



Intraproteomic Networks: New Forays Into Predicting Interaction Partners

L. Aravind and Lakshminarayan M. Iyer

Genome Res. 2002 12: 1156-1158

Access the most recent version at doi:[10.1101/gr.353302](https://doi.org/10.1101/gr.353302)

References This article cites 17 articles, 9 of which can be accessed free at:
<http://genome.cshlp.org/content/12/8/1156.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center is a white box with the text "LEARN MORE". On the right is a woman wearing a red and white superhero cape and mask, with the Cellecta logo (a green molecular structure) and the word "CELLECTA" below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Intraproteomic Networks: New Forays Into Predicting Interaction Partners

L. Aravind¹ and Lakshminarayan M. Iyer

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, 20894, USA

Biological treasure troves of complete protein complements of diverse organisms (proteomes) have been unveiled in the past few years as a result of the tremendous success of genome projects. The fundamental fascination of most biochemists and molecular biologists is how the different polypeptides comprising the proteome interact to conduct “business” in various biological systems. The flood of genomic data has made large-scale attacks on this problem through computational and experimental methods very feasible. On the computational side, the main progress has been in the form of identification and classification of the individual protein domains, thereby helping to narrow down to the actual determinants of the intraproteomic interactions (Ponting et al. 2000; Lander et al. 2001). On the experimental side, high-throughput proteomic analysis has yielded protein-interaction maps for different organisms at an unprecedented level of detail (Matthews et al. 2001; Tucker et al. 2001). Initial analysis of this data reveals that the interactions within the proteome of an organism constitute a scale-free network characterized by hubs of highly connected polypeptides, each of which interact with several proteins with few or no further connections (Snel et al. 2002; Wolf et al. 2002).

Despite these advances, the precise set of changing interactions that are related to the organism’s responses to changing environments, or those that are involved in development and differentiation of multicellular organisms, is not easily deduced from these studies. Furthermore, the exact determinants of the interactions in a polypeptide and the effects of modifications on them cannot be extrapolated directly from these large-scale studies. This is where a new genre of computational studies could provide potentially interesting results. Essentially, these studies would need to go beyond the identification of the individual modules involved in inter-

actions and predict some of the actual interactions themselves. While the great structural diversity of the protein domains makes this task rather enormous, the current availability of large amounts of structural data makes this, in part, tractable. Computational analysis of this problem also is likely to uncover several general principles behind protein interactions that are unlikely to be directly uncovered through other methods.

Proteins mediate interactions with each other and other molecules via a great diversity of interfaces that span the whole range of structural complexity from simple α -helical surfaces, through repetitive α -helical or β -propeller superstructures, to complex binding pockets. One of the simplest interaction interfaces seen in proteins is the coiled coil that comprises two α -helical stretches winding around each other to form a double-helical superstructure (Fig. 1) (Lupas 1996; Burkhard et al. 2001). The coiled-coil regions are characterized by heptad periodicity and typically contain hydrophobic residues (like leucine) that lie on the same side of the helix and stabilize the superstructure through hydrophobic interactions (Fig. 1). As a result, these structures often are referred to as leucine zippers (O’Shea et al. 1989) and are utilized extensively in homo- or heterodimerization or oligomerization in all life forms, especially in eukaryotes. A number of eukaryotic transcription factors combine a DNA-

