Classification of multi-helical DNA-binding domains and application to predict the DBD structures of σ factor, LysR, OmpR/PhoB, CENP-B, Rap1, and XylS/Ada/AraC

Masashi Suzuki*, Steven E. Brenner

MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

Received 12 July 1995; revised version received 17 August 1995

Abstract We have systematically compared structures of multi-helical DNA-binding domains (DBDs) which have been determined by crystallography or NMR spectroscopy. All the known multi-helical DBDs are very similar. The core of these structures consists of two α-helices in the helix–turn–helix combination, associated with one or two other helices. The structures can be classified according to either additional structural compositions or the configuration of the helices. Many DBDs, whose structures are currently unknown, have sequences which resemble those of known structures, permitting outlines of the new structures to be predicted.

Key words: Transcription factor; DNA recognition; DNA–protein interaction; Protein folding

1. Introduction

The vast majority of transcription factors use an α-helix for DNA-recognition (the recognition helix) [1,2]. These proteins employ a variety of folds, which can be understood in terms of structural requirements, i.e., the folds must stabilise a recognition helix with a limited number of amino acid residues and must expose the recognition helix on the surface so that it can bind to DNA. Recognition helices found in crystal/solution structures have only a small number of amino acid residues (about ten residues) but if a peptide of only the ten residues is synthesised, it is unlikely to fold into a stable α-helix.

Some transcription factors use a metal ion to stabilise the recognition helix: the zinc finger, C4 zinc binding and C6 zinc cluster families [3–8]. The recognition helix can bind to a metal ion through Cys or His residues when it adopts an appropriate structure. Some other factors, the basic domain-leucine zipper and helix–loop–helix families [9–13], consist of a continuous α-helix, of which different regions recognise DNA and form a zipper. The use of a single element of secondary structure permits the zipper to stabilise the DNA recognition region of the helix. A few transcription factors (E2 and p53) incorporate a recognition helix into folds which have many β-strands [14,15].

The remaining α-helix type transcription factors, a large number of factors, use globular folds which are predominantly α-helical. Standard zinc fingers are found only in eukaryotes, and C6 proteins are used only by fungi, but the multi-helical globular folds are found in DNA-binding domains (DBDs) of prokaryotes as well as in eukaryotes.

*Corresponding author.

Resemblance between individual multi-helical DBDs has been repeatedly noticed [16–23]. In particular, the helix–turn–helix (HTH) motif was proposed to group some prokaryotic factors [16,22]. Ramakrishanan et al. [18] pointed out that a prokaryotic factor, CAP, and an eukaryotic DNA-binding protein, histone H5, have similar structural compositions, and Clark et al. [19] expanded the similarity to Engrailed homeodomain and HNF3.

In this paper, by using crystal/NMR coordinates, we compare 27 multi-helical DNA-binding domains systematically and quantitatively to classify them and to understand the stereochemical characteristics common among the domains, aiming to predict strucures of some other DNA-binding domains which currently remain undetermined.

2. Materials and methods

The crystal/NMR structures examined are summarised in Table 1. The helices of each protein were defined using the helix records in the PDB [24] entries shown in table I, or, if these records were not present, using the DSSP [25] algorithm implemented by RasMol [26]. An exception was λcro whose helices were defined by inspection.

As a measure of configurational similarity between pairs of proteins, we summed the differences in the angles (measured in radians) between the three DBD core helices. When four core helices were present, either helix 1 or 4 was excluded so as to produce the lowest (best) score. (Helices 1 and 4 are placed in similar positions; see section 3). This measure was used rather than root mean square (RMS) deviation, as it better reflected the configurational variation and avoided arbitrary erroneous assignment of equivalent positions in variable-length elements of structure. Other approaches could also be used, however inter-helix angles provide a simple and intuitive method of examining structures. We note that distances between helix centroids is not a useful marker, because extents of helices from the centroids vary among these structures.

3. Results and discussion

3.1. Similarity among the multi-helical DBDs

All the DNA-binding domains (DBDs) studied here (Table 1) have at least three α-helices (see crystal/NMR structures in Fig. 1 and schematic drawings in Fig. 2). Two of the three helices (here referred to as helices 2 and 3) are combined roughly in the manner of helix–turn–helix (HTH) [22] and another is positioned either N-terminal (referred to as helix 1) or C-terminal (referred to as helix 4) to the HTH (Fig. 2). Helix 3 binds to DNA (the recognition helix).

The DBDs of LacR and PurR contain helices 2–4 (Fig. 2a) but lack helix 1, while those of Myb, TetR, Engl, Antp, Matα2, Oct1/2 homeodomain, and LFB contain helices 1–3 but lack helix 4 (Fig. 2b,c). Helix 1 in the former group and helix 4 in

