Elsevier Editorial System(tm) for Current Opinion in Structural Biology
Manuscript Draft

Manuscript Number:

Title: United we stand: combining structural methods

Article Type: Biophysical methods

Corresponding Author: Dr Nathan Cowieson,

Corresponding Author's Institution:

First Author: Nathan Cowieson

Order of Authors: Nathan Cowieson; Bostjan Kobe, PhD; Jennifer L Martin, PhD

United We Stand:

Combining Structural Methods

Nathan P. Cowieson[1]*, Bostjan Kobe[2, 3] and Jennifer L. Martin[3, 2]*

[1]Monash Centre for Synchrotron Science, Monash University, Clayton, Victoria 3800 Australia

[2]School of Molecular and Microbial Sciences and [3]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072 Australia

*Corresponding authors:

NPC: Monash Centre for Synchrotron Science,

Monash University, Clayton, Victoria 3800 Australia

Phone +61 (3) 9902 0117 Fax: +61 (3) 9905 3637

JLM: Institute for Molecular Bioscience,

University of Queensland, Brisbane, Queensland 4072 Australia

Phone: +61 7 3346 2016 Fax: +61 7 3346 2101

Email addresses:

NPC Nathan.Cowieson@sync.monash.edu.au;

BK b.kobe@uq.edu.au;

JLM j.martin@imb.uq.edu.au

Short title: Combining structural methods

SUMMARY

High-resolution techniques are the mainstay of structural biologists; however, to address challenging biological systems many are now turning to hybrid approaches that use complementary structural data. In this review we outline the types of structural problems that benefit from combining results of many methods, we summarise the types of data that can be generated by complementary approaches, and we highlight the application of combined methods in structural biology with recent structural studies of membrane proteins, mega-complexes and inherently flexible proteins.

INTRODUCTION

Structural biologists benefit enormously by combining structural approaches to tackle biological systems. This is evident in the increasing use of complementary methods combined with the traditional structural biology techniques of macromolecular X-ray crystallography (MX), nuclear magnetic resonance (NMR) and electron microscopy (EM) to generate structural information. New approaches include mass spectrometry of intact complexes [1], synchrotron radiation circular dichroism spectroscopy [2], electron paramagnetic resonance spectroscopy (EPR) combined with site-directed spin labelling [3], and a combination of cross-linking, mass spectrometry and computational docking with sparse distance restraints [4,5]. The most effective process to integrate data from diverse sources takes advantage of computational modelling, and can be summarized as follows: (i) data collection; (ii) conversion of data into spatial restraints; (iii) generation of structural models that meet the restraints; and (iv) assessment of the accuracy and precision of the resulting structures. Spatial features that can be restrained include positions, contacts, proximities, shapes and symmetries of individual atoms, domains, macromolecules or (sub)assemblies.

Many biochemical, biophysical and proteomic techniques can generate useful structural information (Table 1, Figure 1). For recent reviews of methodology, sample requirements and interpretation of data, see [1,6-10]. Common reasons for combining methods include (1) technical limitations of

individual methods – *eg* the molecule doesn't crystallize; (2) mega-complexity – *eg* multi-protein complexes and (3) flexibility - *eg* inherently disordered proteins. These three themes clearly overlap, but represent convenient categories for us to highlight studies from the past two years that used multiple approaches with spectacular success.

1. TECHNICAL LIMITATIONS

Combining multiple methods is particularly valuable when sample requirements for high-resolution structural biology techniques cannot be met or when only low resolution data can be obtained. Non-traditional methods can also enable interpretation of the results of traditional structural methods and help advance these through specific bottlenecks. Membrane proteins are a typical example because they are difficult to produce and crystallize. A recent review highlighted how biochemical and computational analyses coupled with low resolution maps from cryo-EM can allow a detailed mechanistic understanding of membrane protein structure and function in the absence of crystallographic data [11]. Moreover, a series of recent papers describing crystal structures of the β2-adrenergic receptor [12-14] represents the culmination of a combined methods *tour de force*. Thorough biophysical characterisation of the protein using fluorescence resonance energy transfer (FRET) [15], cross-linking, chemical reactivity studies and pharmacological evaluation of the effect of ligand binding [16] identified an unstructured C-terminus and a protease sensitive loop, both hypothesised to inhibit crystallisation. Removal of the C-terminus and stabilisation of the flexible loop, either through binding a monoclonal antibody or by replacement of the loop with engineered lysozyme, allowed crystallisation and structure determination at 3.4-3.7 and 2.4 Å resolution, respectively [12-14].

