Secondary structure and folding topology of the DNA binding domain of interferon regulatory factor 2, as revealed by NMR spectroscopy

Koichi Uegaki[a], Masahiro Shirakawa[a], Hisashi Harada[b], Tadatsugu Taniguchi[b], Yoshimasa Kyogoku[a,*]

[a]*Institute for Protein Research, Osaka University, Suita, Osaka 565, Japan*
[b]*Institute for Molecular and Cellular Biology, Osaka University, Suita, Osaka 565, Japan*

Received 1 December 1994; revised version received 9 January 1995

Abstract The secondary structure elements of the DNA-binding domain of mouse interferon regulatory factor 2 [IRF-2(113)] were determined by heteronuclear multidimensional NMR spectroscopy. The sequential NOE connectivities, amide proton exchange rates, and $^3J_{HN\alpha}$ coupling constants indicated the presence of three α-helical regions and four short β-strands connected through relatively long loops. The long range NOEs indicated the four strands form an antiparallel β-sheet and the three α-helices form a bundle on the sheet. The arrangement of the secondary structure elements and the overall folding topology resemble those of the DNA binding domains of bacterial activator CAP, heat shock transcription factors, and fork-head proteins, although there is no sequence homology among them.

Key words: Interferon regulatory factor; Nuclear magnetic resonance; Secondary structure; DNA binding motif; Folding topology

1. Introduction

Interferon regulatory factors were originally identified by Fujita et al. [1]. Two DNA-binding factors are known to regulate the expression of the type I interferon (IFN) and IFN-inducible genes. These are interferon regulatory factor-1 (IRF-1), a transcriptional activator, and IRF-2, its antagonistic repressor [2,3]. Recently, two other IRF-related transcriptional factors were cloned [4,5]. Each factor binds to the specific target promoter sequences of the IFN-α and -β genes as well as IFN-inducible genes. These factors are structurally similar, particularly in the amino-terminal region, which confers the DNA binding character. The DNA binding domain of IRF-2 is located in the N-terminal region, which consists of 112 amino acid residues, and binds to the consensus base sequence [6]. The amino acid sequence of this DNA binding domain shows repeated proline residues, but does not show apparent homology with other transcription factors containing known DNA binding motifs (helix-turn-helix (HTH), basic-leucine zipper (bZip), Zn-finger, etc.). Therefore, interferon regulatory factors may have a unique structure which is characteristic of the IRF family.

In this communication, we report determination of the secondary structure of the IRF-2(113) protein and the folding topology of the secondary structure determined by heteronuclear multidimensional NMR analysis. The determined secondary structure elements do not contain a typical HTH motif, but the overall folding topology is that of the α + β HTH-like structure family. This is the first experimental evidence regarding the structure of an interferon regulatory factor.

2. Materials and methods

2.1. Sample preparation

Uniformly ^{15}N-labeled recombinant mouse IRF-2(113) was expressed in BL21(DE3) in M9 medium with $^{15}NH_4Cl$ (0.5 g/l) as the sole nitrogen source, and the uniformly $^{15}N/^{13}C$-doubly labeled protein was expressed in M9 medium with $^{15}NH_4Cl$ (0.5 g/l) and [$^{13}C_6$]D-glucose (1 g/l) as the nitrogen and carbon sources. The purification procedure was described in the previous paper [6]. The samples for NMR measurements comprised 1–3 mM protein in 90% H_2O/10% D_2O, or in D_2O, containing 10 mM phosphate buffer and 50 mM KCl, pH 5.5 (direct meter reading).

For titration with DNA, two strands of 12 bp DNA (dGAAAGTGAAAGT and dACTTTCACTTTC) were synthesized to form a duplex, which corresponds to the consensus sequence of the binding site [7].

2.2. NMR measurements

All NMR spectra were recorded at 25°C, pH 5.5, using a Bruker AMX-500 spectrometer equipped with an H-X inverse probe or a triple-resonance $^1H(^{15}N, ^{13}C)$ probe for ^1H detection. Proton chemical shifts were referenced to the water signal at 25°C (4.75 ppm), and the ^{15}N and ^{13}C chemical shifts were indirectly referenced to $HCO^{15}NH$ and ^{13}C-aspartic acid.

