References

ALLEN, F.H., BELLARD, S., BRICE, M.D., CARTWRIGHT, B.A., DOUBLEDAY, A., HIGGS, H., HUMMELINK, T., HUMMELINK-PETERS, B. G., KENNARD, O., MOTHERWELL, W. D. S., RODGERS, J. R. & WATSON, D. G. (1979). *Acta Cryst.* B**35**, 2331–2339.

ALLINGER, N. L. (1976). *Adv. Phys. Org. Chem.* **13**, 1–82.

AZARNIA, N., JEFFREY, G. A. & SHEN, M. S. (1972). *Acta Cryst.* B**28**, 1007–1013.

CECCARELLI, C., JEFFREY, G. A. & TAYLOR, R. (1981). *J. Mol. Struct.* **70**, 255–271.

HAMILTON, W. C. & IBERS, J. A. (1968). *Hydrogen Bonding in Solids*. New York: Benjamin.

JEFFREY, G. A. (1969). *Acc. Chem. Res.* **2**, 344–352.

JEFFREY, G. A., GRESS, M. E. & TAKAGI, S. (1977). *J. Am. Chem. Soc.* **99**, 609–611.

JEFFREY, G. A. & KIM, H. S. (1970). *Carbohydr. Res.* **14**, 207–216.

JEFFREY, G. A. & MALUSZYNSKA, H. (1982). *Int. J. Biol. Macromol.* **4**, 173–185.

JEFFREY, G. A. & SHIONO, R. (1977). *Acta Cryst.* B**33**, 2700–2701.

JEFFREY, G. A. & WOOD, R. A. (1982). *Carbohydr. Res.* **108**, 205–211.

KOLLMAN, P. A. & ALLEN, L. C. (1969). *J. Chem. Phys.* **51**, 3286–3293.

LINDNER, K. & SAENGER, W. (1982). *Acta Cryst.* B**38**, 203–210.

MALUSZYNSKA, H., KINOSHITA, Y. & JEFFREY, G. A. (1982). *Carbohydr. Res.* **100**, 17–28.

MALUSZYNSKA, H., RUBLE, J. R. & JEFFREY, G. A. (1981). *Carbohydr. Res.* **97**, 199–204.

NEWTON, M. D. (1983). *Acta Cryst.* B**39**, 104–113.

NEWTON, M. D., JEFFREY, G. A. & TAKAGI, S. (1979). *J. Am. Chem. Soc.* **101**, 1997–2002.

NORRESTAM, R., BOCK, K. & PEDERSEN, C. (1981). *Acta Cryst.* B**37**, 1265–1269.

PICKERING, S. U. (1893). *J. Chem. Soc.* **63**, 141–195.

PIMENTEL, G. C. & MCCLELLAN, A. L. (1960). *The Hydrogen Bond*, p. 100. San Francisco: Freeman.

TAKAGI, S. & JEFFREY, G. A. (1978*a*). *Acta Cryst.* B**34**, 1591–1596.

TAKAGI, S. & JEFFREY, G. A. (1978*b*). *Acta Cryst.* B**34**, 3104–3107.

TAKAGI, S. & JEFFREY, G. A. (1978*c*). *Acta Cryst.* B**34**, 2551–2555.

TAKAGI, S. & JEFFREY, G. A. (1979*a*). *Acta Cryst.* B**35**, 902–906.

TAKAGI, S. & JEFFREY, G. A. (1979*b*). *Acta Cryst.* B**35**, 1482–1486.

TSE, Y.-C. & NEWTON, M. D. (1977). *J. Am. Chem. Soc.* **99**, 611–613.

Acta Cryst. (1983). B**39**, 480–490

The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and Its Complexes with Inhibitors

BY MARKUS MARQUART, JOCHEN WALTER, JOHANN DEISENHOFER, WOLFRAM BODE AND ROBERT HUBER

Max-Planck Institut für Biochemie, D-8033 *Martinsried bei München, Federal Republic of Germany*

(*Received* 26 *August* 1982; *accepted* 8 *December* 1982)

Abstract

Sixteen protein crystal structures of the system bovine trypsin, trypsinogen, bovine basic pancreatic trypsin inhibitor (PTI) and Kazal inhibitor (porcine secretory inhibitor and Japanese quail ovomucoid), which had been refined by energy-restrained crystallographic refinement (EREF) at high resolution (1·4 to 1·9 Å), were analysed with respect to the geometries of the active site and the peptide groups in general. It was found that the conformation of Asp 102–His 57–Ser 195 is well conserved, irrespective of the different functional states of the enzyme. The His 57–Ser 195 hydrogen bond, however, improves considerably in the complexes as regards bond length and angle. The inhibitor binding at the main chain and side chain of P1 is also well conserved, including solvent molecules involved in the network of hydrogen bonds. The small-molecule inhibitor APPA (amidinophenyl pyruvate) bound to trypsin was also studied at high resolution. It binds by forming a tetrahedral adduct. The relation between the distance of the Ser 195 O_γ to the carbonyl C of the inhibitors and the pyramidalization at the carbonyl C obeys the Bürgi–Dunitz–Shefter relation for nucleophile–electrophile interaction. Analysis of the peptide groups suggests improvements of the standard geometry of several interbond angles. It also shows that peptide groups may be substantially non-planar and this non-planarity can be reliably analysed under the conditions of this study.

1. Introduction

Refinement of crystal structures of proteins requires geometrical constraints or restraints as the number of observed diffraction intensities is insufficient to determine atomic positions and 'thermal' parameters, even at 1·5 Å resolution.

The crystalline order of protein crystals often limits resolution to less than 2 Å. Consequently, geometrical restraints become even more important. With geo-

0108-7681/83/040480-11$01.50

metrical restraints refinement is possible even at about 3 Å resolution in crystal structures of large protein molecules (Deisenhofer, 1981). The number of parameters to be varied may be drastically reduced when large molecular fragments of known geometry are allowed to move as rigid bodies (*CORELS*, Sussman, Holbrook, Church & Kim, 1977). However, this must be regarded as a first approximation, which usually does not allow refinement below $R \sim 0{\cdot}40$.

The restraints used for refinement are based on peptide and amino acid geometries and non-bonded contacts derived mostly from small-molecule crystal structures (EREF, Levitt, 1974). Geometric constraints are used in the real-space refinement procedure (Diamond, 1971, 1974). The number of parameters varied is reduced to dihedral and selected interbond angles in this case.

The necessity to constrain or restrain polypeptide conformation in protein crystal structural refinement causes an obvious dilemma: functional properties, like enzyme catalysis, may involve sterical strain (Jencks, 1975) and distorted geometries of the reactive site of enzyme or substrate. These must remain undetected if the conformation is constrained or strongly restrained

Table 1. *Summary of the refined bovine trypsin, trypsinogen, trypsin/trypsinogen–bovine PTI/porcine PSTI structures*

OMJPQ3 is Japanese quail ovomucoid domain 3. The number of crosses in the line under the name indicates the isomorphous family. A definition of bonds and angles is given in the text (§2.3). R.m.s. deviations of individual measurements are given in brackets. The significance of the deviations between target and mean measured values is discussed in the text.