SSDI 0014-5793(95)00988-4

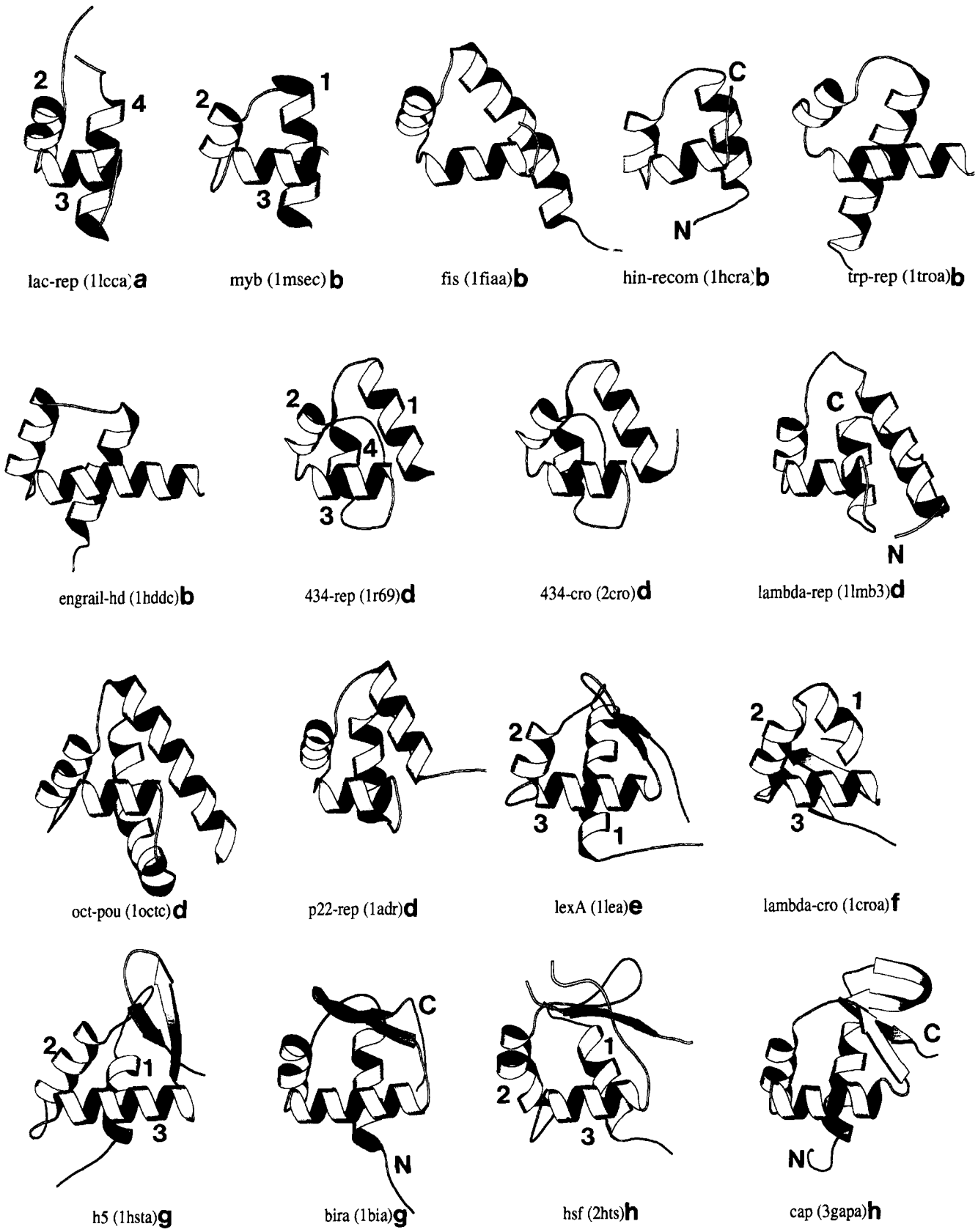


Fig. 1. Crystal/NMR structures of multi-helical DNA-binding domains. All proteins are oriented so that the DNA-binding helix (helix 3 in Fig. 2) runs horizontally from left to right in the N to C direction, and helix 2 runs in the N-to-C direction from top to bottom on the left side of helix 3. For the sake of clarity, β-strands with only 1 or 2 residues are omitted. In some examples the N and C termini are indicated. Classification, a–h, (see Fig. 2) is also shown with the names of the proteins. The PDB code names are shown in the parentheses.

