REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
8/30/06	Final Report	

4. TITLE AND SUBTITLE
Development of combined computational and experimental approaches for using molecular engineering in the design, construction, and analysis of integrated biosensor microsystems

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-02-1-0063

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Homme W Hellinga

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University Medical Center
Department of Biochemistry, Box 3711
Durham, North Carolina 27710

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR
875 N Randolph St
Arlington VA 22203

AFRL-SR-AR-TR-06-0405

12. DISTRIBUTION / AVAILABILITY STATEMENT

Unrestricted Approved for public release: Distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

SEE ATTACHMENT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*

FINAL REPORT

LRIR: F49620-02-1-0063

Title: Development of combined computational and experimental approaches for using molecular engineering in the design, construction, and analysis of integrated biosensor microsystems

Laboratory Task Manager: **Dr. Homme W. Hellinga**

Commercial Phone: **919-681-5885** DSN: _____ FAX: **919-684-8885**

Mailing Address: **Department of Biochemistry, Box 3711**
Duke University Medical Center
Durham, North Carolina 27710

E-Mail Address: **hwh@biochem.duke.edu**

AFOSR Program Manager: **Dr. Hugh C. De Long**

OBJECTIVES

The broad objective of the project was to develop and experimentally validate computational methods for the design of protein-based biosensors that selectively bind to a wide variety of small molecules or proteins, to construct a family of biosensors for the detection of chemical or biological threats. More specifically, the project aimed to deliver biosensors for nerve agent surrogates, to widen the scope ofthe computational design methodology to a) tackle ligands of increasing conformational complexity, b) address protein-protein, and c) protein-DNA interactions.

APPROACH

Structure-based protein design methods constitute the basis for this project. These methods aim to describe molecular recognition using semi-empirical potential functions that capture van der Waals, hydrogen-bonding, electrostatic, and solvation contributions. These descriptions are combined with representations of the dominant degrees of freedom in a protein design calculations: the sequence and structure of amino acid side-chains placed within the three-dimensional frame work of a parent protein ("the scaffold"), and the translations/rotational degrees of freedom of a ligand. Discrete combinatorial optimization algorithms are then used to identify a combination of amino acids and docked ligand conformation that represents the global energy minimum of the potential function.

Computationally generated solutions are then tested experimentally by making the specified mutations in the gene encoding the parent protein using oligonucleotide-directed mutagenesis, producing the mutant protein by heterologous over-expression in E. coli followed by protein purification and appropriate biochemical assays to test for the presence of the desired function.

ACCOMPLISHMENTS

The scope of the design of receptors with drastically altered ligand-binding properties was successfully accomplished. The computational methods start with a high-resolution X-ray structure and use this to predict the requisite mutations needed to change the binding specificity of a ligand-binding site. The calculations needed to predict the mutations necessary for converting a binding site from recognizing a natural ligand to being complementary to a radically different molecule are non-trivial. About 10-12 residues are involved in forming a complementary surface for ligands in the typical 300-400 dalton size. First, the new ligand is docked in place of the old one; second, all of the residues in the complementary surface are mutated simultaneously to identify the appropriate combination that forms an optimal lock-and-key fit between the protein and the new ligand. At each mutable position, there are 20 possible mutations, and for each mutant there are several possible structures. In practice ~6,500 structures represent all 20 amino acids. Thus there are $6{,}500^{n}*m$ possible combinations, where n is the number of mutable positions, and m represents all the possible conformations of the docked ligand (typically $\sim 10^{6}$). For redesigning the PBPs this means that there are typically $\sim 10^{150}$ possible combinations within which a small number of solutions needs to be identified that are predicted to have good lock-and-key fits. Prior to the start of the project, we had developed deterministic algorithms that can tackle these enormous combinatorial problems to identify the global energy minimum in reasonable compute time (a few days on a fast Pentium processor), using amino acid packing in protein interiors as the test case. During the project period these were adapted and improved to tackle protein-ligand interactions. Of particular note is the introduction of constraints that ensure that all possible hydrogen bonds are satisfied in a ligand.

As an experimental system we used members of the periplasmic-binding protein (PBP) superfamily. Many of these proteins are monomeric soluble receptors that are easily expressed in *E. coli* and purified by immobilized metal affinity chromatography. They consist of two domains linked by a hinge region, and undergo large, ligand-mediated conformational changes. Previously we have been able to exploit these conformational changes to construct reagentless fluorescent and electrochemical sensors that link ligand binding to changes in fluorescence intensity emission or electrochemical activity. The high-resolution structures of several E. coli PBPs have been determined by X-ray crystallography, and form the starting points for the design calculations.

