J. Mol. Biol. (1993) **233**, 705–715

The 2·5 Å Structure of Pokeweed Antiviral Protein

**Arthur F. Monzingo[1], Edward J. Collins[1]†, Stephen R. Ernst[1]
James D. Irvin[2] and Jon D. Robertus[1]‡**

[1]*Department of Chemistry and Biochemistry
University of Texas, Austin, TX 78712, U.S.A.*

[2]*Department of Chemistry
Southwest Texas State University, San Marcos, TX 78666, U.S.A.*

(Received 19 October 1992; accepted 22 January 1993)

The pokeweed antiviral protein (PAP), isolated from the leaves of *Phytolacca americana*, is one of a family of plant and bacterial ribosome-inhibiting proteins (RIPs) which act as specific *N*-glycosidases on rRNA. Here we report the three-dimensional structure of PAP determined to 2·5 Å resolution by X-ray crystallography. After 14 rounds of refinement, the *R* factor is 0·17 for 5·0 to 2·5 Å data. The protein is homologous with the A chain of ricin and exhibits a very similar folding pattern. The positions of key active site residues are also similar. We also report the 2·8 Å structure of PAP complexed with a substrate analog, formycin 5′-monophosphate. As seen previously in ricin, the formycin ring is stacked between invariant tyrosines 72 and 123. Arg179 bonds to N-3 which is thought to be important in catalysis.

Keywords: PAP; X-ray structure; ricin homolog; ribosome inhibiting protein

1. Introduction

Many plants produce enzymes which inhibit ribosomes (RIPs§). If these proteins are delivered to the cytoplasm they can kill a cell at very low doses. *Phytolacca americana*, the pokeweed plant, produces three isozymes which are members of this enzyme class. Pokeweed antiviral protein, PAP, can be isolated in good yield from spring leaves of the plant while a variant, PAP-2, is expressed in low levels in spring leaves and in high levels in summer leaves. A third isozyme found in the seeds is called PAP-S. The proteins are called antiviral proteins because they were discovered through the observation that soluble extracts of pokeweed could retard replication of tobacco mosaic virus (Duggar & Armstrong, 1925; Kassanis & Kleczkowski, 1948). The pokeweed proteins are all representatives of the class-1 family of RIPs which are single chain proteins of about 30,000 molecular weight.

In addition, there is a family of heterodimers called class-2 RIPs. These have an A chain which is homologous to the class-1 proteins (Ready *et al.*, 1984) and one or more B chains which normally act to bind the surface of target cells. These heterodimeric proteins are about 10,000 times as toxic as class-1 proteins by virtue of their facilitated uptake and are true cytotoxins. The best studied example of this class is ricin.

It is now known that RIPs act as specific *N*-glycosidases, removing an invariant adenine base from a loop region of 28 S rRNA (Endo & Tsurugi, 1988; Endo *et al.*, 1988). The adenine base is apparently involved in binding translocation factors, since RIP intoxicated ribosomes show a diminished binding of EF-1 and EF-2. Initial rate kinetics for PAP activity against ribosomes from *Artemia salina* have been described (Ready *et al.*, 1983). The K_m for ribosomes is 0·2 to 1 μM, depending on salt concentrations, and k_{cat} is 350 min^{-1}. Electron microscopy showed that PAP is localized in the plant cell wall (Ready *et al.*, 1986). This led to the conclusion that PAP is a defensive protein; it enters the cell when the wall is breached, inhibiting cellular ribosomes and retarding viral replication within the compromised cell. The demonstration that PAP, indeed, inhibits pokeweed ribosomes lends support to this hypothesis (Bonness, 1992). The amino acid sequence of PAP has recently been deduced from the sequence of a cDNA clone (Lin *et al.*, 1991). PAP

† Present address: Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, MA 02138, U.S.A.

‡ Author to whom all correspondence should be addressed.

§ Abbreviations used: RIP, ribosome-inhibiting protein; PAP, pokeweed antiviral protein; MIR, multiple isomorphous replacement; RTA, ricin toxin A chain; FMP, formycin-5′-monophosphate; NED, no electron density.

0022-2836/93/200705-11 $08.00/0

is frequently used in the construction of tissue-specific therapeutic agents called immunotoxins. Recently HIV-infected T cells have been treated with PAP conjugated to anti-CD4$^+$ antibodies (Zarling *et al.*, 1990).

PAP was crystallized for X-ray studies some years ago (Robertus *et al.*, 1977), but the triclinic crystal form was difficult to work with. Considerable effort has since been spent on the heterodimeric RIP, ricin. The X-ray structure of ricin has been refined to 2·5 Å resolution (Rutenber *et al.*, 1991) and the details of both the A chain (Katzin *et al.*, 1991) and the B chain (Rutenber & Robertus, 1991) have been described. Site-directed mutagenesis (Ready *et al.*, 1991; Kim *et al.*, 1992) and X-ray analysis of substrate analogs have allowed a mechanism of action for all RIP enzymes to be proposed (Monzingo & Robertus, 1992). In this paper we describe the solution of the crystal structure of PAP by a combination of MIR and molecular replacement methods and compare it to the ricin toxin A chain (RTA).

2. Materials and Methods

Potassium tetracyanoplatinate ($K_2Pt(CN)_4 \cdot 3H_2O$) and potassium tetranitritoplatinite ($K_2Pt(NO_2)_4$) were purchased from Apache Chemicals, Inc. Formycin 5′-monophosphate (FMP) was purchased from Sigma Chemical Company.

PAP was prepared and crystallized as described previously (Irvin, 1975; Robertus *et al.*, 1977). The crystals are triclinic, space group $P1$, with $a=49{\cdot}5$ Å, $b=50{\cdot}1$ Å, $c=65{\cdot}2$ Å, $\alpha=80{\cdot}0°$, $\beta=113{\cdot}2°$, $\gamma=116{\cdot}5°$. There are two 30 kDa monomers per unit cell; and rotation function studies have shown that they are approximately related by a 2-fold nearly coincident with the crystallographic a axis.

An artificial mother liquor consisting of 30% (w/v) PEG 8000, 0·05 M Tris (pH 7·4), was used in derivative preparation. The $Pt(CN)_4^{2-}$ derivative was prepared by soaking crystals in artificial mother liquor containing 60 mM $K_2Pt(CN)_4$ for 10 days. The $Pt(NO_2)_4^{2-}$ derivative was prepared by soaking crystals in artificial mother liquor containing 8 mM $K_2Pt(NO_2)_4$ for 6 weeks. The PAP crystal with FMP bound was prepared by soaking in artificial mother liquor containing 5 mM FMP for 6 weeks. Data sets for the native crystal and 3 derivatives were collected to a nominal resolution of 1·9 Å using a San Diego Multiwire Systems area detector (Hamlin, 1985; Howard *et al.*, 1985). The X-ray source was a graphite-monochromatized GX-20 rotating anode generator operated at 40 kV, 40 mA.

