A helix–turn–helix structure unit in human centromere protein B (CENP-B)

Junji Iwahara1,2, Takanori Kigawa2, Katsumi Kitagawa3, Hiroshi Masumoto3, Tuneko Okazaki3,4 and Shigeyuki Yokoyama1,2,5

1Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, 2Cellular Signaling Laboratory, Institute of Physical and Chemical Research (RIKEN), Wako-shi, Saitama 351-01 and 3Department of Molecular Biology, School of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan 4Present address: Institute for Comprehensive Medical Science, Fujita Health University, School of Medicine, Toyoake, Aichi, 470-11, Japan 5Corresponding author e-mail: yokoyama@y-sun.biochem.s.u-tokyo.ac.jp

CENP-B has been suggested to organize arrays of centromere satellite DNA into a higher order structure which then directs centromere formation and kinetochore assembly in mammalian chromosomes. The N-terminal portion of CENP-B is a 15 kDa DNA binding domain (DBD) consisting of two repeating units, RP1 and RP2. The DBD specifically binds to the CENP-B box sequence (17 bp) in centromere DNA. We determined the solution structure of human CENP-B DBD RP1 by multi-dimensional 1H, 13C and 15N NMR methods. The CENP-B DBD RP1 structure consists of four helices and has a helix–turn–helix structure. The overall folding is similar to those of some other eukaryotic DBDs, although significant sequence homology with these proteins was not found. The DBD of yeast RAP1, a telomere binding protein, is most similar to CENP-B DBD RP1. We studied the interaction between CENP-B DBD RP1 and the CENP-B box by the use of NMR chemical shift perturbation. The results suggest that CENP-B DBD RP1 interacts with one of the essential regions of the CENP-B box DNA, mainly at the N-terminal basic region, the N-terminal portion of helix 2 and helix 3.

Keywords: CENP-B/CENP-B box/centromere/NMR structure/protein–DNA interaction

Introduction

The centromere plays a major role in segregation of eukaryotic chromosomes in mitosis and meiosis, by serving as the site for kinetochore assembly and sister chromatid attachment (Clarke, 1990; Willard, 1990). In addition, the centromere is considered to regulate the cell cycle checkpoint for the metaphase–anaphase transition (Nicklas et al., 1995; Rieder et al., 1995). To understand these functions of the centromere at the molecular level, the DNAs and the proteins in centromeres have been characterized for a variety of eukaryotes (Pluta et al., 1995). In human cells several proteins localized in the centromere have been identified with centromere-specific autoantibodies (Earnshaw and Rothfield, 1985). CENP-A (17 kDa), CENP-B (80 kDa) and CENP-C (140 kDa) are such proteins with DNA binding activities (Masumoto et al., 1989; Palmer et al., 1991; Sugimoto et al., 1994; Sullivan et al., 1994). Thus far a sequence-specific DNA binding activity has been found only for CENP-B (Muro et al., 1992).

In the CENP-B amino acid sequence deduced from the cloned gene (Sullivan and Glass, 1991) there is a DNA binding domain (DBD) within the N-terminal 125 residues (Yoda et al., 1992). This domain of CENP-B binds with high affinity to a 17 bp sequence, the CENP-B box, as shown in Figure 1A (Masumoto et al., 1989; Muro et al., 1992). Three regions, consisting of 4, 1 and 4 bp (Figure 1A), form the essential core recognition sequence for CENP-B to bind to the CENP-B box sequence (Masumoto et al., 1993; Yoda et al., 1996). The CENP-B box sequence frequently exists in human α-satellite (alphoid) DNA, the human centromere-specific repeating DNA family composed of 171 bp monomer units with chromosome-specific sequence variations (Willard, 1990; Yoda and Okazaki, 1997). The distribution of CENP-B boxes was studied in human chromosome 21 and it was found that CENP-B boxes are regularly spaced over a 1.3 Mbp region in one of the two adjacent α-satellite DNA arrays (Ikeno et al., 1994). CENP-B has a dimerization domain in the C-terminal 59 residues, which is separate from the DBD (Kitagawa et al., 1995), and the CENP-B dimer forms either a complex containing two DNA molecules with a CENP-B box (Muro et al., 1992) or a loop structure on a DNA strand containing two CENP-B boxes (K.Yoda, A.Okuda, A.Kikuchi and T.Okazaki, submitted for publication). These results suggested that the function of CENP-B in vivo may be to organize the long centromeric satellite arrays with frequent CENP-B boxes into a higher order chromatin structure by juxtaposing pairs of CENP-B box sequences and that this structure may then become the foundation for the centromere/kinetochore structure and activity (Muro et al., 1992; Ikeno et al., 1994; Yoda et al., 1996).

