Acta Cryst. (1986). B42, 379-387

Comparison of Two Independently Refined Models of Ribonuclease-A

BY ALEXANDER WLODAWER

Center for Chemical Physics, National Bureau of Standards, Gaithersburg, MD 20899, *USA and Laboratory of Molecular Biology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD* 20205, *USA*

AND NIVEDITA BORKAKOTI, DAVID S. MOSS AND BRENDAN HOWLIN

Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England

(*Received* 4 *February* 1985, *accepted* 3 *February* 1986)

Abstract

The X-ray structure of ribonuclease-A refined in London and the joint X-ray and neutron structure refined in Washington are compared. The two structures are shown to be very similar, the mean difference between the protein atoms of the two structures being 0·374 Å. The two models differ in assignments of atomic positions of the active-site side-chain atoms in Lys 41 and His 119 and these differences may be due to different conformations in the two structures. Strongly bound solvent occupies the same position in both structures but there is some evidence of other solvent structure being different, possibly owing to the presence of different alcohols or the use of partial deuteration in Washington. Temperature factors in both structures show similar trends.

Introduction

The structure of bovine pancreatic ribonuclease-A (RNase-A) has been refined independently in two laboratories. The structure was refined using first only X-ray and later both X-ray and neutron data extending to 2·0 Å in Washington (Wlodawer, Bott & Sjölin, 1982; Wlodawer & Sjölin, 1983), while 1·45 Å X-ray data were used by the London group (Borkakoti, Moss & Palmer, 1982; Borkakoti, Moss, Stanford & Palmer, 1984). The common starting model used in both investigations was refined using 2·5 Å X-ray data by Wlodawer (1980). Both refinements, using data at different resolutions, were accomplished using different computer algorithms, with crystals grown under different conditions from different solvents and, in the case of the Washington structure, utilizing not only X-ray but also neutron structure amplitudes. Nevertheless, the final structures were expected to be similar, and the aim of the following comparison is to investigate whether this was actually the case, as well as to try to rationalize the discrepancies.

The problem of the level of confidence one can place in the refined protein structures, as well as the correlation between the errors in the model and such measures of the fit of coordinates and diffraction data as R factors, have been repeatedly addressed before. For example, Derewenda, Brzozowski, Stepien & Grabowski (1982) have concluded that the estimate of root-mean-square errors in the atomic coordinates obtained using the formulas of Luzzati (1952) is close to being correct, even though it was usually assumed to be an upper limit. Reliable estimation of standard deviations from least-squares refinement requires the determination of the inverse of a normal matrix which includes the significant off-diagonal terms. This is not usually available for protein structures. A further difficulty in the assessment of errors in these structures is the presence of static and dynamic disorder. The true structure is neither the mean nor the modal conformation and may require multimodal distributions of the atomic positions for its correct description. This problem presents the crystallographer with special difficulties when designing a realistic structure-factor model for protein molecules.

Resolution of these problems is assisted when more than one estimate of the structure exists. A gain in utilizing two independently refined protein structures in estimating the accuracy of atomic positions of either one has been shown by Chambers & Stroud (1979). They have compared the highly refined model of diisopropyl fluorophosphate-inhibited trypsin of Chambers & Stroud (1977) with a model of the benzamidine-inhibited enzyme obtained by Bode, Schwager & Huber (1976). The comparison has shown that the r.m.s. deviation between large parts of the respective models was in good agreement with the internal estimates of the errors, provided by methods such as Luzzati's (1952) or Cruickshank's (1949). While the comparison of the two models answered a number of questions regarding the relative accuracy and the relationship between calculated and observed errors, it failed in one important respect, namely in the evaluation of the reasons for disagreement in those parts of the structure where the differen-

0108-7681/86/040379-09$01.50

ces were the largest. This was principally because the comparisons were performed by only one group, using the published coordinates of the competing structure. Selected atoms in the model of Bode, Schwager & Huber (1976) were flagged as not visible in the Chambers & Stroud (1977) electron density map. However, since two very different methods of refinement were employed (*viz* difference Fourier techniques with idealization of molecular models by Chambers & Stroud and real-space refinement by Bode, Schwager & Huber), there were different distributions of errors in the two structures. Further, individual temperature factors were used in the Bode, Schwager & Huber model whereas the Chambers & Stroud model employed an overall temperature factor, causing ambiguity in the assignment of proper confidence levels to the doubtful parts of the structure. Nevertheless, the comparison of these two independently refined models of trypsin has led to a better understanding of the level of confidence one should place in the refined structure of a protein.

More recently, Marquart, Walter, Deisenhofer, Bode & Huber (1983) compared sixteen crystal structures of the system which included trypsin, trypsinogen and several inhibitors. All the refinements were performed using identical least-squares methods. The final discrepancies were small and could usually be directly traced to the differences caused by crystal packing. While providing valuable information about the geometry of highly refined proteins, this study did not deal in the strict sense with completely independent structure determinations.

For the reasons summarized below, we have considered that the comparison of the Washington (W) and London (L) structures of RNase-A might yield even more information than the comparison of the structures of trypsin. Both structures were refined starting from an identical model which, in retrospect, was shown to exhibit serious shortcomings. Thus the agreement of those parts which had to be completely rebuilt would indicate that the procedures followed in each laboratory were sound. While both structures were refined using least-squares methods, the actual programs were different, and the comparison of the results was of interest. Both models have a similar degree of departure from ideal geometry, facilitating the comparisons (Table 1). In addition to answering the question about the relationship between the estimated and actual accuracy of the models, we were particularly interested in investigating those features which might potentially be truly different between them. The differences were expected to be caused mostly by the requirements of maximal deuteration imposed by the desire to lower the incoherent background in the neutron measurements (Wlodawer, Bott & Sjölin, 1982), which led to the soaking for six months of the crystals used in Washington in a fully deuterated mother liquor. The degree of perturbation of the structure by such a procedure was not known but should be of general interest in neutron protein crystallography, where deuteration is a norm rather than an exception. The availability of the neutron data should also remove the ambiguity of orienting some side chains which appear symmetric in the X-ray maps (histidine, glutamine, asparagine), but which should be quite asymmetric in the neutron maps. Finally, the solvents used in both investigations were different, and that fact, together with the deuteration of the W crystals, was expected to affect the observed solvent structure. All of these questions will be addressed below.

