Secondary structure of the C-terminal domain of the tyrosyl-transfer RNA synthetase from *Bacillus stearothermophilus*: a novel type of anticodon binding domain?

Alessandro Pintar[a], Valerie Guez[b], Claire Castagné[a], Hugues Bedouelle[b], Muriel Delepierre[a,*]

[a]*Nuclear Magnetic Resonance Laboratory, Pasteur Institute, 28 Rue du Dr. Roux, 75015 Paris, France*
[b]*Cellular Biochemistry Unit, URA1129, Pasteur Institute, 28 Rue du Dr. Roux, 75015 Paris, France*

Abstract The tyrosyl-tRNA synthetase catalyzes the activation of tyrosine and its coupling to the cognate tRNA. The enzyme is made of two domains: an N-terminal catalytic domain and a C-terminal domain that is necessary for tRNA binding and for which it was not possible to determine the structure by X-ray crystallography. We determined the secondary structure of the C-terminal domain of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* by nuclear magnetic resonance methods and found that it is of the $\alpha+\beta$ type. Its arrangement differs from those of the other anticodon binding domains whose structure is known. We also found that the isolated C-terminal domain behaves as a folded globular protein, and we suggest the presence of a flexible linker between the two domains.
© 1999 Federation of European Biochemical Societies.

Key words: Aminoacyl-tRNA synthetase; Anticodon; Nuclear magnetic resonance; Secondary structure; *Bacillus stearothermophilus*

1. Introduction

Aminoacyl-tRNA synthetases (reviewed in [1–3]) are key enzymes in protein biosynthesis. They catalyze the activation of the amino acid through its reaction with ATP and the transfer of the aminoacyl-adenylate to the cognate tRNA. Their very high specificity is the basis of the correct translation of the genetic code and involves the recognition of such different molecules as the amino acid and the corresponding tRNA.

The complex, multisite interaction between aaRSs and their cognate tRNAs involves large contact surfaces and both basic and apolar residues [4–9,13]. The main points of contact of the tRNA with the enzyme are the acceptor arm, which will be coupled to the amino acid, a few base pairs in the T and D stems, and the anticodon loop, which contains the base triplet encoding the amino acid type. The tRNA anticodon is usually recognized by specifically dedicated domains in the aaRSs.

The crystal structure of the tyrosyl-tRNA synthetase from the prokaryote *Bacillus stearothermophilus* (TyrRS) has been determined at 2.3 Å resolution [10]. It revealed the presence of a N-terminal catalytic domain (residues 1–319) and of a disordered C-terminal domain (residues 320–419) for which the electron density was very weak and where it was not possible to trace the polypeptide chain. However, recent biochemical studies suggest that the C-terminal domain is folded in solution [11]. The C-terminal domain of TyrRS is necessary for tRNA binding, as shown by deletion mutagenesis [12]. Furthermore, site-directed mutagenesis of the C-terminal domain allowed the identification of six basic residues involved in tRNA binding [13].

Here, we report the secondary structure of a recombinant protein, TyrRS(Δ4), corresponding to the isolated C-terminal domain of the TyrRS from *B. stearothermophilus*, as determined by nuclear magnetic resonance (NMR) methods. We compare this secondary structure with the structure of the anticodon binding domains of the other aaRSs, and analyze the positions of the six basic residues involved in the binding of the tRNA anticodon. We also report the backbone dynamics of TyrRS(Δ4), as derived from heteronuclear ^{1}H-^{15}N NOE, and discuss them in relation to the compactness of the C-terminal domain.

