BUTTERWORTH HEINEMANN

Substrate recognition by ribosome-inactivating protein studied by molecular modeling and molecular electrostatic potentials

Gianpaolo Bravi, Giuseppe Legname, and A.W. Edith Chan

Italfarmaco Research Center, Cinisello Balsamo, Italy

A computer model of dianthin 30, a type 1 ribosome-inactivating protein (RIP), is constructed by homology modeling using two known X-ray structures; a type 1 RIP, pokeweed antiviral protein (PAP), and chain A of a type 2 RIP, ricin. The 3D structure is refined by molecular dynamics and its binding site compared with those of PAP and ricin using molecular electrostatic potential mapping. The differences in the maps obtained clearly show how, despite the similarity of the topology of the binding site, differences in electrostatic potential can account for the experimentally observed differences in substrate recognition and binding. This demonstrates the potential of these techniques for guiding further experimental analyses.

Keywords: ribosome-inactivating protein, RIP, dianthin, ricin, pokeweed antiviral protein, PAP, protein structure, homology modeling, molecular dynamics, molecular electrostatic potential, MEP

INTRODUCTION

Most plant tissues contain ribosome-inactivating proteins (RIPs) that inhibit protein synthesis through modification of ribosomal RNA. These proteins are highly specific RNA N-glycosidases; they catalytically cleave the glycosidic bond of a specific adenine located on a loop that is highly conserved in both prokaryotic and eukaryotic ribosomal RNA. The mechanism for the inactivation has been investigated in yeast, where the removal of adenine apparently inhibits the binding of elongation factor Tu to the eukaryotic ribosomes.[1]

Two major classes of RIPs have been reported: type 1 and type 2. Type 1 has a single polypeptide chain whereas type 2 contains two chains, A and B. The A chain of the latter contains the enzymatic moiety, which is homologous to the single chain of type 1. The B chain, also known as the lectin chain, confers high, nonspecific cytotoxicity.[1]

The X-ray crystal structures of both types of RIPs, represented by pokeweed antiviral protein (PAP, type 1)[2] and ricin (type 2),[3–6] have been obtained in the past 2 years. More recently two additional type 1 crystal structures have been published.[7,8] These structures confirm that the A chain from type 2 and the single chain from type 1 share the same three-dimensional fold. The adenine-binding site has been identified in the structures, and is almost identical in all cases. Yet, this appears to be in conflict with their functional role in inactivating the different kinds of ribosomes.

In this article, we investigate this apparent anomaly using homology modeling and molecular electrostatic potential (MEP) distributions. First, a homology model of type 1 RIP, dianthin 30 (dian),[9] was modeled on the known structures of ricin A chain and PAP and its structure analyses. Second, we compared and analyzed those residues close to the catalytic site of these three structures using MEP distributions projected on a surface map, the aim being to locate which regions could be responsible for substrate recognition and binding. The differences in the maps reflected a specific recognition by different RNA ribosomes. It is clear that it is this region that needs to be targeted for site-directed mutagenesis as a means of altering the specificity of binding of the RIPs to different RNA molecules.

METHODS

All the homology building, energy, and dynamics calculations were done using Quanta 3.3[10] (Protein Design Pack-

Color Plates for this article are on p. 109.

Address reprint requests to Dr. Chan at Hoechst Roussel, Kingfisher Drive, Covingham, Swindon, Wilts SN3 5BZ, UK. e-mail: edith@hrplsr1.demon.co.uk

Received 11 October 1994; accepted 25 October 1994

655 Avenue of the Americas, New York, NY 10010

0263-7855/95/$10.00
SSDI 0263-7855(94)00014-J

age) and CHARMM[11] on a Silicon Graphics Crimson machine. First, the sequences of ricin A chain (Brookhaven code: 1rtc, 2.3 Å)[6] and PAP (1paf, 2.5 Å)[7] were aligned based on $C_\alpha \times C_\alpha$ distance homology. The final alignment (Table 1) was obtained by matching dianthin to ricin and PAP, based on their sequence homology, followed by manual adjustment. There were six deletions and six insertions in the unknown dianthin structure. Residue matching was done in three ways: by identity, by similarity using the Dayhoff matrix,[12] and by similar secondary structure. The C_α root mean square (rms) differences among the three sequences using the three matching methods are reported in Table 2.

