Article

Symmetry at the Cellular Mesoscale

David S. Goodsell ^{1,2}

¹ Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA; goodsell@scripps.edu

² RCSB Protein Data Bank, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Received: 23 August 2019; Accepted: 6 September 2019; Published: 16 September 2019

Abstract: Symmetry plays a functional role in the structure and action of biomolecules and their associations and interactions in living cells. This symmetry is a natural consequence of the evolutionary mechanisms that lead to the development of life, and it ranges from perfect point-group symmetry in protein oligomers to more approximate symmetries in the higher-order mesoscale structure of cellular environments.

Keywords: protein symmetry; protein function; protein structure; cellular mesoscale; quasisymmetry; molecular evolution

1. Introduction

Biology is filled with symmetry, where regularity of form leads to functional features. The cellular mesoscale—the level between the nanometer scale of atomic structure and the micrometer scale of cellular ultrastructure—is particularly rich in functional symmetries. Some examples are perfect expressions of classic point-group symmetry, whereas others are looser expressions of regularity or periodicity. As with much of biology, there are no hard-and-fast rules, and it is often hard to draw a sharp line between what is symmetrical and what is not. This makes the study of biomolecular symmetry both challenging and appealing, since there always seem to be interesting exceptions in the gray area between perfect symmetry and complete asymmetry.

The topic of symmetry has been a perennial favorite among biomolecular researchers, particularly among the pioneers of biomolecular structure determination, where symmetry arguments often shed light on the structures and functions of the molecules that were being studied. Historical examples include Pauling’s insight that a symmetrical alpha-helical structure of polypeptides would maximize the number of internal hydrogen bonds [1], Watson and Crick’s symmetrical double-helical model of DNA based on the X-shaped X-ray diffraction pattern [2], and Caspar and Klug’s concept of quasisymmetry to unify diverse observations about the sizes and stoichiometries of viral capsids [3]. Monod [4] synthesized these diverse observations, postulating that symmetrical biomolecular assemblies are a consequence of the functional need for “finiteness, stability and self-assembly”. More recently, there has been a resurgence of interest in biomolecular symmetry [5–8], driven by the explosion of structural results and the ability to engineer new protein assemblies, which requires, of course, a deep understanding of the basic principles that drive protein structure and assembly.

Experimental and theoretical methods are rapidly improving, allowing for the research community to build from examination of individual biomolecules to levels of supramolecular assembly and the molecular details of cellular ultrastructure. CryoEM methods allow for the resolution of large molecular assemblies, such as ribosomes and microtubules, in their cellular context [9]. Computational modeling and simulation are now shedding light on genomes [10], entire cells [11], and even cells over their lifecycle [12]. In this short review, I will discuss the evolutionary constraints that lead to the prevalence of symmetrical structures in biology, and then survey some exemplary functional symmetries that appear between the atomic and cellular levels.

2. Evolutionary Constraints

Remarkably, mesoscale symmetry is a natural consequence of the way that life developed from simple precursors on the Earth. The early evolution of life (or alternatively, the whims of an intelligent designer) yielded a central paradigm of biomolecular construction, where the information is stored in a polymer (DNA) and is used to synthesize self-assembling biomachines (proteins), which then construct (and self-assemble) all of the other components of the cell. This paradigm imposes several constraints on the designs of biomolecules and the organisms built from them:

1. the fidelity of polymerases and ribosomes effectively limit the size of information-specified proteins, and there is limited information space in the genome, leading to a typical protein size range of 250 to 450 amino acids;
2. larger structures are typically built in a modular way, through non-covalent association of subunits or covalently-attached polymers of modular components;
3. cellular structures rely on many copies of these biomolecules to improve the statistics of error incorporation and of thermodynamics; and,
4. most processes in cellular biology proceed through stochastic interaction of components; directed motion at the cellular mesoscale is the exception, not the rule.

It is instructive to spend a moment to contrast these characteristics with the familiar machinery of our macroscale world. Most mechanical devices are constructed with a defined plan of components and they have a defined set of allowable motions: for example, automobile engines have a defined number of pistons that cycle through a highly-predictable set of actions. Since we control much of our machinery, the stochastic approach is only rarely applied (as in autonomous vacuum cleaners, such as the Roomba). However, there is a key similarity: industrial machinery uses a modular approach that is similar to that of biomachinery, allowing for ready replacement of faulty components.

