ABSTRACT

Motivation: Progress in large-scale experimental determination of protein–protein interaction networks for several organisms has resulted in innovative methods of functional inference based on network connectivity. However, the amount of effort and resources required for the elucidation of experimental protein interaction networks is prohibitive. Previously we, and others, have developed techniques to predict protein interactions for novel genomes using computational methods and data generated from other genomes.

Results: We evaluated the performance of a network-based functional annotation method that makes use of our predicted protein interaction networks. We show that this approach performs equally well on experimentally derived and predicted interaction networks, for both manually and computationally assigned annotations. We applied the method to predicted protein interaction networks for over 50 organisms from all domains of life, providing annotations for many previously unannotated proteins and verifying existing low-confidence annotations.


Contact: admin@bioverse.compbio.washington.edu

Supplementary information: A supplemental appendix gives additional details not in the main text. (http://data.compbio.washington.edu/misc/downloads/nannotation_data/supplement.pdf).

INTRODUCTION

Advances in large-scale protein interaction determination methods have made the elucidation of protein interaction networks for entire organisms possible. Protein interaction networks have been experimentally determined for Caenorhabditis elegans (Li et al., 2004), Drosophila melanogaster (fly) (Giot et al., 2003), Helicobacter pylori (Rain et al., 2001) and Saccharomyces cerevisiae (yeast) (Fromont-Racine et al., 1997; Schwikowski et al., 2000; Uetz et al., 2000; Ho et al., 2002). Although incomplete, these networks have been used to predict lethal mutations in yeast (Jeong et al., 2001), provide evolutionary comparisons between organisms (Wuchty et al., 2003) and identify functional modules and network motifs (Ravasz et al., 2002; Spirin and Mirny, 2003).

These networks have also been used to provide functional annotations based on network connectivity: metabolic and protein interaction networks have been shown to be functionally modular in nature (Schwikowski et al., 2000; Ravasz et al., 2002; Rives and Galitski, 2003), and proteins that interact have been observed to be more likely to share function and cellular location (Schwikowski et al., 2000; von Mering et al., 2002). This fact is exploited in the 'majority-rule' method of network annotation in which a protein is annotated based on the most commonly occurring functions in its interaction partners (Schwikowski et al., 2000). In the yeast interaction network, it was reported that 42 cellular role annotation categories (Schwikowski et al., 2000) could be assigned with an accuracy of ~70% using the majority-rule method. Approaches using Markov random field analysis (Lefortsky and Kasif, 2003; Deng et al., 2004), global network connectivity (Vazquez et al., 2003) and clustering (Brun et al., 2003; Samanta and Liang, 2003) have all been applied to the yeast protein interaction network with equal or greater success. Functional predictions at this level of accuracy are useful for annotation of proteins for which there is little functional information and for providing novel functional predictions, such as involvement in specific pathways, or confirming existing annotations provided by other methods, for proteins that have been characterized.

Since protein interaction datasets do not exist for most organisms, including several important ones, computational methods have been developed to predict protein interactions or functional relationships between proteins in experimentally uncharacterized organisms (Pazos et al., 1997; Marcotte et al., 1999; Matthews et al., 2001; Goh and Cohen, 2002; Yu et al., 2004). In the interolog method (Walhout et al., 2000; Matthews et al., 2001; Yu et al., 2004), for example, interactions are predicted between two proteins based on their sequence similarity to protein pairs known to interact (Fig. 1A). A considerable benefit of this method is that the prediction of protein interaction networks for one organism is accomplished by integrating over protein interactions from a large number of diverse sources. Previous results indicate that proteins predicted to interact also have similar functions (figure 1 in von Mering et al., 2002) but did not provide detailed analysis of this observation.

The Bioverse database and computational biology framework (http://bioverse.compbio.washington.edu), an integrated, knowledge-based resource to facilitate the understanding of the relationships between molecular and organismal biology, includes predicted protein interactions as well as functional annotations for over 50 organisms. The Bioverse is unique in part because each prediction is assigned a heuristic quality score, which can be used to integrate information from different sources and to calibrate the resulting predictions. We used predicted protein interaction networks that combine information from a large number of sources and developed a network annotation method, the ‘neighborhood weighting method’,...
We show that this combination of methods can be used to annotate proteins in organisms for which little or no experimental interaction information is available, and that weights generated can be used to estimate the precision of these predictions.

**METHODS**

**Protein annotation**

Protein sequences in the Bioverse were obtained from the NCBI sequence repository (Benson et al., 2000) and from collaborators (Kikuchi et al., 2003; Yu et al., 2005). Starting (primary) automated functional annotations (automated primary annotations; APAs) were performed by applying a variety of domain/family/motif classification methods to each sequence and mapping these individual results to Interpro categories, similar to the iprscan program (Apweiler et al., 2000), then to Gene Ontology (GO) categories (The Gene Ontology Consortium, 2001) using the ipr2go mapping provided by Interpro (see Supplementary Material). With respect to the current study it is important to note that APAs are based only on matches to the category databases outlined above and do not include information derived from interolog determinations, the described network-based annotation, or direct transfer of annotation from similar sequences. Since APAs are based on sequence similarity to conserved domains/motifs it is impossible to eliminate the presence of some circularity in the annotation process; however, this method avoids the overt circularity found in some other methods.