CENP-B genes have also been cloned from mouse, hamster and African green monkey (Sullivan and Glass, 1991; Bejarano and Valdivia, 1996; Yoda et al., 1996). The amino acid sequences of human CENP-B and the mammalian homologs are highly conserved, particularly in the DBD (100% identity) (Figure 1C) and the dimerization domain (98% identity). The 9 base pairs (sites 1–3 in Figure 1A) recognized by human CENP-B are conserved in the other mammalian CENP-B binding sites studied thus far (Figure 1B) (Muro et al., 1992; Kipling et al., 1995; Yoda et al., 1996), although the frequency of
occurrence of CENP-B binding sites in centromeres is quite different among species (Goldberg et al., 1996; Romanova et al., 1996). Therefore, the mechanisms of DNA binding and dimerization of CENP-B are probably conserved among mammals. Furthermore, two CENP-B homologs (Abp1/Cbp1 and Cbh) have been identified in the fission yeast \textit{Schizosaccharomyces pombe} (Murakami et al., 1996; Lee et al., 1997) and their binding sites have been found in the K repeat and/or central core sequence in the essential centromere DNA (Halverson et al., 1997; Lee et al., 1997; Ngan and Clarke, 1997).

The CENP-B DBD was proposed to have a helix–loop–helix structure consisting of ~60 residues at its N-terminal region, on the basis of the low sequence similarity with helix–loop–helix proteins such as Myc and MyoD (Sullivan and Glass, 1991). On the other hand, the helix–loop–helix DNA binding motif mediates dimerization (Ferre et al., 1993; Ma et al., 1994), while the CENP-B DBD binds to the CENP-B box sequence as a monomer (Yoda et al., 1992). Recently it was suggested, based on the amino acid sequence, that the CENP-B DBD consists of two repeating units (RP1 and RP2), as shown in Figure 1C (Suzuki and Brenner, 1995; Suzuki et al., 1995). CENP-B DBD RP1 mostly corresponds to the region that was suggested to have sequence similarity with the helix–loop–helix proteins.

To our knowledge there have been no reports about the tertiary structure of any protein involved in centromere function. In order to gain insight into the molecular mechanism of the interaction between CENP-B and CENP-B box DNA at atomic resolution we have begun to study the CENP-B DBD. In this paper we report the tertiary structure of CENP-B DBD RP1, as determined by heteronuclear multidimensional NMR.

\section*{Results and discussion}

The first repeat (RP1) of the CENP-B DBD was expressed under control of the T7 promoter in \textit{Escherichia coli} as a polypeptide with 14 N-terminal vector-derived residues, including a 6× His tag (MRGSHHHHHHGMAS), followed by the CENP-B DBD RP1 moiety (amino acid residues 1–56; Figure 1C). About 6 mg/l culture purified CENP-B DBD RP1 was obtained using M9 minimal medium. The circular dichroism (CD) data showed that CENP-B DBD RP1 is predominantly α-helical; the α-helix content at 20°C was estimated to be 54%. In addition, CENP-B DBD RP1 exhibited cooperative thermal denaturation, centered at 58°C, as monitored by CD at 222 nm under conditions of 20 mM potassium phosphate, 400 mM Na$_2$SO$_4$, 0.25 mM EDTA and 0.01% NaN$_3$ at pH 6.0 (data not shown). These results indicate that CENP-B DBD RP1 itself is a well-structured unit.

\section*{NMR assignments}

NMR measurements for resonance assignments and structure determination of CENP-B DBD RP1 were performed under conditions of 20 mM potassium phosphate, 400 mM Na$_2$SO$_4$, 0.25 mM EDTA and 0.01% NaN$_3$ at pH 6.0. Figure 2A shows the 2D $^1$H–$^1$N HSQC spectrum of CENP-B DBD RP1. First, ~80% of the $^1$H and $^{15}$N resonances of CENP-B DBD RP1 were assigned by use of 3D $^1$H–$^1$N NOESY-HSQC, 3D $^1$H–$^{15}$N TOCSY-HSQC, 3D HNHB, 2D $^1$H–$^{15}$N HSQC, 2D NOESY, 2D TOCSY and 2D DQF-COSY spectra. We completed the resonance assignments using the 3D CT-HNCA, 3D CT-HNCOCA, 3D HCCO-TOCSY and 3D $^1$H–$^{15}$C NOESY-HSQC spectra. Figure 2B shows an example of sequential assignment along the Leu42–Leu53 backbone using the 3D CT-HNCA spectrum.
A helix-turn-helix structure unit in CENP-B

