PROTEINS: Structure, Function, and Genetics 37:628–640 (1999)

Locating Interaction Sites on Proteins: The Crystal Structure of Thermolysin Soaked in 2% to 100% Isopropanol

Andrew C. English,[1] Sarah H. Done,[1] Leo S.D. Caves,[1] Colin R. Groom,[2] and Roderick E. Hubbard[1]*
[1]*Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, United Kingdom*
[2]*Pfizer Central Research, Sandwich, Kent, United Kingdom*

ABSTRACT **Multiple-solvent crystal structure determination (MSCS) allows the position and orientation of bound solvent fragments to be identified by determining the structure of protein crystals soaked in organic solvents. We have extended this technique by the determination of high-resolution crystal structures of thermolysin (TLN), generated from crystals soaked in 2% to 100% isopropanol. The procedure causes only minor changes to the conformation of the protein, and an increasing number of isopropanol interaction sites could be identified as the solvent concentration is increased. Isopropanol occupies all four of the main subsites in the active site, although this was only observed at very high concentrations of isopropanol for three of the four subsites. Analysis of the isopropanol positions shows little correlation with interaction energy computed using a molecular mechanics force field, but the experimentally determined positions of isopropanol are consistent with the structures of known protein-ligand complexes of TLN. Proteins 1999;37:628–640.** © 1999 Wiley-Liss, Inc.

Key words: organic solvent; binding sites; drug design; X-ray crystallography; inhibitors

INTRODUCTION

Structural information is becoming available for an increasing number of protein molecules that are potential targets for therapeutic agents. A key requirement for the optimal use of these X-ray and NMR structures in structure-assisted drug design are tools for the prediction of the position, orientation and binding affinity of functional groups in an active site.[1] While there are computational approaches to this problem, new chemical leads are still generally discovered by screening large numbers of compounds. As design and synthesis of focused libraries of compounds using combinatorial chemistry techniques has the potential to greatly increase the chance of finding new lead compounds, the use of protein structural information in the design of these libraries is an area of considerable interest.[2]

Two experimental approaches have emerged recently to allow the direct incorporation of structural information in inhibitor design. The group of Fesik[3,4] has pioneered a technique known as SAR by NMR (structure activity relationships by nuclear magnetic resonance) in which NMR of a ^{15}N-labeled protein can screen a library of compounds. Changes in selected amide chemical shifts of the protein during a ligand titration experiment can identify the positions of ligands in binding sites in close proximity to each other. Once structural information for the optimized ligands has been derived they can be linked together. This technique can reduce the number of compounds to be synthesized and has demonstrated success where more conventional techniques of drug discovery have failed.

Multiple-solvent crystal structure determination (MSCS)[5,6] involves soaking protein crystals in organic solvents and locating the positions and orientations of any bound molecules using X-ray crystallography. By employing a variety of organic solvents it is possible to build up an experimental functionality map of the protein binding site, in a similar way to theoretical techniques such as multiple copy-simultaneous searching (MCSS).[7] Therefore, MSCS could have applications in structure-assisted inhibitor design, where organic solvents could be used to direct the design of new ligands against pharmaceutically relevant proteins. In addition, if the solvent molecules chosen are representative of the molecules used in constructing combinatorial libraries, then MSCS may aid the design of more focused libraries. The experimental results from MSCS could also be used to improve computational techniques, for example by comparison with MCSS or by the incorporation of empirical data into the MCSS program.

In this paper, we investigate whether the MSCS method can provide a semi-quantitative measure of binding by studying the effect of solvent concentration on the number of binding sites identified. There are a number of considerations in choosing a system to develop MSCS. The method is applicable only if the protein crystals are able to withstand soaking in organic solvents, and in the past, gluteraldehyde has been used to cross-link [5,8] the crystals. A key criterium is that the soaked crystals must diffract to a sufficient resolution such that any bound solvent molecules can be distinguished from water molecules in the electron density maps.

Thermolysin (TLN) was selected as the initial target protein.[9] The protein is a bacterial zinc endopeptidase of

Grant sponsor: Biotechnology and Biological Sciences Research Council.
Sarah H. Done's present address is Pfizer Central Research, Sandwich, Kent, UK, CT13 9NJ
*Correspondence to: Roderick E Hubbard, Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, United Kingdom, YO10 5DD. E-mail: rod@yorvic.york.ac.uk
Received 27 May 1999; Accepted 23 July 1999

TABLE I. Data Collection Statistics

	% Isopropanol v/v								
	Native	2	5	10	25	60	80	90	100
Resolution[a] (Å)									
High	1.65	2.05	1.95	1.95	1.95	2.10	2.10	1.95	2.20
Low	20.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0
Cell dimensions (Å)									
a = b	94.15	93.93	93.99	93.99	94.14	94.23	94.12	93.27	93.04
c	131.62	130.91	130.87	130.88	130.80	131.03	130.68	130.36	130.42
Number of unique									
Reflections	41876	20770	23955	24568	24598	20404	19860	23150	16875
Completeness (%)									
Overall	100	95	95	97	97	99	97	93	97
Highest resolution shell	100	93	92	95	95	98	95	91	95
R_{merge}[b] (%)									
Overall	7.6	11.2	7.8	8.9	9.3	12.6	10.8	7.3	10.8
Highest resolution	35.2	34.6	23.6	26.8	37.2	40.4	36.8	26.0	36.3
Shell									
Overall multiplicity	5.0	7.2	6.6	5.7	6.5	7.8	4.5	5.8	5.8
I/σI									
Overall	6.2	5.5	6.4	5.9	5.9	5.2	6.0	6.2	5.4
Highest resolution shell	1.8	2.2	3.2	2.8	2.0	1.9	2.1	2.9	2.1

[a]Resolution was determined by cutting off the data when I/σI < 2.0, or when the completeness of the highest resolution bin was <80%.
[b]$R_{merge} = \Sigma_{hkl}|I_i - I_n|/\Sigma_{hkl}I_n$ where I_i is an observed intensity *hkl* and I_n is the average of the observed equivalents.

34.5 kDa, with one half of the enzyme folded mainly into α-helices and the other half mainly β-sheets. TLN has a large, rigid active site cleft consisting of at least four subsites (S_2, S_1, S_1^1, S_2^1),[10] with the main specificity pocket at the S_1^1 subsite which is known to prefer hydrophobic groups.

The native crystalline enzyme has a Val-Lys dipeptide[11] bound in the active site interacting with the S_1^1 and S_2^1 subsites. It is thought that during crystallization, one molecule of TLN cleaves off the Val-Lys dipeptide from the C-terminus of another molecule. Since the Val-Lys C-terminus of TLN is found at full occupancy it is thought that molecules lacking these last two residues are not incorporated into the crystal lattice. Isopropanol was chosen as the initial probe molecule as it is a water miscible organic solvent which is a mimic for both the hydroxyl and hydrophobic functionality in larger ligands and has a size and shape that would allow it to be distinguished from water molecules in electron density maps.

MATERIALS AND METHODS

Crystallization of TLN

TLN was purchased from Calbiochem (San Diego, CA) and crystals grown using a modified method based on that of Holmes and Matthews.[9] A stock solution (4.6 mM) of TLN was prepared in 45% DMSO, 50 mM Tris-HCl, 2.5 M CsCl pH 7.5. The solution was centrifuged and 10 µl of the resultant supernatant was placed in a microbridge in the well of a Linbro tray. Five hundred (500) µl of water was added to the well, and clumps of hexagonal rods grew to full size in 5–6 days (0.8 × 0.3 × 0.3mm).

Soaking of TLN Crystals in Isopropanol

Crystals were transferred into a buffer containing 7% DMSO, 3.2 mM calcium acetate, 10 mM Tris chloride, pH 7.5. The crystals were then placed into a 25 µl dialysis button and sealed. These were transferred to 5 ml of a solution of isopropanol, at the required percentage (v/v) and left for 18–24 h to soak. Single crystals were mounted in 0.7 mm glass capillaries for X-ray data collection.

Data Collection and Refinement

X-ray data for the native enzyme were collected to a resolution of 1.65 Å at room temperature from a single crystal at Daresbury SRS (station PX9.6), using a CCD detector. The data were integrated using MOSFLM,[12] scaled and processed using CCP4[13] software and the structures refined using the maximum likelihood program REFMAC.[14] The initial phases were calculated using the deposited structure of native TLN (1lnf)[11] with the water molecules, active-site dipeptide and side chains of Met 120, Glu 143, Leu 144, and Tyr 157 removed. During refinement, water molecules were added manually in the X-SOLVATE module of QUANTA[15] using maximum likelihood weighted mF_o-DF_c maps contoured at 3.0σ. A solvent continuum was included in the model as implemented within REFMAC. During subsequent refinement water molecules with a temperature factor above 75 Å^2 were removed from the model unless electron density persistently returned on removal of the water. Manual rebuilding of protein side chains was carried out in the X-AUTOFIT module of QUANTA[15] using maximum likelihood weighted mF_o-DF_c maps contoured at 3σ. The occupancies of disordered residues were carefully weighted until there was no residual positive or negative difference density in the mF_o-DF_c map on either conformer. Data and refinement statistics can be found in Tables I and II, respectively. This structure is referred to as TLN_0.

