CMLS, Cell. Mol. Life Sci. 55 (1999) 472–486
1420-682X/99/030472-15 $ 1.50 + 0.20/0
© Birkhäuser Verlag, Basel, 1999

CMLS Cellular and Molecular Life Sciences

## Research Article

# Repetitive use of a phosphate-binding module in DNA polymerase β, Oct-1 POU domain and phage repressors

**K. Yura, M. Shionyu, K. Kawatani and M. Gō***

Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602 (Japan), Fax + 81 52 789 2977, e-mail: go@bio.nagoya-u.ac.jp

Received 11 January 1999; accepted 13 January 1999

**Abstract.** Motifs for sequence specific-protein-DNA interactions, such as helix-turn-helix, zinc finger and leucine zipper, are now better understood as a result of extensive studies of three-dimensional (3D) structures of transcription factors. On the other hand, little attention has been paid to motifs for sequence nonspecific binding, namely DNA-phosphate binding. To address the question whether different transcription factors and DNA manipulation enzymes, that is enzymes that work on DNA, share a similar mode of phosphate binding, we surveyed interactions between DNA and protein module, a structural unit of a globular protein. We analyzed the modular organization of DNA polymerase β and found that residues making contact with DNA phosphates were localized to five modules. Structural comparison of these phosphate-binding modules against others in transcription factors and DNA manipulation enzymes revealed that DNA polymerase β, the Oct-1 POU domain, 434 Cro and the Arc repressor have a phosphate-binding module with 3D structures similar to one another. This newly detected module, the phosphate-binding helix-turn-helix (pbHTH) module, named for its function and 3D structure, interacts with DNA by (i) making hydrogen bonds between a DNA phosphodiester oxygen and an amino hydrogen of the main chain located at the N-terminus of a C-terminal α-helix, and (ii) making electrostatic interactions between DNA phosphates and side chains of lysine or arginine. Finding structurally and functionally similar phosphate-binding units in different transcription factors and DNA manipulation enzymes suggests that shuffling of modules is not limited to the DNA base-recognition motif. Phosphate-binding modules are apparently also shuffled in DNA-binding proteins.

**Key words.** 434 Cro; Arc repressor; DNA polymerase β; DNA-protein hydrogen bond; HTH module; module shuffling; Oct-1 POU domain; sodium ion.

Protein-DNA interactions are classified as sequence-specific and -nonspecific interactions, the former being mainly achieved by interactions of protein side chains with DNA bases [1]. Four kinds of bases in DNA have unique atom positions that can form hydrogen bonds to or electrostatic interactions with side chain atoms of amino acid residues. Nonspecific interactions with DNA occur by interactions of main or side chains of proteins and DNA phosphates. Sugars of the DNA backbone have the potential to form hydrogen bonds to electron donors. When a main or a side chain of proteins forms a hydrogen bond to a sugar, the bond often bifurcates to a DNA phosphate [2]. As DNA has

*Abbreviations:* 3D = three-dimensional; brHTH = base-recognition helix-turn-helix; CO = backbone carbonyl oxygen; HhH = helix-hairpin-helix; HTH = helix-turn-helix; NH = backbone amino hydrogen; pbHTH = phosphate-binding helix-turn-helix; PDB = Protein Data Bank; RHH = ribbon-helix-helix; RMSD = root mean square deviation; sfHTH = scaffold helix-turn-helix.

* Corresponding author.

phosphodiester oxygens in its backbone, the existence of which does not depend on DNA sequences, proteins interact nonspecifically with DNA phosphates. This interaction often aids in properly locating specific binding sites of the protein onto DNA [3].

Protein-DNA specific interactions have been thoroughly examined, and common three-dimensional (3D) structures in interaction regions have been identified. The well-known DNA-binding motifs are helix-turn-helix (HTH), zinc finger, basic leucine zipper and basic helix-loop-helix motifs. These motifs, which have one $\alpha$-helix inserted into a major groove of DNA 'read' DNA sequences [4]. Sequence-specific interaction on a minor groove found in purine repressor [5] utilizes the $\alpha$-helix, and in TATA box binding protein [6] and integration host factor [7] both utilize $\beta$-sheets.

Unlike DNA sequence-specific binding motifs, a structural similarity in DNA nonspecific binding regions has not been found. Pabo and Sauer [3] remarked that 'there does not appear to be any simple "rule" or pattern describing which residues are used for backbone contact.'

We reported that the HTH motif is composed of a single module [8]. A globular domain was found to be partitioned into compact substructures, that is modules [9]. The average length of a module is about 15 residues. Correspondence of module boundaries and intron positions of the genes of hemoglobin, lysozyme and other proteins was reported [9–16]. Thus, modules are likely to be units of protein evolution. Modules are also units of function, as was shown experimentally by RNase activity in single modules [17], by exchanging modules of hemoglobin $\alpha$- and $\beta$-subunits, resulting in a swap of the function [18, 19], and by exchanging a single module in isocitrate dehydrogenase that binds nicotinamide adenine dinucleotide phosphate (NADP) to a module in NAD-specific isopropylmalate dehydrogenase, resulting in exchange of coenzyme specificity without loss of

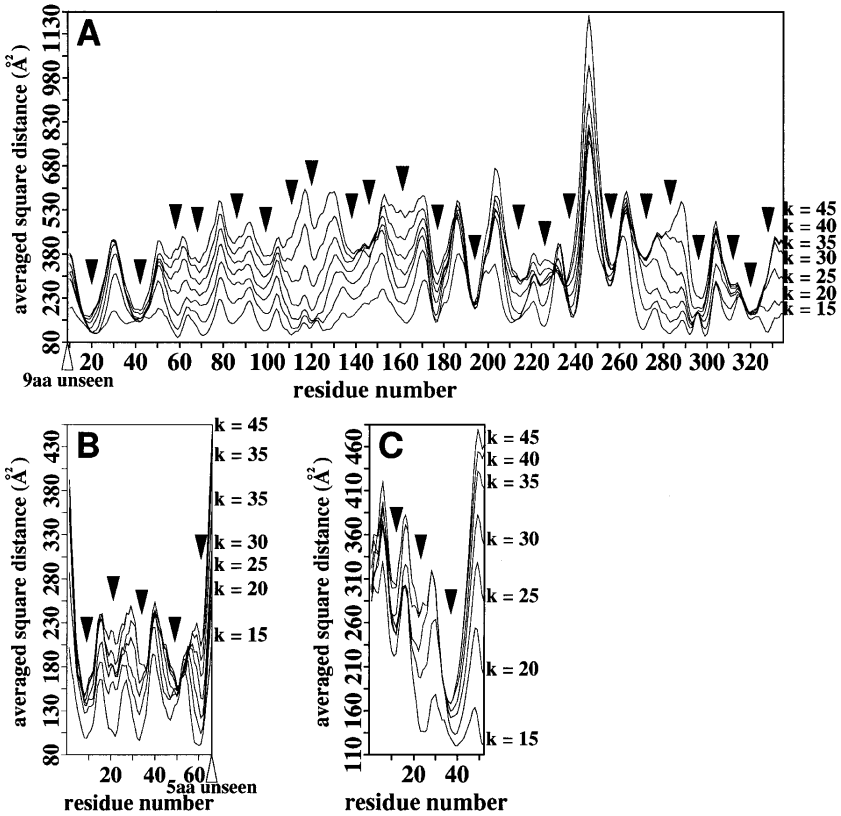


Figure 1. Centripetal profiles of (*A*) human DNA polymerase $\beta$ (PDB code: 1BPY) [57], (*B*) 434 phage 434 Cro protein (3CRO) [28] and (*C*) P22 phage Arc repressor (1PAR) [29]. Module boundaries are described by black arrowheads. Each profile has seven lines that corresponded to seven values of $k$, a window length of C$\alpha$ atoms around $i$th C$\alpha$ atom. See module boundary determination in 'Materials and methods'. Nine residues at the N-terminus of DNA polymerase $\beta$ and five residues at the C-terminus of 434 Cro were not observed by X-ray crystallographic analysis [57, 28].

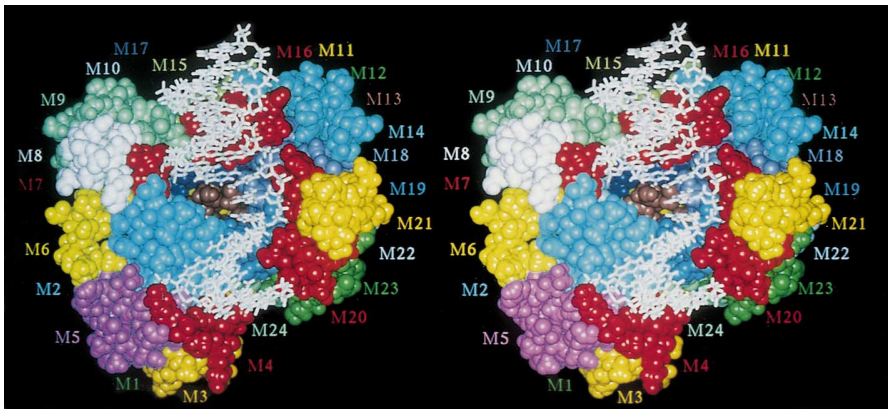


Figure 2. Module organization of DNA polymerase $\beta$ (PDB code: 1BPY) [57]. DNA is depicted by the stick model in white. The active site residues are on brown module M13 and dark blue module M17 at the center. Phosphodiester oxygens are mainly bound to modules in red, and in purple at the bottom left corner.

activity [20]. In HTH proteins, common HTH modules recognize specific DNA sequences [8].

