*J. Mol. Biol.* (1986) **190**, 259–267

# Stereochemistry of Binding of the Tetrapeptide Acetyl-Pro-Ala-Pro-Tyr-$NH_2$ to Porcine Pancreatic Elastase

## Combined Use of Two-dimensional Transferred Nuclear Overhauser Enhancement Measurements, Restrained Molecular Dynamics, X-ray Crystallography and Molecular Modelling

A nuclear magnetic resonance study of the conformation of the tetrapeptide acetyl-Pro-Ala-Pro-Tyr-$NH_2$ bound to porcine pancreatic elastase is presented. From two-dimensional transferred nuclear Overhauser enhancement measurements, a set of 23 approximate distance restraints between pairs of bound ligand protons, indicative of an extended type structure, is derived. The structure of the bound tetrapeptide is then refined from two different starting structures (an extended $\beta$-strand and a polyproline helix) by restrained molecular dynamics, in which the interproton distances are incorporated into the total energy of the system in the form of effective potentials. Convergence to essentially the same average restrained dynamics structures is achieved. The refined structures are then modelled into the active site of elastase by interactive molecular graphics. The determination of the anchor point of the bound tetrapeptide on the enzyme was aided by a simultaneous crystallographic study which, despite the fact that only electron density for a Pro-X dipeptide fragment was visible, enabled both the approximate position and orientation of binding to be determined. It is found that the tetrapeptide is bound in the S′ binding site in the reverse orientation found in other serine protease–inhibitor complexes and is stabilized both by hydrogen-bonding and by van der Waals' interactions.

Elastase belongs to a class of enzymes known as the serine proteases whose members include, amongst others, trypsin and chymotrypsin. The design of novel elastase inhibitors has been the subject of recent interest, as the elastases have been linked to degenerative diseases such as pancreatitis, emphysema and arthritis (Powers, 1976; Janoff, 1983; Lungarella *et al.*, 1981). Because the design of such inhibitors is dependent on the interaction of inhibitor functional groups with receptor subsites on the enzyme, a range of biophysical methods has been used to analyse the interactions of substrates, products, and peptide and heterocyclic inhibitors with elastase (Shotton *et al.*, 1972; Dimicoli *et al.*, 1980; McRae *et al.*, 1980; Bizzozero & Dutler, 1981; Hughes *et al.*, 1982; Harper & Powers, 1985; Harper *et al.*, 1985; Meyer *et al.*, 1985, 1986). The present study focuses its attention on one such "natural" oligopeptide inhibitor; namely, the tetrapeptide acetyl-Pro-Ala-Pro-Tyr-$NH_2$.

A recent crystallographic study of a complex of the tripeptide acetyl-Ala-Pro-Ala with porcine elastase at 1·65 Å resolution (Meyer *et al.*, 1986) revealed that two tripeptide molecules were bound per molecule of enzyme in an orientation backwards with respect to the direction of binding seen in the trypsin–bovine pancreatic trypsin inhibitor complex (Huber *et al.*, 1974). Backward binding was likewise exhibited (but not recognized) in the elastase complex with the trifluoroacetylated inhibitor $CF_3CO$-Lys-Ala-NH-$C_6H_4$-*p*-$CH_3$ (Hughes *et al.*, 1982). These unexpected results suggest that the stereochemistry of interaction of elastase with small peptide inhibitors is somewhat unpredictable. In order to extend our understanding of such interactions as well as to gain additional information that could contribute to drug design methodologies, we have investigated the stereochemistry of binding of the weak tetrapeptide inhibitor acetyl-Pro-Ala-Pro-Tyr-$NH_2$ ($K_I \sim 10^{-3}$ M; J. C. Powers, personal communication) by a combination of: (1) n.m.r.† spectroscopy using two-dimensional transferred nuclear Overhauser enhancement measurements to demonstrate the proximity of bound ligand protons; (2) restrained molecular dynamics to refine the structure of the bound tetrapeptide on the basis of the interproton

† Abbreviations used: n.m.r., nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser enhancement or effect; TRNOE, transferred nuclear Overhauser enhancement or effect; r.m.s., root-mean-square; RD, restrained molecular dynamics. N.B. the non-SI unit cal = 4·184 J.



