

# 2.2 Å resolution structure of the amino-terminal half of HIV-1 reverse transcriptase (fingers and palm subdomains)

Torsten Unge<sup>1\*</sup>, Stefan Knight<sup>1</sup>, Rama Bhikhabhai<sup>1</sup>, Seved Lövgren<sup>1</sup>,  
Zbigniew Dauter<sup>2</sup>, Keith Wilson<sup>2</sup> and Bror Strandberg<sup>1</sup>

<sup>1</sup>Department of Molecular Biology, Biomedical Center, Uppsala University, Box 590, S-751 24 Uppsala, Sweden and <sup>2</sup>EMBL Outstation, c/o Desy, Notkestrasse 85, 22603 Hamburg, Germany

**Background:** HIV-1 reverse transcriptase (RT) catalyzes the transformation of single-stranded viral RNA into double-stranded DNA, which is integrated into host cell chromosomes. The molecule is a heterodimer of two subunits, p51 and p66. The amino acid sequence of p51 is identical to the sequence of the amino-terminal subdomains of p66. Earlier crystallographic studies indicate that the RT molecule is flexible, which may explain the difficulty in obtaining high-resolution data for the intact protein. We have therefore determined the structure of a fragment of RT (RT216), which contains only the amino-terminal half of the RT molecule ('finger' and 'palm' subdomains).

**Results:** The crystal structure of RT216 has been refined at 2.2 Å resolution to a crystallographic R-value of 20.8%. The structure is very similar to that of the

corresponding part of the p66 subunit in the p66/p51 heterodimer, although there is a small difference in the relative orientation of the two subdomains compared with the structure of an RT-DNA-antibody fragment complex. There are a large number of stabilizing contacts (mainly hydrogen bonds and hydrophobic interactions) between the subdomains. The locations of conserved amino acids and the position of some important drug-resistant mutations are described.

**Conclusions:** The RT216 structure provides detailed three-dimensional information of one important part of HIV-1 RT (including the critical active site residues). We propose a model to explain the inhibitory effect of non-nucleoside inhibitors, which partially accounts for their effect in terms of conformational changes of active site residues.

**Structure** 15 October 1994, 2:953-961

Key words: amino-terminal domains, crystal structure, drug design, high resolution, HIV-1 reverse transcriptase

## Introduction

HIV-1 reverse transcriptase (RT) is composed of two subunits of 66 kDa (560 amino acids) and 51 kDa (440 amino acids), named p66 and p51, respectively. The amino acid sequence of p51 is identical to the sequence of the amino-terminal 440 amino acids of p66, but the corresponding three-dimensional structures are different [1]. The polymerase active site is found within these first 440 amino acids, whereas the carboxy-terminal 120 residues comprise the RNase H domain, which catalyzes the removal of the template RNA strand from the RNA-DNA heteroduplex by use of exonucleolytic and endonucleolytic activities.

The polymerase activity of RT is carried out by the p66 subunit of the p66/p51 heterodimer. The role of the p51 subunit is mainly to support the structure of p66, but it may also be involved in the binding of the tRNA primer (tRNA<sup>Lys3</sup>) [1]. Analysis of the structure of the HIV-1 RT heterodimer has suggested that certain drug-resistance mutations may exert their effect through changes in the p51 subunit [2].

The emergence of drug-resistant strains of HIV-1 has been described in which mutations have occurred in the RT sequence [2]. The problem of the development of

resistant strains is severe, and combination therapy, though initially promising, has resulted in new mutants [3]. Not many residues in RT seem to be strictly conserved, and an inhibitor which only binds to strictly conserved residues that are directly involved in the polymerase reaction would be toxic due to similarities of the enzyme to cellular polymerases. Recent experiments indicate that an extremely high dose of nonnucleoside inhibitors can prevent development of resistant viral strains [4]. Since these inhibitors act exclusively on the HIV-1 RT molecule this observation indicates that it might be possible to develop the nonnucleoside inhibitors in such a way that they could prevent development of resistant viral strains.

It is evident that a high-resolution structure of RT will be of significant importance for the design of new inhibitors and for improvement of the inhibitors in use today. Several laboratories are engaged in determining the structure of the HIV-1 molecule. A 3.5 Å resolution structure of HIV-1 RT in complex with the nonnucleoside inhibitor nevirapine has been reported by Kohlstaedt *et al.* [1] and a 3.0 Å structure of a ternary complex between RT, double-stranded DNA and an antibody fragment (Fab) has been described by Jacobo-Molina *et al.* [5]. The p66 subunit of the p66/p51 heterodimer

\*Corresponding author.

contains five subdomains, named fingers, palm, thumb, connection, and RNase H [1]. The molecule is highly flexible, making crystallographic studies difficult. With the aim of obtaining a less flexible molecule than the entire p66/p51 heterodimer, we have expressed and crystallized the amino-terminal half (fingers and palm subdomains) of the polymerase domain of the RT molecule. Here we report the structure at 2.2 Å resolution of this part of the molecule, which we call RT216 and which comprises the first 216 amino acids of p66 (identical to the first 216 amino acids of p51).

The structure, which has a fold very similar to the fingers and palm subdomains of the p66 subunit in the previously described structures, has been refined to an R-value of 20.8%. Details of the stabilizing contacts between the fingers and palm subdomains are described as well as the locations of conserved amino acids and the positions of the mutations giving rise to resistance to nucleoside and nonnucleoside inhibitors.

## Results and discussion

The RT216 molecule clearly has a rigid overall structure, since the crystals (space group  $P4_3$  with cell dimensions  $a=b=98.2$  Å and  $c=31.7$  Å; one molecule per asymmetric unit) diffract beyond 2.2 Å. The structure was determined, using five heavy atom derivatives, independently of the earlier published structures of the heterodimer [1,5]. It provides new and more detailed structural information on the fingers and palm subdomains. As is seen in Fig. 1, not only the side chains but also the carbonyl groups are well defined in the electron density map.

