Protein Surface Recognition

Approaches for Drug Discovery

Edited by

ERNEST GIRALT

Department of Organic Chemistry, University of Barcelona and Institute for Research in Biomedicine, Barcelona, Spain

MARK W. PECZUH

Department of Chemistry, University of Connecticut, USA

XAVIER SALVATELLA

ICREA and Institute for Research in Biomedicine, Barcelona, Spain

WILEY

A John Wiley and Sons, Ltd., Publication

Protein Surface Recognition

Protein Surface Recognition

Approaches for Drug Discovery

Edited by

ERNEST GIRALT

Department of Organic Chemistry, University of Barcelona and Institute for Research in Biomedicine, Barcelona, Spain

MARK W. PECZUH

Department of Chemistry, University of Connecticut, USA

XAVIER SALVATELLA

ICREA and Institute for Research in Biomedicine, Barcelona, Spain

WILEY

A John Wiley and Sons, Ltd., Publication

This edition first published 2011

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data

Protein surface recognition : approaches for drug discovery / editors, Ernest Giralt, Mark Peczuh, Xavier Salvatella.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-05905-0 (cloth)
1. Drugs–Design. 2. Protein-protein interactions 3. Proteins–Inhibitors. 4. High throughput screening (Drug development) I. Giralt, Ernest. II. Peczuh, Mark. III. Salvatella, Xavier.
[DNLM: 1. Proteins–metabolism. 2. Drug Discovery–methods. 3. Enzyme Inhibitors–pharmacology. 4. Protein Binding. 5. Surface Properties. QU 55 P9689 2010]
RS420.P76 2010
615'.19–dc22

2010018781

A catalogue record for this book is available from the British Library.

ISBN 978-0-470-05905-0

Set in 10/12 pt Times by Thomson Digital, Noida, India
Printed in Singapore by Fabulous Printers Pte Ltd

Contents

Preface

By and large, current drugs fall into two broad categories: small molecules and protein therapeutics (biologics). While specific notions of 'small molecule' may vary, they can generally be characterized by their low (<1 kD) molecular weight, high functional group density, and often the presence of heterocycles as part of the core structure. As such, small molecules may be derived from, or be inspired by, natural products or they may be the product of organic synthesis. Such 'synthetic drugs' are at the origin of the pharmaceutical industry itself.

From a financial perspective, small molecules are presently the *bread and butter* of the industry with worldwide annual sales in the hundreds of billions (USD). Biologics, however, are themselves a multibillion dollar annual market and are seen by some as having a high potential for growth. The success of biologics has been mainly the result of advances in biotechnology that have facilitated the identification and subsequent expression of the appropriately tailored proteins. To be active, both small molecules and protein therapeutics must bind to a target biomolecule. It is *how* each of these types of molecules binds its partner that further differentiates them. Small molecules usually bind at an interior active site whereas proteins are involved in protein-protein interaction (PPI) that involve the exterior surfaces of proteins.

An example of each is illustrative of this point. Atorvastatin (Lipitor), a second generation statin derived from the related fungal metabolite pravastatin (Prevachol) is the number one small molecule therapeutic in the US as determined by 2009 retail sales. It is a competitive inhibitor of HMG-CoA reductase, the rate-limiting enzyme in cholesterol biosynthesis. This reductase focuses functional groups of the peptide backbone and side chains in a convergent manner inward toward the drug. In contrast, binding to the protein therapeutic Trastuzumab (Herceptin *ca* 1.3b USD/yr) by the HER2 cellular receptor covers a sizeable exterior surface area (ca. 1600 $\mathring{A}^2$). Functional group presentation by the two binding partners in the Herceptin-Her2 interaction is more divergent in nature, especially when compared to the interaction between Lipitor and HMG-CoA reductase.

A growing body of evidence suggests that a middle ground – using small molecules to bind the exterior protein surface and inhibit PPIs – can be a powerful strategy for the development of new tools for chemical biology and medicinal chemistry. A small molecule with these properties, a binder of Bcl-X_L and other anti-apoptotic Bcl proteins, is already in Phase 2a clinical trials (ABT-263); notably, the molecule is a result of a fragment-based drug discovery effort. It binds the Bcl proteins in a groove usually occupied by an α-helix of the native protein binding partners. Molecules such as ABT-263 can combine the mode of action of protein therapeutics with the synthetic accessibility, ease of administration, bioavailability, and robustness of traditional small molecule drugs. The development of protein surface binders by structure-based drug design or similar approaches should be contrasted with the discovery of small molecules that inhibit a given PPI by an allosteric

mechanism, which have largely arisen via serendipity and are testament to our still crude understanding of the physico-chemical principles that govern the interactions between proteins.

In between the contrasting worlds of small molecules and therapeutic proteins, peptides are likely to play an important role as therapeutic agents that modulate protein-protein interactions. Compared to protein therapeutics, the most important advantages offered by small molecules are their relatively straightforward synthetic accessibility and bioavailability. Protein therapeutics, by contrast, have specificity as a key asset, which arises from their ability to establish a large number of noncovalent interactions with the surface of the target. Nowadays, peptides combine the advantages of therapeutic proteins with those of small molecules due, among other developments, to recent progress in their modification to improve their bioavailability profile.

This book provides both a context and a guidepost for the development of molecules that alter protein function by inhibiting protein-protein interactions (PPIs) as opposed to conventional active site inhibition. The subject material has been broken into four broad sections: principles, approaches, techniques, and case studies. The principles section provides a general description of the biophysical properties of PPIs with an emphasis on those that are relevant to drug design; in Chapter 1, 'The Discovery and Characterization of Protein–Protein Interactions', we provide an overview of the methods used for identifying and characterizing PPIs, a survey of the main structural and dynamical properties of protein-protein complexes and a discussion of the challenges and opportunities inherent to inhibiting their formations whereas in Chapter 2, 'Biophysics of Protein–Protein Interactions', we provide, instead, a detailed account of the noncovalent interactions that provide the driving force for complex formation and of the thermodynamics and kinetics of the process.

Following this overview, the approaches section reviews established strategies for the inhibition of PPIs in terms of the small molecule inhibitor. Chapter 3, 'On the Logic of Natural Product Binding in Protein–Protein Interactivity' presents a rationale on how natural products bind protein surfaces and the functional consequences of the interactions. Chapters 4 and 5, 'Interface Peptides', and 'Inhibition of Protein–Protein Interactions by Peptide Mimics', detail the progression of a strategy whereby peptide sequences from the protein-protein interface are used as inhibitors and then subsequently serve as models for the development of peptide mimics with the same activity. Secondary structures such as turns and α-helices are common in the collection of interface peptides. As such, mimicry of these elements has received significant attention. Chapter 6, 'Discovery of Inhibitors of Protein–Protein Interactions by Screening Chemical Libraries', collects examples of small molecule inhibitors of protein–protein interactions that have come about via screening efforts.