Table 1. *Details of the Washington and London refinements of ribonuclease-A*

X and *N* refer to X-ray and neutron data respectively.

	Washington	London
Resolution of data	2·0 Å	1·45 Å
Number of reflections	7708 (*X*) 4132 (*N*)	19 098 (*X*) —
R factors*	0·159 (*X*) 0·183 (*N*)	0·223 (*X*) —
R.m.s. deviation of bond lengths (Å)	0·022	0·018
R.m.s. deviation of angle distances (Å)	0·062	0·038
R.m.s. deviation from planarity (Å)	0·019	0·013
R.m.s. deviation of thermal parameters of bonded atoms (Å^2)	0·095	—

* *R* factors for the Washington data excluded reflections with values of less than three standard deviations. All reflections were included in the London data.

Experimental procedures

(*a*) *Crystals and data collection*

The details of the procedures followed by the Washington group are given by Wlodawer (1980), Wlodawer, Bott & Sjölin (1982) and Wlodawer & Sjölin (1983). Crystals of RNase were grown from a solution containing 43% 2-methyl-2-propanol, pH = 5·3, and had unit-cell parameters $a = 30{\cdot}18$, $b = 38{\cdot}4$, $c = 53{\cdot}22$ Å, $\beta = 105{\cdot}85°$, space group $P2_1$. These crystals were soaked for about six months in a fully deuterated synthetic mother liquor, with several complete changes of the solvent. Neutron data were collected from one crystal using the flat-cone diffractometer (Prince, Wlodawer & Santoro, 1978), while the X-ray data set was measured using three crystals on a four-circle diffractometer. Of all the reflections in the 10 to 2 Å shell, 7708 (95·6%) were observed $[I > 3\sigma(I)]$ in the X-ray data set, while only 4132 (51%) of the neutron intensities were present. This disparity was caused by the low flux of neutrons and by the poor signal-to-noise ratio inherent in the neutron diffraction of proteins (Wlodawer, Bott &

Sjölin, 1982). Absorption corrections were applied using the method of North, Phillips & Mathews (1968).

The detailed data on the crystals and procedures used by the London group are given by Carlisle, Palmer, Mazumdar, Gorinsky & Yeates (1974) and by Borkakoti *et al.* (1982). Crystals of RNase-A were grown from 40% ethanol and their unit-cell parameters were $a = 30{\cdot}31$, $b = 38{\cdot}26$, $c = 52{\cdot}91$ Å and $\beta = 105{\cdot}91°$. The differences in the unit-cell parameters compared to the W crystals are much larger than the errors in their measurements. A Hilger & Watts four-circle diffractometer was used to measure intensities. 35 702 intensities from five crystals were measured in shells of 200 to 400 reflections each. A semi-empirical absorption correction (North *et al.*, 1968) was applied. Scaling and averaging of the data sets gave 19 098 intensities with intensity greater than zero (82% of the theoretical maximum at 1·45 Å resolution).

(*b*) *Initial model*

Both refinements discussed in this paper were initiated from an identical model, and some properties of this initial structure are of importance in the understanding of their course. The preliminary model of ribonuclease-A (Wlodawer, 1980) was refined using X-ray data extending between 10 and 2·5 Å, collected using deuterated crystals, as described below for the W structure. The starting point was provided by the unpublished coordinates of RNase determined as described by Carlisle *et al.* (1974) and made available to AW. The attempts to refine this model without rebuilding proved unsuccessful. Several cycles of refinement using the method of Hendrickson & Konnert (1981) resulted only in the reduction of the crystallographic R factor from 45 to 42% at 3 Å resolution. Rather than rebuilding the model at this stage, it was decided that a faster approach could be provided by rotating the refined model of ribonuclease-S (Powers, 1976) onto Carlisle's coordinates of RNase-A, using Cα positions as guides. This was indeed done, using a program of Hendrickson (1979), matching the positions of Cα's 2–16 and 24–124. Thus the starting model consisted of residues 1–16 and 24–124 from the rotated structure of Powers while the residues 17–23 were taken unmodified from Carlisle *et al.* The model synthesized in this manner refined to an R factor of 0·31 without manual rebuilding and to an R factor of 0·252 after several cycles of rebuilding using a computer graphics program *BILDER* (Diamond, 1981), followed by least-squares refinement. The final model did not include any solvent other than the phosphate present in the active site [shown later by Borkakoti *et al.* (1982) to be a sulfate], and had an overall temperature factor $U_{iso} = 0{\cdot}07$ Å^2.

The orientation of residues which appear symmetric in the electron density maps but in reality are not (such as the side chains of glutamine, asparagine and histidine) was not checked with the view of assigning more probable hydrogen bonds, and thus these groups were randomly oriented. Another major problem of the starting model, confirmed independently by the L and W models, was in the misplacement of residues 17–20. These residues (Thr-Ser-Ala-Ala) were shifted by one location up the chain, bulging out at residue 21 and skipping the true course of the chain between residues 16 and 17.

While the problems discussed above clearly affected the course of the refinements by both groups, other doubtful regions were also present in the initial model and will be discussed further below.

