Acta Cryst. (1996). D52, 327–344

Ribonuclease from *Streptomyces aureofaciens* at Atomic Resolution

Jozef Sevcik,[a] Zbigniew Dauter,[b] Victor S. Lamzin[b] and Keith S. Wilson[b]

[a]*Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta, 842 51 Bratislava, Slovak Republic, and* [b]*European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany*

(Received 3 January 1995; accepted 12 June 1995)

Abstract

Crystals of ribonuclease from *Streptomyces aureofaciens* diffract to atomic resolution at room temperature. Using synchrotron radiation and an imaging-plate scanner, X-ray data have been recorded to 1.20 Å resolution from a crystal of native enzyme and to 1.15 Å from a crystal of a complex with guanosine-2′-monophosphate. Refinement with anisotropic atomic temperature factors resulted in increased accuracy of the structure. The *R* factors for the two structures are 10.6 and 10.9%. The estimated r.m.s. error in the coordinates is 0.05 Å, less than half that obtained in the previous analysis at 1.7 Å resolution. For the well ordered part of the main chain the error falls to below 0.02 Å as estimated from inversion of the least-squares matrix. The two independent molecules in the asymmetric unit allowed detailed analysis of peptide planarity and some torsion angles. The high accuracy of the analysis revealed density for a partially occupied anion in the nucleotide binding site of molecule *A* in the native structure which was not seen at lower resolution. The anisotropic model allowed correction of the identity of the residue at position 72 from cysteine to threonine. Cys72 SG had been modelled in previous analyses with two conformations. The solvent structure was modelled by means of an automated procedure employing a set of objective criteria. The solvent structure for models refined using different programs with isotropic and anisotropic description of thermal motion is compared.

1. Abbreviations

RNase Sa, ribonuclease from *Streptomyces aureofaciens*; ARP, automated refinement procedure; 2′,3′-GCPT, guanosine-2′,3′-cyclophosphorothioate; 2′-GMP, guanosine-2′-monophosphate; 3′-GMP, guanosine-3′-monophosphate; MKH, model of the native enzyme refined using *PROLSQ* with isotropic atomic temperature factors; MSI, model of the native enzyme refined using *SHELXL*93 with isotropic atomic temperature factors; MSA, model of the native enzyme refined using *SHELXL*93 with anisotropic atomic temperature factors; MGMP, model of the 2′-GMP complex refined using *SHELXL*93 with anisotropic atomic temperature factors.

2. Introduction

For small molecules there are sufficient X-ray data, typically to 1.0 Å resolution or better, to refine the atomic parameters against these data alone. For an anisotropic atomic model, for each atom there are three positional and six thermal parameters to be refined. Nevertheless, at 1.0 Å the X-ray observation:parameter ratio is about 5:1 for a non-centrosymmetric structure. This excess is quite sufficient to define a good least-squares minimum. The situation is quite different for proteins where the resolution is generally limited to less than atomic and in addition the crystal contains about 50% aqueous solvent.

A severe complication in protein crystallography for both structure solution and refinement is the limited number of X-ray data. At less than atomic resolution, this necessitates additional restraints effectively to increase the number of observations. The X-ray data are therefore complemented by, for example, stereochemical restraints based on known structures of small molecules (Engh & Huber, 1991). The latter, however, may not be completely valid for proteins (Lamzin, Dauter & Wilson, 1995) and one aim of the present series of atomic resolution protein structures at EMBL is the derivation of a library based on real proteins. For the ordered regions of the present structure it should, in principle, be possible to carry out completely unrestrained refinement. However, the information from the X-ray terms alone is not enough to define residues or side chains which are poorly ordered or present in multiple conformations, as is also found for small-molecule structures and weak restraints are therefore required.

Proteins play major roles in the structure and especially in the function of living organisms. To do so they must often interact with other small or large molecules. Elucidation of the mechanism of interactions based on knowledge of accurate three-dimensional structures provides clues for better understanding of chemical reactions in living organisms and forms a basis for constructing enzymes with modified properties, specific inhibitors, drugs and vaccines. This has clear applications in many fields such as medicine, biotechnology and the food industry.

Ribonuclease from *Streptomyces aureofaciens*, a bacterium which is an industrial source of chlorotetracycline, is a guanylate endoribonuclease (RNase Sa, E.C. 3.1.4.8) which highly specifically hydrolyses the phosphodiester bonds of RNA at the 3′ side of guanosine nucleotides. RNase Sa belongs to the prokaryotic subgroup of the microbial ribonuclease family. RNase Sa is active with RNA as a substrate in the pH range from 4.7 to 9.3 with a maximum at pH 7.0. The isoelectric point is pH 4.3 (Zelinkova, Bacova & Zelinka, 1971; Bacova, Zelinkova & Zelinka, 1971). The molecule consists of 96 amino-acid residues (Shlyapnikov *et al.*, 1986) and its molecular weight is 10.5 kDa.

The crystal structures of native RNase Sa and its complex with 3′-GMP were previously determined by multiple isomorphous replacement and refined against 1.8 Å synchrotron data collected on film to *R* factors of 17.2 and 17.5%, respectively (Sevcik, Dodson & Dodson, 1991). Refinement of RNase Sa using more accurate 1.8 Å synchrotron data collected with an imaging-plate scanner ($R_{\rm merge}$ = 3.3% compared to film data for which $R_{\rm merge}$ is 5.6%) and refinement of the complex with 2′-GMP against 1.7 Å data ($R_{\rm merge}$ = 3.3%) converged with *R* factors of 13.9 and 13.2%, respectively (Sevcik, Hill, Dauter & Wilson, 1993). This second refinement of native RNase Sa was carried out for a proper comparison with the 2′-GMP complex structure. The major differences between the two models of native RNase Sa were in the accuracy of the structures and the number of solvent molecules. The effect of these water molecules was especially reflected in a drop of *R* factor in the low-resolution shells. Improvement of the whole model was reflected particularly in clearer electron density for poorly ordered residues.

The structures of RNase Sa and its complexes with 2′-GMP and 3′-GMP confirmed the basis of guanine specificity of the enzyme. The base of the nucleotide is bound to the main-chain NH groups of residues Gln38, Asn39 and Arg40 and to the side-chain atoms OE1 and OE2 of Glu41. Asn39 has an important role in maintaining the conformation of the main-chain loop which binds the base. The structure of the complex with *exo*-guanosine-2′,3′-cyclophosphorothioate (2′,3′-GCPT), a thio-analogue of the intermediate of the two-step reaction, was refined at 2.0 Å (*R* factor is 11.9%) resolution with synchrotron data (Sevcik, Zegers, Wyns, Dauter & Wilson, 1993). The ligand was bound in a non-functional mode. Nevertheless, modelling of a cyclic intermediate which mimics the substrate in the second step of the reaction catalysed by the enzyme through residues Glu54 and His85 agrees with the hypothesis of the enzyme's mechanism. Asn39, Glu41, Glu54, Arg69 and His85 are fully conserved in the sequences of all microbial ribonucleases. The overall sequence similarity is relatively low in spite of the highly conserved structural core in all of these enzymes (Sevcik, Sanishvili, Pavlovsky & Polyakov, 1990).

Studies of ribonucleases as model systems (small size, solubility, stability, accessibility) have made a major contribution to understanding the structure, function and stability of enzymes. This RNase Sa is one of the few protein structures refined to atomic resolution. It is important for critical evaluation of previously obtained results and contributes to general knowledge of protein structures. The main aim of this work is to provide improvements in the protein and solvent model and in addition to compare results from different refinement procedures. The high accuracy of the structure and modelling of the solvent based on objective criteria are the most important results.

3. Methods

3.1. *Crystallization*

RNase Sa crystals have been obtained by vapour diffusion from solutions of 3.0% protein by weight in 0.1 *M* phosphate buffer at pH 7.2 at room temperature. 22% saturated ammonium sulfate was used as precipitant (Sevcik, Gasperik & Zelinka, 1982) giving a final pH for the mother liquor of 6.7. Crystallization experiments at other pH were unsuccessful. The space group is $P2_12_12_1$. There are two enzyme molecules in the asymmetric unit, referred to throughout as molecules *A* and *B*. As in previous experiments the complex of RNase Sa was prepared by diffusion of 2′-GMP into native crystals in mother liquor. The soaking time in the presence of 2′-GMP at a concentration of about 20 mg ml^{-1} was 2 d.

3.2. *Data collection*

Data were collected at room temperature from a single crystal of native enzyme and two crystals of the complex with 2′-GMP (the size of the crystals was approximately 1.2 × 0.6 × 0.4 mm) on the EMBL X11 beamline at the DORIS storage ring, DESY, Hamburg, with an MAR Research (Hamburg) imaging-plate scanner. Three sets of data were collected with different exposures: to 1.2 (1.15), 1.6 (1.75) and 2.5 (2.8) Å resolution (numbers in parentheses refer to that of the crystal of the complex) with oscillation angle per image of 0.75, 1.0 and 1.5°, respectively. The low-resolution sets for the complex were collected from a second crystal from the same batch on the EMBL X31 beamline. The two crystals appeared to be not entirely isomorphous and this may have negatively influenced the quality of the data. Native data were processed using the *MOSFLM* package (Leslie, 1992) as described in Dauter, Terry, Witzel & Wilson (1992) and the complex data with *DENZO* (Otwinowski, 1993). A summary of data collection and processing is given in Table 1. The percentage completeness of the data is shown in Fig. 1. The native crystal was oriented so as to minimize the fraction of the unique reflections in the blind region, but the complex crystal was oriented with its *c* axis too closely aligned with the rotation axis which resulted in the loss of some reflections, especially at high resolution. The merging *R* factor for symmetry-equivalent reflections as a function

Table 1. *Data collection and processing*

	Native	Complex
No. of crystals	1	2
Beamline	X11	X11 + X31
Wavelength (Å)	0.92	0.92
Maximum resolution (Å)	1.20	1.15
No. of measurements	246637	445145
Independent reflections	60670	62845
Overall completeness (%)	95.3	87.3
$R(I)_{\rm merge} = \sum \lvert I - \langle I \rangle \rvert / \sum I(\%)$	3.9	6.6

of resolution is shown in Fig. 2. The overall temperature factor estimated from the Wilson plot (Wilson, 1942) is 10.8 and 11.2 Å^2 for native enzyme and complex.

The merging of native data produced a smooth curve, whereas for the complex the curve shows a maximum corresponding to the region where weak intensities from the medium-resolution data were merged with strong data from the high-resolution set, Fig. 2. The difference in scale of high- and medium-resolution images was about 100. To confirm that this effect was not because of non-isomorphism of the two crystals, a statistical test was applied. The three data sets, the previous 1.7 Å complex (Sevcik *et al.*, 1993), the high-resolution data from beamline X11 and separately processed medium- and low-resolution data from the second crystal measured on the X31 beamline were scaled in pairs and the normalized $\chi^2 = (1/N)\sum[(F_1 - F_2)^2/(\sigma_1^2 + \sigma_2^2)]$ (goodness of fit) was calculated. The new high- and low-resolution

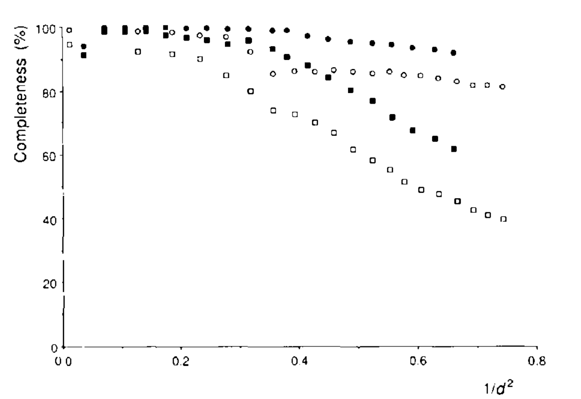


Fig. 1. Percentage completeness of the data. All reflections (circles) and those with $I > 3\sigma(I)$ (squares) as a percentage of the total theoretical number of reflections as a function of resolution for native (filled points) and 2′-GMP complex data (open points).