A review of technologies that enable the evaluation of protein surface binding constitutes the next section of the book. This details aspects that range from organic synthesis to screening methods. Chapter 7 'High-throughput Methods of Chemical Synthesis Applied to the Preparation of Inhibitors of Protein-Protein Interactions', describes methods for the preparation of small molecule inhibitors of PPIs and the strategies behind their synthesis. Chapters 8, 'In Silico Screening', and 9.1 '*In Vitro* Screening: Screening by Nuclear Magnetic Resonance', provide accounts of how computational tools and Nuclear Magnetic Resonance can provide key information, while aspects of high throughput screening in terms

of *in vitro* and cell-based assays are outlined in Chapter 9.2, '*In Vitro* Screening: Methods of High-throughput Screening'.

Finally, the integration of the previous concepts is illustrated through two case studies in the final section of the book. These case studies include 'Inhibitors of the MDM2-p53 Protein–Protein Interaction' (Chapter 10) and 'The Discovery of Potent LFA-1 Antagonists' (Chapter 11).

We trust that the readers of the book will find it a source of valuable information in addressing the challenges and potential rewards associated with the inhibition of protein-protein interactions and we wish to thank all our co-workers and co-authors for their enthusiastic contributions in making the book possible. We specifically would like to thank Brendan Orner and David Bolstad for very valuable discussions and Paul Deards for initiating the project.

Ernest Giralt, Mark W. Peczuh and Xavier Salvatella

List of Contributors

Xavier Barril, ICREA and University of Barcelona, Barcelona, Spain

Jorge Becerril, Department of Chemistry, Yale University, New Haven, CT, USA

Denzil Bernard, Comprehensive Cancer Center and Departments of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

C. W. Bertoncini, Department of Chemistry, University of Cambridge, UK

Richard T. Desmond, Department of Chemistry, University of Connecticut, Storrs, CT, USA

Annaliese K. Franz, Department of Chemistry, University of California at Davis, Davis, CA, USA

Tom Gadek, SARcode Corporation, San Francisco, CA, USA

Carlos García-Echeverría, Novartis Institutes for Biomedical Research, Basel, Switzerland

Ernest Giralt, Department of Organic Chemistry, University of Barcelona and Institute for Research in Biomedicine, Barcelona, Spain

Andrew D. Hamilton, Department of Chemistry, Yale University, New Haven, CT, USA

A. Higueruelo, Department of Biochemistry, University of Cambridge, UK

James J. La Clair, Xenobe Research Institute, San Diego, CA, USA

Qing Lin, Department of Chemistry, State University of New York at Buffalo, Buffalo, NY, USA

Francisco Javier Luque, Department of Physical Chemistry, Faculty of Pharmacy, University of Barcelona, Barcelona, Spain

Irene Luque, Department of Physical Chemistry and Institute of Biotechnology, Faculty of Sciences, University of Granada, Granada, Spain

Mark W. Peczuh, Department of Chemistry, University of Connecticut, Storrs, CT, USA

Johanna M. Rodriguez, Department of Chemistry, Yale University, New Haven, CT, USA

Xavier Salvatella, ICREA and Institute for Research in Biomedicine, Barcelona, Spain

Sanjeev Shangary, Comprehensive Cancer Center and Departments of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

Jared T. Shaw, Department of Chemistry, University of California at Davis, Davis, CA, USA

Wenjiao Song, Department of Chemistry, State University of New York at Buffalo, Buffalo, NY, USA

Yuchen Tang, Department of Chemistry, University of California at Davis, Davis, CA, USA

Shaomeng Wang, Comprehensive Cancer Center and Departments of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

Pauline N. Wyrembak, Department of Chemistry, Yale University, New Haven, CT, USA

Part I
Principles

1

The Discovery and Characterization of Protein–Protein Interactions

C. W. Bertoncini[1], A. Higueruelo[2] and X. Salvatella[3]
[1]Department of Chemistry, University of Cambridge, UK
[2]Department of Biochemistry, University of Cambridge, UK
[3]ICREA and Institute for Research in Biomedicine, Barcelona, Spain

1.1 Introduction

The regulation of protein–protein interactions (PPIs) is fundamental for cellular function because PPIs are involved in virtually all biological processes. A complete and detailed description of the interaction map for proteins, known as interactome, is therefore one of the most important challenges in molecular biology, one that will provide great opportunities for therapeutic intervention in the complex diseases that challenge the biomedical community and the pharmaceutical industry. In this chapter we provide an overview of the different techniques that are currently available for the discovery and structural and thermodynamic analysis of PPIs as well as a survey of the general structural and dynamical properties of proteins and protein complexes that affect drug design. Rather than a comprehensive survey of the technical literature on methods to screen and characterize PPIs we present here a general discussion of these tools and refer the reader to the reviews and examples of application that we cite to identify the primary literature.

Protein Surface Recognition: Approaches for Drug Discovery Ernest Giralt, Mark W. Peczuh and Xavier Salvatella.

1.2 Techniques to Identify Protein–Protein Interactions

Many methods have been developed for the isolation and characterization of protein complexes, both *in vitro* and *in vivo*. Among them five methodologies are particularly suitable for high-throughput, and account for the majority of proteome-wide studies.

1.2.1 The Yeast Two Hybrid Assay (Y2H)

This system exploits the formation of a stable complex between interacting proteins to bring together two modules of a cis-acting transcriptional promoter, stimulating the expression of a reporter gene. It requires the construction of two hybrid genes, one encoding the DNA-binding domain (BD) of the transcription factor fused to a target protein (the bait) and a second encoding its transcription-activation domain (AD) fused to a different protein (the prey). If the prey and bait proteins interact through a PPI the two modules of the transcription factor (BD and AD) are brought together to reconstitute the transcription activity. Provided that the interaction between the prey and bait proteins is sufficiently strong, the now functional transcription factor will bind to the promoter sequence in the proximity of the reporter gene, via its DNA binding domain (BD), and recruit the transcriptional machinery, via its transcription-activation domain (AD, Figure 1.1A). The most commonly employed DNA-binding domains are derived from the yeast Gal4 and LexA transcription factors, while activating domains come also from Gal4 or from the viral activator VP16. Expression of the reporter gene gives the yeast a unique characteristic which allows identification of a successful PPI interaction between the bait and prey proteins. Reporter genes commonly employed are *lacZ*, that codifies for the enzyme β-galactosidase, that metabolizes X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) to give a distinctive blue color, or auxotrophic genes such as *HIS3*, *LEU2* or *URA3*, which confer positive colonies the ability to grow in media lacking specific nutrients [1].

