Crystal structure of a caricain D158E mutant in complex with E-64

Nikolaos A. Katerelos, Mark A.J. Taylor, Mandy Scott, Peter W. Goodenough, Richard W. Pickersgill*

Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, UK

Received 15 April 1996; revised version received 17 June 1996

Abstract The structure of the D158E mutant of caricain (previously known as papaya protease omega) in complex with E-64 has been determined at 2.0 Å resolution (overall *R* factor 19.3%). The structure reveals that the substituted glutamate makes the same pattern of hydrogen bonds as the aspartate in native caricain. This was not anticipated since in the native structure there is insufficient room to accommodate the glutamate side chain. The glutamate is accommodated in the mutant by a local expansion of the structure demonstrating that small structural changes are responsible for the change in activity.

Key words: Kinetically influential mutant; Cysteine protease; Site directed mutagenesis; X-ray crystallography; Enzyme-inhibitor complex

1. Introduction

Since the structure of the plant cysteine protease papain was solved by Drenth et al. in 1968 [1] there has been considerable discussion of the role of Asp-158 (papain numbering) in the activity of the cysteine proteases. The active site cysteine and histidine have clearly identifiable roles in hydrolysis of the peptide bond the former acting as the nucleophile and the latter as the general acid/base. It is now generally accepted that Cys-25 and His-159 exist in ionized form in the active site [2–5]. Caricain (previously known as papaya protease omega) is the most basic cysteine protease from the latex of *Carica papaya* with 69.5% sequence identity with papain [6]. Refined structures of the papain superfamily are: papain [7], actinidin [8], human liver cathepsin B [9], caricain [10], cruzain [11] and glycyl endopeptidase [12]. There are also a number of cysteine protease structures determined in complex with the inhibitor E-64 (see Section 2): papain/E-64 [13], papain/E-64-c [14] and actinidin/E-64 [15]. The reader is referred to Brocklehurst ([16] and references contained within) for a review of the kinetics of papain and caricain.

The electrostatic component of the interaction between the side chain of Asp-158 and Cys-25 in papain was calculated to be 0.7 pK_a units [17,18] and for caricain 0.8 pK_a units [19]. Site-directed mutants of papain at position 158 have been reported by Ménard et al. [20,21]. In the first of these papers Asp-158 was demonstrated not to be an essential catalytic residue in papain. The second paper reported three additional mutants at 158, namely Gly, Ala and Glu. The k_{cat}/K_m values ranked as follows: Asp > Asn > Glu > Gly from which they concluded that the hydrogen bonds made by residue 158 are important in stabilizing the ion pair since Asn could make 'similar' hydrogen bonds to the Asp but the Glu could not. These authors clearly consider that Asn could fit neatly into the unperturbed structure whilst the Glu side chain is too long to fit. Taylor et al. [19] report the effect of substituting Ala, Asn and Glu for Asp-158 in caricain and highlight the potential multiple ionizations involved in activity. The k_{cat}/K_m values for the caricain mutants rank Asp > Glu > Asn > Ala. This suggests that for caricain the Glu-158 mutant is the most similar in structure as it has the highest activity. Clearly, the same hydrogen bonds could be made by the glutamate if it could be accommodated within the structure.

To establish the position of the glutamate side chain at position 158 in caricain we have determined the structure by X-ray crystallography. We discuss this structure and activity, and the implications for understanding the effect of mutations at the corresponding position in papain.

2. Materials and methods

2.1. Production of the caricain mutant D158E/E-64 complex

The expression of procaricain, without the 26 amino acid signal sequence, in *E. coli* has been described [22] as has the site-directed mutagenesis to produce the D158E mutant [19]. The protein was produced as inclusion bodies which were solubilized, refolded and processed to mature enzyme as described previously [19,23]. The activated caricain mutant was then reacted with a molar excess of the specific inhibitor L-*trans*-epoxysuccinyl-L-leucylamido 4-guanidinobutane ('E-64', Fig. 1) extracted from cultures of *Aspergillus japonicus* [24,25] obtained from the Peptide Institute (Japan). The E-64 was solubilized in ethanol and any non-bound E-64 was removed by washing over a dialysis membrane.

