Chapter 3

Intrinsic Disorderness in Transient Interactions of Hub Proteins
3.1 Introduction

Protein-Protein Interactions (PPIs) are crucial for most cellular processes. These PPIs together form PPI networks. PPI networks have been constructed using manually curated data from the literature, high throughput experiments and prediction methods. The PPI networks of different organisms, like many other biological and non-biological networks, display scale-free topology [Barabasi and Oltvai, 2004], indicating the presence of ‘hubs’ in the network, which interact with large number of partners. In yeast PPI, hubs were classified based upon their temporal dynamics with their interaction partners, inferred from gene expression data [Han et al, 2004]. For each hub, average Pearson Correlation Coefficient (PCC) was calculated between the expression level of the hub and the expression level of each of its respective interaction partner across a variety of experimental conditions. It was inferred that hubs with lower average PCC bind to different partners at different times (date hubs) while hubs with higher PCC interact with most of their partners simultaneously (party hubs) [Han et al, 2004].

Disordered regions have been implicated in transient binding interactions. Disordered regions generally undergo disorder to order transition as they bind to their interaction partners. This transition is associated with a decrease in conformational entropy of the disordered region, which uncouples binding strength from specificity, allowing highly specific interactions to become reversible [Tompa, 2002; Dyson and Wright, 2005], which is a prerequisite for transient interactions. (See “Transient Binding” in Chapter 1).

Disordered regions have also been implicated in multiple binding interactions as their structural plasticity allows them to efficiently interact with different binding partners (See “Binding To Multiple Partners” in Chapter 1). It was shown that hub proteins were more disordered compared to non hub proteins [Dosztanyi et al, 2006; Haynes
3.2 Materials and Methods

3.2.1 Data

Protein interaction network data was obtained from Han et al, 2004. This dataset involves 1,379 proteins with 2,493 high-confidence interactions. Among the 2,493 interactions, 931 interactions involved date hubs and 907 interactions involved party hubs. In this set, hubs were defined as proteins with >5 interacting partners. The average Pearson Correlation Coefficient (PCC) for the subnetwork comprising of the hub and its interacting partners was obtained from Han et al, 2004.

A dataset of 70,647 derived binary interactions along with the socio affinity index score for the interaction was obtained from the genome-wide yeast protein complex data [Gavin et al, 2006]. This list is available at http://yeast-complexes.embl.de/complexview.pl?rm=download.

3.2.2 Prediction of Disordered Regions

I used four different disorder prediction methods namely, modified Uversky’s method, IUPred, GlobPlot and Disembl (REM465).

Modified Uversky method is based on the observation that IDPs have high mean net charge and/or low mean hydrophobicity [Uversky et al, 2000]. It is a global disorder predictor and gives a Netscore to each protein. This was calculated for each protein using following formula:

Netscore = [(mean net charge) − 2.785 * (mean net hydrophobicity) + 1.151] / 2.952
Intrinsic Disorderness in Transient Interactions of Hub Protein

A protein with positive Netscore is predicted to be IDP, while negative score predicts ordered protein.

IUPred is based on the assumption that disordered regions do not form sufficient favorable interactions to fold and thus have high estimated energy content [Dosztanyi et al, 2005a; Dosztanyi et al, 2005b].

GlobPlot is based on propensities derived from non-globular regions in PDB [Linding et al, 2003a].

Remark 465 of DisEMBL is an artificial neural network based method trained on regions with non-assigned electron densities in PDB [Linding et al, 2003b].

3.2.3 Secondary Structure Prediction

Consensus secondary structure for proteins was predicted using NN, SOPM, DPM, DSC, GOR4, PHD, PREDA and SIMPA96 algorithms available at NPS server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html).

3.3 Results

3.3.1 Intrinsic Disorder Inversely Correlates With Average PCC of Hubs

I calculated the predicted intrinsic disorder of the hub protein and studied its correlation with the corresponding average PCC of the hub. The plot between average PCCs and intrinsic disorder of hub proteins showed a weak yet significant negative correlation as determined by two different disorder prediction programs (Figure3.1a and b; Modified Uverksy’s method-PCC of -0.253, P value =3.2E-04; IUPred-PCC of -0.3155, P value = 5.6E-06).
Intrinsic Disorderness in Transient Interactions of Hub Protein

Fraction of residues disordered in hub proteins (IUPred)

Average PCC

Net Score of the hub protein (Modified Uversky’s method)

Average PCC

b
Figure 3.1 Correlation between fraction of intrinsic disorder in hub proteins with the average PCC of the hub. a. Intrinsic disorder was predicted by modified Uversky’s method, wherein positive scores indicate disordered protein, while negative score indicates ordered protein. b. Intrinsic disorder was predicted by IUPred method.

3.3.2 Predicted Intrinsic Disorder in Date and Party Hub Proteins

A significantly high number of date hubs were predicted to be disordered by modified Uversky’s method compared to the party hubs (Table 3.1a; $\chi^2=13.28$, degree of freedom 1, $P$ value = 0.0003). Mean netscore derived from modified Uversky’s method for date hubs was more than twice than that for party hubs (-0.014 and -0.033 respectively, t-test $P$ value = 3E-04).

Table 3.1 Prediction of disorder in date and party hub proteins by various disorder prediction methods. a. Disorder prediction by global disorder prediction method-modified Uversky's method. b. Disorder prediction by local disorder predictors.

<table>
<thead>
<tr>
<th>Hubs</th>
<th>No. of disordered proteins (%)</th>
<th>No. of ordered proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date (91)</td>
<td>28 (30.8)</td>
<td>63</td>
</tr>
<tr>
<td>Party (108)</td>
<td>11 (10.2)</td>
<td>97</td>
</tr>
</tbody>
</table>
b.