A primary annotation is the initial functional annotation for a protein, either manually assigned (manual primary annotation, MPA) or assigned by the Bioverse using automated methods (APA), before any network-based predictions have been made. For APAs, quality scores for annotations from individual methods were calculated as the percentage sequence identity (SI) between the protein sequence and the matched pattern or sequence. The overall scores of Interpro (and thus GO) annotations (annotation scores) were calculated as the score of the best individual method contributing to that Interpro category.

The GO is a vocabulary for functional description of proteins and is arranged in a directed acyclic graph (DAG) of categories of different levels of functional specificity. For instance, the GO category for receptor tyrosine kinases (GO:0004716) is a member of several other GO categories including protein kinases (GO:0004713), which is in turn a kinase (GO:0016301), which is an enzyme (GO:003824). Thus proteins with specific GO annotations can be described at varying functional specificity levels (GO levels, i.e., distance of the category from the DAG root; see Supplementary Table I) that describe a path (or multiple paths) to the DAG root. Because categories in a DAG may have multiple parents, some categories have several different specificity levels; however, the level is generally correlated with functional specificity (The Gene Ontology Consortium, 2001).

GO MPAs were obtained from the Saccharomyces Genome Database (SGD) (Weng et al., 2003) for yeast, from FlyBase (The FlyBase Consortium, 2003) for fly and from GenBank (Benson et al., 2000) for human. For the purposes of network-based annotation (see below), the quality score of these annotations was considered to be 1.0. The human APAs from Bioverse were combined with the MPAs from GenBank for the examples in Table 2. In cases where a protein had an identical annotation from both APA and MPA sources, the MPA category had precedence and was assigned a functional quality score of 1.0. MPAs with a source of inferred by electronic annotation were not considered to be manually curated (The Gene Ontology Consortium, 2001) and hence were discarded.

**Interaction prediction and network generation**

Protein–protein interactions were predicted by a method very similar to the previously described ‘interolog’ method (Walhout et al., 2000; Matthews et al., 2001; Yu et al., 2004). A database of protein interactions was compiled from the Database of Interacting Proteins (DIP) (Xenarios et al., 2000; 19,251 interactions), the General Repository for Interaction Datasets (GRID)
[Breitkreutz et al. (2003); 20,985 interactions], and crystallized complexes in the Protein Data Bank (PDB) [Berman et al. (2000); 8,835 interactions]. All the proteins encoded by a target genome were compared to the sequences in the interaction database using three iterations of PSI-BLAST and considering all matches with Z-scores better than 5 and E-values <1.0. The SI score for the match was then calculated as the percentage of identical residues matched, relative to the entire length of the matched sequence from the interaction database. In contrast to the original interolog method [Walhout et al., 2000], all similar sequences were considered, and no attempt to distinguish orthologs and paralogs was made. Although use of these parameters does not exclude very low quality matches the interaction quality score, which incorporates the SI of the matches, defines the contribution of the matches in the neighborhood weighting method (described below); so inappropriate interactions provide a minimal contribution to the prediction process. PSI-BLAST was chosen over slower but more sensitive methods of determining sequence similarity, because of the aim of providing annotations for large numbers of sequences.

Interactions were predicted when each partner (A and B) in an experimentally derived interaction was found to be similar (A: I and B: II) to different numbers of sequences.

sequence similarity, because of the aim of providing annotations for large numbers of sequences.

Interactions were predicted when each partner (A and B) in an experimentally derived interaction was found to be similar (A: I and B: II) to different proteins (I and II) in the target organism (Fig. 1A). A score was calculated by multiplying the SI from each comparison together (SI_{A \times I} \times SI_{B \times II}; interaction score), analogous to the ‘joint similarity’ measure previously described by Yu et al. (2004). The experimental yeast interactions are from the yeast core interaction set from the DIP.

Recently, experimental protein interaction data from a large-scale study of the fruit fly (Giot et al., 2003) have become available. We chose not to include these data in our protein interaction database in order to ensure that the results obtained for the predicted protein interaction networks for the fly were not unfairly biased by their presence. Additionally, this dataset is fairly small relative to that of the yeast. Inclusion of this and similar experimental data, in the future, will improve results for the organism in question as well as for related organisms.

The purpose of this study was to generate the most accurate network-based functional annotations for as many proteins as possible in a given target organism, not necessarily to generate the most accurate predicted protein interaction network [estimates of the accuracy of in silico predicted interactions can be found (von Mering et al., 2002)]. We, therefore, considered all similar proteins found (not just the highest scoring match) and used an inclusive score threshold for network generation. For the network size numbers in Table 1, predicted interactions with scores >0.15 [equivalent to a joint similarity (Yu et al., 2004)] of 0.38] were included as edges in the predicted networks. Although this is a more inclusive limit than previously determined for high confidence interolog mapping [0.80, Yu et al. (2004)], it still represents a conservative threshold for interaction inclusion (figure 2B from Yu et al., 2004). For purposes of network annotation prediction, all interaction scores were considered and were used in the calculation of the neighborhood weight (see below), so that lower scoring interologs have a smaller impact on the final predictions.