Heavy atom sites were located by difference Patterson and Fourier analysis. Heavy atom parameters were refined and phases were calculated using a suite of programs assembled by Dr Gregory Petsko.

The MIR phases were improved by solvent flattening using the method of Wang (1985). Rigid body refinement was done using CORELS (Sussman, 1985). A symmetry-averaged map was generated using the method of Bricogne (1976). Model building was done using the program FRODO (Jones, 1982) on an Evans and Sutherland PS390 graphics system.

Crystallographic refinement of PAP used the X-PLOR package (Brunger, 1988). The strategy of refinement was similar to that reported for the ricin heterodimer (Rutenber *et al.*, 1991). We defined a round of refinement as a rebuilding, by hand, of the model followed by a variable number of cycles of automated refinement using the molecular dynamics option of X-PLOR on a CRAY Y-MP8/864. The automated refinement used terms from 5 Å to the nominal resolution limit of the run, typically 2·8 or 2·5 Å.

At the end of each round of refinement, a Fourier map was computed using structure factor amplitudes and phases calculated from the newly refined model. As described previously (Rutenber *et al.*, 1991), such a map is potentially biased by the calculated phases and so a "no electron density" (NED) map was calculated using an in-house program, NEDFFT. A NED map is similar to an omit map but is easier to systematically generate for the entire model. A slab portion of the starting density is omitted and the remaining map is back-transformed to obtain new calculated structure factors ($F_{calc'}$) and phases. These phases are then combined with amplitudes of the form $(2F_{nat,obs}-F_{calc'})$, to produce an unbiased slab of density, corresponding to that which was omitted. The process is repeated until the entire NED map is constructed. These maps were computed using terms from 10 Å to the nominal resolution of the map. Difference Fourier maps with amplitudes of the form $(F_{nat,obs}-F_{calc})$ and phases for the current model were also calculated to aid in rebuilding.

For the first 4 rounds of refinement, the 2 monomers were constrained to the symmetry relationship that had been obtained from the CORELS rigid body refinement. For rounds 5 to 13, weighted restaints were used to impose non-crystallographic symmetry between the monomers with the backbone being more strongly restrained than side-chains, or, in the later rounds, bound water molecules. Through these later rounds, the 2 monomers remained virtually identical about the non-crystallographic symmetry axis despite being restrained rather than constrained. For rounds 5 to 12, rebuilding was done in one monomer and transformed into the other monomer. The transformed model was inspected to make sure it conformed to the NED map. Generally, with each round, both monomers were inspected in the NED map but no significant differences were observed. Finally, non-crystallographic symmetry was not restrained in round 14.

Beginning with round 12, bound water molecules were added to the model. Waters were added which could form at least one hydrogen bond and had corresponding peaks in both the difference Fourier map and NED map. Initially, added waters were required to be binding to both protein molecules of the asymmetric unit; that is, they had to obey the non-crystallographic symmetry operator. Later, waters were added on the surface of the protein molecule which did not obey the non-crystallographic symmetry operator.

Using data collected from a PAP crystal soaked with FMP, difference electron density maps were computed with amplitudes of the form $(F_{complex}-F_{nat,obs})$ and phases from the native refinement. Omit maps were computed with amplitudes of the form $(2F_{complex}-F_{nat,calc})$, where both calculated structure factors and phases are from the refined model, but with certain active site residues deleted, to diminish biasing in the ligand fitting. A model with FMP bound to each of the monomers of the asymmetric unit was built and refined by energy minimization using X-PLOR.

Rotation searches were performed using the fast rotation function program of Tanaka (1977) and the

Table 1
Data collection statistics

	Native	$Pt(CN)_4^{2-}$	$Pt(NO_2)_4^{2-}$	FMP
No. of observations	77,556	128,906	123,430	72,900
No. of reflections	19,127	14,074	13,874	13,177
Resolution limit (Å)	2·5	2·8	2·8	2·8
Completedness (%)	97	99	98	94
R_{merge}†	0·056	0·128	0·114	0·106
Mean isomorphous difference (%)		20·1	17·0	21·7

† $R_{merge} = \Sigma |I - \bar{I}| / \Sigma \bar{I}$ and gives the overall agreement between reflections measured more than once.

MERLOT package provided by Dr Paula Fitzgerald (Fitzgerald, 1988).

The refined coordinates of native PAP and of PAP bound with FMP have been deposited in the Brookhaven Protein Data Bank have been assigned the accession numbers 1PAF and 1PAG, respectively.

3. Results and Discussion

(a) *Data collection*

The area detector was used to collect data to a nominal resolution of 1·9 Å resolution. Derivative data were used to calculate MIR phases to 2·8 Å resolution. Native data to 2·5 Å resolution were used to refine the protein structure. 2·8 Å data of the PAP·FMP complex were used to refine that structure. Data collection statistics are summarized in Table 1.

(b) *MIR phasing and maps*

With no Harker plane in space group $P1$, solution of the derivative heavy atom positions proceeded with difficulty. The difference Patterson of the $Pt(CN)_4^{2-}$ derivative gave an interpretation that was consistent with one determined several years ago using data collected from a diffractometer (Monzingo, 1979). Unfortunately, no two pairs of the sites obeyed the same 2-fold operator. The $Pt(NO_2)_4^{2-}$ derivative was also initially screened from diffractometer data and, again, the two sites did not obey any 2-fold operator that could be defined from the $Pt(CN)_4^{2-}$ sites. This was disconcerting but for both compounds difference Fouriers using phases from one derivative did return the sites of the other derivative; this indicated that our solution was probably correct.

The final refined heavy atom parameters are summarized in Table 2. The overall figure of merit of MIR phases calculated to 2·8 Å resolution was 0·63. The MIR phasing statistics are shown in Table 3.

The MIR phases were improved by the method of

Table 2
Heavy atom parameters

Derivative	Site no.	*X*	*Y*	*Z*	Relative occupancy	Residues bound to†
$Pt(CN)_4^{2-}$	1	0·083	0·117	0·188	0·703	Lys A246 Lys B59‡
	2	−0·912	−0·883	−0·823	0·455	Lys B33 Lys B115‡
	3	0.323	0.334	−0·103	0.417	Lys A15
	4	1·033	0·502	1·169	0·396	Lys B15
$Pt(NO_2)_4^{2-}$	1	1·102	1·068	0·198	0·430	Met A156
	2	0·365	0·388	0·523	0·606	Lys A210 Lys B210

† A and B indicate respective monomers of the PAP dimer.
‡ Indicates a residue of a symmetry-related molecule.

Table 3
Multiple isomorphous replacement statistics

d_{min} (Å)	8·85	6·25	5·10	4·42	3·95	3·61	3·34	3·13	2·95	2·80
$Pt(CN)_4^{2-}$										
No. of reflections	354	725	959	1112	1276	1403	1527	1616	1730	1765
f_H	73	68	63	59	54	50	47	43	40	37
E	28	24	25	30	34	35	31	28	27	27
$Pt(NO_2)_4^{2-}$										
No. of reflections	354	713	935	1090	1249	1372	1507	1596	1711	1753
f_H	53	50	46	43	39	37	34	32	29	27
E	47	30	29	34	37	36	33	29	26	26
Mean figure of merit	0·73	0·66	0·65	0·67	0·68	0·66	0·63	0·60	0·57	0·57

f_H is the r.m.s. heavy atom structure factor. *E* is lack of closure.