Is there a shared structural unit for DNA nonspecific binding, namely DNA-phosphate binding? We used module structure, one of the structural units in protein, of transcription factors and DNA manipulation enzymes including modification, restriction, repair, replication enzymes and polymerases to address the question. It would be ideal to find 3D structures of proteins that bind DNA without sequence specificity. DNA polymerase $\beta$ is one of the best targets for analyzing how a protein binds nonspecifically to DNA. DNA polymerase $\beta$ was found to interact with DNA mostly by sequence-nonspecific hydrogen bonds [21]. As phosphate-binding modules were found in DNA polymerase $\beta$, we searched for similar modules in other transcription factors and DNA manipulation enzymes.

## Materials and methods

**Module boundary determination.** Module boundaries were determined by centripetal and extension profiles [12, 22]. Candidates for module boundaries were obtained by locating local minima of $F_i$. $F_i$ is an index of the centripetal character of the $i$th residue, calculated by a mean-square distance between the C$\alpha$ atom of the $i$th residue and C$\alpha$ atoms within a window of $(2k+1)$ residues ($k = 15$, 20, 25, 30, 35, 40 and 45) centered by the $i$th residue. The centripetal profile detects the local center of a protein. The candidates were confirmed by an extension profile that checked the extendedness of the chain around the candidate module boundaries. N- and C-terminal module boundaries were determined separately because one could not set a large $k$ at the terminus. Final determination of module boundaries was automated (Y. Sato et al., unpublished). Treatment of the N- and C-termini of a protein resulted in detection of a new module boundary in 434 Cro. It was reported to have five modules by Yura et al. [8], but the C-terminal module decomposed into two (fig. 1).

**Defining a protein-DNA contact.** Heavy atoms in protein and DNA were defined to be in contact when they were located within the range of 4.0 Å distance. An amino acid residue and a nucleotide were defined to be in contact when three or more pairs of contact were found between them. These cutoffs were set to choose residues that were obviously in contact with DNA. DNAs were divided into two parts: base and phosphate plus sugar. When atoms in a residue were in contact with a base, the residue was defined to be in contact with the DNA base, even when other atoms in the same residue were in contact with phosphate and/or sugar. When atoms in a residue were in contact with phosphate and/or sugar but not with base, the residue was defined to be in contact with DNA phosphate. A module that contained a base-contact residue was called a base-recognition module, and that containing only a phosphate-contact residue was called a phosphate-binding module.

**Superimposition of modules.** The 3D structure of modules that interact with DNA phosphate in all the DNA-binding proteins in the Protein Data Bank [23] were compared by superimposing C$\alpha$ atoms. If the length of amino acid sequences differed between two modules, all possible sequential correspondences of C$\alpha$ atoms of the two modules were considered. Details were described in Yura et al. [8].

In the present work, modules with helices on both termini were treated. Cutoff for similarity was determined to be around 2.0 Å. In 574 proteins of sequence identity, less than 30% in the Protein Data Bank [23], root mean square deviation (RMSD) calculation resulted in 5.04 Å on average, with a standard deviation of 1.62 Å, when 1387 modules with a helices on both termini were randomly chosen and superimposed in every possible pair (data not shown). RMSD $\lesssim$ 2.0 Å falls within 3% of the entire number of module pairs. Therefore, two modules with RMSD $<$ 2.0 Å are similar, with a statistical significance at the 3% level. Wintjens et al. [24] evaluated the significance of RMSD of a protein fragment with two $\alpha$-helices connected by a $\beta$-turn structure and found that 2.5 Å was the reasonable cutoff for classification of peptides having 20-amino acid residues.

**Analysis of the module-DNA hydrogen bond.** Since locations of hydrogens are usually not given in the Protein Data Bank [23], they were geometrically calculated. Hydrogen bond analysis was based on the method of Baker and Hubbard [25].

## Results

**DNA phosphate-binding and ion-binding modules in DNA polymerase $\beta$.** A black arrowhead in figure 1A indicates a module boundary of DNA polymerase $\beta$. There are 24 modules in DNA polymerase $\beta$. The length of a module in DNA polymerase $\beta$ is about 14 residues on average. Module organization in the 3D structure of DNA polymerase $\beta$ is depicted in figure 2. DNA polymerase $\beta$ has been subdivided into four subdomains, based on structural comparisons among nucleic acid polymerases [21]. The N-terminal 8-kD subdomain is composed of five modules, the fingers subdomain of five modules, the palm subdomain of seven and the thumb subdomain of seven (fig. 3). The subdomain junctions correspond approximately to module boundaries.

DNA polymerase $\beta$ made contacts with DNA, as shown in table 1. An atom of DNA polymerase $\beta$ and that of DNA were defined to be in contact when they are located within a 4.0-Å distance. The contact was localized to seven modules, M2, M4, M5, M7, M16, M19 and M20. Modules M4, M5, M7, M16 and M20 make contact with DNA phosphates and sugars, but

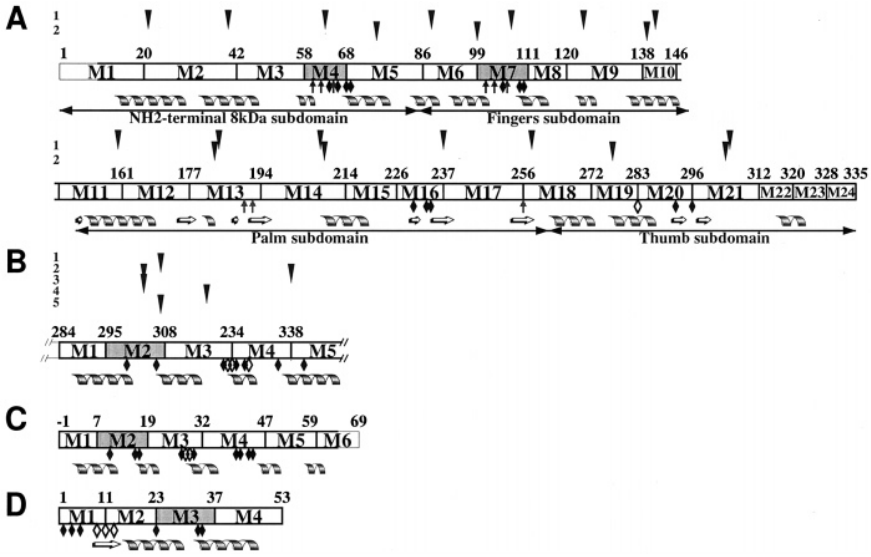


Figure 3. Diagrams of module organization of (*A*) human DNA polymerase $\beta$, (*B*) Oct-1 POU-specific domain, (*C*) 434 phage 434 Cro protein and (*D*) P22 phage Arc repressor. Shadowed boxes are similar phosphate-binding HTH modules. Coils below the boxes are $\alpha$-helices, and thick arrows are $\beta$-strands. Black diamonds on boxes are DNA phosphodiester oxygen hydrogen-bonding sites, and white diamonds are DNA base hydrogen-bonding sites. Intron positions of DNA polymerase $\beta$ and the Oct-1 POU-specific domain are indicated by arrows over the boxes. Of DNA polymerase $\beta$, the introns in row 1 are from human DNA polymerase $\beta$ (accession number: U10526) [43], and the ones in row 2 are from human terminal deoxynucleotidyl transferase (M20703) [44]. Of the Oct-1 POU domain, the intron in row 1 is from human and mouse Oct-2 (X81030, X81031) [58], those in row 2 are from *Caenorhabditis elegans* CEH-6 (Z75711) [59], row 3 from rat GHF1(X65364) [60] and from turkey Pit1 (U18928) [61], row 4 from mouse Oct-3 (S235987) [62], and row 5 from *C. elegans* unc-86 (M22363) [63]. Metal-binding residues are indicated by arrows below the boxes of DNA polymerase $\beta$. Two sodium ions are bound independently to modules M4 an M7. One magnesium ion is bound to module M13, and the other magnesium ion is bound to modules M13 and M17. N-terminal nine residues in DNA polymerase $\beta$ and C-terminal five residues in 434 Cro shown in thin lines were not determined by X-ray crystallography [57, 28]. For Oct-1, the POU-specific domain only is shown.

