

Protein Complexes – Challenges and Opportunities for Drug Discovery

Uwe Schulte, Logopharm GmbH, Germany

Introduction

The new discipline of proteomics has begun to profoundly change the drug discovery process. Novel platform technologies and comprehensive databases now allow for accelerated and more systematic identification of targets, biomarkers and compounds. Nevertheless, several important implications are still neglected, and so far, proteomic studies have delivered large amounts of data but few results have found their way into commercial development. Above all, the high complexity of biomolecules and their interactions poses major analytical and conceptual challenges. Protein-protein-interactions have emerged as being essential for physiological function and potentially affecting pharmacological behaviour, extending the classical definition of drug targets. Which technologies are most successful in their identification? Can protein complexes be effectively modulated e.g. by small molecules? And if so, what are the anticipated risks and potential benefits for the drug discovery process?

Re-defining drug targets

The traditional gene-based view of drug targets that still governs the industry predicted a maximum of 2000-3000 target genes. Consequently, discovery focuses on a limited set of genes, and compounds are optimized for binding to individual proteins. However, the high attrition rates observed when going from *in vitro* to animal systems suggest that physiological targets are indeed different. Evidence has accumulated over the last years that proteins exert their function as part of larger assemblies like protein complexes or so-called microdomains. Modern proteomic approaches have revealed the prevalence and complexity of such assemblies in biological systems: more than 50% of proteins form stable complexes with five or more partners as demonstrated in studies on yeast. And a vastly growing amount of literature highlights the multiple implications of the modular organization of the proteome. Association of proteins not only alters their biochemical and structural properties, but may regulate the primary function, provide links to intracellular pathways ensuring functional specificity, determine trafficking and targeting, influence protein half-life and stability, and even change the pharmacological profile. Moreover, protein complex composition is often specific for cellular compartments or individual cell types. In conclusion, physiological functions are carried by protein complexes and interaction networks, and these form the actual targets for potential pharmacological intervention. This new definition of targets also implies new drug discovery strategies.

Interfering with protein-protein interactions

The idea of developing compounds acting on protein complexes still faces strong scepticism. Two of the most commonly heard concerns are

(a) “Why should one do it?” Although directing compounds against individual proteins has been successful in a lot of cases, there is a long list of examples where this strategy has failed. Among the most prominent examples are ion channels, which were regarded as a promising target class several years ago. Meanwhile, many compounds acting on pore-forming alpha subunits have been identified, but few of them have been successful in clinical studies due to lack of specificity and critical side effects. Ironically, many marketed drugs that have originally been developed against individual targets were later shown to owe their success to unanticipated selectivity for target (complex) subpopulations or effects on different side targets.

(b) “It is not possible.” Still a common notion in drug development is that small molecules can not interfere efficiently with protein-protein interactions. This is largely based on misconception of how protein interactions work. Even in constitutive complexes there is conformational flexibility of the protein interfaces. In addition, the allosteric nature of proteins implicates that ligands can induce structural changes at protein-protein interfaces (or in associated partner proteins) even when their binding occurs at distant sites. Allosteric ligands can be as effective as classical competitive (ant)agonist, and offer the advantage that their binding sites are not sterically restricted. Such compounds are therefore promising candidates for drugs acting on protein complexes, and a growing number of such examples is described in the literature. Of course, their targeted screening and development remains a demanding task. Modern technologies like high-content screening and cell-based assays open promising perspectives.

Resolving the challenge of complexity

It is estimated that up to 250.000 protein products are encoded by our genome, and even if only a minor portion is expressed at relevant levels in any type of cell, the number of potential interactions and assemblies is beyond imagination. Thus, their systematic identification is a major technical challenge. Two distinct strategies are currently used:

(a) Recombinant methods like yeast-two hybrid screens or co-purification analysis of tagged proteins offer the advantage of standardized procedures and high throughput. Both are based on expression of gene construct libraries in yeast (or other cell systems), but even when applied to the proteome of this rather simple organism error rates are quite high. More than 50% of reported interactions are estimated to be false-positive, and the majority - especially those involving membrane proteins - remains undetected. *In vitro* binding assays like protein arrays use recombinant proteins and can also achieve very high throughput. However, they are restricted to soluble proteins or domains and prone to artefacts.