2.2. Preparation and crystallization

The crystals were grown by the hanging drop method at 18°C. The D158E mutant complexed with E-64 was concentrated to approx. 13 mg/ml in 10 mM Tris pH 8.0. The drops contained 3 µl of the protein complex and 3 µl of the reservoir solution which contained 29% PEG 1000 and 0.3 M $(NH_4)_2SO_4$. The crystals were rather small needles of dimensions $0.15 \times 0.05 \times 0.05$ mm belonging to the monoclinic space group C2 with $a = 53.45$ Å, $b = 65.33$ Å, $c = 64.37$ Å and $\beta = 111.60°$ with one molecule in the asymmetric unit.

2.3. Data collection and processing

Data were collected from a single crystal of the caricain D158E mutant in a complex with E-64 using synchrotron radiation at the Photon Factory, KEK, Tsukuba, Japan. These data were collected using beam line BL-6A2 equipped with a Weissenberg camera [26]. The cassette used was of radius 429.7 mm with two 20×40 cm image plates placed horizontally one above the other to enable efficient collection of data to at least 2.2 Å. The rotation range used was 10.5°, with 0.5° overlap with the preceding range, and the coupling constant was 1.5°/mm. A single exposure consisted of 30 repeats of the oscillation range at a rate of 2°/s. The total oscillation range covered was 180.5° corresponding to 36 image plates. Oscillation images of 1.0° were used to determine the crystal parameters and the orientation of the crystal. The data were reduced using the program DENZO [27] running on a Silicon Graphics Indigo^2 extreme workstation and CCP4 programs [28] for the final internal scaling and merging of the data. The 42446 measurements were reduced to 12010 independent reflec-

*Corresponding author. Fax: (44) (734) 267917.
E-mail: richard.pickersgill@bbsrc.ac.uk

PII S0014-5793(96)00697-7

tions with an R_{merge} of 6.0%. The internal agreement and completeness of the data as a function of resolution are presented in Table 1.

2.4. Molecular replacement

The X-PLOR program package, running on a VAX 4400, was used to solve the rotation and translation functions [29]. The search model was native caricain [10] with the inhibiting mercury atom removed. Data in the range 15.0–4.0 Å were used in the rotation function; the solution (202.8, 85.0, 265.8) was at a peak height of 5.1 σ above the next highest peak and 7.7 σ above the mean. The translation function solution (−0.7, 0.0, 22.4) was at 6.9 σ above the next highest peak and 8.9 σ above the mean. Rigid body refinement using data in the range 10.0–2.0 Å reduced the R factor from 0.462 to 0.387. The first $2F_{obs}-F_{calc}$ map calculated using the rigid body refined molecular replacement solution showed clear evidence of E-64 bound to the active site cysteine.

2.5. Refinement

E-64 was fitted into this electron density map and glutamate was substituted in place of aspartate at position 158. Electron density fitting was performed using FRODO [30] run on an Evans and Sutherland PS390 graphics system. At this stage a new entry for E-64 was introduced into the X-PLOR topology and parameter files. Cycles of positional and temperature factor refinement using X-PLOR reduced the R factor for protein, water and E-64 to 0.211. The automated refinement procedure [31] was then used in restrained mode using protein and water. 50 cycles of restrained refinement reduced the R factor from 0.202 to 0.153. The resulting $3F_{obs}-2F_{calc}$ map again showed clear evidence for E-64 with the additional water molecules mimicking the shape of E-64. Some elongated density was suggestive of ethanol molecules. ARP is a relatively unbiased procedure, but to ensure that the position of the E-64 and Glu-158 were not biased by the previous modelling a simulated annealed omit map was calculated using X-PLOR. This map again gave clear density for the part of E-64 and density corresponding to the glutamate side chain (Fig. 2). This side chain was clearly not an aspartate as seen in the native enzyme. The glutamate, E-64, 94 water molecules and one ethanol molecule (with good density in the $2F_{obs}-F_{calc}$ map) were included in a final cycle of positional and B factor refinement using X-PLOR. The final model comprises 216 residues, E-64 and 95 solvent molecules for which the R factor is 0.193. All residues lie in the most favoured (87.6%) or additional allowed (12.4%) regions of the Ramachandran plot, as defined by PROCHECK [32]. The overall G factor for the refined structure is +0.17. Atomic coordinates and structure factor amplitudes have been submitted to the protein data bank (1MEG and R1MEGSF).

