The three-dimensional structure of capsule-specific CMP: 2-keto-3-deoxy-*manno*-octonic acid synthetase from *Escherichia coli*

Stefan Jelakovic[a], Klaus Jann[b], Georg E. Schulz[a,*]

[a]*Institut für Organische Chemie und Biochemie, Albert-Ludwigs-Universität, Albertstr. 21, 79104 Freiburg im Breisgau, Germany*
[b]*Max-Planck-Institut für Immunbiologie, Stübeweg 51, 79108 Freiburg im Breisgau, Germany*

Received 30 May 1996; revised version received 24 June 1996

Abstract CMP-Kdo synthetases from Gram-negative bacteria activate Kdo for incorporation into lipo- and capsule-polysaccharides. Here we report the crystal structure of the capsule-specific synthetase from *E. coli* at 2.3 Å resolution. The enzyme is a dimer of 2×245 amino acid residues assuming C_2 symmetry. It contains a central predominantly parallel β-sheet with surrounding helices. The chain fold is novel; it is remotely related to a double Rossmann fold. A large pocket at the carboxyl terminal ends of the central β-strands most likely accommodates the catalytic center. A putative phosphate binding site at the loop between the first β-strand and the following helix is indicated by a bound iridium hexachloride anion.

Key words: Capsular polysaccharide; CMP-Kdo synthetase (*Escherichia coli*); X-ray analysis; Crystal structure; Saccharide activation

1. Introduction

The saccharide 2-keto-3-deoxy-*manno*-octonic acid (Kdo) is a characteristic component of lipopolysaccharides of Gram-negative bacteria [1–3]. Important enzymes of Kdo metabolism are CMP-Kdo synthetase (CKS), determined by the *kdsB* gene [4–6], and Kdo transferase, determined by the *kdtA* gene [7,8]. CKS catalyzes the activation of Kdo according to: $CTP + Kdo \xrightarrow{Mg^{2+}} CMP{-}Kdo + PP_i$, whereas Kdo transferase participates in the synthesis of the lipid A moiety of lipopolysaccharide. Kdo and sialic acid are unique in as far as they are activated by coupling to monophospho- rather than to the usual diphosphonucleosides [9].

More recently, it was found that pathogenic *E. coli* with group II capsules [1,10,11] exhibit an elevated CKS activity at capsule permissive temperatures above 20°C that is clearly correlated with capsule expression [12–14]. This activity increase is caused by a second CKS determined by one of the temperature regulated capsule genes, *kpsU* [15]. The capsule-specific second enzyme was termed K-CKS in order to distinguish it from the isozyme L-CKS engaged in lipopolysaccharide synthesis. K-CKS has recently been isolated and characterized [13–16]. It contains 245 amino acid residues among which 44% are identical with L-CKS. The kinetic data of the two isozymes differ distinctly [16].

The rapid spread of antibiotic resistance among bacterial strains and the fact that Kdo is absent from mammalian cells [17] make K-CKS an attractive target molecule for drug design as an anti-infective measure. Cloning of the *kpsU* gene from the K5 antigen gene cluster [16] permitted structural analysis of this enzyme. Here, we describe the overexpression, purification, crystallization and X-ray structure determination of K-CKS. To our knowledge this is the first actually presented structure of a sugar-activating enzyme. A short note announcing an L-CKS crystal structure has been given earlier [18]. The AMP-transferring enzyme kanamycin nucleotidyltransferase [19] and the sugar phosphate-transferring enzyme galactose-1-phosphate uridylyltransferase [20] use similar substrates, but they do not activate sugars and they differ structurally from K-CKS.

2. Materials and methods

2.1. Materials

The restriction endonucleases *Eco*RI and *Hin*dIII used in the subcloning of the DNA fragment were purchased from Boehringer-Mannheim and New England Biolabs, respectively. *Pfu* DNA polymerase and its buffer used in PCR amplification were from Stratagene. The PCR purification kit and the gel extraction kit were supplied by Qiagen. T4-Ligase was obtained from Boehringer-Mannheim and Benzonase from Merck. All other chemicals were of the highest available purity.