Table 1. Residues of DNA polymerase $\beta$ in contact with DNA.

| DNA polymerase $\beta$ | | DNA | | No. of contacts | | |
|---|---|---|---|---|---|---|
| Module | residue | chain | nt | phosphate | sugar | base |
| M2 | His34 | T | C5 | 0 | 0 | 23 |
| M2 | His34 | D | G1 | 0 | 0 | 3 |
| M2 | Lys35 | D | G1 | 5 | 2 | 0 |
| M2 | Ala38 | D | G1 | 0 | 2 | 1 |
| M2 | Tyr39 | D | G1 | 1 | 3 | 0 |
| M4 | Gly64 | D | T2 | 0 | 7 | 0 |
| M4 | Gly64 | D | C3 | 4 | 1 | 0 |
| M4 | Val65 | D | C3 | 4 | 0 | 0 |
| M4 | Gly66 | D | T2 | 8 | 4 | 0 |
| M4 | Thr67 | D | T2 | 3 | 0 | 0 |
| M4 | Lys68 | D | G1 | 11 | 2 | 0 |
| M4 | Lys68 | D | T2 | 7 | 0 | 0 |
| M5 | Ile69 | D | T2 | 7 | 0 | 0 |
| M7 | Gly105 | P | C8 | 0 | 5 | 0 |
| M7 | Gly105 | P | G9 | 3 | 0 | 0 |
| M7 | Ile106 | P | G9 | 3 | 0 | 0 |
| M7 | Gly107 | P | C8 | 10 | 3 | 0 |
| M7 | Pro108 | P | C8 | 5 | 0 | 0 |
| M7 | Ser109 | P | G7 | 0 | 4 | 0 |
| M7 | Ser109 | P | C8 | 8 | 0 | 0 |
| M7 | Ala110 | P | C8 | 3 | 0 | 0 |
| M16 | Ser229 | T | A11 | 4 | 1 | 0 |
| M16 | Lys230 | T | A11 | 4 | 3 | 0 |
| M16 | Gly231 | T | C10 | 5 | 5 | 0 |
| M16 | Thr233 | T | C10 | 7 | 2 | 0 |
| M16 | Lys234 | T | C10 | 4 | 2 | 0 |
| M19 | Lys280 | T | G6 | 7 | 1 | 5 |
| M19 | Arg283 | T | G6 | 0 | 5 | 5 |
| M19 | Arg283 | T | G7 | 0 | 4 | 1 |
| M20 | Leu287 | T | G7 | 3 | 0 | 0 |
| M20 | Thr292 | T | G7 | 2 | 1 | 0 |
| M20 | Glu295 | T | C8 | 1 | 10 | 0 |

Two atoms within 4.0 Å were considered to be in contact. A residue with fewer than three pairs of contacts was omitted. Chains in the DNA are depicted as D (downstream), T (template) and P (primer).

not with DNA bases. These five modules can be called phosphate-binding modules. Modules M2 and M19 made contact with DNA phosphates and sugars and also with DNA bases. When an amino acid residue contacts a DNA base located deep in a groove, the same residue inevitably contacts DNA phosphates and sugars located at sides of the groove. Therefore, modules M2 and M19 were not regarded as phosphate-binding modules. Modules M4, M5, M7, M16 and M20 were placed around DNA, as if to grasp DNA (fig. 2 in red, except for M5 in purple).

In DNA polymerase $\beta$, two sodium ions and two magnesium ions bind to become an active enzyme. Based on crystal-soaking experiments, sodium ions could be potassium ions [26]. Residues that bind metal ions are depicted in figure 3. One of the sodium ions was exclusively bound to module M4, whereas the other was bound to module M7. Modules M4 and M7 are $Na^+$-binding modules. These two modules are also phosphate-binding modules, and their 3D structures are strikingly similar (fig. 4). One of the magnesium ions was bound to oxygen atoms of two aspartate residues in module M13. The other magnesium ion was bound to the same aspartate residues with different oxygen atoms of the side chains and to an aspartate residue on module M17. Modules M13 and M17 are $Mg^{2+}$-binding modules. Even though these two modules had different 3D structures, they use oxygen atoms of aspartate residues located close to the C-terminal of the modules to coordinate $Mg^{2+}$.

**A search for a similar phosphate-binding module in DNA-binding proteins.** Thirty-nine transcription factors and DNA manipulation enzymes whose 3D structures were solved with DNA were broken down into modules. Modules that contact DNA phosphates were compared against the five phosphate-binding modules of DNA polymerase $\beta$. As a result, module M2 of the Oct-1 POU domain [27], module M2 of the 434 Cro protein

[28] and module M3 of the Arc repressor [29] were found to have structures similar to modules M4 and M7 of DNA polymerase $\beta$ (figs 5 and 6A). The similarity in 3D structures of those modules was not apparent from their sequence comparison (fig. 6B). DNA polymerase $\beta$ modules M4 and M7 coincided with a helix-hairpin-helix (HhH) sequence motif, one of the DNA-binding motifs first found in endonuclease III [26, 30–32]. An amino acid sequence search, however, did not turn up any structurally or functionally similar unit in the Oct-1 POU domain, 434 Cro or Arc repressor. The pbHTH module turned out to include the HhH motif and also to include modules with a structural and functional similarity without sequence similarity to the HhH motif. Oct-1 and 434 Cro proteins are HTH proteins. The overall structural similarity of the Oct-1 POU-specific domain and 434 Cro became known when the 3D structure of Oct-1 was determined [33]. The Oct-1 POU-specific domain was decomposed into five modules (figs 3 and 5C). Details of the modular organization of Oct-1 will be discussed elsewhere (K. Yura et al., unpublished).

434 Cro was broken down into six modules (fig. 1B). Modules M2 and M4 are phosphate-binding modules, whereas module M3 is a base-recognition module (fig. 3). The spatial arrangement of modules is depicted in figure 5D. 434 Cro and 434 repressor [34] essentially have similar 3D structures. 434 repressor has phosphate-binding modules and a base-recognition module corresponding to those of 434 Cro [8].

The Arc repressor has a DNA-binding ribbon-helix-helix (RHH) motif [35]. DNA bases are recognized by an antiparallel $\beta$-sheet created by dimerization. The monomer broken down into four modules (figs 1C and 5E), and DNA phosphates were mainly bound to modules M1 and M3 (fig. 3). Module M1 was hydrogen-bonded to bases and phosphodiester oxygens, whereas module M3 was hydrogen-bonded only to phosphodiester oxygens.

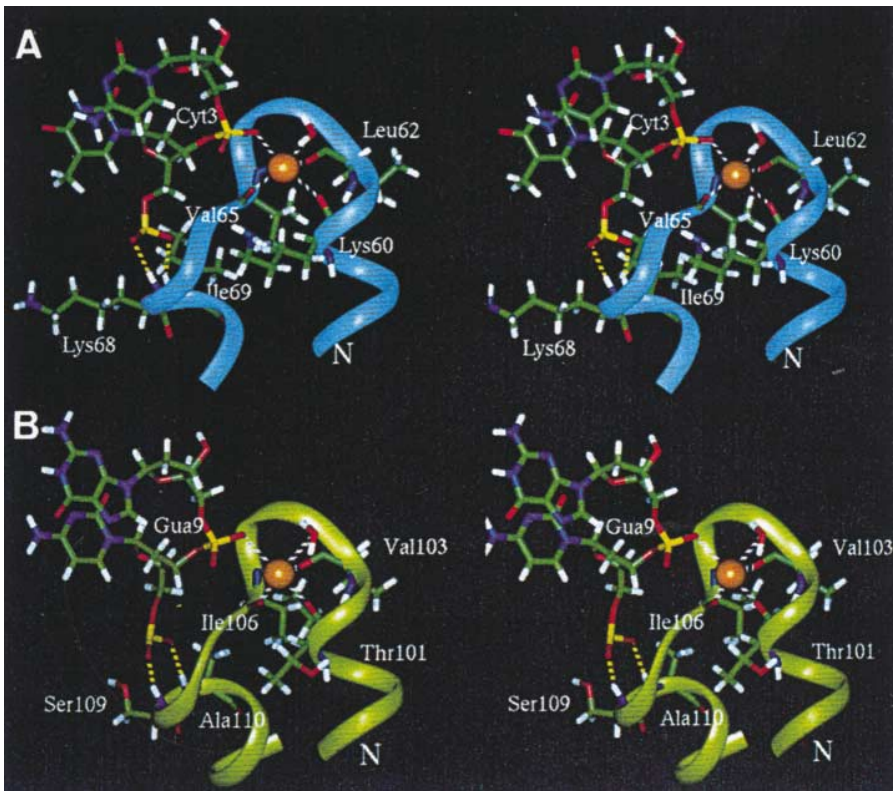


Figure 4. $Na^+$-binding modules in DNA polymerase $\beta$. (*A*) Module M4 and (*B*) module M7. Orange atoms are sodium. Residues that have atoms to coordinate the sodium atom are shown in the stick model. Pink dotted lines indicate coordination. Phosphodiester oxygens of DNA make hydrogen bonds to $\alpha$-helices at the C-terminal, as shown in yellow dotted lines. Residues with donor and acceptor atoms are also shown in the stick model.

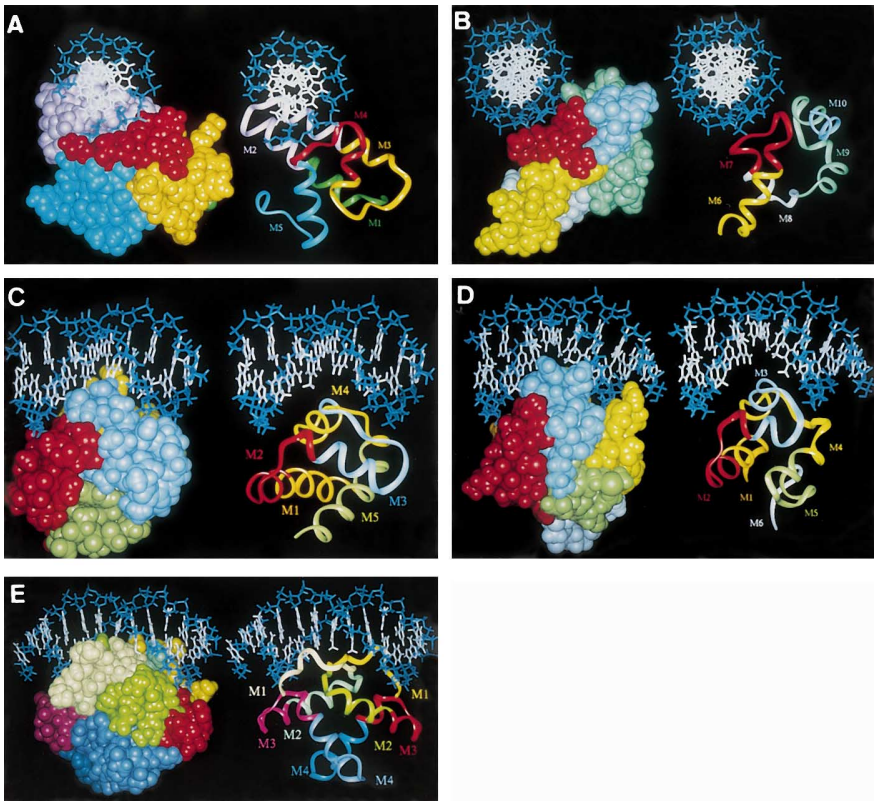


Figure 5. Module organization of (*A*) DNA polymerase $\beta$ from modules M1 to M5, (*B*) DNA polymerase $\beta$ from modules M6 to M10, (*C*) Oct-1 POU-specific domain, (*D*) 434 Cro protein monomer and (*E*) Arc repressor dimer depicted by space filling (left) and tube (right) models. Each protein is colored by a module. DNAs are depicted by the stick model. DNA backbones are colored blue, and bases are in white.

