PROTEIN DESIGN AND SIMULATION

Part I : Protein Design

Part II: Protein Simulation

Thesis by

Changmoon Park

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1993

(Submitted May 17, 1993)

To my parent

Acknowledgement

I would like to thank my advisor, Bill Goddard, for providing the research opportunities which made this thesis possible. His insight, enthusiasm, and guidance will always be appreciated.

I would also like to thank Judy Campbell and Mel Simon for their advice and support during the experiments. I also thank all those members of Goddard Group, Campbell Group, Simon Group, and Biopolymer Synthesis Center of Caltech who have contributed to my research in both scientific and personal matters.

I am grateful to Kyoung-Joon Oh and Hogyu Han for their friendship and valuable discussions about the experiments.

Finally, my special gratitude goes to my wife, Euisook, and my children, Shul and Sun, for their love, encouragement, and patience over the past years.

Abstract

Since specific DNA binding proteins play many important roles in the regulation of cellular reactions including replication, transcription, and translation by the specific interactions of DNA binding proteins with DNA, the design and synthesis of sequence-specific DNA binding proteins is of great interest in modern chemistry and biology.

Chapter 1 introduces a strategy by which to design new protein structures recognizing new sequences of DNA. The results of experiments using new protein show that there is cooperation between the monomers in binding to DNA and each monomer recognizes the half-site of the dimer binding site.

Chapter 2 describes the advantage of palindromic sites and dimerization in DNA recognition according to the experimental results. The results also show that each monomer in a dimer recognizes one half-site of the dimer binding site irrespective of the relative orientation monomer in the dimer and the dimer binding site depends on the relative orientation of the two monomer in the dimer.

Chapter 3 shows that the monomer of DNA binding region of the v-Jun leucine zipper protein recognizes the dimer binding site. Our results support the possibility that two monomers of v-Jun might bind sequentially to the dimer site with dimerization of v-Jun occurring while bound.

Chapter 4 describes the design of a new protein recognizing a new 16 bp site in DNA. Our results show that there is cooperation between monomers and all three monomers in the new peptide recognize unique half-sites in the proposed trimeric binding site.

Chapter 5 describes quantum mechanical calculations on the active site of cytochrome P-450cam to improve force constants for the molecular simulations of cytochrome P-450cam. Our results show that the size of iron ion is a function of its spin and oxidation state and plays a key role in the process of oxygenation.

Table of Contents

Part I

Protein Design

Chapter 1

Protein Stitchery: Design of a Protein for Selective Binding to a Specific DNA Sequence

Changmoon Park*†‡, Judy L. Campbell†‡, and William A. Goddard III*†

*Materials and Molecular Simulation Center, Beckman Institute 139-74

†Division of Chemistry and Chemical Engineering (Contribution No.8547),

and ‡Division of Biology

California Institute of Technology, Pasadena, CA 91125

Abstract

We present a general strategy for designing proteins to recognize DNA sequences, and illustrate this with an example based on the "Y-shaped scissors grip" model for leucine-zipper gene regulatory proteins. The designed protein is formed from two copies, in tandem, of the basic (DNA binding) region of v-Jun. These are coupled through a tripeptide to yield a "dimer" expected to recognize the sequence TCATCGATGA (the v-Jun homodimer recognizes ATGACTCAT). We synthesized the protein and oligonucleotides containing the proposed binding sites and used gel retardation assays and DNase I footprinting to establish that the dimer binds specifically to the new DNA sequence but not the wild-type DNA sequences, nor to oligonucleotides in which the recognition half-site is modified by single-base changes.

These results also provide strong support for the "Y-shaped scissors grip" model for binding of leucine-zipper proteins.

Introduction

We propose a general strategy for designing new proteins to recognize specific DNA binding sites, it is to select segments of proteins, each of which recognizes particular DNA segments, and to stitch these together via a short peptide with a cysteine crosslink in a way compatible with each peptide being able to bind to its own DNA segment. This leads to a new protein that recognizes the composite site.

As a starting point we consider the gene regulatory leucine-zipper proteins. They are characterized by two structural motifs (1-3):
(i) the leucine zipper which is responsible for dimerization and
(ii) the basic region for DNA binding (4-7). The basic regions of unbound leucine-zipper dimers are unfolded, but fold into the alpha-helix conformation upon binding to the specific site (8-10). The most plausible model for the conformation of leucine-zipper protein is the "Y-shaped scissors grip" model (1,2), in which the basic region of each monomer interacts with DNA on either side of the dyad axis of the binding site. Thus, for GCN4 each arm recognizes the half-site AGTA (11,12).

The Design

Our design strategy assumes this "Y-shaped scissors grip" model (Figure 1a). We form new proteins by crosslinking (stitching together) various binding arms so as to be consistent with the orientation of the recognition helix in each half-site. Here we build upon the results of

Kim and coworkers (5,6), who showed that the leucine zipper of GCN4 can be replaced with linkers (Gly-Gly-Cys) at the C-terminus of the DNA binding segment which upon oxidation dimerize and bind to the same site (ATGACTCAT) as GCN4. As a model system to explore the design of new DNA binding proteins, we have chosen the v-Jun leucine zipper dimer (Figure 1a) which also binds to the site ATGACTCAT as a homodimer with itself or as a heterodimer with Fos (4,13-16), another member of this DNA binding protein family. We will reverse the sequence relationship of the alpha-helix to the target nucleotide of the binding arms by adding the Gly-Gly-Cys linker to the N-terminus (rather than C). As illustrated in Figure 1b the new protein is designed to recognize the DNA sequence TCAT**X**ATGA, where **X** represents 0 to 2 additional bases to accommodate the loop region of the new dimer.

Several criteria were used in selecting v-Jun as the starting point:

(i) To prevent nonspecific disulfide bond formation, the protein must not contain cysteine in its basic region.

(ii) Since we want to reverse the alpha-helix relative to the target DNA sequence, the protein should have no residues (especially proline and probably glycine) that would interrupt alpha-helices.

(iii) Since we want to ensure that the protein can form the alpha-helix when joined with the linker, the composition of amino acids in its basic region should strongly favor a-helices [by the Chou-Fasman criterion (17)].

We considered 14 leucine-zipper proteins and found that v-Jun best satisfies the above criteria.

We took as our standard protein the 31 residues at the N-terminus of v-Jun joined with the linker (Gly-Gly-Cys) (Figure 2a). The new protein (v-Jun-NN) is designed to bind specifically to the site TCAT**X**ATGA, where **X** might contain 0 to 2 base pairs (bp). As indicated in Figure 2b, we considered four cases for **X**:

(i) **X** = F (no base pairs), denoted as NNS-F

(ii) **X** = C (which is equivalent to X = G), denoted as NNS-C/G

(iii) **X** = CG, denoted as NNS-CG

(iv) **X** = GC, denoted as NNS-GC

We excluded using A or T on the assumption that the methyl group of T (which sites in the major groove) might interfere with binding of the protein.

Materials and Experiments

Preparation of Peptide and Oligonucleotides

Peptide v-Jun-N was chemically synthesized at The Biopolymer Synthesis Center of Caltech. The automated stepwise solid phase syntheses were performed on an ABI 430A peptide synthesizer employing an optimized synthetic protocol of the f-Boc chemistry. The peptide was purified by reverse phase HPLC on a Vydac C18 column. A linear gradient of 0-50% aqueous/acetonitrile/0.1% TFA was run over 120 minutes. The purity of peptide was checked by mass spectroscopy: calculated, 4039.3; experimental, 4041.8. All

oligonucleotides used were chemically synthesized at The Biopolymer Synthesis Center of Caltech.

Gel retardation assays

The binding solution contains bovine serum albumin at 50 mg/ml, 10% (vol/vol) glycerol, 20 mM Tris.HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$ and 1.56 nM v-Jun-NN in 10 µl reaction volume. After adding 5,000 cpm of each 5'-^{32}P-labeled probe DNA (25-26 bp), the solutions were stored at 4° C for 1 hr and loaded directly on an 8% nondenaturing polyacrylamide gel in TE buffer at 4° C.

DNase I footprinting assays

The footprinting assay solution (in 50 µl) contains bovine serum albumin at 50 mg/ml, 5% glycerol, 20 mM Tris-HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 20,000 cpm of each 5'-^{32}P-labeled probe DNA (60-62 bp) and 50 nM v-Jun-NN. This solution was stored at 4° C for 1 hr. After adding 5 µl of DNase I diluted in 1x footprinting assay buffer, the solutions were stored 1 min more at 4° C. the DNase I digestion was stopped by addition 100 µl of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, 25 µg/ml sonicated salmon sperm DNA and 25 µg/ml of yeast tRNA. This was phenol/chloroform extracted, ethanol precipitated, and washed with 70% ethanol. The pallet was resuspended in 5 µl of formamide loading buffer, denatured at 90° C for 4 min, and analyzed on 10% polyacrylamide sequencing gel (50% urea).

Results

We carried out gel retardation assays using four DNA sequences

(α) the new sequence TCATCGATGA (case c above), NNS-CG

(β) the binding sequence for v-Jun, ATGACTCAT,

(γ) the complementary double base pair substitution (C2 -> A2 and G9 -> T9) of α: TAATCGATTA.

(δ) the complementary double base pair substitution (A3 -> C3 and T8 -> G8) of α: TCCTCGAGGA.

The results (Figure 3) indicate that v-Jun-NN binds to the new DNA sequence as a homodimer with a dissociation constant of less than 1 nanomolar at 4^{0} C. On the other hand, the new protein does not bind significantly to the wild-type site β of the mutant sites γ and δ.

In order to establish the specific binding site for v-Jun-NN, we carried out deoxyribonuclease (DNase) I footprinting. These results (see Figure 4) show that the v-Jun-NN protects the exact binding site predicted for the new protein. Thus we conclude that each arm of DNA-bound v-Jun-NN retains the same structure as in native v-Jun. The DNase I footprinting results (Figure 4) also indicate that NNS-CG has the strongest binding affinity for v-Jun-NN. The decrease in binding for shorter X may result from the strain required at the loop region of the dimer to place the binding segments along the binding region. Particularly interesting is the difference in specificity observed between NNS-GC and NNS-CG, (Figure 4, compares lanes 9 and 12 for

top strands and lanes 21 and 24 for bottom strands). These results indicate that X plays more than the role of spacer.

Discussion

These results support the idea that the N-terminal region of v-Jun contributes to the binding to DNA through specific interaction with the DNA (since in v-Jun-NN this region is forced to contact the DNA). This supports the angulated bend conformation (1). Our results help differentiate the respective roles of the basic region and of the leucine zipper region in recognition and binding. The basic region of v-Jun is sufficient for specific binding. Although the leucine zipper region is not directly involved in DNA binding, our results indicate that its position relative to the basic region plays an important role in determining which target sequence of DNA the protein recognizes.

Summarizing, we have designed a new protein (stitched together from segments derived from the natural protein) to recognize a specific new DNA binding site, and we have established specific binding of the new protein this site. Note that use of the--Cys linker is not essential in the design. We could just as well replace the cysteine and make a continuous ~70 amino acid protein that should recognize a predictable site (14). In addition this strategy is not limited to two arms. We could have stitched together three, four, or more arms with appropriate linkers to design proteins that would recognize DNA sequences with 15, 20 or 25 base pairs. Such systems with EDTA-Fe (18) or other nucleases would presumably cut very

specific sites, allowing the genome to be cut into much longer segments. The design is not limited to v-Jun. Any protein or other molecule that recognizes a specific DNA sequence by binding along the major groove could be a candidate. Many such cases are now known so that we already have a collection of available partial bindings sites that could be combined to form composite target binding sites for designing new binding proteins. Of course the segments of the new proteins should be designed so that the intramolecular interactions are not so strong as to compete with binding to the DNA.

Our results support the idea that each monomer arm of the dimer binds along the major groove to the half of the binding site of the dimer (11,12). This strongly supports the "Y-shaped scissors grip" model for leucine zipper proteins (1). Our new strategy can be used to investigate the interaction between DNA and protein and the structure of DNA/protein complex.

References

1. C. R. Vinson, P. B. Sigler, S. L. McKnight, *Science* **246**, 911 (1989);
2. W. T. Pu, K. Struhl, *Proc. Natl. Acad. Sci. USA* **88**, 6901 (1991).
3. W. H. Landschultz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
4. M. Neuberg, M. Schuermann, J. B. Hunter, R. Muller, *Nature* **338**, 589 (1989).
5. R. V. Tlanian, C. J. McKnight, P. S. Kim, *Science* **249**, 769 (1990);
6. E. K. O'Shea, R. Rutkowski, P. S. Kim, *Science* **243,** 538C (1989).
7. P. Agre, P. F. Johnson, S. L. McNight, *Science* **246,** 922 (1989).
8. M. A. Weiss, *Biochemistry* **29**, 8020 (1990)
9. M. A. Weiss, T. Ellenberger, C. r. Wobbe, J. P. Lee, S. C. Harrison, K. Struhl, *Nature* **347**, 575 (1990)
10. V. Saudek, H. S. Pasley, T. Gibson, H. Gausepohl, R. Frank, A. Pastore, *Biochemistry* **30**, 1310 (1991).
11. M. G. Oakley, P. B. Dervan, Science 248, 847 (1990);
12. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *Science* **249,** 774 (1990).
13. K. Struhl, *Cell* **50**, 841 (1987);
14. T-J. Bos, F. J. Ramscher III, T. Curran, P. K. Vogt, *Oncogene* **4**, 123 C (1989)

15. C. Abate, D. Luk, R. Gentz, F. J. Ramscsher, III, T. Curran, *Proc. Natl. Acad. Sci. USA* **87**, 1032 (1990)

16. R. Turner, R. Tijan, *Science*, **243**, 1689 (1989)

17. P. Y. Chou and G. D. Fasman, *J. Mol. Biol.* **74**, 263 (1973).

18. D. P. Mack, B. L. Iverson, P. B. Dervan, *J. Am. Chem. Soc.* **110**, 7572 (1988).

19. Y. Nakabeppu, D. Nathans, *EMBO J.* **8**, 3833 (1989).

Figure 1. (a) The "Y-shaped scissors grip" model for the v-Jun homodimer bound to the ATGACTCAT site. (b) The designed protein v-Jun-NN. After removing the leucine zipper region (blue and light-blue) of each monomer, the upper arm (green) and its DNA binding site (pink, ATGA) was shifted just below the lower arm (orange). In Figure 1b the shifted upper arm and DNA fragment retain their original green and pink colors, respectively. Different linkers (Gly-Gly-Cys, purple) were added at the N-termini of both arms, and a disulfide bond was made. v-Jun-N was dimerized by oxidized dithiothreitol and purified by HPLC. Protein concentration was determined by the method of Bradford with Bio-Rad protein assay kit. Thus the designed protein is expected to bind to TCAT**X**ATGA (where cyan joins to pink in 1b) and does not to ATGACTCAT (where pink joins to cyan in 1a), where **X** (filling bases) fit the new loop introduced between the peptide monomers to make the dimer.

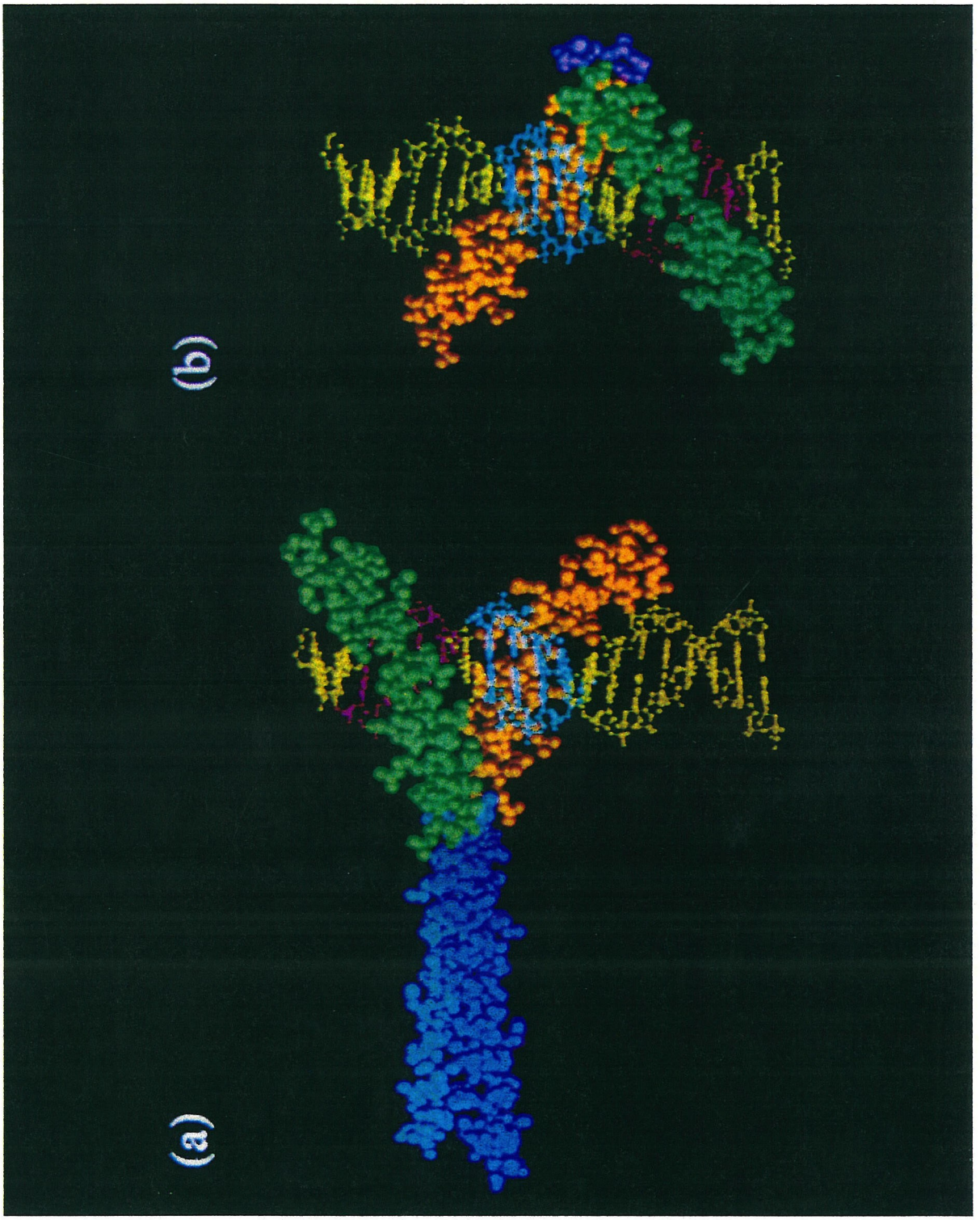

(b)
(a)

Figure 2. Sequences of protein and oligonucleotides used in gel retardation and footprinting. v-Jun-br contains the basic region of v-Jun corresponding to the residues from 218 to 346, and Gly-Gly-Cys is added to make v-Jun-N. Single-letter amino acid code is used.

(a) Protein

vJun-br	:		S	QERIKAERKR	MRNRIAASKS	RKRKLERIAR
vJun-N	:	CGG	S	QERIKAERKR	MRNRIAASKS	RKRKLERIAR

(b) DNA

NNS-Φ : ctcagatccggatccatggttagaccTCATATGA cggtactggtcgagaattcggatcct

NNS-C/G : ctcagatccggatccatggttagaccTCAT(C/G)ATGAcggtactggtcgagaattcggatcct

NNS-CG : ctcagatccggatccatggttagaccTCATCGATGA cggtactggtcgagaattcggatcct

NNS-GC : ctcagatccggatccatggttagaccTCATGCATGA cggtactggtcgagaattcggatcct

α : gttagaccTCATCGATGAcggtactg

β : gttagaccATGACTCAT cggtactg

γ : gttagaccTAATCGATTAcggtactg

δ : gttagaccTCCTCGAGGAcggtactg

Figure 3. Gel retardation assay for the binding of v-Jun-NN to various DNA segments. The dimer selectively binds to the predicted site, NNS-CG. Odd-numbered lanes, no protein; even-numbered lanes, v-Jun-NN. Lanes 1 and 2, probe DNA a (NNS-CG); lanes 3 and 4, probe DNA b (the wild type site); lanes 5 and 6 probe DNA g, lanes 7 and 8, probe DNA d (Figure 2b). As determined by titration of the gel shift, v-Jun-NN binds to the predicted new sequence, NNS-CG with a dissociation constant (K_d) of ~ 0.3 nanomolar at 4° C.

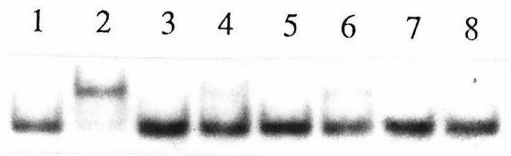
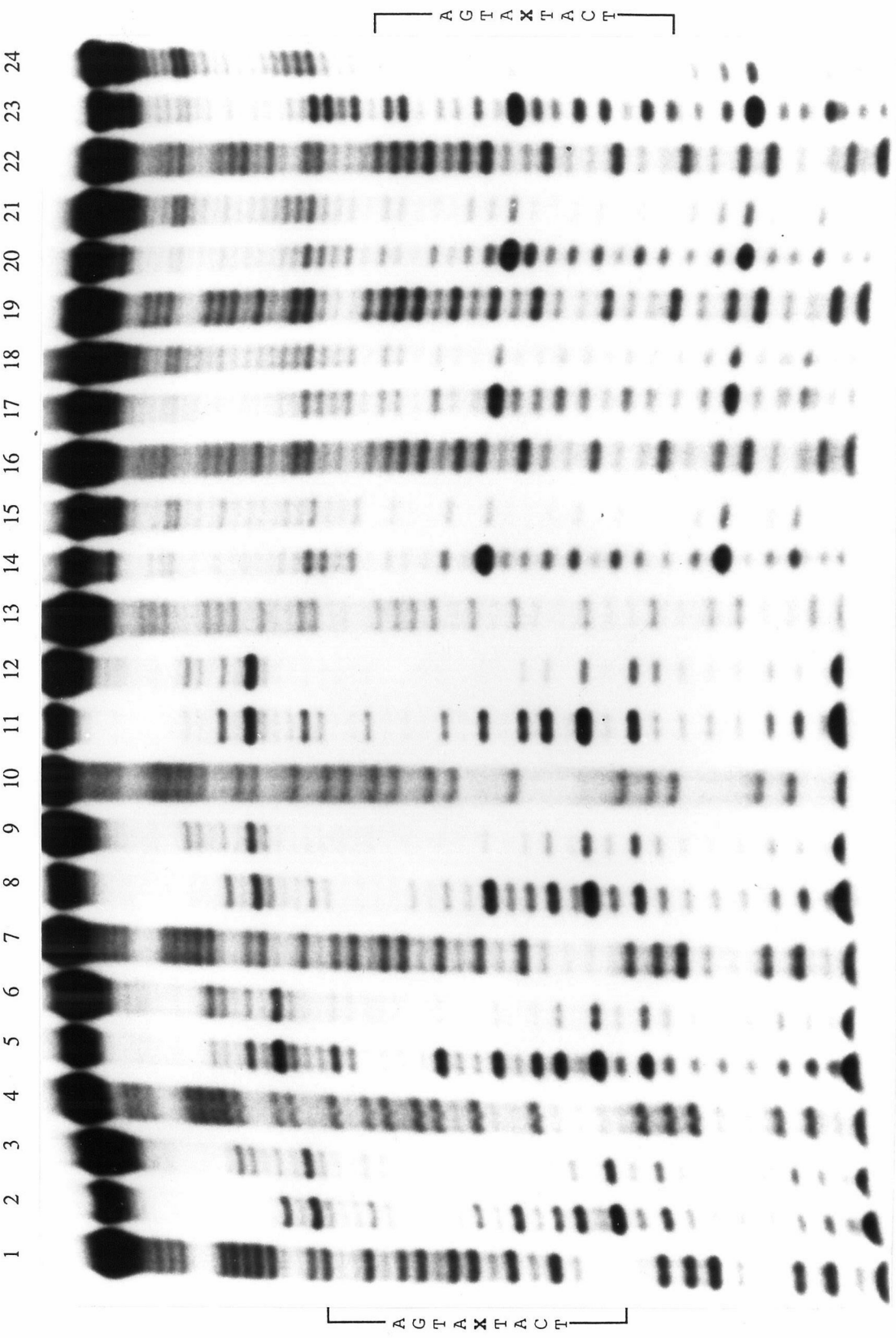

1 2 3 4 5 6 7 8

Figure 4. DNase I footprinting assay of v-Jun-NN with DNA containing the predicted binding sites, NNSs. This shows that the protein protects the target binding site with the best protein for NNS-CG. lanes 1-12, labeled at 5' end of top strand; lanes 13-24, labeled at the 5' end of bottom strand. Lanes 1-3 and 13-15, NNS-F; lanes 4-6 and 16-18, NNS-C/G; lanes 7-9 and 19-21, NNS-GC; lanes 10-12 and 22-24, NNS-CG. The first lane for each probe DNA (lanes 1,47,10,13,16,19, and 22), G+A marker; the middle lane for each probe DNA (lanes 2,5,8,11,14,17,20, and 23), no protein; the last lane for each probe DNA (lanes 3,6,9,12,15,18,21, and 24), v-Jun-NN.


A G T A X T A C T
A G T A X T A C T
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Chapter 2

Design Superiority of Palindromic DNA Sites for Site-Specific Recognition of Proteins: Tests Using Protein Stitchery

Changmoon Park*†‡, Judy L. Campbell†‡, and William A. Goddard III*†

*Materials and Molecular Simulation Center, Beckman Institute 139-74
†Division of Chemistry and Chemical Engineering (Contribution No.8776),
and ‡Division of Biology
California Institute of Technology, Pasadena, CA 91125

Abstract

Using Protein Stitchery with appropriate attachment of cysteines linking to either carboxy (C) and amino (N) ends of the basic region of the v-Jun leucine-zipper gene-regulatory protein, we constructed three dimers, pCC, pCN, and pNN. All three bind specifically to the appropriately rearranged DNA recognition sites for v-Jun: ATGAcgTCAT, ATGAcgATGA, and TCATcgATGA, respectively (dissociation constants of about 4 nM at 4° C). Results of DNase I footprinting provide strong support for bent recognition helices in leucine-zipper protein/DNA complexes. Comparison of the results for pCC and pNN with those for pCN shows the design superiority of palindromic sequences for protein recognition.

