Transient Protein-Protein Interactions: Structural, Functional and Network Properties

Author list: James Richard Perkins[1*] Ilhem Diboun[1*] Benoit Hugues Dessailly[1] Jon Lees[1] Christine Orengo[1]

[1] Department of Structural and Molecular Biology, University College of London, Gower Street, WC1E 6BT London, United Kingdom

[*] Contributed equally to this work

Abstract

Transient interactions, which involve protein interactions that are formed and broken easily, are important in many aspects of cellular functions. Here we describe structural and functional properties of transient interactions between globular domains and between globular domains and short peptides from disordered protein regions. The importance of post translational modifications in transient interactions is also considered. We review techniques used in the detection of the different types of transient protein-protein interactions. We also look at the role of transient interactions within protein-protein interaction networks and consider their contribution to different aspects of these networks.

Introduction

Interactions between proteins play an essential role in the proper function of living cells. Different types of protein-protein interactions have been defined in the literature (Nooren and Thornton, 2003a) and these definitions include distinctions between obligate and non-obligate complexes, where the former cover complexes of protomers

that cannot exist independently as opposed to non-obligate complexes. Another distinction can be made between permanent and transient complexes. Permanent protein-protein interactions (PPI) are strong and irreversible, whereas a complex qualifies as transient if it readily undergoes changes in oligomeric state.

Transient complexes can be further subdivided into weak and strong. Weak transient complexes show a dynamic mixture of different oligomeric states in vivo, whereas strong transient complexes change their quaternary state only when triggered, for example by ligand-binding. Weak transient interactions are characterised by a dissociation constant Kd from the micromolar range ($> 10^{-4}$ M) and lifetimes of seconds. Strong transient interactions, stabilised by binding of an effector molecule, may last longer and have a lower Kd in the nanomolar range ($< 10^{-6}$ M) (Nooren and Thornton; 2005). Obligate interactions are usually permanent whilst non-obligate interactions can be transient or permanent (Nooren and Thornton, 2003a).

This classification scheme, also summarised in Figure 1, can prove very useful but these distinctions are not entirely clear-cut and there is a continuum between the different types of interactions.

Structural properties of transient interactions between globular domains

Structural data has been exploited to characterise and distinguish different types of proteins interaction (Lo et al., 1999; Nooren and Thornton, 2003a; Nooren and Thornton, 2003b; Chakrabarti and Janin, 2002; Mintseris and Weng, 2003; Ansari and Helms, 2005). These indicate that transiently interacting proteins have interfaces that are smaller in size than permanent interfaces and have amino acid compositions that are

not drastically different from the rest of the protein surface - the proportion of hydrophobic residues is the same as the rest of the surface but interfaces are slightly richer in neutral polar groups. These transient interfaces also tend to be quite compact, comprising a central core that is fully buried during interaction, surrounded by a rim that presents features more similar to the rest of the surface than the interface core. Transient interfaces also tend to be rich in waters.

Many proteins undergo large conformational changes upon transient interaction, with the most extreme cases involving disorder-to-order transitions, as discussed in a recent review (Janin et al., 2008). Furthermore, it has also been observed that strong transient dimers tend to have larger, less planar and sometimes more hydrophobic interfaces than weak transient dimers, and that they often undergo more extensive conformational changes upon interaction (Nooren and Thornton, 2003a). Interfaces from weak transient complexes can also be recognised on the basis that their residues are more conserved than residues on the rest of the surface (Nooren and Thornton, 2003b). Finally, a recent study comparing transient (or weak) and permanent (or strong) homodimers from the PDB suggests that the interfaces of the weak homodimers are loosely packed and that this property may contribute to their lower stability (Dey et al., 2010).

When comparing the sequences of proteins involved in transient and obligate complexes of known structures, it also appears that transient interfaces evolve much faster, and therefore show greater sequence plasticity and smaller evidence of correlated mutations across the interface because the protomers have less time to adjust to changes in the partner's interface (Mintseris and Weng, 2005).

Based on the characterisation of transient interfaces and their specific features, a number of structure-based methods are being developed with the aim of predicting interactions and interfaces between proteins, and in some cases, to look specifically for transient complexes. Machine-learning algorithms have been shown to have high potential in discriminating permanent and transient PPIs (Liu et al., 2010). A large number of state–of-the-art approaches in protein –protein interface (not necessarily only transient) prediction have been recently reviewed (Ezkurdia et., 2009).

Also, a link has been found between the stability of residues in the unbound state and whether they are part of a transient interface or not, thus suggesting possible new approaches for the prediction of interfaces from unbound protein structures (Bonet et al., 2006). Furthermore, a group recently showed that structural changes are very frequent upon complex formation, by using an approach based on a structural alphabet representing possible backbone conformations, and the CAPRI benchmark dataset of proteins for which structures are available both in unbound and bound forms, *i.e.* generally transient interactions (Martin et al., 2008). This supports the importance of the induced-fit model of protein-protein interaction according to which proteins undergo conformational changes upon interaction. This model is also popular with enzymes where a conformational change helps stabilise their initial weak contact with ligands.

Importance of structural disorder in transient interfaces

It has been shown recently that a significant number of interfaces in the PDB can be characterised as protruding or interwound, whereby a segment of the partner protrudes deep inside the other partner (Yura and Hayward, 2009). This phenomenon relates closely to the notion of disorder as it has been observed that the regions that are interwound in interfaces are often more likely to be disordered in the unbound state.

Intrinsic disorder is thought to confer structural flexibility or 'fuzziness' to protein complexes (Tompa and Fuxreiter, 2008). Tompa and colleagues defined two levels of structural fuzziness: static and dynamic. With static fuzziness, a disordered protein or segment adopts one or more stable ordered conformations resulting in structural polymorphism of the complex. However, a protein or segment may remain disordered in the bound state, though shifting between a number of different conformations in rapid equilibrium (Tompa and Fuxreiter, 2008). Dynamic fuzziness is illustrated by scaffold proteins and disordered segments serving as flexible linkers of two or more binding regions.

Intrinsic disorder has been largely implicated in transient binding interactions. With disorder at the binding interface, there is often a disorder-to-order transition upon binding that is associated with a decrease in conformational entropy, causing the interaction to be low affinity (Singh et al., 2007). Interactions involving proteins or regions that remain largely disordered in the bound state often feature few molecular contacts (see later discussion on motifs) and are thought to be largely transient (Tompa and Fuxreiter, 2008). Furthermore, disorder was found to be enriched amongst

functions that have a temporal nature, as indicated by the observation that long regions of disorder are found in 66% of signalling proteins (Iakoucheva et al., 2002)..

Increased recognition of the importance of structural disorder in proteins has led to the development of a large number of bioinformatics tools to predict disordered regions from protein sequence (Dosztanyi et al., 2010). More details on existing tools for the study of protein disorder are shown in Table 1, and the user is directed to more recent reviews (Dosztanyi et al., 2010; He et al., 2009). Disorder predictors have been used in order to predict protein regions involved in interactions with other proteins (Oldfield et al., 2005), with the recent web server, ANCHOR (Dosztanyi et al., 2009) combining disorder prediction with prediction of binding regions in its output.

Transient Interactions between disordered and ordered regions mediated by Linear Motifs

Linear Motifs (LMs) or SLiMs have attracted a lot of attention in recent years (Ren et al., 2008). LMs are short (generally 2-8 residues in length) conserved amino acid sequences which can interact with globular domains from the same and/or other proteins (Diella et al., 2008). They may account for up to 15-40% of interactions (Petsalaki and Russell, 2008; Neduva et al., 2005). Though usually found in intrinsically disordered regions, LMs can also be found in exposed flexible loops within globular domains where they do not interfere either structurally or functionally with the domain in which they are found (Gould et al., 2010). Examples of LMs occurring in the protein p53 are shown in Figure 2.

Interactions mediated by LMs have very different properties from those between two globular domains as they bind with much smaller interface areas (Stein et al., 2009) and weaker affinities (Diella et al., 2008). Such properties help explain why LM-mediated interactions are frequently involved in signalling and regulatory transient interactions where the ability to form and break interactions readily is essential in order to respond quickly to cellular perturbations and changes in environment (Stein et al., 2009). Figure 2 shows a comparison of the LM(s) obtained from the ELM database (Gould et al., 2010) and ANCHOR binding region predictions for the tumour suppressor protein p53 protein.