Fig. 2. (A) The $^1$H-$^{15}$N HSQC spectrum of CENP-B DBD RP1. The negatively numbered residues are from the N-terminal vector-derived region. The cross-peaks due to sidechain resonances are indicated by sc. The spectral width in the $^{15}$N dimension was 775 Hz and aliased cross-peaks are displayed without sign discrimination. (B) Sequential resonance assignment for Leu42–Leu53 in the 3D CT-HNCA spectrum of CENP-B DBD RP1. The connectivities of the inter-residue cross-peaks are indicated with arrows. The sample used for the NMR measurements was CENP-B DBD RP1 (1.5 mM) in 20 mM potassium phosphate buffer, pH 6.0, containing 400 mM Na$_2$SO$_4$, 0.25 mM EDTA and 0.01% NaN$_3$.

Structure determination
The 14 vector-derived residues exhibited only a few inter-residue and negative $[^1]$H-$^{15}$N NOEs, indicating that the region was unstructured. Therefore, structure determination was performed only for the 56 CENP-B-derived residues. The secondary structure elements of CENP-B DBD RP1 are shown in Figure 3, together with a summary of the sequential and medium range NOEs, the $^3J_{HHH}$ coupling constants, observation of HN-H$_2$O cross-peaks, $^{13}$C$_\alpha$ chemical shift indices and $[^1]$H-$^{15}$N NOE values. For calculation of tertiary structure inter-proton NOE information was obtained from the 3D $^1$H-$^{15}$N NOESY-HSQC, 3D $^1$H-$^{13}$C NOESY-HSQC and 2D homonuclear NOESY spectra, with 662 restraints (466 inter-residue distance restraints, 160 intra-residue distance restraints and 36 dihedral angle restraints) being used. Superimposition of the final family of 20 refined structures is shown in Figure 4A. The backbone root mean square (r.m.s.) deviation of the family was 0.48 ± 0.17 Å. Other statistics of these structures are listed in Table I. Figure 4B shows overall folding of the energy minimized average structure of CENP-B DBD RP1.

Solution structure of CENP-B DBD RP1
Four $\alpha$-helices (helix 1, Phe10–Asn23; helix 2, Lys28–Phe35; helix 3, Ser40–Asn48; helix 4, Lys49–Ala54) exist in the solution structure of CENP-B DBD RP1. A canonical helix-turn-helix (HTH) structure is observed in the region from helix 2 to helix 3. This result is roughly consistent with the prediction by Suzuki et al. (1995). CENP-B DBD
small as 1.4 Å (Figure 5A). However, the orientation of helix 1 relative to the HTH region of CENP-B DBD RP1 is different from that of the *fushi tarazu* homeodomain (Figure 5A). Setting aside the turn regions, the relative orientations between helices 1, 2 and 3 of CENP-B DBD RP1 are more similar to those of RAP1 and the Myb DBD repeating units than to that of the *fushi tarazu* homeodomain (Figure 5B and C). In particular, the structure of RAP1 DBD domain 1 superimposes well upon that of CENP-B DBD RP1, with an r.m.s. deviation value of 2.1 Å for the Cα atoms of the three helices.

