PROTEIN REGULATION, PROTEIN-PROTEIN INTERACTIONS AND STRUCTURAL GENOMICS†

Bostjan Kobe

*Department of Biochemistry and Molecular Biology, Institute for Molecular Bioscience,
Special Research Centre for Functional and Applied Genomics, and Cooperative Research Centre for
Chronic Inflammatory Diseases, University of Queensland, Brisbane, Queensland 4072 Australia*

Abstract

The new technical developments and the success of genome sequencing projects have prompted a new approach to scientific investigation and discovery in every field of biochemistry and molecular biology, including structural biology. One of the most prominent recent developments is the birth of structural genomics, a world-wide initiative that aims to provide the three-dimensional structures of all representative proteins. However, structural biology faces an exciting future beyond structural genomics; if we are to understand how the proteome works and use the genomic information for therapeutic purposes, studies of protein-protein interactions and macromolecular complexes, mechanism and regulation of macromolecular function, membrane protein structure, and structure-based therapeutic design must be pursued in parallel. Successful approaches will combine large-scale, high-throughput approaches developed through structural genomics with more traditional hypothesis-driven approaches, supported by integrative bioinformatics tools.

The limited funding resources and limited opportunities for involvement in large consortia in a country of the size of Australia require creative strategies in approaching structural biology problems. This article reviews some of the directions pursued by our laboratory, including a 'focused' structural genomics program suited for smaller-scale teams, and studies of protein-protein interactions (exemplified by the work on nuclear transport proteins and protein kinases) and protein regulation (exemplified by the work on nuclear transport proteins and phenylalanine hydroxylase).

Introduction

This article is based on a lecture in the 'Perspectives' session of the 1st Central European Conference 'Chemistry towards Biology'. The article gives the author's subjective view on the perspectives in structural biology in the coming decade, and attempts to link these perspectives to the research in the author's laboratory. The research develops necessarily as a compromise between (i) what the author and his coworkers find exciting and significant, and (ii) the restrictions imposed by the funding

situation and the research environment at an Australian University. Despite being much larger in size, Australia is similar in population to many central European countries, and therefore the circumstances may be relevant to central European communities.

The new technical developments and the success of genome sequencing projects have prompted a new approach to scientific investigation and discovery in every field of biochemistry and molecular biology, including structural biology. One of the most prominent recent developments is the birth of structural genomics, a world-wide initiative that aims to provide the three-dimensional structures of all representative proteins.[1] However, structural biology faces an exciting future beyond structural genomics. The determination of all representative structures is an important yet only a small step towards understanding the molecular basis of biological processes. Strategic directions taking place in parallel and beyond the current stage of structural genomics will include the studies of protein-protein interactions and macromolecular complexes, mechanism and regulation of macromolecular function, and membrane protein structure, as well as structure-based therapeutic design. Successful approaches will combine large-scale, high-throughput approaches developed through structural genomics with more traditional hypothesis-driven approaches, supported by integrative bioinformatics tools.

The limited funding resources and opportunities for involvement in large consortia in a country such as Australia require creative strategies in approaching structural biology problems. Our group is developing a 'focused' structural genomics program suited for smaller-scale teams, and in parallel pursuing smaller scale projects in protein-protein interactions and protein regulation, applying the high-throughput approaches developed for structural genomics, to other projects. Our efforts will be illustrated by our structural genomics of macrophage proteins, the studies of active site-directed protein regulation (nuclear transport proteins, phenylalanine hydroxylase), and the studies of protein-protein interactions (nuclear transport proteins and protein kinases).

Structural genomics of macrophage proteins

The Human Genome Project and other high-throughput genome sequencing efforts result in the identification of large numbers of proteins, a large portion with unknown functions (~40% in the human genome). The next big issue in biology is to define the structures and functions of all these proteins. The function of a protein directly depends

on its three-dimensional (3D) structure. Sequence alignments offer the first approach for functional annotation of a novel protein; however, the evolutionary constraints for 3D structures are known to be even higher than for sequences. The knowledge of the 3D structure of a protein is therefore one of the most powerful avenues for inferring functional information (*e.g.* [2-5]). This notion led to the development of a new field of structural biology termed structural genomics. The goal of structural genomics is to provide a comprehensive view of protein structure universe, through determining the structure of at least one representative protein from every protein family.[6] High-throughput structure determination required to make such an approach feasible has recently been demonstrated, through technological advances in recombinant technology and protein expression, structure determination (in particular X-ray crystallography; X-ray detectors, cryogenic data collection and tunable synchrotron radiation sources) and high-performance computing. The structures of representative proteins subsequently allow the prediction of 3D structures of a large number of related proteins.[7]

Achieving the goals of structural genomics requires large teams and substantial funding. However, the methodology of the structural genomics approach, in terms of pursuing the more manageable projects ('low-hanging fruit') first, can also be applied to projects of a smaller scale, and promises faster and more cost-effective progress. Furthermore, a smaller team can identify a niche in the world-wide structural genomics initiative through intelligent protein target selection.

We applied these ideas to a project involving structural characterization of proteins with roles in macrophages. Macrophages are cells that play a crucial role in innate immunity and are consequently associated with inflammatory disease and cancer. We use gene expression information obtained via DNA microarray technology to identify proteins with putative roles in macrophage function. Targets for structure determination are chosen from this large set of proteins using a set of criteria that will maximize the insight into protein function (preference is given to proteins with novel structural motifs, proteins with unknown molecular functions, and proteins with stronger evidence for the role in macrophages; discussed in more detail below).

