Minireview

Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors

Chuck Chuong Nguyen, Milton H. Saier Jr.*

Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA

Received 8 September 1995; revised version received 7 November 1995

Abstract Phylogenetic tree construction for 25 sequenced members of the LacI-GalR family (LGF) of transcription factors revealed that almost all branches are similar in length, radiating essentially from a single point. This observation suggests that most of these proteins arose by duplication events which occurred at a specific time in evolutionary history, and that further duplication events were rare. Analyses of the multiple alignment of the LGF proteins lead to suggestions regarding structure-function relationships and reveal that the helix-turn-helix DNA-binding motif of LGF proteins is similar in sequence to those of numerous non-homologous DNA-binding proteins.

Key words: Transcription; Regulation; DNA-binding protein; Helix-turn-helix motif; Evolution; Phylogenetic tree

1. Introduction

Numerous transcriptional regulatory proteins of bacteria, archaea and eukarya function to control rates of RNA polymerase-mediated transcriptional initiation [1]. These proteins fall into several classes based on their modes of DNA binding. One such class that is particularly prevalent in bacteria binds the DNA via helix-turn-helix (H-T-H) motifs [2,3]. Among these proteins are some of the most thoroughly investigated transcription factors including the lactose repressor of *Escherichia coli* (LacI), the cyclic AMP receptor protein of *E. coli* (CRP), and the Cro protein of phage lambda [4]. Although all of these factors possess H-T-H motifs, many of them are not demonstrably homologous and probably arose independently of each other.

In 1991, a report from our laboratory presented a phylogenetic analysis of the ten then sequenced members of a family of H-T-H transcriptional regulatory proteins which included the repressors of the lactose, galactose, fructose, and purine operons of *E. coli* (LacI, GalR, FruR, and PurR, respectively) [5]. These proteins possess small N-terminal H-T-H domains that are involved in DNA binding as well as large C-terminal ligand binding domains that are homologous to the periplasmic sugar binding receptors specific for galactose, arabinose, and ribose [5–9]. Ligand binding to the C-terminal domain generally results in a conformational change in the N-terminal domain that causes dissociation of the protein from the DNA. The sequences of these proteins have been analyzed in detail [10], their structures have been predicted based on X-ray diffraction, NMR, and modeling studies [11,12], and recently the three-dimensional structure of one of the members of this family, PurR, in ternary complex with its corepressor and DNA operator has been solved [13].

In this report, we present an update of our previous phylogenetic analysis of the LacI-GalR family (LGF). These analyses lead to an interesting suggestion regarding the evolutionary pathway taken for the diversification of this protein family. Sequence comparisons allow extrapolation of functional assignments in PurR to other members of the family. We note that many nonhomologous prokaryotic DNA binding proteins exhibit sequence similarities in their H-T-H regions although significant sequence similarity elsewhere in these proteins is lacking. The significance of these findings is briefly discussed.

2. The LacI-GalR family (LGF)

Many currently sequenced members of the LGF (excluding the homologous periplasmic sugar-binding receptors) are listed in Table 1. When two or more proteins of the same function from closely related organisms have been sequenced (i.e. FruR from *E. coli* and *S. typhimurium*; ScrR from *S. typhimurium* and *K. pneumoniae*, or LacI from *E. coli* and *K. pneumoniae*), only one of these proteins was selected for presentation [5,10]. Similarly, if two very similar proteins of essentially the same function have been sequenced from a single organism (i.e. GalR and GalS of *E. coli*; [14]), only one was selected for study. Incompletely sequenced proteins were also omitted from our study. It is interesting to note that the chromosomally-encoded sucrose repressor of *E. coli* (CscR Eco; [15]) is only distantly related to other sequenced sucrose repressors such as ScrR Kpn (Table 1; see below).

