Survey of Large Protein Complexes in *D. vulgaris* Reveals Great Structural Diversity

Bong-Gyoon Han[a,1], Ming Dong[b,1], Haichuan Liu[c], Lauren Camp[d], Jil Geller[d], Mary Singer[d], Terry C. Hazen[d], Megan Choi[b], H. Ewa Witkowska[c], David A. Ball[a], Dieter Typke[a], Kenneth H. Downing[a], Maxim Shatsky[e,f],Steven E. Brenner[e,f], John-Marc Chandonia[e,f], Mark D. Biggin[b], and Robert M. Glaeser[f,2]

[a]Life Sciences, [b]Genomics, [d]Earth Sciences, and [f]Physical Biosciences Divisions, Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720; [c]OB/GYN Department, University of California San Francisco Sandler-Moore Mass Spectrometry Core Facility, University of California, San Francisco, CA 94143; [e]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720

An unbiased survey has been made of the stable, most abundant multi-protein complexes in *Desulfovibrio vulgaris* Hildenborough (*Dv*H) that are larger than Mr ~400 k. The quaternary structures for 8 of the 16 complexes purified during this work were determined by single-particle reconstruction of negatively stained specimens, a success rate ~10 times greater than that of previous ''proteomic'' screens. In addition, the subunit compositions and stoichiometries of the remaining complexes were determined by biochemical methods. Our data show that the structures of only two of these large complexes, out of the 13 in this set that have recognizable functions, can be modeled with confidence based on the structures of known homologs. These results indicate that there is significantly greater variability in the way that homologous prokaryotic macromolecular complexes are assembled than has generally been appreciated. As a consequence, we suggest that relying solely on previously determined quaternary structures for homologous proteins may not be sufficient to properly understand their role in another cell of interest.

Key words: comparative evolutionary analysis, single-particle electron microscopy, structural homology

Large, multiprotein complexes are known to form crucial, modular units of function in all cells. The view that cells, at least those as small as bacteria, are little more than ''bags of second-order chemical reactions'' has thus been replaced with the view that they are better described as a collection of ''protein machines'' (1). As a result, stable multiprotein complexes have become key targets for research to define the protein-interaction networks that exist in model organisms such as yeast or *E. coli*. In the case of other cells, however, it is not yet known (*i*) to what extent the properties of multiprotein complexes can be well-modeled on the basis of what is known for previously characterized, homologous complexes, and (*ii*) the extent to which cells, especially those as ''simple'' as bacteria, organize these machines within the cytoplasm, i.e., the extent to which even the simplest cells are more than just bags of multiprotein complexes. The emerging tool of cryo-EM tomography is uniquely well-suited to address these questions by ''imaging the entire proteome'' (2, 3) with sufficient resolution to distinguish many of the larger macromolecular complexes. The goal toward which the characterization of subcellular ultrastructure is thus moving, illustrated schematically in *SI Appendix*, Fig. S1, is to search a given tomographic volume with a variety of known templates to locate specific multiprotein complexes of interest (4, 5). Part of the anticipated power of this approach is that it would build upon knowledge gained in previous work using X-ray crystallography, multidimensional NMR spectroscopy, single-particle electron microscopy (EM), and sophisticated modeling (2, 6, 7).

As part of a large program to characterize and image the ensemble of macromolecular complexes in *Desulfovibrio vulgaris* Hildenborough (http://pcap.lbl.gov), a bacterium of potential use in bioremediation of soils contaminated by toxic heavy metals (8–10), we have undertaken a survey of the most abundant complexes large enough to be distinguishable from one another within tomographic reconstructions of single cells. Fifteen different macromolecular complexes with particle weights of at least 400 kDa were isolated by a ''tagless'' strategy (11), which is ''unbiased'' in that it makes no prior assumptions about which protein complexes should be purified. Instead, purification used a high-throughput pipeline that includes differential solubility in ammonium sulfate, ion-exchange chromatography, hydrophobic interaction chromatography, and size-exclusion chromatography. Also, *Dv*H ribosomes were isolated by a special-purpose protocol similar to that used for the purification of 70S *E. coli* ribosomes.

It has been thought that reliable templates for a large fraction of a cell's multiprotein complexes could be obtained from information already available. (We used the following resources in the current work: Protein Data Bank, www.rcsb.org/pdb/home/home.do; Electron Microscopy Data Bank, www.ebi.ac.uk/msd-srv/emsearch/index.html; 3D Complex.org web site, http://supfam.mrc-lmb.cam.ac.uk/elevy/3dcomplex/Home.cgi; and Protein Quaternary Structure database, http://pqs.ebi.ac.uk.) Although this premise may be true for prokaryotic machines such as ribosomes, chaperonins, or the core enzyme of RNA polymerase, we have found that conservation of subunit composition and quaternary structure is not at all guaranteed. On the contrary, we now report that quaternary structures of such complexes vary much more than has generally been appreciated.

Results

Purification and Identification of the Most Abundant Large Protein Complexes That Exist Within *Desulfovibrio vulgaris* Hildenborough (*Dv*H). Because multiprotein complexes within *Dv*H had not been catalogued, we used a ''tagless'' method to purify, identify, and structurally characterize those complexes that remain stable upon cell lysis. This method makes no assumptions about what proteins might exist in the form of multiprotein complexes, or what subunit stoichiometries and quaternary structures of these complexes should be. Instead,

comigrating protein subunits are separated on the basis of their physical properties, and the constituent polypeptides are identified by mass spectroscopy. The resulting proteomic survey reported here is intentionally limited to complexes with molecular mass greater than ≈400 kDa and copy number greater than ≈100 per cell, because these would be the easiest ones to identify in EM tomograms due to their size and abundance. One of the complexes (phosphoenolpyruvate synthase) that was isolated in this way proved, however, to be a 265 kDa homodimer, which eluted during size-exclusion chromatography (SEC) as an ≈370 kDa particle. Electron microscopy of this particle subsequently showed it to have an elongated shape, thus explaining its anomalously high apparent molecular weight in SEC.

The biochemical identities and subunit compositions of the 15 "largest, most abundant particles" that we found within *Dv*H are given in Table 1. Three of this set proved to be homo-oligomeric complexes of proteins (DVU0631, DVU0671, and DVU1012, respectively) for which no biochemical function could be identified or for which only weak similarity to proteins with known functions could be detected. Ten of the remaining 12 protein complexes whose biochemical functions could be identified with confidence are ones involved either in energy metabolism or in pathways of intermediary metabolism. The two remaining particles, GroEL and RNA polymerase, were already expected to be among the set of abundant particles in the desired size range.

Structural Characterization of Purified Multiprotein Complexes. Three-dimensional reconstructions were obtained at a resolution of 3 nm or better for 70S ribosomes in addition to 7 of 15 complexes purified by the high-throughput, tagless pipeline. In addition, the values of particle weight obtained by size-exclusion chromatography and native gel electrophoresis were used to estimate the subunit stoichiometries of those complexes for which single-particle EM reconstructions were not successful. Images of the eight 3D reconstructions that were successful, shown in Fig. 1, illustrate the fact that each such particle has a characteristic size and shape by which it could be identified. The extent to which diverse particles can be distinguished on the basis of their sizes and shapes supports the proposal that it will be possible to identify and localize a large number of different macromolecular complexes within cryo-EM tomograms, provided that these are obtained with a resolution in the range of 3 nm or better.

The preparation of samples for electron microscopy does not always produce specimens suitable for obtaining three-dimensional reconstructions, and as a result structures were not obtained for 8 of 15 complexes purified by the tagless approach. In some cases it appeared that the particles might be inherently flexible or polymorphic in structure, but in other cases we believe that the particles were easily damaged at some step during preparation for electron microscopy. Our success rate in producing informative 3D reconstructions is nevertheless at least 10 times higher than that reported in an earlier survey of complexes in the yeast proteome (12), possibly because our focus on characterizing only the largest such complexes. In addition, we took further time to optimize the details of preparing EM grids for each type of protein whenever the initial results looked promising, but there nevertheless was more heterogeneity than expected. Although the fraction of purified complexes for which we were able to get good three-dimensional reconstructions was thus relatively high, we believe that generic improvements in preparing single-particle samples for electron microscopy (rather than further biochemical purification of samples) could further improve the success rate and throughput.

Apart from GroEL, phosphoenolpyruvate synthase, and the 70S ribosome, all of the 10 remaining complexes whose biochemical identities can be assigned with confidence were found to have subunit stoichiometries or quaternary structures that are not fully conserved, even within bacteria, as is shown in column 7 and 8 of Table 1. The extent to which quaternary structures vary between different bacteria is quite surprising, because tertiary structure is normally well conserved over great evolutionary distance and because the quaternary structures of some well-known homomeric (e.g., GroEL) and heteromeric (e.g., RNAP II core enzyme) protein complexes have also been found to be conserved over long evolutionary distances.

The striking nature of our observation is highlighted by a further description of the following four examples. First, the majority of *Dv*H RNAP II is purified as an unusual complex containing two copies of both the core enzyme and NusA (particle E shown in Fig. 1; for further details see *SI Appendix*, Figs. S11 and Fig. S31), a particle that has not been seen in other bacteria. Second, *Dv*H pyruvate:ferredoxin oxidoreductase (PFOR), whose structure was determined earlier by Garczarek et al. (13), is an octomer (particle B in Fig. 1; for further details see *SI Appendix*, Figs. S17 and Fig. S28), but in another species of the same genus, *Desulfovibrio africanus*, it is a dimer (14). As we reported in ref. 13, the insertion of a single valine residue into a surface loop of the dimer appears to account for the assembly of the *Dv*H protein into the higher oligomer. Third, although lumazine synthase (also known as riboflavin synthase β subunit) forms an icosahedral complex in *Dv*H as it does in *B. subtilis* (15) and *Aquifex aeolicus* (16), the pentameric subunit is rotated by ≈30° relative to its orientation in the previously reported icosahedral structures, as shown in Fig. 2. As a result of this rotation, the diameter of the *Dv*H icosahedron is increased, and the interaction interface between pentamers is clearly not conserved. Instead, the vertices of the *Dv*H pentamers make head-to-head contact with one another at the icosahedral threefold axis rather than the side-by-side contact between edges of the pentamers that is seen in the previously described structures. Fourth, a *Dv*H homolog of the carbohydrate phosphorylase family is a ring-shaped complex, as is shown in Fig. 3. Although it was not possible to obtain a 3D reconstruction for this particle at a resolution high enough to determine the subunit stoichiometry, its particle weight on size-exclusion chromatography suggests that it is at least a hexamer (Table 1). Because previously described members of the carbohydrate phosphorylase family are either monomers or dimers, these ring-shaped particles represent a unique quaternary structure for this family.

A further, unexpected result that emerged in our survey is the fact that GroEL was initially purified from *Dv*H as a ≈400 kDa complex (*SI Appendix*, Fig. S14), which electron microscopy (data not shown) demonstrated to be a C7-symmetric single-ring heptamer rather than the expected D7-symmetric double-ring 14-mer. However, when Mg^{2+} and ATP are added to the buffer, these single-ring particles assemble efficiently into double-ring complexes (particle D in Fig. 1; for further details see *SI Appendix*, Fig. S30) that appear identical in structure to *E. coli* GroEL. Similar behavior was shown with *Thermobacter brockii* GroEL (17). Although it is thus unlikely that the single-ring form of GroEL exists at any appreciable concentration within cells of *Dv*H, the result serves as a caution that homologous complexes that have conserved subunit stoichiometries and quaternary structures may nevertheless have substantially different stabilities during purification. We note, for example, that GroEL is also purified as a single ring from mitochondria (18) and from a few other bacteria (17, 19–21). A more detailed survey of structure and activity of GroEL orthologs purified from 10 bacteria and 3 mitochondria is presented in *SI Appendix*, Phylogenetic and functional analysis of GroEL quaternary structure.

Discussion

Quaternary Structures of Large Complexes in *Dv*H Are Not Easily Predicted from the Structures of Known Homologs. Within the set of 13 large complexes with identifiable homologs obtained in our

Table 1. Biochemical identity and composition of large macromolecular complexes purified from *Desulfovibrio vulgaris* Hildenborough by the tagless strategy

Gene*	Database annotation	EC number	Molecular mass of polypeptide, kDa	Particle weight estimated by SEC (weight estimated from EM structure when known) (kDa)	Approx. no. of particles per cell	Stoichiometry (symmetry when known)†	Examples of stoichiometry in other bacteria‡
DVU0460	Predicted phospho-2-dehydro-3-deoxyheptonate aldolase	2.5.1.54 or 4.1.2.13	28.4	530	200	α_{16-20}	α_{10}§
DVU0631	Putative protein	—	55.7	600	100	α_{10-14}	—
DVU0671	Putative protein	—	59.1	440 (473)	700	α_8, (*D4*)	—
DVU1012	Hemolysin-type calcium-binding repeat protein	—	316.4	800	1,400	α_{2-3}	—
DVU1044	Inosine-5′-monophosphate dehydrogenase	1.1.1.205	52.2	440 (418)	800	α_8 (*D4*)	α_4
DVU1198	Lumazine synthase (riboflavin synthase β-subunit)	2.5.1.9	16.6	600 (996)¶	300	$\alpha_7\beta_{60}$ (*I*)	$\alpha_3\beta_{60}$ (15); β_5 (29); β_{10} (30)
DVU1200	Riboflavin synthase α-subunit		23.6				
DVU1329	RNA polymerase β-subunit	2.7.7.6	153.2	1,100 (885)	500	$[\beta\beta'\alpha_2\omega$NusA$]_2$ (*C2*)	$\beta\beta'\alpha_2\omega$
DVU2928	RNA polymerase β'-subunit		154.8				
DVU2929	RNA polymerase α-subunit		38.9				
DVU3242	RNA polymerase ω-subunit		8.8				
DVU0510	NusA		47.8				
DVU1378	Ketol-acid reductoisomerase	1.1.1.86	36.1	370	600	α_{8-12}	α_4; α_{12} (31)
DVU1833	Phosphoenolpyruvate synthase	2.7.9.2	132.6	370 (265)	1,200	α_2 (*C2*)	α_2
DVU1834	Pyruvate carboxylase‖	6.4.1.1	136.4	340	800	$[\underline{\alpha\beta}]_2$ or $[\underline{\alpha\beta}]_4$**	$[\underline{\alpha\beta}]_4$ (32); $[\alpha\beta]_4$ (33); $[\alpha\beta]_{12}$ (34)
DVU1976	60 kDa chaperonin (GroEL)	—	58.4	530 (409 and 818)	700††	α_7 and $[\alpha_7]_2$ (*C7* and *D7*)	$[\alpha_7]_2$
DVU2349	Carbohydrate phosphorylase	2.4.1.1	97.4	670 (≥584)	700	α_{6-7} (Ring-shaped)	α_2
DVU2405	Alcohol dehydrogenase	1.1.1.1	41.8	370	12,000	α_{9-10}	α_2
DVU3025	Pyruvate:ferredoxin oxidoreductase ‡‡	1.2.7.1	131.5	1,000 (1,052)	4,000	$[\underline{\alpha\beta\delta\gamma}]_8$, (*D4*)	$[\underline{\alpha\beta\delta\gamma}]_2$; $[\alpha\beta\delta\gamma]$; (35)
DVU3319	Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase	1.5.99.8 and 1.5.1.12	119.0	300	1100	α_3	α_2; α_2 or α_4 (36)

Homologs from other bacteria listed in the rightmost column are members of the same Pfam families (28) as the *D. vulgaris* protein.

*Entries in bold font indicate protein complexes for which three-dimensional reconstructions were obtained by single-particle electron microscopy (EM) of negatively stained samples.

†Stoichiometry is derived from EM data where we have determined the structure. In other cases, the stochiometry is derived from the SEC size estimation.

‡Unless indicated by a specific literature citation, information about subunit stoichiometry was obtained from http://biocyc.org

§*E. coli* also contains three DAHP synthetases (AroF, AroH and AroG) with stoichiometry α_2, α_2 and α_4, respectively. *M. tuberculosis* AroG has stoichiometry $\underline{\alpha_5}$ (29). Although Pfam lists Class I aldolases such as DVU0460 in a different family than DAHP synthetases, they are all classified in the same superfamily (Aldolase) in SCOP (37), based on structural evidence of remote homology.

¶Contribution of the Riboflavin synthase α-subunit to the particle weight is not included.

‖ Pyruvate carboxylase is present in some bacteria as a single polypeptide chain and in other bacteria as α and β chains that are homologous to the C- and N-terminal parts, respectively, of the single-chain form of the enzyme. In cases shown here, the α and β chains from other bacteria comprise the same Pfam domains as the single *Dv*H protein. We use $\underline{\alpha\beta}$ to represent the single-chain form.

**EM result indicates either a dimer or tetramer. Size-exclusion chromatography cannot distinguish between these possibilities.

†† Particle copy number estimated on the assumption that the protein is present in the cell as a D7 14-mer rather than as the C7 heptamer isolated in our standard buffer conditions.

‡‡Homologs of pyruvate: ferredoxin oxidoreductase are sometimes fused and sometimes split into multiple chains. In the case shown here, the α, β, γ, and δ chains from *T. maritima* comprise the same Pfam domains as the single *Dv*H protein. We use $\underline{\alpha\beta\delta\gamma}$ to represent the single-chain form.

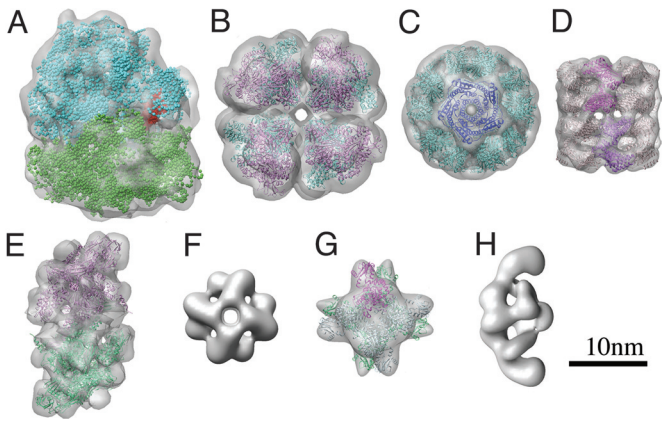


Fig. 1. Gallery of three-dimensional reconstructions obtained by single-particle electron microscopy for eight different, large macromolecular complexes isolated from *Desulfovibrio vulgaris* Hildenborough. Whenever pseudoatomic-resolution models could be created by docking known atomic structures for homologous proteins (or homology models reflecting the *Dv*H sequence), these are shown in color, embedded within the gray isosurfaces for the reconstructed volumes. Homology models were created by using the MODBASE (27) server located at http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi. In the case of the 70s ribosome, however, united-atom representations of X-ray atomic model structures were used. (*A*) 70S ribosome, Mr $\approx$ 3×10^6. The 30S subunit from a X-ray atomic structure (PDB entry 1GIX) is shown in green whereas the 50S subunit (PDB entry 1GIY) is shown in cyan. There is extra density in the EM map at the E site for binding of tRNA, shown in red. (*B*) Octomeric complex, from (13), of pyruvate:ferredoxin oxidoreductase, Mr $\approx 1 \times 10^6$. The top half of the homology model is represented in orchid ribbons and the bottom in turquoise. (*C*) Icosahedral complex of lumazine synthase (beta subunits of riboflavin synthase), Mr $\approx 1 \times 10^6$. One of the pentameric subunits is shown in blue ribbon whereas all others are shown in turquoise. (*D*) GroEL double ring, Mr $\approx$ 800 k. Two ribbon diagrams of the pseudoatomic homology model are shown in purple (bottom ring) and magenta (top ring) respectively, whereas all others are shown in pink. (*E*) Dimer of RNA polymerase, including the transcription elongation factor NusA, Mr $\approx$ 800 k. Two monomers of the heteromeric core enzyme (PDB entry 2PPB) are shown as pink and green ribbons, respectively. (*F*) Homo-octomer of putative protein (DVU0671), Mr $\approx$ 470 k. (*G*) Homo-octomer of inosine-5′-monophosphate dehydrogenase, Mr $\approx$ 416 k. The tetramer at the bottom is shown as light green ribbons and the one at the top as light blue ribbons, with a single monomer shown in magenta. (*H*) Dimer of phosphoenolpyruvate synthase. Mr $\approx$ 265 k. Although an X-ray crystal structure is available for an homologous protein from *Neisseria meningitides* (PDB entry 2OLS), the molecular weight of that protein is only $\approx$2/3 that of the *Dv*H enzyme. As a result, we have not attempted to dock this X-ray structure into the EM map.

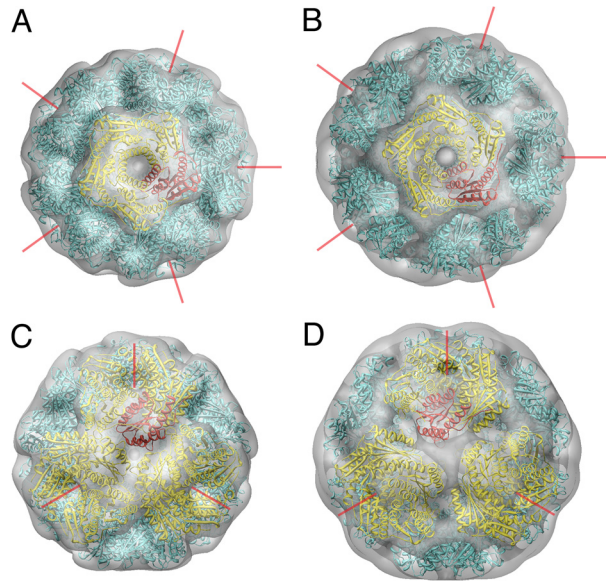


Fig. 2. Comparison of the two types of icosahedral complexes of lumazine synthase (riboflavin synthase beta subunit) formed by the proteins from *Aquifex aeolicus* (*A* and *C*) and from *D. vulgaris* Hildenborough (*B* and *D*), respectively. Note that the vertices of the pentameric subunits of the *Dv*H enzyme meet at the icosahedral threefold axis, thereby resulting in an icosahedral shell with a larger diameter than that produced when the pentamers of the *A. aeolicus* enzyme interact in a more edge-to-edge fashion. The positions and directions of some of the fivefold axes are indicated with red lines to facilitate the comparison of the two structures. (*A* and *C*) Transparent isosurface representations of the X-ray crystal structure of the enzyme from *A. aeolicus*, computed at the same resolution as that estimated for the structure obtained by electron microscopy for the enzyme from *Dv*H, are shown looking down both the fivefold axis (*A*) and down the threefold axis (*C*). A ribbon diagram of the atomic model of the enzyme is shown embedded within the low-resolution isosurface. (*B* and *D*) Transparent isosurface representations of the 3D reconstruction of the *Dv*H enzyme obtained by electron microscopy are shown looking down both the fivefold axis (*B*) and down the threefold axis (*D*). The ribbon diagram of a homology model of the *Dv*H enzyme shown in this panel was rotated by $\approx$30° about the icosahedral fivefold axis to produce a good manual fit within the EM density map. The homology model was created by using the MODBASE (27) server located at http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi .

survey, only GroEL and the 70S ribosome can be modeled with confidence based on the structures of homologous macromolecular complexes. Ten of the remainder have at least one homolog whose known structure is different from that which we have determined for the *Dv*H complex, and the eleventh (phosphoenolpyruvate synthase) has an amino acid sequence that is $\approx$50% longer than that of the homologous protein whose structure is available in the PDB.

A total of 9 of 13 identifiable complexes have sequence identities between the *Dv*H proteins and their homologs that are >30%. Even within this group, only four of the nine complexes have stoichiometries that are the same as those of their most similar homologous complex (Table 1). Although the <50% conservation of structure is less in this case than the $\approx$70% conservation found by Levy et al. (22) (for protein complexes with sequence identity in the range $\approx$30–40%), this difference in the degree of conservation is not statistically significant. Furthermore, in agreement with a generalization reported by Levy et al. (22), that a high percentage of proteins with known quaternary structure are homomeric complexes, all of the complexes purified by our tagless strategy proved either to be, or at least to include, homomeric complexes.

Large Diversity in Subunit Stoichiometry and Quaternary Structure Suggests That These Are Subject to Considerable Selective Pressure. There are many reasons why the quaternary structures of multiprotein complexes might be optimized differently in various bacteria, thereby resulting in versions of the respective enzymes whose performance is best suited to the biochemistry and physiology of a given organism. By extension, environmental changes that produce measurable reprogramming of expression profiles might equally well cause remodeling of either the assembled configurations of certain protein complexes [such as the 100S ribosome-dimers that are produced in *E. coli* under stress (23, 24)] or their spatial organization within cells.

The effect that different degrees of oligomerization can have on the kinetics and regulation of enzymatic activity is no doubt an important factor that contributes to the structural diversity that we have observed for multiprotein complexes. In addition to the well-known effects that the formation of homo-oligomers can have on substrate cooperativity and allosteric regulation, it is likely that the particular way in which multimeric complexes are assembled would affect the productive fraction of diffusion-limited collisions with substrates or with transient partners. We

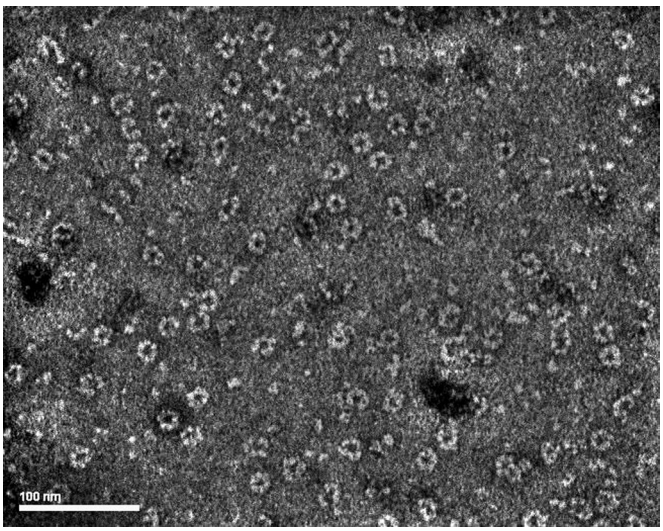


Fig. 3. Electron micrograph of a putative carbohydrate phosphorylase complex isolated from *Dv*H and negatively stained with uranyl acetate. The basic ring-shaped nature of this protein complex is quite apparent, thereby establishing that the subunit stoichiometry and quaternary structure of this complex is completely unlike that of any known member of the carbohydrate phosphorylase family. Beyond that, however, many of the rings are broken open or deformed in other ways. It is difficult to know if this structural heterogeneity just reflects a native flexibility and polymorphism of this protein complex or whether the particles were damaged either during purification or during EM sample preparation. In addition, even a subset of the most circular of the intact rings appears to show heterogeneity in particle diameter. These factors, together with the highly preferred orientation adopted by these particles, have made it impractical to obtain a trustworthy 3D reconstruction of this protein complex.

have also suggested that adjusting the oligomer size (number of monomers) might be used to tune the localized strengths of source-sink relationships in spatially distributed networks of metabolic reactions (13). Other reasons why some enzymes form large, homo-oligomeric complexes could include self-compartmentalization of enzymatic function as a way (*i*) to protect late-stage intermediates from unwanted (off-pathway) reactions, or, more generally, (*ii*) to provide a structural mechanism for channeling of intermediates.

Independent Determinations of Subunit Stoichiometries and Quaternary Structures Are Required. The diversity of subunit stoichiometries and quaternary structures observed in our experiments with *Dv*H is not just relevant to understanding how different bacteria optimize the kinetics and performance of their respective biochemical networks. It is also necessary to use accurate models of the sizes and shapes of multiprotein complexes as templates to determine their spatial locations within cells through the analysis of tomographic reconstructions. Although search templates for some multiprotein complexes, such as the ribosome or GroEL, could be derived from already known structures, it is clear that single-particle electron microscopy or other methods should be used to establish the sizes and shapes of most of the complexes that exist in a new organism of interest. To not do so would be to risk searching for instances of a specific complex and finding none of them, simply because one had been searching with an invalid template. For the specific case of *Dv*H, for example, we have found—as stated above—that templates for only 2 of the 13 largest and most abundant complexes with recognizable function (GroEL and the 70S ribosome, respectively) could have been modeled with high confidence, based on the structures of known macromolecular complexes.

Experimental Procedures

Protein Purification and Identification. Protein complexes were isolated from cells grown as mid-logarithmic cultures in 5-L or 400-L fermentors, which were run as turbidostats. As mentioned above, up to 4 orthogonal separation methods were used to purify multiprotein complexes solely on the basis of differences in their physical properties. The subunit compositions of samples containing purified complexes that ran on native-gel electrophoresis as predominantly a single band with Mr $>$ 400 k were characterized by SDS PAGE, and mass spectroscopy was used to identify the component proteins. Further details about cell growth, the purification of each respective complex, and the identification of proteins by mass spectroscopy are provided in *SI Appendix*.

Electron Microscopy. Aliquots of the purified complexes were examined by single-particle electron microscopy (EM) (25) of negatively stained samples. Uranyl acetate was used as the negative stain in the majority of cases, but ammonium molybdate was tried as a second choice when the results obtained with uranyl acetate were not acceptable. Particles were selected from areas of relatively thick stain to minimize the risk of flattening of particles, and images were recorded on film, using a JEOL 4000 microscope operated at 400 keV. Initial models of particle structures were obtained by the random conical tilt (RCT) method (26) whenever either low-pass filtered density maps of homologous structures (e.g., the 70S ribosome) or intuitive models were not an option. Further details are provided in *SI Appendix*, including representative micrographs, details of the reconstruction and refinement strategies, evaluation of the resolution of reconstructions by means of the FSC curve, and validation of results whenever possible by docking either known structures or homology models.

ACKNOWLEDGMENTS. We thank all members of the Protein Complex Analysis Project (PCAP) at Lawrence Berkeley National Laboratory, whose contributions have been vital to the conduct of this research. This work was supported in part by U.S. Department of Energy Contract DE-AC02-05CH11231 and has been conducted in affiliation with the Virtual Institute for Microbial Stress and Survival.

