Computing the g-dependent interfacial polarization

In this section, we derive an auxiliary analytic result that enables us to rigorously compute the protein-water interfacial tension (PWIT), as described in Methods. The physical analysis is based on two local parameters, the function $g(r)$ representing the time-averaged coordination (number of hydrogen bonds) associated with a water molecule at position $r$ (Figure 1a), and $P=P(r)$, the polarization field generated by water dipoles as they interact with net charges on the protein surface or with the resultant electrostatic field arising from confinement of interfacial water. In this section we obtain the g-dependence of $P$; that is, we show that $P(r)$ is actually $P(g(r))$.

In order to determine the g-dependence of $P$, we introduce the electrostatic potential $\Psi=-\Phi(r)$, with $E=\nabla \Phi$. Thus, the Debye equation (Eq. 4, Methods):
\[ \nabla \left[ \int F^{-1}(K)(r-r')E(r')dr' \right] = \rho(r) \quad (S1) \]

reads in \( \omega \)-space\(^1\):

\[ F(\Phi)(\omega) = F(\rho)(\omega)/[|\omega|^2K(\omega)] , \quad (S2) \]

where \( |\omega|^2 \) is the Fourier conjugate of \( \nabla^2 \). Thus, \( E \) is obtained from the inverse Fourier transformation of Eq. S2:

\[ E(r) = \nabla F^{-1}\{ F(\rho)(\omega)/[|\omega|^2K(\omega)]\}(r) = (2\pi)^{3/2} \nabla \int d\omega e^{-i\omega \cdot r} F(\rho)(\omega)/[|\omega|^2K(\omega)] \quad (S3) \]

Given a charge density distribution of the form:

\[ \rho(r) = \sum_{m \in L} 4\pi q_m \delta(r-r_m) , \quad F(\rho)(\omega) = (2\pi)^{3/2} \sum_{m \in L} 4\pi q_m \exp(i\omega \cdot r_m) , \quad (S4) \]

with \( L \)=set of charges on the protein surface labeled by index \( m \), we obtain:

\[ \Phi(r) = (2\pi)^{-3} \sum_{m \in L} \int d\omega e^{-i\omega \cdot (r-r_m)} 4\pi q_m/[|\omega|^2K(\omega)] , \quad (S5) \]

which gives the explicit form of the polarization:

\[ P(r) = \int F^{-1}(K_p)(r-r')E(r')dr' = \]

\[ (2\pi)^{-3} \sum_{m \in L} \int d\omega e^{-i\omega \cdot (r-r_m)} 4\pi q_m/[|\omega|^2K(\omega)] , \quad (S6) \]

with \( F^{-1}(K_p)(r) = \int d\omega e^{i\omega \cdot r} ([\epsilon_b-\epsilon_o]/[1+(\tau(r)c)^2]|\omega|^2]) \).

**Protein-water interfacial tension is a molecular determinant of protein interactions**

In this section, we analyze protein-protein (P-P) interfaces of complexes with reported structure and, by comparing P-P interfaces with protein-water (P-W) interfaces, we show that protein associations are promoted by regions on the
surface of the free subunits that generate unstable P-W interfaces. The analysis upholds the view that PWIT is a molecular indicator of complexation propensity.

The PWIT of a protein with stable soluble structure is determined by the position \(r\)-dependent parameter \(g=g(r)\). Thus, the solvent accessible envelope of the protein surface\(^2,3\) may be covered by a minimal set \(W\) of water-confining osculating (first-order contact) spheres \(D_j\), \(j \in W\), each with its average value \(g=g_j\), which depends on the physicochemical nature of protein group(s) in the contact region.

We now unravel the nature of interfacial hot spots. Interfacial tension arises in \(D_j\) when \(\Delta G_j > 0\), where \(\Delta G_j\) is the interfacial surface tension associated with spanning contact region \(j\), computed using Eq. 2 with the integral restricted to the domain \(D_j\). To validate our results, we examined free subunits in 28 protein complexes (Supplementary Table 1) with defined contact topologies. The most common patches generating interfacial tension involve SABHBs, that is, backbone hydrogen bonds of the protein that are partially exposed to water so that the amide and carbonyl are hydrated by low density water\(^4\) (Fig. 1a). Thus, the most common osculating spheres contact polar patches arising from incomplete shielding of the backbone hydrogen bonds. These spheres constitute \(~73\pm5\%) of the interfacial spheres that fulfill \(\Delta G_j > 0\) in the database of free subunits from the 28 complexes. They are described by the range \(g_j=2.5\pm0.5\), yielding \(\Delta G_j=(1.54\pm0.40)k_BT\). Furthermore, the correlation between the PWIT for the free subunits and their structural deficiency \(\nu\), defined by the ratio of SABHBs
to BHBs, implies that SABHBs are a determinant of interfacial tension (Supplementary Fig. 1).

Nonpolar interfacial cavities with curvature radius $\theta \geq 7\text{Å}$ yield the highest interfacial tension at $\Delta G_j \approx -T \Delta S_j = 13.17 k_B T$. The same contribution is associated with each disjoint $\theta = 7\text{Å}$-sphere covering convex regions ($\theta < -2.7\text{Å}$). By contrast, the "clathrate" range $0 \geq \theta \geq -2.7\text{Å}$, where nonpolar moieties can be hydrated while preserving the tetrahedral hydrogen-bond lattice ($g=4$) do not generate tension, as expected.

To support the intuitively obvious claim that PWIT is a determinant of the propensity of the protein to associate, we examined free subunits in 28 protein complexes (Supplementary Table 1) with defined contact topologies. The total area of the P-P interfaces for each complex is computed after identification of the residues in each subunit that are engaged in intermolecular contacts. On the other hand, the protein-water (P-W) interfacial tension of each free subunit is computed by numerical integration of Eqs. 2, 6 with charge and atomic radii assigned using the program PDB2PQR. We define filtered sets $W_n = \{j \in W: \Delta G_j \geq nk_BT\}$ $(n=1,2,...)$ of contributors to the P-W interfacial tension in the free subunits $(W_n \subset W_{n'}$ for $n > n')$, with $S_n$, the total P-W interface associated with $W_n$. In all 28 complexes examined, the total P-P interface distributed among the subunits within the complex is 100% contained in $S_2$, and it is only 60±7% contained in $S_3$.

These results suggest that interfacial solvent cavities spanned at a thermodynamic cost $\Delta G \geq 2k_B T$ are the promoters of protein associations. This
conclusion is confirmed by the tight correlation ($R^2=0.83$) between $A_2$, the surface area of $S_2$, and the total P-P interfacial area of each complex (Supplementary Fig. 2a). This correlation becomes weaker for $S_1$ with area $A_1$ ($R^2=0.60$) and nearly vanishes ($R^2=0.09$) when we consider (as control) the total solvent-exposed surface area (SESA) of free subunits.

Hot spots in P-P associations were identified by scanning the interface through site-specific substitution of amino acids for alanine (effectively a side-chain truncation). The most extensively scanned interface corresponds to the 1:1 human growth hormone (hGH)-receptor complex$^7$ (Supplementary Fig. 2b). To validate our molecular marker, the hot spots obtained by alanine scanning were contrasted with an in-silico shaving procedure that determines the reduction $\Delta \Delta G_{if} = \Delta G_{if}(m) - \Delta G_{if}(0)$ in interfacial tension introduced by site-directed mutation ($0$=wild type $\rightarrow$ m=mutation). The latter materializes as side-chain truncation at the $\beta$-carbon (equivalent to alanine substitution). A strong correlation exists between $-\Delta \Delta G_{if}$ and the experimental association free energy difference $\Delta \Delta G_a$ produced by substitution of interfacial residues (Supplementary Fig. 2c). The interfacial scanning results validate the nanoscale-thermodynamics derivation of PWIT and upheld the view that this physical attribute is a selective promoter of protein associations.
**Supplementary Table 1.** Exhaustive catalog of representative topologies of inter-subunit contacts for protein complexes with up to six subunits reported in the protein data bank (PDB). The topology of the inter-subunit contact pattern for a complex is given by a graph where a ball represents a protein subunit and a stick between two balls represents a protein-protein (P-P) interface. The PDB accession codes for 28 nonhomologous complexes illustrating the different topologies are indicated by their PDB-accession codes (four alphanumeric characters). Each PDB entry designates the crystal structure of a protein complex. The species represented pertain to the entire spectrum of population sizes, from virus (1htg) to human (1dvz). The numbers in parentheses denote % decrease in overall PWIT upon complexation starting with free subunits. The dataset is used to unravel the role of P-W interfacial tension patches as promoters of protein associations.

| Topology | PDB Accession | Topology | PDB Accession | Topology | PDB Accession |
|----------|---------------|----------|---------------|----------|---------------|
| ![Topology 1](image1.png) | 2af7(63) | ![Topology 2](image2.png) | 1fr3(61) | ![Topology 3](image3.png) | 2siv(59) |
| ![Topology 4](image4.png) | lub4(53) | ![Topology 5](image5.png) | 2alb(51) | ![Topology 6](image6.png) | 2c5r(40) |
| ![Topology 7](image7.png) | 4otc(57) | ![Topology 8](image8.png) | 1qoh(54) | ![Topology 9](image9.png) | 1dvz(57) |
| ![Topology 10](image10.png) | 1jtk(55) | ![Topology 11](image11.png) | 1hho(64) | ![Topology 12](image12.png) | 1nys(39) |
| ![Topology 13](image13.png) | 1uo5(54) | ![Topology 14](image14.png) | 1vwe(53) | ![Topology 15](image15.png) | 1fvu(50) |
| ![Topology 16](image16.png) | 1zij(52) | ![Topology 17](image17.png) | 1lufy(51) | ![Topology 18](image18.png) | 1wud(42) |

1u4j(48) 1hpo(56) 1hvg(50) 1lve(47) 2oqq(60) 1v2z(51) 1btn(62) 2ien(59) 1ihb(57) 1p6z(72) 3orc 1b8x 2gfh 1goc 2b65 1h5m 1may 1uga 1quz
Supplementary Figure 1. Correlation between structural deficiency (\(v\), in percent) and protein-water interfacial tension for the free subunits from the 28 complexes in the database indicated in Supplementary Table 1. In the well-curated yeast (\(S.\ cerevisiae\)) interactome, structural deficiency has been shown to be a molecular indicator of interactivity\(^8\), implying that PWIT is also a good marker for complexation propensity.
**Supplementary Figure 2.** Protein-water interfacial tension promotes biomolecular associations. 

**a.** Total area of the surfaces that generate protein-water (P-W) interfacial tension in the free subunits (uncomplexed state) plotted against the total protein-protein (P-P) interfacial area for 28 protein complexes (Supplementary Table 1). The total surface area of cavities with PWIT ≥ k_B T (region A₁, k_B T = baseline energy of thermal fluctuations) correlates with the P-P interfacial area with R^2 = 0.60. This correlation improves to R^2 = 0.83 when we consider region A₂, comprised only of cavities generating higher interfacial tension (PWIT ≥ 2k_B T), and virtually vanishes (R^2 = 0.09) when we consider simply the solvent-exposed surface areas (SESAs).

**b.** The P-P interface between the human growth hormone (hGH) and its receptor in the 1:1 complex. 

**c.** Comparison with experimental alanine scanning of the P-P interface of the hGH receptor validates the interfacial tension computation. Correlation between association free energy differences ΔΔG_a between mutant (m) and wild type (0)
for the alanine substitution of each residue from the hGH receptor at the P-P interface and the interfacial free energy difference between wild-type and mutant for the free protein subunit (uncomplexed hGH receptor). The association free energy is computed as $\Delta G_a = -RT\ln K_a = RT\ln(K_d)$, where $K_a$, $K_d$ are the association and dissociation equilibrium constants, respectively. The hot-spot residues (biggest contributors to affinity) obtained by alanine scanning were contrasted with an in-silico shaving procedure that gives the change in PWIT ($-\Delta \Delta G_{if}$) introduced by a side-chain truncation at the $\beta$-carbon. A strong correlation exists between $-\Delta \Delta G_{if}$ and the association free energy difference $\Delta \Delta G_a$ generated by site-specific substitution of interfacial residues.

**PWIT is an indicator of complexation propensity: Orthologs with different levels of homo-oligomerization**

In this section, we compare the level of homo-oligomerization for pairs of orthologous proteins in order to demonstrate that the PWIT of the free protein subunit is the determinant of the oligomerization complexity (Fig. 1b).

|                | D2 symmetry | C2 symmetry | seq. identity | P-P interface ratio | PWIT ratio |
|----------------|-------------|-------------|---------------|---------------------|------------|
| 3mds           | 1gv3        | 58          | 86%           | 73%                 |
| 3pgm           | 1 e58       | 48          | 81%           | 88%                 |
| 1o58           | 1d6s        | 52          | 78%           | 71%                 |
| 1nl            | 1mjf        | 35          | 78%           | 48%                 |
| 1uwt           | 1cbg        | 24          | 75%           | 80%                 |
| 1rli           | 1x77        | 30          | 58%           | 65%                 |
| 1non           | 1a3c        | 73          | 55%           | 50%                 |
| 1eyi           | 1dk4        | 30          | 46%           | 34%                 |
| 1j2r           | 1nf9        | 27          | 43%           | 56%                 |
| 1ub3           | 1o0y        | 45          | 38%           | 51%                 |
| 1vjp           | 1j5p        | 56          | 37%           | 31%                 |
| 1x94           | 1nri        | 26          | 31%           | 27%                 |
| D3 symmetry | C2 symmetry | seq. identity | P-P interface ratio | PWIT ratio |
|------------|------------|--------------|---------------------|------------|
| 1pjc       | 1qp8       | 36           | 91%                 | 70%        |
| 1tzf       | 1vgz       | 36           | 82%                 | 72%        |
| 1u1z       | 1mka       | 27           | 73%                 | 73%        |
| 1cks       | 1puc       | 51           | 68%                 | 71%        |
| 1pmm       | 1ajr       | 22           | 65%                 | 50%        |
| 1nw4       | 1jys       | 28           | 62%                 | 60%        |
| 1a3g       | 1daa       | 29           | 60%                 | 67%        |
| 1tqj       | 1h1y       | 44           | 50%                 | 57%        |
| 1on3       | 1od2       | 21           | 50%                 | 46%        |
| 1kr2       | 1kam       | 40           | 34%                 | 42%        |
| 1nqb       | 1moe       | 69           | 18%                 | 15%        |
| 1k3p       | 1aj8       | 36           | 15%                 | 11%        |
| 1p8c       | 1knc       | 37           | 76%                 | 69%        |
| 1pi2       | 1k9b       | 67           | 50%                 | 57%        |
| 1mkz       | 1uuy       | 32           | 42%                 | 37%        |
| 1uiy       | 1jxz       | 34           | 39%                 | 41%        |
| 1t3d       | 1xat       | 39           | 14%                 | 17%        |

**Supplementary Table 2.** PDB-reported ortholog proteins with different levels of homo-oligomerization and different complex symmetry. The ratio of P-P interfaces (small to large)$^9$ and the ratio of PWITs for the respective free subunits are given for pairs of orthologous proteins with different levels of oligomerization.

**Species Key**
| PDB Complex | Species                     | PDB Complex | Species                     |
|-------------|-----------------------------|-------------|-----------------------------|
| 3mds        | Thermus thermophilus        | 1gva        | Anabaema                    |
| 3pgm        | S. cerevisiae               | 1e58        | E. coli                     |
| 1058        | Thermotoga maritima         | 1d6s        | Salmonella typhimorium      |
| 1inl        | Thermotoga maritima         | 1mjf        | Pyrococcus furiosus         |
| 1uwt        | Sulfolobus solfataricus     | 1c8g        | Trifolium repens            |
| 1rli        | Bacillus subtilis           | 1x77        | Pseudomonas aeruginosa      |
| 1non        | Bacillus caldolyticus       | 1a3c        | Bacillus subtilis           |
|             |                             |             | Methanocaldococcus jannaschii|
| 1eye        | Sus scrofa                  | 1dk4        | Pseudomonas aeruginosa      |
| 1j2r        | E. coli                     | 1nf9        | Pseudomonas aeruginosa      |
| 1ub3        | Thermus thermophilus        | 1o0y        | Thermotoga maritima         |
| 1vjp        | Thermotoga maritima         | 1j5p        | Thermotoga maritima         |
| 1x94        | Vibrio cholerae             | 1nri        | Haemophilus influenzae       |
| 1a16        | E. coli                     | 1pv9        | Pyrococcus furiosus         |
| 1sjw        | Streptomyces nogalater      | 1ocv        | Pseudomonas aeruginosa      |
| 1sru        | E. coli                     | 1se8        | Deinococcus radiodurans     |
|             |                             |             | Geobacillus                 |
| 1b9b        | Thermotoga maritima         | 2btm        | stearothermophilus           |
|             |                             |             | Geobacillus                 |
| 1bj4        | Homo sapiens                | 1kkn        | stearothermophilus           |
| 1qsm        | S. cerevisiae               | 1bo4        | Serratia marcescens         |
| 1f8w        | Enterococcus faecalis       | 1lvi        | Pseudomonas putida           |
| 1pjc        | Phormidium lapideum         | 1qp8        | Pyrobaculum aerophilum      |
| 1tf8        | Salmonella enterica         | 1vqz        | Rieisseria gonorrhoeae      |
|             | Pseudomonas aeruginosa      |             |                             |
| 1u1z        | E. coli                     | 1mka        | Schyzosaccharomyces pombe    |
|             |                             |             |                             |
| 1cks        | Homo sapiens                | 1puc        |                             |
| 1pmm        | E. coli                     | 1ajr        | Sus scrofa                  |
|             |                             |             | E. coli                     |
| 1nw4        | Plasmodium falciparum       | 1jys        |                             |
| 1a3g        | E. coli                     | 1daa        | Bacillus sp. ym-1           |
| 1tqj        | Synechocystis sp. Propionibacterium | 1h1y | Oryza sativa                 |
|             | frederichii                 |             |                             |
| 1on3        | E. coli                     | 1od2        | S. cerevisiae               |
| 1kr2        | Homo sapiens                | 1kam        | Bacillus subtilis           |
| 1nqb        | Mus musculus                | 1moe        | Mus musculus                |
| 1k3p        | E. coli                     | 1aj8        | Pyrococcus furiosus         |
| 1p8c        | Thermotoga maritima         | 1knc        | Mycobacterium tuberculosis  |
| 1pi2        | Glycine max                 | 1k9b        | Glycine max                 |
| 1mkz        | E. coli                     | 1uuy        | Arabidopsis thaliana        |
| 1uiy        | Thermotoga maritima         | 1jxg        | Pseudomonas sp.              |
| 1t3d        | E. coli                     | 1xat        | Pseudomonas aeruginosa      |
Accuracy of the homology-based prediction

In this section, we test the homology-based prediction of PWIT for proteins that have orthologs with reported structure. Using as test cases proteins with reported structures possessing orthologs that can serve as homology templates, we determine the accuracy of the prediction as a function of the level of sequence identity between test and template.

Supplementary Figure 3. Accuracy of the homology-based prediction (units in both axis are percentages). A testing set was constructed based on the fact that each of the 106 ortholog groups across the 36 species examined has two PDB representatives for 2 different species (the species are typically *E. coli* and *H. sapiens*, the most representative species in PDB). Thus, the testing set consists of 106 proteins with PDB-reported structures and their respective PDB-reported orthologs were used as templates to build 106 homology models that were contrasted against the actual PDB structures. The accuracy of the homology model was assessed by computing its PWIT, denoted $\Delta G_{\text{if}}(\text{hom})$, and comparing it with the value $\Delta G_{\text{if}}$ obtained from the actual protein structure. Thus, the accuracy of the models was determined by plotting the relative deviation
$|\Delta G_{\text{dd}}(\text{hom})-\Delta G_{\text{dd}}|/\Delta G_{\text{dd}}$ ($\Delta G_{\text{dd}}>0$ for any PDB structure) expressed as % against the % sequence identity between the test protein and its orthologous template. Error bars indicate dispersions across pairs of proteins with identical % sequence identity. The results reveal that a sequence identity larger than 35% guarantees over 90% accuracy of the model.
Relative structural deficiency of orthologous proteins

In this section we establish the statistical fact that the extent of structural degradation of orthologous proteins correlates negatively with the population size of the species to which the orthologs belong.

