The EMBO Journal Vol.16 No.15 pp.4689–4697, 1997

Crystal structure of PHO4 bHLH domain–DNA complex: flanking base recognition

Toshiyuki Shimizu, Atsuki Toumoto, Kentaro Ihara, Masato Shimizu[1], Yoshimasa Kyogoku[2], Nobuo Ogawa[3], Yasuji Oshima[3] and Toshio Hakoshima[4]

Department of Molecular Biology, Nara Institute of Science and Technology (NAIST), 8916-5 Takayama, Ikoma, Nara 630-01, [1]Biomolecular Engineering Research Institute, Furuedai, Suita, Osaka 565, [2]Institute for Protein Research, Osaka University, Yamadaoka, Suita, Osaka 565 and [3]Faculty of Engineering, Osaka University, Yamadaoka, Suita, Osaka 565, Japan

[4]Corresponding author
e-mail: hakosima@bs.aist-nara.ac.jp

The crystal structure of a DNA-binding domain of PHO4 complexed with DNA at 2.8 Å resolution revealed that the domain folds into a basic–helix–loop–helix (bHLH) motif with a long but compact loop that contains a short α-helical segment. This helical structure positions a tryptophan residue into an aromatic cluster so as to make the loop compact. PHO4 binds to DNA as a homodimer with direct reading of both the core E-box sequence CACGTG and its 3′-flanking bases. The 3′-flanking bases GG are recognized by Arg2 and His5. The residues involved in the E-box recognition are His5, Glu9 and Arg13, as already reported for bHLH/Zip proteins MAX and USF, and are different from those recognized by bHLH proteins MyoD and E47, although PHO4 is a bHLH protein.
Keywords: crystal structure/E-box recognition/flanking base recognition/helix–loop–helix motif/PHO4

Introduction

Transcription of a set of the genes relevant to the phosphatase (PHO) system in the yeast *Saccharomyces cerevisiae* is regulated by intracellular levels of the essential nutrient phosphate (Oshima, 1991; Johnston and Carlson, 1992 for review). Several genes have been identified in the PHO system. Those under regulation of phosphate are *PHO5* (encoding p60, a major fraction of repressible acid phosphatase), *PHO8* (repressible alkaline phosphatase), *PHO84* (phosphate transporter), *PHO10* and *PHO11*. The transcription of these genes is controlled by a system composed of at least five gene products: PHO4, PHO80, PHO81, PHO85 and PHO2. PHO4 is one of the regulatory proteins indispensable for transcription of the *PHO5*, *PHO81* and *PHO84* genes. Transcription of *PHO5* requires (in addition to PHO4) the homeodomain protein PHO2 (Bürglin, 1988). In phosphate-rich medium, phosphorylation of PHO4 by a complex between the cyclin PHO80 and the cyclin-dependent kinase (CDK) PHO85 (Kaffman *et al.*, 1994) causes the accumulation of PHO4 predominantly in cytoplasm, which results in inhibition of *PHO5* transcription (O'Neill *et al.*, 1996). The kinase activity of the PHO80–PHO85 complex is down-regulated by the CDK inhibitor PHO81 in the starvation of phosphate (Schneider *et al.*, 1994).

The PHO4 protein consists of 312 amino acid residues (Yoshida *et al.*, 1989) and has four functional domains (Ogawa and Oshima, 1990). The carboxy-terminal region is a DNA binding domain composed of a basic (b) region, followed successively by a helix–loop–helix (HLH) motif that was first identified in an immunoglobulin enhancer binding factor and then in other regulatory proteins (Murre *et al.*, 1989). The bHLH motif occurs in a wide range of diverse regulatory proteins found in eukaryotes from yeast to human. In a class of regulatory proteins, the bHLH motif is followed by a Leu zipper (Zip) motif at the carboxy-terminus. These bHLH and bHLH/Zip motifs mediate dimerization that results in the formation of both heterodimers and homodimers. The consensus DNA sequence targeted by several bHLH proteins is reported to be a 5′-CANNTG-3′ element (Blackwell and Weintraub, 1990) known as the E-box (Baxevanis and Vinson, 1993 for review), in which the central two base pairs are specified by each protein. PHO4 can bind as a homodimer to its upstream activation sites (UASs) containing an E-box motif in the promoter regions of *PHO5* (Ogawa *et al.*, 1994; Vogel *et al.*, 1989), *PHO8* (Barbaric *et al.*, 1992; Hayashi and Oshima, 1991) and *PHO81* (Ogawa *et al.*, 1993) genes via the bHLH motif (Ogawa and Oshima, 1990). The recognition sequences by PHO4 can be classified into two types: CACGTG (type 1) and CACGTT (type 2). The type 1 sequence is suggested to be more efficient than the type 2 for both PHO4 binding (Fisher *et al.*, 1991; Fisher and Goding, 1992) and PHO8 expression (Hayashi and Oshima, 1991). Recently, Fisher and Goding (1992) showed that bases outside the E-box were also involved in determining the specificity of binding. Moreover, Ogawa *et al.* (1994, 1995) determined the UAS sequences of several genes and proposed consensus sequences of PHO4 binding sites containing their flanking bases, GCACGTGGG for type 1 and GCACGTTTT for type 2. Hence, important questions remain in regard to the bHLH proteins, although some three-dimensional structures of bHLH and bHLH/Zip domain–DNA complexes have provided valuable information (Ferré-D'Amaré *et al.*, 1993, 1994; Ellenberger *et al.*, 1994; Ma *et al.*, 1994). It is important to elucidate the subtle differences in binding site preferences, as well as to refine our insights into specific protein–DNA interactions. It is of particular interest whether or not the recognition of bases on the flanking side of the E-box occurs in the PHO4 protein, as proposed by Ogawa *et al.* (1994). Here, we have determined the three-dimensional structure of a complex of the bHLH domain of the PHO4 protein with DNA by X-ray crystallography.