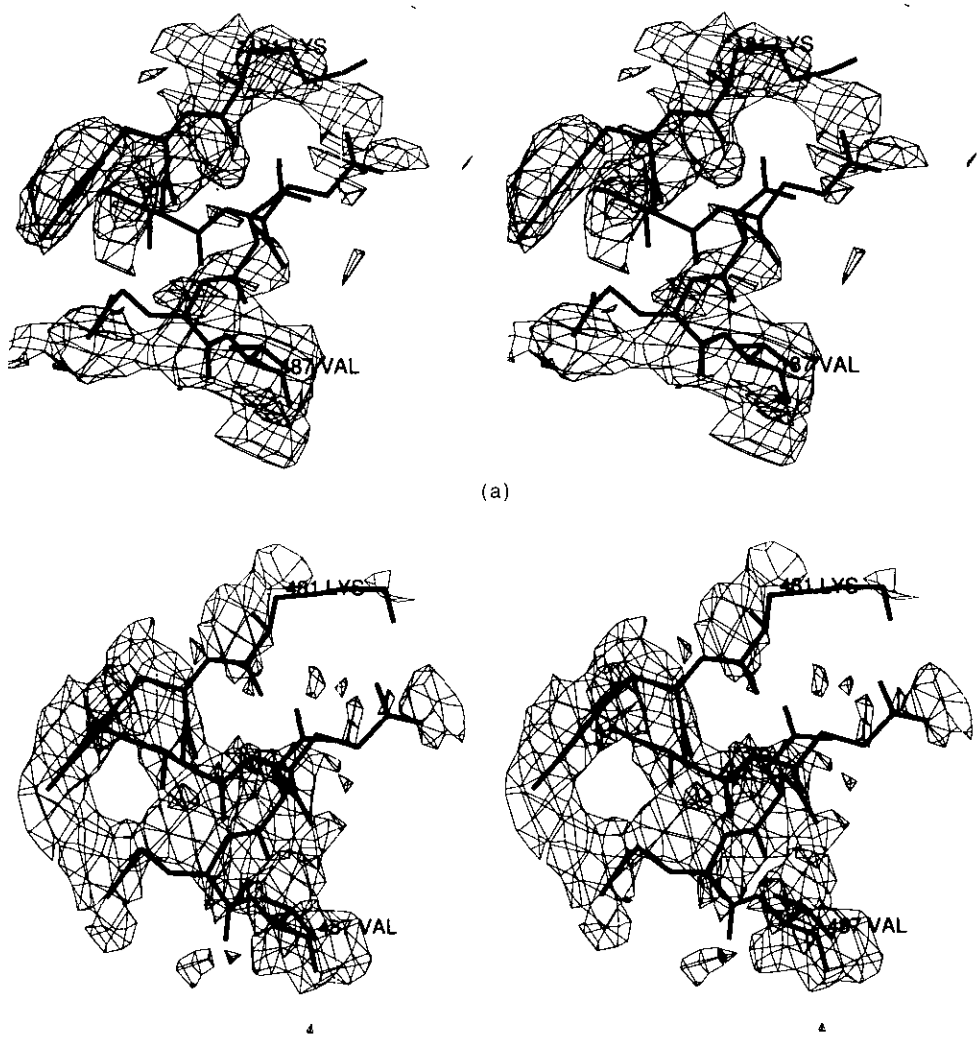


Fig. 1.

Wang (1985). Inspection of the resulting electron density map allowed density to be assigned to two protein molecules. The strongest features of the map appeared to be two helices, one in each monomer, which could be related by a 2-fold operation about the crystallographic *a* axis. Two of the $Pt(CN)_4^{2-}$ heavy atom sites, numbers 3 and 4, obeyed this same symmetry operator. In addition, several other apparent helices could be defined in each of the monomers. The relative orientation of three of these helices enabled the placement of two ricin A chain (RTA) monomers. The helices defined in the map overlapped with the A, E and G helices of RTA (Katzin *et al.*, 1991), and the boundaries of the map matched very well with the boundaries of the RTA model monomers. Figure 1(a) and (b), shows regions of the MIR map corresponding to the two molecules; segments of the final two PAP models are superimposed. The two RTA test molecules were initially related by a 2-fold rotation axis coincident with the crystallographic *a* axis. In an effort to refine the symmetry operator, the two RTA monomers were refined as rigid bodies against the PAP diffraction data using CORELS (Sussman, 1985). This refinement improved the fit of the monomers to the "solvent-flattened" MIR map. Using the symmetry operator resulting from the rigid body refinement, a cyclically improved symmetry-averaged map was generated using the method of Bricogne (1976). Figure 1(c) shows a section of this averaged map corresponding to the regions shown in Figure 1(a) and (b). From this map, most of the chain could be traced, and the monomer was rebuilt to fit the density using what was then known of the PAP amino acid sequence (110 residues) along with the PAP-S sequence (Kung *et al.*, 1990).

(c) *Model refinement*

Crystallographic refinement was undertaken in rounds. Each round included hand rebuilding of the model using difference Fourier and NED maps followed by automated simulated annealing as described in Materials and Methods. The progress of the refinement is summarized in Table 4 and the *R* factor is plotted as a function of round of crystallographic refinement in Figure 2.

