Volume 280, number 1, 11–16 FEBS 09455 March 1991

ADONIS 001457939100162F

Correct intron splicing generates a new type of a putative zinc-binding domain in a transcriptional activator of *Aspergillus nidulans*

Peter Kulmburg[1], Thierry Prangé[2], Martine Mathieu[1], Daria Sequeval[1], Claudio Scazzocchio[1] and Béatrice Felenbok[1]

[1]*Institut de Génétique et Microbiologie, Bâtiment 409, Centre Universitaire Paris-Sud, 91405 Orsay Cedex, France* and [2]*Laboratoire pour l'Utilisation du Rayonnement Electromagnétique, Bâtiment 209D, Centre Universitaire Paris-Sud, 91405 Orsay Cedex, France*

Received 17 December 1990

alcR is the pathway-specific transcriptional activator of the ethanol regulon in the filamentous fungus, *Aspergillus nidulans*. The deduced amino acid sequence of a cDNA clone, including the 5' part of the *alcR*-mRNA, shows that a putative Zn-binding domain of the all-cysteine class, exemplified by GAL4 is present. This structure presents some striking features. At variance with other structures of this class, the binding domain is strongly asymmetrical. Model building indicates that the zinc-binding motif of alcR could adopt an helix-turn-helix structure. We propose that the DNA-binding motif of alcR could participate in two types of DNA-binding structures: the zinc-cluster and the helix-turn-helix.

Zn-binding domain; Ethanol regulon; Model building; *Aspergillus nidulans*

1. INTRODUCTION

In the ascomycete fungi, several transcriptional activators contain a highly conserved amino-terminal DNA-Zn-binding domain, of the general structure Cys-Xaa_2-Cys-Xaa_6-Cys-Xaa_{5-8}-Cys-Xaa_6-Cys reviewed in [1,2]. It has been suggested that these structures could chelate a Zn(II) atom through the coordination of the metal by the thiol groups of 4 cysteines [3] as was found by Klug and co-workers for the TFIIIA transcription factor in which two cysteines and two histidines are involved in the 'zinc finger structure' [4].

The prototype of this 6 cysteines Cys_2-Cys_2 class in fungi is the product of the GAL4 gene, a positive regulatory gene of *S. cerevisiae* [3]. Evidence that Zn(II) is actually chelated by these peptide-motives and is necessary for the binding of the protein to the DNA was demonstrated not only for GAL4 [3] and for the glucocorticoid receptor which belongs to the Cys_2-Cys_2 class [5], but also for factors of the Cys_2-His_2 class, TFIIIA [6,7] and ADR1 [8].

Two types of structures have been suggested for the DNA-binding domain of the GAL4 protein. A zinc-finger motif analogous to the *Xenopus* Cys_2-His_2 transcription factor IIIA [4] has been proposed by Johnston [3]. The second model proposed recently by Pan and Coleman [9] involves the binding of the 6 cysteines to two Zn(II) atoms folded in a $Zn(II)_2$-Cys_6-binuclear-complex. The presence of two zinc atoms has been demonstrated in the DNA-binding domains of GAL4 [9,10] and LAC9 [11].

The *alcR* gene is the pathway-specific, positively-acting regulatory gene of the ethanol utilisation pathway of *Aspergillus nidulans*. An active *alcR* gene product is necessary for the induction of both alcohol dehydrogenase I and aldehyde dehydrogenase [12,13]. The expression of the *alcR* gene is itself inducible, autoregulated and subjected to carbon catabolite repression [14,15]. We have published previously the sequence of this gene [16]. It is clear that in the second exon a number of residues appear which constitute the second half of a canonical Zn-finger of the Cys_2-Cys_2 class, and moreover that the residues which follow the finger structure, and which are conserved in the fungal genes of this class are also present. It was pointed out (Creaser, personal communication, [17]) that a perfect 'finger' would appear if the first exon was read in a −1 open reading frame. The two halves of the 'finger' would be separated by an intron, and in order to generate a canonical finger, splicing would have to occur at a sequence quite different to the fungal intron 3' consensus sequences [18].