(*c*) *Refinement and rebuilding*

The W structure was refined using the algorithm of Hendrickson & Konnert (1981), modified by Wlodawer & Hendrickson (1982) to allow for the simultaneous utilization of both X-ray and neutron diffraction data in the evaluation of the refinement matrix. This method of refinement introduces stereochemical and other prior knowledge about the structure into the least-squares minimization. These geometrical 'observations' serve as restraints on the atomic parameters. In the refinement described below, they included ideal bond lengths and angles, planarity of peptide groups, of imidazole and phenyl rings, and of side-chain amide and guanidinium groups, chirality at asymmetric centres, energy penalty for close nonbonded contacts, restricted torsion angles and limitations on the deviations of the isotropic temperature factors of linked atoms. While hydrogen and deuterium positions and temperature factors were allowed to vary, their contribution was removed from the calculations of the X-ray structure factors, since their inclusion was not warranted at 2 Å resolution.

The model was modified repeatedly between the refinement runs with the help of computer graphics programs *BILDER* (Diamond, 1981) and *FRODO* (Jones, 1982). Two types of maps were calculated, using as coefficients either $(2F_o - F_c)$ or $(F_o - F_c)$. In the latter case, about 8% of the atoms were removed from the F_c calculations for each 'fragment ΔF' map (Wlodawer *et al.*, 1982). The final model was deposited with the Brookhaven Protein Data Bank as file 4RSA.

The L structure was refined by the least-squares procedure *RESTRAIN* of Haneef, Moss, Stanford & Borkakoti (1985) alternating with model-building sessions using the program *FRODO* (Jones, 1982). The least-squares method applied geometrical restraints to bonded distances, distances across bond angles and planar restraints to the same groups as in

the W structure. The distance restraints were applied by including appropriate squared terms in the residual minimized as in the algorithm of Hendrickson & Konnert (1981) but the planarity restraints were applied in a different way in the L structure as described by Haneef *et al.* (1985). No H-bond or nonbonded interactions were restrained and independent isotropic mean-square displacements were refined for protein and solvent atoms. The side group of His 119 was refined in two positions using group occupancy factors coupled together so that their sum was unity. Table 1 contains some parameters for the models which resulted from the respective restrained refinements. This model was a result of further refinement of the data set deposited with the Brookhaven Protein Data Bank as 1RN3.

Results

(*a*) *Atomic coordinates*

A comparison of the positional coordinates was facilitated by least-squares minimization of the r.m.s. difference between the L and W sets of protein atoms. The root-mean-square (r.m.s.) differences between the two protein models after this minimization are shown in Fig. 1, are summarised in Table 2 and are discussed below.

The deviation between the L and W main-chain coordinates has been subjected to analysis of variance

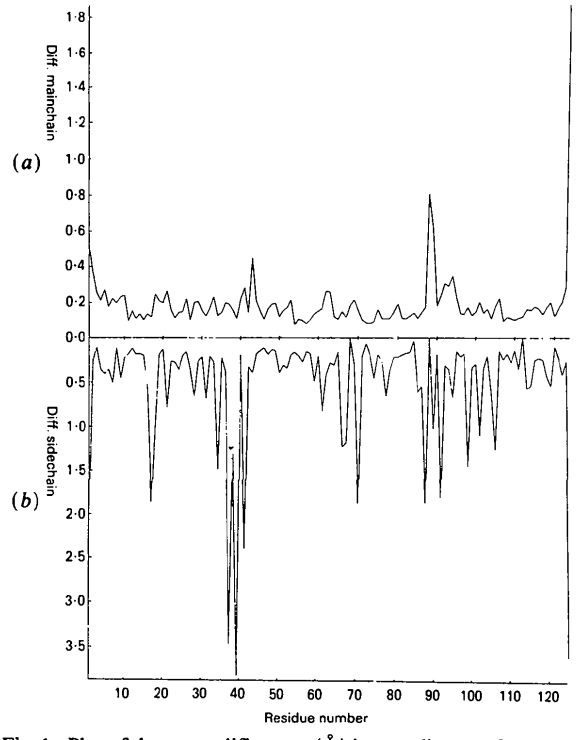

Fig. 1. Plot of the r.m.s. difference (Å) in coordinates after translation and rotation for the W and L structures against residue number. (*a*) Main-chain atoms only. (*b*) Side-chain atoms only.

Table 2. *Root-mean-square differences* (Å) *between London and Washington protein coordinates*

Figures in parentheses are mean differences (Å)

	Protein atoms including outliers	Protein atoms minus outliers	Numbers of outliers
Main chain + side chain atoms	0·799 (0·374)	0·357 (0·247)	44
Main chain only	0·223 (0·173)	0·191 (0·166)	5
Side groups only	1·111 (0·596)	0·484 (0·341)	39

Table 3. *Analysis-of-variance comparison between the main-chain atomic coordinates of the L and W structures*

Source of variation	Sum of squared coordinate difference (Å^2)	Degrees of freedom	Mean sum of squares (Å^2)
Translation	35·53	3	11·84
Rotation	0·68	3	0·22
Residual	18·00	1470	0·012
Total variation	54·21	1476	0·037

and the results are displayed in Table 3. In this analysis outliers (see *Discussion*) were removed from the coordinate differences so as to exclude as far as possible effects due to differences of interpretation. The total sum of squared differences between the coordinates of the 492 pairs of atoms is 54·2 Å^2, of which 35·5 Å^2 is accounted for by a relative translation of the two main chains, in a direction perpendicular to the unique polar direction in the crystals. The relative rotation of the two main chains is less than 0·5°. After translation and rotation the residual r.m.s. coordinate difference is 0·110 Å. This residual is a measure of the random errors in the main-chain models and also of any conformational differences between the two main chains. It may not take into account systematic errors present in both structures (*e.g.* the same incorrect target distance used in restraining both structures) but it does enable the calculation of an upper limit for the r.m.s. random error in the coordinates of the two main chains. If the L and W main chain are assumed to be identical and errors are assumed to be equally distributed between the two structures then the e.s.d. of the main-chain atomic coordinates is 0·08 Å. The other extreme assumption is that the variation is entirely due to genuine differences between the two structures. In this case the r.m.s. difference between the two main chains is 0·11 Å. This figure shows that the main-chain conformations are generally very similar.

