The crystal structure of an intact human Max–DNA complex: new insights into mechanisms of transcriptional control

P Brownlie†, TA Ceska, M Lamers, C Romier, G Stier, H Teo and D Suck*

Background: Max belongs to the basic helix-loop-helix leucine zipper (bHLHZ) family of transcription factors. Max is able to form homodimers and heterodimers with other members of this family, which include Mad, Mxi1 and Myc; Myc is an oncoprotein implicated in cell proliferation, differentiation and apoptosis. The homodimers and heterodimers compete for a common DNA target site (the E box) and rearrangement amongst these dimer forms provides a complex system of transcriptional regulation. Max is also regulated by phosphorylation at a site preceding the basic region. We report here the first crystal structure of an intact bHLHZ protein bound to its target site.

Results: The X-ray crystal structure of the intact human Max protein homodimer in complex with a 13-mer DNA duplex was determined to 2.8 Å resolution and refined to an R factor of 0.213. The C-terminal domains in both chains of the Max dimer are disordered. In contrast to the DNA observed in complex with other bHLH and bHLHZ proteins, the DNA in the Max complex is bent by about 25°, directed towards the protein. Intimate contacts with interdigitating sidechains give rise to the formation of tetramers in the crystal.

Conclusions: The structure confirms the importance of the HLH and leucine zipper motifs in dimerization as well as the mode of E box recognition which was previously analyzed by X-ray crystallography of shortened constructs. The disorder observed in the C-terminal domain suggests that contacts with additional protein components of the transcription machinery are necessary for ordering the secondary structure. The tetramers seen in the crystal are consistent with the tendency of Max and other bHLHZ and HLH proteins to form higher order oligomers in solution and may play a role in DNA looping. The location of the two phosphorylation sites at Ser1 and Ser10 (the latter is the N-cap of the basic helix) suggests how phosphorylation could disrupt DNA binding.

Address: EMBL, Structural Biology Programme, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

†Present address: MRC Protein Engineering Centre, Hills Road, Cambridgeshire, UK.

*Corresponding author.
E-mail: suck@embl-heidelberg.de

Key words: DNA binding, helix-loop-helix, leucine zipper, Max, transcription factor, X-ray crystallography

Received: 24 December 1996
Revisions requested: 21 January 1997
Revisions received: 26 February 1997
Accepted: 26 February 1997

Electronic identifier: 0969-2126-005-00509

Structure 15 April 1997, 5:509–520

Introduction

Max is one member of an interacting family of transcription factors belonging to the basic helix-loop-helix leucine zipper (bHLHZ) class of DNA-binding proteins [1]. One member of this family, Myc, is an oncoprotein implicated in cell proliferation, differentiation and apoptosis [2]. All known activities of Myc require interaction with Max and the resulting heterodimer activates transcription through the *trans*-activating domain of Myc. Dimerization of this family of proteins is necessary for DNA binding, and as with the Fos–Jun transcription factors, this Myc–Max heterodimer is the principal transcriptional activator [3]. Max, but not Myc, forms homodimers which bind the same hexanucleotide (E box) recognition site and represses transcription; partitioning between heterodimers and homodimers plays a role in the control of transcription.

Recently Max has been found to dimerize with two other members of this family of transcription factors, Mxi1 and Mad [4,5]. These species do not homodimerize and the heterodimers, like the Max homodimers, seem to have a repressive effect on transcription. This repression may work through a sequestering of Max away from Myc and through binding of the recognition site by Max dimers or Mad–Max heterodimers, thus preventing binding of the activating Myc–Max heterodimer. Whilst Max is a long lived protein and present throughout the cell cycle at constant levels, Myc is short lived and its level fluctuates through the cell cycle [6,7]. More recent data seem to indicate that the efficient repression of transcription requires ternary complex formation of the Max–Mad heterodimer with the mSin3 corepressor [8,9].

Three structural elements give rise to the name of this class of proteins: the basic region (b), the helix-loop-helix (HLH) and the leucine zipper (Z). These three elements coexist in different combinations in a total of four classes of transcription factors: bHLH (e.g. MyoD), bZ (e.g. GCN4), bHLHZ (e.g. Max) and HLH (e.g. Id). A truncated form of the Max protein was expressed and its structure, in complex with a

22 base pair DNA duplex, has been described by Burley and co-workers [10]. This first crystal structure showed that the three sequentially contiguous elements form two long helices separated by an eight residue loop. The first helix comprises the b region and the first helix of the HLH, whilst the second and parallel helix comprises the second helix of the HLH continuing into the helical Z element. Specific, major groove base contacts from the b region explain recognition of the palindromic DNA sequence CACGTG (E box). The HLH and Z elements form the dimerization interface with both elements participating in dimerization.

The crystal structures of three other proteins containing the bHLH motif show a high conservation of structure [11–13]. A conserved E box binding motif is apparent, consisting of a glutamate residue making specific base contacts. This glutamate residue forms a salt bridge to an arginine sidechain which in turn forms a hydrogen bond with the phosphate backbone. The crystal structures contain varying lengths of DNA and show different crystal packing; the DNA is not significantly bent in any of the structures.

In all of the crystal structures discussed so far, the protein has been trimmed to encompass only the DNA-recognition and dimerization motifs, motifs that were initially predicted from sequence conservation analyses. There are, however, known additional functional regions outside these motifs and it was expected that the crystal structure of a full length Max dimer in complex with DNA would provide information regarding the structure and function of these additional regions.

The Max homodimer has been reported to activate transcription in a yeast system at low levels and is dependent on intact C and N termini for maximal efficiency [14]. Phosphorylation at Ser1 and Ser10 is known to affect DNA interaction [15–17]. A nuclear localization site exists in the C-terminal region. The entire region C-terminal to the Z element is very polar, and contains a large number of serine residues.

Natural variants of Max exist, corresponding to variations at the N and C termini [1,18]. An insertion of nine amino acids between Asp11 and Ala12, encoded on a separate exon, is sometimes incorporated by alternate splicing giving rise to two forms of Max known as p21 (with no insertion) and p22 (with insertion). Additional variation arises through the absence of a large part of the C terminus (63 amino acids) giving rise to a protein called deltaMax. The C terminus carries the nuclear localization signal [19], though deltaMax may still be carried to the nucleus if dimerized with another member of this family containing an intact C terminus. The role of the Max variants is unclear.

The regulation of transcription by this family of proteins is clearly complex. Myc–Max is the principle transcriptional activating species but Max may be sequestered as Max–Max, Mxi1–Max or Mad–Max. Phosphorylation, the generation of natural variants through alternate splicing, and varying protein half-lives and transcription levels all contribute to a complex system of regulation.

Here, we present the structure of Max p21, a non-phosphorylated form of 150 amino acids, lacking the N-terminal methionine residue, and present as a homodimer bound to a 13-mer DNA duplex containing the CACGTG E box elements.

