

# Studies on crystal structures, active-centre geometry and depurinating mechanism of two ribosome-inactivating proteins

Qichen HUANG,\*† Shenping LIU,\* Youqi TANG,\* Shanwei JIN† and Yu WANG†

\*Department of Chemistry, Peking University, Beijing 100871, China, and †Shanghai Institute of Organic Chemistry, Academia Sinica, Shanghai 200032, China

Two ribosome-inactivating proteins, trichosanthin and  $\alpha$ -momorcharin, have been studied in the forms of complexes with ATP or formycin, by an X-ray-crystallographic method at 1.6–2.0 Å (0.16–0.20 nm) resolution. The native  $\alpha$ -momorcharin had been studied at 2.2 Å resolution. Structures of trichosanthin were determined by a multiple isomorphous replacement method. Structures of  $\alpha$ -momorcharin were determined by a molecular replacement method using refined trichosanthin as the searching model. Small ligands in all these complexes have been recognized and built on the difference in electron density. All these structures have been refined to achieve good results, both in terms of crystallography and of ideal geometry. These two proteins show considerable similarity in their three-dimensional folding and to

that of related proteins. On the basis of these structures, detailed geometries of the active centres of these two proteins are described and are compared with those of related proteins. In all complexes the interactions between ligand atoms and protein atoms, including hydrophobic forces, aromatic stacking interactions and hydrogen bonds, are found to be specific towards the adenine base. The relationship between the sequence conservation of ribosome-inactivating proteins and their active-centre geometry was analysed. A depurinating mechanism of ribosome-inactivating proteins is proposed on the basis of these results. The N-7 atom of the substrate base group is proposed to be protonated by an acidic residue in the active centre.

## INTRODUCTION

Ribosome-inactivating proteins (RIPs) are proteins which can inhibit protein synthesis in eukaryotic cells by catalytically damaging their ribosomes [1–4]. According to the definition of Stirpe and Barbieri [3], there are two types of RIP: type-1 RIPs are single-peptide-chain proteins, and type-2 RIPs are double-peptide-chain proteins (these two peptide chains are covalently linked by a disulphide bond). The two chains of type-2 RIPs are termed the A-chain (active chain) and B-chain (binding chain) respectively. The B-chain of type-2 RIPs is a lectin peptide; it can bind to the cell membrane and facilitate easy entry of the A-chain into the cell [1]. When type-1 RIPs, or the A-chain of type-2 RIPs, enter a eukaryotic cell, the protein can enzymically inactivate the cell's ribosome by rendering its 60 S subunit unable to bind to elongation factor 2, thus inhibiting protein synthesis. RIPs are so effective in doing this that it was estimated that once inside the cell just one molecule can kill a cell. This is why most type-2 RIPs are among the most toxic natural substances known. Type-1 RIPs show much less toxicity, because it is difficult for them to penetrate cells without a lectin domain.

RIPs are widely distributed among plants, where they were found in taxonomically unrelated families. In some plants the amounts of RIPs recoverable are very high [4]. To date, samples from hundreds of plant species have been found to have RIP activities, and RIPs may be present in several parts of the same plant. Also more than one kind of RIP can be isolated from some plants. All RIPs, even from very different plant species, are highly related. It was first pointed out by Zhang and Wang that the sequence of trichosanthin (TCS) is homologous with the A-chain of ricin [5]. Now it is clear that all type-1 RIPs of which the

sequences are known and the A-chains of all type-2 RIPs are homologous [6].

RIPs are of great interest both theoretically and because of their potential medical usefulness. Both type-1 and type-2 RIPs have been used for different purposes for a long time [1,7]. In China, TCS has been used as a mid-term aborting reagent since ancient times [7,8]. The type-2 RIPs, such as ricin, have been used as potential anti-cancer agents [1,9]. Recently, interest in RIPs has been growing on account of their anti-viral activities. In 1989, McGrath et al. found that TCS was specifically toxic to human-immunodeficiency-virus (HIV)-infected cells [10]. Many other RIPs have also been found to have this interesting property [11,12]. Their biological role in the life of the plant is also of great interest [4]. For these reasons a lot of effort has been devoted to studying RIPs. More than ten RIPs have been sequenced by different methods (for references, see the review by Stirpe et al. [4]). After a long search for a modification of eukaryotic-cell ribosomes caused by treatment with RIPs, these proteins were found to be specific RNA N-glycosidases which cleave only a single glycosidic bond of adenine residue 4324 among thousands of nucleotide residues of rRNA in the 28 S subunit of the eukaryotic ribosome [13,14]. Their substrates consist of a highly conserved -GAGA- loop, and intact ribosome particles show much higher substrate activity than does naked rRNA [14,15].

Crystallographic studies have also been done on several RIPs, e.g. ricin, for which the B- and A-chain structures have been reported at 2.5 Å (0.25 nm) resolution [16–18]. TCS molecules crystallized in different space groups have been studied for several years [19–21]. A primary X-ray-crystallographic study on  $\alpha$ -momorcharin (MMC) has been done independently in our

laboratory (Q. Huang, unpublished work) and in other laboratories [22]. Recently Monzingo and Robertus have reported the 2.5 Å resolution X-ray-crystallographic structures of complexes of ricin A-chain with formycin monophosphate and with a dinucleotide [23]. They have also deposited the structures of pokeweed (*Phytolacca americana*) anti-viral protein, a type-1 RIP, and similar complexes of it, with the Protein Data Bank. In their paper the binding mode of small ligands with ricin's active centre was described and a depurinating mechanism was proposed. Very recently, several groups claimed that they have finished the refinements of the structures of TCS and MMC and the complexes they form with some nucleotides. They proposed a mechanism for RIPs similar to that put forward by Monzingo and Robertus [23]. All these studies have provided much useful information on these useful proteins, but apparently the mechanism of depurination by RIPs presented by Monzingo and Robertus [23] has raised many doubts when some experimental facts were considered (see the Discussion).

The two RIPs studied here, TCS from the root tubes of *Trichosanthes kirilowii* (Tian Hua Fen) and MMC from the seeds of *Momordica charantia* (bitter gourd), are both type-1 RIPs. These two proteins were isolated from the same plant family (Cucurbitaceae). TCS and MMC were found to consist of 247 and 250 amino acid residues respectively [6]. In the present paper, five crystallographic structures of these two RIPs are reported. They are structures of the complex formed by TCS with ATP (called TCA), the complex of TCS with formycin (formycin is a kind of antibiotic and is an analogue of adenosine, but has a C–C glycosidic bond) (TCF), complexes formed by MMC with the above two small ligands (termed MCA and MCF respectively) and the native MMC (MCN). We determined these structures using multiple-isomorphous-replacement and molecular-replacement methods. They have been refined at 1.6–2.2 Å resolution, and good fits have been achieved to both X-ray and ideal-geometry terms. Small ligands in all complexes have been recognized and built on the basis of difference-Fourier-electron-density maps. These structures provide high-resolution structures of complexes of RIPs. The whole folding of these proteins was compared. We have also given a detailed description of the geometry of the active centres of these proteins. Interactions between ligand atoms and RIP atoms were analysed. They also provide interesting information about protein–nucleotide interactions. Those amino acids conserved among RIPs, and their relationships with active-centre structures, are discussed. A depurinating mechanism was proposed on the basis of these structures. This mechanism is different from that proposed by Monzingo and Robertus [23]. It shows clearly how the transition state of reaction is formed and stabilized. We also discuss the substrate specificity of RIPs. Our results will provide deeper and clearer understanding of these useful proteins.

## EXPERIMENTAL

### Materials and crystallization

TCS and MMC were isolated and purified at the Shanghai Institute of Organic Chemistry, Academia Sinica. All crystals used in the present study were grown using hanging-drop vapour-diffusion methods in multiwell plastic plates. All complexes were obtained by co-crystallization with the small ligands. The crystallizing conditions for MCN were as follows: the concentration of protein was 1% (w/v); the precipitant solution was 14% (w/v) poly(ethylene glycol) 6000 (purchased from Merck Co.), buffered with 0.3 M sodium citrate, pH 5.7. The crystallization conditions for MCF and MCA were the same as for MCN, except that their precipitant solutions were 26% (w/v)

**Table 1** X-ray data collection parameters

Parameter	Crystal name ...	MCN	MCA	MCF	TCA	TCF
Space group		<i>R</i> 3	<i>R</i> 3	<i>R</i> 3	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
<i>a</i> (Å)		130.97	131.60	131.60	38.24	38.24
<i>b</i> (Å)		130.97	131.60	131.60	76.74	76.74
<i>c</i> (Å)		40.14	39.70	39.70	79.25	79.25
Method*		AFC-5	AFC-5	AXIS	PF( <i>a</i> *)	AXIS
Resolution (Å)		2.6	2.4	2.0	1.6	1.9
Method*		AXIS	PF( <i>a</i> * + <i>b</i> *)	–	PF( <i>c</i> *)	PF( <i>b</i> *)
Resolution (Å)		2.2	1.8	–	1.6	1.6
Indep. reflection		11731	14716	13935	21538	22182
<i>R</i> <sub>merg</sub> † (%)		6.2	8.3	7.2	6.9	5.6
Completeness (%)		83.3	61.5	72.8	66.9	70.3

\* The reciprocal axis after PF represents the spindle axis along which the crystal oscillates.

†  $R_{merg} = \frac{\sum |I(i) - \langle I \rangle|}{\sum N(h) \cdot I(h)}$ , where  $I(i)$  is the intensity of an individual measurement and  $\langle I \rangle$  is the mean value of symmetry-related measurements. The summation is over all measurements.