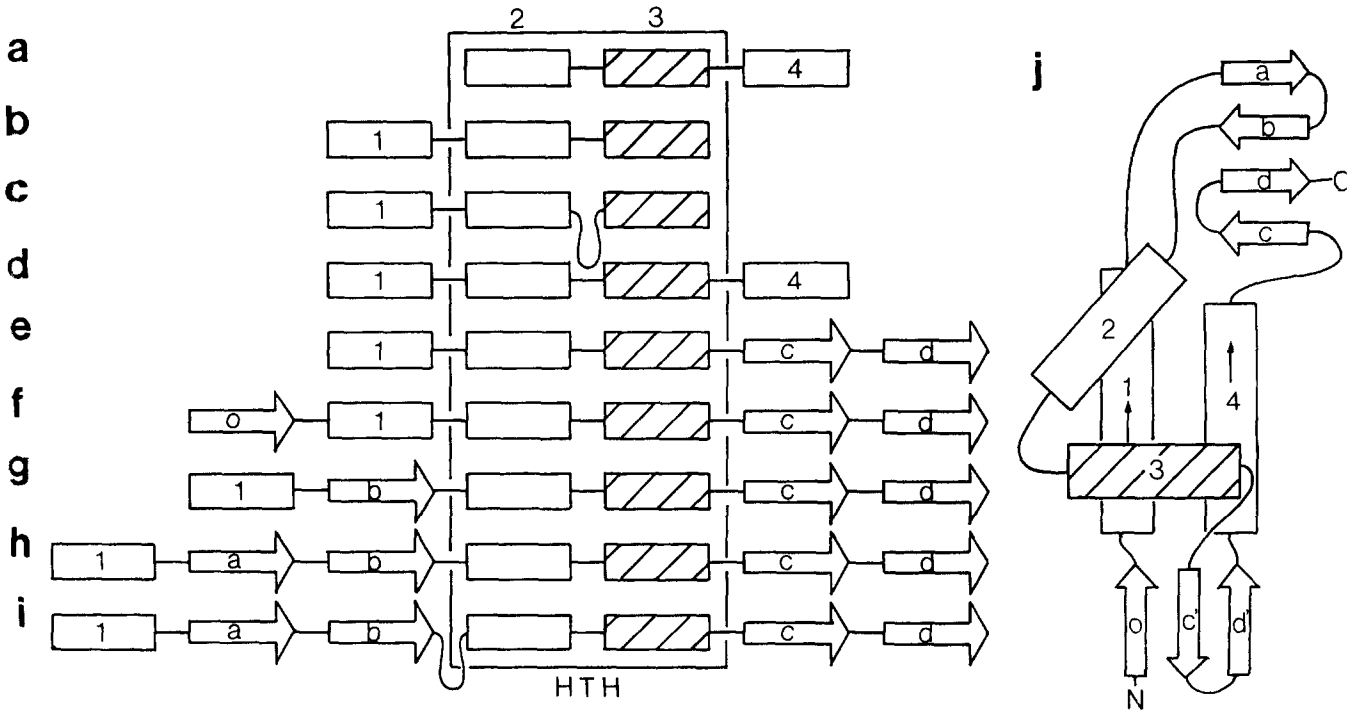


Fig. 2. Schematic drawing of multi-helical DBDs. (a)–(i) Secondary structure composition (α-helices 1–4, and β-strands o, a–d) of the DBDs: a-LacR, PurR, b-Myb, TetR, Antp, Engl, Matα2, Oct1 homeodomain, FIS, HIN, TrpR, c-LFB, d-Oct1/2 POU, λR, 434C, 434R, P22R, e-LexA, DtxR, f-λcro, g-H5, BirA, HNF3, h-CAP, HSF, ETS, i-IRF2. Helices 2 and 3 are often called the helix–turn–helix (HTH) motif and helix 3 binds to the DNA. Note that in (c) number of residues inserted between helices 2 and 3 is larger than that in (a) and that in (i) that between helices 1 and 2 is larger than that in (h). (j) A hypothetical structure which has all the structural elements. β-strands, c and d can pair with strands a and b or strand o and thus two sets of c and d are drawn.

the latter are oriented similarly, forming a multi-helical domain with the HTH (compare subfigures in the top line of Fig. 1).

Some proteins (Oct1/2 POU, λR, 434C, 434R) have both helices 1 and 4. The two helices are both roughly perpendicular to helix 3 and their N-C directions are generally similar (com-

Table 1

Name	Abbr	PDB Code	Reference	Clas*
λ repressor	λR	1LMB3***	[36–38]	d
λ cro	λcro	1CROA***	[40]	f
434 cro	434C	2CRO***, 3CRO	[41–43]	d
434 repressor	434R	1R69***, 2OR1, 1RPE, 1PER	[44–49]	d
Antennapedia	Antp	1AHD	[50]	b
BirA	BirA	1BIA***, 1BIB	[51]	g
CAP	CAP	3GAPA***, 1CGP	[53]	h
Diptheria toxin repressor	DtxR	N.D.**	[20]	e
Engrailed	Engl	1HDDC***	[54]	a
ETS domain, Fli-1	ETS	N.D.**	[55]	h
FIS	FIS	1FIA***, 3FIS	[56,57]	b
Heat shock factor	HSF	2HTS***	[58,59]	h
Hin Recombinase	HIN	1HCRA***	[23]	b
Histone H5	H5	1HST***	[18]	g
HNF3	HNF3	N.D.**	[19]	g
IRF2	IRF2	N.D.**	[60]	i
Lac repressor	LacR	1LCAA***	[61,62]	a
LexA	LexA	1LEA***	[63,64]	e
LFB1	LFB	N.D.**	[65,66]	c
Mat α2	Mat	N.D.**	[67]	b
Myb	Myb	1MSEC***	[68]	b
Oct 1&2 homeodomain	OctH	1OCT, 1HDP	[71,72]	b
Oct 1&2 Pou domain	Oct	1OCTC***, 1POU	[69–72]	d
P22 repressor	P22R	2ADR***	[73]	d
Pur repressor	PurR	N.D.**	[74]	a
Tet repressor	TetR	N.D.**	[75]	b
Trp repressor	TrpR	1WRP, 2WRP, 3WRP, 1TRR, 1TROA***	[76–78]	b

Clas*, classification shown in Fig. 2; N.D.**, not deposited; ***, the coordinates used for the score calculation in Fig. 3.

	434C	434R	OCT1	H5	λR	P22R	LexA	FIS	MYB	CAP	LacR	BirA	HSF	HIN	λC	TrpR	Engl
434C	0.00	0.16	0.53	0.58	0.68	0.84	0.91	1.13	1.07	1.25	1.27	1.42	1.31	1.16	0.87	1.29	2.15
434R	0.16	0.00	0.42	0.54	0.65	0.68	0.87	0.98	0.73	1.00	0.93	1.07	1.09	0.89	0.71	0.95	1.81
OCT1	0.53	0.42	0.00	0.84	0.40	0.56	0.93	0.82	1.09	1.27	1.29	1.43	1.38	1.18	0.74	1.79	2.17
H5	0.58	0.54	0.84	0.00	0.67	1.12	0.57	1.50	0.94	0.91	0.93	1.11	1.57	1.37	1.19	0.96	1.81
λR	0.68	0.65	0.40	0.67	0.00	0.44	0.27	0.70	0.42	0.54	0.56	0.70	0.90	0.81	0.98	1.42	1.47
P22R	0.84	0.68	0.56	1.12	0.44	0.00	0.73	0.43	0.32	0.53	0.42	0.25	0.49	0.67	0.86	1.32	1.37
LexA	0.91	0.87	0.93	0.57	0.27	0.73	0.00	1.07	0.51	0.35	0.37	0.68	1.14	0.97	1.15	1.15	1.25
FIS	1.13	0.98	0.82	1.50	0.70	0.43	1.07	0.00	0.64	0.87	0.78	0.67	0.65	0.78	0.54	1.64	1.69
MYB	1.07	0.73	1.09	0.94	0.42	0.32	0.51	0.64	0.00	0.63	0.55	0.40	0.81	0.47	0.65	1.00	1.08
CAP	1.25	1.00	1.27	0.91	0.54	0.53	0.35	0.87	0.63	0.00	0.11	0.49	0.94	1.10	1.28	1.13	1.18
LacR	1.27	0.93	1.29	0.93	0.56	0.42	0.37	0.78	0.55	0.11	0.00	0.37	0.83	1.02	1.21	1.14	1.20
BirA	1.42	1.07	1.43	1.11	0.70	0.25	0.68	0.67	0.40	0.49	0.37	0.00	0.58	0.87	1.06	1.22	1.28
HSF	1.31	1.09	1.38	1.57	0.90	0.49	1.14	0.65	0.81	0.94	0.83	0.58	0.00	0.94	1.08	1.80	1.86
HIN	1.16	0.89	1.18	1.37	0.81	0.67	0.97	0.78	0.47	1.10	1.02	0.87	0.94	0.00	0.61	1.16	0.99
λC	0.87	0.71	0.74	1.19	0.98	0.86	1.15	0.54	0.65	1.28	1.21	1.06	1.08	0.61	0.00	1.10	1.60
TrpR	1.29	0.95	1.79	0.96	1.42	1.32	1.15	1.64	1.00	1.13	1.14	1.22	1.80	1.16	1.10	0.00	0.86
Engl	2.15	1.81	2.17	1.81	1.47	1.37	1.25	1.69	1.08	1.18	1.20	1.28	1.86	0.99	1.60	0.86	0.00
HTH angle (°)	46.0	51.7	67.1	53.5	77.3	87.5	75.1	84.7	75.4	84.5	85.5	91.9	105.5	74.3	63.7	53.2	76.4

