Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 9841–9845, December 1989
Biophysics

Predictions of structural elements for the binding of Hin recombinase with the *hix* site of DNA

(protein–DNA interactions/helix–turn–helix motif/molecular modeling/structure–function relationships/*Salmonella typhimurium*)

KEVIN W. PLAXCO, ALAN M. MATHIOWETZ, AND WILLIAM A. GODDARD III

Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, CA 91125

Contributed by William A. Goddard III, September 25, 1989

ABSTRACT **Molecular dynamics simulations were coupled with experimental data from biochemistry and genetics to generate a theoretical structure for the binding domain of Hin recombinase complexed with the *hix* site of DNA. The theoretical model explains the observed sequence specificity of Hin recombinase and leads to a number of testable predictions concerning altered sequence selectivity for various mutants of protein and DNA.**

A critical problem for fully exploiting the opportunities in protein engineering is to understand the principles determining why a protein binds selectively to a particular base-pair sequence of DNA. Advances in this understanding have been made by a number of indirect studies; however, the difficulties associated with crystallization and analysis of protein–DNA complexes limit the opportunities to obtain structural information directly from crystallography. Our research objectives are to elucidate such interactions by using a combination of molecular mechanics and molecular dynamics simulations constrained by knowledge-based structural predictions. Because of the vast amount of solution-phase experimental data accumulated about the DNA-binding characteristics of Hin recombinase (refs. 1–6; J. Sluka, A. C. Glasgow, M. I. Simon, and P. B. Dervan, personal communication; D. Mack and P. B. Dervan, personal communication), we selected this system for application of a constrained simulations approach (developed by A.M.M. and W.A.G.; to be published at a later date). Utilizing these theoretical techniques in conjunction with information gleaned from various experiments, we have derived a theoretical model of the Hin–DNA binding that is consistent with current experimental data. This model suggests a number of new experiments to test and refine the ideas about the interactions determining site-specific protein–DNA binding.

These studies illustrate what we believe will be an effective mode of elucidating the mechanisms of sequence-specific protein–DNA binding. Experimental techniques such as chemical and enzymatic footprinting, affinity cleavage, and genetics can specify the regions (both sequence and groove location) of DNA sites involved in protein–DNA recognition, and define the structural motifs involved in protein binding. However, these techniques do not provide detailed atomic-level information about the interactions responsible for site specificity. Theoretical molecular dynamics calculations can provide useful information about interactions at the atomic level, but with current techniques these studies are only practical if the region of protein and DNA involved in the interaction is specified. The detailed model can then be used to design experiments that can distinguish subtle differences in the nature of the specific interactions and further refine the theoretical model.

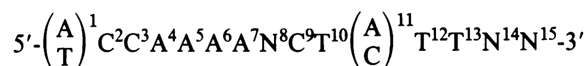
The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Hin recombinase is a 190-amino acid protein from *Salmonella typhimurium* containing both a specific DNA-binding activity and a DNA-recombination activity. Hin recombinase mediates a site-specific recombination between two 26-base-pair elements (*hix*L and *hix*R) separated by 993 base pairs of *Salmonella* chromosomal DNA. Hin binds to the pseudo-dyad symmetric *hix* binding sites as a dimer, with one molecule of the protein at each of the two half-sites comprising the dyad repeat. Purified Hin is able to catalyze a phosphodiester cleavage at the center of symmetry of each *hix* site *in vitro* and in the presence of the proteins Fis and Hu is able to perform strand exchange and religation of the DNA between the two hix sites (reviewed in ref. 4).

Hin is a member of a large family of site-specific recombinases from widely divergent organisms that are homologous members of the helix–turn–helix family of proteins (1). This family includes λ Cro and λ repressor, which have been structurally determined (7, 8), as indicated in Fig. 1.

By analogy to the closely related protein γδ resolvase, Hin was postulated to contain two domains: a catalytic domain comprising the amino-terminal 138 residues of the protein and a DNA-binding domain consisting of the carboxyl-terminal 52 residues* (1) and containing the putative helix–turn–helix motif. Experiments with a chemically synthesized DNA-binding domain [Hin-(139–190)] confirmed that this region is sufficient for binding to the symmetric halves of each *hix* element and that it does so with a binding energy at least half that of dimeric wild-type protein (refs. 1 and 6; J. Sluka, A. C. Glasgow, M. I. Simon, and P. B. Dervan, personal communication). Because proteins containing 55 and 60 amino acids of Hin bind with an affinity comparable to that of the 52-residue protein, it is reasonable that Hin-(139–190) contains all that is necessary for maximal binding. On the other hand, experiments with the 31-residue polypeptide of Hin-(160–190), thought to correspond to the entire helix–turn–helix domain, demonstrated that it is not sufficient for binding, suggesting that other elements in the DNA-binding domain play an additional role in Hin binding and selectivity (1). Fis and Hu are not necessary for Hin binding (4).