0022-2836/86/140259-09 $03.00/0

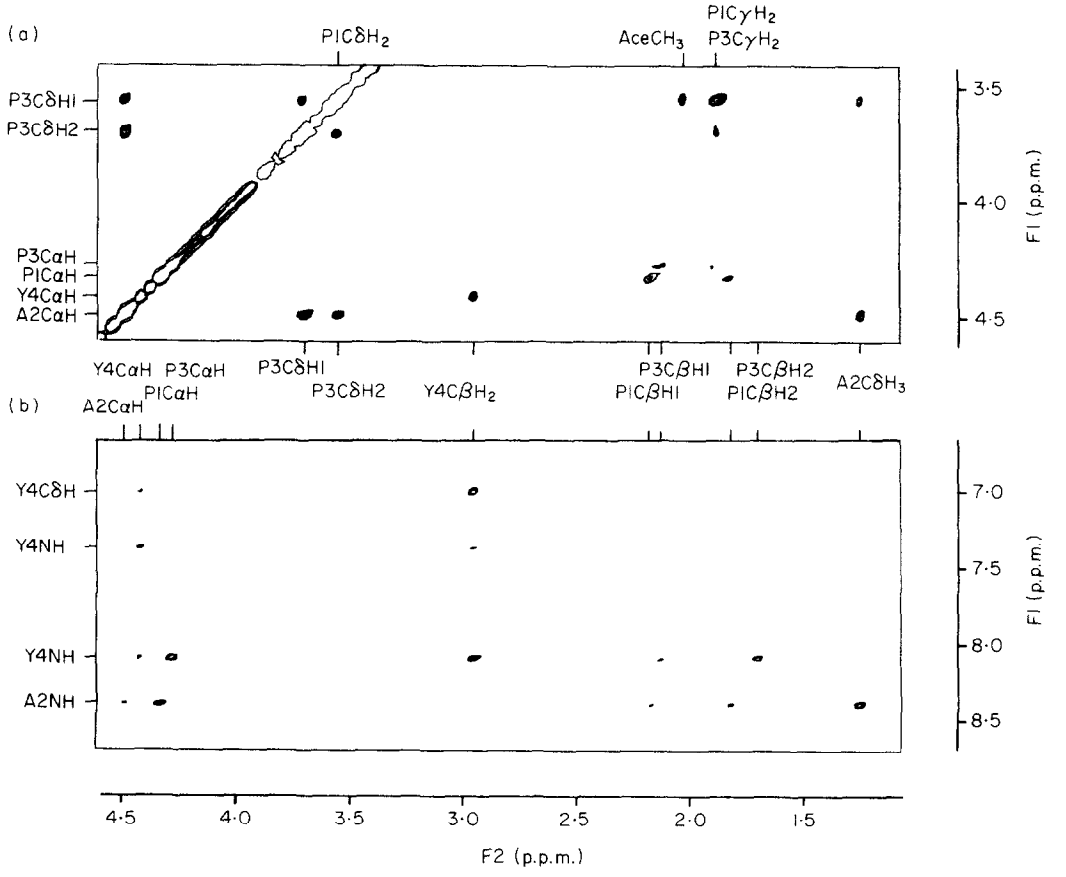


**Figure 1.** (a) NH/aromatic (F1 axis)–aliphatic (F2 axis) and (b) CαH/CδH (F1 axis)–aliphatic (F2 axis) regions of the pure-phase absorption 500 MHz 2-dimensional TRNOE spectra (mixing time 0·1 s) of 11 mM-acetyl-Pro-Ala-Pro-Tyr in the presence of 1 mM-elastase at 5°C. The experimental conditions were as follows: 11 mM-tetrapeptide and 1 mM-porcine pancreatic elastase in (a) 99·96% $^2H_2O$ and (b) 90% $H_2O$/10% $^2H_2O$ buffer containing 25 mM-potassium phosphate (pH 5·0), 250 mM-KCl, 0·05 mM-EDTA. All n.m.r. spectra were recorded on a Bruker AM500 spectrometer. The pulse sequence for a 2-dimensional TRNOE spectrum is identical to that for a conventional 2-dimensional NOE spectrum (Jeener *et al.*, 1979). The 2-dimensional TRNOE spectra at a mixing time of 100 ms were recorded in the pure-phase absorption mode using the time-proportional phase incrementation method (Redfield & Kuntz, 1975; Bodenhausen *et al.*, 1980) as described by Marion & Wuthrich (1983). Appropriate phase cycling was used for the suppression of axial peaks and of peaks due to multiple quantum coherence transfer; in addition, a 10% variation in the mixing time was used to eliminate zero quantum coherence transfer (Macura *et al.*, 1981); 128 transients were collected for each of 512 increments with a relaxation delay of 1 s between successive transients. The sweep width employed was 7042 Hz with the carrier placed in the middle of the spectrum. The digital resolution employed was 6·89 Hz/point in both dimensions. This was achieved by appropriate zero-filling in the $t_1$ dimension only. For the measurements in 90% $H_2O$/10% $^2H_2O$, the solvent resonance was suppressed by selective irradiation during the relaxation delay and the mixing time (Wider *et al.*, 1984). In addition to the 100 ms mixing time 2-dimensional TRNOE spectra, spectra were recorded with a mixing time of 150 ms. Further, several time-dependent 1-dimensional TRNOE measurements (Clore & Gronenborn, 1983) were carried out, irradiating a few selected resonances to ensure that at 100 ms the initial rate approximation $N_{ij}(t) \sim (1-a)\sigma_{ij}^{BB}t$ was fulfilled. Porcine pancreatic elastase was purchased from Serva (20909, Control C) and used without further purification. The tetrapeptide was a gift from Professor J. C. Powers. The amino acids are labelled according to the 1-letter code and the chemical shifts are given relative to 4,4-dimethylsilapentane-1-sulphonate. p.p.m., parts per million.

distance data derived from the TRNOE experiments; (3) X-ray crystallography to locate the anchor point of the bound tetrapeptide on the enzyme; and (4) interactive molecular graphics to model the structure of the bound tetrapeptide, as determined from the TRNOE measurements and restrained molecular dynamics calculations, into the active site of elastase.

The theory of the time-dependent TRNOE has been discussed in detail (Clore & Gronenborn, 1983) and used extensively to determine the conformations of ligands bound to proteins (e.g. see Gronenborn & Clore, 1982*a*, *b*; Gronenborn *et al.*, 1984*a*, *b*; Clore *et al.*, 1984, 1986), so only the pertinent points will be summarized here. When exchange is fast on the chemical shift scale so that only a single set of exchanged broadened average ligand resonances is observed, as in the present case, the observed initial slope of the time development of the TRNOE, $N_{ij}(t)$, between two protons $i$ and $j$ is given by $a\sigma_{ij}^{FF} + (1-a)\sigma_{ij}^{BB}$, where $\sigma_{ij}^{FF}$ and $\sigma_{ij}^{BB}$ are the cross-relaxation rates between protons $i$ and $j$ in the free and bound states, respectively, and $a$ is the mole fraction of free

**Table 1**

*Proton resonance assignments of the tetrapeptide acetyl-Pro-Ala-Pro-Tyr-$NH_2$ at 5°C*

| | Resonance | | | | | | |
|---|---|---|---|---|---|---|---|
| Residue | NH | CαH | CβH | CγH | CδH | CεH | Others |
| Acetyl | | | | | | | $CH_3$ 2·04 |
| Pro | | 4·33 | 2·21, 1·85 | 1·91, 1·91 | 3·56, 3·56 | | |
| Ala | 8·48 | 4·51 | 1·27 | | | | |
| Pro | | 4·29 | 8·15, 1·71 | 1·91, 1·91 | 3·71, 3·56 | | |
| Tyr | 8·18 | 4·42 | 2·98, 2·98 | | 7·09 | 6·79 | $NH_2$ 7·46, 7·09 |

The resonance positions of the free tetrapeptide were assigned by a combination of 2-dimensional pure-phase absorption MLEV 17 Hartmann–Hahn spectroscopy (Davis & Bax, 1985; Bax & Davis, 1985) at 2 mixing times (15 and 40 ms) in both $^2H_2O$ and 90% $H_2O$/10% $^2H_2O$ in order to demonstrate direct and relayed through-space connectivities. The buffer conditions used were the same as those for the TRNOE experiments. The specific assignment of the 2 proline residues was made by combining the through-bond connectivities obtained from the Hartmann–Hahn spectra with the sequential through-space connectivities observed in the 2-dimensional TRNOE spectra.