Crystal	Trypsinogen	Trypsinogen	Trigonal trypsin	Orthorhombic trypsin	Trypsinogen at 173 K	Trypsinogen at 103 K	Orthorhombic trypsin + BA	Orthorhombic trypsin
Name	TNME	NATI	TRYB	DEBA	TGHH	TGHS	BA	DEBB
	+	+	+	++	+	+	++	++
Solvent	50% CH_3OH–50% water pH 7·0	2·4 *M* $MgSO_4$ pH 6·9	2·4 *M* $(NH_4)_2SO_4$ pH 6·9	2·4 *M* $(NH_4)_2SO_4$ pH 8·0	70% CH_3OH–30% water pH 7·0	70% CH_3OH–30% water pH 7·0	2·4 *M* $(NH_4)_2SO_4$ pH 7·0	2·4 *M* $(NH_4)_2SO_4$ pH 5·0
Bond lengths (Å) (target)	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
CA (1·525)	1·524 (4)	1·524 (4)	1·525 (5)	1·525 (5)	1·524 (6)	1·525 (5)	1·53 (1)	1·53 (1)
AO (1·257)	1·256 (2)	1·256 (3)	1·256 (3)	1·255 (4)	1·255 (5)	1·256 (4)	1·26 (1)	1·254 (9)
AN (1·318)	1·313 (7)	1·313 (8)	1·315 (7)	1·315 (9)	1·32 (1)	1·32 (1)	1·32 (2)	1·32 (2)
CN (1·468)	1·465 (7)	1·465 (7)	1·467 (6)	1·467 (8)	1·47 (1)	1·469 (9)	1·47 (2)	1·47(2)
Bond angles (°) (target)								
ACN (113·1)	112 (2)	112 (2)	112 (2)	112 (2)	112 (2)	112 (2)	111 (2)	111 (3)
CAO (119·7)	120 (1)	120 (1)	120 (1)	120 (1)	120 (1)	120 (1)	120 (2)	119 (2)
CAN (115·2)	115 (1)	115 (1)	116 (1)	116 (1)	116 (1)	116 (1)	116 (2)	117 (2)
NAO (124·1)	124 (1)	124 (1)	124 (1)	124 (1)	124 (1)	124 (1)	124 (2)	123 (2)
ANC (123·8)	123 (2)	122 (2)	123 (2)	122 (2)	122 (2)	123 (2)	121 (3)	122 (3)
Torsion angles (°) (target)								
ω_1 (180·0)	179 (7)	179 (7)	179 (6)	179 (7)	179 (7)	179 (7)	179 (8)	179 (8)
ω_3 (0·0)	−2 (4)	−2 (4)	−1 (3)	−1 (3)	−2 (4)	−2 (4)	−2 (6)	−2 (6)

Crystal	Trypsin + APPA	Trypsinogen + PSTI	Trypsin + PTI	Trypsinogen + PTI	Anhydrotrypsin + PTI	Trypsinogen + IV + PTI	BPTI	OMJPQ3
Name	APPA	TGKZ	TTNA	TTOG	TTAN	TTIV	BPTI	OMJPQ3
	++	+++	++++	++++	++++	++++	+++++	++++++
Solvent	3·0 *M* $(NH_4)_2SO_4$ pH 7·6	1·5 *M* $MgSO_4$ pH 7·6	2·5 *M* K/Na phosphate pH 7·0	2·4 *M* $MgSO_4$ pH6·9	2·5 *M* K/Na phosphate pH 7·0	2·4 *M* $MgSO_4$ pH 6·9	2·25 *M* K/H phosphate pH 10·0	1·5 *M* Citrate pH 9·0
Bond lengths (Å) (target)	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
CA (1·525)	1·53 (1)	1·525 (9)	1·524 (9)	1·52 (1)	1·52 (1)	1·52 (1)	1·53 (1)	1·520 (6)
AO (1·257)	1·26 (1)	1·255 (7)	1·256 (6)	1·253 (8)	1·255 (8)	1·254 (9)	1·255 (9)	1·254 (4)
AN (1·318)	1·32 (2)	1·31 (2)	1·31 (2)	1·31 (2)	1·31 (2)	1·31 (2)	1·32 (2)	1·31 (1)
CN (1·468)	1·47 (2)	1·47 (1)	1·47 (1)	1·47 (2)	1·47 (2)	1·47 (2)	1·47 (2)	1·461 (9)
Bond angles (°) (target)								
ACN (113·1)	111 (3)	112 (2)	112 (3)	112 (3)	112 (3)	112 (3)	111 (3)	112 (2)
CAO (119·7)	120 (2)	120 (2)	120 (2)	119 (2)	119 (2)	120 (2)	120 (2)	120 (1)
CAN (115·2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)	116 (2)
NAO (124·1)	123 (2)	124 (2)	124 (2)	123 (2)	123 (2)	123 (2)	124 (1)	124 (1)
ANC (123·8)	121 (3)	122 (2)	122 (3)	121 (4)	122 (3)	122 (3)	121 (2)	122 (3)
Torsion angles (°) (target)								
ω_1 (180·0)	179 (8)	178 (7)	179 (8)	180 (8)	179 (8)	179 (9)	180 (7)	179 (8)
ω_3 (0·0)	−2 (8)	−1 (5)	−2 (6)	−2 (8)	−2 (7)	−2 (7)	−1 (6)	−1 (5)

to standard geometry. The use of restraints in the refinement of protein structures in addition prohibits a straightforward evaluation of the positional and angular errors of the model. Only estimates of the errors [*e.g. via* the Luzzati (1952) plot] are possible.

We will demonstrate in this paper that, despite the problems described, reliable information on peptide geometry can be obtained in favourable cases, *i.e.* when several structure determinations can be compared at resolutions better than 2 Å. The degree of internal consistency observed provides an objective estimate of the errors in the models.

The trypsinogen, trypsin, PTI, porcine pancreatic secretory trypsin inhibitor (PSTI) structures are an appropriate system which has been analysed in detail. The separate molecules were investigated as well as their complexes [Rühlmann, Kukla, Schwager, Bartels & Huber (1973) (TTNA); Huber *et al.* (1974) (TTNA); Deisenhofer & Steigemann (1975) (PTI); Huber, Bode, Kukla & Kohl (1975) (TTAN); Bode & Schwager (1975) (DEBA, BA, DEBB); Bode, Schwager & Huber (1976) (BA, DEBB); Bode, Fehlhammer & Huber (1976) (NATI, TTOG, TTIV); Bode & Huber (1976) (TTOG); Fehlhammer, Bode & Huber (1977) (TRYB, TTIV); Bode, Schwager & Huber (1978) (TTIV); Huber & Bode (1978) (NATI, TRYB, TTNA, TTOG); Bode & Huber (1978) (NATI, TTOG); Walter, Steigemann, Singh, Bartunik, Bode & Huber (1982) (TNME, NATI, TRYB, DEBA, TGHH, TGHS); Bolognesi *et al.* (1982) (TGKZ); Walter (1982) (APPA)]. In addition to these different crystal structures many isomorphous variants have been analysed at differing pH values, solvent compositions or temperatures [Walter *et al.* (1982) (TNME, NATI, TRYB, DEBA, TGHH, TGHS); Singh, Bode & Huber (1980) (TNME); Bode & Schwager (1975) (DEBA, BA, DEBB)]. These structures provided different views of these molecules (in non-isomorphous crystal structures) and similar views based on different data sets (in isomorphous crystal structures). In addition, hydrogen bonding, nucleophile–electrophile addition, plays an important role in protease function. The relevant geometric parameters will also be compared.

The analyses were performed at resolutions between 1·4 and 1·9 Å as crystalline order permitted. They are summarized in Table 1.*

2. Methods

2.1. *Refinement of the complexes*

The complexes of PTI with trypsin (TTNA), trypsinogen (TTOG), anhydrotrypsin (TTAN), the ternary complex with trypsinogen and the Ile-Val dipeptide (TTIV) as well as the benzamidine-inhibited trypsin (BA), the free trypsin at pH 5·0 (DEBB) and PTI had been analysed and refined by real-space refinement in the references given.