KCl, buffered with 0.75 M sodium citrate containing 2 mM ligands (all ligands used in these experiments were purchased from Sigma Co., except where explicitly stated otherwise). Crystallization conditions for TCF or TCA were the same as for MCF or MCA, except that their precipitant was 20% (w/v) KCl solution and the pH value was 6.1. Usually large crystals suitable for X-ray analysis were obtained within 1 week. All crystallizations were done at 24 °C. All hanging drops contained 6 µl of protein solution and 4 µl of precipitant solution. The covered reservoir wells contained 1 ml of precipitant solution without ligands.

The heavy-atom-derivative crystals of TCA were obtained by following the procedure of Ma et al. on the X-ray analysis of the orthorhombic crystal form of TCS [21]. We soaked TCA crystals in 10 mM K<sub>2</sub>HgI<sub>4</sub> solution or 0.4 mM K<sub>2</sub>PtCl<sub>4</sub> solution for about 1 week each.

### Data collection and process

Crystal sizes were in the range 0.1–0.7 mm; the space groups and cell parameters of these crystals were determined with a Rigaku AFC-5 diffractometer or by precession photographs. Crystal data sets were collected using different methods. Table 1 lists the data statistics and the collecting methods of all data sets, except those of heavy-atom derivatives. Several sets of medium-resolution data of MCN and MCA were collected using an AFC-5 diffractometer. Medium-resolution data sets of TCF, MCF and MCN were collected on a Rigaku R-AXIS IIC area detector system [24]; when using this system the crystal-detector distances were set to 85 mm, at which the highest available resolution is 1.9 Å, with Cu–Kα radiation. The total oscillation angle ranges for TCF, MCF and MCN were 180, 167.5 and 160.0° respectively. The oscillation angle per frame was 2.5°, with about 35–40 min exposure time per frame. The reflections were integrated and reduced on a VAX station. High-resolution data sets for TCA, TCF, MCA, MCF were collected at the BLA62 hatch of Photon Factory in Tsukuba, Japan, using a large-radius screenless Wissenberger camera and synchrotron radiation [25]. Between 15 and 25 imaging plates were used to record reflections, and these imaging plates were read by a Fuji BA100 scanner. The radius of the chosen camera was about 430 mm. The X-ray wavelength was 0.900 or 1.000 Å. The oscillation angle for all crystals was 180° and the oscillation angle per imaging plate was

5–7°. Exposure time per frame was 10–20 s. Reflection data were then processed using the WEIS program on a FACOM computer.

All data sets from different sources for the same crystal were merged to give a unique set of data using the program PROTEIN [26]. Apparently the data sets collected using synchrotron radiation are of the highest quality and resolution. The absolute scale factors of these data were calculated by the program WILPLOT. It should be pointed out that differences in cell parameters between TCA and TCF or MCA and MCF are very small. They are smaller than the deviations between different sources for the same crystal or between different data frames for the same source. We calculated their mean values respectively and treated them the same in the following calculations. The mean cell parameters of TCA and TCF are almost the same as those reported in the orthorhombic native TCS crystal. The mean parameters of MCN showed, however, slight, but significant, differences from those of MCA and MCF. We found that, under the same conditions, MCA or MCF crystals diffracted more strongly than MCN and they had a greater endurance to X-ray exposure. It seems that the addition of small ligands to MMC made the MMC protein crystal more stable.

Unique sets of Hg-derivative and Pt-derivative crystals of TCA data were collected to 2.8 Å resolution on a diffractometer. The number of reflections was 5430 and 5294 respectively, with completeness 81% and 79% at 2.8 Å.

### Model building and refinement

The two sets of heavy-atom-derivative crystal data of TCA were used in different Patterson calculations. As reported in the 5 Å resolution crystal analysis of native TCS by Ma et al. [21], the  $K_2HgI_4$ -derivative crystal and  $K_2PtCl_4$ -derivative crystal give two Hg and two Pt sites respectively. In our crystals we found that the heavy atoms occupy almost the same positions as in 5 Å analysis. A set of multiple isomorphous replacement phases were calculated and the heavy-atomic parameters were refined by the program PROTEIN. The mean figure of merit of the phase set was 0.75. Fourier maps were then calculated using  $F_o$  as Fourier coefficients and weighted by the figure of merit of each reflection. A  $C^\alpha$ -tracing was made on mini-maps. It should be pointed out that the description of the ricin A-chain structure [18] and the peptide tracing report of native TCS [20] made our interpretation of the map much easier, especially in recognizing those higher-order secondary-structure regions. A protein model of TCA was then built on an SGI 70/GT graphic station using TOM/FRODO software [27]. Since differences of two small ligands do not make great changes on protein atoms in TCA and in TCF, we started the refinements of TCA and TCF from the same initial model. Alternating refinement cycles of energy-restrained least-square minimization and slow cooling protocol, simulating annealing, were carried out using the program package XPLOR [28]. In the refinements we always used the energy parameters for protein atoms suggested by Engh and Huber [29], since we have found these parameters give the best results for the ideal bond lengths and bond angles. After the ligands had been interpreted, based on difference Fourier electron density, and their model had been built, we had to set up the topology and parameter files for adenine base and for formycin respectively, since there are no standard files for these two ligands in XPLOR's topology and parameter files directory. The topology and parameters for adenine are derived from file TOPH11.DNA and PARAM11.DNA of XPLOR's topology parameter directory. These two original files contain all the hydrogen atoms and treat hydrogen bonds explicitly. We define adenine as an N-9 protonated tautomer. The ideal-geometry parameters of formycin

are from results of the crystal structure of formycin monohydrates [30]. Potential energy parameters for formycin are derived from PARAM11.DNA. Then these ligands were also involved in the refinements. The data were gradually extended to high resolutions. After the conventional  $R$  value decreased to about 0.26, unrestrained temperature factors of single atoms were also refined on these structures. Finally solvent molecules were located at stereochemically reasonable positions with different electron densities greater than three times the root-mean-square value. Within refinements many cycles of manual model rebuilding were made to correct global errors which could not be corrected by automatic refinement. Those atoms which could not be defined by the final electron density were set as dummy atoms which did not take part in Fourier calculations but whose positions were still restrained by ideal-geometry terms.

MMC structures were determined by molecular-replacement methods using refined TCS as the searching model. The sequence alignment of TCS with MMC shows that there are no insertions and deletions within the first 247 residues, except for 16 additional residues at the C-terminus of MMC [6]. Therefore we carried out the search using the whole TCS model. The rotation function was calculated using the fast rotation function program ALMN [31] in the CCP4 suite. Only one round of calculation was carried out before we were sure of obtaining the correct answer. The TCS model was put in an artificial triclinic unit cell with all cell parameters equal to 90, and a set of structure factors were calculated. Then they were combined with the observed structure factors of MCA to calculate the rotation function. The integrated Patterson radii were from 6 Å to 25 Å, and the resolution range of structure factors used was from 8 Å to 3 Å. The order of the Bessel function used was 60. The highest rotation function peak appeared at:  $\alpha = 63.9$ ,  $\beta = 86.8$ ,  $\gamma = 241.3$  (degrees, Euler angle). Its peak height was about 3.4 times of that of the second peak and 14 times that of the root-mean-square value. Then the TCS model was rotated by the above angles and was put in the unit cell of MCA to calculate the translation function. The translation function was calculated using the program TFSGEN in the CCP4 suite, which made use of all symmetry operations [32]. In the translation function map, only one peak in the translation function asymmetry unit appeared as positive, and this was about 1.7 times higher than the root-mean-square value. It appeared at  $X = 0.468$ ,  $Y = 0.599$  (fractional co-ordinates).

These rotation angles and translation vectors were confirmed by the following refinements. First the TCS model was transferred by the above rotation angle and translation vector into the unit cells of MCN, MCA and MCF. Then we truncated the 87 amino acid residues that differ between TCS and MMC in the refined TCS model to Ala and then carried out several cycles of rigid-body refinements. The following refinements were carried out as in TCA and TCF. Omitted side-chain atoms in MMC and small ligands in the complexes were located on the electron-density map. At the final step the unrestrained temperature factor refinements of single atoms were carried out on MCA. Bond- and angle-restrained temperature factor refinements of single atoms were carried on MCF. Group temperature factor (atoms of each residue were divided into two groups: CA, C and N form one group; the other atoms form another group) refinements were carried out on MCN.

Table 2 lists the refinement characteristics of these structures. Apparently these structures had achieved good results, both in crystallographic terms and in standard-geometry values. The estimated root-mean-square co-ordinate errors for MCN, MCA, MCF, TCA and TCF are 0.30, 0.25, 0.25, 0.18 and 0.20 Å respectively, according to Luzatti's method [33]. It was noticed that the data sets of TCA and TCF were of the highest quality





**Figure 2** Superimposition of CA atoms of TCS (thin line) on those of MMC (thick line)

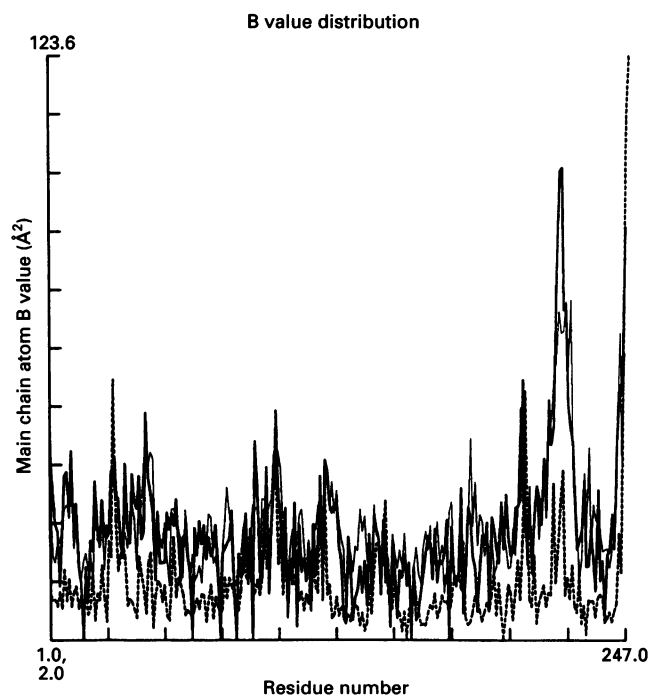
A binding adenine is shown in the cleft formed by two domains of TCS or MMC.

ion-pairs and hydrophobic interactions observed between main-chain atoms or side-chain atoms. These interactions stabilize secondary structures and the whole folding of these two proteins.