Network-based annotation

Network-based annotation was performed for each protein in the network by first compiling a list of GO categories associated with all proteins connected to it by a predicted or experimental interaction. A neighborhood weight was calculated for each category in the list based on the frequency of occurrence of the category (majority-rule method) or by summing the individual contributions from neighbors as

\[
\text{Neighborhood weight} = \sum_{i=1}^{N} A_i^C \times \text{Interaction}_i^C \times \text{Annotation}_i^C,
\]

where \( C \) is the category being scored, and \( N \) is the number of contributors to that category. Interaction is the protein interaction quality score and Annotation is the functional annotation quality score from that contributor. The source factor \( A \) is used to adjust the contribution of predicted interactions derived from the DIP and GRID. Supplementary Figure I shows that DIP and GRID have a significantly different contribution to network annotation of predicted interaction networks than do predicted interactions derived from the PDB. For the figures in this paper a source factor of 0.1 was used, meaning that predicted interactions derived from DIP or GRID contribute significantly less to the neighborhood weight than do PDB-derived interactions. This value was chosen from visual inspection of graphs of precision versus coverage at various source factor values and minimum weight thresholds (Supplementary Figure I) and provides a reasonable balance between precision and coverage. The list of categories was then sorted according to the neighborhood weight and the top-ranked categories used as predictions. For Figures 2B and 3, only predictions with a neighborhood weight of >0.025 were considered.

Comparison of Figure 2A and B shows that precision is correlated with both the magnitude of the neighborhood weight of a prediction (Fig. 2B) and the rank of the prediction (Fig. 2A). This rank is determined according to its weight but is only relative to the weights of the other predictions made for the same protein. The neighborhood weighting method represents an advantage over the majority-rule method in its ability to incorporate interactions and annotations that have quality scores, and to distinguish between sources of interaction data.

To generate random control networks, the primary annotations for all proteins in the network were randomly reassigned to another protein in the network. This provides the most conservative control network for annotation, as it retains network structure and distribution of annotations. Local network connectivity properties that are important for the majority-rule and neighborhood weighting methods, such as the average number of neighbors and the distribution of \( k \) in the network, are identical between the experimentally derived (yeast) or predicted networks and their randomized control networks.

Evaluation of results

To evaluate the performance of the methods used, we chose to use the precision measure described by Deng et al. (2004). Precision and recall, commonly used in the evaluation of information retrieval (Donaldson et al., 2003; Zhou et al., 2004), are combined measures of true positive (TP), false positive (FP) and false negative (FN) rates. Standard precision is the number of correct predictions made out of all predictions (\( N \)), and can be expressed as

\[
\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.
\]

However, comparison of two GO categories would underestimate precision at 0 if the categories were different but shared a common path. For instance a clathrin vesicle coat annotation is more similar to vesicle coat than to ATPase, and an evaluation measure should account for this type of similarity (see Supplementary Table I). The precision measure described in detail by Deng et al. (2004) addresses this issue by expressing the degree of agreement between GO paths, rather than simply the agreement between categories. So, summarized from Deng et al. (2004), precision is defined as

\[
\text{precision} = \frac{\sum_{i=1}^{N} 4^{-|P_i - O_i|}}{N},
\]

where \( N \) is the number of predicted GO paths, \( P_i \) is the length of GO path \( i \) and \( O_i \) is the maximal overlap between GO path \( i \) and all the known GO paths. Precision results for each protein in the network were averaged to provide overall values. Behavior of this measure is somewhat problematic when considering very short GO paths. For example, the precision given for a predicted GO path with length 1 that has no overlap with any known GO paths is 25%. However, the measure is well-behaved in the range of GO path lengths actually observed in our data (4–14), and is a good measure of accuracy.

Recall is the percentage of the known annotations that were predicted by the method and the common counterpart to precision, and is defined as

\[
\text{recall} = \frac{\sum_{i=1}^{M} 4^{-|P_i - O_i|}}{M},
\]

where \( M \) is the number of known GO paths, \( P_i \) is the length of GO path \( i \) and \( O_i \) is the maximal overlap between GO path \( i \) and all the known GO paths. Precision results for each protein in the network were averaged to provide overall values. Behavior of this measure is somewhat problematic when considering very short GO paths. For example, the precision given for a predicted GO path with length 1 that has no overlap with any known GO paths is 25%. However, the measure is well-behaved in the range of GO path lengths actually observed in our data (4–14), and is a good measure of accuracy.
Table 1. Network-based annotation results for selected organisms in the Bioverse

