers structure dynamic improvement of the initial motif alignment. *Biochemistry* 62, 578–582

33 Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41, 95–98

34 McGuffin, L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404–405

35 Karplus, K. *et al.* (1999) Predicting protein structure using only sequence information. *Proteins* 37 (Suppl. 3), 121–125

36 Rost, B. and Eyrich, V.A. (2001) EVA: large-scale analysis of secondary structure prediction. *Proteins* 45 (Suppl. 5), 192–199

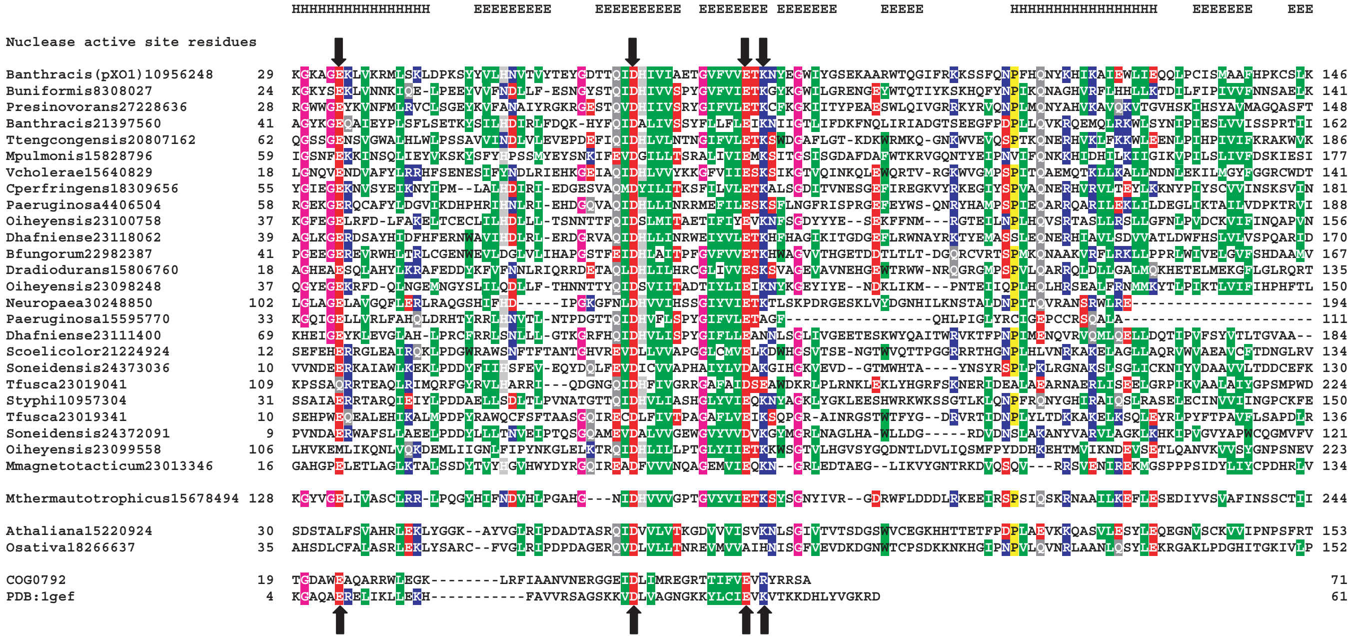
37 States, D.J. and Gish, W. (1994) Combined use of sequence similarity and codon bias for coding region identification. *J. Comput. Biol.* 1, 39–50

38 Gish, W. and States, D.J. (1993) Identification of protein coding regions by database similarity search. *Nat. Genet.* 3, 266–272

39 Letunic, I. *et al.* (2002) Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* 30, 242–244

40 Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815

41 The PyMOL Molecular Graphics System (2002) http://www.pymol.org



```
                                                HHHHHHHHHHHHHHHH     EEEEEEEEE     EEEEEEEEEE  EEEEEEEE EEEEEEE     EEEEE            HHHHHHHHHHHHHHHHH     EEEEEEE     EEE

Nuclease active site residues

Banthracis(pXO1)10956248      29   KGKAGEKLVKRMLSKLDPKSYYVLHNVTVYTEYGDTTQIDHIVIAETGVFVVETKNYEGWIYGSEKAARWTQGIFRKKSSFQNPFHQNYKHIKAIEWLIEQQLPCISMAAFHPKCSLK 146
Buniformis8308027             24   KGKYSEKLVNNKIQE-LPEEYVVFNDLLF-ESNGYSTQIDHIVVSPYGVFVIETKGYKGWILGRENGEYWTQTIYKSKHQFYNPIKQNAGHVRFLHHLLKTDILFIPIVVFNNSAELK 141
Presinovorans27228636         28   RGWWGEYKVNFMLRVCLSGEYKVFANAIYRGKRGESTQVDHIIVSRYGIFVLETKCFKGKIITYPEAESWLQIVGRRKYRVQNPLMQNYAHVKAVQKVTGVHSKIHSYAVMAGQASFT 148
Banthracis21397560            41   AGYKGEQAIEYPLSFLSETKYSILHDIRLFDQK-HYFQIDALIVSSYFLLFLEIKNIIGTLIFDKFNQLIRIADGTSEEGFPDPLLQVKRQEMQLRKWLSYNIPIESLVVISSPRTII 162
Ttengcongensis20807162        62   QGSSGENSVGWALHLWLPSSAVVINDLVLEVEPDEFIQIDHLVLNTNGIFVLETKSWDGAFLGT-KDKWRMKQ-GNKWVEVQSPTKQNERHVKLFKKWLEENLPIHPIVIFKRAKWVK 186
Mpulmonis15828796             59   IGSNFEKKINSQLIEYVKSKYSFYHPSSMYEYSNKIFEVDGILLTSRALIVIEMKSITGSISGDAFDAFWTKRVGQNTYEIPNVIFQNKKHIDHILKIIGIKVPILSLIVFDSKIESI 177
Vcholerae15640829             18   LGNQVENDVAFYLRRHFSENESIFYNDLRIEHKGEIAQIDHLVVYKKGFVIIESKSIKGTVQINKQLEWQRTV-RGKWVGMPSPITQAEMQTKLLKALLNDNLEKILMGYFGGRCWDT 141
Cperfringens18309656          55   YGIEGEKNVSYEIKNYIPM--LALHDIRI-EDGESVAQMDYILITKSFILVLETKALSGDITVNESGEFIREGKVYRKEGIYSPVAQNERHVRVLTEYLKKNYPIYSCVVINSKSVIN 181
Paeruginosa4406504            58   RGEKGERQCAFYLDGVIKDHPHRIHNLRI-EHDGQVAQIDHLLINRRMEFILESKSFLNGFRISPRGEFEYWS-QNRYHAMPSPIEQARRQARILEKLILDEGLIKTAILVDPKTRVI 188
Oiheyensis23100758            37   KGFEGELRFD-LFAKELTCECLILHDLLL-TSNNTTFQIDSLMITAETIFIYEVKNFSGDYYYE--SEKFFNM---RGTEILNPLHQVSRTASLLRSLLGFNLPVDCKVIFINQAPVN 156
Dhafniense23118062            39   AGLKGERDSAYHIDFHFERNWAVIHDLRL-ERDGRVAQIDHLLINRWEIYVLETKHFHAGIKITGDGEFLRWNAYRKTYEMASSLEQNERHIAVLSDVVATLDWFHSLVLVSPQARID 170
Bfungorum22982387             41   PGEEGEREVRWHLTRLCGENWEVLDGLVLIHAPGSTFEIDHLAITPFGVFVVETKHWAGVVTHGETDDTLTLT-DGQRCVRTSPMKQNAAKVRFLRKLLPPRLWIVELGVFSHDAAMV 167
Dradiodurans15806760          18   AGHEAESQLAHYLKRAFEDDYKFVFNNLRIQRRDETAQLDHLILHRCGLIVVESKSVAGEVAVNEHGEWTRWW-NRQGRGMPSPVLQARRQLDLLGALMQKHETELMEKGFLGLRQRT 135
Oiheyensis23098248            37   QGYEGEKRFD-QLNGEMNGYSLILQDLLF-THNNTTYQIDSVIITADTIYLIEIKNYKGEYIYE--NDKLIKM---PNTEIIQPLHQLHRSEALFRNMMKYTLPIKTLVIFIHPHFTL 150
Neuropaea30248850            102   LGLAGELAVGQFLERLRAQGSHIFHD-----IPGKGFNLDHVVIHSSGIYVIETKTLSKPDRGESKLVYDGNHILKNSTALDNPITQVRANSRWLRE--------------------- 194
Paeruginosa15595770           33   KGQIGELLVRLFAHQLDRHTYRRLHNVTL-NTPDGTTQIDHVFLSPYGIFVLETAGF----------------QHLPIGLYRCIGEPCCRSQALA--------------------- 111
Dhafniense23111400            69   KHEIGEYKLEVGLAH-LPRCFRRLSNLLL-GTKGRFHQIDHVLISPYGIFLLEANNLSGLIVGEETESKWYQAITWRVKTFPNPIMENQVRVQMLQELLDQTIPVFSYVTLTGVAA-- 184
Scoelicolor21224924           12   SEFEHERRGLEAIRQKLPDGWRAWSNFTFTANTGHVREVDLLVVAPGGLCMVELKDWHGSVTSE-NGTWVQTTPGGRRRTHGNPLHLVNRKAKELAGLLAQRVWVAEAVCFTDNGLRV 134
Soneidensis24373036           10   VVNDEERKAIAWLKEKLPDDYFIIHSFEV-EQYDQLFEVDICVVAPHAIYLVDAKGIHGKVEVD--GTMWHTA----YNSYRSPLPKLRGNAKSLSGLICKNIYVDAAVVLTDDCEFK 130
Tfusca23019041               109   KPSSAQRRTEAQLRIMQRFGYRVLHARRI---QDGNGQIDHFIVGRRGAFAIDSEAWDKRLPLRNKLEKLYHGRFSKNERIDEALAEARNAERLISEELGRPIKVAALAIYGPSMPWD 224
Styphi10957304                31   SSAIAERRTARQIEIYLPDDAELLSDLTLPVNATGTTQIDHVLIASHGLYVIEQKNYAGKLYGKLEESHWRKWKSSGTLKLQNPFRQNYGHIRAIQSLRASELECINVVIINGPCKFE 150
Tfusca23019341                10   SEHPWEQEALEHIKALMPDPYRAWQCFSFTAASGQIRECDLFIVTPAGAFLVEIKSQPGR-AINRGSTWTFYG--DRVRTIDNPLYLTDKKAKELKSQLEYRLPYFTPAVFLSAPDLR 136
Soneidensis24372091            9   PVNDAERWAFSLLAEELPDDYLLLTNVEIPTQSGQAMEVDALVVGEWGVYVVDVKGYMGRLNAGLHA-WLLDG-----RDVDNSLAKANYVARVLAGKLKHKIPVGVYAPWCQGMVFV 121
Oiheyensis23099558           106   LHVKEMLIKQNLVQKDEMLIIGNLFIPYNKGLELKTRQIDHLILLPTGLYIIETKKWSGTVLHGVSYGQDNTLDVLIQSMFPYDDHKEHTMVIKNDEVSETLQANVKVVSYGNPSNEV 223
Mmagnetotacticum23013346      16   GAHGPELETLAGLKTALSSDYTVYHGVHWYDYRGQIREADFVVVNQAGEMVIEQKN--GRLEDTAEG--LIKVYGNTRKDVQSQV---RRSVENIREKMGSPPPSIDYLIYCPDHRLV 134

Mthermautotrophicus15678494 128   KGYVGELIVASCLRR-LPQGYHIFNDVHLPGAHG---NIDHVVVGPTGVYVIETKSYSGNYIVR--GDRWFLDDDLRKEEIRSPSIQSKRNAAILKEFLESEDIYVSVAFINSSCTII 244

Athaliana15220924             30   SDSTALFSVAHRLEKLYGGK--AYVGLRIPDADTASRQIDVVLVTKGDVVVISVKNLSGIVTVTSDGSWVCEGKHHTTETFPDPLAEVKKQASVLESYLEQEGNVSCKVVIPNPSFRT 153
Osativa18266637               35   AHSDLCFALASRLEKLYSARC--FVGLRIPDPDAGERQVDLVLLTNREVMVVAIHNISGFVEVDKDGNWTCPSDKKNKHGIPNPVLQVNRLAANLQSYLEKRGAKLPDGHITGKIVLP 152

COG0792                       19   TGDAWEAQARRWLEGK--------LRFIAANVNERGGEIDLIMREGRTTIFVEVRYRRSA                                                            71
PDB:1gef                       4   KGAQAERELIKLLEKH-----------FAVVRSAGSKKVDLVAGNGKKYLCIEVKVTKKDHLYVGKRD                                                    61
```

Figure 1. An alignment of a sample set of NERD (nuclease-related domain) sequences. The alignment was generated using AliBee (http://www.genebee.msu.su/services/malign_full.html) [32] and colored in BioEdit [33]. PSI–BLAST [8] searches of the nonredundant protein database using the *Bacillus anthracis* pXO1–01 protein (gi: 10956248) as query were performed using the default parameters. After five rounds of searching, representatives of all subgroups of NERD were found. The highest E value was 2e-06. The uppermost group is composed of prokaryotic proteins, the middle protein is the sole example of an archaeal NERD-containing protein and the lowermost are two plant proteins. The shading threshold is 40%. The alignment is colored according to identity and similarity according to the default BioEdit amino acid similarity scoring matrix. The secondary-structure prediction is given for pXO1–01 as a combined result of PSIPRED [34], Sam-T99–2d [35] and Profsec [36] at the MetaServer (http://bioinfo.pl/Meta/) [28]. The results for other members of the NERD family are almost identical. Arrows indicate the conserved residues that are important for the endonuclease activity of resolvases. Two shorter sequences (*Nitrosomonas europaea* Q82W50 and *Pseudomonas aeruginosa* Q9I5W3) are coded by genomic sequences with a stop codon and no sequence homology beyond the stop codon, when checked using the BLASTX program [37,38] at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). It seems that the N-terminal catalytic domain is sufficient for their function. Sequences shown are (species name, gi number, in brackets are the first and last positions in the sequences aligned): Banthracis (pXO1)10956248, *Bacillus anthracis* Q8KYT4 (29–146); Buniformis8308027, *Bacteroides uniformis* Q9KIA1 (24–141); Presinovorans27228636, *Pseudomonas resinovorans* Q8GHQ8; Banthracis21397560, *Bacillus anthracis* Q81XB5 (41–162); Ttengcongensis20807162, *Thermoanaerobacter tengcongensis* Q8RBY3 (62–186); Mpulmonis15828796, *Mycoplasma pulmonis* Q98QN6 (59–177); Vcholerae15640829, *Vibrio cholerae* Q9KTS7 (18–141); Cperfringens18309656, *Clostridium perfringens* Q8XML4 (55–181); Paeruginosa4406504, *Pseudomonas aeruginosa* AAD20003 (58–188); Oiheyensis23100758, *Oceanobacillus iheyensis* Q8ELC6 (37–156); Dhafniense23118062, *Desulfitobacterium hafniense* ZP_00101791 (39–170); Bfungorum22982387, *Burkholderia fungorum* ZP_00027654 (41–167); Dradiodurans15806760, *Deinococcus radiodurans* Q9RTK3 (18–135); Oiheyensis23098248, *Oceanobacillus iheyensis* Q8ES50 (37–150); Neuropaea30248850, *Nitrosomonas europaea* Q82W50 (102–194); Paeruginosa15595770, *Pseudomonas aeruginosa* Q9I5W3 (33–111); Dhafniense23111400, *Desulfitobacterium hafniense* ZP_00097061 (69–184); Scoelicolor21224924, *Streptomyces coelicolor* O86560 (12–134); Soneidensis24373036, *Shewanella oneidensis* Q8EGX7 (10–130); Tfusca23019041, *Thermobifida fusca* ZP_00058754 (109–224); Styphi10957304, *Salmonella typhi* Q9L5M7 (31–150); Tfusca23019341, *Thermobifida fusca* ZP_00059052 (109–224); Soneidensis24372091, *Shewanella oneidensis* Q8EJH0 (9–121); Oiheyensis23099558, *Oceanobacillus iheyensis* Q8EPJ8 (106–223); Mmagnetotacticum23013346, *Magnetospirillum magnetotacticum* (16–134); Mthermautotrophicus15678494, *Methanothermobacter thermautotrophicus* O26566 (128–244); Athaliana15220924, *Arabidopsis thaliana* Q9SS58 (30–153); Osativa18266637, *Oryza sativa* Q8W3G9 (35–152). This multiple sequence alignment (alignment number ALIGN_000650) has been deposited with the European Bioinformatics Institute (ftp://ftp.ebi.ac.uk/pub/databases/embl/align/ALIGN_000650).
]

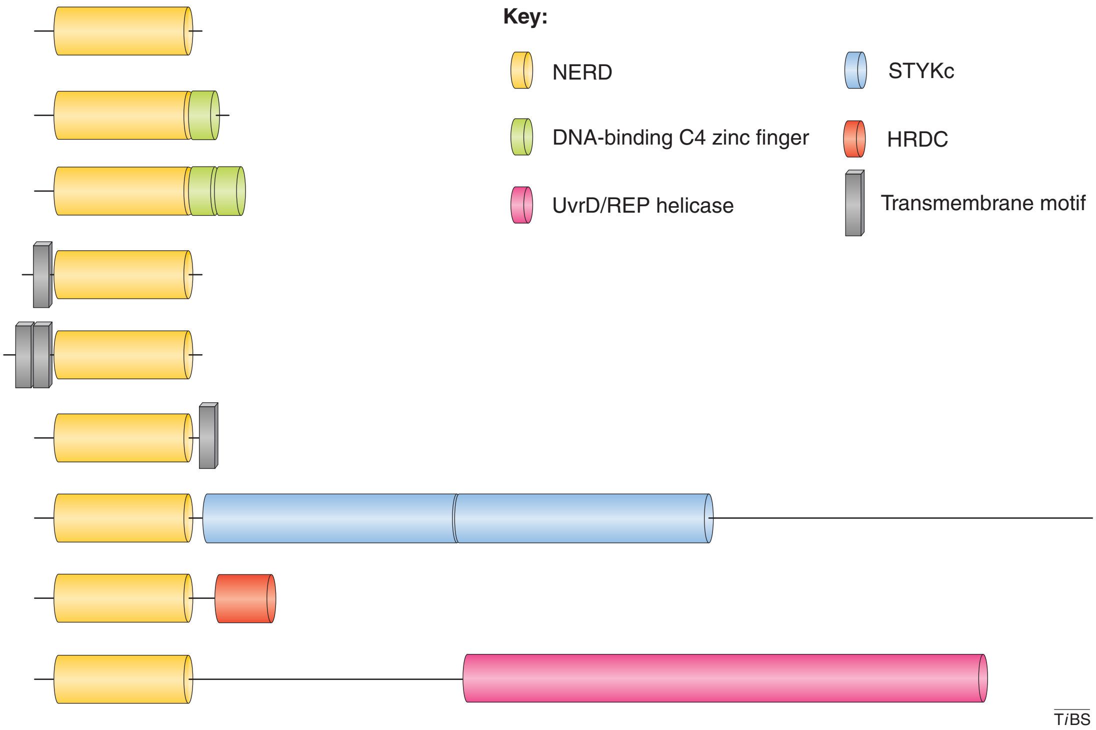


Figure 2. The domain architecture of NERD (nuclease-related domain)-containing proteins. In all cases of multidomain proteins, NERD is located in the N terminus. All domains were recognized using the simple modular architecture research tool (SMART) server (http://smart.embl-heidelberg.de/ or http://smart.ox.ac.uk/) [39]. In case of long proteins, the size of domains is not proportional to protein length.

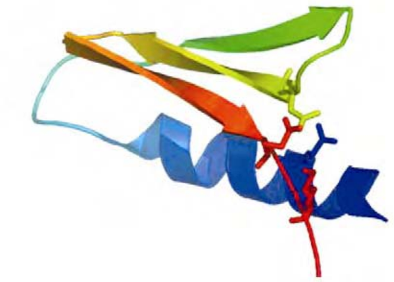


Figure 3. The predicted structure of NERD (nuclease-related domain). The pXO1–01 model was obtained with the Modeller comparative modelling suite [40], on the basis of the FFAS (fold and function assignment system) [27] alignment. The ribbon diagram was prepared using Pymol [41].

DRAFT MANUSCRIPT TO BE SUBMITTED

Crystal structure of virulence plasmid-encoded sensor domains inhibiting sporulation in *Bacillus anthracis*

Gudrun R. Stranzl [1], Marcin Grynberg [1,3], Chandra La Clair [2], Dorinda Shoemaker [2],
Robert Schwarzenbacher [1], Eugenio Santelli [1], Adam Godzik [1], Marta Perego [2],
Robert C. Liddington [1,4]

[1]Infectious & Inflammatory Disease Research Center (IIDC), The Burnham Institute,
La Jolla, CA 92037, USA

[2]The Scripps Research Institute, Department of Molecular and Experimental Medicine,
Divison of Cellular Biology, La Jolla, CA 92037, USA

[3]Present Address: Department of Genetics, Institute of Biochemistry and Biophysics,
Pawinskiego 5A, 02106 Warsaw, Poland

[4]Correspondence: rlidding@burnham.org

Running Title: Structure of *B. anthracis* sensor domains

Key Words: *Bacillus anthracis,* sporulation, histidine kinase, sensor domain, two-component system

ABSTRACT

Sporulation in *B. anthracis* is a highly regulated process requiring the integration of positive and negative stimuli through a network of signal transduction pathways. Two virulence plasmid-encoded proteins, pXO1-118 and pXO2-61, sharing homology with the sensor domain of kinase BA2291, were shown to inhibit sporulation during pathogenesis, providing a molecular link between these two antagonistic processes. Here we report the crystal structures of these proteins and suggest that competition with BA2291 for the binding of an unidentified signalling molecule provides a simple mechanism for their inhibitory effect.

INTRODUCTION

Regulation of sporulation in *Bacilli* is coordinated by multiple signals converging on the transcription factor Spo0A through the phosphorelay signal transduction system for sporulation initiation {Burbulys, 1991 3660 /id}. A balance between two phosphorylation states of Spo0A, resulting in either repression or activation of sporulation in response to environmental stimuli, is achieved through the action of several histidine sensor kinases and aspartyl phosphate phosphatases {Perego, 1994 4902 /id;LeDeaux, 1995 5051 /id;Jiang, 2000 6391 /id;Perego, 2001 6450 /id}. Histidine sensor kinases of bacterial two component systems are enzymes generally made of a divergent sensor domain coupled to a dimerization/phosphoacceptor and an ATPase domain (for reviews see {Hoch, 1995 5464 /id}). Binding of a ligand to the sensor domain is believed to activate the protein to autophosphorylate and subsequently transfer the phosphoryl group to a downstream response regulator, in this case Spo0A, through the intermediate activity of the Spo0F response regulator and the Spo0B phosphotransferase of the phosphorelay. Activation of Spo0A by phosphorylation initiates the sporulation developmental program. Many kinases possess phosphatase activity when not engaged by a ligand and can therefore drive the process in the opposite direction in the absence of their specific signal. Nine possible sporulation histidine sensor kinase encoding genes were identified on the chromosome of *Bacillus anthracis* {Brunsing, 2005 7039 /id}. The gene product of five of them was inferred to be capable of inducing sporulation in *in vivo* studies. In particular, BA2291 was found to be capable of complementing *Bacillus subtilis* sporulation kinase-deficient mutants while its absence strongly affected the ability of *B. anthracis* to sporulate. However, its overexpression in *B. subtilis* completely prevented sporulation suggesting that BA2291 can act as a phosphatase on the sporulation phosphorelay.

Spore production is essential for survival in the environment and initiation of infection by *B. anthracis*, but potentially detrimental once infection is established and vegetative growth and toxin production peak. Fully virulent strains of *B. anthracis* contain two plasmids, pXO1 and pXO2, that carry the genes encoding for the major virulence determinants, namely toxin (protective antigen, PA; edema factor, EF; lethal factor, LF) and capsule (*cap*) production, and their regulators at the transcriptional level AtxA, AcpA and AcpB {Okinaka, 1999 6807 /id;Vietri, 1995 6835 /id;Uchida, 1993 7059 /id;Drysdale, 2004 7111 /id}. Previous studies have identified two highly homologous plasmid encoded proteins, pXO1-118 and pXO2-61 with greater than 30% amino-acid identity with the sensor domain of the major sporulation histidine kinase BA2291 {White, 2006 7105 /id} (Fig.1). Furthermore, it was shown that these proteins modulate the activity of BA2291, converting it into a repressor of sporulation when overexpressed in *B. anthracis* or *B. subtilis.* The genes encoding the pXO1-118 and pXO2-61 proteins are located in close proximity and divergently transcribed to those encoding AtxA (pXO1-118) or an *acpA* pseudogene located between the *cap* and *acpA* loci (pXO2-61), suggesting that they might be relevant to the virulence of *B. anthracis* (Fig. 2).

The precise molecular mechanism by which pXO1-118 and pXO2-61 modulate the function of BA2291 has not yet been elucidated, nor have the nature of the ligand inducing autophosphorylation of BA2291 and the mechanism by which the signal is transduced from its sensor to its catalytic domain. To begin addressing these points, the crystal structure of both pXO1-118 and pXO2-61 was solved and features were identified that are consistent with a sensor domain function.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Functional analysis was carried out in the *B. anthracis* Sterne strain 34F2 (pXO1$^+$, pXO2$^-$). Cells were grown in LB medium or Schaeffer's sporulation medium [Schaeffer, 1965 #65]. Transformation by electroporation was carried out according to Koehler *et al.* [Koehler, 1994 #62]. Unmethylated DNA was obtained by passing plasmid constructs into the *dam*$^-$ strain SCS110 (Stratagene). *E. coli* DH5α was used for plasmid construction and propagation. Antibiotics were used at the following concentrations in *E. coli* or *B. anthracis*, respectively: kanamycin 30μg/ml and 7.5μg/ml; chloramphenicol 10μg/ml and 7.5μg/ml; spectinomycin 100μg/ml and 200μg/ml. Ampicilin was used at 100μg/ml for *E. coli* only. The β-galactosidase assays were carried out as previously described [Brunsing, 2005 #60; Ferrari, 1985 #61; Miller, 1972 #63]. Protein interaction analysis was carried out essentially as described by the Clontech Yeast Two-hybrid system manual.

Plasmid constructions

The plasmid for *E. coli* over expression and purification of ORF118 was obtained by cloning the PCR amplified coding sequence using oligonucleotides BaORF1185'Nde (5'-GAGTGGA<u>CATATG</u>GAAGCAACAAAACG-3') and BaORF1183'Bam (5'-CTATA<u>GGATCC</u>AAAAATTTCAAGGTG-3') into plasmid pET28a (Stratagene) digested with *Nde*I and *Bam*HI thus generating a fusion to 6 histidine codons at the 5' end of the gene. A synthetic gene encoding full-length pXO2-61 was purchased from

GenScript Co., NJ, USA and subcloned in pET28 as described for pXO1-118. Transcriptional fusions to the *E. coli lacZ* gene were constructed in the replicative vector pTCVlac [Poyart, 1997 #64]. The promoter region of pXO1-118 was amplified using oligonucleotides p1185'Eco2 (5'-CTATTGAATTCATTGATAAAGTGTAG-3') and p1183'Bam2 (5'-TAAATGGATCCTGGCTTTCTTTTAGG-3'). The promoter region of pXO2-61 was PCR amplified using oligonucleotides pX0261-5'Eco (5'-GTTTAGAATTCTGAAATATTTTAATAGAC-3') and pX0261-3'Bam (5'-CTTTTGGATCCAATCAGATATAAATTTTTC-3'). The fragments were digested with *Eco*RI and *Bam*HI and cloned in pTCVlac similarly digested. The promoter region of *atxA* was PCR amplified using oligonucleotides Delta 118Eco2 (5'-TTCCAGAATTCCACTCCTTAATTCC-3') and AtxA3'Bam (5-CAAATGGATCCAGGGCATTTATATTATC-3'): the fragment was digested with *Eco*RI and *Eco*RV (the latter is naturally present in the *atxA* gene) and the 360bp fragment was cloned in pTCVlac digested with *Eco*RI and *Sma*I. This fragment contains all the promoter determinants required for *atxA* transcription according the Dai *et al.* {Dai, 1995 6637 /id}. Plasmid pORICm was used for the construction of the pXO1-118 deletion strain [Brunsing, 2005 #60]. A 720bp fragment downstream pXO1-118 was PCR amplified using oligonucleotides Delta 118Kpn (5'-AATAAGGTACCTTAAGTAATAAATAC-3') and Delta 118Bam (5'-ATATTGGATCCTAAAAAAGAAATATAAC-3') and cloned in pORICm at the *Kpn*I and *Bam*HI sites. An 860bp fragment upstream of pXO1-118 was also PCR amplified using oligonucleotides Delta118Sal (5'-CATAAGTCGACTCCTTAATTCCTTAAAAATC-3') and Delta118Pst (5'-

TATTACTGCAGGGAAACGGCCAATAATC-3’) and cloned in the resulting plasmid at the *Sal*I-*Pst*I sites. Finally, a blunt-ended spectinomycin cassette was cloned at the *Hinc*II site positioned in between the two cloned fragments in the vector multiple cloning site. The resulting plasmid was transformed into strain 34F2 and used to generate a deletion-spectinomycin replacement of pXO1-118 essentially as described [Brunsing, 2005 #60].

Plasmid pORICm was also used for the construction of the *atxA* deletion strain. The *atxA* coding region and upstream sequences were PCR amplified using oligonucleotide Ba118delta (5’-TTAATGAATTCTCGCATATACATTGTGAATAC-3’) and AtxA3’Bam (5’-CAAATGGATCCAGGGCATTTATATTATC-3’) and cloned in the *Eco*RI-*Bam*HI sites of pORICm. The resulting plasmid was digested with *Bcl*I and *Eco*RV and the 670bp excised fragment was replaced by the spectinomycin cassette as a *Bam*HI-*Hinc*II fragment. The resulting plasmid was used to transform strain 34F2 and generate a deletion-replacement of the *atxA* gene essentially as described [Brunsing, 2005 #60].

The gene encoding pXO1-118 was cloned in the two hybrid system vector pGBT9 and pGAD424 (Clontech) as an *Eco*RI-*Bam*HI fragment obtained by PCR amplification using oligonucleotides THS1185’Eco (5’-AATTAGAATTCGGAGGAATGGAAGCAACAAAACGATAC-3’) and BaORF1183’Bam described above. The *atxA* gene was cloned in the pGBT9 and PGAD424 plasmids using oligonucleotides AtxA5’*Eco*RI (5’-TTATAGAATTCCTAACACCGATATCCATA-3’) and AtxA3’Bam (5’-

CAAATGGATCCAGGGCATTTATATTATC-3’). An *Eco*RI linker with the sequence (5’-GAATTCTTGCCGGGACCTCTTCCGGGTCCGGAACTTCCTGGACCGGAGGGAATTC-3’) was then inserted in the *Eco*RI site to provide flexibility to the fusion protein. All PCR reactions were carried out on the full genome of strain 34F2 extracted using the UltraClean Microbial DNA Isolation Kit (Mo Bio, Solana Beach, California) or on purified pXO2 plasmid DNA (generously provided by Philip Hanna).

Protein expression and purification

Expression was obtained in *E. coli* BL21(DE3) grown in LB medium after induction with 0.1mM IPTG for 4 hours at 32 °C. The protein was purified from the soluble fraction of the cell extract and purified by Nickel affinity chromatography on a His trap chelating column (Pharmacia), thrombin removal of the tag and size exclusion on a Superdex75 column (Amersham, Pharmacia). The protein was stored frozen in 20mM Tris HCl buffer pH 7.4, 1.0 M NaCl, 50μM KCl, 5mM dithiothreitol. Selenomethionine labelled protein was purified with essentially the same protocol from cells grown in…. The pXO2-61 protein was expressed and purified as described for pXO1-118, except that 500 mM NaCl were sufficient to keep the protein soluble for freezing and long-term storage. The molecular weight of both proteins was confirmed by SDS-Page and MALDI-TOF mass spectrometry.

Crystallization, data collection, and structure solution

Native and selenomethionine-labelled pXO1-118 were crystallized by sitting or hanging drop vapor diffusion at room temperature by mixing of 3 μl of precipitant solution (40% (v/v) PEG-300, 100mM Tris-HCl pH 5.4, 5% (w/v) PEG-1000) and 3μl of protein solution at 14mg/ml . Rod-shaped crystals grew within 3 days and belong to space group $P3_221$ with unit cell a=b= 89.86 Å, c= 35.25 Å, $\alpha=\beta=$ 90°, $\gamma=$ 120°. The Matthews coefficient is 2.2 for 1 molecule/asymmetric unit (ASU), corresponding to 44.2% solvent content. One native and one Selenium-SAD (Single Anomalous Dispersion) datasets were collected at 100 K at SSRL beamline 9-2 and NSLS beamline X26C, respectively. Diffraction images were processed and scaled with the HKL package [Otwinowski, 1997 #23]. SOLVE [Terwillinger, 1999 #24] was used to locate four Se sites in the pXO1-118 ASU, leading to a set of initial phases with figure of merit (FOM)=0.32. Density modification (FOM=0.60) and automatic model building using RESOLVE [Terwillinger, 2001 #25] resulted in a 77% complete model. Further model building was performed manually in O [Kleywegt, 2001 #20] and refinement was carried out with REFMAC5 [Murshudov, 1997 #40] with simulated-annealing using CNS [Brunger, 1998 #21]. Electron density for a buried ligand was located in each subunit, and was tentatively filled with a molecule of undecanoic acid. The final model contains residues 1-150, three non-native N-terminal residues, one molecule of undecanoic acid, 95 water molecules and one K^+ ion. R_{work}= 17.7 % and R_{free}=22.5 % for data between 76.7 and 1.76 Å resolution.

Purified pXO2-61 was crystallized by the microbatch method under paraffin oil. Well diffracting crystals were obtained in two days using a buffer containing 1M NaI, 20% (v/v) PEG3350. They belong to space group $P2_12_12_1$ with unit cell parameters a=44 Å, b=62 Å, c=124 Å, $\alpha = \beta = \gamma = 90°$. Data were collected at SSRL to 1.49 Å and processed with the HKL package [Otwinowski, 1997 #23]. The structure was solved with molecular replacement using the structure of pXO1-118 as search model. Model building and refinement were carried out in O [Kleywegt, 2001 #20] and REFMAC5 [Murshudov, 1997 #40]. The final model has R_{work}= 17.7 % and R_{free}=20.9 % for data between 62.02 and 1.49 Å resolution. The asymmetric unit contains two molecules (residues 5-136), 364 water molecules, 26 I^- and 8 Na^+ ions, with 56.7% solvent content (Matthews coefficient of 2.9). The two polypeptide chains superpose with a root-mean-square deviation (rmsd) of 0.43 Å for all C carbons.

Data collection and refinement statistics are summarized in Table 1. The stereochemical quality of both models was assessed with PROCHECK [Laskowski, 1993 #32] with excellent results.

RESULTS

Overall structure of pXO1-118 and pXO2-61

We solved the structure of pXO1-118 to 1.76 Å resolution by Single-Wavelength Anomalous Dispersion (SAD) using a Selenomethonine substituted crystal. The asymmetric unit contains one polypeptide chain that forms a dimer across the crystallographic dyad in a single compact globular domain (Fig 1 and 3A). The dimeric state of pXO1-118 is consistent with the results of gel-filtration experiments (data not shown,) and yeast two hybrid analysis (see below). Each subunit adopts a globin-like fold that dimerizes via an antiparallel, left-handed four-helix bundle formed by the two C-terminal helices G and H from each subunit, burying 3500 $Å^2$ surface area per dimer. Following the standard nomenclature for the globin fold, helices A, B, E, F, G, H are present in pXO1-118 while helices C and D are missing. Despite the absence of significant sequence homology, the ternary and quaternary structure of the dimer is similar to that seen in the *B. subtilis* non-heme globin stress response regulator RsbR {Murray, 2005 7113 /id} and the *B. subtilis* oxygen sensor HemAT {Zhang, 2003 7108 /id}. When compared to HemAT however, pXO1-118 lacks helices Z, C and part of A, while helix H has a four-turn C-terminal extension, the latter compensating for loss of dimerization contacts made by helix Z. The absence of helix C allows helix E to pack directly against helix F, a displacement that fills a void left by the missing heme ligand.

As expected, the structure of pXO2-61 (Fig 2 and 3B) is very similar to that of pXO1-118 in both polypeptide fold and dimer arrangement, the main difference between the two structures being a C-terminal truncation resulting in a shorter helix H, a more compact overall shape and a decrease in dimerization interface to 2200 Å^2. The models for the two dimers superimpose with an rmsd of 1.16 Å for 264 C carbons.

A putative buried ligand binding cavity within each subunit of the dimer

A largely hydrophobic cavity running between helices E and G of pXO1-118 is filled by continuous electron density that stretches for about 12 Å and is significantly bent towards helix E, approximately occupying the position of heme pyrrole rings B and C in HemAT. Based on the density itself and electrostatic considerations, we modeled a molecule of undecanoic acid in the cavity, although its apparent decreased order when compared to the surrounding protein side chains possibly indicates the presence of a mixture of ligands and/or ligand conformations. A further, more hydrophilic section of the cavity is lined by helices E and G and the loop connecting helices D and E and is filled by four buried water molecules and one putative potassium ion in our structure. Overall, the cavity runs roughly through the center of each subunit and extends for about 20 Å with two sharp bends at the point of transition between the two predominately hydrophobic and hydrophilic sections (Fig. 4A). The same feature is observed in the structure of pXO2-61, but no clear electron density attributable to a ligand could be found, rather the space is filled with eight additional solvent molecules hydrogen-bonded to each other and to protein main-chain groups. 30 residues (out of 133) are fully conserved among pXO1-118, pXO2-61, BA2291 and their *B. cereus* homologs (Fig.1) . Of these, 12 directly face such cavity (of 22) and appear to be involved in interacting with a potential ligand, suggesting the presence of a bona fide binding or catalytic conserved site. Most notably, a

Marta Perego
Comment: Ref to Fig4 B&C??

very well conserved motif, ^{69}KIAxER^{74} at the end of helix F (x is any residue) is crucial for the recognition of the polar head of the acid ligand and appears to be a signature motif for this subclass of proteins. The side chain of Arg 74 forms a buried salt bridge with that of Asp 33 and water-mediated interactions with those of Glu 73 and Asn 87 as well as the carboxylic moiety of the fatty acid ligand, as part of a complex hydrogen bonding network, as shown in Fig. 4B and legend.

Dimerization interface

The dimerization interface comprises both hydrophobic and hydrophilic interactions. When moving along the dimerization axis in the direction shown in figure 1, we encounter three areas of mostly hydrophobic contact surface, interspersed with two sections of mainly hydrophilic residues making polar, electrostatic and water-mediated interactions (see Fig. 5). A total of 34 and 20 water molecules are buried in the intersubunit space of the dimer in pXO1-118 and pXO2-61 respectively. A partial view of the interface with superimposed electron density map is shown in figure 5. A three-dimensional superposition of the dimers in both crystal structures shows no obvious substitutions that could prevent the formation of a heterodimer for steric or electrostatic reasons. Due to the lack of a three-dimensional structure for the sensor domain of BA2291, a direct comparison with the binding interface of pXO1-118 and pXO2-61 cannot be made. A simple homology-based comparison shows that amino-acid conservation is lower than for the plasmid-encoded pair, reflecting the overall reduced sequence identity, and in contrast to the striking similarity in the putative ligand-binding pocket. However, we could find no specific reason to believe that BA2291 should not homodimerize in a similar fashion or even heterodimerize with either pXO1-118 or pXO2-61.

Functional analysis of sensor domains

The products of the pXO1-118 and pXO2-61 genes were shown to negatively regulate *in vivo* the activity of the major sporulation histidine kinase BA2291 both in *B. anthracis* and *B. subtilis* {White, 2006 7105 /id}. However, the mechanism of this regulation is still unknown and this did not allow us to rule out the possibility that the genetic location of the pXO1-118 gene, adjacent to and divergently transcribed from the *atxA* gene, could have a functional significance. To test this possibility we constructed a 34F2 mutant strain carrying a spectinomycin gene replacement for pXO1-118 gene. The strain, named 34F2Δ118, did not show any growth or sporulation defect when compared to the parental strain 34F2 (data not shown). In order to test whether the pXO1-118 protein had any role in *atxA* transcription, both strains were transformed with a pTCVlac construct carrying the *atxA* promoter and the transcription of this gene was analyzed by means of β-galactosidase assays. As shown in Fig. 6A, no difference in transcription was observed between the parental strain and 34F2Δ118 indicating that the pXO1-118 protein does not affect AtxA production. As a consequence, the product of pXO1-118 did not affect the transcription of the *pagA* gene encoding the protective antigen (data not shown).

To test the possibility that pXO1-118 could affect the activity of AtxA, the yeast two-hybrid system (Clontech) was used to detect protein:protein interaction. Both genes were singly cloned in the bait plasmid pGBT9 and in the prey plasmid pGAD424. When the interaction assays were carried out in the yeast strain AH109, we detected interaction in the control strain carrying pXO1-118 on both pGBT9 and pGAD424 plasmids but we did not detect any interaction of pXO1-118 with AtxA either as bait or as prey. These results confirmed that pXO1-118 can dimerize but did not support the hypothesis that it may interact with AtxA.

Gene transcription analysis

The transcription profile of the pXO1-118 and pXO2-61 promoters were determined by means of β-galactosidase analysis carried out on promoter-*lacZ* fusion constructs. The pTCVlac plasmid derivatives carrying either the pXO1-118 or the pXO2-61 promoters were transformed in the Sterne strain 34F2. In order to determine whether the AtxA virulence transcription factor had any role in regulating the transcription of the sensor domains, the *lacZ*-fusion constructs were also transformed in the 34F2 derivative carrying a deletion of the *atxA* gene (34F2ΔatxA). The results of this analysis are shown in Figure 6B. The transcription from both promoters was induced in late exponential phase and it increased during the early hours of stationary phase. The absence of AtxA prevented this induction from the pXO2-61 promoter but not from the pXO1-118 promoter. A similar pattern of transcription was observed when the cells were grown in Schaeffer’s sporulation medium which induces sporulation of *B. anthracis* cells at a faster rate than the LB medium (data not shown). Thus while transcription of the pXO1-118 gene is independent of AtxA, the transcription of pxO2-61 depends on this transcription activator as previously indicated by a microarray study [Bourgogne, 2003 #45].

DISCUSSION

In this paper we report the crystal structures of two dimeric proteins from *Bacillus anthracis*, pXO1-118 and pXO2-61, which have been previously shown to be capable of inhibit the histidine kinase BA2291-dependent sporulation in this organism as well as in *B. subtilis* {White, 2006 7105 /id}. Based on the presence of a large, highly conserved cavity and the approximately 30% identity with the sensor domain of BA2291, we predict the latter to adopt a similar fold and all three proteins to be able to bind the same molecule(s) that regulate its function. This allows us to make predictions on the mechanism by which such repression occurs. The most straightforward model consistent with the available data is one in which pXO1-118 and pXO2-61 compete for binding the same as yet unidentified signalling molecule(s) that ultimately activate the kinase function of BA2291. As the identity of the ligand remains unknown, we cannot predict whether any catalytic activity by pXO1-118 or pXO2-61 is involved in decreasing its cellular levels. In any case, sufficient amounts of plasmid encoded protein would make the signal unavailable to the sensor domain of BA2291, thus turning this histidine kinase into a phosphatase of the Spo0F intermediate of the phosphorelay. This results in decreasing the level of phosphorylated Spo0A (Spo0A~P) and largely suppressing the sporulation phenotype. A prerequisite for this model is the presence of limiting amounts of activating signal in the cell, and the observation that BA2291 represses, rather than activate, sporulation when its gene is present in multicopy {Brunsing, 2005 7039 /id} suggests that this might be the case. On a different note, the apparent lack of any obvious ligand-induced conformational change as inferred by comparing the structures of ligand-bound pXO1-118 and free pXO2-61 is consistent with, albeit not necessary for, a passive role such as the one proposed (but see below). Of course this model, as any other that could be envisioned on the basis of existing data, needs experimental validation and can only be regarded as speculative in view of the current knowledge. As discussed elsewhere {White, 2006 7105 /id}, a model based on the ability of pXO1-118 or pXO2-61 to heterodimerize with BA2291 appears to be inconsistent with available evidence.

As hinted to above, with all the limitations inherent in such a comparison, we could observe no major structural changes between pXO1-118 and pXO2-61 despite the binding of a ligand at a conserved site in the former. In contrast, although little structural information is available on how a buried ligand bound to a sensor kinase exerts its regulatory function, it is widely believed that ligand-induced conformational changes in the sensor domain will lead to quaternary structure rearrangements that allow autophosphorylation to take place. Perhaps, major rearrangements occur only in the linker region that connects the sensor domain to the histidine kinase domain, and this linker is missing in pXO1-118 and pXO2-61. The discrepancy could also be explained on the basis of the amino acid differences that exist between pXO1-118/pXO2-61 and the sensor domain of BA2291, or between the *in vivo* relevant ligand and the one bound to the E.coli produced pXO1-118 protein, and any speculation on the subject will have to await the identification of the molecules involved in the signalling process.

Despite sharing only 7% and 4% identity respectively at the amino-acid level after structure-based alignment, the overall fold and dimer arrangement of pXO1-118 and pXO2-61 is strikingly similar to that of the *B. subtilis* stress response regulator RsbR, the only other non-heme globin so far identified, whose structure has been recently reported [ref]. The positions of all six helices in RsbR is remarkably similar to the helix positions in both *B. anthracis* sensor domains, although differences exist in some of the connecting loops and notably in part of the dimerization helices G and H, which are significantly bent only in RsbR, leading to an rmsd of 3.6 and 3.4 Å for 120 and 118 $C\alpha$ atoms respectively when a single subunit of RsbR was superimposed to pXO1-118 and pXO2-61 using the DALI algorithm [ref] (see Figure 3). Interestingly, as noted for pXO1-118 and pXO2-61, the dimerization interface of RsbR, and to a lesser extent that of HemAT, is made of distinct sections dominated by either hydrophobic or hydrophylic/water mediated interactions. However, although all three proteins dimerize through the same interface, the relative orientation of the subunits is significantly different in RsbR when compared to both *B. anthracis* proteins, as shown in fig. 3.

Final paragraph?

Something on the general use of the globin fold?

Table 1 Statistics of data collection and refinement.

	SAD phasing	Model refinement	
	pXO1-118	pXO1-118	PXO2-61
Wavelength (Å)	0.9781	0.97923	0.97923
Resolution range (Å)	50-2.5	76.7-1.76	62.02-1.49
Observations	70890	173637	408633
Unique reflections	5888	16461	56717
Completeness[1] (%)	99.8(100.0)	99.5(95.0)	98.9(94.5)
R_{sym}[1,2] (%)	6.9(24.7)	5.7(46.3)	6.8(31.5)
R_{cryst}[3]/R_{free}[4] (%)		18.1/22.5	17.7/20.9
Protein atoms		1408	2628
Water molecules		95	364
Ions		1	34
Ligand molecules		1	
r.m.s.d.[5]			
Bonds (Å)		0.009	0.012
Angles (°)		1.14	1.37
Average B-factors (Å^2)			
Protein			
Main chain		26.8	18.9
Side chain		34.1	23.9
Water		34.5	35.9
Ligand		39.1	
Ramachandran plot			
Most favoured		142	249
Additionally allowed		4	5
Generously allowed		0	0
Disallowed		0	0

[1] Values in parentheses refer to the highest resolution shell. [2] $R_{sym}=\Sigma|I_h-\langle I_h\rangle|/\Sigma I_h$, where $\langle I_h\rangle$ is the average intensity over symmetry equivalent reflection. [3] R-factor $=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$, where the

summation is over the data used for refinement. [4]R_{free} was calculated using 5% of data excluded from refinement. [5]Root-mean-square deviations [Engh, 1991 #36]

ACKNOWLEDGEMENTS

We thank the U.S. Army (DAMD17-03-2-0038) and NIH (P01 AI 55789) for support of this work. We would like to thank Annie Heroux for measuring crystals taking advantage of the FedEx crystallography access from the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, under Contract No. DE-AC02-98CH10886. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. We would like to thank the staff of the SSRL Structural Molecular Biology Program, which is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. G.R. Stranzl received an Erwin Schrödinger fellowship from the Austrian Science Fund. M.G. was supported in part by grant 2 PO4B 026 28 from the Polish State Committee for Scientific Research. Work at The Scripps Research Institute was supported in part by grant GM055594 from the National Institute of General Medical Sciences and AI055860 from the National Institute of Allergies and Infectious Diseases, National Institutes of Health. The Stein Beneficial Trust supported in part oligonucleotide synthesis and DNA sequence. The authors also thank James A. Hoch for helpful discussion.

FIGURE LEGENDS

Fig. 1. Sequence alignment and domain architecture of *B. anthracis* sensor domain proteins. Amino acid sequences of *B. anthracis* pXO1-118,pXO2-61 and of *B. cereus* pBC218-0049 were aligned with the sensor domain of the sporulation histidine kinase BA2291 of *B. anthracis* and its highly conserved ortholog (100% identity) from *B. thuringiensis* (Hkna) and *B. cereus* (Bc51976636). A more distantly related protein histidine kinase identified in Geobacillus kaustophilus is also aligned (GK56379900) (25% of identical residues). Sequences were aligned by the MUSCLE program (http://phylogenomics.berkeley.edu/cgi-bin/muscle/input_muscle.py)(R.C.Edgar, Nucleic Acids Research 32, 5 (2004).) and colored in BioEdit (T.A.Hall, Nucl Acids Symp Ser, 41 (1999).). The shading threshold is 70%. The coloring of the alignment reflects identity and similarity, according to the default BioEdit amino acid similarity scoring matrix. Black color defines identical positions whereas gray - similar positions.

Helical elements as derived from the structure pf PXO1-118 are shown as cylinders, named according to the standard nomenclature for globins. The domain organization of the full-length proteins is also shown (HisKA = histidine kinase domain, HTPase = ATP-binding domain).

Fig. 2. Physical map (not in scale) of the pXO1, pXO2 and pBC218 plasmid regions containing the sensor domain encoding genes described in this report.

The bp positions of the fragments shown and the ORF numbering are from GenBank accession numbers AF065404 (pXO1), NC_002146 (pXO2) and NZ_AAEK01000004 (pBC218). Plasmid pBC218 is from *B. cereus* strain G9241 {Hoffmaster, 2004 6999 /id}. The arrows indicate the direction of transcription of the open reading frames according to the annotations in the database.

Fig. 3. Ribbon representation of pXO1-118 and pXO2-61. Structures are colored blue to red from the N to the C terminus. Helices are marked on one subunit of pXO1-118 according to the standard naming for globins. RsbR is shown for comparison. The structures are shown in the same orientation after alignment of the right-hand subunit using the DALI algorithm [ref].

A. pXO1-118
B. pXO2-61
C. RsbA

Fig. 4. Putative ligand binding cavity in pXO1-118 and pXO2-61

A. The fatty acid ligand modeled in the structure of pXO1-118 shown in its 2fo-fc electron density map. Arginine 74 and the two visible conformations of Phenylalanine 19 are shown.

B, C. The hydrogen bonding network in the hydrophylic part of the cavity for pXO1-118 and pXO2-61 respectively. Water molecules are shown as red spheres, a putative potassium ion in pXO1-118 and an iodine ion from the crystallization buffer in pXO2-61 are shown as green and yellow spheres, respectively.

Fig. 5. Partial view of the dimerization interface of pXO1-118 with superimposed final 2fo-fc electron density map contoured at 1.3σ.

Fig. 6: Transcription analysis of promoter-*lacZ* transcriptional fusions in *B. anthracis*. β-galactosidase assays were carried out on *B. anthracis* cultures grown in LB supplemented with kanamycin at 7.5μg/ml.

Open symbols: growth curves; closed symbols: Miller Units

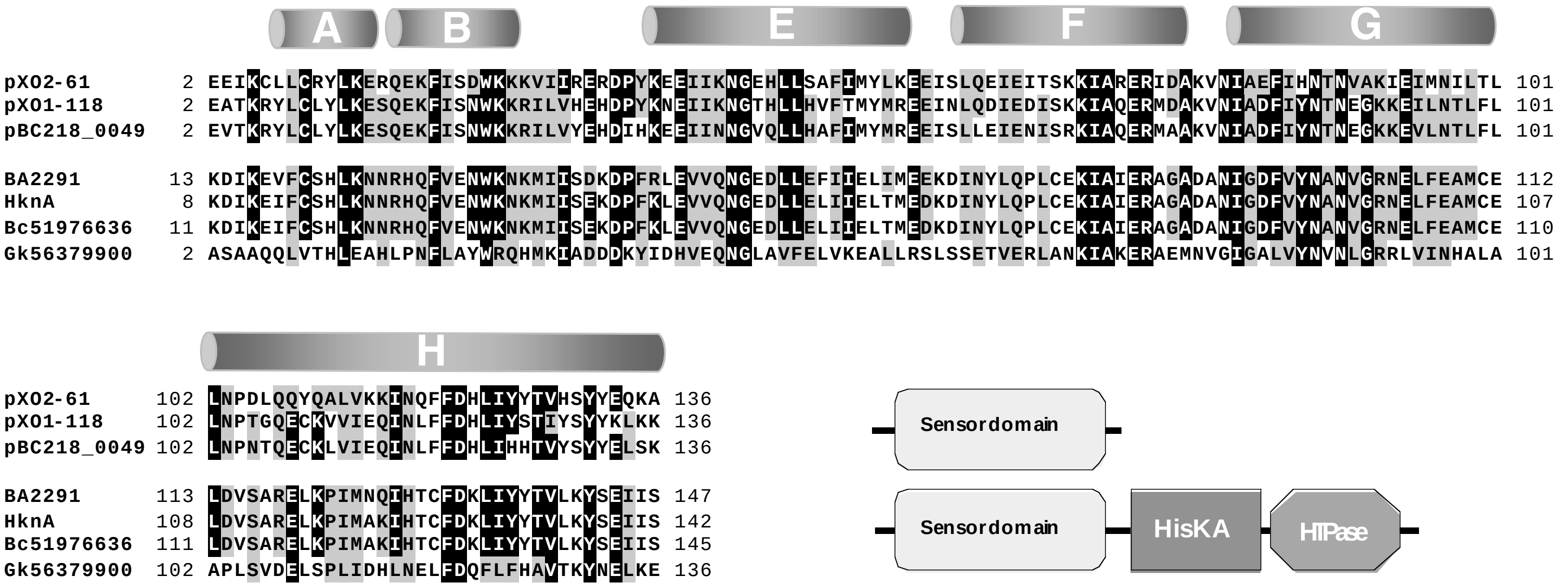
A. Transcription analysis of the *atxA* promoter in the pXO1-118 deletion strain.

Strains and symbols: 34F2/pTCVlac-atxA: - ○ -; 34F2Δ118/pTCVlac-atxA: - ▽ -.

B. Transcription analysis of the pXO1-118 and pXO2-62 promoters in the 34F2 and the 34F2Δ*atxA* deletion strains.

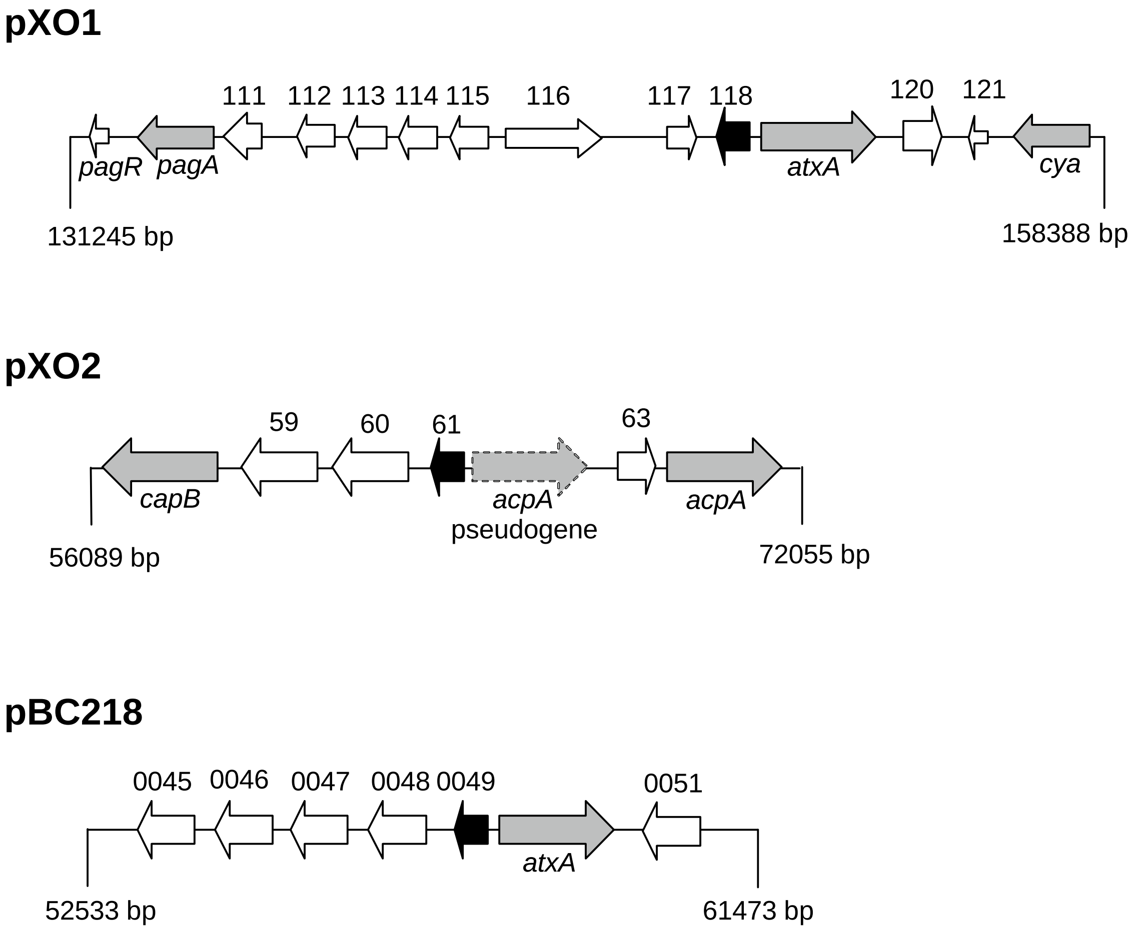
Strains and symbols: 34F2/pTCVlac-118: - ▽ -; 34F2ΔatxA/pTCVlac-118: - ◇-;

34F2/pTCVlac-62: - ○ -; 34F2ΔatxA/pTCVlac62: - △ -.

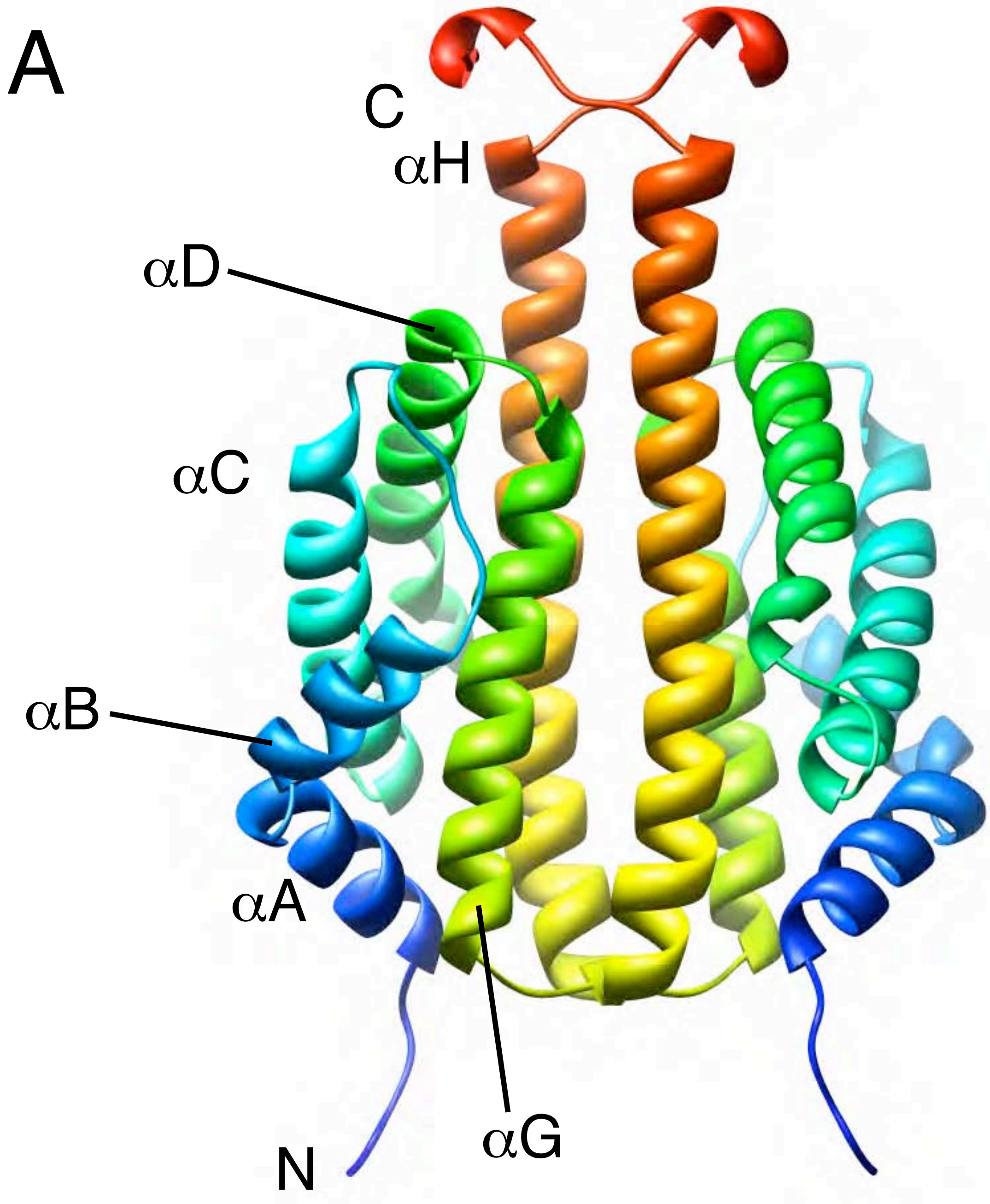


```
pXO2-61       2 EEIKCLLCRYLKERQEKFISDWKKKVIIRERDPYKEEIIKNGEHLLSAFIMYLKEEISLQEIEITSKKIARERIDAKVNIAEFIHNTNVAKIEIMNILTL 101
pXO1-118      2 EATKRYLCLYLKESQEKFISNWKKRILVHEHDPYKNEIIKNGTHLLHVFTMYMREEINLQDIEDISKKIAQERMDAKVNIADFIYNTNEGKKEILNTLFL 101
pBC218_0049   2 EVTKRYLCLYLKESQEKFISNWKKRILVYEHDIHKEEIINNGVQLLHAFIMYMREEISLLEIENISRKIAQERMAAKVNIADFIYNTNEGKKEVLNTLFL 101

BA2291       13 KDIKEVFCSHLKNNRHQFVENWKNKMIISDKDPFRLEVVQNGEDLLEFIIELIMEEKDINYLQPLCEKIAIERAGADANIGDFVYNANVGRNELFEAMCE 112
HknA          8 KDIKEIFCSHLKNNRHQFVENWKNKMIISEKDPFKLEVVQNGEDLLELIIELTMEDKDINYLQPLCEKIAIERAGADANIGDFVYNANVGRNELFEAMCE 107
Bc51976636   11 KDIKEIFCSHLKNNRHQFVENWKNKMIISEKDPFKLEVVQNGEDLLELIIELTMEDKDINYLQPLCEKIAIERAGADANIGDFVYNANVGRNELFEAMCE 110
Gk56379900    2 ASAAQQLVTHLEAHLPNFLAYWRQHMKIADDDKYIDHVEQNGLAVFELVKEALLRSLSSETVERLANKIAKERAEMNVGIGALVYNVNLGRRLVINHALA 101

pXO2-61     102 LNPDLQQYQALVKKINQFFDHLIYYTVHSYYEQKA 136
pXO1-118    102 LNPTGQECKVVIEQINLFFDHLIYSTIYSYYKLKK 136
pBC218_0049 102 LNPNTQECKLVIEQINLFFDHLIHHTVYSYYELSK 136

BA2291      113 LDVSARELKPIMNQIHTCFDKLIYYTVLKYSEIIS 147
HknA        108 LDVSARELKPIMAKIHTCFDKLIYYTVLKYSEIIS 142
Bc51976636  111 LDVSARELKPIMAKIHTCFDKLIYYTVLKYSEIIS 145
Gk56379900  102 APLSVDELSPLIDHLNELFDQFLFHAVTKYNELKE 136
```

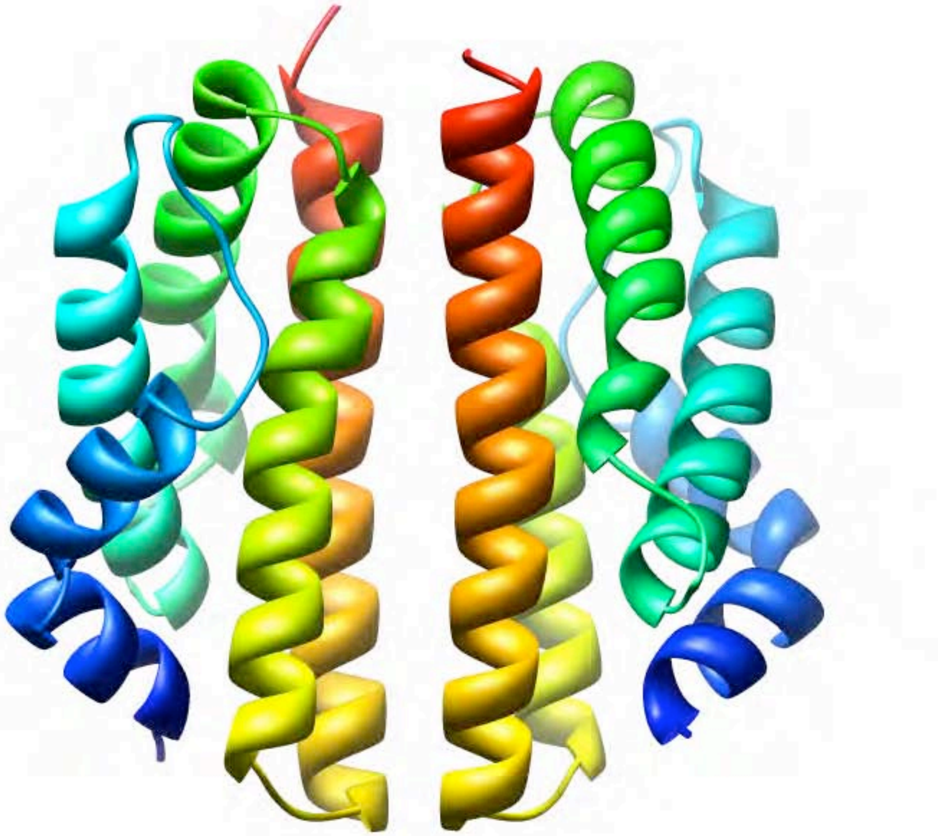
Stranzl et al.
Fig.1



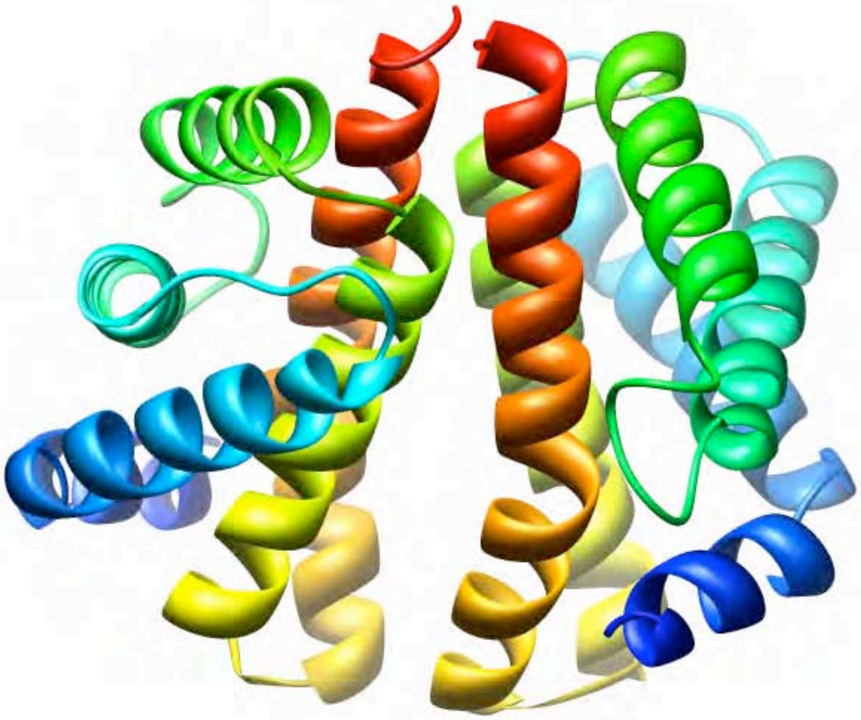
Stranzl et al.
Fig.2


A
C
αH
αD
αC
αB
αA
N
αG

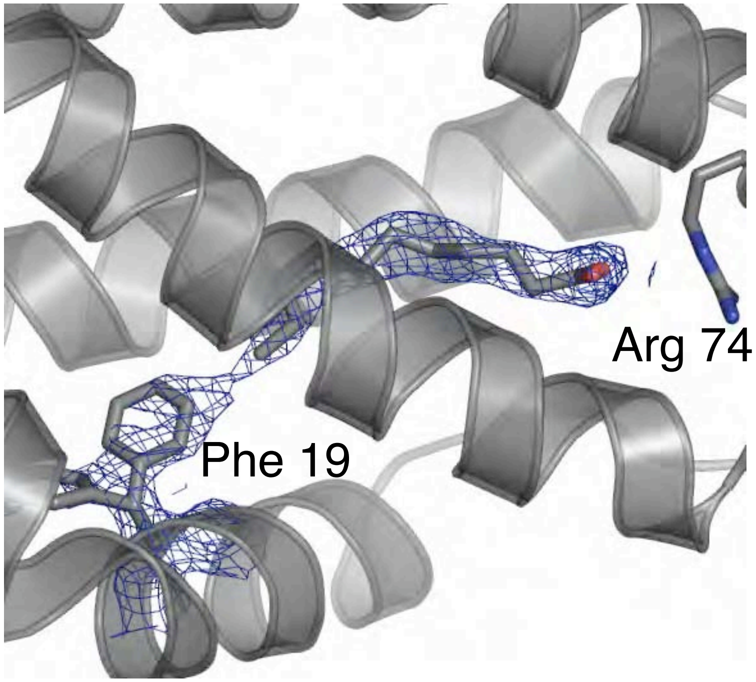
B



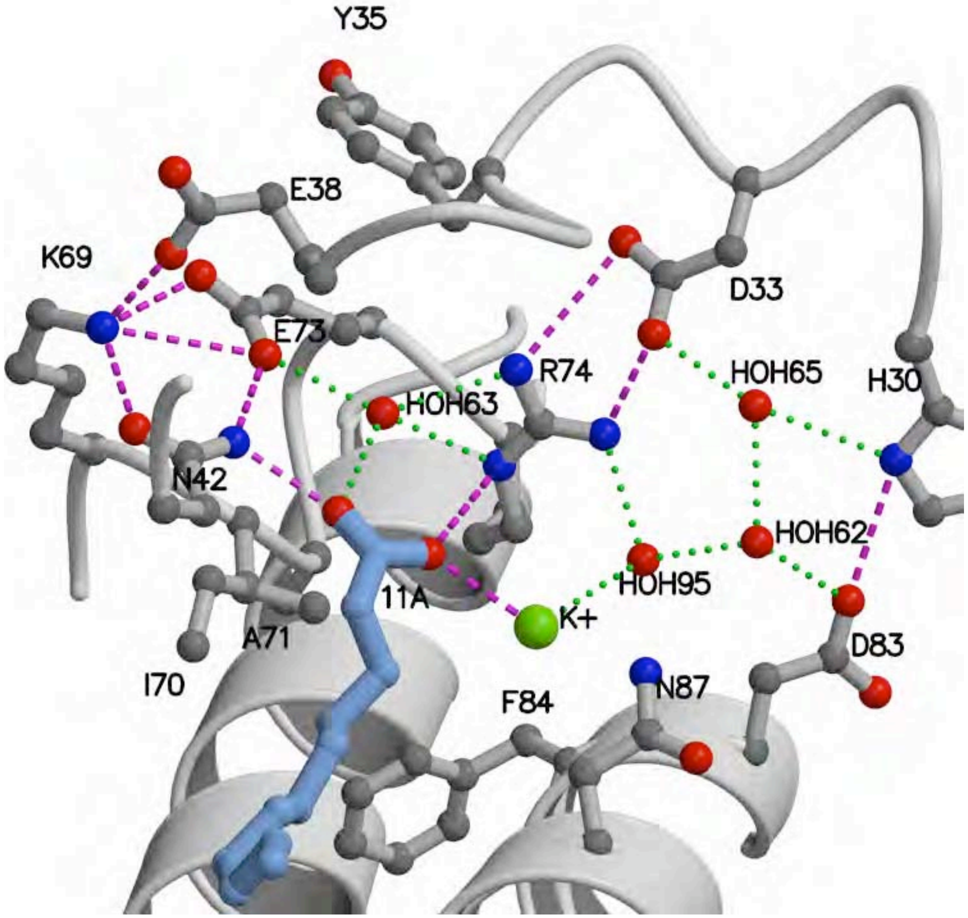
C

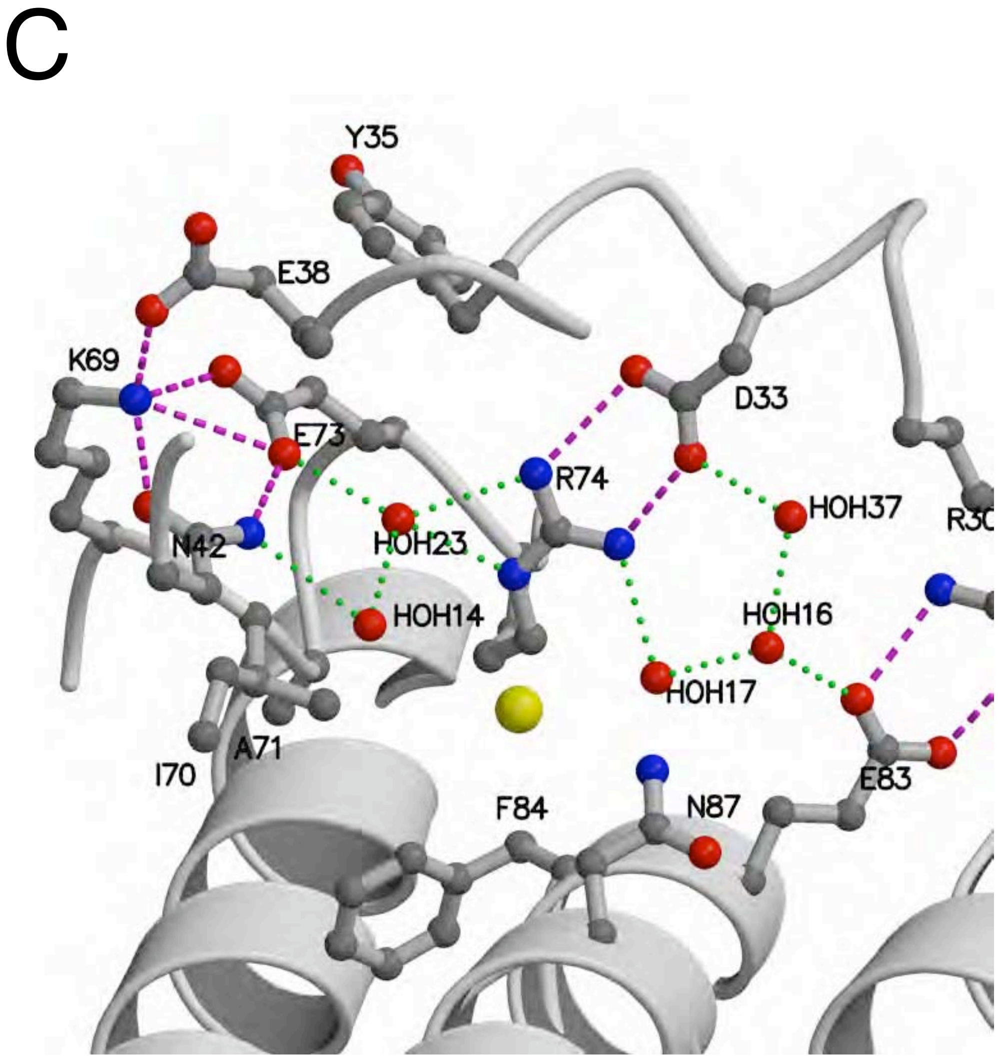


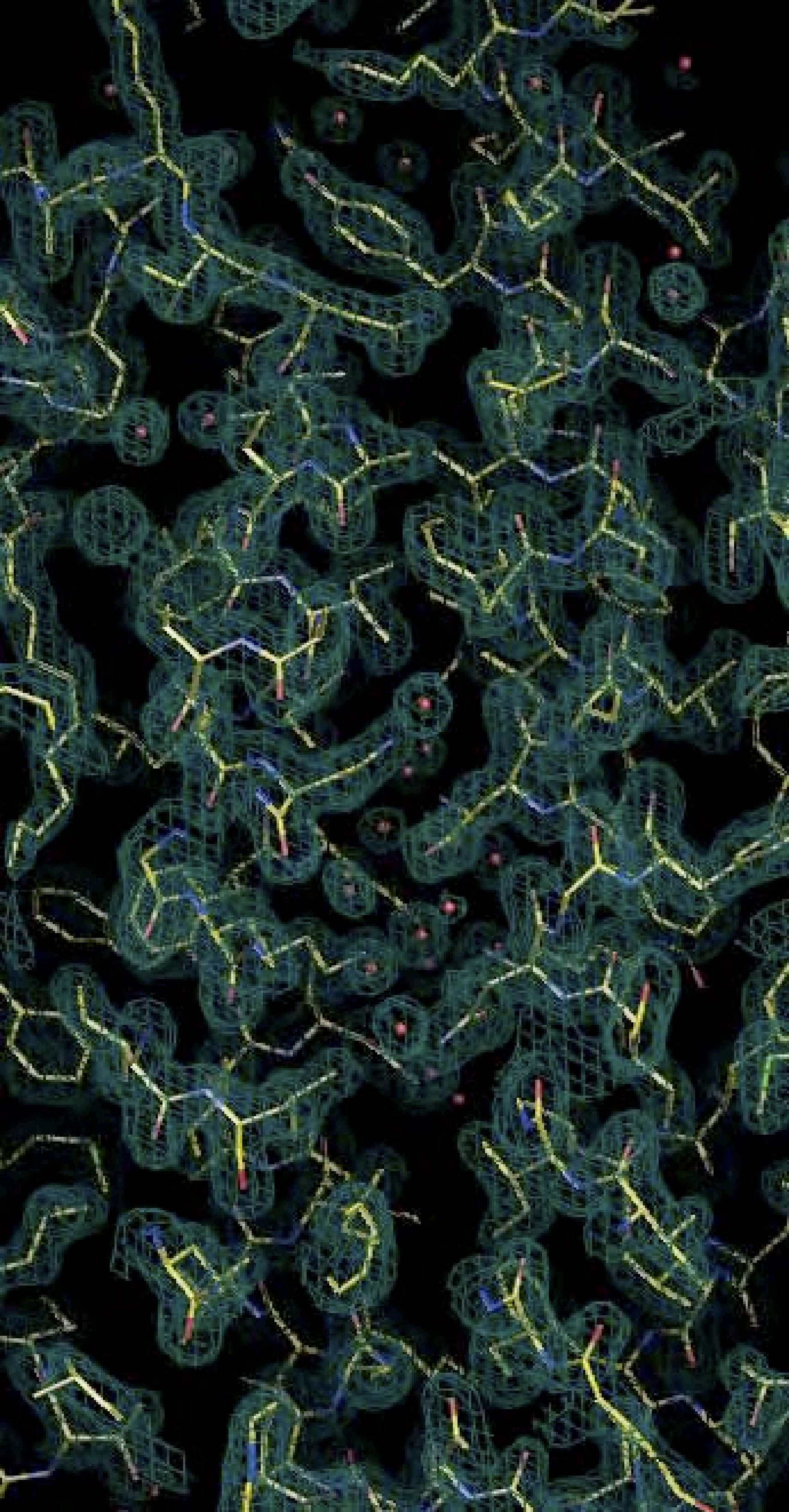
A

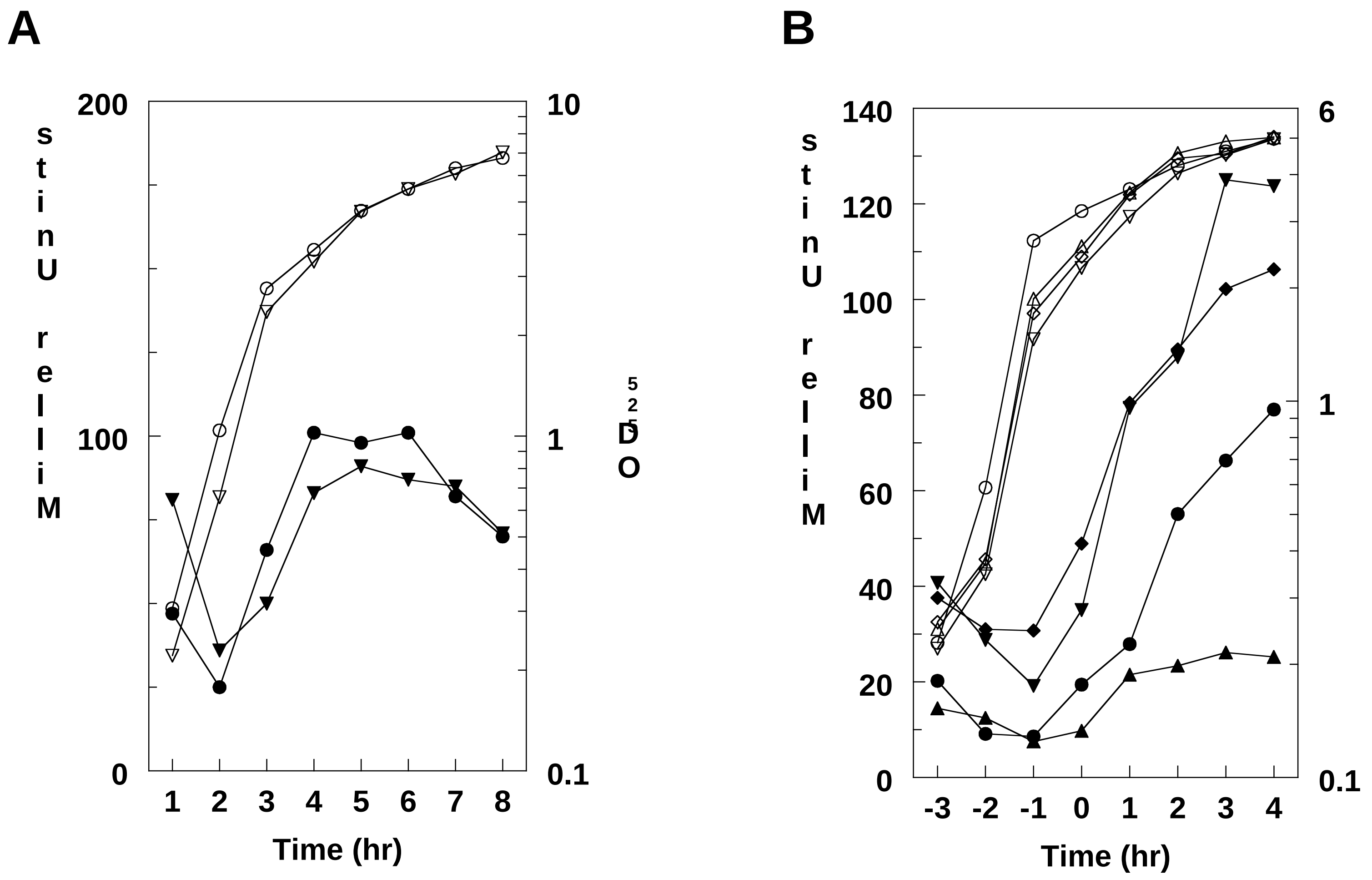


B


Y35
E38
K69
D33
E73
R74
HOH65
H30
HOH63
N42
HOH62
11A
HOH95
K+
A71
D83
I70
N87
F84


C
Y35
E38
K69
E73
R74
D33
HOH37
R3
N42
HOH23
HOH14
HOH16
HOH17
E83
A71
I70
F84
N87





Stranzl et al.
Fig.6

nature structural & molecular biology

The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor

Benjamin E Turk[1,5], Thiang Yian Wong[2,5], Robert Schwarzenbacher[2], Emily T Jarrell[1], Stephen H Leppla[3], R John Collier[4], Robert C Liddington[2] & Lewis C Cantley[1]

Recent events have created an urgent need for new therapeutic strategies to treat anthrax. We have applied a mixture-based peptide library approach to rapidly determine the optimal peptide substrate for the anthrax lethal factor (LF), a metalloproteinase with an important role in the pathogenesis of the disease. Using this approach we have identified peptide analogs that inhibit the enzyme *in vitro* and that protect cultured macrophages from LF-mediated cytolysis. The crystal structures of LF bound to an optimized peptide substrate and to peptide-based inhibitors provide a rationale for the observed selectivity and may be exploited in the design of future generations of LF inhibitors.