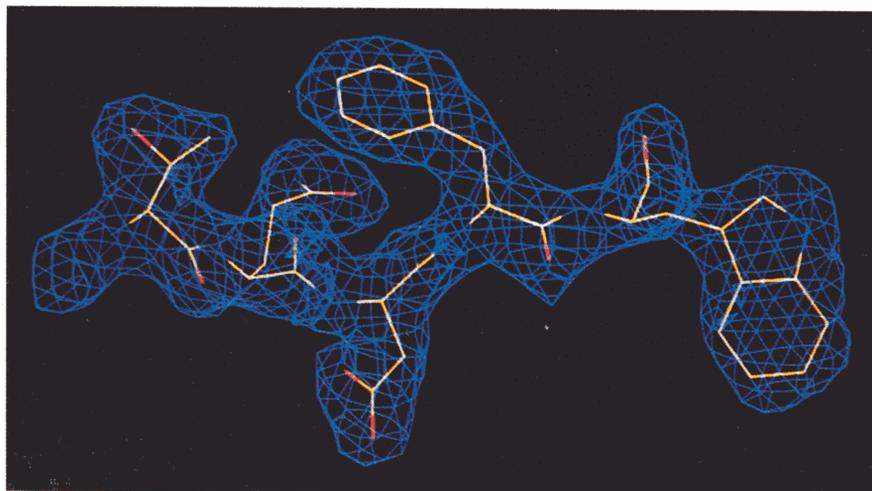
### Overall structure

From the structure of HIV-1 RT in a complex with the inhibitor nevirapine [1] it was evident that the interaction between the fingers and palm subdomains in the p66 monomer was strong but that, on the other hand, relatively few contacts existed between these two subdomains and the other subdomains of the p66 subunit.

In the p51 subunit, the connection subdomain packs against the fingers and palm subdomains with an extensive interaction area, and, to accommodate this interaction, the relative orientation of the fingers and palm subdomains is different from that in the p66 subunit. We therefore surmised that a polypeptide chain encompassing the fingers and palm subdomains would fold into a stable globular structure similar to that found in the p66 subunit, although it could not be excluded that the two subdomains would be arranged as in the p51 subunit. Binding studies show that RT216 binds the phenyl ethyl thiazole thiourea (PETT) inhibitor (K Bäckbro, B Strandberg and T Unge, unpublished data and C Agren, *et al.*, & X-X Zhou, unpublished data), indicating that the fold is similar to the corresponding part of the p66 subunit. Therefore it was not surprising to find that the fingers and palm subdomains in RT216 have a similar arrangement and similar folds to the intact p66 subunit [1,5] (Fig. 2).

Folding patterns very similar to the palm subdomain have been observed in several other RNA and DNA-binding proteins. These include the palm subdomain of the Klenow fragment [6] as well as the ribosomal protein S6 [7], the RNA-binding domain of the U1A small nuclear ribonucleoprotein particle protein [8] and the DNA-binding carboxy-terminal domain of the E2 protein from papilloma virus [9].

In order to quantify the similarities between our structure and that of Jacobo-Molina *et al.* [5], we have superimposed our model of RT216 on the corresponding regions of the p66 and p51 subunits (coordinates kindly provided by E Arnold), using the least squares fit program LSQMAN (GJ Kleywegt and TA Jones, unpublished program). The individual subdomains superimpose slightly better on the corresponding subdomains in the p66 subunit than in the p51 subunit. Comparing RT216 with the p66 subunit, the root mean square deviation (rmsd) in  $C\alpha$  positions is 1.17 Å for 114 residues in the fingers subdomain, and 1.44 Å for 81 residues in the palm subdomain. The corresponding numbers for the

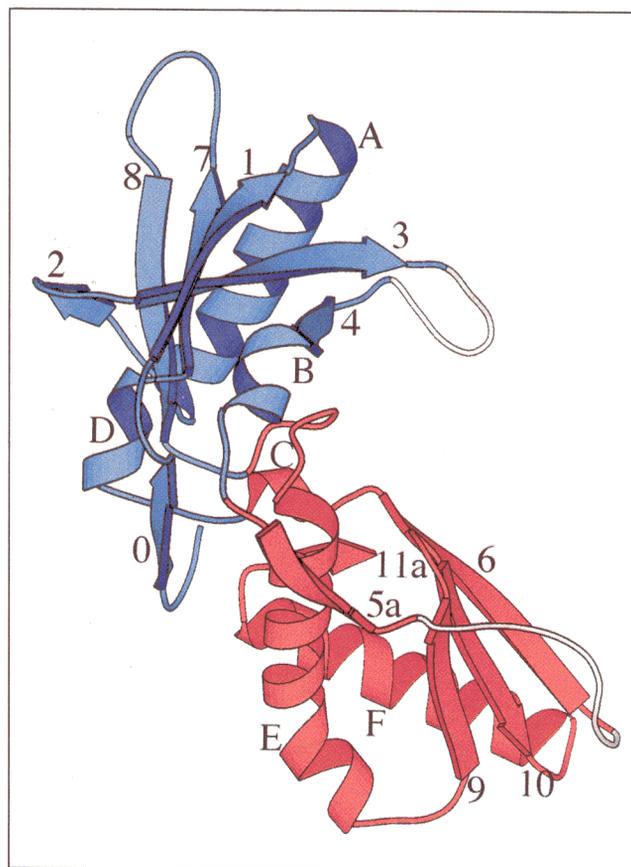


**Fig. 1.** Part of the  $2F_o - F_c$  electron density map of RT216 at 2.2 Å resolution contoured at  $1\sigma$  above the mean. Superimposed on the electron density is a pentapeptide with the sequence, from left to right, Thr84, Gln85, Asp86, Phe87 and Trp88. The positions of the carbonyl groups are well defined, here and in the rest of the model.

comparison with the p51 subunit are 1.47 Å for the fingers, and 1.76 Å for the palm subdomain. As is evident from Fig. 2 in Kohlstaedt *et al.* [1], the relative orientation of the subdomains in RT216 is very similar to that in the p66 subunit and drastically different to that in the p51 subunit. The rmsd in C $\alpha$  positions is 2.01 Å for 195 atoms in the p66 subunit. The difference between the RT216 molecule and the fingers and palm subdomains in the p66 subdomain can be seen largely as a rigid-body rotation of the subdomains with respect to each other. In RT216, the angle at the hinge (near helix C, see Fig. 2) between the two subdomains is 13.3° larger than in the p66 subunit of the RT–DNA–Fab complex. This difference could be explained by crystal packing interactions in the RT216 crystal and the effect of the interaction with the p51 subunit in the RT–DNA–Fab complex. Comparison of the RT–nevirapine C $\alpha$  coordinates with the RT–DNA–Fab C $\alpha$  coordinates shows that there is only a small difference between the relative positions of the fingers and palm subdomains (T Steitz, personal communication). Thus this difference cannot be explained by the interaction with DNA.

