

# STRUCTURAL SYMMETRY AND PROTEIN FUNCTION

---

David S. Goodsell and Arthur J. Olson

*Department of Molecular Biology, Scripps Research Institute, La Jolla, California 92037;  
 e-mail: goodsell@scripps.edu, olson@scripps.edu*

**Key Words** oligomeric proteins, protein symmetry, protein structure/function relationships

■ **Abstract** The majority of soluble and membrane-bound proteins in modern cells are symmetrical oligomeric complexes with two or more subunits. The evolutionary selection of symmetrical oligomeric complexes is driven by functional, genetic, and physicochemical needs. Large proteins are selected for specific morphological functions, such as formation of rings, containers, and filaments, and for cooperative functions, such as allosteric regulation and multivalent binding. Large proteins are also more stable against denaturation and have a reduced surface area exposed to solvent when compared with many individual, smaller proteins. Large proteins are constructed as oligomers for reasons of error control in synthesis, coding efficiency, and regulation of assembly. Symmetrical oligomers are favored because of stability and finite control of assembly. Several functions limit symmetry, such as interaction with DNA or membranes, and directional motion. Symmetry is broken or modified in many forms: quasisymmetry, in which identical subunits adopt similar but different conformations; pleomorphism, in which identical subunits form different complexes; pseudosymmetry, in which different molecules form approximately symmetrical complexes; and symmetry mismatch, in which oligomers of different symmetries interact along their respective symmetry axes. Asymmetry is also observed at several levels. Nearly all complexes show local asymmetry at the level of side chain conformation. Several complexes have reciprocating mechanisms in which the complex is asymmetric, but, over time, all subunits cycle through the same set of conformations. Global asymmetry is only rarely observed. Evolution of oligomeric complexes may favor the formation of dimers over complexes with higher cyclic symmetry, through a mechanism of prepositioned pairs of interacting residues. However, examples have been found for all of the crystallographic point groups, demonstrating that functional need can drive the evolution of any symmetry.

## CONTENTS

INTRODUCTION .....	106
THE SYMMETRY OF OLIGOMERIC PROTEINS .....	107
Characteristics and Natural Occurrence of Symmetry Groups .....	108

Structural Mechanisms of Oligomerization . . . . .	111
WHY BUILD SYMMETRICAL PROTEINS? . . . . .	112
Why Build Large Proteins? . . . . .	114
Why Build Oligomeric Proteins? . . . . .	115
Why Build Symmetrical Proteins? . . . . .	116
Functional Niches for Small Monomeric Proteins . . . . .	119
SYMMETRY AND COOPERATIVITY . . . . .	120
Allosteric Regulation . . . . .	120
Multivalent Binding . . . . .	122
MORPHOLOGICAL FUNCTIONS OF SYMMETRY . . . . .	123
Rulers, Rings, and Containers . . . . .	123
The Need to Be Large . . . . .	127
FUNCTIONS THAT LIMIT SYMMETRY . . . . .	128
Directional Motion . . . . .	128
Interaction with DNA . . . . .	129
Membrane Interactions . . . . .	129
SYMMETRY BREAKING IN OLIGOMERIC PROTEINS . . . . .	130
Quasisymmetry . . . . .	130
Pleomorphic Symmetry . . . . .	135
Pseudosymmetry . . . . .	137
Symmetry Mismatch . . . . .	138
ASYMMETRY . . . . .	141
Local Asymmetry . . . . .	142
Reciprocating Mechanisms . . . . .	143
Global Asymmetry . . . . .	144
EVOLUTION OF OLIGOMERIC PROTEINS . . . . .	146
THE AESTHETICS OF SYMMETRY . . . . .	148

## INTRODUCTION

Symmetry has played an important role in science from its very origins. The Greeks, fascinated by the symmetry of vibrating strings, developed a quantitative understanding of pitch and harmony, and Kepler formulated a simple mathematical description of gravity that was based on the elliptical geometry of planetary orbits. Today, symmetry continues to permeate scientific thought. Physicists are looking for symmetries to unify an ever-growing menagerie of subatomic particles, and developmental biologists are discovering how the simple symmetries of molecular diffusion may combine to form complex body plans during embryogenesis. Some fields seem ripe for such symmetries, but when these fields are critically analyzed, the proposed symmetries never materialize. Kepler's attempt to rationalize the positions of the planets in the solar system based on Platonic solids is a historical case in point.

CA Coulson, a theoretical chemist and mathematician, described the utility and the seduction of symmetry in his own field: "It is when symmetry interprets facts that it serves its purpose; and then it delights us because it links our study of chemistry with another world of the human spirit—the world of order, pattern, beauty,

satisfaction” (as quoted in 34). In the many studies, both proven and spurious, in which researchers have attempted to find symmetry, the assumption has been made implicitly that such unifying symmetries *do* exist—that Nature herself is built by symmetry from simpler components. Historically, many searches for symmetry were motivated by belief in a divine creator with aesthetic sensibilities similar to our own. Most modern researchers, however, see symmetry as an emergent feature of the general parsimony of our observed universe, resulting from the limited modes of interaction between a small number of building blocks as they assemble (or are assembled) into structures of greater complexity.

Symmetry has played a central role in biomolecular science since its earliest triumphs. The structure of DNA reported by Watson and Crick in 1953, with its direct relationship of double-helical symmetry to genetic function, set the stage and perhaps overshadowed all that has followed. Indeed, Kendrew is said to have been disappointed in the “visceral” nature of myoglobin at low resolution, a disappointment that was more than compensated for by the symmetrical spiral tubes of  $\alpha$ -helices in the atomic-resolution structure. In this review, we explore the functional roles played by structural symmetry in macromolecules. For discussion of other types of symmetry in molecular processes, such as the inherent symmetry of reversible reactions, the reader might begin with the discussion by Garcia-Bellido (26).

## THE SYMMETRY OF OLIGOMERIC PROTEINS

Symmetry is the rule rather than the exception for proteins. Most of the soluble and membrane-bound proteins found in living cells form symmetrical oligomeric complexes with two or more identical subunits, and nearly all structural proteins are symmetrical polymers of hundreds to millions of subunits. Svedberg has been credited with the idea that proteins are composed of discrete subunits (90). In 1967, Klotz presented a list of proteins presumed to form oligomers (48). This list was expanded to ~300 entries (primarily soluble enzymes) in a 1975 review (49), underscoring the prevalence of oligomeric proteins in cells. In that compilation, over half of the oligomeric proteins are homodimers or homotetramers, presumed to form symmetrical complexes, and only ~15% were heterooligomers of different chains. Klotz et al (49) also noted the relative scarcity of oligomers with odd numbers of subunits. Goodsell attempted to quantify the prevalence of oligomeric proteins in cells based on the concentration of soluble proteins in *Escherichia coli*, obtaining an average oligomerization state of about four (27), and a visual survey of soluble proteins in the Protein Data Bank (PDB) underscored the prevalence of symmetrical, oligomeric species (28). Jones & Thornton tabulated the multimeric states of proteins in the July 1993 release of the PDB (43), finding a predominance of monomers; of 970 total proteins, 66% were monomeric, 15% were dimeric, 12% were tetrameric, and the remainder adopted other oligomeric states. Jones & Thornton noted, however, that the PDB over-represents small monomers, owing to the difficulties involved in protein crystallization.

**TABLE 1**    Natural occurrence of oligomeric proteins in *Escherichia coli*<sup>a</sup>

Oligomeric state	Number of homooligomers	Number of heterooligomers	Percent
Monomer	72		19.4
Dimer	115	27	38.2
Trimer	15	5	5.4
Tetramer	62	16	21.0
Pentamer	1	1	0.1
Hexamer	20	1	5.6
Heptamer	1	1	0.1
Octamer	3	6	2.4
Nonamer	0	0	0.0
Decamer	1	0	0.0
Undecamer	0	1	0.0
Dodecamer	4	2	1.6
Higher oligomers	8		2.2
Polymers	10		2.7

<sup>a</sup>These data were compiled by using information at the SWISS-PROT annotated protein sequence database (on the World Wide Web at [www.expasy.ch/sport](http://www.expasy.ch/sport)), with search tools developed by Michel Sanner. The list of *Escherichia coli* K12 chromosomal entries (compiled by Amos Bairoch including release 35.0 of the database and updates to May 1998) was searched for entries with explicit “subunit” annotations, yielding 617 entries. This corresponds to 16% of the total list, or 30% if “hypothetical” proteins are omitted. These individual protein chains were then processed manually to create a list of 372 oligomeric species.

A survey of *E. coli* proteins in the SWISS-PROT annotated protein sequence databank is included in Table 1. This survey includes soluble proteins, membrane-bound proteins, and structural proteins. Monomers are in the minority, composing only about one fifth of the protein species. Dimers and tetramers are far more common. Homooligomers also predominate: 79% of oligomers with from 2 to 12 subunits are homooligomers, whereas only 21% form heterooligomeric complexes. As discussed in sections below, these homooligomeric complexes, when structures are known, associate by closed point group or helical symmetry. Asymmetric homooligomers are virtually unknown.

## Characteristics and Natural Occurrence of Symmetry Groups

Early in the evolution of life, protein was selected as the basic material for building the cellular machinery. With this selection came a choice of “handedness”—choosing one chirality of the  $\alpha$  carbon over the other. The reason for the choice of L-amino acids instead of D-amino acids, and chiral amino acids instead of an achiral analog, has been the subject of much scientific and philosophical discussion. One of the major consequences of the adoption of exclusively L-amino-acid proteins is

that modern oligomeric proteins adopt only enantiomorphic symmetries; mirror and inversion symmetries are disallowed.

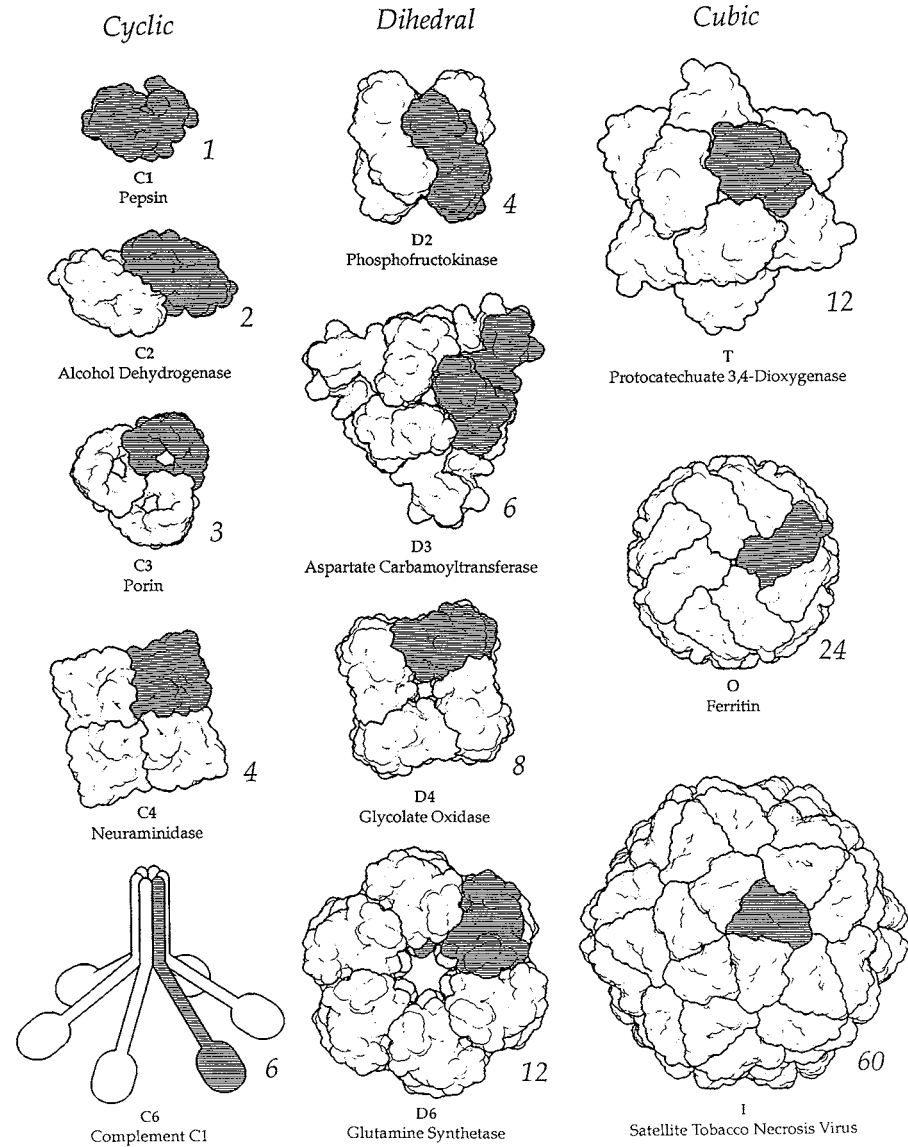
Examples of most of the low-copy-number enantiomorphic point groups have been observed in naturally occurring proteins. All of the crystallographic point groups have been used, as shown by the examples in Figure 1. The choice of a particular symmetry group can have a profound effect on the function and stability of the complex.

**Cyclic Groups** The cyclic groups contain a single axis of rotational symmetry, forming a ring of symmetrically arranged subunits. C1 symmetry (monomeric proteins) and C2 symmetry (dimeric proteins) are common among proteins of diverse function. The higher cyclic groups are much more rare. Typically they are involved in functions that require directionality or sidedness, such as interaction with membranes or rotational motion, or functions that require formation of a hollow tube or chamber.

**Dihedral Groups** The dihedral groups contain an axis of rotational symmetry and a perpendicular axis of two-fold symmetry. Dihedral symmetry is common among soluble cytoplasmic enzymes, particularly tetramers with D2 symmetry. Oligomers with dihedral symmetry have several different types of interface, including interfaces between oligomers related by the main rotational symmetry and dimeric interfaces related by the perpendicular two-fold axes. This provides a rich infrastructure from which to build allosteric control.

The choices for stability and interaction are potentially greater for dihedral oligomers than those available in cyclic oligomers with the same number of subunits. Cyclic groups of four-fold or greater symmetry limit contacts between subunits. In most cases, there will be few cross contacts in a cyclic group, and only neighboring subunits around the cyclic ring will be in contact. Imagine a C4 complex with subunits numbered sequentially around the ring. The nature of the symmetry makes difficult any contact between subunit 1 and subunit 3 and between subunits 2 and 4. In the dihedral group D2, however, these contacts are allowed (although observed less frequently than one might expect), and the complex is formed around three different dimeric interfaces, allowing many options for regulation of this interaction and giving more opportunity for contact between all subunits, thereby increasing stability. This problem is even worse in the higher symmetries, where even larger ring structures are formed.

**Cubic Groups** Cubic symmetries contain three-fold symmetry combined with another, nonperpendicular rotational axis, with three possibilities: tetrahedral, with three- and two-fold axes; octahedral, with three- and four-fold axes; and icosahedral, with three- and five-fold axes. Cubic symmetries, with their exacting structural constraints, primarily play specialized roles in storage and transport. Crick & Watson (18, 19) first proposed that cubic symmetries, and icosahedral symmetry in particular, are uniquely suited to creation of hollow shells, such as the protein coats of simple spherical viruses.



**Figure 1** Crystallographic point group symmetries. Examples of proteins with each of the crystallographic point group symmetries have been found. Point group symbols are included below each protein structure (e.g. *C1* and *D2*), and the number of identical subunits in each group is included below and to the right of the structure (e.g. 24 in octahedral group *O*). One subunit is shaded in each example. Note that other noncrystallographic point groups are consistent with the enantiomorphic nature of proteins, including cyclic symmetries *C5* and *C7* or higher and dihedral symmetries *D5* and *D7* or higher. Protein Data Bank accession codes for all structures used in the figures (101) are accessible via the Internet (<http://www.rcsb.org/pdb>).

**Line, Plane, and Space Groups** The addition of translational symmetry to rotational symmetries forms helical structures, symmetrical planes, and space-filling crystals. These symmetries are unbounded, in that they may be extended indefinitely until the organism runs out of room, runs out of subunits, or mechanically stops the growth.

Line symmetries combine rotation with translation along the rotation axis, forming a helix. A perpendicular two-fold rotation axis may also be incorporated, to form a double helix or higher-order intertwined helices. Helices formed of protein subunits are widely used as structural elements. Pauling proposed that a subunit with two complementary binding surfaces would form a hollow helical fibril (76). Structures from electron microscopy reveal this helical symmetry in, for example, microtubules, flagella, and tobacco mosaic virus. Helical interactions are also used to build tighter, narrower fibers, without the central hollow, by orienting the binding surfaces such that only a small number of subunits compose each turn. Examples of this include actin fibrils and intermediate filaments.

Plane symmetries are formed when translation is applied in two spatial directions and combined with rotational elements. Plane symmetries abound in decorative artwork; the elaborate tiling designs of MC Escher are prime examples. Plane arrays of proteins are found in biological membranes, such as connections at cellular gap junctions, which form a tight hexagonal array.

Space groups, although playing an indispensable role in protein structure determination, are relatively rare in vivo. Collagen forms a natural three-dimensional lattice in connective tissue fibers, and a mutation in hemoglobin favors the formation of long, fibrous three-dimensional lattices that distort red blood cells in sickle cell anemia. Small crystalline arrays may be found in hormone storage granules and in peroxisomes. Perhaps the rigid uniformity of three-dimensional lattices precludes their widespread use as biological motifs; life is built on a more malleable plan, allowing greater diversity of structure.