Inhalational anthrax progresses rapidly to a highly fatal systemic infection[1]. The causative bacterium *Bacillus anthracis* secretes three plasmid-encoded toxin proteins that contribute to pathogenesis: protective antigen (PA), edema factor (EF) and lethal factor (LF)[2]. PA binds to a cell surface receptor and forms an oligomeric pore that translocates both EF and LF into the cytosol of target cells. The combination of PA and LF is known as lethal toxin (LeTx), and intravenous delivery of LeTx alone causes death in rodents[2,3]. In addition, *B. anthracis* strains deficient in either component of LeTx are greatly attenuated, suggesting an important role for the toxin in the disease[4]. As antibiotics alone typically fail against systemic anthrax unless administered at an early stage, LeTx has been proposed as a potential target for anthrax drugs to be used with antibiotics in combination therapy[1]. Several experimental approaches to LeTx neutralization based on inhibition of cellular LF uptake have shown efficacy in animal models[5,6].

LF is a zinc-dependent metalloproteinase that cleaves most MAP kinase kinase (MKK) enzymes at sites near their N termini[7–10]. Cleavage impairs the ability of the MKK to interact with and phosphorylate its downstream MAP kinase substrates by disrupting or removing a docking site known as the D-domain[11]. Inhibition of MAP kinase pathways by LF impairs dendritic cell and macrophage function and may help to establish infection[9,12]. Higher levels of toxin are cytotoxic specifically to macrophages and probably contribute to fatality later in the course of the disease[1,2,13,14]. Although the mechanisms by which MKK cleavage leads to macrophage cell death are not entirely known, p38 family MAP kinases seem to be required for survival of macrophages upon activation by bacterial endotoxins[15].

Efficient cleavage of MKKs requires interaction between an LF exosite that has not yet been characterized and a region in the MKK catalytic domain distal from the cleavage site[16]. However, mutation of residues surrounding the scissile bond in MKKs abolishes proteolysis, indicating that cleavage site recognition is also crucial to substrate selection by LF[7,15]. Accordingly, LF can cleave short peptides, and efficient substrates have been generated based on a consensus motif derived from MKK cleavage sites[17–19]. It is not clear, however, which positions surrounding the cleavage site are most critical for efficient catalysis, nor whether residues found in MKKs are optimal for cleavage by LF. Such information is important for the design of therapeutically useful small molecule LF inhibitors, as thus far only rather long (more than ten residues) peptide hydroxamates have been reported to specifically inhibit LF[19]. Here we take an unbiased approach to the discovery of LF substrates and inhibitors by selection from random pools of millions of peptides, and report the crystal structures of LF in complex with optimized substrates and small molecule peptide-based inhibitors.

RESULTS

Determination of the optimal peptide cleavage motif for LF

To gain insight into substrate recognition by LF and to facilitate the development of LF inhibitors, we applied a mixture-based peptide library approach that produces extended cleavage site motifs for proteases[20,21]. Initially we prepared a partially degenerate peptide mixture, acetyl-KKKPTPXXXXXAK (See **Table 1** for explanation of nomenclature), in which we fixed six positions with the residues found N-terminal to the LF cleavage site in MKK-1 and followed them by a number of degenerate positions. Partial digestion of the library with LF followed by Edman sequencing of the mixture provided the specificity for the positions C-terminal to the cleavage site (**Table 1**).

---

[1] Division of Signal Transduction, Department of Medicine, Beth Israel Deaconess Medical Center, and Department of Cell Biology, Harvard Medical School, 330 Brookline Avenue, Boston, Massachusetts 02215, USA. [2] The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. [3] National Institute of Allergy and Infectious Diseases, 9000 Rockville Pike, Bethesda, Maryland 20892, USA. [4] Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. [5] These authors contributed equally to this work. Correspondence should be addressed to L.C.C. (lewis_cantley@hms.harvard.edu) or R.C.L. (rlidding@burnham.org).

Published online 29 December 2003; doi:10.1038/nsmb708

Table 1 LF cleavage site specificity and cleavage sites of known protein substrates

	Cleavage position P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′
Consensus	R (2.1)	K (2.0)	K (2.0)	V (1.5)*	Y (3.1)	P	Y (3.0)	P (1.9)	N (1.4)	E (1.6)
	S (2.1)	R (1.9)	R (1.9)	P (1.5)*	R (1.6)		L (2.2)	Q (1.4)	M (1.3)	A (1.5)
	K (1.7)	S (1.7)	H (1.6)	F (1.4)*	F (1.4)		I (2.1)	R (1.4)		
	H (1.4)	H (1.5)	S (1.4)	A (1.4)*	L (1.3)		M (1.8)	K (1.3)		
							F (1.8)	G (1.3)		
							V (1.4)			
MKK-1	K	K	K	P	T	P	I	Q	L	N
MKK-2	R	K	P	V	L	P	A	L	T	I
MKK-3	R	K	K	D	L	R	I	S	C	M
MKK-4	K	R	K	A	L	K	L	N	F	A
MKK-4	F	K	S	T	A	R	F	T	L	N
MKK-6	R	N	P	G	L	K	I	P	K	E
MKK-7	P	R	P	T	L	Q	L	P	L	A
MKK-7	P	R	H	M	L	G	L	P	S	T

Positions surrounding the scissile bond are defined as (…P3-P2-P1-P1′-P2′-P3′…) where cleavage occurs between the P1 and P1′ residues. Top: LF selectivity as determined using the peptide libraries acetyl-KKKPTPXXXXXAK (for the P1′–P4′ positions) and MXXXXXPYPMEDK(K-biotin) (for the P6–P2 positions). Selectivity values were determined by dividing the molar amount of a given residue within a sequencing cycle by the average molar amount of all residues within that cycle, so that a value of 1 is average and would thus indicate no selectivity. Only positive selections of ≥1.3 are shown. Values at the P3 position marked with an asterisk reflect the proportional increase of that residue from the previous cycle. Bottom: Residues present at positions surrounding the LF cleavage sites in MKK proteins.

To obtain selectivity information for sites N-terminal to the scissile bond, we constructed a secondary library, MXXXXXPYPMEDK (K-biotin), in which we fixed the residues most highly selected by LF at the primed positions. We also fixed proline at the P1 position, as an MKK-1 mutant bearing alanine at this position is not cleaved by LF[7]. Partial cleavage of this library was followed by removal of the undigested peptides and C-terminal fragments with immobilized avidin. Sequencing of the N-terminal fragments subsequently provided the specificity for LF at the unprimed positions (**Table 1**).

LF seems to be most selective at the P1′ position (immediately C-terminal to the scissile bond), where the enzyme requires a hydrophobic amino acid, and can accommodate both aliphatic and aromatic residues. Other features of the motif include a general selection for hydrophobic residues at the P2 position and an unusual selectivity for basic residues at multiple positions N-terminal to the cleavage site. Notably, sequence comparisons and mutagenesis studies have indicated that at least two basic residues and a downstream ΦXΦ sequence (where Φ indicates a hydrophobic amino acid and X any amino acid) are essential features of D-domains for mediating interactions with MAP kinases[22–24]. This similarity provides an evolutionary rationale for the targeting of these particular sites within the MKKs by LF: adaptive mutations in MKKs that would render them uncleavable would necessarily produce nonfunctional enzymes, thus making the acquisition of anthrax resistance unlikely.

Although general features of the selected consensus LF cleavage motif are reflected in the residues surrounding the cleavage sites within the MKKs (**Table 1**), specific aspects of the motif, such as the selection of tyrosine over other hydrophobic residues at the P1′ position, could not have been predicted based on consideration of known cleavage sites. Accordingly, a ten-residue peptide based on the consensus cleavage site (LF10) is cleaved ~50-fold more efficiently than an analogous MKK-1 cleavage site–spanning peptide (**Table 2**). We further substantiated the library selections by preparing additional peptides with alanine substitutions at various sites within the consensus. In each case, the substitution led to a substantial decrease in cleavage efficiency (**Table 2**). An extended 15-residue consensus peptide (LF15) provided a marked increase in cleavage efficiency over LF10, while maintaining favorable spectral properties (an eight-fold increase in fluorescence upon exhaustive cleavage). This peptide has the highest specificity constant of any LF peptide substrate thus far reported[17–19], allows detection of very low quantities of LF, and should therefore be useful in high-throughput screens for LF inhibitors.

Evaluation of peptide-based LF inhibitors

Substrate-derived inhibitors for metallo proteinases have been produced by incorporating a metal-chelating group either to the C terminus of a peptide corresponding to the unprimed positions, or to the N terminus of a peptide covering the primed positions[25,26]. As LF has substantial selectivity on either side of the scissile bond, we prepared both types of inhibitors and tested them for their ability to inhibit cleavage of the consensus peptide by LF. As in a previously reported study[19], we found that a relatively long C-terminal peptide hydroxamate is a potent LF inhibitor, whereas short peptide analogs such as acetyl-KVYP-hydroxamate inhibit the enzyme poorly (**Table 3**). Conversely, measurable inhibition was found with a small compound incorporating primed side residues, 2-thioacetyl-YPM-amide (SHAc-YPM, **Table 3**). This compound bears an N-terminal metal chelating group followed by a hydrophobic residue at the P1′ position, an arrangement shared by compounds previously reported to inhibit matrix metalloproteinases (MMPs)[27,28]. This relationship prompted us to test several similar MMP inhibitors for potency against LF. One such compound, GM6001 (3-(*N*-hydroxycarboxamido)-2-isobutylpropanoyl-Trp-methylamide)[29], an N-terminal hydroxamic acid with a P1′ leucine mimetic, a P2′ tryptophan and a C-terminal methyl group, inhibited LF more potently than did the other compounds tested (**Table 3** and data not shown). The enhanced potency of GM6001 over SHAc-YPM, despite the presence of predicted suboptimal residues, is probably attributable to the favorable substitution of the hydroxamic acid moiety for the thioacetyl group[28,30].

Table 2 Catalytic parameters for cleavage of substrate peptides by LF

Peptide	Sequence	k_{cat} / K_m (M^{-1} s^{-1})
MKK-1	Mca-KKPTPIQLN-Dnp	2,500 ± 800
LF10	Mca-KKVYPYPME-Dnp	130,000 ± 20,000
LF10-P5 Ala	Mca-**A**KVYPYPME-Dnp	7500 ± 500
LF10-P2 Ala	Mca-KKV**A**PYPME-Dnp	60,000 ± 10,000
LF10-P1′ Ala	Mca-KKVYP**A**PME-Dnp	22,000 ± 2,000
LF15	Mca-RRKKVYPYPME-Dnp-TIA	$4 \times 10^7 \pm 1 \times 10^7$

Residues in bold indicate substitutions to the consensus peptide. Substrate peptides contain N-terminal Mca (7-methoxycoumarin-4-acetyl) fluorescent groups and Dnp (2,4-dinitrophenyldiaminopropionic acid) quenching residues C-terminal to the cleavage site, allowing reaction progress to be followed fluorometrically by observing the increase in coumarin fluorescence upon cleavage (excitation 325 nm, emission 393 nm). For all peptides except LF15, the k_{cat}/K_m was determined by measuring the cleavage rate at 1 μM peptide (where [S] << K_m; [S] represents concentration of substrate). For the LF15 peptide, k_{cat} (3.4 s^{-1}) and K_m (85 nM) were determined individually by measuring the initial rate at various peptide concentrations. Values reflect the average of three separate determinations ± s.d.

Table 3 Potency of peptide-based LF inhibitors

Compound	K_i^{app} (μM)
Acetyl-KVYP-hydroxamate	>100
PLG-hydroxamate	>100
MKARRKKVYP-hydroxamate	0.0011 ± 0.0002
SHAc-YPM	11 ± 3
GM6001	2.1 ± 0.2

K_i^{app} values were determined by measuring inhibition of peptide cleavage (1 μM LF15 for the 10-mer hydroxamate or 1 μM LF10 for all other compounds) over a range of inhibitor concentrations. Values are the mean ± s.d. of three separate determinations, each done in triplicate.

Both SHAc-YPM and GM6001 inhibited cleavage of MKK proteins by LF *in vitro* with potency comparable to their ability to inhibit cleavage of the peptide substrate (**Fig. 1a** and data not shown). GM6001 also partially inhibited cleavage of MKKs in a LeTx-treated macrophage cell line (**Fig. 1b**). Notably, LF inhibition by GM6001 in cultured cells was sufficient to protect them from LeTx-induced cell death (**Fig. 1c**,**d**). Neither the thioacetyl compound nor the long C-terminal peptide hydroxamate was active in cell culture, presumably owing to poor cell permeability or metabolic instability (data not shown). We also found that the inhibitory potency of the C-terminal peptide hydroxamate (but not that of any of the other compounds) was substantially poorer when evaluated at physiological salt concentrations, which are much higher than for standard assay conditions for LF *in vitro* (data not shown). GM6001 could also prevent cell death when added as late as 3 h after LeTx, suggesting that it can protect cells subsequent to internalization of the toxin (**Fig. 1e**). These results indicate that small molecule metalloproteinase inhibitors provide a means to neutralize the biological activity of anthrax toxin.

Structures of LF in complex with peptides and inhibitors

To understand the molecular basis for substrate selectivity by LF and to guide further inhibitor design, we solved the X-ray crystal structures of LF in complex with a consensus peptide, LF20 (both in a zinc-free state and in an active site mutant with zinc), and with two of the inhibitors reported here, GM6001 and SHAc-YPM, both in the presence of zinc (**Fig. 2a–c** and **Table 4**). Crystals soaked in the MKARRKKVYP C-terminal hydroxamate showed additional electron density around the active site, but this was not interpretable as a single atomic model.

The LF20 peptide (MLARRKKVYPYPMEPTIAEG-amide) incorporates consensus residues (P5–P4′) surrounding the scissile bond based on the peptide library screen, flanked by residues of authentic MKK2. In the crystal structure of the zinc-free LF20 complex, nine peptide residues (from the P3 valine to the P6′ threonine) are defined by electron density; in the zinc-bound active site mutant, the peptide lies in the same location, and a further two residues at the N terminus are visible (lysines P5 and P4); whereas residues downstream of the cleavage site are in general less well defined, suggestive of partial cleavage. The peptide binds in an extended conformation, along the 40 Å–long substrate recognition groove (formed by domains II–IV) that was previously defined by soaking an MKK2-derived peptide into LF crystals[31] (**Fig. 2a**,**d**,**e**). However, the present complex structure is at substantially higher resolution than that of the earlier study, and, as expected, the LF20 binds more strongly than the MKK2 peptide. The new crystallographic data unequivocally demonstrate that the binding

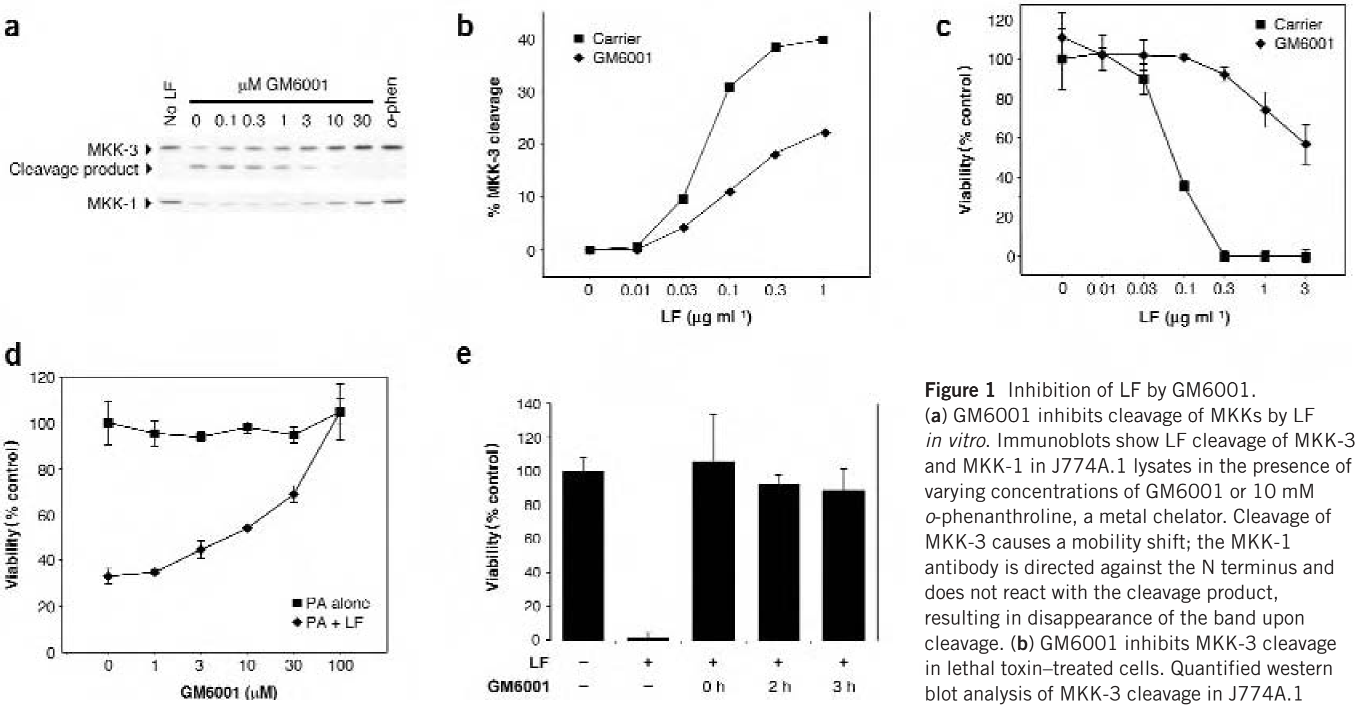


Figure 1 Inhibition of LF by GM6001. (**a**) GM6001 inhibits cleavage of MKKs by LF *in vitro*. Immunoblots show LF cleavage of MKK-3 and MKK-1 in J774A.1 lysates in the presence of varying concentrations of GM6001 or 10 mM *o*-phenanthroline, a metal chelator. Cleavage of MKK-3 causes a mobility shift; the MKK-1 antibody is directed against the N terminus and does not react with the cleavage product, resulting in disappearance of the band upon cleavage. (**b**) GM6001 inhibits MKK-3 cleavage in lethal toxin–treated cells. Quantified western blot analysis of MKK-3 cleavage in J774A.1 treated with lethal toxin (0.5 μg ml^{-1} PA with the indicated concentrations of LF) in the absence or presence of 100 μM GM6001. (**c**) Protection of J774A.1 cells from lethal toxin–mediated cell death by GM6001. Cell viability as determined by MTT assay after lethal toxin treatment in the presence of 100 μM GM6001 or 0.2% (v/v) DMSO carrier. (**d**) Dose-dependent neutralization of lethal toxin by GM6001. J774A.1 cell viability determined by MTT assay after treatment with lethal toxin (0.5 μg ml^{-1} PA + 0.3 μg ml^{-1} LF) or PA alone (0.5 μg ml^{-1}) in the presence of the indicated concentrations of GM6001. (**e**) GM6001 protects J774A.1 cells when added subsequent to LeTx. Cell viability is shown after treatment with PA alone (0.4 μg ml^{-1}) or PA with LF (25 ng ml^{-1}), with GM6001 added to 100 μM at the indicated time after toxin addition.

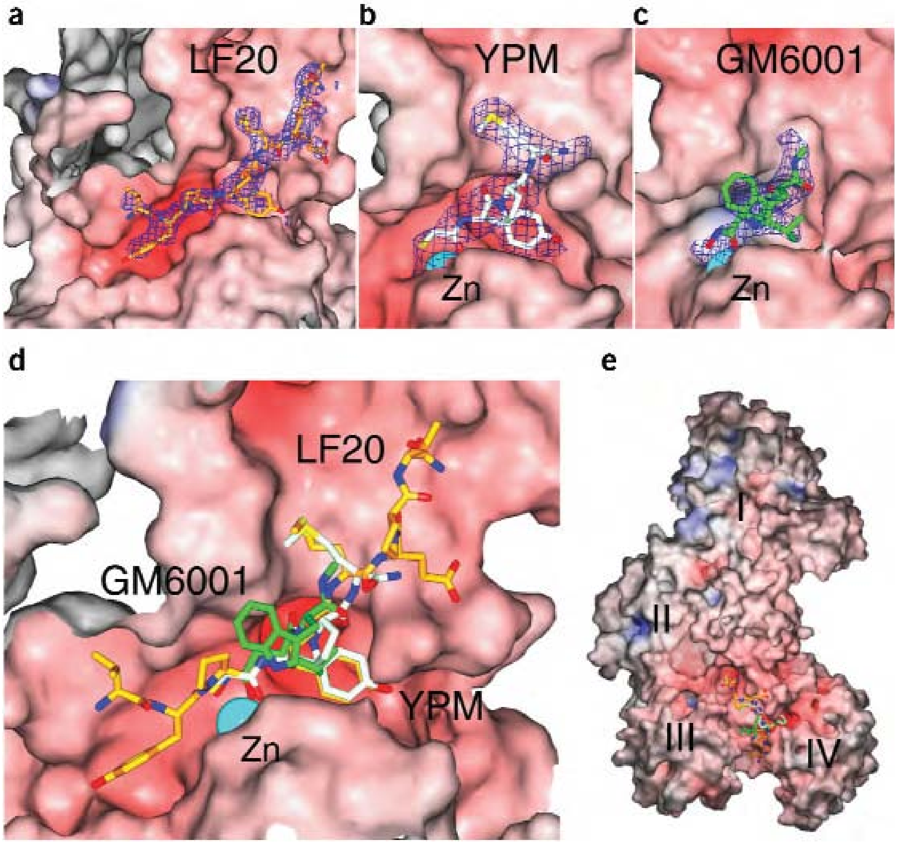


Figure 2 Structures of LF in complex with peptides and inhibitors. Molecular surface of LF is colored by charge (red, negative; blue, positive), with Zn^{2+} as a solid sphere (cyan) and the model of the peptide or inhibitor in ball-and-stick representation. The individual electron density surrounding each molecule is a $2F_o - F_c$ difference map calculated at the respective final resolution and contoured at 1.0 σ. (**a**) LF20 (yellow) in the absence of Zn^{2+}, resolution limit 2.85 Å. The model of bound LF20 shows the sequence VYPYPMEPT (residues 8–16 of the 20-residue-long LF20). This is the ordered region, and the electron density is clearly visible in difference maps ($2F_o - F_c$ and $F_o - F_c$) calculated from crystal X-ray diffraction data. (**b**,**c**) SHAc-YPM (white, labeled YPM), resolution limit 3.50 Å, and GM6001 (green), resolution limit 2.70 Å, respectively. Continuous electron density extends from the zinc atom to the metal-chelating moieties of the inhibitors (hydroxamate and thioacetyl, respectively). (**d**) The superposed individual complex structures of all three target molecules from **a**–**c** in the substrate-binding groove of LF, using the surface calculated for LF–LF20. The targets are all bound in the same N-to-C peptide orientation. (**e**) An overview of LF bound to the targets LF20, GM6001 and SHAc-YPM, superposed and colored as in **d**. The molecular surface was calculated from the LF–LF20 complex. The domains in LF are labeled I–IV. The catalytic site is in domain IV, where the zinc atom (not shown in this figure) is bound. These figures were prepared using SPOCK (http://mackerel.tamu.edu/spock/).

mode conforms to the canonical thermolysin substrate-binding mode[32]. The LF20 peptide is bound in a productive conformation, in contrast to that previously inferred from the LF–MKK2 structure[31], where the peptide is bound in a nonproductive mode (the reverse orientation and 6 Å distant from the active site). Therefore, the new complex structures, Protein Data Bank (PDB) entries 1PWV and 1PWW, supersede PDB entry 1JKY.

The ordered sequence of LF20 binds closely to the LF main chain and secondary structures surrounding the catalytic zinc-binding site. The P5 and P4 lysine residues lie close to a strongly acidic patch at the entrance to the active site, rationalizing the preference for basic residues at multiple positions upstream of the cleavage site. Residues P3–P1 form antiparallel β-sheet-like interactions with strand 4β3 of LF. The P2 tyrosine side chain occupies a fairly narrow hydrophobic pocket; this may explain the preference for tyrosine at this site. The P1′ tyrosine residue is buried within a deep hydrophobic S1′ pocket in LF, adjacent to the active site center. The pocket expands substantially on binding peptide (induced fit), including a ~3.5-Å shift of the main chain at Glu676 at the bottom of the pocket. Additionally, there is a ~3.0-Å shift of the side chain of Phe329, which is positioned along the substrate recognition groove, in close proximity to the active site and the bound peptide (this is also seen for all other bound ligands). The depth and plasticity of the S1′ cavity presumably allow the enzyme to accommodate large hydrophobic residues at the P1′ position; this explains why LF is most selective at this site.

The SHAc-YPM inhibitor shares three residues with the LF20 peptide downstream of the cleavage site, and the corresponding peptide electron density and derived model are markedly similar, with the P1′ tyrosine buried in the S1′ pocket (**Fig. 2b**,**d**,**e**). The thioacetyl moiety was modeled in a bidentate conformation[33,34] with the carbonyl oxygen atom and thiol sulfur atom directed toward the zinc. For the LF(E687C)–GM6001–Zn^{2+} complex (**Fig. 2c–e**), where LF(E687C) represents the LF E687C mutant, the peptide binds in a similar location. We modeled the hydroxamate moiety in the conventional bidentate planar conformation[27,32,33,35–37], with the carbonyl and hydroxyl oxygen atoms directed toward the zinc. The P1′ side chain is a leucine mimetic and binds in the S1′ pocket. The smaller side chain induces correspondingly less expansion of the S1′ pocket. The tryptophan side chain at the P2′ position makes no specific contacts with the protein, suggesting that it does not contribute to specificity.

DISCUSSION

The three independent LF-complex structures reported here indicate several common features essential for optimized substrate and inhibitor binding. The long hydrophobic substrate-binding groove and deep S1′ pocket adjacent to the catalytic Zn^{2+}-binding site seem to be the main determinants for strong target affinity. This strong hydrophobic selectivity has also been indicated by experimental data from nonpeptidic small molecule drug library screens of Panchal *et al.*[38] (this issue). These structures will enable the design of compounds with greater complementarity to the S1′ pocket and substrate recognition groove, combined with metal chelating groups spaced appropriately to allow for highly potent inhibition of LF.

Given the success of protease inhibition in the treatment of cardiovascular disease and AIDS, small molecule LF inhibitors would seem to be the most likely source for new drugs to treat anthrax. The possibility of encountering either naturally occurring or engineered antibiotic-resistant strains suggests that the availability of such

Table 4 Data collection summary for LF-complex crystals

	LF–LF20	LF(E687C)–LF20–Zn	LF–SHAc-YPM–Zn	LF(E687C)–GM6001–Zn
Data collection				
Space group	*P*2$_1$	*P*2$_1$	*P*2$_1$	*P*2$_1$
Cell dimensions (Å)				
a	96.70	96.70	96.70	96.70
b	137.40	137.40	137.40	137.40
c	98.30	98.30	98.30	98.30
Wavelength (Å)	1.07	0.98	1.08	0.97
Resolution range (Å)	50.0–2.85	30.0–2.80	30.0–3.50	50.0–2.70
Total reflections	96,701	94,088	91,831	255,861
Unique reflections	55,398	54,931	28,731	72,275
Completeness (%)[a]	92.2 (90.0)	86.8 (76.0)	90.8 (84.9)	99.6 (98.8)
R_{sym} (%)[a,b]	10.5 (48.6)	6.6 (40.9)	15.9 (45.1)	8.3 (48.0)
$I / \sigma I$[a]	6.7 (1.4)	12.2 (2.2)	7.4 (2.5)	15.6 (2.5)
Refinement statistics				
R_{work} (%)[b,c]	23.1	23.0	23.2	23.0
R_{free} (%)[b,c]	28.3	27.7	29.5	26.8

[a]Values in parentheses are for the highest-resolution shell. [b]$R_{sym} = \Sigma|I - \langle I\rangle| / \Sigma\langle I\rangle$, where I is the observed intensity and $\langle I\rangle$ is the average intensity from multiple observations of symmetry-related reflections. [c]R-factor $= \Sigma||F_o| - |F_c|| / \Sigma|F_o|$; R_{work} represents reflections not in R_{free} set; R_{free} represents 5% of a random selection of data not used during refinement.

compounds would be crucial in minimizing potentially large numbers of deaths. The work described here creates many paths toward the production of such drugs, both by enabling the rapid screening of chemical libraries and by providing a structural basis for rational drug design. Our results suggest in particular that sizable libraries of MMP inhibitors already in existence are likely to contain additional LF inhibitors, perhaps with increased potency and specificity. This work also illustrates the utility of peptide libraries for both the rapid optimization of substrate peptides and the generation of lead compounds. Such methods should be generally applicable to any protease of interest as a therapeutic target.

METHODS

Peptide library methods. Cleavage site selectivity for LF was determined by modification of described methods[21]. Libraries were custom synthesized at the Tufts University Core Facility (Boston). Degenerate positions ('X') were prepared using isokinetic mixtures to produce equimolar amounts of the 19 proteogenic amino acids excluding cysteine. For determination of the primed side selectivity, the library acetyl-KKKPTPXXXXXXAK (1 mM) was digested with LF[39] to 5–10% completion in a 10 μl reaction containing 20 mM HEPES, pH 7.4, 100 mM NaCl. The reaction products were analyzed by N-terminal peptide sequencing on an Applied Biosystems Procise 494 automated Edman sequencer. To determine the unprimed side selectivity, the library MXXXXXPYPMEDK(K-biotin) (20 μl at 1 mM) was digested to 5% completion as above, and quenched by adding an equal volume of 10 mM *o*-phenanthroline. The reaction products were incubated in batches with 500 μl avidin agarose (Sigma) in 500 μl of 25 mM ammonium bicarbonate with tumbling for 1 h, at which time the slurry was transferred to a column. The flowthrough and wash were combined, evaporated under reduced pressure and analyzed by Edman sequencing as described above.

Peptide cleavage assays. All peptides were synthesized at the Tufts University Core Facility except C-terminal peptide hydroxamates (Genemed Synthesis). Concentrations were determined based on the absorbance of the coumarin group (ε_{328} = 12,900 M^{-1} cm^{-1}) for the peptides and on tyrosine absorbance (ε_{280} = 1,200 M^{-1} cm^{-1}) for the inhibitors. Peptide cleavage assays were carried out in a Molecular Devices Spectramax Gemini XS fluorescence plate reader in black 96-well plates using LF10 digested to completion (which results in a 12-fold increase in fluorescence) as a standard. Reactions were run at 25 °C in 20 mM HEPES, pH 7.4, 0.1 mg ml^{-1} BSA (plus 1 mM DTT for assays of the thioacetyl inhibitor or 0.01% (v/v) Brij 35 for assays of the ten-residue hydroxamate inhibitor). For k_{cat} / K_m determinations, LF was used at 2–20 nM and the rates were determined from the linear range of the reaction progress curve (<10% substrate turnover). For the LF15 peptide, rates were determined in a continuous assay at varying substrate concentrations using a Photon Technology International Fluorescence system using 2 nM LF under the conditions described above, using the peptide at 1 μM digested to completion (eight-fold increase in fluorescence) as a standard. Data were corrected for the inner filter effect by measuring the quenching of an Mca-peptide standard at each substrate concentration. Data were fitted directly to the Michaelis-Menten equation. Peptide cleavage sites were confirmed by Edman sequencing of the reaction products.

Analysis of MKK cleavage. For *in vitro* MKK cleavage, J774A.1 cells were lysed in 0.5% (v/v) Igepal CA-630, 20 mM HEPES, pH 7.4, 100 mM NaCl, 1 mM DTT, 5% (v/v) glycerol, 1 mM PMSF, and 4 μg ml^{-1} each of leupeptin, pepstatin and aprotinin. LF was preincubated for 30 min at 25 °C with varying concentrations of inhibitor before the addition of J774A.1 cell lysate. After an additional 30 min the reaction was quenched by adding SDS-PAGE loading buffer. To analyze cleavage in cultured cells, J774A.1 cells in six-well plates were pretreated with GM6001 (CALBIOCHEM) or DMSO carrier alone (0.2% (v/v) final concentration in complete media) for 30 min at 37 °C before adding PA (to 0.5 μg ml^{-1}) and LF (to the indicated concentration). Cells were incubated at 37 °C for an additional 90 min, washed once with PBS and then lysed directly in SDS-PAGE loading buffer (100 μl per well) and boiled 10 min. Samples were fractionated by SDS-PAGE and transferred to PVDF membrane for immunoblotting with anti-MKK-3 (Santa Cruz Biotechnology C-19) or anti-MKK-1 N terminus (Upstate Biotechnology, catalog no. 06-269). MKK-3 cleavage was quantified using NIH Image software (http://rsb.info.nih.gov/nih-image/).

Lethal toxin assays. J774A.1 cells were plated in 96-well dishes at 3 × 10^5 cells per well and allowed to recover for 16 h, after which the medium was removed and replaced with fresh complete medium (100 μl per well) containing the indicated concentration of GM6001 or carrier alone (0.2% (v/v) DMSO). After 30 min, PA and/or LF were added to the indicated concentrations and incubation continued for an additional 4 h. To assay viability, 10 μl of 5 mg ml^{-1} MTT in PBS was added to each well, and incubation was continued for 2 h before aspirating the supernatant and extracting with 0.1 M HCl in isopropanol. Absorbance at 570 nm with a background correction at 690 nm was determined in an absorbance plate reader.

Crystallization. LF wild type and E687C active site mutant protein crystals were grown in 1.7 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0, 2 mM EDTA by the hanging-drop vapor diffusion method, at 25 ± 4 °C, using a protein concentration of 13 mg ml^{-1} (ref. 31). Cocrystals of LF with GM6001 grew under similar conditions. All crystals used are monoclinic, in space group *P*2$_1$, with unit cell dimensions a = 96.7 Å, b = 137.4 Å, c = 98.3 Å, α = 90°, β = 98.0°, γ = 90°, and contain two molecules per asymmetric unit. In general, similar features were observed at the two active sites, but the density for Molecule B was stronger.

LF–substrate and LF–inhibitor complexes. Native LF or LF E687C monoclinic *P*2$_1$ single crystals were harvested and bathed in several rounds of crystallization buffer prior to soaking in their respective target peptide or inhibitor solutions. Soaks were done at room temperature, 23 °C ± 2 °C. The treated crystals were then individually flash-frozen in liquid nitrogen. All data was collected was at 100 K, in a nitrogen cryostream.

The wild-type LF–LF20 peptide complex was obtained by soaking crystals in a solution of 10 mM LF20, 1.8 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0, 2 mM EDTA for 8 min. Each crystal was then transferred into a cryoprotectant solution of 10 mM LF20, 2.4 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0, 2 mM EDTA, 25% (v/v) glycerol, and bathed for a further 1 min before mounting in a cryoloop and flash-freezing. The LF(E687C)–LF20–Zn^{2+} crystal complex was first soaked in a solution of 1 mM $ZnSO_4$, 1.8 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0 for 5 min, followed by the treatment as described for the wild-type LF–LF20 complex.

The LF–SHAc-YPM inhibitor–Zn^{2+} complex was obtained by soaking crystals in 1 mM $ZnSO_4$, 1.8 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0 for 5 min; then in 5 mM SHAc-YPM, 1.8 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0 for a further 5 min; and then in 5 mM SHAc-YPM, 2.4 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0, 2 mM EDTA, 25% (v/v) glycerol for 1 min before mounting and freezing.

The LF–GM6001 and LF(E687C)–GM6001 inhibitor complex crystals were grown from a 1:2 molar ratio of LF to inhibitor and crystallized as for native. Crystals were soaked in 1 mM $ZnSO_4$, 1.8 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0 for 5 min, then in 0.1 mM GM6001 (0.7% (v/v) DMSO), 1.8 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0 for 2 min, and finally in 0.1 mM GM6001 (0.7% (v/v) DMSO), 2.4 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0, 2 mM EDTA, 25% (v/v) glycerol for <1 min before mounting and freezing. Using a LF(E687C)–GM6001 cocrystal, the LF(E687C)–GM6001–Zn^{2+} inhibitor complex crystal was also prepared with the method described here. No substantial differences in target binding or active site conformation between wild type or mutant LF–GM6001–Zn^{2+} complexes were observed (residue 687 is not involved directly in inhibitor or zinc binding). As the LF(E687C)–GM6001–Zn^{2+} complex gave higher-resolution data, this complex was used in further refinement.

Data collection. Data for the LF–LF20, LF(E687C)–LF20–Zn^{2+}, and LF–SHAc-YPM complexes were collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Menlo Park, California, USA), on beamlines 1-5 (wavelength = 1.07 Å), 9-1 (wavelength = 0.98 Å) and 7-1 (wavelength = 1.08 Å). Data for the LF(E687C)–GM6001–Zn^{2+} complex were collected at the National Synchrotron Light Source (NSLS, Brookhaven, New York, USA) on beamline x12c (wavelength = 0.97 Å). X-ray diffraction data were collected for LF–LF20, LF(E687C)–LF20–Zn^{2+}, LF–SHAc-YPM–Zn^{2+}, and LF(E687C)–GM6001–Zn^{2+} to resolution limits of 2.85 Å, 2.80 Å, 3.50 Å and 2.70 Å, respectively.

Data processing and refinement. Crystallographic data were processed using the HKL package[40]. Refinement and model building were done in CNS[41] and O[42]. The high-resolution model of LF (PDB entry 1J7N)[31] was used as the starting model. The model was put through rigid body refinement and then minimization, and initial maps were calculated. Additional electron density at ≥1.0 σ in $2F_o - F_c$ and 2 σ in $F_o - F_c$ maps was clearly seen in the active site groove of LF for all cases. The model of the peptide or inhibitor with zinc was then built into this position and further refined in CNS[40]. Difference maps of the LF models, including peptide or inhibitor, and also omitting the peptide or inhibitor, were calculated in subsequent rounds of model rebuilding and refinement. Composite omit maps were also used. The final *R*-factors for each complex were as follows: LF–LF20 (Zn^{2+}-free), $R_{free} = 28.3\%$ and $R_{work} = 23.1\%$; LF(E687C)–LF20–Zn^{2+}, $R_{free} = 27.7\%$ and $R = 23.0\%$; LF–SHAc-YPM–Zn^{2+}, $R_{free} = 29.5\%$ and $R = 23.2\%$; and LF(E687C)–GM6001–Zn^{2+}, $R_{free} = 26.8\%$ and $R = 23.0\%$. The final models fall within or exceed the limits of all the quality criteria of PROCHECK from the CCP4 suite[43].

Coordinates. Coordinates and structure factors have been deposited in the Protein Data Bank (accession codes: 1PWQ, LF–YPM–Zn^{2+}; 1PWU, LF(E687C)–GM6001–Zn^{2+}; 1PWV, LF–LF20; 1PWW, LF(E687C)–LF20–Zn^{2+}).

ACKNOWLEDGMENTS

Thanks to P. Bartlett (University of California Berkeley), D. Tronrud and B. Matthews (University of Oregon) and B. Rupp (Lawrence Livermore National Laboratory) for pointing out the canonical binding mode for Zn metalloproteases and to E. Garman (University of Oxford) and A. Gonzalez (SSRL) for discussions on the crystallography. We also thank H. Robinson and S. Vaday for collecting data at the National Synchrotron Light Source (NSLS). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the US Department of Energy (DOE), Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the DOE, Office of Biological and Environmental Research, and by the NIH, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. Data for this work were also collected at the NSLS, Brookhaven National Laboratory, which is supported by the DOE, Division of Materials Sciences and Division of Chemical Sciences, under contract no. DE-AC02-98CH10886. This work was supported by NIH (R.J.C., R.C.L. and L.C.C.), the US National Science Foundation (L.C.C.) and the US Department of the Army (DAMD17-03-1-0062 to L.C.C.). The US Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick, Maryland 21702-5014 is the awarding and administering acquisition office. The contents of this manuscript do not necessarily reflect the position or policy of the US government, and no official endorsement should be inferred. B.E.T. is a Leukemia and Lymphoma Society special fellow.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Structural & Molecular Biology* website for details).

Received 2 July; accepted 23 October 2003

---

1. Dixon, T.C., Meselson, M., Guillemin, J. & Hanna, P.C. Anthrax. *New Engl. J. Med.* **341**, 815–826 (1999).
2. Duesbery, N.S. & Vande Woude, G.F. Anthrax toxins. *Cell. Mol. Life Sci.* **55**, 1599–1609 (1999).
3. Moayeri, M., Haines, D., Young, H.A. & Leppla, S.H. *Bacillus anthracis* lethal toxin induces TNF-α-independent hypoxia-mediated toxicity in mice. *J. Clin. Invest.* **112**, 670–682 (2003).
4. Pezard, C., Berche, P. & Mock, M. Contribution of individual toxin components to virulence of *Bacillus anthracis. Infect. Immun.* **59**, 3472–3477 (1991).
5. Sellman, B.R., Mourez, M. & Collier, R.J. Dominant-negative mutants of a toxin subunit: an approach to therapy of anthrax. *Science* **292**, 695–697 (2001).
6. Mourez, M. *et al.* Designing a polyvalent inhibitor of anthrax toxin. *Nat. Biotechnol.* **19**, 958–961 (2001).
7. Duesbery, N. *et al.* Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**, 734–737 (1998).
8. Vitale, G. *et al.* Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/threonine phosphorylation of MAPKs in cultured macrophages. *Biochem. Biophys. Res. Commun.* **248**, 706–711 (1998).
9. Pellizzari, R., Guidi-Rontani, C., Vitale, G., Mock, M. & Montecucco, C. Anthrax lethal factor cleaves MKK3 in macrophages and inhibits the LPS/IFNγ-induced release of NO and TNFα. *FEBS Lett.* **462**, 199–204 (1999).
10. Vitale, G., Bernardi, L., Napolitani, G., Mock, M. & Montecucco, C. Susceptibility of mitogen-activated protein kinase kinase family members to proteolysis by anthrax lethal factor. *Biochem. J.* **352**, 739–745 (2000).
11. Enslen, H. & Davis, R.J. Regulation of MAP kinases by docking domains. *Biol. Cell* **93**, 5–14 (2001).
12. Agrawal, A. *et al.* Impairment of dendritic cells and adaptive immunity by anthrax lethal toxin. *Nature* **424**, 329–334 (2003).
13. Friedlander, A.M. Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process. *J. Biol. Chem.* **261**, 7123–7126 (1986).
14. Hanna, P.C., Acosta, D. & Collier, R.J. On the role of macrophages in anthrax. *Proc. Natl. Acad. Sci. USA* **90**, 10198–10201 (1993).
15. Park, J.M., Greten, F.R., Li, Z.W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048–2051 (2002).
16. Chopra, A.P., Boone, S.A., Liang, X. & Duesbery, N.S. Anthrax lethal factor proteolysis and inactivation of MAPK kinase. *J. Biol. Chem.* **278**, 9402–9406 (2003).
17. Hammond, S.E. & Hanna, P.C. Lethal factor active-site mutations affect catalytic activity *in vitro. Infect. Immun.* **66**, 2374–2378 (1998).
18. Cummings, R.T. *et al.* A peptide-based fluorescence resonance energy transfer assay for *Bacillus anthracis* lethal factor protease. *Proc. Natl. Acad. Sci. USA* **99**, 6603–6606 (2002).
19. Tonello, F., Seveso, M., Marin, O., Mock, M. & Montecucco, C. Screening inhibitors of anthrax lethal factor. *Nature* **418**, 386 (2002).
20. Songyang, Z. *et al.* SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767–778 (1993).
21. Turk, B.E., Huang, L.L., Piro, E.T. & Cantley, L.C. Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nat. Biotechnol.* **19**, 661–667 (2001).
22. Tanoue, T., Adachi, M., Moriguchi, T. & Nishida, E. A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat. Cell Biol.* **2**, 110–116 (2000).
23. Enslen, H., Brancho, D.M. & Davis, R.J. Molecular determinants that mediate selective activation of p38 MAP kinase isoforms. *EMBO J.* **19**, 1301–1311 (2000).
24. Xu, B., Stippec, S., Robinson, F.L. & Cobb, M.H. Hydrophobic as well as charged residues in both MEK1 and ERK2 are important for their proper docking. *J. Biol. Chem.* **276**, 26509–26515 (2001).
25. Holmquist, B. & Vallee, B.L. Metal-coordinating substrate analogs as inhibitors of

metalloenzymes. *Proc. Natl. Acad. Sci. USA* **76**, 6216–6220 (1979).
26. Moore, W.M. & Spilburg, C.A. Purification of human collagenases with a hydroxamic acid affinity column. *Biochemistry* **25**, 5189–5195 (1986).
27. Gowravaram, M.R. *et al.* Inhibition of matrix metalloproteinases by hydroxamates containing heteroatom-based modifications of the P1′ group. *J. Med. Chem.* **38**, 2570–2581 (1995).
28. Baxter, A.D. *et al.* A novel series of matrix metalloproteinase inhibitors for the treatment of inflammatory disorders. *Bioorg. Med. Chem. Lett.* **7**, 897–902 (1997).
29. Grobelny, D., Poncz, L. & Galardy, R.E. Inhibition of human skin fibroblast collagenase, thermolysin, and *Pseudomonas aeruginosa* elastase by peptide hydroxamic acids. *Biochemistry* **31**, 7152–7154 (1992).
30. Levy, D.E. *et al.* Matrix metalloproteinase inhibitors: a structure-activity study. *J. Med. Chem.* **41**, 199–223 (1998).
31. Pannifer, A.D. *et al.* Crystal structure of the anthrax lethal factor. *Nature* **414**, 229–233 (2001).
32. Holmes, M.A. & Matthews, B.W. Binding of hydroxamic acid inhibitors to crystalline thermolysin suggests a pentacoordinate zinc intermediate in catalysis. *Biochemistry* **20**, 6912–6920 (1981).
33. Grams, F. *et al.* X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors. Implications for substrate binding and rational drug design. *Eur. J. Biochem.* **228**, 830–841 (1995).
34. Gaucher, J.F. *et al.* Crystal structures of α-mercaptoacyldipeptides in the thermolysin active site: structural parameters for a Zn monodentation or bidentation in metalloendopeptidases. *Biochemistry* **38**, 12569–12576 (1999).
35. Dhanaraj, V. *et al.* X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. *Structure* **4**, 375–386 (1996).
36. Chen, L. *et al.* Crystal structure of the stromelysin catalytic domain at 2.0 Å resolution: inhibitor-induced conformational changes. *J. Mol. Biol.* **293**, 545–557 (1999).
37. Roswell, S. *et al.* Crystal structure of human MMP9 in complex with a reverse hydroxamate inhibitor. *J. Mol. Biol.* **319**, 173–181 (2002).
38. Panchal, R. *et al.* Identification of small molecule inhibitors of anthrax lethal factor. *Nat. Struct. Mol. Biol.* **11**, 67–72 (2004).
39. Roberts, J.E., Watters, J.W., Ballard, J.D. & Dietrich, W.F. Ltx1, a mouse locus that influences the susceptibility of macrophages to cytolysis caused by intoxication with *Bacillus anthracis* lethal factor, maps to chromosome 11. *Mol. Microbiol.* **29**, 581–591 (1998).
40. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
41. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
42. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
43. Bailey, S. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

nature
structural &
molecular biology

Identification of small molecule inhibitors of anthrax lethal factor

Rekha G Panchal[1], Ann R Hermone[1,5], Tam Luong Nguyen[1,5], Thiang Yian Wong[2], Robert Schwarzenbacher[2], James Schmidt[3], Douglas Lane[1], Connor McGrath[1], Benjamin E Turk[4], James Burnett[1], M Javad Aman[3], Stephen Little[3], Edward A Sausville[1], Daniel W Zaharevitz[1], Lewis C Cantley[4], Robert C Liddington[2], Rick Gussio[1] & Sina Bavari[3]

The virulent spore-forming bacterium *Bacillus anthracis* secretes anthrax toxin composed of protective antigen (PA), lethal factor (LF) and edema factor (EF). LF is a Zn-dependent metalloprotease that inactivates key signaling molecules, such as mitogen-activated protein kinase kinases (MAPKK), to ultimately cause cell death. We report here the identification of small molecule (nonpeptidic) inhibitors of LF. Using a two-stage screening assay, we determined the LF inhibitory properties of 19 compounds. Here, we describe six inhibitors on the basis of a pharmacophoric relationship determined using X-ray crystallographic data, molecular docking studies and three-dimensional (3D) database mining from the US National Cancer Institute (NCI) chemical repository. Three of these compounds have K_i values in the 0.5–5 μM range and show competitive inhibition. These molecular scaffolds may be used to develop therapeutically viable inhibitors of LF.

Anthrax, a disease caused by *Bacillus anthracis*, has recently been the subject of intense interest because of its use as a biological weapon against human populations. The inhalation of *B. anthracis* spores is often fatal if the condition is not properly diagnosed and treated with antibiotics during the early stages of infection. In many cases antibiotic regimes may not be effective, especially if there is bacterium overload, which causes large amounts of lethal toxin to be released. Hence, a new level of adjunct treatment is needed to inactivate the toxins released by *B. anthracis*.

Anthrax toxin (AT) consists of three proteins: lethal factor, protective antigen and edema factor, all of which work in concert to kill host cells. Initially, PA binds to an AT receptor[1,2] on the host cell surface, where it is subsequently cleaved by furin (or furin-like proteases) to produce a 20-kDa N-terminal fragment (PA_{20}) and a 63-kDa C-terminal fragment (PA_{63})[3,4]. After cleavage, seven PA_{63} monomers assemble to form a heptameric prepore capable of binding both LF and EF. Upon binding of LF or EF, the entire complex undergoes receptor-mediated endocytosis. It is hypothesized that the acidic endosomal environment causes a conformational change in the PA_{63} heptamer to produce a functional pore that traverses the membrane and translocates the two enzymatic moieties LF and EF into the cell cytosol. EF is a calmodulin-dependent adenylate cyclase[5]; LF is a Zn-dependent metalloprotease that cleaves several members of the MAPKK family near the N terminus[6,7]. This cleavage prevents interaction with, and phosphorylation of, downstream MAPK[8], thereby inhibiting one or more signaling pathways. Through a mechanism that is not yet well understood, this results in the death of the host. Recent studies suggest that the inactivation of p38 MAPK induces apoptosis in LF-exposed macrophages, thereby preventing the release of chemokines and cytokines, and preventing the immune system from responding to the pathogen[9].

Based on the current understanding of the mechanism of anthrax toxin, methods may be developed to inhibit various steps in toxin assembly and/or function. In one antitoxin therapy approach, dominant-negative PA mutants have been generated that coassemble with the wild-type PA protein, blocking the translocation of LF and EF across the cell membrane. Such PA mutants are potent inhibitors of anthrax toxin in both cell-based assays and *in vivo* animal models[10,11]. In a second approach, a peptide inhibitor that binds to the heptameric PA and prevents the interaction of PA with LF and EF has shown efficacy in animals[12].

The lethal action of anthrax toxin may also be inactivated by molecules that inhibit the protease activity of LF. So far, the only known small molecule inhibitors of LF are nonspecific hydroxymates that are effective at >100 μM concentration[13] and more recently reported hydroxymate derivatives of peptide substrate that inhibit LF at nanomolar concentrations[14]. In this study, we identified several small (nonpeptidic) compounds that inhibit anthrax LF protease activity with K_i values in the 0.5–5 μM range. We approached anthrax therapeutic development (in parallel with the peptidomimetic approach used by Turk *et al.*[15]; this issue) using structure-based discovery to

---

[1]Developmental Therapeutics Program, NCI Frederick, Frederick, Maryland 21702-1201, USA. [2]The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. [3]United States Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Frederick, Maryland 21702, USA. [4]Division of Signal Transduction, Beth Israel Deaconess Medical Center, Harvard Institutes of Medicine, Room 1007, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to R.G.P. (panchal@dtpax2.ncifcrf.gov) or S.B. (bavaris@ncifcrf.gov).

doi:10.1038/nsmb711

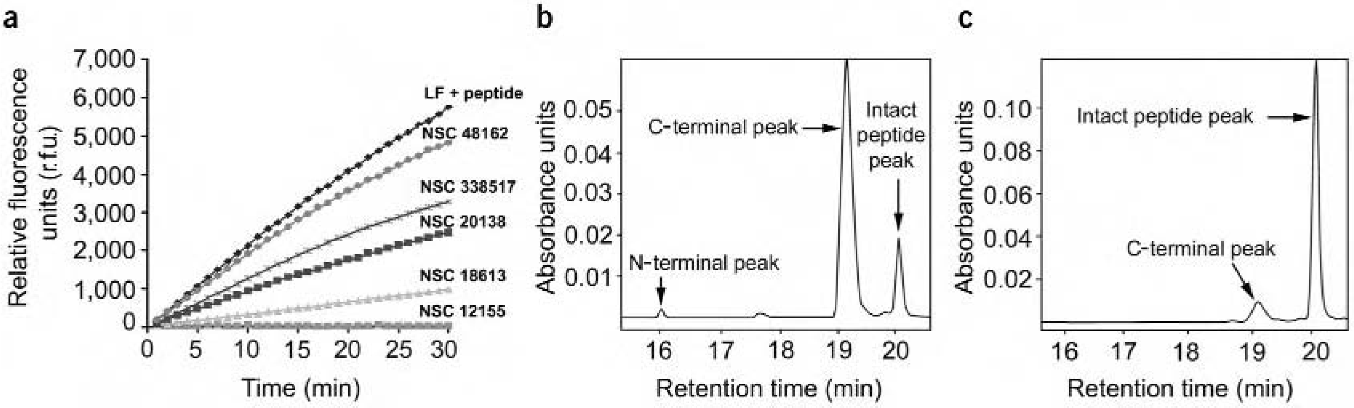


Figure 1 A two-stage assay for screening and validating small molecule inhibitors of anthrax lethal factor. (**a**) Representative data from a fluorescent plate reader assay showing different degrees of inhibition by compounds from the NCI Diversity Set. (**b**) HPLC-based assay without inhibitor, showing the N- and C-terminal cleavage products after incubation of the substrate with LF for 30 min. (**c**) HPLC-based assay with inhibitor NSC 12155 showing a reduced C-terminal peak area at 365 nm, indicating strong inhibition of LF activity.

identify small organic molecules as lead candidates. Specifically, we used molecular diversity screening combined with 3D database searching and molecular modeling. The LF X-ray crystal structure reported by Pannifer *et al.*[16] was useful during the structure-based drug discovery portion of these studies.

The first phase of this study involved a high-throughput screen (HTS) of small molecules from the NCI Diversity Set to identify LF inhibitors. Hits identified from the HTS were verified with an HPLC-based assay. Afterwards, we used X-ray crystallography and molecular modeling (conformational sampling, database mining and molecular docking) to identify additional lead therapeutics. Based on an iterative process of compound selection and biological testing, a pharmacophore for LF inhibitors was developed.

RESULTS

High-throughput screening and hit validation

To screen and identify compounds that inhibit LF activity, we developed a high-throughput fluorescence-based assay. An optimized peptide (KKVYPYPME; B.E.T. *et al.*, unpublished data) with a fluorogenic coumarin group at the N terminus and a 2,4-dinitrophenyl (dnp) quenching group at the C terminus was used as LF substrate for *in vitro* assays. After cleavage by LF, fluorescence increased (excitation and emission wavelengths, 325 and 394 nm, respectively). After standardization of the high-throughput assay, the 1,990 compounds in the NCI Diversity Set were tested (**Fig. 1a**). Compounds that showed >75% inhibition were selected for validation using an HPLC-based assay. This eliminated false positives due to fluorescence quenching by some of the test compounds. Using the HPLC-based assay (**Fig. 1b**,**c**), compounds that showed >50% inhibition were selected for further study. The HPLC assay, in addition to eliminating false positives, was a more rigorous test of LF inhibition, as a lower inhibitor concentration (20 µM) was used (compared with 100-µM concentration used in the fluorescence-based assays). Furthermore, the identified LF inhibitors did not inhibit a range of different proteases, thus confirming that these compounds did not inhibit LF promiscuously (see **Supplementary Fig. 1** online).

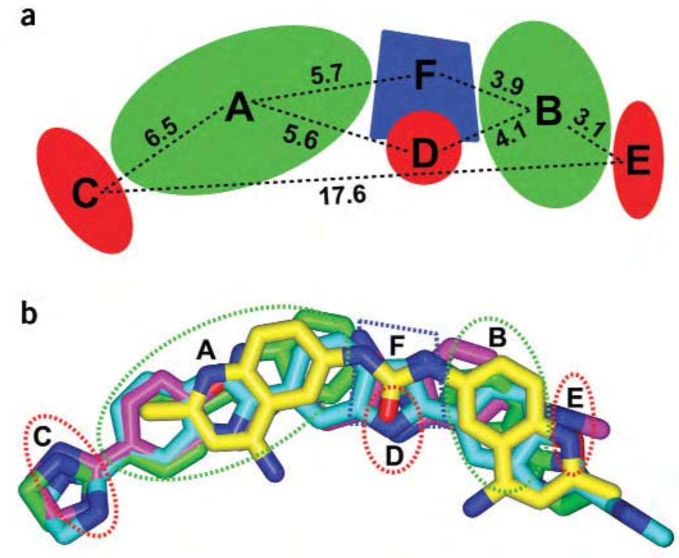
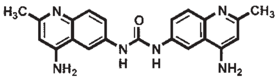
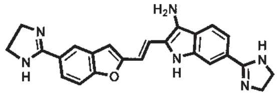
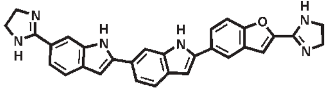
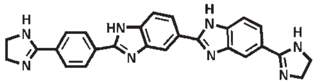
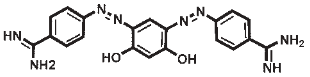
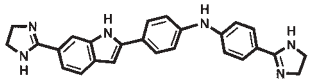
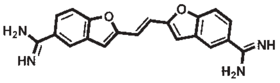


Figure 2 General pharmacophore model of the LF inhibitors. (**a**) Black dashed lines depict the distances between the various centroids of the pharmacophore centers. Green ellipses (A and B) are aromatic centers; red ellipses (C, D and E) are polar centers (hydrogen bond donors or acceptors); blue region (F) is a neutral linker that may include a variety of polar or hydrophobic groups. (**b**) Pharmacophoric overlap of LF inhibitors (stick rendering) and their correspondence to the general LF inhibitor pharmacophore shown in **Figure 2a**. The pharmacophoric overlap regions of compounds are highlighted in dashed lines (green, aromatic centers; blue, neutral (polar or hydrophobic groups acceptable) linker region; red, polar centers. For all structures: nitrogen, blue; oxygen, red. Carbon atoms for NSC 12155, yellow; for NSC 357756, magenta; for NSC 369721, green; for NSC 369728, light blue. The pharmacophore is based on the energy-refined X-ray conformation of NSC 12155 bound to LF. These data were combined with molecular docking studies of structurally related analogs (**Table 1**) from 3D database mining studies.

Table 1 Two-dimensional chemical representations of LF inhibitors with percent inhibition at a compound concentration of 20 μM, K_i values and type of inhibition

Structure	NSC number	% inhibition	K_i (μM)	Inhibition type
	12155	95	0.5 ± 0.18	Competitive
	357756	90	4.9 ± 1.7	Competitive
	369718	90	N.D.	N.D.
	369721	90	4.2 ± 0.21	Competitive
	359465	48	N.D.	N.D.
	377362	33	N.D.	N.D.
	240899	0	N.D.	N.D.

N.D., not determined.

Pharmacophoric features of anthrax LF inhibitors

We identified 19 compounds with >50% LF inhibition (at 20 μM inhibitor concentration) from the NCI Diversity Set screen. These included several organometallic and charged molecules. Here, we chose to concentrate on only relatively small organic compounds for structure-based studies, as these molecules are more likely to show therapeutic potential. The conformational spaces of two leads, NSC 12155 and NSC 357756, were subsequently explored to generate multiple pharmacophoric hypotheses, which were then used in 3D database mining studies to identify additional LF inhibitors. We carried out several iterations of this process, which consisted of 3D database mining of the entire NCI repository (as well as commercially available chemical repositories including the Available Chemicals Directory, MayBridge and BioByte) and subsequent biological testing, to identify new inhibitors. During this process >60 compounds were tested and most of them were inactive. However, six of the compounds, which showed a range of LF inhibitory potency, were used to develop and refine a consistent pharmacophore (**Fig. 2a**). A 3D superimposition of four of the most potent LF inhibitors (NSC 12155, NSC 357756, NSC 369718 and NSC 369721) (**Fig. 2b**) exhibits an excellent overlay of the polar heteroatoms and hydrophobic substituents of these molecules. The chemical structures of a range of identified LF inhibitors are shown in **Table 1**.

Kinetic studies

To determine the K_i values and types of inhibition mediated by the inhibitors (competitive, noncompetitive or uncompetitive), we determined kinetic constants of the peptide substrate and compared them with those obtained in the presence of different inhibitor concentrations. The K_m and V_{max} values for the LF-catalyzed hydrolysis of the peptide substrate were 19 μM and 1.1 μmol min^{-1} mg^{-1} of LF, respectively. NSC 12155, NSC 357756 and NSC 369721 showed competitive inhibition (**Table 1**), as they had no effect on the V_{max}, but $K_{m(app)}$ increased with inhibitor concentration (see **Supplementary Fig. 2** online).

Anthrax LF–NSC 12155 cocrystal structure

The crystal structure of LF in complex with NSC 12155 (the most potent inhibitor) was determined at a resolution of 2.9 Å (electron density map, **Fig. 3a**). NSC 12155 binds to the catalytic site of LF with its urea moiety close to the catalytic Zn atom (within 4 Å). One quinoline ring shows strong electron density near the side chain of His690, suggesting a favorable π-stacking interaction between the histidine's side chain imidazole and the quinoline ring (**Fig. 3b**). Conversely, the second quinoline showed poor electron density, indicating that there is more rotational freedom about its quinoline-urea bond. Despite the overall lack of a strong positional preference for this quinoline, a more consistent density was detected near its amino substitution, indicating a slightly greater preference for a 'C-shaped' conformation of NSC 12155 when bound to LF. This is consistent with the pharmacophoric overlap shown in **Figure 2b**.

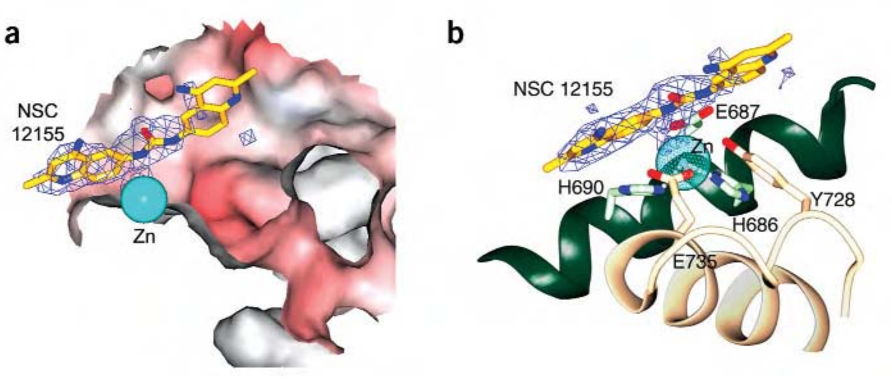


Figure 3 X-ray crystal structure of the LF–NSC 12155–Zn complex. The electron density surrounding NSC 12155 shown in these figures are $2F_o - F_c$ difference maps (see Methods) calculated at 2.9-Å resolution. (**a**) Detailed view of the electron density trace and overall model fit of NSC 12155. Molecular surface of LF colored by charge (red, negative; blue, positive), with Zn^{2+} (cyan), and the model of the inhibitor molecule NSC 12155 (yellow) in stick representation. The difference map, $2F_o - F_c$, is contoured at 1.1 σ level. (**b**) The inhibitor NSC 12155 bound in the active site of LF. The difference map, $2F_o - F_c$, is contoured at 1.0 σ. A portion of NSC 12155 appears nonrigid owing to a rotatable bond, and almost full electron density coverage is seen for this portion at a contour level of 0.6 σ. Inhibitor molecule (yellow), zinc-coordinating residues (H686, H690, E735) and catalytic residues (E687, Y728) are in stick representation. The Cα atoms of residues 680–694 (green, background) and 726–742 (beige, foreground) are in ribbon representation. The Zn^{2+} ion (cyan) is a lined sphere, and its hydrogen bonds with His686, His690 and Glu735 are represented as aligned small white spheres. These figures were prepared using SPOCK (http://mackerel.tamu.edu/spock/).

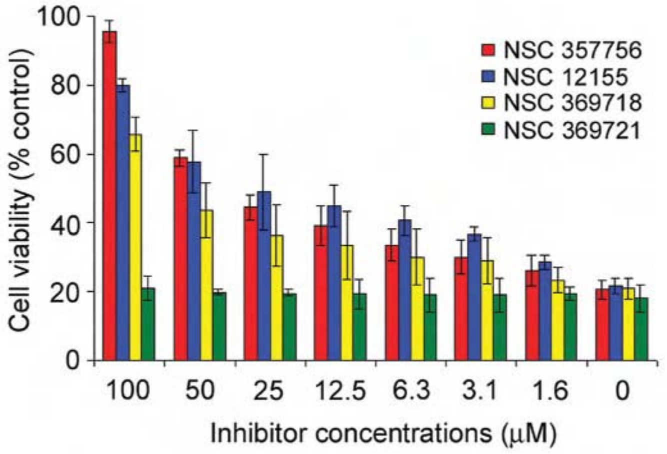


Figure 4 Efficacy of LF inhibitors in a cell-based toxicity assay. J774A.1 cells were pretreated with either DMSO control or various concentrations of inhibitors, and then incubated with anthrax lethal toxin. After 4 h, cell viability was determined with MTT dye.