Two loops, located between strands  $\beta$ 3 and  $\beta$ 4 (residues 64–71) and strand  $\beta$ 5a– $\beta$ 6 (residues 94–104), have no electron density in the RT216 map and have not been modelled. There is only weak electron density for the loop that connects  $\beta$ 7 to  $\beta$ 8. This loop has been tentatively modelled as a polyalanine chain. Analysis with SDS-PAGE showed, however, that the crystals contained intact RT216 (data not shown). In the p66 subunit of the RT–DNA–Fab complex the  $\beta$ 3– $\beta$ 4 and  $\beta$ 5a– $\beta$ 6 loops are ordered because of the interaction with DNA and thus could be successfully modelled.

The contact area at the interface between fingers and palm subdomains is about 900 Å<sup>2</sup>. The helices  $\alpha$ B,  $\alpha$ C,  $\alpha$ D and  $\alpha$ E, as well as all three loops connecting the fingers and palm subdomains (namely  $\alpha$ B– $\beta$ 5a,  $\alpha$ C– $\alpha$ D and  $\beta$ 8– $\alpha$ E) are involved in an extensive interaction network between the two subdomains (Fig. 3a). The loop  $\beta$ 8– $\alpha$ E (amino acids 148–154) at the ‘base’ of the palm subdomain takes part in most of these interactions. The loop packs into a shallow bowl on the surface of the fingers domain in a way reminiscent of a ball-and-socket joint. Most of the interactions are hydrophobic, but a few hydrogen bonds are also present. There are two hydrogen bonds from the side chain of Asn81, one from the N $\delta$ 2 atom to the carbonyl oxygen of Gly152, and one from the O $\delta$ 1 atom to the amide nitrogen of Lys154. The side chain of Lys154 in turn forms a hydrogen bond to the carbonyl oxygen of Thr84, thus capping the carboxyl terminus of helix B. A hydrogen bond between O $\gamma$  of Ser156 and the carbonyl oxygen of Pro150, although not an interdomain interaction, could be important for determining the conformation of the  $\beta$ 8– $\alpha$ E loop. One further interaction of special interest is a hydrogen bond (connecting Ser3 and Trp212) between the closely located amino- and carboxy-terminal ends of RT216.

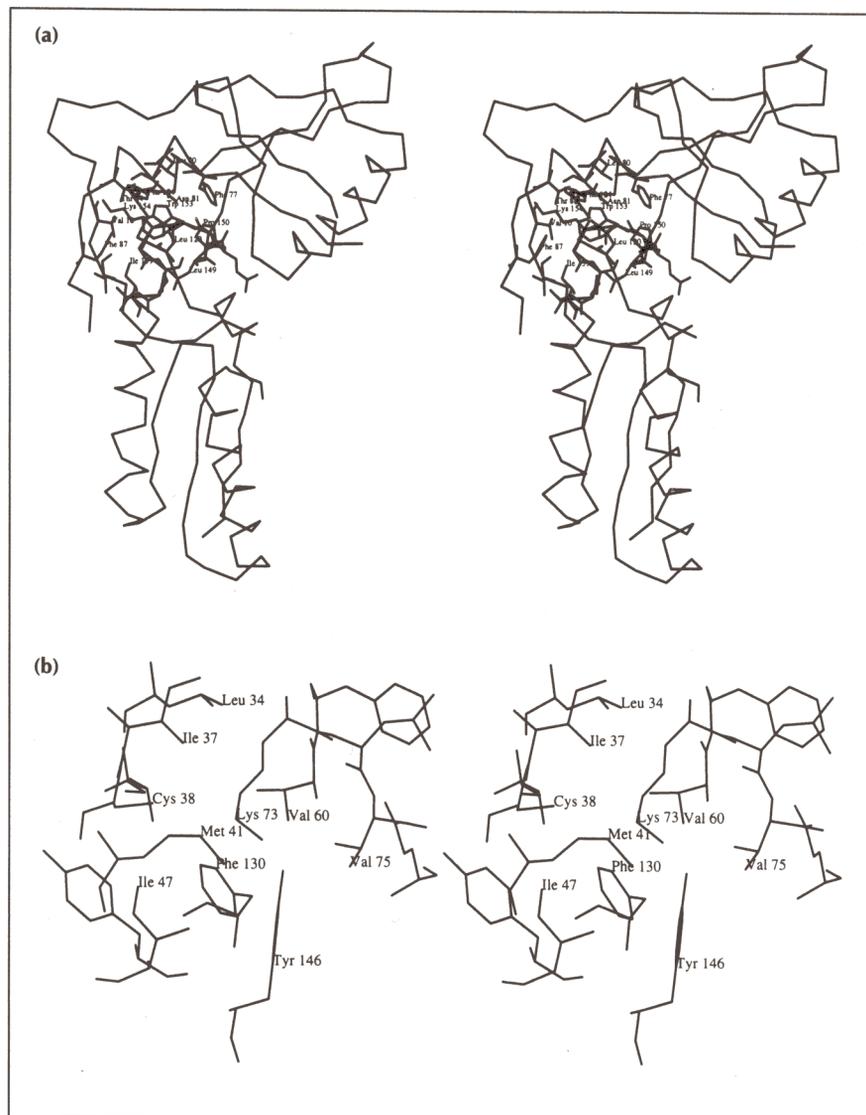


**Fig. 2.** The fold of RT216. The names and colours of the subdomains, fingers (blue) and palm (red), and the enumeration of the secondary structure elements follow the nomenclature used for the structure of HIV-1 RT complexed with hevirapine [1] and with DNA and an antibody fragment [5]. The  $\beta$ -strands are marked 0 to 11a and helices are labelled A to F. Two loops which have not been modelled are coloured grey and are located between strands  $\beta$ 3 and  $\beta$ 4 (residues 64–71) and strands  $\beta$ 5a– $\beta$ 6 (residues 94–104). This drawing and Fig. 5 were prepared using the program MOLSCRIPT [22].