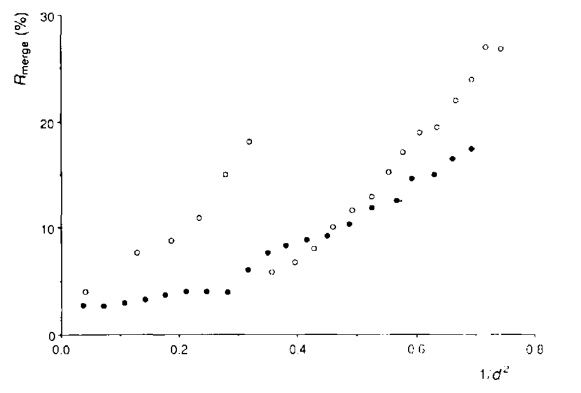


Fig. 2. The merging $R(I)$ factor for symmetry-related reflections, $R = \sum\lvert I - \langle I \rangle\rvert / \sum I$, as a function of resolution for native (filled circles) and complex data (open circles).

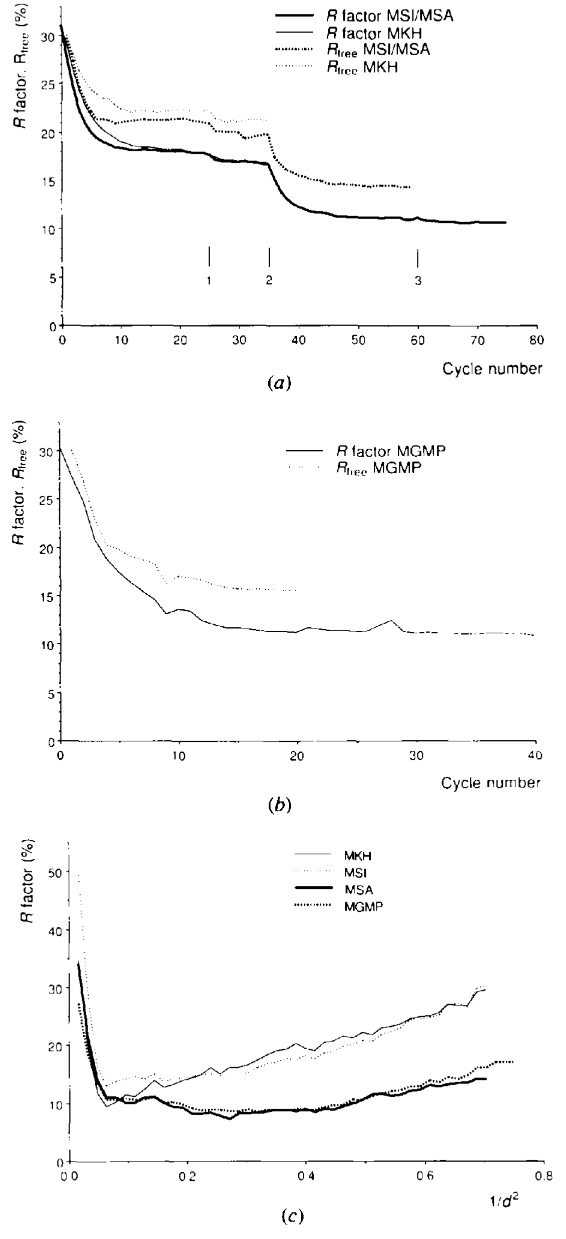


Fig. 3. (*a*) R factor (95% of data) and $R_{\rm free}$ (5% of data) for the refinement of the native structure with *PROLSQ* and *SHELXL93*. The main steps are shown: inclusion of H atoms (1), refinement with anisotropic atomic temperature factors (2) and refinement against all data (3). (*b*) R and $R_{\rm free}$ for MGMP. From cycle 20 all data were used. (*c*) R factor as a function of resolution for all four models.

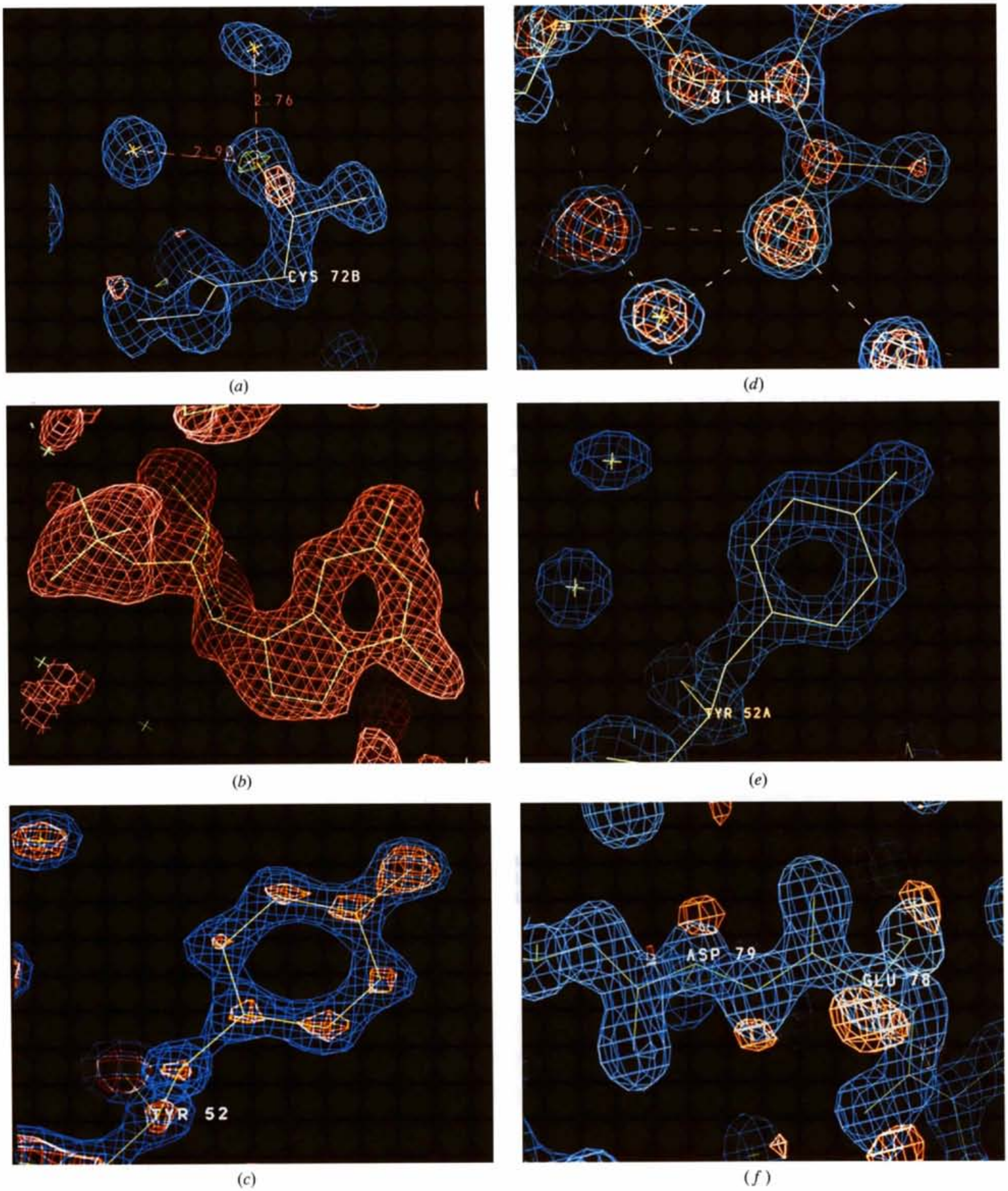


Fig. 4. (*a*) Cys72 in the $(3F_o - 2F_c, \alpha_c)$ electron-density map, the contour level is 0.9 Å e^{-3} (2σ). Peaks of positive (red) and negative (green) difference electron density contoured at 0.25 and –0.25 Å e^{-3} (3σ), respectively, indicate that the SG atom is incorrectly placed. (*b*) Active site of molecule *B*. Electron density contoured at 2σ level corresponds to the difference between the previous 1.7 Å data from a crystal, where 2′-GMP was present with an occupancy of 0.5, and the current MGMP data. This shows that in the present structure 2′-GMP the occupancy is negligible. (*c*) Tyr52 in $(3F_o - 2F_c, \alpha_c)$ electron density, blue 1.35 Å e^{-3} (3σ), red 2.25 Å e^{-3} (5σ). (*d*) Thr18 in $(3F_o - 2F_c, \alpha_c)$ electron density, blue 1.35 Å e^{-3} (3σ), red 2.25 Å e^{-3} (5σ). (*e*) Tyr52 in $(3F_o - 2F_c, \alpha_c)$ from the native structure (PDB coordinate set 1GMQ) refined to 1.8 Å at 0.9 Å e^{-3} (2σ). (*f*) Difference electron density (red) contoured at 0.25 Å e^{-3} (3σ) indicates positions of H atoms. $(3F_o - 2F_c, \alpha_c)$ density is shown in blue.

Table 2. *Refinement statistics*

		MKH		MSI	MSA	MGMP
Resolution limits (Å)		10–1.2		10–1.2	10–1.2	10–1.15
Protein atoms in molecule *A*/*B*		745/745		745/745	745/745	745/745
Water molecules		385		288	332	439
Atoms in sulfate ions		10		10	10	5
2′-GMP atoms		—		—	—	24
R factor (%)		16.7		16.7	10.6	10.9
Positional σ_A error estimate (Å)		0.15		0.13	0.05	0.08
Stereochemistry	Target		Target			
Bond length (1–2) (Å)	0.020	0.023	0.030	0.024	0.024	0.020
Angle distance (1–3) (Å)	0.040	0.039	0.030	0.025	0.031	0.029
Planar distance (1–4) (Å)	0.050	0.040	0.030	0.044	0.044	0.045
Chiral volumes (Å^3)	0.15	0.17	0.20	0.22	0.20	0.19
Planar torsion angles (°)	3.0	3.6	6.0	5.6	6.0	5.6
Deviations from planes (Å)	0.020	0.018	0.03	0.027	0.027	0.027

data gave a value of about 3.6 suggesting compatibility, whereas combinations with the previously measured 1.7 Å data gave values higher than 25. Minor differences between the two complex crystals, such as occupancy of the ligand, cannot be ruled out completely.