One key advantage of the Y2H assay is its *in vivo* nature, that allows the investigation of PPIs under physiological conditions. Additional advantages of this method are its high sensitivity – it can detect very weak and therefore transient interactions, with Kd as low as 10^{-7} M – its scalibility and its easy automation. The Y2H assay can also be used in a quantitative fashion to determine the strength of the interaction between the bait and prey proteins by monitoring the amount of reporter protein produced, for example by measuring, when using the lacZ reporter gene, the β-galactosidase activity. The main disadvantage of the Y2H approach to the identification of PPIs is the number of control experiments that it requires, that are mainly aimed at determining whether the bait and prey proteins have affinity for DNA and are indeed capable of self-activating the transcription of the reporter. An additonal concern when using this approach, one that is directly linked to the its *in vivo* nature, is the possibility that a third protein mediates, in the assay, the interaction between the pray and bait proteins; it is therefore important to validate all PPIs derived from this assay by other methods, including those discussed in this chapter. Other limitations of the Y2H assay are due to its use of yeast, as some post-translational modifications are different in yeast to those in other eukaryots, and to the localization of of interactions in the nucleus, where some target PPIs may experience an incorrect cellular environment. Membrane proteins are obviously not suitable for this assay, but interactions between the cytoplasmic domains of extracellular receptors can be screened, using this approach, to study signal transduction pathways. PPIs

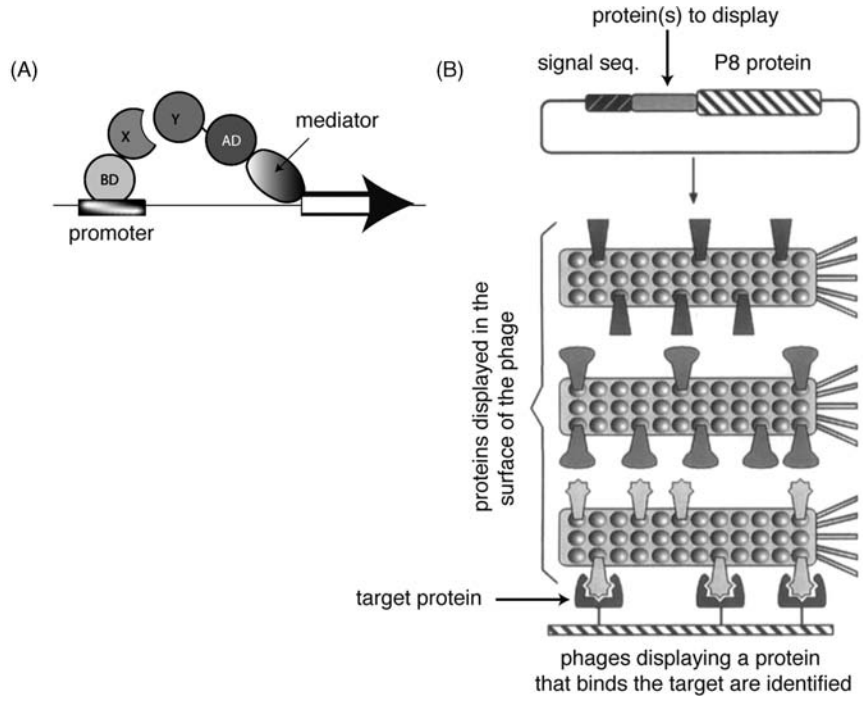

Figure 1.1 *Methods to study protein-protein interactions I. (A) Yeast two hybrid. The formation of a complex between the bait (X) and prey (Y) proteins brings together the binding domain (BD) and the activation domain (AD) of the transcription factor, which stimulates the expression of a reporter gene (adapted from Shoemaker and Panchenko [4]) (B) Phage display. The genome of a bacteria-specific virus, a phage, is engineered to carry the DNA of an exogenous protein. This protein is also displayed on the outer envelope of the phage, making the phage particle a unique carrier of both the genetic information and functional polypeptide for a given gene. Positive interaction partners are isolated by incubating the phages with the target protein immobilized on a solid support (adapted from [5] with permission from Elsevier)*

identified from genomic-scale Y2H analysis are expected to have a success rate of 50%, and bioinformatic analysis tools to refine the results with co-expression and co-localization analysis can very significantly increase the accuracy of the results [2, 3].

1.2.2 Phage Display

This method was one of the earliest tools developed for screening PPIs, before the recent spread of mass spectrometry-assisted protein identification. Phages are bacteria-specific viruses which carry the viral DNA enclosed in an envelope of viral proteins. Phage particles are therefore unique in that they contain both DNA and protein copies of a given gene in a single entity [6]. This singularity provided molecular biologist with a unique tool to isolate simultaneously both the protein displayed in the exterior of a phage particle and its DNA sequence.

The phage display technique involves the construction of a DNA library where the sequences that code for the proteins to be screened are fused to the sequence of a bacteriophage coat protein (P8 or P3) in a plasmid containing the rest of the components of the phage genome (6.5 Kbp for the filamentous bacteriophage M13). Upon infecting an *E. coli* host the phages display the chimeric proteins in their outer surface and bear inside the DNA sequences that correspond to such proteins. Phages are produced in *E. coli* individually, and the particles are then assayed for binding to the immobilized target protein in an ELISA (Enzyme-Linked Immunsorbent Assay) fashion. In order to reduce background bound phages are collected and re-amplified in *E. coli* and, after two to three rounds of binding, the DNA of phages strongly interacting with the target is isolated and sequenced, leading to the identification of the proteins interacting with the target protein [7]. A concise recollection of several random peptide and genomic libraries constructed in different phage vectors, as well as different kind of proteins successfully displayed in filamentous phages was published in 1997 by Smith and Petrenko [6]. Despite being slightly outdated, this survey highlights the range of proteins that withstand phage display that includes enzymes, hormones, receptors, cytokines and DNA binding proteins and account for more than 50 publications.

The range of target molecules that have been subjected to phage display-based screening is very wide, and is by no means restricted to polypeptides. Smith and Petrenko also collected published data on the range of target proteins that have been screened, that includes antibodies, Calmodulin, the tumor suppressor P53, Hsc70, integrins and hormones [6]. Three applications are worth mentioning in the context of PPIs:

(i) the identification of epitopes to monoclonal antibodies, by constructing naive phage peptide libraries of 10 to 40 amino acids [8];
(ii) the analysis of interfaces at the residue level by alanine scanning mutagenesis [9];
(iii) the high-throughput determination of surfaces and free energies of binding [10, 11];
(iv) the identification and construction of new scaffolds for PPIs, like single domain β-sandwich proteins (FN3 and V_HH), ankyrin repeats, WW or SH3 domains, and four helix boundels [7, 12].

Phage display is not, however, without disadvantages. An important one is, for certain applications, the need for the constructed library to be a representative sample of the whole genome, that may be challenging for some laboratories [5]; the libraries can, however, now be obtained commercially, and once produced, can be replicated by passage through a bacterial host. An additional major limitation arises when the displayed proteins fold rapidly because the chimeric fusion protein needs to be remain unfolded for efficient secretion to the bacterial periplasm, prior to display; this has however been recently overcome by the use of alternative translocation pathways and by signal recognition particles [13]. A third major potential problem concerns the immobilization of the target protein, which may hinder the interaction surface: GST or His tags are therefore desirable to aid in immobilization.

