ANTHONY ARTALE
MED LIB NATHAN CUMMINGS CTR (S-46)
MEMORIAL SLOAN KETTERING CANCER CTR
1275 YORK AVENUE
NEW YORK, NY 10021
UNITED STATES

Title: **PROTEIN ENGINEERING**

DB Ref. No.: IRN10409592
ISSN: ISSN02692139
Vol./Issue: 7
Date: 1994
Pages: 1449-1453
Article Title: LEXA REPRESSOR AND IRON UPTAKE REGULATOR FROM ESCHERICHIA COLI
Article Author: HOLM, L, RUTERJAINS H, SCHNARR M, FOGH R, BOELENS, R
Report Number: IRN10409592
Publisher: IRL PRESS,

Protein Engineering vol.7 no.12 pp.1449–1453, 1994

LexA repressor and iron uptake regulator from *Escherichia coli*: new members of the CAP-like DNA binding domain superfamily

Liisa Holm[1], Chris Sander, Heinz Rüterjans[2], Manfred Schnarr[3], Rasmus Fogh[4], Rolf Boelens[4] and Robert Kaptein[4]

European Molecular Biology Laboratory, D-69012, Heidelberg, [2]Institut für Biophysikalische Chemie, Johann Wolfgang Goethe-Universität, Sandhofstrasse, Gebäude 75A, D-60528 Frankfurt am Main, Germany, [3]Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 15, Rue Descartes, 67084 Strasbourg, France and [4]Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[1]To whom correspondence should be addressed

Comparison of structures can reveal surprising connections between protein families and provide new insights into the relationship between sequence, structure and function. The solution structure of LexA repressor from *Escherichia coli* reveals an unexpected structural similarity to a widespread class of prokaryotic and eukaryotic regulatory proteins, which is typified by catabolite gene activator protein (CAP). The use of combined sequence profiles allows the identification of two new prokaryotic members of the superfamily: listeriolysin regulatory protein (PrfA) and ferric uptake regulatory protein (Fur). LexA, PrfA and Fur are the first examples of prokaryotic regulatory proteins in which DNA recognition is mediated by a variant of the classical helix–turn–helix motif, with an insertion in the turn region.

Key words: CAP/helix–turn–helix/iron uptake/LexA repressor/sequence comparison

Introduction

Gene regulation is central to life processes. The structures of several classes of DNA binding domains found in regulatory proteins reveal a variety of designs for recognizing a specific site on DNA (Harrison, 1991). One of the most prolific solutions is the helix–turn–helix (HTH) motif which has already been seen in ~20 structures solved by crystallography or NMR and in diverse structural contexts. A powerful sequence pattern for the HTH motif (Dodd and Egan, 1990) has allowed the identification of putative DNA binding sites in hundreds of regulatory proteins, in many cases in advance of structural confirmation (e.g. biotin repressor; Wilson *et al.*, 1992). A number of recently solved structures of DNA binding domains have been surprising in two respects. First, several DNA binding domains of different overall folding type contain variant HTH motifs. The variants have inserts varying between 1 and 21 residues in the turn region (Ogata *et al.*, 1992; Assa-Munt *et al.*, 1993; Ceska *et al.*, 1993; Clark *et al.*, 1993; Dekker *et al.*, 1993; Leiting *et al.*, 1993; Ramakrishnan *et al.*, 1993; Harrison *et al.*, 1994). In these cases sequence search methods have problems identifying the HTH motif as exemplified by the LexA repressor (Lamerichs *et al.*, 1989; Dodd and Egan, 1990). Second, structural classes are emerging in which similarity extends beyond the presence of an HTH motif to the fold of an entire structural domain (Assa-Munt *et al.*, 1993; Clark *et al.*, 1993; Dekker *et al.*, 1993; Harrison *et al.*, 1994). Such findings have interesting implications for protein function and evolution.

