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

Junji Iwahara^{1,2}, Takanori Kigawa²,
Katsumi Kitagawa³, Hiroshi Masumoto³,
Tuneko Okazaki^{3,4} and
Shigeyuki Yokoyama^{1,2,5}

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 character-

ized 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

Fig. 1. (A) Human CENP-B box sequence. The base pairs recognized by CENP-B, elucidated by *in vitro* binding experiments using CENP-B and variants of the CENP-B box (Masumoto *et al.*, 1993) are boxed. U and Y represent purine (A or G) and pyrimidine (C or T) respectively. (B) Base sequence alignment of CENP-B binding sites among human, *Mus musculus*, *Mus caroli* and African green monkey (AGM). The CENP-B binding sites of *Mus musculus* and *Mus caroli* are present in the centromere-specific repeating sequences, which are composed of monomer units of 120 and 79 bp respectively (Masumoto *et al.*, 1989; Kipling *et al.*, 1995). The CENP-B binding sites of African green monkey were found in the α -satellite DNA (Yoda *et al.*, 1996), but the number of CENP-B binding sites in the α -satellite DNA is much smaller in African green monkey than in humans (Goldberg *et al.*, 1996). (C) Amino acid sequence alignment between CENP-B DBD RP1 and RP2. Identical residues are boxed and similar residues are bridged with bars.

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 *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.

Results and discussion

The first repeat (RP1) of the CENP-B DBD was expressed under control of the T7 promoter in *Escherichia coli* as a polypeptide with 14 N-terminal vector-derived residues, including a 6 $\times$ 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 Na₂SO₄, 0.25 mM EDTA and 0.01% NaN₃ at pH 6.0 (data not shown). These results indicate that CENP-B DBD RP1 itself is a well-structured unit.

Resonance 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 Na₂SO₄, 0.25 mM EDTA and 0.01% NaN₃ at pH 6.0. Figure 2A shows the 2D ¹H-¹⁵N HSQC spectrum of CENP-B DBD RP1. First, ~80% of the ¹H and ¹⁵N resonances of CENP-B DBD RP1 were assigned by use of 3D ¹H-¹⁵N NOESY-HSQC, 3D ¹H-¹⁵N TOCSY-HSQC, 3D HNHB, 2D ¹H-¹⁵N HSQC, 2D NOESY, 2D TOCSY and 2D DQF-COSY spectra. We completed the resonance assignments using the 3D CT-HNCA, 3D CT-HN(CO)CA, 3D HCCH-TOCSY and 3D ¹H-¹³C NOESY-HSQC spectra. Figure 2B shows an example of sequential assignment along the Leu42–Leu53 backbone using the 3D CT-HNCA spectrum.

Fig. 2. (A) The ^1H - ^{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_2SO_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 $\{^1\text{H}\}$ - ^{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 $^3J_{\text{HNH}\alpha}$ coupling constants, observation of $\text{HN-H}_2\text{O}$ cross-peaks, $^{13}\text{C}_\alpha$ chemical shift indices and $\{^1\text{H}\}$ - ^{15}N NOE values. For calculation of tertiary structure inter-proton NOE information was obtained from the 3D ^1H - ^{15}N NOESY-HSQC, 3D ^1H - ^{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 α -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

Fig. 3. The secondary structure elements of CENP-B DBD RP1 and a summary of the sequential and medium range NOEs, $^3J_{\text{HNH}\alpha}$ coupling constants, presence of HN-H₂O cross-peaks, $^{13}\text{C}_\alpha$ chemical shift indices and $\{^1\text{H}\}$ - ^{15}N NOE values. Δ , a $^3J_{\text{HNH}\alpha}$ coupling constant >7 Hz. Residues that exhibited a cross-peak between HN and H₂O resonances in the 3D ^1H - ^{15}N NOESY-HSQC spectrum are indicated by $\blacklozenge$. Dotted lines indicate the NOE connectivities in which one resonance overlapped with another. The $\star$ in the bar graph of the $\{^1\text{H}\}$ - ^{15}N NOE values indicates a proline residue.

RP1 does not exhibit any structural similarity to the helix-loop-helix domain of MyoD complexed with DNA (Ma *et al.*, 1994), in spite of the partial sequence similarity (Sullivan and Glass, 1991).