3. Phylogenetic tree for the LGF

The phylogenetic tree for the twenty-five fully sequenced members of the LGF listed in Table 1 is presented in Fig. 1. Almost all proteins included in the study appear on non-forked branches that radiate from points near the center of the tree. The few exceptions other than two trivial orthologues (i.e. CcpA from *B. subtilis* and *B. megaterium*, and GalR from *E. coli* and *H. influenzae*) are: (a) the ribose (RbsR) and purine (PurR) repressors of *E. coli* which evidently branched from each other late in the evolutionary process; (b) the *Bacillus* catabolite repressor protein, CcpA, and the *Clostridium* amy-

*Corresponding author. Fax: (1) (619) 534-7108.
E-mail: msaier@ucsd.edu

SSDI 0014-5793(95)01344-X

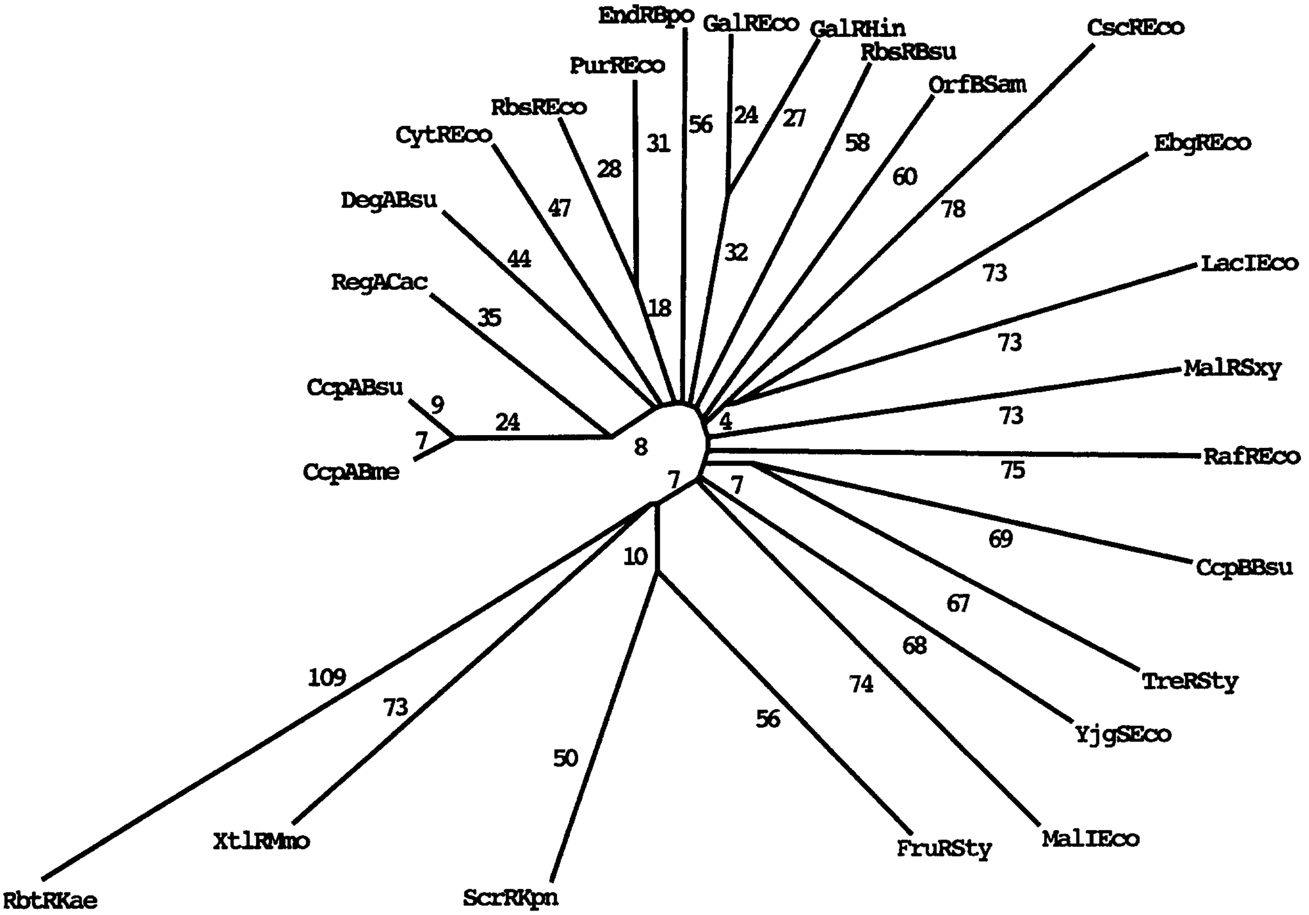

Fig. 1. Phylogenetic tree for the proteins of the LacI-GalR family (LGF) listed in Table 1. Branch lengths are approximately proportional to phylogenetic distance and are expressed in arbitrary units. Construction of the tree and estimation of the relative evolutionary distances between members of the family were as described by Reizer and Reizer [21] using the progressive alignment method (TREE program) of Feng and Doolittle [22]. This method utilizes the minimum mutation matrix of Dayhoff et al. [23], thereby permitting matching of identical residues as well as similar residues [9]. Neutral elements are inserted in gaps. Abbreviations are as indicated in Table 1.