Supplementary Figure 4. Average structural deficiency across species relative to *E. coli*, given as %. The structural deficiency $S_n$ for species $n$ relative to *E. coli* is defined as $S_n = \langle S_{j,n}\rangle = \langle (v_{j,n} - v_{j,1})/v_{j,1}\rangle_j$, where $v_{j,n} = \nu$-value of protein from species $n$ in ortholog group $j$ ($j=1,2,\ldots,106$) and the average is taken over all 106 ortholog groups.
Sequence-based identification of SABHBs

In this section, we show that the structure degradation parameter \( \psi \) can be determined from the protein sequence through a tight correlation between the extent of shielding of the backbone on a particular region of the chain and the disorder propensity of the region.

![Supplementary Figure 5](image)

**Supplementary Figure 5.** Correlation between the disorder score of a residue (\( f_d \)) and the extent of protection (\( \zeta \)) of the backbone hydrogen bond engaging that particular residue (if any). The parameters are defined in Methods. The disorder score on each individual residue was obtained for 2,806 nonredundant nonhomologous PDB domains. Residues have been independently grouped in 45 bins of 400 residues each, according to the extent of protection (\( 7 \leq \zeta \leq 52 \)). The average score has been determined for each bin (square), and the error bars represent the dispersion of disorder scores within each bin. The strong correlation between the disorder score and the extent of wrapping and the dispersions obtained imply that SABHBs can be safely inferred in regions where the disorder score is \( 0.35 \leq f_d \leq 0.95 \) (actually, \( f_d = 0.8 \) is the realistic threshold for disorder, hence the more realistic range is \( 0.35 \leq f_d \leq 0.8 \)). The rectangle represents the region of order-disorder intermediate region where the existence of SABHBs (\( 7 \leq \zeta \leq 19 \), for desolvation radius 6Å) may be inferred from disorder.
score with absolute certainty. No hydrogen bond in monomeric domains reported in PDB was found to possess less than 7 protectors, implying a threshold for structural sustainability in soluble proteins.

**Cross-validation of disorder-based and homology-based inference of v-values**

In this section, we validated the disorder-based estimation of v-values by comparing them with homology-based inferences.

![Graph](image)

**Supplementary Figure 6.** Cross-validation of disorder-based and homology-based inference of v-values (units in both axis are percentages). To cross-validate the homology-based and sequence-based v-value prediction, an exhaustive set of 2,792 homology models for the 106 orthologous groups considered was constructed using the 325 PDB-reported templates (Supplementary Tables 3, 4). Each group has on average 32 orthologous proteins. The v-value for a homology model structure, denoted v(hom) was compared with v(dis), the v-value obtained directly from the sequence-based disorder prediction (Supplementary Figure 5) for different levels of sequence identity between template and query protein. The results reveal that the homology-based inference can be assessed using the disorder score prediction with 85% reliability for >35% sequence identity.
Relative partial disorder across 105 species

In this section we compare the disorder-based inferences of structural degradation for orthologous proteins across 105 species to corroborate the broad pattern of small-population-high-degradation.

Supplementary Figure 7. Relative partial disorder across 105 species. The original set of 106 ortholog groups with PDB representatives was expanded to a set of 541 ortholog groups (most of which now lack PDB representation), so that each group contains proteins in at least 100 of the 105 species indicated below. The $\nu$-value for each ortholog protein was determined from the sequence-based disorder score (Supplementary Figure 5) by identifying regions of partial disorder and the relative protein-complexation propensity for species $n$ was determined as $S_n = \langle S_{i,n} \rangle_j = \langle (v_{j,n} - v_{j,1})/v_{j,1} \rangle_j$, where $v_{j,n} = \nu$-value of protein from species $n$ in ortholog group $j$ ($j=1,2,\ldots,541$). The relative protein-complexation propensity is
shown to increase with decreasing population size, in qualitatively agreement with the trend shown in Figure 2a. The species numbering is as follows:

| species number | species name                        |
|----------------|-------------------------------------|
| 1              | Tetrahymena thermophila             |
| 2              | Haloferax volcanii                  |
| 3              | Salmonella enterica                 |
| 4              | Escherichia coli                    |
| 5              | Yersinia pestis                     |
| 6              | Cyanidioschyzon merolae             |
| 7              | Vibrio cholera                      |
| 8              | Chlamydomonas reinhardtii           |
| 9              | Campylobacter jejuni                |
| 10             | Cryptosporidium hominis             |
| 11             | Cryptosporidium parvum              |
| 12             | Ostreococcus tauri                  |
| 13             | Coccidioides immitis                |
| 14             | Plasmodium falciparum               |
| 15             | Plasmodium vivax                    |
| 16             | Aspergillus fumigatus               |
| 17             | Theileria annulata                  |
| 18             | Theileria parva                     |
| 19             | Saccharomyces cerevisiae            |
| 20             | Schizosaccharomyces pombe           |
| 21             | Ustilago maydis 521                 |
| 22             | Magnaporthe grisea 70-15            |
| 23             | Monosiga brevicollis                |
| 24             | Cryptococcus neoformans             |
| 25             | Trypanosoma cruzi                   |
| 26             | Batrachochoytrium dendrobatidis      |
Yarrowia lipolytica
Entamoeba histolytica
Kluyveromyces lactis
Fusarium graminearum FG3
Puccinia graminis f. sp. tritici
Neurospora crassa
Stagonospora nodorum SN15
Candida albicans SC5314
Candida glabrata
Rhizopus oryzae
Phytophthora sojae
Trichomonas vaginalis
Giardia lamblia
Debaryomyces hansenii
Leishmania major Friedlin
Thalassiosira pseudonana
Dictyostelium discoideum
Schistosoma mansoni
Phytophthora ramorum
Trichoplax adhaerens
Coprinopsis cinereus okayama7#130
Branchiostoma floridae
Ciona intestinalis
Ciona savignyi
Daphnia pulex
Pediculus humanus
Helobdella robusta
Drosophila pseudoobscura
Drosophila virilis
Acrystosiphon pisum
Drosophila willistoni
Drosophila melanogaster
|   | Scientific Name                       |
|---|---------------------------------------|
| 59| *Drosophila ananassae*                 |
| 60| *Drosophila grimshawi*                 |
| 61| *Drosophila mojavensis*                |
| 62| *Tribolium castaneum*                  |
| 63| *Sorghum bicolor*                      |
| 64| *Bombyx mori*                          |
| 65| *Arabidopsis thaliana*                 |
| 66| *Pristionchus pacificus*               |
| 67| *Lottia gigantea*                      |
| 68| *Caenorhabditis japonica*              |
| 69| *Caenorhabditis remanei*               |
| 70| *Caenorhabditis elegans*               |
| 71| *Caenorhabditis briggsae*              |
| 72| *Brugia malayi*                        |
| 73| *Caenorhabditis brenneri*              |
| 74| *Capitella spI*                        |
| 75| *Culex pipiens*                        |
| 76| *Nasonia vitripennis*                  |
| 77| *Anopheles gambiae*                    |
| 78| *Ixodes scapularis*                    |
| 79| *Aedes aegypti*                        |
| 80| *Nematostella vectensis*               |
| 81| *Physcomitrella patens*                |
| 82| *Oryzias latipes*                      |
| 83| *Oryza sativa*                         |
| 84| *Apis mellifera*                       |
| 85| *Populus trichocarpa*                  |
| 86| *Gasterosteus aculeatus*               |
| 87| *Pinus sylvestris*                     |
| 88| *Danio rerio*                          |
| 89| *Fugu rubripes*                        |
| 90| *Tetraodon nigroviridis*               |
1. Fernández, A., Sosnick, T. R. & Colubri, A. Dynamics of hydrogen-bond desolvation in folding proteins. *J. Mol. Biol.* **321**, 659-675 (2002).

2. Street, A. G. & Mayo, S. L. Pairwise calculation of solvent accessible surface areas. *Fold. Des.* **3**, 253-258 (1998).

3. Zhang, N., Zeng, C. & Wingreen, N. S. Fast accurate evaluation of protein solvent exposure. *Proteins Strct. Funct. Bioinf.* **57**, 565-576 (2004).

4. Pietrosemoli, N., Crespo, A. & Fernández, A. Dehydration propensity of order-disorder intermediate regions in soluble proteins. *J. Proteome Res.* **6**, 3519-3526 (2007).
5. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D Complex: a structural classification of protein complexes. *PLoS Comput. Biol.* **2**, e155 (2006).

6. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nuc. Ac. Res.* **32**, W665-W667 (2004).

7. Clackson, T., Ultsch, M. H., Wells, J. A. & de Vos, A. M. Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J. Mol. Biol.* **277**, 1111-1128 (1998).

8. Fernández, A. Molecular basis for evolving self-dissimilarity in the yeast protein network. PLoS Computational Biology 3, e226 (2007).

9. Levy, E. D., Boeri Erba, E., Robinson, C. V. & S. A. Teichmann. Assembly reflects evolution of protein complexes. *Nature* **453**, 1262-1265 (2008).
Supplementary Material - Part II

Supplementary Tables 3-7

Nonadapative origins of interactome complexity

Ariel Fernández¹,² and Michael Lynch³

¹Department of Bioengineering, Rice University, Houston, TX 77005, USA
²Department of Computer Science, The University of Chicago, Chicago, IL 60637
³Department of Biology, Indiana University, Bloomington, IN 47405

Supplementary Table 3.

36 species containing orthologs in at least 90 out of 106 groups of orthologous proteins identified with OrthoMCL for which there are PDB representatives from at least 2 species.

Euryarchaeota (halobacteria)
ARCH: hvo
hvo   Haloferax volcanii

Gamma proteobacteria
GAMA: eco sty sfl ype vch
eco   Escherichia coli W3110 GenBank
sty   Salmonella enterica
ype   Yersinia pestis CO92
vch   Vibrio cholerae O1 biovar eltor str. N16961

Epsilon proteobacteria
EPSILON: caj
caj   Campylobacter jejuni

Ciliate
CIL: tth
tth  Tetrahymena thermophila

ddi (eukaryota)  Dictyostelium discoideum

Fungi (ascomycota, basidiomycota)
ASCO:  sce  ncr  dha  yli  kla  cgl  spo
  sce  Saccharomyces cerevisiae
  ncr  Neurospora crassa
  dha  Debaryomyces hansenii
  yli  Yarrowia lipolytica
  kla  Kluyveromyces lactis
  cgl  Candida glabrata
  spo  Schizosaccharomyces pombe

Plants and Algae
PLAL:  psy  osa  ath  cre  tps  cme  ota
  psy  Pinus sylvestris
  osa  Oryza sativa
  ath  Arabidopsis thaliana
  cre  Chlamydomonas reinhardtii
  tps  Thalassiosira pseudonana
  cme  Cyanidioschyzon merolae
  ota  Ostreococcus tauri

Nematodes
NEMA:  cel  cbr
  cel  Caenorhabditis elegans
  cbr  Caenorhabditis briggsae

Arthropoda
ARTH:  aga  dme  aed  ame  dpu
  aga  Anopheles gambiae
  dme  Drosophila melanogaster
  aed  Aedes aegypti
  ame  Apis mellifera
  dpu  Daphnia pulex

Fish and Bird
FISH:  tni  fru  dre  /gga
  tni  Tetraodon nigroviridis
  fru  Takifugu rubripes
  dre  Danio rerio
  gga  Gallus gallus

Mammals (mouse, rat, human)
MAMM:  mmu  rno  hsa
### Supplementary Table 4.
**Extent of PDB representation for each ortholog group**

| OrthoMCL identifier | Number of species | Protein description in E. coli proteome                               |
|----------------------|-------------------|------------------------------------------------------------------------|
| OG2_71139            | 3                 | Full=UPF0076 protein yjgF                                             |
| OG2_71139            | 3                 | Full=Protein tdcF                                                      |
| OG2_70953            | 2                 | Full=Dihydrolipoyl dehydrogenase                                       |
| OG2_71130            | 3                 | Full=Adenylate kinase                                                 |
| OG2_71331            | 2                 | Full=GTP cyclohydrolase 1                                             |
| OG2_71036            | 2                 | Full=tRNA Delta                                                       |
| OG2_70694            | 2                 | Full=Biotin carboxylase                                               |
| OG2_70752            | 2                 | Full=50S ribosomal protein L14                                         |
| OG2_70698            | 2                 | Full=Chaperone protein htpG                                           |
| OG2_71321            | 2                 | Full=Pyridoxine kinase                                               |
| OG2_71321            | 2                 | Full=Pyridoxamine kinase                                             |
| OG2_70755            | 2                 | Full=Pyruvate kinase II                                              |
| OG2_70755            | 2                 | Full=Pyruvate kinase I                                               |
| OG2_70925            | 3                 | Full=Ribonucleoside-diphosphate reductase 2 subunit alpha             |
| OG2_70925            | 3                 | Full=Ribonucleoside-diphosphate reductase 1 subunit alpha             |
| OG2_71349            | 2                 | Full=ATP-independent RNA helicase dbpA                                 |
| OG2_70922            | 2                 | Full=Carbamoyl-phosphate synthase large chain                         |
| OG2_70928            | 3                 | Full=Signal recognition particle protein                              |
| OG2_71657            | 3                 | Full=Ribonuclease D                                                  |
| OG2_70948            | 3                 | Full=UDP-glucose 4-epimerase                                          |
| OG2_70949            | 2                 | Full=Inositol-1-monophosphatase                                       |
| OG2_71140            | 2                 | Full=UPF0001 protein yggS                                             |
| OG2_71025            | 3                 | Full=Adenylosuccinate synthetase                                      |
| OG2_70967            | 3                 | Full=Sulfite reductase [NADPH] flavoprotein alpha-component            |
| OG2_71999            | 2                 | Full=Uncharacterized sugar kinase ydjH                                |
| OG2_70819            | 3                 | Full=Aromatic-amino-acid aminotransferase                             |
| OG2_70819            | 3                 | Full=Aspartate aminotransferase                                      |
| OG2_72499            | 2                 | Full=Protein bolA                                                    |
| OG2_71318            | 2                 | Full=Adenine phosphoribosyltransferase                                |
| OG2_71403            | 2                 | Full=Galactokinase                                                   |
| OG2_70765            | 3                 | Full=Serine hydroxymethyltransferase                                  |
| OG2_71569            | 2                 | Full=D-tyrosyl-tRNA                                                  |
| OG2_71038            | 2                 | Full=Adenylytransferase thiF                                          |
| OG2_71038            | 2                 | Full=Molybdopterin biosynthesis protein moeB                           |
| OG2_71565            | 2                 | Full=Aldose 1-epimerase                                               |
| OG2_71094            | 2                 | Full=DNA topoisomerase 4 subunit B                                   |
| OG2_71497            | 2                 | Full=Gamma-glutamyl phosphate reductase                               |
| OG2_70864            | 2                 | Full=Threonyl-tRNA synthetase                                         |
| OG2_70867            | 4                 | Full=Phosphoglycerate kinase                                         |
| Accession   | Count | Description                                           |
|------------|-------|-------------------------------------------------------|
| OG2_70861  | 2     | Full=DNA polymerase IV                                |
| OG2_70737  | 3     | Full=Deoxyuridine 5'-triphosphate nucleotidohydrolase |
| OG2_70862  | 2     | Full=2-oxoglutarate dehydrogenase E1 component        |
| OG2_71162  | 3     | Full=Aspartate ammonia-lyase                          |
| OG2_70951  | 2     | Full=Cold-shock DEAD box protein A                    |
| OG2_71000  | 4     | Full=Ribonucleoside-diphosphate reductase 1 subunit beta |
| OG2_71007  | 3     | Full=Glutaredoxin-3                                  |
| OG2_70708  | 2     | Full=Thioredoxin-1                                   |
| OG2_70708  | 2     | Full=Thioredoxin-2                                   |
| OG2_72501  | 2     | Full=Phosphoserine aminotransferase                   |
| OG2_71163  | 2     | Full=GMP synthase [glutamine-hydrolyzing]            |
| OG2_70858  | 2     | Full=CTP synthase                                    |
| OG2_70726  | 4     | Full=FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase |
| OG2_70726  | 4     | Full=FKBP-type peptidyl-prolyl cis-trans isomerase fpkA |
| OG2_71266  | 3     | Full=Superoxide dismutase [Cu-Zn]                    |
| OG2_71070  | 2     | Full=ATP synthase subunit alpha                       |
| OG2_70995  | 2     | Full=ATP synthase subunit beta                        |
| OG2_70994  | 3     | Full=Thymidylate kinase                              |
| OG2_71370  | 2     | Full=Protease 3                                      |
| OG2_70996  | 3     | Full=DNA polymerase III subunit tau                  |
| OG2_70990  | 2     | Full=Lysyl-tRNA synthetase, heat inducible            |
| OG2_70990  | 2     | Full=Lysyl-tRNA synthetase                           |
| OG2_70796  | 2     | Full=30S ribosomal protein S13                        |
| OG2_70795  | 2     | Full=30S ribosomal protein S5                         |
| OG2_73374  | 2     | Full=S-formylglutathione hydrolase yeiG              |
| OG2_73374  | 2     | Full=S-formylglutathione hydrolase frmB              |
| OG2_70963  | 2     | Full=Spermidine synthase                             |
| OG2_70712  | 2     | Full=DNA mismatch repair protein mutS                |
| OG2_70822  | 3     | Full=Glutathione reductase                           |
| OG2_71067  | 2     | Full=ATP synthase gamma chain                        |
| OG2_71942  | 3     | Full=Methionyl-tRNA synthetase                       |
| OG2_71100  | 2     | Full=Coenzyme A biosynthesis bifunctional protein coaBC |
| OG2_70978  | 2     | Full=Guanylate kinase                               |
| OG2_70785  | 2     | Full=S-adenosylmethionine synthetase                 |
| OG2_70787  | 2     | Full=30S ribosomal protein S11                        |
| OG2_71188  | 3     | Full=Catalase HPII                                   |
| OG2_71183  | 4     | Full=Fructose-6-phosphate aldolase 2                 |
| OG2_71183  | 4     | Full=Transaldolase A                                |
| OG2_71183  | 4     | Full=Fructose-6-phosphate aldolase 1                 |
| OG2_71183  | 4     | Full=Transaldolase B                                |
| OG2_70778  | 2     | Full=Methionine aminopeptidase                       |
| OG2_70856  | 2     | Full=Chaperone protein dnaK                          |
| OG2_70851  | 3     | Full=Orotate phosphoribosyltransferase               |
| OG2_70706  | 2     | Full=D-erythrose-4-phosphate dehydrogenase           |
| OG2_70706  | 2     | Full=Glyceraldehyde-3-phosphate dehydrogenase A      |
| OG2_73272  | 3     | Full=Protein cyaY                                    |
| OG2_70916  | 2     | Full=Chaperone protein dnaJ                          |
| OG2_70913  | 2     | Full=Copper-exporting P-type ATPase A                |
| OG2_71350  | 2     | Full=Thymidylate synthase                            |
| OG2_71352  | 2     | Full=Dihydroorotate dehydrogenase                    |
| OrthoMCL identifier | Species (hsa = human i.e. Homo sapiens etc) | UniProt identifier | Protein Description | PDB entries |
|---------------------|--------------------------------------------|-------------------|---------------------|-------------|
| OG2_71680 2         | Full=Succinate-semialdehyde dehydrogenase [NADP+] |                |                    |             |
| OG2_71680 2         | Full=Lactaldehyde dehydrogenase |                |                    |             |
| OG2_71201 2         | Full=50S ribosomal protein L17 |                |                    |             |
| OG2_70777 4         | Full=Superoxide dismutase [Mn] |                |                    |             |
| OG2_70777 4         | Full=Superoxide dismutase [Fe] |                |                    |             |
| OG2_70770 3         | Full=Enolase |                |                    |             |
| OG2_70959 2         | Full=Translation initiation factor IF-2 |                |                    |             |
| OG2_70828 2         | Full=Curved DNA-binding protein |                |                    |             |
| OG2_71614 2         | Full=Ribokinase |                |                    |             |
| OG2_70904 2         | Full=Phosphoribosylamine--glycine ligase |                |                    |             |
| OG2_70904 2         | Full=Phosphoribosylformylglycinamidine cyclo-ligase |                |                    |             |
| OG2_70900 3         | Full=Cystathionine gamma-synthase |                |                    |             |
| OG2_70804 3         | Full=Nucleoside diphosphate kinase |                |                    |             |
| OG2_71042 2         | Full=L-lactate dehydrogenase [cytochrome] |                |                    |             |
| OG2_71043 2         | Full=3-oxoacyl-[acyl-carrier-protein] synthase 2 |                |                    |             |
| OG2_71327 2         | Full=Ribosome-recycling factor |                |                    |             |
| OG2_71048 2         | Full=Nucleoside-triphosphatase rdgB |                |                    |             |
| OG2_71532 2         | Full=Aminomethyltransferase |                |                    |             |
| OG2_71430 2         | Full=Ribose-5-phosphate isomerase A |                |                    |             |
| OG2_70886 2         | Full=Acyll carrier protein |                |                    |             |
| OG2_70885 2         | Full=Aminotransferase ybdL |                |                    |             |
| OG2_71233 2         | Full=D-3-phosphoglycerate dehydrogenase |                |                    |             |
| OG2_71129 3         | Full=Glycogen phosphorylase |                |                    |             |
| OG2_71129 3         | Full=Maltodextrin phosphorylase |                |                    |             |
| OG2_70998 4         | Full=Triosephosphate isomerase |                |                    |             |
| OG2_70872 2         | Full=Uncharacterized protein ycgM |                |                    |             |
| OG2_71232 2         | Full=Tyrosyl-tRNA synthetase |                |                    |             |
| OG2_70651 2         | Full=2,5-diketo-D-gluconic acid reductase A |                |                    |             |
| OG2_70651 2         | Full=Uncharacterized protein yeaE |                |                    |             |
| OG2_70651 2         | Full=2,5-diketo-D-gluconic acid reductase B |                |                    |             |
| OG2_70727 2         | Full=Glutaredoxin-4 |                |                    |             |
| OG2_71607 3         | Full=Malate dehydrogenase |                |                    |             |
| OG2_70939 2         | Full=Glucose-6-phosphate isomerase |                |                    |             |
| OG2_70931 2         | Full=Fumarate reductase iron-sulfur subunit |                |                    |             |
| OG2_70931 2         | Full=Succinate dehydrogenase iron-sulfur subunit |                |                    |             |