1. Alberts B (1998) The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* 92:291–294.
2. Robinson CV, Sali A, Baumeister W (2007) The molecular sociology of the cell. *Nature* 450:973–982.
3. Nickell S, Kofler C, Leis AP, Baumeister W (2006) A visual approach to proteomics. *Nat Rev Mol Cell Biol* 7:225–230.
4. Frangakis AS, et al. (2002) Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc Natl Acad Sci USA* 99:14153–14158.
5. Ortiz JO, Forster F, Kurner J, Linaroudis AA, Baumeister W (2006) Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *J Struct Biol* 156:334–341.
6. Sali A, Glaeser R, Earnest T, Baumeister W (2003) From words to literature in structural proteomics. *Nature* 422:216–225.
7. Alber F, et al. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450:683–694.
8. Hazen TC, Tabak HH (2005) Developments in bioremediation of soils and sediments polluted with metals and radionuclides: 2. Field research on bioremediation of metals and radionuclides. *Rev Environ Sci Biotechnol* 4:157–183.
9. Loffler FE, Edwards EA (2006) Harnessing microbial activities for environmental cleanup. *Curr Opin Biotechnol* 17:274–284.
10. Wall JD, Krumholz LR (2006) Uranium reduction. *Annu Rev Microbiol* 60:149–166.
11. Dong M, et al. (2008) A ''Tagless'' strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J Proteome Res* 7:1836–1849.
12. Aloy P, et al. (2004) Structure-based assembly of protein complexes in yeast. *Science* 303:2026–2029.
13. Garczarek F, et al. (2007) Octomeric pyruvate-ferredoxin oxidoreductase from Desulfovibrio vulgaris. *J Struct Biol* 159:9–18.
14. Chabriere E, et al. (2001) Crystal structure of the free radical intermediate of pyruvate: Ferredoxin oxidoreductase. *Science* 294:2559–2563.
15. Ritsert K, et al. (1995) Studies on the lumazine synthase/riboflavin synthase complex of Bacillus subtilis—Crystal-structure analysis of reconstituted, icosahedral beta-subunit capsids with bound substrate-analog inhibitor at 2.4 angstrom resolution. *J Mol Biol* 253:151–167.
16. Zhang XF, Meining W, Fischer M, Bacher A, Ladenstein R (2001) X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 angstrom resolution: Determinants of thermostability revealed from structural comparisons. *J Mol Biol* 306:1099–1114.
17. Truscott KN, Hoj PB, Scopes RK (1994) Purification and Characterization of Chaperonin-60 and Chaperonin-10 from the Anaerobic Thermophile Thermoanaerobacter-Brockii. *Eur J Biochem* 222:277–284.

18. Dubaquie Y, Schatz G, Rospert S (1998) *Methods in Enzymology*, eds Lorimer G, Baldwin T (Elsevier, Amsterdam), Vol 290, pp 193–202.
19. Pannekoek Y, Van Putten JPM, Dankert J (1992) Identification and molecular analysis of a 63-kilodalton stress protein from Neisseria gonorrhoeae. *J Bacteriol* 174:6928–6937.
20. Ishii N, Taguchi H, Sasabe H, Yoshida M (1995) Equatorial split of holo-chaperonin from thermus-thermophilus by Atp and K+. *Febs Letters* 362:121–125.
21. Ferrer M, et al. (2004) Functional consequences of single:double ring transitions in chaperonins: Life in the cold. *Mol Microbiol* 53:167–182.
22. Levy ED, Erba EB, Robinson CV, Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* 453:1262–1266.
23. Ortiz JO, Etchells S, Leis A, Hartl FU, Baumeister W (2007) Ribosomes under stress: Structural variability of the 100S particles studied by cryoelectron tomography. *J Biomol Structure Dynamics* 24:633.
24. Ueta M, et al. (2005) Ribosome binding proteins YhbH and YfiA have opposite functions during 100S formation in the stationary phase of Escherichia coli. *Genes Cells* 10:1103–1112.
25. Frank J (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies—Visualization of Biological Molecules in Their Native State* (Oxford Univ Press, New York) 2nd Ed.
26. Radermacher M, Wagenknecht T, Verschoor A, Frank J (1987) 3-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50s-ribosomal subunit of Escherichia coli. *J Microsc Oxford* 146:113–136.
27. Pieper U, et al. (2006) MODBASE: A database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res* 34:D291–D295.
28. Finn RD, et al. (2006) Pfam: Clans, web tools and services. *Nucleic Acids Res* 34:D247–D251.
29. Morgunova E, et al. (2005) Crystal structure of lumazine synthase from Mycobacterium tuberculosis as a target for rational drug design: Binding mode of a new class of purinetrione inhibitors. *Biochemistry* 44:2746–2758.
30. Klinke S, et al. (2005) Crystallographic studies on decameric Brucella spp. lumazine synthase: A novel quaternary arrangement evolved for a new function? *J Mol Biol* 353:124–137.
31. Ahn HJ, et al. (2003) Crystal structure of class I acetohydroxy acid isomeroreductase from Pseudomonas aeruginosa. *J Mol Biol* 328:505–515.
32. Dunn MF, Araiza G, Finan TM (2001) Cloning and characterization of the pyruvate carboxylase from Sinorhizobium meliloti Rm1021. *Arch Microbiol* 176:355–363.
33. Lai HF, Kraszewski JL, Purwantini E, Mukhopadhyay B (2006) Identification of pyruvate carboxylase genes in Pseudomonas aeruginosa PAO1 and development of a P-aeruginosa-based overexpression system for alpha(4)- and alpha(4)beta(4)-type pyruvate carboxylases. *Appl Environ Microbiol* 72:7785–7792.
34. Aoshima M, Ishii M, Igarashi Y (2004) A novel biotin protein required for reductive carboxylation of 2-oxoglutarate by isocitrate dehydrogenase in Hydrogenobacter thermophilus TK-6. *Mol Microbiol* 51:791–798.
35. Kletzin A, Adams MWW (1996) Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from Pyrococcus furiosus and pyruvate ferredoxin oxidoreductase from Thermotoga maritima. *J Bacteriol* 178:248–257.
36. Meile L, Leisinger T (1982) Purification and Properties of the Bifunctional Proline Dehydrogenase 1-Pyrroline-5-Carboxylate Dehydrogenase from Pseudomonas-Aeruginosa. *Eur J Biochem* 129:67–75.
37. Andreeva A, et al. (2008) Data growth and its impact on the SCOP database: New developments. *Nucleic Acids Res* 36:D419–D425.

Supporting Information for

Survey of Large Protein Complexes in *D. vulgaris* Reveals Great Structural Diversity

Bong-Gyoon Han, Ming Dong, Haichuan Liu, Lauren Camp, Jil Geller, Mary Singer, Terry C. Hazen, Megan Choi, H. Ewa Witkowska, David A. Ball, Dieter Typke, Kenneth H. Downing, Maxim Shatsky, Steven E. Brenner, John-Marc Chandonia, Mark D. Biggin, Robert M. Glaeser*

* To whom correspondence should be addressed. E-mail: rmglaeser@lbl.gov

Methods

1. Electron-microscope tomography of unstained, whole cells

Individual *Dv*H cell specimens for cryo-microscopy were grown in 10 ml of stock LS4D media in 14 ml Falcon tubes or as 100 ml cultures in stoppered glass serum vials. Cultures were grown in an anaerobic chamber at 30 °C until a density greater than ~$3x10^8$ cells/ml (OD 0.4) was reached. Before inoculation the LS4D was reduced with titanium citrate (0.5%) and Thaurer's vitamins were added (0.1%).

For cryo electron microscopy 200 mesh lacey carbon grids (Ted Pella 01881-F) were pre-treated by glow discharge for 20 seconds. The Formvar support was not removed from the lacey carbon. 4 µl of a 1:4 mix of 10 nm nano gold (Ted Pella 15703-20) and deionized H_2O was added to the grid for 3 minutes then blotted on Whatman #1 filter paper. The *Dv*H cells were mixed 4:1 with 10 nm nano gold, 4 µl were instantly placed directly onto the grids, which were manually blotted for 4 seconds and then immediately plunged into liquid ethane by a compressed air piston. All samples were stored in liquid nitrogen.

All images were acquired on a JOEL 3100 series electron microscope equipped with a field emission gun (FEG) operating at 300 kV, an omega energy filter, a cryo-transfer stage cooled with liquid nitrogen to 80K and a Gatan 795 2Kx2K CCD camera. Images were recorded using nominal microscope magnifications of 30K, 25Kx, or 20K giving a pixel size at the CCD of 0.84 nm, 1.0 nm, or 1.2 nm respectively. Underfocus values ranged from ~10 µm to ~12 µm, and energy filter widths were typically around 25 eV.

All tomographic single-tilt series were recorded under low dose conditions, using a maximum dose per complete tilt series of 150 $e^-/Å^2$, with typical values of approximately 100-130 $e^-/Å^2$. Typical angular ranges were between +65 degree and -65 degrees with increments of 1 degree. Tilt series data sets were recorded semi-automatically with the program SerialEM (*1*) (http://bio3d.colorado.edu/), adapted for JEOL microscopes.

Images were binned 2-fold, and tomographic reconstructions, such as the example shown in Figure S1, were computed with the assistance of the eTomo graphical user interface (http://bio3d.colorado.edu/imod/doc/UsingEtomo.html) for the IMOD Tomography package (*2*). The resulting three-dimensional volumes were viewed using IMOD (http://bio3d.colorado.edu/).

2. Cell culture and biomass production

D. vulgaris Hildenborough (DvH) (ATCC 29579) was obtained from the American Type Culture Collection (Manassas, VA). A defined lactate-sulfate medium, LS4D (*3*) is used in all cultures. The medium is sterilized by autoclaving for 45 minutes at 121^{o}C. Before inoculation, phosphate, vitamins and reducing agent (titanium citrate) are added to the medium. Stock cultures of DvH were prepared by growing the ATCC culture to log phase, and storing at -80^{o}C. Starter culture is prepared inside an anaerobic chamber (Coy Laboratory Products, Inc., Grass Lake, MI) using stock culture at a ratio of 1 ml stock/100 ml LS4D. The starter culture is incubated at 30^{o}C and allowed to grow for 48 hrs to log phase (optical density at 600 nm of ~0.3-0.4; ~3x10^{8} cells/ml). From the starter culture, a 10% subculture for inoculating the production culture is made in LS4D, in the anaerobic chamber, and incubated at 30^{o}C until log phase growth is reached (around 15 hours).

The production culture is grown in 5 L customized fermentors (Electrolab, Fermac 360, United Kingdom), run as turbidostats. PEEK headplates and agitators were specially manufactured so that there are no metallic wetted parts. The fermentor is autoclaved with 4.5 L LS4D medium and cooled on the bench under a nitrogen gas blanket. Once cooled, vitamins, phosphate and reducing agent are injected to the fermentor, followed by ten percent subculture (500 mL). The fermentor is continuously agitated at 200 rpm, maintained at 30^{o}C, with nitrogen flowing through the headspace at 100 mL/min.

Once log phase is reached, fresh medium is pumped to the fermentor at a dilution rate of 0.3 1/hr, maintaining an optical density of 0.6 (at 600 nm). The effluent passes through a chilling coil and is collected in a 20 L carboy where the temperature in maintained at 2-4 ^{o}C. Effluent is collected over 12-15 hours, and then centrifuged at 11,000 g for 10 minutes, with refrigeration at 4^{o}C (Beckman Coulter, Avanti J-25). The supernatant is discarded, and the pellets are stored at -80^{o}C until further processing.

3. Purification of protein complexes

Overview

The tagless purification strategy was based on previously described work (*4, 5*). All complexes were purified from cells derived from either a small scale culture of 20 L or a large scale culture of 400 L. Proteins were first bound to and then batch eluted from a Q-Sepharose clean up column to remove many nonprotein impurities. 400 L scale preparations were then fractionated into six parts by ammonium sulfate precipitation. The ammonium sulfate fractions from the large preparation or the cleaned up small scale preparations were then fractionated by MonoQ chromatography. All the fractions from each MonoQ column were analyzed by both native and SDS PAGE to identify abundant protein bands that migrated at approximately 400 kDa or greater (Figure S2). In addition, proteins that did not bind to the Q-Sepharose cleanup column were further fractionated by size exclusion chromatography (SEC) and then analyzed by SDS PAGE (Figure S3). Fractions containing each putative protein complex were pooled and subjected to hydrophobic interaction chromatography (HIC) and/or SEC until sufficiently pure for EM analysis. 15 protein complexes were successfully purified to at least 75% purity as estimated by SDS PAGE (Figures S2 and S3); a further 5 complexes proved either to migrate at less than 300 kDa on an SEC column or to be duplicates of other protein bands and thus were not analyzed by EM. Suitable fractions were buffer exchanged into 10 mM HEPES, pH 7.6, 2 mM DTT, 0.01% NP-40 for EM as described previously (*4, 5*).

Experimental Methods

Extracts were prepared as described previously (*5*). 20 L bacterial cultures yielded crude extracts of 340 mg of protein and 400 L cultures yielded 10 g of protein. Chromatography was done using a AKTA FPLC system. All chromatography columns and media were from GE Healthcare. All separations were performed at 4°C except hydrophobic interaction chromatography (HIC), which was run at room temperature. The concentrations of proteins were monitored by UV light at 280 nm. Mixtures of two buffers were used for ion exchange chromatography (IEC) and HIC. For IEC, buffer A contained 25 mM HEPES pH 7.6, 0 M NaCl, 10% (v/v) glycerol, 2 mM DTT, 0.01% (v/v) NP-40 and buffer B contained buffer A plus 1 M NaCl. For HIC, buffer A' contained 25 mM HEPES pH 7.6, 10% (v/v) glycerol, 2 mM DTT and buffer B' contained buffer A' plus 2 M $(NH_4)_2SO_4$. For SEC, the buffer used contained 25 mM HEPES pH 7.6, 0.05 M NaCl, 10% (v/v) glycerol, 2 mM DTT, 0.01% (v/v) NP-40.

Q-Sepharose clean-up: Protein extract supernatants were loaded onto either a 1.6 x 20 cm (small scale) or 5.0 x 30 cm (large scale) Q-Sepharose Fast Flow column equilibrated with 5% buffer B, and the bound proteins were eluted together with 50% buffer B. All fractions containing significant amounts of protein were pooled. The total protein amount obtained was 240 mg and 7 g for the small and large scale preparations respectively.

Ammonium sulfate precipitation: After the Q-Sepharose clean-up step, the large scale extract was fractionated into 6 parts by ammonium sulfate precipitation: 0-38%, 38-48%, 48-53%, 53-57%, 57-63% and greater than 63% ammonium sulfate saturation. Each cut, which contained between 568 mg to 1028 mg protein, was desalted into 5%

buffer B by buffer exchange using a G25 desalting column (5.0 x 30 cm).

Anion exchange chromatography: The post clean-up step small scale extracts were applied to a 20 ml 1.6 x 10 cm, 20 ml MonoQ column. Each desalted ammonium sulfate precipitation cut from large scale preparations was loaded to a 3.5 x 10 cm, 96 ml MonoQ column. All MonoQ columns were pre-equilibrated with 5% buffer B and developed with a linear gradient from 5% to 50% buffer B in 25 column volumes. For the 20 ml and 96 ml columns, the flow rates were 4 ml/min and 10 ml/min and fraction sizes were 4 ml and 24 ml respectively.

Protein complex survey: To quickly locate high abundance large molecular weight protein complexes, the Mono Q fractions were analyzed by native PAGE (e.g. Figure S2). In addition, those proteins that did not bind the Q-Sepharose column were fractionated by SEC and the resulting fractions also analyzed by native PAGE (e.g. Figure S3). 20 strong protein bands, which migrated at approximately 400 kDa or greater on native PAGE were picked and subjected for further purification. The fractions containing these chosen target complexes were further fractionated by HIC and/or SEC until EM grade purity were reached. Specific details of the HIC and SEC steps are described below for each factor.

Protein complex molecular weight calculation: The molecular weights of purified protein complexes were determined from their migration on a 1.0 x 30 cm Superose6 column or a 1.6 x 60 Superdex200 column in SEC buffer. The molecular weight standards used to calibrate the SEC column were BSA (67 kDa), aldolase (158 kDa), catalase (223 kDa), ferritin (440 kDa), and thyroglobulin (669 kDa).

Protein copy number estimation: The copy numbers of protein complexes per cell listed in Table 1 were estimated from the amount of protein in the flow through of the Q-Sepharose cleanup column and the Mono Q fractions; the estimated yield of total protein present after chromatography; and the number of cells used in the preparation. The amount of each complex in the MonoQ fractions or the Q-Sepharose flow through was estimated from native PAGE by comparing the target protein bands with known amounts of a BSA standard.

Electrophoresis and silver staining: Chromatographic fractions were analyzed by PAGE using Criterion Precast gels (Bio-Rad): 4-15% gradient gels for native PAGE and 4-20% gradient gels for SDS PAGE. Gels were stained using a SilverQuest^{TM} silver staining kit (Invitrogen).

<u>Specific details for each protein complex</u>

DVU0460: Predicted phospho-2-dehydro-3-deoxyheptonate aldolase. Aldolase was purified from the 48 to 53% ammonium sulfate precipitation cut from a 400 L culture preparation. 1.2 mg of protein from a single MonoQ fraction containing aldolase was diluted with an equal volume of buffer B' and loaded to a 5 ml HiTrap Phenyl HP column equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0 % buffer B' in 15 column volumes. The flow rate was 1 ml/min and the fraction size was 2.5 ml. A fraction containing 10 μg/ml of purified protein was picked for EM analysis (Figure S4). The apparent molecular weight of aldolase was estimated as 530 kDa by SEC.

DVU0631: Putative protein. Putative protein DVU0631 was purified from a 20 L culture preparation. 1.6 mg of protein from a single Mono Q fraction containing putative protein DVU0631 was separated using a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. A fraction containing 5 μg/ml of purified protein was picked for EM analysis (Figure S5). The molecular weight of putative protein DVU0631 was estimated as 600 kDa by SEC.

DVU0671: Putative protein. Putative protein DVU0671 was purified from 20 L culture preparation. 8.7 mg of protein from several MonoQ fractions containing putative protien DVU0671 were diluted with an equal volume of buffer B' and loaded to a 0.46 cm x 10 cm (1.7 ml) Source 15PHE 4.6/100 PE column, which was equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0% buffer B in 20 column volumes. The flowrate was 0.5 ml/min and the fraction size was 0.5 ml. HIC fractions containing putative protein DVU0671 were further separated using a 1.6 x 60 cm Superdex200 column. Two fractions containing 25 μg/ml of purified protein were picked for EM analysis (Figure S6). The molecular weight of putative protein DVU0671 was estimated as 440 kDa by SEC.

DVU1012: Hemolysin-type calcium-binding repeat protein. Hemolysin-type calcium-binding repeat protein was was purified from the 0-38% ammonium sulfate precipitation cut from a 400 L culture preparation. 3.4 mg of protein from a single MonoQ fraction containing hemolysin-type calcium-binding repeat protein was diluted with an equal volume of buffer B' and loaded to a 5 ml HiTrap Phenyl HP column equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0 % buffer B' in 15 column volumes. The flow rate was 1 ml/min and the fraction size was 2.5 ml. HIC fractions containing hemolysin-type calcium-binding repeat protein were further separated by a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. A fraction containing 120 μg/ml of purified protein was picked for EM analysis (Figure S7). The molecular weight of hemolysin-type calcium-binding repeat protein was estimated as 800 kDa by SEC.

***DVU1044: Inosine-5`-monophosphate dehydrogenase.*** Inosine-5`-monophosphate dehydrogenase (IMP dehydrogenase) was purified from the 0-38% ammonium sulfate precipitation cut from a 400 L culture preparation. 24 mg of protein from a single MonoQ fraction containing IMP dehydrogenase was diluted with an equal volume of buffer B' and loaded to a 5 ml HiTrap Phenyl HP column equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0 % buffer B' in 15 column volumes. The flow rate was 1 ml/min and the fraction size was 2.5 ml. HIC fractions containing IMP dehydrogenase were further separated by a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. Two fractions containing 4 μg/ml of purified protein were picked for EM analysis (Figure S8). The molecular weight of IMP dehydrogenase was estimated as 440 kDa by SEC.

DVU1198 and DVU1200: Riboflavin synthase. Riboflavin synthase, DVU1198 and DVU1200, was purified from the greater than 63% ammonium sulfate precipitation cut

from a 400 L culture preparation. 15.6 mg of protein from a single MonoQ fraction containing riboflavin synthase was concentrated with an Amicon® Ultra-4 centrifugal filter unit with Ultracel-10 membrane (Millipore) and fractionated using a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. Sizing column fractions containing riboflavin synthase were diluted with an equal volume of buffer B' and loaded to a 2 ml HiTrap Phenyl HP column (two 1 ml HiTrap Phenyl HP column connected in tandem), which was equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0% buffer B in 15 column volumes. The flowrate was 0.5 ml/min and the fraction size was 1 ml. Three fractions containing 4 μg/ml of purified protein were picked for EM analysis (Figure S9). The apparent molecular weight of riboflavin synthase was estimated as 600 kDa by SEC.

DVU1378: Ketol-acid reductoisomerase. Ketol-acid reductoisomerase was purified from a 20 L culture preparation. 3.3 mg of protein from several MonoQ fractions containing ketol-acid reductoisomerase were diluted with an equal volume of buffer B' and loaded to a 2 ml HiTrap Phenyl HP column (two 1 ml HiTrap Phenyl HP column connected in tandem), which was equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0% buffer B in 15 column volumes. The flowrate was 0.5 ml/min and the fraction size was 1 ml. HIC fractions containing ketol-acid reductoisomerase were further separated using a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. Two fractions containing 20 μg/ml of purified protein were picked for EM analysis (Figure S10). The molecular weight of ketol-acid reductoisomerase was estimated as 370 kDa by SEC.

DVU1329, DVU2928, DVU2929, DVU3242: DNA-directed RNA polymerase. DNA-directed RNA polymerase was purified from the 0-38% ammonium sulfate precipitation cut from a 400 L culture preparation. 7.2 mg of protein from a single MonoQ fraction containing DNA-directed RNA polymerase was diluted with an equal volume of buffer B' and loaded to a 5 ml HiTrap Phenyl HP column equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0 % buffer B' in 15 column volumes. The flow rate was 1 ml/min and the fraction size was 2.5 ml. HIC fractions containing DNA-directed RNA polymerase concentrated with an Amicon® filter, using an Ultracel-10 membrane, and were further separated by a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. A fraction containing 150 μg/ml of purified protein was picked for EM analysis (Figure S11). The molecular weight of DNA-directed RNA polymerase was estimated as 1,100 kDa by SEC.

DVU1833: Phosphoenolpyruvate synthase. Phosphoenolpyruvate synthase was purified from a 20 L culture preparation. 7 mg of protein from several MonoQ fractions containing phosphoenolpyruvate synthase were diluted with an equal volume of buffer B' and loaded to a 0.46 cm x 10 cm (1.7 ml) Source 15PHE 4.6/100 PE column, which was equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0% buffer B in 20 column volumes. The flowrate was 0.5 ml/min and the fraction size was 0.5 ml. HIC fractions containing phosphoenolpyruvate synthase were further separated using a 1.6 x

60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. Two fractions containing 40 µg/ml of purified protein were picked for EM analysis (Figure S12). The molecular weight of phosphoenolpyruvate synthase was estimated as 370 kDa by SEC.

DVU1834: Pyruvate carboxylase. Pyruvate carboxylase was purified from the 0-38% ammonium sulfate precipitation cut from a 400 L culture preparation. 16 mg of protein from a single MonoQ fraction containing pyruvate carboxylase was diluted with an equal volume of buffer B' and loaded to a 5 ml HiTrap Phenyl HP column equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0 % buffer B' in 15 column volumes. The flow rate was 1 ml/min and the fraction size was 2.5 ml. HIC fractions containing pyruvate carboxylase were concentrated with an Amicon® filter as described above and further separated by a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. A fraction containing 20 µg/ml of purified protein was picked for EM analysis (Figure S13). The molecular weight of pyruvate carboxylase was estimated as 340 kDa by SEC.

DVU1976: 60 kDa chaperonin. 60 kDa chaperonin (GroEL) was purified from a 20 L culture preparation. 4.5 mg of protein from a single MonoQ fractions containing GroEL was diluted with an equal volume of buffer B' and loaded to a 0.46 cm x 10 cm (1.7 ml) Source 15PHE 4.6/100 PE column, which was equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0% buffer B in 20 column volumes. The flowrate was 0.5 ml/min and the fraction size was 0.5 ml. HIC fractions containing 60 kDa chaperonin were further separated using a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. A fraction containing 75 µg/ml of purified protein was picked for EM analysis (Figure S14). The molecular weight of GroEL was estimated as 530 kDa by SEC, which EM analysis confirmed corresponded to the single-ring, heptamer form of GroEL. Inclubation of this sample with 7 mM $MgCl_2$, 0.5 mM ATP for 10 min leads to formation of the double-ringed tetradecamer form of GroEL that was used to obtain the EM structure shown in Figure 1.

DVU2349: Phosphorylase (glycogen phosphorylase family). Phosphorylase was purified from the 0-38% ammonium sulfate precipitation cut from a 400 L culture preparation. 24 mg of protein from a single MonoQ fraction containing phosphorylase was diluted with an equal volume of buffer B' and loaded to a 5 ml HiTrap Phenyl HP column equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0 % buffer B' in 15 column volumes. The flow rate was 1 ml/min and the fraction size was 2.5 ml. HIC fractions containing phosphorylase were further separated by a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. Two fractions containing 40 µg/ml of purified protein were picked for EM analysis (Figure S15). The molecular weight of phosphorylase was estimated as 670 kDa by SEC.

DVU2405: Alcohol dehydrogenase. Alcohol dehydrogenase was purified from a 400 L culture preparation. 50 mg protein that did not bind to the Q-Sepharose clean up column was diluted with an equal volume of buffer B' and loaded to 5 ml HiTrap Phenyl HP column equilibrated with 50% buffer B', which was equilibrated with 50% buffer B'.

After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0% buffer B in 15 column volumes. The flowrate was 1 ml/min and the fraction size was 2.5 ml. HIC fractions containing alcohol dehydrogenase were further separated using a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. A fraction containing 950 µg/ml of purified protein was picked for EM analysis (Figure S16). The molecular weight of alcohol dehydrogenase was estimated as 400 kDa by SEC.

DVU3025: Pyruvate-ferredoxin oxidoreductase. Pyruvate-ferredoxin oxidoreductase (PFOR) was purified from a 20 L culture preparation. 4.4 mg of protein from a single MonoQ fraction was separated using a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. A fraction containing 45 µg/ml of purified protein was picked for EM analysis (Figure S17). The molecular weight of PFOR was estimated as 1,000 kDa by SEC.

DVU3319: Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase. Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase was purified from a 20L culture preparation. 0.7 mg of protein from several MonoQ fractions containing proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase were diluted with an equal volume of buffer B' and loaded to a 0.46 cm x 10 cm (1.7 ml) Source 15PHE 4.6/100 PE column, which was equilibrated with 50% buffer B'. After washing with 2 column volumes of 50% buffer B', the column was developed with a linear gradient from 50% to 0% buffer B in 20 column volumes. The flow rate was 0.5 ml/min and the fraction size was 0.5 ml. HIC fractions containing proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase were further separated using a 1.6 x 60 cm Superdex200 column. The flow rate was 0.4 ml/min and the fraction size was 2.5 ml. Two fractions containing 40 µg/ml of purified protein were picked for EM analysis (Figure S18). The molecular weight of Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase was estimated as 400 kDa by SEC.

70S Ribosome. 70 S ribosomes were not expected to be purified as part of the tagless survey for purifying protein complexes described above, as it is known that they are unstable under the buffer conditions used for that general-purpose protocol. Instead, a separate purification method was developed in order to isolate ribosomes, based on the protocol established for the *E.coli* 70 S particle (*6*)(*7*). All purification steps were performed at 4°C. Cell pellets from a 5 L *D. vulgaris* culture were resuspended in 40 ml buffer A (20 mM Tris HCl pH 7.5, 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, 5 mM beta-mercaptoethanol, 0.1 mM phenylmethane sulfonyl fluoride, and 0.15 M sucrose) and disrupted using a Beadbeater™ (BioSpec Products, Inc.). Cell debris was removed by centrifugation at 18,000 rpm for 2 hours in a Sorvall SS-34 rotor. The supernatant was overlaid onto a two-layer sucrose gradient. The upper layer contained 0.5 M sucrose in buffer B (20 mM Tris HCl pH 7.5, 500 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, and 5 mM beta-mercaptoethanol). The lower layer contained 0.7 M sucrose in buffer C (20 mM Tris HCl pH 7.5, 60 mM NH_4Cl, 6 mM $MgCl_2$, 0.5 mM EDTA, 5 mM beta-mercaptoethanol). The samples were centrifuged at 28,000 rpm for 15 hours in a Beckman Ti45 rotor. The pellet containing 70 S ribosomes was resuspended in buffer C and separated on a 25%-45% linear sucrose gradient in buffer C by centrifugation at 28,000 rpm for 16 hours in a Beckman SW28 rotor. Gradient fractions were collected

from the bottom of centrifuge tubes using a peristaltic pump, resulting in the elution profile shown in Figure S19. EM analysis of the gradient fractions identified the fractions containing intact 70 S ribosome.

4. Identification of protein components by mass spectroscopy

Reagents used

ACS/HPLC grade acetonitrile (AcCN) and HPLC water were from Honeywell Burdick & Jackson; trifluoroacetic acid (TFA) was from Pierce, Suprapur formic acid was from EMD Biosciences; sequencing grade modified porcine trypsin was from Promega; C18 ziptips and MultiScreen IP 0.45µm Clear Non-sterile plates were from Millipore; guanidine hydrochloride, [tris-(2-carboxyethyl)-phosphine], iodoacetamide, polyvinylpyrrilodone 360 and ammonium bicarbonate were from Sigma.

Protein digestion

In-gel digestion of candidate proteins was performed according to the established protocol (*8*). Modified porcine trypsin from Promega was used at a final concentration of 12.5 ng/□l. In few cases, polypeptide components of protein complexes were not separated on the gel but directly digested with trypsin utilizing a 98-well PVDF plate format that we have adapted from Papac *et al.* (*9*). Briefly, protein was captured onto PVDF membrane of a MultiScreen IP 0.45µm Clear Non-sterile plate, thoroughly washed, reduced and alkylated with iodoacetamide. Membrane was then blocked with polyvinylpyrrolidone 360, trypsin was added and digestion proceeded at 37^{o}C for 4 hr. Mixtures of proteolytic peptides were desalted using C18 ziptips, peptides were eluted with 50% AcCN/0.1% TFA.

Sample preparation for MS

For peptide mass fingerprinting (PMF) (*10-14*) and MS/MS analyses, desalted mixtures of proteolytic peptides were mixed with matrix solution (α-cyano-4-hydroxycinnamic acid 5 mg/ml in 50% ACN/0.1% TFA/10 mM dibasic ammonium phosphate) at a 1:1 ratio directly on a stainless steel target. For MALDI LC MS/MS analysis, samples were separated off-line, as reported previously (*4*), with the modifications outlined below. The Ultimate 3000 HPLC (Dionex Corporation, Sunnyvale, CA, USA) that was custom plumbed to accommodate a dual parallel column arrangement was employed. Tryptic digests were separated on monolithic columns (200 µm I.D., 5 cm length, LC Packings, Dionex Corporation, Sunnyvale, CA, USA) that alternated between a separation and clean up/re-equilibration stage. Following a 5 min isocratic step at 0% B, a linear gradient of 0-70% B in 14 min at a flow rate of 2.5 µl/min was used (A: 0.05% TFA; B: 95% AcCN/0.05% TFA). A SunCollect spotter (SunChrom, Friedrichsdorf, Germany) was used to collect eluate at a rate of one fraction (spot) per five seconds; collection started at 9 min and ended at 19.8 min, counting from the point of injection (129 spots total). Matrix was delivered at a 2.5 µl/min rate and mixed with the column eluate right before spotting onto the MALDI target.