Introduction

The mechanism by which cells respond to external stimuli is a fundamental problem in modern biology. Transcriptional regulatory proteins are known to play a key role in several systems evolved by cells to convert extracellular signals into altered gene expression (1). They operate by specifically binding to DNA target sites which regulate the transcription of particular genes. Prominent among transcriptional regulatory proteins is the leucine-zipper family, which recognize the DNA binding site either homodimers or heterodimers (2-4). The leucine-zipper proteins are characterized by two functional segments:

i. the leucine zipper region, a helical region containing 4 or 5 leucines spaced at seven amino acid intervals, and

ii. the basic region containing many basic residues (5-10).

The basic region appears to be unfolded in solution, but assumes an alpha-helical structure binding to its recognition site (11-13). Site-directed mutagenesis (6,7) and domain swapping (8-10) experiments show that the leucine zipper region mediates dimerization and that the basic region is responsible for DNA binding. Experiments replacing the leucine zipper region with a three peptide linker (14,15) showed that: the dimerized basic region recognizes the same site as the native protein supporting the scissors grip model (5), where each monomer recognizes the half-site of the symmetrical DNA binding site. Recently, we showed that the normal dimer (denoted pCC), which selectively recognizes the sequence ATGAcgTCAT, can be

inverted to form a protein (denoted pNN) which selectively recognizes the inverted site, TCATcgATGA (15).

Gel electrophoresis experiments (22) with Jun homodimer and with Jun-Fos heterodimer showed that Jun and Fos induce DNA bending in the opposite direction upon binding to the specific site. To explain this it was proposed that the basic region of Jun has a bent α-helix while the basic region of Fos has a *straight* helix. However, a recent X-ray crystal structure (21) for the complex between GCN4 and DNA containing the GRE site (ATGACTCAT) showed a straight single α-helix for the basic region of GCN4. Our current results support the interpretation that the v-Jun homodimer bound to its specific site has bent α-helices.

Peptide design

Using Protein Stitchery, we have now made three kinds of v-Jun (16,17) homodimers (denoted pCC, pNN, and pCN) and show herein that each selectively recognizes the appropriately reorganized DNA binding sites ATGAcgTCAT, TCATcgATGA, and ATGAcgATGA (Figure 3). The concept of Protein Stitchery (15) is that the individual basic arms (half-sites) of the dimer and the individual half-sites of the DNA can be recombined or stitched together in various sequences to form new proteins selective for binding to the new DNA sites. Thus we use here the recognition helix v-Jun-br of Figure 1a with a cysteine linker either at the amino end (v-Jun-N) or at the carboxy end (v-Jun-C). These can be combined to form either pNN, pCC, or pCN dimers as

illustrated in Figure 1c. Formation of pNN and pCC (via pathway I) is straightforward since each involves dimerization of identical monomers. In order to ensure formation of pCN, the cysteine at the carboxy end of v-Jun-C was reacted with excess 2,2'-dithiodipyridine to form thiopyridyl-(v-Jun-C) (18,19) and then coupled with the cysteine at the amino end of v-Jun-N to form the pCN dimer (v-Jun-C)-S-S-(v-Jun-N) (pathway III, Figure 1c). We also verified pathway II for forming pCC.

Results and Discussion

We carried out gel retardation assays (15) for each of the three peptide dimers with oligonucleotides (Figure 1b) corresponding to each of the three proposed binding sites. These results (Figure 2a) show that each dimer recognizes the appropriate binding site specifically with no detectable binding to the other sites. It is important to note that *this strong preference for dimer occurs even though all oligonucleotides contain proper sites for binding a single arm of each dimer.* Therefore at 3 nM peptide concentration the dimer does not make a stable complex with DNA unless *both arms in the dimer recognize their proper sites.* This implies cooperation between the monomers in recognizing the binding site (20). Since all three dimers have similar (2 to 6 nM) binding affinities with their own sites and since all three lead to the same length protected region from DNase I digestion, we conclude that (i) all three cases involve similar conformations in the complex between DNA and peptide, and (ii) the

monomer arm retains the same contact region in various dimers. This occurs despite the changing orientation of the monomers in the various peptide dimers (15).

There are two major models for the bound conformation of leucine zipper protein to the specific site. One is the 'induced helical fork' model (13) which propose a *straight single α-helix* for the basic region, and the other is 'scissors grip' model (5) proposing a *bent α-helix* for the basic region. The recent X-ray crystal structure (21) for the complex of GCN4 containing only the basic and leucine zipper region and DNA containing GRE site showed that *the basic region of each protein has a straight α-helix conformation* recognizing each half site of the dimer binding site. There was no DNA bending caused by protein binding (21). However, there remain many problems with assuming that the basic region is in all cases a straight α-helix:

(i) the bases flanking the active site affect the binding of leucine zipper protein even though the crystal structure shows no direct contacts with protein (21)

(ii) gel electrophoresis experiments using Jun homodimer and Jun-Fos heterodimer showed that Jun and Fos induce DNA bending in opposite directions upon binding to their site (22), whereas GCN4 does *not* induce DNA bending (21,27). (iii) Even though GCN4-br (a peptide containing the basic region of GCN4 protein) showed no specific binding [for details see ref. 14], we find that the monomer v-Jun-br (a peptide containing only the basic region of v-Jun, see Figure 1a) specifically binds to the dimer site and shows the same protection

as the dimer. Our conclusion then is that there is no universal model for the DNA bound conformation of the basic region of leucine zipper proteins. Whether it is linear (as in GCN4) or bent (as in Jun) depends on the specific primary sequence and the properties of the solutions (stabilizers, pH, etc.) used in the experiments.

The result that all three dimers pCC, pNN, and pCN bind strongly to the appropriate combination of oligonucleotide sites implies that the helical binding arm is bent (5,22) (see Figure 3). Our argument is as follows. The optimum binding site for both Jun homodimer and the Jun-Fos heterodimer is known to be ATGAcgTCAT (or ATGAcgTCAT) where the inner seven (or eight) bases play an important role in recognition (23,24)

The X-ray crystal structure of GCN4 bound to DNA leads to straight a-helices which have direct contacts with only the inner seven bases of the GRE site (ATGACTCAT). Thus each arm recognizes the half-site (gATGAc or gTCATc) of the dimer binding site asymmetrically. If the same were true for v-Jun and if the same contacts are maintained between the protein and bases for the bound conformations of pCC/CC, pNN/NN, and pCN/CN (as expected since the binding constants and protection are the same), then the orientations of the binding arms would have very different orientations (Fig. 3e-f). This should result in different protection from DNase I digestion (not observed). In addition for the pNN/NN complex, this would lead to N-termini of the two arms too distant to be connected by the added linker, GGCCGG. The alternative to Fig 3def is for each

dimer to have the same angle (as in Fig. 3d). Thus the actual contact region would not be equivalent in the three cases and it would be difficult to explain the gel retardation and footprinting results. On the other hand, with the recognition helix bent roughly at the middle of the helix (as indicated in Figures 3a,b,c), it is plausible that the contact regions are ATGAcgTCAT for pCC, TCATcgATGA for pNN, and ATGAcgATGA for pCN. This leads to equivalent contact regions for all three cases and to the roughly equivalent binding energies apparent in Figure 2a. Thus we conclude that for v-Jun the basic region becomes bent upon binding to DNA. In addition, footprinting (15) of the three peptide dimers (Figure 2b), each with the appropriate oligonucleotide dimer, suggests that the complexed peptide dimers protect the full specific site (all 10 bp) from DNase I digestion. These results strongly support the bent recognition helix model for the basic dimers considered herein and hence also for the leucine-zipper parent proteins (22).

For the pCN/CN complex, the footprinting (Fig. 2b) shows incomplete protection on the binding site and partial protection on the bases flanking the binding site, whereas for pCC/CC and pNN/NN this does not happen. This occurs even though gel retardation assays indicate specific binding for all complexes. This occurs even though all three complexes show specific binding in gel retardation assays. Our explanation of this (Figure 4) suggests why palindromic sequences are so common for selective binding of regulatory proteins (25,26). This reasoning is supported by recent results we have observed

showing that (i) the *monomer* of v-Jun containing only the basic region (v-Jun-br) specifically protects both pCC and pNN binding sites identically to the protection provided by the dimers pCC and pNN, respectively, and (ii) at 3 nM concentration gel retardation showed that pCC (and pCN) has lower binding affinity for the DNA probe carrying a sequence of cgATGAcgTCATcgTCATcg [containing pCC and pCN binding sites overlapping half of each dimer binding site in the center] than for CC (and CN) probe DNA. These results imply that the half-site, gTCATc (or gATGAc), added next to the pCC (or pCN) binding site interferes with the binding of pCC (or pCN) to the dimer binding site [because the half-site can be used as a binding site for each arm of the dimer if the orientation between the site and arm fits each other]. Details of these results will be published elsewhere. Figure 4 indicates the strength of binding for all three peptide dimers at or near their DNA recognition sites. Here O represents good binding while X represents nonspecific binding. The palindromic sites for pNN and pCC lead to binding only when the protein is exactly at the recognition site whereas pCN can recognize both full site (both arms bound) and half sites (one arm bound). In DNase I footprinting and gel retardation assays, semispecific binding compete with specific binding. This occurs because one arm of the semispecifically bound peptide would cover half of the specific binding site, preventing another dimer from binding and providing full protection. This explains (i) why gel retardation assays (Fig. 2a) show lower binding affinity for the pCN/CN complex compared to the pCC/CC and

pNN/NN complexes and (ii) why footprinting assays (Fig. 2b) show incomplete protection on the binding site and partial protection on a few bases flanking the binding site. Such semispecific binding interferes with the site specific binding and would eventually result in low production and abnormally slow growth. However, gel retardation shows no detectable nonspecific or semispecific binding at low peptide concentration, indicating that semispecific binding is significantly weaker than specific binding. After dimerization, the proteins suitable for palindromic dimer binding sites avoid semispecific DNA binding, leading to more selective recognition of the specific sites. Thus *palindromic dimer binding sites provide good design for selective molecular recognition* and for further flexibility the link can align sites (Figure 3) to modify recognition.

The results on the three dimers considered here provide encouragement that this Protein Stitchery approach is feasible for designing and synthesizing proteins to recognize long DNA sequences. Thus for trimers to recognize 15 bp sequences, we are using an approach similar to Figure 1c involving appropriate use of cysteine linkages and transfer activators. It seems possible to design proteins for 20 bp and longer.

Summarizing:

(i). We find that Protein Stitchery of v-Jun leads to three dimers pCC, pNN, and pCN each of which binds specifically to the appropriate rearrangement of DNA sites. Thus, there is cooperation between the

two monomers of the dimer in binding to DNA which depends on the relative orientation of two monomers in the dimer.

(ii). These results provide strong support for the bent α-helix model of the basic region when bound to DNA.

(iii). These results provide an explanation for the advantage of dimerization and the use of palindromic sites in the site selective binding of proteins to DNA.

(iv). These results show Protein Stitchery to be useful for establishing the conformation and mechanism for binding of proteins to their DNA binding sites.

References

1. Curran, T. & Franza, Jr., B. R. (1988) Cell **55**, 395-397.
2. Kouzarides, T. & Ziff, E. (1989) Nature(London) **340**, 568-571.
3. Turner R. & Tjian, R. (1989) Science **243**, 1688-1694.
4. Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) Science **243**, 1695-1699.
5. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) Science **246**, 911-916.
6. Neuberg, M., Schermann, M., Hunter, J. B. & Muller, R. (1989) Nature(London) **338**, 589-590.
7. Ransone,L. J., Visvader, J., Wamsley, P. & Verma, I. M. (1990) Proc. Natl. Acad. Sci. USA. **87**, 3806-3810.
8. Ransone, L. J., Wamsley, P., Mosley, K. L. & Verma, I. M. (1990) Mol. Cell. Biol. **10**, 4565-4573.
9. Agre, P., Johnson, P. K. and S. L. McKnight, (1989) Science **246**, 922-926.
10. Neuberg, M., Adamkiewicz, J., Hunter,J. P & Muller, R. (1989) Nature(London) **341**, 243-245.
11. Weiss, M. A. (1990) Biochemistry **29**, 8020-8024.
12. Patel, L., Abate, C. & Curran, T. (1990) Nature(London) **347**, 572-575.
13. O'Neil,K. T., Hoess, R. H. & DeGrado, W. F. (1990) Science **249**, 774-778.

14. Talanian, R. V., McKnight, C. J. & P. S. Kim, P. S. (1990) Science **249**, 769-771.

15. Park, C., Campbell, J. L. & Goddard III, W. A. (1992) Proc. Natl. Acad. Sci. USA **89**, 9094-9096.

16. Kaki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. (1987) Proc. Natl. Acad. Sci. USA. **84**, 2848-2852.

17. Bos, T. J., Bohmann, D., Tsuchie, H., Tjian,R. & Vogt, P. K. (1988). Cell **52**, 705-712.

18. Corey D. & Schultz, P. G. (1987) Science **238**, 1401-1403.

19. Zuckermann, R., Corey, D, & Schultz, P. G. (1987) Nucleic Acids Res. **15**, 5305-5321.

20. Abate, C., Luk ,D., Gagne, E., Roeder, R. G. & Curran ,T. (1990) Mol. Cell. Biol. **10**, 5532-5535.

21. Ellenberg, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) Cell **71**, 1223-1237.

22. Kerpolla, T. K. & Curran, T. (1991) Science **254**, 1210-1214.

23. Ryseck, R. & Bravo, R. (1991) Oncogene **6**, 533-542 (1991).

24. Schule, R., Umesono, K., Mangelsdorf, D. J. Bolado, J., Pike, J. W., & Evans, R. M. (1990) Cell **61**, 497-504.

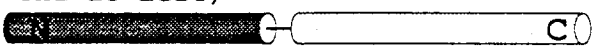
25. Pabo, C. O. & Sauer, R. T. (1984) Annu. Rev. Biochem. **53**, 293-321.

26. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) Science **240**, 1759-1764.

27. Gartenberg, M. R., Ampe, C., Steitz, T. A. & Crothers, D. M. (1990) Proc. Natl. Acad. Sci. USA **87**, 6034-6038.

Figure 1. Sequences of protein (a) and oligonucleotides (b) used in the gel retardation and footprinting studies. The total length of each oligonucleotide is 62. v-Jun-br contains the basic region of v-Jun. Cys-Gly-Gly is added to the N-terminus of v-Jun-br to make v-Jun-N and Gly-Gly-Cys is added to the C-terminus of v-Jun-br to make v-Jun-C. Proteins were chemically synthesized and checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology (15). (c) Shows the strategy for making pCC (and pNN) and pCN dimers. v-Jun-C was incubated with 10 mM dithiothreitol (DTT) for 5 hours at room temperature and purified directly into 10 mM 2,2'-dithiodipyridine, 100 mM sodium phosphate (pH 5.5) containing 30% acetonitrile. The resulting thiopyridyl-(v-Jun-C) was purified by HPLC. Purified monomer v-Jun-N was reacted with 2 equivalent of thiopyridyl-(v-Jun-C) in solution containing 100 mM TEAA buffer(pH 7.5) and 15% acetonitrile for 12 hours at room temperature. The final product, pCN, was purified by HPLC (15).

(a) Peptides (Amino end at left)



v-Jun-br: S QERIKAERKR MRNRIAASKS RKRKLERIAR

v-Jun-N : CGG S QERIKAERKR MRNRIAASKS RKRKLERIAR

v-Jun-C : S QERIKAERKR MRNRIAASKS RKRKLERIAR GGC

(b) Oligonucleotides

CC: 5'-ctcagatccggatcctaggttaaacg**ATGACGTCAT**cggtataggtcgagaattcggatcct-3'
3'-gagtctaggcctaggatccaatttgc**TACTGCAGTA**gccatatccagctcttaagcctagga-5'

CN: 5'-ctcagatccggatcctaggttaaacg**ATGACGATGA**cggtataggtcgagaattcggatcct-3'
3'-gagtctaggcctaggatccaatttgc**TACTGCTACT**gccatatccagctcttaagcctagga-5'

NN: 5'-ctcagatccggatcctaggttaaacg**TCATCGATGA**cggtataggtcgagaattcggatcct-3'
3'-gagtctaggcctaggatccaatttgc**AGTAGCTACT**gccatatccagctcttaagcctagga-5'

(c) Disulfide bond formation to make pCC (or pNN) and pCN

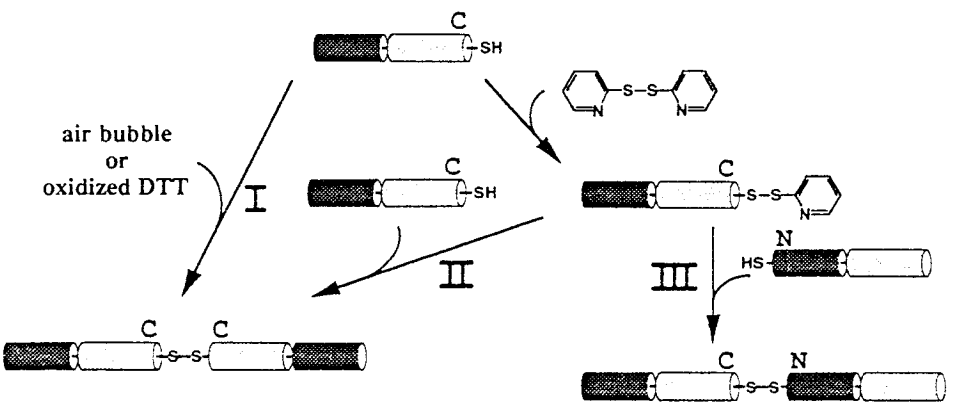


Figure 2a. Gel retardation assay for binding of pCC, pCN, and pNN to the CC, CN and NN probe DNAs. The binding solution contains bovine serum albumin (BSA) at 50 μg/ml, 10% (vol/vol) glycerol, 20 mM Tris.HCl (pH 7.5), 4 mM KCl, 4 mM $MgCl_2$ and 3 nM of appropriate peptide where indicated in a 10 μl reaction volume. After 5000 cpm of each 5'-32P-labeled probe DNA was added, the solutions were stored at 4° C for 40 min and loaded directly on an 8% nondenaturing polyacrylamide gel in Tris/EDTA buffer at 4° C. As determined by the titration of the gel shift, K_d ~ 2 nM for pCC/CC, K_d ~ 6 nM for pCN/CN, and K_d ~ 4 nM for pNN/NN all at 4° C. These results show that each peptide binds specifically to its proposed binding site and not to the other sites.

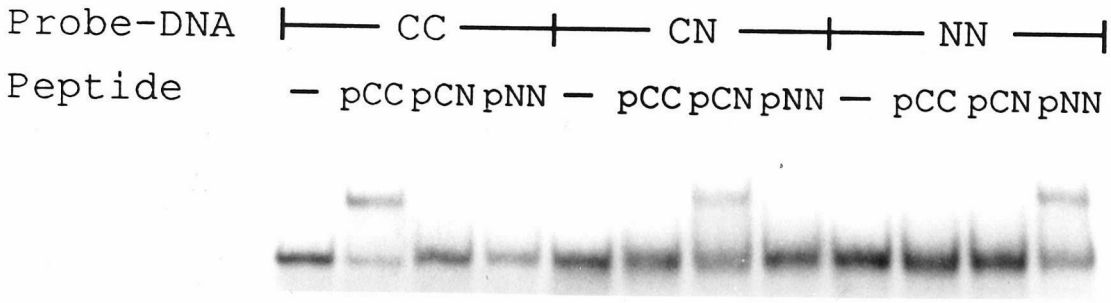

Probe-DNA
CC
CN
NN
Peptide
pCC pCN pNN
pCC pCN pNN
pCC pCN pNN

Figure 2b. DNase I footprinting assay of pCC, pCN and pNN peptide with DNA containing the predicted binding sites for, pCC, pCN, and pNN, respectively. The footprinting assay solution (in 50 µl) contains bovine serum albumin (BSA) at 50 µg/ml, 5% (vol/vol) glycerol, 20 mM Tris.HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 20,000 cpm of each 5'-32P-labeled probe DNA and 50 nM of appropriate peptide (where indicated). This solution was stored at 4° C for 40 min. After 5 µl of DNase I diluted in 1X footprinting assay buffer was added, the solutions were stored 1 min more at 4° C. The DNase I digestion was stopped by adding 100 µl of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, and sonicated salmon sperm DNA at 40 µg/ml. This mixture was phenol/chloroform-extracted, ethanol-precipitated and washed with 70% (vol/vol) ethanol. The pallet was resuspended in 5 µl of formamide loading buffer, denatured at 90° C for 4 min and analyzed on 10% polyacrylamide sequencing gel (50% urea). These results show that each peptide specifically binds to the proposed binding site and protects the whole site except for the case of pCN/CN which shows some incomplete protection on the binding site. This exception is explained as due to binding to semispecific (half) sites by single arms as discussed in the text (see Figure 4).

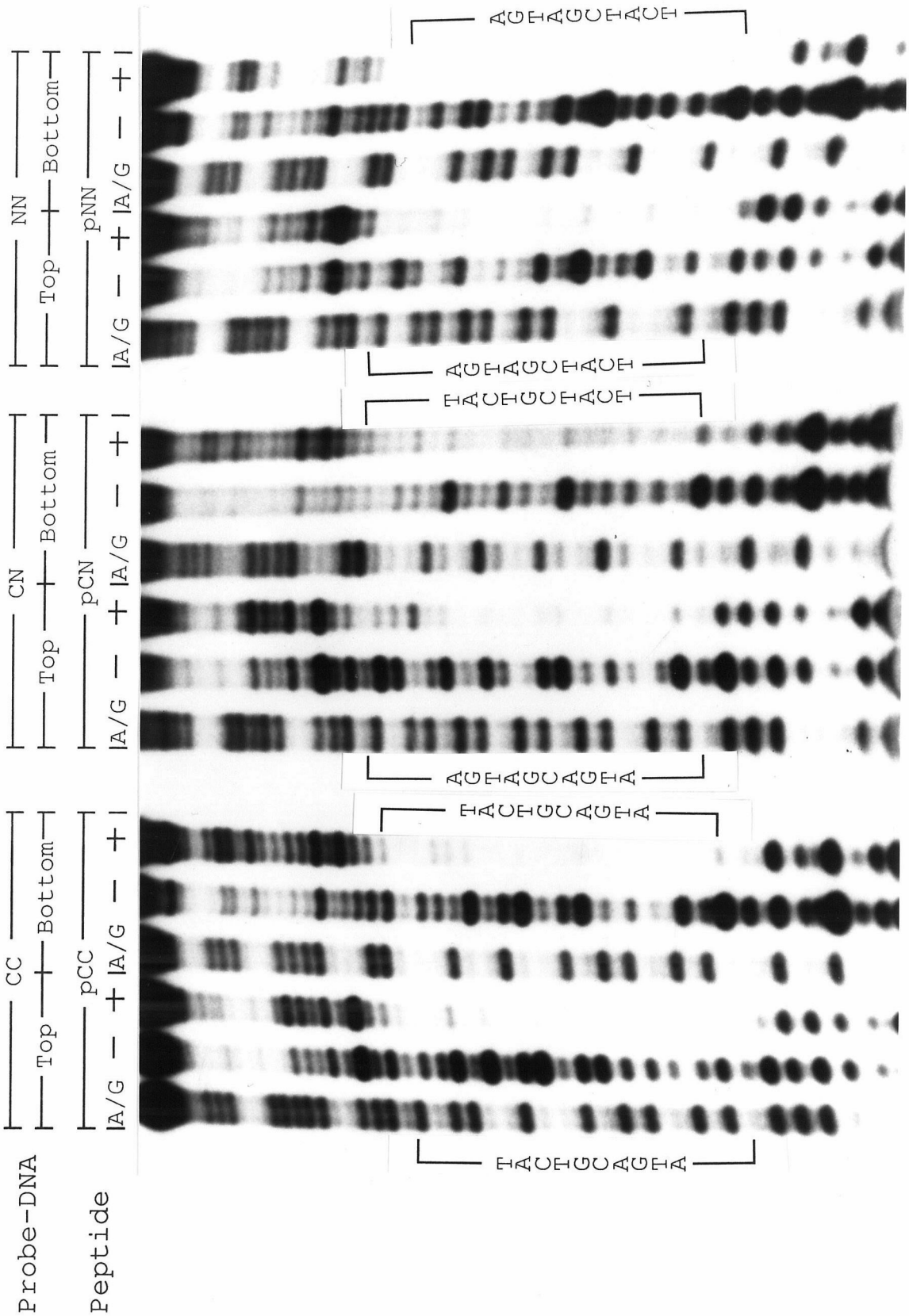

Probe-DNA
CC
Top
Bottom
Peptide
pCC
|A/G
−
+
|A/G
−
+
|
CN
Top
Bottom
pCN
|A/G
−
+
|A/G
−
+
|
NN
Top
Bottom
pNN
|A/G
−
+
|A/G
−
+
|
AGTAGCAGTA
TACTGCAGTA
AGTAGCTAACT
TACTGCTAACT
AGTAGCTAACT
TACTGCTAACT

Figure 3. Schematic diagram for the complex between peptides and their corresponding DNA sites assuming a bent recognition helix for (a)-(c) and a straight recognition helix for (d)-(f) based on the X-ray crystal structure of GCN4/DNA complex (21). (a) and (d) The complex between pCC and probe DNA CC. (b) and (e) The complex between pCN and probe DNA CN. (c) and (f) The complex between pNN and probe DNA NN. The linker connecting two monomers indicates a disulfide bond between the cysteines attached to the end of peptides. For each case the side view is on the top and the top view is on the bottom. The outer and inner circles of the top view represent the outer and inner major groove surfaces of the top strand for the proposed binding site projected onto an imaginary plane perpendicular to the axis of DNA and running through the center of peptide and binding site. Shading is used with the peptide and DNA contacts to ease the tracking of these regions for different cases. This diagram shows that a bent recognition helix can contact the same four bases for all three peptide dimers, while a linear recognition helix would contact different bases in the three peptide dimers. [This diagram is not meant to imply an exact correlation between where the basic region is bent and where the bases are positioned.]