3. Results

3.1. Structure of the caricain D158E/E-64 complex

The structure has been refined to reasonable stereochemistry and R factor including E-64, 94 water molecules and one ethanol in the final model. The electron density maps show the epoxy oxygen O3 in E-64 becomes a hydroxyl bound to C3 with the C2 atom covalently bound to the sulphur (SG) atom of cysteine 25. The SG to C2 bond length is 1.82 Å.

The RMSD* for 216 alpha carbon atoms between the

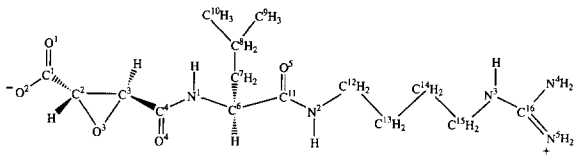

Fig. 1. Schematic drawing of the structure of E-64 to show the atom numbering used in the text and tables.

D158E mutant in complex with E-64 and caricain is 0.23 Å. Widening of the active site cleft on substrate or inhibitor binding was originally suggested by Drenth et al. in 1976 [33]. Active site widening on E-64 binding has been demonstrated in papain and actinidin [13–15] and overall the structure of the caricain D158E/E-64 complex expands slightly in response to both the mutation and E-64 binding.

The glutamate at position 158 in the caricain D158E mutant is well defined in the final electron density map and in 'model bias free' maps (Fig. 2). The side chain of the substituted glutamate forms the same hydrogen bonds to Lys-137, Ser-136 and the main chain amides of 136 and 137 as the aspartate side chain in native caricain. The hydrogen bonds and distances are given in Table 4. The structure opens up slightly to accommodate the glutamate side chain as indicated by the alpha-carbon distances from 158 to 136 and from 158 to 137 increasing from 6.2 and 7.1 Å in native caricain to 6.6 and 7.7 Å in the D158E mutant. The conformation of Asp-158 in caricain is similar to Ponder's rotamer 3 which accounts for 16% of conformers and the glutamate of the mutant is closest to rotamer 7 which is only 5% of conformers (Fig. 3) [34].

The mean temperature factors for protein and solvent are 10.2 and 22.5 Å^2, respectively (Table 2). The mean temperature factor for the 18 well defined atoms in E-64, from C1 to C12 (Fig. 1), is 15.1 Å^2. Beyond C12, which has a temperature factor of 25.1 Å^2, there is no electron density at the error level in the map and the temperature factors increase by approx. 5 Å^2 for each additional carbon such that C16 has a temperature factor of 46.3 Å^2. Below the error level in the electron density map there is contiguous density suggestive of two conformations for the non-defined part of E-64.

4. Discussion

Table 3 lists the hydrogen bonds between E-64 and caricain D158E, and in parentheses the distances for the papain/E-64 complex are given [13]. There are few surprises here; the contacts for O1 to N2 are the same as for the papain/E-64 com-

Fig. 2. Stereo picture of the electron density for Glu-158 in the unbiased slowcool map (the electron density is contoured at 1 σ). This is an unbiased map resulting from a simulated annealed run with E-64 and the Glu-158 omitted.

Fig. 3. Stereo pictures of the hydrogen bonds involving the side chain at position 158. (A) Native caricain with D158 hydrogen bonding to the main chain amides of 136 and 137 and the side chains of Ser 136 and Lys 137. (B) D158E caricain mutant in complex with E-64 showing E158 making hydrogen bonds to the main chain amides of 136 and 137, Ser 136 and Lys 137. (C) Papain with D158 making hydrogen bonds to the main chain amides of 136 and 137. See Table 4 for details of the hydrogen bond distances.