The solution structure of LexA repressor from *Escherichia coli* was recently solved by means of high-resolution NMR (Fogh *et al.*, 1994). LexA represses a number of genes involved in the response to DNA damage (SOS response), including RecA and LexA itself. The DNA binding region of LexA is near the N-terminal (residues 1–84). The C-terminal region (residues 85–202) is involved in the dimerization of the protein, which reinforces the binding of LexA to the operator. Here we show that the DNA binding domain of LexA repressor belongs to a rapidly growing structural superfamily of DNA binding domains typified by catabolite gene activator (CAP) from *E.coli* (Schultz *et al.*, 1991). Similar DNA binding domains are known also in biotin repressor (BirA) from *E.coli* (Wilson *et al.*, 1992), histone H5 (GH5) from chicken (Ramakrishnan *et al.*, 1993), hepatocyte nuclear factor HNF-3 from rat and a heat shock transcription factor (HSF) from yeast (Harrison *et al.*, 1994). Sequence space around and between the structurally known subfamilies is scanned by profile searches. As a result, the superfamily is expanded by two new proteins of yet unknown structure.

Methods

Families of homologous proteins were collected from the Swissprot sequence database (Bairoch and Boeckmann, 1992) using standard tools (Prosite, Bairoch, 1992; Fasta, Lipman and Pearson, 1985; Filter_Fasta, Sander and Schneider, 1991; Blastp, Altschul *et al.*, 1990). The multiple alignments (Higgins *et al.*, 1992) were edited manually as needed. Structural alignments by the program Dali (Holm and Sander, 1993) were used to combine alignments of the different families. The combined alignments were used to generate sequence profiles (Gribskov *et al.*, 1987), where closely related sequences were downweighted using Voronoi weights (Sibbald and Argos, 1990).

Profile searches were used to identify sequences of possible remote homologues (~20% sequence identity) with the same fold as the families of known structure. The hypotheses were scrutinized (i) by comparing consensus patterns of the structure-based multiple alignment and a multiple family alignment of all clear homologues with the candidate, (ii) by comparing the predicted secondary structure (Rost and Sander, 1993) for the candidate family with the observed secondary structure in families of known 3-D structure and (iii) by visual and numerical (energy) evaluation of packing and distribution of hydrophobic versus polar residues (Holm and Sander, 1992) in rough 3-D models built using the MaxSprout program (Holm and Sander, 1991). While these criteria are insufficient to prove a hypothetical model correct, they allow rejection of grossly inadequate proposals.

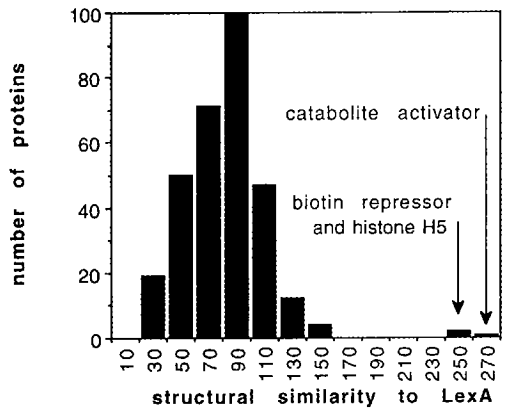

Fig. 1. Structure database search. The structure of the LexA repressor was compared with 300 representative (30% sequence identity cut-off) structures (Hobohm *et al.*, 1992) selected from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). The strongest (arrows) structural similarities of LexA are with the DNA binding domains of catabolite gene activator protein (entry 3GAP), biotin repressor (entry 1BIA) and histone H5 (entry 1HST). The coordinates of HNF-3/forkhead (Clark *et al.*, 1993) and HSF (Harrison *et al.*, 1994), two additional members of this structural class, were not available in the Brookhaven Protein Data Bank. Structural similarity, S, forming the horizontal axis, is defined as the sum of similarities of equivalent intramolecular distances:

$$S = \Sigma_i\Sigma_j[(0.20 - (|d^A_{ij} - d^B_{ij}|)/d^*_{ij})^*e - (d^*_{ij}/20.0\ \text{Å})^2],$$

where the summation is over all residues i,j of the common core, d^*_{ij} denotes the arithmetic mean of the Cα–Cα distances d_Aij and d_Bij in proteins A and B, a relative deviation of 0.20 (20%) is the threshold of similarity and the exponential factor downweights contributions from pairs at longer distances. The common core is found by maximizing S (Holm and Sander, 1993).