Proteins that interact with actin are notoriously difficult to study because they are often large, flexible and multi-domain. One such example is talin, a ~2500 residue protein that links members of the integrin family of cell adhesion molecules to filamentous action (F-actin); Gingras *et al* [17] tackled this protein using hybrid methods. Secondary structure prediction and NMR of multiple constructs enabled structure determination of the C-terminal actin-binding domain; the adjacent

dimerisation helix was studied by MX and mutagenesis confirmed that dimerisation was required for F-actin binding. The NMR and crystal structures were docked into a small angle X-ray scattering (SAXS) envelope of the polypeptide comprising both domains, showing that the full-length talin dimer likely adopts a wide range of conformations. Finally, differential scanning calorimetry (DSC) and actin co-sedimentation assays indicated that the two-domain polypeptide binds F-actin, and EM of the complex showed that the interaction involves three actin monomers along the long pitch helix of the F-actin filament [17]. Hybrid methods were also employed to study the multi-domain structure of two other actin-binding proteins, cortactin and gelsolin. MX and SAXS were used to demonstrate how the six domains of gelsolin convert from a compact to an extended form in the presence of calcium [18] and our labs used bioinformatics, SAXS and cross-linking with mass spectrometry to show that cortactin adopts a globular rather than an extended structure in solution [19] (Figure 2).

Combining methods may also be necessary when traditional approaches give ambiguous results, as was the case for our work on acyl-CoA thioesterase 7 [20]. The intact two-domain enzyme could not be crystallized, but the individual structures of each domain could not explain the catalytic activity. Mutagenesis, analytical ultracentrifugation (AUC), cross-linking with mass spectrometry and molecular modelling were used to determine the full-length structure revealing how the active sites are generated (Figure 2) [20].

Another application of combining methods is to use one method to help advance another through a bottleneck. A good example is the recent *de novo* structure prediction of a protein by the Rosetta program, using cpu time donated from 70,000 home computers [21]. The model generated was so accurate it was able to phase crystallographic data of the same protein by molecular replacement [21] suggesting that this approach could be used more broadly for phasing crystal structures. Other examples used SAXS and EM information to phase crystallographic data [22] or SAXS data to resolve discrepancies between MX and EM structures [23].

2. MEGA-COMPLEXES

Most proteins in the cell are thought to function, at least transiently, as part of complexes or as functional modules [24]. Understanding these biological systems – protein:protein complexes, mega-complexes, or even the entire cell – requires spanning several orders of magnitude in spatial and temporal dimensions [25]. Although MX can be used to tackle large assemblies (strategies reviewed in [26]), combining approaches with EM is the dominant means of studying such complexes [27] because the interactions between components are often weak and transient and the complexes very large, heterogeneous or only available in limited amounts. The classic example of combining approaches to study large complexes is the elucidation of the ribosome structure (reviewed recently in [28]).

Arguably the most spectacular recent application of the integration of diverse data to generate structural information has been the determination of the architecture of the nuclear pore complex (NPC) [29]. One of the largest macromolecular assemblies in eukaryotic cells, the NPC comprises no less than 456 proteins (resulting from multiple copies of over 30 different proteins). Major challenges were the large size and high degree of flexibility of the NPC. Alber *et al* [30] used an iterative four step approach involving (i) experimental data generation using AUC, quantitative immunoblotting, affinity purification, overlay assays, EM, immuno-EM, membrane fractionation and bioinformatics; (ii) translation of data into spatial restraints; (iii) structure calculation by satisfying these restraints; and (iv) analysis of the calculated ensemble of structures to yield a final structure. The resulting structure provided insights into the evolutionary origins of NPC assembly and the mechanism of cargo transport through the pore [29]. The Integrative Modelling Platform software developed for the NPC project facilitates the integration of diverse types of structural data and has the potential to assist in many other applications [30].

Another challenging system for structural biologists is the proteasome. Sharon *et al* [31] recently characterized one of the two major sub-complexes of the 19S regulatory particle of the proteosome, the peripheral lid, using a combination of (i) tandem mass spectrometry of the intact nine-component complex and (ii) chemical cross-linking. The results were incorporated with yeast-two-

hybrid and mutant data to develop a comprehensive interaction map. The combined data enabled the identification of a four-subunit scaffold, elucidation of a regulatory mechanism for complex assembly, and comparative analysis of the sub-complex with the related COP9 signalosome [31].

3. Flexibility and Dynamics

Flexibility can represent a critical extra dimension for many proteins. In such cases, combining techniques provides a more comprehensive description of structure and dynamics than using individual methods alone. Indeed, several recent high impact papers coupled high-resolution structure with biophysical approaches to describe protein flexibility and dynamics [32,33]. A recent review [34] describes how dynamic motion can be assessed in different ways, for example by trapping different states of a dynamic process, evaluating the structural ensemble, complementing structural data with kinetic information, or studying the structures and kinetics simultaneously (Figure 1). Figure 3 shows some of the structural tools that can be used to generate information about flexibility and dynamics.

