Overview of Protein Structural and Functional Folds

INTRODUCTION TO PROTEIN STRUCTURE

Proteins fold into stable three-dimensional shapes, or conformations, that are determined by their amino acid sequence. The complete structure of a protein can be described at four different levels of complexity: primary, secondary, tertiary, and quaternary structure.

As a multitude of protein structures are rapidly being determined by X-ray crystallography and by nuclear magnetic resonance (NMR), it is becoming clear that the number of unique folds is far less than the total number of protein structures. Not only do functionally related proteins generally have similar tertiary structures (see below), but even proteins with very different functions are often found to share the same tertiary folds. As a consequence, structural conservation at the tertiary level is perhaps more profound than it is at the primary. The identification of the fold of a protein has therefore become an invaluable tool since it can potentially provide a direct extrapolation to function, and may allow one to map functionally important regions in the amino acid sequence.

Several groups have already attempted to classify protein structures into fold families and superfamiles without focusing on function (Orengo et al., 1993; Murzin et al., 1995). The scope of this unit is not to enumerate all the existing folds and tertiary structures determined to date, but rather to provide a comprehensive overview of some commonly observed protein fold families and commonly observed structural motifs which have functional significance. Likewise, the PDB-entry tables given in this unit provide some examples of various folds, but are not comprehensive lists. The unit is organized into sections based on both structural and functional relations.

Primary Structure

Primary structure is defined as the linear amino acid sequence of a protein's polypeptide chain. In fact, the term ***protein sequence*** is often used interchangeably with primary structure. In 1973, Chris Anfinsen demonstrated that the primary amino sequence of a protein uniquely determines the higher orders of structure for a protein and is thus of fundamental importance (Anfinsen, 1973). It is noteworthy, however, that changes in the local biological environment of a protein molecule can sometimes perturb its three-dimensional structure. For example, interactions with ligands, substrates, or other proteins can bring about controlled conformational changes producing potentially profound effects. Furthermore, a few proteins have been found to have intrinsically unstructured regions (Wright and Dyson, 1999; Tompa, 2002). Hence, although structural uniqueness associated with a protein sequence is a powerful principle, it can sometimes depend on the local environment and is not rigidly followed in every case.

Secondary Structure

Secondary structure is defined as the local spatial conformation of the polypeptide backbone excluding the side chains. Regular secondary structures (also referred to as ***secondary structure elements***) common to many proteins include α-helices, β-sheets, and turns (see below). They can vary widely in length, from as few as three to five residues in short helices and sheets, to over fifty residues in some coiled-coil helices (see Frequently Observed Secondary Structure Assemblies or Structural Motifs). Such structures are generally defined both by characteristic main chain ϕ and ψ dihedral angles (i.e., the torsion angles between backbone atoms C_{i-1}–N_i–$C_i\alpha$–C_i and N_i–$C_i\alpha$–C_i–N_{i+1}, respectively; Fig. 17.1.1A) and by regular main chain hydrogen bonding patterns. When discussing the structure of a protein, the term ***topology*** is often used to refer to the connectivity of secondary structure elements. For example, a portion of a protein containing a β-strand connected to an α-helix and then another β-strand is said to have a βαβ topology. The complete secondary structure of a polypeptide chain is often represented in two dimensions by a topology diagram (e.g., Fig. 17.1.1H) which shows both the connectivity and the relative orientation of neighboring secondary structure elements. Such diagrams are particularly useful in classifying β-sheets.

Contributed by Peter D. Sun, Christine E. Foster, and Jeffrey C. Boyington
Current Protocols in Protein Science (2004) 17.1.1-17.1.189

α-*helices*

An α-helix is formed when a protein backbone adopts a right-handed helical conformation with 3.6 residues per turn and a set of hydrogen bonds formed between the main chain carbonyl (CO) of the i^{th} residue and the main chain NH of the $(i+4)^{th}$ residue (Fig. 17.1.1B). Occasionally, hydrogen bonds are observed between residues *i* and *i+3*, resulting in a 3_{10} helix. Compared to the more common α-helices, 3_{10} helices are much shorter, usually comprising only one to three turns. The optimal backbone ϕ and ψ dihedral angles for a right-handed α-helix are −57° and −47°, while for a 3_{10} helix they and −49° and −26°, respectively.

β-*sheets*

β-sheets (also referred to as β-pleated sheets) make up another major secondary structural element in proteins. A β-sheet consists of at least two β-strands, each approaching an extended backbone conformation with dihedral angles confined to the region where the ϕ torsion angle is between −60° and −180°, and the ψ torsion angle is between 30° and 180° (i.e., $-180° < \phi < -60°$, $30° < \psi < 180°$). All main-chain CO-to-NH hydrogen bonds lie between adjacent strands (Fig 17.1.1C). A parallel β-sheet is formed when all the sheet-forming strands run parallel to each other and in the same direction from N to C termini, whereas an antiparallel β-sheet is formed when the strands are still parallel to each other, but run in opposite directions (Fig. 17.1.1C); a β-sheet formed from both parallel and antiparallel strands in referred to as a mixed β-sheet. A characteristic feature of β-sheets is the right handed twist which is visible when the sheet is viewed edge-on (Fig. 17.1.1D).

Turns and loops

Loops and turns connect helices and β-sheets in protein structures. Most turns and loops assume irregular secondary structures, but important exceptions are the type I and II reverse turns (also referred to as β-turns or β-bends), which are tight turns connecting adjacent, antiparallel β-strands. Shown in Figure 17.1.1E, type I and II turns each have a hydrogen bond between the CO of the first residue in a turn and the NH of the fourth residue (also the last residue) in the turn. The difference between the type I and II turn is a 180° flip of the peptide plane in the middle of the turn. The type I turn is more frequently observed than the type II. The terms loops and coils generally refer to secondary structure elements that display less regular hydrogen bonding patterns than those observed in α-helices, β-sheets, and reverse turns.

Tertiary Structure

Tertiary structure refers to the three-dimensional arrangement of all the atoms that constitute a protein molecule. It relates the precise spatial coordination of secondary structure elements and the location of all functional groups of a single polypeptide chain.

Domains, Folds, and Motifs

The tertiary structure of a protein can often be divided into domains—i.e., distinct compact folding units usually comprising 100 to 200 residues. Small proteins may contain a single domain, whereas larger proteins often contain multiple domains. A fold refers to a characteristic spatial assembly of secondary structure elements into a domain-like structure that is common to many different proteins. Particular folds are often, though not always, related to a certain function (e.g., nucleotide-binding folds). A structural motif is similar to a fold, but is generally smaller, tending to form the building blocks of folds. Some structural motifs (e.g., β-barrels) are observed in a vast array of unrelated proteins with many variations, while others, especially those related to a unique biochemical function (e.g., zinc fingers; see Zinc-Containing DNA-Binding Motifs), are highly conserved, generally occurring in protein domains of similar function. Highly conserved structural motifs often display a characteristic signature at their amino acid sequence level, the so-called ***sequence motif***. Very often the term ***motif*** is used alone to refer to either a ***structural motif*** or a ***sequence motif***. This can lead to confusion since structural motifs do not always have a unique amino acid signature. It is important to note that the terms ***domain***, ***fold***, and ***motif*** are often used interchangeably with blurred distinctions. This happens especially for larger, highly conserved structural motifs which may be as equally well known as a particular type of domain or fold.

Quaternary Structure

The structure of many proteins, especially those >100 kDa in mass, are oligomers consisting of more than one polypeptide chain. The precise spatial arrangement of the subunits within a protein is referred to as the quaternary structure.

Frequently Observed Secondary Structure Assemblies Or Structural Motifs

This section describes some structural motifs commonly observed in a wide variety of proteins. Other structural motifs of particular functional significance, such as DNA- and RNA-binding motifs, are described in later sections of this unit.

α-*helix bundles*

Helix bundles (or helical bundles) usually refer to the packing of several helices in a nearly parallel or antiparallel orientation. In some cases the term also applies to the general packing of helices regardless of their orientation to each other. There are many examples of protein structures that contain three-, four-, or multiple-helix bundles, such as cyclin (three-helix bundle), hematopoietic cytokines (four-helix bundle), globins (six-helix bundle), and cytochrome c oxidase (22-helix bundle). Helix bundles are often distinguished by the number of helices in the bundle, the relative direction of the helices, and the characteristic twist of each helix bundle. In describing the direction of the helices in a bundle, both parallel/antiparallel and up/down terminologies are used in the literature, although the latter terms give a more precise description (Fig. 17.1.1F). In the first convention, two helices are said to be parallel if their N- to C-terminal vectors, or peptide directions, are parallel to each other (i.e., pointing in the same direction). The opposite is referred to as antiparallel. In the up/down notation, the first helix in a bundle is defined as the 'up' helix, and all subsequent helices are assigned either 'up' or 'down' depending on whether they are parallel or antiparallel, respectively, to the first helix. In most helical bundles, the axes of individual helices are inclined to the representative axis of the entire bundle in a systematic manner, forming a twist with defined handedness that resembles a spiral. Most helix bundles carry a left-handed twist.

Coiled-coil

As a special case of helix bundles, ***coiled-coil*** refers to two or more helices intertwined with each other to form a supercoiled helical structure. The helices involved in a coiled-coil configuration are normally long, with certain characteristic repeats in their amino acid sequences, and are usually parallel rather than antiparallel to each other. In most cases, the participating helices of a coiled-coil are from separate subunits but reside in a topologically equivalent part of a protein.

β-*hairpin*

When an antiparallel β-sheet contains only two strands connected by a single tight turn, it is termed a β-hairpin (Fig. 17.1.1G). A β-hairpin is usually highly twisted and less regular compared to a more extended β-sheet.

Greek key and jelly-roll motifs

In an antiparallel β-sheet assembly, four adjacent strands are sometimes arranged in a pattern similar to the repeating unit of one of the ornamental patterns used in ancient Greece. Thus, this type of strand assembly is termed a ***Greek key*** motif (Fig. 17.1.1H). The immunoglobulin fold is one example of a fold that contains the Greek key motif. A certain arrangement of two Greek key motifs form the so-called ***jelly roll*** (Fig. 17.1.1I). In this case, strands 1, 2, 7, and 8 form one Greek key, while strands 3, 4, 5, and 6 form the other. The eight strands fold into a nearly closed barrel shape. A characteristic feature in the topology of any β-sheet is the arrangement of all the strands in the sheet, namely the ***strand order*** within the hydrogen bonding of the sheet. This hydrogen bonding-directed strand order is also called ***strand connectivity***. In the case of a jelly roll motif, this strand connectivity has an order of 1-2-7-4-5-6-3-8.

β-*sandwich*

β-sandwich refers to a set of β-strands arranged into two β-sheets packed face-to-face against each other (Fig. 17.1.1J). The two sheets can be made out of parallel, antiparallel, or mixed strands. In many β-sandwiches, the strands of the two sheets are orientated either parallel/antiparallel or approximately orthogonal to each other. A common feature in some types of β-sandwiches is ***strand switching***, which involves an edge strand that hydrogen bonds to one sheet at its N-terminal end and the other at its C-terminal end.

Mixed α/β *sandwich*

Aside from the β-sandwich, other layered secondary structure assemblies include α/β, α/β/α, and β/β/α sandwiches. For example, an α/β/α-sandwich is defined as structures with one layer of β-sheet (middle layer) sandwiched in between two layers of α-helix. Both layers of helix pack closely against the middle β-sheet layer. The α//β/α sandwich is commonly ob-

served in nucleotide-binding proteins (see DNA-Binding Motifs and Domains; also see RNA-Binding Structural Motifs and Domains).

β-*barrel*

β-barrels are assemblies of β-strands folded into closed barrel-like structures when viewed from the ends of the strands (Fig. 17.1.1K). The size and shape of a barrel vary widely in protein structures. A given barrel can have as few as five strands, such as in the src homology SH3 domain (see Modular Domains Involved in Signal Transduction), or as many as 22 strands, as observed in the structures of bacterial outer membrane proteins FepA and FhuA (see β-Barrel Membrane Proteins). In addition, their strands can be arranged in parallel, antiparallel, or mixed orientations.

α/β-*barrel*

In contrast to a β-barrel, which consists entirely of β-strands, an α/β-barrel is made out of alternating α and β structures. A typical example of an α/β-barrel fold is found in the structure of triosephosphate isomerase (TIM; see The TIM-Barrel Fold). The structure has eight repeats of an alternating unit consisting of a β-strand followed by an α-helix (Fig. 17.1.1L). It is organized such that the eight β-strands form an inner β-barrel which is then surrounded by eight α-helices. α/β-barrels are frequently observed in enzyme structures.

β-*propeller*

β-propellers consist of a disk-shaped circular arrangement of four to eight small β-sheets around a central tunnel (for reviews see Fulop and Jones, 1999; Jawad and Paoli, 2002). Each blade of the propeller comprises a twisted, four-stranded antiparallel sheet with β-meander topology (i.e., one strand directly follows another, forming an array of β-hairpins; also see Lipocalins) that packs face to face with neighboring blades, commonly through hydrophobic interactions (Fig. 17.1.1M). β-propeller blades generally align with 40- to 60-residue sequence repeats that start with either the second, third, or fourth β-strand, and continue into the next sheet for a total of four strands. Ring closure usually occurs with the N- and C-terminal β-strands coming together within the final propeller blade in an association sometimes referred to as 'molecular Velcro'. β-propeller functions vary widely and include enzymatic catalysis (e.g., collagenase, galactose oxidase, sialidase), ligand binding (e.g., tachylectin-2, haemopexin), signaling (e.g., seven-bladed G protein β-subunit), protein-protein interaction (e.g., integrin extracellular segment), and scaffolding functions (e.g., constitutive photomorphogenic 1 (COP1) protein). Well known repeating sequence motifs observed in β-propeller proteins include the WD motif (also called WD40), the kelch motif, the RCC1 repeat, the TolB repeat, the YWTD repeat, the aspartate box, the tachylectin-2 repeat, and the tryptophan docking motif (Fulop and Jones, 1999).

WEB-BASED STRUCTURAL BIOINFORMATICS

The rapid growth of protein structural data along with the simultaneous growth of the World Wide Web is facilitating the assembly and posting of numerous structural biology databases for researchers throughout the world. Concurrently, web servers that provide interactive computations such as protein structure alignments, graphics-based structural analysis, and pattern matching are also being developed at a swift pace. This section summarizes several helpful bioinformatics websites that relate to protein structure, classification, and structural folds; however, due to the large number of sites available, some excellent ones are not discussed in this unit. Instead, compiled here are some representative web sites that appear suited for a broad range of researchers interested in structural biology; sites focused on more narrow topics or technical issues are not discussed. The databases are arranged into the following categories:

1. *Identification and annotation of protein domains and/or folds.* Describes websites that serve as databases of protein folds and motifs and also allows new protein sequences submitted by the user to be scanned for domain/motif profiles.
2. *Protein structure databases.* Discusses worldwide databases for protein structures and their services.
3. *Molecular graphics programs for PC and Macintosh personal computers.* Brief descriptions and listings of freely available graphics programs for displaying and manipulating macromolecular structures on PCs and Macintosh (Mac) computers.
4. *Classification of protein structures.* Databases that classify protein folds and domains by various protocols.
5. *Structural alignments.* Databases that provide structural alignments of proteins as well

as servers that align user-submitted structures interactively.

For more detailed descriptions of the databases as well as instructions on how to use them, the reader is referred to *Current Protocols in Bioinformatics* (Baxevanis et al., 2004).

Identification and Annotation of Protein Domains and/or Folds

There are now numerous database projects with interactive web sites that serve to categorize proteins, or portions of proteins, according to families, domains, motifs, or even functional sites using various types of sequence clustering methods or profile matching, usually using Hidden Markov Models (HMMs). These database web sites enable the user to upload sequences to be scanned interactively for real time diagnostic classification and annotation. Most sites also allow the user to browse the site's database to view annotation on a particular type of domain or motif, including a short descriptive paragraph and key literature references, or to view the complete annotation of a particular protein in the database. Below are brief descriptions of some of the more well known databases.

