<table>
<thead>
<tr>
<th>Category</th>
<th>Symbol</th>
<th>Definition of Symbol</th>
<th>Attributes in Category</th>
<th>ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome Occurrence</td>
<td>G(x)</td>
<td>Number of times a particular PART occurs in genome x. (These are based on PSI-blast comparisons between PDB and the genomes with an e-value cutoff in these comparisons of .0001.)</td>
<td>20</td>
<td>(19,31,32)</td>
</tr>
<tr>
<td></td>
<td>L(e)</td>
<td>average expression level over all genes that contain this PART.</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>C(e)</td>
<td>PART composition of the yeast transcriptome in expression level experiment e. This refers to the fraction of the mRNA population with this PART as opposed to all other parts. (This is only applicable to expression experiments, such as SAGE and GeneChips, that measure absolute mRNA levels in copies per cell.)</td>
<td>8</td>
<td>(33)</td>
</tr>
<tr>
<td>Expression</td>
<td>E(e)</td>
<td>Transcriptome enrichment compared to genome in experiment e. (Transcriptome enrichment is defined as percentage difference of PART composition in the transcriptome and the genome. In symbols: E(e) = [C(e)-G(Scer)] / G(Scer) .)</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F(r)</td>
<td>Expression level fluctuation in experiment r. (This is the standard deviation in the expression ratio measurement R(i,t) over a timecourse, viz: &lt;(R(i,t)&lt;R(l,i,t)&gt;)) where one averages over all times t and genes i that have a particular PART.</td>
<td>7</td>
<td>(34)</td>
</tr>
<tr>
<td>Alignments</td>
<td>V(f)</td>
<td>The number of aligned pairs in pair-set f.</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U(f)</td>
<td>RMS deviation in Cα atoms averaged over all alignments in pair-set f.</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R(f)</td>
<td>Similar to U(f) for pair-set f but only the best fitting half of the atoms are included in the calculation</td>
<td>2</td>
<td>(35)</td>
</tr>
<tr>
<td></td>
<td>S(f)</td>
<td>Average percentage identity between pairs of aligned proteins in pair-set f</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P(f)</td>
<td>Average sequence P-value for pair-set f</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Q(f)</td>
<td>Average structural P-value for pair-set f</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Compositions</td>
<td>N(p)</td>
<td>The number of structures associated with a particular PART in dataset p.</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B(a,p)</td>
<td>Composition of amino acid a in a particular PART where one averages over all structures in dataset p associated with the PART</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Motion</td>
<td>M(s,d)</td>
<td>The maximum value of statistic s derived from surveying set of motions d in the Macromolecular Motions Database for a particular PART, where s is only calculated from the entries in the database that are associated with the PART.</td>
<td>7</td>
<td>(36,37)</td>
</tr>
<tr>
<td></td>
<td>A(s,d)</td>
<td>Similar to M(s,d) but now we take the average instead of the maximum.</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>I(y,c)</td>
<td>For a given PART, the number of types of protein-protein interactions in interaction dataset y subject to the restriction c regarding whether or not the proteins are on the same chain. The number of interaction types is the number of distinctly different PARTs that interacts with a given PART .</td>
<td>24</td>
<td>(38,39)</td>
</tr>
<tr>
<td></td>
<td>J(y,c)</td>
<td>For a given PART, the total number of types of interactions in interaction dataset y subject to the restriction c regarding whether or not the proteins are on the same chain. Here we show all interactions observed not just the number of distinct PART-PART interactions tabulated in I(y,c).</td>
<td>24</td>
<td>(38,39)</td>
</tr>
<tr>
<td>Transposon</td>
<td>T(b)</td>
<td>The sensitivity of the cell to a transposon inserted into genes containing a particular PART under different growth condition b. The sensitivity was indicated by negative logarithm of a P-value, which measures the degree to which the observations for one particular gene could have resulted from wild-type cells that randomly change their phenotype.</td>
<td>20</td>
<td>(40)</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>X(q)</td>
<td>Various miscellaneous ranks</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>182</td>
</tr>
<tr>
<td>Attributes</td>
<td>Value</td>
<td>Description</td>
<td>ref.</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-------------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td><strong>x</strong> =</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aful</td>
<td>Archaeoglobus fulgidus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mjan</td>
<td>Methanococcus jannaschii</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mthe</td>
<td>Methanobacterium thermoautotrophicum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phor</td>
<td>Pyrococcus horikoshii</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scer</td>
<td>Saccharomyces cerevisiae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cele</td>
<td>Caenorhabditis elegans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aseo</td>
<td>Aquifex aeolicus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>syne</td>
<td>Synechocystis sp.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecol</td>
<td>Escherichia coli</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bsub</td>
<td>Bacillus subtilis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mtub</td>
<td>Mycobacterium tuberculosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hinf</td>
<td>Haemophilus influenzae Rd</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hpyl</td>
<td>Helicobacter pylor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mgen</td>
<td>Mycoplasma genitalium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mpne</td>
<td>Mycoplasma pneumoniae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bbur</td>
<td>Borrelia burgdorferi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tpal</td>
<td>Treponema pallidum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ctra</td>
<td>Chlamydia trachomatis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cpne</td>
<td>Chlamydia pneumoniae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpro</td>
<td>Rickettsia prowazeki</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>e</strong> =</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vegsam</td>
<td>GeneChip mRNA expression analysis of 5200 yeast ORFs under vegetative growth conditions.</td>
<td>(41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vegyou</td>
<td>GeneChip mRNA expression analysis of 5455 yeast ORFs under vegetative growth conditions.</td>
<td>(42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sage</td>
<td>mRNA expression analysis of 3788 yeast ORFs determined by Serial Analysis of Gene Expression.</td>