lase repressor, RegA, which have been reported to exhibit cross reactivity [16]; and (c) the sucrose (ScrR) and fructose (FruR) repressors of *K. pneumoniae* and *S. typhimurium*, respectively, which branched from each other relatively early. The very early branching of TreR of *S. typhimurium* from CcpB of *B. subtilis* (Fig. 1) is probably too close to the trunk of the tree to indicate that these two proteins are more closely related to each other than to other members of the family. Thus, for the twenty-five proteins depicted, there are twenty primary branches, all stemming from a point close to the center of the tree.

4. Multiple alignment of LGF proteins

The multiple alignment of the 25 sequences was examined for regions of striking sequence similarity. By far the most conserved regions were the H-T-H regions of the N-terminal DNA binding domains. This portion of the alignment, portrayed to the left of Fig. 2, includes two fully conserved residues (A at alignment position 16 and S at alignment position 27). In PurR, A8, corresponding to the alanine at position 16, is internal, comprising part of the hydrophobic core. It allows proper packing between helices 1 and 2 of the H-T-H motif. In PurR, S19, corresponding to the fully conserved S at alignment position 27, makes a contact with phosphate in the DNA and also forms a weak hydrogen bond to N23 (alignment position 31 in Fig. 2). It therefore provides both a liganding function and a structural function. This dual role may be characteristic of the equivalent residue in all or most members of the LGF.

These two residues (A at position 16 and S at position 27 in Fig. 2) were the *only* fully conserved residues in the multiple alignment of the 25 LGF proteins included in our study. Other residues in the H-T-H region (Fig. 2) were largely conserved as follows: alignment positions 12, 15, 21, 26 and 30, all hydrophobic; position 14, all D or E with 3 exceptions; position 21, all Vs with 1 exception; position 22, all Ss with 2 exceptions; position 25, all Ts with 1 exception, and position 26, all Vs with 3 exceptions. The functional significance of some of these residues in specific members of the LGF is known and has been discussed [6,7,10,13,17]. For example, in PurR, S14 (alignment position 22) hydrogen bonds to T17 (alignment position 25) accounting in part for the conservation of these residues.