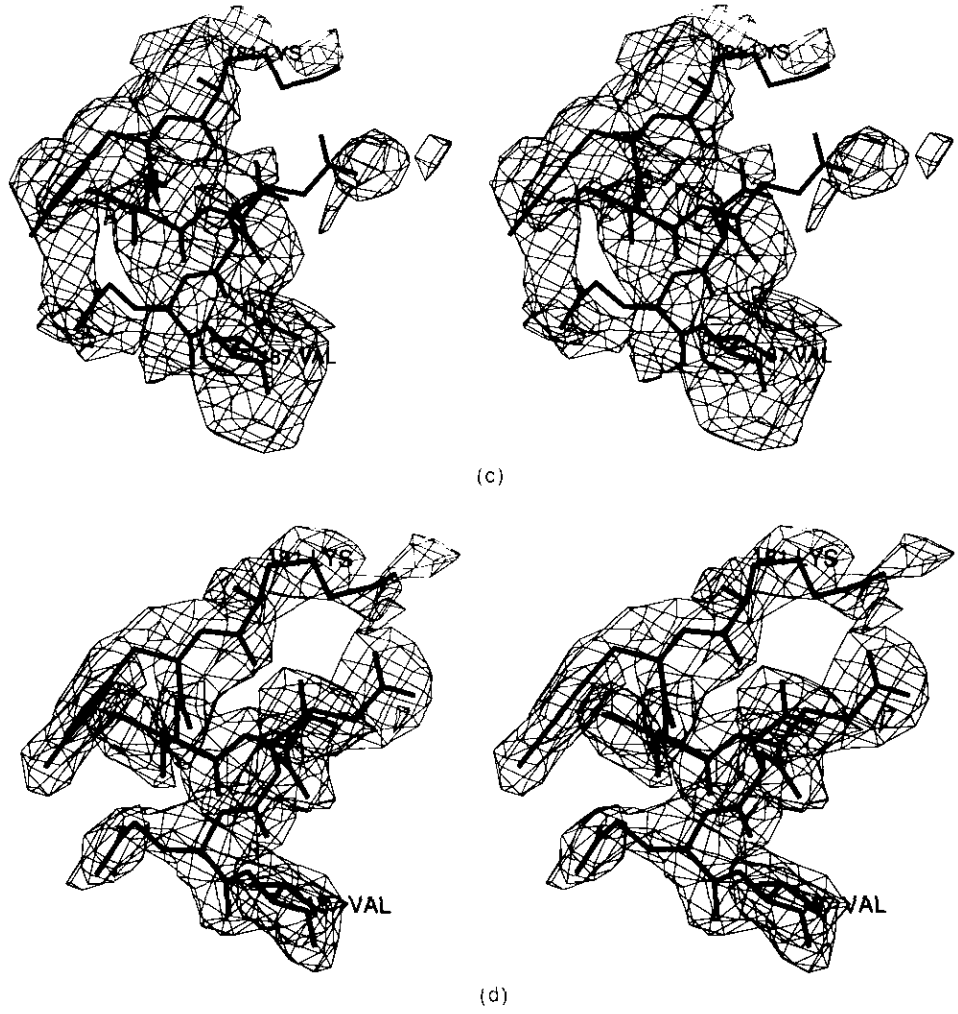


Figure 1. Electron density maps of PAP during refinement. The MIR maps were made using 10·0 to 2·8 Å resolution data; the NED map is from 10·0 to 2·5 Å data. For each map, the contour encloses 15% of the cell volume. (a) The original MIR map for a region of the protein in molecule A. (b) The corresponding region of the MIR map for molecule B. (c) The same region of the protein in the symmetry-averaged MIR map. (d) The NED map of the region at the end of crystallographic refinement.

Table 4

Summary of model statistics

Model	R†	r.m.s. deviation from ideality: Bonds (Å)	r.m.s. deviation from ideality: Angles (°)	r.m.s. difference from model 14 (Å): Main chain	r.m.s. difference from model 14 (Å): Side chain	⟨ΔΦ⟩ from model 14(°)	Comment
1	0·32	0·031	5·7	4·94		67	PAP-S sequence, constrained NCS
2	0·32	0·029	5·5	5·08		62	
3	0·32	0·028	5·4	5·05		59	
4	0·29	0·027	5·1	4·46	5·69	56	PAP sequence
5	0·27	0·019	4·4	4·34	5·49	48	Restrained NCS
6	0·24	0·024	4·7	4·30	5·40	44	
7	0·22	0·023	4·3	1·14	2·19	37	Major sequence realignment
8	0·21	0·021	4·0	0·95	1·90	33	
9	0·22	0·022	3·9	0·94	1·81	30	Extended to 2·5 Å
10	0·21	0·021	3·9	0·47	1·00	29	
11	0·18	0·021	3·8	0·31	0·76	26	Refine individual B values
12	0·17	0·020	3·6	0·28	0·67	23	Add bound waters
13	0·16	0·020	3·5	0·25	0·53	21	
14	0·15	0·019	3·6				No NCS restraints

† R factor calculated for 2σ data in 5·0 to 2·8 Å shell. ⟨ΔΦ⟩ = mean |ΔΦ| from model 14.

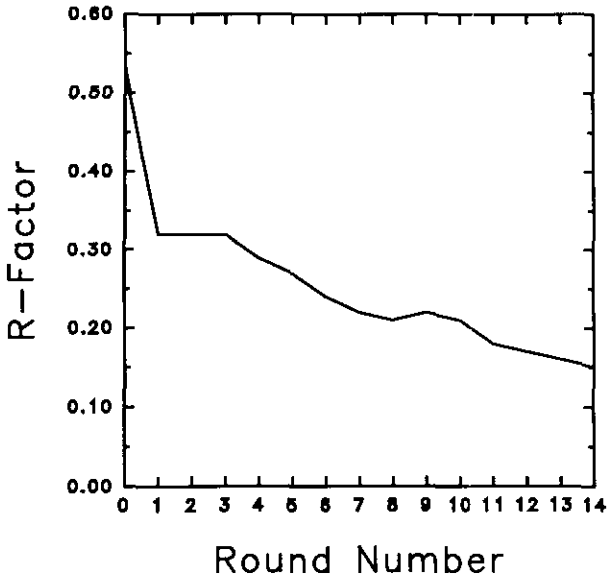


Figure 2. The R factor as a function of rounds of crystallographic refinement.

Progress through the first three rounds was slow with virtually no change in the refined R factor (=0·32). When the amino acid sequence of PAP became known from the DNA sequence (Lin *et al.*, 1991), it was built into model 4. The chemical sequence determination of the isozyme from pokeweed seeds, PAP-S, indicated the presence of two disulfide bonds. Alignment of the sequences of the two isozymes, which are 80% identical, showed that PAP probably has the same disulfides. Electron density for two disulfides was located, increasing our confidence in the chain tracing. Inclusion of the correct amino acid sequence in model 4 resulted in a drop in the R factor to 0·29.

As mentioned in Materials and Methods, restrained rather than constrained non-crystallographic symmetry was used beginning with round 5. The resulting refined monomers were still virtually identical (r.m.s. difference for main-chain atoms = 0·08 Å), but the symmetry relationship between monomers changed slightly and the R factor decreased to 0·27.

The NED map of model 6 revealed problems fitting residues 106 to 161. In particular it showed a loop of unaccounted density at the C-terminal end and insufficient density to accommodate the sequence at the other end of this stretch of residues. A sliding shift of four residues resulted in a better fit of the chain along the entire segment. The R factor from this refined model was 0·22.

With round 9, the C-terminal amino acid, which appeared to be somewhat disordered, was added to the model and the resolution limit of the refinement was extended to 2·5 Å. Individual atomic temperature factors were refined beginning with round 11 and water molecules were added beginning with model 12. The final model contains 87 bound water molecules and has an R factor of 0·17 for the 5·0 to 2·5 Å shell of data. Figure 1(d) shows the portion of the NED map from the final model corresponding to the maps displayed in Figure 1(a), (b) and (c); together they illustrate the progress made during model refinement.