ligand. This simple relationship applies to both the one-dimensional and two-dimensional TRNOE experiments. For a small ligand molecule such as the tetrapeptide (524 $M_r$) and a large macromolecule such as elastase ($\sim$22,000 $M_r$), the free and bound states correspond to two distinctive regimes; namely, the $\omega\tau_c < 1$ and $\omega\tau_c > 1$ regimes, respectively (where $\omega$ is the spectrometer frequency and $\tau_c$ is the correlation time). As a result, the cross-relaxation rates in the free and bound states are of opposite sign, such that the NOEs in the free state are positive, whereas the TRNOEs are negative. (Note that in a pure-phase absorption two-dimensional TRNOE experiment, a negative NOE gives rise to a cross-peak with the same sign as that of the diagonal peaks, whereas a positive NOE gives rise to a cross-peak of opposite sign.) In the case of the free tetrapeptide, no positive NOEs could be observed at mixing times below 0·5 second. Thus, at short mixing times (<0·2 s), the magnitude of the negative TRNOE, $N_{ij}(t)$, in our experiments, is simply given by the approximation $N_{ij} \sim (1-a)\sigma_{ij}^{BB}t$. As the cross-relaxation rates are proportional to $\langle r^{-6}\rangle$, distance information can be obtained from the relative intensities of the TRNOE cross-peaks using cross-peaks corresponding to known distances as internal references (e.g. the distance between methylene protons or between the Hδ1/Hδ2 and Hε1/Hε2 protons of tyrosine).

Figure 1 shows two regions of the pure-phase absorption two-dimensional TRNOE spectra obtained with a ratio of free to bound ligand of 10 and a mixing time of 0·1 second. The first region (Fig. 1(a)) illustrates TRNOE connectivities between aliphatic protons, and the second (Fig. 1(b)) illustrates TRNOE connectivities from the NH and aromatic protons to the aliphatic protons. The assignments of the resonances of the tetrapeptide are given in Table 1. Three points are worth mentioning with respect to the two-dimensional TRNOE spectra. (1) Only cross-peaks between averaged ligand resonances are observed; no protein–protein or protein–ligand cross-peaks are seen. (The absence of any protein–protein cross-peaks is hardly surprising, given the low concentration of elastase employed and the relatively broad linewidths of the protein resonances.) (2) All cross-peaks exhibit the same sign as the diagonal peaks, indicating that all observed effects correspond to negative TRNOEs and thus arise from the bound tetrapeptide. (3) The effects seen are highly selective and no evidence of spin-diffusion is observed. A further important point is that during the entire course of the TRNOE experiments (which lasted several days), no sign of cleavage of the tetrapeptide could be detected. A summary of the observed TRNOEs is given in Table 2.

Because of the relatively large errors involved in two-dimensional integration of cross-peaks, we decided to classify the cross-peaks into three categories; strong (relative intensity >100); medium (relative intensity 40 to 100) and weak (relative intensity <40). As internal standards, we used the cross-peaks between the Hδ1/Hδ2 and Hε1/Hε2 protons of tyrosine ($r = 2{\cdot}5$ Å; note that as the tyrosine ring flips in free solution at a very fast rate, only a single averaged resonance is observed for the Hδ1 and Hδ2 protons, and for the Hε1 and Hε2 protons; consequently, the observed cross-peak intensity divided by 2 corresponds to the distance of 2·5 Å) between the methylene Cβ protons of the proline residues ($r = 1{\cdot}8$ Å), and between the Cα and methyl protons of alanine ($\langle r^{-6}\rangle)^{-1/6} = 2{\cdot}5$ Å). The distances were then calibrated, essentially as described by Williamson *et al.* (1985) and Clore *et al.* (1985), as follows: strong cross-peaks were assigned to a distance of 2·0(±0·5) Å, medium cross-peaks to 2·5(±0·5) Å and weak cross-peaks to 3·0 (+1·0/−0·5) Å. In addition, in the case of distances involving methylene protons for which stereospecific assignments could not be made, methyl protons and the CδH1/CδH2 protons of tyrosine, a single $(\langle r^{-6}\rangle)^{-1/6}$ distance was used. These distances are summarized in Table 2.

The effective correlation times $\tau_{eff}(ij)$ for fixed

**Table 2**
*Distance restraints (Å) derived from the two-dimensional TRNOE spectra*

| | | | |
|---|---|---|---|
| A. *Intraresidue* | | | |
| Pro1 | | Pro3 | |
| CαH–CβH1† | 2·5 | CαH–CβH1† | 2·5 |
| CαH–CβH2† | 3 | CαH–CβH2† | 3 |
| CαH–$C\gamma H_2$ | 3 | | |
| Ala2 | | Tyr4 | |
| CαH–$C\beta H_3$ | 2·5 | CαH–NH | 3 |
| NH–CαH | 3 | CαH–$C\beta H_2$ | 2·5 |
| NH–$C\beta H_3$ | 2 | CαH–Cδ2H2‡ | 3 |
| | | $C\beta H_2$–Cδ1H1/H2 | 2·5 |
| | | NH–$C\beta H_2$ | 2·5 |
| | | NH–Cδ1H1‡ | 3 |
| B. *Interresidue* | | | |
| Acetyl–Pro1 | | Ala2–Pro3 | |
| CH3(Ace)–$C\delta H_2$(1) | 2·5 | CαH(2)–$C\delta H_2$(3) | 2 |
| | | $C\beta H_3$(2)–$C\delta H_2$(3) | 2·5 |
| Pro1–Ala2 | | Pro3–Tyr4 | |
| CαH(1)–NH(2) | 2 | CαH(3)–NH(4) | 2 |
| CβH1(1)–NH(2) | 3 | CβH1(3)–NH(4) | 3 |
| CβH2(1)–NH(2) | 3 | CβH2(3)–NH(4) | 3 |

The error estimates in the interproton distances are as follows: for $r_{ij} = 2$ Å, 2·5 Å and 3 Å, they are +0·5/−0·5 Å, +0·5/−0·5 Å and +1·0/−0·5 Å, respectively.

† In the case of the Cβ protons of proline, stereospecific assignments are easily made as the Cα proton is always closer to the *cis* Cβ proton than to the *trans* Cβ proton.