Results and discussion

Overall fold of the molecule

The crystal structure (Table 1) of p21 Max at 2.8 Å resolution confirms the bHLHZ motif, as previously observed in a truncated form of Max [10] and in the upstream stimulatory factor (USF) [11], another bHLHZ transcription factor. Although p21 Max is present in the crystals in its entirety, together with the 13 base pair duplex of DNA, in our crystals a large part of the C terminus is disordered. The present model includes residues 3–82 and 10–82 in monomers A and B, respectively. In addition, we observe density for 22 of the 26 nucleotides present in the cognate DNA. Figure 1 shows the orientation of the 11 base pair duplex DNA as bound to the Max dimer. In other proteins of the bHLHZ and bHLH classes, tetramer formation has been shown [11,20–23]. There are suggestions that Max may also form tetramers [11,24] adding a further possible level of complexity to the regulation. We observe a large contact region between two Max dimers, which lie around a crystallographic twofold axis (Fig. 2a). The space group of our crystals is $P6_522$, with 12 DNA-bound Max dimers in the unit cell. Two dyad-related dimers form intimate pairs giving rise to six homotetramers in the unit cell. There are two contact points between these dimers (indicated in Fig. 2a). The first of these is centred on the crystallographic twofold axis at residue Glu59 and extends about one helix turn in each direction (Fig. 2b): this contact point occurs, therefore, at the helix 2 zipper boundary (H2–Z). The second point of contact occurs between helix 1 of one dimer and the zipper of the other dimer (H1–Z) (Fig. 2c). Both contact points give helix crossing angles of 60–70°, thus falling within one of the favourable regions previously documented for the packing of helices [25,26]. Two DNA duplexes are bound by one tetramer.

The tetramer interface shows intimate contacts with interdigitating sidechains. Nearly 5% of the solvent accessible surface area, representing about 1000 Å^2, is lost upon two dimers forming a tetramer. At the H1–Z interface the residues Asp31, His34, Ser35 and Asp38, of one dimer, form contacts with His69, Thr70, His71, Gln73 and Asp74 of the second dimer. At the H2–Z interface the contacts involve residues Asp55, Glu59, Gln62, Tyr63 and Arg66 from both dimers; Ser35 from the second chain of the dimer also

Table 1

Summary of the data collection and phasing statistics.

Data set	Detector*	Resolution (Å)	Completeness (%)	R_{sym}† (%)	R_{iso}‡	Sites major (minor)
Native (1)	MAR(HD)	3.1	98	6.3		
Native (2)	MAR(X11)	2.8	96	9.8		
K_2HgI_4	CCD(X8C)	2.8	91	6.0	12.0	1 (2)
K_2HgI_4	MAR(X11)	4.0	94	13.3	22.0	1 (2)
$K_3UO_2F_5$	CCD(X8C)	2.8	83	9.2	35.0	4 (5)
$K_3UO_2F_5$	MAR(X11)	5.0	91	8.3	31.0	4 (5)

*HD = Heidelberg in house, MAR detector; X11 = Hamburg beamline, MAR detector; X8C = Brookhaven beamline, CCD detector.
†$R_{sym} = \Sigma||-<I>|/\Sigma I$, where I is the observed intensity and $<I>$ is the average intensity for multiple measurements. ‡$R_{iso} = \Sigma||F_{PH}|-|F_P||/\Sigma|F_P|$, where F_P is the native structure factor and F_{PH} the derivative structure factor.

participates in the interface. At both points of contact there are hydrogen bonds stabilizing the interaction, namely B55Asp Oδ1–symB63Tyr OH (Fig. 2b), A34His Nε2–symB73Gln Oε1 and A31Asp Oδ1–symB71His Nδ1, where 'sym' designates the atoms of the symmetry-related dimer (Fig. 2c).

It is also possible that simultaneous binding of a tetramer to two DNA sites, leading to DNA looping, may be functionally important. Indeed, tandem E boxes are found in many systems, including some of the known target genes of c-Myc [27–29]. Stable tetramers have been demonstrated in another protein of the bHLHZ class, USF. This homotetramer has been purified to homogeneity and shown to bind two DNA sites simultaneously, possibly leading to DNA looping [11]. Tetramer formation was shown to be dependent upon an intact leucine zipper. In other studies, the peptide corresponding to the leucine zipper of Max alone has shown no tendency to form a tetramer suggesting that the HLH region is also important [30]. The tetramer that we see is in full agreement with these observations.

DNA structure

The sequence of the DNA duplex used in the crystallization [31] is shown Figure 3a. We observed only 22 of

Figure 1

Stereo diagram of the Max dimer bound to an 11-base pair duplex DNA; Max is shown as a Cα trace and every tenth residue is labelled.

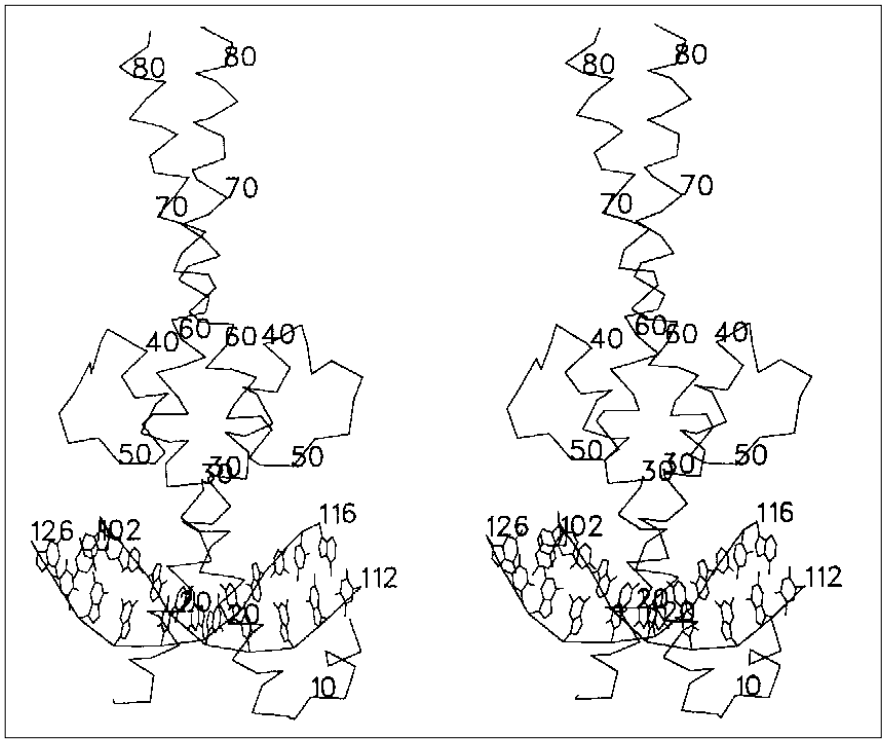
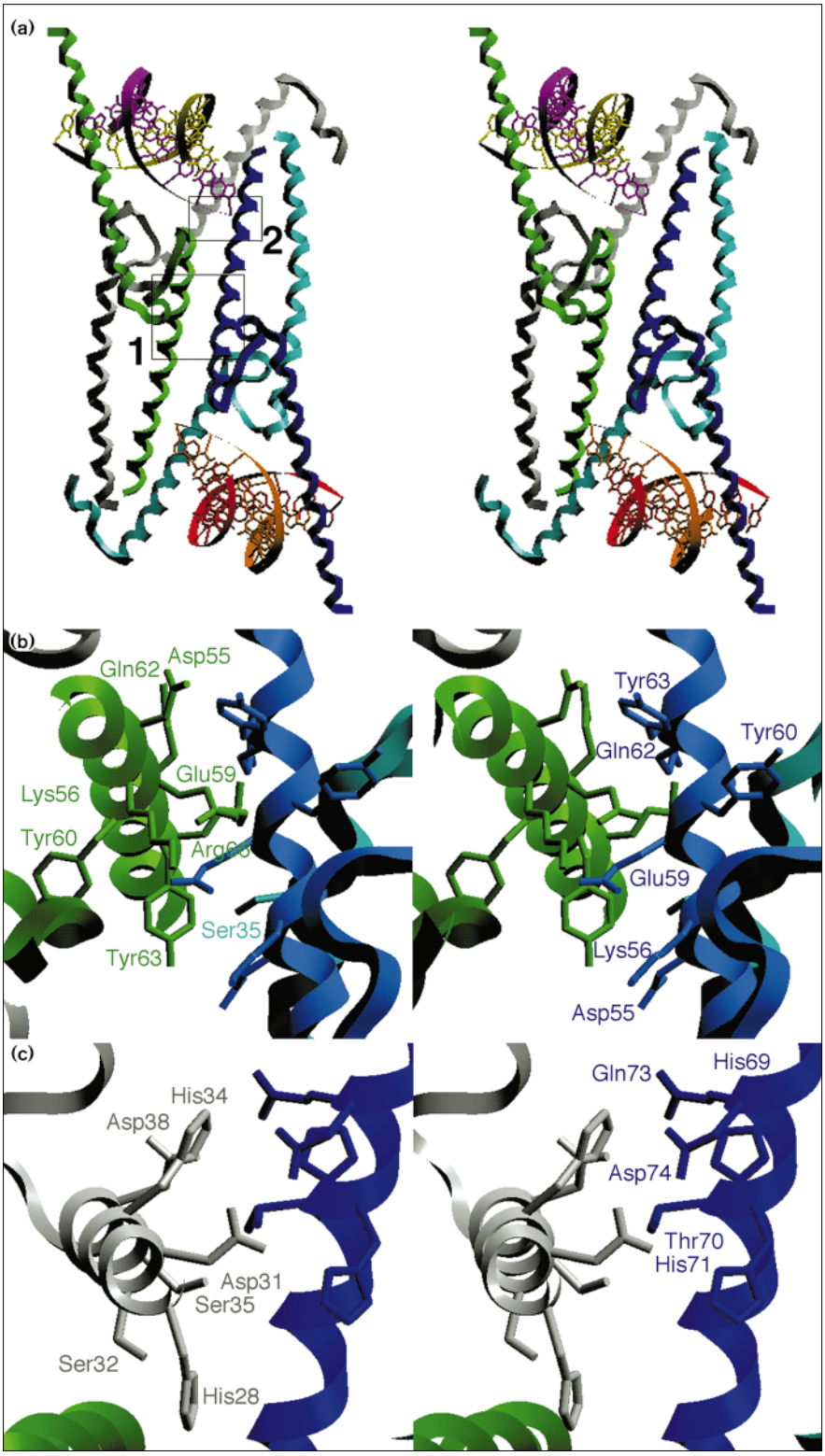