MALDI TOF MS and MS/MS

Applied Biosystems 4800 Proteomics Analyzer (AB 4800) mass spectrometer (Applied Biosystems, Foster City, CA, USA/MDS Sciex, Concord, ON, Canada) equipped with TOF/TOF^{TM} ion optics and a 200 Hz NdYag laser (*15*) and controlled by 4000 Series Explorer Software V3.5.28193 was utilized. MS settings were: m/z range = 800 – 6000 Da; total shots per spectrum = 800 – 1500; single shot protection on (signal

intensity range = 0 – 95000); fixed laser intensity = 3800 – 4500. MS/MS data were generated using collision-induced dissociation (CID). MS/MS settings were: m/z range = [60-(10% below the precursor m/z)]; resolution of precursor ion selector = 400 FWHM; metastable suppressor: on; total shots per spectrum = 1500 – 4000 with stop conditions (1500 shots in maximum collected for spectra containing ≥6 peaks with S/N ≥ 80); fixed laser intensity = 4700 – 5500; the collision cell was floated at 1 kV; no collision gas was used. AB 4800 MS mode was externally calibrated using Plate Model and Default MS Calibration Update software and employed a combination of six peptide standards (des-Arg^1-bradykinin, angiotensin I, Glu^1-fibrinopeptide B and three ACTH clips: 1-17, 18-39 and 7-38) with the requirement of at least four standards passing the criteria of S/N of 300, mass tolerance of 0.5 Da, and maximum outlier error of 25 ppm. Default calibration of AB 4800 MS/MS data was based on minimum five matched fragment ions of angiotensin I detected with a minimum S/N of 120, mass tolerance of 2 Da and maximum outlier error of 20 ppm. Automated acquisition of MS and MS/MS data in the batch mode employed an interpretation method with the following settings: number of shots per spot = 12; minimum S/N filter = 50 – 80; minimum chromatogram peak width = 1 fraction; resolution of precursor exclusion window = 200 FWHM; trypsin autolysis peaks were excluded.

MS and MS/MS data analysis

PMF: Mass spectra were processed (baseline adjustment, noise filtering and monoisotopic peak filtering) using Data Explorer Software (Applied Biosystems, Foster City, CA, USA/MDS Sciex, Concord, ON, Canada) to produce a list of monoisotopic molecular ion masses. Monoisotopic mass peak lists were submitted to the Aldente search engine (*16, 17*) (http://expasy.org/tools/aldente/) for protein identification. A combination of two taxa; *Desulfovibrio vulgaris subsp. vulgaris str.* Hildenborough (*Dv*H) and mammalia (taxon 40674) within UniProtKB/Swiss-Prot (Release 54.8 of 05-Feb-2008) and UniProtKB/TrEMBL (Release 37.8 of 05-Feb-2008) were searched using the following parameters: enzyme trypsin: one missed cleavage; fixed modification on Cys: carbamidomethyl (1 allowable; scoring factor 0.9), variable modification on Met: methionine sulfoxide (2 allowable; scoring factor 0.9); thresholds: shift=0.2, slope=200, error=25, minimum hits=4); mass range: 0-250,000 for all polypeptides but DVU101 for which mass range of 0-350,000 was used. Polypeptide identification was considered to be confident when its score was higher than a threshold value which was equal to a score generated by searching a random database, using pValue of 0.05 as a cutoff point; pValue was the probability of finding, for a given spectrum, a protein with the same score in a random protein database. Identities of selected polypeptides that demonstrated relatively low (DVU0460) or below-threshold scores (DVU3242) were confirmed by MS/MS.

MS/MS data were manually matched to the expected sequences. In accordance with the guidelines for publication of proteomics data (*18*), detailed information on MS-evidence leading to polypeptide identification is provided in Table S1 and Figures S20 to S26, as is indicated below, including PMF data on PMF-only identifications and MS/MS data on identifications based upon single peptides (“one hit wonders”).

LC MALDI MS/MS: Data analysis was performed using ProteinPilot software (Version 2.0, Revision 50861, Applied Biosystems, Foster City, CA, USA/MDS Sciex, Concord, ON, Canada) with Paragon search engine (*19*). The custom database that contained all *DvH* polypeptides and a selection of common contaminants, the latter from Applied Biosystems, was interrogated. The following parameters for ProteinPilot searchwere utilized: Sample Type: protein identification; Cys alkylation: iodoacetamide; ID Focus: biological modifications and amino acid substitutions; Species: none; Search Effort: thorough; Detection Protein Threshold: 1.3 (95%). Hits were considered to be of high confidence if at least one of at least two distinct peptides had a score of 2 (99% confidence). Polypeptides identified on the basis of less stringent criteria are also reported; their diagnostic MS/MS spectra are contained in the figures indicated below.

MS Identification of *DvH* Polypeptides

1. Summary of MS-evidence of polypeptide identification – Table S1.
2. Identification based on PMF only (DVU1833 and DVU3319, Figures S20 and S21, respectively).
3. Identification based on low scoring PMF and MS/MS data (DVU0460 and DVU 3242, Figures S22 and S23, respectively).
4. MS/MS-based identification: low score hit DVU0927 in Figure S24 and “single hit wonders” DVU1314 and DVU0928 in Figure S25 and S26, respectively.

5. Characterization by single-particle electron microscopy

EM grid preparation

Protein samples were initially provided at concentrations of around 0.3 mg/ml and diluted with 10 mM pH 7.5 HEPES buffer. The concentration for each sample was adjusted over the range 0.003 to 0.3 mg/ml to optimize the particle distribution on the EM grid. (Final values used for each of the specimens were: Ribosome, 0.03 mg/ml; Lumazine synthase, 0.3 mg/ml; PFOR, 0.025 mg/ml; GroEL, 0.03 mg/ml; RNA polymerase, 0.03mg/ml; Hypothetical protein Q72EA7, 0.08 mg/ml; IMP dehydrogenase, 0.2 mg/ml; and PEP synthase, 0.015 mg/ml.) For each sample, three microliter of sample was applied to the carbon-coated and glow-discharged EM grid and incubated for 1 minute. The grids were washed several times by touching to drops of 10 mM pH 7.5 HEPES buffer for 10 seconds with the exception of the ribosome sample. In the case of ribosomes, the sample was washed with a pH 7.5 buffer containing 20 mM Tris with 60 mM NH_4Cl, 6 mM $MgCl_2$, 0.5 mM EDTA and 1 mM DTT. The excess solution was removed by a micropipette, leaving about 0.5 microliter on the EM grid to keep the sample from drying. Three microliter of the negative stain of choice was applied to the EM grid for 1 minute. Later the grid was blotted with filter paper and air-dried. Series of EM images were recorded for each grid at three different magnifications, evaluated for homogeneity of stained samples and particle distribution, and uploaded into a data-base for archiving and sharing results of the appearance of the negatively stained samples. In the earlier stages of the project, three different negative stains (5% ammonium molybdate with 1% trehalose, 2% uranyl acetate, and 2% neutralized phosphotungstate, respectively) were tested, and the best stain was chosen on the basis of the sample preservation and homogeneous texture of the stain. After having screened 10 different samples, uranyl acetate was found to give more consistent success in terms of the homogeneity of the sample and good stain distribution, and it has thus been used to prepare EM grids for data collection for most of the samples. Ammonium molybdate stain also produced reasonable EM samples and has been used for the collection of PFOR data (*5*). Phosphotungstate failed to produce EM grids of acceptable quality in the earlier screening stage and was excluded from the standard screening method.

Data collection

EM images were recorded on Kodak SO-163 film at a magnification of 30,000 or 40,000 with under focus values ranging from 1 to 2 μm on a JEOL 4000 microscope operated at 400 kV. The areas with deep stain were scanned to find a suitable place for data collection. Some of the more labile proteins introduced a large amount of background noise due to the formation of protein aggregates or broken and denatured protein particles, distinct from the background noise of homogeneous stain that is seen with more favorable protein particles. When an area was found with homogenous particle sizes and a good particle distribution, in which the particle-to-particle distance was about three times or more than the particle diameter, low-dose images were collected from a new, adjacent area with a final electron dose of ~20 electrons / $Å^2$ on the sample plane. When the random conical tilt (RCT) method was needed, tilt pair images were collected from the same area by recording data from untilted and tilted specimens. Initially, the specimen stage was tilted and the images from tilted view were collected by using low

dose technique at 45 or 60 degrees and later the specimen stage was tilted back for the collection of images of untilted specimens from the same area.

Data analysis

The recorded images were digitized by a film-scanning robot equipped with the Nikon Super Coolscan 8000 ED densitometer (*20*). The images were scanned with a resolution of 6.35 μm per pixel and later averaged 2 fold in each direction, resulting in a resolution of 4.23 or 3.18 Å/pixel at the sample level.

When data were collected from untilted specimens, particles were selected by using the program BOXER in EMAN (*21*) and the coordinates were later used to window out particles from the original micrographs by the program package SPIDER (*22*). When tilt image pairs were collected for the RCT method, corresponding image pairs were displayed side by side and particle pairs were picked by using the program xmipp_mark (*23*). The particle-coordinate data from tilted and untilted images were then used to window out particles from the original micrographs by the program SPIDER. When the expected resolution from the EM images was close or beyond the first zero of the CTF, the defocus value of each micrograph was determined by ctffind3 (*24*), and phase-flipping was applied for each particle by using the SPIDER software package. All the particles were aligned by using a reference free alignment SPIDER command AP SR before classification. Classification was performed for aligned particles and class averages were calculated by using the IMAGIC (*25*) command MSA.

Initial starting model structures of PEP synthase (Fig S32) and a hypothetical protein (Fig S33) were obtained by RCT method by using the SPIDER software package. The Euler angles were assigned for all the particles in each class in SPIDER, and starting models were built by the SPIDER command bp 32f.

In other cases, we could start with intuitive models based on prior knowledge and/or on the appearance of class-average images. These models were built by using a SPIDER command mo 3 with simple geometric volumes consisting mainly of multiple spheres in corresponding scales to match class average views. These initial models are in all cases shown along with the EM images, FSC curves, and the final refinement results, in Figures S27 to S34. Once an initial model structure was available through RCT method or intuitive building technique, the orientations and translational alignments of the particle images were refined iteratively against the model by projection matching in SPIDER. The progress of the iterative refinement was initially monitored by the improvement of the FSC curve, which usually changed little after around 10 cycles. When the FSC curve began to rise at higher resolution, indicating artificial fitting of the background noise in the updated model structure, the refinement was stopped.

When atomic-model structures were available for proteins with high sequence identity to those in the DvH protein complexes, homology models were built using the utility provided by MODBASE (*26*). The docking of atomic models into EM densities and display were done by using CHIMERA (*27*).

Use of the nominal magnification of the electron microscope resulted in 3-D reconstructions whose sizes were a reasonable approximation to those of the corresponding atomic models, when available. The size of the EM reconstructions was nevertheless adjusted slightly for about half of the reconstructions in order to optimize the docking. The need to optimize the size of the EM map, which generally required a

change of less than 5% (except for lumazine synthase, which was adjusted by 10%), may arise from a number of different factors such as flatting of some particles during negative staining, positive staining of the surfaces of some particles, and uncertainty of the density value to use when selecting the value of the isosurface for docking.

Further comments about individual complexes

The Lumazine synthase (riboflavin synthase β subunit) EM density map showed extra density inside the icosahedral cage. The riboflavin synthase α subunit has been known to comigrate during the purification in the previous work of *Bacillus subtilis* (*28*). While mass spectroscopy did show the presence of the α subunit in our purified lumazine synthase sample, the stoichiometry was, as expected, too low to measure quantitatively.

GroEL sample needed the presence of Mg-ATP to form the conventional double rings. When purified by chromatography without the presence of Mg-ATP, the double ring forms completely dissociated into the single ring forms. The purified single ring forms could be converted reversibly into the double ring forms by the addition of Mg-ATP, as described in the section above on purification of protein complexes.

RNA polymerase samples were purified in two biochemical states with and without a transcription regulating factor NusA. Only the RNA polymerase sample with NusA produced homogenous samples good enough for electron microscopy. In addiiton, one of the RNA polymerase samples, in which the NusA was stripped off with the ion exchange column, produced quite inhomogeneous EM samples when prepared with negative stain.

The stoichiometry of the PEP synthase complex was not clear from the shape of the EM density map. Both the native gel electrophoresis result and the volume of the EM density favored a tetrameric rather than dimeric structure, however. As a result, D2 symmetry was imposed during refinement.

Phylogenetic and functional analysis of GroEL quaternary structure

GroEL belongs to a family of molecular chaperones that are required for the proper folding of various proteins. GroEL creates a chamber for a not yet properly folded substrate while its co-chaperonin GroES serves as a lid to that chamber. Textbooks describe the canonical structure of bacterial GroEL as a homo-tetradecameric structure composed of two homo-heptameric rings (*29*). These two rings are bound together back-to-back, allowing each ring to create its own chamber capped by homo-heptameric GroES, which can bind to either or both of the GroEL rings. The double-ring structure is thought to be essential to the mechanism of GroEL action (*30*). However, this structure is not universally shared amongst all chaperones homologues to GroEL. For example, *Cricetulus griseus* Hsp60, a mammalian mitochondrial ortholog of GroEL, has been purified and shown to function as a single heptameric ring (*31*).

Although GroEL and Hsp60 have been extensively studied for the last two decades, uncertainty still exists as to whether their structure, and consequently their mechanism of function, differs between species. The canonical structure and function of the $GroEL_{14}$-$GroES_7$ complex is largely based on the extensive studies of *Escherichia coli* GroEL. However, several lines of evidence suggest that GroEL may exist and function in alternative forms, such as a single ring. In some cases the observed quaternary structure of GroEL delicately depends on experimental conditions, which makes it hard to draw conclusions about the functional structure of GroEL *in vivo* from structural studies to date. Therefore, two major questions remain unsolved: 1) are there native single-ring GroEL isoforms that do not create double rings *in vivo*, and, if the answer is yes, 2) what is the functional difference between single and double ring forms? GroEL properties from several species are summarized in Table S2, and detailed discussion of some of the structural and functional properties is given below.

GroEL/Hsp60 quaternary structure depends on purification conditions

Cricetulus griseus mitochondrial chaperonin Hsp60 was first purified as a single heptameric ring (*31*). However, it was later observed that, in the presence of physiological concentrations of ATP and Hsp10, Hsp60 forms a mixture of single-ring and double-ring structures (*32*). The majority (70-90%) of structures forms a double-ring "football" shape, with a Hsp60 tetradecamer core and a Hsp10 heptamer bound to each side. In the same set of experiments, in the presence of ATP, but in the absence of Hsp10, ~90% of Hsp60 was found as the single heptameric ring, while the rest (~10%) formed a double-ring structure. In the absence of ATP and Hsp10, about 80% of Hsp60 formed single-ring heptamers, while the rest were monomers. In addition, Hsp60 readily breaks down into monomers at low temperature (0°C) in the presence of ATP (*33*).

In at least in three bacterial species, *Thermoanaerobacter brockii*, *Thermus thermophilus*, and *Desulfovibrio vulgaris*, GroEL quaternary structure changed with an addition of

cofactors or ions such as ATP, Mg^{++}, K^{+}, and GroES. GroEL from *Thermoanaerobacter brockii* was initially purified as a single heptameric ring (*34*). However, it was later purified as a double ring with addition of 600 nM $GroES_7$, 2mM ATP, and between 5 and 30 mM of Mg^{++} (*35*). In the double-ring form, $GroES_7$ is bound to one side of GroEL(*35*). In the case of *Thermus thermophilus*, GroES and GroEL were initially purified as a large heteromeric "football" shaped complex, $GroEL_{14}$-$GroES_{14}$ (*36*). Interestingly, when later purified in near-physiological conditions (with 90 mM K^{+} and 1.5 mM Mg-ATP), the $GroEL_{14}$-$GroES_{14}$ complex breaks apart and forms single-ring $GroEL_7$-$GroES_7$ complexes (*37*). We purified *E. coli* GroEL as a double-ring structure using the buffer indicated in Table S2. However, when we purified *Desulfovibrio vulgaris* GroEL in the same buffer, it appeared as a single ring. Yet, when we better approximated physiological conditions by adding 0.5 m*M* ATP and 7 m*M* Mg^{++} to the buffer, the *D. vulgaris* GroEL appeared as a double ring. From these studies, it is apparent that quaternary structure of GroEL and its interaction with GroES is profoundly dependent upon purification conditions.

Double-ring structure is essential for E. coli GroEL, while mitochondrial Hsp60 is able to function as a single-ring structure

E. coli GroEL is composed of two equivalent heptameric rings (*30*). In order to ascertain whether the *E. coli* GroEL double-ring structure is essential for proper function of chaperonin-mediated folding (*38*), a mutant, $GroEL_{SR1}$, was created. $GroEL_{SR1}$ has four changed amino acids at the ring-to-ring interface, which prevent it from forming a double-ring structure. $GroEL_{SR1}$ was found to be unable to release a bound GroES heptamer, thus trapping substrate inside its chamber (*39*). This effect is consistent with the prevailing GroEL model, which argues that the second ring is required for ATP binding in order to trigger the release of GroES (*40*).

Mitochondrial chaperonin Hsp60 from Chinese hamster (*Cricetulus griseus*) and a testis-specific mitochondrial isoform of Hsp60 from moth (*Heliothis virescens*) were purified as a single heptameric ring (*31, 41*). However, in contrast to *E. coli* GroEL, it was shown that *Cricetulus griseus* mitochondrial Hsp60 is able to function as a single ring (*42-44*). Therefore the functional mechanism of *E. coli* GroEL and mitochondrial Hsp60 appear different. It is unknown whether Hsp60 is able to function as a double ring as well.

Single-ring GroEL from other species can functionally substitute for double-ring GroEL in E. coli

A heterologously expressed functional single-ring GroEL can functionally substitute for wild type GroEL in *E. coli*. It thus appears that the native double-ring form is not strictly required for chaperon-mediated folding in *E. coli*. It has been shown that either mammalian mitochondrial single-ring Hsp60 together with Hsp10 (the ortholog of GroES), or other single-ring *E. coli* GroEL mutants (where mutations in addition to those made to $GroEL_{SR1}$ restore the chaperonin properties in $GroEL_{SR1}$), are able to functionally substitute for wild type GroEL-GroES in *E. coli (42-44)*. In another experiment, GroEL/GroES from *Oleispira antarctica* RB8 (*45*) (discussed below)

enhanced the viability (141-fold faster growth) of *E. coli* K-12 at low temperature of 8°C (*46*). At temperatures below 10°C, *Oleispira antarctica* RB8 GroEL has a single-ring structure. The *E. coli* transgenic strain (with GroEL/GroES from *Oleispira antarctica* RB8) experienced a growth even at temperatures below 4°C, whereas the wild type *E. coli* does not grow below 8°C.

Wild-type transition between single and double rings as an adaptation to changing environment

Probably the most interesting studied example of GroEL quaternary structure and function is from the psychrophilic bacterium, *Oleispira antarctica* RB8 (*45*), where it has been shown that GroEL is functional as either a single or a double ring depending on temperature. At physiological temperatures of 4–10°C, GroEL is predominantly a single ring, while when stressed with higher temperature, >10°C, GroEL forms a double-ring complex. At 4°C, in the presence of GroES and in the absence of denatured substrates, GroEL ATPase activity was completely inhibited. Therefore, the authors suggest that at this temperature, the organism reduces energy consumption by switching GroEL to the more efficient single-ring form, which does not use ATP when chaperonin activity is not required. To answer the question of why the double ring structure is observed at higher temperature if the single ring GroEL is fully functional under physiological conditions, the authors created a single ring GroEL mutant, O.GroEL$_{SR}$. The mutations to the wild type GroEL from *Oleispira antarctica* RB8 were introduced at the same ring-to-ring contacts as in the single-ring *E. coli* GroEL$_{SR1}$ mutant. While at 4–10°C O.GroEL$_{SR}$ substitutes for wild type GroEL, at higher temperature the mutant loses its refolding function. The authors suggest that the double-ring mechanism is required to release GroES at higher temperatures. In this study, all structural states observed in *in vitro* experiments were also confirmed by analysis of GroEL from cell-free extracts.

Phylogenetic analysis of GroEL/Hsp60 proteins

In our attempt to understand the evolutionary history of single/double-ring GroEL forms, we performed two studies. One is based on the analysis of specific residues at the GroEL-GroES interface hypothesized to be responsible for differences in binding affinity of single and double-ring structures. In the second study, we reconstructed a phylogenetic tree (Figure S35) of the GroEL/Hsp60 proteins from the species presented in Table S2.

There is a principal difference in cofactor binding specificity of *E.coli* GroEL and mammalian mitochondrial Hsp60. *E. coli* double-ring GroEL is functional in combination with either *E. coli* GroES or mammalian mitochondrial Hsp10 (the ortholog of GroES) (*31*). However, the single-ring mammalian mitochondrial Hsp60 only functions in combination with mammalian Hsp10 (*47*) (although the Hsp60 and Hsp10 do not have to be from the same species (*31*)). According to Richardson and colleagues (*47*), a mobile loop of Hsp10 is responsible for Hsp60's specificity for Hsp10. After mutating three residues in the bacterial GroES loop to match the sequence of Hsp10, the mutant GroES acquires the ability to function with Hsp60, substituting for native Hsp10. This suggests a correlation between the mobile loop sequence and the single/double ring property of

GroEL. We used this tripeptide sequence motif to classify GroES sequences from the Pfam Cpn10 family (*48*). First, we observed that the tripeptide motifs of *E. coli* GroES and mammalian Hsp10 are specific to *E. coli* and mammals, respectively. Second, we noticed that the motif from *Desulfovibrio vulgaris* appears in many species from various lineages, including Archaea (for details see Table S3). This suggests that the *Desulfovibrio vulgaris* motif is likely to be ancestral to more species than the *E. coli* motif. Therefore, to the extent that these three residues are significant, GroEL from *Desulfovibrio vulgaris* may serve as a basis for further studies of GroEL mechanism in other species.

We hypothesized that a phylogenetic analysis would reveal the history and distribution of distinct quaternary structures for GroEL. We reconstructed a phylogenetic tree of the thirteen GroEL orthologs from Table S2 (Figure S35). We labeled each protein with its quaternary structure and GroES/Hsp10 binding loop tripeptide signature. To our surprise, there is no evidence that either the single-ring or double-ring form is ancestral to the other, nor is there evidence that either form is dominant within any major clade. The tripeptide signature did not clearly correlate with known GroEL's quaternary structure as had been proposed by Richardson et al. (*47*). Thus, it is not possible to use GroES binding loop sequence as a predictor for GroEL quaternary state. Furthermore, the data do not reveal a clear evolutionary history of GroEL's quaternary structure with a single transition between quaternary structures. Rather, the quaternary structure seems evolutionary labile; however, whether this is a consequence of purification protocols or reflects the underlying biological activity remains to be determined.

Methods

Phylogenetic analysis

Thirteen GroEL and Hsp60 proteins from the species of Table S2 were used for a phylogenetic tree inference. These are taken from UniProtKB/Swiss-Prot Release 56.2 (*49*): CH60_BUCAI (P25750), CH60_CHRVI (P31293), CH60_DESVH (Q72AL6), CH60_ECOLI (P0A6F5), CH60_NEIGO (P29842), CH60_OLEAN (Q8KM30), CH60_PARDE (Q9Z462), CH601_RHOSH (P20110), CH60_THEBR (Q60024), CH60_THET2 (P61490), CH63_HELVI (P25420), CH60_CRIGR (P18687), and HSP60_YEAST (P19882).

A multiple sequence alignment was obtained from Pfam (version 22.0) (*48*), where GroEL and Hsp60 belong to the Cpn60_TCP1 family. We removed columns with more than 90% gaps. Phylogenetic tree inference was done using PhyML (*50*) and QuickTree (*51*) with default parameters. PhyML applies a maximum likelihood approach and QuickTree applies a variant of the Neighbor-Joining method. Bootstrap analysis (1000 bootstraps) was used to estimate robustness of the phylogenetic trees. The clades within reconstructed trees with PhyML and QuickTree programs are identical besides the placement of GroEL ortholog from Archaea relative to *Thermoanaerobacter brockii* and *Thermus thermophilus*.

Analysis of GroES mobile loop

The GroES mobile loop that binds to GroEL (*47*), comprised of eight amino acids, was used to analyze the reconstructed phylogenetic tree (Figure S35) and the Pfam species tree (Table S3). We used three loop sequences taken from *E. coli*, *Mus musculus*, and *D. vulgaris*: "SAGGIVLT" from CH10_ECOLI (P0A6F9), "TKGGIMLP" from CH10_MOUSE (Q64433) and "TAGGLYIP" from CH10_DESVH (Q72AL5). The tripeptide motifs "SxxxxVxT" from *E. coli*, "TxxxxMxP" from *Mus musculus* and "TxxxxYxP" from *D. vulgaris* were used to identify all proteins in the Cpn10 Pfam family that have an exact occurrence of one of these motifs in the corresponding loop region. The proteins identified were analyzed in the context of the Pfam species tree.

Figures S1 to S37

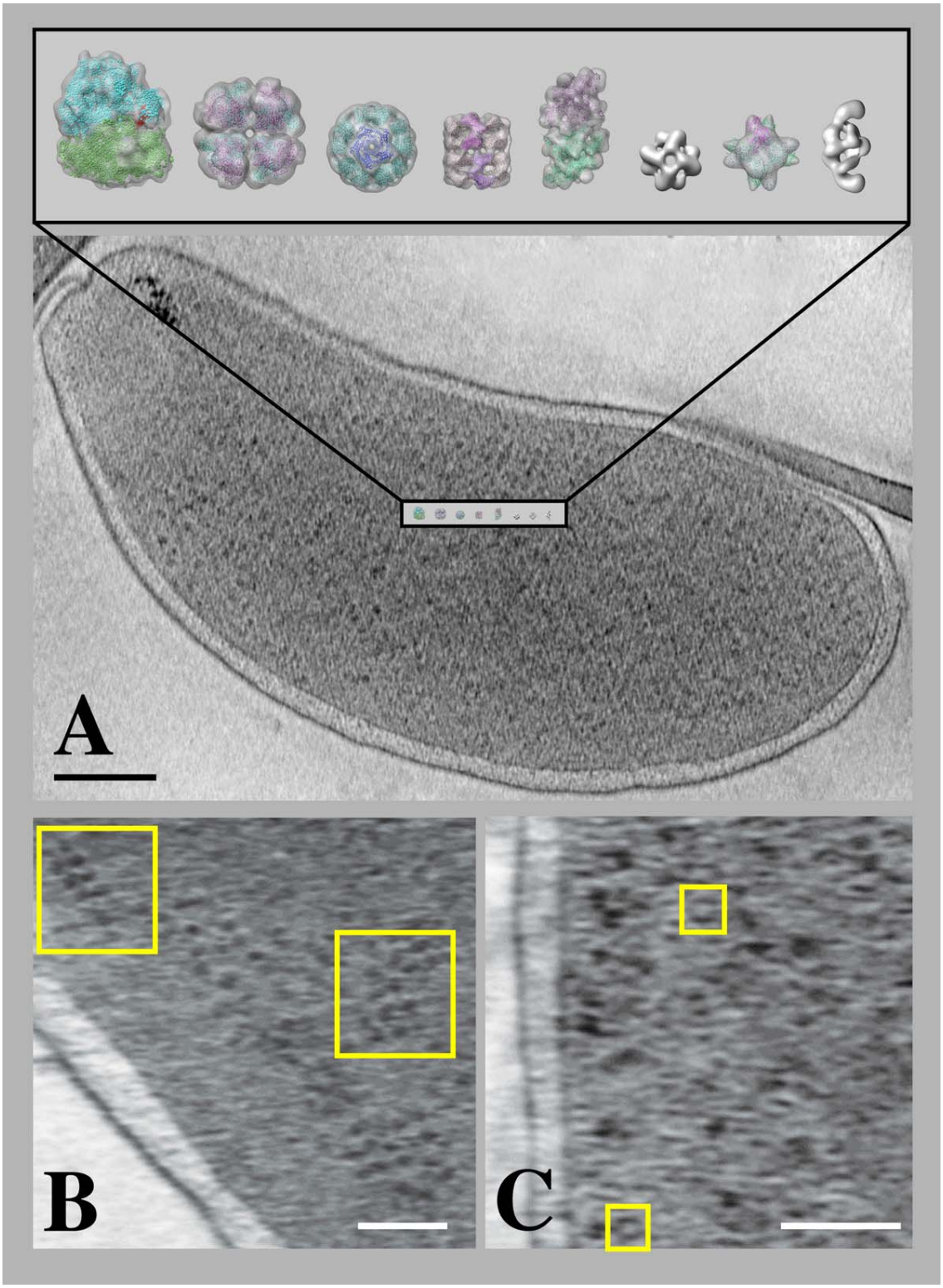

A
B
C

Fig. S1. Illustration of the variety of sizes and shapes of different types of multi-protein complexes purified from *Desulfovibrio vulgaris* Hildenborough. Three-dimensional density maps are shown for eight complexes with Mr >400 k, shown first on a scale large enough to make it possible to appreciate the differences in their sizes and shapes, and then shown again at the same scale as the image of a whole *Dv*H cell. (A) The whole-cell image is a 16.8 nm thick section taken from a region close to the center of a cryo-EM tomogram, selected so as to include the base of the flagellum, top left. The scale bar represents 200 nm. The small size of many of the protein complexes relative to the size of the cell emphasizes the magnitude of the template-matching task involved in searching the volume for instances of each of these particles. (B) Zoomed-in portion of a 16.8 nm section, in which the scale bar represents 100 nm. Yellow boxes highlight clusters of putative ribosomes within the cytoplasm. The size and shape of these clusters indicates a higher level of organization than that expected simply from polysomes. (C) Another zoomed-in portion of a 16.8 nm section, in which the scale bar again represents 100 nm. In this case yellow boxes highlight circular, ring-shaped densities of unknown composition. The diameter of these rings is larger than that of lumazine synthase, for example, a hollow shell that would otherwise appear to be a ring in a section of a tomogram.