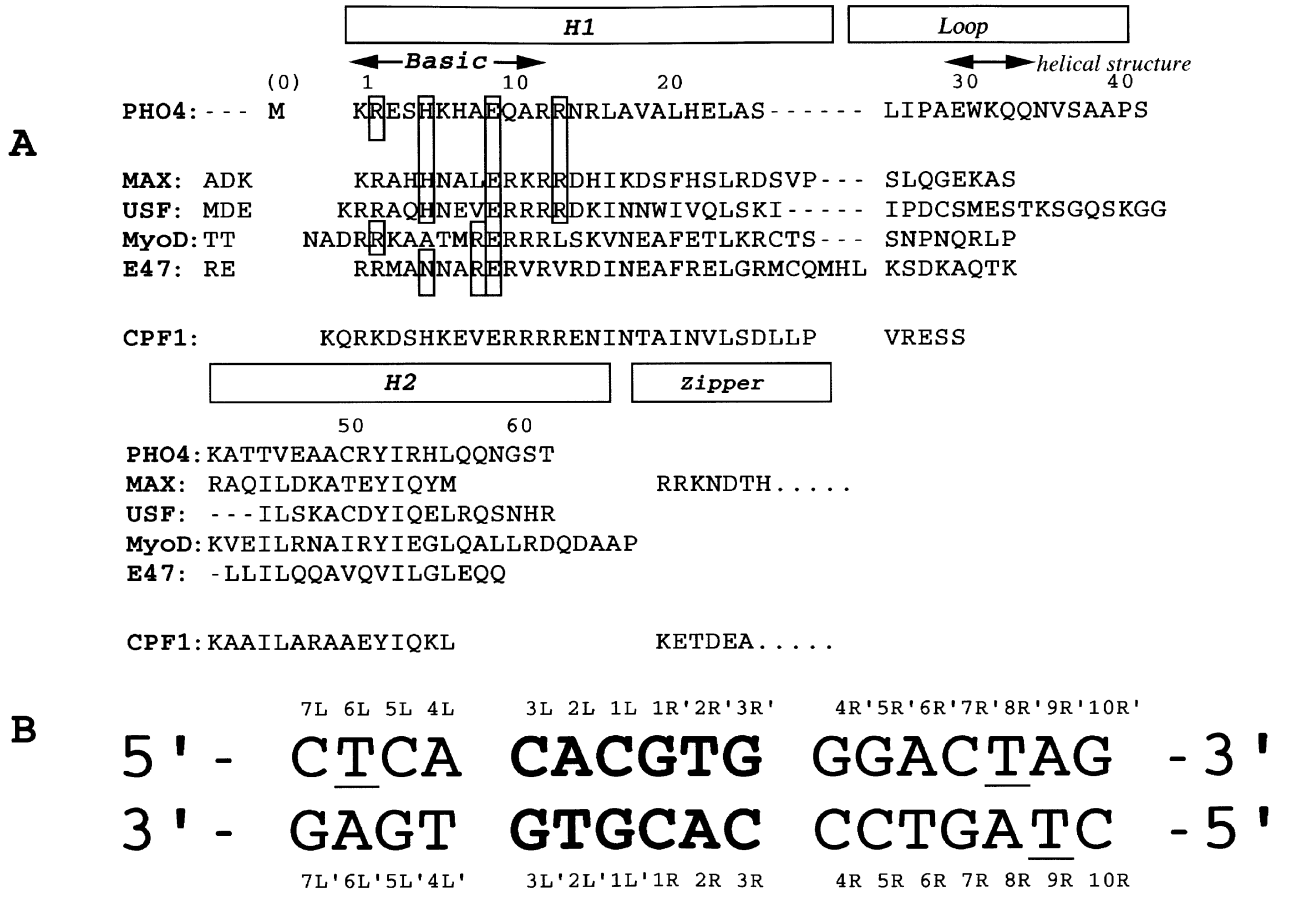

Fig. 1. Sequences of PHO4(63) protein and UASp2(17) DNA. (**A**) A sequence alignment of bHLH, MyoD and E47, and of bHLH/Zip, MAX and USF, the structures of which have been reported. The crystal structure of USF lacks the Leu zipper region. The sequence of Cpf1 also aligned. Amino acid residues in one-letter codes are numbered according to Fisher *et al.* (1991). Secondary structures of PHO4 are determined by PROCHECK. Residues recognizing the bases are enclosed by solid lines. (**B**) Sequence of the duplex oligonucleotide of UASp2(17) DNA. The core E-box element is in bold. The thymine bases that were replaced by 5-iodouracil for isomorphous derivatives are underlined. The numbering scheme is that in Ferré-D'Amaré *et al.* (1993).

Results

Overall structure of the PHO4–DNA complex

The crystal structure of a complex that contains two bHLH domains of PHO4(63), which consists of 126 residues as a dimer, and a double-stranded DNA fragment of 17 bp was determined by the multiple isomorphous replacement (MIR) method and refined at 2.8 Å resolution. The amino acid sequence of PHO4 is compared in Figure 1A with those of other bHLH and bHLH/Zip proteins, the three-dimensional structures of which were previously reported. Each peptide chain of the PHO4 dimer is designated A and B. The 17 bp oligomer used for the present study, designated UASp2(17), was derived from the second site of the UASs of the *PHO5* gene, UASp2, since the site has one of the highest affinities with PHO4 (Ogawa *et al.*, 1994). The core sequence of UASp2(17) contains the symmetrical E-box, CACGTG (Figure 1B). The numbering systems of amino acid residues and base pairs of DNA are as described in Fisher *et al.* (1991) and Ferré-D'Amaré *et al.* (1993), respectively.

The DNA-binding domain of PHO4 consists of two helices, designated H1 and H2, separated by a long loop that contains a novel α-helical region. PHO4 binds to DNA as a homodimer and the two monomers fold into a parallel, left-handed four-helix bundle (Figure 2). The

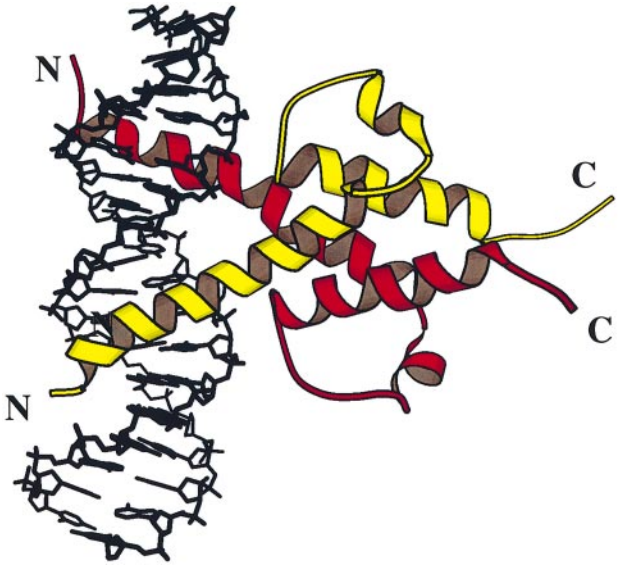

Fig. 2. Overview of the PHO4–DNA complex drawn with the program MOLSCRIPT (Kraulis, 1991). Helical regions are represented by ribbons; non-regular secondary structure elements by thin tubes. Molecules A and B are colored red and yellow, respectively. Helical structure is clearly seen in the loop region.

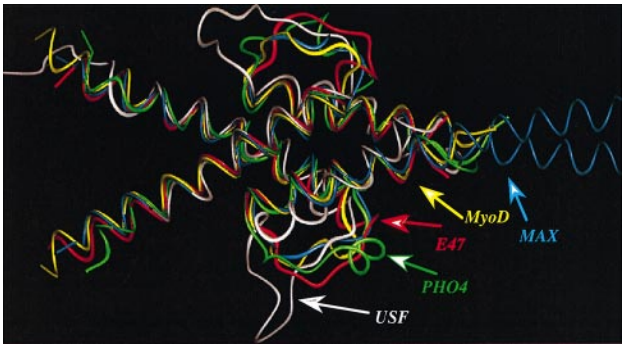

Fig. 3. Three-dimensional comparison of the PHO4 (green) and the other bHLH (E47, red; MyoD, yellow) and bHLH/Zip (MAX, cyan; USF, white). MAX has a zipper region.