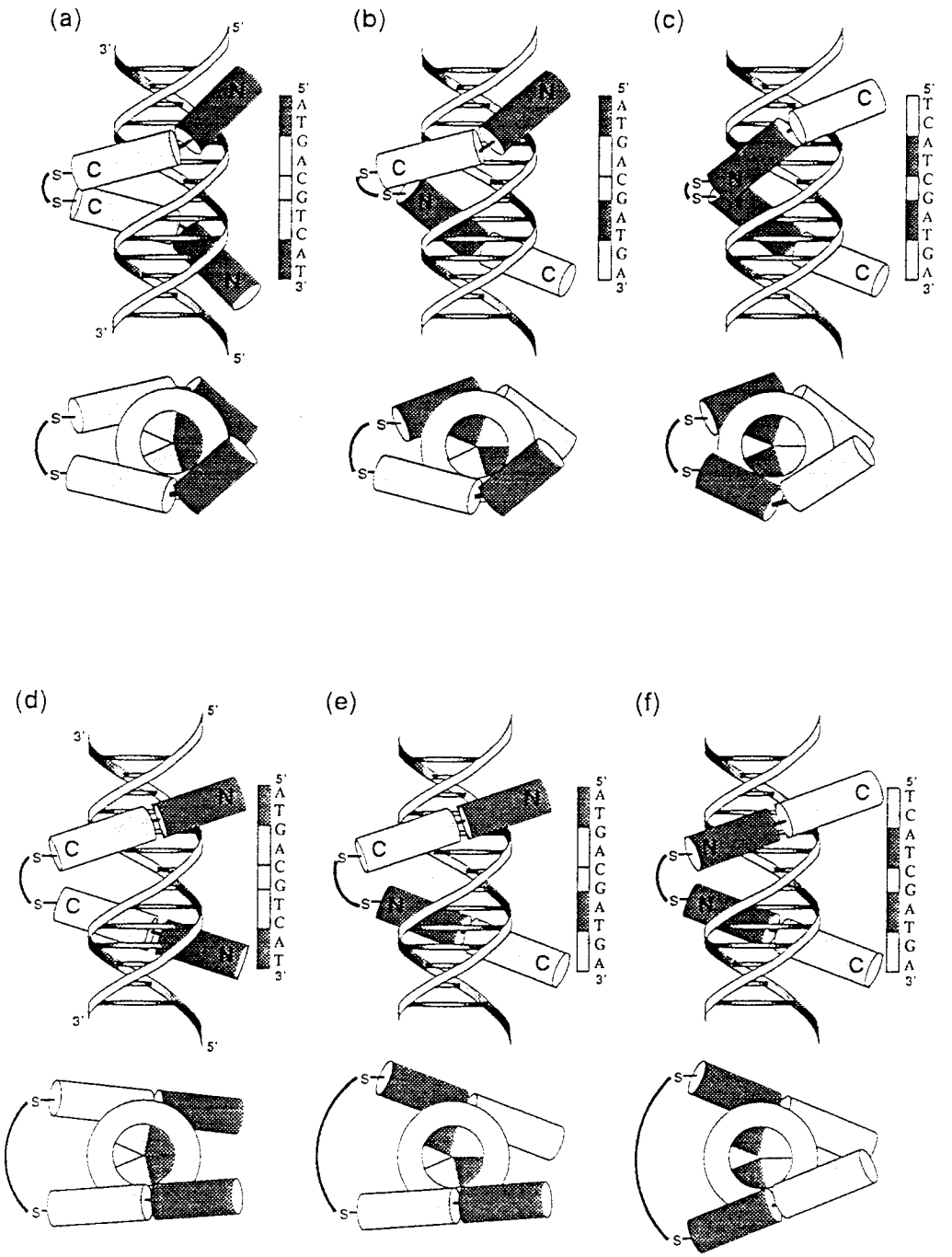
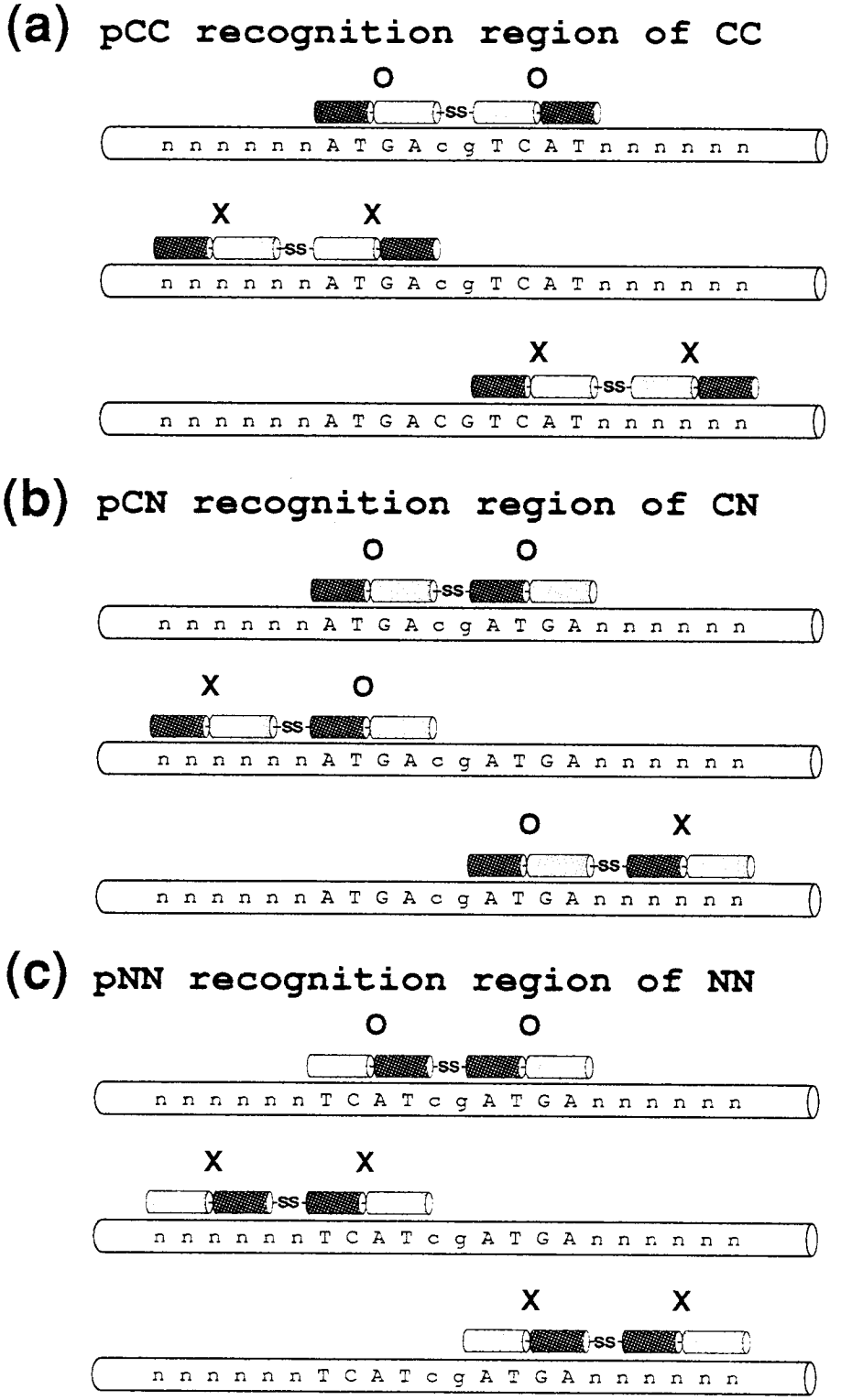

(a)
(b)
(c)
(d)
(e)
(f)
5'
3'
N
C
S
ATGACGTCAT
TCATCGATGA

Figure 4. Specific binding of protein at and near the corresponding DNA binding site. (a) The complex pCC/CC. (b) The complex pCN/CN. (c) The complex pNN/NN. Here O represents specific binding while X represents nonspecific binding. The case of pCN/CN has one specific binding site and two sites for semispecific (half-site) binding near its (non-parlindromic) binding site. However, the cases of pCC/CC and pNN/NN do not allow semispecific binding near their (parlindromic) binding sites.


(a) pCC recognition region of CC
O
O
SS
n n n n n n A T G A c g T C A T n n n n n n
X
X
SS
n n n n n n A T G A c g T C A T n n n n n n
X
X
SS
n n n n n n A T G A C G T C A T n n n n n n
(b) pCN recognition region of CN
O
O
SS
n n n n n n A T G A c g A T G A n n n n n n
X
O
SS
n n n n n n A T G A c g A T G A n n n n n n
O
X
SS
n n n n n n A T G A c g A T G A n n n n n n
(c) pNN recognition region of NN
O
O
SS
n n n n n n T C A T c g A T G A n n n n n n
X
X
SS
n n n n n n T C A T c g A T G A n n n n n n
X
X
SS
n n n n n n T C A T c g A T G A n n n n n n

Chapter 3

The Monomer of the DNA Binding Region of the v-Jun Leucine Zipper Protein Recognizes the Dimer Binding Site without Dimerization.

Changmoon Park*†‡, Judy L. Campbell†‡, and William A. Goddard III*†

*Materials and Molecular Simulation Center, Beckman Institute 139-74
†Division of Chemistry and Chemical Engineering, and ‡Division of Biology
California Institute of Technology, Pasadena, CA 91125

Abstract

It has been believed that leucine zipper proteins recognize the binding site as a dimer preformed in solution and that the monomer does not does not bind specifically to this site. We synthesized a 31 residue peptide, v-Jun-br, which contains only the DNA binding region of the v-Jun monomer. Footprinting assays show that v-Jun-br specifically recognizes and protects the binding site of v-Jun homodimer pCC (and the rearranged binding site) almost identically as do dimers. This means that the monomer recognizes the half-site of the dimer binding site and dimerization does not affect the bound conformation of each monomer. These results support the possibility that two monomers of v-Jun might bind sequentially to the dimer site with dimerization of v-Jun occurring while bound.

I. Introduction

Many DNA binding proteins are characterized by specific binding motifs (e.g., helix-loop-helix, zinc finger, leucine zipper) many of which recognize their parlindromic DNA binding sites as dimers (1). Leucine zipper proteins have common structural motifs confined on a region of about 60 residues with a leucine zipper region (4 or 5 leucines occurring exactly every 7 residues) on the C-terminus responsible for dimerization and a basic region on the N-terminus which is responsible for DNA binding (2).

The basic region has no fixed conformation in solution, but changes into an α-helix when bound to the specific site (3-6). Experiments using the basic region of GCN4 (7) or v-Jun (8,9) showed that without the leucine zipper region the basic region alone will recognize the dimer binding site (denoted CC) when dimerized at the carboxy termini (denoted pCC) by an added linker. In addition dimerization at the amino termini leads to recognition of the rearranged pNN binding site (8,9).

The X-ray crystal structure of GCN4 bound to DNA shows a straight α-helix conformation for the basic region and no bending of the DNA induced by protein binding (10). However, Jun homodimer and Jun-Fos heterodimer do induce DNA bending (in opposite directions) upon binding to the specific site (11). To explain this DNA bending it was proposed for the basic region that Jun has a bent α-helix while Fos has a straight α-helix. The equivalent binding and footprinting using three v-Jun homodimers, pCC, pCN, and pNN, was

interpreted as evidence for a bent helix for the bound basic region of v-Jun (9).

It is widely believed that protein dimerization is essential for leucine zipper protein to effect specific DNA recognition. Evidence in favor of this view are the observations that:

(i) most mutations that prevent dimerization also prevent DNA binding (12-14),

(ii) GCN4 can make stable dimer in the absence of specific DNA (15), and

(iii) the oxidized dimer of the GCN4 basic region specifically recognizes the dimer binding site, but the reduced monomer does not (7).

On the other hand NMR experiments showed that in the absence of the specific DNA binding site the lifetime of GCN4 homodimer is between 10 ms and 1 s (6). Consequently GCN4 dimer is not stable in solution in the absence of specific DNA.

In order to obtain a direct test of whether predimerization is essential for the binding of leucine zipper protein, we synthesized a peptide v-Jun-br (Figure 1a) containing only the basic region of v-Jun monomer and did footprinting assays for oligonucleotides containing the dimer binding site.

II. Experimental Details

A. Oligonucleotide and peptide preparation

The probe DNAs CC and NN (Figure 1b) were prepared as described (9). Peptide v-Jun-br was chemically synthesized and purified as described (8) (see Caption, Figure 1). The dimers pCC and pNN were prepared by oxidization of the v-Jun-N or v-Jun-C peptides as described (9). Protein concentration was determined as described (8).

B. Footprinting

The probe DNAs for footprinting assays were prepared as described (8). Peptide concentrations were determined as described (8) (see Caption Figure 2 for details).

III. Results

Figure 2 columns 3 and 7 show that for the CC probe DNA (top and bottom) the dimer pCC leads to recognition of the CC binding site. Figure 2 columns 4 and 8 show that monomer v-Jun-br also *protects the complete CC dimer binding site.* Figure 2 columns 11 and 15 show that for the NN probe DNA (top and bottom), the dimer pNN leads to recognition of the NN binding site. figure 2 columns 12 and 16 show that the monomer v-Jun-br also *protects the complete NN dimer binding site.*

Because v-Jun-br contains only the basic region, there is no possibility of dimerization. Since the C-termini become positioned

near each other when two monomers bind to the pCC binding site while the N-termini of both monomers are positioned near each other when two monomers bind to the pNN binding site, the similarity in the results between monomers and dimers shows that there are no specific interactions between the two monomers when bound to the site. These results show also that the added linkers in pCC (and pNN) have little influence on the DNA bound conformations of monomers of pCC (and pNN).

It has been believed that the monomers for leucine zipper proteins have no ability for specific DNA recognition (7,12-14). However, the footprinting assays of Figure 2 show that *v-Jun-br specifically binds to and protects the same dimer binding site of pCC (and pNN).* These results show that there are no specific interactions between the two monomers on the binding site which influence monomer binding to the dimer binding site. Thus binding of monomers to NN versus CC probe DNAs must lead to different ends of the two monomers coming together, therefore, the monomer of v-Jun basic region specifically binds to the half-site of the dimer binding site.

The results of footprinting assays (Figure 2) show that *the monomer v-Jun-br protects identically as the dimer pCC (and pNN).* These results mean that the added linkers (Gly-Gly-Cys or Cys-Gly-Gly) when oxidized to make dimer do not change the bound conformations of the monomers on the binding site of pCC (and pNN) and so, each monomer retains the same contacts with DNA on both sites (9). These

results also mean that oxidization and covalent bond of the thiol groups of the linkers to make the pCC and pNN dimers do not lead to sufficient tension to change the contacts between the monomer and DNA. These results support the bent α-helix for the basic region of v-Jun (9,11). The reason is that with straight helices placing the monomers for pNN at the sites corresponding to the bound pCC would put the N-termini of two monomers of pNN in positions which can not be reached without bending the basic region (9).

IV. Discussion

Leucine zipper proteins dimerize using the leucine zipper region to form a coiled-coil structure for the dimer (10,16). Most mutant leucine zipper proteins which are unable to carry out dimer formation fail to recognize the binding site (12-14). Therefore, it has been believed that the dimerization of leucine zipper protein is a prerequisite to specifically recognizing the binding sites. This idea is supported by the results that the oxidized dimer of GCN4 basic region specifically recognizes the GCN4 dimer binding site but the monomer does not (7,17).

However, our current results (figure 2) show that the monomer of v-Jun basic region (v-Jun-br) specifically binds to both halves of the dimer binding sites of pCC and pNN. Because the v-Jun-br has no functional motif to become a dimer and because it recognizes both the binding site of pCC and of pNN, we conclude that v-Jun-br recognizes the half-site of the dimer binding site as a monomer, this contrasts

with the situation for GCN4. This result implies that v-Jun might dimerize on the binding site, removing the prerequisite of dimerization before binding of course. It could be that v-Jun bind to DNA in a different conformation from that of GCN4.

These results are consistent with a recent study (18) on the DNA binding protein RexA which recognizes a site with dyad symmetry as a dimer. RexA has similar dissociation constant with leucine zipper protein for the complex formation (18,19) between protein and DNA, and for protein dimerization (6,20). However, recent experiments show that the monomer of RexA recognizes the half-site of the full dimer binding site, and it was proposed that two monomers of RexA bind sequentially to the binding site through making the dimer on the binding site.

For both pCC and pNN binding site the monomer and dimer both lead to complete protection of the binding site and the same length of protection. This result supports the bent model for the basic region of v-Jun (11), Thus if the basic region has a straight α-helix, the orientations of each monomers would become quite different on both binding sites resulting in different protections (9). These results support the results of gel electrophoresis (11) using Jun homodimer, since the X-ray crystal structure of GCN4 (10) bound to DNA shows straight α-helices. This indicates that the v-Jun and GCN4 proteins behave differently.

References

1. Pabo, C. O. & Sauer, R. T. (1984) Annu. Rev. Biochem. **53**, 293-321.
2. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) Science **246**, 911-916.
3. Weiss, M. A. (1990) Biochemistry **29**, 8020-8024.
4. Patel, L., Abate, C. & Curran, T. (1990) Nature(London) **347**, 572-575.
5. O'Neil,K. T., Hoess, R. H. & DeGrado, W. F. (1990) Science **249**, 774-778.
6. Weiss, M. A., Ellenberg, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990) Nature(London) **347**, 575-578
7. Talanian, R. V., McKnight, C. J. & P. S. Kim, P. S. (1990) Science **249**, 769-771.
8. Park, C., Campbell, J. L. & Goddard III, W. A. (1992) Proc. Natl. Acad. Sci. USA **89**, 9094-9096.
9. Park, C., Campbell, J. L. & Goddard III, W. A. (1993) Proc. Natl. Acad. Sci. USA (in print)
10. Ellenberg, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) Cell **71**, 1223-1237.
11. Kerpolla, T. K. & Curran, T. (1991) Science **254**, 1210-1214.
12. Neuberg, M., Adamkiewicz, J., Hunter,J. P & Muller, R. (1989) Nature(London) **341**, 243-245.
13. Turner R. & Tjian, R. (1989) Science **243**, 1688-1694.

14. Heeckeren, W. J., Sellers, J. W. & Struhl, K. (1992) Nucleic Acids Res. **20**, 3721-3724.

15. Hope, I. A. & Struhl, K. (1987) EMBO J. **6**, 2781-2784.

16. Rasmussen, R., Benregnu, D., O'shea, E. K., Kim, P. S. & Alber, T. (1991) Proc. Natl. Acad. Sci. USA **88**, 561-564.

17. Talanian, R. V., McKnight, C. J., Rufkovski, R. & Kim, P. S. (1992) Biochemistry **31**, 6871-6875.

18. Kim, B. & Little, J. W. (1992) Science **255**, 203-255.

19. Hope, I. A. & Struhl, K. (1985) Cell **43**, 177-188.

20. Schnarr, M., Ponyet, J., Granger-Schnarr, M. & Daune, M. (1985) Biochemistry **24**, 2812-2818.

Figure 1. Sequences of protein (a) and oligonucleotides (b) used in the gel retardation and footprinting studies. The total length of each oligonucleotide is 62. v-Jun-br contains the basic region of v-Jun. Peptide v-Jun-br and v-Jun-C were prepared as descrived (8,9). Peptide v-Jun-br was chemically synthesized and purified and the purity was checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology (8) : calculated; 3822.3, experimental; 3824.6.

a Peptides

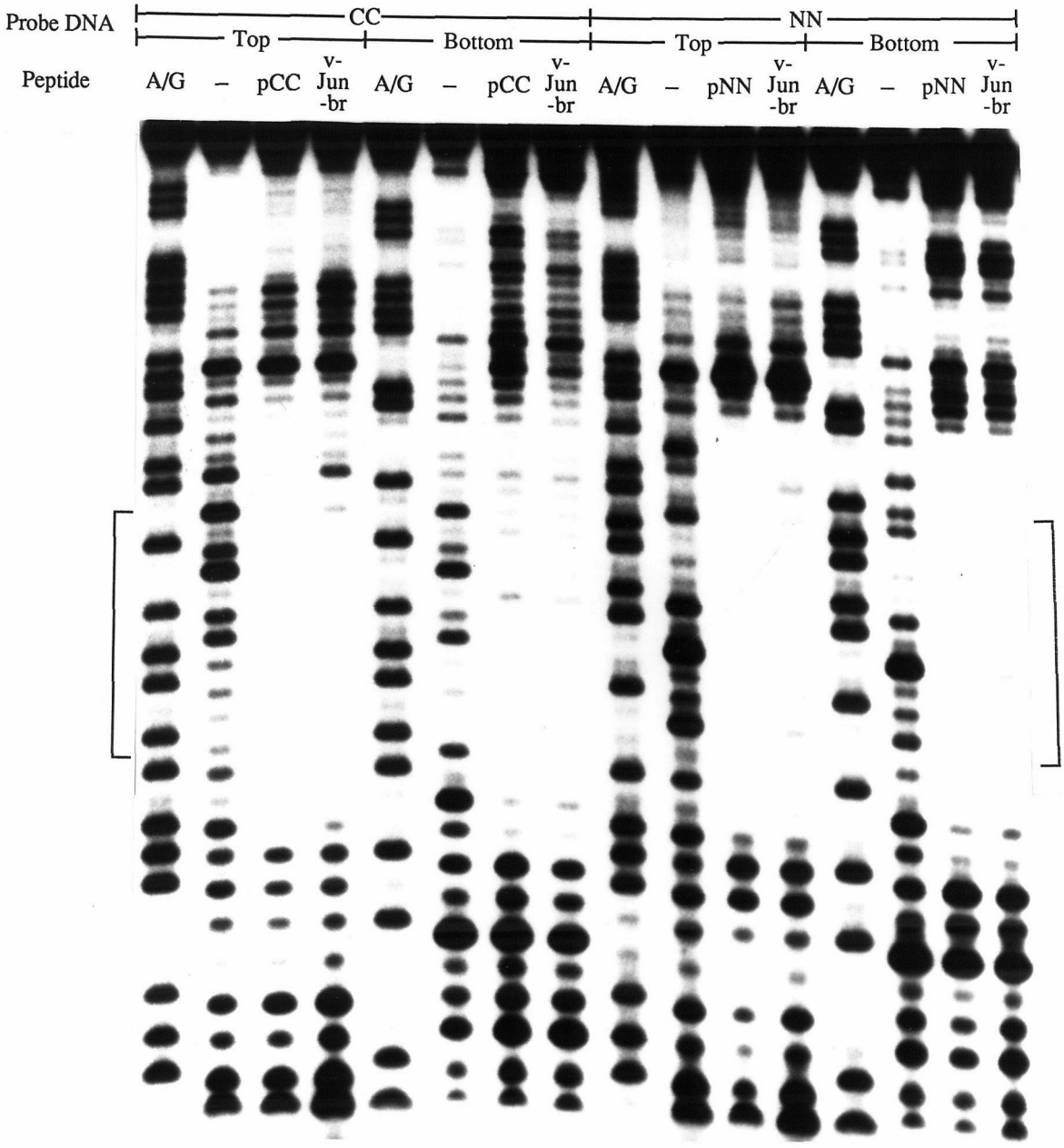
```
v-Jun-br:        S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
V-Jun-N : CGG    S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
v-Jun-C :        S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR  GGC
```

b Oligonucleotides

```
CC: 5'-ctcagatccggatcctaggttaaacgATGAcgTCATcggtataggtcgagaattcggatcct-3'
    3'-gagtctaggcctaggatccaatttgcTACTgcAGTAgccatatccagctcttaagcctagga-5'

NN: 5'-ctcagatccggatcctaggttaaacgTCATcgATGAcggtataggtcgagaattcggatcct-3'
    3'-gagtctaggcctaggatccaatttgcAGTAgcTACGgccatatccagctcttaagcctagga-5'
```

Figure 2. DNase I footprinting assays of v-Jun-br with probe DNAs CC and NN. In order to compare the results of protections between monomer and dimer, DNase I Footprinting assays of pCC and pNN were also carried out together with probe DNAs CC and NN, respectively. The brackets show the proposed dimer binding sites for probe DNAs (see figure 1b). Peptide concentrations were determined as described (8). The reaction solution (in 50 μl) contains 5% (vol/vol) glycerol, 20 mM Tris.HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 50,000 cpm of each 5'-32P-labeled probe DNA, bovine serum albumin (BSA) at 0.1 %, poly(dI.dC) at 2 μg/ml, and 600 nM of pCC (or pNN) or 3 μM v-Jun-br (where indicated). This solution was stored at 4° C for 40 min. After 5 μl of DNase I diluted in 1X footprinting assay buffer was added, the solutions containing v-Jun-br were digested for 10 min and the solutions containing pCC (or pNN) were digested for 6 min at 4° C. After DNase I digestion, all the remaining procedures were followed as described (8).


Probe DNA
CC
NN
Top
Bottom
Top
Bottom
Peptide
A/G
–
pCC
v-Jun-br
A/G
–
pCC
v-Jun-br
A/G
–
pNN
v-Jun-br
A/G
–
pNN
v-Jun-br

Chapter 4

Design and Synthesis of a New Peptide Recognizing a Specific 16 bp Site of DNA.

Changmoon Park*†‡, Judy L. Campbell†‡, and William A. Goddard III*†

*Materials and Molecular Simulation Center, Beckman Institute 139-74
†Division of Chemistry and Chemical Engineering, and ‡Division of Biology
California Institute of Technology, Pasadena, CA 91125

Abstract

We designed a peptide to recognize a new 16 bp site (about 1.5 turn) of DNA. With DNase I footprinting we show that the new peptide specifically recognizes the proposed site with a dissociation constant of about 3 nM at 4° C. Gel retardation shows that all three arms of the new peptide specifically recognizes the proposed site (which consists of three 5 bp half-sites), but not to the dimer sites which have two half-sites for two arms. Footprinting assays show that the new peptide binds and protects the exact proposed site. These results also provide information about the reaction of specific and nonspecific binding in the recognition between protein and DNA.