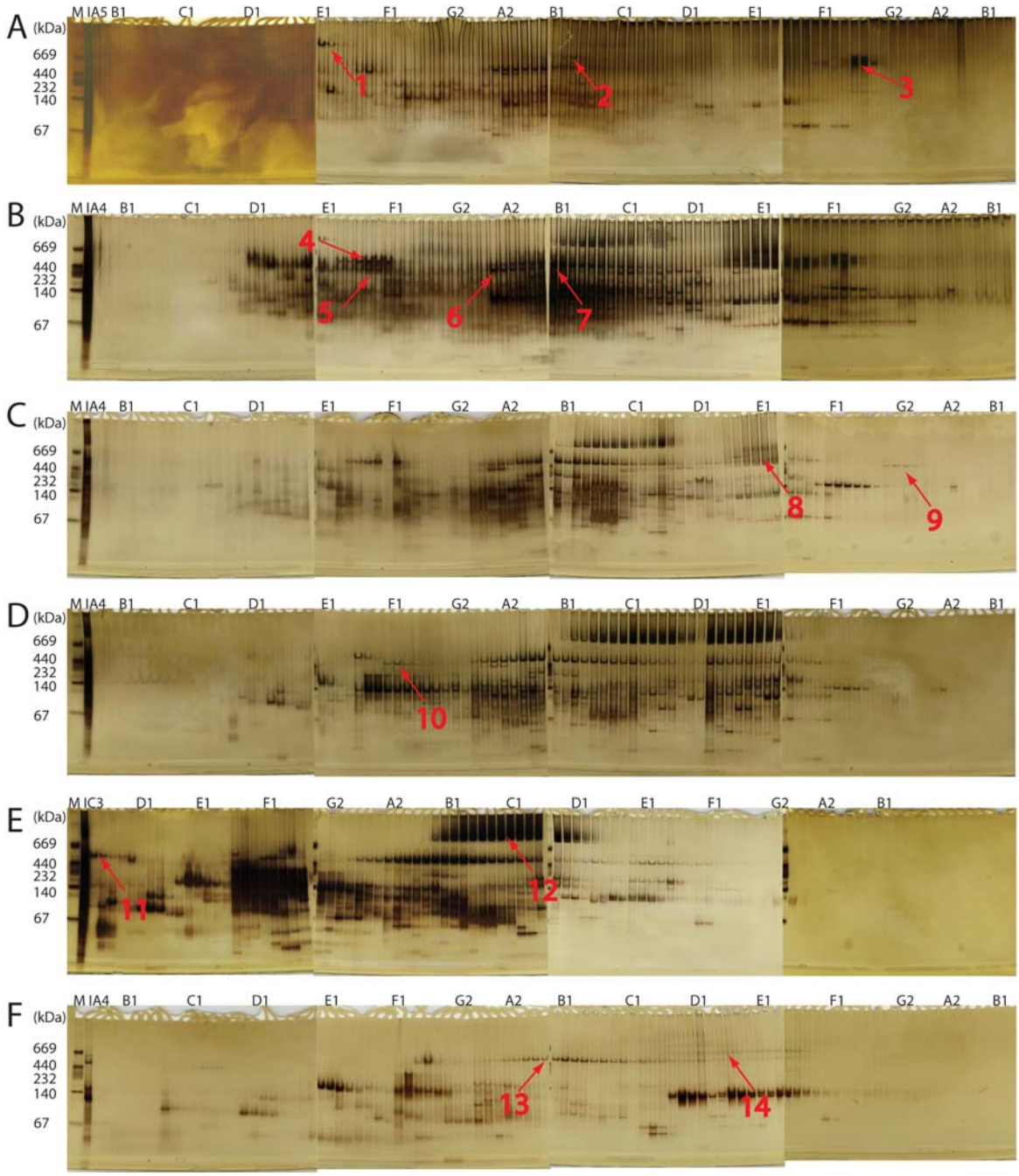


Figure S2. Tagless survey of large *D. vulgaris* protein complexes that bind to Q-Sepharose resin from a 400 L culture preparation. MonoQ column fractions from 6 ammonium sulfate precipitation cuts were each analyzed by native PAGE (4-15% acrylamide): ammonium sulfate saturations of A. 0-38%; B. 38-48%; C. 48-53%; D. 53-57%; E. 57-63%; F. greater than 63%. Arrows show the 14 protein complexes that were sufficiently purified for EM analysis after further fractionation: 1. Putative protein (DVU0631); 2. Phosphorylase (DVU2349); 3. Hemolysin-type calcium-binding repeat protein (DVU1012); 4. Phosphoenolpyruvate synthase (DVU1833); 5. Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase (DVU3319); 6. Pyruvate carboxylase (DVU1834); 7. Inosine-5`-monophosphate dehydrogenase (DVU1044); 8. RNA polymerase (DVU1329, DVU2928, DVU2929, DVU3242); 9. Predicted phospho-2-dehydro-3-deoxyheptonate aldolase (DVU0460); 10. Putative protein (DVU0671); 11. Ketol-acid reductoisomerase (DVU1378); 12. Pyruvate-ferredoxin oxidoreductase (DVU3025); 13. 60 kDa chaperonin (GroEL, DVU1976); 14. Riboflavin synthase (DVU1198, DVU1200).

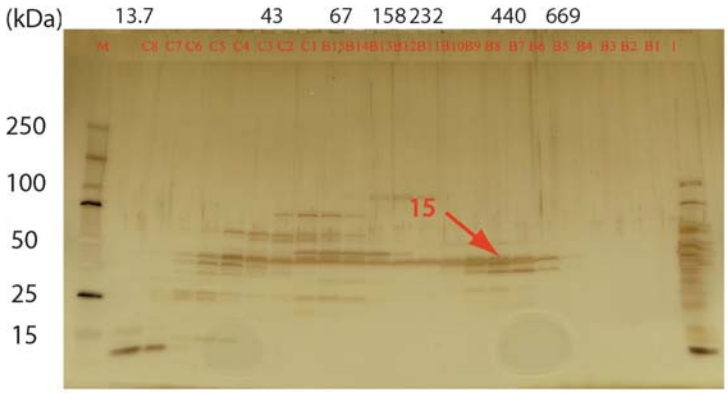


Figure S3. Tagless survey of large *D. vulgaris* protein from a 400 L culture preparation that did not bind to Q-Sepharose resin. These proteins were analyzed by SEC and the fractions then separated by SDS PAGE. The arrow shows alcohol dehydrogenase, which was sufficiently purified for EM analysis by further fractionation. Size markers for the SDS PAGE are shown at the left of the gel and the positions of size makers on the SEC column are shown at the top of the gel.

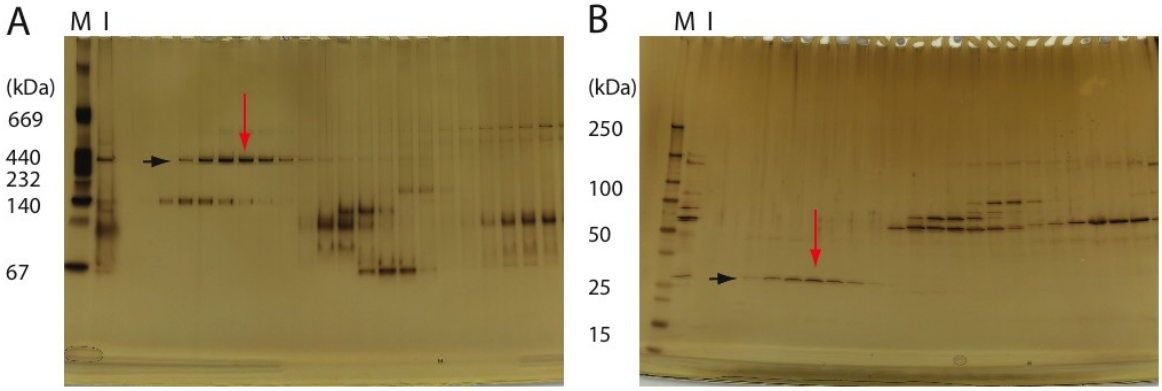


Figure S4. Purification of aldolase. Fractions of the HIC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the aldolase protein. Red arrows show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the HIC column is shown in lane I.

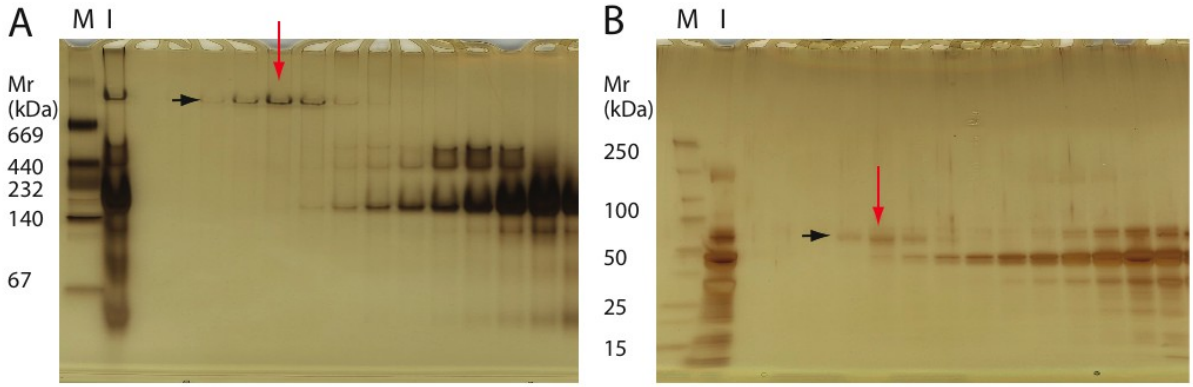


Figure S5. Purification of putative protein (DVU0631). Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the putative protein. Red arrows show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

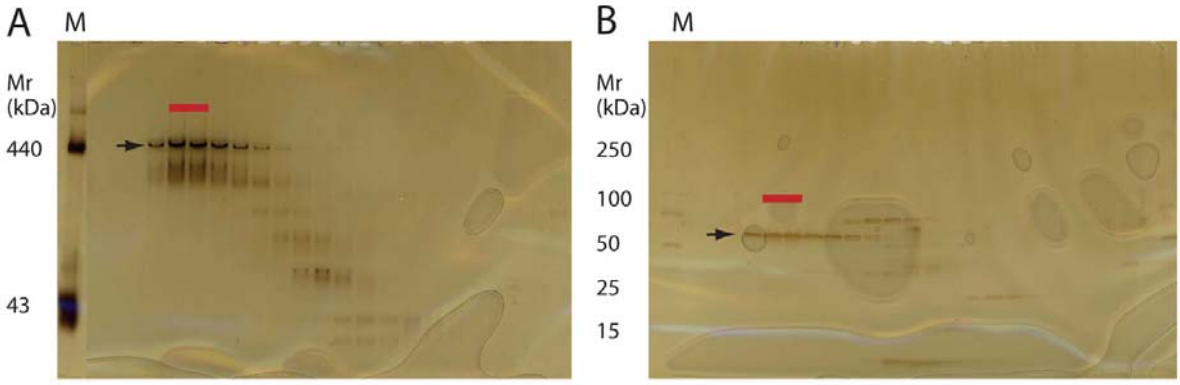


Figure S6. Purification of putative protein (DVU0671). Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the putative protein. Red bars show the fraction used for EM analysis. Molecular weight standards are shown in lane M.

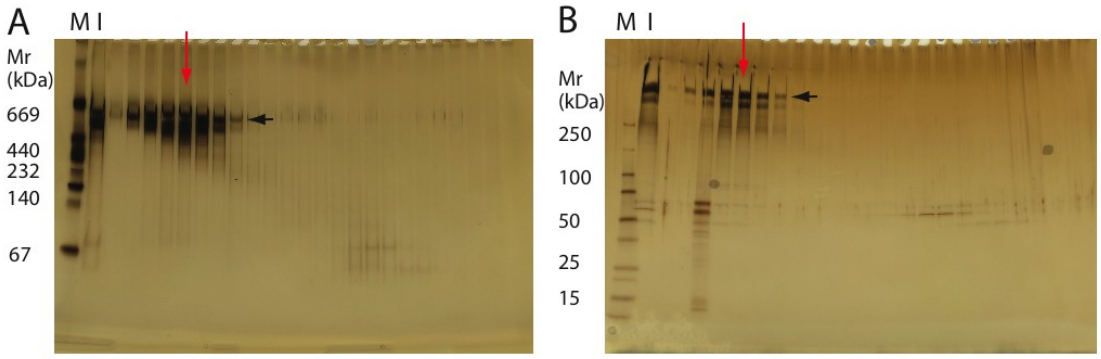


Figure S7. Purification of hemolysin-type calcium-binding repeat protein. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the putative protein. Red arrows show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

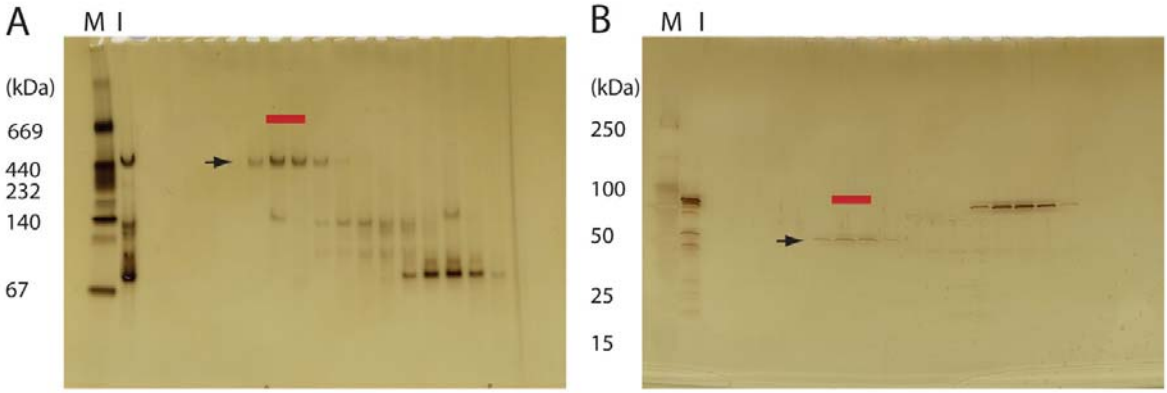


Figure S8. Purification of IMP dehydrogenase. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the putative protein. Red bars show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

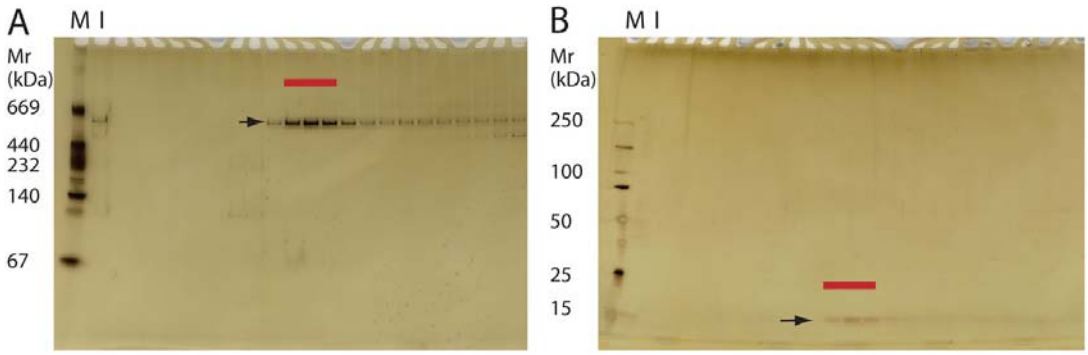


Figure S9. Purification of riboflavin synthase. Fractions from the HIC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrow in A shows the riboflavin synthase and in B shows riboflavin synthase β subunit (DVU1198). α subunit (DVU1200) couldn't be seen on PAGE and was identified by MS/MS. Red bars show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the HIC column is shown in lane I.

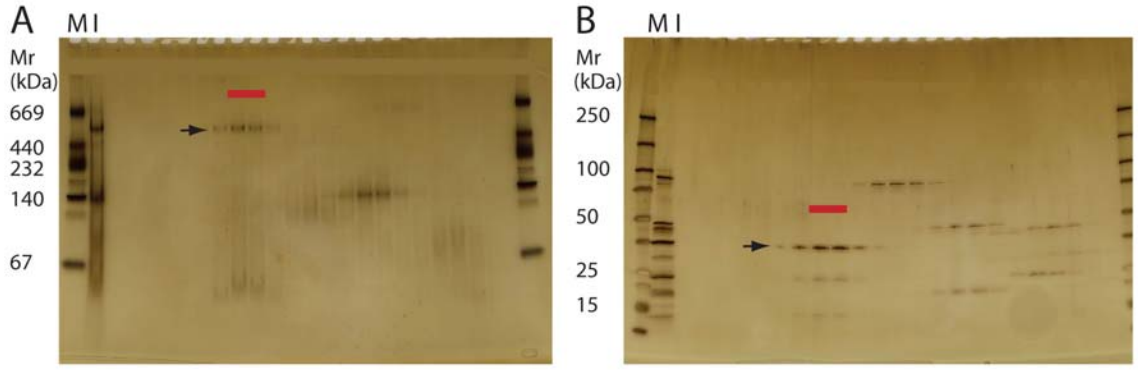


Figure S10. Purification of ketol-acid reductoisomerase. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the ketol-acid reductoisomerase. Red bars show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

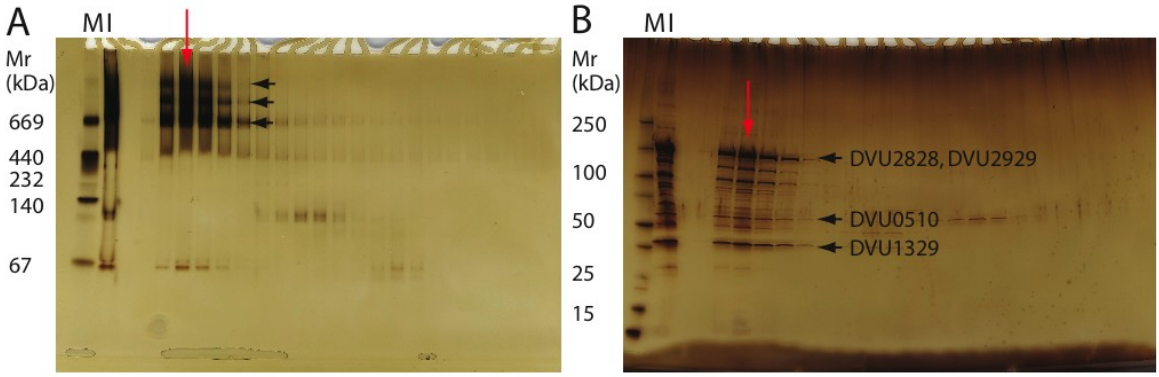


Figure S11. Purification of DNA-directed RNA polymerase. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows in A show DNA-directed RNA polymerase and in B show the subunits of DNA-directed RNA polymerase. β: DVU2928; β': DVU2929; α: DVU1329. ω subunit was detected by MS but can't be seen in the gel. σ was not detected by MS. NusA (DVU0510) was co-purified with DNA-directed RNA polymerase. Red arrows show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

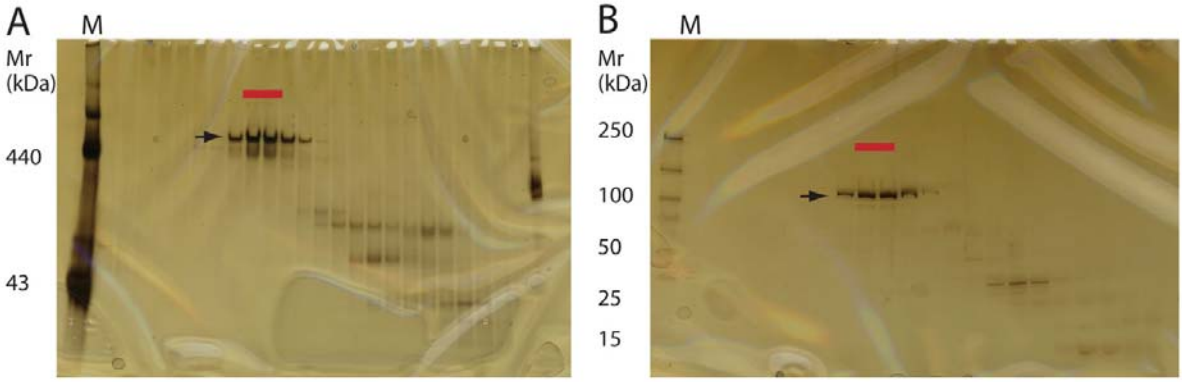


Figure S12. Purification of phosphoenolpyruvate synthase. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the phosphoenolpyruvate synthase. Red bars show the fraction used for EM analysis. Molecular weight standards are shown in lane M.

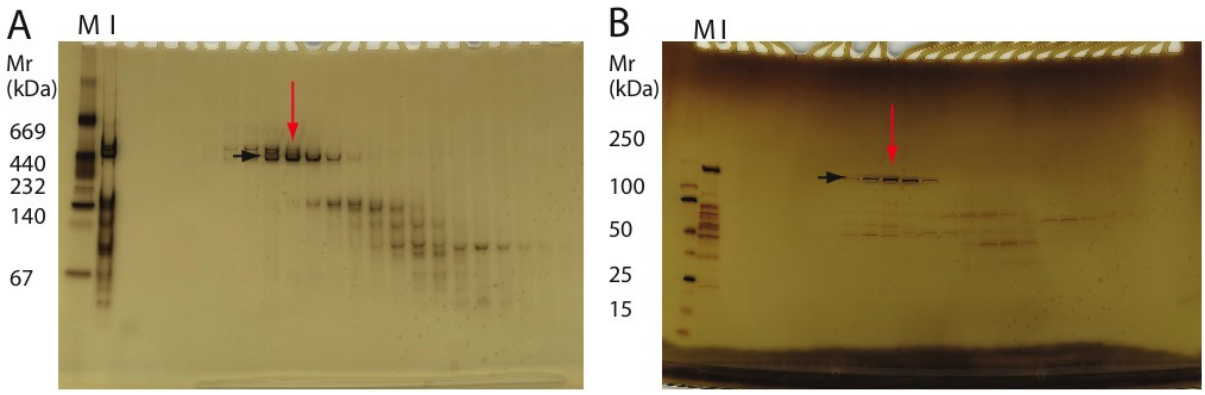


Figure S13. Purification of pyruvate carboxylase. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the pyruvate carboxylase. Red arrows show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

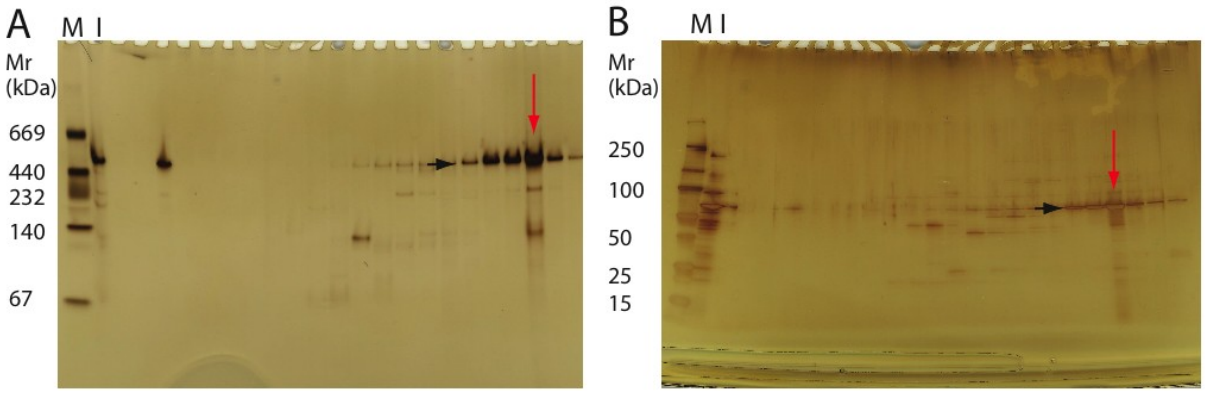


Figure S14. Purification of GroEL. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the 60 kDa chaperonin. Red arrows show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the HIC column is shown in lane I.

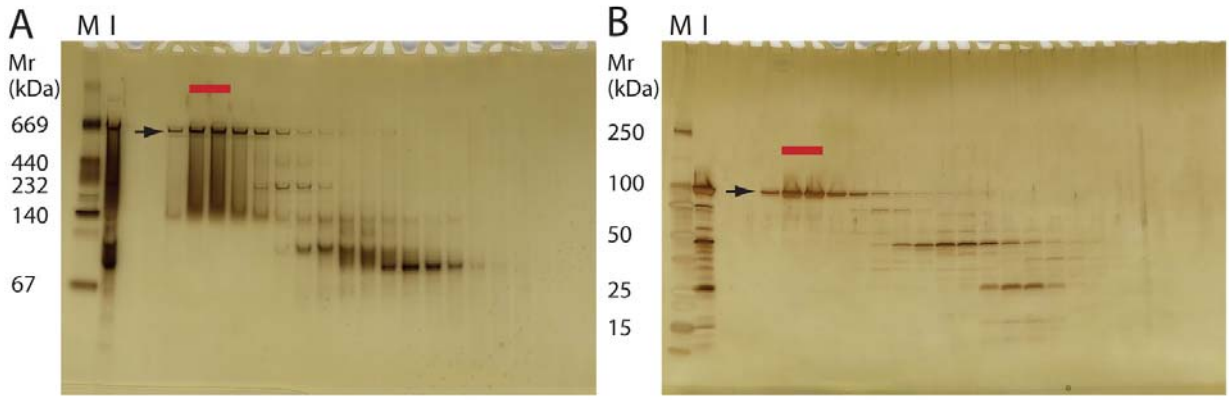


Figure S15. Purification of phosphorylase. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the phosphorylase. Red bars show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

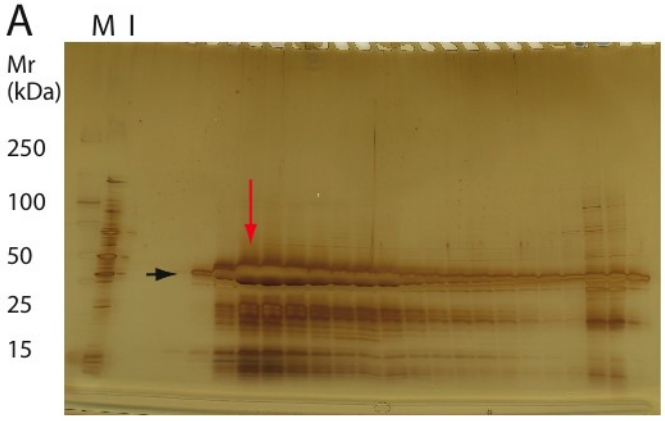


Figure S16. Purification of alcohol dehydrogenase. Fractions of the SEC column were analyzed by SDS PAGE (A). Black arrow shows the alcohol dehydrogenase. Red arrow shows the fraction used for EM analysis. Native gels for this protein did not show a band for this protein, consistent with the expectation, based on its failure to bind to the Q column, that it is a positively charged protein. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

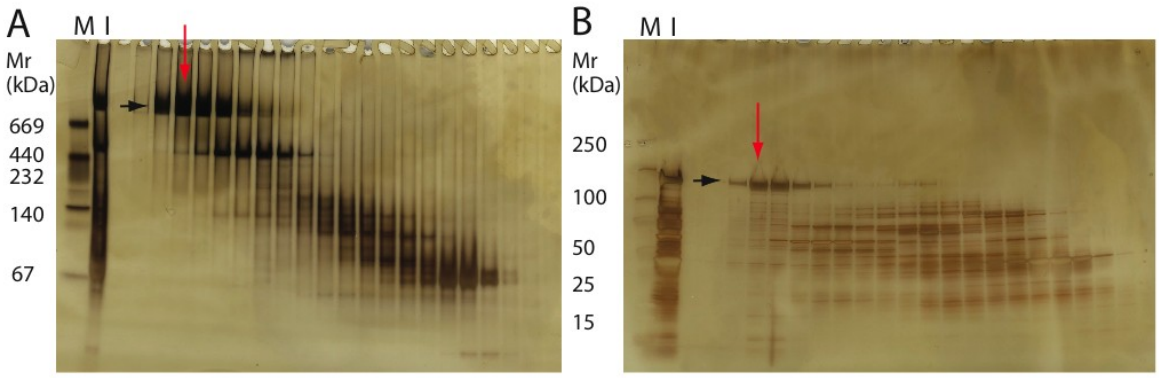


Figure S17. Purification of PFOR. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrows show the PFOR. Red arrows show the fraction used for EM analysis. Molecular weight standards are shown in lane M. The input fraction to the SEC column is shown in lane I.

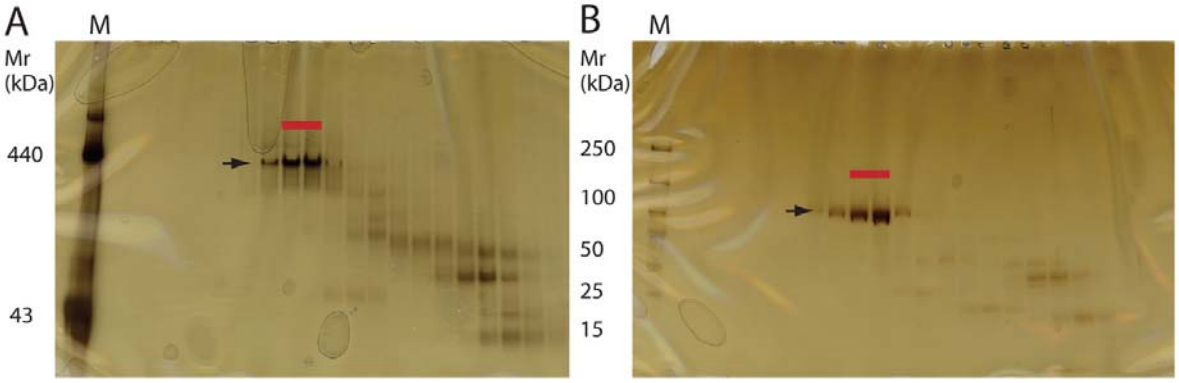


Figure S18. Purification of proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase. Fractions of the SEC column were analyzed by native PAGE (A) and SDS PAGE (B). Black arrow shows the proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase. Red bars show the fraction used for EM analysis. Molecular weight standards are shown in lane M.

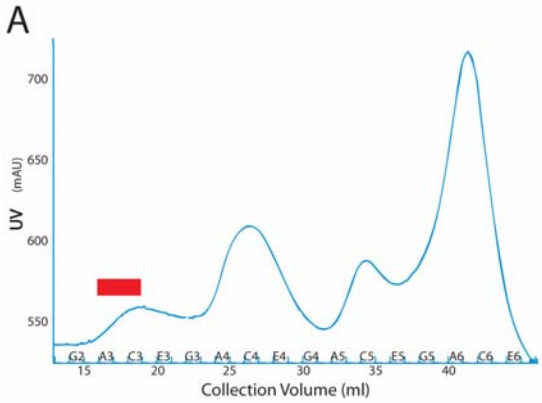


Figure S19. Purification of 70S ribosomes. Ribosome collection from sucrose gradient was monitored by UV. Red bar shows the fractions used for EM analysis.