Fig. 3. Score of differences in the DBDs. Engl(1HDD) and TrpR(1TRO) are most different from the others, while HSF(2HTS), HIN(1HCR) and λcro(1CRO) are closer. The rest of the structures are subclassified into, [434C(2CRO), 434R(1R69), Oct1 POU(1OCT), H5(1HST), λR(1LMB), P22R(1ADR)] and [λR(1LMB), P22R(1ADR), LexA(1LEA), FIS(1FIA), Myb(1MSE), CAP(3GAP), LacR(1LCC), BirA(1BIA)]. Note that λR(1LMB) and P22R(1ADR) are found on the border of the two subgroups.

pare subfigures in the second line of Fig. 1, see also Fig. 2j, see also [20] for comparison of helix 1 in DtxR and helix 4 in λR).

Some of the transcription factors have two β-strands which are C-terminal to helix 3 (here referred to as strands c and d, Fig. 2e–i). Among them λcro has another β-strand at the N-terminus (strand o) which pairs with strand c (Fig. 2f,j). Others (BirA, CAP, H5, ETS, HNF3, IRF2) have one or two β-strands (strands a and b) between helices 1 and 2, and strand b pairs with strand c (Figs. 2g–j). The two types of β-sheets, the λcro type and the type of BirA etc., are formed on different sides of helix 3 (Fig. 2j).

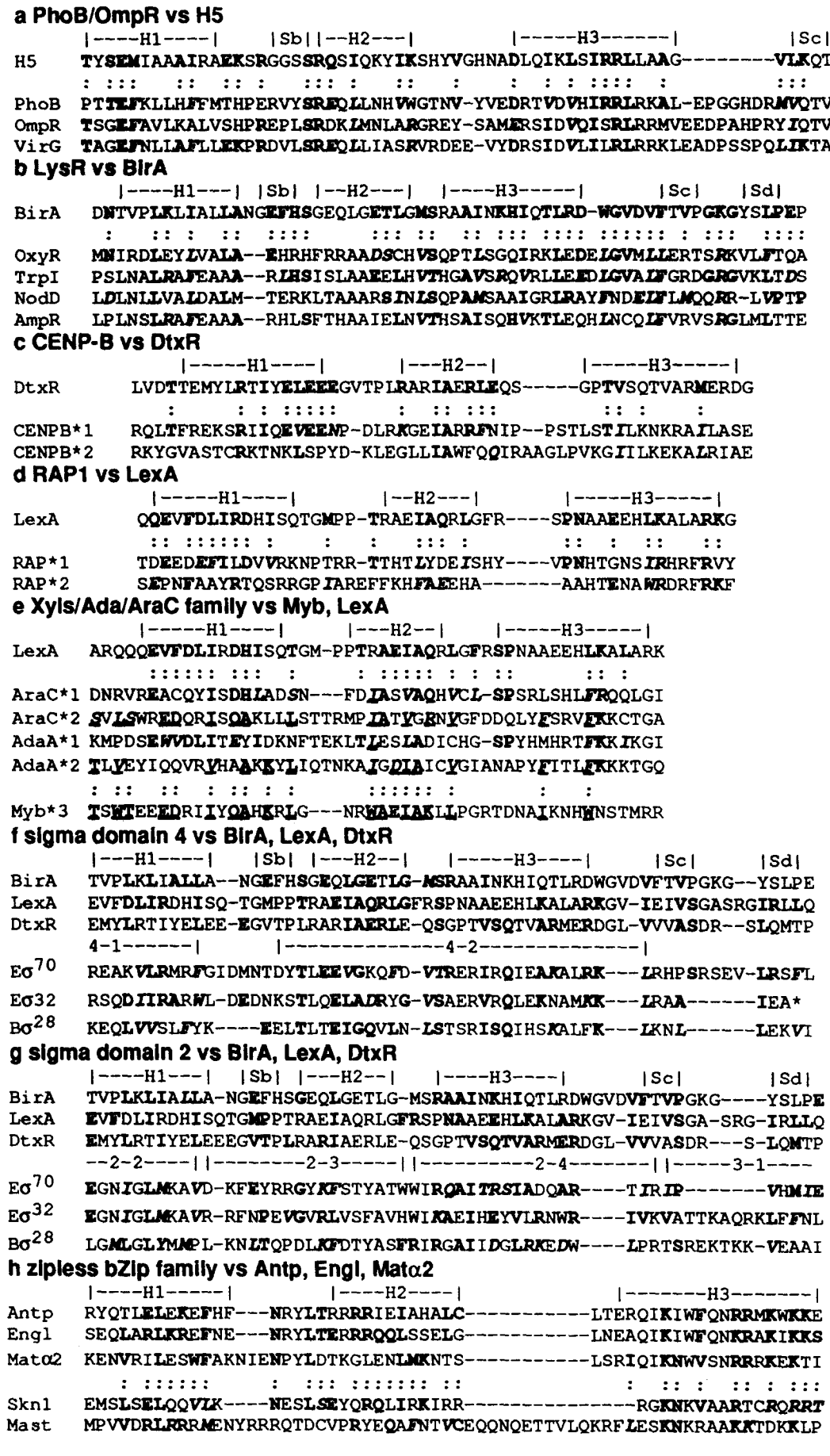
To understand similarities and differences in the structures, in particular, in the ways how the three α-helices are oriented with each other, we have calculated a score of difference between the structures in the angles between pairs of helices (Fig. 3, see also section 2). More than one structure has been determined for some of the transcription factors. The difference between two such structures is generally small. For example, the scores between different 434R structures are 0.61 or smaller, and those between TrpR structures, 2WRP, 3WRP (apo-repressor), 1TRR, 1TRO, are 0.30 or less, (another TrpR structure, 1WRP, seems more different and the score between 1WRP and 2WRP is 0.70).