‡ As the tyrosine ring flips rapidly, only a single averaged resonance for the Cδ1H1 and Cδ2H2 protons can be observed. However, simple model building shows that a single CδH proton cannot be within 4 Å of the Cα proton and the NH proton of the same residue simultaneously. For this reason, we assigned the TRNOE between the CδH resonance and the CαH and NH resonances as arising from 2 different CδH protons. This ensures the same orientation of the tyrosine ring in the refined structures. In the case of the $C\beta H_2$–Cδ1H1/Cδ2H2 distance, however, a single $(\langle r^{-6}\rangle)^{-1/6}$ mean distance is used.

interproton distance vectors were determined from the initial slope of the time development of the corresponding TRNOEs measured by one-dimensional difference spectroscopy (as described by Clore & Gronenborn, 1983) using the equation (Solomon, 1955):

$$\frac{\sigma_{ij}^{obs}}{(1-a)} = \frac{\gamma^4\hbar^2}{10r_{ij}^6}\left(\tau_{eff}(ij) - \frac{6\tau_{eff}(ij)}{1+4\omega^2\tau_{eff}(ij)^2}\right), \quad (1)$$

where $\sigma_{ij}^{obs}$ is the observed cross-relaxation rate, $\gamma$ is the gyromagnetic ratio of the proton, and $\hbar$ Planck's constant divided by $2\pi$. The values of $\sigma_{ij}^{obs}$ measured for the Pro1(CβH1–CβH2), Ala2(CαH–$C\beta H_3$), Pro3(CβH1–CβH2) and Tyr4(Hδ1–Hε1)/(Hδ2–Hε2) interproton vectors were 2, 0·5, 2 and 0·5 second$^{-1}$, respectively. These values correspond to an effective correlation time of approximately 22 nanoseconds for the Ala2(CαH–$C\beta H_3$) and Tyr4(Hδ1–Hε1)/(Hδ2–Hε2) vectors and of approximately 12 nanoseconds for the CβH1–CβH2 vectors of Pro1 and Pro3. The effective correlation time of 22 nanoseconds at 5°C corresponds to a 1:1 ligand–protein complex of 20,000 to 25,000 $M_r$. The smaller effective correlation time for the CβH1–CβH2 proline vectors implies a degree of internal flexibility for the proline rings of the bound tetrapeptide, analogous to that observed for the sugar rings in DNA fragments (Clore & Gronenborn, 1984; Clore *et al.*, 1986). It is important to note, however, that the choice of distance ranges and errors used above is such that there is sufficient overlap between adjacent distance ranges to ensure that any uncertainty in distance assignments arising from these differences in effective correlation times has no effect on the end result. Thus, for example, when the Tyr4(Hδ1–Hε1) or Ala2(CαH–$C\beta H_3$) distances are calculated from ratios of cross-relaxation rates using the CβH1–CβH2 cross-relaxation rate and distance as an internal reference from the relationship $r_{kl}/r_{ij} = (\sigma_{ij}/\sigma_{kl})^{1/6}$, a value of 2·3 Å is obtained which represents an underestimate of only 0·2 Å, well within the error ranges used.

The approximate conformation of the bound tetrapeptide is easily deduced from a qualitative interpretation of the interresidue TRNOEs. In particular, the strong CαH($i$)–NH($i+1$) and weak CβH($i$)–NH($i+1$) TRNOEs involving the Pro1/Ala2 and Pro3/Tyr4 pairs, coupled with the weak CαH($i$)–NH($i$) intraresidue TRNOEs for the Ala2 and Tyr4 residues, and the strong TRNOE between the CαH proton of Ala2 and the CδH protons of Pro3 are indicative of an extended-type conformation.

In order to obtain a more precise structure for the bound tetrapeptide, we proceeded to carry out restrained molecular dynamics calculations (Clore *et al.*, 1985; Nilsson *et al.*, 1986). This involves the simultaneous solution of the classical equations of motion for all atoms of the tetrapeptide for a suitable time period (McCammon *et al.*, 1974, 1979; Karplus & McCammon, 1983) with the experimental interproton distances incorporated into the total energy function of the system in the form of effective potentials. The effective potential used is the quadratic biharmonic potential described by Clore *et al.* (1985):

$$E_{NOE}(r_{ij}) = \begin{cases} c_1(r_{ij}-r_{ij}^0)^2, & \text{if } r_{ij} > r_{ij}^0 \\ c_2(r_{ij}-r_{ij}^0)^2, & \text{if } r_{ij} < r_{ij}^0 \end{cases} \quad (2)$$

where $r_{ij}^0$ and $r_{ij}$ are the target and calculated interproton distances, respectively, and $c_1$ and $c_2$ are force constants given by:

$$c_1 = \frac{kTS}{2(\Delta_{ij}^+)2}, \quad c_2 = \frac{kTS}{2(\Delta_{ij}^-)^2}, \quad (3)$$

where $k$ is the Boltzmann constant, $T$ is absolute temperature, $S$ is a scale factor, and $\Delta_{ij}^+$ and $\Delta_{ij}^-$ are the positive and negative error estimates of $r_{ij}$, respectively. In the energy minimization and molecular dynamics calculations, the scale factor $S$ was set to 6, so that error estimates of 0·5 and 1·0 Å correspond to force constants of 7·15 and 1·79 kcal/mol, respectively.

Two refinement calculations were carried out starting from two idealized extended-type

**Table 3**

*Energies of the initial (Ini I, Ini II) and average restrained dynamics (RDI, RDII) structures of the tetrapeptide*

| | Ini I | Ini II | RDI† | RDII† |
|---|---|---|---|---|
| Energies (kcal/mol)‡ | | | | |
| Total | 870 | 345 | 125 (0·05) | 50 (0·09) |
| Potential | 870 | 345 | 78 (5) | 78 (5) |
| Kinetic | — | — | 47 (5) | 52 (5) |
| Bond (68) | 211 | 12 | 11 (3) | 12 (3) |
| Angle (124) | 129 | 22 | 46 (4) | 50 (5) |
| Torsion (27) | 26 | 39 | 19 (3) | 18 (3) |
| Improper (21) | 0 | 0 | 3 (1) | 4 (1) |
| van der Waals | 463 | 235 | −5 (2) | −6 (2) |
| Electrostatic | −13 | 3 | −7 (4) | −10 (4) |
| H-bond | 0 | 0 | −1 (1) | −1 (1) |
| Restraints (24) | 55 | 35 | 12 (2) | 10 (2) |
| Temperature (K) | | — | 282 (27) | 312 (29) |
| The r.m.s. difference between calculated and target inter-proton distances | 0·71 | 0·60 | 0·29 | 0·30 |

All energy minimization and molecular dynamics calculations were carried out on a CRAY 1A computer using the program CHARMM (Brooks *et al.*, 1983) optimized for the CRAY (Brunger & Karplus, unpublished). The calculations were carried out as described (Clore *et al.*, 1985), except that all hydrogen atoms were treated explicitly in the empirical energy potential.