Weak LM mediated interactions are often stabilised by a mechanism of cooperativity in large dynamic complexes (an example is given in Figure 5). This is particularily important with regulatory processes to ensure that a small starting signal is not lost as might happen if messenger molecules had to diffuse through the cell unaided, binding only with low affinity when they meet their intended targets (Gibson, 2009). The stabilisation of LM mediated interactions in complex structures under physiological conditions is hidden in databases such as the PDB, that show bias to binary interaction data.

Although weak in nature, LM mediated interactions show high specificity with their binding partners. In part, this is conferred by cellular location eliminating certain interactions through the absence of partners involved (Stein et al., 2009). At a finer level, specificity is determined by regions adjacent to LMs. Stein and Aloy analysed

the relative contributions of LMs and their flanking regions to the interaction binding energy, and found that flanking regions contribute 21% of the binding energy on average (Stein and Aloy, 2008). Further evidence to support the significance of the flanking regions is the observation that they are often structurally conserved (Chica et al., 2009).

Although an LM might not form a tertiary structure on its own, it will usually fold into a stable structure (along with its flanking regions) once bound to a partner domain. The different types of structures adopted by peptide sequences upon binding have been classified into categories (Mohan et al., 2006): those forming α-helices, β-strands (see also (Remaut and Waksman, 2006)), and a last class forming irregular structure elements (see (London et al., 2010) for a more recent analysis). Some example transient peptide-protein interactions are shown in Figure 3.

Molecular recognition features (MoRFs) are another type of protein-protein interaction-mediating peptide (Oldfield et al., 2005; Oldfield et al., 2005; Mohan et al., 2006). They are longer than LMs (10-70 residues) and are always located within disordered regions. MoRFs are able to mediate the interaction between two proteins, and undergo a disorder-to-order transition upon binding (Mohan et al., 2006). MoRFs often overlap with LMs and their flanking sequences but that is not always the case (Fuxreiter et al., 2007; Dunker et al., 2008; Ren et al., 2008).

Multispecificity and the role of post translational modifications (PTM)

Several LMs may co-exist together within a given disordered region and it is not unusual for them to overlap. The tumour suppressor protein p53 provides a well

characterised example, for which available structures show that the same disordered region binds to different interacting partners. Interactions with two such partners, i.e. SIR2 and Cyclin A2, are shown in Figure 4. Both these proteins bind to the same natively unstructured region in the C-terminal region of p53. This region has been predicted to contain ELMs, and several proteins are predicted to bind to it. These interactions are therefore mutually exclusive, allowing p53 to perform different roles depending on which of its partners are present in its environment (more examples in Figure 3), and provides a key mechanism for the cell to switch between different processes (Neduva and Russell, 2005; Oldfield et al., 2008). This 'molecular switch', is often modulated by post translational modifications (PTMs).

Examples of PTM mediated molecular switching are numerous. A kinase might phosphorylate a tyrosine residue in a LM, thereby changing its propensity to interact with other proteins, such as those containing an SH2 domain (Liu et al., 2006). Also, with dystroglycan, where tyrosine phosphorylation of certain residues can affect its interaction partners (Moore and Winder, 2010).

The transient often reversible nature of PTMs combined with their ability to modulate the binding specificity of LMs allows them to play a major role as regulators of cellular processes. A recent computational analysis showed that the set of protein complexes involved in the cell cycle is similar for all eukaryotes, but that they differ in PTMs and transcriptional control (Jensen et al., 2006). Recent reviews provide more details on the different existing types of PTMs, and on the role of PTMs in PPIs and signalling pathways (Stein et al., 2009; Boehme and Blattner, 2009; Lin et al., 2010).

Transient interactions in protein interaction networks

Transient versus stable hubs

Transient interactions have been suggested to play a role in the promiscuity of hub proteins in PINs (Kim et al., 2006; Singh et al., 2007; Higurashi et al., 2008). In a pioneering study, Han and colleagues showed that while some hub proteins are highly co-expressed with their partners implicating aggregation in stable complexes, other hubs have partners that vary in expression suggesting that they may bind to them at different times, in a transient manner (Han et al., 2004). These two types of hubs were termed ‘party’ and ‘date’ hubs respectively.

Further understanding of the temporal nature of certain hub interactions later emerged from structural work (Tuncbag et al., 2009). Kim and colleagues classified hubs into singlish (with one or two interfaces) and multi-interface (with more than two interfaces) hub proteins and postulated that singlish interface hubs only bind their partners interchangeably, in a transient fashion (Kim et al., 2006). Another study, exploring the binding state rather than the number of interface surfaces, classified proteins as ‘sociable’ transient hubs if they are found in three or more different binding states, or as ‘non-sociable’ if they present a unique pattern of partnership corresponding to a single binding state. A sub-type of ‘non-sociable’ stable hubs adopt a unique binding state, in which they bind to many partners simultaneously (Higurashi et al., 2008).

Disorder as a mechanism for hub promiscuity

The low affinity transient interactions mediated by some hub proteins are also characterised by multi-specificity to a range of targets. This uncoupling of specificity

and affinity is a characteristic of intrinsic disorder. Thus, whilst the folding of intrinsically disordered domains on binding allows a large surface area of interaction to form providing specificity, the initial disorder to order transition is energetically unfavourable, which reduces the binding affinity (Dyson and Wright, 2005). Also, as previously explained, locally disordered segments (including those with LMs) often use post translational modification mechanisms to modulate the binding affinity, whilst specificity is conferred by the flanking regions (Stein and Aloy, 2008; Stein et al., 2009).

In addition, intrinsic disorder facilitates binding to multiple partners owing to structural polymorphism of disordered proteins/regions in the bound folded state (as previously explained). In turn, locally disordered regions often contain overlapping interfaces allowing binding to several partners exclusively (Gibson, 2009) (Figure 4).

Transient hubs tend to be enriched in disorder in comparison to stable hubs (Singh et al., 2007; Kim et al., 2008; Ekman et al., 2006). Although, a recent study failed to show this difference at the level of the interface; which could be explained by the established role of disordered regions in the molecular assembly of stable complex proteins. Further evidence supporting the role of intrinsic disorder in explaining the transient nature of certain hub interactions in PINs comes from comparative interactomics, which indicates that the increase in complexity and extensiveness of PINs along the tree of life from simple prokaryotes to unicellular and finally higher eukaryotes is concomitant with an increase in disorder content in their proteome (Ward et al., 2004). A higher level of disorder in complex eukaryotic proteomes may be associated with the emergence of transient rather than stable interactions. While the latter are often involved in core functions and tend to be evolutionary conserved, the former are able

to mediate functions that are more specific to complex systems such as fine-tuned cross-talking/regulation of various biological processes (Levy and Pereira-Leal, 2008; Haynes et al., 2006).

The role of transient interactions in elucidating the dynamics of PINs

PINs are dynamic in nature as not all interactions between proteins occur in the same time and space. This is because, in living cells, PPIs are subject to tight regulation achieved through a number of genetic, molecular and physico-spatial mechanisms including regulation of gene expression, post-translational modification (PTM), timed molecular degradation and diffusion/compartmentalisation. Many binary interactions operate together under physiological conditions to form supramolecular machines capable of mediating complex biological functions (Schmid and McMahon, 2007). PPI in large complexes can be further regulated by structural mechanisms such as allostery and cooperativity (Pardee and Reddy, 2003). All these mechanisms, as well as the ability to form transient interactions, allow dynamic assemblies of supramolecular machine complexes, tuned to specific stages of the overall mediated function.

A case which illustrates how the dynamics of protein complex machine assembly can be brought about by structural mechanisms, is the Clathrin/AP2 endocytic vesicle complex; which features a sequence of intermediate oligomeric states ultimately leading to the formation of a mature vesicle (Schmid and McMahon, 2007). The process is illustrated in Figure 5, but for now it suffices to mention that the driving force for the assembly of the vesicle is a combination of cooperativity and structural flexibilty

mediated by a number of accessory proteins that use transient interactions to help recruit various components of the complex in a timed fashion (Schmid and McMahon, 2007).

As a final note, it has been suggested that hubs that help bridge the functional modules in PINs engage mostly in dynamic (possibly transient) interactions with partners. This support the idea that transient interactions may act as a mechanism to organise and regulate the functional flow in protein networks (Komurov and White; 2007)

The importance of transient interactions to proposed models of PIN evolution

Our knowledge of the different types of PPIs has important implications for existing theories on the evolution of PINs. The duplication-divergence model (Ispolatov et al., 2005) proposes that a duplicate protein inherits some or all of the associations from its ancestor. This model may not make sense for obligate complexes as the duplication of a subunit participating in an obligate complex is not desirable since it would lead to a stoichiometric imbalance in the concentration of the individual constituents of the complex (Veitia, 2003). Such a constraint is not observed with transient homomer complexes which were shown to have evolved in higher organisms to incorporate additional copies of their constituent protomers (Nooren and Thornton, 2003a).