<table>
<thead>
<tr>
<th>Organism</th>
<th>Number of proteins</th>
<th>Largest predicted network</th>
<th>Number of proteins</th>
<th>Number of interactions</th>
<th>Proteins with no primary annotation</th>
<th>Network-annotated proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>25 328</td>
<td>3216</td>
<td>38 412</td>
<td>871</td>
<td>322</td>
<td></td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>21 882</td>
<td>2156</td>
<td>38 116</td>
<td>353</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>15 879</td>
<td>2799</td>
<td>54 884</td>
<td>421</td>
<td>326</td>
<td></td>
</tr>
<tr>
<td>Encephalitozoon cuniculi</td>
<td>1908</td>
<td>380</td>
<td>874</td>
<td>33</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>36 996</td>
<td>6779</td>
<td>39 274</td>
<td>1726</td>
<td>790</td>
<td></td>
</tr>
<tr>
<td>Orca sativa</td>
<td>25 840</td>
<td>3458</td>
<td>50 743</td>
<td>700</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>Magnaporthe grisea</td>
<td>11 042</td>
<td>1842</td>
<td>19 076</td>
<td>1257</td>
<td>583</td>
<td></td>
</tr>
<tr>
<td>Plasmodium falciparum</td>
<td>4584</td>
<td>262</td>
<td>3391</td>
<td>49</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Saccharomyces cerevisae</td>
<td>6278</td>
<td>5118</td>
<td>10 380</td>
<td>2789</td>
<td>890</td>
<td></td>
</tr>
<tr>
<td>Agrobacterium tumefaciens</td>
<td>5396</td>
<td>296</td>
<td>690</td>
<td>28</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Bacillus anthracis</td>
<td>5309</td>
<td>264</td>
<td>732</td>
<td>60</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Bacillus subtilis</td>
<td>4112</td>
<td>247</td>
<td>707</td>
<td>27</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Brucella melitensis</td>
<td>3188</td>
<td>238</td>
<td>652</td>
<td>68</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Brucella suis</td>
<td>3247</td>
<td>225</td>
<td>611</td>
<td>64</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Campylobacter jejuni</td>
<td>1634</td>
<td>334</td>
<td>10 288</td>
<td>138</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Clostridium perfringens</td>
<td>2659</td>
<td>125</td>
<td>390</td>
<td>42</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>4257</td>
<td>400</td>
<td>1473</td>
<td>126</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>1562</td>
<td>771</td>
<td>5647</td>
<td>337</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Listeria monocytogenes</td>
<td>2844</td>
<td>176</td>
<td>485</td>
<td>53</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Mycobacterium tuberculosis</td>
<td>4176</td>
<td>107</td>
<td>375</td>
<td>28</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Neisseria meningitidis</td>
<td>2020</td>
<td>103</td>
<td>400</td>
<td>14</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Pseudomonas aeruginosa</td>
<td>5555</td>
<td>333</td>
<td>903</td>
<td>37</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Rickettsia conorii</td>
<td>1374</td>
<td>77</td>
<td>216</td>
<td>22</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Rickettsia prowazekii</td>
<td>834</td>
<td>67</td>
<td>181</td>
<td>20</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Salmonella typhimurium</td>
<td>4429</td>
<td>332</td>
<td>1359</td>
<td>107</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>Shigella flexneri</td>
<td>3848</td>
<td>383</td>
<td>4548</td>
<td>108</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>2632</td>
<td>113</td>
<td>316</td>
<td>34</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Vibrio cholerae</td>
<td>3788</td>
<td>275</td>
<td>1021</td>
<td>86</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Vibrio para-hydrolyticus</td>
<td>4821</td>
<td>365</td>
<td>1520</td>
<td>54</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Vibrio vulniatus</td>
<td>4484</td>
<td>372</td>
<td>1557</td>
<td>72</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Yersinia pestis</td>
<td>3729</td>
<td>352</td>
<td>1100</td>
<td>106</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Halobacterium</td>
<td>2425</td>
<td>397</td>
<td>1545</td>
<td>143</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Methanococcus jannaschii</td>
<td>1714</td>
<td>45</td>
<td>133</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Methanococcus maripaludis</td>
<td>1722</td>
<td>17</td>
<td>41</td>
<td>17</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Pyrococcus abyssi</td>
<td>1769</td>
<td>72</td>
<td>196</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Shown for each organism is the number of proteins encoded by the genome; the number of proteins in the predicted protein interaction network; the number of interactions in the predicted network; the number of proteins with neither functional annotation nor primary annotation scores < 0.25 (proteins with no primary annotation); and the number of such proteins that could be assigned a GO category using the neighborhood weighting method with a confidence (estimated precision) > 60% (network-annotated proteins) by correlating the neighborhood weight with prediction precision (as in Fig. 2B). This table demonstrates the applicability of the neighborhood weighting method to predicted protein interaction networks from any organism. It also shows that a substantial number of proteins with no APA can be assigned functions with high confidence (estimated precision) using our approach. Results for over 50 organisms can be found on the Bioverse website (http://data.compbio.washington.edu/misc/downloads/nannotation_data/).

RESULTS

We applied network annotation methods to the predicted protein interaction networks in the Bioverse to predict functional annotations and evaluated the precision of predictions made using this method under various conditions (Fig. 1B). A variation of the majority-rule method that exploited the quality scores associated with both the predicted protein interactions and starting (primary) functional annotations was developed. In the ‘neighborhood weighting’ method, the contribution of each annotation from neighboring proteins (i.e. proteins predicted to interact) is weighted based on a heuristic combination of the interaction and functional annotation quality scores, which produces a weight for each functional category predicted. This neighborhood weight allows ranking and filtering of the predictions made by the method.

Annotations were assigned to proteins before applying any network-based predictive methods (primary annotations) by manual
Filter predictions to an arbitrary precision. Results show the manner in which the neighborhood weight can be used to comparable axes, scales and ranges) using manual primary annotations from proteins with at least one prediction (coverage) out of all proteins evaluated (2037) versus the threshold is shown as a solid area. Shown in the inset graph is precision versus rank using the majority-rule method (circles). This measure evaluates the precision of all predictions by comparing them with the primary annotations for the protein in question, but allows partial matches for predictions where a portion of the GO path matches a primary annotation (see Methods section). Multi-forward cross-validation is not necessary since our method does not use any training. Results were compared against control networks generated by randomly reassigning the primary annotations for a particular protein to another protein in the network, and repeating this for all proteins in the network.