3.3. *Determination of accurate unit-cell parameters*

Unit-cell parameters for RNase Sa derived from data processing at atomic resolution are $a = 64.82$, $b = 78.56$, $c = 39.05$ Å for the native and $a = 64.80$, $b = 78.76$, $c = 39.13$ Å for the complex crystals. They are very similar suggesting the crystals are isomorphous. Small deviations may arise from binding the inhibitor or from errors in wavelength and crystal-to-detector distance determination, particularly when synchrotron radiation and area detectors are used.

To measure cell parameters precisely, a separate experiment was carried out. Hereafter, the standard deviations are shown in parentheses. The wavelength was calibrated to be 0.90259 (5) Å using a silicon powder on the EMBL X31 beamline. Four crystals of native RNase Sa from different crystallization batches were used. From each crystal two diffraction images orthogonal to one another were recorded between 6.0 and 1.4 Å resolution. Each image was processed separately with *DENZO*. The crystal-to-detector distance was refined. The unit-cell parameters so obtained are $a = 64.73\,(4)$, $b = 78.56\,(7)$, $c = 38.99\,(5)$ Å. The errors are about 0.1%. The new unit-cell parameters are marginally smaller than those derived during data processing. The differences are probably a consequence of using a well calibrated wavelength whereas that used in data processing was less accurately measured. The new accurate unit-cell parameters were used in the last refinement steps of the native enzyme and its complex.

3.4. *Refinement*

3.4.1. *General strategy*. Every refinement cycle consisted of least-squares minimization, Fourier-map calculation and automated updating of the model. Atomic positions and temperature factors were refined by restrained least-squares minimization using the *CCP*4 (Collaborative Computational Project, Number 4, 1994) version of *PROLSQ* (Konnert & Hendrickson, 1980) and independently using *SHELXL*93 (Sheldrick, 1993). In both programs the diagonal approximation and conjugate-gradient algorithm were used.

Refinement was carried out against 95% of the data. The remaining 5% (approximately 3000 reflections) randomly excluded from the full data were used for cross-validation in which the free *R* factor ($R_{\rm free}$) was calculated to follow the progress of refinement (Brünger, 1993). All data were included in the last refinement steps. After each refinement cycle an automated refinement procedure (ARP, Lamzin & Wilson, 1993) was applied for modelling and updating solvent structure. The sum of occupancies for each set of pairs of atoms in disordered residues was semt to unity. The fractional occupancy was adjusted manually when using *PROLSQ* but was refined when using *SHELXL*93. H-atom positions were not refined, but calculated according to established geometrical criteria. The target deviations against the stereochemical restraints are given in Table 2. The program *FRODO* (Jones, 1978) on Evans & Sutherland PS300 and ESV graphics stations was used for visualizing and rebuilding the models.

3.4.2. *Least-squares minimization*. In *PROLSQ* the atomic model is refined against structure-factor amplitudes. Temperature factors for main-chain 1–2 and 1–3 neighbours and side-chain 1–2 and 1–3 neighbours, were restrained to differ by r.m.s. values of 3, 5, 6 and 8 Å^2, respectively. Temperature factors assigned to H atoms were set equal to those of the host atom.

*SHELXL*93 refines against intensities rather than amplitudes enabling direct use of all measured observations including those with negative values. Restraints were again applied, Table 2. Although *SHELXL*93 works with U_{ij} atomic thermal parameters (where $B = 8\pi^2 U$) for convenience of comparison with conventional protein crystallography, *B* values are used throughout the present text. Bonded atoms were restrained in both isotropic and anisotropic refinement to have the same B values with an effective standard deviation of 4.0 Å^2. For terminal atoms the deviation was 8.0 Å^2. For anisotropic refinement six parameters describing thermal vibration were refined for each atom. For pairs of chemically bonded atoms, the components of the anisotropic displacement in the direction of the bond (1–2 and 1–3 neighbours) were restrained to a deviation of 0.8 Å^2. For solvent atoms a weak anti-bumping restraint was used with diameters of 3.2, 2.7, 2.6 and 3.5 Å for C, N, O and all other atoms, respectively. Temperature factors for H atoms were set to 1.2 times the value of the atoms to which they were attached (1.5 for OH and methyl groups).

3.4.3. *Automated refinement procedure.* Building of the solvent structure was performed automatically using ARP in an iterative manner. Recently introduced ARP options such as density shape analysis, merging of atoms, automatic determination of the difference electron-density threshold and real-space fit (ARP, Version 4, Lamzin & Wilson, to be published) were employed. In each cycle ARP identified and removed those water molecules which were most likely to be wrong and added new sites. At the beginning of refinement up to 30 atoms were added or removed each cycle. This was reduced to ten in the last cycles when refinement had essentially converged. All solvent sites were modelled as fully occupied. A problem with modelling solvent molecules, which are weak, not fully occupied and represent overlapping alternative hydrogen-bonded networks, remains. Such solvent molecules contribute mostly to the low-resolution region of the reciprocal lattice (Langridge *et al.*, 1960). Their parameters are not overdetermined even at atomic resolution.

Removal of solvent molecules was carried out on the basis of the $(3F_o - 2F_c, \alpha_c)$ Fourier synthesis coupled with distance constraints. An atom was considered for rejection if the electron density at its centre was lower than a defined cutoff level and the density shape around the atom was highly non-spherical. In our experience the method is not highly sensitive to the exact value assumed for this cutoff. Use of a cutoff 0.5 r.m.s. above the mean density was found to give better convergence than a value of 1.0. A merging option was used to prevent solvent atoms from approaching too closely to each other or to protein atoms. Water molecules which were closer to protein atoms than 2.0 Å were removed. This keeps intact features in the difference density in the protein region and simplifies their identification for manual rebuilding. Pairs of water molecules (which may arise from modelling a highly anisotropic site) separated by 2.0 Å or less were merged to give one site with averaged x, y, z and B values.

Addition of new atoms employed the $(F_o - F_c, \alpha_c)$ density map. The shortest acceptable distance for hydrogen-bonded and non-bonded contacts for solvent–solvent and solvent–protein pairs of atoms was set to 2.2 Å and maximum hydrogen-bond distance to 3.3 Å. No further restraints to these distances were applied for solvent atoms which were already included in the model. An electron-density threshold for selecting new atoms was determined automatically by ARP in an objective manner. The noise in the difference-density histogram is expected to follow a Gaussian distribution in the later stages of refinement (Main, 1990). From the observed histogram the threshold was chosen so that no more than 10% of the density points selected above this level arise from random noise. Thus, if the density map is highly noisy, as at the start of refinement, and the maximum number of sites allowed to be accepted is relatively high, then it may be expected that the model will have about 10% randomly placed solvent atoms. This is supposed to be overcome by removal of poorly defined atoms and by accepting only a few new atoms in each cycle.

The density threshold used for selection of new atoms had throughout refinement approximately the same value of 0.25 Å e^{-3}, as in the later stages of refinement the noise level mostly arises from errors in the X-ray data and therefore remained constant, but changed from 2.8 at the beginning of refinement to about 4.0 r.m.s. density at the end, because the r.m.s. difference density decreased as refinement progressed.

A sphericity-based real-space refinement of x, y, z parameters of water molecules matching expected and actual $(3F_o - 2F_c, \alpha_c)$ atomic density shape was employed and found to greatly enhance the procedure. This provides better refinement of weakly ordered molecules, which (especially if they are in fact not fully occupied) often tend to drift to the edge of the density peak. The r.m.s shift, resulting from the real-space fit, applied every cycle to solvent molecules was about 0.1 Å.

4. Results

4.1. *Refinement of the native structure*

4.1.1. *Refinement with isotropic atomic temperature factors using PROLSQ: MKH.* Coordinates of the protein atoms from the structure of RNase Sa previously refined against 1.8 Å data, PDB (Bernstein *et al.*, 1977) coordinate set 1GMQ (Sevcik *et al.*, 1993), were used as a starting model. The initial R factor was 31.1% for 95% of the data between 10.0 and 1.2 Å. Refinement of the two molecules (1490 protein atoms in total out of which five side chains were in two alternative conformations), 385 water molecules and two sulfate ions (one near Arg63*A* and the other at the active site of molecule *A*) converged in 38 cycles to an R factor of 16.7%, Fig. 3(*a*). The side chains modelled with two conformations were Val6*A*, Ser42*A*, Glu54*A*, Cys 72*A*, Ser3*B*, Glu54*B* and Cys 72*B*. The number of such residues was reduced to five after Cys72 in both molecules was corrected, see below. The occupancy of the two conformations was estimated from the electron density and atomic temperature factors. The

Table 3. *Native structure refinement using PROLSQ (MKH) and SHELXL93 with isotropic temperature factors (MSI)*

Step	1	2	3	4	5
Cycles	1–20	21–25	26–30	31–35	36–38
No. of solvent molecules	330/272	337/270	372/281	384/290	385/288
R factor (%)	18.2/18.1	17.7/17.7	17.0/16.9	16.8/16.7	16.7/16.7

Notes: 95% of data were used. First numbers refer to MKH, second to MSI. Step 1. Previously refined RNase Sa at 1.8 Å without solvent molecules was used as a starting model. Step 2. Introduction of the two sulfate ions. Step 3. Inclusion of partial contributions of H atoms at their idealized positions. Step 4. Adjustment of Asp1*A*, Asp25*A*, Glu74*A*, Leu91*A*, Glu41*B*, Thr76*B* in *PROLSQ* and Asp25*A*, Arg40*A*, Glu74*A*, Leu91*A*, Gln38*B*, Arg40*B*, Glu41*B* in *SHELXL*. Double conformations were assigned for Val6, Ser42*A*, Glu54*A*, Ser3*B*, Glu54*B* (double conformations of Cys72*A* and Cys72*B* were used from the beginning of refinement). Step 5. Replacement of Cys72 by threonine.

Table 4. *Native structure refinement using SHELXL93 with anisotropic temperature factors (MSA)*

Step	5	6	7	8	9	10
Cycles	36–43	44–50	51-55	56–59	60–63	64-75
No. of solvent molecules	312	314	311	322	336	332
R factor (%)	11.7	11.1	11.1	10.9	10.8	10.6

Notes: 95% of data were used in steps 5–8, all data in steps 9 and 10. Step 5: The coordinates of all protein atoms and water molecules from *SHELXL93* isotropic refinement were used as a starting model. Step 6. Rebuilding of Asp1*A*, Gln32*A*, Arg40*A*, Asp25*B*, Gln38*B* Arg40*B*, Ser42*B*, Gln77*B*. Built double conformation for Leu91*A* Step 7. Substitution of Cys72*A* and Cys72*B* by Thr. Step 8. Double conformations for Thr5*A*, Pro13*A*, Pro27*A*, His85*A*, Pro60*B*. Rebuilding of His53*A*, Glu54*A*, Gln38*B*, His53*B* and main chain at Gly4*A*. Step 9. Rebuilding of main chain at Gly4*B*. Modelling double conformation for HIS85*B*. Accurate unit-cell parameters used. Step 10. Adjustment of several residues.

number of water molecules represents about 50% of the theoretical number expected. The course of refinement is shown in Table 3.

4.1.2. *Refinement with isotropic temperature factors using SHELXL93: MSI.* The structure was refined in parallel using *SHELXL*93 with isotropic atomic temperature factors in a similar way to MKH and with the same starting model. The *R* factor dropped more quickly, Fig. 3(*a*), compared to *PROLSQ*, however the CPU time per cycle is about ten to 20 times greater. Default values of gradient step sizes were used in the refinement and there was no attempt to investigate the influence of their values on the speed of convergence.