2.2. Expression and purification

The gene encoding K-CKS from *E. coli* was PCR-amplified from plasmid pCR3 [16] using the following primers: 5′-CCGGAATTCATGAGCAAAGCAGTTATTGTC-3′ containing an *Eco*RI site (underlined) and 5′-GCTGCAAGCTTCTTGTGGAACAGGGC-3′ with a *Hin*dIII site. Primer design was based on the DNA sequence [15]. Digestion with *Eco*RI and *Hin*dIII yielded a 750 bp amplification product which was inserted into the expression vector pUAKyI [21]. The resulting plasmid pUCKS was transformed into *E. coli* HB101 cells.

Transformed cells were grown at 37°C in LB-medium supplemented with 0.1 $g \cdot l^{-1}$ ampicillin. Expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) after the cell culture reached an optical density of 0.6 at 578 nm. After a further 5 h incubation at the same temperature, cells were harvested by centrifugation and sonicated (model 7100, Measuring and Scientific Equipment) twice for 5 min in 20 mM Tris-HCl pH 8.0 containing 5 mM $MgCl_2$, 2 mM DTT, 5 $mg \cdot l^{-1}$ phenylmethylsulfonylfluoride (PMSF) and 500 $U \cdot l^{-1}$ Benzonase. The cell debris was removed by centrifugation at $13000 \times g$ for 10 min and then at $80000 \times g$ for 1 h. After precipitation with 1.8 M ammonium sulfate, the protein was resuspended in buffer A (20 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 2 mM DTT), dialyzed against the same buffer, and applied to a Reactive Red affinity column (Sigma) equilibrated with buffer A. The protein was eluted from the column with a linear gradient of 0–300 mM NaCl in buffer A. Fractions containing overexpressed protein were identified by enzymatic activity [22] and SDS-PAGE. They were pooled, concentrated, equilibrated against buffer-A and further purified using a Resource-Q column

*Corresponding author. Fax: (49) (761) 203-5987.
E-mail: schulz@bio5.chemie.uni-freiburg.de

Abbreviations: CMP, cytidine 5′-monophosphate; CKS, CMP-Kdo synthetase; DTT, dithiothreitol; K-CKS, capsule-specific CKS; Kdo, 2-keto-3-deoxy-*manno*-octonic acid; PEG, polyethylene glycol; MIR, multiple isomorphous replacement; NCS, non-crystallographic symmetry

PII S0014-5793(96)00724-7

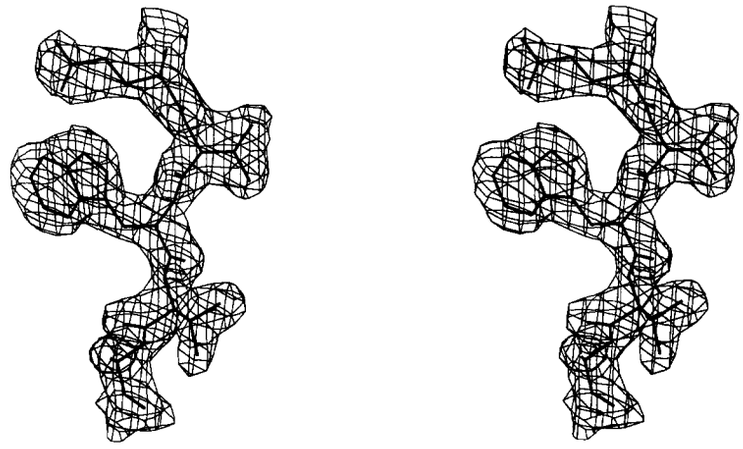

Fig. 1. Stereoview of a representative portion of the final ($2F_o$-F_c) electron density map at 2.3 Å resolution. The figure shows a region around strand β4 of the model including residues ^{45}Glu-Val-Trp-Val-Ala. The density map is contoured at the 1.3 σ level.

(Pharmacia). The purified protein was concentrated to about 30 $g \cdot l^{-1}$ and stored with 40% glycerol at −20°C. The typical yield was between 40 and 50 mg of pure protein from 1 liter culture.