**Supplementary Table 5.**

**Structural database for the 106 ortholog groups**

First column: species (hsa = human i.e. Homo sapiens etc),
second column: OrthoMCL identifier,
third column: UniProt identifier,
fourth column: protein description,
fifth column: list of PDB files and chains that represent the protein (PDB entries divided with “ | ”).
eco NP_418664.2 P0AF93 UPF0076 protein yjgF 1QU9,A/B/C=1-128|1QU9,A=1-128
eco NP_417583.4 P0AGL2 Protein tdcF 2UYJ,A/B/C=1-129|2UYK,A/B/C=1-129|2UYN,A=1-129, B/C=1-129|2UYP,A/B/C=1-129
sce YIL051C P40185 Protein MMF1, mitochondrial
sce YER057C P40037 Protein HMF1 1JD1,A/B/C/D/E/F=1-129
cel WBGene00016011 Q10121 UPF0076 protein C23G10.2
mmu ENSMUSP00000022946 P52760 Ribonuclease UK114
hsa ENSP00000254878 P52758 Ribonuclease UK114 1ONI,A/B/C/D/E/F/G/H/I=2-137|

### OG2_70953 ###
eco NP_414658.1 P0AP90 Dihydrolipoyl dehydrogenase
sce YFL018C P09624 Dihydrolipoyl dehydrogenase, mitochondrial 1JEH,A/B=22-499|1V59,A/B=22-499
cel WBGene00010794 O17953 Dihydrolipoyl dehydrogenase
mmu ENSMUSP0000002980 O08749 Dihydrolipoyl dehydrogenase, mitochondrial
hsa ENSP00000205402 P09622 Dihydrolipoyl dehydrogenase, mitochondrial 12MC,A/B/C/D/E/F/G/H=36-509|1ZMD,A/B/C/D/E/F/G/H=36-509|1ZY8,A/B/C/D/E/F/G/H/I/J=36-509|2F5Z,A/B/C/D/E/F/G/H/I/J=36-509|

### OG2_71130 ###
eco NP_415007.1 P69441 Adenylate kinase 1A6T5, mitochondrial 1A8R,A/B=1-214|1ANK,A/B=1-214|1E4V,A/B=1-214|1E4Y,A/B=1-214|2ECK,A/B=1-214|4AKE,A/B=1-214|
sce YDR226W P07170 Adenylate kinase 1 1AKY,A=3-221|1DVR,A/B=3-221|2AKY,A=3-221|3AKY,A=3-221|
cel WBGene00016205 P34346 Adenylate kinase 2, mitochondrial
mmu ENSMUSP00000030583 Q9WTP6 Adenylate kinase 2, mitochondrial
hsa ENSP00000346921 P54819 Adenylate kinase 2, mitochondrial 2C9Y,A=2-238|

### OG2_71331 ###
eco NP_416658.1 P0A6T5 GTP cyclohydrolase 1 1A8R,A/B/C/D/E/F/G/H/I/J/K/L/M/N/O=2-222|1A9C,A/B/C/D/E/F/G/H/I/J/K/L/M/N/O=2-222|1FBX,A/B/C/D/E/F/G/H/I/J/K/L/M/N/O=2-222|1GTP,A/B/C/D/E/F/G/H/I/J/K/L/M/N/O=2-222|1N3R,A/B/C/D/E/F/G/H/I/J/K/L/M/N/O=2-222|1N3S,A/B/C/D/E/F/G/H/I/J/K/L/M/N/O=2-222|
sce YGR267C P51601 GTP cyclohydrolase 1
cel WBGene00000298 Q19980 GTP cyclohydrolase 1
mmu ENSMUSP00000087405 Q05915 GTP cyclohydrolase 1
hsa ENSP00000254299 P30793 GTP cyclohydrolase 11FB1,A/B/C/D/E=55-250|

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
eco NP_417722.1 P24182 Biotin carboxylase 1BNC,A/B=1-449|1DV1,A/B=1-449|1DV2,A/B=1-449|1K69,A=1-447|2GPS,A/B=1-449|2GPW,A/B/C/D=1-449|2J9G,A/B=1-449|2V58,A/B=1-449|2V59,A/B=1-449|2V5A,A/B=1-449|2VR1,A/B=1-449|2W6M,A/B=1-449|2W6N,A/B=1-449|2W6O,A/C=1-449|2W6P,A/B=1-449|2W6Q,A/B=1-449|2W6Z,A/B=1-449|2W70,A/B=1-449|2W71,A/C=1-449|3G8C,A/B=1-444|3G8D,A/B=1-444|
sce YGL062W P11154 Pyruvate carboxylase 1
sce YBR218C P32327 Pyruvate carboxylase 2
sce YBR208C P32528 Urea amidolylase
cel WBGene00009319 Q45430 Protein F32B6.2, partially confirmed by transcript evidence
cel WBGene00017864 Q19842 Propionyl-CoA carboxylase alpha chain, mitochondrial
cel WBGene0004258 O17732 Pyruvate carboxylase 1
mmu ENSMUSP00000038763 Q91ZA3 Propionyl-CoA carboxylase alpha chain, mitochondrial
mmu ENSMUSP00000029259 Q99MR8 Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial
mmu ENSMUSP00000063825 Q05920 Pyruvate carboxylase, mitochondrial
hsa ENSP00000365463 B4DKY8 cDNA FLJ59564, highly similar to Propionyl-CoA carboxylase alpha chain, mitochondrial
hsa ENSP00000347900 P11498 Pyruvate carboxylase, mitochondrial
hsa ENSP00000265594 Q96RQ3 Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial 2EJM,A=640-725#
# # # # # # # OG2_70752 # # # # # #
eco NP_417769.1 P0ADY3 50S ribosomal protein L14 1P85,I=1-123|1P86,I=1-123|1VS6,K=1-123|1VS8,K=1-123|2AW4,K=1-123|2AWB,K=1-123|2GYA,I=2-123|2GYC,I=2-123|2I2T,K=1-123|2I2V,K=1-123|2J28,K=2-122|2QAM,K=1-123|2QAO,K=1-123|2QBA,K=1-123|2QBC,K=1-123|2QBE,K=1-123|2QBG,K=1-123|2QBI,K=1-123|2QBK,K=1-123|2QOV,K=1-123|2QOX,K=1-123|2QOZ,K=1-123|2QP1,K=1-123|2RDO,K=1-123|2Z4L,K=1-123|2Z4N,K=1-123|3BBX,K=1-123|3FIK,K=2-122|
sce YKL170W P35996 54S ribosomal protein L38, mitochondrial
sce YER117W P04451 60S ribosomal protein L23 1K5Y,R=7-137|1S1I,R=1-137|
sce YBL087C P04451 60S ribosomal protein L23 1K5Y,R=7-137|1S1I,R=1-137|
cel WBGene000004435 P48158 60S ribosomal protein L23
mmu ENSMUSP00000099435 P62830 60S ribosomal protein L23
hsa ENSP00000367336
# # # # # # # OG2_70698 # # # # # #
eco NP_415006.1 P0A6Z3 Chaperone protein htpG
1SF8,A/B/C/D/E/F/G/H=511-624|1Y4S,A/B=1-559|1Y4U,A/B=1-559|2GQ0,A/B=230-495|2IOP,A/B/C/D=1-624|2IOQ,A/B=1-624|2IOR,A=1-215|
sce YPL240C P02829 ATP-dependent molecular chaperone HSP82
1A4H,A=1-220|1AH6,A=1-220|1AH8,A/B=1-220|1AM1,A=2-214|1AMW,A=1-214|1BGQ,A=1-214|1HK7,A/B=273-560|1US7,A=1-214|1USU,A=273-530|1USV,A/C/E/G=272-530|1ZW9,A=1-220|1ZWH,A=1-220|2AKP,A/B=25-210|2BRC,A=1-214|2BRE,A/B=1-219|2CG9,A/B=1-677|2CGE,A/B/D=273-677|2CGF,A=1-214|2FXS,A=1-220|2IWS,A=1-214|2IWU,A=1-214|2IXW,A=1-214|2IW5,A/B/C/D=1-214|2VWC,A=1-219|2WEP,A=1-220|2WEQ,A=1-220|2WER,A/B=1-220|3C0E,A=1-220|3C11,A=1-220|
sce YMR186W P15108 ATP-dependent molecular chaperone HSC82
cel WBGene00011480 Q22235 Protein T05E11.3, confirmed by transcript evidence
cel WBGene00000915 Q18688 Heat shock protein 90
mmu ENSMUSP00000085380
mmu ENSMUSP0000052746
mmu ENSMUSP0000024739 P11499 Heat shock protein HSP 90-beta
mmu ENSMUSP0000021698 P07901 Heat shock protein HSP 90-alpha
mmu ENSMUSP0000020238 P08113 Endoplasmin
hsa ENSP00000369893 Q58FF6 Putative heat shock protein HSP 90-beta 4
hsa ENSP00000368186 Q58FF3 Putative endoplasmin-like protein
hsa ENSP00000359433 Q58FF2 Heat shock protein 94c
hsa ENSP00000369649 Q58FF7 Putative heat shock protein HSP 90-beta-3
hsa ENSP00000299767 P14625 Endoplasmin
hsa ENSP00000329390
hsa ENSP00000335153 Q14568 Putative heat shock protein HSP 90-alpha A2

################################ OG2_70755 ################################
ecO NP_416368.1 P21599 Pyruvate kinase II
eCO NP_416191.1 P0AD61 Pyruvate kinase I 1E0T,A/B/C/D=1-470|1E0U,A/B/C/D=1-470|1PKY,A/B/C/D=1-470|
sce YOR347C P52489 Pyruvate kinase 2
cel YAL038W P00549 Pyruvate kinase 1 1A3W,A/B=1-500|1A3X,A/B=1-500|
cel WBGene00014001 Q23539 Pyruvate kinase
cel WBGene00009126 B7WNA0 Pyruvate kinase
mmu ENSMUSP0000035417 P53657 Pyruvate kinase isozymes R/L
mmu ENSMUSP0000034834 P52480 Pyruvate kinase isozymes M1/M2
hsa ENSP00000271946 P30613 Pyruvate kinase isozymes R/L 2VGB,A/B/C/D=47-574|2VGF,A/B/C/D=47-574|2VGG,A/B/C/D=47-574|2VGL,A/B/C/D=47-574|
hsa ENSP00000373740

################################ OG2_70785 ################################
eco NP_417417.1 P0A817 S-adenosylmethionine synthetase 1FUG,A/B=2-384|1MXA,A=2-384|1MXB,A=2-384|1MXC,A=2-384|1P7L,A/B/C/D=2-383|1RG9,A/B/C/D=2-383|1XRA,A=2-384|1XRB,A=2-384|1XRC,A=2-384|
sce YLR180W P10659 S-adenosylmethionine synthetase 1
| Organism | Accession | Gene Name | Description |
|----------|-----------|-----------|-------------|
| sce      | YDR502C   | P19358    | S-adenosylmethionine synthetase 2 |
| mmu      | ENSMUSP0000065081 | Q6PE05 | Mat2a protein |
| hsa      | ENSP00000303147 | P31153 | S-adenosylmethionine synthetase isoform type-2 |
| eco      | NP_414574.1 | P00968 | Carbamoyl-phosphate synthase large chain |
| sce      | YJR109C   | P03965    | Carbamoyl-phosphate synthase arginine-specific large chain |
| mmu      | ENSMUSP0000027144 | Q8C196 | Carbamoyl-phosphate synthase [ammonia], mitochondrial |
| mmu      | ENSMUSP0000013773 | B2RQC6 | Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase |
| hsa      | ENSP00000264705 | P27708 | CAD protein |
| eco      | NP_417101.1 | P0AGD7 | Signal recognition particle 1DUL,A=371-430 |
| sce      | YPR088C  | P20424    | Signal recognition particle subunit SRP54 |
| mmu      | ENSMUSP0000093400 | P14576 | Signal recognition particle 54 kDa protein 1RY1,W=326-434 |
| hsa      | ENSP00000216774 | P61011 | Signal recognition particle 54 kDa protein 1MFQ,C=323-441 |
**eco**  NP_416318.1  P09155  Ribonuclease D  1YT3,A=1-375  
**sce**  YOR001W  Q12149  Exosome complex exonuclease RRP6  1MOY,A=207-382|2HBJ,A=129-536|2HBK,A=129-536|2HBL,A=129-536|2HBM,A=129-536  
**cel**  WBGene00000796  Q17951  Protein C14A4.4a, confirmed by transcript evidence  
**mmu**  ENSMUSP0000017408  P56960  Exosome component 10  
**hsa**  ENSP00000366135  Q01780  Exosome component 10  2CPR,A=483-593  

----------  
**eco**  NP_415280.3  P09147  UDP-glucose 4-epimerase  1A9Y,A=1-338|1A9Z,A=1-338|1KVQ,A=1-338|1KVR,A=1-338|1KVS,A=1-338|1KVT,A=1-338|1KUT,A=1-338|1KVU,A=1-338|1LRJ,A=1-338|1LRK,A=1-338|1LRL,A=1-338|1NAH,A=1-338|1NAI,A=1-338|1UDA,A=1-338|1UDB,A=1-338|1UDC,A=1-338|1XEL,A=1-338|2UDP,A/B=1-338  
**sce**  YBR019C  P04397  Bifunctional protein GAL10  1Z45,A=1-699  
**cel**  WBGene00008132  O62107  Protein C47B 2.6a, confirmed by transcript evidence  
**mmu**  ENSMUSP00000095453  
**hsa**  ENSP00000313026  Q14376  UDP-glucose 4-epimerase  1EK5,A=1-348|1EK6,A/B=1-348|1HZJ,A/B=1-348|1I3K,A/B=1-348|1I3L,A/B=1-348|1I3M,A/B=1-348|1I3N,A/B=1-348  

----------  
**eco**  NP_415728.1  P0ADG4  Inositol-1-monophosphatase  2QFL,A=1-267  
**sce**  YHR046C  P38710  Inositol monophosphatase 1  
**sce**  YDR287W  Q05353  Inositol monophosphatase 2  
**cel**  WBGene00008765  Q19420  Inositol monophosphatase ttx-7  
**mmu**  ENSMUSP0000006174  O55023  Inositol monophosphatase  
**mmu**  ENSMUSP0000025403  Q91UZ5  Inositol monophosphatase 2  
**hsa**  ENSP00000269159  Q14732  Inositol monophosphatase 2  2CZH,A/B=1-288|2CZI,A=1-288|2CZK,A=1-288|2DDK,A/B=1-288|2FVZ,A/B/C/D=16-288|2FVZ,A/B/C/D=16-288  
**hsa**  ENSP00000256108  Q29218  Inositol monophosphatase  1AWB,A/B=2-277|1IMA,A/B=1-277|1IMB,A/B=1-277|1IMC,A/B=1-277|1IMD,A/B=1-277|1IME,A/B=1-277|1IMF,A=1-277|2HHM,A/B=2-277  

----------  
**eco**  NP_417426.1  P67080  UPF0001 protein yggS  1W8G,A=1-234  
**sce**  YBL036C  P38197  UPF0001 protein YBL036C  1B54,A=1-257|1CT5,A=2-257  
**cel**  WBGene00017286  P52057  Proline synthetase co-transcribed bacterial homolog protein  
**cel**  WBGene00017285  Q19257  Putative uncharacterized protein  
**mmu**  ENSMUSP00000033875  Q922Y8  Proline synthetase co-transcribed bacterial homolog protein  
**hsa**  ENSP00000333551  O94903  Proline synthetase co-transcribed bacterial homolog protein  

----------
eco  NP_418598.1  P0A7D4  Adenylosuccinate synthetase 1ADE,A/B=2-432|1ADI,A/B/1CG0,A=2-432|1CG1,A=2-432|1CG3,A=2-432|1CG4,A=2-432|1CH8,A=2-432|1CIB,A=2-432|1G1M,A=2-432|1GIN,A=2-432|1HON,A/B=2-431|1HO0,A/B=2-431|1HOP,A/B=2-431|1JUY,A=2-432|1KJX,A=1-432|1KKB,A=1-432|1KKS,A=1-432|1NT,H,A=2-431|1QF4,A=2-432|1QF5,A=2-432|1SON,A=2-431|1SOO,A=2-431|2GCQ,A=2-431|

sce  YNL220W  P80210  Adenylosuccinate synthetase
cel  WBGene00016509  P91134  Probable adenylosuccinate synthetase
mmu  ENSMUSP00000016105  P46664  Adenylosuccinate synthetase isozyme 2

mmu  ENSMUSP00000021726  P28650  Adenylosuccinate synthetase isozyme 1

hsa  ENSP00000333019  Q8N142  Adenylosuccinate synthetase isozyme 1

hsa  ENSP00000355493  P30520  Adenylosuccinate synthetase isozyme 2

############ OG2_71999 ##############
eco  NP_416286.4  P77493  Uncharacterized sugar kinase ydjH  3H49,A/B=2-315|
sce  YJR105W  P47143  Adenosine kinase
cel  WBGene00011128  Q93934  Protein R07H5.8, confirmed by transcript evidence

mmu  ENSMUSP00000047665  P55264  Adenosine kinase

hsa  ENSP00000286621  P55263  Adenosine kinase 1BX4,A=22-362|2I6A,A/B/C/D=22-362|2I6B,A/B=22-362|