Further evidence for a high degree of similarity between the L and W coordinates was provided by refining the W coordinate set without the waters with the L reflection data and performing a similar refinement on the L coordinates, again without the

waters, also with the L reflection data, using the program *RESTRAIN*. The crystallographic R factors obtained from these refinements were $R = 0{\cdot}274$ for the W structure and $R = 0{\cdot}273$ for the L structure. Both sets of data give very similar agreement with the L data, which is to be expected from the small degree of difference found from the comparison of the coordinates. The agreement found is good, considering that both structures were derived from X-ray only and X-ray and neutron data for the L and W structure respectively and refined using different algorithms. The agreement leads to the conclusion that both models are good solutions to two very similar structures.

The comparison of the main-chain torsion angles ψ and φ is shown in Fig. 2. The largest difference was found for the ψ angle of residue 89, which was of opposite sign in the two structures. As observed by the W group in their investigation of a complex between RNase and a transition-state analogue (Wlodawer, Miller & Sjölin, 1983) this peptide has to be flipped, even though both interpretations agreed with the electron density map. While this action did not improve the R factor, some of the dihedral angles became closer to ideality and it was concluded that the original orientation of this peptide in the W model was incorrect. The comparison of the torsion angles of the L and W models reinforced this result.

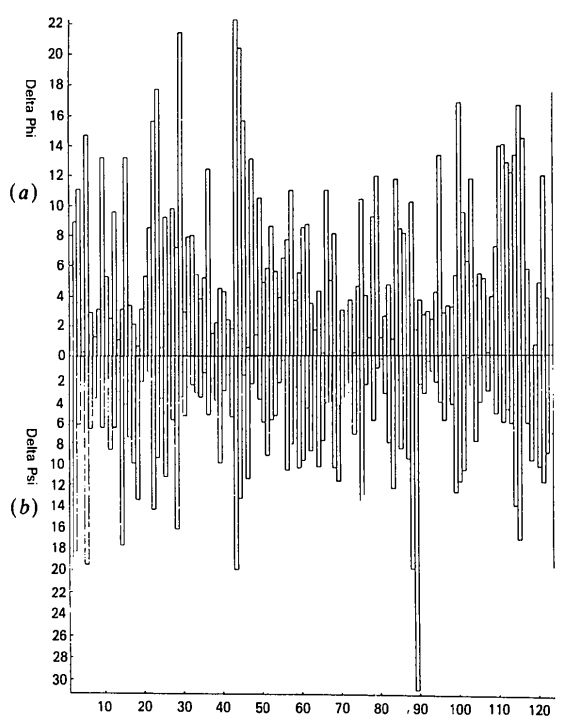

Fig. 2. Histogram of the difference in torsion angles between the W and L structures against residue number (in degrees). (*a*) Difference in φ. (*b*) Difference in ψ.

In spite of the good overall agreement in the protein conformations of the two models, there are two significant differences in the active-site area of the two models. In view of the catalytic importance of the residues involved (Lys 41 and His 119), the arguments used in interpreting these areas are presented.

(i) *Lys* 41. The starting position of Lys 41 was not kept in either refinement. As reported by Borkakoti *et al.* (1982), the density for the side chain was only well defined as far as Cγ. The partially visible side-chain density in the L map was used as a guide to the rebuilding of the side chain. In the final L model Nζ is 5 Å from the initial position. An attempt by

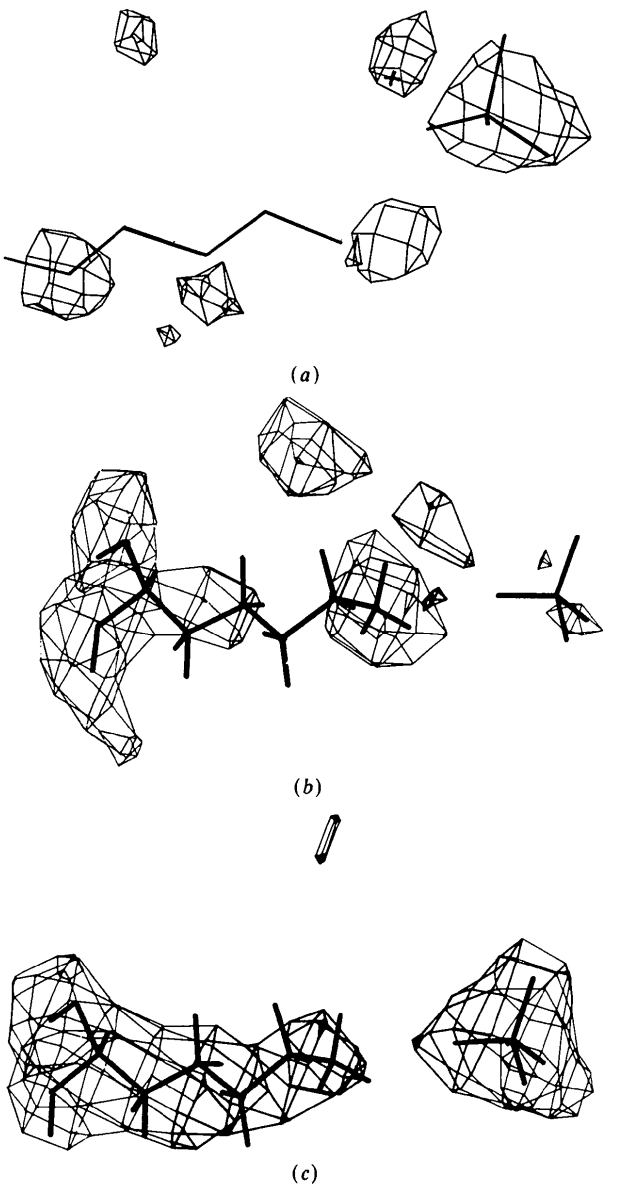

Fig. 3. $2F_o-F_c$ difference Fourier maps for Lys 41, in the extended conformation, excluding the coordinates of the side chain from the calculation. (*a*) X-ray data, London. (*b*) Neutron data, Washington. (*c*) X-ray data, Washington.

the L group to position the terminal Nζ atom in the density occupied by a solvent molecule resulted in a fully extended chain with most atoms lying outside electron density (Fig. 3*a*).