In this communication, we establish that the regulatory *alcR* gene product of the ascomycete *Aspergillus nidulans* contains a putative Zn-binding domain. This domain belongs to the general class exemplified by GAL4 but presents some striking new features. It is strongly asymmetrical with respect to the conserved central cysteine and moreover, we suggest that this extended asymmetrical loop forms an arm of an helix-turn-helix motif, as found in a number of prokaryotic and eukaryotic regulatory proteins.

Correspondence address: B. Felenbok, Institut de Génétique et Microbiologie, Bâtiment 409, Centre Universitaire Paris-Sud, 91405 Orsay Cedex, France

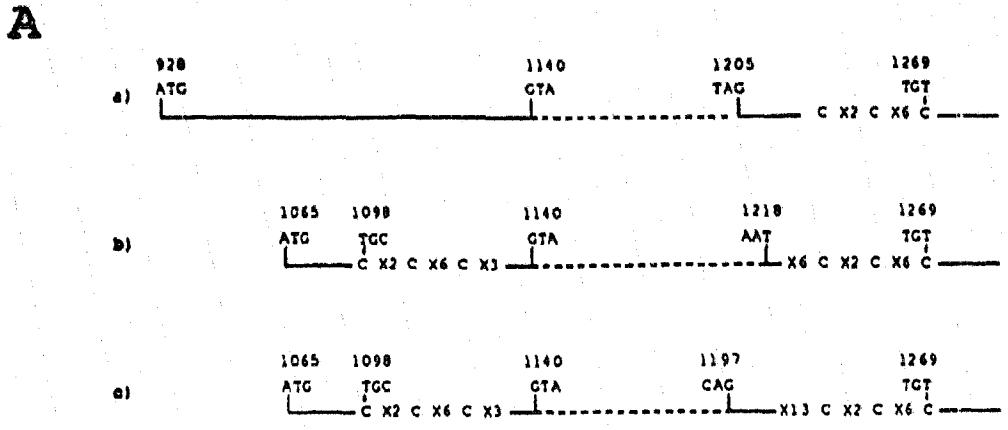
2. MATERIALS AND METHODS

2.1. *cDNA clone construction*

Two oligodeoxynucleotide primers were used to synthesize the cDNA clone including the 5' part of the *alcR* mRNA following the Saiki procedure [19]. These are underlined on the sequence in Fig. 1B. We thank D. Gwynne of Allellix (Toronto, Canada) for providing the oligonucleotides.

2.2. *Modelling of the alcR putative Zn-binding domain*

Our model for the alcR Zn-binding domain was obtained using the Hydrophobic Cluster Analysis (HCA), [20] as a preliminary step for discriminating the structurally defined regions. The MANOSK [21] and the FROD [22] graphic programs with internally wired amino-acid dictionaries were employed to build the molecular structure, starting from the X-ray coordinates of aspartyl-transcarbamylase of *Escherichia coli* which contains a Cys-X_2-Cys-X_2-Cys-X_4-Cys motif