## Structural Mechanisms of Oligomerization

The subunits in oligomeric proteins interact through highly specific contact surfaces. Monod et al identify two types of contact surfaces in oligomeric complexes: isologous (or homologous) contacts and heterologous contacts (72). They define isologous interfaces as those where identical surfaces on the two subunits interact and heterologous interfaces as those formed by different surfaces on the two subunits. As they mention, "In a heterologous association, the domain of bonding has no element of symmetry." They note that isologous interfaces are limited to dimeric associations, where a two-fold axis crosses through the middle of the interface, and all other associations between two subunits are necessarily heterologous. The major consequence of the type of an interface, isologous or heterologous, is for the evolution of the interface, as described below in the section on evolution. The terms isologous and heterologous have fallen out of use in recent years. This is perhaps owing to the discovery of many oligomeric proteins with noncontiguous interfaces, such as the  $\beta$ -subunit of the DNA polymerase III holoenzyme. Domain-swapped



dimers, in particular, strain the definition; these interfaces are split into two discrete units, and often the flexibility of the linkers connecting the two domains may relax the strict two-fold symmetry of the entire split interface.

The structural features of protein-protein interfaces have received extensive study. Crane presented two ideas based on insights from physics, before atomic structures of proteins were known. "For a high degree of specificity, the contact or combining spots on the two particles must be *multiple* and *weak*. Furthermore, those on one particle must have a geometrical arrangement which is complementary to the arrangement of those on the other" (17). Chothia & Janin revealed, 25 years later, how these two principles were manifested in three protein structures: the insulin dimer, the trypsin-PTI complex, and the  $\alpha\beta$  oxy-hemoglobin dimer (14). They noted that these three interfaces are complementary in shape and rely on the shielding of many hydrophobic groups for stabilization. Subsequent surveys, incrementally larger as more structures were available, honed these principles (see 60 and references therein).

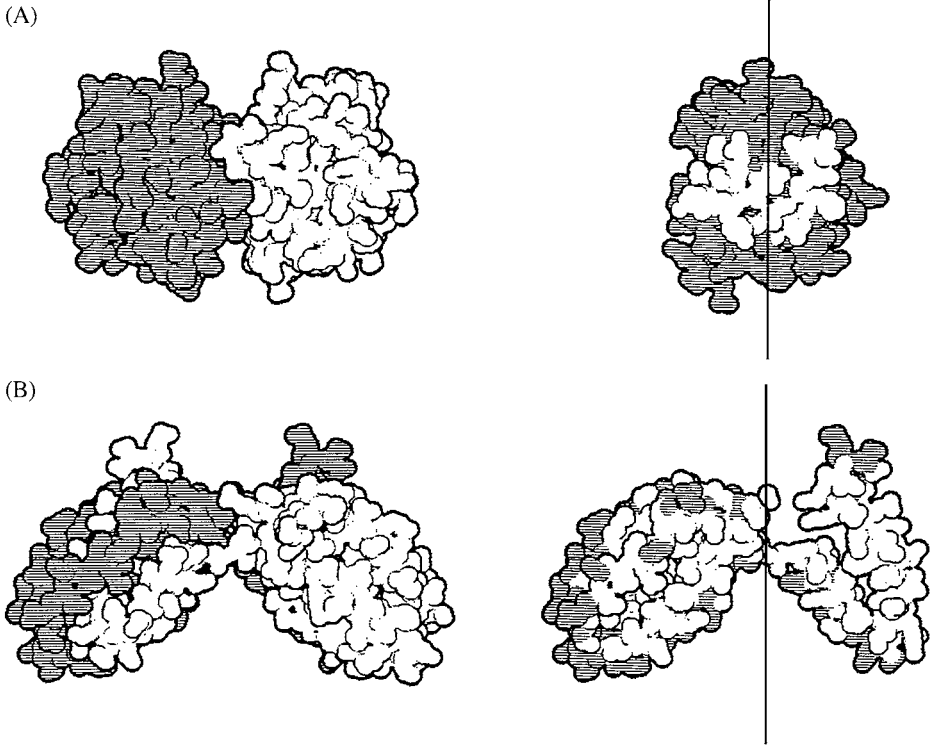
Interfaces may be broken into two broad classes: interfaces between globular subunits and interlocked interfaces. Note that the line dividing these two classes is often fuzzy. Most interfaces are formed between globular subunits that presumably fold as single subunits and then associate to form the oligomer. Superoxide dismutase (Figure 2a) is an example of this type of complex; the two subunits are essentially squashed spheres, pressed together to form the dimer. Interlocked interfaces, the second type, are composed of subunits that adopt much of their folded structure only after forming the complex. Dimeric cytokines such as interleukin 10 (Figure 2b) are excellent examples; it is hard to imagine any structure in these subunits in the absence of the dimeric complex. The characteristics of these types of interfaces are quite distinct (60). Interfaces between globular subunits may or may not have hydrophobic cores, and most have a surprising amount of water scattered throughout the interface. Interlocked interfaces are indistinguishable from the interior of proteins: They often show  $\beta$  strands intercalated into sheets of their neighbors; together, the subunits form extensive hydrophobic cores; and buried water is relatively rare.

Interfaces that require extra stability may incorporate covalent attachments or metal sites. Antibodies use disulfide linkages to glue their two flexible halves together. Metal ions are particularly common at the center of cyclic and dihedral oligomers with three-fold or greater symmetry. These ions are typically coordinated to symmetry-related sidechains from each subunit, sitting directly on the rotational axis of symmetry. Examples include a zinc ion that stabilizes the insulin hexamer (2, 7) and divalent cations that stabilize viral capsids around three-fold axes (74).

## WHY BUILD SYMMETRICAL PROTEINS?

Given that oligomeric proteins are very common, there must be some selective advantage driving the evolution of monomeric species into oligomers. This question has been discussed by many authors (see 49 and references therein) and was





**Figure 2** Interfaces of oligomeric proteins. Protein-protein interfaces are highly specific, being formed of several dozen amino acids on the surface of each subunit. They are typically complementary in shape and chemical nature. Interfaces are most commonly formed between two globular subunits, as in superoxide dismutase (A). On the left is the dimeric protein, with one subunit shaded. On the right is one subunit, rotated to show the interface region, with interface amino acids in white and solvent-exposed amino acids shaded. A minority of interfaces are formed by chains that interlock extensively, as in interleukin 10 (B). Again, the dimer is on the left and the single subunit, colored to show the interface, is on the right.

perhaps most succinctly answered by Monod, who considered the driving forces for formation of symmetrical oligomeric complexes to be "...finiteness, stability, and self-assembly" (71). The only major topic missing from this list is the novel functional possibilities presented by oligomers. We divide the problem into a series of questions: (a) Why build large proteins; (b) why build oligomeric proteins, and (c) why build *symmetrical*, oligomeric proteins?

The evolution of oligomeric proteins is bounded by two opposing forces. On one side, protein function typically drives evolution toward larger sizes. On the other hand, the mechanisms of protein synthesis tend to limit the length of polypeptide chains, favoring smaller proteins over larger ones. The typical size of 30,000–50,000 Daltons (87) is a compromise between these opposing forces.

## Why Build Large Proteins?

Large proteins have many advantages over smaller ones:

1. *Morphological function* Many proteins have functions that require creation of very large, stable structures. These include long, thin structural elements and large, hollow capsids and rings. Other proteins simply need to be sterically large. These functions are described in detail in a later section.
2. *Cooperative function* Allosteric and multivalent associations are other functions that create an evolutionary force selecting large proteins with several identical active sites rather than monomeric proteins with a single active site. These functions are described in detail in a later section.
3. *Stability against denaturation* Large proteins, with their extensive internal interactions, have more stable folded structures than very small proteins. Monod articulated this advantage: "...wherever order depends on very weak interactions, it must be bought at the price of increasing number of these interactions," where by "order," Monod referred to a stable, stereospecific globular protein that folded spontaneously to the active conformation (71). Protein stability involves a fine balance between the enthalpic stabilization by many weak nonbonded interactions and the competing effect of various entropic factors of conformational mobility and solvation (65). In small proteins, the enthalpic compensations of ordered structure are not sufficient to offset the entropic cost of conformational restriction. Proteins that are restrained to small sizes by their function (described below) thus use more extreme means to achieve stability, such as covalent disulfide linkages or specific metal sites.
4. *Reduction of surface area* In general, it is preferable to reduce the protein surface area that is exposed to solvent, by creating a large protein with several identical active sites, versus several individual proteins. This may be accomplished in several ways: by creating a long polypeptide of one chain with several functions, by creating an oligomer of several nonidentical subunits, or by creating a homooligomer of identical subunits.

Reduction of surface area reduces the amount of solvent needed to hydrate proteins. A quick calculation can estimate the magnitude of this problem. Assume that the aqueous cytoplasm is composed entirely of 20,000-Dalton subunits and that all oligomers are spherical in shape. Thus a monomer would be a sphere of radius 1.8 nm, a dimer would have a radius of 2.2 nm, and so on. Assume that the cytoplasm is 20% protein (24) and that the bound water of hydration is  $\sim 1.4$  g/g of protein (15) or a hydration shell  $\sim 0.6$  nm thick. If the aqueous cytoplasm is composed entirely of monomers, the hydrated proteins occupy 47% of the total volume, over twice the 20% volume occupied by the protein alone. Assuming this same 0.6-nm layer of hydration, the volume of the hydrated protein drops to 40% for dimers, 35% for tetramers, and 30% for

dodecamers. Thus, oligomerization can significantly reduce the amount of water bound to protein surfaces. However, this may not be a major evolutionary driving force, because Clegg has shown that many cells can lose over half of their water without adverse effects (15).

The reduced surface area provided by an oligomeric protein provides protection from degradation. Both insulin and proinsulin form hexamers in storage granules, stabilized by central zinc ions (7). In proinsulin, the outer surface is well covered by the connecting peptide, which is cleaved on maturation. The mature insulin, still in hexameric form, then forms a crystalline granule within the storage vesicle, further shielding the molecule from protease digestion. The crystallization also serves to enhance the conversion reaction, by removing mature insulin from the soluble pool. When released into the blood, dilution, reduced levels of zinc, and higher pH cause the hexamer to dissociate into the biologically active monomers, which are rapidly degraded.

Reduced surface area also has been postulated to improve the diffusion of substrates to enzyme active sites. Substrates are thought to perform a two-dimensional diffusive random walk along the surface upon encountering an enzyme, leading to more productive encounters with the active site than simple three-dimensional diffusive encounters. The process has been documented in simulations of superoxide dismutase, in which the surface around the active site forms a "funnel" that collects substrate (84). It has been postulated that the dimeric state of the protein serves to hide the "unproductive" side of the enzyme.

Proteins with conserved function but a nonconserved oligomerization state might be examples of evolutionary selection based simply on the advantages gained by reduction of surface area. For example, protocatechuate-3,4-dioxygenase contains a conserved ( $\alpha$ - $\beta$ -Fe<sup>3+</sup>) heterodimer core, but forms oligomeric complexes with 4, 5, or 12 ( $\alpha$ - $\beta$ ) heterodimers in different species, all with no apparent cooperative interactions.

## Why Build Oligomeric Proteins?

As noted above, many functions favor large proteins, requiring either one physically large protein or favoring one large protein with several identical active sites over many smaller proteins with individual active sites. These large proteins may be constructed in one of three ways: as long, single chains; as heterooligomers of several smaller chains; or as homooligomers of identical chains. As shown in Table 1, Nature favors the latter choice, most often constructing large proteins from many identical building blocks. Describing viral capsids, Crick & Watson write "...the virus, when in the cell, finds it easier to control the production of a large number of identical small protein molecules rather than that of one or two very large molecules to act as its shell" (18). Several reasons that homooligomers are favored have been proposed:

1. *Error control* By building a large complex from many small subunits, translation errors may be reduced by discarding subunits with defects, providing an extra step for proofreading. These errors have been characterized and quantified in prokaryotes (reviewed in 57, 75). Missense errors, which change an amino acid at a given position, have been estimated to occur at an average frequency of  $\sim 5 \times 10^{-4}$  per codon. Approximately one in four proteins of 500 amino acids have a substituted amino acid, and proteins with 2000 amino acids nearly always have an error. Missense errors, however, are fairly harmless; the vast majority of single-site mutations cause only a modest decrease in the protein's functionality. In a study of mutant bacteria with error-prone EF-Tu, a doubling of the missense error rate causes only a 10% decrease in bacterial growth rate.

Processivity errors, in which translation is terminated prematurely to yield a truncated protein, have a more significant impact. The prokaryotic processivity error frequency has been estimated at an average of  $3 \times 10^{-4}$ /codon, so about one in seven proteins with 500 amino acids will be released before it is fully translated, and a protein with 3000 amino acids will only rarely be translated in full.

2. *Coding efficiency* Homooligomers provide a genetically compact way to encode the information needed to build a large protein: Association of many individual small subunits allows the creation of a large structure with a minimum of genetic space. Crick & Watson (19) proposed this idea and predicted that spherical plant virus capsids are composed of many identical subunits on these grounds. They argued that the amount of RNA in these small viruses—making the then-unproven assumption of a three-nucleotide codon—is insufficient to encode a capsid composed of a large, single protein; therefore, they went on to predict (correctly) that these capsids are composed of subunits arranged with icosahedral symmetry. However, the large amount of noncoding DNA in eukaryotic genomes argues against this being a major driving force in higher organisms (95).
3. *Regulation of assembly* Large assemblies built of many identical subunits have attractive regulatory properties, because they are subject to sensitive phase transitions. For instance, actin is involved in many dynamic processes at the cell surface. A collection of actin-binding proteins control the nucleation, growth, termination, and disassembly of actin filaments, allowing fine spatial and temporal control (88). Similarly, microtubules spontaneously switch between phases of growth and shrinkage, in a behavior termed “dynamic instability” (69). This dynamic regulation of microtubule length may have important physiological implications for mitosis (70).

## Why Build Symmetrical Proteins?

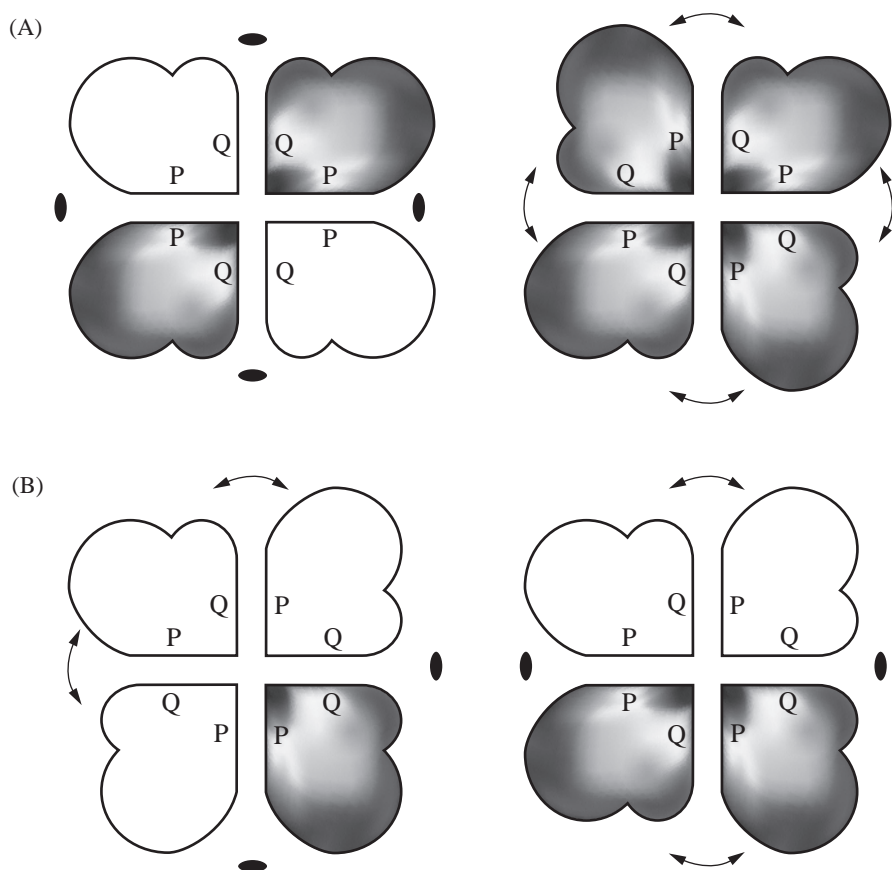
The homooligomeric proteins found in modern cells are also highly symmetrical, with soluble oligomers forming closed complexes related by simple point groups, and extended polymers showing helical symmetry. Several features favor

symmetrical complexes rather than asymmetric aggregates in the evolution of oligomeric proteins:

1. *Stability of association* Blundell & Srinivasan make an evocative comment in the proceedings of a recent symposium (8): "...generally, the lowest energy state of an assembly is a symmetrical one." This observation has been demonstrated in systems of identical particles, in which the particles interact by nondirectional forces. For instance, in clusters of noble gas atoms, certain highly symmetrical assemblies are favored (40). However, given that the symmetry of oligomeric proteins is under the control of evolution of function, this principle may not apply. For instance, if a nonsymmetrical protein is essential for function, a complex will evolve in which the nonsymmetric association is the lowest energy state. Just the same, many early analyses of quaternary structure tacitly assume that homooligomers adopt closed point group symmetries because they are optimally stable, with existence proofs as the justification (11, 49, 50, 72).