2. Materials and methods

2.1. Sample preparation
The isolated C-terminal domain of TyrRS from *B. stearothermophilus* was produced as a recombinant protein, TyrRS(Δ4), from a derivative of the expression vector pET20, in the *Escherichia coli* strain BL21(DE3). This derivative encodes the sequence M(A321–A419)LE(H)$_6$, where the segment A321–A419 corresponds to the 99 C-terminal residues of TyrRS. The initiator methionine of this sequence is cleaved in vivo (V. Guez, S. Nair et al., unpublished). TyrRS(Δ4) was purified by affinity chromatography on a Ni^{2+} ion column. The ^{15}N and ^{15}N/^{13}C labelled samples were prepared by growing the bacteria in M9 medium enriched with vitamins and using ^{15}NH$_4$Cl and U-^{13}C$_6$ glucose as the only nitrogen and carbon sources, respectively. After their elution from the affinity column with imidazole, the protein fractions were pooled, extensively dialyzed against 20 mM potassium phosphate buffer (pH = 6.8), and concentrated on Centricon membranes. The final concentration was 1.2 mM for the ^{15}N labelled sample and 0.8 mM for the ^{15}N/^{13}C doubly labelled sample.

2.2. Nuclear magnetic resonance
NMR experiments were acquired at 308 K on a three-channel Varian Unity 500 spectrometer operating at 499.84 MHz for ^{1}H and equipped with a triple resonance *z*-gradient probe. Sequential assignments were obtained from the combination of the three-dimensional ^{15}N-edited NOESY-HSQC and TOCSY-HSQC [14] and of the triple resonance CBCA(CO)NH and HNCACB experiments [15]. Short and medium range NOE interactions were obtained from the ^{15}N-edited NOESY-HSQC recorded with a 150 ms mixing time. ^{1}Hα, ^{13}Cα, and ^{13}C′ chemical shifts were obtained from the TOCSY-HSQC, CBCA-(CO)NH/HNCACB, and HNCO [16] spectra, respectively. Chemical shifts were referenced to trimethyl-silylpropionate (TSP). $^3J_{NH-H\alpha}$

*Corresponding author: Fax: (33) (1) 4568 8885.
E-mail: murield@pasteur.fr

coupling constants were evaluated from the HMQC-J spectrum [17]. The heteronuclear ^{1}H-^{15}N NOE values [18] were calculated as NOE/ NOE_ref, where NOE is the peak intensity in the spectrum recorded with ^{1}H presaturation and NOE_ref is the corresponding peak intensity in the reference spectrum. Errors were evaluated from base plane noise. Apparent exchange rates for amide protons were estimated from the ratio of the intensity of NH cross peaks in ^{1}H-^{15}N HSQC spectra recorded with and without solvent presaturation [19]. Spectra were recorded in the States-TPPI mode, and data transformed using the VNMR software. Spectral analysis was performed using XEASY [20].

3. Results

The good dispersion of NH resonances (Fig. 1) and the easily identifiable NOEs between aromatic and aliphatic side chains (data not shown) are indicative of the presence of both secondary and tertiary structure. The peaks corresponding to the N-terminal residues L2, F3 and S4, to the C-terminal histidine tag, and to the positively charged residues K47, R48, K90 and K91 are missing or of very low intensity in the ^{1}H-^{15}N HSQC spectrum.

The secondary structure elements were first identified from the ^{1}Hα, ^{13}Cα and ^{13}C' chemical shift index (CSI) plots [21] (Fig. 2). The ^{1}Hα CSI plot suggests the presence of three helical regions (residues 12–20, 33–43, 49–55) and of four stretches of polypeptide chain in the β conformation (59–61; 72–75; 83–88; 93–98). The ^{13}Cα CSI plot confirms the existence of the three helices, although with slightly different limits (12–17; 34–40; 48–55), and of the extended regions (59–61; 72–74; 83–87; 94–98). It also allows us to identify two additional extended regions corresponding to residues 24–26 and 66–68. Finally, the CSI plot for the peptide bond ^{13}C' confirms the presence of the extended regions (24–26, 64–67, 84–87 and 91–97), and of the three helices (12–18; 36–40; 48–57).