I. Introduction

Specific DNA binding proteins play important roles in biological system. The regulation of cellular reactions including replication, transcription, and translation are mostly mediated by the specific interactions of DNA binding proteins with DNA (1). Therefore, design and synthesis of sequence-specific DNA binding proteins is of great interest in modern chemistry and biology.

Synthesis of peptides specifically recognizing long sequences (more than 10 bp) of DNA is also important in mapping large genomes. Most known restriction enzymes recognize 4 to 8 bp sites, so they create too many fragments to be handled when they are used to digest genomic DNA. Many attempts have been developed to recognize (and cleave) a specific site on a large size of DNA (2,3). Recent methods (4-6) include a set of steps to recognize (and cut) a specific restriction site among many identical restriction sites existing on the same DNA. These involve:

(i) protection of specific restriction sites by making a complex between the DNA binding protein and duplex DNA or by making a triple helix between synthetic nucleic acid and duplex DNA,

(ii) methylation using a specific methylase on all corresponding restriction sites except the specific site,

(iii) denaturation of the complex made in step (i), and

(iv) digestion with an appropriate restriction enzyme.

This procedure creates a single site-specific cleavage when used even in large genomic DNAs (6). However, this procedure requires

complete protection of the specific site from mathylation, or it will create nonspecific cleavages.

Another approach to recognize a long site of DNA is to design and synthesize new protein specifically recognizing a long site of DNA. Such a synthetic protein could also be used as a DNA nuclease by covalently attaching iron chelator (EDTA) (7) or copper-binding tripeptide (GlyGlyHis) on the amino terminus (8), not always possible for the natural DNA binding protein.

To illustrate the design a protein specifically recognizing a long 16 bp site of DNA, we stitch together a new peptide from three v-Jun basic regions. v-Jun is a member of leucine zipper protein which regulates DNA transcription and binds a site with dyad symmetry as a homodimer or a heterodimer with Fos (9,10). A recent X-ray crystal structure for the complex of GCN4 and DNA (11) shows that the dimerization is mediated by the leucine zipper region, and that each basic region recognizes the half-site of the dimer binding site. Gel retardation and footprinting assays were used to investigate the properties of new peptide. These results also give further insight on the interaction between leucine zipper protein and DNA.

II. Materials and Experiments

A. Oligonucleotides and peptides preparation

Probe DNAs CC and CN were prepared as described (13). Oligonucleotides CC•NC (Figure 1b) was chemically synthesized and purified at the Biopolymer Synthesis Center at the California Institute

of Technology as described (12). Peptides pCC and pCN were prepared as described (13). Peptide v-Jun-NC (Figure 1a) was chemically synthesized and purified as described (12).

B. Protein Stitchery to make trimer

The procedures to make trimer (Figure 1c) are the same as the procedures to make dimer pCN except that v-Jun-NC was used instead of v-Jun-N. 2,2'-dithiodipyridine was used to activate the thiol group of v-Jun-C ensure disulfide bond formation between specific thiol groups (13-15). v-Jun-C was incubated with 10 mM dithiothreitol (DTT) for 5 hours at room temperature and purified directly into 10 mM 2,2'-dithiodipyridine, 100 mM sodium phosphate (pH 5.5) containing 30% acetonitrile. The resulting thiopyridyl-(v-Jun-C) was purified by HPLC. Purified monomer v-Jun-NC was reacted with 3 equivalents of thiopyridyl-(v-Jun-C) in solution containing 100 mM TEAA (tetra ethyl ammonium acetate) buffer(pH 7.5) and 15% acetonitrile for 12 hours at room temperature. The final trimer product, pCC•NC, was purified by HPLC. The HPLC analysis showed that reduction of pCC•NC gives two peaks corresponding to v-Jun-C and v-Jun-NC in the expected1:2 ratio.

C. Gel retardation

Gel retardation assays were carried out as described (12). 5000 cpm of each 5'-32P-labeled probe DNA and 3 nM of appropriate peptide were used as indicated in Figure 2. The results show that each

peptide binds specifically to its proposed binding site and not to the other sites.

D. Footprinting

Footprinting assays were performed as described (12). 50,000 cpm of each 5'-32P-labeled probe DNA and 50 nM of appropriate peptide were used as indicated. This solution was stored at 4° C for 40 min. After 5 μl of DNase I diluted in 1X footprinting assay buffer was added, the solutions were digested for 4 min at 4° C. The remaining procedures were performed as described (12).

III. Results

A. pCC•NC recognizes the proposed site specifically.

To determine the binding properties of pCC•NC, gel retardation assays were carried out for probe DNAs CC•NC, CC, and CN. The results (Figure 2) show that pCC•NC binds to probe DNA CC•NC which has the exact site for all three arms of pCC•NC, but does not to probe DNAs CC or CN, each of which has a site for two arms of pCC•NC. This means that all three arms of pCC•NC are participating in the recognition and there is cooperation among all three arms in the pCC•NC (it was previous shown that there is cooperation in the dimers, pCC, pCN, and pNN) (12,13). Combined with the results for v-Jun homodimers, this indicates that pCC•NC makes contacts with about 16 bp of DNA (about 1.5 turn of duplex DNA) along the major groove.

B. The half-site interferes with dimers to bind the dimer binding sites.

pCC (and pCN) shows lower binding affinity for the probe DNA CC•NC than for the probe DNA CC (and CN) even though the probe DNA CC•NC has a binding site for pCC (and pCN) (Figure 2). If the half-site did not interfere with dimers to bind the dimer binding site, there would be no difference in the binding affinities of pCC (or pCN) between to CC•NC and CC (or CN). However, about two times higher affinity was detected for pCC (or pCN) to the site CC (or CN) than to the site of CC•NC. This means the half-site added next to the binding site of pCC (or pCN) to make the binding site of pCC•NC interferes with pCC (or pNN) to bind the dimer binding site.

C. pCC•NC specifically protects the proposed binding site.

To determine the specific site for binding of pCC•NC to DNA, footprinting assays were performed. The results (Figure 3) show that the new peptide pCC•NC protects the exact proposed site, confirming the results from gel retardation assays. These results also indicate that each arm of pCC•NC binds to the proposed half-site and protects that half-site from DNase I digestion.

For the top strand of probe DNA CC•NC there is a partial protection on the 3' end bases flanking the binding site. This partial protection can be explained as the result of semispecific binding occurring between the binding site and peptide pCC•NC. The probe DNA CC•NC can provide two major sites for pCC•NC (Figure 4):

(i) one is a site for all three arms of pCC•NC resulting in specific bindings for all three arms, and

(ii) the other is a site for two arms of pCC•NC, that is, a binding site of pCC, which creates a semispecific binding that two arms have specific binding and one arm has nonspecific binding.

IV. Discussion

The observation that

(i) the monomer of v-Jun basic region specifically recognizes the dimer binding site (16) and

(ii) the half-site of the dimer binding site added next to the dimer binding site interferes with the dimer, reducing binding affinity to the dimer binding site. Both results support the idea of semispecific binding (13). If there are no interactions between the dimer and the half-site, the probe DNA CC (and CN) and CC•NC would behave identically to pCC (and pCN) except for a six bp difference in size (which does not seem to influence gel retardation). However, the results of gel retardation assays (Figure 2) show that the binding affinity of pCC (and pCN) to DNA increased upon changing the probe DNA from CC•NC to CC (and CN). This means there is some interaction between the half-site and the dimer. Within the concept of semispecific binding, we expect that probe DNA CC•NC provides (i) 1 specific binding and 1 semispecific binding for peptide pCC, and (ii) 1 specific binding and 4 semispecific bindings for peptide pCN. However, probe DNA CC provides only 1 specific binding, while probe

DNA CN provides 1 specific binding and 2 semispecific bindings. Therefore, the observation of a double binding affinity of pCC (and pCN) to DNA by changing probe DNA from CC•NC to CC (and CN) support the idea of semispecific binding (13). The footprinting assays (Figure 3) indicate a partial protection on the 5' end bases flanking the binding site of pCC•NC for the top strand. This also supports the concept of semispecific binding. This is because probe DNA CC•NC can bind in two ways (Figure 4): (i) a specific binding where all three arms make specific interactions, and (ii) a semispecific binding which has 2 arms making specific interactions and 1 arm making nonspecific interactions.

Summarizing,

(i) the new peptide trimer pCC•NC specifically binds to the proposed trimer binding site and not to the dimer binding sites.

(ii) The results of gel retardation assays footprinting assays support the idea of semispecific binding.

(iii) the results show the power of using Protein Stitchery both to design new peptides (in this case for recognizing new and longer sites) and to investigate the details of protein DNA interactions.

References

1. Pabo, C. O. & Sauer, R. T. (1984) Annu. Rev. Biochem. **53**, 293-321.
2. McClelland, M., Kessler, L. & Bittner, M. (1984) Proc. Natl. Acad. Sci. USA **81**, 983-987.
3. Pei, D., Corey, D. R. & Schultz, P. G. (1990) Proc. Natl. Acad. Sci. USA **87**, 9858-9862.
4. Koob, M., Grimes, E. & Szybalski, W. (1988) Science **241**, 1084-1086.
5. Strobel, S. A. & Dervan, P. B. (1990) Science **249**, 73-75.
6. Ferrin, L. J. & Camerini-Otero, R. D. (1991) Science **254**, 1494-1497.
7. Sluka, J. P., Hovath, S. J., Bruist, M. F., Simon, M. I. & Dervan, P. B. (1987) Science **238**, 1129-1132.
8. Mack, D. P. Iverson, B. L. & Dervan, P. B. (1988) J. Am. Chem. Soc. **110**, 7572-7574.
9. Kaki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. (1987) Proc. Natl. Acad. Sci. USA. **84**, 2848-2852.
10 Bos, T. J., Bohmann, D., Tsuchie, H., Tjian,R. & Vogt, P. K. (1988). Cell **52**, 705-712.
11. Ellenberg, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) Cell **71**, 1223-1237.
12. Park, C., Campbell, J. L. & Goddard III, W. A. (1992) Proc. Natl. Acad. Sci. USA **89**, 9094-9096.

13. Park, C., Campbell, J. L. & Goddard III, W. A. (1993) Proc. Natl. Acad. Sci. USA (in print)

14. Corey D. & Schultz, P. G. (1987) Science **238**, 1401-1403.

15. Zuckermann, R., Corey, D, & Schultz, P. G. (1987) Nucleic Acids Res. **15**, 5305-5321.

16. Park, C., Campbell, J. L. & Goddard III, W. A. "The monomer of the DNA binding region of the v-Jun leucine zipper protein recognizes the dimer binding site without dimerization." (To be submitted for publication)

Figure 1. Sequences of protein (a) and oligonucleotides (b) used in the gel retardation and footprinting studies. The total length of each oligonucleotide is 62. v-Jun-br contains the basic region of v-Jun. To make v-Jun-CN, both Cys-Gly-Gly and Gly-Gly-Cys is added to the N- and C-terminus of v-Jun-br, respectively. v-Jun-N and v-Jun-C were prepared as descrived (12,13). v-Jun-CN was chemically synthesized and purified and , and the purity was checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology as described (12): calculated; 4256.4, experimental; 4259.1. (c) Shows the strategy for making pCC•NC trimer.

(a) Peptides (amino terminus on the left)

```
v-Jun-br:       S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
V-Jun-N : CGG   S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR
v-Jun-C :       S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR  GGC
v-Jun-CN: CGG   S  QERIKAERKR  MRNRIAASKS  RKRKLERIAR  GGC
```

(b) Oligonucleotides

```
CC·NC: 5'-ctcagatccggatcctaggttaaacgATGAcgTCATcgTCATcggtataggtcgagaattcggatcct-3'
       3'-gagtctaggcctaggatccaatttgcTACTgcAGTAgcAGTAgccatatccagctcttaagcctagga-5'

CC   : 5'-ctcagatccggatcctaggttaaacgATGAcgTCATcggtataggtcgagaattcggatcct-3'
       3'-gagtctaggcctaggatccaatttgcTACTgcAGTAgccatatccagctcttaagcctagga-5'

NN   : 5'-ctcagatccggatcctaggttaaacgATGAcgATGAcggtataggtcgagaattcggatcct-3'
       3'-gagtctaggcctaggatccaatttgcTACTgcTACTgccatatccagctcttaagcctagga-5'
```

(c) Procedure to make pCC·NC

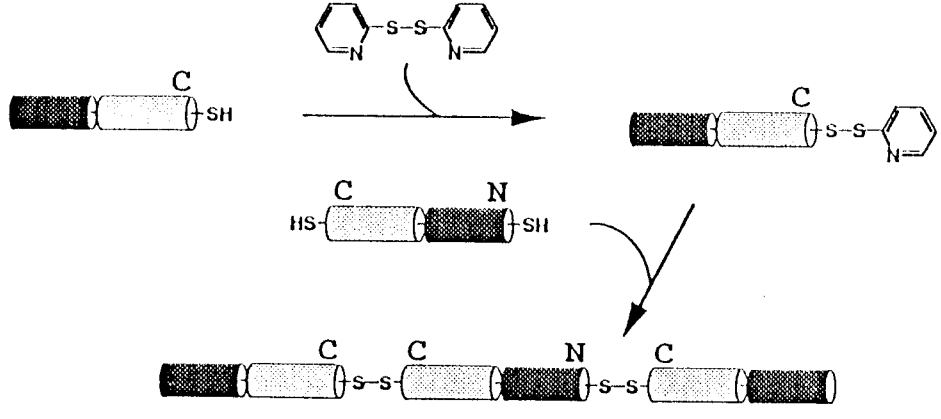


Figure 2. Gel retardation assays for binding of pCC•NC pCC, and pCN to the CC•NC, CC, and CN probe DNAs were carried out as described (12). 3 nM of each peptide was used in 10 μl reaction volume where indicated with each of 5000 cpm appropriate probe DNA. As determined by the titration of the gel shift, Kd ~ 2 nM for pCC/CC, Kd ~ 5 nM for pCC/CC•NC, Kd ~ 6 nM for pCN/CN, Kd ~ 15 nM for pCC/CC•NC, and Kd ~ 5 nM for pCC•NC/CC•NC all at 4° C.

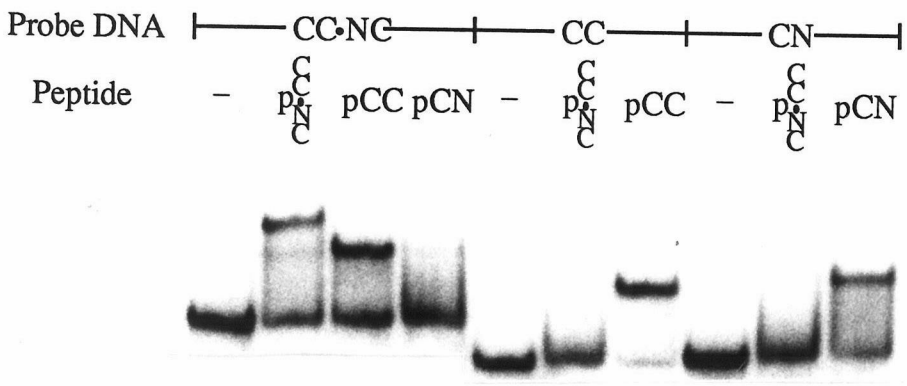

Probe DNA
CC•NC
CC
CN
Peptide
–
pCC pCN
–
pCC
–
pCN

Figure 3. DNase I footprinting assays of pCC•NC with probe DNA CC•NC were performed as described (12). 3 nM of pCC•NC was used with 50,000 cpm of probe DNA CC•NC in 50 μl reaction volume. The brackets show the proposed trimer pCC•NC binding sites for probe DNAs (see Figure 1b).

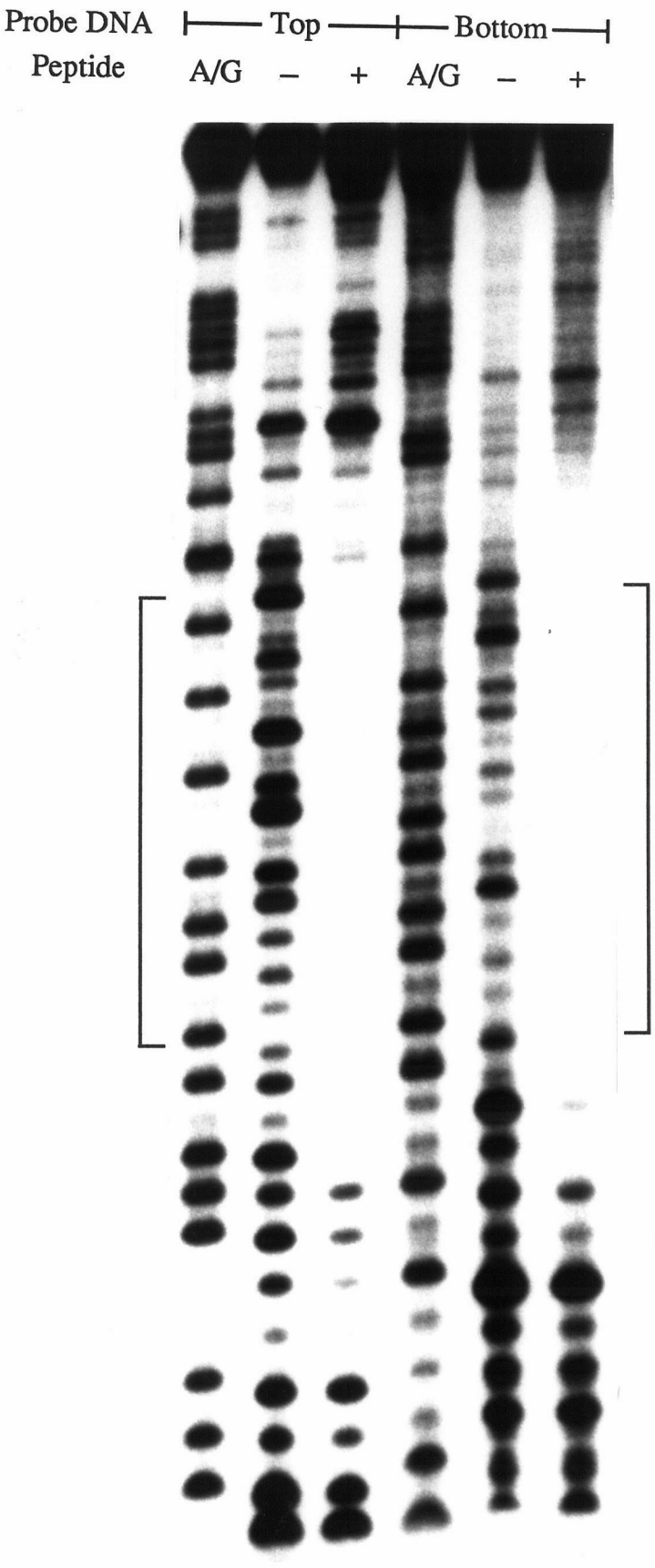
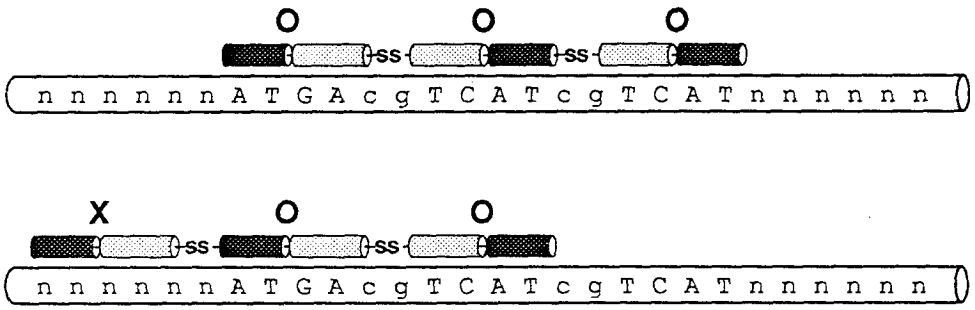

Probe DNA
Top
Bottom
Peptide
A/G
–
+
A/G
–
+

Figure 4. Specific binding of pCC•NC at and near the proposed DNA binding site in probe DNA CC•NC. Here **O** represents specific binding while **X** represents nonspecific binding between each arm and each half-site of dimer binding site. Only the two major contributions for the interaction of pCC•NC/CC•NC complex were shown. As indicated, for the upper case all three arms of pCC•NC have specific interaction with prove DNA CC•NC, but for the lower case two arms of pCC•NC have specific interactions and one arm of pCC•NC has nonspecific interaction with the proposed binding site.


O
O
O
ss
ss
n n n n n n A T G A c g T C A T c g T C A T n n n n n n
X
O
O
ss
ss
n n n n n n A T G A c g T C A T c g T C A T n n n n n n

Part II

Protein Simulation

Chapter 5

Molecular Descriptions for the Active Site of P-450cam

Changmoon Park and William A. Goddard III

Materials and Molecular Simulation Center, Beckman Institute 139-74
Division of Chemistry and Chemical Engineering
California Institute of Technology, Pasadena, CA 91125

Abstract

Cytochrome P-450cam which catalyzes the hydroxylation of camphor at the 5-exo-position has been a model system for the other P-450s in their mechanism and structure. However, because of its high sensitive and complex geometric changes during the reaction, still there are many aspects to be investigated which are hard to extract from experiments. To solve these problems we decided to do molecular dynamics simulations by using Biograf. To do molecular dynamics simulations, first of all, we need such good force constants that can describe the changes in geometry, according to the spin state and oxidation state of Fe as well as the strength and kind of ligands.

We did *ab initio* calculations with a simplified model system to obtain good force constants. Our results show that the size of Fe depends on its spin and oxidation state, the out-of-plane movement of Fe is dominated by the coulombic repulsion between the axial ligand and the porphyrin ring nitrogens, and Fe moves from the porphyrin toward the ligand with greater ligand field strength of the ligand. From the calculations, we find that this simple model system can give many insights into the behaviors of P-450cam, which are difficult to get from expriments.

1. Introduction

1.1. Cytochrome P-450cam

Because of the fact cytochrome P-450 selectively hydroxylates unactivated alkanes, P-450 has been the focus of sustained attention and has played a central role in the study of oxygen activation. For a long time, however, no great advances in the understanding of the P-450 system were made due to the high instability of the cytochrome upon solubilization and the difficulty of separating the integral proteins from one another without denaturation because they are bound to the microsomal membranes (2). This problem was solved when Gunsalus and coworkers successfully purified and crystallized soluble cytochrome P-450 (3), especially P-450cam from *Pseudomonas putida* (*P. putida*). Also, Unger et al. succeeded in cloning the gene for the P-450cam (4), making it possible to do site-directed mutagenesis experiments. Among the P-450s, P-450cam from *P. putida* is by far the best-characterized bacterial system having its own perfect x-ray crystal structure (Figure 1). The P-450cam reaction cycle serves as a functional model for all the P-450s, and recent studies show that there is a strong structural similarity between many P-450s and P-450cam (5). Therefore, by studying P-450cam we can get valuable information about the structure and function of other P-450s.

P. putida uses the relatively inert camphor as the sole source of carbon and energy (10), and the P-450cam system carries out the first step: D-camphor to 5-exo-hydroxycamphor. The P-450cam system consists of three components, NADH-specific FAD-containing

putidaredoxin reductase (Pd R), the iron-sulfur protein putidaredoxin (Pd), and P-450cam.

There are two interesting areas of research for this system. The first area is about the electron transfer between NADH, Pd R, Pd, and P-450 heme iron. The second area lies in the chemical details about the events occurring around the active site during oxidation. Even though all the intermediate species in the P-450 reaction cycle are not identified exactly, the generally accepted mechanism is as in Figure 2 (1).

1.2. Substrate-free and bound state of P-450cam

In the absence of substrate, the iron atom is in a low-spin state having a water molecule or hydroxide ion as its sixth ligand trans to the sulfur ligand, that is six-coordinated, and about five hydrogen-bonded solvent molecules fill the substrate pocket (17). By binding substrate, the iron atom changes into a high-spin state which is five-coordinated (Figure 3) (6,7). There is also a reduction potential shift from -300 to -173 mV (8). By considering the reduction potential of Pd (-196 mV), this shift makes the reduction of substrate-bound P-450cam favorable (9,10). By comparing the x-ray structures of substrate-free and bound P-450cam, however, no significant differences are found between the camphor-free and bound structures (17), while substrate binding results in a large decrease in thermal motion of a few segments which are associated with substrate binding and the substrate-access channel.

1.3. Stereochemistry of P-450cam

From the experiments of Gelb et al., it was found that there was absolutely no stereoselectivity at the hydrogen-abstracting step, while subsequent hydroxylation always occurs at the 5-exo position (13). The latter can be explained by examination of the high-resolution (1.63 Å) x-ray crystal structure (14). The camphor molecule fits well in the active site making a lot of contacts with its neighbors in the site. In that orientation the closest point of camphor molecule to the oxygen atom bounded to the heme iron atom is the 5-exo position. Figure 4 shows amino acids positioned near the active site with substrate camphor and heme. The heme itself provides the largest interactions with the substrate. Val 295 is also important, since it fits well with the *gem* dimethyl groups, C-8 and C-9. In the absence of Val 295, increased motion at the active site results in multiple products (15). Tyr 96 is responsible for the high stereospecificity of P-450cam. This was tested by using the site-directed mutant in which Tyr 96 has been changed to a phenylalanine and the substrate analogous thiocamphor and camphane (16). The results of these experiments suggest that the carbonyl group of camphor plays an important role in the stereochemistry of P-450cam. By using the mutant enzyme with camphor, several hydroxylated products occurred in addition to the 5-exo-hydroxy isomer. The same results were observed when thiocamphor and camphane were used with the wild-type enzyme.

Recently Poulos et al. (31) presented a crystal structure of the P-450cam-CO with camphor (camphor-P-450cam-CO). They showed that camphor in the P-450cam-CO complex moved away from the CO molecule by about 0.8 Å to allow more space for CO. The average temperature factor for camphor atoms increases in the transition from the ferric camphor-P-450cam binary complex to the ferrous camphor-P-450cam-CO ternary complex, indicating that the camphor is not as firmly held in place in the CO complex. If a similar situation occurs in the camphor-P-450cam-O_2 complex, the increased motion of the camphor can explain the complete lack of stereospecificity in the abstraction of the C5 hydrogen atom of camphor.

