Acta Cryst. (1996). D**52**, 1125–1135

Subtilisin BPN′ at 1.6 Å Resolution: Analysis for Discrete Disorder and Comparison of Crystal Forms

TRAVIS GALLAGHER,[a] JOEL OLIVER,[b] RICHARD BOTT,[c] CHRISTIAN BETZEL[d] AND GARY L. GILLILAND[a]*

[a]*Center for Advanced Research in Biotechnology of the University of Maryland Biotechnology Institute and the National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, MD 20850, USA,* [b]*The Procter & Gamble Company, Corporate Research Division, Miami Valley Laboratories, PO Box 538707, Cincinnati, OH 45253-8707, USA,* [c]*Department of Biochemistry, Genencor International Inc., 180 Kimball Way, South San Francisco, CA 94080, USA, and* [d] *Institut fur Physiologische Chemie, DESY Hamburg, Gebaaeude 22a; Notkestrasse 85, 22603 Hamburg, Germany. E-mail: gary@ibm3.carb.nist.gov*

(*Received 15 August 1995; accepted 5 June 1996*)

Abstract

The three-dimensional structure of the serine protease subtilisin BPN′ (SBT) has been refined at 1.6 Å resolution in space group $C2$ to a final R value of 0.17. 17 regions of discrete disorder have been identified and analyzed. Two of these are dual-conformation peptide units; the remainder involve alternate rotamers of side chains either alone or in small clusters. The structure is compared with previously reported high-resolution models of SBT in two other space groups, $P2_12_12_1$ and $P2_1$. Apart from the surface, there are no significant variations in structure among the three crystal forms. Structural variations observed at the protein surface occur predominantly in regions of protein–protein contact. The crystal packing arrangements in the three space groups are compared.

1. Introduction

Members of the subtilase family of serine proteases are expressed in a diverse range of organisms from prokaryotes to humans (Siezen, de Vos, Leunissen & Dijkstra, 1991). The family's founders, subtilisins (E.C. 3.4.21.14), are secreted by several species of bacteria, particularly *Bacillis subtilis* and *B. amyloliquefaciens*, for the digestion of nutrients in their native soil milieu. Subtilisin BPN′ (SBT) is the name given to the enzyme from *B. amyloliquefaciens*. SBT has 275 residues and a molecular weight of 28 kDa. The same molecule, originally isolated from a different bacterial strain but shown to be identical, is also called subtilisin Novo. Subtilisin Carlsberg is a similar enzyme from *B. subtilis* with 83 residue differences from SBT.

Several decades of scientific work place SBT among the best studied enzymes. Its active site and mechanism share many features with the trypsin family of proteases, although the protein fold is different (Wright, 1972). Substrate specificity has been analyzed and rationally modified (Estell *et al.*, 1986). Numerous site-specific mutations that increase thermostability have been described (Pantoliano *et al.*, 1989; Gilliland, Gallagher & Bryan, 1996). The stronger of SBT's two calcium-binding sites binds Ca^{2+} with $K_a = 7 \times 10^6 M^{-1}$, and is important in folding and stability (Bryan *et al.*, 1992). The other site is weaker and less specific, consisting of two overlapping subsites for monovalent and divalent cations (Pantoliano *et al.*, 1988). Although the wild-type enzyme refolds poorly *in vitro*, demonstration of reversible unfolding in an engineered variant has opened the way for folding studies (Strausberg *et al.*, 1993). The enzyme is used both as a laboratory reagent and as a protein-degrading additive in industrial and household detergents.

The three-dimensional structure of subtilisin was first reported by Wright, Alden & Kraut (1969) at a resolution of 2.5 Å. The crystals belonged to space group $C2$ and were grown at pH 6 with 1.3 M ammonium sulfate as the precipitant. Drenth, Hol, Jansonius & Koekoek (1972) then reported the crystal structure of SBT to 2.8 Å resolution in a second crystal form, space group $P2_1$, using acetone as the precipitant. The resolution of this crystal form has recently been extended to 1.75 Å (Gilliland *et al.*, 1996). Bott *et al.* (1988) reported another crystal form for SBT at 1.8 Å resolution, in space group $P2_12_12_1$. These crystals were obtained under conditions similar to the original $C2$ form, using ammonium sulfate, but with added $CaCl_2$. One source of variation in SBT crystal growth is the use of different active-site inhibitors. Crystallization of SBT usually requires some method of inhibiting enzymatic activity to prevent autodigestion. The Wright *et al.* (1969) structure used phenylmethanesulfonyl fluoride (PMSF) while the Drenth *et al.* (1972) structure used diisopropylfluorophosphate; both of these agents bind covalently to the active-site serine. Bott *et al.* (1988) also used PMSF in their structure determination. In order to examine and compare the high-resolution structures of SBT in these three different crystal

forms, we sought to extend the resolution of the original $C2$ structure to match the others. The resulting $C2$ data resolution in fact surpassed the others and gave the additional bonus of revealing several discretely disordered regions in the SBT molecule.

2. Materials and methods

2.1. *Crystallization and structure determination.*

Crystals in space group $P2_1$ were grown from 55% acetone, pH 9, as described in Bryan *et al.* (1986). This method is a slight variation on the original $P2_1$ crystal conditions given by Drenth & Hol (1967). Data collection utilized a rotating-anode X-ray source and a Siemens multiwire proportional counter* as described in Gallagher, Bryan & Gilliland (1993). The $P2_1$ crystal form utilized a double mutant of SBT. The Ser → Cys mutation at position 221 reduces enzymatic activity by about four orders of magnitude to prevent autolysis (Abrahmsen *et al.*, 1991; Gallagher *et al.*, 1993). The Asn → Ser mutation at residue 218 enhances stability and improves crystallization behavior for this crystal form (Bryan *et al.*, 1986). Both mutations have been structurally characterized and found to produce only local changes at the mutation site (Gilliland *et al.*, 1996). The crystallization, structure solution and refinement of the $P2_12_12_1$ crystal form (2ST1) are described by Bott *et al.* (1988). These crystals were grown at pH 6 under conditions otherwise similar to those producing the $C2$ form (see below).