PROSITE

This database (Bairoch and Bucher, 1994; Falquet et al., 2002) is part of the ExPASy (Expert Protein Analysis System) proteomics project (Appel et al., 1994) operated by the Swiss Institute for Bioinformatics (SIB). The PROSITE web server (*http://us.expasy.org/prosite*) uses protein-sequence-profile matching to identify a wide variety of functional regions in protein sequences. In addition to some large domains, PROSITE can also scan for very short stretches of less than five residues (e.g., regions of posttranslational modifications). At the time of this writing, PROSITE included 1178 documentation entries, and 1614 different patterns, rules, and profiles/matrices.

SMART

The Simple Modular Architecture Research Tool (SMART; Schultz et al., 1998; Letunic et al., 2002) specializes in identifying and annotating mobile eukaryotic protein domains. The general focus of this database is on regulatory domains and extracellular modules with little emphasis on enzymatic domains. Over 650 different domains (HMMs) are currently included in the SMART database. SMART can detect signal peptides, transmembrane regions, coiled-coils, and internal repeats as well. The SMART web site (*http://smart.embl-heidelberg.de*) provides numerous annotation hyperlinks for each type of domain, including the current number of proteins identified with the domain, lists of proteins containing the domain, a comprehensive sequence alignment of domain family members, a consensus sequence for the domain family, the species distribution of proteins containing the domain, the PDB entry codes (hyperlinked to PDBsum) for solved structures containing the domain, and links to the InterPro database (see below).

Pfam

The Pfam database is a manually curated (for the most part) collection of multiple sequence alignments and HMMs that describe various protein families (Bateman et al., 2002). It is maintained by the Wellcome Trust Sanger Institute. Approximately one-quarter of Pfam is automatically generated from the PRODOM database and is referred to as Pfam-B. Pfam is built solely from proteins within the SwissProt/TrEMBL protein sequence database releases. Over 5200 groups of homologous sequences are included in Pfam. They are each classified as either a family, domain, repeat, or motif. So-called nondomain regions, such as signal sequences, transmembrane regions, and coiled-coils, are also predicted through third party software. The Pfam web site (*http://www.sanger.ac.uk/Software/Pfam*) and its four mirror sites provide an interface for uploading sequences and searching the database. Annotations include an InterPro abstract (when available), the protein family sequence alignment, hyperlinks to the HMMs that were used, the known species distribution, and various hyperlinks (if available) to entries in other databases such as PROSITE, RCSB PDB, SCOP, HOMSTRAD, PDBsum, COGS, SYSTERS, and InterPro. When structures are available, they can be viewed interactively on graphics programs (if installed on the user's PC or Mac) by selecting them and clicking on the Chime or RasMol web page buttons (see below for databases and software).

InterPro

The InterPro database (*http://www.ebi.ac.uk/interpro*) is a consortium of several different database projects for annotating protein families, domains, and motifs (Apweiler et al., 2001). It is maintained by the European Bioinformatics Institute (EBI) which is part of the European Molecular Biology Laboratory (EMBL). Member databases include Pfam,

SMART, PROSITE, TIGRFAM, PRINTS, and ProDom, in addition to the protein sequence database SwissProt/TrEMBL. InterPro has combined data from these seven member databases to form a much larger, coherent database. Redundant, overlapping entries are merged into single unique InterPro entries and hierarchical family relationships are modified accordingly. Through the InterPro web site, text searches of the InterPro database can be made or sequences can be uploaded and scanned against the database. The web page for each InterPro entry provides a brief descriptive abstract (frequently used by or linked to other databases), links to member database entries, and links to related InterPro domain entries, in addition to literature reference links.

Databases not discussed

Other similar databases not discussed in this section include TIGRFAM (*http:// www.tigr.org/ TIGRFAMs/index.shtml*), PRINTS (*http:// www.bioinf.man.ac.uk/dbbrowser/ PRINTS*), ProDom (*http://prodes.toulouse. inra.fr/prodom/ doc/prodom.html*), Blocks (*http://www. blocks.fhcrc.org*), and CDD (*http://www.ncbi. nlm.nih.gov/Structure/cdd/ cdd.shtml*).

Protein Structure Databases

RCSB PDB

The Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) is a comprehensive database of three-dimensional protein and nucleic acid structures determined by X-ray crystallography, NMR, cryoelectron microscopy, and theoretical modeling (Berman et al., 2000). It is maintained by Rutgers, The State University of New Jersey; the San Diego Supercomputer Center (SDSC) at the University of California, San Diego; and the National Institute of Standards and Technology (NIST). The PDB was established in 1971 at Brookhaven National Laboratories (BNL) with the deposition of seven structures. Since then, management of the PDB has been transferred to the RCSB (fall, 1998) and well over 20,000 structures have been deposited. The PDB has become an international resource for macromolecular structural coordinates and most peer-reviewed journals now require the deposition of coordinates in this database prior to publishing structural data.

PDB structures can be easily accessed through the main web site (*http://www. rcsb.org/pdb*) or six other international mirror sites by searching key words or PDB entry codes. An advanced search form is also available at *http://www.rcsb.org/pdb/cgi/queryForm.cgi*. From the resulting list, entries can be chosen by clicking on the EXPLORE hyperlink on the right-hand side of the page. At this point, a wealth of information is available; the following is a list of the hyperlinks available from this page (also refer to the documentation available from the web site):

1. *Download/Display File.* Allows downloading or displaying the coordinate file.
2. *Medline.* Provides the MedLine abstract of the primary publication describing the structure.
3. *View Structure.* Permits interactive viewing of the structure. On this page, links are available for coordinates that are formatted specifically to be parsed into interactive graphics programs such as RasMol, Chime, Swiss PdbViewer, VRML, or MICE (see below) by simply clicking on the name of the program. Links are provided on the same page that explain the downloading and installation of these programs. If there are no graphics programs installed on the user's computer, a simple Java applet called QuickPDB is also available requiring no installation. Clicking on the QuickPDB button provides a red colored wire α-carbon trace of the structure that can be rotated, translated, or magnified (zoomed) with a mouse. The protein sequence(s) is displayed as well, enabling residues to be selected by highlighting regions of the sequence. As of early 2003, QuickPDB can display only protein structures; DNA, RNA, and ligand structure display is not supported. Additionally, prearranged static images of the molecule created by RASTER3D are available in jpg and tiff formats from the View Structure web page.
4. *Structural Neighbors.* Clicking on this will provide links for that particular structure within the CATH, CE, FSSP, SCOP, or VAST databases (see below).
5. *Geometry.* This provides considerable structural analysis of the structure including outputs from WHAT CHECK and PROCHECK (comprehensive geometric analyses), PROMOTIF (secondary structure analysis), castP (pocket and cavity analysis), LPC (ligand contact analysis), and CSU (contact analysis), in addition to a Ramachandran plot, and tabulated analyses of average bond length, bond angles, and dihedral angles.
6. *Sequence Details.* This link includes sequence(s) from the coordinate file in FASTA format and links to sequence databases.

Other Existing Structural Databases

MSD. The Macromolecular Structure Database (MSD; *http://www.ebi.ac.uk/msd/index.html*) at the European Bioinformatics Institute (EBI) manages and distributes macromolecular structural data. It collaborates closely with the RCSB PDB (see above), providing services for the deposition (AutoDep; *http://autodep.ebi.ac.uk/*, retrieval (OCA browser; *http://oca.ebi.ac.uk/oca-bin/ocamain*) and analysis (Biotech validation server; *http://biotech.ebi.ac.uk:8400*) of PDB data. Other MSD services are listed below.

1. SSM. Through the Secondary Structure Matching (SSM) server (*http://www.ebi.ac.uk/msd-srv/ssm/ssmstart.html*) one can pick a structure from the PDB or upload a coordinate file and then scan the PDB or SCOP databases for the best C^{α} three-dimensional superpositions. These superpositions can then be viewed immediately using RasMol or RasTop.
2. The Ligand Chemistry Search Service. This site provides access to the Ligands and Small Molecules Dictionary of the MDS database encompassing all chemically distinct components found within the PDB (*http://www.ebi.ac.uk/msd-srv/chempdb/cgi-bin/cgi.pl*). Included are ligands, small molecules, hetero groups, and standard and nonstandard amino acids. Various chemical details such as formula, atom connectivity, and charge can be retrieved in addition to the coordinates of real or idealized structures.
3. Sequences of structural genomics projects. A list of structural genomics targets from The National Institute of General Medical Sciences (NIGMS)–supported research centers has been compiled, converted into FASTA format, and made accessible at this site (*http://www.ebi.ac.uk/msd/projects/SGtargets.html*).
4. PQS. The Protein Quaternary Structure (PQS) file server (*http://pqs.ebi.ac.uk*) uses an automated procedure (Henrick and Thornton, 1998) to apply crystallographic symmetry to PDB crystallographic coordinates. The resulting oligomers are analyzed by multiple criteria to predict which interchain contacts are nonspecific crystal contacts and which specifically contribute to an oligomeric complex. These results are then used to predict the most likely quaternary state of the macromolecule. Advanced or simple search forms are available to look up the results for current PDB entries; however, currently, users cannot upload files for analysis.
5. The 3D-EM Macromolecular Structure Database. This is a newly launched database for macromolecular electron microscopy data such as volume maps, sections, structure factor files, layer-line data, figures, and textual descriptors (*http://www.ebi.ac.uk/msd/iims/3D_EMdep.html*).

MMBD. The Molecular Modeling Database (MMDB; *http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml*) is maintained by the National Center for Biotechnology Information (NCBI), part of the National Institutes of Health (NIH). It is a monthly updated database containing experimentally determined macromolecular structure coordinates downloaded form the RCSB PDB (see above). Theoretical models are excluded. Using NCBI's Entrez data retrieval system, one can run key word searches to find protein structures. Searching results in a list of hyperlinked PDB entry codes with brief one- or two-line descriptions. Upon choosing a particular hyperlink, an MMDB Structural Summary is presented. From this page, several hyperlinked options are available. These include:

1. The PDB web page for that particular entry.
2. A list of related PubMed references.
3. NCBI's Taxonomy browser.
4. A three-dimensional view of the protein using NCBI's own molecular graphics program, Cn3D (obtained by clicking View 3D Structure; program must be installed locally).
5. A list of structurally homologous proteins (referred to as structural neighbors) generated by the program VAST (Vector Alignment Search Tool; obtained by clicking on the colored bar representing the protein chain or domain of interest). Superpositions of these structures with the original structure can be viewed using the Cn3D or MAGE programs.
6. Protein sequence alignments between various domains in the structure and other sequences within the same domain family from NCBI's Conserved Domain Database (CDD).

Molecular Graphics Programs for PC and Macintosh Platforms

High-end graphics work stations are no longer needed to view three-dimensional models of macromolecular structures or perform simple manipulations in real time. This can now be accomplished with PC or Mac personal computers running any of a variety of operating systems. Below is a sample listing from the growing inventory of personal computer-friendly molecular graphics programs that are freely available from the World Wide Web. Unless otherwise noted, each program listed runs on both PC and Mac platforms. These

programs enable the user to input PDB-formatted coordinates and then view ribbon, CPK, wireframe, or combination models of macromolecules which can be rotated, translated, or magnified. The user can then select certain residues or atoms for manipulations and output static images for future reference. Most programs have several additional features, such as selecting atom-atom distances, displaying hydrogen bonds, and viewing in a stereodisplay or dot-surface format. Despite the growing popularity of these small graphics programs, many structural biologists often use more sophisticated programs (e.g., O, Ribbons, MolScript, Raster3D, GRASP) on UNIX or LINUX platforms to model, build, and analyze structures, as well as generate figures for publication. These programs are not discussed here.

RasMol

RasMol is probably the most popular program for viewing macromolecular structures on PC and Mac computers (Sayle and Milner-White, 1995; Bernstein, 2000; *http://www.bernstein-plus-sons.com/software/ rasmol*). Web sites that provide access to macromolecular coordinates frequently have hyperlinks that will activate a local copy of RasMol and parse coordinates into the viewer.

Chime

This program, currently available from MDL Information Systems (*http://www.mdlchime.com/chime*), was built partly from RasMol code and runs as a web browser plug-in using Netscape or Internet Explorer as the viewing platform. Like RasMol, database web pages providing macromolecular coordinates often have hyperlinks specifically for running Chime; however, unlike RasMol, Chime only reads in the coordinates supplied by the sponsoring website. There are several Chime-interfacing programs such as Protein Explorer (*http://www.proteinexplorer.org*), STING (*http://honiglab.cpmc.columbia.edu/STING/help*), and Noncovalent Bond Finder (*http://www.umass.edu/microbio/chime/find-ncb/index.htm*) that provide enhanced or additional features to Chime, including different interfaces for program control, simple animation, automation of tasks, highlighting atomic interactions, or enabling users to upload their own PDB files for viewing. Further documentation and Chime resources are available at *http://www.umass.edu/microbio/chime*.

Mage

This program is particularly well suited for use as a teaching or presentation tool and can run simple animation. It is also the oldest of the macromolecular graphics freeware available for personal computers (Richardson and Richardson, 1992; *http://kinemage.biochem.duke.edu/kinemage/kinemage.php*).

Deep View/Swiss-PdbViewer

This program was formerly known as simply the Swiss-PdbViewer. Deep View is one of the more sophisticated programs freely available for use on personal computers (*http://us.expasy.org/spdbv*). Features include performing torsion manipulations and mutations of residues, displaying electron density maps, energy minimization, computing and displaying electrostatic potential maps, displaying molecular surfaces, and generating POV-ray scenes. It is also closely linked with the molecular modeling server, SWISS-MODEL (*http://www.expasy.org/swissmod*).

RasTop

This program is adapted from RasMol; however, it provides a menu-driven graphics interface rather than the command-line interface that RasMol uses (*http://www.geneinfinity. org/rastop*). As of early 2003, RasTop runs on Windows but not Mac platforms (except within environments such as virtual PC).

Cn3D

This program is specifically designed for NCBI's MMDB system and reads only ASN.1-formatted coordinate files (Wang et al., 2000b; *http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml*). In addition to being a molecular-graphics viewer, Cn3D also displays the protein sequence in a second window. As regions of the sequence are highlighted with a mouse, the corresponding residues are also highlighted in the structure. Furthermore, homologous sequences can be imported for alignment, enabling one to map sequence differences directly onto the three-dimensional structure. By accessing NCBI's VAST database, the user can also view superpositions of structurally related proteins in Cn3D.

MolMol

This software has been developed with a special emphasis on displaying protein or nucleic acid structures determined by NMR (Koradi et al., 1996; (*http://www.mol. biol.ethz.*

ch/wuthrich/software/molmol). The program runs only on UNIX or Windows platforms.

PocketMol

This software runs on Pocket PCs running Windows CE 3.0 or higher, using a stylus for molecule manipulations (*http://birg.cs.wright.edu/pocketmol/pocketmol.html*).

MolView

This is a general purpose molecular viewer that runs on Mac platforms only (Smith, 1995; *http://www.danforthcenter.org/smith/MolView/molview.html*). MolView can also create QuickTime movies.

Qmol

This viewer generates molecular surfaces and AVI movies (Gans and Shalloway, 2001; *http://www.mbg.cornell.edu/shalloway/jason/qmol.html*.

WebMol

A JAVA based viewer that can run as an applet or a stand-alone application (Walther, 1997; *http://www.cmpharm.ucsf.edu/~walther/webmol.html*).

Classification of Protein Structures

Several web-accessible databases have been developed that classify the domains of known three-dimensional structures of protein molecules in a hierarchical manner. Three of the most well known databases, SCOP, CATH, and FSSP, are briefly described below.

SCOP

The Structural Classification of Proteins (SCOP) database classifies protein domains and multidomain regions with known structure according to evolutionary and structural relationships (Murzin et al., 1995; *http://scop.mrc-lmb.cam.ac.uk/scop/index.html*). For most entries, hyperlinks are available to run RasMol or Chime (see above) or to display a static jpg image of the molecule. This database is periodically curated/modified in a manual fashion with the aid of various automatic programs (Lo et al., 2002). SCOP release 1.61 (November 2002) includes 44,327 domains from 17,406 proteins. Structures are classified at four major levels (i.e., class, folds, superfamily, family).