We reported that 434 Cro protein had three modules, M2, M3 and M5, that were structurally similar (table 2). Module M3 was almost equivalent to the HTH motif [8]. Hence, we named modules M4 and M7 of DNA polymerase $\beta$, M2 of the Oct-1 POU domain, M2 of 434 Cro and M3 of the Arc repressor phosphate-binding HTH (pbHTH) modules.

**Characteristics of pbHTH modules.** The five modules—M4 and M7 of DNA polymerase $\beta$, M2 of the Oct-1 POU domain, M2 of 434 Cro and M3 of the Arc repressor—bound a DNA phosphate in a similar manner and all used backbone amino hydrogens (NHs) at the N-terminus of the C-terminal $\alpha$-helix of the module. An $\alpha$-helix is formed by hydrogen bonds between a backbone carbonyl oxygen (CO) belonging to $i$th residue and an NH of $i+3$rd residue. At the N-terminus of the $\alpha$-helix, however, three NHs are free from making hydrogen bonds with CO atoms. Free NH has the potential to form hydrogen bonds with a hydrogen acceptor. DNA phosphodiester oxygen formed a hydrogen bond to one of the free NHs in these five modules.

In modules M4 and M5 of DNA polymerase $\beta$, there were three free NHs, namely Thr67, Lys68 and Ile69, at the N-terminus of the C-terminal $\alpha$-helix of module M4. NHs of Lys68 and Ile69 formed hydrogen bonds to phosphodiester oxygens of thymine. In addition to these hydrogen bonds, main-chain NHs of Gly64 and Gly66 and the side chain of Lys68 hydrogen bonded to DNA phosphodiester oxygens (table 3; fig. 7A).

In module M7 of DNA polymerase $\beta$, there were two free NHs, namely NHs of Ser109 and Ala110 at the N-terminus of the C-terminal $\alpha$-helix. The residue at 108 was proline; a side chain of proline is covalently bonded to backbone NH. The NHs of Ser109 and Ala110 hydrogen bonded to phosphodiester oxygens of cytosine. In addition, the main chain NH of Gly105 hydrogen bonded to phosphodiester oxygen of guanine (table 3; fig. 7B).

In the Oct-1 module M2, two out of three NHs of the C-terminal $\alpha$-helix were free. A DNA phosphodiester oxygen was hydrogen-bonded to an NH of Gln306, the first of the three NHs. The side chain of Arg299, a

residue on the other $\alpha$-helix, was also hydrogen-bonded to another phosphodiester oxygen of thymine (table 3; fig. 7C).

In the 434 Cro module M2, all three NHs on the N-terminus of the C-terminal $\alpha$-helix were free. A DNA phosphodiester oxygen was hydrogen-bonded to an NH of Gln17, the first of the three free NHs. The side chain of Thr16 was also hydrogen-bonded to the phosphate (table 3; fig. 7D).

In the Arc repressor module M3, two of the three NHs, namely the NHs of Val33 and Asn34 on the N-terminus of the C-terminal $\alpha$-helix were free. Both NHs hydrogen-bonded to the phosphodiester oxygen atoms of the same adenine. The NH of Ser35 was assumed to be free, yet it was hydrogen-bonded to O$\gamma$ of Ser32. This side chain-main chain hydrogen bond is a typical Ncap [36] (table 3; fig. 7E).

A common feature in all five modules was the hydrogen bond between the phosphodiester oxygen atoms of DNA and one of the NHs at the N-terminus of the C-terminal $\alpha$-helix. One more hydrogen bond donor was used for DNA binding out of the three NHs and/or side chains. In total, at least two hydrogen bond donors were required. The N-terminus of an $\alpha$-helix is positively charged because of the dipole moment of the helix [37, 38]. This dipole moment likely functions as an attractive force for DNA phosphates.

Three other common features among the pbHTH modules were found when we examined the 3D structure of proteins without DNA and the amino acid sequences of pbHTH modules (fig. 6B). First, the main chain NHs of the C-terminal $\alpha$-helix of pbHTH modules used for DNA phosphodiester binding in complex with DNA formed no hydrogen bonds to other parts of proteins when the proteins did not bind DNA. Typically, there is an Ncap at the N-terminus of an $\alpha$-helix [36]. The NH of the N-terminus of the C-terminal $\alpha$-helix in pbHTH modules was, however, not capped, and presumably formed hydrogen bonds to water molecules, as if waiting for DNA to present itself. Second, the size of a side

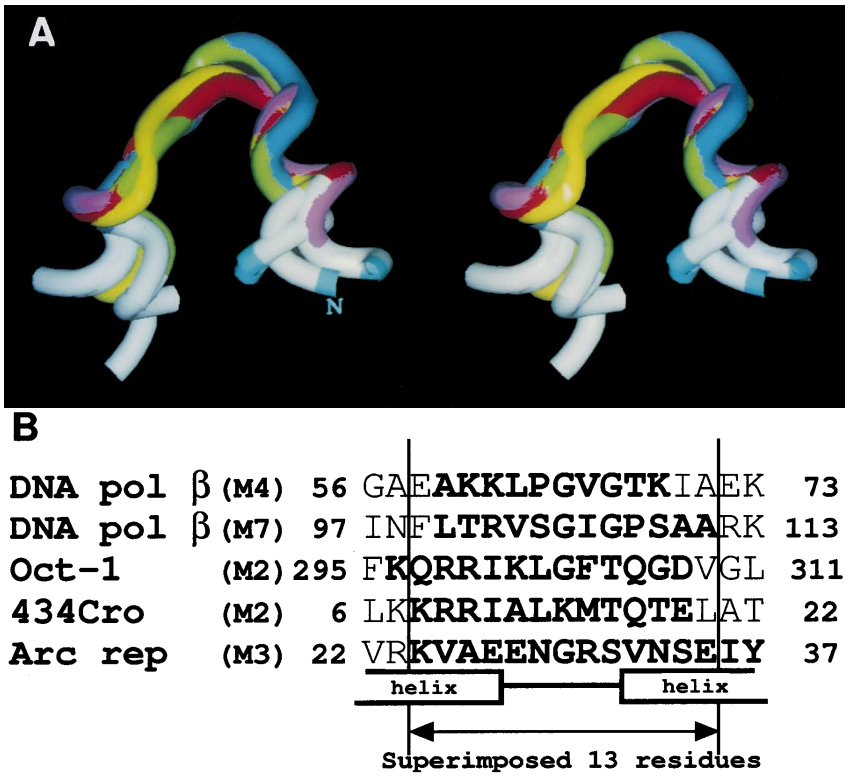


Figure 6. (*A*) Superimposition of phosphate-binding HTH modules. Each module is depicted by the C$\alpha$ trace tube model. White regions are part of the previous and next modules. Light blue indicates the N-terminal side. DNA polymerase $\beta$ modules M4 and M7 are in blue and green, respectively, the Oct-1 module M2 is in purple, the 434 Cro module M2 is in red and the Arc repressor module M3 is in yellow. (*B*) Structural alignment of phosphate-binding HTH modules in (*A*). Amino acid residues in boldface are included in the module. Residues in thin fonts are those of juxtaposing modules.

Table 2. RMSD of HTH modules (Å).

| Module | M4 (pol) | M7 (pol) | M2 (Oct) | M2 (434) | M3 (Arc) | M3 (434) | Function |
|---|---|---|---|---|---|---|---|
| M4 (DNA pol. $\beta$) | | | | | | | phosphate |
| M7 (DNA pol. $\beta$) | 0.60 | | | | | | phosphate |
| M2 (Oct 1) | 2.46 | 2.46 | | | | | phosphate |
| M2 (434 Cro) | 2.38 | 2.32 | 0.66 | | | | phosphate |
| M3 (Arc rep.) | 2.18 | 2.04 | 1.29 | 0.99 | | | phosphate |
| M3 (434 Cro) | 2.08 | 2.06 | 1.53 | 1.30 | 0.61 | | base |
| M5 (434 Cro) | 2.38 | 2.30 | 1.51 | 1.26 | 0.71 | 0.71 | scaffold |

RMSD was calculated by taking the average C$\alpha$ deviation of 13 residues. Correspondence of residues is shown in figure 6B. Phosphate in the column of function indicates that the module is a pbHTH module, base indicates a brHTH module and scaffold indicates an sfHTH module.