DNA footprinting has delineated the region of DNA involved in recognition (1), while genetic studies have been used to define the sequence requirements for Hin binding (5). DNA methylation interference and protection patterns have been used to elucidate specific contacts between Hin and its binding site (3, 6). These data plus comparison of the four naturally occurring *hix* half-sites provide the consensus sequence for the DNA sequence recognized by Hin (5):



$$5'\text{-}\binom{A}{T}^{1}C^{2}C^{3}A^{4}A^{5}A^{6}A^{7}N^{8}C^{9}T^{10}\binom{A}{C}^{11}T^{12}T^{13}N^{14}N^{15}\text{-}3'$$

*The complete sequence of the DNA-binding domain of Hin is as follows (in the standard one-letter code): GRPRA INKHE QEQIS RLLEK GHPRQ QLAII FGIGV STLYR YFPAS SIKKR MN (residues 139–190).

	STABILIZATION HELIX	TURN	RECOGNITION HELIX	
λ Repressor	31 - Leu Ser Gln Glu Ser Val Ala Asp Lys Met	Gly Met Gly	Gln Ser Gly Val Gly Ala Leu Phe Asn	Gly Ile Asn - 55
	P^{9}		A^{2} G^{-4} P^{-5} P^{8}	G^{-6}
λ Cro	14 - Phe Gly Gln Thr Lys Thr Ala Lys Asp Leu	Gly Val Tyr	Gln Ser Ala Ile Asn Lys Ala Ile His Ala	Gly Arg - 38
	P^{8} P^{9}		A^{2} A^{-3} P^{8} $G^{-4}T^{-5}$ P^{7}	G^{-6}
Hin Recombinase	161 - Pro Arg Gln Gln Leu Ala Ile Ile Phe	Gly Ile Gly	Val Ser Thr Leu Tyr Arg Tyr Phe Pro	Ala Ser Ser - 184
	P^{11} P^{12} P^{13}	T^{13}	T^{12} A^{-10} P^{10} G^{-9} P^{-9}	

FIG. 1. DNA-binding domains of λ repressor (phage λ), λ Cro (phage λ), and Hin recombinase (from *S. typhimurium*). The important interactions with DNA are indicated below each amino acid sequence. The nucleotides and phosphates contacted are numbered from the center of symmetry of the appropriate sites. Data are from protein/DNA cocrystal structure (λ repressor; ref. 6), model building from known protein structure (λ Cro; ref. 5), and this work (Hin).

For our calculations, we have used the sequence

$$5'\text{-}A^{4}A^{5}A^{6}A^{7}C^{8}C^{9}T^{10}C^{11}T^{12}T^{13}G^{14}G^{15}\text{-}3'.$$

This element corresponds to that portion of the consensus sequence that has been shown to be contacted by Hin as determined by DNase I and methylation protection assays (1, 3).

Affinity-cleavage studies utilizing proteins equipped with nonspecific cleaving moieties [Fe(II)-EDTA] have defined the location of the amino and carboxyl termini of the Hin DNA-binding domain. Sluka *et al.* (2) have put forward a model based on a helix–turn–helix motif where the amino terminus of the protein is located in the minor groove near the symmetry axis of the *hix* site. The residues Gly^{139}, Arg^{140}, Pro^{141}, and Arg^{142}, located in the minor groove, participate in sequence-specific recognition. Additional sequence-specific interactions are provided by the putative recognition helix, which is also oriented toward the symmetry axis of the site (D. Mack and P. B. Dervan, personal communication). These data serve to define the orientation of the interaction but do not indicate the detailed atomic interactions responsible for recognition.

