

Three-dimensional structure of *Serratia marcescens* nuclease at 1.7 Å resolution and mechanism of its action

V.Yu. Lunin^a, V.M. Levdikov^b, S.V. Shlyapnikov^c, E.V. Blagova^b, V.V. Lunin^d, K.S. Wilson^e,
A.M. Mikhailov^{b,*}

^aInstitute of Mathematical Problems in Biology, Russian Academy of Sciences, 142292 Pushino, Moscow region, Russia

^bInstitute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia

^cEngelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilov str. 32, 117984 Moscow, Russia

^dChemical Department, Moscow State University, 119899 Moscow, Russia

^eEuropean Molecular Biology Laboratory, c/o DESY, 2000 Hamburg 52, Germany

Received 2 April 1997; revised version received 16 April 1997

Abstract The three-dimensional crystal structure of *Serratia marcescens* (*Sm*) nuclease has been refined at 1.7 Å resolution to the R-factor of 17.3% and R-free of 22.2%. The final model consists of 3678 non-hydrogen atoms and 443 water molecules. The analysis of the secondary and the tertiary structures of the *Sm* nuclease suggests a topology which reveals essential inner symmetry in all the three layers forming the monomer. We propose the plausible mechanism of its action based on a concerted participation of the catalytically important amino acid residues of the enzyme active site.

© 1997 Federation of European Biochemical Societies.

Key words: *Serratia marcescens*; Endonuclease; X-ray structure; Sequence analysis; Catalytic mechanism

1. Introduction

Serratia marcescens (*Sm*) nuclease is an extracellular sugar non-specific endonuclease and catalyzes the splitting of the 3'-O-P bonds of both single- and double-stranded DNA and RNA, producing mono-, di-, tri- and tetra-5'-monophosphate-terminated nucleotides [1–3]. The nuclease is a dimer (Fig. 1) of identical subunits of 245 amino acid residues [4]. Three isoforms of *Sm* nuclease with slightly dishevelled N-terminal structures have been characterized: *Sm*₁, *Sm*₂ and *Sm*₃. The only difference between the whole enzyme (*Sm*₂) and the other two isoforms, *Sm*₁ and *Sm*₃, was heterogeneity in their N-terminus, the latter lacking three and one amino acids, respectively [5–7].

The nuclease requires divalent cations for its activity. The optimal enzyme activity was achieved with 5–10 mM Mg²⁺ ion within the pH range of 6.5–8 and Mg²⁺ can be substituted on Mn²⁺ but the efficiency of Mn²⁺ was much lower.

Sm nuclease belongs to an expanding family of sugar non-specific magnesium-dependent nucleases that have been identified in several taxonomic species of microorganisms. These homologous eu- and prokaryotic proteins include two mitochondrial nucleases from *Saccharomyces cerevisiae* [8,9] and *Bos taurus* [10] and nuclease *Anabaena* sp. PCC 7120 [11].

Some time ago, similar highly ordered crystal forms of *Sm* nuclease which were suitable for the X-ray studies were grown

in our [6] and Krause et al.'s [12] groups. Recently, the structure of the enzyme has been solved by a crystallographic analysis at 2.1 Å resolution [13] and it remains the only member of homologous related proteins for which the 3D structure is available.

In this communication, we describe the tertiary structure of *Sm* nuclease which has been determined by an X-ray diffraction analysis at 1.7 Å resolution and propose the plausible mechanism of its action based on a concerted participation of the catalytically important amino acid residues of the enzyme active site.

2. Materials and methods

The nuclease *Sm*₁ isoform without the first three residues (D-T-L), as compared to the original enzyme (*Sm*₂), was used for the crystallization experiments, but numeration of the amino acid residues in the primary structure of the protein has remained as in the nuclease *Sm*₂, viz. the N-terminal residue of the nuclease *Sm*₁ is marked as E4 and so on.

The protein was purified from the crude commercial preparation of the *Sm* nuclease as published previously [5] and the nuclease *Sm*₁ form was isolated by FPLC on Mono Q column (0.5×5 cm, Pharmacia) in accordance with a procedure, slightly altered from that described in detail elsewhere [6].