Accordingly, because of the need for a modular approach to biomolecules, self-similarity is built into mesoscale structure at many levels (Figure 1). We find strong geometric symmetry at the molecular level, and softer regularities at higher levels while looking at the scale ranges from atoms to cells. At all of these levels, the evolution of symmetry is often driven by functional needs.

3. The Atomic Level: Self-Similarity in Biological Polymers

Cells take a parsimonious approach to construction, building polymers with a defined set of repeated units: roughly 20 amino acids for proteins and eight nucleotides for DNA and RNA. This has many benefits, including the need for only a limited alphabet of information in the genome and minimizing the number of synthetic pathways needed to construct the building blocks. This polymeric construction leads to classic symmetries that are based on common elements of the repeated chemical unit, such as the protein alpha helix and the DNA double helix.

Repeated chemical structure at higher levels is used to create biomolecules with new functional symmetries (Figure 2, top). For example, a repeated sequence of three amino acids, glycine-X-proline (where X can be any amino acid), forms the characteristic triple helix of collagen, which is used as a structural element in connective tissue. Glycine is the smallest amino acid, allowing for close contact between the three strands on the interior of the helix.

Figure 1. Artistic conception of the mesoscale structure of a cell. The process of autophagy is shown, where membranes (green) are built around internal portions of a cell, such as the large aggregates of digestive enzymes shown here (red and magenta), and then delivered to lysosomes (darker region at bottom). Many of the biomolecules are symmetrical complexes of identical subunits, including point group symmetries in the enzymes and helical symmetry in the long actin filaments shown in blue. We can also see regularity of distribution at higher scale levels, for example in the scattering of ribosomes (purple) and enzymes (blue) in the cytoplasm, and the arrays of proteins embedded in the membranes. Illustration from a feature at RCSB PDB-101 (<http://pdb101.rcsb.org/motm/203>).

There are also numerous examples of repeated domains in proteins, where a gene is duplicated to form a larger multi-domain protein. The two (or more) domains might then diverge during evolution according to functional need. Figure 2 (bottom) shows a repeated domain that is used in many structural tasks, termed the “armadillo” repeat after a fruit fly mutation, where it was discovered. The human genome includes many examples with 4–12 consecutive repeats of this motif, which together form a helical structure. The inner surface of this helix interacts with an extended conformation of a target protein. The structure that is shown here is an engineered version that is designed to bind to a repeated peptide of lysine and arginine amino acids with a defined length [13].

Figure 2. Tandem repeats in proteins. (Top) The sequence of collagen has many repeated triplets that include a glycine (highlighted in blue), which packs into the tight interior of the collagen triple helix (shown in spheres in the molecular image). (Bottom) An engineered protein with nine armadillo repeats (red and magenta), designed to bind to a peptide (blue) of defined length. Molecular illustrations created using JSmol (jmol.sourceforge.net) with PDB entries 1bkv, 5mfmm (www.rcsb.org).

4. The Molecular Level: Symmetries of Biological Assemblies

Biological symmetry is arguably at its best at the level of biomolecular assemblies [5]. In many cases, these symmetries have evolved for specific functional consequences. Figure 3 shows a few examples. Symmetry is commonly used to create biomolecules with specific shapes, for example, cyclic symmetries are commonly found in pores, icosahedral symmetries are found in nano-containers, and helical symmetries are used to build filaments. Symmetry is also central in allosteric regulation, where conformational switches between two symmetrical conformations of an oligomeric protein are triggered by the binding of effector molecules, and then modulate the activity of the protein.

Figure 3. Symmetric assemblies with specific functions. Complement C9 forms pores with cyclic symmetry that pierce cell membranes; pyruvate dehydrogenase is a tetrameric enzyme regulated by an allosteric change in shape; actin filaments with helical symmetry provide infrastructure in cells. Figure created with Illustrate (<http://ccsb.scripps.edu/illustrate>) with PDB entries 6dlw, 1e0u, 1a3w, 5mvy.

The vast majority of soluble oligomeric proteins have point-group symmetry. A point group is a group of geometric symmetries that keep at least one point fixed. Because of the enantiomorphic nature of most biological molecules, biological point groups are restricted to groups of rotational symmetries, with all of the axes passing through a single point. These fall into several classes (Figure 4), including cyclic symmetries with a single axis of rotation, dihedral groups with a cyclic axis and perpendicular two-fold axes, and octahedral and icosahedral groups. These symmetries are the physical expression of Monod's principles of "finiteness, stability and self-assembly". Point groups ensure that a finite, defined number of subunits are included in the assembly; they maximize the number of contacts (and thus stability) within an assembly and they yield multiple association pathways to promote self-assembly. Helical symmetries are also common, for example in structural filaments of the cytoskeleton, but they typically require a collection of nucleating and severing machinery to manage and regulate their construction.