5. Sequence comparisons of the H-T-H motifs in various DNA-binding protein families

An attempt was made to derive an LGF specific signature sequence [18] from the H-T-H regions of these proteins. The following sequence was derived:

```
                 10                  20                  30                  40                  50                  60
CcpBBsu   (1)  M A N I K E I A R L A N V S V S T V S R V L N   H H   P   Y V S E E K R K L V H Q V M K E L D Y T P N R T A I D L I
CcpABme   (2)  N V T I Y D V A R E A S V S M A T V S R V V N   G N   P   N V K P S T R K K V L E T I E R L G Y R P N A V A R G L A
CcpABsu   (3)  N I T I Y D V A R E A N V S M A T V S R V V N   G N   P   N V K P T T R K K V L E A I E R L G Y R P N A V A R G L A
DegABsu   (2)  K T T I Y D V A K A A G V S I T T V S R V I N   N T   G   R I S D K T R Q K V M N V M N E M A Y T P N V H A A A L T
PurREco   (1)  M A T I K D V A K R A N V S T T T V S H V I N   K T   R   F V A E E T R N A V W A A I K E L H Y S P S A V A R S L K
RbsREco   (1)  M A T M K D V A R L A G V S T S T V S H V I N   K     R   F V S E A I T A K V E A A I K E L N Y A P S A L A R S L K
CytREco   (9)  A A T M K D V A L K A K V S T A T V S R A L M   N P   D   K V S Q A T R N R V E K A A R E V G Y L P Q P M G R N V K
RegACac   (2)  A T S I K D V A R E A G V S I A T V S R V L N   D I   D   V V N E D T K K K V L D A I K E L G Y R P N I V A R S L K
GalREco   (1)  M A T I K D V A R L A G V S V A T V S R V I N   N S   P   K A S E A S R L A V H S A M E S L S Y H P N A N A R A L A
GalRHin   (1)  M S T I R D V A K L A N V S V A T V S R V L N   H S   L   S V S E N T R L A V E R A I A Q L A Y Q P N A N A Q A L A
MalRSxy   (1)  M V T I K D I A Q A A N V S T S T V S R V I S   G N   P   R I S M Q T R E K V K A T M K S F N Y Q P N R A A R T L A
YjgSEco   (5)  R I S L Q D I A T L A G V T K M T V S R Y I R   S P   K   K V A K E T G E R I A K I M E E I N Y I P N R A P G M L L
RbsRBsu   (1)  M A T I K D V A G A A G V S V A T V S R N L N   D N   G   Y V H E E T R T R V I A A M A K L N Y Y P N E V A R S L Y
OrfBSam   (2)  T R R L A Q V A K K V G V S E A T V S R V L N   D K   P   G V S E A T R Q S V L S A L D V L G Y       E R P T Q L R
EndRBpo   (2)  V T T M K E V A E R A G V S K S T V S Q F L Q   K R Y N   Y M S E N T K K K I E Q A I E D L S Y I P N E V A R S L K
TreRSty   (4)  R L T I K D I A R L S G V G K S T V S R V L N   N E   S   G V S E R T R E R V E A V M N Q H G F S P S R S A R A M R
LacIEco   (3)  P V T L Y D V A E Y A G V S Y Q T V S R V V N   Q A     S H V S A K T R E K V E A A M A E L N Y I P N R V A Q Q L A
MalIEco   (6)  K I T I H D V A L A A G V S V S T V S L V L S G K G       R I S T A T G E R V N A A I E E L G F V R N R Q A S A L R
FruRSty   (1)    M K L D E I A R L A G V S R T T A S Y V I N G K A K Q Y R V S D K T V E K V M A V V R E H N Y H P N A V A A G L R
XtlRMmo   (8)  K V T L E N V A A I A N V S K I T A S R A F S Q P D       K V H P E T L R R I L D A A D K I G Y V V N A A A R S L R
RafREco   (1)    M S L K A I A T T L G I S V T T V S R A L G G F S       D V A A S T R E R V E A E A R R R G Y R P N T Q A R R L K
CscREco   (1)  M A S L K D V A R L A G V S M M T V S R V M H N A E       S V R P A T R D R V L Q A I Q T L N Y V P D L S A R K M R
ScrRKpn   (5)  R V T I K D I A E L A G V S K A T A S L V L N G R G K E L R V A Q E T R E R V L A I A R E Q H Y Q P S I H A R S L R
EbgREco   (1)  M A T L K D I A I E A G V S L A T V S R V L N D D P   T L N V K E E T K H R I L E I A E K L E Y K T S   S A R K L Q
RbtRKae   (3)  K I T I Y D L A E L S G V S A S A V S A I L N G N W K K R R I S A K L A E K V T R I A E E Q G Y A I N R Q A S M L R

Consensus      - - T I K D V A - - A G V S - - T V S R V L N   - -   -   - V S - - T R - - V - - A - - - L - Y - P N - - A R - L -
                     Helix          Turn        Helix
```

Fig. 2. Multiple alignment of the most conserved regions of the proteins of the LGF encompassing most of the DNA binding regions. Abbreviations for the family are as provided in Table 1. Alignment positions are indicated above the multiple alignment and do not correspond to the residue numbers of any one protein. Residue numbers of the twenty-five aligned proteins are provided in parentheses at the beginning of each sequence. The consensus sequence (Consensus) (at least 13 of the 25 residues at any one position conserved) is provided below the alignment. The (putative) helix-turn-helix motif is indicated below the consensus sequence. Fully conserved residues are presented in bold type. The alignment was generated using the progressive alignment program of Feng and Doolittle [22]. The complete multiple alignment for the 25 LGF proteins is available upon request from MHS.