The crystallization conditions used were reported previously [6]. Three crystal forms of *Sm* nuclease have been obtained, two of which diffract to about 1.1–1.6 Å resolution.

X-ray data to 1.7 Å resolution were collected on the EMBL X11 beam line during the parasitic mode of operation of DORIS storage ring (EMBL, c/o DESY, Notkestraße 85, 2000 Hamburg 52, Germany). The detector was an imaging-plate scanner developed at EMBL by J. Hendrix and A. Lentfer. The results of the data collection are presented in Table 1. Data were integrated using the program DENZO (Z. Otwinowski).

The model of *Sm* nuclease deposited to PDB by Miller et al. was used to obtain a start model. This model at 2.1 Å resolution was refined to 16.8% value of crystallographic R-factor [13]. The start value of R-factor in the same resolution zone but calculated with new experimental data was 32.4% and the exclusion of water molecules increased the R-factor value up to 36%.

To provide reliable control in the refinement process, the test set [14] consisted of 10% of the all present reflections chosen randomly. It must be emphasized that this test set being defined once was never changed later and the test reflections were never included, neither into the refinement nor into calculation of Fourier syntheses which were used to correct periodically the model. To reduce the level of noise caused by phase errors, the weights defined on the base of maximization of marginal likelihood function [15,16] were used when calculating Fourier syntheses.

The first stage of the refinement was performed with the use of the FROG program [17,18]. The refinement was carried out with step-by-step extension of resolution zone 3.0, 2.1, 1.9, 1.7 Å and simultaneous increasing of the number of degrees of freedom in the model. As a

*Corresponding author.

E-mail: amm@biostr.crystal.msk.ru

Abbreviations: *Sm* nuclease, endonuclease of *Serratia marcescens* (EC 3.1.4.9)

Table 1
Data on crystals and X-ray diffraction measurements for *Sm* nuclease

<i>Crystals</i>	
Space group	P2 ₁ 2 ₁ 2
Unit cell (Å)	
a	106.7
b	74.8
c	69.0
Z	2
<i>Diffraction measurement</i>	
Resolution range (Å)	30.3–1.7
Measured reflections ($I > \sigma$)	219 270
Unique reflections ($I > \sigma$)	57 095
Completeness (%)	94.6
Rmerge (%)	7.0
Rst (%)	5.3

$R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$, $R_{\text{st}} = \sum \sigma / \sum I$, where I_i is the intensity measurement of a reflection and $\langle I \rangle$ is the mean value for this reflection.

result, the R-factor was reduced from 36.0% at 2.1 Å to 24.1% at 1.7 Å resolution. At the same time, the R-free value was equal 29.2 and mean deviation of bond lengths and angles was 0.004E and 1.06°. At this point 358 newly placed water molecules were included into the model. The program ASSIR [19] was used which has analyzed the environment of the picks and chosen those capable of forming hydrogen bonds with protein atoms or other water molecules.

The next stages of refinement were performed with the use of the X-PLOR system [20]. Stages of automatic refinement alternated with inspection of Fourier syntheses and manual correction of the model using FRODO [21] and O [22] programs. At the last steps of refinement, additional water molecules were included into the model, so that the whole number of water molecules became as high as 443. The final R-factor values and model quality are reflected in Table 2.

It must be noted that terminal residues were found to be poorly determined in the final model. The atoms in these parts of the chain had high-temperature factors and electron density maps exhibited lower density in these regions with respect to the other parts of the structure. The inspection of the structure of *Sm* nuclease with the use of PROCHECK program [23] has not displayed serious errors. All the overall quality criteria were estimated as 'inside' or 'better' in this program.

3. Results and discussion

The analysis of the *Sm* nuclease final model suggests a topology which is somewhat different from that previously published [4,13] and reveals essential inner symmetry in all three layers forming the monomer (Figs. 1 and 2).

The modifications concern the top layer part of the chain, which runs from residue D49 through residue R136, and are considered to be responsible for the substrate binding and the carboxy-terminal subdomain, which runs from A198 through G242, and is supposed to be involved into the building of the dimer. In the initial interpretation, the D49–R136 segment was attributed as a very long kinked helix S116–D135. Nevertheless, the visual analysis of the D49–K115 segment (Figs. 1 and 2), dihedral angles φ and ψ values and possible hydrogen bond network (Table 3) reveal a rich secondary structure.