Regarding the angles of the three helices, the homeodomain structure (represented by Engl in Fig. 3) is very different from the others – it is even more different than TrpR (this has been confirmed by comparing the distances between the three helices). Therefore, grouping of the homeodomain with histone H5 and CAP [19] may be problematic. Hin recombinase and λcro are more similar to the rest of proteins but still seem slightly different (this is consistent with Fig. 10 in [23]). The remaining proteins can be clustered into two subgroups; one contains proteins like 434R and histone H5, while the other includes BirA (Fig. 3). λ repressor and P22R position on the border of the two subgroups. Intriguingly, proteins of different structural compositions (Fig. 2a–i) are mixed in each subgroup. Also, classic HTH proteins, such as 434R and CAP, are dispersed through the two subgroups. Other methods of comparison might yield different results for the subgrouping.

The angle between helices 2 and 3 (here referred to as the HTH angle) was calculated for the structures (Fig. 3). To our surprise, the angles found in the classic HTH proteins (λcro, 434C, 434R, λR, P22R, CAP, and BirA) are almost the same as those found in the others, even though the others have varying number of residues between the two helices. On the other hand, the HTH angles in the classic HTH proteins vary as much as those in the others. The HTH angle averaged for the classic HTH proteins is 70.5° ± 16.8° and that for all the

→

Fig. 4. Amino acid sequences of some DBDs shown in comparison with those of known DBD structures. The amino acid positions occupied by the same residues as found in the reference protein(s) are shown bold, those occupied by closely related residues are shown bold and Italic. In (e) those residues are shown underlined in reference to Myb. In (f) and (g) σ subdomains 2–2 to 4–2 [28,29] are indicated. The sequences shown here for each family are not extensive. See [28–35] for more sequences. Asterisks (*) are used to number domains, i.e. AraC*1 indicates domain 1 of the AraC protein. Colons (:) are used to indicate the positions at which residues similar to those of the reference protein are found in the sequences. The Swissprot codes of the sequences are PHOB_ECORI[PhoB], OMPR_ECOLI[OmpR], VIRG_ECOLI[VirG], OXYR_ECOLI[OXYR], TRPI_PSEAE[TrpI], NODD_RHILT[NodD], AMPR_ENTCL[AmpR], AMPR_ENTCL[AmpR], DTXR_CORDI[DtxR]. RAP1_YEAST[Rap*1, Rap*2], ARAC_ECOLI[AraC*1, AraC*2], ADAA_BACSU[Ada*1, Ada*2]. RP70_ECOLI[Eσ^{70}], RP32_ECOLI[Eσ^{32}], RPSD_BACSU[Bσ^{28}], SKN1_CAEEL[Skn1], and MAM_DROME[Mast].