The 2D ^1H-^{15}N HSQC [8,9,10], 3D HMQC-NOESY-HMQC [11], 3D ^1H-^{15}N TOCSY-HSQC [12], and ^1H-^{15}N NOESY-HSQC [12] spectra were recorded for an ^{15}N-labeled protein. The 3D ^1H-^{13}C NOESY-HSQC [13] experiment was performed for a uniformly $^{15}N/^{13}C$ labeled protein. 3D CT-HNCA, CT-HN(CO)CA, CT-HNCO, and CT-HN(CA)CO triple-resonance experiments were performed in essentially the same way as described by Grzsiek and Bax [14]. CT-HCACO triple-resonance experiments were done in the same way as described by Powers et al. [15].

For measurement of the coupling constant, $^3J_{HN\alpha}$, we employed the ^1H-^{15}N HMQC-J experiment [16]. In order to determine the exchange rates of the amide protons of individual residues, we measured the ^1H-^{15}N HSQC spectra of the ^{15}N-uniformly labeled IRF-2(113) protein in a D_2O solution at pH 5.5 and 25°C.

3. Results

Sequential assignments for the backbone atom were completed mainly through the combined use of the HNCA, HN(CO)CA, HCACO, HNCO and HN(CA)CO experiments. The side chain assignments were performed by means of TOCSY-HSQC, but the aliphatic ^1H resonances of many residues have not yet been assigned because extensive chemical shift overlapping occured in the finger print region. The ob-

*Corresponding author. Fax 81-6-879-8599.

SSDI 0014-5793(95)00040-2

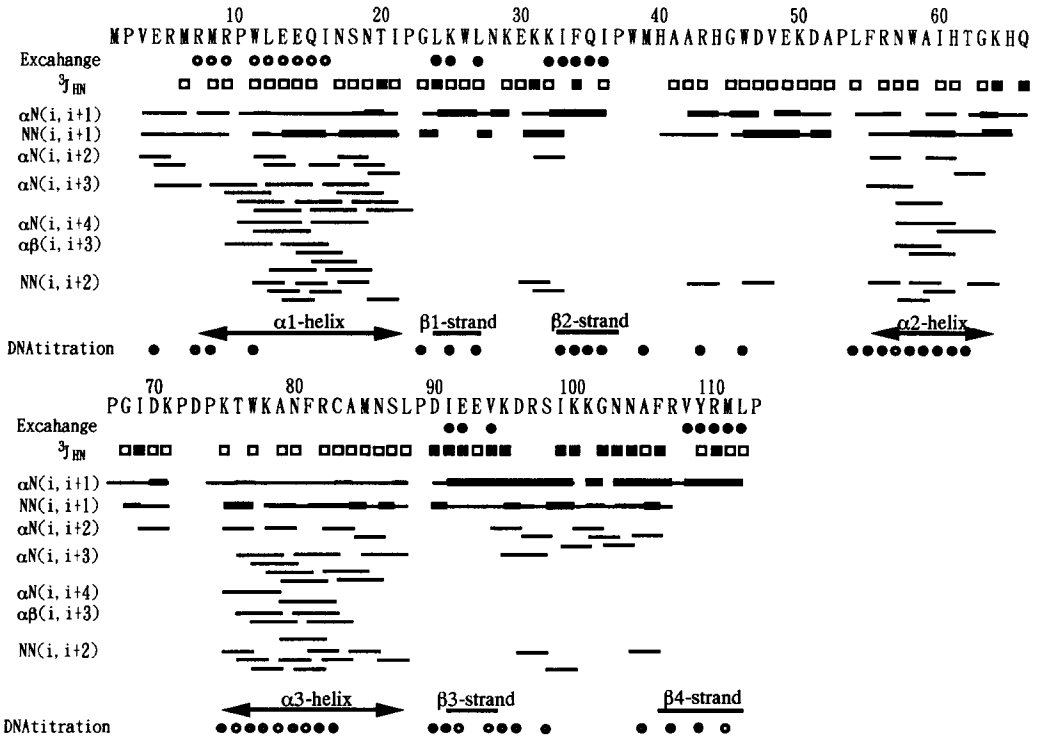

Fig. 1. Summary of the sequential and medium range NOEs between the HN, CαH, and CβH protons, together with the amide proton exchange rates and $^3J_{HN\alpha}$ coupling constant data. The thickness of the lines reflects the intensity of NOEs. The residues whose amide protons were still present 10 min after dissolution in D_2O, but exchanged by 60 min, are indicated by open circles. The residues whose amide protons remained at 60 min after dissolution in D_2O are shown by closed circles. The closed and open squares indicate $^3J_{HN\alpha}$ coupling constant values of more than 7 Hz and less than 6 Hz, respectively. The residues whose amide proton signals vanished at DNA/protein ratios, 0.1, 0.2 and 0.4, are indicated as closed, shaded and open circles, respectively.