The HTH region (from helix 2 to helix 3) and helix 1 of CENP-B DBD RP1 form a hydrophobic core with the sidechains of Ile16, Val20, Ile31, Phe35 and Leu42 (Figure 4C). The Thr9-Phe10-Arg11-Glu12 sequence at the N-terminus of helix 1 forms a helix-stabilizing T-X-X-E type 'capping box' (Harper and Rose, 1993; Gronenborn and Clore, 1994) with a hydrogen bond between the amide group of Thr9 and the γ -carboxyl group of Glu12, which is reflected in the characteristic pattern of $^{13}\text{C}_\alpha$ chemical shifts (Figure 3) and NOEs. The sidechain of Leu8 forms a hydrophobic interaction with those of Lys13, Ile37 and Thr41. Consequently the region from Arg5 to Thr9 is structured near the turn between helix 2 and helix 3 (Figure 4B) and therefore corresponds to the 'N-terminal arm' in other eukaryotic DNA binding proteins with a HTH motif (Qian *et al.*, 1989; Hirsch and Aggarwal, 1995; König *et al.*, 1996).

The tertiary structure from the N-terminus to helix 3 of CENP-B DBD RP1 is quite similar to those of other eukaryotic HTH domains, such as homeodomains, the Myb DBD repeating units (Ogata *et al.*, 1992, 1994) and the RAP1 DBDs (König *et al.*, 1996). Figure 5 shows superimposition of the structures of these DBDs on that of CENP-B DBD RP1. The HTH region (helix 2, the turn and helix 3) of CENP-B DBD RP1 is best superimposed on that of the *fushi tarazu* homeodomain (Qian *et al.*, 1994); the r.m.s. deviation value of the C_α atoms is as

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 C_α 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- C_α bond of Asn48 (with a dihedral angle $\phi = -97^\circ$ for the minimized average structure). Long range ^1H - ^1H 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 $^3J_{\text{HNH}\alpha}$ value >8 Hz, whereas other residues exhibited $^3J_{\text{HNH}\alpha}$ values <5 Hz (Figure 3). Second, all of the residues from Ser40 to Ile52 except Asn48 had positive $^{13}\text{C}_\alpha$ 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 $\{^1\text{H}\}$ - ^{15}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.

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

X-PLOR energies (kcal/mol)	
$E_{\text{van der Waals}}^a$	26.9 ± 5.1
$E_{\text{Lennard-Jones}}^b$	-118.7 ± 29.4
r.m.s. deviation from experimental restraints	
Distance (Å)	0.037 ± 0.003
Dihedral angle (degree)	0.29 ± 0.25
Deviations from ideal covalent geometry used within X-PLOR	
Bonds (Å)	0.0034 ± 0.0002
Angles (degree)	0.61 ± 0.02
Improvers (degree)	0.45 ± 0.02

^aThe quadratic van der Waals term was calculated with a force constant of 4 kcal/mol/Å⁴ and the van der Waals radii were set to 0.75 times the standard value used in CHARMM.

^bLennard-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. ¹⁵N-labeled CENP-B DBD RP1 was mixed with CB21 under conditions of 250 mM Na₂SO₄, 20 mM potassium phosphate, 0.25 mM EDTA and 0.01% NaN₃, 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 ¹H-¹⁵N HSQC spectra were measured at different molar ratios (0, 0.38, 0.75, 1.13 and 1.50) of CB21 to ¹⁵N-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 ¹H-¹⁵N 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 ¹⁵N chemical shift changes larger than 0.25 p.p.m. or ¹H 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 C_{α} 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 C_{α} atoms of helices 1–3 were fitted. (C) Superimposition of CENP-B DBD RP1 (yellow) on RAP1 DBD domain 1 (green). The C_{α} 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 $0.25 \text{ p.p.m.} < |\Delta\delta_{15N}| < 0.50 \text{ p.p.m.}$ or $0.05 \text{ p.p.m.} < |\Delta\delta_{1HN}| < 0.10 \text{ p.p.m.}$ and magenta for residues with $0.5 \text{ p.p.m.} < |\Delta\delta_{15N}|$ or $0.1 \text{ p.p.m.} < |\Delta\delta_{1HN}|$, where $\Delta\delta_{15N}$ and $\Delta\delta_{1HN}$ represent the ^{15}N and ^1H chemical shift differences between the absence and presence of CB21 DNA. 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

Fig. 7. Chemical shift changes in the base pair imino protons of CB21 DNA upon addition of CENP-B DBD RP1. Each value is the absolute difference between in the absence and presence of a 1.33 molar equivalent of CENP-B DBD RP1. The imino proton resonances that were too broad to observe are indicated by asterisks. The NMR samples were measured in 20 mM potassium phosphate buffer, pH 6.0, containing 250 mM Na₂SO₄, 0.25 mM EDTA and 0.01% NaN₃.