Further insights into difficulties in applying the duplication-divergence evolution model to stable complexes comes from the work of Kim and colleagues (Kim et al., 2006), who have found that two proteins are significantly likely to be paralogs if they use the same interface to bind a common partner. In contrast, proteins interacting at different sites with a common hub protein, as in the case of multi-interface stable

complexes, are rarely paralogs as this would have required co-evolution of the hub and the duplicate protein to form a new interface (Kim et al., 2006).

However, the duplication-divergence model draws support from structural observations which suggest that a significant proportion of protein complexes evolve via duplications of proteins involved in homomeric interactions (Pereira-Leal et al., 2007). Such duplications, followed by divergence of one or the other duplicated copy, result in heteromeric interactions of paralogous proteins. More recent studies exploiting additional experimental interaction data and protein family information confirmed this trend but failed to detect the trend in E. Coli, suggesting other specific constraints might be operating in prokaryotic organisms (Reid et al., 2010).

Experimental Detection of transient PPI

Owing to their unstable nature, weak PPIs can be technically difficult to study and our ability to identify them is lagging behind that of the more stable interactions currently detectable by a multitude of experimental techniques. The high-throughput Y2H is able, in principle, to detect binary transient interactions (Shoemaker and Panchenko, 2007a). However, as with other experimental PPI detection methods, Y2H has its own weaknesses. These include a high false positive rate and attempts to deal with this problem often result in reducing the overall ability of the method to detect transient interactions (Bruckner et al., 2009).

Despite its high false positive rate, Y2H remains the most popular method for large scale detection of PPI owing to its scalable and accessible nature (Bruckner et al., 2009).

This interest has led to technological developments which enable detection of additional types of interactions, such as those involving cytosolic, membrane-bound proteins or extracellular proteins (Bruckner et al., 2009). Other research efforts have focussed on counteracting the problem of false negatives. For instance, a recent study advised on the benefit of pooling interaction data readouts from repeated Y2H screens in order to achieve higher detection coverage (Venkatesan et al., 2009). This approach was shown to have the additional benefit of highlighting the transient interactions as these are likely to be detected by single screens in contrast to the more permanent type of interactions whose stable nature allows them to be detected repeatedly in multiple screens (Vinayagam et al., 2010).

Another recognised high throughput PPI detection method is Tandem affinity purification run in conjunction with mass spectroscopy (TAP-MS) (Collins and Choudhary, 2008). The method has been traditionally unable to detect transient interactions, which are often lost during the washing steps necessary to remove non-specific binding. Luckily, with the advent of chemical cross-linking, it has become possible to freeze transiently formed complexes by causing covalent-bond formation between interacting partners in vivo (Worthington et al., 2006). The successful integration of chemical cross-linking with TAP-MS was dependent on the optimisation of the tagging and purification steps of TAP; which proved necessary to preserve the cross link during the washing phase (Tagwerker et al., 2006; Stingl et al., 2008).

The application of TAP-MS with cross linking has been limited to small scale detection of transiently formed complexes (Tagwerker et al., 2006; Stingl et al., 2008) whilst its large scale efficacy remains to be demonstrated. Scalability remains an issue with other

technical advances for detecting transient interactions. For instance, a protocol applying single-step affinity purification combined with stable isotope labelled amino acids approach (SILAC) has been recently shown to detect the dynamic components of protein complexes (Mousson et al., 2008).

A functionally important subset of transient interactions are dependent on post-translational modification events. These interactions are often missed when PPI screening is performed in yeast. This limitation has been tackled in different ways, including the use a mammalian cell culture system, but this remains constrained by the troublesome nature of mammalian cell transfection (Bruckner et al., 2009). Amongst other methods that feature the use of endogenous cell systems for PPI screening, the Bimolecular Fluorescence Complementation (BiFC) assay is worth mentioning owing to its ability to detect transient (as well as permanent) PPI in intact cells eliminating the need for purification (Morell et al., 2007), at a potentially high throughput rate (Kerppola, 2006). The ability of various experimental techniques for identifying transient PPI is captured further in Figure 6.

Computational prediction of transient PPI

A rigorous assessment of high throughput as well as literature curated PPI data has shown that these data are both prone to error and uncomprehensive in nature (Venkatesan et al., 2009; Braun et al., 2009; Cusick et al., 2009). In this regard, computational methods that can predict novel PPI are certainly useful.

There are now many types of computational tools to predict interactions between proteins, and these include methods that rely on structural data, genomic context

information, experimental data, and methods that combine these data sources. Interaction prediction methods have been the subject of several good recent reviews (Shoemaker and Panchenko, 2007b; Skrabanek et al., 2008; Valencia and Pazos, 2008). Therefore we do not aim to review the different approaches here, but briefly consider how these methods can be enhanced by combining them and by including predictions based on experimental data. We will then briefly review methods aimed at detecting peptide-mediated interactions, as these are specifically involved in transient interactions and have been less extensively covered in the literature.

Computational tools that exploit microarray data can be used for predicting protein interactions. The underlying idea for these approaches is to look for gene co-expression across different conditions, assuming that proteins displaying similar patterns of expression are more likely to be associated functionally. Although the co-expression based approach is more suited to predicting strongly correlated stable complexes (Figure 6), the development of statistical frameworks for detecting temporal co-expression of genes in a subset of conditions has broadened the applicability of the method to predicting transient interactions (Adler et al., 2009). However, as with the genomic context methods, this approach aims at predicting functionally associated proteins that are not necessarily physically interacting (Figure 6).

Integrating different data sources can deliver significantly increased performance (Jansen et al., 2003; Hwang et al., 2005; Myers and Troyanskaya, 2007; McDowall et al., 2009), (Ranea et al., 2010 – unpublished data). Furthermore, it is clear that these integrated approaches are identifying protein associations not detected by the experimental methods (Brown and Jurisica, 2005), (Ranea et al., 2010 – unpublished

data). Only 1% of the protein associations predicted using an integrated approach were matched by experimental data whilst associations hidden from experimental studies were found to feature sets of highly connected proteins likely to be involved in signalling and regulation and therefore likely to be transient (Ranea et al., 2010 – unpublished data). However, generally these prediction methods identify protein associations not necessarily involving physical interaction (Harrington et al., 2008; Lu et al., 2007). Further more, they do not attempt to distinguish between transient and permanent types of physical interactions.

Prediction of transient peptide-mediated PPI

Current high-throughput experimental techniques for the discovery of protein-protein interactions can show significant underrepresentation of LM mediated peptide-protein interactions. For example, they account for as little as 1% of all putative interactions in some yeast-two hybrid screens, much lower than the number of interactions they are estimated to account for in vivo (Neduva and Russell, 2006).

Experimental detection of peptide-mediated protein-protein interactions is difficult for several reasons. The fact that many partners might bind to many regions of the protein being analysed makes it difficult to determine which peptide is responsible for interaction with a given partner. In addition, the bound complexes can be difficult to capture because they are involved in transient processes with low affinity interfaces (Diella et al., 2008). Finally, a general problem with any transient interaction is that the conditions under which the interaction is being detected are critical. This is especially true for some peptide-mediated interactions where post-translational modifications of

the peptide might be critical for binding to occur. In order to guide such experiments, a number of computational methods have therefore been developed to predict peptide-protein interactions, the domains involved, and the LMs they interact with.

Computational methods to predict LMs generally search for conserved stretches of residues that may be responsible for binding globular domains. However, this process is very difficult for LMs as their short length, with sometimes only a few key residues being conserved, means that there can be large numbers of false positives. Biological information, such as cellular compartmentalisation of the query protein and whether or not the part of the protein in which the motif is found is ordered, must be taken into account to address this difficulty. The reader is directed to the recent review by Diella et al. for a more in depth description of LM prediction methods (Diella et al., 2008). Alternative approaches to find peptide mediators of protein-protein interactions include the PONDR-VXLT based disorder prediction method (Cheng et al., 2007) as well as the ANCHOR methods (Dosztanyi et al., 2010). More details of these methods along with others can be found in Table 2.

A structure based method to predict potential interactions between LMs and globular domains exists, based on the work of Serrano et al., which tries to predict peptide interactors for a given globular domain (Sanchez et al., 2008). Results from such predictions can be found in the ADAN database (Encinar et al., 2009). See Table 2 for further details. Given a peptide-protein interactions, the publically available PepSite method (Petsalaki et al., 2009) predicts further details, for example binding position and orientation.