bundle topology is identical to the structures of the other bHLH/Zip proteins, such as Max and USF (Ferré-D'Amaré *et al.*, 1993, 1994, respectively), and to those of bHLH proteins such as MyoD (Ma *et al.*, 1994) and E47 (Ellenberger *et al.*, 1994). Superposition of these structures yields relatively small root-mean-square (r.m.s.) deviations in a range of 1.3–1.4 Å for α-carbon atoms (68 residues from 5 to 23 and from 42 to 56), excluding the loop region and both terminal segments (Figure 3). Six amino-terminal residues from positions 3 to 8 of helix H1 in the A-chain are slightly extended and form a 3_{10} helix, which lies in a different crystalline environment from that of the amino-terminal region of the B-chain: Ser4 of the B-chain contacts with Gln34 of the A-chain of the symmetry-related molecules. However, the N-terminal regions around Arg2 of both A- and B-chains have no contact with any symmetry-related molecule. Therefore, the contact around Ser4 seems to have no interference with the DNA recognition by Arg2 as described below. The different crystalline environment around Gln34, which locates at the loop, causes a slightly different conformation of the flexible loop of the A-chain from that of the B-chain. The carboxy-terminal region around 60–62 is also in the different environment. The PHO4 dimer is stabilized by van der Waals interactions between helices H1 and H2. About three-quarters of the hydrophobic residues participate in formation of the hydrophobic core. The hydrophobic core of the four-helix bundle closely resembles those of other bHLH and bHLH/Zip proteins. The buried surface area of the PHO4 bHLH dimer, 1739 $Å^2$ calculated with a 1.4 Å radius probe using the methods described by Connolly (1983), is nearly the same as those of MAX, MyoD and E47 with a few percent deviation.

DNA recognition

Each half-site of the symmetrical E-box is recognized directly by His5, Glu9 and Arg13 (Figures 4 and 5A–F). The interactions are essentially symmetrical, although the amino-terminal region of the A-chain (positions 0–8) is slightly extended, as described above. Glu9, which is absolutely conserved among all the bHLH and bHLH/Zip proteins, makes a bidentate contact with the conserved CA base step of each half-site [Figure 4, A(2L); A(2R); C(3L); C(3R)]. This result is consistent with the site-directed mutagenesis experiment in which substitution of Gln, Asp or Leu into this position abolished DNA binding

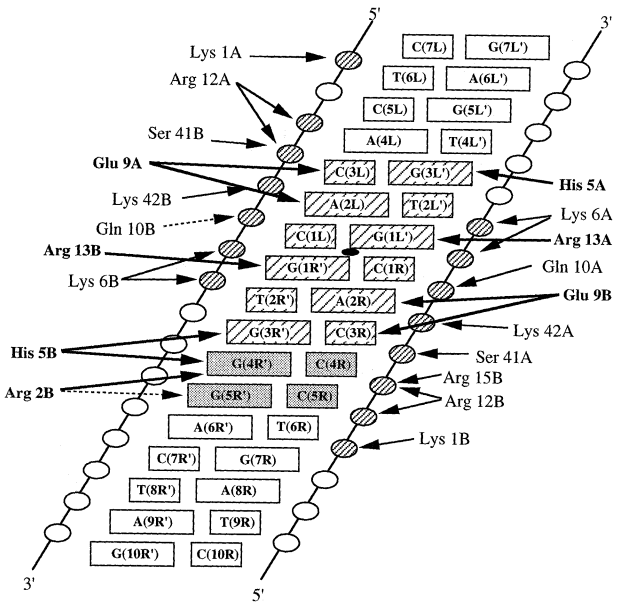

Fig. 4. Summary of contacts of PHO4 residues with DNA bases and phosphate groups. Schematic summary of the base and phosphate contacts made by each monomer. The DNA is represented as a cylindrical projection with phosphates indicated by circles. The E-box bases are stippled and recognized flanking bases are hatched. Base pair recognitions are indicated by bold-lined arrows, and phosphate recognitions by thin-lined arrows. The weak interaction is shown by dashed-lined arrows. All contacts are via side chains.

by PHO4 protein (Fisher and Goding, 1992). Similar contacts were also observed in the crystals of MAX, MyoD and E47 complexed with DNA.

It has been shown by mutation analysis (Dang *et al.*, 1992) that an Arg residue at position 13 confers specificity for CACGTG (class B) versus CAGCTG (class A) E-boxes. Interestingly, the PHO4 protein, like the bHLH/Zip proteins, has an Arg residue at this position, despite its lacking a Leu-zipper. The other bHLH proteins have a hydrophobic residue (Figure 1A). This Arg13 of PHO4 makes a direct, symmetrical contact with the central G(1L′/1R′) of the E-box (Figure 5C and D), in a manner similar to that of bHLH/Zip proteins, MAX and USF (Ferré-D'Amaré *et al.*, 1993, 1994, respectively). For these central 2 bp of the E-box, water-mediated, not direct, contact was observed in MyoD (Ma *et al.*, 1994), and the asymmetrical contact was pointed out in E47 (Ellenberger *et al.*, 1994).

The recognition of bases by His5 of each chain of PHO4 is asymmetric. His5 in the A-chain contacts with G(3L′) in the E-box, but, in the B-chain, forms bifurcated hydrogen bonds with G(3R′) in the E-box and G(4R′) flanking the 3′ end of the E-box (Figures 5F and G and 6A). The stereochemical parameters of the bifurcated hydrogen bonds are acceptable when compared with those already reported (Taylor *et al.*, 1984; Preißner *et al.*, 1991) (Figure 6B). Most bHLH and bHLH/Zip proteins have a His or Asn residue at this position (Figure 1A). Each of the His5 residues of MAX and USF makes a weak contact with the corresponding guanine base because the contact distance is relatively long for a hydrogen bond (3.8–3.9 Å). In E47, the corresponding residue is Asn, and it contacts the guanine base in one monomer, though the residue in the other monomer makes no base contact. In