**Helix 4 of CENP-B DBD RP1**

Helix 4 is characteristic of CENP-B DBD RP1. Helices 3 and 4 are directly connected in the amino acid sequence and may be regarded as one long helix kinked at the N–Cα bond of Asn48 (with a dihedral angle $\phi = -97^\circ$ for the minimized average structure). Long range $^1$H–$^1$H NOEs between helices 1 and 4 were observed for several residues; in particular the aromatic ring protons of Phe10 (helix 1) gave 13 NOEs to the backbone and sidechain protons of Ile52 and Leu53 (helix 4). Note that helix 1 also makes a hydrophobic contact with helix 3, as described above. Therefore, hydrophobic interaction of helix 1 with helices 3 and 4 defines their relative locations, with an angle of 129° between the helix axes (Figure 4C). The anomalous structure at Asn48, at the junction of helices 3 and 4, was also supported by two independent sources of experimental data that were not included as constraints in the structure determination. First, in the region from Ser40 to Ala54 only Asn48 exhibited a $^3J_{HNH}$ value $>8$ Hz, whereas other residues exhibited $^3J_{HNH}$ values $<5$ Hz (Figure 3). Second, all of the residues from Ser40 to Ile52 except Asn48 had positive $^{13}$Cα chemical shift indices (Wishart and Sykes, 1994), while that of Asn48 was negative (Figure 3). The slow solvent exchange of the Asn48 and Ile52 backbone amide protons indicates that the backbone amide and carbonyl groups of Asn48 are involved in hydrogen bonds within helices 3 and 4 respectively. Helix 4 seems to be more flexible than the other three helices in CENP-B DBD RP1, because the average value of the backbone amide $^1$H,$^1$N NOEs of helix 4 (0.62) is smaller than those of helices 1–3 (0.76, 0.74 and 0.75 respectively) at 600 MHz.

*The Antennapedia* homeodomain (Qian et al., 1989) and domain 2 of the RAP1 DBD (König et al., 1996) also have helix 4 connected directly to helix 3. Helices 3 and 4 of the Antennapedia homeodomain are nearly coaxial (with an angle of 161°), while those of RAP1 DBD domain 2 make an angle of 92°. In between are helices 3 and 4 of CENP-B DBD RP1, which makes an angle of 129°. Nevertheless, because helix 3 of RAP1 DBD domain 2 is six residues longer than that of CENP-B DBD RP1, helix 4 of RAP1 DBD domain 2 makes a hydrophobic contact with the N-terminal portion of helix 1, as in the case of CENP-B DBD RP1. These hydrophobic interactions between helices 1 and 4 may contribute to the stability of domain folding (König et al., 1996). In addition, the angle between helices 3 and 4 seems to be important for location of the following repeat, RP2, of CENP-B DBD on the DNA, as discussed below.
A helix–turn–helix structure unit in CENP-B

Fig. 4. Solution structure of CENP-B DBD RP1. (A) Superimposition of the backbones of 20 NMR-derived structures of CENP-B DBD RP1 (stereoview). (B) Ribbon representation of the energy-minimized average structure derived from the 20 refined CENP-B DBD RP1 structures, as viewed from the same angle as in (A). Helix 1, blue; helix 2, green; helix 3, light green; helix 4, orange. (C) Hydrophobic contacts between helices. The N-terminal nine residues are omitted for simplicity. The colors of the helices are the same as in (B). The molecular graphics package MOLSCRIPT (Kraulis, 1991) was used.

Table I. Structural statistics for the 20 refined structures of CENP-B DBD RP1

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X-PLOR energies (kcal/mol)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_{\text{van der Waals}}$</td>
<td>26.9 ± 5.1</td>
<td></td>
</tr>
<tr>
<td>$E_{\text{Lennard–Jones}}$</td>
<td>$-118.7$ ± 29.4</td>
<td></td>
</tr>
<tr>
<td>r.m.s. deviation from experimental restraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance (Å)</td>
<td>0.037 ± 0.003</td>
<td></td>
</tr>
<tr>
<td>Dihedral angle (degree)</td>
<td>0.29 ± 0.25</td>
<td></td>
</tr>
<tr>
<td>Deviations from ideal covalent geometry used within X-PLOR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.0034 ± 0.0002</td>
<td></td>
</tr>
<tr>
<td>Angles (degree)</td>
<td>0.61 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>Impropers (degree)</td>
<td>0.45 ± 0.02</td>
<td></td>
</tr>
</tbody>
</table>

$^a$The quadratic van der Waals term was calculated with a force constant of 4 kcal/mol/Å$^4$ and the van der Waals radii were set to 0.75 times the standard value used in CHARMM.

$^b$Lennard–Jones potential calculated using the CHARMM empirical energy function.