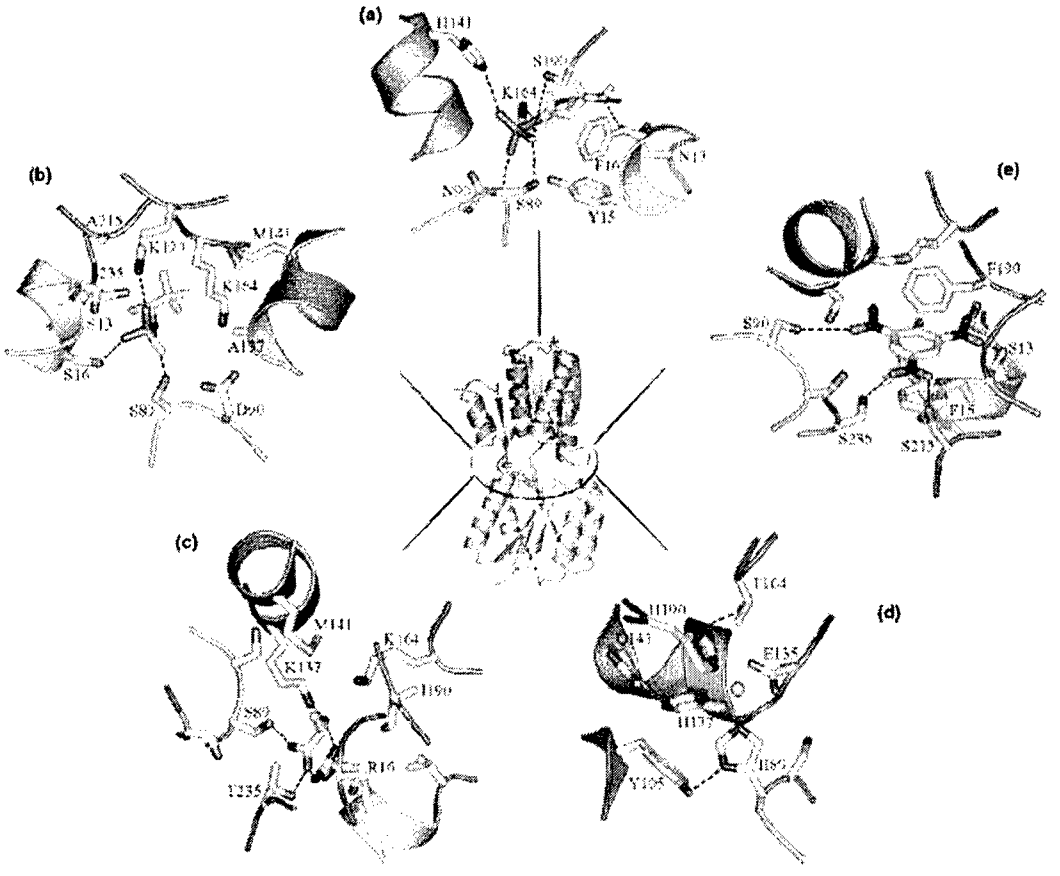

Figure 1. *Diversification of ribose-binding protein by computational protein design. A, lactate; b, PMPA (nerve agent surrogate); c, lactate; d, zinc(II); e, TNT.*

To address the aim of constructing sensors for nerve agent surrogates, we designed fifteen PMPA receptors: three in ribose-binding protein (GBP); twelve in glucose-binding protein (GBP). These receptors bind PMPA with affinities ranging from 45 nM – 10 μM. We have analyzed the protein-ligand interactions of two GBP-based designs in more detail using alanine-scanning mutagenesis. These studies have established that a) relatively high-affinity receptors have been designed that may function as reagentless recognition elements in novel sensors for nerve agents, b) the interactions predicted by computational design are present in the actual proteins, c) the rank ordering predicted by the computational design algorithm correlates roughly with that observed experimentally.

To test the scope of the algorithm we constructed several receptors for other ligands, including glyphosate and ibuprofen. These studies extend the demonstration of the scope of the design methodology, and may provide further useful recognition elements for biosensor development (Figure 1).