The pattern for sequential and medium range NOEs and the values of ^{3}J$_{NH-Hα}$ coupling constants (Fig. 3) are in good

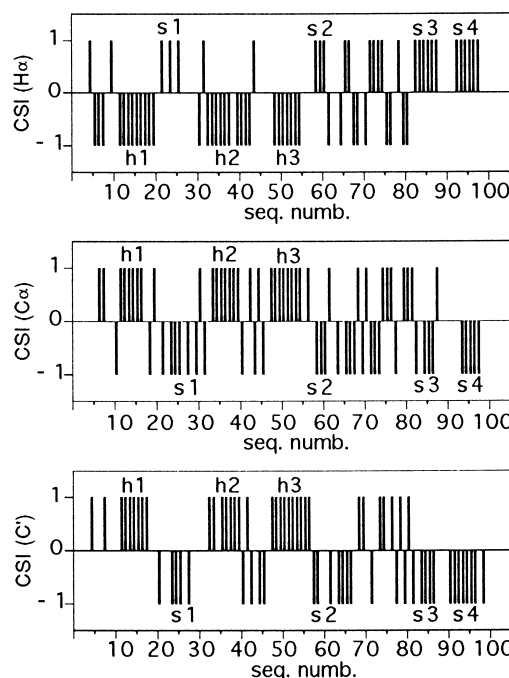Fig. 2. Chemical shift index plots based on the ^{1}H chemical shifts of α protons (top), on the ^{13}C chemical shifts of α carbons (center), and on the ^{13}C chemical shifts of peptide bond carbonyls (bottom). Consensus secondary structure regions are marked by h (helix) or s (strand).

agreement with the CSI plots with the following exceptions. In the region between residues 62 and 82, the secondary structure is not well defined. A region of extended secondary structure (64–68) can be detected by the ^{13}Cα and ^{13}C' CSI but is not supported by the ^{1}Hα CSI and by the relative intensities of sequential NN(i, i+1) and αN(i, i+1) NOEs. An additional region of extended structure (72–74) is suggested by ^{1}Hα CSI, ^{13}Cα CSI, and NOE pattern, but is not confirmed by the ^{13}C' CSI.

The measurement of apparent exchange rates for amide protons was particularly effective in identifying the helical regions (Fig. 3). Furthermore, residues that exchange quite rapidly in the second time scale could be identified (28–32; 62–67; 74–76).

The consensus secondary structure, as determined by CSI plots, NOE patterns, coupling constants and apparent exchange rates was defined as follows: helix 1 (12–18), strand 1 (24–28), helix 2 (33–42), helix 3 (48–55), strand 2 (59–61), strand 3 (83–87) and strand 4 (93–98).

The pairing of strands was obtained from the three-dimensional, ^{15}N-edited NOESY spectrum. Several NOEs between strand 3 and strand 4 (NH-NH: T84–I96, I86–Y94; Hα-NH: F83–Y98, V85–I96, R87–Y94), between strand 1 and strand 4 (NH-NH: F25–R97, H27–R97; Hα-NH: S24–L95, V26–R97), and between strand 2 and strand 3 (NH-NH: Y60–R87; Hα-NH: V61–R87) could be identified. They show that the C-terminal strands 3 and 4 are anti-parallel, and are flanked by strand 2 and strand 1, respectively.

To estimate the local backbone dynamics, we measured the heteronuclear ^{1}H-^{15}N NOE (Fig. 4). Starting from the N-terminus, the NOE values steadily increase from negative or near zero values up to 0.63, attained at residue 11. They remain
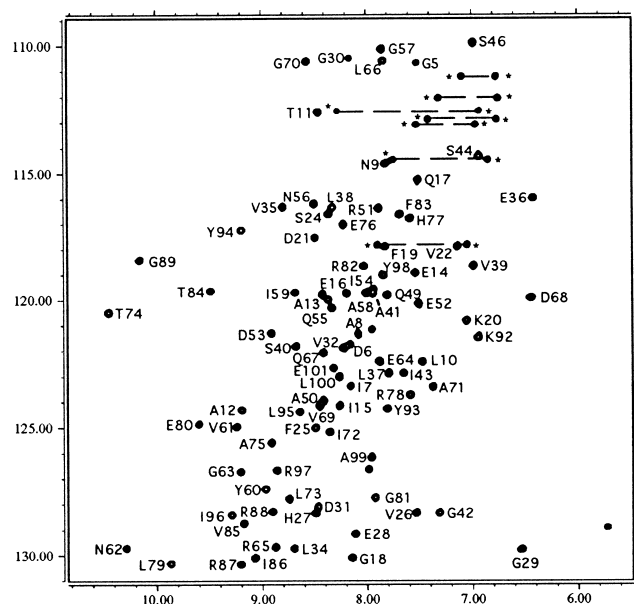