Only the averaged backbone coordinates of the structurally conserved regions (regions with defined secondary structures as stated in Table 1) of ricin and PAP were copied

Table 1. Alignment for ricin, pokeweed antiviral protein, and dianthin[a]

```
ric   1 IFPKQYPIINFTTAGATVQSYTNFIRAVRGRLTTGADVRHEIPVLPNRV GLPINQR
PAP   1       VNTIIYNVGSTTISKYATFLNDLRNEAKDPSLKCYGIPMLPNT  NTN  PK
dian  1    DAATAYTLNLANPSASQYSSFLDQIRNNVRDTSLIYGGTDVAVIGAPSTT  DK
                  β1              α1

ric  57 FILVELSNHAELSVTLALDVTNAYVVGYRAG      N SAYFFH PDN
PAP  49 YVLVELQGSNKKTITLMLRRNNLYVMGYSDPFETN KCRYHIFNDI S
dian 53 FLRLNFQGP RGTVSLGLRRENLYVVAYLAMDNANVNRAYYFKN QIT
           β2          β3          β4          β5

ric  99 QEDAEAITHL FTDV   QNRYTFAFGGNYDRLEQLA    GN LRENIELG
PAP  95 GTERQDVETTLCPNA NSRVSKNINFDSRYPTLESKA    GVKSRSQVQLG
dian 99 SAELTALFP  EVVV   ANQKQLEYGEDYQAIEKNAKITTGDQSRKELGLG
           α2              β6        α3

ric 141 NGPLEEAISALYYYSTGGTQLPTLARSFIICIQMISEAARFQYIEGEMRTRIRY
PAP 142 IQILDSNIGKIS GV MSFTEKTEAEFLLVAIQMVSEAARFKYIENQVKTNF
dian 146 INLLITMIDGV  NKKVR VVKDEARFLLIAIQMTAEAARFRYIQNLVTKNF
            α4                  α5                  α6

ric 195 NRRSAPDPSVITLENSWGRLSTAIQESNQGAFASPIQLQRRNGSKFSVYDV
PAP 192 NRAFNPNPKVLNLQETWGKISTAIHDAKNGVLPKPLELVDASGAKWIVLRV
dian 195 PNKFDSENKVIQFQVSWSKISTAIFGDCKNGVFNKDYDFGFGVRQAKDLQM
                        α7

ric 246 SILIPIIALMVYRCAPPPSSQF
PAP 243 DEIKPDVALLNYVGGSCQTT
dian 246 GLLKYLGRPKSSSIEANSTDDTADVL
            α8
```

[a]Numbering for ricin (1rtc) and PAP (1paf) is derived from their X-ray structures; that of dian is from dianthin 30 (Swiss-Prot accession number P24476). Secondary structure denotations are as follows: **loop fragment,** β-strand, α-helix. Residues in italic are those used in MEP calculations.

Table 2. C_α root mean square differences among the three structures ricin, dianthin, and PAP.[a]

Method of atom matching	Ricin–PAP	Ricin–dianthin	PAP–dianthin
Identical[b]	0.86 (72)	1.69 (53)	1.88 (76)
Similar[c]	1.57 (141)	2.16 (120)	2.00 (152)
Secondary structure[d]	1.43 (161)	1.65 (153)	1.70 (156)
breakdown of secondary structure			
β1 (7–13)[e]	1.35	0.97	1.13
α1 (18–32)	0.88	1.43	1.12
β2 (57–64)	1.33	2.10	1.40
β3 (69–75)	1.14	1.30	1.05
β4 (80–86)	1.12	1.90	1.59
β5 (89–93)	0.87	0.95	0.93
α2 (98–106)	**3.48**	**2.66**	**4.05**
β6 (114–116)	**2.63**	**3.49**	**1.39**
α3 (123–130)	1.06	1.26	0.86
α4 (141–151)	1.57	1.33	1.25
α5 (161–180)	0.60	1.26	1.15
α6 (183–191)	0.71	1.35	1.45
α7 (202-220)	1.09	1.73	1.40
α8 (246–251)	1.40	1.77	2.13

[a]All atom matches were done using the alignment in Table 1.
[b]Atom matching by identical residues.
[c]Atom matching by similar residues (according to the Dayhoff matrix).
[d]Atom matching by comparing the same secondary structures and the breakdown of each secondary elements. In the first line of each comparison, the first number is the C_α rms value (Å) while the value inside parentheses indicates the number of residues matched by that particular method.
[e]Numbering refers to ricin.
Bold numbers indicate variable regions.