An important question in understanding protein flexibility is whether structural differences between holo- and apo-enzymes represent induced-fit or selection of a pre-existing state. This question was addressed recently for maltose-binding protein [35]. High-resolution crystal structures of the holo- and apo- forms of the protein showed that the two enzyme domains are rotated by ~35° with respect to each other in the two structures. Application of paramagnetic NMR (PM-NMR) relaxation enhancement to spin-labeled holo- and apo-enzyme solutions demonstrated for the first time the presence of a pre-existing holo-form-like conformation (at ~5%) in the apo-enzyme.

SAXS and NMR data can generate structural ensembles for flexible macromolecules, and these ensembles are thought to represent the molecule's range of motion; how realistic is this assumption? One recent study focused on the enzyme matrix metalloprotease 9 that incorporates a putative flexible linker [36]. A combination of SAXS and high-resolution domain structures generated a number of full-length structures. Atomic force microscopy (AFM) was used to measure molecular dimensions one molecule at a time, thereby confirming the range of motion. Similarly,

the full-length, flexible, multi-domain p53 protein was studied by a combination of techniques [37]. SAXS and single particle analysis EM (SPA-EM) showed that unliganded p53 is characterized by a heterogeneous conformational population. When p53 is complexed with DNA, both techniques indicated a considerable reduction in flexibility.

The recently defined class of natively unfolded proteins (reviewed in [38]) is not amenable to crystallographic methods. However, NMR is particularly suited to their study [39] especially when combined with other methods. Recent work has combined NMR, CD and cross-linking [40]; NMR, CD and SAXS [41] and NMR, CD, SEC, AUC, dynamic light scattering (DLS) and cross-linking [42]. One example where NMR was not required is the study of the N-terminal regions of the Msh6 and Msh3 proteins (that recognize mismatched DNA bases) [43]. These regions were evaluated by comparing SAXS data with theoretical models of random peptide chains, to demonstrate their native disorder [43]. Furthermore, the C-terminal domain of the *Shaker* voltage-activated potassium channel was shown to be intrinsically disordered by using a combination of SEC, AUC and CD [44]. In both cases, mutagenesis indicated that inherent flexibility is required for function.

CONCLUSIONS

On their own, individual types of structural data can have considerable limitations or uncertainties, but these can often be overcome or minimized by combining synergistic data. When all 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 by satisfying all restraints, by the variability of the generated structures, by cross-validating through omitting portions of the data, by including incorrect data (which should lead to poorly-resolved structures), and by evaluating the model in the light of other data not included in the structure calculation.

Now that high-resolution macromolecular structure determination has become almost commonplace for standard targets, structural biologists routinely incorporate biological data to gain a better

understanding of function. Similarly, coupling high-resolution structure with data from techniques that describe dynamics also value-adds to our understanding of function. The routine nature of modern high-resolution structural biology means that many "low-hanging fruit" macromolecules are already well-characterised in structural terms. Using the same analogy, we then need a "ladder" to reach the more difficult "high-hanging fruit", such as membrane proteins, mega-complexes or natively disordered proteins. If current trends are any indication, combining data from multiple methods is a means of providing such a ladder, enabling structural biologists to tackle ever larger and more challenging systems.

ACKNOWLEDGEMENTS

We thank Bonnie Wallace for contributing to early discussions on this paper, and Bohumil Maco and Thomas Huber for helpful discussions. We apologise to researchers whose work relevant to this topic could not be cited due to word limits. Our work is supported by an Australian Research Council (ARC) grant to JLM and BK. NPC is a Monash Centre for Synchrotron Science Fellow, BK is an ARC Federation Fellow and a National Health and Medical Research Council (NHMRC) Honorary Research Fellow, and JLM is a NHMRC Senior Research Fellow.

REFERENCES

1. Benesch JL, Ruotolo BT, Simmons DA, Robinson CV: **Protein complexes in the gas phase: technology for structural genomics and proteomics**. *Chem Rev* 2007, **107**:3544-3567.

2. Cowieson NP, Miles AJ, Robin G, Forwood JK, Kobe B, Martin JL, Wallace BA: **Evaluating protein:protein complex formation using synchrotron radiation circular dichroism spectroscopy**. *Proteins* 2008, **70**:1142-1146.

3. Alexander N, Bortolus M, Al-Mestarihi A, McHaourab H, Meiler J: **De novo high-resolution protein structure determination from sparse spin-labeling EPR data**. *Structure* 2008, **16**:181-195.

4. King G, Ross I, Jones A, Kobe B, Huber T, Mouradov D, Hume D: **Rapid identification of disulfide-containing chemical cross-links in proteins using MALDI-TOF/TOF-mass spectrometry**. *Anal Chem* 2008:in press.

5. Tang C, Clore GM: **A simple and reliable approach to docking protein-protein complexes from very sparse NOE-derived intermolecular distance restraints**. *J Biomol NMR* 2006, **36**:37-44.

6. Minor DL, Jr.: **The neurobiologist's guide to structural biology: a primer on why macromolecular structure matters and how to evaluate structural data**. *Neuron* 2007, **54**:511-533.

