Biochemical Society Transactions www.biochemsoctrans.org

Crystallography and protein–protein interactions: biological interfaces and crystal contacts

Bostjan Kobe*†[1], Gregor Guncar*‡, Rebecca Buchholz*, Thomas Huber*§, Bohumil Maco*, Nathan Cowieson†‖, Jennifer L. Martin†, Mary Marfori* and Jade K. Forwood*¶

*School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, QLD 4072, Australia, †Institute for Molecular Bioscience and Special Research Centre for Functional and Applied Genomics, The University of Queensland, Brisbane, QLD 4072, Australia, ‡Department of Biochemistry, Molecular and Structural Biology, Jozef Stefan Institute, Ljubljana, Slovenia, §Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, QLD 4072, Australia, ‖Monash Centre for Synchrotron Science, Monash University, Clayton, VIC 3800, Australia, and ¶School of Biomedical Sciences, Charles Sturt University, Wagga Wagga, NSW 2650, Australia

Abstract

Crystallography is commonly used for studying the structures of protein–protein complexes. However, a crystal structure does not define a unique protein–protein interface, and distinguishing a 'biological interface' from 'crystal contacts' is often not straightforward. A number of computational approaches exist for distinguishing them, but their error rate is high, emphasizing the need to obtain further data on the biological interface using complementary structural and functional approaches. In addition to reviewing the computational and experimental approaches for addressing this problem, we highlight two relevant examples. The first example from our laboratory involves the structure of acyl-CoA thioesterase 7, where each domain of this two-domain protein was crystallized separately, but both yielded a non-functional assembly. The structure of the full-length protein was uncovered using a combination of complementary approaches including chemical cross-linking, analytical ultracentrifugation and mutagenesis. The second example involves the platelet glycoprotein Ibα–thrombin complex. Two groups reported the crystal structures of this complex, but all the interacting interfaces differed between the two structures. Our computational analysis did not fully resolve the reasons for the discrepancies, but provided interesting insights into the system. This review highlights the need to complement crystallographic studies with complementary experimental and computational approaches.

Biological interfaces and crystal contacts

MX (macromolecular X-ray crystallography) is the most popular method to obtain high-resolution structural information on protein–protein complexes, and it yields the most detailed structural information about the interaction. However, there is a major issue when using MX to study protein–protein interactions: a crystal structure does not define a unique protein–protein interface. A crystal is a regular 3D (three-dimensional) array of molecules, and a number of protein–protein contacts are present that do not occur in solution. If we define the biologically relevant protein–protein interfaces as the 'biological interface', this interface may or may not be present in the crystal; on the other hand, several other 'crystal-packing' interfaces (also called 'crystal contacts') are present, which are responsible for the packing of the protein molecules in the crystal lattice (Figure 1). Assuming that, under the crystallization conditions used, the expected protein–protein complex is formed (and therefore that one of the protein–protein interfaces corresponds to the biological interface), the task is to identify this biological interface among the crystal contacts.

Distinguishing biological interfaces from crystal contacts is a challenging problem. There is a fundamental difficulty in differentiating biological and crystal contacts, because the physical basis of their formation is the same. A biological interface typically co-exists with 6–12 different crystal-packing interfaces [1,2]. The situation is complicated further by the fact that, for technical reasons, truncated proteins and incomplete complexes lacking certain binding partners are often used for crystallization. The PDB (Protein Data Bank) contains information on the biological unit (BIOMT records), but these records are inaccurate (error rate estimated at 9% [3]). A number of computational methods have been developed for distinguishing biological and crystal contacts based on the interface area and the fraction of surface involved, numbers and properties of atomic pair contacts, residue propensities, evolutionary conservation and estimates of free energies [4–10]. For example, crystal-packing interfaces are generally smaller, the chemical composition is slightly different, and the packing of atoms is not as tight as for biological interfaces; however, the distributions of these properties overlap between biological and crystal-packing interfaces. A recent approach based on a machine learning classifier, which considered a number of interface properties,