Cornish-Bowden & Koshland performed a thermodynamic analysis of oligomeric proteins to justify the prevalence of point group symmetries (16). They arranged four subunits into a square and defined two binding surfaces on each subunit, denoted P and Q. Two planar point group symmetries are possible, one related by twofold axes in the plane of the page, with P-P interfaces and Q-Q interfaces, and one related by four-fold rotation, with all P-Q interfaces (Figure 3). Two asymmetric closed complexes are also possible by combining local two-fold axes with local 90° rotations, containing a mixture of P-P, P-Q, and Q-Q interfaces. By surveying many different values for the binding strength of P-P, P-Q, and Q-Q interfaces, Cornish-Bowden & Koshland were able to show that the symmetric arrangements are favored quite strongly, even given only modest differences in the binding energy between the three types of interface interaction. In reality, the difference in interaction energies will be very large, so that oligomers with mixed symmetries requiring the formation of two types of interactions, like the P-P and P-Q pairings, are rarely observed.

The stability of closed, symmetrical oligomers is a consequence of two factors: (a) the specificity of protein-protein interfaces favors symmetrical complexes, and (b) the maximum numbers of intersubunit interactions are formed in closed complexes. The many structures of oligomeric proteins have revealed that protein-protein interaction sites, because they are composed of extended, complementary two-dimensional surfaces, are highly specific and directional. The directionality of protein-protein interfaces ensures that all homooligomers are symmetrical. A protein subunit is bound in one specific location and orientation relative to its mate; no relative rotation, slipping around this surface like a clutch, is allowed. This can be compared to the difference between carbon atoms in cycloalkanes and those in benzene. In cycloalkanes, the C-C bonds are relatively free to rotate, and most conformations of the molecule



**Figure 3** Four planar arrangements for a tetramer. Subunits with two interfaces, P and Q, may be arranged in a plane of the page in four unique conformations. Shaded subunits are flipped around two-fold axes in the plane compared with unshaded subunits. As described in the text, Cornish-Bowden & Koshland (16) used this model to validate the greater thermodynamic stability of the upper two symmetrical complexes over the lower two asymmetrical complexes.

are asymmetric. In benzene, however, delocalization over the C-C bonds disallows rotation, enforcing symmetry. The specificity and directionality of protein-protein interfaces ensures that each subunit will interact identically with its neighbors, limiting the transformations between neighbors to combinations of helical and cyclic (a special case with helical rise equal to zero) symmetries. (But, see the section on symmetry breaking below).

Given that homooligomers are symmetrical, with helical and cyclic symmetries, closed point group symmetries will give the maximal stability

over the entire oligomer. Caspar (11) proposed that “Specific bonding between the [identical] units necessarily leads to a symmetrical structure, since there will be only a limited number of ways to form the maximum number of most stable bonds,” where “bonds” refers to each unique protein-protein interaction surface. For instance, six subunits in an extended chain will have only five stabilizing interactions, with two less-stable subunits at the dangling ends, whereas a ring of six will have six protein-protein interfaces.

2. *Finite assembly* Proteins must avoid unwanted aggregation. Point group symmetry provides a method to create oligomers of defined copy number. Helical symmetries and other symmetries with translational elements are not bounded, however, and require special mechanisms to terminate growth. Several disease states seem to be the result of pathological aggregation of mutant proteins, such as sickle-cell anemia, Alzheimer’s disease, and prion-related diseases.
3. *Folding efficiency* Wolynes has speculated that symmetric protein structures provide fewer kinetic barriers to folding than do asymmetric structures (95). Based on analogies with simple clusters of atoms, he argues that the energy landscape for folding of symmetric complexes may be smoother than that of completely asymmetric structures.

## Functional Niches for Small Monomeric Proteins

In most cases, evolution appears to drive proteins to larger size and thus to symmetric, oligomeric complexes. In some specialized classes of proteins, however, functional considerations have the opposite effect, favoring small, monomeric proteins:

1. *Rapid diffusion* Cytochrome *c*, ferredoxin, plastocyanin, and other soluble electron transport proteins must be small and streamlined to diffuse rapidly to their sites of action in the crowded environment inside cells. Extracellular hydrolases, hormones, and many toxins are small for the same reason. These proteins are by and large monomeric, for the simple reason that it is difficult to create an oligomeric protein this small that can still fold to form a functional protein and remain stable under harsh environmental conditions.
2. *Stability at low concentrations* Oligomeric proteins are unstable at very low concentrations, so secreted proteins are commonly monomeric. Ricin is an interesting exception. It is a heterodimer, with a B chain that binds to the target cell surface and an A chain that inactivates eukaryotic ribosomes. The two subunits are connected by a disulfide bridge, but reduced ricin, if applied at concentrations at which the subunits associate, is even more toxic than the disulfide-linked complex (61). Apparently, the disulfide bridge serves primarily to hold the subunits together at the low concentrations found as the toxin diffuses to its target.



## SYMMETRY AND COOPERATIVITY

Oligomerization of proteins provides the opportunity for cooperative interaction between subunits. Allosteric regulation and multivalent binding are two advantages exploited by proteins.

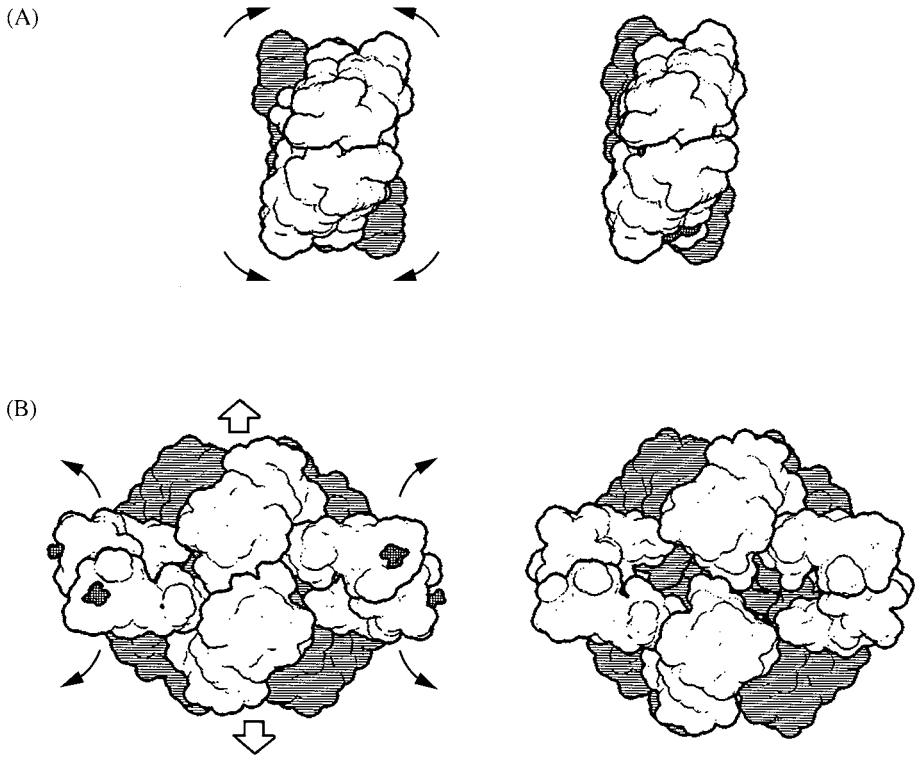
### Allosteric Regulation

Allosteric regulation encompasses two classes: (a) *homotropic*, in which binding of a molecule to one subunit modulates the binding of the same type of molecule to the other subunits, and (b) *heterotropic*, in which binding of an effector molecule changes the conformation of the protein, modulating the binding of a second type of molecule. Hemoglobin is a familiar example of homotropic cooperativity, and enzymes such as phosphofructokinase and aspartate carbamoyltransferase are examples of heterotropic regulation by allostery. The structural basis of allosteric regulation has been the subject of several reviews (23, 63, 67, 79).

Symmetry arguments played a central role in the formulation of the concept of allosteric regulation. The original model of Monod et al postulated two states, the relaxed (R)-state and the tense (T)-state (72). These two states are different in conformation and in their affinity for substrates, but the subunits within a given state are related by perfect symmetry. In the model of Monod et al, the oligomer cycles between two symmetrical states, all R to all T, and by assuming strict symmetry, an attractively simple mathematical model may be used to describe the behavior of the system. Soon thereafter, Koshland et al proposed a sequential model in which one subunit at a time converts from the R to the T state, forming a series of asymmetric intermediates between the fully symmetrical all-R and all-T states (55). The mathematics are necessarily more complex with sequential models. It remains a surprise that the original two-state model of Monod et al worked so well and continues to be a reasonable first approximation in many cases.

Allosteric regulation requires a molecular geometry that allows the passing of messages from one subunit to the next or, better, from one subunit to all of the rest [although allosteric regulation in monomers has been proposed, based on slow conformational changes and "memory" of the enzyme for its product-bound state (81)]. Perutz noted that there are few a priori constraints on the possible motions, even if one requires identical symmetry for the T and the R states (79). As long as all subunits shift similarly relative to the point group symmetry axes, the overall symmetry will be preserved. Nonetheless, many of the cooperative enzymes that show allosteric motions use dihedral symmetry and show motions that are easily related to the dihedral axes.

Many allosteric enzymes are composed of two rigid rings of subunits, which then associate back-to-back around a perpendicular two-fold axis to form the dihedral complex. The allosteric regulation occurs by rotating these two rings in relation to one another about the cyclic axis and/or translating them along the axis. Fructose-1,6-bisphosphatase, a tetramer with D<sub>2</sub> symmetry, shows a large motion



**Figure 4** Allosteric motions. Two types of allosteric motion are common in oligomeric proteins. The first is a rotational motion, exemplified by fructose-1,6-bisphosphatase (A). The upper dimer rotates by  $15^\circ$  in the T-to-R transition. The second is a pincher motion, exemplified by aspartate carbamoyltransferase (B). Regulatory domains at the right and left of the complex flex open and separate the catalytic subunits at the center.

of this sort (Figure 4a). The tetramer is formed of two stable dimers, and the R-to-T transition involves a rotation of  $15^\circ$  around the two-fold axis, piercing the stable dimers. Lifting the requirement for identical subunits, hemoglobin also fits this model, with the two stable  $\alpha\text{-}\beta$  pairs rotating independent units.

A second approach uses a "pincher" motion similar to the changes seen in bacterial repressors. In the repressors, the effector binds at the interface between two subunits in the dimer, causing the molecule to flex, repositioning the DNA-binding elements at the tips of the subunits. In allosteric proteins with dimeric symmetry, this type of motion is used for heterotropic regulation. The effector binds at or near the dimer interface, causing the two subunits to flex, and changes in conformation propagate to the active site, which may be quite distant from the effector site. Examples include glycogen phosphorylase (3) and chorismate mutase (89).

Allosteric enzymes with dihedral symmetry also show this type of pincher movement. Dimeric units within the complex perform a similar flexing upon effector binding. Often the active site is at the dimer interface that is remodeled in the transition. In aspartate carbamoyltransferase (29) with D3 symmetry, two trimers stack on one another and are connected by three pincher-type interactions at the points of a triangle. The regulatory motion separates and rotates the two trimers along the threefold-symmetry axis (Figure 4*b*). Bacterial L-lactate dehydrogenase (41) with D2 symmetry may be thought of as two dimers. Each dimer contains a single effector binding site that lies on a dimer axis. Binding of fructose-1,6-bisphosphate to this site effects a pincher-type motion between two subunits, shifting the orientation of one dimer relative to the next by  $\sim 6^\circ$  (note that the binding of a single fructose-1,6,-bisphosphate molecule to a dimeric enzyme site is an example of symmetry mismatch and pseudosymmetry, as described below).

Other enzymes with more complex functions use less easily characterized motions. These enzymes still use dihedral symmetry, presumably because of the intimate nature of contacts possible, but they add intrasubunit-domain motions to increase the vocabulary of motion that may be used. GroEL, with 14 subunits in D7 symmetry, is an excellent example. The subunits within one ring show positive cooperativity in the binding of ATP, whereas the two rings show negative cooperativity, with ATP hydrolysis in one ring promoting ATP binding on the other. Large domain rotations have been observed both in ATP binding and in the binding of the protein effector GroES (97). Pyruvate kinase similarly shows extensive rearrangement of domains within each subunit during the allosteric transition (68).

Dihedral symmetries appear to be far more conducive to allosteric regulation than do the higher cyclic symmetries alone. The two solutions used by many allosteric enzymes—rotation of two rings around the highest symmetry axis in a dihedral group and pincherlike motions—are both consistent with dihedral symmetry, but not with cyclic symmetry alone. The rarity of allosteric proteins with exclusively cyclic symmetry may result from the relative inefficiency of information transfer; the allosteric transition must propagate one subunit at a time around the cyclic ring. Cyclic symmetries are used in allosteric proteins only when necessitated by the function. The gap junction is an example; an irislike motion regulates the diameter of the pore.

## Multivalent Binding

Cross-linking proteins rely on two or more functional sites arranged to maximize their interaction with their targets. Structural cross-linkers have very specific shapes and symmetries that suit their function, and flexibility is often a key feature, allowing some latitude in the relative orientations of the objects linked together. Actin-binding proteins of several shapes are used to build different cellular structures (66). Actin-bundling proteins, such as  $\alpha$ -actinin, are short rods with binding sites at each end. They link actin filaments into parallel bundles, for use in motility and the shaping of cellular membranes. Gelation proteins, on the other hand,

are typically large, flexible molecules with several actin-binding sites, which link filaments into polyhedral networks and give cytoplasm its gell-like nature.

Multivalent binding also increases the binding strength of a molecule to a single target, when the target displays multiple sites for binding and the protein contains several discrete binding sites. The overall binding strength is improved by reduction of entropy. Once one site of the molecule has bound, the other sites are held in close proximity to the target, making binding far more likely. Immune recognition takes advantage of this cooperativity (20, 30). Immunoglobulin, complement C1, and C-lectin mannose-binding proteins are designed to recognize a bacterial or viral surface and thus search for targets with several sites of binding within a given distance.

This same type of “entropically favored” binding has been proposed as a possible means of inducing curvature in membranes (39). A polyvalent molecule binding to several sites on a membrane will induce a curvature concave towards the side of ligand binding. The energy of this interaction was estimated at about 0.1 kcal/mol for a divalent molecule, suggesting that an array of these interactions, such as the array of matrix proteins that mediate viral budding in retroviruses, would be needed for a biologically significant effect.

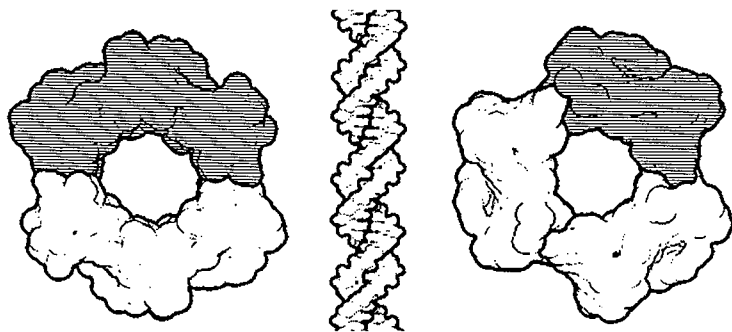
The functional roles of these molecules place severe restrictions on their shapes. For molecules that bind multivalently to large targets, such as antibodies to cell surfaces, the most efficient design has the binding sites oriented in a similar direction. In this way, the binding sites are arranged to bind to adjacent sites, and the “tail” of the complex is available for recognition by subsequent steps in immune recognition. The complement C1 protein is a case in point: It uses C6 symmetry to recognize bacterial surfaces. Given that the molecule will be built of six subunits, the C6 symmetry is far more effective than a D3 multipointed “jack.” Thus, molecules of the immune system are often built with rotational symmetry, but not higher symmetries. Cross-linkers, on the other hand, are most efficient when formed of two or more oppositely oriented binding sites, and thus they often show dihedral symmetries. Thus, the S-lectins from plants are effective agglutinators, because their active sites point in opposite directions and bind to targets on different cells.

## MORPHOLOGICAL FUNCTIONS OF SYMMETRY

Symmetrical oligomers of identical subunits are used for a wide range of morphological functions, in which the symmetrical shape of the complex is functionally useful.

### Rulers, Rings, and Containers

Symmetry is often used to create objects of a given size, for use as rulers to measure nanoscale distances, rings to surround molecular targets, or containers to enclose objects of a given size.



**Figure 5** Sliding clamps of DNA polymerase. A ring of six average-sized subunits or domains is perfect for surrounding a DNA strand, shown at the center. Two methods of creating this ring have been discovered: The  $\beta$  subunit of bacterial DNA polymerase III (*left*) is a dimer, with three similar domains in each subunit, and the eukaryotic processivity factor PCNA (*right*) is a trimer with two domains in each subunit. A single subunit is shaded in each.

Repressors use cyclic two-fold symmetry to create allosteric “rulers” that accurately measure the repeat length of DNA. The two-fold symmetry makes allosteric regulation particularly straightforward; like a pair of calipers, the binding of an inducer or effector at the “hinge” can change the distance between the two functional “tips.” These repressors carry with them the requirement for a palindromic DNA-binding site, matching the C2 symmetry of the repressor with the local C2 symmetry of DNA. A more detailed discussion of the restrictions imposed by DNA symmetry is included in a section below.