They were further refined in this investigation compared to the original studies applying EREF (Jack & Levitt, 1978; Levitt, 1974). The other crystal structures listed in Table 1 had already been refined with this method (Walter *et al.*, 1982; Bolognesi *et al.*, 1982; Papamokos *et al.*, 1982).
EREF minimizes the function:

$$R = E + kX$$

where X is the crystallographic residual $\sum (|F_o| - |F_c|)^2$ and E the potential-energy function:

$$\begin{aligned} E = &\sum K_b(b_i - b_o)^2 && \text{(bonds)} \\ &+ \sum K_\tau(\tau_i - \tau_o)^2 && \text{(bond angles)} \\ &+ \sum K_\theta[1 + \cos(m\theta + \delta)] && \text{(torsion angles)} \\ &+ \sum (Ar_i^{-12} + Br_i^{-6}) && \text{(non-bonded interactions).} \end{aligned}$$

The factor k, which controls the relative contribution of the crystallographic residual to the total residual, was chosen in such a way that both terms E and X decreased simultaneously.

There is no explicit term for hydrogen-bond energies in the above energy function. Attractive hydrogen-bond energy is instead modelled by choosing constants A and B in the 6–12 potential such that for possible hydrogen-bonding partners at a distance of 2·9 Å there is an energy minimum of about −5·0 kJ mol^{-1}.

The equilibrium values of bond lengths, inter-bond angles, torsion angles and the corresponding force constants used in refinement were those suggested by Levitt (1974). As an example for the energy parameters applied, Table 2 contains the equilibrium values b_o, τ_o, the minimum-energy torsion angles and the force constants K_b, K_τ and K_θ for bonds, bond angles and torsion angles occurring in a peptide unit.

* A more detailed version of Table 1 (additionally comprising statistics on data collection, overall temperature factors and R factors), a more detailed version of Table 3 (all C_α–C_α distances) and the product–moment correlation coefficients between every pair of structures of Table 1 have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 38356 (10 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. Additionally, atomic coordinates and structure factors for all structures listed in Table 1 have been deposited with the Protein Data Bank, Brookhaven National Laboratory (References: 1OVO, 2PTN, 1TPO, 3PTN, 3PTB, 1TPP, 4PTI, 2PTC, 1TPA, 2TGA, 1TGC, 1TGT, 2TGT, 2TGP, 3TPI, 1TGS, R1OVOSF, R2PTNSF, R1TPOSF, R3PTNSF, R3PTBSF, R1TPPSF, R4PTISF, R2PTCSF, R1TPASF, R2TGASF, R1TGCSF, R1TGTSF, R2TGTSF, R2TGPSF, R3TPISF, R1TGSSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centers at Cambridge, Melbourne or Osaka. The data for all structures except OMJPQ3 have also been deposited with the British Library Lending Division as Supplementary Publications Nos. 37004 (14 microfiche) and 37009 (18 microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 2. *Equilibrium values and force constants for bonds, bond angles and torsion angles occurring in a peptide unit*

A definition of the bonds and angles is given in the text (§2.3). The statistical quantities were calculated under the assumption that the structures of Table 1 represent a sample of 3491 independent observations of the bonds and angles in column 1. R.m.s. deviations of individual measurements are given in brackets. The significance of the deviations between target and mean, measured values is discussed in the text.

Bond type	Force constant (kJ Å^{-2})	Equilibrium value (Å)	Mean values over 3491 residues (Å)
CA	1959·4224	1·525	1·53 (1)
AO	2381·4510	1·257	1·255 (7)
AN	964·6383	1·318	1·31 (2)
CN	1055·0736	1·468	1·47 (2)
Bond-angle type	**(kJ deg^{-2})**	**(°)**	**(°)**
ACN	263·9371	113·1	112 (2)
CAO	288·5350	119·7	120 (2)
CAN	233·5703	115·2	116 (2)
NAO	237·5327	124·1	124 (2)
ANC	213·6499	123·8	112 (3)
Torsion-angle type	**(kJ deg^{-1})**	**(°)**	**(°)**
X–A–N–X	42·0	180·0/ 0·0	179 (8)/ −2 (6)

* The torsion angle *X–A–N–X* describes ω_1 (first row) and ω_3 (second row).

In order not to limit the approach of Ser 195 O_γ towards C′ (C′ denotes the carbonyl carbon) of the scissile bond in the complexes by repulsive non-bonded forces, non-bonded interaction terms for the C′ of the scissile bond were excluded from energy calculations.

2.2. *APPA*

We add the analysis of the small-inhibitor *p*-amidinophenyl pyruvate complex with trypsin (APPA, Walter, 1982): orthorhombic trypsin crystals were soaked in 3*M* ammonium sulphate (pH = 7·6) containing 5 m*M* APPA. {Intensity data up to 1·4 Å resolution, rotation method and cylindrical film cassette, Rigaku rotating anode (5 kW), spot size 3 × 0·3 mm, modified Nonius (Delft, Holland) precession camera or a rotation precession camera (Huber, Rimsting, Federal Republic of Germany), focus-to-crystal distance 30 cm. Collection and evaluation (Schwager & Bartels, 1977) of more than 100000 data (two crystals), 31376 significant unique reflexions, 71·7% of all possible reflexions up to 1·4 Å resolution, R(merge) value 7·7% [$\sum (I - \langle I \rangle)/\sum \langle I \rangle$, summation over all measurements, $\langle I \rangle$ averaged intensity, I intensity of individual measurement].}

A difference Fourier synthesis calculated between APPA-trypsin and free trypsin (DEBA) using the DEBA phases exhibited a high positive difference density for the APPA group (Fig. 1). At a graphics-display system (Jones, 1978) an APPA model group was constructed and fitted into the difference density. This APPA model combined with the refined structure of free trypsin (DEBA) served as starting coordinate set for the EREF refinement procedure.

Except for the carbonyl C(9) (see Fig. 6 for atom numbering) of APPA we used the energy parameters recommended by Levitt (1974) (see Table 2). To prevent repulsion of the carbonyl C(9) of APPA from Ser 195 O_γ of trypsin, no van der Waals interactions were applied for both atoms (the non-bonded interaction coefficients A and B were set to 0·0). The equilibrium bond lengths of the C(8)–C(9) and C(9)–C(10) bonds were chosen as 1·518 and 1·579 Å respectively, as found for the Na pyruvate structure (Tavale, Pant & Biswas, 1961). To favour neither tetrahedral nor trigonal geometry of the carbonyl C(9), the equilibrium bond length of C(9)–O(1) was set at 1·35 Å and the equilibrium bond angles of C(8)–C(9)–O(1), C(8)–C(9)–C(10) and O(1)–C(9)–C(10) were set at 115°. These are values intermediate between those tetrahedral (1·43 Å and 110°) and trigonal conformations (1·26 Å and 120°). By applying weak force constants (one third of those given by Levitt) the influence of the energy and geometric terms was reduced for the carbonyl group. This ensures that the position of C(9) is predominantly determined by the crystallographic contributions.

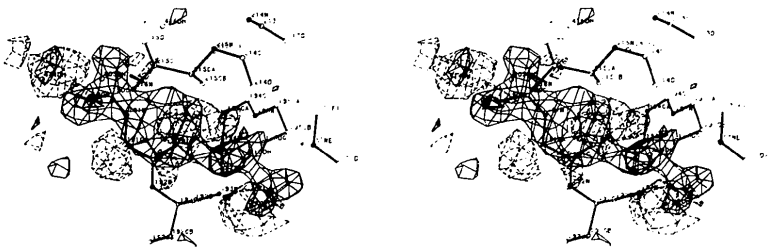

Fig. 1. Initial difference electron density map between APPA and DEBA calculated with coefficients ($|F$(APPA)$| - |F$(DEBA)$|$) exp [$-i\alpha$(DEBA)] at the active site, overlaid with the trypsin chain segments in this region.

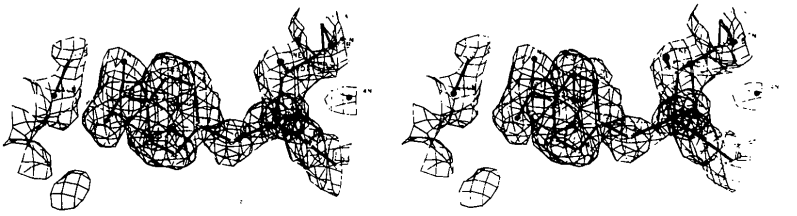

Fig. 2. Final $2F_o - F_c$ electron density map of APPA with the final model overlaid.