The central part of the top layer (Figs. 1 and 2) is formed by a two-stranded β -sheet twisted through 180° (V85–H89 and I111–I115) flanked by a symmetric system of helices. In so doing, the three short helices, h1a (P67–D69, 3_{10} helix), h1b (P73–Y76, 3_{10} helix) and h1c (A79–L83, α -helix) preceding the β -sheet resemble a partially destroyed long helix, h4, which, in turn, may be divided into three connected but slightly skewed helices, h4a (S116–E120), h4b (A122–R131) and h4c (K132–D135). The central helix, h4b, is almost an ideal α -helix while h4c is closer to a 3_{10} helix and h4a is an intermediate one, as there exists here the possibility for hydrogen bonds to close both 10- and 13-member cycles. The ends of β -strands H89 and I111 are connected by a pin formed by two short 3_{10} helices, h2 (L93–L96) and h3 (W102–N106) and a turn. The loop T50–K60 has a conformation close to an extended one and forms an arc in the plane perpendicular to the plane of the layer P73–D135. Hydrogen bonds stabilize the ends of this fragment while its central part possesses more conformational freedom and may be subjected to changes in the binding process.

Suggestion of an additional helix, h5 (F200–F203), in the topology of the carboxy-terminal subdomain allows to underline the inner symmetry present in this fragment as well as in

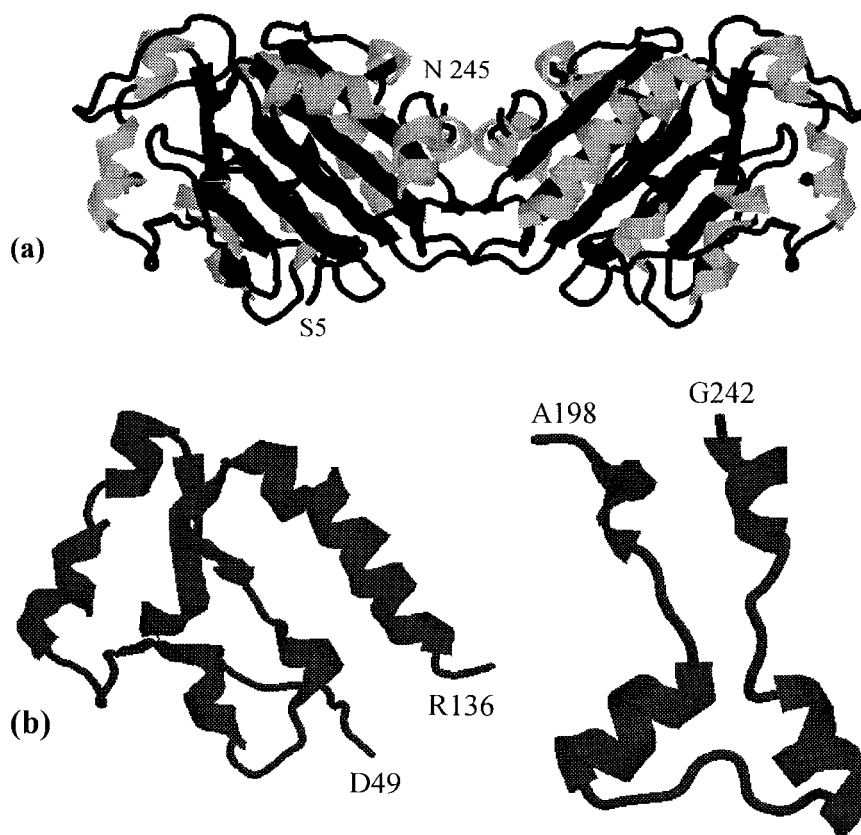
Table 2
Characteristics of the refined structure

