A common core for binding single-stranded DNA: structural comparison of the single-stranded DNA-binding proteins (SSB) from *E. coli* and human mitochondria

Gordon Webster[a,*], Jochen Genschel[a,b], Ute Curth[a,b], Claus Urbanke[a,b], ChulHee Kang[a,c], Rolf Hilgenfeld

[a] *Department of Structural Biology and Crystallography, Institute of Molecular Biotechnology, Beutenbergstrasse 11, D-07745 Jena, Germany*
[b] *Biophysikalisch-Biochemische Verfahren, Medizinische Hochschule Hannover, Carl-Neuberg Strasse 1, D-30623 Hannover, Germany*
[c] *Department of Biochemistry and Biophysics, Washington State University, Pullman, WA 99164-4660, USA*

Received 30 April 1997; revised version received 3 June 1997

Abstract The crystal structure of the DNA-binding domain of *E. coli* SSB (*Eco*SSB) has been determined to a resolution of 2.5 Å. This is the first reported structure of a prokaryotic SSB. The structure of the DNA-binding domain of the *E. coli* protein is compared to that of the human mitochondrial SSB (*Hs*mtSSB). In spite of the relatively low sequence identity between them, the two proteins display a high degree of structural similarity. *Eco*SSB crystallises with two dimers in the asymmetric unit, unlike *Hs*mtSSB which contains only a dimer. This is probably a consequence of the different polypeptide chain lengths in the *Eco*SSB heterotetramer. Crucial differences in the dimer-dimer interface of *Eco*SSB may account for the inability of *Eco*SSB and *Hs*mtSSB to form cross-species heterotetramers, in contrast to many bacterial SSBs.

Key words: SSB; Single-stranded DNA-binding protein; Crystal structure

1. Introduction

From bacteria to man, single-stranded DNA-binding proteins (SSBs) are essential for DNA replication, repair and recombination. In spite of this, three-dimensional information on these proteins became available only very recently: The X-ray structures of the DNA-binding domains of human RPA and human mitochondrial SSB (*Hs*mtSSB) were reported earlier this year [1,2] and shown to bear little relation to one another, as well as being very different from the previously known structures of some viral SSBs [3–6]. However, whereas the molecular size and oligomeric state of these proteins vary dramatically between the members of the functional family, it has recently been proposed that portions of their DNA-binding domains show some resemblance to the oligomer-binding (OB) fold [7]. The SSB from *E. coli* (*Eco*SSB) consists of 177 amino acids (excluding the N-terminal methionine) and forms a very stable homotetramer [8–11]. Crystallisation studies by us [12] and others [13–16] were complicated by the fact that of the four polypeptide chains, at least two have to be truncated at their carboxy termini to obtain X-ray-grade crystals [12–14] (Pasutto et al., in preparation). Removal of the C-terminal fragments actually results in a protein with a higher affinity for single-stranded DNA (ssDNA) [17,18]. However, the C-terminal domain is essential for the in vivo functioning of *Eco*SSB, possibly by virtue of interactions with other proteins in the cell [18].

We have now determined the crystal structure of an *Eco*SSB mixed tetramer, containing two full-length polypeptide chains and two chains terminating at residue 151. Comparison of the N-terminal DNA-binding domain of *Eco*SSB with *Hs*mtSSB [2] revealed an overall similarity as well as interesting differences.

2. Materials and methods

2.1. Crystallisation

Crystals of *Eco*SSB were obtained by equilibrating a 3–4 mg/ml solution of the protein against a reservoir containing 5% PEG 400, 40 mM sodium cacodylate and 10 mM β-mercaptoethanol at pH 6.4. The protein used was a 1:1 mixture of native (177 residues) and carboxy-terminal truncated (151 residues) polypeptide chains, which had been incubated for at least 36 h at room temperature [18,19]. The monoclinic crystals produced are of space group C2 ($a = 104.72$, $b = 63.86$, $c = 98.42$ Å, $\beta = 112.30°$) and are isomorphous to *Eco*SSB crystals described earlier [12–15]. From packing considerations [20], supported by SDS gel electrophoresis (Pasutto et al., in preparation), the content of one asymmetric unit was determined to be two intact and two truncated SSB chains.