The final R factor was 19·1% for all reflexions with RR = 1·2 [$2|F_o| - |F_c||/(|F_o| + |F_c|) < 1·2$] in the resolution range 1·4 to 6·5 Å. The refined model and $2F_o$–F_c electron density of APPA are shown in Fig. 2.

The status of refinement of all the structures [maximal resolution: 1·9 Å (TTAN, TTOG, TTIV, TTNA, OMJPQ3) to 1·4 Å (APPA); ratio of significantly measured to possible number of structure factors up to the maximal resolution: 65·3% (TNME) to 85·3% (PTI); R factors: 20·9% (TGHS) to 16·2% (PTI)] is shown in Table 1* in terms of mean and r.m.s. values for all bonds, bond angles and restrained torsion angles in the peptide units.

2.3. *Peptide geometry*

Using the convention introduced by Winkler & Dunitz (1971) we define the four torsion angles of a peptide unit as follows:

$\omega_1 = \omega(C_\alpha{-}C'{-}N{-}C_\alpha)$ $\quad$ $\omega_3 = \omega(O{-}C'{-}N{-}C_\alpha)$
$\omega_2 = \omega(O{-}C'{-}N{-}H)$ $\quad$ $\omega_4 = \omega(C_\alpha{-}C'{-}N{-}H)$.

Bonds and interbond angles within a peptide unit are defined as follows:

$CA = C_\alpha{-}C'$ $\quad$ $\tau = ACN = \angle(N{-}C_\alpha{-}C')$
$AO = C'{-}O$ $\quad$ $CAO = \angle(C_\alpha{-}C'{-}O)$
$AN = C'{-}N$ $\quad$ $CAN = \angle(C_\alpha{-}C'{-}N)$
$CN = C_\alpha{-}N$ $\quad$ $NAO = \angle(N{-}C'{-}O)$
$\quad$ $ANC = \angle(C'{-}N{-}C_\alpha)$.

As the position of the H of the peptide NH group is not determined in the crystal structures described only the ω_1 and ω_3 torsions can be calculated.

From ω_1 and ω_3 we calculate the torsion angle describing the deformation at C′, χ_C (Warshel, Levitt & Lifson, 1970); $\chi_C = \omega_1 - \omega_3 + \pi$(modulo 2π) (Winkler & Dunitz, 1971) (for an example see Fig. 3).

The deformation of the peptide at C′ can also be described by the out-of-plane bend angle θ[4] (the plane being defined by C_α, C′, N) (Huber *et al.*, 1974), and the corresponding out-of-plane distance of the carbonyl O.

For all residues of the structures listed in Table 1, the torsion angles and interbond angles defined above were calculated. Mean values and r.m.s. deviations over all 3491 peptides with a temperature factor lower than 40·0 Å [mean(all), r.m.s.(all)] as well as separate mean values and r.m.s. deviations for each of the 16 structures listed in Table 1 [mean(struc.), r.m.s.(struc.)] were evaluated (see column 4 in Table 2). Residues with temperature factors above 40·0 Å^2 were excluded as their conformation is not determined with sufficient accuracy.

To estimate the degree of correlation between the different structures with respect to peptide torsion angles, pairwise product–moment correlation coefficients r_{ij} for structures i and j were calculated.*

$$r_{ij} = \sum (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)/[\sum (x_{ik} - \bar{x}_i)^2 \times \sum (x_{jk} - \bar{x}_j)^2]^{1/2}.$$

x_{ik} stands for ω_1, ω_3, or χ_C respectively of residue k in structure i and $\bar{x}_i$ is the ω_1, ω_3 or χ_C mean(struc.) value respectively for structure i; x_{ik} and x_{jk} are corresponding residues in structures i and j.

Fig. 4 in addition gives a very detailed representation of the ω_1 and χ_C angles calculated. It should be noted that residues of the activation domain in the free trypsinogen structures are disordered and not included in the comparison.

* See deposition footnote.

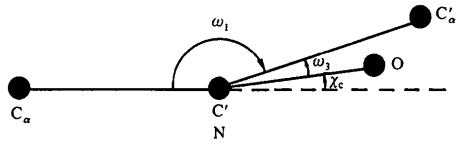

Fig. 3. Example for the definition of the peptide-unit torsion angles. The peptide unit is seen projected along the C′–N bond. ω_1 is 160°, ω_3 is −10° and χ_C is −10°.

* See deposition footnote.

Table 3. *Mean* C_α–C_α *distance* (Å) *between DEBA and the other structures of Table* 1 *after optimal superposition*

DEBB	0·083	BA	0·084	APPA	0·093
TRYB	0·182	NATI	0·209	TNME	0·209
TGHH	0·224	TGHS	0·238	TTAN	0·282
TTNA	0·285	TTIV	0·292	TTOG	0·292
TGKZ	0·337				