Figs. 3(*b*) and 3(*c*) show the same region in the W X-ray and neutron maps. The strong neutron scattering density at Nζ was consistent with diffraction from an ND_3^+ group and this evidence led to the fully extended conformation in the W model ending in density occupied by a water molecule in the L model. The final X-ray map is consistent with this interpretation.

It is apparent that Lys 41 is less well ordered in the L native structure than in the W structure. The W map shows clearer density in a similar position to that found for Lys 41 in inhibitor complexes of the L structure (Borkakoti, 1983).

(ii) *His* 119. The electron density for the side chain of His 119 in the L map is shown in Fig. 4(*a*) where atoms beyond Cβ had been omitted from the calculated phases. This residue also shows more disorder in the L map and has been interpreted in terms of multiple sites by the L group. Extra density also exists in the W map (Fig. 4*b*), which has been interpreted in terms of a solvent molecule. The distance between the oxygen of this solvent and Cβ is 2·8 Å, less than previously observed for a hydrogen bond.

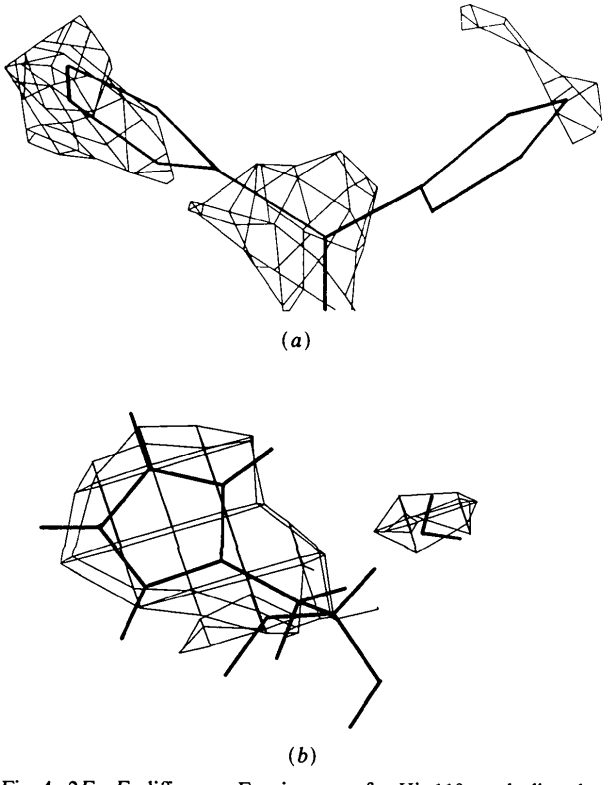

Fig. 4. $2F_o-F_c$ difference Fourier maps for His 119, excluding the coordinates of the side chain from the calculation. (*a*) X-ray data, London, showing alternative site for His 119. (*b*) X-ray data, Washington, showing D_2O site.

(*b*) *Temperature factors*

Temperature factors calculated as $U_{iso} = B/8\pi^2$ Å^2 for individual atoms in the protein molecule were on the average much higher in the L structure than in the W structure. This reflected the higher resolution of the L data than the W data (temperature factors becoming increasingly positive as the resolution of the data increases). In the L structure each atom was allowed to find its own individual temperature factor while in the W structure the temperature factors of neighbouring atoms were coupled. Hence no direct comparison of the observed temperature factors is possible but the trends of both the L and the W temperature factors for the same atoms in both structures should indicate regions of agreement in the two models.

Fig. 5(*a*) shows a plot of the average temperature factor taken from both structures for the main-chain atoms only. High average values are found for atoms at either end of the main chain, *i.e.* Lys 1 and Val 124. Residues which are externally situated on the molecule and lacking in H-bond contacts would also be expected to show high average values and this is indeed the case for the main chains of Ser 21, Lys 37, Asn 67, Lys 91, Asn 94 and Asn 113. Thr 36 also shows large atomic displacements. Disagreement between the L and W structures occurred at Glu 49 and Ala 52 where the L structure showed much larger displace-

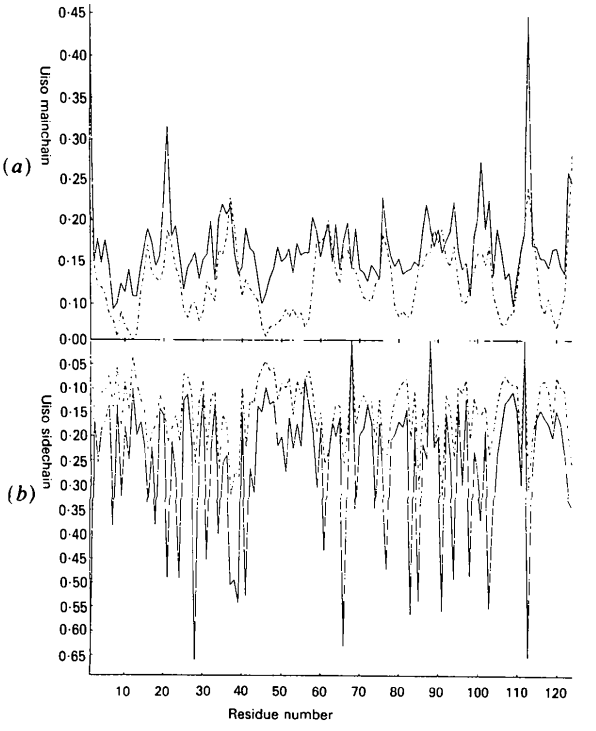

Fig. 5. Comparison of the isotropic temperature factors, averaged over each residue for the W and L structures for (*a*) main-chain atoms, (*b*) side-chain atoms, where (---) = Washington and (——) = London.