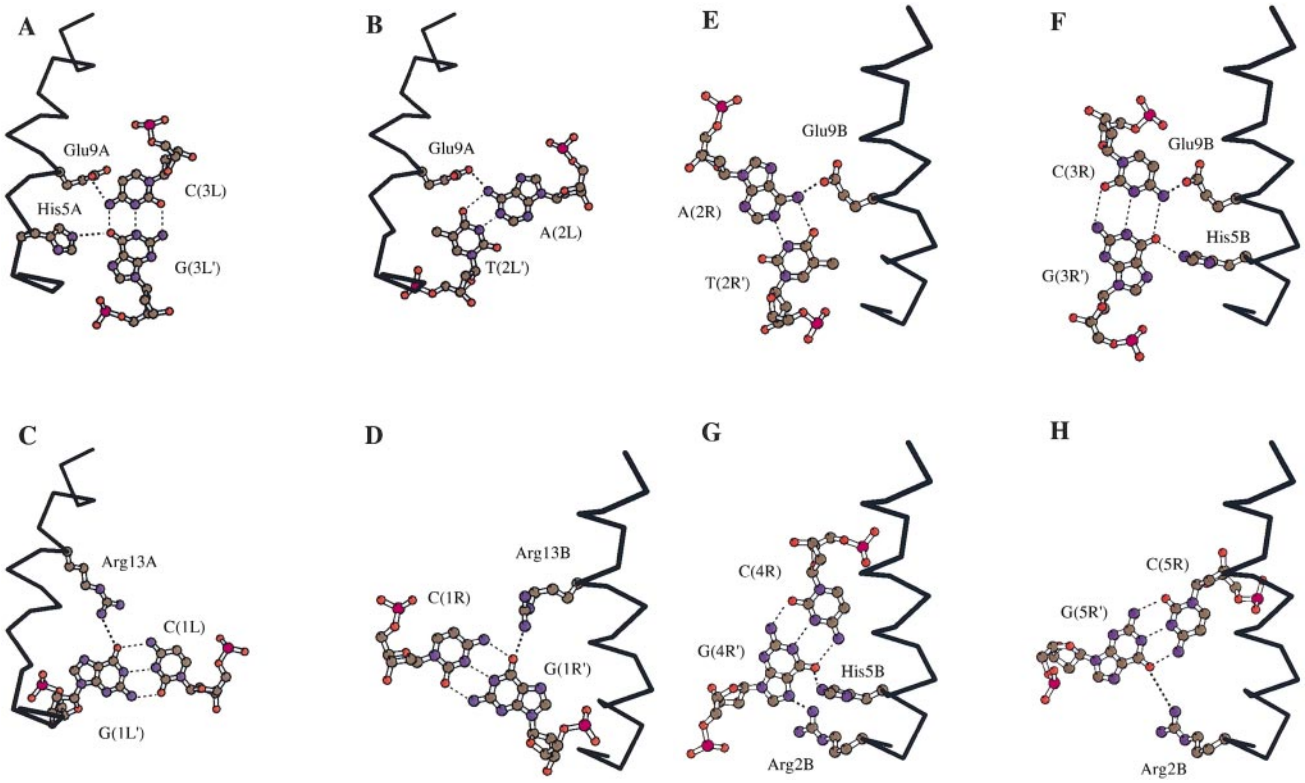

Fig. 5. The base-specific interactions. The views are down the overall DNA helix axis and hydrogen bonds (dotted) are indicated.

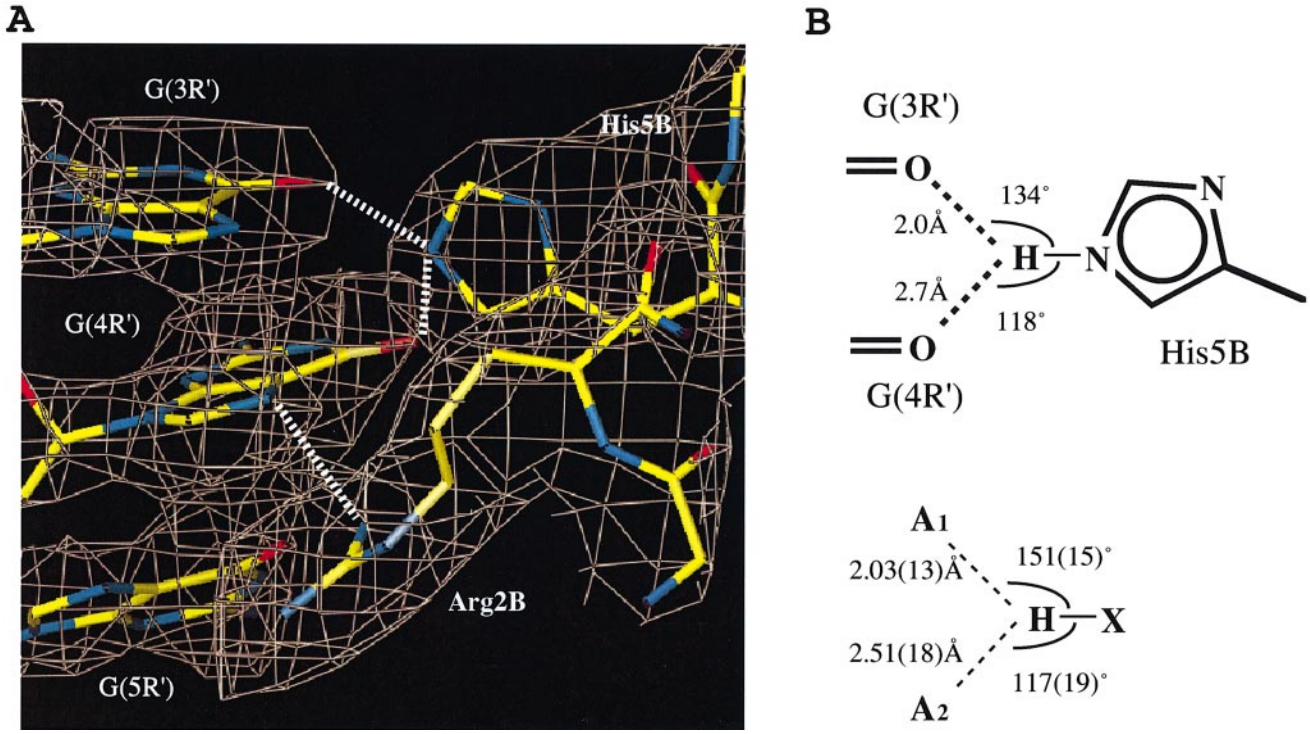

Fig. 6. Recognition of the 3′ flanking bases by His5B and Arg2B. (**A**) $2F_o$-F_c electron density map around G(3R′). Electron density is contoured at 1σ above the mean. Dashed lines indicate hydrogen bonds. His5B forms bifurcated hydrogen bonds with G(3R′) and G(4R′). Arg2B contacts with G(4R′) and also makes a weak contact with G(5R′). (**B**) Geometry of the bifurcated hydrogen bond found in this study (upper). The mean values of the geometry of the bifurcated hydrogen bonds observed in crystals (Preißner *et al.*, 1991) are shown (lower). Two acceptors A1 and A2 oppose the donor X-H.

MyoD, the corresponding residue is an Ala, and it is buried in the major groove without any direct interaction with DNA. These structural features suggest that base recognition by residues at this position is variable in comparison with the other conserved residues. Moreover, Arg2 of the B-chain also makes a direct contact with G(4R′) and a weak contact (~4 Å) with G(5R′) (Figures 5G and H and 6A).

A number of residues involving Lys1, Lys6, Gln10, Arg12 and Arg15 in the basic region make contacts with phosphate groups. In addition, Ser41 in the loop region and Lys42 at the start of helix H2 also make contact with the phosphate groups. These features are consistent with the observation that PHO4(53), lacking most of the basic region, is unable to bind to DNA (Shimizu, 1995).

Loop structure

The bHLH proteins have loops connecting helices H1 and H2 in various length, sequence and amino acid compositions. Although the four-helix bundles are very similar, there are significant differences in the loop structures of bHLH and bHLH/Zip proteins (Figure 3). PHO4 has a long but compact loop that forms a helical structure (Figure 1A). This short stretch of α-helix has never been found in any crystal structure of bHLH or bHLH/Zip proteins, but coincides with the observation of free PHO4 structure in solution by NMR (Shimizu, 1995). This indicates that the short α-helix is not induced in crystal or on DNA binding, but rather is inherent. The PHO4 loop contains a Trp residue that is positioned by this helical structure such that it faces the other aromatic rings of Tyr52 and His55 of helix H2 and Pro28 within each monomer. As previously discussed with regard to MAX structure (Ferré-D'Amaré *et al.*, 1993; Ellenberger *et al.*, 1994), a proline residue at the carboxy-terminal end of helix H1 packs against the tyrosine corresponding to Tyr52 of PHO4, but the notable cap structure of the aromatic cluster is observed in the PHO4 protein.