Fig. 1. ^{1}H-^{15}N HSQC spectrum of TyrRS(Δ4). Peaks corresponding to side chains of asparagines and glutamines are labelled with asterisks and connected by broken lines. Peaks corresponding to G18, G29, G42, G63, G81 and L66 are folded in the ^{15}N dimension.
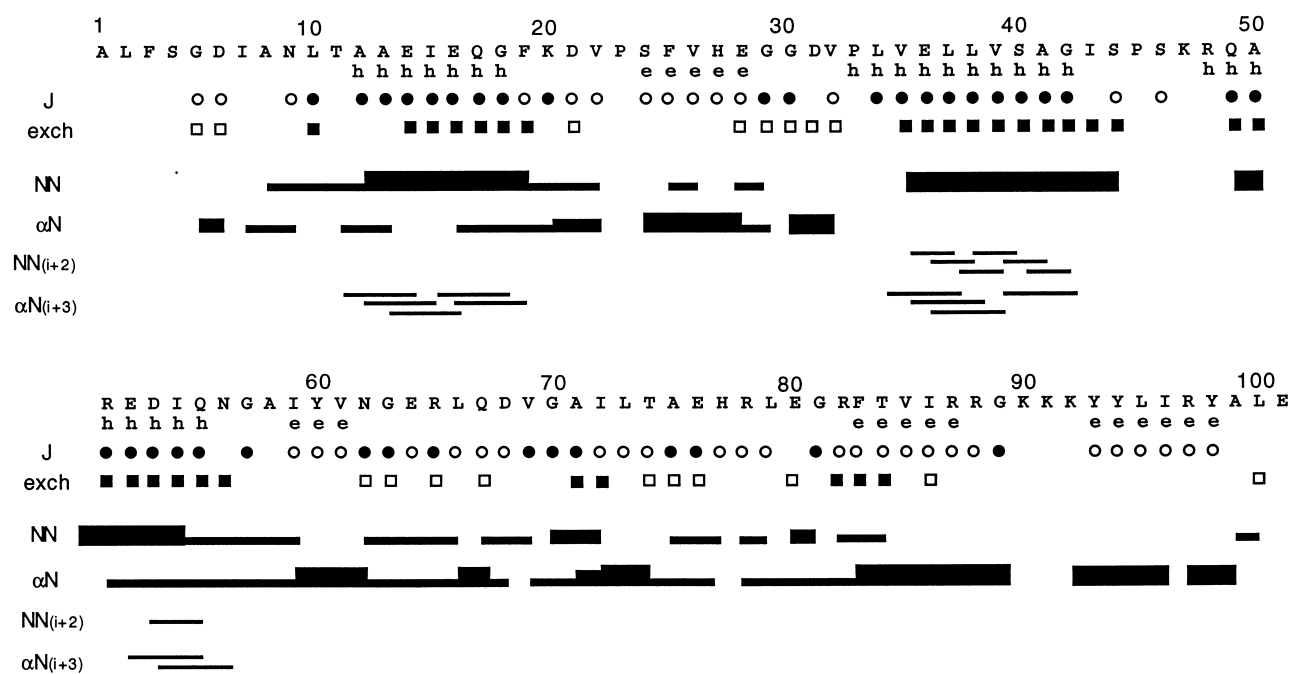
Fig. 3. Amino acid sequence, consensus secondary structure (h, helix; e, extended), $^3J_{NH-H\alpha}$ coupling constants, apparent exchange rates for amide protons, and NOE patterns for TyrRS(Δ4). The 6 histidine tag at the C-terminus is not shown in the sequence. Empty circles: large coupling constants (>9 Hz); filled circles: small coupling constants (<6 Hz). Empty square: fast exchange; filled square: slow exchange.