TABLE II. Refinement Statistics

	% Isopropanol v/v								
	Native	2	5	10	25	60	80	90	100
Number of water molecules	207	149	153	149	140	135	121	120	99
Number of isopropanol molecules	—	0	1	2	2	2	3	8	11
R_{factor}[a] (%)									
Initial	28.8	22.5	23.3	23.3	25.7	27.3	29.4	49.8	51.3
Final	16.3	16.5	16.6	16.6	16.4	15.7	16.8	16.6	17.1
Free R_{factor}[b] (%)									
Initial	28.2	22.7	22.7	22.7	25.6	27.1	29.1	47.8	48.9
Final	18.8	21.9	21.2	21.3	20.7	19.3	22.2	21.3	22.3
RMS differences									
Bond lengths (Å)[c]	0.016	0.015	0.014	0.015	0.014	0.014	0.015	0.014	0.016
Bond angles (°)	0.8	1.9	1.6	1.6	1.7	2.1	2.0	1.8	2.4
Average B factors (Å^2)									
Protein	17.6	22.9	26.0	24.5	27.3	23.7	24.9	26.0	31.7
Water	37.1	35.9	39.3	37.5	39.7	38.0	37.7	41.1	45.8
Isopropanol	—	—	60.0	49.2	48.0	42.8	44.7	48.0	57.0

[a]R_{factor} (95% of reflections) = $\Sigma_{hkl}||F_o| - |F_c||/|F_o|$, where $|F_o|$ and $|F_c|$ are the magnitudes of the observed and calculated structure factor amplitudes of a reflection *hkl*, respectively.
[b]Free R_{factor} (5% of reflections) = $\Sigma_{hkl}||F_o| - |F_c||/|F_o|$, for 5% of all reflections omitted from the refinement process.
[c]Largest value set to (0.020).

X-ray diffraction data for crystals soaked in isopropanol were collected from single crystals on a 30-cm MAR image plate mounted on a Rigaku RU200 rotating anode generator. The data were integrated using DENZO,[16] scaled and processed using CCP4 software and the structures refined using REFMAC. The isopropanol-soaked structures were solved prior to collection of the native data set at 1.65 Å. As such, the starting model was an earlier structure of native TLN (1ltx) using data, which was collected, in this laboratory (2.1 Å resolution). The initial phases were calculated after removal of the water molecules, active-site dipeptide and side chains of Met 120, Glu 143, Leu 144, and Tyr 157. Despite a general shrinkage in all cell dimensions accompanying increasing concentration of isopropanol (see Table I), all structures could be solved using straightforward difference Fourier techniques. Including weak data at high resolution is valuable for refinement, particularly with maximum likelihood techniques, and the relatively high values of R_{merge} reflect this. The resulting structures are hereafter referred to as TLN_X, where X is the percentage of isopropanol in the crystal soaking experiments. The protocols used in fitting isopropanol molecules to electron density and the procedures adopted during refinement are described in detail below.

Fourier Maps

A range of different Fourier maps (F_o-F_o, F_o, $\rho_o-\rho_o$, F_o-F_c and $2F_o-F_c$) were calculated to identify, model, and validate isopropanol molecules bound to TLN in the crystal structures. When such an analysis is undertaken, it is important that no model bias is introduced into the results. Therefore, all maps phased with the isopropanol-soaked structures were calculated after omitting all isopropanol molecules, and water molecules within 4 Å of each isopropanol from the model, and a further two rounds of refinement performed in their absence. Similarly, all maps phased with the refined native model were calculated after removal of the dipeptide, and water molecules within 4 Å of isopropanol positions taken from the soaked structures, and a further two rounds of refinement performed in their absence.

F_o maps were calculated with coefficients $m|F_{o(ipa,x)}|\exp -i\phi_{(ipa,x)}$, where $m|F_{o\,(ipa,x)}|$ are the maximum likelihood weighted observed amplitudes of the isopropanol soaked protein (at percentage x). The phases, $\phi_{(ipa,x)}$ were calculated from the refined isopropanol soaked model, after deletion of the atoms described above.

$F_{o(ipa,x)}-F_{o(nat)}$ maps were calculated with coefficients $(m|F_{o(ipa,x)}|-m|F_{o(nat)}|)\exp -i\phi_{(nat)}$, where $m|F_{o(ipa,x)}|$ and $m|F_{o(nat)}|$ are the maximum likelihood weighted observed amplitudes of isopropanol soaked (at percentage x) and native TLN, respectively. The phases, $\phi_{(nat)}$ were calculated from the refined higher resolution native model (TLN_0, 1.65 Å), after deletion of the atoms described above. To calculate an F_o-F_o map it is essential that the two protein structures are isomorphous.

At concentrations of 80% isopropanol and above, the crystals became less isomorphous with the native crystal, and calculation of conventional $F_{o(ipa,x)}-F_{o\,(nat)}$ maps was not possible. Instead, $\rho_{o(ipa,x)}-\rho_{o\,(nat)}$ maps were constructed, which is the difference between two F_o maps in real space. To construct the map, the $\rho_{o(ipa,x)}$ map is transposed onto the $\rho_{o(nat)}$ map by using a transposition derived by overlapping the main chain atoms of the independently refined models. The $\rho_{o(nat)}$ map is then subtracted from the $\rho_{o(ipa,x)}$ map in real space.

Protocol to Assign Solvent Structure

The $F_{o(ipa,x)}-F_{o(nat)}$ and $F_{o(ipa,x)}$ electron density maps were used to distinguish between isopropanol molecules and water. $F_{o(ipa,x)}-F_{o(nat)}$ difference maps arguably provide the best way to validate the presence of molecules of

bound solvent. They are highly sensitive in the region of interest and are not subject to phase bias. In addition, $F_{o(ipa,x)}$ maps are highly diagnostic and have a number of advantages. F_o maps do not have problems associated with scaling and usually have lower noise levels than other Fourier maps. Fragments omitted from the model return theoretically at half height in the map. $\rho_{o\ (ipa,x)} - \rho_{o\ (nat)}$ maps were also constructed although since the calculation involves substracting two maps on different crystal lattices these maps tend to be more susceptible to noise.

Isopropanol has the formula $CH_3CH(OH)CH_3$ and the four heavy atoms total 26 electrons, whereas the single heavy atom of water (oxygen) has eight electrons. Therefore, in general, isopropanol will give rise to stronger peaks in the electron density maps, with a distinctive shape. While a comparison of shape is best achieved at lower contour levels, maps were also contoured at high levels to assess whether a peak is significant. Higher contour levels also help distinguish isopropanol from disordered water, as peaks corresponding to regions of disordered water tend to fragment. Where isopropanol is displacing the dipeptide in the active-site region, positions of bound molecules cannot be identified in the $F_{o(ipa,x)} - F_{o(nat)}$ and $\rho_{o(ipa,x)} - \rho_{o\ (nat)}$ maps. Instead, $F_o - F_c$ maps were used (in addition to F_o maps) to identify the bound isopropanol molecules.

Model Building and Refinement

Once positions of isopropanols were identified, molecules were built into the model using $mF_{o(ipa,x)} - DF_c$ difference electron density maps within the X-LIGAND module of QUANTA.[15] At a resolution of 1.95–2.2 Å the oxygen and carbon atoms of isopropanol are not readily distinguished in the various electron density maps. Therefore, orientation of each molecule was aided by optimizing the hydrogen bonding between the hydroxyl group of isopropanol and the surrounding protein or water molecules, and the hydrophobic interactions between the isopropyl group and nearby hydrophobic residues. The occupancies of isopropanol molecules included in the model were set to 1.0, and no attempt was made to refine them. The B factors were freely refined as implemented within REFMAC. In general, the B factors for molecules of isopropanol located at lower soak concentrations (IPA 1 and 2) decrease as the concentration of isopropanol increases, which would indicate a gradual increase in the occupancy at these sites. A number of the isopropanol binding sites (especially those located only at high concentrations) display several possible binding modes. At such mixed population sites the occupancy of a particular binding mode will be less than 1.0, and the B factors for these sites will be less accurate.

