

Decamers observed in the crystals of bovine
pancreatic trypsin inhibitorJacek Lubkowski and Alexander
Wlodawer*Macromolecular Structure Laboratory, ABL-
Basic Research Program, Frederick Cancer
Research and Development Center, National
Cancer Institute, Frederick, MD 21702, USA

Correspondence e-mail: wlodawer@ncifcrf.gov

The structure of bovine pancreatic trypsin inhibitor (BPTI) has been solved at 2.1 Å resolution in a new crystal form (space group $P6_422$ with unit-cell dimensions $a = b = 95.0$, $c = 158.1$ Å). The asymmetric unit is a pentamer, but a decamer is created by application of crystallographic symmetry. The decamer of BPTI is only the fourth such assembly reported to date in the Protein Data Bank.

Received 2 June 1998

Accepted 18 August 1998

We dedicate this paper to Sir
Francis Crick, in recognition
of his amazing foresight.

PDB Reference: bovine
pancreatic trypsin inhibitor,
2hex.

1. Introduction

Crystal packing of macromolecules creates symmetric assemblies, but neither very low nor very high symmetry is common. For example, out of the 3258 entries reported to date in the Biological Macromolecule Crystallization Database (Gilliland *et al.*, 1994), only 86 (~2.6%) belong to space group $P1$ with pure translational symmetry. The same is true for the contents of the Protein Data Bank (PDB; Bernstein *et al.*, 1977), in which 163 out of 6178 structures belong to space group $P1$. Moreover, this fraction is likely to be overestimated after taking into account the non-crystallographic symmetry (NCS). Of the higher symmetries, fivefold symmetry is also very poorly represented. Therefore, it is quite unexpected to find one of the best known small proteins, bovine pancreatic trypsin inhibitor (BPTI), may also fall into this unique category.

BPTI is a small globular protein consisting of a single polypeptide chain containing 58 residues. Owing to its small size, stability and easy availability, BPTI has been used as a model system for many experimental and theoretical studies. Therefore, there is a large collection of information describing the structure of BPTI and its mutants (Huber *et al.*, 1970; Deisenhofer & Steigemann, 1975; Wlodawer *et al.*, 1987; Danishefsky *et al.*, 1993) and a wide variety of the other properties of this protein (Levitt, 1981; Creighton, 1985).

2. Materials and methods

2.1. Crystallization and X-ray data collection

Crystals of form IV of BPTI were grown from the preparation purchased from Sigma (catalog No. T-0256). The protein sample was dialyzed against HEPES buffer (pH 7.0) for 3 d and then concentrated to ~14 mg ml⁻¹. The

solution obtained was equilibrated against 2.0 M ammonium sulfate by sitting-drop diffusion (solution #32 in Crystal Screen I, Hampton Research). Crystals appeared after several months in two morphologically distinct forms: hexagonal bipyramids and hexagonal rods. Crystals of both types grew to average linear dimensions of ~0.1 mm. In both cases, the space group was $P6_422$ with unit-cell parameters $a = b = 95.0$, $c = 158.1$ Å. These crystals have a unit-cell volume similar to that of crystals previously characterized by Crick (1953), who postulated that this form contains five molecules in the asymmetric unit. However, with the molecular mass of BPTI being 6500 Da, any number of molecules within the range 4–8 could be accommodated in the asymmetric unit.

The X-ray data were collected using synchrotron radiation on beamline X9B at the National Synchrotron Light Source (NSLS), Brookhaven with $\lambda = 0.99$ Å. A single crystal with bipyramidal morphology was frozen at 100 K in a nitrogen-gas stream, using an Oxford Cryosystems low-temperature device. Prior to flash freezing, the crystal was transferred to mother liquid enriched with 10% (v/v) glycerol. X-ray data were recorded with an MAR Research 345 mm image-plate detector. Even though observable diffraction extended to ~1.9 Å resolution, data reduction with the *HKL* suite of programs (Otwinowski & Minor, 1997) resulted in acceptable statistics only for data to ~2.1 Å resolution (Table 1).