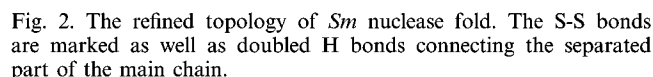
Resolution (Å)	6.0–1.7
σ -Cutoff	2.0
Number of working reflections	50 268
Number of control reflections	5679
Number of residues	2 × 239
Number of non-H protein atoms	3678
Number of water molecules	443
R factor (%)	17.3
R-free factor (%)	22.2
<i>R.m.s./maximal deviations from standard</i>	
For bond lengths (Å)	0.011/0.069
For bond angles (degrees)	1.8/10.9
<i>Mean/maximal B factors (residues 7–243)</i>	
For main chain (Å ²)	12.9/31.4
For side chains (Å ²)	16.0/50.5
For water molecules (Å ²)	35.1/62.9
<i>Mean B factors (residues 5–6 and 244–245)</i>	
For main chain (Å ²)	39.1
For side chains (Å ²)	43.8
Residues in most favoured regions of Ramachandran plot	91.2%
Residues in additional allowed regions	7.8%

Table 3
Hydrogen bonds in D49–R136 and D199–G242 main chain fragments

Main chain atoms	Distance (Å)	Main chain atoms	Distance (Å)	Main chain atoms	Distance (Å)
<i>Helix h1a</i>		<i>Helix h4a</i>		<i>Helix h6</i>	
N66 O – D69 N	2.9	K115 O – L118 N	3.0	T206 O – I210 N	3.2
P67 O – T79 N	3.6	K115 O – N119 N	3.0	V207 O – Q211 N	3.0
		S116 O – N119 N	3.2	D208 O – K212 N	2.9
		S116 O – E120 N	2.9	E209 O – R213 N	3.1
<i>Helix h1b</i>		R117 O – G121 N	3.3	I210 O – T214 N	3.0
A72 O – D75 N	2.9	L118 O – G121 N	3.2	Q211 O – G215 N	2.9
P73 O – Y76 N	3.2	L118 O – N123 N	3.2		
A74 O – T77 N	3.2			<i>Helix h7</i>	
				P223 O – Q227 N	3.1
<i>Helix h1c</i>		<i>Helix h4b</i>		D224 O – A228 N	2.9
G78 O – A82 N	3.1	G121 O – R125 N	3.0	D225 O – S229 N	3.1
A79 O – L83 N	2.9	A122 O – L126 N	3.0	V226 O – L230 N	2.9
N80 O – K84 N	2.8	W123 O – E127 N	3.0	Q227 O – K231 N	3.1
		A124 O – D128 N	3.0		
<i>β-Sheet s4–s5</i>		R125 O – E129 N	3.1	<i>Helix h8</i>	
K84 O – S116 N	2.9	L126 O – E130 N	2.9	V236 O – Q239 N	3.1
D86 N – Q114 O	2.9	E127 O – R131 N	3.1	L237 O – L240 N	2.9
D86 O – Q114 N	3.0			P238 O – M241 N	2.9
G88 N – T112 O	3.0	<i>Helix h4c</i>		Q239 O – G242 N	3.6
G88 O – T112 N	3.1	E130 O – L133 N	3.2		
Q90 N – N112 O	2.9	R131 O – I134 N	3.4		
		K132 O – D135 N	3.0		
<i>Helix h2</i>				<i>Immobilization of T50–K60 loop ends</i>	
L93 O – L96 N	3.2	<i>Helix h5</i>		T50 O – G98 N	2.8
L96 O – V99 N	3.1	D199 O – Q202 N	3.2	A52 N – A94 O	3.0
		P200 O – F203 N	3.0	K60 N – N106 O	3.3
<i>Helix h3</i>		C201 O – R204 N	3.6	K60 O – L108 N	3.0
A101 O – S104 N	3.1				
W102 O – L105 N	2.9				
E103 O – N106 N	3.0				
S104 O – Y107 N	3.1				

The data for the A chain are present. The differences for B chain are not significant.





Sequence comparison of *Sm* nuclease to three non-specific functionally related enzymes revealed five main regions where residues of *Sm* nuclease were invariant to one or several compared proteins. These regions include residues 24–57, 86–94, 107–150, 165–181 and 199–235 (Fig. 3). In addition, the *Sm* nuclease residues P67 and A68, and L3 and N245 are identical to those in the *Anabaena* and *S. cerevisiae* nucleases, respectively, and the A74 is retained also in the *B. taurus* nuclease. On the whole, there are 23 residues, which are conserved in all

Fig. 3. Sequence alignment of the *S. marcescens* (*Sm*), *Anabaena* sp. (*As*), *S. cerevisiae* (*Sc*) and *B. taurus* (*Bt*) nuclease conservative regions. Identical residues are indicated by uppercase letters. The *Sm* nuclease amino acid residues conserved in some sequences shown by bold letters. Amino acid numbering for the *Sm* nuclease is used.