Assignment of Electron Density to Isopropanol in Thermolysin

In TLN_2, the $mF_{o(ipa,2)} - mF_{o(nat)}$ difference map confirms that the dipeptide in the active site has been displaced, with the region enclosed in negative difference density (Fig. 1A). However, there are no molecules of isopropanol visible in the $mF_{o(ipa,\ 2)} - DF_c$ or $2mF_{o(ipa,\ 2)} - DF_c$ maps, the latter of which is shown in Figure 1B. In TLN_5, the region of the dipeptide remains enclosed in negative difference density in the $mF_{o\ (ipa,\ 5)} - mF_{o\ (nat)}$ map (not shown), and a single molecule of isopropanol (IPA 1) is visible in the $S_1{}^1$ subsite as the largest volume peak in the $mF_{o\ (ipa,\ 5)} - DF_c$ omit map (Fig. 1C). As IPA 1 is binding in the region that the Val side chain of the dipeptide occupies in the TLN_0 structure, the isopropanol is not visible in the $mF_{o\ (ipa,\ 5)} - mF_{o(nat)}$ map.

In TLN_10, a second molecule of isopropanol (IPA 2) is visible in the $2mF_{\ o(ipa,\ 10)} - DF_c$ and $mF_{o(ipa,\ 10)} - mF_{o(nat)}$ density maps (Fig. 2B and C). It is not until 80% isopropanol that a third binding site is identified along the active site-cleft, and it is fully enclosed within the 3σ level in the $mF_{o\ (ipa,\ 80)} - DF_c$ omit map.

In TLN_90, an additional five molecules of isopropanol are identified (IPA 4, 5, 6, 7 and 8), and the subsites where they bind have been grouped together in their ability to bind isopropanol (Fig. 3). IPA 4–8 are fully enclosed within positive difference density in the $mF_{o(ipa,\ 90)} - DF_c$ omit map (contoured in the range 3.0–3.5σ). In addition, IPA 4–8 are fully enclosed within electron density in the $mF_{o\ (ipa,\ 90)}$ map (contoured in the range 0.3–0.5σ). Although this might be considered a low contouring, the noise level for the $mF_{o(ipa,\ 90)}$ map was estimated to be 0.05σ so these peaks are significant. A typical F_o map is shown in Figure 3.

In TLN_100, a further four molecules of isopropanol are identified (IPA 9,10,11 and 12), and we suggest these represent the lower affinity sites for TLN. IPA 9–12 are fully enclosed within positive difference density in the $mF_{o(ipa,\ 100)} - DF_c$ omit map (contoured in the range 2.5–3.0σ). In addition, IPA 9–12 are fully enclosed within electron density in the $mF_{o(ipa,100)}$ map (contoured in the range 0.3–0.4σ). IPA 7 (first identified in TLN_90) is not clearly visible in the electron density maps for TLN_100. The $mF_{o(ipa,100)} - DF_c$ omit and $mF_{o(ipa,100)}$ maps show only a minor positive peak at the site of IPA 7 in TLN_90.

Quantifying Electron Density of the Solvent Structure

In addition to calculating $F_{o(ipa,x)} - F_{o(nat)}$ and $F_{o\ (ipa,x)}$ maps to validate bound molecules of isopropanol, we have attempted to quantify the electron density associated with

Figure 1. (Overleaf.) Stereo drawings of the active-site region of TLN showing the displacement of the VK dipeptide, and the binding of isopropanol. A and B show the refined TLN_2 structure superimposed with the dipeptide from TLN_0 (colored yellow). C shows the refined TLN_5 structure. Red cpk spheres denote bound water molecules in A, B and C. **A.** $mF_{o(ipa,\ 2)} - mF_{o\ (nat)}$ difference electron density map (contoured at ±3.5σ), at a resolution of 2.05 Å. Positive and negative difference density are shown in green and red, respectively. The phases were calculated from the refined TLN_0 model (see Materials and Methods). **B.** $2mF_{o(ipa,\ 2)} - DF_c$ map (contoured at 1σ) shows weak density in the region of the dipeptide (2.05 Å). **C.** $mF_{o(ipa,\ 5)} - DF_c$ omit map (contoured at +2.5σ), shows a molecule of isopropanol bound in the $S_1{}^1$ subsite (1.95 Å). Positive difference density is shown in green. No negative density was observed at this contour level. A nearby molecule of water was also omitted from the model when calculating the map to highlight the differences in shape between the electron density of isopropanol and water.

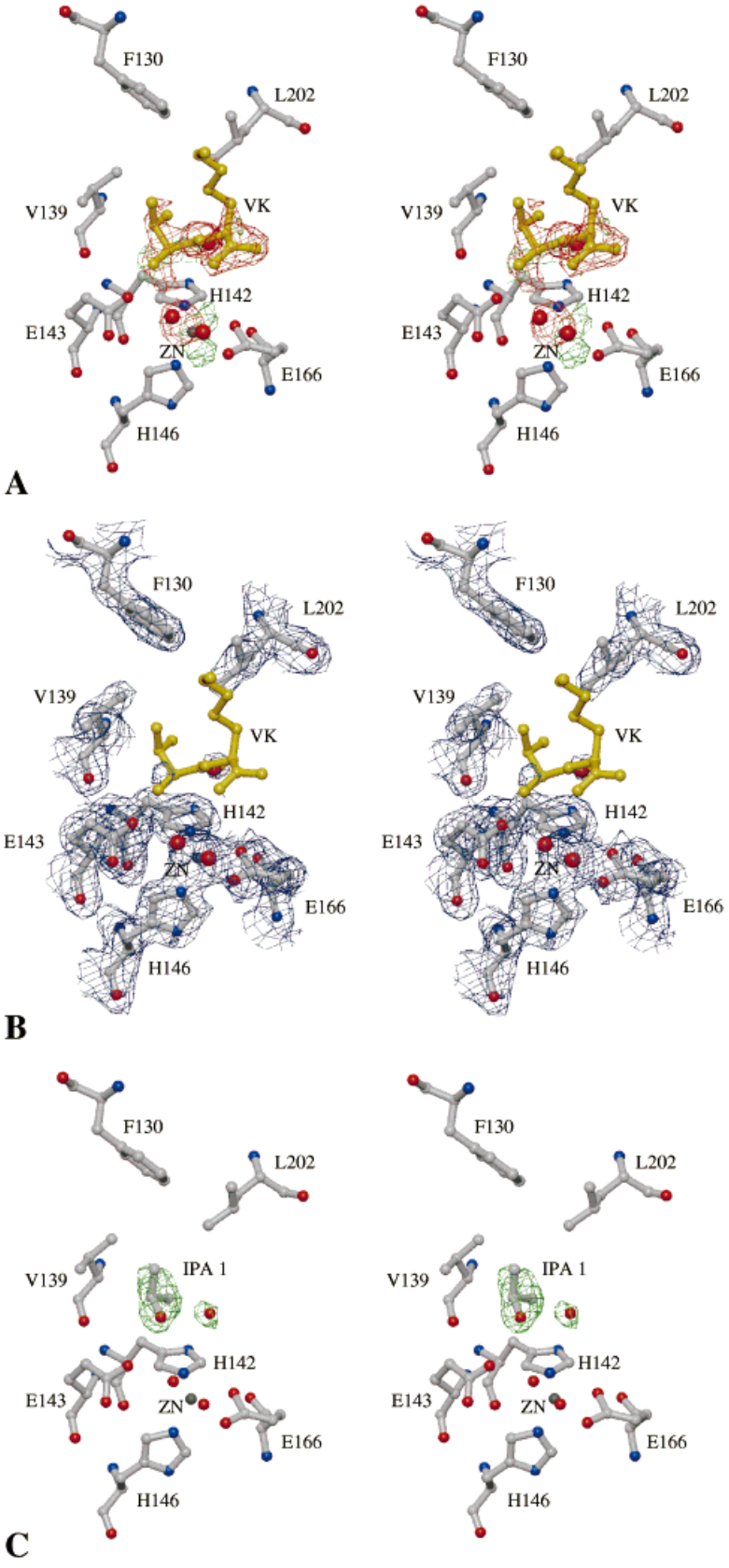


Figure 1. (Legend on preceding page.)