Figure 2



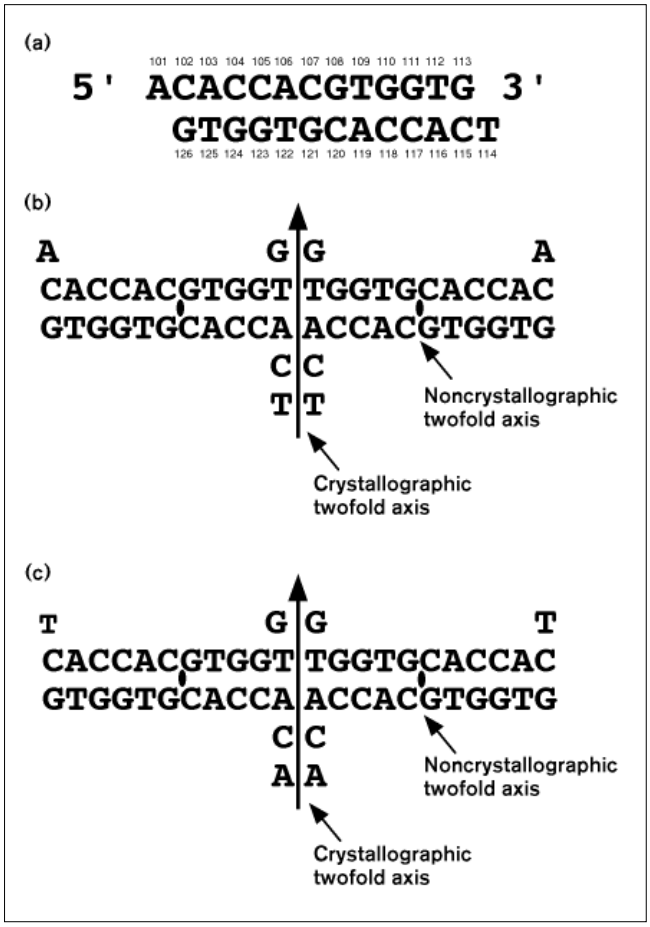
Overall Max structure and tetramer packing. (a) Stereo view of two Max dimers, with monomers coloured green and grey in one dimer and blue and cyan in the other, showing the contacts made between the two dimers that are suggestive of a tetramer contact site. The two contact regions are outlined by rectangles labelled 1 and 2. Region 1, around residues 55–66, forms the most extensive interaction region at the twofold axis of the two dimers. (b) Stereo view of residues 55–66 showing the interdigitation between the two dimers. (c) Stereo view of residues 28–38 from one chain of one dimer interacting with residues 69–74 of the chain of the second dimer. The colour-coding of the chains is the same as in (a). This second region shows less extensive interactions.

the 26 bases, with the end bases turned out of the helical conformation. Two 13-mer DNA duplexes are stacked end-to-end across a crystallographic twofold axis and the GC base pairs on either side of the dyad as well as the

overhanging A and T were not visible in the electron density. Thus the anticipated Watson–Crick base pair formation between the overhanging A and T (Fig. 3a) does not occur and instead they are forced out of the stack along with an additional GC base pair. The result is the formation of a continuous stack of 22 base pairs. To our knowledge, disruption of a DNA duplex in this manner has not been observed before. The arrangement of the end-to-end 13-mer duplexes is shown schematically in Figure 3b with G113–C115 disordered, or alternatively C102–G126 disordered as in Figure 3c (for numbering scheme see Fig. 3a). The 11-mer duplex observed is palindromic; only the end bases are unique and identify each of the two strands, leading to an ambiguity in the orientation of the two 13-mer duplexes.

The DNA in the intact Max dimer complex is bent by about 25° (over 11 base pairs) towards the bound protein, that is, towards the major groove (Figs 1 and 4). The direction of bending is opposite to that deduced from a gel mobility and phasing analysis study which was interpreted to indicate a bend of up to 80° towards the minor groove [32]. One should note, however, that there exists considerable controversy about the reliability of such measurements [33]. As we cannot rule out crystal packing effects, it is at present unclear whether the bending observed in our crystals is biologically relevant. All crystal structures to date containing the HLH motif show essentially unbent DNA. However, the observed bending, taken together with the tetramer formation suggested from the crystal, would be consistent with possible DNA looping.

Figure 3



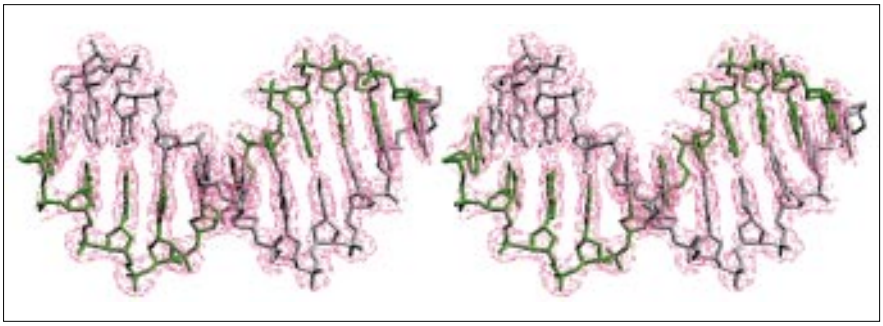
The DNA structure. (a) The DNA sequence used in the crystallization. The bases of the duplex are labelled 101 to 126 starting from the 5′-end of one strand and ending at the 3′-end of the other strand. (b) A schematic of the DNA structure showing which bases are base paired and which bases have been forced out of a helical conformation. The bases which have been forced out of the helix are poorly defined in the electron-density maps, and lead to an ambiguity in the identity of these bases. (c) The alternative sequence arrangement of the DNA packing.