Three-fold and higher rotational axes of symmetry form pores or cavities that are often put to functional use. Kelman et al note that several proteins that encircle double-stranded DNA have six-fold rotational symmetry, with either six individual subunits or six similar domains (47). They argue that six-fold symmetry is the best compromise for the size of subunits (small enough to be economical and large enough to fold). The processivity factors of DNA polymerase (Figure 5) show approximate six-fold symmetry, with six average-sized domains arranged in a ring that clamps around the DNA strand. In the *Escherichia coli*  $\beta$ -subunit of polymerase III holoenzyme, two subunits assemble to form a ring, each with three domains (53). The eukaryotic processivity factor of DNA polymerase  $\delta$ , on the other hand, adopts an identical ring shape, but is composed of three subunits, each with two domains (56). Perhaps a hexameric ring of single-domain subunits has yet to be discovered in another organism.

The size of the cavity may be estimated by using a simple approximation. First, the radius of the subunit ( $R_{SU}$ ) is calculated (86):

$$R_{SU} = 3 \sqrt{\frac{3\nu M}{4\pi N_A}} = 0.0665(nm)3\sqrt{M},$$

where  $v$  is the protein partial specific volume of 0.74 ml/g,  $M$  is the molecular mass in Daltons, and  $N_A$  is Avogadro's number. They then assume that these spheres just touch in the complex, yielding a cavity size ( $R_{cavity}$ ):

$$R_{cavity} = R_{SU} \left( \frac{1}{\sin(\pi/n)} - 1 \right),$$

where  $n$  is the number of subunits in the ring. This probably overestimates the size of the cavity. We use a second approximation, which simplifies the extension of the approximation to cubic symmetries. We sum the diameters of the  $n$  subunits and use the sum as the circumference of the oligomeric ring on which the subunit centers lie. The cavity size is then estimated as:

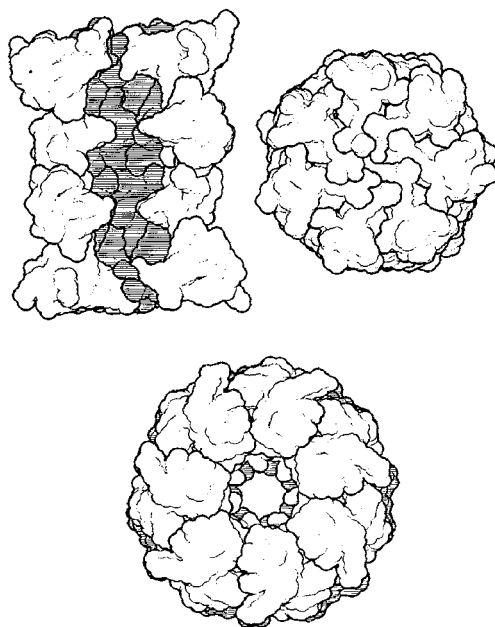
$$R_{cavity} = R_{SU} \left( \frac{n}{\pi} - 1 \right).$$

This calculation yields smaller cavity sizes, particularly for rings with three to six subunits, and it approximates the extensive contact between subunits. In fact, the calculation for a trimer yields a nonphysical negative value, which is consistent with many observed trimeric structures such as porin, which have protein atoms extending to the three-fold axis.

Oligomeric rings are used as pores through lipid bilayers. Examples include the connexon, which forms a six-fold ring; the complement membrane attack complex and perforin, both of which form rings of variable size; and the nuclear pore, a large complex of proteins with eight-fold symmetry. Surprisingly, the trimeric bacterial porin does not use the oligomeric symmetry to form its pore—instead, each subunit forms a separate pore through the membrane, bounded by a large  $\beta$ -barrel. Rings are also important for the creation of rotary motors. The large flagellar motor complexes of *E. coli* and *Salmonella typhimurium* are examples of motors used to power rotary motion, and ATP synthase is an example that is used oppositely as a generator. As discussed below, rotary motors are limited to cyclic symmetry and lower symmetries. The trp RNA-binding attenuation protein, TRAP, may hold a surprise. It is a ring-shaped complex of 11 identical subunits, which, under control of tryptophan concentration, negatively regulates the trp genes by binding to RNA. Mutagenesis studies have suggested that the RNA wraps around the perimeter of the ring, rather than threading through the hole (98).

Approximate ring structures may be built with other symmetries. For example, human  $\beta$ -trypsin is a tetramer with approximate D2 symmetry, stabilized by binding to heparin, which forms a large pore along one of the two-fold axes (78). The active sites are oriented inwardly, opening onto the pore and limiting access. This has been proposed as the explanation for the resistance of  $\beta$ -trypsin to most endogenous protease inhibitors. The processivity factors of DNA polymerase, mentioned above, are another example, with approximate C6 symmetry.

Monomeric proteins are used to store and transport individual small molecules—for example, the bacterial periplasmic binding proteins and metallothioneins, but, for trapping and storing larger molecules or larger numbers of molecules,



**Figure 6** D7 symmetry to enclose a protein. Two proteins at opposite ends of the life of protein have similar symmetry. *Top*, the chaperone GroEL shows D7 symmetry, forming a protein-sized cavity that assists in the folding of nascent proteins. *Bottom*, the proteasome also shows D7 symmetry, forming a cavity that degrades obsolete proteins.

symmetrical oligomers are used. These containers have attractive properties; they may be built at a defined size to act as a “sieve,” trapping only molecules of the proper size. They also may be built with defined chemical characteristics, creating a custom environment within the enclosed space.

The simplest solution is the use of cyclic symmetry to form a “cup,” but, more commonly, two of these cups assemble back-to-back, using dihedral symmetry. Two proteins with identical symmetry but opposite function incorporate this motif (Figure 6). The bacterial chaperonin GroEL shows D7 symmetry, forming two cups that guide the folding of immature proteins (97). The bacterial proteasome also uses D7 symmetry, forming two cups that degrade obsolete proteins (64). Eukaryotic proteasomes use a similar overall morphology, but they are composed of a collection of similar proteins arranged in pseudo-D7 symmetry (32). The choice of this unusual symmetry group by both of these proteins is not as surprising as it might seem; it is dictated by two structural constraints: the need for a cavity of 40–50 Å to house the substrate protein and the need to build the cavity from typically sized proteins. From calculations like those above, we might expect that D6, D7, or D8 would be able to accommodate these functional constraints.



Cubic symmetry is used to build containers of even larger size. Ferritin (Figure 1) uses octahedral symmetry to build a container for iron ions (91), and virus capsids use icosahedral symmetry to create even larger containers. When more space is needed, quasisymmetry may be used (see below). The lumazine synthase-riboflavin synthase complex of *Bacillus subtilis* is a particularly unusual application of an icosahedral shell. The complex performs the two final reactions in the synthesis of riboflavin. The complex is composed of an icosahedral capsid of 60  $\beta$  subunits, which carries out a condensation reaction that surrounds a trimer of  $\alpha$  subunits, which then perform the final dismutation to form riboflavin (59).

The size of the cavity at the center of icosahedral capsids may be approximated as above, by summing the areas of a great circle through the center of the subunits and then calculating a sphere with similar overall surface area. The radius of the icosahedral cavity ( $R_{ico}$ ) is then:

$$R_{ico} = R_{SU} \left( \sqrt{\frac{60T}{4}} - 1 \right),$$

where  $T$  is the triangulation number (described below). Calculated values for  $R_{ico}$  showed an rms error of  $\sim 15\%$  over a test set of 22 crystallographically determined capsids (data not shown). Similarly, cavities for tetrahedral and octahedral complexes may be estimated by using values of 12 and 24 instead of 60 and using an appropriate triangulation number.

## The Need to Be Large

Occasionally, the outer diameter of the complex may be functionally important. Koshland suggests that early cells needed large proteins to reduce loss through their leaky membranes (54). He postulates that early cells had not yet developed methods for active transport, so that large, primitive pores were the major method of transporting molecules into and out of the cell. Examples in modern organisms of the “need to be large” are difficult to find. The oxygen-carrying proteins of invertebrates, such as hemocyanin, form very large complexes. These proteins show allosteric control, but Perutz suggests that their high oligomerization state is to “prevent their passage through cell membranes,” or perhaps to prevent loss at cellular junctions (79).

**Structural Elements** Cells often use translational symmetries to build large structural elements from average-sized protein units. The most common method is the use of helical symmetry to construct filaments. Examples come in all sizes: thin filaments of actin, intermediate filaments of keratin and desmin, and thick filaments of tubulin. Filaments with simple helical symmetry are directional, in that one end may be distinguished from the other. This directionality is put to functional use in both actin and tubulin in their use as tracks for the molecular motors myosin and kinesin. Filaments that do not require directionality can incorporate

two-fold symmetry perpendicular to the helical axis. For instance, intermediate filaments are formed of dimeric subunits that assemble in an approximate  $4_2$  helix, overlapping in a lap joint for maximal strength. Bacterial flagella are unusual cases; they form corkscrew-shaped superhelices, even though they are composed of a single type of flagellin subunit. As described below, a clever method of symmetry breaking is thought to be the mechanism.

One potential problem with helical symmetries is the lack of boundaries—how does the cell choose the proper length? For cytoskeletal elements, a complex series of initiation and termination proteins controls the assembly and disassembly of subunits. For tobacco mosaic virus, the solution is more direct. The RNA packaged in the virion acts as a “ruler,” building a virion of defined length.

A remarkable exception to the use of modular helices as structural elements is the giant protein titin (58), the protein that limits the extension of muscle sarcomeres. With >27,000 amino acids and a molecular mass of 3 million Daltons, titin proteins extend over a micrometer in length. Because each one is a single protein, it is important as a ruler for defining the size of muscle sarcomeres. It also thought to contain a stretchable element, which adds elasticity to muscle cells.

## FUNCTIONS THAT LIMIT SYMMETRY

Many biochemical functions limit the level of symmetry that is possible, working in opposition to the gains obtained from higher oligomerization and thus higher symmetry states. The result is an evolutionary tug of war, yielding the optimal state for a given functional niche.

### Directional Motion

Directional linear motion places functional limits on the symmetry adopted by processive protein machinery. Polymerases and ribosomes perform a directional, asymmetric reaction and thus are themselves without point group symmetry. Note that individual subunits within these polymerases may have local symmetry, as for the sliding clamps mentioned above. The functions of these subunits, however, are nondirectional. Bacterial DNA polymerase III is an unusual exception to this observation; it is a large complex of enzymes that has, overall, approximately two-fold symmetry (37). The two polymerase subunits, one acting on the leading strand and one on the lagging strand, associate in the active complex, along with a helicase and several subunits that orchestrate the special needs of the lagging strand. This is an example of multivalent binding: The driving force for the dimerization of two polymerases is the advantage of having two enzymes tethered in one place, for acting on two strands of DNA that are guaranteed to be spatially close to one another.

For filaments used as tracks for molecular motors, simple helical symmetry is the highest that will allow unidirectional motion. Unidirectional motion along an

intermediate filament or a double-stranded DNA helix is disallowed by symmetry. The motors themselves are also limited in symmetry; they cannot have symmetry elements that intersect with the helix axis. Similarly, rotary motors are limited to cyclic symmetry. Motors with dihedral symmetry would destroy the directionality of motion, leading to a frustrated random walk.

## Interaction with DNA

One might speculate on the aspects of B-DNA structure that are not involved directly in its function. The helical symmetry of DNA is not specified by its function; it could adopt any form—left- or right-handed—and still transfer genetic information. The antiparallel orientation of the two strands, with local two-fold axes perpendicular to the helix axis at each base pair, is also not specified by function. Two strands in parallel, with a two-fold symmetry parallel with the helix axis, would also provide a mechanism for information transfer and would remove the need for discontinuous replication of the lagging strand (but might reduce the opportunities for control imposed by the helically wound antiparallel double helix). The B-DNA structure is an example of a design locked in at an early stage of evolution, perhaps not optimal, but unchangeable once incorporated.

The double-helical symmetry of DNA places limits on the symmetry of DNA-binding proteins. The sugar-phosphate backbones form a symmetrical double helix, but the local two-fold axes running through the center of each base pair, perpendicular to the helix axis, are broken at the atomic level by the nonidentity of bases in each pair. Thus, DNA-binding protein may interact at several levels. Proteins that bind nonspecifically to the backbone might be expected to show dimeric symmetry. In fact, however, this is relatively rare. Most non-base-specific functions use other symmetries: Nucleases are primarily monomeric, and remodeling proteins, such as the nucleosome, are complex oligomers. The dimeric symmetry of DNA is used primarily in the binding of bacterial repressors and restriction endonucleases, where the symmetry of the DNA backbone is mirrored in a self-complementary base sequence.

Alternatively, many monomeric proteins interact specifically with a given DNA site, binding directionally to the local sequence and ignoring the overall symmetry of the DNA backbone. Many eukaryotic transcription factors fall into this class. The TATA-binding protein is an interesting hybrid of the two approaches (44). At one level, it has two-fold pseudosymmetry, with two similar domains binding to symmetric DNA backbones. At another level, however, it recognizes and bends DNA in a sequence-specific and directional manner, initiating transcription in the proper direction from the TATA sequence.

## Membrane Interactions

Biological membranes are nearly always asymmetric in function, separating “inside” from “outside.” The functions of membrane proteins are also asymmetric.

Receptors must distinguish between extracellular and intracellular sides, and directional transport must be mediated by a protein that knows inside from outside. Thus the proteins interacting with these surfaces and those embedded in the membrane are nearly always limited to cyclic symmetries, with the cyclic axis perpendicular to the plane of the membrane.

A few functions do not require this specificity, and one might expect to find examples with higher symmetries. Channels that allow passive bidirectional transport might show dihedral symmetries. The gramicidin channel is one example, with two-fold symmetry such that the axis of symmetry is parallel to the surface instead of perpendicular. Thus far, larger protein channels such as aquaporins, have shown cyclic symmetry.

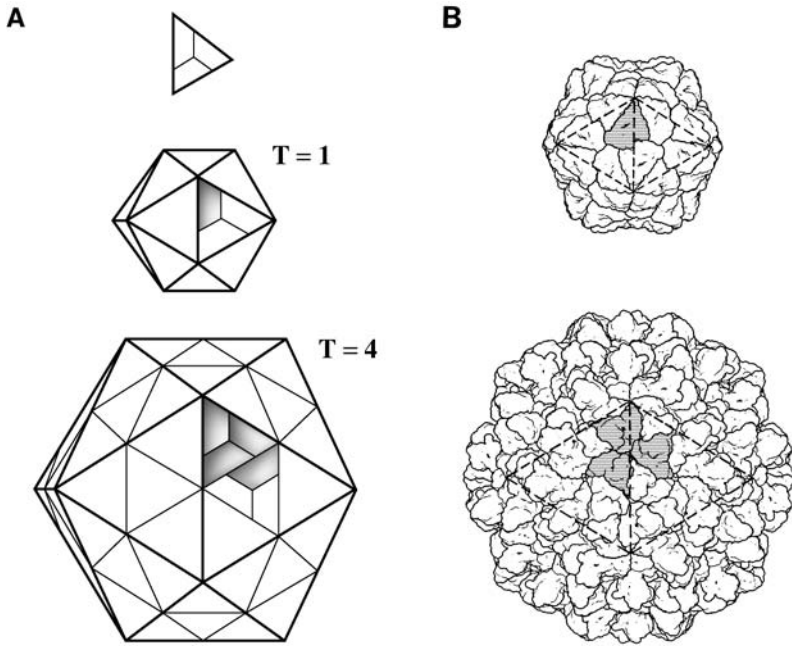
## SYMMETRY BREAKING IN OLIGOMERIC PROTEINS

Biological molecules often break from perfect symmetry to accomplish specific functional goals. There are cases in which identical subunits adopt similar but different positions—this is termed quasisymmetry. Taken to extremes, identical subunits may be used to build several different structures, which is termed pleomorphic symmetry. On the other hand, there are cases in which similar but different subunits perform identical roles—this is termed pseudosymmetry. And finally, in large complexes, components with different symmetry may be fitted together, forming a symmetry mismatch locally at the interface.

### Quasisymmetry

Most viruses require capsids that are larger than can be created by 60 identical, moderately-sized proteins in perfect icosahedral symmetry. Some viruses answer this need by creating capsids with multiple chains: For instance, poliovirus and rhinovirus capsids are composed of 60 copies of each of four different chains, all arranged in perfect icosahedral symmetry. Other viruses, however, have taken a more creative approach, using a single protein in several different quasiequivalent structural roles. In these viruses, each of the 60 symmetrically identical positions in the icosahedron is filled by a number of identical chains.