Molecular docking studies

To further investigate whether the C conformation has an important role during the binding of NSC 12155 to LF, we used molecular docking to study the conformational preference of the freely rotating quinoline in the NSC 12155–LF model. Results from these analyses suggest that the NSC 12155 scaffold does prefer the planar C conformation to the 'L-shaped' conformation when bound to LF. This is further supported by the following: (i) quantum mechanical calculations at the level of density functional theory, as well as analysis of related crystal structures (data not shown), support a planar preference (either L or C shaped) for NSC 12155; (ii) rotation of the 'free' quinoline out of plane to its planar L conformation results in unfavorable hydrophobic-polar interactions between the amino groups of NSC 12155 and the side chain of Val675; (iii) in the planar C conformation, the urea oxo and quinoline amino substituents of NSC 12155 are more likely to engage in favorable intramolecular acid-base interactions; (iv) molecular docking studies of 32 substituted quinoline and urea derivatives (chemoinformatically mined from the NCI repository), which were inactive in the LF assay (data not shown), indicate that these scaffolds are either incapable of forming the preferred C conformation of NSC 12155 or lack features that would enable favorable binding; and (v) additional modeling studies of NSC 12155 indicate that the urea nitrogens are within range to form favorable acid-base interactions with the carboxylate of Glu687 (supported by X-ray data: distances of the urea nitrogens of NSC 12155 are 4.12 Å and 4.72 Å from OE1 and OE2 of Glu687, respectively).

Cytotoxicity assay

To determine the ability of the small molecule inhibitors to protect macrophages against LF, we pretreated the cells with NSC 12155, NSC 357756, NSC 369718 or NSC 369721 at concentrations ranging from 1 to 100 µM and further incubated them in the presence of anthrax lethal toxin. Cell viability was determined using MTT dye (**Fig. 4**). NSC 357756 showed 96% protection at 100 µM, whereas NSC 12155 and NSC 369718, the most potent of the LF inhibitors *in vitro*, showed lower protection at 100 µM. These three compounds showed some protection <25 µM, suggesting that they might be good leads against lethal toxin *in vivo*. Additionally, NSC 369721 was ineffective even at 100 µM in the cell-based toxicity assay. The moderate protection of these inhibitors is probably attributable to their limited ability to penetrate the macrophage cell membrane. The cell-based data will aid in the development of second-generation LF inhibitors.

DISCUSSION

Molecular docking studies of both inactive and active analogs of the compounds shown in **Table 1** are consistent with the common pharmacophore (**Fig. 2a**) proposed in this study. For example, the amidine groups of NSC 240899 formed unfavorable steric and polar interactions when docked in the NSC 12155-binding site, which may explain this compound's complete lack of LF inhibition despite its structural similarity to NSC 357756. NSC 357756, NSC 369718 and NSC 369721 did not engage in unfavorable interactions when docked in the NSC 12155-binding site, supporting this hypothesis. However, the large size and solvent-exposed nature of the LF-binding groove also allows NSC 357756, NSC 369718 and NSC 369721 to assume several different binding modes near the enzyme's active site.

The X-ray structure of the LF–NSC 12155 complex and the extensive molecular docking studies with LF inhibitors also allow for the identification of favorable structural modifications that may enhance the potency of these compounds. For example, X-ray and molecular modeling studies of NSC 12155 indicate that the 0.5-µM K_i of this inhibitor could be improved by replacing one of the quinoline moieties with a pyrrole. Such a modification would provide an additional hydrogen bond with the carboxylate of Glu687. The planar C conformation of NSC 12155 could be stabilized by replacing its amino substituents with nitro groups, thus facilitating resonance throughout this scaffold. Additionally, our study in concert with Turk *et al.*[15] suggests that replacement of one of NSC 12155's quinoline rings with a tetra-aza-benzo[a] fluorene would enhance binding by placing additional molecular volume in the S1′ site of LF. Moreover, the deep S1′ pocket (visible in **Fig. 3a**, next to zinc) seems highly selective, such that a large hydrophobic ring structure would probably increase the affinity of an inhibitor for the LF active site.

In summary, these studies describe a first critical phase in generating therapeutically viable, small molecule (nonpeptidic) countermeasures for anthrax lethal toxin. During the next phase of inhibitor optimization, information obtained from the cell-based assay will guide the incorporation of structural components that will increase inhibitor bioavailability, while at the same time allowing for optimal binding affinity in the LF substrate-binding cleft.

METHODS

Diversity set. In brief, the NCI Diversity Set is a collection of 1,990 compounds chosen (from 71,756 open compounds in the NCI chemical repository with ≥1 g inventory) to cover a large, diverse range of molecular scaffolds and pharmacophore features, while also being relatively rigid (all compounds in the Diversity Set have five or fewer rotatable bonds, facilitating pharmacophore development and conformational sampling). For a detailed description of the Diversity Set compound selection and criteria see http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html.

Fluorescent plate-based assay. For high-throughput screening in 96-well plates, the reaction volume was 100 µl per well. Master mix containing 40 mM HEPES, pH 7.2, 0.05% (v/v) Tween 20, 100 µM $CaCl_2$ and 1 µg ml^{-1} of LF was added to each well containing 100 µM of NCI Diversity Set compound. The reaction was initiated by adding the optimized peptide substrate (MCA-KKVYPYPME[dnp]K amide), to a final concentration of 20 µM. Kinetic measurements were obtained every minute for 30 min using a fluorescent plate reader (Molecular Devices, Gemini XS). Excitation and emission maxima were 324 nm and 395 nm, respectively.

Table 2 Data collection summary of LF–NSC 12155–Zn complex crystal

Resolution range (Å)	25.0–2.90
Reflections	
Total	175,849
Unique	56,384
Completeness (%)[a]	99.5 (99.3)
R_{sym} (%)[a,b]	10.6 (49.8)
$I / \sigma I$ [a]	11.7 (2.9)

[a]Values in parentheses are for the highest-resolution shell. [b]$R_{sym} = \Sigma| I - \langle I \rangle | / \Sigma \langle I \rangle$, where I is the observed intensity and $\langle I \rangle$ is the average intensity from multiple observations of symmetry-related reflections.

HPLC-based assay. An HPLC-based assay was used to validate the hits from the primary screen and eliminate the false positives obtained owing to fluorescence quenching. Reaction mix (30 µl total volume) containing 40 mM HEPES, pH 7.2, 0.05% (v/v) Tween 20, 100 µM $CaCl_2$, LF substrate (20 µM final concentration), with or without the inhibitor (20 µM final concentration), was incubated with LF (1 µg ml^{-1}) for 30 min at 30 °C. The reaction was stopped by adding 8 M guanidine hydrochloride in 0.3% (v/v) TFA. Substrate and products were separated on a Hi-Pore C18 column (Bio-Rad) using 0.1% (v/v) TFA (solvent A) and 0.1% (v/v) TFA + 70% (v/v) acetonitrile (solvent B). The column effluent was monitored at 365 nm, where the substrate and C-terminal cleavage products showed greater absorbance.

The HPLC-based assay was used for enzyme kinetic studies. Kinetic constants were obtained from plots of initial rates with seven concentrations of the substrate. For the best inhibitors, K_i and the type of inhibition were evaluated using seven different concentrations of the substrate ranging from 2 to 40 µM and four different concentrations of the inhibitor. K_i values for the competitive inhibitors were calculated using the equation $K_i = [I] / [(K_{m(app)} / K_m) - 1]$, where [I] is the inhibitor concentration[17]. K_i values in **Table 1** are the averages ± s.d.

LF refinement and inhibitor docking. The structure of LF was energy-refined using the Discover (Accelrys) program's cff91 force field. Our strategy entailed using a step-down, template forced minimization procedure with the Zn coordination site fixed. This process was repeated until coordinates of the final model were within the experimentally determined X-ray crystallographic resolution. The inhibitor–enzyme structure coordinates were subsequently tether-minimized in the same manner as described above, and the final structure was subjected to hydropathic analysis using HINT (eduSoft).

Conformer generation. Conformational models of inhibitors were generated using Catalyst 4.7 (Accelrys). A 'best-quality' conformational search was used to generate conformers within 20 kcal mol^{-1} of the global energy minimum.

Data mining. Catalyst 4.7 (Accelrys) was used for all database mining. Briefly, the imidazole rings of NSC 357756 were used to form a three-dimensional search query (A.R.H. *et al.*, unpublished data). Subsequent molecular docking studies (see above) were used to suggest candidates for biological testing.

Quantum mechanical calculations. The conformations (L and C shaped) of NSC 12155 were fully optimized (until the norm of the gradient was $<5.0 \times 10^4$) using DGauss (Oxford Molecular Group). Local spin density (LSD) correlation potentials were approximated by the Vosko-Wilk-Nusair method[18] and gaussian analytical functions were used as basis sets. LSD-optimized orbital basis sets of double ζ-split valence polarization quality[19] were used. In final optimizations, the BLYP exchange-correlation functional[20,21] was applied as a nonlocal gradient correction after each self-consistent field cycle.

Crystallization. Native, wild-type LF protein was crystallized using 13 mg ml^{-1} LF. Crystals were grown from 1.7 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 7.5–8.0, 2 mM EDTA, using hanging-drop vapor diffusion[16]. Monoclinic crystals appeared after four days to two weeks, and were then harvested for experiments. The LF crystals belong to the monoclinic space group $P2_1$, with unit cell dimensions a = 96.70 Å, b = 137.40 Å, c = 98.30 Å, $\alpha = \gamma = 90°$, $\beta = 98°$, containing two molecules per asymmetric unit.

LF–inhibitor complexes. LF native crystals were harvested from the hanging drops in which they were grown, bathed in several rounds of fresh buffer without EDTA containing 1.9 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0, and left to soak in this solution for a further 30 min. These crystals were then used to obtain the protein–inhibitor–zinc complexes. All manipulations were done at room temperature (23–26 °C).

The LF–NSC 12155–Zn complex was obtained by soaking an individual native LF monoclinic $P2_1$ crystal in a solution of 1 mM $ZnSO_4$, 1.9 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0 for 5 min. The crystal was then transferred to a solution of 1.0 mM NSC 12155, 1% (v/v) DMSO, 1.9 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0 for 15 min. Finally, the crystal was transferred into a cryoprotectant solution of 1.0 mM NSC 12155, 2.4 M $(NH_4)_2SO_4$, 0.2 M Tris-HCl, pH 8.0, 2 mM EDTA, 25% (v/v) glycerol, and soaked at room temperature for 1 min. The crystal was then immediately mounted onto a cryoloop and flash-frozen in liquid nitrogen. All data were collected at 100 K.

Data collection. Datasets for the LF complexes were collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Menlo Park, California, USA) on beamline 9-1 (wavelength = 0.983 Å). X-ray diffraction data were collected for the LF–NSC 12155–Zn complex to a resolution limit of 2.90 Å. Data collection statistics are shown in **Table 2**.

Structure solution and refinement. Collected data were processed in the HKL package[22]. Refinement and model building were done in CNS[23] and O[24], respectively. Using PDB entry 1J7N as the starting model, the model of LF alone was put through rigid body refinement and then minimization before the first initial maps were calculated for model building and further refinement. Excess electron density at 1.0 σ indicated the binding location of the inhibitor in the active site of LF. The model of the inhibitor was then built into this position and further refined in CNS[23]. The final R-factors were R_{free} = 27.58% and R_{work} = 22.38%. The final model falls within or exceeds the limits of all the quality criteria of PROCHECK from the CCP4 suite[25].

Cytotoxicity assay. J774A.1 cells were preincubated with DMSO control or compounds for 30 min and then treated with PA (50 ng ml^{-1}) and LF (14 ng ml^{-1}). After 4 h incubation with the toxin, 25 µl of MTT (1 mg ml^{-1}) dye was added and the cells were further incubated for 2 h. The reaction was stopped by adding an equal volume of lysis buffer (20% (v/v) DMF and 20% (w/v) SDS, pH 4.7). Plates were incubated overnight at 37 °C and absorbance was read at 570 nm in a multiwell plate reader. Experiments were done in duplicate and repeated three independent times for each of the inhibitors tested. The results are the averages ± s.d.

Coordinates. The coordinates and structure factors for the LF–NSC 12155–Zn complex have been deposited in the Protein Data Bank (accession code 1PWP).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS
This research was sponsored by the Medical Biological Defense Research Program, US Army Medical Research and Materiel Command, project no. 0242C012. We acknowledge the US National Cancer Institute for the allocation of computing time and staff support at the Advanced Biomedical Computing Center of the Frederick Cancer Research and Development Center. We thank the staff of the Stanford Synchrotron Radiation Laboratory (SSRL) for assistance during data collection. We also thank S. Leppla (US National Institutes of Health (NIH)) and D. Hsu for the LF protein preparation used in the crystal structure. Portions of this research were carried out at the SSRL, a national user facility operated by Stanford University on behalf of the US Department of Energy (DOE), Office of Basic Energy Sciences (BES). The SSRL Structural Molecular Biology Program is supported by the DOE, Office of Biological and Environmental Research, and by the NIH, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 26 August; accepted 30 October 2003

1. Bradley, K.A., Mogridge, J., Mourez, M., Collier, R.J. & Young, J.A. Identification of the cellular receptor for anthrax toxin. *Nature* **414**, 225–229 (2001).
2. Scobie, H.M., Rainey, G.J., Bradley, K.A. & Young, J.A. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *Proc. Natl. Acad. Sci. USA* **100**, 5170–5174 (2003).
3. Klimpel, K.R., Molloy, S.S., Thomas, G. & Leppla, S.H. Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. *Proc. Natl. Acad. Sci. USA* **89**, 10277–10281 (1992).
4. Molloy, S.S., Bresnahan, P.A., Leppla, S.H., Klimpel, K.R. & Thomas, G. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. *J. Biol. Chem.* **267**, 16396–16402 (1992).
5. Leppla, S.H. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. *Proc. Natl. Acad. Sci. USA* **79**, 3162–3166 (1982).
6. Vitale, G. *et al.* Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/threonine phosphorylation of MAPKs in cultured macrophages. *Biochem. Biophys. Res. Commun.* **248**, 706–711 (1998).
7. Duesbery, N.S. *et al.* Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. *Science* **280**, 734–737 (1998).
8. Duesbery, N.S. *et al.* Suppression of ras-mediated transformation and inhibition of tumor growth and angiogenesis by anthrax lethal factor, a proteolytic inhibitor of multiple MEK pathways. *Proc. Natl. Acad. Sci. USA* **98**, 4089–4094 (2001).
9. Park, J.M., Greten, F.R., Li, Z.W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048–2051 (2002).
10. Sellman, B.R., Mourez, M. & Collier, R.J. Dominant-negative mutants of a toxin subunit: an approach to therapy of anthrax. *Science* **292**, 695–697 (2001).
11. Singh, Y., Khanna, H., Chopra, A.P. & Mehra, V. A dominant negative mutant of *Bacillus anthracis* protective antigen inhibits anthrax toxin action *in vivo*. *J. Biol. Chem.* **276**, 22090–22094 (2001).
12. Mourez, M. *et al.* Designing a polyvalent inhibitor of anthrax toxin. *Nat. Biotechnol.* **19**, 958–961 (2001).
13. Hammond, S.E. & Hanna, P.C. Lethal factor active-site mutations affect catalytic activity *in vitro*. *Infect. Immun.* **66**, 2374–2378 (1998).
14. Tonello, F., Seveso, M., Marin, O., Mock, M. & Montecucco, C. Screening inhibitors of anthrax lethal factor. *Nature* **418**, 386 (2002).
15. Turk, B.E. *et al.* The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. *Nat. Struct. Mol. Biol.* **11**, 60–66 (2004).
16. Pannifer, A.D. *et al.* Crystal structure of the anthrax lethal factor. *Nature* **414**, 229–233 (2001).
17. Segel, I.H. *Enzyme Kinetics* (Wiley, New York, 1975).
18. Vosko, S.J., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations. *Can. J. Phys.* **58**, 1200–1211 (1980).
19. Godbout, N., Salahub, D., Andzelm, J. & Wimmer, E. Optimization of Gaussian-type basis-sets for local spin-density functional calculations. 1. Boron through neon, optimization technique and validation. *Can. J. Chem.* **70**, 560–571 (1992).
20. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
21. Lee, C., Yang, W. & Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **37**, 785–789 (1988).
22. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
23. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
24. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
25. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D.* **50**, 760–763 (1994).

Crystal structure of a complex between anthrax toxin and its host cell receptor

Eugenio Santelli[1], **Laurie A. Bankston**[1], **Stephen H. Leppla**[2] **& Robert C. Liddington**[1]

[1] *Program on Cell Adhesion, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA*
[2] *Microbial Pathogenesis Section, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 20892, USA*

Anthrax toxin consists of the proteins protective antigen (PA), lethal factor (LF) and oedema factor (EF)[1]. The first step of toxin entry into host cells is the recognition by PA of a receptor on the surface of the target cell. Subsequent cleavage of receptor-bound PA enables EF and LF to bind and form a heptameric PA$_{63}$ prepore, which triggers endocytosis. Upon acidification of the endosome, PA$_{63}$ forms a pore that inserts into the membrane and translocates EF and LF into the cytosol[2]. Two closely related host cell receptors, TEM8 and CMG2, have been identified. Both bind to PA with high affinity and are capable of mediating toxicity[3,4]. Here, we report the crystal structure of the PA–CMG2 complex at 2.5 Å resolution. The structure reveals an extensive receptor–pathogen interaction surface mimicking the non-pathogenic recognition of the extracellular matrix by integrins[5]. The binding surface is closely conserved in the two receptors and across species, but is quite different in the integrin domains, explaining the specificity of the interaction. CMG2 engages two domains of PA, and modelling of the receptor-bound PA$_{63}$ heptamer[6–8] suggests that the receptor acts as a pH-sensitive brace to ensure accurate and timely membrane insertion. The structure provides new leads for the discovery of anthrax antitoxins, and should aid the design of cancer therapeutics[9].

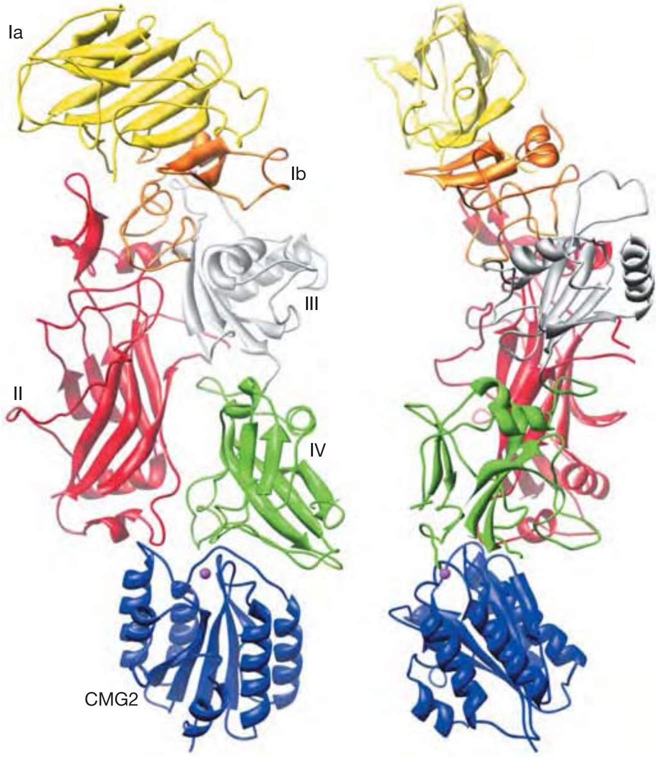


Figure 1 Structure of the PA–CMG2 complex. Two orthogonal views are shown in ribbon representation. PA is coloured by domain (I–IV). CMG2 is blue; the metal ion is shown as a magenta ball. PA domain I is cleaved after receptor binding, leading to the loss of domain Ia (yellow) and the formation of PA$_{63}$. All molecular graphics images were generated using the UCSF Chimera package[29] (http://www.cgl.ucsf.edu/chimera).

Both TEM8 and CMG2 contain a domain that is homologous to the I domains of integrins, which comprise a Rossmann-like α/β-fold with a metal-ion-dependent adhesion site (MIDAS) motif on their upper surface[10]. Crystal structures of the CMG2 I domain and full-length PA proteins have previously been determined[6,11]. The PA monomer is a long slender molecule comprising four distinct domains. In the PA–CMG2 I domain complex, two of these four domains (II and IV) pack together at the base of PA and engage the upper surface of the CMG2 I domain surrounding the MIDAS motif (Fig. 1), burying a large protein surface (1,900 Å^2), consistent with the very high affinity (sub-nanomolar dissociation constant) of this interaction[12]. The I domain adopts the 'open' conformation, typical of integrin–ligand complexes[5,13]. PA mimics the ligand recognition mechanism of the integrins[5] by contributing an aspartic acid side chain that completes the coordination sphere of the MIDAS magnesium ion, as predicted by mutagenesis[14,15] (Fig. 2a, b). This single interaction contributes substantially to binding, as mutation of the aspartic acid to asparagine completely eliminates toxicity, as does mutation of a metal-coordinating residue on the receptor.

However, the MIDAS bond does not fully explain the specificity of the interaction, as it does not distinguish between CMG2 and integrins. Further specificity arises from two additional interactions. First, PA domain IV docks onto the surface of CMG2 adjacent to the MIDAS motif. Domain IV comprises a β-sandwich with an immunoglobulin-like fold, but the mode of binding is quite different from that of antibody–antigen recognition. One of the receptor loops (α2–α3) emanating from the MIDAS motif forms a hydrophobic ridge that inserts into a groove formed by one edge of the β-sandwich where its hydrophobic core is exposed. Flanking this ridge-in-groove arrangement are two further loops from CMG2, which make a number of specific polar interactions and salt bridges (Figs 3 and 4a). Together with the MIDAS contact, CMG2 and PA domain IV bury 1,300 Å^2 of surface area, a value very similar to two integrin–ligand interactions that have affinities in the sub-micromolar range[5,13]. CMG2 and TEM8 share 60% identity in their I

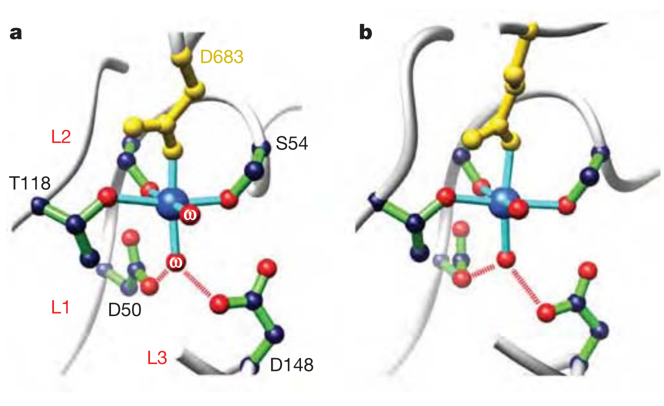


Figure 2 The MIDAS motifs of the PA–CMG2 complex (**a**) and the collagen–integrin α2β1 complex[5] (**b**). Coordinating side chains and two water molecules (ω) are shown in ball-and-stick representation. The metal is shown in blue. D683 from PA, and a collagen glutamic acid, are in gold. Bond distances to the metal are 2.1 ± 0.2 Å in both cases. The three MIDAS loops (L1–L3) are labelled in **a**.

domains, and homology modelling based on the CMG2 structure shows that this ridge is well conserved in TEM8 and their murine counterparts, implying that they will bind PA in a similar fashion; however, the structure and sequence of the ridge are very different in integrins, explaining their weak binding (Fig. 4b).

The interaction between PA domain II and CMG2 was not anticipated. A β-hairpin from a well-ordered loop (β3–β4) at the bottom of domain II inserts into a pocket on the receptor, burying 600 Å^2 of protein surface (Fig. 4b, c). This additional contact may explain the very high affinity of the PA–CMG2 interaction. The pocket is adjacent to the MIDAS motif and is formed by two exposed tyrosine residues (Y119 and Y158) and the β4–α4 loop, which line the sides of the pocket, and by a histidine (H121) at its base. The pocket is conserved in TEM8, but does not exist in the I domains of integrins, thus providing further specificity (Fig. 4b, c). The importance of this loop was shown by systematic mutation of the PA molecule, which revealed three mutations in this loop that reduced toxicity by >100-fold, including G342 at the tip of the β-hairpin that inserts into the pocket[16].

Biophysical studies of channel conductance by PA$_{63}$ pores indicate that the entire region encompassed by residues 275–352 (strands β2 and β3 and flanking loops; see Fig. 3) in domain II rearranges to form a long β-hairpin that lines the channel lumen[7,8]. This requires that the β2 and β3 strands and the β3–β4 loop peel away from the side of domain II. For this to happen, domain IV, which packs against them in the pre-pore, must separate at least transiently from domain II. Thus, by binding to both domains II and IV, CMG2 may restrain the conformational changes that lead to membrane insertion. Indeed, whereas PA$_{63}$ heptamers insert into artificial planar bilayers (in the absence of receptor) when the pH is reduced to 6.5, the pH requirement for receptor-mediated insertion on cells is more stringent, requiring a pH of 5.5 (ref. 17). Thus, we propose that the binding of CMG2 to the β3–β4 loop stabilizes the pre-pore conformation at neutral pH; that is, the receptor may act as a brace to prevent premature membrane insertion on the cell surface before endocytosis. The pH profile of membrane insertion is consistent with the titration of histidine residues, and seven of the nine histidines within PA$_{63}$ cluster at the domain II–IV interface (Fig. 3). In addition, the histidine at the base of the CMG2 pocket

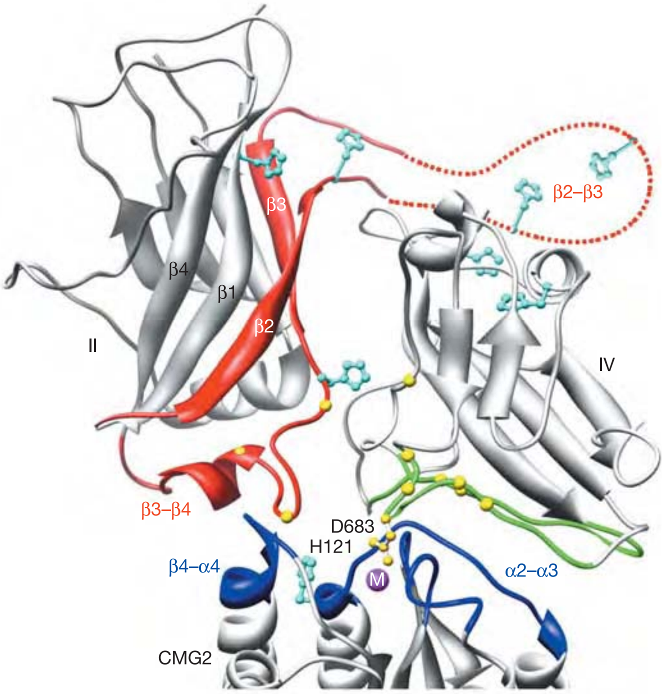


Figure 3 Intermolecular contacts between PA domains II and IV and CMG2. Contacting regions are coloured blue and green for CMG2 and PA domain IV, respectively. The β2–β3 loop and flanking regions of PA domain II, which are implicated in pore formation, are highlighted in red. The β2–β3 loop is disordered in monomeric PA and is shown schematically as a dashed line. The histidine residues within PA domains II and IV and within the CMG2 I domain are shown coloured cyan and are in ball-and-stick representation. Mutation sites that reduce binding by >100-fold (D683, S337, G342, W346, I656, N657, I665, Y681, N682, P686, L687) are highlighted in gold.

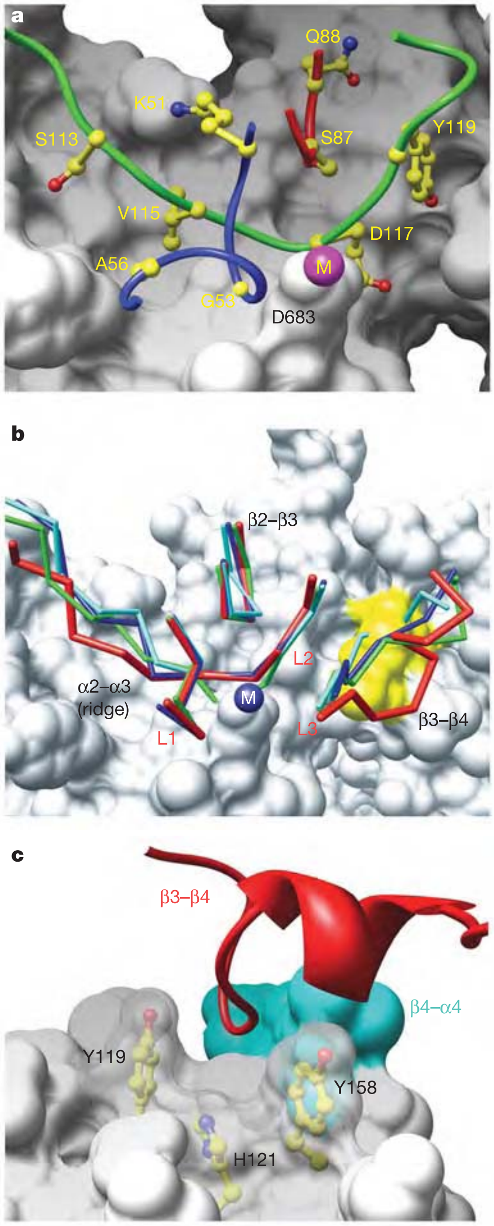


Figure 4 Key elements of the PA–CMG2 interaction **a**, Solvent-accessible surface of the PA domain IV groove, with key side chains from three CMG2 loops (β1–α1, blue; β2–β3, red; α2–α3, green) shown in ball-and-stick representation. The α2–α3 loop forms the ridge. The MIDAS metal is labelled (M). **b**, Comparison with integrin I domains in the 'open' conformation (CMG2, red; αM, cyan; α2, green; αL, blue) overlaid on the MIDAS motif. **c**, Surface of the CMG2 pocket into which the PA β3–β4 loop (red ribbon) inserts, formed by three CMG2 side chains (shown in ball-and-stick representation) and the β4–α4 loop (cyan).

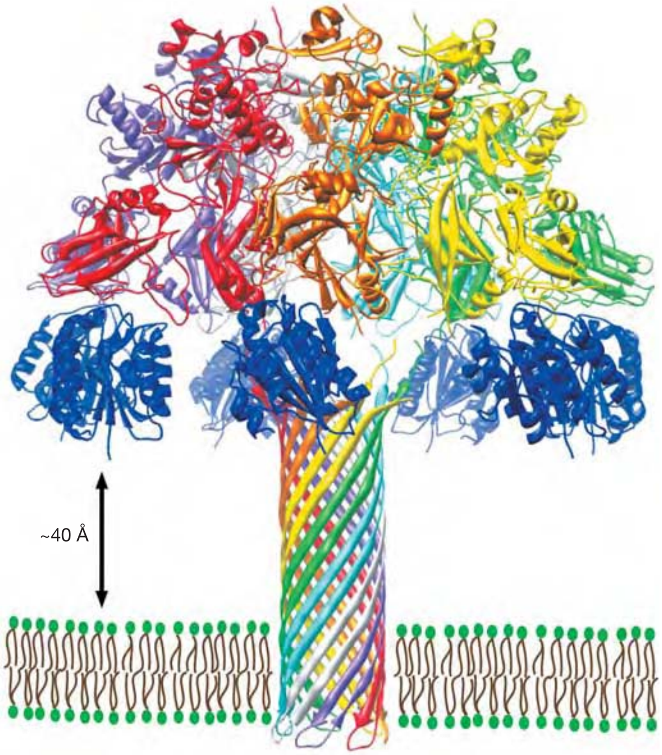


Figure 5 Hypothetical model of the receptor-bound, membrane-inserted PA pore. The model is based on the pre-pore PA_{63} crystal structure[6], channel conductance studies[8], and the crystal structure of α-haemolysin[19]. The barrel is formed by rearrangement in each monomer of the segment shown in red in Fig. 3. Each PA_{63} monomer is shown in a different colour. Residues 303–324 form the membrane-spanning region of the barrel. Seven copies of the CMG2 I domain bound to the heptamer are in blue. The ~40 Å gap between the CMG2 I domain and the membrane may be occupied by a ~100-residue domain of CMG2, C-terminal to the I domain, which precedes its membrane-spanning sequence.

(conserved in TEM8) has no H-bonding partners, and is close to an arginine side chain from the β3–β4 loop of PA. Histidine protonation provides a plausible trigger for the release of domain II from CMG2 in the acidified endosome. Indeed, we have shown that the structure of the β3–β4 loop is pH-sensitive, as it becomes disordered when crystals of PA grown at pH 7.5 (in the absence of receptor) are reduced to pH 6.0 (ref. 18).

It is straightforward to model the 7:7 heptameric PA_{63}–CMG2 complex, as the crystal structure of the pre-pore is known[6] (Fig. 5). Seven CMG2 I domains lie at the base of the heptameric ‘cap’, increasing its height by 35 Å. The I domains are well separated, consistent with a 7:7 binding stoichiometry[12], and their amino- and carboxy termini point downwards, towards the membrane. In the transition from pre-pore to pore, the seven hairpin loops, one from each PA monomer[6,8], are predicted to create a 14-stranded, membrane-spanning β-barrel. Assuming an α-haemolysin-like structure[19], the barrel extends ~75 Å below the I domains, with the bottom 30 Å spanning the membrane. This leaves ~40 Å between the bottom of the I domains and the membrane surface, which may be occupied by the second domain of CMG2, which comprises ~100 residues between the I domain and its C-terminal transmembrane sequence. Thus, the receptor may support the heptamer at the correct height above the membrane for accurate membrane insertion, which is stoichiometric on cells but less efficient in the absence of receptor[17].

Table 1 **Data collection and refinement statistics**

Parameter	Value	
Space group	$P2_12_12_1$	
Unit cell (Å)	$a = 88.2, b = 94.1, c = 135.6$	
Resolution (Å)	30–2.5	
Wavelength (Å)	0.892	
R_{merge} (%)	17.6 (99.1)	
I/σ	11.5 (2.4)	
σ-cutoff	None	
Average redundancy	5.3 (5.2)	
Completeness (%)	99.9	
Mosaicity	0.4	
R_{work} (last shell)	20.7 (27.5)	
R_{free} (last shell)	26.6 (37.2)	
σ -cutoff	None	
B factors ($Å^2$)*	32.9, 21.4, 23.3	
r.m.s.d. bond lengths (Å)	0.17	
r.m.s.d. bond angles (°)	1.65	
Ramachandran plot (residues, %)		
Most favoured	655	86.3%
Additionally allowed	101	13.3%
Generously allowed	3	0.4%
Disallowed	0	0%

Values in parentheses refer to the highest resolution shell (2.59–2.50 Å).
*The three values are for Wilson, main chain and side chain, respectively.

Soluble versions of the CMG2 and TEM8 I domains protect against anthrax (*Bacillus anthracis*) toxin by acting as decoys[3,15], and our structure will allow for the design of new therapeutic agents that disrupt the PA–receptor interaction. TEM8 is strongly upregulated on the surface of endothelial cells that line the blood vessels of tumours[20,21], allowing for the development of anthrax toxin as an anti-tumour agent[22]; however, toxicity may arise as CMG2 is expressed in most tissues. Although we expect the interactions of TEM8 and CMG2 with PA to be very similar, there are significant differences that may be exploited in the design of PA molecules that would bind better to TEM8 than to CMG2, thus minimizing the side effects from toxin binding to normal tissues. For example, V115 of CMG2, which lies at the heart of the interface with PA domain IV, is a glycine in TEM8, whereas the rim of the pocket that accepts the PA domain II loop has the sequence DGL in CMG2 but is replaced by the sequence HED in TEM8. ☐

Methods

Protein expression and purification

Full-length PA (residues 1–735) was prepared as previously described[14]. The I domain of human CMG2 was cloned as an N-terminal His-tag fusion in pET15b (Novagen) and expressed in *Escherichia coli* strain BL21(DE3). After induction of cell cultures with 0.5 mM IPTG for 2 h at 37 °C, CMG2 was purified from the soluble fraction of the cell lysate by nickel affinity chromatography (HiTrap chelating HP, Pharmacia), followed by removal of the tag with thrombin (Sigma), ion exchange (HiTrap monoQ, Pharmacia) and gel filtration (Superdex S75, Pharmacia), affinity removal of thrombin (HiTrap benzamidine FF, Pharmacia) and incubation in a buffer containing 100 mM EDTA to strip-bound metal. The final product was dialysed and concentrated to 15–20 mg ml^{-1} and flash-frozen in 150 mM NaCl, 20 mM TrisCl pH 7.5, and comprises residues 40–218 of $CMG2^{386}$ (GenBank accession number AAK77222) plus an N-terminal extension of sequence GSHMLEDPRG as a result of the cloning strategy. The molecular mass was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. To prepare the PA–CMG2 complex, PA was mixed at a final concentration of 4 mg ml^{-1} with a threefold molar excess of CMG2 and a twofold excess of $MnCl_2$, incubated for 20 min at room temperature and purified by gel filtration (Superdex S200, Pharmacia). The complex was extensively dialysed and exchanged, and concentrated to 6 mg ml^{-1} in 20 mM TrisCl pH 7.5, 10 μM $MnCl_2$ for crystallization trials.

Crystallization and structure solution

Needle-like crystals grew to a size of 10 × 10 × 500 μm in 5–10 days at room temperature in a sitting-drop vapour diffusion set-up using a reservoir buffer containing 50–100 mM CHES pH 9.0–9.2, 25% PEG400. Crystals were flash-frozen at 4 °C in liquid nitrogen using the crystallization buffer with 40% PEG400 as a cryo-protectant before diffraction analysis. The crystals belong to space group $P2_12_12_1$ with unit cell parameters $a = 88.2$ Å, $b = 94.2$ Å, $c = 135.6$ Å. There is one PA–CMG2 complex in the asymmetric unit. A complete native data set to 2.5 Å was collected at beamline 9-1 at SSRL on a ADSC Quantum-315 CCD detector and processed with the HKL package[23] (see Table 1). PA was positioned in the unit cell by Molecular Replacement (Protein Data Bank (PDB) ID code 1acc)[6] using MOLREP, and refined with REFMAC version 5.0 (ref. 24). Density for the MIDAS Mn^{2+} ion and upper loops of the receptor was evident in this map, and a molecule of CMG2 (PDB ID code 1SHT)[11] was manually placed in the electron density. Model building was performed with O[25] and TURBOFRODO (A. Roussel and C. Cambillau, Silicon Graphics), and the solvent structure was built with ARP/wARP 6.0 (ref. 26). Although the random errors in the diffraction data are high, owing to the small crystal size, the final refinement statistics and maps are excellent (Table 1). Thus, the final *R*-factors are $R_{\text{free}} = 26.6\%$ and $R_{\text{work}} = 20.7\%$ overall, and $R_{\text{free}} = 37.2\%$ and $R_{\text{work}} = 27.5\%$ in the outer resolution bin, with root-mean-square deviations (r.m.s.d.) from ideal values of 0.017 Å for bond lengths and 1.65° for angles. Stereochemistry is excellent as assessed with PROCHECK[24], and the model is consistent with composite simulated annealing omit maps (3,000 °C) calculated in CNS[27]. The model comprises residues 16–735 of PA; 41–210 of CMG2, with the exception of three loops (residues 159–174, 276–287 and 304–319) in PA for which no electron density was observed; 139 water molecules; two Ca^{2+} ions in PA domain I; two Na^{+} ions; one PEG molecule; and one Mn^{2+} ion at the MIDAS site. The *B* factors for the Ca^{2+} and Mn^{2+} ions (27–33 $Å^2$) are higher than for the coordinating residues (16–20 $Å^2$). Although the MIDAS metal ion *in vivo* is likely to be Mg^{2+}, we have previously shown for integrin I domains that the stereochemistry of the open conformation is not dependent on the nature of the metal ion[5]. The bond lengths to the Mn^{2+} ion are 2.1 ± 0.2 Å, identical to those observed in integrin–ligand complexes[5,13,28]. PA domain I (residues 16–258) undergoes a small rotation as a consequence of crystal constraints when compared with the structure of isolated PA such that the r.m.s.d. values for the superposition of the two molecules are 1.44, 0.58 and 0.79 Å for residues 16–735, 259–735 and 16–258 respectively. CMG2 residues 41–200 superimpose with a r.m.s.d. of 0.60 with the isolated protein[11], while the C-terminal helix (residues 201–210) shifts downwards by one helical turn.

Received 6 May; accepted 18 June 2004; doi:10.1038/nature02763.

1. Moayeri, M. & Leppla, S. H. The roles of anthrax toxin in pathogenesis. *Curr. Opin. Microbiol.* **7,** 19–24 (2004).
2. Abrami, L., Liu, S., Cosson, P., Leppla, S. H. & Vander Goot, F. G. Anthrax toxin triggers endocytosis of its receptor via a lipid raft-mediated clathrin-dependent process. *J. Cell Biol.* **160,** 321–328 (2003).
3. Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414,** 225–229 (2001).
4. Scobie, H. M., Rainey, G. J., Bradley, K. A. & Young, J. A. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *Proc. Natl Acad. Sci. USA* **100,** 5170–5174 (2003).
5. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. & Liddington, R. C. Structural basis of collagen recognition by integrin α2β1. *Cell* **101,** 47–56 (2000).
6. Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **385,** 833–838 (1997).
7. Benson, E. L., Huynh, P. D., Finkelstein, A. & Collier, R. J. Identification of residues lining the anthrax protective antigen channel. *Biochemistry* **37,** 3941–3948 (1998).
8. Nassi, S., Collier, R. J. & Finkelstein, A. PA63 channel of anthrax toxin: an extended β-barrel. *Biochemistry* **41,** 1445–1450 (2002).
9. Frankel, A. E., Koo, H.-K., Leppla, S. H., Duesbury, N. S. & Vande Woude, G. F. Novel protein targeted therapy of metastatic melanoma. *Curr. Pharm. Des.* **9,** 2060–2066 (2003).
10. Lee, J.-O., Rieu, P., Arnaout, M. A. & Liddington, R. C. Crystal structure of the A-domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell* **80,** 631–635 (1995).
11. Lacy, D. B., Wigelsworth, D. J., Scobie, H. M., Young, J. A. & Collier, R. J. Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: An anthrax toxin receptor. *Proc. Natl Acad. Sci. USA* **101,** 6367–6372 (2004).
12. Wigelsworth, D. J. *et al.* Binding stoichiometry and kinetics of the interaction of a human anthrax toxin receptor, CMG2, with protective antigen. *J. Biol. Chem.* **279,** 23349–23356 (2004).
13. Shimaoka, M. *et al.* Structures of the αL I domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell* **112,** 99–111 (2003).
14. Rosovitz, M. J. *et al.* Alanine scanning mutations in domain 4 of anthrax toxin protective antigen reveal residues important for binding to the cellular receptor and to a neutralizing monoclonal antibody. *J. Biol. Chem.* **278,** 30936–30944 (2003).
15. Bradley, K. A. *et al.* Binding of anthrax toxin to its receptor is similar to α integrin-ligand interactions. *J. Biol. Chem.* **278,** 49342–49347 (2003).
16. Mourez, M. *et al.* Mapping dominant-negative mutations of anthrax protective antigen by scanning mutagenesis. *Proc. Natl Acad. Sci. USA* **100,** 13803–13808 (2003).
17. Miller, C. J., Elliott, J. L. & Collier, R. J. Anthrax protective antigen: prepore-to-pore conversion. *Biochemistry* **38,** 10432–10441 (1999).
18. Petosa, C. in *Crystal Structure of the Anthrax Protective Antigen*. Thesis, Harvard Univ (1995).
19. Song, L. *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274,** 1859–1866 (1996).
20. Nanda, A. & St Croix, B. Tumor endothelial markers: new targets for cancer therapy. *Curr. Opin. Oncol.* **16,** 44–49 (2004).
21. Nanda, A. *et al.* TEM8 interacts with the cleaved C5 domain of collagen alpha 3(VI). *Cancer Res.* **64,** 817–820 (2004).
22. Liu, S., Schubert, R. L., Bugge, T. H. & Leppla, S. H. Anthrax toxin: structures, functions and tumour targeting. *Expert Opin. Biol. Ther.* **3,** 843–853 (2003).
23. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
24. Collaborative Computational Project, No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
25. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjelgaard, M. Improved methods for building protein models into electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
26. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374,** 229–244 (2003).
27. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
28. Lee, J.-O., Bankston, L. A., Arnaout, M. A. & Liddington, R. C. Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure* **3,** 1333–1340 (1995).
29. Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38,** 305–320 (1996).

Acknowledgements We thank the NIH and the DOD for financial support, and the DOE and staff at the SSRL for synchrotron access and support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.C.L. (rlidding@burnham.org). The atomic coordinates have been deposited in the Protein Data Bank under accession code 1T6B.

Cell cycle regulation of central spindle assembly

Masanori Mishima[1], **Visnja Pavicic**[1], **Ulrike Grüneberg**[2], **Erich A. Nigg**[2] **& Michael Glotzer**[1]

[1] *Research Institute of Molecular Pathology, Dr. Bohrgasse 7, A-1030 Vienna, Austria*
[2] *Max-Planck-Institute für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany*

The bipolar mitotic spindle is responsible for segregating sister chromatids at anaphase. Microtubule motor proteins generate spindle bipolarity and enable the spindle to perform mechanical work[1]. A major change in spindle architecture occurs at anaphase onset when central spindle assembly begins. This structure regulates the initiation of cytokinesis and is essential for its completion[2]. Central spindle assembly requires the centralspindlin complex composed of the *Caenorhabditis elegans* ZEN-4 (mammalian orthologue MKLP1) kinesin-like protein and the Rho family GAP CYK-4 (MgcRacGAP). Here we describe a regulatory mechanism that controls the timing of central spindle assembly. The mitotic kinase Cdk1/cyclin B phosphorylates the motor domain of ZEN-4 on a conserved site within a basic amino-terminal extension characteristic of the MKLP1 subfamily. Phosphorylation by Cdk1 diminishes the motor activity of ZEN-4 by reducing its affinity for microtubules. Preventing Cdk1 phosphorylation of ZEN-4/MKLP1 causes enhanced metaphase spindle localization and defects in chromosome segregation. Thus, phosphoregulation of the motor domain of MKLP1 kinesin ensures that central spindle assembly occurs at the appropriate time in the cell cycle and maintains genomic stability.

At the metaphase–anaphase transition, the anaphase-promoting complex triggers proteolysis of cyclin B (an activating subunit of the mitotic kinase Cdk1) and sister chromatid separation. Chromosomes move polewards and non-kinetochore spindle microtubules become bundled, initiating assembly of the central spindle, a structure that has important roles in cytokinesis. In *C. elegans* embryos and other animal cells, central spindle assembly requires centralspindlin[3]. Many proteins that regulate mitosis and cytokinesis re-localize upon anaphase onset. For example, Aurora B and its associated subunits dissociate from centromeres and concentrate on the central spindle[4–6]. Similarly, anaphase onset triggers redistribution of centralspindlin (Fig. 1a, b). In metaphase, centralspindlin is diffuse and in anaphase it localizes to the microtubules positioned between the separating chromosomes, as seen previously[7–10]. ZEN-4 (also known as CeMKLP1) colocalizes with the proline-directed phosphatase CDC-14 (ref. 11) and depletion of CDC-14 prevents ZEN-4 localization[12]. Non-degradable cyclins stabilize Cdk1 activity and prevent central spindle assembly[13,14]. Together these data

JOURNAL OF BACTERIOLOGY, Sept. 2006, p. 6354–6360 Vol. 188, No. 17
0021-9193/06/$08.00+0 doi:10.1128/JB.00656-06

Sensor Domains Encoded in *Bacillus anthracis* Virulence Plasmids Prevent Sporulation by Hijacking a Sporulation Sensor Histidine Kinase

Andrea K. White,[1] James A. Hoch,[1] Marcin Grynberg,[2]† Adam Godzik,[2] and Marta Perego[1]*

Division of Cellular Biology, Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037,[1] and Infectious and Inflammatory Disease Research (IIDC), The Burnham Institute, La Jolla, California 92037[2]

Received 8 May 2006/Accepted 13 June 2006

Anthrax toxin and capsule, determinants for successful infection by *Bacillus anthracis*, are encoded on the virulence plasmids pXO1 and pXO2, respectively. Each of these plasmids also encodes proteins that are highly homologous to the signal sensor domain of a chromosomally encoded major sporulation sensor histidine kinase (BA2291) in this organism. *B. anthracis* Sterne overexpressing the plasmid pXO2-61-encoded signal sensor domain exhibited a significant decrease in sporulation that was suppressed by the deletion of the BA2291 gene. Expression of the sensor domains from the pXO1-118 and pXO2-61 genes in *Bacillus subtilis* strains carrying the *B. anthracis* sporulation sensor kinase BA2291 gene resulted in BA2291-dependent inhibition of sporulation. These results indicate that sporulation sensor kinase BA2291 is converted from an activator to an inhibitor of sporulation in its native host by the virulence plasmid-encoded signal sensor domains. We speculate that activation of these signal sensor domains contributes to the initiation of *B. anthracis* sporulation in the bloodstream of its infected host, a salient characteristic in the virulence of this organism, and provides an additional role for the virulence plasmids in anthrax pathogenesis.

The etiological agent of anthrax, *Bacillus anthracis*, is a uniquely pervasive and persistent environmental pathogen due to its ability to form dormant spores that are resistant to adverse environmental conditions such as extremes of temperature, UV radiation, and antimicrobial chemical agents (9, 22). The spore is essential to the organism not only for its persistence in the environment but also for the ability of this organism to infect its hosts. Infection is initiated when spores are introduced into the host body and phagocytosed by macrophages, or perhaps other phagocytic cells (5, 10, 11). It is believed that this is followed by germination of the spores into vegetative cells, with subsequent toxin gene expression and capsule production, resulting in the onset of anthrax disease (11).

Interestingly, while the spore is required to initiate the infection, once vegetative growth is established, sporulation does not occur in the bloodstream of the infected host (17). This might be explained by the observation that macrophages can take up spores and destroy them as soon as they start to germinate, while encapsulated vegetative cells are able to evade the immune system (14, 16). Thus, while the transition to and maintenance of vegetative growth, which accompanies toxin and capsule production and progression of the disease, are advantageous to the pathogenic lifestyle of *B. anthracis*, sporulation within the host may not be.

The observation that sporulation and progression of the anthrax disease are potentially mutually exclusive events requires that regulatory networks must exist to ensure that while one is occurring, the other does not. The major deciding factor in orchestrating which of these events occurs is the level of phosphorylated Spo0A (Spo0A~P) response regulator-transcription factor. Spo0A is the phosphorylation target of the *Bacillus* species' phosphorelay signal transduction system that controls sporulation initiation (7). In addition to its role in upregulating the expression of genes required to initiate sporulation, in *B. anthracis*, phosphorylated Spo0A indirectly regulates expression of the anthrax toxin genes *pagA* (protective antigen), *cya* (edema factor), and *lef* (lethal factor) via its negative regulation of the transition state regulator AbrB (3, 23). Thus, while some low level of Spo0A~P is required for repression of AbrB and maximal anthrax toxin production, too much Spo0A~P would result in the onset of sporulation, which has been speculated to be antithetical to successful pathogenesis (6). The regulatory mechanism(s) that results in the appropriate levels of Spo0A~P formation in *B. anthracis* during an infection has yet to be elucidated.

Given the pivotal role played by Spo0A~P in the decision between sporulation and virulence in *B. anthracis*, surprisingly little was known until recently of the signals or the sporulation sensor kinase(s) that feeds into the sporulation phosphorelay in this organism. Functional analysis of nine putative sporulation sensor histidine kinase-encoding genes recently identified in *B. anthracis* indicated several with likely roles in sporulation. Of particular interest is the chromosomally encoded sensor histidine kinase BA2291 (Ames strain designation). Deletion of the gene for BA2291 results in a delay in sporulation in *B. anthracis*, and this protein is able to complement sporulation kinase-deficient mutants (Δ*kinA* Δ*kinB* mutants) of *Bacillus subtilis* when introduced in a single copy, supporting its role as a bona fide sporulation histidine kinase (6).

In this communication we report the identification and char-

---

* Corresponding author. Mailing address: Division of Cellular Biology, Mail Code MEM-116, Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. Phone: (858) 784-7912. Fax: (858) 784-7966. E-mail: mperego@scripps.edu.

† Present address: Department of Genetics, Institute of Biochemistry and Biophysics, PAS 5A Pawinskiego St., 02-106 Warsaw, Poland.

acterization of two virulence plasmid-encoded proteins with strong similarity to the sensor domain of BA2291 and with a role in regulating the activity of this sporulation kinase.

MATERIALS AND METHODS

Bacterial strains and growth conditions. All *B. subtilis* strains used in this study are derivatives of JH642 (*trpC2 phe-1*). *B. subtilis* strains JH11422 (Δ*kinA*::*cat*) and JH16567 (Δ*kinA*::*cat* Δ*kinB*::*tet*) were transformed with plasmid pCm::Spc (25) in order to replace the chloramphenicol resistance gene with the spectinomycin resistance gene, giving rise to strains JH19190 and JH19191, respectively. These strains were transformed with plasmid pJAK2291 (6) so that the BA2291 gene and its promoter were integrated into the chromosome by double crossover recombination at the *amyE* gene selecting for chloramphenicol resistance. The resulting strains were named JH19192 and JH19193, respectively. All *B. anthracis* strains are derivatives of the Sterne strain 34F2 (pX01^{+} pX02^{-}). The construction of *B. anthracis* ΔBA4223 and ΔBA2291 strains was described by Brunsing et al. (6). The transformation of *B. anthracis* strains with pHT315 and its derivatives was done as previously described (15). The transformation of *B. subtilis* strains was done as described by Anagnostopoulos and Spizizen (1).

Bacterial strains were grown in Schaeffer's sporulation medium (SM) (24) or Luria-Bertani (LB) medium with the appropriate antibiotics. For *B. subtilis*, spectinomycin was used at 50 μg/ml, and chloramphenicol was used at 5 μg/ml. For both *B. anthracis* and *B. subtilis* strains harboring plasmid pHT315 and its derivatives, erythromycin and lincomycin were used at 5 and 25 μg/ml, respectively.

Spore assays. Images of live sporulating cells were captured after growth in 5 ml SM broth supplemented with erythromycin and lincomycin for 17 h at 37°C with shaking. Sporulation phenotypes were examined on SM agar plates by streaking isolated colonies of the desired strains onto SM agar plates containing erythromycin and lincomycin. The plates were incubated at 37°C for 48 h.

Liquid sporulation assays were carried out in SM supplemented with erythromycin and lincomycin. Cultures (5 ml) were grown for 48 h at 37°C. Cells were plated as duplicate serial dilutions before and after treatment with chloroform (10%, vol/vol, final concentration). The percentage of sporulation was calculated as the ratio of the spore count after $CHCl_3$ treatment to the total viable count.

Plasmid construction. Construction of pXO1-118 and pXO2-61 expression vectors in pHT315 (copy number, approximately 15) (2) was carried out by PCR amplification of the genes using genomic DNA of *B. anthracis* 34F2 or plasmid pXO2, respectively, as the template. The respective amplification reactions were carried out with the following pairs of oligonucleotide primers (the restriction site used for cloning is underlined): 5′-CGATGGATATCGGTGTTAGCATGTC-3′ and 5′-ATTGAGAATTCTATAACTCCCAAAAATTTC-3′; and 5′-ATCACCTGCAGTTTATTATTCTGAAATATTTTAATAG-3′ and 5′-CAATAAAGCTTAACAATCATGCTTTTTGTTC-3′. The fragment containing the pXO1-118 gene was digested with EcoRI and EcoRV and cloned in pHT315 digested with EcoRI and SmaI, obtaining plasmid pHT315-118. The fragment carrying the pXO2-61 gene was digested with PstI and HindIII and cloned in similarly digested plasmid pHT315, obtaining plasmid pHT315-61. The fidelity of the PCR was verified by DNA sequence analysis.

Construction of BA2291 overexpression vector. The coding sequence for BA2291 was amplified by PCR from the chromosome of *B. anthracis* 34F2 using the following primers: 5′-TATTCGTCATATGGAAATGGAGGGAATG-3′ and 5′-GACCCTTCGAAGCTTAGAAGCAGTTATACTTAC-3′. The PCR product was digested with NdeI and HindIII and ligated into the same sites of vector pET28 (Novagen), resulting in a fusion to six histidine codons at the 5′ end of the gene (plasmid pET28-BA2291). The insertion sequence was verified by sequencing analysis.

Expression and purification of BA2291. pET28-BA2291 was overexpressed in *Escherichia coli* BL21(DE3) in 1 liter of LB broth containing kanamycin at 30 μg/ml. The culture was grown at 37°C with shaking to an optical density at 600 nm of approximately 0.6. Expression was induced by the addition of a 0.4 mM final concentration of isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were incubated for an additional 3 hours at 37°C. Approximately 5.9 g (wet weight) of cells was harvested by centrifugation and resuspended in binding buffer (50 mM Tris-HCl [pH 8.0], 0.3 M NaCl, 10 mM β-mercaptoethanol). Cells were broken by two passages through a French pressure cell at 16,000 lb/in^2, and the cell extract was cleared of the cellular debris and membrane fraction by ultracentrifugation. The resulting cleared lysate was incubated with 3 ml of preequilibrated Ni-nitrilotriacetic acid nickel resin (QIAGEN) for 16 h at 4°C on an orbital rocker. Unbound protein was removed by washing the resin with 150 column volumes of binding buffer followed by 50 column volumes of binding buffer containing 30 mM imidazole. Pure protein was eluted in binding buffer containing 250 mM imidazole and collected in 1-ml fractions. Fractions containing the most pure preparations of BA2291 (98% purity) as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were pooled and concentrated by ultrafiltration with a membrane with a molecular weight cutoff of 30,000. The amino-terminal six-His tag was removed by digestion with thrombin (10 mg of N-terminal six-His–BA2291 and 24 U of thrombin) during dialysis in 1 liter of 50 mM Tris-Cl (pH 8.0), 10% glycerol, and 1 mM dithiothreitol using Spectra/Por dialysis tubing with a molecular weight cutoff of 12,000 to 14,000. Digestion was carried out for 16 h at 4°C. The digested protein was stored at a final concentration of 0.6 mg/ml (14.6 μM) at −80°C.

Autophosphorylation and phosphotransfer assays. Phosphorylation reactions and purification of *B. subtilis* KinA, Spo0F, and Spo0F∼P were performed as previously described (19, 26). Autophosphorylation assays of KinA and BA2291 used 1 μM and 5 μM concentrations of proteins, respectively. Assays for KinA to Spo0F phosphotransferase activity used the enzymes at 0.2 and 2 μM final concentrations, respectively. When BA2291 was included in these assays, it was used at a final concentration of 5 μM. These assays were carried out in a 30-μl reaction volume at room temperature. Aliquots of 12 μl were removed and added to 2.4 μl of SDS-PAGE sample buffer at 0 min and 60 min of incubation. Samples were analyzed on 15% SDS-PAGE gels. The gels were dried, exposed to a PhosphorImager screen, and analyzed by using ImageQuant software (Molecular Dynamics).

RESULTS

Bioinformatic identification of virulence plasmid-encoded sensor domains. Whole-genome sequence analysis of *B. anthracis* resulted in the identification of two virulence plasmid-encoded proteins with significant sequence similarity to the sensor domain only of the BA2291 sporulation histidine sensor kinase. Proteins encoded by pXO1-118 (GenBank accession number AAT28889.2) (18) and pXO2-61 (GenBank accession number AAT29005.2) share 62% identical and conserved residues in predicted amino acid sequence with the sensor domain of BA2291 (residues 1 to 161) (Fig. 1). The pXO1-118 gene is located in very close proximity to (358 nucleotides) and divergently transcribed from the gene encoding the *trans*-acting virulence gene regulator AtxA on the pathogenicity island of virulence plasmid pXO1. The pXO2-61 protein is encoded by a gene adjacent to an *atxA* pseudogene located on virulence plasmid pXO2. This amplification of signal domain-encoding genes is unique to BA2291 and *B. anthracis*, as proteins similar to the sensor domains of the other sporulation histidine kinases were not found to be encoded elsewhere in the genome and, to the best of our knowledge, such amplification is not known to occur in other organisms. The only exception would be the *Bacillus cereus* strain associated with an illness resembling inhalation anthrax, strain G9241, which carries a gene orthologue to the pXO1-118 and pXO2-61 genes on its virulence plasmid pBC218 (pBC218_0049, accession number NZ_AAEK01000004) (12). The presence of the pXO1-118 and pXO2-61 genes for these sensor domain proteins on the virulence plasmids of *B. anthracis* suggests a possible regulatory mechanism allowing the coordinate regulation of sporulation and virulence.

BA2291-dependent inhibition of *B. anthracis* and *B. subtilis* sporulation by overproduction of the sensor domains. To determine if any regulatory effect was exerted by the virulence plasmid-encoded sensor domains on BA2291 in *B. anthracis*, each of the sensor domains was expressed from its own promoter on multicopy plasmid pHT315 (2) and introduced into several *B. anthracis* strains. The sporulation phenotypes of these strains were examined using phase-contrast microscopy

```
pXO2-61       -----------MEEIKCLLCRYLKERQEKFISDWKKKVIIRERDPYKEEIIKNGEHLLSA 49
pXO1-118      -----------MEATKRYLCLYLKESQEKFISNWKKRILVHEHDPYKNEIIKNGTHLLHV 49
Ba2291        MEMEGMEVFPIDKDIKEVFCSHLKNNRHQFVENWKNKMIISDKDPFRLEVVQNGEDLLEF 60
                          :  *  :* :**: : :*: :**::::: ::**:: *:::**  **

pXO2-61       FIMYLKEEISLQEIEITSKKIARERIDAKVNIAEFIHNTNVAKIEIMNILTLLNPDLQQY 109
pXO1-118      FTMYMREEINLQDIEDISKKIAQERMDAKVNIADFIYNTNEGKKEILNTLFLLNPTGQEC 109
Ba2291        IIELIMEEKDINYLQPLCEKIAIERAGADANIGDFVYNANVGRNELFEAMCELDVSAREL 120
              :   : **  :: ::   :*** **  *  ** :*::*:*  : *::: :  *:   ::

pXO2-61       QALVKKINQFFDHLIYYTVHSYYEQKA-------------- 136
pXO1-118      KVVIEQINLFFDHLIYSTIYSYYKLKKEYIHSYYELKKKYN 150
Ba2291        KPIMNQIHTCFDKLIYYTVLKYSEIISRNLEEKQQYINETH 161
              : ::::*:  **:*** *:  *  :     :    :
```

FIG. 1. Amino acid sequence alignment of the *B. anthracis* sensor domains encoded by pXO1-118 and pXO2-61 and the BA2291 sensor domain (residues 1 to 161). Sequences were aligned by the ClustalW program. Asterisks indicate identical residues in all three sequences; colons denote conserved substitutions. Paired scores resulted in 34% identity between the pXO2-61 and BA2291 sensor domains, 29% identity between the pXO1-118 and BA2291 sensor domains, and 62% identity between the pXO1-118 and pXO2-61 sensor domains.

of whole cells after 17 h of growth at 37°C in SM broth (Fig. 2). Expression of pXO2-61 (Fig. 2, pHT315-61) resulted in a marked decrease in sporulation in wild-type *B. anthracis* compared to that of the strain carrying the vector control pHT315 (Fig. 2A and B). The ability of *B. anthracis* carrying pHT315-61 to continue sporulating, albeit at a lower level, might be explained by the existence of seven putative sporulation sensor kinases active in this organism. A single deletion of any of these sporulation kinases results in only a minor reduction in sporulation, at least in laboratory media (6). However, when a deletion of the other major sporulation kinase in this organism, BA4223 (6), is combined with expression of pXO2-61 in *B. anthracis*, the inhibition of sporulation is complete (Fig. 2D and E). The inhibition of sporulation observed due to the

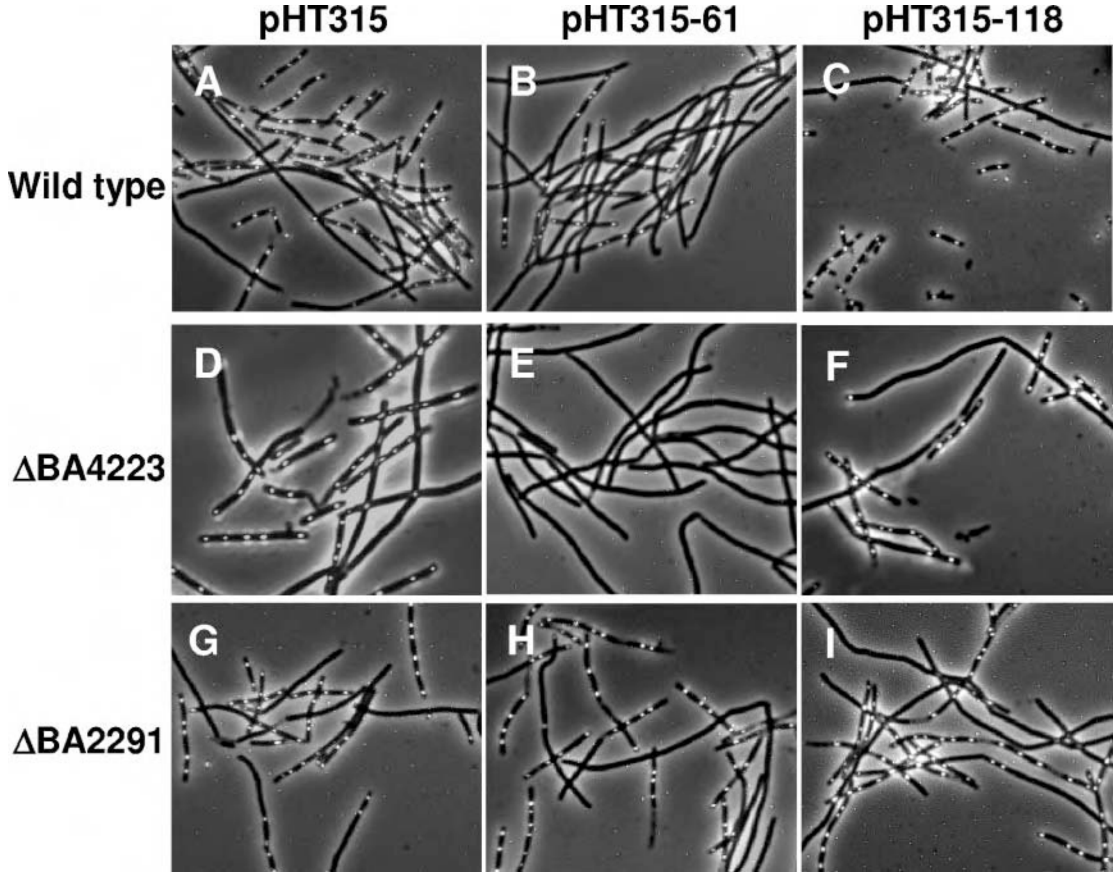


FIG. 2. Sporulation phenotypes of *B. anthracis* parental 34F2, Δ*BA4223*, and Δ*BA2291* strains harboring sensor domains encoded by pX01-118 (pHT315-118) and pXO2-61 (pHT315-61) expressed from their native promoters on multicopy plasmid pHT315. Cultures of each strain were grown in 5 ml of Schaeffer's sporulation medium (24) with the appropriate antibiotics for 17 h at 37°C with shaking.

presence of pXO2-61 is dependent on the presence of BA2291, because the level of sporulation in a *B. anthracis* ΔBA2291 strain carrying pHT315-61 was comparable to the one in the parental strain carrying pHT315 (Fig. 2G and H).

The regulatory effect of pXO1-118 on sporulation in *B. anthracis* was less clear. Overexpression of pXO1-118 on pHT315 did not result in a significant decrease in sporulation in any of the *B. anthracis* strains tested, based on microscopic analysis (Fig. 2C, F, and I) or plate phenotypes on SM agar (data not shown). Because of the tendency of *B. anthracis* cells to remain in long chains rather than break into single units even after the initiation of the sporulation process, a reliable and reproducible quantitation of sporulation efficiency could not be carried out by the spore assay described in Materials and Methods.

To further explore a possible regulatory role for pXO1-118 and pXO2-61 in sporulation initiation, the effects of the sensor domains encoded by both virulence plasmids on the function of BA2291 were analyzed in the case of BA2291-dependent complementation of sporulation in *B. subtilis*. Each virulence plasmid-encoded sensor domain and its native promoter were cloned into the replicative vector pHT315 and transformed into *B. subtilis* sporulation sensor histidine kinase Δ*kinA* and Δ*kinA* Δ*kinB* mutants, respectively, carrying the gene encoding BA2291 integrated into the chromosome in a single copy. The sporulation phenotype of each strain was compared to that containing only the pHT315 vector by examining plate phenotypes on Schaeffer's sporulation agar (Fig. 3) and by carrying out sporulation assays in liquid cultures (Table 1).

Introduction of either sensor domain into *B. subtilis* Δ*kinA* or Δ*kinA* Δ*kinB* mutants in the absence of BA2291 had no significant effect on sporulation compared to the vector-only control, as determined by the level of opacity within the streaks and isolated colonies on SM agar plates (Fig. 3, streaks 1, 2, 3, 7, 8, and 9). In *B. subtilis*, colony opacity increases with the level of sporulation; Spo0 mutant colonies are transparent. In contrast, when either sensor domain was introduced into the strains expressing BA2291, a significant decrease in BA2291-dependent sporulation, marked by a severe decrease in opacity, was observed (Fig. 3, streaks 4, 5, 6, 10, 11, and 12). This effect was much more severe in the presence of the sensor domain encoded by pXO2-61 than with that encoded by pXO1-118. However, in the presence of either sensor domain and BA2291, the level of sporulation was less than what was observed in the same strains in the absence of BA2291 (Fig. 3, compare streak 5 to streak 2, 6 to 3, 11 to 8, and 12 to 9). This indicates that not only was the BA2291-dependent complementation of sporulation previously observed inhibited, but the sporulation process induced by other sporulation kinases was actually blocked by the presence of the sensor domains in a BA2291-dependent manner.

The ability of the sensor domains to inhibit sporulation in *B. subtilis* in a BA2291-dependent manner was further demonstrated by examining the plate phenotypes of wild-type *B. subtilis* in the presence and absence of BA2291 and each sensor domain (Fig. 3). In the absence of BA2291, sporulation appears normal in strains in which either of the two sensor domains is expressed (Fig. 3, streaks 13, 14, and 15). However, in the presence of BA2291, sporulation is completely abolished when pXO2-61 is introduced (Fig. 3, streak 17), and sporulation is diminished with the introduction of pXO1-118 (Fig. 3, compare streak 18 to streak 16, in particular in the area with single colonies). Quantitation of sporulation efficiencies in liquid cultures essentially concurred with the visual analysis of agar plates (Table 1), except that the effect of pXO1-118 did not seem to be as detectable as it was when cells were grown on a solid surface. Perhaps growth in a liquid versus in a solid medium differentially affects the level of expression of pXO1-118, thus resulting in seemingly different phenotypes.

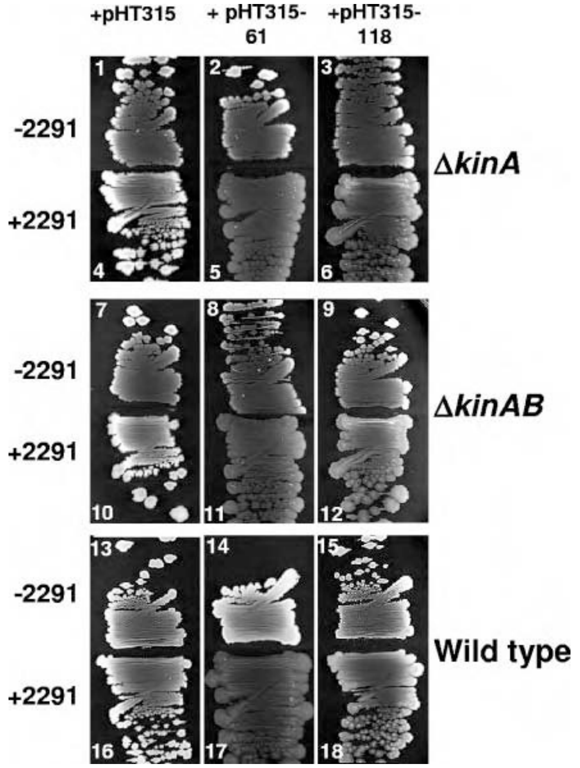


FIG. 3. Effects of overexpression of the sensor domains encoded by pX01-118 and pX02-61 on the sporulation phenotypes of *B. subtilis* wild-type, Δ*kinA* mutant, and Δ*kinA* Δ*kinB* mutant strains with (+2291) and without (−2291) the *B. anthracis* sporulation sensor kinase BA2291 integrated in a single copy on the chromosome. Strains were streaked on Schaeffer's sporulation medium agar (24) and incubated at 37°C for 48 h. Opaque sporulating strains appear white, and nonsporulating strains appear black/gray. The streak numbers correspond to the numbers in column 1 of Table 1.

By examining the effect of each sensor domain on BA2291-dependent sporulation in *B. subtilis* rather than in *B. anthracis*, we were able to isolate the regulatory effects of both sensor domains on BA2291 independently from any additional regulatory networks that might exist in the native host.

The sensor domains convert the BA2291 kinase to an inhibitor of sporulation. A mechanism by which the pXO1-118- and pXO2-61-encoded sensor domains regulate the activity of BA2291 to be either a contributor to or an inhibitor of sporulation is suggested by the observation that, while BA2291 in a single copy complements sporulation kinase-deficient mutants of *B. subtilis*, BA2291 expressed in multicopy completely abolishes the normally high levels of sporulation in wild-type *B. subtilis* (6). A possible explanation for this observation is that with a single copy, there is adequate signal available to activate BA2291 for autophosphorylation and subsequent phospho-

TABLE 1. Effects of pX01-118 and pX02-61 on sporulation in *B. subtilis*[a]

No.[b]	Strain	Relevant genotype	Vector	Spores/ml	Viable cells/ml	% Sporulation
1	JH19190	Δ*kinA*	pHT315	5.2×10^4	1.7×10^8	0.03
2	JH19190	Δ*kinA*	pHT315-61	1.6×10^5	3.1×10^8	0.05
3	JH19190	Δ*kinA*	pHT315-118	1.3×10^5	2.7×10^8	0.05
4	JH19192	Δ*kinA amyE*::*B2291*	pHT315	1.4×10^8	5.0×10^8	28.0
5	JH19192	Δ*kinA amyE*::*B2291*	pHT315-61	2.5×10^4	3.2×10^8	0.008
6	JH19192	Δ*kinA amyE*::*B2291*	pHT315-118	1.0×10^8	4.0×10^8	25.0
7	JH19191	Δ*kinA* Δ*kinB*	pHT315	0	1.8×10^8	0
8	JH19191	Δ*kinA* Δ*kinB*	pHT315-61	0	2.8×10^8	0
9	JH19191	Δ*kinA* Δ*kinB*	pHT315-118	0	2.1×10^8	0
10	JH19193	Δ*kinA* Δ*kinB amyE*::*B2291*	pHT315	5.5×10^7	2.8×10^8	19.0
11	JH19193	Δ*kinA* Δ*kinB amyE*::*B2291*	pHT315-61	0	1.4×10^8	0
12	JH19193	Δ*kinA* Δ*kinB amyE*::*B2291*	pHT315-118	7.7×10^7	3.5×10^8	22.6
13	JH642	Wild type	pHT315	1.4×10^8	4.8×10^8	29.2
14	JH642	Wild type	pHT315-61	1.1×10^8	4.0×10^8	27.5
15	JH642	Wild type	pHT315-118	1.2×10^8	3.9×10^8	30.7
16	JH19169	*amyE*::*B2291*	pHT315	1.3×10^8	2.5×10^8	52.0
17	JH19169	*amyE*::*B2291*	pHT315-61	6.8×10^7	4.0×10^8	17.0
18	JH19169	*amyE*::*B2291*	pHT315-118	1.6×10^8	3.5×10^8	45.7

[a] Strains were grown for 48 h at 37°C in SM plus erythromycin-lincomycin, and the spore assay was carried out as described in Materials and Methods. Values are representative of four independent experiments.