DNA–protein interactions

In contrast to the truncated Max–DNA complex [10], the DNA in our structure is not averaged and strict noncrystallographic constraints were used only in the early stages of refinement (see Materials and methods). Nevertheless, the contacts made to the CACGTG core sequence are essentially symmetric indicating that they are not distorted by crystal packing effects. Specific recognition of the E box is essentially identical to that seen in the truncated Max complex. Contacts formed by Glu22 and His18 define the outer E box bases CA and GT, while Arg26 determines the identity of the central CG dinucleotide, a contact pattern which is characteristic for HLH proteins recognizing the CACGTG E-box element (Fig. 5a). In the truncated Max–DNA complex similar contacts occur between the corresponding residues p22-Glu32, p22-His28 and p22-Arg36 (the prefix 'p22' indicates the residue numbering used for the truncated Max–DNA complex in [10]).

The backbone contacts within the CACGTG sequence are also essentially the same in the two Max complexes, strongly suggesting that they are not affected by the crystal packing (Fig. 5a). Some differences outside the E box region might be expected due to the differing sequences of the oligonucleotides used for cocrystallization and because the outer parts of the DNA in the crystals of the truncated Max complex were averaged. In addition, the DNA bending observed in the intact Max complex will potentially lead to somewhat different protein–DNA contacts, particularly outside the central CACGTG sequence. We do indeed observe an additional contact (2.9Å) to the 5′-flanking guanine involving His18 Nε2 and Gua124 O6. This interaction may be of biological relevance, as biochemical data suggest that flanking bases influence the binding site selection [14,34]. A further interaction between the DNA and the Max homodimer occurs at the end of the duplex DNA where Tyr63 stacks on Gua126, as shown in Figure 5b.

Figure 4



A stereo view of the DNA showing the electron density ($2F_o–F_c$) contoured at 1σ; the DNA shows a bend of about 25°. The loop part of the Max structure is located above the DNA.

Structure of the N terminus

Electron density was only visible for one of the two N termini, presumably due to crystal contact differences. The residues 3–82 are defined in chain A, and residues 10–82 have electron density in chain B. Both monomers show a different structure in this region, as compared to the previously published model [10]. Instead of the basic region helix starting at p22-Arg25, as in the previous model, in this model the helix starts at Ser10. The sequence around this N-cap residue is common whereas the previous trimmed structure is not well N-capped and this may explain the 'fraying' of the helix ends. While the highly conserved basic p22-Arg25 (Arg15) sidechains have similar positions, the highly conserved basic p22-Lys24 (Lys14) has a totally different position. These two residues are clearly involved with DNA binding, though the disordered sidechains observed in the second monomer demonstrate that hydrogen bonding is not necessary when the DNA is already bound. The different conformations of this region, observed in the two Max structures, are due to the loss of helical structure at the N terminus in truncated Max [10].

The observed N terminus of the basic region corresponds closely with that of the MyoD structure [13], which has a capping threonine instead of serine at an equivalent position. The mainchain atoms of the two conserved basic residues mentioned above are, therefore, in similar positions in the intact p21 Max and MyoD structures and can fulfil the same function.

The whole basic region of Max and other bHLH proteins is known to undergo conformational changes upon binding DNA: a random coil structure becomes helical, demonstrating a dramatic form of induced fit [11,35,36]. Moreover, comparison of the truncated Max structure with that of USF [11] suggests plasticity in the basic region, shown by the effect of crystal contacts. The region preceding the basic region also shows flexibility as highlighted by the effects of different crystal contacts between monomers.

Phosphorylation site(s)

The residues Ser1 and Ser10 are known to be phosphorylated by casein kinase in both p21 and p22 Max [16], although the insertion of nine residues C-terminal to the phosphorylation sites in p22 will certainly alter the location of these sites. Phosphorylation of Max reduces its ability of repressing Myc transcriptional activation [17]. The structural consequences of phosphorylation/dephosphorylation are not well known and differing structural effects are seen for those proteins whose structure is known for both states. The positions of the phosphorylated serine residues are depicted in Figure 5c. The crystallographically-related protein–DNA complex is shown here to give an idea of the approximate placement of these sites relative to the DNA. Ser10 is well ordered, and is located about 16 Å from the DNA phosphate backbone. Phosphorylation of this residue would probably disrupt DNA interaction, partially by charge repulsion, but also by changing the character of the N-cap of the helix. A recent thermal denaturation study using circular dichroism suggested that phosphorylation substantially reduces the α helix forming propensity [37]. Ser1 is not visible in this structure but its approximate position can be estimated: one possible position is shown by a small sphere in Figure 5c. The position of Ser1 is clearly much closer to the DNA than Ser10, and phosphorylation of this site would clearly have its main effect by electrostatic repulsion of the protein away from the DNA.

It is interesting to note that the only other known protein structure with a casein kinase phosphorylation site, Ser101 of calmodulin [38], also has the phosphorylated serine as an N-cap residue of a helix. This secondary structural motif may represent a common casein kinase II substrate-recognition motif. Ser1 of Max may also form a helix N-cap as the first N-terminal residues that are visible, starting at Asn3, are in a helical conformation.

Comparison with other bHLHZ proteins

The structures of Max (determined from this study), MyoD [13] and the truncated form of Max [10] can be

Figure 5

DNA–protein interactions. (a) A schematic showing the interactions of one of the monomers of the dimer with the DNA. The residue numbers are shown along the sides and nucleotides which form the E box are shaded. (b) Stereo view of the symmetry-related residue Tyr63 stacking against the end DNA base Gua126. This interaction of the DNA with the protein blocks DNA–DNA stacking. The electron density ($2F_o–F_c$) is contoured at 1σ. (c) Stereo view showing the phosphorylation site Ser10 and its relation to the DNA; a symmetry-related molecule of DNA is shown. The second phosphorylation site, Ser1, is disordered in our structure, but an approximate position is indicated by a small sphere.