ments. It is interesting to note that the agreement on high average temperature factors is found for main-chain atoms in residues with relatively large 'mobile' side chains, *e.g.* Lys, where a greater degree of thermal disorder is to be expected. The main-chain displacements in Fig. 5(*a*) are relatively small indicating a reasonable degree of order, confirming the restricted mobility and 'buried' nature of most of these atoms.

Average U_{iso} values for the side-chain atoms (Fig. 5*b*) are much higher than the main-chain atoms, indicating the influence of the greater flexibility of the exposed side chains. High average values are found for Lys 1, Thr 17, Ser 18, Ser 21, Ser 24 and residues 37–41. High temperature factors were present in Lys 41 in both models even though the atomic positions were different. Residues that are externally situated, Lys 61, Lys 66, Gln 69, Gln 74, Tyr 76, Ser 77, Lys 91, Asn 94, Asn 113 and Val 124, also have high average values in both models. Groups with 'mobile' side chains such as Lys 7, Glu 9, Gln 28, Lys 31, Asp 83 and Arg 85 show better agreement than that found for high temperature factors in the main chain.

Owing to the restraints employed in the W refinement, the range of values spanned by the W structure is not as great as that of the L structure. However, the comparison between the two sets of data, measured by different methods and computed by different algorithms, is such that generalizations can be drawn on the relative motions of the atoms in the protein.

(*c*) *Solvent structure*

The structure of the solvent (other than the sulfate present in the active site) was determined completely independently in the W and L refinements, based on the evaluation of difference Fourier maps. The number of solvent peaks (all assumed to be D_2O) assigned in the W structure was 128, while 123 H_2O molecules were placed in the L structure. While both the oxygens and deuteriums were placed in the W structure, in the following discussion we will consider only the positions, temperature factors and occupancies of the oxygens, disregarding the deuteriums completely.

The ordered list of all solvents from the W structure with their L equivalents is shown in Table 4. Of all the solvents whose coordinates were determined, 58 were found to be in close proximity (distance less than 1 Å) in the two structures. In addition, seven solvents were displaced by between 1 and 1·5 Å and clearly occupied the same electron density, while two other similar solvents were displaced between 1·5 and 1·9 Å. Thus just over half of the solvents were found to be in similar positions (and making similar contacts with the atoms belonging to the protein), while the other half of the positions determined in each structure did not match.

Table 4. *Comparison of the solvent positions in the Washington and London models of RNase*

Solvent positions found in the symmetry-related molecule are marked with ≠. W solvent is listed first and is ordered on the basis of U_{iso}/occupancy. Solvents found closer than 3·2 Å from an oxygen or nitrogen belonging to the protein are considered to be in the first shell ($s = 1$), those within 3·2 Å of the first-shell solvent are assigned to the second shell, and all the others to the third shell.

D_2O	U_{iso}	s	H_2O	Distance	D_2O	U_{iso}	s	H_2O	Distance
166	0·065	1	218	0·39	242	0·426	1		
189	0·063	1	326	0·65	216	0·430	1	239	1·00
200	0·091	1	250	0·28	138	0·377	1	255	0·83
184	0·111	1	212	0·31	195	0·418	1	322	0·64
167	0·136	1	244	0·24	233	0·435	1		
168	0·141	1	313	0·83	133	0·440	1		
199	0·148	1	243	0·34	178	0·446	2	231	0·62
180	0·153	1	290	0·36	191	0·397	1		
160	0·159	1	202	0·31	161	0·301	1		
190	0·137	1	251	0·36	163	0·467	2		
142	0·173	1	207	0·09	153	0·476	3		
294	0·162	1			171	0·278	1	317	0·33
169	0·156	1			267	0·479	1		
181	0·173	1	≠ 244	0·23	205	0·481	1		
183	0·137	1	242	0·47	249	0·383	1		
172	0·182	1	214	0·50	218	0·482	1		
175	0·070	1	295	0·17	185	0·485	1	332	0·81
186	0·183	1	327	0·61	210	0·486	1	265	0·33
271	0·193	1	292	0·20	212	0·420	1	279	0·45
239	0·165	1	336	0·55	198	0·496	2	252	0·46
179	0·212	1	≠ 213	0·35	140	0·418	1		
203	0·212	1			296	0·444	1	319	0·63
194	0·159	1	330	0·75	320	0·421	1	325	0·24
182	0·214	1	299	0·38	251	0·461	1		
187	0·230	1	216	0·40	301	0·520	2		
211	0·182	1	296	0·18	149	0·526	1		
192	0·261	2	280	0·60	297	0·266	1		
206	0·200	1	276	0·33	306	0·538	1	≠ 228	1·44
165	0·265	2	245	0·26	217	0·496	1	≠ 235	1·63
321	0·265	1	318	0·58	154	0·472	1		
214	0·185	1	307	0·44	170	0·331	1		
145	0·279	1	289	0·28	202	0·557	1	215	1·15
196	0·278	1	266	0·22	303	0·557	1		
243	0·286	1	337	0·35	155	0·148	2		
228	0·293	1	278	0·23	254	0·561	1		
302	0·208	1			150	0·417	2		
164	0·298	1			208	0·574	2		
238	0·321	1			244	0·576	1		
270	0·326	1			151	0·583	1	≠ 223	0·32
223	0·284	1	309	0·04	286	0·585	1	271	0·41
316	0·167	2			177	0·589	1		
126	0·100	2			137	0·590	3		
246	0·171	1	310	0·38	261	0·600	2		
258	0·136	1			221	0·606	1	294	0·45
248	0·198	1			252	0·607	1	211	1·08
213	0·348	1	267	0·40	136	0·478	1	334	0·41
222	0·349	1	237	0·36	275	0·556	1		
253	0·354	3			255	0·647	1		
219	0·181	3			284	0·647	2	328	0·94
197	0·312	1	236	0·76	278	0·450	1	287	1·20
131	0·222	3			162	0·662	1	311	0·51
135	0·262	1	≠ 257	0·32	319	0·339	1		
313	0·215	1			143	0·621	1		
209	0·377	1	≠ 248	0·50	129	0·349	1		
207	0·210	2	314	0·50	235	0·662	1		
158	0·355	1	312	0·93	229	0·685	1		
240	0·386	1			268	0·237	1		
201	0·373	2			147	0·410	3		
188	0·390	1	229	0·69	250	0·800	2		
220	0·183	1			141	0·625	2	258	0·94
276	0·186	1			260	0·614	1	≠ 269	1·20
128	0·403	2			215	0·545	1		
157	0·407	1	232	0·69	156	0·519	1		
287	0·407	1	293	1·02	259	0·595	2		