Whilst much progress has been made on the prediction of disordered regions mediated interactions, it must be stressed that most prediction methods suffer from a high rate of false positives. Some increased confidence might be gained by combining the more orthologous prediction methods, such as sequence based with disorder tendency based, but in all cases predictions must be experimentally verified to be worthy; as examplified in the study by Grigoryan and colleagues (Grigoryan et al, 2009).

CONCLUSION

Transient protein protein interactions mediate essential functional roles in biological systems, notably in regulating the dynamics of biological networks. However they are unfortunately not easy to detect experimentally. Also, it has traditionally proved rather difficult to solve their three-dimensional structure; for instance, the efficacy of X-ray crystallography is limited because of the difficulty in crystallizing weak interactions. However, recent technological improvements, particularly in the field of NMR, have resulted in more and more structures of transient protein-protein interactions being solved (Vaynberg and Qin, 2006). Also, the fact that many transient interactions are mediated by interactions between domains and small peptides suggests that NMR, which is able to solve structures of weak interactions but is limited in the size of the systems it can deal with, can provide very useful structural insights into these transient PPIs. Stabilisation of weak interactions by chemical cross-linking has also been used in combination with other approaches such as mass spectrometry to obtain insights into structures of transient PPIs (Singh et al., 2010).

Another point of interest to experimental structural biologists regarding transient interactions is the fact that such interactions often re-use the same domain-peptide interaction patterns. For example, if domain A is able to interact with two different domains B and C, but is interacting with both of them via the same peptide on their surface, solving the structure of the complex of domain A with the interacting peptide from B would be sufficient to give insights into the interaction mechanism between both A-B and A-C (see also (Bravo and Aloy, 2006)).

It is becoming clear that more recently discovered types of interactions, such as those involving disordered regions or linear motifs play an extremely important role in protein interaction networks, and that such interactions are very often transient, and involved in regulation of cell processes. These types of interactions are particularly important in higher organisms and understanding them better may have important implications, notably for our understanding of disease and for the discovery of new drug targets (Russell and Gibson, 2008; Neduva and Russell, 2006). Computational and experimental approaches are already being developed to recognise these motifs.

The emerging picture of the cell is one in which weak interactions between biomolecules are made possible by the existence of molecular scaffolds close to which biomolecules can be found in high concentrations, thereby promoting interactions despite their weak affinity (Gibson, 2009). In this context, it is likely that transient interactions will be found to play an even more important role than is currently understood.

Methods to identify transient interactions between proteins can only be improved if we continue refining our understanding of the properties of such interactions, at the structural, sequence and systems biology levels. This process will be facilitated by the accumulation of interaction data from systems biology and structural biology. Continuous development of technologies that are fine-tuned for the detection of weak protein interactions will help us gain further knowledge about the structural basis of transient interactions and will improve our ability to computationally predict them.

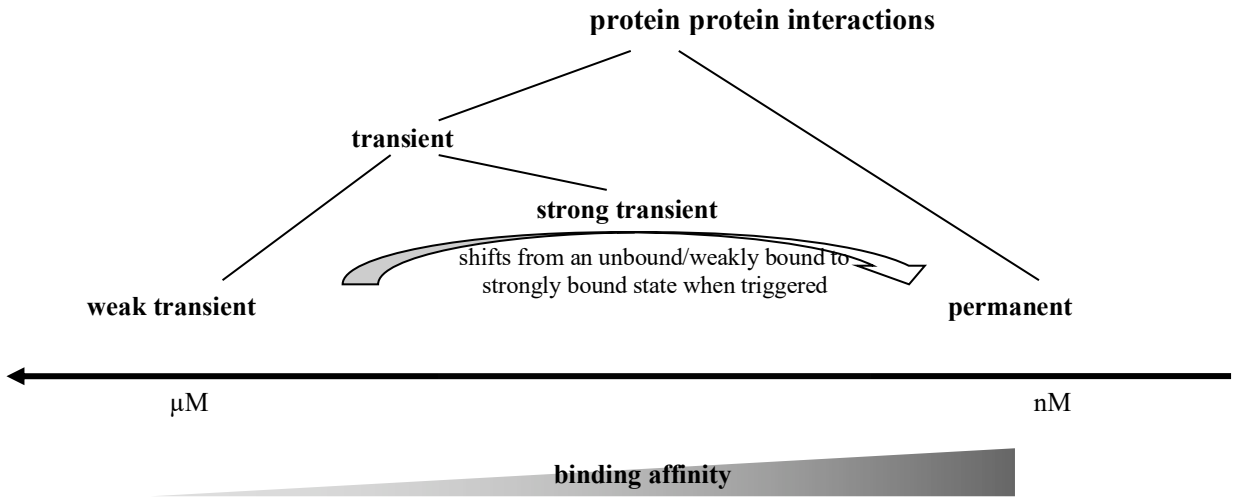


Figure 1. Illustrating the different types of protein protein interactions on the basis of varying binding affinities. The binding affinity is denoted by being disproportionally related to the diassociation constant (k_{off}/k_{on}) K_D**.** Whilst the permanent interactions feature strong binding affinities (K_D from the nM range), proteins interacting in a weakly transient manner show a fast bound-unbound equilibrium with KD values typically from the μM range. The strong transient category of interactions illustrates the continnum that exists between the weak and the more permanent interactions. This category refers to interactions that are triggered/stabilised by an effector molecule or conformational change. An example are the Ras proteins that form tight complexes with their partners when GTP-bound and only weak complexes when GDP bound.

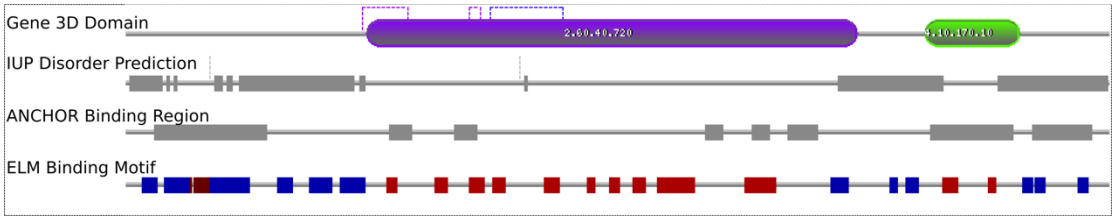
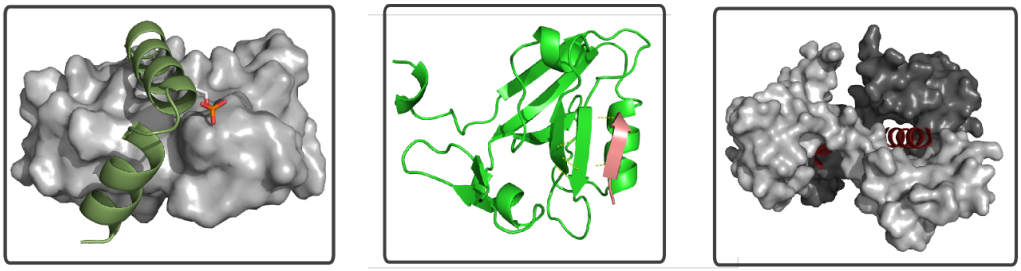


Figure 2. Potential binding sites predicted for the tumour suppressor p53. The Gene 3d Domain prediction shows the position of potential structural domains on the protein, ANCHOR Binding Regions and ELM Binding Motifs show potential binding sites for the protein. ELM predictions in red are those predicted in regions covered by a predicted structural domain in Gene3D, and therefore of lower confidence. Using the ELM resource itself might give better predictions since they use additional filters that were not used here.



(a) KIX domain of CBP with KID domain of CREB

(b) PSD-95 PDZ domain with its peptide

(c) Calcineurin - Calmodulin complex

Figure 3. Details of different types of transient protein-peptide interactions. (a) Peptide (green cartoon) from the KID domain of CREB binding to a planar part of the surface of the KIX domain of CBP (grey surface); a phosphate represented as sticks plays an important role in binding (pdb:1KDX). These figures were made using PyMol (reference). (b) Cartoon representation of the third PDZ domain from the synaptic protein PSD-95 (green cartoon) interacting with a peptide (pink cartoon). An extra beta-strand is added to a beta-sheet in the domain, in a mechanism coined beta-augmentation. H-bonds between strands in the beta-sheet are represented as dashed yellow lines (1BE9). (c) Alpha-helical peptide (red cartoon) from the alpha subunit of calcineurin binding in a deep groove at the interface between two subunits of the calmodulin homodimer (grey and dark grey surfaces) (2W73).