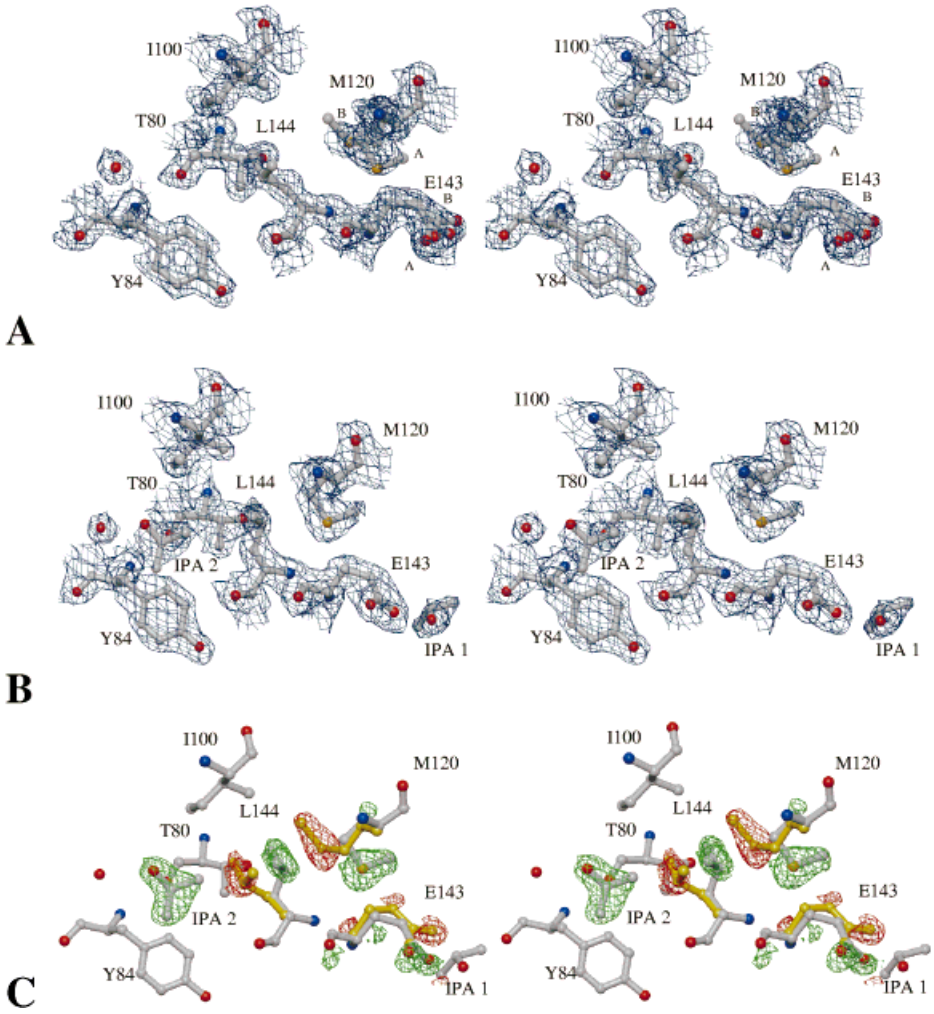


Fig. 2. Stereo drawings to show isopropanol binding reducing the disorder in the interior of TLN. A shows the refined TLN_0 structure. B and C show the refined TLN_10 structure. Red cpk spheres denote bound water molecules in A, B and C. **A.** $2mF_{o(nat)}-DF_c$ map (contoured at 1.2σ), at a resolution of 1.65 Å. The side chains of M120 and E143 are disordered. **B.** $2mF_{o(ipa,10)}-DF_c$ map (contoured at 1.2σ), at a resolution of 1.95 Å. The side chains of M120 and E143 are in a single conformer, and L144 has changed conformation, highlighting a "concerted re-packing".[11] **C.** $mF_{o(ipa,10)}-mF_{o(nat)}$ difference density map (contoured at $\pm 5.0\sigma$) (1.95 Å). Positive and negative difference density are shown in green and red, respectively. For clarity, the alternate B conformers of the M120 and E143 side chains, and the original conformation of L144 in the TLN_0 structure have been superimposed (colored yellow). The phases were calculated from the refined TLN_0 model (see Materials and Methods).

the solvent structure. Figure 4 compares the volume of the electron density in the $mF_{o(ipa,100)}-DF_c$ omit map (above a threshold of 4.0σ) assigned to isopropanol, with peaks assigned to water molecules of similar B factor. The $mF_{o(ipa,100)}-DF_c$ omit map was calculated omitting molecules of isopropanol and water molecules with B factors in the range 45–70 $Å^2$ from the model (33 waters), with an additional round of refinement performed without these atoms. Although absolute volumes for peaks in electron density maps are difficult to calculate, for this analysis we are interested only in comparing the relative volume of peaks. Thirty-four discrete peaks were found in the $mF_{o(ipa,100)}-DF_c$ omit map using X-LIGAND, above a search threshold of $+4\sigma$. Of the 34 peaks found, 10 correspond to modelled isopropanol molecules (colored red), 17 to waters (colored green), and 7 to peaks of ambiguous density (not shown in figure). Several of the unassignable peaks have a relatively large volume and could be molecules of isopropanol binding in a disordered manner. The mean volumes for the modelled isopropanol and water molecules (above 4.0σ) were 1.3 $Å^3$ (range 0.4–2.4 $Å^3$) and 0.5 $Å^3$ (range 0.2–1.0 $Å^3$), respectively. There is a clear trend that molecules of isopropanol tend to have a larger volume than bound water molecules of similar B factor. Of the waters that have a volume comparable to molecules of isopropanol, a number form networks and as such have a larger volume, since the peak extends over several water positions. Although X-ray crystallography cannot unambiguously identify these peaks as being due to isopropanol, the shape and volume of the electron density in the range of Fourier maps calculated are all consistent with protein bound molecules of isopropanol.

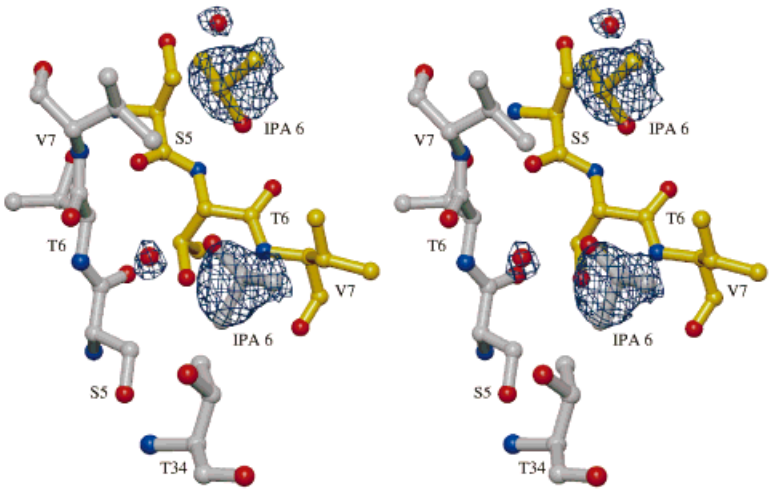


Fig. 3. Stereo drawing showing isopropanol (IPA 6) binding at a crystal contact in TLN. The figure shows the $mF_{o(ipa,90)}$ map (contoured at 0.3σ), at a resolution of 1.95 Å superimposed with the refined TLN_90 model. Red cpk spheres denote bound water molecules, and yellow bonds indicate the symmetry-related molecule. A nearby molecule of water (B value 28.5 Å^2) was also omitted from the model when calculating the map, to highlight the differences in shape between the electron density of isopropanol and water. For clarity, the $mF_{o(ipa,90)}$ map is displayed only over IPA 6 and a nearby water molecule.

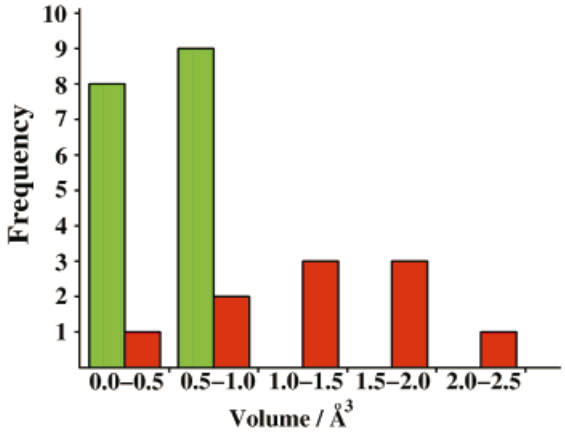


Fig 4. Comparison of the volume of the electron density for peaks in the $mF_{o(ipa,100)}-DF_c$ omit map assigned to isopropanol, with peaks assigned to water molecules of similar B in the TLN_100 structure (see Materials and Methods). Peaks in the map were found using the X-LIGAND module of QUANTA, at a threshold of +4.0σ. Sites assigned to molecules of isopropanol (colored red) tend to have a larger volume than water molecules of similar B factor (colored green).