Results

Structure of the DNA binding domain of LexA repressor

The structure consists of three helices followed by a small antiparallel sheet. The second and third helix form a variation of the classical HTH motif in that the turn region contains one extra residue. The overall fold is similar to that of CAP, BirA, histone H5, HFN-3/forkhead and HSF (Figures 1 and 2). The common structural core, defined by comparison of distance matrices (Holm and Sander, 1993), includes the elements of secondary structure while the intervening loops vary in length and conformation. Three-dimensional superimposition of LexA yields positional r.m.s.d.s (for the Cα atoms) of 1.9 Å for 60 equivalent pairs with CAP, 2.5 Å for 61 equivalent pairs with BirA and 2.7 Å for 63 equivalent pairs with histone H5. Allowing for gaps in the variants, all available HTH structures typically superimpose with <1 Å r.m.s.d. over 21 equivalent Cα pairs. The loop between α1 and α2 in LexA is of intermediate length between CAP and BirA; helix α1 of LexA is extended by one turn compared with CAP. In CAP and HSF, the α1–α2 loop is long enough to form an extension of the β-sheet (only the two C-terminal β-strands are common to all structures in the class). The length of the 'turn' region in the HTH motif is variable as well (Figure 3). The connection between the two strands of LexA corresponds to the W1 ('wing') loop of HNF-3/forkhead, histone H5 and HSF (Brennan, 1993). This loop is longer in LexA than in either CAP or BirA, but shorter than in the eukaryotic proteins (HSF, HNF-3/forkhead, histone H5). In LexA, it probably does not play a role in DNA recognition although the 'wing' motif is important in HNF-3 (Clark *et al.*, 1993).

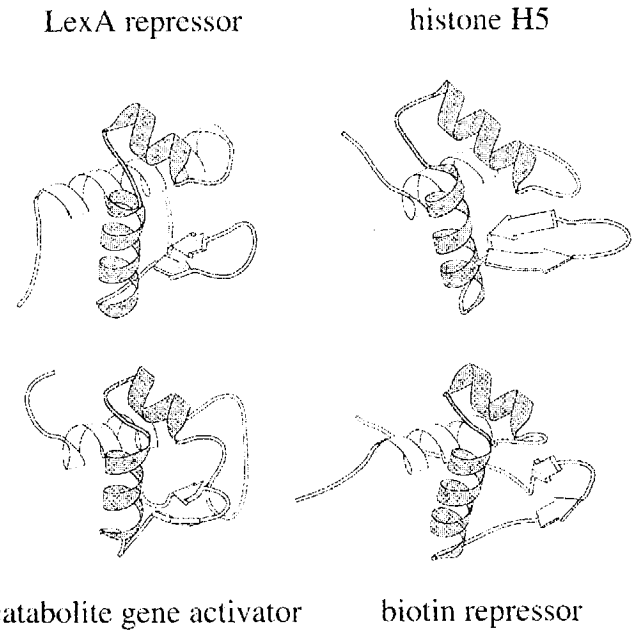

Fig. 2. CAP-like DNA binding domains. These domains have in common an N-terminal helix, a helix–turn–helix motif (darker shading) and a two-stranded β-sheet (arrows). The vertical helix is the DNA recognition helix, which docks into the major groove of B-DNA. The turn ('T') of the HTH motif has a one-residue insert in LexA and a seven residue insert in histone H5 (top row), relative to the classical HTH motif as found in catabolite gene activator and biotin repressor (bottom row). CAP in addition has a long insert after the first helix that forms an extension of the β-sheet. Drawn using Molscript (Kraulis, 1991).

Superfamily of CAP-like DNA binding domains

The six known structures represent protein families with quite different sequences. The structural alignment (Figure 3) forms the core of a multiple alignment, which was enlarged by including sequences with detectable homology to the proteins of known structure. Figure 4 gives an overview of the protein families which contain DNA binding domains with similar folds. Structurally important core residues appear similarly conserved in all six families (Figure 5). Strict conservation of exposed DNA binding residues is not expected nor observed, as each family recognizes a specific DNA sequence. Using combined sequence profiles to search sequence databases, we identified two additional proteins that are predicted to be members of the same structural class.