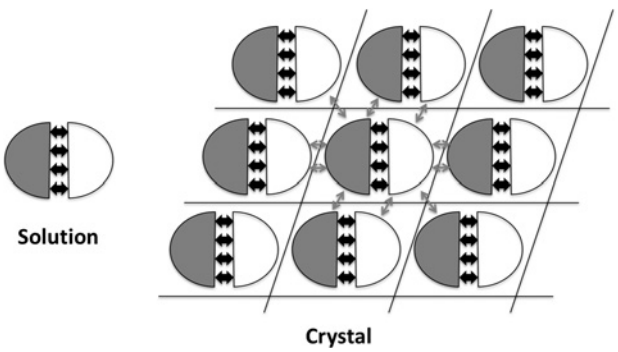
Key words: acyl-CoA thioesterase 7 (Acot7), platelet glycoprotein Ibα, protein–protein interaction, structural biology, thrombin, X-ray crystallography.

Abbreviations used: 3D, three-dimensional; Acot7, acyl-CoA thioesterase 7; AUC, analytical ultracentrifugation; EM, electron microscopy; GpIbα, glycoprotein receptor subunit Ibα; MX, macromolecular X-ray crystallography; SEC, size-exclusion chromatography.

[1]To whom correspondence should be addressed (email b.kobe@uq.edu.au).

Figure 1 | Schematic diagram of a crystal of a protein–protein complex, highlighting the biological interface and the crystal contacts

Two interacting proteins, depicted as half-ovals in white and grey, interact in solution (left on the Figure) through the flat interface highlighted with black block-arrows. In the crystal schematically depicted as a 2D array on the right of the Figure, each protein forms a number of other interfaces with neighbouring molecules in the crystals, as highlighted by grey thin arrows.



was reported to have reduced the error rates of 16–24% for the earlier methods to 11% [3]. The error rate therefore remains high, emphasizing the need to obtain further information on the biological interface from complementary structural and functional approaches.

To complicate matters even further, protein crystals can often only be grown under conditions quite different from the physiological environment, in solutions with high protein concentrations, non-physiological pH, high ionic strength, low temperatures, or in the presence of compounds that aid crystallization. The crystallization conditions may therefore promote protein–protein contacts that are not biologically relevant, or the biological interface may not be present at all.

The only way to test these possibilities is to combine crystallographic studies with complementary structural and functional methods. We briefly review complementary structural approaches that can be used for this purpose, and highlight two interesting examples of biological interface ambiguity, one from our recent work and one interesting unresolved example of different protein complex structures solved by different groups, before concluding in the 'Outlook' section.

Combining structural methods

The growth in combining other techniques with the traditional approaches of MX, NMR and EM (electron microscopy) shows both the benefits and the necessity of considering complementary approaches. There are a number of reasons for using a combination of approaches, including technical and feasibility issues (e.g. the macromolecule is too big for NMR, too small for EM or does not crystallize), and the need for acquiring additional data not accessible by the traditional techniques (e.g. data on the dynamic behaviour of the system). The most common reason, however, is to facilitate interpretation and resolve ambiguities, including MX-specific ambiguities about interaction stoichiometries and interfaces discussed in this review.

Different techniques access different types of information, some such that we may not be able to deduce 3D positions of atoms or molecules directly, but we may infer spatial information indirectly. Just to list a few examples: SAXS (small angle X-ray scattering), neutron and light scattering, AFM (atomic force microscopy), MS and AUC (analytical ultracentrifugation) yield information on the shape, size and mass of macromolecules; chemical cross-linking, FRET (fluorescence resonance energy transfer) and EPR yield data on proximities of different parts of macromolecules; CD informs about the secondary structural content of a protein; and various labelling techniques inform about the composition and stoichiometry of macromolecules [11]. A number of additional techniques exist to yield data on interaction energies and dynamic behaviour, as well as the functional aspects of the biological system.

The most effective process for integrating hybrid data from diverse sources takes advantage of computational modelling, and can be streamlined by linking the steps of data collection, data conversion into spatial restraints, generation of structural models that meet these restraints, and assessment of the accuracy and precision of the resulting structures [12]. When computer-generated structures that satisfy various restraints cluster together, the data are adequate to define a unique state of the macromolecule. Calculated structures can be assessed for self-consistency in terms of satisfying the restraints and the variability of the generated structures, by cross-validating through omitting portions of the data, and by evaluating the model in the light of other data not included in the structure calculation.