1.2.3 Protein Microarrays

A recently proposed method to analyse PPIs on a genomic scale uses functional protein microarrays [14, 15], where thousands of recombinantly expressed and purified proteins are individually spotted on a surface, by chemical derivatization, to constitute the panel of proteins to be screened (Figure 1.1B). A single fluorescently labelled protein (or ligand), is

then put in contact with the array in buffered aqueous solution, and subsequently washed with incrementing stringency. Following this the microarray slide is read by a scanner with laser excitation and fluorescence detection capabilities to identify fluorescent spots indicative of the occurrence of a PPI between the labelled ligand and a protein of the microarray. An important advantage of this method is that variable solution conditions can be easily assayed; this makes it possible, for example, to characterize binding at different concentrations in a high-throughput fashion to report on the thermodynamic stability of the PPIs detected. This approach has been successfully employed to identify and characterize proteins interacting with the Erb receptor family, where affinities using the protein microarray where comparable to those determined by surface plasmon resonance [16].

The main advantage of this technique lies on its ability to screen thousand of interactions simultaneously on a single chip [17]. However care must be taken when interpreting some interactions, in particular low affinity ones, as the chemical derivatization process may affect the properties of immobilized proteins. In addition, checks for correct expression and adequate immobilization have to be carried out; for this purpose and it is common for proteins to carry an extra peptide tag which allows identification in western blots and in the microarray slide.

1.2.4 Affinity-based Methods

A number of methods have been developed to specifically isolate protein complexes formed *in vivo* and further analyse them by mass spectrometry [18]. The main idea is to fuse the protein of interest (bait) to a peptide tag which confers affinity to a ligand immobilized on a solid support. Proteins that establish a PPI with the tagged protein can in this way be co-isolated upon incubation with the ligand matrix, and complexes can then be eluted by incubation with free ligand. Modern MS methodologies are key in this approach as they are used to analyse the bound proteins.

The general procedure involves the construction of the gene for the chimera that fuses the coding sequence of the bait to the desired tag. The plasmid is then transfected into a eukaryotic cellular host, where it is expressed, producing large amounts of the protein. Cells are then lysed, and the lysates are subjected to affinity chromatography, where protein complexes involving the tagged bait are specifically isolated. Proteins composing the complexes are resolved by polyacrilamide gel electrophoresis (SDS-PAGE), bands are excised and then subjected to tryptic digestion to produce peptides suitable to MS analysis. Such standard methodologies include the use of Matrix Assisted Laser Desorption Ionization (MALDI) MS, or liquid chromatography coupled to Electro Spray Ionization (ESI) MS [19].

This is a simple methodology that is recommended for most laboratories, as it is relatively inexpensive, requires neither complex equipment nor commercial services and can be carried out with the help of commercial kits. Affinity-based methods normally identify high affinity interactions i.e. with slow kinetics of dissociation, and one of their great advantages is that they allow the isolation of multiprotein complexes, that is not possible when using the Y2H or phage display assays or protein microarrays. It is however important to acknowledge that the use of a peptide tag can promote or impair certain PPIs, affect the normal localization of the bait protein as well as impair the isolation of the protein–protein complexes if the tag becomes buried as they form. All these problems are, however, easily overcome by the use of

a second unrelated tag in further similar experiments aimed at confirming the PPI. Depending on the nature of the tag, it is useful to classify affinity methods in three groups:

1.2.4.1 Single Tag Affinity Purification

This method involves the use of a unique peptide motif at the N- or C-terminus of the bait protein to detect protein–protein interactions that occurr *in vivo* by co-sedimentation of the interacting partners. One of the most extended methodologies involves the use of the gluthatione-S-transferase (GST) protein as a fusion of one of the assayed proteins [1, 21, 22]. The GST tag confers the bait protein high affinity to gluthatione, which is immobilized on agarose beads to *pull down* interacting proteins from cellular extracts. The disadvantages of this technique lie in the considerable size of GST (27 KDa) that can perturb the structure of the fused protein, and in the co-isolation of proteins interacting with GST itself rather than with the bait. Similar approaches employ a poly-Histidine tagged protein with high affinity to metal-chelated beads; this tag only slightly perturbs the structure of proteins, but usually results in the isolation of His-rich proteins that are false positives. Other motifs widely used as tags include maltose binding protein (MBP), immunoglobulin binding domains (protein A or G), and the Strep-tag, which is based on the high affinity biotin/streptavidin interaction [23].

1.2.4.2 Tandem Affinity Purification (TAP)

TAP is a modified version of the single tag affinity method, and involves two different peptide motifs in tandem, separated by a protease cleavage site (Figure 1.2B) [20]. The improvement in TAP in respect to the single tag method lies on the usage of two affinity purification steps which reduces the presence of spurious interacting proteins that can lead to false positives. The initial combination of tags featured a tandem of protein A and a Calmodulin binding peptide (CaMBP), separated by a Tobacco Etch Virus (TEV) protease cleavage site [24]. Protein complexes involving the tagged-bait protein are first isolated with immunoglobulin-agarose beads that have high affinity for protein A. Digestion with TEV protease releases the complex and exposes the Calmodulin Binding Peptide (CaMBP). Incubation with CaM-coated beads followed by and elution with EGTA or free CaMBP allows isolation of the purified complex. A new generation of tags involves high efficiency cloning vectors, the use of inducible promoters of expression, tetracysteine motifs suitable for in cell fluorescence imaging, and streptavidin tags [18, 25]. Proteome wide scale studies in yeast by the TAP method have recently identified more than 500 protein complexes of physiological relevance, demonstrating the high-throughput capabilities of the technique [26, 27].

1.2.4.3 Co-immunoprecipitation (Co-IP)

A protein complex stabilized by PPIs present in a cellular or tissue homogenate can be isolated by means of an appropriate antigen-antibody pair followed by affinity chromatography with protein A or G-coated beads [28]. Antibodies suitable for Co-IP studies can be raised against the protein of interest, or against a small peptide tag fused to the protein of interest; the second option is preferable since it ensures the absence of cross reactions and allows the use of already characterized commercial antibodies. Commonly employed tags for Co-IP studies are HA ($_1$YPYDVPDY^{A_9}), c-Myc ($_1$EQKLISEED$^{L_{10}}$), FLAG ($_1$DYKDDDD^{K_8}), all of which have plasmids and antibodies commercially available.