B

```
tcggaactccgggaaccccaaggcctcactgatcagccctaaaatgagtggtggatattccatg  832

ttccccacattaagcttcttctcattggcctttgggcactccgactccgctcgaccgcgacgaa  896

ccctggccaggggaacctgggtatccaggtcatgtgctggaaccccgcatgctccaggcctcat  960

tccccagatttcgactccatggctgacagcatccagaccgcacgtttttggacttctgatacca 1024

                                          Met Ala Asp Thr Arg Arg     6
tctatacagatatcactgtcgatattctcctgcacacagc ATG GCA GAT ACG CGC CGA 1082

Arg Gln Asn His Ser Cys Asp Pro Cys Arg Lys Gly Lys Arg Arg Cys     22
CGC CAG AAT CAT AGC TGC GAT CCC TGT CGC AAG GGC AAG CGA CGC TGT  1130

Asp Ala Pro                                                        25
GAT GCC CCG gtaggttgccgatatcggctccccagcgtgtgcactgacagtcgctgagatg 1191

         Glu Asn Arg Asn Glu Ala Asn Glu Asn Gly Trp Val Ser Cys    39
taacacag GAA AAT AGA AAC GAG GCC AAT GAA AAC GGC TGG GTT TCG TGT 1241

Ser Asn Cys Lys Arg Trp Asn Lys Asp Cys Thr Phe Asn Trp Leu Ser     55
TCA AAT TGC AAG CGT TGG AAC AAG GAT TGT ACC TTC AAT TGG CTC TCA  1289

Ser Gln Arg Ser Lys Ala Lys Gly Ala Ala Pro Arg Ala Arg Thr Lys     71
TCC CAA CGC TCC AAG GCA AAA GGG GCT GCA CCT AGA GCG AGA ACA AAG  1337

Lys Ala Arg Thr Ala Thr Thr Thr Ser Glu Pro Ser Thr Ser Ala Ala     87
AAA GCC AGG ACC GCA ACA ACC ACC AGT GAA CCA TCA ACT TCA GCT GCA  1385

Thr Ile Pro Thr Pro Glu Ser Asp Asn His Asp Ala Pro Pro Val Ile    103
ACA ATC CCT ACA CCG GAA AGT GAC AAT CAC GAT GCG CCT CCA GTC ATA  1433

Asn Ser His Asp Ala Leu Pro Ser Trp Thr Gln Gly Leu Leu Ser His    119
AAC TCT CAC GAC GCG CTC CCG AGC TGG ACT CAG GGG CTA CTC TCC CAC  1481
```

Fig. 1. Deduced amino-acid sequence of the N-terminal part of the alcR protein. Panel A shows the 3 different possibilities of splicing described in the text. Panel B shows the sequence of the cDNA clone at the 5' part of the *alcR* mRNA. The oligodeoxynucleotide primers used to synthesize the cDNA are underlined. The complete sequence has already been published [16]. The sequence of the intron and the 5' flanking region is indicated in small characters. The translated amino-acid sequence is shown above the DNA sequence.

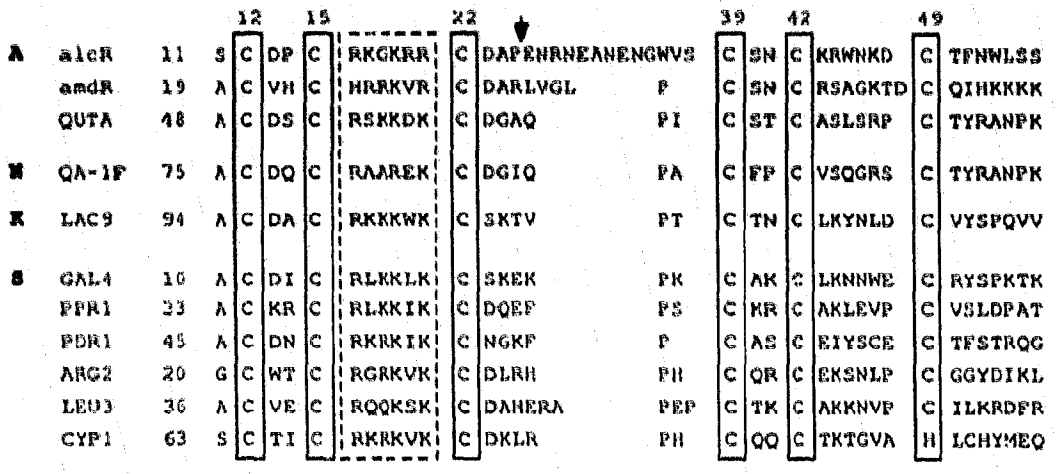


Fig. 2. Aligned sequences of 'Zn-binding motif' of the Cys_2-Cys_2 binding domains from fungi. A, *Aspergillus nidulans:* alcR (this paper); amdR [26]; QUTA [27]; N, *Neurospora crassa:* QA-1F [28]; K, *Kluyveromyces lactis:* LAC9 [29]; S, *Saccharomyces cerevisiae:* GAL4 [3]; PPR1 [30]; PDR1 [31]; ARG2 [32]; LEU3 [33]; CYP1 [34]. The cysteine residues and the conserved basic region are boxed. The arrow indicates the location of the splicing point of the intron of *alcR*.