Refinement of two protein molecules, two sulfate ions and 288 water molecules converged in 38 refinement cycles to an *R* factor of 16.7%, the same as for MKH. Double conformations were modelled for the same residues as for *PROLSQ*. The course of refinement is given in Table 3 and *R* factors in Fig. 3(*a*). The statistics, Table 2, indicate that the geometry differs more from the idealized library compared to the use of *PROLSQ* as the restraints are weaker.

4.2.3. *Refinement with anisotropic temperature factors using SHELXL93: MSA.* As a starting set the MSI coordinates were used. The course of refinement is shown in Table 4. Three positional and six thermal parameters for each of the 1490 non-H protein atoms (two molecules) and for about 350 water molecules give a ratio of input to refined parameters of 3.5. Therefore, it was reasonable to employ anisotropic atomic temperature factors.

Refinement of two protein molecules, two sulfate ions and 332 water molecules with all data between 10.0 and 1.2 Å converged with an *R* factor of 10.6%. The drop in $R_{\rm free}$ during the course of refinement is shown in Fig. 3(*a*). Two conformations were built for 13 residues (seven from the starting model plus six new), Table 4. This was reduced to 11 residues after correcting the identity of the Cys72 residue. Introduction of double conformations for the extra amino-acid residues compared to the isotropic refinement reduced the *R* factor only slightly. For most of these residues the two conformations were built within a continuous piece of electron density.

After 15 cycles of refinement it became clear that the residue in position 72 is not cysteine but threonine. This residue is far from the active site and apparently does not participate in the catalytic function of the enzyme. Cys72 was originally modelled according to the amino-acid sequence (Shlyapnikov *et al.*, 1986). In all but one of the previously refined structures of RNase Sa and its complexes with mononucleotides this residue was modelled in two conformations for the SG atom in both molecules. CB–SG distances in both conformers showed significant discrepancy from the target value of 1.808 Å taken from Engh & Huber (1991). The deviation was increasingly pronounced in structures with higher accuracy and weaker geometrical restraints, Table 5. The difference electron density shows a peak between the SG and CB atoms and a hole at the SG position, clearly indicating that the actual bond length is shorter, Fig. 4(*a*). Furthermore, there are two well defined water molecules at hydrogen-bonding distances of 2.8 Å from Thr72 OG1 which could not be formed with an SH group. Therefore, Cys72 was replaced by Thr72 and more refinement cycles were submitted, Table 3 and 4. Thr72 fits nicely to the electron density and the distances CB–CG, CB–OG fit well to target values, Table 6.

This has clarified long-standing doubts about the presence of Cys72 in RNase Sa beginning with unsuccessful attempts to use mercury compounds specific to cysteine residues in preparing isomorphous derivatives (J. Sevcik, unpublished) as well as from experiments aimed to modify it chemically for other purposes (I. Rybajlak, personal communication). Experiments to confirm chemically the identity of residue 72 are underway.

4.1.4. *Anisotropic refinement of the 2′-GMP complex with SHELXL93: MGMP.* Refinement of the RNase Sa/2′-GMP structure followed the same protocol as that of the native structure using *SHELXL*93 with anisotropic temperature factors. As starting model the MSA coordinates after cycle 25 were used without water molecules. The refinement of the two protein molecules, 24 inhibitor

Table 5. *Side-chain bond distances (Å) for Cys72*

	Resolution (Å)	*R* (%)	Molecule *A* CB—SG	Molecule *A* CB—SG′	Molecule *B* CB—SG	Molecule *B* CB—SG′	Reference
PROLSQ isotropic refinement							
Native	1.8	17.2	1.803	—	1.815	—	(1)
Complex/3′-GMP	1.8	17.5	1.762	1.782	1.746	1.780	(1)
Native	1.8	13.9	1.726	1.776	1.730	1.754	(2)
Complex/2′-GMP	1.7	13.2	1.703	1.766	1.742	1.776	(2)
Complex/2′,3′-GCPT	2.0	11.9	1.719	1.744	1.675	1.749	(3)
MKH	1.2	16.7	1.604	1.705	1.664	1.730	(4)
SHELXL93 isotropic refinement							
MSI	1.2	16.7	1.622	1.614	1.600	1.652	(4)
SHELXTL93 anisotropic refinement							
MSA	1.2	10.6	1.537	1.527	1.579	1.477	(4)

Target value is 1.808 Å.

References: (1) Sevcik, Dodson & Dodson (1991). (2) Sevcik, Hill, Dauter & Wilson (1993). (3) Sevcik, Zegers, Wyns, Dauter & Wilson (1993). (4) Present work.

Table 6. *Side-chain bond distances (Å) for Thr72*

Target values are 1.433 Å for CB—OG1 and 1.521 Å for CB—CG2 distances.

Molecule	*A* CB—OG1	*B* CB—CG2	*A* CB—OG1	*B* CB—CG2
MKH	1.448	1.558	1.473	1.533
MSI	1.493	1.537	1.464	1.534
MSA	1.420	1.521	1.419	1.508
MGMP	1.439	1.517	1.410	1.514

Table 7. *Refinement of the complex structure (MGMP) using SHELXL93 with anisotropic temperature factors*

Step	1	2	3	4	5	6	7
No. of cycles	1–8	9–14	15–20	21–25	26–30	31–35	36–40
No. of solvent molecules	241	359	383	405	425	433	439
R factor (%)	14.6	11.7	11.2	11.3	11.1	11.0	10.9

Notes: 95% of data were used in steps 1–3, all data in steps 4–7. Step 1. MSA protein coordinates without solvent molecules and double conformations of disordered residues as a starting model. Step 2. Modelling of two conformations for Val6*A*, Ser42*A*, Leu91*A*, Ser3*B*, Glu54*B*. Step 3. Adjustment of 5 residues. Rebuilding of main chain at Gly4*A* and Gly4*B*. Step 4. Refinement with all data. Step 5. Modelling of two conformations for Thr5*B*. Step 6. Modelling of two conformations for Pro13*A* and Pro60*B*. New unit-cell parameters. Step 7. Modelling of two conformations for His85*B*.

Table 8. *Residues with two conformations and disordered residues.*

(*a*) Residues with two conformations in MSA and MGMP

MSA Molecule *A*	MSA Molecule *B*	MGMP Molecule *A*	MGMP Molecule *B*
	Ser3*B*		Ser3*B*
Thr5*A*		Thr5*A*	Thr5*B*
Val6*A*		Val6*A*	
Pro13*A*		Pro13*A*	
Ser42*A*		Ser42*A*	
Glu54*A*	Glu54*B*		Glu54*B*
	Pro60*B*		Pro60*B*
His85*A*	His85*B*		His85*B*
Leu91*A*		Leu91*A*	

(*b*) Disordered residues (*B* > 45 Å^2)

MSA Molecule *A*	MSA Molecule *B*	MGMP Molecule *A*	MGMP Molecule *B*
Asp1*A*			
		Ser3*A*	
Asp25*A*	Asp25*B*	Asp25*A*	
Gln32*A*			
	Gln38*B*		Gln38*B*
Arg40*A*	Arg40*B*	Arg40*A*	Arg40*B*
	Glu41*B*		Glu41*B*
	Thr76*B*		Thr76*B*
	Gln77*B*		Gln77*B*

atoms, one sulfate ion and 439 water molecules converged in 40 refinement cycles to an *R* factor of 10.9%. The course of refinement is given in Table 7 and *R* factors in Fig. 3(*b*).

Clear density corresponding to 2′-GMP was observed at the molecule *A* active site and all its atoms were refined with unit occupancy. H atoms belonging to the 2′-GMP molecule were not introduced. There was no electron density at the active site of molecule *B* for 2′-GMP. The difference map calculated between the previously collected 1.7 Å resolution data (Sevcik *et al.*, 1993) where 2′-GMP was present at the active site of molecule *B* with occupancy 0.5 and the current MGMP data shows clear electron density at the active site of molecule *B*, Fig. 4(*b*). This suggests that in the present crystal 2′-GMP is not bound to molecule *B*. Moreover the previously measured 1.7 Å data and the present MGMP data are not fully equivalent as the goodness of fit is much greater than 1, see above.

It has been observed that Glu54 is disordered in structures of the native enzyme but in structures of complexes with inhibitors it has only one well defined conformation. In the present model Glu54*B* is disordered which supports the observation that 2′-GMP is not present at the active site of molecule *B*. Those residues for which double conformations were built and those which were disordered in the final MSA and MGMP are shown in Table 8.

The active site of molecule *B* in the crystal is less accessible to inhibitor than that of molecule *A*. It was shown previously that the close proximity of a neighbouring molecule in the crystal prevented access of 3′-GMP to the active site of the *B* molecule. However, 2′-GMP and 2′,3′-GCPT were observed there with low occupancy. As crystals of the complexes were prepared by soaking, it is likely that small variation in the concentration of the inhibitor influences its binding to the active site of molecule *B*.

5. Discussion

5.1. *Accuracy of the models*

The final *R* factors for all four native models is shown as a function of resolution in Fig. 3(*c*), refinement parameters, restraint weighting scheme and deviations from ideality after the last cycle of refinement in Table 2 and the statistics for electron-density maps in Table 9. The σ_A plot (Read, 1986) was used to estimate the overall coordinate error for all models, Table 2. The structures with anisotropic atomic temperature factors have significantly better accuracy (0.05–0.08 Å) than structures with isotropic factors (0.13–0.15 Å) showing the advantage of anisotropic temperature factors when there are sufficient X-ray data. In addition the drop of R_{free}, Figs. 3(*a*) and 3(*b*), is significant. The high accuracy of the structures is confirmed by the Ramachandran plot for both molecules of MSA, Fig. 5, using *PROCHECK* (Morris, MacArthur, Hutchinson & Thornton, 1992).

95% of the X-ray data were used in most refinement cycles, 5% were omitted and used to monitor Rfree. All data were included in the last stages of refinement. One might expect that if extra data are included and the observations to parameters ratio increases the conventional *R* factor should go up. The situation in this case is the other way round: the *R* factor falls by 0.3% for both MSA and MGMP, Tables 4 and 7. Thus, use of incomplete data (even if only a small percentage of reflections is randomly omitted) limits the convergence of the refinement even at atomic resolution. This clearly shows one problem of using R_{free} as a cross-validation parameter: in order to check whether overfitting has occurred, part of the data is omitted and this itself introduces a degree of overfitting.

For the well defined parts of the structure there are resolved peaks in ($3F_o-2F_c$, α_c) electron density at the 2.25 e Å^{-3} (5σ) level for individual atoms, Figs. 4(*c*) and 4(*d*). For comparison, the same residue, Tyr52, taken from the structure refined at 1.8 Å is shown, Fig. 4(*e*). The difference electron density showed the positions of many H atoms before their contribution was included in the refinement, Fig. 4(*f*). Difference electron-density peaks observed at main-chain carbonyl groups before introduction of anisotropic atomic temperature factors represent their thermal vibrations and/or π electron pairs.