<table>
<thead>
<tr>
<th>Prediction method</th>
<th>IUPred</th>
<th>Total no. of proteins</th>
<th>No. of disordered segments of ≥30 residues</th>
<th>No. of proteins with disordered segments of ≥30 residues</th>
<th>Average length of disordered segments</th>
<th>Average no. of disordered segments per protein</th>
<th>% residues in disordered segments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>date hubs</td>
<td>91</td>
<td>118</td>
<td>49</td>
<td>90.39</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>party hubs</td>
<td>108</td>
<td>69</td>
<td>41</td>
<td>67.74</td>
<td>0.64</td>
</tr>
<tr>
<td>Globplot</td>
<td></td>
<td>date hubs</td>
<td>91</td>
<td>74</td>
<td>46</td>
<td>91.22</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>party hubs</td>
<td>108</td>
<td>39</td>
<td>34</td>
<td>77.54</td>
<td>0.36</td>
</tr>
<tr>
<td>DISOPRED (REM465)</td>
<td></td>
<td>date hubs</td>
<td>91</td>
<td>48</td>
<td>33</td>
<td>43.96</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>party hubs</td>
<td>108</td>
<td>32</td>
<td>20</td>
<td>42.63</td>
<td>0.30</td>
</tr>
</tbody>
</table>
Date hubs also showed significantly higher number of disordered segments as compared to party hubs, by three local disorder prediction methods used (IUPRED, Globplot, Disembl; Table 1b). Compared to party hubs, date hubs had almost twice as many long disordered regions (≥30 residues) per protein, their average length was longer and covered twice as much portion of proteins (Table 3.1b). Not only the number of long disordered segments was higher in date hubs but also the number of proteins having such long disordered segments was significantly higher (Table 3.1b).

I also did secondary structure analysis of date and party hub proteins. Since only few of these proteins had representation in PDB (5 full length date hub and 14 party hub proteins), I analyzed secondary structure by prediction methods. Coils are not necessarily disordered, but prevalence of coiled residues reflect prevalence of intrinsic disorder. The data shows that fraction of coils in date hub proteins is significantly higher as compared to the party hub proteins (Figure 3.2; t-test for fraction of coil residues - \( P \) value = 1E-05).

![Figure 3.2 Predicted secondary structure of date and party hub proteins.](image-url)
3.3.3 Transient Nature of Date Hub Interactions as Revealed by Protein Complex Data

Gavin et al., 2006, identified protein complexes in yeast at genome wide level by affinity purification of tagged proteins followed by mass spectrometry. Because this technique isolates stable protein complexes, it identifies stable interactions. Thus I analyzed the distribution of date and party hub interactions in this data. I searched the hub–interaction partner pairs of Han et al (2004) in the protein-protein interaction pairs provided by Gavin et al (2006). 42% of the interaction pairs involving date hubs (387/931) and 67% of interaction pairs involving party hubs (603/907) mapped to the derived binary interactions provided by Gavin et al, 2006 ($\chi^2$ value=113.8, degree of freedom1, $P$ value=1.4E-26, Figure 3.3).

Figure 3.3 Lower percentage of date hub interactions mapped to protein complex data compared to party hubs.
Gavin et al. (2006) defined a term ‘socio-affinity index’ which measured the propensity of proteins to form partnerships. This index was tentatively correlated to the available dissociation constant for the interaction pairs [Gavin et al., 2006]. The distribution of socio-affinity index was significantly different between interactions involving date hubs and party hubs (Figure 3.4; t-test $P = 0.001$, Wilcoxon rank sum test $P = 0.006$).

![Figure 3.4](image.png)

**Figure 3.4** The difference in mean socio-affinity index between date and party hubs. Error bars indicate $\pm$ SE.

### 3.4 Discussion

Molecular recognition involving reversible but specific interaction poses particular problems to rigid body assumption of proteins. In order to have specific recognition, interaction partners should have large number of contacts and thus high binding affinity but this also reduces the dissociation rate (and thus reversibility) of interaction [Dunker et al., 2001]. In other words binding affinity and binding specificity are coupled. Intrinsic disorder has been suggested to allow specific interaction in reversible/weak manner by uncoupling binding affinity and binding specificity [Dunker et al., 2001; Dyson and

I show that average PCC of hubs, correlates weakly yet significantly with measures of intrinsic disorder in hubs. Classification of hubs into date and party hubs also showed clear over-representation of intrinsic disorder in date hubs as analyzed by a global and three local prediction methods. The significant prevalence of party hub interactions in the genome wide protein complex dataset supports the transient nature of date hub interactions, which makes it less likely to be identified in a screen for stable protein complexes. This is further corroborated by my data wherein date hubs show significantly low socio-affinity indices as compared to party hubs, since previously, socio-affinity index has been tentatively correlated to available dissociation constants in the literature [Gavin et al, 2006]. Further, compared to party hubs, date hubs were enriched in ‘cell signaling’ and ‘transcription’ categories [Han et al, 2004]. The enrichment of disorder in date hubs might enable them to perform their functions in signaling pathways, which are extensively mediated by transient interactions.

It should be noted that the classification of hubs into date and party hubs has been considerably debated [Batada et al, 2007; Bertin et al, 2007]. But the existence of difference in the binding properties of hubs is supported by 1) biological examples of signaling hubs like p53, which transiently interact with their binding partners and hubs in protein complexes, which interact with their partners in a more stable manner, 2) structural analyses of hubs and their classification into single and multi-interface hubs, which correspond to date and party hubs respectively [Kim et al, 2008], and 3) my own analyses, which shows clear distinction of date and party hub proteins in their disorder content.