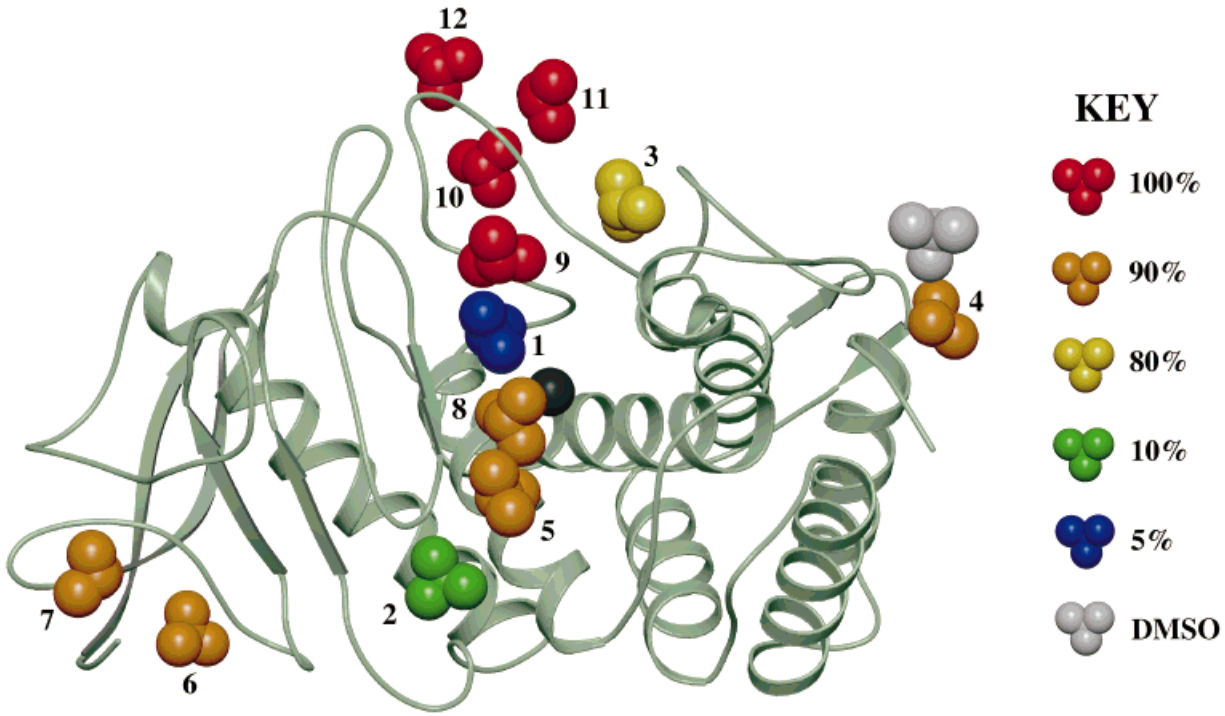


Fig. 5. Ribbon diagram of thermolysin (TLN_100) showing the binding sites for isopropanol. Molecules of isopropanol are shown in a cpk sphere representation and are colored according to the order of binding, from blue to red (see key). With the exception of IPA 7, the binding of isopropanol to TLN is additive. A bound molecule of dimethyl sulfoxide (DMSO) and the active site zinc ion are shown in cpk sphere representations and are colored grey and black, respectively.

CHARMM Interaction Energy

A CHARMM[17] interaction energy for each of the experimentally determined isopropanol positions was calculated. The TLN_100 structure, along with three symmetry-related copies, which were generated to account for isopropanol molecules located at crystal contacts, was used as the starting model. Polar hydrogens were added with the hydrogen positions placed optimally within HBUILD.[18] The overall system was initially subject to 100 steps of steepest descent minimization, fixing all non-hydrogen atoms and using a nonbonded cutoff of 8.0 Å. The whole system was then subject to five iterations of ten steps of steepest-descent minimization, and 100 steps of conjugate gradients minimization. The root-mean-square deviations (RMSD) between the starting model and minimized model was 0.3 Å, based on all atoms. After minimization, the interaction energy of each isopropanol with the rest of the system was evaluated. During evaluation of interaction energy the hydrogen of the hydroxyl group of each isopropanol was allowed to precess in steps of 2°. All calculations were performed using version 22 of the CHARMM program with the polar hydrogen model (PARAM19). The CHARMM interaction energy was plotted against mean B factor for each isopropanol, the height of peaks in the $mF_{o(ipa,100)}-DF_c$ omit electron density map at the site of each isopropanol, and the percentage of solvent at which each isopropanol was first observed. For peak height, an $mF_{o(ipa,100)}-DF_c$ omit map was calculated omitting the isopropanol molecules from refinement, and the CCP4 program PEAKMAX[13] used to give a relative value of the peak height at each isopropanol position. Values for correlation coefficients were obtained using the method of least-squares. The CHARMM interaction energy for IPA 7 was not calculated since it is not located in the TLN_100 structure.

HBPLUS[19] and LIGPLOT[20] were used to produce schematic diagrams of each of the binding sites for isopropanol in the TLN_100 structure, and IPA 7 in the TLN_90 structure. The cut off of the hydrogen bond and hydrophobic contacts used in the HBPLUS program was 2.5–3.3 Å and 2.9–4.3 Å, respectively. Diagrams for this paper were prepared using the programs MOLSCRIPT,[21] BOBSCRIPT,[22] and RASTER3D.[23,24]

RESULTS

As the conditions used to crystallize TLN were different to those of Holmes and Matthews,[11] a new native structure (hereafter TLN_0) (PDB code: 2tlx) was determined at 1.65 Å resolution, and refined from the coordinates of the original structure (PDB code:1lnf). The overall root-mean-square deviations (RMSD) between the 1lnf and TLN_0 structures is 0.1 Å and 0.2 Å for the main-chain and side-chain atoms, respectively. Both structures contain a dipeptide in the same conformation within the active site. There are 159 waters in 1lnf and 207 waters in TLN_0, with 145 of the waters in analogous positions ($<$2.0 Å apart). The main differences arise in the nature of the disorder at Met 120, Leu 144, and Tyr 157. For 1lnf, the weighting of side-chain conformers of Met 120 (A:B) and Leu 144 (A:B) are 0.2:0.8 and 0.1:0.7, respectively. In TLN_0, the weighting of the side-chain conformers of Met 120 (A:B) is 0.4:0.6, and Leu 144 has a single conformer of occupancy 0.8. In 1lnf, Tyr 157 has been modelled in a single conformer, whereas in TLN_0 the residue is disordered with a weighting for A:B of 0.5:0.3. The differences may be a consequence of the different crystallization conditions.

Unlike studies on subtilisin[8] and elastase,[5] it was not necessary to cross-link TLN crystals prior to soaking in organic solvents. Datasets were collected on TLN crystals soaked in 2, 5, 10, 25, 60, 80, 90, and 100% isopropanol solution to a resolution of 2.1 Å or better, except for the crystals soaked in 100% isopropanol where some of the crystals developed fine cracks, limiting diffraction to a resolution of 2.2 Å. Although the crystals survived the soaking procedure, the isopropanol variably affected the crystals, and in general led to an increase in mosaicity and decrease in maximum resolution from which data could be collected.

The isopropanol soaked crystals remain in space group $P6_122$ at all concentrations of solvent and difference Fourier techniques were used to solve each structure. However, the cell dimensions do shrink, such that in 100% isopropanol each cell dimension is reduced by approximately 1 Å compared to TLN_0, probably caused by an overall loss of water from the interstitial region. Following refinement, the structures were carefully inspected and isopropanol positions identified and modelled principally using F_o-F_o, F_o, and F_o-F_c difference Fourier techniques (see Materials and Methods). Details of the data collection and refinement are summarized in Tables I and II, respectively.