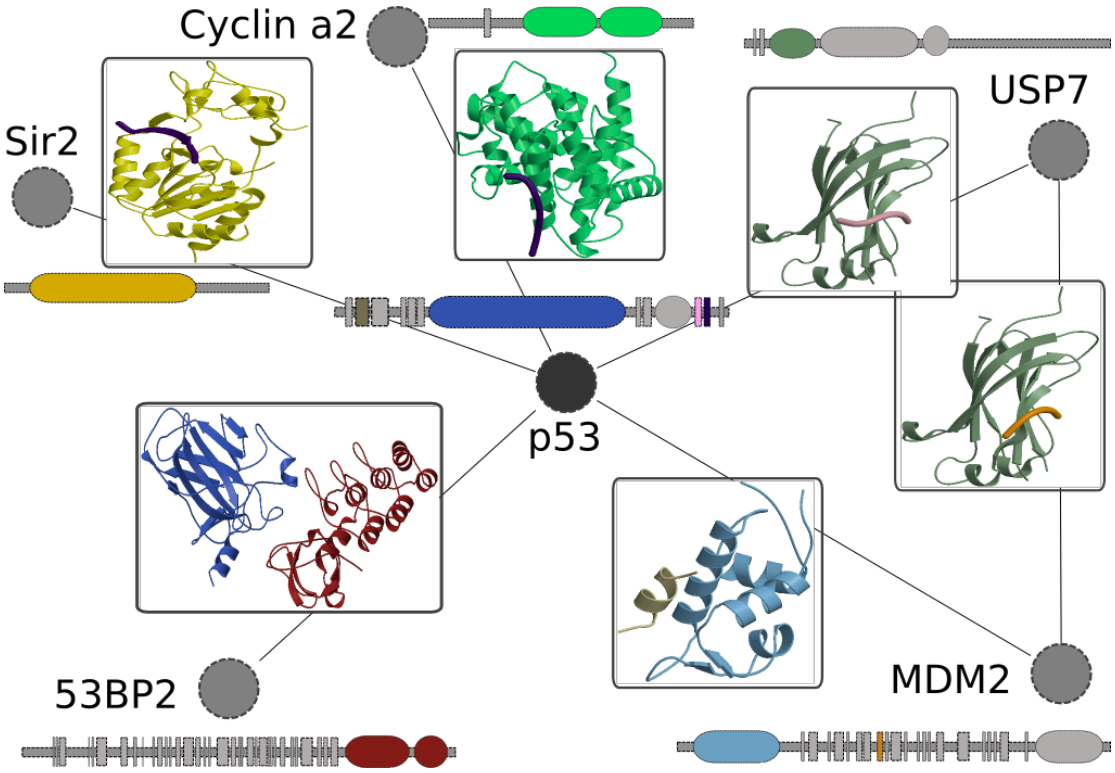


Figure 4: Different types of transient protein-protein interactions. Nodes represent different proteins, edges represent interactions between the proteins. Nodes are annotated with protein names and 2D representations of the protein showing putative domains from Gene3D or Pfam (longer, rounded edges) and putative ELM binding regions (shorter, square edges). Each edge is illustrated with a 3D structure of the complex. Colours in the structures correspond to those in the sequence, e.g. pink ELMs in the 2D representation of p53 correspond to the pink segment in the structure of the interaction between p53 and USP7. Interactions between p53 (beige peptide) and MDM2 (light-blue domain) (PDB:1YCQ), p53 (pink peptide) with USP7 (dark green domain) (2FOJ), and USP7 (dark green domain) with MDM2 (orange peptide) (2FOP) represent protein-peptide transient interactions, mediated by linear motifs predicted from ELM. USP7 interacts with peptides from p53 and MDM2 using the same

interface, suggesting these two interactions cannot occur simultaneously. More details on the biological relevance of the interactions between P53, MDM2 and USP7 are given in the text. The interaction between the core domain of p53 (blue domain) and 53BP2 (red domain) (2YCS) represents a transient domain-domain interaction, where binding of 53BP2 enhances the DNA-binding and transactivation functions of P53 on the promoters of proapoptotic genes. p53 can also bind with either Cyclin a2 (light green) (1H26) or Sir2 (yellow) (1MA3) using the same region, shown in purple.

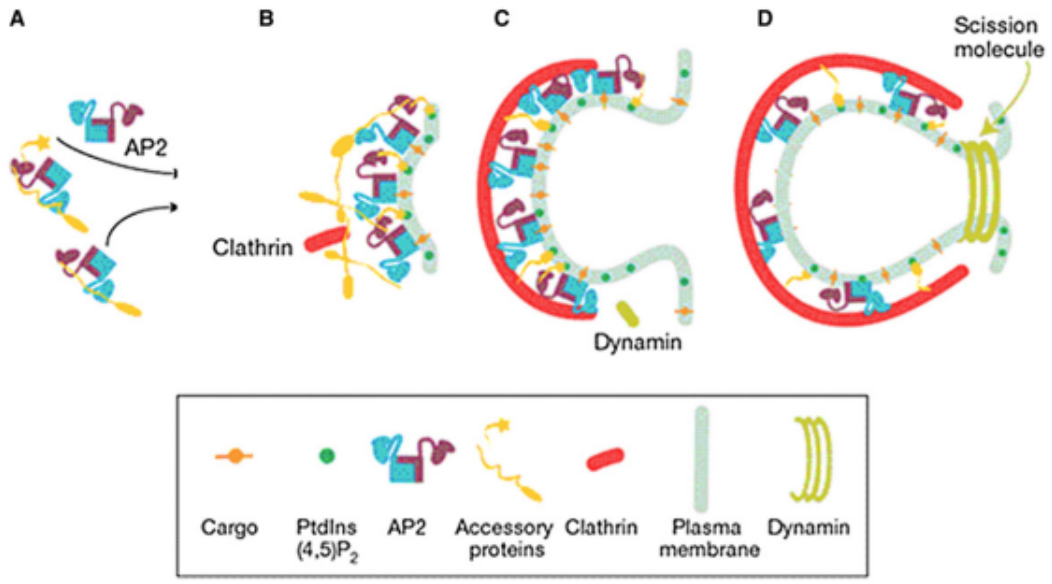


Figure 5. The Clathrin/AP2 complex. An important component of the complex is a set of accessory proteins that, in the initial phase of the vesicle formation interact simultaneously with many cargo bound AP2 molecules to help them cluster along the membrane. These accessory proteins cooperatively stabilise the resulting complex by further binding one another. Only then is Clathrin recruited to the growing complex and many of the accessory proteins change their interaction partner shifting from AP2 to Clathrin, allowing recruitment of plug-in module dynamin that drives vesicle scission. Importantly, many of the accessory proteins in question exist in low concentrations and mediate only low affinity interactions; yet they are stabilised within the complex by a

mechanism of cooperativity owing to the synergy between their many interactions with proteins AP2 or Clathrin and between themselves (Schmid and McMahon, 2007).