Since TCS and MMC are from the same plant family and there are no amino-acid-residue deletions or insertions between them, it is expected that there will be no large deviations between their structures. Figure 2 shows the superimposition of TCS with MMC. The root-mean-square deviation between all the CA atoms of these two proteins is 0.8 Å after least-squares fitting. The largest deviation occurs at external loop 218–222, which connects sheets  $\beta 2_1$  and  $\beta 2_2$ . Both in TCS and MMC, this external loop is highly disordered (see Figure 3). Loops 42–45, 141–145, 176–178 also show large deviations. After omitting these loops the root-mean-square deviation between CA atoms of the two proteins decreased to 0.5 Å. We found that other residues which show large deviations between TCS and MMC all possess high temperature factors and are at the surface, and not within, secondary structures.

The small ligands did not change TCS and MMC conformations greatly. The root-mean-square deviations between all CA atoms of MCF, MCA and MCN after least-squares fitting were about 0.2 Å, and the largest deviation was 1.3 Å. The root-mean-square deviation between all CA atoms of TCA and TCF was 0.12 Å and the largest deviation was 0.5 Å.

Both in TCS and MMC, residues within the secondary structure regions are much more ordered than those within loops. Figure 3 illustrates the temperature-factor distribution of MCN, MCA and TCA per residue. We can see from it that the high temperature-factor peaks appear at the external loop 218–221 and the C-terminus both in MMC and TCS. In all MMC structures we cannot locate the residues after 246. They may be completely disordered. It is clear that TCS shows more order than MMC and that the native MMC is more disordered than its complexes. It confirms the observation that MCA and MCF diffract more strongly than MMC.



**Figure 3** Comparison of temperature-factor distribution per residue of MCA (thick line), MCN (thin line) and TCA (thick broken line)

#### Active-centre geometry of TCS and MMC

In all four complexes we have located and built the small ligands. The positions of ligands in TCS and MMC and the interactions

of their atoms with the atoms of the proteins clearly indicate active centres of these two proteins. The residues constituting the active centres of TCS and MMC are the same. They are: 69–71, 83–85, 108–114 and 155–163. Residues ranging from 189 to 192 form the outer edge of the active centres. These residues are located at, or near, sheets  $\beta_4$ ,  $\beta_5$  and  $\beta_6$ , and at or near helices  $\alpha_2$ ,  $\alpha_3$  and  $\alpha_4$  respectively. Most of them are at the start or end of the secondary-structure region or are at the joint of two secondary-structure regions.

Figure 4 gives a series of detailed pictures of the active centres of TCS and MMC. Those residues mentioned above form a 10 Å-deep and 20 Å-long active slit at the bottom of the cleft formed by the interface of two domains. Sheets  $\beta_4$  and  $\beta_5$  form two walls of the slit, with sheet  $\beta_5$  as the bottom. Helix  $\alpha_5$ , which crosses over the slit roughly vertically, forms one end of the slit. In complexes of both TCS and MMC, except for TCF, this slit is divided into two parts by the side chain of Tyr-70; each has its opening to the cleft. The first part is like a half-opened box. Side chains of residue 155 and the side chain of residue 70 at two alternative conformations form two opposite walls of this box. Main chains and side chains of residues 69–71, the main chain of residue 109 and side chains of residues 110–114, and the side chains of residues 160 and 163 form the other three walls. Side chains of residues 83 and 85 are positioned at its far corner. Its opening is around the side chains of residues 111, 160 and 163 and faces towards the cleft formed by the protein's two domains. Residues 189–192 are positioned near this opening. The second part of the slit has an opening around the side chain of residue Glu-85, the side chain of residue 70 in one of the alternative conformations and the side chain of residue 108. This opening also faces towards the cleft. This part is jointed with the first part at the main chain of residue 109. It is formed by the peptide plane of residue 109, the main chains and side chains of residues 70, 83, and 85. The side chains of residues 71, and 155 are positioned at its far corner.

Compared with the MCN, the largest changes in active centres of the complexes occur at the side chain of residue Tyr-70: its side chain has two alternative conformations in complexes TCA, MCA and MCF, but only one in MCN (see Figure 4a and the analysis below). Figure 4(b) shows that active centres of TCS and MMC are almost the same, except at residue 159. In MMC the amino acid of residue 159 is Ala, but in TCS it is Ser [6,34,35]. This difference changes a hydrophobic residue in the first part of the active centre of MMC into a hydrophilic residue in that of TCS.

In complexes formed with ATP (MCA and TCA), it was found that only adenine, the hydrolytic product of ATP, existed in the active centres of TCS and MMC; in other words the nucleotide existed in the active centres of TCS and MMC with its N-glycoside bond cleaved. There were several reasons why we draw this conclusion. First, in difference Fourier maps only the electron density corresponding to adenine base can be seen. It was observed that the difference electron density smoothly enveloped the base (Figure 4c). The second reason was that if these ligands in TCA or MCA were not adenine, there would be great steric hindrance between ligand atoms and protein atoms. We built the adenine model on the electron density data and found that adenine occupies very similar positions in TCA and MCA (Figure 4b). At these positions we found that, if the N-glycoside bond of ATP were not cleaved, no matter how the N-glycoside torsion angle C-4–N-9–C-1'–O-4' rotates and what the puckering mode of the ribose ring is, the ribose would cause steric hindrance with the protein's atoms. (For example, the C-1' atom of ribose would be too close to the OE1 atom of Glu-160 and NH-1 atom of Arg-163 with adenine at those positions).

With adenine at those positions, the distance between the ligand's N-7 atom and the carbonyl O atom of residue 109 of the protein would be too short if the glycoside bond of the nucleotide were uncleaved (see Table 3 and analysis below).

Table 3 lists all the interactions involved between active centre atoms and ligand binding. From it we can see that interactions involved in the binding of adenine with active-centre atoms of TCS and MMC include hydrophobic force, aromatic stacking and hydrogen bonds. Most of these interactions are also observed in ricin and pokeweed complexes [23]. It should be pointed out that residue Tyr-70 in TCS and in MMC and, as we think, in all RIPs, plays a special role. As mentioned above, in complexes TCA and MCA the side chain of Tyr-70 occupies two alternative positions (Figure 4a). At both of these two positions its side chain is stacked with adenine. One position was near that in native protein MCN. At this position the stacking interaction between the Tyr-70 side chain and adenine is weak: there is a small tilt angle between the two planes and the inter-plane distance is relative large (within the range of 3.6–3.9 Å in TCA or MCA). The same occurs at Tyr-111. At another position which is far from that in MCN the side chain of Tyr-70 rotates about 57° around its CA–CB single bond from the first position. This rotation makes the stacking interaction with adenine more favourable: the phenoxy plane is almost parallel with the adenine plane and the inter-plane distance becomes shorter (between 3.0 and 3.4 Å; a typical value is 3.4 Å [37]). We have refined the occupations of these two conformations of Tyr-70 and found they were about 0.4 and 0.6 respectively. In the former conformation, temperature factors of Tyr-70 side-chain atoms are higher than those for the latter one. The rotation of the Tyr-70 side chain also makes the first part of the active site a complete hydrophobic binding pocket: the phenoxy group blocks the connection between the two parts of the active slit by forming an interaction between it and the Gly-109 C–O bond. That means that the adenine is trapped in the first part covered by the Tyr-70 side chain. Transferring from these two states the Tyr-70 side chain and Ile-155 sandwich the adenine. Careful analysis shows that the latter conformation of Tyr-70 is not only stabilized by stronger stacking with adenine, but also by hydrogen bonds between its OH atom with OE1 and OE2 of Glu-85 (see Figure 4a). At this position the side chain of Tyr-70 positions itself with its OH atom just vertically above the C=O double bond of Gly-109, which is almost co-planar with the adenine base. The OH atom was about 3.0–3.3 Å from these two atoms in TCA and MCA. So we can say that the Tyr-70 side chain also interacts with the C=O double bond through their  $\pi$ -electrons. This interaction may also be a part of a large aromatic stacking system that involves stacking of adenine, and it also stabilizes the latter conformation.

Apparently the hydrogen-bond patterns between adenine and protein in TCA and MCA are specific towards adenine base (Figure 4b and Table 3). It is interesting that the distance between N-7 and the Gly-109 carbonyl O atom in TCA and MCA is about 0.5–0.7 Å shorter than the minimum van der Waals contact distance of O and N atoms (3.2 Å). It can only occur when the N-7 atom of adenine is protonated and a hydrogen bond is formed between these two atoms.