7. Neylon C: **Small angle neutron and X-ray scattering in structural biology: recent examples from the literature**. *Eur Biophys J* 2008.

8. Wlodawer A, Minor W, Dauter Z, Jaskolski M: **Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures**. *Febs J* 2008, **275**:1-21.

9. Whitten AE, Trewhella J: **Small-angle scattering and neutron contrast variation for studying bio-molecular complexes**. *Methods in Molecular Biology* 2008, **Microbiology,**

Nanotechnologies, and Physical Chemistry in Separation, Detection, and Analysis of Biomolecules.:Ed James W. Lee, Humana Press, USA.

10. Miles AJ, Wallace BA: **Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics**. *Chem Soc Rev* 2006, **35**:39-51.

11. Fleishman SJ, Unger VM, Ben-Tal N: **Transmembrane protein structures without X-rays**. *Trends Biochem Sci* 2006, **31**:106-113.

12.** Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, et al.: **GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function**. *Science* 2007, **318**:1266-1273.

These three papers describe a landmark in structural biology, the first structure of a drug-bound G-protein coupled receptor. The result of a long-running research program that used a variety of methods to fully characterise the protein, the crystal structure determination was underpinned by the identification of flexible regions that required modification prior to crystallisation.

13.** Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, et al.: **High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor**. *Science* 2007, **318**:1258-1265.

14.** Rasmussen SG, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, et al.: **Crystal structure of the human beta2 adrenergic G-protein-coupled receptor**. *Nature* 2007, **450**:383-387.

15. Granier S, Kim S, Shafer AM, Ratnala VR, Fung JJ, Zare RN, Kobilka B: **Structure and conformational changes in the C-terminal domain of the beta2-adrenoceptor: insights from fluorescence resonance energy transfer studies**. *J Biol Chem* 2007, **282**:13895-13905.

16. Kobilka BK: **G protein coupled receptor structure and activation**. *Biochim Biophys Acta* 2007, **1768**:794-807.

17.* Gingras AR, Bate N, Goult BT, Hazelwood L, Canestrelli I, Grossmann JG, Liu H, Putz NS, Roberts GC, Volkmann N, et al.: **The structure of the C-terminal actin-binding domain of talin**. *EMBO J* 2008, **27**:458-469.

To address how this mammoth 2500 residue protein interacts with F-actin, the time-honoured approach of divide and conquer was combined with multiple structural methods including MX, NMR, SAXS and EM.

18. Ashish, Paine MS, Perryman PB, Yang L, Yin HL, Krueger JK: **Global structure changes associated with Ca2+ activation of full-length human plasma gelsolin**. *J Biol Chem* 2007, **282**:25884-25892.

19. Cowieson N, King G, Cookson D, Ross I, Huber T, Hume D, Kobe B, Martin J: **Cortactin adopts a globular conformation and bundles actin into sheets**. *J Biol Chem* 2008, **in press**.

20. Forwood JK, Thakur AS, Guncar G, Marfori M, Mouradov D, Meng W, Robinson J, Huber T, Kellie S, Martin JL, et al.: **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* 2007, **104**:10382-10387.

21. Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, Baker D: **High-resolution structure prediction and the crystallographic phase problem**. *Nature* 2007, **450**:259-264.

22. Hao Q: **Macromolecular envelope determination and envelope-based phasing**. *Acta Crystallogr D Biol Crystallogr* 2006, **62**:909-914.

23. Vestergaard B, Sanyal S, Roessle M, Mora L, Buckingham RH, Kastrup JS, Gajhede M, Svergun DI, Ehrenberg M: **The SAXS solution structure of RF1 differs from its crystal structure and is similar to its ribosome bound cryo-EM structure**. *Mol Cell* 2005, **20**:929-938.

24. Robinson CV, Sali A, Baumeister W: **The molecular sociology of the cell**. *Nature* 2007, **450**:973-982.

25. Stein M, Gabdoulline RR, Wade RC: **Bridging from molecular simulation to biochemical networks**. *Curr Opin Struct Biol* 2007, **17**:166-172.

26. Mueller M, Jenni S, Ban N: **Strategies for crystallization and structure determination of very large macromolecular assemblies**. *Curr Opin Struct Biol* 2007, **17**:572-579.

27. Hanein D: **Structural analysis of supramolecular assemblies by hybrid methods**. *J Struct Biol* 2007, **158**:135-136.

28. Liljas A: **On the complementarity of methods in structural biology**. *Acta Crystallogr D Biol Crystallogr* 2006, **62**:941-945.

29.** Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, et al.: **The molecular architecture of the nuclear pore complex**. *Nature* 2007, **450**:695-701.

In these two papers, a diverse set of biochemical, biophysical and proteomic data were integrated computationally to derive an improved model of the nuclear pore complex. The first paper focuses on the integration strategy, which is applicable to structure determination of other macromolecular assemblies. The biological implications of the NPC structure are discussed in the second paper.