2.2. Structure determination

Native diffraction data were collected using a Mar 18 cm image plate mounted on the X-ray diffraction beamline at the Elettra synchrotron in Trieste, Italy. The data were collected at a wavelength of 1.0 Å, to a maximum resolution of 2.5 Å, and were processed using the HKL suite of programs [21] to yield a scaled set of diffraction intensities (Table 1). The structure was determined by the technique of molecular replacement, using the program AMoRe [22] with the human mitochondrial SSB structure [2] (36% sequence identity) as the search model (see Table 2). The *Hs*mtSSB model, oriented according to the molecular replacement solution found with AMoRe (with an initial *R*-factor of 49.3%) was used as a template upon which to build the *Eco*SSB sequence into electron density maps. These maps were calculated with weighted Fourier coefficients obtained from the SIG-

Table 1
Summary of X-ray diffraction data for *Eco*SSB

Space group	C2 (unique axis *b*)
Unit cell parameters	104.72, 62.19, 97.41 Å
	90.0, 112.63, 90.0°
Resolution range	30.0–2.5 Å
Total number of reflections	51 106
Number of unique reflections	18 545
R_{merge}	0.169
Completeness	91.9%
Mean I/σ(I)	3.2
Overall redundancy	2.8

Statistics quoted are for all data to 2.5 Å (no reflections excluded).

*Corresponding author. Fax: (49) (3641) 656062.
E-mail: gwebster@imb-jena.de

PII S0014-5793(97)00747-3

MAA procedure [23] to reduce the effects of model bias. All model-building operations were carried out using the programs O [24] or FRODO [25]. The resultant model was subjected to rigid-body, positional, simulated annealing and B-factor refinement using X-PLOR 3.1 [26] with the Engh and Huber geometric and force parameters [27]. After further model refitting and refinement, the *R*-factor was reduced to 25.5% for all data from 10 to 2.5 Å, for the model containing 107 residues of the structured N-terminal DNA-binding domain. The C-terminal domain is currently still under construction, and it is clear that much of this region of the protein is disordered in the crystal form described here.

2.3. Structure comparison

The crystal structures of *Eco*SSB and *Hs*mtSSB were compared using the program LSQKAB [28]. Visual comparisons were made using O [24].

3. Results and discussion

We successfully used the three-dimensional model for *Hs*mtSSB [2] to determine the crystal structure of *Eco*SSB by molecular replacement (see Table 2). Despite their relatively low sequence identity (36%), the two proteins share very similar biophysical properties [29]. They form stable tetramers with an overall D_2 symmetry, comprised of two dimers interacting head to head, across a relatively flat interface (Fig. 1). In both structures, the tetramer is comprised of two dimers related by the twofold crystallographic symmetry axis. However, while the asymmetric unit of the *Hs*mtSSB crystals contains only one dimer [2], there are two dimers per asymmetric unit in the *Eco*SSB crystals. This is probably due to the difference in polypeptide chain lengths in the mixed *Eco*SSB tetramers.

In both structures, the individual monomers are composed of a small α-helical segment (residues 63–68 in *Eco*SSB) and two β-pleated sheets that form a β-barrel whose front and back faces consist of three and four antiparallel strands respectively. The overall r.m.s deviation of Cα atoms between the DNA-binding domains of the two structures is 1.56 Å. In the *Hs*mtSSB structure [2], there are two long hairpin loops between antiparallel strands (residues 30–37 and 103–110) exhibiting large atomic temperature factors. Not unexpectedly, these loops (residues 21–28 and 88–95) show a significantly

Table 2
Summary of molecular replacement procedure for *Eco*SSB

Resolution range	20.0–4.0 Å		
Rotation and translation function (1st dimer)			
Best solution	α = 337.40	β = 99.91	γ = 12.79
	t_x = 0.2946	t_y = 0.0000	t_z = 0.2500
Correlation coefficient	0.238		
R-factor	53.8%		
Rotation and translation function (2nd dimer)			
Best solution	α = 173.87	β = 72.56	γ = 247.85
	t_x = 0.0230	t_y = 0.4135	t_z = 0.1935
Correlation coefficient	0.320		
R-factor	51.7%		
Refinement of combined solution			
Dimer 1	α = 338.19	β = 100.76	γ = 13.55
	t_x = 0.29622	t_y = 0.00013	t_z = 0.25197
Dimer 2	α = 172.80	β = 72.10	γ = 248.93
	t_x = 0.01915	t_y = 0.41481	t_z = 0.19640
Correlation coefficient	0.431		
R-factor	49.5%		