tained chemical shifts and assignment procedures will be given elsewhere [17]. The sequential NOE connectivities were obtained from the 3D ^1H-^{15}N (^{13}C) NOESY-HSQC and ^1H-^{15}N HMQC-NOESY-HMQC spectra. For the degenerate HN(i)-HN(i + 1, 2) NOEs, we employed the H-^{15}N HMQC-NOESY-HMQC experiment. This experiment is particularly useful for obtaining the NOE connectivities between amide protons with degenerate or nearly degenerate chemical shifts. The results are summarized in Fig. 1, together with the $^3J(HN\alpha)$ values and amide exchange rates.

From the data in Fig. 1 we can identify the locations of the secondary structure elements in IRF-2(113). From the N-terminal, the arrangement of the secondary structures is α-helix (α1), turn, β-strand (β1), turn, β-strand (β2), long extended loop, α-helix (α2), loop or turn, α-helix (α3), β-strand (β3), loop, and β-strand (β4).

The short and long range NOE networks among the β1, β2, β3 and β4-strands are shown in Fig. 2. The arrangement of the β-strands is β1-β2-β4-β3, and these β-strands form an antiparallel β-sheet. Eventually, the β4-strand turns back and goes into the space between the β2- and β3-strands.

For the interaction between the α1-helix and the β-sheet, NOEs were observed for the side chains of the pairs, Ile[16]-Trp[26] and Ile[16]-Phe[34] and between the side chain of Ile[16] and the C_α proton of Lys[25]. For the α3-helix and β-sheet, NOEs were seen between Phe[81]-Tyr[109], and for the interaction of the α2-helix with the α1- and α3-helices, NOEs appeared for the pairs, Leu[54]-Phe[81], and His[61]-Trp[11]. Based on these NOEs, we could predict the folding topology of the secondary structure elements. The three helices form a bundle and are located on one side of the β-sheet.

The ^1H-^{15}N HSQC spectra were measured for a protein solution with synthesized DNA added at different DNA/protein ratios. Several peaks were found to change in intensity and signal width. The residues which showed broadening effect are indicated in Fig. 1. The broadening was remarkable for some residues in the α2-helix, the N terminal region of the α1-helix, and the α3-helix. However, some residues in the β-sheet and loops also changed.

4. Discussion

IRF2(113) exhibits the following arrangement of the secondary structure elements, α1-β1-β2-loop-α2-loop-α3-β3-loop-β4. The four β-strands form an antiparallel β-sheet and three a helices are located on one side of the sheet forming a bundle. Some of the amide proton signals of the α3-helix broadened out on the addition of DNA. According to the concept of a recognition helix, this α3-helix seems to be the helix which directly interacts with DNA.

The long loop connecting the β2-strand and the α2-helix is characteristic. It extends from the end of the β2-strand, folds back at the middle of the loop, and then continues to the α2-helix, forming a distorted helical structure. This folded loop may be involved in the interaction with DNA, as judged from the broadening on DNA titration. There are other loops con-

