

Structural insight into gene duplication, gene fusion and domain swapping in the evolution of PLP-independent amino acid racemases

Lijun Liu^a, Kousuke Iwata^b, Masafumi Yohda^b, Kunio Miki^{a,c,*}

^aDepartment of Chemistry, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

^bDepartment of Biotechnology and Life Science, Faculty of Engineering, Tokyo University of Agriculture and Technology, Naka-cho, Koganei, Tokyo 184-8588, Japan

^cRIKEN Harima Institute/SPring-8, Koto 1-1-1, Mikazukicho, Sayo-gun, Hyogo 679-5148, Japan

Received 6 August 2002; accepted 14 August 2002

First published online 2 September 2002

Edited by Hans Eklund

Abstract The X-ray crystal structure has revealed two similar $\alpha\beta$ domains of aspartate racemase (AspR) from *Pyrococcus horikoshii* OT3, and identified a pseudo mirror-symmetric distribution of the residues around its active site [Liu et al. (2002) J. Mol. Biol. 319, 479–489]. Structural homology and functional similarity between the two domains suggested that this enzyme evolved from an ancestral domain by gene duplication and gene fusion. We have expressed solely the C-terminal domain of this AspR and determined its three-dimensional structure by X-ray crystallography. The high structural stability of this domain supports the existence of the ancestral domain. In comparison with other amino acid racemases (AARs), we suggest that gene duplication and gene fusion are conventional ways in the evolution of pyridoxal 5'-phosphate-independent AARs. © 2002 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Aspartate racemase; Amino acid racemase; PLP-independent; Gene duplication and gene fusion; Domain swapping

1. Introduction

The mirror image of an enzymatic reaction can take place under specific conditions. A good example of such a reaction has occurred in the study of HIV-1 retroviral proteinases, which has shown that the natural enzymes are active only on L-peptides, whereas the D-enzymes (artificially synthesized from D-amino acids) digest only D-peptides [1,2]. In contrast, some enzymes such as racemases can interact with both L- and D-enantiomers despite the lack of mirror symmetry in enzymes due to the homochiral evolutionary selection. Thus, it should be of interest to determine how racemases have evolutionally adapted themselves to the mirror-symmetrical reactions that they catalyze. As such, studies on amino acid racemases (AARs) might provide direct insight into this question.

Amino acid racemizations are a class of proton-transfer reactions that provide the D-amino acid substrates for some

specific physiological functions such as mediation of the mammalian nervous transmission and maintenance of bacterial cell-wall rigidity and strength [3–5]. Depending upon the existence of cofactors, amino acid racemases can be divided into two classes, pyridoxal 5'-phosphate (PLP)-dependent and -independent types [6]. Aspartate racemase (AspR) [7], glutamate racemase (GluR) [8], diaminopimelate epimerase (DapE) [9], and proline racemase [10] catalyze the corresponding racemizations in a PLP-independent manner, whereas alanine racemase [11] does so in a PLP-dependent manner. All of the above-mentioned PLP-independent AARs have a pair of cysteine residues at their active sites as the catalytic acid–base pair.

Among these PLP-independent AARs, *Haemophilus influenzae* DapE (*H. DapE*) appears to have two superimposable $\alpha\beta$ domains composed primarily of β -strands [12], and *Aquifex pyrophilus* GluR (*A. GluR*) is made up of two $\alpha\beta$ domains, although the two domains are not equal in size, with one having a four-stranded and the other a six-stranded β -sheet [13]. We have recently determined the crystal structure of *Pyrococcus horikoshii* OT3 AspR (*P. AspR*), which was found to have two $\alpha\beta$ domains. The *P. AspR* structure also exhibits a pseudo mirror-symmetrical arrangement of the amino acid residues in the active site, providing insight into how racemase catalyzes a mirror-symmetric reaction [14].

In the present study, by examining the structural features of the two domains of *P. AspR* and comparing these features with those of other AARs, we have revealed an evolutionary aspect of PLP-independent AARs for gene duplication, gene fusion, and domain swapping. We have also succeeded in expressing the C-terminal domain of *P. AspR* (C-domain), and determined the three-dimensional structure of this domain, which was found to be almost identical to that of the monomer structure. The structural stability of the solely expressed domain supports the existence of the ancestral domain and provides us with a rough structural image of that domain.