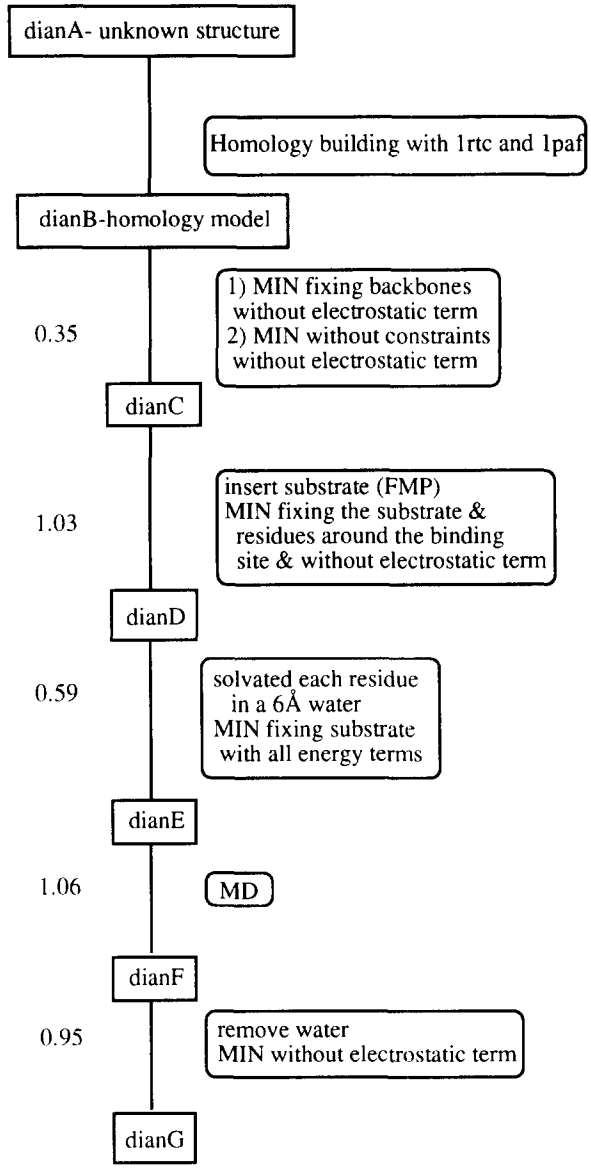


Figure 1. Complete model refinement procedure for dianthin. Values on the left, between models, indicate their C_α–rms differences (in angstroms).

onto the dianthin unknown structure (dianA). In this model, the side chains of the dianthin residues that must be transformed in the unknown structure were built as far as possible on the basis of the coordinates of the corresponding residues in the template structures.

The choices of fragments for the insertions were selected from a fragment database containing a $C_\alpha \times C_\alpha$ distance matrix generated from a set of known protein structures of resolution better than 3.0 Å.[13] Fragments were least squares fitted to the corresponding regions and ranked on the basis of compatibility with the overall structure, a good least squares fit, and a good geometrical fit in the overlap regions. Each fragment (Table 3) was selected primarily on the basis of the rms values from the least-squares fit, unless obvious steric hindrance with the remainder of the protein was indicated by visual inspection.

The refinement of the dianthin model was done in several stages of minimization and dynamics. The whole procedure, together with the C_α–rms distances between each model, is shown in Figure 1. All the minimizations (MIN) were performed using the Adpoted-Basis Newton Raphson (ABNR) method until the rms gradient was equal to or less than 0.05 kcal/mol Å. The homology model (dianB) was minimized twice in vacuo with the electrostatic energy term turned off: first with its backbone fixed, then without any constraints. The model dianC was obtained in the same way.