Table 3. Hydrogen bonds between phosphate-binding HTH modules and DNA.

| | Protein | | | DNA | | |
|---|---|---|---|---|---|---|
| | module | residue | atom | chain | nt | atom |
| DNA polymerase $\beta$ | M4 | Gly64 | **N** | D | C3 | **O2P** |
| | M4 | Gly66 | **N** | D | T2 | **O2P** |
| | M4 | Lys68 | **N** | D | T2 | **O1P** |
| | M4 | Lys68 | N$\zeta$ | D | G1 | **O1P** |
| | M5 | Ile69 | **N** | D | T2 | **O2P** |
| | M7 | Gly105 | **N** | P | G9 | **O1P** |
| | M7 | Ser109 | **N** | P | C8 | **O2P** |
| | M7 | Ala110 | **N** | P | C8 | **O1P** |
| Oct-1 | M2 | Arg299 | N$\eta$2 | A | T3 | **O1P** |
| | M2 | Gln306 | **N** | A | T3 | **O2P** |
| 434 Cro | M2 | Arg10 | N$\eta$1 | A | T4 | **O2P** |
| | M2 | Thr16 | O$\gamma$1 | B | T3 | **O1P** |
| | M2 | Gln17 | **N** | B | T4 | **O1P** |
| Arc repressor | M3 | Val33 | **N** | E | A4 | **O1P** |
| | M3 | Asn34 | **N** | E | A4 | **O2P** |

The atoms in the left column form hydrogen bonds to atoms in the right column in the same row. Atoms of protein and DNA backbones are printed in boldface to emphasize that most bonds are between backbones. Chain identifications of the DNAs are as the ones in the Protein Data Bank [23].

chain of a residue located one residue before the C-terminal $\alpha$-helix is small. This is perhaps to avoid being a steric hindrance when the module binds to the DNA backbone. The side chain came close to DNA, since the main-chain NH located next to the residue formed a hydrogen bond to a DNA backbone. Third, the pbHTH module has a number of arginines or lysines, residues with positively charged side chains within the module or juxtaposing ones. This seems to contribute to electrostatic interactions with the DNA backbone (fig. 7).

The repetitive appearance of the pbHTH module in DNA polymerase $\beta$, the Oct-1 POU domain, 434 Cro and the Arc repressor indicates (i) that sequence nonspecific binding of DNA is achieved by the protein module and (ii) that the mode of sequence nonspecific binding of DNA is not arbitrary; a common structure used for sequence nonspecific binding was noted. The extent of these findings remains to be surveyed. Our study challenges the view regarding nonspecific interaction of DNA-binding proteins in that there is no rule or pattern for DNA backbone contact in proteins [3].

**Orientation of DNAs against pbHTH modules.** Orientation of the DNA region against module M4 and that against module M7 in DNA polymerase $\beta$ was similar. The orientation of the DNA region against the Oct-1 POU domain module M2, the 434 Cro module M2 and the Arc repressor module M3 was also quite similar. However, the orientation of DNA to DNA polymerase $\beta$ module M7 and that of Arc repressor M3 clearly differed (fig. 8); the locations of NH contributing to hydrogen bonds in modules M4 and M7 of DNA polymerase $\beta$ differ from those in module M2 of the Oct-1 POU domain, in module M2 of 434 Cro and in module M3 of the Arc repressor. In DNA polymerase $\beta$, the second and the third NHs of the N-terminus of

the C-terminal $\alpha$-helix form hydrogen bonds with DNA, whereas in the Arc repressor, the first and second NHs form hydrogen bonds with DNA. Two residues in an $\alpha$-helix are related by a rotation of about 100 degrees along the $\alpha$-helix axis. It is, therefore, geometrically calculated that the difference in orientation of the DNA axis in DNA polymerase $\beta$ module M7 and Arc repressor module M3 is about 80°. The angle of DNA bound to DNA polymerase $\beta$ and that bound to the Arc repressor was actually about 80°, when the modules were superimposed.

The difference in angle of bound DNA in DNA polymerase $\beta$ modules M4 and M7, and the Arc repressor module M3, the 434 Cro module M2 or the Oct-1 POU domain module M2 seems to relate to additional hydrogen bonds in pbHTH modules of DNA polymerase $\beta$. In modules M4 and M7, NHs of Gly64 and Gly105, respectively, form hydrogen bonds to DNA backbones. These hydrogen bonds did not exist in the Oct-1 POU domain module M2, 434 Cro module M2 or Arc repressor module M3 (fig. 7).

## Discussion

**Types of HTH module: base-recognition, phosphate-binding and scaffold modules.** We reported that 434 Cro has three HTH modules, modules M2, M3 and M5 [8]. RMSD of these modules and against the ones found here are given in table 2. Module M2 of 434 Cro is a DNA phosphate-binding module shared by DNA polymerase $\beta$, the Oct-1 POU domain and the Arc repressor. Module M3 recognizes DNA bases. Module M5 seems to have no obvious function [8]. The consequence is that there are at least three types of HTH modules (fig. 9).

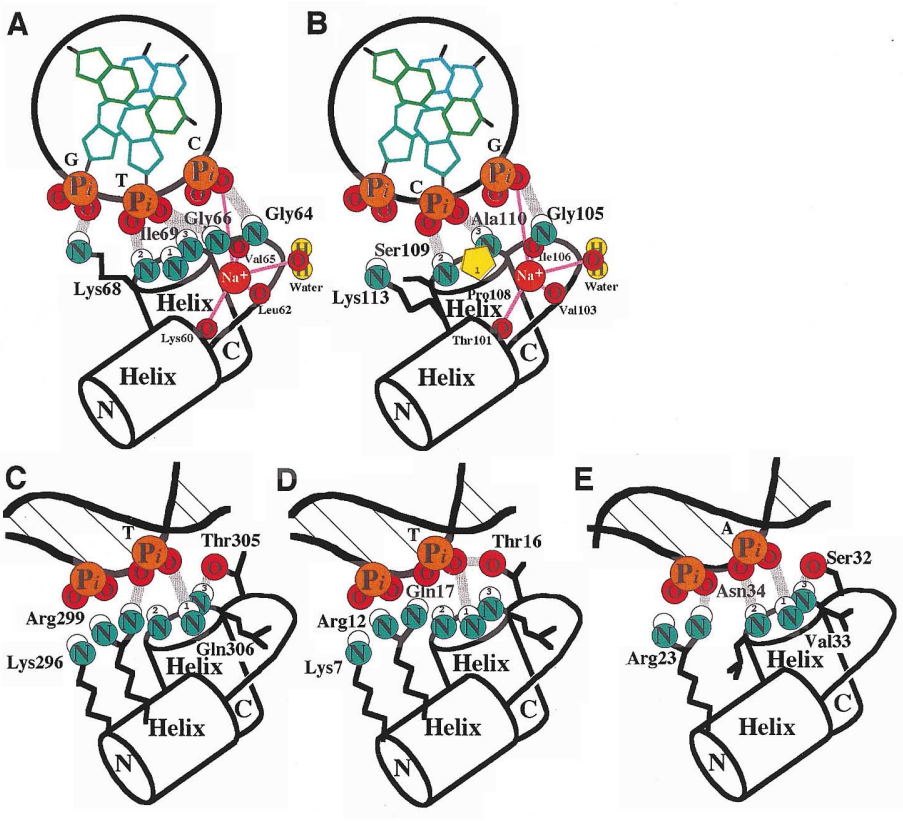


Figure 7. A schematic representation of DNA-pbHTH module interactions. (*A*) DNA polymerase $\beta$ module M4, (*B*) DNA polymerase $\beta$ module M7, (*C*) Oct-1 POU domain module M2, (*D*) 434 Cro module M2 and (*E*) Arc repressor module M3. One or two of the backbone NHs at the N-terminal of the C-terminal $\alpha$-helix form hydrogen bonds to phosphodiester oxygens of DNA (gray lines). Lysines or arginines are located on the N-terminal helix in the case of the Oct-1 POU domain M2, 434 Cro M2 and Arc repressor M3. They seem to interact electrostatically with DNA phosphodiester oxygens. Lysines were found on the C-terminal helix in the case of DNA polymerase $\beta$ M4 and M7. In (*A*) and (*B*), sodium ions are shown in orange. Coordination is indicated by pink dotted lines.