2.2. Structure solution and refinement

Structure solution by molecular replacement (MR) was complicated by the moderate quality of the X-ray data and by the small size of the protein. Before the initial MR trials, we searched the PDB for all available structures of BPTI in the free form or as complexes with

proteases. Consequently, we extracted all the independent BPTI coordinate sets, superimposed them and selected the common subset as the probe model for MR. This subset consisted of residues 3–55 with atomic positions and *B* factors as described in PDB entry 1BPI (Parkin *et al.*, 1996). We noted that the discrepancies between the coordinates extracted from different structures were small, with r.m.s. deviations for the backbone atoms within the range 0.2–0.4 Å. The only exceptions were the structures of some mutants (*e.g.* PDB entry 8PTI; Housset *et al.*, 1991) or those determined at lower resolution. Owing to the expected high similarity between the molecule of

BPTI in the hexagonal crystals and the probe, the molecular-replacement searches were performed against 3.0 Å X-ray data, a relatively high resolution for MR. All MR calculations were performed with the program *AMoRe* (Navaza, 1994). Careful analysis of the results allowed us to unambiguously identify the contents of the asymmetric unit as five molecules of BPTI. The correct solution corresponded to a crystallographic *R* factor of 0.365 and a correlation coefficient of 0.579, after rigid-body refinement as implemented in *AMoRe*. The next listed peak was described by an *R* factor of 0.432 and a correlation coefficient of 0.412. After extracting 10% of randomly

selected data for free *R*-factor calculations (Brünger, 1992*a*), refinement was performed using *X-PLOR* (Brünger, 1992*b*). Initially, the strict NCS constraints were used, which were subsequently substituted by the NCS restraints. NCS was not used in the final stages of refinement, since the side chains of several arginine residues had very well defined, but different, conformations in each monomer. Finally, all solvent molecules that could be located in the electron density were included in the refinement, as well as 7.5 sulfate anions (one structurally conserved for each monomer and 2.5 between monomers on special positions). The final cycles of refinement were performed with *SHELX97*