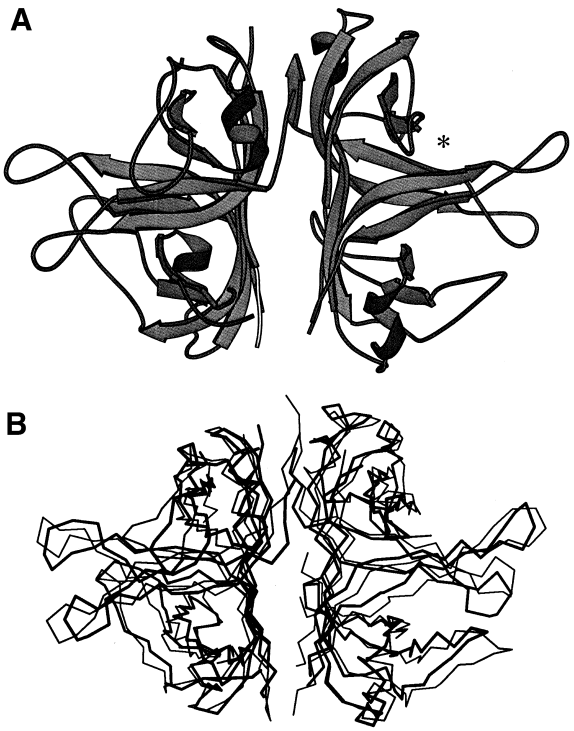

Fig. 1. (A) Secondary structure ribbon diagram of the DNA-binding domain of the *Eco*SSB tetramer, generated using MOLSCRIPT [39]. The asterisk (*) denotes the position of the putative DNA-binding cleft. (B) The superposition of the *Eco*SSB (shown in bold) and *Hs*mtSSB tetramers, plotted using MOLSCRIPT [39].

different structure in *Eco*SSB. The segment between residues 51 and 65 is apparently very mobile in *Hs*mtSSB, lacking electron density altogether [2]. *Eco*SSB has a shorter but still flexible loop at this point (residues 42–51), exhibiting only weak and discontinuous electron density for the outermost two or three residues.

In both structures, the polypeptide chains show no secondary structural features beyond residue 107 (*Eco*SSB) and 122 (*Hs*mtSSB), respectively. This seems to hold true for much of the C-terminal domain in *Eco*SSB which is not present in *Hs*mtSSB. Such a lack of secondary structure and a possible corresponding degree of disorder is not unexpected in view of the amino acid sequence of this domain, which consists of 31% glycine, 15% proline, and 20% glutamine residues for the segment 113–167. Only for the ultimate 10 residues of the intact *Eco*SSB polypeptide chain can regular secondary structure (perhaps α-helical) be expected, since there is a high proportion of negatively charged side chains in this segment. However, our electron density maps have so far failed to indicate the position of the two copies of this segment present in the 2:2 heterotetramer form of *Eco*SSB that was crystallised here. Interestingly, it has been shown that removal of this negatively charged C-terminal region leads to an enhanced affinity of the resulting truncated *Eco*SSB protein for ssDNA [18].

While the C-terminal domain is neither required nor present in *Hs*mtSSB and seems to play no role in DNA binding, it is required for the in vivo functioning of *Eco*SSB [18]. Its importance is illustrated by the *ssb113* mutation of the penultimate proline 176 to a serine, giving rise to a UV and temperature-sensitive phenotype [30,31]. Since it seems that these last