(coordinates from the Protein Data Bank), [23]. However, important structural modifications had to be made in the loop as the hydrophobicity of the two fingers (alcR vs ATCase) are completely different. The modelling implied also a short deletion between the two cysteines in the X_4 region of ATCase structure and the insertion of two additional cysteines at positions 22 and 49 for building the binuclear Zn complex.

The overall structure motif was submitted to energy minimization algorithm to relieve steric clashes and improve the side chain packing and hydrogen bonding (Program XPLOR, [24] with the energy parameters described by Brooks et al., [25]). The calculations were started at a value of total energy equal to $E_{total} = 4.5 \times 10^5$ kcal, the final value after 80 cycles of energy minimization, was lowered as $E_{total} = 1.5$ kcal. The total energy term used in the calculations comprises the covalent bond, bond angle, dihedral angle, planar (e.g. Phe, Trp,...), intra-molecular van der Waals and explicit hydrogen bond energy terms.

3. RESULTS AND DISCUSSION

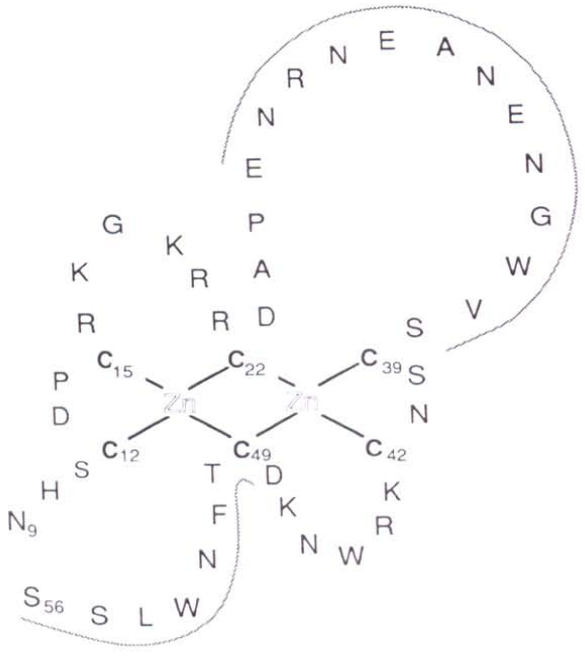
3.1. *Sequence of the N-terminal part of the alcR protein*

A cDNA clone including the 5′ part of *alcR*-mRNA has been constructed in order to know if such a 'Zn-finger' structure could exist. We are faced with 3 possibilities (Fig. 1A): (a) the published reading frame (0) of the first exon is correct [16], splicing occurs at the place indicated in the published sequence and the second exon contains a residual, perhaps 'fossil' sequence of canonical 'Zn-finger'; (b) a new reading frame (−1) is the correct one, there is a canonical 'Zn-finger', but splicing occurs at the 3′ end of the intron at an heterodox site; (c) the −1 reading frame of the first exon is correct, splicing occurs at a different but still orthodox site, and an asymmetric structure in relation to the conserved central cysteine containing 23 residues is generated in between the consensus cysteine residues. Alternatives b and c imply that the peptide starts at an ATG located at nucleotide 1065. The sequence of the corresponding genomic DNA and the deduced amino-acid sequence of the cDNA clone are shown in Fig. 1B. Alternatives a and b are excluded since the splicing occurs at the 3′ end of the intron at an orthodox CAG site, located at 1197, consistent with the S1 mapping experiments (which have an accuracy of ±10 nucleotides) but at a variance (8 nucleotides upstream) with the site previously published.