(TSRKN)(LIVM)X(DENQA)(LIV)AX$_4$(LIV)(STG)X$_2$(TA)(VA)S

Residues in parentheses indicate alternative possibilities at a particular position; X = any residue. This motif was screened against the current SwissProt database with either zero or one mismatch. In addition to the proteins of the LGF, two non-homologous proteins in the SwissProt database proved to contain this motif. These two proteins were the StrR regulatory protein of the streptomycin biosynthetic operon of *Streptomyces griseus* exhibiting the sequence: (211)SLRQIAAQAGVSPSTAS, and an open reading frame of *E. coli*, YehA, exhibiting the sequence: (302)NLKEVAAKSKLTDTTVS. We suggest that in both cases, the regions detected form H-T-H DNA binding motifs. Experimental evidence bearing on this possibility is not yet available.

With a single mismatch, many additional non-homologous proteins were identified. Almost all those of known function proved to be prokaryotic H-T-H DNA binding proteins, and

Table 1
Proteins of the LacI-GAlR family (LGF)

Abbreviation	Name	Organism	Length	Accession No.
CcpABsu	Catabolite control protein A	*Bacillus subtilis*	334	sp P25144
CcpABme	Catabolite control protein A	*Bacillus megaterium*	332	gp L26052
CcpBBsu	Catabolite control protein B	*Bacillus subtilis*	311	sp P37517
DegABsu	Degradation enzyme activator	*Bacillus subtilis*	337	sp P37947
PurREco	Purine repressor	*Escherichia coli*	341	sp P15039
RbsREco	Ribose repressor	*Escherichia coli*	329	gp L10328
CytREco	Cytidine repressor	*Escherichia coli*	341	sp P06964
RegACac	Amylase repressor	*Clostridium acetobutylicum*	332	gp L14685
GalREco	Galactose repressor	*Escherichia coli*	343	sp P03024
GalRHin	Galactose repressor	*Haemophilus influenzae*	332	sp P31766
MalRSxy	Maltose repressor	*Staphylococcus xylosus*	337	pir S44187
YjgSEco	Unidentified ORF	*Escherichia coli*	332	sp P39343
RbsRBsu	Ribose repressor	*Bacillus subtilis*	325	sp P36944
OrfBSam	Unidentified ORF	*Streptomyces ambofaciens*	332	pir S33361
EndRBpo	Endogluconase repressor (putative)	*Bacillus polymyxa*	340	sp P27871
TreRSty	Trehalose repressor	*Salmonella typhimurium*	315	sp P36674
LacIEco	Lactose repressor	*Escherichia coli*	360	sp P03023
MalIEco	Maltose repressor	*Escherichi coli*	325	sp P18811
FruRSty	Fructose repressor	*Salmonella typhimurium*	334	sp P21930
XtlRMmo	Xylitol repressor	*Morganella morganii*	335	gp L34345
RafREco	Raffinose repressor	*Escherichia coli*	335	sp P21867
CscREco	Sucrose repressor	*Escherichia coli*	331	sp P40715
ScrRKpn	Sucrose repressor	*Klebsiela pneumoniae*	334	sp P37076
EbgREco	Evolved β-galactosidase repressor	*Escherichia coli*	327	sp P06846
RbtRKae	Ribitol repressor	*Klebsiella aerogenes*	270	sp P07760

Table 2
Representative prokaryotic DNA binding proteins that are not demonstrably homologous to LGF proteins but exhibit helix-turn-helix DNA binding motifs that are the same as those of LGF proteins with a single mismatch

Protein	Species	Function	Sequence[d]
FixK[a]	*Rhizobium meliloti*	N_2 fixation regulatory protein	(167)S$\underline{R}$QDIADYLGLTIETVS
FimZ[b]	*Escherichia coli*	Fimbrial protein regulation	(192)S$\underline{N}$KEIADKLLLSNKTVS
BvgA[b]	*Bordetella pertusssis*	Activator of virulence	(170)S$\underline{N}$KDIADSMFLSNKTVS
MoxX[b]	*Paracoccus denitrificans*	Regulator of methanol utilization	(178)S$\underline{Y}$RDIADRACISYKTVS
Cro[c]	Phage P22	Regulatory protein	(17)T$\underline{Q}$RAVAKALGISDAAVS
Sigam 54[c]	*Bacillus subtilis*	RNA polymerase sigma factor	(329)TLREVADCLSL$\underline{H}$ESTVS
SmtB[c]	*Synechococcus* sp.	Repressor	(66)$\underline{C}$VGDLAQAIGVSESAVS
TrpI[c]	*Pseudomonas syringae*	*trpBA* operon activator	(27)SVSQ$\underline{A}$AEQLHVTHGAVS