Crystals belonging to space group $C2$ were grown likewise by the hanging-drop vapor-diffusion method, under the following conditions. The protein solution contained 7.5 mg ml^{-1} SBT in 20 m*M* MES buffer, pH 7.0, with the inhibitor PMSF present at 2 m*M*. The reservoir contained 35–40% saturated ammonium sulfate, 0.5% polyethylene glycol 8000, 1% β-octylglucoside and 2% 2-methyl-2,4-pentanediol in 50 m*M* MES buffer, pH 7.0. Crystals grew from equal mixtures of protein and reservoir solutions within a week. Both the $C2$ and $P2_12_12_1$ crystal forms often grew in the same drop. Small plate-shaped seed crystals were transferred to new hanging drops for continued growth. The new drops were made by mixing protein solution as above but at pH 6.5 and with 5 m*M* $CaCl_2$ with reservoir solution consisting of 38–40% saturated ammonium sulfate and 1 m*M* sodium azide in either 50 m*M* MES at pH 6.5 or 50 m*M* MOPS at pH 7.0. Before macroseeding, the target drops were allowed to equilibrate for two hours. The transferred crystals would then grow for about a week. For the large crystals used in X-ray data collection, this macroseeding procedure was repeated three times. This method differs from the original $C2$ crystal conditions of Wright *et al.* (1969) mainly in the pH and buffer used. Data to 1.55 Å resolution were collected using synchrotron radiation from the EMBL beamline X31 at the DORIS storage ring, DESY, Hamburg, Germany. The synchrotron ring was operating in a single bunch mode at 3.7 GeV and a current of 50 to 70 mA with approximately five hours between injections. Data were recorded at a wavelength of 1.07 Å at 4 using the EMBL image-plate scanner at a crystal-to-plate distance of 240 mm. The final $R_{\rm merge}$, defined as $\sum|I-\langle I\rangle|/\sum I$, was 5.0%. The reflections were phased by use of the molecular replacement technique (Fitzgerald, 1988), using as probe the Bott *et al.* (1988) $P2_12_12_1$ structure.

2.2. *Refinement*

The refinements of all three crystal forms are summarized in Table 1. All refinement was carried out by the restrained-parameter least-squares procedure of *PROLSQ* (Hendrickson & Konnert, 1981). Refinements of the $P2_12_12_1$ and $P2_1$ structures are described elsewhere. The first part of the $C2$ refinement utilized the vectorized '*gprlsa*' version (Furey, Wang & Sax, 1982), running on a Cray supercomputer. After 15 rounds (about 250 refinement cycles), the crystallographic residual R had fallen from 0.275 to 0.170. A total of 175 water molecules were added to the model during this phase of the refinement. Subsequent rounds of refinement, involving discretely disordered residues with two or more modeled conformations, utilized the version of *PROFFT* modified by Finzel (1987) to incorporate the fast Fourier transform algorithms of Ten Eyck (1973), Agarwal (1978) and Sheriff (1987), to restrain contacts between different molecules. The *FRODO* (Jones, 1978) program was used on an Evans and Sutherland PS390 graphics system to examine F_o-F_c and $2F_o-F_c$ difference electron-density maps and omit maps, to adjust the model and to add solvent molecules. Solvent molecules were added to the model at chemically reasonable positions where difference density exceeded 3σ. Water molecules whose omit density fell below 3σ or whose $2F_o-F_c$ density fell below 1σ were removed from the model during the course of the refinement, while new water molecules were added as indicated by difference map maxima. The final model contains 194 water molecules which have been ranked by occ^2/B (James & Sielecki, 1983).

Discrete disorder was treated as follows. Residues whose difference electron density suggested multiple conformations were analyzed using omit maps. These

* Certain commercial equipment, instruments, and materials are identified in this paper in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment are necessarily the best available for the purpose.

Table 1. *Crystallographic and refinement data for subtilisin BPN′ in three crystal forms*

Crystal form	I	II	III
Space group	$C2$	$P2_12_12_1$	$P2_1$
Protein	Subtilisin BPN′	Subtilisin BPN′	Subtilisin BS3 (= subtilisin BPN′, mutant N218S,S221C)
Pecipitant	$(NH_4)_2SO_4$ (40%)	$(NH_4)_2SO_4$ (40%)	Acetone (55%)
Other crystal conditions	7.5 mg ml^{-1} protein pH 7.0 0.02 *M* MES 2 m*M* PMSF inhibitor	3–5 mg ml^{-1} protein pH 6.0 0.5 *M* sodium acetate 10 m*M* $CaCl_2$ 1 μ*M* PMSF inhibitor	8 mg ml^{-1} protein pH 9.0 0.05 *M* glycine–NaOH 1.0 m*M* DTT
Unit cell dimensions (Å,)	$a = 66.56$ $b = 54.15$ $c = 62.73$ $\beta = 91.87$	$a = 39.25$ $b = 72.86$ $c = 75.20$	$a = 41.77$ $b = 79.52$ $c = 37.22$ $\beta = 114.97$
V (Å^3)	2.26×10^5	2.15×10^5	1.12×10^5
Molecules per a.u.	1	1	1
V_M (Å^3 Da^{-1})	2.12	2.01	2.10
Refinement method	*PROLSQ*	*PROLSQ*	*PROLSQ*
Resolution (Å)	8.0–1.6	10.0–1.8	8.0–1.75
Reflections	26,220	15,433	19,788
Total atoms	2144	2099	2144
Solvent atoms	194	154	200
Final R value	0.17	0.14	0.15
R.m.s. ideal dist (Å)	0.019	0.015	0.017

maps enabled initial modeling of the conformers and provided the basis for a starting estimate of the partial occupancies. Using the alternate-chains feature of *PROLSQ*, the coordinates and temperature parameters of the individual conformers were refined, and difference and omit maps were used to evaluate the correctness of the occupancies. After 5–10 iterations, coordinates converged and difference electron density diminished to background levels. As the refinement progressed, additional discretely disordered residues were detected. The final model contains 18 alternate conformers of 15 side chains, and two alternate conformers of the main chain, for a total of 80 atoms with alternate positions modeled.

2.3. *Structure comparison*

Pairwise superpositions of the models were performed using all 275 Cα atoms as guides (Satow, Cohen, Padlan & Davies, 1986), and distances between equivalent atoms were calculated. Significant structural variations were identified using the method of Bott & Frane (1990). The basis of this method is a linear fit between the logarithm of the atoms' positional deviations and their mean B factors; thus equivalent structural variations are more significant when B factors are low. For each atom, the excess deviation over that expected for its mean B factor is normalized to give a 'Z score'. Residue Z scores are calculated as the sum of atomic Z scores divided by the square root of the number of atoms in the residues. Residue Z scores over 3.0 represent residues with significantly different positions between the models. This method provides an unbiased assessment of structural differences between identical or similar proteins.