1. *SCOP class*. Includes, alpha, beta, alpha/-beta (mostly parallel β-sheets), alpha/beta (mostly antiparallel β-sheets), multidomain alpha/beta, membrane, cell surface proteins (excluding proteins of the immune system), and small proteins.
2. *Folds*. Classified according to the arrangement of secondary structure elements and topological connections.
3. *Superfamilies*. Proteins or domains with low sequence identity are grouped into superfamilies according to structural and functional similarities to reflect potential evolutionary relationships.
3. *Families*. Structures are clustered into families if they have sequence identities ≥30% or if they display compelling functional and structural similarity.

CATH

The Class, Architecture, Topology, Homologous (CATH) Superfamily database is assembled from crystallographically determined protein structures with resolution better than 3.0 Å and from NMR structures, using a largely automated set of procedures (Orengo et al., 1997, 2002; *http://www.biochem.ucl.ac.uk/bsm/cath_new/index.html*). After delineating domain boundaries, protein domains are systematically classified at the five different levels listed below, forming a hierarchical tree. Domain entries can be viewed via RasMol hyperlinks (see above) or static gif images. CATH database release 2.4 (January 2002) contains 39,480 domains.

1. *Class*. Each domain is designated a specific class, namely alpha, beta, alpha/beta, or structures with few secondary elements.
2. *Architecture*. Domains are grouped by architecture according to the general three-dimensional orientations of their secondary structure elements.
3. *Topology*. Domains are categorized into topology or fold group by the connectivity of their secondary structure elements.
4. *Homologous superfamily*. Topology groups are further divided into homologous superfamiles by homology utilizing sequence, and structural and functional data.
5. *Family*. Final classification is by sequence families according to sequence identity.

FSSP

The Families of Structurally Similar Proteins (FSSP) database comprises all protein chains from the PDB (see above) that are longer than 30 residues (Holm and Sander, 1996, 1998; *http://www.ebi.ac.uk/dali/fssp*). Sequence redundancy is removed by generating a representative set of sequence-unique proteins, none of which have >25% sequence iden-

tity with any others in the set. Each member of the representative set represents a group of sequence homologs with >25% sequence identity. Pairwise structural alignments are made (using the Dali program) between all members of the representative set and between each representative protein and its sequence homologs. The individual alignments, complete with RasMol links (see above) to view three-dimensional superpositions, color-coded sequence alignments, and various alignment statistics such as the root mean square deviation (RMSD), Z-scores, and percent sequence identity, are available on the web. These structural comparisons are used in a hierarchical clustering procedure to create a ***fold tree*** within FSSP for the classification of protein structures. A *Z*-score cutoff of ≥2 is used to define individual folds at the top of the hierarchy. Subsequent *Z*-score cutoffs of 4, 8, 16, 32, and 64 are used to designate further classification levels with increasing structural similarity. Moreover, another database, referred to as a supplement to FSSP and called the Dali Domain Dictionary (*http://www.ebi.ac.uk/dali/domain*), has recently been built in a very similar manner but focuses on protein domains rather than whole proteins (Dietmann and Holm, 2001).

Structural Alignments

Several databases accessible through the World Wide Web offer comprehensive collections of structural alignments between known protein structures. Some servers also allow the user to upload coordinate files for interactive structural alignments. Proper cutoff values applied to alignment parameters result in lists of structural neighbors for each entry. Three of these database servers are briefly described below.

HOMSTRAD

The Homologous Structure Alignment Database (HOMSTRAD) groups protein structures determined by crystallography or NMR into several hundred families according to sequence identity and provides structural alignments within the families that can be downloaded or viewed via RasMol (see above) hyperlinks (Mizuguchi et al., 1998; *http://www-cryst.bioc.cam.ac.uk/~homstrad*). As of early 2003, 1033 multimember families and 3187 single member families were included in the database.

VAST

The Vector Alignment Search Tool (VAST) provides a systematic algorithm for aligning and comparing the three-dimensional structures of proteins (Madej et al., 1995; Gibrat et a1., 1996; *http://www. ncbi.nlm.nih.gov/ Structure/VAST/vast.shtml*). More than 18,000 protein domains within the NCBI's Molecular Modeling Database (MMD) have been compared with each other in a pairwise fashion using VAST and are stored in an NCBI database. From these results a list of 'Structural Neighbors' has been tabulated for each domain within the MMD. The alignments of structural neighbors can be viewed using NCBI's Cn3D or the MAGE program (see above). Furthermore, the user can upload new structures using the ***VAST Search*** web page (*http://www.ncbi. nlm.nih.gov/Structure/VAST/ vastsearch.html*) to run VAST interactively and obtain a list of structural neighbors.

CE

The Combinatorial Extension (CE) web server uses a structural alignment method that involves aligning fragments of proteins and then searching for continuous paths of aligned fragment pairs (AFPs; Shindyalov and Bourne, 1998; *http://cl.sdsc.edu/ce.html*). Several options available for the user are listed below.

1. Choose a protein from the PDB (see above) and scan the entire PDB or representative proteins from the PDB for structural neighbors.
2. Upload a coordinate file and scan against the PDB
3. Calculate the structural alignment between two proteins (PDB entries or user-uploaded).
4. Perform multiple sequence alignment of proteins in the PDB using a Monte Carlo–based method.

The aligned structures generated can be downloaded as a PDB file or viewed interactively using RasMol (see above), Protein Explorer (requires Chime; see above), or the web site's own Java applet called Compare3D that requires no locally installed program.

PROTEINS INVOLVED IN THE FUNCTION OF IMMUNE SYSTEMS

Immunoglobulins and Immunoglobulin-Like Superfamilies

Traditionally, the term immunoglobulin (Ig) fold refers to the homologous domain structures present in the variable and constant regions of all immunoglobulins. In recent years, however, with the rapid growth of sequence and

structure knowledge, many other proteins whose functions are not related to immunoglobulins have nonetheless been found to contain one or more Ig-like domains in their tertiary structures. Examples of such proteins include the cell surface glycoprotein receptors such as growth hormone receptor, CD4 and CD8, adhesion molecules such as cadherins and type III fibronectins, class I and class II major histocompatibility proteins, chaperone protein PapD, the transcriptional factor NF-κB, and even some enzymes such as Cu and Zn superoxide dismutase and β-galactosidase. The discovery of the Ig-like domains in many proteins not only broadens the definition of superfamily, but more importantly, demonstrates the versatility of the fold and the ability of these domains to function as modular units in many diverse systems.

The immunoglobulin fold is defined as a domain that is formed from seven to nine β-strands folded into a Greek key β-sandwich (Fig. 17.1.2; also see above). The topology consists of one β-sheet formed by strands A, B, E, and D, while the opposing sheet is formed by strands C, C′, and occasionally C″, as well as F and G. Based on the strand hydrogen-bonding pattern and the number of strands in each domain, the superfamily is further divided into five commonly observed subclasses: V, C1, C2, I, and E (Fig. 17.1.2). The V-type domain has the same topology as immunoglobulin variable (V) regions. It has nine strands with a typical strand switch occurring in the first strand, which usually breaks into two halves (A and A′) with the first half (strand A) hydrogen bonding to strand B of the B-E-D β-sheet and the second half (strand A′) hydrogen bonding to strand G of the opposing β-sheet. Strands C′ and C″ are usually shorter than the other strands. In immunoglobulins, the antigen binding site is composed primarily of three loop regions termed complementary determining regions (CDRs). They are located between strands B and C, C′ and C″, and F and G. The C1 type of immunoglobulin fold refers to the topology observed in the structure of constant regions of immunoglobulins. It has seven strands, of which strands A-B-E-D form one β-sheet and strands C-F-G form the other. Most members of the C2-type subclass have seven strands and differ from the C1 type by a simple switch whereby strand D hydrogen bonds with strand C in the C-F-G sheet rather than strand E in the A-B-E sheet. Hence strand D is labeled C′ in the C2-type subclass. Some members of the C2 family also have additional strands at either the N or C termini. Based on tertiary structure similarities, there are two frequently adopted tertiary folds in the C2 class; one is similar to the C2 domain of T-lymphocyte cell surface glycoproteins CD2 and CD4, while the other is similar to the fibronectin type III domain, and hence is also referred to as the fibronectin type III-like fold, shown as C2(Fn) in Figure 17.1.2C. The I-type domains closely resemble the topology of immunoglobulin-variable domains except that they lack strands C′ and C″. The members of the I-type subclass also have the characteristic strand switching phenomena in the first strand, as do the immunoglobulin variable domains. The E-type immunoglobulin domains also present a topology similar to those of the V type with the difference being the lack of C″ and D strands. Table 17.1.1 summarizes all the observed immunoglobulin subclasses.

Among the members of the immunoglobulin superfamily, occasionally, additional strands are observed at the N and C termini and sometimes insertion strands which lie outside the core immunoglobulin fold can be found connecting the adjacent immunoglobulin strands.

The MHC Peptide Binding Fold

The three-dimensional structures of class I and class II major histocompatibility complex (MHC) molecules reveal a distinct fold associated with the peptide-binding domain. In addition to MHC molecules, the structures of neonatal Fc receptor (FcRn), hereditary haemochromatosis protein HFE, MHC-I chain homolog (MIC), UL16 binding protein (ULBP), and Rae-1 have all been shown to contain an MHC-like fold (Radaev et al., 2001b; Boyington and Sun, 2002). MHC molecules are heterodimers consisting of an α/β peptide-binding domain and two Ig C1-type domains (Fig. 17.1.3). In class I MHC molecules, the peptide-binding α1α2 domain and the α3 Ig domain reside on one chain (the heavy chain), while the second Ig domain is contributed by an associated β2m molecule (Fig. 17.1.3A). Class II MHC molecules consist of an α and β chain (Fig. 17.1.3B). The N-terminal part of each chain contributes to half of the peptide-binding domain and the C-terminal half of each chain provides a single Ig domain. The MHC Ig domains all have Ig C1-type folds (see Immunoglobulins and Immunoglobulin-Like Superfamilies). In class I MHC molecules, the peptide-binding domain is formed by the α1 and α2 regions of the heavy chain whereas

in class II molecules it is formed by the α1 and β1 regions of both α and β chains. The peptide-binding domain consists of two α-helices (one from each region), forming the two sides of the groove and an eight-stranded antiparallel β-sheet (four strands from each region), forming the floor of the groove (Fig. 17.1.3C). The size of the groove is ~30 Å long and 12 Å wide in the middle. Both class I and II MHC molecules bind peptides in extended conformations with anchor residues buried inside pockets of the groove. The distribution of these pockets along the binding site determines the peptide selectivity of a particular allele of class I or class II molecules. Class I MHC molecules are restricted to binding peptides of 8 to 11 residues, whereas class II MHC molecules bind longer peptides of 15 to 24 residues. In contrast, none of the MHC-like proteins, such as FcRn, HFE, MICs, ULBPs, and Rae-1s associate with peptides. Their putative peptide-binding grooves are narrower or even closed compared to the MHC peptide-binding groove (Table 17.1.1).

The C-Type Lectin-Like Receptor Fold

In addition to immunoglobulin-like superfamilies, many recently identified immunoreceptors are members of the C-type lectin-like superfamily. Examples include mannose-binding protein, FcεRII (CD23), CD94, members of NKG2 family receptors, Ly49 receptors, and DC-SIGN. Their structures resemble that of C-type lectin, a calcium-dependent carbohydrate-binding protein (see C-type Lectins). What is unique to many of these C-type lectin-like receptors is the absence of a functional calcium-binding site and their ability to form a common disulfide-bonded dimer (Boyington et al., 1999). Table 17.1.1 lists the known structures of C-type lectin-like receptors.

Structure of Immunoreceptor-Ligand Complexes

The goal of structural immunology in recent years has been to define molecular structures of immunoreceptor-ligand complexes rather than revealing new protein folds. The notable milestones include the publications of several T cell receptors and their cognate MHC-peptide complexes (Garboczi et al., 1996; Garcia et al., 1998; Reinherz et al., 1999)—i.e., T cell coreceptor CD8 or CD4 and their class I or II MHC complexes (Gao et al., 1997; Wang et al., 2001), T cell coreceptor CTLA-4 and ligand B7 complex (Stamper et al., 2001; Zhang et al., 2001), NK cell killer immunoglobulin-like receptor (KIR) and its MHC ligand complexes (Fan et al., 2001; Boyington et al., 2000), NK cell activating receptor NKG2D and its ligand complexes (Li et al., 2001; Radaev et al., 2001b; Li et al., 2002b), and Fc receptors in complex with their Fc ligands (Garman et al., 2000; Radaev et al., 2001a; Sondermann et al., 2000).

Proteins in the Complement System

The complement system is composed of a large number of plasma proteins which participate in an innate immune response to bacteria. Antibodies, mannose-binding lectins, or other complement components binding to the bacterial cell surface can initiate a cascade of complement reactions ultimately leading to recruitment of inflammatory cells, and the opsonization and killing of bacterial pathogens. The following section reviews the structural folds of fragments of several complement components (Table 17.1.2).

C3 fragment C3d

Complement protein C3 is a central component in the opsonization of bacteria. One of its natural cleavage products, C3dg can act as a ligand for the B cell receptor CR2 (CD21) when covalently attached to the pathogen surface. The three-dimensional structure of C3d, a protease-resistant fragment (residues 996 to 1303) of C3dg, reveals an α/α barrel composed of twelve α helices (Fig. 17.1.4; Nagar et al., 1998). The barrel is composed of two concentric rings of six α helices each. Each ring is made of parallel helices inclined slightly from the main axis of the barrel. The topology of the molecule causes consecutive helices to alternate from outside to inside with the orientation of the outer ring helices antiparallel to those of the inner ring. This same fold is also observed in enzymes such as glucoamylase, endoglucanase, and the β subunit of protein farnesyltransferase. The three residues essential for covalent attachment to the pathogen, Cys17 (mutated to Ala in the crystal structure to avoid incorrect disulfide bond formation), Gln20, and His133, are situated at the rim of one end of the barrel (right panel of Fig. 17.1.4A).

C5a

C5a is a 74-residue glycoprotein derived from C5 and is very potent in mediating local inflammatory responses within the complement cascade. The solution NMR structures show a five α helix antiparallel bundle (Fig. 17.1.4B; Zhang et al., 1997). Mutational analysis indicates that the C-terminal twelve residues, including the C-terminal helix, contribute

to agonist activity and receptor binding (Zhang et al., 1997).

Complement factor D

Complement factor D is a serine protease which cleaves complement factor B only when complexed with C3b. The crystal structure shows a typical trypsin fold (see Protease Folds) consisting of two nearly identical six-stranded antiparallel β barrels with Greek key topology (see above), two flanking α helices, and an active site situated between the two barrels (Fig. 17.1.4C; Narayana et al., 1994). The active site is very unusual in that the Asp and His residues of the canonical Asp-His-Ser triad are misoriented from the conformation that is favorable for catalytic activity of a serine protease. It is proposed that complement factor D activity is regulated by a reversible induced-fit mechanism which activates the enzyme only in the presence of its natural substrate, factor B/C3b. Mutation and X-ray studies indicate that three residues, Ser 94, Thr 214, and Ser 215 are responsible for the altered conformation of the active site relative to other serine proteases (Kim et al., 1995a).

Complement regulatory protein CD59

CD59 (also known as MEM-43, MIRL, H19, MACIF, HRF20, and protectin) is a glycophosphatidylinositol (GPI)-anchored glycoprotein that functions as a potent inhibitor of complement-mediated cell lysis and may also play a role in T-cell activation. CD59 belongs to a group of cysteine-rich cell surface molecules known as the Ly6 superfamily which includes the urokinase plasminogen activator receptor, mouse thymocyte antigen ThB, and murine Ly6 antigens. Two NMR solution structures of the extracellular portion (residues 1 to 70) of the protein reveal a disk-shaped molecule with five disulfide bridges comprising a three- and two-stranded antiparallel β-sheet abutting each other, and a two-turn α-helix packed against one side of the larger β-sheet (Fig. 17.1.4D; Fletcher et al., 1994; Kieffer et al., 1994). This structure is also very similar to those of snake venom toxins (see Small Single-Subunit Toxins) with which the Ly6 superfamily shares sequence homology.