a PhoB/OmpR vs H5
|----H1----| |Sb||--H2---| |-----H3------| |Sc| |Sd|
H5 TYSEMIAAAIRAEKSRGGSSRQSIQKYIKSHYVGHNADLQIKLSIRRLLAAG---------VLKQTKGVGASGSFRLA
PhoB PTTEFKLLHFFMTHPERVYSREQLLNHVWGTNV-YVEDRTVDVHIRRLRKAL-EPGGHDRMVQTVRGTGYRFSTRF-
OmpR TSGEFAVLKALVSHPREPLSRDKLMNLARGREY-SAMERSIDVQISRLRRMVEEDPAHPRYIQTVWGLGYVFVPDGS
VirG TAGEFNLLAFLLEKPRDVLSREQLLIASRVRDEE-VYDRSIDVLILRLRRKLEADPSSPQLIKTARGAGYFFDADVQ
b LysR vs BirA
|----H1---| |Sb| |--H2---| |----H3-----| |Sc| |Sd|
BirA DNTVPLKLIALLANGEFHSGEQLGETLGMSRAAINKHIQTLRD-WGVDVFTVPGKGYSLPEP
OxyR MNIRDLEYLVALA--EHRHFRRAADSCHVSQPTLSGQIRKLEDELGVMLLERTSRKVLFTQA
TrpI PSLNALRAFEAAA--RLHSISLAAEELHVTHGAVSRQVRLLEEDLGVALFGRDGRGVKLTDS
NodD LDLNILLVALDALM--TERKLTAAARSINLSQPAMSAAIGRLRAYFNDELFLMQQRR-LVPTP
AmpR LPLNSLRAFEAAA--RHLSFTHAAIELNVTHSAISQHVKTLEQHLNCQLFVRVSRGLMLTTE
c CENP-B vs DtxR
|-----H1----| |---H2--| |-----H3-----|
DtxR LVDTTEMYLRTIYELEEEGVTPLRARIAERLEQS-----GPTVSQTVARMERDG
CENPB*1 RQLTFREKSRIIQEVEENP-DLRKGEIARRFNIP--PSTLSTILKNKRAILASE
CENPB*2 RKYGVASTCRKTNKLSPYD-KLEGLLIAWFQQIRAAGLPVKGIILKEKALRIAE
d RAP1 vs LexA
|-----H1----| |--H2---| |-----H3-----|
LexA QQEVFDLIRDHISQTGMPP-TRAEIAQRLGFR----SPNAAEEHLKALARKG
RAP*1 TDEEDEFTLDVVRKNPTRR-TTHTLYDEISHY----VPNHTGNSIRHRFRVY
RAP*2 SEPNFAAYRTQSRRGPIAREFFKHFAEEHA-------AAHTENAWRDRFRKF
e XylS/Ada/AraC family vs Myb, LexA
|-----H1----| |---H2--| |-----H3-----|
LexA ARQQQEVFDLIRDHISQTGM-PPTRAEIAQRLGFRSPNAAEEHLKALARK
AraC*1 DNRVREACQYISDHLADSN---FDIASVAQHVCL-SPSRLSHLFRQQLGI
AraC*2 SVLSWREDQRISQAKLLLSTTRMPIATVGRNVGFDDQLYFSRVFKKCTGA
AdaA*1 KMPDSEWVDLITEYIDKNFTEKLTLESLADICHG-SPYHMHRTFKKIKGI
AdaA*2 TLYEYIQQVRVHAAKKYLIQTNKAIGDIAICVGIANAPYFITLFKKKTGQ
Myb*3 TSHTEEEDRIIYQAHKRLG---NRWAEIAKLLPGRTDNAIKNHWNSTMRR
f sigma domain 4 vs BirA, LexA, DtxR
|---H1----| |Sb| |---H2--| |-----H3----| |Sc| |Sd|
BirA TVPLKLIALLA--NGEFHSGEQLGETLG-MSRAAINKHIQTLRDWGVDVFTVPGKG--YSLPE
LexA EVFDLIRDHISQ-TGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGV-IEIVSGASRGIRLLQ
DtxR EMYLRTIYELEE-EGVTPLRARIAERLE-QSGPTVSQTVARMERDGL-VVVASDR--SLQMTP
4-1------| |--------------4-2--------------|
Eσ70 REAKVLRMRFGIDMNTDYTLEEVGKQFD-VTRERIRQIEAKALRK---LRHPSRSEV-LRSFL
Eσ32 RSQDIIRARWL-DEDNKSTLQELADRYG-VSAERVRQLEKNAMKK---LRAA------IEA*
Bσ28 KEQLVVSLFYK----EELTLTEIGQVLN-LSTSRISQIHSKALFK---LKNL------LEKVI
g sigma domain 2 vs BirA, LexA, DtxR
|----H1---| |Sb| |---H2--| |----H3----| |Sc| |Sd|
BirA TVPLKLIALLA-NGEFHSGEQLGETLG-MSRAAINKHIQTLRDWGVDVFTVPGKG----YSLPE
LexA EVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGV-IEIVSGA-SRG-IRLLQ
DtxR EMYLRTIYELEEEGVTPLRARIAERLE-QSGPTVSQTVARMERDGL-VVVASDR---S-LQMTP
--2-2----||--------2-3-----||----------2-4-------||-----3-1----
Eσ70 EGNIGLMKAVD-KFEYRRGYKFSTYATWWIRQAITRSIADQAR----TIRIP-------VHMIE
Eσ32 EGNIGLMKAVR-RFNPEVGVRLVSFAVHWIKAEIHEYVLRNWR----IVKVATTKAQRKLFFNL
Bσ28 LGMLGLYMMPL-KNLTQPDLKFDTYASFRIRGAIIDGLRKEDW----LPRTSREKTKK-VEAAI
h zipless bZip family vs Antp, Engl, Matα2
|----H1-----| |----H2----| |-------H3-------|
Antp RYQTLELEKEFHF---NRYLTRRRRIEIAHALC----------LTERQIKIWFQNRRMKWKKE
Engl SEQLARLKREFNE---NRYLTERRRQQLSSELG----------LNEAQIKIWFQNKRAKIKKS
Matα2 KENVRILESWFAKNIENPYLDTKGLENLMKNTS----------LSRIQIKNWVSNRRRKEKTI
: :::::: : ::: ::::::: :: : :: : :: : :::
Skn1 EMSLSELQQVLK----NESLSEYQRQLIRKIRR--------------RGKNKVAARTCRQRRT
Mast MPVVDRLRRRMENYRRRQTDCVPRYEQAFNTVCEQQNQETTVLQKRFLESKNKRAAKKTDKKLP

proteins excluding of HSF, 70.9° ± 13.7° (the HTH angle of HSF is very different from the others as its helix 2 is kinked, see Fig. 1, the HTH angle averaged including of HSF is 72.7° ± 15.4°). The proteins belonging to each subgroup naturally have similar HTH angles; the subgroup of Oct1 POU has lower values, 46.0°-87.5°, than the subgroup of Myb, 75.4°-105.5°. Therefore, in this regard, the classic helix–turn–helix motif does not appear to be so important or independent as it was once believed [27], but the whole globular domain must be considered for understanding the structure.

3.2. Comparison of amino acid sequences

We have also compared amino acid sequences of some other transcription factors of currently unknown structures with those of crystal/NMR structures (Fig. 4). When structural information is taken into account, transcription factors in the PhoB/OmpR family resemble histone H5, those in the LysR family-BirA, CENP-B-DtxR, RAP1-LexA, factors in the XylS/Ada/AraC family-Myb/LexA, σ domains 2 and 4-BirA/LexA/DtxR, and the zipless bZip family-Antp/Engl/Matα2. For example, the PhoB/OmpR family has the (Arg/Trp)-Gly-Hydrophobic-Gly sequence, which is aligned with the Lys-Gly-Val-Gly sequence of H5. Arg, Trp, Lys are the residues often used for two functions in crystal structures: binding to a DNA phosphate and creating hydrophobic environment with their stems. The two Gly residues are likely to be important for making a particular turn between the two β-strands. Although 7–8 residues are inserted between the third helix and the β-sheet, this might not cause a serious problem as the number of residues found in this part varies among the proteins which have the same secondary composition as that of H5.