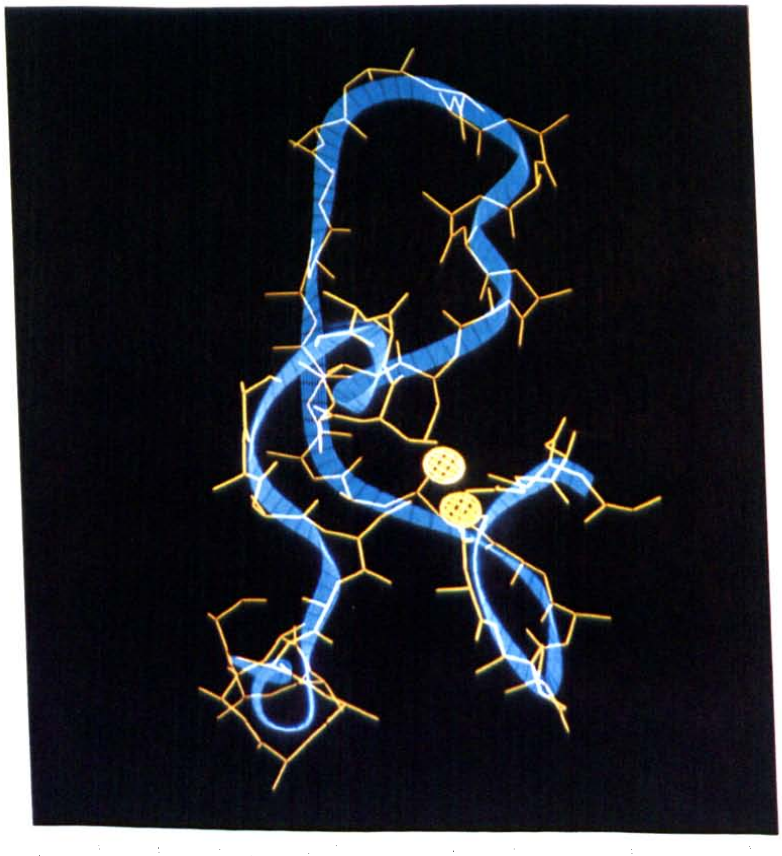
The results show conclusively that the correct reading frame is that indicated in Fig. 1A as −1 (possibility c), and that the translated sequence contains a putative Zn-domain.

3.2. *Comparison of 'Zn-binding' motives from ascomycetes fungi*

The similarity between the putative Zn-domain of the *alcR* gene product and other structures of the same kind is shown in Fig. 2 [26–34]. A strong similarity, including the presence of universally conserved residues, is evident up to the 13th residue of the amino-terminal end of the motif with a preponderance of positive charges. Then 15 residues are found, where usually 6 are present in most of the other fungal regulatory gene products of this class. It is interesting that among the universal residues of the loop of these proteins, there is a proline, one (or exceptionally two) residues before the 4th cysteine, this residue is absent in alcR, but a proline is present at position 25 after the 3rd cysteine. Intron splicing occurs between this proline and the following residue, with conservation of the reading frame. At the carboxy terminus of the motif a weaker conservation of aminoacids is observed among the proteins depicted in Fig. 2.



A



B

Fig. 3. Proposed binuclear metal complex formed by the alcR protein. Panel A, amino-acid sequence of alcR putative binuclear-Zn(II)-complex. The grey lines indicate the predicted α-helical structures. Panel B, a view of the C$^{\alpha}$ backbone (orange) underlined by ribbon representation is shown around the two putative Zn(II) atoms coordinated to the 6 available cysteines. One Zn(II) atom (upper on the figure) serves as a link between the two α helices. The N terminal is on the right and the C terminal on the left.