2.3. Crystallization

Crystals were grown by the hanging drop method at 20°C. The protein was concentrated in a Centricon-10 centrifugation chamber (Amicon), dialyzed for 12 h against 5 mM $MgCl_2$ and 2 mM DTT in water and adjusted to a concentration of 16 $g \cdot l^{-1}$. Drops contained equal amounts of this protein solution and a solution of 16% PEG-20000, 100 mM Tris-HCl pH 9.4, 200 mM calcium acetate, 5 mM $MgCl_2$ and 2 mM DTT. The reservoir contained only 2–4% PEG-20000. Crystals appeared generally within 2 days and grew to maximum dimensions of $800 \times 600 \times 150$ μm^3 after one week. The crystals were stored in 16% PEG-8000, 100 mM Tris-HCl pH 9.4, 200 mM calcium acetate and 5 mM $MgCl_2$.

2.4. Data collection and phase determination

All X-ray diffraction data were collected at room temperature with a multi-wire area detector (model X1000, Siemens) using Cu-K α radiation from a rotating anode (model RU200B, Rigaku). Images were processed with the program XDS [23]. The structure was solved by multiple isomorphous replacement (MIR). Heavy atom derivatives were produced by soaking crystals in doped storage solutions (Table 1). The heavy atom positions were located by difference Patterson maps and confirmed by cross-difference Fourier maps. For refinement of heavy atom parameters and phase calculation with solvent flattening we used the program PHASES (W. Furey, 1991; University of Pittsburgh, PA, USA).

3. Results and discussion

3.1. Crystal structure determination

The M_r of one subunit of K-CKS is 27 028. Gel permeation chromatography showed that the enzyme is dimeric. K-CKS crystals belong to space group $P2_1$ with unit cell parameters $a = 46.1$ Å, $b = 133.5$ Å, $c = 48.3$ Å and $\beta = 102.5°$ and diffract to 1.9 Å resolution on a rotating anode. With one dimer per asymmetric unit, the crystals contain 54% solvent ($V_M = 2.6$ $Å^3$/Da). Native diffraction data have been collected to 2.3 Å resolution and initial phases were derived from 5 heavy atom derivatives up to 2.6 Å resolution resulting in a figure of merit of 0.60 (Table 1). They were improved by solvent flattening.

The resulting electron density map revealed some secondary structure elements. After skeletonization with the program BONES [24], the molecular envelope was identified and used to generate a mask with the program MAMA [25], which also yielded an initial non-crystallographic relation (NCS) between the two subunits of the enzyme. The NCS matrix was im-

Table 1
Data collection and phase determination by multiple isomorphous replacement

	Native	Derivatives				
		MMA[a]	$PtCl_4$[b]	Na_3IrCl_6[c]	$KAuCl_4$[d]	$PbCl_2$[e]
Resolution[f] [Å]	2.3	2.6	3.2	3.8	2.8	3.7
Completeness						
total [%]	95	97	94	94	80	71
outermost shell [%]	65	82	72	72	71	62
R_{sym}[g] [%]	6.8	6.1	10.2	5.2	6.3	4.1
R_{iso}[h] [%]		21.4	13.2	10.6	12.4	13.4
Sites		.16/.00/.01/81	.34/.16/.49/61	.35/.16/.50/69	.82/.35/.08/63	.87/.36/.11/56
x/y/z/rel. occupancy		.89/.07/.42/76		.09/.35/.24/59		
Phasing power[i]		1.7	1.2	1.3	1.2	1.2

[a] Methylmercury acetate, 1 mM in storage buffer soaked for 3 days, the sites are at Cys-126 and Cys-126′.
[b] Soaked, 2 mM in storage buffer for 1 day, the site is near Arg-10 and Lys-19.
[c] Soaked, 1 mM in storage buffer for 1 day, the sites are near Arg-10, Lys-19 and Arg-10′, Lys-19′.
[d] Soaked, 5 mM in storage buffer for 2 days, the site is at Cys-229′.
[e] Soaked, 1 mM in storage buffer for 1 day, the site is at Cys-229′.
[f] Each data set was measured from a single crystal.
[g] $R_{sym} = \Sigma_h \Sigma_i | I(h)_i - < I(h) > | / \Sigma_h \Sigma_i I(h)_i$.
[h] $R_{iso} = 2 \cdot \Sigma | F_{der} - F_{nat} | / \Sigma (F_{der} + F_{nat})$.
[i] Values are for the total resolution range of the respective derivative.