The amino acid sequences listed in Fig. 4 indeed have features which coincide with the structures suggested by the reference proteins. The three helices predicted are short and have only a few turns. In many of the predicted helices, one side is occupied by hydrophobic residues. For example, helix 3 of the PhoB/OmpR family has the sequence, Ile/Val-XXX-Ile-XX-Leu. The position which is C-terminal to the conserved first hydrophobic position in the predicted helix 2 is occupied by small residues, Ala or Gly, except for in that of H5 and the PhoB/OmpR family. This might be important for particular packing angles of the three helices. Insertion of residues is found only between two secondarly structural components, which might not disturb the packing much. Gly residues are found at some putaive β-hairpins (for example, between strands c and d in the PhoB/OmpR and LysR families). Detailed discussion will be given elsewhere.

The above features can be used not only to predict the structural compositions of the transcription factors but also suggest the configuration of their three putative α-helices. For example, histone H5 and BirA share the same structural composition, three α-helices and three β-strands, but the two proteins have slightly different configurations (Figs. 1 and 3). Since the PhoB/OmpR family resembles H5, and since the LysR family resembles BirA, the PhoB and LysR families probably share the same structural composition but may have slightly different packing arrangements.

In this paper we have discussed similarities and differences among the multi-helical DNA-binding domains. Three α-helices can form probably the smallest globular structure and thereby stabilise the recognition helix: two α-helices can be placed on a plane but three α-helices can create a truely three-dimensional structure, placing hydrophobic residues inside the structure and hydrophilic residues outside. This economical structural motif is repeatedly found among the known DNA-binding proteins. These can be classified into subgroups which do not always correlate with evolutionary relationships, which suggests how well the multi-helical fold is designed and that many DBDs evolutionally converged into the fold.

Acknowledgements: SEB is principally supported by St. John's College Benefactors' Scholarship.

References

[1] Harrison, S.C. (1991) Nature 353, 715–719.
[2] Harrison, S.C. and Aggarwal, A.K. (1990) Annu. Rev. Biochem. 59, 933–969.
[3] Pavletich, N.P. and Pabo, C.O. (1991) Science 252, 809–817.
[4] Pavletich, N.P. and Pabo, C.O. (1993) Science 261, 1701–1707.
[5] Luisi, B. F., Xu, X.W., Otwinowski, Z., Freedman, J.P., Yamamoto, K.R. and Sigler, P.B. (1991) Nature 352, 497–505.
[6] Schwabe, J.W., Chapman, L., Finch, J.T. and Rhodes, D. (1993) Cell 75, 567–578.
[7] Rastinejad, F., Perlmann, T., Evans, R.M. and Sigler, P.B. (1995) Nature 375, 203–211.
[8] Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C. (1992) Nature 356, 409–414.
[9] Ellenberger, T.E., Brandle, C.S., Struhl, K. and Harrison, S.C. (1992) Cell 71, 1223–1237.
[10] König, P. and Richmond, T. (1993) J. Mol. Biol. 233, 139–154.
[11] Ferré-D'Amaré, A.R., Pognone, P., Roeder, R.G. and Burley, S.K. (1993) Nature 363, 38–45.
[12] Ferré-D'Amaré, A.R., Pognone, P., Roeder, R.G. and Burley, S.K. (1994) EMBO J. 13, 180–189.
[13] Ma, P.C., Rould, M.A., Weintraub, H. and Pabo, C.O. (1994) Cell 77, 451–459.
[14] Hegde, R.S., Grossman, S.R., Laimins, L.A. and Sigler, P.B. (1992) Nature 359, 505–512.
[15] Cho, Y., C., Gorina, S., Jeffrey, P.D. and Pavletich, N.P. (1994) Science 265, 346–355.
[16] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–321.
[17] Ohlendolf, D.H., Anderson, W.F., Lewis, M., Pabo, C.O. and Matthews, B.W. (1983) J. Mol. Biol. 169, 757–769.
[18] Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993) Nature 362, 219–223.
[19] Clark, M.L. Halay, E.D., Lai, E. and Burley, S.K. (1993) Nature 364, 412–420.
[20] Qiu, X., Verlinde, C.L.M.J., Zhang, S., Schmitt, M.P., Holmes, R.K. and Hol, W.G.J. (1995) Structure 3, 87–100.
[21] Holm, L., Dander, C., Rüterjans, H., Schnarr, M., Fogh, R., Boelens, R. and Kaptein, R. (1994) Protein Eng. 7, 1449–1453.
[22] Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) Nature 290, 754–758.
[23] Feng, J.A., Johnson, R.C. and Dickerson, R.E. (1994) Science 263, 348–355.
[24] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 525–542.
[25] Kabsch, W. and Sander, C. (1983) Biopolymers. 22, 25777–2637.
[26] Sayle, R.A. (1995) RasMol. ftp://ftp.dcs.ed.ac.uk/pub/rasmol/.
[27] Brennan, G. and Mathews, B.W. (1989) J. Biol. Chem. 264, 1903–1989.
[28] Helmann, J.D. and Chamberlin, M.J. (1988) Annu. Rev. Biochem. 57, 839–872.
[29] Lonetto, M., Bribskov, M. and Gross, C.A. (1992) J. Bacteriol. 174, 3834–3849.
[30] Bowerman, B., Eaton, B.A. and Priess, J.R. (1992) Cell, 68, 1061–1075.
[31] Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) J. Mol. Biol. 190, 37–44.