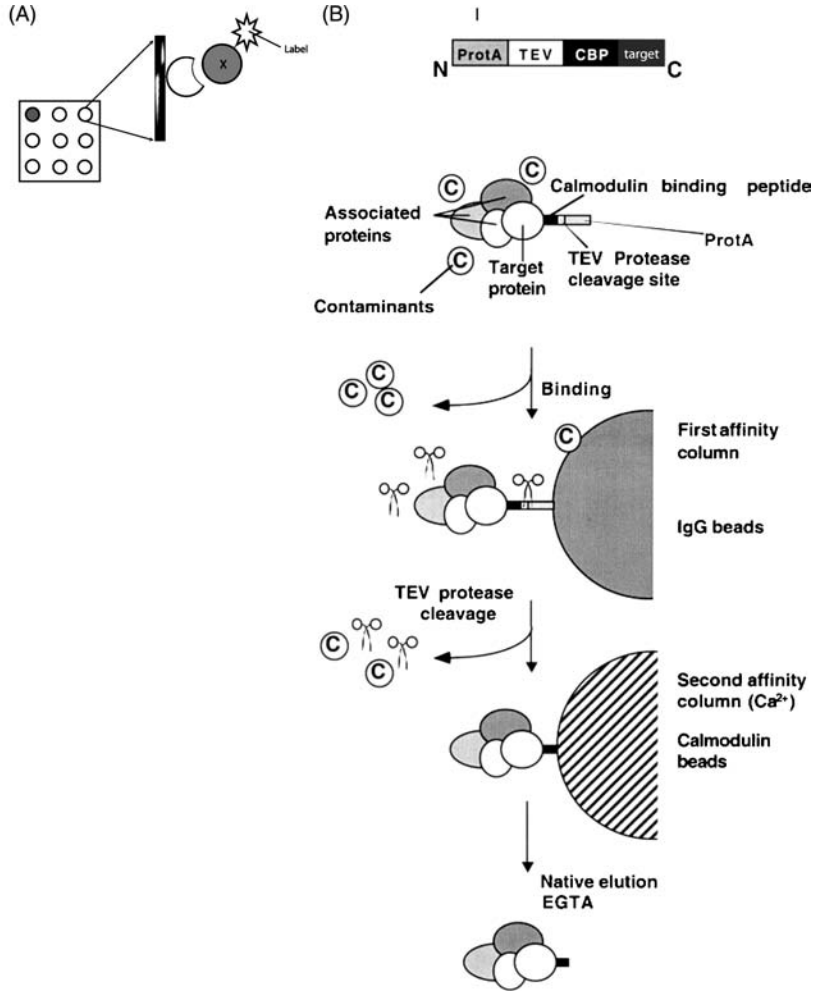

Figure 1.2 *Methods to study protein-protein interactions II (A) Protein microarrays. Proteins are immobilized on a slide by chemical derivatization, and the array of proteins is screened by a fluorescently labelled protein X (or ligand). Fluorescent spots, indicative of a protein complex, are identified by laser-based scanning (adapted from Shoemaker and Panchenko [4]) (B) Tandem affinity purification (TAP). In order to isolate interaction partners for a desired protein, the TAP method genetically fuses the protein of interest to a tandem of peptide tags. The standard TAP tag possesses a protein A (ProtA) and a Calmodulin binding peptide (CBP), linked by a Tobacho Etch Virus (TEV) protease cleavage site. Complexes with the tagged protein are purified in first instance with immunoglobulin-coated beads (affinity to protein A) followed by digestion with TEV protease. The released protein complex is subjected to a second purification step with CaM-coated beads (affinity to CBP) and eluted by addition of EGTA or free peptide (adapted from [20] with permission from Elsevier). Proteins are resolved by gel electrophoresis and identified by MALDI-MS*

1.2.5 FRET-based Detection of Protein–Protein Interactions

When two fluorophores are sufficiently close in space, a nonradiative transfer of energy termed Föster resonance energy transfer (FRET) can occur between them. The efficiency of FRET varies strictly with the sixth power of the distance between the two fluorophores (Figure 1.3A) and this provides FRET-based methodologies with the ability to efficiently assess the distance between two appropriately labelled molecules. For most biologically useful fluorophores, FRET occurs in the range of distances between 10 and 80 nm, that is the same order of magnitude as the size of macromolecules, thus certifying that FRET can be used for detecting the formation of protein complexes through PPIs. FRET has indeed long been exploited to study interactions between proteins by exploiting the intrinsic fluorescence of amino acids or by labeling reactive groups with extrinsic fluorophores [29]. However, it was not until the cloning and expression of the green fluorescent protein (GFP) from the jellyfish Aequorea victoria and its variants [30, 31] that the possibility of utilizing FRET to study protein–protein interactions *in vivo* became a reality [32, 33]. The use of GFPs is now well established for imaging protein complexes but the size of this protein (27 KDa) makes it desirable to design new tools, such as genetic tags with small dyes [34, 35], to fluorescently label proteins in the cellular environment.

The screening and identification of PPIs using fluorescent technologies has promising use in high-throughput cell analysis but FRET still has limitations, in particular concerning acquisition time and automation [36]. However recent technological advances have allowed

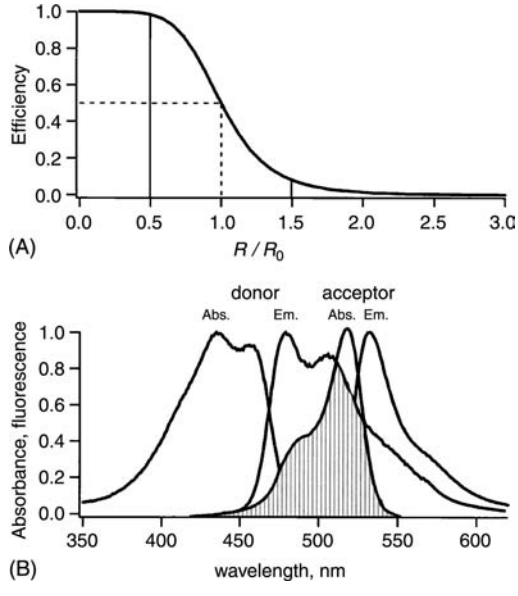

Figure 1.3 *FRET as a tool to identify protein-protein interactions (A) Forster theory shows that FRET efficiency (**E**) varies inversely with the sixth power of the distance between donor and acceptor molecules (**R**), where $\mathbf{R_0}$ is the characteristic distance at which transfer efficiency is 50%. $\mathbf{R_0}$ (nm) depends on the relative orientation between the transition dipoles of the donor and acceptor, on the spectrum overlap integral between the region of emission of the donor and the region of excitation of the acceptor (B) as well on the refractive index of the medium and the quantum yield of the donor (adapted from [39] with kind permission from Springer Science + Business Media)*

the combination of FRET with high resolution optical microscopy and flow cytometry, providing the capability of observing and screening PPIs in *E. Coli* [37]. In the near future this will be certainly applied to screening protein complexes in yeast, mammalian cells in culture and possibly in organisms such as *C. Elegans*, in particular with the use of fluorescence lifetime-based methods [38]. Published methodologies involve the creation of a library of plasmid DNA containing the protein to be studied fused to YFP and a variable gene fused to CFP. Upon transfection and protein expression, cells displaying high FRET efficiency, and hence reporting an interaction between the two labelled proteins, are sorted and isolated in a flow cytometer. Positive cells are cultured further and subjected to a second sorting step, to reduce background. Clones with high FRET are plated individually in a well plate and the plasmid DNA is isolated and sequenced to identify the identity of the interaction partners [37].