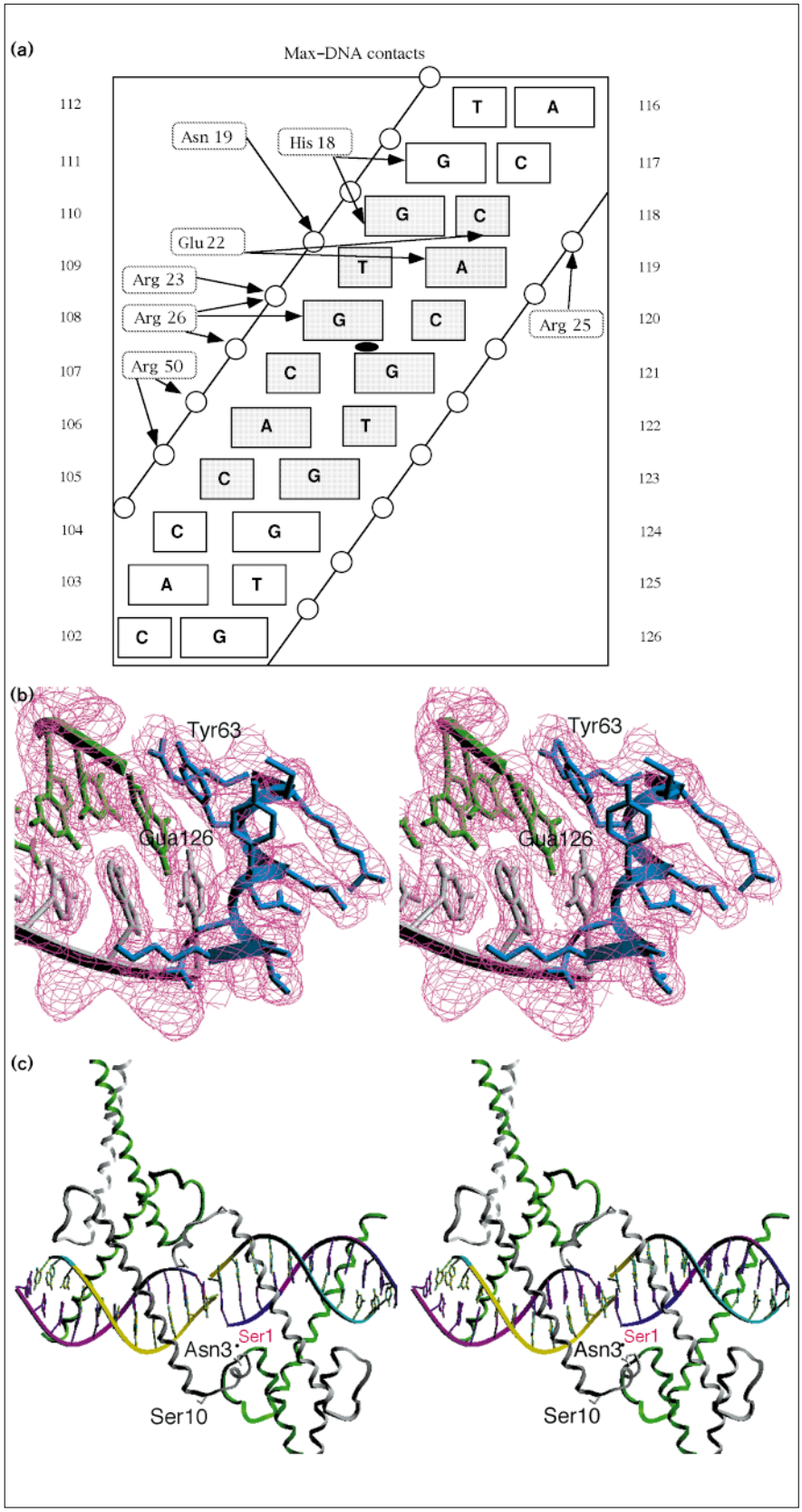
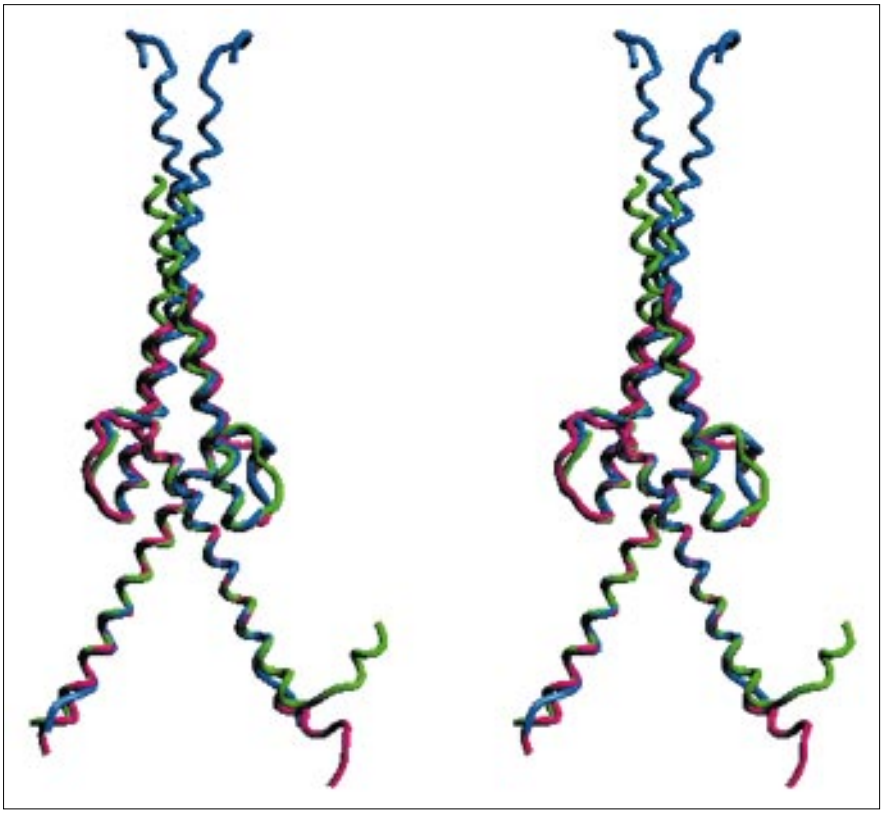


Figure 6



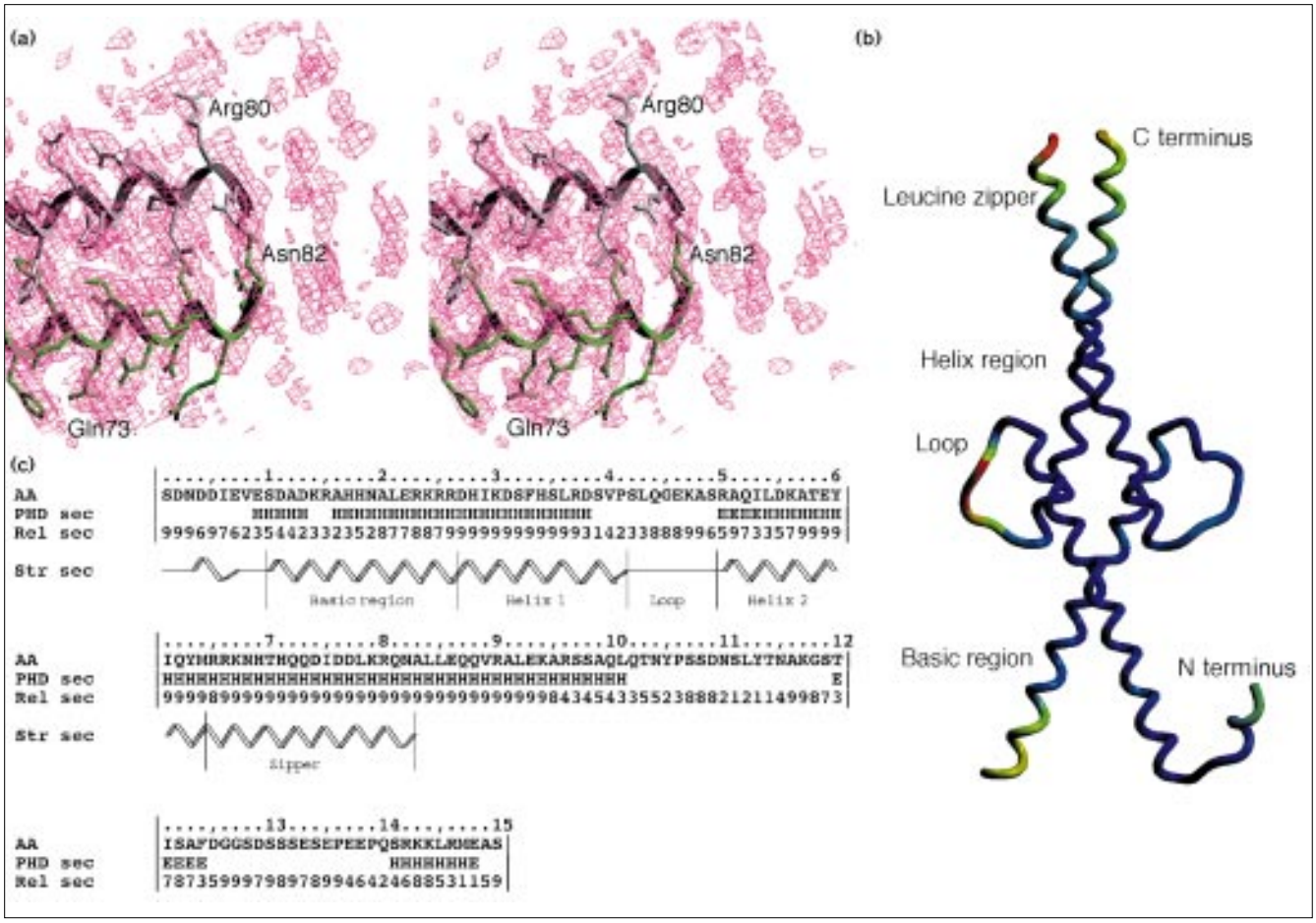
A stereo view superposition of intact Max (this study), MyoD and truncated Max (from the Burley study) showing the overlap of the three structures. Max from this study is in green, MyoD is shown in magenta and Max from Burley's group [10] is shown in blue.