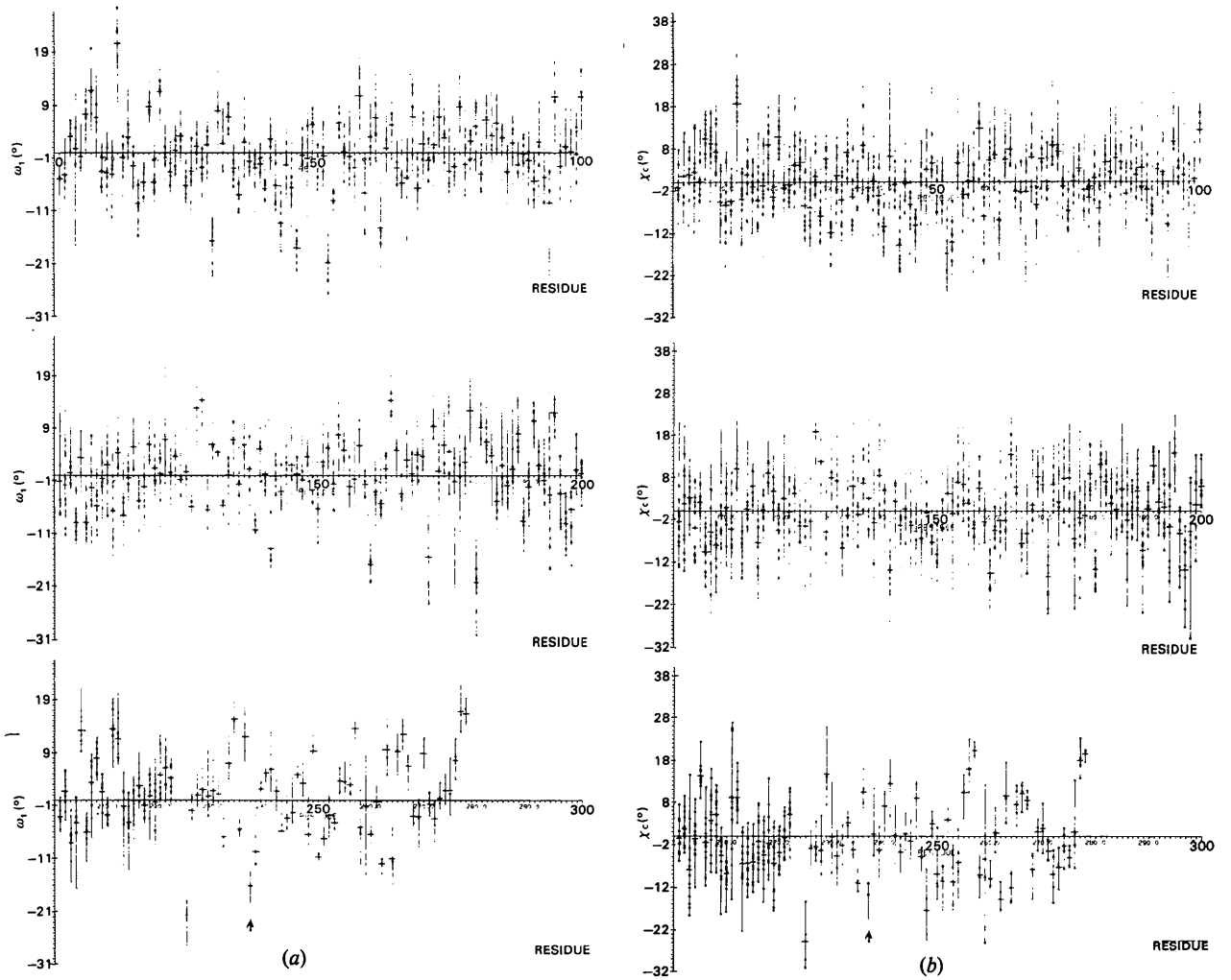

Fig. 4. Deviation of individual torsion angles from their mean(all) values. Vertical lines connect corresponding residues of the structures listed in Table 1. Residues are numbered 1 to 278; residue number 1 is Ile 16 of the trypsin(ogen) component; 223 is the first residue of PTI. Horizontal rules represent the mean values over corresponding residues. (*a*) $\omega_1 - \omega_1$ [mean(all)]. (*b*) $\chi_C - \chi_C$ [mean(all)].

The positional differences between the structures were analysed by optimal superposition of the C_α atoms of equivalent residues using a least-squares procedure (Table 3).* Only those atoms which are defined with a temperature factor smaller than 40·0 Å^2 were compared. Residues 15–17, 22–26 and 142–154 in TGKZ, 18 in TGHH and 17–18 in TGHS were omitted from the positional comparison, as they are known (Walter *et al.*, 1982; Bolognesi *et al.*, 1982) to adopt a conformation different from that in the other trypsin(ogen)s.

* See deposition footnote.

3. Results and discussion

3.1. C_α *comparison*

Mean distances of C_α atoms between each pair of structures of Table 1 are in the range 0·36 to 0·08 Å. This demonstrates clearly the general correspondence between the structures. In detail the mean distances within the isomorphous families are smaller than between structures of different relationship within the isomorphous families. Table 3, which lists the mean C_α distances between DEBA and the other structures of Table 1, may serve as an example for this trend. TRYB (trigonal trypsin), however, is as closely related to the

orthorhombic trypsin as it is to the isomorphous trypsinogen family.

This comparison provides an objective measure of the upper limit of the mean error in C_α positions. The values contain contributions from both errors and true structural differences caused by lattice packing, solvent environment and/or different functional states. The smallest values (around 0·1 Å) may be close to the true error. The Luzzati plot for APPA or TGKZ indicates errors of 0·18 and 0·21 Å respectively. This value refers to all atoms and a smaller value for C_α is to be expected, well in accord with the notion above.

A residue-by-residue comparison shows smaller deviations for internal residues and larger differences for external loops in particular when non-isomorphous series are compared. This indicates clearly the contribution of real structural differences to the distances of Table 3.

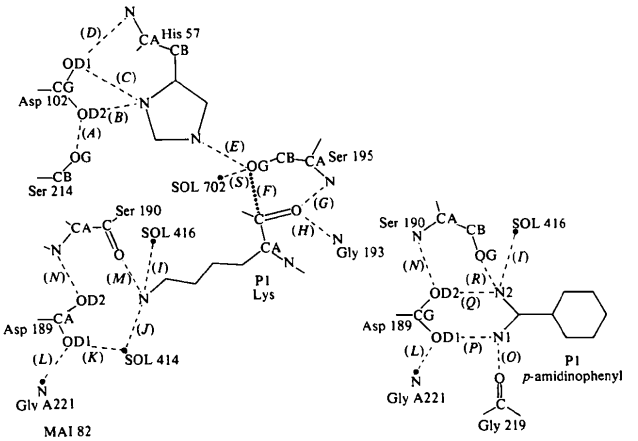

Fig. 5. Schematic drawing of the reactive site of trypsin and the P1 binding environment. Hydrogen bonds are indicated by (-----). *P*1 Lys and *P*1 *p*-amidinophenyl binding is shown side by side.

3.2. *Active site and* P1 *binding site*

Fig. 5 is a schematic drawing of the reactive site of trypsin and its hydrogen-bonding environment and the P1 binding environment of the inhibitors discussed. All possible proton donors and acceptors whose distance and angular relationship is appropriate for hydrogen bonding are included in this scheme. We differentiate between lysine and amidino groups at P1.

Table 4 lists the distances of atoms involved in hydrogen bonds. The Ser 214–Asp 102–His 57 interactions are very similar in all compounds. The carboxylate is hydrogen bonded to Ser 214 O_γ, His 57 N_δ and His 57 N. The bond (*C*) is very long to indicate a weak hydrogen-bonding interaction. There is no conspicuous difference between the various trypsinogens, trypsins or complexes to reflect the different functional states.

However, there is a clear distinction in the His 57 N_ε–Ser 195 O_γ contact between free and liganded compounds. TGKZ to TTIV have a short N_ε–O_γ hydrogen bond indicating strong interaction while there is a long bond in all non-liganded compounds. The last column in Table 4 contains the Ser 195 C_β–O_γ–His 57 N_ε angle. Assuming rotational freedom of the H atom around the 195 C_β–O_γ bond, one expects a Ser 195 C_β–O_γ–His 57 N_ε angle of *ca* 111° for a linear 195 O_γ–57 N_ε hydrogen bond. As can be seen from Table 4, the hydrogen-bond angle improves considerably in the complexes compared to the unliganded enzymes correlated with the distance (*E*). The relatively small angle (90·8°) for APPA is due to the bifurcated hydrogen bond discussed below.

In the complexes the active site is completely shielded and inaccessible to solvent. In the unliganded compounds His 57 and Ser 195 may be transiently hydrogen bonded to solvent molecules weakening the O_γ–N_ε interaction. Solvent indeed influences the 195 O_γ–57 N_ε interaction as shown by the inverse correlation between the length of hydrogen bond (*E*) (195 O_γ–57 N_ε) and the interaction with the ordered solvent molecule 702 (*S*) (195 O_γ–Sol 702): in trypsin at pH 8

Table 4. *Hydrogen-bond distances* (Å) *at the reactive site in trypsin and the P1 bonding environment*

Letters *A* to *S* correspond to the labels in Fig. 5.

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	∠195C_β–O_γ–57N_ε
NATI	2·8	2·6	3·3	3·0	3·0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	81·2°
TNME	2·7	2·6	3·4	3·0	3·0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	76·6
TGHS	2·8	2·7	3·3	3·0	2·9	—	—	—	—	—	—	—	—	—	—	—	—	—	—	83·7
TGHH	2·8	2·7	3·4	2·9	2·8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	86·0
TRYB	2·8	2·7	3·4	2·9	3·1	—	—	—	—	—	2·8	3·0	—	2·8	—	—	—	—	3·0	82·6
DEBA	2·8	2·7	3·3	2·9	3·1	—	—	—	—	—	2·9	3·0	—	2·8	—	—	—	—	2·9	83·7
DEBB	2·7	2·6	3·3	2·8	2·8	—	—	—	—	—	2·9	2·7	—	3·0	—	—	—	—	3·4	93·7
BA	2·8	2·7	3·5	2·8	3·0	—	—	—	3·1	—	—	2·8	—	2·9	2·8	2·9	2·9	3·0	3·3	87·7
APPA	2·6	2·6	3·4	2·8	2·8	1·7	2·8	3·0	2·9	—	—	2·9	—	2·9	2·9	3·0	3·0	2·9	—	90·8
TGKZ	2·7	2·8	3·2	3·0	2·7	2·7	3·2	2·6	2·8	3·0	2·9	2·9	2·8	2·8	—	—	—	—	—	97·8
TTNA	2·7	2·6	3·2	3·1	2·6	2·7	2·8	2·8	2·8	2·8	2·6	3·0	3·0	2·8	—	—	—	—	—	103·2
TTOG	2·7	2·7	3·5	3·1	2·5	2·9	3·0	2·7	3·0	2·9	2·6	2·8	3·0	2·7	—	—	—	—	—	117·8
TTIV	2·9	2·7	3·3	3·1	2·6	2·8	2·9	2·8	2·8	2·7	2·6	3·0	2·8	2·7	—	—	—	—	—	113·2
TTAN	2·7	2·7	3·4	3·0	—	—	2·9	2·9	2·9	2·8	2·6	3·0	2·9	2·7	—	—	—	—	—	—