2. Materials and methods

2.1. Structural determination of *P. AspR*

Three-dimensional structure determination of the intact *P. AspR* has been reported previously [14].

2.2. Structure determination of the C-domain

2.2.1. Cloning, expression, and purification. PCR amplification was performed with the oligonucleotide primers, 5'-AACATATGATAGAAAGAACTGCAAA-3' and 5'-AAGGATCCTTAAGGATC-

*Corresponding author. Fax: (81)-75-753 4032.

E-mail address: miki@kuchem.kyoto-u.ac.jp (K. Miki).

Abbreviations: PLP, pyridoxal 5'-phosphate; AAR, amino acid racemase; AspR, aspartate racemase; GluR, glutamate racemase; DapE, diaminopimelate epimerase; C-domain, C-terminal domain of *P. AspR*; SC-domain, solely expressed C-terminal domain of *P. AspR*

TATGAGGGGAAGTT-3', using pPH0670E [15] as a template to obtain a DNA fragment encoding the C-domain encompassing the residues from Met103 to Pro213 with restriction digestion sites at both ends. The amplified DNA fragment was subcloned into the pT7Blue T vector (Novagen). After sequence confirmation, the gene was excised by *NdeI* and *BamHI*, and introduced into pET23a (Novagen) to construct a plasmid, pPH0670d2E. *Escherichia coli* BL21(DE3) transformed with pPH0670d2E was cultured at 37°C in a Luria–Bertani medium containing 100 µg ml⁻¹ of ampicillin. The *E. coli* cells were collected by centrifugation and disrupted by sonication. The crude extract was heated at 70 °C for 30 min to denature host proteins. After removal of most host proteins by centrifugation, the extract was applied onto a Butyl-Toyopearl column (Tosoh, Japan) equilibrated with buffer A (50 mM Tris–HCl, pH 8.0, 1 mM 2-mercaptoethanol) containing 1.6 M ammonium sulfate, and the column was then eluted with a linear ammonium sulfate gradient from 1.6 to 0 M. Fractions containing the C-domain were merged, concentrated, equilibrated with buffer A by ultrafiltration (Centriprep YM10, Millipore, USA), and then applied to a UnoQ6 column (Bio-Rad, USA) equilibrated with buffer A. Finally the column was eluted with buffer A in which a linear NaCl gradient from 0 to 500 mM was introduced, and the fractions containing the solely expressed C-terminal domain (SC-domain) were merged and concentrated to 30 mg ml⁻¹ in buffer A for crystallization.

2.2.2. Crystallization. The concentration of the SC-domain for crystallization was 30 mg ml⁻¹ in the purification buffer (50 mM Tris–HCl, pH 8.0, 2 mM mercaptoethanol). Crystallization was carried out at 20°C by vapor-diffusion sitting-drop method. Initial crystallization trials were carried out by Hampton Research Crystal Screen kits, and microcrystals were obtained within 2 days. The reservoir solution condition was finally optimized to 16% polyethylene glycol 8000, 0.10 M MES, pH 6.5, 0.20 M calcium acetate. The crystallization drop was prepared by mixing 3 µl of protein solution and 3 µl of reservoir, equilibrated against 0.8 ml of reservoir. Within 3 days, tetragonal plate-like crystals could be harvested at a maximum size of 0.2×0.2×0.04 mm.

2.2.3. Diffraction data collection and processing. X-ray diffraction data were collected at 100 K using a MARCCD detector on BL44B2 at SPring-8 (Harima, Japan) at 2.04 Å resolution. Crystals were directly flash-frozen in a nitrogen stream, and 20% glycerol was used as a cryoprotectant. The X-ray wavelength was 1.000 Å. Data were processed with MOSFLM [16] and subsequently merged with SCALA of the CCP4 suite [17]. The statistics for data collection and processing are shown in Table 1.