<td>(43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>matea</td>
<td>GeneChip mRNA expression analysis of yeast mating type a strain grown on glucose.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mateal</td>
<td>GeneChip mRNA expression analysis of yeast mating type alpha strain grown on glucose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gal</td>
<td>GeneChip mRNA expression analysis of yeast mating type a strain grown on galactose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>heat</td>
<td>GeneChip mRNA analysis of yeast mating type a strain grown on glucose at 30 degree before a 39 degree heat shock.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ref</td>
<td>Reference transcriptome. This is a scaling and merging of the above experiments.</td>
<td>(33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>r</strong> =</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdc28</td>
<td>cDNA microarray genome-wide characterization of mRNA transcript levels for CDC28 synchronized yeast cells during the cell cycle.</td>
<td>(45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdc15</td>
<td>cDNA microarray genome-wide characterization of mRNA transcript levels for CDC15 synchronized yeast cells during the cell cycle.</td>
<td>(45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alpha</td>
<td>Analysis using cDNA microarrays of yeast mRNA levels after synchronization of cell cycle via alpha arrest factor.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diaux</td>
<td>Genome-wide cDNA microarray analysis of the temporal program of yeast mRNA expression accompanying the metabolic shift from fermentation to respiration.</td>
<td>(46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spor</td>
<td>cDNA microarray genome-wide analysis to assay changes in gene expression during sporulation.</td>
<td>(47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>heaterc</td>
<td>cDNA microarray experiment and analysis on 4290 E.coli ORFs after exposure of the bacteria to heat shock.</td>
<td>(48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>deve</td>
<td>Analysis of genome wide changes during successive larval stages using cDNA microarrays of ~12000 C. elegan ORFs.</td>
<td>(49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>f</strong> =</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>All pairs within a PART included in the calculations in Wilson et al. (For example, for fold rankings this would be the total number of pairs within a fold.)</td>
<td>(35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>foldonly</td>
<td>A subset of the pair-set &quot;all&quot; that only includes pairs between structures that are in the same PART but different sub-PART. (If PART is fold, then sub-PART is superfamily; If PART is superfamily, then sub-PART is family.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>a</strong> =</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pdb100</td>
<td>All structures within the fold (as defined by SCOP pdb100d)</td>
<td>(50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pdb40</td>
<td>Similar to pdb100 but now using a version of the PDB clustered at 40% similarity (as defined by SCOP pdb40d)</td>
<td>(50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributes</td>
<td>Value</td>
<td>Description</td>
<td>ref.</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-------------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td><strong>Interaction type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>y=</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction restriction</td>
<td>c=</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intra</td>
<td>The interaction must occur between PARTs in different chains</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>none</td>
<td>The union of &quot;inter&quot; and &quot;intra&quot;. Interactions can occur in PARTs on the same or different chains.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motion statistic</strong></td>
<td>s=</td>
<td></td>
<td>(36,37)</td>
<td></td>
</tr>
<tr>
<td>nresidue</td>
<td>Number of residues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rmsoverall</td>
<td>Overall RMS of two structures after they are superimposed by a sieve-fit technique. Note that they are larger than traditionally used RMS (details see ref.).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rhinges</td>
<td>Number of hinges involved in the motion.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kappa</td>
<td>Torsion energy of the motion (maximum energy less minimum energy over the motion) (in kcal/mole).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trans</td>
<td>Absolute difference in energy between one mRNA and one tRNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deltaxe</td>
<td>Distance between atoms in the coordinate files. Five or more contacts between atoms separated by less than 5 Å was considered a valid PART-PART contact.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motion dataset</strong></td>
<td>d=</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>auto</td>
<td>List of ~220 &quot;gold-standard&quot; manually curated motions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Transposon conditions</strong></td>
<td>b=</td>
<td></td>
<td>(40)</td>
<td></td>
</tr>
<tr>
<td>benr</td>
<td>Benomyl resistance: YPD + 20µg/ml benomyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bcp1</td>
<td>YPD + 5-bromo-4-chloro-3-indolyl phosphate at 37°C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mb</td>
<td>YPD + 0.01% methylene blue at 37°C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>egta</td>
<td>YPD + 2m EGTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mns</td>
<td>YPD + 0.008% MMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bens</td>
<td>Benomyl hypersensitivity: YPD + 1µg/ml benomyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bens</td>
<td>YPD + 0.003%SDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ben</td>
<td>Benomyl hypersensitivity: YPD + 10µg/ml benomyl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yp37</td>
<td>YPD at 37°C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Misc. quantities</strong></td>
<td>q=</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pseudogenes</td>
<td>Number of pseudogenes in worm genome matching a particular PART</td>
<td>(53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>func</td>
<td>Total number of functions associated with this PART (the survey on non-enzyme functions were lumped into a single category).</td>
<td>(54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>enz</td>
<td>Total number of enzymatic functions associated with this PART.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>size</td>
<td>Average length of a PART in the pdb404d clustering of the PDB.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>The year of the first structure that is part of the PART was determined.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>