3. Results

3.1. *The 1.6Å structure*

With the highest resolution reported for any subtilisin BPN′ structure, this model has been deposited in the Protein Data Bank (Bernstein *et al.*, 1977).* Effects of the active-site inhibitor, the 17 regions of discrete disorder, and crystal packing provide the only important deviations from previous descriptions of SBT. The high-affinity Ca^{2+} site (Site *A*) is as previously described (Gallagher *et al.*, 1993), while the weaker and less specific *B* site is occupied by a cation in the monovalent subsite (Pantoliano *et al.*, 1988). This cation appears to be an ammonium ion with full occupancy (based on the crystal growth conditions and the previously described promiscuity of the site); the site has octahedral coordination with five protein oxygen ligands (the carbonyl groups of 169, 171, 174, 195 and a carboxyl O atom of Asp197) and a water ligand (coordination lengths from 2.8 to 3.1 Å). Table 2 lists the 15 discretely disordered side chains, which in most cases involve variation of the χ_1 torsion angle among the usual staggered conformations (Bhat, Sasisekharan & Vijayan, 1979; Ponder & Richards,

* Atomic coodinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1SUP, R1SUPSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: PT0003).

1987). 40% of the disordered side chains belong to serine residues on or near the molecular surface. In the active site, discrete disorder is observed for residue His64 and the neighboring Ser63 (see Fig. 1). Here the alternate conformations of the histidine correlate with the partial occupancy (0.7) of the inhibitor. This 'static' disorder represents a superposition of molecules in the crystal, some of which have the inhibitor while some do not. Previous SBT structures have shown that PMS can cause the His64 side chain to be displaced (Wright *et al.*, 1969). Fig. 1 also shows that residue Ser63 is discretely disordered among the three possible rotamers; the occupancies are 0.7, 0.15 and 0.15. The main conformation, with $\chi_1 = 76$, is stabilized by a short hydrogen bond (2.5 Å) to the hydroxyl group of Tyr217; the other conformations have no hydrogen bonding. The threefold disorder cannot be attributed simply to presence/absence of the inhibitor and reflects some dynamic mobility in this side chain. The two conformations of residue Met222, like His64, are also correlated with the partial occupancy of the inhibitor. A few other dual-conformation residues near the active site, including Gln59, Thr71, and Lys94, probably represent static disorder. The three instances of triple-conformation residues, and several dual-conformation residues far from the active site, probably represent single-molecule dynamic disorder.

Two of the discretely disordered regions are dual-conformation peptides (residues 97–98 and 257–258; see Table 3 and Figs. 2 and 3). In both of these cases, the lesser occupied conformer is energetically less favorable, both according to standard Ramachandran analysis (Ramachandran, Ramakrishnan & Sasisekharan, 1963) and with regard to hydrogen bonding. In the case of the peptide 97–98 where the residues are in a loop that belongs to the enzyme's substrate-binding region, the effect is probably a result of the partial occupancy of the inhibitor. In the main conformation, there is a 2.9 Å hydrogen bond between 97 O and 100 N, stabilizing this exterior loop in a tight turn. The other disordered peptide, 257–258, is also on the molecular surface and the main conformation has two hydrogen bonds to water molecules, while the minor conformation has none. Interestingly, this peptide was observed to rotate about 90 when the enzyme was inactivated by peroxide (Bott *et al.*, 1988).

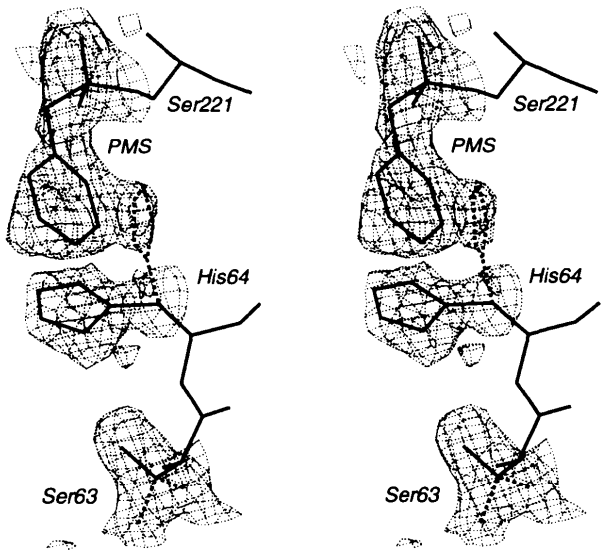

Fig. 1. Stereoview of omit electron density showing multiple conformations of residues in the active site of SBT in space group *C*2. The side chains of Ser63 and His64, along with the PMS inhibitor, were omitted from calculated structure factors. His64 has two conformations while Ser63 has three.

Table 2. *Side chains with multiple conformations in C2 subtilisin BPN′*

Residue type	Sequence number	Occupancy	Thermal parameters (Å²)*	Torsion angle† (°)	
Ser	3	0.72	17.9	χ_1	66
	3B	0.28	15.4		−58
Ser	24	0.70	14.8	χ_1	−60
	24B	0.30	13.6		−142
Gln‡	59	0.68	24.6	χ_1	−137
	59B	0.32	20.3		56
Ser§	63	0.70	14.0	χ_1	76
	63B	0.15	12.3		−156
	63C	0.15	12.3		−47
His‡§	64	0.80	18.9	χ_1	−63
	64B	0.20	14.5		−151
Thr‡	71	0.85	6.1	χ_1	−61
	71B	0.15	6.1		54
Ser	78	0.80	21.4	χ_1	−83
	78B	0.20	17.0		74
Lys‡	94	0.68	14.4	χ_3	−97
	94B	0.32	11.9		172
Lys	170	0.68	17.1	χ_3	−170
	170B	0.32	13.3		158
Val	177	0.85	6.9	χ_1	−175
	177B	0.15	6.4		−68
Met‡	222	0.70	11.6	χ_2	67
	222B	0.30	7.2		72
Ser	248	0.55	13.8	χ_1	53
	248B	0.45	13.3		−61
Ser	249	0.45	12.6	χ_1	−162
	249B	0.37	10.6		68
	249C	0.18	10.8		−56
Thr	255	0.75	11.3	χ_1	−64
	255B	0.25	10.9		52
Val	270	0.45	10.1	χ_1	70
	270B	0.38	11.0		−167
	270C	0.17	10.2		−57

* Averaged over the atoms with alternate positions. † The values of the first torsion angle differing by more than 10° are given. ‡ Active-site region; disorder probably due to PMS partial occupancy. § These residues are shown in Fig. 3.