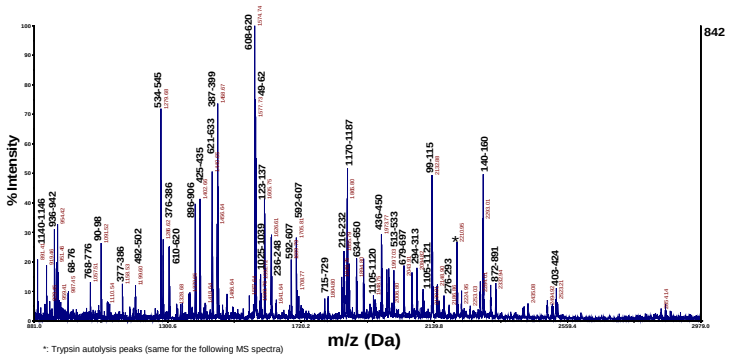
DVU1833 Q72B07 ppsA

MS Peak List

839.3069
842.4485
859.3933
874.3741
891.4056
919.4583
929.4477
944.4466
951.4567
954.424
959.4117
964.4175
987.4547
1057.5081
1083.5593
1091.5172
1110.5382
1112.5132
1118.451
1158.5325
1198.5535
1199.6007
1279.6814
1286.6178
1305.5729
1328.6779
1341.5608
1346.5524
1367.6666
1370.6519
1387.5759
1402.6646
1418.6361
1440.6504
1456.6385
1458.6711
1462.6487
1473.6139
1486.6431
1557.6381
1570.743
1574.7408
1577.7346
1590.7552
1592.7167
1605.7542
1626.6051
1641.6393
1684.5385
1689.7922
1705.8108
1708.7698
1713.7037
1720.8252
1795.7008
1804.7955
1848.7457
1855.7877
1865.7976
1871.7606
1894.8046
1916.8263
1948.7451
1953.8015
1971.7705
1973.7677
1989.7312
1997.0251
2006.7968
2011.8917
2043.9148
2068.8394
2084.825
2104.8262
2131.9177
2132.8845
2146.9089
2148.9011
2154.884
2169.8477
2186.8606
2210.9504
2224.9529
2253.0254
2284.0117
2293.0071
2316.9397
2332.9377
2420.0605
2435.0784
2494.9224
2510.8818
2523.2136
2527.8867
2854.1365
2867.0769

A. MS spectrum

External calibration



*: Trypsin autolysis peaks (same for the following MS spectra)

B. PMF result

Band
5

Aldente version 11/02/2008

Spectrum Peaks 96 / Mass [839.31; 2867.08] / Intensity [1; 1] / pI - / Mw -
Release UniProtKB/Swiss-Prot - Release 54.8 of 05-Feb-2008
UniProtKB/TrEMBL - Release 37.8 of 05-Feb-2008
Proteins -In range 268080 / After digestion 76695
-First Analysis on 76695 sequences : After alignment 49420
-Second Analysis on best 50 of first analysis : After alignment 50 / Displayed 50
Peptides Generated 32035090 / Matching a peak 740778 / Average per protein 119
Statistics Threshold 17.10

Rank	Score	Hits	AC	ID	Name	MW	pI	Cov %	TaxId
1	**719.98**	**54**	Q72B07	**Q72B07_DESVH**	**Phosphoenol**	**133**	**6.0**	**50**	**882**
2	12.76	8	Q725Z2	Q725Z2_DESVH	ADP-ribosylgly	32	5.9	35	882

Phosphoenolpyruvate synthase, putative

Score	Mw	pI	Hits	Coverage	Shift (Da)	Slope (ppm)
719.98	132643	5.99	54	50%	0.008	-72

	Exp Da	Theo Da	Intensity	%	rank	Delta Da	Delta ppm	Dev ppm	MC	Modifs CAM	Modifs MSO	Modifs PTM	Position start	Position end	Sequence
	859.393	859.446	1	100	1	-0.05	-60	2	-	-	-		343	348	FYVFQR
	891.406	891.468	1	100	1	-0.06	-69	-5	-	-	-		1140	1146	QELAQFR
*	944.447	944.495	1	100	1	-0.05	-50	13	-	-	-		936	942	YLEQVHR
	959.412	959.479	1	100	1	-0.07	-69	-5	-	-	-		349	357	DDGAEVVVR
	987.455	987.537	1	100	1	-0.08	-82	-17	-	-	-		68	76	AISSVAFHR
	1057.508	1057.575	1	100	1	-0.07	-62	2	1	-	-		768	776	RAEIEANVR
	1083.559	1083.652	1	100	1	-0.09	-85	-19	-	-	-		27	36	LILNGADIVR
	1091.517	1091.584	1	100	1	-0.07	-61	4	1	-	-		90	98	AIVDKEFNR
	1110.538	1110.59	1	100	1	-0.05	-46	19	-	-	-		702	711	GTEEVLAVHR
	1112.513	1112.569	1	100	1	-0.06	-50	15	-	-	-		736	745	TSVDLADHVR
	1118.451	1118.497	1	100	1	-0.05	-40	25	-	-	0/2		503	512	TDPDMVPGMR
*	1158.533	1158.615	1	100	1	-0.08	-71	-4	-	-	-		377	386	VPVPDLEAYR
	1198.553	1198.606	1	100	1	-0.05	-43	22	-	-	-		193	204	EAGADDVPVAVR
	1199.601	1199.674	1	100	1	-0.07	-60	5	1	-	-		492	502	INKGDILAAER
	1279.681	1279.773	1	100	1	-0.09	-71	-4	-	-	-		534	545	ELGIPAIIGIQR
	1286.618	1286.71	1	100	1	-0.09	-71	-4	1	-	-		376	386	KVPVPDLEAYR
	1305.573	1305.669	1	100	1	-0.1	-72	-5	-	-	-		610	620	EVEDFEVGLLR
	1346.552	1346.655	1	100	1	-0.1	-75	-8	-	-	-		551	563	ALDGQDVTVDGTR
	1367.667	1367.753	1	100	1	-0.09	-62	4	-	-	-		782	794	AEQIPAVQEALAK
	1370.652	1370.727	1	100	1	-0.08	-54	12	-	-	-		13	25	GRPAEADIAELTK
	1387.576	1387.664	1	100	1	-0.09	-63	4	-	-	-		896	906	NIHDWEVEAFK
	1402.665	1402.759	1	100	1	-0.09	-66	0	-	-	-		425	435	LWFVQARPETR
	1440.65	1440.742	1	100	1	-0.09	-62	4	-	-	0/1		621	633	AEFMLGNIGVHPR
	1456.639	1456.737	1	100	1	-0.1	-66	0	-	-	1/1		621	633	AEFMLGNIGVHPR
	1458.671	1458.77	1	100	1	-0.1	-67	0	-	-	-		387	399	WSLSLAQAEQVAR
*	1574.741	1574.854	1	100	1	-0.11	-71	-3	1	-	-		608	620	LREVEDFEVGLLR
	1577.735	1577.828	1	100	1	-0.09	-58	9	-	-	-		49	62	NYNTALISQVEGIR
	1592.717	1592.821	1	100	1	-0.1	-65	2	-	1/1	-		1025	1039	VGFCGQGVSNSVILR
	1605.754	1605.86	1	100	1	-0.11	-65	2	-	-	-		123	137	DIHAEVQANPGTLIK
	1626.605	1626.712	1	100	1	-0.11	-65	2	-	1/1	-		236	248	AYHWDCASAYNLR
	1689.792	1689.936	1	100	1	-0.14	-84	-16	-	-	0/1		592	607	VGLILADVGQAMFLSR
	1705.811	1705.931	1	100	1	-0.12	-69	-1	-	-	1/1		592	607	VGLILADVGQAMFLSR
	1720.825	1720.916	1	100	1	-0.09	-52	15	1	1/1	-		1024	1039	KVGFCGQGVSNSVILR
	1804.795	1804.919	1	100	1	-0.12	-67	0	1	-	-		715	729	ELDHKLDEHVELATR
	1855.788	1855.901	1	100	1	-0.11	-60	8	-	-	0/1		216	232	AFAGLQDTYLNMVGEAR
	1865.798	1865.914	1	100	1	-0.12	-61	6	-	-	-		1170	1187	QAGFASFAEQAEALAAQR
	1871.761	1871.896	1	100	1	-0.14	-71	-2	-	-	1/1		216	232	AFAGLQDTYLNMVGEAR
	1894.805	1894.929	1	100	1	-0.12	-65	3	-	-	-		634	650	ALEAYDNGELEHVVHAK
	1948.745	1948.867	1	100	1	-0.12	-62	6	-	-	-		1105	1120	HTQAEDLSDWYEGELR
	1973.768	1973.896	1	100	1	-0.13	-64	4	-	-	0/1		436	450	WNEEFETHPHTIFMR
	1989.731	1989.891	1	100	1	-0.16	-79	-10	-	-	1/1		436	450	WNEEFETHPHTIFMR
*	2011.892	2012.041	1	100	1	-0.15	-73	-4	-	-	-		513	533	VASAILADVGGDTSHAAITSR
	2043.915	2044.035	1	100	1	-0.12	-58	10	-	-	-		679	697	EYVGHVTGLAAEIEELASR
	2068.839	2068.979	1	100	1	-0.14	-66	2	-	1/1	0/1		294	313	MINPVISGTAFSADTATGCR
	2084.825	2084.974	1	100	1	-0.15	-70	-1	-	1/1	1/1		294	313	MINPVISGTAFSADTATGCR
	2104.826	2104.968	1	100	1	-0.14	-66	2	1	-	-		1105	1121	HTQAEDLSDWYEGELRR
	2132.885	2133.025	1	100	1	-0.14	-65	4	1	-	-		99	115	IDWNDTEINRDPEFLQK
	2169.848	2170.009	1	100	1	-0.16	-73	-4	1	1/1	1/2		276	293	AKQEWAIENTSLSVCMMR
	2293.007	2293.146	1	100	1	-0.14	-60	9	-	-	-		140	160	TFVNNVVEGFATSPEGIDQLR
	2316.94	2317.103	1	100	1	-0.16	-69	0	-	-	0/1		872	891	NLLGGNLFEQHEDNPMLGYR
	2332.938	2333.098	1	100	1	-0.16	-68	1	-	-	1/1		872	891	NLLGGNLFEQHEDNPMLGYR
	2494.922	2495.089	1	100	1	-0.17	-66	3	-	1/1	0/2		403	424	NISVAYGSMIMDTEFCIDSGER
	2510.882	2511.084	1	100	1	-0.2	-79	-10	-	1/1	1/2		403	424	NISVAYGSMIMDTEFCIDSGER
	2867.077	2867.286	1	100	1	-0.21	-72	-2	-	-	1/2		967	991	QFIEEFDGFSIGSNDMTQMVLATDR
						avg	**-65.2**								
						stdev	**9.3**								

```
   1 mgktqaekpa akGRPAEADI AELTKkLILN GADIVRiged aellvggkNY NTALISQVEG IRtpqfrAIS SVAFHRllde
  81 tkvnaslirA IVDKEFNRID WNDTEINRDP EFLQKfvrnl akDIHAEVQA NPGTLIKlrT FVNNVVEGFA TSPEGIDQLR
 161 krsvlvqagi lsvdlpkeve devkrayrdi ckEAGADDVP VAVRssaage dsrkkAFAGL QDTYLNMVGE ARvarAYHWD
 241 CASAYNLRsm tyrreailda laraeatgde siaikAKQEW AIENTSLSVC MMRMINPVIS GTAFSADTAT GCRgtsrrdl
 321 vsidasyglg eavvggmvtp dkFYVFQRDD GAEVVVRqmg ckdkkivyde kggtkKVPVP DLEAYRWSLS LAQAEQVARg
 401 vrNISVAYGS MIMDTEFCID SGERLWFVQA RPETRWNEEF ETHPHTIFMR rlevdpkAAA GAEIIVEGNG ASRGAGQGRv
 481 kylrsaleln kINKGDILAA ERTDPDMVPG MRVASAILAD VGGDTSHAAI TSRELGIPAI IGIQRleilr ALDGQDVTVD
 561 GTRgrvyrgm lplrEVGGEM DLSKLPATKt kVGLILADVG QAMFLSRLRE VEDFEVGLLR AEFMLGNIGV HPRALEAYDN
 641 GELEHVVHAK LKELDVNLSK llreqlsagl igidmklrEY VGHVTGLAAE IEELASRdnl rGTEEVLAVH RrMRELDHKL
 721 DEHVELATRr mdilkTSVDL ADHVRvimgy ddelallagt dpdilkrRAE IEANVRahve rAEQIPAVQE ALAKiaslrh
 801 evglrsglqt amddvravpe kirllirsrg frtgkehyvq tlagglalfa mafygkdivy rttdfksney rNLLGGNLFE
 881 QHEDNPMLGY RgvsrNIHDW EVEAFKlarg vyggvnlqim lpfvrtleea rsmkrYLEQV HRlkSGVDGL KVILMSEIPA
 961 NAVLAKQFIE EFDGFSIGSN DMTQMVLATD Rdnaslghiy deedpavvwa ilvtiftgqk ygrKVGFCGQ GVSNSVILRg
1041 lvaiagivsa svvpdtyyqt kldmaevesh niptekIgaw lreqhfgklr qlldtkgygh ilkkHTQAED LSDWYEGELR
1121 RLHEQLResm dtpkekfyrQ ELAQFRaafh kpviyatwdw nrtvedamrQ AGFASFAEQA EALAAQRakk
```

Figure S20: Identification of DVU1833 (gene ppsA) by PMF using Aldente search engine. Peak list that was submitted to the PMF query is given in the left hand side column.

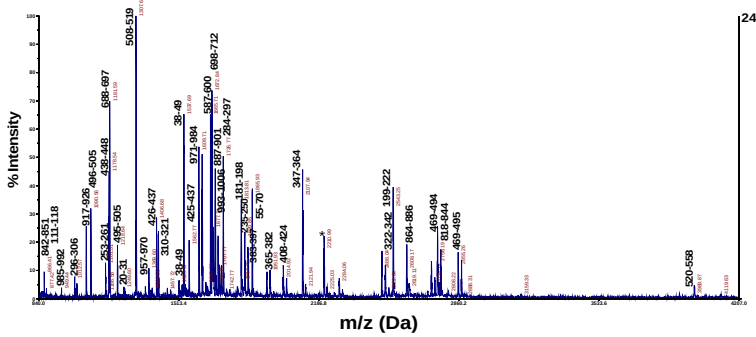
DVU3319 Q725V6 putA

MS Peak List

803.3683
825.0594
842.4854
855.0151
861.4355
866.4055
877.0071
877.4232
949.4435
1012.5703
1019.4773
1022.5712
1068.5537
1090.5619
1161.5079
1166.5759
1176.6289
1178.5422
1181.5863
1218.6449
1249.6017
1253.5824
1307.6654
1352.6011
1368.5634
1369.5957
1385.5939
1406.6829
1415.6232
1457.7244
1473.7035
1513.5734
1529.5758
1536.6971
1537.6918
1545.5519
1550.7874
1562.7668
1609.7058
1625.689
1643.7866
1665.7051
1672.838
1677.7109
1679.719
1688.8323
1701.759
1707.7715
1717.7563
1726.7725
1742.7748
1794.7462
1812.809
1813.8062
1817.7581
1826.798
1829.8479
1845.8381
1865.9252
1935.9567
1951.9348
2014.9235
2030.8994
2107.9373
2121.9351
2210.9902
2225.0317
2283.0725
2284.0605
2300.0674
2488.041
2505.0552
2522.0442
2543.2488
2560.137
2608.1709
2619.1094
2623.2019
2648.9902
2708.2285
2727.1599
2743.1531
2756.186
2762.2246
2764.2617
2772.1553
2779.2583
2808.2229
2855.2629
2871.229
2888.314
3159.3291
3989.8677
3994.7737
4119.8345

A. MS spectrum

External calibration



B. PMF result

Band 1

Aldente version 11/02/2008

Spectrum Peaks 95 / Mass [803.37; 4119.83] / Intensity [1; 1] / pI - / Mw -
Release UniProtKB/Swiss-Prot - Release 54.8 of 05-Feb-2008
UniProtKB/TrEMBL - Release 37.8 of 05-Feb-2008
Proteins -In range 268080 / After digestion 69174
-First Analysis on 69174 sequences : After alignment 43590
-Second Analysis on best 50 of first analysis : After alignment 50 / Displayed 50
Peptides Generated 32035090 / Matching a peak 667765 / Average per protein 119
Statistics Threshold 17.10

Rank	Score	Hits	AC	ID	Name	MW	pI	Cov %	TaxId
1	896.29	62	Q725V6	Q725V6_DESVH	Proline dehydrogen	112	5.9	61	882

Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase

Score	Mw	pI	Hits	Coverage	Shift (Da)	Slope (ppm)
931.96	111946	5.88	62	61%	0.051	-76

Exp Da	Theo Da	Intensity	%	rank	Delta Da	ppm	Dev ppm	MC	Modifs CAM	MSO	PTM	Position start	end	Sequence
803.368	803.389	1	100	1	-0.02	-25	-11	-	-	-		852	858	IAEEEGR
861.436	861.436	1	100	1	0	0	18	1	1/1	-		578	584	RAADICR
866.405	866.437	1	100	1	-0.03	-35	-16	-	-	-		845	851	NVSDYIR
877.423	877.42	1	100	1	0	4	23	-	-	0/1		111	118	SNIEGMAR
949.443	949.477	1	100	1	-0.03	-34	-11	-	-	0/1		985	992	VVTENTMR
1012.57	1012.615	1	100	1	-0.04	-43	-16	1	-	-		298	306	SEKLPIGIR
1019.477	1019.527	1	100	1	-0.05	-48	-20	1	-	-		262	269	LRSDPEFR
1022.571	1022.611	1	100	1	-0.04	-37	-10	-	-	-		460	468	LLAARPEPR
1068.554	1068.584	1	100	1	-0.03	-27	2	-	-	-		917	926	FALTGAVFSR
1090.562	1090.589	1	100	1	-0.03	-24	6	-	-	-		496	505	AFVEALADVR
1166.576	1166.624	1	100	1	-0.05	-40	-7	1	-	0/1		253	261	EMTLELFKR
1176.629	1176.662	1	100	1	-0.03	-27	6	-	-	-		801	810	VIVLDAIYDR
1178.542	1178.58	1	100	1	-0.04	-31	2	-	-	-		438	448	QSFAEGAALER
1181.586	1181.631	1	100	1	-0.05	-37	-3	-	-	-		688	697	IGYNLAEVFR
1218.645	1218.684	1	100	1	-0.04	-31	3	1	-	-		495	505	KAFVEALADVR
1249.602	1249.642	1	100	1	-0.04	-31	4	-	-	-		20	31	SISGEAPSIFNK
1253.582	1253.627	1	100	1	-0.04	-35	1	-	-	-		288	297	DLDDLLHWAR
1307.665	1307.711	1	100	1	-0.05	-34	4	-	-	-		508	519	FGQTVPLYIGGR
1369.596	1369.657	1	100	1	-0.06	-44	-4	-	-	0/1		957	970	QPFGGFAMSGVGSK
1385.594	1385.652	1	100	1	-0.06	-41	0	-	-	1/1		957	970	QPFGGFAMSGVGSK
1406.683	1406.727	1	100	1	-0.04	-31	9	-	-	-		426	437	LLENTANESFLR
1415.623	1415.684	1	100	1	-0.06	-42	-1	-	-	-		310	321	GAYWDVETVIAK
1513.573	1513.645	1	100	1	-0.07	-46	-3	-	-	0/2		38	49	VMDWAMQNEDFK
1529.576	1529.64	1	100	1	-0.06	-41	2	-	-	1/2		38	49	VMDWAMQNEDFK
1536.697	1536.765	1	100	1	-0.07	-43	0	-	-	-		902	916	AATFDEALSIANGTR
1537.692	1537.755	1	100	1	-0.06	-40	3	-	-	-		635	647	APGEHNHLFYQPK
1545.552	1545.635	1	100	1	-0.08	-53	-8	-	-	2/2		38	49	VMDWAMQNEDFK
1550.787	1550.844	1	100	1	-0.06	-35	8	1	-	-		506	519	SRFGQTVPLYIGGR
1562.767	1562.829	1	100	1	-0.06	-39	5	1	-	-		425	437	RLLENTANESFLR
1609.706	1609.768	1	100	1	-0.06	-38	7	-	-	0/1		971	984	TGGPDYLLQFMDPR
1625.689	1625.763	1	100	1	-0.07	-44	0	-	-	1/1		971	984	TGGPDYLLQFMDPR
1643.787	1643.879	1	100	1	-0.09	-55	-9	-	-	-		587	600	IWELSAWQVVEVGK
1665.705	1665.769	1	100	1	-0.06	-37	8	-	1/1	-		698	712	EAGLPEGVFNYCPGR
1672.838	1672.909	1	100	1	-0.07	-41	4	-	-	0/1		887	901	IAQEEIFGPVLAVMR
1677.711	1677.776	1	100	1	-0.06	-38	8	1	-	-		993	1006	RGFTPIDEDDWIV
1679.719	1679.839	1	100	1	-0.12	-70	-24	1	-	0/3		104	118	LMGMTIRSNIEGMAR
1688.832	1688.904	1	100	1	-0.07	-41	5	-	-	1/1		887	901	IAQEEIFGPVLAVMR
1701.759	1701.831	1	100	1	-0.07	-41	5	-	-	0/1		383	396	YEFQVLYGMAEPVR
1707.772	1707.848	1	100	1	-0.08	-44	3	-	-	-		74	88	EYFATEDADIPPVLK
1717.756	1717.825	1	100	1	-0.07	-39	7	-	-	1/1		383	396	YEFQVLYGMAEPVR
1726.773	1726.839	1	100	1	-0.07	-38	9	1	-	-		284	297	DTEKDLDDLLHWAR
1812.809	1812.892	1	100	1	-0.08	-45	4	-	-	-		181	198	ALPGNGPVEGFDWGATPK
1826.798	1826.885	1	100	1	-0.09	-47	2	-	1/1	1/2		235	250	VVAMGGFLCIDMEQLK
1829.848	1829.925	1	100	1	-0.08	-41	7	1	-	0/1		383	397	YEFQVLYGMAEPVRK
1845.838	1845.92	1	100	1	-0.08	-44	5	1	-	1/1		383	397	YEFQVLYGMAEPVRK
1865.925	1866.001	1	100	1	-0.08	-39	9	-	-	-		55	70	FVDVLPYLNTSESLLR
1935.957	1936.032	1	100	1	-0.08	-38	12	-	-	0/1		365	382	TIAAVMETALALNVPEHR
1951.935	1952.027	1	100	1	-0.09	-46	4	-	-	1/1		365	382	TIAAVMETALALNVPEHR
2014.924	2015.013	1	100	1	-0.09	-43	7	-	1/1	0/1		408	424	LYCPYGELIPGMAYLVR
2030.899	2031.008	1	100	1	-0.11	-52	0	-	1/1	1/1		408	424	LYCPYGELIPGMAYLVR
2107.937	2108.023	1	100	1	-0.09	-40	12	-	1/1	-		347	364	ILENSDIVYFACASHNVR
2505.055	2505.193	1	100	1	-0.14	-54	2	-	-	-		322	342	QNGWEIPVWTDKPESDIAYEK
2543.249	2543.383	1	100	1	-0.13	-52	4	-	-	-		199	222	VNVSIKPSALYSQAKPVDVEGSVR
2608.171	2608.319	1	100	1	-0.15	-56	1	-	1/1	-		864	886	TDLPAEGCYVPLTIVGDIRPEHR
2727.16	2727.308	1	100	1	-0.15	-53	4	-	-	0/1		469	494	AVEPGPGGLPPFTNDAMIDFTVPDNR
2743.153	2743.303	1	100	1	-0.15	-54	4	-	-	1/1		469	494	AVEPGPGGLPPFTNDAMIDFTVPDNR
2756.186	2756.32	1	100	1	-0.13	-47	10	-	-	0/1		818	844	AASSIHIGPSEDPSNYMGPVADATLQK
2764.262	2764.42	1	100	1	-0.16	-56	1	1	1/1	-		863	886	RTDLPAEGCYVPLTIVGDIRPEHR
2772.155	2772.315	1	100	1	-0.16	-56	1	-	-	1/1		818	844	AASSIHIGPSEDPSNYMGPVADATLQK
2855.263	2855.403	1	100	1	-0.14	-48	10	1	-	0/1		469	495	AVEPGPGGLPPFTNDAMIDFTVPDNRK
2871.229	2871.398	1	100	1	-0.17	-58	0	1	-	1/1		469	495	AVEPGPGGLPPFTNDAMIDFTVPDNRK
3989.868	3990.049	1	100	1	-0.18	-44	19	-	1/1	-		520	558	DVTTADLIPTTNPAKPAEVVASICQAGRPEIDDAIAAAK
					avg	-40.8								
					stdev	11.8								

```
   1 mdqqhldgkv vergkeffrS ISGEAPSIFN KgwwtgkVMD WAMQNEDFKv qlfrFVDVLP YLNTSESLLR hirEYFATED
  81 ADIPPVLKwg agkagiggal takLMGMTIR SNIEGMARqf iigdnskeav kglaklrkdg ftftvdllge atvseeesea
 161 yaqgyhevvd aiareqekwk ALPGNGPVEG FDWGATPKVN VSIKPSALYS QAKPVDVEGS VRgilsrlvp iyrkVVAMGG
 241 FLCIDMEQLK ykEMTLELFK RLRSDPEFRh yphlsivlqa ylrDTEKDLD DLLHWARSEK LPIGIRlvkG AYWDYETVIA
 321 KQNGWEIPVW TDKPESDIAY EklahrILEN SDIVYFACAS HNVRTIAAVM ETALALNVPE HRYEFQVLYG MAEPVRKglk
 401 nvagrvrLYC PYGELIPGMA YLVRRLLENT ANESFLRQSF AEGAALERll enpqktlhrL LAARPEPRAV EPGPGGLPPF
 481 TNDAMIDFTV PDNRKAFVEA LADVRSRFGQ TVPLYIGGRD VTTADLIPTT NPAKPAEVVA SICQAGRPEI DDAIAAAKka
 561 altwrdtspa draaylrRAA DICRkrIWEL SAWQVVEVGK QWDQAYHDVT EGIDFLEYYA Remlrlgapr rmgrAPGEHN
 641 HLFYQPKgia aviapwnfpf aiaigmasaa ivtgnpvifk psssisrIGY NLAEVFREAG LPEGVFNYCP GRssimgdyl
 721 vehpdislic ftgsmevglr iqekaakvqp gqrqckrvia emggknatii ddddadldeav lqvlysafgf qgqkcsacsr
 801 VIVLDAIYQR flerlvkAAS SIHIGPSEDP SNYMGPVADA TLQKNVSDVI RIAEEEGRvl lkRTDLPAEG CYVPLTIVGD
 881 IRPEHRIAQE EIFGPVLAVM RAATFDEALS IANGTRFALT GAVFSRspeh ldkarrEFRV GNLYLNKgst galverQPFG
 961 GFAMSGVGSK TGGPDYLLQF MDPRVVTENT MRRGFTPIDE DDDWI
```

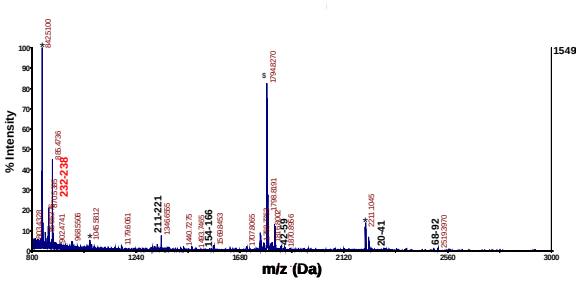
Figure S21: Identification of DVU3319 (gene putA) by PMF using Aldente search engine. Peak list that was submitted to the PMF query is given in the left hand side column.

DVU0460 Q72EV8

MS Peak List
870.5385
885.4736
1110.4879
1151.7205
1179.6061
1302.6804
1307.6875
1320.6133
1330.6626
1346.6555
1362.6346
1373.6764
1378.6421
1440.7275
1475.785
1493.7465
1558.718
1569.8453
1707.8065
1720.8778
1762.8268
1764.8417
1765.7625
1766.7831
1768.7754
1778.0041
1780.8025
1783.8281
1792.8202
1794.8271
1798.8191
1810.8218
1812.8135
1814.8063
1816.8002
1826.8115
1827.9003
1839.9486
1851.9344
1854.918
1868.8979
1870.8958
1892.8901
2119.1951
2207.1238
2211.1047
2225.1362
2234.114
2291.2227
2300.1904
2383.9741
2502.3018
2518.3567

A. MS spectrum

Internally calibrated using two trypsin autolysis peaks at m/z = 842.51 and 2211.1046



$ This peak is unidentified.

B. PMF result

Band 13 Aldente version 11/02/2008

Spectrum Peaks 53 / Mass [870.54; 2518.36] / Intensity [1; 1] / pI - / Mw -
Release UniProtKB/Swiss-Prot - Release 54.8 of 05-Feb-2008
UniProtKB/TrEMBL - Release 37.8 of 05-Feb-2008
Proteins -In range 268080 / After digestion 35051
-First Analysis on 35051 sequences : After alignment 17939
-Second Analysis on best 50 of first analysis : After alignment 50 / Displayed 50
Peptides Generated 32035090 / Matching a peak 403191 / Average per protein 119
Statistics Threshold 17.10

Rank	Score	Hits	AC	ID	Name	MW	pI	Cov %	TaxId
1	22.12	9	Q72EV8	Q72EV8_DESVH	Predicted phos	28	6	34	882

Predicted phospho-2-dehydro-3-deoxyheptonate aldolase

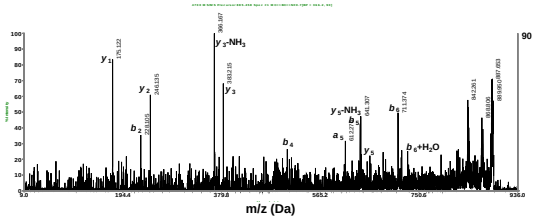
Score	Mw	pI	Hits	Cov	Shift (Da)	Slope (ppm)
22.12	28438	6.04	9	34%	-0.021	16

Exp Da	Theo Da	Intensity	%	rank	Delta Da	Delta ppm	Dev ppm	MC	Modifs CAM	Modifs MSO	Modifs PTM	Position start	Position end	Sequence
1330.663	1330.657	1	100	1	0.01	4	3	-	-	0/1		211	221	DFLQMVHDAVR
1346.656	1346.652	1	100	1	0	2	1	-	-	1/1		211	221	DFLQMVHDAVR
1558.718	1558.707	1	100	1	0.01	7	4	-	1/1	-		154	166	DQYDPQVVAHCAR
1827.9	1827.892	1	100	1	0.01	5	0	1	1/1	-		152	166	IRDQYDPQVVAHCAR
1854.918	1854.901	1	100	1	0.02	9	3	-	-	0/1		42	59	DTVNQVAEGGADAVLMHK
1870.896	1870.896	1	100	1	0	0	-5	-	-	1/1		42	59	DTVNQVAEGGADAVLMHK
2119.195	2119.176	1	100	1	0.02	9	2	-	-	-		72	92	DVGLIVHLSASTSLSPLPNAK
2291.223	2291.206	1	100	1	0.02	7	0	-	-	1/1		20	41	SVIVPLDHGVSVGPIDGLVDMR
2518.357	2518.362	1	100	1	-0.01	-1	-10	1	-	-		68	92	EGGRDVGLIVHLSASTSLSPLPNAK
						avg	-0.2							
						stdev	4.5							

```
  1 mhigkkirme rlfnrttgrS VIVPLDHGVS VGPIDGLVDM RDTVNQVAEG GADAVLMHKg lvrcghrEGG RDVGLIVHLS
 81 ASTSLSPLPN AKtltatved aikhgadgvs vhvnlgdete rdmladlgrv atiandwgvp llammyargp rIRDQYDPQV
161 VAHCARvgve lgadvvkvpy tgdmdtfahv vqsccvpvvi aggpkldstr DFLQMVHDAV Raggsglsvg rNIFQHARps
241 qlvkavrglv hedwdveqai aivge
```

C. MS/MS spectra

885.5 NIFQHAR (232-238)



Note: This peptide was not identified in PMF search but was identified in MS/MS seach using ProteinPilot 2.0 search engine.