At this stage, the substrate formycin 5′-monophosphate (FMP, geometry taken from 1pag)[7] was added into the model because we wanted to observe the effect of the substrate inside the RIP during dynamics. This model (dianD) was minimized in a 6-Å water shell[14] without constraints and with all energy terms considered: the result is labeled as dianE.

The molecular dynamics (MD) were performed at 300 K. During 6 ps of heating, 10 ps of equilibration, and the first 10 ps of simulation, the coordinates of the substrate were fixed. Then another 60 ps of simulation at 300 K was carried out without any constraints on the complex. The structure at the last step of the second simulation stage was selected as dianF. The final structure, dianG, was obtained

by minimizing dianF without water and with the electrostatic term turned off.

The prediction of the isoelectric point (p*I*) of the protein was done using the program MacDNASIS Pro.[15] The MEP distributions were computed from partial atomic charges generated from CHARMM.

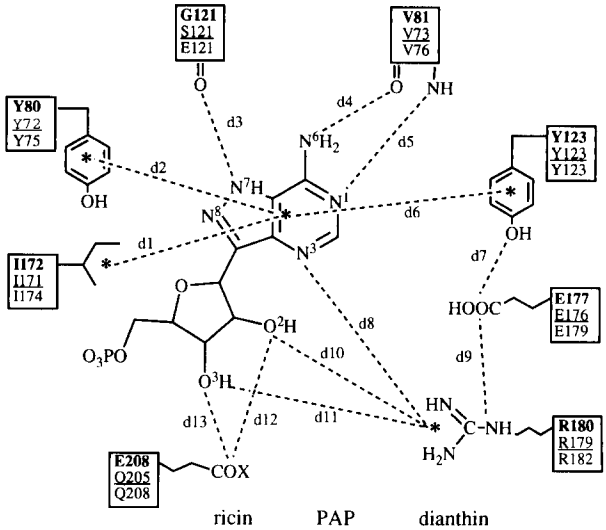
RESULTS AND DISCUSSION

The refined model structure of dianthin (dianG) retains the same fold as PAP and ricin, namely a six-stranded β sheet and eight α helices. The superimposed C_α traces of the three proteins are shown in Color Plate 1. The regions containing the secondary structures overlap well (Table 2), whereas the loop regions show some diversity. The program PROCHECK[16] was used to examine various geometrical parameters of these three structures (1rtc, 1paf, and dianG). The percentages of residues that reside in allowed regions in the Ramachandran plots are 98.7, 98.7, and 98.4%, respectively. Although several chi values for 1rtc (ricin) and 1paf (PAP) were outside the norm, those of the refined dianthin structure were within a reasonable range. The results produced by PROCHECK were consistent with a 2.5-Å resolution structure. From Table 2, all the important elements of the structure are in agreement with the template proteins.

The bonding interactions between FMP and the three proteins are similar. In particular the formycin ring of FMP is situated in a hydrophobic environment enclosed by residues Y75, V76, I174, V209, W211, A178, L170, and Y123 (dianthin numbering). A more detailed bonding description of FMP in the active site of the three proteins is illustrated in Figure 2. In the dianthin structure, there are some extra hydrogen bonds detected: the side chain of Q208 is hydrogen bonded with the oxygen at position 2 (O2) of FMP and the side chain of R182 is bonded to both the side chain of E179 and the nitrogen at position 3 (N3) of the formycin ring. Additional polar interactions involve N7 and N1 of the substrate with the backbone NH of E121 and V76, respectively.

In the PAP structure, there are two disulfide bridges (residues 85–106 and 34–259, the C_α–C_α distances being 5.8 and 5.7 Å, respectively) but neither ricin nor dianthin has any cysteine in the same regions that can form a disulfide bridge. One might expect these disulfides to be conserved, as all these proteins share the same protein fold. We have measured the distances between the corresponding residues, for ricin (N88–F108 [10.0 Å], R39–S264 [8.9 Å]) and dianthin (R89–V110 [8.5 Å], Y36–S263 [4.4 Å]). The results obtained contrast with our expectations, being all 3 or 4 Å longer. The differences in the first disulfide bridge seem to be more significant because it is located between the end of α2 and the beginning of β5. This difference plus the fact that α2 belongs to a highly variable region among the three enzymes (see Table 2) may suggest the importance in ribosome binding and recognition. The second disulfide bridge is located close to the C-terminal region. In the case of ricin, it is involved in the interaction with chain B; thus it is reasonable that some variations among different types of RIP exist at this region.