1.3 Techniques to Characterize Protein–Protein Interactions

Once a protein–protein complex is unequivocally identified, several biophysical methods can be used to characterize the macromolecular assembly, understand its properties and suggest, in a structure-based fashion, strategies to inhibit its formation. Low resolution techniques, that are key to determine the stoichiometry and overall topology of the complex, should ideally be combined with high resolution approaches that report at atomic resolution on the structure and dynamics of the binding interface as well as with methods to determine its thermodynamic stability. Since proteins often operate as multiprotein complexes it may be necessary to use multidisciplinary approaches to dissect how multiprotein complexes are formed; one recent example of such an approach has been the determination of the molecular architecture of the nuclear pore complex, a 50 MDa macromolecular machine consisting of 456 proteins. The authors of this study employed a combination of high-resolution structures, cryo-electron microscopy, mass spectrometry and analytical ultracentrifugation to determine the stoichiometry and position of each protein in the complex. This information was then used in a computational analysis to reveal the overall morphology of the nuclear pore complex, as determined by cryo-electron microscopy [40, 41].

It is not within the scope of this chapter to provide a thorough description of the biophysical methods applicable to the study of PPIs, as more specific chapters in this book will address them; we just intend to comparatively assess the information that each technique is capable to provide, consider their applications and limitations, and describe how they can be combined to provide a complete analysis of a given protein–protein complex.

1.3.1 X-ray Crystallography

It is the most widely employed method to determine structure of macromolecular complexes with atomic resolution. The method strongly depends on the quality and diffraction capabilities of the crystals obtained, but nowadays commercial kits are available for the screening of various crystallization conditions. In addition, modern crystallography experiments need to be performed at synchrotrons, where the flux and characteristics of the X-ray beams are optimal for biomolecules. When resolution permits, analysis of the crystals provides a full snapshot of the subunit-subunit contact surface, including side-chain

to side-chain contacts. In some cases the stoichiometry of the complex can be affected by crystal contacts in the lattice, it is therefore necessary to use a complementary biophysical method to confirm the stoichiometry of the protein–protein complex in solution. There appears not to be a limit in the size or nature of proteins capable of being crystallized, as the recent study showing a 340 KDa multimeric transmembrane ion channel at atomic resolution permits infer [42]. However, poorly structured polypeptides, which represent almost 30% of the genome, or flexible regions in folded proteins, which are dynamically important, are not visualized by this method and need to be characterized by other techniques such as Nuclear Magnetic Resonance.

1.3.2 Nuclear Magnetic Resonance (NMR)

This spectroscopic method can be used to provide atomic resolution models of macromolecules in solution. For the study of folded proteins NMR methodologies have a narrower range of applicability than X-ray crystallography because the NMR signals of proteins containing more than 400 amino acids are too broad to be detected efficiently. NMR spectroscopy is however an extremely powerful tool for the study of macromolecular complexes that present flexible regions and for the dynamical characterization of macromolecules. A detailed description of this technique is provided in chapter 9.1, authored by E. Giralt.

1.3.3 Isotermal Titration Calorimetry (ITC)

This method is based on the measurement of the heat absorbed or generated upon the interaction of macromolecules in solution. As heat quantities involved in binding events are considerably small, the ITC equipment relies on the accurate quantization of such heat changes. The information generated by a single ITC experiment comprises the association constant Ka (or its inverse, the dissociation constant Kd) and the stoichiometry of binding (n). In addition, thermodynamics of the binding event are also characterized, as both changes in free energy (ΔG) and enthalpy ($\Delta;H$) are measured [43]. If titrations at varying temperatures are performed a third thermodynamic variable can be determined, the heat capacity upon binding (ΔCp), which can be related to changes in solvent accessible area upon complex formation [44]. This technique as well as its potential for the thermodynamic characterization PPIs are described in detail in chapter 2, authored by I. Luque.

1.3.4 Other Techniques

From a drug discovery perspective the most important aspects of protein–protein complexes that need to be characterized are their structure at high resolution (Section 1.4.4) and thermodynamic stability as a function of sequence (Sections 1.4.4.7 and 1.5.1.1 of this chapter). A number of complementary techniques are however having an important impact in the characterization of PPIs in that they allow to determine the structure of complexes that are challenging to crystallize, such as cryo-electron microscopy and mass spectrometry [45–47], and to determine very accurately the kinetics of protein–protein binding, such as surface plasmon resonance [48].

1.4 Structure and Dynamics of Protein Complexes

Structure-based strategies for the inhibition of PPIs for therapeutic purposes rely on an accurate understanding of the structure of the binding interface and require an awareness of the general properties of protein complexes. These are, as will become evident in this section, clearly distinct from complexes formed by proteins and small organic molecules such as enzyme inhibitors.

1.4.1 Functional Classification of Protein–Protein Complexes

Functional classifications of protein–protein complexes stabilized by PPIs typically divide them in four groups [49, 50]: antibody-antigen complexes, enzyme-inhibitor complexes, electron-transfer complexes and complexes involved in signal transduction and cell cycle regulation.

1.4.1.1 Antibody-Antigen Complexes

These complexes play a key role in the immune system. The structures of antibodies contain six complementary-determining regions (CDR) that identify the antigen with high specificity; although enriched in Ser and Thr these regions are highly variable. Interaction surfaces are of medium size (1200–2000 $Å^2$) [49] and, in most cases, binding occurs with minimal conformational change in the antigen, suggesting that structural adaptation operates on the surface of the antibody.

1.4.1.2 Enzyme-Inhibitor Complexes

These complexes can be further divided in two subsets depending on their interface size, small (1200–2000 $Å^2$) or large (>2000 $Å^2$) [51]. Usually small interfaces show a single recognition patch, whereas larger interfaces present more than one recognition site.

1.4.1.3 Electron-Transfer Complexes

These complexes are transient and of low stability. They are therefore difficult to obtain in crystal form. Most electron-transfer proteins characterized until now have interfaces of between 900 and 1200 $Å^2$ [49].

1.4.1.4 Complexes Involved in Signal Transduction and Cell Cycle Regulation

These include G-proteins and protein-receptor assemblies. These complexes exhibit exquisite sensitivity to environmental changes, usually forming transient interactions and presenting low to medium affinity range (low mM to high nM) [49].

1.4.2 Differentiation Between Crystallographic and Functional Complexes

Although X-ray crystallography is an extremely powerful tool to extract high-resolution structural formation about the protein–protein interface it is important to acknowledge that crystal contacts can, due to the high concentration of samples, lead to the formation of crystallographic complexes that are present, in solution, neither *in vitro* nor *in vivo*. Discerning whether a complex is crystallographic or functional solely from knowledge of the sequence of the binding partners and, possibly, of the structure of the complex is highly

nontrivial. In an interesting study of the determinants of specificity in PPIs Janin and coworkers compared a set of specific interactions (without including short-lived assemblies or electron-transfer) against a set of nonspecific interactions; in their analysis crystal contacts were found to have a smaller interface than specific ones (with an average interface area of 570 $Å^2$), presented a lower number of hydrogen bonds per surface unit. and were less closely packed than interfaces from specific protein–protein contacts [52]. Predictive models based on the size and composition of the protein–protein interface that use this information can distinguish these two types of complexes with a certain degree of accuracy, especially when sequence conservation from related proteins is utilized.