compared by aligning the 20 residues in the basic region to the start of the loop (Fig. 6). The three structures superimpose well, the differences being in the details of the loops and in the extremities of the helices. Superposition of the Cα positions of residues 16–83 of our model with the corresponding positions in the truncated Max structure yields a root mean square deviation (rmsd) of 1.27 Å, with the largest differences occurring in the loop regions. There is a different conformation in the MyoD N-terminal region compared to the Max structure, but as we only observe residues 3–9 in one of the Max monomers, this suggests that the region may be conformationally flexible. The most striking difference is in the length of the helix at the C terminus of the protein. In the truncated form of Max [10] the helix is very long, continuing even after the coiled-coil interactions have ended; in the full length Max the C-terminal helix ends at residue 82 (i.e. with not all of the leucine zipper intact).

The electron density in the C-terminal region is shown in Figure 7a, some density is seen extending beyond the end of the helix, but it is not interpretable. This observation suggests that the C-terminal region may have an influence on this part of the leucine zipper structure. To try and rationalize this, a Max dimer is shown in Figure 7b colour-coded according to B factor (blue represents a low B factor; red represents a high B factor). The most flexible regions in the molecule are in the loops (especially in the longer chain) and at the extremities of the chains. The C-terminal leucine zipper is the part of the molecule that has the overall highest B factor. Thus, in the full length Max, the C-terminal part of the leucine zipper does not form a tight interaction. The truncated Max structure was described as having poorly defined density for the C-terminal region and also shows poor geometry in this region, including the leucines comprising the heptad repeat of the coiled-coil zipper region. Our structure also shows progressively worsening electron density moving towards the C terminus of the leucine zipper. The final two leucine residues of the heptad repeat, representing over two turns of helix, cannot be confidently built in simulated annealing omit maps with phases coming from the structure lacking these residues.

The question as to why this class of protein requires two dimerization motifs, when only one of the two is sufficient in other transcription factors, has not been completely answered. It seems that the leucine zipper provides the greater degree of dimerization specificity. Nevertheless, there may not be a requirement for a very stable leucine zipper in these proteins. For this reason it is possible that the full length of the leucine zipper is not necessary for

Figure 7



The structural features observed in the Max protein. (a) Stereo view electron-density ($2F_o–F_c$) map, contoured at 0.75σ, around the C terminus showing the model as it has been built and the density beyond the model which was not readily interpretable. The colour scheme is the same as in Figure 2. (b) A view of the Max dimer colour-coded according to B factor showing that the region containing the leucine zipper is the least well ordered: blue represents the lowest B factor, red the highest. (c) A secondary structure prediction of MAX using the program PHD, along with secondary structure assignments from the X-ray structure. The C-terminal region shows long regions which have little propensity for helical or β secondary structure formation. The sequence and predictions are labelled: AA, amino acid sequence; PHD sec, secondary structure prediction from PHD [42] (H, helix; E, β sheet; blank, loop); Rel sec, the reliability of the secondary structure prediction (0, lowest; 9, highest); Str sec, secondary structure observed in the X-ray structure.

dimerization, an idea that is supported by some mutation experiments. If Leu95 or Leu102 are mutated to proline residues dimerization is still observed [39]. If the first three positions of the heptad repeat are mutated, however, then dimerization is not observed. The Max variant, deltaMax, lacks Leu102 and the new residues incorporated around this position will lead to certain loss of helicity at this point. In addition, studies of isolated peptides corresponding to the zipper region show a marked decrease in dimer stability moving from GCN4 to Myc–Max to Max–Max zipper peptides [30,40]. If the formation of Max–Max homodimers is principally to act as a reservoir for the subsequent formation of activating or repressing heterodimers then the dimer stability should not be too strong. However, the backbone chain cannot be readily retraced in the leucine zipper region and because the B factors show a consistent rise in value along the zipper the protein may remain in the form of a zipper but simply be firmly fixed only at the N terminus.

The leucine zipper dimerization domain in Max clearly represents a less stable structure than the corresponding domain in GCN4 or in the Fos–Jun dimer. From the structure described here one might predict that the C-terminal half of the zipper could be deleted with minimal loss of dimer stability. While deletion of the entire zipper of USF did not prevent dimer formation, as shown in the crystal structure [11], solution studies of USF with partial deletions showed no dimer formation in the absence of DNA [41].

Finally, regarding the C-terminal domain comprising residues 93–150, this entire region is disordered in the Max structure reported here. We have used PHD [42], a secondary structure prediction program, to try and give some clues as to what sort of secondary structure might be anticipated for this region. The prediction (Fig. 7c) shows a long region of undetermined structure after the leucine zipper helical region. The undefined region is followed by two short stretches of secondary structure separated by an equally long region of indeterminate structure. As the predicted position of the bHLHZ loop is in good agreement with the structural observations, we conclude that this C-terminal region may be flexible with no apparent secondary structural core. Presumably this C terminus is required for interaction with additional components of the transcriptional machinery along with its known function of nuclear localization.

Biological implications

The human Max protein, a member of the basic helix-loop-helix leucine zipper (bHLHZ) family of transcription factors, plays a central role in a network of interacting bHLHZ proteins which regulate the biological activities of the transcription factor Myc. Myc, itself a member of the bHLHZ family, is an oncoprotein implicated in cell proliferation, differentiation and apoptosis. All known activities of Myc require interaction with Max and the resulting heterodimer activates transcription through the transactivation domain of Myc; the Myc–Max heterodimer binds to the sequence CACGTG (the so-called E box element). DNA binding occurs through the basic region of the HLH motif, while both the HLH motif and the leucine zipper form part of the dimerization interface in the bHLHZ class of proteins. Max homodimers, as well as Max–Mad or Max–Mxi1 heterodimers, bind to the same E box elements and act as transcriptional repressors. Repression results from the sequestration of Max (which thereby prevents Myc–Max heterodimer formation) and from competition for the common DNA target site. The output signal of the regulatory network will therefore depend on the relative concentrations of the individual proteins, as well as their dimerization and DNA-binding affinities. Additional levels of regulation appear to operate through phosphorylation, alternative splicing and possibly higher order oligomerization. While Max is a long lived protein and present throughout the cell cycle at constant levels, Myc is short lived and its level fluctuates.