DNA binding sites on the CENP-B DBD RP1 structure

A gel mobility shift analysis was performed with DNA fragments of various lengths and the entire CENP-B DBD (residues 1–129, RP1–RP2). The minimum DNA fragment for CENP-B binding was determined to be a 21 bp fragment (CB21, Figure 7) in which the 17 bp CENP-B box sequence is centered (data not shown). We used the CB21 fragment to analyze changes in the NMR spectra upon interaction of CENP-B DBD RP1 with DNA. $^{15}$N-labeled CENP-B DBD RP1 was mixed with CB21 under conditions of 250 mM Na$_2$SO$_4$, 20 mM potassium phosphate, 0.25 mM EDTA and 0.01% NaN$_3$, pH 6.0, where the ratio of specific to non-specific interactions was maximized and co-precipitation of CENP-B DBD RP1 and CB21 was avoided. Five 2D $^1$H-$^{15}$N HSQC spectra were measured at different molar ratios (0, 0.38, 0.75, 1.13 and 1.50) of CB21 to $^{15}$N-labeled CENP-B DBD RP1. Upon addition of the CB21 DNA fragment, a set of amino acid residues exhibited large chemical shift perturbations in a dose-dependent manner. Each of these residues showed only one $^1$H–$^{15}$N cross-peak at a given DNA concentration, indicating that the interaction between CENP-B DBD RP1 and CB21 was of the ‘fast exchange’ type on the NMR chemical shift timescale under these conditions. The cross-peaks due to Arg6, Gly29 and Thr41 became too broad to observe upon addition of CB21. Nine cross-peaks exhibited either $^{15}$N chemical shift changes larger than 0.25 p.p.m. or $^1$H chemical shift changes $>$0.05 p.p.m. (Arg5, Leu8, Asp25, Lys28, Ser43, Ile45, Leu46, Lys47 and Leu53). The 12 DNA-perturbed residues are mapped on the tertiary structure of CENP-B DBD RP1 in Figure 6A. It is likely that these residues of CENP-B DBD RP1 are located either in the DNA binding interface or in regions of DNA-induced conformation...
Fig. 5. Comparison of the structure of CENP-B DBD RP1 with those of the *fushi tarazu* homeodomain, Myb DBD R2 and RAP1 DBD domain 1. (A) Superimposition of CENP-B DBD RP1 (yellow) on the *fushi tarazu* homeodomain (magenta). The Cα atoms of the HTH region (from helix 2 to helix 3) were fitted between the two molecules. (B) Superimposition of CENP-B DBD RP1 (yellow) on Myb DBD R2 (blue). The Cα atoms of helices 1–3 were fitted. (C) Superimposition of CENP-B DBD RP1 (yellow) on RAP1 DBD domain 1 (green). The Cα atoms in helices 1–3 were fitted. These pictures were generated with the MidasPlus system (Ferrin et al., 1988; Huang et al., 1991).

Fig. 6. (A) Mapping of CENP-B DBD RP1 residues that exhibited chemical shift perturbations in the HSQC spectra upon addition of a 1.5 molar equivalent of CB21 DNA. The results are shown on ribbon representations of the CENP-B DBD RP1 solution structure. The colors used are blue for residues with 0.25 p.p.m. < |Δδ15N| < 0.50 p.p.m. or 0.05 p.p.m. < |Δδ1HN| < 0.10 p.p.m. and magenta for residues with 0.5 p.p.m. < |Δδ15N| or 0.1 p.p.m. < |Δδ1HN|. Residues with HSQC cross-peaks that became too broad to observe upon addition of CB21 DNA are colored red. This drawing was generated with the MidasPlus system (Ferrin et al., 1988; Huang et al., 1991). (B) Surface electrostatic potential of CENP-B DBD RP1, with the peptide backbone viewed inside. Positively and negatively charged areas are colored blue and red respectively. Orientation of the molecule is as in (A). This figure was produced by the program GRASP (Nicholls et al., 1991; Nicholls, 1993).

change. The most probable example of the latter case is Ile45, which is nearly completely buried inside the molecule (Figure 4C), while other DNA-perturbed residues are much more exposed. All of these DNA-perturbed residues, except for Leu53 in helix 4, are localized in three regions, the N-terminal arm and the N-terminal portions of helix 2 and helix 3. The same three regions of the homeodomains, the Myb repeats and the RAP1 domains, whose HTH folds superimpose well onto that of CENP-B DBD RP1, are actually in contact with DNA in the complex structures (Kissinger et al., 1990; Billeter et al., 1993; Ogata et al., 1994; König et al., 1996). In each case helix 3 contacts the major groove of the double-stranded DNA and the positively charged amino acid residues interact with the phosphate groups. In fact, in the electrostatic potential profile of CENP-B DBD RP1 (Figure 6B) several
positively charged areas surround helix 3. Therefore, it is likely that helix 3 contacts the major groove of the DNA, like other HTH DBDs.