Structure prediction for the positive regulatory factor of listeriolysin

The haemolytic toxin listeriolysin is the major virulence factor of *Listeria*, a ubiquitous Gram-positive bacterium that can cause meningoencephalitis in humans. The positive regulator of listeriolysin (PrfA) was identified as a gene deletion in an avirulent strain of *Listeria*. PrfA can be reliably aligned over its entire length (237 residues) with the CAP family. The CAP family includes a large number of bacterial regulatory proteins, e.g. regulators of nitrogen fixation. The regulatory signal is transmitted by a conformational change of the protein, which is composed of an N-terminal ligand binding domain and a C-terminal DNA binding domain. Sequence identity between PrfA and other members of the CAP family is weak, at best just above 20%. Significantly, however, conserved sequence positions in the CAP family, which mostly map to the hydrophobic core, are similarly conserved in PrfA and its close

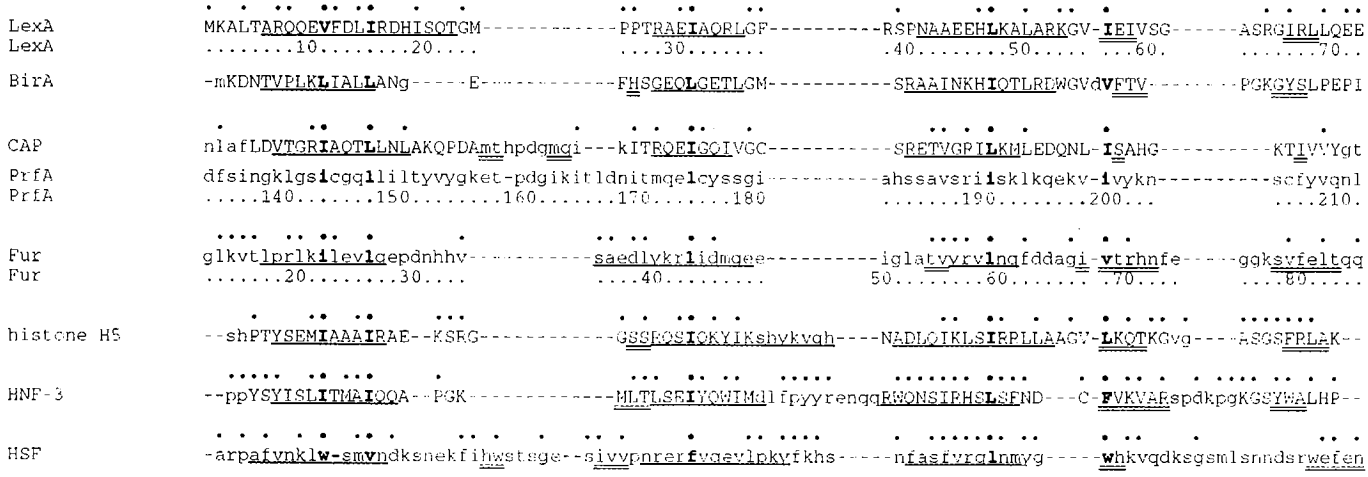

Fig. 3. Structural alignment of CAP-like DNA binding domains. The results from structure and sequence database searches are summarized in the form of a multiple alignment. Six families have a known 3-D structure: LexA, BirA, CAP, histone HS, HSF and HFN-3/forkhead. PrfA and Fur are predicted to adopt the same fold (new result). Residues structurally equivalent with LexA repressor are in upper case, otherwise sequences are in lower case. The structural alignment of HNF-3/forkhead to histone H5 is taken from Clark *et al.* (1993). The structure of HSF (Harrison *et al.*, 1994) was aligned by eye. Secondary structure (prediction in case of Fur) is given as helix (single underline) and strand (double underline). Sequence families (data not shown) around each protein of known structure were compiled by standard sequence database searches. Bullets (·) indicate positions at which residues are conserved independently in each sequence family (normalized sequence variability <20, calculated as in Sander and Schneider, 1991). The fact that the conserved positions (·) line up is strong sequence-based confirmation of the structural homology. Residue positions where all five subfamilies of known structure have a conserved residue are in bold type. The established members of the superfamily almost exclusively have valine, isoleucine and leucine at these core positions and so do the PrfA and Fur family. Figure 5 shows how these positions map in 3-D. The number of homologous sequences was four in the LexA family, 16 in the catabolite gene activator (crp/flp) family, 56 in the histone 5/histone 1 family, 11 in the HNF-3/forkhead family, eight in the heat shock factor family and six in the Fur family. No sequence homologues of the DNA binding domain in biotin repressor were found, so it has no bullets.