In contrast to the situation in TCA and MCA, in TCF and MCF crystals formycin occupies different positions in their active centres. In TCF the electron density corresponding to formycin is very low. The ligand shows high disorder. Only those atoms involving hydrogen bonding can be accurately defined. Also formycin's position is very different from the adenine residue in TCA or MCA (Figure 4d). Its base is located in the second part of the active slit. Its C-8 and N-7 atoms are

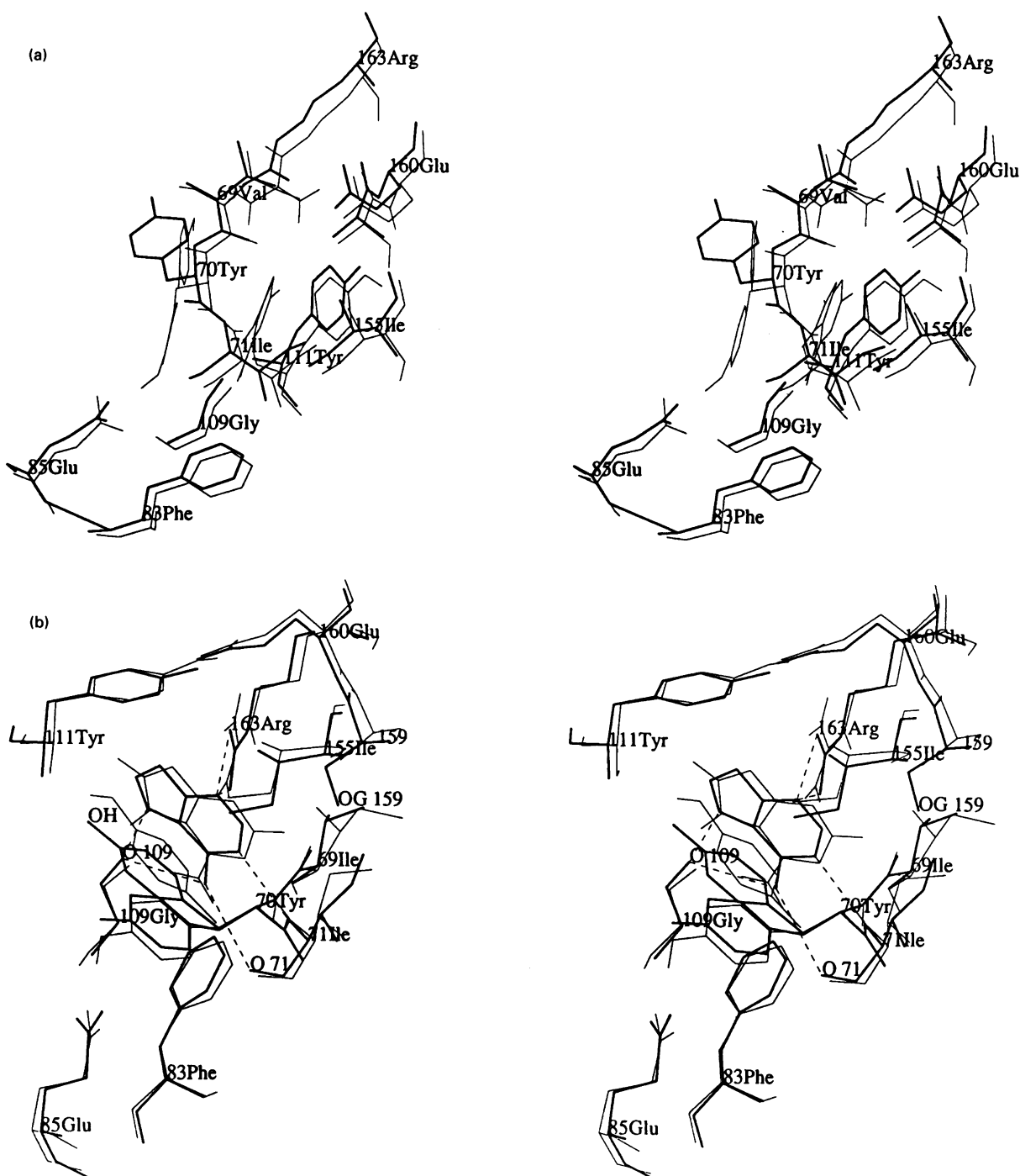
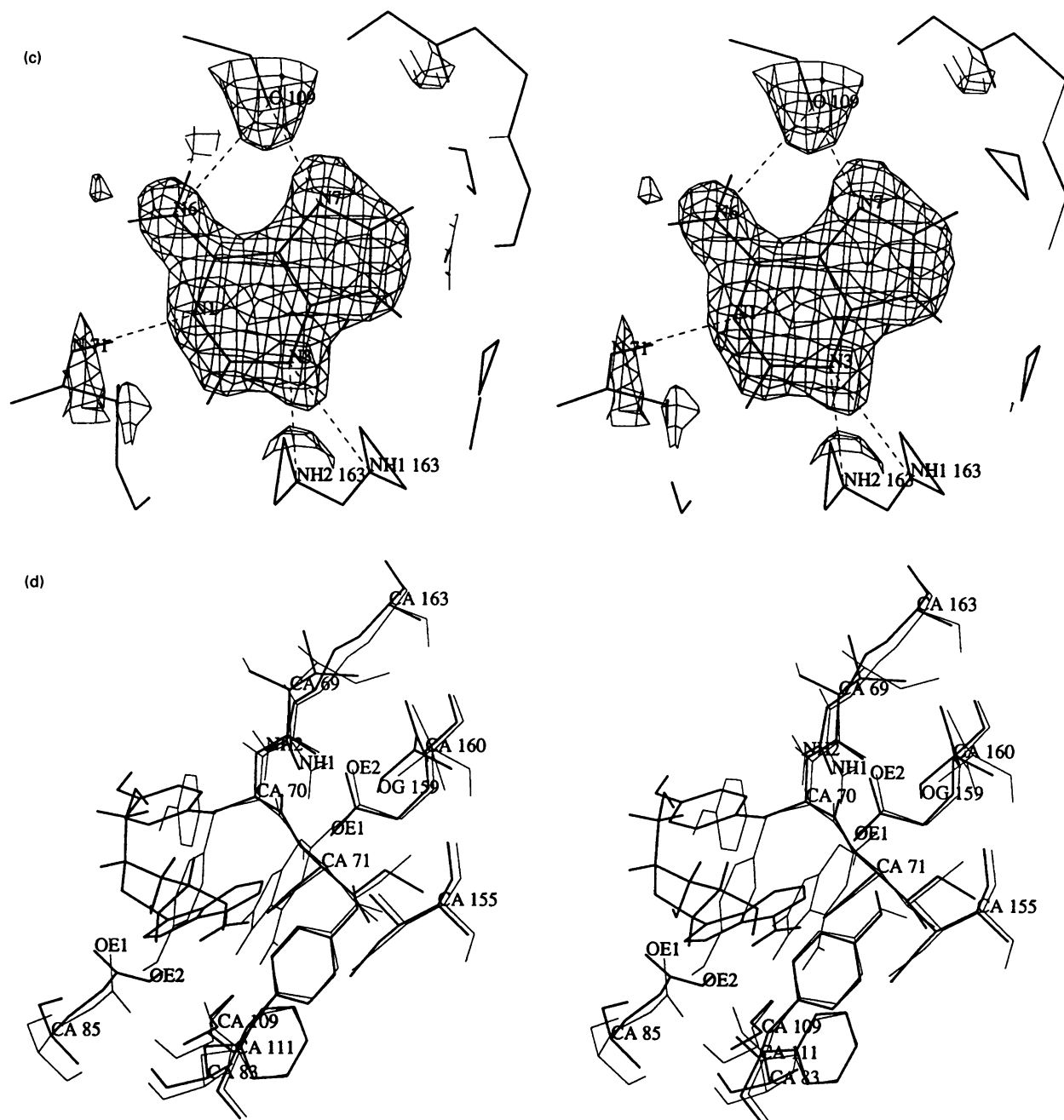


Figure 4(a, b) For legend see page 293.

near the opening of that part and its ribose extends into the cleft through the opening of the first part. Apparently the interactions between its base and atoms of the second part are also very different from those in TCA and MCA (see Table 3). In TCF the side chain of Tyr-70 occupies only one position near that in MCN. With the base at this position the rotation of the side chain of Tyr-70 mentioned above is apparently prevented by the base. This means that the first part of the active centre of TCF can no longer form a hydrophobic pocket. On the other hand, in MCF formycin was well defined by electron density compared

with in TCF. Atoms of the ligand base in MCF were ordered with a very low B value. It occupies a position in the first part of the active slit near that of adenine in TCA or MCA, but a shift between them can be seen (see Figure 4e). This shift of formycin can be said to be a slight rotation around the N-6 atom with respect to adenine in MCA. The largest shift distance of the base is 1.4 Å, at the C-9 atom. The base positions itself apparently shallower in the protein's binding pocket than adenine does in TCA or MCA. The side chain of Tyr-70 was also observed to have two conformations, and they are at almost the same