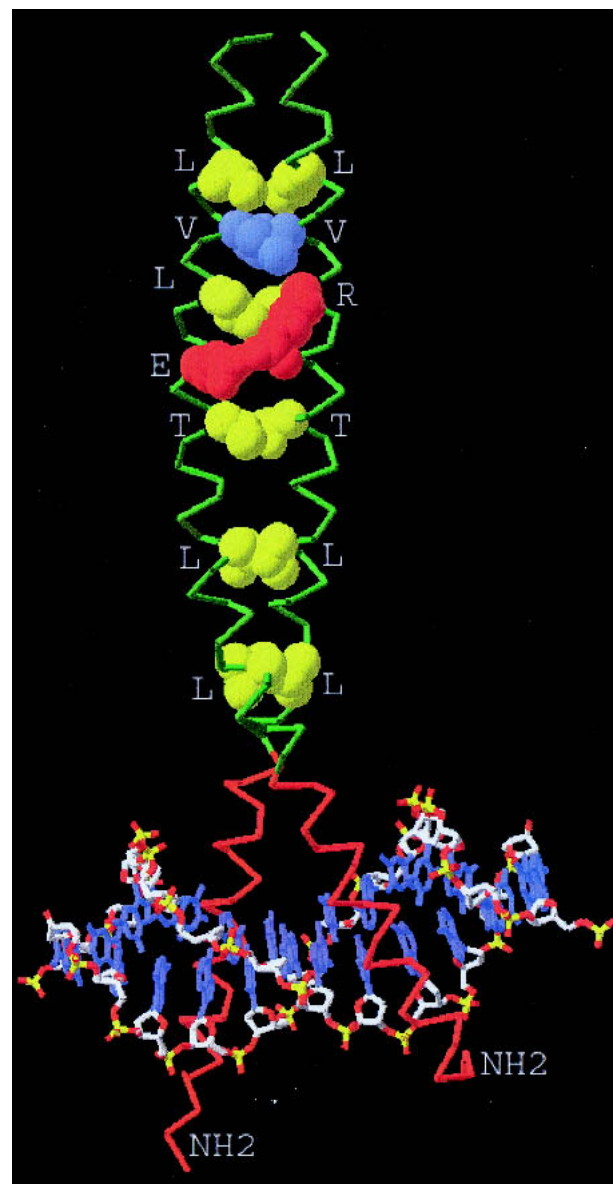


Figure 1 A representation of interactions in a coiled coil. This coiled coil is the basic stretch (B-ZIP) module of the transcription factor Pap1 and shows three different kinds of stabilizing interactions between residues of the heptad of each interacting partner. The leucines of the leucine zipper in position four of the heptad are shown in yellow, an example of the attractive interaction between the residues of the fifth and seventh position is shown in red, and the interaction between the residue pair in the first position of the heptad is shown in violet.

¹Corresponding author.

E-MAIL aravind@ncbi.nlm.nih.gov; FAX (301) 480-9241.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.353302>.