Complement control protein modules

Complement control protein (CCP) modules, also known as complement protein (CP) modules, short consensus repeats (SCRs), or sushi repeats, are small domains of ~60 amino acids containing four invariant cysteines forming intramodule disulfide bridges and several other highly conserved residues. These modules are present in tandem arrays in many complement regulatory proteins in addition to a number of cell surface proteins. Three-dimensional structures have been determined for the fifth, fifteenth, and sixteenth CCP domains of factor H, CCP modules 3 and 4 of a Vaccinia virus complement control protein mimic, and CCP modules 1 and 2 from the complement cofactor CD46 (Table 17.1.2; Norman et al., 1991; Barlow et al., 1992, 1993; Wiles et al., 1997; Casasnovas et al., 1999). The domain structure consists of an elongated, twisted β-sandwich formed from a three- or four-stranded antiparallel β-sheet and a two-stranded antiparallel β-sheet (Fig. 17.1.4E). Although the known structures of SCR modules all have essentially the same backbone fold, differences in the number of β strands between various modules arise from differences in hydrogen bonding patterns. Structures containing two tandem CCPs show an elongated head to tail arrangement with elbow angles that vary significantly from structure to structure.

PROTEINS INVOLVED IN SIGNAL TRANSDUCTION PATHWAYS

Cytokines

Four-helix bundles

Members of the hematopoietic cytokines possess a common four-helix bundle topology in their structures. The four helices (labeled a, b, c, d) always pack in a left-handed up-up-down-down manner, with helices a and b, and c and d forming two pairs of parallel helices juxtaposed in opposite directions (Sprang and Bazan, 1993; Wlodawer et al., 1993). The position of the four helices and the packing angles between them are generally well conserved. The superfamily is further divided into long-chain, short-chain, and interferon (IFN) γ–like families of cytokines.

Long-chain family. Members in this family include growth hormone (GH), granulocyte colony-stimulating factor (G-CSF), erythropoietin (EPO), leukemia inhibiting factor (LIF), IL-6, IL-12 α-chain, leptin, and ciliary neurotrophic factor. The four helices of this family are each ~25-residues long (Fig. 17.1.5A and D).

Short-chain family. The short-chain family includes most of the interleukins (i.e., IL-2, -3, -4, -5, -7, -9, and -13), granulocyte/macrophage-colony-stimulating factor (GM-CSF),

macrophage-colony-stimulating factor (M-CSF), and stem-cell factor (SCF). The helices are shorter and uneven in length (eight to fifteen amino acids long) compared to members of the long-chain family. Another signature characteristic of short-chain hematopoietic cytokines is the presence of a pair of short, twisted antiparallel β-strands, one in the loop connecting helices a and b, and the other in the loop connecting helices c and d (Fig. 17.1.5B and D). Many of them also have disulfide bridges in their structures; however, the locations of these disulfide bonds are not well conserved.

Interferon γ–like. This family includes IFN-γ and IL-10. Both contain six helices (a, b, c, d, e, and f) and function as homodimers (Ealick et al., 1991; Zdanov et al., 1995). The dimerization is particularly interesting in this family. The two domains of the dimer are formed by swapping the last two helices (e and f) between the two monomers. The four-helix bundle forming helices in each domain are made out of helices a, c, and d from one monomer, and helix f from the other (Fig. 17.1.5C and E).

β-*Trefoil*

Members of the β-trefoil family include not only cytokines such as interleukin-1α, interleukin-1β, and fibroblast growth factor, but also proteins such as soybean trypsin inhibitor, ricin B, amaranthin, and hisactophilin, which display the same fold (Priestle et al., 1989; Graves et al., 1990; Clore and Gronenborn, 1991; Eriksson et al., 1991; Zhang et al., 1991; Zhu et al., 1991). The overall structure is a six-stranded closed β-barrel (A-D-E-H-I-L) sitting atop three β-hairpins (B-C, F-G and J-K). The β-trefoil has internal pseudo three-fold symmetry (Fig. 17.1.6A) relating three four-stranded antiparallel β-sheets. Each β-sheet is made of four consecutive strands. For example, the three β-sheets of IL-1β are comprised of strands L-A-B-C, D-E-F-G, and H-I-J-K, respectively.

The cystine-knot growth factor superfamily

The recently solved crystal structures of four different growth factors, nerve growth factor (NGF), transforming growth factor-β (TGF-β), platelet-derived growth factor (PDGF), and human chorionic gonadotropin (hCG), have revealed a common overall topology among these functionally diverse protein families (Daopin et al., 1993; Murray-Rust et al., 1993; Murzin et al., 1995; Sun and Davies, 1995). These proteins share very little sequence homology. Nevertheless, they all have an unusual arrangement of six cysteines that are disulfide linked to form a knotted conformation and this core structure, the cystine-knot, is flanked by four strands of an antiparallel β-sheet (McDonald et al., 1991; Oefner et al., 1992; Daopin et al., 1993; Schlunegger and Grutter, 1993; Lapthorn et al., 1994; Wu et al., 1994). Fig. 17.1.6B shows the crystal structure of human TGF-β2 with the common flanking β-strands (1, 2, 3, and 4) shown in green and the six disulfide-linked cysteines shown in ball-and-stick models. Outside the four structurally conserved β-strands, the helices and loops observed in the structure of TGF-β2 are not conserved in NGF, PDGF, and hCG. The three disulfide bonds are formed between the six cysteines of TGF-β2 as follows: Cys 15–Cys 78, Cys 44–Cys 109, and Cys 48–Cys 111, creating a linkage pattern of I–IV, II–V, and III–VI among the six cysteines. A topological knot is formed by threading disulfide I–IV through a ring formed with disulfides II–V and III–VI as its two sides, and the peptides in between cystine II and III, and V and VI as its other two sides. Another common structural feature among the cystine-knot cytokines is that the active forms of these proteins are always dimers, either homo- or heterodimers; however, despite the overall topological similarity between the monomers, each family of growth factors form their own distinct dimers with a unique dimer interface that is characteristic to each family of growth factors.

EGF-like domain

EGF-like domains are often observed in tandem repeats in many proteins involved with a receptor-related signal transduction component (Groenen et al., 1994). Members of this family include epidermal growth factor, transforming growth factor α, the EGF-domain of E-selectin, Factor X, Factor IX, plasminogen activator, prostaglandin H2 synthase-1, and thrombomodulin. An EGF module has ~50 amino acids and is stabilized by three disulfide bonds (Graves et al., 1994; Kohda and Inagaki, 1992; Rao et al., 1995). Its structure elements consist of two small β-sheets: a three-stranded antiparallel β-sheet close to the N terminus and a two-stranded antiparallel β-sheet near the C-terminus (Fig. 17.1.6C). The disulfide bonds are formed between cysteines I–III, II–IV, and V–VI in sequence.

Chemokines

There are two subfamilies of chemokines, the CXC type and the CC type (also known as the α and β type, respectively), referring to the spacing between the first two conserved cyste-

ines in sequence. They are potent chemotactic activators for neutrophils (CXC type), monocytes, and lymphocytes (CC type). The structures of several of these chemokines are available, such as interleukin 8 (IL-8; Baldwin et al., 1990; Clore et al., 1990) of the CXC type and neutrophil-activating peptide-2 (NAP-2) of the CXC type (Malkowski et al., 1995), as well as the macrophage inflammatory protein-1β (MIP-1β), RANTES, and monocyte and chemoattractant proteins (MCP-1 and -3) of the CC type (Table 17.1.3; Lodi et al., 1994; Skelton et al., 1995; Handel and Domaille, 1996; Lubkowski et al., 1997). The two disulfide bonds are formed between the four conserved cysteines in a pattern of I–III and II–IV. Both subfamilies of chemokine are composed of dimeric proteins which share the same structural fold at the monomer level, but they differ drastically in their dimerization (Fig. 17.1.6D). The basic fold consists of a C-terminal α-helix packing against a three-stranded antiparallel β-sheet. The CXC chemokines dimerize on the edge of the first β-strand, forming an extended six stranded antiparallel β-sheet with the two C-terminal helices packing on one side of the sheet. In contrast, the CC chemokines form the dimer using their N-terminal loop region, resulting an elongated dimer with the β-sheets of the monomers opposing each other (Fig. 17.1.6E).

Receptors

TNF receptor family and its associated proteins

The structure of a human 55 kDa tumor necrosis factor (TNF) receptor shows a unique fold consisting of four homologous cysteine repeat domains (Banner et al., 1993). There are no helices and very few regular β-sheets in its secondary structure (Fig. 17.1.7A). Instead, the fold is composed mostly of loops and β-hairpins. Upon complex formation with TNF, there are three receptor molecules symmetrically bound to a TNF trimer.

In addition to TNF, the structure of a TNF-associated protein, TRAF2 (Park et al., 1999), and the structure of a TNF receptor superfamily member, TRAIL, as well as a TRAIL-DR5 complex (Cha et al., 1999; Mongkolsapaya et al., 1999) have also been determined.

Receptors for cystine-knot growth factors

The structures of the ligand-binding domains for several cystine-knot growth factor receptors have been solved to date. They include a type II murine activin receptor (ActRII; Greenwald et al., 1999), a type I human bone morphogenic protein receptor (BRIA; Kirsch et al., 2000), a type II human transforming growth factor receptor (TBRII; Fig. 17.1.7B; Hart et al., 2002), the TrkA subunit of a human nerve growth factor (Wiesmann et al., 1999), and domain 2 of a vascular endothelial growth factor (VEGF) receptor, Flt-1 (Wiesmann et al., 1997; Table 17.1.3). Two of these, TrkA and Flt-1, are tyrosine kinase receptors, whereas the other three are serine/threonine kinase receptors. Structurally, the ligand-binding domains of TrkA and Flt-1 are members of the immunoglobulin superfamily whereas the ligand-binding domains of ActRII, BRIA, and TBRII display a characteristic three-finger toxin fold. The latter three are also part of the so-called TGF-β receptor superfamily.

Nuclear and nuclear transport receptors

The nuclear receptor superfamily is composed of steroid receptors, the thyroid hormone receptor, retinoic-acid receptor, and a number of orphan receptors (Weatherman et al., 1999). All have two domains, an N-terminal DNA-binding domain that functions as a transcription activator, and a C-terminal ligand-binding domain that is specific for different ligands. The DNA-binding domain, rather conserved among the members of the nuclear-receptor superfamily, uses two Zn finger motifs to recognize short DNA repeats (see DNA-Binding Motifs and Domains). The ligand-binding domain is more variable than the DNA-binding domain, but nevertheless displays a similar fold of mostly α-helical segments among its members. To date, the ligand-binding domains of thyroid hormone receptor (TR; Fig. 17.1.7C; Darimont et al., 1998), retinoic-acid receptor (RAR; Renaud et al., 1995), estrogen receptor (ER; Shiau et al., 1998), progesterone receptor (PR; Williams and Sigler, 1998), and peroxisome proliferator-activated receptor (PPAR; Xu et al., 2002) have been determined (Table 17.1.3).

Importin-α and -β, also known as karyopherin-α and -β, are nuclear import receptors that recognize nuclear localization signals and bind and transport proteins from the cytoplasm to the nucleus through the nuclear pore complex. The structures of importin-α and -β consist of homologous ARM and HEAT repeats (Table 17.1.3; Conti et al., 1998; Chook and Blobel, 1999; Cingolani et al., 1999; Kobe, 1999). The structural fold of ARM and HEAT repeats are reviewed in the section below (see

Modular Domains Involved in Signal Transduction).

Integrins and adhesion-related proteins

The structures of the I-domain from CD11b and CD11a integrin α-chains show that these domains possess a flavodoxin-like fold similar to members of the dinucleotide-binding family of proteins (Lee et al., 1995; Qu and Leahy, 1995). The tertiary structure consists of a six-stranded mostly parallel β-sheet with only one antiparallel strand on one edge of the sheet (Fig. 17.1.7D; Table 17.1.3). The connectivity of the strands follows a 3-2-1-4-5-6 order. Flanking the central β-sheet are several α-helices on each side. A bound divalent metal ion at the end of the β-sheet is postulated to play an important role in its activity. The structure of an intact αVβ3 integrin exhibits a multidomain organization of the molecule (Xiong et al., 2001). The adhesion domain consists of a β-propeller domain from the αV-chain and the I-domain from the β3-chain. Following the adhesion domain, the αV-chain has three immunoglobulin-like domains, whereas the β3-chain possess four EGF-like modules.

β-catenins are cytosolic proteins essential for cell adhesion and signal transduction. The structure of β-catenin consists of a central domain of twelve armadillo (ARM) repeats (see Modular Domains Involved in Signal Transduction), which mediate binding to cadherins and Tcf family transcription factors (Graham et al., 2000; Huber and Weis, 2001).

Scavenger receptor cysteine-rich domain

The cysteine-rich domain of the scavenger receptor is a conserved domain of 110 amino acids. It is found widely in cell surface receptors and secreted proteins that participate in innate immune functions. The structure of a scavenger receptor cysteine-rich domain (SRCR) from the Mac-2 binding protein has been determined (Hohenester et al., 1999). The fold consists of a curved six-stranded β-sheet cradling an α-helix. There are three conserved disulfide bonds (Fig. 17.1.7E; Table 17.1.3): one bridging between strand C and the EF loop (Cys 31-Cys 95), one between the α-helix and strand F (Cys 44-Cys 105), and one within the EF loop (Cys 75-Cys 85).

Glutamate receptor

Glutamate is an important neurotransmitter of mammalian central nervous systems mediating long-term potentiation/depression, learning, and memory. There are two forms of glutamate receptors, the metabotropic (mGluR) and ionotropic (iGluR) glutamate receptors. While metabotropic glutamate receptors are members of G-protein-coupled seven transmembrane receptors, ionotropic glutamate receptors contain ligand-gated channels. The structures of the extracellular ligand-binding domains of mGluR and iGluR are similar (Armstrong et al., 1998; Kunishima et al., 2000), with two domains each assuming an α/β fold topology (Fig. 17.1.7F; Table 17.1.3). Glutamate is bound in an interdomain crevice.

Transferrin receptor

Transferrin receptors (TfR) facilitate iron uptake in vertebrates by binding to iron-loaded transferrins on the cell surface, and then internalizing and releasing the iron in an endosome. The TfR-bound apo-transferrin then recycles to the cell surface where it is exchanged with iron-bound transferrin. The crystal structure of a human TfR shows a receptor dimer with each monomer consisting of three domains (Fig. 17.1.7G; Lawrence et al., 1999). The first protease-like domain has a fold similar to that of the carboxy- or aminopeptidases with a central seven-stranded β-sheet flanked by α-helices. The second apical domain folds into a β-sandwich similar to domain four in aconitase. The third domain is all α-helical and forms an irregular four-helical bundle (Table 17.1.3).

Modular Domains Involved in Signal Transduction

Domains that bind phosphotyrosine

SH2 domains. SH2 domains were initially identified in the oncoprotein v-Fps, and subsequently in the Src and Abl tyrosine kinases (Pawson et al., 2001). The name is derived from the designation Src homology domain 2. The function of this module is to bind phosphotyrosine peptides (Table 17.1.4). Through this very specific phosphorylation-dependent interaction, various types of cellular signaling processes can be controlled, such as enzyme regulation, sequestering or recruiting of specific proteins, linking proteins together through adaptor molecules, and protein multimerization. Most SH2 domains are involved in some aspect of the cytoplasmic tyrosine kinase signaling pathways (for reviews of the SH2 domain, see Kuriyan and Cowburn, 1997; Pawson et al., 2001). Examples of proteins with SH2 domains include protein tyrosine kinases (PTKs), phosphatases (PTPases), a subunit of phospholipase C (PLC-γ1), the p85 subunit of

phosphatidylinositol-3OH kinase, the cytoplasmic tail of cell surface receptor proteins, and adaptor proteins like Grb2; however, as of this writing, the total number of SH2-containing proteins is well over 1100 according to the SMART database (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*).