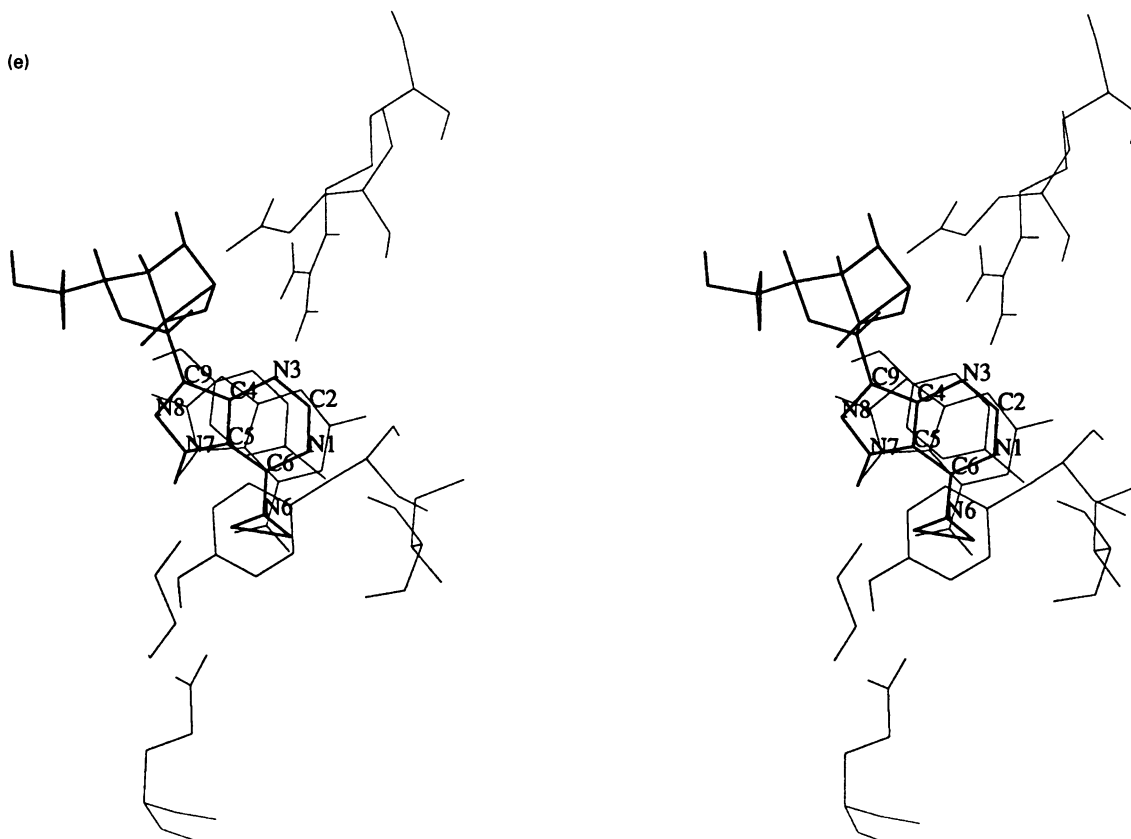
**Figure 4(c, d)** For legend see facing page.

positions as in TCA or MCA. Apparently the relative shift of formycin base in MCF with respect to adenine in MCA is caused by avoiding the steric hindrance between ribose atoms and atoms of the active centre of MMC. This confirms our conclusion that, in MCA and TCA, only adenine base exists (after superimposition, we also found the relative shifts of the base atoms of formycin monophosphate in complexes of ricin A-chain and pokeweed anti-viral protein).

Interactions between the ribose moiety of formycin and protein were weak both in TCF and MCF, but especially in TCF. Since high disorder of ribose atoms occurs in TCF, interactions between ribose atoms and protein atoms cannot be defined accurately. We can only say that the ribose of formycin in TCF is suspended

in the cleft towards which the opening of the first part of the active centre faces, and very few contacts between its atoms and protein atoms are observed. The configuration of ribose cannot be accurately defined in X-ray terms either. In MCF, however, although the atoms of ribose possessed high temperature factors, some interactions can be found between them and the protein atoms. The puckering mode of ribose was envelope C-2'-endo and the glycoside torsion angle C-4-C-9-C-1'-O-4' was  $-163^\circ$ , an *anti* conformation according to the definition of [37]. The observation that the conformation of formycin ribose is an *anti* one is different from that observed in free formycin. In free formycin, a *syn*, or an intermediate between *syn* and *anti*, conformation of the glycoside torsion angle was observed [30].





**Figure 4** Series of pictures of active centres of MCN, MCA, MCF, TCA and TCF

In these pictures of protein polar hydrogen atoms and all ligand hydrogen atoms, only those in MCA or MCF are shown. (a) Comparison of active slit of MCN (thick line) and MCA (thin line). Residues constituting the active centre are labelled. The binding adenine in TCA locates in the first part of the active slit. (b) Comparison of active slit of MCA (thin line) and TCA (thick line). The hydrogen bonds between binding adenine atoms and atoms of the protein's active slit are illustrated by a broken line (except between N-3 and Arg-163 NH1 and NH2). (c) Superimposition of adenine model in MCA on the final electron density. (d) Comparison of active slit of TCF (thick line) and MCA (thin line). Residues forming the second part of the active slit at which the base of formycin in TCF locates are labelled. (e) Superimposition of formycin in MCF (thick line) on active slit of MCA (thin line).

The transfer from a *syn* to an *anti* conformation of formycin ribose in MCF is caused by the active-centre structure of MMC: the ribose will cause steric hindrance in the *syn* conformation.

After we had found the differences in the mode of binding of ligands with different proteins, we tried to account for the different binding modes. As mentioned above, the only difference between active centres of MMC and TCS is the replacement of Ala-159 in MMC by Ser-159 in TCS (Figure 4a). This replacement changes a hydrophobic residue into a hydrophilic one. Furthermore, the OG atom of Ser-159 in TCS is located in the binding pocket formed by the first part of the active centre of TCS and forms a strong hydrogen bond with the Val-69 O atom. We found that if formycin had occupied the same position in TCS as in MMC, its C-2 atom would be in contact with Ser-159 OG at a distance of about 3.2 Å. There would be a slight steric hindrance between them. The binding energy was further decreased by burying a hydrophilic atom in the interface. Except for these, since the C-2 atom would be just between Ser-159 OG and Val-69 O, the hydrogen atom attached to formycin C-2 would be near the H atom attached to Ser-159 OG and would cause great steric hindrance to it. In other words, it would disturb the above-mentioned hydrogen bond. This would greatly decrease the binding energy of formycin with TCS. Since the occupation and temperature factors are high-correlated para-

meters, in these complexes we did not refine the occupation factor of the small ligands.

#### The relationship between conserved residues and the active centre of RIPs

As pointed out by many researchers, all type-1 RIPs and A-chains of type-2 RIPs are homologous [3,5,6]. There are many residues which are conserved among all RIPs. In Table 4 we select several RIPs and list parts of their amino acid sequences that are conserved among all RIPs or which form the active centre. The selected sequences are: TCS [34], MMC [35], ricin A-chain [38], saporin-6 [39], pokeweed anti-viral protein [40], luffin [41], abrin A-chain [42] and barley RIP [43].

From Table 4 we found that many conserved residues among RIPs are involved in forming the active centre described by us. Among those conserved residues, we found Tyr-70, Glu-160, Arg-163 and Trp-192, which directly form the active centres of TCS and MMC. As mentioned above, Tyr-70 plays an important role in binding with adenine. In TCS and MMC conserved residues Glu-160 and Arg-163 are directly involved in forming the active centres of these two proteins: the side chains of these two residues form the opening of the binding pocket and the ribose moiety of the ligand is near their atoms. Arg-163 is also

**Table 3 Interactions involved in ligand binding of TCS and MMC**

Values in parentheses after aromatic stacking residues are minimum distances of atoms in two plains (in Å). Values in parentheses after hydrogen bonds are distance (in Å) between hydrogen bond donor and acceptor and the angle of donor-H...acceptor (in °). 70' represents the alternative conformation far from that in MCN. 70 represents near that in MCN.

Crystal	Interaction		
	Hydrophobic	Aromatic stacking (Å)	Hydrogen bonds (Å, °)
MCA	Ile-155, Ile-71 Ile-69, Leu-114 Phe-83, Ala-159	Tyr-111 (3.8–4.6) Tyr-70 (3.6–4.4) Tyr-70' (3.4–3.6)	N-6-Gly-109 O (2.65, 172) N-7-Gly-109 O (2.74, > 160) Ile-71 N-N-1 (2.74, 176) Arg-163 NH-1-N-3 (3.2, 144) Arg-163 NH-2-N-3 (3.4, 130)
TCA	Ile-155, Ile-71 Ile-69, Ile-114 Phe-83	Tyr-111 (3.8–4.4) Tyr-70' (3.0–3.5) Tyr-70 (3.8–4.1)	N-6-Gly-109 O (3.25, 151) N-7-Gly-109 O (2.47, > 175) Ile-71 N-N-1 (2.84, 166) Arg-163 NH-1-N-3 (3.1, 141) Arg-163 NH-2-N-3 (3.0, 143)
TCF	Ile-155, Phe-83 Ile-71	Gly-109 C, O, N, CA (3.1–3.4)	Ile-71 N-N-1 (3.05, 143) N-6-Glu-85 OE2 (3.0, 156)
MCF	Tyr-111, Gly-109 Ile-71, Phe-83 Ile-155, Ala-159 (base) Ile-69, Leu-114	Tyr-70 (3.8–4.1) Tyr-70' (3.4–3.9) Tyr-70 (3.8–4.4) Tyr-111 (4.1–4.6)	N-6-Ile-71 O (2.67, 168) N-6-Gly-109 O (2.7, 177) N-7-Gly-109 O (2.8, 156) Ile-71 N-N-1 (3.13, 165) Arg-163 NH-1-N-3 (3.2, 143) Arg-163 NH-2-N-3 (3.3, 147)
(ribose)	Trp-192, Tyr-111		O2'-Arg-163 NH-2 (3.15, 114) O2'-Arg-163 NH-1 (3.47, 156) O2'-Try-70' OH (2.53, 155) O3'-Glu-189 O (3.12, 161) O3'-Arg-163 NH-1 (3.35, 143)