############ OG2_70819 ##############
eco  NP_418478.1  P04693  Aromatic-amino-acid aminotransferase 3FSL,A/B/C/D/E/F=1-397|3TAT,A/B/C/D/E/F=1-397|
eco  NP_415448.1  P00509  Aspartate aminotransferase 1AAM,A=1-396|1AAW,A=1-396|1AHE,A/B=1-396|1AHF,A/B=1-396|1AHG,A/B=1-396|1AHX,A/B=1-396|1AHY,A/B=1-396|1AIA,A/B=1-396|1AIB,A/B=1-396|1AIC,A/B=1-396|1AMQ,A=1-396|1AMR,A=1-396|1AMS,A=1-396|1ARG,A/B=1-396|1ARH,A/B=1-396|1ARI,A/B=1-396|1ARS,A=1-396|1ART,A=1-396|1ASA,A=1-396|1ASB,A=1-396|1ASC,A=1-396|1ASD,A=1-396|1ASE,A=1-396|1ASF,A=1-396|1ASG,A=1-396|1ASL,A/B=1-396|1ASM,A/B=1-396|1ASN,A/B=1-396|1B4X,A=1-396|1BQA,A/B=1-396|1BQD,A/B=1-396|1C9C,A=1-396|1CQ6,A=1-396|1GCQ,A=1-396|1CQ8,A=1-396|1CZC,A=1-396|1CZE,A=1-396|1G4V,A=1-396|1G4X,A=1-396|1G7W,A=1-396|1G7X,A=1-396|1IX6,A=1-396|1IX7,A=1-396|1IX8,A=1-396|1QIR,A=1-396|1QIS,A=1-396|1QIT,A=1-396|1SPA,A=1-396|1TOE,A=1-388|1TOG,A/B=1-388|1TOI,A=1-388|1TOJ,A=1-388|1TOK,A/B=1-388|1X28,A/B=1-396|1X29,A/B=1-396|1X2A,A/B=1-396|1YOO,A=1-396|2AAT,A=1-396|2D5Y,A=1-396|2D61,A=1-396|2D63,A=1-396|2D64,A=1-396
sce YLR027C P23542 Aspartate aminotransferase, cytoplasmic
  1YAA,A/B/C/D=2-412|
cel WBGene00020146 Q22067 Probable aspartate aminotransferase, cytoplasmic
cel WBGene00016652 O01804 Aspartate aminotransferase
cel WBGene00015778 Q17994 Aspartate aminotransferase
mmu ENSMUSP00000065813
mmu ENSMUSP00000034097 P05202 Aspartate aminotransferase, mitochondrial
mmu ENSMUSP00000026196 P05201 Aspartate aminotransferase, cytoplasmic
hsa ENSP00000245206 P00505 Aspartate aminotransferase, mitochondrial
hsa ENSP00000359539 P17174 Aspartate aminotransferase, cytoplasmic

############################ OG2_72499 ##############################

eco NP_414969.4 P0ABE2 Protein bolA 2DHM,A=1-100|
sce YAL044W-A Q3E793 Uncharacterized bolA-like protein YAL044W-A
cel WBGene00019658 P91375 Putative bolA-like protein K11H12.1
mmu ENSMUSP0000016087 Q9D8S9 BolA-like protein 1 1V60,A=13-128|
hsa ENSP00000358149 Q9Y3E2 BolA-like protein 1

############################ OG2_71036 ##############################

eco NP_418592.1 P16384 tRNA Delta 2ZM5,A/B=1-316|2ZXU,A/B=1-316|3FOZ,A/B=1-316|
sce YOR274W P07884 tRNA isopentenyltransferase 3EPH,A/B=13-421|3EPJ,A/B=13-421|3EPK,A/B=13-421|3EPL,A/B=13-421|
cel WBGene00001740 Q9GYG3 Abnormal growth rate protein 1
mmu ENSMUSP00000030406
hsa ENSP00000321810 Q9H3H1 tRNA isopentenyltransferase, mitochondrial

############################ OG2_71318 ##############################

eco NP_415002.1 P69503 Adenine phosphoribosyltransferase 2DY0,A/B=1-183|
sce YML022W P49435 Adenine phosphoribosyltransferase 11G2P,A=3-187|1G2Q,A/B=3-187|
cel WBGene00020557 P91455 Adenine phosphoribosyltransferase
mmu ENSMUSP0000006764 P08030 Adenine phosphoribosyltransferase
hsa ENSP00000355098

############################ OG2_71403 ##############################

eco NP_415278.1 P0A6T3 Galactokinase
sce YDR009W P13045 Protein GAL3
sce YBR020W P04385 Galactokinase 2AJ4,A/B=2-527|
cel WBGene00006461 O01969 Temporarily assigned gene name protein 96
mmu ENSMUSP0000092186 Q68FH4 N-acetylgalactosamine kinase
mmu ENSMUSP00000021114 Q80UL3 Galactokinase 1
hsa ENSP00000316632 Q01415 N-acetylgalactosamine kinase2A2C,A=1-458|2A2D,A=1-458|
hsa ENSP00000364334 P51570 Galactokinase 1WUU,A/B/C/D=2-392|YH7,A=1-392|

############################ OG2_70765 ##############################
eco NP_417046.1 P0A825 Serine hydroxymethyltransferase
1DFO,A/B/C/D=1-417|EQB,A/B/C/D=1-417|
sce YLR058C P37291 Serine hydroxymethyltransferase, cytosolic
sce YBR263W P37292 Serine hydroxymethyltransferase, mitochondrial
cel WBGene00003214 P50432 Serine hydroxymethyltransferase
mmu ENSMUSP00000082202
mmu ENSMUSP0000026470 Q99K87 Serine hydroxymethyltransferase
mmu ENSMUSP0000018744 P50431 Serine hydroxymethyltransferase, cytosolic
hsa ENSP00000333667 P34897 Serine hydroxymethyltransferase, mitochondrial
2A7V,A=17-504|
hsa ENSP00000318868 P34896 Serine hydroxymethyltransferase, cytosolic
1BJ4,A=11-480|

############################ OG2_71569 ##############################
eco NP_418323.1 P0A6M4 D-tyrosyl-tRNA 1JKE,A/B/C/D=1-145|
sce YDL219W Q07648 D-tyrosyl-tRNA
cel WBGene00004151 Q9XUP4 D-tyrosyl-tRNA
mmu ENSMUSP0000028917 Q9DD18 D-tyrosyl-tRNA
hsa ENSP00000366672 Q8TEA8 D-tyrosyl-tRNA 2OKV,A/B/C/D=1-209|

############################ OG2_71038 ##############################
eco NP_418420.4 P30138 Adenylyltransferase thiF 1ZFN,A/B/C/D=1-251|ZKM,A/B/C/D=1-251|ZUD,1/3=1-251|
eco NP_415347.1 P12282 Molybdopterin biosynthesis protein moeB
1JW9,B=1-249|JWA,B=1-249|JWB,B=1-249|
sce YHR111W P38820 Adenylyltransferase and sulfurtransferase UBA4
cel WBGene00018357 O44510 Adenylyltransferase and sulfurtransferase MOCS3
mmu ENSMUSP00000096670 A2BDX3 Adenylyltransferase and sulfurtransferase MOCS3
hsa ENSP00000244051 Q95396 Adenylyltransferase and sulfurtransferase MOCS3
3I2V,A=335-460|

############################ OG2_71565 ##############################
eco NP_415277.1 P0A9C3 Aldose 1-epimerase
sce YNR071C P53757 Uncharacterized isomerase YNR071C
1YGA,A/B=1-342|
sce YHR210C P38893 Uncharacterized isomerase YHR210C
cel WBGene00044734 Q9UAT6 Putative uncharacterized protein
1LUR,A/B=1-330|
cel WBGene00015270 Q9UAT6 Putative uncharacterized protein 1LUR,A/B=1-330
mmu ENSMUSP00000040580 Q8K157 Aldose 1-epimerase
hsa ENSP00000272252 Q96C23 Aldose 1-epimerase ISNZ,A/B=1-342|ISO0,A/B/C/D=1-342

############################ OG2_71094 ############################
eco NP_417502.1 P20083 DNA topoisomerase 4 subunit B 1S14,A/B=1-217|1S16,A/B=1-390
sce YNL088W P06786 DNA topoisomerase 2 21BGW,A=409-1201|1BJT,A=409-1201|1PVG,A/B=1-413|1QZR,A/B=1-413|2RGR,A=419-1177

 cel WBGene00022854 Q27537 Chromosome instability 4
 cel WBGene00019876 P34534 Putative DNA topoisomerase 2, mitochondrial
 cel WBGene00010785 Q23670 Probable DNA topoisomerase 2
 mmu ENSMUSP00000099427
 mmu ENSMUSP00000017629 Q64511 DNA topoisomerase 2-beta
 hsa ENSP00000269577 P11388 DNA topoisomerase 2-alpha 1LWZ,A=431-1200|1ZXM,A/B=29-428|1ZXN,A/B/C/D=29-428
 hsa ENSP00000264331 Q02880 DNA topoisomerase 2-beta

############################ OG2_71497 ############################
eco NP_414778.1 P07004 Gamma-glutamyl phosphate reductase
sce YOR323C P54885 Gamma-glutamyl phosphate reductase 1VLU,A/B=1-456

 cel WBGene00011938 P54889 Probable delta-1-pyrroline-5-carboxylate synthetase
 mmu ENSMUSP00000025979 Q9Z110 Delta-1-pyrroline-5-carboxylate synthetase
 hsa ENSP00000360268 P54886 Delta-1-pyrroline-5-carboxylate synthetase

 char

 s2H2G,A/B=362-795

############################ OG2_70864 ############################
eco NP_416234.1 P0A8M3 Threonyl-tRNA synthetase 1EVK,A/B=242-642|1EVL,A/B/C/D=242-642|1FYF,A/B=242-642|1KOG,A/B/C/D/E/F/G/H=242-642|1QF6,A=1-642|1TJE,A=1-224|1TKG,A=1-224|1TKY,A=1-224
sce YIL078W P04801 Threonyl-tRNA synthetase, cytoplasmic

 cel WBGene00006617 P52709 Threonyl-tRNA synthetase, cytoplasmic
 mmu ENSMUSP00000032728 Q8BLY2 Probable threonyl-tRNA synthetase 2, cytoplasmic
 mmu ENSMUSP00000029752 Q3UQ84 Threonyl-tRNA synthetase, mitochondrial
 mmu ENSMUSP00000022849 Q9D0R2 Threonyl-tRNA synthetase, cytoplasmic
 hsa ENSP00000338093 A2RTX5 Probable threonyl-tRNA synthetase 2, cytoplasmic
 hsa ENSP00000358060 Q9BW92 Threonyl-tRNA synthetase, mitochondrial
hsa ENSP00000265112 P26639 Threonyl-tRNA synthetase, cytoplasmic
1WWT,A=79-153|

#OG2_70867######
eco NP_417401.1 P0A799 Phosphoglycerate kinase 1ZMR,A=1-387|
sce YCR012W P00560 Phosphoglycerate kinase IFW8,A=2-416|1QPG,A=2-416|3PGK,A=2-416|
cel WBGene00020185 P91427 Probable phosphoglycerate kinase
mmu ENSMUSP00000088647
mmu ENSMUSP0000082884
mmu ENSMUSP0000080302 P09411 Phosphoglycerate kinase l
mmu ENSMUSP0000033585 P09041 Phosphoglycerate kinase 2
2P9Q,A/B=2-417|2P9T,A=2-417|2PAA,A/B=2-417|
hsa ENSP00000305995 P07205 Phosphoglycerate kinase 2
hsa ENSP00000362413 P00558 Phosphoglycerate kinase 1
2ZGV,A=1-417|3C3A,A/B=1-417|3C3B,A/B=1-417|3C3C,A/B=1-417|

#OG2_70861######
eco NP_414766.1 Q47155 DNA polymerase IV 1OK7,C=336-351|1UUNN,C/D=243-351|
sce YOR346W P12689 DNA repair protein REV1 2AQ4,A=305-738|3BJY,A=305-738|
cel WBGene00014066 Q09615 Protein ZK675.2, partially confirmed by transcript evidence
cel WBGene00017696 P34409 DNA polymerase kappa
mmu ENSMUSP0000027251 Q920Q2 DNA repair protein REV1
mmu ENSMUSP0000022172 Q9QUG2 DNA polymerase kappa
hsa ENSP00000258428 Q9UBZ9 DNA repair protein REV1
2EBW,A=44-133|3GQC,A/B/C/D=330-833|
hsa ENSP00000353425

#OG2_70737######
eco NP_418097.1 P06968 Deoxyuridine 5'-triphosphate nucleotidohydrolase
1DUD,A=1-151|1DUP,A=1-151|1EU5,A=1-151|1EUW,A=1-151|1RN8,A=1-151|1RNJ,A=1-151|1SEH,A=1-151|1SYL,A=1-151|2HR6,A=1-151|2HRM,A=1-151|
sce YBR252W P33317 Deoxyuridine 5'-triphosphate nucleotidohydrolase
3F4F,A/B=C=1-147|3HHQ,A/B/C/D/E/F/G/H/I/J/K/L/M/N/O/P/Q/R/S/T/U/V/W/X=1-147|
cel WBGene00010609 Q001934 Protein K07A1.2, confirmed by transcript evidence
mmu ENSMUSP0000057854 Q9CQ43 Deoxyuridine triphosphatase
hsa ENSP00000370376 P33316 Deoxyuridine 5'-triphosphate
nucleotidohydrolase, mitochondrial 1Q5H,A/B/C=112-252|1Q5U,X/Y/Z=112-252|2HQU,A/B/C=94-252|3EHW,A/B/C/X/Y/Z=94-252|

#OG2_70862######
eco NP_415254.1 P0A799 2-oxoglutarate dehydrogenase E1 component
2JGD,A=1-933,B=1-933|
| Species | Accession | Description | Targets | Information |
|---------|-----------|-------------|---------|-------------|
| sce     | YJL138C   | ATP-dependent RNA helicase eIF4A1 | UFK,A=231-395 [1FUU,A/B=2-395] | QDE,A=9-232 [1QVA,A=2-224] VSO,A/B=1-395 [2VSX,A/B=1-395] |
| cel     | WBGene0000283 | Eukaryotic initiation factor 4A | P27639 | |
| mmu     | ENSMUSP0000074886 | Eukaryotic initiation factor 4A | |
| mmu     | ENSMUSP0000072748 | Eukaryotic initiation factor 4A-I | P60843 | |
| mmu     | ENSMUSP00000009649 | Eukaryotic initiation factor 4A-II | P10630 | |
| hsa     | ENSP00000309842 | Eukaryotic initiation factor 4A-I | 2G9N,A/B=20-236 [2ZU6,A/C/D/F=20-406] | IQIQ,A/D=1-406 |
| hsa     | ENSP00000326381 | Eukaryotic initiation factor 4A-II | Q14240 | |
| eco     | NP_416738.1 | Ribonucleoside-diphosphate reductase 1 subunit beta | 1AV8,A/B=2-341 [1BIQ,A=2-376] | JPR,A/B=1-376 [1JQC,A/B=2-376] MRR,A/B=2-376 [1MXR,A/B=2-375] PFR,A/B=2-341 [1PM,A/B=2-375] PIU,A/B=2-375 [1PYI,A/B=2-375] IzA,B=2-375 [1PZ,A/B=2-375] |
| eco     | NP_416913.1 | Pyridoxine kinase | 2DDM,A/B=1-283 [2DDO,A/B=1-283] | 2DDW,A/B=1-283 |
| eco     | NP_416153.1 | Pyridoxamine kinase | 1TD2,A/B=1-287 [1VI9,A/B/C/D=1-287] |
| sce     | YNR027W    | Putative pyridoxal kinase | P53727 | |
| sce     | YEL029C    | Putative pyridoxal kinase | P39988 | |
| mmu     | ENSMUSP0000038540 | Pyridoxal kinase | Q8K183 | |
mmu ENSMUSP0000002868
hsa ENSP00000029156 Q00764 Pyridoxal kinase 2AJP,A/B=6-312|2F7K,A/B=1-312|2YXT,A/B=1-312|2YXU,A/B=1-312|3FHX,A/B=1-312|3FHY,A/B=1-312

############ OG2_73272 ############

type ECO P27838 Protein cyA 1EW4,A=1-106|ISOY,A=1-106|2EFF,A=1-106|2P1X,A=1-106

type SCE YQL20W Q07540 Frataxin homolog, mitochondrial 2FQL,A=52-174|2GA5,A=53-174

type CEL WBGene00001486 Q9TY03 Frataxin homolog

mmu ENSMUSP00000008081 O35943 Frataxin, mitochondrial

hsa ENSP00000366482 Q16595 Frataxin, mitochondrial 1EKG,A=88-210|1LY7,A=91-210

############ OG2_71007 ############

type ECO NP_418067.1 P0AC62 Glutaredoxin-3 1FOV,A=2-83|1ILB,A=1-83|1UQ8,G=2-83|1UQ9,G=2-83|1UQI,G=2-83|1UQJ,G=2-83|1UQK,G=2-83|1UQL,G=2-83|1UQM,G=2-83|1UQO,G=2-83|1UQQ,G=2-83|3GRX,A=2-83

type SCE YQL513W P07695 Glutaredoxin-2, mitochondrial 3CTF,A=35-143|3CTG,A=35-143|3D4M,A=35-143|3D5J,A/B=35-143

type CEL WBGene00021331 Q9N456 Glutaredoxin protein 10

mmu ENSMUSP0000022082 Q9QUH0 Glutaredoxin-1

mmu ENSMUSP0000024500

hsa ENSP0000369314 P35754 Glutaredoxin-1 1B4Q,A=2-106|1JHB,A=1-106

hsa ENSP0000356410 Q9NS18 Glutaredoxin-2, mitochondrial 2CQ9,A=48-164|2FLS,A=56-164|2HT9,A/B=41-164

############ OG2_70708 ############

type ECO NP_418228.2 P0AA25 Thioredoxin-1 1F6M,C/D/G/H=2-109|1KEB,A/B=2-108|1MT7,A=|1OAZ,A/B=1-109|1SKR,B=2-108|1SKS,B=2-108|1SKW,B=2-108|1SL0,B/D=2-108|1SL1,B=2-108|1SL2,B=2-108|1SRX,A=2-109|1T7P,B=2-108|1T8E,B=2-108|1THO,A=2-108|1TK0,B=2-108|1TK5,B=2-108|1TK8,B=2-108|1TKD,B=2-108|1TXL,A=2-109|1X9M,B=2-108|1X9S,B=2-108|1X9W,B=2-108|1XOA,A=2-109|1XOB,A=2-109|1ZCP,A/B/C/D=2-108|1ZYO,B=2-108|1ZYY,A/B=1-109|2AJ0,B/I=2-108|2BTO,T=2-109|2EIO,A/B/C/D=2-109|2E1Q,A/B=2-109|2EIR,A/B/C/D=2-109|2FCH,A/B/C/D/E/F/G=2-108|2FD3,A/B=2-108|2H6X,A/B=2-109|2H6Y,A/B=2-109|2H6Z,A/B=2-109|2H70,A/B=2-109|2H71,A/B=2-109|2H72,A/B=2-109|2H73,A/B=2-109|2H74,A/B=4-109|2H75,A/B=2-109|2H76,A/B=2-109|2O8V,B=2-109|2TIR,A=2-109|2TRX,A/B=2-109|3D3YR,A/B=2-109