3.3. *Modelling of the alcR putative Zn-binding domain*

The question remains whether this heterodox structure is able to adopt a folded domain as in other DNA-binding proteins. Comparison of the Zn-binding domain of DNA-binding proteins with metalloproteins of known 3D conformation, prior to crystallographic or spectroscopic studies, has led to detailed structural models [35,36]. We have constructed a structural model for the alcR putative binuclear Zn-domain starting from the X-ray coordinates of aspartyl-transcarbamylase of *Escherichia coli* which contains a Cys-X_2-Cys-X_{21}-Cys-X_4-Cys motif [23]. The cysteines linkages (Cys-22 and Cys-49) were extended to a second Zn(II) atom (Zn-S distance: 2.08 Å) using the MANOSK and FRODO programs, as indicated in the experimental section. We have determined that there are no constraints that could prevent this motif to adopt a folded structure in which two Zn(II) atoms are tetra-coordinated by 6 cysteines in a folded structure compatible with those determined by NMR studies on binuclear Cd(II)-complexes [9]. This resulted in the asymmetrical clover-leaf structure ribbon representation given in Fig. 3, with side chains omitted for purpose of clarity. The extended loop of ATCase comprises an antiparallel β-sheet with two turns (in which a proline is involved) [23]. In alcR, the corresponding extended loop was predicted as an α-helix for the C-terminal part of the loop. The sequence from residue 16 to 24, following the two first cysteines, is structurally not well defined, since it contains a high number of polar basic residues, however this region is only involved in two small loops connecting these complexed cysteines. A β-turn engendered by Pro-25 connects this region to the predicted α-helical structure which extends from residue 26 to 38. The 'post Zn-finger' sequence following Cys-49 is also predicted to be an α-helical region.

It was shown that in ADR1 the Cys-Cys region of the structure comprises a loosely structured tract [8]. The preponderance of positively charged residues which is at the root of the loose structure of the tract between cysteines 15 and 22 in alcR is also a general characteristic of the 'GAL4' class of Zn-binding domains (Fig. 2). The structural model of the alcR putative DNA-binding region shares the features that an helix-turn-helix appears but with the turn region replaced by a putative bi-nuclear Zn domain. It is noteworthy that an α-helix is also present in the C-terminal structures of TFIIIA [35], XFin [37] and ADR1 [8]. In the Zn-binuclear domain of GAL4 [9], helical tracts have been predicted. In the glucocorticoïd receptor, a structure encompassing two α-helices has been found by Cd(II) NMR studies. The α-helix after the first zinc finger would be the recognition helix located in the DNA major groove [38].

A number of different relative orientations are equally possible for the two predicted α-helices compatible with the orientation determined for the helix-turn-helix of the CRO-operator DNA complex [39]. Thus the helical structures described above would fit into the major groove of DNA [40].

Homeobox domains also contain helix-turn-helix structures which share some apparent similarities with the procaryotic repressors [41]. It would be interesting if these predicted binuclear Zn-clusters were just another way of bringing about the classic DNA-binding motif shared by a number of prokaryotic and eukaryotic regulatory products (reviewed in [40,42]). The unusually asymmetric domain of the alcR gene product could be considered to bridge both types of structures.

Acknowledgement: This work was supported by the CNRS.