(d) *Model description*

Non-crystallographic restraints were not used in the final round of model refinement allowing the two PAP molecules to move independently. Figure 3 shows a superposition of the two C^{α} traces; the r.m.s. deviation was 0·43 Å while the deviation of all atoms was 0·86 Å. For reference, the r.m.s. deviations between the two molecules before the final round of refinement were 0·04 Å and 0·10 Å, respectively, for main-chain and side-chain atoms. There

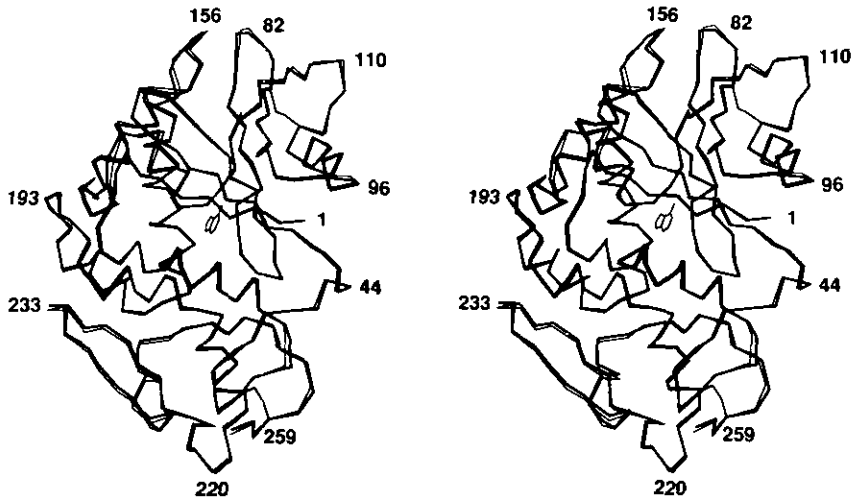


Figure 3. Least squares superposition of the 2 PAP molecules. The 2 molecules of the asymmetric unit were allowed to refine in an unrestrained fashion during the last round of refinement. The 2 disulfides are shown with medium width bonds. The binding site of the adenine-like pyrazolopyrimidine ring of FMP is also shown. The C^{α} traces show that no major differences in structure exist between the two.

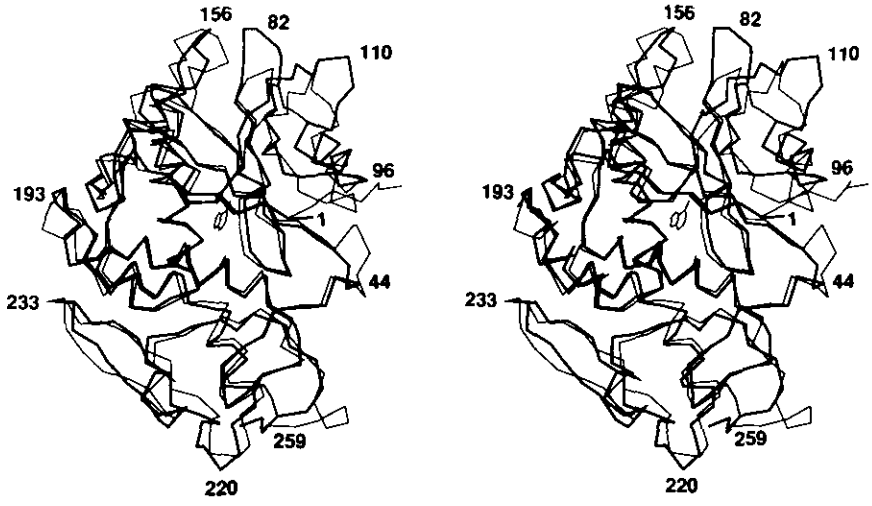


Figure 4. A comparison of PAP and ricin A chain. The 2 C^{α} traces have been superimposed by least squares. PAP is shown as the dark bonds and RTA as light. The 2 disulfides of PAP are shown with medium width bonds. The PAP binding site of the pyrazolopyrimidine ring of FMP is also shown. The overall folding and secondary structure patterns are very similar.

are no major conformational differences between the two molecules of the asymmetric unit; and for purposes of structural description and comparison with other toxins, the average of the two models will be used.

The sequence of PAP deduced from the cDNA sequence contains 313 amino acid residues, but the crystal structure shows that PAP contains 262 residues, similar to the 261 residues found in the sequence determination of PAP-S. From the chemically determined sequence of the amino terminus (Ready *et al.*, 1984), we know that the cDNA sequence contains an amino-terminal extension of 22 amino acid residues. Based on the crystal structure, we can conclude that the cDNA sequence also contains a carboxyl-terminal extension of 29 amino acid residues. These sequences may play a role in targeting PAP for the cell wall and are, most likely, removed to form the mature PAP. Amino and carboxyl-terminal extensions have been observed with several other RIPs (Irvin & Uckun, 1992).

The archetype of the RIP family is RTA and its structure has been described in detail (Katzin *et al.*, 1991). Figure 4 shows a least squares superposition of the C^{α} traces of RTA and PAP; the r.m.s. deviation is 2·3 Å. It is clear that the two folds are very similar and possess the same elements of secondary structure, that is, eight alpha helices and a beta sheet consisting of six strands. This notion is amplified in Figure 5(a), which displays the amino acid sequence alignment of the two proteins and the elements of secondary structure observed from the crystallographic models. Figure 5(b) shows the deviation between corresponding C^{α} atoms of PAP and RTA as a function of the amino acid sequence. Unlike RTA, PAP contains two disulfide bonds, one between residues 34 and 259 and the other between residues 85 and 106.

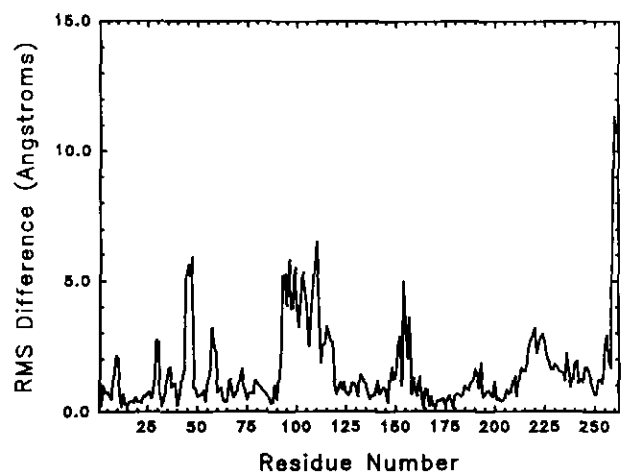
The active site of RTA has been described (Katzin *et al.*, 1991), and site-directed mutagenesis used to confirm the significance of various residues to enzyme activity (Frankel *et al.*, 1990; Ready *et al.*, 1991). An overall mechanism of action has been proposed incorporating the observed binding of substrate analogs (Monzingo & Robertus, 1992). In RTA nomenclature the key active site residues include Glu177 (176 in PAP) and Arg180 (179), involved in catalysis and tyrosines 80 (72) and 123 (123) involved in binding the target adenine base; all of these residues are invariant in the RIP family of toxins. The backbone carbonyl oxygen and nitrogen of Val81 (73) also makes specific hydrogen bonding interactions with the bound substrate adenine ring. Figure 6 shows a least squares superposition of the active-site regions of RTA and PAP and, again, shows that the toxins are very similar in structure. The largest variation is with PAP Tyr72 (RTA Tyr80) which rotates slightly between the proteins. In fact, we have seen some movement of this residue in RTA in order to accommodate the binding of substrate analogs (Monzingo & Robertus, 1992).

