The 2.5 Å Structure of Pokeweed Antiviral Protein

Arthur F. Monzingo1, Edward J. Collins1†, Stephen R. Ernst1
James D. Irvin2 and Jon D. Robertus1†

1Department of Chemistry and Biochemistry
University of Texas, Austin, TX 78712, U.S.A.
2Department 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 50 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; Kasanis & Kleczkowski, 1948). The pokeweed proteins are all representatives of the class-I 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-I 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-I 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 28S 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
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 (Zurling 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 R1P, 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 (K₂Pt(CN)₄·3H₂O) and potassium tetratritrilotitanate (K₂Pt(NO₃)₂) were purchased from Apache Chemicals, Inc. Formylin 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 = 49.5 Å, b = 50.1 Å, c = 65.2 Å, α = 86.0°, β = 113.2°, γ = 116.5°. 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 K₂Pt(CN)₄ for 10 days. The Pt(NO₃)₂⁻² derivative was prepared by soaking crystals in artificial mother liquor containing 5 mM K₂Pt(NO₃)₂ for 4 weeks. The PAP crystal with FMP bound was prepared by soaking in artificial mother liquor containing 5 mM FMP for 4 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 sources 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 COBRELS (Sussman, 1985). A symmetry-averaged map was generated using the method of Bricogne (1976). Model building was done using the program PRODO (Jones, 1982) on an Evans and Sutherland PS90 graphics system.

Crystallographic refinement of PAP used the X-PLOR package (Brünger, 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, NEDFPT. 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_{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_{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 COBRELS rigid body refinement. For rounds 5 to 13, weighted restraints 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_{native}) and phases from the native refinement. Omit maps were computed with amplitudes of the form (2F_{complex} - F_{native}), 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
The Structure of PAP

Table 1

Data collection statistics

<table>
<thead>
<tr>
<th></th>
<th>Native Pt(CN)(^{2-})</th>
<th>Pt(NO(_{3}))(^{2-})</th>
<th>FMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of observations</td>
<td>77,556</td>
<td>128,906</td>
<td>128,430</td>
</tr>
<tr>
<td>No of reflections</td>
<td>19,127</td>
<td>14,074</td>
<td>13,874</td>
</tr>
<tr>
<td>Resolution limit (Å)</td>
<td>2.8</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>(R_{merge})(†)</td>
<td>0.056</td>
<td>0.128</td>
<td>0.114</td>
</tr>
<tr>
<td>Mean isomorphous difference (%)</td>
<td>2.0</td>
<td>1.70</td>
<td>2.17</td>
</tr>
</tbody>
</table>

\(† R_{merge} = \Sigma |I - \bar{I}|/\Sigma I\) 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 and have been assigned the accession numbers 1PAF and 1PAQ, 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)\(^{2-}\) 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\(_{3}\))\(^{2-}\) 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)\(^{2-}\) sites. This was disconcerting but for both compounds difference Fourier's 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

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Site no.</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Relative occupancy</th>
<th>Residues bound to$^{\dagger}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt(CN)(^{2-})</td>
<td>1</td>
<td>0.093</td>
<td>0.117</td>
<td>0.188</td>
<td>0.073</td>
<td>Lys A246</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.912</td>
<td>-0.883</td>
<td>-0.823</td>
<td>0.455</td>
<td>Lys B33</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.323</td>
<td>0.334</td>
<td>-0.103</td>
<td>0.417</td>
<td>Lys B115</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.003</td>
<td>0.592</td>
<td>1.169</td>
<td>0.396</td>
<td>Lys B15</td>
</tr>
<tr>
<td>Pt(NO(_{3}))(^{2-})</td>
<td>1</td>
<td>1.102</td>
<td>1.068</td>
<td>0.198</td>
<td>0.430</td>
<td>Met A156</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.365</td>
<td>0.388</td>
<td>0.523</td>
<td>0.606</td>
<td>Lys A210</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lys B210</td>
</tr>
</tbody>
</table>

$^{\dagger}$ A and B indicate respective monomers of the PAP dimer.