As quasisymmetry was originally conceived, these chemically identical chains adopt positions that are approximately identical in the local environment, and small elastic deformations of the subunits allow similar contacts to be formed between each. As stated by Klug: "... if each subunit in the final structure still forms the same types of sets of bonds with its neighbors, then, although the units are no longer exactly equivalently related, they may be said to be quasi-equivalently related" (51). Caspar & Klug described a method for tiling icosahedra with triangular networks, creating steadily larger capsids composed of quasiequivalent triangular subunits, thus introducing the concept of a "triangulation number" (12), as shown in Figure 7A and 7B. Using these triangular lattices, they showed that



**Figure 7** Triangulation number and quasisymmetry. *A*. With the basic triangular unit shown at top, composed of three subunits, a wide variety of quasisymmetrical icosahedral shells can be formed. The basic  $T = 1$  icosahedron is the simplest. All subunits are identical—one subunit is shown shaded at center. A larger shell may be formed by creating an icosahedron with four triangular units tiled within each icosahedral face, as shown at the bottom. Then, each of the triangles is not identical, and there are four unique subunits (shown shaded) in different local environments. Many other triangulation numbers are possible, by tiling the triangular units differently within the icosahedral geometry. *B*. Two examples of viruses. *Top*, satellite tobacco necrosis virus, with  $T = 1$  symmetry and a single subunit in the asymmetric unit; *bottom*, tomato bushy stunt virus, with  $T = 3$  symmetry and three subunits in the asymmetric unit. Two faces of the underlying icosahedron are shown with *dotted lines*.

quasiequivalent lattices could be constructed for shells with  $60T$  subunits, where  $T = h^2 + hk + k^2$  and  $h$  and  $k$  are integers.

Some hint of the tolerances involved in assembly of large complexes from subunits that allow small local deformations is provided by the study of a bacteriophage portal protein (92). This protein forms 12- or 13-fold cyclical oligomers. It assembles sequentially, with subunits adding at an angle of  $25.8^\circ$  around the ring, just shy of the angle needed for a 14-fold ring. Apparently, small deviations in this angle allow the ring to close 40% of the time into 12-fold rings ( $30^\circ$ /subunit) and 60% of the time into 13-fold rings ( $27.7^\circ$ /subunit).

The structures of viral capsids have revealed that this ideal for quasiequivalence is only rarely achieved. Although the triangular-network model has been

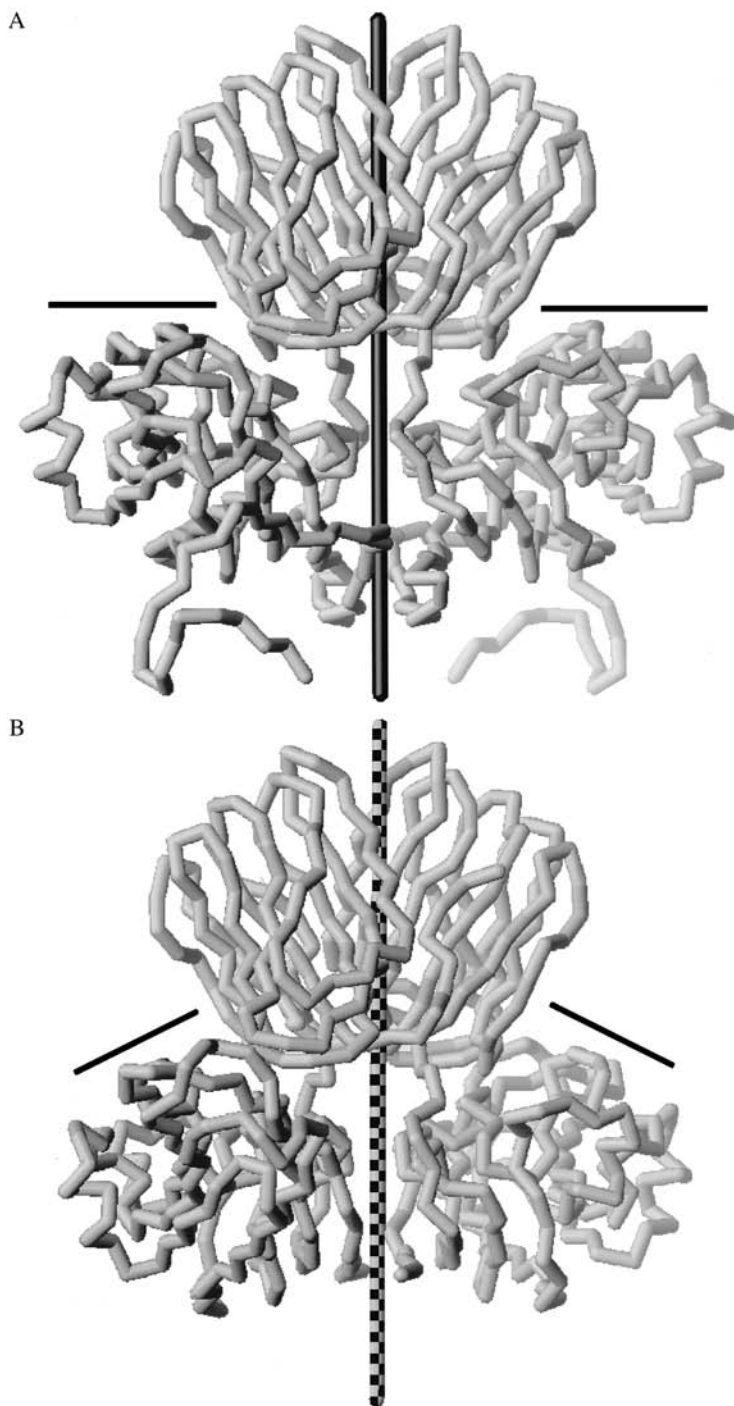
observed in nearly every icosahedral virus structure, the concept of minimal changes and elastic deformations in structure between different quasiequivalent subunits has fared less well (82). In reality, the viral protomers tend to show structural “switches,” adopting two or more significantly different conformations that mediate the different quasiequivalent contacts. There is considerable literature describing the structural features used by viruses, as observed by electron micrograph reconstruction and X-ray crystallography. Based on these works, there has been a recent rebirth of interest in the field, as mutagenesis and computational chemistry are applied. We do not attempt to review this burgeoning field, but we do touch on a few interesting highlights. For more information, the reader might start with a recent review by Johnson & Speir (42).

Tomato bushy stunt virus (TBSV) is a classic example of a structural switch (Figure 8), allowing a single type of subunit to adapt to three different environments in a  $T = 3$  quasisymmetrical capsid (35). The subunit is composed of an N-terminal shell domain located at the interior of the capsid, connected by a flexible linker to a C-terminal domain that projects from the capsid surface. An N-terminal region in the subunit undergoes an order-disorder transition based on the local symmetry environment, being ordered in subunits that associate around icosahedral two-fold axes and disordered in subunits that associate through quasisymmetrical twofold axes. The subunits make extensive contact with one another, and additional stability is obtained by the ordered N-terminal arms, which associate between three subunits at the quasi-six-fold axis.

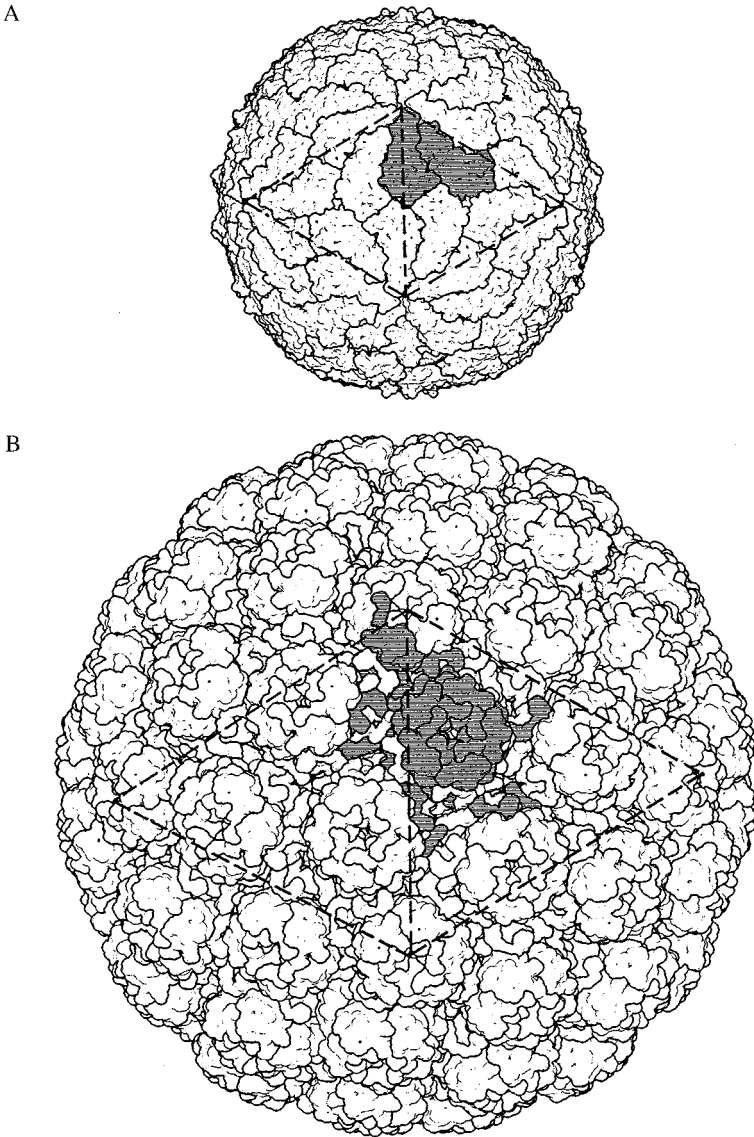
Bluetongue virus (BTV) shows an interesting variation on quasisymmetry (Figure 9a). The virus is composed of several concentric protein shells (31). The outer shell adopts a  $T = 13$  arrangement of surprising regularity, conforming closely to the quasisymmetrical ideal. The inner shell is more unusual. It adopts a  $T = 1$  arrangement, but has two separate subunits in each equivalent position. These two subunits are bean shaped and are packed back-to-back in nonequivalent, asymmetric conformations. This arrangement is discussed in more detail below, under the heading of “Global Asymmetry.”

**Figure 8** Structural switch in tomato bushy stunt virus. The three subunits in the asymmetric unit of tomato bushy stunt virus adopt slightly different conformations to accommodate the geometric requirements of the quasisymmetrical positions. Two pairs of subunits are shown here, with the inside of the virion facing downward in each. Each subunit is composed of two domains, connected by a flexible linker. The C-terminal domain (at the top in each) forms a structure that protrudes from the spherical capsid, which is formed by the N-terminal domain (at the bottom in each). A, subunits arranged around the strict dimer axis that runs through each icosahedral edge, with ordered N-terminal arms extending from the bottom of the complex; B, subunits arranged around a quasisymmetrical two-fold axis that intersects the icosahedral face, with their disordered N-terminal arms. Notice the difference in the orientation of the N-terminal domains, forming a flatter capsid surface in A (shown by the lines) and a more curved surface in B.









**Figure 9** Broken quasisymmetry in viruses. A, bluetongue virus, which can be thought of as a  $T = 1$  virus with two subunits in the asymmetric unit (*shaded*). These two subunits are packed back-to-back and adopt significantly different conformations. B, simian virus 40 can be thought of as a  $T = 1$  virus, with six subunits in the asymmetric unit, shown shaded. One of the six forms a pentamer with neighbors at the icosahedral five-fold axis, and five form a pentamer on the icosahedral face. The pentamers are connected by flexible arms, seen surrounding the shaded pentamer.

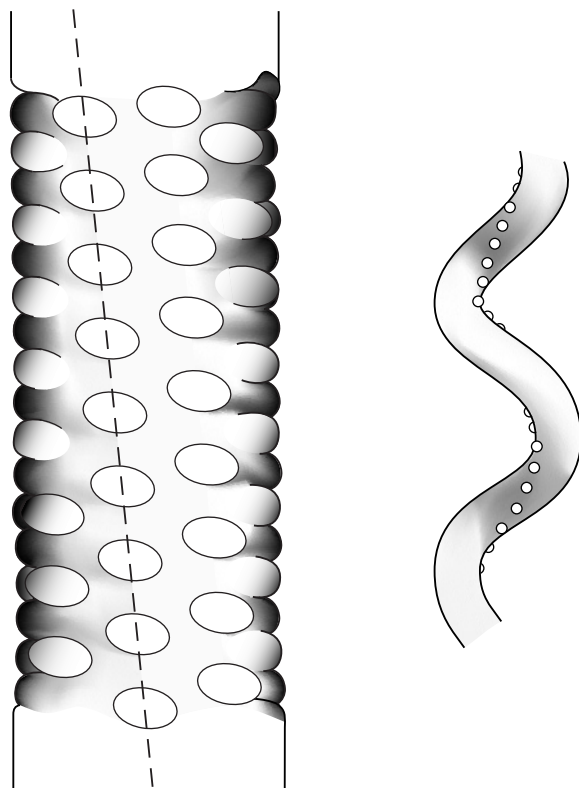
Simian virus 40 (SV40) is an extreme case, straining the concept of quasisymmetry to its limits. The subunits form stable pentamers with extensive intersubunit contacts, but nearly all of the pentamer-pentamer contacts are mediated through swapping of C-terminal arms (62). The pentamers form a shell with apparent  $T = 4$  quasisymmetry, but with a full pentamer at each vertex of the triangular lattice instead of an expected trimer. The flexibility of the C-terminal arms allows this unusual break of symmetry to occur, as shown in Figure 9*b*. In some cases, arms are swapped between two pentamers, around a true icosahedral two-fold axis or a quasisymmetrical two-fold axis. Interactions between pentamers sitting on the icosahedral five-fold axes and their neighbors, however, form a three-way swap between three pentamers.

Bacterial flagella provide another interesting example of quasisymmetry. Flagella are long superhelical filaments used to propel bacteria through their environment (93). Amazingly, these superhelical structures are often composed of a single type of protein subunit, termed flagellin. This requires that identical subunits adopt nonidentical environments, such that the subunits on the inner face of the superhelix are more crowded than those on the outer face. The current model of flagellar structure requires flagellin to adopt two states (10, 45). When flagellin self-assembles into the filament, it forms distinctive longitudinal columns, seen by electron micrograph reconstructions and X-ray diffraction (Figure 10). If an entire column switches conformation from a longer form to a shorter form, one side of the filament is compressed, forcing the entire filament to adopt a superhelical form. This model has been quite successful in predicting the various polymorphs observed under different experimental conditions. As different numbers of longitudinal columns switch from one conformation to the other, different straight or curly filaments are formed.

## Pleomorphic Symmetry

Nature has also taken advantage of the idea of building several different structures from a single type of modular subunit. The term “pleomorphic” has been borrowed from chemistry, referring to compounds that crystallize in several different habits. In the present context, the term refers to subunits that assemble into different structures.

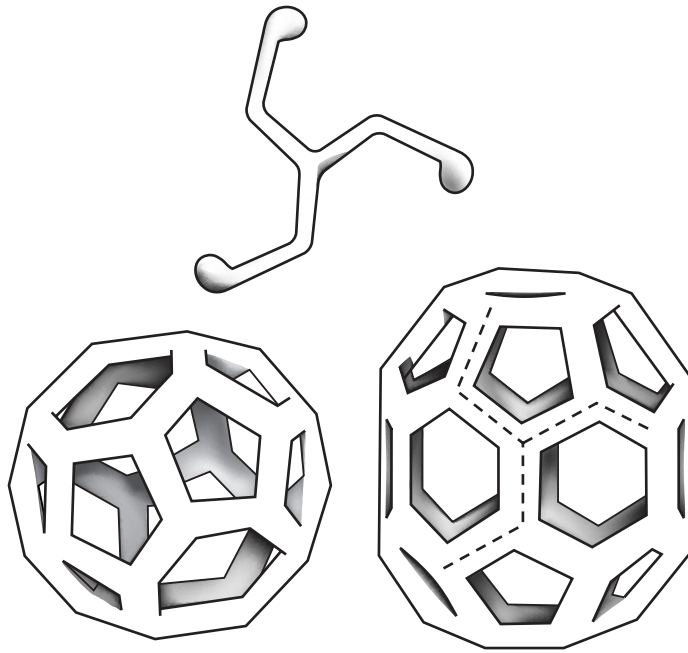
A popular example of pleomorphism with covalent bonds is the spectrum of different buckminsterfullerines, in which a single carbon atom is the subunit. By construction of lattices with six-fold and five-fold rings, a wide range of symmetrical closed spheres and tubes, as well as diverse of asymmetric structures, may be constructed. The key that makes this diversity of structure possible is the ability to form C-C linkages with a range of different C-C-C bond angles. If the covalent bonding chemistry were so specific that bond angles were rigid at  $120^\circ$ , only hexagonal nets would be possible. But since the allowable range is wider, easily allowing six-fold and five-fold rings to form, a larger range of geometries is available to the final structures.



**Figure 10** Model of flagellar superhelices. *Left*, flagella are formed from helical polymers of flagellin. Flagellin adopts two conformations, one of which has a shorter helical rise per subunit. *Right*, if an entire row of subunits, shown with the dotted line, shifts to the shorter conformation, it will distort the filament into a superhelix.

Several biological molecules show similar pleomorphism. Perhaps the most familiar example is clathrin (Figure 11). The clathrin subunit is a trimeric triskelion, with partially flexible arms. These triskelions then assemble, arm binding to arm, to form closed geodesic structures that mediate the invagination of coated pits. The structures show a wide range of geometry, from simple icosahedra to elongated ellipsoids, with five-fold and six-fold rings reminiscent of the smallest buckminsterfullerenes (38).