Fig. 4. Stereo view of the electron density for E-64 calculated using F_{obs}-F_{calc}, α_{calc} with E-64 omitted from F_{calc}, α_{calc} (the electron density map is contoured at 1σ).

Fig. 2

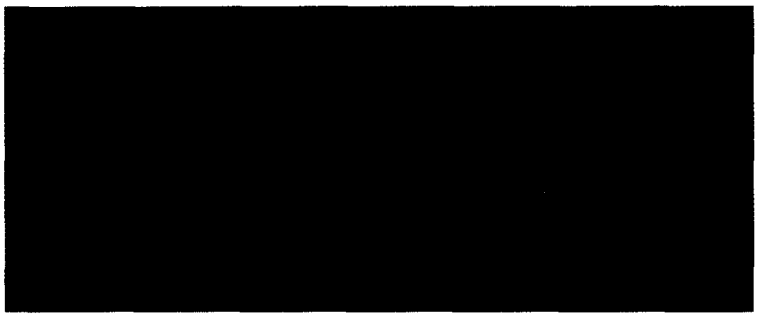

Fig. 3

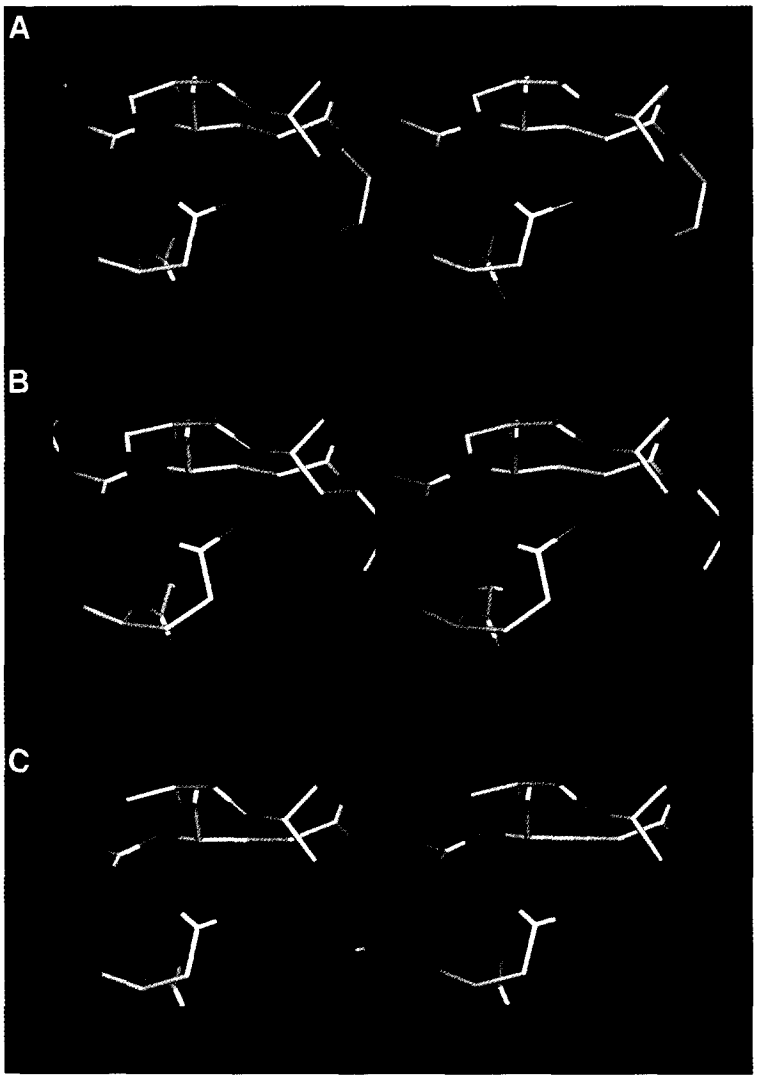

Fig. 4

Table 1
Summary of the data quality and completeness as a function of resolution.