In the loop region, PHO4 protein lacks an inner hydrogen network corresponding to those observed in both E47 (Ellenberger *et al.*, 1994) and MAX (Ferré-D'Amaré *et al.*, 1993). The loop of E47 is stabilized partly by the hydrogen bond network of the Gln triad, one Gln residue of which belongs to the loop. The PHO4 loop makes its sole contact with a phosphate group by Ser41. Contrastingly, USF has a long, extended loop that traverses the adjacent minor groove and contacts with phosphate groups and sugar moieties within the minor groove. Generally, the loops of both bHLH and bHLH/Zip motifs are not always functionally interchangeable in a swap experiment (Pesce and Benezra, 1993). The conformation of the loop region of MAX is stabilized by some interactions within the protein and between protein and DNA, while the MyoD loop, the length of which is almost similar to MAX, exhibits a different conformation.

DNA structure

UASp2(17) DNA has an essentially B-DNA form, as observed in the other complexes of bHLH and bHLH/Zip proteins. The average helical twist is 33.4° and the mean rise per base pair is 3.30 Å, implying 10.77 bp per turn. This result is in good agreement with the finding of Shimizu (1995) that the circular permutation mobility assay showed a weak tendency for DNA bending. Although the overall structure of UASp2(17) resembles that of B-DNA form, the r.m.s. deviations of UASp2(17) in the protein interaction region [C(6L)-C(7R′)] from the canonical B-DNA are relatively large: 2.61 Å for all atoms, 1.78 Å for bases and 3.15 Å for sugar–phosphate backbone atoms. As reported in MyoD (Ma *et al.*, 1994), the major groove is rather narrow (10.3 Å in PHO4 versus 12.3 Å in B-DNA), and the minor groove is rather wide (8.3 Å in PHO4 versus 4.8 Å in B-DNA). The narrowed major groove and widened minor groove are observed in all DNA oligomers complexed with bHLH and bHLH/Zip proteins. The DNA structure is stabilized by stacking interactions of base pairs with symmetry-related DNA molecules.

Discussion

The present structural studies of the PHO4 bHLH–DNA complex have revealed the anticipated flanking-base recognition. Ogawa *et al.* (1994, 1995) pointed out that GG bases flanking the 3′ end of the PHO4 E-box are almost conserved in the type 1 UAS in the PHO regulon. In the present crystal, the first G base flanking the 3′ end of the E-box was found to be recognized through a hydrogen bond to His5. In addition, Arg2 recognized the second G base flanking the 3′ end, G(5R′), by a weak contact. This is the first report of these interactions with the 3′-flanking bases, while an Arg residue at the other position (position 8) in MyoD and E47 recognized the 5′-flanking base (Figure 1A). Interestingly, an Arg residue at position 2 in MyoD is known to make a contact with an E-box base (Ma *et al.*, 1994).

It would be interesting to see if the unique loop of PHO4 influences specific dimer formation. The folding of each chain in the dimer is stabilized by the hydrophobic core formed by the loop described above. Since the interactions involve the residues from helix H2 and Pro28 which is adjacent to helix H1, the loop structure seems to influence the mutual orientation of helices H1 and H2. The angle between the helical axes of helices H1 and H2 of PHO4 is 47°, which is different from the corresponding angles of MyoD (42°), MAX (41°) and USF (75°). These differences may influence specific dimer formation, though no loop residue locates at the interface between the monomers. It is notable that Tyr52 of helix H2 forms a hydrogen bond with Gln57 of helix H2 of the other monomer. These two residues are conserved in MyoD and a similar interhelical hydrogen bond was observed. In E47, the corresponding residues are Val and Glu, respectively, and there is no interhelical hydrogen bond between H2 helices. Alternatively, this Glu residue forms a hydrogen bond with a His residue located at the C-terminus of helix H1 of the other monomer. It is notable that there is no interhelical hydrogen bond between two monomers of MAX and USF. These differences may also contribute to specify the dimerization partner.

It is of considerable interest that the homodimer of PHO4 exhibits asymmetry in the DNA recognition sequence. The DNase I footprinting experiments of several PHO4-binding sites showed that the sequences protected by PHO4 expand up to more than 20 bp long with semi-invariant 3′-flanking bases rather than the 5′-flanking

bases (Ogawa *et al.*, 1994). The asymmetry of the protected region was also observed by the bHLH domain (Shimizu, 1995). The present structure exhibits several features of its asymmetric binding. The angle between the helical axes of DNA and helix H1 is considerably different between the A- and B-chains (9°). In contrast, the differences of the corresponding angles for MyoD, E47, MAX and USF are small (0–2°). The angle for helix H1 of the B-chain (124°), which recognized the 3′-flanking bases, is greater than that of the A-chain and enables the helix H1 to run along the major groove so as to reach the 3′-flanking bases. If compared with helix H1 of the B-chain, the angle for the A-chain is relatively close to a right angle (115°) and the helix H1 seems to run across the major groove rather than along the groove. The buried surface area of the B-chain upon DNA binding is 10% greater than that of the A-chain. These observations indicate that the PHO4–bHLH domain dimerized strongly with two H1 helices spaced so that only one helix H1 would interact properly with the flanking bases. The asymmetry of the binding is stabilized by a hydrogen bond between Arg15B with the phosphate group of C(4R), whereas Arg15A has no hydrogen bond with the corresponding phosphate group. Moreover, Arg12B interacts with two phosphate groups that are shifted toward the 3′ end, if compared with the two phosphate groups interacted with Arg12A. This also contributes to the stabilization of the asymmetric binding.

One of the residues contacting the 3′-flanking base is Arg2, which is almost conserved. However, the corresponding residues of Arg2 in MAX, USF and E47 have no contact with any DNA base. Alternatively, the corresponding residues of MyoD contact with G of the E-box. This observation indicates that the conserved residues of each protein could make contact with different bases. PHO4 is a bHLH protein but His5 is conserved in bHLH/Zip protein but not in other bHLH proteins, including MyoD and E47. Compared with two known bHLH/Zip proteins, it should be pointed out that the mutual orientations of the two helices H1 in these proteins are different. The interhelical angle of the two H1 helices of PHO4 is 61°, which is larger than that of MAX (55°), but smaller than that of USF (67°). These differences in the geometry could cause different contacts of the conserved residues.

In the UASs of both types 1 and 2 in the PHO regulon, a G base flanking the 5′ end of the PHO4 E-box motif is frequently observed. However, PHO4 exhibits a high affinity to *PHO5* UASp2, which has A in this position. Moreover, the binding activity is retained when the G at this position is replaced with A or C in *PHO84* UAS site D (Ogawa *et al.*, 1995). In the crystal, Arg2 of the A-chain seems to be pushed away from the DNA by a contact with the methyl group of T(4L′). This unfavorable contact can be avoided when T(4L′) is replaced with C, which is paired with the G flanking the 5′ end of the PHO4 E-box. Given GG at the 3′ end, the additional interactions with the 3′-flanking bases could result in loose contacts of the helix H1 of the A-chain with the 5′-flanking bases.