[b] Numbers correspond to the streak numbers in Figure 3.

transfer to the sporulation phosphorelay, resulting in sporulation. However, when BA2291 is present in multicopy, only a small portion of BA2291 in the cell is activated, due to insufficient levels of activating signal. The remaining portion of BA2291 that is not bound by activating signal is in a form that inhibits sporulation.

In vitro studies of purified BA2291 demonstrate that this histidine kinase does not autophosphorylate at a detectable level in vitro, yet it retains the ability to remove phosphoryl groups from Spo0F~P that has been produced by phosphoryl group transfer from KinA~P (the major sporulation kinase in *B. subtilis*) (7, 13) to Spo0F (Fig. 4A). We propose that the ability of BA2291 to inhibit sporulation is due to its ability to remove phosphoryl groups from the phosphorelay at the level of Spo0F and that it is this activity that predominates when BA2291 is not activated to autophosphorylate by activating signal, even when other sporulation kinases are activated by their own signals and feeding phosphoryl groups into the phosphorelay, as is the case in wild-type *B. subtilis* carrying multicopy BA2291. The BA2291-dependent inhibition of

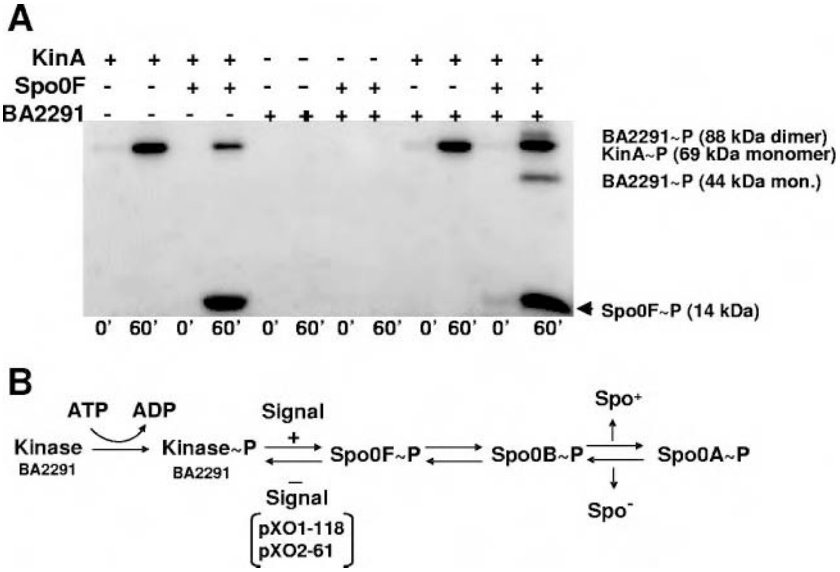


FIG. 4. In vitro activity of the BA2291 histidine sensor kinase. (A) Autophosphorylation and phosphoryl transfer activity assays of BA2291 purified from *E. coli* were carried out as described in Materials and Methods. Autophosphorylation and phosphoryl transfer activities of purified *B. anthracis* BA2291 (5 μM) were compared to those observed for *B. subtilis* proteins KinA (0.2 μM) and Spo0F (2 μM) in the presence of γ-^{32}P-labeled ATP at 0 and 60 min. The samples were run on 15% SDS-PAGE gels. (B) Schematic representation of the phosphorelay signal transduction system for sporulation initiation (7). Emphasized is the role of BA2291 in inducing sporulation (in the presence of activating signal) or inhibiting sporulation (in the absence of activating signal or in the presence of pXO1-118 and pXO2-61) by removing phosphoryl groups from Spo0F~P.

sporulation observed upon introduction of the virulence plasmid-encoded sensor domains is very similar to the inhibition of sporulation observed when BA2291 is present in wild-type *B. subtilis* in multicopy. We suggest that pXO1-118 and pXO2-61 interfere with the ability of BA2291 to perceive and/or transmit a signal for activation, which results in its conversion to an inhibitor of sporulation rather than a contributor to sporulation.

DISCUSSION

We have identified a novel sensor domain family that includes three single-domain virulence plasmid-encoded proteins (those encoded by pXO1-118, pXO2-61, and pBC218_0049) and several multidomain histidine sensor kinases, orthologues of the *B. anthracis* BA2291 that is involved in sporulation initiation in the *Bacillus cereus-B. anthracis-B. thuringiensis* group of spore-forming organisms. There is a high level of amino acid sequence similarity between the pXO1-118- and pXO2-61-encoded proteins and the sensor domain of the sporulation histidine kinase BA2291 of *B. anthracis* (Fig. 1). Structural studies indicate that the plasmid-encoded sensor domains exist as homodimers and exhibit the same globin fold, characterized by a highly hydrophobic pocket suggestive of ligand-binding capabilities (G. Stranzl et al., unpublished data).

It is clear from the studies described in this report that the virulence plasmid-encoded sensor domains have a strong effect on the activity of sporulation sensor kinase BA2291. This effect results in the conversion of BA2291 from a normally functioning sporulation kinase that contributes phosphoryl groups to the sporulation phosphorelays of *B. subtilis* and *B. anthracis* to an enzyme that is able to inhibit sporulation. BA2291 becomes such a potent inhibitor of sporulation in the presence of pXO1-118 and pXO2-61 that it is able to abolish sporulation even in the presence of additional functional and active sporulation sensor kinases that can phosphorylate the Spo0F response regulator (6, 13). Deletion of the pXO1-118 gene does not result in a detectable sporulation phenotype as would be expected for a negative regulator. This is expected, given that only extreme sporulation defects are qualitatively and quantitatively detectable in *B. anthracis* and that the negative regulators of sporulation in *B. subtilis* (for example, KipI, Sda, Rap, or Spo0E phosphatases) give rise to often undetectable phenotypes when deleted (data not shown) (8, 20, 21, 27).

Because of the similarities among these sensor domains, there exist several possible mechanisms by which pXO1-118 and pXO2-61 might interrupt signaling to BA2291. It seems possible that heterodimers might form between monomers of either of the two virulence plasmid-encoded sensor domains and the sensor domain of BA2291. This would prevent normal homodimer formation between two BA2291 monomers, thus preventing the *trans*-autophosphorylation activity required for the input of phosphoryl groups into the sporulation phosphorelay upon binding by activating signal. Although this model is theoretically possible, it seems unlikely, due to the fact that the heterodimer proposed would still have to be able to interact appropriately with Spo0F in order to remove phosphoryl groups from the phosphorelay to inhibit the sporulation as observed. In addition, in pull-down assays in which overexpressed pXO1-118 or pXO2-61 was purified from *B. anthracis*, BA2291 failed to copurify with either protein (data not shown). This suggests that neither virulence plasmid-encoded sensor domain forms a strong heterodimer with BA2291 in vivo.

A more likely model is that the pXO1-118- and pXO2-61-encoded sensor domains competitively bind the same activating signal/receptor as the sensor domain of BA2291. In this manner, expression of pXO1-118 and pXO2-61 would result in the sequestering of BA2291 signal/receptor, resulting in the sporulation-inhibiting form of BA2291 (Fig. 4B).

Additional studies are required to understand the biochemical mechanism of pXO1-118 and pXO2-61 conversion of BA2291, but the fact that the BA2291 protein purified from *E. coli* is inactive as a kinase (presumably because of the lack of activating signal) has so far hampered our attempts to define a mechanism. However, it is clear that these virulence plasmid-encoded sensor domains regulate the activity of sporulation sensor kinase BA2291 and thus regulate sporulation. The fact that each virulence plasmid-encoded sensor domain is located within a pathogenicity island suggests that the regulation of the function of BA2291 by these domains may be the missing link in coordinating the onset of pathogenesis to the inhibition of sporulation required for pathogenesis. This is supported further by the observation that the *trans*-acting virulence gene regulator, AtxA, is also a regulator of pXO2-61 expression (4; Stranzl et al., unpublished). This illuminates a direct tie between inhibition of sporulation and toxin gene expression and adds to the increasingly complicated network of regulation between plasmid-encoded and chromosome-encoded functions in *B. anthracis*.

ACKNOWLEDGMENTS

This study was supported in part by grant AI055860 from the National Institute of Allergy and Infectious Disease and grants GM019416 and GM055594 from the National Institute of General Medical Sciences, National Institutes of Health, United States Public Health Service. M.G. was supported in part by grant 2 PO4B o26 28 from the Polish State Committee for Scientific Research. Oligonucleotide synthesis and DNA sequencing costs were underwritten in part by the Stein Beneficial Trust.

We thank Robert Liddington for the gift of plasmid pXO2.

This article is manuscript number 18131-MEM from the Scripps Research Institute.

REFERENCES

1. **Anagnostopoulos, C., and J. Spizizen.** 1961. Requirements for transformation in *Bacillus subtilis*. J. Bacteriol. **81:**741–746.
2. **Arantes, O., and D. Lereclus.** 1991. Construction of cloning vectors for *Bacillus thuringiensis*. Gene **108:**115–119.
3. **Baillie, L., A. Moir, and R. Manchee.** 1998. The expression of the protective antigen of *Bacillus anthracis* in *Bacillus subtilis*. J. Appl. Microbiol. **84:**741–746.
4. **Bourgogne, A., M. Drysdale, S. G. Hilsenbeck, S. N. Peterson, and T. M. Koehler.** 2003. Global effects of virulence gene regulators in a *Bacillus anthracis* strain with both virulence plasmids. Infect. Immun. **71:**2736–2743.
5. **Brittingham, K. C., G. Ruthel, R. G. Panchal, C. L. Fuller, W. J. Ribot, T. A. Hoover, H. A. Young, A. O. Anderson, and S. Bavari.** 2005. Dendritic cells endocytose Bacillus anthracis spores: implications for anthrax pathogenesis. J. Immunol. **174:**5545–5552.
6. **Brunsing, R. L., C. La Clair, S. Tang, C. Chiang, L. E. Hancock, M. Perego, and J. A. Hoch.** 2005. Characterization of sporulation histidine kinases of *Bacillus anthracis*. J. Bacteriol. **187:**6972–6981.
7. **Burbulys, D., K. A. Trach, and J. A. Hoch.** 1991. The initiation of sporulation in *Bacillus subtilis* is controlled by a multicomponent phosphorelay. Cell **64:**545–552.
8. **Burkholder, W. F., I. Kurtser, and A. D. Grossman.** 2001. Replication initiation proteins regulate a developmental checkpoint in *Bacillus subtilis*. Cell **104:**269–279.
9. **Gould, G. W.** 1977. Recent advances in the understanding of resistance and dormancy in bacterial spores. J. Appl. Bacteriol. **42:**297–309.

10. **Guidi-Rontani, C., and M. Mock.** 2002. Macrophage interactions. Curr. Top. Microbiol. Immunol. **271:**115–141.
11. **Guidi-Rontani, C., M. Weber-Levy, E. Labruyere, and M. Mock.** 1999. Germination of *Bacillus anthracis* spores within alveolar macrophages. Mol. Microbiol. **31:**9–17.
12. **Hoffmaster, A. R., J. Ravel, D. A. Rasko, G. D. Chapman, M. D. Chute, C. K. Marston, B. K. De, C. T. Sacchi, C. Fitzgerald, L. W. Mayer, M. C. Maiden, F. G. Priest, M. Barker, L. Jiang, R. Z. Cer, J. Rilstone, S. N. Peterson, R. S. Weyant, D. R. Galloway, T. D. Read, T. Popovic, and C. M. Fraser.** 2004. Identification of anthrax toxin genes in a Bacillus cereus associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. USA **101:**8449–8454.
13. **Jiang, M., W. Shao, M. Perego, and J. A. Hoch.** 2000. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. Mol. Microbiol. **38:**535–542.
14. **Kang, T. J., M. J. Fenton, M. A. Weiner, S. Hibbs, S. Basu, L. Baillie, and A. S. Cross.** 2005. Murine macrophages kill the vegetative form of *Bacillus anthracis*. Infect. Immun. **73:**7495–7501.
15. **Koehler, T. M., Z. Dai, and M. Kaufman-Yarbray.** 1994. Regulation of the *Bacillus anthracis* protective antigen gene: CO_2 and a *trans*-acting element activate transcription from one of two promoters. J. Bacteriol. **176:**586–595.
16. **Makino, S.-I., I. Uchida, N. Terakado, C. Sasakawa, and M. Yoshikawa.** 1989. Molecular characterization and protein analysis of the *cap* region, which is essential for encapsulation in *Bacillus anthracis*. J. Bacteriol. **171:** 722–730.
17. **Mock, M., and A. Fouet.** 2001. Anthrax. Annu. Rev. Microbiol. **55:**647–671.
18. **Okinaka, R. T., K. Cloud, O. Hampton, A. R. Hoffmaster, K. K. Hill, P. Keim, T. M. Koehler, G. Lamke, S. Kumano, J. Mahillon, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P. J. Jackson.** 1999. Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. J. Bacteriol. **181:**6509–6515.
19. **Perego, M., S. P. Cole, D. Burbulys, K. Trach, and J. A. Hoch.** 1989. Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins Spo0A and Spo0F of *Bacillus subtilis*. J. Bacteriol. **171:**6187–6196.
20. **Perego, M., C. G. Hanstein, K. M. Welsh, T. Djavakhishvili, P. Glaser, and J. A. Hoch.** 1994. Multiple protein aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *Bacillus subtilis*. Cell **79:**1047–1055.
21. **Perego, M., and J. A. Hoch.** 1987. Isolation and sequence of the *spo0E* gene: its role in initiation of sporulation in *Bacillus subtilis*. Mol. Microbiol. **1:**125–132.
22. **Russell, A. D.** 1990. Bacterial spores and chemical sporicidal agents. Clin. Microbiol. Rev. **3:**99–119.
23. **Saile, E., and T. M. Koehler.** 2002. Control of anthrax toxin gene expression by the transition state regulator *abrB*. J. Bacteriol. **184:**370–380.
24. **Schaeffer, P., J. Millet, and J. P. Aubert.** 1965. Catabolic repression of bacterial sporulation. Proc. Natl. Acad. Sci. USA **54:**704–711.
25. **Steinmetz, M., and R. Richter.** 1994. Plasmids designed to alter the antibiotic resistance expressed by insertion mutations in *Bacillus subtilis*, through *in vivo* recombination. Gene **142:**79–83.
26. **Tzeng, Y.-L., and J. A. Hoch.** 1997. Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. J. Mol. Biol. **272:**200–212.
27. **Wang, L., R. Grau, M. Perego, and J. A. Hoch.** 1997. A novel histidine kinase inhibitor regulating development in *Bacillus subtilis*. Genes Dev. **11:**2569–2579.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 280, NO. 42, pp. 35433–35439, October 21, 2005
© 2005 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

Structure and Lytic Activity of a *Bacillus anthracis* Prophage Endolysin*[S]

Received for publication, March 11, 2005, and in revised form, August 5, 2005 Published, JBC Papers in Press, August 15, 2005, DOI 10.1074/jbc.M502723200

Lieh Yoon Low[‡1], Chen Yang[‡1], Marta Perego[§], Andrei Osterman[‡], and Robert C. Liddington[‡2]

From the [‡]Infectious and Inflammatory Disease Center, The Burnham Institute and the [§]Division of Cellular Biology, Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037

We report a structural and functional analysis of the λ prophage Ba02 endolysin (PlyL) encoded by the *Bacillus anthracis* genome. We show that PlyL comprises two autonomously folded domains, an N-terminal catalytic domain and a C-terminal cell wall-binding domain. We determined the crystal structure of the catalytic domain; its three-dimensional fold is related to that of the cell wall amidase, T7 lysozyme, and contains a conserved zinc coordination site and other components of the catalytic machinery. We demonstrate that PlyL is an *N*-acetylmuramoyl-L-alanine amidase that cleaves the cell wall of several *Bacillus* species when applied exogenously. We show, unexpectedly, that the catalytic domain of PlyL cleaves more efficiently than the full-length protein, except in the case of *Bacillus cereus*, and using GFP-tagged cell wall-binding domain, we detected strong binding of the cell wall-binding domain to *B. cereus* but not to other species tested. We further show that a related endolysin (Ply21) from the *B. cereus* phage, TP21, shows a similar pattern of behavior. To explain these data, and the species specificity of PlyL, we propose that the C-terminal domain inhibits the activity of the catalytic domain through intramolecular interactions that are relieved upon binding of the C-terminal domain to the cell wall. Furthermore, our data show that (when applied exogenously) targeting of the enzyme to the cell wall is not a prerequisite of its lytic activity, which is inherently high. These results may have broad implications for the design of endolysins as therapeutic agents.

Endolysins are bacteriophage-encoded enzymes that lyse the host bacterial cell wall during the lytic phase of the phage infectious cycle. They typically consist of an N-terminal catalytic domain and a C-terminal domain that targets the enzyme to the cell wall, providing high species and strain specificity (1, 2). For example, the *Listeria monocytogenes* lysins, Ply118 and Ply500, specifically hydrolyze *Listeria* cells but are inactive in the absence of the cell wall-binding domain (1).

A comparative genome analysis of *Bacillus anthracis* revealed a gene encoding a putative endolysin within the integrated copy of the λ Ba02 prophage, which we will call PlyL. PlyL has a high degree of sequence similarity in its catalytic domain with an endolysin from the bacteriophage γ (PlyG) (3, 4), which specifically lyses and kills *B. anthracis* and closely related species when added exogenously to bacterial cultures. For this reason, PlyG is being developed as a diagnostic and therapeutic agent (5).

Here we describe a structural and functional analysis of PlyL. We show that the N-terminal (catalytic) domain is an amidase with high inherent lytic activity against the cell wall of several *Bacillus* species. In contrast to many previously described lysins, we find that the C-terminal domain plays a dual role, not only as a cell wall targeting domain but also as an inhibitor of catalytic activity in the absence of the cognate target.

MATERIALS AND METHODS

Cloning and Expression of Full-length Endolysin and C-terminal Domain—Full-length PlyL was cloned by PCR from the Bacillus anthracis Ames strain total DNA extract provided by Dr Phil Hanna (University of Michigan Medical School) using the oligonucleotide primers 5′-AAAGGAGATATACATATGGAAATCAGAAAAAAATTAGTT-3′ (forward) and 5′-GAATTCGGATCCTCATTATTTATCATCATACCACCAATC-3′ (reverse). We used the forward primer 5′-GGAGATATACATATGGCAAGTGCAACGGTAACCCCTAAA-3′ with the same reverse primer. PCR products were cloned into pET22b (Novagen) via NdeI and BamHI restriction sites (without tag). The resulting plasmids were transformed into BL21DE3 (Novagen) for protein expression. All protein constructs were expressed using the same protocol. Transformed cells from overnight plates were used to inoculate 1 liter of 2× TY medium (16 g/liter Tryptone, 10 g/liter yeast extract, and 5 g/NaCl supplemented with 100 μg/ml ampicillin), and allowed to grow to A_{600} of 1.0 at 37 °C. 1 mM isopropyl 1-thio-β-D-galactopyranoside was added to induce protein expression over 3 h at 37 °C.

Full-length PlyL Purification—Cells were harvested by centrifugation at 4 °C. 30 ml of lysis buffer (50 mM Na-Mes,[3] pH 6.0, 10 mM β-mercaptoethanol, 0.1% Triton X-100, and 0.1 mM $ZnSO_4$) was used to resuspend the cell pellet. Resuspended cells were lysed by sonication and clarified by centrifugation for 1 h at 4 °C. Clarified lysate was loaded directly into a HITRAP 5 ml SP column on an Akta FPLC (Amersham Biosciences) equilibrated with 50 ml of buffer A (50 mM Na-Mes, pH 6.0, 10 mM β-mercaptoethanol, and 0.1 mM $ZnSO_4$). Unbound protein was eluted by washing the column with 50 ml of buffer A. A gradient of 0–1 M NaCl in buffer A with a total volume of 50 ml was applied to the column to elute the protein. Fractions containing the full-length PlyL, more than 90% pure as verified by SDS-PAGE, were pooled and concentrated to 10–20 mg/ml. A final gel filtration column, Superdex 75 (16/60; Amersham Biosciences), was then used to further purify the protein.

Cloning, Expression, and Purification of Ply21 Constructs—The gene encoding Ply21 was provided by Dr. Martin Loessner (Inst.f.Lebensmit-

---

* This work was supported by Grant DAMD17-03-2-0038 from the United States Army Medical Research and Materiel Command. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1YB0) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Fig. 1.

[1] Both authors contributed equally to this work.

[2] To whom correspondence should be addressed: The Burnham Institute, 10901 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-646-3136; Fax: 858-713-9925; E-mail: rlidding@burnham.org.

[3] The abbreviations used are: Mes, 4-morpholineethanesulfonic acid; GFP, green fluorescent protein; CBD, cell wall-binding domain.

tel-u.Ernährungswissens ETH-Zentrum Zürich). An internal NdeI was silence-mutated using an overlap-extension PCR technique. Full-length Ply21 (263 amino acids) and its N-terminal domain (amino acids 1–159) were subcloned into pET15b via NdeI and BamHI sites. Bacterial cell extracts were prepared as described above. The supernatant was loaded onto the equilibrated nickel-nitrilotriacetic acid column and washed with 10 column-volumes of wash buffer (50 mM Tris-Cl, 300 mM NaCl and 30 mM imidazole at pH 7.5). The elution buffer was similar to the wash buffer but included 300 mM imidazole. Thrombin was used to remove the N-terminal His-tag at room temperature for at least 24 h. The cleaved proteins was then purified by Superdex 75 (16/60).

Purification and Crystallization of the PlyL Catalytic Domain—The N-terminal catalytic domain was generated by limited proteolysis of the full-length PlyL using elastase at a ratio of 1:100 at room temperature for 16 h. A Superdex S75 16/60 column (Amersham Biosciences) was used as a final column to purify the catalytic domain. The buffer was 20 mM Tris-Cl, pH 7.0, 100 mM NaCl, 10 mM β-mercaptoethanol. The final purified protein was concentrated to 20 mg/ml. Mass spectrometry and amino acid analysis revealed that elastase cleaved after residue Val-159. The protein appeared as a single band on SDS-PAGE, and the molecular weight was confirmed by matrix-assisted laser desorption ionization-mass spectrometry. Crystals were obtained by hanging drop vapor diffusion at 20 °C, using a reservoir of 0.6 M NaH_2PO_4, 1.0 M K_2HPO_4, 0.1 M acetate at pH 6.7. Each drop consisted of 2 μl of protein and 1 μl of buffer. Crystals grew as hexagonal rods to 0.1 mm × 0.1 mm × 0.3 mm in 3 days at room temperature. They adopt space group $P6_1$ with cell dimensions a = 163.2 Å, c = 37.3 Å. To prepare for cryo-x-ray data collection, the crystals were soaked in a series of steps with crystallization buffer containing glycerol to a final concentration of 20%. All x-ray data sets were collected at 100 K.

PlyL C-terminal Domain Purification and Crystallization—The purification protocol was identical to that of the His-tagged Ply21 constructs described above. Crystals of the C-terminal 75 amino acid domain were obtained by equilibration against 1.5 M $(NH_4)_2SO_4$ and 10% glycerol in Tris-Cl, pH 7.0, by hanging drop vapor diffusion. The crystal grew to a size of 0.1 × 0.1 × 0.3 mm^3 in 7 days at room temperature; they diffract to 2.7 Å resolution using a Rigaku FR-E High Brilliance X-Ray generator and adopt space group $P4_12_12$ with cell dimensions a = 52.5 Å, c = 224.2 Å.

Structure Determination of the PlyL Catalytic Domain—Multiwavelength anomalous diffraction data sets were collected at beamline 9-2 at the Stanford Synchrotron Radiation Laboratory using a MAR345 image plate and processed using the programs DENZO and SCALEPACK (6). The presence of a zinc ion in the crystal was confirmed by a fluorescence scan at the zinc L-I edge. 18 selenomethionine sites were found using SOLVE (7) and used for phase calculation to a resolution of 2.0 Å. An initial model was generated by RESOLVE (8), further model building was done using O (9), and the model was refined with CNS (10) (version 1.1 on Mac OS X). Native crystals were obtained under identical conditions. A data set was collected in-house with a Rigaku FR-E High Brilliance X-Ray generator using the R-axis IV detector. The CNS-refined model of the selenomethionine structure was used as the input template for native refinement to a resolution of 1.86 Å. There are three molecules (A–C) in the asymmetric unit with essentially identical structure (root mean square deviations on Cα coordinates of 0.29 Å) and a solvent content of 46%. Density for the last two amino acids in the molecules A and C are missing. Molecule B has the most complete density throughout, and its B-factors are lower than for the other two molecules. Refinement statistics are presented in TABLE ONE. The coordinates and structure factors have been deposited with the PDB with accession code 1YB0. The catalytically inactive mutant E90A crystallized isomorphously with the wild-type and showed only small differences in the vicinity of the mutation site.

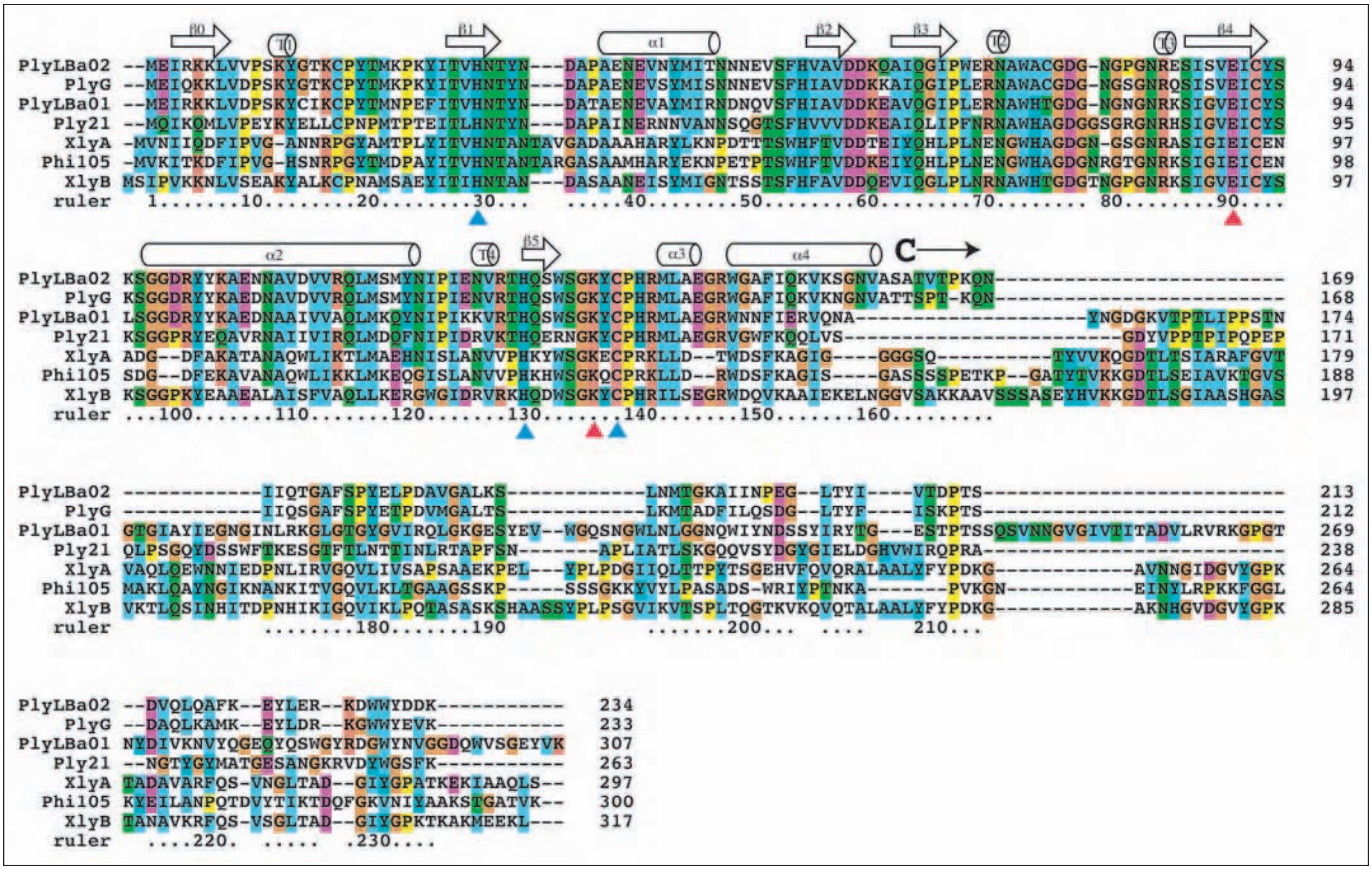
Assay of Lytic Activity—The activity of PlyL and Ply21 when applied exogenously to cultures of B. anthracis (Sterne 34F2), *Bacillus cereus* ATCC 4342, *Bacillus megaterium* WH320, *Bacillus subtilis* 168 and *Escherichia coli* CFT073 were tested. Cultures were grown to mid-exponential phase, and cells were harvested and resuspended in 10 mM sodium phosphate, pH 7.0). The lysis of cell suspensions upon addition of 2–4 μM pure endolysin samples was monitored at 600 nm.

Determination of the Cleavage Site in Peptidoglycan—Peptidoglycan suspension (0.5 mg/ml) from B. subtilis (Fluka) was incubated at 37 °C with purified PlyL (0.4 μM) in 10 ml of Good's buffer (20 mM Na-MES, pH 6.5) containing 100 mM KCl. Boiled PlyL was used as a control. After incubation for 30, 60, and 120 min, samples were boiled and centrifuged at 13,000 rpm/min, clear supernatants were analyzed for the release of free amino acids using a modified protocol described in Ref. 12. 100-μl aliquots were mixed with 12 μl of 10% $K_2B_4O_7$, and 10 μl of 1-fluoro-2,4-dinitrobenzene solution (0.1 M in ethanol) was added, and the mixture was heated at 65 °C for 45 min in the dark. Following acid hydrolysis in 4 M HCl for 12 h at 95 °C, the 2,4-dinitrophenyl-labeled compounds were analyzed by HPLC on a reverse-phase column (C_{18}, 4.6 × 150 mm, Vydac). The labeled amino acids were eluted with a linear gradient from 90% A + 10% B to 30% A + 70% B (A, 10% acetonitrile in 20 mM acetic acid; B, 90% acetonitrile in 20 mM acetic acid), and detected at 365 nm. The release of free reducing groups during the enzymatic reaction was measured by a modified Morgan-Elson reaction (12) using *N*-acetylglucosamine as the standard.

C-terminal Domain Cell Binding Assay—A modified green fluorescent protein (GFP) gene (gift of Dr Ruchika Gupta) was PCR-amplified using the following oligonucleotide primers: 5′-CGCGGCAGCCATATGGTGAGCAAGGGCGAGGAGCTGTTC-3′ and 5′-GCCCGGATCCTCGAGTTACTTGTACAGCTCGTCCATGCC-3′. The resulting fragment was digested by NdeI and XhoI (underlined) and ligated with the XhoI-BamHI fragment of the C-terminal domain of PlyL, which was amplified using 5′-AGCCATATGCTCGAGATGGCAAGTGCAACGGTAACCCCT-3′ (forward) and the same reverse oligonucleotide that was used for the cloning of the full-length protein. The GFP-C-terminal domain fusion and a GFP control were cloned into a pET15b vector via NdeI and BamHI or XhoI, respectively. Both proteins were expressed and purified using nickel-nitrilotriacetic acid affinity chromatography and gel-filtration as described above. Cell samples for the binding assays were obtained by growing Bacilli cultures to late log phase. Cells were harvested, washed with PBS-T (phosphate-buffered saline + 0.1% Tween 20), and incubated with 0.4 mM protein samples (GFP-C-domain fusion or GFP control) for 5 min at room temperature, prior to three washes with PBS-T. The washed cells were smeared onto a microscope slide for confocal image analysis with the Bio-Rad Radiance 2100 Multiphoton Laser Scanning Confocal Microscope system equipped with Argon laser (Image Analysis and Histology Facilities, The Burnham Institute). The objective used was 60× LSM with oil immersion and zoom 5 on the N.A.1.0 (Olympus) microscope. The wavelength of 488 nm was used to excite the GFP.

RESULTS

Identification and Characterization of PlyL—A Blast search (http://ncbi.nlm.nih.gov/BLAST/) using the γ phage endolysin, PlyG, as the query sequence identified two genes encoding putative endolysins located within an integrated prophage of B. anthracis. The λ Ba01 and λ Ba02 endolysins are annotated as BA3767 and BA4073 ("PlyL"), respec-



```
PlyLBa02 --MEIRKKLVVPSKYGTKCPYTMKPKYITVHNTYN---DAPAENEVNYMITNNNEVSFHVAVDDKQAIQGIPWERNAWACGDG-NGPGNRESISVEICYS   94
    PlyG --MEIQKKLVDPSKYGTKCPYTMKPKYITVHNTYN---DAPAENEVSYMISNNNEVSFHIAVDDKKAIQGIPLERNAWACGDG-NGSGNRQSISVEICYS   94
PlyLBa01 --MEIRKKLVDPSKYCIKCPYTMNPEFITVHNTYN---DATAENEVAYMIRNDNQVSFHIAVDDKEAVQGIPLERNAWHTGDG-NGNGNRKSIGVEICYS   94
   Ply21 --MQIKQMLVPEYKYELLCPNPMTPTEITLHNTYN---DAPAINERNNVANNSQGTSFHVVVDDKEAIQLIPFNRNAWHAGDGGSGRGNRHSIGVEICYS   95
    XlyA -MVNIIQDFIPVG-ANNRPGYAMTPLYITVHNTANTAVGADAAAHARYLKNPDTTTSWHFTVDDTEIYQHLPLNENGWHAGDGN-GSGNRASIGIEICEN   97
  Phi105 -MVKITKDFIPVG-HSNRPGYTMDPAYITVHNTANTARGASAAMHARYEKNPETPTSWHFTVDDKEIYQHLPLNENGWHAGDGNRGTGNRKSIGIEICEN   98
    XlyB MSIPVKKNLVSEAKYALKCPNAMSAEYITIHNTAN---DASAANEISYMIGNTSSTSFHFAVDDQEVIQGLPLNRNAWHTGDGTNGPGNRKSIGVEICYS   97
   ruler   1.......10........20........30...   .....40........50........60........70........ 80........90....

PlyLBa02 KSGGDRYYKAENNAVDVVRQLMSMYNIPIENVRTHQSWSGKYCPHRMLAEGRWGAFIQKVKSGNVASATVTPKQN-------------------------  169
    PlyG KSGGDRYYKAEDNAVDVVRQLMSMYNIPIENVRTHQSWSGKYCPHRMLAEGRWGAFIQKVKNGNVATTSPT-KQN-------------------------  168
PlyLBa01 LSGGDRYYKAEDNAAIVVAQLMKQYNIPIKKVRTHQSWSGKYCPHRMLAEGRWNNFIERVQNA----------------------YNGDGKVTPTLIPPSTN  174
   Ply21 KSGGPRYEQAVRNAIIVIRQLMDQFNIPIDRVKTHQERNGKYCPHRMLAEGRVGWFKQQLVS------------------------GDYVPPTPIPQPEP  171
    XlyA ADG--DFAKATANAQWLIKTLMAEHNISLANVVPHKYWSGKECPRKLLD--TWDSFKAGIG----GGGSQ-----------TYVVKQGDTLTSIARAFGVT  179
  Phi105 SDG--DFEKAVANAQWLIKKLMKEQGISLANVVPHKHWSGKQCPRKLLD--RWDSFKAGIS----GASSSSPETKP--GATYTVKKGDTLSEIAVKTGVS  188
    XlyB KSGGPKYEAAEALAISFVAQLLKERGWGIDRVRKHQDWSGKYCPHRILSEGRWDQVKAAIEKELNGGVSAKKAAVSSSASEYHVKKGDTLSGIAASHGAS  197
   ruler ...100.......110.......120.......130.......140.......150.......160.........

PlyLBa02 ------------IIQTGAFSPYELPDAVGALKS------------LNMTGKAIINPEG--LTYI----VTDPTS-------------------------  213
    PlyG ------------IIQSGAFSPYETPDVMGALTS------------LKMTADFILQSDG--LTYF----ISKPTS-------------------------  212
PlyLBa01 GTGIAYIEGNGINLRKGLGTGYGVIRQLGKGESYEV--WGQSNGWLNLGGNQWIYNDSSYIRYTG---ESTPTSSQSVNNGVGIVTITADVLRVRKGPGT  269
   Ply21 QLPSGQYDSSWFTKESGTFTLNTTINLRTAPFSN-------APLIATLSKGQQVSYDGYGIELDGHVWIRQPRA-------------------------  238
    XlyA VAQLQEWNNIEDPNLIRVGQVLIVSAPSAAEKPEL---YPLPDGIIQLTTPYTSGEHVFQVQRALAALYFYPDKG-------------AVNNGIDGVYGPK  264
  Phi105 MAKLQAYNGIKNANKITVGQVLKLTGAAGSSKP-----SSSGKKYVYLPASADS-WRIYPTNKA-------PVKGN-----------EINYLRPKKFGGL  264
    XlyB VKTLQSINHITDPNHIKIGQVIKLPQTASASKSHAASSYPLPSGVIKVTSPLTQGTKVKQVQTALAALYFYPDKG-------------AKNHGVDGVYGPK  285
   ruler             ........180.......190         .......200...  ....    210...

PlyLBa02 --DVQLQAFK--EYLER--KDWWYDDK-----------  234
    PlyG --DAQLKAMK--EYLDR--KGWWYEVK-----------  233
PlyLBa01 NYDIVKNVYQGEQYQSWGYRDGWYNVGGDQWVSGEYVK  307
   Ply21 --NGTYGYMATGESANGKRVDYWGSFK-----------  263
    XlyA TADAVARFQS-VNGLTAD--GIYGPATKEKIAAQLS--  297
  Phi105 KYEILANPQTDVYTIKTDQFGKVNIYAAKSTGATVK--  300
    XlyB TANAVKRFQS-VSGLTAD--GIYGPKTKAKMEEKL---  317
   ruler   ....220.  .....  .230....
```

FIGURE 1. **Sequence alignment of a family of *Bacillus* endolysins.** The residues involved in zinc binding and catalysis are conserved among the *Bacillus* endolysins, and indicated by *blue* and *red triangles*, respectively. Secondary structural elements for the PlyL catalytic domain are indicated. XlyB, XlyA, Ply21, and ϕ-105 are endolysins from *Bacillus licheniformis*, *B. subtilis*, *B. cereus*, and *B. subtilis*, respectively. The catalytic domain ends after helix α4, and the beginning of the C-terminal domain is indicated. Alignment was created using the program ClustalX version 1.82 (20).

tively, in the genome sequence of *B. anthracis* Ames (NCBI accession number NC_003997). Additional endolysins from other *Bacillus* species and their phages were also detected in this search. Those with greater than 30% identity over their catalytic domains are shown in Fig. 1. PlyL is most closely related to PlyG in both the enzymatic (93% identity) and C-terminal (60% identity) domains. BA3767 is also very similar but lacks the C-terminal domain.

We cloned and expressed a *B. anthracis* gene encoding BA4073/PlyL. Crystallization trials of the full-length protein were unsuccessful. However, limited proteolysis using elastase allowed us to isolate a stable N-terminal fragment (residues 1–159). Cleavage occurs at the junction between the predicted catalytic and cell wall-binding domains. This fragment was much more soluble than the full-length protein (>40 mg/ml *versus* <3 mg/ml), and crystallized readily. We also crystallized the C-terminal domain; although we have not yet solved its structure, the existence of crystals that diffract to high resolution indicates that it is an autonomously folded domain.

N-Acetylmuramoyl-L-alanine Amidase Activity of PlyL Resides in Its N-terminal Domain—To assess the enzymatic activity of PlyL, peptidoglycan from B. subtilis was treated with full-length PlyL and the elastase-generated N-terminal fragment. No increase in free reducing groups derived from peptidoglycan could be observed, indicating that the enzyme is neither a glucosaminidase nor a muramidase. The free amino groups of the digested (solubilized) products were labeled with 1-fluoro-2,4-dinitrobenzene. After acid hydrolysis, the 2,4-dinitrophenyl-labeled compounds were separated by high pressure liquid chromatography. Only the amount of 2,4-dinitrophenyl-alanine was increased significantly (supplemental Fig. 1*A*), which indicates that the enzyme is an *N*-acetylmuramoyl-L-alanine amidase, specifically cleaving the amide bond between *N*-acetylmuramic acid and L-alanine. The same result was observed for the N-terminal proteolytic fragment, showing that it comprises a complete catalytic domain. The N-terminal domain was more active than the full-length protein in this assay (supplemental Fig. 1*B*), providing the first indication that the C-terminal domain is autoinhibitory.

Structure of the PlyL N-terminal Domain—We solved the structure of the PlyL catalytic domain (residues 1–159) at 1.86 Å resolution using multiwavelength anomalous diffraction phasing from a selenomethionine-substituted protein (TABLE ONE). The fold is most similar to those of the T7 lysozyme (13), Citrobacter AmpD (14), and the *Drosophila* peptidoglycan recognition protein PGRP-LB (15), with which it shares 10–20% identity. For consistency, we have followed the strand and helix nomenclature of T7 lysozyme. The overall fold consists of a six-stranded β-sheet flanked by four long α-helices (one at the front (α1) and three at the back (α2, α3, and α4) as well as a number of elaborate loops with short α-helical segments (Figs. 2 and 3*A*). Compared with T7 lysozyme, an N-terminal extension creates an additional β-strand (β0) at one end of the sheet. A zinc ion binds to the front face of the molecule at the center of the active site, coordinated by His-29 from strand β1, and by two residues, His-129 and Cys-137, on either side of strand β5. The fourth ligand is a phosphate (or sulfate) ion from the crystallization buffer.

Active Site—The active site is solvent-exposed and lies in a shallow groove on the protein surface, consistent with the ability to cleave a

TABLE ONE

Crystallographic statistics

Figures in parentheses refer to the highest resolution shell.

Data collection	Selenomethionine			Native
	Peak	Remote	Inflection	
Wavelength (Å)	0.9792	0.8919	0.9794	1.54
Resolution (Å)	2.03	2.03	2.03	1.86
Resolution range	30–2.03 (2.07–2.03)	30–2.03 (2.07–2.03)	30–2.03 (2.07–2.03)	30–1.86 (1.89–1.86)
Total observations	190,756	182,585	191,392	188,742
Unique reflections	39,226	39,277	39,375	53,283
Completeness	98.9 (96.8)	98.4 (95.5)	96.9 (98.9)	100 (99.9)
Average I/σ	19.1 (3.0)	17.0 (2.6)	18.4 (2.7)	23.6 (2.2)
R_{sym}[a]	10.8 (44.2)	9.2 (45.3)	9.2 (46.5)	8.7 (52.4)
Figure of merit after SOLVE = 0.41				
Refinement		Native		
Refinement range		30.0–1.86		
Number of reflections		48,365		
R_{work}[b]		20.8		
R_{free}[c]		24.3		
Number of refined residues		479		
Number of water molecules		283		
Root mean square deviation from ideality				
Bonds lengths (Å)		0.007		
Bond Angles (°)		1.5		
Average *B*-value (Å^2)	A	B	C	
Protein	27.8	25.9	41.6	
Main chain	26.2	24.3	40.5	
Side chain	29.3	27.4	42.7	
Solvent		34.5		
Ramachandran Plot (%)				
Most favored		85.6		
Additionally allowed		14.2		
Generously allowed		0.2		
Disallowed		0.0		

[a] $R_{sym} = \Sigma|I_h - \langle I_h \rangle|/\Sigma I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent reflection.
[b] $R_{work} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$, where the summation is over the data used for refinement.
[c] R_{free} was calculated using 5% of data excluded from refinement (23).

highly cross-linked and branched polymer. Helix α1 packs more closely against the β-sheet in PlyL than in T7 lysozyme, so that the pronounced substrate-binding groove observed for T7 lysozyme is not seen for PlyL. The active site can be overlaid closely with that of T7 lysozyme (Fig. 3*B*). The three zinc-coordinating residues (His-29, His-129, and Cys-137) are conserved between PlyL and T7 lysozyme (the third zinc-coordinating residue is an Asp in *Citrobacter* AmpD). PlyL Lys-135 is structurally analogous to Lys-128 of T7 lysozyme, which has been shown to be important for catalysis (13), perhaps by stabilizing the developing negative charge on the amide carbonyl in the transition state; however, PGRP-LB has a threonine at this position. Tyr-46 in T7 lysozyme and Tyr-78 in PGRP-LB are important for catalysis and are thought to act as the general base to activate the nucleophilic water molecule. On the basis of sequence alignment the analogous residue in PlyL was predicted to be Phe-53. However, in the crystal structure the side chain of Phe-53 adopts a different orientation and the carboxylate group of Glu-90 (from a neighboring strand) occupies the space analogous to the T7 tyrosine. To demonstrate a catalytic role for Glu-90 in PlyL, we mutated it to alanine, and indeed this mutation completely abolished the amidase activity. The mutant is correctly folded as judged by its ability to crystallize isomorphously with the wild-type protein (data not shown).

There are only 10 amino acid residues different within the N-terminal domains of PlyL and PlyG, so that their three-dimensional structures should be almost identical. These differences are plotted on the three-dimensional model of PlyL (Fig. 3*A*). Most of the differences are located on the surface of the molecule, and all of them are distant from the active site and a putative substrate binding cleft, suggesting that the two catalytic domains should have similar or identical substrate specificity and catalytic activity.

Lytic Activity of PlyL and Ply21—We next examined the lytic activity of PlyL on whole cells of several bacilli, as measured by light scattering (A_{600}) (Fig. 4) and confirmed by microscopy. We found that the full-length PlyL lysed *B. cereus* with an efficiency comparable with that reported for PlyG on *B. anthracis* and some strains of *B. cereus* (5). However, in marked contrast with PlyG, a relatively high lytic activity of PlyL was established on *B. megaterium* and lower but detectable activity on *B. subtilis* and *B. anthracis.*

We found, unexpectedly, that the N-terminal catalytic domain of PlyL is more active than the full-length protein in lysing *B. subtilis*, *B. megaterium*, and *B. anthracis* cells. The strongest enhancement was observed on *B. subtilis* (Fig. 4, *B* and *C*). By contrast, the removal of the C-terminal domain had a barely significant effect on the lytic activity toward *B. cereus*.

To explore the generality of this observation, we studied the lytic activity of the endolysin, Ply21, from the *B. cereus* phage, TP21. Ply21 is 61% identical with PlyL in its catalytic domain, whereas its C-terminal

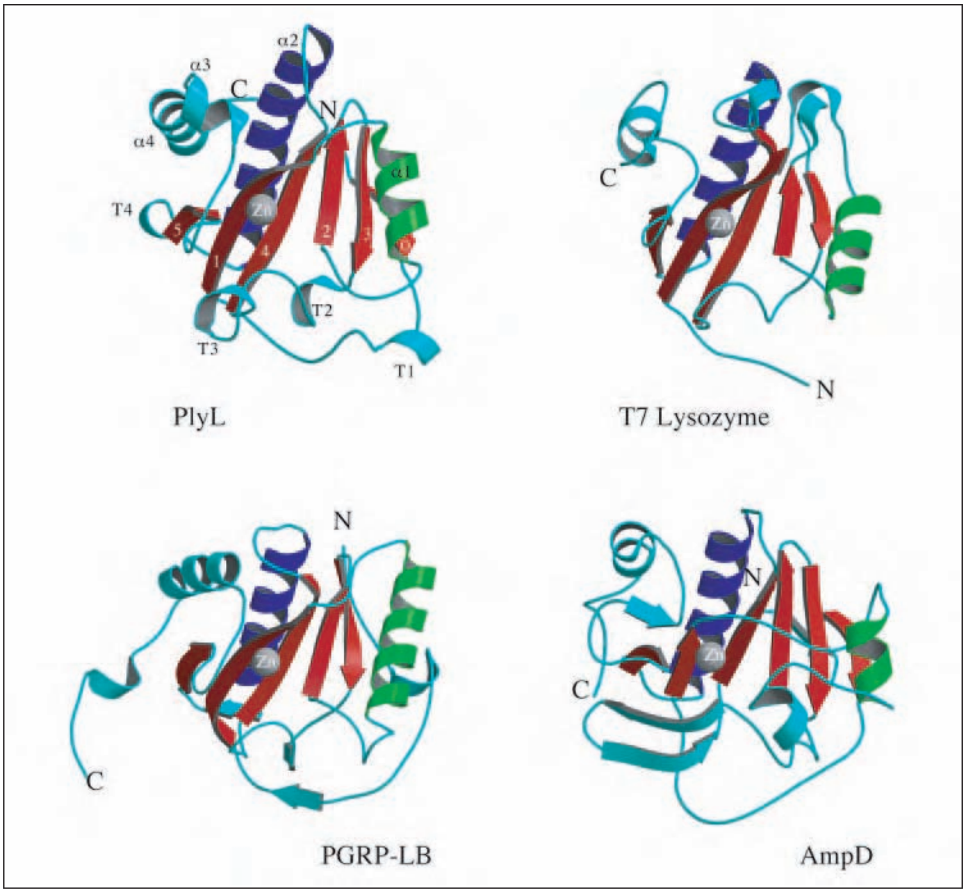


FIGURE 2. **Three-dimensional structure of PlyL catalytic domain and related amidases.** Molscript (version 2.1 (21, 22)) ribbon representations of the structures of PlyL, T7 lysozyme (PDB: 1LBA), PGRP-LB (PDB: 1OHT), and AmpD (PDB: 1J3G). The zinc ion is shown as a *gray sphere*. The colors represent the secondary structure arrangement. The backbone RMS differences with T7 lysozyme and PGRP-LB, are 1.8 Å (for 107 atoms) and 2.0 Å (for 106 atoms), respectively. The N and C termini are labeled *N* and *C*. The N termini of PlyL, PGRP-LB, and AmpD are at the back of the β-sheet.

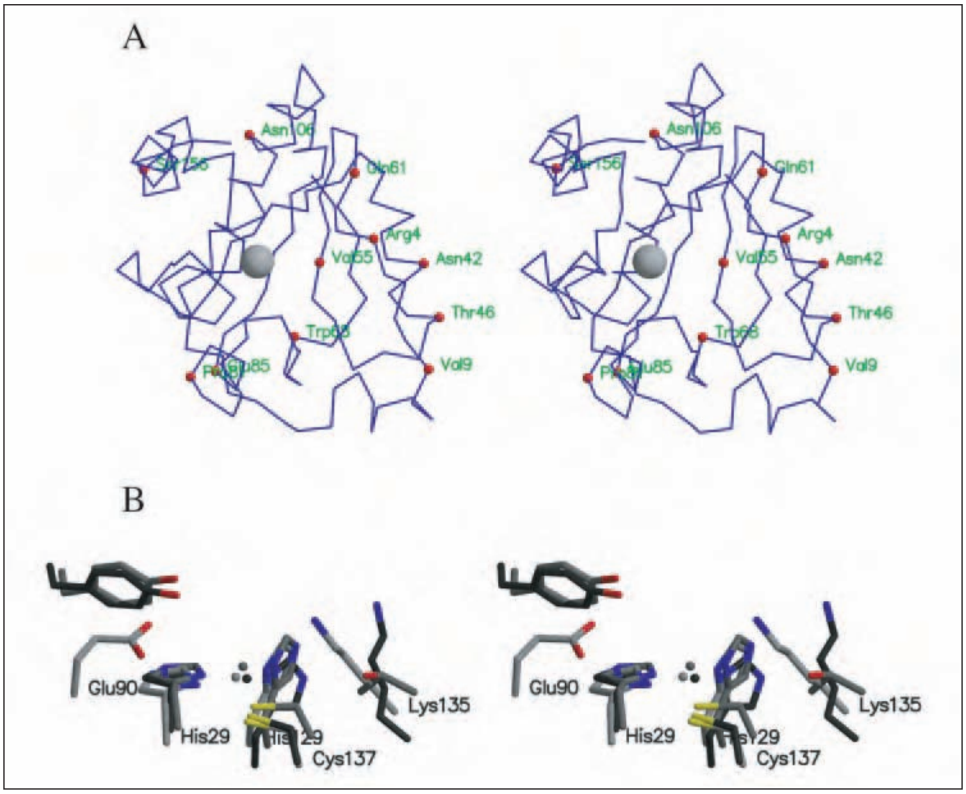


FIGURE 3. **Stereo views of PlyL and active site comparisons.** *A*, stereo Cα representation of PlyL. Amino acids differences between PlyL and PlyG are indicated. Most of these are surface-exposed except for Val-55, which makes hydrophobic contacts with Trp-68 in PlyL. In PlyG, the Val-55 is replaced by the larger residue Ile, but this is complemented by a change to the smaller Leu in place of Trp-68. *B*, stereo view of the active site residues of PlyL (*light gray*), T7 lysozyme (PDB: 1LBA) (*medium gray*), and PGRP-LB (PDB: 1OHT) (*dark gray*).

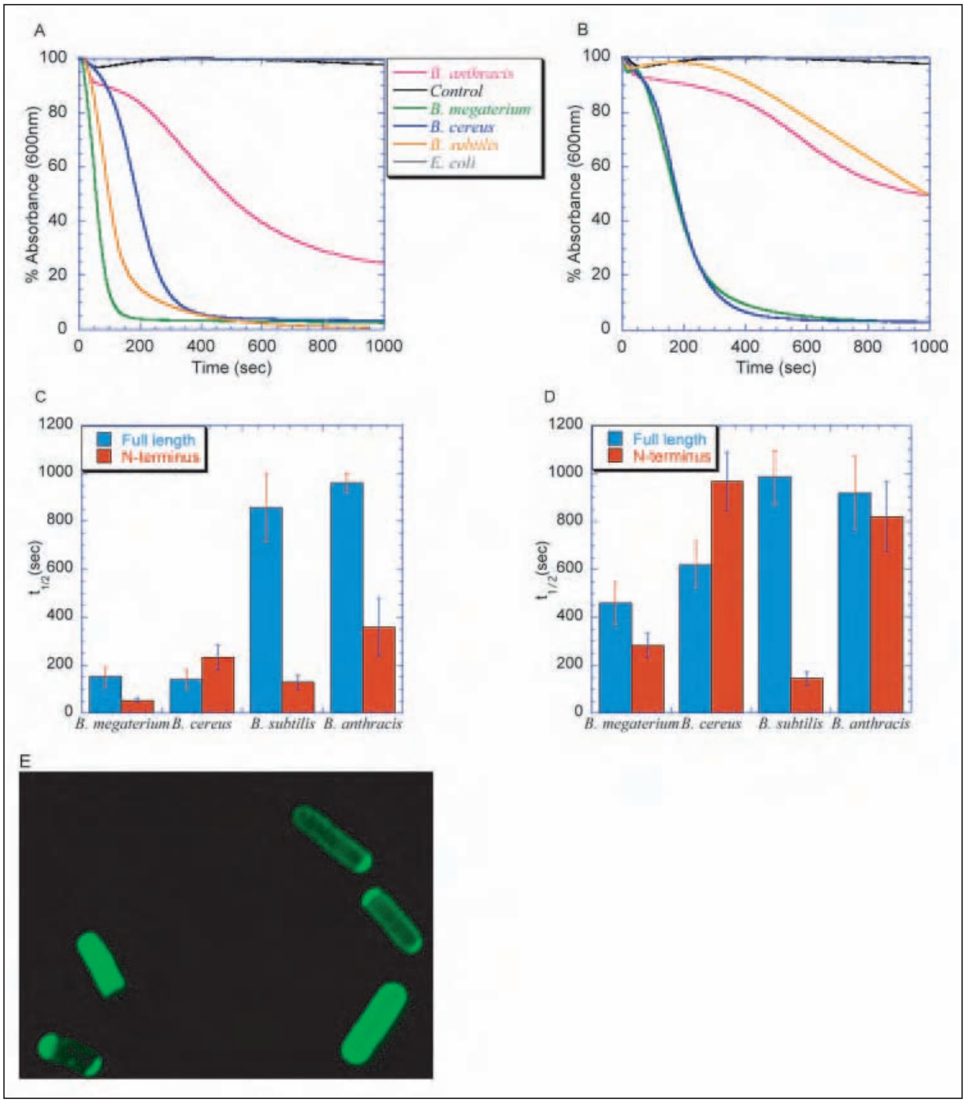


FIGURE 4. **Lytic and cell wall binding activity of PlyL and Ply21.** Lysis of viable cells of four different *Bacillus* species by full-length PlyL (*A*) and its N-terminal catalytic domain (*B*). The protein concentration was 0.4 μM except for *B. anthracis* where 2 μM was used. *C*, the time required for the full-length and catalytic domain of PlyL to reduce the A_{600} by half ($t_{1/2}$). *Error bars* indicate the S.D. from at least three independent experiments. *D*, lysis of Bacillus species by the *B. cereus* phage lysin, Ply21, and its N-terminal domain. Experimental conditions are the same as in *A*. *E*, confocal image of the GFP-PlyL CBD fusion protein binding to the cell wall of *B. cereus,* showing the rod-shape cells with green fluorescence. No fluorescence was observed for other *Bacillus* species or for the control with GFP alone (data not shown).

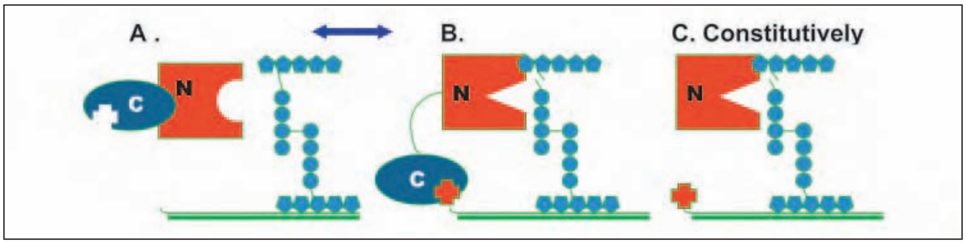


FIGURE 5. **A proposed model of species-specific activation of PlyL.** *A*, in full-length PlyL, the C-terminal domain (*gray oval*) binds to and suppresses the catalytic activity of the N-terminal domain (*blue square*) allosterically. *B*, binding of the C-terminal domain to a cell-wall component (shown by *black cross*) characteristic of a cognate bacterium (such as *B. cereus*) releases the constraints on the catalytic domain, allowing it to adopt an alternative, active, conformation. In the absence of such an interaction partner, as in the case of *B. subtilis*, *B. megaterium* or a free peptidoglycan *in vitro*, the full-length PlyL would exist mostly in the inactive conformation. *C*, a truncation of the C-terminal domain maintains the enzyme in a constitutively active form.

domain has no obvious homology. Ply21 has previously been shown to lyse *B. cereus* strains when added exogenously, whereas *B. subtilis* cells are resistant (16). We confirmed this specificity for the full-length endolysin, but found a dramatic increase in lytic activity toward *B. subtilis* by the free N-terminal domain. In contrast, lysis of *B. cereus* by the N-terminal domain was significantly reduced compared with the full-length enzyme (Fig. 4*D*).

To further assess the role of the C-terminal domain of PlyL, we performed cell binding studies using a recombinant C-terminal domain fused with GFP. When added to *B. cereus* and viewed under a confocal microscope, a clear green fluorescence can be observed around the cells (Fig. 4*E*). No binding was observed with *B. megaterium* or *B. subtilis* (*B. anthracis* was not tested).

DISCUSSION

We have shown that the endolysin from the *B. anthracis* λ prophage Ba02, PlyL, is a *bona fide* cell wall lytic amidase with a modular organization comprising an N-terminal catalytic domain and a C-terminal cell wall-binding domain. We determined the three-dimensional atomic resolution structure of the catalytic domain and showed that the overall fold and active site are similar to but distinct from that of T7 lysozyme and other amidases. The zinc-coordinating residues, His-29, His-129, and Cys-137 are invariant among the *Bacillus* endolysins listed in Fig. 1, as are the other active-site residues, Glu-90 and Lys-135. The role of Glu-90 was not predicted from sequence alignments with T7 lysozyme, but its side chain occupies a similar spatial location to the general base Tyr in T7 lysozyme, and we demonstrated a critical role for Glu-90 in catalysis by mutagenesis.

Our results suggest that all of the enzymes listed in Fig. 1 should have an *N*-acetylmuramoyl-L-alanine amidase activity and a similar catalytic mechanism. This was already demonstrated for the TP21 endolysin, Ply21 (1), and we showed that its isolated N-terminal domain has a similar lytic specificity against *Bacillus* species as PlyL. It seems very likely then that the catalytic domain of the *B. anthracis*-specific phage lysin, PlyG, will also have a similar inherent activity. In particular, the 10 residues that differ between the PlyL and PlyG catalytic domains do not lie close to the active site, so that the distinct lytic specificities of the full-length proteins are presumably endowed by the C-terminal domain, which is less well conserved.

We showed that the C-terminal domain is indeed a cell wall-binding domain (CBD) and that it interacts specifically with *B. cereus* cells. We further showed that the presence of the CBD within the full-length PlyL has an inhibitory effect on the lytic activity of the catalytic domain when tested with peptidoglycan or with the whole cells of *B. subtilis* and, to a lesser extent, with *B. megaterium* and *B. anthracis.* By contrast, the presence of the CBD had a barely significant effect on the activity of PlyL toward *B. cereus.* We established that this behavior is not a peculiarity of PlyL, because we observed a very similar pattern of activity for a phage endolysin that is specific for *B. cereus*, Ply21. This protein has the same domain organization as that of the PlyL and a similar catalytic domain but has little sequence similarity in the C-terminal cell wall-binding domain. It appears that in both cases the cell wall-binding domain serves as an additional level of selectivity by negatively regulating the catalytic domain and only allows the catalytic domain to function effectively in the presence of a specific cell wall.

We therefore propose the following model (Fig. 5); the CBDs of PlyL and Ply21 have dual functions. (i) In the absence of specific interaction with cognate cell wall, the CBD plays an autoinhibitory role, similar to a propeptide in zymogens. Given the structure of the catalytic domain, it is likely that the inhibition is allosteric, because the C terminus of the domain protrudes from a surface that is distal to the active site. (ii) The CBD participates in species-specific cell wall binding (recognition), which disrupts the interaction between the CBD and the catalytic domain thus relieving the inhibitory effect. For example, the marked difference in the activity of the full-length PlyL and the free N-terminal domain against *B. subtilis* can be explained by very weak binding of the CBD to the *B. subtilis* cell wall, whereas the cell wall is intrinsically sensitive to the amidase activity. In the case of *B. cereus* where the full-length and truncated enzymes have an almost equally high activity, we propose that strong binding of the CBD to the target cell wall releases the constraints on the catalytic domain. Similar results were found for Ply21. In that case, localization of the enzymatic domain to the cell surface significantly enhances the rate of lysis, presumably via a local concentration effect.

Endolysins are generally observed to be highly specific toward a particular species of bacteria, by virtue of their distinct CBDs that recognize variable cell wall structures (1, 2). Our observation that the catalytic domain of PlyL has strong lytic activity against a number of different *Bacillus* species and that this activity does not require (or is inhibited by) the CBD suggests either that the PlyL family of endolysins are atypical or that the kinetics of lysis are different when the lysin is applied exogenously rather than endogenously. We note however that there are precedents for such behavior; thus, certain phage hydrolases have been shown to maintain or even increase their exogenous lytic activity when the C terminus is truncated (17–19). These findings may have important implications for the development of lysins as therapeutic agents.

Acknowledgment—We thank Martin J. Loessner at Swiss Federal Institute of Technology Zurich for the gift of the Ply 21 clone and Edward Monosov of the cell imaging facility, The Burnham Institute, for carrying out the confocal microscopy experiments.

REFERENCES

1. Loessner, M. J., Kramer, K., Ebel, F., and Scherer, S. (2002) *Mol. Microbiol.* **44,** 335–349
2. López, R., García, E., García, P., and García, J. L. (1997) *Microb. Drug Resist.* **3,** 199–211
3. Inglesby, T. V., O'Toole, T., Henderson, D. A., Bartlett, J. G., Ascher, M. S., Eitzen, E., Friedlander, A. M., Gerberding, J., Hauer, J., Hughes, J., McDade, J., Osterholm, M. T., Parker, G., Perl, T. M., Russell, P. K., and Tonat, K. (2002) *J. Am. Med. Assoc.* **287,** 2236–2252
4. Brown, E. R., and Cherry, W. B. (1955) *J. Infect. Dis.* **96,** 34–39
5. Schuch, R., Nelson, D., and Fischetti, V. A. (2002) *Nature* **418,** 884–889
6. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
7. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **55,** 849–861
8. Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **56,** 965–972
9. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjelgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
10. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **54,** 905–921
11. Deleted in proof
12. Ghuysen, J.-M., Tipper, D. J., and Schneewind, O. (1966) *Methods Enzymol.* **8,** 685–699
13. Cheng, X., Zhang, X., Pflugrath, J. W., and Studier, F. W. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 4034–4038
14. Liepinsh, E., Genereux, C., Dehareng, D., Joris, B., and Otting, G. (2003) *J. Mol. Biol.* **327,** 833–842
15. Kim, M. S., Byun, M., and Oh, B. H. (2003) *Nat. Immunol.* **4,** 787–793
16. Loessner, M. J., Maier, S. K., Daubek-Puza, H., Wendlinger, G., and Scherer, S. (1997) *J. Bacteriol.* **179,** 2845–2851
17. Loessner, M. J., Gaeng, S., Wendlinger, G., Maier, S. K., and Scherer, S. (1998) *FEMS Microbiol. Lett.* **162,** 265–274
18. Loessner, M. J., Gaeng, S., and Scherer, S. (1999) *J. Bacteriol.* **181,** 4452–4460
19. Baba, T., and Schneewind, O. (1996) *EMBO J.* **15,** 4789–4797
20. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882
21. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
22. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. Sect. D* **50,** 869–873
23. Kleywegt, G. J., and Brünger, A. T. (1996) *Structure* **4,** 897–904

Structure of a GH-25 *N*-acetylmuramidase from *Bacillus anthracis* prophage LambdaBa04 at 1.4 Å

Lieh Yoon Low, Chen Yang, Andrea Osterman, and Robert Liddington.

The Burnham Institute for Medical Research. Centre for infectious and Inflammatory Dieseases.
10901 North Torrey Pines Road, La Jolla, CA 92037.

Abstract

Another endolysin with the similar C-terminal cell wall binding (CWB) as the PlyL was found in LambdaBa04 prophage region of the *Bacillus anthracis* str. Ames. The selectivity is similar to that of the PlyL, but kill bacilli four fold faster than PlyL. We solved the structure of the N-terminal catalytic domain by single-isomorphous replacement to a native resolution of 1.4Å. The fold consists of central eight β-barrel surrounded by five α-helices, a typically fold for the glycosyl hydrolase 25 family (EC 3.2.1.17). Using differential scanning calorimetry, we show that the catalytic domain is more active and less selective than the full-length enzyme, which highlight the usefulness of using only the catalytic domain for therapeutic agent to treat acute bacillus infection.

Introduction

Bacteriophage uses endolysins to break open the cell wall of its host when the assembled progenitor particles are ready to be released. They target different bonds on the peptidoglycan, the major component of the bacterial cell wall. The two major types of endolysin are amidase and glycosidase, which target peptide linker and the sugars moieties of the peptidoglycan respectively. The simplest modular design of endolysin consists of only two domains: catalytic and cell wall binding (CWB) domain. CWB domain is believed to function as targeting the enzyme to the cell-wall substrates. The sequence of the catalytic domain is generally more conserved than the CWB domain. This is because the main components of bacteria cell wall are very similar among different species. However, CWB usually binds to species specific markers on the bacterial cell wall surface, such as the choline on the teichoic acid (Hermoso et al. 2003; Loessner 2005).

In previous studies, we solved the structure of the N-terminal catalytic domain of the plyL, an amidase endolysin (Low et al. 2005). We found that the catalytic domain is more active than the full-length enzyme, suggesting that the C-terminal CWB domain (70 amino acids), not only binds to the cell wall of *B. cereus*, but also could be a negative regulator. Sequence searches using BLAST with the amino acid sequence of this CWB domain as query sequence, we found another endolysin (glycosyl hydrolase 25 family, GH-25) from prophage Lambda Ba04 shares 68 % (69 amino acids) identities in the CWB domain. We called this new endolysin PlyBa04. There is no leader sequence and holin gene at the upstream of the endolysin, so the passage of this enzyme through cell wall is unknown.

The GH-25 hydrolase, EC 3.2.1.17, is an *N*-acetylmuramidase, which cleaves the β-1,4 glycosylic bond between the *N*-acetylmuramic acid and *N*-acetylglucosamine of the peptidoglycan. The first member of the GH-25 family identified and characterized was the lysozyme Ch from a fungus, Chalaropsis species (Hash 1963). The residues involved in catalysis consist mainly of an Asp and a Glu amino acid (Fouche and Hash 1978). There are two existing structures of the GH-25: Cellosyl from *Streptomyces Coelicolor* (1JFX: Rau et al. 2001) and the Cp-1 from *Streptococcus pneumoniae* (1H09: Hermoso et al. 2003). The general fold of the catalytic domain adopts a “non-perfect” alpha/beta barrel, formed by eight parallel β-strands surrounded by five α-helices. The substrate-binding groove is located at the C-terminal end of the β-barrel. The sequence homology between the catalytic domains of the GH-25 hydrolase is very low (Figure 1), with mainly the identical residues found within the β-barrel. There is no sequence similarity beyond the 200 amino acids from the N-terminus.

There is interest in using endolysin as therapeutic agent to treat acute bacillus infection. The basic requirement for the use of endolysin is that the protein be extremely efficient and thermo-stable. In this report, we present the high-resolution structure of the PlyBa04 catalytic domain, and show that the catalytic domain is more stable than the full length protein.

Materials and methods

Cloning, expression and purification. The gene of the N-acetylmuramidase endolysin, protein ID NP_843024.1, was PCR amplified from *Bacillus anthracis* str. Ames (provided by Dr Phil Hanas, University of Michigan Medical School) using the oligos: 5'-CCG CGC GGC AGC CAT ATG GGA CAT ATT ATT GAT ATT TCA-3' and 5'-TTA GCA GCC GGA TCC TTA TGC CGA TTC TGT AAA CCA AGA TAG-3'. An internal NdeI site was removed by silence-mutation (using the overlapped PCR extension method) with the oligos 5'-GGT TTA TAT GTT GGT CAT CAC ATG TAT ACA CCT TTC GGT-3' and 5'-ACC GAA AGG TGT ATA CAT GTG ATG ACC AAC ATA TAA ACC-3'. The amplified DNA product was double digested using NdeI and BamHI, and ligated into pET15b vector (Novagen) before transforming into XL1-BLUE (Stratagene). The correct ligation construct was selected by restriction digest and confirmed by DNA sequence analysis.

The expression vector containing the endolysin gene was transformed into BL21DE3 using $CaCl_2$ method. The freshly transformed cells were grown in 2xTY (16g/L Tryptone, 10 g/L yeast extract, and 5 g/ NaCl; with 100 μg/ml ampicillin) until OD_{600} of 1.0 at 37°C. 1 mM IPTG was added to induce protein expression for three hours at 37°C. Cells were harvested by centrifugation at 4°C. Lysis buffer consisting of 20 mM MES (2-Morpholinoethanesulfonic acid, titrated with NaOH) pH 6.5, 300 mM NaCl and 1 % Triton X-100, was used to resuspend the cell pellet. Cells were lysed by French-press with 3 passes at 1000 PSI and clarified by centrifugation for 1 hour at 4°C. Clarified lysate was loaded directly into a 5 ml HITRAP Chelating column (Amersham Biosciences), which had been charged with $NiCl_2$ and equilibrated with 50 ml buffer A (20 mM MES, 300 mM NaCl, pH 6.5, 1 mM β-mercaptoethanol). Unbound protein was eluted by washing the column with 50 ml in buffer A containing 30 mM imidazole. The His-tagged protein was eluted by 0.3 M imidazole in buffer A. Fractions containing the pure protein were pooled and thrombin (Sigma) added at concentration of 1 unit per mg protein substrate. Thrombin digestion was incubated with rocking at 4°C for 16 hours. Superdex S200 (Amersham Biosciences) gel filtration was used to further clean-up the protein. The first three vector derived residues, Gly-Ser-His-, of final construct are remained after the thrombin removal of the N-terminal His-tag.

Differential Scanning Calorimetry. The DSC was carried out with a VP-DSC differential scanning microcalorimeter from MicroCal, LLC (USA). Protein concentration used was 0.4

mg/ml, and at a scan rate of 1°/min. 0.8 mg/ml of individual protein concentration, added 1:1 volume ratio, was used for the co-melting experiments. The protein constructs were buffer-exchanged into 50 mM Phosphate and 100 mM NaCl using a 5-ml HITRAP (Amersham Biosciences) desalting column just before the experiments.

Bacilli killing experiments. The live cell killing experiments were similar as described (Low et al. 2005). The cells used are *B. cereus* ATCC 4342, *B. megaterium* WH320 and *B. subtilis* 168. Cultures were grown to mid-exponential phase, and cells were harvested and resuspended in 10 mM sodium phosphate (pH 7.0). The lysis of cell suspensions upon addition of 0.4 M pure endolysin samples was monitored at 600 nm using Beckman Coulter DTX880 Multimode Detector. Three samples per experiments were measured on a 96-well plate simultaneously.

Crystallization. Purified protein was concentrated to 30 mg/ml. The protein appeared as a single band on SDS-PAGE, and the molecular weight was confirmed by MALDI-MS. Crystals were obtained by sitting-drop vapor-diffusion, using a reservoir of 25% PEG4000, 0.2 M NaCl, 0.1 M Na-acetate at pH 4.5. Each drop consisted of 4 μl protein and 2 μl buffer.

Data collection. Both native and heavy atoms soaked crystals were collected with a Rigaku FRE High Brilliance X-Ray generator using the R-axis IV detector. Denzo and Scalepack (Otwinowski et al. 1997) were used for indexing and integration. For heavy atom soaking, 10 mM Methylmercuric nitrate (MMN) in the crystallization buffer was added to the drop with crystal at about 1:1 volume ratio, and allowed to equilibrate at 20°C for 2 hours. Both native and heavy atom soaked crystal were mounted onto the goniometer under the liquid nitrogen stream immediately after 10 seconds soaked in cryo-buffer (identical to the precipitant with 25% PEG4000). All X-ray data sets were collected at 100 K.

Single isomorphous replacement. In order to solve the phase problem by isomorphous replacement techniques, heavy metals solutions were added directly to the drops with crystals. First 10 images of the potential derivatized crystals were scaled with the native dataset using Scalepack (Otwinowski et al. 1997). Only MMN added to final concentration of about 10 mM for 2 hours produced a significant difference in the Chi-square value: 44 at 30Å, 15 at 3Å and overall average of 27. SOLVE (Terwilliger and Berendzen 1999) found one site within 12 minutes with figure-of-merit at 0.6 (2 Å) and overall z-score of 7.14. RESOLVE (Terwilliger 2000) built 110 out of 190 amino acid automatically. There is only one protein molecular per asymmetric unit.