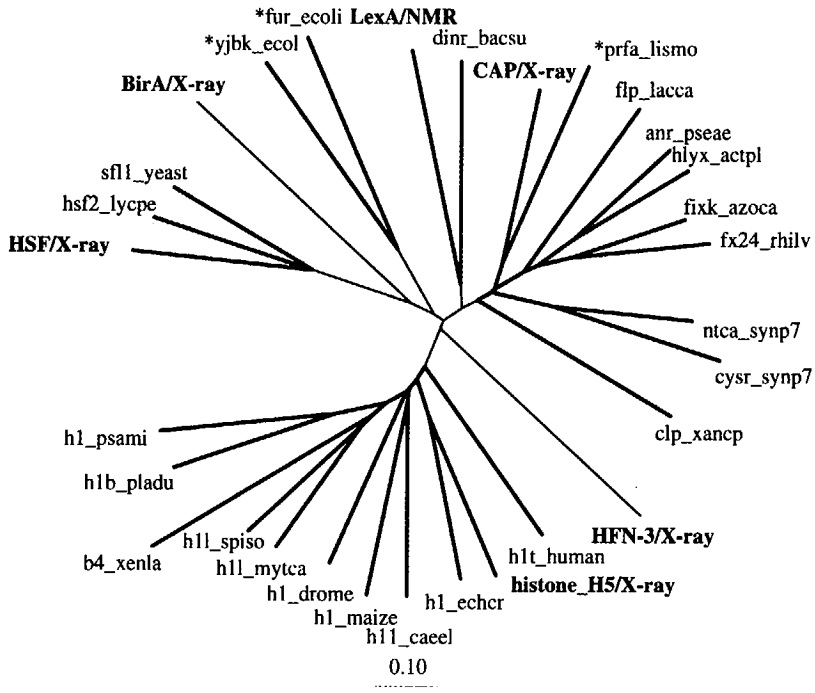

Fig. 4. Unrooted sequence tree for the superfamily. The tree shows the Swissprot database codes of representative sequences (at most 50% pairwise sequence identity) identified as members of the superfamily in a scan against the Swissprot database. Proteins of known structure are in bold type. Families branch off from the centre. Sequence identities between the families are in the 20% range. Non-trivial new findings are marked with an asterisk (*); these are the ferric uptake regulator family, represented by fur_ecoli and its homologue yjbk_ecoli and positive regulators of listeriolysin, represented by prfa_lismo. Drawn using Treetool (M.Maciukenas, University of Illinois, 1991).

relatives (see Figure 3). PrfA has an insertion of one residue in the turn of the HTH motif, like LexA repressor. A 3-D model for PrfA implied by the alignment is of comparable quality as models for established members of the CAP family (data not shown).

Structure prediction for ferric uptake regulatory protein

Iron is essential to cells, but detrimental at too high concentrations. The mechanism of the bacterial ferric uptake regulatory protein (Fur) has been thought unique because of the lack of sequence homology to any other known DNA binding protein.

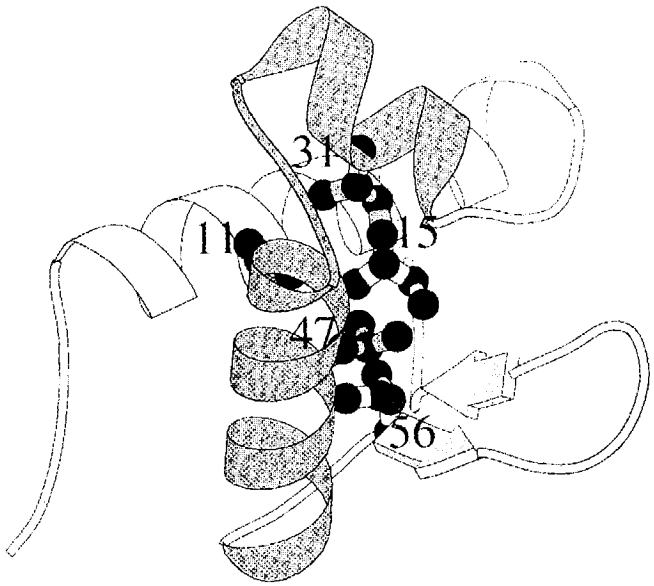

Fig. 5. Sequence conservation mapped in 3-D. Conserved residues typically fall into two classes. If they are solvent exposed, they are probably important for external function. If they are in the protein interior, they are probably important for maintaining the correct fold. There are five residues (side chains shown attached to LexA ribbon) which are conserved in all six families of known 3-D structure (see Figure 3). These conserved residues form the hydrophobic core of the DNA binding domain. The orientation is the same as in Figure 2.