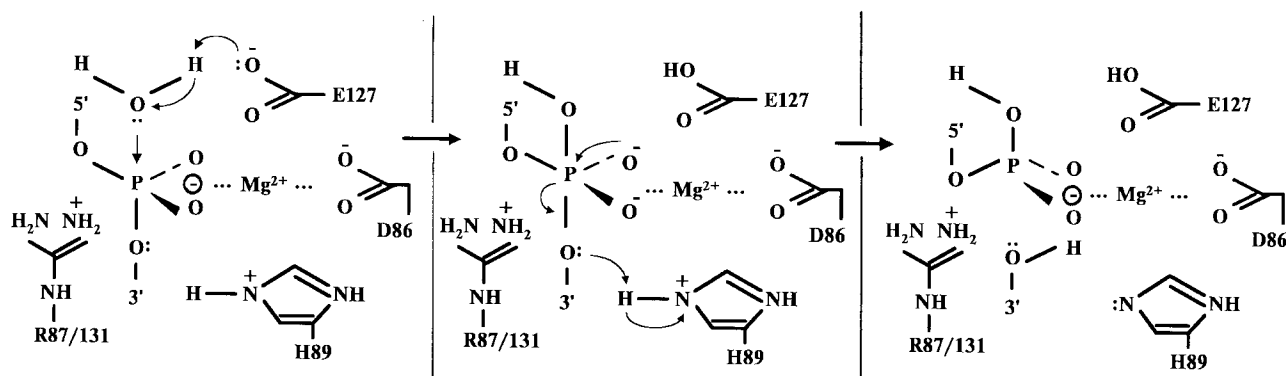


Fig. 4. A general mechanism of *Sm* nuclease-catalyzed cleavage of the phosphodiester bonds.

four proteins, and six invariant residues are placed in the most conserved *Sm* nuclease sequence 86–94.

Based on the sequence homology, it is very likely that the nuclease active sites are formed by the same amino acid residues and their geometry is similar. Among the 23 general conserved residues, six, viz. D86, R87, H89, E127, R131 and E211, have been analyzed primarily as potential candidates for direct participation in the chemistry of *Sm* nuclease-catalyzed reaction [27] but, so far, unfortunately, we have no data on the possible functional role of the invariant K172 residue. The validity of the sequence alignments in predicting active site residues was born from the *Sm* nuclease site-directed mutagenesis studies when the three above-mentioned carboxylic residues, two arginine and H89, were replaced on Ala. So, the H89A substitution results in total inactivation of the enzyme and the efficiency (k_{cat}/K_m value) of hydrolysis by the E127A, R87A, E211A, D86A and R131A mutants, respectively, $>10^3$ -, 5×10^2 -, 2×10^2 -, 10^2 - and 30-fold lower than of the wild-type nuclease [27]. Only the E211A mutant (and, to a certain degree, the R131A mutant) retains the substrate binding properties as the native enzyme. For other mutants, K_m values are as much as two orders of magnitude higher.

The results of the *Sm* nuclease site-directed mutagenesis do not prove that these residues are part of its active site but the likelihood considerably increases when these residues are strictly conserved in the sequences of the related enzymes.

At the same time, these data agree essentially (not entirely) with the results of our 3D structure studies of the free nuclease, i.e. nuclease that has no nucleotide bound at its active site, and corroborate the modeling simulation scheme of the DNA–nuclease association proposed earlier [12]. Nevertheless, the functional role of the catalytic group of the active site of *Sm* nuclease is still a matter of discussion.