Acot7 (acyl-CoA thioesterase 7)

An interesting example of ambiguity resulting from crystallography, from our own recent work, involves the enzyme Acot7 [13]. Acots catalyse the hydrolysis of fatty acyl-CoA to free fatty acid and CoA, and thereby regulate lipid metabolism and cellular signalling [14]. While prokaryotic homologues possess a single thioesterase (hot-dog) domain, mammalian Acot7 contains a pair of domains in tandem. Because the intact two-domain enzyme could not be crystallized, we pursued a 'divide-and-conquer' approach popular with structural biologists and crystallized each (N- and C-) domain separately. Both structures showed a hexameric arrangement of domains (Figure 2). However, the individual domains showed no catalytic activity and needed to be combined to restore function. The individual domain structures shed little light on the structural basis of this observation, and suggested that these structures may not represent functionally relevant states. Indeed, using a combination of (i) functional information, based on mutagenesis of the probable active-site residues, and (ii) structural information, obtained using SEC (size-exclusion chromatography), AUC and the chemical

Figure 2 | **Structure determination of Acot7**

Acot7 is a two-domain protein that trimerizes in solution. The full-length protein could not be crystallized, but both domains were solved independently using MX, each revealing a hotdog domain in a hexameric arrangement. However, neither domain has enzymatic activity on its own. We used both N- and C-domain structures (individual domains shown in different colours in surface representation with active site residues from the N- and C-domains coloured red and blue respectively) plus AUC, SEC, cross-linking (X-link)/MS (cross-links indicated as black lines in the ribbon diagram) and mutagenesis data to generate the model of full-length Acot7, showing that catalytic residues from both domains (red and blue) are required to generate the three active sites in the trimer. The Figure was prepared using PyMol (DeLano Scientific; http://pymol.sourceforge.net/).

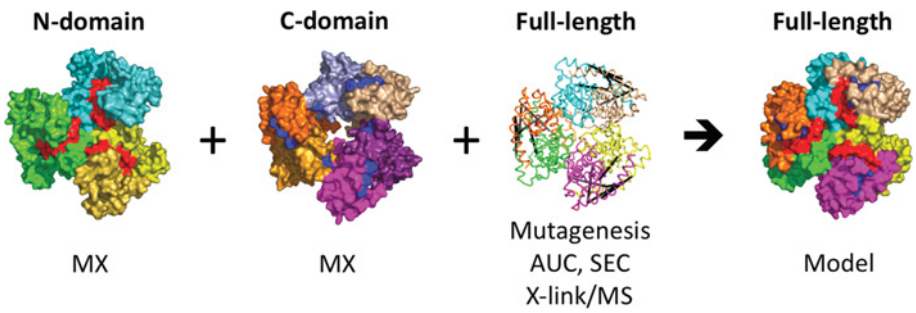
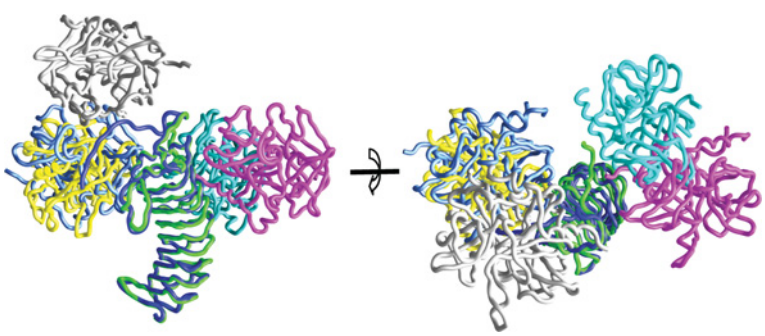


Figure 3 | **Comparison of glycoprotein Ibα (GpIbα)–thrombin complexes in the structures 1OOK and 1P8V**