The X-ray structure of intact human Max bound to cognate DNA confirms the general architecture seen in truncated versions of Max and other bHLHZ and bHLH proteins. The C-terminal part of the leucine zipper is not well defined in the crystal structure suggesting that it does not represent a tight interaction site in the Max homodimer. This is consistent with the observation that the two C-terminal leucine residues in the zipper are not essential for dimerization and is also in line with the concept of a dynamic network of interacting factors requiring the transient formation of homodimers and heterodimers. The C-terminal domain following the leucine zipper is disordered in the crystal structure, suggesting that it is intrinsically flexible and that other components of the transcriptional machinery may be required for secondary structure formation.

The crystal dyad-related Max dimers form tetramers with tight interhelical 'knobs-into-holes' contact regions involving both the HLH and the leucine zipper motif. This is consistent with the tendency of free, intact Max protein to aggregate in solution and of truncated Max– DNA complexes to form higher order oligomers. Tetramer formation has been observed for other members of the bHLHZ and HLH family (e.g. USF) and the potential role of these bivalent homotetramers in DNA looping has been discussed.

The protein–DNA contacts involved in recognition of the E box element are essentially identical to those seen in the structure of the truncated Max–DNA complex. We do observe, however, a contact to the guanine base just outside the central core which may influence the preference for flanking bases. In contrast to the structure of the DNA observed in complex with other bHLH proteins, the DNA in the intact Max–DNA complex is not straight, but bent by about 25° towards the bound protein (towards the major groove). It remains to be seen whether the observed bending is an intrinsic property of the intact Max protein or a consequence of crystal packing forces.

Another difference to the previous Max structure concerns the N termini. In the intact Max–DNA complex the basic region helix is extended by a further helical turn starting at residue Ser10, which represents a proper helical N-cap. As a consequence of crystal contacts additional residues, starting from Asn3, are visible in one monomer highlighting the flexibility of this region. However, the proximity of the two casein kinase phosphorylation sites, Ser10 and particularly Ser1, to the DNA backbone suggest that electrostatic repulsion is partly responsible for the inhibitory effect of phosphorylation on DNA binding by Max homodimers. In addition, changing the character of the N-cap residue of the basic region helix by phosphorylation will affect DNA binding.

Materials and methods

Full length human Max (p21) was over-expressed in *Escherichia coli* and purified using heparin sepharose and mono S columns. The protein was concentrated to 18 mg ml^{-1} by vacuum dialysis, combined with annealed 13-mer DNA (1:2; protein dimer:double-stranded DNA) and crystallized by the hanging-drop vapour diffusion method. The well solution

Table 2

Refinement statistics.

Resolution (Å)	8.0–2.8
Number of reflections	10 256
Number of reflections at 2σ	9496
Rms bond length (Å)	0.013
Rms bond angle (°)	1.9
Crystallographic R factor (%)	21.3
Free R factor for 950 reflections (%)	27.3
Total number of atoms	2188
Total number of waters	18
Overall figure of merit	0.464

contained 21% 2-methyl-2,4-pentanediol (MPD). 100 mM sodium citrate pH 3.9–4.0 with an initial drop volume of 2 μl. Crystals appeared after about one week and grew for several weeks to a maximum size of 0.8 mm × 0.2 mm × 0.2 mm in the form of hexagonal rods. The space group is $P6_522$ with cell dimensions a = 107.8, c = 126.9 with a solvent content of 53%. One dimer–DNA complex exists in the asymmetric unit. SDS gel electrophoresis and HPLC analysis of carefully washed crystals confirmed the presence of intact Max protein and 13-mer DNA in the hexagonal crystals (data not shown).

Native data was collected on a MAR image plate, with either CuKα radiation from an Elliot GX21 rotating anode or with synchrotron radiation at 1.0 Å. One data set from each derivative was collected on a MAR image plate at station X11 of the Hamburg synchrotron and the others with a CCD detector at the Brookhaven synchrotron, station X8C. All data was collected at 0–5°C. Data processing was done with XDS [43]. Molecular replacement and phase calculations were carried out with the CCP4 package [44].

The coordinates from the trimmed Max crystal structure [10] were modified to give a search model representing about 60% of scattering mass. Outstanding solutions for both the rotation function and the translation function positioned the dimer complex with its long axis almost parallel to the c axis. Rigid-body refinement against the same subset of data (15–3.2 Å) showed the R factor to drop from 52% to 49%. All refinement procedures were carried out with X-PLOR [45]. After one round of atomic positional refinement with tight twofold restraints between the monomers, model phases were used in a difference Fourier map to locate all of the heavy-atom sites listed in Table 1, which were then refined. The overall figure of merit for these multiple isomorphous replacement (MIR) phases was 0.54 at 4 Å and a map at this resolution showed the two long helices of the Max monomers to be consistent with the previous model, as well as revealing the DNA. All later electron-density maps were calculated using these experimental phases combined with the current model phases using the program SIGMAA [46].

Several rounds of refinement with X-PLOR incorporating data to 2.8 Å, with easing of the restraints and eventual refinement of individual tightly restrained isotropic atomic B factors allowed additional residues and bases to be built [47]. Calculation of difference maps, the free R factor and use of geometry checks were used throughout to monitor the refinement. The final conventional R factor is 21.3% and the free R factor is 27.3% with an rms deviation in bond lengths of 0.013 Å and bond angles of 1.90° (Table 2).

Solvent flattening, histogram matching, connectivity restraints and twofold averaging of the density [48] had no significant effect in revealing additional structure.

Accession numbers

The atomic coordinates have been deposited at the Brookhaven Protein Data Bank.

Acknowledgements

The cDNA originating from the laboratory of RN Eisenman was given to us by A Wenzel. We would like to thank C Ziepluch for initial biochemical studies, S Burley for providing the truncated Max coordinates, S Ginnell for assistance collecting data at Brookhaven, and the staff at the Hamburg EMBL outstation for assistance in collecting data at DESY.