Table 9. *Final electron-density characteristics (e Å^{-3})*

	MKH	MSI	MSA	MGMP
($3F_o-2F_c$, α_c)*				
Maximum	9.98	9.42	8.41	8.33
Minimum	−1.46	−1.37	−1.12	−1.32
R.m.s.	0.50	0.47	0.45	0.46
(F_o-F_c, α_c)				
Maximum	0.56	0.57	0.30	0.52
Minimum	−0.41	−0.48	−0.37	−0.36
R.m.s.	0.07	0.07	0.05	0.05

* F_{000}/V = 0.28 e Å^{-3} was used.

Several cycles of full-matrix refinement in blocks containing parameters of 20 residues and overlapping by one residue were used to estimate the coordinate errors of MSA and MGMP. The r.m.s. error for main-chain atoms of both molecules in the two structures, Fig. 6, is about 0.02 Å. O and N atoms have smaller errors than C atoms because they have more electrons. O atoms have errors similar to those for N atoms. This reflects higher mobility for carbonyl O atoms as they make only one covalent bond. The overall r.m.s. coordinate error for protein atoms is 0.03 Å and for all atoms (including waters and ligands) 0.05 Å for MSA and 0.07 Å for MGMP. These are in excellent agreement with values derived from the σ_A plots. The agreement between different methods of error estimation has been discussed in Daopin, Davies, Schlunegger & Grütter (1994).

The coordinate errors, Fig. 6, are different for molecules *A* and *B* but similar for the two *A* molecules and the two *B* molecules from the two models. The highest inaccuracies are around the most 'problematic' parts of the structure, *e.g.* the Ser3 carbonyl O atom, which is highly disordered in all molecules.

For much of the structure data to atomic resolution allow correct identification of the atomic type directly from the density map. A plot of (F_o, α_c) density interpolated at atomic centres as a function of atomic

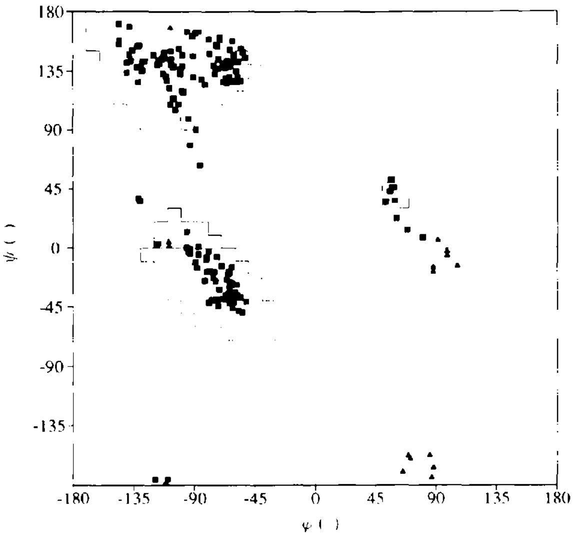


Fig. 5. Ramachandran plot for the native structure (MSA). 91.9% of the residues are in the most favoured regions. Glycine residues are shown as triangles.

temperature factors is given in Fig. 7(*a*) for the two molecules of MSA. If the atomic temperature factor is low the C, N and O atoms are clustered and distinct from each other. The density for each atom type is approximately proportional to the atomic number. The points located below the 'carbon curve' correspond to side chains in double conformations. S atoms of cysteine residues are clearly distinguished from other atoms. S atoms of the two supposed sulfate anions bound to protein fall on the same curve even for a temperature factor as high as 80 Å^2. This serves as an additional proof that the anion bound to the active site has an atom heavier than oxygen but is not sufficient to resolve the ambiguity as to whether the anion is sulfate or phosphate.

A similar plot for water molecules is given in Fig. 7(*b*). Most of the solvent sites have density significantly higher than 1σ above the mean. The density for solvent sites with higher temperature factors approaches the mean value of the map. These solvent sites are of marginal significance but most of them form reasonable hydrogen-bond contacts and are located in separated peaks of electron density.

5.2. *Temperature factors*

The statistics of temperature factors for all models is given in Table 10. Average protein atomic temperature factors are very similar in all models. Temperature factors of *B* molecules are higher than for *A* molecules, resulting from crystal contacts which allow more freedom for molecule *B*. The differences were about 4.5 Å^2 in all previously refined models at lower resolution while in the present ones they are only 1.5 Å^2.

5.3. *Comparison of the models*

R.m.s. and maximum deviations between corresponding main-chain atoms based on least-squares superposition of CA atoms of molecules *A* and *B* for all four models are given in Table 11. Molecules *A* and *B* in all models of the native structure are essentially identical with r.m.s. deviation between CA atoms of 0.07 Å or less. The maximum displacement of 0.29 Å corresponds to the N-terminus.

Differences between molecules *A* of the native and complex structure are caused by the presence of in-

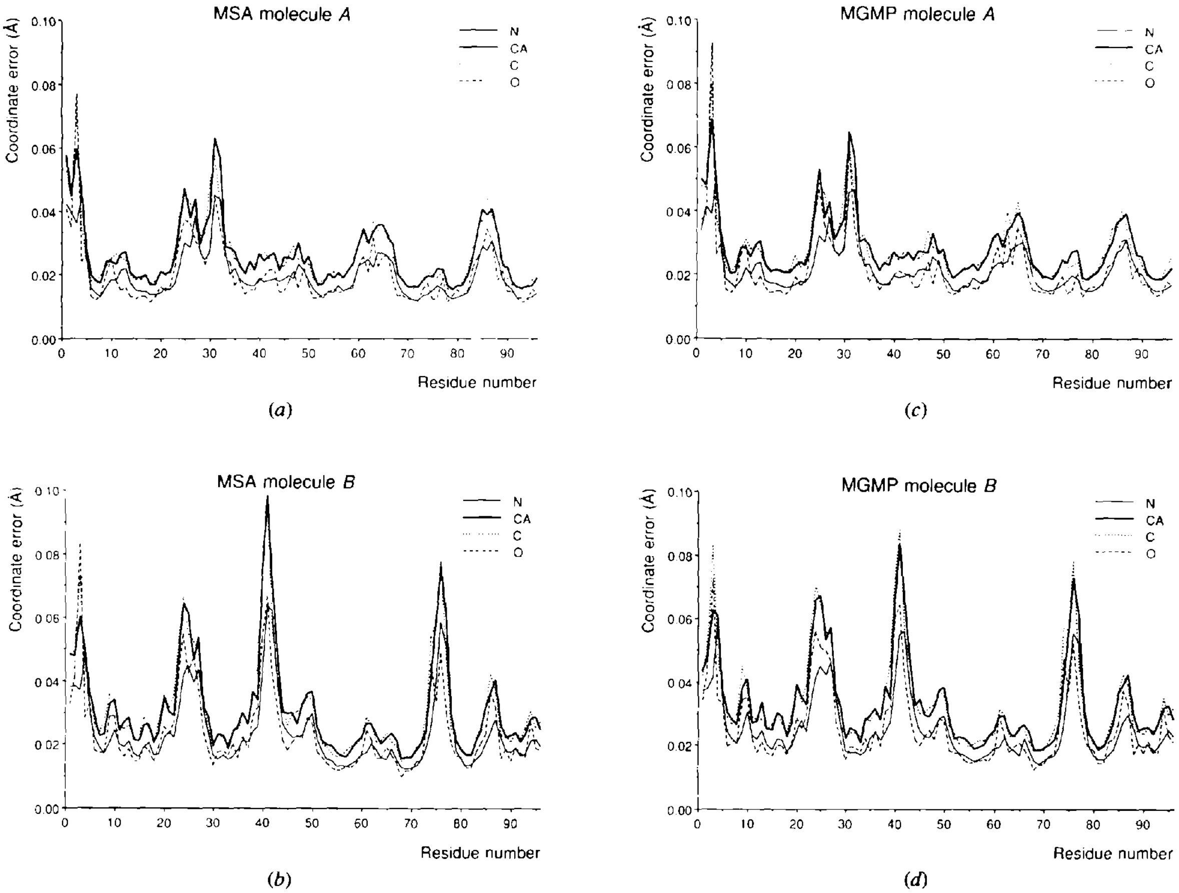


Fig. 6. Coordinate errors for main-chain atoms estimated from matrix inversion. (*a*) Molecule *A* of MSA, (*b*) molecule *B* of MSA, (*c*) molecule *A* of MGMP, (*d*) molecule *B* of MGMP.

Table 10. *Average isotropic atomic temperature factors (Å^2)*

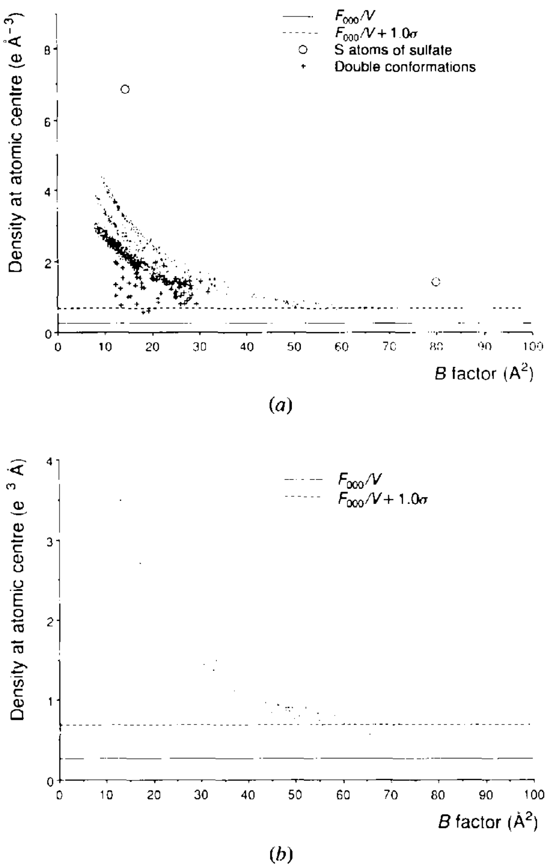
	MKH		MSI		MSA*		MGMP*	
Molecule	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
Main chain	10.4	12.5	11.4	13.2	13.4	15.3	12.1	13.9
Side chain	17.2	18.5	16.9	18.3	18.2	20.2	16.2	17.2
Protein	13.8	15.4	14.0	15.7	15.8	17.7	14.1	15.5
Ions	14.0/41.3†	—	14.4/79.1	—	16.4/76.9	—	15.7	—
2′-GMP	—	—	—	—	—	—	21.7	—
Overall protein	14.6		14.9		16.8		14.8	
Water molecules	42.0		36.6		42.0		43.0	
Wilson plot estimate			10.8				11.2	

* Isotropic equivalents of the anisotropic tensors are given. † The occupancy for all atoms of the second anion in MKH was 0.5.

hibitor. After formation of the complex the surrounding atoms of the active site move so that the active site cleft is more closed while the rest of the structure remains intact. The movement of individual atoms is small but as they all move towards the active site, added together they make a considerable change in the structure. MGMP *B* in comparison to the native molecule *B* shows a relatively large displacement which is caused by a different conformation of the main chain around Ser3–Gly4, but the rest of the molecule, including the active-site region is very close to the others. The density for the carbonyl O atom of Ser3 suggests two possible conformations. Only the stronger was modelled in each structure.