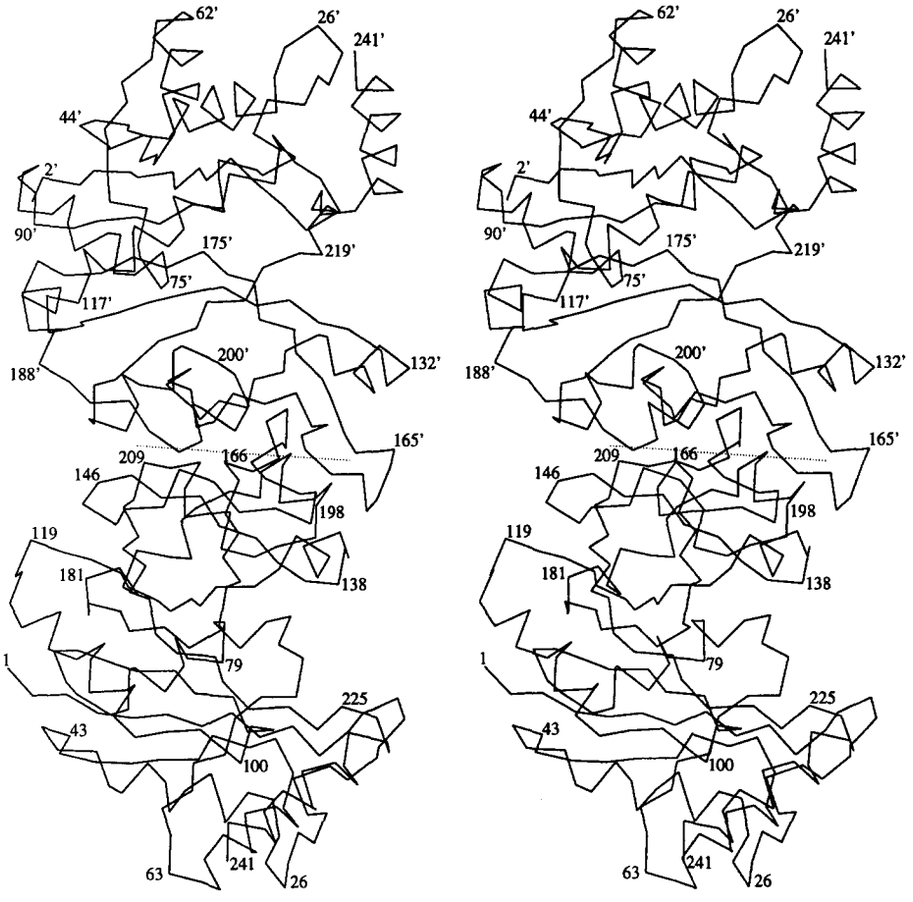

Fig. 2. Stereoview of the C_α backbone of dimeric capsule-specific CMP-Kdo synthetase from *E. coli* (K-CKS). Some residue numbers are given. The molecular twofold axis is a dotted line.

proved with the program IMP [25]. Subsequent density averaging with the program DM [26] yielded a map that revealed an unambiguous polypeptide chain trace.

3.2. Model building and refinement

The chain tracing was converted to a model using the program O [27] together with the DNA-derived amino acid sequence of Pazzani et al. [15]. One subunit was fitted, and the other one was generated by NCS. The crystallographic *R*-factor for this initial model was 38.5%. The model was then refined using the program X-PLOR [28] with NCS restraints for all residues that are not involved in interface or packing contacts. No solvent molecules were included at this stage. The refinement converged at an *R*-factor of 22.5% for all data in the resolution range 10–2.3 Å. The free *R*-factor as based on a 5% test set of about 1100 reflections was 26.6%.