4. Discussion

4.1. Secondary structure

Anticodon binding domains of aaRSs show quite different folds. In class IIa synthetases (ProRS [4], GlyRS [22], HisRS [23]), they are C-terminal and have a mixed α/β fold. In class IIb synthetases (AspRS [8], AsnRS [24], and LysRS [5,7,25]), they are placed at the N-terminus and their core is composed of a five-stranded β-barrel, which is surrounded by three helices in LysRS. In MetRS (class Ia) [26], GluRS (class Ib) [27], and TrpRS (class Ic) [28], they are C-terminal and form an α-helix bundle. In GlnRS (class Ib) [9], the anticodon recognition site is formed by two C-terminal domains folding into two adjacent β-barrels. In PheRS [29], the putative anticodon

more or less constant through the sequence up to residue 99 with the only exception of G29, which shows a very low NOE value. The average NOE for residues 11–99 (not including G29) is 0.68.

Fig. 4. Plot of heteronuclear ^1H-^{15}N NOEs as a function of the amino acid sequence.

binding domain is structurally similar to that of AspRS. In SerRS [6], there is no distinct anticodon binding domain. The secondary structure of representative anticodon binding domains is sketched in Fig. 5. The C-terminal domain of TyrRS can be described as a mixed α+β protein. Its secondary structure bears no similarity with those found in the anticodon binding domains of other aaRSs, i.e. those whose structure has been determined. Class Ic contains only two members: TrpRS and TyrRS. The crystal structure of TrpRS from B. stearothermophilus has been determined [28]. TyrRS and TrpRS share a high structural homology despite their low sequence homology. From a structure-based alignment of the amino acid sequences of TrpRS and TyrRS, one would expect the 30 N-terminal residues of TyrRS(Δ4) to adopt an α-helical conformation. Although conformational changes induced by tertiary interactions between the N- and C-terminal domain cannot be excluded, our NMR results do not support this prediction. Residues 12–18 of TyrRS(Δ4) indeed adopt an α-helical conformation (helix h1), but residues 1–10 are flexible and do not show a well defined conformation, and residues 24–28 adopt a β conformation (s1). Moreover, TyrRS has 70 additional residues that are not present in TrpRS. The structural homology between TrpRS and TyrRS seems to be limited to the catalytic domain. Our results thus suggest the existence of a novel structural class of anticodon binding domains within the family of aaRSs.

4.2. Backbone dynamics

Local backbone dynamics were estimated from the heteronuclear ^1H-^{15}N NOE values. These values show that TyrRS(Δ4) behaves as a folded, globular protein and suggest the presence of a flexible linker between the N- and C-terminal domains of TyrRS. The linker could correspond to residues 1–10 in TyrRS(Δ4). In the N-terminal region of
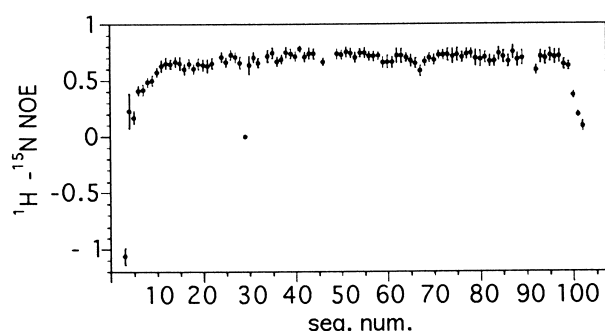
Bst-TyrRS (Ic)

321 ALFSGDIANLTAAEIEQGFKDVPSFVHEGGDVPLVELLVSAGISPSKRQAREDIQNGAIYVNGERLQDVGAILTAEHRLEGRFTVIRRGKKKYYLIRYA 419

Bst-TrpRS (Ic)

208 DDAKTIEKKIKSAVTDSEGIIRYDKEAKPGISNLLNIYSILSGQSIEELERQYEGKGYGVFKADLAQVVIETLRPIQERYHHWMESEELDRVLDEGAEKANRVASEMVRKMEQAMGLGRBR 328

Tth-ProRS (IIa)

288 IQVVIVPIYKDESRERVLEAAQGLRQALLAQGLRVHLDDRDQHTPGYKFHEWELKGVPFRVELGPKDLEGGQAVLASRLGGKETLPLAALPEALPGKLDA 387

Eco-LysU (IIb)