3.2. *Structure comparison*

In the $P2_12_12_1$ structure (2ST1), use of a substoichiometric level of inhibitor resulted in very low occupancy of PMS in the active site; the final model features a

Table 3. *Alternate conformations of the protein backbone in C2 subtilisin BPN'*

Residues	Occupancy	Thermal parameters (Å^2)*	Conformational angles of peptide unit (°) φ_1	ψ_1	φ_2	ψ_2
97–98†	0.67	33.3	−80	−176	−59	−27
97B–98B†	0.33	32.8	−112	60	89	−57
257–258‡	0.72	14.1	−116	−34	176	−169
257B–258B‡	0.28	11.9	−102	−124	−94	−151

* Averaged over the five non-H atoms of the central peptide (C$\alpha\cdots$Cα). † See Fig. 2. ‡ See Fig. 3.

naked Ser221 and no inhibitor (Bott *et al.*, 1988). The $P2_1$ structure (1SUB) features the active-site mutation Ser → Cys at residue 221. This cysteine is found oxidized to the sulfone as in Abrahmsen *et al.* (1991). A second mutation previously shown to increase the enzyme's thermostability and to improve crystal quality, Asn-218-Ser, is also in the active-site region (Bryan *et al.*, 1986). This residue is 3.8 Å from Ser145 of a symmetry-related molecule. The weak cation-binding *B* site is occupied by a sodium ion in the monovalent subsite in this structure, which also includes one molecule of acetone, stacking against the side chain of Phe58.

Pairwise r.m.s. deviations among the three superimposed models are given in Table 4. The mean coordinate error for the *C*2 structure has been estimated at 0.15 Å based on Luzzati analysis (Luzzati, 1952), and between 0.15 and 0.20 Å for the other two structures. Deviation of Cα positions is close to that expected for identical structures with observational coordinate errors in this range. The method of Bott & Frane (1990), which takes into account the temperature factors of compared structures, was applied to determine which residues differ significantly among the three crystal forms (see *Materials and methods*). The foundation of this analysis is a linear regression that gives a best-fit line to the distribution of log deviations *versus* mean *B* factors [Fig. 4(*a*)]. The *y* intercept of this line gives the value of the log of the deviation expected for atoms with $B = 0$ Å^2. For the *C*2 *versus* $P2_12_12_1$ comparison, the intercept and slope are −2.51 and 0.09. For the *C*2 *versus* $P2_1$ comparison, the intercept and slope are −2.63 and 0.10. The corresponding values for the $P2_12_12_1$ *versus* $P2_1$ comparison are −2.27 and 0.10. Residue *Z* scores derived from the analysis report the degree of deviation of each residue [see Fig. 4(*b*)]. In each of the three pairwise comparisons, the highest residue *Z* scores correspond to residues whose χ_1 rotamer is altered by contact with another protein in the crystal. Leading the list are Tyr6 and Ser37, whose residue *Z* scores in the *C*2 *versus* $P2_12_12_1$ comparison are 8.4 and 7.7, respectively. Tyr6 provides a dramatic example of crystallization-induced conformational change. In both *C*2 and $P2_1$ this side chain occupies an inward-pointing rotamer, but in space group $P2_12_12_1$ it is displaced to an outward rotamer by a crystal contact (Fig. 5). In an interesting correlation involving protein stability, this alternate rotamer for Tyr6 was also observed in a mutant SBT lacking the calcium *A* site (Gallagher *et al.*, 1993). Determined in the $P2_1$ crystal form (where the crystal contact is absent), this calcium-free version of SBT had

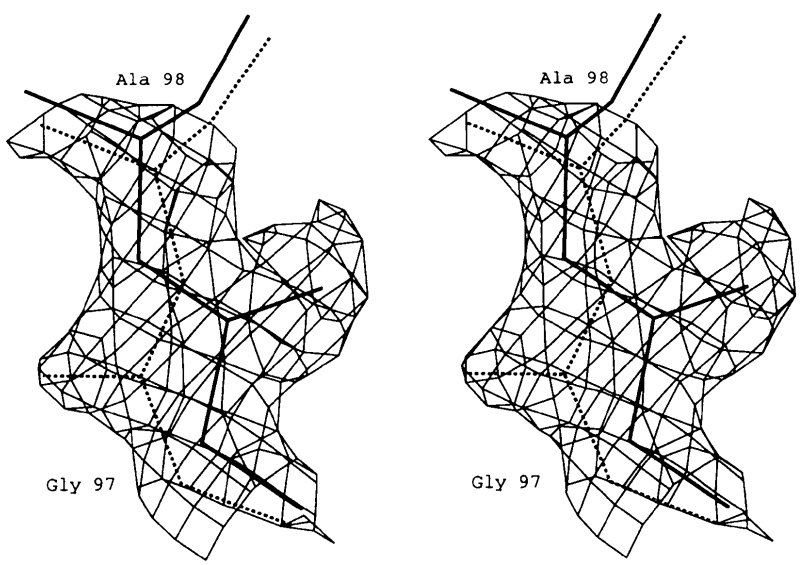

Fig. 2. Stereoview of omit electron density for the dual-conformation peptide 97–98 in SBT, space group *C*2. Omit maps are contoured at 1.5σ

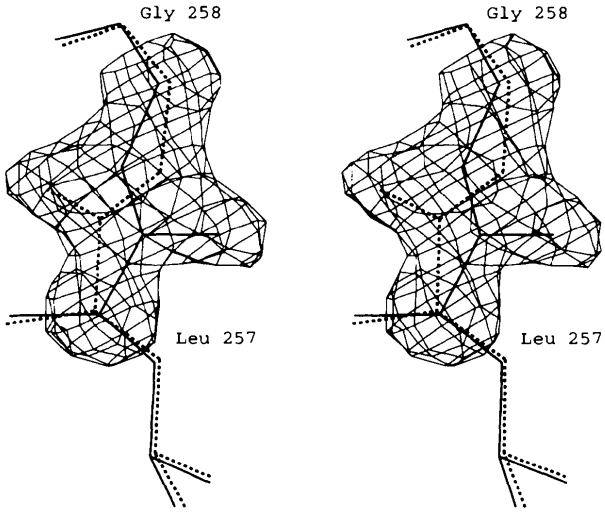

Fig. 3 Stereoview of omit electron density for the dual-conformation peptide 257–258 in SBT, space group $C2$.