Structure refinement. O (Jones et al. 1991) and CNS (Brunger et al. 1998) programs were used to rebuild and refine the native model. Refinement statistics are presented in Table 2. The coordinates and structure factors have been deposited with the PDB with accession code 2H87.

Results

Thermodynamic studies. Both the full-length PlyL and PlyBa04 enzymes are less soluble than individual domains. When mixed at high concentration, at more than 3 mg/ml, visible protein precipitation was observed. At lower protein concentrations, the two domains did not co-migrate on a gel filtration column (data not shown). Differential scanning calorimetry (DSC) was carried out to study the thermal stability of the endolysin with and without the CWB domain. The thermodynamic parameters are presented in Figure 2 and Table 1. The catalytic and CWB domains alone are more stable than the full-length protein. Only a single peak was observed for the full-length protein suggests that the folding of the whole protein is highly cooperative. Interestingly, when the two separately purified domains are mixed, the melting profile showed clearly two peaks, with melting temperature corresponding approximately to the values obtained with individual domains alone. The latter result implies that the two domains when separated, do not interact with each other. Although the full-length protein is a monomer, the CWB alone is a dimer, as estimated from the gel filtration column (data not shown). The dimerisation of the CWB could be stronger than the inter-domain interaction, so that, when added as separate proteins, no complex could be formed.

Enzymatic activities. The PlyBa04 endolysin is at least 4 fold more active than the PlyL and Ply21 described in our previous study (Low et al. 2005) as shown in Figure 3. The catalytic domain can kill *B. cereus* and *B. megaterium* within 100 seconds at 0.1 μM (2 mg/L), compared to average of 200 seconds from amidase endolysin at 0.4 μM. This could be due to the accessibility and/or the glycosidic bonds are more crucial than the amide bonds in terms of cell wall integrity. The species specificity of the PlyBa04 is similar to that of the PlyL. The catalytic domain kills the bacillus quicker than the full-length enzyme, as observed for the PlyL (Low et al. 2005). The CWB of the two proteins may be acting as negative regulator to the catalytic domains.

X-ray Structure Determination. Like PlyL, the N-terminal catalytic domain has pI of near neutral pH (7.7) and C-terminal CWB is acidic. This narrows the buffer pH for solubilizing the full-length protein. Mes buffer (at 20 mM) at pH of 6.5 was used for lysis, purification to storage, as other common buffer with pH at 1 unit higher or lower resulted in protein precipitation.

The crystal grew to about 0.2x0.1x0.1 mm^3 in dimension within 2 weeks. No crystal was obtained when protein was exchanged into a phosphate buffer at pH 6.5 before crystallization setup. The crystal can only be grown using sitting-drop vapor diffusion method, where the crystal actually attached to the plastic surface. Hanging-drop vapour diffusion method either produced no crystal, or a single big (0.5x0.5x0.5 mm^3) multiply-stacked crystals with irregular shape. Most crystals have to be gently scrapped off the surface, and usually results in cracking. Intact crystals sometimes produced smear diffraction spots that were difficult to index with high-resolution spots unusable. Only about 1 in 10 crystals survived and produced good diffraction images.

A MMN derivative dataset was used as the single heavy atom for Single isomorphous replacement technique to solve the crystallographic phase problem. The programs SOLVE (Terwilliger and Berendzen 1999) and RESOLVE (Terwilliger 2000) were used to calculate the phase and build the intial model, respectively. By inspecting the heavy-atoms coordinate together with the model, the Hg was found to be binding to Cys157, which is relatively exposed to solvent. A native dataset at 1.4 Å was used for refinement. The crystallographic statistics are presented in Table 2.

Overall fold. The overall fold is very similar to the existing structures of the GH-25 family (Figure 4). The rmsd of the backbone alignment between Cellosyl (1JFX) and Cp-1 endolysin (1H09) are 1.62 Å (163 backbone carbon atoms) and 1.56 Å (145 backbone carbon atoms), respectively. Eight β-strands form a barrel in the interior of the enzyme. However, unlike a typical alpha-beta-barrel, there are only five long alpha helices on the exterior covering only about half of the circumference around the β-barrel. A negatively charged substrate-binding grove is located at the C-terminal end of the β-barrel. The other end of the β-barrel is “closed” by a short α-helix.

Active site. As shown in Figure 5, a density resembles that of a 6-member ring with an approximate 3-carbon side chain appeared at the substrate-binding cleft. The density was assumed to be a MES buffer molecule as this buffer was used during purification and storage. Glu95 OE2 is hydrogen bonding with the $-NH^+-$ group of the MES buffer (2.68 Å), while the -O- group is hydrogen bonding with GLN161 (2.79 Å). The $-SO_3$ group is forming ionic interaction with Arg32 (3.38 Å).

There are two pairs of highly conserved hydrogen bonded negatively charged residues found in the active site. The side chain distances between the Asp9-Asp175 and Asp93-Glu95 are both 2.6 Å. As suggested by Hermoso et al (Hermoso et al. 2003), these pairs of residues behave as low barrier hydrogen bonds, which have been proposed to be involved in proton trafficking, thus ensuring regeneration of the protonated states of the catalytic residues. The corresponding general acid and base for the catalysis are Asp175 and Asp9, respectively. The second pair, Asp93 and Asp95, may be hydrogen bonded with the substrate/intermediate.

Discussion

The structure of the prophage PlyBa04 cloned from the *B. anthracis* adopts a similar fold as the rest of the GH-25 family. In this family, only a few amino acids are highly conserved. Two pairs of negatively charged amino acid residues have been identified as the essential catalytic residues. They are located at the C-terminal end of the parallel β-barrel, contributing to a highly negatively charged substrate-binding groove. Intriguingly, a MES buffer molecule was found at the active site of the PlyBa04. The bound MES is mainly stabilized by electrostatic and hydrogen bonding with the side chain of the active site residues. No crystal could be obtained, even at exact pH and ionic conditions, without the MES buffer, suggest that the active site is flexible as expected for a typical enzyme with a deep substrate binding groove.

The endolysin PlyL present a interesting case where the CWB domain of the protein exert a negative regulation upon the N-terminal catalytic domain when the specific cell-wall ligand was absent (Low et al. 2005). Based on the structures of the PlyL and the PlyBa04, the C-terminal CWB domain with 70 amino acids should be located on the opposite side of the substrate binding grooves. It is unclear how the polypeptide with similar sequence could have a same allosteric effect on the two kinds of enzymes with different fold.

DSC experiments showed that the full-length two-domain PlyBa04 fold in a highly co-operative manner showing only a single transition on the DSC profile. However, when they are added in as separate proteins, no interactions could be detected. This may be due to the fact that the CWB is a dimer and that the dimerization is stronger than the inter-domain interactions. This might actually mean that there is an opposing force in the folding of this two-domain protein. 1) in the absence of a specific cell-wall ligand, the CWB interacts with the catalytic domain; 2) when a specific ligand is present, the force of binding to the ligand (and perhaps require dimerization), resulted in the release of the binding of the CWB to the catalytic domain. The release of the catalytic domain will then allow the catalysis to proceed efficiently. The CWB of the two *B. anthracis* endolysins are not completely conserved, possibly due to the fact that they have to bind enzymes with different folds.

The alternate hypothesis is that the CWB destabilizes the catalytic domains, when the cell-wall ligand is absent. Both the full-length protein and the CWB alone exhibit a single thermal unfolding transition, suggest that the folding of the two-domain protein and the dimer CWB are both highly co-operative. The isoelectric points (pI's) of the PlyL full length, catalytic and CWB domains are 7.6, 8.8, and 4.8, respectively, whereas, the pI's of the PlyBa04 are 6.5, 7.7, and 4.9, respectively. The opposite charge distribution between the domains may be resulting in a non-specific inter-molecular charge-charge aggregation that may decrease the protein solubility. A specific cell-wall ligand is required to bind to the CWB and break the unfavorable protein complexes.

The latter hypothesis is favorable as it could accommodate the fact that there is no requirement for sequence or structure similarity in the catalytic domains, as long as the overall charge of the domains is positively charged.

Conclusion

The structure of a *B. anthracis* prophage GH-25 has been solved at high resolution of 1.4Å. Together with the sequence comparison and active site conserved amino acids, it is confirmed that the endolysin is a *N*-acetylmuramidase. The specificity of this endolysin is similar but bacillus killing activity is faster than the PlyL. The CWB of this endolysin may also be an auto-inhibitory domain as the full-length enzyme is less active than the catalytic domain alone. The mechanism of the allosteric control is still unclear, but the non-specific charge-charge interaction between the domains of opposite charge could be a destabilizing force that down regulates the activity of the endolysins.

Acknowledgements

The authors would like to thank Dr Phil Hanas, University of Michigan Medical School, for providing the *B. anthracis* DNA material; Dr Andrey Bobkov for performing the DSC experiments.

References

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallographica Section D* **54:** 905-921.

Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. 2004. The Jalview Java alignment editor. *Bioinformatics* **20:** 426-427.

DeLano, W.L. 2005. MacPyMOL: A PyMOL-based Molecular Graphics Application for MacOS X *DeLano Scientific LLC, South San Francisco, CA, USA.* **http://www.pymol.org**.

Fouche, P.B., and Hash, J.H. 1978. The N,O-diacetylmuramidase of Chalaropsis species. Identificaiton of aspartyl and glutamyl residues in the active site. *J Biol Chem* **253:** 6787-6793.

Hash, J.H. 1963. Purification and Properties of Staphylolytic Enzymes from Chalaropsis Sp. *Arch Biochem Biophys* **102:** 379-388.

Hermoso, J.A., Monterroso, B., Albert, A., Galan, B., Ahrazem, O., Garcia, P., Martinez-Ripoll, M., Garcia, J.L., and Menendez, M. 2003. Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* **11:** 1239-1249.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica Section A* **47:** 110-119.

Loessner, M.J. 2005. Bacteriophage endolysins--current state of research and applications. *Curr Opin Microbiol* **8:** 480-487.

Low, L.Y., Yang, C., Perego, M., Osterman, A., and Liddington, R.C. 2005. Structure and lytic activity of a Bacillus anthracis prophage endolysin. *J Biol Chem* **280:** 35433-35439.

Otwinowski, Z., Minor, W., and Charles W. Carter, Jr. 1997. [20] Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology,* pp. 307-326. Academic Press.

Rau, A., Hogg, T., Marquardt, R., and Hilgenfeld, R. 2001. A new lysozyme fold. Crystal structure of the muramidase from Streptomyces coelicolor at 1.65 A resolution. *J Biol Chem* **276:** 31994-31999.

Terwilliger, T. 2000. Maximum-likelihood density modification. *Acta Crystallographica Section D* **56:** 965-972.

Terwilliger, T.C., and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallographica Section D* **55:** 849-861.

Figure Legends

Figure 1. Sequence comparison of the catalytic domain of the GH-25 family endolysins where structure are known. Residues highlighted in red color are completely conserved. The sequences beyond about 200 amino acids are likely to be the species-specific CWB domain. The alignment figure was generated by Jalview version 2.08 (Clamp et al. 2004).

Figure 2. DSC profiles of the full length, N-terminal catalytic and the C-terminal CWB domains of the PlyBa04. Refer to Table 1 for the thermodynamic parameters of the constructs.

Figure 3. The time required to kill the three selected bacilli by the PlyBa04. The full-length enzyme takes longer time to lyse the bacilli compared to the catalytic domain. The error shown represent the standard deviation of three separate experiments.

Figure 4. Electrostatic surface potential and ribbon model of the PlyBa04 generated by Pymol (DeLano 2005). A Mes buffer molecule sits on the highly negative charged active site groove.

Figure 5. Stereo view of showing the Mes molecule binding to the active site of the endolysin. This figure is generated by PyMol (DeLano 2005).

Table 1. Thermodynamic parameters obtained from DSC.

Construct	T_m (°C)	ΔH (kcal/mol)
Full length	48.4	133.7
Catalytic domain	54.1	101.3
CWB domain	60.8	33.6
Catalytic and CWB domains	52.9/60.9	--

Table 2 Crystallographyic data

Data Collection		
Detector	Rigaku R-axis 4	
Wavelength (Å)	1.5418	
Resolution (Å)	50.0-1.60	30.0-1.40
Number of observation	87 504	126 396
Number of unique observation	24 605	36 080
Completeness (%)	98.0 (95.0)	96.5 (85.1)
I/σ	39.4 (18.1)	35.8 (3.5)
R_{sym}[b] (%)	3.7 (12.1)	5.0 (39.0)
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters	48.536, 56.405, 67.348	48.682, 56.454, 67.361
FOM after SOLVE	0.61 (0.61 at 2.06 Å)	
Refinement		
Resolution	30.0-1.4	
Total number of Reflection	35 073	
Number of reflections in test set	1 732	
R_{work}[c] (%)	20.71	
R_{free}[d] (%)	22.79	
Average B-factor (Å^2)	14.8	
Wilson Plot B-factor (Å^2)	13.0	
Number of protein atoms	1 523	
Number of water molecules	231	
Ramachandran statistics[e] (%)		
Most favored	91.4	
Additionally allowed	8.6	
Generously allowed	0.0	
Disallowed	0.0	
RMSD from ideal geometry		
Bond length (Å)	0.005	
Bond angles (°)	1.264	

[a] Numbers in parentheses refer to the highest resolution shell.

[b] $R_{sym} = \Sigma|I_h-\langle I_h\rangle|/\Sigma I_h$, where $\langle I_h\rangle$ is the average intensity over symmetry equivalent reflection.

[c] $R_{work} = \Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$, where the summation is over the data used for refinement.

[d] R_{free} was calculated using 10% of data excluded from refinement (Kleywegt, 1996).

[e] Calculated using PROCHECK (Laskowski R A 1993).

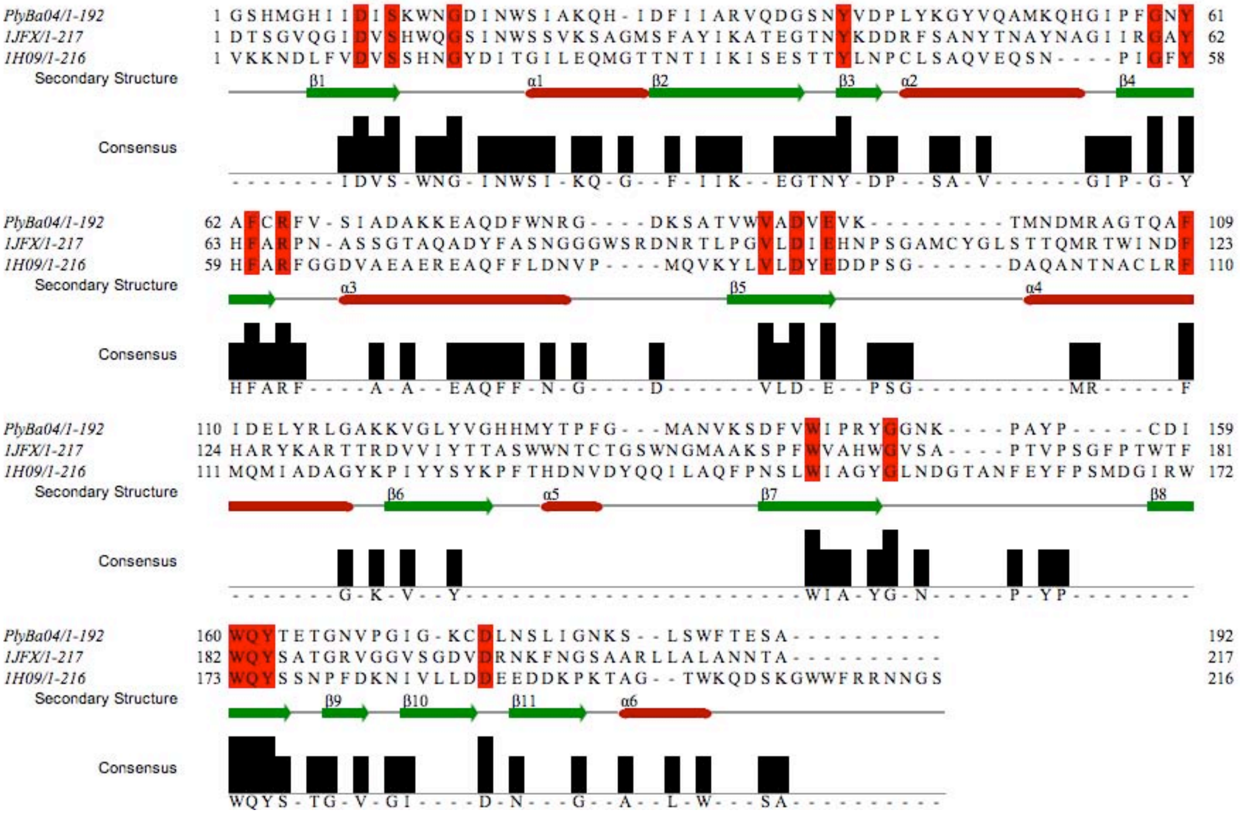

PlyBa04/1-192 1 GSHMGHIIDISKWNGDINWSIAKQH-IDFIIARVQDGSNYVDPLYKGYVQAMKQHGIPFGNY 61
1JFX/1-217 1 DTSGVQGIDVSHWQGSINWSSVKSAGMSFAYIKATEGTNYKDDRFSANYTNAYNAGIIRGAY 62
1H09/1-216 1 VKKNDLFVDVSSHNGYDITGILEQMGTTNTIIKISESTTYLNPCLSAQVEQSN----PIGFY 58
Secondary Structure
β1
α1
β2
β3
α2
β4
Consensus
-------IDVS-WNG-INWSI-KQ-G--F-IIK--EGTNY-DP--SA-V------GIP-G-Y
PlyBa04/1-192 62 AFCRFV-SIADAKKEAQDFWNRG----DKSATVWVADVEVK---------TMNDMRAGTQAF 109
1JFX/1-217 63 HFARPN-ASSGTAQADYFASNGGGWSRDNRTLPGVLDIEHNPSGAMCYGLSTTQMRTWINDF 123
1H09/1-216 59 HFARFGGDVAEAEREAQFFLDNVP----MQVKYLVLDYEDDPSG------DAQANTNACLRF 110
Secondary Structure
α3
β5
α4
Consensus
HFARF----A-A--EAQFF-N-G----D------VLD-E--PSG-----------MR-----F
PlyBa04/1-192 110 IDELYRLGAKKVGLYVGHHMYTPFG---MANVKSDFVWIPRYGGNK----PAYP-----CDI 159
1JFX/1-217 124 HARYKARTTRDVVIYTTASWWNTCTGSWNGMAAKSPFWVAHWGVSA----PTVPSGFPTWTF 181
1H09/1-216 111 MQMIADAGYKPIYYSYKPFTHDNVDYQQILAQFPNSLWIAGYGLNDGTANFEYFPSMDGIRW 172
Secondary Structure
β6
α5
β7
β8
Consensus
-------G-K-V--Y---------------------WIA-YG-N-----P-YP---------
PlyBa04/1-192 160 WQYTETGNVPGIG-KCDLNSLIGNKS--LSWFTESA---------- 192
1JFX/1-217 182 WQYSATGRVGGVSGDVDRNKFNGSAARLLALANNTA---------- 217
1H09/1-216 173 WQYSSNPFDKNIVLLDDEEDDKPKTAG--TWKQDSKGWWFRRNNGS 216
Secondary Structure
β9
β10
β11
α6
Consensus
WQYS-TG-V-GI----D-N---G--A--L-W---SA----------

Figure 1

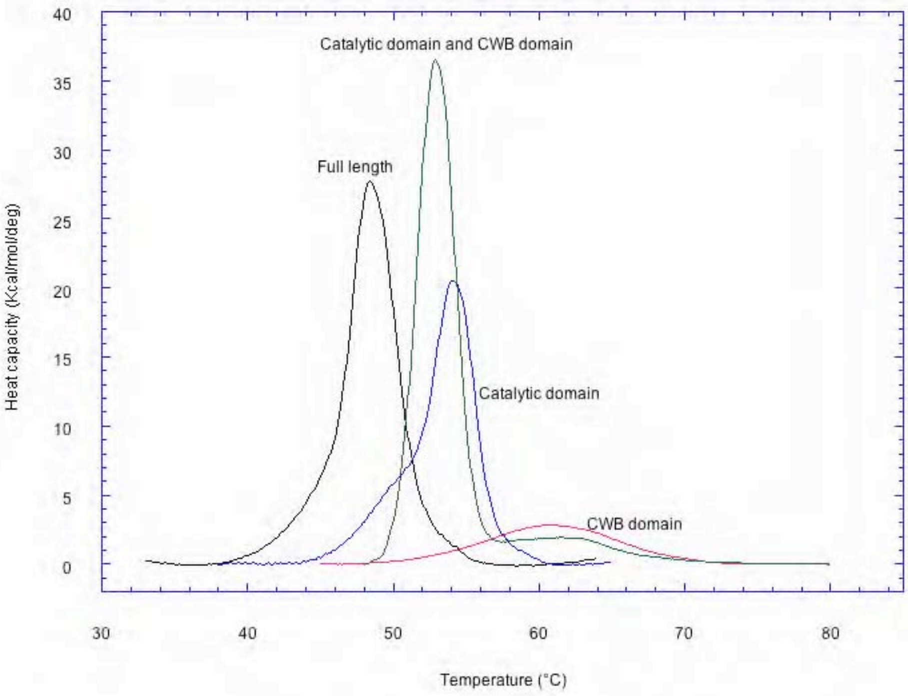

Catalytic domain and CWB domain
Full length
Catalytic domain
CWB domain
Heat capacity (Kcal/mol/deg)
0
5
10
15
20
25
30
35
40
Temperature (°C)
30
40
50
60
70
80

Figure 2

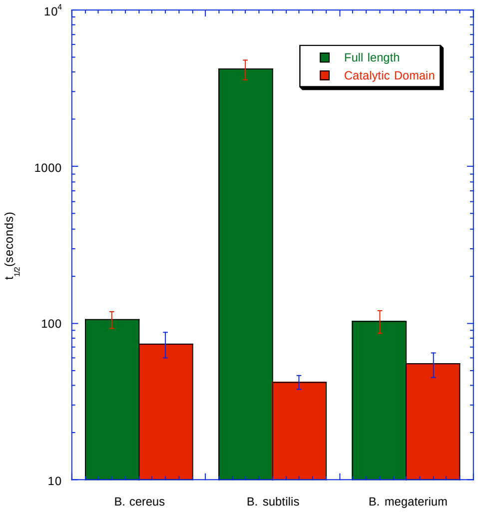

10^4
1000
100
10
t1/2 (seconds)
Full length
Catalytic Domain
B. cereus
B. subtilis
B. megaterium

Figure 3

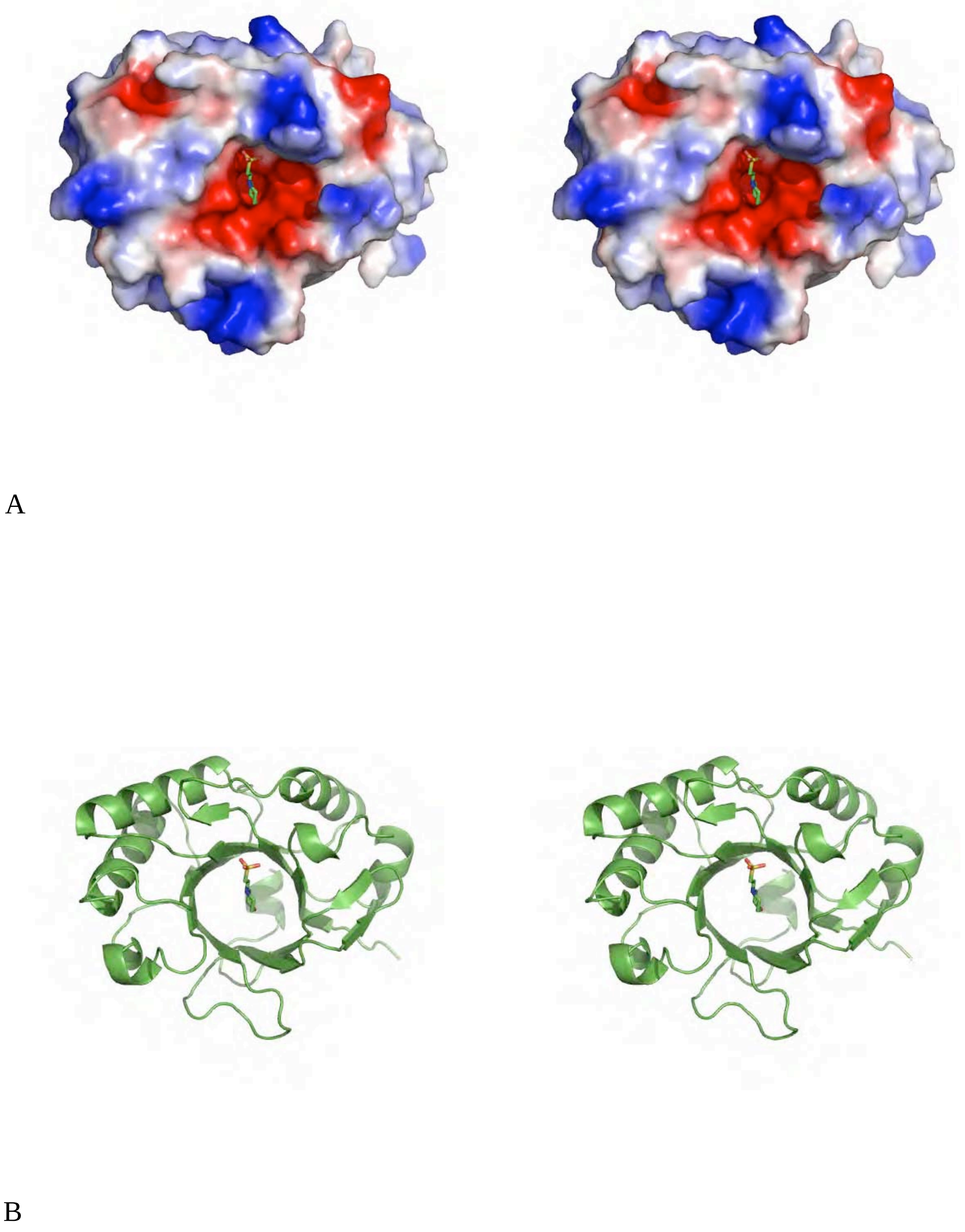


Figure 4

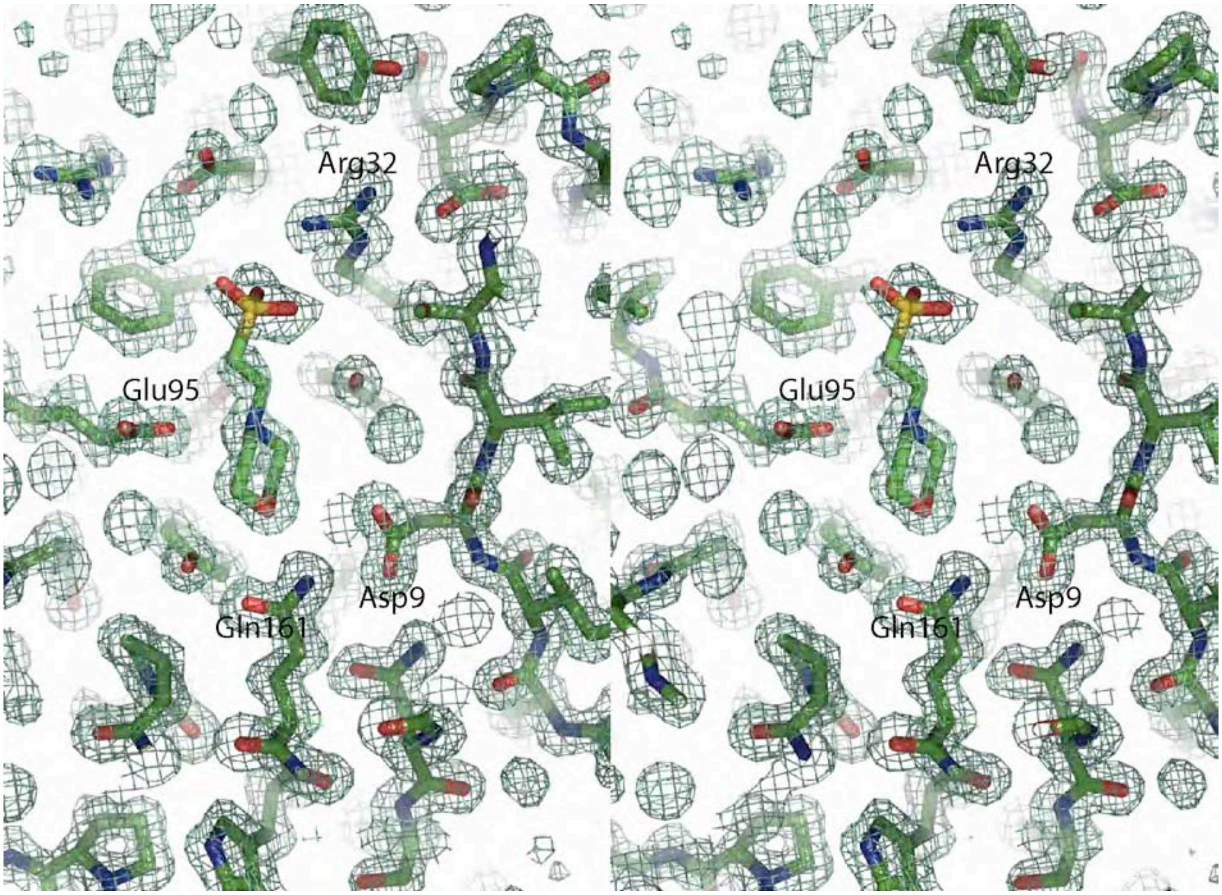


Figure 5

JBC Papers in Press. Published on September 5, 2006 as Manuscript M603275200
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M603275200

STRUCTURAL BASIS OF NEUTRALIZATION BY A HUMAN ANTI-SARS SPIKE PROTEIN ANTIBODY, 80R*

William C. Hwang[1,4], Yaqiong Lin[1,4], Eugenio Santelli[1], Jianhua Sui[2], Lukasz Jaroszewski[1], Boguslaw Stec[1], Michael Farzan[3], Wayne A. Marasco[2] and Robert C. Liddington[1#]

From the [1]Infectious & Inflammatory Disease Center, Burnham Institute for Medical Research, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA; [2] Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA; and [3] Department of Microbiology and Molecular Genetics, Harvard Medical School, New England Primate Research Center, Southborough, MA 01772, USA; [4]These authors contributed equally to the work.

Running title: SARS S1-antibody complex structure

Address correspondence to: Robert C. Liddington, Burnham Institute for Medical Research, 10901 North Torrey Pines Road, La Jolla, CA 92037, Tel. 858-646-3136; Fax 858-792-1768; E-Mail: rlidding@burnham.org.

Severe Acute Respiratory Syndrome (SARS)[1] is a newly emerged infectious disease that caused pandemic spread in 2003. The etiological agent of SARS is a novel coronavirus (SARS-CoV). The coronaviral surface spike protein S is a type I transmembrane glycoprotein that mediates initial host binding via the cell surface receptor angiotensin-converting enzyme 2 (ACE2), as well as the subsequent membrane fusion events required for cell entry. Here we report the crystal structure of the S1 receptor binding domain (RBD) in complex with a neutralizing antibody, 80R, at 2.3 Å resolution, as well as the structure of the uncomplexed S1 RBD at 2.2 Å resolution. We show that the 80R binding epitope on the S1 RBD overlaps very closely with the ACE2 binding site, providing a rationale for the antibody's strong binding and broad neutralizing ability. We provide a structural basis for the differential effects of certain mutations in the spike protein on 80R versus ACE2 binding, including escape mutants, which should facilitate the design of immunotherapeutics to treat a future SARS outbreak. We further show that the RBD of S1 forms dimers via an extensive interface that is disrupted in receptor- and antibody-bound crystal structures, and we propose a role for the dimer in virus stability and infectivity.

Severe Acute Respiratory Syndrome (SARS), a newly emerged infectious disease, claimed 813 lives from ~8000 patients during a 2003 global epidemic. In severe illness, influenza-like symptoms quickly progress to pneumonia, hypoxia, acute respiratory distress and failure, resulting in 10% overall death rate with exceptionally high mortality among the elderly (1). A novel coronavirus (SARS-CoV) has been identified as the etiological agent of SARS. The SARS-CoV surface spike protein S mediates viral entry into the host cell (2), and comprises two functional domains: S1 (G13-R667) and S2 (S668-T1255). S1 contains the host-specific receptor binding domain (RBD) while S2 mediates fusion between viral and host cell membranes (3). Angiotensin-converting enzyme 2 (ACE2) was identified as a functional receptor for the SARS-CoV (4). The recently determined structure of the S1-RBD in complex with the extracellular domain of ACE2 (5) illustrates the structural basis for the initial step of virus-host recognition.

As the mediator of host-specific SARS infection and a major viral surface antigen, the S protein is an attractive candidate for both vaccine development and immunotherapy. Marasco, Farzan, Sui and colleagues (6) previously identified a potent neutralizing human monoclonal antibody against the S1 RBD, designated "80R",

Supplemental Material can be found at:
http://www.jbc.org/cgi/content/full/M603275200/DC1

from two non-immune (i.e., not restricted by B cell recombination) human antibody libraries. 80R binds S1 with nanomolar affinity, blocks the binding of S1 to ACE2, prevents the formation of syncytia *in vitro* (6), and inhibits viral replication *in vivo* (7). Deletion studies have shown that the 80R epitope on S1 is located in the minimal ACE2 binding domain, between residues 324 and 503 (6,7).

Here, we report the crystal structure of the S1-RBD both alone and in complex with 80R. The complex structure reveals the basis of the broad neutralizing ability of 80R, and will facilitate the design of immunotherapeutics in the case of a future SARS outbreak. We further show that the S1-RBD forms dimers by means of an unexpected reorganization of the region distal to the receptor-binding surface. The dimers are disrupted by complex formation, and we discuss the possibility that receptor binding plays an active role in the initial steps of viral uncoating.

EXPERIMENTAL PROCEDURES

Protein expression, purification, and crystallization: The gene encoding single-chain (VH-linker-VL) antibody 80R (scFv) was cloned into pET22b (Novagen) containing an N-terminal periplasmic secretion signal pelB, and a thrombin-removable C-terminal 6xHis tag. 80R was overexpressed in BL21(DE3) cells at 30°C for 15 hours with 1 mM IPTG. Protein was purified by Hisbind Ni-NTA (Novagen) column and Superdex 200 gel filtration chromatography (Amersham Biosciences) after thrombin digestion.

The gene encoding S1-RBD (residues 318-510) was cloned into vector pAcGP67A (Pharmingen) containing an N-terminal gp67 secretion signal and a thrombin-cleavable C-terminal 6xHis tag. It was expressed in sf9 cells (Invitrogen) with Multiplicity of Infection=5 for 72 hours. Similar to 80R, S1-RBD was purified from the media with Hisbind Ni-NTA and Superdex 200 columns, with thrombin digestion. N-linked glycosylation was removed by incubation with PNGase F (New England Biolab) at 23°C, as monitored by SDS-PAGE. S1 RBD-80R complexes were formed by mixing the two purified components, and isolated by gel filtration with Superdex 200 in 10 mM Tris-HCl, 150 mM NaCl, pH 7.4. Peak fractions were pooled and concentrated to ~7 mg/ml. For S1-RBD crystal growth, the protein was also concentrated to ~7 mg/ml.

Crystals grew by the hanging drop vapor diffusion method at 17°C over ~21 days. For S1-RBD, 2 μl of S1-RBD was mixed with an equal volume of well solution containing 4% w/v polyethylene glycol 4000, 0.1 M sodium acetate, pH 4.6. For the S1-RBD-80R complex, 2 μl of the complex was mixed with an equal volume of well solution containing 12.5% w/v polyethylene glycol 4000, 0.1 M sodium acetate, 0.2 M ammonium sulfate, pH 4.6.

Data collection, structure determination, and refinement - X-ray diffraction data were collected at the National Synchrotron Light Source beam-line X6A and X29A for S1-RBD crystals, and at the Stanford Synchrotron Radiation Laboratory beam-line 11.1, and the Advanced Light Source beam-lines 5.0.3 and 12.3.1 for crystals of the S1-RBD-80R complex. Glycerol (25%) was used as a cryoprotectant in both cases. All the data were processed with DENZO and SCALEPACK, or with the HKL2000 package (8). Crystals of S1 RBD adopt space group $P4_32_12$ with unit cell dimensions a=75.9, c=235.8 (Table I).

Crystals of the S1-RBD-80R complex adopt space group $P2_1$ with unit cell dimensions a=47.5, b=175.9, c=67.6, β=96.6°. The crystals display a lattice-translocation defect in which a fraction of the layers have a translational offset, resulting in periodic sharp and diffuse rows of reflections (Fig. 1). Similar defects were first described by Bragg and Howells (9). Different crystals displayed different degrees of lattice defects, and data merged poorly between crystals. Using a single crystal we were able to collect a data set of good quality with a final R_{MERGE}=0.145 and completeness of 93.8% to 2.3 Å resolution. Processing the data required careful optimization of integration profiles and the imposition of a fixed mosaicity (0.45°). Correlation between the offset layers caused the appearance of a strong off-origin peak (65% of the origin) in the native Patterson map at (1/3, 0, 0), indicating that the dislocation occurred along the a* direction.

Additional features of the Patterson map were visible at approximately 1/10 of the origin peak, and provided a measure of the severity of the defect among different crystals. The averaged intensity for the layers of reflections showed a periodic variation that corresponded to the sharp and diffuse layers, and we used the procedure developed by Wang et al. (10) to correct for the intensity modulation (Fig. 2). We calculated average intensities for individual h layers and applied a correction to the intensities using the formula:

$$I_{COR} = I_{MEAS}/(A+B\cos(2\pi h\Delta x))$$

where A and B were obtained by least square fitting of the averaged measured intensities. The ratio of the parameters B and A (B/A=0.65) coincided with the height ratio of the Patterson peak at (1/3, 0, 0), as required by the lattice-translocation theory presented by Wang. The corrected intensity distribution (Fig 2b) was used for the structure solution and the refinement.

The structure of the S1-RBD-80R complex was determined using the Joint Center for Structural Genomics molecular replacement pipeline (11) which employs a modified version of MOLREP (12), and independently using PHASER (13), with the S1-RBD domain from the S1-RBD-ACE2 complex and the scFv domain from the scFv-turkey egg-white lysozyme complex (PDB code 1DZB) as search models. The asymmetric unit contains two molecules of S1 RBD-80R. The final model includes residues 318 to 505 (molecule 1) and 319 to 509 (molecule 2) of S1 RBD and residues 1 to 245 (molecule 1) and 1 to 244 (molecule 2) of 80R, and 470 water molecules. No electron density is observed for the artificial poly-Gly/Ser inter-domain linker. Initial solutions from molecular replacement were subjected to several rounds of refinement with the program REFMAC5 (14) with simulated annealing in CNS (15) and manual model rebuilding with programs O (16) and Coot (17).

The structure of uncomplexed S1-RBD (which showed no lattice defects) was determined by molecular replacement with PHASER (13) using S1-RBD from the structure of the S1-RBD-ACE2 complex (PDB code 2AJF) as the search model. The asymmetric unit contains two molecules of S1-RBD arranged as a symmetric dimer. The final model includes residues 320 to 503 of both monomers and 152 water molecules.

Geometric parameters are excellent as assessed with PROCHECK (18) (Table I). Final R_{WORK}/ R_{FREE} values are 18.2/21.3 and 24.8/29.5 for uncomplexed S1-RBD and the S1-RBD-80R complex, respectively. The higher R values for the S1 RBD-80R complex can likely be explained by the limitations of the lattice defect model and the integration of weak, elongated spots, as previously discussed (10). Notwithstanding, the final electron density map for the S1 RBD-80R complex is of excellent quality (Fig. 3), and the model-to-map correlation is above 0.9 for most of the residues at 2.3 Å resolution. Coordinates have been deposited in the Protein Data Bank with codes 2GHV (S1-RBD) and 2GHW (S1-RBD-80R complex).

RESULTS AND DISCUSSION

Structure of the S1-RBD-80R complex - We determined the crystal structure of the S1-RBD-80R complex at 2.3 Å resolution (Figs. 3 and 4, Table I). The S1 RBD has a very similar structure to that in the ACE2 complex (Fig. 4b). The complex interface involves all 6 antibody complementarity-determining region (CDR) loops, which protrude into the concave surface on the S1 receptor binding motif. Chothia and colleagues (19,20) showed that there exists only a small repertoire of main-chain conformations for 5 of the 6 CDR loops (excluding H3, which is often long, and highly variable in structure), and that these structures can be predicted from their amino acid sequences. For the CDR loops of 80R, loops L2, L3, H1 and H2 adopt main chain conformations close to those predicted by Chothia. However, the L1 loop is atypical, although similar to that of the anti-HIV-gp41 antibody (code 1DFB). The H3 loop is short and well-ordered. A loop that is classically considered part of the framework (between β strands D and E) (Fig. 4c, Table II) also plays a major role in the interface: the “extended loop” (5) of S1 wraps around this framework loop making multiple contacts.

Although ACE2 employs a different recognition mode (dominated by a helix that lines the concave surface of S1) the 80R epitope on S1 overlaps very closely with the ACE2 binding surface (Fig. 4b). Thus, of the 29 residues (between 426 and 492) on S1 that contact 80R, 17 of these also make interactions in the S1-ACE2 interface. The S1-80R interface buries ~2200 Å^2 of protein surface, compared with ~1700 Å^2 for S1-ACE2. The "gap-volume", a measure of shape complementarity (21), is ~4000 Å^3 for the S1-80R interface, about half that of S1-ACE2 (~7000 Å^3). The larger buried surface and smaller gap-volume provide a rationale for the stronger binding and neutralizing activity of the antibody.

The structure provides a rationale for previous mutagenesis studies. Thus, residue N479 is involved in both interfaces, and, accordingly, mutations decrease both 80R binding and ACE2 binding (2-10-fold (7,22)). Two further mutational sites that reduce antibody binding, at D454 and E452, are not directly involved in the interface; their effect can be explained by the participation of these acidic residues in a salt-bridge network that anchors the receptor/antibody binding interface to the S1 RBD core and the extended loop that wraps around the framework hairpin (Fig. 4c). One key difference between the two interfaces lies in the role of S1 residue D480: thus, D480A or D480G mutations completely abolish binding to 80R, but have no effect on ACE2 binding (7). Consistently, D480 lies at the heart of the S1-80R interface, making an intermolecular salt bridge to R162 (see Supplementary Fig. 1) and an H-bond to N164 of 80R, while D480 makes no contacts in the S1-ACE2 complex. Binding of S1 RBD to either ACE2 or 80R is independent of glycosylation (23,24); accordingly, all three potentially glycosylated asparagines (N318, N330, N357) in the S1 RBD are remote from the binding interfaces.

Structure of the uncomplexed S1-RBD - We also determined the crystal structure of the uncomplexed S1-RBD (residues 318 to 510) at 2.2 Å resolution (Figs. 5 and 6, Table I). Compared with its structure in complex with either 80R or ACE2, the receptor-binding surface, including the extended loop, is essentially identical to its structure in the complexes. However, there are extensive rearrangements and increased ordering of the region distal to the 80R/ACE2 binding surface (Fig. 5), which lead to the formation of an extensive dimer interface with a buried surface area of ~2200 Å^2 (Fig. 6). The major reorganization occurs in three structural elements (secondary structure nomenclature as in (5)): (i) the loop between strands 2 and 3 containing Helix B reorganizes such that the new helix B is one turn longer and lies orthogonal to its position in the complexed structures; (ii) Helix B from the neighboring monomer packs tightly across the dimer interface, causing Helix A to shift by 10-12 Å to a new position adjacent to the C-terminus; and (iii) the C-terminus also undergoes a small concerted shift (~ 4 Å). The dimer is formed by the pairing of the β sheets (via their β2 strands) and B helices from each monomer and is largely hydrophobic in nature.

This dimer is predicted by the DCOMPLEX server (phyyz4.med.buffalo.edu/czhang/complex.html) to be physiologically relevant, and to have a binding energy comparable with the S1-80R and S1-ACE2 binary complexes. In agreement with the structural data, gel filtration studies of S1 RBD indicate a monomer-dimer equilibrium in solution at μM concentrations (data not shown). Of note, it has been reported that the murine hepatitis coronavirus S1 domain also exists as a stable dimer (25).

The C-termini of S1 RBD's lie on the "lower" surface of the dimers (Fig. 6a), topologically consistent with their connection to the membrane-spanning S2 domain. At the lower surface of the dimer interface, two cysteine residues, from apposing B helices (Cys 378), come into close proximity (Sγ-Sγ distance = 3.2 Å), but do not form a disulfide bond. We propose a model in which S1 dimers present two preformed receptor-binding motifs pointing outward from the viral membrane surface. A plausible role for the S1 dimers is to cross-link S protein trimers (which trimerize via their S2 domains) on the viral surface (26), thus contributing to the structural integrity of the virion. Modeling two ACE2 receptors onto the S1 dimer leads to steric clashes between the receptors (Fig. 6c), which could explain why S1 is monomeric in crystals of the S1-ACE2 complex. Interestingly, in the S1-80R complex, S1 dimers still form, and in this case the two Cys378 residues remain in close apposition. However, the monomers are twisted with respect to their positions in the uncomplexed S1 dimers, and the hydrophobic interface is largely disrupted. In silico modeling of two 80R fragments onto the uncomplexed S1 dimer does not lead to steric clashes, and in this case the dimer rearrangement is presumably driven by competing lattice forces.

These observations raise the intriguing hypothesis that binding of multiple receptors *in vivo* promotes disruption of S protein dimers, perhaps in a redox-dependent fashion, thus priming S for subsequent membrane fusion events mediated by the S2 domains. A role for receptor-promoted viral uncoating is well established in the (non-enveloped) picornaviruses (27), and has also been described for the Env protein of avian leukosis virus (28). Clearly, further experiments are required to explore this hypothesis.

Prospects for Immune Therapy – Marasco, Farzan, Sui and colleagues (22) previously demonstrated that 80R IgG can neutralize all SARS-CoVs and SARS-like CoVs that evolved during the 2002/2003 outbreak. Because the 80R epitope on S1 overlaps so closely with the ACE2 binding site, we suggest that, for most residues on S1 at the binding site, antigenic drift on S that makes 80R ineffective is likely to abolish binding to ACE2 as well. A notable exception is the D480G mutation, which was found in the SARS-like CoVs from civet cats during the 2003-2004 winter season. These CoVs were likely responsible for an independent interspecies transmission that resulted in the infection of four patients in a mini 2003-2004 outbreak (29). The 80R antibody does not bind these mutants, as noted above; however, these viruses were also less pathogenic and no cases of human-to-human transmission were reported.

By establishing the susceptibility and resistance profiles of newly emerging SARS-CoVs through early S1 genotyping of the neutralizing epitope of 80R, which we have now mapped in atomic detail, an effective immunotherapeutic strategy with 80R should be possible in a future SARS outbreak. In this setting, administration of 80R IgG would provide immediate protection for individuals; subsequently, the innate immune response would take effect, resulting in reduced virus titers and "superspreader" events, crucial for effective containment of the disease.

REFERENCES

1. Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S., Abu-Raddad, L. J., Ho, L. M., Thach, T. Q., Chau, P., Chan, K. P., Lam, T. H., Tse, L. Y., Tsang, T., Liu, S. H., Kong, J. H., Lau, E. M., Ferguson, N. M., and Anderson, R. M. (2003) *Lancet* **361**(9371), 1761-1766
2. Holmes, K. V. (2003) *J Clin Invest* **111**(11), 1605-1609
3. Xu, Y., Lou, Z., Liu, Y., Pang, H., Tien, P., Gao, G. F., and Rao, Z. (2004) *J Biol Chem* **279**(47), 49414-49419
4. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003) *Nature* **426**(6965), 450-454
5. Li, F., Li, W., Farzan, M., and Harrison, S. C. (2005) *Science* **309**(5742), 1864-1868
6. Sui, J., Li, W., Murakami, A., Tamin, A., Matthews, L. J., Wong, S. K., Moore, M. J., Tallarico, A. S., Olurinde, M., Choe, H., Anderson, L. J., Bellini, W. J., Farzan, M., and Marasco, W. A. (2004) *Proc Natl Acad Sci U S A* **101**(8), 2536-2541
7. Sui, J., Li, W., Roberts, A., Matthews, L. J., Murakami, A., Vogel, L., Wong, S. K., Subbarao, K., Farzan, M., and Marasco, W. A. (2005) *J Virol* **79**(10), 5900-5906
8. Otwinowski, Z., and Minor, W. (1997) *Methods in Enzymology* **276**, 307-326
9. Bragg, W. L., and Howells, E. R. (1954) *Acta Crystallogr.* **7**, 409-411

10. Wang, J., Kamtekar, S., Berman, A. J., and Steitz, T. A. (2005) *Acta Crystallogr D Biol Crystallogr* **61**(Pt 1), 67-74
11. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K., and Jaroszewski, L. (2004) *Acta Crystallogr D Biol Crystallogr* **60**(Pt 7), 1229-1236
12. Vagin, A., and Teplyakov, A. (1997) *J. Appl. Cryst.* **30**, 1022-1025
13. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) *Acta Crystallogr D Biol Crystallogr* **61**(Pt 4), 458-464
14. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr D Biol Crystallogr* **53**(Pt 3), 240-255
15. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-921
16. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjelgaard, M. (1991) *Acta Crystallogr.* **A47**, 110-119
17. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1), 2126-2132
18. Laskowski, R. J., MacArthur, N.W., Moss, D.S., and Thornton, J.M. (1993) *Journal Applied Crystallography* **26**, 283-290
19. Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B., and Winter, G. (1992) *J Mol Biol* **227**(3), 799-817
20. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., and et al. (1989) *Nature* **342**(6252), 877-883
21. Jones, S., and Thornton, J. M. (1996) *Proc Natl Acad Sci U S A* **93**(1), 13-20
22. Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S. K., Huang, I. C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W. A., Guan, Y., Choe, H., and Farzan, M. (2005) *Embo J* **24**(8), 1634-1643
23. Sui J, L. W., Murakami A, Tamin A, Matthews LJ, Wong SK, Moore MJ, Tallrico AS, Olurinde M, Choe H, Anderson LJ, Bellini WJ, Farzan M, Marasco WA. (2004) *Proc Natl Acad Sci USA* **101**(8), 2536-2541
24. Chakraborti, S., Prabakaran, P., Xiao, X., and Dimitrov, D. S. (2005) *Virol J* **2**, 73
25. Lewicki, D. N., and Gallagher, T. M. (2002) *J Biol Chem* **277**(22), 19727-19734
26. Lin, Y., Yan, X., Cao, W., Wang, C., Feng, J., Duan, J., and Xie, S. (2004) *Antivir Ther* **9**(2), 287-289
27. Nurani, G., Lindqvist, B., and Casasnovas, J. M. (2003) *J Virol* **77**(22), 11985-11991
28. Mothes, W., Boerger, A. L., Narayan, S., Cunningham, J. M., and Young, J. A. (2000) *Cell* **103**(4), 679-689
29. Song, H. D., Tu, C. C., Zhang, G. W., Wang, S. Y., Zheng, K., Lei, L. C., Chen, Q. X., Gao, Y. W., Zhou, H. Q., Xiang, H., Zheng, H. J., Chern, S. W., Cheng, F., Pan, C. M., Xuan, H., Chen, S. J., Luo, H. M., Zhou, D. H., Liu, Y. F., He, J. F., Qin, P. Z., Li, L. H., Ren, Y. Q., Liang, W. J., Yu, Y. D., Anderson, L., Wang, M., Xu, R. H., Wu, X. W., Zheng, H. Y., Chen, J. D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L. Y., Li, H., Chen, F., Di, B., He, L. J., Lin, J. Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X. J., Spiga, O., Guo, Z. M., Pan, H. Y., He, W. Z., Manuguerra, J. C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y. X., Wu, C. I., and Zhao, G. P. (2005) *Proc Natl Acad Sci U S A* **102**(7), 2430-2435

FOOTNOTES

*We thank V. Stojanoff and J. Jakoncic at beamline X6A and H. Robinson and G. Shea-McCarthy at beamline X29A (NSLS), and K. Frankel at beam-lines 12.3.1 and C. Trame at 5.0.3 (ALS) for help with data collection, and C. Bakolitsa and L. Bankston for helpful discussion. This work is supported by NIH grants to RCL (DAMD17-03-2-0038) and WAM (AI28785, AI48436, AI061318 and AI053822). [1]The abbreviations used are: SARS, Severe Acute respiratory Syndrome; CoV, coronavirus; ACE2, angiotensin-converting enzyme 2; RBD, receptor-binding domain; CDR, complementarity-determining region.

Table I. Data collection and refinement statistics

	S1-RBD	S1-RBD-80R
Data Collection		
Cell parameters	a = 75.9, c = 235.9 Å	a=47.5, b=175.9, c=67.6 Å; β=96.6°
Space group	$P4_32_12$	$P2_1$
Resolution (Å)	2.2	2.3
Total reflections	233011	159047
Unique reflections	36036	51915
Completeness (%)*	99.9 (99.9)	93.8 (87.0)
Average I/σ(I)*	24.7 (2.0)	8.8 (1.9)
R_{MERGE}*	0.098 (0.739)	0.145 (0.571)
Redundancy	6.5	3.1
Refinement		
R_{WORK}#	0.182 (0.230)	0.248 (0.301)
R_{FREE} (5% data)#	0.213 (0.289)	0.295 (0.391)
RMSD bond distance (Å)	0.013	0.009
RMSD bond angle (°)	1.49	1.22
Average B value	50.0	37.1
Solvent atoms	152	470
Ramachandran plot		
Residues in most favored regions	276	631
Residues in additional allowed regions	35	81
Residues in generously allowed regions	3	5
Residues in disallowed regions	0	0

*Numbers in parentheses correspond to the highest resolution shell (2.28-2.20 Å for S1 RBD; 2.29-2.38 Å for S1 RBD-80R).

#Numbers in parentheses correspond to the highest resolution shell (2.26-2.20 Å for S1 RBD; 2.29-2.38 Å for S1 RBD-80R).

Table II. Contact residues between 80R and S1 RBD.

CDR H1			CDR H2						CDR H3					FRL1	CDRL1		
S31	Y32	A33	V50	I51	S52	Y53	N57	Y59	D99	R100	S101	Y102	D105	R150	V161	R162	S163
T487	Y491	T486	T486	T486	T486	R426	R426	S432	T487	Y491	Y436	S432	Y491	L472	D480	Y436	L478
G488		T487				T485	T485	T485	Y491		G482	Y436				N437	N479
I489		G488				T486	T486	T486			Y484	Y484				K439	D480
						T487					T487	T486				D480	
						G488					Y491	T487					
						Q492											

CDR L1	CDR L2				FR L3										CDR L3		
N164	D182	S184	T185	R186	S195	G196	S197	G198	S199	D202	F203	T204	T206	S208	R223	S224	W226
Y436	Y440	Y442	Y442	Y475	L472	C474	P469	W436	W476	P470	P470	P470	L472	L472	Y436	Y436	S432
D480	Y442	Y475	N479		N473	Y475	P470	P469				A471			D480	Y484	T433
	N479				Y475		A471	P470							Y481		Y484
	D480						C474	C474							G482		
							Y475								Y484		

80R residues are listed on the top line and grouped under CDR or Framework Region (FR). S1-residues in contact with 80R residues are listed in subsequent lines. Hydrogen-bonded residues are in red. S1-ACE2 and S1-80R interfaces share many common S1 residues, except for 5 residues (404, 443, 460, 462, 463) which are found only at the S1-ACE2 interface, while 12 residues (433, 437, 439, 469, 470, 471, 474, 476, 478, 480, 485, 492) are found only at the S1 - 80R interface.

FIGURE LEGENDS

Fig. 1 Diffraction patterns of complex crystal. The complex crystals display a lattice-translocation defect caused by translocations in the crystal packing between neighboring layers along the a* direction. *Panel a,* a* is nearly vertical, in the plane of the paper, and the defect results in periodic sharp-diffuse-diffuse rows of diffraction intensities (the bottom left quadrant is a zoom-in of the boxed area). *Panel b,* a* is nearly parallel to the X-ray beam, perpendicular to the paper, and the defect is not evident.

Fig. 2 *h* layer intensities before and after correction. *Panel* a, The lattice defect results in a strong-weak-weak pattern of intensities along h, which were corrected *(Panel b)* according to the procedure of Wang et al. (10).

Fig. 3: **Stereo 2Fo-Fc electron density map of the S1-RBD-80R complex at the S1-80R interface.** S1 and 80R residues are shown in red and blue, respectively, with selected residues labeled. Contour level = 1.5 σ.

Fig. 4. Structure of the S1-RBD-80R complex. *Panel a,* Overall structure of the complex. Antibody variable region light chain is in blue and heavy chain is in magenta; S1-RBD is in red. *Panel b,* Comparison between the S1 RBD-80R complex (red and yellow) and the S1 RBD-ACE2 complex (blue and green) overlaid on the S1-RBD domain. *Panel c,* Close-up of the interface. Selected S1 side-chains are in red; 80R in blue; hydrogen bonds in cyan. CDR's (L1-L3, H1-H3) and the framework (FW) loop (interacting with the "extended loop" of S1) are labeled. There is an aromatic ring stacking between Y484 (S1) and Y102 (80R). Y484 and Y102 are in turn coordinated by hydrogen bonds between T486 (S1) and Y102 (80R) and Y53 (80R), and Y484 (S1) and Y436 (S1), respectively. Another intermolecular hydrogen bond occurs between L478 (S1) and S163 (80R). N164 (80R) makes intramolecular hydrogen bonds with R223 (80R). Intramolecular hydrogen bonding between Y103 (80R) and D182 (80R) may be important for maintaining the 80R structure at the interface and may be important for S1 RBD-80R binding. C474 (S1), A471 (S1) and S197 (80R) form another intermolecular hydrogen bonds that may stabilize the S1 RBD-80R interface.

Fig. 5: **Stereo comparison of the S1-RBD domain.** Uncomplexed (dimeric) S1-RBD is in red, complex with 80R antibody is in green; complex with ACE2 is in blue. Helical elements A and B, and the C-terminus, are labeled. The receptor-binding surface, including the "extended loop", is highly conserved in all three structures, lies at the back of the field and is not visible in this view. RMSD values for pairwise comparisons are 0.9-1.1 Å for main chain residues excluding helix A (residues 350-360) helix B (370-381), and N- and C- termini before residue 323 or after residue 502. The small differences in these regions of the complexed S1-RBD's presumably arise from the different crystal environments. The large changes in the uncomplexed S1-RBD are a consequence of dimer formation.

Fig. 6 Structure of the S1-RBD dimer. *Panel a,* S1 monomers are in red and blue, related by a vertical two-fold axis. The receptor binding surfaces and C-termini are indicated. *Panel b,* Same as in *a* but rotated by 90° about a horizontal axis to show the molecular dyad. *Panel c,* Hypothetical model of the S1-RBD dimer with two molecules of ACE2 bound, showing steric overlap (circled). The view is rotated about a vertical axis compared with *a* in order to minimize the overlap in projection. Two Fab fragments can bind the dimer without steric hindrance (not shown). A full-length dimeric antibody could presumably cross-link neighboring dimers on the viral surface, but the geometry is inappropriate for binding both sites on a single dimer simultaneously.

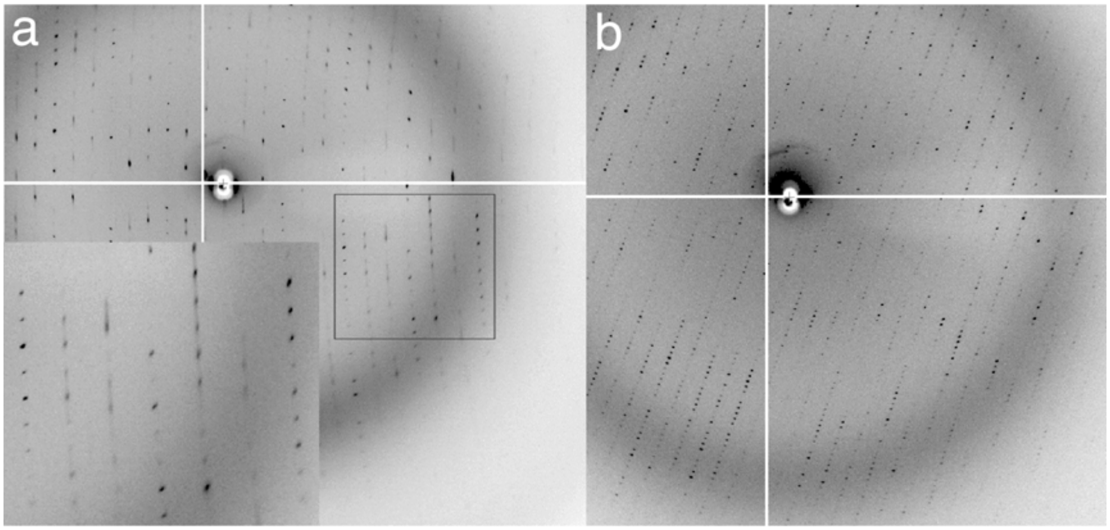


Figure 1: Hwang et al.

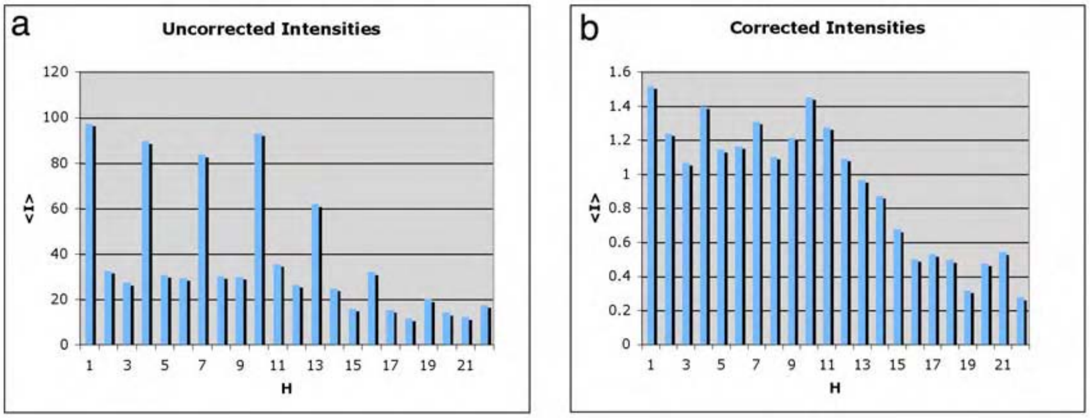


Figure 2: Hwang et al.

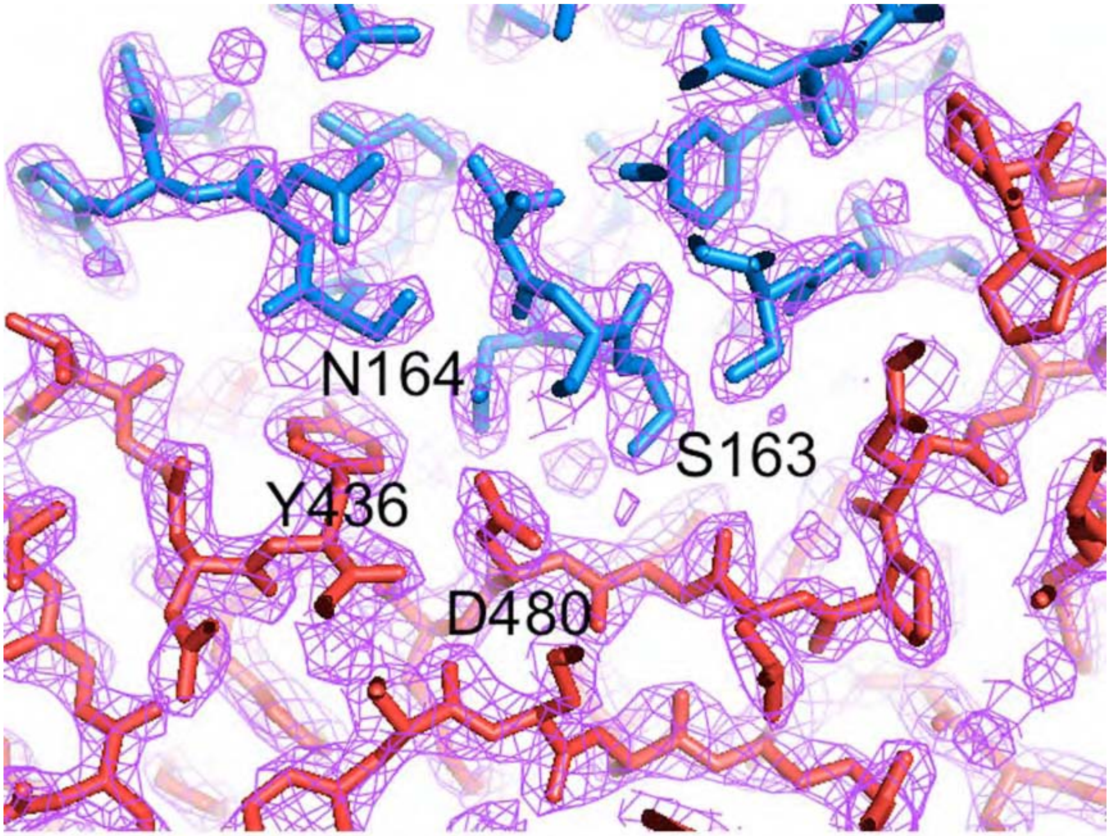


Figure 3: Hwang et al.

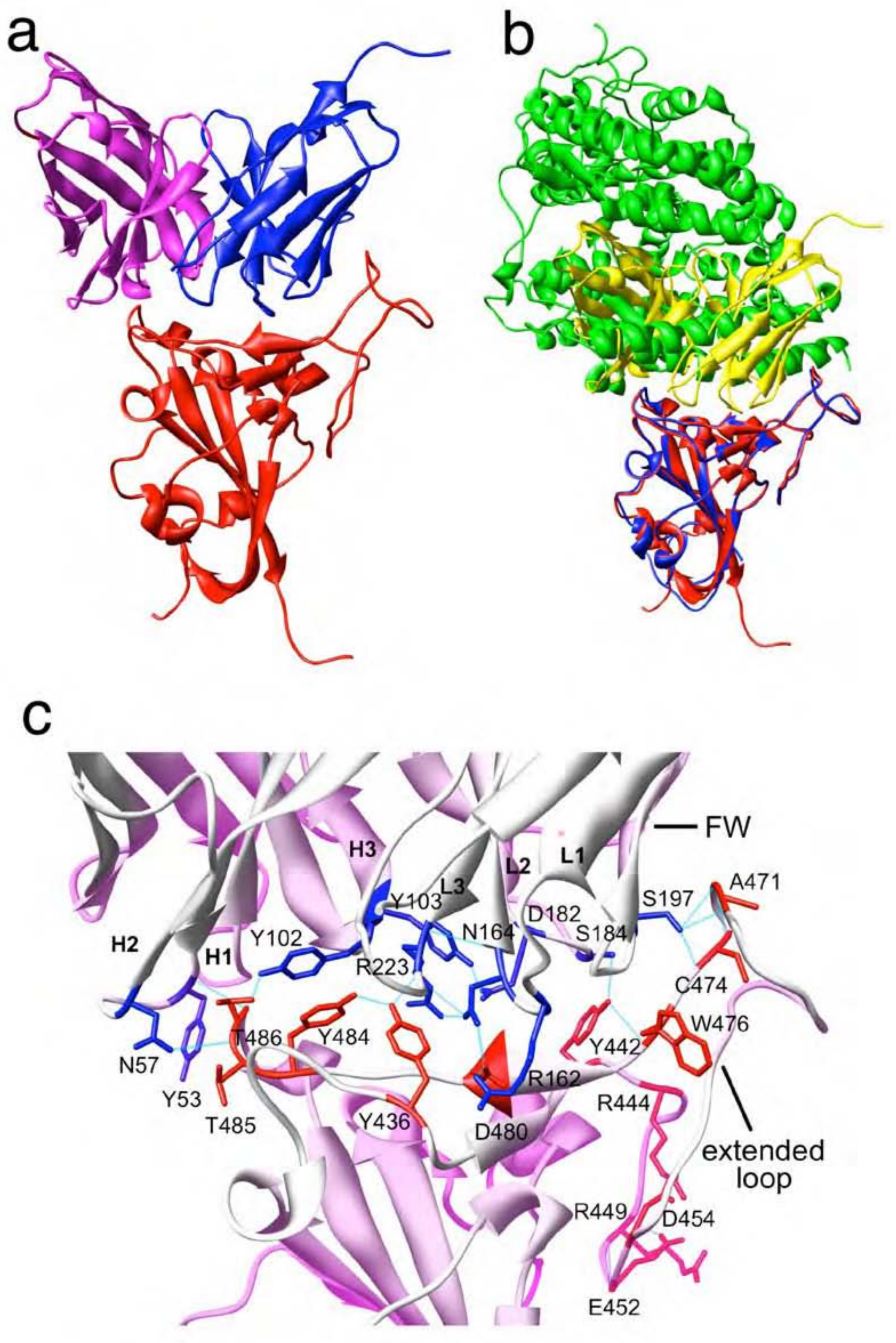


Figure 4: Hwang et al.

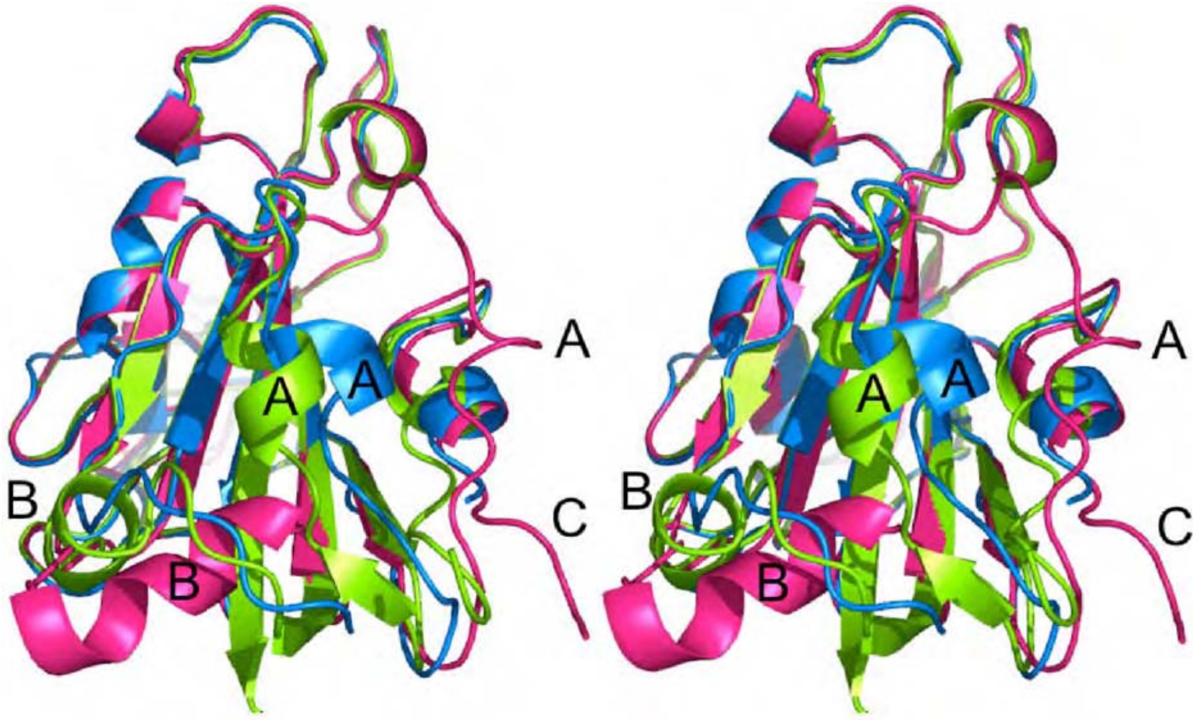


Figure 5: Hwang et al.

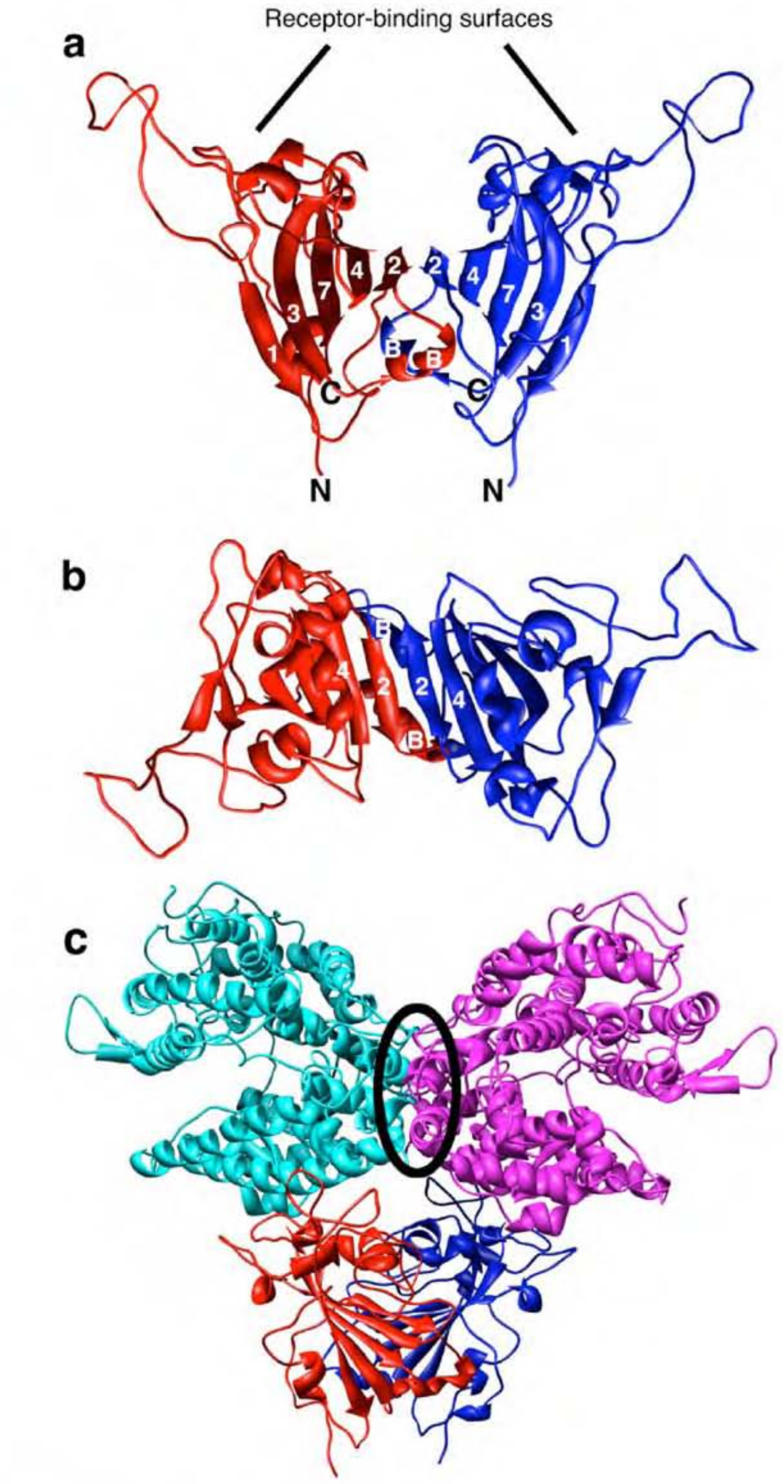


Figure 6: Hwang et al.