Most pathogens that attempt to invade mammalian cells fail at the very first stage due to the remarkable effectiveness of innate immunity. The presence of potential pathogens is detected via receptors that recognize generic non-mammalian structures

including cell wall components (lipopolysaccharide (LPS), peptidoglycans, lipotechoic acids) and microbial DNA (*e.g.* unmethylated CpG motifs).[8] The first line of defense is the macrophage, which comprises 15-20% of the cells in most organs, and is particularly abundant at the routes of pathogen entry such as lung, skin, gut and genitourinary tract.[9] Upon recognition of a potential pathogen, the macrophage engulfs and attempts to destroy the foreign organism. At the same time, it activates a remarkable spectrum of genes creating a hostile extracellular environment (via the acute phase response, fever, local blood coagulation, natural antibiotics/defensins), recruits additional cells to the site of invasion (via secretion of a wide range of chemotactic factors and proinflammatory agents) and primes an appropriate acquired immune response specific to the class of pathogen (through actions of specific cytokines such as tumor necrosis factor-α and interleukins 1, 6, 12 and 18). A successful pathogen overcomes these defenses; many even take advantage of the macrophage as a portal of infection and replicate within the cell. Failure of innate defense does not preclude continued secretion of macrophage products. Acute infections lead to life-threatening effects, disseminated intravascular coagulation, hypotension and pathological fever. In chronic local infections, or in response to inflammation caused by non-infectious agents that activate macrophages but cannot be cleared, the less acute actions of macrophage products still cause local tissue destruction and wasting disease (cachexia).

The knowledge of regulation of macrophage function will form the basis of two classes of therapeutics. On the one hand, we may want to amplify the toxic function of macrophages to destroy microorganisms or tumor cells more effectively. On the other hand, selective suppression of components of the macrophage activation response offers approaches to treatment of septicemia and toxic shock, arthritis, atherosclerosis and other chronic inflammatory diseases.

We are using the following experimental procedure (Figure 1). The major fundamental criterion for target selection is the evidence of either macrophage-specific expression or induction by macrophage-activating agents. Proteins with sequence similarity to known protein structures, and transmembrane regions of proteins, are discarded. The targets are prioritized to maximize insight into protein function. Next, the target proteins are subjected to expression and purification, the protocol consisting of two major steps: (i) a small-scale screen for soluble protein expression; and (ii) larger

scale protein expression and purification. We are using the Gateway cloning technology (Invitrogen) to construct the expression vectors. The proteins are expressed using the hexa-histidine tag, and purified using affinity chromato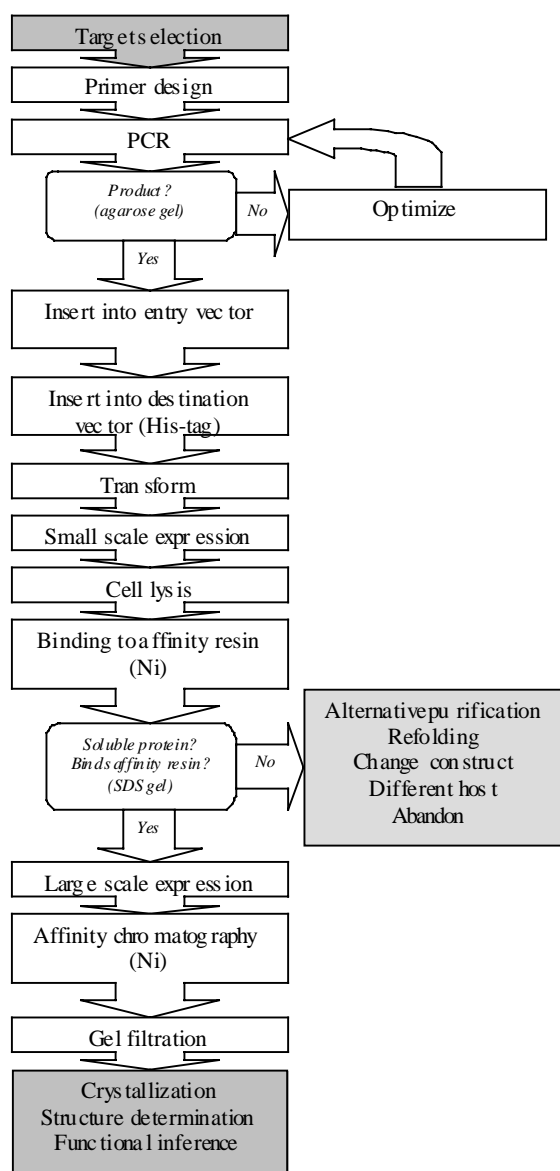graphy (nickel resin) followed by size exclusion chromatography. The proteins are finally subjected to crystallization screening (with sparse-matrix crystallization screens) using hanging-drop vapor diffusion in 96-well plates. The structures are planned to be determined primarily by the multiwavelength anomalous dispersion (MAD) method using seleno-methionine-labelled proteins. The results of all stages of the experimental work are recorded using a computer project management system LISA.[10]

The combination of gene-expression analysis and 3D structure determination provides unprecedented possibilities for functional annotation of proteins with unknown or poorly characterized functions. Gene expression analysis provides information about involvement in cellular processes (the so-called cellular or biological function), while 3D structures provides information about possible enzymatic or binding activities (the so-called biochemical or molecular function).