Table 4. *R.m.s. differences (Å) among subtilisin BPN′ structures*

		$P2_12_12_1$	$P2_1$
$C2$	Cα	0.28	0.28
	Main chain*	0.32	0.33
	All atoms	0.69	0.71
	Excl. contacts†	0.40	0.56
$P2_12_12_1$	Cα	–	0.33
	Main chain	–	0.34
	All atoms	–	0.73
	Excl. contacts	–	0.43

*Main-chain calculations include N, Cα, C, and O atoms. †All atoms except those in contact residues. A contact residue is defined as a residue containing an atom within 4.0 Å of a different protein.

a disordered *N* terminus (residues 1–3 not seen), and the displacement of Tyr6 from the inward to the outward rotamer was attributed to this destabilization.

The active site is similar in the three structures. The PMS inhibitor in space group $C2$ causes the greatest changes, by altering the conformation of His64 and

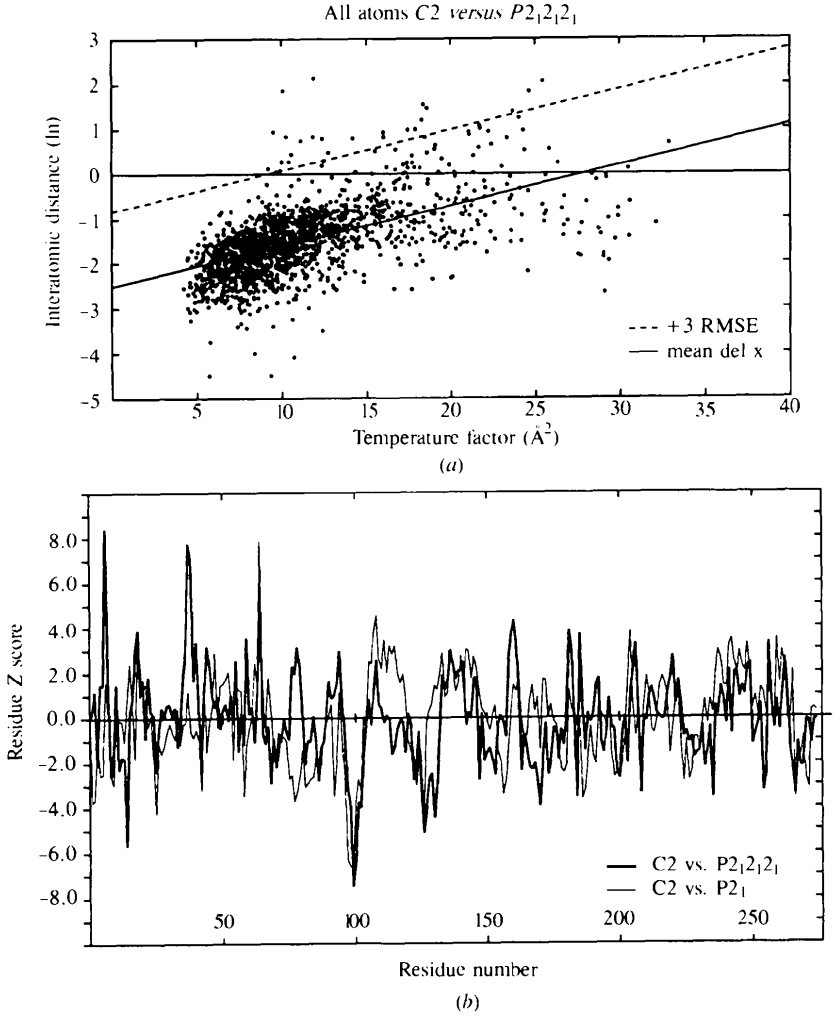

Fig. 4. Analysis of deviations among superimposed models, using the method of Bott & Frane (1990). In part (*a*) the logarithm of the distance between equivalent atoms is plotted against their mean temperature factors for all atoms comparing the $C2$ with the $P2_12_12_1$ crystal form. The linear fit to this distribution (mean value of distance as a function of temperature factor) is plotted as a solid line. The dashed line shows three time the r.m.s. error added to the mean value, corresponding to an atomic Z score of 3. In part (*b*) residue Z scores are plotted *versus* residue number for both the $C2$ *versus* $P2_12_12_1$ and the $C2$ *versus* $P2_1$ comparisons.

Met222 as noted above (see Fig. 1). Although the final occupancy for PMS is 0.7, it is well ordered, with the S atom close to the Cα position of the *P*1 substrate-binding site (McPhalen & James, 1988). The PMS phenyl ring points toward the *P*1′ binding site but is short and wide, resulting in the steric displacement of His64 as described above. The inhibitor displaces the two water molecules that are nearest to Ser221 in the $P2_12_12_1$ structure, one of which coincides with the position of the carbonyl O atom of the scissile bond and is hydrogen bonded to the side-chain N atom of Asn155. Two other water molecules, near the positions of the *P*2 and *P*3 site carbonyl O atoms, are also absent in the *C*2 structure though present in the other two. In these cases the cause of the displacement must be more subtle since the PMS does not sterically intrude on these sites. In space group $P2_1$ the two mutations, both in the active-site region, have only local effects on the protein structure. The positions of the two sulfonyl O atoms bound to Cys221 correspond closely with the two water sites in the naked active site of the $P2_12_12_1$ structure mentioned above, and one of them is hydrogen bonded to the side-chain N atom of Asn155.

3.3. *Conserved water*

Water molecules were grouped according to the distance to the nearest water site in the other crystal forms, and the mean *B* factor in each group computed. This analysis is similar to that performed by Bott *et al.* (1988) when they compared the $P2_12_12_1$ form with the older models of the other two forms. The distribution shows a group of water molecules with low *B* factors and close positional alignment. The histogram then has a minimum near 1 Å; there are few water molecules with this inter-water distance. Then the function increases again and the *B* factors are higher. The group of water molecules that align to within 1.0 Å are considered to represent conserved water sites, and their low *B* factors are consistent with their being firmly bound. This analysis identified 72 water sites that are conserved to within 1 Å among the three models. About a third of these are completely buried within the protein, while another third are bound in deep surface cavities with little solvent access. If the distance criterion is reduced to 0.5 Å, 45 water molecules are included. These 45 tightly conserved water sites include numerous buried water molecules that have been described previously (McPhalen & James, 1988; Bott *et al.*, 1988). Most of the conserved waters are partially exposed on the protein surface, and some of these are used to form bridging waters in crystal contacts. A more detailed comparison of hydration in the three models is in preparation.