[a] Several homologous FixK proteins from different bacterial species exhibit this motif.
[b] These proteins are homologous.
[c] These proteins are not demonstrably homologous to other proteins listed in this table.
[d] The underlined amino acid is the one that differs from the LGF helix-turn-helix sequence motif (see text).

in many such cases, the regions detected correspond to recognized H-T-H regions. Some of these proteins are listed in Table 2 and include the FixK proteins of many nitrogen fixing bacteria (homologous to the cyclic AMP receptor proteins (CRP) and fumarate-nitrate regulators (FNR) of Gram-negative bacteria), the FimZ protein of *E. coli* and several of its homologues, the Cro transcriptional regulator of phage P22 and the alternative sigma factor of the *B. subtilis* RNA polymerase, σ^{54}. Because several of these proteins (see footnotes 1–3 to Table 2) were not demonstrably homologous to each other, it can be concluded that the sequence motif used for the construction of the DNA binding H-T-H structure in the LGF proteins has been utilized by several families of DNA binding proteins in addition to the LGF. This marked sequence similarity, restricted to the H-T-H motifs, may either have arisen by evolutionary convergence in order to generate a high affinity DNA binding structure or by domain shuffling followed by extensive sequence divergence outside of the H-T-H motif.

6. DNA interactions external to the H-T-H regions of LGF proteins

Adjacent to the H-T-H motif and to the right hand side of Fig. 2 is a region exhibiting moderate conservation. For example, the Y at alignment position 56 (residue #45 in PurR) is conserved in all but two proteins where F is found; the A at alignment position 62 (residue #51 in PurR) is conserved in all but 3 proteins where G or P is found, and the L at alignment position 65 (residue #54 in PurR) is conserved in all but 3 proteins where V or M is found (Fig. 2). Residue #45 in PurR is in the short loop separating helix 3 from helix 4, and residues #51 and #54 in PurR are in helix 4, the 'hinge helix', which constitutes the second (minor groove) DNA binding element in PurR [13].

7. Sequence divergence of the large, C-terminal, ligand-binding domains of LGF proteins

The C-terminal domains of the LGF proteins are homologous to several bacterial periplasmic sugar binding receptors. The 3-dimensional structures of both the liganded and the unliganded forms of some of these proteins are known [6–8,19,20]. Outside of the regions of the LGF protein sequences depicted in Fig. 2, only two largely conserved residues were found to be of particular note. The D at alignment position 187 in the complete multiple alignment (residue #160 in PurR) is conserved in all but 3 proteins where N, S and A are found, respectively, and the R at alignment position 226 (residue #196 in PurR) is conserved in all but 4 proteins where K is found. The former residue in PurR is critical for stabilizing the tertiary structure of the open, unliganded form of the protein (R.G. Brennan, personal communication). The latter residue is important for effector binding [10,13].

8. Conclusions and perspectives

We have noted unusual phylogenetic relationships that characterize the LGF. Because almost all branches in the tree (Fig. 1) bear a single protein, and almost all branches stem from positions near the trunk of the unrooted tree, we suggest that at a very early time in evolutionary history, not long after fusion of the DNA binding domain with the ligand binding domain [5], there was tremendous pressure for gene duplication events that gave rise to most of the current, functionally dissimilar members of the LGF. Presumably this time correlated with a time when bacterial carbon metabolism first became subject to stringent and specific transcriptional control by repressors. Subsequently, very few additional gene duplications occurred, and most of the diversification that did arise in the LGF was due to vertical transmission of genetic material as the bacterial species themselves proliferated and diversified. During this diversification process, the N-terminal DNA binding domains retained the primordial sequence most strikingly, the C-terminal ligand binding domains were least well-conserved as they diversified in order to accommodate a variety of dissimilar ligands, and the regions adjacent to the H-T-H motifs retained an intermediate degree of sequence similarity, presumably because these regions play a unified role in contributing to the stability of the DNA–protein interaction by forming noncovalent bonds with the DNA in the minor groove.