† The energies for the dynamics structures are obtained by averaging the energies over the structures from the last 10 ps of the dynamics run. The r.m.s. fluctuations in the energy terms are given in parentheses.

‡ The number of terms for the bond, angle, torsion, improper and restraints energies is given in parentheses.

structures: (1) an extended $\beta$-strand known as Ini I ($\phi = -139°$, $\psi = 180°$), and (2) a polyproline helix known as Ini II ($\phi = -80°$, $\psi = 135°$). Each structure was then subjected to the following steps: (1) 500 cycles of restrained energy minimization; (2) 1 picosecond of equilibration, during which time the structure was heated up from 200 K to 300 K in steps of 10 K every 0·1 picosecond; (3) 6 picoseconds of thermalization in which the initial velocities at 300 K were reassigned every 0·6 picosecond; and (4) 20 picoseconds of restrained dynamics without adjusting the temperature of the system. The final averaged structures were then

**Table 4**

*Atomic r.m.s. difference (Å) between the initial (Ini I, Ini II) and average restrained dynamics (RDI, RDII) structures*

| | Atomic r.m.s. difference (Å) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | All atoms | | | All atoms excluding tyrosine ring | | | Backbone atoms | | |
| | Ini I | RDI | RDII | Ini I | RDI | RDII | Ini I | RDI | RDII |
| Ini II | 2·65 | 2·56 | 2·21 | 2·08 | 0·85 | 0·81 | 1·22 | 0·67 | 0·63 |
| Ini I | | 3·16 | 2·84 | | 2·01 | 2·05 | | 0·99 | 1·01 |
| RDI | | | 0·60 | | | 0·10 | | | 0·11 |

**Table 5**

*The $\phi$ and $\psi$ angles of the average restrained dynamics structures RDI and RDII*

| | $\phi$ | | $\psi$ | |
|---|---|---|---|---|
| Residue | RDI | RDII | RDI | RDII |
| Pro1 | −98 | −99 | 170 | 170 |
| Ala2 | −116 | −113 | 179 | −178 |
| Pro3 | −108 | −112 | 173 | 160 |
| Tyr4 | −117 | −126 | — | — |

obtained by averaging the co-ordinates over the last 10 picoseconds of the dynamics run. The average restrained dynamics structure known as RDI is derived from Ini I, and that known as RDII is derived from Ini II.

The energies of the initial and average restrained dynamics structures are given in Table 3; the r.m.s. difference over all atoms, over all atoms excluding the tyrosine ring and over the backbone atoms are given in Table 4; the values of the $\phi$ and $\psi$ angles are given in Table 5; and stereo views of the initial and average restrained dynamics structures are shown in Figure 2. From these data, it is clear that the two initial starting structures (atomic r.m.s. differences of 2·7 and 2·1 Å with and without the tyrosine ring, respectively) have converged to essentially the same final structures (atomic r.m.s. differences of 0·6 and 0·1 Å with and without the tyrosine ring, respectively). Thus, the only portion of the structure that is relatively ill-defined is the exact location of the tyrosine ring. Convergence is achieved both through an improvement in the restraints energy and in the van der Waals energy.

In considering the converged restrained dynamics structures, it is important to bear in mind two points: (1) the interproton distances used for the refinement represent $(\langle r^{-6} \rangle)^{-1/6}$ means rather than arithmetic means, so that their values are heavily weighted towards fluctuations with the shortest interproton distances; and (2) the interproton distances themselves are highly correlated (this is easily ascertained by simple model building). Consequently, if the TRNOE effects arise from the presence of several significantly different conformations, a single structure would not be able to provide an acceptable fit to the experimental data. In the present case, all the interproton distance restraints are satisfied within the errors specified, and the r.m.s. difference between the calculated and target interproton distances is $\lesssim 0{\cdot}3$ Å. We therefore conclude that the bound tetrapeptide occupies either a smooth global minimum energy region (i.e. a single conformation) or a global minimum energy region containing many local minima (i.e. multiple but very similar conformations). If the latter is indeed the case, as seems likely from the crystallographic data (see below), then the differences between the two average restrained dynamics structures provide a measure of the region of conformational space that the bound tetrapeptide can sample. This region of space is relatively small