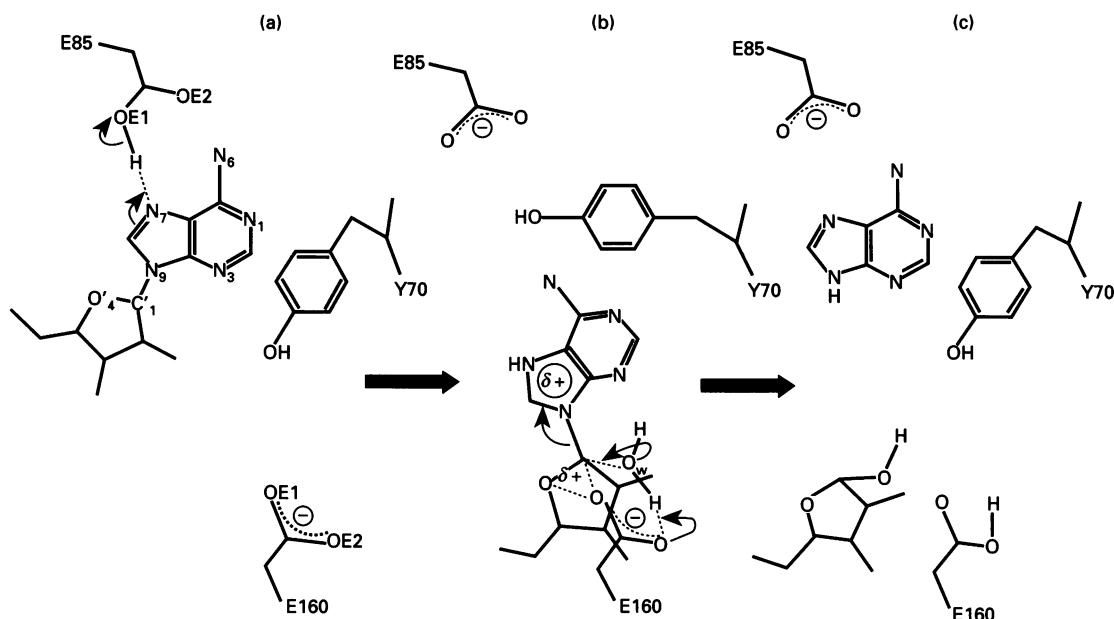
**Table 4 Constant residues and residues in active centre of some RIPs**

The Table lists residues constant in all RIPs or residues forming the active centre. Numbers are put on constant residues or the start residues of residue regions forming the active centre of TCS and MMC with TCS as reference. Asterisks at the top of columns denote residues directly forming the active centre. Asterisks in sequences denote constant residues. + Means residue(s) insertion and - means residue(s) deletion. The sequence alignment is based on [6] and our modification. Abbreviations: TCS, trichosanthin (*Trichosanthes kirilowii* Maxim, Cucurbitaceae); MMC,  $\alpha$ -momorcharin (*Momordica charantia*, Cucurbitaceae); RTA, ricin, A-chain (*Ricinus communis*, Euphorbiaceae); SAP, saporin SO-6 (*Saponaria officinalis*, Caryophyllaceae); PAP, pokeweed anti-viral proteins (*Phytolacca americana* seeds, Phytolaccaceae); LUF, luffin- $\alpha$  (*Luffa cylindrica*, Cucurbitaceae); ABR, abrin A-chain (*Abrus precatorius*, Leguminosae); BPI, barley protein-synthesis inhibitor (*Hordeum vulgare*, Gramineae).

	14	17	22	70	83	85	108	114	155	160	163	189	192											
			*	*	*	*	*	*	*	*	*	*	*											
TCS	Y	G	V	F...L	R...V	Y	I...F	N	-	E...S	G	N	Y	E	R	L...I	Q...S	E	A	A	R...E	N	S	W
MMC	*	G	M	*...L	*...I	*	I...F	N	-	E...S	G	N	Y	E	R	L...I	Q...A	*	A	A	*...E	N	S	*
RTA	*	T	N	*...V	*...A	*	V...F	H	P	D...G	G	N	Y	D	R	L...I	Q...S	*	A	A	*...E	N	S	*
SAP	*	S	S	*...I	*...L	*	V...F	R	S	E...E	Y	T	Y	D	S	I...I	Q...A	*	V	A	*...E	V	S	*
PAP	*	A	T	*...L	*...L	*	V...F	N	-	D...N	G	L	Y	P	T	L...I	Q...S	*	A	A	*...E	N	-	*
LUF	*	S	E	*...L	*...L	*	I...F	N	-	E...S	G	N	Y	E	K	L...L	Q...A	*	A	S	*...E	N	S	+
ABR	*	K	Q	*...L	*...A	*	V...L	R	-	D...Y	G	T	Y	G	D	L...I	Q...A	*	A	A	*...E	N	N	*
BPI	*	A	T	*...I	*...I	*	L...W	W	-	E...-	G	D	T	D	K	L...L	-...N	*	A	T	*...Q	+N	G	*

observed to be directly involved in binding adenine (Table 3). Figure 4(a) shows that the OE2 of Glu-160 and NH2 of Arg form an ion-pair. This ion-pair may fix the Glu-160 side chain to fulfil its enzymic task (see below). Tyr-70, Glu-160 and Arg-163 were also proved, by site mutagenesis, to be directly involved in the RIPs' enzymic activities [44,45]. Conserved residue Trp-192 is located near the opening of the first part of the active centre of RIPs. We guess it may interact with atoms near the reactive residue of the ribosome. Among those residues forming the active centre we have also found several half-conserved residues. Residue 155, which contacts with the adenine plane directly, is

either Ile or Leu (the side chains of which are all hydrophobic) in all RIPs. The same applies to residue 114. Among all RIPs residue 69 is one of Ile, Val, Leu and Ala, and residue 71 is one of Ile, Val and Leu. Most RIPs have a Glu and an Asn at residues 189 and 190. Residue 111 is Tyr in all RIPs, except for barley RIP. So in most RIPs, the residue will make its contribution to the binding of substrate by aromatic stacking. These three residues are around the opening of the first part of RIPs' active centres. Since these residues form the first part of the active centre, we can say that, in all RIPs, the geometry of that part of the active centre is highly conserved. Residue 109 is Gly in all



**Figure 5** A schematic illustration of the proposed de-adenination mechanism of RIPs

A complete description is given in the text. **(a)** The N-7 atom of the adenine is protonated by OE1 of Glu-85; at this time the adenine is positioned at the second part of the active centre of RIPs. **(b)** The side chain of Tyr-70 rotates, then opens the path between the first and second parts of the active centre. Then the adenine becomes trapped in the hydrophobic pocket formed by the first part. The bond is breaking between ribose and adenine. The positive charge on the adenine is decreased by the N-9–C-1' bond electron-pair transfer to N-9, and the cation character increases on ribose and is stabilized by OE1 of Glu-160. **(c)** Bond formation between C-1' and an attacking water molecule which is activated by the OE2 of Glu-160. Then one product, a free adenine, moves to the second part of the active centre, and the side chain of Tyr-70 rotates back; the path closes off. The main role of the side chain of Tyr-70 is to act like a valve between the two parts of the active centre.

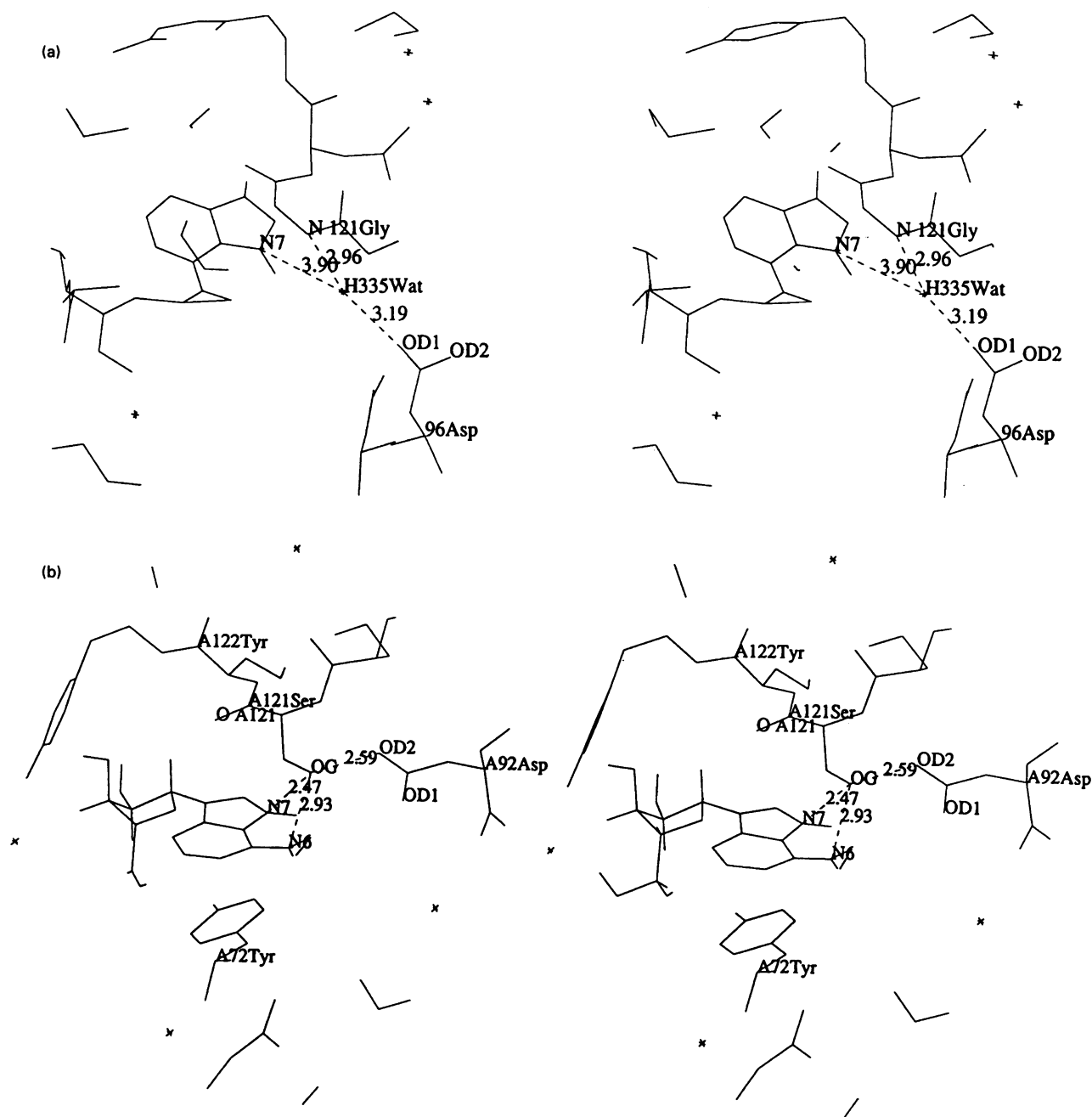
RIPs, except saporin 6, in which it is a Tyr. This means in most RIPs the base can also stack with the peptide plane of residue 109. Most RIPs have a hydrophobic residue at 83 and an acidic amino acid residue at 85. The former residue may make a contribution to hydrophobic forces and the latter residue may provide a proton for protonating adenine N-7. In all RIPs except ricin A-chain, residue 108 is a hydrophilic amino acid. In ricin A-chain residue 108 is a Gly. It seems that in RIPs the geometries of the second part of the active centres are also conserved.