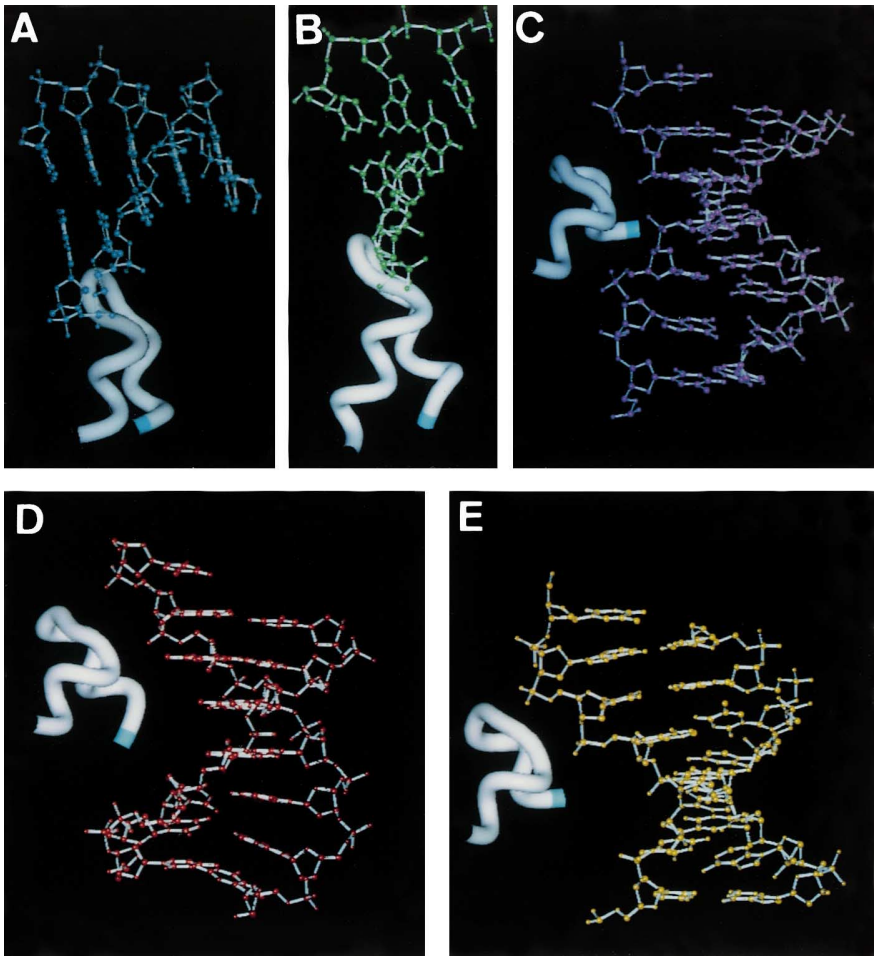


Figure 8. DNA-binding mode against phosphate-binding HTH modules. (*A*) DNA polymerase $\beta$ M4, (*B*) DNA polymerase $\beta$ M7, (*C*) Oct-1 POU domain M2, (*D*) 434 Cro M2 and (*E*) Arc repressor M3. The pbHTH modules are viewed from the same side. DNAs for DNA polymerase $\beta$ are located at positions similar to each other (*A*, *B*). DNAs for the Oct-1 POU domain, 434 Cro and Arc repressor are also located in a position similar to one another (*C*, *D*, *E*), but different from those of DNA polymerase $\beta$.

The first HTH module is the newly identified pbHTH module. The backbone atoms on the N-terminus of the C-terminal $\alpha$-helix are used to form hydrogen bonds with the DNA backbone. The pbHTH module binds DNA, without sequence specificity. The second is a base-recognition HTH (brHTH) module which is almost equivalent to the HTH motif. The C-terminal $\alpha$-helix is inserted into a major groove of DNA, and side chains on the $\alpha$-helix hydrogen bond with DNA bases. The brHTH module binds DNA with sequence specificity. The third HTH module is a scaffold HTH (sfHTH) module with no obvious function; it might possibly serve as a scaffold for other modules. The sfHTH module might function as a foundation to properly allocate other modules [39].

Motif-based predictions of DNA-bound 3D structures of transcription factors and DNA manipulation enzymes have been made [40]. Characteristics of the pbHTH module are also likely to be useful for predicting protein-DNA interactions. Finding a module structurally similar to HTH modules in DNA-binding proteins of which the DNA-complex structure is unknown, and finding a module possessing features of the pbHTH module, but not of brHTH nor sfHTH modules, leads to the prediction that DNA may bind to the module the way it does in DNA polymerase $\beta$, the Oct-1 POU domain, 434 Cro or the Arc repressor.

**Repetitive use of a structurally similar pbHTH module.** A module similar in 3D structure and function to modules M4 and M7 of DNA polymerase $\beta$ was found in the Oct-1 POU domain, 434 Cro and Arc repressor.

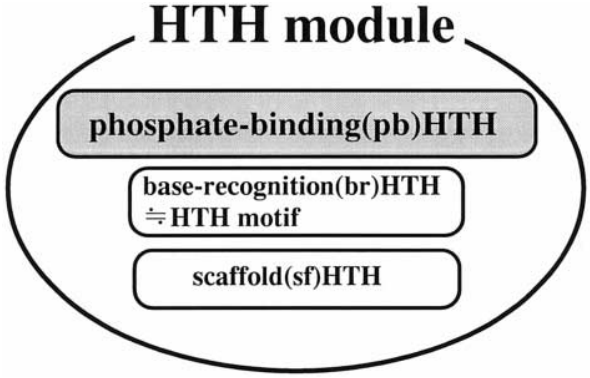


Figure 9. Classification of HTH modules based on their functions. There are at least three functions in HTH modules. HTH modules are functionally grouped into pbHTH, brHTH, and sfHTH modules.

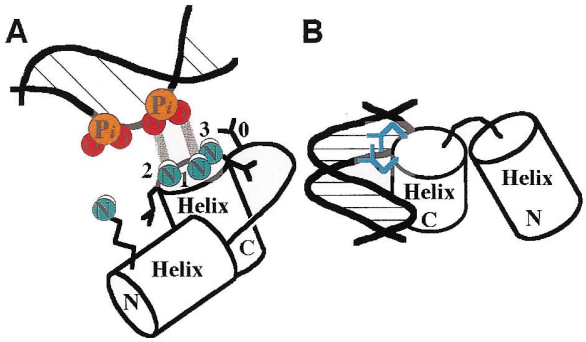


Figure 10. Difference in binding mode of (*A*) pbHTH and (*B*) brHTH of prokaryote to DNA. Side chains of residues that form hydrogen bonds or seemingly strong electrostatic interaction are drawn. The C-terminal $\alpha$-helix of the pbHTH module is almost perpendicular to the DNA backbone, whereas that of brHTH is inserted into a major groove.

However, overall 3D structures of the four proteins differ (fig. 5). This finding supports the notion that the pbHTH module was shuffled into different proteins at the time of protein creation. One of the conceivable mechanisms here is exon shuffling [41, 42]. Of the four proteins studied, introns have been found in the genes of human DNA polymerase $\beta$ [43] and its homolog, human terminal deoxynucleotidyl transferase [44], and Oct-1. DNA polymerase $\beta$ was divided into four subdomains [21]. Two out of 20 introns are located close to two of three subdomain boundaries (fig. 3). The correspondence supports the widely accepted view that domains were indeed combined by exon shuffling [45]. However, the other 18 introns are also located in subdomains of DNA polymerase $\beta$ (fig. 3), and these introns tend to locate close to module boundaries, with some exceptions. Accumulation of protein 3D structures and their genomic structures resulted in a statistically significant correlation of module boundary and intron position [15, 16]. A statistically significant correlation between module boundaries and intron positions in transcription factors was also noted (K. Yura et al., unpublished). In DNA polymerase $\beta$, an intron was found on the N-terminus of pbHTH module M7, and in Oct-1 introns were found on the C-terminus of the

Table 4. Hydrogen bonds between brHTH modules and DNA bases.

| Protein name | 1st res. | 2nd res. |
|---|---|---|
| Catabolite gene activator | - | E181(O$\varepsilon$2)–$C_5$(N4) |
| Hin recombinase | - | S174(O$\gamma$)–$A_{10}$(N7) |
| $\lambda$ Repressor | Q44(O$\varepsilon$1)–$A_4$(N6) | S45(O$\gamma$)–$G_{16}$(N7) |
| Purine repressor | - | T16(O$\gamma$1)–$A_6$(N6) |
| 434 Cro | Q28(O$\varepsilon$1)–$A_5$(N6) | Q29(N$\varepsilon$2)–$G_{16}$(O6) |
| Lac repressor | Y17(O$\eta$)–$T_4$(O4) | Q18(O$\varepsilon$1)–$C_5$(N4) |

'1st res.' and '2nd res.' indicate the first and second residues of the C-terminal $\alpha$-helix of the brHTH module, respectively. In each column, the left side is a protein residue (atom) and the right side is a DNA nucleotide (atom). Atoms in parenthesis are a hydrogen donor or an acceptor. Protein-DNA hydrogen bonds were calculated based on the PDB entry of 1CGP [53] for catabolite gene activator, 1HCR [54] for hin recombinase, 1LRD [55] for $\lambda$ repressor, 1PNR [5] for purine repressor, 3CRO [28] for 434 Cro and 1LCC [56] for lac repressor. Hydrogen bonds were calculated by the method of Baker and Hubbard [25].

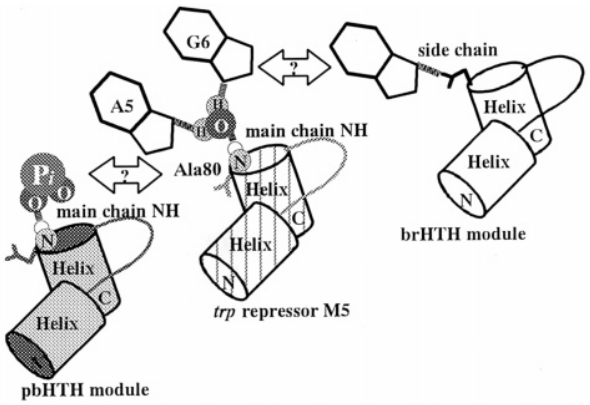


Figure 11. Schematic drawing to emphasize characteristics of the brHTH module in the *trp* repressor. The brHTH module in the *trp* repressor forms hydrogen bonds to DNA bases, using protein backbone NH via the water molecule. The manner in which a protein backbone contributes to hydrogen bonds between protein and DNA is similar to the case of the pbHTH module.