type ECO NP_417077.1 P0AGG4 Thioredoxin-2

type SCE YLR043C P22217 Thioredoxin-1 2I9H,A=1-103

type SCE YGR209C P22803 Thioredoxin-2 2FA4,A/B=1-104|2HSY,A=1-104
| species | gene_id | protein_id | description |
|---------|---------|------------|-------------|
| cel     | WBGene00021933 | Q9N357 | Putative uncharacterized protein |
| cel     | WBGene00015062 | Q09433 | Thioredoxin-1 |
| cel     | WBGene00007099 | Q17424 | Probable thioredoxin-2 |
| mmu     | ENSMUSP00000030051 | P10639 | Thioredoxin |
| mmu     | ENSMUSP0000005487 | P97493 | Thioredoxin, mitochondrial |
| hsa     | ENSP00000363641 | P10599 | Thioredoxin 1AIU,A=1-105|1AUC,A=1-105|1CQG,A=1-105|1CQH,A=1-105|1E1T,A=1-105|1ERV,A=1-105|1E22,A=1-105|1E24,A=1-105|1E26,A=1-105|1TRU,A=1-105|1TRV,A=1-105|1TRW,A=1-105|1W1C,A=1-105|1W1E,C=1-105|2HSH,A=1-105|2HXK,A/B/C=1-105|2IFQ,A/C=1-105, B=1-105|2IIY,A=1-105|3TRX,A=1-105|4TRX,A=1-105| |
| hsa     | ENSP00000216185 | Q99757 | Thioredoxin, mitochondrial |
| eco     | NP_418553.1 | P0A8N5 | Lysyl-tRNA synthetase, heat inducible |
| eco     | NP_417366.1 | P0A8N3 | Lysyl-tRNA synthetase 1BBU,A=2-504|1BBW,A=2-504|1KRS,A=31-149|1KRT,A=31-149| |
| sce     | YDR037W | P15180 | Lysyl-tRNA synthetase |
| cel     | WBGene00002238 | Q22099 | Lysyl-tRNA synthetase |
| mmu     | ENSMUSP00000090808 | Q99MN1 | Lysyl-tRNA synthetase |
| hsa     | ENSP00000325448 | Q15046 | Lysyl-tRNA synthetase |
| eco     | NP_418628.4 | P0A9L3 | FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase |
| eco     | NP_417806.1 | P45523 | FKBP-type peptidyl-prolyl cis-trans isomerase fkpA |
| sce     | YNL135C | P20081 | FK506-binding protein 1 1YAT,A=2-114| |
| cel     | WBGene00001427 | Q9U2Q8 | Protein Y18D10A.19, confirmed by transcript evidence |
| cel     | WBGene00001431 | Q45418 | Protein F31D4.3, confirmed by transcript evidence |
| mmu     | ENSMUSP00000032508 | P30416 | FK506-binding protein 4 |
| mmu     | ENSMUSP00000078382 | Q64378 | FK506-binding protein 5 |
| mmu     | ENSMUSP0000020964 | Q9Z212 | Peptidyl-prolyl cis-trans isomerase FKBP1B |
| hsa     | ENSP00000370373 | P68106 | Peptidyl-prolyl cis-trans isomerase FKBP1B |
| hsa     | ENSP00000349811 | Q13451 | FK506-binding protein 5 1KT0,A=1-457| |
| hsa     | ENSP00000371161 | P62942 | Peptidyl-prolyl cis-trans isomerase FKBP1A 1A7X,A/B=2-107|1B6C,A/C/E/G=2-107|1BKF,A=2-108|1BL4,A/B=2-107|1D6O,A/B=2-108|1D7H,A/B=2-108|1D7I,A/B=2-108|1D7J,A/B=2- |
108|EYM,A/B=2-108|1F40,A=2-108|1FAP,A=2-107|1FKB,A=2-108|1FKD,A=2-107|1FKF,A=2-108|1FKG,A=2-108|1FKH,A=2-108|1FKI,A/B=2-108|1FKJ,A=2-107|1FKR,A=2-108|1FKT,A=2-108|1J4H,A=2-107|1J4L,A=2-107|1J4R,A/B/D=2-107|1NSG,A=2-107|1QPF,A/D=2-107|1QPL,A/C=2-107|2DG3,A=2-108|2DG4,A=2-108|2DG9,A=2-108|2FAP,A=2-107|2FKE,A=2-107|2PPN,A=2-108|2PPP,A=2-108|3FAP,A=2-107|3H9R,B=1-108|4FAP,A=2-107|hsa ENSP0000001008 Q02790 FK506-binding protein 4 1N1A,A/B=2-139|1P5Q,A/B/C=146-458|1Q1C,A=2-259|1QZ2,A/B/C=145-458| eco NP_416171.1 P0AC69 Glutaredoxin-4 1YKA,A=1-115|2WCI,A/B=1-115| sce YPL059W Q02784 Monothiol glutaredoxin-5, mitochondrial| sce YER174C P32642 Monothiol glutaredoxin-4| sce YDR098C Q03835 Monothiol glutaredoxin-3 3D6I,A/B=37-148| cel WBGene00013029 Q9XTU9 Protein Y49E10.2, confirmed by transcript evidence| cel WBGene00017062 Q6EZG4 Putative uncharacterized protein| mmu ENSMUSP00000066621 Q9CQM9 Glutaredoxin-3 1WIK,A=241-336| mmu ENSMUSP00000021522 Q80Y14 Glutaredoxin-related protein 5| hsa ENSP00000328570 Q86SX6 Glutaredoxin-related protein 5 2WEM,A/B/C/D=35-150| hsa ENSP00000357634| eco NP_418415.1 P0AC69 DNA-directed RNA polymerase subunit beta' 2AUK,A/B/C/D/E=944-1129| sce YOR116C P04051 DNA-directed RNA polymerase III subunit RPC1| sce YDL140C P04050 DNA-directed RNA polymerase II subunit RPB1 1I3Q,A=1-1733|1I50,A=1-1733|1I6H,A=1-1733|1K83,A=1-1733|1NIK,A=1-1733|1NT9,A=1-1733|1PQV,A=1-1733|1R5U,A=1-1733|1R95,A=1-1733|1R9T,A=1-1733|1SFO,A=1-1733|1TWA,A=1-1733|1TWC,A=1-1733|1TWF,A=1-1733|1TWG,A=1-1733|1TWH,A=1-1733|1WCM,A=1-1733|1Y1V,A=1-1733|1Y1W,A=1-1733|1Y1Y,A=1-1733|1Y77,A=1-1733|2B63,A=1-1733|2B8K,A=1-1733|2E2H,A=1-1733|2E2I,A=1-1733|2E2J,A=1-1733|2JA5,A=1-1733|2JA6,A=1-1733|2JA7,A/M=1-1733|2JA8,A=1-1733|2NVQ,A=1-1733|2NVQ,A=1-1733|2NTB,A=1-1733|2NVX,A=1-1733|2NVY,A=1-1733|2NVZ,A=1-1733|2R7Z,A=1-1733|2R92,A=1-1733|2R93,A=1-1733|2VUM,A=1-1733|2YU9,A=1-1733|3CQZ,A=1-1733|3FKL,A=1-1733|3GTM,A=1-1733|3GTJ,A=1-1733|3GTK,A=1-1733|3GTL,A=1-1733|3GTM,A=1-1733|3GTO,A=1-1733|3GTP,A=1-1733|3GTLQ,A=1-1733|3H3V,B=1-1733|3HOA,A=1-1733|3HOA,A=1-1733|3HOV,A=1-1733|3HOU,A=1-1733|3HOV,A=1-1733|3HOZ,A=1-1733| cel WBGene00000123 P16356 DNA-directed RNA polymerase II subunit RPB1
cel WBGene00004411 Q18566 DNA-directed RNA polymerase
mmu ENSMUSP00000026322 B2RXC6 DNA-directed RNA polymerase
mmu ENSMUSP00000050771 P08775 DNA-directed RNA polymerase II subunit RPB1
hsa ENSP00000314949 P24928 DNA-directed RNA polymerase II subunit RPB1 2GHQ,C/D=1795-1803|2GHT,C/D=1796-1803|
hsa ENSP00000365773 Q9Y617 DNA-directed RNA polymerase III subunit RPC1

!!!!!!!!!!!!!! OG2_72501 !!!!!!!!!!!!!!!!
eco NP_415427.1 P23721 Phosphoserine aminotransferase 1BJN,A/B=3-362|1BJO,A=3-362,B=3-362|
sce YOR184W P33330 Phosphoserine aminotransferase
cel WBGene00009177 P91856 Probable phosphoserine aminotransferase
mmu ENSMUSP00000025542 Q99K85 Phosphoserine aminotransferase
hsa ENSP00000365773 Q9Y617 Phosphoserine aminotransferase 3E77,A/B/C=17-370|

!!!!!!!!!!!!!! OG2_70925 !!!!!!!!!!!!!!!!
eco NP_417161.1 P39452 Ribonucleoside-diphosphate reductase 2 subunit alpha
eco NP_416737.1 P00452 Ribonucleoside-diphosphate reductase 1 subunit alpha 1QFN,B=737-761|1R1R,A/B/C=1-761|1RLR,A=1-761|2R1R,A/B/C=1-761|3R1R,A/B/C=1-761|4R1R,A/B/C=1-761|5R1R,A/B/C=1-761|6R1R,A/B/C=1-761|7R1R,A/B/C=1-761|
sce YIL066C P21672 Ribonucleoside-diphosphate reductase large chain 2
sce YER070W P21524 Ribonucleoside-diphosphate reductase large chain 1 1ZYZ,A/B=1-888|1ZZD,A=1-888|2CVS,A=1-888|2CVT,A=1-888|2CVU,A=1-888|2CVV,A=1-888|2CVW,A=1-888|2CVX,A=1-888|2CVY,A=1-888|2EUD,A=1-888|2ZLF,A=1-888|2ZLG,A=1-888|
cel WBGene00004391 Q03604 Ribonucleoside-diphosphate reductase large subunit
mmu ENSMUSP00000033283 P07742 Ribonucleoside-diphosphate reductase large subunit
hsa ENSP00000300738 P23921 Ribonucleoside-diphosphate reductase large subunit2WGH,A/B=75-742|

!!!!!!!!!!!!!! OG2_71266 !!!!!!!!!!!!!!!!
eco NP_416163.1 P0AGD1 Superoxide dismutase [Cu-Zn] 1ESO,A=20-173|
sce YJR104C P00445 Superoxide dismutase [Cu-Zn] 1B4L,A=2-154|1B4T,A=2-154|1F18,A=1-154|1F1A,A=1-154|1F1D,A=1-154|1F1G,A/B/C/D/E/F=1-154|1JCV,A=2-154|1JK9,A/C=2-154|1SDY,A/B/C/D=2-154|1YAZ,A=2-154|1YSO,A=2-154|2JCW,A=2-154|
cel WBGene00007036 Q27538 Superoxide dismutase [Cu-Zn]
cel WBGene00004933 P34461 Extracellular superoxide dismutase [Cu-Zn]
cel WBGene00004930 P34697 Superoxide dismutase [Cu-Zn]
mmu ENSMUSP00000057271
mmu ENSMUSP0000023707 P08228 Superoxide dismutase [Cu-Zn]
212|1NMZ,A=1-212|1NN0,A=1-212|1NN1,A=1-212|1NN3,A=1-212|1NN5,A=1-212|

############ OG2_71370 ###############
eco NP_417298.1 P05458 Protease 3 1Q2L,A=24-962|
sce YLR389C Q06010 A-factor-processing enzyme
cel WBGene00013492 O62499 Protein Y70C5C.1, partially confirmed by
transcript evidence
cel WBGene00018426 Q10040 Putative zinc protease C28F5.4
cel WBGene00015360 Q17593 Putative uncharacterized protein
cel WBGene00015359 Q17592 Putative uncharacterized protein
mmu ENSMUSP00000025747 Q9JHR7 Insulin-degrading enzyme
hsa ENSP00000026598 P14735 Insulin-degrading enzyme

############ OG2_70996 ###############
eco NP_415003.1 P06710 DNA polymerase III subunit tau
  1JR3,A/B/C=1-373|1NJF, A/B/C/D=1-243|1NJG,A/B=1-243|1XXH,B/C/D/G/H=1-373|1XXI,B/C/D/G/H=1-368|2AYA,A=499-625|3GLF,B/C/D/G/H=1-373|3GLG,B/C/D/G/H=1-373|3GLH,B/C/D/G/H=1-373|3GLI,B/C/D/G/H=1-373|
sce YNL290W P38629 Replication factor C subunit 3 1SXJ,C=1-340|
cel WBGene00018409 P34429 Probable replication factor C subunit 5
mmu ENSMUSP00000083652 Q9D0F6 Replication factor C subunit 5
hsa ENSP000000229043 P40937 Replication factor C subunit 5
  1LFS,A=1-340|

############ OG2_70796 ###############
eco NP_418354.1 P0A858 Triosephosphate isomerase 1TMH,A/B/C/D=1-238|1TRE,A/B=1-255|
sce YDR050C P00942 Triosephosphate isomerase 1I45,A/B=2-247|1NEY,A/B=2-248|1NF0,A/B=2-248|1YPI,A/B=2-248|2YPI,A/B=2-248|3YPI,A/B=2-247|7TIM,A/B=2-247|
cel WBGene00006601 Q10657 Triosephosphate isomerase 1MO0,A/B=2-247|
mmu ENSMUSP00000024223 P17751 Triosephosphate isomerase
hsa ENSP00000037170 P60174 Triosephosphate isomerase 1HTI,A/B=2-249|1WY1,A/B=2-249|2IAM,P=23-37|2IAN,C/H/M/R=23-37|2JK2,A/B=2-249|2VOM,A/B/C/D=2-249|

############ OG2_70796 ###############
eco NP_417757.1 P0A759 30S ribosomal protein S13 1M5G,M=2-118|1P6G,M=2-118|1P87,M=2-118|1VS5,M=1-118|1VS7,M=1-118|2AVY,M=2-118|2AW7,M=2-118|2GY9,M=2-115|2GYB,M=2-115|2I2P,M=2-117|2I2U,M=2-117|2QAL,M=2-118|2QAN,M=2-118|2QB9,M=2-118|2QBB,M=2-
118|2QBD,M=2-118|2QBF,M=2-118|2QBH,M=2-118|2QBJ,M=2-118|2QOU,M=2-118|2QW,|2QOY,M=2-118|2QP0,M=2-118|2VHO,M=2-118|2VHP,M=2-118|2Z4K,M=2-118|2Z4M,M=2-118|3DF1,M=2-117|3DF3,M=2-117|3FIH,M=2-114
sce YNL081C P53937 37S ribosomal protein SWS2, mitochondrial
dsce YML026C P35271 40S ribosomal protein S18 1K5X,M=15-145|1S1H,M=15-146|1S1H,M=15-146|1S1H,M=15-146|1S1H,M=15-146|1S1H,M=15-146
sce YDR450W P35271 40S ribosomal protein S18 1K5X,M=15-145|1S1H,M=15-146|1S1H,M=15-146|1S1H,M=15-146|1S1H,M=15-146
sce YGL123W P25443 40S ribosomal protein S2 1K5X,E=75-223|1S1H,E=75-223|1S1H,E=75-223|1S1H,E=75-223|1S1H,E=75-223
sce YOL155C Q05164 Haze protective factor 1
sce YMR317W Q04893 Uncharacterized protein YMR317W
sce YKR102W P36170 Flocculation protein FLO10
sce YIR019C P08640 Flocculation protein FLO11
| Species | Gene ID | Accession | Description |
|---------|---------|-----------|-------------|
| *Saccharomyces cerevisiae* (sce) | YIL169C | P40442 | Putative uncharacterized protein YIL169C |
| *S. cerevisiae* (sce) | YHR211W | P38894 | Flocculation protein FLO5 |
| *S. cerevisiae* (sce) | YGR014W | P32334 | Protein MSB2 |
| *S. cerevisiae* (sce) | YCR089W | P25653 | Factor-induced gene 2 protein |
| *S. cerevisiae* (sce) | YAR050W | P32768 | Flocculation protein FLO1 |
| *S. cerevisiae* (sce) | YAL063C | P39712 | Flocculation protein FLO9 |
| *Candida elegans* (cel) | WBGene00022536 | O17084 | Putative uncharacterized protein |
| *C. elegans* (cel) | WBGene000021761 | Q76602 | Putative uncharacterized protein |
| *Mus musculus* (mmu) | ENSMUSP00000092953 | Q80T03 | Mucin-6 |
| *M. musculus* (mmu) | ENSMUSP00000049941 | Q6PZe0 | Mucin-19 |
| *M. musculus* (mmu) | ENSMUSP00000039699 | Q71sp1 | Gastric mucin |
| *M. musculus* (mmu) | ENSMUSP00000026590 | Q9Jky4 | Intestinal mucin |
| *M. musculus* (mmu) | ENSMUSP00000009631 | Q55225 | Otagelin |
| *M. musculus* (mmu) | ENSMUSP00000001995 | Q8Ciz8 | von Willebrand factor |
| *Homo sapiens* (hsa) | ENSP00000351956 | Q02817 | Mucin-2 |
| *H. sapiens* (hsa) | ENSP00000347118 | Q6W4X9 | Mucin-6 |
| *H. sapiens* (hsa) | ENSP00000341666 | Q8Wxi7 | Mucin-16 |
| *H. sapiens* (hsa) | ENSP00000304549 | Q9Hc84 | Mucin-5B |
| *H. sapiens* (hsa) | ENSP00000261405 | P04275 | von Willebrand factor 1A03,A/B=1686-1872,1ATZ,A/B=1685-1873,1AUQ,A=1261-1468,1FE8,A/B/C=1683-1874,1FNS,A=1271-1465,1JB,A=1263-1464,1JK,A=1263-1464,1M10,A=1261-1468,1OK,A=1271-1465,1SQ0,A=1260-1472,1U0n,A=1261-1468,1Uex,C=1260-1468,2ADF,A=1683-1874,3Gxb,A/B=1495-1671 |
| *Escherichia coli* (eco) | NP_417886.1 | P0AC86 | Glycogen phosphorylase |
| *E. coli* (eco) | YP_026218.1 | P00490 | Maltodextrin phosphorylase |
| *S. cerevisiae* (sce) | YPR160W | P06738 | Glycogen phosphorylase |
| *C. elegans* (cel) | WBGene00020696 | Q86nc1 | Phosphorylase |
| *M. musculus* (mmu) | ENSMUSP00000035743 | Q8c194 | Glycogen phosphorylase, brain form |
| *M. musculus* (mmu) | ENSMUSP00000047564 | Q9wub3 | Glycogen phosphorylase, muscle form |
| *M. musculus* (mmu) | ENSMUSP00000071231 | Q9et01 | Glycogen phosphorylase, liver form |
| *H. sapiens* (hsa) | ENSP00000216962 | P11216 | Glycogen phosphorylase, brain form |
| *H. sapiens* (hsa) | ENSP00000216392 | P06737 | Glycogen phosphorylase, liver form |
hsa ENSP00000164139 P11217 Glycogen phosphorylase, muscle form 1Z8D,A=1-842

################### OG2_71942 ###################

eco NP_416617.1 P00959 Methionyl-tRNA synthetase 1F4L,A=1-551|1MEA,A=139-164|1MED,A=139-164|1P7P,A=2-551|1PFU,A=2-551|1PFV,A=2-551|1PFW,A=2-551|1PYF,A=2-551|1PG0,A=2-551|1PG2,A=2-551|1QQT,A=2-552
sce YGR264C P00958 Methionyl-tRNA synthetase, cytoplasmic 2HSN,A=2-159
cel WBGene00003415 Q20970 Methionyl-tRNA synthetase, cytoplasmic

mmu ENSMUSP00000034646 Q68FL6 Methionyl-tRNA synthetase, cytoplasmic 2DJV,A=2-551

hsa ENSP00000262027 P56192 Methionyl-tRNA synthetase, cytoplasmic 2O05,A/B=1-302|2O06,A/B=1-302|2O07,A/B=1-302|2O0L,A/B=1-302

################### OG2_70963 ###################

cel WBGene00003422 Q9N3T8 Spermidine synthase

mmu ENSMUSP0000006611 Q64674 Spermidine synthase

hsa ENSP00000366156 P19623 Spermidine synthase 2O05,A/B=1-302|2O06,A/B=1-302|2O07,A/B=1-302|2O0L,A/B=1-302

################### OG2_70712 ###################

eco NP_417213.1 P23909 DNA mismatch repair protein mutS 1E3M,A=B=1-800|1NG9,A/B=1-800|1OH5,A/B=1-800|1OH6,A/B=1-800|1OH7,A=B=1-800|1OH8,A/B=1-800|1W7A,A/B=1-800|1WB9,A/B=1-800|1WBB,A/B=1-800|1WBD,A/B=1-800|2OK2,A/B=1-800

cel WBGene00003422 Q9N3T8 Spermidine synthase

mmu ENSMUSP00000024967 P43247 DNA mismatch repair protein Msh2

mmu ENSMUSP0000005503 P54276 DNA mismatch repair protein Msh6

hsa ENSP00000234420 P52701 DNA mismatch repair protein Msh6 2GFU,A=68-201|2O8B,B=341-1360|2O8C,B=341-1360|2O8D,B=341-1360|2O8E,B=341-1360|2O8F,B=341-1360

hsa ENSP00000233146 P43246 DNA mismatch repair protein Msh2 2O8B,A=1-934|2O8C,A=1-934|2O8D,A=1-934|2O8E,A=1-934|2O8F,A=1-934