There are also some conserved residues in RIPs which are not directly involved in forming the active centres; they are Phe-4, Tyr-14, Phe-17, Arg-22, Leu-52 and Arg-122. We checked them and found that all of them form some kind of interactions with residues of, or near, the active centres. These residues can be thought of as being involved in the stabilization of the geometry of active centres. For example, the aromatic side chains of Phe-4, Tyr-14 and Phe-17 are buried in TCS's or MMC's core and form a hydrophobic cluster, around which are Ile-18 and Ala-161. The OH group of Tyr-14 also forms a hydrogen bond with O of residue 157. Between NH<sub>2</sub> of Arg-22 and 168OE2, NE of 22 and 162 O there are two pairs of hydrogen bonds. Leu-52, Ile-60, residue 50, and Leu-154 form a hydrophobic core which may fix the position of Ile-155. As mentioned above, residues Phe-4, Tyr-14, Phe-17 and Arg-22 are located at  $\beta$ -sheet  $\beta_1$  and  $\alpha$ -helix  $\alpha_1$  respectively. Residues 50, 52, and 60 are located at sheets  $\beta_{1,2}$  and  $\beta_{1,3}$  respectively. Residues 154, 157, 161 and 162 are located at helix  $\alpha_{1,5}$ , at which residues 155, 160 and 163 are also located. So those interactions mentioned above may be involved in fixing the relative positions of helix  $\alpha_{1,5}$ . The guanidiny plane of Arg-122 lies parallel with the phenoxy plane of Tyr-111 and its NH1 and NH2 atoms form ion-pairs with 189OE1 and 115OE1. These interactions may fix the positions of residues 111 and 189.

Generally speaking, most conserved residues in the RIP family are around the active centre described by us. The geometries of the active centre and its binding mode with ligands, especially of its first part, are also conserved in all RIPs. In other words, the geometries of the active centre and its ligand binding mode are described above and the depurinating mechanism proposed below may represent all RIPs.

#### A proposed depurination mechanism for RIPs

Although no direct experiments have been done on the transition state of the RNA N-glycosidase mechanism of RIPs, detailed experiments on acid-catalysed solvolysis and AMP glycosidase catalytic dynamic characteristics have been done [46–48]. In these experiments, it was observed that the N-7 atom of the adenine base of the reactive residue was protonated. In TCA and MCA the hydrolytic products' adenines are all N-7-protonated. In MCF the N-7 atom of formycin forms a hydrogen bond with the carbonyl O atom of residue 109. If the N-7 atom of the adenine were not protonated, the base could not approach its end state, because of steric hindrance between N-7 and the carbonyl O atom of residue 109 without the ability to form a hydrogen bond, and it must not position at the binding pocket formed by the first part of the active centre of RIPs. However, it is possible for the substrate to bind with RIPs with the base positioning at the second part of the active centre. We have also compared structures of TCS and MMC with structures of ricin A-chain (2.5 Å, Protein Data Bank entry 1RTC [18]) and pokeweed anti-viral protein (Protein Data Bank entry 1PAF) (J. D. Robertus, unpublished work). First we superimposed CA atoms of several residues of ricin A-chain or pokeweed anti-viral protein, which are constant in all RIPs, on TCSs and thus got the



**Figure 6** Equivalent residues in ricin A-chain and pokeweed anti-viral protein that can transfer a proton to N-7 of reactive adenine

Thin lines represent the formycin model in TCF. Broken lines represent hydrogen bonds. (a) Ricin A-chain after superimposition on TCS. (b) Pokeweed anti-viral protein after superimposition on TCS.

initial fitting matrix. Then we picked out three-dimensionally equivalent residues, i.e. those with CA atoms are closer than 1.5 Å after superimposition. Between CA, C and N atoms of 197 such residues of ricin A-chain and TCS the root-mean-square deviation is 0.73 Å. Between TCS and pokeweed anti-viral protein (molecule A), the corresponding deviation of 156 equivalent residues is 0.8 Å. On the basis of these observations we propose the following depurination catalytic mechanism for RIPs (Figure 5).

The first step of the depurination mechanism is the approach of substrate to the cleft of RIPs. There may be many interactions

between protein atoms in the cleft and atoms of the substrate, since there are many charged amino acid residues in the cleft. The cleft was observed to be large enough to contain a single-stranded RNA-chain. To avoid the steric hindrance between N-7 and 109 O, the adenine base may position itself at the second part of the active centre, when N-7 is not protonated. Its atoms may form those interactions in TCF described above with atoms in that part. The ribosomal proteins may also be involved at this step; they may contribute to the binding energy (see the Discussion). Substrate is enriched around the active centre of RIPs after this step.

The second step is the protonation of the N-7 atom of the reactive adenine base by an acidic residue of RIPs (Figure 5). As described above, in TCS and MMC the side chain of Glu-85 is positioned near the N-7 atom when the ligand binds at that position (also see Table 3). In ricin A-chain, the Asp-96 OD1 through 335Wat O, and in pokeweed anti-viral protein, the Asp-92 OD2 through Ser-121 OG, are equivalent acidic residues in their active centres which can protonate the N-7 atom (Figure 6). Then the side chain of Tyr-70 rotates about 57° around its CA–CB single bond and the path between the first and second part of the active centre opens. Thus the adenine is trapped in the hydrophobic pocket formed by the first part of the active centre. This time the base position is near formycin in MCF; this movement can be thought of as a rotation around hydrogen bond 71 N–H...N-1. Then the ribose may rotate around O-2' and this rotation brings O-4' and C-1' atoms near OE1 of Glu-160. The bond is breaking between ribose and adenine. The cation character of adenine is decreased by the N-9–C-1' bond electron pair transfer to N-9, and the cation character increases on ribose; this positive charge accumulation establishes the oxycarbonium character of C-1' and O-4'. That is, the bond order of C-1'–O-4' is somehow increased to greater than 1.0 and the positive charge of C-1' distributed among C-1' and O-4'. This positively charged oxycarbonium ion is stabilized by the negative charge of OE1 of Glu-160.

The final step is the occurrence of the reaction and the leaving of products. The configuration of C-1' flips and the N-9–C-1' bond breaks; at the same time a new bond is formed between C-1' and an attacking water molecule (named 86Wat and 83Wat respectively in MCA and TCA; Figure 4a). This water molecule is found just below C-1', opposite to the N-9 atom of adenine in TCS or MMC and near the OE2 atom of Glu-160. It may be activated by the OE2 of Glu-160, and approaches towards the positively charged C-1' atom. Then a new bond is formed between the C-1' atom and the hydroxy group of the water. After reaction, the van der Waals repulsive force between protein atoms and configuration-flipped ribose atoms pushes the modified ribosome away from the active centre. Another product adenine, a free adenine, moves to the second part of the active centre and remains in the binding pocket. The side chain of Tyr-70 rotates back and the path closes off. As pointed out above, the adenine binds very tightly with the protein; we think it is very difficult and slow for adenine to diffuse into the solution environment. We suppose it is the next ribosome approaching the RIP's active centre that causes some changes to the group of active-centre atoms and thus releases the adenine. The simplest and most reasonable change may occur at Tyr-70: ribosome may fix the side chain of Tyr-70 at the position observed in native RIPs by some unknown interactions, for example the interaction of aromatic stacking of the side chain of Tyr-70 with some bases of ribosome rRNA or the hydrogen bond between its OH group and a charged phosphate group. Then the binding pocket disappears, and the adenine may shift to the second part of the active centre and be pushed away through the other opening (Figure 5). Generally speaking, it is a typical acid catalytic mechanism. A charge-transfer triad can describe this mechanism: a proton from OE1 of Glu-85, through the adenine base of the ribosome, transfers to the OE2 of Glu-160, during the depurination process.

## DISCUSSION

The first thing we noted in our structure determination was that ATP was cleaved and only the hydrolytic product adenine could exist in the active centre of TCS and MMC. Since the solvolysis

of ATP happened very slowly at the pH of our crystallization, we think it may be cleaved by TCS and MMC. But biochemical experiments show that no hydrolytic activity of TCS on adenosine was observed (Q. Huang, S. Liu, Y. Tang, S. Jin and Y. Wang, unpublished work). There are two reasons why the mononucleotide is not the substrate of RIPs. First, the binding of mononucleotides with RIPs is relatively weak, at least with TCS, as shown by structure analysis of TCF. The small ligand cannot bind strongly in the active centre in reaction state. In fact, as the  $pK_a$  value of N-7 is  $< 2$ , at neutral pH there is very little N-7 protonated nucleotide in solution. Secondly, the reaction may be inhibited by adenine occupying the active centre. From our analysis above we can deduce an order of binding affinity between different ligands: adenine  $>$  formycin  $>$  adenosine phosphate. Since intact eukaryotic ribosome can be cleft by RIPs very quickly ( $V_{max}$  about  $1500 \text{ min}^{-1}$  and  $K_m > 1 \text{ mM}$  [15]), we can say that the binding power of intact ribosome with RIPs is stronger than that of adenine. We should point out that we have tried data collection on crystals obtained by co-crystallization from MMC with 5'-AMP or 3'-ATP. There is no difference from crystals of 5'-ATP in the electron density. We have also co-crystallized MMC with 2'-deoxyadenosine 5'-monophosphate, 5'-ADP, 3'-AMP, 3'-ADP and 3'-ATP and collected part of its X-ray data. We cannot find the ribose moiety in the active centre there either.