A sequence profile combined from the CAP, BirA and LexA families identified Fur as a possible remote homologue. There is no easily detectable sequence similarity to any of the families separately, but striking agreement with the combined consensus pattern derived from the structural comparison (Figure 3). One of two conserved glycines with positive backbone phi angle in CAP, BirA and LexA (Gly36 and Gly54), is also conserved in the sequence of the Fur family (Gly66). The HTH motif is predicted to be a similar variation as in LexA and PrfA, with one residue inserted in the turn compared with the classical HTH motif.

Five homologues of Fur with 84–26% sequence identity over ~150 residues have the highest frequency of substitutions in the loop following the N-terminal helix, which is also structurally variable between the known structures. The putative recognition helix is highly conserved. Secondary structure prediction is currently the most accurate form of *ab initio* structural prediction. The secondary structure predicted directly from sequence (Rost and Sander, 1993) agrees remarkably well with the proposed alignment to the known structures (Figure 3). Model building in 3-D resulted in a well-packed hydrophobic core (data not shown). An interesting compensatory mutation was observed: the pair Phe12–Leu50 of LexA is replaced by Leu23–Phe62 in Fur.

Experimental data support the proposed 3-D model of Fur. Proteolytic cleavage at residue 77 yields two functional domains (Coy and Neilands, 1991). Our profile alignment spans the DNA binding domain. Circular dichroism spectra show that the protein is mostly α-helical with some β-structure, which fits well with the proposed CAP-like fold. Some structural features of Fur have been probed by solution ^{1}H-NMR studies (Saito *et al.*, 1991). In spite of solubility problems that forced the use of high pH conditions and partly unfolded specimen, it was possible to outline a fold for Fur schematically. The location of turns inferred from the NMR data agrees with our secondary structure predictions. Reassuringly, in the HTH region our structural model is even consistent with some of a dozen or so assigned long-sequence range NOEs (Saito *et al.*, 1991). Incompatible NOEs involve residues in loop regions, which we cannot model accurately.

Discussion

LexA repressor is the sixth protein of known structure in a widespread class of DNA binding domains typified by CAP. Evolutionary relationships between these diverse families are difficult to reconstruct either from sequence similarity, architectural detail or mode of DNA binding. The HTH motif is one structural solution—not the only possible one—for creating a protein surface complementary to that of the major groove of DNA. The recognition helix is believed to make the most important DNA contacts in proteins that have this motif. In different families, the specificity of recognition can be enhanced by various means, such as auxiliary DNA contacts made by flexible arms (e.g. in homeodomains) or 'wing' loops (e.g. in the monomeric HFN-3/forkhead), bending of DNA (e.g. by CAP), dimerization (many prokaryotic repressors; LexA, lambda repressor etc.) or combination of different types of DNA binding domains (POU domain, composed a POU-specific and a homeodomain).

The six CAP-like proteins illustrate how DNA binding domains are used as plug-in modules in functionally diverse proteins. Histone H5 is a monomeric component of the nucleosome and apparently binds DNA non-specifically. HSF has distinct trimerization and activation domains in a chain of ~700 residues. HNF-3/forkhead domains occur in several tissue-specific transcription activators. The DNA binding domain is C-terminal in CAP and N-terminal in BirA and LexA. Although lambda repressor and LexA repressor use DNA binding domains with different overall folds (the lambda repressor domain is structurally related to the POU-specific domain) (Assa-Munt *et al.*, 1993; Dekker *et al.*, 1993), their C-terminal domains appear to be derived from a common ancestor and to respond to the same cellular signals (Perry *et al.*, 1985): for both repressors the induction process involves a specific proteolysis reaction catalysed by RecA, which splits off the dimerization domain.

The rapidly growing pool of known 3-D structures allows more and more sequences to be classified into superfamilies with structural and functional analogies (Holm and Sander, 1994). In the present case study, we identified two new DNA binding proteins (Fur and PrfA) with non-classical HTH motifs as probably structurally related to LexA, CAP, BirA, HSF, HFN-3/forkhead and histone H5. The proposed similarity of Fur to LexA, CAP and BirA might lead to a reappraisal of the previous solution studies (Saito *et al.*, 1991) and, perhaps, a more detailed structural model. Independently, the predicted DNA binding sites of Fur and PrfA can be tested by site-directed mutagenesis.