1.4.3 Classification Based on the Nature of the Constituents and the Lifetime of the Complex

Nooren and Thornton defined three categories to classify PPIs according to these criteria [53]. The main division is that between homo-oligomeric (composed by identical chains) and hetero-oligomeric (nonidentical chains) complexes. Homo-oligomers can be further subdivided into those that are isologous, where binding interfaces are composed by the same region of each monomer, and those that are heterologous, where monomers interact through different regions. Heterologous homo-oligomers can form cyclic structures or aggregate into an endless repeated structure. Another important distinction Nooren and Thornton considered is one between complexes where the monomers can exist independently *in vivo* (nonobligate) and those where they cannot (obligate). The latter need to be denatured to dissociate, whereas the former can form from stable self-standing monomers. Examples of nonobligate complexes include antibody-antigen, enzyme-inhibitor and signal transduction complexes. Finally a third distinction can be made based on lifetime, to distinguish between permanent and transient complexes *in vivo*. Usually obligate interactions are permanent, e.g. those between monomers in homodimers, whereas transient interactions present a wide range of affinities and kinetics.

It is apparent from these definitions that interfaces of permanent complexes are more similar to the protein interior than the rest of the protein surfaces. Indeed permanent interfaces are dryer, more hydrophobic and larger than the interfaces of transient complexes [49, 54]. However, as mentioned before, the vast diversity in function, flexibility, affinity and specificity that protein assemblies present is difficult to capture in general rules.

1.4.4 Descriptors and Topology of Protein Complexes

We will here provide a brief review of the structural features typically studied in protein–protein interfaces. Classical computational characterization of interfaces include size, shape, packing, electrostatic interactions such as hydrogen bonds and salt bridges, amino acid composition, amino acid pairing preferences and hotspots.

1.4.4.1 Size

The size of the interfaces is commonly expressed as the change in the ASA (solvent-accessible surface area) between monomers and complex. For example, for a dimer, the interface size B, is $B = ASA_1 + ASA_2 - ASA_{12}$. Some authors prefer to report B/2 in spite of the fact is not exactly half B for both surfaces unless they are flat. Standard sizes for protein

complexes interfaces lie between 1200 and 2000 Å^2 averaging 23 residues in each monomer [49].

1.4.4.2 Shape

Although interacting surfaces are most often flat, they can be concave or convex. In general, the partner of smaller size shows some convexity, binding to the concave cavity in the partner of larger size. An exception to this trend are antibody-antigen complexes where the antibody antigenic site is generally convex independently of the antigen size [49]. For interfaces larger than 2000 Å^2 it has been found that the binding site is closer to the centre of mass of the protein than the average location of the surface [55].

1.4.4.3 Packing

Another structural feature of the interacting protein surfaces that is often computed is the degree packing as it is useful order to estimate the degree of steric complementarity between protomers. The most reported packing indices are Shape Complementarity score (Sc) [56] and Gap Volume index (GV) between proteins [57]. It has been found that homodimers, enzyme-inhibitor and permanent hetero-complexes are more closely packed than antibody-antigen and transient hetero-complexes.

1.4.4.4 Electrostatic Interactions

It is clear that electrostatic complementarity between partners in protein–protein complexes confers specificity. On average, there is one hydrogen bond (HB) per 200 Å^2 of interface area (B), or 100 Å^2 if one references to a single component (B/2)[58, 59]. Permanent protein complexes have typically less intermolecular hydrogen bonds per buried ASA than non-obligate complexes, 0.9 HB per 100 Å^2 in homodimers where enzyme-inhibitor complexes show 1.4 HB and antibody-antigen 1.1 per 100 Å^2 [58]. Additionally, protein–protein interfaces also have water-mediated hydrogen bonds, which present the same average distribution as the direct protein–protein hydrogen bonds, that is 10 water molecules per 1000 Å^2 (B/2) [60]. Salt bridges or hydrogen bonds involving at least one charged residue do occur: Lo Conte *et al.* found in their data set that 30% of the hydrogen bonds accounted at the interfaces are salt bridges [61]. However, almost half of the homodimeric structures analysed do not present this type of interaction. Disulfide bonds can also be found between interacting proteins but they are quite rare [58].

1.4.4.5 Amino Acid Composition

In analyzing the amino acid composition of protein–protein interfaces and their pairing preferences different studies find different frequencies due to the sets they analyse and how the interfaces were defined. Ofran and Rost, for example, divided their data set in six different types of protein–protein interfaces and, while they found some generalities, their main finding was that each interface type had distinct residue propensities [62]. For example, while Lys was underrepresented in all types of interfaces, Arg was overrepresented (Arg is common in all protein surfaces, not only interfaces). Large hydrophobic amino acids such as His, Met and Tyr were favoured in all interfaces while Ser, Ala and Gly were underrepresented. The authors corroborated previous findings that hydrophobic residues were more frequent at homo-multimers than hetero complexes; however, when they further divided into transient and obligate interaction, this distinction no longer held.

1.4.4.6 Pairing Preferences

With respect to interactions at the interface, Ofran and Rost found that hydrophobic-hydrophilic contacts were preferred at intradomain, domain-domain and transient hetero-complexes interfaces and disulfide bonds between Cys residues occurred more often than expected [62]; they also found that salt bridges were less frequent at the interfaces of homocomplexes and that interactions between identical amino acid were favored at obligate homo-complexes. More recently, Headd and colleagues, studied 135 transient hetero-complexes and found that 32% of all contacts at the interfaces involve backbone atoms [63] and, focusing on side chains, he found Glu, Ser, Asp, Lys and Arg to be the most frequent interacting side chains at the interface, each having more than 7% presence, and Met, Cys, Trp and His to be the least frequent, with less than 3.5% representation. Concerning interchain pair contacts in this data set, the most frequent occurring pairs were salt bridges (Glu-Arg, Asp-Arg, Glu-Lys and Asp-Lys); this evidence highlights the importance of electrostatic complementarity between interacting surfaces. After the charge–charge interaction the next most frequent interactions were found to be Tyr with Arg, Asn, Lys and Glu, followed by Arg with Trp and Asn.