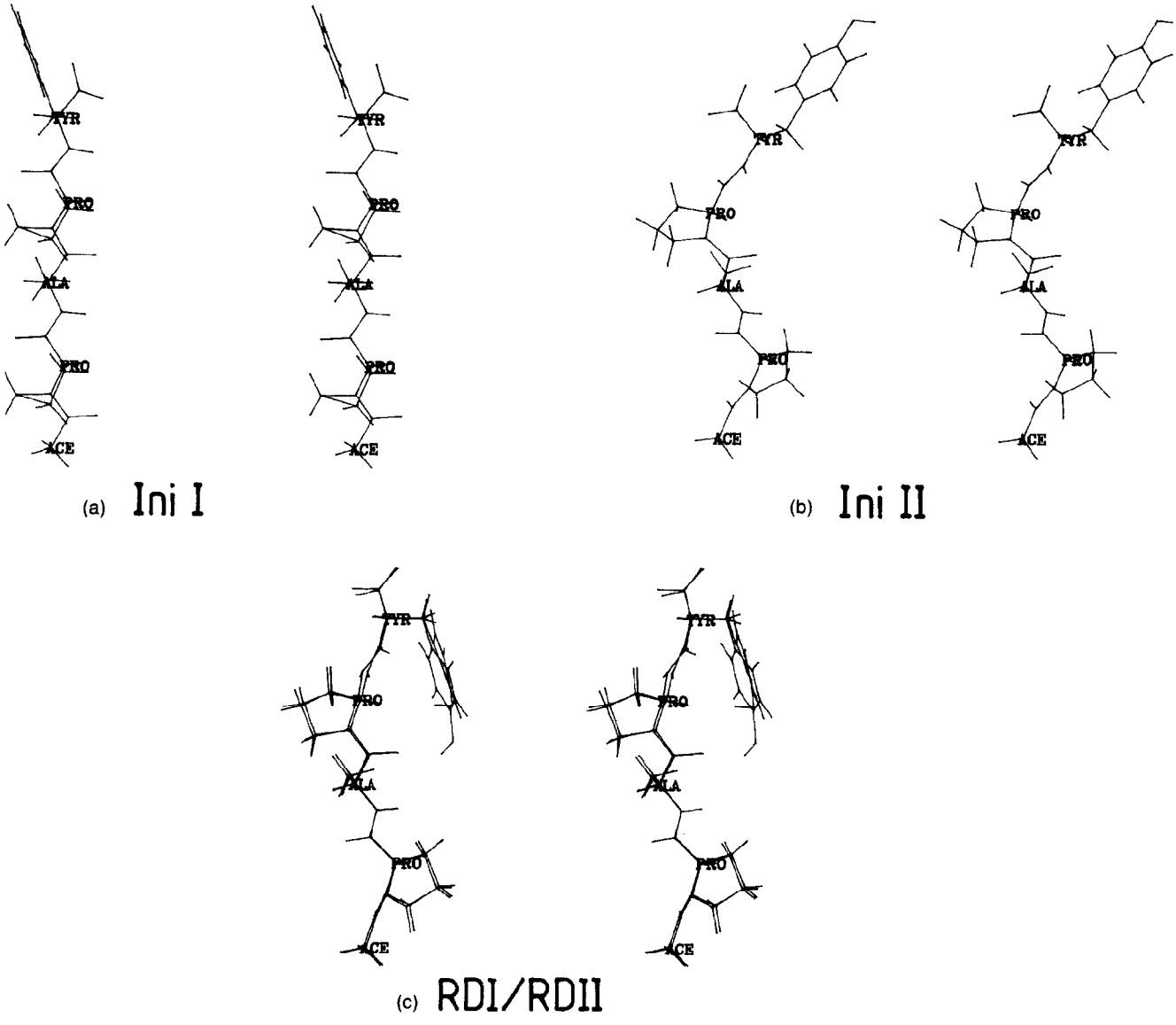


**Figure 2.** Stereo views of (a) the extended $\beta$-strand initial structure Ini I, (b) the polyproline helix initial structure Ini II, and (c) the best fit superposition of the 2 final average restrained dynamics structures RDI and RDII.

and well-defined by the NOE distances (see Tables 4 and 5, and Fig. 2). The relatively high precision with which this region can be defined is due to the fact that the NOE distances used, although necessarily imprecise, are correlated.

In order to gain further insight into the interaction of the tetrapeptide with elastase, we then proceeded to use interactive molecular graphics to model the inhibitor into the active site of the enzyme. This essentially involves a two-step procedure: (1) the determination of a suitable anchor point(s); and (2) the subsequent manipulation of the inhibitor molecule to maximize favourable hydrogen bonding, electrostatic and hydrophobic interactions.

Determination of an anchor point was facilitated by a simultaneous attempt to solve the crystal structure of this complex. The only residual electron density in the Fourier difference map above the $1\sigma$ confidence level was found in the S′ "leaving group" location of elastase. Indeed, there is only density for a dipeptide, but it is sufficiently good at 1·8 Å resolution to clearly identify a proline residue as well as the tetrahedral geometry of its C$\alpha$ atom, thereby providing an unambiguous assignment of the binding direction. The second visible residue of the ligand is C-terminal to this proline residue, its C$\alpha$ atom is also clearly tetrahedral and the electron density for its C$\beta$ atom is equally good. No electron density attributable to a tyrosine ring, however, could be seen. Thus, at this stage, assignment of the visible electron density to the Pro3-Tyr4 moiety of the tetrapeptide rather than to the Pro1-Ala2 moiety could not be made. However, comparison of the tetrapeptide Fourier difference map with that of the product complex of

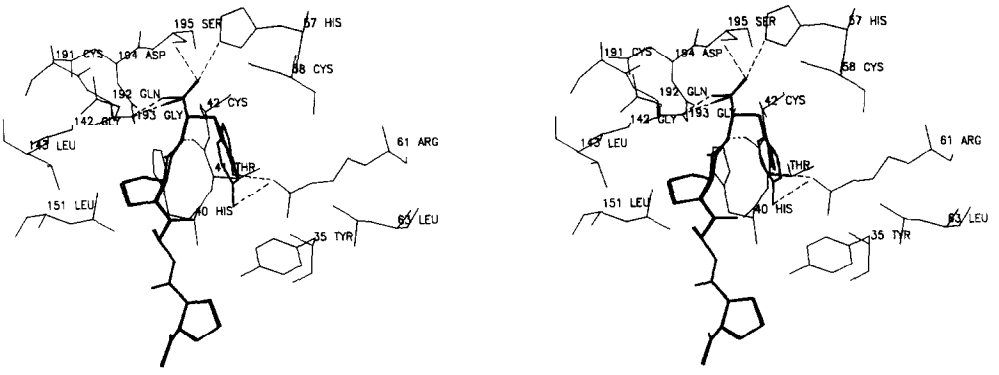


**Figure 3.** Stereo view of the proposed binding of acetyl-Pro-Ala-Pro-Tyr in the S′ "leaving group" location of elastase. Hydrogen bonds are indicated by broken lines, the 2 docked average restrained dynamics structures for the bound tetrapeptide are represented by thick lines, and the protein residues are shown as thin lines. Model building was carried out initially on a Vector General 3DM display using the programs FIT and PIX (Morimoto & Meyer, 1976) and then completed on an Evans & Sutherland PS 330 Colour Graphics System interfaced to a VAX 11/780 computer using the interactive molecular graphics program FRODO (Jones, 1978; Pflugrath *et al.*, 1984). The co-ordinates of elastase were taken from the 1·65 Å resolution crystal structure of the elastase–acetyl-Ala-Pro-Ala complex (Meyer *et al.*, 1986). The average restrained molecular dynamics structures RDI and RDII of the bound tetrapeptide were then modelled into the extended binding site of elastase as described in the text. During the modelling procedure, van der Waals contacts were assessed by means of a dot surface of the binding site at twice the van der Waals radius of the constituent atoms, such that optimal van der Waals' contact occurs when the atoms of the ligand touch this surface. In the model, the Arg61 side-chain is so positioned that its guanidinium group can make a hydrogen bond with the OH group of Tyr4. It should be noted, however, that this particular hydrogen bond is only a possibility as the C$\beta$ and C$\gamma$ side-chain atoms have very high thermal factors ($B \sim 45$ Å$^2$), and the other atoms of the side-chain are not visible in the electron density map (Meyer *et al.*, unpublished results).