Table 11. *Least-squares superposition of molecules A and B: r.m.s. and maximal displacement for CA atoms (Å)*

	MSI	MSA	MGMP
Molecule *A*			
MKH	0.06/0.26	0.06/0.23	0.12/0.42
MSI	—	0.05/0.17	0.12/0.39
MSA	—	—	0.10/0.35
Molecule *B*			
MKH	0.06/0.13	0.07/0.24	0.12/0.68
MSI	—	0.06/0.29	0.12/0.72
MSA	—	—	0.09/0.44



(*a*)

(*b*)

Fig. 7. (F_o, α_c) electron densities at atomic centres as a function of temperature factor. (*a*) All atoms; S atoms from the two sulfates are marked by circles. (*b*) Water molecules only.

The deviation between CA atoms of superimposed *A* and *B* molecules is similar for both MSA and MGMP. The r.m.s. deviation is 0.38 Å (MSA) and 0.39 Å (MGMP) with a maximum displacement of 1.5 Å for the loop formed by Gly61–Thr64. This loop protrudes from the surface of the protein, which gives it flexibility reflected in different positions in different crystal environments. In addition, in molecule *A* in both MSA and MGMP it interacts with a strongly bound sulfate anion (see below).

The r.m.s. deviation between main- and side-chain atoms for molecules *A* and *B* (MSA) is plotted in Fig. 8(*a*) as a function of residue number. With the exception of the flexible loop Gly61–Thr64, the main-chain atoms deviate with an r.m.s. of 0.30 Å. This is substantially higher than the estimated error in the atomic positions. There is no correlation between deviations in main-chain atoms and their temperature factors, Fig. 8(*b*). The 'residue-based' correlation coefficient between r.m.s. deviations in side-chain atoms and their average temperature factors is 17%, which is very low. Thus, the deviations between main-chain atoms in molecules *A* and *B* reflect a real difference between the two crystallographically independent molecules.

Side-chain atoms deviate much more, with an r.m.s. of about 1.0 Å. Some of these deviations are due to high temperature factors or partial disorder of side chains for

Asp25, Gln32, Gln38, Arg40 and Glu41. Many deviations reflect different crystal environment (Ser48, Glu74, Thr76, Gln77 and Gln94). In molecule *A* these residues are part of a well ordered hydrogen-bond network with symmetry-related molecules while in molecule *B* they are oriented towards solvent and have higher temperature factors and weaker density. Arg63 belongs to the loop Gly61–Thr64, which is different in the two molecules. The Arg63 side chain interacts with bound sulfate anion in molecule *A* and acquires a different conformation in molecule *B*. Deviations for side-chain atoms correlate to some extent with temperature factors, Fig. 8(*b*), and the 'residue-based' correlation coefficient is 58%.

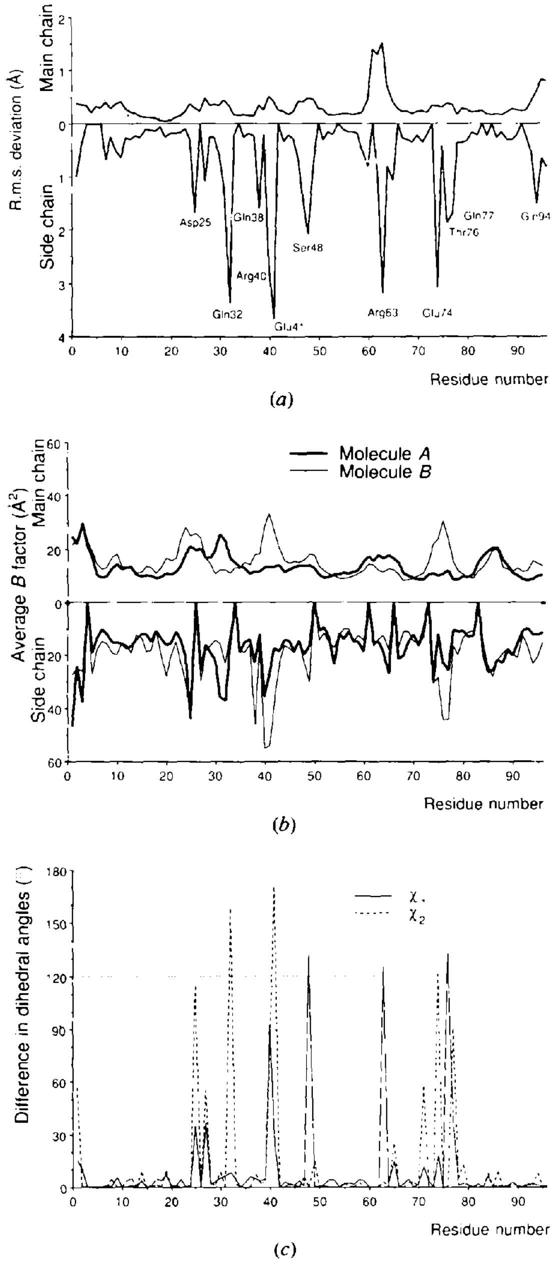


Fig. 8. Comparison of molecules *A* and *B* in MSA. Side chains in multiple conformations are excluded. (*a*) R.m.s. deviation between main-chain and side-chain atoms as a function of residue number. Residues with highly deviating side chains are labelled. (*b*) Average temperature factor for main- and side-chain atoms as a function of residue number. (*c*) Absolute difference of χ^1 and χ^2 dihedral angles for side chains. The dotted line marks the 120° difference between preferred rotamer conformations.

Comparison of dihedral χ_1 and χ_2 angles for both molecules in MSA is shown in Fig. 8(*c*) as an alternative indication of side chains in different conformations. All deviations in dihedral angles correspond to residues with side chains deviating highly from non-crystallographic symmetry. Deviations are generally clustered around 120 which corresponds to alternative 'staggered' rotamers typical for sp^3-hybridized C—C bonds.

To conclude, the deviations from non-crystallographic symmetry in the RNase crystal structure are significant for main-chain atoms and quite large for side chains and some main-chain atoms. It is mostly affected by different crystal environment but also by the presence of bound ligands. Implementation of non-crystallographic symmetry (NCS) as a constraint in the refinement of RNase structure would be beneficial for most of the backbone atoms only at resolution lower than about 2.0 Å, when the coordinate error is expected to be comparable with the r.m.s. deviation. If side-chain atoms are also to be considered, introduction of NCS constraints is expected to hold only at resolution lower than about 3.0 Å, where the coordinate error can be as poor as 0.5 Å.

5.4. *Residues in double conformations*

For some residues two discrete side-chain conformations were modelled, Table 8. For the proline residues the second conformation was built only for the CG atom. The two conformations for His85 in both molecules of MSA are related by tilting the imidazole ring by about 20°. One of these positions is the same as the single conformation found in the complexes of RNase Sa with mononucleotides. A similar situation arises for the other catalytic residue, Glu54. This has two alternative conformations in the native *A* and *B* molecules. In molecule *A* of the complex with 2′-GMP, where the ribose is oriented towards the outside of the active site, Glu54*A* has only one conformation essentially identical to one of those observed in the native protein. Glu54*B* has two conformations as in the native (inhibitor is not bound there). A third conformation of Glu54, different to both of those in the native was observed in the complex with the product of reaction, 3′-GMP, due to the ribose ring being buried in the active site.

According to Smith, Hendrickson, Honzatko & Sheriff (1986) residues for which multiple conformations are likely are exposed to solvent and are either polar or charged. This is not the case for Val6*A* and Leu91*A* which are non-polar and together with Thr5*A* only partially exposed to solvent. They are located close to each other (the distance between Leu91 CD1 and Thr5 OG1 is only 3.5 Å) and oriented in such a way that the less occupied conformation of Leu91 faces the more highly occupied conformation of Thr5. These three residues form a disordered group. Similarly, a pair of disordered residues is found between the two molecules in the crystal, namely Ser42*A* and Pro60*B*. The shortest distance between them is 4.5 Å. Similar groupings of two disordered residues have been observed in other structures, *e.g.* that of crambin refined at 0.83 Å resolution (Teeter, Roe & Heo, 1993).

For MGMP two conformations were built for essentially the same residues as in MSA, Table 8, except Glu54*A* and His85*A* which have well defined single conformations as the active site is occupied by inhibitor. The second conformation of Thr5*B* in MGMP was not seen in MSA.

5.5. *Disordered side chains*

The electron density is generally well defined but there are several poorly defined residues in MSA for which electron density for at least one side-chain atom in the $(3F_o - 2F_c, \alpha_c)$ map is less than 0.5σ and/or the temperature factor higher than 45 Å^2, Table 8. For the side chains of three residues, Gln38*B*, Arg40*B* and Glu41*B*, there is hardly any density at all in contrast to their counterparts in molecule *A* where mobility of these residues is limited by the close proximity of a neighbouring protein molecule in the crystal lattice. As Glu41 takes part in binding the base, its conformation in complexes of RNase Sa with nucleotides is clear. Asp25 in both molecules is highly disordered in all previous as well as present models. The structures neither confirm its identity nor suggest clearly an alternative amino acid.

5.6. *Bound anions*

A sulfate ion bound to Arg63*A* was identified in all structures. In the active site of molecule *A* of the present native structure there was electron density which by its characteristic tetrahedral shape suggested the presence of an ion which occupies the same site as the phosphate group in complexes of RNase Sa with mononucleotides. This feature was not observed in the electron density in previous RNase Sa structures at lower resolution. There is no direct chemical evidence as to the identity of the ion. It could be a sulfate or phosphate ion as both were present in the crystallization liquor (phosphate buffer, ammonium sulfate as precipitant). The average distances between the central atom and the four O atoms are 1.50, 1.49 and 1.53 Å for MKH, MSI and MSA, respectively. The expected values are 1.49 Å for S—O and 1.56 Å for P—O (Wells, 1975). As the average temperature factor for this anion is about 75 Å^2 implying it is much less than fully occupied, the bond lengths could be influenced by restraints (they were set for a sulfate anion) and, therefore, its nature cannot be unambiguously identified. However, as the average length in the anisotropic MSA model increases towards that typical of a phosphate, this may well be the true identity of the ion. The anion near Arg63*A* is clearly a sulfate as the average S—O bond length is 1.49 Å (r.m.s. 0.02 Å). It is well defined with average temperature factor of 14 Å^2.

5.7. *Water structure*

The building of water structure for all models was carried out in essentially the same manner (see ARP, above). For the native structures refined isotropically (MKH and MSI) the same initial protein model without water molecules was used. However, a significant difference in the number of water molecules in these two models was found: 385 in MKH and 288 in MSI. A plot of the number of waters as a function of atomic temperature factor, Fig. 9(*a*), shows very similar profiles for different models for sites with temperature factors less than 40 Å^2.