The MEP distribution of each RIP was calculated for residues close to the catalytic site and then mapped onto the solvent surface. The MEP is a representation of how a molecule is perceived by a positive charge in its vicinity,[17] and the results help to explain the specificities of these proteins in binding different ribosomes (i.e., how the ribosome sees the entrance to the binding site). The three solvent surfaces colored by the MEP are illustrated in Color Plate 2. Residues 121–124 and 209–212 (dianthin numbering), which are located at the entrance to the adenine-binding site, show considerable sequence variations among the three enzymes.



	ricin	PAP	dianthin
d1	4.48	4.62	4.13
d2	3.80	3.85	3.78
d3	2.70	2.73	3.08
d4	3.13	3.26	5.04
d5	3.10	3.03	3.49
d6	5.16	5.68	6.59
d7	3.12	2.88	4.21
d8	3.23	3.49	3.67
d9	4.14	2.98	3.09
d10	4.02	3.02	4.30
d11	3.55	4.55	5.80
d12	6.49	8.40	3.04
d13	9.29	7.36	3.28

Figure 2. van der Waals contacts and polar interactions involving FMP in the binding site. Residues highlighted in bold, underlined, and in lightface type refer to ricin, PAP, and dianthin, respectively. Asterisk denotes centroid of a fragment.

In Color Plate 2, blue represents areas more negative in energy, and red is more positive in energy. The three enzymes show a similar, overall MEP topology as well as local differences. In dianthin the proximal presence of the two carboxylate moieties of E121 and D122 gives rise to a highly negative MEP value (the two large deep blue regions in Color Plate 2a), while the nearby Q124 (the green region between the blue regions in Color Plate 2a) has a less negative value on the solvent surface. The same spatial region in PAP (Color Plate 2b) is occupied by the side chain of R122 giving rise to a highly positive MEP value (the large red and orange regions), while in ricin (Color Plate 2c) this region represents the most negative one (the largest dark blue region) due to the presence of the carboxylate of D124.

Another significant difference is at the lower right-hand corner of the MEP. In dianthin (Color Plate 2a) this region is occupied by V209, which is electrically neutral (green), whereas it becomes progressively more negative at N209 of

Table 3. Selected fragments from the database for dianthin

Sequences[a]	Protein (pdb code)	rms (least squares, Å)	Fit (dist, Å)
Apstt (46–50) EVVGA (120–124)	Leghemoglobin (2lh3)	1.31	0.30
anv (85–87) SLN (95–97)	Immunoglobulin Fab (2fb4)	0.84	0.49
pevvva (107–112) LFSPFV (116–121)	Mengo E. virus coat protein (2mev)	1.09	0.77
kittgdg (131–137) ASKAGVK (127–133)	PAP (1pag)	1.13	0.59
nkkvr (157–161) SKPIX (420–424)	Hexochinase B (2yhx)	1.10	0.27

[a]Letters in lower and upper case denote residues from dian (without coordinates) and selected fragment (with coordinates) found in the database, respectively.

ricin (light blue; Color Plate 2c) and E206 of PAP (dark blue; Color Plate 2b). All these differences are consistent with the different interactions in ricin, PAP, and dianthin in the presence of the target RNA sequence resulting in different ribosome specificity.

More information could be obtained from the model. For example, knowledge of the isoelectric point (p*I*) of a protein is necessary for carrying out the separation process known as isoelectric focusing and provides clues to its cellular function. Experimental p*I* values for type 1 RIP, and for the A chain of type 2 RIPs are around p*I* 9–10 and 4–7,[18] respectively. We have calculated the solvent accessibility of the three structures and selected those basic and acidic residues that are at least 50% accessible. These selected residues were used in the p*I* calculations. The results obtained are in agreement with experiment, being p*I* 6.4, 9.9, and 9.6, for ricin, PAP, and dianthin, respectively. Thus, when the p*I* cannot be determined experimentally, a homology model constructed using the above-described procedure could provide a theoretical p*I* for further experimental work.