1.4. Substrate-access channel

From the high-resolution x-ray structures of P-450cam, the active site is placed deep inside of the enzyme and there is no apparent substrate-access channel. However, the most plausible candidate seems to be a small opening above the camphor molecule which consists of three sharp reversals of the polypeptide chain (Figure 5). From the camphor-free structure (17), this opening is smaller than the camphor molecule, but there occurs a significant increase in thermal motion which can allow the camphor to enter the active site. This hypothesis is supported by the fact that substances much bigger than camphor molecule can access the active site. A second possible candidate lies near the heme propionate groups and is filled with ordered water molecules. However, by considering the nonpolar

property of camphor, this hydrophilic route could be used as an exit channel for the product rather than the entrance route.

1.5. Pd-P450cam complex

During the oxygenation, Pd is essential for the product formation in the second electron transfer, the rate-limiting step. In the absence of Pd, the reduced P-450cam-O_2 is readily isolated, but gradually degenerates without formation of product. However, in the presence of either reduced or oxidized Pd, the reduced P-450cam-O_2 undergoes oxygenation readily. With excess reduced Pd, the amount of product is equal to the amount of reduced P-450cam available, while one-half the total P-450cam available is reduced when we used oxidized Pd (18). In both cases, a complex between Pd and P-450cam is required for the product formation (19). Recent experiments (IR and resonance Raman spectra, and ^{15}N NMR measurements (30)) show that the binding of Pd weakens the bond between C-O for the case of P-450cam-CO (or C-N for the case of P-450cam-CN) by moving the thiolate ligand toward the Fe, probably making it favorable to cleave the O-O bond in the P-450cam-O_2 system (20).

1.6. Proposal and research work

Although a lot of information from x-ray structures of P-450cam is now available, there are still many aspects of the system to be investigated. Also, x-ray data cannot generally give dynamic information even though they give valuable information about the static

structure. Moreover, because of the complexity and instability of P-450's, there are many restrictions in experiments which make it difficult to get certain types of information from experiments. To overcome these difficulties, we decided to do molecular dynamics simulations by using Biograf. From the previous statements, however, it becomes obvious that for the molecular dynamic simulation of cytochrome P-450cam, the essential prerequisite is a high quality set of force constants for the P-450cam system, particularly around the active site. This high quality description is necessary for describing the highly sensitive behaviors of P-450cam as a function of environmental changes. For example, the description must take into account the change of oxidation and spin state of Fe as well as the out of plane movement of Fe observed as a function of the number and nature of ligands. To obtain the best force constants we carried out ab initio quantum chemical calculations up to the level of GVB (Generalized Valence Bond) with an appropriately chosen model complex.

2. Calculations

2.1. Introduction

Our current studies focus on obtaining detailed information on how bond distances, bond angles, and force constants depend upon oxidation state, spin state, and ligands. The objectives here are to obtain:

(a) detailed geometries and energy information that can be used in conjunction with current experimental data to elucidate mechanistic issues in these systems; and

(b) force fields for the active site region that can be used in molecular dynamics simulations of the full P-450 enzyme, including substrate and reaction intermediates.

We use ab initio quantum chemical methods [Hartree Fock (HF), Generalized valance bond (GVB)] in conjunction with a model for heme of P-450cam to study how changes in geometry and spin state depend on the nature of the axial ligands. In order to provide essential detail on the role of the heme in the mechanisms of P-450, we developed a computational model for the active site of P-450cam. We find that this model leads to useful information about the role of

(i) the metal oxidation state (Fe^{+3} or Fe^{+2}),

(ii) the spin state (low spin, high spin, intermediate spin),

(iii) the porphyrin oxidation state (neutral or cation),

(iv) the axial ligand (Cl^- used here to represent cystein), and

(v) the sixth ligand (oxygen or H_2O)

on the geometry and energetics of these systems.

2.2. Model and Calculations

2.2.1. Basis sets

The basis sets used for this calculation are as follows: the Fe and Cl were described with ab initio core effective potentials to represent the Ne core electrons (1s2s2p). The remaining seven electrons of Cl

and 16 electrons of Fe were treated explicitly using a valance double zeta basis for Fe and a triple zeta basis for Cl.(21) All electrons of the H, N, and O atoms were treated explicitly with a valence double zeta basis set(22). In addition, d polarization functions were used for Cl and O (z_{dO} = 0.85, z_{dCl} = 0.6).

2.2.2. The model

Following Olafson and Goddard (OG)(23), we used four NH_2 groups (along the +-x and +-y axes) to represent the porphyrin, as illustrated in Figure 6a. The following constraints were used to ensure that this model best represents a porphyrin system:

(1) The H-N-H angle was taken as 120° and the four NH_2 groups were constrained to have D_{4h} symmetry. The distance of Fe from the N plane is denoted R_{Fe-Np}, where Np denotes the midpoint of the four N atoms.

(2) For five-coordinate systems, the Cl axial ligand was placed a distance of R_{Cl-Np} from the midpoint of the nitrogen plane. In these calculations we optimized both R_{Cl-Np} and R_{Fe-Np} and carried out a number of calculations with either R_{Cl-Fe}, R_{Cl-Np} or R_{Fe-Np} fixed.

(3) For six-coordinate systems, the O or H_2O sixth ligand was placed a distance of R_{O-Np} from the nitrogen plane. The H_2O was oriented in the xz plane. This leads to C_{4v} symmetry for O ligand and C_{2v} symmetry for H_2O. The geometry of H_2O was kept fixed (R_{HO} = 0.958 Å, Θ_{HOH} = 104.5°). We optimized all three parameters, R_{Cl-Fe}, R_{Fe-Np}, and R_{Np-O}, independently.

(4) All the calculations for the five- and six-coordinate systems were carried out with a fixed porphyrin hole radius of 2.0044 Å (based on recent high resolution data (14)).

(5) For ferrous Fe, we find the iron to have six electrons in d orbitals (formal oxidation state Fe^{2+}) leading to a formal charge of -2 for the four NH_2 groups. Thus the orbitals of the NH_2 group are as sketched in Figure 6b where eight electrons are involved in s lone pairs on the N pointed at the Fe and six electrons are in N π orbitals perpendicular to the plane system. This leads to two singly-occupied N π orbitals for the model system (Figure 6a), whereas in the porphyrin these unpaired electrons on the N π orbital would be spin paired to the carbon π orbitals to form bonds. In order to prevent the N orbitals from spin-pairing with each other or with unpaired electrons on the Fe, we describe these six electrons in terms of four self-consistent orbitals using Fock operators derived from the energy expression

$$E_{Np} = S\ 2f_i < i\ l\ h\ l\ i > + S\ (a_{ij}J_{ij} + b_{ij}K_{ij}) \qquad (a)$$

where

$$f_i = 3/4, \quad a_{ij} = 13/12, \quad b_{ij} = -7/12. \qquad (b)$$

This corresponds to averaging over all six *triplet configurations* of Figure 6. The interaction of the N π electrons with all other electrons is described using the energy expression

$$E_{Np,s} = S\ S\ f_i\ f_j\ [4J_{ij} - 2K_{ij}], \qquad (c)$$

where the sum j is over all the orbitals except the four N π orbitals. This ensures the same type of interaction as if the full porphyrin were present. The remaining energy expression is the normal one for the various HF and GVB calculations.

(6) To mimic porphyrin oxidized systems we removed one electron from the N π orbitals, leaving a total of five electrons distributed over four orbitals. The coefficients (b) for the energy expression (a) were thus modified to

$$f_i = 5/8, \quad a_{ij} = 0.75, \quad b_{ij} = -0.5. \qquad (d)$$

2.2.3. Calculations

We focused on the three spin states with configurations in (e) and (f) for the ferric and ferrous oxidation states of Fe:

Ferric Fe

S = sextet (S = 5/2) = $(xz)^1$ $(xy)^1$ $(yz)^1$ $(z^2)^1$ $(x^2\text{-}y^2)^1$

Q = quartet (S = 3/2) = $(xz)^1$ $(xy)^2$ $(yz)^1$ $(z^2)^1$ (e)

D = doublet (S =1/2) = $(xz)^2$ $(xy)^1$ $(yz)^2$

Ferrous Fe

q = quintet (S = 2) = $(xz)^1$ $(xy)^1$ $(yz)^2$ $(z^2)^1$ $(x^2\text{-}y^2)^1$

t = triplet (S =1) = $(xz)^1$ $(xy)^2$ $(yz)^2$ $(z^2)^1$ (f)

s = singlet (S = 0) = $(xz)^2$ $(xy)^2$ $(yz)^2$

and as indicated in Figure 7. These states are lowest lying states for most of the four-, five-, and six-coordinate systems considered herein. When it is necessary to discuss higher states of these spins, we will

denote the Q and D in (e) as Q_{xy} and D_{xy} and the q and t in (f) as q_{yz} and t_{xz}. Other low lying states generally involve redistributing electrons between xy, xz and yz leading, for example, to Q_{yz}, D_{xz}, q_{xy}, and t_{xy}.

2.3. Results

We carried out calculations for a number of molecular geometries of the four-, five - and six-coordinate systems with ferric or ferrous Fe and with H_2O or oxygen atom at the sixth site (see Table 1,2,3 and 4).

2.3.1. Four - coordinate Systems

From the calculation of the four-coordinate system, we find R_{Fe-N} = 2.15 Å for the q state which is close to the value 2.19 Å estimated by Hoard (24). We find R_{Fe-N} = 2.05 Å for the t state (Figure 8a), 2.03 Å for the S state, and 1.95 Å for the Q state (Figure 8b). Thus removing the electron from the (antibonding) $d_{x^2-y^2}$ orbital decreases the size by 0.10 Å (ferrous) or 0.08 Å (ferric). Considering the same occupancy of $d_{x^2-y^2}$ for ferrous and ferric, we see that for high-spin states the size of ferrous Fe is about 0.12 Å larger than that of ferric Fe while for low spin the difference is 0.10 Å. This is due to the higher charge of Fe^{+3}, which (i) leads to slightly smaller d orbitals and (ii) polarizes the N lone-pair orbitals (Figure 6b) to mix in more Fe character. As a result, the in-plane force constant for the oxidized state is about 60% higher than that of the ferrous state. Independent of spin and oxidation state, the optimal position of Fe is always in the plane (Figure 9). This is

because the attraction between Fe and the porphyrin nitrogens is larger than the antibonding interactions between the Fe $dx^2\text{-}y^2$ orbital and the nitrogen lone pairs (In agreement with OG(23)). For ferrous Fe, we find high-spin (q) to be the ground state for $R_{Fe\text{-}N} > 1.84$ Å, with the intermediate-spin (t) stable for smaller distances (see Figure 8a). For ferric Fe, the high-spin (S) is stable for $R_{Fe\text{-}N} > 1.83$ Å. The strong dependence of Fe size on spin and oxidation states allows the axial ligands and level of oxidation to control the geometry and spin of five- and six-coordinate systems, which is the focus in later sections.

2.3.2. Five - coordinate Systems

For five-coordinate system, we used Cl^- to represent cystein sulfur (P-450cam has a cysteine sulfur as a fifth ligand to heme Fe, and other P-450's probably also have such a ligand.). We did calculations on five-coordinate system under three different restrictions, (i) $R_{Fe\text{-}Cl} = 2.2$ Å (Figure 10), (ii) $R_{Np\text{-}Cl} = 2.63$ Å (Figure 11), and (iii) $R_{Fe\text{-}Np} = 0.43$ Å (Figure 12) (we also fixed $R_{N\text{-}Np} = 2.0044$ Å for all these cases.) For the case of $R_{Fe\text{-}Cl} = 2.2$ Å, the ground state is high spin (q and S) with $R_{Fe\text{-}Np}$ of 0.69 Å and 0.44 Å for ferrous and ferric Fe, respectively. This is in good agreement with the X-ray data that finds 0.43 Å for the ferric state(14). We also find that Fe is 0.46 Å out of the plane for t, 0.33 for Q, 0.39 for s , and 0.20 Å for D. That is, the forces exerted by the porphyrin to prevent out-of-plane motion of Fe are in the following order:

$$D > Q > s > S \sim t > q \qquad \text{(figure 10 and 11).}$$

Thus it is easiest to move Fe out of the plane in the q state and hardest in the D state. These results are expected on the basis that occupation of $d_{x^2-y^2}$ leads to a large decrease of the out-of-plane force constant, occupation of d_{z^2} leads to a small decrease, and oxidation of Fe leads to an increase. This variation in preference for the planarity plays a key role on the spin equilibrium for the active site of substrate-free P-450. We also find the large out-of-plane displacement of Fe is caused by the nonbonded interactions between the axial ligand (Cl) and the porphyrin nitrogens. This is in agreement with the original suggestion by OG (23). However, we find that it is the coulombic repulsion between the fifth ligand and porphyrin ring nitrogens that dominates(26), as illustrated by Figure 13 [which shows Cl- and Cl interacting with a charged $(NH_2)_4$ unit without Fe].

2.3.3. Six-coordinate Systems - H_2O as Sixth Ligand

There is abundant evidence that the resting P-450 system has water (29) in the sixth coordination site (probably six H_2O molecules) and that the spin state is low-spin (S=1/2) (6), as described in the introduction. Using H_2O in the six-coordinate system, we find two low lying states (D_{xz} and D_{xy} essentially degenerate). For D_{xz} (or D_{xy}) the optimum position of Fe is 0.12 (or 0.16 Å) out of the plane (toward Cl) and the distance between Fe and oxygen of water is 2.26 Å (or 2.26 Å) (Figure 14). This is in excellent agreement with the high resolution x-ray experiments (17) on the H_2O ligand case which give R_{Fe-Np} = 0.30 Å and R_{Fe-O} = 2.28 Å. Since x-ray crystallography is more

sensitive to the mean plane of the porphyrin rather than the N plane, the actual R_{Fe-Np} may be closer to the theoretical value of 0.14 Å than 0.30 Å. For the high-spin state (S), the repulsion between the electron in the Fe d_{z^2} orbital (occupied in high-spin but not in low-spin) and the lone pair orbitals of water leads to no bonding in the sixth coordination site (Table 3). As expected from its larger size, we find that high-spin ferric iron is displaced 0.35 Å out of the plane (toward Cl), while low-spin ferric iron is 0.12 Å out of plane, as indicated in Figure 14 (with R_{Cl-Np} fixed at 2.63 Å). Even though with our model the high spin state (S) is calculated to be more stable than the low-spin state (D) which is not in agreement with experiment, we calculate no net bonding to the water in the sixth coordinate site for the high-spin state (S). But the interaction between Fe and water that leads to a bond strength of 8 kcal for the low-spin state (D), which can force H_2O to remain bonded to the Fe (by the protein and by interactions within the water cluster (17)). This might stabilize the low-spin state. This can be tested after the force fields from these calculations are incorporated into full molecular dynamics calculations on the protein.

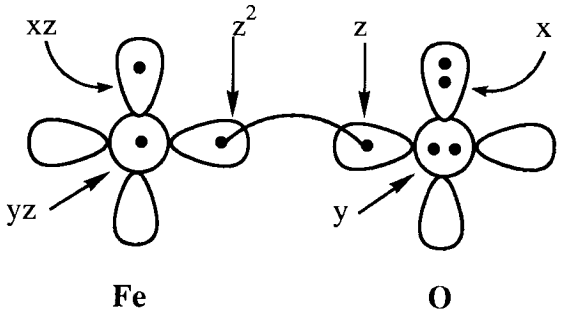
2.3.4. Six-coordinate Systems - Oxygen as Sixth Ligand

The putative reaction intermediate in P-450 catalyzed by hydroxylase is an Fe-O species (formally either Fe^{IV} $O^{=}$ or Fe^{V} $O^{=}$). Among several states (which have equal number of electron with $Fe^{IV}O^{=}$ or $Fe^{V}O^{=}$), the ground state is best visualized as

$$Cl^{-}P^{=}\ Fe_S^{3+}O^{-}$$

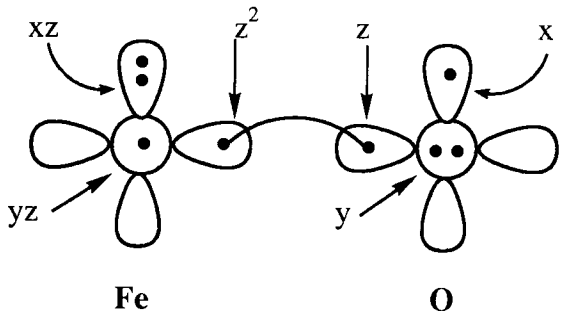
where the Fe^{3+} is in the S state. For the unoxidized porphyrin ($P^{=}$) we obtain R_{Fe-O} = 1.84 Å and R_{Fe-Np} = 0.10 Å (Figure 15) while for the oxidized porphyrin (P^{-}) we obtain R_{Fe-O} = 1.81 Å, and R_{Fe-Np} = 0.06 Å (Figure 16). Thus the oxygen pulls the Fe away from the Cl, shifting the equilibrium position from -0.35 to +0.10 Å for $P^{=}$ and from -0.30 to +0.06 Å for P^{-} systems, respectively. For P^{-} this illustrates that oxygen is a stronger ligand than Cl and explains why strong ligands denature P-450 into P-420. By binding a strong ligand in the absence of substrate, the Fe moves in the direction of the sixth strong ligand, thereby weakening the bond between Fe and the fifth ligand and allowing the fifth ligand to move more freely than before. Consequently the fifth ligand might make a new bond with neighboring atoms. Because the position of the soret band at about 420 nm for the reduced cytochrome P-450-CO is typical of heme proteins lacking a cysteinyl axial ligand (27), the formation of cytochrome P-420 is thought to involve the displacement of thiolate ligand. Mercurials (which are believed to attack sulfhydryl groups) convert P-450 to P-420(3b), supporting the above views.

For the monooxidized porphyrin system, the ground state is S=2 and the bonding can be pictured as in (g) with spin pairing of the O_{pz} singly-occupied orbital of O^{-} to the Fe d_z^2 singly-occupied orbital of the S state (high-spin ferric). This state would be expected to be unreactive for homolytic processes since the oxygen orbitals are strongly spin-paired.



(g)

There is abundant evidence that the oxygenation of P-450's leads to a p-radical cation (28). To determine the possible roles of this species, we calculated states with one electron extracted from the p orbitals of the pseudoporphyrin. Oxidizing the porphyrin leads to a dramatic change in the orbitals. The positive charge in the porphyrin p system stabilizes the Fe d_{xz} and d_{yz} orbitals sufficiently to transfer one of the electrons from O to Fe leading to the state (h). The result is two three electron FeO p bonds much like in free O_2 molecule. The Fe-O charge transfer reduces 1-3 coulombic repulsion (oxygen-nitrogen), allowing the oxygen to move more easily into the plane (Figure 16 and 17 and Table 4) from +1.90 to +1.80. This shortens and strengthens the Fe-O bond for the oxidized porphyrin, leading to R_{Fe-O} = 1.86 Å, and R_{Fe-Np} = -0.05 Å, allowing radical character on the oxygen [see (h)].



(h)

This radical character on the oxygen in (h) leads to a much more reactive oxygen while the ferrous iron has a smaller force constant for the out-of-plane movement (than ferric) which allows the oxygen to position itself appropriately for reactions with substrate (Figure 17). For systems with two-center, three-electron bonds, it is important to properly include the resonance of various states. We have to do GVB-CI calculations for these states to include these resonance effects. Without these resonance effects, the Fe^{+3}-O^{-} state (g) would be lower than the Fe^{+2}=O state (h) by 5.4 kcal/mol.

2.4. Summary

From calculations on model systems, we conclude:

(i) The size of Fe depends on its spin and oxidation state, changing by as much as 0.2 Å. This plays a key role in the process of oxygenation.

(ii) The out-of-plane movement of Fe is dominated by the coulombic repulsion between the axial ligand and the porphyrin ring nitrogens.

(iii) Fe moves from the porphyrin toward the ligand with greater ligand field strength of the ligand.

(iv) For unoxidized porphyrin ($P^{=}$) the Fe-O bond is unreactive but oxidation of the porphyrin π-system (P^{-}), leads to an Fe-O bond that

should be quite active for oxygenation. This indirect control of Fe-O activity through modifying the porphyrin π systems provide the protein with an indirect control mechanism.

(v) An appropriately designed model complex can provide a great deal of useful information about the active site of a complex system, providing insight that is difficult to extract from experiments.

3. Future research

We have already done the calculation on the model system of P-450 active site to get good force constants for the molecular dynamics simulation of the full protein and incorporated the force constant from the potential curve of our calculation into Biograf. Therefore future plans are the following:

(i) Do molecular dynamics simulation for the full protein. In this process we are going to investigate the role of various residues, especially in the active site.

(ii) Do site-directed mutagenesis and predict the modification of the substrate for the modified protein.

(iii) Find the substrate access channel and exit channel. This process can also give some insight into the mechanism of the reaction.

(iv) We are going to design modified enzymes which are specific to the new substrates, which can be used for other purposes.

(v) We will design non-biological catalysts that can do equally selective catalysis.

4. References

1. (a) Gunsalus, J. T., Adv. Inorg. Biochem. 1979, **1** 119-145. (b) White, R. E.; Coon, M. J., Annu. Rev. Biochem. 1890, **49** 315-356.
2. Omura, T.; Sato, R., J. Biol. Chem. 1964, **239** 2379-2385.
3. (a) Lindnmayer, A.; Smith, L., Biochim. Biophys. Acta. 1964, **93** 445-461. (b) Das, K.; Katagiri, M.; Yu, G. A.; Erbes, D. L.; Gunsalus, I. C., Biochem. Biophys. Res. Commu. 1970, **40** 1431-1436.
4. Unger, B. P.; Gunsalus, I. C.; Sligar, ß. G., J. Biol. Chem. 1986, **261** 1158-1163.
5. (a) Nelson, D. R.; Strobel, H. W., J. Biol. Chem. 1988, **262** 6038-6050. (b) Nelson, D. R.; Strobel, H. W., Biochemistry 1989, **28** 6656-660.
6. Sharrock, M.; Debrunner, P. G.; Schultz, C.; Lipscomb, G. D.; Marshall, V.; Gunsalus, I. C., Biochim. Biophys. Acta. 1976, **420** 8-26.
7. (a) Philson, S. B., Ph D Thesis 1976, University of Illinois, Urbana, IL. (b) Sligar, S. G., Biochemistry 1976, **15** 5399-5406.
8. Sligar, S. G.; Cinty, D. L.; Gibson, G. G., Schenkman, Biochem. Biophys. Res. Commun. 1979, **90** 925-932.
9. Sligar, S. G.; Gunsalus, I. C., Proc. Natl. Acad. Sci. U.S.A. **73** 1078-1082.

10. Sligar, S. G.; Murry, R. I., "Cytochrome P-450: Structure, Mechanism, and Biochemistry" (Oritiz de Montellano, P. R., ed.) 1986, pp. 429-503, Plenum, New York.

11. Lewis, B. A.; Sligar, S. G., J. Biol. Chem. 1983, **258** 3591-3601.

12. Groves, J. T.; Watanabe, Y., J. Am. Chem. Soc. 1988, **110** 8443-8452.

13. Glb, M. H.; Heinbrook, D. C.; Malkonen, P.; Sligar, S. G., Biochemistry 1982, **21** 370-377.

14. (a) Poulos, T. L.; Finzel, b. C.; Gunsalus, I. C.; Wagner, G, C.; Kraut, J., J. Biol. Chem. 1985, **260** 16122-16130. (b) Poulos, T. L.; Frinzel, B.; Howard, A. J., J. Mol. Biol. 1987, **195** 687-700.

15. (a) Atkins, W. M.; Sligar, S. G., J. Am. Chem. Soc. 1989, **111** 2715-2717. (b) Iami, M.; Shimada, H.; Watanabe, Y.; Matsushima, H. Y.; Makino, R.; Koga, H.; Hiriuchi, T.; Ishimura, Y., Proc. Natl. Acad. Sci. U.S.A. 1989, **86** 7823-7827.

16. Atkins, A. M.; Sigar, S. G., J. Biol. Chem. 1988, **263** 18842-18849.

17. Poulos, T. L.; Finzel, B. C.; Howard, A. J., Biochemistry 1986, **25** 5314-5322.

18. (a) Ishimura, Y.; Ullrich, V; Peterson, J. A., Biochem. Biophys. Res. Commun. **42** 140-146. (b) Tyson, C. A.; Lipscomb, J. D.; Gunsalus, I. C., J. Biol. Chem. 1972, **247** 5777-5784.

19. (a) Lipscomb, J. D.; Sligar, S. G.; Namtvedt, M. J.; Gunsalus, I. C., J. Biol. Chem. 1976, **251** 1116-1124. (b) Gunsalus. I. C.; Sligar, S. G.; Debrunner, P. G., Biochem. Soc. Trans. 1975, **3** 821-835.

20. (a) Eisenstein, L.; Debey, P.; Douzou, P., Biochem. Biophys. Res. Commun. 1977, **77** 1377-1383. (b) Sligar, S. G.; Debrunner, D. G.; Lipscomb, J. D.; Namtvedt, M. J.; Gunsalus, I. C., Proc. Natl. Acad. Sci. U.S.A. 1974, **71** 3906-3910.

21. (a) Hay, P. J.; Wadt, W. R., **Fe**. (b) Hay, P. J.; Wadt, W. R., **Cl**.

22. (a) Dunning / Huzinaga. (b) Huzinaga.

23. (a) Olafson, B. D.; Goddard III, W. A., Proc. Natl. Acad. Sci. U.S.A. 1977, **74** 1315-1319. (b) Olafson, B. D., Ph. D. Thesis, California Institute of Technology, Pasadena, CA.

24. Hoard, J. L., Science 1971, **174** 1295-1302.

25. (a) Dawson, J. H.; Eble, K. S., Adv. Inorg. Bioinorg. Mech. 1986, 4 1-64. (b) Dawson, J. H.; Kau. L-S.; Penner-hahn, J. E.; Sono, M.; Bruce, G. S.; Hager, L. P.; Hodgson, K. O., J. Am. Chem. Soc. 1986, **108** 8114-8116.