Figure 1. A flowchart showing the basic steps in protein production in the structural genomics appproach. The two major decision points are indicated.

Since the start of the project in 2001, we have subjected 40 proteins to the pipeline. The cDNAs of most of these proteins have been successfully cloned into expression

vectors. Around a quarter of proteins show soluble expression, consistent with observations by other investigators.[11] These proteins are currently undergoing crystallization studies.

Intrasteric (active site-directed) protein regulation

The term intrasteric regulation was introduced to describe autoregulation of protein kinases and phosphatases by internal sequences resembling substrates ('pseudosubstrates'), and acting directly at the active site.[12] Although indirect biochemical evidence supported the intrasteric regulation hypothesis, unequivocal confirmation has only become available relatively recently through structural studies of autoinhibited enzymes, such as cAMP-dependent protein kinase with the bound peptide inhibitor,[13] twitchin,[14] calmodulin-dependent protein kinase-1,[15] and the protein phosphatase calcineurin.[16] In the basic scheme of intrasteric regulation, the protein is maintained in an inactive state through the binding of an autoregulatory sequence that masks the active site. In this way, intrasteric regulation is the converse of the better known allosteric regulation.[17] Activation is achieved through an activatory ligand or protein, or post-translational modifications, resulting in the release of the autoregulatory sequence from the active site.

An interesting example of intrasteric regulation is observed in the metabolic enzyme phenylalanine hydroxylase (PAH). PAH converts phenylalanine to tyrosine. It is structurally related to tyrosine hydroxylase (TH) and tryptophan hydroxylase (TPH), both involved in the biosynthesis of the neurotransmitters.[18] PAH needs to be regulated very tightly, because it manages the level of phenylalanine, an essential amino acid, which is subject to large fluctuations as a result of dietary intake. On the one hand, an uncontrolled enzyme would rapidly deplete the phenylalanine stores in the liver; on the other hand, the metabolites of phenylalanine are toxic to the developing brain. Therefore, PAH is regulated via activation by phenylalanine and phosphorylation, and inhibition by tetrahydrobiopetrin (BH_4).[18] Activation by the substrate phenylalanine is considered the major regulatory event, and is accompanied by large conformational changes.

We determined the crystal structure of rat PAH_{1-428} (containing a short truncation at the C-terminus), revealing two domains: a C-terminal catalytic domain, and an N-terminal regulatory domain.[19] The very N-terminal sequence comprising amino acids

19-29 reached into the active site of the catalytic domain and appeared to autoinhibit the enzyme. We tested this autoinhibitory role of the N-terminal sequence by expressing a protein lacking the 29 N-terminal amino acids (PAH$_{30-428}$) and confirmed that PAH$_{30-428}$ was constitutively active (*i.e.* it does not require phenylalanine activation).[20] The mutant also showed an altered structural response to phenylalanine. Similar results were obtained using PAH lacking the first 26 residues.[21]

A surprising observation revealed by the structure of PAH$_{1-428}$ was that residues 1-18, containing the phosphorylation site Ser16, showed no defined structure in both phosphorylated and un-phosphorylated forms;[19] this was difficult to reconcile with the established role of phosphorylation in activating the enzyme.[18] We used nuclear magnetic resonance (NMR) to follow the dynamics of the N-terminal mobile region. Our results confirm that this region is mobile in absence of phenylalanine, but a significant loss of mobility is observed for a portion of the sequence after the addition of phenylalanine.[22] This observation suggests that upon activation, the N-terminal sequence becomes associated with the folded core of the molecule. According to our working model, the binding of phenylalanine to its regulatory site causes conformational changes, during which the N-terminal sequence moves away from the active site, with phosphorylation aiding this transition through stabilizing the phenylalanine-activated form (Figure 2). However, a structural characterization of the various ligand-bound states will be required for a complete understanding of the regulation of PAH.

Another example of intrasteric regulation involves the nuclear transport factor importin-α (Impα). Nuclear proteins are synthesized in the cytoplasm, and need to be transported into the nucleus through the nuclear pore complexes (NPCs) spanning the nuclear envelope. Most macromolecules require an active, signal-mediated transport process. The first and best characterized nuclear targeting signals are the 'classical' nuclear localization sequences (NLSs) that contain one or more clusters of basic amino acids.[23] The NLSs do not conform to a specific consensus sequence, and fall into two distinct classes termed monopartite NLSs, containing a single cluster of basic amino acids, and bipartite NLSs, containing two basic clusters. Despite the variability, the classical NLSs are recognized by the same receptor protein termed importin or karyopherin, a heterodimer of α and β subunits.[24] Impα contains the NLS-binding site and importin-β (Impβ) is responsible for the translocation of the importin-substrate