N	Dmin Å	$R_{merge}(I)$	I/sigma	Completeness (%)	Multiplicity
1	7.72	0.041	15.8	84.4	3.2
2	5.47	0.039	16.9	98.8	3.5
3	4.47	0.039	15.8	96.3	3.5
4	3.87	0.042	15.4	97.2	3.5
5	3.46	0.047	14.2	98.5	3.6
6	3.16	0.052	12.9	99.0	3.6
7	2.93	0.059	10.6	98.7	3.6
8	2.74	0.066	10.5	99.1	3.6
9	2.58	0.075	9.0	98.8	3.6
10	2.45	0.079	8.4	97.4	3.7
11	2.34	0.085	8.1	94.6	3.6
12	2.24	0.094	7.4	96.8	3.6
13	2.15	0.108	6.3	75.6	3.3
14	2.07	0.117	5.9	54.2	3.4
15	2.00	0.138	4.8	44.1	3.3
Totals		0.060	10.4	86.0	3.5

There were 42,446 measuremets of 12010 unique reflections.
$R_{merge}(I) = \Sigma_h \Sigma_i (<I(h)> - I(h)_i) / \Sigma_h \Sigma_i I(h)$
where there are i equivalents with average intensity $<I(h)>$.

plex and the leucyl moiety occupies the S2 subsite. The major difference is that the 4-guanidinobutane moiety is not seen in the caricain D158E structure. In the papain complex N4 of E-64 makes a hydrogen bond with the hydroxyl of Tyr-61 and N5 with the hydroxyl of Tyr-67. Residue 61 in caricain is a histidine so the hydrogen bond to N4 cannot be made; this may explain why the 4-guanidinobutane moiety is not seen in the caricain complex.

Comparison of the D158E caricain/E-64 complex with the caricain structure shows that the structure expands in response to both E-64 binding and to the D158E mutation. The hydrogen bonds made by the glutamate in the mutant structure are the same as those made by the aspartate in the native structure. In both proteins the side chain at position 158 makes hydrogen bonds to the main chain amides of 136 and 137 and to the side chains of Ser-136 and Lys-137 (Fig. 3A,B). It was not clear that these hydrogen bonds would be conserved in the glutamate mutant since without local expansion the glutamate side chain would be too long to fit. The alternative was that the glutamate might be more solvent exposed with water filling the cavity left by the movement of the side chain. The refined structure of the D158E mutant shows that substitution of glutamate for aspartate opens up the structure locally allowing the hydrogen bonds to be made. It could be argued that E-64 affects the conformation of the glutamate side chain and that in the absence of E-64 the glutamate moves out of the structure but this seems unlikely given the number and strength (distances) of the hydrogen bonds involved. The hydrogen bonding distances involving the side chain at position 158 are presented in Table 4. Structurally, the major effect of the mutation is not a major repositioning of the charge of 158 but a subtle shift in the polypeptide chain forming a part of one side of the active site and substrate binding cleft. The carboxylate of the glutamate is slightly reorientated towards the active site histidine. The subtle change in the position of the charge and the change in solvation of the active site groups due to the local expansion would then be expected to be responsible for modified pH activity profile [19].

Table 2
Stereochemical details of the final model

Resolution range (Å)	10.0–2.0
Number of reflections	11858
Number of atoms	1764
Number of protein atoms	1642
Number of E-64 atoms	25
Number of solvent molecules	95 (including 1 ethanol)
R value	0.193
Overall *G* factor	+0.17
Ramachandran plot	87.6% (allowed), 12.4% (additionally allowed)
RMSD[a] in bond length	0.012 Å
RMSD in bond angle	1.7°
RMSD in dihedral angle	25.6°
RMSD in improper angle	1.7°
Mean *B* factor for protein	10.2 Å^2
Mean *B* factor for solvent	20.7 Å^2
Mean *B* factor for E-64	15.2 Å^2 (18 atoms to C12) 22.5 Å^2 (all E-64 atoms)

[a]RMSD is root mean square deviation from the parameters of Engh and Huber [35].