Fisher and Goding (1992) showed that the presence of a T base flanking the 5′ end of the CACGTG motif inhibits its binding to PHO4, but not to Cpf1 protein, which is involved in both centromere function and methionine

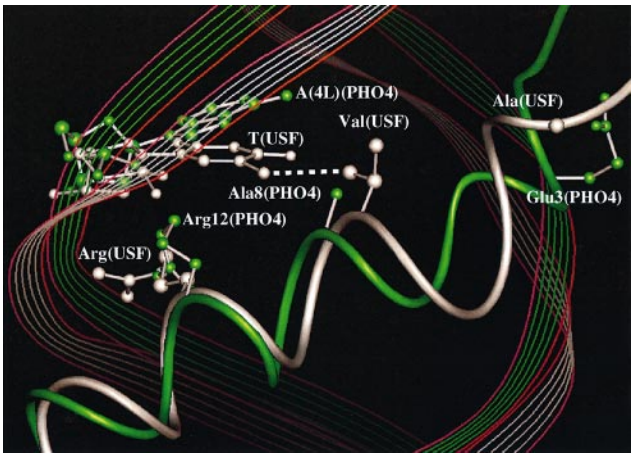

Fig. 7. Surroundings of the 5′-flanking base in PHO4 (green) and USF (white). A van der Waals contact shown by a dashed line was observed between the methyl group of the flanking thymine (at position 4L) and Val8 of USF. Glu3 (PHO4) and Ala3 (USF) are far from the 5′-flanking base in PHO4 and USF.

biosynthesis, and which also has a bHLH structure recognizing RTCACRTG (where R is a purine residue) (Baker and Masison, 1990; Cai and Davis, 1990; Mellor *et al.*, 1990). This difference may be caused by a possible contact between the methyl group of the T base and the proteins. A preliminary model-building study showed that the methyl group of T at position 4L would interfere with the side chain conformation of Arg12, which contacts nearby phosphate groups (Figure 7). Since Cpf1 also has an Arg residue at the position corresponding to Arg12 of PHO4, two Args of PHO4 and Cpf1 may have different conformations, though discussion of the conformational differences of the residues must wait until we have the three-dimensional structure of Cpf1 complexed with DNA. Val8 of Cpf1 protein may have a van der Waals contact with the methyl group of T at the 4L position. In fact, a van der Waals contact was observed between a flanking T base and Val8 in USF (Ferré-D'Amaré *et al.*, 1994), which can bind to DNA in the presence of T flanking 5′ end (Bendall and Molloy, 1994) (Figure 7). In contrast, PHO4 has an Ala residue at this position and could make no contact with the methyl group of the T base.

Substitution of Glu3 with an aspartic acid residue completely prevented the inhibition by the flanking T base (Fisher and Goding, 1992). However, Glu3 is far away from the T base and makes no contact with it in the crystal (Figure 7). Moreover, Ala residues at position 3 of MAX and USF are also far away from the flanking T base (Ferré-D'Amaré *et al.*, 1993, 1994) (Figure 8). It is therefore unlikely that this residue is involved in the interaction of the flanking T base unless a conformational change occurs. As the PHO4 protein used in the binding assay experiment (Fisher and Goding, 1992) is ~30 residues longer than that in the present study, it is possible that the conformation around Glu3 is different. A further structural study will be required.

The PHO4 binding preference for CACGTG over CACGTT could arise from possible distortions of the interactions observed in the current study. In particular, the replacement of G with T at the 3′-end position, i.e. the replacement of C with A at position 3R, could

Table I. Statistics of structure determination and refinement

Structure determination statistics					
Data set	Native1	Native2	IdU(6L,8R′)	IdU(8R′)	IdU(6L,8R′,9R)
Instrument	PF (resolution 2.8 Å)	*R*-axisIIc (resolution 3.0 Å)			
Unique/measured	9398/36943	6156/15397	5788/14516	4899/10515	5165/12579
Completeness (%)	92.0/76.8[a]	70.9	65.7	56.1	58.6
R_{merge} (%)	8.7/20.2[a]	7.22	6.32	10.74	8.02
R_{iso} (%)			18.2	19.8	20.7
Phasing power (12.0–3.5 Å)			1.35	1.02	1.62
R_{Cullis} (12.0–3.5 Å)			0.73	0.79	0.66
Mean figure of merit (∞–3.5 Å)		0.40			
Refinement statistics: 8.0–2.8 Å					
Data with $F>3\sigma F$					
No. of reflections		7581			
R factor (%)		23.0			
Free *R* factor (%)		28.2			
No. of non hydrogen atoms		1687			
No. of solvent molecules		80			
R.m.s. bond length (Å)		0.013			
R.m.s. bond angle		2.28°			
R.m.s. improper angle		1.89°			

$R_{merge} = \Sigma\Sigma|I_{h,j}-\langle I_h\rangle|/\Sigma\Sigma\langle I_h\rangle$; $R_{iso} = \Sigma||F_{PH}|-|F_P||/\Sigma|F_P|$; $R_{cullis} = \Sigma||F_{PH} \pm F_P|-F_{H(CALC)}|/\Sigma|F_{PH}-F_P|$ (for centric reflections);
$I_{h,j}$ = measured diffraction intensity; $\langle I_h\rangle$ = mean value of all intensity measurements of (h,k,l) reflection. Phasing power: the ratio of the root mean square (r.m.s.) heavy atom scattering factor amplitude to the r.m.s. lack of closure error.
[a]For the highest resolution shell (2.86–2.8 Å).

have abolished the interaction between C(3R) and Glu9 (Figure 5A).

NMR studies of PHO4(63) in a DNA-free condition (Shimizu, 1995) indicated that there is no α-helical structure in the amino-terminal region of helix H1 (basic region from Met0 to Arg13). This is a case of induced fit, as already suggested for both USF (Ferré-D'Amaré *et al.*, 1994) and MyoD (Starovasnik *et al.*, 1992). Recent work indicates that the basic region of PHO4 may mask the activation domain (Shao *et al.*, 1996). The flexible structure might be appropriate for this masking.