CALCULATIONS

The theoretical studies involved torsion-space and traditional molecular mechanics simulations aided by constraints imposed to bias the conformations to fit experimental data and insights. The initial model for the binding domain of Hin was constructed by aligning the sequence of the carboxyl-terminal 52 residues of Hin with the sequence of Cro. With this alignment (partially illustrated in Fig. 1), the helix–turn–helix domain of Hin (residues 146–190) was built onto the C^{α} coordinates of Cro (9) by using the C^{α}-constrained torque mechanics approach to structure prediction. The structure was created one residue at a time starting at residue 146 and proceeding through residue 190. As each residue was added to the growing chain, the structure was optimized by minimizing normal valence and nonbond potential energy terms, in conjunction with harmonic potentials constraining the Hin C^{α} atoms to the positions in Cro. The structure thus created was then optimized in the absence of constraints and allowed to equilibrate with molecular dynamics.

Comparison of the sequence selectivity of Hin and Cro provided clues for the initial docking of Hin to DNA. The binding elements of both share a sequence, CTNT, for which in Cro the suggested structure (7) involves Lys·G and Ser·A hydrogen bonding to the first two base pairs. Since the corresponding residues of Hin are Arg^{178} and Ser^{174}, initial docking was performed by allowing Ser^{174} to produce a bridging hydrogen bond with A^{-10}†, in an orientation analogous to that suggested for Cro (7). This orientation is consistent with the carboxyl-terminal affinity-cleavage experiments of Mack and Dervan (personal communication), which localized the carboxyl end of the domain to a region proximal to the dyad center.

Once the helix–turn–helix element of Hin was docked in this orientation, the structure was optimized in three steps: (*i*) energy minimization of the protein–DNA complex with the

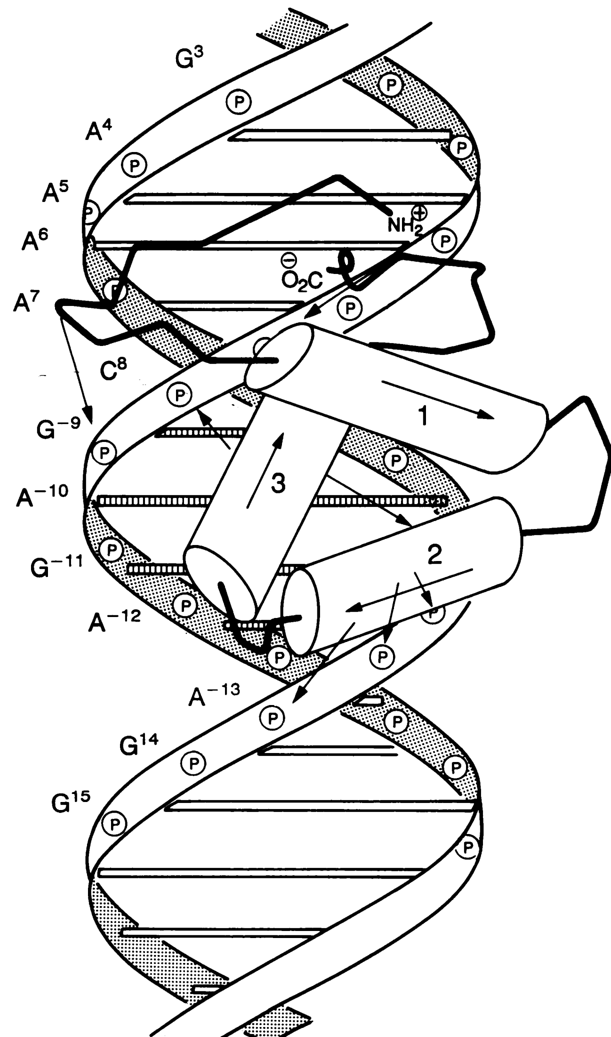


FIG. 2. A model of the proposed structure of the DNA-binding domain of Hin recombinase with indications of the backbone structure and phosphate contacts (hydrogen bonding to phosphates is illustrated by arrows to the phosphates contacted). Hin helices 1 (no sequence-specific interactions), 2 (stabilization helix), and 3 (recognition helix) are shown with arrows indicating the overall direction of the polypeptide backbone.

†Nucleotides in the strand opposite to that shown in the consensus sequence will be designated with negative superscripts corresponding to the numbers of the nucleotides to which they are base-paired.

use of defined docking constraints (artificial bonds involving Ser^{174}·A^{-10} and Arg^{178}·G^{-9}) to reduce steric interactions, (*ii*) equilibration of the minimized complex by molecular dynamics, and (*iii*) unconstrained dynamics and minimization to produce a low-energy conformation. During the first two levels of calculation, the *hix* DNA was held fixed in a standard B-form DNA structure.