complex through the NPC. The transfer through the pore is facilitated by other factors including the GTPase Ran (Ras-related nuclear protein). Once inside the nucleus, Impβ binds to Ran-GTP, which causes the dissociation of the import complex (Figure 3).
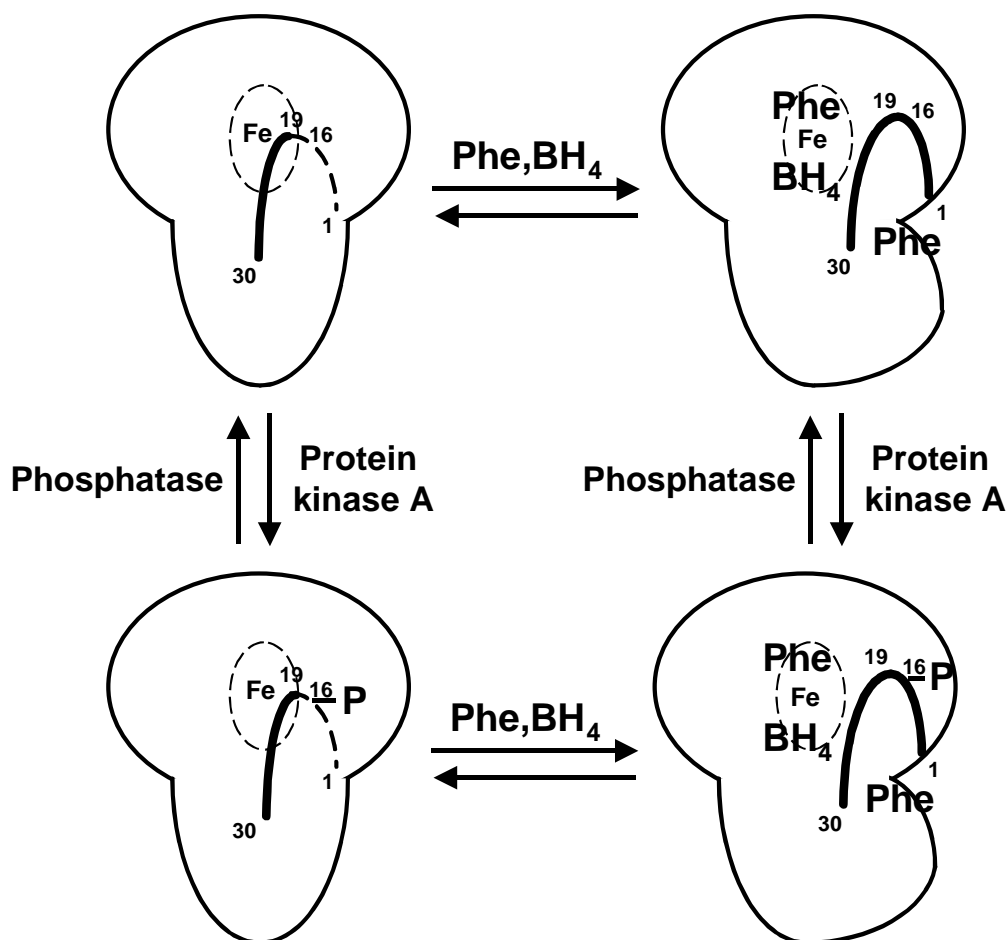

Figure 2. Schematic diagram of the regulation of PAH by phenylalanine, BH_4 and phosphorylation. The large object represents a monomer of PAH, with the large protrusion as the catalytic domain and the small protrusion as the regulatory domain. The dashed ellipse with Fe is the active site, and the thick curved line is the N-terminal autoregulatory sequence. The dashed line represents mobile regions, and the solid line represents ordered regions. 'Phe' and 'BH_4' roughly indicate phenylalanine and BH_4 binding sites. The right column represents active forms of PAH, and the left column autoinhibited forms of PAH. Phosphorylation (bottom row) facilitates the phenylalanine-induced conversion from the autoinhibited to the active form.

The crystal structure of mouse Impα revealed a large elongated domain corresponding to the majority of the protein (Figure 4).[25] However, a portion of the N-terminal sequence was observed binding along this domain. The binding site for this sequence corresponded to the NLS-binding site, revealing another example of intrasteric

regulation. In this case, the autoregulatory sequence (residues 44-54) is a clear case of a 'pseudosubstrate', as it shows close similarity with NLSs, and forms interactions analogous to the NLS with the binding site.[26]
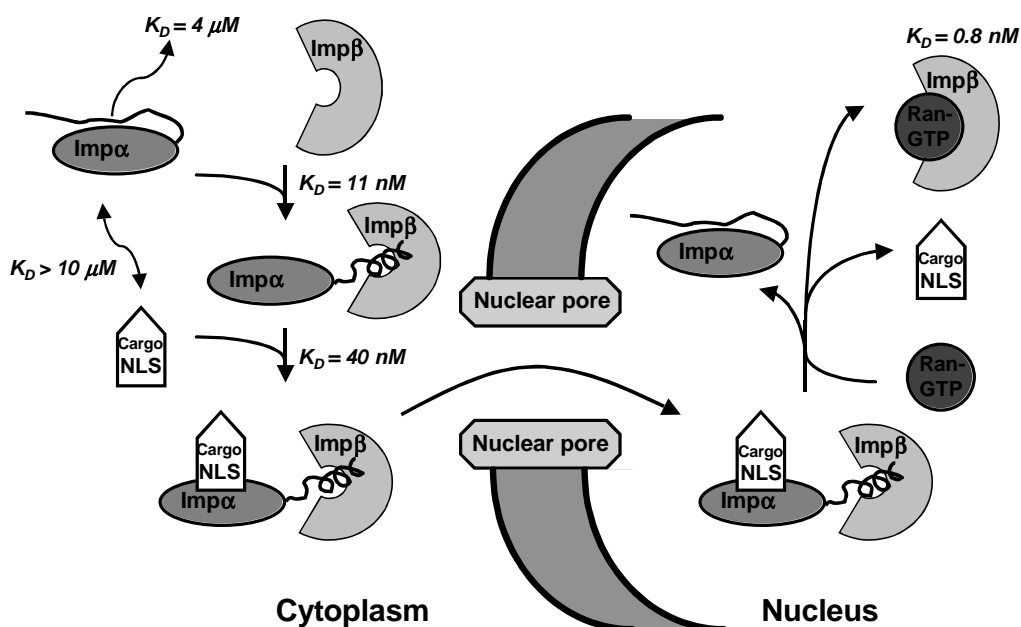