The fold of the ~100 residue SH2 domain is characterized by an α/β barrel-like sandwich with a central core formed by a four-stranded antiparallel β-sheet followed by a smaller three-stranded antiparallel β-sheet (Fig. 17.1.8A). The β-barrel is flanked by two α-helices, one on each side of the larger β-sheet (Booker et al., 1992; Overduin et al., 1992; Waksman et al., 1992). A phosphopeptide-binding site is located on a surface formed by the fourth strand of the central β-sheet, the first helix, the loop between strands 5 and 6, and the loop at the end of helix 2. The phosphate moiety of the phosphotyrosine forms an ion pair with an arginine residue from the second strand of the central β-sheet. Another arginine from the first helix and a lysine from strand 4 appear to have favorable interaction with the phosphotyrosine ring. The peptide binding specificity of SH2 domains are defined largely by amino acids at the +1, +2, and +3 site of the phosphotyrosine peptide.

PTBs. Phosphotyrosine-binding domains (PTBs), also known as protein-interaction domains (PIs), function as signaling domains which bind to proteins containing the consensus sequence Asn-Pro-X-(pTyr/Tyr). For reviews, see Yaffe (2002b) and Yan et al. (2002). Most PTBs target sequences with a phosphorylated tyrosine; however, some recognize non-phosphorylated sequences. Furthermore, some PTB domains are reported to recognize peptide ligands containing no tyrosine residue or even phospholipid ligands (Yaffe, 2002b; Yan et al., 2002). Despite functional diversity within this family, structural homology is strongly conserved and many members display very high ligand specificity.

The PTB fold consists of a seven-stranded antiparallel β-sandwich flanked by an α-helix (Fig. 17.1.8B; Zhou et al., 1995, 1996; Eck et al., 1996). Remarkably, this is the same fold observed for the pleckstrin homology (PH) domain and the EVH1 domain (see Polyproline Binding Domains). In the case of the PTB domain, the peptide ligand generally forms an additional β-strand between β-strand 5 and the α-helix, essentially extending the second β-sheet (Yan et al., 2002). First identified in the adaptor protein SCH (Blaikie et al., 1994; Kavanaugh and Williams, 1994; van der Geer et al., 1995) and the insulin receptor substrate 1 (IRS-1; Gustafson et al., 1995), over 430 PTB domains are now listed in the SMART database (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*).

Phosphoserine- and phosphothreonine-binding domains

Forkhead-associated domains (FHA). The forkhead-associated domain (FHA) is a small modular phosphothreonine-binding domain widely found, from prokaryotes to eukaryotes, in numerous types of multidomain proteins, including transcription factors, kinases, phosphatases, kinesins, RNA-binding proteins, and some metabolic enzymes (Li et al., 2000; Durocher and Jackson, 2002). Initially identified by Hofmann and Bucher (1995) in the family of forkhead-associated transcription factors, this domain has since been found on over 400 different proteins (*http://smart.embl-heidelberg.de/browse.shtml*). In addition to recognizing phosphothreonine, some FHA domains are reported to bind phosphotyrosine (Liao et al., 1999; Wang et al., 2000a) or even unphosphorylated peptides (Li et al., 2002a). The structure of the FHA domain (Fig. 17.1.8C) comprises a ~100 residue eleven-stranded β-sandwich (Liao et al., 1999; Durocher et al., 2000; Wang et al., 2000a). Some FHA domains have small helices within the loop regions as well. Structurally, this domain is most closely related to the MH2 domain of the SMAD family that also functions as a protein interaction module. Interestingly, the FHA fold includes nearly twice as many residues as the 55– to 75–amino acid FHA homology region initially studied. This FHA homology region covers only β-strands 3 to 10 on one face and includes the ligand recognition site (Durocher and Jackson, 2002). The phosphothreonine-containing peptide ligand binds in an extended conformation across three loops (β3/4, β4/5 and β6/7) along the end of one face of the β-sandwich.

14-3-3 proteins. 14-3-3 proteins constitute a large array of multifunctional phosphoserine/phosphothreonine-binding proteins that play significant roles in signal transduction for numerous eukaryotic-cell processes, including regulation of gene expression, cell-growth control, apoptosis, and neuronal development (Fu et al., 2000b; Yaffe, 2002a). Many different types of proteins are targeted by 14-3-3 proteins, including kinases, acetylases, acetyl transferases, hydroxylases, cell surface recep-

tors, cytoskeletal and scaffolding proteins, small G-proteins, and transcription factors. Indeed over 100 different proteins have been reported to interact with 14-3-3 proteins (Yaffe, 2002a). The name 14-3-3 refers to the elution position of these proteins from a DEAE cellulose column and the subsequent gel electrophoresis migration pattern during isolation from bovine brain by Moore and Perez in 1967.

14-3-3 proteins are relatively abundant, acidic proteins found in eukaryotic organisms from fungi to plants and mammals. In mammals, seven different isoforms have been characterized, referred to as β, γ, ε, η, σ, τ, and ζ. They function as homo- and heterodimers with a monomeric molecular weight of ~30 kDa. The first three-dimensional structures of 14-3-3 determined in 1995 revealed a crescent shaped dimer (Fig. 17.1.8D) composed of a bundle of antiparallel α-helices (nine per monomer; Liu et al., 1995; Xiao et al., 1995). The inner concave surface (or central channel) is more negatively charged and more highly conserved than the outer surface. Subsequent crystal structures of peptide-ligand complexes revealed an amphipathic binding groove in each monomer within the central channel that binds peptides in an extended conformation (Yaffe, 2002a). The recent crystal structure of 14-3-3 ζ complexed with serotonin *N*-acetyltransferase bound to a bisubstrate analog in addition to isothermal titration calorimetry (ITC) and activity studies provide intriguing evidence that 14-3-3 binding structurally modulates serotonin *N*-acetyltransferase such that substrate binding affinity and activity increase (Obsil et al., 2001). Thus, 14-3-3 proteins are not simply protein interaction domains, but can also regulate enzymatic activity.

Polyproline-binding domains

SH3 domains. SH3 (src homology 3) domains are protein interaction modules of 50 to 75 residues that recognize proline-rich sequences. Like SH2 domains, SH3 domains are perhaps best known for their role in the regulation of tyrosine kinases; however, they are also found in numerous other proteins, including the adaptor proteins Grb2 and Crk, as well as PTPases such as the γl subunit of phopholipase C (PLC-γl), certain cytoskeletal proteins such as myosin-1B and fodrin, and the yeast actin-binding protein ABP-1. The SMART database now lists over 2200 proteins containing SH3 domains (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*).

The core of an SH3 domain consists of two perpendicular antiparallel β-sheets folded into a β-barrel like structure. One sheet is formed with strands B, C, and D, the other with A and E (Fig. 17.1.9A; Musacchio et al., 1992; Yu et al., 1992). The relatively flat, nonpolar recognition site is formed primarily by two loops of variable length, one between strand A and B (the so-called RT loop), and one between B and C (the so-called n-Src loop). The core binding motif is Φ-Pro-X-Φ-Pro, where Φ is a hydrophobic residue and x is any residue. Each of the ΦP dipeptides occupies a conserved hydrophobic binding pocket. An interesting feature of the SH3 domain is that it recognizes its ligands in either of two orientations, depending on the placement of positively charged specificity residues flanking the Φ-Pro-X-Φ-Pro sequence. For reviews detailing the structural and functional aspects of SH3 domains see Cesareni et al. (2002), Kuriyan and Cowburn (1997), and Mayer (2001).

WW domains. WW domains (also known as WWP or rsp5 domains) are very small modules consisting of only 35 to 40 residues folded into a three-stranded antiparallel β-sheet (Fig. 17.1.9B; Macias et al., 1996). As its name suggests, this domain contains a pair of signature tryptophan residues spaced ~20 to 22 residues apart. Additionally, there is a highly conserved proline three residues beyond the second tryptophan. WW domains have been observed throughout eukaryotes, including fungi, plants, nematodes, and various vertebrates. They function specifically in binding proline-rich sequences and have been classified into four groups based on ligand specificity. Group I domains bind to PPxY-containing sequences, group II bind PPLP sequences, group III bind polyproline sequences flanked by Arg or Lys, and group IV domains bind to ligands containing phospho-(S/T)P. Structural solutions of several ligand complexes show that the polyproline ligand interacts with the convex face of the β-sheet. For reviews, see Macias et al. (2002) and Sudol and Hunter (2000).

EVH1 domains. Enabled/vasodilator-stimulated phosphoprotein homology 1 (EVH1) domains (also known as WH1 domains) are protein-interaction modules critical for a diverse array of signal transduction pathways related to actin cytoskeleton remodeling (Ball et al., 2002). EVH1 domains of ~115 residues are widely present in eukaryotic species from fungi to mammals and are generally part of large multidomain proteins. The EVH1 fold comprises a small seven-stranded antiparallel β-

sandwich flanked by a C-terminal α-helix (Fig. 17.1.9C; Fedorov et al., 1999; Prehoda et al., 1999). Interestingly, this fold closely resembles both the pleckstrin homology (PH) and phosphotyrosine binding (PTB) domains. EVH1 domains function by recognizing and targeting ligands containing the polyproline consensus sequence Phe-Pro-Pro-Pro-Pro; however, based on ligand specificity, EVH1 domains can be divided into class I proteins that recognize the sequence Phe-Pro-X-Φ-Pro, (X represents any residue and Φ represents a hydrophobic residue) and class II proteins that recognize the sequence Pro-Pro-X-X-Phe (Ball et al., 2002). The peptide ligand binds across the concave surface formed by β-strands 1, 5, 6, and 7. PH and PTB domains also bind their ligands on this same face of the β-sandwich, albeit in different regions (Fedorov et al., 1999).

GYF domains. The GYF domain was first identified in CD2BP2, a signaling protein that binds to a proline-rich region of the cytoplasmic tail of the T cell adhesion protein CD2 (Nishizawa et al., 1998). The SMART database now lists 54 occurrences of GYF domains including sequences from fungi, plants, and metazoans (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*) . The NMR solution structure of the C-terminal 62 residues of *Drosophila* CD2BP2 reveals a small four-stranded antiparallel β-sheet adjacent to a small α-helix (Fig. 17.1.9D; Freund et al., 1999). NMR and mutational data suggest that the binding site is centered around the highly conserved triplet Gly-Tyr-Phe situated in the loop between the α-helix and the third β-strand (Freund et al., 1999). Hence, this domain has been named the GYF domain.

Phospholipid-binding domains

PH domains. Pleckstrin homology (PH) domains are modules of ~100 amino acids that are present in a wide range of signaling proteins, including many kinases, isoforms of phospholipase C, GTPases, GTPase-activating proteins, and nucleotide-exchange factors. The majority of characterized PH domains recognize and bind phosphoinositides or inositol phosphates. Indeed, most PH domains are part of membrane-associated proteins and it has been suggested that PH domains function as membrane adaptors, targeting proteins to specific membranes (Hurley and Misra, 2000; Maffucci and Falasca, 2001); however, recent evidence suggests that some PH domains, such as those of the β-adrenergic receptor kinase (β-ARK), the TecII tyrosine kinase, insulin receptor substrate 1 (IRS-1), and the Etk tyrosine kinase, may bind protein ligands (see Maffucci and Falasca, 2001, and references therein).

Each PH domain folds into a seven-stranded antiparallel β-sandwich flanked by a C-terminal long α-helix at one edge of the sandwich (Downing et al., 1994; Ferguson et al., 1994; Macias et al., 1994; Yoon et al., 1994). Strands A, B, C, and D form one sheet of the β-sandwich while strands E, F, and G form the other (Fig. 17.1.10A). The identified IP_3 binding site is located at the two loop regions, loop AB (the loop between strand A and B) and CD (the loop between strand C and D), opposite of the long helix. The basic PH fold has also been observed in PTB and EVH1 domains and the Pan-binding domain, despite the fact that these domains have no sequence homology with the PH domain. This has led to its reference by some as the PH superfold (Blomberg et al., 1999).

C1 domains. The C1 domain is a cysteine-rich compact structure of ~50 residues. It was first identified as a conserved region of protein kinase C (PKC) isozymes that binds diacylglycerol (DAG) and phorbol esters and participates in allosteric regulation. Since then, several hundred C1 domains have been found in other proteins. Not all of these bind DAG but may instead bind proteins or other lipids. For reviews, see Cho (2001) and Hurley and Misra (2000).

Structurally, the C1 domain (Fig. 17.1.10B) comprises a two-stranded β-sheet, a three-stranded β-sheet, a small α-helix and two bound Zn^{2+} ions (Hommel et al., 1994; Zhang et al., 1995). Each Zn^{2+} ion is coordinated by three Cys and one His side chain. The DAG/phorbol ester binding site is situated at one end of the molecule in a region surrounded by hydrophobic residues. A belt of basic residues abuts this region, encircling the middle part of the molecule.

C2 domains. C2 domains were initially identified as a Ca^{2+}-binding conserved homology region of protein kinase C (PKC) isozymes. To date, however, C2 domains have been observed in over 1600 different proteins according to the SMART database (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*), most of which function in signal transduction or membrane trafficking (see Cho, 2001, and Hurley and Misra, 2000 for reviews). The majority bind membrane phospholipids in a Ca^{2+}-dependent manner; however, not all require Ca^{2+} and some domains bind proteins as well. The

structure of C2 domains consists of ~130 residues folded into an eight-stranded β-sandwich with topologies resembling an Ig fold (Fig.17.1.10C and Fig. 17.1.2, respectively; Sutton et al., 1995; Grobler and Hurley, 1997). Although the β-strands are consistently in the same location, two different circularly permuted C2 domain topologies are observed. The first β-strand (strand a) in synaptotagmin-like C2 domains becomes the last β-strand in the phospholipase C–like C2 domains. The rest of the topology is identical between various C2 domains. Three variable loops at one end of the domain, referred to as Ca^{2+}-binding regions (CBRs), are intimately involved in the binding of two to three Ca^{2+} ions and in phospholipid binding. Interestingly, the CBRs also correspond topologically to the complementarity-determining regions (CDRs) of the variable domains of antibodies. Structure-function studies of C2 domains from PKCα and cytosolic phospholipase A_2 ($cPLA_2$) have shown that the two bound Ca^{2+} ions in each C2 domain play distinct roles: one provides a bridge between the C2 domain and anionic phospholipids, whereas the second Ca^{2+} induces a conformational change that enhances membrane-protein interactions (Cho, 2001; Stahelin and Cho, 2001).

FYVE domains. FYVE fingers are modular domains of ~60 residues that specifically bind to phosphatidylinositol 3-phosphate (PI3P). They are observed in numerous eukaryotic proteins from yeast to human, and function primarily in the regulation of endocytic membrane trafficking (for a review, see Stenmark et al., 2002). The name originates from the first four FYVE finger-containing proteins identified: Fab1p, YOTB, Vac1p, and EEA1. The three-dimensional structure of the FYVE finger consists of two double-stranded antiparallel β-sheets, a C-terminal helix and two long loops (Fig. 17.1.10D; Misra and Hurley, 1999; Mao et al., 2000). The core structure is further stabilized by two Zn^{2+} ions that are coordinated by four conserved Cys-X-X-Cys motifs. A 2.2-Å crystal structure of the homodimeric EEA1 FYVE finger bound to inositol 1,3-bisphosphate (a soluble analogue of PI3P) reveals the probable PI3P binding site, which includes residues from the following conserved FYVE signature motifs: the N-terminal Trp-X-X-Asp, the C-terminal RCV, and the Arg-(Arg/Lys)-His-His-Cys-Arg motif (Dumas et al., 2001). A lower resolution NMR structure of EEA1 bound to di-C_4-PI3P shows a similar binding site, but with different atomic contacts (Kutateladze and Overduin, 2001). Furthermore, a hydrophobic N-terminal loop (referred to as the turret loop) adjacent to the PI3P binding site is proposed to partially penetrate and interact with the membrane surface.