**Base pairs recognized by CENP-B DBD RP1**

The chemical shift changes of the base paired imino proton resonances of the CB21 DNA (Figure 7) were analyzed upon addition of CENP-B DBD RP1. Eighteen imino proton resonances were assigned sequentially (Wüthrich, 1986) by 2D NOESY, whereas imino proton resonances of base pairs 1, 20 and 21 (G:C, T:A and T:A respectively) were not observed. Chemical shifts of the assigned imino proton resonances of the CB21 DNA in the presence and absence of CENP-B DBD RP1 were compared (Figure 7). Proton chemical shift changes >0.05 p.p.m. were observed for base pairs in the TCGTT sequence. This region overlaps with ‘site 1’ (TTCG), one of the essential sites in the CENP-B box for CENP-B binding (Figure 7). In contrast, such large chemical shift changes were not observed for sites 2 and 3 (Figure 7). Therefore, CENP-B DBD RP1 is concluded to interact with the 4 bp of site 1 in the CENP-B box sequence. Note that other HTH structures similar to CENP-B DBD RP1 can recognize up to 5 bp (Kissinger et al., 1990; Billeter et al., 1993; Ogata et al., 1994; König et al., 1996). Consequently, sites 2 and 3 are considered to be recognized by the rest of the CENP-B DBD, including RP2.

**A model of association between CENP-B and the CENP-B box**

In the present study we found that CENP-B DBD RP1 has an HTH structure, which is a typical structure found in DNA binding proteins. The distribution of positively charged areas surrounding helix 3 and the chemical shift perturbations of the amide $^{1}H$ and $^{15}N$ resonances, as mentioned above, suggest that helix 3 of CENP-B DBD RP1 contacts the major groove side of B-form DNA, like other HTH DBDs, such as homeodomains, Myb and RAP1. As judged from the imino proton chemical shift perturbations, CENP-B DBD RP1 is considered to recognize site 1 (TTCG) of the CENP-B box sequence. The other repeat, RP2, of the CENP-B DBD is quite homologous to RP1 (Figure 1C) and may therefore include an HTH structure similar to that of RP1. Thus CENP-B DBD RP2 is likely to bind with site 2 (A) and site 3 (CGGG) of the CENP-B box (Figure 1A), possibly with the N-terminal ‘ linker’ region and with the following HTH structure respectively (Suzuki et al., 1995).

On these assumptions we tried to build a preliminary docking model of the complex of the entire CENP-B DBD (RP1–RP2) and CENP-B box DNA (sites 1–3) in the B-form. Two opposite orientations of helix 3 relative to site 1 were possible; the vector of helix 3 (from the N- to the C-terminus) lies either in the same direction as that of the major groove vector from sites 1 to 3 (Figure 8A) or in the opposite direction (Figure 8B). However, RP2 can interact with site 3 only when the RP1 orientation is as shown in Figure 8A, because the linker between the two HTH structures is characteristically short. Therefore, we placed helix 3 of RP1 in contact with the major groove side of site 1 in the orientation shown in Figure 8A, using the MidasPlus program (Ferrin et al., 1988; Huang et al., 1991). As the structure of RP2 has not yet been determined, the structure of RP2 was modeled on that of RP1 and was then placed in contact with the major groove side of site 3. Again, one of the two possible orientations of RP2 helix 3 relative to site 3 was selected, to allow both RP1 and RP2 to bind to sites 1 and 3 respectively at the same time. The helix 3 vector of RP2 lies in the direction opposite from that of RP1 along the major groove (Figure 8A). The present docking model of the complex is shown in Figure 8C. In this model the angle between helices 3 and 4 of RP1 appears to be essential for both RP1 and RP2, with a characteristically short linker in between to interact with the CENP-B box DNA. The Arg57 residue in the linker included in RP2 is located on the minor groove side of the A:T base pair of site 2 in this model. Intriguingly, this residue corresponds with Arg6 in RP1, whose $^{1}H$ and $^{15}N$ resonances were largely perturbed upon addition of DNA. However, it should be noted here that the details of the intermolecular recognition mechanisms have yet to be experimentally elucidated.