$^{\ddagger}$ Indicates a residue of a symmetry-related molecule.

Table 3

Multiple isomorphous replacement statistics

<table>
<thead>
<tr>
<th></th>
<th>8.85</th>
<th>6.25</th>
<th>5.10</th>
<th>4.42</th>
<th>3.95</th>
<th>3.61</th>
<th>3.34</th>
<th>3.13</th>
<th>2.95</th>
<th>2.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_{min}) (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pt(CN)(^{2-})</td>
<td>354</td>
<td>725</td>
<td>959</td>
<td>1112</td>
<td>1276</td>
<td>1266</td>
<td>1403</td>
<td>1527</td>
<td>1616</td>
<td>1730</td>
</tr>
<tr>
<td>(f_o)</td>
<td>73</td>
<td>68</td>
<td>93</td>
<td>99</td>
<td>59</td>
<td>54</td>
<td>50</td>
<td>47</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td>(E)</td>
<td>28</td>
<td>24</td>
<td>25</td>
<td>30</td>
<td>34</td>
<td>34</td>
<td>35</td>
<td>31</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>Pt(NO(_{3}))(^{2-})</td>
<td>354</td>
<td>713</td>
<td>935</td>
<td>1060</td>
<td>1249</td>
<td>1372</td>
<td>1507</td>
<td>1590</td>
<td>1711</td>
<td>1753</td>
</tr>
<tr>
<td>(f_o)</td>
<td>53</td>
<td>50</td>
<td>46</td>
<td>43</td>
<td>39</td>
<td>37</td>
<td>34</td>
<td>32</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>(E)</td>
<td>47</td>
<td>30</td>
<td>29</td>
<td>34</td>
<td>37</td>
<td>36</td>
<td>33</td>
<td>29</td>
<td>26</td>
<td>25</td>
</tr>
</tbody>
</table>

Mean figure of merit | 0.73 | 0.66 | 0.65 | 0.97 | 0.68 | 0.60 | 0.60 | 0.57 | 0.57 |

\(f_o\) is the r.m.s. heavy atom structure factor. \(E\) is lack of closure.
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 rein 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 100 to 2.8 Å resolution data; the XED map is from 100 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 XED map of the region at the end of crystallographic refinement.