Many viruses also show pleomorphic forms, particularly when the environmental conditions are changed during assembly. Cowpea chlorotic mottle virus is normally a  $T = 3$  icosahedron, but changes in pH and ionic strength can cause the virion to reassemble into a variety of tubes, sheets, and multishelled particles (42). The HIV capsid may also be an example of pleomorphism. In a recent model based on electron microscopy of reconstructed cores, the cone-shaped HIV core



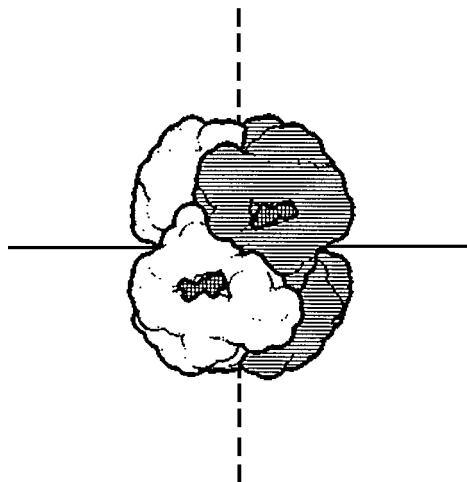
**Figure 11** Clathrin pleomorphism. Clathrin coats are formed of three-armed triskelions (*top*). These arrange into a variety of geodesic structures. Smaller structures are rich in five-fold rings (*left*), and larger structures add sixfold rings (*right*). The proposed location of a triskelion is shown in *dotted lines* in the right structure.

is proposed to be constructed of two different structures: a hexagonal network, rolled to form a cone, and two fullerene-type hemispherical caps at each end (25).

## Pseudosymmetry

Pseudosymmetry refers to oligomers composed of two or more types of similar chains, such that the entire complex resembles an homooligomer. Hemoglobin is a familiar example, composed of two  $\alpha$  and two similar  $\beta$  subunits. The  $\alpha_2\beta_2$  tetramer is strictly dimeric, but shows approximate D<sub>2</sub> symmetry, as shown in Figure 12. Mammalian lactate dehydrogenase is another example, with oligomers formed with various mixtures of muscle and heart isoenzymes.

In the retroviral proteases, a symmetrical enzyme interacts with a pseudosymmetric substrate. These proteases are homodimers that form a long, two-fold-symmetric active site, with alternating hydrogen bond acceptors and hydrogen bond donors arrayed along its length. The substrate is a typical peptide with no internal symmetry. The peptide binds in extended form, interacting in a pseudosymmetrical manner with all of the hydrogen-bonding groups available from the protein (94). Bacterial L-lactate dehydrogenase is another example. The enzyme



**Figure 12** Hemoglobin pseudosymmetry. Hemoglobin is a heterotetramer of two  $\alpha$  subunits, shown in *white*, and two  $\beta$  subunits, shown *shaded*. The heme groups are shown in *darker shading*. A perfect two-fold axis relates the two  $\alpha$  and the two  $\beta$  subunits, shown as the horizontal line. Since the  $\alpha$  and  $\beta$  subunits are similar in sequence and structure, they may be related by pseudosymmetry, forming pseudo-two-fold axes vertically, shown with a *dotted line*, and perpendicular to the plane of the paper through the center of the tetramer.

is a homotetramer with D<sub>2</sub> symmetry, but only two molecules of the allosteric activator fructose-1,6-bisphosphate bind per oligomer. The binding site is formed by two subunits and is located exactly on one of the two-fold axes. The approximate dimeric symmetry of the substrate, with two phosphate groups extending in opposite directions from a compact sugar, facilitates binding to the two-fold symmetric site (41). Note that these interactions are also examples of symmetry mismatch, described below.

The human growth hormone receptor is an example of a pseudosymmetric association (22). The hormone itself is a small, monomeric protein. At the surface of a cell, it binds to two receptor molecules, forming a complex with approximate dimeric symmetry. The two receptor molecules bind to different faces of the hormone molecule, using similar binding surfaces on the receptor. One might expect that a dimeric hormone would be easier to develop evolutionarily. Perhaps the need for a small hormone, combined with the difficulty in creating a stable dimeric protein of small size, favors the evolution of a monomer.

## Symmetry Mismatch

Large molecular complexes are often built of many different protein species. In some cases, such as the ribosome, the complex is built of many different proteins (and nucleic acids) associating in an asymmetric manner. In other cases, the entire complex adopts a given symmetry, with identical numbers of each protein

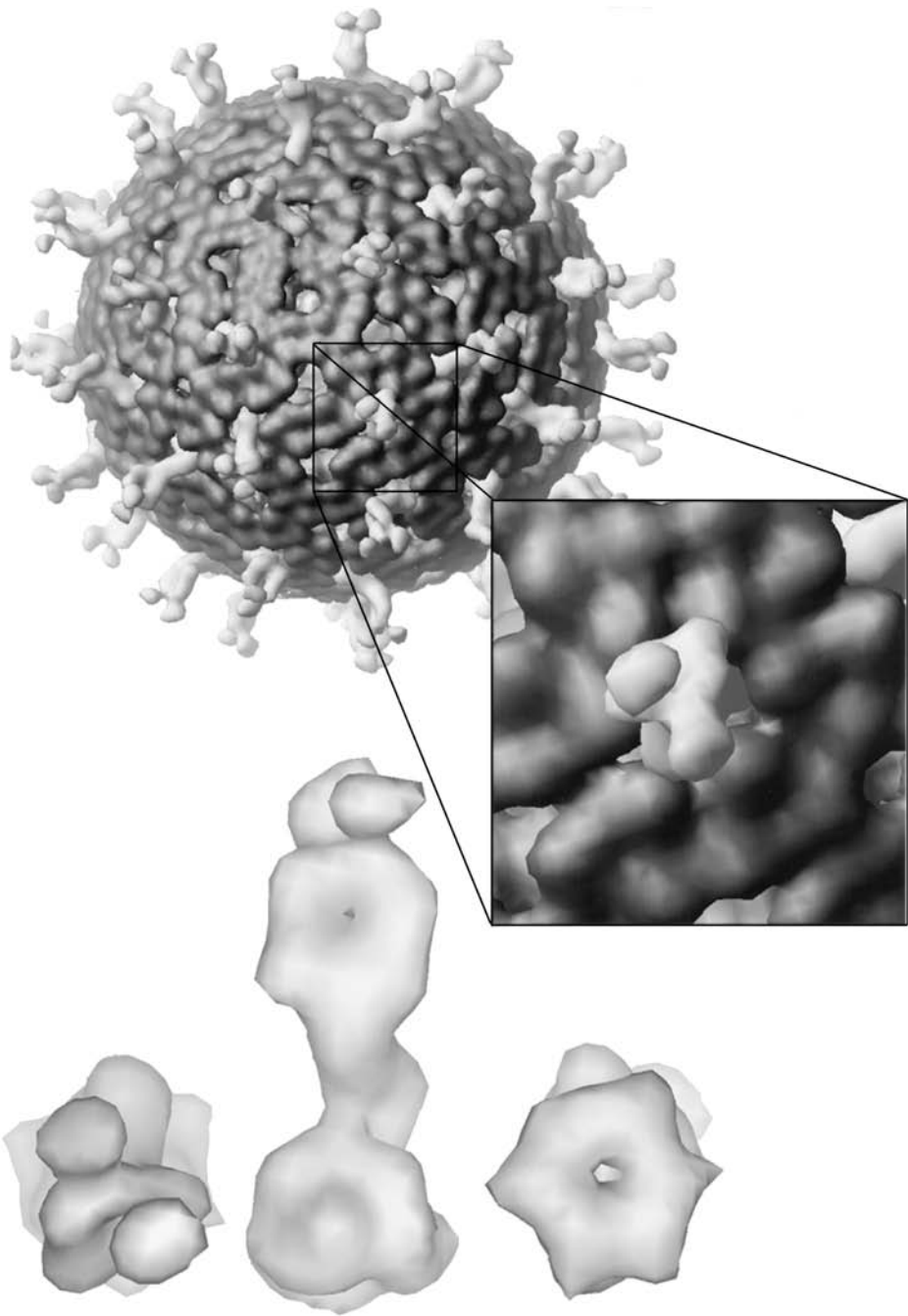
chain. The picornavirus capsids, formed of 60 copies of four different chains, are an example. Occasionally, however, the different components of these large complexes may adopt different symmetries, which then associate, forming a symmetry mismatch at the site of interaction.

The  $\alpha$ -keto acid dehydrogenase complexes, such as the pyruvate dehydrogenase complex linking glycolysis to the citric acid cycle, are good examples (80). They are composed of three different subunits, denoted E1, E2, and E3, which perform sequential steps in the oxidative decarboxylation reaction. The pyruvate dehydrogenase complex from *E. coli* is composed of a core of 24 E2 subunits arranged in octahedral symmetry, forming a cube-shaped structure. A hole is formed at each of the six faces of the cube, each with four-fold molecular symmetry. A dimer of E3 is thought to bind within each of these holes, forming a mismatch of two-fold symmetry within a four-fold symmetric environment.

Rotavirus is another example of a knob of lower symmetry fitting into a hole of higher symmetry (Figure 13). The viral capsid is composed of two concentric shells with typical  $T = 13$  icosahedral quasisymmetry. Many small holes are formed between subunits, and the holes are in register between the two shells. Sixty dimeric hemagglutinin VP4 molecules bind within one class of these holes, extending from the capsid surface. The outer end of these molecules shows typical dimeric symmetry, but the inner end forms a globular structure with approximate hexagonal shape, fitting perfectly within the quasihexagonal holes (99).

Many examples of symmetry mismatch between monomeric proteins and oligomers with cyclic symmetry may be found. The pseudosymmetrical interactions of dimeric enzymes with monomeric substrates, described above, are examples. In those, the approximate dimeric nature of the substrate softens the mismatch. This is not so in ATP synthase, which has a three-fold symmetric complex composed of three  $\alpha$  and three  $\beta$  subunits, pierced by a monomeric  $\gamma$  subunit. As described below, the interaction distorts the ring of  $\alpha$  and  $\beta$  subunits, forcing each  $\alpha$ - $\beta$  pair into a different conformation and mediating the unusual rotary mechanism of action (1). Cholera toxin (100) and the heat-labile enterotoxin of *E. coli* (85) also each show a ring of subunits surrounding a monomeric subunit.

Symmetry mismatch has been evoked to explain several functional features of protein complexes. Hendrix proposed that a symmetry mismatch in bacteriophage could be the mechanism used for DNA injection (36). The tails of these phages show six-fold rotational symmetry, but they attach to the icosahedral heads at one unique vertex, through an axis of five-fold rotational symmetry. He proposed that this mismatch of symmetry might allow an ATP-driven motor to be formed at the interface, which, by turning, would forcibly eject DNA from the head. He further proposed that similar mismatches might be important in the flagellar motor. This is certainly the case in ATP synthase, in which the monomeric rotor is turned inside a three-fold symmetric ring (as described below), and the possibility of rotation in the mechanism of intracellular proteolysis has been proposed based on the seven-fold/six-fold symmetry mismatch of the components of the *E. coli* Clp chaperone-assisted protease (6).





A “Vernier” mechanism that relies on symmetry mismatch has been proposed as the structural mechanism determining the shape of bacteriophage T4 heads (5,77). The head of T4 is icosahedral, but it is extended in one direction by the addition of hexamers in an equatorial band around one five-fold axis. The amount of elongation is controlled by the interaction of an internal scaffolding with the shell during assembly (46). This scaffolding is later disassembled to form the mature phage. The Vernier method relies on the interaction of two concentric helical structures with different helical repeats. If formation of the caps that close each end of the oblate head are tied to a given alignment of the scaffolding core to the shell, then the length of the head could be determined by the repeat distance at which the two structures periodically come into phase. This mechanism is successful in explaining the aberrant head structures formed by mutant phage.

## ASYMMETRY

Protein monomers are strongly asymmetric. As summarized by Chothia (13), the asymmetry of L-amino acids gives rise to a preferred handedness for  $\alpha$ -helices and  $\beta$ -sheets. The packing of these units of secondary structure, which show a small number of preferred modes, then gives rise to asymmetric folded structures, including twisted  $\beta$ -sheets, curving  $\beta$ -ribbons, and tilted  $\alpha$ -helical bundles. Cases with high internal symmetry, such as cylindrical  $\alpha$ - $\beta$  barrels, are relatively rare. Often, they are the result of gene duplications and might be thought of as a form of linked quaternary structure.

The asymmetry imposed by the limitation to L-amino acids does not appear to extend to the level of quaternary structure, as noted by Chothia (13). Instead, symmetry is the rule in protein association: Overall, oligomeric proteins adopt closed point group symmetry, and polymers adopt helical symmetry. Asymmetry is observed, however, at several levels. Local asymmetry, in which individual amino acids show different conformations when comparing different subunits in an assembly, is ubiquitous. Asymmetry is also a key element of allosteric interactions, in which individual subunits can adopt one of several alternate conformations. However, true global asymmetry, in which subunits with identical primary sequences adopt positions within a complex that are not related by symmetry, is rarely observed.

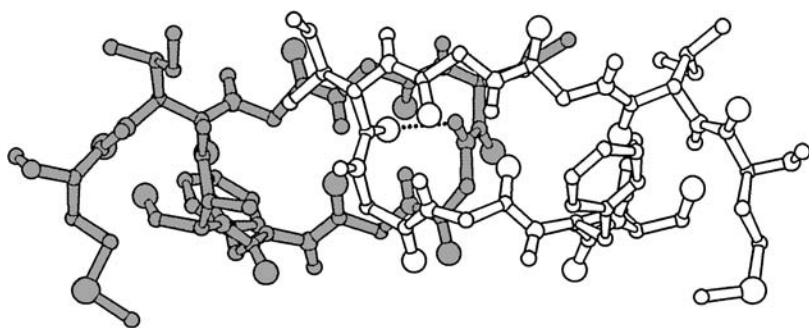
←  
**Figure 13** Symmetry mismatch in rotavirus. Rotavirus hemagglutinin VP4 is a dimer that binds to the viral capsid and extends into the surrounding environment. Many hemagglutinin spikes are shown in the reconstruction at top, and a detail of one is shown in the *inset*. As shown at *bottom*, the hemagglutinin shows typical dimeric symmetry at the outer end, with two separate lobes. At the inner end, however, the two chains form a hexagon-shaped globule, designed to fill the hexagonal holes left between subunits in the capsid. Cryoelectron microscope reconstruction data are courtesy of M Yeager at the Scripps Research Institute.

## Local Asymmetry

Nearly every crystal structure of an homooligomeric protein will show local differences in sidechain conformation and occasional small differences in backbone conformation. These may be observed as actual differences in atomic coordinates, if more than one subunit is found in the asymmetric unit, or as disordered residues, if the subunits are related by crystallographic symmetry. Often, these can be attributed to lack of constraints on the sidechain. Sidechains or even entire loops that are exposed to solvent will show largely different conformations in different subunits. Alternatively, crystal packing may order sidechains and flexible loops in different conformations. These differences can provide valuable observations of the mobility of proteins and the deformability of surface residues and loops.

Location near a symmetry axis can force breaking of local symmetry in a way that is necessary for structural integrity. Often subunits will adopt different conformations when near symmetry elements to optimize interactions across the axis. This was noted at the two-fold interface of insulin: "The two-fold-axis is not exactly obeyed, probably as a consequence of the very congested packing of residues which is observed here" (2). The flaps of HIV-1 protease provide another example, as shown in Figure 14. The peptide between ILE-50 and GLY-51 interacts with its symmetry mate across the two-fold axis; in one subunit, the amine hydrogen faces across the axis; in the other, the peptide is rotated so that the carbonyl oxygen faces across the axis, allowing formation of a hydrogen bond.

Local asymmetry seems to play a functional role in half-of-sites reactivity, or negative cooperativity. In proteins that show half-of-sites reactivity, binding of ligand or effector to one subunit disfavors binding to the other subunit. Examples include D-glyceraldehyde-3-phosphate dehydrogenase, alcohol dehydrogenase,



**Figure 14** Local asymmetry in HIV-1 protease. Asymmetry of sidechains is often observed close to symmetry elements. The flaps of HIV-1 protease show asymmetry at a hydrogen bond (*dotted line*), that passes through the two-fold symmetry axis of dimeric protein, which runs top to bottom in the plane of the page, relating the flap in white to the shaded flap.

arginine kinase, tyrosyl-tRNA synthetase, and many others (83). The crystal structures of these enzymes show globally symmetric structures, with local asymmetry of loops, particularly around the active sites. The observation made by Moras et al seems to apply to the entire class: "The tetrameric molecule of glyceraldehyde-3-phosphate dehydrogenase has long been shown to have asymmetric properties which almost certainly reflect themselves in small but definite differences of conformation in the four polypeptide chains" (73).

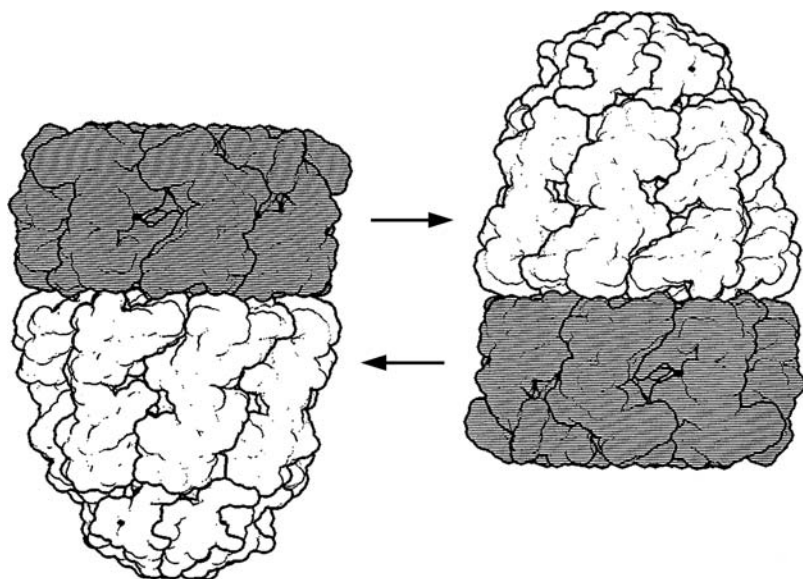
Based on biochemical results that show half-of-sites reactivity, Degani & Degani propose that arginine kinase adopts a globally asymmetric structure (21). Crystal structures of this enzyme, however, show a symmetrical dimer. Crystal structure analyses have been reported for other proteins that show half-of-sites reactivity, including alcohol dehydrogenase, tyrosyl-tRNA synthetase, and D-glyceraldehyde-3-phosphate dehydrogenase. In all, the subunits are related by two-fold symmetry. The negative cooperativity appears to be effected through subtle motion sidechains, causing the two subunits to adopt different local structures at the active site when complexed, but still retaining an overall symmetrical structure when uncomplexed.