The RAP1 and Myb proteins involve two HTH structures for recognition of base pairs from the major groove side (Ogata et al., 1994; König et al., 1996). In the case of Myb the two HTH structures (repeats R2 and R3) in contact with each other cover a relatively short range (8 bp) of the DNA and the vectors of their helices 3 are...
in the same direction along the major groove. On the other hand, the RAP1 DBD covers a 16 bp range of DNA; domains 1 and 2, with a substantially long linker in between, contact two separate regions of the major groove. In this case the directions of the two helix 3 vectors are the same along the major groove. Correspondingly, the RAP1 DBD recognizes a direct repeat of two identical 5 bp sequences. In contrast, in our model structure of the CENP-B DBD complexed with the CENP-B-box (17 bp) the two repeats, RP1 and RP2, are arranged to have opposite helix 3 directions (Figure 8A) and may be related by a pseudo-2-fold symmetry axis in between (Figure 8C). Interestingly, opposite helix 3 directions of two HTH structures are known for dimeric proteins, such as the λ and trp repressors, which recognize inverted repeat DNAs (Branden and Tooze, 1991). Actually, in the CENP-B box the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991). The DNA encoding CENP-B DBD RP1 (CENP-B:1–56) was amplified by PCR with pETCENP-B (Yoda et al., 1992) as the template DNA and was subcloned into expression vector pRSETA (Schoepfer, 1993).

**Materials and methods**

**Expression and purification of CENP-B DBD RP1**

The DNA encoding CENP-B DBD RP1 (CENP-B:1–56) was amplified by PCR with pETCENP-B (Yoda et al., 1992) as the template DNA and was subcloned into expression vector pRSETA (Schoepfer, 1993). *Escherichia coli* strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C and the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991). Actually, in the CENP-B box the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991). The DNA encoding CENP-B DBD RP1 (CENP-B:1–56) was amplified by PCR with pETCENP-B (Yoda et al., 1992) as the template DNA and was subcloned into expression vector pRSETA (Schoepfer, 1993).

**Expression and purification of CENP-B DBD RP1**

The DNA encoding CENP-B DBD RP1 (CENP-B:1–56) was amplified by PCR with pETCENP-B (Yoda et al., 1992) as the template DNA and was subcloned into expression vector pRSETA (Schoepfer, 1993).

The DNA encoding CENP-B DBD RP1 (CENP-B:1–56) was amplified by PCR with pETCENP-B (Yoda et al., 1992) as the template DNA and was subcloned into expression vector pRSETA (Schoepfer, 1993). *Escherichia coli* strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C and the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991). Actually, in the CENP-B box the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991). The DNA encoding CENP-B DBD RP1 (CENP-B:1–56) was amplified by PCR with pETCENP-B (Yoda et al., 1992) as the template DNA and was subcloned into expression vector pRSETA (Schoepfer, 1993). *Escherichia coli* strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C and the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991). Actually, in the CENP-B box the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991). The DNA encoding CENP-B DBD RP1 (CENP-B:1–56) was amplified by PCR with pETCENP-B (Yoda et al., 1992) as the template DNA and was subcloned into expression vector pRSETA (Schoepfer, 1993). *Escherichia coli* strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C strain BL21(DE3) pLysS was transformed with the resultant plasmid, pHRP1. The transformed strain was cultured at 37°C and the 4 bp sequences of sites 1 and 3 constitute a pseudo-inverted repeat, provided that TT is equivalent to CC (Branden and Tooze, 1991).
experiments used for resonance assignments of CENP-B DBD RP1 were 2D NOESY (Jeener et al., 1979; Macura and Ernst, 1980) with a mixing time of 120 ms, 2D DQF-COSY (Rance et al., 1983), 2D TOCSY (Bax and Davis, 1984) with a mixing time of 200 ms, 3D 1H-15N HMBC-J (Kay et al., 1993), 3D 1H-15N-NOESY-HSQC with a mixing time of 120 ms, 3D 1H-15N TOCSY-HSQC (Marion et al., 1989), 3D HNHB (Archer et al., 1991), 1H,15N NOE measurement (Grzesiek and Bax, 1993), 2D 1H-13C CT-HSQC, 3D CT-HNCA (Grzesiek and Bax, 1992), 3D HCC-H-TOCSY (Kay et al., 1993) and 3D 1H-13C NOESY-HSQC with a mixing time of 120 ms. The NMR data were processed using the NMRpipe system (Delaglio et al., 1995). The NMRView software package (Johnson and Blelins, 1994) was used to analyze the processed spectra.