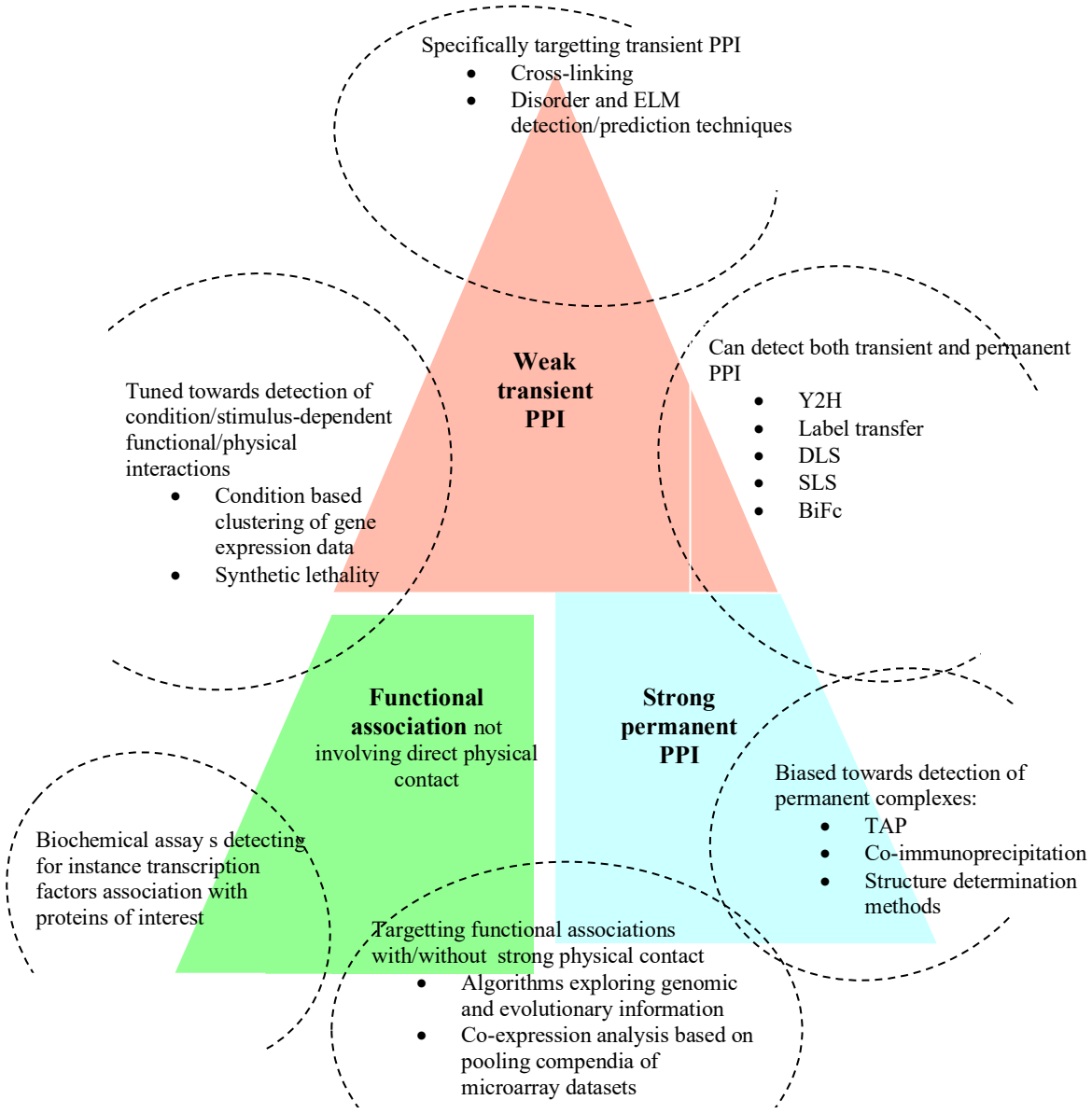


Figure 6: Categorising the different types of data captured by various PPI detection and predication methods. Single approaches can be biased and limited but together, they are able to provide a more comprehensive view of biological networks. BiFC (bimolecular fluorescence complementation), TAP (Tandom affinity purification), DLS (Dynamic light scattering), SLS (static light scattering)

Reference List

Adler,P., Kolde,R., Kull,M., Tkachenko,A., Peterson,H., Reimand,J., and Vilo,J. (2009). Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods. Genome Biol. *10*, R139.

Ansari,S. and Helms,V. (2005). Statistical analysis of predominantly transient protein-protein interfaces. Proteins *61*, 344-355.

Block,P., Paern,J., Hullermeier,E., Sanschagrin,P., Sotriffer,C.A., and Klebe,G. (2006). Physicochemical descriptors to discriminate protein-protein interactions in permanent and transient complexes selected by means of machine learning algorithms. Proteins *65*, 607-622.

Boehme,K.A. and Blattner,C. (2009). Regulation of p53--insights into a complex process. Crit Rev. Biochem. Mol. Biol. *44*, 367-392.

Bonet,J., Caltabiano,G., Khan,A.K., Johnston,M.A., Corbi,C., Gomez,A., Rovira,X., Teyra,J., and Villa-Freixa,J. (2006). The role of residue stability in transient protein-protein interactions involved in enzymatic phosphate hydrolysis. A computational study. Proteins *63*, 65-77.

Bonvin,A.M., Boelens,R., and Kaptein,R. (2005). NMR analysis of protein interactions. Curr. Opin. Chem. Biol. *9*, 501-508.

Braun,P., Tasan,M., Dreze,M., Barrios-Rodiles,M., Lemmens,I., Yu,H., Sahalie,J.M., Murray,R.R., Roncari,L., de Smet,A.S., Venkatesan,K., Rual,J.F., Vandenhaute,J., Cusick,M.E., Pawson,T., Hill,D.E., Tavernier,J., Wrana,J.L., Roth,F.P., and Vidal,M. (2009). An experimentally derived confidence score for binary protein-protein interactions. Nat. Methods *6*, 91-97.

Bravo,J. and Aloy,P. (2006). Target selection for complex structural genomics. Curr. Opin. Struct. Biol. *16*, 385-392.

Brown,K.R. and Jurisica,I. (2005). Online predicted human interaction database. Bioinformatics. *21*, 2076-2082.

Bruckner,A., Polge,C., Lentze,N., Auerbach,D., and Schlattner,U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int. J. Mol. Sci. *10*, 2763-2788.

Chakrabarti,P. and Janin,J. (2002). Dissecting protein-protein recognition sites. Proteins *47*, 334-343.

Cheng,Y., Oldfield,C.J., Meng,J., Romero,P., Uversky,V.N., and Dunker,A.K. (2007). Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry *46*, 13468-13477.

Collins,M.O. and Choudhary,J.S. (2008). Mapping multiprotein complexes by affinity purification and mass spectrometry. Curr. Opin. Biotechnol. *19*, 324-330.

Cusick,M.E., Yu,H., Smolyar,A., Venkatesan,K., Carvunis,A.R., Simonis,N., Rual,J.F., Borick,H., Braun,P., Dreze,M., Vandenhaute,J., Galli,M., Yazaki,J., Hill,D.E., Ecker,J.R., Roth,F.P., and Vidal,M. (2009). Literature-curated protein interaction datasets. Nat. Methods *6*, 39-46.

Dey,S., Pal,A., Chakrabarti,P., and Janin,J. (2010). The Subunit Interfaces of Weakly Associated Homodimeric Proteins. J. Mol. Biol.

Diella,F., Haslam,N., Chica,C., Budd,A., Michael,S., Brown,N.P., Trave,G., and Gibson,T.J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. Front Biosci. *13*, 6580-6603.

Dosztanyi,Z., Meszaros,B., and Simon,I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics. *25*, 2745-2746.

Dosztanyi,Z., Meszaros,B., and Simon,I. (2010). Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Brief. Bioinform. *11*, 225-243.

Dunker,A.K., Oldfield,C.J., Meng,J., Romero,P., Yang,J.Y., Chen,J.W., Vacic,V., Obradovic,Z., and Uversky,V.N. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. BMC. Genomics *9 Suppl 2*, S1.

Dyson,H.J. and Wright,P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197-208.

Ekman,D., Light,S., Bjorklund,A.K., and Elofsson,A. (2006). What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? Genome Biol. *7*, R45.

Encinar,J.A., Fernandez-Ballester,G., Sanchez,I.E., Hurtado-Gomez,E., Stricher,F., Beltrao,P., and Serrano,L. (2009). ADAN: a database for prediction of protein-protein interaction of modular domains mediated by linear motifs. Bioinformatics. *25*, 2418-2424.

Fuxreiter,M., Tompa,P., and Simon,I. (2007). Local structural disorder imparts plasticity on linear motifs. Bioinformatics. *23*, 950-956.

Gibson,T.J. (2009). Cell regulation: determined to signal discrete cooperation. Trends Biochem. Sci. *34*, 471-482.

Gould,C.M., Diella,F., Via,A., Puntervoll,P., Gemund,C., Chabanis-Davidson,S., Michael,S., Sayadi,A., Bryne,J.C., Chica,C., Seiler,M., Davey,N.E., Haslam,N., Weatheritt,R.J., Budd,A., Hughes,T., Pas,J., Rychlewski,L., Trave,G., Aasland,R., Helmer-Citterich,M., Linding,R., and Gibson,T.J. (2010). ELM: the status of the 2010 eukaryotic linear motif resource. Nucleic Acids Res. *38*, D167-D180.

Han,J.D., Bertin,N., Hao,T., Goldberg,D.S., Berriz,G.F., Zhang,L.V., Dupuy,D., Walhout,A.J., Cusick,M.E., Roth,F.P., and Vidal,M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88-93.

Harrington,E.D., Jensen,L.J., and Bork,P. (2008). Predicting biological networks from genomic data. FEBS Lett. *582*, 1251-1258.

Haynes,C., Oldfield,C.J., Ji,F., Klitgord,N., Cusick,M.E., Radivojac,P., Uversky,V.N., Vidal,M., and Iakoucheva,L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS. Comput. Biol. *2*, e100.