## Reciprocating Mechanisms

Several examples of "reciprocating" mechanisms have been observed. In these, each subunit may adopt one or more states, but the state of one subunit is dependent on the states of the neighbors. For instance, envision a dimer in which each subunit may adopt two conformations, A and B. The dimer interaction will be such that if one subunit is in conformation A, the other must be in conformation B, and vice versa. Thus, at any given time, they do not adopt identical states, just like pistons on a crankshaft are all in different positions. However, if we take a time-averaged structure, the three A-B pairs are identical. Also, in analogy with the pistons and crankshaft, the reciprocating motion is processive; the conformation of one subunit will be optimized for catalysis, and the conformation of the other subunit will be optimized for binding to the next substrate.

ATP synthase is the best known example of this type of reciprocating engine. The reciprocating mechanism was proposed by Boyer (9), and the mechanism has been revealed at the atomic level (1). ATP synthase is composed of a  $\alpha_3\beta_3$  ring of subunits, encircling a multisubunit membrane-bound axle. The reciprocating mechanism cycles between three conformations: an O site with very low affinity for ligands and no catalytic activity, an L site that binds ligands loosely and is inactive, and a T site that binds ligands tightly and performs the catalytic step. In the complex, one of the three  $\alpha\beta$  heterodimers adopts each conformation at any given time, and physical turning of the axle converts one to the next.

The chaperonin GroEL acts, to a first approximation, through a two-state reciprocating mechanism (Figure 15). The asymmetric complex is composed of a 14-subunit GroEL, arranged as two rings of seven subunits bound back-to-back

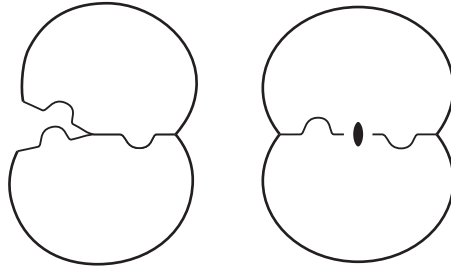


**Figure 15** Reciprocating mechanism of GroEL:GroES. The asymmetric complex is formed of a 14-subunit GroEL and a conical GroES cap formed of seven subunits. To a first approximation, the complex cycles between two conformations. *Upper conformation*, the right half of GroEL is open, and the left half is capped by GroES, which induces a significant conformational change in the GroEL subunits that it contacts. *Lower conformation*, the roles have reversed. The left side has lost GroES and is open, allowing the folded protein to exit and new proteins to enter, and the right side is capped by GroES. The complex is always asymmetric for the two sides, but, looking at a time average, both sides cycle through identical conformations.

with approximate  $D_7$  symmetry, and GroES, which forms a  $C_7$ -symmetric cap on one ring of the GroEL complex (97). GroEL undergoes large deformations upon binding of GroES, cycling between two conformations: a *cis* conformation, to which GroES is bound, which provides a closed space within which proteins fold, and the *trans* conformation, which is open to the surrounding solvent, releases and takes up polypeptides. One cycle allows about 15 seconds for the polypeptide to fold, and requires 7 ATP molecules. Communication between the two halves of the complex is mediated through a smooth tilt of subunits relative to the seven-fold axis, breaking the  $D_7$  symmetry of GroEL complex, but preserving the protein-protein interactions that bind one ring to the other.

## Global Asymmetry

Homooligomers with asymmetrically arranged subunits are remarkably rare. One can easily imagine an oligomer with nonintegral rotation symmetry or screw symmetry that forms a complex of limited size, but without identical environments for



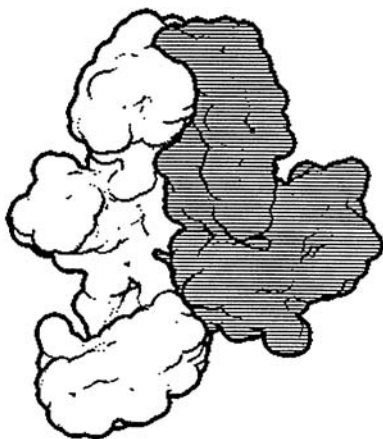
**Figure 16** Evolutionary argument against asymmetry. *Right*, in a symmetrical complex as shown, a portion of the protein surface is optimized through evolution to form an interface. Both subunits have a “knob” and a complementary “hole” that match across a two-fold-symmetry axis, shown with the small football symbol. *Left*, in the asymmetric complex, the surface forming the interface must be simultaneously optimized for interface contact and for interaction with solvent: the upper subunit in the complex has its “knob” forming an interface contact and its “hole” exposed to solvent, and in the lower subunit, the roles are reversed. Similar diagrams can be envisioned for other cyclic and screw-related asymmetric complexes.

each subunit. One can envision a strong evolutionary force against such asymmetric dimers. Compare the two dimers in Figure 16. In the two-fold related dimer, the interaction surface is identical in the two subunits: evolution would proceed by optimizing the face to be complementary with its mate. In the asymmetric dimer, on the other hand, the interface surface must be simultaneously optimized to perform two roles. In one subunit, half of the “interface” will be in contact with the neighbor, and half will be exposed to solvent; in the other subunit, these roles are reversed. Thus, a large area must be evolutionarily optimized for interaction with the neighbor *and* with solvent, and only half of this surface actually makes contact in the dimer.

Such a complex has been proposed for hexokinase, based on screw-related subunits observed in crystal structures. However, the screw relationship is quite different in two different crystal habits, arguing against this as the unique dimerization mode. A similar screw relationship has been proposed for arginine kinase, based on biochemical data showing half-of-sites reactivity.

The BTV core (Figure 9a) is the first X-ray crystallographic example of this type of global asymmetry (31). In this structure, each position in a  $T = 1$  icosahedron is filled with two subunits, bound back-to-back. The subunit is bean shaped and composed of three domains, and shows different conformations in the two different asymmetric positions. It has yet to be determined whether the back-to-back dimer is a stable intermediate in the process of assembly.

HIV-1 reverse transcriptase shows a severe example of asymmetry, as shown in Figure 17. The active complex is a dimer of a 66,000-Dalton subunit and a 51,000-Dalton subunit, the latter of which is a proteolyzed version of the first. The complex has a single polymerase active site, a single RNase H proofreading site,



**Figure 17** Asymmetry in human immunodeficiency virus-1 reverse transcriptase. The reverse transcriptase of HIV-1 is composed of two subunits, a 66,000-Dalton subunit (*white*) and a 51,000-Dalton subunit (*shaded*), which is a proteolyzed version of the larger. The two subunits adopt entirely different conformations in the complex.

and a single tRNA-binding site. Each subunit is composed of four similar subdomains, which pack together in entirely different orientation in the two subunits (52). Owing to the strictures of viral economy, here one sequence plays two very different structural roles.

## EVOLUTION OF OLIGOMERIC PROTEINS

The remarkable solutions that cells have found to functional problems show that there are few limitations to what can be developed, given enough time. Just the same, there have been attempts to justify the observed forms of symmetry with underlying evolutionary arguments. Monod et al (72) presented one of the first ideas—that dimeric (isologous) interfaces are easier to create by mutation of existing monomers than are interfaces in complexes of higher symmetry (heterologous interfaces). They note that, on the surface of any monomer, pairs of complementary residues are prepositioned for formation of dimeric contacts owing to symmetry. For instance, a monomer with an arginine at one point on the surface and a glutamate some small distance away can immediately form two salt bridges with a second monomer, arginine on one to glutamate on the other, and glutamate on one to arginine on the other. The distance between the arginine and the glutamate on one subunit is necessarily identical with the distance between these two residues on the second subunit, providing the seed for a dimeric interface. This is not the case with higher symmetries. In complexes with higher cyclic symmetry, the distance and orientation of a putative “seeding” pair are not rigidly defined around the ring, so there exists the possibility of many nonproductive pairings.



In polymeric complexes, the existence of a prepositioned pair is not possible, because the two interacting surfaces are different, so evolution of a high-symmetry complex must begin with pairing of a single residue on each subunit. To give the same strength of binding for this nascent oligomer, the high-symmetry interfaces must have a fortuitous alignment of four residues on each subunit, two on one side of the interface and two on the other.

Evolution of a D2 tetramer may be favored over that of a C4 tetramer by this same argument. The D2 tetramer may be formed in two evolutionary steps, using an existing complementary pair to seed a dimer interface between two monomers, and then, after this dimer has been optimized by evolution, using a second complementary pair on the dimer to seed a tetramer interface. The creation of a C4 tetramer requires a concerted set of evolutionary steps, bringing the entire complex together at once. Hanson proposed that the requirement of evolution of an open dimer as a first step towards creation of a ring would be an improbable event and suggested that cyclic structures evolved sequentially by altering the relative orientation of the binding faces: "In this way, a D2 isologous ring could evolve successively into a D3 and then a D4 ring, or ring contraction could take place" (33). Looking to the many structures of oligomeric proteins, we see that this type of remodeling is probably not possible. It is difficult to postulate an evolutionary pathway that would change the angle between interfaces on a given subunit from  $180^\circ$  to  $120^\circ$ , either by repositioning an interface along the surface of a protein or by changing the angle by contracting the underlying protein fold.

Additional mutations favoring contact within interfaces are magnified in effect in oligomeric complexes. Monod et al write: "Because of the inherent cooperativity of their structure, symmetrical oligomers should constitute particularly sensitive targets for molecular evolution, allowing much stronger selective pressures to operate in the random pursuit of functionally adequate structures" (72). A single mutation will form two new hydrogen bonds in a dimer, and single mutations will simultaneously relieve two close steric contacts. The effect is further magnified in higher-order oligomers.

These evolutionary forces are easily overcome by a specific functional need. Two observations give insight into the strength of these intrinsic structural forces on evolution relative to the strength of functional need. Arguing for a strong evolutionary preference for dimer interfaces, dimers and tetramers with D2 "dimer-of-dimers" symmetry are the most prevalent symmetries for soluble enzymes, even for those not showing significant allosteric behavior. Arguing against a strong force, the processivity factors of DNA polymerase (Figure 5) use two different symmetries to fulfill the same need. In the *E. coli*  $\beta$ -subunit of polymerase III holoenzyme, two subunits assemble to form a ring with two interface patches. The eukaryotic processivity factor of DNA polymerase  $\delta$ , on the other hand, adopts an identical ring-shape, but it is composed of three subunits. An identical functional need is fulfilled by a dimer and by a trimer. Obviously, the evolutionary forces seeding new interfaces are not strong enough to favor the dimer exclusively over the trimer or vice-versa.



Xu et al discuss three possible modes for evolution of dimeric proteins (96), based on a survey of modern dimeric proteins. The first is the traditional mechanism, in which a stable monomeric species develops a dimerization site, perhaps through the use of complementary pairs of residues as described above. We might expect this mechanism to account for most of the oligomeric species observed today, such as superoxide dismutase in Figure 2a. It is not hard to image superoxide dismutase as a primitive species, stable as a monomer. The second pathway is through domain swapping, as first proposed by Bennett et al (4). The interleukin shown in Figure 2b may have evolved in this manner, with a swap across the narrow linker at the center of the dumbbell-shaped molecule. The final mode for evolution of dimeric proteins is the most difficult of all—a one-step process, in which a dimeric species emerges fully formed. Xu et al propose several highly interlocked dimers, such as the gene V protein and the Trp aporepressor, as examples of this mechanism. Because of their extensive interdigitation, they are probably not stable as monomers, and the geometry of the backbone is not consistent with domain swapping.

Unexpected symmetries may reveal a “frozen accident” of evolution (95)—a symmetry formed in the past and then unable to evolve further without loss of function and a compromised organism. Hemoglobin provides a familiar example: One might expect that an  $\alpha_4$  tetramer would be more genetically compact than the observed  $\alpha_2\beta_2$  heterotetramer. Presumably, early in the evolution of vertebrates, a gene duplication formed two copies of subunits for a tetrameric hemoglobin, and of the several competing forms— $\alpha_4$ ,  $\beta_4$ , or  $\alpha_2\beta_2$ —the heterotetramer was selected.

## THE AESTHETICS OF SYMMETRY

Looking at much of decorative and fine art, we find that symmetry has a strong aesthetic appeal for many cultures. Symmetrical patterns, arrangements, and objects adorn our rooms and define our architecture, and breaks from symmetry are incorporated deliberately and with artistic intent, because they will surprise and shock our expectations. When choosing vegetables or flowers, when looking at plants and animals, or when admiring the latest movie star, we favor the most symmetric, seeking out the most “perfect” individuals. This predilection for symmetry may have its roots far back in our evolution, where symmetry was often a sign of a healthy mate or fresh food.

Nature also selects for symmetry, for reasons of economy and control. These symmetries occur at the molecular level, as described in this review, and extend to the cellular and organismal level, for much the same reasons. Because symmetrical complexes are functionally more successful and economical than asymmetric objects in many cases, our world is filled with five-pointed starfish, Y-shaped antibodies, spiral flowers, circular tree rings, geodesic diatoms, lenticular red blood cells, and perfectly icosahedral viruses. These symmetries were perfected through

functional pressures on evolution; their appeal to our senses is merely a happy side effect.

Of course, there remain symmetries that likely exist simply through serendipitous combinations of physical forces—not selected for functional need, but appearing fully formed once the universe took its current shape. The crystalline beauty of minerals is one example, resulting from the minimization of inherent directional interactions of molecules. Similarly, the  $\alpha$ -helix is another, with a structure defined not through evolutionary forces but through the intrinsic chemical geometry of the polypeptide chain. The rainbow is perhaps the most glorious example—a gift of Nature existing simply through combination of the surface tension and refractive index of water.

## ACKNOWLEDGMENTS

This work was funded by grant DEFG03 96ER2272 from the U.S. Department of Energy. This is publication 12361-MB from the Scripps Research Institute.

**Visit the Annual Reviews home page at [www.AnnualReviews.org](http://www.AnnualReviews.org)**

## LITERATURE CITED

1. Abrahams JP, Leslie AGW, Lutter R, Walker JE. 1994. Structure at 2.8 angstroms resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370:621–28
2. Adams MJ, Blundell TL, Dodson EJ, Dodson GG, Vijayan M, et al. 1969. Structure of rhombohedral 2 zinc insulin crystals. *Nature* 224:491–95
3. Barford D, Johnson LN. 1989. The allosteric transition of glycogen phosphorylase. *Nature* 340:609–16
4. Bennett MJ, Schlunegger MP, Eisenberg D. 1995. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* 4:2455–68
5. Berger B, Shor PW. 1998. On the structure of the scaffolding core of bacteriophage T4 and its role in head length determination. *J. Struct. Biol.* 121:285–94
6. Beuron F, Maurizi MR, Belnap DM, Kocsis E, Booy FP, et al. 1998. At sixes and sevens: characterization of the symmetry mismatch of the ClpAP chaperone-assisted protease. *J. Struct. Biol.* 123:248–59
7. Blundell T, Dodson G, Hodgkin D, Mercola D. 1972. Insulin: the structure in the crystal and its reflection in chemistry and biology. *Adv. Protein Chem.* 26:279–402
8. Blundell TL, Srinivasan N. 1996. Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. *Proc. Natl. Acad. Sci. USA* 93:14243–48
9. Boyer PD. 1993. The binding site change mechanism for ATP synthase—some probabilities and possibilities. *Biochim. Biophys. Acta* 1140:215–50
10. Calladine CR. 1978. Change of waveform in bacterial flagella: the role of mechanics at the molecular level. *J. Mol. Biol.* 118:457–79
11. Caspar DLD. 1966. Design and assembly of organized biological structure. *Molecular Architecture in Cell Physiology, Symp. Soc. Gen. Physiol.*, pp. 191–207. New York: Prentice Hall
12. Caspar DLD, Klug A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* 27:1–24
13. Chothia C. 1991. Asymmetry in protein