Figure 3. Schematic diagram of the NLS-dependent nuclear import pathway, highlighting the various binding affinities. Impα, oval light-grey object 'α'; Impβ, medium-grey object 'β'; NLS-containing cargo protein, white pentagonal object 'NLS'; RanGTP, round dark-grey object 'Ran GTP'. For simplicity, other factors involved in the pathway have been omitted from the diagram. The numbers correspond to the dissociation constants for the different binding events, based on biosensor studies.[27]

Figure 4. Structure of Impα. The majority of Impα is drawn as a ribbon diagram (with the programs Moslcript[38] and Raster3D[39]). The autoinhibitory region (residues 44-54) is shown in a ball-and-stick representation (dark grey).[25] Superimposed is the peptide corresponding to the NLS of nucleoplasmin (light grey).[28]

The autoregulatory sequence is a part of the N-terminal region of the protein, also called the 'IBB' (importin-β binding) domain. Impβ therefore functions not only to transport Impα into the nucleus, but also as its activator in the cytoplasm. The following model explains the regulation of nuclear import (Figure 3). In the nucleus, binding of Impα to nuclear proteins containing NLSs is not desired; the autoinhibitory IBB domain therefore prevents the binding of various nuclear proteins, and RanGTP prevents Impβ from binding to Impα. Once transported to the cytoplasm, however, Impβ binds to the IBB domain, removing it from the NLS-binding site and activating Impα. In the cytoplasm, the Impα-Impβ complex can therefore collect NLS-containing proteins destined for the nucleus and transport them there. Once the trimeric transport complex reaches the nucleus, however, the protein RanGTP binds to Impβ and displaces Impα, and Impα can release its cargo. The directionality of nuclear import is thought to be conferred by an asymmetric distribution of the GTP- and GDP-bound forms of Ran between the cytoplasm and the nucleus. This distribution is in turn controlled by various Ran-binding regulatory proteins.

We studied the thermodynamics and kinetics of various binding steps in the nuclear import pathway using surface plasmon resonance.[27] There appears to be an increase of at least 250-fold in affinity for NLS binding by Impα when Impβ is present (the dissociation constant increases from 40 nM to at least 10 μM). However, the affinity of a peptide corresponding to the autoinhibitory sequence of Impα, to a truncated Impα lacking the entire IBB domain, is only 4 μM! It is clear that the entropic contribution of the autoinhibition (in other words, the high local concentration of the autoinhibitory sequence, resulting from it being tethered to Impα) is an important factor determining the efficiency of autoinhibition and achieving the optimal balance of binding affinities during nuclear transport.

Protein-protein interactions

It used to be puzzling how a single receptor protein, Impα, can bind a diverse set of NLSs, including monopartite NLSs (*e.g.* P<u>KKRK</u>V, basic cluster underlined in the single letter amino acid code), and bipartite NLSs (*e.g.* <u>KR</u>PAATKKAGQA<u>KKKK</u>, both underlined basic clusters required). Furthermore, either group of NLSs contains a diverse set of sequences, with no obvious consensus. Our structures of complexes of mouse

Impα with peptides corresponding to NLSs,[28] and similar studies on yeast Impα,[29] explain the puzzle. The two clusters of basic residues in bipartite NLSs bind to two distinct regions on the surface of Impα, using electrostatic, polar and hydrophobic interactions, while the linker sequence between the two clusters makes fewer favorable contacts and therefore does not need to be highly conserved. We determined the structures of complexes of Impα with peptides corresponding to several different bipartite NLSs and find that the linker sequence can form a diverse set of interactions, depending on its sequence and length (unpublished results). The basic cluster in monopartite NLSs can interact with either binding region used by the bipartite NLS, but the one used by the C-terminal basic cluster of the bipartite NLSs is the high affinity site. The binding strategy used is extremely elegant, and explains the 'promiscuous specificity' of NLS binding; individual side chain-binding pockets can often accommodate either a lysine or arginine residue, determining the specificity of binding, but a significant part of the interaction is contributed by the main chain of the peptide.[28]

Phosphorylation in the vicinity of NLSs provides another opportunity for the regulation of nuclear import.[30] One system under complex phosphorylation control is the simian virus 40 large T-antigen (T-Ag); phosphorylation N-terminal to the NLS increases the efficiency of nuclear import. We determined the structures of the complexes of Impα with the phosphorylated and un-phosphorylated peptides corresponding to the relevant region of T-Ag, revealing that Impβ may play a role in the importin complex discriminating between the two forms of the peptide (unpublished results).

