REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 14, 1995	

4. TITLE AND SUBTITLE
Novel Approaches to the Characterization of Specific Protein-Protein Interactions Important in Gene Expression

5. FUNDING NUMBERS

6. AUTHOR(S)
Ronald L. Somerville

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Purdue Research Foundation
Division of Sponsored Programs
Hovde Hall
West Lafayette, IN 47907

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)* Through the application of a range of biochemical, immunochemical and genetic techniques, we have accomplished the following goals: First, we characterized a previously undescribed protein, WrbA and showed that it functions as an accessory protein in the binding of TrpR to its operator target; second, we demonstrated that WrbA is a stationary-phase protein, and that the regulation of promoters controlled by TrpR becomes more stringent in stationary phase cells; third, we developed a method for the purification of the TyrR protein, a dual-specificity transcription factor that can both activate and inhibit the action of RNA polymerase. The domain structure of TyrR was established by chemical procedures.

14. SUBJECT TERMS
Helix-turn-helix motif, transcriptional regulation, TrpR protein, TyrR protein, Escherichia coli.

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Novel Approaches to the Characterization of Specific Protein-Protein Interactions Important in Gene Expression

Final Technical Report

Ronald L. Somerville, Ph.D.

October 14, 1995

U.S. Army Research Office

Grant No. DAAL 03-91-G-0302

Purdue University
West Lafayette, IN

A. STATEMENT OF PROBLEM

The regulation of transcription initiation at promoters has until recently been considered to involve the binding to operator targets in DNA of a variety of different *trans*-acting protein factors, each having distinct operator-binding specificity. Depending on the system, the outcome of such protein-DNA interaction was either an enhancement or a reduction in the rate of formation of a particular mRNA species. This relatively crude view of transcriptional regulation is now undergoing modification in light of recent findings that many promoters contain upstream target sequences for multiple *trans*-acting factors. Regulatory proteins that modulate gene expression in these cases do so by touching one another and/or some part of the transcriptional machinery. Sometimes these effects involve mechanisms that include topological changes in DNA or that may require the participation of additional proteins not capable of interacting directly with DNA. The biological advantage of multifactorial promoters comes from the resultant ability of the organism to control gene expression in sophisticated ways in response to subtle environment or developmental cues.

The object of this proposal was to perfect and apply methods for assaying and purifying proteins whose principal distinguishing characteristic is their ability to interact with known DNA-binding regulatory proteins.

B. SUMMARY OF MOST IMPORTANT RESULTS

1. Studies of the WrbA protein, an accessory factor in specific protein-DNA interaction

Highly purified preparations of Trp repressor (TrpR) protein derived from *Escherichia coli* strains that were engineered to overexpress this material were found to contain another protein, of 21 kDa. The structural gene for the second protein, designated WrbA [for tryptophan (W) repressor-binding protein], was cloned and sequenced. A method for the overproduction and purification of the WrbA protein was developed. The WrbA protein was found by several criteria to enhance the formation and/or stability of noncovalent complexes between TrpR holorepressor and its primary operator target. The WrbA protein alone does not interact with the *trp* operator. The *wrbA* gene was found to be preferentially expressed in cells that had exited the logarithmic phase of growth. During the stationary phase, cells deficient in the WrbA protein were inferior to wild type in their ability to repress the *trp* promoter. It appears that one role for the WrbA protein is to enhance the blockage of TrpR-specific transcriptional processes, thereby imparting a currently undefined physiological advantage during the nongrowing phase of the life cycle of the organism.

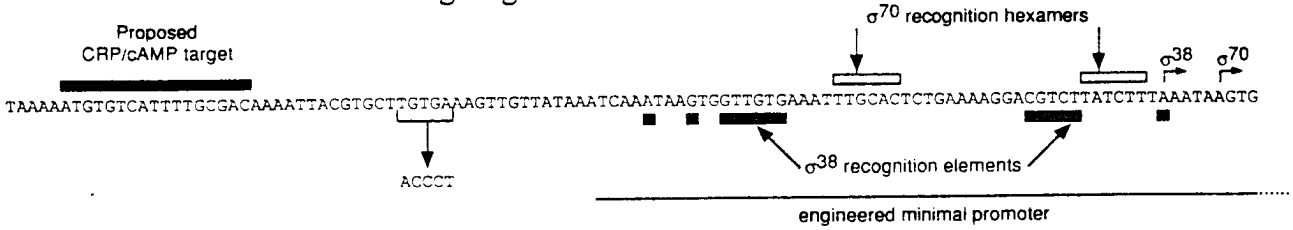
2. Structure-Function Analysis of the wrbA Promoter

There is relatively little information about what structural features enable a segment of DNA to selectively act as a site for transcription initiation in cells that have ceased to divide. This issue has been investigated for the *wrbA* promoter in several ways. By chemically determining the startpoint for *wrbA* messenger RNA at different stages of growth it was shown that the *wrbA* promoter is a compound punctuation element. Within

a short (60 bp) segment of DNA lie targets for both the standard (σ^{70}) form of RNA polymerase and the stationary phase-specific (σ^{38}) form of the enzyme. The 5' ends of the two principal *wrbA* messenger RNA species are separated by 5 nucleotides, so it is reasonable to assume that the different forms of RNA polymerase approach the *wrbA* promoter from opposite faces of the DNA helix.