With the aim of finding a correlation between the presence of equivalent solvents in both structures and the confidence that could be placed in them, we have ordered the solvents on the basis of the temperature factors divided by occupancy (W structure) or temperature factors only (L structure). It was immediately

obvious that the solvents with the lowest temperature factors were most likely to be in common between the two structures. In the W structure, 79% of the one-third of solvents with the lowest effective *B* were common, while 69% of those found in the first half but only 37% of those found in the second half were common. The r.m.s. deviation for the 34 common solvents found in the top third was 0·41 Å. In the L structure, 83% of the top third, 76% of the top half, and 33% of the bottom half were found to be common.

While the temperature factors were a clear guide in the comparison, other properties of the solvents were found to be important. We have divided the solvents according to their distance from the potential hydrogen donors and acceptors in the proteins. Solvent found within 3·2 Å of such atoms were considered to be first shell; those further than 3·2 Å from the protein but closer than 3·2 Å to first-shell water molecules were considered to belong to the second shell, while the remaining waters were placed in the third category. Of all the solvent in the W structure, 101 belonged to class 1, 22 to class 2, and 5 to class 3. For the first-shell solvents, 60% were common, while 32% of the second shell and none of the third class were common with the L structure. It is very likely that the solvents belonging to the third category were not real.

The distribution of solvents into these three classes is very similar for the L structure. Of the 123 water molecules placed in the model, 87 are found in the first shell, 26 in the second shell, and ten belong to the third category. For the first shell, 59 (68%) are common, for the second shell seven (27%) are common and only one common solvent belongs to the third class. The equivalent of this H_2O 258 is D_2O 198, which is found within the second shell of hydration in the W structure. Of the seven equivalents of the second-shell waters, five also belong to the second shell in the W data, while two others (D_2O 195 and 278) are considered in that case to belong to the first shell.

In order to pinpoint the differences between the individual assignments of the solvents found in each structure, we have examined their positions with respect to the $(2F_o-F_c)$ Fourier maps of the other structure. This comparison enabled us to pinpoint several reasons for the disagreement. D_2O 294 was found to be occupying a clear density in the L map, but that density was claimed by the Nζ of Lys 41 in the W map. The electron density for Lys 41 was not defined beyond Cε in the L map. D_2O 169 was about 2 Å from the symmetry mate of H_2O 285, but the latter solvent has a high temperature factor and was only 1·2 Å from the Oγ 21 and its position was clearly suspect. Seven other positions corresponding to D_2O molecules with the lowest temperature factors (203, 302, 164, 238, 270, 316 and 126) did not correspond to any density in the L map. D_2O 126 was not refining properly and is quite possibly an artifact, while the rest of these solvents may represent true differences between the structures. The correlation is much less clear for those solvent molecules which had even higher temperature factors, and it was usually impossible to make a clearcut decision about which of them represented true differences and which were due to misassignments. One interesting difference is provided by D_2O 205; this solvent occupies density assigned in the L model to the alternate position of His 119. It is quite clear that this region of the unit cell has some electron density in both crystal forms, and only the interpretation of its meaning was different.

Only five of the top third of the H_2O positions determined in the L model did not have their counterparts in the W model. Four of these solvents occupied areas free of density in both the X-ray and neutron difference Fourier maps, while one (H_2O 339) was close to, but not in, a small density. Again, it is possible that these positions represent true differences between the two structures. On the other hand, the positions in the lowest third of the solvents agreed very poorly, with only eight common sites. The difference of interpretation could be seen in some poorly delineated solvent clusters, such as the one containing H_2O 228, 229 and 230. Only two solvents (D_2O 188 and the symmetry mate of 306) were occupying this area in the W model, and this made the agreement appear even poorer than it was in reality. It is clear that some solvent is present at this site, but its exact location remains by no means unambiguous.

This comparison shows convincingly that the strongly bound solvent occupies the same sites regardless of the nature of the alcohol in the mother liquor or the exchange of D_2O for H_2O. It is also clear that the positions of these solvent molecules can be determined at the resolution and the state of refinement of the present studies with a high degree of probability. On the other hand, the comparisons have shown a few solvent sites which were clearly misassigned in each of the studies, and a small number of sites where the positions assigned to the solvent were different because of the different interpretation of the protein in that area. The large majority of the sites which were assigned as solvent in one study but not in the other simply had no corresponding density in the opposite difference Fourier maps and thus may have represented true differences in the solvent structure. It is interesting to note that the number of common solvent sites was similar to the 40 waters assigned in the very conservative neutron studies of carbonmonoxymyoglobin (Schoenborn & Hanson, 1980), but much smaller than the 316 solvents placed in the X-ray studies of oxymyoglobin (Phillips, 1980).