(DEBA) and in trigonal trypsin (TRYB) the distance (*E*) is 3·1 Å, while (*S*) is 2·9 and 3·0 Å, whereas in trypsin at pH 5 (DEBB) (*E*) is 2·8 Å, while (*S*) is 3·4 Å. Benzamidine-inhibited trypsin is intermediate with a hydrogen-bond length (*E*) of 3·0 Å and an (*S*) of 3·3 Å. The Ser 195 C_β–O_γ–His 57 N_ε angles in DEBA, TRYB, DEBB and BA nicely fit into this correlation with values of 83·7, 82·6, 93·7 and 87·7° respectively. The comparatively short distance in TGHH remains unexplained.

In APPA the His 57 N_ε is directed between Ser 195 O_γ and the APPA carboxylate indicating a bifurcated hydrogen bond (see Fig. 6 for a detailed representation of the interaction of APPA with trypsin). As shown below proton transfer from Ser 195 O_γ to His 57 N_ε has probably occurred in APPA, but not in the proteinase inhibitor complexes.

The Ser 195 O_γ–P1 Lys 15 C distance (*F*) in the PTI and PSTI complexes is between 2·7 and 2·9 Å. In TTOG it is exceptionally long. Previous real-space refinement (Huber & Bode, 1978; Bode, Schwager & Huber, 1978) had resulted in a distance of 2·6 Å in the PTI complexes.

In APPA this distance is considerably shorter at 1·7 Å. The severe overlap of the O_γ and C′ electron densities makes this value somewhat less reliable than in the previous cases although the resolution is higher. These distance relations will be discussed later.

The P1 carbonyl oxygen binding in the oxyanion binding hole is similar in all liganded structures with hydrogen bonds (*G*, *H*) to Gly 195 NH and Ser 195 NH respectively.

In the Lys P1 complexes the positively charged N_ζ is tetrahedrally coordinated to H_2O 414 and 416 and Ser 190 O uniformly in all compounds. The solvent is an integral and invariant constituent part of the P1 pocket. The lysine side chain is too short to interact directly with the negatively charged Asp 189. The Asp 189 carboxylate group is hydrogen bonded to H_2O 414 in the free trypsins and the lysine complexes and to Ser 190 N and Gly A221 N in all trypsin-containing compounds. Asp 189 and Ser 190 are part of the activation domain and disordered in the free trypsinogens.

In the P1 amidinophenyl complexes (BA and APPA) the amidino group is symmetrically bound to the Asp 189 carboxylate. In addition there are interactions with H_2O 416, Ser 190 O_γ and Gly 219 O. The amidino group may be regarded as a model for the arginine side chain bound at P1.

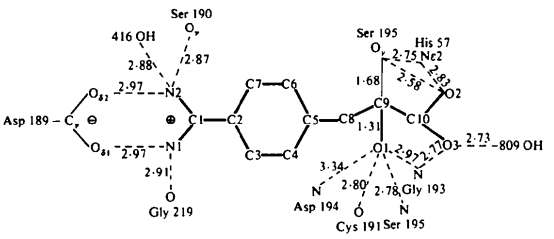

Fig. 6. Detailed schematic drawing of the APPA binding geometry. (Distances are in Å.)

3.3. *Peptide geometry*

It is clear that the geometric restraints as given by the force constants and the equilibrium values influence the results, as do the overall scaling factor, weighting, diffraction and geometric terms. The values used in this study under the conditions of this study (in particular, high-resolution data) allow deviations from the restrained geometry to fit the X-ray contributions as shown in the following.

Mean values for bond lengths and the interbond angles CAO and NAO differ only slightly from their target values, but there are significant differences for ACN (= τ), CAN and ANC bond angles. These differences are reflected in the overall number mean(all) and in the group averages [mean(struc.)]. The deviations observed may be a consequence of the predominant β secondary structure of the proteins in our sample. A preliminary analysis of a large all-α protein (citrate synthase; Remington, Wiegand & Huber, 1982), however, yields very similar values. The deviations of τ, CAN and ANC average(all) values from the restrained values are considered significant: r.m.s.(all) deviations of these bond angles (2·49, 1·75, and 2·57°) are of the order of the differences (1·22, 1·04 and 1·99° respectively). The expected errors in the average values (0·04, 0·03 and 0·04°) are thus a factor 30 to 50 smaller than the differences. The target values of Table 2, which had been derived from small peptide crystal structures, need modification.

The ω_1 and ω_3 values of the peptide group are of particular interest as they indicate deformations from planarity. Remarkably, the mean values deviate slightly from 180 and 0° by −1·0 and −1·78°. The minimum-energy configuration of a peptide group is assumed to be planar (Ramachandran & Sasisekharan, 1968; Levitt & Lifson, 1969). However, theoretical studies on oligomers of α-aminoisobutyric acid (Paterson, Rumsey, Benedetti, Nemethy & Scheraga, 1981), a molecule that adopts 3_{10}- or α-helical conformations, showed small but significant deviations from planarity in the lowest-energy conformations. The authors find a correlation between the handedness of the helix and the sign of ω_1; negative ω_1 values are associated with right-handed helical conformations and *vice versa*. The sign of ω_1 (+179°) in our study may be characteristic of β secondary structure, but an analysis of the all-α protein citrate synthase (Remington *et al.*, 1982) gives nearly the same average ω_1 (+179·1°), in contradiction to the above-stated correlation of signs. Interestingly the average ω_3 for citrate synthase ($\langle\omega_3\rangle$ = 0·41°) is different from the mean(all) value

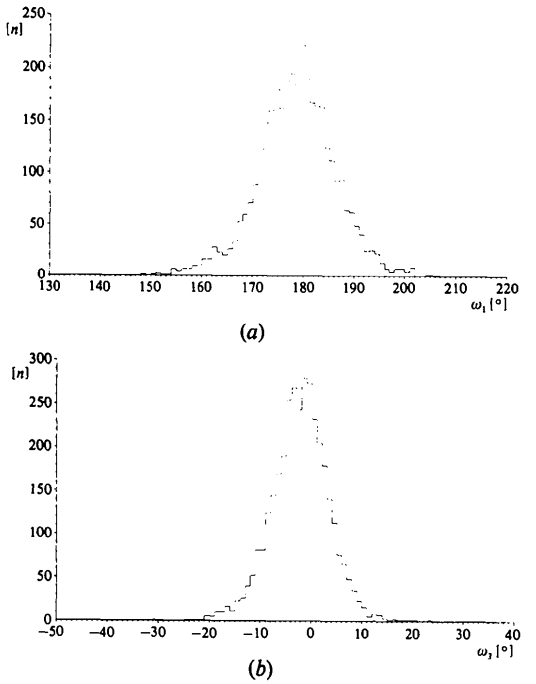

Fig. 7. Histograms showing the distribution of (*a*) ω_1, and (*b*) ω_3.

$\{\omega_3[\text{mean(all)}] = -1{\cdot}78°\}$ found for the structures listed in Table 1.