In our proposed mechanism, we think one of the product adenines cannot diffuse into the environment easily or quickly. In fact, we have soaked the MCF crystal in its mother liquid containing no ligands for about 10 days, then collected its X-ray data, but formycin still stays stably in the MMC active site. In our proposed mechanism the glycoside bond seems to be broken before a water molecule attacks C-1'. It implies that, in this catalytic reaction, an early  $S_N1$  transition state is involved. This implication is consistent with the observation of the hydrolysis reaction catalysed by AMP nucleosidase [48]. We have noted the differences of ligand binding between TCS and MMC: the binding of formycin and very possibly adenine is weaker in TCS than in MMC. It is consistent with the fact that MMC's cell-free inhibitory effect on protein synthesis is about 60 times as high as TCS's [3].

On the basis of the geometry of the active centre and the shape of the cleft, we tried to find some reasons for the secondary structure specificity of RIPs. As reported by Endo and Tsurugi [15], the cleaved N-glycoside bond of rRNA is located within a loop with a stem. We checked the cleft which was thought to bind with nearby residues of rRNA and found only a single-stranded helix of RNA could fit into it. From the described geometry of the active centre we found the reactive residue must be a loop-out residue. This excludes double-helix nucleotides or single-helix nucleotides as substrates.

Apparently our mechanism is different from that represented by Monzingo and Robertus [23]. In Monzingo and Robertus' mechanism, the N-3 atom of the reactive base is supposed to be protonated by the side chain of Arg-163 (R-180 in ricin A-chain). This mechanism raised doubts when considering several experimental facts. First, the experiments on acid solvolysis of adenine nucleosides showed that modifying the pyrimidine part of the base has minor effects on the solvolysis reactivity, i.e. the protonation of the N-3 atom cannot contribute very much to the reaction [46]. Secondly, arginine is a strong base, and the positive side chain of Arg-163 forms a strong ion-pair with the side chain of Glu-160. It is difficult to imagine that Arg-163 can transfer a proton to N-3. On the other hand, the N-7 protonation mechanism is consistent with experimental findings on the acid solvolysis of nucleosides. In TCS and MMC also we find an

acidic residue in the active centre, Glu-85, which can protonate the N-7 atom. Similarly, in ricin A-chain and in pokeweed antiviral protein, we find that there are equivalent acidic residues in their active centres, which can protonate the N-7 atom (Figure 6).

We thank Professor Y. Katsube, Professor Y. Matsuura, Dr. Y. Hata and Dr. M. Sato in the Institute for Protein Research, Osaka University, Japan, for their kindly helping in the early stages of this work. We also thank Professor N. Sakabe and Dr. T. Nakagawa at the Photon Factory in Tsukuba, Japan, for their help in our X-ray data collecting. This work was supported by the Chinese Natural Science Foundation for 'Important Chemical Problems in Life Processes, III-3-2'. Some early-stage studies of this subject were carried out in the Institute for Protein Research, Osaka University, Osaka, Japan. All final refined co-ordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory. [Entry codes: MRG (MCA), MRH (MCF), MRI (MCN), MRJ (TCA) and MRK (TCF).

## REFERENCES

- Olsnes, S. and Phil, A. (1982) in *Molecular Action of Toxins and Viruses* (Cohen, P. and Van Heyningen, S., eds.), pp. 51–105, Elsevier Biomedical Press, Amsterdam
- Roberts, W. K. and Selltrechnikoff, C. P. (1986) *Biosci. Rep.* **6**, 19–29
- Stirpe, F. and Barbieri, L. (1986) *FEBS Lett.* **295**, 1–8
- Stirpe, F., Barbieri, L., Battelli, M. G., Soria, M. and Lappi, D. A. (1991) *Bio/Technology* **10**, 405–412
- Zhang, X. J. and Wang, J. H. (1986) *Nature (London)* **321**, 477–478
- Jing, S., Gu, Z. and Wang, Y. (1992) *Kexue Tongbao (China)* **38**, 671
- Wang, Y., Qian, R.-Q., Gu, Z.-W. et al. (1986) *Pure Appl. Chem.* **58**, 789–798
- Yeung, H. W., Li, W. W., Feng, Z., Barbieri, L. and Stirpe, F. (1988) *Int. J. Pept. Protein Res.* **31**, 265–268
- Vitetta, E. S., Fulton, R. J., May, R. D., Till, M. and Uhr, J. W. (1987) *Science* **238**, 1098–1104
- McGrath, M. S., Hwang, K. M., Caldwell, S. E. et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2844–2848
- Lee-Huang, S., Huang, P. L., Nara, P. L. et al. (1990) *FEBS Lett.* **272**, 12–18
- Lee-Huang, S., Huang, P. L., Kung, H.-F. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6570–6574
- Endo, Y. and Tsurugi, K. (1987) *J. Biol. Chem.* **262**, 8128–8130
- Endo, Y., Mitsui, K., Motizuki, M. and Tsurugi, K. (1987) *J. Biol. Chem.* **262**, 5908–5912
- Endo, Y. and Tsurugi, K. (1988) *J. Biol. Chem.* **263**, 8735–8739
- Rutenber, E., Katzin, B. J., Collins, E. J. et al. (1991) *Proteins* **10**, 240–250
- Rutenber, E. and Robertus, J. D. (1991) *Proteins* **10**, 260–269
- Katzin, B. J., Collins, E. J. and Robertus, J. D. (1991) *Proteins* **10**, 251–259
- Pan, K., Zhang, Y., Lin, Y. et al. (1982) *Sci. Sin. (B)* **6**, 522–526
- Pan, K., Zhang, Y., Lin, Y. et al. (1985) *Sci. Sin. (B)* **8**, 718–723
- Ma, X., Wang, Y. and Wang, J. (1987) *Sci. Sin. (B)* **30**, 692
- Zhang, F., Li, W. W., Yeung, H. W. (1990) *J. Mol. Biol.* **214**, 625–626
- Monzinger, A. F. and Robertus, J. D. (1992) *J. Mol. Biol.* **227**, 1136–1145
- Sato, M. (1992) *Rigaku Res.* **23**, 13–26
- Sakabe, N. (1991) *Nucl. Instrum. Methods* **A303**, 448–463
- Steigerman, W. (1974) Ph.D. Thesis, University of Munich
- Jones, T. A. (1978) *J. Appl. Crystallogr.* **11**, 262–272
- Brunger, A. (1992) *XPLOR Version 3.0 Manual*, Yale University, New Haven
- Engh, R. A. and Huber, R. (1991) *Acta Crystallogr.* **A47**, 492–500
- Prusiner, P., Brennan, T. and Sundaralingam, M. (1973) *Biochemistry* **12**, 1196–1201
- Blow, D. M. (1985) in *Molecular Replacement* (Compiled by Machin, P. A.), Daresbury Laboratory, Warrington
- Tickle, I. J. (1985) in *Molecular Replacement* (Compiled by Machin, P. A.), Daresbury Laboratory, Warrington
- Luzatti, P. V. (1952) *Acta Crystallogr.* **5**, 802–810
- Chow, T. P., Feldman, R. A., Lovett, M. and Piatak, M. (1990) *J. Biol. Chem.* **265**, 8670–8674
- Ho, W. K. K., Liu, S. C., Shaw, P. C., Yeung, H. W., Ng, T. B. and Chan, W. Y. (1991) *Biochim. Biophys. Acta* **1088**, 311–314
- Minami, Y. (1993) *Biosci. Biotechnol. Biochem.* **57**, 1141–1144
- Saenger, W. (1984) in *Principles of Nucleic Acid Structure* (Cantor, C. R., ed.), pp. 9–28, Springer-Verlag, New York
- Halling, K. C., Halling, A. C., Murry, E. E., Ladin, B. F., Houston, L. L. and Waever, R. F. (1985) *Nucleic Acids Res.* **13**, 8019–8033
- Maras, B., Ippoliti, R., DeLuca, E., Lendaro, E., Bellelli, A., Barra, D., Bossa, F. and Brunori, M. (1990) *Biochem. Int.* **21**, 631–638
- Kung, S.-S., Kimura, M. and Funatsu, G. (1990) *Agric. Biol. Chem.* **54**, 3301–3318
- Islam, M. R., Nishida, H. and Funatsu, G. (1990) *Agric. Biol. Chem.* **54**, 1343–1345
- Funatsu, G., Taguchi, Y., Kamenosono, M. and Yanaka, M. (1988) *Agric. Biol. Chem.* **52**, 1095–1097
- Asano, K., Svensson, B., Svendsen, I., Poulson, F. M. and Roepstorff, P. (1986) *Carlsberg Res. Commun.* **51**, 129–141
- Ready, M. P., Kim, Y. and Robertus, J. D. (1991) *Proteins* **10**, 270–278
- Habukam, N., Miyano, M., Kataoka, J., Tsuge, H. and Noma, M. (1992) *J. Biol. Chem.* **267**, 7758–7760
- Garrett, E. R. and Mehta, P. J. (1972) *J. Am. Chem. Soc.* **94**, 8532–8541
- DeWolf, W. E., Jr., Fullin, A. F. and Schramm, V. L. (1979) *J. Biol. Chem.* **254**, 10868–10875
- Mentch, F., Parkin, D. W. and Schramm, V. L. (1987) *Biochemistry* **26**, 921–930