Figure 4. Point group symmetries of several protein assemblies. Rotational symmetry axes are shown with lines, with a small oval on two-fold axes, a triangle on three-fold axes, etc. The sliding clamp uses cyclic symmetry to create a three-fold ring that encircles DNA, glutamine synthetase contains twelve identical subunits with dihedral symmetry that use allosteric motions to regulate their action, and ferritin and the virus capsid use higher symmetries to create hollow shells. Visualized using online Jmol tools at the RCSB PDB, with entries 1plq, 2lgs, 1fha, 2buk.

Molecular symmetries have been a boon for structural biology. X-ray crystallography has been the workhorse of structure determination for decades. It relies on obtaining well-ordered single crystals of a purified biomolecule, which is an empirical art that is enabled by the intrinsic point-group symmetry of many of these molecules. Crystallographers employ all manner of symmetry-related tricks during structure determination, such as averaging electron density maps by non-crystallographic symmetry. Structural biologists are only now being released from this central reliance on the symmetry of crystal lattices with recent improvements in cryo-electron microscopy. These methods gather data on thousands of randomly-oriented particles and use it to build a tomographic image of the molecule. Symmetry averaging can assist in this synthesis, and it has been used for years in tomograms of viral capsids. However, recent improvements in image classification and experimental resolution have opened the door to completely asymmetric structures, such as ribosomes.

The nuts-and-bolts of crystallographic symmetry can cause difficulties for non-experts. One persistent source of confusion is the concept of the asymmetric unit of a crystal lattice. Typically, coordinates for the asymmetric unit are deposited in the Protein Data Bank archive, and larger symmetrical assemblies might be generated from the crystal lattice parameters. Depending on the details of the crystal form, this asymmetric unit may include one subunit of a biologically relevant assembly, the entire assembly, or some other fraction of the assembly. The RCSB Protein Data Bank (rcsb.org), in a large remediation effort, recently identified the likely functional biological units, which makes them easily accessible to non-expert users.

Since this is biology, there are also many exceptions to formal symmetry, which were explored in detail in a previous review [5]. These include:

1. pseudosymmetry of hemoglobin, where two alpha subunits and two similar beta subunits form a complex with approximate D2 symmetry;
2. symmetry mismatch in molecular motors such as ATP synthase, which links a proton-driven motor with C12 symmetry with an ATP-driven motor with C3 symmetry, which is further distorted to C1 symmetry by an asymmetric axle; and,
3. quasisymmetry of viral capsids, discussed in more detail in the next section.

Figure 5 shows a type of pseudosymmetry that occasionally shows up when comparing prokaryotic bacterial cells and eukaryotic cells like our own. The prokaryotic protein is a simple symmetrical complex of several identical subunits, and the eukaryotic protein has a similar form, but it is composed of multiple domains that are connected into one long chain. This longer protein has presumably developed through gene duplication of the simpler bacterial protein, and the longer protein allows additional evolutionary latitude to tune the function of the different portions of the assembly.

Figure 5. Voltage-gated sodium channel from bacteria with C4 symmetry, and a human channel with similar overall folding pattern but composed of one continuous chain (and two auxiliary subunits in red). Jmol illustration using PDB entries 3rw0, 6j8i.

5. The Supramolecular Level: Quasisymmetry

There is one additional stopping point on our journey from atoms to the cellular mesoscale. Proteins with segmental flexibility are used to create assemblies that are larger than possible while using strict point-group symmetry (Figure 6). Much like the geodesic domes of Buckminster Fuller, Caspar and Klug [3] conceived viral quasisymmetry as an approximate tiling of an icosahedron with triangles of slightly different sizes. They define a triangulation number: $T = 1$ is a perfect icosahedron with one triangle on each face, $T = 4$ has four triangles on the face, and so on. Subunits then pack into these slightly distorted triangular lattices, with small levels of segmental flexibility that accommodate the slight variations in triangle size and dihedral interactions. As additional viruses have been studied, this beautiful conception has needed some stretching, for example, with the cone-shaped capsid of HIV, which has traditional hexamers and pentamers, but forms non-equilateral faces, and even viruses, such as SV40, which are entirely composed of pentameric subunits, connected with flexible arms.