Materials and methods

Expression, purification, DNA synthesis and crystallization

The PHO4(63) was overexpressed in *Escherichia coli* BL21 (DE3) (Studier and Moffatt, 1986) using the T7 RNA polymerase system. The DNA oligomers for co-crystallization and the iodinated DNA were synthesized by the solid-phase phosphotriester method on a DNA synthesizer (model 391, Applied Biosystems). Purification of protein and DNA and co-crystallization were carried out as previously reported (Toumoto *et al.*, 1997). A preliminary gel retardation assay, and DNA binding experiments using surface plasmon resonance measurement with BIAcore (Pharmacia Biosensor, Uppsala, Sweden) and a Beacon Fluorescence Polarization System (PanVera Corp., Madison, WI) have shown that the binding affinity of the truncated form PHO4(63) is almost the same as that of intact PHO4 (unpublished data). The best crystals of PHO4(63)–UASp2(17) complex were grown when the 15 ml drops containing 0.4 mM protein, 0.2 mM DNA, 1% (w/v) PEG6000 and 20 mM Na citrate buffer (pH 3.6) were equilibrated with a 500 ml reservoir solution of 1% (w/v) PEG6K and 20 mM Na citrate buffer (pH 3.6). The present crystal obtained by the truncation of the amino-terminal diffracts better than crystals of UASp2(17)–PHO4(85) complex (Hakoshima *et al.*, 1993), although the solvent content of this crystal is remarkably high (~71%). Isomorphous derivatives of the complex were obtained by using DNA duplexes with 5-iodouracil substituted for thymine at position 8R′, at positions 6L and 8R′, and at positions 6L, 8R′ and 9R, as indicated by the numbering scheme of Figure 1B. Because crystals grow at a low pH, possibly neutralization of carboxyl groups or phosphate groups may produce artificial interactions or eliminate ion pairs. We performed UV-CD spectra measurements of PHO4 with or without DNA in both pH 7.0 and pH 3.6 (our unpublished data). The conformational changes of PHO4 upon DNA binding is well monitored by the changes in the CD spectra. In particular, induced helical structure of the basic region drastically changes the CD spectra. The results obtained indicated that there are no essential differences between pH 7.0 and pH 3.6. Therefore, we believe that the present structure is essentially representative of the PHO4–DNA complex.

Data collection and structure determination

Three-dimensional data sets to 3.0 Å resolution were collected for native and all derivative crystals by an R-Axis IIc (Rigaku, Japan) on a Rigaku RU300H rotating-anode generator operated at 40 kV, 100 mA with CuKα radiation. The crystal belongs to the orthorhombic space group $P2_12_12_1$, with unit cell dimensions a = 53.51 Å, b = 68.30 Å and c = 108.77 Å. Assuming one protein dimer–DNA complex in the asymmetric unit, the packing density (V_m) of the crystal (Matthews, 1968) is evaluated to be 4.2 Da/Å^3, which indicates that the solvent content is quite high (71%). For crystallographic refinement, diffraction data were collected to 2.8 Å, using a Weissenberg camera for macromolecules (Sakabe, 1991) installed on the beam line 18B at the Photon Factory (Tsukuba, Japan). The wavelength was set to 1.00 Å and the diffraction path was filled with helium gas to avoid air scattering. The data were processed by the program WEIS (Higashi, 1989). The iodine sites in the crystals were located by difference Patterson methods. All calculations were performed using programs from the CCP4 program package (Collaborative Computational Project, Number 4, 1994). The positions of iodine atoms were confirmed by difference Fourier analysis and refined using the program MLPHARE (CCP4 program package) at 3.5 Å. Phases for the last MIR map had a mean figure-of-merit of 0.40 for data from ∞–3.5 Å resolution. The phases were improved by the program DM (Cowtan, 1994) with a combination of solvent flattening/histogram matching and were extended to 3.0 Å. This procedure was very effective due to the high solvent content of this crystal, and the map revealed clear electron density for the DNA and the α-helices of PHO4. A two-fold averaging was not applied, but the quality of the MIR map is sufficient to place the model in helix regions.

Model building and structure refinement

A partial model was built into the map using the program O (Jones *et al.*, 1991). The positions of the iodine atoms were used to properly position the DNA model. The partial model was subjected to energy minimization. All crystallographic refinement was carried out using the X-PLOR package, version 3.1 (Brünger, 1992). The phase combined map from the partial structure showed the electron density of loop regions. Conventional energy minimization and simulated annealing with molecular dynamics (Brünger *et al.*, 1987) were performed. In the

simulated annealing step, the slow-cooling protocol was applied (Brünger *et al.*, 1990), starting at 3000 K and continuing to 300 K (time-step, 0.5 fs; decrement of temperature, 25 K; number of steps at each temperature, 50; tolerance, 0.2 Å). Restraints for the base planarity, the sugar puckers and the hydrogen atoms in the oligonucleotide were used in X-PLOR (Parkinson *et al.*, 1996). An anisotropic overall *B*-factor refinement with X-PLOR was performed. After several rounds of refinement and manual rebuilding, the crystallographic *R*-factor of the final model was 23.0% with the data $F > 3\sigma(F)$ and 25.1% with all of the data from 8 Å to 2.8 Å. The corresponding free *R* values are 28.2% and 33.6%, respectively. 80 water molecules were included in the current model and were accepted with the following criteria: the peaks derived from water molecules appeared with strong density ($>4\sigma$) in F_o-F_c map; water molecules with a *B*-factor of no more than 50 Å^2 only were accepted. *B*-factors of 95% water molecules were relatively low (<45 Å^2) and the highest value was 49.2 Å^2 (overall *B*-factor is 26.2 Å^2). These water molecules are in the surroundings of the basic region of PHO4 and the DNA, and hence are of biological interest. The stereochemical quality of the model was monitored using PROCHECK (Laskowski *et al.*, 1993). DNA conformation was analyzed using the program NEWHEL93 (R.Dickerson, personal communication). The structure determination and the refinement statistics are summarized in Table I.

Acknowledgements

We acknowledge Drs S.K.Burley and T.Ellenberger for providing the coordinates of MAX/USF and E47, respectively. We thank N.Sakabe and M.Suzuki for their assistance in data collection and S.Fujii for providing information on DNA conformational analysis programs. This work was supported by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture of Japan to T.H. (06276104, 05244102 and 07458171).

References

Baker,R.E. and Masison,D.C. (1990) Isolation of the gene encoding the *Saccharomyces cerevisiae* centromere-binding protein CP1. *Mol. Cell. Biol.*, **10**, 2458–2467.

Barbaric,S., Fascher,K.-D. and Hörz,W. (1992) Activation of the weakly regulated PHO8 promoter in *S. cerevisiae*: chromatin transition and binding sites for the positive regulatory protein PHO4. *Nucleic Acids Res.*, **20**, 1031–1038.

Baxevanis,A.D. and Vinson,C.R. (1993) Interactions of coiled coils in transcription factors: where is the specificity? *Curr. Opin. Genet. Dev.*, **3**, 278–285.

Bendall,A.J. and Molloy,P.L. (1994) Base preferences for DNA binding by the bHLH-Zip protein USF: effects of $MgCl_2$ on specificity and comparison with binding of Myc family members. *Nucleic Acids Res.*, **22**, 2801–2810.

Blackwell,T.K. and Weintraub,H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science*, **250**, 1104–1110.

Brünger,A.T.(1992) *X-PLOR Manual, Version 3.1: A System for Crystallography and NMR.* Yale University Press, New Haven, CT.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) Crystallographic R factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Brünger,A.T., Krukowski,J. and Erickson,J. (1990) Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr.*, **A46**, 585–593.

Bürglin,T.R. (1988) The yeast regulatory gene PHO2 encodes a homeo box. *Cell*, **53**, 339–340.

Cai,M. and Davis,R.W. (1990) Yeast centromere binding protein CBF1, of the helix–loop–helix protein family, is required for chromosome stability and methionine prototrophy. *Cell*, **61**, 437–446.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Connolly,M.L. (1983) Solvent accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.