30.** Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, et al.: **Determining the architectures of macromolecular assemblies**. *Nature* 2007, **450**:683-694.

31.* Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV: **Structural organization of the 19S proteasome lid: insights from MS of intact complexes**. *PLoS Biol* 2006, **4**:e267.

A comprehensive interaction map of the 19S proteasome lid complex was created using a combination of tandem mass spectrometry and chemical crosslinking, and previously available yeast two-hybrid data.

32. Overall CM, Butler GS: **Protease yoga: extreme flexibility of a matrix metalloproteinase**. *Structure* 2007, **15**:1159-1161.

33. Lipfert J, Doniach S: **Small-angle X-ray scattering from RNA, proteins, and protein complexes**. *Annu Rev Biophys Biomol Struct* 2007, **36**:307-327.

34. Henzler-Wildman K, Kern D: **Dynamic personalities of proteins**. *Nature* 2007, **450**:964-972.

35.** Tang C, Schwieters CD, Clore GM: **Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR**. *Nature* 2007, **449**:1078-1082.

Here, PM-NMR was used to demonstrate that a small percentage of a population of unliganded maltose binding protein exists in a conformation similar to that of the liganded form. Quantitative measurements of alternate conformations are rare in structural biology and this seminal paper is the first direct evidence of a 'pre-existing state' mode of protein binding.

36. Rosenblum G, Van den Steen PE, Cohen SR, Grossmann JG, Frenkel J, Sertchook R, Slack N, Strange RW, Opdenakker G, Sagi I: **Insights into the structure and domain flexibility of full-length pro-matrix metalloproteinase-9/gelatinase B**. *Structure* 2007, **15**:1227-1236.

37. Tidow H, Melero R, Mylonas E, Freund SM, Grossmann JG, Carazo JM, Svergun DI, Valle M, Fersht AR: **Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex**. *Proc Natl Acad Sci U S A* 2007, **104**:12324-12329.

38. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK: **Intrinsic disorder and functional proteomics**. *Biophys J* 2007, **92**:1439-1456.

39. Dyson HJ, Wright PE: **Elucidation of the protein folding landscape by NMR**. *Methods Enzymol* 2005, **394**:299-321.

40. Bruns K, Studtrucker N, Sharma A, Fossen T, Mitzner D, Eissmann A, Tessmer U, Roder R, Henklein P, Wray V, et al.: **Structural characterization and oligomerization of PB1-F2, a proapoptotic influenza A virus protein**. *J Biol Chem* 2007, **282**:353-363.

41. Makowska J, Rodziewicz-Motowidlo S, Baginska K, Vila JA, Liwo A, Chmurzynski L, Scheraga HA: **Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins**. *Proc Natl Acad Sci U S A* 2006, **103**:1744-1749.

42. Batra-Safferling R, Abarca-Heidemann K, Korschen HG, Tziatzios C, Stoldt M, Budyak I, Willbold D, Schwalbe H, Klein-Seetharaman J, Kaupp UB: **Glutamic acid-rich proteins of rod photoreceptors are natively unfolded**. *J Biol Chem* 2006, **281**:1449-1460.

43.* Shell SS, Putnam CD, Kolodner RD: **The N terminus of Saccharomyces cerevisiae Msh6 is an unstructured tether to PCNA**. *Mol Cell* 2007, **26**:565-578.

This paper describes the use of theoretical models of random peptides and SAXS data to show that regions of two homologous proteins are disordered. This method of data interpretation represents a powerful approach to demonstrate native disorder.

44. Magidovich E, Orr I, Fass D, Abdu U, Yifrach O: **Intrinsic disorder in the C-terminal domain of the Shaker voltage-activated K+ channel modulates its interaction with scaffold proteins**. *Proc Natl Acad Sci U S A* 2007, **104**:13022-13027.

FIGURE LEGENDS

Figure 1. Schematic diagram highlighting the synergies and integration of different structural methods. The traditional methods in the blue box generate 3D structure and symmetry information. Non-traditional and hybrid approaches can give rise to the types of data listed in the yellow box that can help advance both structural and dynamic studies of macromolecules. See Table 1 for more information on individual methods, and abbreviations used in text and figures.

Figure 2. Two examples from our labs that used hybrid methods to generate structural information for protein targets. **A.** Cortactin, a multidomain protein that regulates actin dynamics. We used a combination of bioinformatic sequence analysis and cross-linking to demonstrate interaction between the actin binding domains (green balls) and the C-terminal SH3 domain (red oval). A low resolution SAXS structure (grey density) confirmed the globular nature of the protein allowing us to develop a model of the structure. Intramolecular binding of the SH3 domain is likely to be a mechanism of autoinhibition. **B.** Acyl CoA thioesterase 7 (Acot7) is a two-domain protein that trimerises in solution. The full–length protein could not be crystallised, but both domains were solved independently by MX, each revealing a hotdog domain in a hexameric arrangement. However, neither domain has enzymatic activity on its own. We used both N-domain and C-domain structures (active site residues from the N- and C-terminal domains coloured red and blue, respectively) plus AUC, SEC, cross-linking/MS (crosslinks indicated as black lines) and mutagenesis data to generate the model of full-length Acot7, showing that catalytic residues from both domains (red, blue) are required to generate the three active sites in the trimer.