(e) *Binding of a substrate analog*

Formycin 5′-monophosphate is a substrate analog with an adenine-like pyrazolopyrimidine moiety. It binds to the RTA active site and accommodates the base in a fashion similar to that of adenine in the dinucleotide ApG (Monzingo & Robertus, 1992). Data were collected from a PAP crystal soaked with FMP, and a 2·8 Å difference electron density map calculated with amplitudes $(F_{\text{complex}} - F_{\text{nat,obs}})$ showed that FMP was bound in the putative active site of both monomers of the asymmetric unit. An omit electron map with amplitudes $(2F_{\text{complex}} - F_{\text{nat,calc}})$ and with phases calculated from the model from which active site residues and solvent had been omitted proved to be more suitable for model building. Figure 7(a) shows the binding site region

```
                    10         20         30         40
PAP          VNTIIYNVGSTTISKYATFLNDLRNEAKDPSLKCYGIPMLPNTNTNP
             bbbbbbb      aaaaaaaaaaaaaaa
             bbbbbbbb     aaaaaaaaaaaaaaa
RTA     IFPKQYPIINFTTAGATVQSYTNFIRAVRGRLTTGADVRHEIPVLPNRVGLPINQ
              10         20         30         40         50

         50         60         70         80         90         100
PAP     KYVLVELQGSNKRTITLMLRRNNLYVMGYSDPFETNKCRYHIFNDISGTERQDVE
          bbbbbbb    bbbbbbbb     bbbbbbbbb    bbbbbbb       aaaaaa
          bbbbbb     bbbbbb         bbbbb        bbbbb       aaaaaa
RTA     RFILVELQNHAEISVTLALDVTNAYVVGYRAG      NSAYFFHPDNQEDAEAI
           60         70         80              90         100

             110        120        130        140        150
PAP     TTLCPNANSRVSKNINFDSRYPTLESKAGVKSRSQVQLGIQILDSNIGKIS  GVM
        aaa          bbbbb      aaaaaaaaa           aaaaaaaa
                       bbb      aaaaaa              aaaaaaaaaaaa
RTA     THLF  TDVQNRYTFAFGGNYDRLEQLAGN LRENIELGNGPLEEAISALYYYSTG
              110        120        130        140        150

          160        170        180        190          200
PAP     SFTEKTEAEFLLVAIQMVSEAARFKYIENQVKTNF  NRAFNPNPKVLNLQETW
           aaaaaaaaaaaaaaaaaaaa  aaaaaaaa            aaaaaaaa a
           aaaaaaaaaaaaaaaaaaaa  aaaaaaaaa           aaaaaaaaaa
RTA     GTQLPTLARSFIICIQMISEAARFQYIEGEMRTRIRYNRRSAPDPSVITLENSW
        160        170        180        190        200        210

       210        220        230        240        250        260
PAP     GKISTAIHDAKNGVLPKPLELVDASGAKWIVLRVDEIKPDVALLNYVGGSCQTT
        aaaaaaaa
        aaaaaaaa
RTA     GRLSTAIQESNQGAFASPIQLQRRNGSKFSVYDVSILIPIIALMVYRCAPPPSSQF
              220        230        240        250        260
```

(a)



(b)

Figure 5. Comparisons of PAP and RTA. (a) The amino acid sequences of PAP and RTA have been aligned according to their structural overlaps. Regions of α helix and β strands are indicated by a and b, respectively, between the sequences. (b) A plot of the C^{α} deviations between PAP and RTA. PAP residue numbers are used in the plot.

of the omit map for the PAP·FMP complex. The compound can be fit readily and refinement statistics suggest that it binds somewhat more tightly to PAP than to RTA. That is, the PAP crystal appears to be 100% occupied while RTA was about 50%. The R factor for the refined PAP·FMP complex is 0·20 for 5·0 to 2·8 Å data. The r.m.s. deviation from ideality for bonds is 0·013 Å and for angles is 3·03°. Figure 7(b) shows the FMP interactions with PAP active site residues; the binding is essentially identical to that seen previously in RTA. The formycin ring is sandwiched between tyrosines 72 and 123 (80 and 123 in RTA nomenclature), and we see a 15° rotation about the C^{α}–C^{β} bond of Tyr72 to facilitate this. N-6 donates a hydrogen bond to the carbonyl oxygen of residue 73, and N-1 receives a bond from the amido nitrogen of Val73 (the analog of Val81 in RTA). As expected from the hypothesized mechan-

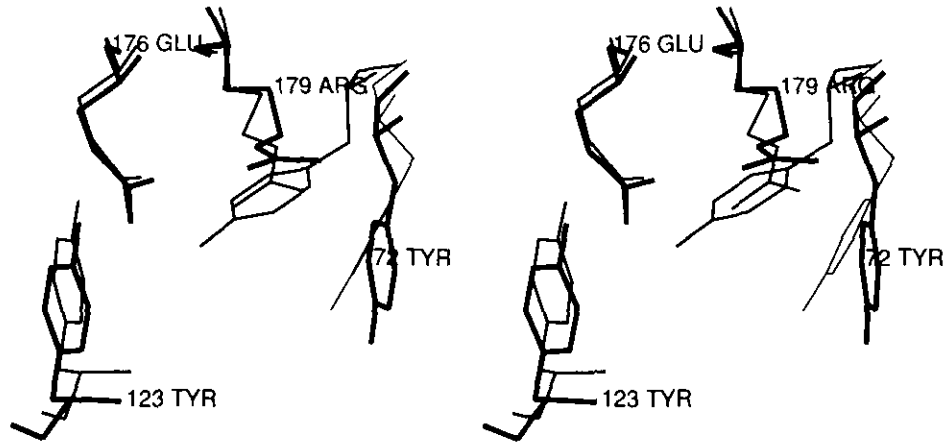


Figure 6. A superposition of the active site residues of PAP and RTA. The active site groups have been aligned by least squares. PAP is the dark bonded structure and RTA the medium. Tyr80 of the ricin · FMP complex is shown with light bonds.