Cowtan,K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**, 34–38.

Dang,C.V., Dolde,C., Gillison,M.L. and Kato,G.J. (1992) Discrimination between related DNA sites by a single amino acid residue of Myc-related basic–helix–loop–helix proteins. *Proc. Natl Acad. Sci. USA*, **89**, 599–602.

Ellenberger,T., Fass,D., Arnaud,M. and Harrison,S.C. (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix–loop–helix dimer. *Genes Dev.*, **8**, 970–980.

Ferré-D'Amaré,A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, **363**, 38–45.

Ferré-D'Amaré,A.R., Pognonec,P., Roeder,R.G. and Burley,S.K. (1994) Structure and function of the b/HLH/Z domain of USF. *EMBO J.*, **13**, 180–189.

Fisher,F. and Goding,C.R. (1992) Single amino acid substitutions alter helix–loop–helix protein specificity for bases flanking the core CANNTG motif. *EMBO J.*, **11**, 4103–4109.

Fisher,F., Jayaraman,P.-S. and Goding,C.R. (1991) C-myc and the yeast transcription factor PHO4 share a common CACGTG-binding motif. *Oncogene*, **6**, 1099–1104.

Hakoshima,T., Teranishi,Y., Ohira,T., Suzuki,K., Shimizu,M., Shirakawa,M., Kyogoku,Y., Ogawa,N. and Oshima,Y. (1993) Crystallographic characterization of a PHO4–DNA complex. *J. Mol. Biol.*, **229**, 566–569.

Hayashi,N. and Oshima,Y. (1991) Specific *cis*-acting sequence for PHO8 expression interacts with PHO4 protein, a positive regulatory factor, in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **11**, 785–794.

Higashi,T. (1989) The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Crystallogr.*, **22**, 9–18.

Johnston,M. and Carlson,M. (1992) Regulation of carbon and phosphate utilization. In Jonew,E.W., Pringle,J.R. and Broach,J.R. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp.193–281.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kaffman,A., Herskowitz,I., Tjian,R. and O'Shea,E.K. (1994) Phosphorylation of the transcription factor PHO4 by a cyclin–CDK complex, PHO80–PHO85. *Science*, **263**, 1153–1156.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Ma,P.C.M., Rould,M.A., Weintraub,H. and Pabo,C.O. (1994) Crystal structure of MyoD bHLH domain–DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell*, **77**, 451–459.

Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.

Mellor,J., Jian,W., Funk,M., Rathjen,J., Barnes,C.A., Hinz,T., Hegemann,J.H. and Philippsen,P. (1990) CPF1, a yeast protein which functions in centromeres and promoters. *EMBO J.*, **9**, 4017–4026.

Murre,C., McCaw,P.S. and Baltimore,D. (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD* and *myc* proteins. *Cell*, **56**, 773–783.

Ogawa,N. and Oshima,Y. (1990) Functional domains of a positive regulatory protein, PHO4, for transcriptional control of the phosphatase regulon in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **10**, 2224–2236.

Ogawa,N., Noguchi,K., Yamashita,Y., Yasuhara,T., Hayashi,N., Yoshida,K. and Oshima,Y. (1993) Promoter analysis of the PHO81 gene encoding a 134kDa protein bearing ankyrin repeats in the phosphatase regulon of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **238**, 444–454.

Ogawa,N., Hayashi,N., Saito,H., Noguchi,K., Yamashita,Y. and Oshima,Y. (1994) Regulatory circuit for phosphatase genes in *Saccharomyces cerevisiae*: specific *cis*-acting sites in *PHO* promoters for binding the positive regulator PHO4p. In Torriani-Gorini,A., Silver,S. and Yagil,E. (eds), *Phosphate in Microorganisms: Cellular and Molecular Biology.* American Society for Microbiology, Washington DC, pp. 56–62.

Ogawa,N., Saitoh,H., Miura,K., Magbanua, J.P.V., Bun-ya,M., Harashima,S. and Oshima,Y. (1995) Structure and distribution of specific *cis*-elements for transcriptional regulation of PHO84 in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **249**, 406–416.

O'Neill,E.M., Kaffman,A., Jolly,E.R. and O'Shea,E.K. (1996) Regulation of PHO4 nuclear localization by the PHO80–PHO85 cyclin–CDK complex. *Science*, **271**, 209–212.

Oshima,Y. (1991) Impact of the Douglas-Hawthorne model as a paradigm for elucidating cellular regulatory mechanisms in fungi. *Genetics*, **128**, 195–201.

Parkinson,G., Vojtechovsky,J., Clowney,L., Brünger,A.T. and Berman,H.M. (1996) New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr.*, **D52**, 57–64.

Pesce,S. and Benezra,R. (1993) The loop region of the helix–loop–helix protein Id1 is critical for its dominant negative activity. *Mol. Cell. Biol.*, **13**, 7874–7880.

Preißner,R., Egner,U. and Saenger,W.(1991) Occurrence of bifurcated three-center hydrogen bonds in proteins. *FEBS Lett.*, **288**, 192–196.

Sakabe,N. (1991) X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl. Instrum. Methods Phys. Res.*, **A303**, 448–463.

Schneider,K.R., Smith,R.L. and O'Shea,E.K.(1994) Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. *Science*, **266**, 122–126.

Shao,D., Creasy,C.L. and Bergman,L.W. (1996) Interaction of *Saccharomyces cerevisiae* Pho2 with Pho4 increases the accessibility of the activation domain of Pho4. *Mol. Gen. Genet.*, **251**, 358–364.

Shimizu,M. (1995). Structure and function of the DNA-binding domain of the yeast PHO4 protein. PhD thesis of Osaka University.

Starovasnik,M.A., Blackwell,T.K., Laue,T.M., Weintraub,H. and Klevit,R.E. (1992) Folding topology of the disulfide-bonded dimeric DNA-binding domain of the myogenic determination factor MyoD. *Biochemistry*, **31**, 9891–9903.

Studier,F.W. and Moffatt,B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. *J. Mol. Biol.*, **189**, 113–130.

Taylor,R., Kennard,O. and Versichel,W. (1984) Geometry of the N-H · · O=C hydrogen bond. 2 three-center ('bifurcated') and four-center ('trifurcated') bonds. *J. Am. Chem. Soc.*, **106**, 244–248.

Toumoto,A., Shimizu,T., Ihara,K., Shimizu,M., Kyogoku,Y., Ogawa,N., Oshima,Y. and Hakoshima,T. (1997) Preliminary X-ray studies of a new crystal form of PHO4–DNA complex. *Acta Crystallogr.*, **D53**, 103–104.

Vogel,K., Hörz,W. and Hinnen,A. (1989) The two positively acting regulatory proteins PHO2 and PHO4 physically interact with PHO5 upstream activation regions. *Mol. Cell. Biol.*, **9**, 2050–2057.

Yoshida,K., Kuromitsu,Z., Ogawa,N. and Oshima,Y. (1989) Mode of expression of the positive regulatory genes *PHO2* and *PHO4* of the phosphatase regulon in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **217**, 31–39.

Received on February 13, 1997; revised on May 6, 1997