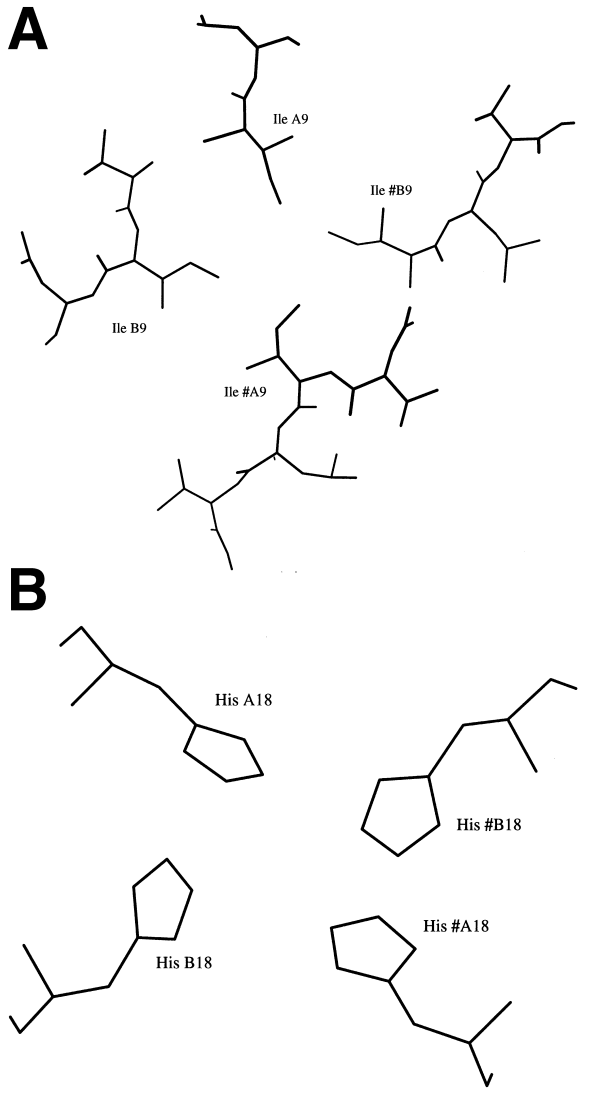

Fig. 2. The cluster of crystallographically and non-crystallographically related Ile-9 and His-18 residues in the interfaces of the *Eco*SSB (a) and *Hs*mtSSB (b) tetramers, respectively.

10 residues of the C-terminal have some role other than DNA binding, perhaps being involved in protein-protein interactions, it has been supposed that the rest of the C-terminal domain in *Eco*SSB serves merely as a spacer, keeping these negative charges away from the bound DNA [18]. This proposal is certainly in accordance with the seeming lack of structure in this region of the protein. It has also been previously shown that the C-terminal third of *Eco*SSB becomes more accessible to proteolysis upon binding of the protein to single-stranded DNA, and that removal of this entire region enhances DNA-binding affinity [17]. The presence or absence of the C-terminal 26 residues also has implications for the crystallisation of *Eco*SSB, as illustrated by the necessity to remove them in (at least) two of the four polypeptide chains of the tetramer, in order to obtain diffracting crystals [12–14].

In *Hs*mtSSB, a cluster formed by the imidazole moieties of four symmetry-related histidine residues (His-18) has been found to stabilise the tetramer interface (Fig. 2b), and it was speculated that these could constitute a binding site for zinc ions, although there is no evidence for a role of divalent cations in SSB assembly and function [2]. In *Eco*SSB, these histidines are substituted for Ile-9, the four copies of which are closely packed across the crystallographic and non-crystallographic two-fold axes (Fig. 2a). This difference in an otherwise very similar dimer-dimer interface might be the reason for the inability of *Eco*SSB and *Hs*mtSSB to form cross-species heterotetramers [18], in contrast to many bacterial SSBs [19]. A central role in the tetramerisation of *Eco*SSB has also been demonstrated for the histidine residue at position 55, replacement of which by a bulkier side chain leads to destabilization of the tetramer [32,33]. This can now be clearly understood in the light of the present structure, since this residue is very tightly enclosed on all sides, in a pocket formed by surrounding residues (Leu-10, Thr-36, Glu-53, and residues Asn-6 and Leu-83 from the other monomer in the non-crystallographic dimer) close to the tetramer interface. Prior to the existence of any structural information on SSB, it was widely accepted that the ssDNA-binding domain was α-helical [34], yet it was known that His-55 was implicated in tetramer formation, whilst the adjacent Trp-54 was a major determinant of ssDNA binding. Whereas the α-helical structure predicted for this region would be inconsistent with this finding, the β conformation that is actually observed does indeed allow two adjacent residues to have very different environments, with one being buried in the tetramer interface, whilst the other projects into the solvent region, making it accessible for ssDNA binding.