He,B., Wang,K., Liu,Y., Xue,B., Uversky,V.N., and Dunker,A.K. (2009). Predicting intrinsic disorder in proteins: an overview. Cell Res. *19*, 929-949.

Higurashi,M., Ishida,T., and Kinoshita,K. (2008). Identification of transient hub proteins and the possible structural basis for their multiple interactions. Protein Sci. *17*, 72-78.

Hwang,D., Rust,A.G., Ramsey,S., Smith,J.J., Leslie,D.M., Weston,A.D., de Atauri,P., Aitchison,J.D., Hood,L., Siegel,A.F., and Bolouri,H. (2005). A data integration methodology for systems biology. Proc. Natl. Acad. Sci. U. S. A *102*, 17296-17301.

Iakoucheva,L.M., Brown,C.J., Lawson,J.D., Obradovic,Z., and Dunker,A.K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. J. Mol. Biol. *323*, 573-584.

Ispolatov,I., Krapivsky,P.L., and Yuryev,A. (2005). Duplication-divergence model of protein interaction network. Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys. *71*, 061911.

Janin,J., Bahadur,R.P., and Chakrabarti,P. (2008). Protein-protein interaction and quaternary structure. Q. Rev. Biophys. *41*, 133-180.

Jansen,R., Yu,H., Greenbaum,D., Kluger,Y., Krogan,N.J., Chung,S., Emili,A., Snyder,M., Greenblatt,J.F., and Gerstein,M. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science *302*, 449-453.

Jensen,L.J., Jensen,T.S., de Lichtenberg,U., Brunak,S., and Bork,P. (2006). Co-evolution of transcriptional and post-translational cell-cycle regulation. Nature *443*, 594-597.

Kerppola,T.K. (2006). Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. Nat. Protoc. *1*, 1278-1286.

Kim,P.M., Lu,L.J., Xia,Y., and Gerstein,M.B. (2006). Relating three-dimensional structures to protein networks provides evolutionary insights. Science *314*, 1938-1941.

Kim,P.M., Sboner,A., Xia,Y., and Gerstein,M. (2008). The role of disorder in interaction networks: a structural analysis. Mol. Syst. Biol. *4*, 179.

Kriwacki,R.W., Hengst,L., Tennant,L., Reed,S.I., and Wright,P.E. (1996). Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc. Natl. Acad. Sci. U. S. A *93*, 11504-11509.

Levy,E.D. and Pereira-Leal,J.B. (2008). Evolution and dynamics of protein interactions and networks. Curr. Opin. Struct. Biol. *18*, 349-357.

Lin,J., Xie,Z., Zhu,H., and Qian,J. (2010). Understanding protein phosphorylation on a systems level. Brief. Funct. Genomic. Proteomic. *9*, 32-42.

Liu,B.A., Jablonowski,K., Raina,M., Arce,M., Pawson,T., and Nash,P.D. (2006). The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. Mol. Cell *22*, 851-868.

Liu,R., Jiang,W., and Zhou,Y. (2010). Identifying protein-protein interaction sites in transient complexes with temperature factor, sequence profile and accessible surface area. Amino. Acids *38*, 263-270.

Lo,C.L., Chothia,C., and Janin,J. (1999). The atomic structure of protein-protein recognition sites. J. Mol. Biol. *285*, 2177-2198.

London,N., Movshovitz-Attias,D., and Schueler-Furman,O. (2010). The structural basis of peptide-protein binding strategies. Structure. *18*, 188-199.

Lu,L.J., Sboner,A., Huang,Y.J., Lu,H.X., Gianoulis,T.A., Yip,K.Y., Kim,P.M., Montelione,G.T., and Gerstein,M.B. (2007). Comparing classical pathways and modern networks: towards the development of an edge ontology. Trends Biochem. Sci. *32*, 320-331.

Martin,J., Regad,L., Lecornet,H., and Camproux,A.C. (2008). Structural deformation upon protein-protein interaction: a structural alphabet approach. BMC. Struct. Biol. *8*, 12.

McDowall,M.D., Scott,M.S., and Barton,G.J. (2009). PIPs: human protein-protein interaction prediction database. Nucleic Acids Res. *37*, D651-D656.

Mintseris,J. and Weng,Z. (2003). Atomic contact vectors in protein-protein recognition. Proteins *53*, 629-639.

Mintseris,J. and Weng,Z. (2005). Structure, function, and evolution of transient and obligate protein-protein interactions. Proc. Natl. Acad. Sci. U. S. A *102*, 10930-10935.

Mohan,A., Oldfield,C.J., Radivojac,P., Vacic,V., Cortese,M.S., Dunker,A.K., and Uversky,V.N. (2006). Analysis of molecular recognition features (MoRFs). J. Mol. Biol. *362*, 1043-1059.

Moore,C.J. and Winder,S.J. (2010). Dystroglycan versatility in cell adhesion: a tale of multiple motifs. Cell Commun. Signal. *8*, 3.

Morell,M., Espargaro,A., Aviles,F.X., and Ventura,S. (2007). Detection of transient protein-protein interactions by bimolecular fluorescence complementation: the Abl-SH3 case. Proteomics. *7*, 1023-1036.

Mousson,F., Kolkman,A., Pijnappel,W.W., Timmers,H.T., and Heck,A.J. (2008). Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. Mol. Cell Proteomics. *7*, 845-852.

Myers,C.L. and Troyanskaya,O.G. (2007). Context-sensitive data integration and prediction of biological networks. Bioinformatics. *23*, 2322-2330.

Neduva,V., Linding,R., Su-Angrand,I., Stark,A., de Masi,F., Gibson,T.J., Lewis,J., Serrano,L., and Russell,R.B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS. Biol. *3*, e405.

Neduva,V. and Russell,R.B. (2005). Linear motifs: evolutionary interaction switches. FEBS Lett. *579*, 3342-3345.

Neduva,V. and Russell,R.B. (2006). Peptides mediating interaction networks: new leads at last. Curr. Opin. Biotechnol. *17*, 465-471.

Nooren,I.M. and Thornton,J.M. (2003a). Diversity of protein-protein interactions. EMBO J. *22*, 3486-3492.

Nooren,I.M. and Thornton,J.M. (2003b). Structural characterisation and functional significance of transient protein-protein interactions. J. Mol. Biol. *325*, 991-1018.

O'Connell,M.R., Gamsjaeger,R., and Mackay,J.P. (2009). The structural analysis of protein-protein interactions by NMR spectroscopy. Proteomics. *9*, 5224-5232.

Oldfield,C.J., Cheng,Y., Cortese,M.S., Romero,P., Uversky,V.N., and Dunker,A.K. (2005). Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry *44*, 12454-12470.

Oldfield,C.J., Meng,J., Yang,J.Y., Yang,M.Q., Uversky,V.N., and Dunker,A.K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC. Genomics *9 Suppl 1*, S1.

Pardee,A.B. and Reddy,G.P. (2003). Beginnings of feedback inhibition, allostery, and multi-protein complexes. Gene *321*, 17-23.

Pereira-Leal,J.B., Levy,E.D., Kamp,C., and Teichmann,S.A. (2007). Evolution of protein complexes by duplication of homomeric interactions. Genome Biol. *8*, R51.

Petsalaki,E. and Russell,R.B. (2008). Peptide-mediated interactions in biological systems: new discoveries and applications. Curr. Opin. Biotechnol. *19*, 344-350.

Petsalaki,E., Stark,A., Garcia-Urdiales,E., and Russell,R.B. (2009). Accurate prediction of peptide binding sites on protein surfaces. PLoS. Comput. Biol. *5*, e1000335.

Reid,A.J., Ranea,J.A., and Orengo,C.A. (2010). Comparative evolutionary analysis of protein complexes in E. coli and yeast. BMC. Genomics *11*, 79.

Remaut,H. and Waksman,G. (2006). Protein-protein interaction through beta-strand addition. Trends Biochem. Sci. *31*, 436-444.

Ren,S., Uversky,V.N., Chen,Z., Dunker,A.K., and Obradovic,Z. (2008). Short Linear Motifs recognized by SH2, SH3 and Ser/Thr Kinase domains are conserved in disordered protein regions. BMC. Genomics *9 Suppl 2*, S26.

Russell,R.B. and Gibson,T.J. (2008). A careful disorderliness in the proteome: sites for interaction and targets for future therapies. FEBS Lett. *582*, 1271-1275.