Automated versus manual annotations in network annotation methods

For the relatively large number of organisms for which the MPAs available are few or none, APAs must be used to provide functional information. Thus, we first wanted to show that network annotation methods can be extended to APAs, such as those provided by the Bioverse, in the context of experimentally derived networks. We compared the performance of MPAs with the performance of APAs, using the majority-rule method on the experimentally derived yeast network. Results using MPAs were similar to those reported previously (Schwikowski et al., 2000; Deng et al., 2004) with a precision of 61% (MPA control precision was 11%), considering only the most commonly occurring predicted category for each protein with two or more “votes” (data not shown). The recall of the method increases as more rank positions are considered, and for MPAs the recall at the top-ranked position was 13% increasing to 35% when considering the top three ranks. Precision using APAs was 49%, but was still well above APA control levels of 18% (data not shown), and the recall was 14% at the top-ranked position and 33% considering the top three ranks. These results show that computationally predicted annotations behave in a manner similar to manually assigned annotations in network annotation.

Network-based annotation of predicted interaction networks

Although network annotation has been shown to work well for experimental protein interaction networks previously (Schwikowski et al., 2000; Brun et al., 2003; Letovsky and Kasif, 2003; Samanta and Liang, 2003; Vazquez et al., 2003; Deng et al., 2004) and in the current study, it remained unclear how well it would work on networks predicted using the interolog method. Accordingly, we evaluated the precision of network annotation methods applied to the predicted protein interaction network from Drosophila. Shown in Figure 2A is precision versus rank using the majority-rule method (circles) on the predicted network and a random control network (dashed lines), where rank is determined by the frequency of curation (here called MPA), as in the case of the SGD annotations for yeast (Weng et al., 2003). For most organisms, however, MPAs are not available or cover only a limited number of proteins in the organism’s proteome. Unassisted computational annotation (here called APA) was used independently to assign primary annotations to proteins (see Methods section).

To evaluate the accuracy of the methods described, we first performed leave-one-out analysis by comparing GO category predictions made for a particular protein to the known annotations for that protein (i.e. primary annotations) and used a precision measure which accounts for the structure of the GO (Deng et al., 2004) (Fig. 1B). This measure evaluates the precision of all predictions by comparing them with the primary annotations for the protein in question, but allows partial matches for predictions where a portion of the GO path matches a primary annotation (see Methods section). Multi-forward cross-validation is not necessary since our method does not use any training. Results were compared against control networks generated by randomly reassigning the primary annotations for a particular protein to another protein in the network, and repeating this for all proteins in the network.
occurrence of the category among a protein’s neighbors. The precision of the method is ~50% for the top-ranked prediction and falls to control levels by the fifth ranked prediction. The recall of the method was 15% at the top-ranked position and 46% when considering the top three ranked positions. The neighborhood weighting method (squares), in which predictions are ranked based on a calculated weight, was developed to make use of the quality scores associated with the predicted interaction and the automated annotation, and to account for differences in source interaction data (see Supplementary Figure I). Figure 2A shows that the neighborhood weighting method yields significantly better precision for predictions in the top-five rank positions than control levels and that the top-ranked prediction is at ~70% precision, with a recall of 13% at the top-ranked position and 43% when considering the top three ranked positions.

Using the neighborhood weight to filter predictions shows that much higher levels of precision can be obtained using this method, albeit with lower coverage. We applied the neighborhood weight method to the predicted fly network, using the method as a filter by only evaluating the precision of predictions with weights above a minimum threshold. Shown in Figure 2B is the average precision of the top-five ranked predictions (closed squares) using different minimum weight thresholds (abscissa) versus random control precision (open squares) when starting with APAs. Also shown is the number of proteins with at least one prediction above the neighborhood weight threshold indicated (coverage; solid area), expressed as a percentage of the total number of proteins that could be evaluated (2037). The inset to Figure 2B shows the corresponding results from the predicted fly network starting with MPAs provided by FlyBase (The FlyBase Consortium, 2003), which are very similar. At the highest weight thresholds, the precision of the MPAs is lower (~85%) than that of the APAs (100%). This is probably on account of the fact that the MPAs have a broader distribution of GO categories than the APA (~2.5 times as many unique categories), making finding a correct GO category by the network annotation method more difficult. In addition, the MPAs have more categories relevant to system-level functions, such as heart development or defense response, which may not be as amenable to this method of annotation as are functions related to individual proteins. These results demonstrate that the neighborhood weighting method can be usefully applied to predicted protein interaction networks.

Approximately 30% of the proteins had at least one prediction with a neighborhood weight >2.5 (corresponding to an estimated precision of 75% and recall of 20%) and ~15% had at least one prediction with a neighborhood weight >5.0 (precision of 83%, recall 11%). Calibration of the neighborhood weight against precision is useful for estimation of the precision of novel network annotations for proteins that have no primary annotations, and allows one to trade genome coverage for precision. For the fly, 326 proteins with no existing APAs could be annotated using this method with a confidence (estimated precision) of 75% or greater, and 96 of these had no annotation assigned by FlyBase either. These examples show the manner in which our approach can be used to provide high-confidence functional predictions for proteins that have not been experimentally characterized, and for which little or no functional information is available. Several of these predictions are listed in Table 2 along with their estimated precision.