Protein interaction domains

PDZ domains. PDZ domains are compact globular domains of ~90 residues that serve as protein-protein interaction modules in the assembly of large complexes. For reviews, see Hung and Sheng (2002) and Ponting et al. (1997). These relatively abundant domains are widely observed in bacteria, plants, and higher eukaryotes, including humans. Several of these domains (up to 13) are often found within one protein. Formerly known as Disc-large homology regions (DHRs) or GLGF repeats, PDZ modules derived their current nomenclature from the first three PDZ-containing proteins identified: PSD-95/SAP90, Disc-large, and ZO-1.

The first PDZ structures determined were the third PDZ of rat PSD-95 (Doyle et al., 1996) and the third PDZ domain of human *Dlg* (Morais Cabral et al., 1996). Structurally, the domain comprises a five- or six-stranded β-sandwich and two α-helices that specifically bind the C-terminal peptide of certain proteins (Fig. 17.1.11A). The peptide ligand binds as an antiparallel β-strand between the β-strand and the αB-helix with up to nanomolar affinity. PDZ domains have also been reported to interact with internal peptide sequences in addition to C-terminal peptides, as observed in the crystal structure of nNOS and syntropin (Hillier et al., 1999). Three classes of PDZ domains have been loosely defined based on their specificity for the last four residues of the bound peptide: class I binds the sequence -X-Ser/Thr-X-Φ, class II binds -X-Φ-X-Φ, and class III binds -X-Asp/Glu-X-Φ, where X represents any amino acid and Φ represents a hydrophobic residue. However, a more comprehensive and complicated classification has also been proposed based on the identity of two residues within the PDZ fold (Bezprozvanny and Maximov, 2001).

VHS domains. VHS domains comprise ~140 residues and reside in the N-termini of certain proteins that play key roles in vesicular trafficking and protein sorting. The name VHS originates from Vps27, Hrs, and STAM, three proteins that harbor this domain; however, to date, well over 100 proteins are known to contain VHS domains. These can be divided into four categories based on their relative placement

with other modular domains (Lohi et al., 2002). Group 1 consists of the STAM/EAST/Hbp family, each of which have the VHS-SH3-ITAM domain organization (see above for SH3). Group 2 proteins have a VHS-FYVE arrangement which includes the Vps27 and Hrs proteins (see above for FYVE). Group 3 contain Golgi-localizing, γ-adaptin ear homology domain, ARF-interacting (GGA) proteins with a VHS-GAT-ear arrangement. Finally, group 4 are VHS-containing proteins that do not fit into groups 1 to 3.

Crystal structures of the VHS domains of *Drosophila* Hrs and human TOM1 (Mao et al., 2000; Misra et al., 2000) revealed a right-handed superhelix of eight α-helices (Fig. 17.1.11B) with striking structural homology to the eight-helix ENTH domain of epsin-1. The first seven helices of the VHS and ENTH domains superimpose with rms deviations of ~1.8 Å (Lohi et al., 2002). Furthermore, the VHS domain displays structural similarities to HEAT and ARM repeats. More recently, crystal structures of the VHS domains of GGA1 and GGA3 complexed with target peptides showed that signal peptides containing an acidic-cluster dileucine (ACLL) motif, Asn-X-X-Leu-Leu, bind to a groove between α-helices 6 and 8 (Fig. 17.1.11B; Misra et al., 2002; Shiba et al., 2002).

SNAREs. Soluble, *n*-ethylmaleimide-sensitive factor attachment-protein receptors (SNAREs) comprise a superfamily of conserved eukaryotic proteins that function in membrane fusion. They are characterized by ~60 residue SNARE domains consisting of a long amphipathic α-helix. Most SNAREs are type II transmembrane proteins with an N-terminal cytoplasmic SNARE domain and a very short extracytoplasmic region; however, some are fastened to the membrane by palmitoylation or prenylation rather than by a transmembrane domain (Hay, 2001). Upon complex formation, SNARE domains anchored to opposite membranes come together to form a helical bundle that promotes membrane fusion. For reviews, see Hay (2001) and Misura et al. (2000a).

First observed to atomic resolution in the crystal structure of the syntaxin-1A/synaptobrevin-II/SNAP-25B complex (Sutton et al., 1998), this complex comprises a structurally conserved coiled-coil of four parallel α-helices (Fig. 17.1.11C). Superficially, it resembles the tetrameric GCN4 four-helix-bundle with a largely hydrophobic core; however, deviations in the heptad repeat pattern and composition, as well as local deviations in helical curvature, set this complex apart from known leucine-zipper or coiled-coil proteins.

SNARE complexes consist of a single SNARE from one membrane termed an R-SNARE and three Q-SNARES associated with the opposing membrane. In some cases, two Q-SNARES are on the same polypeptide (e.g., SNAP-25) while in other cases all three Q-SNARES are on separate molecules. The "R" and "Q" designations are based on the presence of conserved Arg or Gln residues in the center (or 0 layer) of each respective SNARE domain. The Arg and three Gln residues interact with each other at the core of the complex to form an intriguing hydrogen bond network completely isolated from solvent. The Q-SNAREs have been further classified by protein profiling (Bock et al., 2001). Group Qa are homologous to syntaxins, group Qb are homologous to the N-terminal helix of SNAP-25 and the group Qc SNAREs are homologous to the C-terminal helix of SNAP-25. Furthermore, the R-, Qa-, Qb-, and Qc-SNAREs each appear to occupy conserved positions within the complex.

In isolation, SNARE domains are unstructured and are referred to as existing in an open conformation that favors SNARE complex formation; however, syntaxin SNAREs have an Habc domain N-terminal to the SNARE domain that has an antiparallel triple-helix structure (Fernandez et al., 1998). The Habc domain can complex with the SNARE domain (referred to as H3 in this context), induce helical structure, and form a four-helix bundle (Munson et al., 2000). This is referred to as a closed SNARE conformation and inhibits complex formation with other SNAREs. Syntaxin appears to exist in equilibrium between open and closed forms. Regulatory proteins such as nSec1 can perturb this equilibrium by binding to the closed conformation, thereby stabilizing it and preventing SNARE complex formation. The crystal structure of yeast syntaxin Sso1p (H3Habc; Munson et al., 2000) shows the H3 helix is very similar in structure to the comparative helix in a SNARE complex, although its orientations with neighboring helices is strikingly different. Curiously, the crystal structure of the nSec1/syntaxin 1a complex (Misura et al., 2000b) reveals the H3 domain adopting an irregular bent helical structure distinct from that of the SNARE complex. Thus, the H3 region of syntaxin can have at least three different conformations depending on its local environment or binding partners.

SAM domains. The sterile alpha motif (SAM) is a modular protein interaction domain present in a wide variety of signaling proteins, including receptor tyrosine kinases, Ser/Thr kinases, diacylglycerol kinases, transcription factors, scaffolding and adaptor proteins, polyhomeotic proteins, and GTPases. The SMART database (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*) lists over 700 proteins containing SAM domains. SAM domains are ~70 residues in size and generally reside in the N- or C-terminal ends of proteins. The basic fold consists of an N-terminal extended region followed by four short α-helices and one long C-terminal α-helix (Fig. 17.1.11D). Several modes of SAM domain interaction have been observed crystallographically, including dimer formation as well as polymerization (Smalla et al., 1999; Stapleton et al., 1999; Thanos et al., 1999; Kim et al., 2002). Furthermore, there are reports of SAM domains binding other non-SAM containing proteins.

MH2 domains (C-terminal domain of Smad proteins). Smad proteins are an integral part of the receptor signal transduction apparatus for the transforming growth factor-β (TGF-β) family of cytokines such as TGF-β, bone morphogenetic protein (BMP), and activins. Smads are divided into three classes: comediator Smads (Co-Smads), receptor regulated Smads (R-Smads), and inhibitory Smads (I-Smads). Both the N- and C-terminal domains of Smads are conserved across species and connected by a variable linker region (L-domain). The N-terminal domain (MH1) interacts with DNA, whereas the C-terminal MH2 domain is involved in phosphorylation-regulated homo- and hetero-oligomerization among Smads as well as interaction with other proteins in the signaling pathway. The fold of the MH2 domain comprises a twisted antiparallel β sandwich made from five- and six-stranded β-sheets capped at one end by a three-helix bundle pointing away from the sandwich, and capped at the other end by a three large loops and a helix (Fig. 17.1.11E; Shi et al., 1997). Signal-dependent phosphorylation of two serines at the C-terminal Ser-Ser-X-Ser sequence of R-Smads is believed to trigger the formation of a heterotrimer with a Co-Smad. Homotrimeric crystal structures of R-Smads 1, 2, and 4 MH2 domains (psuedophosphorylated, phosphorylated, and unphosphorylated, respectively) have provided models for this interaction and indicate a phosphoserine binding site at the L3 loop. (Fig. 17.1.11F; Shi et al., 1997; Qin et al., 2001; Wu et al., 2001). There are also crystal structures of complexes of the MH2 domains of both Smad2 and Smad3 with extended structures of SARA (Smad anchor for receptor activation; Fig 17.1.11G; Wu et al., 2000; Qin et al., 2002).

Structural repeat motifs

In addition to the numerous modular domains that form independently folded functional units, a number of much smaller structural motifs of 20 to 40 residues are observed that exist as tandem arrays of 5 to 25 repeats (Groves and Barford, 1999). In proteins containing these arrays, the functional unit is a set of repeats rather than the individual motifs. The stacking of adjacent motifs often results in curving superhelical structures, many of which function in the context of macromolecular binding interactions. Four of the more well known of these structural repeat motifs are discussed below.

Leucine-rich repeats. The leucine-rich repeat (LRR) was first identified in 1985 in the sequence of α2-glycoprotein (Takahashi et al., 1985). Now, however, over 2900 proteins are reported to contain LRRs according to the SMART database (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*). Most of these are eukaryotic proteins, but >100 are in prokaryotic sequences. Although these proteins have diverse functions, a common element that most LRRs share is mediating protein-protein interaction. LRRs comprise repeats of 20 to 30 amino acids with the following eleven-residue repeat consensus sequence Leu-X-X-Leu-X-Leu-X-X-(Asn/Cys)-X-Leu, where X represents any residue and some L positions can be occupied by Val, Ile, or Phe (Kajava, 1998; Kobe and Kajava, 2001). The three-dimensional structure of the repeat consists of an α-helix followed by a single β-strand (Kobe and Deisenhofer, 1993). Subsequent repeats stack directly on top of each other such that the β-strands form a curved parallel β-sheet and the helices line up parallel to a common axis (Fig. 17.1.12A). LRR structures are generally crescent shaped with the β-sheet along the inner concave surface and the helices on the outside convex surface with very little twist along the central axis. So far, the β-sheet region and adjacent loops of LRRs appear to be the most frequent areas for protein-protein interaction (Kobe and Kajava, 2001).

HEAT repeats. The HEAT repeat derives its name from four of the first proteins identified to contain it: Huntingtin, elongation factor 3, the PR65/A subunit of protein phosphatase 2A,

and lipid kinase TOR (Andrade and Bork, 1995). Andrade and coworkers (2001) reported the detection of 71 HEAT repeat-containing eukaryotic proteins in the Swissprot database. HEAT repeats are 37 to 43 residues long and consist of two antiparallel α-helices designated A and B. Between 3 and 22 repeats stack directly on top of each other and thus form two layers of parallel helices with an overall superhelical curvature (Fig. 17.1.12B; Groves et al., 1999). B helices line the concave surface and A helices the convex surface. The crystal structures of importin β1 and β2 bound with different protein ligands each reveal protein recognition sites among the inner B helices (Chook and Blobel, 1999; Cingolani et al., 1999; Vetter et al., 1999). For recent reviews about HEAT repeats, see Andrade et al. (2001), Groves and Barford, (1999), and Kobe et al. (1999).

ARM repeats. ARM repeats are named after the armadillo protein, the *Drosophila* segment polarity gene product in which they were first identified. These repeats are very similar to HEAT repeats and are ~40 residues in length; however, instead of the two helices observed in each HEAT repeat, ARM repeats consist of three helices, designated H1, H2, and H3 (Groves and Barford, 1999; Andrade et al., 2001). The SMART database reports nearly 400 proteins containing ARM repeats (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml).* The crystal structures of β-catenin and karyopherin α (importin α) revealed stacking of repeats to form three ladders of parallel helices arranged in a superhelical structure (Fig. 17.1.12C; Conti et al., 1998; Huber et al., 1997). The superhelical twist is formed by an ~30° rotation of each repeat from its neighbors. Within individual repeats, the antiparallel helices H2 and H3 have orientations similar to helices A and B of HEAT repeats. Helix H1, however, is smaller and lies perpendicular to the H2/H3 pair. H3 and its flanking loops are observed to form a groove lined with conserved residues along the entire length of the ARM repeat structure. The structure of β-catenin complexed with the cytoplasmic domain of E-cadherin provides a striking example of how this groove can be used for protein recognition (Huber and Weis, 2001). In the complex, the C-terminal 100 residues of cadherin bind in a predominantly extended conformation along the entire length of the H3 groove of β-catenin. A variant of the ARM motif is also observed to bind to specific RNA sequences in the Pumilio family of mRNA translation regulators (Wang et al., 2002).

Ankyrin repeats. The ~33-residue ankyrin (ANK) repeat is found in numerous proteins with varied functions, such as cyclin-dependent protein kinase inhibitors, the notch membrane receptor, the inhibitory subunit (IκBα) of nuclear factor κB (NF-κB), the transcriptional regulator GABPβ, the p53-binding protein 53BP2, certain cytoskeletal organizing proteins, and toxins (Groves and Barford, 1999; Sedgwick and Smerdon, 1999). The SMART database reports over 2700 proteins containing ankyrin repeats (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*). The ankyrin repeat receives its name from the cytoskeletal protein ankyrin in which 24 repeats were found (Lux et al., 1990); however, it was first identified three years earlier in several cell cycle and developmental regulators (Breeden and Nasmyth, 1987). Amidst the diverse assortment of proteins containing ankyrin repeats, the common functional theme that has developed is that of protein-protein interaction.

The ankyrin fold was first observed in the crystal structure of 53BP2 bound to the p53 tumor suppressor (Gorina and Pavletich, 1996). The structural repeating unit consists of a β-hairpin loop followed by a pair of antiparallel α-helices that are roughly perpendicular to the β-hairpin (Fig. 17.1.12D). However, the repeating amino acid sequence actually begins with the second β-strand in one unit and ends with the first β-strand in the next. Stacking of repeats results in a curving structure with a left-handed twist, consisting of two rows of parallel helices and a narrow extended β-sheet. Ankyrin repeat proteins recognize their target proteins in varied ways, but a strikingly common theme involves interaction through the ends of the β-hairpin fingers.

Protein Kinase Fold

Eukaryotic protein kinases (PK) comprise an enormous superfamily of homologous proteins that catalyze the transfer of the γ phosphate of ATP to specific hydroxyamino acids on a protein substrate. These proteins play numerous roles directly regulating and participating in a diverse array of signal transduction cascades. PKs can be broadly categorized as tyrosine kinases and serine/threonine kinases; however, a comprehensive classification system was published in 1995 based primarily on sequence homology (Hanks and Hunter, 1995). This classification has recently been expanded

to include a total of seven major groups of PK families (Manning et al., 2002a, 2002b). Tyrosine kinases (TK) comprise one group and include both membrane-spanning receptors, such as the insulin receptor, and cytosolic proteins, such as Src kinases. The Ser/Thr kinases comprise six main groups. The AGC group includes cyclic nucleotide-regulated PKs and diacylglyceride-regulated/phospholipid-depe ndent PKs (PKC), among others. The CaMK group consists of calcium/calmodulin-dependent PKs. The CMGC group includes cyclin-dependent PKs, mitogen activated protein (MAP) kinases, glycogen synthase kinase-3, and casein kinase-II. The STE grouping includes several MAPK cascade families. The CK1 group comprises CK1, tau tubulin kinase (TTBK), and vaccinia related kinase (VRK) families. The sixth group of serine/threonine kinases comprises tyrosine kinase-like (TKL) families. These include the mixed lineage kinase (MLK), interleukin-1 receptor-associated kinase (IRAK), LIMK/TESK, Raf, receptor interacting protein kinase (RIPK), and the activin and TGF-β receptor kinases. At least 50 proteins have been identified that have sequence similarity with eukaryotic protein kinases and yet are either reported or predicted to lack catalytic activity (Manning et al., 2002b). Finally, more than a dozen atypical protein kinases (aPKs) have been identified that are reported to have protein kinase activity, but have no sequence homology with the eukaryotic protein kinase superfamily (Manning et al., 2002b).

