Supplementary Material

Supplementary Note 1. Identifying high-quality binary interactions

A binary interaction is a direct, biophysically feasible interaction between two proteins. However, it has been shown that databases that systematically curate protein interactions from the literature could contain erroneous pairs or non-binary interactions (i.e., co-complex associations) (Cusick, et al., 2009). Because it is of paramount importance to identify only the high-quality binary interactions for a variety of biological purposes (Das and Yu, 2012), INstruct obtains only binary interactions from the aforementioned databases and filters them rigorously. PSI-MI evidence codes (Hermjakob, et al., 2004) for each reported interaction from these databases indicate what experiment was used to ascertain the interacting pair. We retained only those interactions indicated as binary by the supporting evidence codes (Supplementary Table 2) because it is only possible to infer interaction interfaces using assays that can detect direct physical interactions (Wang, et al., 2012).

From the aforementioned databases, we compiled a comprehensive list of all publications of high-throughput (HT) experiments (single large-scale studies in which many protein pairs are systematically tested for interactions) that identify protein interactions. Because there are fewer of these studies than conventional small-scale studies, we were able to inspect each publication manually. We only included interactions from those publications whose HT experiments have been verified by traditional orthogonal assays (e.g., co-immunoprecipitation). Additionally, if the authors indicate a subset of their final dataset as "high-quality" or "core," we retained only the interactions meeting the authors' high-confidence criteria. However, for small-scale studies, since it is impossible to manually inspect all papers, we used a well-validated criterion to identify high-quality interactions – it has been shown that interactions supported by two or more publications are of high quality (Das and Yu, 2012; Venkatesan, et al., 2009; Yu, et al., 2008). Using these criteria, we obtained 61,108 high-quality binary interactions for all seven organisms (full stats in Supplementary Table 1).

Supplementary Note 2. Interaction Interface Inference and Validation

Given a pair of interacting proteins for which there exists a co-crystal structure, we integrated the information from both 3did and iPfam to identify their interacting domains. However, the majority of interactions in INstruct do not fall into this category and therefore must be resolved using our interaction interface inference method.

This approach identifies high-confidence domains catalogued by Pfam in each of two interacting proteins. Pfam curates its Pfam-A set of domain families by constructing seed alignments for each family from a nonredundant, functionally verified set of domain-sequences trusted to be representative of the family. Hidden Markov models (currently based on the package HMMER3) are used to grow each domain family from a set of representative seed domains to include closely homologous domains in other proteins. The resulting family of domains is given a single accession identifier of the form PFXXXXX. To ensure the quality of our method, INstruct only annotates a protein with Pfam-A domains, which have been found “in-full and significant” in the protein, subject to Pfam curation criteria. (Punta, et al., 2012)

Once annotated with Pfam-A domains, if proteins in INstruct (which have already been shown to interact in our high-quality binary network) are found to contain domains from families that have been shown to interact in another pair of proteins as indicated by co-crystal evidence in the PDB and catalogued by 3did

or iPfam, then the domain pair is also predicted to be an interface which supports the protein interaction for which no co-crystal evidence is available.

Such parsimonious methods of assigning domain interfaces could potentially result in the prediction of domain-domain interactions that do not facilitate a given protein-protein interaction, however we have verified that our predicted domain-domain interactions are high quality. We performed three-fold cross-validation to verify the reliability of the domain-domain interactions inferred by transferring domain interaction interfaces supported by co-crystal structures to interacting protein pairs without co-crystal structures. Into three subsets we split 1,456 human protein interaction pairs that have co-crystal structures. Using two of the subsets at a time as a training set and one as a test set, we predicted domain-domain interactions in the test set using our comparative modeling approach, rather than taking advantage of the co-crystal evidence. We repeated the procedure using each of the three subsets as the test set in order to ascertain how accurately we could predict interaction interfaces when deprived of co-crystal evidence. We found that we can correctly infer the protein-protein interaction interfaces in over 90% of the 1,456 interaction pairs, indicating high confidence in our method and in the data supplied by 3did and iPfam (Wang, et al., 2012).

References

Cusick, M.E.*, et al.* (2009) Literature-curated protein interaction datasets, *Nat Methods*, **6**, 39-46.

Das, J. and Yu, H. (2012) HINT: High-quality protein interactomes and their applications in understanding human disease, *BMC Syst Biol*, **6**, 92.

Hermjakob, H.*, et al.* (2004) The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data, *Nat Biotechnol*, **22**, 177-183.

Punta, M.*, et al.* (2012) The Pfam protein families database, *Nucleic Acids Res*, **40**, D290-301.

Venkatesan, K.*, et al.* (2009) An empirical framework for binary interactome mapping, *Nat Methods*, **6**, 83-90.

Wang, X.*, et al.* (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease, *Nat Biotechnol*, **30**, 159-164.

Yu, H.*, et al.* (2008) High-quality binary protein interaction map of the yeast interactome network, *Science*, **322**, 104-110.