In the first eight cycles the number of solvent sites in MKH and MSI were the same, Fig. 9(*b*), as 30 molecules were allowed to be added each cycle and only a very few were removed. The numbers start deviating when

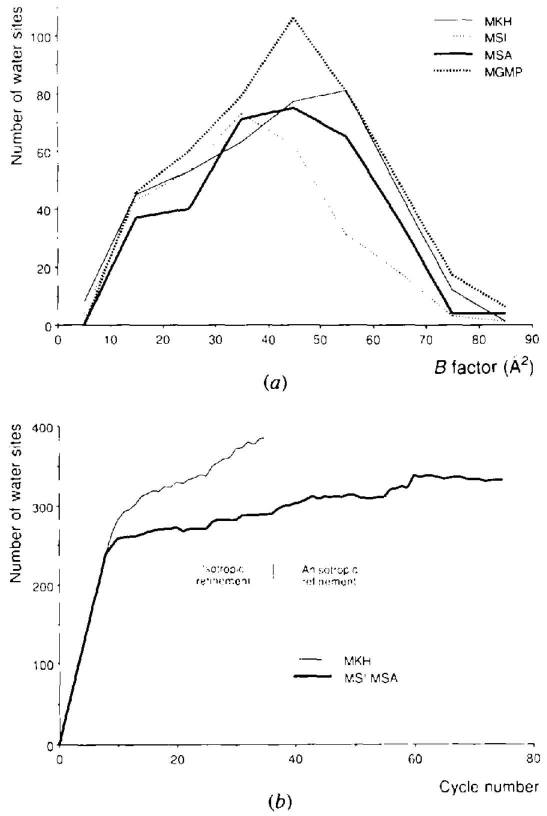


Fig. 9. (*a*) Number of water molecules from all models in temperature-factor ranges. (*b*) Number of water molecules as a function of refinement cycle of the native structure.

the total number of waters reached about 240 and the number of new sites suggested by ARP became less than 30. Subsequent refinement cycles increased the number of water molecules in MSI only slightly in comparison to MKH.

The solvent structures from the MKH and MSI models were compared. Atoms were classified as equivalent if they lay within 1.1 Å of one another in the two models (about half of the ARP merging distance). After eight cycles of refinement MKH contained 238 water sites and MSI 239. Of these, 194 were within 0.5 Å and 201 within 1.1 Å of one another. Only 37 water molecules (15%) did not have an equivalent in the other model. By the end of isotropic refinement (38 cycles) the number of waters increased to 385 for MKH and 288 for MSI. The number of equivalent sites increased to 248 within 0.5 Å and 271 within 1.1 Å. 104 sites (27%) from MKH and 17 sites (6%) from MSI were not paired. A plot of the number of equivalent waters as a function of the distance between them is given in Fig. 10(*a*).

The non-paired water molecules are mostly those with high (more than 40 Å^2) temperature factors. Non-paired waters in MKH mimic the solvent continuum as this model was refined without bulk solvent contribution. If these waters are added to MSI they form a typical hydrogen-bond network as their distances to MSI water sites cluster around 2.9 Å, Fig. 10(*b*). The temperature factors of paired water sites from MKH and MSI are plotted as a function of the distance between them in Fig. 10(*c*). Pairs of waters with low temperature factors generally superimpose very closely while waters with higher temperature factors deviate more.

Ten more refinement cycles of MSI with the solvent continuum option turned off gave 322 solvent sites. When this revised MSI was compared to MKH there were more equivalent sites (267 within 0.5 Å and 298 within 1.1 Å). Only 88 sites (23%) in MKH and 25 sites (6%) in MSI were not equivalent. The revised MSI model without solvent continuum still contains less solvent sites and refines to about 0.4% higher *R* factor than MKH, perhaps an indicator of memory effects in the refinement.

Comparison of the solvent models refined isotropically (MSI) and anisotropically (MSA), both with the solvent continuum option, is presented in Fig. 10(*a*). The profile differs from the two isotropically refined models (MKH and MSI). There are 332 solvent sites in MSA and 288 in MSI. 254 of these deviate by less than 0.5 Å and 276 by less than 1.1 Å. There are now less paired water sites with deviation less than 0.1 Å. An apparent maximum appears at a deviation of about 0.1 Å. This suggests that the MSI and MSA solvent structures are systematically different, due to the anisotropic description of thermal vibration in MSA. Comparison of MSI (288 water molecules) and MGMP (439 water molecules) gives similar results. There are 220 water sites that deviate by less than 0.5 Å and 273 by less than 1.1 Å. The reason for the lower number of paired water molecules in the range below 0.5 Å can be explained by the fact that molecule *A* of the complex is more closed.

Comparison of the two anisotropically refined models, MSA (332 water molecules) and MGMP (439 water

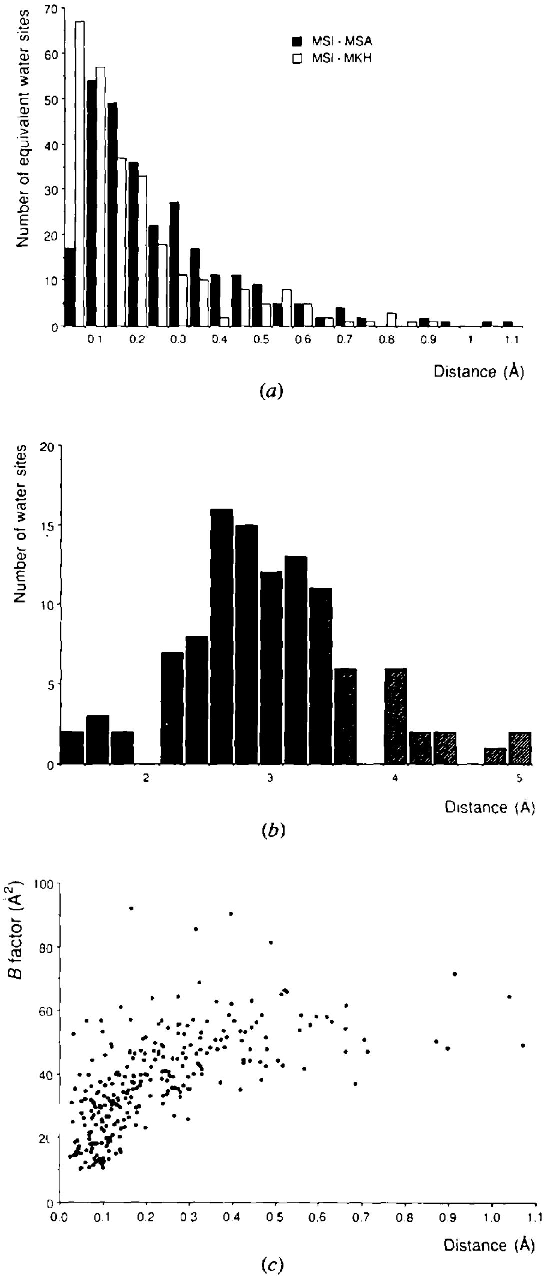


Fig. 10. (*a*) Histogram of numbers of equivalent water molecules between MSI and MKH (open columns) and between MSI and MSA (black columns) as a function of distance between them. (*b*) Histogram showing distribution of distances between unpaired MKH water molecules and MSI molecules. (*c*) Plot of temperature factors of MSI and MSA paired water sites as a function of distance between them.

molecules) gives 259 and 308 water sites that deviate by less than 0.5 and 1.1 Å, respectively.

In all refinements of the native structure the same crystal data, the same starting conditions and the same procedure for solvent structure were used. The protein structures were essentially identical, but different results were obtained for solvent structure. Thus, the refinement procedure itself is responsible for different electron-density distribution in marginal regions. This may result from different least-squares programs. In the refinement, structure-factor amplitudes were used in *PROLSQ* and intensities in *SHELXL*93. In addition the restraints are slightly different for the two programs. An important difference is that *SHELXL*93 refines the water continuum by default and such an option was not used in the refinement of the MKH model. The most different water sites are those with the highest temperature factors. These are usually located either in higher solvent shells or make contacts with disordered residues. Nevertheless, exceptions are quite frequent. These weak waters do not seem to be important for understanding the molecular function of the protein but they do contribute to the quality of the model.

Modelling of solvent flatness by acquiring more high temperature factor waters in MKH is equivalent to a model with essentially the same *R* factor where the solvent continuum has been employed. However, absence of solvent continuum option when refining with *SHELXL*93 results in increase of *R* factor. This may be related to the use of intensities (rather than amplitudes as in *PROLSQ*) in the refinement, as the reflections, mostly responsible for the solvent structure, are those at low resolution (above 6 Å) and have high intensity values. Therefore, for isotropic refinement the solvent continuum gives a better *R* factor for refinement against intensities. This agrees with the recommendations in the use of *SHELXL*93 (Sheldrick, 1993). When amplitudes are used, however, a lower *R* factor may be obtained if bulk solvent is modelled as a set of extra water sites with high temperature factors.

5.8. *The models viewed from different angles*

MSA was refined with weak restraints on peptide planarity. The chiral volumes of main-chain C atoms were restrained to zero with a standard deviation of 0.23 Å^3 which corresponds to a deviation of the C atom out of the CA—O—N plane of about 0.04 Å. The histogram of the ω angles for MSA is shown in Fig. 11(*a*), where molecules *A* and *B* in the native structure

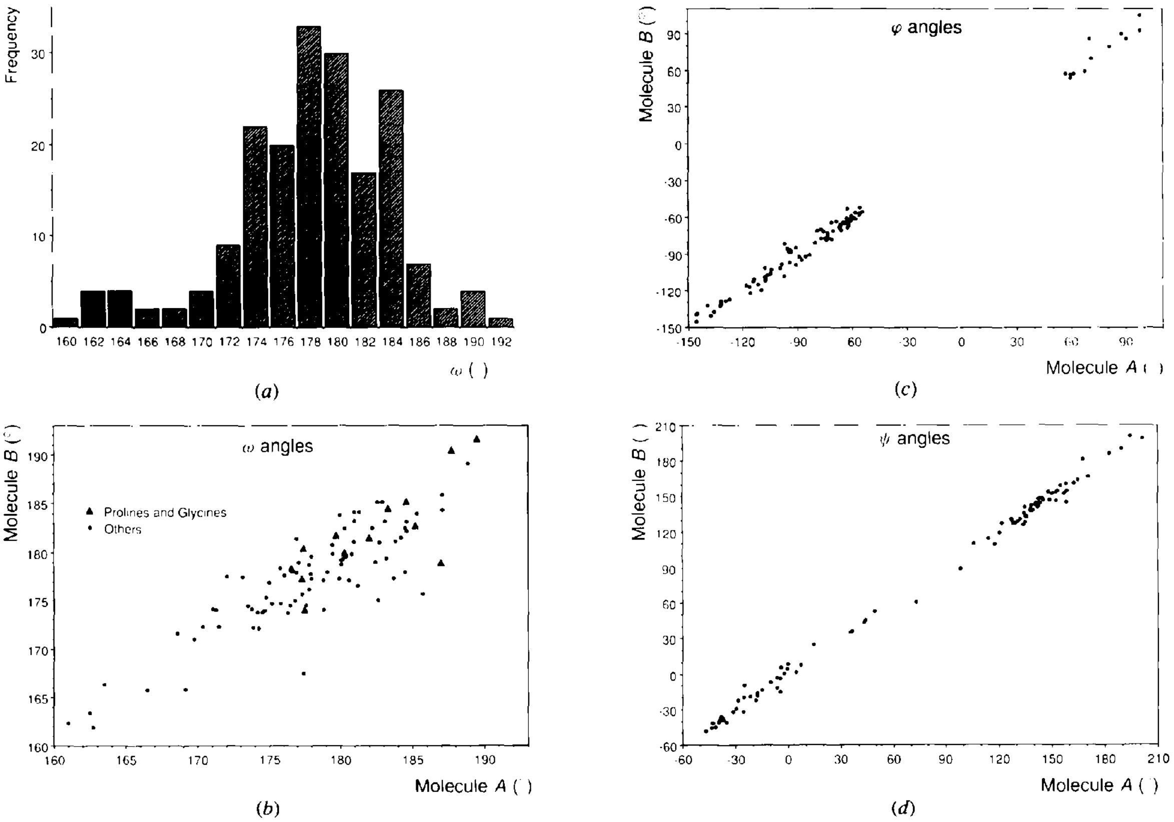


Fig. 11. (*a*) Histogram of ω angles for MSA. (*b*) Comparison of ω angles from molecule *A versus* molecule *B*. Glycines and prolines are indicated. (*c*) φ angles from molecule *A versus* molecule *B*. (*d*) ψ angles from molecule *A versus* molecule *B*.

gave 188 contributors in total. The distribution is not symmetric: it has a mean value of 178.0°, with a standard deviation of 5.8°. Thus, the peptide is not quite planar and the average value of ω deviates by 2° from planarity. Some peptides deviate from planarity by up to 20°.