CONCLUSION

A homology computer model of dianthin 30 has been constructed using the X-ray structures of ricin A chain and PAP. The model obtained has a convincing structure and is in agreement with the two template proteins. An analysis of their adenine-binding site shows strong similarities in the binding of FMP across the three structures. Comparison of the folds suggests that the two disulfide bondings in PAP do not have a strong structural role.

Molecular electrostatic potential maps revealed similar topologies of the binding sites of all three proteins, but striking differences in the local disposition of the charged groups. This helps to account for the differences in substrate recognition and binding despite the overall similarities in the binding site.

The study shows the value of homology modeling and the use of MEP maps to help understanding experimental results and to provide insight to guide further experiments.

REFERENCES

1 Stirpe, F., Barbieri, L., Battelli, M.G., Soria, M., and Lappi, D.A. Ribosome-inactivating proteins from plants: Present status and future prospects. *Bio. Technol.* 1992, **10,** 405–412

2 Monzingo, A.F., Collins, E.J., Ernst, S.R., Irvin, J.D., and Robertus, J.D. The 2.5 Å structure of pokewood antiviral protein. *J. Mol. Biol.* 1993, **223,** 705–715

3 Montfort, W.M., Villafranca, J.E., Monzingo, A.F., Ernst, S.R., Katzin, B., Rutenber, E., Xuong, N.H., Hamlin, R., and Robertus, J.D. The three-dimensional structure of ricin at 2.8 Å. *J. Biol. Chem.* 1987, **262,** 5389–5403

4 Mlsna, D., Monzingo, A.F., Katzin, B., Ernst, S.R., and Robertus, J.D. The structure of recombinant ricin at 2.3 Å. *Protein Sci.* 1993, **2,** 429–435

5 Monzingo, A.F. and Robertus, J.D. X-ray analysis of substrate analog in the ricin A chain active site. *J. Biol. Chem.* 1992, **227,** 1136–1145

6 Katzin, B., Collins, E.J., Robertus, J.D. Structure of ricin A chain at 2.5 Å *Proteins Struct. Funct. Genet.* 1991, **10,** 251–259

7 Husain, J., Tickle, I.J., and Wood, S.P. Crystal structure of momordin, a type I ribosome inactivating protein from the seeds of *Momordica charantia*. *FEBS Lett.* 1994, **342,** 154–158

8 Zhou, K., Fu, Z., Chen, M., Lin, Y., and Pan, K. Structure of trichosanthin at 1.88 Å resolution. *Proteins* 1994, **19,** 4–13

9 Legname, G., Bellosta, P., Gromo, G., Modena, D., Keen, J.N., Roberts, L.M., and Lord, J.M. Nucleotide sequence of cDNA coding for dianthin 30, a ribosome inactivating protein from *Dianthus caryophyllus*. *Biochim. Biophys. Acta* 1991, **1090,** 119–122

10 Quanta 3.3 (Molecular Simulations, Inc., Burington, MA, 1992)

11 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M.J. CHARMM: A program for macromolecular energy, minimization

and dynamics calculations. *J. Comp. Chem.* 1983, **4,** 187–217

12 Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. A model of evolutionary change in proteins. In: *Atlas of Protein Sequence and Structure* (Dayhoff, M.O. Ed.), Vol. 5, Suppl. 3, Silver Spring, Washington, D.C.,1978, pp. 345–352

13 Orengo, C.A., Flores, T.P., Taylor, W.R., and Thornton, J.M. Identification and classification of protein fold families. *Protein Eng.* 1993, **6,** 485–500

14 Arnold, G.E. and Ornstein, R.L. An evaluation of implicit and explicit solvent model systems for the molecular dynamics simulation of bateriophage T4 lysozyme. *Proteins* 1994, **18,** 19–33

15 MacDNASIS version 3.0 (Hitachi Software Engineering Co., Ltd., Plymouth, MN, 1991)

16 Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 1993, **26,** 283–291

17 Politzer, P., Lawrence, P.R., and Iayasuriya, K. Molecular electrostatic potential: An effective tool for the elucidation of biochemical phenomena. *Environ. Health Persp.* 1985, **61,** 191–202

18 Barbieri, L. and Stirpe, F. Ribosome-inactivating proteins from plants: Properties and possible uses. *Cancer Surv.* 1982, **1**(3), 489–520