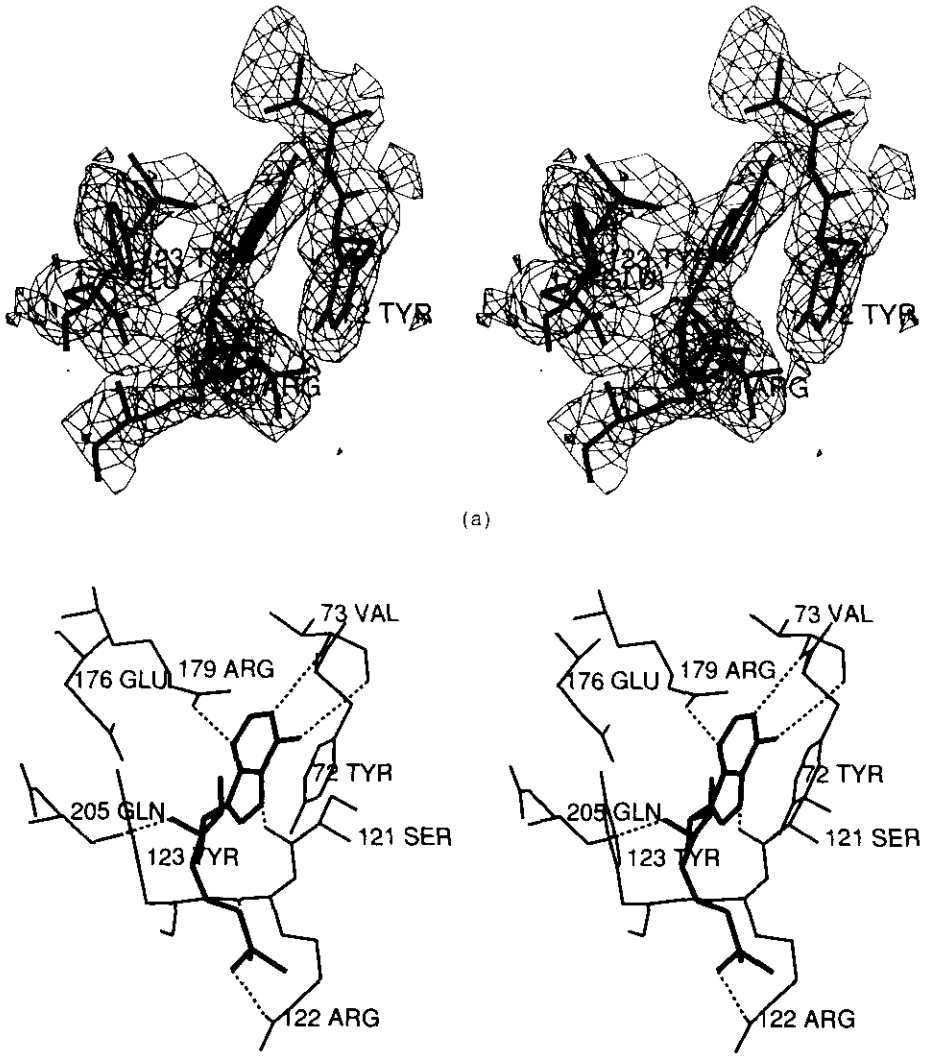


Figure 7. The bonding of FMP to PAP. (a) A 2·8 Å $2F_{complex} - F_{nat,calc}$ omit map showing the electron density for bound FMP and for active site residues. (b) The binding of FMP in the PAP active site. Hydrogen bonds are shown as broken lines.

ism of this enzyme class, N-3 of formycin receives a hydrogen bond from Arg179 (180 in RTA). In addition, N-7 appears to donate a hydrogen bond to the carbonyl oxygen of Ser121.

The similar fold, active site configuration, and analog binding between RTA and PAP suggest that they function in a similar manner. The mechanism of action proposed for RTA (Ready *et al.*, 1991; Molnzingo & Robertus, 1992) is appropriate for PAP. Indeed, the PAP structure lends credence to that proposal in that all the key features thought to be important in depurination by ricin are conserved.

One interesting difference between PAP and ricin is the observation that PAP can attack bacterial ribosomes, whereas RTA cannot (Hartley *et al.*, 1991). Since the binding sites for adenine and the catalytic sites are so similar, it appears that this difference results from modest structural differences at other positions in the proteins. We have previously suggested that ribosomal recognition probably involves interactions remote from the adenine binding site (Rutenber *et al.*, 1991). The topography of bacterial and eukaryotic ribosomes undoubtedly varies near the susceptible rRNA target sequence. Minor structural differences between members of the RIP family may allow or prevent binding of the target loop. Figures 4 and 5(b) show comparisons of the C^{α} backbones of RTA and PAP; several regions show enough difference to warrant consideration as causes for the differences in ribosome recognition. Residues 120 to 124 form a flap over the adenine-binding site and show considerable structural variation between the enzymes. The close proximity to the known adenine-binding site makes this a prime suspect to account for differences in ribosomal binding. On the back side of the protein the loops connecting β strands e and f (residues 89 to 93 and 115 to 117, respectively, in RTA) also show relatively large differences between PAP and RTA. Because ribosome binding may involve regions of the toxin remote from the catalytic site, these differences may also be worth exploring.

(f) *Rotation function*

With 30% sequence homology and the large extent of structural homology observed between PAP and RTA, one might expect that a traditional molecular replacement solution would have been possible. In fact, rotation searches using the RTA model and PAP diffraction data reveal two plausible solutions. Using the 6 to 8 Å shell of diffraction data and a radius of integration of 16 Å, the top two peaks correspond to orientations of the RTA molecule which are related by a 2-fold rotation about the a axis. These two peaks are also among the highest with radii of integration 18 and 20 Å. As it turns out, this rotation solution is incorrect. Using the 4 to 5 Å shell of data and integration radius 18 or 20 Å, the top two peaks correspond to what we have independently found to be the correct orientations.

(g) *Crystal packing*

As mentioned above, only two of the six heavy atom sites ($Pt(CN)_4^{2-}$ sites 3 and 4) obey the non-crystallographic symmetry operator. These atoms are bound by the side-chains of Lys15 of the respective monomers. The $Pt(NO_2)_4^{2-}$ site 2 lies very nearly on the pseudo-2-fold between the monomers, being bound by Lys210 from both monomers. The triclinic packing of the crystal apparently causes the asymmetry of the remaining three sites. These heavy atoms all bind in a channel formed at the intersection of three dimers (x, y, z; $x-1, y-1, z$; $x-1, y, z-1$). $Pt(CN)_4^{2-}$ sites 1 and 2 each bind to two lysine side-chains from adjacent molecules. The $Pt(NO_2)_4^{2-}$ site 1 binds to the side-chain of Met156.

The two monomers of the dimer are related by a rotation of 177° about an axis which is only 0·6° from the crystallographic a axis with a translation of about 0·09 Å. The dimer is related to an adjacent dimer ($x+1, y, z+1$) by a rotation of $-177°$ about the same axis with a translation of 24·8 Å, virtually half the length of the cell along a. Thus, the crystal packing is nearly monoclinic (space group $C2$ or $P2_1$) with two dimers in each unit cell. The pseudo-monoclinic cell can be constructed from the triclinic lattice points and has dimensions $a=49·5$ Å, $b=75·0$ Å, $c=74·8$ Å, $\alpha=106·3°$, $\beta=88·9°$, $\gamma=87·5°$, where a is the "unique" axis.