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' (TTCCG), 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 ¹H and ¹⁵N 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 (TTCCG) 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 ¹H and ¹⁵N 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

Fig. 8. (A and B) Two possible orientations of CENP-B DBD RP1 on site 1 of the CENP-B box DNA. Helix 3 of RP1 is suggested to contact the major groove side of site 1 in this study. The large arrows in cylinders indicate the helix 3 vectors (from the N- to the C-terminus). RP2, which is linked to RP1 by a relatively short linker, can contact site 3 in (A) but not in (B). (C) A model of the complex of the entire CENP-B DBD (RP1–RP2) and the CENP-B box DNA (sites 1–3) in the B-form. A B-form structure of the CENP-B box DNA and a structural similarity between RP1 and RP2 are postulated. RP1 and RP2 were placed at sites 1 and 3 respectively, so that each helix 3 contacts the major groove of the DNA. The helix 3 vectors of RP1 and RP2 are opposite to each other along the major groove. The colors of the RP1 helices are as in Figure 4. The model building and the graphics drawing were performed with the MidasPlus system (Ferrin *et al.*, 1988; Huang *et al.*, 1991).

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 (Figure 8C).

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 in 1.5 l M9 medium supplemented with ampicillin (100 µg/ml) and chloramphenicol (34 µg/ml). For preparation of ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ -labeled CENP-B DBD RP1 samples, 1 g/l $^{15}\text{NH}_4\text{Cl}$ and/or 2 g/l $[^{13}\text{C}_6]\text{D-glucose}$ were used as the sole nitrogen and/or carbon sources. When the cell density reached an OD_{600} of 0.9, isopropyl β -D-thiogalactopyranoside was added to a final concentration of 1.2 mM to induce expression of

CENP-B DBD RP1. Cells were harvested 3 h after induction and were lysed by sonication in 20 mM potassium phosphate buffer, pH 8.0, containing 300 mM NaCl. After centrifugation of the lysate at 12 000 g the supernatant was applied to a 4 ml Ni-NTA (Qiagen) column, which was washed with 20 mM potassium phosphate buffer, pH 6.0, containing 300 mM NaCl (buffer B). CENP-B DBD RP1 was eluted with a gradient of imidazole of 0–500 mM in buffer B. The fractions that contained CENP-B DBD RP1 were pooled and loaded onto a 5 ml P11 phosphocellulose (Whatman) column equilibrated with buffer B. After a wash with 250 mM potassium phosphate buffer, pH 7.0, containing 300 mM NaCl, a gradient of potassium phosphate buffer, pH 7.0, of 250–750 mM in 300 mM NaCl was used to elute CENP-B DBD RP1. The molecular mass of the purified polypeptide was measured by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and was confirmed to agree with the expected value. The buffer used to purify CENP-B DBD RP1 was exchanged for NMR spectroscopy using a Centricon-3 (Amicon) 3 kDa cut-off ultrafiltration membrane.

CD analysis of CENP-B DBD RP1

CD measurements of purified CENP-B DBD RP1 were performed with a Jasco J-720 spectropolarimeter. The protein concentration was 13 µM in 20 mM potassium phosphate buffer, pH 6.0, containing 400 mM Na_2SO_4 , 0.25 mM EDTA and 0.01% NaN_3 . The CD spectrum of CENP-B DBD RP1 was measured at 20°C in the wavelength range 195–255 nm with a 1.0 mm path length cell. The α -helix content of CENP-B DBD RP1 was estimated from CD spectra at 208 nm (Greenfield and Fasman, 1969). The melting curve of CENP-B DBD RP1 was recorded at 222 nm using a 1.0 mm path length cell in the temperature range 20–80°C.