By covalently attaching the cleavage reagent Fe(II)-EDTA to the peptide, Sluka *et al.* (2) showed that the amino-terminal end of the binding domain (residues 139–146) is located in the minor groove, a conclusion confirmed by methylation studies (ref. 6; J. Sluka, A. C. Glasgow, M. I. Simon, and P. B. Dervan, personal communication). Thus, initially we treated the amino-terminal region (residues 139–146) of the DNA-binding domain as an independent motif, generated the structure separately in an extended conformation, and optimized it independent of the remainder of the polypeptide. The amino terminus was then docked to the minor groove by following the model of Sluka *et al.* (ref. 6; J. Sluka, A. C. Glasgow, M. I. Simon, and P. B. Dervan, personal communication) [as derived from chemical modification data (2)]. Constraints were provided between Arg^{140} and Arg^{142} and atoms in the minor groove (at base pairs 4, 5, and 6) and the structure was optimized as described above. The low-energy minor-groove (residues 139–146) and major-groove (residues 146–190) portions were then linked with an artificial distance constraint and the full 52-residue polypeptide was optimized.

Although there are approximations and restrictions inherent in such calculations, we believe that they account for proper steric and hydrogen bonding interactions and lead to a number of new structural details that help formulate experiments to test the structural elements.

RESULTS

The predicted structure of the Hin binding domain is sketched in Fig. 2, while Fig. 3 contains a stereo image including some specific contacts. The recognition helix (helix 3, residues 173–180) lies within the major groove and involves four major site-specific sets of interactions, stabilized by two tyrosine–phosphate hydrogen bonds. Optimal tyrosine–phosphate hydrogen bonding seems to produce a somewhat wider major groove (and narrower minor groove) than that of the classic B-form DNA [this is consistent with the observation that methylation of A^{-10} at the minor-groove atom N3 is reduced by Hin–*hix* complex formation (3), presumably because of this modification of DNA conformation].

The recognition helix is held in place by the stabilization helix (residues 162–169) that has hydrophobic interactions with the recognition helix plus hydrogen bonds between side chains of Arg^{162}-Gln^{163}-Gln^{164} and P^{13}–P^{11} of the phosphate backbone. Additional phosphate contacts are made by Lys^{146} (with P^{-9}) and Lys^{186} (with P^{-5}), further increasing the non-sequence-specific energy of interaction. Fig. 2 shows the orientation of Hin in the complex together with phosphate contacts made in the complex.

The specific interactions of Arg^{140} and Arg^{142} with the minor groove of $A^5A^6A^7$ are supported by the region 143–161, which includes a third helix lacking sequence-specific interactions with the DNA (Lys^{146} has a hydrogen bond to P^{-9}). A number of hydrophobic interactions between groups on the three helices provides additional hydrophobic stabilization of the overall structure.

The protein–DNA contacts summarized in Fig. 4 are sufficient to explain the known sequence selectivity and the observed methylation interference and protection patterns characteristic of the Hin protein (2, 3) with only one exception noted below. Key points are as follows.

(*i*) The strong selectivity for $T^{12}T^{13}$ is generated by complementarity between the hydrophobic surface created by the side chains of residues in the turn region of the peptide, Ile-Gly-Val (residues 171–173), and the C5 methyl groups of the thymines.

(*ii*) The model structure shows no significant interaction between position 11 and the Hin protein (this agrees with the lack of sequence conservation at position 11 in *hix* sites and with the apparent lack of sequence selectivity associated with this site for Hin proteins).

(*iii*) The model has T^{10} strongly preferred due to a bridging set of hydrogen bonds between the complementary base A^{-10} and Ser^{174} of Hin. Such hydrogen bonding has been postulated to produce adenine specificity upon Cro binding, as suggested by Ohlendorf *et al.* (7). Replacement of A^{-10} with either C or G would reduce optimal hydrogen bonding (neither has hydrogen bond donor capability in the major groove), while an $A^{-10} \rightarrow T$ transversion would eliminate the hydrogen bonding potential at this site.