The analyses and postulates presented in this minireview open up new vistas that provide guides for future experimentation. First, by estimating the divergence times of various bacterial species and including LGF orthologues of the same function from evolutionarily divergent bacteria in phylogenetic analyses analogous to that depicted in Fig. 1, it should be possible to estimate the time in evolutionary history when gene duplication gave rise to most of the LGF paralogues. Second, by determining the 3-dimensional structures of functionally dissimilar LGF paralogues, it should be possible to establish the extent to which functional constraints dictated sequence conservation. Finally, by conducting further sequence investigations as well as protein–DNA interaction studies and 3-dimensional structural analyses, clues may be forthcoming as to common structural features shared by nonhomologous H-T-H DNA binding domains that confer high affinity DNA binding. Such studies will undoubtedly clarify our understanding of the evolutionary process that gave rise to the plethora of present day transcription factors and allow more precise definition of structure–function relationships.

Acknowledgments: We thank Richard G. Brennan for valuable discussions, Jonathan and Aiala Reizer for examining the LGF H-T-H motif sequence, and Mary Beth Hiller and Jacqueline Richardson for valuable assistance in the preparation of this manuscript. This work was supported by US Public Health Service Grants 5RO1 AI21702 and 2RO1 AI14176 from the National Institutes of Allergy and Infectious Diseases.

References

[1] Reznikoff, W.S., Siegele, D.A., Cowing, D.W. and Gross, C.A. (1985) Annu. Rev. Genet. 19, 355–387.

[2] Takeda, Y., Ohlendorf, D.H., Anderson, W.F. and Matthews, B.W. (1983) Science 221, 1020–1026.

[3] Brennan, R.C. (1991) Curr. Opin. Struct. Biol. 1, 80–88.

[4] Harrison, S.C. and Aggarwal, A.K. (1990) Annu. Rev. Biochem. 59, 933–969.

[5] Vartak, N.B., Reizer, J., Reizer, A., Gripp, J.T., Groisman, E.A., Wu, L.-F., Tomich, J.M. and Saier Jr., M.H. (1991) Res. Microbiol. 142, 951–963.

[6] Mauzy, C.A. and Hermodson, M.A. (1992) Protein Sci. 1, 831–842.

[7] Mauzy, C.A. and Hermodson, M.A. (1992) Protein Sci. 1, 843–849.

[8] Tam, R. and Saier Jr., M.H. (1993) Microbiol. Rev. 57, 320–346.

[9] Saier Jr., M.H. (1994) Microbiol. Rev. 58, 71–93.

[10] Weickert, M.J. and Adhya, S. (1992) J. Biol. Chem. 267, 15869–15874.

[11] Schumacher, M.A., Macdonald, J.R., Bjorkman, J., Mowbray, S.L. and Brennan, R.G. (1993) J. Biol. Chem. 268, 12282–12288.

[12] Hsieh, M., Hensley, P., Brenowitz, M. and Fetrow, J.S. (1994) J. Biol. Chem. 269, 13825–13835.

[13] Schumacher, M.A., Choi, K.Y., Zalkin, H. and Brennan, R.G. (1994) Science 266, 763–770.

[14] Weickert, M.J. and Adhya, S. (1993) Mol. Microbiol. 10, 245–251.

[15] Bockmann, J., Heuel, H. and Lengeler, J.W. (1992) Mol. Gen. Genet. 235, 22–32.

[16] Davison, S.P., Santangelo, J.D., Reid, S.J. and Woods, D.R. (1995) Microbiology 141, 989–996.

[17] Friedman, A.M., Fischmann, T.O. and Steitz, T.A. (1995) Science 268, 1721–1727.

[18] Bairoch, A. (1992) Nucleic Acids Res. 20, 2013–2018.

[19] Vyas, N.K., Vyas, N.M. and Quiocho, F.A. (1991) J. Biol. Chem. 266, 5226–5237.

[20] Flocco, M.M. and Mowbray, S.L. (1994) J. Biol. Chem. 269, 8931–8936.

[21] Reizer, A. and Reizer, J. (1994) in: Methods in Molecular Biology: Computer Analysis of Sequence Data, Part II (Griffin, A.M. and Griffin, H.G. eds.) pp. 319–325, Humana Press, Totowa, NJ.

[22] Feng, D.-F. and Doolittle, R.F. (1990) Methods Enzymol. 183, 375–387.

[23] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure, vol. 5, suppl. 3 (Dayhoff, M.O. ed.) pp. 345–352, National Biomedical Research Foundation, Silver Spring, MD.