3.4. *Crystal packing*

The three crystals forms pack with similar solvent fractions (see Table 1). Among the three there are no identical crystal contacts, though the $P2_12_12_1$ and $P2_1$ forms have two pairs of similar contacts as described below. Crystal contacts have been analyzed using a 4 Å distance cutoff. In space group *C*2, SBT packs with 11 contacting neighbors. Three of these are twofold related while the other eight comprise four pairs of relations involving translation, two with the centering operation and two with the screw axis. Table 5 gives information on the symmetry and components of each contact. In these Tables, contacts are ranked by the total number of contact atoms, beginning with the strongest contact. The $P2_12_12_1$ form packs with ten neighbors (see Table 6). In the $P2_1$ crystal form, each protein has eight contacting neighbors (see Table 7).

In both the $P2_12_12_1$ and $P2_1$ crystal forms there is a contact that places the *C*-terminal end of the protein's central helix (residues 220–238) near the *N*-terminal

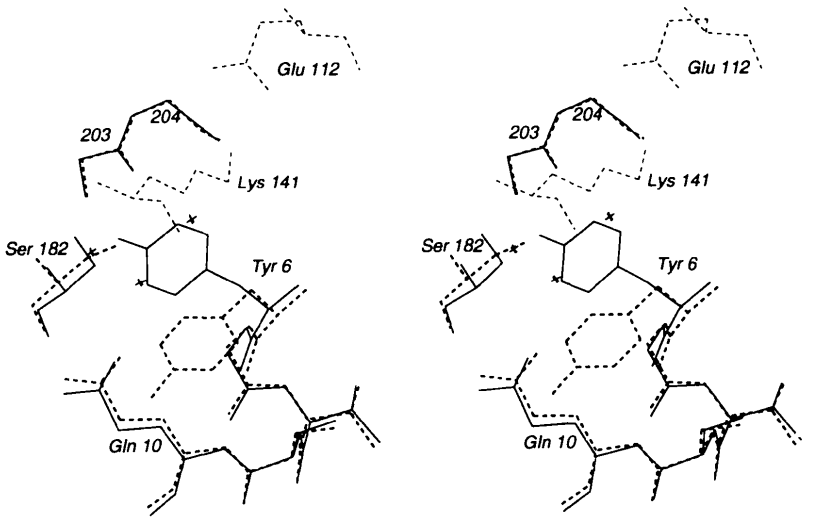

Fig. 5. Stereoview of the displacement of the Tyr6 side chain by a crystal contact. Solid lines show the normal position (space groups *C*2 and $P2_1$). Dashed lines show the side chains of Glu112 and Lys141 from a symmetry-related molecule and the new conformation of Tyr6 in space group $P2_12_12_1$. Three new water sites replacing the Tyr side chain are also shown.

Table 5. *Molecular contacts in subtilisin BPN′ crystals, space group C2*

Contact distance maximum 4.0 Å; hydrogen bonds and salt links maximum = 3.2 Å.

Contact rank	Symmetry relation to contacting protein	Contact atoms*	Contact residues	Number of hydrogen bonds	Salt links	Bridging waters
1	$-x+\frac{1}{2}, y+\frac{1}{2}, -z+1$	16:15	7:8	1	0	7
2	$-x+\frac{1}{2}, y+\frac{1}{2}, -z$	8:12	4:5	1	0	2
3	$x+\frac{1}{2}, y+\frac{1}{2}, z$	12:6	7:3	1	0	6
4	$x-\frac{1}{2}, y+\frac{1}{2}, z$	8:8	4:4	2	0	3
5	$-x, y, -z$	8:8†	4:4	0	0	3
6	$-x+1, y, -z$	3:3†	1:1	0	0	0
7	$-x, y, -z+1$	2:2†	2:2	0	0	0
	Totals	98	49	5	0	21

* Atoms within 4.0 Å of another protein molecule. † Twofold-related contact; the contacting surfaces are identical.

Table 6. *Molecular contacts in subtilisin BPN′ crystals, space group $P2_12_12_1$*

Contact distance maximum 4.0 Å; hydrogen bonds and salt links maximum = 3.2 Å.

Contact rank	Symmetry relation to contacting protein	Contact atoms*	Contact residues	Number of hydrogen bonds	Salt links	Bridging waters
1	$-x, y+\frac{1}{2}, -z-\frac{1}{2}$	41:36	14:10	10	0	4
2	$x+1, y, z$	15:18	5:5	2	0	4
3	$-x+\frac{1}{2}, -y+1, z+\frac{1}{2}$	18:21	7:8	3	0	5
4	$x+\frac{1}{2}, -y+\frac{3}{2}, -z$	9:10	5:6	0	1	1
5	$-x-\frac{1}{2}, -y+1, z+\frac{1}{2}$	10:4	4:2	0	0	1
	Totals	182	66	15	1	15

* Atoms within 4.0 Å of another protein molecule.

Table 7. *Molecular contacts in subtilisin BS3 crystals, space group $P2_1$*

Contact distance maximum 4.0 Å; hydrogen bonds and salt links maximum = 3.2 Å.

Contact rank	Symmetry relation to contacting protein	Contact atoms*	Contact residues	Number of hydrogen bonds	Salt links	Bridging waters
1	$x+1, y, z+1$	26:23	10:8	5	0	6
2	$-x, y+\frac{1}{2}, -z$	17:13	7:7	0	0	5
3	$x, y, z+1$	10:19	5:10	3	0	2
4	$x+1, y, z$	10:11	5:5	3	0	4
	Totals	129	57	11	0	17

* Atoms within 4.0 Å of another protein molecule.

ends of the two parallel surface helices (residues 103–117 and 132–145). In both cases the contact relates molecules along a screw axis: contact 5 in $P2_12_12_1$ along the *c* direction and contact 2 in $P2_1$ along the *b* direction. Although the angular relation of the contacting molecules differs, the unit-cell dimension produced by the screw translation is similar in both cases (75.2 *versus* 79.5 Å). A sulfate ion mediates the contact in $P2_12_12_1$, with hydrogen bonds to helix-terminal residues in both contacting proteins. This common contact, which is illustrated in Fig. 6, suggests that helix dipoles have a role in forming crystal contacts. However, in the *C*2 crystals there is not only no similar contact, but contact 5 involves residues from the C-terminal ends of the three relevant helices, an unfavourable dipole arrangement.

Another contact is similar between the $P2_12_12_1$ and $P2_1$ crystals: contact 2 in the former and contact 3 in the latter. Both of these contacts relate molecules with identical orientation, so that the vector from one contact surface to the other, through a single molecule, determines a lattice translation. Thus, lattice constant *a* in $P2_12_12_1$ and lattice constant *c* in $P2_1$ both correspond to the approximate diameter of the protein. The residues involved are similar but not identical in the two crystal forms. Fig. 7 illustrates this similar contact in the two crystal forms.