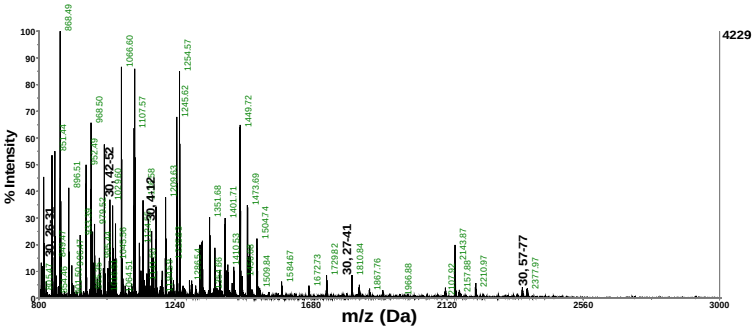
Figure S22: Identification of DVU0460 (gene DVU_0460) by PMF using Aldente search engine and by MS/MS using Protein Pilot. Peak list that was submitted to the PMF query is given in the left hand side column.

DVU3242 Q725M7 rpoZ

MS Peak List
811.4186
815.4746
823.4127
842.4973
849.46
851.4299
856.5207
868.4758
870.5374
896.4978
906.4606
933.3818
951.4811
952.4865
968.4998
973.5036
979.5214
995.4415
996.4513
997.4429
1008.4958
1011.4523
1022.5255
1027.4198
1029.6052
1032.5925
1038.4845
1045.5616
1047.4977
1066.6083
1071.607
1107.5753
1124.5598
1129.5926
1136.5936
1141.5133
1148.5712
1164.6119
1178.6268
1198.5651
1208.6586
1225.6515
1245.6473
1254.5933
1320.7158
1351.7111
1383.7492
1400.7739
1401.7498
1405.6572
1410.5711
1429.7753
1456.7209
1476.7593
1481.7276
1584.7318
1672.8251
1728.9088
1729.8945
1794.8433
1810.9208
1826.918
1867.8463
1879.9279
1910.9114
1966.9832
2015.0253
2078.0739
2107.078
2113.0937
2162.0236
2163.126
2211.1236
2284.2209
2346.1201
2362.1346
2378.1169
2383.2057
2395.1103
2435.1812

A. MS spectrum

External calibration



B. PMF result

Band 83-3 Aldente version 11/02/2008

Spectrum Peaks 80 / Mass [811.42; 2435.18] / Intensity [1; 1] / pI - / Mw -
Release UniProtKB/Swiss-Prot - Release 54.8 of 05-Feb-2008
UniProtKB/TrEMBL - Release 37.8 of 05-Feb-2008
Proteins -In range 268080 / After digestion 70421
-First Analysis on 70421 sequences : After alignment 42196
-Second Analysis on best 50 of first analysis : After alignment 50 / Displayed 50
Peptides Generated 32035090 / Matching a peak 691202 / Average per protein 119
Statistics Threshold 17.10

Rank	Score	Hits	AC	ID	Name	MW	pI	Cov %	TaxId
1	58.77	19	A4FUZ0	A4FUZ0_B	MGC139925 protei	54	5.4	29	9913
2	58.06	19	P15241	K2M2_SHE	Keratin, type II micr	54	5.5	31	9940
3	53.43	19	Q148H4	Q148H4_B	Similar to keratin, h	55	5.6	29	9913
4	48.24	19	A0JNT2	A0JNT2_H	KRT83 protein.	50	5.3	30	9606
5	45.39	18	Q28582	Q28582_S	Hair type II keratin	55	6.1	28	9940
6	43.63	16	A6NCN2	A6NCN2_H	Uncharacterized pr	39	5	33	9606
7	35.58	15	A1A4M3	A1A4M3_BLOC	S35232 proteir	40	5.1	36	9913
8	35.08	18	P78385	KRT83_HU	Keratin type II cutic	54	5.3	26	9606
9	35.08	18	A1A4S9	A1A4S9_H	Keratin 83.	54	5.4	26	9606
10	34.62	16	P25691	K2M3_SHE	Keratin, type II micr	55	6.2	26	9940
11	34.41	18	Q6NT21	Q6NT21_H	Keratin 83.	54	5.5	24	9606
12	30.19	16	A7M776	A7M776_R	Type II keratin Kb2	53	5.6	26	10116
13	28.46	14	Q9ERE2	Q9ERE2_M	Type II hair keratin	44	5.1	34	10090
14	27.53	16	Q6IMF0	Q6IMF0_M	Type-II keratin Kb2	55	5.9	24	10090
15	26.05	16	Q6IG08	Q6IG08_R	Type II keratin Kb2	62	7.8	23	10116
16	25.44	15	Q0VD04	Q0VD04_B	Similar to keratin, h	56	6.2	23	9913
17	24.89	16	Q43790	KRT86_HU	Keratin type II cutic	54	5.6	22	9606
18	24.89	16	A8K872	A8K872_H	cDNA FLJ77849, h	53	5.5	22	9606
19	23.95	16	Q14533	KRT81_HU	Keratin type II cutic	55	5.5	21	9606
20	22.84	14	P97861	KRT86_MC	Keratin type II cutic	53	5.6	27	10090
21	21.47	14	Q61726	Q61726_M	Keratin type II (Frag	55	6.1	23	10090
22	19.96	12	Q0VDM9	Q0VDM9_M	Krt78 protein (Frag	55	5.9	24	10090
23	19.31	12	Q0VDN0	Q0VDN0_M	Krt78 protein (Frag	57	6	23	10090
24	19.06	14	A5A6N5	A5A6N5_P	Keratin, hair, basic,	56	6.3	24	37012
25	18.78	14	Q06018	Q06018_S	Intermediate filame	53	8.6	24	9940
26	18.68	14	P78386	KRT85_HU	Keratin type II cutic	56	6.3	23	9606
27	18.29	15	Q17R48	Q17R48_H	Keratin, hair, basic,	55	5.5	20	9606
28	17.67	13	A7M746	A7M746_R	LOC681126 proteir	55	6.1	22	10116
29	14.82	13	Q9Z2T6	KRT85_MC	Keratin type II cutic	56	6.2	19	10090
30	**14.25**	**7**	**Q725M7**	**RPOZ_DE**	**DNA-directed RNA**	**9**	**5.2**	**71**	**882**

30. DNA-directed RNA polymerase subunit omega

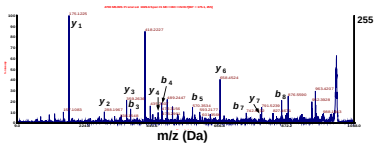
Score	Mw	pI	Hits	Cov	Shift (Da)	Slope (ppm)
14.25	8851	5.2	7	71%	-0.04	32

Exp Da	Theo Da	Intensity	%	rank	Delta Da	Delta ppm	Dev ppm	MC	Modifs CAM	MSO	PTM	Position start	Position end	Sequence
849.46	849.469	1	100	1	-0.01	-10	4	1	-	-		26	31	RVQQYR
1029.605	1029.605	1	100	1	0	0	6	1	-	-		42	50	NKEVVTALR
1148.571	1148.573	1	100	1	0	0	0	1	1/1	-		4	12	ITVEDCQKR
1810.921	1810.897	1	100	1	0.02	13	2	1	-	-		27	41	VQQYREGYEPLVDSK
2346.12	2346.099	1	100	1	0.02	9	-6	-	-	0/2		57	77	VMPEDLALYRPAEGEEMPVAE
2362.135	2362.094	1	100	1	0.04	17	1	-	-	1/2		57	77	VMPEDLALYRPAEGEEMPVAE
2378.117	2378.089	1	100	1	0.03	12	-3	-	-	2/2		57	77	VMPEDLALYRPAEGEEMPVAE
					avg	**5.9**								
					stdev	**9.5**								

1 marITVEDCQ KRidnrfllv qmaikRVQQY REGYEPLVDS KNKEVVTALR eiaagkVMPE DLALYRPAEG EEMPVA

C. MS/MS spectra

1029.6 NKEVVTALR (42-52)



1810.9 VQQYREGYEPLVDSK (27-41)

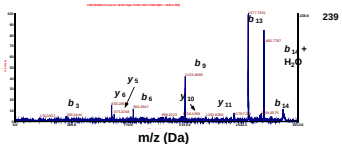


Figure S23: Identification of DVU3242 (gene rpoZ) by PMF using Aldente search engine and by MS/MS with manual interpretation. Peak list that was submitted to the PMF query is given in the left hand side column.

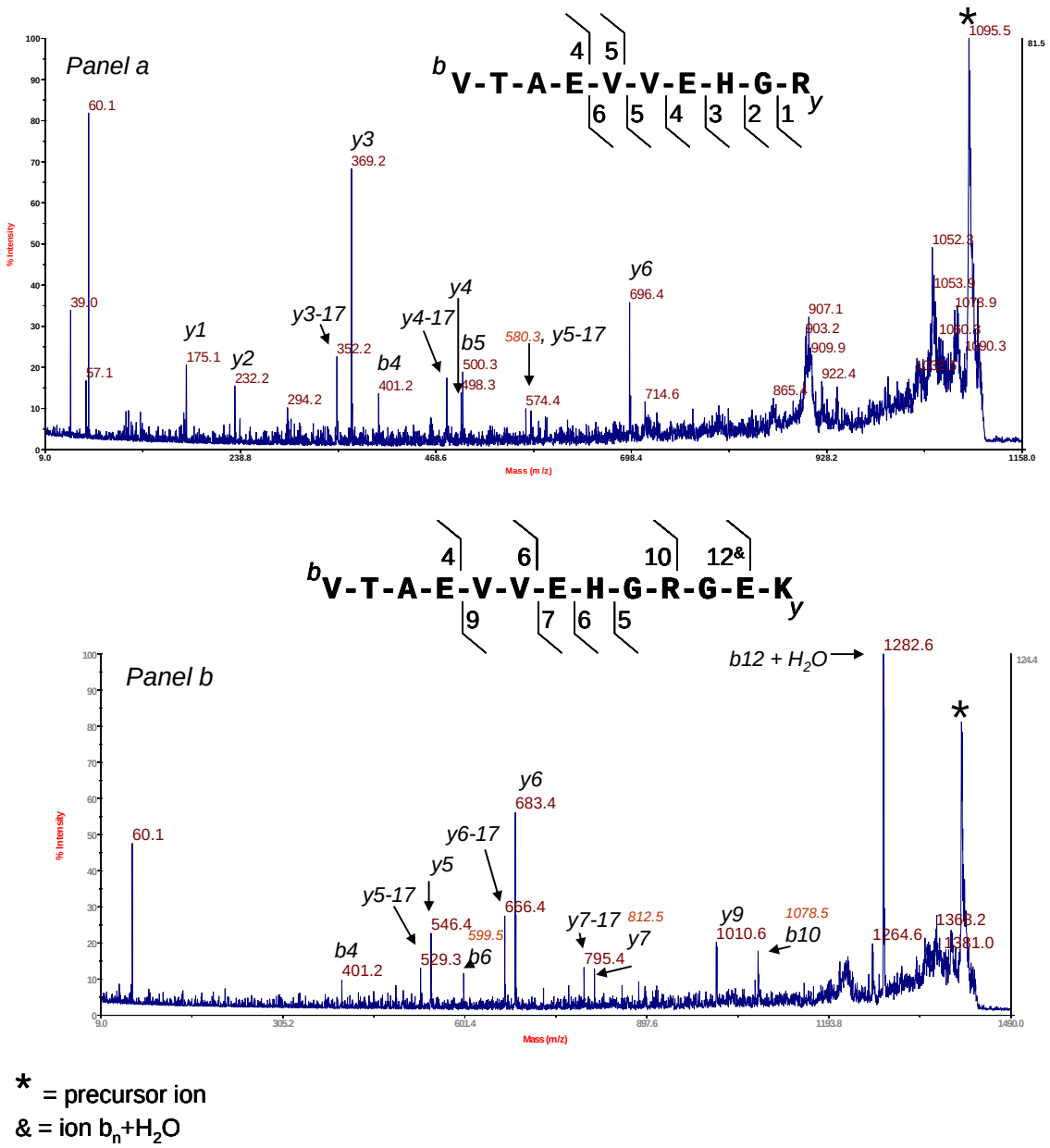
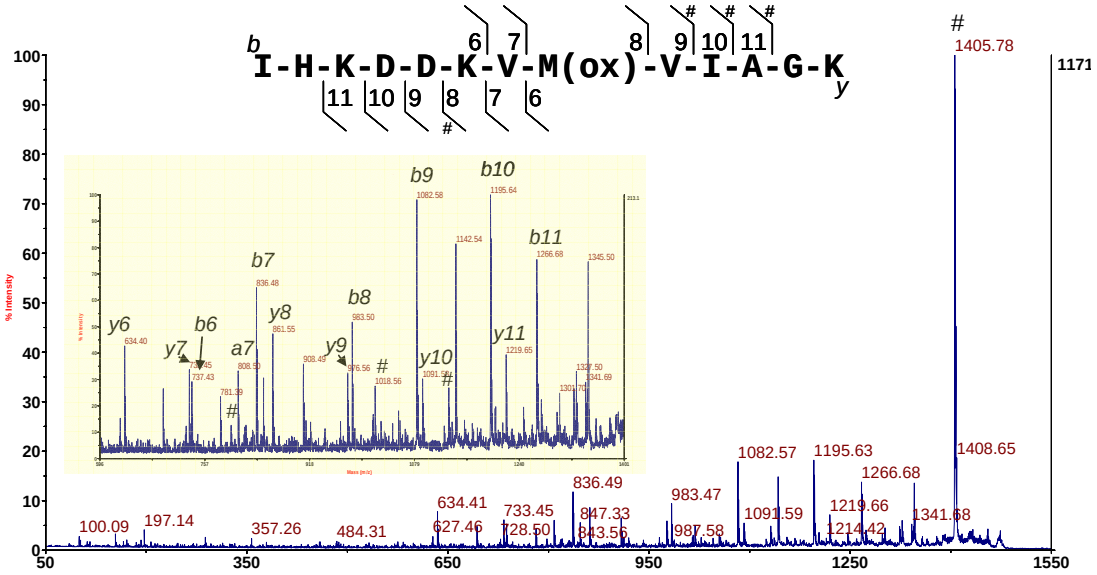


Figure S24: Identification of 50S ribosomal protein L21 (DVU0927, Q72DK2, gene rplU) by MS/MS analysis of molecular ions m/z 1096.58 (VTAEVVEHGR) and m/z 1410.73 (VTAEVVEHGRGEK), *panel a* and *panel b*, respectively.



#Annotates ions for which loss of $S(O)CH_4$ was detected

Figure S25: Identification of 50S ribosomal protein L24 (DVU1314, Q72CG9, gene rplX) by MS/MS analysis of molecular m/z 1469.76: IHKDDKVM(ox)VIAGK.

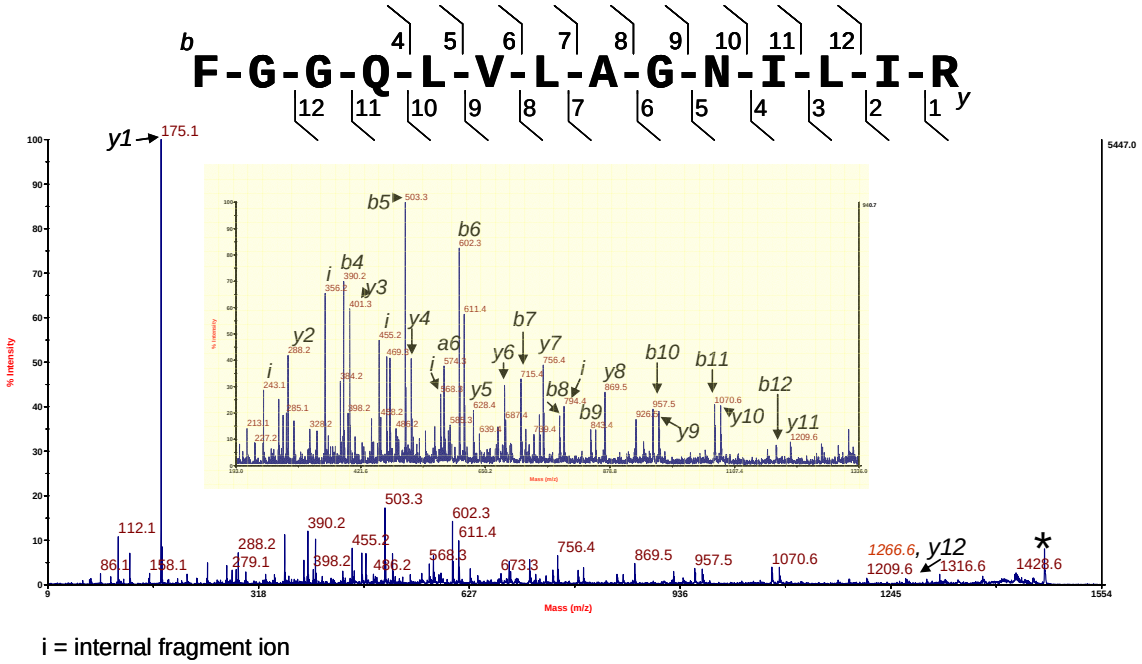


Figure S26: Identification of 50S ribosomal protein L27 (DVU0928, Q72DK1, gene rpmA) by MS/MS analysis of molecular m/z 1470.8: FGGQLVLAGNILIR.

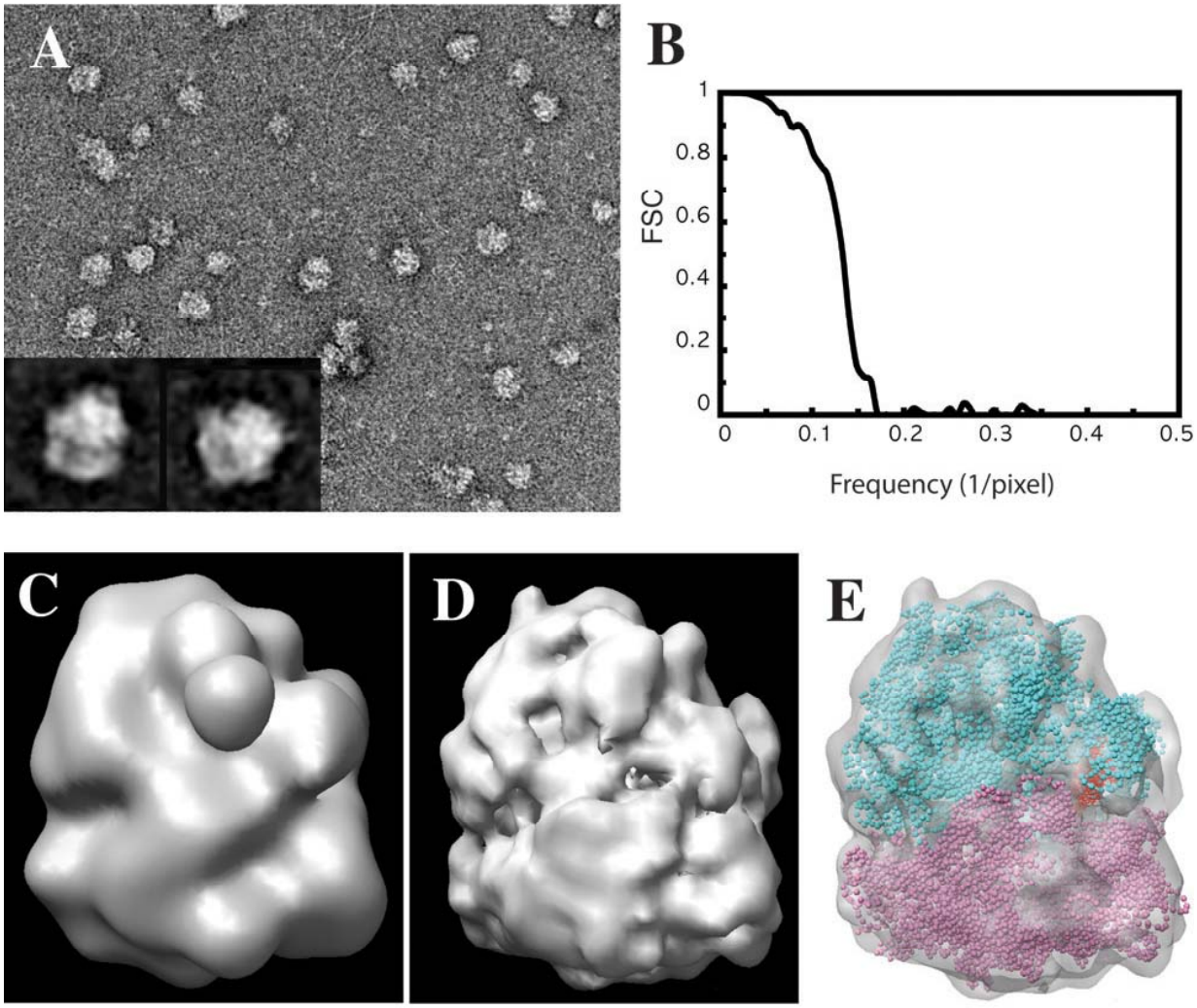


Figure S27. Ribosome EM data, refinement result, and docking of a pseudo-atomic model structure. 11,000 particles were used for the refinement. The standard FSC curve falls to 0.5 at a resolution of 24 Å after 15 cycles of refinement. (a) Electron micrograph of the negatively stained sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation for the final reconstruction. (c) The x-ray crystal structure of PDB accession codes 1GIX and 1GIY, filtered to a resolution of 50 Å, was used as the initial model for refinement. (d) The refined structure. (e) A X-ray model structure of 1GIX and 1GIY docked into the semitransparent EM density. The 30S subunit is shown in purple while the 50S subunit is shown in cyan. There is extra density in the EM map at the E site for binding of tRNA, shown in red.

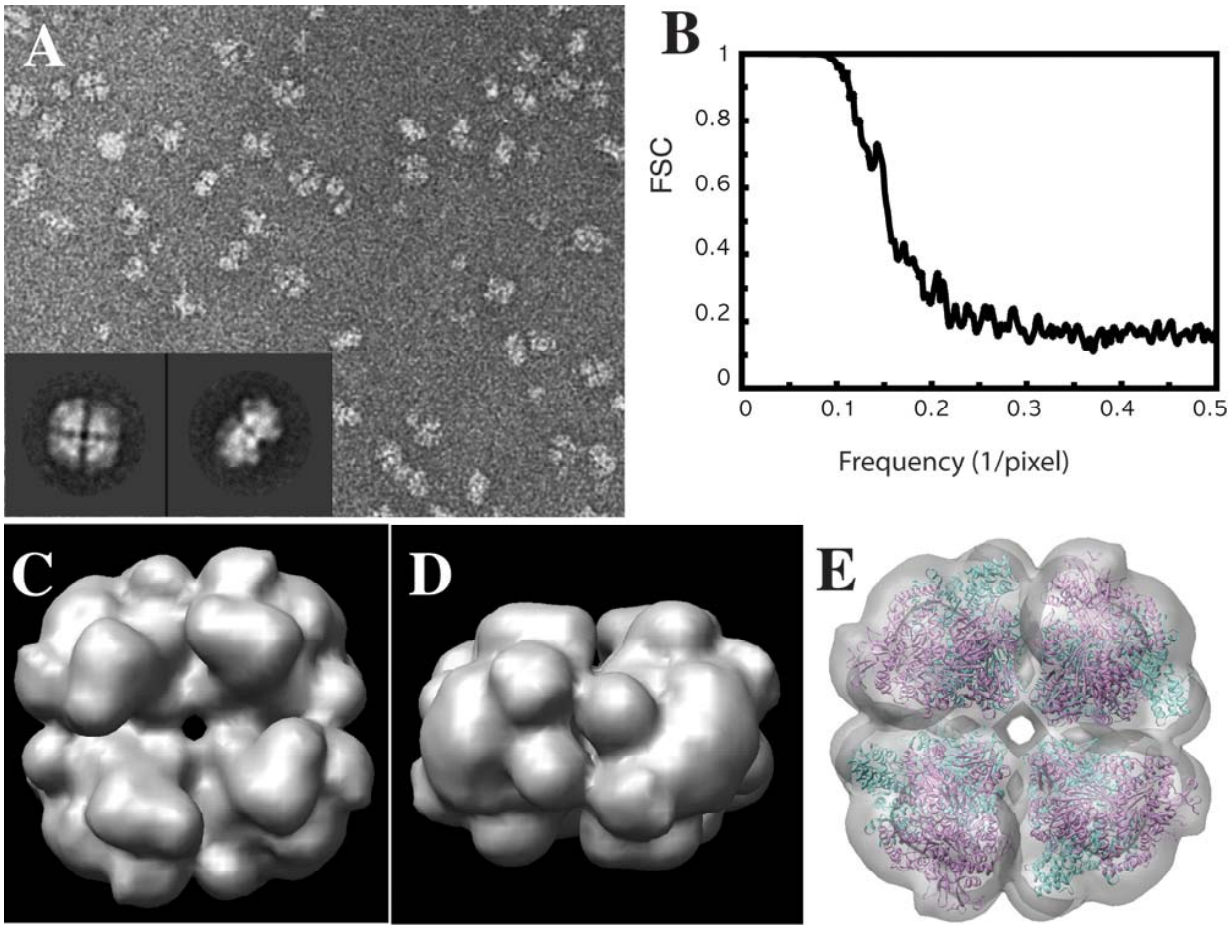


Figure S28. PFOR EM image and refinement published previously by Garczarek et al. (*5*), and shown here only for completeness. 12402 particles were used for the refinement. D4 symmetry was imposed during the refinement. Comparison of separate reconstructions from two halves shows a resolution of 17 Å at 0.5 FSC after refinement. (a) Electron micrograph of the negatively stained PFOR sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation curve. (c) A top view of the refined EM density. (d) A side view of the refined structure. (e) A homology model for DvH PFOR, based on a PDB template 1KEK, docked into the semitransparent EM density. The top half of the complex is represented as orchid ribbons and the bottom turquoise.

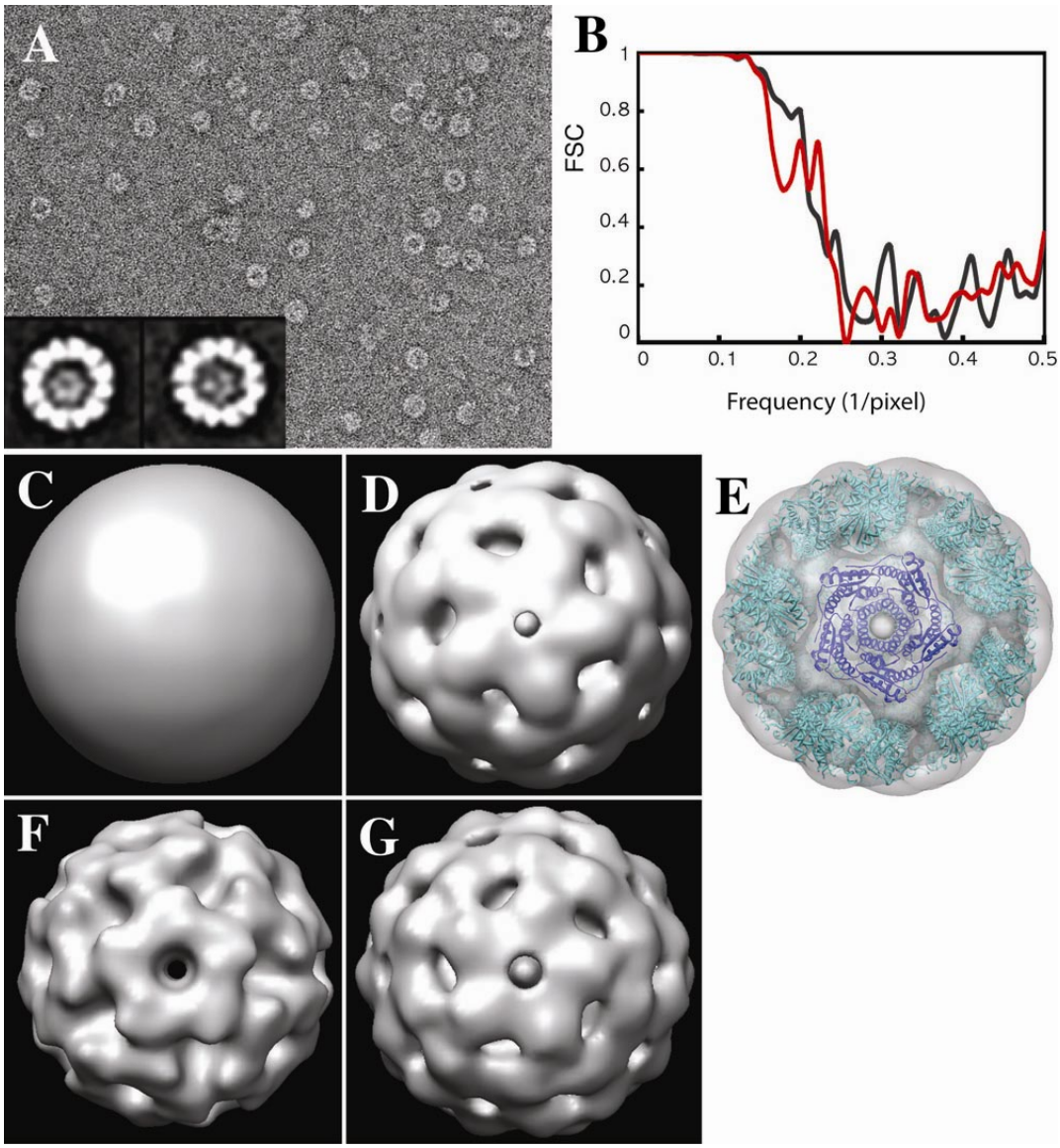


Figure S29. Lumazine synthase EM data, refinement result, and docking of a pseudo-atomic model structure. 3174 particles were used for the refinement. Icosahedral symmetry was imposed during the refinement, based on the size of the particle relative to the molecular weight of the monomer. When a uniform spherical shell is used as the initial model, the standard FSC curve (shown in black) falls to 0.5 at a resolution of 15 Å after two rounds of refinement, each of which consisted of 15 cycles. Refinement proceeded more slowly when a low-resolution X-ray structure (panel f) was used as the initial model, but the final result converged (FSC curve shown in red) to a structure similar to that obtained by starting with a spherical shell. We interpret the slower convergence in this case as indicating that the low-resolution X-ray structure started the refinement in a false minimum. (a) Electron micrograph of the negatively stained sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation for the final reconstruction obtained when the initial model was a spherical shell. (c) The uniform, spherical shell with an outer radius of 9 nm and an inner radius of 5.4 nm that

was used as the initial model for refinement. (d) The refined structure. (e) A homology model based on a PDB of the *A. aeolicus* enzyme structure (1HQK) docked into the semitransparent EM density. A pentameric subunit (shown as blue ribbons) was docked into the EM density map, and the positions of all other pentamers (shown as turquoise ribbons) were then generated in accord with the icosahedral symmetry. (f) The initial model obtained by truncating the resolution of the X-ray crystal structure (1RVV) for the *B. subtilis* enzyme at 1.9 nm. (g) Refined EM density map obtained when the low-resolution version of the X-ray crystal structure in (f) is used as an initial model. The rotation angle of pentamers in the EM map is the same as it is when a uniform sphere is used as the initial model.