Protein Science (in Press 8/06)

Vaccinia virus N1L protein resembles a B cell lymphoma-2 (Bcl-2) family protein

Mika Aoyagi, Dayong Zhai, Chaofang Jin, Alexander Aleshin, Boguslaw Stec, John C. Reed and Robert C. Liddington

Infectious and Inflammatory Disease Center, Burnham Institute for Medical Research, La Jolla, California 92037, U. S. A.

Corresponding author: Robert C. Liddington, Infectious and Inflammatory Disease Center, Burnham Institute for Medical Research, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA; e-mail: rlidding@burnham.org; tel: 858-646-3136; fax: 858-713-9925

Running title: Structure of vaccinia virus N1L protein

Total manuscript pages: 25 (including 1 table and 3 figures); no supplementary data

Abstract

Poxviruses encode immuno-modulatory proteins capable of subverting host defenses. The poxvirus, vaccinia, expresses a small 14 kDa protein, N1L, that is critical for virulence. We report the crystal structure of N1L, which reveals an unexpected but striking resemblance to host apoptotic regulators of the B cell lymphoma-2 (Bcl-2) family. Although N1L lacks detectable Bcl-2 homology motifs at the sequence level, we show that N1L binds selectively to pro-apoptotic Bcl-2 family proteins in vitro, consistent with a role for N1L in modulating host antiviral defenses.

Keywords: Poxvirus/vaccinia virus/virulence/crystal structure/Bcl-2/apoptosis

Poxviruses, such as vaccinia and variola (smallpox), are among the largest animal viruses, carrying a linear double-stranded DNA genome (150-350 kb) with ~200 distinct genes (Moss 2000). Poxviruses express their own machinery for DNA replication, mRNA transcription and virion assembly (Moss 2000). They also encode proteins that manipulate host defense mechanisms for efficient viral replication (Johnston and McFadden 2003; Seet et al. 2003; Shchelkunov 2003).

A 14 kDa vaccinia protein, N1L, was initially identified from an attenuated spontaneous deletion mutant (6/2) of vaccinia virus (Kotwal and Moss 1988). N1L is a potent virulence factor, which when deleted caused the strongest attenuation observed for any gene that was not essential for growth in culture (Kotwal et al. 1989; Bartlett et al. 2002). Thus, deletion of the N1L gene reduced mortality of intracranially infected mice by a factor of 10^4 (Kotwal et al. 1989). Furthermore, in the highly attenuated vaccinia Ankara strain, N1L is truncated with a distinct C-terminus (Antoine et al. 1998). Although initially described as a secreted "virokine" (Kotwal et al. 1989), N1L is now believed to localize predominantly within the host cell (Bartlett et al. 2002). N1L has 94% sequence identity between vaccinia and variola orthologs (Massung et al. 1993), but appears to be unique to poxviruses (Bartlett et al. 2002).

Understanding the molecular mechanisms of viral immuno-modulatory proteins furthers our insights into the delicate interplay between pathogen and host, illuminates pathways of cellular immunity, and provides new leads for the development of antiviral therapeutics and vaccines. Towards these goals, we report here the crystal structure of N1L, which reveals a compact α-helical architecture characteristic of the Bcl-2 family of host cell apoptotic regulators. *In vitro* binding studies demonstrate binding to several cellular pro-apoptotic BH3 domains, suggesting a direct role for N1L in the modulation of host cell apoptotis.

Results and Discussion

Bcl-2-like structure of vaccinia N1L

We determined the crystal structure of vaccinia N1L at 2.2Å resolution (Table 1). The crystals contain six molecules in the asymmetric unit arranged as 3 symmetric dimers. Conformational heterogeneity occurs in an N-terminal loop (Asn13-Phe24) and at the C-terminus (Leu109-Gly115). Otherwise, the six copies are very similar, with RMS main-chain deviations of ≤ 0.7 Å in pair-wise comparisons. The refined models include N1L residues 1-114, with additional residues at the N-terminus from the expression vector (Ser-1 and His0); the three C-terminal residues (Gly115-Lys117) have not been modeled owing to poor or absent electron density.

N1L forms a compact α-helical bundle (Fig. 1A). The N-terminal helix, $\alpha1$, is connected by a short loop to the $\alpha2$ with an interhelical angle of ~80°. The last five residues of $\alpha2$ (Leu29-Leu33) form a 3_{10} helix followed by a short turn that orients $\alpha3$ at ~100° from $\alpha2$. Three helices, $\alpha4$, $\alpha5$ and $\alpha6$, are nearly antiparallel to each other. The C-terminal helix $\alpha6'$ (Glu103-Leu113) is contiguous with $\alpha6$, except that a single 3_{10} helical turn at Glu103 creates a bend, rotating $\alpha6'$ clockwise by 80° with respect to $\alpha6$. $\alpha6'$ is positioned almost perpendicular to the central $\alpha5$ helix. In the overall organization, the two central $\alpha5$ and $\alpha6$ helices are surrounded by two helices ($\alpha1$ and $\alpha2$) on one side, and two helices ($\alpha3$ and $\alpha4$) on the other.

The N1L fold closely resembles that of the Bcl-2 family of cellular apoptotic regulators (Petros et al. 2004), despite a very low sequence identity of 11% (Fig. 1B, C). The DALI server (Holm and Sander 1993) identifies several anti-apoptotic Bcl-2 family proteins as the closest structural neighbors of N1L: mouse myeloid cell leukemia-1 (Mcl-1, PDB code 1WSX (Day et al. 2005); Z-score=9.6, RMSD=2.8Å), Kaposi sarcoma virus Bcl-2 homolog (1K3K (Huang et al. 2002); Z=7.9, RMSD=3.2Å), human Bcl-X_L (1MAZ (Muchmore et al. 1996); Z=7.0, RMSD=3.1Å), *C. elegans* Bcl-2 protein CED-9 (1OHU (Woo et al. 2003); Z=6.5, RMSD=3.7Å), and mouse Bcl-X_L (1PQ0 (Liu et al. 2003); Z=6.4, RMSD=4.2Å). In contrast to the Bcl-2 family members, N1L lacks a C-terminal transmembrane helix (Petros et al. 2004), and in general contains shorter secondary structural elements (Figs. 1B & 1C). Indeed, vaccinia N1L is the smallest known protein that maintains the Bcl-2-like fold.

Dimeric assembly of N1L

The N1L crystal structure reveals a homo-dimeric assembly (Fig. 2A) distinct from the monomeric structures reported for host Bcl-2 family members (Petros et al. 2004). Dimerization buries 2100 Å^2 of surface, accounting for 30% of the total (6900 Å^2) of each subunit. The DCOMPLEX server (Zhou et al. 2005) (phyyz4.med.buffalo.edu/czhang/complex.html) predicts the N1L dimer to be biologically relevant (rather than a crystallization artifact). Gel filtration analysis also suggests that N1L is dimeric in solution at μM concentrations (data not shown), consistent with earlier biochemical studies (Bartlett et al. 2002).

Molecular contacts at the dimer interface are provided by the α1 and α6 helices (Fig. 2A & B). Bulky hydrophobic α1 residues, Ile6 and Leu10, pack against their counterparts across the dimer interface. In a similar manner, charged α1 residues, Arg7 and Asp14, of one subunit interact with their counter-ions in the second subunit. α6 also provides complementary hydrophobic (Phe95 and Phe99) and charged (Arg90 and Glu103) residues across the dimer interface. Comparable hydrophobic and charged residues are absent in other Bcl-2 family proteins (Fig. 2C). Notably, this antiparallel N1L homodimer is distinct from a recently described Bcl-X_L dimer, in which C-terminal halves are swapped between two monomers by formation of a single continuous α5-α6 helix (O'Neill et al. 2006). However, as in the case of the domain-swapped Bcl-X_L dimer (O'Neill et al. 2006), the dimer interface of N1L excludes a putative functional face of the molecule, namely the hydrophobic binding groove (Petros et al. 2004) prominent among the Bcl-2 family proteins.

Bcl-2 homology (BH) motifs

The Bcl-2 family of proteins contain at least one of the four “Bcl-2 homology” (BH1-4) regions (Fig. 1B, C) that structurally and functionally support their regulatory roles in apoptosis (Cory and Adams 2002; Danial and Korsmeyer 2004; Petros et al. 2004). Structure-based alignment (Fig. 1C) demonstrates a lack of apparent sequence homology of N1L in regions structurally equivalent to the BH domains. Nevertheless, several key BH-domain interactions appear to be maintained in N1L. For example, an “NIED” sequence found at the beginning of α5 in N1L serves the same structural role as the “NWGR” signature motif of the Bcl-2 BH1 domain. In both N1L and Bcl-2 family proteins, the conserved Asn (Asn65 in N1L and Asn136 in Bcl-X_L) at the first position of the motif N-terminally caps the central helix α5. In addition, analogous to the Trp residue at the second position (Trp137 in Bcl-X_L), Ile66 of N1L forms hydrophobic contacts with α6/α6’ residues (Tyr105 and Leu109), possibly contributing to the overall structural integrity (Huang et al. 2002). On the other hand, the last two residues in the “NWGR” motif, Gly and Arg, which are crucial for protein-protein interactions among the Bcl-2 related proteins (Sattler et al. 1997), are replaced by Glu67 and Asp68 in N1L.

Although N1L lacks consensus BH motifs, its molecular surface contains an elongated hydrophobic patch comparable to that found in the anti-apoptotic Bcl-2 family proteins. In these Bcl-2 proteins, α5 (BH1), α7 (BH2), α2 (BH3), α3 and α4 form a long hydrophobic groove (Fig. 3A) where the BH3 region from another Bcl-2 protein binds to form a hetero-dimer (Yin et al. 1994; Sattler et al. 1997; Liu et al. 2003). For N1L, a hydrophobic groove is located on the same face of the molecule as in the Bcl-2 family proteins (Fig. 3B), but N1L’s groove is narrower and shorter owing to additional charged residues (Glu32, Asp35, Asp38, Glu67, Asp68 and Arg71) and the closer packing of α2 against α5.

N1L binds BH3 peptides in vitro

Heterodimerization between pro- and anti-apoptotic Bcl-2 family proteins is a crucial step in regulating apoptosis, and is mediated by the binding of BH3 domains from the pro-apoptotic members to the hydrophobic groove of the anti-apoptotic members (Cory and Adams 2002; Danial and Korsmeyer 2004). We explored potential interactions between N1L and the pro-apoptotic Bcl-2 members using fluorescence polarization assays (Fig. 3C, D). We found that

N1L interacts with peptides comprising the BH3 domains of three different pro-apoptotic Bcl-2 proteins (Bid, Bim and Bak) with affinities similar to those of the anti-apoptotic Bcl-X_L. Curiously, no significant binding was detected between N1L and Bad.

Bcl-2 like proteins in poxviruses

The crystal structure of vaccinia N1L demonstrates the existence of a Bcl-2-like structural fold in the orthopoxviruses. Bcl-2-like proteins have been identified only in fowlpox and canarypox viruses of the avipoxviruses (Afonso et al. 2000; Tulman et al. 2004). Avipoxvirus-encoded Bcl-2 homologs (FPV039 and CNPV058) show sequence homology (~25% identity and ~50% similarity) to cellular Bcl-2 proteins, and contain recognizable BH1 and BH2 domains as well as a C-terminal transmembrane domain (Afonso et al. 2000; Tulman et al. 2004). Owing to the absence of detectable Bcl-2 homologs, most other poxviruses have been assumed to utilize other proteins for controlling host apoptosis (Cuconati and White 2002; Hardwick and Bellows 2003; Taylor and Barry 2006). Vaccinia F1L and myxoma (leporipoxvirus) M11L proteins, for instance, share little sequence homology with Bcl-2 family proteins, yet block apoptosis by inhibiting pro-apoptotic Bak, possibly via their putative BH3-like domain (Wang et al. 2004; Wasilenko et al. 2005; Postigo et al. 2006; Su et al. 2006).

An ortholog search of N1L against other poxviral genomes (www.poxvirus.org) yielded a set of uncharacterized proteins from the distantly related non-orthopoxvirus members. Goatpox, sheeppox and “lumpy skin disease” viruses encode proteins (GTPV_Pellor114, SPPV_A115, and LSDV_WARM144) that share sequence homology (~20% identity and ~50% similarity) with the vaccinia N1L. Elucidating the function of these putative orthologs in modulating host immunity will likely provide insights into the molecular basis of host range and virulence across the poxvirus family.

Functional implication and conclusion

Recent studies have demonstrated that N1L targets several components of a multi-subunit IκB kinase complex in NF-κB signaling pathways (DiPerna et al. 2004) and reduces cytokine secretion (Zhang et al. 2005). How might this finding be linked to our structural and biochemical observations? One observation that may be pertinent is that a cellular Bcl-2 protein, in addition to regulating mitochondrial-mediated apoptosis, also controls the activation of multiple transcription factors, including NF-κB (Regula et al. 2002; Massaad et al. 2004). Interestingly, the cowpox anti-apoptotic protein, CrmA, inhibits NF-κB activation by suppressing the caspase-dependent processing of pro-inflammatory cytokines (Ray et al. 1992), suggesting that apoptotic and NF-κB signaling pathways are linked at the molecular level (Bowie et al. 2004). Our identification of a Bcl-2-like protein in vaccinia with the ability to bind BH3 peptides will thus generate testable hypotheses to probe the molecular mechanisms by which N1L counteracts host antiviral defenses.

Materials and methods

N1L expression, purification and crystallization.

The vaccinia N1L coding sequence (Western Reserve strain VACWR028) was PCR amplified and subcloned into the NdeI/BamHI site on the pET15b vector (Novagen). Recombinant N1L protein (117 amino acids), with an N-terminal His$_6$-tag, was expressed in *Escherichia coli* BL21(DE3) CodonPlus RIL (Strategene) overnight at 15ºC by adding isopropyl-β-D-thiogalactopyranoside. Following cell lysis by sonication, the His$_6$-N1L protein was purified through HiTrap Ni^{2+}-chelating and Superdex 200 gel filtration columns (GE Healthcare Bio-Sciences AB). The protein purity was confirmed by SDS-PAGE and peptide mapping mass spectroscopy, and stored in 20 mM TrisHCl pH 8, 150 mM NaCl, 4 mM β-mercaptoethanol (β-ME) at –80 ºC. Seleno-L-methionine (SeMet)-labeled N1L was prepared using minimal M9 medium under metabolic inhibition, as described elsewhere (Van Duyne et al. 1993). Incorporation of seven SeMet residues (including the first Met residue) per N1L molecule into the protein was confirmed by electrospray mass spectrometry analysis.

All crystallization experiments were performed using the hanging-drop vapor diffusion method at 20ºC. 2 μl of the N1L or SeMet-N1L (30 mg/ml) were mixed with an equal volume of reservoir containing 5-10% (w/v) polyethylene glycol 4000, 100 mM Na-K tartrate, 100 mM TrisHCl pH 8 and 20 mM β-ME. Monoclinic crystals appeared after 1 week and continued to grow over a period of 1-2 months.

Data collection and structure determination.

The SeMet and native data sets (Table 1) were collected from a flash-cooled crystal (100 K) at beam lines 9-2 (Stanford Synchrotron Radiation Laboratory, CA) and 12.3.1 (Advanced Light

Source, CA), respectively. The cryoprotectant solution consisted of the equilibrated crystallization solution augmented with 30% (v/v) 2-methyl-2,4-pentanediol. The diffraction data were processed with HKL2000 (Otwinowski and Minor 1997). Forty-two Se sites - seven sites for each of the six molecules in the asymmetric unit - were identified by SHELXD (Schneider and Sheldrick 2002) and refined using SHARP (de La Fortelle and Bricogne 1997). After density modification by SOLOMON (Abrahams and Leslie 1996), maps calculated to 3Å resolution were used for manual model building using XFIT (McRee 1999) and COOT (Emsley and Cowtan 2004).

The initial model, comprising six copies of residues 1–113, was refined through cycles of model building and refinement using XFIT (McRee 1999) and CNS (Brünger et al. 1998). Rigid body refinement against the 2.2 Å native data, treating the six molecules as separate rigid groups, was followed by cycles of minimization, simulated annealing and *B*-factor refinement, resulting in R_{WORK}=0.296 and R_{FREE}=0.312. Next, the flexible terminal and loop residues, as well as water molecules, were modeled based on $2F_o–F_c$ and $F_o–F_c$ maps, and subjected to further crystallographic refinement without non-crystallographic symmetry restraints, leading to final values of R_{WORK}=0.21 and R_{FREE}=0.25. The model has excellent stereochemistry as defined by PROCHECK (Laskowski et al. 1993) (Table 1). PDBFIT (McRee 1999), CE (Shindyalov and Bourne 1998) and DALI (Holm and Sander 1993) were used to obtain superposition and RMS deviations of the models. Interhelical angles in the final models were calculated using INTERHLX (K. Yap, University of Toronto). The structure factors and coordinates, comprising six copies of N1L and 238 water molecues, have been deposited into the PDB with the accession code XXX.

Fluorescence Polarization Assays (FPAs).

Binding of N1L to the Bcl-2 homology-3 (BH3) domains of several Bcl-2 family proteins was quantified using fluorescence polarization-based peptide binding assays (Zhai et al. 2005). Recombinant human Bcl-X_L, lacking the 20-residue C-terminal transmembrane tail, was prepared as previously described (Zhai et al. 2005), and used as a control. Fluorescein isothiocyanate (FITC)-conjugated synthetic peptides comprising the BH3 domains of pro-apoptotic Bcl-2 proteins (BH3-Bid, FITC-aminohexanoyl (Ahx)-EDIIRNIARHLAQVGDSMDR; BH3-Bim, FITC-Ahx-DMRPEIWIAQELRRIGDEFNAYYAR; BH3-Bak, FITC-Ahx-PSSTMGQVGRQLAIIGDDINRRYDS) were prepared at the Burnham Institute's medicinal chemistry core facility, while the FITC-BH3-Bad peptide (NLWAAQRYGRELRRMSD-K(FITC)-FVD) was purchased from Synpep Corporation, CA. Varying concentrations of N1L and Bcl-X_L were incubated with 5-15 nM of the FITC-BH3 peptides, and the resulting fluorescence polarization (Analyst TM AD assay Detection system, LJL Biosystem) was used to calculate EC_{50} values.

Acknowledgement

We thank the Stanford Synchrotron Radiation Laboratory (supported by the Department of Energy and the National Institutes of Health) and Advanced Light Source (supported by the U.S. Department of Energy) for use of data collection facilities, L, Bankston, G. Wei, L. Low for helpful discussions, and S.-J. Cho, C. Bakolitsa, E. Santelli, Y. Zong for assistance with data analysis. This work was supported by DAMD17-03-2-0038 and NIH U01 AI061139 (to RCL)

References

Abrahams, J.P., and Leslie, A.G. 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52:** 30-42.

Afonso, C.L., Tulman, E.R., Lu, Z., Zsak, L., Kutish, G.F., and Rock, D.L. 2000. The genome of fowlpox virus. *J Virol* **74:** 3815-3831.

Antoine, G., Scheiflinger, F., Dorner, F., and Falkner, F.G. 1998. The complete genomic sequence of the modified vaccinia Ankara strain: comparison with other orthopoxviruses. *Virology* **244:** 365-396.

Bartlett, N., Symons, J.A., Tscharke, D.C., and Smith, G.L. 2002. The vaccinia virus N1L protein is an intracellular homodimer that promotes virulence. *J Gen Virol* **83:** 1965-1976.

Bowie, A.G., Zhan, J., and Marshall, W.L. 2004. Viral appropriation of apoptotic and NF-kappaB signaling pathways. *J Cell Biochem* **91:** 1099-1108.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr*. **D54:** 905-921.

Cory, S., and Adams, J.M. 2002. The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer* **2:** 647-656.

Cuconati, A., and White, E. 2002. Viral homologs of BCL-2: role of apoptosis in the regulation of virus infection. *Genes Dev* **16:** 2465-2478.

Danial, N.N., and Korsmeyer, S.J. 2004. Cell death: critical control points. *Cell* **116:** 205-219.

Day, C.L., Chen, L., Richardson, S.J., Harrison, P.J., Huang, D.C., and Hinds, M.G. 2005. Solution structure of prosurvival Mcl-1 and characterization of its binding by proapoptotic BH3-only ligands. *J Biol Chem* **280:** 4738-4744.

de La Fortelle, E., and Bricogne, G. 1997. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* **276:** 472-494.

DiPerna, G., Stack, J., Bowie, A.G., Boyd, A., Kotwal, G., Zhang, Z., Arvikar, S., Latz, E., Fitzgerald, K.A., and Marshall, W.L. 2004. Poxvirus protein N1L targets the I-kappaB kinase complex, inhibits signaling to NF-kappaB by the tumor necrosis factor superfamily of receptors, and inhibits NF-kappaB and IRF3 signaling by toll-like receptors. *J Biol Chem* **279:** 36570-36578.

Emsley, P., and Cowtan, K. 2004. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126-2132.

Hardwick, J.M., and Bellows, D.S. 2003. Viral versus cellular BCL-2 proteins. *Cell Death Differ* **10 Suppl 1:** S68-76.

Holm, L., and Sander, C. 1993. Protein Structure Comparison by Alignment of Distance Matrices. *J Mol Biol* **233:** 123-138.

Huang, Q., Petros, A.M., Virgin, H.W., Fesik, S.W., and Olejniczak, E.T. 2002. Solution structure of a Bcl-2 homolog from Kaposi sarcoma virus. *Proc Natl Acad Sci U S A* **99:** 3428-3433.

Johnston, J.B., and McFadden, G. 2003. Poxvirus immunomodulatory strategies: current perspectives. *J Virol* **77:** 6093-6100.

Kotwal, G.J., Hugin, A.W., and Moss, B. 1989. Mapping and insertional mutagenesis of a vaccinia virus gene encoding a 13,800-Da secreted protein. *Virology* **171:** 579-587.

Kotwal, G.J., and Moss, B. 1988. Vaccinia virus encodes a secretory polypeptide structurally related to complement control proteins. *Nature* **335:** 176-178.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: a program to check the stereochemical euqlity of protein structures. *J Appl Cryst* **26:** 283-291.

Liu, X., Dai, S., Zhu, Y., Marrack, P., and Kappler, J.W. 2003. The structure of a Bcl-xL/Bim fragment complex: implications for Bim function. *Immunity* **19:** 341-352.

Massaad, C.A., Portier, B.P., and Taglialatela, G. 2004. Inhibition of transcription factor activity by nuclear compartment-associated Bcl-2. *J Biol Chem* **279:** 54470-54478.

Massung, R.F., Esposito, J.J., Liu, L.I., Qi, J., Utterback, T.R., Knight, J.C., Aubin, L., Yuran, T.E., Parsons, J.M., Loparev, V.N., et al. 1993. Potential virulence determinants in terminal regions of variola smallpox virus genome. *Nature* **366:** 748-751.

McRee, D.E. 1999. XtalView/Xfit – a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125:** 156-165.

Moss, B. 2000. Poxviridae: The viruses and their replication. In *Fields Virology*, 2nd ed. (eds. D.M. Knipe, and P.M. Howley), pp. 2849-2883. Lippincott Williams & Wilkins, Philadelphia.

Muchmore, S.W., Sattler, M., Liang, H., Meadows, R.P., Harlan, J.E., Yoon, H.S., Nettesheim, D., Chang, B.S., Thompson, C.B., Wong, S.L., et al. 1996. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* **381:** 335-341.

O'Neill, J.W., Manion, M.K., Maguire, B., and Hockenbery, D.M. 2006. BCL-XL dimerization by three-dimensional domain swapping. *J Mol Biol* **356:** 367-381.

Otwinowski, Z., and Minor, W. 1997. Processing x-ray diffraction data collected in oscillation mode. *Methods. Enzymol.* **276:** 307-326.

Petros, A.M., Olejniczak, E.T., and Fesik, S.W. 2004. Structural biology of the Bcl-2 family of proteins. *Biochim Biophys Acta* **1644:** 83-94.

Postigo, A., Cross, J.R., Downward, J., and Way, M. 2006. Interaction of F1L with the BH3 domain of Bak is responsible for inhibiting vaccinia-induced apoptosis. *Cell Death Differ* **Advanced online publication:** 1-12.

Ray, C.A., Black, R.A., Kronheim, S.R., Greenstreet, T.A., Sleath, P.R., Salvesen, G.S., and Pickup, D.J. 1992. Viral inhibition of inflammation: cowpox virus encodes an inhibitor of the interleukin-1 beta converting enzyme. *Cell* **69:** 597-604.

Regula, K.M., Ens, K., and Kirshenbaum, L.A. 2002. IKK beta is required for Bcl-2-mediated NF-kappa B activation in ventricular myocytes. *J Biol Chem* **277:** 38676-38682.

Sattler, M., Liang, H., Nettesheim, D., Meadows, R.P., Harlan, J.E., Eberstadt, M., Yoon, H.S., Shuker, S.B., Chang, B.S., Minn, A.J., et al. 1997. Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science* **275:** 983-986.

Schneider, T.R., and Sheldrick, G.M. 2002. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58:** 1772-1779.

Seet, B.T., Johnston, J.B., Brunetti, C.R., Barrett, J.W., Everett, H., Cameron, C., Sypula, J., Nazarian, S.H., Lucas, A., and McFadden, G. 2003. Poxviruses and immune evasion. *Annu Rev Immunol* **21:** 377-423.

Shchelkunov, S.N. 2003. Immunomodulatory proteins of orthopoxviruses. *Mol Biol (Mosk)* **37:** 41-53.

Shindyalov, I.N., and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* **11:** 739-747.

Su, J., Wang, G., Barrett, J.W., Irvine, T.S., Gao, X., and McFadden, G. 2006. Myxoma virus M11L blocks apoptosis through inhibition of conformational activation of Bax at the mitochondria. *J Virol* **80:** 1140-1151.

Taylor, J.M., and Barry, M. 2006. Near death experiences: poxvirus regulation of apoptotic death. *Virology* **344:** 139-150.

Tulman, E.R., Afonso, C.L., Lu, Z., Zsak, L., Kutish, G.F., and Rock, D.L. 2004. The genome of canarypox virus. *J Virol* **78:** 353-366.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229:** 105-124.

Wang, G., Barrett, J.W., Nazarian, S.H., Everett, H., Gao, X., Bleackley, C., Colwill, K., Moran, M.F., and McFadden, G. 2004. Myxoma virus M11L prevents apoptosis through constitutive interaction with Bak. *J Virol* **78:** 7097-7111.

Wasilenko, S.T., Banadyga, L., Bond, D., and Barry, M. 2005. The vaccinia virus F1L protein interacts with the proapoptotic protein Bak and inhibits Bak activation. *J Virol* **79:** 14031-14043.

Woo, J.S., Jung, J.S., Ha, N.C., Shin, J., Kim, K.H., Lee, W., and Oh, B.H. 2003. Unique structural features of a BCL-2 family protein CED-9 and biophysical characterization of CED-9/EGL-1 interactions. *Cell Death Differ* **10:** 1310-1319.

Yin, X.M., Oltvai, Z.N., and Korsmeyer, S.J. 1994. BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax. *Nature* **369:** 321-323.

Zhai, D., Luciano, F., Zhu, X., Guo, B., Satterthwait, A.C., and Reed, J.C. 2005. Humanin binds and nullifies Bid activity by blocking its activation of Bax and Bak. *J Biol Chem* **280:** 15815-15824.

Zhang, Z., Abrahams, M.R., Hunt, L.A., Suttles, J., Marshall, W., Lahiri, D.K., and Kotwal, G.J. 2005. The Vaccinia Virus N1L Protein Influences Cytokine Secretion in Vitro after Infection. *Ann N Y Acad Sci* **1056:** 69-86.

Zhou, H., Zhang, C., Liu, S., and Zhou, Y. 2005. Web-based toolkits for topology prediction of transmembrane helical proteins, fold recognition, structure and binding scoring, folding-kinetics analysis and comparative analysis of domain combinations. *Nucleic Acids Res* **33:** W193-197.

Table 1. Crystallographic data collection, phasing and refinement statistics

	Se-peak	**Se-remote**	**Se-edge**	**Native**
Data collection				
Wavelength (Å)	0.97916	0.91841	0.97948	1.11587
Resolution (Å)	40.0 - 3.0 (3.11 – 3.00)[a]			30.0 – 2.2 (2.28 – 2.20)[a]
Space group	$P2_1$			$P2_1$
Unit cell dimensions (Å)	a=68.7, b=109.4, c=70.2, β=110.6°			a=68.6, b=110.0, c=69.6, β=110.9°
Total reflections	143216	95038	111262	127223
Unique reflections	37519[b]	37514[b]	36309[b]	47333
Completeness (%)	98.1 (88.1)[a]	97.4 (83.6)[a]	93.7 (63.0)[a]	95.7 (79.3)[a]
R_{MERGE}	0.072 (0.45)[a]	0.074 (0.48)[a]	0.071 (0.56)[a]	0.050 (0.28)[a]
$\langle I/\sigma I\rangle$[d]	17.3 (2.3)[a]	13.3 (1.6)[a]	15.6 (1.4)[a]	17.8 (2.1)[a]
Phasing				
Phasing power:				
anomalous differences	2.1	0.93	0.92	
dispersive differences	1.00 / 0.74[e]	- / -	1.77 / 1.68	
Overall figure-of-merit:				
acentric / centric		0.53 / 042		
Refinement				
Resolution (Å)				30.0 – 2.2
Reflections				47043
R_{WORK}/R_{FREE}[g]				0.21 / 0.25
No. protein / water atoms				5885 / 238
$\langle$Overall B-factor$\rangle$ (Å^2)				41.0
RMSD bond length (Å)				0.0064
RMSD bond angle (°)				1.08
Ramachandran plot				
Most favored (%)				95.6
Additional allowed (%)				4.4

[a]Highest resolution shell; [b]Friedel pairs not merged; [c]$R_{MERGE} = \sum\sum_j |I_j - \langle I\rangle| / \sum\sum_j I_j$; [d]Average signal-to-noise ratio; [e]Acentric/centric; [f]$R = \sum||F_o| - |F_c|| / \sum|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively; [g]5 % of the reflections were set aside randomly for R_{FREE} calculation.

Figure legends

Fig. 1. Structure of vaccinia N1L and comparison with Bcl-2 family proteins. (A) Stereo view of the N1L monomer. Helices and termini are labeled. **(B)** Stereo superposition of N1L (navy) and Bcl-X_L (gray; 1MAZ (Muchmore et al. 1996)). N1L helices are labeled. Functionally important BH regions of Bcl-X_L are colored in magenta (BH4), green (BH3), orange (BH1) and cyan (BH2). **(C)** Structure-based sequence alignment of N1L with Bcl-2 family members: Mcl-1, mouse myeloid cell leukemia-1 (PDB code 1WSX (Day et al. 2005)); KSHV, Kaposi sarcoma virus Bcl-2 homolog (1K3K (Huang et al. 2002)); human Bcl-X_L (1MAZ (Muchmore et al. 1996)). Hydrophobic residues are highlighted in green, acidic/basic residues are in red/blue. Secondary structures of N1L and Bcl-X_L below the sequence; and consensus BH motifs are indicated above with the same color scheme as in (B). The highly conserved Bcl-2 signature motif, NGWR, is boxed.

Fig. 2 N1L adopts a dimeric structure. (A) Stereo view of the N1L homodimer. The $\alpha1$ and $\alpha6$ helices from one N1L monomer (blue) interact in an antiparallel way with equivalent helices in another monomer (green). N- and C-termini and helices of each subunit are labeled. **(B)** Specific $\alpha1$ and $\alpha6$ residues at the N1L dimer interface. In the anti-parallel N1L dimer, Ile6, Leu10, Phe95 and Phe99 constitute a critical hydrophobic patch whereas Arg7/Asp14 and Arg90/Glu103 pairs form complementary electrostatic surfaces, not present in Bcl-X_L. The N1L monomer (blue) in this view is related to the blue monomer in (A) by 90° about a vertical axis. (**C**) The same view for Bcl-X_L showing analogous residues, which are either not hydrophobic or not complementary in charge. BH1-4 domains are colored in magenta (BH4), green (BH3), orange (BH1) and cyan (BH2), as in Fig. 1B and C.

Fig. 3. N1L binds to BH3 peptides. **(A)** and **(B)** Hydrophobic surfaces of N1L and Bcl-X_L. The solvent-accessible surface of N1L indicates the presence of a small hydrophobic groove on the same face of the molecule as the BH3 binding groove in Bcl-X_L. Phe, Trp, Tyr, Met, Ile, Leu, Val and Ala are colored in yellow. Approximate positions of helices surrounding the groove are indicated on the surfaces. The orientation of each protein is similar to that in Fig 1B. (**C**) Fluorescence polarization plots of FITC-labeled BH3 domains (Bid, Bim, Bak and Bad) in the presence of varying concentrations of N1L (■) or Bcl-X_L (◘) (**D**) Tabulated EC_{50} values from these plots.

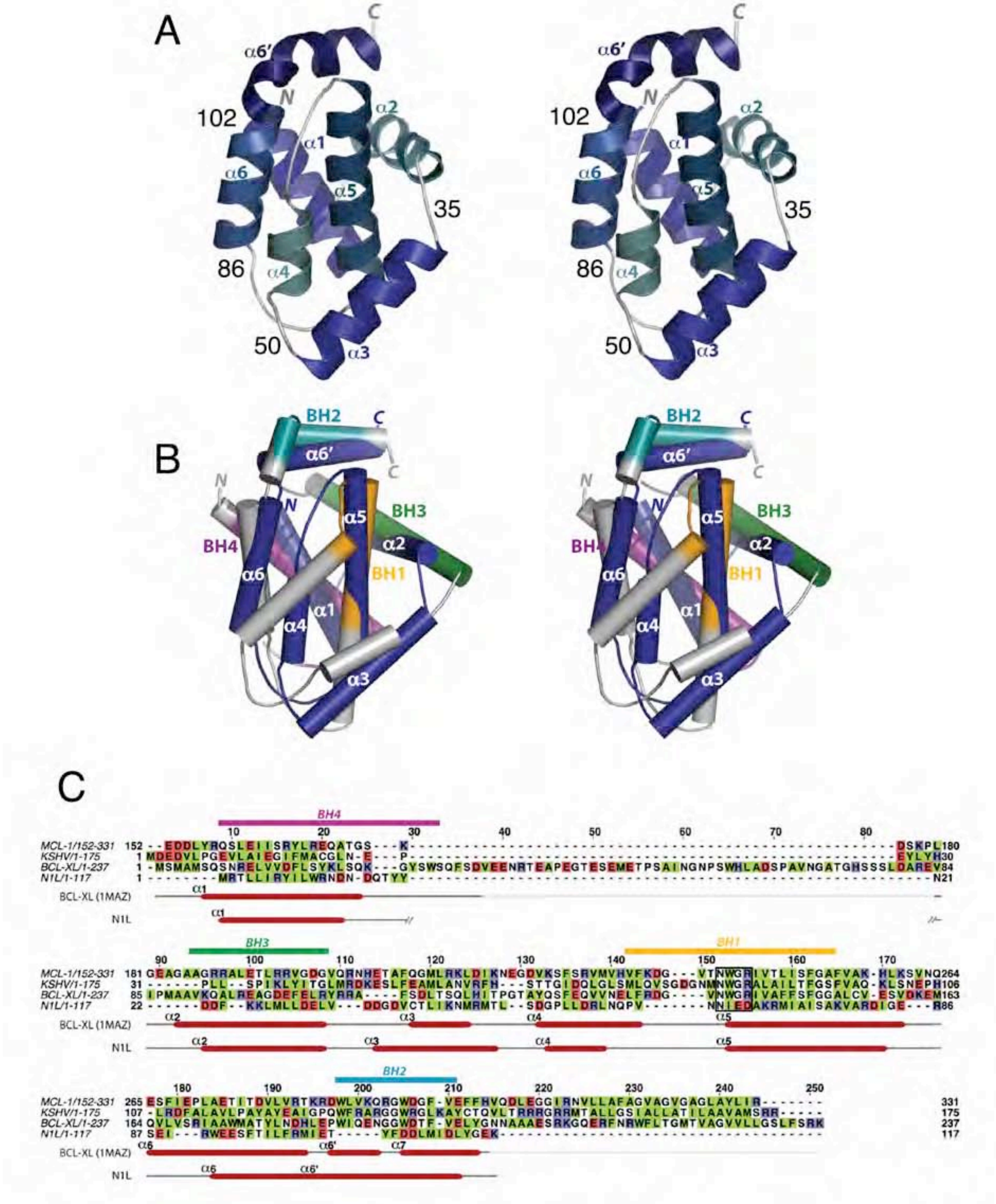


Figure 1: Aoyagi et al.

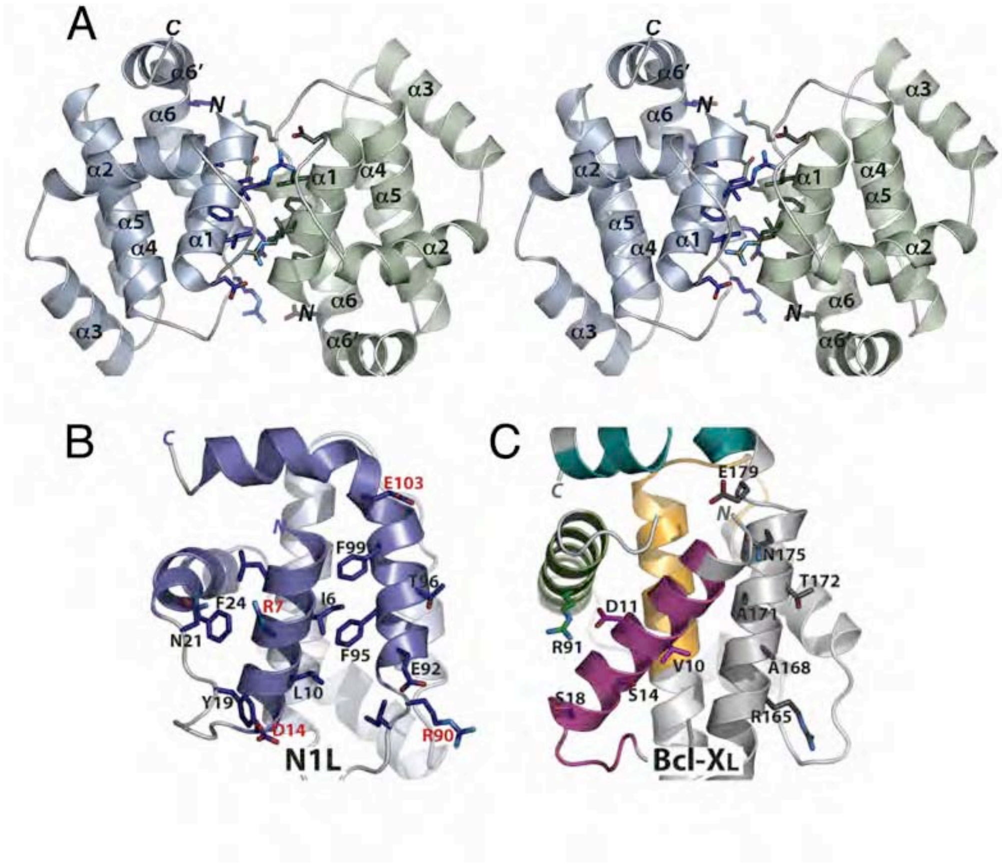
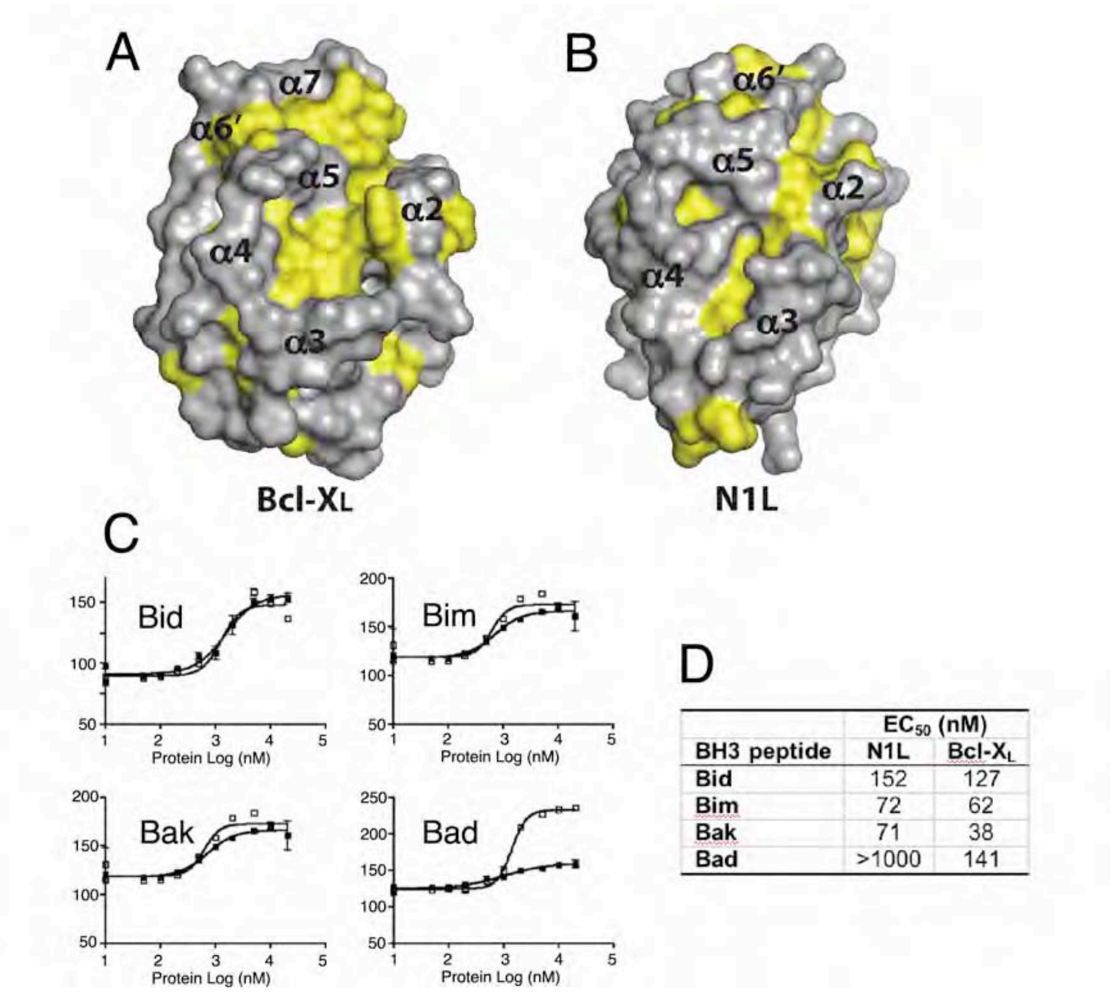


Figure 2: Aoyaga et al.



BH3 peptide	EC$_{50}$ (nM) N1L	EC$_{50}$ (nM) Bcl-X$_L$
Bid	152	127
Bim	72	62
Bak	71	38
Bad	>1000	141

Figure 3: Aoyagi et al.

VirFact: a relational database of virulence factors and pathogenicity islands (PAIs).

Adrian Tkacz, Leszek Rychlewski and Adam Godzik

ABSTRACT

The VirFact database (http://virfact.burnham.org) contains information on microbial virulence factors and pathogenicity islands (PAIs) from major pathogens. The database collects information from literature and combines them with results obtained by genome context analysis and distant homology recognition. The database can be browsed by virulence factor, PAI or organism name. The annotations, including multiple alignments of proteins homologous to virulence factors, genomic context, models of three dimensional structures (if available) are presented using graphical web interface and standard visualization tools. The VirFact can also be used as a tool to recognize the presence of homologs of known virulence factors in the genome delivered by the user.

INTRODUCTION

Recent development of comparative genomic analysis and experimental molecular biological techniques made it possible to identify specific genes responsible for virulence of pathogenic microbes. Despite some discussions (1), it is widely accepted that virulence of a pathogenic microbe is imparted by a specific set of genes, often localized together on a plasmid (virulence plasmids) or on a genome (pathogenicity islands). Virulence factors are typically identified by comparing genomic sequences of pathogenic and non-pathogenic strains or by studying virulence of deletion mutants. While building VirFact we adhered to a broad definition of a virulence factor that includes genes specifically involved in interactions between a pathogen and its host, but also genes supporting pathogenic lifestyle and many genes of unknown function if they are part of the genomic structure related to pathogenicity. Virulence factors of many organisms are well studied, but the information about them is usually available only is specialized literature and then usually only in the context of a specific organism. We believe that this scattering of information makes it difficult to study general questions involving pathogenicity, such as for instance similarity between virulence apparatus of unrelated pathogens. At the same time, sequence analysis and annotations of many virulence related genes is very uneven and tools such as distant homology analysis, fold recognition or modeling are seldom used. The goal of the VirFact project is the development of a well annotated database containing information about pathogenicity systems from different organisms and providing a uniform level of annotation, including annotations with most sensitive algorithms.

THE DATABASE

The VirFact database (http://virfact.burnham.org) is implemented as a relational database containing a collection of virulence factors and pathogenicity islands from major microbial pathogens. The current release of VirFact is divided into five main areas (discussed below) providing different approaches and views to data analysis:

- a collection of individual virulence factors
- a collection of pathogenicity islands
- source genomes
- annotations and prediction results
- links

The first section contains basic information about individual virulence factors, such as their amino acid sequences, annotations collected from literature and links to other fields in database. This area is de facto the core of the system.

Individual virulence factors from a given organisms often form operon like structures called pathogenicity islands (PAIs) – information about them forms the next area of the VirFact database. Additional data, such as a PAI position at the genome, its short characterization and lists of genes it contains is provided here. Since PAIs usually evolve by lateral transfer, they differ by many features from the host genome. To aid in identifying novel PAIs, the user can view a chart (deposited in database) showing genomic regions that deviate most from the rest of the genome. This diversity is based on three compositional criteria: G+C content, dinucleotide frequency and codon usage (2).

For individual virulence factors, the annotations and results of analysis and prediction tools provide information about homologs and genomic context and other information about a chosen virulence factor, as discussed in detail below.

Finally, the links to sections described above and various addresses that are useful for the user or necessary for the service are listed in a separate area of the website. The current (July 20, 2004) release of VirFact contains about 400 proteins, 12 pathogenicity islands (PAIs) and 7 completely sequenced genomes and it is increasing constantly.

THE WEB SITE

VirFact is publicly available on the web at http://virfact.burnham.org. The database can be browsed by virulence factor name, PAI or genome using links on the top of the main web page.

- the "Virulence Factors" link: lets the user to see all virulence factors deposited in the database
- the "PAIs" link: allows to display all PAIs that are contained in VirFact. After selection of a specific PAI, the composition of PAI proteins is shown.
- the "Genome" link: leads user to an interface, which allows to check all VirFact proteins that are encoded in selected genome. An additional feature is a chart showing genomic regions that deviate most from the rest of the genome, which could form new, as yet unrecognized PAIs.

For each displayed virulence factor, on the right side of a webpage, there are links to annotation and prediction results, to sequence in FASTA format or to other links that could be potentially useful, like to NCBI PubMed. The link called "Homologs", allows user to view PSI-BLAST (3), FFAS03 (4) or T-Coffee (5) results. PSI-BLAST is used to compare a query sequence with those contained in non redundant protein database at NCBI by performing the iterative BLAST search. It is the most sensitive widely used program for recognizing homologs, making it useful for finding very distantly related proteins. The "FFAS" link shows the results of FFAS03 server, a profile-profile alignment algorithm used for super-sensitive recognition of distant homologs and fold assignments. Finally, links called "Alignment" and "Tree" leads to T-Coffee results, where a multiple alignment was built using proteins found by the PSI-BLAST search. The T-Coffee results can be visualized with the "JalView" (multiple sequence alignment viewer, 6) and the "A Tree Viewer (ATV)" (phylogenetic tree viewer, 7) applications (Java Virtual Machine is required by both programs).

The "Genomic Context" interface was designed to perform the analysis of the genomic context using The SEED system (http://theseed.uchicago.edu/FIG/index.cgi) for genome annotations. As described by Overbeek et al. SEED is designed to help a researcher study a specific subsystem (set of genes), supporting community-wide annotation of genomes and searching for specific missing genes. SEED focuses on conservation of a genomic context between homologs of the specific gene. In VirFact, we compared genomic context of close homologs of the virulence factor being studied. It is important to note that SEED uses its own definition of a homolog, typically much more conservative then would result from a PSI-BLAST search.

The VirFact can also be queried using the Web-based interface called "Scan" for a presence of homologs of virulence factors covered by VirFact in the genome provided by the user. The search takes some time, up to several minutes, depending on a genome size. The output page shows potential virulence factors in the user genome, with information about the similarity score to known virulence factors, the position on a genome and the sequence alignment to the "parent" virulence factor in the FASTA format. For example, we show here a short analysis of *Francisella tularensis* genome. In the example presented here we focus on the information on how to use VirFact website, the full analysis of the potential virulence factors in *F. tularensis* genome will be presented elsewhere. As is showed in the chart (Fig. 1), there is a peak around 45 kb indicating high diversity of this region from the rest of the genome. In the same region VirFact found a protein similar to "Z0262 gene product" of Escherichia coli. Further analysis indicates that this hypothetical protein of E. coli has a homolog described only in the case of *Francisella tularensis*, called IglB. The last protein is acknowledged as associated in intracellular growth (8). Moreover, a neighborhood of "Z062 gene product" shows the functional coupling with other unknown proteins often present in other pathogens.

UPDATES

Parsing, annotation and data updates have been automated to minimize human intervention. The VirFact database will be updated at least once per two months to ensure current report of data. The information about PAIs is manually curated.

FUTURE PERSPECTIVES

VirFact was developed as a relational database of PAIs and virulence factors for the comprehensive representation of pathogenicity in various prokaryotic organisms. A web interface was designed to easy access the various features. To our knowledge, this is the only database devoted exclusively to pathogenicity island and virulence factors that provides a variety of tools for data analysis. We plan to expand the VirFact database to incorporate all annotated PAIs from all completely sequenced genomes and all virulence-related genes/proteins described in the literature. In near future we would like to broad VirFact of new tools predicting surface regions of the proteins and trans-membrane regions. We believe the VirFact will be useful tool for the investigation of the bacterial virulence and for the detection of virulence factors in newly sequenced genomes.

ACKNOWLEDGEMENTS

We would like to thank Dr Ross Overbeek for The SEED: an Annotation/Analysis Tool that makes possible a development a genome context part of VirFact service. The authors also thank Zhanwen Li for her help in FFAS calculations. The work was partially funded by the 6FP grant MicrobeArray (to LR) and United States Army Medical Research and Materiel Command Grant DAMD17-03-2-0038 (to AG).

REFERENCES

1. Wassenaar,T.M. and Gaastra,W. (2001) Bacterial virulence: can we draw the line? FEMS Microbiol Lett., 201, 1-7.

2. Tu,Q. and Ding,D. (2003) Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis. FEMS Microbiol. Lett., 221, 269-275.
3. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25, 3389-3402.
4. Rychlewski,L., Jaroszewski,L., Li,W. and Godzik,A. (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Science, 9, 232-241.
5. Notredame,C., Higgins,D.G. and Heringa,J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol., 302, 205-217.
6. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. Bioinformatics, 20, 426-427.
7. Zmasek,C.M. and Eddy,S.R. (2001) ATV: display and manipulation of annotated phylogenetic trees. Bioinformatics, 17, 383-384.
8. Gray,C.G., Cowley,S.C., Cheung,K.K. and Nano,F.E. (2002) The identification of five genetic loci of Francisella novicida associated with intracellular growth. FEMS Microbiol Lett., 215, 53-56.

Figure 1. Graphic illustration of the using the VirFact for a search of virulence homologs in the genome delivered by the user. The chart of discriminant scores shows a region that deviates most from the rest of the genome. The VirFact has found in this place a homolog similar to “Z062 gene product” of *Escherichia coli*. The PSI-BLAST result show that “Z062 gene product” has a similar sequence: IglB [*Francisella tularensis*]. Moreover, the “Genomic Context” interface shows a significant neighborhood of Z062 with other proteins (in table, the “Z062 gene product” is no. 1, called as “hypothetical protein”).

Opinion TRENDS in Microbiology Vol.13 No.11 November 2005
Full text provided by www.sciencedirect.com

Convergent evolution as a mechanism for pathogenic adaptation

Sergey Sikora, Alex Strongin and Adam Godzik

The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

The survival of human pathogens depends on their ability to modulate defence pathways in human host cells. This was thought to be attained mainly by pathogen specific 'virulence factors'. However, pathogens are increasingly being discovered that use distant homologs of the human regulatory proteins as virulence factors. We analyzed several cases of this approach, with a particular focus on virulence proteases. The analysis reveals clear cases of bacterial proteases mimicking the specificity of their human counterparts, such as strong similarities in their active and/or binding sites. With more sensitive tools for distant homology recognition, we could expect to discover many more such cases.

The undercover agents of bacterial invasion

The success of human pathogens largely depends on their adaptation to the human organism and, in particular, on the ability of the pathogen to influence and modulate human pathways involved in defence mechanisms. Many virulence factors are pathogen-specific proteins but many are distant homologs of human proteins. With the development of more sensitive tools for distant homology recognition and sequence–structure relationship determination, we could expect to discover more cases from the class of human protein homologs and perhaps even move some cases from the first to the second group. The main purpose of the current studies is to determine how a pathogenic bacterium adapts bacterial proteins that are homologs of host proteins in its invasion. One of the most intriguing examples in this respect is a group of virulent proteases. Because of their function, proteases are ideal drug targets and, as a result, have drawn special attention from the scientific community. Virulent proteases have an important role in the pathology of such dangerous diseases as anthrax [1], tuberculosis [2] and smallpox [3]. The recent focus of the US National Institute of Health emphasized the importance of proteases in human diseases and established an interdisciplinary Center for Proteolytic Pathways (CPPs), where the studies reported here were conducted. Although mimicking host proteins by bacterial virulent factors has been discussed for some time [4], what the evolutionary and molecular features of such mimicry are has remained unknown. Here, we have been able to determine such features in several specific systems.

In the case of bacterial homologs of human proteins, the evolutionary distance between the organisms results in clear differences in structure, specificity and function between the human and bacterial proteins. Thus, to interfere with the human pathways successfully, bacterial virulence factors have to solve the apparent contradiction between their distant evolutionary relation, which usually implies functional divergence, and the functional similarity needed for their virulence function. We can imagine that selective pressure would result in convergent, function-driven evolution and, here, we show that this is indeed the case.

Convergent evolution of the structural architecture, as well as active site sequences of virulent bacterial factors toward its human homologs, is not a novel idea. The active-site convergent evolution has been demonstrated for several virulent factors, such as SptP from *Salmonella* and the invasin from *Yersinia* [4]. In one example, to be internalized by the host cell, *Salmonella* delivers SopE and SopE2 proteins into the cell using the type III secretion system [5–8]. These proteins have a guanine nucleotide exchange factor (GEF) activity, which activates Rac1 and CDC42 specifically. Activation of these proteins leads to cytoskeleton rearrangement and subsequent bacterial internalization. When inside the cell, *Salmonella* reverses the cytoskeleton rearrangement to normalize the function of the host cell by delivering an SptP protein, which reverses the SopE and SopE2 effect in a manner highly specific for the host cell. The SptP protein is a GTPase activating enzyme (GAP), which induces GTP hydrolysis and thus shuts down the Rac1/CDC42-dependent skeleton rearrangement pathway. The SptP analysis suggests convergent evolution of the *Salmonella* GAP with the host GAPs [5,6,8,9] and, specifically, with the active site of the host GAPs. The second example comes from *Yersinia pseudotuberculosis*, which uses invasin protein to bind the integrin receptor on the host cell surface, subsequently leading to internalization of the bacteria by the host cell. The invasin protein achieves this effect by means of convergent evolution with the integrin-binding surface of fibronectin [10–12]. Yet another example of the convergent evolution of the bacterial virulence factor toward the host homolog is presented by the YopJ protein from *Yersinia pestis*. This protein mimics the activity of the human sumo protease to disrupt posttranslational modification of the host proteins and, subsequently, to inhibit various signalling pathway in the host cell [13–16].

Corresponding author: Godzik, A. (adam@burnham.org).

www.sciencedirect.com 0966-842X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.tim.2005.08.010

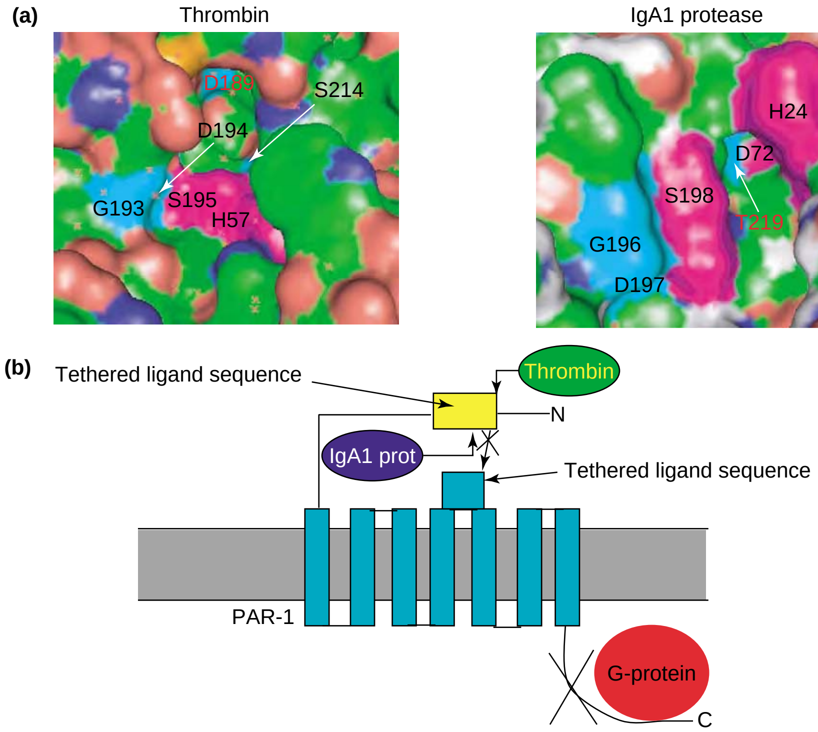


Figure 1. Modulation of the inflammation response by the IgA1 protease. **(a)** Structural representation of thrombin and the IgA1 protease catalytic and/or substrate-binding pocket. Residues that correspond to the catalytic triad are coloured magenta and conserved surface residues in the catalytic pocket are coloured cyan. The negatively charged aspartate of thrombin D189 is substituted on the surface topologically by T219 in IgA1 protease. Both D189 and T219 are labelled in red font. Both catalytic pockets of thrombin and IgA1 protease have similar topology and most probably accommodate a loop structure due to the tightness of both pockets. Both thrombin and the IgA1 protease have the same substrate specificity because they have identical substrate-binding pockets. However, elimination of negatively charged D189 from the surface and the appearance of polar T219 leads to the altered scissile bond specificity of the IgA1 protease, compared to thrombin. **(b)** The probable mechanism of the inhibition of PAR-1-mediated signalling. Activated thrombin digests away the N-terminal portion of PAR-1 inducing the conformational change in the N-terminal part of the PAR-1 receptor and subsequent binding of the N-terminal tethered ligand sequence to the effector tethered ligand sequence, which activates G-protein-dependent signalling. The IgA1 protease cuts the tethered ligand sequence downstream from the thrombin scissile bond, thus eliminating signalling and the subsequent inflammation response, even in the presence of active thrombin.

The analysis of two virulent proteases from *Haemophilus influenza* and *Mycobacterium tuberculosis*, which is presented in subsequent sections, showcases the molecular level convergent evolution of the protease active sites toward their human counterparts as a mechanism of virulent adaptation. We also demonstrate an example of a bacterial toxin that mimics the human signalling domains to ensure a particular compartmentalization. In the two examples, we show that active site mutations in virulence factors make them nearly identical to their human homologs, thus making them able to interact with human substrates. In another example, we show how other parts of the virulence factors not associated with the active site structurally evolve to enable particular compartmentalization and subsequent inclusion of the bacterial homolog into the corresponding human pathway. The data presented here suggest that both these mechanisms are employed successfully in the course of evolution of bacterial pathogens.

Does IgA1 protease modulate thrombin-dependent inflammation response?

During bacterial infection a human organism responds to the invasion through various defence mechanisms, including inflammation. The inflammation response signalling pathway is regulated by thrombin (Figure 1b), which activates protease-activated receptors (PARs). PAR-1 is a transmembrane receptor, which is activated when thrombin digests its N-terminal extracellular part located next to the tethered ligand sequence [17]. Conformational change within the N-terminal part of PAR-1 leads to a signalling cascade, in turn leading to the inflammation response (Figure 1b). We suggest that IgA1 protease from *Haemophilus influenzae* modulates this response by means of convergent evolution with the thrombin substrate-binding site (with some important modifications) and subsequently by blocking the inflammation response.

The known activity of IgA1 virulent protease is to digest IgA1 immunoglobulin at the hinge region [18]. However, given the modest role of IgA1 in human humoral defence, it is unclear how this mechanism could explain the virulence of the *Haemophilus* species. IgA1 protease contains a domain that is distantly related to thrombin (PDB code 1h8d, FFAS score $= -10.100$). The similarity is especially evident in the substrate-binding and/or catalytic region (Figure 1a). The model of the IgA1 protease on a thrombin template revealed an almost complete conservation of the catalytic pocket between the two proteins; the single exception was a negatively charged aspartate of the thrombin becoming a polar threonine on the IgA1 protease surface (Figure 1a). This minor difference provides a plausible explanation for the respective sequence specificity of these proteases: IgA1 cleaves specifically at the PS (proline-serine) motif [19],

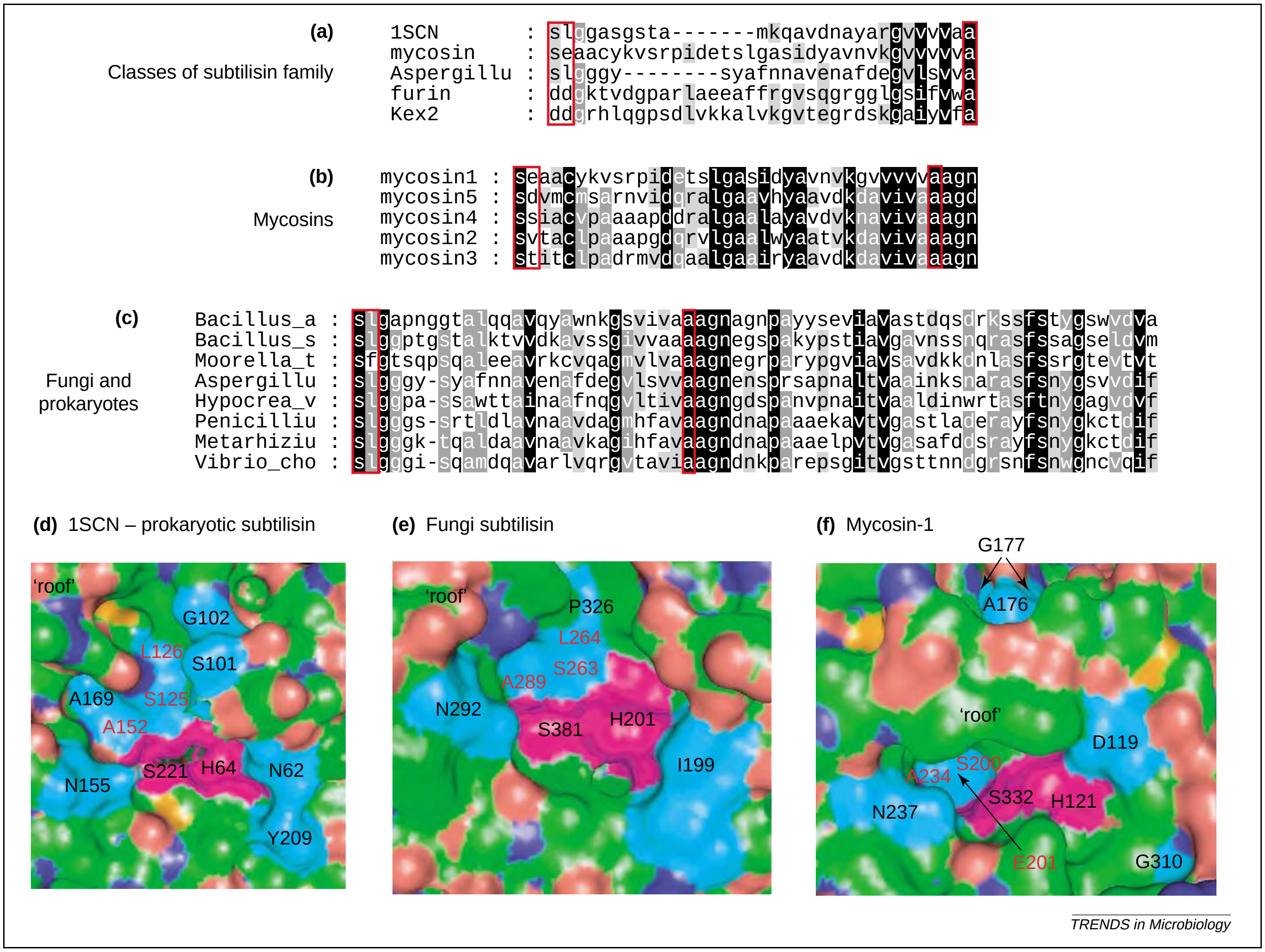


Figure 2. Structural analysis of the mycosin-1 active site. **(a)** Multiple sequence alignment of various classes of the subtilisin family of proteases. The part of the multiple sequence alignment that possesses three residues responsible for substrate specificity is shown. The subtilisins from prokaryotes (PDB code 1SCN), mycosin-1, subtilisin from *Aspergillus fumigatus* (fungi), furin and kex-2 are labelled. Multiple sequence alignment was performed using T-COFFEE. Positions corresponding to the identified substrate-specific residues are inscribed in red rectangles. **(b)** Multiple sequence alignment of five mycosin paralogs from *Mycobacterium tuberculosis*. Positions corresponding to the identified substrate-specific residues are inscribed in the red rectangle. **(c)** Multiple sequence alignment of prokaryotic and fungi subtilisins. Positions corresponding to the identified substrate-specific residues are inscribed in the red rectangle. **(d,e,f)** Structural representation of catalytic and/or substrate-binding pockets of prokaryotic subtilisin, fungi subtilisin from *Aspergillus fumigatus* and mycosin-1, respectively. Catalytic residues are coloured in magenta. Other conserved surface residues, which are part of the proteolytic pocket, are coloured in cyan. Substrate-specific residues located at the bottom of the pocket are labelled in red, whereas the rest of the residues are labelled in black.

whereas thrombin cleaves at RS (arginine-serine) [20]. We hypothesize that, as suggested by the overall similarity of the binding sites, IgA1 proteases mimic thrombin and bind to the thrombin-binding region of PAR-1 tethered ligand but probably cleave it at a different site. Cleavage of PAR-1 tethered ligand at a non-native site would lead to the inactivation of PAR-1 and to the blocking of an inflammation response (Figure 1b).

The IgA1 proteases represent a large group of virulent proteases, members of which share sequence and topological similarity within the substrate-binding pocket. The catalytic and/or substrate-binding pocket is largely conserved throughout the group. Almost all of the residues within the catalytic pocket are conserved throughout all the bacterial IgA1 protease homologs. The last residue – T219 – is conserved in some members of the IgA1 protease family, such as IgA1 protease from *Haemophilus influenzae,* haemoglobin and Tsh proteases from *Escherichia coli,* but it is aspartate in other members of the IgA1 virulent protease family. A recently released atomic-resolution structure of one of the members of the group, haemoglobin protease from *E. coli,* supports our prediction of the high degree of structural and substrate-binding-pocket similarity between thrombin and IgA1 proteases [21].

Mycosins: the road from bacterial to human specificity

The second example of bacterial adaptation is provided by recently discovered mycosin proteins from *Mycobacterium tuberculosis* [2]. There are five paralogs in *M. tuberculosis* – mycosin 1–5 – but mycosin-1 orthologs have been found only in the pathogenic species of *Mycobacterium* [2]. Mycosins belong to the subtilisin family, which includes various prokaryotic and fungi subtilisins as well as the vertebrate furin and kex-2.

The bottom of the substrate-binding and/or catalytic pocket of the prokaryotic subtilisin (PDB code 1scn)

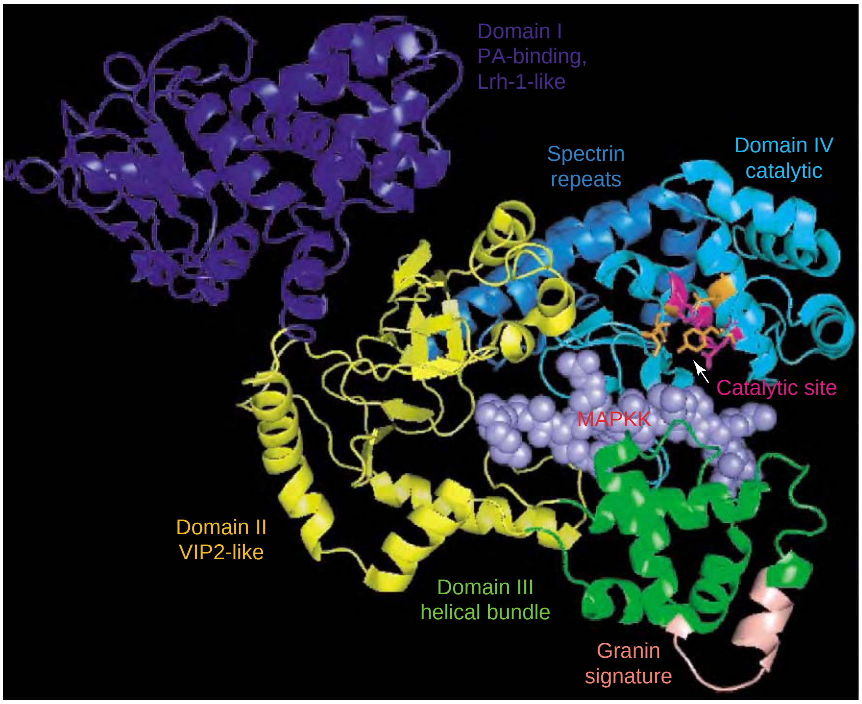


Figure 3. Structural representation of the domains of lethal factor (PDB code 1jky). The N-terminal domain I (dark blue) was found to be structurally similar to the Lrh-1 protein. Domain II (yellow) was found to have a fold similar to the VIP-2 ADP ribosylase. Domain III (green) is a helical bundle containing a stretch of amino acids identified by ExPASy as a granin signature (coloured salmon). Domain IV (cyan) contains the catalytic centre, in which the catalytic triad (shown in sticks) is coloured magenta; other residues in the catalytic centre important for the proteolytic activity are coloured orange. Domain IV also contains spectrin repeats outside the catalytic centre (coloured marine). The substrate (MAPKK) is represented as purple spheres.

contains three key residues S125, L126 and A152 (Figure 2d, labelled in red), which are conserved in all prokaryotes (except mycosin-1) and in fungi. These residues determine the substrate specificity of each particular member of the family. The conservation of the residues of the substrate-binding pocket throughout various representatives of the subtilisin family – fungi, *Mycobacterium tuberculosis* other prokaryotes and vertebrate subtilisins (furin and kex2) – are analyzed in Figure 2a. Residue A152 is conserved throughout the entire family, whereas S125 is conserved in prokaryotes (including *Mycobacterium* mycosins) and is substituted by a negatively charged residue in vertebrate subtilisins (Figure 2a). The L126 of the prokaryotic subtilisins is substituted by a negatively charged residue in both vertebrate subtilisins and mycosin-1 (Figure 2b). Modelling shows that the binding pockets have nearly identical shapes between regular prokaryotic and fungi subtilisins (Figure 2d and e). Only the mycosin-1 pocket is much tighter and more similar to the furin catalytic pocket (Figure 2a, d and f). We suggest that the shape and mutation pattern of the mycosin-1 substrate-binding pocket shows its convergent evolution toward human furin, suggesting that its pathogenic mechanism might involve digesting furin substrates.

Anthrax lethal factor adaptation

Our final example comes from the structural and sequence analysis of lethal factor (LF) and protective antigen (PA) from *Bacillus anthracis*. The LF, together with its transporter (activated protective antigen or PA63), enters the macrophages and causes macrophage apoptosis by inactivating the MAPKK functionality [1]. Anthrax produces full size versions of protective antigen, PA83, which is subsequently digested by human protease furin into PA63 and PA20 [22]; this in turn leads to PA63 heptamerization, binding to the LF and subsequent endocytosis of the PA63–LF complex by macrophages [1]. In the cytosol, LF digests MAPKK, which leads to the induction of macrophage apoptosis; however, the exact mechanism of this is unclear [23].

Possible hints about the function of LF come from structure analysis. In addition to known structural similarities of LF domains, we found that the N-terminal domain I structure is significantly similar to Lrh-1 nuclear receptor ligand-binding protein (PDB code 1pk5, RMSD = 2.64A) [24], suggesting possible nuclear localization of the LF–PA63 complex (Figure 3). This is further supported by the presence of the nuclear localization signal (^{225}VKNKRT^{230}) within PA63 proteins and the presence of granin-1 signature (positions 346–355) in domain III. The role of granins is in directing secretory vesicles to the destination and granin signature serves as a sorting signal [25]. Furthermore, a part of the catalytic domain IV of the LF (Figure 3) is distantly similar to a spectrin repeat (PDB code 2spc, RMSD = 1.14A); these are involved in intracellular transport, including nuclear transport [26]. Taken together, these three observations suggest a possibility for nuclear localization of the LF–PA complex, suggesting a mechanism of LF action that is quite different from the current consensus. In addition, domain IV human-like spectrin repeats, which surround the catalytically active region of domain IV, mask the protelolytic centre of the LF and possibly have a role in protecting the protease domain from degradation by cellular proteases.

Experimental co-localization studies of the PA–LF complex have proved our hypothesis because these two proteins were found to be localized to the nucleus (A. Strongin, unpublished results). Next, we identified potential targets of the LF in the nucleus using PHI-BLAST and the LF-specific consensus sequence of MAPKKs [27]. Two potential targets have been identified, both of which are human hypothetical proteins: Protein 1 (gi:37181688) is homologous (FFAS score = −44.8) to the Ran GTPase-activating protein, which participates in RNA processing; and Protein 2 (gi:7022460), which is similar (FFAS score = −58.3) to the transcription initiation factor 4 (IF4). Currently, experimental work is being performed to verify the cleavage of these proteins by LF and to identify potential additional targets of LF in the human nucleome.

This analysis suggests a second possible mechanism of bacterial virulence factor adaptation to the human organism: the evolution of the non-catalytic portion of LF [nuclear localization signal (NLS) signatures, granin signature, spectrin repeats] to be transported into the nucleus and to screen the catalytic domain and binding site from the potential recognition by human cellular defences.

Functional convergence with host

The first two examples described previously reveal a mechanism of bacterial adaptation to a human organism by convergent, function-driven evolution of specific molecular features of virulence factors. Although bacterial proteins have diverged from eukaryotic proteins over millions of years, localized functional features enable fast adaptation to the life in the respective host. The virulent proteases described here are the most prominent examples of such evolution. The anthrax lethal factor, the virulent protease and the main anthrax toxin, is a more extreme example of bacterial proteins adopting typically human folds to mask its protease domain and to become inserted in a specific cell compartment. Convergent functional features would enable the IgA1 protease to block the human inflammation pathway by mimicking the thrombin active site. It would also allow mycosin-1 to attain furin substrate specificity to digest proteins that are integral for humoral defence. At the same time, when full protein sequences are used in the phylogenetic analysis, these proteases fit perfectly into the bacterial phylogeny, suggesting that no lateral transfer took place and that the functional convergence was achieved by small, localized changes. Clearly, natural selection favoured pathogens that could quickly develop the human-like specificity of secreted proteins and the subsequent ability to manipulate the human pathway. It is tempting to speculate that this mechanism could also be used in other virulence-related genes.