(*iv*) In the theoretical model, Arg^{178} is responsible for recognition of G^{-9}, donating hydrogen bonds to the N7 nitrogen and the C5 carbonyl groups. Consistent with this, methylation of the N7 nitrogen of G^{-9} is deleterious to Hin binding (3). Replacement of G^{-9} could be tolerated since the guanidinium group of arginine could also form a set of bridging hydrogen bonds across the N7 atoms of positions −9 and −8. Thus, alternative sequences providing full hydrogen bonding to Arg^{178} would also include N^8C^9 as well as $\binom{T}{C}^8 T^9$ $\left[\text{corresponding to } \binom{A}{G}^{-8} A^{-9}\right]$.

(*v*) Position 7, a conserved adenine in the four known *hix* half-sites, does not appear to be contacted by the protein. This lack of selectivity is consistent with the observation that methylation of the N3 of A^7 does not reduce Hin binding, and

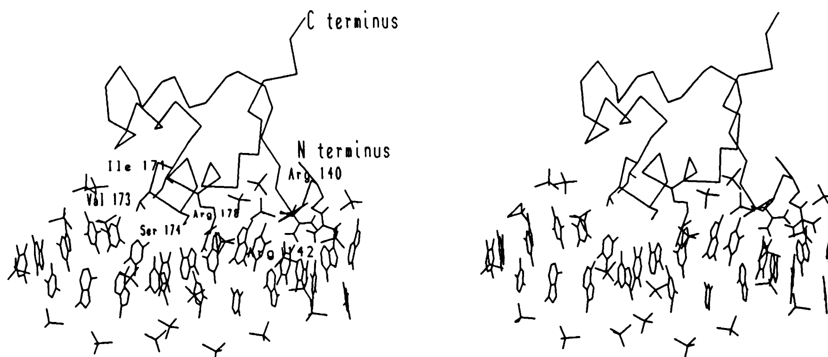


FIG. 3. A stereo diagram of proposed sequence-specific contacts in the major and minor grooves. The C^α trace for the entire binding domain is shown with the side chains of those residues implicated in sequence-specific contacts. The deoxyribose sugars have been omitted for clarity.