Another interesting point emerging from the comparison is that while almost all of the best determined

waters were in the first shell of hydration, seven common solvent positions were found in the second shell. It appears that not only the individual solvent positions but also some elements of the solvent network persist during the change of crystal environment. The relative sparsity of such common sites indicates, though, that the coordinates of the second-shell solvent should be taken with a large degree of caution, especially if used as a basis for theoretical calculations.

Discussion

There are five identifiable causes of the observed differences between the L and W coordinates.

(1) Errors present in the W model led to an erroneous interpretation of the main chain at residue 89. This error has already been reported by the authors (Wlodawer *et al.*, 1983).

(2) Some differences are due to uncertainties in electron-density-map interpretation such as in the orientation of the terminal groups such as Gln 101 and His 105 in the L model. In these cases the L and W electron density maps were similar and the ambiguities were resolved by reference to the neutron scattering density maps.

(3) There are differences of interpretation based on actual differences in the L and W maps. The differences in the L and W coordinates in the side chains of Lys 41 and in three Thr residues (17, 70 and 87) fall into this category. Other such differences between the models are due to the use of deuteration in the W crystals. Some of the better defined solvent molecules in each model had no counterpart in the other model and probably represent genuine differences between the two structures.

(4) Some differences in the two models arise from arbitrary assignments of atomic positions where there was weak density in all three maps. The side chain of Lys 37 illustrates such a case.

(5) A few differences arise from different interpretations of similar electron density. This is illustrated by the positioning of a solvent molecule in the W model in density occupied by the minor site of His 119 in the L model.

The root-mean square and mean differences between the L and W models are shown in Table 2. The second column of figures in this table is for comparisons where atoms involved in categories 1 to 5 above were omitted. This exclusion of outliers from these latter figures means that they give a better indication of the random differences between the two models. The five outlying atoms in the main chain occurred in the two terminal residues and in Gly 88 and Ser 89. It is interesting to note how the root-mean-square differences are more sensitive to outlying values than are the mean differences.

Where the positioning of atoms in density maps was problematical, both the W and L groups removed such atoms from the refinement before producing further maps so that there would be no bias towards a possibly erroneous interpretation. The use of different least-squares procedures has not resulted in significant differences between the two models.

Two of us (NB and BH) would like to thank the SERC for financial support during the course of this work. A NATO grant RG 040.80 is gratefully acknowledged.

References

BODE, W., SCHWAGER, P. & HUBER, R. (1976). *Proteolysis and Physiological Regulation*, Miami Winter Symposia, Vol. 11, pp. 43-76. New York: Academic Press.

BORKAKOTI, N. (1983). *Eur. J. Biochem.* **132**, 89-94.

BORKAKOTI, N., MOSS, D. S. & PALMER, R. A. (1982). *Acta Cryst.* B**38**, 2210-2217.

BORKAKOTI, N., MOSS, D. S., STANFORD, M. & PALMER, R. A. (1984). *J. Spectrosc. Crystallogr. Res.* **14**, 467-494.

CARLISLE, C. H., PALMER, R. A., MAZUMDAR, S. K., GORINSKY, B. A. & YEATES, D. G. R. (1974). *J. Mol. Biol.* **85**, 1-18.

CHAMBERS, J. L. & STROUD, R. M. (1977). *Acta Cryst.* B**33**, 1824-1837.

CHAMBERS, J. L. & STROUD, R. M. (1979). *Acta Cryst.* B**35**, 1861-1874.

CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 154-157.

DEREWENDA, Z. S., BRZOZOWSKI, A. M., STEPIEN, A. & GRABOWSKI, M. J. (1982). *Acta Cryst.* A**38**, 432-438.

DIAMOND, R. (1981). In *Biomolecular Structure, Conformation, Function and Evolution*, edited by R. SRINIVASAN, Vol 1, pp. 567-588. New York: Pergamon Press.

HANEEF, I., MOSS, D. S., STANFORD, M. & BORKAKOTI, N. (1985). *Acta Cryst.* A**41**, 426-433.

HENDRICKSON, W. A. (1979). *Acta Cryst.* A**35**, 158-163.

HENDRICKSON, W. A. & KONNERT, J. H. (1981). In *Biomolecular Structure, Conformation, Function and Evolution*, edited by R. SRINIVASAN, Vol. 1, p. 43. New York: Pergamon Press.

JONES, T. A. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 303-317. Oxford: Clarendon Press.

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

MARQUART, M., WALTER, J., DEISENHOFER, J., BODE, W. & HUBER, R. (1983). *Acta Cryst.* B**39**, 480-490.

NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351-359.

PHILLIPS, S. E. V. (1980). *J. Mol. Biol.* **142**(4), 531-554.

POWERS, T. B. (1976). PhD thesis, Yale Univ.

PRINCE, E., WLODAWER, A. & SANTORO, A. (1978). *J. Appl. Cryst.* **11**, 173-178.

SCHOENBORN, B. P. & HANSON, J. C. (1980). *Am. Chem. Soc. Symp. Ser.* (*Water Polym.*), **127**, 215-224.

WLODAWER, A. (1980). *Acta Cryst.* B**36**, 1826-1831.

WLODAWER, A., BOTT, R. & SJÖLIN, L. (1982). *J. Biol. Chem.* **257**(3), 1325-1332.

WLODAWER, A. & HENDRICKSON, W. A. (1982). *Acta Cryst.* A**38**, 239-247.

WLODAWER, A., MILLER, M. & SJÖLIN, L. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 3628-3631.

WLODAWER, A. & SJÖLIN, L. (1983). *Biochemistry*, **22**(11), 2720-2728.