Two other examples of interesting and important interactions are as follows. Firstly, there is a hydrogen bond (2.54 Å) between the amino acids Lys73 and Tyr146, both of which are buried in a hydrophobic environment (Fig. 3b). Lys73 is located at the amino-terminal end of strand  $\beta$ 4 and should be of importance for a solid anchoring of the  $\beta$ 3– $\beta$ 4 loop. The strand  $\beta$ 4 and the  $\beta$ 3– $\beta$ 4 loop have been modelled to be in contact with the nucleic acid template strand [5]. Secondly there is a stabilizing hydrogen bond (2.67 Å) between the main chain O of Ile167 and N $\epsilon$ 2 of His208 at the point where helix E makes a bend, caused by the presence of Pro170 in the helix.

#### Structure of the active site

Before describing the details of the active site, we consider the limitations of using a truncated form of HIV-1 RT. Unlike the RT heterodimer, RT216 does not have a significant enzymatic activity, probably because this truncated form of the enzyme is missing regions required for polymerase activity (e.g. the thumb



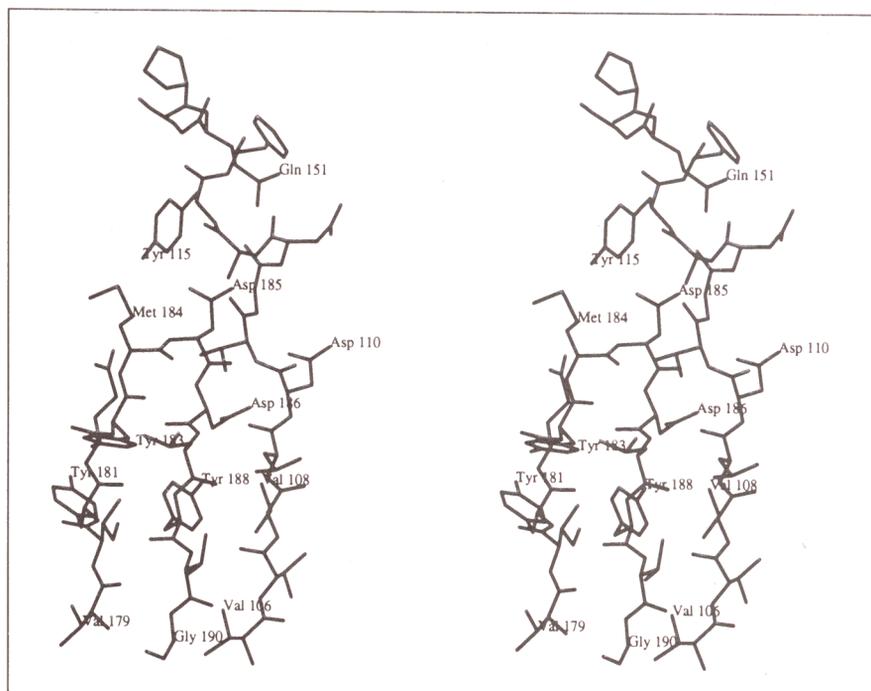
**Fig. 3.** Stereoviews depicting some interactions of particular interest in the RT216 structure. **(a)** The interface between the palm and the fingers subdomains. The C $\alpha$  trace of RT216 is shown together with side chains in the hydrophobic core between the two subdomains. **(b)** The side chain of Lys73 at the amino terminus of strand  $\beta$ 4 is completely buried in a hydrophobic environment where it makes a hydrogen bond to the side chain of Tyr146. This interaction could be important for anchoring the  $\beta$ 3– $\beta$ 4 loop.

subdomain). RT216 does not form any dimeric structure, and RT216 may not bind nonnucleoside inhibitors as strongly as the RT heterodimer, since parts of the binding sites (e.g. the  $\beta$ 12– $\beta$ 13 hairpin) of these inhibitors are missing. However, binding studies (K Bäckbro, B Strandberg and T Unge, unpublished data) show that RT216 binds the PETT inhibitor. Furthermore, the fold of RT216 is very similar to the fold of the corresponding part of the p66 subunit in the RT heterodimer, and we therefore argue that this structure can be used to deduce detailed structural information relevant to the physiologically active heterodimer. Obviously, the detailed information obtained from this high-resolution structure should be used in combination with the structures of the RT heterodimer [1,5,10].

The active site (Figs 4 and 5) is situated on a three-stranded antiparallel  $\beta$ -sheet composed of strands  $\beta$ 9,  $\beta$ 10 and  $\beta$ 6. In addition, critical active site residues are located in the turn between strands  $\beta$ 9 and  $\beta$ 10, and in the loop between strand  $\beta$ 8 and helix  $\alpha$ E. The residues in the  $\beta$ 9– $\beta$ 10 turn and on the surface of the sheet

are arranged as four parallel rows. Row number 1 contains the conserved residues Met184 and Asp185. Aspartic acid 185 is strictly conserved and critical for polymerase activity, as are two other aspartate residues in row 2, namely Asp110 and Asp186 [11]. In addition, row 2 contains the conserved residue Tyr183, which is strictly conserved in the retrovirus group; thus all residues in row 2 are highly conserved. Row 3 contains residues Tyr181, Tyr188 and Val108. All these have been shown to be involved in the binding of non-nucleoside inhibitors and in development of drug resistance [2,10]. The same is true for Val106 in row 4. The non-conserved residues Val179 and Gly190 are also included in row 4. One of the carboxylate oxygens of Asp185 has a hydrogen bond interaction with Tyr115 OH (3.18 Å), which places the other carboxylate oxygen in a position which should be suitable for enzymatic activity.