GpIbα residues 1–265 were used in the superposition. The molecules present in the asymmetric unit of the respective crystal structures, as defined in the deposited coordinate files, are termed GpIbα 1 and thrombin 1. 1OOK: GpIbα 1, blue; thrombin 1, light blue; thrombin 2, grey; thrombin 3, magenta. 1P8V: GpIbα 1, green; thrombin 1, cyan; thrombin 2, yellow. The molecules are shown in ribbon representation. The two views are related by a 90° rotation around the *x*-axis. The Figure was produced using Grasp [24].



cross-linking/MS/molecular modelling hybrid approach [15], explained that the active sites in the full-length enzyme are generated through an N-domain/C-domain interaction that replaces the N-/N- and C-/C-domain interactions observed in the crystals of individual domains (Figure 2).

Glycoprotein Ibα–thrombin interaction

The second example we highlight here involves the structure of the complex between platelet GpIbα (glycoprotein receptor subunit Ibα) and the protease thrombin. This receptor regulates the adhesion of blood platelets to damaged blood vessel walls and the subsequent platelet aggregation. The GpIbα subunit binds thrombin, a serine protease with both pro-coagulant and anti-coagulant activities [16]. Two groups reported the crystal structures of the complex between thrombin and the N-terminal extracellular domain [LRR (leucine-rich repeat) domain] of GpIbα (PDB ID 1OOK [17] and 1P8V [18]), but the interaction interfaces in the two structures were completely different [19–22] (Figure 3).

We performed a detailed comparison of the two structures to look for clues that may explain the differences [23]. The analysis included the examination of the properties of different interfaces in the crystals, the examination of the functional implications of the different interactions occurring in the crystal, the examination of the interactions with a C-terminal-sulfated region previously implicated in the GpIbα–thrombin interaction, comparison of the recombinant proteins used in the two studies, comparison of the crystallization conditions used for producing the crystals in the two studies, and the examination of the electrostatic properties of the two crystal structures. The analysis

leads to several conclusions: (i) the 1:1 complex observed in solution is likely to be the only long-lived interaction; (ii) the anionic GpIbα sequence is likely to be responsible for the initial interaction between the two proteins; (iii) the interaction with the rest of GpIbα occur subsequently and may alternate between different binding modes; (iv) the interaction is pH-dependent and pH may regulate the conversion between different binding modes; (v) the primary interface observed in the 1P8V crystals is most likely to be the strongest interaction based on integrating data from mutational, functional and computational studies. While our analysis could not unambiguously point to one interface or structure to be the 'correct' one, it suggests mutagenesis experiments that could shed further light on the interaction, and suggests that the interaction plasticity uncovered by crystal structures may be biologically significant. This is a very complex biological system where the thrombin–GpIbα interaction could lead to opposing functional consequences (pro-thrombotic or anti-thrombotic effects), based on the site of interaction and the presence of other macromolecules. The interactions and the consequent effects will clearly depend on concentrations of various components.

Outlook

One important shortcoming of MX, when used for studying interactions between macromolecules, is that it does not define a unique protein–protein interface.

Complementary methods can help interpret MX results and should therefore be used in combination with MX in structural studies. In the future, it would be of interest to systematically analyse all crystal structures in the PDB to find structures of similar molecules and complexes that differ from each other. Computational approaches can be used for highlighting the interesting cases that may require further experimental studies to identify the biological interfaces.

B.K. is an Australian Research Council Federation Fellow and an NHMRC (National Health and Medical Research Council) Honorary Research Fellow. J.L.M. is an NHMRC Research Fellow.