It should be noted that the functional consequences of the convergent features of the virulent proteases presented here are still hypothetical and are only now being tested experimentally (Box 1).

Box 1. Present and future experimental work

There are two outstanding questions that are currently being resolved. First, how general is the trend of virulent proteins from pathogenic bacteria to mimic host proteins to survive? Second, can our computerized predictions-based hypothesis be validated experimentally? The first question is being answered by further computational studies to identify more examples of mimicking of host proteins by bacterial pathogens. Further experimental work that has been performed or is underway answers the question of the validity of the hypothesis presented previously.

To answer the question of the validity of anthrax lethal factor predictions, we have been able to prove the nuclear localization of LF and PA proteins, hypothesized earlier. Immunofluorescent studies of LF and PA trafficking identified nuclear localization of these proteins. Experimental studies were taken one step beyond the validation of the earlier hypothesis and also suggest the probable role of LFin the nucleus of human cells.

The hypothesis on convergent evolution of mycosin-1 with the human furin protein is currently being tested. The main objective of this experimental work is to determine the currently unknown role of the mycosin-1 protein in tuberculosis infection and to compare mycosin DNA sequences of various pathogenic and non-pathogenic strains of *Mycobacterium tuberculosis.* This work is being performed in collaboration with the A. Sloutsky laboratory, Massachusetts State Laboratory Institute (http://www.massparks.net/dph/bls/labsite.htm).

Finally, the hypothesis on IgA1 protease from *Haemophilus influenzae* mimicking thrombin active site and its ability to digest the N-terminal part of the PAR-1 receptor can also be tested experimentally. Cloned IgA1 protease can be tested for its ability to cleave PAR-1 N-terminal peptides *in vitro*.

References

1 Ascenzi, P. *et al*. (2002) Anthrax toxin: a tripartite lethal combination. *FEBS Lett.* 531, 384–388

2 Brown, G.D. *et al*. (2000) The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases. *Gene* 254, 147–155

3 Ansarah-Sobrinho, C. and Moss, B. (2004) Role of the I7 protein in proteolytic processing of vaccinia virus membrane and core components. *J. Virol.* 78, 6335–6343

4 Stebbins, C.E. and Galan, J.E. (2001) Structural mimicry in bacterial virulence. *Nature* 412, 701–705

5 Roy, C.R. and van der Goot, F.G. (2003) Eukaryotic cells and microbial pathogens: a familiar couple take centre stage. *Nat. Cell Biol.* 5, 16–19

6 Ahmadian, M.R. *et al*. (2002) The actin filament architecture: tightly regulated by the cells, manipulated by pathogens. International Titisee Conference on the actin cytoskeleton: from signalling to bacterial pathogenesis. *EMBO Rep.* 3, 214–218

7 Galan, J.E. and Zhou, D. (2000) Striking a balance: modulation of the actin cytoskeleton by *Salmonella*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8754–8761

8 Hayward, R.D. and Koronakis, V. (2002) Direct modulation of the host cell cytoskeleton by *Salmonella* actin-binding proteins. *Trends Cell Biol.* 12, 15–20

9 Stebbins, C.E. and Galan, J.E. (2000) Modulation of host signaling by a bacterial mimic: structure of the Salmonella effector SptP bound to Rac1. *Mol. Cell* 6, 1449–1460

10 Ellison, D.W. *et al*. (2004) Invasin and beyond: regulation of Yersinia virulence by RovA. *Trends Microbiol.* 12, 296–300

11 Koga, H. *et al*. (2005) Observation of the intracellular behavior of recombinant *Yersinia pseudotuberculosis* invasin protein. *Microbiol. Immunol.* 49, 297–302

12 Van Nhieu, G.T. and Isberg, R.R. (1991) The *Yersinia pseudotuberculosis* invasin protein and human fibronectin bind to mutually exclusive sites on the alpha 5 beta 1 integrin receptor. *J. Biol. Chem.* 266, 24367–24375

13 Orth, K. (2002) Function of the Yersinia effector YopJ. *Curr. Opin. Microbiol.* 5, 38–43

14 Orth, K. *et al*. (2000) Disruption of signaling by Yersinia effector YopJ, a ubiquitin-like protein protease. *Science* 290, 1594–1597

15 Yoon, S. *et al*. (2003) Yersinia effector YopJ inhibits yeast MAPK signaling pathways by an evolutionarily conserved mechanism. *J. Biol. Chem.* 278, 2131–2135

16 Zhang, Y. *et al*. (2005) Inhibition of MAPK and NF-{kappa}B pathways is necessary for rapid apoptosis in macrophages infected with Yersinia. *J. Immunol.* 174, 7939–7949

17 Macfarlane, S.R. *et al*. (2001) Proteinase-activated receptors. *Pharmacol. Rev.* 53, 245–282

18 Chintalacharuvu, K.R. *et al*. (2003) Cleavage of the human immunoglobulin A1 (IgA1) hinge region by IgA1 proteases requires structures in the Fc region of IgA. *Infect. Immun.* 71, 2563–2570

19 Senior, B.W. *et al*. (2000) Cleavage of a recombinant human immunoglobulin A2 (IgA2)-IgA1 hybrid antibody by certain bacterial IgA1 proteases. *Infect. Immun.* 68, 463–469

20 Derian, C.K. *et al*. (2002) Thrombin regulation of cell function through protease-activated receptors: implications for therapeutic intervention. *Biochemistry (Mosc.)* 67, 56–64

21 Otto, B.R. *et al*. (2005) Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J. Biol. Chem.* 280, 17339–17345

22 Klimpel, K.R. *et al*. (1992) Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. *Proc. Natl. Acad. Sci. U. S. A.* 89, 10277–10281

23 Park, J.M. *et al*. (2002) Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* 297, 2048–2051

24 Ye, Y. and Godzik, A. (2004) FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic Acids Res.* 32, W582–W585

25 Gorr, S.U. (1996) Differential storage of prolactin, granins (Chromogranin B and secretogranin II) and constitutive secretory markers in rat pituitary GH4C1 cells. *J. Biol. Chem.* 271, 3575–3580

26 Djinovic-Carugo, K. *et al*. (2002) The spectrin repeat: a structural platform for cytoskeletal protein assemblies. *FEBS Lett.* 513, 119–123

27 Turk, B.E. *et al*. (2004) The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. *Nat. Struct. Mol. Biol.* 11, 60–66

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 279, No. 28, Issue of July 9, pp. 29320–29324, 2004
Printed in U.S.A.

Structure of the *Chlamydia* Protein CADD Reveals a Redox Enzyme That Modulates Host Cell Apoptosis*

Received for publication, February 4, 2004, and in revised form, April 12, 2004
Published, JBC Papers in Press, April 15, 2004, DOI 10.1074/jbc.M401268200

Robert Schwarzenbacher‡§, Frank Stenner-Liewen¶, Heike Liewen¶, Howard Robinson‖, Hua Yuan‡, Ella Bossy-Wetzel‡, John C. Reed‡, and Robert C. Liddington‡**

From the ‡The Burnham Institute, La Jolla, California 92037, the ¶University of the Saarland, Medical Dept. I, Homburg, 66421, Germany, and the ‖Biology Department, Brookhaven National Laboratory, Upton, New York 11973-5000

The *Chlamydia* protein CADD (Chlamydia protein associating with death domains) has been implicated in the modulation of host cell apoptosis via binding to the death domains of tumor necrosis factor family receptors. Transfection of CADD into mammalian cells induces apoptosis. Here we present the CADD crystal structure, which reveals a dimer of seven-helix bundles. Each bundle contains a di-iron center adjacent to an internal cavity, forming an active site similar to that of methane mono-oxygenase hydrolase. We further show that CADD mutants lacking critical metal-coordinating residues are substantially less effective in inducing apoptosis but retain their ability to bind to death domains. We conclude that CADD is a novel redox protein toxin unique to *Chlamydia* species and propose that both its redox activity and death domain binding ability are required for its biological activity.

Chlamydiae are obligate intracellular bacteria and the causative agents of important sexually transmitted and disabling ocular (blinding trachoma) human diseases (1). *Chlamydia* engages in a unique relationship with its host. Upon entering host cells, the parasite starts a biphasic developmental cycle from the infectious form, called an elementary body, to a non-infectious, vegetative growth form, called a reticulate body, and then eventually back to the replication-incompetent infectious form (2). After the transition back to the infectious form, the host cell dies and releases its infectious load (3). To accommodate its life cycle, *Chlamydia* may inhibit apoptosis during the early stages of infection (4, 5) and promote apoptosis at later stages (6, 7).

Recently, the *Chlamydia* protein CADD[1] has been shown to associate with tumor necrosis factor family proteins and to induce apoptosis when transfected into a variety of mammalian cell lines (8). CADD has no close homologues but does show 18% sequence identity with coenzyme PQQ (pyrrolo-quinoline-quinone) synthesis protein C (PqqC) family members, which are part of the six-step PQQ synthesis pathway in bacteria (9). However, homologues of other members of the pathway are not found in *Chlamydia* species for which genome information is available. Indeed, ectopic expression of PqqC from *Klebsiella pneumoniae* failed to cause apoptosis, demonstrating the specificity of CADD-induced cell death (8). CADD is expressed late in the infectious cycle of *Chlamydia trachomatis* and is secreted into the host cytoplasm, where it co-localizes with tumor necrosis factor receptors in the proximity of the inclusion body. Sequence comparisons had suggested that CADD contains a death domain.

Here we present the crystal structure of CADD, which reveals an iron-containing redox enzyme that bears no resemblance to death domains. Mutagenesis of the active site of CADD reduced but did not eliminate its apoptotic activity, suggesting that both its catalytic activity and death domain binding activities contribute to its biological activity.

EXPERIMENTAL PROCEDURES

Mutation, Expression, and Purification of CADD—The open reading frame encoding CADD, CT610 (GI: 3329055) from *C. trachomatis* was subcloned into pcDNA3-hemagglutinin (Invitrogen), pGEX-4T (Amersham Biosciences), pet21d (Invitrogen), and PEGFP-C2. The following mutations, Y170F (CADD-mut1) and E81A/H88A/Y170F/H174A (CADD-mut2) were introduced using the QuikChange kit (Stratagene), confirmed by DNA sequencing, subcloned into pet21d (Invitrogen), pGEX-4T, PEGFP-C2 and pDS-RED-C2, and transformed into *Escherichia coli* XLBlue. Glutathione *S*-transferase (GST) fusion proteins were obtained by induction with 0.1 mM isopropyl-β-D-thiogalactopyranoside at 25 °C for 8 h and then purified using glutathione-Sepharose (Amersham Biosciences). After thrombin cleavage, CADD was further purified on an s200 gel filtration column (Aekta-FPLC, Amersham Biosciences), concentrated to 12 mg/ml (AMICON), and flash-frozen in liquid nitrogen for long term storage at −80 °C. The selenomethionine-substituted protein was expressed as described (10) and purified as for the wild type, except that 5 mM tris(2-carboxyethyl)phosphine was added to the dialysis and gel filtration buffers.

Crystallization—Purified CADD was crystallized by the vapor diffusion method at room temperature using a sparse matrix screen (Hampton). Sitting and hanging drops consisting of 3 μl of precipitant solution (10% (v/v) polyethylene glycol 12000, 20 mM cacodylate, pH 6.5) and 3 μl of protein solution (12 mg/ml protein) yielded crystals within 3–5 days. Crystals grew as very thin plates with dimensions of 200 × 200 × 20 μm in space group $C222_1$. The crystal structure was determined by a selenium MAD experiment using a seleno-methionine substituted protein (10). For data collection, crystals were transferred into cryobuffer (crystallization buffer with 25% (v/v) glycerol) and flash-cooled in liquid nitrogen.

Data Collection, Structure Solution, and Refinement—The three-wavelength MAD data set was collected from one single crystal, using synchrotron radiation at beamline X12B of the National Synchotron

---

* This work was supported by United States Army Medical Research and Materiel Command Grant DAMD17-03-2-0038 (to R. C. L.) and National Institutes of Health Grant CA69381 (to J. C. R.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1RCW) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

§ Supported by Fonds zur Förderung der Wissenschaftlichen Forschung fellowship J2209-B04. To whom correspondence may be addressed: The Burnham Institute, 10901 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-822-3637; E-mail: roberts@sdsc.edu.

** To whom correspondence may be addressed: The Burnham Institute, 10901 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-646-3100; E-mail: rlidding@burnham.org.

[1] The abbreviations used are: CADD, *Chlamydia* protein associating with death domains; MAD, multiwavelength anomalous diffraction; GFP, green fluorescent protein; EGFP, enhanced GFP; GST, glutathione *S*-transferase; r.m.s.d., root mean square deviation; MMOH, methane monooxygenase; ICP-AAS, inductively coupled atomic absorption spectroscopy; PQQ, pyrrolo-quinoline-quinone; wt, wild type.

TABLE I
Crystallographic statistics of CADD

	Native	Se-λ_1	Se-λ_2	Se-λ_3
Data collection				
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Cell dimensions (Å)				
a	77.55	77.63	77.62	77.51
b	192.97	193.33	193.38	193.70
c	93.74	93.99	93.97	94.11
National Synchotron Light Source beamline	X9B	X12B	X12B	X12B
Wavelength (Å)	0.954	0.9793	0.9785	0.9611
Resolution (Å)	95–2.5	30–3.1	30–3.1	30–3.1
Reflections (observed)	98,532	95,734	88,847	93,278
Reflections (unique)	24,069	13,117	13,079	13,212
Completeness (%)	96.8 (94.2)	99.8 (100.0)	99.7 (100.0)	99.7 (99.8)
$I/\sigma(I)$	11.2 (2.3)	7.1(1.4)	7.4 (1.5)	6.3 (1.2)
R_{merge} [a] (%)	6.1 (40.1)	15.1 (49.3)	15.7 (53.6)	18.1 (62.3)
Phasing (MAD)				
Resolution range	30–3.5			
Number of selenium sites	12			
Figure of merit	0.44			
Refinement				
Resolution range (Å)	95–2.5			
R_{cryst} [b] (%)	21.66			
R_{free} [c] (%)	25.85			
Protein atoms	5,189			
Iron atoms	6			
Solvent molecules	176			
r.m.s. deviations				
Bond angles (°)	1.51			
Bond lengths (Å)	0.015			

[a] $R_{\text{merge}} = \Sigma I - [I] / \Sigma I$, where I is the observed intensity and $[I]$ is the average intensity from multiple observations of symmetry-related reflections, values in parentheses correspond to the highest resolution shell.

[b] $R_{\text{cryst}} = \Sigma|F_o - F_c|/\Sigma|F_o|$.

[c] R_{free} = same as R_{cryst} but comprises a test set (5% of total reflections), which was not used in model refinement.

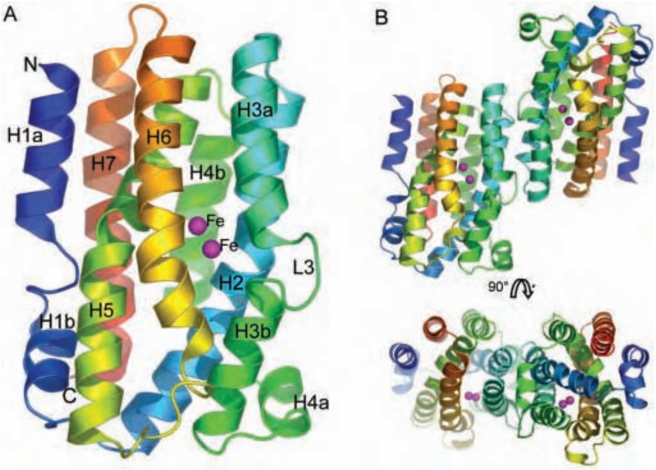


FIG. 1. **The overall structure of CADD.** *A*, CADD depicted in *ribbon representation*, rainbow color-coded from N terminus (*blue*) to C terminus (*red*), with helices H1–H7, the two iron ions, and loop L3 labeled. *B*, the CADD dimer is shown normal and parallel to its long axis.

Light Source. Oscillation data were recorded in frames of 1° through a continuous angular range of 120° for the peak (λ = 0.9791 Å), the high energy remote (λ = 0.925 Å), and the inflection point (λ = 0.9794 Å). The native data set was collected at beamline X9B of National Synchotron Light Source. All data were processed with the programs DENZO and SCALEPACK (11). The CADD structure was phased and traced using the program SOLVE/RESOLVE (12). Model building and refinement were carried out in O (13) and REFMAC5 (14). The final CADD model comprises three protein monomers (residues A7–A219, B7–B219, C7–C219), 6 Fe^{2+} ions with 3 closely bound putative water molecules, and 176 water molecules. Residues 1–6 and 220–231 were not visible in the electron density maps and therefore were not included in the model. Statistics for data collection, refinement, and model quality are summarized in Table I. Surface calculations were carried out with the CASTP server (15) and the protein-protein-interaction server (16). Figures were drawn with SPOCK (17) and PYMOL (DeLano Scientific LLC).

Cell Culture, Transfections, and Apoptosis Measurements—HeLa cells were maintained in Dulbecco's modified Eagle's medium (Irvine Scientific) and supplemented with 10% fetal bovine serum, 1 mM L-glutamine, and antibiotics. Cells (10^6) were transfected with PEGFP-C2 plasmids containing CADDwt, CADD-mut1, and CADD-mut2, using LipofectAMINE (Invitrogen) following the vendor's protocol. Both floating and adherent cells were recovered 1 day later and pooled, and the percentage of transfected (green fluorescent) cells with nuclear apoptotic morphology was determined by staining with 0.1 μg/ml 4′,6-diamidino-2-phenylindole (mean ± S.D.; n = 3). Cytosolic extracts from HeLa cells were subjected to immunoblotting and probed with rabbit polyclonal anti-green fluorescent protein (GFP) antibody (Invitrogen) for the presence of GFP-CADD fusion proteins.

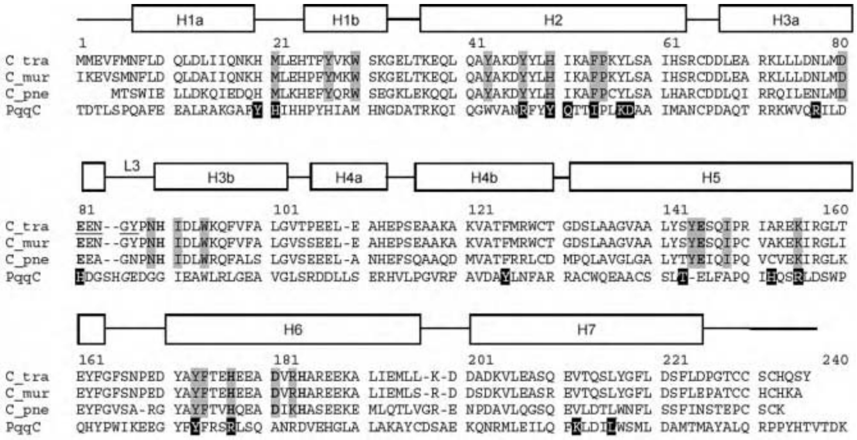
Protein Binding Assays—A plasmid containing DR5 was *in vitro* transcribed and translated in the presence of L-[^{35}S]methionine using the TNT kit from Promega. GST-CADD, GST-CADD-mut1(data not shown), GST-CADD-mut2, and control GST-CD40 (cytosolic domain) fusion proteins were immobilized on glutathione-Sepharose at 1 μg/μl and incubated with *in vitro* translated target proteins for 2 h at 4 °C. Beads were then washed four times in 1 ml of 140 mM KCl, 20 mM Hepes, pH 7.5, 5 mM $MgCl_2$, 2 mM EGTA, 0.5% Nonidet P-40, and analyzed by SDS-PAGE/fluorography.

Mass Spectrometry and ICP-AAS—Matrix-assisted laser desorption/ionization-time of flight, peptide mapping, and ICP-AAS-spectrometric analysis on the purified CADD protein were accomplished using standard techniques at the Facility for Mass Spectrometry at the Scripps Research Institute in La Jolla.

Coordinates—Coordinates and structure factors for CADD have been deposited with the Protein Data Bank (www.rcsb.org/pdb) under accession code 1RCW.

RESULTS

CADD Structure—Recombinant CADD from *C. trachomatis* was expressed in *E. coli*, purified, and crystallized. The crystal structure was determined by a selenium MAD experiment (10). CADD is a 231-residue protein, molecular mass = 26,734 Da, which forms a homo-dimer in solution, as judged by gel filtration. The CADD monomer is cylindrical with approximate dimensions of 45 × 29 × 37 Å. CADD folds into a seven-helix



```
        1                     21                    41                    61                   80
C tra   MMEVFMNFLD QLDLIIQNKH MLEHTFYVKW SKGELTKEQL QAYAKDYYLH IKAFPKYLSA IHSRCDDLEA RKLLLDNLMD
C_mur   IKEVSMNFLD QLDAIIQNKH MLEHPFYMKW SKGELTKEQL QAYAKDYYLH IKAFPKYLSA IHSRCDDLEA RKLLLDNLMD
C_pne      MTSWIE LLDKQIEDQH MLKHEFYQRW SEGKLEKQQL QAYAKDYYLH IKAFPCYLSA LHARCDDLQI RRQILENLMD
PqqC    TDTLSPQAFE EALRAKGAFY HIHHPYHIAM HNGDATRKQI QGWVANRFYY QTTIPLKDAA IMANCPDAQT RRKWVQRILD

        81                    101                   121                   141                  160
C_tra   EEN--GYPNH IDLWKQFVFA LGVTPEEL-E AHEPSEAAKA KVATFMRWCT GDSLAAGVAA LYSYESQIPR IAREKIRGLT
C_mur   EEN--GYPNH IDLWKQFVFA LGVSSEEL-E AHEPSEAAKA KVATFMRWCT GDSLAAGVAA LYSYESQIPC VAKEKIRGLI
C_pne   EEA--GNPNH IDLWRQFALS LGVSEEEL-A NHEFSQAAQD MVATFRRLCD MPQLAVGLGA LYTYEIQIPQ VCVEKIRGLK
PqqC    HDGSHGEDGG IEAWLRLGEA VGLSRDDLLS ERHVLPGVRF AVDAYLNFAR RACWQEAACS SLT-ELFAPQ IHQSRLDSWP

        161                   181                   201                   221                  240
C_tra   EYFGFSNPED YAYFTEHEEA DVRHAREEKA LIEMLL-K-D DADKVLEASQ EVTQSLYGFL DSFLDPGTCC SCHQSY
C_mur   EYFGFSNPED YAYFTEHEEA DVRHAREEKA LIEMLS-R-D DSDKVLEASR EVTQSLYGFL DSFLEPATCC HCHKA
C_pne   EYFGVSA-RG YAYFTVHQEA DIKHASEEKE MLQTLVGR-E NPDAVLQGSQ EVLDTLWNFL SSFINSTEPC SCK
PqqC    QHYPWIKEEG YFYFRSRLSQ ANRDVEHGLA LAKAYCDSAE KQNRMLEILQ FKLDILWSML DAMTMAYALQ RPPYHTVTDK
```

FIG. 2. **Multiple sequence alignment of CADD proteins from different bacterial sources.** Shown are sequences of CADD from *C. trachomatis* (*C tra*, NP_220127.1; GI:15605341) *Chlamydia muridarum* (*C_mur*, NP_297273.1; GI:15835514), *Chlamydia pneumoniae J138* (*C_pne*, NP_300818.1; GI:15836294), and PqqC Eubacterium *K. pneumoniae* (P27505 GI:130800). Helices and loops as observed in the structure of CADD are indicated *above* its sequence. Regions participating in the dimer interface are *underlined*. Metal-coordinating residues are *bold*, and residues lining the internal cavity are highlighted in *gray*. For comparison, the PqqC active site residues are shown in *white letters* with a *black background*.

mostly parallel/anti-parallel bundle, where six α-helices (H1, H2, H3, H4, H5, H7) partly embrace the seventh helix (H6) (see Fig. 2*A*). According to the Structural Classification of Proteins Data Base (18), CADD belongs to the "heme-oxygenase" fold. Helices H1, H3, H4 are kinked and can therefore be represented as separate shorter α-helices denoted as A and B. This is especially true for helix H3, where a hairpin loop, residues 82–87, is inserted (Figs. 1*A* and 2).

The CADD dimer is formed through an interaction via helices H2 and H3A, residues 59–85 (Figs. 1*B* and 2). The interface-accessible surface area is 915 Å^2/monomer, which accounts for 9.2% of the accessible surface area of the CADD dimer. The interaction is predominantly hydrophobic (55% non-polar atoms) but also includes a number of polar interactions and salt bridges. The most similar structures found using the DALI server (19) are: PqqC (20), with an r.m.s.d. of 2.8 Å for the superposition of 221 Cα atoms and 18% sequence identity; human heme-oxygenase (21), with an r.m.s.d. of 2.9 Å for 199 Cα atoms and 11% sequence identity; the R2 subunit of ribonucleotide reductase (R2-RNR) (22), with an r.m.s.d. of 3.2 Å for 178 Cα atoms and 12% sequence identity; and the α-subunit of methane monooxygenase (MMOH) (23), with an r.m.s.d. of 3.1 Å for 174 residues and 9% sequence identity. Although none of the active sites are conserved, each of these enzymes appears to be a redox enzyme, suggesting that this fold is particularly suitable for this type of enzyme. According to sequence similarity searches with the bioinformatics server Fold and Function Assignment System (24), CADD shares distant sequence homology with transcription enhancement gene A transcription factors (25) and can be used as a template to obtain homology models for these proteins.

The Active Site—The seven helices of CADD provide the scaffold for a narrow internal cavity equipped with a di-metal center (Figs. 2 and 3*A*). The experimental electron density map clearly indicates the presence of two metal ions coordinated by 6 residues (Glu-81, His-88, Glu-142, His-174, Asp-178, His-181) (Figs. 2 and 3*A*). The di-metal site is located in the center of the molecule adjacent to the cavity, which most likely serves as the active site. Atomic absorption measurements using ICP-AAS revealed the presence of iron and small but significant amounts of zinc in the protein. This indicates the presence of a di-iron site, which, judged by difference maps and elevated *B*-factors, is not fully occupied in the CADD crystals. The small amounts of zinc might be due to oxidation and partial replacement of iron for zinc, which has been observed in several crystal structures of di-iron-containing proteins (26). The di-metal center appears to be octahedrally coordinated and bridged by a glutamate residue (Glu-81) and a water molecule or hydroxide ion. Fe1 is coordinated by two histidines (His-174, His-88) and the glutamate (Glu-81), as well as the putative water, which it shares with Fe2. Fe2 is coordinated by histidine (His-181), two glutamates (Glu-142, Glu-81), aspartate (Asp-178), and the bridging water molecule (Fig. 3*A*). All 6 active site residues coordinating the metal ions are strictly conserved among CADD proteins from *Chlamydia* species (Fig. 2). The water molecule or hydroxide ion is coordinated by both iron atoms at a distance of 2.2 Å (Figs. 3*A* and 4*B*). It is adjacent to Asp-178, His-174, and Tyr-170 and faces the internal active site cavity. The elliptically shaped density (3σ peak in $2F_o - F_c$ map) for the water molecule/hydroxide ion is obscured by the electron-rich iron atoms nearby and therefore not unambiguously interpretable. The electron density and resulting *B*-factors are also consistent with a reactive oxygen species bound to the di-iron site. The cavity next to the di-iron site shows an overall positive charge and measures 5 × 7 × 14 Å, with a volume of 340 Å^3 (15). The cavity is lined with 15 conserved hydrophilic or aromatic residues (His-50, Ile-51, Phe-54, Glu-81, Ile-89, Glu-142, Asp-178, Tyr-141, Ile-145, Lys-152, Tyr-170, His-174, Pro-55, Glu-82, and Asn-87 (Fig. 2)). Below this cavity, the hydrophobic

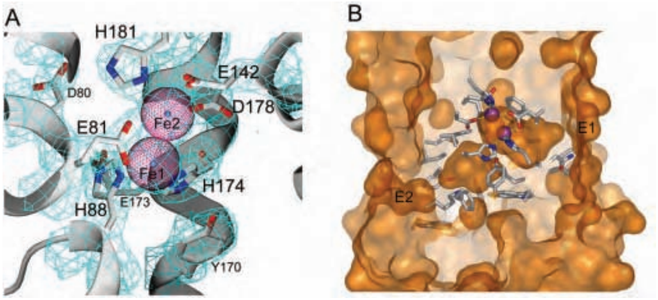


FIG. 3. **Active site analysis.** *A*, the di-iron site. The metal-coordinating active site residues Glu-81, His-88, Glu-142, His-174, Asp-178, and His-181 as well as the putative tyrosyl radical carrying Tyr-170 and the putative OH$^-$ molecule are depicted in *ball and stick* format. The $2F_o - F_c$ electron density map is contoured at 1.5 σ. *B*, close-up view of the CADD molecule in a transparent surface representation (*orange*) showing the internal cavities, the di-metal site (*purple spheres*), and surrounding residues in *ball and stick* format. The two possible entrances are located on the *right side* of the metal site (E1) next to loop L3 and on the *bottom* of the molecule (E2) close to helix H1b-H6.

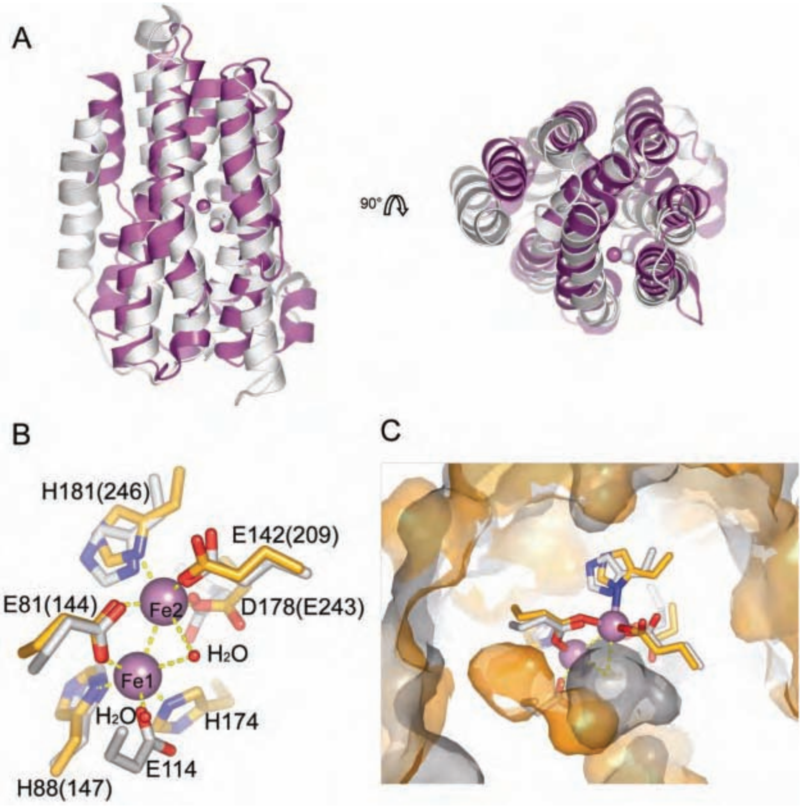


FIG. 4. **Comparisons of CADD and MMOH.** *A*, superposition of CADD (*purple*) and the α-subunit of MMOH (Protein Data Bank accession code 1mhyD; *white*) in *ribbon representation*. The metal ions in the active site are shown as *spheres* colored *purple* (CADD) and *white* (MMOH). *B*, same as *A* showing a superposition of the two di-iron sites. The Fe^{2+} ions are shown as *purple spheres* with the coordinating residues in sticks (CADD in *orange*, 1mhyD in *white*). *C*, comparison of the active site cavities in transparent surface representation from CADD (*orange*) and 1mhyD (*gray*). The di-metal site is shown for reference.

core is largely aromatic and also contains a buried lysine (Lys-152). A system of cavities spans across the core of the molecule, with two potential openings next to loop L3 and between helices H1B and H5. One opening, E1, penetrates the surface of the protein between helices H2, H3, and the unique loop L3 (Figs. 1*A* and 3*B*). It is lined by residues Ile-51, Pro-55, Ile-89, and Glu-82. An alternative access path, E2, leads from the di-iron site through a narrow opening into a second cavity lined by residues Met-21, Tyr-43, Tyr-47, Trp-92, Ile-148, Ala-149, Phe-171, Ala-149, Lys-152, and Tyr-27 and from there to the surface next to residues Trp-30 and Asp-151 (helices H1B and H5). The size of the active site cavity openings restricts the substrates to small compounds such as O_2, H_2O_2, CH_4, CH_3OH, CO, or CO_2. Larger molecules could only pass through by means of a conformational change.

The active site of CADD is similar to that found in RNR-R2 from *E. coli* (Protein Data Bank accession code 1xsm). The helices forming the core that contains the active site can be superimposed with an r.m.s.d. of 2.8 Å. The function of RNR-R2 is to generate a tyrosyl radical on an adjacent tyrosine with the help of its di-iron center. The organic free radical is transferred to the RNR-R1 subunit, which catalyzes the *de novo* production of deoxy nucleotides (22). Interestingly, CADD also contains a tyrosine (Tyr-170) next to the di-iron center. These similarities raise the question of whether the physiological function of CADD is the production of radicals for RNR-R1. However, no equivalents are seen for Asp-84, Asp-237, and Trp-48, which are critical residues for the radical initiation pathway proposed in RNR-R2 (Tyr-122-Asp-84-Fe1-His-118-Asp-237 to Trp-48) (26). Taken together, these findings indicate that CADD cannot function as a RNR-R2 but might use a tyrosyl-radical for catalysis.

Cellular Activity of CADD—To test whether Tyr-170 and the di-metal site are involved in the toxicity of CADD, we generated two active site mutants by PCR mutagenesis and tested their apoptotic activity through transfection experiments in mammalian cells. The role of Tyr-170 was tested with a Y170F (CADD-mut1) mutant. To prevent the formation of a functional di-metal center, we made the quadruple mutant of the metal-coordinating residues: E81A/H88A/H174A/Y170F (CADD-mut2). When equivalents of each plasmid DNA were transfected into HeLa cells, CADD-mut1 showed a decrease in toxicity of about 5–15% when compared with the wild-type. CADD-mut2 showed more than 60% reduction in apoptotic activity (Fig. 5*A*). Immunoblotting shows (Fig. 5*B*) that both CADD mutants are expressed at similar or higher levels to wild-type. This indicates that the mutants, especially CADD-mut2, are better tolerated by the transfected mammalian cells than the wild-type. To address the question of whether the active site mutant proteins still bind to death receptors, we carried out an *in vitro* DR5 binding assay (Fig. 5*C*), comparing GST-CADD-wt, GST-CADD-mut1 (data not shown), and GST-CADD-mut2. CADD wild type and active site mutants show comparable binding to death receptor DR5, indicating that the active site mutations do not alter the DR5 binding activity of CADD.

DISCUSSION

The crystal structure shows that CADD shares similarity to heme-oxygenase and PqqC enzymes. The sequence similarity and "PqqC-like" annotation for CADD proteins are reflected by the same fold, but the active sites are not conserved, and the two proteins are therefore functionally and most likely also evolutionarily unrelated (20). CADD is consequently an orphan unique to *Chlamydia* species, which further emphasizes its role as a highly specific toxin that evolved in this intracellular parasite. Comparison with the more distant structural homologues, RNR-R2 (22) and MMOH (23), reveals di-iron active sites in a strikingly similar structural context. Although the three proteins belong to different fold subclasses (CADD shows the heme-oxygenase fold, whereas RNR-R2 and MMOH belong to the ferredoxin fold), the helices forming the core containing the active site can be superimposed with an r.m.s.d. <2.8 Å. The active site of CADD is structurally similar to that in

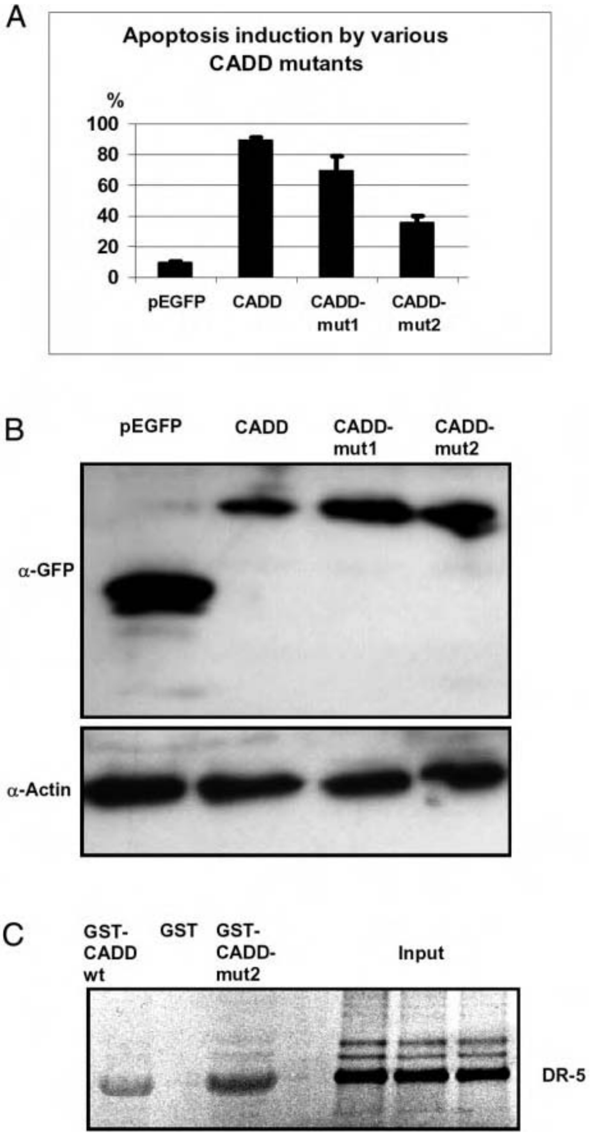


FIG. 5. **Cellular activity of CADD mutants.** *A*, *in vivo* cell death assay using HeLa cells, comparing CADD-wt, CADD-mut1, and CADD-mut2. At 1 day after transfection, the percentage of GFP-positive apoptotic cells was determined by 4′,6-diamidino-2-phenylindole staining (mean ± S.D.; $n = 3$). Note that the active site mutant 2 displays the lowest toxicity in independent experiments ($n = 3$). Plasmid DNA was transfected in equivalent amounts, resulting in a similar percentage of GFP-positive cells (not shown). *B*, a representative immunoblot analysis of cell lysates from transfectants is shown. Samples were normalized for total protein content, and GFP fusion proteins were detected with anti-GFP antibody by an ECL method. Note that higher levels of GFP-CADD mut1 and mut2 suggest that cells tolerate the mutants more, when compared with CADD-wt (*upper panel*). The blot was reprobed with an antibody recognizing α-actin to confirm loading of equivalent amounts of cellular protein (*lower panel*). *C*, *in vitro* DR5 binding assay, comparing GST-CADD-wt, GST (control), and GST-CADD-mut2: E81A, H88A, Y170F, H174A. The active site mutant GST-CADD-mut2 bound death receptor DR5 to a similar degree when compared with the wild type. GST (control) shows no binding with DR5.

RNR-R2 but does not contain the conserved residues of the radical pathway. CADD can therefore not serve as an RNR-R2, but it is tempting to speculate that CADD, like RNR-R2, may generate and use a free tyrosyl radical on Tyr-170 to facilitate redox reactions. However, mutagenesis studies with an Y170F mutant show only a 5–15% decrease in toxic activity, indicating that Tyr-170 is not essential for CADD function. The central cavity of CADD contains several tyrosines, and it is possible that another one (Tyr-47, Tyr-141) may substitute for the loss of Tyr-170.

A structural comparison with the di-iron center in MMOH from *Methylococcus capsulatus* (Protein Data Bank accession code 1mhyD) (23) reveals strong conservation of the metal-coordinating residues, except for a difference in the coordination of Fe1 in CADD, where Glu-114 is replaced on the other side of Fe1 with His-174 (Fig. 4, *A* and *B*). A detailed analysis of the active sites further reveals that in contrast to RNR-R2, MMOH and CADD contain an internal cavity next to the di-iron center (Fig. 4*C*). In MMOH, the cavity functions as the site of catalysis, where substrate and product access the di-iron center through the tunnel-like cavity from the bottom of the molecule. CADD contains a similar tunnel when the entrance next to Trp-92, between H1B and H5, is used (Fig. 3*B*). On the other hand, the opening next to the loop L3 is a potential region for a conformational change that could open the cavity to the outside for the exchange of substrate and product. Thus, CADD is most likely an enzyme similar to MMOH (23), which uses an internal active site equipped with a di-iron center to catalyze redox reactions on small molecule substrates. Further biochemical studies are needed to determine the reaction catalyzed by CADD.

Transfection assays with a CADD mutant lacking critical metal-coordinating residues establish a direct connection between the di-iron site and the apoptotic activity of CADD. Alterations at the active site, which is buried within the molecule, do not abolish interaction with death receptors, which suggests that the optimal induction of apoptosis by CADD requires both the intracytoplasmic cross-linking of death receptors as well as its redox activity.

Acknowledgment—We thank Jose Maria de Pereda for valuable discussions and helpful suggestions.

REFERENCES

1. Antilla, T. (2001) *J. Am. Med. Assoc.* **285**
2. Schachter, J. (1988) *Curr. Top. Microbiol. Immunol.* **138,** 109–139
3. Belland, R. J., Scidmore, M. A., Crane, D. D., Hogan, D. M., Whitmire, W., McClarty, G., and Caldwell, H. D. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 13984–13989
4. Fischer, S. F., Schwarz, C., Vier, J., and Hacker, G. (2001) *Infect. Immun.* **69,** 7121–7129
5. Geng, Y., Shane, R. B., Berencsi, K., Gonczol, E., Zaki, M. H., Margolis, D. J., Trinchieri, G., and Rook, A. H. (2000) *J. Immunol.* **164,** 5522–5529
6. Schoier, J., Ollinger, K., Kvarnstrom, M., Soderlund, G., and Kihlstrom, E. (2001) *Microb. Pathog.* **31,** 173–184
7. Perfettini, J. L., Hospital, V., Stahl, L., Jungas, T., Verbeke, P., and Ojcius, D. M. (2003) *Biochimie (Paris)* **85,** 763–769
8. Stenner-Liewen, F., Liewen, H., Zapata, J. M., Pawlowski, K., Godzik, A., and Reed, J. C. (2002) *J. Biol. Chem.* **277,** 9633–9636
9. Meulenberg, J. J., Sellink, E., Loenen, W. A., Riegman, N. H., van Kleef, M., and Postma, P. W. (1990) *FEMS Microbiol. Lett.* **59,** 337–343
10. Harrison, C. J., Bohm, A. A., and Nelson, H. C. (1994) *Science* **263,** 224–227
11. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
12. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 849–861
13. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
14. CCP4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
15. Liang, J., Edelsbrunner, H., and Woodward, C. (1998) *Protein Sci.* **7,** 1884–1897
16. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13–20
17. Christopher, J. A. (1998) *SPOCK*, The Center for Macromolecular Design, Texas A&M University, College Station, TX
18. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.* **247,** 536–540
19. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
20. Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7913–7918
21. Schuller, D. J., Wilks, A., Ortiz de Montellano, P. R., and Poulos, T. L. (1999) *Nat. Struct. Biol.* **6,** 860–867
22. Eriksson, M., Jordan, A., and Eklund, H. (1998) *Biochemistry* **37,** 13359–13369
23. Rosenzweig, A. C., Brandstetter, H., Whittington, D. A., Nordlund, P., Lippard, S. J., and Frederick, C. A. (1997) *Proteins* **29,** 141–152
24. Jaroszewski, L., Li, W., and Godzik, A. (2002) *Protein Sci.* **11,** 1702–1713
25. Pang, A. S., Nathoo, S., and Wong, S. L. (1991) *J. Bacteriol.* **173,** 46–54
26. Stubbe, J., Nocera, D. G., Yee, C. S., and Chang, M. C. (2003) *Chem. Rev.* **103,** 2167–2201

Quinone biogenesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone

Olafur Th. Magnusson*, Hirohide Toyama†, Megumi Saeki†, Ana Rojas‡, John C. Reed‡, Robert C. Liddington‡, Judith P. Klinman*, and Robert Schwarzenbacher‡§

*Departments of Chemistry and of Molecular and Cell Biology, University of California, Berkeley, CA 94720; †Department of Biological Chemistry, Faculty of Agriculture, Yamaguchi University, Yamaguchi 753-8515, Japan; and ‡The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037

Contributed by Judith P. Klinman, April 14, 2004

The biosynthesis of pyrroloquinoline quinone (PQQ), a vitamin and redox cofactor of quinoprotein dehydrogenases, is facilitated by an unknown pathway that requires the expression of six genes, *pqqA* to *-F*. PqqC, the protein encoded by *pqqC*, catalyzes the final step in the pathway in a reaction that involves ring cyclization and eight-electron oxidation of 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic-acid to PQQ. Herein, we describe the crystal structures of PqqC and its complex with PQQ and determine the stoichiometry of H_2O_2 formation and O_2 uptake during the reaction. The PqqC structure(s) reveals a compact seven-helix bundle that provides the scaffold for a positively charged active site cavity. Product binding induces a large conformational change, which results in the active site recruitment of amino acid side chains proposed to play key roles in the catalytic mechanism. PqqC is unusual in that it transfers redox equivalents to molecular oxygen without the assistance of a redox active metal or cofactor. The structure of the enzyme-product complex shows additional electron density next to R179 and C5 of PQQ, which can be modeled as O_2 or H_2O_2, indicating a site for oxygen binding. We propose a reaction sequence that involves base-catalyzed cyclization and a series of quinone-quinol tautomerizations that are followed by cycles of O_2/H_2O_2-mediated oxidations.

Pyrroloquinoline quinone [4,5-dihydro-4,5-dioxo-1*H*-pyrrolo-[2,3-*f*]quinoline-2,7,9-tricarboxylic acid; PQQ (Fig. 1)] is an aromatic, tricyclic *ortho*-quinone that serves as the redox cofactor for several bacterial dehydrogenases. Among the best known examples are methanol dehydrogenase and glucose dehydrogenase (1, 2). PQQ belongs to the family of quinone cofactors that has been recognized as the third class of redox cofactors following pyridine nucleotide- and flavin-dependent cofactors (3). Although plants and animals do not produce PQQ themselves, PQQ has invoked considerable interest because of its presence in human milk and its remarkable antioxidant properties (4–6). Recently, the first potential eukaryotic PQQ-dependent enzyme [aminoadipic 6-semialdehyde-dehydrogenase (AASDH; U26)] has been identified, indicating that PQQ may function as a vitamin in mammals as well (7).

Quinone cofactors are generally covalently linked to the polypeptide chain and derived posttranslationally from precursor amino acid residues encoded within their parental polypeptide chain. For example, in copper amine oxidases, topaquinone is formed by a "self-processing" oxidation of a specific Tyr residue in the presence of copper ion and molecular oxygen (8, 9). PQQ is distinct from the other quinone cofactors in that its biogenesis is independent of its site of action. PQQ is constructed from the amino acids glutamate and tyrosine, as shown in Fig. 1 (10, 11). The PQQ biosynthesis pathway in *Klebsiella pneumoniae* requires the expression of six genes, designated *pqqABCDEF* (12). PqqA encodes a 23-residue peptide with conserved glutamate and tyrosine residues that most likely serves as the precursor for PQQ biosynthesis (10, 13, 14). Transformed *Escherichia coli* cells carrying a plasmid that contains the PQQ operon from *K. pneumoniae* lacking *pqqC*, as well as a *pqqC* mutant strain of *Methylobacterium extorquens* AM1, accumulate an intermediate that can be converted to PQQ upon addition of PqqC (15, 16). These results demonstrate that PqqC catalyses the last step in PQQ biosynthesis.

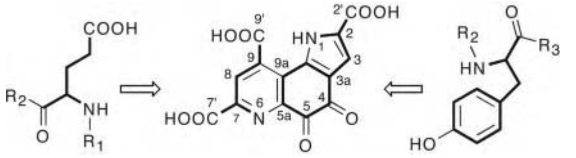


Fig. 1. Chemical structure of PQQ (4,5-dihydro-4,5-dioxo-1*H*-pyrrolo-[2,3-*f*]quinoline-2,7,9-tricarboxylic acid) with atom nomenclature. All carbon and nitrogen atoms of PQQ are derived from conserved tyrosine and glutamate residues of the PqqA peptide. R_1 and R_3 represent the N- and C-terminal portions of PqqA, respectively. R_2 represent a three-amino-acid linker between Glu and Tyr.

It has been shown that the PqqC reaction is accelerated in the presence of molecular oxygen (17) and that it requires NADPH and an uncharacterized activating factor (16) for sustained catalytic activity. Most recently, we elucidated the structure of the PqqC substrate, allowing the overall reaction catalyzed by PqqC to be inferred (18). The substrate is 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid (**1** in Fig. 5*A*), a fully reduced derivative of PQQ, which has not yet undergone ring-cyclization at the pyrrole moiety. Herein, we present the structure of PqqC in complex with PQQ. Based on the enzyme active site environment, together with the stoichiometry of O_2 uptake and H_2O_2 formation in a single-enzyme turnover, we propose a multistep catalytic mechanism for the reaction catalyzed by PqqC.

Methods

Expression and Purification of PqqC. PqqC from *K. pneumoniae* (NCBI accession no. X58778) was expressed in *E. coli* and purified as described elsewhere (19).

Single-Turnover Kinetics, Peroxide Formation, and Oxygen Uptake. The substrate for PqqC was purified from a *pqqC* knock-out mutant of *M. extorquens* AM1 as described elsewhere (17, 18). All experiments were performed in 0.1 M potassium phosphate buffer (pH 8.0) at 20°C, and all data fitting was done by using KALEIDAGRAPH (Synergy Software, Reading, PA).

Abbreviations: PQQ, pyrroloquinoline quinone; AR, Amplex red; HRP, horseradish peroxidase.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 1OTV and 1OTW).

§To whom correspondence should be addressed. E-mail: roberts@burnham.org.

www.pnas.org/cgi/doi/10.1073/pnas.0402640101

PNAS | May 25, 2004 | vol. 101 | no. 21 | 7913–7918

Reactions were initiated by addition of substrate (12.4 μM) to a solution containing enzyme (5–100 μM). Reaction mixtures were quenched at designated time points by addition of HCl (0.5 M) and analyzed by reversed phase HPLC. A Beckman HPLC system equipped with a diode-array detection system was used, and substrate and product were separated by using a Vydac (Hesperia, CA) C_{18} column (5μ, 4.6 × 250 mm). A linear gradient of 0.1% trifluoroacetic acid from 0% to 80% CH_3CN in 25 min was used. Substrate and product elute with retention times of 13.1 min and 14.6 min under these conditions, respectively. The amount of PQQ formed was determined by comparison with a standard curve of authentic material obtained from Fluka. The concentration of PQQ was determined spectrophotometrically in an aqueous solution at pH 7 (20). PQQ formation at low enzyme concentration ($\leq$1 μM) was measured by using an enzymatic assay based on the activation of glucose dehydrogenase as described elsewhere (16).

The production of H_2O_2 was assessed with an enzymatic assay by using Amplex red (AR; 10-acetyl-3,7-dihydroxyphenoxazine) and horseradish peroxidase (HRP) obtained from Molecular Probes. The HRP-mediated oxidation of the colorless AR reagent by peroxide produces a chromophore, which can be measured either colorimetrically or by fluorimetry (21). Two different methods were used. In method A, H_2O_2 was measured in a continuous fashion by using 50 μM AR and 1 unit/ml HRP in samples containing 12.4 μM substrate and 90 μM PqqC. Reactions were initiated by the addition of substrate, and H_2O_2 formation was monitored spectrophotometrically. The amount of peroxide generated in these reactions was deduced by comparison with a standard curve generated from authentic H_2O_2 treated in the same manner as the samples above. In method B, aliquots were withdrawn from reaction mixtures containing substrate (12.4 μM) and PqqC (90 μM) and quenched in HCl (0.5 M) at designated time points. Samples from each quench were diluted 60-fold, and the amount of H_2O_2 produced was measured fluorimetrically upon treatment with AR/HRP and compared with an H_2O_2 standard (excitation at 530 nm, emission at 582 nm).

Oxygen consumption was measured by using a Clark oxygen electrode (YSI model 5300; YSI Inc., Yellow Springs, OH). Reactions were initiated by adding 5 μl of substrate (12.4 μM final concentration) to 995 μl of PqqC (90 μM) that had been equilibrated for 10–20 min at 20°C to obtain a stable baseline. The effective concentration of O_2 in the protein solution was measured in a separate experiment by monitoring the consumption of oxygen during turnover of protocatechuic acid by protocatechuic dioxygenase as described elsewhere (22).

Sequence Alignment. The PqqC sequence alignment was obtained by using the program T-COFFEE (23) on a converged PSI-BLAST (24) search of the microbial genome database at the National Center for Biotechnology Information (NCBI) with the PqqC sequence (gi 130800) from *K. pneumoniae*. Sequences were clustered at 80% identity and assembled into four different groups by using program CD-HIT (25). A representative sequence of each cluster is shown to display conservation of active site residues.

Crystallization. Crystals were grown at 25°C by the sitting drop vapor diffusion method in droplets composed of one part protein solution (8 mg/ml in 20 mM Tris·HCl, at pH 8.0 and 1 mM DTT) and one part reservoir solution (1.2 M ammonium sulfate, Mes/NaOH at pH 6.0). The orthorhombic crystals are of space group $P2_12_12$ and contain two molecules in the asymmetric unit. The PqqC/PQQ complex structure was obtained by soaking PqqC crystals in a crystallization solution containing 1 mM PQQ. For data collection, these crystals were transferred into cryobuffer [crystallization buffer with 25% (vol/vol) glycerol] and flash-cooled in liquid nitrogen.

Data Collection, Structure Solution, and Refinement. The PqqC structure was determined with a selenium-MAD experiment as described (19). Native and complex datasets were collected at beamline 9.1 of the Stanford Synchrotron Radiation Laboratory and processed with the HKL suite (26). The structure of the PqqC/PQQ complex was determined by molecular replacement by using MOLREP (27) with the native PqqC structure as a search model. Crystallographic refinement and model building was performed by using REFMAC5 (27) and O (28). The PqqC model comprises two protein monomers (residues 1–249) and 130 water molecules. In both chains, the region between residues 152 and 160 showed weak density and higher B-factors, indicative of partial disorder. The model for the PqqC/PQQ complex includes two PqqC monomers (residues 1–249), two PQQ molecules, two putative H_2O_2 molecules, and 119 water molecules. The native and complex structures were solved at 2.1 Å and 2.3 Å resolution, respectively. Further information concerning data collection and refinement statistics is available in Table 1, which is published as supporting information on the PNAS web site. Figures were drawn with PYMOL (DeLano Scientific, San Carlos, CA).

Coordinates. Coordinates and structure factors for PqqC and the PqqC/PQQ complex have been deposited with the Protein Data Bank (www.pdb.org/pdb) under accession codes 1OTV and 1OTW, respectively.

Results

The PqqC Reaction. The PqqC reaction was analyzed for the production of PQQ by separation of substrate and product by HPLC. PqqC from *K. pneumoniae* produces 1 mol of PQQ per mol of enzyme in a single turnover. The reaction displays first-order kinetics for PQQ formation with regard to substrate under conditions of both excess and substoichiometric enzyme (data not shown). Even at low PqqC concentration (0.1 μM) in the presence of excess substrate, only one enzyme turnover could be detected by using a sensitive assay based on activation of the apo-form of glucose dehydrogenase. Under saturating conditions where all of the substrate is enzyme bound, the observed rate constant at 20°C is 0.38 ($\pm$ 0.03) min^{-1} (Fig. 2, curve A).

To probe the putative role of O_2 in the reaction, we looked for the consumption of molecular oxygen and production of hydrogen peroxide. As shown in Fig. 2, curve B, the rate of O_2 uptake ($k_{\text{obs}} = 0.38 \pm 0.02$ min^{-1}) is similar to the rate of PQQ formation. The measured stoichiometry of the reaction shows that the enzyme consumes $\approx$3 mol of O_2 per mol of PQQ produced (2.77 ± 0.49). The production of H_2O_2 was measured by two different methods. With AR and HRP present in the PqqC assay mixture, approximately one equivalent (0.89 ± 0.12) of H_2O_2/PQQ was detected (Fig. 2, curve C). However, in a discontinuous assay, in which H_2O_2 was measured after acid denaturation of the protein, approximately two equivalents (1.86 ± 0.23) of H_2O_2/PQQ were formed (Fig. 2, curve D). Both methods yield identical first-order rate constants (within error) 0.40 ± 0.02 min^{-1} and 0.39 ± 0.03 min^{-1} for the continuous and discontinuous assays, respectively. The difference in the amount of H_2O_2 obtained by the two methods suggests that the enzyme tightly binds one equivalent of peroxide because denaturation of the protein is required to make the remaining peroxide accessible to detection. Note, however, that the data for the continuous assay (Fig. 2, curve C) were fit to two exponentials, in which the second slower and smaller amplitude phase may represent slow dissociation of the second equivalent of peroxide. In any case, the data show that PqqC consumes 3 mol of O_2 per PQQ

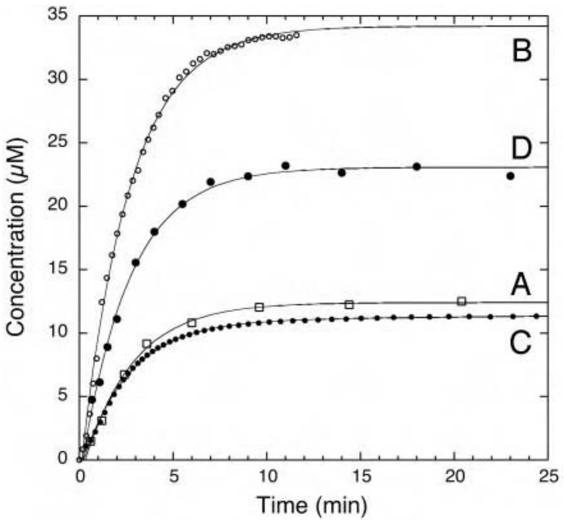


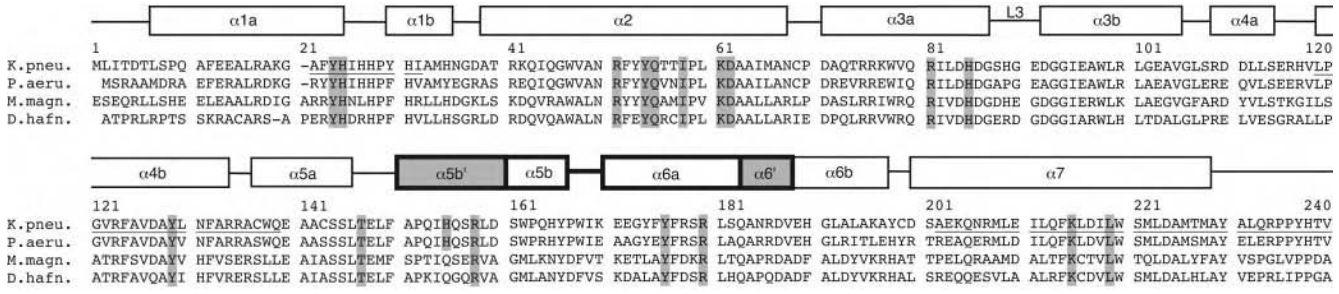
Fig. 2. Single-turnover kinetics of the PqqC reaction. Curve A illustrates formation of PQQ as determined by quantitative HPLC analysis after acid quenching at designated time points. Curve B shows dioxygen consumption measured polarographically using a Clark-type electrode. Curve C illustrates formation of H_2O_2 measured in a continuous coupled assay with AR/HRP. Curve D shows formation of H_2O_2 measured in a discontinuous assay using AR/HRP after acid quenching at designated time points. See *Results* for details.

and generates 2 mol of H_2O_2 per PQQ in a single-turnover reaction.

Overall Structure of PqqC. PqqC is a 251-residue protein (molecular mass = 28.91 kDa) that forms a homodimer in solution, as evident from gel filtration experiments (data not shown). PqqC folds into a compact seven-helix bundle, with six circular aligned helices (α1, α2, α3, α4, α5, and α7), partly embracing a seventh hydrophobic helix (α6) (Figs. 3 and 4). Analysis of the PqqC structure shows that the seven α-helices provide the scaffold for an active site cavity (19). The cavity is lined with 42 mostly hydrophilic and aromatic residues that are highly conserved within PqqC proteins from different bacteria (Fig. 4). The cavity shows a distinct overall positive charge, measures 9 Å × 13 Å × 23 Å and embraces a molecular surface volume of 2,200 Å^3 (29). Two openings connect it to the outside. A structural similarity search performed with the atomic coordinates of PqqC using the DALI server (30) yields human heme-oxygenase (31) with 9% sequence identity and a rms distance of 3.1 Å for the superposition of 194 Cα atoms.

Structure of the PqqC/PQQ Complex. The PqqC/PQQ complex structure was obtained by soaking PqqC crystals in a crystallization solution containing 1 mM PQQ. The structure shows that PQQ binds in the center of the active site cavity, accompanied by a large conformational change in the protein. The structural rearrangement is almost entirely executed in the region of helices α5a to α6b (residues 142–193) and includes the elongation of helix α5b to form α5b′ and the fusion of helices α6a and α6b into one long helix, α6′. Starting with α5a, a detailed description of the conformational change includes the following rearrangements depicted in Figs. 3 and 4. Helix α5a accommodates small shifts (≈0.7 Å) in the main-chain atoms, resulting in a slight compression of the helix. The coil region between residues 151 and 158 folds into α5b′ and closes the main entrance to the active site cavity. This rearrangement shifts H154 and R157 directly into the active site, ready to coordinate the carboxylic group C7′ of PQQ. The main chain in the region between residues 159 and 169, (from α5b to the beginning of α6a) shifts ≈1.8 Å upwards, keeping the overall main chain and side-chain conformation unchanged. The largest shift occurs in the region of helices α6a and α6b (residues 170–187, Figs. 3 and 4 *A* and *B*). Helix α6a rotates ≈90° around its long axis and shifts ≈3.5 Å upwards, 2 Å toward α5b′, and 4 Å toward the center of the molecule. This shift is facilitated by loop L6 winding up in a spring-like manner to connect helices α6a and α6b into one long helix, α6′. The Cα of Y175 moves 6.9 Å from a solvent exposed location to a position directly in the center of the molecule where it interacts with the oxygen atoms of C4 and C5 of the ligand. The new position forces the side-chain of W97 into another rotamer that is stabilized by H-bonds between the main-chain nitrogen of Y175 and a water molecule, which itself interacts with the carbonyl oxygen of A96. The same holds for R179, which moves ≈7.7 Å to interact with O4 of the substrate and the sidechains of E147 and D186. Thus, PQQ recognition in PqqC exhibits a classical induced fit mechanism, resulting in the ligand being completely buried from the solvent.

Active Site Interactions with the Ligand. PQQ binds in the center of the active site cavity coordinated by 18 highly conserved residues (Y23, H24, R50, Y53, Q54, I57, K60, D61, R80, H84, Y128, T146, H154, R157, Y175, R179, K214, and L218; see Fig. 4). Fig. 4C shows the coordination sphere of PQQ. The four residues H154, R157, Y175, and R179 move into the active site upon substrate binding (Fig. 4*B*). Critical sites of interaction between ligand and protein are the three carboxyl groups, the amino



```
          1                     21                    41                    61                    81                    101                   120
K.pneu.   MLITDTLSPQ AFEEALRAKG -AFYHIHHPY HIAMHNGDAT RKQIQGWVAN RFYYQTTIPL KDAAIMANCP DAQTRRKWVQ RILDHDGSHG EDGGIEAWLR LGEAVGLSRD DLLSERHVLP
P.aeru.    MSRAAMDRA EFERALRDKG -RYYHIHHPF HVAMYEGRAS REQIQGWVAN RFYYQVNIPL KDAAILANCP DREVRREWIQ RILDHDGAPG EAGGIEAWLR LAEAVGLERE QVLSEERVLP
M.magn.   ESEQRLLSHE ELEAALRDIG ARRYHNLHPF HRLLHDGKLS KDQVRAWALN RYYYQAMIPV KDAALLARLP DASLRRIWRQ RIVDHDGDHE GDGGIERWLK LAEGVGFARD YVLSTKGILS
D.hafn.    ATPRLRPTS SKRACARS-A PERYHDRHPF HVLLHSGRLD RDQVQAWALN RFEYQRCIPL KDAALLARIE DPQLRRVWRQ RIVDHDGERD GDGGIARWLH LTDALGLPRE LVESGRALLP

          121                   141                   161                   181                   201                   221                   240
K.pneu.   GVRFAVDAYL NFARRACWQE AACSSLTELF APQIHQSRLD SWPQHYPWIK EEGYFYFRSR LSQANRDVEH GLALAKAYCD SAEKQNRMLE ILQFKLDILW SMLDAMTMAY ALQRPPYHTV
P.aeru.   GVRFAVDAYV NFARRASWQE AASSSLTELF APQIHQSRLD SWPRHYPWIE AAGYEYFRSR LAQARRDVEH GLRITLEHYR TREAQERMLD ILQFKLDVLW SMLDAMSMAY ELERPPYHTV
M.magn.   ATRFSVDAYV HFVSERSLLE AIASSLTEMF SPTIQSERVA GMLKNYDFVT KETLAYFDKR LTQAPRDADF ALDYVKRHAT TPELQRAAMD ALTFKCTVLW TQLDALYFAY VSPGLVPPDA
D.hafn.   ATRFAVQAYI HFVRERSLLE AIASSLTELF APKIQGQRVA GMLANYDFVS KDALAYFDSR LHQAPQDADF ALDYVKRHAL SREQQESVLA ALRFKCDVLW SMLDALHLAY VEPRLIPPGA
```

Fig. 3. Multiple sequence alignment of PqqC proteins from different bacterial sources. K.pneu., eubacterium *K. pneumoniae* (protein ID P27505); P.Aeru., *Pseudomonas aeruginosa* (protein ID NP_250677); M.magn., proteobacterium *Magnetospirillum magnetotacticum* (protein ID ZP_00052131), and D.hafn. the Gram-positive bacterium *Desulfitobacterium hafniense* (protein ID ZP_00101389). α-Helices and loops as observed in the structure of PqqC from *K. pneumoniae* are indicated with boxes above their sequences, where structural elements involved in the conformational change (α5b′, α5b, α6a, and α6′) are drawn in bold. Regions participating in the dimer interface are underlined. The helices formed upon ligand binding, as well as the active site residues coordinating the ligand, are highlighted in gray.

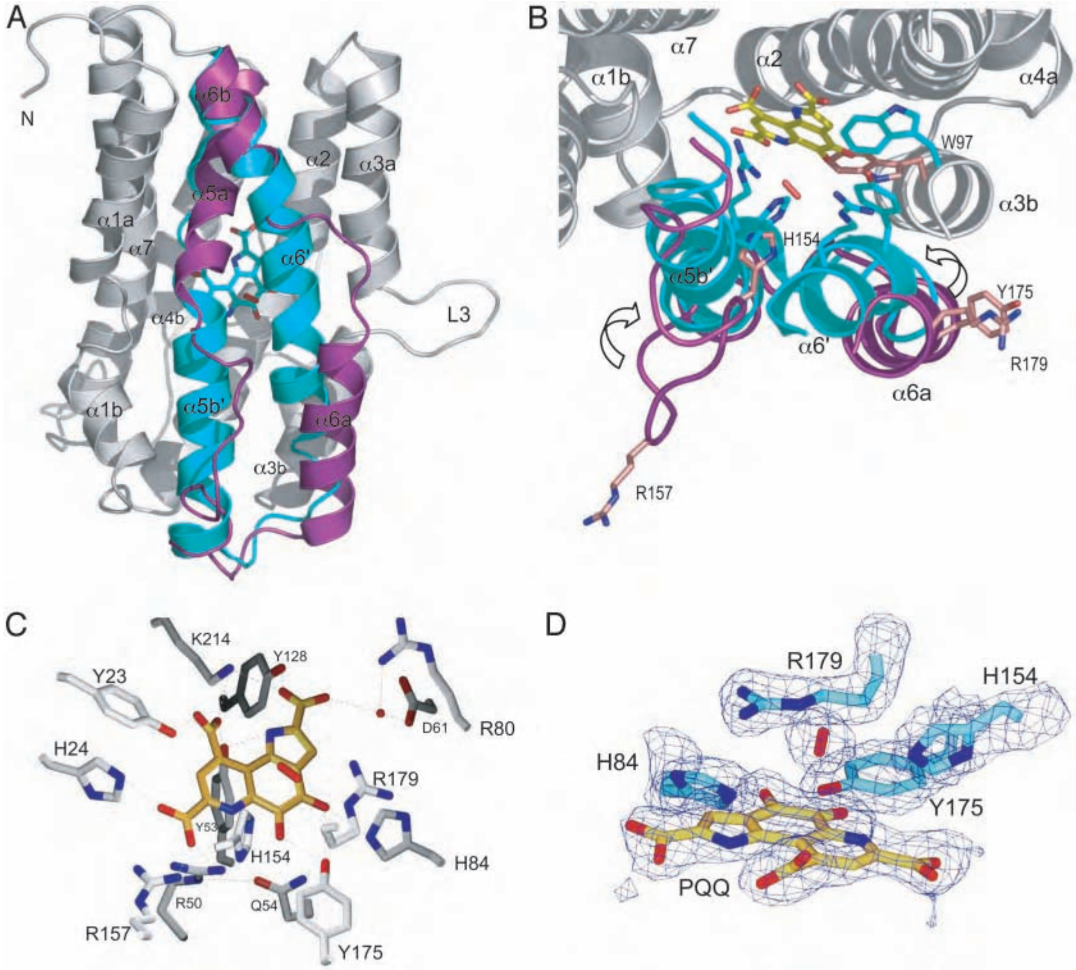


Fig. 4. Active site analysis. (*A*) The conformational change upon formation of the PqqC/PQQ complex (see *Results* for further details). For comparison, structures of the unligated and ligated PqqC are superimposed and shown in ribbon representation (gray). The moving parts (residues 142–193) are depicted as colored ribbons: unligated (magenta) and ligated (cyan). The PQQ (cyan) and the putative H_2O_2 (red) molecule are shown in sticks. (*B*) Same as *A* but tilted 90° around the horizontal axis, showing the active site residues H154, R157, Y175, R179, and W97 as sticks in the open (magenta) and closed (cyan) structure. The PQQ molecule is depicted in yellow sticks, and the putative H_2O_2 is shown as a red stick. (*C*) The PqqC active site in the closed state with PQQ (yellow) and the putative H_2O_2 molecule (red). All PqqC active site residues, within a distance of 3.3 Å to PQQ, are shown in gray sticks. (*D*) The oxygen binding site. The PQQ molecule and active site residues H84, H154, Y175, and R179, as well as the putative H_2O_2 molecule, are depicted in sticks. The 2Fo-Fc electron density map is contoured at 1.5 σ.

group in the pyrrole-ring N1, and the two quinone oxygens O4 and O5. The electron density maps at 2.3 Å resolution for the ligand show the tricyclic ring system and especially the C5—O5 bond in an approximately planar conformation. The three carboxylate groups of the ligand are twisted out of the plane of the ring system and form polar interactions with several active site residues. R157 interacts with the carboxyl group COO7′, which is also coordinated by two other positively charged residues, H24 and R50. Part of this network is a 12-Å-long and 3-Å-wide tunnel that bends from the COO7′ carboxyl group of PQQ, via R157 and five water molecules, out to the protein surface. Y175 coordinates the oxygen atoms of C5 and C4 of the ligand and creates another extended hydrogen bonding network with Q54 and R50. This site is connected to the outside via a 3 Å wide tunnel that goes 9 Å from PQQ-O5′ out to the protein surface near loop L3. The tube, blocked by water O31, is guided by Y175, R179, H84, Q182, and S178, all of which are in hydrogen bonding distance to each other. These multiple interactions lock the ligand into the active site and create a microenvironment that facilitates the complex enzymatic reaction.

An interesting feature of the PqqC/PQQ structure is the density located 3 Å above the C5 atom of PQQ (Fig. 4*D*). It is in close contact to R179 and H154 located at a distance of 2.8 Å and 3.3 Å, respectively. The elliptically shaped density is too strong to be modeled as a single water molecule, resulting in unrealistically low B-factors and strong positive density (3σ peak in 2FoFc map). Furthermore, it is too weak and does not have a sphere of coordinating residues as expected for a small ion, such as Na^+ or Cl^-, which are present in the crystallization buffer. Interestingly, the location of this density next to the reactive C5 of PQQ coincides with the substrate position in respect to PQQ in PQQ enzymes like glucose dehydrogenase and methanol dehydrogenase (32), strongly suggesting an important role in oxidation. Therefore and according to the PqqC reaction, which is oxygen dependent and produces hydrogen peroxide, the density most likely represents an O_2 or H_2O_2 molecule, the latter being more probable. Modeling O_2 or H_2O_2 into the density yields B-factors around 28 Å^2 consistent with the range of B-factors displayed by surrounding atoms. Although these observations raise the possibility of a trapped O_2 or H_2O_2 molecule, this interpretation clearly requires higher resolution data and additional experimental evidence. Currently there is no crystal

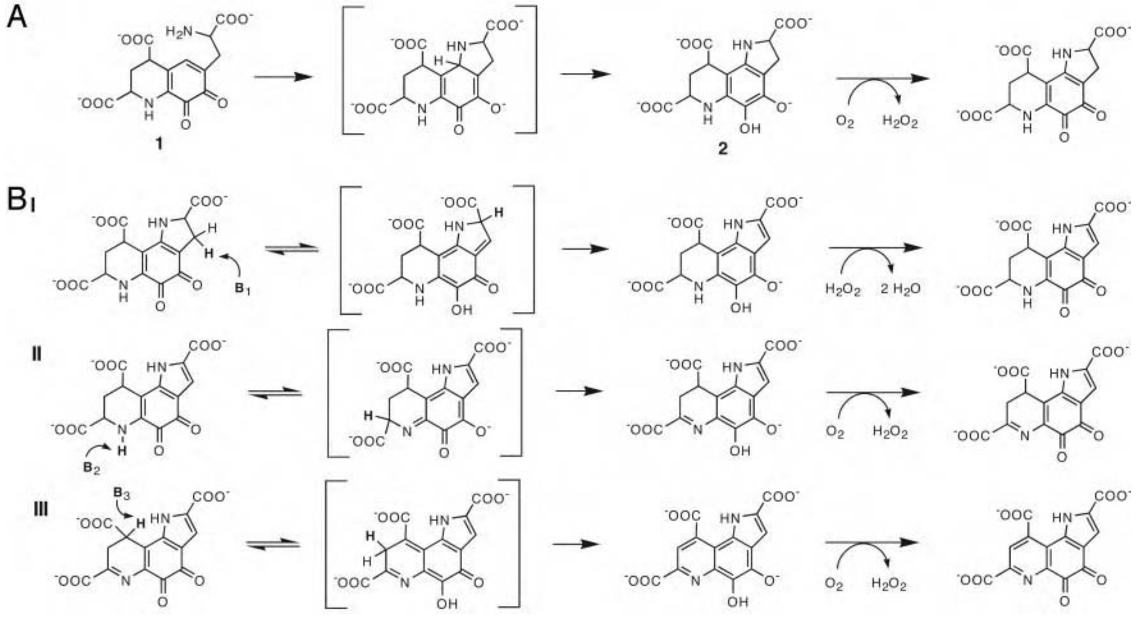


Fig. 5. The proposed reaction mechanism of PqqC. (*A*) Ring cyclization of 3a-(2-amino-2-carboxy-ethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid (**1**) to an obligatory intermediate (**2**) and its subsequent oxidation. (*B*) Base-catalyzed tautomerization and the six electron H_2O_2/O_2-mediated oxidation of **2**. Tautomerization is initiated by enzyme-catalyzed proton abstraction at C3, N6, and C9, respectively. Oxidation of quinol intermediates presumably occurs by means of two single-electron transfer steps (not shown). The order of double-bond formation (I-III) is arbitrary and cannot be determined at this point. See *Discussion* for further details, including the identity of proposed catalytically important residues.

structure of a protein with O_2 or H_2O_2 bound at a non-metal site, but recent kinetic studies on several variants of the quino-protein copper amine oxidase have provided evidence for specific binding of O_2 to a side-chain-generated pocket (33).