Table 4
Summary of model statistics

<table>
<thead>
<tr>
<th>Model</th>
<th>R†</th>
<th>Bond (Å)</th>
<th>Angle (°)</th>
<th>Main chain</th>
<th>Side chain</th>
<th>ΔΦ (°)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.32</td>
<td>0.021</td>
<td>57</td>
<td>4.94</td>
<td>5.08</td>
<td>67</td>
<td>PAP-S sequence, constrained NCS</td>
</tr>
<tr>
<td>2</td>
<td>0.32</td>
<td>0.029</td>
<td>56</td>
<td>5.05</td>
<td>5.05</td>
<td>66</td>
<td>PAP sequence</td>
</tr>
<tr>
<td>3</td>
<td>0.32</td>
<td>0.028</td>
<td>54</td>
<td>5.05</td>
<td>5.05</td>
<td>59</td>
<td>Restricted NCS</td>
</tr>
<tr>
<td>4</td>
<td>0.29</td>
<td>0.027</td>
<td>51</td>
<td>4.96</td>
<td>5.09</td>
<td>56</td>
<td>PAP sequence</td>
</tr>
<tr>
<td>5</td>
<td>0.27</td>
<td>0.019</td>
<td>44</td>
<td>4.34</td>
<td>5.49</td>
<td>46</td>
<td>Restricted NCS</td>
</tr>
<tr>
<td>6</td>
<td>0.24</td>
<td>0.023</td>
<td>47</td>
<td>4.30</td>
<td>5.40</td>
<td>44</td>
<td>Major sequence realignment</td>
</tr>
<tr>
<td>7</td>
<td>0.22</td>
<td>0.023</td>
<td>45</td>
<td>1.14</td>
<td>2.19</td>
<td>37</td>
<td>Extended to 2.5 Å</td>
</tr>
<tr>
<td>8</td>
<td>0.21</td>
<td>0.021</td>
<td>40</td>
<td>0.95</td>
<td>1.90</td>
<td>33</td>
<td>Add bound waters</td>
</tr>
<tr>
<td>9</td>
<td>0.22</td>
<td>0.022</td>
<td>39</td>
<td>0.94</td>
<td>1.81</td>
<td>30</td>
<td>Add bound waters</td>
</tr>
<tr>
<td>10</td>
<td>0.21</td>
<td>0.021</td>
<td>39</td>
<td>0.47</td>
<td>1.00</td>
<td>29</td>
<td>Refine individual B values</td>
</tr>
<tr>
<td>11</td>
<td>0.18</td>
<td>0.020</td>
<td>36</td>
<td>0.31</td>
<td>0.76</td>
<td>26</td>
<td>No NCS restraints</td>
</tr>
<tr>
<td>12</td>
<td>0.17</td>
<td>0.020</td>
<td>36</td>
<td>0.28</td>
<td>0.47</td>
<td>23</td>
<td>No NCS restraints</td>
</tr>
<tr>
<td>13</td>
<td>0.16</td>
<td>0.020</td>
<td>35</td>
<td>0.25</td>
<td>0.53</td>
<td>21</td>
<td>No NCS restraints</td>
</tr>
<tr>
<td>14</td>
<td>0.15</td>
<td>0.019</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† R factor calculated for 25 data in 3.0 to 2.8 Å shell. ΔΦ = mean |ΔΦ| from model 14.
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 54 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 $C^\alpha$ 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 0.64 Å and 0.10 Å, respectively, for main-chain and side-chain atoms. There

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

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 pyrazolo-pyrimidine ring of FMP is also shown. The $C^\alpha$ traces show that no major differences in structure exist between the two.
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 Cα 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 Cα 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 250 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 (Mozingo & Robertus, 1992). In RTA nomenclature the key active site residues include Glu177 (176 in PAP) and Arg180 (178), 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 carboxyl 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 (Mozingo & 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 (Mozingo & 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_{obs} - F_{calc}) showed that FMP was bound in the putative active site of both monomers of the asymmetric unit. An omit electron map with amplitudes (2F_{complex} - F_{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
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 Cα 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 50 to 2.8 Å data. The r.m.s. deviation from ideality for bonds is 0.013 Å and for angles is 3.0°. 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 Cα-Cβ 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<sub>obs</sub>−F<sub>calc</sub> 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; Molinizio & 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 Cα 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)42- 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(NO3)42- site 2 lies very nearly on the pseudo-2-fold between the monomers, being bound by Lys210 from both monomers. The trichinic 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)42- sites 1 and 2 each bind to two lysine side-chains from adjacent molecules. The Pt(NO3)42- 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 0.99 Å. The dimer is related to an adjacent dimer (x+1, y, z+1) by a rotation of 180° about the same axis with a translation of 24.3 Å, virtually half the length of the cell along a. Thus, the crystal packing is nearly monoclinic (space group C2 or P2₁) with two dimers in each unit cell. The pseudomonoclinic cell can be constructed from the trichinic lattice points and basis dimensions a = 49.5 Å, b = 75.0 Å, c = 74.3 Å, α = 106.3°, β = 88.9°, γ = 87.5°, where a is the “unique” axis.

An increase in favorable packing contacts appears to be the reason for the trichinic 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 trichinic 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


Duggar, B. M. & Armstrong, J. K. (1965). The effect of treating the virus of tobacco mosaic with the juices of

Edited by W. Hendrickson