- structures. *CIBA Found. Symp.* 162:36–57
14. Chothia C, Janin J. 1975. Principles of protein-protein recognition. *Nature* 256: 705–8
  15. Clegg JS. 1984. Properties and metabolism of the aqueous cytoplasm and its boundaries. *Am. J. Physiol.* 246:R133–51
  16. Cornish-Bowden AJ, Koshland DE Jr. 1971. The quaternary structure of proteins composed of identical subunits. *J. Biol. Chem.* 246:3092–102
  17. Crane HR. 1950. Principles and problems of biological growth. *Sci. Monthly* June:376–89
  18. Crick FHC, Watson JD. 1956. Structure of small viruses. *Nature* 177:473–75
  19. Crick FHC, Watson JD. 1957. Virus structure: general principles. *CIBA Found. Symp. "The Nature of Viruses"*, pp. 5–13. Boston: Little, Brown
  20. Crothers DM, Metzger H. 1972. The influence of polyvalency on the binding properties of antibodies. *Immunochemistry* 9:341–57
  21. Degani Y, Degani C. 1980. Enzymes with asymmetrically arranged subunits. *Trends Biochem. Sci.* 5:337–41
  22. DeVos AM, Ultsch M, Kossiakoff AA. 1992. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 255:306–12
  23. Evans PR. 1991. Structural aspects of allostery. *Curr. Opin. Struct. Biol.* 1:773–79
  24. Fulton AB. 1982. How crowded is the cytoplasm? *Cell* 30:345–47
  25. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI. 1999. Assembly and analysis of conical models for the HIV-1 core. *Science* 283:80–83
  26. Garcia-Bellido A. 1996. Symmetries throughout organic evolution. *Proc. Natl. Acad. Sci. USA* 93:14229–32
  27. Goodsell DS. 1991. Inside a living cell. *Trends Biochem. Sci.* 16:203–6
  28. Goodsell DS, Olson AJ. 1993. Soluble proteins: size, shape and function. *Trends Biochem. Sci.* 18:65–68
  29. Gouax JE, Lipscomb WN. 1990. Crystal structures of phosphoacetamide ligated T and phosphoacetamide and malonate ligated R states of aspartate carbamoyltransferase at 2.8-Å resolution and neutral pH. *Biochemistry* 29:389–402
  30. Greenbury CL, Moore DH, Nunn LAC. 1965. The reaction with red cells of 7S rabbit antibody, its sub-units and their recombinants. *Immunology* 8:420–31
  31. Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, et al. 1998. The atomic structure of the bluetongue virus core. *Nature* 395:470–78
  32. Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, et al. 1997. Structure of the 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386:463–71
  33. Hanson KR. 1966. Symmetry of protein oligomers formed by isologous association. *J. Mol. Biol.* 22:405–9
  34. Hargittai I, Hargittai M. 1995. *Symmetry through the Eyes of a Chemist*. New York: Plenum
  35. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. 1978. Tomato bushy stunt virus at 2.9 Angstroms resolution. *Nature* 276:368–73
  36. Hendrix RW. 1978. Symmetry mismatch and DNA packaging in large bacteriophages. *Proc. Natl. Acad. Sci. USA* 75:4779–83
  37. Herendeen DR, Kelly TJ. 1996. DNA polymerase III: running rings around the fork. *Cell* 84:5–8
  38. Heuser J, Kirchhausen T. 1985. Deep-etch views of clathrin assemblies. *J. Ultrastruct. Res.* 92:1–27
  39. Hewitt JA. 1977. On the influence of polyvalent ligands on membrane curvature. *J. Theor. Biol.* 64:455–72
  40. Hoare MR, Pal P. 1975. Physical cluster mechanics: statistical thermodynamics and nucleation theory for monatomic systems. *Adv. Phys.* 24:645–78
  41. Iwata S, Kamata K, Yoshida S, Minowa T, Ohta T. 1994. T and R states in the crystals of bacterial L-lactate dehydrogenase reveal

- the mechanism for allosteric control. *Nature Struct. Biol.* 1:176–85
42. Johnson JE, Speir JA. 1997. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol.* 269:665–75
43. Jones S, Thornton JM. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* 93:13–20
44. Juo ZS, Chiu TK, Leiberman PM, Baikalov I, Berk AJ, Dickerson RE. 1996. How proteins recognize the TATA box. *J. Mol. Biol.* 261:239–54
45. Kamiya R, Asakura S, Wakabayashi K, Namba K. 1979. Transition of bacterial flagella from helical to straight forms with different subunit arrangements. *J. Mol. Biol.* 131:725–42
46. Kellenberger E. 1990. Form determination of the heads of bacteriophages. *Eur. J. Biochem.* 190:233–48
47. Kelman Z, Finkelstein J, O'Donnell M. 1995. Why have six-fold symmetry? *Curr. Biol.* 5:1239–42
48. Klotz IM. 1967. Protein subunits: a table. *Science* 155:697–98
49. Klotz IM, Darnall DW, Langerman NR, eds. 1975. *Quaternary Structure of Proteins*, pp. 293–411. New York: Academic
50. Klotz IM, Langerman NR, Darnall DW. 1970. Quaternary structure of proteins. *Annu. Rev. Biochem.* 39:25–62
51. Klug A. 1968. Point groups and the design of aggregates. *Nobel Symp. "Symmetry and Function of Biological Systems at the Macromolecular Level" 11th, Stockholm*, pp. 425–36. Wiley & Sons, New York
52. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256:1783–89
53. Kong X-P, Onrust R, O'Donnell M, Kuriyan J. 1992. Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* 69:425–37
54. Koshland DE. 1976. The evolution of function in enzymes. *Fed. Proc.* 35:2104–11
55. Koshland DE, Nemethy G, Filmer D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5:365–85
56. Krishna TSR, Kong X-P, Gary S, Burgers PM, Kuriyan J. 1994. Crystal structure of the eukaryotic DNA processivity factor PCNA. *Cell* 79:1233–43
57. Kurland CG. 1992. Translational accuracy and the fitness of bacteria. *Annu. Rev. Genet.* 26:29–50
58. Labeit S, Kolmerer B. 1995. Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* 270:293–96
59. Ladenstein R, Schneider M, Huber R, Bartunik H-D, Wilson K, et al. 1988. Heavy riboflavin synthase from *Bacillus subtilis*: crystal structure analysis of the icosahedral beta-60 capsid at 3.3 Å resolution. *J. Mol. Biol.* 203:1045–70
60. Larsen TA, Olson AJ, Goodsell DS. 1998. Morphology of protein-protein interfaces. *Structure* 6:421–27
61. Lewis MS, Youle RJ. 1986. Ricin subunit association: thermodynamics and the role of the disulfide bond in toxicity. *J. Biol. Chem.* 261:11571–77
62. Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. 1991. Structure of simian virus 40 at 3.8-Ångstrom resolution. *Nature* 354:278–84
63. Lipscomb WN. 1991. Structure and function of allosteric enzymes. *CHEMTRACTS-Biochem. Mol. Biol.* [Data Trace Chem. Publ.] 2:1–15
64. Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R. 1995. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268:533–39
65. Lumry R, Rajender S. 1970. Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers* 9:1125–227

66. Matsudaira P. 1991. Modular organization of actin crosslinking proteins. *Trends Biochem. Sci.* 16:87–92
67. Mattevi A, Rizzi M, Bolognesi M. 1996. New structures of allosteric proteins revealing remarkable conformational changes. *Curr. Opin. Struct. Biol.* 6:824–29
68. Mattevi A, Valentini G, Rizzi M, Speranza ML, Bolognesi M, Coda A. 1995. Crystal structure of *Escherichia coli* pyruvate kinase type I: molecular basis of the allosteric transition. *Structure* 3:729–41
69. Mitchison T, Kirschner M. 1984. Dynamic instability of microtubule growth. *Nature* 312:237–42
70. Mitchison TJ. 1988. Microtubule dynamics and kinetochore function in mitosis. *Annu. Rev. Cell Biol.* 4:527–49
71. Monod J. 1968. On symmetry and function in biological systems. *Nobel Symp. Symmetry Funct. Biol. Syst. Macromol. Lev., 11th, Stockholm*, pp. 15–27. New York: Wiley
72. Monod J, Wyman J, Changeux J-P. 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12:88–118
73. Moras D, Olsen KW, Sabesan MN, Buehner M, Ford GC, Rossmann MG. 1975. Studies of asymmetry in the three-dimensional structure of lobster D-glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.* 250:9137–62
74. Olson AJ, Bricogne G, Harrison SC. 1983. Structure of tomato bushy stunt virus IV. The virus particle at 2.9 Å resolution. *J. Mol. Biol.* 171:61–93
75. Parker J. 1989. Errors and alternatives in reading the universal genetic code. *Microbiol. Rev.* 53:273–98
76. Pauling L. 1953. Protein interactions: aggregation of globular proteins. *Discov. Faraday Soc.* 13:170–76
77. Paulson JR, Laemmli UK. 1977. Morphogenetic core of the bacteriophage T4 head. Structure of the core in polyheads. *J. Mol. Biol.* 111:459–85
78. Pereira PJB, Bergner A, Macedo-Ribeiro S, Huber R, Matxchiner G, et al. 1998. Human beta-trypsin is a ring-like tetramer with active sites facing a central pore. *Nature* 392:306–11
79. Perutz MF. 1989. Mechanisms of cooperativity and allosteric regulation in proteins. *Q. Rev. Biophys.* 22:139–236
80. Reed LJ, Hackert ML. 1990. Structure-function relationships in dihydrolipoamide acyltransferases. *J. Biol. Chem.* 265:8971–74
81. Ricard J, Noat G. 1985. Kinetic cooperativity of monomeric mnemonic enzymes. *Eur. J. Biochem.* 152:557–64
82. Rossmann MG. 1984. Constraints on the assembly of spherical virus particles. *Virology* 134:1–11
83. Seydoux F, Malhotra OP, Bernhard SA. 1974. Half-site reactivity. *CRC Crit. Rev. Biochem.* 2:227–57
84. Sharp K, Fine R, Honig B. 1987. Computer simulations of the diffusion of a substrate to an active site of an enzyme. *Science* 236:1460–64
85. Sixma TK, Pronk SE, Kalk KH, Wartna ES, Zanten BAM, et al. 1991. Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature* 351:371–77
86. Srere PA. 1981. Protein crystals as a model for mitochondrial proteins. *Trends Biochem. Sci.* 6:4–7
87. Srere PA. 1984. Why are enzymes so big? *Trends Biochem. Sci.* 9:387–90
88. Stossel TP. 1989. From signal to pseudopod: How cells control cytoplasmic actin assembly. *J. Biol. Chem.* 264:18261–64
89. Strater N, Hakansson K, Schnappauf G, Braus G, Lipscomb WN. 1996. Crystal structure of the T state of allosteric yeast chorismate mutase and comparison with the R state. *Proc. Natl. Acad. Sci. USA* 93:3330–34
90. Svedberg T. 1929. Mass and size of protein molecules. *Nature* 123:871
91. Theil EC. 1987. Ferritin: structure, gene regulation, and cellular function in

- animals, plants and microorganisms. *Annu. Rev. Biochem.* 56:289–315
92. vanHeel M, Orlova EV, Dube P, Tavares P. 1996. Intrinsic versus imposed curvature in cyclical oligomers: the portal protein of bacteriophage SPPI. *EMBO J.* 15:4785–88
93. Wilson DR. 1993. Bacterial flagellar filaments and their component flagellins. *Can. J. Microbiol.* 39:451–72
94. Wlodawer A, Erickson JW. 1993. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* 62:543–85
95. Wolynes PG. 1996. Symmetry and the energy landscapes of biomolecules. *Proc. Natl. Acad. Sci. USA* 93:14249–55
96. Xu D, Tsai C-J, Nussinov R. 1998. Mechanism and evolution of protein dimerization. *Protein Sci.* 7:533–44
97. Xu Z, Horwich AL, Sigler PB. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* 388:741–50
98. Yang M, Chen X-p, Militello K, Hoffman R, Fernandez B, et al. 1997. Alanine-scanning mutagenesis of *Bacillus subtilis* trp RNA-binding attenuation protein (TRAP) reveals residues involved in tryptophan binding and RNA binding. *J. Mol. Biol.* 270:696–710
99. Yeager M, Berriman JA, Baker TS, Belamy AR. 1994. Three-dimensional structure of the rotavirus haemagglutinin VP4 by cryo-electron microscopy and difference map analysis. *EMBO J.* 13:1011–18
100. Zhang R-G, Scott DL, Westbrook ML, Nance S, Spangler BD, et al. 1995. The three-dimensional crystal structure of cholera toxin. *J. Mol. Biol.* 251:563–73
101. Protein Data Bank accession codes for structures used in the figures. The PDB may be reached at <http://www.rcsb.org/pdb>. Figure 1: pepsin, 5pep; alcohol dehydrogenase, 2ohx; porin, 2por; neuraminidase, 1ivb; phosphofructokinase, 1pfk; aspartate carbamoyltransferase, 1at1; glycolate oxidase, 1gox; glutamine synthetase, 2gls; protocatechuate-3,4-dioxygenase, 3pcg; ferritin, 1hrs; satellite tobacco necrosis virus, 2stv. Figure 2 superoxide dismutase, 1xso; interleukin 10, 1ilk. Figure 4: fructose-1,6-bisphosphatase, 4fbp (T) and 5fbp (R); aspartate carbamoyltransferase, 4at1 (T) and 1at1 (R). Figure 5:  $\beta$  subunit of RNA pol III, 2pol; PCNA, 1axc. Figure 6: GroEL, 1der; proteasome, 1ryp. Figure 7: satellite tobacco necrosis virus, 2stv; tomato bushy stunt virus, 2tbv. Figure 8: bluetongue virus, 2btv; simian virus 40, 1sva. Figure 11: hemoglobin, 4hhb. Figure 13: HIV-1 protease, 7hvp. Figure 14: GroEL:GroES, 1aon. Figure 16: reverse transcriptase, 3hvt.



## CONTENTS

MEASURING THE FORCES THAT CONTROL PROTEIN INTERACTIONS, <i>Deborah Leckband</i>	1
STRUCTURE AND FUNCTION OF LIPID-DNA COMPLEXES FOR GENE DELIVERY, <i>S. Chesnoy, L. Huang</i>	27
SIGNALING AND SUBCELLULAR TARGETING BY MEMBRANE-BINDING DOMAINS, <i>James H. Hurley, Saurav Misra</i>	49
GCN5-RELATED N-ACETYLTRANSFERASES: A Structural Overview, <i>Fred Dyda, David C. Klein, Alison Burgess Hickman</i>	81
STRUCTURAL SYMMETRY AND PROTEIN FUNCTION, <i>David S. Goodsell, Arthur J. Olson</i>	105
ELECTROKINETICALLY CONTROLLED MICROFLUIDIC ANALYSIS SYSTEMS, <i>Luc Bousse, Claudia Cohen, Theo Nikiforov, Andrea Chow, Anne R. Kopf-Sill, Robert Dubrow, J. Wallace Parce</i>	155
DNA RECOGNITION BY Cys2His2 ZINC FINGER PROTEINS, <i>Scott A. Wolfe, Lena Nekludova, Carl O. Pabo</i>	183
PROTEIN FOLDING INTERMEDIATES AND PATHWAYS STUDIED BY HYDROGEN EXCHANGE, <i>S. Walter Englander</i>	213
QUANTITATIVE CHEMICAL ANALYSIS OF SINGLE CELLS, <i>D. M. Cannon Jr, N. Winograd, A. G. Ewing</i>	239
THE STRUCTURAL BIOLOGY OF MOLECULAR RECOGNITION BY VANCOMYCIN, <i>Patrick J. Loll, Paul H. Axelsen</i>	265
COMPARATIVE PROTEIN STRUCTURE MODELING OF GENES AND GENOMES, <i>Marc A. Martí-Renom, Ashley C. Stuart, Andrés Fiser, Roberto Sánchez, Francisco Melo, Andrej Sali</i>	291
FAST KINETICS AND MECHANISMS IN PROTEIN FOLDING, <i>William A. Eaton, Victor Muñoz, Stephen J. Hagen, Gouri S. Jas, Lisa J. Lapidus, Eric R. Henry, James Hofrichter</i>	327
ATOMIC FORCE MICROSCOPY IN THE STUDY OF MACROMOLECULAR CRYSTAL GROWTH, <i>A. McPherson, A. J. Malkin, Yu. G. Kuznetsov</i>	361
A DECADE OF CLC CHLORIDE CHANNELS: Structure, Mechanism, and Many Unsettled Questions, <i>Merritt Maduke, Christopher Miller, Joseph A. Mindell</i>	411
DESIGNED SEQUENCE-SPECIFIC MINOR GROOVE LIGANDS, <i>David E. Wemmer</i>	439
PULSED AND PARALLEL-POLARIZATION EPR CHARACTERIZATION OF THE PHOTOSYSTEM II OXYGEN-EVOLVING COMPLEX, <i>R. David Britt, Jeffrey M. Peloquin, Kristy A. Campbell</i>	463



ELECTROSTATIC MECHANISMS OF DNA DEFORMATION, <i>Loren Dean Williams, L. James Maher III</i>	497
STRESS-INDUCED STRUCTURAL TRANSITIONS IN DNA AND PROTEINS, <i>T. R. Strick, J.-F. Allemand, D. Bensimon, V. Croquette</i>	523
MOLECULAR MECHANISMS CONTROLLING ACTIN FILAMENT DYNAMICS IN NONMUSCLE CELLS, <i>Thomas D. Pollard, Laurent Blanchoin, R. Dyche Mullins</i>	545
UNNATURAL LIGANDS FOR ENGINEERED PROTEINS: New Tools for Chemical Genetics, <i>Anthony Bishop, Oleksandr Buzko, Stephanie Heyeck-Dumas, Ilyoung Jung, Brian Kraybill, Yi Liu, Kavita Shah, Scott Ulrich, Laurie Witucki, Feng Yang, Chao Zhang, Kevan M. Shokat</i>	577