Fig. 7 is a histogram of the ω_1 and ω_3 distributions. The broad distribution of ω_1 compared to ω_3 [r.m.s.(all) 7·58 and 5·60° respectively] is obvious. ω_1 is determined by C_α, C, N, C'_α. As the C_α positions are defined also through the side-chain electron densities, the weight of the crystallographic terms should be larger than for ω_3 which is defined by O, C, N, C'_α. The O atom is connected to C′ only and its position is determined by the corresponding electron density and the geometric restraints, whose influence is expected to be larger in this case.

Having shown that the refinement procedure can indicate deviations from the target geometry we now analyse the three series (trypsinogen, trypsin, trypsin-PTI) and the single species Tg-PSTI with respect to ω_1, ω_3, and χ_C.

The average ω_1, ω_3 and χ_C product–moment correlation coefficients between all the structures listed in Table 1 are 0·62, 0·34 and 0·45 respectively.* According to Hamilton (1964) the validity of these coefficients is difficult to evaluate, but the so-called null hypothesis (*i.e.* the probability that the true correlation between structures i and j is zero for a given r_{ij}) can be tested by calculating the quantity $t = [(n-2)r_{ij}^2/(1-r_{ij}^2)]^{1/2}$ and using a Student's t distribution with $n-2$ degrees of freedom. Thus for $n = 223$ (=number of residues in trypsin) and $r_{ij} = 0{\cdot}13$ one calculates a t value which allows one to reject the null hypothesis at a 95% confidence level. Therefore only structures with correlation coefficients which are considerably smaller than 0·13 can be considered uncorrelated. This can be verified by calculating correlation coefficients between Kol-Fab (Marquart, Deisenhofer, Huber & Palm, 1980) and citrate synthase (Remington *et al.*, 1982), two structures which have nothing in common besides their residue numbers. The correlation coefficients for these uncorrelated structures range from −0·009 for ω_1 to 0·020 for χ_C.

The correlation coefficients for ω_1 are generally higher than for ω_3 and χ_C, which is derived from ω_1 and ω_3. The reason for this observation is probably the same as discussed above for the different widths of the ω_1 and ω_3 distributions: ω_1 is more defined by crystallographic terms, whereas ω_3 is more influenced by energy restraints, which lead to a more random distribution of ω_3 values among the different structures.

The ω_1 and partly also the ω_3 correlation coefficients confirm the observation already stated above in connection with the C_α comparison: There is higher correlation within the isomorphous families than between them. TRYB (trigonal trypsin) however is as highly correlated with the orthorhombic trypsin as with the isomorphous trypsinogen family.

It can also be seen from Fig. 4 that there is a high degree of correlation between the structures particularly concerning the strongly non-planar peptide groups.

EREF refinement reduced the θ^4 angle of the scissile peptide in the trypsin/trypsinogen–PTI complexes to *ca* −11° ($\chi_C \simeq -13{\cdot}0°$), a value smaller than had been found previously after (less extensive) constrained real-space refinement (Huber *et al.*, 1974). Fig. 4 (ω_1 and χ_C values of the scissile peptide are marked with an arrow) clearly shows that this value is highly significant, as both ω_1 and χ_C strongly deviate from their mean(all) values and adopt similar values (ω_1: 163·9, 165·6, 159·5, 162·8°; χ_C: −12·9, −10·4, −10·3, −18·5°) in TTNA, TTIV, TTAN and TTOG respectively.

The deformation at C′ of the scissile peptide in the trypsinogen–PSTI complex ($\theta^4 = -12{\cdot}5°$, $\chi_C = -14{\cdot}5°$) is very similar to that in the trypsin/trypsinogen–PTI complexes, although in TGKZ ω_1 differs by about 10° ($\omega_1 = 172{\cdot}8°$) from ω_1 of the other complexes.

χ_C for free PTI is close to 0° (−0·4°), indicating a planar carbonyl carbon of the scissile peptide Lys 15. Complex formation involves slight pyramidalization of this atom, too small to be observed by ^{13}C NMR techniques (Richards, Tschesche & Wüthrich, 1980).

In APPA the pyruvate carbonyl carbon is still more deformed ($\theta^4 = -37{\cdot}2°$) but not yet tetrahedral ($\theta^4 = -54°$). This is interesting as the Ser 195 O_γ–C′

* See deposition footnote.

distance of 1·7 Å indicates covalent-bond formation. It seems as if binding of the carbonyl group in the oxyanion binding hole favours a geometry intermediate between trigonal and tetrahedral, quite independent of the O_γ–C′ bond formation.

This is particularly obvious in the anhydrotrypsin–PTI complex which lacks the Ser 195 O_γ as this residue is converted to dehydroalanine. Yet, the pyramidalization of C′ is very similar to that found in other complexes.

We emphasize that in the EREF refinement 'normal' force constants had been used throughout to restrain to standard geometry with the exception of the pyruvate carbonyl group of APPA. The very high resolution justified weaker restraints here. It is clear that, in general, restraints oppose deviations from standard geometry. The true distortions may be larger than observed in this study.

APPA is a very potent inhibitor of trypsin with K_i = 0·0016 mM l^{-1} (Stürzebecher, Markwardt, Richter, Voigt, Wagner & Walsmann, 1976). Pyruvate derivatives and other long-chain α-keto carboxylic acids have been analysed by single-crystal X-ray analysis (Tavale *et al.*, 1961; Jain, Tavale & Biswas, 1966). These showed that the α-keto group is trigonal planar. The bond between the keto carbon and the carboxyl carbon is very long, with, in particular, a value of 1·6 Å in the acid salts. In Fig. 6 APPA is illustrated in detail. It differs from the PSTI and PTI complexes by a strong interaction between the pyruvate carboxylate and His 57. This favours proton transfer from Ser 195 O_γ to N_ε His 57 to yield a positively charged imidazole. The Ser 195 O_γ oxyanion approaches C′ closer than the seryl O_γ–H does in the PTI and PSTI complexes.

We have regarded the bonding within the PTI–trypsin complexes as intermediate states of tetrahedral adduct formation in analogy to the amine carbonyl carbon addition compounds described by Bürgi, Dunitz & Shefter (1973). In analogy to their study we defined Δ as the out-of-plane distance of C′ (the plane being defined by C_α N and O) and d_1 as the O_γ–C′ distance. Fig. 8 is a plot of the d_1, Δ relation and the fitting curve derived from equation (1) in Bürgi *et al.* (1973). The circles denote the small molecules from which Bürgi *et al.* derived the relation between Δ and d_1, and the crosses mark the complexes of our study. It is clear that the four complexes TTNA, TGKZ, TTIV and APPA fit the curve remarkably well; TTOG deviates. This correspondence is interesting as the fitting curve was derived from compounds with an amine nitrogen as nucleophile interacting with a keto carbonyl carbon. In the protein complexes an alcoholic oxygen (polarized by Asp 102–His 57) interacts with a peptide carbonyl group.

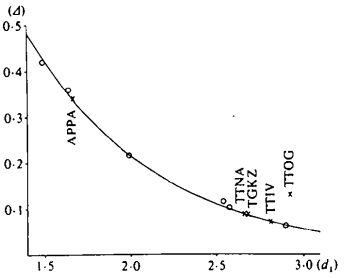

Fig. 8. Relation between Δ and d_1 as described in equation (1) in Bürgi, Dunitz & Shefter (1973).

References

BODE, W., FEHLHAMMER, H. & HUBER, R. (1976). *J. Mol. Biol.* **106**, 325–335.

BODE, W. & HUBER, R. (1976). *FEBS Lett.* **68**, 231–235.

BODE, W. & HUBER, K. (1978). *FEBS Lett.* **90**, 265–269

BODE, W. & SCHWAGER, P. (1975). *J. Mol. Biol.* **98**, 693–617.