(b) Native source based approaches couple biochemical fractionation or isolation of protein assemblies with mass spectrometric identification. Preparation of source material can be tedious and sample amounts become a limiting factor. The biochemical method of choice is affinity-based purification of native protein complexes. Strong and selective enrichment of target protein(s) considerably reduces complexity and at the same time allows the detection of very low abundant proteins. High-affinity ligands like specific antibodies are required as well as optimization of conditions that include membrane proteins and adequate controls. In addition, the high sensitivity and reliability of mass spectrometry needs to be matched with more advanced bioinformatics methods. These shortcomings have so far prevented large-scale application and standardization of this approach. However, when carried out properly, it can deliver highly reliable and comprehensive results.

It is evident that we are still far from understanding the complex protein interaction networks in living organisms. Current studies deliver snapshots that often show little overlap and - a key problem - lack independent verification, for example by functional experiments. Targeted approaches like affinity-based purification from native tissue can provide the most reliable results, and major impacts can be expected from future advances in related technologies. Whatever methodology is used for studying protein-protein interactions, drug discovery and development can benefit in several ways:

- Target validation and subunit composition

The search for novel drug targets often starts with genes identified in patients or disease models. Detailed information on tissue distribution, regulatory and functional mechanisms, cell-type specificity and molecular structure of the corresponding proteins is required for validation but often not readily accessible. A third of our genome is still functionally unassigned, and about the same fraction is annotated only on a preliminary basis. Thus,

identification of associated proteins can provide direct access to molecular pathways and function and complement genetic knockout strategies that are often inconclusive due to lethality or compensatory effects. Another aspect is that several proteins require assembly with other proteins to be functional: Associated subunits may be mandatory for proper folding, stability, trafficking or biological activity. Prominent examples are found among G-protein coupled receptors and ion channels. Finally, even when interaction partners have been generally described, the actual subunit composition in the cell type or tissue of interest is often unknown. Identification of cell-type specific assemblies provides a basis for developing more specific drugs.

- **Co-targets and modulators**

Associated proteins can be highly specific for a given target or confer regulatory mechanisms that are relevant under pathological conditions. Such co-targets offer the chance to develop compounds with new functional properties, like conditional modulators or blockers/mimics of specific regulatory input pathways. Resulting drugs may have improved selectivity and safety profiles. The importance of another type of co-targets has become evident in attempts to interfere with intracellular signalling. Many of the underlying pathways contain redundancies, branchings or convergencies that may bypass the effect of drugs. Thus, proteomic analysis of protein interaction cascades helps to identify key signalling mediators and biological backup “co”-targets, thereby improving pharmacological and therapeutic strategies, as for example in the treatment of cancer. Finally, co-targets can offer alternative access to the development of marketable drugs in cases where the original target is protected by competitors’ patents.

- **Structural design**

Protein 3D structures elucidated by crystallography or nuclear magnetic resonance spectroscopy draw a rather static picture of proteins biased towards independently folding domains. So-called “unstructured regions” - that actually represent a major portion of proteins - have been largely neglected. Studies of protein complexes have shown that unordered domains play an important role in protein assembly and often adopt structures within higher order complexes. Although it appears technically challenging, more meaningful and novel target structures could be obtained by analysis of native protein complexes rather than of individual subunits or sub-domains. Furthermore, structure-based design and screening focuses on known ligand binding sites since much less is known about conformational dynamics and sites for modulatory input. Consequently, allosteric compounds are rather found by chance. Proteomic analysis of target complexes can provide clues for regulatory domains or allosteric sites, for example by revealing molecular determinants for functional interactions or protein modifications such as phosphorylation sites.

- **Screening assays**

Cellular assays provide the opportunity to study the target protein in a cellular context and using multiple readouts. Main challenges are to reconstitute the physiological target phenotype and to establish meaningful and readily detectable readout signals. One possibility is to select a cell line that is closely related to the native target cell and provides a comparable protein background, but this requires additional validation and is not always feasible. Alternatively, established cell lines with known background can be stably transfected with a target gene construct. Expression levels, stability and functionality of the target may be improved by co-transfection of accessory subunits. In addition, native complexes provide insight into potential effector pathways that can be used as readouts. Ultimately, differential screening of cell lines expressing different target subunit compositions would provide a means for systematic development of complex-specific compounds.