Figure 3. Schematic diagram showing some of the types of structural data that can be generated for flexible molecules. A hypothetical two-domain protein with freedom of movement between the domains is represented as a grey fan. The structure of such a flexible protein can be represented in different ways and several possible models are shown in red with examples of techniques that give information about each kind of model.

Figure 1
Click here to download high resolution image

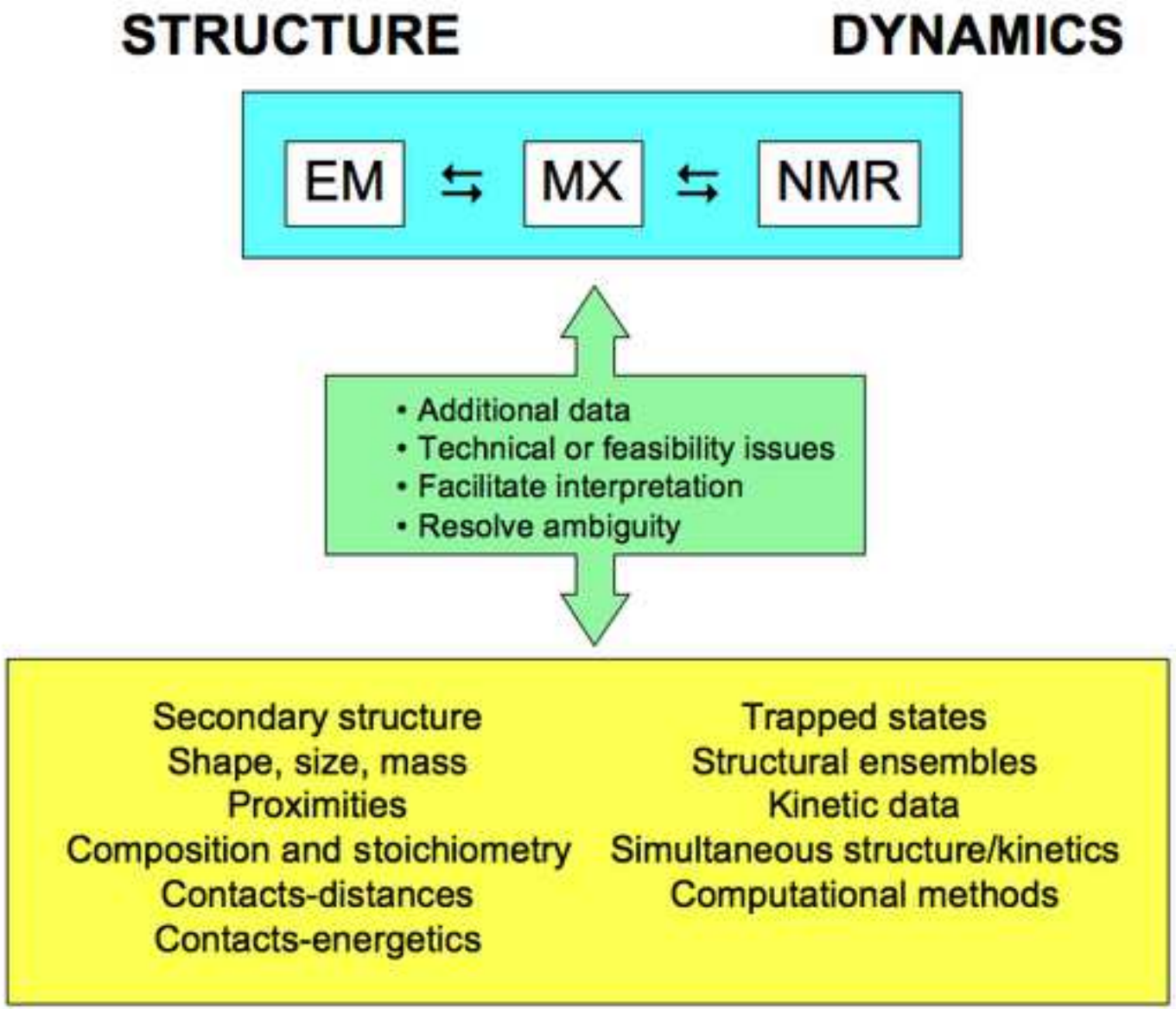

Figure 2
Click here to download high resolution image

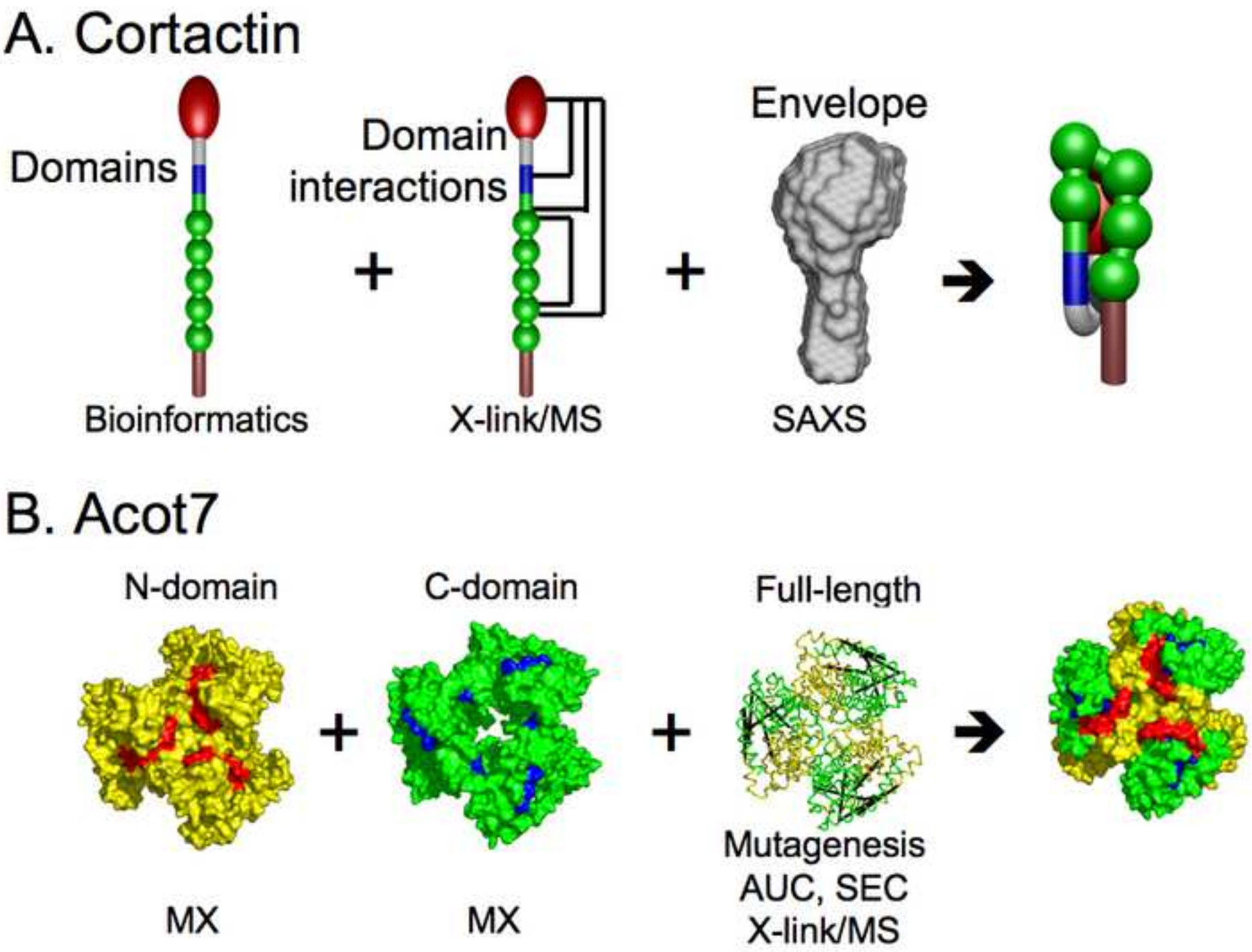

Figure 3

Click here to download high resolution image

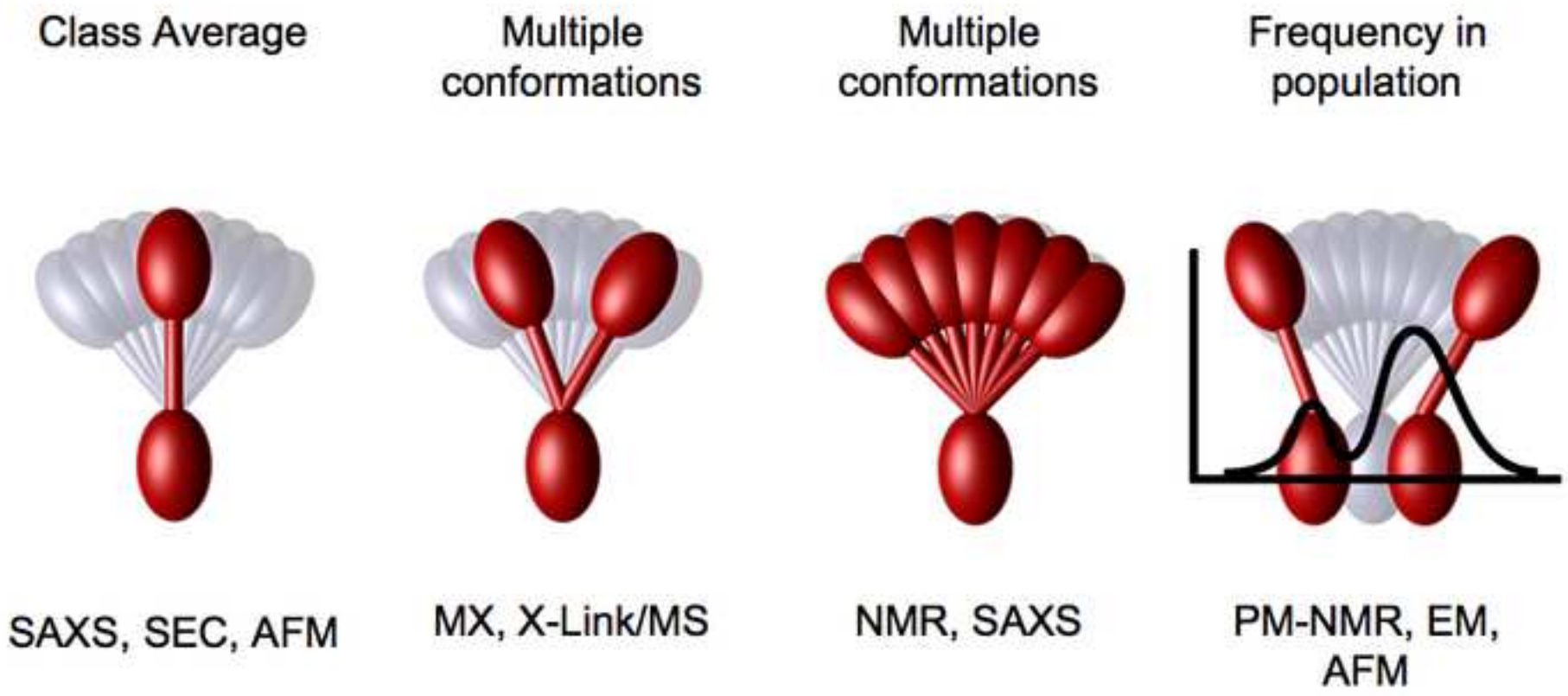

Table 1. Examples of types of structural data and methods commonly used to generate the data

Structural data generated	Methods that can be used
Molecular structure: medium to high resolution information on the 3D position of atoms in a macromolecule	Macromolecular X-ray crystallography (MX), nuclear magnetic resonance (NMR, *eg* nuclear Overhaiuser effect (NOE))*, single particle analysis cryo-electron microscopy (SPA cryo-EM)*, electron crystallography, neutron crystallography, homology modelling