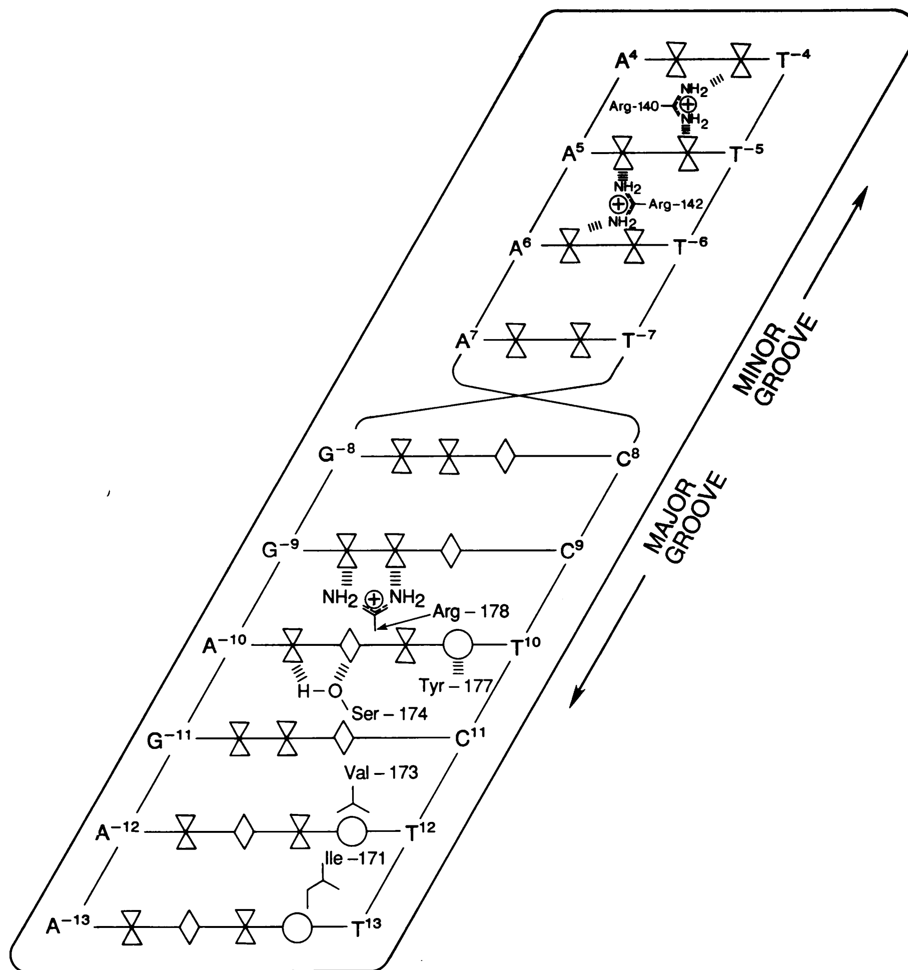


FIG. 4. A diagram of the sequence-specific contacts of the Hin–*hix* complex. Hydrogen bond donors are designated by diamonds, acceptors by hourglasses, and the 5-methyl group of thymine by a circle.

Hin binding does not prevent methylation at this site (3). We propose that the observed conservation at this site is coincidental and plays no role in DNA recognition.

(*vi*) Recognition at positions 4, 5, and 6 is provided by sequence-specific contacts between the amino terminus of the binding domain and atoms in the minor groove (Arg^{140} with the O3 atoms of $T^{-5}T^{-4}$ and Arg^{142} with the N3 nitrogens of $A^{6}A^{5}$).

(*vii*) The final positions, 3–1, are conserved among all known *hix* sites, but DNase I footprinting suggests that these positions are not contacted by the 52-residue binding domain (4), and they are not contacted in our structure. It is possible that this region plays a role in site-specific recombination by interacting with the catalytic domain of recombinase.

SUGGESTED TESTS OF THE STRUCTURE

The theoretical model of the Hin binding domain suggests mutations (see Fig. 5 for a summary) and base-pair substitutions that should alter binding specificity.

The model predicts that residues 146, 162, 163, 164, and 186 donate hydrogen bonds to the phosphate backbone and that these positions can be satisfied with any residues with similar hydrogen bonding potential (such as arginine, asparagine, glutamine, or lysine). Tyr^{177} and Tyr^{179} are involved in hydrogen bonding to the phosphate backbone but may also play a role in modifying the width of the major groove. If this latter effect is important in recognition, their substitution with amino acids having equivalent hydrogen bond donor ability but lacking such a rigid side chain would lead to reduced binding.

Arginine is the only amino acid capable of bridging the minor-groove hydrogen bond acceptors and providing adenine specificity at positions 4, 5, and 6. Any other substitution should alleviate selectivity at these positions and greatly

Hin Sequence									
Ile	Gly	Val	Ser	Thr	Leu	Tyr	Arg	Tyr	Phe
Functionally Acceptable Substitutes									
						Arg		Arg	
				Ala		Asn		Asn	Ile
Val				Cys	Ile	Gln		Gln	Leu
Leu	---	---	---	Ser	Val	Lys	Lys	Lys	Met
Substitutions that Alter Selectivity									
Ala		Ala							
Gln		Thr	Ala				Ala		

FIG. 5. Predicted effects of various point mutations in the putative recognition helix. The wild-type Hin sequence is shown in the upper row. Conservative mutations, those that are predicted to affect neither the structure nor the binding characteristics of Hin, are listed in the middle rows. Mutations predicted to change sequence selectivity without structurally disrupting Hin are shown in the lower rows; the symbol Ala actually refers to any small residue that will not disrupt the overall structure of the binding domain. All other symbols represent the standard three-letter code.

diminish binding (due to the loss of hydrogen bonding). The calculations suggest that flexibility of the arginine side chain would also allow equivalent hydrogen bonding to A·T transversions. However, the C2 amino group of guanine, located in the minor groove, would prevent hydrogen bonding to G·C-containing sequences.

Substitutions of Ile^{171}, which is responsible for selective hydrophobic interactions with T^{13}, are probably limited to the sterically conservative replacements leucine and valine. Replacement of Ile^{171} with the hydrophilic amino acid glutamine may lead to adenine selectivity at position 13 by providing a hydrogen bond donor and acceptor of the correct geometry.

Val^{173} (responsible for T^{12} selectivity) is a highly constrained position that can only be satisfied by valine. Replacement with a small amino acid such as alanine should remove selectivity at position 12 and reduce overall Hin affinity, and replacement of Val^{173} with threonine may generate A^{12} selectivity by providing a hydrogen bond acceptor and donor of the appropriate geometry.