2.2.4. Structure determination and refinement. The molecular replacement method was employed for phasing. The primary solution was obtained with AmoRe [18] with a model of the C-domain of the intact *P. AspR* structure [14]. The structure was initially refined by a rigid-body treatment, followed by simulated annealing using CNS [19]. Thereafter, the manual graphical rebuilding using O [20] was performed and the structure was subsequently refined using CNS. The final model was refined by the individual B-factor refinement at 2.04 Å resolution. The atomic coordinates of the domain structure have been deposited in the Protein Data Bank with an accession code of 1iu9. The statistics for structure determination and refinement are also shown in Table 1.

3. Results

3.1. Pseudo symmetry between two domains of *P. AspR*

The three-dimensional structure of the *P. AspR* monomer consists of two compact α/β domains; the active site of this enzyme is located in the cleft between the two domains where two strictly conserved cysteine residues serve as a catalytic acid–base (Fig. 1A) [14]. A pseudo mirror symmetry has also been identified in the active site, by which the active site could be regarded as two moieties, each of which selectively binds one enantiomeric substrate [14].

Structurally, the two *P. AspR* domains are hinged by domain swapping with two α-helices, α7 and α13. Because helices α7 and α13 are located at a pair of perfectly superimposable positions, superimpositions of the two *P. AspR* domains

before and after domain swapping give essentially the same RMS deviation of Cα atoms of 1.9 Å, based on the secondary structures, which implies a two-fold pseudo symmetry between the two domains (Fig. 1A). Sequence alignment between these two domains revealed 26% identity and 36% similarity (Fig. 1C). A relatively large structural difference between the two domains can be observed at the regions around a short α-helix (α2) in the N-domain and around a π-helix (π1) in the C-domain.

The C-terminal moieties in both domains, residues 59–111 and residues 172–224, share very high sequence homology (40% identity and 47% similarity) (Fig. 1C). In particular, two α/β motifs containing the catalytic cysteine residues have higher identity (> 50%), and the positions of both catalytic cysteine residues are conserved in the primary structure (Fig. 1C) and in the three-dimensional structure (Fig. 1B). A superimposition of the two α/β motifs reveals a much lower RMS deviation of Cα atoms of approximately 1.1 Å (Fig. 1B).

3.2. Structure of the SC-domain

The SC-domain contains residues 103–213, corresponding to the sequence posterior to domain swapping. The SC-domain crystallizes as a monomer in the space group of *P*₂₁₂₁₂₁. In the crystals, each asymmetric unit contains one molecule of the SC-domain. All amino acid residues could be located from the electron density map at 2.04 Å resolution. The overall structure of the SC-domain is shown in Fig. 2A. The SC-domain structure has the same folding as that of the C-domain in the intact enzyme, containing a four-stranded parallel β-sheet flanked by six α-helices from two sides. In both cases, all secondary structure elements, including the π1 helix as well as the C- and N-termini, showed no significant structural differences. A superimposition of the SC-domain and C-domain reveals an RMS deviation of Cα atoms of 0.50 Å (Fig. 2B), implying that the domain structure is highly stable. A total of 56 water molecules were found to be located in the SC-domain structure, with most having spatial positions that are also found in the intact enzyme structure. Additionally, a calcium was found in the intermolecular interface, by which Asp204 (δO1 and δO2) is bridged to Gly115' (O) in the neighboring molecule.

Table 1
Statistics of data collection and structure determination for the SC-domain of *P. AspR*

Data collection	
Beamline	BL44B2, SPring-8
Temperature (K)	100
Unit cell (Å)	<i>a</i> , 36.83
	<i>b</i> , 45.33
	<i>c</i> , 57.37
Space group	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁
Resolution limits (Å) ^a	100–2.04 (2.15–2.04)
Total observed reflections	57 302
Unique reflections	6508
Completeness (%) ^a	99.3 (96.3)
<i>R</i> _{merge} (%) ^a	10.8 (25.2)
Structure refinement	
<i>R</i> -factor (%)	22.6
<i>R</i> -free (%)	27.1
RMS derivative of bond lengths (Å)	0.005
RMS derivative of bond angles (°)	1.07

^aValues corresponding to the highest resolution shell in the parentheses.