As discussed in the text, determination of the initial anchor point for the tetrapeptide was aided by a simultaneous crystallographic study. A large crystal of elastase (0·5 mm × 0·5 mm × 0·9 mm in cross-section), grown as described by Meyer *et al.* (1986) was soaked for 2 days in the presence of excess tetrapeptide (0·1 M-sodium acetate, 0·5 M-ammonium sulphate, pH 5·0) and then mounted along the *c* axis in a thin-walled capillary with a minimum meniscus of buffer but in physical contact with excess ligand. (Space group $P2_12_12_1$; cell dimensions $a = 51{\cdot}4$ Å, $b = 58{\cdot}2$ Å, $c = 75{\cdot}4$ Å.) Data collection and film evaluation were as described (Meyer *et al.*, 1986). A total of 38,594 reflections were measured to be above the 1$\sigma$ confidence level; 17,206 unique reflections to 1·74 Å gave a completeness ratio of 70% (the 1·78 to 1·74 Å shell had a completeness ratio of 24%). The internal agreement (overall $R_{\text{MERGE}}$) of the data was 8·42%. For the calculation of the phases and the difference Fourier map, the 1·65 Å resolution structure of the elastase–acetyl-Ala-Pro-Ala complex (Meyer *et al.*, 1986) was stripped of all ligand and water molecules in the extended binding site and used as the native structure. No ligand density was found in the S and especially S1 sites of elastase; clear density in the S1′–S2′ region, however, was seen and used as the basis for model building (see the text).

the tripeptide acetyl-Ala-Pro-Ala (Meyer *et al.*, 1986) revealed that the observable electron density for the tetrapeptide was directly superimposable on the Pro-Ala electron density for the tripeptide bound at the APA2 site in the vicinity of Tyr35, Thr41, His57, Gly193 and Ser195. Moreover, with the visible tetrapeptide electron density assigned to the Pro3-Tyr4 moiety, comparison of the resulting "docked" tetrapeptide with the APA2 tripeptide showed that the two inhibitor molecules were also more or less superimposable.

Apparent identity between the binding of these two complexes is significant, because the acetyl-Ala-Pro-Ala complex shows an additional inhibitor molecule (APA1) bound on the other side of the primary specificity pocket in the vicinity of His57, Ser214 and Val216 (Meyer *et al.*, 1986) and, hence, suggests that this site too should be investigated for a possible binding mode. A look at this alternative site, however, quickly eliminates it as a possibility because the corresponding binding of the tetrapeptide in sites S4 to S1 (notation of Schecter & Berger, 1967) is precluded by the S1 site aversion to large, branched C$\beta$ side-chains. (Note that in the crystal structure the S4 subsite is also blocked by the Glu62 side-chain of a symmetrically related lattice neighbour; Meyer *et al.*, 1986). The S2–S2′ binding mode is also excluded primarily on the basis of an unfavourable fit of proline in the S1′ subsite. This is due both to the size of the proline residue and to the inflexibility of the pyrrolidine ring preventing the P1′–P2′ residues from adapting to the S1′–S2′ site geometries. That is to say, it is not possible to bend the molecule to fit the constraints of this binding site.

Subsequent adjustment of the acetyl-Pro-Ala Pro-Tyr molecule in the chosen site yielded a model (see Fig. 3) exhibiting a polarity of binding opposite to that observed in several serine protease–inhibitor and product complexes (Ruehlman *et al.*, 1973; James *et al.*, 1980; Bolognesi *et al.*, 1982; Chen & Bode, 1983), although identical to that observed in the acetyl-Ala-Pro-Ala complex (Meyer *et al.*, 1986). Additionally, this binding mode is non-productive, because the scissile bond is too far removed from the catalytic residues. The non-

productivity of the complex is confirmed by the n.m.r. results, which indicate lack of cleavage over a period of several days.

The model of the bound tetrapeptide exhibits the following features (Fig. 3). The acetyl-Pro1 terminus is pointed towards the solvent with possible hydrogen bonding between the acetyl group and solvent molecules. The Pro3 pyrrolidine ring is positioned between the backbone atoms of residues His40, Gln192 and Gly193, and the side-chain atoms of Leu143 and Leu151. A hydrogen-bond network is set up at the C terminus involving the carboxyamide group of Tyr4: the acid amide group is hydrogen-bonded to the His57 Nε and the Ser195 Oγ atoms, and the carboxyl oxygen atom is hydrogen-bonded to the backbone nitrogen atom of Gly193. Additionally, the backbone nitrogen atom of Tyr4 is hydrogen-bonded to the carboxyl oxygen atom of Thr41. Two further hydrogen bonds may be present. One is between the Tyr4 OH group and the guanidinium group of Arg61. This interaction is likely to be weak, as neither the tyrosine ring nor the guanidinium group of Arg61 are visible in the electron density map, and the average *B* factor for the visible portion of the Arg61 side-chain is very large (with a value of $\sim 45$ Å$^2$; Meyer *et al.*, unpublished results). The other becomes part of the C terminus network if a water molecule present in the 1·65 Å resolution crystal structure of elastase (Meyer *et al.*, unpublished results) adjacent to His57 and Ser195 is hypothesized to also exist in solution. It is important to stress that this model should not be considered as a rigid one. Indeed, the crystallographic observation that only Pro3 and the Tyr4 backbone are visible in the Fourier difference map suggests that the acetyl-Pro1-Ala2 moiety and Tyr4 ring are sufficiently mobile to be statistically disordered in the crystal.

This result is contrary to our initial expectation of a single productive binding mode with the Pro1, Ala2, Pro3 and Tyr4 ligand residues located in the S2, S1, S1′ and S2′ sites, respectively. This incorrect prediction was based on kinetic and binding data showing that elastase has a preference for proline in the P2 ligand position and for small hydrophobic side-chains (such as leucine or alanine) in the P1 position (Atlas & Berger, 1972; Atlas, 1975). Thus, porcine pancreatic elastase continues to be unpredictable in its binding of small ligands. This can be attributed directly both to weak stabilizing interactions in the primary S1 specificity site and to rather weak interactions provided by enzyme side-chains in other subsites.

From the point of view of drug design methodologies, the present study emphasizes that there is no theoretical substitute for experiment. In this respect, it is hoped that the approach presented in this paper, in which complementary but incomplete data from n.m.r. studies (namely, the conformation of the bound ligand) and X-ray crystallography (namely, the approximate location and orientation of the bound ligand) are combined into a complete whole for subsequent use in molecular modelling studies, should prove extremely powerful in the future.

This work was supported by the Max-Planck Gesellschaft (G.M.C. and A.M.G.), the Robert A. Welch Foundation, the Office of Naval Research and the Petroleum Research Fund as administered by the American Chemical Society (G.C. and E.F.M.). We thank the Max-Planck-Institut fur Plasma Physik (Garching) for computer facilities on the CRAY 1, Drs J. Pflugrath and A. T. Brunger for useful discussions, and Professor J. C. Powers for the gift of the tetrapeptide.