26. Saito, M.; Kashiwaki, H., Inter. J. Quan. Chem. (Quantum Chemical Symposium) 1987. **21** 661-668.

27. Libscomb, J. D., Biochemistry 1980, **19** 3590-3597.

28. (a) Rutter, R.; Valentime, M.; Hendrich, M. P.; Hager, L. P.; Dbrunner, P. G., Biochemistry 1983, **22** 4769-4774. (b) Penner-Hahn, J, E.; McMurry, J. J., J. Biol. Chem. Comm. 1983, **258** 12761-12764. (c) Roberts, J. E.; Hoffman, B. M., J. Biol. Chem. Comm. 1981, 256 2118-2121. (d) La Mar, G. N.; de Rapp, J. S., J. Biol. Chem. 1981, **256** 237-243. (e) Groves, J. T.; Nomo, T. E., J. Am. Chem. Soc. 1983, **105** 5786-5791.

29. (a) Kumaki, K.; Sato, M., J. Biol. Chem. 1978, **253** 1048-1058. (b) White, R. E.; Coon, M., J. Ann. Rev. Biochem., 1980, **49** 315-356.

30. Shiro, Y.; Iizuka, T.; Makino, R.; Ishimura, Y.; Morishima, I., J. Am. Chem. Soc. 1989, **111** 7707-7711.

31. Raag, R.; Pouls, T. L., Biochemistry 1989, **28** 7586-7592.

Table 1. Energetics and Force constants for Four-coordinate Complexes.

	Planar R_{FeNp}(Å)	K_{FeNp}[a] (kcal/mol)/Å^2	Energy[e] kcal/mol	Out of Plane[b] $K_\perp$[d] kcal/mol)/Å^2	K_{NFeN}[c] (kcal/mol)/rad^2
Ferrous					
q(q_{yz})	2.150	271.107	0	121.909	61.779
t(t_{xz})	2.054	296.659	42.333	228.933	116.093
s	2.058	298.551	75.786	231.250	116.907
q_{xy}	2.139	263.672	3.926	111.450	56.507
t_{xy}	2.062	300.805	58.228	246.387	124.942
t_{z2}	2.053	194.745	64.398	_	_
Ferric					
S	2.026	446.048	0	263.486	133.573
Q(Q_{xy})	1.947	500.113	29.750	368.336	186.785
D(D_{xy})	1.956	497.082	85.845	449.898	228.191
Q_{yz}	1.957	461.525	34.058	402.467	204.112
Q_{z}^2	1.953	478.169	35.034	_	_
D_{xz}	1.947	503.518	96.981	415.469	210.942
D_{z2}	1.964	460.178	129.713	_	_

[a] For each Fe-N bond

b With porphyrin hole fixed at R_{NNp} = 2.0044 Å

[c] $E = E_o + 1/2 \Sigma_{K>L} K_{NFeN}(\theta_{KL}-\theta^e_{KL})^2$, where $\theta^e_{KL} = \pi$ for two bonds and $\theta_{KL} = \pi/2$ for our bonds. The two linear bonds domnate.

[d] $E = E_o + 1/2 K_\perp R_\perp^2$, where $R_\perp$ is the R_{FeNp} distance.

[e] Relative energy with each state at the optimum value of R_{FeNp}.

Table 2. Energetics and Force constants for Five-coordinate Complexes.

In all cases $R_{N\text{-}Np}$ = 2.0044 Å

State	Constraint: R_{FeCl} = 2.20 Å				Constraint: R_{ClNp}= 2.63 Å				Constraint: R_{FeNp}= 0.43 Å		
	R_{FeCl} Å	Ke (kcal/mol)/Å^2	E_{min}[a] (kcal/mol)	$E_{barrier}$[b] (kcal/mol)	R_{FeCl} Å	Ke (kcal/mol)/Å^2	E_{min}[a] (kcal/mol)	$E_{barrier}$[b] (kcal/mol)	R_{FeCl} Å	Ke (kcal/mol)/Å^2	E_{min}[a] (kcal/mol)
Ferrous											
$q(q_{xy})$	0.692	189.453	0	44.304	0.324	293.959	0	11.243	2.650	70.206	0
t_{xz}	0.461	207.312	52.441	25.379	0.200	288.642	39.686	5.204	2.704	51.665	44.995
s	0.394	224.534	70.862	20.403	0.240	307.584	58.254	7.811	2.538	88.319	70.244
q_{yz}	0.653	185.499	7.748	46.219	0.285	260.202	4.284	8.528	2.696	59.600	4.225
t_{xy}	0.427	146.527	77.574	45.923	0.167	284.892	61.383	3.547	2.735	60.784	65.053
Ferric											
S	0.437	315.739	0	30.775	0.354	436.633	0	20.681	2.329	251.938	0
$Q(Q_{xy})$	0.326	370.550	49.545	20.915	0.257	399.396	46.510	12.277	2.353	228.789	50.613
$D(D_{xz})$	0.198	342.155	84.664	6.938	0.116	154.074	84.877	0.837	2.218	253.193	95.543
Q_{yz}	0.294	376.993	53.902	32.440	0.218	410.349	59.578	9.186	2.379	202.314	64.870
D_{xy}	0.207	394.184	90.030	14.134	0.225	428.204	91.952	9.881	2.279	230.847	101.007

[a] Energy at minimum (relative to lowest energy of each oxidation state).

[b] Barrier to push Fe into N plane, given the constraints.

Table 3. Energetics for six-coordinate ferric system with Cl as fifth ligand and H_2O as sixth ligand. $R_{Cl\text{-}Np}$ = 2.63 Å; $R_{N\text{-}Np}$ = 2.0044 Å

State	R^{Opt}_{FeNp}[a] Å	R^{Opt}_{FeO} Å	K_{FeO}[d] (kcal/mol)/Å^2	D_{FeO} (kcal/mol)	E (kcal/mol)
Dxy	-0.130	2.284	67.847	8.526	-0.935
Dxz	-0.156	2.260	56.290	0.820	0.000
S	-0.354	(2.284)[b]		[c]	-87.028

[a] Positive is toward O.

[b] Not bound.

[c] H_2O is not bound to the S state, the energy was calculated with the H_2O at the same location (with respect to the porphyrin) as for D_{xy} (R_{NpO} = 2.284 - 0.130 = 2.154 Å); leading to R_{FeO} = 2.154 + 0.354 = 2.505 Å.

[d] At R_{FeNp} = -0.15 Å.

Table 4. Energetics for six-coordinate ferric systems with Cl as fifth ligand and oxygen as sixth ligand. $R_{Cl\text{-}Np}$ = 2.63 Å; $R_{N\text{-}Np}$ = 2.0044 Å

Porphyrin Oxidized	State	Net Charge Distribution	R^{Opt}_{FeNp}[a] Å	R^{Opt}_{FeO}[b] Å	K_{FeO} (kcal/mol)/Å^2	D_{FeO} (kcal/mol)	E^c (kcal/mol)
No	S	$Cl^-P{=}Fe^{3+} - O^-$	0.140	1.843	286.474	12.837	0
	q	$Cl^-P{=}Fe^{2+} = O$					9.693
	Q	$Cl^-P{=}Fe^{3+} - O^-$					23.647
	t	$Cl^-P{=}Fe^{2+} = O$					18.360
Yes	S	$Cl^-P{-}Fe^{3+} - O^-$	0.089	1.806	320.867		30.268
	q	$Cl^-P{-}Fe^{2+} = O$	-0.054	1.858	367.102		35.673
	Q	$Cl^-P{-}Fe^{3+} - O^-$					55.673
	t	$Cl^-P{-}Fe^{2+} = O$					44.387

[a] Positive is toward O.

[b] Calculated using R_{FeNp} = 0.0Å; R_{FeO} = 1.8 Å.

[c] Calculated using R_{FeNp} = 0.0Å.

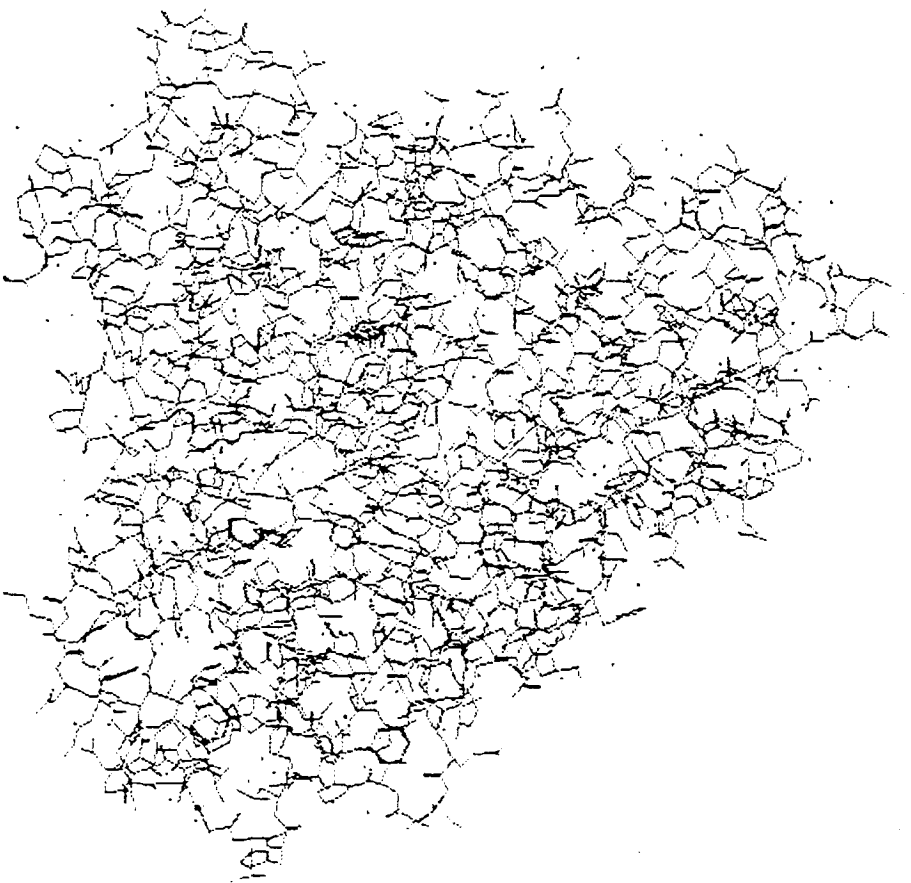


Figure 1. X-ray crystal structure of P-450cam

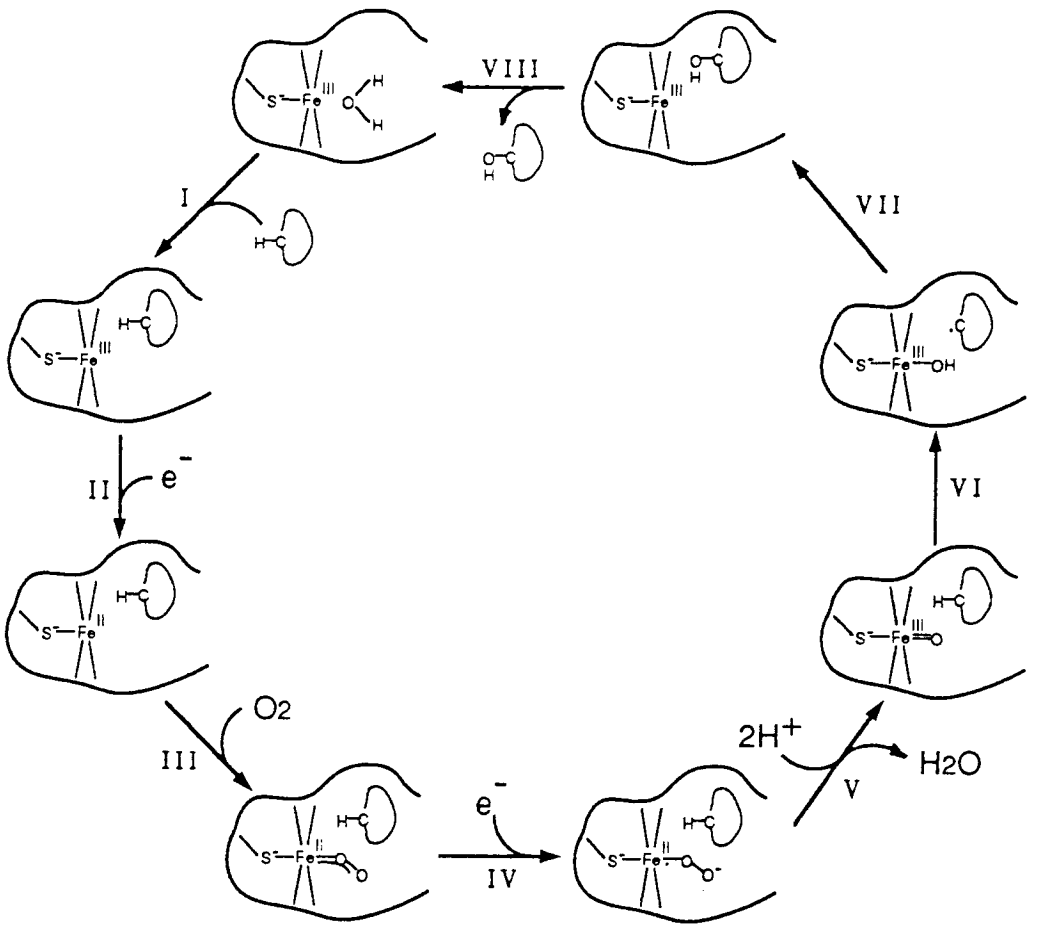


Figure 2. The reaction mechanism of cytochrome P-450cam
When there are no substrate bound in the active site Fe is in a low-spin ferric state and a cluster of waters occupies the active site.
I. Substrate binding changes Fe into a high-spin ferric state.
II. Fe becomes high-spin ferrous state by one electron reduction.
III. Dioxygen binds to Fe.
IV. A second, one electron reduction.
V. H_2O is generated consuming two protons.
The details of steps from VI to VII are unknown.
VIII. Generation of a oxygenized product.

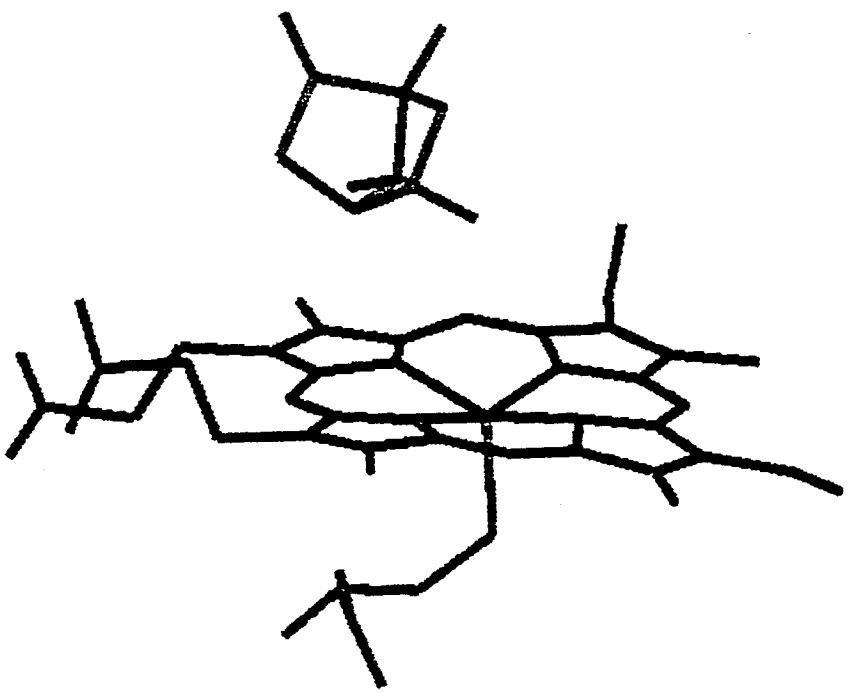


Figure 3. A model for the active site of camphor bound P-450cam. Fe is at high-spin ferric state and five-coordinated.

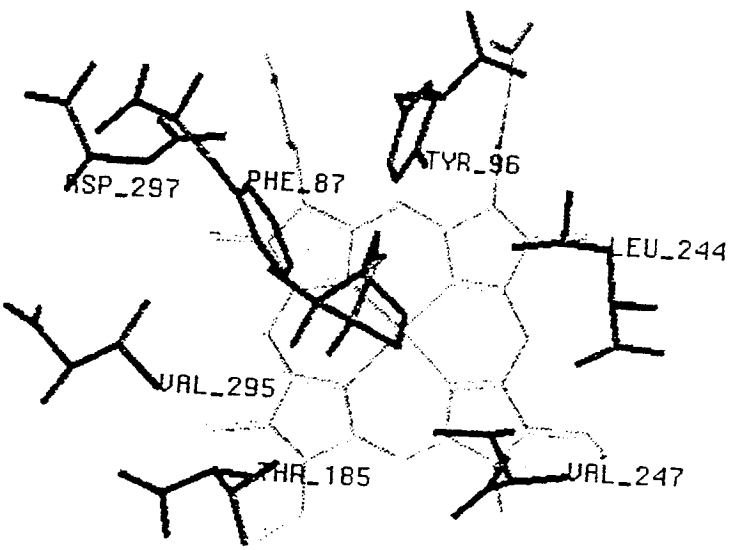


Figure 4. A model for the active site showing all amino acids playing important roles to keep the substrate in correct orientation.

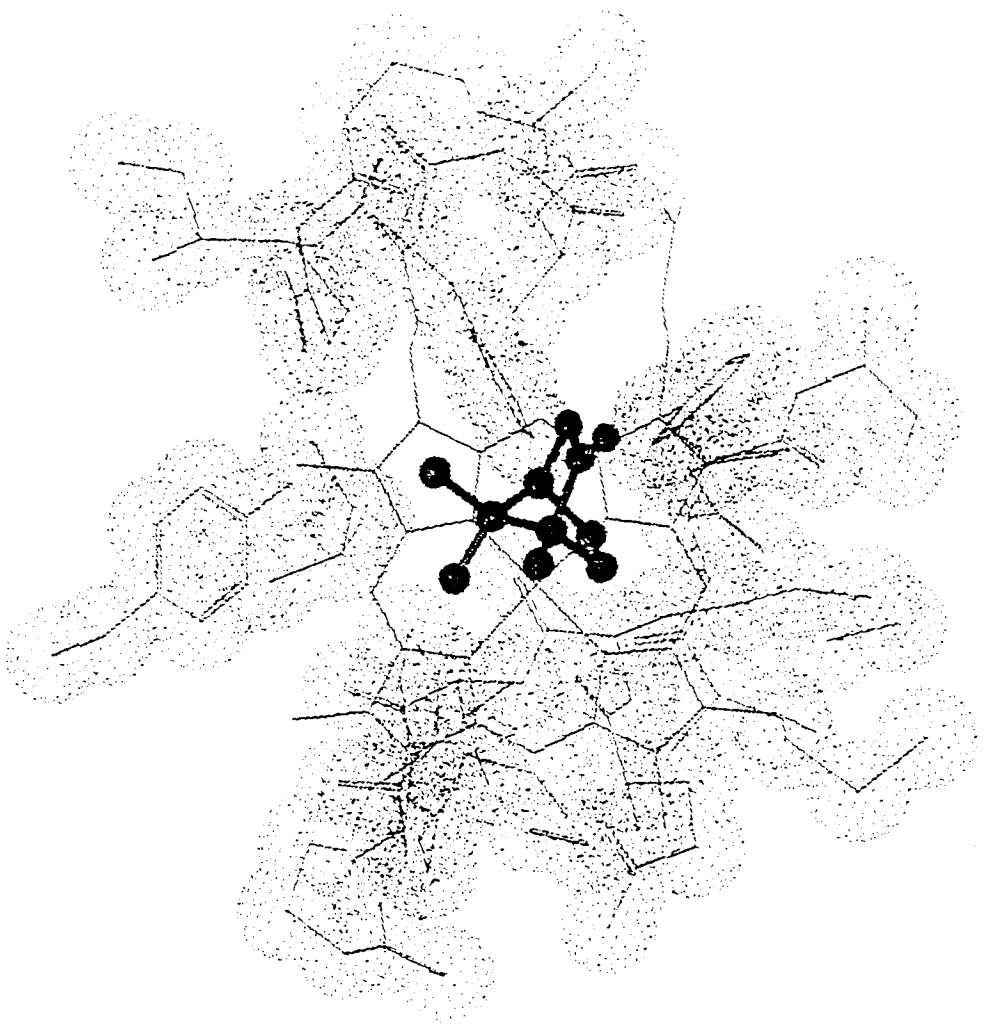


Figure 5. The most plausible substrate access channel of P-450cam with van der Waals surface. The substrate camphor is shown through the channel.

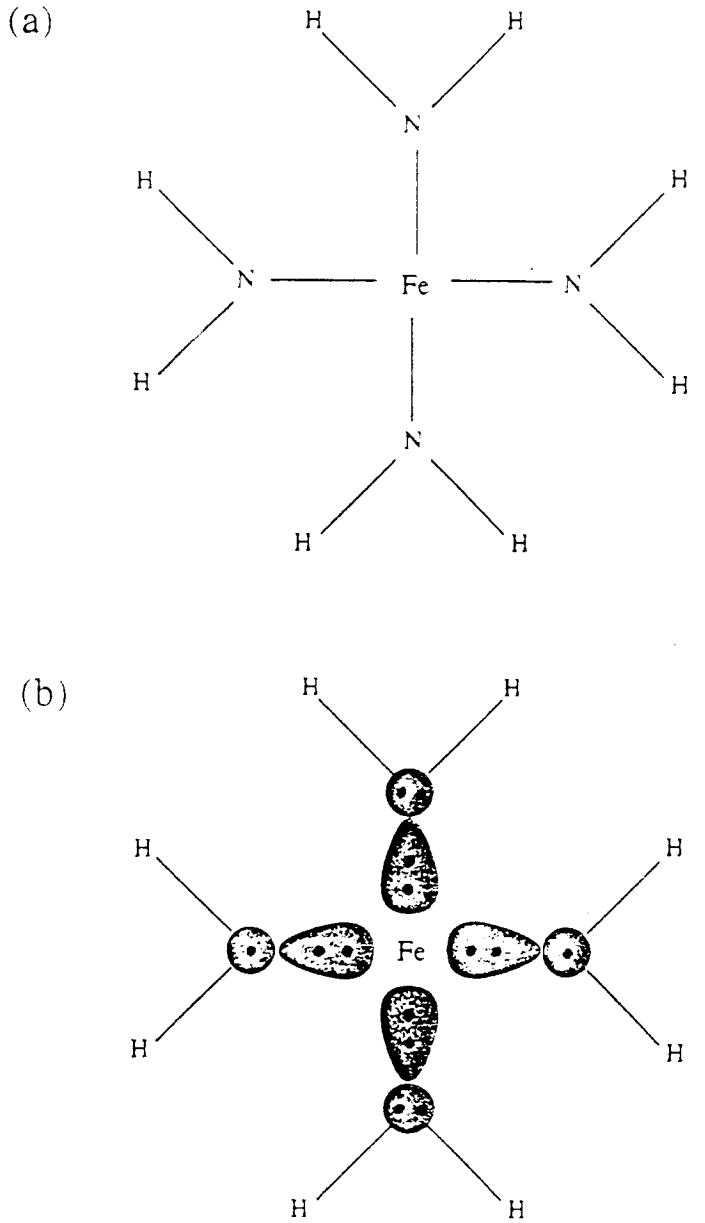


Figure 6. (a) The $(NH_2)_4$ model for porphyrin and (b) A schematic representation of the nonbonding orbitals.

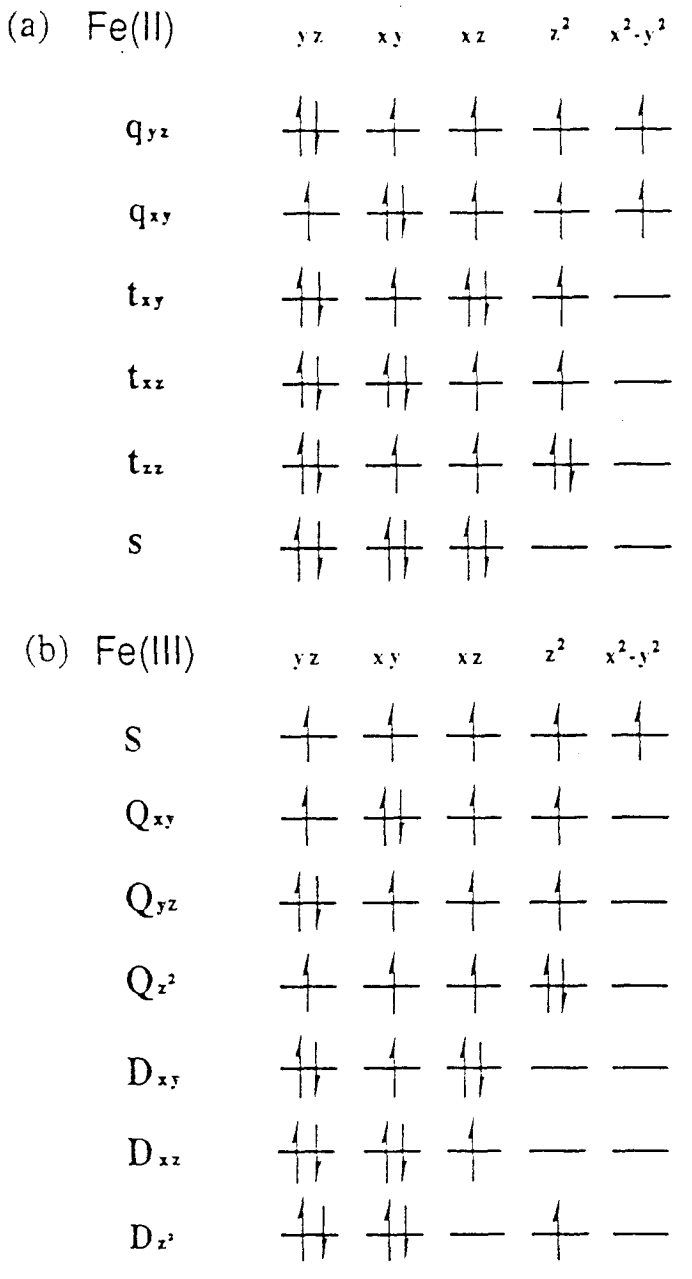


Figure 7. Electronic configurations for various ferrous (a) and ferric (b) states. The states labeled Q, T, Q and D are used without subscript in other sections of this report.