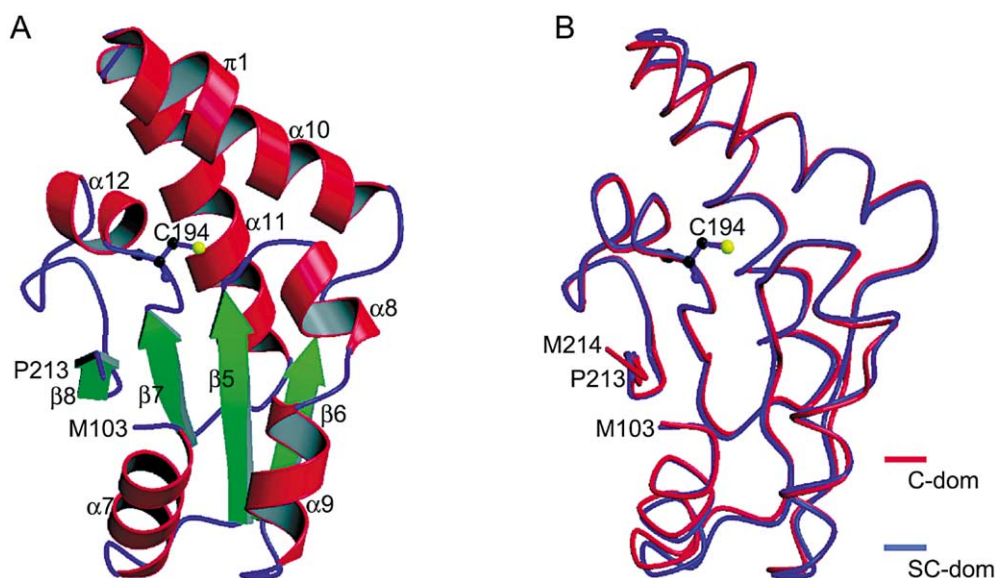


Fig. 2. The structures of the C-terminal domains (posterior to domain swapping). A: A ribbon representation of the SC-domain. Cys194 is shown in a ball-and-stick model, and secondary structure elements are labeled. B: A superimposition of structures of the C-domain and the SC-domain. The C-domain and the SC-domain are colored red and blue, respectively. Panels A and B are drawn with the same view orientation.

The structural conservation between the C-domain and SC-domain not only supports the hypothesis that this two-domain enzyme evolved from an ancestral domain, but also provides an outline of the topological image for the ancestral domain.

4.2. PLP-independent AARs

The structures of another two PLP-independent AARs, *H. DapE* [12] and *A. GluR* [13], are composed of two α/β domains. Although the two domains of *H. DapE* do not show remarkable sequence identity, they share high structural similarity with a superimposition of 1.78 Å based on secondary structure elements [12]. Interestingly, domain duplication was also found in *H. DapE*, but it was at a different point from that in *P. AspR*, with the swapping in this enzyme being related by a pair of β -strands instead of α -helices. In the case of *A. GluR*, the two domains are not equal in size [13]. As expected from the similarity of their substrates and the conservation of their active sites, *A. GluR* has a similar three-dimensional structure to that of *P. AspR*. It has been shown by SCOP [25] that *A. GluR* and *P. AspR* belong to the same family and that *A. GluR* has an extended C-terminus in comparison with *P. AspR*. If approximately 40 residues are truncated at the C-terminal site of *A. GluR*, the two domains become equal in size, with each containing a four-stranded β -sheet similar to those of *P. AspR*. Two *A. GluR* domains (posterior to the truncation of the C-terminus) could also be superimposed with an RMS deviation of ~ 2.0 Å based on secondary structure elements, with the domain duplication having similar structural appearance to that in *P. AspR*. All three of these enzymes have their active sites between their two domains, and a pair of cysteine residues is responsible for the catalysis as a general acid–base, suggesting a similar catalytic mechanism for these racemases.