1.4.4.7 Hot Spots

The most striking feature of protein–protein interacting surfaces is the existence of hot spots. In 1995 Clackson and Wells, using a technique called Ala scanning mutagenesis, systematically mutated to Ala all the receptor residues at the interface between the human growth hormone and its receptor and measured the free energy of binding of the resulting complex mutants [64]. In this pioneering work, the authors found that certain residues were responsible for most of the interaction free energy of the complex. A number of other experimental studies have proved that this is a common characteristic for almost all interfaces of the protein complexes [65]. Moreover, a public accessible database (ASEdb – http://nic.ucsf.edu/asedb/) holds most of the current experimental data for Ala scanning mutagenesis [66]. The accepted criterion to define a residue as a hot spot is that upon its mutation to Ala the free energy of binding increases by at least 2 $kcal.mol^{-1}$. Bogan and Thorn analysed a data set from alanine scanning mutagenesis and found all the hot spots shared common characteristics, which led them to postulate the 'O-ring' arrangement for the hot spot residues in protein–protein binding interfaces. [67] Hot spot residues are usually clustered at the centre of the interface and are surrounded by energetically neutral residues; the role of these neutral residues is to maintain the hot spots shielded from the solvent by creating a micro environment around the hotspot with lower dielectric constant, enhancing electrostatic interactions and reducing the desolvation cost of binding. It is no surprise then that the most frequent hot spots residues (Trp, Tyr and Arg) are capable of both hydrophobic and electrostatic interactions. They also found that hot spots are self-complementary across the interfaces. Nussinov and coworkers found that structural conserved residues at the interfaces of protein complexes correlate with experimental Ala scanning data and studied the organization of these computational hotspots [68]. They found hot spots are not evenly distributed in the interface, but rather they cluster together in 'hot regions'. These regions are highly packed and, within a region, hot spots form networks of cooperative interactions, whereas between hot regions the contribution to the global energy of binding is additive. They suggested that the clustering of hot spots in dense hot regions makes it easier to remove water molecules and strengthens electrostatic interactions in agreement with the O-ring arrangement. Furthermore these regions are more rigid and therefore have a relatively

low entropic penalty upon binding. In conclusion, PPIs are locally optimized in these hot regions, whereas the rest of the interface is more tolerant and lees specific, a fact that could explain the diversity in protein binding.

1.5 Protein–Protein Complexes as Therapeutic Targets

The ubiquitous nature and central role of PPIs make them very attractive targets for therapeutic intervention. However, PPIs have long been believed to be undruggable, supported by the logical assumption that a small molecule will be unable to substitute one of the partners in a multiprotein complex where the average standard interfaces are 2000 $Å^2$, with an average of 23 residues in each interacting polypeptide unit.

1.5.1 Challenging Undruggability [69, 70]

As recently summarized by Witty and Kumaravel [70] two main risks need to be assessed in the selection of therapeutic targets for a given indication. Biological risk tries to determine the probability that the modulation of the activity of the target will lead to the desired pharmacological effect whereas chemical risk tries to determine the probability that it will be possible to identify a bio-available small molecule that will effectively bind the surface of the therapeutic target and affect its function. From the biological risk point of view PPIs are very attractive targets for drug discovery, i.e. are of low biological risk, because their ever-presence in biological processes suggests that their modulation will be therapeutically relevant. In fact, in the case of extracellular targets, antibody drugs represent a key validation of this concept. In the case of PPIs the key question is therefore to assess the chemical risk for protein complexes, or in other words how likely is to find a small molecule capable of disrupting the interactions between proteins. Without considering the possibility of allosteric modulation, two experimental findings have lowered this chemical risk: the existence of energetic hot spots at the interfaces and site adaptability of surface patches.

1.5.1.1 Hot Spots

The existence of localized regions responsible for most of the binding free energy in PPIs (Section 1.4.4.7) suggests that small molecules that target the key regions of the interface can have the ability to inhibit PPIs and modulate the activity of protein complexes, i.e. decreases the chemical risk of using PPIs as therapeutic targets. An increasing number of studies report that small molecules can bind directly to protein interfaces [71, 72]; examples include the inhibition of the p53-MDM2 interaction [73], antagonists for the Bcl-2 anti-apoptotic family of proteins [74], inhibitors of ZipA-FtsZ interaction [75] and disruption of the interaction between IL-2 and its receptor IL-2Ra [76]. Alanine scanning is a relatively costly approach to the detection of hot spots in protein–protein interactions but knowledge of the structure of the complex from either X-ray crystallography or NMR spectroscopy can be exploited for this purpose by using programs that predict hot spots.

1.5.1.2 Site Dynamics

As previously described, PPIs surfaces tend to be quite flat. This represents an additional challenge for drug design because it limits the number of noncovalent interactions that can be

established between potential drugs and protein surfaces. Recent structural evidence of flexible adaptability in these regions, like in the classical example of IL-2 [77, 78], opens the prospect for more druggable protein complexes because they suggest that dynamics in the surface of the free proteins will offer conformations that are less flat and therefore more druggable that the corresponding bound structure. Indeed surface flexibility is now included in assessments of the druggability of surface patches, as shown recently [79, 80]. The availability of methods to describe proteins as dynamics ensembles rather than as rigid structures [81–83] will undoubtedly play a key role in future developments in drug discovery, especially in the discovery of compounds that target surface patches, as they explicitly incorporate flexibility in the structure determination process.

1.6 Conclusions

The technologies for the systematic screening and structural characterization of PPIs are now very mature and are providing a large number of very attractive therapeutic targets for the drug discovery and pharmaceutical industries. PPIs continue however to be considered very challenging therapeutic targets, i.e. non druggable, because the interaction surfaces involved in their stabilization are large, relatively flat and difficult to address using small molecules. As we have seen, however, two key features of PPIs, namely the uneven distribution of free energy and the dynamics of the free state, suggest that it will be possible, in the near future, to identify synthetically accessible molecules to harness the potential of inhibiting PPIs.

References

1. Singh CR, Asano K (2007) Localization and characterization of protein–protein interaction sites. *Meth Enzymol* 429:139–161.
2. Kemmeren P. *et al.* (2002) Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol Cell* 9:1133–1143.
3. von Mering C. *et al.* (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417:399–403.
4. Shoemaker BA, Panchenko AR (2007) Deciphering protein–protein interactions. Part I. Experimental techniques and databases. *PLoS Comput Biol* 3:337–344.
5. Sidhu SS, Lowman HB, Cunningham BC, Wells JA (2000) Phage display for selection of novel binding peptides. *Meth Enzymol* 328:333–363.
6. Smith GP, Petrenko VA (1997) Phage display. *Chem Rev* 97:391–410.
7. Sidhu SS, Koide S (2007) Phage display for engineering and analyzing protein interaction interfaces. *Curr Opin Struct Biol* 17:481–487.
8. Yip YL, Ward RL (1999) Epitope discovery using monoclonal antibodies and phage peptide libraries. *Comb Chem High Throughput Screen* 2:125–138.
9. Morrison KL, Weiss GA (2001) Combinatorial alanine-scanning. *Curr Opin Chem Biol* 5:302–307.
10. Kotz JD, Bond CJ, Cochran AG (2004) Phage-display as a tool for quantifying protein stability determinants. *Eur J Biochem* 271:1623–1629.
11. Pál G, Kouadio JL, Artis DR, Kossiakoff AA, Sidhu SS (2006) Comprehensive and quantitative mapping of energy landscapes for protein–protein interactions by rapid combinatorial scanning. *J Biol Chem* 281:22378–22385.
12. Magliery TJ, Regan L (2004) Combinatorial approaches to protein stability and structure. *Eur J Biochem* 271:1595–1608.