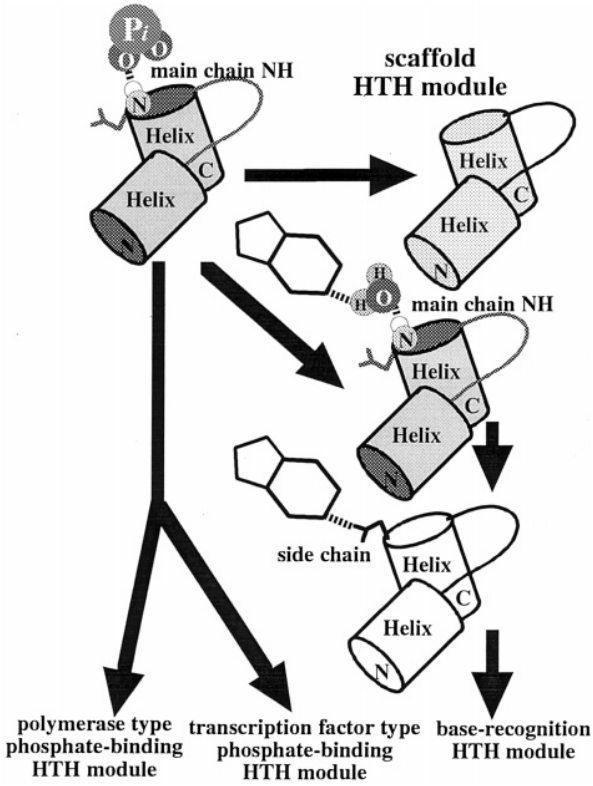


Figure 12. Putative functional divergence of HTH modules. A dark gray module is the pbHTH module a white module is the brHTH module and a light gray one is the sfHTH module. Initial pbHTH module hydrogen-bonds with phosphodiester oxygen by a main-chain NH. A brHTH module might be modified out of the pbHTH module by changing the hydrogen acceptor from a phosphate to a base with help from a water molecule. A side chain probably took on the roles of NH and water molecule, at the last stage. An sfHTH module emerged as a result of recruitment of the pbHTH module or brHTH module for a foundation for other functional modules to make fused modules, as a stable globular domain.

pbHTH module M2 in human, mouse and nematode genes (fig. 3).

In 434 Cro, two phosphate-binding modules—M2 and M4—sandwich the base-recognition module M3 (fig. 5D). This sandwich arrangement seems to be an effective mode for interacting with DNA. The DNA backbone seems to be held and pulled toward the protein by the two phosphate-binding modules. As a result, the base-recognition module located between the phosphate-binding modules seems to be inserted into a major groove of DNA and interacts with DNA bases. The division of labor in 434 Cro, namely base recognition by module M3 and phosphate binding by modules M2 and M4, could be an evolutionary remnant of the functional combination of modules. The combination of two different phosphate-binding modules and a base-recognition module as well as scaffold modules possibly created 434 Cro.

Module shuffling could shuffle a function to different proteins; in this case, a function to bind phosphates. Isolated and incorporated modules are likely to have similar functions. Function of an isolated module was described by Yanagawa et al. [17]. An isolated module of barnase, one of the RNases, cleaved RNA. Incorporation of a module into different combinations of modules may not affect the 3D structure of the module. Ikura et al. [46] showed by nuclear magnetic resonance (NMR) measurements that an isolated module had a tendency to take a 3D structure similar to its structure in intact protein. Conformation of a module is mostly determined by intramodule interaction [22]. The pbHTH module may possibly have been incorporated into DNA polymerase $\beta$, the POU-specific domain of the Oct-1 POU domain, 434 Cro and the Arc repressor.

**Putative evolutionary relationship of pbHTH, brHTH and sfHTH modules.** The pbHTH module exists in DNA polymerase $\beta$ and the Oct-1 POU domain of eukaryotes, and 434 Cro and the Arc repressor of prokaryotes. The brHTH module as an HTH motif exists in eukaryotes and prokaryotes [47]. Therefore, pbHTH and brHTH modules probably existed before the divergence of eukaryotes and prokaryotes.

The brHTH and pbHTH modules could be evolutionarily related, for they use residues located at the same position on the C-terminal $\alpha$-helix. The pbHTH module forms hydrogen bonds to DNA phosphodiester oxygens by using at least one of the three residues at the N-terminus of a C-terminal $\alpha$-helix. The brHTH module of prokaryotes forms hydrogen bonds to DNA bases by using one or two residues located at the N-terminus of the C-terminal $\alpha$-helix (table 4). The difference in interaction mode between pbHTH and brHTH modules is (i) that the pbHTH module uses nitrogen atoms of the main chain to form hydrogen bonds, whereas the brHTH module uses atoms of side chains, and (ii) that the C-terminal $\alpha$-helix of the pbHTH module is located perpendicular to the backbone of DNA, whereas the C-terminal $\alpha$-helix of the brHTH module is inserted into a major groove of DNA (fig. 10).

The HTH module of the *trp* repressor has characteristics of both brHTH and pbHTH modules. The DNA complex of the *trp* repressor revealed that water-mediated contacts were the most important contacts for recognizing DNA [48]. One of the most important water-mediated DNA-base recognitions was found at the C-terminal $\alpha$-helix of the HTH motif. The main-chain NH of Ala80, which is located at the N-terminus of the C-terminal $\alpha$-helix in the HTH motif, formed hydrogen bonds to guanine and adenine bases via the water molecule. This HTH motif is a brHTH module because

it recognizes DNA bases [8]. However, the manner in which it carried out the base-recognition function is similar to that for the pbHTH module because it used the backbone NH of the N-terminus of the C-terminal $\alpha$-helix of the module. We believe that the HTH module in the *trp* repressor corresponds to the evolutionary intermediate between pbHTH and brHTH modules (fig. 11).

Which type of HTH module appeared first? Phosphate is a fundamental component of DNA/RNA, but not of proteins. To bind specifically to DNA and RNA, but not to proteins, finding phosphorus atoms is considered to be the most effective. Binding to phosphates could therefore be fundamental to interacting with DNA and RNA. During the course of prebiological evolution, the pbHTH modules might have appeared first, since the pbHTH module binds to phosphates of DNA. Some pbHTH modules might have acquired the potential to bind to DNA bases, using a water molecule, as in the case of the *trp* repressor. Those pbHTH modules could become brHTH modules by using their side chains instead of a water molecule to interact with the nucleotide. That could be the evolutionary process of the brHTH module from pbHTH module. We cannot rule out the possibility that the brHTH module appeared first. The sfHTH modules might emerge as a result of recruitment from the pbHTH or brHTH module. The sfHTH modules perhaps at first interacted with DNA and RNA, but possibly lost the ability and specialized as scaffolding. The sfHTH module was probably needed to stabilize a globular domain (fig. 12).

DNA is apparently a latecomer in evolution, compared with RNA [49]. The first pbHTH and brHTH modules could interact with RNA. There is no direct evidence of HTH module-RNA interaction, but several lines of evidence, such as homeodomain-RNA interaction [50], HTH motif-like structures in L11 ribosomal protein [51] and glutamyl-transfer RNA (tRNA) synthetase [52] suggest the possibility of RNA-HTH module interaction.

*Acknowledgments.* We thank M. Ohara for helpful comments on the manuscript. This work was supported by Grants-in-Aid for Scientific Research on Priority Areas, for Scientific Research (B) and for Encouragement of Young Scientists from the Ministry of Education, Science, Sports and Culture of Japan.