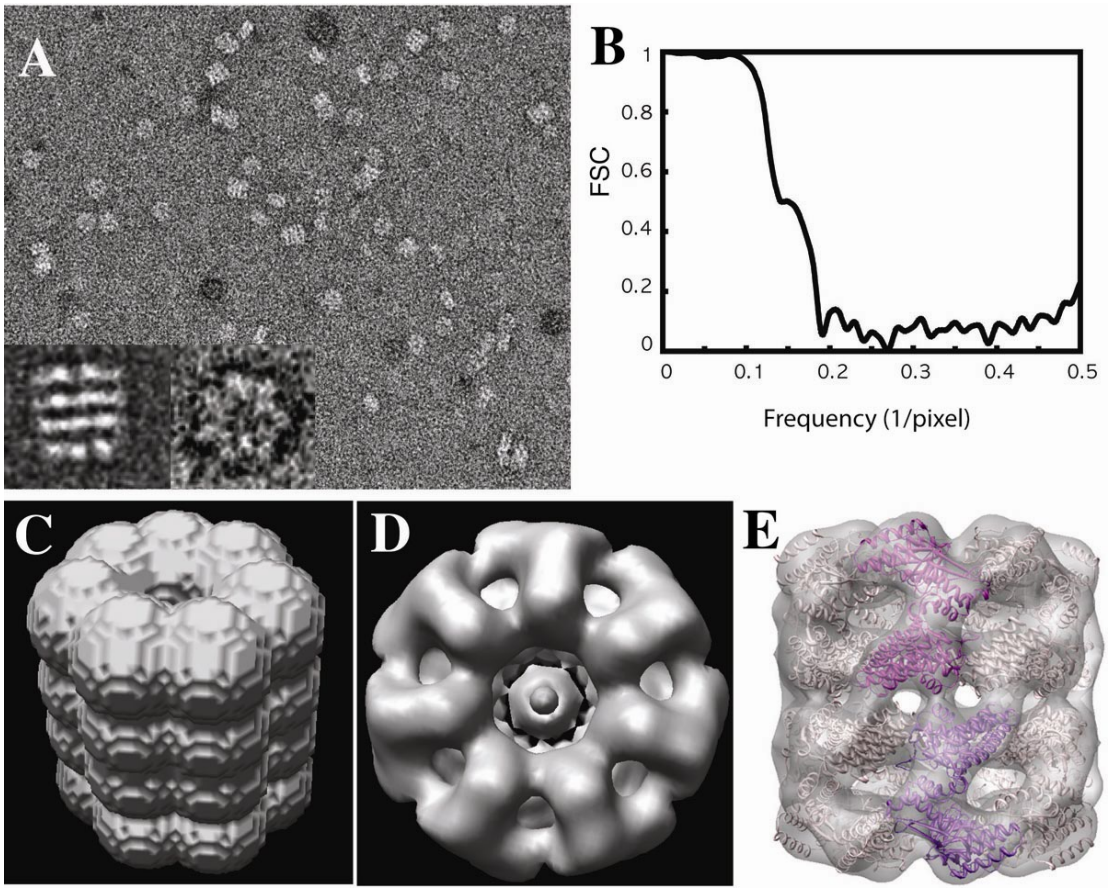


Figure S30. GroEL EM image, refinement result, and docking of a pseudo-atomic model structure. 3029 particles were used for the refinement. D7 symmetry was imposed during the refinement. The standard FSC curve falls to 0.5 at a resolution of 21 Å after two rounds of refinement, each of which consisted of 15 cycles. (a) Electron micrograph of the negatively stained sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation curve. (c) An intuitive starting model was built from 28 stacked spheres, based on prior knowledge of the structure of GroEL. (d) A top view of the refined structure. (e) A homology model, built from a PDB template 1KP8, was docked into the semitransparent EM density. Two ribbon diagrams of the pseudo-atomic model are shown as purple (bottom ring) and magenta (top ring) ribbons respectively, while all others are shown as pink ribbons.

.

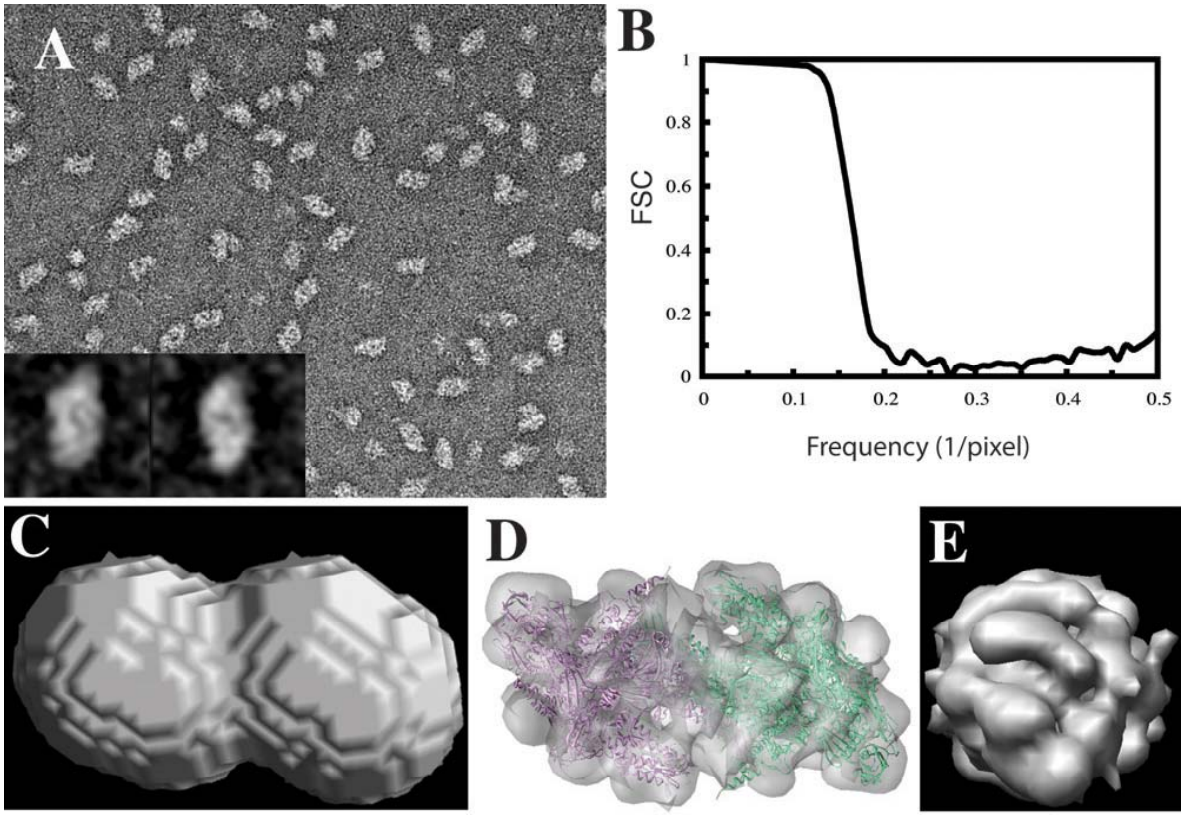


Figure S31. RNA polymerase EM image, refinement result, and docking of a pseudo-atomic model structure. 21420 particles were used for the refinement. C2 symmetry was imposed during the refinement. The standard FSC curve falls to 0.5 at a resolution of 19 Å after two rounds of refinement, each of which consisted of 15 cycles. (a) Electron micrograph of the negatively stained sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation curve. (c) An intuitive starting model was built from 4 partially overlapping spheres in a way that resembles the outline of some of the particle views seen in the original micrographs. (d) A side view of the refined structure in which a X-ray crystal structure of the PDB accession number 2PPB was docked into the semitransparent EM density. Two monomers are shown in pink and green ribbons. (e) Another view, 90° rotated around Y-axis relative to (d).

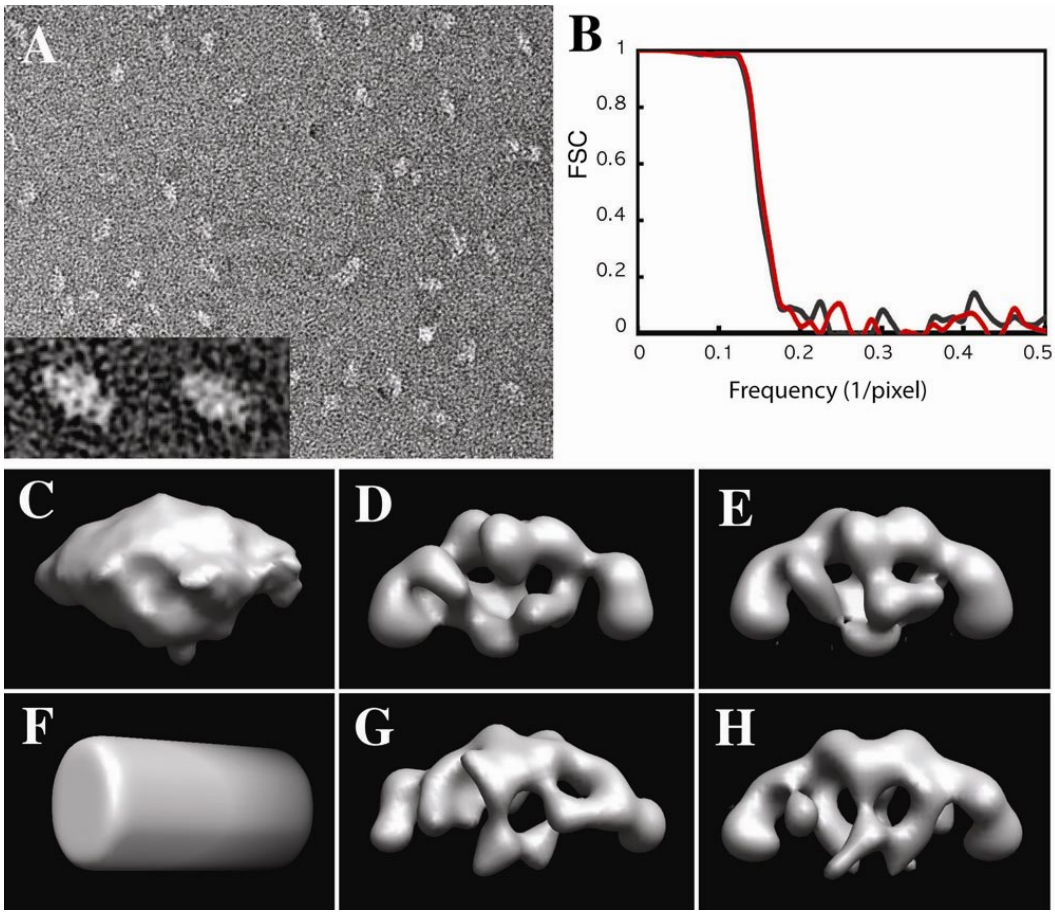


Figure S32. PEP synthase EM image and refinement result. 2105 particles were used for the refinement, and results are compared for an initial model built by the RCT method and for a simple, uniform cylinder. Although the internal details of the refined reconstructions vary depending upon the initial model, the resulting overall sizes and shapes of the reconstructed particles are very similar. When the RCT model is used and C2 symmetry is imposed, the standard FSC curve (shown in black) falls to 0.5 at a resolution of 29 Å after two rounds of refinement, each of which consisted of 12 cycles. (a) Electron micrograph of the negatively stained sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation curve for refinement based on the RCT model, with C2 symmetry imposed. (c) The initial model obtained by the RCT method, using a single class of particles from images of untilted specimens. The initial model obtained by merging volumes from 10 classes of particles was very similar, and thus these additional results are not shown here. (d) The refined structure obtained without imposing symmetry, when using the initial model in (c). The only available X-ray structure (PDB accession code 2ols) of a phosphoenolpyruvate synthase is that of a protein ~2/3 the length of the *Dv*H enzyme, and thus no attempt is made to dock that structure into the EM density map. (e) The refined structure obtained when enforcing C2 symmetry, again using the initial model in (c). Note that there are only small differences between (d) and (e), which means that the refined structure obtained without enforcing symmetry already shows clear C2 symmetry. (f) The uniform cylinder used as an initial

model. (g) The refined structure obtained without imposing symmetry, when using the initial model in (f). (h) The refined structure (FSC curve shown in red) obtained when enforcing C2 symmetry, again using the initial model in (f). The shapes and sizes of the refined structures remain similar to those of the structure in (e) although the internal details of (g) and (h) differ to some extent from one another as well as from (e).

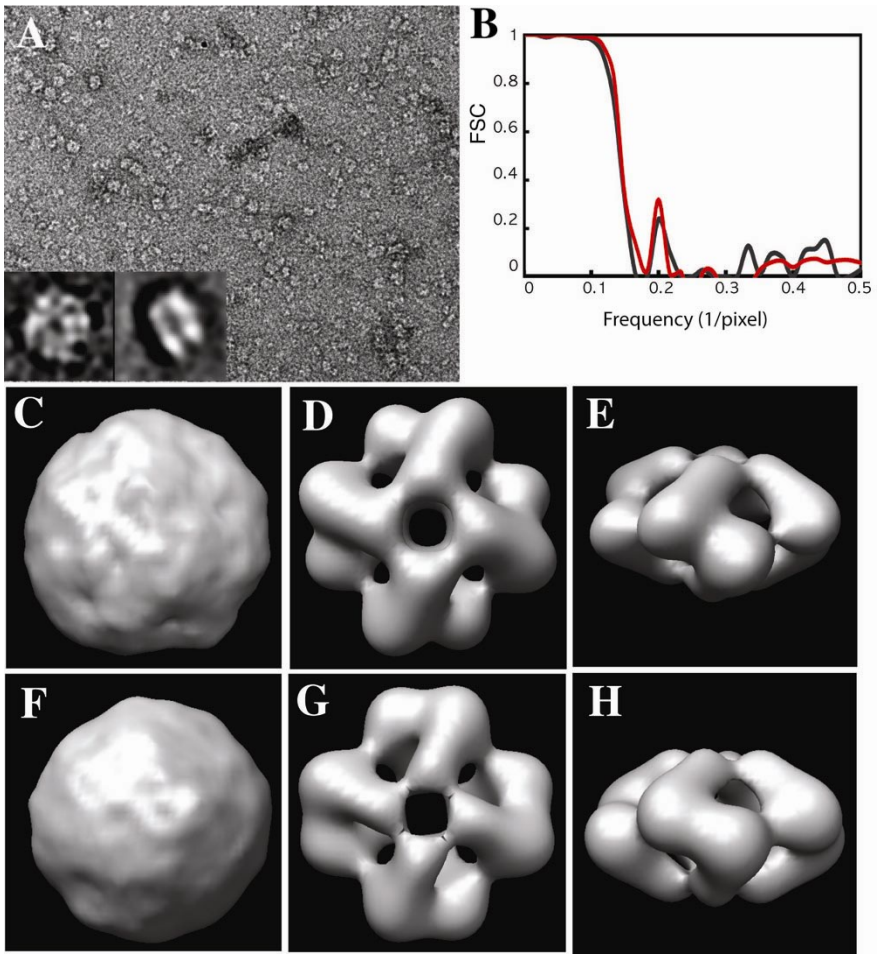


Figure S33. Putative protein EM image and refinement result. 2142 particles were used for the refinement, and results are compared for two initial models built by the RCT method, one (shown in (c)) that was produced from a single class of particles in the images of untilted specimens and the other (shown in (f)) which was obtained by merging volumes from 7 classes according to the guidance given on the SPIDER web site for "Pawel's methods" [http://www.wadsworth.org/spider_doc/spider/docs/techs/rancon/recn.html]. D4 symmetry was imposed during refinement. When using the initial model shown in (c), the standard FSC curve (shown in black) falls to 0.5 at a resolution of 23 Å after two rounds of refinement, each of which consisted of 15 cycles. (a) Electron micrograph of the negatively stained sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation curve for the refinement that that began with the initial model in (c). (c) The initial model obtained by the RCT method, using a single class of particles from images of untilted specimens. (d) Top view of the refined structure. (e) Side view of the refined structure, rotated 90 ° around the X-axis relative to (d). (f) The initial model obtained by the RCT method in which volumes obtained for 7 classes were merged. (g) Top view of the refined structure using the initial model in (f). The FSC curve for this structure is shown in red. (h) Side view of the refined structure shown in (g), rotated 90 ° around the X-axis.

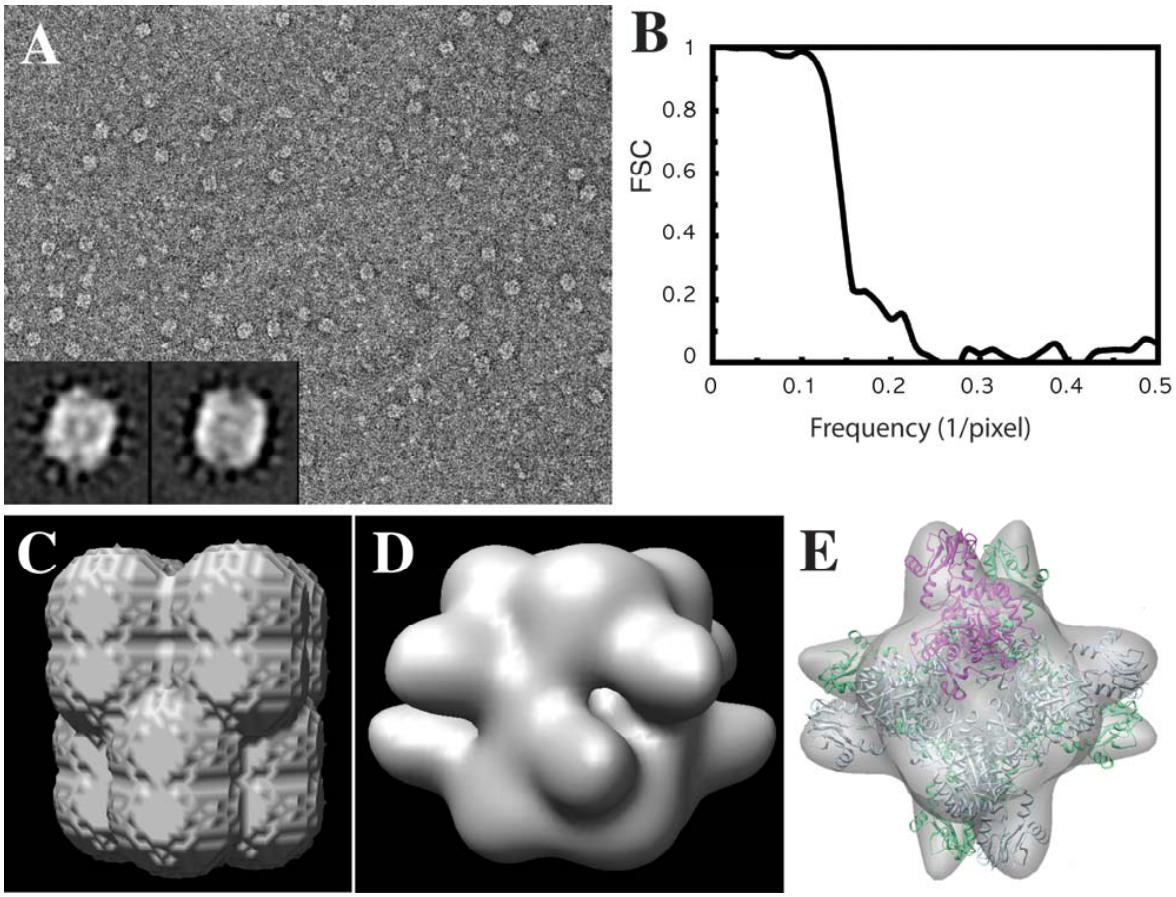


Figure S34. IMP dehydrogenase EM data, refinement result, and docking of a pseudo-atomic model structure. 5814 particles were used for the refinement C4 symmetry was imposed during the refinement. The standard FSC curve falls to 0.5 at a resolution of 19 Å after two rounds of refinement, each of which consisted of 15 cycles. (a) Electron micrograph of the negatively stained IMP dehydrogenase sample. Two examples of class averages are shown at the bottom left. (b) Fourier shell correlation curve. (c) An intuitive starting model was built from 16 partially overlapping spheres in a way that reflected the number of subunits expected from the estimated particle weight (and the known molecular weight of the protein monomer) and some of the views seen in the class averages. (d) A side view of the refined structure. (e) A top view in which a homology model, built from a PDB template 1ZFJ, was docked into the semitransparent EM density. Tetramer at the bottom is shown as light green ribbons and the one at the top as light blue ribbons with a monomer shown as magenta ribbon.

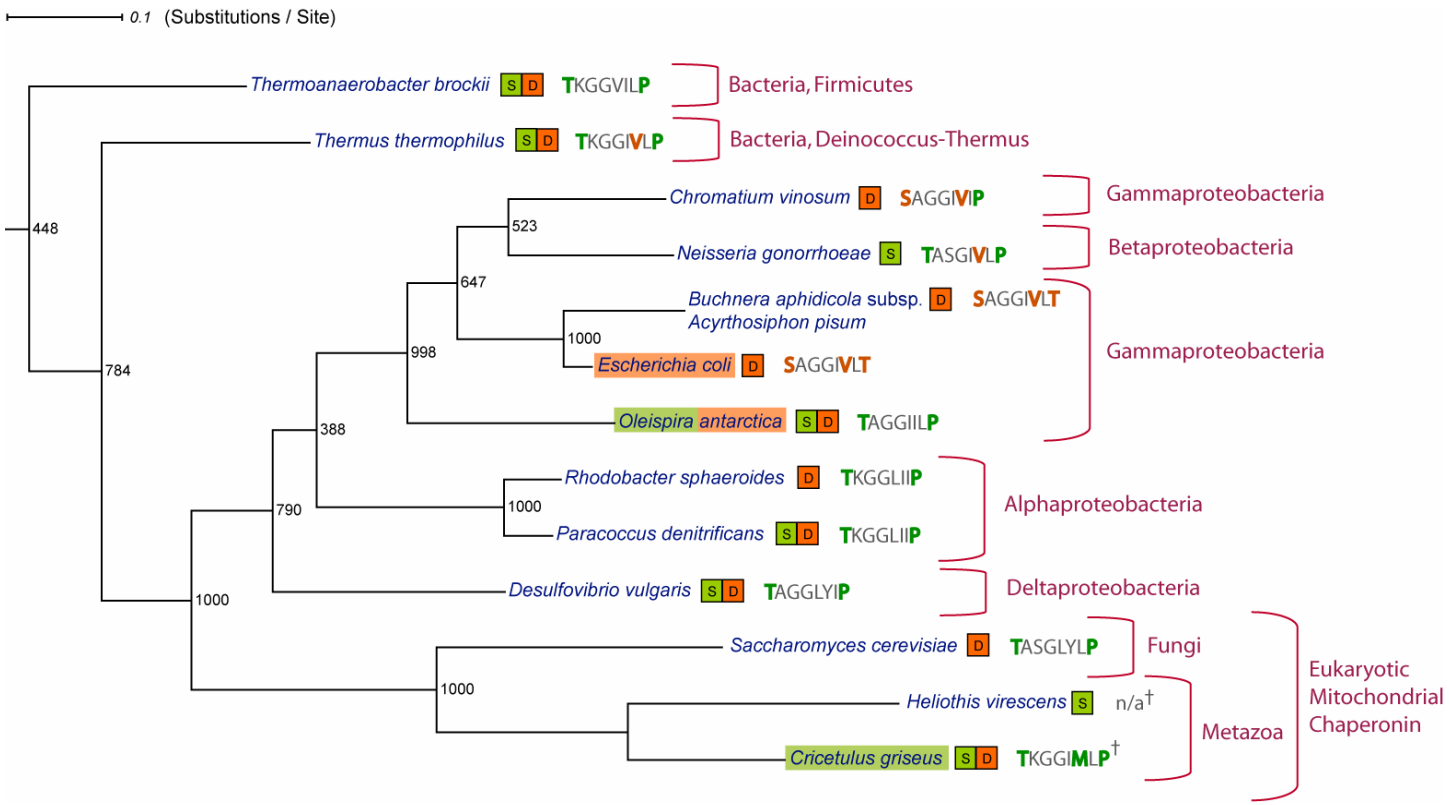


Figure S35. Phylogenetic tree of GroEL and Hsp60 proteins from Table S2, constructed using PhyML (default parameters). Bootstrap values, out of 1000, are displayed. Complexes purified as a double ring are labeled with an orange square box marked with 'D'; those purified as a single ring are labeled with a green square box marked with 'S'. Species in which a double ring is required for normal function (*E. coli*) are highlighted with an orange background; those for which a single ring is known to be sufficient for function (*Cricetulus griseus* mitochondria) are highlighted with a green background. *Oleispira antarctica* is highlighted in green and orange since its GroEL is functional as double or single ring structure depending on temperature. The binding loop sequence from GroES/Hsp10 is also shown for each protein. This sequence fragment has been demonstrated by mutational analysis to determine the specificity of binding of mammalian Hsp60 to either mammalian Hsp10 or *E.coli* GroES. † - There are currently no available sequences of an Hsp10 ortholog from *Heliothis virescens* or *Cricetulus griseus*. For *Cricetulus griseus* we used Hsp10 binding loop from *Mus musculus* (Hsp60 sequences from *Cricetulus griseus* and *Mus musculus* are 98% identical).

Table S1

Evidence of Polypeptide Identification by Mass Spectrometry

ID#	Polypeptide name	DVU#	UniProt accession#	Gene[1]	MW Theor. (kDa)	ID Category[2]	MS/MS Experiment[3]	MS/MS Score[4] / Pept.#[5]	MS/MS Seq. Cov.[6] (%)	PMF Score[7]	MS Cal.[8]	Mass error[9]	PMF Seq. Cov. (%)	# Mol. Ions Matched[11a] / Unmatched[11b]	Average Expression[10]	Percentile Rank[12b]
I	Predicted phospho-2-dehydro-3-deoxyheptonate aldolase[13]	0460	Q72EV8	DVU_0460	28.4	PMF+ [ID1]	No sep / Manual			22.1	Int.	-0.2±4.5	34	9 / 44	5.8	87
II	Putative uncharacterized protein	0631	Q72EE7	DVU_0631	55.7	PMF+ [ID2+]	No sep / Manual			81.5	Int.	7.2±9.7	29	13 / 36	5.1	75
III	Putative uncharacterized protein	0671	Q72EA7	DVU_0671	59.2	PMF+ [ID2+]	No sep / Manual			284.5	Ext.	54.3±10.3	67	34 / 56	4.9	72
IV	Hemolysin-type calcium-binding repeat protein	1012	Q72DB7	DVU_1012	316.4	PMF+ [ID2+]	No sep / Manual			142.5	Int.	6.6±8.0	28	62 / 131	6.6	94
V	Inosine-5`-monophosphate dehydrogenase	1044	Q72D85	guaB	52.2	PMF+ [ID2+]	No sep / Manual			459.0	Int.	-1.9±5.7	55	34 / 35	5.7	86
VI	Riboflavin synthase alpha chain	1200	Q72CT3	ribE	23.6	[ID2+]	No sep / Auto	4.0 / 2	9.5						5.4	81
VI	Riboflavin synthase beta chain	1198	P61940	ribH	16.6	PMF+ [ID2+]	No sep / Auto	14.0 / 7	55.5	61.1	Int.	4.9±5.3	91	17 / 82	5.6	84
VII	DNA-directed RNA polymerase subunit alpha	1329	Q72CF4	rpoA	38.9	PMF+ [ID2+]	LC MALDI	20.1 / 10	42.9	432.3	Int.	7.2±7.0	65	34 / 36	6.6	93
VII	DNA-directed RNA polymerase subunit beta	2928	Q727C7	rpoB	153.2	PMF+ [ID2+]	LC MALDI	51.0 / 24	32.1	331.1	Int.	-1.3±8.3	62	86 / 97[14]	5.8	87
VII	DNA-directed RNA polymerase subunit beta'	2929	Q727C6	rpoC	154.8	PMF+ [ID2+]	LC MALDI	44.9 / 22	31.2	270.4	Int.	-3.9±12.9	58	98 / 97[14]	5.7	85
VII	DNA-directed RNA polymerase subunit omega[13]	3242	Q725M7	rpoZ	8.9	PMF+ [ID2+]	No sep / Manual			*14.25*[15]	Ext.	5.9±9.5	71	7 / 73	6.7	94
VII	N utilization substance protein A	0510	Q72EQ9	nusA	47.8	[ID2+]	LC MALDI	64.0 / 30	69.9						5.3	79
VIII	Ketol-acid reductoisomerase	1378	Q72CA6	ilvC	36.1	PMF+ [ID2+]	No sep / Manual			84.7	Int.	-4.7±6.3	50	19 / 50	7.3	97
IX	Phosphoenolpyruvate synthase[13]	1833	Q72B07	ppsA	132.6	PMF	not performed			720.0	Ext.	-65.2±9.3	50	54 / 42	5.3	80
X	Pyruvate carboxylase	1834	Q72B06	pyc	136.4	PMF+ [ID2+]	No sep / Manual			628.9	Int.	3.2±7.8	42	49 / 40	6.7	94
XI	60 kDa chaperonin	1976	Q72AL6	groL	58.4	PMF+ [ID2+]	No sep / Manual			290.3	Int.	3.2±6.2	59	37 / 53	7.7	98
XII	Phosphorylase (glycogen phosphorylase family)	2349	Q729K1	DVU_2349	97.4	PMF+ [ID2+]	No sep / Manual			354.0	Ext.	2.3±7.7	53	52 / 108	5.0	74
XIII	Alcohol dehydrogenase, iron containing	2405	Q729E6	DVU_2405	41.8	PMF+ [ID2+]	No sep / Manual			65.3	Ext.	44.9±9.2	55	29 / 148	8.3	99
XIV	Pyruvate-ferredoxin oxidoreductase	3025	Q726T1	poR	131.5	PMF+ [ID2+]	No sep / Manual			247.5	Ext.	-16.8±8.1	48	70 / 134	7.1	96
XV	Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase[13]	3319	Q725V6	putA	111.9	PMF	not performed			932.0	Ext.	-40.8±11.8	61	62 / 33	5.2	78
30S	30S ribosomal protein S2	0874	Q72DQ5	rpsB	28.47	[ID2+]	LC MALDI	12.4 / 5	30.7						6.8	94