The current model consists of 2×241 amino acid residues, the 4 C-terminal residues are mobile and invisible. A representative portion of the electron density map is shown in Fig. 1. The quality of the model is satisfying because all non-glycine residues have main-chain dihedral angles in 'allowed' regions with 93% of them in 'most favored' regions [29]. The root mean square deviations from standard are below 0.01 Å and 1.4° for bond lengths and angles, respectively. A *cis*-peptide precedes Pro-159 in both subunits. The NCS relates the two subunits by an exact 180° rotation without translation.

3.3. Structure and catalysis

The chain fold of dimeric K-CKS is shown in Fig. 2. One subunit is a single α/β domain of approximate dimensions $55\times45\times40$ Å^3. About 60% of the amino acids are in secondary structural elements, i.e. in nine α-helices, two short 3_{10}-helices and 11 β-strands. A topology graph is given in Fig. 3.

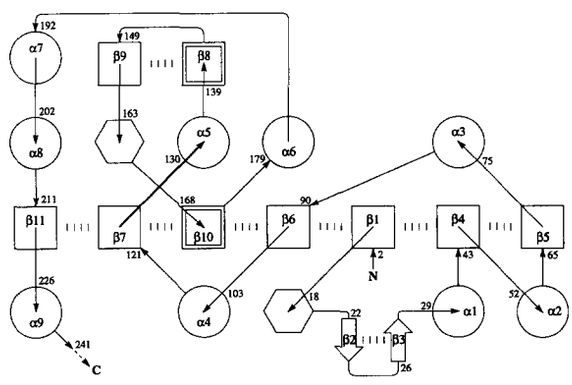

Fig. 3. Topology diagram of one subunit of the capsule-specific CMP-Kdo synthetase from *E. coli* (K-CKS). The β-strands are represented by single squares if they point towards the viewer and as double squares if they point in the opposite direction; β-strands lying approximately in the plane of the paper are drawn as arrows. The α-helices are circles and the 3_{10}-helices are hexagons. The positional numbers denote residues preceding a secondary structure element.

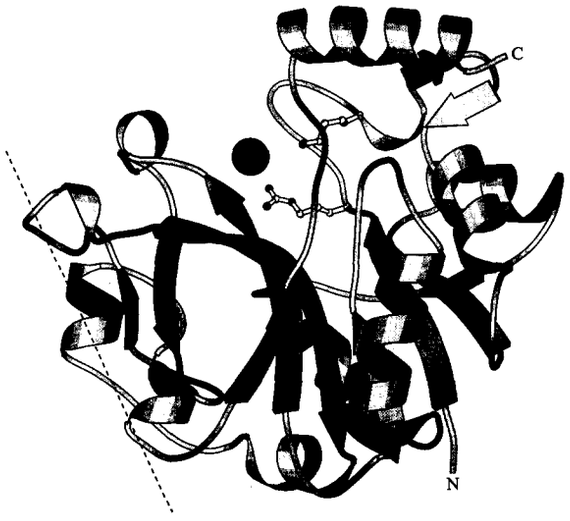

Fig. 4. Ribbon plot of one subunit of the dimeric capsule-specific CMP-Kdo synthetase produced with MOLSCRIPT [36]. Helices are given as coils and β-strands as arrows. The molecular twofold axis is a broken line. Arg-10 and Lys-19 are depicted together with the $IrCl_6^{3-}$ anion at a putative phosphate binding site.

The dominant feature is a seven-stranded predominantly parallel β-sheet flanked on both sides by α-helices. The β-sheet shows the usual twist. Strand β10, which contains a bulge at residue Gly-174, is antiparallel to the other six strands. It forms additional hydrogen bonds with strand β8. Strands β6 and β7 are joined by helix α4 forming an unusual left-handed βαβ connection. At one end, the major β-sheet is extended to a further smaller β-sheet that participates in a 40 residue chain excursion connecting strands β7 and β10. This excursion forms most of the subunit interface (Fig. 2). The chain ends in helix α9 far away from the interface.

