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[a,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 Fransisco Sandler-Moore Mass Spectrometry Core Facility, University of California, San Francisco, CA 94143; and [e]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720

Edited by David S. Eisenberg, University of California, Los Angeles, CA, and approved August 17, 2009 (received for review January 3, 2009)

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.

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 that has already been 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 larger 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 that are large enough to be distinguishable from one another within tomographic reconstructions of single cells. A total of 15 different macromolecular complexes with particle weights of at least 400 kDa were isolated by a "tagless" strategy (11), which is "unbiased" in the sense 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. In addition, *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 reasonable to think that reliable templates for a large fraction of a cell's multiprotein complexes could be obtained from information that is 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 the conservation of subunit composition and quaternary structure is not at all guaranteed. On the contrary, we now report that the quaternary structures of such complexes vary to a much greater extent 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 the collection of multiprotein complexes within *Dv*H had not been cataloged, 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 the subunit stoichiometries and quaternary structures of these complexes should be. Instead,

Author contributions: B.-G.H., M.D., T.C.H., H.E.W., K.H.D., S.E.B., J.-M.C., M.D.B., and R.M.G. designed research; B.-G.H., M.D., H.L., L.C., J.G., M. Singer, M.C., D.A.B., D.T., and M. Shatsky performed research; B.-G.H., M.D., H.L., H.E.W., D.A.B., M. Shatsky, S.E.B., and J.-M.C. analyzed data; and M.D., T.C.H., H.E.W., D.A.B., K.H.D., M. Shatsky, S.E.B., J.-M.C., M.D.B., and R.M.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]B.-G.H. and M.D. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: rmglaeser@lbl.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0813068106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0813068106

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.

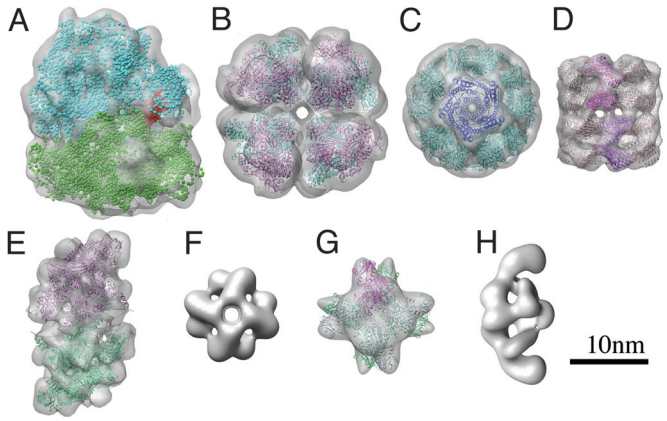

Fig. 1. Gallery of three-dimensional reconstructions obtained by single-particle electron microscopy for eight different, large macromolecular complexes isolated from *Desulfovibrio vulgaris* Hildenborough. Whenever pseudoatomic-resolution models could be created by docking known atomic structures for 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.

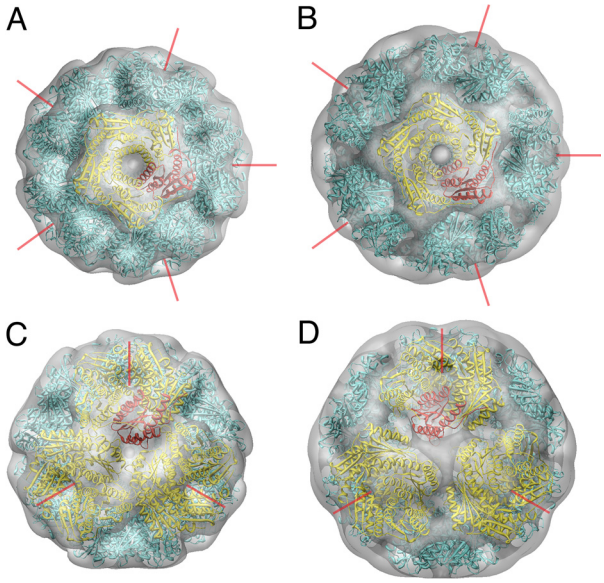

Fig. 2. Comparison of the two types of icosahedral complexes of lumazine synthase (riboflavin synthase beta subunit) formed by the proteins from *Aquifex aeolicus* (*A* and *C*) and from *D. vulgaris* Hildenborough (*B* and *D*), respectively. Note that the vertices of the pentameric subunits of the *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

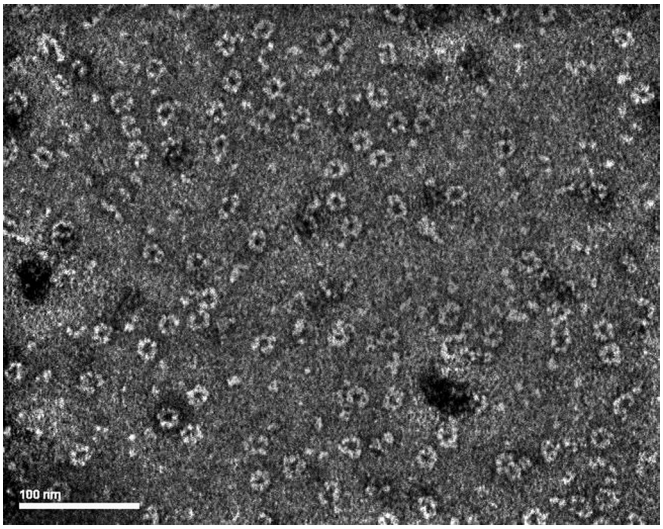

Fig. 3. Electron micrograph of a putative carbohydrate phosphorylase complex isolated from *Dv*H and negatively stained with uranyl acetate. The basic ring-shaped nature of this protein complex is quite apparent, thereby establishing that the subunit stoichiometry and quaternary structure of this complex 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.