A rich source of insight into protein-protein interactions is provided by the family of protein kinases. Protein kinases are the enzymes responsible for protein phosphorylation, the most abundant type of cellular regulation. Phosphorylation affects essentially every cellular process including metabolism, growth, differentiation, motility, membrane transport, learning and memory, and defects in protein kinase function result in a variety of diseases. Protein kinases are a major target for drug design. To ensure signaling fidelity, kinases must be sufficiently specific and act only on a defined subset of cellular targets. Defining a substrate for a protein kinase defines its role in a particular cellular process. However, experimental approaches for determining specificity and particularly identifying *in vivo* substrates are laborious and expensive.

We reasoned that we can take advantage of quite extensive structural information on protein kinases to develop computational methods that predict substrate specificities of uncharacterized kinases. All protein kinases show a common fold, consisting of two lobes hinged through a short linker region. The active site is located in the cleft between the lobes. Although different enzymes in different states of activation show quite diverse conformations, the active forms of all protein kinase structures show a comparable 'closed' conformation, suggesting that any structural inferences can be extrapolated to the entire family.

Based on an analysis of the crystal structures of peptide complexes of protein Ser/Thr kinases,[31-33] we identified twenty enzyme residues ('determinants') that contact the side chains of the residues surrounding the phosphorylation site (only substrate positions (-3), (-2), (-1), (+1), (+2) and (+3) were considered). Using molecular modeling and sequence analysis of kinases and substrates, we extracted a set of rules that guide the specificity of binding to these positions. We implemented these rules in a web-interfaced computer program PREDIKIN that performs an automated prediction of optimal substrate peptides, using only the amino acid sequence of the protein kinase as input.[34]

PREDIKIN accepts a protein kinase sequence and outputs predictions of possible heptapeptide substrate sequences. First, it locates a characteristic conserved kinase motif and extracts the kinase catalytic domain from the protein sequence provided. Next, it locates other (semi)conserved kinase motifs, and based on the proximity to these motifs locates the determinant residues. It then applies the specificity rules and predicts an optimal heptapeptide sequence. To run the program, the user inputs the kinase type and sequence into a form in the browser window. Output consists of the locations of key kinase motifs, the type of kinase, a list of the determinant residues, a list of possible substrate heptapeptide sequences, and commentary text. Substrate data is passed to another window (automatically opened via a link) which contains substrate sequence data formatted for protein database searching. The program is available on http://www.biosci.uq.edu.au/kinsub/home.htm and is functional within Internet Explorer 5.
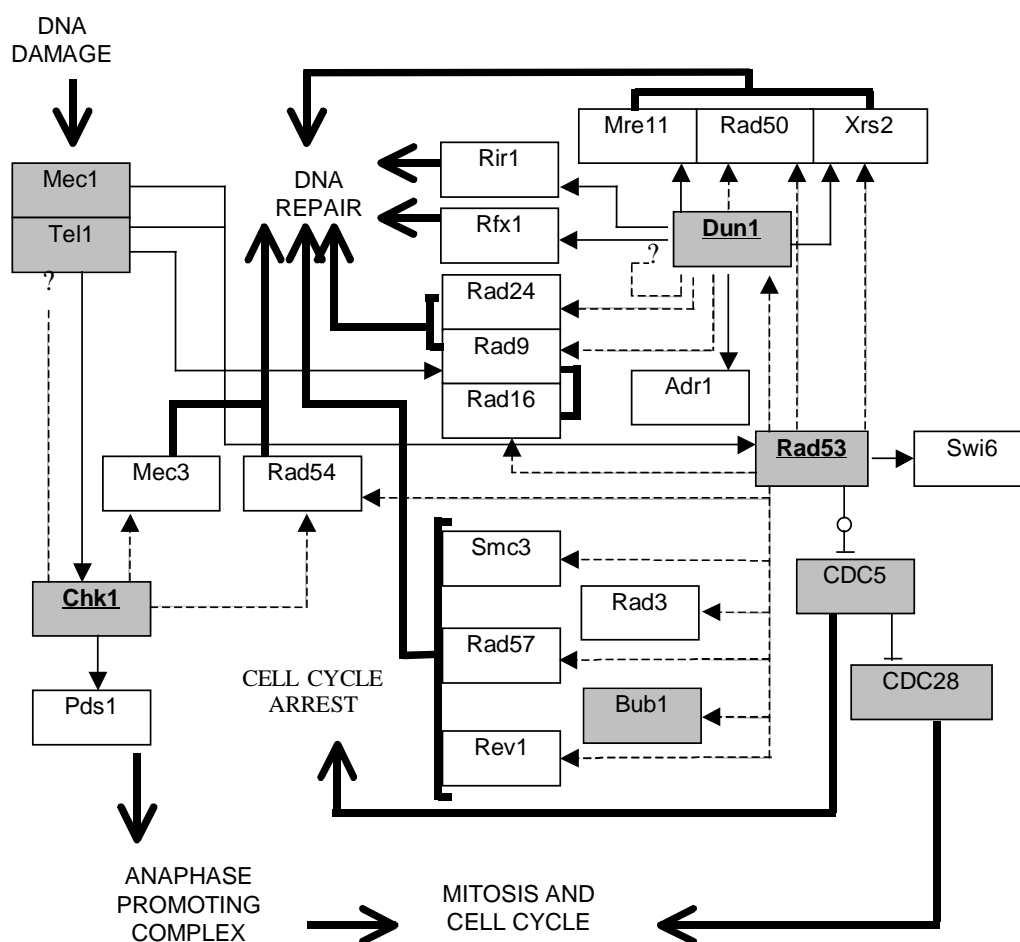
Figure 5. Schematic diagram of signaling connections linked to DNA damage checkpoints in *S. cerevisiae*. Grey boxes, protein kinases; solid and dashed connections, known and predicted phosphorylations, respectively; circles: predicted sites in known substrates; thick open arrows, general connections between processes. The joined boxes represent complexes. For the protein kinases analyzed (bold and underlined), all known interactions shown were also successfully predicted using PREDIKIN.