To ensure that the deviations in ω angles are real and do not merely reflect random errors in the model and that the observed values do not suffer from the relatively small (188) sample size, the correlation between ω angles in the *A* and *B* molecules is shown, Fig. 11(*b*). The 94 points fit well to a straight line, which would represent identity. The r.m.s. deviation between ω angles in the *A* and *B* molecules is 3.0°. Thus, the deviations of ω are meaningful and maintained in the two molecules in the two different environments in the asymmetric unit.

ω angles corresponding to proline and glycine residues are well clustered and have average values of 181 and 183°, respectively. The number of these residues (ten prolines in the *trans* conformation and 16 glycines) is not enough to derive a statistically significant histogram. As they show average values which are higher than 180°, it is clear that these 'special' residues are not responsible for the shift of the average angle (178°) for all residues. The average ω angle for all residues excluding glycines and prolines is 177°. The correlations between φ angles, Fig. 11(*c*) and ψ angles, Fig. 11(*d*) for residues of molecules *A* and *B* are excellent. The r.m.s. deviations between pairs of angles in the two molecules is 5.3° for φ and 5.0° for ψ. There are no significant outliers in the comparison of any of the three main-chain torsion angles between molecules *A* and *B*.

A plot of staggered χ^1 angles for all residues (except alanines, glycines, prolines and residues in double conformations) taken from MSA and MGMP shows three peaks at 64 (7), 180 (5) and –64 (9)° corresponding to three possible rotamers, Fig. 12. The mean values and standard deviations (in parentheses) were derived by fitting three Gaussian functions in the distribution. The χ^1 angles were not restrained. The rotamer preferences, *i.e.* area under each of the three peaks, agree with those derived from 61 protein structures (McGregor, Islam & Sternberg, 1987). However, the mean values and standard deviations are different. Application of g^-, t and g^+ rotamer distributions from this publication to RNase Sa secondary structure and amino-acid sequence gave values of 66, 186 and –66°. The values for g^+ and g^- are comparable, while t is different. The standard deviations given by McGregor *et al.* (1987) vary from 10 to 42° for different residues and are about 20° on average. The standard deviations for rotamer preferences derived from the two RNase Sa structures (304 contributors) are about three times smaller, which could be related to relatively small sample size but rather reflect the high quality of the models refined at atomic resolution. Detailed analysis of ω, φ, ψ and χ^1 angles will be published separately. In the light of the present findings revision of restraint parameters for these angles in refinement may be appropriate.

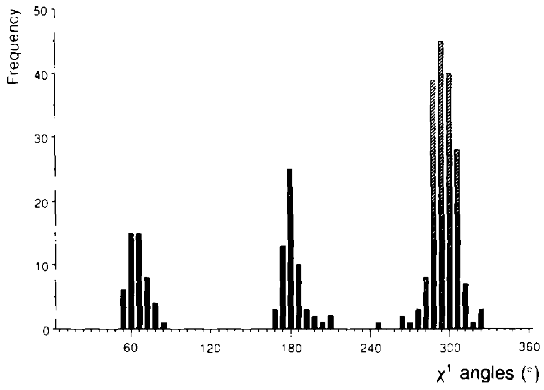


Fig. 12. Histogram of χ^1 angles in MSA and MGMP.

The RNase coordinate and X-ray data have been deposited with the Protein Data Bank (Bernstein *et al.*, 1977).*

6. Concluding remarks

Until recently there was a clear division between macromolecular and small-molecule crystallography. While small-molecule crystals usually diffract to at least the edge of the Cu *K*α sphere which is possible to explore fully using diffractometer techniques, diffraction to better than 2 Å is considered as high resolution for proteins. Indeed, the mean resolution quoted for structures now available from the Protein Data Bank is 2.2 Å and only rarely have protein structures been refined to resolution higher than 1.5 Å.

In order to see individual atomic features in a crystal structure they should be distinguishable in the electron density. The shortest distance between covalently bonded non-H atoms in organic structures is about 1.2 Å. This agrees with an empirical definition of atomic resolution as at least 1.2 Å (Sheldrick, 1990) for it to be possible, in principle, to solve the structure by direct methods.

The main lessons which can be learnt from refinement of the two crystal forms of RNase at atomic resolution are as follows.

(1) The strategy of refinement at atomic resolution implemented for RNase comprises several steps. Initially

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory [Reference: 1RGE, R1RGESF (MGMP); 1RGF, R1RGFSF (MKH); 1RGG, R1RGGSF (MSA); 1RGH, R1RGHSF (MSI)]. Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: JN0011). At the request of the authors, the atomic coordinates and structure factors will remain privileged until 31 September 1996.

restrained refinement was carried out using the diagonal (sparse-matrix) approximation with isotropic description of atomic thermal motion. This rapidly led to a model which was generally correct. Subsequently, more details were introduced: H atoms as fixed contributors, double conformations and anisotropic treatment of thermal vibration. Automated modelling of solvent sites facilitated faster convergence. Such a scheme substantially minimized the degree of manual intervention. All X-ray data were used in the final refinement cycles. In addition, block-matrix least-squares minimization allowed the estimation of individual atomic coordinate errors.

(2) Because of the high accuracy of the model the identity of most of the residues can be conclusively confirmed. For example, only refinement at atomic resolution clarified the identity of residue 72 as threonine rather than as cysteine in a double conformation. Double conformations of side chains and the proper conformation of carboxamides were clearly evident at atomic resolution.

(3) Several approaches to model solvent structure have been tried. Well behaved water molecules can easily be modelled and treated in an objective and automatic manner. Weak (presumably partially occupied) water sites and those belonging to overlapping hydrogen-bonding networks present a problem. Even at this resolution it is not possible to model such solvent satisfactorily. Introduction of bulk solvent contribution provided in essence the same effect as a set of extra weak solvent sites. Relatively poor agreement between observed and calculated structure factors at very low resolution indicates inadequacy of the modelling of weak solvent.

(4) Chemically identical but crystallographically independent molecules, subjected to different crystal environment and lattice contacts, display significant conformational differences. Such differences even for main-chain atoms often exceed the estimated coordinate error typical for high-resolution X-ray crystal structure analyses. Many side chains, especially on the protein surface, show different rotamer conformations. The accuracy in atomic positions should be taken into account when non-crystallographic symmetry constraint is applied.

(5) In spite of significant differences in the conformation of individual residues, the main-chain conformational angles ω, φ and ψ are extremely similar in the two independent molecules. In addition, the ω angle has an average value of 178° and may differ from planarity by up to 20°. This observation reveals a general property of protein structure and suggests it is appropriate to restrain ω during refinement to 178° rather than 180° with a standard deviation of about 6° at any resolution as the nature of protein does not depend on the resolution of the crystallographic analysis.

(6) As bright synchrotron radiation sources, sensitive area detectors and the use of cryogenic techniques are widely available, recording of protein diffraction to very high and sometimes atomic resolution has become tractable. To date, atomic resolution data have been collected for more than 20 proteins at the EMBL Hamburg Outstation alone. RNase Sa is one of these structures. The goals of such studies is to bridge the gap between the X-ray crystallography of small and large structures and to explore the detailed stereochemistry of proteins. We believe that micro- and macromolecular crystallography will to a large extent converge in future.

This work was financed in part by the Slovak Academy of Sciences grant 2/999138/93, the East European Cooperation grant ERB-CIPA-CT-2029 and the EU Biotechnology grant BIO2–CT-92–0524. The authors are grateful to Uwe Carstens for providing excellent library facilities.

References

Bacova, M., Zelinkova, E. & Zelinka, J. (1971). *Biochim. Biophys. Acta*, **235**, 335–342.

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

Brünger, A. T. (1993). *Acta Cryst.* **D49**, 24–36.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.

Daopin, S., Davies, D. R., Schlunegger, M. P. & Grütter, M. G (1994). *Acta Cryst.* **D50**, 85–92.

Dauter, Z., Terry, H., Witzel, H. & Wilson, K. S. (1992). *Acta Cryst.* **B46**, 833–841.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.

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

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* **A36**, 344–350.

Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *J. Mol. Biol.* **2**, 38–64.

Lamzin, V. S., Dauter, Z. & Wilson, K. S. (1995). *J. Appl. Cryst.* **28**, 338–340.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* **D49**, 129–147.

Leslie, A. G. W. (1992). In *CCP4 ESF-EACMB Newslett. Protein Crystallogr.* Vol. 26. Warrington, England: Daresbury Laboratory.

Main, P. (1990). *Acta Cryst.* **A46**, 507–509.

McGregor, M. J., Islam, S. A. & Sternberg, J. E. (1987). *J. Mol. Biol.* **198**, 295–310.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.

Otwinowski, Z. (1993). *DENZO, An Oscillation Data Processing Program for Macromolecular Crystallography.* Yale University, New Haven, CT, USA.

Read, R. J. (1986). *Acta Cryst.* **A42**, 140–149.

Sevcik, J., Dodson, E. J. & Dodson, G. G. (1991). *Acta Cryst.* **B47**, 240–253.

Sevcik, J., Hill, C. P., Dauter, Z. & Wilson, K. S. (1993). *Acta Cryst.* **D49**, 257–271.

Sevcik, J., Sanishvili, R. G., Pavlovsky, A. G. & Polyakov, K. M. (1990). *Trends Biochem. Sci.* **5**, 158–162.

Sevcik, J., Zegers, I., Wyns, L., Dauter, Z. & Wilson, K. S. (1993). *Eur. J. Biochem.* **216**, 301–305.

Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.

Sheldrick, G. M. (1993). *SHELXL93, Program for Crystal Structure Refinement*. University of Göttingen, Germany.

Shlyapnikov, S. U., Both, V., Kulikov, V. A., Dementiev, A. A., Sevcik, J. & Zelinka, J. (1986). *FEBS Lett.* **209**, 335–339.

Smith, J. L., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. (1986). *Biochemistry*, **25**, 5018–5027.

Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.

Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.

Wells, A. F. (1975). *Structural Inorganic Chemistry*, 4th ed, pp. 570–673. Oxford: Clarendon Press.

Zelinkova, E., Bacova, M. & Zelinka, J. (1971). *Biochim. Biophys. Acta*, **235**, 343–344.