Sanchez,I.E., Beltrao,P., Stricher,F., Schymkowitz,J., Ferkinghoff-Borg,J., Rousseau,F., and Serrano,L. (2008). Genome-wide prediction of SH2 domain targets using structural information and the FoldX algorithm. PLoS. Comput. Biol. *4*, e1000052.

Schmid,E.M. and McMahon,H.T. (2007). Integrating molecular and network biology to decode endocytosis. Nature *448*, 883-888.

Shi,Y. and Wu,J. (2007). Structural basis of protein-protein interaction studied by NMR. J. Struct. Funct. Genomics *8*, 67-72.

Shoemaker,B.A. and Panchenko,A.R. (2007a). Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS. Comput. Biol. *3*, e42.

Shoemaker,B.A. and Panchenko,A.R. (2007b). Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS. Comput. Biol. *3*, e43.

Singh,G.P., Ganapathi,M., and Dash,D. (2007). Role of intrinsic disorder in transient interactions of hub proteins. Proteins *66*, 761-765.

Singh,P., Panchaud,A., and Goodlett,D.R. (2010). Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique. Anal. Chem. *82*, 2636-2642.

Skrabanek,L., Saini,H.K., Bader,G.D., and Enright,A.J. (2008). Computational prediction of protein-protein interactions. Mol. Biotechnol. *38*, 1-17.

Stein,A. and Aloy,P. (2008). Contextual specificity in peptide-mediated protein interactions. PLoS. One. *3*, e2524.

Stein,A., Pache,R.A., Bernado,P., Pons,M., and Aloy,P. (2009). Dynamic interactions of proteins in complex networks: a more structured view. FEBS J. *276*, 5390-5405.

Stingl,K., Schauer,K., Ecobichon,C., Labigne,A., Lenormand,P., Rousselle,J.C., Namane,A., and de Reuse,H. (2008). In vivo interactome of Helicobacter pylori urease revealed by tandem affinity purification. Mol. Cell Proteomics. *7*, 2429-2441.

Tagwerker,C., Flick,K., Cui,M., Guerrero,C., Dou,Y., Auer,B., Baldi,P., Huang,L., and Kaiser,P. (2006). A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. Mol. Cell Proteomics. *5*, 737-748.

Tompa,P. and Fuxreiter,M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem. Sci. *33*, 2-8.

Tuncbag,N., Kar,G., Gursoy,A., Keskin,O., and Nussinov,R. (2009). Towards inferring time dimensionality in protein-protein interaction networks by integrating structures: the p53 example. Mol. Biosyst. *5*, 1770-1778.

Valencia,A. and Pazos,F. (2008). Computational Methods to Predict Protein Interaction Partners. In Protein-protein Interactions and Networks, A.R.Panchenko and T.Przytycka, eds. Springer), pp. 67-82.

Vaynberg,J. and Qin,J. (2006). Weak protein-protein interactions as probed by NMR spectroscopy. Trends Biotechnol. *24*, 22-27.

Veitia,R.A. (2003). Nonlinear effects in macromolecular assembly and dosage sensitivity. J. Theor. Biol. *220*, 19-25.

Venkatesan,K., Rual,J.F., Vazquez,A., Stelzl,U., Lemmens,I., Hirozane-Kishikawa,T., Hao,T., Zenkner,M., Xin,X., Goh,K.I., Yildirim,M.A., Simonis,N., Heinzmann,K., Gebreab,F., Sahalie,J.M., Cevik,S., Simon,C., de Smet,A.S., Dann,E., Smolyar,A., Vinayagam,A., Yu,H., Szeto,D., Borick,H., Dricot,A., Klitgord,N., Murray,R.R., Lin,C., Lalowski,M., Timm,J., Rau,K., Boone,C., Braun,P., Cusick,M.E., Roth,F.P., Hill,D.E., Tavernier,J., Wanker,E.E., Barabasi,A.L., and Vidal,M. (2009). An empirical framework for binary interactome mapping. Nat. Methods *6*, 83-90.

Vinayagam,A., Stelzl,U., and Wanker,E.E. (2010). Repeated two-hybrid screening detects transient protein-protein interactions. Theoretical Chemistry Accounts *125*, 613-619.

Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F., and Jones,D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. *337*, 635-645.

Worthington,A.S., Rivera,H., Torpey,J.W., Alexander,M.D., and Burkart,M.D. (2006). Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. ACS Chem. Biol. *1*, 687-691.

Yura,K. and Hayward,S. (2009). The interwinding nature of protein-protein interfaces and its implication for protein complex formation. Bioinformatics. *25*, 3108-3113.

Tables

Tool Type:	Name	Details	URL

Database	Database of Protein Disorder (Disprot)	Contains details on intrinsically disordered proteins, defined as proteins with at least one experimentally determined region of disorder. Currently contains 1183 disordered regions for 517 proteins, all manually curated. Proteins are searchable and browseable, and each entry contains annotations and further details for each disordered region within the protein	http://www.disprot.org
Prediction Server	Predictor of Naturally Disordered Regions (PONDR)	Takes an amino-acid sequence as input. Uses neural-networks, trained on sequences of known disordered and ordered regions, using properties of the sequences in windows of amino-acids to generate inputs values for the predictor	http://www.pondr.com/
Prediction Server	Disopred	Takes an amino acid sequence and uses an combination of PSI-blast, neural networks and a support vector machine to classify parts of the sequence as ordered or disordered	http://bioinfadmin.cs.ucl.ac.uk/disopred/

Table 1 – Examples of tools related to predicting regions of intrinsic disorder in a protein

Focus	Name	Details	URL
Predict Interactions between peptide and domain	ADAN	Database, accessible online. Contains functionally annotated protein domain strucures involved in interactions with linear motifs, where possible alongisde their peptide ligands. In addition it performs a structural prediction of the putative ligands for a given domain using a domain if available, or homology modelling if not, and then the FOLDX algorithm, and stores the predictions.	http://adan-embl.ibmc.umh.es
Predict Interactions between peptides and domains	PepSite	Software, accessible online. Predicts potential binding areas in peptides that bind to a known, and identify the rough orientation and position of the binding sites on the protein.	http://pepsite.embl.de/
Predict interactions between peptides and domains	SMALI	Software, accessible online. Takes as input a peptide, and looks for potential domain binding partners, or takes an SH2 domain as input and looks for potential binding domains, from a user-specified list.	http://lilab.uwo.ca/SMALI.htm
Predict interactions between peptides and domains	FOLDX	Software. Has been used to predict peptide interaction partners for SH2 domains, and also is employed by ADAN to predict partner domains.	http://foldx.crg.es/
Viewing peptide-	3DID	Database, accessible online. Contains high resolution solved structures of domain-domain protein interactions, and	http://3did.irbbarcelona.org/

protein interactions		currently (as of January 28th 2010) 829 transient, peptide mediated interactions. The interactions were taken from the protein databank by taking all PDB entries containing at least two interacting proteins, and filtering these entries using information from the protein families database (Pfam) and data on known protein binding linear motifs from the Eukaryotic Linear Motif database.	
Viewing peptide-protein interactions	PEPX	Database. Accessbible online. Contains protein-peptide complexes. Browsable by several methods including CATH domain superfamily. To populate the database the PDB was mined to find 1431 protein-peptide complexes with peptides of length 35 residues or less, which were clustered using three-dimensional similarity of the protein-peptide interface.	http://pepx.switchlab.org/
Detection of disordered regions involved in interactions	ANCHOR	Uses disorder prediction methods (specifically IUPred) and identifies potential binding sites where there is a dip in predicted disorder	http://anchor.enzim.hu/
Novel motif detection	DiLiMot	Server. takes a set of protein sequence which care a common feature and looks for consensus motifs between the sequences.	http://dilimot.embl.de/

Novel motif detection	D-motif/D-star	Software (downloadable). Uses interaction data alongside sequence information to improve results & reduce false positives	http://www.comp.nus.edu.sg/~bioinfo/hugowill/DSTAR.html
Find already known motif	ELM	Database and Server. An ELM motif is a regular expression that describes an LM. The database allows the different instances of the motifs to be browsed, and links to experimental methods used to determine the instance, and also gives a quality score. When searching the database with a query set of proteins in order to predict LMs in a query protein, ELM uses a number of filtering steps to reduce false positives and ensure only biologically plausable instances are returned	http://elm.eu.org/
Find already known motif	Scansite	Database and server, for a number of domains, predicts potential interactors with domains involved in signalling, important for identifying potential transient interactors. User can search with their own motif or use predetermined ones within scansite	http://scansite.mit.edu/

Table 2 – A selection of computational tools for the prediction and analysis of interactions involving disordered regions