An increase in favorable packing contacts appears to be the reason for the triclinic packing being favored over a true monoclinic packing. There are four salt links (Arg A122-Asp B217, Asp A217-Arg B122, Glu A223-Lys B236 and Lys B236-Glu A223) and three hydrogen bonds formed between the two monomers of the PAP dimer. In a model dimer with a true 180° rotation, only two salt links (A122-B217 and A217-B122) are observed. When the observed PAP dimer is modeled into a monoclinic cell, there are four salt links, 14 hydrogen bonds, and four hydrophobic contacts potentially formed with symmetry-related molecules. In the triclinic cell, the same number of salt links and hydrophobic contacts are observed, but there are 24 additional hydrogen bonds.

We are grateful to Raquelle Smalley for her help in preparing the Figures. This work was supported by grant GM30048 from the National Institutes of Health, and by a grant from the Foundation for Research.

References

Bonness, M. S. (1992). Doctoral dissertation, University of Texas.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallogr. sect. A*, **32**, 832–847.

Brunger, A. T. (1988). Crystallographic refinement by simulated annealing. In *Crystallographic Computing 4: Techniques and New Technologies* (Isaacs, N. W. & Taylor, M. R., eds), Clarendon Press, Oxford.

Duggar, B. M. & Armstrong, J. K. (1925). The effect of treating the virus of tobacco mosiac with the juices of

various plants. *Ann. Missouri Bot. Garden*, **12**, 359–366.

Endo, Y. & Tsurugi, K. (1988). The RNA N-glycosidase activity of ricin A-chain: the characteristics of the enzymatic activity of ricin A-chain with ribosomes and with rRNA. *J. Biol. Chem.* **263**, 8735–8739.

Endo Y., Chan, Y. L., Lin, A., Tsurugi, K. & Wool, I. G. (1988). The cytotoxins α-sarcin and ricin retain their specificity when tested on a synthetic oligoribonucleotide (35-mer) that mimics a region of the 28 S ribosomal ribonucleic acid. *J. Biol. Chem.* **263**, 7917–7920.

Fitzgerald, P. M. (1988). MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. *J. Appl. Crystallogr.* **21**, 273–278.

Frankel, A., Welsh, P., Richardson, J. & Robertus, J. D. (1990). The role of arginine 180 and glutamic acid 177 of ricin toxin A chain in the enzymatic inactivation of ribosomes. *Mol. Cell. Biol.* **10**, 6257–6263.

Hamlin, R. (1985). Multiwire area X-ray diffractometers. *Methods Enzymol.* **114**, 416–452.

Hartley, M. R. Legname, G., Osborn, R., Chen, Z. & Lord, J. M. (1991). Single-chain ribosome inactivating proteins from plants depurinate *Escherichia coli* 23S ribosomal RNA. *FEBS Letters*, **290**, 65–68.

Howard, A. J., Nielsen, C. & Xoung, N. H. (1985). Software for a diffractometer with multiwire area detector. *Methods Enzymol.* **114**, 453–472.

Irvin, J. D. (1975). Purification and partial characterization of the antiviral protein from *Phytolacca americana* which inhibits eukaryotic protein synthesis. *Arch. Biochem. Biophys.* **169**, 522–528.

Irvin, J. D. & Uckun, F. M. (1992). Pokeweed antiviral protein: ribosome inactivation and therapeutic applications. *Pharmacol. Therapeut.* **55**, 279–302.

Jones, T. A. (1982). FRODO: A graphics fitting program for macromolecules. In *Computational Crystallography* (Sayre, D., ed.), pp. 303–317, Oxford University Press, Oxford.

Kassanis, B. & Kleczkowski, A. (1948). The isolation and some properties of a virus-inhibiting protein from *Phytolacca esculenta*. *J. Gen. Microbiol.* **2**, 143–153.

Katzin, B. J., Collins, E. J. & Robertus, J. D. (1991). The structure of ricin A chain at 2·5 Å. *Proteins*, **10**, 251–259.

Kim, Y., Mlsna, D., Monzingo, A. F., Ready, M. P., Frankel, A. & Robertus, J. D. (1992). The structure of a ricin mutant showing rescue of activity by a noncatalytic residue. *Biochemistry*, **31**, 3294–3296.

Kung, S., Kimura, M. & Funatsu, G. (1990). The complete amino acid sequence of antiviral protein from the seeds of pokeweed (*Phytolacca americana*). *Agri. Biol. Chem.* **54**, 3301–3318.

Lin, Q., Chen, Z. C., Antoniw, J. F. & White, R. F. (1991). Isolation and characterization of a cDNA clone encoding the anti-viral protein from *Phytolacca americana*. *Plant Mol. Biol.* **17**, 609–614.

Monzingo, A. F. (1979). Doctoral dissertation, University of Texas.

Monzingo, A. F. & Robertus, J. D. (1992). X-ray analysis of substrate analogs in the ricin active site. *J. Mol. Biol.* **227**, 1136–1145.

Ready, M. P., Bird, S., Rothe, G. & Robertus, J. D. (1983). Requirements for antiribosomal activity of pokeweed antiviral protein. *Biochim. Biophys. Acta*, **740**, 19–28.

Ready, M. P., Wilson, K., Piatak, M. & Robertus, J. D. (1984). Ricin-like plant toxins are evolutionarily related to single-chain ribosome-inhibiting proteins from *Phytolacca*. *J. Biol. Chem.* **259**, 15252–15256.

Ready, M. P., Brown, D. T. & Robertus, J. D. (1986). Extracellular localization of pokeweed antiviral protein. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 5053–5056.

Ready, M. P., Kim, Y. & Robertus, J. D. (1991). Site-directed mutagenesis of ricin A chain and implications for the mechanism of action. *Proteins*, **10**, 270–278.

Robertus, J. D., Monzingo, A. F. & Irvin, J. D. (1977). Preliminary X-ray diffraction studies on an anti-viral protein. *Biochem. Biophys. Res. Commun.* **74**, 775–779.

Rutenber, E. & Robertus, J. D. (1991). The structure of ricin B chain at 2·5 Å resolution. *Proteins*, **10**, 260–269.

Rutenber, E., Katzin, B. J., Collins, E. J., Mlsna, D., Ernst, S., Ready, M. P. & Robertus, J. D. (1991). The crystallographic refinement of ricin at 2·5 Å resolution. *Proteins*, **10**, 240–250.

Sussman, J. L. (1985). Constrained-restrained least-squares (CORELS) refinement of proteins and nucleic acids. *Methods Enzymol.* **115**, 271–302.

Tanaka, N. (1977). Representation of the fast-rotation function in a polar coordinate system. *Acta Crystallogr. sect. A*, **33**, 191–193.

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112.

Zarling, J. M., Moran, P. A., Haffar, O., Sias, J., Richman, D. D., Spina, C. A., Myers, D. E., Kuebelbeck, V., Ledbetter, J. A. & Uckun, F. M. (1990). Inhibition of HIV replication by pokeweed antiviral protein targeted to $CD4^+$ cells by monoclonal antibodies. *Nature (London)*, **347**, 92–95.

Edited by W. Hendrickson