Glutamine 151 is located close to the active site and possibly is involved in template binding [5]. In the RT216 structure there is a hydrogen bond between the amide nitrogen (Ne2) of Gln151 and the main chain

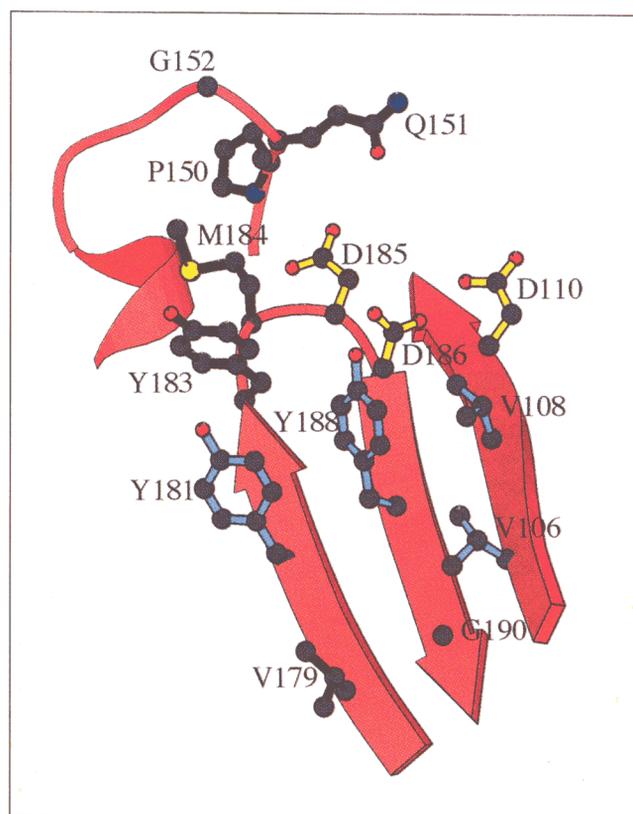


**Fig. 4.** Stereo diagram of the active site region of RT216. Most of the active site residues are located in a three-stranded antiparallel  $\beta$ -sheet built from strands  $\beta 9$ ,  $\beta 10$  and  $\beta 6$ . Part of the long connection between strand  $\beta 8$  and helix  $\alpha E$ , residues 150–152, which has been implicated in template binding [5], is also shown. The side chain of the conserved residue Gln151 is oriented by a hydrogen bond between the amide nitrogen and the main chain carbonyl oxygen of Gly112, thus leaving the amide oxygen free to interact with template DNA.

O of Gly112 (2.95 Å). Thus the amide oxygen of Gln151 is oriented in a way which permits binding to the template.

The residues in the  $\beta$ -sheet are in close contact with each other. Thus two of the strictly conserved catalytic residues, Asp186 and Asp110 in row 2, are in van der Waals contact with each other (3.86 Å). Moreover there is close packing (within van der Waals distance) between residues in rows 2, 3 and 4: Asp186 is in contact with Val108 (3.9 Å) and with Tyr188 (3.7 Å), Val108 contacts Tyr188 (3.4 Å) and Tyr183 contacts Tyr181 (3.3 Å). Other pairs of residues are close, but at slightly more than van der Waals distance [Tyr181 to Val179 (4.6 Å) and Val106 to Val108 (4.7 Å)].

Binding of a hydrophobic nonnucleoside inhibitor in contact with the residues Tyr181, Tyr188 and Val108 of row 3 [1] will probably cause a conformational change of all the residues in row 3 and also indirectly change the conformation of the residues in row 2 and row 1 that are critical for transcriptional activity. Another possible effect of the binding of the nonnucleoside inhibitors is a restriction of the motions of the p66 thumb subdomain [1,2]. Since the point mutations of residues in row 3 and 4 which are found in nonnucleoside drug-resistant mutants of RT (Tyr181→Cys, Tyr188→His, Val108→Ile and Val106→Ala [2]) have been shown to influence the activity only marginally [12], it seems as if the active site can tolerate small perturbations. Thus the most likely cause for the inhibitory effect of non-nucleoside inhibitors is the restriction of the relative movements of the subdomains. However, we cannot exclude the possibility that the distortions at the active site caused by inhibitor binding have an effect on enzymatic activity. We are currently carrying out a detailed comparison



**Fig. 5.** Positions of active site and nonnucleoside inhibitor binding residues. The residues on the surface of the  $\beta$ -sheet and in the turn  $\beta 9$ – $\beta 10$  are lined up in 4 rows. Row 1: Met184 and Asp185; Row 2: Tyr183, Asp186 and Asp110; Row 3: Tyr181, Tyr188 and Val108; Row 4: Val179, Gly190 and Val106. Strictly conserved residues are marked with yellow sticks and residues which are mutated in nonnucleoside drug resistant mutants of RT are marked with blue sticks. Atoms are in the following colour scheme: carbon black, oxygen red, nitrogen blue and sulphur yellow.

between the RT216 structure and the RT-nevirapine structure, aiming to describe the changes of the active site residues through inhibitor binding and to evaluate the effect on the enzyme activity.

#### The interaction between RT and DNA

The 3 Å resolution studies of the RT-DNA-Fab complex [5] provided important information on the interactions between RT and DNA. Using the higher-resolution structure of RT216 it has been possible to obtain a more detailed description of this interaction. By superimposing the palm subdomains of RT216 and RT-DNA-Fab (the p66 subunit), the phosphate backbone of DNA as located by Jacob-Molina *et al.* [5] could be positioned in relation to the RT216 molecule. All distances less than 7 Å between the centres of phosphate groups and atoms in the palm subdomain of RT216 have been listed in Table 1. A corresponding superposition of the fingers subdomains and a listing of distances between atoms was also done (Table 1). As is shown in Table 1, the secondary structure elements of RT216 which are involved in the interaction with DNA are  $\beta$ 5a,  $\beta$ 9, and  $\beta$ 10 in the palm subdomain,  $\alpha$ B in the fingers subdomain and the  $\beta$ 8- $\alpha$ E loop between the subdomains. This is in good agreement with the results from the studies of the RT-DNA-Fab complex. It is interesting to note that two of the residues that take part in hydrogen bonds at the subdomain interface, Lys154 and Asn81, may also be involved in DNA binding. Both of these residues are close to phosphate groups in the RT-DNA-Fab structure (Table 1). The  $\beta$ 3- $\beta$ 4 loop, which probably also takes part in the interaction with DNA [5], is flexible in the RT216 structure.