Table 3
Hydrogen bonds between E64 and caricain

E-64	Caricain	Distance (Å)
O1	N(25)	3.01 (2.95)
O1	NE2(19)	2.74 (2.87)
O2	ND1(159)	2.88 (2.90)
O4	N(66)	2.78 (2.88)
N1	O(158)	3.27 (3.40)
N2	O(66)	2.94 (3.04)

The figures in parentheses are the corresponding distances in the papain-E-64 complex [13]. Atom names for E-64 are given in Fig. 1.

The effect of the same mutation (D158E) on the conformation of papain may be different since the aspartate is hydrogen bonded only by the main chain amides of 136 and 137 and not by the side chains of these residues since they are alanine (Fig. 3C). The glutamate could be accommodated with a similar local expansion to that seen in caricain or alternatively the side chain could occupy a more solvent exposed position as there are fewer hydrogen bonds to hold it in position. This caricain result opens up the possibility that the glutamate might be accommodated in a similar way in the mutant papain. The reversed order of k_{cat}/K_m for mutants Glu > Asn in caricain and Asn > Glu in papain [19,21] might be because it is easier to accommodate Asn in papain than in caricain without any change in structure as fewer hydrogen bonds between residue 158 and the rest of the protein are perturbed by the substitution.

Comparison of the caricain D158E/E-64 complex with the papain/E-64 complex shows conservation of hydrogen bonds close to the active site amino acids and the S2 subsite and a similar widening of the substrate binding cleft. The conservation of hydrogen bonds only breaks down around the S3 subsite where a sequence change in caricain (Tyr-61 in papain is a histidine in caricain) implies a change in hydrogen bonding pattern.

Acknowledgements: This work was funded by the BBSRC. We thank Professor Sakabe for the provision of beam time at the Photon Factory (KEK, Tsukuba, Japan). N.A.K. acknowledges funding from the Alexander S. Onassis Public Benefit Foundation. We thank Dr. Victor Lamzin for assistance running ARP.

Table 4
Hydrogen bonds involving the side chain of amino acid 158 in caricain, caricain D158E mutant and papain

Hydrogen bond	Caricain (D158) (Å)	D158E caricain E-64 complex (E158) (Å)	Papain (D158) (Å)
OD2(OE1)-NZ(137)	2.82	3.14	[a]
OD2(OE1)-N(137)	3.10	2.83	3.22
OD1(OE2)-N(136)	2.74	2.91	2.81
OD1(OE2)-OG(136)	2.73	2.80	[a]

[a]No side chain hydrogen bond since 136 and 137 are alanines in papain.