Application of network-based annotation to a number of organisms

The neighborhood weighting method is only useful if it can be used on any organism, well characterized or not. Therefore, we applied it to predicted networks from over 50 organisms whose genomes have been sequenced (Table 1). This is possible, since the prediction of interaction networks integrates interaction information from all known networks in a number of organisms. The average precision of the neighborhood weighting method on the predicted and control networks by prediction rank is shown in Figure 3, with error bars representing the standard error of the mean. As shown for the predicted fly network (Fig. 2) higher precision can be reached using the neighborhood weight as a filter. No significant differences were observed in precision averages between organisms or between group averages of eukaryotes, bacteria and archaea, though coverage differences were present. The average recall of the method considering the top three ranked positions was 36% indicating that over one-third of the known annotations were reiterated at this level. The number of proteins with no annotation assigned by the Bioverse and the number of these proteins that could be assigned a GO annotation with an estimated precision of better than 60% in each organism is shown in Table 1. These results are available as part of the Bioverse (http://bioverse.compbio.washington.edu) and as individual data files (http://data.compbio.washington.edu/misc/downloads/nannotation_data/).

Several examples of proteins in the fly and human predicted interaction networks annotated using the neighborhood weighting method (Table 2) are discussed below. The fly network was annotated using only APAs assigned by the Bioverse, whereas the human network was annotated using a combination of APAs assigned by the Bioverse and MPAs from GenBank (Benson et al., 2000) (see Methods section).
### Table 2. Examples of annotations predicted by the neighborhood weighting method

<table>
<thead>
<tr>
<th>Organism</th>
<th>Name</th>
<th>GO predictions</th>
<th>Other evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fly</strong></td>
<td>bio:830 Fbg0002744</td>
<td>Protein amino acid phosphorylation (99%) Protein modification (90%) Ubiquitin cycle (90%) Intracellular signaling (85%)</td>
<td>FlyBase GO: protein metabolism</td>
</tr>
<tr>
<td></td>
<td>bio:1126 Fbg0030062 Fbg001608 Fbg0030438</td>
<td>Nucleic acid binding (82%) ATP-dependent helicase (78%) Proteolysis (85%) Extracellular (65%) Serine endopeptidase (60%)</td>
<td>FlyBase GO: N-acetyltransferase activity, nucleus FlyBase: function unknown</td>
</tr>
<tr>
<td><strong>Human</strong></td>
<td>bio:29482 NP_660316 LOC146894</td>
<td>Immune response (65%) Membrane (63%)</td>
<td>Triggering receptor expressed on myeloid cells</td>
</tr>
<tr>
<td></td>
<td>bio:3143 Fbg0003884</td>
<td>Motor (90%) Microtubule associated (88%)</td>
<td>FlyBase GO: tubulin binding, cell motility, microtubule</td>
</tr>
<tr>
<td></td>
<td>bio:3417 Fbg0037652 Fbg38349 Fbg0034084</td>
<td>2-Component signal transduction (70%) Nucleic acid binding (78%) ATP-dependent helicase (74%)</td>
<td>FlyBase: function unknown</td>
</tr>
<tr>
<td></td>
<td>bio:25888 NP_060121 AHI1</td>
<td>Intracellular signaling cascade (85%) Protein kinase (83%) Peripheral plasma membrane (78%) Oncogenesis (70%)</td>
<td>Function unknown Jouberin; leukemia-causing protein</td>
</tr>
<tr>
<td></td>
<td>bio:28479 NP_473455 TAGAP</td>
<td>GTP binding (85%) Rho small GTPase (83%) GTPase-mediated signal transduction (81%)</td>
<td>T-cell activation Rho GTPase-activating protein</td>
</tr>
<tr>
<td></td>
<td>bio:28269 NP_573444 EPS8L3</td>
<td>Intracellular signaling cascade (84%) Protein tyrosine kinase (83%)</td>
<td>Function unknown EGF receptor pathway substrate 8 related</td>
</tr>
<tr>
<td></td>
<td>bio:22580 NP_112191 KAZALD1</td>
<td>Proteolysis and peptidolysis (87%) Trypsin (87%) Extracellular (90%) Blood coagulation (85%)</td>
<td>Kazal-type serine protease inhibitor Secreted</td>
</tr>
<tr>
<td></td>
<td>bio:22221 NP_573399 OSCAR</td>
<td>Immune response (85%) Integral plasma membrane (83%)</td>
<td>Osteoclast-associated receptor Regulation of innate and adaptive immune responses</td>
</tr>
<tr>
<td></td>
<td>bio:38054 NP_062550 CRTAM</td>
<td>Protein kinase (75%)</td>
<td>Function unknown Class-I MHC-restricted T-cell associated protein</td>
</tr>
<tr>
<td></td>
<td>bio:7708 NP_060413 FBX034 bio:6779 NP_057687</td>
<td>Cytochrome c oxidase (82%) Electron transport (80%) Respiratory chain complex IV (80%) Rho small GTPase (84%)</td>
<td>Function unknown</td>
</tr>
<tr>
<td></td>
<td>bio:8054 Csortf</td>
<td>GTP binding (82%) GTPase-mediated signal transduction (81%)</td>
<td></td>
</tr>
</tbody>
</table>