4. Discussion

The structure of subtilisin BPN′ at 1.6 Å resolution reveals several regions of discrete disorder, where two

or more distinct conformations are simultaneously observed. In general the major conformer coincides with that observed in the lower resolution structures; this is strictly true in the three cases where the discretely disordered side chain is internal and shielded from solvent (Thr71, Val177 and Val270). Discrete disorder provides evidence of dynamic motions that are known to occur in proteins but are difficult to observe. Motions correlated among multiple residues or correlated with the presence/absence of the active-site inhibitor are particularly informative, since they provide data on the propagation of conformational adjustments which can serve as the basis of allostery and concerted binding/catalysis mechanisms.

In general, the SBT structure in the three crystal forms is very similar and the largest differences are

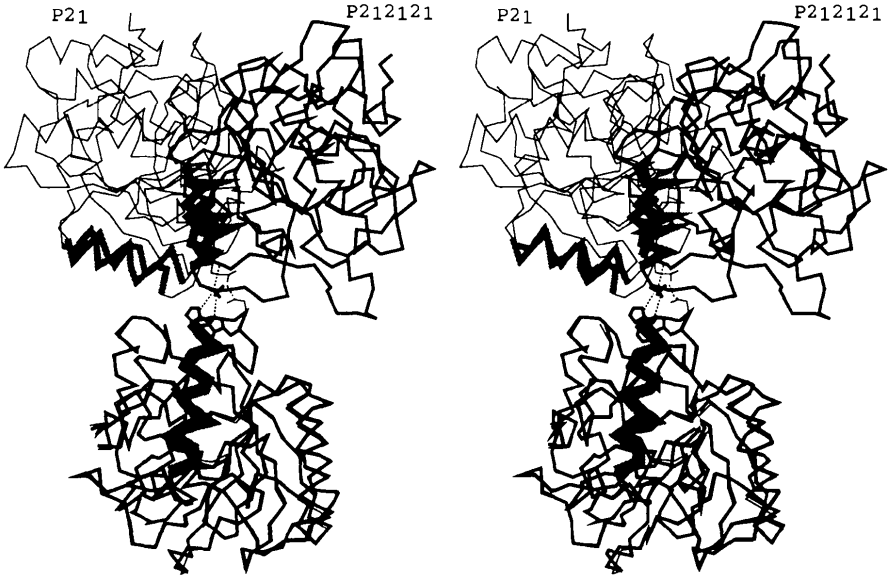

Fig. 6. Stereoview showing the helix dipole contacts in space group $P2_12_12_1$ and $P2_1$. The upper molecules show how the $P2_12_12_1$ (thick lines, on right) and $P2_1$ (thin lines, on left) arrangements place molcules with helix termini in close proximity. The central helix of the lower (reference) molecule and parallel surface helices of the upper molecules are highlighted. Although the angular relation of the molecules varies, the C-terminus of the central helix is near the N-termini of the surface helices in both space groups. Dotted lines show hydrogen bonds in the contacts. In space group $P2_12_12_1$ the contact incorporates a bridging sulfate ion.

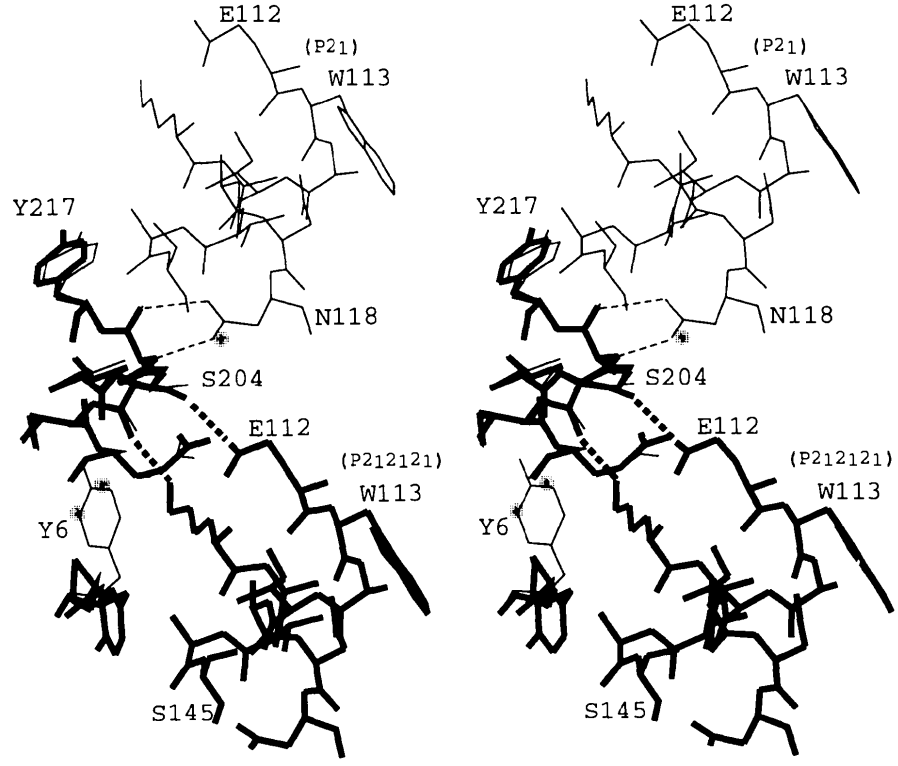

Fig. 7. The similarity between contact 2 in space group $P2_12_12_1$ and contact 3 in $P2_1$. Within each half of the stereo image, on the left are the superimposed reference structures; on the right the contacting neighbors. Thin lines, $P2_12_12_1$; thick lines, $P2_1$. In both crystals the contact relates molecules by pure translation.

because of crystal-packing effects. The structures were compared using the method of Bott & Frane (1990). The chief advantage of this method over simple coordinate comparison is that, by taking *B* factors into account, it identifies small but significant perturbations in well-ordered regions while ignoring insignificant variations in relatively disordered residues, *e.g.* surface lysines. In the context of subtilisin, where numerous engineered mutants and inhibitor complexes are under study, it is a useful and unbiased tool for highlighting important structural effects. The method produces a *Z* score for each atom (and for each residue), such that a *Z* score over 3.0 reflects a significant structural change. Each pairwise comparison produces about ten residues with *Z* greater than 3.0. The largest differences occur at crystal contacts where side chains, and in some cases the backbone, are displaced by crystal-packing forces. All the residues with residue *Z* scores of 4 and above belong to crystal contacts in at least one of the crystal forms in each comparison except His64, whose structural variation is due to the inhibitor in space group *C*2. The largest *Z* score (8.4) belongs to Tyr6, which adopts a different rotamer because of the crystal packing in the $P2_12_12_1$ crystals (Fig. 5).