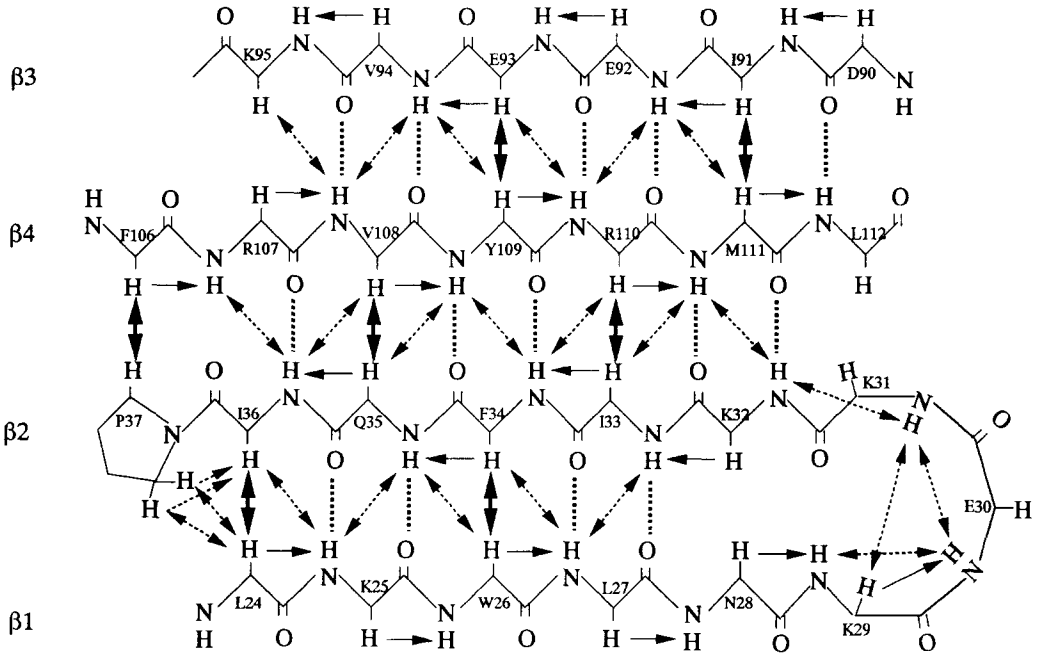

Fig. 2. Short- and long-range backbone NOE pattern indicating the presence of an anti-parallel β-sheet involving the residues from Leu^{24} to Leu^{27}, Lys^{32} to Pro^{37}, Asp^{90} to Val^{94}, and Phe^{106} to Leu^{112}. Thick and dotted double-end arrows indicate interstrand NOE connectivities. Thin single-end arrows indicate the sequential NOE connectivities. Thick dotted lines show the probable positions of hydrogen bonds estimated from the amide proton exchange rates.

necting the α2- and α3-helices, and β3- and β4-strands, respectively.

One of the unique features of the IRF-2(113) protein is that it contains many proline residues (ten in total). Proline residues often disrupt regular secondary structures and occur at the edges of the secondary structures, forming turns or unordered structures. In fact, nine of them are distributed in unordered or loop regions. Four of them are located at the terminal positions of secondary structures. The contribution of Pro^{10} is unique. In contrast to the other prolines, the Pro^{10} residue is included in the α1-helix, although it is not clear whether the helical structure is disturbed or not at this position.

There are some transcription factors (IRF-1, ISGF3γ [4], and ICSBP [5]) which participate in the regulation of IFN and

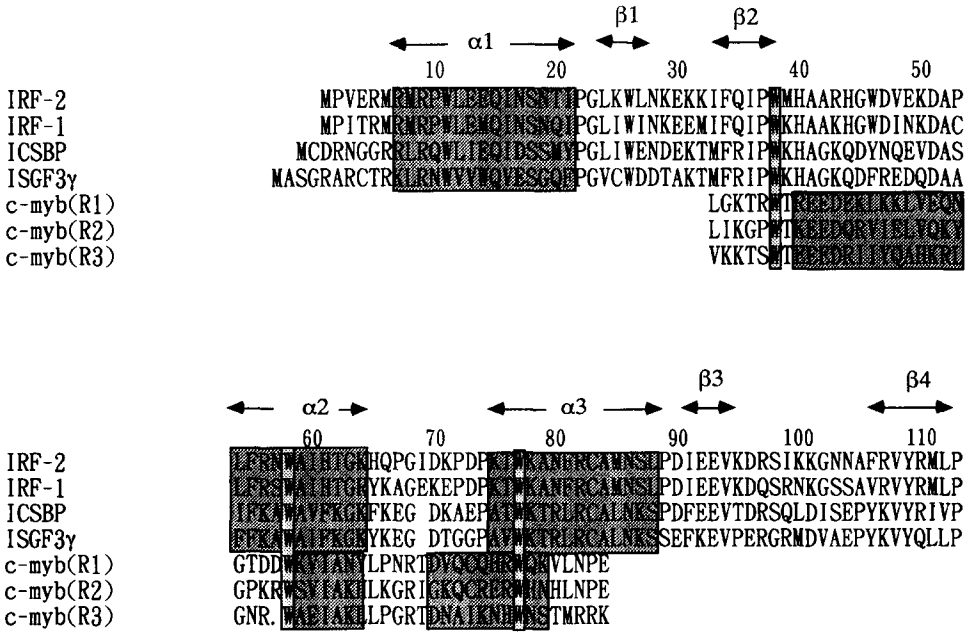

Fig. 3. Sequence alignment of several IRF family members and c-myb subdomains. The numbering of residues is according to the IRF-2 sequence. The expected secondary structure elements and conserved tryptophan residues among these family members are indicated by shaded boxes.