If the antiparallel strand β10 of the central sheet were ignored (taking residues 125–211 as the connection between β7 and β11) and if β6 and β7 were connected in a right-handed manner, the chain topology would resemble the double Rossmann fold of the lactate dehydrogenase family [30]. The topology (Fig. 3) indicates that during early evolution these two changes could have occurred concomitantly in a continuous manner, rendering the CKS-fold a derivative of the more abundant double Rossmann fold.

While the monomer is reasonably compact, the dimer is very extended (Fig. 2) with approximate dimensions of $90 \times 45 \times 40$ Å^3. It is formed through tight interactions, involving mostly helix α7 and the long loop between strand β9 and the following 3_{10}-helix. These form nonpolar contacts together with a total of 18 hydrogen bonds between the two subunits, most of which are between side chain atoms. As judged from Fig. 2, the subunit can be subdivided into two wings that line a large pocket in between. The first wing consists of residues 1–100 together with the C-terminal α-helix, while the second wing contains the remaining residues and forms the dimer interface.

The pocket is at the carboxy terminal ends of the central predominantly parallel β-sheet, where numerous other enzymes of such a construction have their catalytic centers. The pocket is large enough to accommodate the substrates CTP and Kdo. Furthermore, several amino acid residues within this pocket are conserved between K-CKS and other CMP-Kdo synthetases. Drawing from other mononucleotide binding enzymes with a central parallel β-sheet like the adenylate kinases [31] one would expect that the loop between strand β1 and the following 3_{10}-helix binds a phosphate of CTP. This suggestion is supported by the observation that the heavy atom anion $IrCl_6^{3-}$ binds at Arg-10 and Lys-19 close to this loop as sketched in Fig. 4 and noted in Table 1. The amino acid sequence after strand β1 is ^{12}Gly-Ser-Ser-Arg-Leu-Pro-Gly-Lys, which resembles the mononucleotide sequence fingerprint Gly-x-x-Gly-x-Gly-Lys only vaguely. It should be kept in mind, however, that CKS transfers a double phosphoryl group to water forming a pyrophosphate and not a single phosphoryl group like the kinases and the G-proteins [31].

A search through the Protein Data Bank using the program DALI [32] indicated that the chain fold of K-CKS is not related to any deposited protein. The sequence of K-CKS is 40–44% identical with CMP-Kdo synthetases involved in lipopolysaccharide synthesis (L-CKS) in various Gram-negative bacteria [6,33,34] and shows 21% identity to the CMP-N-acetylneuraminic acid synthetase encoded by the K1 antigen region-2 of *E. coli* [35]. All these enzymes catalyze the generation of CTP-activated sugar nucleotides. The presence of K-CKS in addition to L-CKS (lipopolysaccharide biogenesis) presumably allows for an independent regulation of group II capsule expression [15]. These capsules constitute important virulence determinants of pathogenic *E. coli* strains causing extraintestinal infections.

Acknowledgements: We thank Dr. C. Rosenow for providing the recombinant plasmid pCR3 used for subcloning the K-CKS gene. This project was supported by the Deutsche Forschungsgemeinschaft under SFB-60. The coordinates and the structure factors will be deposited in the Protein Data Bank.