[32] Suzuki, M. and Makino, K. (1995) Proc. Japan Acad. 71B, 132–137.
[33] Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988) Proc. Natl. Acad. Sci. USA 85, 6602–6606.
[34] Nitta, Y. and Suzuki, M. (1995) Proc. Jpn. Acad., 71 B, 193–197.
[35] Suzuki, M. and Makino, K. (1995) Proc. Jpn. Acad., 71 B, 132–137.
[36] Pabo, C.O. and Lewis, M. (1992) Nature 298, 443–447.
[37] Clarke, N.D., Beamer, J.L., Goldberger, H.R., Berkower, C. and Pabo, C.O. (1991) Science 254, 267–270.
[38] Jordan, S.R. and Pabo, C.O. (1988) Science 242, 893–899.
[40] Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) Nature 290, 754–759.
[41] Mondragón, A., Wolberger, C. and Harrison, S.C. (1989) J. Mol. Biol. 205, 179–188.
[42] Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1990) Cell 67, 517–528.
[43] Mondragón, A. and Harrison, S.C. (1991) J. Mol. Biol. 219, 321–334.
[44] Neri, D., Billeter, M. and Wüthrich, K. (1992) J. Mol. Biol. 223, 743–767.
[45] Mondragón, A., Subbiah, S., Almo, S.C., Drottar, M. and Harrison, S.C. (1989) J. Mol. Biol. 205, 189–205.
[46] Anderson, J.E., Ptashne, M. and Harrison, S.C. (1987) Nature 326, 846–852.
[47] Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. and Harrison, S.C. (1988) Nature 242, 899–907.
[48] Rodgers, D.W. and Harrison, S.C. (1993) Structure 1, 227–239.
[49] Shimon, L.J.W. and Harrison, S.C. (1993) J. Mol. Biol. 232, 826–838.
[50] Billeter, M., Quian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. (1993) J. Mol. Biol. 234, 1084–1094.
[51] Wilson, K.P., Schewchuk, L.M., Brennan, R.G., Otsuka, A.J. and Matthews, B.W. (1992) Proc. Natl. Acad. Sci. USA 89, 9257–9261.
[52] Weber, I.T. and Steitz, T.A. (1987) J. Mol. Biol. 198, 311–326.
[53] Schultz, S.C., Shields, G.C. and Steize, T.A. (1991) Science 253, 1001–1007.
[54] Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) Cell 63, 579–590.
[55] Liang, H., Mao, X., Olejniczak, E.T., Nettesheim, D.G., Yu, L., Meadows, R.P., Thompson, C.B. and Fesik, S.W. (1994) Nature Struct. Biol. 1, 871–875.
[56] Kostrewa, D.K., Granzin, J.G., Koch, C., Choe, H.-W., Raghunathan, S., Wolf, W., Labahn, J., Kahmann, R. and Sanger, W. (1991) Nature 366, 178–182.
[57] Yuan, H.S., Finkel, S.E., Feng, J.-A., Kaczor-Grzeskowiak, M., Johnson, R.C. and Dickerson, R.E. (1991) Proc. Natl. Acad. Sci. USA 88, 9558–0562.
[58] Harrison, C.J., Bohm, A.A. and Nelson, C.W. (1994) Science 242, 893–899.
[59] Damberger, F.E., Pelton, J.G., Harrison, C.J., Nelson, H.C. and Wemmer, D.E. (1994) Protein Science 3, 1806–1821.
[60] Uegaki, K., Shirakawa, M., Harada, H., Taniguchi, T. and Kyogoku, Y. (1995) FEBS Lett. 359, 184–188.
[61] Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M. Boelens, R. and Van Gunsteren, W.F. (1985) J. Mol. Biol. 182, 179–182.
[62] Chuprina, V.P., Rullmann, J.A.C., Lamerichs, R.M.J.N., Van Boom, J.H., Boelens, R. and Kaptein, R. (1993) J. Mol. Biol. 234, 446–462.
[63] Lamerichs, R.M.J.N., Padilla, A., Boelens, R., Kaptein, R., Ottleben, G., Rüterjans, H., Granger-Schnarr, M., Oeltal, P. and Schnarr, M. (1989) Proc. Natl. Acad. Sci. USA 86, 6863–6867.
[64] Fogh, R.H., Ottleben, G., Rüterjans, H., Schnarr, M., Boelens, R. and Kaptein, R. (1994) EMBO J. 13, 3936–3944.
[65] Ceska, T.A., Lamers, M., Monaci, P., Nicosia, A., Cortese, R. and Suck, D. (1993) EMBO J. 12, 1805–1810.
[66] Lehming, B., de Francesco, R., Tomei, L., Cortese, R., Otting, G. and Wüthrich, K. (1993) EMBO J. 12, 1797–1803.
[67] Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991) Cell 67, 517–528.
[68] Ogata, K., Morikawa, S., Nakamura, H., Serikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y., (1994) Cell 79 639–648.
[69] Dekker, N., Cox, M., Boelens, R., Verrijzer, C.P., Van der Vilet, P.C. and Kaptein, R. (1993) Nature 362, 852–855.
[70] Assa-Munt, N., Mortishire-Smith, R.J., Aurora, R., Herr, W. and Wright, P.E. (1993) Cell 73, 193–205.
[71] Sivaraja, M., Botfield, M.C., Mueller, M., Jancso, A. and Weiss, M.A. (1994) Biochemistry 33, 9845–9855.
[72] Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O. (1994) Cell 77, 21–32.
[73] Sevilla-Sierra, P., Otting, G. and Wüthrich, K. (1994) J. Mol. Biol. 235, 1003–1020.
[74] Schumacher, M.A., Choi, K.Y., Zalkin, H. and Brennan, R.G. (1994) Science 266, 763–770.
[75] Hinrichs, W., Kisker, C., Düvel, M., Müller, A., Tovar, K., Hillen, W. and Sanger, W. (1994) Science 264, 418–420.
[76] Otwinowski, Z., Shevitz, R.W., Zhang, R.G., Lawson, C., Joachimiak, A., Mamorstein, R.Q., Luisi, B.F. and Sigler, P.B. (1988) Nature 335, 321–329.
[77] Lawson, C.L. and Carrey, J. (1993) Nature 366, 178–182.
[78] Zhang, H., Zhao, D., Revingston, M., Lee, W., Jia, X., Arrowsmith, C. and Jardezky, O. (1994) J. Mol. Biol. 238, 592–614.