**G. Marius Clore**
**Angela M. Gronenborn**

Max-Planck-Institut für Biochemie
D–8033 Martinsried bei Munchen, F.R.G

**Gail Carlson**
**Edgar F. Meyer**

Department of Biochemistry and Biophysics
Texas A & M University
College Station, Tex. 77843, U.S.A.

Received 29 January 1986, and in revised form
20 March 1986

## References

Atlas, D. (1975). *J. Mol. Biol.* **93**, 39–53.
Atlas, D. & Berger, A. (1972). *Biochemistry*, **11**, 4719–4723.
Bax, A. & Davis, D. G. (1985). *J. Magn. Reson.* **65**, 355–360.
Bizzozero, S. A. & Dutler, H. (1981). *Bioorg. Chem.* **10**, 46–62.
Bodenhausen, G., Vold, R. L. & Void, R. R. (1980). *J. Magn. Reson.* **37**, 93–106.
Bolognesi, M., Gatti, G., Menegatti, E., Guarneri, M., Marquart, M., Papamokos, E. & Huber, R. (1982). *J. Mol. Biol.* **162**, 839–868.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swamanithan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.
Chen, Z. & Bode, W. (1983). *J. Mol. Biol.* **64**, 283–311.
Clore, G. M. & Gronenborn, A. M. (1983). *J. Magn. Reson.* **53**, 423–442.
Clore, G. M. & Gronenborn, A. M. (1984). *FEBS Letters*, **172**, 219–225.
Clore, G. M., Gronenborn, A. M. & McLaughlin, L. W. (1984). *J. Mol. Biol.* **174**, 163–173.
Clore, G. M., Gronenborn, A. M., Brunger, A. T. & Karplus, M. (1985). *J. Mol. Biol.* **186**, 435–455.
Clore, G. M., Gronenborn, A. M., Greipel, J. & Maass, G. (1986). *J. Mol. Biol.* **187**, 119–124.
Davis, D. G. & Bax, A. (1985). *J. Amer. Chem. Soc.* **107**, 2821–2822.
Dimicoli, J.-L., Renaud, A. & Bieth, J. (1980). *Eur. J. Biochem.* **107**, 423–432.
Gronenborn, A. M. & Clore, G. M. (1982*a*). *J. Mol. Biol.* **157**, 155–160.
Gronenborn, A. M. & Clore, G. M. (1982*b*). *Biochemistry*, **21**, 4040–4048.
Gronenborn, A. M., Clore, G. M. & Jeffrey, J. (1984*a*). *J. Mol. Biol.* **172**, 559–572.
Gronenborn, A. M., Clore, G. M., Brunori, M., Giardina, B., Falcioni, G. & Perutz, M. F. (1984*b*). *J. Mol. Biol.* **178**, 731–742.

Harper, J. W. & Powers, J. C. (1985). *Biochemistry*, **24**, 7200–7213.

Harper, J. W., Hemmi, K. & Powers, J. C. (1985). *Biochemistry*, **24**, 1831–1841.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974). *J. Mol. Biol.* **89**, 73–101.

Hughes, D. L., Sieker, L. C., Bieth, J. & Dimicoli, J. L. (1982). *J. Mol. Biol.* **162**, 645–658.

James, M. N. G., Siebeki, A. R., Brayer, G. D. & Delbaere, L. T. J. (1980). *J. Mol. Biol.* **144**, 43–89.

Janoff, A. (1983). In *Molecular Basis of Degradative Processes* (Berlin, R. D., Herrman, H., Lepow, I. H. & Tanzer, J. M., eds), pp. 225–260, Academic Press, New York.

Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). *J. Chem. Phys.* **71**, 4546–4553.

Jones, T. A. (1978). *J. Appl. Crystallogr.* **11**, 268–272.

Karplus, M. & McCammon, J. (1983). *Annu. Rev. Biochem.* **52**, 263–300.

Lungarella, G., Gardi, C., Desanti, M. M. & Luzi, P. (1981). *Exp. Mol. Pathol.* **42**, 44–59.

Macura, S., Huang, Y., Suter, D. & Ernst, R. R. (1981). *J. Magn. Reson.* **43**, 259–281.

Marion, D. & Wuthrich, K. (1983). *Biochem. Biophys. Res. Commun.* **113**, 967–974.

McCammon, J. A., Gelin, B. R. & Karplus, M. (1974). *Nature (London)*, **267**, 585–590.

McCammon, J. A., Wolynes, P. G. & Karplus, M. (1979). *Biochemistry*, **18**, 927–942.

McRae, B., Nakajima, K., Travis, J. & Powers, J. C. (1980). *Biochemistry*, **19**, 3973–3978.

Meyer, E. F., Presta, L. G. & Radhakrishnan, R. (1985). *J. Amer. Chem. Soc.* **107**, 4091–4093.

Meyer, E. F., Radhakrishnan, R., Cole, G. M. & Presta, L. G. (1986). *J. Mol. Biol.* **189**, 533–540.

Morimoto, C. N. & Meyer, E. F. (1976). In *Crystallographic Computing Techniques* (Ahmed, F. R., ed.), pp. 488–496. Munksgaard Publishing Co., Copenhagen.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T. & Karplus, M. (1986). *J. Mol. Biol.* **188**, 455–476.

Pflugrath, J. W., Saper, M. A. & Quiocho, F. A. (1984). In *Methods and Applications in Crystallographic Computing* (Hall, S. & Ashiaki, T., eds), pp. 407–420, Clarendon Press, Oxford.

Powers, J. C. (1976). *Trends Biochem. Sci.* **1**, 211–214.

Redfield, A. G. & Kuntz, S. D. (1975). *J. Magn. Reson.* **19**, 250–254.

Ruehlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. (1973). *J. Mol. Biol.* **77**, 417–436.

Schecter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.

Shotton, D. M., White, N. J. & Watson, H. C. (1972). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 91–105.

Solomon, I. (1955). *Phys. Rev.* **90**, 559–565.

Wider, G., Macura, S., Kumar, A., Ernst, R. R. & Wuthrich, K. (1984). *J. Magn. Reson.* **56**, 207–234.

Williamson, M. P., Havel, T. F. & Wuthrich, K. (1985). *J. Mol. Biol.* **182**, 295–315.

*Edited by M. F. Moody*