11 AIDFNDELRNRREKLAALRQQGVAFPNDFBRDHTSDQLHEEFDAKDNQELESLNIEVSVAGRMMTRRIMGKASFVTLQDVGGRIQLYVABDSLPEGVYNDQFKKWDLGDIIGARGTLFKIQTGELSIHCIELRLLTKAL 149
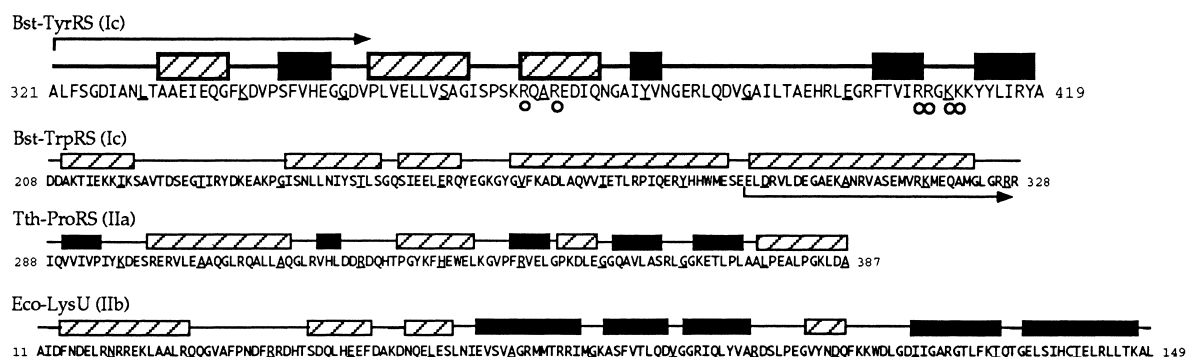
Fig. 5. Amino acid sequence and secondary structure of anticodon binding domains of different aaRSs. From top: C-terminal domain of TyrRS from *B. stearothermophilus* (Bst-TyrRS); C-terminal domain of TrpRS from *B. stearothermophilus* (Bst-TrpRS); C-terminal domain of ProRS from *Thermus thermophilus* (Tth-ProRS); N-terminal domain of the LysU isoenzyme of LysRS from *Escherichia Coli* (Eco-LysU). Dashed boxes: helices; black boxes: strands. Every tenth residue is underlined. The secondary structure of Bst-TyrRS is derived from NMR data (this work); the others are derived from the 3D structures determined by X-ray crystallography (see references in the text). The class of each aaRS is shown in parentheses. The arrows show the region of TrpRS and TyrRS that have been suggested as being structurally similar. For Bst-TyrRS, the residues identified by mutagenesis studies as important for tRNA binding are indicated with a circle.

TyrRS(Δ4), the NH cross peaks for residues 2–4 are hardly detectable because of rapid exchange with the solvent; the heteronuclear NOE values for residues 5–10 are lower than the average value observed for TyrRS(Δ4), and they steadily increase from G5 to L10. In the C-terminal region, the ^1H-^{15}N NOE values remain close to the average, down to the last residue present in the native sequence (A99), then the histidine tag is undetectable. The first ten residues are thus more flexible and less structured than the rest of the protein and the dynamic behaviour of TyrRS(Δ4) is different in its N- and C-terminal regions. The presence of a flexible linker might be important for the interaction of TyrRS with its cognate tRNA.

4.3. Localization of the basic residues involved in the recognition of the tRNA anticodon

It has been shown that a truncated TyrRS, deleted of its C-terminal domain, maintains its capability of forming the tyrosyl adenylate but no longer binds tRNA^Tyr or transfers tyrosine to tRNA [12]. Site-directed mutagenesis studies have shown [13] that three positively charged residues (K151, R207 and K208) in the N-terminal catalytic domain interact with the acceptor stem of tRNA^Tyr whereas six basic residues in the C-terminal domain (R368, R371, R407, R408, K410 and K411) interact with the anticodon arm. These six basic residues correspond to R48, R51, R87, R88, K90 and K91 in the sequence of TyrRS(Δ4). Arginines R48 and R51 are located in the third α-helix (h3). A helical wheel representation shows that h3 is an amphipathic helix, with R48 and R51 forming a positively charged cluster on its hydrophilic side. The other four residues are located at the end of strand 3 and in the loop connecting strand 3 and 4. The NH cross peaks corresponding to K90 and K91 cannot be detected in the HSQC spectra, thus suggesting that this loop is very exposed to the solvent or undergoes slow conformational changes.