References

[1] Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M. and Wolthers, B.G. (1968) Nature 218, 929–932.
[2] Polgar, L. (1974) FEBS Lett. 47, 15–18.
[3] Lewis, S.D., Johnson, F.A. and Shafer, J.A. (1976) Biochemistry 15, 5009–5017.
[4] Lewis, S.D., Johnson, F.A. and Shafer, J.A. (1981) Biochemistry 20, 48–51.
[5] Sluyterman, L.A. and Wijdenes, J. (1976) Eur J Biochem 71, 383–391.
[6] Dubois, T., Kleinschmidt, T., Schneck, A.G., Looze, Y. and Braunitzer, G. (1988) Biol. Chem. Hoppe-Seyler 369, 741–754.
[7] Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A. and Drenth, J. (1984) J. Mol. Biol. 179, 233–256.
[8] Baker, E.N. and Dodson, E.J. (1980) Acta Crystallogr. A36, 559–572.
[9] Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T. Katunuma, N. and Bode, W. (1991) EMBO J. 10, 2321–2330.
[10] Pickersgill, R.W., Rizkallah, P., Harris, G.W. and Goodenough, P.W. (1991) Acta Crystallogr. B47, 766–771.
[11] McGrath, M.E., Eakin, A.E., Engel, J.C., McKerrow, J.H., Craik, C.S. and Fletterick, R.J. (1995) J. Mol. Biol. 247, 251–259.
[12] O' Hara, B.P., Hemmings, A.M., Buttle, D.J. and Pearl, L.H. (1995) Biochemistry 34, 13190–13195.
[13] Varughese, K.I., Ahmed, F.R., Carey, P.R., Hasnain, S., Huber, C.P. and Storer, A.C. (1989) Biochemistry 28, 1330–1332.
[14] Yamamoto, D., Matsumoto, K., Ohishi, H., Ishida, T., Inoue, M., Kitamura, K. and Mizuno, H. (1991) J. Biol. Chem. 266, 14771–14777.
[15] Varughese, K.I., Su, Y., Cromwell, D., Hasnain, S. and Xuong, N.-h. (1992) Biochemistry 31, 5172–5176.
[16] Brocklehurst, K. (1994) Protein Eng. 7, 291–299.
[17] Pickersgill, R.W. (1988) Protein Eng. 2, 247–248.
[18] Pickersgill, R.W., Sumner, I.G., Collins, M.E. and Goodenough, P.W. (1989) Biochem. J. 257, 310–312.
[19] Taylor, M.A.J., Baker, K.C., Connerton, I.F., Cummings, N.J., Harris, G.W., Henderson, I.M., Jones, S.T., Pickersgill, R.W., Sumner, I.G., Warwicker, J. and Goodenough, P.W. (1994) Protein Eng. 7, 1267–1276.
[20] Ménard, R., Khouri, H.E., Plouffe, C., Dupras, R., Ripoll, D., Vernet, T., Tessier, D.C., Laliberte, F., Thomas, D.Y. and Storer, A.C. (1990) Biochemistry 29, 6706–6713.
[21] Ménard, R., Khouri, H.E., Plouffe, Laflamme, P., C., Dupras, R., Vernet, T., Tessier, D.C., Thomas, D.Y. and Storer, A.C. (1991) Biochemistry 30, 5531–5538.
[22] Revell, D.F., Cummings, N.J., Baker, K.C., Collins, M.E., Taylor, M.A.J., Sumner, I.G., Pickersgill, R.W., Connerton, I.F. and Goodenough, P.W. (1993) Gene 127, 221–225.
[23] Taylor, M.A.J., Pratt, K.A., Revell, D.F., Baker, K.C., Sumner, I.G. and Goodenough, P.W. (1992) Protein Eng. 5, 455–459.
[24] Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J. and Tanaka, I. (1978) Agric. Biol. Chem. 42, 523–538.
[25] Hanada, K., Tamai, M., Ohmura, S., Sawada, J., Seki, T. and Tanaka, I. (1978) Agric. Biol. Chem. 42, 529–536.
[26] Sakabe, N., Ikemizu, S., Sakabe, K., Higashi, T., Nakagawa, A., Watanabe, N., Adachi, S. and Sasaki, K. (1995) Rev. Sci. Instrum. 66, 1276–1281.
[27] Otwinowski, Z. (1993) in: Data Collection and Processing (Sayer, L., Isaacs, N. and Bailey, S. eds.) pp. 56–62, Daresbury Laboratory, UK.
[28] Collaborative Computational Project 4 (1994) Acta Crystallogr. D50, 760–763.
[29] Brünger, A.T., Kuryjian, J. and Karplus, M. (1987) Science 235, 458–460.
[30] Jones, T.A. (1978) J. Appl. Crystallogr. 1, 268–272.
[31] Lamzin, V.S. and Wilson, K.S. (1993) Acta Crystallogr. D49, 129–147.
[32] Laskowski, R.A., MacArthur, M.W. and Thornton, J.M. (1994) in: From First Map to Final Model (Bailey, S., Hubbard, R. and Waller, D.A. eds.) pp. 149–159, Daresbury Laboratory, UK.
[33] Drenth, J., Kalk, K.H. and Swen, H.M. (1976) Biochemistry 15, 3731–3738.
[34] Ponder, J.W. and Richards, F.M. (1987) J. Mol. Biol. 193, 775–791.
[35] Engh, R.A. and Huber, R. (1991) Acta Crystallogr. A47, 392–400.