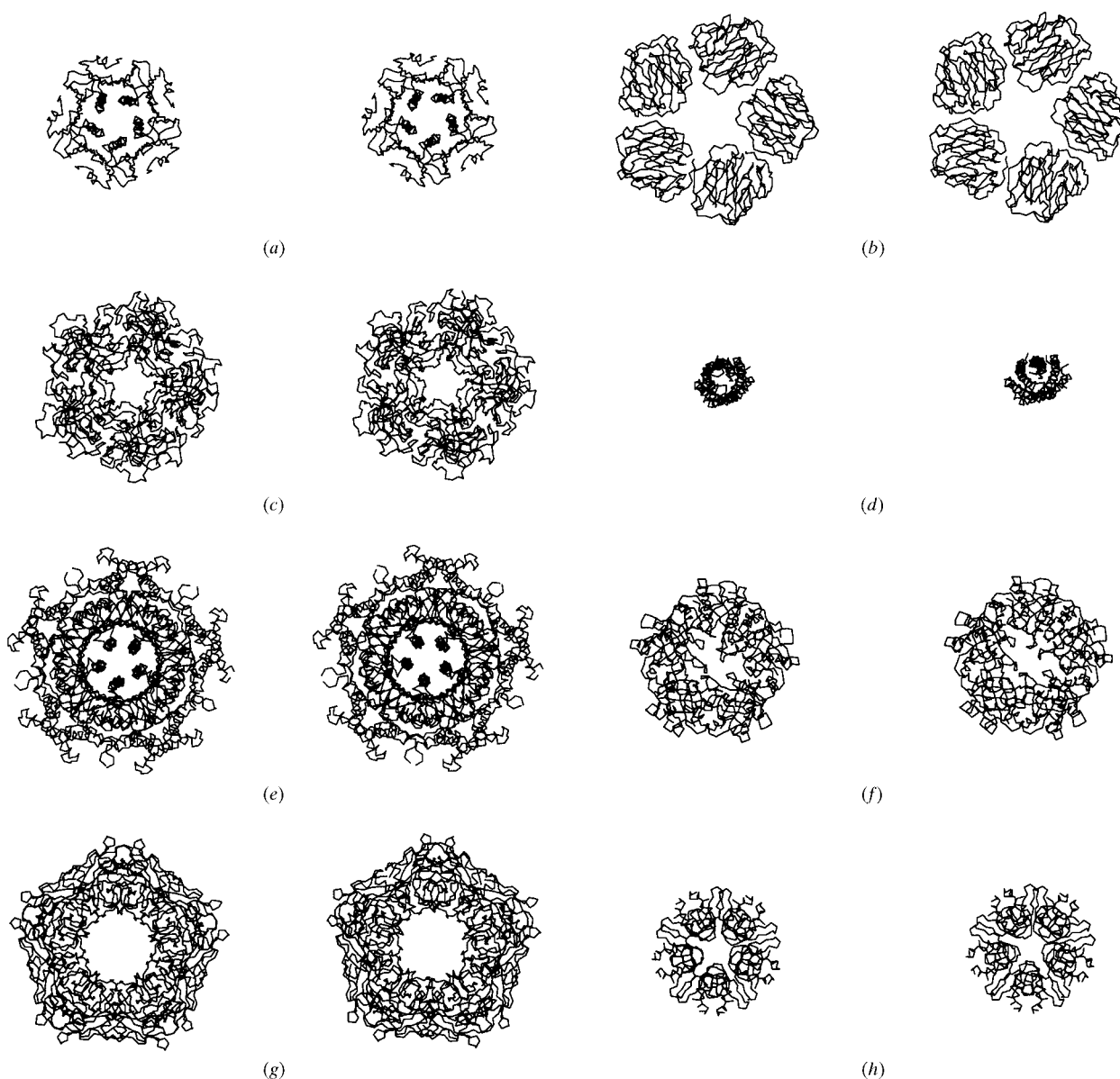


Figure 1

Stereoscopic drawing of the α traces for four representative previously determined pentamers and two decamers, compared with the currently obtained structure of BPTI. (a) Cholera toxin, (b) C-reactive protein, (c) mouse polyomavirus VP1 coat protein, (d) cartilage oligomeric matrix protein, (e) GTP cyclohydrolase I, (f) muconolactone isomerase, (g) cyclophilin A, (h) BPTI (this work).

Table 1
Data-collection and refinement statistics.

Data collection	
Space group	$P6_422$
Cell dimensions (Å)	$a = b = 95.0, c = 158.1$
Resolution range (Å)	20–2.1
Number of collected reflections	333309
Number of unique reflections	25235
R_{sym} (%)†	7.40
Overall completeness (%)	99.6
Completeness in the last shell (2.17–2.10 Å)	99.0
$\langle I/\sigma_I \rangle$	23.0
$\langle I/\sigma_I \rangle$ in the last shell	4.3
Refinement	
Resolution range (Å)	10–2.1
Number of reflections (working set)	21757 (19087 > 4 σ)
R factor (no cutoff)†	0.205
Free R (no cutoff)	0.288
R factor (4 σ cutoff)†	0.192
Free R (4 σ cutoff)	0.270
Number of protein atoms	2229
Number of solvent O atoms	181
Number of sulfate anions	7.5
R.m.s. deviations from ideal values (Å)	
Bond distances	0.006
Angle distances	0.021
Planes	0.025
Average B_{iso} (Å ²)	
Protein	27.6
Waters	41.2
Sulfates	48.8

† $R_{\text{sym}} = \sum |I_i - \langle I \rangle| / \sum I_i$, where I_i is the intensity of an individual measurement of the reflection i and $\langle I \rangle$ is the mean intensity of this reflection. † R factor = $\sum |F_o - F_c| / \sum |F_o|$, where F_o and F_c denote observed and calculated structure amplitudes, respectively, and summation extends over all observed reflections.

(Sheldrick & Schneider, 1997). The refinement statistics are summarized in Table 1. The coordinates and the structure factors have been deposited with the PDB.