**Table 1.** Interactions between RT and DNA.

Phosphate groups	Atoms in RT216	Distances (Å) <sup>a</sup>	Secondary structure elements
C2P	Phe77 O	6.9	$\alpha$ B
	Arg78 O	5.7	$\alpha$ B
	Arg78 N	6.2	
	Arg78 Ne	5.5	
C3P	Asn81 O	5.1	$\alpha$ B
	Asn81 N $\delta$ 2	5.1	
	Asn81 O $\delta$ 1	6.2	
	Lys82 N	6.0	$\alpha$ B
	Lys154 N $\zeta$	5.7	$\beta$ 8- $\alpha$ E
C4P	Asp86 O $\delta$ 1	6.4	$\beta$ 5a
	Glu89 O $\epsilon$ 2	4.5	$\beta$ 5a
	Lys154 N $\zeta$	6.9	$\beta$ 8- $\alpha$ E
D18P	Tyr183 OH	6.9	$\beta$ 9
	Met184 O	6.3	$\beta$ 9- $\beta$ 10
	Asp186 O $\delta$ 1	6.8	$\beta$ 10
	Asp186 O $\delta$ 2	5.1	
	Tyr188 OH	5.9	$\beta$ 10

<sup>a</sup>All distances less than 7 Å between atoms in RT216 and centres of phosphate groups obtained from [5] are listed.

#### Biological implications

HIV-1 reverse transcriptase (RT) is essential for the virus, catalyzing the transformation of the single-stranded viral RNA genome into an RNA-DNA hybrid and then, after hydrolysis of the RNA part by RNase H, into double-stranded DNA. Earlier crystallographic studies of an RT-inhibitor complex and of an RT-DNA-antibody fragment complex have provided important contributions towards the understanding of the general mechanism of nucleic acid polymerization and of drug inhibition and resistance.

We have now determined the crystal structure of the amino-terminal half of RT (RT216) to 2.2 Å resolution. This structure reveals considerably more detailed three-dimensional information about the 'fingers' and 'palm' subdomains (including active site residues) than the two earlier structures.

The two subunits of the native RT heterodimer, p51 and p66, have different three dimensional structures, although each contains amino acid sequences at the amino terminus that are identical to RT216. The structure of RT216 shows that, when expressed alone, this fragment folds and orients the fingers and palm subdomains into the arrangement seen in the p66 subunit of the native RT p66/p51 heterodimer. This indicates that the arrangement of the p51 subdomain is induced by the interaction with the p66 subunit. The interactions (mainly hydrogen bonds and hydrophobic contacts) between the fingers and palm subdomains are strong and well defined in the RT216 structure.

Critical active site residues are located on a three-stranded antiparallel  $\beta$ -sheet and in a turn between two of these strands. These residues are arranged as four parallel rows. Mainly hydrophobic contacts provide the stabilizing forces between residues within the rows as well as between residues in neighbouring rows. Binding of nonnucleoside inhibitors to the residues Tyr181, Tyr188 and Val108 in row 3 and other residues in their vicinity would be expected to cause conformational changes of the critical active site residues in neighbouring rows, and thus have some influence on the polymerase activity. These inhibitors are potential therapeutic agents for use against HIV infection. The goal is to obtain maximal inhibitory capacity at the same time as minimizing side effects and development of drug resistance. Thus, detailed structural information about the mechanism of the inhibitors would facilitate the design of improved anti-viral compounds.

## Materials and methods

### Crystallization

RT216 was expressed in *Escherichia coli* with an extra amino-terminal methionine residue (unpublished data). Crystallization was performed by vapour diffusion. Drops consisting of 6  $\mu$ l of 0.9% (w/v) RT216 plus 3  $\mu$ l of the crystallization buffer [14% (w/v) PEG 8000, 0.1 M Tris HCl pH 8.0 and 0.02% (w/v)  $\text{NaN}_3$ ] were equilibrated against the same buffer at 4°C. Crystals usually appeared within two weeks and grew to a size of 0.7 mm $\times$ 0.2 mm $\times$ 0.15 mm within two months. The crystals belong to the tetragonal space group  $P4_3$ . The cell dimensions are  $a=b=98.2$  Å and  $c=31.7$  Å. Platinum and mercury heavy atom derivatives were made by soaking the crystals for three days in crystallization buffer containing  $\text{K}_2\text{PtCl}_4$ , tetrakis(acetoxymethyl)mercuric methane (TAMM) and di-acetoxymethylmercuric acetate (DMA), respectively (Table 2).

### X-ray data

X-ray data were collected using the synchrotron beam lines DW32 at LURE, Orsay and X11 at EMBL, Hamburg. The facilities at LURE [13] were kindly placed at our disposal by Professors R Fourme and J-P Benoît. Imaging plate scanners were used as detectors and the program DENZO [14] was employed to integrate the reflection intensities from the raw images. Essential details of data collection and processing are given in Table 2. Several native and heavy-atom soaked crystals were checked and many were discarded on the basis of poor diffraction quality. All crystals were cooled by a stream of air to about 4°C. The crystals were elongated along the  $c$ -axis and, although they tended to crack, it was often possible to shift them along the spindle axis to obtain more data from one crystal. The native data to 2.2 Å resolution were merged and used to determine and refine the structure.

The three heavy atom reagents  $\text{K}_2\text{PtCl}_4$ , TAMM and DMA were used to prepare five derivative crystals, called PT2, PT3, TAMM2, TAMM5 and DMA6, respectively. Each derivative data set was collected from a single crystal and processed with the Friedel pairs kept separate.