NMR spectroscopy for structure determination of CENP-B DBD RP1

The sample solutions for determination of CENP-B DBD RP1 structure contained 1.5 mM protein in 20 mM potassium phosphate buffer, pH 6.0, containing 400 mM Na_2SO_4 , 0.25 mM EDTA, 0.01% NaN_3 and 10% $^2\text{H}_2\text{O}$. NMR measurements were performed at 30°C on either a Bruker AMX-600 or DMX-500 spectrometer equipped with a pulsed-field gradient unit. The $^1\text{H}_2\text{O}$ signal was eliminated by use of the z -axis gradient in the WATERGATE technique (Piotto *et al.*, 1992). The NMR

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, 1985) with a mixing time of 65 ms, 2D ^1H - ^{15}N HSQC (Bodenhausen and Ruben, 1980), 2D ^1H - ^{15}N HMQC-J (Kay and Bax, 1990), 3D ^1H - ^{15}N NOESY-HSQC with a mixing time of 120 ms, 3D ^1H - ^{15}N TOCSY-HSQC (Marion *et al.*, 1989), 3D HNHB (Archer *et al.*, 1991), $\{^1\text{H}\}$ - ^{15}N NOE measurement (Grzesiek and Bax, 1993), 2D ^1H - ^{13}C CT-HSQC, 3D CT-HNCA (Grzesiek and Bax, 1992), 3D CT-HN(CO)CA (Grzesiek and Bax, 1992), 3D HCCH-TOCSY (Kay *et al.*, 1993) and 3D ^1H - ^{13}C 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 Blevins, 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 $\times$ His tag are unstructured. The NOEs between the protons were assigned from the 2D NOESY, 3D ^1H - ^{15}N NOESY-HSQC and 3D ^1H - ^{13}C 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 angles based on $^3J_{\text{HNH}\alpha} < 5$ Hz, as measured from the 2D HMQC-J spectrum, and 12 χ_1 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 - ^{15}N 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 17, 19, 21 and 25 bp (CB17, CB19, CB21 and CB25 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 - ^{32}P]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) column chromatography. Binding of the ^{32}P -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'-GCCTTCGTTGGAACGGGATT-3' and 5'-AATCCCGTTTCCAACGAAGGC-3'), with a 4,4'-dimethoxytrityl (DMT) group at the 5'-terminus, was synthesized using an Expedite DNA 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 Tris–HCl buffer, pH 7.5, containing 1 mM EDTA and 150 mM NaCl, were annealed by cooling from 85°C to room temperature over 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 potassium phosphate buffer, pH 6.0, containing 250 mM Na₂SO₄, 0.25 mM EDTA and 0.01% Na₂N₃. To map the DNA binding sites on CENP-B DBD RP1, perturbations of the backbone ^1H and ^{15}N chemical shifts that occurred upon addition of CB21 DNA were measured for the 2D ^1H - ^{15}N HSQC spectra of ^{15}N -labeled CENP-B DBD RP1 at DNA:protein molar ratios of 0, 0.38, 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. Assignment of the imino ^1H resonances of each sample were obtained from the 2D NOESY spectra.

Acknowledgements

We are grateful to Drs K.Takio and N.Domae for the mass spectrometry and to Dr P.König 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 Riken.

References

- Archer, S.J., Ikura, M., Torchia, D.A. and Bax, A. (1991) An alternative 3D NMR technique for correlating backbone ^{15}N with side chain ^1H resonances in larger proteins. *J. Magn. Resonance*, **95**, 636–641.
- Bax, A. and Davis, D.G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Resonance*, **65**, 355–360.
- Bejarano, L.A. and Valdivia, M.M. (1996) Molecular cloning of an intronless gene for the hamster centromere antigen CENP-B. *Biochim. Biophys. Acta Gene Struct. Expression*, **1307**, 21–25.
- Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wuthrich, K. (1993) Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain–DNA complex. *J. Mol. Biol.*, **234**, 1084–1093.
- Bodenhausen, G. and Ruben, D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69**, 185–189.
- Branden, C. and Tooze, J. (1991) *Introduction to Protein Structure*. Garland Publishing, New York, NY.
- Brown, T. and Brown, D.J. (1992) Purification of synthetic DNA. *Methods Enzymol.*, **211**, 20–35.
- Brünger, A.T. (1993) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
- Clarke, L. (1990) Centromeres of budding and fission yeasts. *Trends Genet.*, **6**, 150–154.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
- Earnshaw, W.C. and Rothfield, N. (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*, **91**, 313–21.
- Ferre, D.A.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.
- Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R. (1988) The MIDAS display system. *J. Mol. Graphics*, **6**, 13–27.
- Goldberg, I.G., Sawhney, H., Pluta, A.F., Warburton, P.E. and Earnshaw, W.C. (1996) Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: implications for CENP-B function at centromeres. *Mol. Cell. Biol.*, **16**, 5156–5168.
- Greenfield, N. and Fasman, G.D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, **8**, 4108–4116.
- Gronenborn, A.M. and Clore, G.M. (1994) Identification of N-terminal helix capping boxes by means of ^{13}C chemical shifts. *J. Biomol. NMR*, **4**, 455–458.
- Grzesiek, S. and Bax, A. (1992) Improved three-dimensional triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Resonance*, **96**, 432–440.
- Grzesiek, S. and Bax, A. (1993) The importance of not saturating H₂O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.*, **115**, 12593–12594.
- Halverson, D., Baum, M., Stryker, J., Carbon, J. and Clarke, L. (1997) A centromere DNA-binding protein from fission yeast affects chromosome segregation and has homology to human CENP-B. *J. Cell Biol.*, **136**, 487–500.
- Harper, E.T. and Rose, G.D. (1993) Helix stop signals in proteins and peptides: the capping box. *Biochemistry*, **32**, 7605–7609.
- Hirsch, J.A. and Aggarwal, A.K. (1995) Structure of the even-skipped

- homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J.*, **14**, 6280–6291.
- Huang, C.C., Pettersen, E.F., Klein, T.E., Ferrin, T.E. and Langridge, R. (1991) Conic: a fast renderer for space-filling molecules with shadows. *J. Mol. Graphics*, **9**, 230–236.
- Ikeno, M., Masumoto, H. and Okazaki, T. (1994) Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range alpha-satellite DNA arrays of human chromosome 21. *Hum. Mol. Genet.*, **3**, 1245–1257.
- Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.*, **71**, 4546–4553.
- Johnson, B.A. and Blevins, R.A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
- Kay, L.E. and Bax, A. (1990) New methods for the measurement of NH- α CH coupling constants in ^{15}N labeled proteins. *J. Magn. Resonance*, **86**, 110–126.
- Kay, L.E., Xu, G.-Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.D. (1993) A gradient enhanced HCCH-TOCSY experiment for recording side-chain ^1H and ^{13}C correlations in H_2O samples of proteins. *J. Magn. Resonance (B)*, **101**, 333–337.
- Kipling, D., Mitchell, A.R., Masumoto, H., Wilson, H.E., Nicol, L. and Cooke, H.J. (1995) CENP-B binds a novel centromeric sequence in the Asian mouse *Mus caroli*. *Mol. Cell. Biol.*, **15**, 4009–4020.
- Kissinger, C.R., Liu, B.S., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. *Cell*, **63**, 579–590.
- Kitagawa, K., Masumoto, H., Ikeda, M. and Okazaki, T. (1995) Analysis of protein–DNA and protein–protein interactions of centromere protein B (CENP-B) and properties of the DNA–CENP-B complex in the cell cycle. *Mol. Cell. Biol.*, **15**, 1602–1612.
- König, P., Giraldo, R., Chapman, L. and Rhodes, D. (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell*, **85**, 125–136.
- Kraulis, P.J. (1991) Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Lee, J.K., Huberman, J.A. and Hurwitz, J. (1997) Purification and characterization of a CENP-B homologue protein that binds to the centromeric K-type repeat DNA of *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA*, **94**, 8427–8432.
- Ma, P.C., 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**, 451–9.
- Macura, S. and Ernst, R.R. (1980) Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.*, **41**, 95–117.
- Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) Overcoming the overlap problem in the assignment of ^1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^1H – ^{15}N Hartmann–Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 β . *Biochemistry*, **28**, 6150–6156.
- Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell Biol.*, **109**, 1963–1973.
- Masumoto, H., Yoda, K., Ikeno, M., Kitagawa, K., Muro, Y. and Okazaki, T. (1993) *Properties of CENP-B and its Target Sequence in a Satellite DNA*. Springer-Verlag, Berlin, Germany.
- Murakami, Y., Huberman, J.A. and Hurwitz, J. (1996) Identification, purification, and molecular cloning of autonomously replicating sequence-binding protein 1 from fission yeast *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA*, **93**, 502–507.
- Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M. and Okazaki, T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.*, **116**, 585–596.
- Ngan, V.K. and Clarke, L. (1997) The centromere enhancer mediates centromere activation in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.*, **17**, 3305–3314.
- Nicholls, A. (1993) *GRASP: Graphical Representation and Analysis of Surface Properties*. Columbia University, New York, NY.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Nicklas, R.B., Ward, S.C. and Gorbisky, G.J. (1995) Kinetochore chemistry is sensitive to tension and may link mitotic forces to a cell cycle checkpoint. *J. Cell Biol.*, **130**, 929–939.
- Nilges, M., Gronenborn, A.M. and Clore, G.M. (1988) Determination of 3-dimensional structures of proteins by simulated annealing with interproton distance restraints—application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor-2. *Protein Engng*, **2**, 27–38.
- Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S. and Nishimura, Y. (1992) Solution structure of a DNA-binding unit of Myb: a helix–turn–helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl Acad. Sci. USA*, **89**, 6428–6432.
- Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, **79**, 639–648.
- Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H. and Margolis, R.L. (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl Acad. Sci. USA*, **88**, 3734–3738.
- Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
- Pluta, A.F., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G. and Earnshaw, W.C. (1995) The centromere: hub of chromosomal activities. *Science*, **280**, 1591–1594.
- Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J. and Wüthrich, K. (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580.
- Qian, Y.Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Muller, M., Gehring, W.J. and Wüthrich, K. (1994) Nuclear magnetic resonance solution structure of the fushi tarazu homeodomain from *Drosophila* and comparison with the Antennapedia homeodomain. *J. Mol. Biol.*, **238**, 333–345.
- Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) Improved spectral resolution in COSY ^1H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.*, **117**, 479–485.
- Rieder, C.L., Cole, R.W., Khodjakov, A. and Sluder, G. (1995) The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J. Cell Biol.*, **130**, 941–948.
- Romanova, L.Y., Deriagin, G.V., Mashkova, T.D., Tumeneva, I.G., Mushegian, A.R., Kisselev, L.L. and Alexandrov, I.A. (1996) Evidence for selection in evolution of alpha satellite DNA: the central role of CENP-B/pJ α binding region. *J. Mol. Biol.*, **261**, 334–340.
- Schoepfer, R. (1993) The pRSET family of T7 promoter expression vectors for *Escherichia coli*. *Gene*, **124**, 83–85.
- Sugimoto, K., Yata, H., Muro, Y. and Himeno, M. (1994) Human centromere protein C (CENP-C) is a DNA-binding protein which possesses a novel DNA-binding motif. *J. Biochem. (Tokyo)*, **116**, 877–881.
- Sullivan, K.F. and Glass, C.A. (1991) CENP-B is a highly conserved mammalian centromere protein with homology to the helix–loop–helix family of proteins. *Chromosoma*, **100**, 360–370.
- Sullivan, K.F., Hechenberger, M. and Masri, K. (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.*, **127**, 581–592.
- Suzuki, M. and Brenner, S.E. (1995) Classification of multi-helical DNA-binding domains and application to predict the DBD structures of sigma factor, LysR, OmpR/PhoB, CENP-B, RapI, and XyIS/Ada/AraC. *FEBS Lett.*, **372**, 215–221.
- Suzuki, M., Masumoto, H. and Okazaki, T. (1995) Prediction of the three dimensional structure of the DBD of centromere protein B (CENP-B) in comparison with the structures of Myb, LexA and DtxR DBDs. *Proc. Jpn. Acad.*, **71**, 153–157.
- Willard, H.F. (1990) Centromeres of mammalian chromosomes. *Trends Genet.*, **6**, 410–416.
- Wishart, D.S. and Sykes, B.D. (1994) The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**, 171–180.
- Wüthrich (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York, NY.

- Yoda,K. and Okazaki,T. (1997) Site-specific base deletions in human α -satellite monomer DNAs are associated with regularly distributed CENP-B boxes. *Chromosome Res.*, **5**, 207–211.
- Yoda,K., Kitagawa,K., Masumoto,H., Muro,Y. and Okazaki,T. (1992) A human centromere protein, CENP-B, has a DNA binding domain containing four potential alpha helices at the NH₂ terminus, which is separable from dimerizing activity. *J. Cell Biol.*, **119**, 1413–1427.
- Yoda,K., Nakamura,T., Masumoto,H., Suzuki,N., Kitagawa,K., Nakano,M., Shinjo,A. and Okazaki,T. (1996) Centromere protein B of African green monkey cells: gene structure, cellular expression, and centromeric localization. *Mol. Cell. Biol.*, **16**, 5169–5177.

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