################### OG2_70967 ###################
eco NP_417244.1 P38038 Sulfite reductase [NADPH] flavoprotein alpha-component 1DDG,A/B=226-599|1DDI,A=226-599|1YKG,A=53-218
sce YHR042W P16603 NADPH-cytochrome P450 reductase 2BF4,A/B=34-691|2BN4,A/B=34-691|2BPO,A/B=34-691|3FJO,A=44-211

mmu ENSMUSP00000005651 P37040 NADPH-cytochrome P450 reductase

hsa ENSP00000265302 P16435 NADPH-cytochrome P450 reductase 1B1C,A=61-241|3FJO,A=232-677

eco NP_417957.1 P06715 Glutathione reductase 1GER,A/B=1-450|1GES,A/B=1-450|1GET,A/B=1-450|1GEU,A/B=1-450
sce YPL091W P41921 Glutathione reductase 2HQM,A/B=17-483

cel WBGene00001262 Q09590 Abnormal embryogenesis protein 8, confirmed by transcript evidence

mmu ENSMUSP00000000561 P37040 NADPH-cytochrome P450 reductase

hsa ENSP00000265302 P16435 NADPH-cytochrome P450 reductase 1B1C,A=61-241|3FJO,A=232-677

mmu ENSMUSP00000081006 Q9JLT4 Thioredoxin reductase 2, mitochondrial 1ZDL,A=31-524|1ZKQ,A=31-524|3DGZ,A=34-521

mmu ENSMUSP0000033992 P47791 Glutathione reductase, mitochondrial

mmu ENSMUSP0000020484 Q9JMH6 Thioredoxin reductase 1, cytoplasmic

mmu ENSMUSP0000000828 Q99MD6 Thioredoxin reductase 3

hsa ENSP00000373506

hsa ENSP00000347020 Q16881 Thioredoxin reductase 1, cytoplasmic 1W1C,A/B=161-649|2CFY,A/B/C/D/E/F=151-649|2J3N,A/B/C/D/E/F=151-649

hsa ENSP00000353329 Q86VQ6 Thioredoxin reductase 3 3H8Q,A/B=162-267

hsa ENSP00000334518

hsa ENSP00000221130 P00390 Glutathione reductase, mitochondrial 1ALG,A=480-503|1BWC,A=45-522|1DNC,A=45-522|1GRA,A=45-522|1GRB,A=45-522|1GRE,A=45-522|1GRF,A=45-522|1GRG,A=45-522|1GRH,A=45-522|1GRT,A=45-522|1GRG,A=45-522|1NSN,A=45-522|1QF,A=62-522|1XAN,A=62-522|2AAQ,A=44-522|2GH5,A/B=45-522|2GRT,A=62-522|3DGJ,X=62-522|3DJJ,A=45-522|3DK4,A=45-522|3DK8,A=62-522|3DK9,A=45-522|3GRS,A=45-522|3GRT,A=62-522|4GR1,A=45-522|4GRT,A=62-522|5GRT,A=62-522

eco NP_415520.1 P36659 Curved DNA-binding protein

sce YNL007C P25294 Protein SIS1 1C3G,A=180-349|2B26,A/B/C=181-352|2O37,A=1-89

cel WBGene00001031 Q20774 Protein F54D5.8, confirmed by transcript evidence
| Species | Gene ID | Accession | Description |
|---------|---------|-----------|-------------|
| mmu     | ENSMUSP00000095716 | O89114 | DnaJ homolog subfamily B member 5 |
| mmu     | ENSMUSP00000059074 | Q80Y75 | DnaJ homolog subfamily B member 13 |
| mmu     | ENSMUSP00000053916 | Q9D832 | DnaJ homolog subfamily B member 4 |
| mmu     | ENSMUSP0000005620 | Q9QYJ3 | DnaJ homolog subfamily B member 1 |
| hsa     | ENSP00000344431 | P59910 | DnaJ homolog subfamily B member 13 |
| hsa     | ENSP00000359799 | Q9UDY4 | DnaJ homolog subfamily B member 4 |
| hsa     | ENSP00000368026 | O75953 | DnaJ homolog subfamily B member 5 |
| hsa     | ENSP00000254322 | P25685 | DnaJ homolog subfamily B member 1 |

1HDJ, A=1-76|2QLD, A=158-340|

#### OG2_70651 ######

| Species | Gene ID | Accession | Description |
|---------|---------|-----------|-------------|
| eco     | NP_417485.4 | Q46857 | 2,5-diketo-D-gluconic acid reductase A |
| eco     | NP_416295.1 | P76234 | Uncharacterized protein yeaE |
| eco     | NP_414743.1 | P30863 | 2,5-diketo-D-gluconic acid reductase B |
| sce     | YOR120W | P14065 | Protein GCY |
| sce     | YJR096W | P47137 | Uncharacterized oxidoreductase YJR096W |
| sce     | YHR104W | P38715 | NADPH-dependent aldose reductase GRE3 |
| sce     | YDR368W | Q12458 | Putative reductase 1 |
| cel     | WBGene00022887 | Q09632 | Uncharacterized oxidoreductase ZK1290.5 |
| cel     | WBGene00013896 | Q23320 | Protein ZC443.1, confirmed by transcript evidence |
| cel     | WBGene00012722 | Q9NAI5 | Protein Y39G8B.1a, confirmed by transcript evidence |
| cel     | WBGene00020369 | Q22352 | Putative uncharacterized protein T08H10.1 |
| cel     | WBGene0009980 | P91997 | Protein F53F1.2, confirmed by transcript evidence |
| cel     | WBGene00016443 | Q18483 | Putative uncharacterized protein |
| cel     | WBGene00015565 | P91020 | Putative uncharacterized protein |

1QWK, A=1-317|

| Species | Gene ID | Accession | Description |
|---------|---------|-----------|-------------|
| mmu     | ENSMUSP0000093525 | Q8BIV6 | Putative uncharacterized protein |
| mmu     | ENSMUSP0000039114 | P21300 | Aldose reductase-related protein 1 |
| mmu     | ENSMUSP0000007449 | Q9DCT1 | 1,5-anhydro-D-fructose reductase |
| mmu     | ENSMUSP0000048830 | Q8VCX1 | 3-oxo-5-beta-steroid 4-dehydrogenase |
| mmu     | ENSMUSP0000040244 | P45377 | Aldose reductase-related protein 2 |
| mmu     | ENSMUSP0000030455 | Q9JI6 | Alcohol dehydrogenase [NADP+] |
| mmu     | ENSMUSP0000100045 | P45376 | Aldose reductase |
| hsa     | ENSP00000352584 | O60218 | Aldo-keto reductase family 1 member B10 |
| hsa     | ENSP00000298375 | Q96JD6 | 1,5-anhydro-D-fructose reductase |
hsa ENSP00000242375 P51857 3-oxo-5-beta-steroid 4-dehydrogenase
3BUR,A/B=1-326|3B UV,A/B=1-326|3BV7,A/B=1-326|3C AQ,A/B=1-326|3C AS,A/B=1-326|3CA V,A/B=1-326|3CMF,A/B=1-326|3COT,A/B=1-326|3D OP,A/B=1-326|3G1R,A/B=1-326
hsa ENSP00000361140 P14550 Alcohol dehydrogenase [NADP+]
2ALR,A=2-325
hsa ENSP00000285930 P15121 Aldose reductase 1ABN,A=2-316|1ADS,A=2-316|1AZ1,A=2-316|1AZ2,A=2-316|1EF3,A/B=2-315|1EL3,A=1-316|1E I1,A=1-316|1MAR,A=2-316|1PWM,A=1-316|1T40,A=1-316|1T41,A=1-316|1USO,A=1-316|1X96,A=1-316|1X97,A=1-316|1X98,A=1-316|1XGD,A=2-315|1Z3N,A=1-316|1Z89,A=1-316|1Z8A,A=1-316|2ACQ,A=2-315|2ACR,A=2-315|2ACS,A=2-315|2ACU,A=2-315|2AGT,A=1-316|2DUX,A=1-316|2DUZ,A=1-316|2DV0,A=1-316|2F2K,A=1-316|2FZ8,A=1-316|2FZ9,A=1-316|2FZB,A=1-316|2FZD,A=1-316|2HV5,A=1-316|2HVN,A=1-316|2HVO,A=1-316|2I16,A=1-316|2I17,A=1-316|2IKG,A=1-316|2IKH,A=1-316|2IKI,A=1-316|2IKJ,A=1-316|2INE,A=2-315|2INZ,A=2-315|2IPW,A=2-315|2IQ0,A=2-315|2I QD,A=2-315|2IS7,A=2-315|2ISF,A=2-315|2J8T,A=1-316|2NV C,A=1-316|2P D5,A=1-316|2P D9,A=1-316|2PD B,A=1-316|2PD C,A=1-316|2PD F,A=1-316|2PD G,A=1-316|2PD H,A=1-316|2PDI,A=1-316|2PD J,A=1-316|2PDK,A=1-316|2PD L,A=1-316|2PD M,A=1-316|2PD N,A=1-316|2PD P,A=1-316|2PD Q,A=1-316|2PD U,A=1-316|2PD W,A=1-316|2PD X,A=1-316|2PD Y,A=1-316|2PEV,A=1-316|2PF8,A=1-316|2PF H,A=1-316|2PZN,A=1-316|2QWX,A=1-316|2R24,A=1-316|3BCJ,A=1-316|3DN5,A=1-316|3G5E,A=1-316|3GHR,A=1-316|3GHS,A=1-316|3GHT,A=1-316|3GHU,A=1-316|

OG2_70788

eco YP_026263.1
sce YMR190C P35187 ATP-dependent helicase SGS1
1D8B,A=1271-1351|
cel WBGene00001865 O18017 Bloom syndrome protein homolog
WBGene000019334 Q5DX50 Putative uncharacterized protein
WBGene00006944 Q19546 Probable Werner syndrome ATP-dependent helicase homolog 1

mmu ENSMUSP00000033990 O09053 Werner syndrome ATP-dependent helicase homolog 2
2E6L,A=31-238|2E6M,A=31-238|

mmu ENSMUSP0000035463

mmu ENSMUSP0000098394 Q3UFL0 Putative uncharacterized protein

hsa ENSP00000349859

hsa ENSP00000298139 Q14191 Werner syndrome ATP-dependent helicase 2
2AXL,A=949-1092|2DGZ,A=1140-1239|2E1E,A=1142-1242|2E1F,A=1142-1242|2FBT,A=38-236|2FBV,A=38-236|2FBX,A=38-236|2FBY,A=38-236|2FCO,A=38-236|
hsa ENSP00000318727 P46063 ATP-dependent DNA helicase Q1
2V1X,A/B=49-616|

OG2_70978
sce  YGR088W  P06115  Catalase T
sce  YDR256C  P15202  Peroxisomal catalase A  1A4E,A/B/C/D=15-502|
cel  WBGene00000830  O61235  Catalase-2
cel  WBGene00000831  Q27487  Peroxisomal catalase 1
cel  WBGene00013220  Q8MYL7  Catalase
mmu  ENSMUSP00000028610  P24270  Catalase
hsa  ENSP00000241052  P04040  Catalase 1DG,B/A/C/D=4-501|1DG,F,A/B/C/D=5-501|1DG,G,A/B/C/D=5-501|1DG,H,A/C=4-501, B/D=4-501|1F4J,A/B/C/D=1-527|1QQW,A/B/C/D=1-527|

 OG2_73734

eco  NP_416659.1  P33018  S-formylglutathione hydrolase yeiG
eco  NP_414889.1  P51025  S-formylglutathione hydrolase frmB
cel  YJL068C  P40363  S-formylglutathione hydrolase

 OG2_71183

eco  NP_418381.1  P32669  Fructose-6-phosphate aldolase 2
eco  NP_416959.1  P0A867  Transaldolase A
eco  NP_415346.4  P78055  Fructose-6-phosphate aldolase 1

 OG2_70778

eco  NP_414710.1  P0AE18  Methionine aminopeptidase 1

cel  YLR244C  Q01662  Methionine aminopeptidase 1
eco NP_417402.1 P0A9B6  D-erythrose-4-phosphate dehydrogenase
eco NP_416293.1 P0A9B2  Glyceraldehyde-3-phosphate dehydrogenase A
sce YJR009C P00358  Glyceraldehyde-3-phosphate dehydrogenase 2
sce YJL052W P00360  Glyceraldehyde-3-phosphate dehydrogenase 1
sce YGR192C P00359  Glyceraldehyde-3-phosphate dehydrogenase 3
cel WBGene00001683 P04970  Glyceraldehyde-3-phosphate dehydrogenase 1

cel WBGene00001684 P17329  Glyceraldehyde-3-phosphate dehydrogenase 2

cel WBGene00001685 P17330  Glyceraldehyde-3-phosphate dehydrogenase 3

cel WBGene00001686 P17331  Glyceraldehyde-3-phosphate dehydrogenase 4

mmu ENSMUSP00000097539
mmu ENSMUSP00000097256
mmu ENSMUSP00000082931
mmu ENSMUSP00000096371
mmu ENSMUSP00000091648
mmu ENSMUSP00000090946
mmu ENSMUSP00000089510
mmu ENSMUSP00000075954 Q64467  Glyceraldehyde-3-phosphate dehydrogenase, testis-specific

hsa ENSP00000229239 P04406  Glyceraldehyde-3-phosphate dehydrogenase

hsa ENSP00000222286 Q14556  Glyceraldehyde-3-phosphate dehydrogenase, testis-specific

eco NP_414556.1 P08622  Chaperone protein dnaJ 1BQ0,A=2-104|1BQZ,A=2-78|1EXK,A=131-209|1XBL,A=2-108|
sce YFL016C P35191  DnaJ homolog 1, mitochondrial
cel WBGene00001028 Q8TA83  DnaJ homolog dnj-10

mmu ENSMUSP0000053842 Q99M87  DnaJ homolog subfamily A member 3, mitochondrial

hsa ENSP00000262375 Q96EY1  DnaJ homolog subfamily A member 3, mitochondrial 2CTT,A=213-303|2DN9,A=93-158|

eco NP_417260.1 P0A7E5  CTP synthase 1S1M,A/B=1-545|2AD5,A/B=1-545|
sce YJR103W P38627  CTP synthase 2
sce YBL039C P28274  CTP synthase 1
cel  WBGene00012316  Q9XXN1  Protein W06H3.3, partially confirmed by
transcript evidence

mmu  ENSMUSP00000033727  P70303  CTP synthase 2

mmu  ENSMUSP00000030381  P70698  CTP synthase 1

hsa  ENSP00000361699  P17812  CTP synthase 1  2VO1,A/B=1-273|

hsa  ENSP00000369590  Q9NRF8  CTP synthase 2  2V4U,A=297-562|2VK,T,A=297-562|

############ OG2_70913 ###############

eco  NP_415017.1  Q59385  Copper-exporting P-type ATPase A

sce  YDR270W  P38995  Copper-transporting ATPase 1FVQ,A=2-72|1FVS,A=2-72|1UV1,B=2-72|1UV2,B=2-72|2GGP,B=2-72|

cel  WBGene000000834  B9DI72  Protein Y76A2A.2b, confirmed by transcript
evidence

mmu  ENSMUSP00000058840  A2AG68  ATPase, Cu++ transporting, alpha
polypeptide

mmu  ENSMUSP00000006742  Q64446  Copper-transporting ATPase 2

hsa  ENSP00000355170  Q04656  Copper-transporting ATPase 1

  1AW0,A=375-446|1KVI,A=1-79|1KVJ,A=1-79|1Q8L,A=164-246|1S6O,A=169-240|1S6U,A=169-240|1Y3J,A=486-558|1Y3K,A=486-558|1YJR,A=562-633|1YJT,A=562-633|1YJU,A=562-633|1YJV,A=562-633|2AW0,A=375-446|2G9O,A=275-352|2GA7,A=275-352|2K1R,A=5-77|3CJL,B=7-77|

hsa  ENSP00000242839  P35670  Copper-transporting ATPase 2

  2ARF,A=1036-1196|2EW9,A=486-633|2ROP,A=238-439|

############ OG2_71350 ###############

eco  NP_417304.1  P0A884  Thymidylate synthase 1AIQ,A/B=1-264|1AJM,A=1-264|1AN5,A/B=1-264|1AOB,A=1-264|1AXW,A/B=1-264|1BDU,A=1-264|1BID,A=1-264|1BJG,A=1-264|1BQ1,A/B=1-264|1BQ2,A=1-264|1DDU,A/B=1-264|1DNA,A/B=1-264|1EV8,A=1-264|1EVF,A=1-264|1EVG,A=1-264|1F4B,A=1-264|1F4C,A/B=1-264|1F4D,A/B=1-264|1F4E,A=1-264|1F4F,A/B=1-264|1F4G,A/B=1-264|1FFL,A=1-264|1FWM,A/B=1-264|1JG0,A/B=1-264|1JTQ,A/B=1-264|1JTU,A/B=1-264|1JUT,A/B=1-264|1KCE,A/B=1-264|1KZI,A/B=1-264|1KZJ,A/B/C/D/E/F=1-264|1NCE,A/B=1-264|1QQQ,A=1-264|1SYN,A/B=1-264|1TDU,A/B=1-264|1TDS,A/B=1-264|1TRG,A=1-264|1TSD,A/B=1-264|1TNN,A=1-264|1TY5,A=1-264|1ZPR,A/B=1-264|2A9W,A/B/C/D=1-264|2BBQ,A/B=1-264|2FTN,A=1-264|2FTO,X=1-264|2FTQ,A=1-264|2G8M,A/B=1-264|2G8O,A/B=1-264|2G8X,A/B=1-264|2KCE,A/B=1-264|2TSC,A/B=1-264|2VET,A=1-264|2VF0,A/B=1-264|3B5B,A/B=1-264|3B9H,A=1-264|3BFI,A=1-264|3BGX,A=1-264|3BHL,A/B=1-264|3BHR,A=1-264|3TMS,A=1-264|

sce  YOR074C  P06785  Thymidylate synthase

cel  WBGene00022455  Q9N588  Thymidylate synthase

mmu  ENSMUSP00000026846  P07607  Thymidylate synthase
hsa ENSP00000315644 P04818 Thymidylate synthase 1HVY,A/B/C/D=26-313|1HW3,A=1-313|1HW4,A=1-313|1HZW,A/B=30-313|1100,A/B=30-313|1IU6,A/B/C/D=1-313|1IJU5,A/B/C/D=1-313|1YPV,A=1-313|2ONB,A=1-313|2RD8,A=1-313, B=1-313|2RDA,A/B/D/E/F=1-313#
## OG2_71352 #######
eco NP_415465.1 P0A7E1 Dihydroorotate dehydrogenase 1F76,A/B/D/E=1-336#
eco NP_415933.1 P25553 Lactaldehyde dehydrogenase 2HG2,A=2-478|2ILU,A=2-478|2IMP,A=2-478|2OPX,A=1-479#
sce YBR006W P38067 Succinate-semialdehyde dehydrogenase [NADP+]#
eco NP_417147.1 P25526 Succinate-semialdehyde dehydrogenase [NADP+]#
eco NP_415933.1 P25553 Lactaldehyde dehydrogenase 2HG2,A=2-478|2ILU,A=2-478|2IMP,A=2-478|2OPX,A=1-479#
sce YBR006W P38067 Succinate-semialdehyde dehydrogenase [NADP+]#
eco NP_416714.1 P25526 Succinate-semialdehyde dehydrogenase [NADP+]#
sce YBR006W P38067 Succinate-semialdehyde dehydrogenase [NADP+]#
eco NP_415933.1 P25553 Lactaldehyde dehydrogenase 2HG2,A=2-478|2ILU,A=2-478|2IMP,A=2-478|2OPX,A=1-479#
sce YBR006W P38067 Succinate-semialdehyde dehydrogenase [NADP+]#
eco NP_416173.1 P0AGD3 Superoxide dismutase [Fe] 1ISA,A/B=1-193|1SB,A/B=1-193|1SC,A/B=1-193|1SA5,A/B=2-192|2BKB,A/B/C/D=2-192|2NYB,A/B/C/D=2-192
sce YHR008C P00447 Superoxide dismutase [Mn], mitochondrial 3BFR,A=27-233|
cel WBGene00004931 P31161 Superoxide dismutase [Mn] 1, mitochondrial 3DC6,A/C=25-221|
cel WBGene00004932 P41977 Superoxide dismutase [Mn] 2, mitochondrial 3DC5,A/C=25-218|
mmu ENSMUSP0000007012 P09671 Superoxide dismutase [Mn], mitochondrial

hsa ENSP00000356022 P04179 Superoxide dismutase [Mn], mitochondrial 1AP5,A/B=25-222|1AP6,A/B=25-222|1EM1,A/B=25-222|1JA8,A/B=25-222|1LLV,A/B=25-222|1LUU,A/B=25-222|1MSD,A/B=25-222|1N0J,A/B=25-222|1N0N,A/B=25-222|1PL4,A/B/C/D=25-222|1PM9,A/B=25-222|1QNM,A/B=25-222|1SXZ,A/B=25-222|1VAR,A/B=25-222|1XDC,A/B=25-222|1XI,A/B=25-222|1ZSP,A/B=25-222|1ZTE,A/B/C/D=25-222|1ZUQ,A/B=25-222|2ADP,A=25-222|2ADQ,B=25-222|2GDS,A/B/C/D=25-222|2P4K,A/B/C/D=25-222|2QKA,A/C=25-220|2QKC,A/C=25-220|3C3S,A/B=25-222|3C3T,A/B=25-222|

eco NP_417259.1 P0A6P9 Enolase 1E9I,A/B/C/D=2-432|2 FYM,A/C/D/F=2-432|
sce YPL281C Q12007 Enolase-related protein 1/2
sce YOR393W Q12007 Enolase-related protein 1/2
sce YMR323W P42222 Enolase-related protein 3
sce YHR174W P00925 Enolase 2
sce YGR254W P00924 Enolase 1 1EBG,A/B=2-436|1EBH,A/B=2-436|1ELS,A=2-436|1L8P,A/B/C/D=2-437|1NEL,A=2-436|1ONE,A/B=2-437|1P43,A/B=2-437|1P48,A/B=2-437|2AL1,A/B=2-436|2AL2,A/B=2-437|2ONE,A/B=2-437|4ENL,A=2-436|5ENL,A=2-436|6ENL,A=2-436|7ENL,A=2-436|8ENL,A=2-436|9ENL,A=2-436|
cel WBGene00011884 Q27527 Enolase

mmu ENSMUSP00000079727 P17182 Alpha-enolase
mmu ENSMUSP00000065681
mmu ENSMUSP00000075513 P17182 Alpha-enolase
mmu ENSMUSP0000004378 P17183 Gamma-enolase

hsa ENSP00000352320 B9ZVW5 Enolase
hsa ENSP00000229277 P09104 Gamma-enolase 1TE6,A/B=2-434|2AKM,A/B=2-433|2AKZ,A/B=2-433|
hsa ENSP00000324105 P13929 Beta-enolase
hsa ENSP00000234590 P06733 Alpha-enolase 2PSN,A/B/C/D=1-434|3B97,A/B/C/D=2-434|