Protein kinase catalytic domain. Despite the diversity of this superfamily, structural studies have consistently revealed a structurally conserved core fold for the catalytic domain of eukaryotic protein kinases. Other modular domains (e.g., SH2, SH3, PH, SAM, C2, and PDZ domains, to name a few) are sometimes fused to the catalytic domain for regulatory or localizing purposes. First elucidated in the crystal structure of cyclic AMP-dependent protein kinase (also known as PKA; Knighton et al., 1991), the bilobal protein kinase fold consists of an N-terminal lobe (N-lobe) comprising a five-stranded antiparallel β-sheet with a flanking α-helix and a larger C-terminal, primarily α-helical lobe (C-lobe) of eight to ten α helices and some short β-strands (Fig. 17.1.13A). The active site is located between the lobes with the N-lobe contributing an ATP-binding glycine-rich loop and the C-lobe providing most of the protein-substrate-binding site. An activation loop in the C-lobe contains one or more phosphorylation sites that help activate the enzyme when phosphorylated. Although the topology and core catalytic domain structures are well conserved across this large superfamily, there are some noticeable differences among various members. These include additions and deletions of secondary structure elements, changes in the orientation of secondary structure elements, N- and C-terminal extensions, and variations in the relative orientation between the N- and C-lobes. Some of these structural differences appear to correspond to the sequence-based classifications of PKs discussed above. Noteworthy variations include a 25- to 30-residue insertion between helices G and H (PKA nomenclature) in the CMGC protein kinases (e.g., cyclin-dependent and MAP kinases) and the extra helix C observed in the cAMP-dependent protein kinases. Two atypical protein kinases, the transient receptor potential (TRP) Ca^{2+}-channel kinase (also referred to as ChaK or TRP-PLIK) domain (Yamaguchi et al., 2001) and the actin-fragmin kinase (Steinbacher et al., 1999) have been found to share significant structural homology with the eukaryotic PK superfamily members despite a lack of sequence similarity; however, the structurally homologous region in these enzymes is smaller, comprising most of the N-lobe but only the N-terminal portion—i.e., helices D and E (PKA nomenclature) and the following β-strands—of the C-lobe. The rest of the C-lobe consists largely of α-helices that do not correspond to those in the PK superfamily. Furthermore, actin-fragmin kinase has an extra two-helix insertion in the N-lobe, and the C-lobe of ChaK has a bound zinc atom that is sequestered from solvent.

Src family kinases. The structure of src tyrosine kinases is briefly discussed here as a special case to illustrate how SH2 and SH3 domains are used in *cis* to regulate and localize protein kinase activity. Most src family tyrosine kinases share a common architecture comprising four Src homology (SH) regions in addition to a variable region. The N-terminal SH4 region contains a myristylation signal required for attachment to the membrane, followed by a variable region that may localize the enzyme more specifically. The subsequent SH3 and SH2 domains bind proline-rich targets and certain phosphotyrosine motifs respectively (for more information, see SH2 Domains and SH3 Domains, respectively). These two domains are involved in both regulation of the enzymatic activity and subcellular localization of the kinase. The catalytic activity of the kinase is found in the C-terminal SH1 region. A short

segment C-terminal to the SH1 domain contains a tyrosine residue that inactivates the kinase when phosphorylated. Similar to other protein kinases, phosphorylation of a tyrosine in the activation loop increases activity of the kinase. The crystal structure of inactive haematopoietic cell kinase (Hck; Sicheri et al., 1997) includes the SH2, SH3, and catalytic domain of the enzyme, while the structures of the catalytic domains alone have been determined for lymphoid cell kinase (Lck) in its activated form (Yamaguchi and Hendrickson, 1996) and C-terminal Src kinase (Csk) in its inactive form (Lamers et al., 1999). The Hck crystal structure shows the SH2 domain interacting with a C-terminal phosphotyrosine and the SH3 domain bound to the linker region between the SH2 and kinase domains, effectively locking the kinase in an inactive state (Fig. 17.1.13B). Despite the fact that both the SH3 and SH2 domains are located distal to the active site of the kinase, their intramolecular interactions in the inactive form results in positioning helix C of the N-lobe such that the orientation of a key residue in the active site is disrupted.

Phosphatases

Protein phosphatases

Eukaryotic protein phosphatases comprise three major families of enzymes, each of which have conserved core folds (Barford et al., 1998). The protein tyrosine phosphatase (PTP) family dephosphorylates phosphotyrosine residues, while the phosphoprotein phosphatase (PPP) and protein phosphatase M/Mg^{2+}-dependent phosphatase (PPM) families consist of enzymes that dephosphorylate phosphorylated serine and threonine residues.

PTPs. Protein tyrosine phosphatases (PTPs) include cytosolic and membrane-spanning, receptor-like, tyrosine-specific phosphatases in addition to dual-specificity phosphatases (DSPs) that can dephosphorylate phosphorylated tyrosine, serine, or threonine. Receptor-like PTPs generally have two tandem PTP domains whereas other PTPs contain one PTP domain.

Structurally, there are three subtypes of PTP catalytic domains, all of which share a three-layer α/β/α sandwich architecture. The low-molecular-weight (subtype I) PTPs are composed of a single domain containing a central four-stranded parallel β-sheet flanked by six helices on both sides (Fig. 17.1.13C). The β-sheet has a strand order of 2-1-3-4. Subtype II PTPs have a central highly curved five- or ten-stranded mixed β-sheet sandwiched between four to seven helices on one side and one or two on the other (Fig. 17.1.13D). Unlike subtype I, the strand order for the central parallel strands of this β-sheet is 1-4-2-3. Subtype II PTPs include the so-called high-molecular-weight PTPs such as PTB1, *Yersinia* PTP, and SHP-1 and -2, in addition to the smaller DSPs such as VHR, MAP kinase phosphatase, kinase-associated phosphatase (KAP), and tumor suppressor PTEN (Table 17.1.5). The high-molecular-weight PTPs have a nine-stranded β-sheet whereas the DSPs have a smaller five-stranded β-sheet composed of the same four central parallel strands and one preceding antiparallel strand, and have an additional one or two extra helices. Although superficially quite similar to other DSP structures, the catalytic domain of CDC25 has a rather different topology and the strand order of its five-stranded parallel β-sheet is 1-5-4-2-3 (Fig. 17.1.13E). Therefore, it may be referred to as a third subtype of protein phosphatase. Interestingly, CDC25 has the same fold as the sulfurtransferases rhodanase, GlpE, and the ERK2-binding domain of MAPK phosphatase-3. Although the topologies of these protein phosphatase subtypes are different, the active site strand-loop-helix regions show a striking structural similarity, and they all feature an essential nucleophilic cysteine residue located in a cleft between the β-sheet and the larger helical bundle.

Protein serine/threonine phosphatases. The PPP and PPM families both comprise metal-dependent serine/threonine protein phosphatases. The fold of the PPP family catalytic domain is a mixed β sandwich flanked by seven α helices on one side, an α/β structure of two to three α helices, and a couple of β-strands on the other side (Fig. 17.1.13F). The active site, including the essential Zn^{2+} and Fe^{3+} cations, is located among the loops at one end of the β sandwich. Enzymes in this family include PP1, calcineurin (PP2B), and the bacteriophage λ protein phosphatase. This fold is also shared by at least three other phosphoryltransfer metalloenzymes that are not protein phosphatases: purple acid phosphatase, the N-terminal domain of UDP-sugar hydrolase, and the N-terminal domain of the DNA double-strand-break repair enzyme MRE11 nuclease (Table 17.1.5).

The PPM family includes both eukaryotic and prokaryotic protein phosphatases and has been defined by protein phosphatase 2C (PP2C). Although PP2C has no sequence ho-

mology with PPP phosphatases, it is also a metalloenzyme with a binuclear metal (in this case Mn^{2+}) center at the active site and has a fold that resembles that of PPP phosphatases. The N-terminal domain consists of an eleven-stranded antiparallel β sandwich surrounded by α helices. As in the PPP family, the active site is located at one end of the β sandwich. The C-terminal domain consists of three antiparallel helices.

Alkaline phosphatases

Alkaline phosphatases (AP) are dimeric enzymes that catalyze the hydrolysis of phosphomonoesters and are common to all organisms. The overall structure of an AP monomer comprises two domains, both of which are α/β/α sandwiches (Fig. 17.1.14A). The smaller domain consists of a three- or four-stranded antiparallel β-sheet sandwiched between two α-helices, one on each side. The larger domain has a central ten-stranded β-sheet in the middle flanked by thirteen helices of various lengths. Except for the eighth strand, this β-sheet is composed of parallel strands. Two other enzymes that share the core AP fold include arylsulfatase and prokaryotic cofactor-independent phosphoglycerate mutase (Table 17.1.5).

Sugar phosphatases

The sugar phosphatase fold is shared among the families of fructose-1,6-bisphosphatase, inositol monophosphatase, and inositol polyphosphate 1-phosphatase. The molecule folds into a five-layered α/β/α/β/α sandwich (Fig. 17.1.14B and Table 17.1.5). The two β-sheets are five and eight-stranded, respectively, and mostly antiparallel. The location of the AMP binding site is at the end of the N-terminal α-helix. This fold also remotely resembles the actin fold.

The Cyclin Fold

The eukaryotic cell cycle is regulated by a series of cyclin-dependent kinases (CDKs) that are activated by association with cyclins. Several different forms of cyclins (i.e., cyclins A, B, C, D, E, and H) are known to regulate separate CDKs. Conserved among these cyclins is a region of ~100 amino acids, termed the cyclin box. The basic cyclin box fold contains five α-helices with the central helix, α3, being surrounded by the remaining four (Jeffrey et al., 1995; Noble et al., 1997). Cyclin A consists of two cyclin box domains in tandem (Fig. 17.1.14C and Table 17.1.5). The cyclin box fold is also observed in transcription factor IIB and retinoblastoma tumor suppressor domains. Based on the structure of a cyclinA-CDK2 complex (De Bondt et al., 1993), the activation of CDK has been postulated to occur through a cyclin-induced conformational change in the PSTAIRE helix (C helix) and the T-loop of the kinase domain.

DNA-BINDING STRUCTURAL MOTIFS AND DOMAINS

Numerous DNA-binding domains are observed to utilize a small number of conserved substructures or structural motifs for binding DNA. In most cases a portion of the structural motif protrudes from the domain and interacts with the major and/or minor groove of DNA and also with the phosphate backbone through hydrogen bonds, salt bridges, and nonpolar van der Waals interactions. By far the most prevalent interactions involve an α-helix binding in the major groove. Other less common interactions include β-strands or loops binding in either the major or minor groove of the DNA. In a few cases, considerable distortion of B-form DNA is induced upon binding. This section summarizes some commonly observed DNA-binding structural motifs and DNA-binding domains.

Helix-Turn-Helix Motifs

The helix-turn-helix (HTH) structural motif, first observed in the crystal structure of the λ Cro protein (Anderson et al., 1981), is one of the most common and well characterized DNA-binding motifs, and is found in a very diverse group of proteins (Table 17.1.6). The HTH motif is part of a three-helix bundle in which the second and third helices (α2 and α3) of the bundle make direct contacts with DNA. The third helix, the recognition helix, protrudes from the rest of the protein into the major groove of DNA, making the majority (but not all) of the base-specific contacts, while the second helix sits above the recognition helix at an ~120° angle to it, forming a bridge across the two phosphate backbones on either side of the major groove (Fig. 17.1.15). The N-terminal helix (α1) lies across the second helix opposite the DNA, completing the helical bundle where it forms interactions with the rest of the protein and may or may not interact with the DNA.

There are now tens of permutations of the DNA-binding HTH motif listed as various superfamilies in the online SCOP database (see Web-Based Structural Bioinformatics). For the purposes of this overview, four of these permu-

tations are discussed here: prokaryotic or phage repressor, eukaryotic homeodomain, winged helix, and PurR repressor, with some structures falling into more than one category (Table 17.1.6; Fig. 17.1.15).

Prokaryotic repressor. The structures of prokaryotic repressor HTHs have served as the prototypical model for the overall HTH motif. In this particular class, the HTH motif by itself is not stable enough to fold or bind DNA, but requires the scaffolding of the rest of the protein. Furthermore, monomers often lack DNA-binding ability and thus dimerization is required for binding to DNA. Interestingly, outside the HTH motif of prokaryotic repressors there is little structural consensus among these DNA-binding proteins. Their structures have been observed to vary from α/β folds to all helical folds.

Eukaryotic homeodomain. The eukaryotic homeodomain HTHs all use an N-terminal arm from helix 1 to bind DNA in the adjacent minor groove (Li et al., 1995). Some also use a C-terminal arm for minor groove binding as well. The additional interactions of the arm(s) enable homeodomains to bind DNA as monomers, although most homeodomains are observed to function as homo- or heterodimers in vivo. Also, some proteins in this class (e.g., yeast RAP1 telomere-binding protein, proto-oncogene product Myb, the Paired class homeodomains, the Oct-1 POU domain) contain two HTH motifs connected in tandem by a polypeptide linker.

Winged-helix motif. Another class of HTH proteins is referred to as the winged-helix motif (Fig. 17.1.15D). First characterized as a motif in the eukaryotic HNF-3/fork head transcription factor (Clark et al., 1993), this motif uses one or two flexible loops (or wings) that protrude from the HTH on either side at approximately right angles to the recognition helix to provide additional DNA contacts. The more prominent wing is usually part of a small two- or three-stranded antiparallel β-sheet, a distinctive feature of the winged-helix motif. For a detailed review of the winged helix motif see Gajiwala and Burley (2000).

PurR repressor. The PurR repressor of the LacI family represents a more extreme variant of the HTH motif with three major differences from the canonical HTH (Schumacher et al., 1994). First, although the spatial orientation is consistent with the HTH, the connectivity of the three helices is completely different so that helices 1 and 2 interact with the DNA instead of helices 2 and 3 (Fig. 17.1.15E). Second, the orientation of the HTH with respect to the DNA dyad is in the reverse direction from what is observed in the classical HTH. Third, in addition to the HTH motif a fourth helix (the hinge helix) is present to bind the DNA in the adjacent minor groove. In a physiological dimer, the two symmetry-related hinge helices interact with each other in an antiparallel manner to pry open the minor groove with a set of leucine levers which forces the DNA axis to bend 45°.

Zinc-Containing DNA-Binding Motifs

Six of the many classes of DNA-binding proteins that use zinc coordination to form the stable core or folding template for DNA-binding will be discussed in this section. These include the TFIIIA type zinc finger, the GATA-1 Cys_4 type zinc finger, the Zn_2Cys_6 binuclear cluster, the nuclear hormone receptor DNA-binding domains, the transcription elongation factor TFIIS, and the zinc binding loop of the p53 tumor suppressor.

Zinc fingers

The Cys_2His_2 zinc finger structural motif is the most common DNA-binding motif among eukaryotic transcription factors, with several thousand protein sequences identified (Berg and Shi, 1996). A canonical zinc finger motif contains the consensus sequence (Tyr/Phe)-X-Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His-X_{3-5}-His where X represents any amino acid, and the Cys and His residues serve as zinc ligands (Berg and Shi, 1996). The structure of the motif is comprised of two short antiparallel β-strands followed by an α-helix, which are stabilized by a small hydrophobic core and a tetrahedrally coordinated zinc atom (Fig. 17.1.16A). The N-terminal end of the α-helix rests in the major groove where it makes base-specific contacts while the β-sheet lies on the other side of the helix opposite the DNA (Fig. 17.1.17A) to form limited contact with the phosphate backbone. Zinc finger-containing proteins generally use this motif in tandem arrays within the DNA-binding domain. Since the first three-dimensional structure determination of a zinc finger-DNA complex (Zif268 from mouse intermediate protein; Pavletich and Pabo, 1991), the structures of several other protein-DNA complexes have been elucidated, including the DNA-binding domains from human oncogene GLI, the *Drosophila* regulator protein tramtrack, and TFIIIa, as well as numerous structures of uncomplexed zinc fingers (Fairall et al., 1993; Pavletich and Pabo, 1993; Foster et al., 1997).