Query one or more proteins

Select organism

P11413

H. sapiens

Search | Random

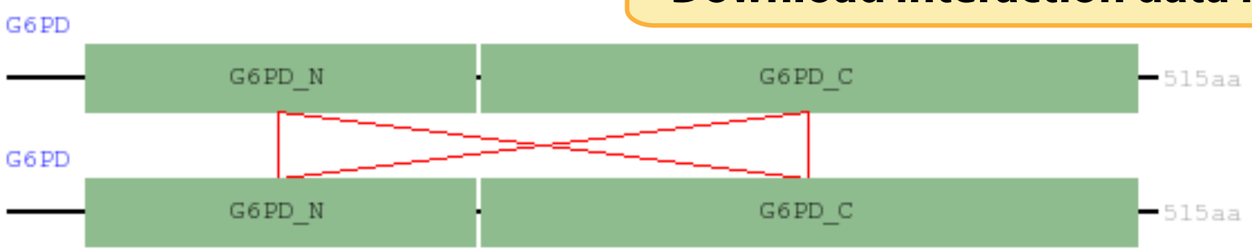
G6PD *(H. sapiens)*

Uniprot: P11413

EntrezID: 2539

Official Symbol: G6PD

Download interaction data for all search terms



Domain on G6PD	Domain on G6PD	Supporting PDB stuctures	Supporting Publications
G6PD_N	G6PD_N	**1qki**	**10745013**, 15858258, 10089300
G6PD_C	G6PD_C	1dpg, **1qki**, **2bhl**	**10745013**, **15858258**, 10089300
G6PD_N	G6PD_C	1dpg, 1e77, 1e7m, 1e7y, 1h93, 1h94, 1h9a, 1h9b, **1qki**, **2bh9**, **2bhl**, 2dpg	**10745013**, **15858258**, 10089300

Download

\t

Click to download this interaction data in tab-delimited format.

HINT

Click here to repeat your search using HINT, a comprehensive database of interactions including ones NOT structurally-resolved.

View all binary interactions

Supplementary Figure 1. Screenshot of the interaction search and retrieval web interface. Shown are the results for a query for the human protein G6PD using its UniProt ID. In this case, the only structurally resolved interaction available is between G6PD and itself. The red edges connecting the domains in this homodimer indicate that the interactions were determined from direct co-crystal evidence; in the table, the PDB structures and publication IDs highlighted in red provide this evidence. PDB structures listed in blue provide homology-based evidence of the domain-domain interaction.

Supplementary Table 1. Interactome network statistics.

	H. sapiens	*A. thaliana*	*C. elegans*	*D. melanogaster*	*M. musculus*	*S. cerevisiae*	*S. pombe*
High-quality binary interactions	27,356	12,068	3,928	4,438	1,222	11,936	160
Domain-domain interactions	11,470	825	180	191	169	1,857	52
Structurally-resolved protein-protein interactions	6,585	644	120	166	119	1,273	37
Unique proteins	3,628	454	144	242	130	978	53

Supplementary Table 2. PSI-MI Accession Codes for assays indicating binary protein interactions

PSI-MI Accession Code	Assay	PSI-MI Accession Code	Assay
889	Acetylation assay	729	Luminescence based mammalian interactome mapping
14	Adenylate cyclase complementation	231	Mammalian protein protein interaction trap
678	Antibody array	515	Methyltransferase assay
8	Array technology	516	Methyltransferase radiometric assay
872	Atomic force microscopy	671	Monoclonal anitbody
10	Beta galactosidase complementation	77	NMR
11	Beta lactamase complementation	81	Peptide array
809	Bimolecular fluorescence complementation	84	Phage display
12	Bioluminescence resonance energy transfer	434	Phosphatase assay
276	Blue native PAGE	841	Phosphotransfer assay
91	Chromatography	953	Polymerization
16	Circular dichroism	435	Protease assay
17	Classical fluorescence spectroscopy	89	Protein array
27	Co-fractionation	90	Protein complementation assay
807	Comigration in gel electrophoresis	31	Protein cross-linking with a bifunctional reagent
404	Comigration in non denaturing gel electophoresis	424	Protein kinase assay
808	Comigration in sds page	55	Resonance energy transfer
405	Competition binding	227	Reverse phase chromatography
25	Copurification	97	Reverse ras recruitment system
28	Cosedimentation in solution	726	Reverse two hybrid
29	Cosedimentation through density gradient	440	Saturation binding
30	Cross-linking studies	99	Scintillation proyimity assay
406	Deacetylase assay	71	Sizing column
870	Demethylase assay	104	Static light scattering
111	dihydrofolate reductase reconstruction	921	Surface plasmon resonance
38	Dynamic light scattering	107	Surface plasmon resonance
40	Electron microscopy	108	T7 phage display
42	Electron paramagnetic resonance	370	Toy-r dimerization assay
410	Electron tomography	232	Transcriptional complementation assay
411	ELISA	20	Transmission electron microscopy
605	Enzymatic footprinting	397	Two hybrid array
415	Enzymatic study	399	Two hybrid fragment pooling approach
47	Far western	398	Two hybrid pooling approach
48	Filamentous phage display	18	Two-hybrid
49	Filter binding	112	Ubiquitin reconstruction
928	Filter binding	113	Western blot
52	Fluorescence correlation spectroscopy	114	X-ray crystallography
416	Fluorescence microscopy	826	Y ray scattering
53	Fluorescence polarization spectroscopy	115	Yeast display
51	Fluorescence technology	825	Y-ray fiber diffraction
54	Fluorescence-activated cell sorting		
728	Gal4 vp16 complementation		
229	Green fluorescence protein complementation assay		
510	Homogeneous time resolved fluorescence		
858	Immunodepleted coimmunoprecipitation		
19	Immunoprecipitation		
492	In vitro binding		
423	In-gel kinase assay		
859	Intermolecular force		
226	Ion eychange chromatography		
65	Isothermal titration calorimetry		
420	Kinase homogeneous time resolved fluorescence		
66	Lambda phage display		
655	Lambda repressor two hybrid		
369	Ley-a dimerization assay		
67	Light scattering		