The algorithms were further developed to address protein-protein and protein-DNA interactions. We chose to tackle the protein of the ab initio design of protein-

protein interactions, that is to say to try and construct an interface between two proteins that are not known to interact. To do this, first "poses" have to be calculated for the two partners. This requires a six-dimensional search, for which we used low-resolution representation of the structures. The low-resolution representation used two spheres to represent an amino acid residue. The first sphere is positioned at the position of the original C_α atom in the structure; the second sphere is positioned along the C_α-C_β axis. A simple, triangular well potential is used to assign a long-range attractive force to the spheres. A Monte-Carlo/simulated annealing protocol is used to identify poses that provide (near-)optimal interdigitation of the spheres in an interface between the two partners (figure 2). Once likely poses have been identified, the low-resolution poses are converted in full atomistic representations, and sequences can be calculated using the algorithms developed for combinatorial optimization of amino acid side-chain structure and sequence.

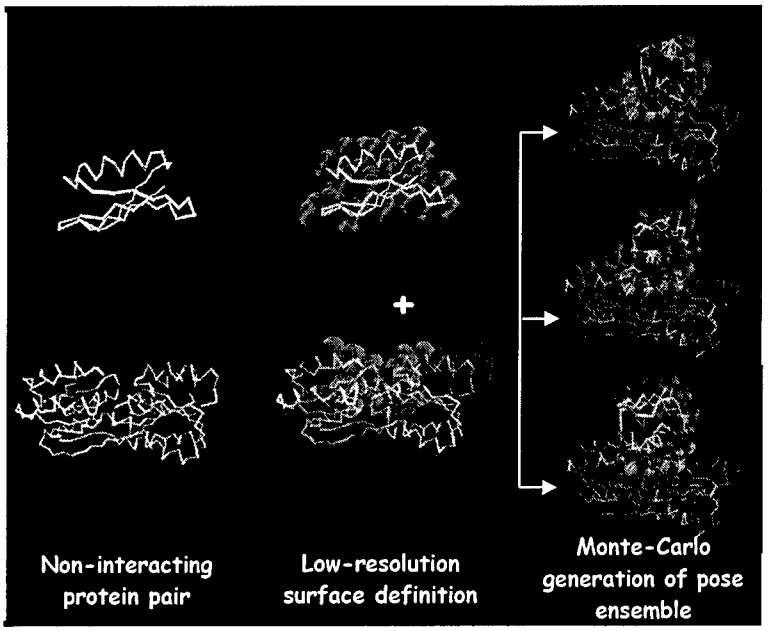

Figure 2. *Initial generation of designed protein-protein interaction poses for design, using a low-resolution representation of the protein surface. The pseudo-surface is sequence independent and is constructed out of a series of impenetrable, "sticky" spheres (short-range attractive well). A Monte Carlo search can be used to generate an ensemble by optimizing an interaction potential between the sticky spheres. Results of such a search are shown for B1 domain (yellow) and MBP (silver). On MBP docking is restricted to the interdomain region. The entire B1 domain surface is potentially available for binding.*

For the protein-DNA interactions, we limited ourselves to the redesign of known protein-DNA interactions, to redesign the sequence specificity of the binding interactions. Here the main challenge is to establish sequences that read out the hydrogen-bonding patterns in the major groove of the DNA helix. Although code was developed that meet both goals, the resulting designs could not be tested experimentally within the time available for the project period.

CONCLUSIONS

The computational design approaches have sufficient predictive accuracy and scope to drastically alter the ligand-binding properties of receptors. This suggests that the essential elements of biomolecular recognition have been successfully captured in molecular design calculations.

SIGNIFICANCE

The developed protein design software is very general. In principle it could therefore be possible to design receptors for a wide variety of ligands, allowing new sensors to be developed rapidly. Furthermore, theoretically the same approach should be able to design enzyme activity, if appropriate models of the transition state are used to represent the desired reaction coordinates.

PUBLICATIONS

L.L. Looger, M.A. Dwyer, J.J. Smith, H.W. Hellinga (2003) Computational design of receptor and sensor proteins with novel functions. *Nature*, **423**:185-190.

M.A. Dwyer, L.L. Looger, H.W. Hellinga (2003) Computational design of a Zn^{2+} receptor that controls bacterial gene expression. *Proc. Natl. Acad. Sci. USA*, **100**:11255-11260.

M. Allert, S.S. Rizk, L.L. Looger, H.W. Hellinga (2004) Computational design of receptors for an organophosphate surrogate of the nerve agent Soman. *Proc. Natl. Acad. Sci. USA*, **101**:7907-7912.

M.A. Dwyer, H.W. Hellinga (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr. Op. Struct. Biol.*, **14**:495-504.

PATENTS

A patent has been filed on the computational design methodology.

TRANSITIONS

This project has been transitioned into a DARPA-sponsored project on the computational design of enzymes, an NIH grant on the design of protein function, and an NIH Director;s Pioneer Award.