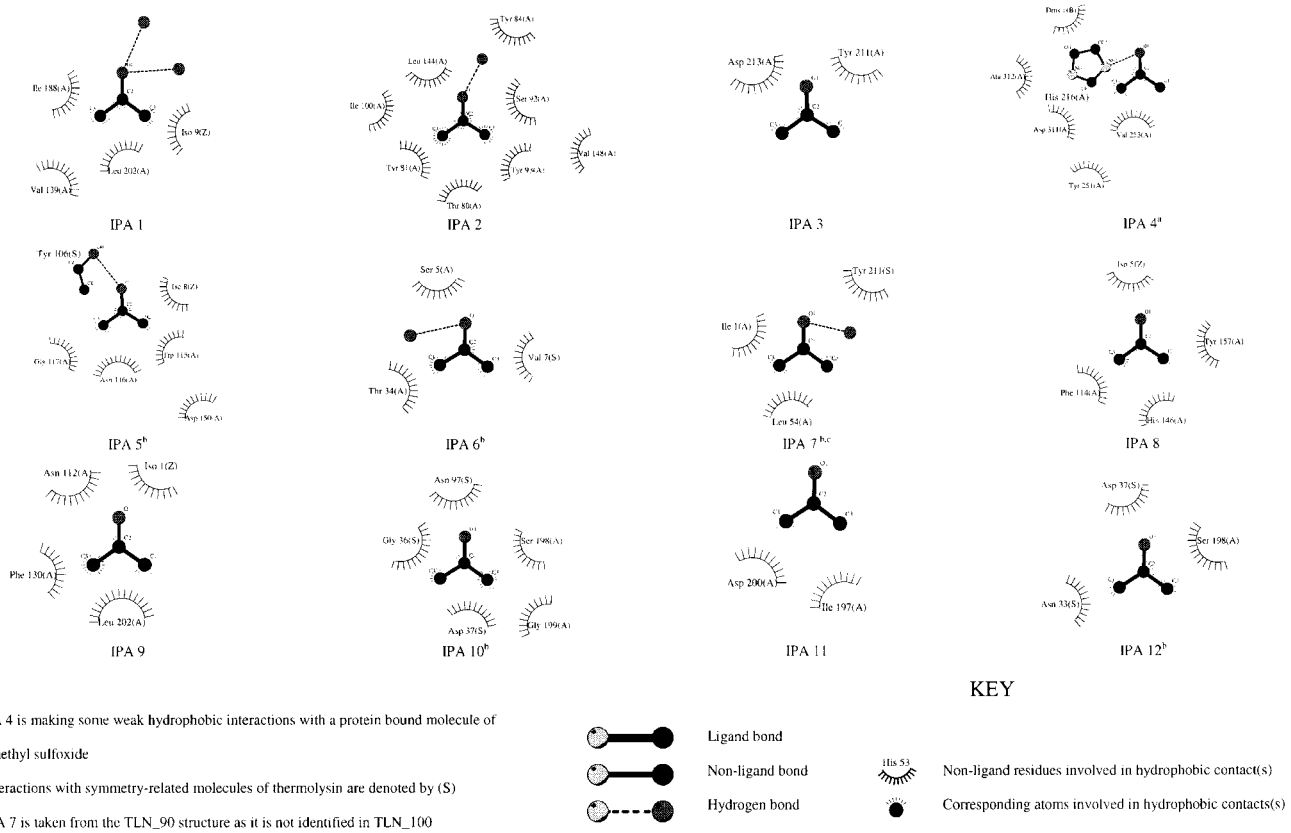
Overall Protein Structure

Root-mean-square deviations (RMSD) calculated between TLN_0 and the isopropanol soaked structures for main-chain atoms and side-chain atoms are in the range 0.1–0.2 Å and 0.3–0.5 Å, respectively, showing that there are only minor changes in the structure of TLN with increasing isopropanol concentration. The binding sites for isopropanol and details of the changes in the disordered residues and dimethyl sulfoxide (DMSO) are analyzed below.

Binding Sites for Isopropanol

As the isopropanol concentration used to soak the crystals increases, an increasing number of protein-bound isopropanol molecules could be identified. Previous work[5,25] has relied principally on the use of F_o-F_c and $2F_o-F_c$ maps to identify protein-bound solvent molecules. The determination of solvent binding to a protein is non-trivial and we have followed a more rigorous protocol, essential for distinguishing between isopropanol molecules and water in the electron density maps. Clearly, the correct identification of isopropanol or water molecules is an essential component of this approach. For clarity, we have included a detailed discussion of the crystallographic methods used to define solvent structure in Materials and Methods.

Table III. Schematic diagrams of the interaction sites for isopropanol with thermolysin (TLN_100)



[a]IPA 4 is making some weak hydrophobic interactions with a protein bound molecule of dimethyl sulfoxide

[b] interactions with symmetry-related molecules of thermolysin are denoted by (S)

[c] IPA 7 is taken from the TLN_90 structure as it is not identified in TLN_100

A total of twelve different binding sites for isopropanol were identified, eleven of which are observed in TLN_100 (summarized in Fig. 5). The molecules of isopropanol are numbered according to the order in which they were observed to bind, from IPA 1 in the TLN_5 structure, through to IPA 9, 10, 11, and 12 which are observed only in the TLN_100 structure. With the exception of IPA 7 (first identified in TLN_90), the binding of molecules of isopropanol to TLN is additive (discussed below). Isopropanol occupies all four of the main subsites in the active-site of TLN, although this was only observed at very high concentrations of isopropanol for three of the four subsites.

In TLN_2, the dipeptide in the active site has been displaced, with the region enclosed in negative difference density (Fig. 1A). However, there are no molecules of isopropanol visible in the various electron density maps, one of which is shown in Figure 1B. In TLN_5, a single molecule of isopropanol (IPA 1) is visible in the $S_1{}^1$ subsite, with the isopropyl group directed towards the interior of the pocket interacting with several hydrophobic residues (Fig. 1C). The O atom of the hydroxyl group is within hydrogen bonding distance of two water molecules, one of which is not present in TLN_0. A schematic diagram of the interactions IPA 1 makes with the protein in the TLN_100 structure are shown in Table III (see Materials and Methods).

Although a binding constant (K_b) for each molecule of isopropanol cannot be measured it is expected that the sites with greatest affinity for isopropanol will be occupied at lower concentrations. We suggest that the $S_1{}^1$ subsite which binds IPA 1 has the greatest affinity for isopropanol, under these conditions, and that the interactions of IPA 1 with several hydrophobic residues are responsible for the relatively strong binding.

In TLN_10, a second molecule of isopropanol (IPA 2) is visible (Fig. 2B and C) that binds in a buried, relatively hydrophobic subsite approximately 15 Å from the active-site zinc. The O atom of IPA 2 is within hydrogen-bonding distance of a water molecule also present in TLN_0. The isopropyl group of IPA 2 is able to form interactions with a number of hydrophobic residues within the pocket (Table III).

It is not until 80% isopropanol (TLN_80) that a third binding site is identified. This third bound molecule of isopropanol is located along the active-site cleft, binding in a location where a water molecule is found in TLN_0 (Table III).

In TLN_90, an additional five molecules of isopropanol are identified (IPA 4, 5, 6, 7, and 8), and the subsites where they bind have been grouped together in their ability to bind isopropanol. IPA 4 binds in a pocket close to a protein-bound molecule of DMSO, replacing a water molecule found in TLN_0. IPA 5 binds at a crystal contact near to the S_2 subsite in the active site displacing three symmetry-related molecules of water found in TLN_0. Both IPA 6 (Fig. 3) and 7 bind at crystal contacts in locations where

nothing was observed in TLN_0. IPA 8 binds in close vicinity to the S_1 subsite in the active site displacing a molecule of water present in TLN_0. The interactions IPA 4, 5, 6, and 8 make with the protein in the TLN_100 structure are shown schematically in Table III. Since IPA 7 is not located in TLN_100, the interactions it makes with the protein in the TLN_90 structure are shown (Table III).

In TLN_100, a further four molecules of isopropanol are identified (IPA 9,10,11, and 12), and we suggest these represent the lower affinity sites for TLN. IPA 9 binds in the $S_2{}^1$ subsite which the Lys side chain of the Val-Lys dipeptide occupies in TLN_0. IPA 10 binds at a crystal contact where nothing was observed in TLN_0. IPA 11 is located in a binding pocket replacing a molecule of water. IPA 12 binds at a crystal contact where nothing was observed in TLN_0. Reliable orientation of this molecule was more difficult since it makes more distant interactions with the protein. The interactions IPA 9, 10, 11, and 12 make with the protein in the TLN_100 structure are shown schematically in Table III.

As noted earlier IPA 7 (first identified in TLN_90) is not clearly visible in the electron density maps for TLN_100. We suggest that isopropanol is still interacting with the site in TLN_100, but there may be a high degree of motion associated with the site, such that the electron density for IPA 7 is weak (see Materials and Methods). In addition, the quality of the data for TLN_100 was not as high as for the other soaked crystals.

Concerted Re-Packing of Met 120, Glu 143, and Leu 144

In TLN_0, the side chain of Met 120 occupies two discrete conformations, and Glu 143 is disordered (Fig. 2A). In contrast, in all the soaked structures Met 120 and Glu 143 each favor only a single conformer, and Leu 144 has changed conformation (Fig. 2B). The $mF_{o(ipa,10)}-mF_{o(nat)}$ difference electron density map (see Materials and Methods) (Fig. 2C) shows that the A conformer side chains of Met 120 and Glu 143 are surrounded by positive difference density, and the regions that the B conformers occupy in TLN_0 are surrounded by negative difference density. In addition, the side chain of Leu 144 (in TLN_10) is surrounded by positive difference density, and the original position of side chain (in TLN_0) is surrounded by negative difference density. This highlights a "concerted re-packing".[11] When IPA 2 binds in the buried subsite, the isopropyl group directly interacts with the side chain of Leu 144, causing a change in conformation of this residue. This then causes a rearrangement of the side chains of Met 120 and Glu 143 such that they occupy a single conformer (A:B 0.4:0.6 to 1.0:0.0 for Met 120, and A:B 0.7:0.3 to 1.0:0.0 for Glu 143). This re-weighting of the side chains is observed in every soaked structure.