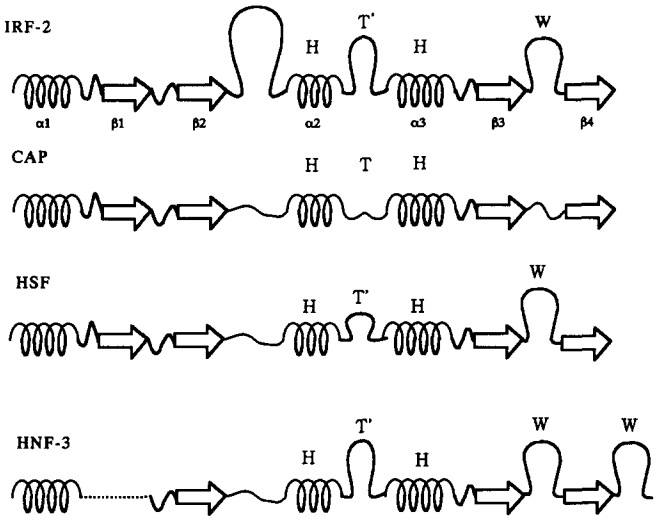

Fig. 4. Secondary structure alignment of IRF-2 and other transcription factors. CAP: cAMP receptor protein, HSF: heat shock transcription factor, and HNF-3: hepatocyte nuclear factor-3γ. H, T and W indicate helix, turn and wing structure, respectively.

IFN-inducible gene expression. Their DNA binding domains show remarkable sequence homology with that of IRF-2. This sequence similarity (especially the conserved amino acids in the secondary structure elements) strongly suggests that the DNA binding domains of these transcription factors in the IRF family have identical three-dimensional structures.

Veals et al. discussed that the IRFs (IRF-1, IRF-2, ICSBP, and ISGF3γ) are related to the myb family of DNA binding proteins, that have three α-helices arranged into a helix-turn-helix (HTH) motif [4]. Their discussion was based on the finding that the positions of three Trp residues among IRF proteins and the c-myb oncoprotein were found to be identical with each other when the IRF and c-myb sequences were aligned. Recent NMR work on DNA binding domain R3 of the c-myb protein showed that it contains the HTH type binding motif, which was supported by the hydrophobic interaction through the side chains of the aromatic residues, Trp and His [18]. When we compared the locations and lengths of the corresponding helices with those in c-myb repeats, there were remarkable differences between them (Fig. 3). The α3-helix of c-myb, R3, the recognition helix, consists of the eleven residues from Asp^{70} and Ser^{80} (numbering follows that of IRF-2). In IRF-2(113), the N terminal region (Asp^{70} to Pro^{74}) of the corresponding part is involved in a turn or a loop caused by the presence of two proline residues. In this regard, the transcription factors in the IFN system belong to a different group of DNA binding proteins from those of the c-myb family.

IRF-2(113) does not contain any homologous amino acid sequence to that of any kind of known DNA binding motif. In IRF-2(113), three α-helices form a bundle on a four-stranded antiparallel β-sheet. Such an overall structural feature is similar to that of the C-terminal DNA binding domain of bacterial activator protein (CAP) [19], and with that of the DNA binding domain of a yeast heat shock transcription factor (klaHSF) [20] (Fig. 4). They share an α1-β1-β2-α2-α3-β3-β4 sequence of secondary structure elements. A similar arrangement of the secondary structure elements of *Drosophila* melanogaster HSF was also reported [21]. The lengths of the elements are similar to each other, although there is no sequence homology among them. The number of amino acid residues of the turn in a typical HTH motif is three, like in that of CAP, but that of the turn between α2 and α3 in IRF-2(113) is 10, i.e. much longer than the 5 of klaHSF. In addition to the long turn, a typical difference in the structures of the CAP domain and IRF-2(113) is the presence of long loops between β2 and α2, and between β3 and β4. Recently, the winged-helix DNA binding motif was proposed to be a variant of the HTH motif [22]. A hepatocyte nuclear factor-3γ (HNF-3γ), which is a member of the fork head family, contains a three α-helix bundle supported by a three stranded antiparallel β-sheet including an 8 residue turn connecting α2 and α3, and two loops (wings) between β3 and β4, and at the C terminal region [23]. From the secondary structure arrangement and the presence of long loops, IRF-2(113) is also similar to the DNA binding domains of these transcription factors.