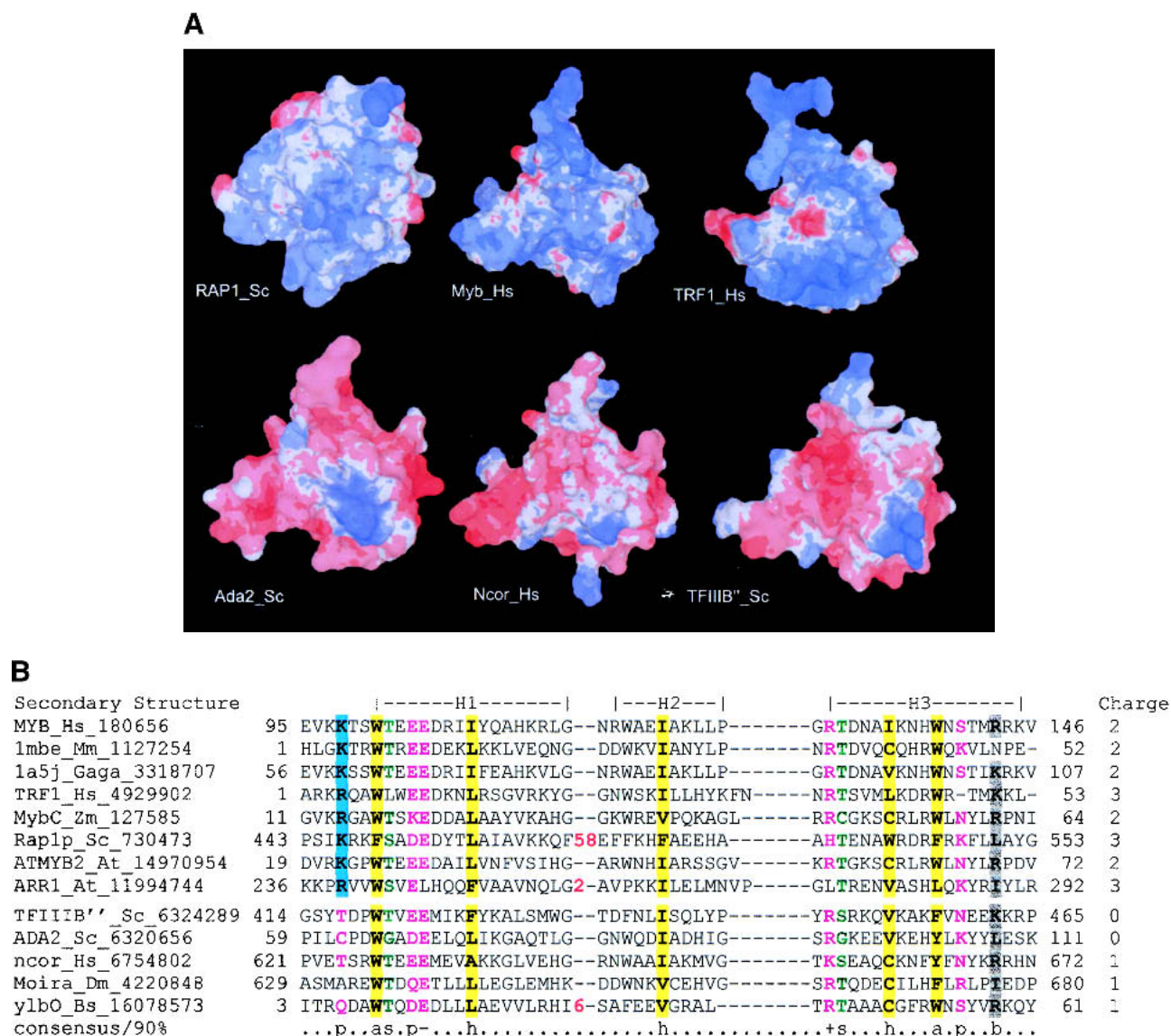


Figure 2 (A) Distribution of surface electrostatic potential on the DNA binding interface of Myb/SANT domains. The top row shows DNA binding Myb domains that have a basic surface charge (blue) on their DNA binding interface, and the bottom row shows SANT domains known to be involved in protein-protein interactions that have an acidic (red) or mixed charged surface in the same region. (B) Multiple sequence alignment of a representative set of DNA binding Myb domains (*top*) and SANT domains (*bottom*). Proteins are denoted by their gene names or pdb id (where a structure is available), their species abbreviations, and the Genbank Identifier (gi). The coloring reflects the amino-acid conservation profile at 90% consensus. The + charge in the minor groove binding site is shown to the right. Species abbreviations are as follows: At, *Arabidopsis thaliana*; Bs, *Bacillus subtilis*; Dm, *Drosophila melanogaster*; Gaga, *Gallus gallus*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Sc, *Saccharomyces cerevisiae*; Zm, *Zea mays*.

binding module, such as basic stretch (B-ZIP) (Landschulz et al. 1988), basic helix-loop-helix domain (bHLH) (Blackwell et al. 1990), or a homeodomain (HD-ZIP) (Schna and Davis 1992), with a coiled-coil region. The simplicity of the interaction interface and availability of extensive biochemical studies on the dimerization of the B-ZIP transcription factors make them attractive targets for prediction of protein-protein interactions through computational analysis of their sequence and structure.

Fassler et al. (2002) present results in this direction by using biochemical studies and thermodynamic measurements to identify two simple principles that govern B-ZIP dimerization. They suggest: (i) The presence of oppositely charged residues on the respective fifth and the seventh positions of the two intertwining heptads result in an attractive interaction favoring formation of a dimeric pair, while residues with the same charge in these positions result in a repulsion that acts against their dimerization (Fig. 1). (ii) Resi-

dues in the first position of the heptad of one monomer interact with the corresponding residues in the first position in the second monomer (Fig. 1). Polar or aliphatic residues in these positions stabilize dimers to a greater extent when they interact with the same kind of residue, as against pairs that may have an aliphatic-polar residue interface. Putting these principles together, Fassler et al. (2002) present extrapolations for the dimerization specificities of B-ZIP proteins from the complete proteome of *Drosophila melanogaster*.

The authors observe that a large number of the *Drosophila* B-ZIP proteins contain a polar residue (usually asparagines) in the first position of their heptad repeat. Combining this with the charge states in the fifth and seventh positions of the heptads they suggest that many of these proteins are more likely to homodimerize rather than heterodimerize with other B-ZIP proteins. *Jra* and *Kay*, the *Drosophila* orthologs of human protooncogene products, Jun and Fos, are predicted to heterodimerize rather than homodimerize based on the presence of repulsive residues in the fifth and seventh positions. Consistent with their model, these *Drosophila* proteins as well as their human orthologs have been experimentally shown to dimerize.

A large number of B-ZIP, bHLH-ZIP, and HD-ZIP transcription factors are encoded by most of the crown-group eukaryotes, especially the plants and humans (Riechmann et al. 2000; Lander et al. 2001), and the majority of them remain uncharacterized in terms of their interaction partners. Thus, extensions of studies such as those presented by Fassler et al. (2002) might aid in uncovering the diversity of their interactions and also understanding how these interactions have changed over evolution. Additionally, in eukaryotes, coiled coils act as interfaces for dimerization in proteins such as the cytoskeletal intermediate filaments; motor proteins like myosin, kinesin, and dynein; the membrane fusion proteins like SNAREs; chromosome condensation proteins like SMC; and the secretory vesicle cargo packaging proteins like P24 (Burkhard et al. 2001). Further analyses on the lines of those carried out on the B-ZIP transcription factors also may be useful in unraveling the range of interactions between these major functional components of the cells.