Dimethyl Sulfoxide

DMSO is present in the crystallization conditions (45% v/v) and the crystal structure of TLN_0 has a molecule of DMSO bound. The O of the molecule of DMSO is able to form a hydrogen bond to a water molecule (3.2 Å). The C1 atom is interacting with Tyr 251 (3.7 Å), and the C2 atom is interacting with His 216 (3.6 Å). The S atom of DMSO is interacting with the ring of Tyr 66 (3.8 Å), of a symmetry-related molecule. The molecule of DMSO is retained in all the crystal structures (Fig. 5), although as the concentration of isopropanol increases the electron density for DMSO is seen to gradually diminish in the $2mF_{o\,(ipa,\,x)}-DF_c$ and $mF_{o(ipa,x)}-DF_c$ omit maps (see Materials and Methods). In TLN_100, DMSO is enclosed in negative difference density in the $\rho_{o(ipa,100)}-\rho_{o(nat)}$ map (contoured at 3σ), although the electron density in the $mF_{o(ipa,100)}-DF_c$ omit map still merits inclusion of DMSO in the model, albeit at a lower occupancy. We suggest that this reduction in electron density is due to the effects of changing the solvent the crystals are soaked in; as the concentration of isopropanol increases, DMSO gradually diffuses out of the crystal lattice.

CHARMM Interaction Energy

A CHARMM[17] energy for the interaction between each of the experimentally determined isopropanol positions and the rest of the system was calculated in vacuo, neglecting solvation effects. The CHARMM interaction energy was plotted against mean B factor for each isopropanol, the height of peaks in the $mF_{o(ipa,100)}-DF_c$ omit electron density map at the site of each isopropanol, and the percentage that each molecule was first observed. Peak height and B factor are both related to the total electron density. While peak height in an F_o-F_c omit map gives some measure of the experimental data, B factor will depend more on the refinement protocol adopted. There is a slight trend that increasing B factor is associated with weaker binding (correlation coefficient 0.62). Such a weak correlation is perhaps not surprising, since a number of the isopropanol binding sites (especially those located only at high concentrations) display multiple binding modes. At such mixed population sites the B factors for these sites will be less accurate (see Materials and Methods). A slight correlation that stronger electron density is associated with stronger binding is observed (correlation coefficient 0.64). There is almost no correlation between the calculated CHARMM interaction energy and the concentration of isopropanol at which each site is observed (correlation coefficient 0.56).

Comparison With Known Inhibitors

IPA 1, 5, 8, and 9 are positioned in, or nearby the known subsites of the active site of TLN,[10] and it is therefore possible to compare the position and orientation of these isopropanol molecules with those obtained for similar functional groups in the published structures of TLN-ligand complexes.

IPA 1 binds in the $S_1{}^1$ subsite that is known to prefer aliphatic residues. All published inhibitors of TLN possess a hydrophobic group at the corresponding $P_1{}^1$ position. Superposition of the TLN_90, TLN_0 and 3tmn[26] shows the isopropyl group of IPA 1 to closely mimic the position of the Val side chain in both dipeptides (Fig. 6A). In addition, a water molecule in the TLN_90

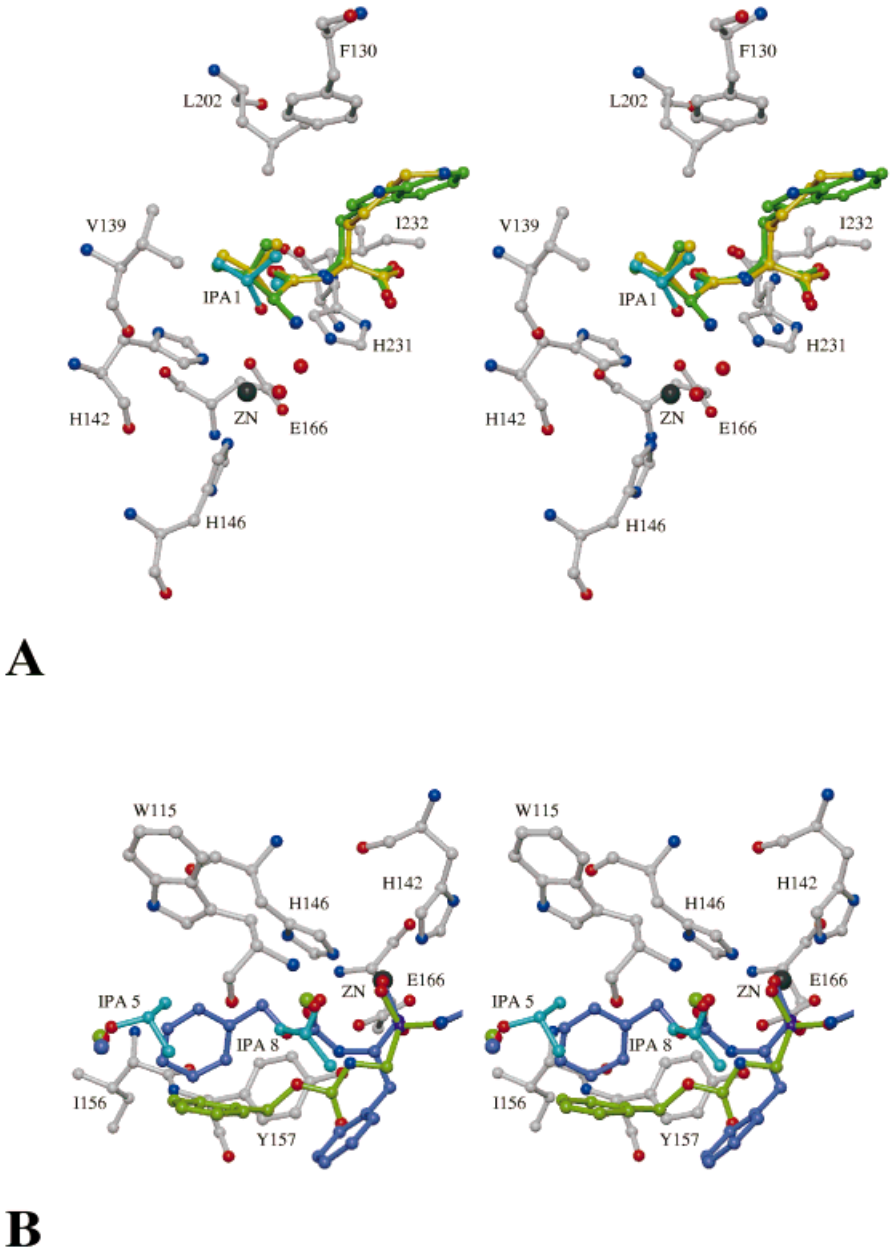


Fig. 6. Stereo diagrams of the active-site of TLN to compare the position and orientation of protein-bound molecules of isopropanol with similar functionality present in published TLN-ligand complexes. **A.** The final TLN_0 model overlaid with TLN_90 and 3tmn. The VK dipeptide is colored yellow and red cpk spheres denote bound water molecules. C atoms and water molecules associated with the other structures are color-coded: TLN_90 (cyan) and 3tmn (green). **B.** The final TLN_0 model overlaid with TLN_90, 4tmn and 6tmn, red cpk spheres denote bound water molecules. C atoms and water molecules associated with the other structures are color-coded: TLN_90 (cyan), 4tmn (lilac) and 6tmn (lime-green). The phosphorous atoms in 4tmn and 6tmn are colored purple. For clarity, some of the solvent structure in A and B and the B conformer of Y157 (in TLN_0) have been omitted. The structures were overlaid in QUANTA using the C-α positions.

structure can be envisaged as mimicking the O atom of the peptide bond in both these inhibitors. Although IPA 1 makes favorable hydrophobic interactions (Table III), the isopropyl group is located towards the top of the pocket, which can accommodate much larger hydrophobic groups.

IPA 5 binds in the S_2 subsite. Superposition of TLN_90, TLN_0, 4tmn[27] and 6tmn[28] shows that the hydroxyl O atom of IPA 5 is located in a conserved site for water molecules (Fig. 6B). In addition, the isopropyl group is within 2.0 Å of the hydrophobic Cbz groups in ZF^{P}LA (4tmn)[27] and ZG^P(O)LL (6tmn),[28] consistent with substituent preferences for the S_2 subsite.

IPA 8 binds near to the S_1 subsite. Superposition of TLN_90, TLN_0, and 6tmn shows that the hydroxyl O atom of IPA 8 is located in a conserved site for water

molecules (Fig. 6B). Also, the hydroxyl O atom in IPA 8 is 0.3 Å from the CBZ ester bond oxygen in ZF^{P}LA (4tmn),[27] so the group is mimicking similar functionality present in a larger inhibitor. In addition, the isopropyl group of IPA 8 can be envisaged as mimicking the backbone of ZF^{P}LA.