Tables 5, 6 and 7 facilitate comparison of the crystal contacts among the three systems. Each crystal employs a set of unique contacts. How the chemical conditions cause a particular set of contacts to form, thus determining the crystal packing, lattice geometry and space-group symmetry, is a dynamic problem, for which the present analysis can supply only static information. This problem is analogous to the protein-folding problem in that the initial and final states are known, but the intermediate states are elusive and key to understanding the transition. Extensive hydration of three protein surfaces, and the predominance of hydrogen bonding (including bridging water molecules) in the major crystal contacts, together suggest that the hydration structure may play an important role in selecting the contacts. According to this view, variations in the occupancies of surface-bound water molecules, which could be sensitive to the protein's chemical environment, could in turn alter the relative strengths of transient contacts. When the hydration pattern promotes a favorable set of transient contacts, supersaturating conditions could lead to condensation of crystal nuclei. One role of a successful crystallization agent would then be to promote desolvation of nascent contacts, so that inter-protein hydrogen bonds can form. An alternative role for hydration in contact selection, which would apply to hydrophobic contacts as well as those involving hydrogen bonds, requires compatibility between the water structures around the nascent contacts. Specifically, compatible water structures adjacent to a contact are expected to strengthen that contact.

Although two of the crystal forms in the present analysis are grown from high salt while the third is from acetone, there is no clear distinction between the character of the resulting crystal contacts. A similar comparison of salt-grown *versus* alcohol-grown crystals of ribonuclease A found that salt-grown crystals had more extensive van der Waals interactions and fewer polar contacts than alcohol-grown crystals (Svensson *et al.*, 1991). This tendency is observed here for SBT, but it is not dramatic. Taken together the results of all the packing comparisons, including the bridging water and hydrogen-bond analysis, do not reveal any distinct rule regarding the effects on packing in high *versus* low ionic strength crystals. The two salt-grown crystals in the present study, though produced under nearly identical conditions, appear as different in molecular detail as either of them are from the acetone-grown higher pH form. In fact, the similarities in packing described in the previous paragraph make the $P2_12_12_1$ and $P2_1$ forms more similar than the two salt-grown forms. Although the acetone-grown crystals have the lowest number of contacts, in other respects such as the total number of contact atoms, hydrogen bonds or bridging water molecules, the acetone values are intermediate between those of the two ammonium sulfate systems. This suggests that the fundamental processes occurring in crystal formation are not profoundly affected by the ionic strength.

References

Abrahmsen, L., Jeffrey, T., Burnier, J., Butcher, K. A., Kossiakoff, A. & Wells, J. A. (1991). *Biochemistry*, **30**, 4151–4159.

Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bhat, T. N., Sasisekharan, V. & Vijayan, M. (1979). *J. Pept. Protein Res.* **13**, 170–184.

Bott, R. & Frane, J. (1990). *Protein Eng.* **3**, 649–657.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T, Katz, B. & Power, S. (1988). *J. Biol. Chem.* **263**, 7895–7906.

Bryan, P. N., Alexander, P., Strausberg, S., Schwarz, F., Wang, L., Gilliland, G. L. & Gallagher, D. T. (1992). *Biochemistry*, **31**, 4937–4945.

Bryan, P. N., Rollence, M. L., Pantoliano, M. W., Wood, J., Finzel, B. C., Gilliland, G. L., Howard, A. J. & Poulos, T. L. (1986). *Proteins Struct. Funct. Genet.* **1**, 326–334.

Drenth, J. & Hol, W. G. J. (1967). *J. Mol. Biol.* **28**, 543–544.

Drenth, J., Hol, W. G. J., Jansonius, J. N. & Koekoek, R. (1972). *Eur. J. Biochem.* **26**, 177–181.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G. & Wells, J. A. (1986). *Science*, **233**, 659–663.

Finzel, B. C. (1987). *J. Appl. Cryst.* **20**, 53–55.

Fitzergerald, P. A. M. (1988). *J. Appl. Cryst.* **21**, 273–278.
Furey, W., Wang, B. C. & Sax, M. (1982). *J. Appl. Cryst.* **15**, 160–166.
Gallagher, D. T., Bryan, P. & Gilliland, G. L. (1993). *Proteins Struct. Funct. Genet.* **16**, 205–213.
Gilliland, G. L., Gallagher, D. T. & Bryan, P. (1996). *Subtilisin Ensymes: Practical Protein Engineering*, edited by R. Bott and C. Betzel, pp. 159–169. New York: Plenum Press.
Hendrickson, W. A. & Konnert, J. (1981). *Biomolecular Structure, Function, Conformation and Evolution*, edited by R. Srinivasian, Vol. 1, pp. 43–47. Oxford: Pergamon Press.
James, M. N. G. & Sielecki, A. R. (1983). *J. Mol. Biol.* **163**, 299–361.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Luzzati, P. V. (1952). *Acta Cryst.* **5**, 802–810.
McPhalen, C. A. & James, M. N. G. (1988). *Biochemistry*, **27**, 6582–6598.
Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S., Hardman, K., Rollence, M. L. & Bryan. P. N. (1989). *Biochemistry*, **28**, 7205–7213.
Pantoliano, M. W., Whitlow, M., Wood, J. F., Rollence, M. L., Finzel, B. C., Gilliland, G. L., Poulos, T. L. & Bryan, P. N. (1988). *Biochemistry*, **27**, 8311–8317.
Ponder, J. W. & Richards, F. M. (1987). *J. Mol. Biol.* **193**, 775–791.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). *J. Mol. Biol.* **190**, 593–604.
Sheriff, S. (1987). *J. Appl. Cryst.* **20**, 55–57.
Siezen, R. J., de Vos, W. M., Leunissen, J. A. M. & Dijkstra, B. W. (1991). *Protein Eng.* **4**, 719–737.
Strausberg, S., Alexander, P., Wang, L., Gallagher, T., Gilliland, G. L. & Bryan, P. N. (1993). *Biochemistry*, **32**, 10371–10377.
Svensson, L. A., Dill, J., Sjolin, L., Wlodawer, A., Toner, M., Bacon, D., Moult, J., Veerapandian, B. & Gilliland, G. L. (1991). *J. Cryst. Growth*, **110**, 119–130.
Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
Wright, C. S. (1972). *J. Mol. Biol.* **67**, 151–163.
Wright, C. S., Alden, R. A. & Kraut, J. (1969). *Nature (London)*, **221**, 235–242.