OG2_70770 OG2_70959
| Species | Accession   | Description                                      |
|---------|-------------|--------------------------------------------------|
| eco     | NP_417637.1 | Translation initiation factor IF-2 1ND9,A=2-50|1Z01,J=388-888 |
| sce     | YOL023W     | Translation initiation factor IF-2, mitochondrial |
| cel     | WBGene00009771 | Translation initiation factor IF-2 |
| mmu     | ENSMUSP00000020749 | Translation initiation factor IF-2, mitochondrial 2CRV,A=621-727 |
| hsa     | ENSP00000263629 | Translation initiation factor IF-2, mitochondrial |
|---------|-------------|--------------------------------------------------|
| eco     | NP_418189.1 | ATP synthase gamma chain 1D8S,G=--|1FS0,G=19-248 |
| sce     | YBR039W     | ATP synthase subunit gamma, mitochondrial         |
| cel     | WBGene00022089 | Putative uncharacterized protein |
| mmu     | ENSMUSP0000026887 | ATP synthase gamma chain |
| hsa     | ENSP00000349142 | ATP synthase subunit gamma, mitochondrial |
|---------|-------------|--------------------------------------------------|
| eco     | NP_418208.1 | Ribokinase 1GQT,A/B/C/D=1-309|1RK2,A/B/C/D=1-309|1RKA,A=1-309|1RKD,A=1-309|1RKS,A=1-309 |
| sce     | YCR036W     | Probable ribokinase                              |
| cel     | WBGene00008548 | Protein F07A11.5, partially confirmed by transcript evidence |
| mmu     | ENSMUSP0000031018 | Ribokinase |
| hsa     | ENSP00000306817 | Ribokinase 2FV7,A/B=11-322 |
|---------|-------------|--------------------------------------------------|
| eco     | NP_416994.1 | Phosphoribosylformylglycinamidine cyclo-ligase    |1CLI,A/B/C/D=2-345 |
| sce     | YGL234W     | Bifunctional purine biosynthetic protein ADE5,7 |
| cel     | WBGene00018174 | Putative uncharacterized protein |
| mmu     | ENSMUSP0000023684 | Trifunctional purine biosynthetic protein adenosine-3 |
| hsa     | ENSP00000371253 | Trifunctional purine biosynthetic protein adenosine-3 1MEJ,A/B/C=810-1010|1MEN,A/B/C=810-1010|1MEO,A=808-1010|1NJS,A/B=808-1010|1RBM,A/B=808-1010|1RBQ,A/B/C/D=808-1010|1RBY,A/B/C/D=808-1010|1RBZ,A/B=808-1010|1RC0,A/B=808-1010|1RC1,A/B=808-1010|1ZLX,A=808-1010|1ZLY,A=808-1010|2QK4,A/B=1-430|2V9Y,A/B=467-794 |
|---------|-------------|--------------------------------------------------|
| eco     | NP_418374.1 | Cystathionine gamma-synthase 1CS1,A/B/C/D=1-386 |
| sce     | YGL184C     | Cystathionine beta-lyase                          |
| sce     | YAL012W     | Cystathionine gamma-lyase 1N8P,A/B/C/D=2-393 |
cel  WBGene00022856  P55216  Putative cystathionine gamma-lyase
cel  WBGene00009048  O45391  Protein F22B.6, confirmed by transcript
evidence
mmu  ENSMUSP00000029830
hsa  ENSP00000359976  P32929  Cystathionine gamma-lyase
2NMP,A/B/C/D=1-402|3COG,A/B/C/D=1-402|3ELP,A/B/C/D=1-405|
#OG2_70804#
eco  NP_417013.1  P0A763  Nucleoside diphosphate kinase
2HUR,A/B/C/D/E/F=2-142|
sce  YKL067W  P36010  Nucleoside diphosphate kinase 3B54,A/B=1-153|
cel  WBGene00009119  Q93576  Nucleoside diphosphate kinase
mmu  ENSMUSP00000024978  Q9WV85  Nucleoside diphosphate kinase 3
mmu  ENSMUSP00000021220  P15532  Nucleoside diphosphate kinase A
mmu  ENSMUSP00000025007  Q9WV84  Nucleoside diphosphate kinase, mitochondrial
mmu  ENSMUSP00000021217  Q01768  Nucleoside diphosphate kinase B
hsa  ENSP00000372398  O00746  Nucleoside diphosphate kinase, mitochondrial 1EHW,A/B=36-175|
hsa  ENSP00000219302  Q13232  Nucleoside diphosphate kinase 3
1IZ6,A/B/D=1-169|
hsa  ENSP00000365572  P22392  Nucleoside diphosphate kinase B
1NSK,L/N/O/R/T/U=1-152|1NUE,A/B/C/D/E/F=2-152|3BBB,A/B/C/D/E/F=2-152|3BBC,A/B/C/D/E/F=2-152|3BBF,A/B/C/D/E/F=2-152|
#OG2_71042#
eco  NP_418062.1  P33232  L-lactate dehydrogenase [cytochrome]
sce  YML054C  P00175  Cytochrome b2, mitochondrial
1FCB,A/B=81-591|1KBI,A/B=81-591|1KBJ,A/B=180-591|1LCO,A/B=81-591|1LDC,A/B=81-591|1LTD,A/B=86-591|1LCW,A/B=182-591|1HZE,A/B=81-591|1ZSF,A/B=81-591|1SZG,A/B=81-591|2OZ0,A/B=81-591|
cel  WBGene00018286  B1GRK5  Putative uncharacterized protein
mmu  ENSMUSP00000029464  Q9NYQ2  Hydroxyacid oxidase 2
mmu  ENSMUSP00000028704  Q9WU19  Hydroxyacid oxidase 1
hsa  ENSP00000358428  Q9NYQ3  Hydroxyacid oxidase 2
hsa  ENSP00000368066  Q9UJM8  Hydroxyacid oxidase 1 2NZL,A=1-370|2RDT,A=1-370|2RDU,A=1-370|2RDW,A=1-370|
#OG2_71043#
eco  NP_415613.1  P0AA15  3-oxoacyl-[acyl-carrier-protein] synthase 2
1B3N,A=2-413|1KAS,A=2-413|2GFX,A=2-412|2GFW,A=2-412|2GFX,A=2-412|2GFY,A=2-412|3G0Y,A=2-413|3G11,A=2-413|
sce  YER061C  P39525  3-oxoacyl-[acyl-carrier-protein] synthase homolog
cel  WBGene00008667  Q6A1T5  Protein F10G8.9b, partially confirmed by transcript
evidence
mmu  ENSMUSP00000022311  Q9D404  3-oxoacyl-[acyl-carrier-protein]
synthase, mitochondrial
hsa ENSP00000280701 Q9NWU1 3-oxoacyl-[acyl-carrier-protein] synthase, mitochondrial 2C9H,A=39-459|2IWY,A/B=38-459|2IWZ,A/B=38-459#

######### OG2_71327 #########

eco NP_414714.1 P0A805 Ribosome-recycling factor 1EK8,A=1-185|1ISE,A=1-185|1ZNO,A=1-185|1ZN1,A=1-185|2RDO,8=1-185#
sce YHR038W P38771 Ribosome-recycling factor, mitochondrial
cel WBGene00020625 P91478 Probable ribosome-recycling factor, mitochondrial

mmu ENSMUSP0000028250 Q9D6S7 Ribosome-recycling factor, mitochondrial 1WIH,A=116-186#

hsa ENSP00000312608

hsa ENSP00000343867 Q96E11 Ribosome-recycling factor, mitochondrial

######### OG2_71048 #########

eco NP_417429.1 P52061 Nucleoside-triphosphatase rdgB 1K7K,A=1-197|2PYU,A=1-197|2Q16,A/B=1-197#
sce YJR069C P47119 Protein HAM1
cel WBGene00001823 Q9GYG4 HAM-1-like protein

mmu ENSMUSP0000097279

mmu ENSMUSP0000099482 Q9D892 Inosine triphosphate pyrophosphatase

hsa ENSP00000369456 Q9BY32 Inosine triphosphate pyrophosphatase 2CAR,A/B=1-194|2ISD,A=1-194|2J4E,A/B/C/D/E/F/G/H=1-194#

######### OG2_71532 #########

eco NP_417381.1 P27248 Aminomethyltransferase 1VLO,A=2-363#
sce YDR019C P48015 Aminomethyltransferase, mitochondrial
cel WBGene00017765 Q22968 Aminomethyltransferase

mmu ENSMUSP0000035230 Q8CFA2 Aminomethyltransferase, mitochondrial

hsa ENSP00000273588 P48728 Aminomethyltransferase, mitochondrial 1WSR,A/B=29-403|1WSV,A/B=29-403#

######### OG2_71430 #########

eco NP_417389.1 P0A720 Ribose-5-phosphate isomerase A 1KS2,A/B=1-219|1LKZ,A/B=1-219|1O8B,A/B=1-219#
sce YOR095C Q12189 Ribose-5-phosphate isomerase 1XTZ,A=1-258#

cel WBGene00015101 P41994 Probable-ribose 5-phosphate isomerase

mmu ENSMUSP0000064158 P47968 Ribose-5-phosphate isomerase

hsa ENSP00000283646 P49247 Ribose-5-phosphate isomerase

######### OG2_70886 #########

eco NP_415612.1 P0A6A8 Acyl carrier protein 1ACP,A=2-78|1LOH,A=2-76|1LOI,A=1-78|1T8K,A=2-77|2FAC,A/B=2-77|2FAD,A/B=2-77|2FAE,A/B=2-77|2FH5,C=1-78|2K92,A=2-78|2K93,A=2-78|2K94,A=2-78|3EJB,A/C/E/G=1-78|3EJD,A/C/E/G=1-78|3EJE,A/C/E/G=1-78#
sce YKL192C P32463 Acyl carrier protein, mitochondrial
cel WBGene00013237 Q9U241 Acyl carrier protein

mmu ENSMUSP0000033157 Q9CR21 Acyl carrier protein, mitochondrial
hsa  ENSP0000007516  O14561  Acyl carrier protein, mitochondrial
  2DNW,A=71-156|

########################################## OG2_70885 ####################################

eco  NP_415133.1  P77806  Aminotransferase ybdL  1U08,A/B=1-386|
sce  YJL060W  P47039  Probable kynurenine--oxoglutarate transaminase
  BNA3  3B46,A/B=1-444|
cel  WBGene00010984  Q5FC69  Protein R03A10.4b, partially confirmed by
  transcript evidence|
cel  WBGene00009232  Q9XX97  Protein F28H6.3, partially confirmed by
  transcript evidence|

mmu  ENSMUSP00000041675  Q8BTY1  Kynurenine--oxoglutarate
  transaminase 1

hsa  ENSP0000302227  Q16773  Kynurenine--oxoglutarate transaminase 1
  1W7L,A=1-422|1W7M,A=1-422|1W7N,A=1-422|3FVS,A/B=1-422|3FVU,A/B=1-422|3FVX,A/B=1-422|

hsa  ENSP0000260508  Q6YP21  Kynurenine--oxoglutarate transaminase 3

########################################## OG2_71233 ####################################

eco  NP_417388.1  P0A9T0  D-3-phosphoglycerate dehydrogenase
  1PSD,A/B=2-410|1SC6,A/B/C/D=7-409|1YBA,A/B/C/D=1-410|2P9C,A/B=1-410|2P9E,A/B/C/D=1-410|2P9G,A/B=1-410|
sce  YIL074C  P40510  D-3-phosphoglycerate dehydrogenase 2|
sce  YER081W  P40054  D-3-phosphoglycerate dehydrogenase 1|
cel  WBGene00007836  O17626  Protein C31C9.2, confirmed by transcript
  evidence|

mmu  ENSMUSP00000064755  Q61753  D-3-phosphoglycerate dehydrogenase

hsa  ENSP0000358417  O43175  D-3-phosphoglycerate dehydrogenase
  2G76,A/B=4-314|

########################################## OG2_71100 ####################################

eco  NP_418096.4  P0ABQ0  Coenzyme A biosynthesis bifunctional protein
  coaBC  1U7U,A=181-406|1U7W,A/B/C=181-406|1U7Z,A/B/C=181-406|1U80,A/B/C=181-406|
sce  YKL088W  P36076  Uncharacterized protein YKL088W|
cel  WBGene00009138  P91988  Protein F25H9.6, confirmed by transcript
  evidence|

mmu  ENSMUSP0000082856  Q8BZB2  Phosphopantothenoylcysteine
decarboxylase|

hsa  ENSP0000343190  Q96CD2  Phosphopantothenoylcysteine decarboxylase
  1QZU,A/B/C/D=1-204|

########################################## OG2_70872 ####################################

eco  NP_415698.1  P76004  Uncharacterized protein ycgM
  1NR9,A/B/C/D=1-219|
sce  YNL168C  P53889  Uncharacterized mitochondrial hydrolase FMP41
  1NKQ,A/B/C/D/E/F=1-259|
cel  WBGene00022798  P34673  Uncharacterized protein ZK688.3
| Species | Accession | ID | Description |
|---------|-----------|----|-------------|
| mmu     | ENSMUSP00000055827 | Q8R0F8 | Fumarylacetoacetate hydrolase domain-containing protein 1 |
| mmu     | ENSMUSP00000028848 | Q3TC72 | Fumarylacetoacetate hydrolase domain-containing protein 2A |
| hsa     | ENSP00000372112 | Q6P587 | Fumarylacetoacetate hydrolase domain-containing protein 1 |
| hsa     | ENSP00000272610 | Q6P213 | Fumarylacetoacetate hydrolase domain-containing protein 2B |
| hsa     | ENSP00000233379 | Q96GK7 | Fumarylacetoacetate hydrolase domain-containing protein 2A |
| eco     | NP_416154.1 | P0AGJ9 | Tyrosyl-tRNA synthetase |
| eco     | NP_415252.1 | P07014 | Succinate dehydrogenase iron-sulfur subunit |
| eco     | NP_417703.1 | P61889 | Malate dehydrogenase |
| sce     | YPL097W | P48527 | Tyrosyl-tRNA synthetase, mitochondrial |
| cel     | WBGene00006968 | Q8MPU1 | Tyrosinyl trna synthetase protein 1, isoform b |
| mmu     | ENSMUSP0000055277 | Q8BYL4 | Tyrosyl-tRNA synthetase, mitochondrial |
| hsa     | ENSP00000320658 | Q9Y2Z4 | Tyrosyl-tRNA synthetase, mitochondrial |
| eco     | NP_418577.1 | P0AC47 | Fumarate reductase iron-sulfur subunit |
| eco     | NP_415252.1 | P07014 | Succinate dehydrogenase iron-sulfur subunit |
| eco     | NP_417703.1 | P61889 | Malate dehydrogenase |
| sce     | YKL085W | P17505 | Malate dehydrogenase, mitochondrial |
| cel     | WBGene00003162 | O02640 | Probable malate dehydrogenase, mitochondrial |
| mmu     | ENSMUSP0000019323 | P08249 | Malate dehydrogenase, mitochondrial |
hsa  ENSP00000327070  P40926  Malate dehydrogenase, mitochondrial
2DFD,A/B/C/D=20-338|

#############################  OG2_70939 #############################

eco  NP_418449.1  P0A6T1  Glucose-6-phosphate isomerase
sce  YBR196C  P12709  Glucose-6-phosphate isomerase
cel  WBGene00013597  Q7K707  Glucose-6-phosphate isomerase
mmu  ENSMUSP00000049355  P06745  Glucose-6-phosphate isomerase

1U0E,A/B=1-558|1U0F,A/B=1-558|1U0G,A/B=1-558|2CVP,A/B=1-558|2CXN,A/B=1-558|2CXO,A/B=1-558|2CXP,A/B=1-558|2CXQ,A/B=1-558|2CXR,A/B=1-558|2CXS,A/B=1-558|2CTX,A/B=1-558|2CXU,A/B=1-558|

hsa  ENSP00000348877  P06744  Glucose-6-phosphate isomerase
1IAT,A=2-558|1IRI,A/B/C/D=1-558|1JIQ,A/B/C/D=1-558|1JLH,A/B/C/D=1-558|1NUH,A=1-558|

Supplementary Table 6.