The *Sm* nuclease DNA-binding site located between two main amino (residues 1–114)- and carboxy-terminal (115–245) structural domains of the protein [12] and the DNA-binding cleft is flanked by two rows of positively charged amino acids that could interact with about one full turn of DNA B-form. This cleft includes, among the other cationic amino acids, the invariant catalytic residues of R87 and R131 and contains also H89 and E127. Two arginine residues, as well as H89 and E127, belong to the different N- and carboxy-terminal subdomains of the nuclease. The latter pair represents the most closed fixed residues with about 5.9 Å distance between the H89ND1 and E127OE atoms.

These structural data, together with the observation that

both H89A and E127A nuclease mutant derivatives possess the worst catalytic activity, are consistent with an assumption about the acid–base catalysis involved in the nuclease-catalyzed phosphodiester bonds cleavage reaction when the E127 functions as a general base and the H89 serves as a general acid.

Two essential residues of *Sm* nuclease are drawn in the enzyme catalytic mechanism in roles analogous to those ascribed to H252 and H134 in the DNase I. There is no obvious way to superimpose both active sites and relate the enzymatic mechanisms but the two active DNase histidines are separated by approximately the same distance (5 Å) as H89 and E127 in *Sm* nuclease, and the visible conservative sequence near H89 of *Sm* nuclease, DRGHQAPL, and near DNase I H131, VALHSAPL, could be considered as important structural motif which provides similar functional role to the corresponding histidine residues.

It should be noted from the analysis of the pH dependencies of the *Sm* nuclease and DNase activities that, since the pK_a of a histidine is generally higher than for a glutamate, the *Sm* nuclease pH optimum, as compared with DNase, is shifted to acidic pH values (see above).

It seems unlikely that side chains of the H89 and E127 are connected via hydrogen bond in the free nuclease but it could be in its complex with DNA. In this case, the pK values of *Sm* nuclease functional groups are significantly higher than those excluded from specific perturbation. According to calculation [28], a pK change of 1.2–1.4 pK units corresponds to about 3 Å distance between the titrating group and the effective change.

The other identified, catalytically important carboxylic groups of *Sm* nuclease are also outside hydrogen-bonded distance from the H89 imidazole ring. The D86 is 9.9 Å removed from the active site histidine and remains a dominant candidate for the coordination Mg^{2+} ions. In contrast, there is no obvious satisfying explanation for the E211 catalytic role without global structural perturbation of the enzyme complexed with substrates, because the E211 (as well as K172) is located in the spatial distant from the active site protein region, it is more than 15 Å away from the essential H89.

While E211 (and K172) in *Sm* nuclease are too far away from the attacking water molecule, and in the absence of any additional information on the role of partially conserved amino acid residues of homologous enzymes (H45, D117, D208 and also H26 are not crucial for activity [26]), we may speculate that *Sm* nuclease catalyzed cleavage of the phosphodiester bonds proceeds within the framework of a general

mechanism of acid–base catalysis in which the histidine and glutamic acid act in concert with the protein-bound Mg^{2+} ion as well as one or two arginine residues.

Thus, based on the 3D structure of the free *Sm* nuclease, the data of sequence analysis and some biochemical characteristics indicating that these residues are essential for enzymatic activity, we propose that the E127 residue acts as a general base, abstracting a proton from the attacking water opposite the O'3-atom (Fig. 4). The subsequent attack at the phosphorus group with inversion of its configuration gives rise to a penta-covalent transition state. The extra negative charge of the penta-coordinated phosphorus in transition state is presumably neutralized by guanidinium groups of the R87 and/or R131. Transition state could be further stabilized by the Mg^{2+} ion, which is coordinated to the scissile phosphate group and D86, correcting position of the phosphate group relative to the enzyme. The H89 functions as a general acid protonating the leaving O3'.

It should be mentioned finally that proposed scheme of *Sm* nuclease functioning is still at present a matter of debate and more detailed information is needed to define the real mechanism of the nuclease-catalyzed phosphodiester bonds digestion.

Acknowledgements: We wish to thank Drs. A. Urzhumtsev, S. Nikonov, N. Nevskaya, N. Lunina, A. Dementiev and F. Antson for their help in the work and useful discussion. PC MAGE program of R. Wess and D. Richardson, and RasWin program of R. Sayle were fruitfully used in the structure analysis. The research described in the publication was made possible by Grants MG9000 and MG9300 from the International Science Foundation and Russian Government and RFBR Grants 94-04-12644 and 97-04-49772.