Figure 6. (Left) Icosahedral quasisymmetry of several virus capsids. Subunits at the pentameric vertices are in red, and subunits in other unique quasisymmetric positions are shown in different colors. (Right) Clathrin triskelions, BAR and F-BAR proteins, and dynamin (in shades of red and purple) use quasisymmetrical assembly to perform the mesoscale task of creating a membrane vesicle inside a “coated pit”. Virus images from the RCSB “Molecule of the Month” (pdb101.rcsb.org) using PDB entries 2buk, 2tbv, 1ohf, 1ohg, and the coated pit illustration was created as part of a collaborative effort with Tim Herman at the Milwaukee School of Engineering (<http://cbm.msosoe.edu/includes/modules/clickableCoatedPit/coatedPit.html>).

Softer forms of quasisymmetry allow for the generation of even larger, but less regular, assemblies. For example, coated pits (Figure 6, right) invaginate regions of cell membranes to form vesicles, driven by a collection of protein assemblies. Clathrin forms three-armed “triskelions” that assemble to form hollow cages with a variety of sizes, depending on the location of hexameric and pentameric associations. BAR and F-BAR proteins are crescent-shaped, and they form arrays that help with membrane bending. Dynamin is a molecular motor that forms a progressively constricting collar to pinch off the membrane. Similarly, the inner surface of entire red blood cells is supported flexible spectrin chains, which link through large membrane-bound protein complexes to form a geodesic network. This network provides a rich palette of interactions, based on the intrinsic flexibility of spectrin and valency of interactions that are possible at each membrane-bound node, which give erythrocytes their distinctive flexibility and resiliency.

6. The Mesoscale Level: Self-Similarity and Regularity

Formal symmetries are rare at the mesoscale, but self-similarity and regularity are the rule. Two-dimensional (2D) and three-dimensional (3D) crystals are occasionally found when high density or limited accessibility is functionally needed. Figure 7 includes two examples. Pancreatic beta cells have small secretory vesicles that store and protect large amounts of insulin, for release into the blood. During maturation of these vesicles, insulin is processed and the mature insulin often forms small crystals inside the vesicle, which quickly dissociate once the vesicle fuses with the cell surface for release. Some membrane proteins, such as bacteriorhodopsin in photosynthetic bacteria, are present at very high concentrations and they form ordered two-dimensional crystals. However, in most cases the heterogeneity of cellular environments precludes these types of beautiful symmetries.

Figure 7. Natural crystal symmetries. Top, artistic rendering of a cross section through an insulin secretory vesicle near the cell membrane of a pancreatic beta cell, with a crystal of roughly 350,000 insulin molecules in yellow. Bottom, two-dimensional (2D) crystal of bacteriorhodopsin (purple, from PDB entry 2brd) from a crystallographic structure of the membrane from a photosynthetic bacterium. Lipids are shown in white.

Mesoscale environments, on the other hand, have a high measure of self-similarity and ultrastructural uniformity. This is a consequence of the design constraints of the mesoscale: the need for statistical homeostasis requires high copy numbers, and the stochastic nature of interaction disallows most large-scale 2D/3D symmetries. Together, these constraints lead to large regions with more-or-less uniform but random distribution of molecules.

This is a boon for study of the mesoscale: in essence, if you study one small region of cytoplasm, you have effectively studied the entire cytoplasm. Taking advantage of this, much of mesoscale modeling has been performed at coarse-grain levels, ranging from modeling proteins with simple shapes to simply representing the local concentrations of molecules within the ultrastructural compartments of

the cell. For structural modeling, we are currently exploring the use of “mesoscale maps” that define regions of the cell with unique structure, to extend the functionality of programs, such as CellPACK [14]. A plausible first pass at a bacterial map has surprisingly few unique environments (Figure 8). Methods, such as LipidWrapper [15], might then be used to populate these regions with structural “tiles” [16]. Of course, this approach completely breaks down for features with large-scale addressability, such as the bacterial genome. In those cases, explicit modeling of the entire molecular assembly is necessary if the addressable nature is the subject of study. However, in other cases, such as capturing the sieve-like nature of the DNA strands, the continuum-map-based approach might be sufficient.

Figure 8. Mesoscale map for *Escherichia coli*. Regions in separate colors in the top half may be modeled using a more-or-less uniform recipe of molecules and concentrations.

7. Prospects

Much of molecular function might be understood and reconciled by applying knowledge of molecular symmetries and how they impact molecular synthesis, maturation, and interactions. Symmetry has also played a central role in the methodology of structural biology, through crystallographic analysis and symmetry averaging in electron microscopy. Softer approaches to symmetry, uniformity, and regularity are now informing the growing field of mesoscale modeling

and experiment, helping to uncover and characterize basic principles in the domain where multiple molecules combine to create a living cell.