1 Choo Y. and Klug A. (1997) Physical basis of a protein-DNA recognition code. Curr. Opin. Struct. Biol. **7:** 117–125
2 Mandel-Gutfreund Y., Schueler O. and Margalit H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. J. Mol. Biol. **253:** 370–382
3 Pabo C. O. and Sauer R. T. (1992) Transcription factors: structural families and principles of DNA recognition. Ann. Rev. Biochem. **61:** 1053–1095
4 Harrison S. C. (1991) A structural taxonomy of DNA-binding domains. Nature **353:** 715–719
5 Schumacher M. A., Choi K. Y., Zalkin H. and Brennan R. G. (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science **266:** 763–770
6 Tan S. and Richmond T. J. (1998) Eukaryotic transcription factors. Curr. Opin. Struct. Biol. **8:** 41–48
7 Rice P. A. (1997) Making DNA do a U-turn; IHF and related proteins. Curr. Opin. Struct. Biol. **7:** 86–93
8 Yura K., Tomoda S. and Gō M. (1993) Repeat of a helix-turn-helix module in DNA-binding proteins. Protein Eng. **6:** 621–628
9 Gō M. (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature **291:** 90–92
10 Gō M. (1983) Modular structural units, exons and function in chicken lysozyme. Proc. Natl. Acad. Sci. USA **80:** 1964–1968
11 Gō M. (1985) Protein structures and split genes. Adv. Biophys. **19:** 91–131
12 Gō M. and Nosaka M. (1987) Protein architecture and the origin of intron. Cold Spring Harbor Symp. Quat. Biol. **52:** 915–924
13 Gilbert W. and Glynias M. (1993) On the ancient nature of introns. Gene **135:** 137–144
14 Tittiger C., Whyard S. and Walker V. K. (1993) A novel intron site in the triosephosphate isomerase gene from the mosquito *Culex tarsalis*. Nature **361:** 470–472
15 Gō M. and Noguti T. (1995) Putative origin of introns deduced from protein anatomy. In: Tracing Biological Evolution in Protein and Gene Structures, pp. 229–235, Gō M. and Schimmel P. (eds), Elsevier, Amsterdam
16 de Souza S. J., Long M., Schoenbach L., Roy S. W. and Gilbert W. (1996) Intron positions correlate with module boundaries in ancient proteins. Proc. Natl. Acad. Sci. USA **93:** 14632–14636
17 Yanagawa H., Yoshida K., Torigoe C., Park J.-S., Sato K., Shirai T. et al. (1993) Protein anatomy: functional roles of barnase module. J. Biol. Chem. **268:** 5861–5865
18 Wakasugi K., Ishimori K., Imai K., Wada Y. and Morishima I. (1994) 'Module' substitution in hemoglobin subunits. J. Biol. Chem. **269:** 18750–18756
19 Inaba K., Ishimori K., Imai K. and Morishima I. (1998) Structural and functional effects of pseudo-module substitution in hemoglobin subunits. New structural and functional units in globin structure. J. Biol. Chem. **273:** 8080–8087
20 Yaoi T., Miyazaki K., Oshima T., Komukai Y. and Gō M. (1996) Conversion of the coenzyme specificity of isocitrate dehydrogenase by module replacement. J. Biochem. (Tokyo) **119:** 1014–1018
21 Pelletier H., Sawaya M. R., Kumar A., Wilson S. H. and Kraut J. (1994) Structures of ternary complexes of rat DNA polymerase $\beta$, a DNA template-primer, and ddCTP. Science **264:** 1891–1903
22 Noguti T., Sakakibara H. and Gō M. (1993) Localization of hydrogen-bonds within modules in barnase. Proteins: Struct. Funct. Genet. **16:** 357–363
23 Bernstein F. C., Koetzle T. F., Williams G. J. B., Meyer Jr E. F., Brice M. D., Rodgers J. R. et al. (1977) The protein databank: a computer-based archival file for macromolecular structures. J. Mol. Biol. **112:** 535–542
24 Wintjens R. T., Rooman M. J. and Wodak S. J. (1996) Automatic classification and analysis of $\alpha\alpha$-turn motifs in proteins. J. Mol. Biol. **255:** 235–253
25 Baker E. N. and Hubbard R. E. (1984) Hydrogen bonding in globular proteins. Progr. Biophys. Mol. Biol. **44:** 97–179
26 Pelletier H., Sawaya M. R., Wolfle W., Wilson H. and Kraut J. (1996) Crystal structures of human DNA polymerase $\beta$ complexed with DNA: implications for catalytic mechanism, processivity and fidelity. Biochemistry **35:** 12742–12761
27 Klemm J. D., Rould M. A., Aurora R., Herr W. and Pabo C. O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell **77:** 21–32

28 Mondragon A. and Harrison S. C. (1991) The phage 434 Cro/OR1 complex at 2.5 Å resolution. J. Mol. Biol. **219:** 321–334
29 Raumann B., Rould M. A., Pabo C. O. and Sauer R. T. (1994) DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. Nature **367:** 754–757
30 Thayer M. M., Ahern H., Xing D., Cunningham R. and Tainer J. A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. EMBO J. **14:** 4108–4120
31 Seeberg E., Eide L. and Bjørås M. (1995) The base excision repair pathway. Trends Biochem. Sci. **20:** 391–397
32 Doherty A. J., Serpell L. C. and Ponting C. P. (1996) The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. Nucl. Acids Res. **24:** 2488–2497
33 Assa-Munt N., Mortishire-Smith R. J., Aurora R., Herr W. and Wright P. E. (1993) The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage $\lambda$ repressor DNA-binding domain. Cell **73:** 193–205
34 Aggarwal A. K., Rodgers D. W., Drottar M., Ptashne M. and Harrison S. C. (1988) Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science **242:** 899–907
35 Raumann B. E., Brown B. M. and Sauer R. T. (1994) Major groove DNA recognition by $\beta$-sheets: the ribbon-helix-helix family of gene regulatory proteins. Curr. Opin. Struct. Biol. **4:** 36–43
36 Richardson J. S. and Richardson D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices. Science **240:** 1648–1652
37 Hol W. G. J., van Duijnen P. T. and Berendsen H. J. C. (1978) The $\alpha$-helix dipole and the properties of proteins. Nature **273:** 443–446
38 Warwicker J. and Watson H. C. (1982) Calculation of the electric potential in the active site cleft due to $\alpha$-helix dipoles. J. Mol. Biol. **157:** 671–679
39 Yura K. and Gō M. (1995) Helix-turn-helix module distribution and module shuffling. In: Tracing Biological Evolution in Protein and Gene Structures, pp. 187–195, Gō M. and Schimmel P. (eds), Elsevier, Amsterdam
40 Ramakrishnan V., Finch J. T., Graziano V., Lee P. L. and Sweet R. M. (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature **362:** 219–223
41 Gilbert W. (1978) Why genes in pieces? Nature **271:** 501
42 Long M., de Souza S. J., Rosenberg C. and Gilbert W. (1996) Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. Proc. Natl. Acad. Sci. USA **93:** 7727–7731
43 Chyan Y. J., Ackerman S., Shepherd N. S., McBride O. W., Widen S. G., Wilson S. H. et al. (1994) The human DNA polymerase beta gene structure. Evidence of alternative splicing in gene expression. Nucleic Acids Res. **22:** 2719–2725
44 Riley L. K., Morrow J. K., Danton M. J. and Coleman M. S. (1988) Human terminal deoxyribonucleotidyltransferase: molecular cloning and structural analysis of the gene and 5′ flanking region. Proc. Natl. Acad. Sci. USA **85:** 2489–2493
45 Doolittle R. F. (1995) The multiplicity of domains in proteins. Ann. Rev. Biochem. **64:** 287–314
46 Ikura T., Gō N., Kohda D., Inagaki F., Yanagawa H., Kawabata M. et al. (1993) Secondary structural features of modules M2 and M3 of barnase in solution by NMR experiment and distance geometry calculation. Proteins Struct. Funct. Genet. **16:** 341–356
47 Luisi B. (1995) DNA-protein interaction at high resolution. In: DNA-Protein: Structural Interactions, pp. 1–48, Lilley D. M. (ed), IRL Press at Oxford University Press, Oxford
48 Sigler P. B. (1992) The molecular mechanism of *trp* repression. In: Transcriptional Regulation, pp. 475–499, McKnight S. L. and Yamamoto K. R. (eds), Cold Spring Harbor Laboratory Press, New York
49 Joyce G. F. and Orgel L. E. (1993) Prospects for understanding the origin of the RNA world. In: The RNA World, pp. 1–26, Gesteland R. F. and Atkins J. F. (eds), Cold Spring Harbor Laboratory Press, New York
50 Dubnau J. and Struhl G. (1996) RNA recognition and translational regulation by a homeodomain protein. Nature **379:** 694–699
51 Hinck A. P., Markus M. A., Huang S., Grzesiek S., Kustonovich I., Draper D. E. et al. (1997) The RNA binding domain of ribosomal protein L11: three-dimensional structure of the RNA-bound form of the protein and its interaction with 23S rRNA. J. Mol. Biol. **274:** 101–113
52 Nureki O., Vassylyev D. G., Katayanagi K., Shimizu T., Sekine S., Kigawa T. et al. (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. Science **267:** 1958–1965
53 Schultz S. C., Shields G. C. and Steitz T. A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science **253:** 1001–1007
54 Feng J. A., Johnson R. C. and Dickerson R. E. (1994) Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. Science **263:** 348–355
55 Jordan S. R. and Pabo C. O. (1988) Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. Science **242:** 893–899
56 Chuprina V. P., Rullmann J. A., Lamerichs R. M., van Boom J. H., Boelens R. and Kaptein R. (1993) Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. J. Mol. Biol. **234:** 446–462
57 Sawaya M. R., Prasad R., Wilson S. H., Kraut J. and Pelletier H. (1997) Crystal structures of human DNA polymerase $\beta$ complexed with gapped and nicked DNA: evidence for an induced fit mechanism. Biochemistry **36:** 11205–11215
58 Matsuo K., Clay O., Kunzler P., Georgiev O., Urbanek P. and Schaffner W. (1994) Short introns interrupting the Oct-2 POU domain may prevent recombination between POU family genes without interfering with potential POU domain 'shuffling' in evolution. Biol. Chem. Hoppe-Seyler **375:** 675–683
59 Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J. et al. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature **368:** 32–38
60 Theill L. E., Hattori K., Lazzaro D., Castrillo J.-L. and Karin M. (1992) Differential splicing of the *GHF1* primary transcript gives rise to two functionally distinct homeodomain proteins. EMBO J. **11:** 2261–2269
61 Wong E. A., Silsby J. L. and Halawani M. E. E. (1992) Complementary DNA cloning and expression of Pit-1/GHF-1 from the domestic turkey. DNA Cell Biol. **11:** 651–660
62 Okazawa H., Okamoto K., Ishino F., Ishino-Kaneko T., Takeda S., Toyoda Y. et al. (1991) The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. EMBO J. **10:** 2997–3005
63 Finney M., Ruvkun G. and Horvitz H. R. (1988) The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein with a homeodomain and extended similarity to transcription factors. Cell **55:** 757–769