Can analogous simple rules be of value in predicting interactions between more complex protein interfaces? Preliminary results suggest that rules, with some degree of discrimination, may be devised for slightly more complex interaction modules, for which some biochemical data exist. One such module is the MYB-like domain that contains a version of the Helix-turn-Helix (HTH) fold. These modules have been known to interact with either DNA or proteins. They interact with DNA by inserting the "recognition helix" of the HTH into the major groove of the DNA and some forms additionally interact

with the minor groove via basic residues from the N-terminal tail (Hanaoka et al. 2001). Based on these properties, it has been proposed that the DNA-binding versions of the MYB domain form a strong basic surface on the side that interacts with DNA, while those that do not bind DNA (often referred to as SANT domains [Aasland et al. 1996]) instead, have a corresponding acidic or mixed charged surface (Hanaoka et al. 2001). Thus one could use models of MYB domains showing the surface electrostatic potential and overall positive charge in the domain as potential predictors for their interactions. Comparison of these properties between classic DNA-binding versions of the MYB domain and the SANT domains, which are found in numerous chromosomal proteins, indicates that many of the latter contain strongly acidic surfaces in place of the basic DNA-binding surfaces of the former (Fig. 2A). Consistent with this, SANT domains of proteins such as ADA2p and TFIIIB'' interact with proteins rather than DNA (Shah et al. 1999; Sterner et al. 2002). The overall acidic surface (Fig. 2) predicts that these are likely to interact with basic targets on the partner proteins. Further biochemical investigations of these domains may help in obtaining more specific rules for their interactions.

Thus, there is some promise that, at least some of the interactions mediated by compact domains with well-characterized structures, such as the HTH, the bHLH domain, the Bromodomain, or the Chromodomain, also could be captured through relatively simple rules. However, this would depend heavily on robust experimental evaluation of specific interactions to provide sufficient precedence to develop useful rules. While this experimental aspect is not particularly advanced for a large number of the characterized domains, future opportunities in this direction could emerge from the collusion of sequence and structure studies of specific protein domains with the protein-interaction maps generated by high-throughput proteomics.

REFERENCES

Aasland, R., Stewart, A.F., Gibson, T. 1996. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem. Sci.* **21**: 87–88.

Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., Weintraub, H. 1990.

Sequence-specific DNA binding by the c-Myc protein. *Science* **250**: 1149–1151.

Burkhard, P., Stetefeld, J., Strelkov, S.V. 2001. Coiled coils: A highly versatile protein folding motif. *Trends Cell Biol.* **11**: 82–88.

Fassler, J., Landsman, D., Acharya, A. Moll, J.R., Bonovich, M., and Vinson, C. 2002. *Genome Res.* **12**: –.

Hanaoka, S., Nagadoi, A., Yoshimura, S., Aimoto, S., Li, B., de Lange, T., Nishimura, Y. 2001. NMR structure of the hRap1 Myb motif reveals a conical three-helix bundle lacking the positive surface charge typical of Myb DNA-binding domains. *J. Mol. Biol.* **312**: 167–175.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Landschulz, W.H., Johnson, P.F., McKnight, S.L. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759–1764.

Lupas, A. 1996. Coiled coils: New structures and new functions. *Trends Biochem. Sci.* **21**: 375–382.

Matthews, L.R., Vaglio, P., Reboul, J., Ge, H., Davis, B.P., Garrels, J., Vincent, S., Vidal, M. 2001. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or interologs. *Genome Res.* **11**: 2120–2126.

O'Shea, E.K., Rutkowski, R., Kim, P.S. 1989. Evidence that the leucine zipper is a coiled coil. *Science* **243**: 538–542.

Ponting, C.P., Schultz, J., Copley, R.R., Andrade, M.A., Bork, P. 2000. Evolution of domain families. *Adv. Protein Chem.* **54**: 185–244.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., et al. 2000. Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**: 2105–2110.

Schena, M. and Davis, R.W. 1992. Zip proteins: Members of an Arabidopsis homeodomain protein superfamily. *Proc. Natl. Acad. Sci.* **89**: 3894–3898.

Shah, S.M., Kumar, A., Geiduschek, E.P., Kassavetis, G.A. 1999. Alignment of the B subunit of RNA polymerase III transcription factor IIIb in its promoter complex. *J. Biol. Chem.* **274**: 28736–28744.

Snel, B., Bork, P., Huynen, M.A. 2002. The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci.* **99**: 5890–5895.

Sterner, D.E., Wang, X., Bloom, M.H., Simon, G.M., Berger, S.L. 2002. The SANT domain of Ada2 is required for normal acetylation of histones by the yeast SAGA complex. *J. Biol. Chem.* **277**: 8178–8186.

Tucker, C.L., Gera, J.F., Uetz, P. 2001. Towards an understanding of complex protein networks. *Trends Cell Biol.* **11**: 102–106.

Wolf, Y.I., Karev, G., Koonin, E.V. 2002. Scale-free networks in biology: New insights into the fundamentals of evolution? *Bioessays* **24**: 105–109.