References

1. Blackwood, E.M. & Eisenman, R.N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* 251, 1211–1217.
2. Amati, B. & Land, H. (1994). Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr. Opin. Cell Biol.* 4, 102–108.
3. Amati, B., Dalton, S., Brooks, M.W., Littlewood, T.D., Evan, G.I. & Land, H. (1992). Transcriptional activation by the human c-Myc protein in yeast requires interaction with Max. *Nature* 359, 423–426.
4. Ayer, D.E., Kretzner, L. & Eisenman, R.N. (1993). Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* 72, 211–222.
5. Zervos, A.S., Gyuris, J. & Brent, R. (1993). Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell* 72, 223–232.
6. Marcu, K.B., Bossone, S.A. & Patel, A.J. (1992). Myc function and regulation. *Annu. Rev. Biochem.* 61, 809–860.
7. Larsson, L.G., Pettersson, M., Oeberg, F., Nilsson, K. & Luescher, B. (1994). Expression of mad, mxi1, max, and c-myc during induced differentiation of hematopoietic cells: opposite regulation of mad and c-myc. *Oncogene* 9, 1247–1252.
8. Ayer, D.E., Lawrence, Q.A. & Eisenman, R.N. (1995). Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of the yeast repressor Sin3. *Cell* 80, 767–776.
9. Ayer, D.E., Laherty, C.D., Lawrence, Q.A., Armstrong, A.P. & Eisenman, R.N. (1996). Mad proteins contain a dominant transcription repression domain. *Mol. Cell. Biol.* 16, 5772–5781.
10. Ferre-D'Amare, A.R., Prendergast, G.C., Ziff, E.B. & Burley, S.K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363, 38–45.
11. Ferre-D'Amare, A.R., Pognonec, P., Roeder, R.G. & Burley, S.K. (1994). Structure and function of the b/HLH/Z domain of USF. *EMBO J.* 13, 180–189.
12. Ellenberger, T., Fass, D., Arnaud, M. & 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.
13. Ma, P.C.M., Rould, M.A., Weintraub, H. & 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.
14. Fisher, F., Crouch, D.H., Jayaraman, P.S., Clark, W., Gillespie, D.A.F. & Goding, C.R. (1993). Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGTG motifs *in vivo*. *EMBO J.* 12, 5075–5082.
15. Berberich, S.J. & Cole, D.M. (1992). Casein kinase II inhibits the DNA-binding activity of Max homodimers but not Myc/Max heterodimers. *Genes Dev.* 6, 166–176.
16. Bousset, K., Henriksson, M., Luescher-Firzlaff, J.M., Litchfield, D.W. & Luescher, B. (1993). Identification of casein kinase II phosphorylation sites in Max: effects on DNA-binding kinetics of Max homo- and Myc/Max heterodimers. *Oncogene* 8, 3211–3220.
17. Koskinen, P.J., Vaestrik, I., Maekelae, T.P., Eisenman, R.N. & Alitalo, K. (1994). Max activity is affected by phosphorylation at two NH2-terminal sites. *Cell Growth Differ.* 5, 313–320.
18. Maekelae, T.P., Koskinen, P.J., Vaestrik, I. & Alitalo, K. (1992). Alternative forms of Max as enhancers or repressors of Myc-Ras cotransformation. *Science* 256, 373–377.
19. Kato, G.J., Lee, M.F.W., Chen, L. & Dand, C.V. (1992). Max: functional domains and interaction with c-Myc. *Genes Dev.* 6, 81–92.
20. Dang, C.V., McGuire, M., Buckmire, M. & Lee, W.M. (1989). Involvement of the 'leucine zipper' region in the oligomerization and transforming activity of human c-Myc protein. *Nature* 337, 664–666.
21. Fisher, D.E., Carr, C.S., Parent, L.A. & Sharp, P.A. (1991). TFEB has DNA-binding and oligomerization properties of a unique helix-loop-helix/leucine zipper family. *Genes Dev.* 5, 2342–2352.
22. Sha, M., Ferre-D'Amare, A.R., Burley, K.S. & Goss, D.J. (1995). Anti-cooperative biphasic equilibrium binding of transcription factor upstream stimulatory factor to its cognate DNA monitored by protein fluorescence changes. *J. Biol. Chem.* 270, 19325–19329.

23. Fairman, R., *et al.*, & Brenner,S.L. (1993). Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides. *Proc. Natl. Acad. Sci. USA* **90**, 10429–10433.
24. Ferre-D'Amare, A.R. & Burley, S.K. (1994). Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. *Structure* **2**, 357–359.
25. Chothia, C., Levitt, M. & Richardson, D. (1977). Structure of proteins: packing of α helices and pleated sheets. *Proc. Natl. Acad. Sci. USA* **74**, 4130–4134.
26. Richmond, T.J. & Richards, F.M. (1978). Packing of helices: geometrical constraints and contact areas. *J. Mol. Biol.* **119**, 536–555.
27. Tobias, K.E., Shor, J. & Kahana, C. (1995). c-Myc and Max transregulate the mouse ornithine decarboxylase promoter through interaction with two downstream CACGTG motifs. *Oncogene* **11**, 1721–1727.
28. Desbarats, L., Gaubatz, S. & Eilers, M. (1996). Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc. *Genes Dev.* **10**, 447–460.
29. Jones, R.M., *et al.*, & Schmidt, E.V. (1996). An essential E-box in the promoter of the gene encoding the mRNA cap-binding protein (eukaryotic initiation factor 4E) is a target for activation by c-myc. *Mol. Cell. Biol.* **16**, 4754–4764.
30. Muhle-Goll, C., Nilges, M. & Pastore, A. (1995). The leucine-zippers of the HLH-LZ proteins Max and c-Myc preferentially form heterodimers. *Biochemistry* **43**, 13554–13564.
31. Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. & Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. *Science* **250**, 1149–1151.
32. Fisher, D.A., Parent, L.A. & Sharp, P.A. (1992). Myc/Max and other helix-loop-helix/leucine zipper proteins bend DNA toward the minor groove. *Proc. Natl. Acad. Sci. USA* **89**, 11779–11783.
33. Hagerman, P.J. (1996). Do basic region-leucine zipper proteins bend their DNA targets...does it matter? *Proc. Natl. Acad. Sci. USA* **93**, 9993–9996.
34. Bendall, A.J. & Molloy, P.L. (1994). Base preference for DNA binding by the bHLH-Zip protein USF: effects of $MgCl_2$ on specificity and comparison with binding of Myc family members. *Nucl. Acids Res.* **22**, 2801–2810.
35. Fisher, D.E., Parent, L.A. & Sharp, P.A. (1993). High affinity DNA-binding Myc analogs: recognition by an α helix. *Cell* **72**, 467–476.
36. Anthony-Cahill, S.J., *et al.*, & DeGrado, W.F. (1992). Molecular characterization of helix-loop-helix peptides. *Science* **255**, 979–983.
37. Szilak, L., Moitra, J., Krylov, D. & Vinson, C. (1997). Phosphorylation destabilizes α helices. *Nat. Struct. Biol.* **4**, 112–114.
38. Quadroni, M., James, P. & Carafoli, E. (1994). Isolation of phosphorylated calmodulin from rat liver and identification of the *in vivo* phosphorylation sites. *J. Biol. Chem.* **269**, 16116–16122.
39. Reddy, C.D., Dasgupta, P., Saikumar, P., Dudek, H., Rauscher, F.J. & Reddy, E.P. (1992). Mutational analysis of Max: role of basic, helix-loop-helix/leucine zipper domains in DNA binding, dimerization and regulation of Myc-mediated transcriptional activation. *Oncogene* **7**, 2085–2092.
40. Muhle-Goll, C., *et al.*, & Pastore, A. (1994). The dimerization stability of the HLH-LZ transcription protein family is modulated by the leucine zippers. A CD and NMR study of TEFB and c-Myc. *Biochemistry* **33**, 11296–11306.
41. Gregor, P.D., Sawadodgo, M. & Roeder, R.G. (1990). The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev.* **4**, 1730–1740.
42. Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584–599.
43. Kabsch, W.J. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown cell constants. *J. Appl. Cryst.* **26**, 795–800.
44. Collaborative Computational Project No.4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
45. Brünger, A.T. (1992). *X-PLOR, Version 3.1*. Yale University Press, New Haven, CT, USA.
46. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149.
47. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
48. Vellieux, F.M.D.A., Hunt, J.F., Roy, S. & Read, R.J. (1995). DEMON/ANGEL: a suite of programs to carry out density modifications. *J. Appl. Cryst.* **28**, 347–351.