IPA 9 binds in the $S_2{}^1$ subsite. Superposition of the inhibitor N-(1-carboxy-3-phenylpropyl)-L-leucyl-L-tryptophan (1tmn)[29] with TLN_100 shows that the C1 and C2 atoms of IPA 9 mimic the Cβ and Cγ atoms of the Trp side chain, respectively. The O atom of the hydroxyl group of IPA 9 in approximately the same location as the Nϵ1 atom of the Trp side chain of the inhibitor which is within hydrogen-bonding distance of the main-chain O atom of Asn 111.

DISCUSSION

We have extended MSCS[5] by the determination of the high resolution crystal structures of thermolysin (TLN), generated from crystals soaked in 2% to 100% isopropanol. We have observed the progressive binding of isopropanol to TLN, and attempts have been made to correlate the experimental positions with theoretical methods.

Pure TLN is commercially available, crystallizes readily, and rapid data collection on many different solvents and solutes is possible (unpublished data). In addition, crystals of TLN do not require cross-linking for stability and despite the crystals developing fine cracks at high concentrations of organic solvent high resolution data can still be collected. This is a similar finding to work on γ-chymotrypsin where crystals were soaked in hexane and cross-linking was not required.[30] The presence of a bound dipeptide[11] in the active site raised some initial concerns but in fact provided a convenient probe for changes in substrate binding.[30] Careful examination of F_o-F_o and F_o electron density maps together, provides a rigorous way of locating and validating bound solvent molecules.

The procedure causes only minor changes to the conformation of the protein, and an increasing number of isopropanol interaction sites could be identified as the solvent concentration is increased. Isopropanol was observed to bind on the surface of the protein, at crystal contacts and in the interior of the protein where disordered protein side chains were stabilized. Binding of isopropanol to the protein is additive, such that molecules located in crystal structures where the soak mixture was predominantly aqueous, bind in the same locations when the crystals are soaked in neat solvent.[25] By varying the concentration of organic solvent the higher affinity sites, more relevant to ligand design, can be identified. Indeed, IPA 1 (first located in the largely aqueous conditions for TLN_5) binds in the main $S_1{}^1$ specificity pocket, presumably reflecting the fact that enzyme active sites evolve to bind substrates.[25] Molecules of isopropanol located only at very high concentrations of solvent (~15 M) are binding very weakly, and display a number of possible binding modes. Clearly, the positioning of isopropanol in one of these very weak binding sites (on the periphery of the active-site) would not be expected to serve as part of a template for the design of a new ligand.

Although MSCS potentially allows an entire search of the protein surface for sites complementary to the probe molecule, there will be regions such as crystal contacts and strongly bound water molecules that will remain inaccessible to organic solvent. In addition, the binding sites identified from soaking protein crystals in organic solvents are not just those that are thermodynamically favorable. The sites also must be kinetically accessible during the crystal soaking (18–24 hours) for them to be observed in the electron density maps.

MSCS has identified a number of binding sites for isopropanol in the active site of TLN, and these are consistent with the structures of known protein-ligand complexes of TLN. For a crystalline protein with no known inhibitors or where detailed structural information about substrate binding is absent, MSCS could be used to identify important subsites within the active site. Recent studies[31] have shown that the binding of ligand fragments may conform to the principles of additivity, suggesting a modular approach to rational drug design. Applied to therapeutically relevant systems MSCS may provide a basis for new leads, or the optimization of known inhibitors. MSCS could also aid the design of more focused combinatorial libraries, by screening a target protein with potential fragments prior to their incorporation into a library.

Attempts to correlate the CHARMM interaction of each experimentally determined isopropanol with the protein against electron density showed little correlation, emphasizing the need for experimental data to address the general problem of ligand binding. Currently, solvation[32] and entropic effects are not properly accounted for computationally but as more empirical data becomes available this may improve.[1] Work is in progress using other systems and solvents (acetone, phenol, acetonitrile, ethanol, etc) to construct experimental functionality maps and to compare the experimentally determined solvent positions with positions predicted by computational approaches.[7,33,34]

ACCESSION NUMBERS

Atomic coordinates for the refined models together with structure factors have been deposited in the Brookhaven Protein Data Bank with the accession codes 2tlx, 1tli, 2tli, 3tli, 4tli, 5tli, 6tli, 7tli and 8tli.

ACKNOWLEDGMENTS

This research was supported by a grant from the Biotechnology and Biological Sciences Research Council. We thank Eleanor Dodson and John Overington for helpful discussions, Alec Tucker for the adapted crystallization conditions, and Steve Mumford for advice in the preparation of the figures. We thank Johan Turkenburg with help during data collection and processing of the 1.65Å native TLN dataset, and acknowledge the use of Daresbury SRS.

REFERENCES

1. Bohm HJ, Klebe G. What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? Ange Chem Int Ed Eng 1996;35:2589–2614.

2. Salemme FR, Spurlino J, Bone R. Serendipity meets precision: the integration of structure-based drug design and combinatorial chemistry for efficient drug discovery. Structure 1997;5:319–324.
3. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins: SAR by NMR. Science 1996;274: 1531–1534.
4. Hajduk PJ, Sheppard G, Nettesheim DG, et al. Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. J Am Chem Soc 1997;119:5818–5827.
5. Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, Ringe D. An experimental approach to mapping the binding surfaces of crystalline proteins. J Phys Chem 1996;100:2605–2611.
6. Mattos C, Ringe D. Locating and characterizing binding sites on proteins. Nat Biotechnol 1996;14:595–599.
7. Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. Proteins 1991;11: 29–34.
8. Fitzpatrick PA, Steinmetz ACU, Ringe D, Klibanov AM. Enzyme crystal structure in a neat organic solvent. Proc Natl Acad Sci USA 1993;90:8653–8657.
9. Holmes MA, Matthews BW. Structure of thermolysin refined at 1.6 Å resolution. J Mol Biol 1982;160:623–639.
10. Matthews BW. Structural basis of the action of thermolysin and related zinc peptidases. Acc Chem Res 1988;21:333–340.
11. Holland DR, Hausrath AC, Juers D, Matthews BW. Structural analysis of zinc substitutions in the active site of thermolysin. Protein Sci 1995;4:1955–1965.
12. Leslie AGW, Brick P, Wonacott AJ. Joint CCP4 and ESF-EACMB newsletter on protein crystallography. Warrington, UK: Daresbury Laboratory. CCP4 Newsletter 1986;18:33–39.
13. CCP4. Collaborative computational project, number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
14. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the maximum likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255.
15. Oldfield TJ. Real space refinement as a tool for model building. In: Macromolecular refinement. Proceedings of the CCP4 study weekend. Warrington, UK:SRS Daresbury Laboratory; 1996a. p 67–74.
16. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
17. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4: 187–217.
18. Brunger AT, Karplus M. Polar hydrogen positions in proteins: empirical energy placement and neutron-diffraction comparison. Proteins 1988;4:148–156.
19. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
20. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 1995;8:127–134.
21. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J App Cryst 1991;24:946–950.
22. Esnouf RM. An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. J Mol Graph 1997;15:132–134.
23. Bacon D, Anderson WF. A fast algorithm for rendering space-filling molecule pictures. J Mol Graph 1988;6:219–220.
24. Merritt EA, Murphy MEP. RASTER3D version 2.0: a program for photorealistic molecular graphics. Acta Crystallogr D Biol Crystallogr 1994;50:869–873.
25. Schmitke JL, Stern LJ, Klibanov AM. Organic solvent binding to crystalline subtilisin in mostly aqueous media and in the neat solvents. Biochem Biophys Res Comm 1998;248:273–277.
26. Holden HM, Matthews BW. The binding of L-valyl-L-tryptophan to crystalline thermolysin illustrates the mode of interaction of a product of peptide hydrolysis. J Biol Chem 1988;263:3256–3260.
27. Holden HM, Tronrud DE, Monzingo AF, Weaver LH, Matthews BW. Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphonamidate transition-state analogues. Biochemistry 1987; 26:8542–8553.
28. Tronrud DE, Holden HM, Matthews BW. Structures of two thermolysin-inhibitor complexes that differ by a single hydrogen bond. Science 1987;235:571–574.
29. Monzingo AF, Matthews BW. Binding of N-carboxymethyl dipeptide inhibitors to thermolysin determined by x-ray crystallography: a novel class of transition-state analogues for zinc peptidases. Biochemistry 1984;23:5724–5729.
30. Yennawar NH, Yennawar HP, Farber GK. X-ray crystal structure of γ-chymotrypsin in hexane. Biochemistry 1994;33:7326–7336.
31. Stout TJ, Sage CR, Stroud RM. The additivity of substrate fragments in enzyme-ligand binding. Structure 1998;6:839–848.
32. Caflisch A. Computational combinatorial ligand design: application to human alpha-thrombin. J Comp Aided Mol Des 1996;10: 372–396.
33. Goodford PJ. A computational-procedure for determining energetically favourable binding sites on biologically important macromolecules. J Med Chem 1985;28:849–857.
34. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161:269–288.