Discussion

PqqC facilitates cyclization and oxidation of 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid to PQQ in a reaction that occurs in the absence of any apparent cofactor. The product is bound and oriented by the fourteen conserved active site residues forming the positively charged active site cavity (Fig. 4*C*). By inference, substrate binding is concluded to induce a large conformational change, which closes the cavity and moves H154, R157, Y175, and R179 into the coordination sphere of the substrate to form an extended network of hydrogen bonds.

In our proposed mechanism, the first chemical step is the formation of the tricyclic ring system to yield **2** as shown in Fig. 5*A*. Cyclization occurs by a 1,4-addition of the primary amine to the conjugated ring structure, which may be facilitated by polar interactions of the substrate with several active site residues. Proton abstraction from N1 of the substrate and its subsequent activation for addition to C1a may be catalyzed by Y53, which is in close proximity to K214 (3.6 Å), making this a putative catalytic diad. Alternatively, the substrate may bind as the neutral amine, and the enzyme may simply provide a scaffold for the cyclization to occur, an event that should proceed rapidly once the amino group is deprotonated.

The next step involves the removal of eight electrons and eight protons from **2** to form PQQ, which means that PqqC needs to function as a dehydrogenase/desaturase. The PqqC/PQQ complex structure does not contain a redox active metal or cofactor, which raises the question of how oxidation is achieved. We propose a mechanism that leads to double bond formation by forming quinol intermediates, which are capable of the direct transfer of electrons to molecular oxygen and/or hydrogen peroxide (Fig. 5*B*) (see below). As shown, this mechanism involves sequential base-catalyzed tautomerizations to produce a reduced quinol from the oxidized quinone, with the introduction of three double bonds. The proposed mechanism predicts multiple proton abstractions occurring at different positions on the substrate. The structure of PqqC with PQQ identifies a number of amino acid residues important for catalysis. The product complex shows three histidine residues ideally placed for general base catalysis. These are H84, H24, and H154, which are all within 5 Å from C3, C9, and N6, respectively. Residues Y53/K214 could also facilitate proton abstraction because the oxygen of Y53 is within 4 Å of C9. Another important residue is Y175, which is within 3 Å of both O4 and O5 and is part of a hydrogen-bonding network that extends to the surface of the protein as discussed above. Y175 is therefore a likely candidate for the shuttling of protons to and from O4 and O5 during the postulated tautomerization and oxidation reactions. Unstable intermediates, which may collapse spontaneously due to the large driving force for aromatization of the quinol, are shown in brackets in the mechanism in Fig. 5*B*. Each time the quinol form is reached, it will transfer electrons to O_2/H_2O_2, possibly in successive steps to produce a superoxide-semiquinone intermediate (not shown). We cannot determine the order of the proposed tautomerization steps at this point, and they are depicted in an arbitrary fashion in Fig. 5*B*.

The biochemical data indicate that molecular oxygen is not the exclusive oxidant in the reaction. The data show that three equivalents of O_2 are consumed and two equivalents of H_2O_2 produced per mole of PQQ formed, respectively (Fig. 2). The stoichiometry can be explained if one equivalent of O_2 is reduced by four electrons to form two equivalents of water whereas the remaining two equivalents of O_2 undergo two-electron reduction to H_2O_2. Because the observed first-order rate constants for PQQ formation, oxygen consumption, and peroxide formation, respectively, are very similar, the first oxidation step seems to be the rate-limiting step in the overall reaction. Also, the release of peroxide after the first oxidation step is proposed to be slow compared with the rate of tautomerization, which would explain its complete reduction to water. The apparent slow release of peroxide at the end of the catalytic cycle, as implied by the different stoichiometries under continuous turnover vs. acid quenching (Fig. 2, curves C and D), may in part explain why the enzyme undergoes only a single turnover under *in vitro* conditions. Another contributing factor may be a slow release of bound PQQ, and experiments will be necessary to address these issues.

The crystal structure of the enzyme–product complex shows additional electron density inside the sealed cavity, next to the reactive C5 of PQQ. This location strongly points to a site of interaction between dioxygen and substrate. Because PQQ is present in the complex, the density may represent a H_2O_2 molecule. Although H_2O_2 was not included in the crystallization buffer, it could have been formed by means of a reaction with $PQQH_2$. Although the crystals were soaked with PQQ, reduction of the cofactor to $PQQH_2$ may have resulted from the presence of DTT in the buffer and/or from the electron beam of the synchrotron radiation. Thiols have been reported to be very efficient in reducing PQQ and derivatives thereof (34). Further support for H_2O_2 bound in the crystal comes from the biochemical studies discussed above, in which peroxide seems tightly bound to the enzyme in solution. We conclude that the structure most likely represents a ternary product complex, providing valuable information regarding the oxidative reaction. First, because the molecule is >3.0 Å away from both C4 and C5 of PQQ, the possibility of a covalent hydroperoxy intermediate is considered unlikely. Second, the close proximity of two basic residues, R179 (2.8 Å) and H154 (3.3 Å), would provide an ideal electrostatic environment to stabilize a superoxide anion produced upon electron transfer from the substrate. The formation of superoxide is often the rate-limiting step in these types of reactions, and electrostatic facilitation of superoxide formation seems to be emerging as a common theme in enzymes that reduce dioxygen (35). For example, the flavoprotein glucose oxidase uses a protonated histidine residue to reduce the activation barrier by $\approx 10^3$-fold for electron transfer from reduced flavin to molecular oxygen (36). Finally, the positive electrostatic potential in the active site is expected to lower the pK_a of the C4-C5 quinol, increasing the oxidation potential toward O_2. The proposed structure-based mechanism introduces an "information-rich" framework for future investigations using site-directed mutagenesis and kinetic and spectroscopic analyses.

The reaction catalyzed by PqqC associates the enzyme with a rare family of cofactor-free oxidases and oxygenases (37). These enzymes include urate oxidase (38), 1*H*-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (39), and ActVA-Orf6 monooxygenase (40). The common feature among these enzymes is their action on substrates that are well suited for direct reaction with dioxygen, i.e., in the absence of an organic cofactor or metal ion. It is known that reduced $PQQH_2$ in aqueous solution is susceptible toward oxygen and produces H_2O_2 very rapidly (41).

The present x-ray crystal structures, in combination with biochemical data, provide a structural study of an enzyme involved in the PQQ-biosynthesis pathway. Our study identifies the final step in PQQ biosynthesis as a multistep reaction that includes cyclization and an overall eight-electron oxidation of 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid to PQQ. The structure of PqqC with PQQ identifies a number of important interactions between enzyme and product. These include K214, Y53, H24, H84, H154, and Y175 as possible participants in the proposed general base catalysis that leads to ring cyclization and tautomerization. Finally, a plausible dioxygen binding site is located next to H154, R179, and C5 of PQQ. Our data suggest a mechanism of oxidation whereby electrons are transferred directly to molecular oxygen and provide a structural foundation for further study of the mechanism of dioxygen activation without the assistance of a redox-active metal or cofactor.

We thank Constantina Bakolitsa and the staff at the Stanford Synchrotron Radiation Laboratory for help with data collection, and Jose Maria de Pereda and Charles Perrin for helpful suggestions. H.T. wants to thank Kimitoshi Sakamoto, Kazunobu Matsushita, and Osao Adachi for advice and support. R.S. was supported by Austrian Science Fund Fellowship J2209-B04, J.P.K. by National Institutes of Health Grant GM39296, and O.T.M. by a postdoctoral fellowship from the Miller Institute for Basic Research in Science, University of California, Berkeley.

1. Anthony, C. (2001) *Antioxid. Redox Signal.* **3,** 757–774.
2. Goodwin, P. M. & Anthony, C. (1998) *Adv. Microb. Physiol.* **40,** 1–80.
3. Duine, J. A. (2001) *Chem. Rec.* **1,** 74–83.
4. He, K., Nukada, H., Urakami, T. & Murphy, M. P. (2003) *Biochem. Pharmacol.* **65,** 67–74.
5. Mitchell, A. E., Jones, A. D., Mercer, R. S. & Rucker, R. B. (1999) *Anal. Biochem.* **269,** 317–325.
6. Steinberg, F., Stites, T. E., Anderson, P., Storms, D., Chan, I., Eghbali, S. & Rucker, R. (2003) *Exp. Biol. Med. (Maywood)* **228,** 160–166.
7. Kasahara, T. & Kato, T. (2003) *Nature* **422,** 832.
8. Kim, M., Okajima, T., Kishishita, S., Yoshimura, M., Kawamori, A., Tanizawa, K. & Yamaguchi, H. (2002) *Nat. Struct. Biol.* **9,** 591–596.
9. Dove, J. E. & Klinman, J. P. (2001) *Adv. Protein Chem.* **58,** 141–174.
10. Houck, D. R., Hanners, J. L., Unkefer, C. J., van Kleef, M. A. & Duine, J. A. (1989) *Antonie Van Leeuwenhoek* **56,** 93–101.
11. Unkefer, C. J., Houck, D. R., Britt, B. M., Sosnick, T. R. & Hanners, J. L. (1995) *Methods Enzymol.* **258,** 227–235.
12. Meulenberg, J. J., Sellink, E., Riegman, N. H. & Postma, P. W. (1992) *Mol. Gen. Genet.* **232,** 284–294.
13. van Kleef, M. A. & Duine, J. A. (1988) *Biofactors* **1,** 297–302.
14. Goosen, N., Huinen, R. G. & van de Putte, P. (1992) *J. Bacteriol.* **174,** 1426–1427.
15. Velterop, J. S., Sellink, E., Meulenberg, J. J., David, S., Bulder, I. & Postma, P. W. (1995) *J. Bacteriol.* **177,** 5088–5098.
16. Toyama, H., Fukumoto, H., Saeki, M., Matsushita, K., Adachi, O. & Lidstrom, M. E. (2002) *Biochem. Biophys. Res. Commun.* **299,** 268–272.
17. Toyama, H., Chistoserdova, L. & Lidstrom, M. E. (1997) *Microbiology* **143,** 595–602.
18. Magnusson, O. Th., Toyama, H., Saeki, M., Schwarzenbacher, R. & Klinman, J. P. (2004) *J. Am. Chem. Soc.* **126,** 5342–5343.
19. Schwarzenbacher, R., Stenner-Liewen, F., Stenner, H., Reed, J. C. & Liddington, R. C. (2004) *Proteins*, in press.
20. Duine, J. A., Frank, J. & Jongejan, J. A. (1987) *Adv. Enzymol.* **59,** 169–212.
21. Zhou, M., Diwu, Z., Panchuk-Voloshina, N. & Haugland, R. P. (1997) *Anal. Biochem.* **253,** 162–168.
22. Schwartz, B., Dove, J. E. & Klinman, J. P. (2000) *Biochemistry* **39,** 3699–3707.
23. Notredame, C., Higgins, D. G. & Heringa, J. (2000) *J. Mol. Biol.* **302,** 205–217.
24. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
25. Li, W., Jaroszewski, L. & Godzik, A. (2001) *Bioinformatics* **17,** 282–283.
26. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
27. CCP4. (1994) *Acta Crystallogr. D* **50,** 760–763.
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991) *Acta Crystallogr. A* **47,** 110–119.
29. Liang, J., Edelsbrunner, H. & Woodward, C. (1998) *Protein Sci.* **7,** 1884–1897.
30. Holm, L. & Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138.
31. Schuller, D. J., Wilks, A., Ortiz de Montellano, P. R. & Poulos, T. L. (1999) *Nat. Struct. Biol.* **6,** 860–867.
32. Oubrie, A., Rozeboom, H. J., Kalk, K. H., Olsthoorn, A. J., Duine, J. A. & Dijkstra, B. W. (1999) *EMBO J.* **18,** 5187–5194.
33. Goto, Y. & Klinman, J. P. (2002) *Biochemistry* **41,** 13637–13643.
34. Itoh, S., Kato, N., Ohshiro, Y. & Agawa, T. (1985) *Chem. Lett.* 135–136.
35. Klinman, J. P. (2001) *J. Biol. Inorg. Chem.* **6,** 1–13.
36. Roth, J. P. & Klinman, J. P. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 62–67.
37. Fetzner, S. (2002) *Appl. Microbiol. Biotechnol.* **60,** 243–257.
38. Kahn, K. & Tipton, P. A. (1998) *Biochemistry* **37,** 11651–11659.
39. Fischer, F. & Fetzner, S. (2000) *FEMS Microbiol. Lett.* **190,** 21–27.
40. Sciara, G., Kendrew, S. G., Miele, A. E., Marsh, N. G., Federici, L., Malatesta, F., Schimperna, G., Savino, C. & Vallone, B. (2003) *EMBO J.* **22,** 205–215.
41. Itoh, S., Ohshiro, Y. & Agawa, T. (1986) *Bull. Chem. Soc. Jpn.* **59,** 1911–1914.

PROTEINS: Structure, Function, and Bioinformatics 56:401–403 (2004)

STRUCTURE NOTE

Crystal Structure of PqqC From *Klebsiella pneumoniae* at 2.1 Å Resolution

Robert Schwarzenbacher,* Frank Stenner-Liewen, Heike Liewen, John C. Reed, and Robert C. Liddington
The Burnham Institute, La Jolla, California

Introduction. The biosynthesis of pyrroloquinoline quinone (PQQ), a novel vitamin[1] and redox cofactor of several quinoprotein dehydrogenases, is facilitated by an unknown pathway that requires the expression of six genes, *pqqA-F.*[2] The final step is catalyzed by PqqC, which uses molecular oxygen to oxidize a quinone intermediate to PQQ.[3] Here we report the purification and crystallization of PqqC and the determination of its X-ray crystal structure at 2.1 Å resolution.

Material and Methods. *Expression and purification of PqqC.* The pqqC open reading frame from Klebsiella pneumoniae NCBI accession no. X58778 (residues Met1 to VAL251) was amplified by PCR and cloned into expression plasmid pet21d (Invitrogen, Carlsbad). The protein was expressed in E.coli strain BL21(DE3). Cells (1 L) were induced with 0.5 mM IPTG (3 h at 37°C) harvested and frozen at −80°C. PqqC was purified by a combination of affinity and gel filtration chromatography. Briefly, after harvesting and lysis of the cells, the crude extract was centrifuged and applied to a 5 ml Ni-affinity column (Pharmacia). PqqC containing fractions were dialyzed against 20 mM Tris/HCl, at pH 8.0, 1 mM DTT, further purified on a Superdex 200 gelfiltration column (Pharmacia, Uppsala, Sweden), and concentrated to a final concentration of ~8 mg/mL in an ultrafiltration cell (Amicon). The seleno-methionine substituted protein was expressed as described[4] and purified accordingly to the wildtype with the exception that there was 5 mM DDT added to the dialysis and gelfiltration buffers.

Crystallization. Crystals were grown at 25°C by the sitting drop vapor diffusion method in droplets composed of one part protein solution (8 mg/ml in 20 mM Tris/HCl, at pH 8.0 and 1mM DTT) and one part reservoir solution (1.2 M ammonium sulfate, MES/NaOH at pH 6.0). The orthorhombic crystals are of space group $P2_12_12$ and contain two molecules in the asymmetric unit. For data collection, these crystals were transferred into cryobuffer (crystallization buffer with 25% (v/v) glycerol) and flash-cooled in liquid nitrogen.

Data collection, structure solution, and refinement. The three-wavelength Selenium MAD data set was collected from one single crystal, using synchrotron radiation at beamline 9.2 of SSRL. Oscillation data were recorded in frames of 1° through a continuous angular range of 120° for the peak (Se-λ_1), the high energy remote (Se-λ_2), and the inflection point (Se-λ_3). Native datasets were collected at beamline 11.1 of SSRL. All data were processed with the programs DENZO and SCALEPACK.[5] The PqqC structure was solved, phased and traced using programs SOLVE and RESOLVE.[6] Crystallographic refinement was carried out using TLS refinement in REFMAC5[7] and program O.[8] The PqqC model comprises two protein monomers (residues A1–A152, A160–A251, B1–B152, B162–B251) and 130 water molecules. No electron density was observed for residues A153–A159 and B153–B161.

Crystallographic statistics are shown in Table I. The structure exhibits excellent stereochemistry with 94.1% and 5.9% of the residues in the most favored and additional allowed regions, respectively, of the Ramachandran plot.[9] Surface calculations were carried out with the CASTP-server[10] and the protein-protein interaction server.[11] The figure was drawn with PYMOL (DeLano Scientific LLC).

Coordinates. Coordinates and structure factors for PqqC have been deposited with the Protein Data Bank (http://www.rcsb.org/pdb) under accession code 1OTV.

Results and Discussion. The PqqC protein from *Klebsiella pneumoniae* has a molecular weight of 28,914 Da (251 residues) and a calculated isoelectric point of 5.6. PqqC was cloned, over-expressed in *E. coli,* purified and crystallized. PqqC crystallizes in the orthorhombic space group $P2_12_12$ with unit cell dimensions a = 67.15 Å, b = 117.03 Å, c = 74.96 Å, and two molecules in the asymmetric unit. The structure of the enzyme was determined by a MAD experiment using a seleno-methionine substituted protein.[4] The PqqC model comprises two protein monomers (residues A1–A152, A160–A251, B1–B152, B162–B251) and 130 water molecules. Statistics for data collection, refinement and model quality are summarized in Table I.

The PqqC molecule is roughly cylindrical, with approximate dimensions of 45 Å × 29 Å × 37 Å [Fig. 1(A)]. PqqC

---

*Correspondence to: Robert Schwarzenbacher, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037. E-mail: roberts@burnham.org.

Received 9 December 2003; Accepted 15 December 2003

TABLE I. Summary of Crystal Parameters, Data Collection, and Refinement Statistics for PqqC (PDB: 1OTV)[†]

Space group	$P2_12_12$			
Unit cell parameters	a = 67.15 Å, b = 117.03 Å, c = 74.96 Å, $\alpha = \beta = \gamma = 90°$			
Data collection	λ_0	Se-λ_1	Se-λ_3	Se-λ_3
Wavelength (Å)	0.954	0.9794	0.9793	0.9150
Resolution range (Å)	69.0–2.10	50.0–2.50	50.0–2.50	50.0–2.50
Highest resolution shell (Å)	2.15–2.10	2.73–2.50	2.73–2.50	2.73–2.50
Number of observations	133,009	31,540	30,978	32,411
Number of reflections	35,743	9,459	9,579	9,482
Completeness (%)	99.2 (96.2)	93.1 (88.7)	93.9 (87.4)	93.2 (85.2)
Mean $I/\sigma(I)$	10.1 (2.3)	11.7 (2.4)	11.4 (2.5)	9.3 (1.7)
R_{meas} on I	14.5 (41.2)	8.4 (45.2)	10.6 (54.1)	10.3 (61.6)
Sigma cutoff	0.0	0.0	0.0	0.0
Model and refinement statistics				
Resolution range (Å)	69.0–2.10	Data set used in refinement		λ_0
Number of reflections (total)	35,734	Cutoff criteria		$\|F\| > 0$
Number of reflections (test)	1,787	R_{cryst}		0.191
Completeness (% total)	99.2	R_{free}		0.230
Stereochemical parameters				
Restraints (RMS observed)				
Bond length		0.018 Å		
Bond angle		1.58°		
Average isotropic B-value		32.9 Å^2		
ESU based on free R value		0.17 Å		
Protein residues/atoms		506/4,156		
Solvent molecules		130		

[†] highest resolution shell.

ESU = Estimated overall coordinate error.[7, 12]

$R_{meas} = [\Sigma_h w \Sigma_i |\langle I_h\rangle - I_{h,i}|]/\Sigma_h \Sigma_i I_{h,i}$ where $w = [n_h/(n_h-1)]^{1/2}$ and $\langle I_h\rangle = [\Sigma_i^n I_{h,i}]/n_h$. This is the multiplicity-weighted R_{sym}.[13]

$R_{cryst} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{free} = as for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from refinement.

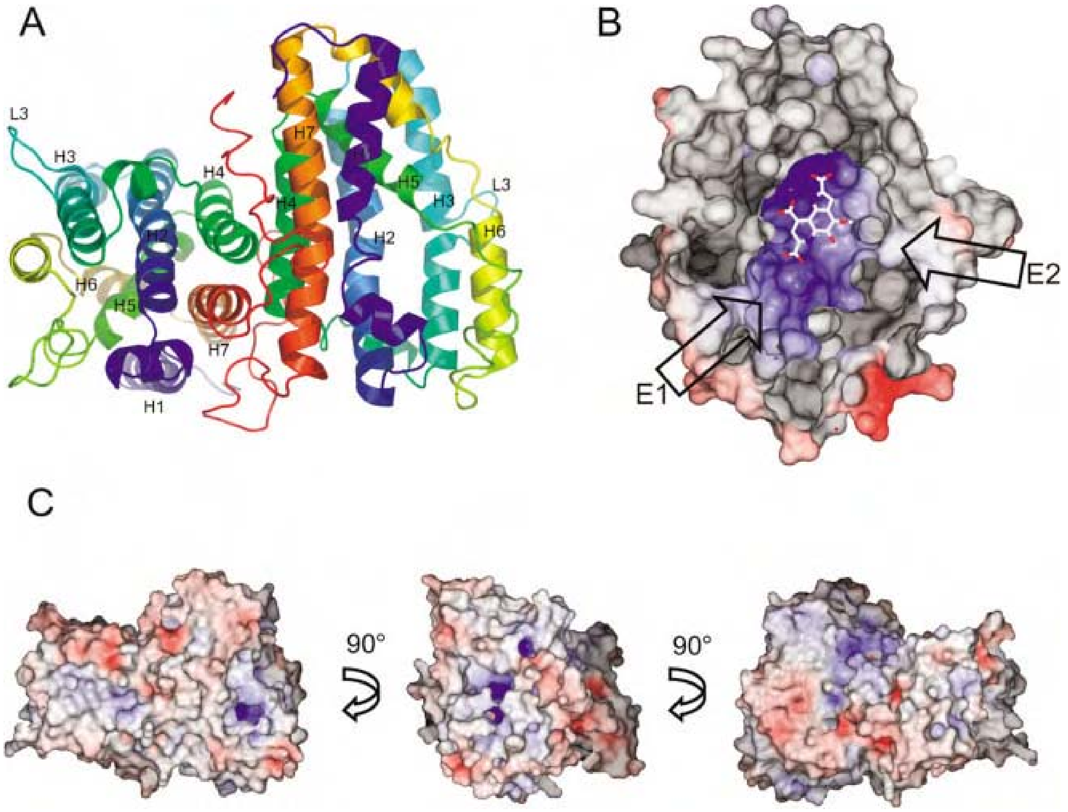


Fig. 1. Crystal structure of PqqC. **A**: The PqqC dimer depicted in ribbon representation rainbow color coded by residue numbering from N-terminus (blue) to C-terminus (red), with Helices H1-H7 and loop L3 labeled. **B**: The active site cavity of PqqC. A cut through the PqqC monomer is shown in surface representation color coded according to surface potential (blue positive; red negative). For size comparison a manually docked PQQ molecule is depicted in sticks. The active site carries a significant positive charge. The two channels connecting the active site to the solvent are marked with arrows E1 and E2. **C**: The PqqC dimer shown in surface representation color coded according to surface potential. A series of views with consecutive 90° rotations are displayed.

folds into a compact seven helix bundle that resembles the heme-oxygenase fold,[14] with six circular aligned helices (H1, H2, H3, H4, H5, H7) partly embracing a seventh hydrophobic helix (H6). Helices H1–H7 form a long up and down helix bundle, where helices H7, H1, H5, H2 and H4 pack parallel and helices H5, H6, H3, H2, H4 and H7 pack anti-parallel to each other. Helices H1, H3, H4, H5, H6 are kinked and can therefore be represented as separate shorter helices denoted as A and B. This is especially true for helix H3, where the peculiar hairpin loop L3, residues 84 to 92, is inserted [Fig. 1(A)].

PqqC forms a homo-dimer in solution, as evident from gel filtration experiments (data not shown). The dimer is formed through an interaction via helices H1, H4b, H7 and the C-terminus with a buried surface area of 2145 Å^2 per monomer [Fig. 1(A,C)]. The nature of the interaction is mainly hydrophobic and facilitated by 63% non-polar atoms. The polar interface residues H26, R206, E209, D216, Y236, T238, H246, T247, T248, L250, and R233 contribute 19 hydrogen bonds.[10]

Analysis of the PqqC structure shows that the seven α-helices provide the scaffold for a unique active site cavity [Fig. 1(B)]. The cavity is lined with 42 mostly hydrophilic and aromatic residues that are highly conserved within PqqC proteins from different bacteria. Interestingly, these residues are contributed from all seven helices H1 to H7. The cavity shows a distinct overall positive charge, measures 9 Å × 13 Å × 23 Å and embraces a molecular surface volume of 2200 Å^3.[11] Two openings connect it to the outside. The larger opening (E1) between helices H1 and H6 is divided by the disordered region between helices H5a and H5b and extends over a molecular surface area of 200 Å^2. The mouth region is lined by 14 residues, including the histidine ladder H24, H30 and H34. The second opening (E2) is much smaller and has a molecular surface of 30 Å^2. It penetrates the surface of the protein between helices H6a, H6b and the unique loop L3. The opening is guided by residues R179, D83, Q182 and N184. Given the size of the two openings, the transfer of substrate and product is only possible via the large entrance. To our surprise the PqqC structure does not appear to contain a metal-site or a redox-cofactor binding site as would be expected for an oxidase, raising the question of how the final oxidative step of PQQ-biosynthesis is achieved.

A structural similarity search performed with the atomic coordinates of PqqC using the DALI server[15] yields human heme-oxygenase,[14] with an RMSD of 3.1 Å for the superposition of 194 Cα atoms and 9% sequence identity, the R2 subunit of ribonucleotide reductase,[16] with an RMSD of 3.7Å for the superposition of 172 Cα atoms and 15% sequence identity, and methane monooxygenase,[17] with an RMSD of 3.9 Å for the superposition of 170 residues and 12% sequence identity. Intriguingly, despite the fact that none of the active site residues are conserved, each of them appears to be a redox enzyme, suggesting that this fold may be particularly suitable for this type of enzyme.

Acknowledgment. RS was supported by FWF-fellowship J2209-B04. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a National user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences.

REFERENCES

1. Kasahara T, Kato T. Nutritional biochemistry: a new redox-cofactor vitamin for mammals. Nature 2003;422:832.
2. Goodwin PM, Anthony C. The biochemistry, physiology and genetics of PQQ and PQQ-containing enzymes. Adv Microb Physiol 1998;40:1–80.
3. Toyama H, Fukumoto H, Saeki M, Matsushita K, Adachi O, Lidstrom ME. PqqC/D, which converts a biosynthetic intermediate to pyrroloquinoline quinone. Biochem Biophys Res Commun 2002;299:268–272.
4. Harrison CJ, Bohm AA, Nelson HC. Crystal structure of the DNA binding domain of the heat shock transcription factor. Science 1994;263:224–227.
5. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
6. Terwilliger TC. Maximum-likelihood density modification using pattern recognition of structural motifs. Acta Crystallogr D Biol Crystallogr 2001; 57:1755–1762.
7. CCP4. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
8. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
9. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;283–291.
10. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci 1996;93:13–20.
11. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 1998;7:1884–1897.
12. Tickle IJ, Laskowski RA, Moss DS. Error estimates of protein structure coordinates and deviations from standard geometry by full-matrix refinement of gammaB- and betaB2-crystallin. Acta Crystallogr D Biol Crystallogr 1998;54:243–52.
13. Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol 1997;4:269–275.
14. Schuller DJ, Wilks A, Ortiz de Montellano PR, Poulos TL. Crystal structure of human heme oxygenase-1. Nat Struct Biol 1999;6:860–867.
15. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993;233:123–138.
16. Eriksson M, Jordan A, Eklund H. Structure of Salmonella typhimurium nrdF ribonucleotide reductase in its oxidized and reduced forms. Biochemistry. 1998;37:13359–13369.
17. Rosenzweig AC, Brandstetter H, Whittington DA, Nordlund P, Lippard SJ, Frederick CA. Crystal structures of the methane monooxygenase hydroxylase from Methylococcus capsulatus (Bath): implications for substrate gating and component interactions. Proteins 1997;29:141–152.

Efficient synthetic inhibitors of anthrax lethal factor

Martino Forino*†, Sherida Johnson*†, Thiang Y. Wong*, Dmitri V. Rozanov*, Alexei Y. Savinov*, Wei Li*, Roberto Fattorusso*, Barbara Becattini*, Andrew J. Orry‡, Dawoon Jung*, Ruben A. Abagyan‡, Jeffrey W. Smith*, Ken Alibek§¶, Robert C. Liddington*, Alex Y. Strongin*, and Maurizio Pellecchia*‖

*Burnham Institute, Cancer Research Center and Infectious and Inflammatory Disease Center, 10901 North Torrey Pines Road, La Jolla, CA 92037; ‡The Scripps Research Institute, Molecular Biology, 10550 North Torrey Pines Road, La Jolla, CA 92037; §National Center for Biodefense, George Mason University, 10900 University Boulevard, PWII Building, Room 160, MSN 1A8, Manassas, VA 20110; and ¶Advanced Biosystems, 5904 Richmond Highway, Suite 300, Alexandria, VA 22303

Edited by Peter K. Vogt, The Scripps Research Institute, La Jolla, CA, and approved May 23, 2005 (received for review April 3, 2005)

Inhalation anthrax is a deadly disease for which there is currently no effective treatment. *Bacillus anthracis* lethal factor (LF) metalloproteinase is an integral component of the tripartite anthrax lethal toxin that is essential for the onset and progression of anthrax. We report here on a fragment-based approach that allowed us to develop inhibitors of LF. The small-molecule inhibitors we have designed, synthesized, and tested are highly potent and selective against LF in both *in vitro* tests and cell-based assays. These inhibitors do not affect the prototype human metalloproteinases that are structurally similar to LF. Initial *in vivo* evaluation of postexposure efficacy of our inhibitors combined with antibiotic ciprofloxican against *B. anthracis* resulted in significant protection. Our data strongly indicate that the scaffold of inhibitors we have identified is the foundation for the development of novel, safe, and effective emergency therapy of postexposure inhalation anthrax.

NMR | protective antigen | fragment-based design | metalloprotease | drug design

The U.S. government has declared that an effective postexposure treatment of anthrax is a key national priority in the fight against bioterrorism. *Bacillus anthracis* (1) is the causative bacterium of anthrax, and its clinical presentation and outcome strongly depend on its entry route in humans. Cutaneous anthrax is rarely lethal. In contrast, inhalation anthrax, a potential weapon of bioterror, is far more dangerous and usually fatal if it is not diagnosed and treated early (2). After anthrax spores are inhaled, they adhere to alveolar macrophages and then germinate. Bacteria migrate to lymph nodes, where they rapidly multiply (3) and excrete a tripartite exotoxin comprised of protective antigen (PA, 83 kDa), lethal factor (LF) Zn^{2+}- metalloproteinase (90 kDa), and calmodulin-activated edema factor adenylate cyclase (EF, 89 kDa). Current knowledge suggests that the concerted activity of PA, LF, and EF kills host macrophages and largely eliminates the host immune system, thereby promoting continual progression of the disease. Unless properly and promptly treated, inhalation anthrax will lead to the death of the host organism (4). To exert its lethal effect, anthrax lethal toxin must enter inside the cell compartment. PA binds to the ubiquitously expressed cellular receptors (5) and, after its proteolytic activation by the furin-like proprotein convertases and the release of the N-terminal 20-kDa fragment, generates the mature PA protein (PA63). PA63 heptamerizes and binds both LF and EF. After endocytosis of the resulting complexes, the engulfed molecules of LF and EF are liberated and exert their toxic action (6). Inside the cell compartment, LF cleaves mitogen-activated protein kinase kinases (MAPKK) (7–9), disrupts signal transduction, and finally leads to macrophage lysis through a mechanism that is not completely understood to date (10). Accordingly, inhibition of LF is the most promising means for treating postexposure anthrax (11, 12).

We describe in this report a fragment-based drug design approach that led us to the discovery of several small-molecule synthetic inhibitors, which have shown a strong and highly specific inhibition of LF protease activity. By using simple enzymatic assays that take advantage of highly sensitive heteronuclear NMR techniques, we have readily identified a preferred inhibitor scaffold for LF. Cell-based and peptide cleavage assays were subsequently used to confirm the potency of the iterated leads. Initial structural analyses of the LF–inhibitor complexes at the atomic resolution level provide insights on the rationale of the potency of the designed inhibitors. The inhibitory potency of the refined leads was validated in *in vitro* as well as cell-based assays. Preliminary *in vivo* studies on the efficacy of our inhibitors combined with antibiotic ciprofloxican against *B. anthracis* (Sterne strain) are also discussed.

Materials and Methods

Reference Compounds and Reagents. All common chemicals, reagents, and buffers were purchased from Sigma–Aldrich, Chembridge (San Diego), or Maybridge (Cornwall, U.K.). Recombinant LF and MAPKKide were both purchased from List Biological Laboratories (Campbell, CA). Fluorinated peptide substrate was from Anaspec (San Jose, CA).

Fluorescence Peptide Cleavage Assay. Cleavage reactions (100 μl each) were performed in a 96-well plate. Each reaction contained MAPKKide (4 μM) and LF (50 nM) in 20 mM Hepes, pH 7.4, and the small-molecule inhibitor. Kinetics of the peptide cleavage was examined for 30 min by using a fluorescent plate reader at excitation and emission wavelength at 485 and 590 nm, respectively.

The K_m and V_{max} values of the MAPKKide cleavage by LF were determined at 25°C by using the same experimental condition described above for the fluorescence screening assay but with increasing MAPKKide concentrations (2, 3, 5, 8, and 10 μM). The K_i and $K_{m(app)}$ were calculated at a fixed 10 μM inhibitor concentration. All constant values were definitely evaluated by fitting the data to the Lineweaver–Burk plot.

NMR Measurements. ^{19}F NMR 1D spectra were acquired on a Bruker (Billerica, MA) Avance 500-MHz spectrometer equipped with a selective ^{19}F/^{1}H probe. Each spectrum was recorded at 25°C in buffers with a 9:1 H_2O:D_2O ratio. All spectra were collected with a sweep width of 5 ppm and an acquisition time of 20 min. The LF assay was performed with 50 nM recombinant LF (List Biological Laboratories) and 20 μM peptide substrate Ac-A-R-R-K-K-V-Y-P-NH-Ph-CF_3 (Anaspec); inhibition activity was detected in the same conditions. Reaction was quenched after 30 min by using 100 μM GM6001 (List Biological Laboratories) at 0°C or **BI-MFM3**.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: LF, lethal factor; PA, protective antigen; MAPKK, mitogen-activated protein kinase kinase; MMP, metalloproteinase.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1ZXV).

[†] M.F. and S.J. contributed equally to this work.

[‖] To whom correspondence should be addressed. E-mail: mpellecchia@burnham.org.

www.pnas.org/cgi/doi/10.1073/pnas.0502733102

PNAS | July 5, 2005 | vol. 102 | no. 27 | 9499–9504

Synthetic Chemistry. ***General procedure for the synthesis of rhodamine derivatives (Table 2).*** Rhodanine acetic acid (0.100 g, 0.523 mmol) was added to a solution of the furfuraldehyde (0.575 mmol) in dimethylformamide (1 ml), and the mixture was stirred until it became homogenous. The mixture was then placed in the microwave (Milestone, Monroe, CT), where it underwent four cycles of 1-min heating (140°C, 1,000 W) and 3 min of cooling (25°C). Water was then added to the solution, where precipitate was formed. The precipitate was collected via filtration, recrystallized from acetone/water, and dried to yield the desired compound. Characterization of each compound was obtained by means of NMR spectroscopy and mass spectrometry, as reported below.

{5-[5-(4-Chloro-phenyl)-furan-2-ylmethylene]-4-oxo-2-thioxo-thiazolidin-3-yl}-acetic acid (BI-11A9). **BI-11A9** was obtained as 0.176 g of reddish orange solid in 88% yield. ^{1}H NMR (300 MHz, *d*-DMSO) δ 7.86 (d, 2H, J = 8.0 Hz), 7.75 (s, 1H), 7.64 (d, 2H, J = 8.0 Hz), 7.41 (s, 2H), 4.74 (s, 2H).

{5-[5-(4-Bromo-phenyl)-furan-2-ylmethylene]-4-oxo-2-thioxo-thiazolidin-3-yl}-acetic acid (BI-11A10). **BI-11A10** was obtained as 0.198 g of reddish orange solid in 89% yield. ^{1}H NMR (300 MHz, *d*-DMSO) δ 7.80 (s, 4H), 7.76 (s, 1H), 7.42 (s, 2H), 4.75 (s, 2H).

{5-[5-(4-Chloro-2-nitro-phenyl)-furan-2-ylmethylene]-4-oxo-2-thioxo-thiazolidin-3-yl}-acetic acid (BI-11A11). **BI-11A11** was obtained as 0.118 g of yellow solid in 53% yield). ^{1}H NMR (300 MHz, *d*-DMSO) δ 8.27 (d, 1H, J = 2.1 Hz), 8.00 (dd, 2H, J = 8.4, 2.1 Hz),7.96 (dd, 1H, J = 8.4, 2.1), 7.76 (s, 1H-), 7.44 (d, 1H, J = 3.9 Hz), 7.34 (d, 1H, J = 3.9 Hz), 4.73 (s, 2H).

{5-[5-(2-nitro-phenyl)-furan-2-ylmethylene]-4-oxo-2-thioxo-thiazolidin-3-yl}-acetic acid (BI-11A1). **BI-11A12** was obtained as 0.089 g of light orange solid in 44% yield. ^{1}H NMR (300 MHz, *d*-DMSO) δ 8.07 (d, 1H, J = 8.0 Hz), 7.99 (d, 1H, J = 8.0 Hz), 7.88 (t, 1H, J = 8.0, 7.5), 7.78 (s, 1H), 7.74 (t, 1H, J = 8.0, 7.5 Hz), 7.46 (d, 1H, J = 3.9 Hz), 7.32 (d, 1H, J = 3.9 Hz), 4.74 (s, 2H).

{5-[5-(3-Chloro-4-methoxy-phenyl)-furan-2-ylmethylene]-4-oxo-2-thioxo-thiazolidin-3-yl}-acetic acid (BI-11B1). **BI-11B1** was obtained as 0.178 g of bright reddish orange solid in 83% yield. ^{1}H NMR (300 MHz, *d*-DMSO/pyridine) δ 7.97 (d, 1H, J = 2.1 Hz), 7.83 (dd, 1H, J = 8.7, 2.1 Hz), 7.77 (s, 1H), 7.43 (d, 1H, J = 4.0 Hz), 7.41 (s, 1H), 7.32 (d, 1H, J = 4.0 Hz), 4.76 (s, 2H), 3.960 (s, 3H); ^{13}C NMR (75 MHz, *d*-DMSO) 193, 170, 166, 157, 156, 155, 153, 149, 146, 124, 122, 121, 121, 118, 56, 44; MALDI-MS *m/z* 431.8886 (M + Na, C17H12ClNO5S2).

{5-[5-(3,4-Dichloro-4-phenyl)-furan-2-ylmethylene]-4-oxo-2-thioxo-thiazolidin-3-yl}-acetic acid (BI-11B2). **BI-11B2** was obtained as 0.065 g of light orange precipitate in 30% yield. ^{1}H NMR (300 MHz, *d*-DMSO) δ 8.14 (d, 1H, J = 1.8 Hz), 7.83 (d, 1H, J = 8.4), 7.83 (dd, 1H, J = 8.4, 1.8 Hz), 7.80 (s, 1H), 7.55 (d, 1H, J = 4.0 Hz), 7.45 (d, 1H, J = 4.0 Hz), 4.76 (s, 2H).

{5-[5-(2-Chloro-5-trifluoromethyl-phenyl)-furan-2-ylmethylene]-4-oxo-2-thioxo-thiazolidin-3-yl}-acetic acid (BI-11B3). **BI-11B3** was obtained as 0.121 g of yellow solid in 59% yield. ^{1}H NMR (300 MHz, *d*-DMSO) δ 8.26 (d, 1H, J = 2.1 Hz), 7.93 (d, 1H, J = 8.1 Hz), 7.87 (dd, 1H, J = 8.1, 2.1 Hz), 7.85 (s, 1H), 7.62 (d, 1H, J = 3.9 Hz), 7.48 (d, 1H, J = 3.9 Hz), 4.73 (s, 2H).

Inhibition of Metalloproteinase (MMP)-2 and MMP-9 Activity and MAPKK Cleavage Assay. MMP-2 and MMP-9 were activated by incubation with APMA (1 mM) at ambient temperature for 4 and 18 h, respectively. Activated proteases (25 nM) were incubated with 50 μM fluorogenic substrate ES001 (R & D Systems) with or without 100 μM each inhibitor. Substrate hydrolysis was measured by obtaining relative fluorescence after a reaction time of 10 min at 37°C using the Gemini EM plate reader (Molecular Devices) at excitation and emission wavelengths of 320 and 405 nm, respectively.

Construction and Expression of MAPKK1. The full-length MAPKK1 cDNA was cloned into the pET15b vector (EMD Biosciences/Novagen, San Diego, CA). The recombinant N-terminally His-tagged MAPKK1 construct was expressed in *Escherichia coli* BL21 cells. The expression of the His-MAPKK1 chimera was induced by isopropyl β-D-thiogalactoside. The soluble His-MAPKK1 protein was purified from the cell lysate on a HiTrap Chelating High Performance Ni-Sepharose column (Amersham Pharmacia Biosciences). His-MAPKK1 was eluted from the column with a linear 0–300 mM imidazole gradient. The high purity of the isolated His-MAPKK1 was confirmed by SDS/PAGE and mass spectrometry analyses.

LF Proteolysis of MAPKK1. His-MAPKK1 (700 ng) was coincubated at 30°C for 2 h with LF (10 ng) in 20 μl of 20 mM Hepes, pH 7.4. The digest reactions were stopped by adding 4 μl of 5% SDS. The digest samples were analyzed on SDS/PAGE on a 10% acrylamide gel. Where indicated, increasing concentrations of LF inhibitors (0.1–20 μM) were added to the samples to inhibit the LF proteolysis of MAPKK1.

Cytotoxicity Assay. Murine macrophage-like cell line RAW 264.7 was a kind gift of M. Fukuda (Burnham Institute, La Jolla, CA). The cells were grown to confluence in wells of a 48-well plate (Costar) in DMEM (Gibco) supplemented with 10% FCS (Sigma). The cells were replenished with fresh medium (0.1 ml per well) and then incubated with LF inhibitors for 4 h to allow the inhibitors to penetrate the cell compartment. PA and LF were then added to the final concentration of 500 ng/ml and 25 ng/ml, respectively. After incubation for an additional 4 h, cell viability was assessed by 3,[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT) staining. Cells were incubated with 0.5 mg/ml MTT in DMEM for 45 min at 37°C; the medium was aspirated, and the blue pigment produced by the viable cells was solubilized with 0.5% SDS/25 mM HCl in 90% isopropyl alcohol. The concentration of oxidized MTT in the samples was measured at 570 nm by using a microplate reader. Each datum point represents the results of at least three independent experiments performed in duplicate. A percentage of viable cells was calculated by using the following equation:

$$\frac{\begin{array}{c}(\mathrm{A}_{570}\text{ of cells treated with LF, PA and inhibitor})\\ -\ (\mathrm{A}_{570}\text{ of cells treated with LF and PA})\end{array}}{\begin{array}{c}(\mathrm{A}_{570}\text{ of cells treated with LF alone})\\ -\ (\mathrm{A}_{570}\text{ of cells treated with LF and PA})\end{array}}.$$

X-Ray Crystallography. LF wild-type native protein was crystallized by using a concentration of 13 mg/ml LF. Crystals were grown from 1.70 M $(NH_4)_2SO_4$/0.2 M Tris·HCl, pH 8.0–7.5/2 mM EDTA by using the hanging drop vapor diffusion method, as described in ref. 30. Monoclinic crystals appeared after 4 days to 2 weeks and were then harvested for experiments. The LF crystals are of the space group monoclinic $P2_1$, with averaged unit cell dimensions a = 96.70 Å, b = 137.40 Å, c = 98.30 Å, α = 90, β = 98, and γ = 90, containing two molecules per asymmetrical unit. The soaked crystals for this crystal complex had unit cell dimensions a = 95.96 Å, b = 136.65 Å, c = 97.90 Å, α = 90, β = 98.23, and γ = 90.

LF native crystals were harvested from the hanging drops in which they were grown and bathed in several rounds of fresh buffer without EDTA, consisting of 1.90 M $(NH_4)_2SO_4$, 0.2 M Tris·HCl, pH 8.0, and finally left to soak in this solution for an additional 30 min. These crystals were then used for obtaining the protein–inhibitor–zinc complexes. All manipulations were done at room temperature (23–26°C).

The LF-**BI-MFM3**-Zn protein–inhibitor crystal complex was obtained by soaking an individual native LF monoclinic $P2_1$ crystal in a solution of 1 mM Zn SO_4/1.90 M $(NH_4)_2SO_4$/0.2 M Tris·HCl, pH 8.0, for 10 min, then transferring the crystal to a solution of 1.0 mM MFM3 inhibitor, 1%(vol/vol) DMSO/1.9M $(NH_4)_2SO_4$/0.2

M Tris·HCl, pH 8.0, for 30 min. Finally, the crystal was transferred into a cryoprotectant solution of 1.0 mM MFM3 inhibitor/2.4 M $(NH_4)_2SO_4$/0.2 M Tris·HCl, pH 8.0/2 mM EDTA/25% glycerol and soaked at room temperature for an additional 1 min. The crystal was then immediately mounted onto a cryoloop and flash-frozen in liquid nitrogen. All data collection was done at 100 K.

The dataset for the LF complexes was collected in the Burnham Institute's in-house x-ray facility on a Rigaku (Tokyo) FR-E rotating copper anode generated x-ray beam (wavelength = 1.5418 Å). X-ray diffraction data were collected for LF-**BI-MFM3-**Zn^{2+} to resolution limits of 2.67 Å.

Using PDB_ID 1J7N as the starting model (without water molecules), the model of LF (with Zn^{2+} ions in the catalytic site) alone was put through rigid body refinement and then minimization, before the first initial maps were calculated for model building and further refinement. Excess electron density at σ level 1.0 indicated the binding location of the inhibitor in the active site of LF. The model of the inhibitor was then built into this position and further refined in CNS (31). The final R factors were $R_{\text{free}} = 27.6\%$ and $R_{\text{work}} = 23.4\%$ for LF-**BI-MFM3**-Zn. The current models fall within the limits of all of the quality criteria of the program PROCHECK (www.biochem.ucl.ac.uk/~roman/procheck/procheck.html) from the CCP4 (www.ccp4.ac.uk/main.html) suite. The coordinates have been submitted to the Protein Data Bank.

***In Vivo* Studies.** *In vivo* studies were conducted in the laboratories of K.A. Three compounds, **BI-11B1**, **BI-11B2**, and **BI-11B3**, were prepared in two steps. One week before the start of the study, 200 mg of each compound was dissolved in 800 μl of DMSO and stored at −20°C. Immediately before injection, each substance was diluted in PBS, resulting in a final concentration of 0.5 mg/ml in 2% DMSO. The animals were challenged on day 0 with 2×10^7 spores per mouse in PBS through i.p. injection. Treatment was started 24 h after challenge. Treatment regimes include ciprofloxacin alone (50 mg/kg) or a combination of ciprofloxacin with **B1-11B1**, **B1-11B2**, or **B1-11B3** LF inhibitors (5 mg/kg). Animals were closely monitored twice per day until day 14 after infection.

Animals were divided into five groups based on their treatment:

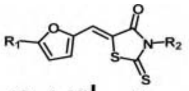
- Control (not treated).
- Ciprofloxacin alone.
- Ciprofloxacin + **B1-11B1**.
- Ciprofloxacin + **B1-11B2**.
- Ciprofloxacin + **B1-11B3**.

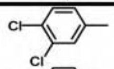
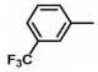
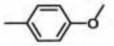
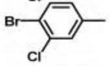
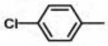
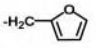
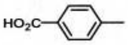
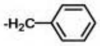
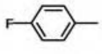
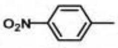
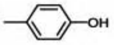
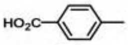
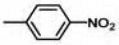
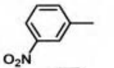
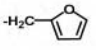
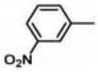
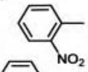
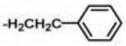
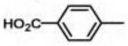
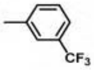
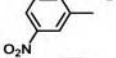
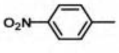
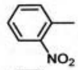
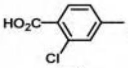
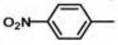
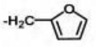
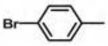
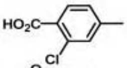
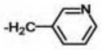
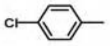
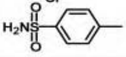
Ciprofloxacin and LF blocking substances were administered through i.p. injection with a volume of 200 μl for each once per day for 10 days. All surviving animals were killed on day 14 by CO_2 inhalation. Sick animals that appeared moribund (exhibiting a severely reduced or absent activity or locomotion level, an unresponsiveness to external stimuli, or an inability to obtain readily available food or water, along with any of the following accompanying signs: ruffled haircoat, hunched posture, inability to maintain normal body temperature, signs of hypothermia, respiratory distress, or any other severely debilitating condition) were killed on the same day.

Results

Peptide cleavage fluorescence assays are straightforward when screening for potent LF inhibitors. Keeping in mind that LF specifically cleaves proteins of the MAPKK family (13) at their N termini (14), we used the optimized peptide MAPKKide as a substrate for our fluorescence assay (List Biological Laboratories). Particularly, MAPKKide is derived from the MAPKK-2 substrate

Table 1. Compounds tested against LF



#	R1	R2	IC50 (μM)	#	R1	R2	IC50 (μM)
BI-9B9b	H	-CH2COOH	140	12		H	7.4
1			300	13		H	7.0
2			150	14			6.0
3		-CH2CH=CH2	50	15		-CH2COOH	5.5
4			37.7	16			4.8
5			36.3	17		-CH2COOH	3.1
6			31.9	18			2.9
7		-(CH2)3COOH	20	19		-CH2CH2COOH	2.7
8		-CH2CH2COOH	12.8	20		-CH2CH=CH2	2.7
9			12.6	21		-(CH2)3COOH	2.3
10			9.9	BI-MFM3		-CH2CH2COOH	0.8 (Ki)
11		-CH2COOH	9.1				

for LF, and it is intramolecularly quenched by fluorescence resonance energy transfer. The C-terminally linked fluorophore is a FITC, and the acceptor chromophore is 4-([4′-(dimethylamino)-phenyl]azo) benzoic acid (DABCYL). After cleavage by LF, it is possible to detect a sensible fluorescence increase in the reaction solution setting excitation and emission wavelengths at 485 and 590 nm, respectively.

Although it would be a sensible strategy to use such an assay to screen several thousand compounds, herein we report a different approach, based on the initial identification of preferred weakly binding scaffolds to be successively used as a starting point for iterative optimizations (15). Although the fluorescence-based assay is a robust technique to search for very potent inhibitors, it becomes more ambiguous in detecting weaker ligands (>100 μM), possibly due to interference introduced by test compounds (normally used at high concentration) in the spectrophotometric assay. For this reason, we relied on a NMR-based enzymatic assay, which is unlikely to lead to false positives (16–23). Recently, the use of ^{19}F-1D NMR to detect enzyme activity and inhibition both in proteases and kinases has been reported (22). NMR experiments based on observation of ^{19}F present several benefits. Above all, this nucleus shows sensitivity comparable to that of ^{1}H, so that it is possible to acquire 1D spectra in a relatively short time. Moreover, because of its large anisotropy, ^{19}F chemical shifts are spread over a wide spectral window; as a consequence, the potential spectral resolution is greatly improved. It is also worth underlining that overlapped signals arising from buffers, solvents, and other reaction components are unlikely to occur in ^{19}F-NMR spectra.

We succeeded in detecting LF inhibition by ^{19}F-NMR using the fluorinated peptide Ac-A-R-R-K-K-V-Y-P-NH-Ph-CF$_3$ as an enzymatic substrate (24, 25). Cleavage of the peptide occurring at the Pro-Xxx position deeply affects chemical environment of ^{19}F nuclei because of the conversion of the amide functionality into an amine with release of pCF$_3$-aniline. Therefore, it is possible to monitor LF kinetics and inhibition by monitoring ^{19}F NMR signals of uncleaved peptide substrate and the reaction product pCF$_3$-aniline.

We applied such a strategy to a small but diversified library of ≈300 compounds representing most of the scaffolds commonly found in drugs (26). This library was designed by selecting compounds on the basis of their drug-like properties, ease of synthesis, and/or availability of several hundred derivatives. Therefore, a library of only 300 scaffolds representative of a chemical space of several hundreds of thousands compounds was tested.

Application of this strategy led to the identification of compound **BI-9B9b** (Table 1), which exerted 50% LF inhibition at 140 μM concentration. Exploring commercially available chemical repositories, such as Maybridge, Chembridge, and those listed by Chemnavigator (San Diego), we spotted the most representative derivatives of **BI-9B9b** (22 among ≈680 analogues identified by using a 2D substructure search). These compounds were selected on the basis of an additional experiment in which we could not detect any appreciable LF inhibition (up to 500 μM) when the furan ring was substituted by a benzene ring, indicating that both rings of **BI-9B9b** are important for binding. All selected compounds (Table 1) have been tested by both NMR- and traditional fluorescence-based assays. Compounds **BI-MFM3** and **17**–**21** emerged as very effective inhibitors with a $>70\%$ LF inhibition at 10 μM concentration.

For each compound, the ^{19}F-1D NMR assay was performed. The results of a representative assay are shown in Fig. 1. The cleavage of the fluorinated peptide (20 μM) by LF (50 nM) led to a strong NMR signal of pCF$_3$-aniline (Fig. 1*a*). A known hydroxamate inhibitor of LF, GM6001 (27), at a concentration of 20 μM, demonstrated a 50% inhibition of the LF activity (Fig. 1*b*). In turn, **BI-MFM3** (20 μM) fully inhibited the cleavage of the fluorinated peptide by LF, thus pointing to **BI-MFM3** as a more potent inhibitor against LF when compared with GM6001 (Fig. 1*c*).

Subsequently, the IC$_{50}$ value of the inhibitors was determined in the MAPKKide peptide cleavage assays (Table 1). The IC$_{50}$ of the most potent inhibitor, **BI-MFM3**, was 1.7 μM (Fig. 1*e*). To confirm and extend these findings, we measured the K_i value and the type of inhibition of LF by **BI-MFM3** (Fig. 1*f*). For these purposes, we initially determined the K_m and V_max of the MAPKKide cleavage by LF, which were 2.22 ± 0.2 μM and 0.0942 ± 0.0007 μmol min^{-1}·mg^{-1}, respectively. We then used a 10 μM concentration of **BI-MFM3** to identify the inhibitor's K_i value, which was determined to be 0.8 ± 0.3 μM in our assay. Because **BI-MFM3** affected the K_m rather than the V_max of the cleavage reactions, **BI-MFM3** is considered to be a competitive inhibitor of LF.

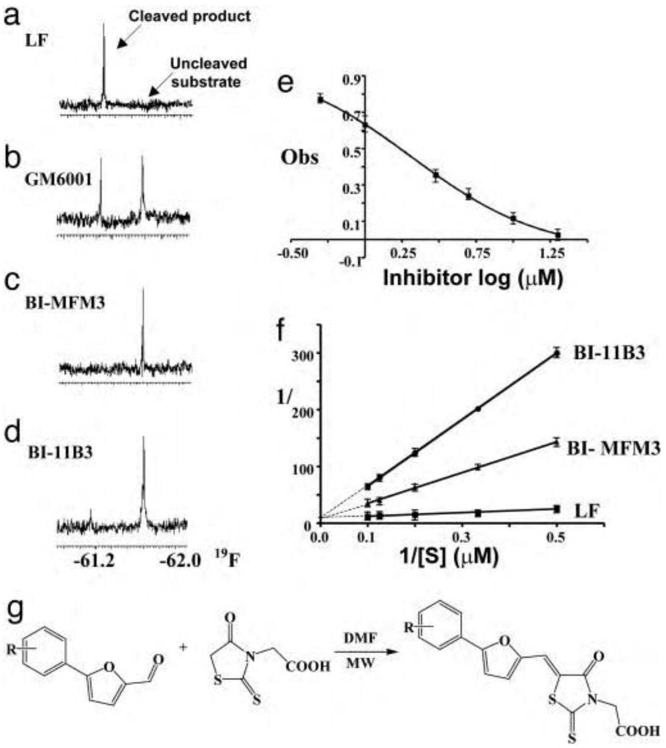


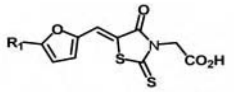
Fig. 1. Inhibition of anthrax LF. (*a*) ^{19}F NMR spectra of the peptide substrate in presence of LF. (*b*) Effect of GM6001 (20 μM). (*c*) Effect of **BI MFM3** (20 μM). (*d*) Effect of **BI-11B3** (0.8 μM). (*e*) IC$_{50}$ evaluation for compound **BI-MFM3**; (*f*) Lineweaver–Burk K_m and $K_{\mathrm{m(app)}}$ evaluation for LF, **BI-MFM3**, and **BI-11B3**, respectively. Each measurement was performed in triplicate. (*g*) Synthetic scheme adopted for the synthesis of compounds listed in Table 2.

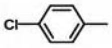
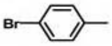
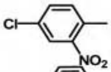
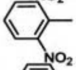
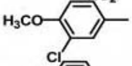
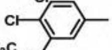
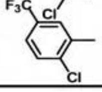
To assess the specificity of our compounds against other MMPs, we tested them against two more related MMPs: MMP2 and MMP9, which appear to be the most functionally important human MMPs (28). Although the IC$_{50}$ of the initial scaffold **BI-9B9b** against MMP-2 and MMP-9 was ≈10 μM, **BI-MFM3** did not inhibit these proteases at concentrations up to 100 μM. To evaluate the activity of **BI-MFM3**, **19**, and **21** in cell-based tests, we used murine RAW264.7 cells, which are sensitive to LF and undergo apoptosis if treated with the bipartite PA-LF toxin. Compounds **19**, **21**, and especially **BI-MFM3** were capable significantly of rescuing cells from the toxic action of LF at micromolar concentration (not shown). These observations have suggested that these three identified leads provide a solid foundation for the design of more effective drugs with improved efficiency against LF.

Encouraged by these data, we sought to design additional compounds with improved inhibitory properties on the basis of structure–activity relationship data reported in Table 1, as follows.

The presence in R$_1$ position of a substituted phenyl with a small electronegative group significantly increases the inhibitory activity, whereas a small group containing a carboxylic moiety in position R$_2$ also seems to improve the potency. On the contrary,

Table 2. Compounds and their measured LF inhibition



#	R_1	% Yield	IC_{50}(nM)
BI-11A9		88	900
BI-11A10		89	850
BI-11A11		53	500
BI-11A12		44	3100
BI-11B1		83	298
BI-11B2		30	265
BI-11B3		66	32(K_i)

a large group, such as a substituted phenyl in R_2, causes a dramatic reduction of activity, especially if not balanced with an effective group in R_1.

In particular, a comparison of activities for compounds **8** and **17** suggested that an acetyl group would be the preferred substituent in R_2. Regarding R_1 group, substitutions in all positions on the phenyl ring seem to be equally effective, thus indicating that compounds with multiple substitutions may result in increased activity. To verify these hypotheses, we elaborated a synthetic scheme (Fig. 1*g*) to afford additional **BI-MFM3** analogues (Table 2). In agreement with the above observations, each of the synthesized compounds showed an increased inhibitory activity compared with **BI-MFM3** in both the fluorescence and NMR-based assays. Particularly, compound **BI-11B3** appeared to be the most potent inhibitor, with a K_i value of 32 ± 22 nM (Fig. 1*f*). A NMR-based assay using the fluorinated peptide also confirmed the potency of **BI-11B3** in inhibiting LF (Fig. 1*d*). Furthermore, to rule out the possibility of eventual nonspecific interactions, we verified that no substantial changes in the IC_{50} values for compounds **BI-11B1** and **BI-11B3** were detected when increasing 7-fold the protein concentration (from 25 to 175 nM), as well as by preincubating the

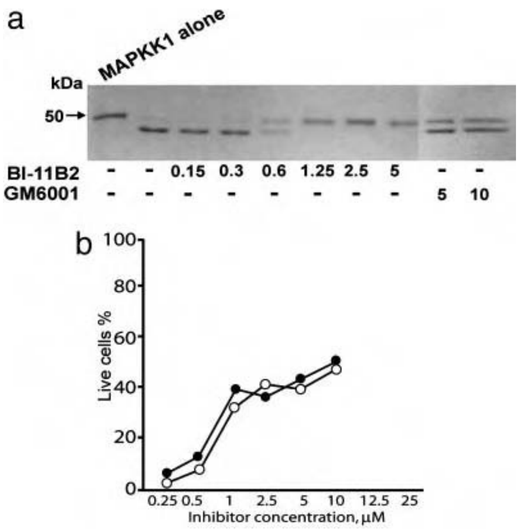


Fig. 2. *In vitro* and cell-based evaluation. (*a*) **BI-11B2** efficiently protects the purified MAPKK-1 against LF cleavage *in vitro*. **BI-11B2** and **GM6001** (as control) were each coincubated with LF and MAPKK1. The digest samples were analyzed by SDS/PAGE to determine the specific conversion of MAPKK1 into the 45-kDa cleavage product. (*b*) Inhibitors **BI-11B2** and **BI-11B3** are effective in protecting MAPKK1 and murine macrophage RAW264.7 cells against LF. Cells were coincubated with anthrax PA (500 ng/ml) and LF (40 ng/ml). The indicated concentrations of the inhibitors were added to the cells. In 4 h, the residual viable cells were measured by adding the tetrazolium salt 3,[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT). The data show that inhibitors **BI-11B2** (open circles) and **BI-11B3** (filled circles) protect cells from the cytotoxic effect by LF and PA.

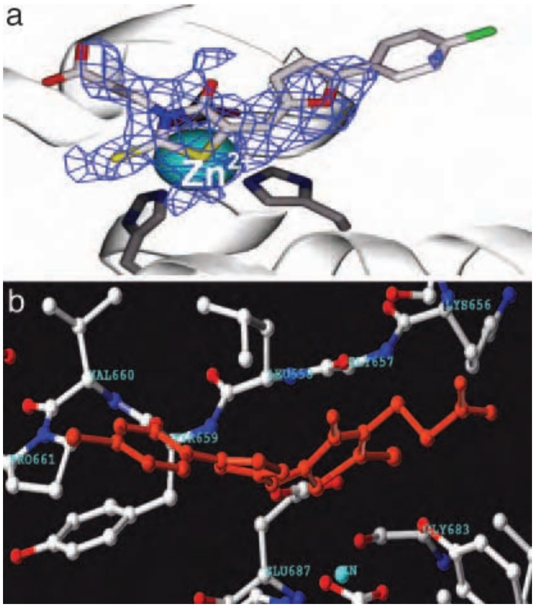


Fig. 3. Crystal structure of the LF-**BI-MFM3**–zinc complex. (*a*) Detailed view of the electron density trace and overall model fit of **BI-MFM3**. (*b*) Detail of the binding site of LF for **MFM3** (both shown in stick representation). These data are at a resolution limit of 2.67 Å. The small molecule appears to be interacting with the zinc atom in the LF active site via an S atom. Additional interactions are mainly of hydrophobic nature involving the aromatic rings of the inhibitor and hydrophobic side chains of LF. Prepared by using SPOCK (http://quorum.tamu.edu/spock) and SYBYL (Tripos Associates, St. Louis).

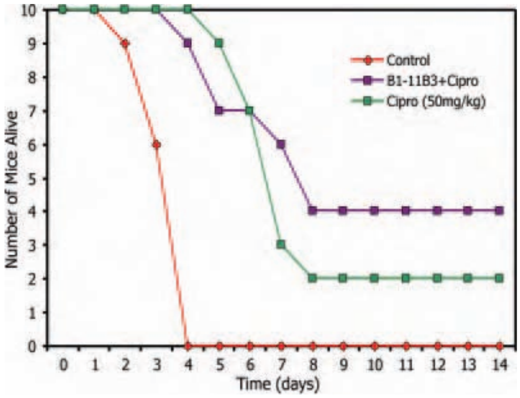


Fig. 4. Comparison of survival rates between different treatments regimes. DBA2 mice were infected with *B. anthracis* Sterne spores at a dosage of 2 × 10^7 per mouse in 200 µl of PBS on day 0 through i.p. injection. Animals were treated with ciprofloxacin alone or in combination with lethal toxin blocking substance **B1-11B3**. Similar data were obtained with compound **BI-11B1** (not shown). Treatment was started 24 h postexposure and continued for 10 days. Nontreated mice were used as a control. Animals were monitored for 14 days after infection.

compounds with LF for 30 min. These simple tests have been shown to give dramatically different IC_{50} values in the presence of nonspecific ligand–protein interactions (29).

To corroborate these findings, we also tested the efficiency of **BI-11B1**, **BI-11B2**, and **BI-11B3** in protecting MAPKK1, a natural protein target of LF, from the LF proteolysis *in vitro*. In the concentration range of 1–2 μM, each of the three inhibitors was capable of protecting MAPKK1 from LF cleavage, and each of the inhibitors was superior relative to the GM6001 hydroxamate (Fig. 2*a*). **BI-11B1** and especially **BI-11B2** and **BI-11B3** were highly potent in protecting the RAW264.7 cells against LF-induced cytotoxicity with IC_{50} values of 2–5 μM (Fig. 2*b*), compared with 50 μM observed with GM6001. Thus, **BI-11B2** and **BI-11B3** were at least 1 order of magnitude more potent in cell-based assays than the GM6001 hydroxamate. In these assays, we could not observe 100% protection with our compounds, probably due to reduced solubility at higher concentrations and/or limited macrophage cell membrane permeability. However, after initial infection, it is reasonable to assume that even a 60% (as shown) or lower rescuing of macrophage activity could be sufficient to combat bacterial proliferation.

To obtain further insights on the mechanism of action of our compounds, we have also initiated a structural characterization of the most potent compounds by means of x-ray crystallography (Fig. 3). We are currently trying to obtain x-ray high-resolution structures for LF in complex with compounds **BI-MFM3** as well as **BI-11B1** and **BI-11B3.** Details of the 3D structure of the complex between LF and **BI-MFM3** are reported in Fig. 3. Analysis of the docked structure revealed that the rhodanine ring is capable of interacting with Zn^{2+} metal-ion via the thiazolidine sulfur atom, which explained the activity of the scaffold **BI-9B9b** (Table 1) against LF and other MMPs (Fig. 3). The carboxylic group of **BI-MFM3** is pointing toward a hydrophilic region of the protein close to its surface, which explains the variability of the substitutions allowed at this position and the increased affinity of the compounds when R_2 is a small charged group (Table 1). In addition, hydrophobic interactions between the phenyl ring group and hydrophobic side chains of LF were also observed, and most likely they are responsible for the increased affinity and selectivity of our compounds for LF vs. other MMPs and the increased affinity with bisubstituted compounds (Table 2). The electron density of the benzene ring is less evident in the structure of **BI-MFM3**, indicating a possible conformational mobility.

To evaluate the efficacy of LF inhibitors when combined with antibiotic against postexposure to *Bacillus anthracis* (Sterne strain), we tested the effect of our compounds in female DBA2 mice (9–11 weeks old) with body weights between 20 and 24 grams (Taconic Laboratories, Germantown, NY). The animals were challenged on day 0 with 2×10^7 spores per mouse in PBS through i.p. injection. Treatment, started 24 h after challenge, included ciprofloxacin alone (50 mg/kg) or a combination of ciprofloxacin with LF inhibitors **B1-11B1**, **B1-11B2**, or **B1-11B3** (5 mg/kg). Animals were closely monitored twice per day until day 14 after infection. Survival rates of mice treated with **B1-11B3** in combination with ciprofloxacin compared with the survival rates of mice treated only with ciprofloxacin are shown in Fig. 4. LF inhibitor **B1-11B1** in combination with ciprofloxican provided 40% protection against the *B. anthracis* Sterne infection compared with the conventional treatment ciprofloxacin that protected only 20% of the animals.

Discussion

Despite the current threat of bioterrorism, there is no specific and effective therapy for inhalation anthrax, a deadly disease in humans. The proteolytic activity of LF MMPs is essential for the onset, progression, and lethality of anthrax. We have applied a fragment-based methodology that has led us to the identification of an initial LF inhibitory scaffold. The iterative optimizations of this scaffold have resulted in a series of *phenylfuran-2-ylmethylenerhodanineacetic acid* derivatives with a nanomolar inhibitory activity against LF. During the past two decades, major efforts from both the academic and pharmaceutical industry sectors have been devoted to the identification of metal–protease inhibitors, given their pivotal role in virtually any human disease (32). A common approach to the development of such inhibitors relied on structure-guided derivatizations of Zn^{2+} chelating compounds, most commonly hydroxamate, to yield potent and possibly selective compounds (33). Likewise, the scaffolds reported here could well be used to derive additional potent and selective inhibitors of several other Zn-metallo-proteases, also aided by our structural analysis and structure–activity relationship data. The LF inhibitors we have derived are capable of protecting macrophages from LF-induced cytotoxicity at concentrations well below those needed with a nonselective hydroxamate-based protease inhibitor and show synergistic protection with ciprofloxacin *in vivo*. Although further in-depth pharmacokinetics studies will be necessary to establish the exact dosage and regimen of the compound and to evaluate the efficacy of the proposed combination therapy against inhalation anthrax, the data reported here provide *in vivo* evidence of the effectiveness of LF inhibitors in the treatment of postexposure anthrax. As such, our lead compounds hold great promise for the development of novel, safe, and effective emergency therapy of postexposure inhalation anthrax.

This project was sponsored in part by National Institutes of Health–National Institute of Allergy and Infectious Diseases Grant 1UO1AI056385 (to R.A.A., A. Y. Strongin, R.C.L., and M.P.).

1. Smith, H. & Keppie, J. (1954) *Nature* **173,** 869–870.
2. Hanna, P. (1998) *Curr. Top. Microbiol. Immunol.* **225,** 13–35.
3. Hanna, P., Acosta, D. & Collier, R. J. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10198–10201.
4. Bradley, K.A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. (2001) *Nature* **414,** 225–229.
5. Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. (1997) *Nature* **385,** 833–838.
6. Scobie, H. M., Rainey, G. J., Bradley, K. A. & Young, J. A. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 5170–5174.
7. Leppla, S. H. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3162–3166.
8. Vitale, G., Pellizzari, R., Recchi, C., Napolitani, G., Mock, M. & Montecucco, C. (1998) *Biochem. Biophys. Res. Commun.* **248,** 706–711.
9. Duesbery, N. S., Webb, C. P., Leppla, S. H., Gordon, V. M., Klimpel, K. R., Copeland, T. D., Ahn, N. G., Oskarsson, M. K., Fukasawa, K., Paull, K. D., *et al*. (1998) *Science* **280,** 734–737.
10. Park, J. M., Greten, F. R., Li, Z. W. & Karin, M. (2002) *Science* **297,** 2048–2051.
11. Sellman, B. R., Mourez, M. & Collier, R. J. (2001) *Science* **292,** 695–697.
12. Min, D. H., Tang, W. J. & Mrksich, M. (2004) *Nature Biotech.* **22,** 717–723.
13. Vitale, G., Bernardi, L., Napolitani, G., Mock, M. & Montecucco, C. (2000) *Biochem. J.* **352,** 739–745.
14. Pellizzari, R., Guidi-Rontani, C., Vitale, G., Mock, M. & Montecucco, C. (1999) *FEBS Lett.* **462,** 199–204.
15. Pellecchia, M., Becattini, B., Crowell, K. J., Fattorusso, R., Forino, M., Fragai, M., Jung, D., Mustelin, T. & Tautz, L. (2004) *Exp. Opin. Ther. Tar.* **8,** 597–611.
16. Becattini, B., Sareth, S., Zhai, D., Crowell, K. J., Leone, M., Reed, J. C. & Pellecchia, M. (2004) *Chem. Biol.* **11,** 1107–1117.
17. Pellecchia, M., Sem, D. S. & Wuthrich, K. (2002) *Nat. Rev. Drug Discov.* **1,** 211–219.
18. Meyer, B. & Peters, T. (2003) *Angew. Chem. Int. Ed.* **42,** 864–890.
19. Zerbe, O. (2002) in *BioNMR in Drug Research* (Wiley, New York).
20. Pellecchia, M., Meininger, D., Dong, Q., Chang, E., Jack, R. & Sem, D. S. (2002) *J. Biomol. NMR* **22,** 165–173.
21. Hajduk, P. J., Olejniczak, E. T. & Fesik, S. W (2001) *J. Am. Chem. Soc.* **123,** 3149–3150.
22. Mayer, M. & Meyer, B. (1999) *Angew. Chem. Int. Ed. Engl.* **38,** 1784–1788.
23. Dalvit, C., Ardini, E., Flocco, M., Fogliatto, G. P., Mongelli, N. & Veronesi, M. (2003) *J. Am. Chem. Soc.* **125,** 14620–14625.
24. Turk, B. E., Wong, T. Y., Schwarzenbacher, R., Jarrell, E. T., Leppla, S. H., Collier, R. J., Liddington, R. C. & Cantley, L. C. (2004) *Nat. Struct. Mol. Biol*. **11,** 60–66.
25. Panchal, R. G., Hermone, A. R., Nguyen, T. L., Wong, T. Y., Schwarzenbacher, R., Schmidt, J., Lane, D., McGrath, C., Turk, B. E., Burnett, J., *et al.* (2004) *Nat. Struct. Mol. Biol.* **11,** 67–72.
26. Bemis G. W. & Murcko, M. A. (1996) *J. Med. Chem.* **39,** 2887–2893.
27. Knight, G. C. (1992) *FEBS Lett.* **296,** 263–266.
28. Kridel, S. J., Chen, E., Kotra, L. P., Howard, E. W., Mobashery, S. & Smith, J. W. (2001) *J. Biol. Chem.* **276,** 20572–20578.
29. McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. (2002) *J. Med. Chem.* **45,** 1712–1722.
30. Pannifer, A. D., Wong, T. Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, D. B., Collier, R. J., Park, S., Leppla, S. H., *et al.* (2001) *Nature* **414,** 229–233.
31. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
32. Rush, T. S., 3rd, & Powers, R. (2004) *Curr. Topics Med. Chem.* **4,** 1311–1327.
33. Levin, J. I. (2004) *Curr. Topics Med. Chem.* **4,** 1289–1310.