Secondary structure: percentage of helix, strand and coil in a protein	Circular dichroism (CD) and synchrotron radiation CD (SRCD), bioinformatics (secondary structure prediction)
Molecular shape, size and mass of macromolecules and assemblies	Small angle X-ray (SAXS) and neutron (SANS) scattering, scanning transmission EM (STEM), negative stain EM, electron and X-ray tomography, mass spectrometry (MS), analytical ultra-centrifugation (AUC), size-exclusion chromatography (SEC), dynamic light scattering (DLS), static light scattering (SLS), atomic force microscopy (AFM), bioinformatics (domain prediction)
Dynamics: flexibility and conformational changes	SAXS, SRCD, NMR (including paramagnetic NMR (PM-NMR); *eg* relaxation data), ultraviolet-visible fluorescence, Raman spectroscopy, hydrogen/deuterium exchange NMR or MS, Laue crystallography, molecular dynamics simulations
Proximities: distances between two points on a macromolecule	NMR (*eg* NOE or PM-NMR), chemical cross-linking/mass spectrometry, fluorescence resonance energy transfer (FRET), electron paramagnetic resonance (EPR)
Composition and stoichiometry of complexes	Immuno-EM, labelling by fusion proteins, subcellular fractionation, quantitative immunoblotting
Contacts-distances: interaction mapping and identification of interacting parts of proteins	NMR (*eg* chemical shifts), chemical cross-linking, affinity purification, yeast two-hybrid, protein-fragment complementation assays, phage display, protein arrays, surface plasmon resonance (SPR), overlay assays, footprinting, limited proteolysis, mutagenesis, hydrogen/deuterium exchange NMR and MS
Contacts-energetics: binding interactions, energetics and kinetics	SPR, isothermal titration calorimetry (ITC), differential scanning calorimetry (DSC), thermal stability measurements

* Many methods, especially NMR and EM, include different approaches that can be used to derive different types of structural and dynamic information. We specifically mention here only some of the more commonly used approaches or measurements.