REFERENCES

[1] Evans, R.M. and Hollenberg, S.M. (1988) Cell 52, 1-3.
[2] Berg, J.M. (1990) J. Biol. Chem. 265, 6513-6516.
[3] Johnston, M. (1987) Microbiol. Rev. 51, 458-476.
[4] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609-1614.
[5] Archer, T.K., Hager, G.L. and Omichinski, J.G. (1990) Proc. Natl. Acad. Sci. USA 87, 7560-7564.
[6] Diakun, G.P., Faitall, L. and Klug, A. (1986) Nature 324, 698-699.
[7] Frankel, A.D., Berg, J.M. and Pabo, C.O. (1987) Proc. Natl. Acad. Sci. USA 84, 4841-4845.
[8] Parraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T. and Klevit, R.E. (1988) Science 241, 1489-1492.
[9] Pan, T. and Coleman, J.E. (1990) Biochemistry 29, 3023-3029.
[10] Povey, J.F., Diakun, G.P., Garner, C.D., Wilson, S.P. and Laue, E.D. (1990) FEBS Lett. 266, 142-146.
[11] Halvorsen, Y.-D.C., Nandabalan, K. and Dickson, R.C. (1990) J. Biol. Chem. 265 22, 13283-13289.
[12] Pateman, J.H., Doy, C.H., Olson, J.E., Norris, U., Creaser, E.H. and Hynes, M. (1983) Proc. R. Soc. London B217, 243-264.
[13] Sealy-Lewis, H.M. and Lockington, R.A. (1984) Curr. Genet. 8, 253-259.
[14] Lockington, R.A., Scazzocchio, C., Sequeval, D., Mathieu, M. and Felenbok, B. (1987) Mol. Microbiol. 1, 275-281.
[15] Felenbok, B. (1991) J. Biotechnol., 17, 11-18.
[16] Felenbok, B., Sequeval, D., Mathieu, M., Sibley, S., Gwynne, D.I. and Davies, R.W. (1988) Gene 73, 385-396.
[17] Felenbok, B., Sophianopoulou, V., Mathieu, M., Sequeval, D., Kulmburg, P., Diallinas, G. and Scazzocchio, C. (1989) in: The molecular genetics of filamentous fungi, Foundations for Biotechnical and Industrial Fermentation Research Vol. 6 (Nevalein, H. ed), pp. 73-83.
[18] Gwynne, D.I., Buxton, F.P., Sibley, S., Davies, R.W., Lockington, R.A., Scazzocchio, C. and Sealy-Lewis, H.M. (1987) Gene 51, 205-216.
[19] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487-491.
[20] Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987) FEBS Lett. 224, 149-155.
[21] Cherfils, J., Vaney, M.C., Morize, I., Surcouf, E., Colloc'h, N. and Mornon, J.P. (1988) J. Mol. Graph. 6, 155-160.
[22] Jones, T.A. (1982) in: Computational Crystallography (Sayre Ed) Clarendon, Oxford, UK.

[23] Bernstein, F.C., Koetze, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, J. (1977) J. Mol. Biol. 112, 535–542.
[24] Brünger, A.T. (1988) 'Xplor V.S. 1.5' software Dep. of Molecular Biophysic and Biochemistry, Yale University, New Haven, CT. 06511, USA.
[25] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. (1983) J. Comp. Chem. 4, 187–217.
[26] Andrianopoulos, A. and Hynes, M.J. (1990) Mol. Cell. Biol. 10, 3194–3203.
[27] Beri, R.K., Whittington, H., Roberts, C.F. and Hawkins, A.R. (1987) Nucleic Acids Res. 15, 7991–8001.
[28] Baum, J.A., Geever, R. and Giles, N.H. (1987) Mol. Cell. Biol. 7, 1256–1266.
[29] Salmeron, J.M. and Johnston, S.A. (1986) Nucleic Acids Res. 14, 7767–7781.
[30] Kammerer, B., Guyonvarch, A. and Hubert, J.C. (1984) J. Mol. Biol. 180, 239–250.
[31] Balzi, E., Chen, W., Vlaszewski, S., Capieaux, E. and Goffeau, A. (1987) J. Biol. Chem. 262, 16871–16879.
[32] Messenguy, F., Dubois, E. and Descamps, F. (1986) Eur. J. Biochem. 157, 77–81.
[33] Friden, P. and Schimmel, P. (1987) Mol. Cell. Biol. 7, 2708–2717.
[34] Creusot, F., Verdière, J., Gaisne, M. and Slonimski, P. (1988) J. Mol. Biol. 204, 263–276.
[35] Berg, J.M. (1988) Proc. Natl. Acad. Sci. USA 85, 99–102.
[36] Klevit, R.E., Herriot, J.R. and Horvath, S.J. (1990) Proteins 7, 215–226.
[37] Lee, M.S. (1989) Science 245, 635–637.
[38] Hard, T., Kellenbach, E., Boelens, R., Maler, B.A., Dahlman, K., Freedman, L.P., Carlstedt-Duke, J., Yamamoto, K.R., Gustafsson, J.-A. and Kaptein, R. (1990) Science 249, 157–160.
[39] Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) Nature 290, 754–758.
[40] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–321.
[41] Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J. and Wüthrich, K. (1989) Cell 59, 573–580.
[42] Johnson, P.F. and McKnight, S.L. (1989) Annu. Rev. Biochem. 58, 799–839.