In conclusion, we have found that a recombinant protein, TyrRS(Δ4), corresponding to the C-terminal domain of TyrRS, is structured in solution, and we propose the presence of a flexible linker between the catalytic and the C-terminal domain. The secondary structure of TyrRS(Δ4) is novel among the anticodon binding domains of aaRSs, and suggests the existence of a new type of anticodon binding domain.

References

[1] Arnez, J.G. and Moras, D. (1997) Trends Biochem. Sci. 22, 211–216.
[2] Francklyn, C., Musier-Forsyth, K. and Martinis, S.A. (1997) RNA 3, 954–960.
[3] Cusack, S. (1997) Curr. Opin. Struct. Biol. 7, 881–889.
[4] Cusack, S., Yaremchuk, A., Krikliviy, I. and Tukalo, M. (1998) Structure 6, 101–108.
[5] Cusack, S., Yaremchuk, A. and Tukalo, M. (1996) EMBO J. 15, 6321–6334.
[6] Cusack, S., Yaremchuk, A. and Tukalo, M. (1996) EMBO J. 15, 2834–2842.
[7] Commans, S., Plateau, P., Blanquet, S. and Dardel, F. (1995) J. Mol. Biol. 253, 100–113.
[8] Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C. and Moras, D. (1991) Science 252, 1682–1689.
[9] Rould, M.A., Perona, J.J., Soell, D. and Steitz, T.A. (1989) Science 246, 1135–1142.
[10] Brick, P., Bhat, T.N. and Blow, D.M. (1989) J. Mol. Biol. 208, 83–98.
[11] Guez-Ivanier, V. and Bedouelle, H. (1996) J. Mol. Biol. 255, 110–120.
[12] Waye, M.Y., Winter, G., Wilkinson, A.J. and Fersht, A. (1983) EMBO J. 2, 1827–1829.
[13] Bedouelle, H. and Winter, G. (1986) Nature 320, 371–373.
[14] Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) J. Biomol. NMR 4, 845.
[15] Muhandiram, D.R. and Kay, L.E. (1994) J. Magn. Reson. B 103, 203–216.
[16] Kay, L.E., Guang, Y.X. and Yamazaki, T. (1994) J. Magn. Reson. A 109, 129–133.
[17] Kay, L.E. and Bax, A. (1990) J. Magn. Reson. 86, 110–126.
[18] Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) Biochemistry 33, 5984–6003.
[19] Spera, S., Ikura, M. and Bax, A. (1991) J. Biomol. NMR 1, 155–165.
[20] Bartels, C., Xia, T., Billeter, M. and Wüthrich, K. (1995) J. Biomol. NMR 5, 1–10.

[21] Wishart, D.S. and Sykes, B. (1994) Methods Enzymol. 239, 363–392.

[22] Logan, D.T., Mazauric, M.-H., Kern, D. and Moras, D. (1995) EMBO J. 14, 4156–4167.

[23] Arnez, J.G., Harris, D.C., Mitschler, A., Rees, B., Francklyn, C.S. and Moras, D. (1995) EMBO J. 14, 4143–4155.

[24] Nakatsu, T., Kato, H. and Oda, J. (1998) Nature Struct. Biol. 5, 15–19.

[25] Onesti, S., Miller, A.D. and Brick, P. (1995) Structure 3, 163–176.

[26] Brunie, S., Zelwer, C. and Risler, J.-L. (1990) J. Mol. Biol. 216, 411–424.

[27] Nureki, O., Vassylyev, D.G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazava, T., Yokoyama, S. and Morikava, K. (1995) Science 267, 1958–1965.

[28] Doublié, S., Bricogne, G., Gilmore, C. and Carter Jr., C.W. (1995) Structure 3, 17–331.

[29] Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M. and Safro, M.G. (1995) Nature Struct. Biol. 2, 537–547.