**Structure calculations**

The solution structure of CENP-B DBD RP1 was calculated using X-PLOR v.3.1 (Brünger, 1993) on DEC alpha workstations. The structure calculation was performed only for residues 1–56 of CENP-B, because the experimental results suggested that the vector-derived residues containing the 6–8 His tag are unstructured. The NOE intensities between the protons were assigned from the 2D NOESY, 3D 1H-15N NOESY-HSQC and 3D 1H-13C NOESY-HSQC spectra. The NOE intensities were classified into three classes, i.e., strong, medium and weak, and were translated into distance constraints in the ranges 1.8–3.0, 1.8–4.0 and 1.8–5.0 Å respectively. In total, 626 distance restraints (466 inter-residue restraints) derived from the NOEs were used for the structural calculation. The dihedral angle constraints were for 24 φ and χ angles as measured from the 2D HMOC-J spectrum, and 12 γ angles. The backbone hydrogen bond restraints in the α-helices were included for residues that lacked cross-peaks between HN and H2O in the 3D 1H-15N NOESY-HSQC spectrum, as measured by the water flip-back method. The 21 hydrogen bonds in the α-helices were characterized by two distance restraints: 2.1–2.3 Å between HN and O and 2.1–3.3 Å between N and O. Eighty structures were generated using the ab initio dynamic simulated annealing protocol (Nilges et al., 1988). The structures with proper three-dimensional folds were further refined using the SA refinement protocol. The atomic coordinates of the 20 refined structures and the restrained energy minimized averaged structure will be deposited in the Brookhaven Protein Data Bank.

**Determination of the minimum DNA fragment for CENP-B DBD binding**

We performed a gel mobility shift analysis to determine the minimum DNA fragment that contains the CENP-B box sequence and binds the entire CENP-B DBD (RP1–RP2). Chemically synthesized DNA fragments of 21 and 25 bp (CMb19 and CMb25, respectively), in which the sequence of the CENP-B box was centered, were labeled at the 5′-termini using T4 polynucleotide kinase (Takara) and [γ-32P]ATP. CENP-B DBD RP1–RP2 (residues 1–129 of CENP-B) was expressed from pETCBN129 (Kitagawa et al., 1995) in *E.coli* BL21 (DE3) and was purified by Butyl-Toyopearl (Toso) and P11 phosphocellulose (Whatman) chromatography. Binding of the 32P-labeled DNA fragment and the CENP-B DBD was analyzed according to Kitagawa et al. (1995).

**Synthesis and purification of CB21 DNA**

Each strand of the CB21 DNA (5′-GCGTCCAACAGGCACAGGC-G-GAGCT-3′ and 5′-AACCGCTTGGGACGCCAGGACGGT-3′) was synthesized using a Perkin Elmer synthesizer (Millipore) on the 1.0 μmol scale. The synthesized single-stranded DNA was purified together with the protecting DMT group by HPLC on a C18 reverse phase column and, after removal of the DMT group, was purified again by the same HPLC method (Brown and Brown, 1992). Equimolar amounts of the two DNA strands, in 10 mM Tri-HCl buffer, pH 7.5, containing 1 mM EDTA and 150 mM NaCl, were annealed by heating to 85°C for 2 h.

**Mapping of interaction sites by NMR spectroscopy**

The chemical shift perturbations that occurred upon mixing of CENP-B DBD RP1 and CB21 DNA were measured at 30°C in 20 mM phosphate buffer (pH 6.0), containing 250 mM NaSO4, 0.25 mM EDTA and 0.01% NaN3. To map the DNA binding sites on CENP-B DBD RP1, perturbations of the backbone 1H and 15N chemical shifts that occurred upon addition of CB21 DNA were measured for the 2D 1H-15N HSQC spectra of 15N-labeled CENP-B DBD RP1 at DNA:protein molar ratios of 0.038, 0.75, 1.13 and 1.50. To map the protein binding sites on the CENP-B box DNA the changes in the imino 1H chemical shifts of the CB21 DNA caused by addition of CENP-B DBD RP1 were measured from the 2D NOESY spectra of the CB21 DNA in the presence and absence of a 1:33 molar equivalent of CENP-B DBD RP1.

**Acknowledgements**

We are grateful to Drs K.Takio and N.Domae for the mass spectrometry and to Dr PKönig for providing the coordinates of the yeast RAP1 DNA binding domain complexed with telomeric DNA before publication by the Brookhaven Protein Data Bank. This work was supported in part by a grant for the Biodesign Research Program from the Riken.

**References**


Acknowl.


A helix-turn-helix structure unit in CENP-B


Received August 7, 1997; revised November 14, 1997; accepted November 17, 1997