30S	30S ribosomal protein S3	1309	Q72CH4	rpsC	24.41	[ID2+]	LC MALDI	25.0/ 11	51.9	6.8	95
30S	30S ribosomal protein S4	1328	Q72CF5	rpsD	23.90	[ID2+]	LC MALDI	32.0 / 15	56.7	6.8	94
30S	30S ribosomal protein S5	1320	Q72CG3	rpsE	17.14	[ID2+]	LC MALDI	20.0 / 8	63.2	6.8	95
30S	30S ribosomal protein S6	0956	Q72DH3	rpsF	11.36	[ID2+]	LC MALDI	10.0 / 4	58.4	7.6	97
30S	30S ribosomal protein S7	1299	Q72CI4	rpsG	17.80	[ID2+]	LC MALDI	20.1 / 7	53.2	6.8	94
30S	30S ribosomal protein S8	1317	Q72CG6	rpsH	13.87	[ID2+]	LC MALDI	16.0 / 8	63.5	6.4	92
30S	30S ribosomal protein S9	2519	Q728T3	rpsI	14.90	[ID2+]	LC MALDI	18.1 / 9	52.3	7.2	96
30S	30S ribosomal protein S10	1302	Q72CI1	rpsJ	11.75	[ID2+]	LC MALDI	10.0 / 4	59.0	7.5	97
30S	30S ibosomal protein S11	1327	Q72CF6	rpsK	13.99	[ID2+]	LC MALDI	16.0 / 6	65.1	7.2	96
30S	30S ribosomal protein S12	1298	Q72CI5	rpsL	13.91	[ID2+]	LC MALDI	6.1 / 3	38.2	7.9	98
30S	30S ribosomal protein S13	1326	Q72CF7	rpsM	13.85	[ID2+]	LC MALDI	18.1 / 9	48.4	7.3	97
30S	30S ribosomal protein S15	0504	Q72ER5	rpsO	10.23	[ID2+]	LC MALDI	8.0 / 4	39.3	7.4	97
30S	30S ribosomal protein S16	0839	P62229	rpsP	9.11	[ID2+]	LC MALDI	6.2 / 3	44.3	8.0	98
30S	30S ribosomal protein S17	1312	Q72CH1	rpsQ	10.30	[ID2+]	LC MALDI	14.1 / 7	71.6	6.8	94
30S	30S ribosomal protein S18	0957	Q72DH2	rpsR	10.28	[ID2+]	LC MALDI	6.0 / 3	32.2	7.5	97
30S	30S ribosomal protein S19	1307	Q72CH6	rpsS	10.49	[ID2+]	LC MALDI	18.0 / 9	61.3	6.8	94
30S	30S ribosomal protein S20	1896	Q72AU4	rpsT	9.65	[ID2+]	LC MALDI	4.0 / 2	37.9	8.9	99
30S	30S ribosomal protein S21	1792	Q72B46	rpsU	8.37	[ID2+]	LC MALDI	8.0 / 4	41.4	9.0	99
50S	50S ribosomal protein L1	2925	Q727D0	rplA	24.78	[ID2+]	LC MALDI	10.2 / 4	29.8	8.0	98
50S	50S ribosomal protein L2	1306	Q72CH7	rplB	30.15	[ID2+]	LC MALDI	22.7 / 10	42.0	6.8	94
50S	50S ribosomal protein L3	1303	Q72CI0	rplC	22.35	[ID2+]	LC MALDI	14.0 / 7	41.1	6.9	95
50S	50S ribosomal protein L4	1304	Q72CH9	rplD	22.60	[ID2+]	LC MALDI	18.0 / 8	55.3	7.3	96
50S	50S ribosomal protein L5	1315	Q72CG8	rplE	20.21	[ID2+]	LC MALDI	22.0 / 9	63.7	5.9	88
50S	50S ribosomal protein L6	1318	Q72CG5	rplF	19.13	[ID2+]	LC MALDI	14.0 / 7	48.6	5.8	87
50S	50S ribosomal protein L9	0958	Q72DH1	rplI	17.95	[ID2+]	LC MALDI	12.2 / 5	64.1	6.8	95
50S	50S ribosomal protein L10	2926	Q727C9	rplJ	18.74	[ID2+]	LC MALDI	8.0 / 3	33.5	7.6	97
50S	50S ribosomal protein L11	2924	P62433	rplK	14.87	[ID2+]	LC MALDI	4.0 / 2	18.6	7.3	97
50S	50S ribosomal protein L7/L12	2927	Q727C8	rplL	13.32	[ID2+]	LC MALDI	10.1 / 4	44.1	6.3	92
50S	50S ribosomal protein L13	2518	Q728T4	rplM	16.45	[ID2+]	LC MALDI	14.0 / 6	47.2	8.1	98
50S	50S ribosomal protein L14	1313	Q72CH0	rplN	13.28	[ID2+]	LC MALDI	10.2 / 4	44.3	6.7	94
50S	50S ribosomal protein L15	1322	Q72CG1	rplO	15.83	[ID2+]	LC MALDI	10.0 / 5	43.9	6.3	91
50S	50S ribosomal protein L16	1310	Q72CH3	rplP	15.06	[ID2+]	LC MALDI	4.3 / 2	19.6	7.1	96

50S	50S ribosomal protein L17	1330	Q72CF3	rplQ	14.58	[ID2+]	LC MALDI	12.3 / 5	33.8	7.0	95
50S	50S ribosomal protein L18	1319	Q72CG4	rplR	13.27	[ID2+]	LC MALDI	4.6 / 2	11.8	6.2	91
50S	50S ribosomal protein L19	0835	Q72DU4	rplS	13.50	[ID2+]	LC MALDI	8.0 / 4	32.2	7.8	98
50S	50S ribosomal protein L20	2535	Q728R8	rplT	13.66	[ID2+]	LC MALDI	4.0 / 2	19.7	7.2	96
50S	50S ribosomal protein L21[13]	0927	Q72DK2	rplU	11.29	[ID2+]	LC MALDI	1.2 / 2	9.8	8.2	99
50S	50S ribosomal protein L22	1308	Q72CH5	rplV	12.48	[ID2+]	LC MALDI	10.0 / 4	57.1	6.6	94
50S	50S ribosomal protein L23	1305	Q72CH8	rplW	12.69	[ID2+]	LC MALDI	6.0 / 3	32.1	6.3	91
50S	50S ribosomal protein L24 [13]	1314	Q72CG9	rplX	12.33	[ID1]	LC MALDI	2.0 / 1	12.1	5.9	88
50S	50S ribosomal protein L25	1574	Q72BR0	rplY	21.54	[ID2+]	LC MALDI	27.0 / 13	77.4	7.3	96
50S	50S ribosomal protein L27 [13]	0928	Q72DK1	rpmA	9.65	[ID1]	LC MALDI	2.0 / 1	16.1	7.1	96
50S	50S ribosomal protein L28	1211	Q72CS2	rpmB	7.64	[ID2+]	LC MALDI	4.0 / 2	34.8	6.8	94
50S	50S ribosomal protein L29	1311	Q72CH2	rpmC	7.11	[ID2+]	LC MALDI	8.0 / 3	52.5	6.7	94
50S	50S ribosomal protein L35	2536	Q728R7	rpmI	7.50	[ID2+]	LC MALDI	4.0 / 2	42.6	8.4	99
30S	*ribosomal protein S1, putative [16]*	*1469*	*Q72C15*	*DVU_1469*	*51.69*	*not detected*				*4.8*	*68*
30S	*30S ribosomal protein S1 [16]*	*3150*	*Q726F8*	*rpsA*	*64.74*	*not detected*				*6.2*	*90*
30S	*30S ribosomal protein S14 type Z[16]*	*1316*	*Q72CG7*	*rpsN*	*7.12*	*not detected*				*6.5*	*93*
50S	*50S ribosomal protein L30 [16]*	*1321*	*Q72CG2*	*rpmD*	*6.18*	*not detected*				*6.3*	*91*
50S	*50S ribosomal protein L31 [16]*	*2912*	*Q727E3*	*rpmE*	*8.14*	*not detected*				*7.9*	*98*
50S	*50S ribosomal protein L32 [16]*	*1209*	*Q72CS4*	*rpmF*	*6.55*	*not detected*				*7.9*	*98*
50S	*50S ribosomal protein L33 [16]*	*2921*	*Q727D4*	*rpmG*	*5.94*	*not detected*				*5.5*	*83*
50S	*50S ribosomal protein L34 [16]*	*1074*	*Q72D55*	*rpmH*	*5.20*	*not detected*				*8.2*	*99*
50S	*50S ribosomal protein L36 [16]*	*1325*	*Q72CF8*	*rpmJ*	*4.40*	*not detected*				*7.7*	*98*

Legend

[1]OrderedLocusName (*e.g.*, DVU_0671) is provided for genes with no name in UniProt.
[2] PMF = Peptide Mass Fingerprinting; [ID1] and [ID2+] = Identification of a single or at least two peptides, respectively, by MS/MS.
[3] Type of MS/MS experiment. 'No sep' = precursor analyzed within whole digest; 'Manual' = manual interpretation of MS/MS spectrum; 'Auto' = Protein Pilot search-engine driven analysis of MS/MS data; 'LC MALDI' = off-line peptide separation by reversed phase HPLC followed by MALDI MS/MS analysis with automated MS/MS data analysis using search Protein Pilot search engine.
[4] Protein Pilot 'unused score' is a sum of scores of individual peptides assigned to the highest ranking polypeptide within a subset of homologous entries in protein database.
[5] Number of peptides that were matched by Protein Pilot with the highest score of 2.0 that corresponds to confidence of 99%, for all but 50S ribosomal protein L21. L21 was identified on the basis of the manually interpreted MS/MS spectra of two overlapping peptides that were assigned as low confidence hits (89% and 40%) - spectra are enclosed. The MS/MS spectra of all reported peptides were manually examined.
[6] Protein sequence coverage calculated using peptides identified with a confidence of 99% only.
[7] Peptide mass fingerprint (PMF) score provided by the Aldente search engine (version 11/02/2008). Aldente statistic threshold for the normalized score was 17.1: proteins with a score greater than this threshold are considered statistically significant hits.
[8] Type of calibration of MS spectra that were used for PMF analysis. All spectra were initially calibrated using AB 4800 Plate Model and Default MS Calibration Update software. Some of the spectra were subsequently internally recalibrated prior to generation of peak lists for PMF analysis, employing two trypsin autolysis molecular ions: m/z = 842.51 and 2211.1046.
[9] Average and standard deviation of differences between experimental and theoretical m/z values of peptides that were matched to the proposed polypeptide sequences, expressed in parts-per-million (ppm).

Percentage of a total number of amino acids within the polypeptide sequence that are represented in peptides matched to this sequence.

[11a] and [11b] Number of molecular ions within the MS spectrum that were matched and not matched to the proposed polypeptide sequence, respectively. The ions that were matched to minor species/contaminants are included in the 'unmatched' category.

[12a] Average expression of each polypeptide based upon (3, 16, 52-57), expressed in the logarithmic scale from 1-10, where 10 is the highest; [12b] corresponding percentile rank.

[13] MS and/or MS/MS spectra of polypeptides identified either by low confidence PMF, single MS/MS spectrum or PMF-only analysis are enclosed in the Supplemental Material.

[14] DVU2728 and DVU2829 were present in a single band: the number of unmatched ions corresponds to those ions that did not match either polypeptide.

[15] PMF identification score is below the statistically significant threshold due to the presence of large number of contaminants (mainly human keratin). MS/MS confirmed the identity of this polypeptide.

[16] Ribosomal proteins not detected by MS. No targeted MS analysis was performed. A failure of detection can be caused by many different reasons and should not be interpreted as a proof of the absence of a polypeptide in the sample.

Table S2: Quaternary structure of bacterial GroEL and mitochondrial Hsp60. Quaternary structure information is supplemented with purification conditions as it appears in the corresponding papers. Single/double ring activity column shows whether it has been demonstrated that GroEL/Hsp60 is functional as a single or/and as a double ring. For example, *Saccharomyces cerevisiae* Hsp60 was purified as a double-ring protein and the purified proteins were shown to be functionally active; however, Hsp60 quaternary structure is not known during its activity.

Bacterial species:	**Quaternary Structure after purification**	**Single ring active**	**Double ring active**
Buchnera aphidicola subsp. *Acyrthosiphon pisum*	Double ring (room temperature, 10 mM potassium phospate buffer (pH 6.9), 200 mM Na_2SO_4) (*58*)	Unknown	Unknown
Chromatium vinosum	Double ring (4°C, 50 mM MOPS, 100 mM NaCl and 1 mM 2-mercaptoethanol (pH 7.5)) (*59*)	Unknown	Unknown
Desulfovibrio vulgaris	Single ring (4°C, 10 mM HEPES, pH 7.6, 2 mM DTT and 0.01% NP40) (current paper) Double ring (4°C, 10 mM HEPES, pH 7.6, 2 mM DTT, 0.01% NP40, 0.5 mM ATP and 7 mM $MgCl_2$) (current paper)	Unknown	Unknown
Escherichia coli	Double ring (4°C, 10 mM HEPES, pH 7.6, 2 mM DTT, 0.01% NP40) (current paper) Double ring (50 mM Tris-HCl (pH 7.4), 400 mM NaCl, 1 mM EDTA, 1 mM DTT) (*60*)	No (*38*)	Yes (*38*)
Neisseria gonorrhoeae	Single ring (1 mM phosphate buffer, pH 7.4, 450 mM NaCl, 1 mM EDTA) (*61*)	Unknown	Unknown
Oleispira antarctica RB8	Single ring (4–10°C, 50 mM MOPS, pH 7.0, 50 mM $MgCl_2$, 50 mM KCl 2 mM ATP, ADP, ATPγS, AMP-PNP, CTP, UTP or GTP) (*45*) Double ring (>10°C, 50 mM MOPS, pH 7.0, 50 mM $MgCl_2$, 50 mM KCl 2 mM ATP, ADP, ATPγS, AMP-PNP, CTP, UTP or GTP) (*45*)	Yes (at temperatures 4-10°C) (*45*)	Yes (at temperatures >10°C) (*45*)
Paracoccus denitrificans	Mostly double rings, with some proportion of single rings (25 mM Tris-HCl, pH 7.5, 100 mM Na_2SO_4, 5 mM $MgCl_2$, 100 μM DTT) (*62, 63*)	Unknown	Unknown
Rhodobacter sphaeroides	Double ring (50 mM TES (pH 7.0), 10% (v/v) ethylene glycol, 20 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 50 mM $NaHCO_3$, 10 mM EDTA, 5 mM 2-mercaptoethanol) (*64*)	Unknown	Unknown

Thermoanaerobac-ter brockii	Single ring (20% 2-propanol, 25 mM Tris/HCl, pH 8.0) (*34*) Single ring (37°C, 20 mM Tris-HCl, pH 7.7, 5-30 mM $MgCl_2$, 50 mM KCl, 100 mM NaCl, plus either 2mM ATP or 600 nM $GroES_7$) (*35*) ~70% double rings, $GroEL_{14}$-$GroES_7$ (37°C, 20 mM Tris-HCl, pH 7.7, 5 mM $MgCl_2$, 50 mM KCl, 100 mM NaCl, 2mM ATP, 600 nM $GroES_7$) (*35*) ~95% double rings, $GroEL_{14}$-$GroES_7$ (37°C , 20 mM Tris-HCl, pH 7.7, *30* mM $MgCl_2$, 50 mM KCl, 100 mM NaCl, 2mM ATP, 600 nM $GroES_7$) (*35*)	Unknown	Unknown
Thermus thermophilus	Double ring (25 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 100 µM dithiothreitol, 100 mM Na_2SO_4) (*36*), Double ring (50 mM Tris-HCl, pH 7.51, 200 mM NaCl, plus either 90 mM KCl or 1.5 Mg-ATP) (*37*) Mostly single rings, with some proportion of double rings (50 mM Tris-HCl, pH 7.51, 200 mM NaCl, 90 mM KCl, 1.5 Mg-ATP) (*37*)	Unknown	Unknown
Mitochondrial:			
Heliothis virescens (a testis-specific mitochondrial isoform)	Single ring (10 mM HEPES KOH (pH 7.4), 1 mM EDTA, 1 mM DTT, 50 mM NaCl, 0.25% NP-40) (*41*)	Unknown	Unknown
Cricetulus griseus	Single ring (0.1 M Tris-HCl (pH 7.7), 10 mM KCl, 10 mM $MgCl_2$, and 0.01% Tween 20) (*31*), Double ring (25°C , 50 mM Tris/HCl, pH 7.4, 50 mM KCl, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM ATP, 90 µM Hsp10 and 30 µM Hsp60) (*32*), ~80% single rings, ~10% double rings and ~10% monomers (25°C, 50 mM Tris/HCl, pH 7.4, 50 mM KCl, 100 mM NaCl, 10 mM $MgCl_2$ and 2 mM ATP) (*32*), ~80% single rings and ~20% monomers (25°C, 50 mM Tris/HCl, pH 7.4, 50 mM KCl, 100 mM NaCl, and 10 mM $MgCl_2$) (*32*)	Yes (*31*)	Unknown
Saccharomyces cerevisiae	Double ring (50 mM HEPES-NaOH (pH 7.4), 50 mM NaCl) (*65*)	Unknown	Unknown

Table S3: Distribution of the tripeptide sequence motif from the GroES mobile loop in the Pfam species tree for the GroES protein family. There is currently no available sequences of an Hsp10 ortholog from *Cricetulus griseus*, therefore we used Hsp10 binding loop from *Mus musculus* (Hsp60 sequences from *Cricetulus griseus* and *Mus musculus* are 98% identical).

The *E. coli* tripeptide motif appears in:

Bacteria -> Proteobacteria -> Gammaproteobacteria ->
Alteromonadales
Vibrionales
Pasteurellaceae
Enterobacteriales (*E. coli*'s clade)
Xanthomonadaceae
Bacteria -> Firmicutes -> Lactobacillales -> Streptococcaceae -> Lactococcus

The *Mus musculus* tripeptide motif appears in:

Eukaryota -> Metazoa ->
Arthropoda
Chordata (mammalian clade)
Nematoda

The *Desulfovibrio vulgaris* tripeptide motif appears in:

Eukaryota->
Fungi
Mycetozoa
Bacteria->
Cyanobacteria
Chloroflexi
Spirochaetales
Nitrospirae
Deltaproteobacteria (*D. vulgaris*' clade)
Actinobacteria
Deinococcus-Thermus
Bacteroidetes
Chlorobi
Archaea -> Methanomicrobia

References

1. Mastronarde, D. N. (2005) *Journal of Structural Biology* **152,** 36-51.
2. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. (1996) *Journal of Structural Biology* **116,** 71-76.
3. Mukhopadhyay, A., He, Z. L., Alm, E. J., Arkin, A. P., Baidoo, E. E., Borglin, S. C., Chen, W. Q., Hazen, T. C., He, Q., Holman, H. Y., Huang, K., Huang, R., Joyner, D. C., Katz, N., Keller, M., Oeller, P., Redding, A., Sun, J., Wall, J., Wei, J., Yang, Z. M., Yen, H. C., Zhou, J. Z. & Keasling, J. D. (2006) *Journal of Bacteriology* **188,** 4068-4078.
4. Dong, M., Yang, L. L., Williams, K., Fisher, S. J., Hall, S. C., Biggin, M. D., Jin, J. & Witkowska, H. E. (2008) *Journal of Proteome Research* **7,** 1836-1849.
5. Garczarek, F., Dong, M., Typke, D., Witkowska, H. E., Hazen, T. C., Nogales, E., Biggin, M. D. & Glaeser, R. M. (2007) *Journal of Structural Biology* **159,** 9-18.
6. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M. & Cate, J. H. D. (2005) *Science* **310,** 827-834.
7. We are pleased to thank Wen Zhang, Raj Pai, and Jack Dunkles for their help in demonstrating the protocol for E.coli 70S ribosomes and for frequent advice as we adapted the protocol for DvH 70S ribosomes.
8. Jimenez, C., Huang, L., Qiu, Y. & Burlingame, A. L. (1998) in *Protocols in Protein Science*, eds. Coligan, J., Dunn, B., Speicher, D. & Wingfield, P. (John Wiley & Sons, pp. Unit 16.4.1-16.4.5.
9. Papac, D. I., Briggs, J. B., Chin, E. T. & Jones, A. J. S. (1998) *Glycobiology* **8,** 445-454.
10. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C. & Watanabe, C. (1993) *Proceedings of the National Academy of Sciences of the United States of America* **90,** 5011-5015.
11. James, P., Quadroni, M., Carafoli, E. & Gonnet, G. (1993) *Biochemical and Biophysical Research Communications* **195,** 58-64.
12. Mann, M., Hojrup, P. & Roepstroff, P. (1993) *Biological Mass Spectrometry* **22,** 338-345.
13. Pappin, D. J. C., Hojrup, P. & Bleasby, A. J. (1993) *Current Biology* **3,** 327-332.
14. Yates, J. R. I., Speicher, S., Griffin, P. R. & Hunkapiller, T. (1993) *Analytical Biochemistry* **214,** 397-408.
15. Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L. & Burlingame, A. L. (2000) *Anal. Chem.* **72,** 552-558.
16. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003) *Nucleic Acids Research* **31,** 3784-3788.
17. Gasteiger, E. & al., e. (2005) in *The Proteomics Protocols Handbook*, ed. Walker, J. (Humana Press, pp. 571-607.
18. Bradshaw, R. A., Burlingame, A. L., Carr, S. & Aebersold, R. (2006) *Molecular & Cellular Proteomics* **5,** 787-788.
19. Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M. & Schaeffer, D. A. (2007) *Molecular & Cellular Proteomics* **6,** 1638-1655.

20. Typke, D., Nordmeyer, R. A., Jones, A., Lee, J., Avila-Sakar, A., Downing, K. H. & Glaeser, R. M. (2005) *Journal of Structural Biology* **149,** 17-29.
21. Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999) *Journal of Structural Biology* **128,** 82-97.
22. Frank, J. (2006) *Three-dimensional electron microscopy of macromolecular assemblies - visualization of biological molecules in their native state* (Oxford University Press, New York).
23. Sorzano, C. O. S., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J. R., Scheres, S. H. W., Carazo, J. M. & Pascual-Montano, A. (2004) *Journal of Structural Biology* **148,** 194-204.
24. Mindell, J. A. & Grigorieff, N. (2003) *Journal of Structural Biology* **142,** 334-347.
25. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. (1996) *Journal of Structural Biology* **116,** 17-24.
26. Pieper, U., Eswar, N., Davis, F. P., Braberg, H., Madhusudhan, M. S., Rossi, A., Marti-Renom, M., Karchin, R., Webb, B. M., Eramian, D., Shen, M. Y., Kelly, L., Melo, F. & Sali, A. (2006) *Nucleic Acids Research* **34,** D291-D295.
27. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004) *Journal of Computational Chemistry* **25,** 1605-1612.
28. Bacher, A., Baur, R., Eggers, U., Harders, H. D., Otto, M. K. & Schnepple, H. (1980) *Journal of Biological Chemistry* **255,** 632-637.
29. Alberts, B. (2008) *Molecular Biology of the Cell* (Garland Sciences.
30. Hartl, F. U. (1996) *Nature* **381,** 571-580.
31. Viitanen, P. V., Lorimer, G. H., Seetharam, R., Gupta, R. S., Oppenheim, J., Thomas, J. O. & Cowan, N. J. (1992) *Journal of Biological Chemistry* **267,** 695-698.
32. Levy-Rimler, G., Viitanen, P., Weiss, C., Sharkia, R., Greenberg, A., Niv, A., Lustig, A., Delarea, Y. & Azem, A. (2001) *European Journal of Biochemistry* **268,** 3465-3472.
33. Viitanen, P. & al., e. (1998) in *Methods in Enzymology*, Vol. 290, pp. 203-217.
34. Truscott, K. N., Hoj, P. B. & Scopes, R. K. (1994) *European Journal of Biochemistry* **222,** 277-284.
35. Todd, M. J., Walke, S., Lorimer, G., Truscott, K. & Scopes, R. K. (1995) *Biochemistry* **34,** 14932-14941.
36. Taguchi, H., Konishi, J., Ishii, N. & Yoshida, M. (1991) *Journal of Biological Chemistry* **266,** 22411-22418.
37. Ishii, N., Taguchi, H., Sasabe, H. & Yoshida, M. (1995) *Febs Letters* **362,** 121-125.
38. Weber, F., Keppel, F., Georgopoulos, C., Hayer-Hartl, M. K. & Hartl, F. U. (1998) *Nature Structural Biology* **5,** 977-985.
39. Weissman, J. S., Rye, H. S., Fenton, W. A., Beechem, J. M. & Horwich, A. L. (1996) *Cell* **84,** 481-490.
40. Saibil, H. R. & Ranson, N. A. (2002) *Trends in Biochemical Sciences* **27,** 627-632.

41. Miller, S. G., Leclerc, R. F. & Erdos, G. W. (1990) *Journal of Molecular Biology* **214,** 407-422.
42. Nielsen, K. L. & Cowan, N. J. (1998) *Molecular Cell* **2,** 93-99.
43. Nielsen, K. L., McLennan, N., Masters, M. & Cowan, N. J. (1999) *Journal of Bacteriology* **181,** 5871-5875.
44. Sun, Z., Scott, D. J. & Lund, P. A. (2003) *Journal of Molecular Biology* **332,** 715-728.
45. Ferrer, M., Lunsdorf, H., Chernikova, T. N., Yakimov, M., Timmis, K. N. & Golyshin, P. N. (2004) *Molecular Microbiology* **53,** 167-182.
46. Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N. & Timmis, K. N. (2003) *Nature Biotechnology* **21,** 1266-1267.
47. Richardson, A., Schwager, F., Landry, S. J. & Georgopoulos, C. (2001) *Journal of Biological Chemistry* **276,** 4981-4987.
48. Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. L. & Bateman, A. (2006) *Nucleic Acids Research* **34,** D247-D251.
49. Bairoch, A., Bougueleret, L., Altairac, S., Amendolia, V., Auchincloss, A., Puy, G. A., Axelsen, K., Baratin, D., Blatter, M. C., Boeckmann, B., Bollondi, L., Boutet, E., Quintaje, S. B., Breuza, L., Bridge, A., Saux, V. B. L., deCastro, E., Ciampina, L., Coral, D., Coudert, E., Cusin, I., David, F., Delbard, G., Dornevil, D., Duek-Roggli, P., Duvaud, S., Estreicher, A., Famiglietti, L., Farriol-Mathis, N., Ferro, S., Feuermann, M., Gasteiger, E., Gateau, A., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hulo, N., Innocenti, A., James, J., Jain, E., Jimenez, S., Jungo, F., Junker, V., Keller, G., Lachaize, C., Lane-Guermonprez, L., Langendijk-Genevaux, P., Lara, V., Le Mercier, P., Lieberherr, D., Lima, T. D., Mangold, V., Martin, X., Michoud, K., Moinat, M., Morgat, A., Nicolas, M., Paesano, S., Pedruzzi, I., Perret, D., Phan, I., Pilbout, S., Pillet, V., Poux, S., Pozzato, M., Redaschi, N., Reynaud, S., Rivoire, C., Roechert, B., Sapsezian, C., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Vitorello, C., Yip, L., Zuletta, L. F., Apweiler, R., Alam-Faruque, Y., Barrell, D., Bower, L., Browne, P., Chan, W. M., Daugherty, L., Donate, E. S., Eberhardt, R., Fedotov, A., Foulger, R., Frigerio, G., Garavelli, J., Golin, R., Horne, A., et al. (2008) *Nucleic Acids Research* **36,** D190-D195.
50. Guindon, S. & Gascuel, O. (2003) *Systematic Biology* **52,** 696-704.
51. Howe, K., Bateman, A. & Durbin, R. (2002) *Bioinformatics* **18,** 1546-1547.
52. Bender, K. S., Yen, H. C. B., Hemme, C. L., Yang, Z., He, Z., He, Q., Zhou, J., Huang, K. H., Alm, E. J., Hazen, T. C., Arkin, A. P. & Wall, J. D. (2007) *Applied and Environmental Microbiology* **73,** 5389-5400.
53. Mukhopadhyay, A., Redding, A. M., Joachimiak, M. P., Arkin, A. P., Borglin, S. E., Dehal, P. S., Chakraborty, R., Geller, J. T., Hazen, T. C., He, Q., Joyner, D. C., Martin, V. J. J., Wall, J. D., Yang, Z. K., Zhou, J. & Keasling, J. D. (2007) *Journal of Bacteriology* **189,** 5996-6010.

54. He, Q., Huang, K. H., He, Z. L., Alm, E. J., Fields, M. W., Hazen, T. C., Arkin, A. P., Wall, J. D. & Zhou, J. Z. (2006) *Applied and Environmental Microbiology* **72,** 4370-4381.
55. Chhabra, S. R., He, Q., Huang, K. H., Gaucher, S. P., Alm, E. J., He, Z., Hadi, M. Z., Hazen, T. C., Wall, J. D., Zhou, J., Arkin, A. P. & Singh, A. K. (2006) *Journal of Bacteriology* **188,** 1817-1828.
56. Stolyar, S., He, Q., Joachimiak, M. P., He, Z. L., Yang, Z. K., Borglin, S. E., Joyner, D. C., Huang, K., Alm, E., Hazen, T. C., Zhou, J. Z., Wall, J. D., Arkin, A. P. & Stahl, D. A. (2007) *Journal of Bacteriology* **189,** 8944-8952.
57. Clark, M. E., He, Q., He, Z., Huang, K. H., Alm, E. J., Wan, X. F., Hazen, T. C., Arkin, A. P., Wall, J. D., Zhou, J. Z. & Fields, M. W. (2006) *Applied and Environmental Microbiology* **72,** 5578-5588.
58. Morioka, M. & Ishikawa, H. (1998) in *Methods in Enzymology. Molecular Chaperones*, eds. Lorimer, G. H. & Baldwin, T. O. (Academic Press, pp. 181-193.
59. Torres-Ruiz, J. A. & McFadden, B. A. (1998) in *Methods in Enzymology. Molecular Chaperones*, eds. Lorimer, G. H. & Baldwin, T. O. (Academic Press, pp. 147-154.
60. Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994) *Nature* **371,** 578-586.
61. Pannekoek, Y., Vanputten, J. P. M. & Dankert, J. (1992) *Journal of Bacteriology* **174,** 6928-6937.
62. Fukami, T. A., Yohda, M., Taguchi, H., Yoshida, M. & Miki, K. (2001) *Journal of Molecular Biology* **312,** 501-509.
63. Ishii, N., Taguchi, H., Sumi, M. & Yoshida, M. (1992) *Febs Letters* **299,** 169-174.
64. Lee, W. T., Watson, G. W. M. F. & Tabita, F. R. (1998) in *Methods in Enzymology. Molecular Chaperones*, eds. Lorimer, G. H. & Baldwin, T. O. (Academic Press, pp. 154-161.
65. Dubaquie, Y., Schatz, G. & Rospert, S. (1998) in *Methods in Enzymology*, eds. Lorimer, G. & Baldwin, T. (Elsevier, Inc., Vol. 290, pp. 193-202.