The tryptophan residues 40, 54, and 88 in *Eco*SSB have been shown to play a central role in DNA binding [35,36]. The fluorescence of Trp-54 is 95% quenched upon binding of

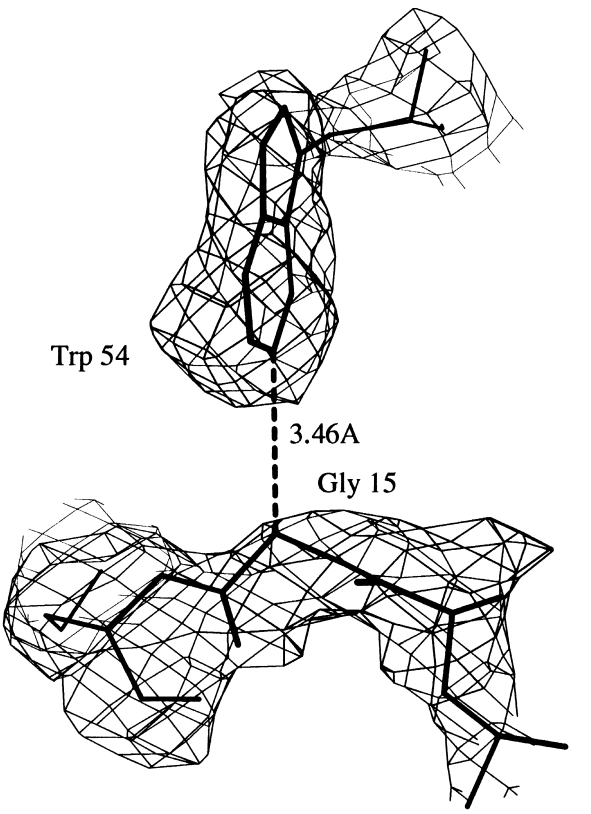

Fig. 3. The interaction of the conserved Trp-54 with Gly-15, accounting for the observed reduction of ssDNA-binding affinity in the *ssb3* mutant where Gly-15 is substituted for Asp. The electron density shown is from a $2F_o-F_c$ map contoured at 1 σ.

single-stranded DNA to *Eco*SSB [36]. This group of Trp residues, along with other important DNA-binding residues such as Phe-60, are distributed along the putative DNA-binding cleft in *Eco*SSB, enabling them to interact with the DNA. Additionally, two of three lysine residues in the N-terminal domain, Lys-49 and Lys-62, are oriented towards the DNA-binding cleft, consistent with the observation that at least one of the N-terminal domain lysines is involved in ssDNA binding [37]. Although it has been previously believed that no arginine residues were involved in ssDNA binding [37], six of them in this region of the protein (arginines 21, 41, 56, 72, 86 and 96) are also oriented into the DNA-binding cleft in a manner that may suggest a possible role in ssDNA binding. Interestingly, the *Eco*SSB *ssb3* mutation, in which Gly-15 is substituted for aspartic acid, giving rise to an extremely UV-sensitive phenotype in affected cells [38], can also now be explained by the *Eco*SSB structure: the observed mutation would produce a striking clash with the essential Trp-54 residue that is one of the key determinants of DNA binding (Fig. 3).

The structure presented here correlates very well with the existing wealth of biochemical and biological data that has been accumulated for *Eco*SSB, and is able to shed light on many aspects of it that remained hitherto unexplained.

Acknowledgements: We thank F. Pasutto and J.R. Mesters for improving the *Eco*SSB crystallisation procedure and D. Geller, U. Gröbner, J. Hoven and L. Litz for technical support. The expert advice of A. Savoia and his coworkers during data collection at Sincrotrone Trieste is gratefully acknowledged. Work at Hannover was supported by the Deutsche Forschungsgemeinschaft.