Although neither *H. DapE* nor *A. GluR* showed significant sequence homology between their two domains, these two enzymes commonly have structural and functional similarity

of their two domains. The low homology in the primary structure between the two domains does not exclude the possibility of an evolutionary relationship similar to that proposed for *P. AspR*, as the three-dimensional structure and function are more conservative than the amino acid sequence in the process of the evolutionary selection [21,26]. This suggests that these PLP-independent AARs share a common evolutionary mechanism, as shown in Fig. 3. In particular, *AspR* and *GluR* are expected to have evolved from a common ancestor

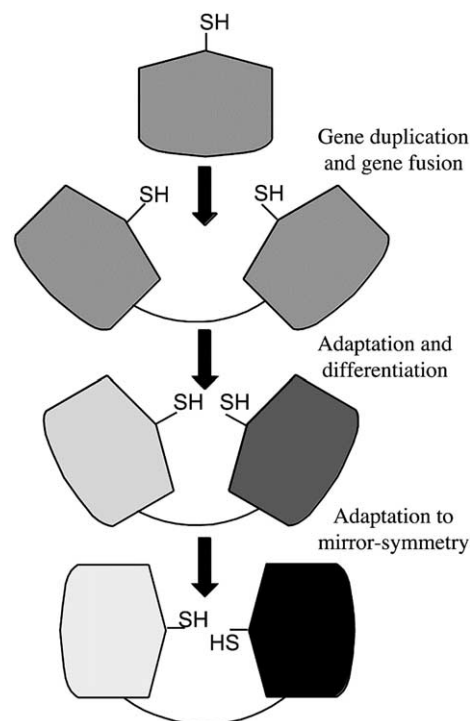


Fig. 3. An evolutionary mechanism proposed for the PLP-independent AARs.

based on their sequence homology and their substrate similarity [27], also providing evidence of descent based on the same evolutionary mechanism.

4.3. PLP-dependent AARs and other non-amino acid racemases

PLP-dependent AARs catalyze the reactions in a different way, with a lysine residue being utilized as a catalytic group and PLP as a cofactor [6]. This kind of enzyme, together with mandelate racemase, a metal ion-dependent α -hydroxyl-carboxylate racemase [28], has a basic $(\beta/\alpha)_8$ barrel domain functioning as a catalytic domain [11,29]. It has been suggested by structural and chemical evidence that these $(\beta/\alpha)_8$ barrel domains have diverged from a common ancestor [30]. In addition, it has been suggested that the $(\beta/\alpha)_8$ barrel domains have evolved by gene duplication and gene fusion from an ancestral half-barrel, which consisted of four β/α units stabilized by dimerization [31–33].

Sequence alignment did not show any marked homology between the α/β domains of the PLP-independent AARs and the half-barrel of PLP-dependent AARs, but the β -sheet in both the α/β domain (β_2 – β_1 – β_3 – β_4) and the half-barrel (β_1 – β_2 – β_3 – β_4) were parallel and four-stranded. No current evidence suggested that the α/β domains of the PLP-independent AARs and the half-barrel of PLP-dependent AARs have evolved from a common ancestor, although they are nearly equal in size and have parallel four-stranded β -sheets.

Structural and functional study of methylmalonyl-coenzyme A epimerase from *Propionibacterium shermanii* also shows two well-superimposable domains without any marked sequence homology [34]. It is also thought that this enzyme evolved from an ancestral domain by gene duplication and domain swapping, providing additional evidence of gene duplication, gene fusion, and domain swapping in the evolution of racemases.

Although it is evident that gene duplication and gene fusion are part of a general phenomenon found in many classes of racemases, more studies are necessary to realize the relationship between the adaptation of racemases to the mirror symmetry of racemizations and the role of gene duplication and gene fusion.

Acknowledgements: This work was supported in part by the 'Research for the Future' Program to K.M. (97L00501) from the Japan Society for the Promotion of Science (JSPS), and by a Grant-in-Aid for Scientific Research to M.Y. (09460053) from the Ministry of Education, Science, Sports and Culture of Japan. We thank Dr. S. Adachi of SPring-8 for his help in the intensity data collection. We also thank Drs. Y. Kawarabayasi and H. Kikuchi at the National Institute of Technology and Evaluation in Japan for their help in gene expression.