In a systematic survey of the effects of various factors known to affect promoter utilization, it was discovered that transcription from the *wrbA* promoter is responsive to the availability of cyclic AMP. The frequency of transcription initiation from the *wrbA* promoter drops by about 50% in cells that are unable, by virtue of a lesion in the *cya* gene, to synthesize cyclic AMP. To pinpoint the target site for cyclic AMP-specific transcriptional activation, a series of biochemical and genetic experiments were carried out. By DNase I footprinting using purified cyclic AMP binding protein it was found that the regulatory site for the cyclic AMP system is centered about a 17 bp DNA segment that lies 91.5 base pairs upstream of the transcriptional startpoint characteristic of stationary phase cells, and 96.5 base pairs upstream of the startpoint that is specific for the σ^{70} form of RNA polymerase. When the DNA segment containing the cyclic AMP target was disconnected from the remainder of the *wrbA* promoter, the cyclic AMP activation effect was abolished.

Our current understanding of the structural features of the *wrbA* promoter are summarized in the following diagram.



3. Isolation and Characterization of Novel trpR Mutants

In an attempt to enlarge our repertoire of tools for understanding how the WrbA protein enhances the affinity of the TrpR holorepressor for its operator target, a number of mutationally altered forms of TrpR, generated by conventional or site-directed mutagenesis, were characterized. During the next grant period, we plan to conduct experiments *in vitro* and *in vivo* to see whether any of the disabling amino acid switches in mutant forms of TrpR are WrbA-dependent, or can be made to become WrbA-dependent through further structural changes.

Operator recognition by the dimeric TrpR protein of *Escherichia coli* is determined by a polypeptide segment of 24 amino acids that is folded into the familiar helix-turn-helix motif. When the primary amino acid sequences of many such motifs are aligned, several positions are seen to be strongly conserved. In TrpR, one highly conserved position it residue 71, occupied by Leu. Using a novel system of oligonucleotide-directed mutagenesis, a complete set of amino acid replacements at residue 71 was generated. The operator-binding activity of each variant form of TrpR was evaluated *in vivo* using a *trp* promoter-*lacZ* operator system. Those TrpR proteins

containing Met or a branched chain amino acid at position 71 were fully active. TrpR proteins having Thr or Cys at position 71 were partially active. All other amino acid replacements led to nonfunctional versions of TrpR. Several amino acids in the latter group (i.e. Ala, Ser, Phe, Tyr and Lys) are found at analogous positions within other fully active helix-turn-helix proteins. Computer modeling studies of the helix-turn-helix segment of several forms of TrpR suggest that residue 71 is part of a deformable, partially hydrated "soft core," that plays a key role in operator recognition by dictating the spatial orientation of a constellation of solvent-exposed residues that actually lie along the protein-DNA interface. The hydrophobic side chain of Leu71 lies in close proximity to Ile57 and Leu75, which are also considered to be part of the "soft core." At both of these locations, disabling amino acid switches were identified within a collection of UV-induced chromosomal *trpR* alleles. In comparative studies of the expression levels of active forms of TrpR, it was found that the inactive species were invariable expressed with far greater efficiency than their active counterparts. This observation suggests that TrpR may translationally repress its own synthesis, and that helix-turn-helix motif is essential to this form of regulation.

All organisms have repeated sequences present in their genomes. REP elements are short (>200 bp), intercistronic, noncoding segments of DNA that are widely distributed amongst prokaryotes (Gilson *et al.*, 1984, *EMBO J.* **3**, 1417-1421). REP elements may constitute upwards of 1% of the *E. coli* chromosome. They are thought to function in chromosome organization. One such REP element lies immediately downstream of the *trpR* gene (coordinate 4665.5) of *E. coli*. Among a large collection of chromosomally-encoded *trpR* mutants that were structurally characterized by sequencing PCR-amplified DNA, there were three independent occurrences of changes within the UGA termination codon. Each change was predicted to extend the *trpR* reading frame by 50 amino acid residues, via the transcription and translation of the downstream REP element. The validity of this prediction was verified by means of *in vitro* translation experiments. MRNA encoding a TrpR:REP chimaera was shown to program the synthesis of a polypeptide of the expected size and composition. These studies show that REP elements can be determinants of protein structure in addition to their currently accepted role in dictating the architecture and topology of the genome.