References

[1] Bochkarev, A., Pfuetzner, R.A., Edwards, A.M. and Frappier, L. (1997) Nature 385, 176–181.
[2] Yang, C., Curth, U., Urbanke, C. and Kang, C. (1997) Nature Struct. Biol. 4, 153–157.
[3] Skinner, M.M., Zhang, H., Leschnitzer, D.H., Guan, Y., Bellamy, H., Sweet, R.M., Gray, C.W., Konings, R.N., Wang, A.H. and Terwilliger, T.C. (1994) Proc. Natl. Acad. Sci. USA 91, 2071–2075.
[4] Folmer, R.H., Nilges, M., Konings, R.N. and Hilbers, C.W. (1995) EMBO J. 14, 4132–4142.
[5] Shamoo, Y., Friedman, A.M., Parsons, M.R., Konigsberg, W.H. and Steitz, T.A. (1995) Nature 376, 362–366.
[6] Tucker, P.A., Tsernoglou, D., Tucker, A.D., Coenjaerts, F.E.J., Leenders, H. and Van der Vliet, P.C. (1994) EMBO J. 13, 2994–3002.
[7] Suck, D. (1997) Nature Struct. Biol. 4, 161–165.
[8] Lohman, T. and Ferrari, M. (1994) Annu. Rev. Biochem. 63, 527–570.
[9] Meyer, R. and Laine, P. (1990) Microbiol. Rev. 54, 342–380.
[10] Greipel, J., Urbanke, C. and Maas, G. (1989) in: Protein-Nucleic Acid Interaction (Saenger, W. and Heinemann, U.), pp. 61–96, Macmillan, London.
[11] Weiner, J., Bertsch, L. and Kornberg, A. (1975) J. Biol. Chem. 250, 1972–1980.
[12] Hilgenfeld, R., Saenger, W., Schomburg, U. and Krauss, G. (1984) FEBS Lett. 170, (1) 143–146.
[13] Ollis, D., Brick, P., Abdel-Maguid, S.S., Murthy, K., Chase, J.W. and Steitz, T.A. (1983) J. Mol. Biol. 170, 797–801.
[14] Ng, J.D. and McPherson, A.J. (1989) Biol. Struct. Dynam. 6, 1071–1076.
[15] Monzingo, A.F. and Christiansen, C. (1983) J. Mol. Biol. 170, 801–801.
[16] Thorn, J.M., Carr, P.D., Chase, J.W., Dixon, N.E. and Ollis, D.L. (1994) J. Mol. Biol. 240, 396–399.
[17] Williams, K., Spicer, E., Lopresti, M., Guggenheimer, R. and Chase, J. (1983) J. Biol. Chem. 258, 3346–3355.
[18] Curth, U., Genschel, J., Urbanke, C. and Greipel, J. (1996) Nucleic Acids Res. 24, 2706–2711.
[19] de Vries, J., Genschel, J., Urbanke, C., Thole, H. and Wackerwagel, W. (1994) Eur. J. Biochem. 224, 613–622.
[20] Matthews, B.W. (1968) J. Mol. Biol. 33, 491–497.
[21] Otwinowski, Z. (1993) Proceedings of the CCP4 study weekend: Data Collection and Processing, 29–30 Jan. 1993 (Sawyer, L., Isaacs, N. and Bailey, S., Compilers), pp. 56–62, SERC Daresbury Laboratory, England.
[22] Navazza, J. (1994) Acta Cryst. D50, 760–763.
[23] Read, R.J. (1986) Acta Cryst. A42, 140–149.
[24] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Cryst. A47, 110–119.
[25] Jones, T.A. (1985) Methods Enzymol. 115, 157–171.
[26] Brunger, A.T. (1987) X-PLOR Manual – Version 3.1. A system for Crystallography and NMR, Yale University Press, New Haven, CT.
[27] Engh, R.A. and Huber, R. (1991) Acta Cryst. A47, 392–401.
[28] The CCP4 Suite: Programs for Protein Crystallography (1994) Acta Cryst. D50, 760–763
[29] Curth, U., Urbanke, C., Greipel, J., Gerberding, H., Tiranti, V. and Zeviani, M. (1994) Eur. J. Biochem. 221, 435–443.
[30] Johnson, B. (1977) Mol. Gen. Genet. 157, 91–97.
[31] Chase, J., L'Italien, J., Murphy, J., Spicer, E. and Williams, K. (1984) J. Biol Chem. 259, 805–814.
[32] Williams, K.R., Murphy, J.B. and Chase, J.W. (1984) J. Biol. Chem. 259, 11804–11811.
[33] Curth, U., Bayer, I., Greipel, J., Mayer, F., Urbanke, C. and Maass, G. (1991) Eur. J. Biochem. 196, 87–93.
[34] Sancar, A., Williams, K.R., Chase, J.W. and Rupp, W.D. (1981) Proc. Natl. Acad. Sci. USA 78, 4274–4278.
[35] Khamis, M.I.J.R., Casas-Finet, A.H., Maki, J.B., Murphy, J.B. and Chase, J.W. (1987) FEBS Lett. 211, 155–159.
[36] Curth, U., Greipel, A., Urbanke, C. and Maass, G. (1993) Biochemistry 32, 2585–2591.
[37] Bandyopadhyay, P.K. and Wu, C.-W. (1978) Biochemistry 17, 4078–4085.
[38] Schmellik-Sandage, C.S. and Tessman, E.S. (1990) J. Bacteriol. 172, 4378–4385.
[39] Kraulis, P.J. (1991) J. Appl. Cryst. 24, 946–950.