Acknowledgements

We would like to thank Christos Ouzounis, Reinhard Schneider, Peer Bork, Michael Scharf and Burkhard Rost for discussions and software. The investigations were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO). A grant from the EC Science programme (No ST2J-0291-C) for the LexA project is gratefully acknowledged.

References

Altschul,S., Gish,W., Miller,W., Myers,E. and Lipman,D. (1990) *J. Mol. Biol.*, **14**, 286–290.

Assa-Munt,N., Mortishire-Smith,R., Aurora,R., Herr,W. and Wright,P. (1993) *Cell*, **73**, 193–205.

Bairoch,A. (1992) *Nucleic Acids Res.*, **20**, 2013–2018.

Bairoch,A. and Boeckmann,B. (1992) *Nucleic Acids Res.*, **20**, 2019–2022.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Brennan,R. (1993) *Cell*, **74**, 773–776.

Ceska,T., Lamers,M., Monaci,P., Nicosia,A., Cortese,R. and Suck,D. (1993) *EMBO J.*, **12**, 1805–1810.

Clark,K., Halay,E., Lai,E. and Burley,S. (1993) *Nature*, **364**, 320–412.

Coy,M. and Neilands,J. (1991) *Biochemistry*, **30**, 8201–8210.

Dekker,N., Cox,M., Boelens,R., Verrijzer,C., van der Vliet,P. and Kaptein,R. (1993) *Nature*, **362**, 852–855.

Dodd,I.B. and Egan,J.B. (1990) *Nucleic Acids Res.*, **18**, 5019–5026.

Fogh,R.H., Ottleben,G., Rüterjans,H., Schnarr,M., Boelens,R. and Kaptein,R. (1994) *EMBO J.*, **13**, 3936–3944.

Gribskov,M., McLachlan,M. and Eisenberg,D. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4355–4358.

Harrison,C., Bohm,A. and Nelson,H. (1994) *Science*, **263**, 224–227.

Harrison,S.C. (1991) *Nature*, **353**, 715–719.

Higgins,D.G., Bleasby,A.J. and Fuchs,R. (1992) *Comput. Appl. Biosci.*, **8**, 189–191.

Hobohm,U., Scharf,M., Schneider,R. and Sander,C. (1992) *Protein Sci.*, **1**, 409–417.

Holm,L. and Sander,C. (1991) *J. Mol. Biol.*, **218**, 183–194.

Holm,L. and Sander,C. (1992) *J. Mol. Biol.*, **225**, 93–105.

Holm,L. and Sander,C. (1993) *J. Mol. Biol.*, **233**, 123–138.

Holm,L. and Sander,C. (1994) *Proteins*, **19**, 165–173.

Kraulis,P. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Lamerichs,R., Padilla,A., Boelens,R., Kaptein,R., Ottleben,G., Ruterjans,H., Granger-Schnarr,M., Oertel,P. and Schnarr,M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 6863–6867.

Leiting,B., Francesco,R., Tomei,L., Cortese,R., Otting,G. and Wutrich,K. (1993) *EMBO J.*, **12**, 1797–1803.

Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.

Ogata,K., Hojo,H., Aimoto,S., Nakai,T., Nakamura,H., Sarai,A., Ishii,S. and Ishimura,Y. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6428–6432.

Perry,K., Elledge,S., Mitchell,B., Marsh,L. and Walker,G. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 4331–4335.

Ramakrishnan,V., Finch,J., Graziano,V., Lee,P. and Sweet,R. (1993) *Nature*, **362**, 219–223.

Rost,B. and Sander,C. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7558–7562.

Saito,T., Wormald,M. and Williams,R. (1991) *Eur. J. Biochem.*, **197**, 29–38.

Sander,C. and Schneider,R. (1991) *Proteins*, **9**, 56–68.

Schultz,S., Shields,G. and Steitz,T. (1991) *Science*, **253**, 1001–1007.

Sibbald,P. and Argos,P. (1990) *J. Mol. Biol.*, **216**, 813–818.

Wilson,K., Shewchuk,L., Brennan,R., Otsuka,A. and Matthews,B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9257–9261.

Received July 20, 1994; revised September 26, 1994; accepted September 26, 1994