### Structure determination

One Pt site of the PT2 derivative was identified from the difference Patterson map and the remaining sites of this and other derivatives were found from appropriate cross-phased difference Fourier syntheses. The heavy atom parameters were refined with the program MLPHARE [15] and phases were calculated to 2.8 Å resolution. At this stage the correct enantiomorph was selected to be  $P4_3$  on the basis of good contrast between protein and solvent regions in the Fourier map. The overall figure of merit was 0.61 for 6166 acentric reflections and 0.75 for 858 centric reflections. The phases were further refined and extended in small steps from 2.8 Å to 2.4 Å with the program SQUASH [16] using simultaneously Sayre's equation, histogram matching and solvent flattening options assuming 45% solvent in the crystal.

The resulting map showed clear density corresponding to most of the protein chain, except for some loop regions. The models of fingers and palm subdomains consisting of  $\text{C}\alpha$  atoms from the available structures of reverse transcriptase [1,5] were overlaid on the map showing good agreement of secondary structure elements within both subdomains.

### Model building and refinement

The protein main chain could easily be traced in the map computed with the SQUASH phases, and correlated well with the  $\text{C}\alpha$  positions of the known structures. Initial chain tracing and modelling were performed using FRODO [17] whereas O [18] was used in all subsequent modelling. The initial model included about 90% of the main chain atoms and 67% of the side chains with an R-factor of 40.1% for data between 8.0 Å and 2.4 Å.

The model was refined by simulated annealing and individual temperature factor refinement with X-PLOR [19]. In total 13 rounds of refinement and remodelling in  $2F_o - F_c$  and  $F_o - F_c$  electron density maps were performed. The progress of refinement was monitored by calculating  $R_{\text{free}}$  [20] for a subset of 10% of the data. Remodelling and model quality control were

**Table 2.** Summary of data collection and processing.

	Native	PT2	PT3	TAMM2	TAMM5	DMA6
Synchrotron	EMBL/LURE	LURE	LURE	EMBL	EMBL	EMBL
Number of crystals	7	1	1	1	1	1
Resolution (Å)	2.2	3.0	3.5	2.7	3.0	2.8
Unique reflections	15 160	6044	3831	5000	5835	6271
Completeness (%)	96 <sup>f</sup>	97	97	58	93	82
Multiplicity	5.7	3.3	2.9	2.2	2.5	2.4
$R_{\text{merge}}(\%)^a$	9.4	9.4	11.1	5.7	8.2	7.3
$R_{\text{deriv}}(\%)^b$	-	22.0	19.2	20.4	27.5	24.1
$K_{\text{emp}}^c$	-	4.6	4.5	5.4	6.4	6.0
Phasing power <sup>d</sup>	-	0.8	1.0	1.2	1.2	1.2
Cullis $R^e$	-	0.80	0.73	0.58	0.58	0.57
Soaking concentration (mM)	-	1	3	1	1	1
Number of sites	-	4	4	1	1	1
Main binding ligands	-	M16 M41 M184 M208	M16 M41 M184 H208	C38	C38	C38

<sup>a</sup> $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$ . <sup>b</sup> $R_{\text{deriv}} = \sum |F_{\text{PH}}| - |F_{\text{P}}| / \sum |F_{\text{P}}|$ . <sup>c</sup> $K_{\text{emp}} = \sum |F_{\text{PH}+}| - |F_{\text{PH}-}| / \sum |F_{\text{PH}}| - |F_{\text{P}}|$ . <sup>d</sup>Phasing power =  $\sum |F_{\text{H}}| / \sum |F_{\text{PH}(\text{obs})}| - |F_{\text{PH}(\text{calc})}|$  for centric reflections. <sup>e</sup>Cullis  $R = \sum |F_{\text{PH}(\text{obs})}| - |F_{\text{PH}(\text{calc})}| / \sum |F_{\text{PH}}| - |F_{\text{P}}|$  for centric reflections. <sup>f</sup>77% completeness in the resolution range 2.24–2.20 Å.

greatly facilitated by using the program OOPS (GJ Kleywegt, and TA Jones, unpublished program). Water molecules were added in  $F_o - F_c$  peaks above  $3\sigma$  at chemically acceptable sites using WATERADD (S Knight, unpublished program) and also visually screened using the water options in O [18].

Refinement was initially carried out against a 2.4 Å data set. At a later stage of the refinement and modelling process a new data set including measurements to 2.2 Å was collected. Since the  $R_{\text{sym}}$  for data in the 2.4–2.2 Å resolution shell was greater than 30% we were uncertain about the information content of the high resolution data. To test whether these data should be included in refinement, the model was refined both with and without the 2.4–2.2 Å data. The free R-factor for data between 5.0 Å and 2.4 Å was then computed from each refined model; the 2.4 Å refined model gave an  $R_{\text{free}}$  of 27.9% whereas the 2.2 Å refined model gave an  $R_{\text{free}}$  of 27.0%. We thus concluded that the 2.2 Å data contributed significant information and should be used, although they were weak and showed poor agreement as judged by the R-factor on symmetry-related reflections.

The current model contains 196 residues and 90 water molecules with an R-factor of 20.8% and excellent stereochemistry (Table 3). Three loop regions in the structure are highly disordered. There is no density for the  $\beta 3$ – $\beta 4$  and  $\beta 5a$ – $\beta 6$  loops. There is weak but persistent density for the loop connecting  $\beta 7$  to  $\beta 8$  (residues 134–139) and this loop has been tentatively modelled as polyalanine.