Ser^{174} is a highly conserved position; no other amino acid is capable of forming such a tightly constrained set of hydrogen bonds. The functionally conservative replacement of a threonine is sterically forbidden due to the highly constrained geometry of the hydrogen bonding, while replacement of this amino acid with one of smaller size (e.g., alanine) would reduce binding by removing two hydrogen bonds.

The next residue, Thr^{175}, does not appear to make any contact with the DNA and presumably is constrained only by steric considerations. Substitution of small amino acids at this position (serine, cysteine, or alanine) should not reduce binding.

Leu^{176} serves a structural role as part of the hydrophobic core of the domain. Residues that participate in hydrophobic interactions are often highly constrained sterically (10), and thus few substitutions at this position would produce stable proteins.

Tyr^{177} and Tyr^{179} play a role in hydrogen bonding to phosphate, as discussed above. For Arg^{178}, the only replacement that would maintain selectivity at position 9 should be lysine, which is also capable of donating two hydrogen bonds.

Phe^{180} also plays a role in maintaining the hydrophobic core of the protein and, as such, is tightly constrained.

To facilitate additional experimental and theoretical investigation of this model, the coordinates of this proposed complex are being submitted to the Brookhaven Protein Database.

SUMMARY

Combining molecular dynamics simulations with constraints based on current knowledge of protein structure leads to a theoretical structure of the binding domain of Hin recombinase with the *hix* site of DNA. The model offers a mechanistic explanation of the presently known characteristics of Hin and predicts the effects of specific mutations of both protein and DNA. The predictions can be tested by currently feasible experiments that should lead to refinements in and improvements on the current theoretical model. Because current experimental and theoretical methods are all limited to providing only partial information about protein–DNA interactions, we believe that this approach of basing molecular simulations on experimental knowledge and using the results of these simulations to design new, more precise experimental tests will be of general utility. These results provide additional evidence for the generality of the helix–turn–helix motif in DNA recognition and stabilization of proteins on DNA.

We thank Professor Peter Dervan for helpful discussions. K.W.P. gratefully acknowledges support from the National Science Foundation in the form of a graduate fellowship. This project was initiated with support from Office of Naval Research/Defense Advanced Research Planning Agency and continued with support from Department of Energy-Energy Conversion and Utilization Technologies. The computations were carried out using BIOGRAF (from BioDesign) with modified routines for torque mechanics and for C^{α} constraints. The computers (Alliant FX8/8 and DEC VAX 8650) and graphics systems (Evans & Sutherland PS 330 and 390) were funded by Office of Naval Research/Defense Advanced Research Planning Agency, National Science Foundation-Materials Research Groups, Office of Naval Research-Special Research Opportunities, and Department of Energy-Energy Conversion and Utilization Technologies. This paper is Contribution 7771 from the Arthur Amos Noyes Laboratory of Chemical Physics.

1. Bruist, M., Horvath, S., Hood, L., Steitz, T. & Simon, M. I. (1987) *Science* **235**, 777–780.
2. Sluka, J., Horvath, S., Bruist, M., Simon, M. I. & Dervan, P. B. (1987) *Science* **238**, 1129–1132.
3. Glasgow, A. C., Bruist, M. F. & Simon, M. I. (1989) *J. Biol. Chem.* **264**, 10072–10082.
4. Glasgow, A. C., Hughes, K. T. & Simon, M. I. (1989) in *Mobile DNA*, eds. Berg, D. & Howe, M. (Am. Soc. Microbiol., Washington, DC).
5. Hughes, K. T., Youderian, P. & Simon, M. I. (1988) *Genes Dev.* **2**, 937–948.
6. Sluka, J. (1988) Ph.D. Thesis (California Institute of Technology, Pasadena, CA), pp. 66–72.
7. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Mathews, B. W. (1982) *Nature (London)* **298**, 718–723.
8. Jordan, S. R. & Pabo, C. O. (1988) *Science* **242**, 839–899.
9. Ohlendorf, D. H., Anderson, W. F., Takeda, Y. & Mathews, B. W. (1983) *J. Biol. Mol. Struct. Dyn.* **2**, 553–563.
10. Reidhaar-Ohlson, J. T. & Sauer, R. T. (1988) *Science* **241**, 53–57.