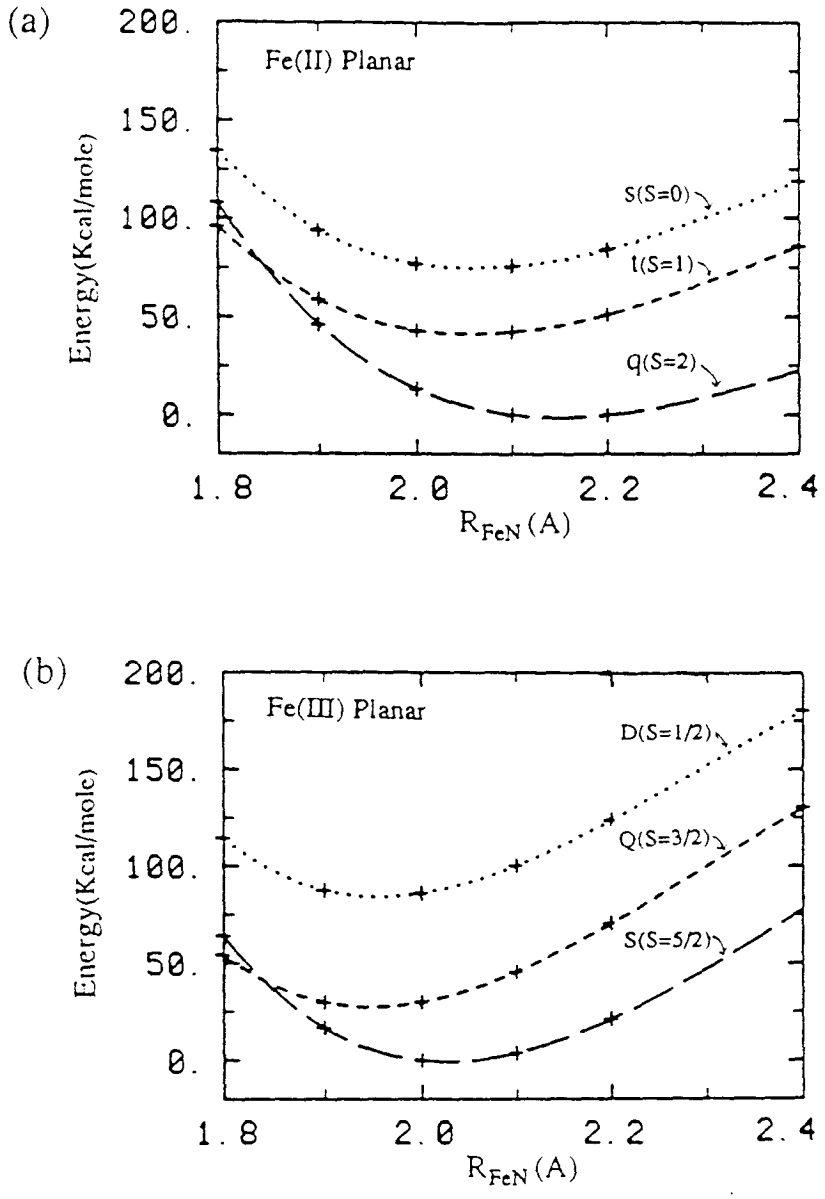


Figure 8. Energetics for *planar* four-coordinate complexes as a function of hole size (R_{Np-N}). (a) Ferrous (Q_{tot} = 0); (b) ferric (Q_{tot} = +1).

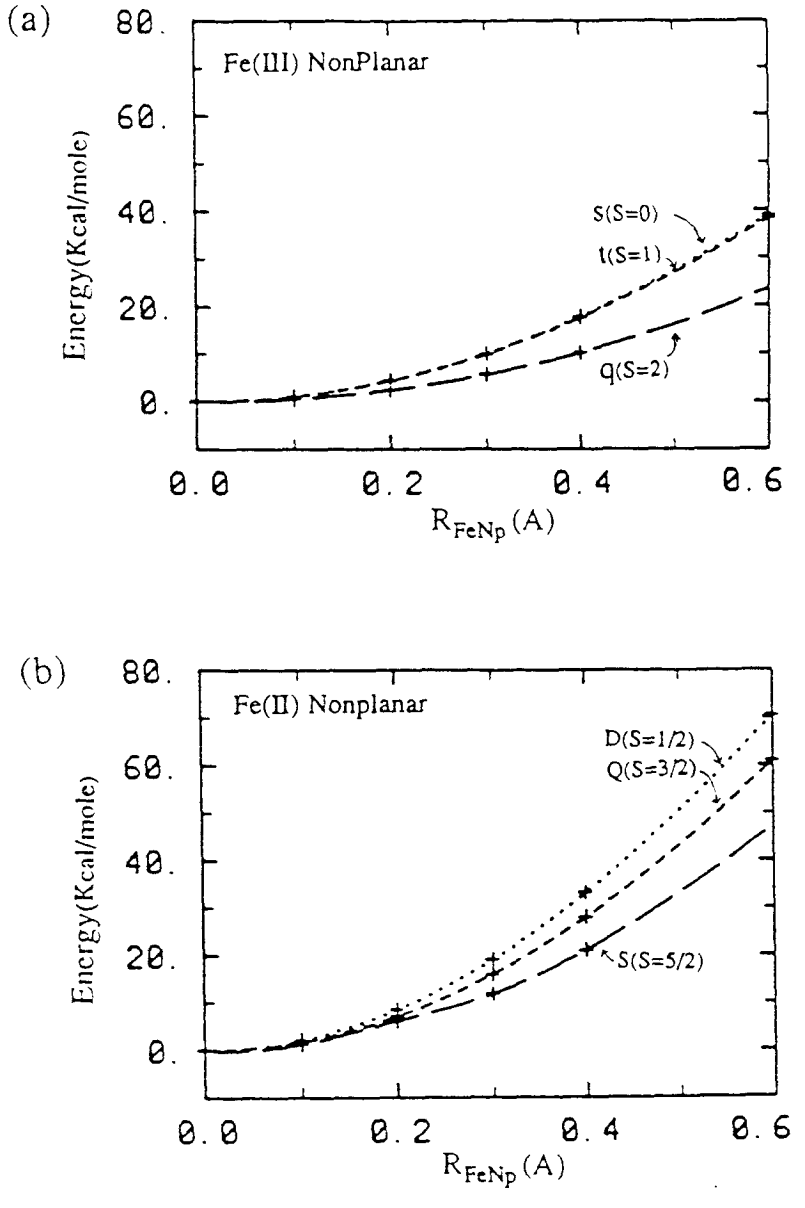


Figure 9. Energetics for *nonplanar* four-coordinate complexes (hole radius fixed at $R_{Np\text{-}N}$ = 2.0044 Å). (a) Ferrous (Q_{tot} = 0); (b) ferric (Q_{tot} = +1).

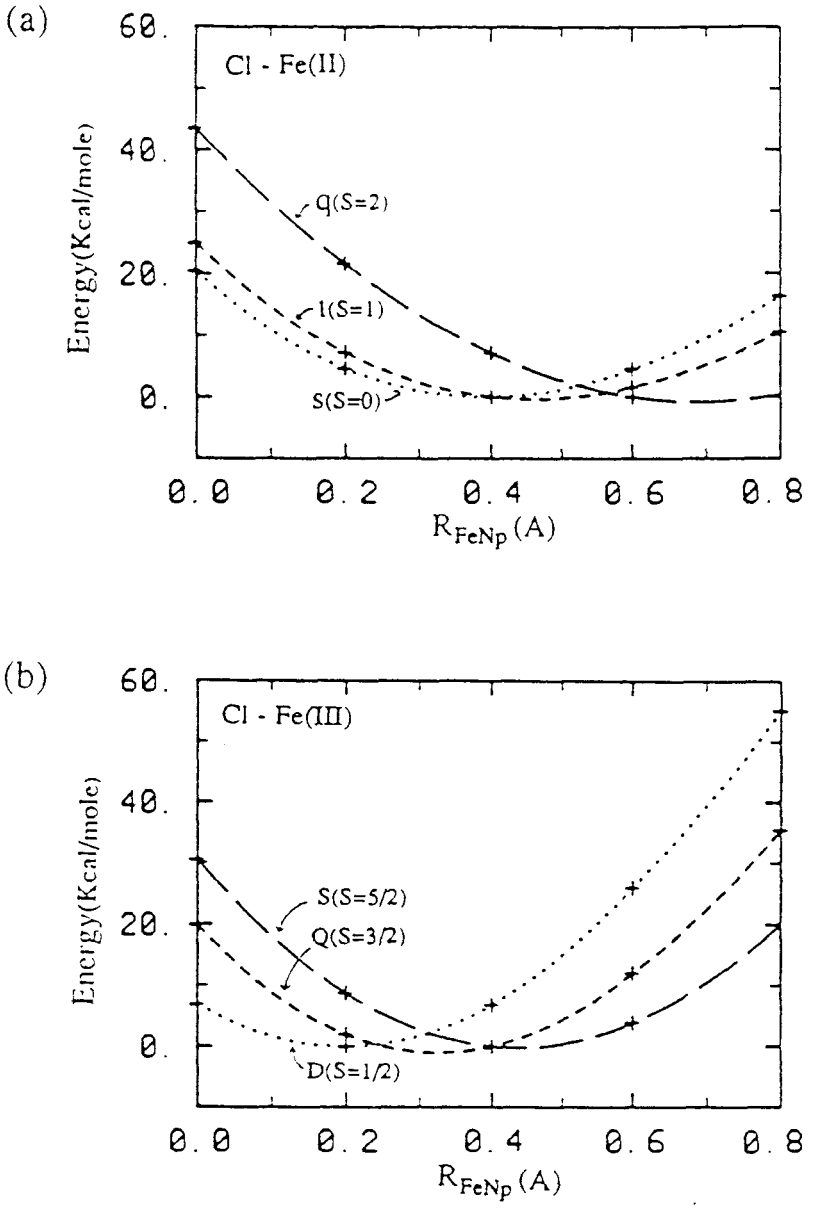


Figure 10. Energetics for five-coordinate complexes (Cl axial ligand) with $R_{Np\text{-}N}$ = 2.0044 Å, $R_{Fe\text{-}Cl}$ = 2.2 Å. (a) Ferrous (Q_{tot} = -1); (b) ferric (Q_{tot} = 0).

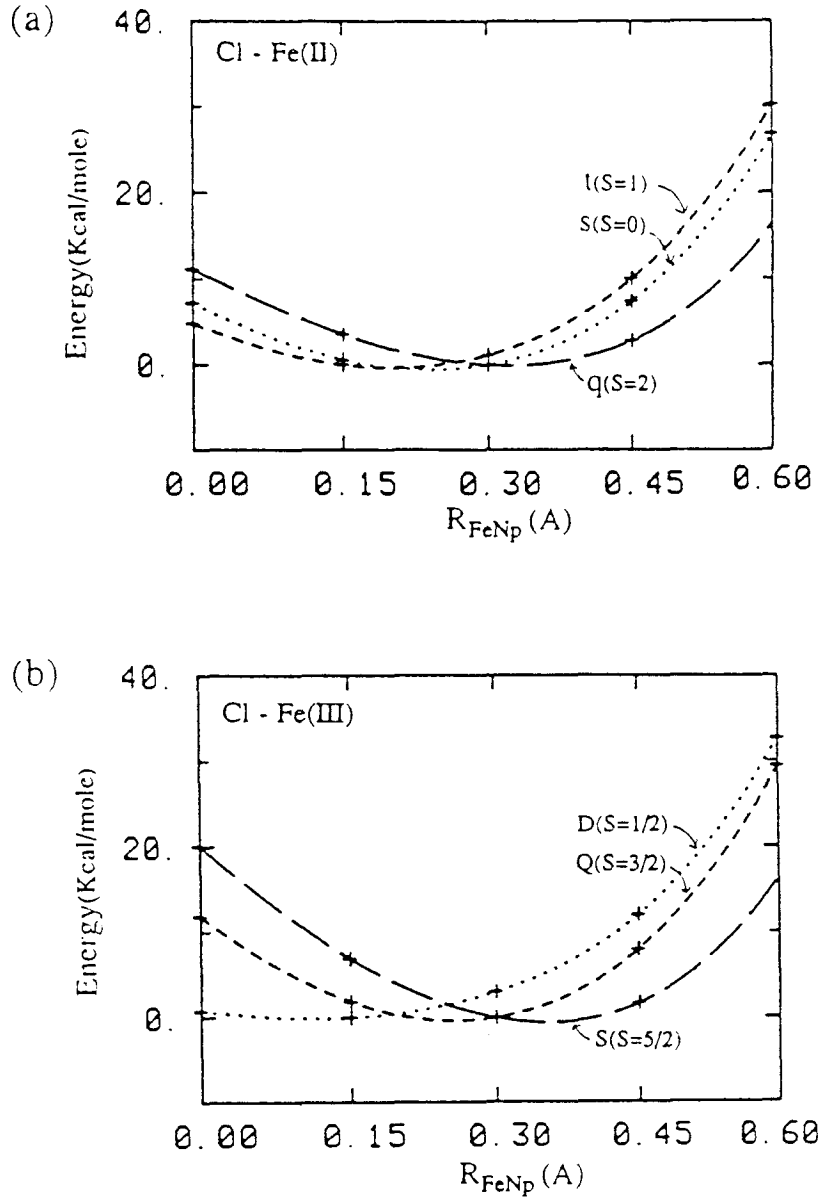


Figure 11. Energetics for five-coordinate complexes (Cl axial ligand) with $R_{Np\text{-}N}$ = 2.0044 Å, $R_{Cl\text{-}Np}$ = 2.63 Å. (a) Ferrous (Q_{tot} = -1); (b) ferric (Q_{tot} = 0).

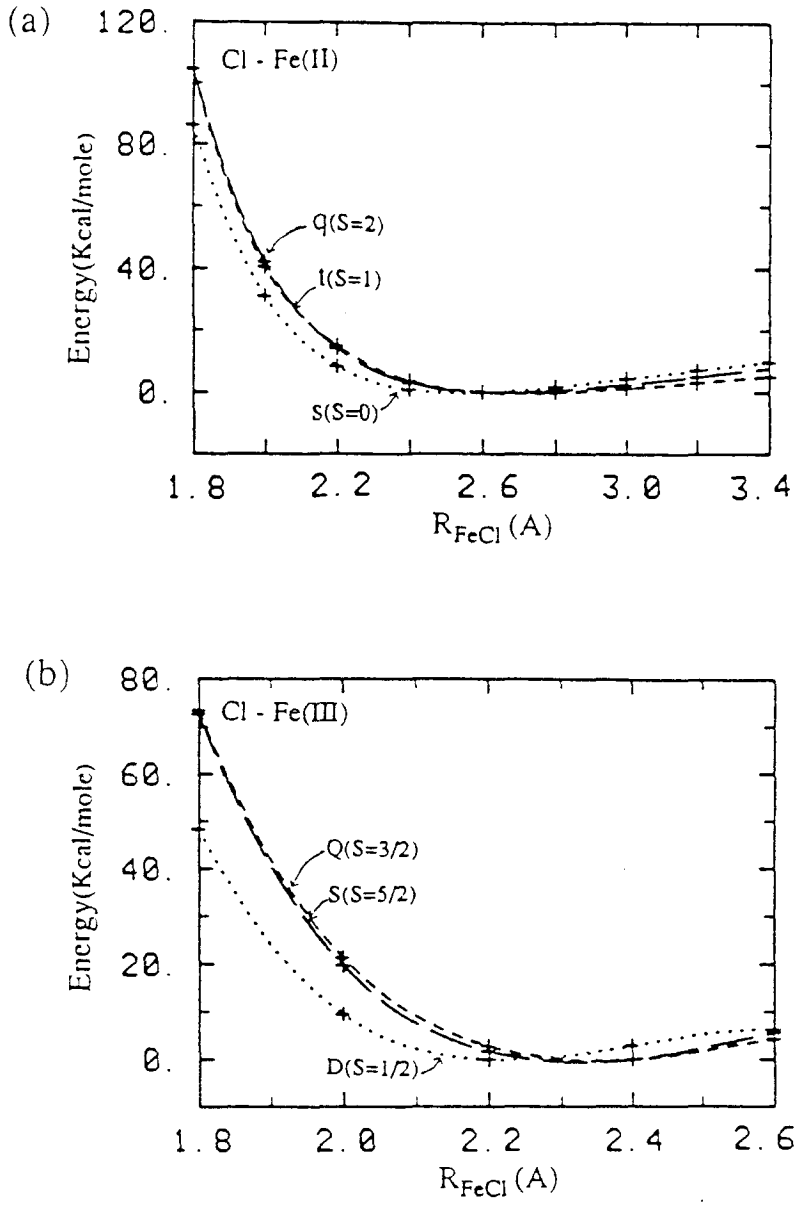


Figure 12. Energetics for five-coordinate complexes (Cl axial ligand) with $R_{Np\text{-}N}$ = 2.0044 Å, $R_{Fe\text{-}Np}$ = 0.43 Å. (a) Ferrous (Q_{tot} = -1); (b) ferric (Q_{tot} = 0).

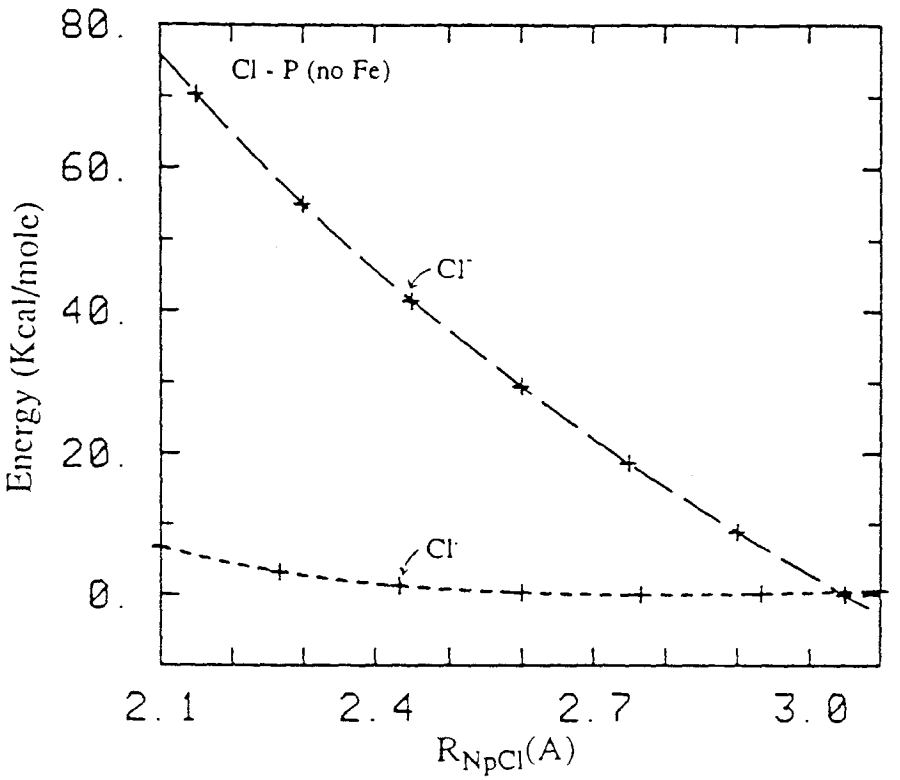


Figure 13. Energetics for five-coordinate complexes with Cl but no Fe (Q_{tot} = -3) R_{Np-N} = 2.0044 Å.

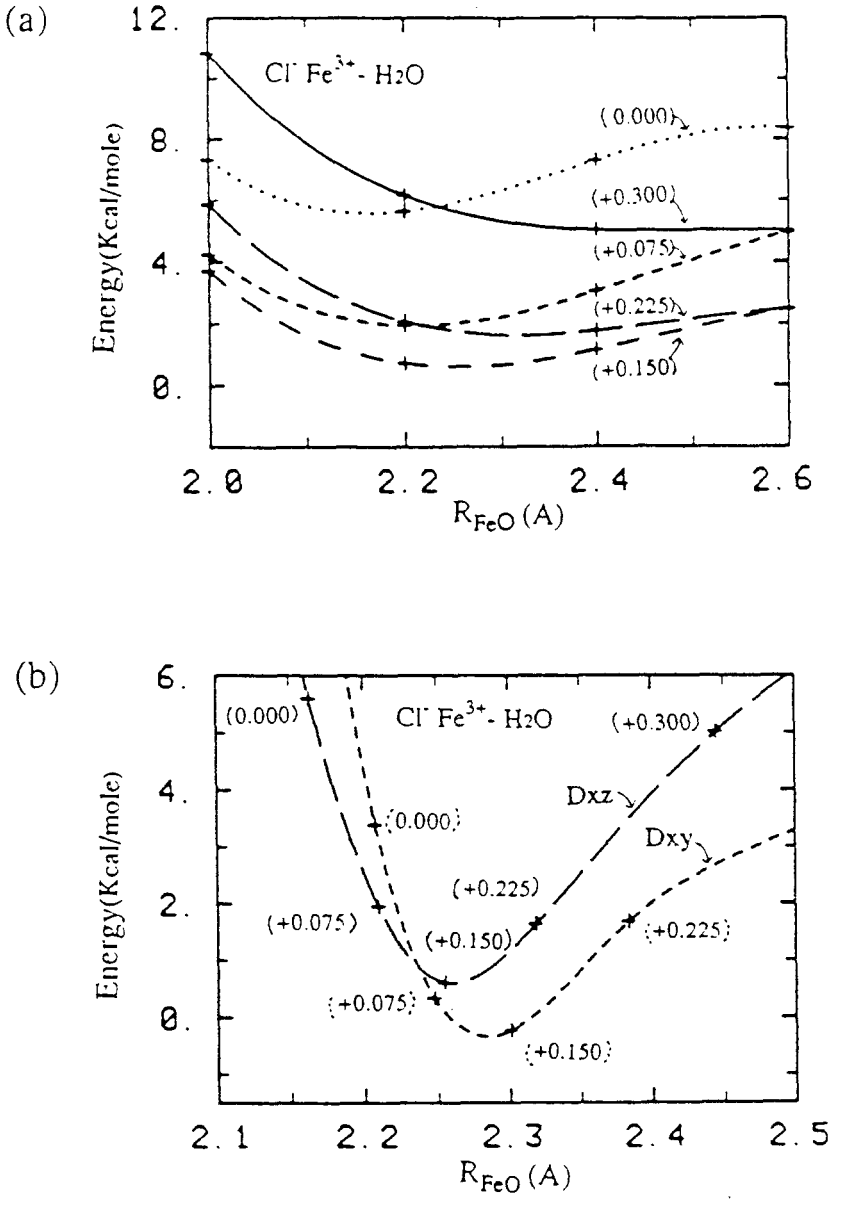


Figure 14. Energetics for ferric six-coordinate complexes with Cl and H_2O axial ligands (Q_{tot} = 0). In each case $R_{Cl\text{-}Np}$ = 2.63 Å, $R_{Np\text{-}N}$ = 2.0044 Å. (a) Dxz (S =1/2) state. Each curve is for fixed $R_{Np\text{-}Fe}$ (positive is toward Cl). (b) Dxz and Dxz states. For each value of $R_{Np\text{-}Fe}$ (see Figure 14a), the optimum $R_{Fe\text{-}O}$ is selected. Thus the left point of Dxz is for $R_{Np\text{-}Fe}$ = 2.0044; the right point is for $R_{Np\text{-}N}$ = -0.30 Å

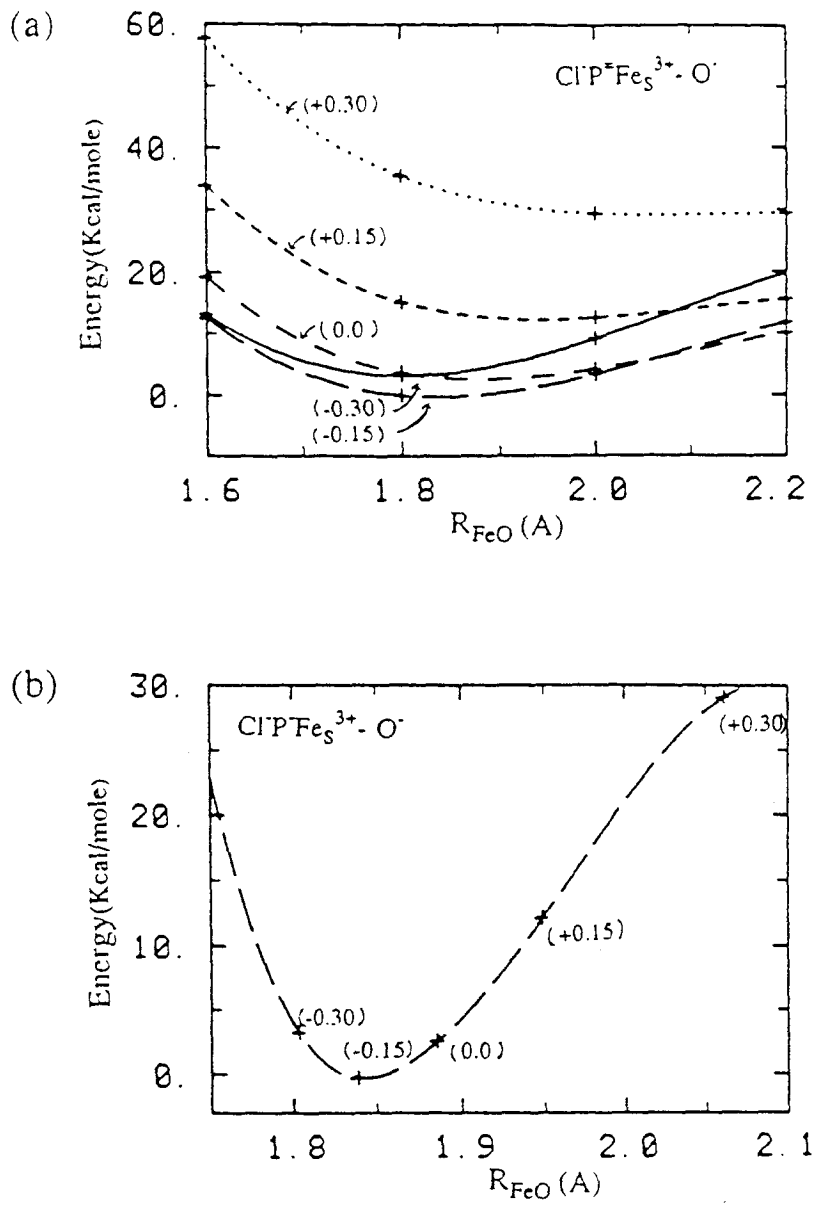


Figure 15. Energetics for ferric six-coordinate complexes with Cl and O axial ligands (R_{Np-N} = 2.0044 Å, R_{Cl-Np} = 2.63 Å). The net charge is Q_{tot} = -1, with the charge distribution best described as $Cl^-P^=Fe^{3+}-O^-$. (a) S state. Each curve is for R_{Np-N} fixed (netive toward oxygen). (b) S State. for fixed value of R_{Np-Fe} (see Figure 15a), the optimum R_{Fe-O} is selected.