The predicted protein interaction networks for *D. melanogaster* (fly) or *H. sapiens* (human) were assigned primary annotations. Only automated annotations assigned by the Bioverse were used for the fly network and a combination of automated and manually assigned annotations from GenBank (Benson et al., 2000) were used for the human network. The ‘bio:’ prefix indicates the Bioverse identifier. Predictions (GO categories) were generated using the neighborhood weighting method and the estimation of precision based on neighborhood weight is shown in parentheses. Redundant or commonly occurring (e.g. ATP binding) predictions are not shown. Examples were chosen from proteins with no assigned primary annotations. The ‘Other evidence’ column shows FlyBase (The FlyBase Consortium, 2003) GO annotations or descriptions for fly examples and GenBank descriptions for human examples. Shaded background indicates examples which support the method, white background indicates novel hypotheses generated for the protein.

Although coverage and precision of the combined annotations are similar to those of APAs alone (data not shown) the predictions made are qualitatively different, with a larger distribution of categories (three times as many categories represented) and more annotations related to system-level phenotypes, such as oncogenesis (Table 2, human protein AHI1) and immune function (Table 2, human protein LOC146894).

### DISCUSSION

Protein interaction and functional relationship networks can provide a large amount of information not readily evident from examining functional annotations of individual proteins themselves (Uetz et al., 2000; Date and Marcotte, 2003; Rives and Galitski, 2003; McDermott and Samudrala, 2004). However, experimental determination of such networks requires a large expenditure of time and resources. The combined annotations provide a richer and more comprehensive view of protein function, allowing for the identification of novel hypotheses and the validation of existing knowledge. This approach can be particularly useful in the context of drug discovery, where understanding the functional roles of proteins in complex biological systems is crucial for developing effective therapies. The ability to predict interactions and annotations with high confidence can guide experimental validation efforts, focusing on the most promising hypotheses and reducing the number of experiments required to corroborate findings. Furthermore, the integration of these networks with other types of data, such as expression profiles and disease-related features, can lead to a more holistic understanding of protein function and its impact on health and disease.
resources. To address this issue, we have used a network-based functional annotation method to predict functional annotations for proteins in computationally predicted protein interaction networks (Fig. 1B). Prediction of functional categories based on network context has been shown to perform well on experimentally derived protein networks with manually curated annotations (Schwikowski et al., 2000; Brun et al., 2003; Letovskiy and Kasif, 2003; Samanta and Liang, 2003; Vazquez et al., 2003; Deng et al., 2004). We first showed that the majority-rule method could be successfully applied to GO annotations assigned in an unsupervised manner to the experimental yeast protein interaction network, using the Bioverse computational biology framework. We then showed that the method also performed well when applied to the predicted fly protein interaction network, using either manually or computationally assigned primary annotations (Fig. 2). Results could be significantly improved by incorporating quality scores from predicted interactions and functional assignments to produce a neighborhood weight for each functional category. Therefore, prior to any experimental characterization, a predicted interaction network can be used to functionally characterize a newly sequenced organism from any domain of life, and improve both the coverage and precision of existing functional annotations, automated or manually curated. While not a replacement for experimental investigation and manual curation of annotations, this process provides a useful framework for more careful annotation of the organism, as well as generating a large number of hypotheses that can be tested experimentally.

We used our approach to improve the functional annotations for over 50 organisms from all domains of life. All these predictions and estimates of their precision have been incorporated into the Bioverse and are available at http://bioverse.compbio.washington.edu. Overall, network-based annotation could be performed for 14% (27,368) of the proteins encoded by all the genomes. These annotations are accessible on the Bioverse web server (Table 1; McDermott and Samudrala (2003)). A total of 8296 proteins in the interaction networks have no GO annotation assigned directly to them by our automated functional annotation method. We were able to provide network-based annotations for 2404 (~30%) of these proteins with an estimated precision of >60%.

In Table 2 we show a number of high-quality (high estimated precision) predictions for the fly and human networks for which no primary GO annotation exists. The examples for the fruit fly were generated by applying the neighborhood weighting method to the predicted fly interaction network, using only APAs assigned by the Bioverse. The ‘Other evidence’ column in Table 2 includes some GO annotations assigned by FlyBase. The examples of predictions for human were generated using the combination of APAs and MPAs from GenBank, and none of these examples has primary GO annotations from either source. A precision estimate for each prediction (in parentheses) was derived by extrapolation from the leave-one-out cross-validation (e.g. Fig. 2B). Examples in which the highest weighted predictions were shown to be largely accurate by agreement with the known functional information about the protein (gray background) demonstrate the validity of the method. For these, there are examples of high-quality predictions, but little or nothing is known about the function of the protein from other sources (white background). These, therefore, represent hypotheses that can be used as starting points for experimental verification. In all, 60 fly proteins and 132 human proteins with no GO primary annotations (automated or manual) could be assigned at least one predicted functional category with an estimated precision >65%.