References

1 Janin, J. and Rodier, F. (1995) Protein–protein interaction at crystal contacts. Proteins **23**, 580–587
2 Dasgupta, S., Iyer, G.H., Bryant, S.H., Lawrence, C.E. and Bell, J.A. (1997) Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. Proteins **28**, 494–514
3 Bordner, A.J. and Gorin, A.A. (2008) Comprehensive inventory of protein complexes in the Protein Data Bank from consistent classification of interfaces. BMC Bioinformatics **9**, 234
4 Ponstingl, H., Kabir, T. and Thornton, J.M. (2003) Automatic inference of protein quaternary structure from crystals. J. Appl. Crystallogr. **36**, 1116–1122
5 Ponstingl, H., Henrick, K. and Thornton, J.M. (2000) Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins **41**, 47–57
6 Valdar, W.S. and Thornton, J.M. (2001) Conservation helps to identify biologically relevant crystal contacts. J. Mol. Biol. **313**, 399–416
7 Bahadur, R.P., Chakrabarti, P., Rodier, F. and Janin, J. (2004) A dissection of specific and non-specific protein–protein interfaces. J. Mol. Biol. **336**, 943–955
8 Henrick, K. and Thornton, J.M. (1998) PQS: a protein quaternary structure file server. Trends Biochem. Sci. **23**, 358–361
9 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. **372**, 774–797
10 Janin, J., Rodier, F., Chakrabarti, P. and Bahadur, R.P. (2007) Macromolecular recognition in the Protein Data Bank. Acta Crystallogr. Sect. D Biol. Crystallogr. **63**, 1–8
11 Cowieson, N.P., Kobe, B. and Martin, J.L. (2008) United we stand: combining structural methods. Curr. Opin. Struct. Biol. **18**, 617–622
12 Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T. et al. (2007) Determining the architectures of macromolecular assemblies. Nature **450**, 683–694
13 Forwood, J.K., Thakur, A.S., Guncar, G., Marfori, M., Mouradov, D., Meng, W., Robinson, J., Huber, T., Kellie, S., Martin, J.L. et al. (2007) Structural basis for recruitment of tandem hotdog domains in acyl-CoA thioesterase 7 and its role in inflammation. Proc. Natl. Acad. Sci. U.S.A. **104**, 10382–10387
14 Hunt, M.C. and Alexson, S.E. (2002) The role acyl-CoA thioesterases play in mediating intracellular lipid metabolism. Prog. Lipid Res. **41**, 99–130
15 Mouradov, D., Craven, A., Forwood, J.K., Flanagan, J.U., Garcia-Castellanos, R., Gomis-Ruth, F.X., Hume, D.A., Martin, J.L., Kobe, B. and Huber, T. (2006) Modelling the structure of latexin–carboxypeptidase A complex based on chemical cross-linking and molecular docking. Protein Eng. Des. Sel. **19**, 9–16
16 Lundblad, R.L. and White, II, G.C. (2005) The interaction of thrombin with blood platelets. Platelets **16**, 373–385
17 Celikel, R., McClintock, R.A., Roberts, J.R., Mendolicchio, G.L., Ware, J., Varughese, K.I. and Ruggeri, Z.M. (2003) Modulation of α-thrombin function by distinct interactions with platelet glycoprotein Ibα. Science **301**, 218–221
18 Dumas, J.J., Kumar, R., Seehra, J., Somers, W.S. and Mosyak, L. (2003) Crystal structure of the GpIbα–thrombin complex essential for platelet aggregation. Science **301**, 222–226
19 Sadler, J.E. (2003) Structural biology: a ménage à trois in two configurations. Science **301**, 177–179
20 Vanhoorelbeke, K., Ulrichts, H., Romijn, R.A., Huizinga, E.G. and Deckmyn, H. (2004) The GPIbα–thrombin interaction: far from crystal clear. Trends Mol. Med. **10**, 33–39
21 Adams, T.E. and Huntington, J.A. (2006) Thrombin–cofactor interactions: structural insights into regulatory mechanisms. Arterioscler. Thromb. Vasc. Biol. **26**, 1738–1745
22 Adam, F., Bouton, M.C., Huisse, M.G. and Jandrot-Perrus, M. (2003) Thrombin interaction with platelet membrane glycoprotein Ibα. Trends Mol. Med. **9**, 461–464
23 Kobe, B., Guncar, G., Buchholz, R., Huber, T. and Maco, B. (2008) The many faces of platelet glycoprotein Ibα–thrombin interaction. Curr. Protein Pept. Sci., in the press
24 Nicholls, A. and Honig, B. (1991) A rapid finite difference algorithm, using successive over-relaxation to solve the Poisson–Boltzmann equation. J. Comput. Chem. **12**, 435–445

Received 11 August 2008
doi:10.1042/BST0361438