PREDIKIN attempts to predict the optimal phosphorylation sequences, analogous to those generated by an oriented peptide library experiment.[35] The predictions agree well with the peptide library results.[34] However, *in vivo* substrates do not necessarily contain the optimal motif. In the cell, the specificity does not depend only on the molecular recognition of a protein kinase for a certain peptide sequence, but is affected by other cellular mechanisms, particularly specific localization. For these reasons, PREDIKIN predictions must be treated prudently and integrated with other available information such as cellular localization, functional information and structural information for substrate proteins, and used with filtering tools such as dual motif

searches.[36] However, the use of PREDIKIN predicted motifs to search protein databases to identify substrates shows comparable statistics to the use of experimentally-determined motifs (based on peptide library experiments). Furthermore, the accuracy is comparable to secondary structure predictions, as well as systematic large-scale experimental methods.[37]

To explore the utility of the method, we used PREDIKIN to analyze the signaling pathways in several cellular processes in yeast. The example of the DNA damage checkpoint pathway shows that PREDIKIN can identify phosphorylation sites for substrates with unmapped sites, and many plausible phosphorylation events within the pathways and between proteins known to interact (Figure 5). The results suggest that PREDIKIN is an extremely useful tool for a rapid, *in silico* construction of signaling pathways and identification of therapeutic targets. Furthermore, our results demonstrate the potential that similar methodology is extended to other proteins which recognize short amino acid motifs, such as modular signal transduction domains (SH2, FHA).

Concluding remarks

Structural biology in the new millennium is not only concerned with the molecular function of proteins, but attempts to place the molecular function in the context of the cellular function. The projects in our laboratory attempt to establish this connection by linking structural information with cell biology by using a number of complementary techniques. This effort is best demonstrated in the cases of (i) the 'targeted structural genomics' approach, where microarray studies provide information on the cellular function, and structural studies provide information on the molecular function, and (ii) the approach used to predict protein kinase substrates, where structural information is used directly to facilitate predictions of cellular functions. Understanding the role of each protein in the proteome requires an integration of data provided by a variety of approaches.

Acknowledgements

David Hume, David Jans, Ian Jennings, Bruce Kemp, Pawel Listwan, Jenny Martin and Trazel Teh. The author is an NHMRC Senior Research Fellow.