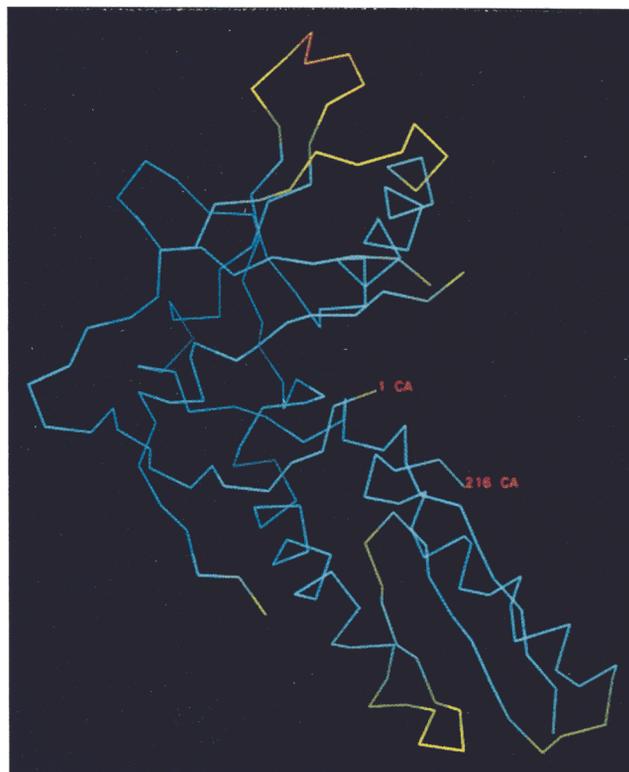
The average atomic temperature factor of the model is 45 Å<sup>2</sup>, indicating a high degree of flexibility in the molecule. Fig. 6 shows the distribution of temperature factors over the molecule. As expected, the most flexible regions are the loops at the distal tips of the two subdomains. The palm subdomain, which contains most of the polymerase active site residues, is more flexible than the fingers subdomain. This is reflected in an average temperature factor of 42 Å<sup>2</sup> for residues in the fingers subdomain as compared to 49 Å<sup>2</sup> for residues in the palm subdomain.

**Table 3.** Model refinement statistics.

Number of non-hydrogen atoms	1632
Number of solvent sites	90
Missing residues	21
Discretely disordered residues	None
Average B-value	45 Å <sup>2</sup>
Resolution limits	7.0–2.2 Å
R-factor	20.8%
Number of reflections	14 650
Free R-factor	27.2%
Number of reflections	1472
Rms deviation in	
Bond lengths	0.012 Å
Bond angles	1.3°

A Ramachandran analysis of the main-chain conformations (data not shown) shows that there are no residues in disallowed regions as defined in the program PROCHECK [21].

The atomic coordinates of all refined atoms in RT216 have been deposited with the Brookhaven Protein Data Bank.



**Fig. 6.** Temperature factor distribution mapped onto the C $\alpha$  trace of RT216. The temperature factor has been colour coded in the colours of the spectrum with blue for the lowest temperature factors and red for the highest.

**Acknowledgements:** This work was supported by the Swedish National Board for Industrial and Technical Development (NUTEK), the Swedish Medical Research Council, the Swedish Natural Science Research Council, the Faculty of Science and Technology, Uppsala University and by Medivir AB, Stockholm and Pharmacia Biotech AB, Uppsala. We are much indebted to our colleagues at Medivir and Pharmacia Biotech for excellent collaboration. We thank Edward Arnold for providing us with the C $\alpha$  coordinates of RT(p66/p51). Our colleagues Kristina Bäckbro, Kerstin Fridborg, Roshan Golmohammadi, Lars Liljas and Karin Valegård are thanked for stimulating discussions. We acknowledge the technical assistance of Thomas Carlsson and Olle Jäger and the valuable administrative help of Gunnell Johansson. We thank Tibor Illeni and Jan Zabielski for their skillful help with large-scale fermentations.

## References

1. Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. & Steitz, T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790.
2. Nanni, R.G., Ding, J., Jacobo-Molina, A., Hughes, S.H. & Arnold, E. (1993). Review of HIV-1 reverse transcriptase three-dimensional structure: implications for drug design. *Perspect. Drug Discovery Design* **1**, 129–150.
3. Emimi, E.A., Graham, D.J., Gotlib, L., Condra, J.H., Byrnes, V.W. & Schleif, W.A. (1993). HIV and multidrug resistance. *Nature* **364**, 679.
4. Balzarini, J., Karlsson, A., Pérez-Pérez, M.J., Camarasa, M.J. & De Clercq, E. (1993). Knocking-out concentrations of HIV-1-specific inhibitors completely suppress HIV-1 infection and prevent the emergence of drug-resistant virus. *Virology* **196**, 576–585.
5. Jacobo-Molina, A., *et al.*, & Arnold, E. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase

- complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA* **90**, 6320–6324.
6. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. & Steitz, T.A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**, 762–766.
  7. Lindahl, M., et al., & Amons, R. (1994). Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **13**, 1249–1254.
  8. Nagai, K., Oubridge, C., Jessen, T.H., Li, J. & Evans, P.R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515–520.
  9. Hegde, R.S., Grossmann, S.R., Laimins, L.A. & Sigler, P.B. (1992). Crystal structure at 1.7 Å resolution of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505–512.
  10. Smerdon, S.J., et al., & Steitz, T.A. (1994). Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **91**, 3911–3915.
  11. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867–3874.
  12. Boyer, P.L., Currens, M.J., McMahon, J.B., Boyd, M.R. & Hughes, S.H. (1993). Analysis of nonnucleoside drug-resistant variants of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.* **67**, 2412–2420.
  13. Fourme, R., et al., & Frouin, J. (1992). Bent crystal, bent multilayer optics on a multipole wiggler line for an X-ray diffractometer with an imaging plate detector. *Rev. Sci. Instrum.* **63**, 982–987.
  14. Otwinowski, Z. (1993). *DENZO: An Oscillation Data Processing Program for Macromolecular Crystallography*. Yale University, New Haven, CT.
  15. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering: Proceedings of the CCP4 Study Weekend 25–26 January 1991*. (Wolf, W., Evans, P.R. and Leslie, A.G.W., eds), pp. 80–86, SERC: Daresbury Laboratory, Warrington, UK.
  16. Zhang, K.Y.J. & Main, P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Crystallogr. A* **46**, 41–46.
  17. Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
  18. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
  19. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460.
  20. Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–474.
  21. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
  22. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of proteins. *J. Appl. Crystallogr.* **24**, 946–950.

Received: 18 Jul 1994; revisions requested: 10 Aug 1994;  
revisions received: 19 Aug 1994. Accepted: 25 Aug 1994.