References

[1] Whitfield, C. and Valvano, M.A. (1993) Adv. Microbiol. Physiol. 35, 135–246.
[2] Raetz, C.R.H. (1990) Annu. Rev. Biochem. 59, 129–170.
[3] Unger, F.M. (1981) Adv. Carbohydr. Res. 38, 323–388.
[4] Ray, P.H., Benedict, C.D. and Grasmuk, H. (1981) J. Bacteriol. 145, 1273–1280.
[5] Goldman, R.C. and Kohlbrenner, W.E. (1985) J. Bacteriol. 163, 256–261.
[6] Goldman, R.C., Bolling, T.J., Kohlbrenner, W.E., Kim, Y. and Fox, J.L. (1986) J. Biol. Chem. 261, 15831–15835.
[7] Raetz, C.R.H. (1987) in: *Escherichia coli* and *Salmonella typhimurium* (Neidhard, F.C., Ed.) pp. 498–603, ASM Publications, Washington, DC.
[8] Clementz, T. and Raetz, C.R.H. (1991) J. Biol. Chem. 266, 9687–9696.
[9] Gabriel, O. (1982) Methods Enzymol. 83, 332–353.
[10] Jann, K. and Jann, B. (1990) Curr. Top. Microbiol. Immunol. 150, 19–42.
[11] Jann, K. and Jann, B. (1992) Can. J. Microbiol. 38, 705–710.
[12] Finke, A., Roberts, I., Boulnois, G., Pazzani, C. and Jann, K. (1989) J. Bacteriol. 171, 3074–3079.
[13] Finke, A., Jann, B. and Jann, K. (1990) FEMS Microbiol. Lett. 69, 129–134.
[14] Bronner, D., Sieberth, V., Pazzani, C., Roberts, I.S., Boulnois, G.J., Jann, B. and Jann, K. (1993) J. Bacteriol. 175, 5984–5992.
[15] Pazzani, C., Rosenow, C., Boulnois, G.J., Bronner, D., Jann, K. and Roberts, I.S. (1993) J. Bacteriol. 175, 5978–5983.
[16] Rosenow, C., Roberts, I.S. and Jann, K. (1995) FEMS Microbiol. Lett. 125, 159–164.
[17] Hammond, S.M., Claesson, A., Jansson, A.M., Larsson, L.G.,

Pring, B.G., Town, C.M. and Ekström, B. (1987) Nature 327, 730–732.
[18] Park, C., Rydel, T., Kohlbrenner, W.M., Lartey, P., Erickson, J.W. and Abad-Zapatero, C. (1992) Meeting abstract W05, Annual ACA meeting, Pittsburgh, PA.
[19] Pedersen, L.C., Benning, M.M. and Holden, H.M. (1995) Biochemistry 34, 13305–13311.
[20] Wedekind, J.E., Frey, P.A. and Rayment, I. (1995) Biochemistry 34, 11049–11061.
[21] Proba, K. (1993) PhD Thesis, Albert-Ludwigs-Universität, Freiburg im Breisgau.
[22] Ray, P.H. and Benedict, C.D. (1982) Methods Enzymol. 83, 535–540.
[23] Kabsch, W. (1988) J. Appl. Cryst. 21, 916–924.
[24] Jones, T.A., Zou, J.Y., Cowan, S.W and Kjeldgaard, M. (1991) Acta Crystallogr. A47, 110–119.
[25] Kleywegt, G.J. and Jones, T.A. (1994) in: From First Map to Final Model (Bailey, S., Hubbard, R. and Waller, R., Eds.), pp. 59–66, SERC Daresbury Laboratory, Warrington.
[26] Cowtan, K. (1994) Joint CCP4 and ESF-EACBM Newslett. Protein Crystallogr. 31, 24–28.
[27] Jones, T.A. and Kjeldgaard, M. (1993) O – The Manual (Version 5.9). Uppsala University, Uppsala.
[28] Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) Science 235, 458–460.
[29] Laskowski, R.A., McArthur, M.V., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Cryst. 26, 283–291.
[30] Schulz, G.E. and Schirmer, R.H. (1979) Principles of Protein Structure, Springer, New York.
[31] Schulz, G.E. (1992) Curr. Opin. Struct. Biol. 2, 61–67.
[32] Holm, L. and Sander, C. (1993) J. Mol. Biol. 233, 123–138.
[33] Fleischmann, R.D. et al. (1995) Science 269, 496–512.
[34] Tipples, G. and McClarty, G. (1995) J. Biol. Chem. 270, 7908–7914.
[35] Zapata, G., Vann, W.F., Aaronson, W., Lewis, M.S., and Moos, M. (1989) J. Biol. Chem. 264, 14769–14774.
[36] Kraulis, P.J. (1991) J. Appl. Cryst. 24, 946–950.