References

1. S. K. Burley, S. C. Almo, J. B. Bonanno, M. Capel, M. R. Chance, T. Gaasterland, D. Lin, A. Sali, F. W. Studier, S. Swaminathan, *Nature Genet.* **1999**, *23*, 151–157.
2. T. I. Zarembinski, L.-W. Hung, H.-J. Mueller-Dieckmann, K.-K. Kim, H. Yokota, R. Kim, S.-H. Kim, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15189–15193.
3. K. Y. Hwang, J. H. Chung, S. H. Kim, Y. S. Han, Y. Cho, *Nature Struct. Biol.* **1999,** *6*, 691–696.
4. K. Volz, *Protein Sci.* **1999**, *8*, 2428–2437.
5. J. M. Thornton, A. E. Todd, D. Milburn, N. Borkakoti, C. A. Orengo, *Nature Struct. Biol.* **2000**, *7 Suppl*, 991–994.
6. S. K. Burley, *Nature Struct. Biol.* **2000**, *7 Suppl*, 932–934.
7. R. Sanchez, U. Pieper, F. Melo, N. Eswar, M. A. Marti-Renom, M. S. Madhusudhan, N. Mirkovic, A. Sali, *Nature Struct. Biol.* **2000**, *7 Suppl*, 986–990.
8. J. A. Hoffman, F. C. Kafatos, C. A. Janeway, R. A. Ezekowitz, *Science* **1999**, *284*, 1313–1318.
9. S. Gordon, P. R. Crocker, L. Morris, S. H. Lee, V. H. Perry, D. A. Hume, *Ciba Found. Symp.* **1986**, *118*, 54–67.
10. P. W. Haebel, V. L. Arcus, E. N. Baker, P. Metcalf, *Acta Crystallogr.* **2001**, *D57*, 1341–1343.
11. R. C. Stevens, *Structure Fold. Des.* **2000**, *8*, R177–R185.
12. B. E. Kemp, R. B. Pearson, *Biochim. Biophys. Acta* **1991**, *1094*, 67–76.
13. D. R. Knighton, J. Zheng, L. F. Ten Eyck, N.-H. Xuong, S. S. Taylor, J. M. Sowadski, *Science* **1991**, *253*, 414–420.
14. B. Kobe, J. Heierhorst, S. C. Feil, M. W. Parker, G. M. Benian, K. R. Weiss, B. E. Kemp, *EMBO J.* **1996**, *15*, 6810–6821.
15. J. Goldberg, A. C. Nairn, J. Kuriyan, *Cell* **1996**, *84*, 875–887.
16. C. R. Kissinger, H. E. Parge, D. R. Knighton, C. T. Lewis, L. A. Pelletier, A. Tempczyk, V. J. Kalish, K. D. Tucker, R. E. Showalter, E. W. Moomaw, N. L. Gastinel, N. Habuka, X. Chen, F. Maldonado, J. E. Barker, R. Backuet, J. E. Villafranca, *Nature* **1995**, *378*, 641–644.
17. J. Monod, J. P. Changeux, F. Jacob, *J. Mol. Biol.* **1963**, *6*, 306–329.
18. P. F. Fitzpatrick, *Annu. Rev. Biochem.* **1999**, *68*, 355–381.
19. B. Kobe, I. G. Jennings, C. M. House, B. J. Michell, K. E. Goodwill, B. D. Santarsiero, R. C. Stevens, R. G. H. Cotton, B. E. Kemp, *Nature Struct. Biol.* **1999**, *6*, 442–448.
20. I. G. Jennings, T. Teh, B. Kobe, *FEBS Lett.* **2001**, *488*, 196–200.
21. G. A. Wang, P. Gu, S. Kaufman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1537–1542.
22. J. Horne, I. G. Jennings, T. Teh, P. R. Gooley, B. Kobe, *Protein Sci.* **2002**, *11*, 2041–2047.
23. C. Dingwall, R. A. Laskey, *Trends Biochem. Sci.* **1991**, *16*, 478–481.
24. S. R. Wente, *Science* **2000**, *288*, 1374–1377.
25. B. Kobe, *Nature Struct. Biol.* **1999**, *6*, 388–397.
26. E. Conti, M. Uy, L. Leighton, G. Blobel, J. Kuriyan, *Cell* **1998**, *94*, 193–204.
27. B. Catimel, T. Teh, M. R. Fontes, I. G. Jennings, D. A. Jans, G. J. Howlett, E. C. Nice, B. Kobe, *J. Biol. Chem.* **2001**, *276*, 34189–34198.
28. M. R. M. Fontes, T. Teh, B. Kobe, *J. Mol. Biol.* **2000**, *297*, 1183–1194.
29. E. Conti, J. Kuriyan, *Structure* **2000**, *8*, 329–338.
30. D. A. Jans, S. Hubner, *Physiol. Rev.* **1996**, *76*, 651–685.
31. Madhusudan, E. A. Trafny, N.-H. Xuong, J. A. Adams, L. F. Ten Eyck, S. S. Taylor, J. M. Sowadski, *Protein Sci.* **1994**, *3,* 176-187.
32. D. J. Owen, M. E. M. Noble, E. F. Garman, A. C. Papageorgiou, L. N. Johnson, *Structure* **1995**, *3*, 467–482.
33. E. D. Lowe, M. E. Noble, V. T. Skamnaki, N. G. Oikonomakos, D. J. Owen, L. N. Johnson, *EMBO J.* **1997**, *16*, 6646–6658.

34. R. I. Brinkworth, R. A. Breinl, B. Kobe, *Proc. Natl. Acad. Sci. USA* **2003**, *100,* 74–79.
35. Z. Songyang, S. Blechner, N. Hoagland, M. F. Hoekstra, H. Piwnica-Worms, L. C. Cantley, *Curr. Biol.* **1994**, *4*, 973–982.
36. M. B. Yaffe, G. G. Leparc, J. Lai, T. Obata, S. Volinia, L. C. Cantley, *Nature Biotechnol.* **2001**, *19*, 348–353.
37. T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, Y. Sakaki, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4569–4574.
38. P. Kraulis, *J. Appl. Cryst.* **1991**, *24*, 946–950.
39. E. A. Merritt, M. E. P. Murphy, *Acta Cryst.* **1994**, *D50*, 869–873.

Povzetek

Novi tehnicni dosezki in uspeh dolocanja sekvenc genomov so vzbudili nov pristop k znanstvenem raziskovanju na vsakem podrocju biokemije in molecularne bniologije, vkljucno z strukturno biologijo. Eden najpomembnejsih dosezkov v zadnjem casu je rojstvo 'strukturne genomike', svetovne iniciative, ki namerava dolociti tridimenzionalne strukture vseh reprezentativnih proteinov. A strukturna biologija se lahko nadeja zivahne prihodnosti, ki se ne bo ustavila s strukturno genomiko; ce hocemo razumeti, kako deluje proteom in uporabiti podatke v terapeutske namene, se bodo morale istocasno nadaljevati raziskave interakcij med proteini in makromolekulskih kompleksov, mehanizmov in regulacije funkcij makromolekul in strukture membranskih proteinov, in strukturne metode razvoja zdravil Uspesni pristopi bodo zdruzili siroko-obsezne pristope visoke zmogljivosti, razvite zaradi strukturne genomike, z bolj tradicionalnimi pristopi, ki temeljijo na specificnih hipotezah, podprte s povezujocimi orodji bioinformatike.

Omejeni viri denarnih sredstev, in omejene priloznosti sodelovanja v velikih konzorcijih, v dezeli s stevilom prebivalstva Avstralije, zahteva ustvarjalne pristope k problemom strukturne biologije. Moj clanek opisuje nekatere pristope nase raziskovalne skupine kot naprimer 'osredotoceno' strukturno genomiko prilagojeno manjsim raziskovalnim timom, in raziskave interakcij med proteini (opisane na primerih transporta v celicno jedro, in proteinskih kinaz) in regulacije proteinov (opisane na primerih transporta v celicno jedro, in hidroksilaze fenilalanina).