BODE, W., SCHWAGER, P. & HUBER, R. (1976). *Proteolysis and Physiological Regulation*, Miami Winter Symposia, Vol. 11, edited by D. W. RIBBONS & K. BREW, pp. 43–76. New York: Academic Press.

BODE, W., SCHWAGER, P. & HUBER, R. (1978). *J. Mol. Biol.* **118**, 99–112.

BOLOGNESI, M., GATTI, G., MENEGATTI, E., GUARNERI, M., MARQUART, M., PAPAMOKOS, E. & HUBER, R. (1982). *J. Mol. Biol.* **162**, 839–868.

BÜRGI, H. B., DUNITZ, J. D. & SHEFTER, E. (1973). *J. Am. Chem. Soc.* **95**, 5065–5067.

DEISENHOFER, J. (1981). *Biochemistry*, **20**, 2361–2370.

DEISENHOFER, J. & STEIGEMANN, W. (1975). *Acta Cryst.* B**31**, 238–250.

DIAMOND, R. (1971). *Acta Cryst.* A**27**, 436–452.

DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.

FEHLHAMMER, H., BODE, W. & HUBER, R. (1977). *J. Mol. Biol.* **111**, 415–438.

HAMILTON, W. C. (1964). In *Statistics in Physical Science*. New York: Ronald Press.

HUBER, R. & BODE, W. (1978). *Acc. Chem. Res.* **11**, 114–122.

HUBER, R., BODE, W., KUKLA, D. & KOHL, V. (1975). *Biophys. Struct. Mech.* **1**, 189–201.

HUBER, R., KUKLA, D., BODE, W., SCHWAGER, P., BARTELS, K., DEISENHOFER, J. & STEIGEMANN, W. (1974). *J. Mol. Biol.* **89**, 73–101.

JACK, A. T. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931–935.

JAIN, S. C., TAVALE, S. S. & BISWAS, A. B. (1966). *Acta Cryst.* **21**, 445.

JENCKS, W. P. (1975). *Adv. Enzymol.* **43**, 219–410.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

LEVITT, M. (1974). *J. Mol. Biol.* **82**, 393–420.

LEVITT, M. & LIFSON, S. (1969). *J. Mol. Biol.* **46**, 269.

LUZZATI, P. V. (1952). *Acta Cryst.* **5**, 802–810.

MARQUART, M., DEISENHOFER, J., HUBER, R. & PALM, W. (1980). *J. Mol. Biol.* **141**, 369–391.

PAPAMOKOS, E., WEBER, E., BODE, W., HUBER, R., EMPIE, M. W., KATO, I. & LASKOWSKI, M. JR (1982) *J. Mol. Biol.* **158**, 515–537.

PATERSON, Y., RUMSEY, S. M., BENEDETTI, E., NEMETHY, G. & SCHERAGA, H. A. (1981) *J. Am. Chem. Soc.* **103**, 2947–2955.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283.

REMINGTON, S., WIEGAND, G. & HUBER, R. (1982). *J. Mol. Biol.* **158**, 111–152.

RICHARDS, R., TSCHESCHE, H. & WÜTHRICH, K. (1980). *Biochemistry*, **19**, 5711–5715.

RÜHLMANN, A., KUKLA, D., SCHWAGER, P., BARTELS, K. & HUBER, R. (1973). *J. Mol. Biol.* **77**, 417–436.
SCHWAGER, P. & BARTELS, K. (1977). *The Rotation Method in Crystallography*, Ch. 10, edited by U. W. ARNDT & A. J. WONACOTT. Amsterdam, New York, Oxford: North-Holland.
SINGH, T. J., BODE, W. & HUBER, R. (1980). *Acta Cryst.* B**36**, 621–627.
STÜRZEBECHER, J., MARKWARDT, F., RICHTER, P., VOIGT, B., WAGNER, G. & WALSMANN, P. (1976). *Acta Biol. Med. Ger.* **35**, 1665–1676.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. (1977). *Acta Cryst.* A**33**, 800–804.
TAVALE, S. S., PANT, L. M. & BISWAS, A. B. (1961). *Acta Cryst.* **14**, 1281–1286.
WALTER, J. (1982). Thesis, Techn. Univ. München.
WALTER, J., STEIGEMANN, W., SINGH, T. P., BARTUNIK, H., BODE, W. & HUBER, R. (1982). *Acta Cryst.* B**38**, 1462–1472.
WARSHEL, A., LEVITT, M. & LIFSON, S. (1970). *J. Mol. Spectrosc.* **33**, 84.
WEBER, E., PAPAMOKOS, E., BODE, W., HUBER, R., KATO, I. & LASKOWSKI, M. JR (1981). *J. Mol. Biol.* **149**, 109–123.
WINKLER, F. K. & DUNITZ, J. D. (1971). *J. Mol. Biol.* **59**, 159–182.

Acta Cryst. (1983). B**39**, 490–494

The Molecular Symmetry of Histidine Decarboxylase and Prohistidine Decarboxylase by Rotation-Function Analysis

BY E. H. PARKS, K. CLINGER AND M. L. HACKERT

Department of Chemistry and Clayton Foundation Biochemical Institute, University of Texas at Austin, Austin, Texas 78712, *USA*

(*Received* 31 *August* 1982; *accepted* 19 *January* 1983)

Abstract

Several crystal forms of histidine decarboxylase (HDC) and prohistidine decarboxylase (pHDC) have been obtained, all of which contain multiple protomers per crystallographic asymmetric unit. A tetragonal crystal form of HDC and a trigonal crystal form of pHDC have been selected for further X-ray analysis. Data to resolutions of 5·6 and 5·5 Å have been collected by diffractometry for HDC and pHDC, respectively. Rotation-function studies were used to locate the non-crystallographic molecular symmetry axes. Both proteins were shown to possess 32 molecular symmetry and a cross rotation function was used to relate the two proteins in their respective unit cells. The presence of pseudo-symmetry-operator peaks complicated the analysis of the rotation-function results. In this case, the interpretation was aided by the use of cross-rotation-function results and rotation-function results based on subunit packing models.

Introduction

Histidine decarboxylase from *Lactobacillus* 30*a* is an unusual amino acid decarboxylase in that it contains a covalently bound pyruvoyl residue at the active site rather than pyridoxal-5′-phosphate (Snell, 1977; and references therein). This pyruvoyl moiety is formed from an internal serine residue during the activation of a proenzyme subunit. This proenzyme (pHDC, subunit molecular weight = 37 000) is cleaved by a seemingly intramolecular process into an α subunit ($M_r = 28\,000$) and a β subunit ($M_r = 9\,000$). It has been demonstrated that the pyruvoyl residue is required for catalysis by the formation of a Schiff-base intermediate with both histidine and histamine in a similar manner to that of pyridoxal-phosphate-dependent enzymes.

A mutant of *Lactobacillus* 30*a* was found which possessed an inactive form of histidine decarboxylase which was very slowly converted to the active histidine decarboxylase (HDC). It has recently been shown that the proenzyme (pHDC) has six associated subunits, $(\pi)_6$, whereas the native enzyme has an $(\alpha\beta)_6$ subunit stoichiometry, both forms having a molecular weight of 208 000 (Hackert, Meador, Oliver, Salmon, Recsei & Snell, 1981).

Histidine decarboxylase has been crystallized in several distinct forms, two of which have been selected for further crystallographic investigations. HDC crystallizes in a tetragonal space group ($I4_122$; $a = b = 222$, $c = 107$ Å). The unit cell contains eight molecules which occupy the 16 general positions of the unit cell. This implies that the HDC molecule has at least twofold molecular symmetry. Mutant pHDC crystallizes in a trigonal space group ($P321$; $a = b = 100$, $c = 164{\cdot}5$ Å). In this system the unit cell contains only two molecules which are positioned on the two internal

0108-7681/83/040490-05$01.50