For the fly prediction examples in Table 2, proteins bio:830 and bio:3143 represent examples in which the neighborhood weighting method recapitulated or expanded (bio:830) annotations from FlyBase, a completely independent source. Human proteins LOC146894, TAGAP, KAZALD1 and OSCAR are proteins for which annotation could be assigned by a manual curator based on existing knowledge about the proteins, but have not yet been. AH11 is an example in which the molecular function of the protein is unknown but our network-based annotations reflect a known system-level phenotype, that of oncogenesis. In this example, both molecular- and systems-level annotations are predicted by the method, providing a basis for further study. Examples FLJ30058, EPS8L3, CRTAM, FBX034 and C5orf5 also provide novel hypotheses about proteins for which little or nothing is known. These examples demonstrate the utility of the neighborhood weighting method in providing novel annotations for proteins using predicted protein interaction networks.

The Bioverse is a unique resource that provides a large number of predictions for each protein and associated quality scores for each prediction. The current report shows how these quality scores can be used to improve the integration of disparate types of data to generate novel predictions and precision estimates. Incorporating the quality scores of the primary annotations and the predicted interactions into the network annotation method provides a neighborhood weight for each prediction. We showed in Figure 2A that predictions ranked using weights generated by our method are better correlated with prediction precision than are predictions ranked using the frequency of category occurrence among interaction partners (majority-rule method). The weights of individual predictions also correlate well with prediction precision (Fig. 2B), and this correlation allows the accuracy of novel predictions to be estimated in cases where the protein has no primary annotations. Questionable predictions can, therefore, be further screened on the basis of their neighborhood weights. Additionally, since the set of known functions for any protein is almost certainly incomplete, estimates of precision based on comparison to this incomplete set, if anything, underestimate the true precision. It is likely that some percentage of the ‘false positive’ predictions represent real functions (e.g. Table 2).

A significant strength of this approach for network generation and annotation of uncharacterized organisms is that it integrates a large amount of data from diverse sources, both in terms of methodology and evolutionary origin, to provide a model for the organism (Bork et al., 2004). Since all interaction datasets from any one experimentally characterized organism (including yeast) are incomplete, the use of complementary and overlapping information from multiple sources creates a better model of the interactions in the target organism. By using a factor based on the source of the predicted interaction (i.e. crystallized complex or other method) to adjust the contribution of predicted interaction quality scores to the neighborhood weight we were able to improve coverage and precision of our method. Continuing improvement in the coverage and quality of experimental interaction databases will further enhance the accuracy of this method.

The performance of the network weighting method on predicted protein interaction networks indicates that the predicted interactions, like real protein interactions, tend to have interacting partners that share similar functions and cellular location. Although this has been observed previously (von Mering et al., 2002), it has
not been rigorously evaluated or used to provide predictions and confidence measures. Our method automatically performs what a biologist might do manually when trying to provide a functional annotation for an uncharacterized protein from a novel genome. That is, determining sequence similarity and then using experimental interaction information to predict possible functions for the protein. The approach of combining APA, prediction of protein interaction networks and network-based annotation of the resulting networks, can do this for large numbers of proteins from novel genomes in a fully automated fashion. This improves the chances of success by integrating strong and/or weak relationships to provide an accurate prediction of function. In addition, our method provides a confidence value, an estimate of the precision, for the prediction that can be evaluated by the biologist. Similar to other functional prediction methods, especially those based on sequence comparison, this method alone may not provide completely accurate predictions for the most difficult proteins, such as paralogs. Ultimately all functional predictions require experimental validation. The highest quality annotations we provide, such as those listed in Table 2, provide new hypotheses for very directed experimental work. We envision an iterative process in which this method (along with others) is used to generate hypothetical interactions and functional annotations; the hypotheses are then tested experimentally, and the results are fed back to improve the interaction and functional predictions of other proteins.

We describe the application of a network-based functional annotation method to predicted protein interaction networks. The neighborhood weighting method relies on predicted interaction and functional assignment quality scores to rank results, and produces functional annotation predictions from primary annotations assigned manually or computationally by the Bioverse. The method has been calibrated to provide precision estimates based on comparison with known annotations. It provides functional predictions for a large number of proteins that cannot be assigned a function computationally and augments manual annotation as well (e.g. Table 2). Additionally, it can be applied using any arbitrary descriptive vocabulary and will improve, in both precision and coverage, as the number and quality of experimentally determined protein interactions grows. We applied the neighborhood weighting method to over 50 organisms demonstrating the broad utility of the method in protein annotation; and these results are available on the Bioverse webserver (http://data.compbio.washington.edu/misc/downloads/annotation_data/). In addition, our annotation method, including network-based annotation from predicted interaction networks, has been used to validate annotations for 28,000 rice cDNAs (Kikuchi et al., 2003), as well as to annotate the completed rice genome from the Beijing Genome Institute (Yu et al., 2005). Our results suggest that network-based annotation methods are not only valuable tools for the study of experimentally derived protein interaction networks, but also represent a significant advance in automated genomic annotation of eukaryotes by employing predicted protein interaction networks.

ACKNOWLEDGEMENTS

This work was supported in part by a Searle Scholar Award and NSF Grant DBI-0217241 to R.S., and the University of Washington’s Advanced Technology Initiative in Infectious Diseases.

Conflict of Interest: none declared.

REFERENCES


