Evidence of probabilistic behaviour in protein interaction networks

Supplementary Information

Protein-protein interaction networks

We studied a total of nine PPI networks comprising six unique organisms: Homo sapiens (human), Drosophila melanogaster (fruit fly), Saccharomyces cerevisiae (yeast), Escherichia coli (bacterium), Caenorhabditis elegans (nematode), and Plasmodium falciparum (malaria-causing parasite). In all cases, protein self-interactions were not considered and an undirected nature for the networks was assumed, i.e., an interaction A−B is equivalent to the interaction B−A.

One PPI network of H. sapiens was analyzed and this was downloaded from the Human Protein Reference Database (HPRD, http://www.hprd.org) [1, 2]. This collection of interactions is manually extracted from the literature.

For E. coli and D. melanogaster, one PPI network of each was studied and both were derived from the Database of Interacting Proteins (DIP, http://dip.doe-mbi.ucla.edu) [3], which curates a diverse body of experimentally-determined interactions for a range of organisms. The list of sources is too vast to describe here but includes data from high-throughput methods, including Y2H [4, 5], protein microarrays [6] and mass spectrometric analysis of highly purified multi-protein complexes [7, 8] as well as from
analysis of protein complexes stored in the Protein Data Bank [9]. For *E. coli*, only interactions between proteins belonging to the K12 strain were considered and these were labelled (or relabelled) by their most recent Swiss-Prot Accession Numbers [10]. *D. melanogaster* proteins were labelled (or relabelled) by their most up-to-date GenInfo identifiers using the National Center for Biotechnology Information (NCBI) Sequence Revision History tool (http://www.ncbi.nlm.nih.gov/entrez/sutils/girevhist.cgi). The use of a particular strain and updating of protein labels to their more recent forms ensures that the PPI networks are as biologically consistent as possible.

Three PPI networks of *S. cerevisiae* were analyzed. Two of these, hereafter referred to as *Yeast-DIP* and *Yeast-CORE*, were downloaded from DIP. The *Yeast-CORE* network is a subset of the *Yeast-DIP* network in that the *Yeast-CORE* set is said to include a more reliable set of interactions that have been verified using the Paralogous Verification Method and the Expression Profile Reliability Index [11]. In both cases, proteins were labelled (or relabelled) using up-to-date Swiss-Prot accession numbers and only interactions between proteins of the strain S288C were considered. The third *S. cerevisiae* network used in this work includes only interactions determined by Y2H screens [5, 12] for a variant of the S288C strain [13]. This network is designated as *Yeast-Y2H*.

Two PPI networks of *C. elegans* were used in this study and both were determined from equivalent Y2H screens [14]. The first, labelled *Worm-Y2H*, is a compilation of all identified interactions and the second, labelled *Worm-CORE*, contains a high-confidence subset of interactions that were consistently identified in three repeated experiments.

The final PPI network analyzed here is that of the malaria-causing parasite *P. falciparum*, which was taken from Y2H experimental data [15].
Supplementary References