From the evidence, the DNA binding domain of IRF-2 will be classified into the $\alpha + \beta$ HTH-like structure class, whose members are not related by sequence homology. However, the absence of sequence homology among these proteins suggests that the IRF family transcriptional factors were derived from other transcriptional factors at an early stage of evolution.

Acknowledgments: This work was supported by Grants-in-Aid for Scientific Research on Priority Areas (04254103) and for Specially Promoted Research (05101004) from the Ministry of Education, Science and Culture of Japan.

References

[1] Fujita, T., Sakakibara, J., Sudo, M., Miyamoto, M., Kimura, Y. and Taniguchi, T. (1988) EMBO J. 7, 3397–3405.
[2] Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., Sudo, Y., Miyata, T. and Taniguchi, T. (1988) Cell 54, 903–913.
[3] Harada, H., Fujita, T., Miyamoto, M., Kimura, Y., Murayama, M., Furia, A., Miyata, T. and Taniguchi, T. (1989) Cell 58, 729–739.
[4] Veals, S.A., Schindler, C., Leonard, D., Fu, X-Y., Aebersold, R., Darnell Jr., J.E. and Levy, D.E. (1992) Mol. Cell. Biol. 12, 3315–3324.
[5] Driggers, P.H., Ennist, D.L., Gleason, S.L., Mak, W.H., Marrks, M.S., Levi, B.Z., Flanagan, J.R., Appella, E. and Ozato, K. (1990) Proc. Natl. Acad. Sci. USA 87, 3743–3747.
[6] Uegaki, K., Shirakawa, M., Fujita, T., Taniguchi, T. and Kyogoku, Y. (1993) Protein Eng. 6, 195–200.
[7] Tanaka, N., Kawakami, T. and Taniguchi, T. (1993) Mol. Cell. Biol. 13, 4531–4538.
[8] Bodenhausen, G. and Ruben, D.J. (1980) Chem. Phys. Lett. 69, 185–189.
[9] Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. (1990) J. Magn. Res. 36, 304–318.
[10] Norwood, T.J., Boyd, J., Heritage, J.E., Sogge, N. and Campbell, I.D. (1990) J. Magn. Reson. 87, 488–501.
[11] Ikura, M., Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) J. Am. Chem. Soc. 112, 9020–9022.
[12] Fairbrother, W.J., Cavanagh, J., Dyson, H.J., Palmer, A.G., Sutrina, S.L., Reizer, J., Saier, M.H. and Wright, P.E. (1991) Biochemistry 30, 6896–6907.
[13] Shirakawa, M., Fairbrother, W.J., Serikawa, Y., Okubo, T., Kyogoku, Y. and Wright, P.E. (1993) Biochemistry 32, 2144–2153.
[14] Grzsiek, S. and Bax, A. (1992) J. Magn. Reson. 96, 432–440.
[15] Powers, R., Gronengorn, A.M., Clore, G.M. and Bax, A. (1991) J. Magn. Reson. 94, 209–213.
[16] Kay, L.E. and Bax, A. (1990) J. Magn. Reson. 84, 72–84.

[17] Uegaki, K., Shirakawa, M., Hayashi, F., Harada, H., Taniguchi, T. and Kyogoku, Y., in preparation.
[18] Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishi, S. and Nishimura, Y. (1992) Proc. Natl. Acad. Sci. USA 89, 6428–6432.
[19] Weber, I.T. and Steitz, T.A. (1989) J. Mol. Biol. 198, 311–326.
[20] Harrison, C.J., Bohm, A.A. and Nelson, H.C.M. (1994) Science 263, 224–227.
[21] Vuister, G.W., Kim, S.J., Orosz, A., Wu, C. and Bax, A. (1994) Struct. Biol. 1, 605–613.
[22] Brennan, R.G. (1993) Cell 74, 773–776.
[23] Clark, K.L., Halay, E.E., Lai, E. and Burley, S.K. (1993) Nature 364, 412–420.