Funding: This work was funded by US National Institutes of Health R01-GM120604 and the RCSB Protein Data Bank (National Science Foundation DBI-1832184, the National Institutes of Health R01GM133198, and the US Department of Energy DE-SC0019749). The APC has been kindly waived by the editors.

Acknowledgments: This is manuscript 29887 from the Scripps Research Institute.

Conflicts of Interest: The author declares no conflict of interest.

References

- Pauling, L.; Corey, R.B. Two hydrogen-bonded spiral configurations of the polypeptide chain. *J. Am. Chem. Soc.* **1950**, *72*, 5349. [[CrossRef](#)]
- Watson, J.D.; Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)] [[PubMed](#)]
- Caspar, D.L.D.; Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **1962**, *27*, 1–24. [[CrossRef](#)] [[PubMed](#)]
- Monod, J. On symmetry and function in biological systems. In Proceedings of the Eleventh Nobel Symposium on Symmetry and Function of Biological Systems at the Macromolecular Level, Sodergarn, Sweden, 26–29 August 1968; Engstrom, A., Strandberg, B., Eds.; John Wiley and Sons: New York, NY, USA, 1969; pp. 15–27.
- Goodsell, D.S.; Olson, A.J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105–153. [[CrossRef](#)] [[PubMed](#)]
- Marsh, J.A.; Teichmann, S.A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **2015**, *84*, 551–575. [[CrossRef](#)] [[PubMed](#)]
- Cannon, K.A.; Ochoa, J.M.; Yeates, T.O. High-symmetry protein assemblies: Patterns and emerging applications. *Curr. Opin. Struct. Biol.* **2019**, *55*, 77–84. [[CrossRef](#)] [[PubMed](#)]
- Pieters, B.J.; van Eldijk, M.B.; Nolte, R.J.; Mecnovic, J. Natural supramolecular protein assemblies. *Chem. Soc. Rev.* **2016**, *45*, 24–39. [[CrossRef](#)] [[PubMed](#)]
- Beck, M.; Baumeister, W. Cryo-Electron Tomography: Can it Reveal the Molecular Sociology of Cells in Atomic Detail? *Trends Cell Biol.* **2016**, *26*, 825–837. [[CrossRef](#)] [[PubMed](#)]
- Rosa, A.; Zimmer, C. Computational models of large-scale genome architecture. *Int. Rev. Cell Mol. Biol.* **2014**, *307*, 275–349. [[PubMed](#)]
- Im, W.; Liang, J.; Olson, A.; Zhou, H.X.; Vajda, S.; Vakser, I.A. Challenges in structural approaches to cell modeling. *J. Mol. Biol.* **2016**, *428*, 2943–2964. [[CrossRef](#)] [[PubMed](#)]
- Karr, J.R.; Sanghvi, J.C.; Macklin, D.N.; Gutschow, M.V.; Jacobs, J.M.; Bolival, B., Jr.; Assad-Garcia, N.; Glass, J.I.; Covert, M.W. A whole-cell computational model predicts phenotype from genotype. *Cell* **2012**, *150*, 389–401. [[CrossRef](#)] [[PubMed](#)]
- Hansen, S.; Ernst, P.; Konig, S.L.B.; Reichen, C.; Ewald, C.; Nettels, D.; Mittl, P.R.E.; Schuler, B.; Pluckthun, A. Curvature of designed armadillo repeat proteins allows modular peptide binding. *J. Struct. Biol.* **2018**, *201*, 108–117. [[CrossRef](#)] [[PubMed](#)]
- Johnson, G.T.; Autin, L.; Al-Alusi, M.; Goodsell, D.S.; Sanner, M.F.; Olson, A.J. cellPACK: A virtual mesoscope to model and visualize structural systems biology. *Nat. Methods* **2015**, *12*, 85–91. [[CrossRef](#)] [[PubMed](#)]
- Durrant, J.D.; Amaro, R.E. LipidWrapper: An algorithm for generating large-scale membrane models of arbitrary geometry. *PLoS Comput. Biol.* **2014**, *10*, e1003720. [[CrossRef](#)] [[PubMed](#)]
- Klein, T.; Autin, L.; Kozlikova, B.; Goodsell, D.S.; Olson, A.; Groller, M.E.; Viola, I. Instant Construction and Visualization of Crowded Biological Environments. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 862–872. [[CrossRef](#)] [[PubMed](#)]

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).