4. Structure-Function Analysis of the TyrR Protein

The TyrR protein of *Escherichia coli*, in response to the binding of tyrosine or phenylalanine, activates certain promoters and represses others. The protein products of the regulated genes catalyze steps in aromatic amino acid biosynthesis or transport. It was shown that promoter activation was selectively abolished in mutant proteins lacking amino acid residues 2-9. An additional series of constructs, encoding mutant TyrR proteins having deletions or point mutations near the N-terminus, were analyzed. Residues Arg2 and Leu3 were shown to be critical for the activation of the *mtr* promoter. In confirmation of previous findings, none of the activation-defective mutant TyrR proteins had lost significant repression function. The TyrR protein was shown by chemical crosslinking to be dimeric. The polypeptide segments critical for dimer formation in vivo were identified by evaluating the negative dominance phenotypes of a series of mutant

proteins, all defective in DNA binding, lacking progressively greater numbers of amino acid residues from either the N-terminus or the C-terminus. Amino acid residues 194-438 were found to contain all the essential dimerization determinants.

The TyrR protein contains 513 amino acids per subunit. Methods were developed for efficiently purifying the TyrR protein to apparent homogeneity. We analyzed the pattern of cleavage of the TyrR protein by trypsin, either in the absence of ligands, or in the presence of saturating levels of L-tyrosine, ATP, or poly(dI-dC). At low (1:200 ratio by weight) trypsin levels, in the absence of ligands, two major digestion products accumulated. These were polypeptides of 22 kD and 31 kD, shown to contain amino acid residues 1-190 and 191-467, respectively. The pattern of trypsin cleavage was unaffected by tyrosine. In the presence of ATP, an intermediate species of 53 kD, probably containing amino acid residues 1-467, was observed. The kinetics of appearance of the 53 kD species were consistent with a role for ATP in accelerating the hydrolysis of the R467-F468 peptide bond. The 53 kD polypeptide underwent further tryptic hydrolysis to yield fragments of 22 kD and 31 kD. When both tyrosine and ATP were present, the rate of formation of the 22 kD and 31 kD fragments was more rapid than in the absence of these ligands. It appears that when both ligands were bound, the rates of hydrolysis of peptide bonds R190-Q191 and R467-F468 are both enhanced. Additional limited proteolysis experiments suggested that polypeptide segment 191-467 contains ATP binding site(s), and that the rate of cleavage of peptide bonds R190-Q191 and R467-F468 s altered when the TyrR protein interacts with poly(dI-dC), an analog of target DNA. Our results reveal the presence of two major structural domains within the TyrR protein. The first domain (amino acid residues 1-190) is extremely resistant to hydrolysis by trypsin. The second domain (residues 191-467), likely to contain ATP binding site(s) is homologous to several other transcriptional activators specific for promoters responsive to the σ^{54} form of RNA polymerase. The remainder of the TyrR protein (residues 468-513) contains the operator recognition elements, probably arranged in the form of a helix-turn-helix motif. This polypeptide segment was not detected as a discrete tryptic hydrolysis product.

C. PUBLICATIONS AND TECHNICAL REPORTS

Somerville, R. L., Shieh, T.-L. N., Hagewood, B. and Cui, J. Gene Expression from Multicopy T7 Promoter Vectors Proceeds at Single Copy Rates in the Absence of T7 RNA Polymerase. Biochem. Biophys. Res. Commun. 181, 1056 (1991).

Cui, J. and Somerville, R. L. Mutational Uncoupling of the Transcriptional Activation Function of the TyrR Protein of Escherichia coli K-12 from the Repression Function. J. Bacteriol. 175, 303 (1993).

Cui, J. and Somerville, R. L. A Mutational Analysis of the Structural Basis for Transcriptional Activation and Monomer-Monomer Interaction in the TyrR System of Escherichia coli K-12. J. Bacteriol. 175, 1777 (1993).

Yang, W., Ni, L. and Somerville, R. L. A Stationary Phase Protein of Escherichia coli that Affects the Mode of Association between the Trp Repressor Protein and Operator-Bearing DNA. Proc. Natl. Acad. Sci., 90, 5796-5800 (1993).

Cui, J., Ni, L. and Somerville, R.L. ATPase Activity of TyrR, a transcriptional regulatory protein for $\ddot{U}^{70}$ RNA polymerase. J. Biol. Chem. **268**, 13023-13025 (1993).

Skrypka, I. and Somerville, R.L. Nucleotide Sequence of the *Salmonella typhimurium trpR* Gene. *DNA Sequence*, 4, 355-360 (1994).

D. PARTICIPATING SCIENTIFIC PERSONNEL

1. Ihor Skrypka (Graduate Research Assistant, 1992-present)
2. Weiping Yang (Graduate Research Assistant, 1991-1993)
3. Jisong Cui (Graduate Research Assistant, 1992)
4. Liuying Ni (Professional Research Associate, 1991-present)
5. Hong Smith (Professional Research Associate, 1994-present)