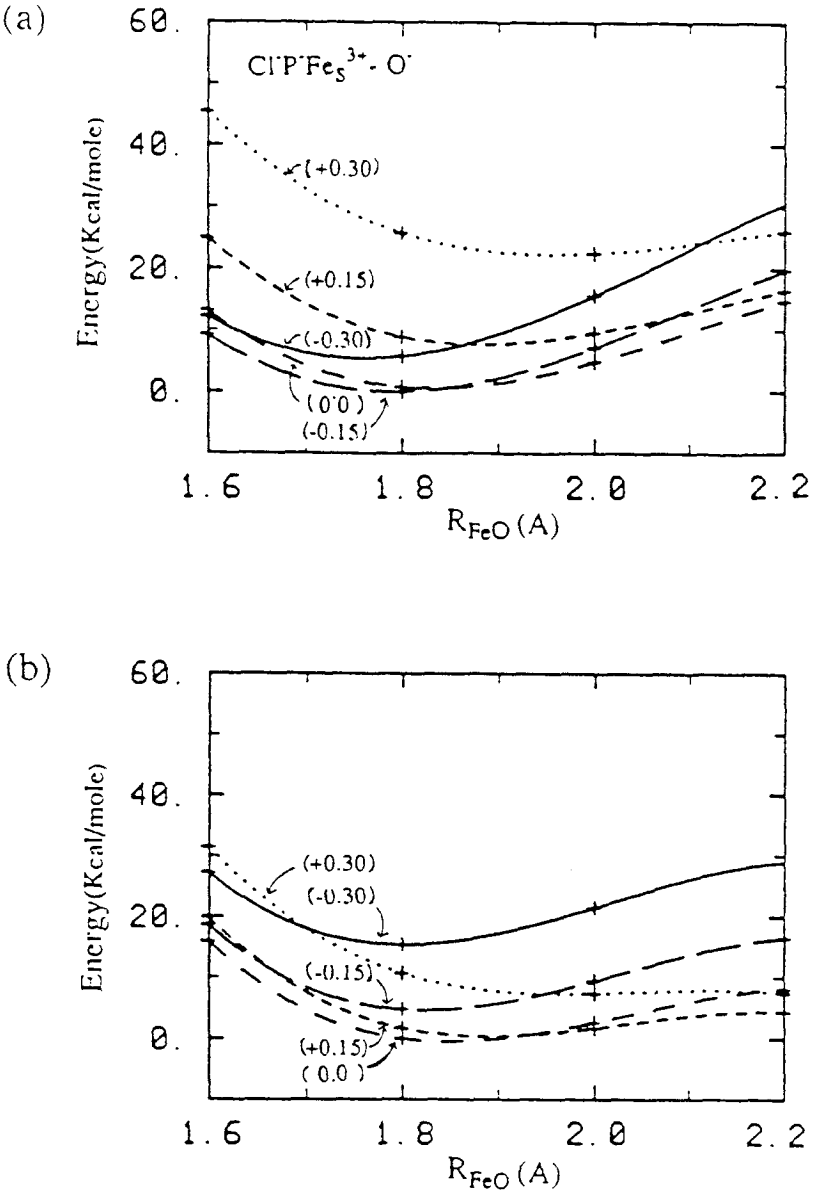


Figure 16. Energetics for ferric six-coordinate complexes with Cl and O axial ligands ($R_{Np\text{-}N}$ = 2.0044 Å, $R_{Cl\text{-}Np}$ = 2.63 Å). The net charge is Q_{tot} = 0, (porphyrin oxidized). (a) Ferric-oxy radical state with the charge distribution best described as $Cl^{-}P^{-}Fe^{3+}\text{-}O^{-}$. (b) Ferrous-oxo state with the charge distribution best described as $Cl^{-}P^{-}Fe^{2+}{=}O$.

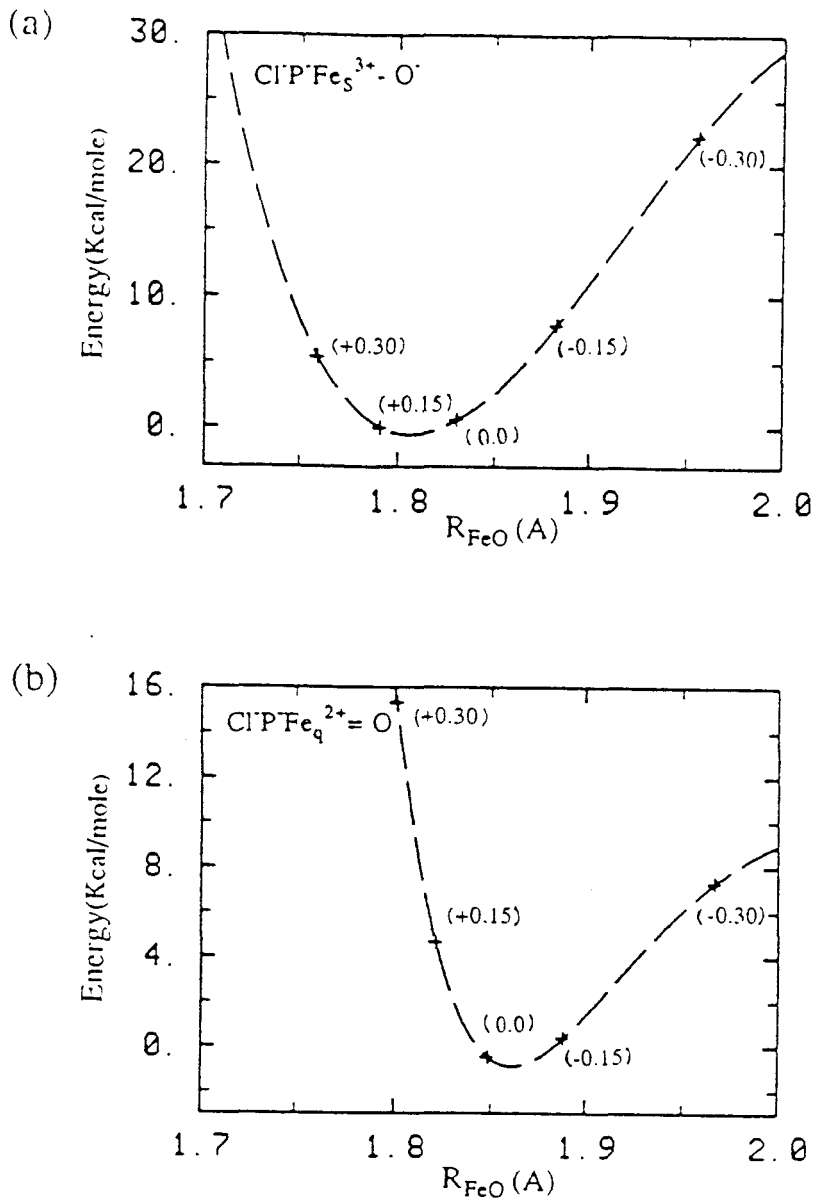


Figure 17. Energetics for six-coordinate complexes with Cl and O axial ligands. The optimum energies of Figure 16 (for each value of R_{Np-Fe}) are plotted versus R_{Fe-O}.

Appendix

Protein Stitchery:
Design of a Protein for Selective Binding
to a Specific DNA Sequence

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 9094–9098, October 1992
Biochemistry

Protein stitchery: Design of a protein for selective binding to a specific DNA sequence

CHANGMOON PARK*†, JUDY L. CAMPBELL†‡, AND WILLIAM A. GODDARD III*†§

*Materials and Molecular Simulation Center, Beckman Institute (139-74), †Division of Chemistry and Chemical Engineering, and ‡Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by William A. Goddard III, June 3, 1992

ABSTRACT **We present a general strategy for designing proteins to recognize DNA sequences and illustrate this with an example based on the "Y-shaped scissors grip" model for leucine-zipper gene-regulatory proteins. The designed protein is formed from two copies, in tandem, of the basic (DNA binding) region of v-Jun. These copies are coupled through a tripeptide to yield a "dimer" expected to recognize the sequence TCATCGATGA (the v-Jun–v-Jun homodimer recognizes ATGACTCAT). We synthesized the protein and oligonucleotides containing the proposed binding sites and used gel-retardation assays and DNase I footprinting to establish that the dimer binds specifically to the DNA sequence TCATCGATGA but does not bind to the wild-type DNA sequences, nor to oligonucleotides in which the recognition half-site is modified by single-base changes. These results also provide strong support for the Y-shaped scissors grip model for binding of leucine-zipper proteins.**

We propose a general strategy for designing proteins to recognize specific DNA-binding sites: this strategy is to select segments of proteins, each of which recognizes particular DNA segments and to stitch these segments together via a short peptide with a cystine crosslink in a way compatible with each peptide being able to bind to its own DNA segment. This technique creates a protein that recognizes the composite site.

As a starting point we consider the gene-regulatory leucine-zipper proteins. They are characterized by two structural motifs (1–3): (*i*) the leucine zipper, which is responsible for dimerization, and (*ii*) the basic region for DNA binding (4–7). The basic regions of unbound leucine-zipper dimers are unfolded but fold into the α-helix conformation upon binding to the specific site (8–10). The most plausible model for the conformation of leucine-zipper protein is the "Y-shaped scissors grip" model (1, 2), in which the basic region of each monomer interacts with DNA on either side of the dyad axis of the binding site. Thus, for yeast transcriptional activator GCN4 each arm recognizes the half-site AGTA (11, 12).

DESIGN

Our design strategy assumes this Y-shaped scissors grip model (Fig. 1*a*). We design proteins by crosslinking (stitching together) various binding arms so as to be consistent with the orientation of the recognition helix in each half-site. Here we build upon the results of Kim and coworkers (5, 6), who showed that the leucine zipper of GCN4 can be replaced with linkers (Gly-Gly-Cys) at the C terminus of the DNA binding segment, which upon oxidation dimerize and bind to the same site (ATGACTCAT) as GCN4. As a model system to explore the design of additional DNA-binding proteins, we have chosen the v-Jun leucine-zipper dimer (Fig. 1*a*), which also binds to the site ATGACTCAT as a homodimer with itself or as a heterodimer with Fos (4, 13–16), another member of this DNA-binding protein family. We will reverse the sequence relationship of the α-helix to the target nucleotide of the binding arms by adding the Gly-Gly-Cys linker to the N terminus (rather than to the C terminus). As illustrated in Fig. 1*b* the designed protein then recognizes the DNA sequence TCATXATGA, where X represents 0–2 additional bases to accommodate the loop region of this dimer.

Several criteria were used in selecting v-Jun as the starting point: (*i*) To prevent nonspecific disulfide-bond formation, the protein must not contain cysteine in its basic region. (*ii*) Because we want to reverse the α-helix relative to the target DNA sequence, the protein should have no residues (especially proline and probably glycine) that would interrupt α-helices. (*iii*) Because we want to ensure that the protein can form the α-helix when joined with the linker, the composition of amino acids in its basic region should strongly favor α-helices [by the Chou–Fasman criterion (17)]. We considered 14 leucine-zipper proteins and found that v-Jun best satisfies the above criteria.

We took as our standard protein the 31 residues at the N terminus of v-Jun joined with the linker (Gly-Gly-Cys) (Fig. 2*a*). The subsequent protein (v-Jun–NN) is designed to bind specifically to the site TCATXATGA, where X might contain 0–2 base pairs (bp). As indicated in Fig. 2*b*, we considered four cases for X: (*i*) X = ϕ (no base pairs), denoted as NNS–ϕ; (*ii*) X = C (which is equivalent to X = G), denoted as NNS–C/G; (*iii*) X = CG, denoted as NNS–CG; and (*iv*) X = GC, denoted as NNS–GC. We excluded using adenine or thymine on the assumption that the methyl group of thymine (which sits in the major groove) might interfere with binding of the protein.

TESTS OF THE DESIGN

We carried out gel-retardation assays using four DNA sequences: (α) the sequence TCATCGATGA (case *iii* above), NNS–CG; (β) the binding sequence for v-Jun, ATGACTCAT; (γ) the complementary double-base-pair substitution ($C^2 \rightarrow A^2$ and $G^9 \rightarrow T^9$) of α: TAATCGATTA; and (δ) the complementary double-base-pair substitution ($A^3 \rightarrow C^3$ and $T^8 \rightarrow G^8$ of α: TCCTCGAGGA.

The results (Fig. 3) indicate that v-Jun–NN binds to the DNA sequence α as a homodimer with a K_d of <1 nM at 4°C. On the other hand, v-Jun–NN does not bind significantly to the wild-type site β or to the mutant sites γ and δ.

To establish the specific binding site for v-Jun–NN, we used deoxyribonuclease (DNase) I footprinting. These results (Fig. 4) show that v-Jun–NN protects the exact binding site predicted for the designed protein. Thus, we conclude that each arm of DNA-bound v-Jun–NN retains the same structure as in native v-Jun. The DNase I footprinting results (Fig. 4) also indicate that NNS–CG has the strongest binding affinity for

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

§To whom reprint requests should be addressed.

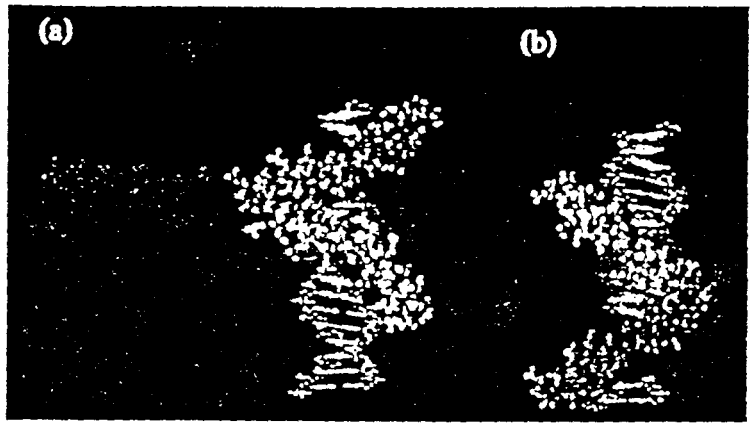


FIG. 1. (*a*) The Y-shaped scissors grip model for the v-Jun–v-Jun homodimer bound to the ATGACTCAT site. (*b*) The designed protein v-Jun–NN. After removing the leucine-zipper region (blue and light-blue) of each monomer, the upper arm (green) and its DNA-binding site (pink, ATGA) were shifted just below the lower arm (orange). In *b* the shifted upper arm and DNA fragment retain their original green and pink colors, respectively. Different linkers (Gly-Gly-Cys, purple) were added at the N termini of both arms, and a disulfide bond was made. v-Jun–N was dimerized by oxidized dithiothreitol and purified by HPLC. Protein concentration was determined by the method of Bradford with the Bio-Rad protein assay kit. Thus, the designed protein is expected to bind to TCATXATGA (where cyan joins to pink in *b*) and does not bind to ATGACTCAT (where pink joins to cyan in *a*), where X (filling bases) fit the loop introduced between the peptide monomers to make the dimer.

v-Jun–NN. The decrease in binding for shorter X may result from the strain required at the loop region of the dimer to place the binding segments along the binding region. Particularly interesting is the difference in specificity observed between NNS–GC and NNS–CG (Fig. 4, compare lanes 9 and 12 for top strands and lanes 21 and 24 for bottom strands). These results indicate that X plays more than the role of spacer.

DISCUSSION

These results support the idea that the N-terminal region of v-Jun contributes to the binding to DNA through specific interaction with the DNA (because in v-Jun–NN this region is forced to contact the DNA). This result supports the angulated bend conformation (1). Our results help differentiate the respective roles of the basic region and of the leucine-zipper region in recognition and binding. The basic region of v-Jun is sufficient for specific binding. Although the leucine-zipper region is not directly involved in DNA binding, our results indicate that its position relative to the basic region plays an important role in determining which target sequence of DNA the protein recognizes.

Summarizing, we have designed a protein (stitched together from segments derived from the natural protein) to recognize a specific DNA-binding site, and we have established specific binding of the designed protein to this site. Note that use of the Gly-Gly-Cys linker is not essential in the design. We could just as well replace the cysteine and make a continuous ≈70-amino acid protein that should recognize a predictable site (14). In addition, this strategy is not limited to two arms. We could have stitched together three, four, or more arms with appropriate linkers to design proteins that would recognize DNA sequences with 15, 20, or 25 bp. Such systems with EDTA-Fe (18) or other nucleases would presumably cut very specific sites, allowing the genome to be cut into much longer segments. The design is not limited to v-Jun. Any protein or other molecule that recognizes a specific

a Protein

vJun-br:	S	QERIKAERKR MRNRIAASKS RKRKLERIAR	
vJun-N:	CGC S	QERIKAERKR MRNRIAASKS RKRKLERIAR	

b DNA

NNS-ϕ:	TCATATGA
NNS-C/G	TCAT(C/G)ATGA
NNS-CG	TCATCGATGA
NNS-GC	TCATGCATGA
α	TCATCGATGA
β	ATGACTCAT
γ	TAATCGATTA
δ	TCCTCGAGGA

FIG. 2. Sequences of protein and oligonucleotides used in gel retardation and footprinting. v-Jun-br contains the basic region of v-Jun, and Gly-Gly-Cys is added to make v-Jun–N. Single-letter amino acid code is used. The protein corresponding to the residues from 218 to 346 of v-Jun was chemically synthesized at the Biopolymer Synthesis Center at the California Institute of Technology. The automated stepwise solid-phase syntheses were done on an Applied Biosystems model 430A peptide synthesizer with an optimized synthetic protocol of the *N*-*t*-butoxycarbonyl (*t*-Boc) chemistry. The peptide was purified by reverse-phase HPLC on a Vydac C_{18} column. A linear gradient of 0–50% aqueous/acetonitrile/0.1% trifluoroacetic acid was run over 120 min. Mass spectroscopy data are as follows: calculated, 4039.3; experimental, 4041.8.

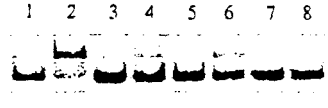


FIG. 3. Gel-retardation assay for binding of v-Jun–NN to various DNA segments. The dimer selectively binds to the predicted site NNS–CG. Even-numbered lanes, no protein; odd-numbered lanes, v-Jun–NN. Lanes: 1 and 2, probe DNA α (NNS–CG); 3 and 4, probe DNA β (wild-type site); 5 and 6, probe DNA γ; 7 and 8, probe DNA δ (Fig. 2*b*). The binding solution contains bovine serum albumin at 50 μg/ml, 10% (vol/vol) glycerol, 20 mM Tris·HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, and 1.56 nM v-Jun–NN in 10-μl reaction volume. After 5000 cpm of each 5′-^{32}P-labeled probe DNA (25- and 26-mer) was added, the solutions were stored at 4°C for 1 hr and loaded directly on an 8% nondenaturing polyacrylamide gel in Tris/EDTA buffer at 4°C. As determined by titration of the gel shift, v-Jun–NN binds to the predicted sequence NNS–CG with a K_d of ≈0.3 nM at 4°C.

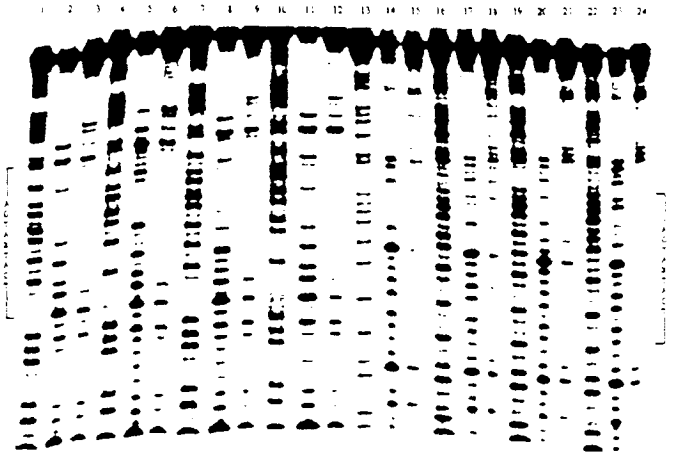


FIG. 4. DNase I footprinting assay of v-Jun–NN with DNA containing the predicted binding sites NNSs. This result shows that the protein protects the target-binding site with the best protection for NNS–CG. Lanes: 1–12, labeled at the 5′ end of top strand; 13–24, labeled at the 5′ end of bottom strand; 1–3 and 13–15, NNS–ϕ; 4–6 and 16–18, NNS–C/G; 7–9 and 9–21, NNS–GC; 10–12 and 22–24, NNS–CG. The first lane for each probe DNA (lanes 1, 4, 7, 10, 13, 16, 19, and 22) contains G+A marker; the middle lane for each probe DNA (lanes 2, 5, 8, 11, 14, 17, 20, and 23) contains no protein; and the last lane for each probe DNA (lanes 3, 6, 9, 12, 15, 18, 21, and 24) contains v-Jun–NN. The footprinting assay solution (in 50 μl) contains bovine serum albumin at 50 μg/ml, 5% (vol/vol) glycerol, 20 mM Tris·HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 20,000 cpm of each 5′-^{32}P-labeled probe DNA (60- to 62-mer), and 50 nM v-Jun–NN. This solution was stored at 4°C for 1 hr. After 5 μl of DNase I diluted in 1× footprinting assay buffer was added, the solutions were stored 1 min more at 4°C. The DNase I digestion was stopped by adding 100 μl of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, sonicated salmon sperm DNA at 25 μg/ml, and yeast tRNA at 25 μg/ml. This mixture was phenol/chloroform-extracted, ethanol-precipitated, and washed with 70% (vol/vol) ethanol. The pellet was resuspended in 5 μl of formamide loading buffer, denatured at 90°C for 4 min, and analyzed on 10% polyacrylamide sequencing gel (50% urea).

DNA sequence by binding along the major groove could be a candidate. Many such cases are now known so that we already have a collection of available partial-binding sites that could be combined to form composite target-binding sites for designing binding proteins. Of course, the segments of these proteins should be designed so that the intramolecular interactions are not so strong as to compete with binding to the DNA.

Our results support the idea that each monomer arm of the dimer binds along the major groove to the half of the binding site of the dimer (11, 12): this strongly supports the Y-shaped scissors grip model for leucine-zipper proteins (1). Our strategy can be used to investigate the interaction between DNA and protein and the structure of DNA–protein complex.

We thank Prof. Mel Simon (California Institute of Technology) for helpful suggestions and discussion and for use of his laboratory resources (funded by National Science Foundation Grant-DMB-90-18536). We also thank Prof. David Eisenberg (University of California, Los Angeles) for helpful comments. This research was supported by a grant from Department of Energy–Advanced Industrial Concepts Division and by funding from the Materials and Molecular Simulation Center of the Beckman Institute. This is contribution number 8547 from the Division of Chemistry and Chemical Engineering.

1. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* 246, 911–916.
2. Pu, W. T. & Struhl, K. (1991) *Proc. Natl. Acad. Sci. USA* 88, 6901–6905.
3. Landschultz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* 240, 1759–1764.
4. Neuberg, M., Schuermann, M., Hunter, J. B. & Muller, R. (1989) *Nature (London)* 338, 589–590.
5. Tlanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* 249, 769–771.
6. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1989) *Science* 243, 538–542.
7. Agre, P., Johnson, P. F. & McKnight, S. L. (1989) *Science* 246, 922–926.
8. Weiss, M. A. (1990) *Biochemistry* 29, 8020–8024.
9. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990) *Nature (London)* 347, 575–578.
10. Saudek, V., Pasley, H. S., Gibson, T., Gausepohl, H., Frank, R. & Pastore, A. (1991) *Biochemistry* 30, 1310–1317.
11. Oakley, M. G. & Dervan, P. B. (1990) *Science* 248, 847–850.
12. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* 249, 774–778.
13. Struhl, K. (1987) *Cell* 50, 841–846.
14. Bos, T. J., Ramscher, F. J., III, Curran, T. & Vogt, P. K. (1989) *Oncogene* 4, 123–126.
15. Abate, C., Luk, D., Gentz, R., Ramscher, F. J., III, & Curran, T. (1990) *Proc. Natl. Acad. Sci. USA* 87, 1032–1036.
16. Turner, R. & Tjian, R. (1989) *Science* 243, 1689–1694.
17. Chou, P. Y. & Fasman, G. D. (1973) *J. Mol. Biol.* 74, 263–281.
18. Mack, D. P., Iverson, B. L. & Dervan, P. B. (1988) *J. Am. Chem. Soc.* 110, 7572–7574.
19. Nakabeppu, Y. & Nathans, D. (1989) *EMBO J.* 8, 3833–3841.