References

- [1] Dunn, B.M., Gustchina, A., Wlodawer, A. and Kay, J. (1994) *Methods Enzymol.* 241, 255–301.
- [2] Milton, R.C. de L., Milton, S.C.F. and Kent, S.B.H. (1992) *Science* 256, 1445–1448.
- [3] Miranda, J.D., Santoro, A., Engelen, S. and Wolosker, H. (2000) *Gene* 256, 183–188.
- [4] Wolosker, H., D'Aniello, A. and Synder, S.H. (2000) *Neuroscience* 100, 183–189.
- [5] Walsh, C.T. (1989) *J. Biol. Chem.* 264, 2393–2396.
- [6] Soda, K. and Esaki, N. (1994) *Pure Appl. Chem.* 66, 709–714.
- [7] Johnston, M.M. and Diven, W.F. (1969) *J. Biol. Chem.* 244, 5414–5420.
- [8] Nakajima, N., Tanizawa, K., Tanaka, H. and Soda, K. (1986) *Agric. Chem. Biol.* 50, 2823–2830.
- [9] Higgins, W., Tardif, C., Richaud, C., Krivanek, M.A. and Cardin, A. (1989) *Eur. J. Biochem.* 186, 137–143.
- [10] Alberly, W.J. and Knowled, J.R. (1986) *Biochemistry* 25, 2572–2577.
- [11] Shaw, J.P., Petsko, G.A. and Ringe, D. (1997) *Biochemistry* 36, 1329–1342.
- [12] Cirilli, M., Zheng, R., Scapin, G. and Blanchard, J.S. (1998) *Biochemistry* 37, 16452–16458.
- [13] Hwang, K.Y., Cho, C.-S., Kim, S.S., Sung, H.-C., Yu, Y.G. and Cho, Y. (1999) *Nat. Struct. Biol.* 6, 422–426.
- [14] Liu, L., Iwata, K., Kita, A., Kawarabayasi, Y., Yohda, M. and Miki, K. (2002) *J. Mol. Biol.* 319, 479–489.
- [15] Liu, L., Iwata, K., Kawarabayasi, Y., Kikuchi, H., Kita, A., Yohda, M. and Miki, K. (2001) *Acta Crystallogr. D* 57, 1674–1676.
- [16] Leslie, A.G.W. (1990) in: *Crystallographic Computing*, Oxford University Press, London.
- [17] Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. Sect. D* 50, 760–763.
- [18] Navaza, J. (1994) *Acta Crystallogr. Sect. A* 50, 157–163.
- [19] Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr. Sect. D* 54, 905–921.
- [20] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* 47, 110–119.
- [21] Murzin, A.G. (1998) *Curr. Opin. Struct. Biol.* 8, 380–387.
- [22] Doolittle, R.F. (1995) *Annu. Rev. Biochem.* 64, 287–314.
- [23] Ohta, T. (2000) *Philos. Trans. R. Soc. London B Biol. Sci.* 355, 1623–1626.
- [24] Miozzari, G.F. and Yanofsky, C. (1979) *Nature* 277, 486–489.
- [25] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) *J. Mol. Biol.* 247, 536–540.
- [26] Dietmann, S. and Holm, L. (2001) *Nat. Struct. Biol.* 8, 953–957.
- [27] Choi, S.-Y., Esaki, N., Ashiuchi, M., Yoshimura, T. and Soda, K. (1994) *Proc. Natl. Acad. Sci. USA* 91, 10144–10147.
- [28] Neidhart, D.J., Howell, P.L., Petsko, G.A., Powers, V.M., Li, R., Kenyon, G.L. and Gerlt, J.A. (1991) *Biochemistry* 30, 9264–9273.
- [29] Gerlt, J.A., Kenyon, G.L., Kozarich, J.W., Neidhart, D.J., Petsko, G.A. and Powers, V.M. (1992) *Curr. Opin. Struct. Biol.* 2, 736–742.
- [30] Farber, G.K. and Petsko, G.A. (1990) *Trends Biol. Sci.* 15, 228–234.
- [31] Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. and Wilmanns, M. (2000) *Science* 289, 1546–1550.
- [32] Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. and Sterner, R. (2001) *Nat. Struct. Biol.* 8, 32–36.
- [33] Hocker, B., Schmidt, S. and Sterner, R. (2002) *FEBS Lett.* 510, 133–135.
- [34] McCarthy, A.A., Baker, H.M., Shewry, S.C., Patchett, M.L. and Baker, E.N. (2001) *Structure* 9, 637–646.
- [35] Kraulis, P. (1991) *J. Appl. Crystallogr.* 24, 946–950.
- [36] Merritt, E.A. and Bacon, D.J. (1997) *Methods Enzymol.* 277, 505–524.