3. Results and discussion

Although a pentamer of BPTI can be identified in the asymmetric unit, we find that pairs of pentamers in the crystal arrange into sandwich-like decameric assemblies (Fig. 1*h*). No biological relevance for BPTI decamers is known or postulated; therefore, they must be considered the result of interactions in the crystal. However, it is interesting to compare some aspects of the pentameric symmetry observed for BPTI with those reported for other proteins. A convenient tool for this purpose is the PQS server (<http://pqs.ebi.ac.uk>) described by Henrick & Thornton (1999). Using this server, we identified 16 entries in the PDB which form detectable pentameric assem-

blies and three which form decamers. The complete list of these entries consists of 15 homopentamers (1BOV, 1CHB, 1CHP, 1CHQ, 1CT1, 1FGB, 1GNH, 1LGN, 1LT5, 1LT6, 1SAC, 1VDF, 1VPN, 1VPS and 2CHB) and one heteropentamer (1HLT). A brief analysis shows that all of these entries represent four different protein folds. The most extensively represented are the toxins, including Verotoxin-1 from *Escherichia coli* (1BOV), cholera toxin (1CHB, 1CHP, 1CHQ, 1CT1, 1FGB, 2CHB) and heat-labile enterotoxin (1LT5, 1LT6). A pentamer representative of all these toxins is shown in Fig. 1(*a*). Two highly homologous proteins, serum amyloid P component (1LGN, 1SAC) and C-reactive protein (1GNH), form a different fold, as shown in Fig. 1(*b*). Mouse polyomavirus VP1 coat protein (1VPN, 1VPS) and the cartilage oligomeric matrix protein (1VDF) represent two different folds, shown in Figs. 1(*c*) and 1(*d*), respectively. In the decameric assemblies, a common feature is the presence of two disk-shaped pentamers arranged into a sandwich-like form. The three entries describing the decamers are GTP cyclohydrolase I (1GTP), muconolactone isomerase (1MLI), and cyclophilin A (2RMB), shown in Figs. 1(*e*), 1(*f*) and 1(*g*), respectively.

The decamer present in the crystal of BPTI is also the result of sandwich-like stacking of two pentamers. Using PQS, we found that the decrease in the accessible surface area (ASA) per monomer upon the formation of the decamer was 1556.3 Å² and the solvation energy (SE) gain due to decamerization was $-46 \text{ kcal mol}^{-1}$. Comparing this structure to other pentamers and decamers, we found that the gain in both ASA and SE calculated for BPTI was within the observed ranges (1112.5–4991.5 Å² for ASA and -5.7 to $-161.0 \text{ kcal mol}^{-1}$ for SA). However, with its 58 residues per chain, BPTI is the smallest globular protein reported to form pentamers. As seen in Fig. 1, there are no common topological features

evident within the monomers forming the different pentamers which could be determinants for the stability of the fivefold rotational symmetry. However, additional calculations may be necessary to settle this point, and thus the structure of BPTI reported here might provide a good subject for further analysis and theoretical considerations.

We thank Drs Janet M. Thornton and Kim Henrick for their help with using the PQS server and Anne Arthur for editorial assistance. Our research was sponsored by the National Cancer Institute, DHHS, under contract with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.
- Brünger, A. T. (1992*b*). *X-PLOR, a System for X-ray Crystallography and NMR*. New Haven: Yale University Press.
- Creighton, T. E. (1985). *J. Phys. Chem.* **89**, 2452–2459.
- Crick, F. H. (1953). *Acta Cryst.* **6**, 221–222.
- Danishefsky, A. T., Housset, D., Kim, K.-S., Tao, F., Fuchs, J., Woodward, C. & Wlodawer, A. (1993). *Protein Sci.* **2**, 577–587.
- Deisenhofer, J. & Steigemann, W. (1975). *Acta Cryst.* **B31**, 238–250.
- Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. (1994). *Acta Cryst.* **D50**, 408–413.
- Henrick, K. & Thornton, J. (1999). *FEBS Lett.* In the press.
- Housset, D., Kim, K.-S., Fuchs, J., Woodward, C. & Wlodawer, A. (1991). *J. Mol. Biol.* **220**, 757–770.
- Huber, R., Kukla, D., Ruhlmann, A., Epp, O. & Formanek, H. (1970). *Naturwissenschaften*, **57**, 389–392.
- Levitt, M. (1981). *Nature (London)*, **294**, 379–380.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Parkin, S., Rupp, B. & Hope, H. (1996). *Acta Cryst.* **D52**, 18–29.
- Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
- Wlodawer, A., Deisenhofer, J. & Huber, R. (1987). *J. Mol. Biol.* **193**, 145–156.