Ortholog groups from alpha-proteobacteria [endosymbionts (*) and free species]

* Rickettsia typhi_wilmington
* Orientia
* Anaplasma
* Wolbachia
Rhodospirillum
Magnetospirillum_magneticum_AMB-1
Silicibacter_TM1040
Erythrobacter_litoralis_HTCC2594

Functional groups for proteins with orthologs in all 8 species

| ORF | gene in Rickettsia typhi wilmington |
|-----|-----------------------------------|
|     | TRANSLATION                       |
| 1   | gi|51473841|ref|YP_067598.1| |
| 51  | gi|51474009|ref|YP_067766.1| |
| 80  | gi|51473231|ref|YP_066988.1| |
| 129 | gi|51473332|ref|YP_067089.1| |
|     | RIBOSOMAL PROTEIN                  |
| 4   | gi|51473838|ref|YP_067595.1| |
| 5   | gi|51473837|ref|YP_067594.1| |
| 6   | gi|51473836|ref|YP_067593.1| |
| 7   | gi|51473835|ref|YP_067592.1| |
|     | RIBOSOME BIOGENESIS                |
| 368 | gi|51473248|ref|YP_067005.1| |
400 gi|51473858|ref|YP_067615.1|

REPLICATION
60 gi|51473761|ref|YP_067518.1|
223 gi|51474004|ref|YP_067761.1|
285 gi|51473779|ref|YP_067536.1|
286 gi|51473607|ref|YP_067364.1|
332 gi|51473371|ref|YP_067128.1|
390 gi|51473372|ref|YP_067529.1|
487 gi|51473946|ref|YP_067703.1|
551 gi|51474034|ref|YP_067790.1|
573 gi|51473426|ref|YP_067183.1|
619 gi|51473901|ref|YP_067658.1|
652 gi|161610765|ref|YP_067311.2|

DNA REPAIR
329 gi|51473311|ref|YP_067068.1|
338 gi|51474012|ref|YP_067769.1|

tRNA MODIFICATION
54 gi|51473604|ref|YP_067361.1|
598 gi|51474015|ref|YP_067772.1|
605 gi|51473287|ref|YP_067044.1|
1091 gi|51473501|ref|YP_067258.1|
1260 gi|51473927|ref|YP_067684.1|

tRNA SYNTHETASE
243 gi|51473351|ref|YP_067108.1|
244 gi|51473350|ref|YP_067107.1|
261 gi|51473521|ref|YP_067278.1|
389 gi|51473803|ref|YP_067560.1|
425 gi|51473344|ref|YP_067101.1|
428 gi|51473609|ref|YP_067366.1|
607 gi|51473263|ref|YP_067020.1|
623 gi|51473951|ref|YP_067708.1|

TRANSCRIPTION
27 gi|51473815|ref|YP_067572.1|
208 gi|51473610|ref|YP_067367.1|
355 gi|51473296|ref|YP_067053.1|
426 gi|51473650|ref|YP_067407.1|
592 gi|51473499|ref|YP_067256.1|
593 gi|51474029|ref|YP_067786.1|
596 gi|51474026|ref|YP_067783.1|

CELL DIVISION
123 gi|51473844|ref|YP_067601.1|
602 gi|51473284|ref|YP_067041.1|
603 gi|51473285|ref|YP_067042.1|
685 gi|51474010|ref|YP_067767.1|
AMINO ACID BIOSYNTHESIS

FATTY ACID BIOSYNTHESIS

REGULATION AND SENSING

HOST RESISTANCE

RESISTANCE

STRESS RESPONSE
Supplementary Table 7.

Ortholog groups from gamma-proteobacteria [endosymbionts (*) and free species]

- Buchnera sp. (Acyrthosiphon pisum)*
- Wigglesworthia brevipalpis*
- Candidatus Blochmannia pennsylvanicus*
- Marinomonas MWYL1
- Escherichia coli
- Pseudomonas aeruginosa

Functional groups for proteins with orthologs in all 6 species

| ORF / gene in Buchnera sp. / description |
|-----------------------------------------|
| **Translation**                          |
| 393 gi|15617022|ref|NP_240235.1| sulfate adenylate transferase subunit 1 |
| 491 gi|28952057|ref|NP_240333.2| elongation factor Tu |
| 443 gi|15617072|ref|NP_240285.1| trigger factor |
| 239 gi|15616868|ref|NP_240081.1| GTP-binding protein Era |
| 492 gi|15617121|ref|NP_240334.1| elongation factor G |
| 507 gi|15617136|ref|NP_240349.1| peptide chain release factor 3 |
| 295 gi|15616924|ref|NP_240137.1| translation initiation factor IF-1 |
| 219 gi|15616848|ref|NP_240061.1| elongation factor Ts |
| 160 gi|15616790|ref|NP_240002.1| peptide chain release factor 1 |
| 179 gi|15616809|ref|NP_240021.1| peptidyl-tRNA hydrolase |
| 406 gi|15617035|ref|NP_240248.1| peptide chain release factor 2 |
| 116 gi|15616746|ref|NP_239958.1| translation initiation factor IF-3 |
| 19 gi|15616649|ref|NP_239861.1| elongation factor P |
| 180 gi|15616810|ref|NP_240022.1| translation-associated GTPase |
| 459 gi|15617088|ref|NP_240301.1| hypothetical protein BU494 |
| 353 gi|15616982|ref|NP_240195.1| translation initiation factor IF-2 |
| **Ribosomal protein**                   |
| 490 gi|15617119|ref|NP_240332.1| 30S ribosomal protein S10 |
| 489 gi|15617118|ref|NP_240331.1| 50S ribosomal protein L3 |
| 488 gi|15617117|ref|NP_240330.1| 50S ribosomal protein L4 |
| 487 gi|15617116|ref|NP_240329.1| 50S ribosomal protein L23 |
| 486 gi|15617115|ref|NP_240328.1| 50S ribosomal protein L2 |
| 485 gi|15617114|ref|NP_240327.1| 30S ribosomal protein S19 |
| 484 gi|15617113|ref|NP_240326.1| 50S ribosomal protein L22 |
| 483 gi|15617112|ref|NP_240325.1| 30S ribosomal protein S3 |
| 482 gi|15617111|ref|NP_240324.1| 50S ribosomal protein L16 |
| 480 gi|15617109|ref|NP_240322.1| 50S ribosomal protein S17 |
| 479 gi|15617108|ref|NP_240321.1| 50S ribosomal protein L14 |
| 478 gi|15617107|ref|NP_240320.1| 50S ribosomal protein L24 |
| 477 gi|15617106|ref|NP_240319.1| 50S ribosomal protein L5 |
| 476 gi|15617105|ref|NP_240318.1| 30S ribosomal protein S14 |
| 475 gi|15617104|ref|NP_240317.1| 30S ribosomal protein S8 |
| 474 gi|161353756|ref|NP_240316.3| 50S ribosomal protein L6 |
Ribosome biogenesis

221 | 15616850| ref| NP_240063.1 | ribosome recycling factor
376 | 15617005| ref| NP_240218.1 | ribosomal large subunit pseudouridine synthase D
402 | 15617031| ref| NP_240244.1 | GTPase EngB
563 | 15617192| ref| NP_240405.1 | GTP-binding protein EngA
22 | 15616652| ref| NP_239864.1 | hypothetical protein BU023
240 | 15616869| ref| NP_240082.1 | ribonuclease III
131 | 15616761| ref| NP_239973.1 | dimethyladenosine transferase
348 | 15616977| ref| NP_240190.1 | ATP-dependent RNA helicase DeaD
370 | 15616999| ref| NP_240212.1 | 16S rRNA processing protein RimM
358 | 15616987| ref| NP_240200.1 | cell division protein FtsJ
383 | 15617012| ref| NP_240225.1 | hypothetical protein BU410

Replication

10 | 15616640| ref| NP_239852.1 | DNA gyrase subunit B
11 | 15616641| ref| NP_239853.1 | DNA polymerase III beta chain
510 | 28952059| ref| NP_240352.2 | replicative DNA helicase
225 | 15616854| ref| NP_240067.1 | DNA polymerase III alpha chain
424 | 15617053| ref| NP_240266.1 | exodeoxyribonuclease V 135 kDa
555 | 15617184| ref| NP_240397.1 | ATP-dependent DNA helicase Rep
| ID   | gi|15616799|ref|NP_240011.1| DNA gyrase subunit A |
|------|---|----------|---|------------|---------------------|
| 512  | gi|15617141|ref|NP_240354.1| hypothetical protein BU548 |

**DNA repair**

| ID   | gi|15616739|ref|NP_239951.1| endonuclease III |
|------|---|----------|---|------------|----------------|
| 516  | gi|15617145|ref|NP_240358.1| A/G-specific adenine glycosylase |

**tRNA modification**

| ID   | gi|15616818|ref|NP_240030.1| tRNA pseudouridine synthase A |
|------|---|----------|---|------------|-----------------------------|
| 161  | gi|15616791|ref|NP_240003.1| HemK protein |
| 308  | gi|15616937|ref|NP_240150.1| hypothetical protein BU328 |
| 101  | gi|15616731|ref|NP_239943.1| cell cycle protein MesJ |
| 16   | gi|28952039|ref|NP_239858.2| tRNA modification GTPase TrmE |
| 351  | gi|15616980|ref|NP_240193.1| tRNA pseudouridine 55 synthase |

**tRNA synthetase**

| ID   | gi|15617007|ref|NP_240220.1| alanyl-tRNA synthetase |
|------|---|----------|---|------------|----------------------|
| 229  | gi|15616858|ref|NP_240071.1| arginyl-tRNA synthetase |
| 501  | gi|15617130|ref|NP_240343.1| tryptophanyl-tRNA synthetase |
| 63   | gi|15616693|ref|NP_239905.1| glutamyl-tRNA synthetase |
| 385  | gi|15617014|ref|NP_240227.1| glutaminyl-tRNA synthetase |
| 296  | gi|15616925|ref|NP_240138.1| aspartyl-tRNA synthetase |
| 336  | gi|15616965|ref|NP_240178.1| asparaginyl-tRNA synthetase |
| 407  | gi|15617036|ref|NP_240249.1| lysyl-tRNA synthetase |
| 543  | gi|15617172|ref|NP_240385.1| lysyl-tRNA synthetase |
| 414  | gi|15617043|ref|NP_240256.1| leucyl-tRNA synthetase |
| 455  | gi|15617084|ref|NP_240297.1| cysteinyl-tRNA synthetase |
| 293  | gi|15616922|ref|NP_240135.1| seryl-tRNA synthetase |
| 100  | gi|15616730|ref|NP_239942.1| methionyl-tRNA synthetase |
| 119  | gi|15616749|ref|NP_239961.1| phenylalanyl-tRNA synthetase subunit |
| 120  | gi|15616750|ref|NP_239962.1| phenylalanyl-tRNA synthetase beta chain |
| 115  | gi|15616745|ref|NP_239957.1| threonyl-tRNA synthetase |

**Transcription**

| ID   | gi|15617093|ref|NP_240306.1| DNA-directed RNA polymerase subunit alpha |
|------|---|----------|---|------------|------------------------------------------|
| 38   | gi|15616668|ref|NP_240880.1| transcription antitermination protein NusG |
| 33   | gi|15616663|ref|NP_239875.1| DNA-directed RNA polymerase subunit beta |
| 32   | gi|15616662|ref|NP_239874.1| DNA-directed RNA polymerase subunit beta' |
| 509  | gi|15617138|ref|NP_240351.1| single-strand binding protein |
| 553  | gi|15617182|ref|NP_240395.1| transcription termination factor Rho |
| 330  | gi|15616959|ref|NP_240172.1| DNA polymerase III delta' subunit |
| 450  | gi|15617079|ref|NP_240292.1| DNA polymerase III subunits gamma and tau |
| 24   | gi|15616654|ref|NP_239866.1| RNA polymerase factor sigma-32 |
| 359  | gi|15616988|ref|NP_240201.1| transcription elongation factor GreA |
| 51   | gi|15616681|ref|NP_239893.1| DNA primase |
| 50   | gi|15616680|ref|NP_239892.1| RNA polymerase sigma factor RpoD |
| 231  | gi|15616860|ref|NP_240073.1| DNA polymerase III epsilon chain |

**Cell division**

| ID   | gi|15616831|ref|NP_240043.1| cell division protein FtsZ |
|------|---|----------|---|------------|----------------------------|
| 306  | gi|15616935|ref|NP_240148.1| septum site-determining protein MinD |
Amino acid biosynthesis

42  gi|15616672|ref|NP_239884.1| acetylornithine deacetylase
86  gi|15616716|ref|NP_239928.1| succinyl-diamino-pimelate desuccinylase
550 gi|28952060|ref|NP_240392.2| diaminopimelate epimerase
87  gi|15616717|ref|NP_239929.1| dihydrodipicolinate synthase

Fatty acid biosynthesis

247 gi|15616876|ref|NP_240089.1| enoyl-[acyl-carrier-protein] reductase (NADH)
327 gi|15616956|ref|NP_240169.1| 3-oxoacyl-[acyl-carrier protein] reductase
328 gi|15616957|ref|NP_240170.1| acyl carrier protein
83  gi|15616713|ref|NP_239925.1| 3-oxoacyl-[acyl-carrier-protein] synthase I

Regulation and sensing

302 gi|15616931|ref|NP_240144.1| cold shock-like protein CspC
457 gi|15617086|ref|NP_240299.1| cold shock protein CspE
242 gi|15616871|ref|NP_240084.1| GTP-binding protein LepA
267 gi|15616896|ref|NP_240109.1| extragenic suppressor protein SuhB

Stress response

178 gi|15616808|ref|NP_240020.1| superoxide dismutase
171 gi|15616801|ref|NP_240013.1| alkyl hydroperoxide reductase
403 gi|15617032|ref|NP_240245.1| GTP-binding protein BipA
176 gi|15616806|ref|NP_240018.1| hypothetical protein BU187
554 gi|15617183|ref|NP_240396.1| thioredoxin
Supplementary Material – Part III

Protocol to compute relative complexation propensities

Nonadaptive origins of interactome complexity

Ariel Fernández¹,² and Michael Lynch³

¹Department of Bioengineering, Rice University, Houston, TX 77005, USA
²Department of Computer Science, The University of Chicago, Chicago, IL 60637
³Department of Biology, Indiana University, Bloomington, IN 47405

To determine the complexation propensity of a protein subunit relative to its ortholog in *E. coli* (species 1, control) we take advantage of the scaling/proportionality shown in Supplementary Figure 1. Thus, we estimate \( M_{j,n} = \frac{[\Delta G_{if}]_{j,n} - [\Delta G_{if}]_{j,1}}{[\Delta G_{if}]_{j,1}} \) as \( M_{j,n} \approx \frac{\nu_{j,n} - \nu_{j,1}}{\nu_{j,1}} \), where \( \nu_{j,n} \) = \( \nu \)-value of protein from species \( n \) in ortholog group \( j \) (\( j=1,2,\ldots,106 \)). The protein sequences in each of the 106 ortholog groups for the 36 species are provided in FASTA format upon request. The estimation proved to be adequate as inferred from the similarity of the trends shown in Fig. 2a and Supplementary Fig. 7 and from the tight correlation shown in Supplementary Fig. 1. In accord with this proportionality, the free energy increment \( \Delta G_{if} \) associated with spanning the P-W interface is estimated at \( \Delta G_{if} = (3.91 \text{kJ/mol})B \nu \), where \( B \) = number of backbone hydrogen bonds in the protein, and therefore, \( B \nu \) = number of SABHBs. The physical quantity -(3.91±0.67)kJ/mol is the experimentally obtained free energy release upon exogenous wrapping of a single SABHB¹, hence the contribution of each individual SABHB to the over-all interfacial free energy cost is (3.91±0.67)kJ/mol.

To compute the \( \nu \)-value of a free protein subunit, we use the program YAPview also known as *Dehydron Calculator* (dehydron is an alternative name for SABHB), whose executable can be obtained upon request or can be freely
The code for the program uses **OpenGL®** for rendering, and the interface was done with **Borland Delphi®**. YAPview is a Windows® application that identifies SABHBs by evaluating the bond microenvironments determined by the structural coordinates of the protein chain contained in a PDB text file. The SABHBs are identified by loading the PDB file in text format or the homology model in the same format of a PDB file (Methods), choosing a protein chain within the PDB entry, a suitable structure display/representation, and enabling the desolvation calculation that becomes operative within the structural representation. The latter is needed to determine the extent of intramolecular dehydration of BHBs. The desolvation computation becomes active through the YAPview selection sequence: Configuration → General Options → Desolvation. Thereafter, one needs to enable the desolvation calculator and select the appropriate parameters, especially desolvation radius and desolvation threshold $\zeta_{\text{thr}}$, according to the specifications provided in Methods. Unless otherwise specified, the computation will use default values for all parameters. The only acceptable defaults are those describing the geometric constraints of the hydrogen bond (bond angles, bond lengths, etc.). YAPview identifies SABHBs using the criterion $\zeta \leq \zeta_{\text{thr}}$, where $\zeta$ is a local parameter quantifying the extent of protection of a BHB that YAPview computes based on the structural coordinates of the side-chain groups contained in the uploaded PDB text file.

By selecting an adequate representation of the protein structure, YAPView displays the SABHBs directly on the protein surface. The easiest and clearest rendering to visualize SABHBs can be obtained by representing residues solely by their alpha carbons, while the backbone conformation is displayed in a simplified manner using virtual bonds joining alpha carbons (see Fig. 2b for illustration). Compatible with this simplified rendering, the BHBs can be displayed as segments (bars) joining the alpha carbons of the residues paired by the BHB. The hydrogen bonds that are poorly dehydrated intramolecularly, that is, below the pre-selected threshold ($\zeta \leq \zeta_{\text{thr}}$) are shown in green in the structure display.
while the well-wrapped ones ($\zeta > \zeta_{\text{thr}}$) are shown in light grey. Higher levels of structural detail are possible, for example full atomistic display of backbone hydrogen-bond donor (amide) and acceptor (carbonyl). In this all-atom backbone rendering, the BHBs should be represented as lines joining the specific atoms O and H, rather than the more schematic representation suited for the alpha-carbon-only residue representation.

For a given set of parameters defining the desolvation threshold, SABHBs and well-wrapped BHBs can be identified using YAPview to yield the parameter $\psi$.

Template-based three-dimensional structures for orthologs lacking PDB-reported structure are constructed using MODELLER, with side chains directly positioned with SCWRL as indicated in Methods, using as templates the PDB-reported structures indicated in Supplementary Table 5. Once the PDB-text file has been generated for the highest ranked homology model, it is uploaded and interrogated in YAPview exactly in the same way as for any text file extracted from PDB, and the computed structural deficiency given by wrapping parameter $\psi$ may be cross-validated against a sequence-based computation of $\psi$, as described below.

The sequence-based computations require that we upload protein sequences from the ortholog groups stored in FASTA format into the predictor of native disordered regions (PONDR®)\textsuperscript{2,3}. The latter can be operated upon registration from the website: [http://www.pondr.com/](http://www.pondr.com/)

For our analysis presented in Fig. 2a,c and Supplementary Fig. 7, the sequence-based identification of SABHBs and relative complexation propensities of proteins with no reported structure are derived from reliable scores of native disorder propensity generated by PONDR-VLXT®. SABHBs belong to a twilight zone between order and native disorder and hence can be obtained from the disorder score ($0 \leq f_d \leq 1$) assigned to each residue within a sliding window, representing the propensity of the residue to be in a disordered region ($f_d=1$, certainty of disorder; $f_d=0.8$=disorder threshold; $f_d=0$, certainty of order). The strong correlation (Supplementary Fig. 5) between the disorder score of a residue and extent of protection of the hydrogen bond engaging the residue (if
any) provides a sequence-based method of inference of SABHBs. Thus, for residues below the disorder threshold $f_d=0.8$, defined by PONDR (tutorial in http://www.pondr.com/), SABHBs occur in regions where the disorder score lies in the range $0.35 \leq f_d < 0.8$ for twilight regions spanning between order ($f_d << 0.5$) and disorder ($f_d >> 0.5$). These regions are loosely structured but with marginal BHB protection at $\zeta \leq 19$ (Supplementary Fig. 5). The wrapping parameter $\nu$ is extracted from the PONDR computation by determining the disorder-based counterpart: $\nu \approx \nu_d = J(\text{twilight})/J$, where $J(\text{twilight}) = \text{number of residues in twilight region } 0.35 \leq f_d < 0.8$, and $J = \text{number of “structured” residues below the disorder threshold } f_d < 0.8$.

As an illustration, the complexation propensities of two orthologs in the first group OG2_70651 (2,5 diketo-D-gluconic acid reductase B) corresponding to E. Coli (ECO ortholog) and H. sapiens (HSA ortholog) are compared using the sequence-based disorder analysis. The PONDR results for the two orthologs are shown below, giving $\nu_d=0.17$ and 0.37 for the E. coli and human protein, respectively.
The thermodynamic parameter $\Delta G_f$ is also accessible from the sequence-based computation on account of the fact that $\Delta G_f = (3.91 \text{kJ/mol})B \approx (3.91 \text{kJ/mol})B_v\phi_d$ which gives the operational equation:

$$\Delta G_f = (3.91 \text{kJ/mol}).B.J(\text{twilight})/J$$
This equation combines disorder-propensity information with the experimentally obtained free energy associated with exogenous wrapping of a SABHB\(^1\) and is used to compute complexation propensities, notwithstanding the absence of PDB-reported protein structures.

1. Fernández, A. & Scott, R. Adherence of packing defects in soluble proteins. *Phys. Rev. Lett.* **91**, 018102 (2003).

2. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).

3. Garner, E., Romero, P., Dunker, A. K., Brown, C. & Obradovic, Z. Predicting binding regions within disordered proteins, *Genome Informatics* **10**, 41-50 (1999).