References

- [1] Filimonova, M.N., Baratova, L.A., Vospelnova, N.D., Zheto-va, A.O. and Leshinskaya, J.B. (1981) *Biokhimiya* [In Russian.] 46, 1660–1665.
- [2] Ball, T.K., Sanruggier, P.N. and Benedik, M.J. (1987) *Gene* 57, 183–192.
- [3] Biedermann, K., Jepsen, P.K., Riise, E. and Svendsen, J. (1989) *Carlsberg Res. Commun.* 54, 17–27.
- [4] Miller, M.D. and Krause, K.L. (1996) *Prot. Sci.* 5, 24–34.
- [5] Filimonova, M.N., Dementiev, A.A., Leshinskaya, J.B., Bakulina, G.Yu. and Shlyapnikov, S.V. (1991) *Biokhimiya* [In Russian.] 56, 508–512.
- [6] Bannikova, G.E., Blagova, E.V., Dementiev, A.A., Morgunova, E.Yu., Mikhailov, A.M., Shlyapnikov, S.V., Varlamov, V.P. and Vainshtein, B.K. (1991) *Biochem. Int.* 23, 813–822.
- [7] Pedersen, J., Filimonova, M., Roepstorff, P. and Biederman, K. (1993) *Biochim. Biophys. Acta* 1202, 13–21.
- [8] Vincent, R.D., Hofmann, T.T. and Zassenhhaus, H.P. (1988) *Nucleic Acids Res.* 16, 3297–3312.
- [9] Zassenhhaus, H.P. and Denniger, G. (1994) *Curr. Genet.* 25, 142–149.
- [10] Cofe, J. and Ruiz-Carrillo, A. (1993) *Science* 261, 765–769.
- [11] Muro-Pastor, A.M., Kuritz, T., Flores, E., Herrero, A. and Wolk, C.P. (1994) *J. Bacteriol.* 176, 1093–1098.
- [12] Miller, M.D., Benedik, M.J., Sillivanova, M.C., Shipley, N.S. and Krause, K.L. (1991) *J. Mol. Biol.* 222, 27–30.
- [13] Miller, M.D., Tanner, J., Alpaugh, M., Benedik, M.J. and Krause, K.L. (1994) *Nat. Struct. Biol.* 1, 461–468.
- [14] Brunger, A.T. (1992) *Nature (London)* 355, 472–474.
- [15] Lunin, V.Yu. and Skovoroda, T.P. (1995) *Acta Cryst.* A51, 880–887.
- [16] Urzhumtsev, A.G., Skovoroda, T.P. and Lunin, V.Yu. (1966) *J. Appl. Cryst.* 29, 741–744.
- [17] Urzhumtsev, A.G., Lunin, V.Yu. and Vernoslova, E.A. (1989) *J. Appl. Cryst.* 22, 500–506.
- [18] Ivanov, M.E. and Urzhumtsev, A.G. Joint CCP4 and ESF-EACBM, (1995) *Newslett. Prot. Cryst.* 31, 20–22.
- [19] Tong, H., Guillemette, J.G., Mauk, A.G., Smith, M. and Brayer, C.D. (1994) *J. Appl. Cryst.* 27, 421–426.
- [20] Brunger, A.T. (1988) *X-PLOR Manual*, Yale University Press, New Haven, CT.
- [21] Jones, T.A. (1978) *J. Appl. Cryst.* 11, 268–272.
- [22] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Cryst.* A47, 110–119.
- [23] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Cryst.* 26, 283–291.
- [24] Worrall, A.F. and Connolly, B.A. (1990) *J. Biol. Chem.* 265, 21889–21895.
- [25] Weston, S.A., Lahm, A. and Suck, D. (1992) *J. Mol. Biol.* 226, 1237–1256.
- [26] Mehdi, S. and Gerlt, J.A. (1984) *Biochemistry* 23, 4844–4852.
- [27] Friedhoff, M.J., Gimadutdinov, O. and Pingond, A. (1994) *Nucleic Acids Res.* 22, 3280–3297.
- [28] Tanford, C. and Roxby, R. (1972) *Biochemistry* 11, 2192–2198.