The zinc-ribbon motif

First observed in the eukaryotic transcription elongation factor TFIIS (Qian et al., 1993), the zinc ribbon motif consists of a small three- or sometimes four- or five-stranded antiparallel β-sheet that coordinates a zinc atom within the loop region at one end of the sheet (Fig 17.1.16B; Pan and Wigley, 2000). The zinc is coordinated by three Cys and one His residue (or four Cys side chains in the case of TFIIS). The mechanism of DNA binding for this motif is controversial and awaits further clarification from future structure determinations of zinc ribbon-DNA complexes.

Zn_2Cys_6 binuclear cluster

A DNA-binding structural motif containing a Zn_2Cys_6 binuclear cluster has been characterized in numerous fragments of fungal transcription factors. The three-dimensional structure of fragments of these proteins—e.g., GAL4 and pyrimidine pathway regulator 1 (PPR1)—complexed with DNA have been determined (Marmorstein et al., 1992; Marmorstein and Harrison, 1994), revealing homodimers each containing three regions: two N-terminal zinc-coordinated DNA-binding modules (one in each monomer), a linker region, and a short coiled-coil made from two C-terminal helices which protrude away from the DNA. The DNA-binding modules are nearly identical in both proteins, consisting of two short helices, each followed by an extended loop (Fig. 17.1.16C). Six cysteines from each module tetrahedrally coordinate two zinc atoms. An internal two-fold symmetry relates the two helix-loop zinc-coordinated motifs within each module. In both GAL4 and PPR1, the C-terminal end of the first helix from each module points into the major groove, mediating base-specific DNA contacts to a CGG triplet (Fig. 17.1.17C and D, respectively). Apart from the DNA-binding motifs, the structures of GAL4 and PPR1 are rather different from each other. The linker region in each monomer of GAL4 is completely extended, making practically no contacts with the rest of the protein, whereas this same region in PPR1 folds into a more compact antiparallel β-ribbon. The GAL4 homodimer is highly symmetric, with the coiled-coil dimerization region projecting perpendicularly from the DNA. The PPR1 homodimer, however, becomes very asymmetric beyond the N-terminal DNA-binding modules, leaning to one side to form a hydrophobic core. Several other complexes of Zn_2Cys_6 motif-containing proteins with DNA show similar perturbations of domain orientation while retaining the basic fold (Table 17.1.6).

GATA-binding motif

The NMR solution structure of the chicken erythroid transcription factor GATA-1 complexed with its target DNA reveals a DNA-binding structural motif distinctly different from the canonical zinc finger structure (Omichinski et al., 1993). The consensus sequence Cys-X_2-Cys-X_{17}-Cys-X_2-Cys characterizes an entire family of regulatory proteins expressed in numerous cell types, each of which recognize a GATA DNA sequence. The DNA-binding domain of GATA-1 consists of two short two-stranded irregular antiparallel β-sheets followed by an α-helix and a long C-terminal loop (Fig. 17.1.16D). The tetrahedrally coordinated zinc is ligated by cysteines from the first β-sheet and helix. The α-helix and loop which connects the two β-sheets are both located in the DNA major groove, and the long C-terminal tail crosses the phosphate backbone to sit in the minor groove (Fig. 17.1.17B).

DNA-binding domains of nuclear hormone receptors

Eukaryotic nuclear hormone receptors form a superfamily of ligand-activated transcription factors that regulate numerous cellular processes. These proteins generally consist of an N-terminal DNA-binding domain and a C-terminal hormone-binding domain responsible for transactivation. The minimal core of the N-terminal DNA-binding domain (60 to 70 residues) consists of two helices separated by a long loop packed against each other at approximately right angles (Fig. 17.1.16E). Two zinc atoms, one at the N-terminus of each helix, are each coordinated by four cysteine side chains. The first helix, also termed the recognition helix, lies in the major groove of bound DNA, making sequence-specific contacts while other portions of the core domain make extensive contacts with the phosphate backbone on either side of the major groove (Fig. 17.1.17E). Nuclear receptors are observed to bind cooperatively to both direct and inverted DNA repeats as either homo- or heterodimers. The three-dimensional structures of the DNA-binding domains of several nuclear receptors complexed with DNA have been elucidated (Table 17.1.6). These include the glucocorticoid receptor (Luisi et al., 1991), estrogen receptor (Schwabe et al., 1993), 9-*cis* retinoic

acid receptor, and thyroid hormone receptor (Rastinejad et al., 1995).

MADS Box Domain

The MADS box (also known as the SRF-like) family of proteins comprises many eukaryotic regulatory proteins, including human serum response factor (SRF), *Drosophila* trachea development factor (DSRF), the METF2 myocyte-specific enhancer factors, and the MCM1 gene regulator from yeast, all of which contain a highly conserved 60-residue core consensus sequence (Table 17.1.6; Pellegrini et al., 1995). SRF was the first of these proteins to have its three-dimensional structure determined (Pellegrini et al., 1995). This protein is homodimeric and consists of three layers (Fig. 17.1.18A). The first layer is a left-handed antiparallel coiled-coil formed from an amphipathic helix from each monomer. The coiled-coil sits above the minor groove near the DNA dyad (the binding site) in an approximate parallel orientation to the minor groove and makes contacts with the phosphate backbone. The DNA is bent by 72° along its helical axis at this point, wrapping around the coiled-coil and enabling the N-terminal portions of the two helices to fit into the major groove of the DNA. In addition, N-terminal arms from each helix extend across the phosphate backbone and bind in the minor groove of the DNA. The coiled-coil is supported by a four-stranded antiparallel β-sheet that lies flat across the helices opposite the DNA. The third layer consists of a small two-helix subdomain above the sheet. Similar to other dimeric DNA-binding proteins, the two-fold symmetry axis of the dimer often coincides with the dyad axis of the DNA binding site. The crystal structure of the MCM1 MADS-box protein bound to DNA and the MATα2 homeodomain core (Tan and Richmond, 1998) reveals an intriguing ternary complex in which a β-hairpin from MATα2 extends the MCM1 β-sheet, and both the MATα2 HTH motif and the MCM1 MADS-box dimer bind DNA.

Basic Region Leucine Zippers

The fold of basic region leucine-zipper (bZIP) domains in eukaryotic transcription factors is relatively simple. First observed in the structure of the GCN4 homodimer (Ellenberger et al., 1992) and the c-Fos/c-Jun heterodimer (Glover and Harrison, 1995), it consists of two, long parallel α-helices that dimerize via a coiled-coil leucine zipper motif at their C-terminal ends and spread apart like partially opened scissors at the N-terminal end, gripping the DNA between them (Fig. 17.1.18B; Table 17.1.6). The primary sequence of leucine zippers contains heptad repeats (**a-b-c-d-e-f-g**) with hydrophobic residues at position **a**, leucine at position **d**, and charged or hydrophilic residues at the other positions (Ellenberger, 1994). Residues at positions **a** and **d** from each helix form extensive hydrophobic interactions at the core of the dimer interface, while charged side chains at positions **e** and **g** form inter- and intrahelical salt bridges at the edge of the interface. The N-terminal primarily basic ends of the helices lie in the major groove on either side of the DNA contacting base-pair edges and the phosphate backbone.

The Helix-Loop-Helix Motif

The helix-loop-helix (HLH) structural motif is similar to the bZIP domain in both its DNA-binding and dimerization strategies. As with bZIP proteins, this structural motif binds DNA as a homo- or heterodimer. Each monomer consists of two long α-helices connected by a loop of variable length (5 to 23 residues) and composition forming a left-handed, parallel, four-helix bundle in the dimer (Fig. 17.1.18C). The first helix in each monomer binds DNA with its N-terminal basic region sitting in the major groove on opposite sides of the DNA helix. Dimer contacts are made through the C-terminal portion of the first helix and a coiled-coil formed by the second helix of each monomer. In some members of the HLH family, the coiled-coil contacts are formed by a leucine zipper; in other HLH proteins the second helix is shorter and the coiled-coil is stabilized primarily through less regular hydrophobic contacts. In both cases, the dimer is required for DNA binding. DNA bound to this motif is primarily B-form with very little distortion observed. Examples of three-dimensional structures of HLH proteins include the eukaryotic transcription factor Max (Ferre-D'Amare et al., 1993), the MyoD transcription activator (Ma et al., 1994), the HLH region of the upstream stimulatory factor (USF; Ferre-D'Amare et al., 1994), and the E47 regulatory protein (Ellenberger et al., 1994; Table 17.1.6).

HMG Box Domains

DNA-binding proteins containing the high-mobility-group (HMG) box comprise a diverse superfamily of eukaryotic regulatory proteins including various transcription factors (e.g., LEF-1), components of chromatin, products of fungal mating type genes, and the mammalian

sex-determining gene product SRY (Table 17.1.6; for reviews see Laudet et al., 1993; Travers, 2000). The HMG consensus box consists of ~80 residues with an average conservation of 25% amino acid identity between boxes. HMG box proteins fall into two subfamilies. The first is observed in all cell types, contains multiple HMG box domains, and tends to bind DNA with little specificity, while the second subfamily is found in fewer cell types, contains only one HMG box domain, and binds DNA specifically. In the mid-1990s, the NMR solution structures of the HMG domains in two proteins of the second subfamily complexed with DNA (Love et al., 1995; Werner et al., 1995) and the solution structures of uncomplexed HMG domains from the first subfamily revealed a three-helix L-shaped structure which binds DNA exclusively in the minor groove and induces profound structural changes in bound DNA (Fig. 17.1.18D). Helices 1 and 2 are antiparallel to each other (with an acute angle between them) forming the base of the L. Helix 3, the longest helix, is almost perpendicular to helix 2 and forms the vertical arm of the L. All three helices participate in forming a concave surface for binding the minor groove and phosphate backbone of the target DNA, causing severe DNA unwinding and bending away from the major groove.

The Histone Fold

A unique fold shared by the H2A, H2B, H3, and H4 subunits of the octameric histone core of the nucleosome was first described by Arents et al. (1991) as the histone fold. Since then, the histone fold has been observed in the *archeael* histones, the TBP-associated factor (TAF) family of transcription factors (e.g., $TAF_{II}42$, $TAF_{II}62$, $TAF_{II}18$, $TAF_{II}28$; Xie et al., 1996; Birck et al., 1998), and the TBP-associated negative cofactors (Kamada et al., 2001). The individual polypeptides display a helix-loop-helix structure, with a central long α-helix flanked on each side by a short α-helix (Fig. 17.1.18E). The histone core subunits and the TAF fragments form similar heterodimeric and heterotetrameric structures. The function of histones in eukaryotic cells is to provide a scaffold for nuclear DNA to wrap around into a highly compacted nucleosome. It has been postulated in TAFs that a similar TAF octameric organization is involved in transcription activation.

DNA-Binding Proteins that use β-Sheet Motifs

The vast majority of DNA-binding proteins structurally characterized to date are observed to interact with DNA through helical structures. However, three different families of DNA-binding proteins have been characterized in which β-sheet structures are used as the primary contacts for protein-DNA interactions.

Ribbon-helix-helix motif

The ribbon-helix-helix motif is found in prokaryotic and phage repressor proteins such as Arc (Raumann et al., 1994), MetJ (Somers and Phillips, 1992) and Mnt (Burgering et al., 1994; Table 17.1.6). These proteins bind DNA as dimers of dimers (homotetramers) and all have homologous structures in which the homodimer consists of a small helical bundle with a symmetric double-stranded antiparallel β-sheet (β-ribbon) on the surface that can fit specifically into the major groove of DNA (Fig. 17.1.19A). In the three-dimensional structures of complexes with DNA, additional contacts are also made through helix and loop residues, but base-specific contacts to DNA are mediated exclusively through side chains of residues in the β-ribbon. Binding of the tetramers to the two tandem operator sites is cooperative and occurs with limited distortion of B-form DNA.

TATA-box binding protein

A second β-sheet motif for binding DNA is observed in the TATA box–binding protein (TBP), a component of the class II general transcription initiation factor TFIID that is required for preinitiation-complex formation with RNA polymerase II (Table 17.1.6). TBP consists of a C-terminal region of ~180 residues which is highly conserved across species, and a variable length N-terminal region. The crystal structure of TBP from *Arabidopsis thaliana* bound to the TATA-box DNA consensus sequence reveals a quasi-symmetric α/β structure composed of a curved ten-stranded β-sheet with four α-helices sitting on the convex side of the sheet with an overall shape resembling a saddle (Fig. 17.1.19B; Burley, 1996). Two distinct loops project away from the protein on each end of the β-sheet forming stirrups for the saddle. DNA containing the TATA box consensus sequence binds to the concave surface formed by the eight central strands of the β-sheet with an induced fit mechanism resulting in remarkably distorted DNA that is partially unwound and bent almost 90°. TBP has marked intramolecular pseudosymmetry with an ap-

proximate two-fold axis passing through the center of the protein perpendicular to the β-sheet. The primary sequence of TBP can be divided into two long repeats (89 to 90 residues each) related by 30% sequence identity. The three-dimensional structure of each half is topologically identical with a root-mean-square (rms) deviation between equivalent Cα positions of 1.1 Å. Several other protein structures are observed to contain a fold similar to one TBP repeat (as opposed to two); however, none of them use the β-sheet to bind DNA.

The histone-like HU family

A third family of proteins that use β-ribbons to bind DNA is the highly conserved histone-like HU protein family (Table 17.1.6). These are prokaryotic dimeric proteins that bind DNA nonspecifically to form nucleosome-like particles containing negatively supercoiled DNA. The crystal structure of a 19-kDa uncomplexed homodimeric HU protein (also known as DNA-binding protein II) from *Bacillus stearothermophilus* reveals a V-shaped protein consisting of a six-helix bundle with two three-stranded, curved, antiparallel β-sheets extending out from the surface to form an open cradle (Fig. 17.1.19C; Tanaka et al., 1984). The crystal structure of a similar protein, heterodimeric integration host factor (IHF) from *E. coli*, complexed with DNA (Rice et al., 1996), reveals that the two β-sheets wrap around the DNA along the minor groove inducing a 160° bend in the DNA (Fig 17.19D).

The Ig-Like Family of Transcription Factors

The structures of several families of Ig-like transcription factors have been elucidated (Table 17.1.6). Each of them uses an Ig-like fold with C2-type (also known as s-type) topology to bind DNA through loop regions at one end of the β sandwich (see Immunoglobulins and Immunoglobulin-Like Superfamilies for a discussion of Ig folds). The SCOP database classifies these proteins as the superfamily of p53-like transcription factors (Murzin et al., 1995; see Web-Based Structural Bioinformatics for SCOP database). Although the individual structures and binding modes vary somewhat among these families, three common loops are used by all six families to contact DNA (Rudolph and Gergen, 2001). The loop between strands A and B (AB loop, sometimes referred to as the recognition loop) interacts with the major groove, making sequence-specific contacts. The loop between strands E and F (EF loop) contacts the DNA backbone, in some cases making base-specific contacts in the minor groove, and the C-terminal tail emanating from strand G sits in the major groove, making base-specific contacts. The structural details of these families are described below.

p53 tumor suppressors

The first of these families to be examined structurally, p53 tumor supressors (Table 17.1.6), can bind DNA as a monomer (Cho et al., 1994). p53 tumor suppressors are key proteins facilitating protection against many types of cancer in animals. It is estimated that up to one-half of all cancers involve genetic inactivation of p53 (Vogelstein, 1990). In this structure, the major groove-binding C-terminal tail is an α-helix and the EF loop is stabilized by a zinc atom (Fig. 17.1.19E).

NF-κB

The NF-κB transcription factors are part of the larger Rel family of transcription factors containing a Rel homology region (RHR). They function as homo- and heterodimers; each monomer being composed of two Ig-like domains. The first domain is a DNA-binding C2-type fold with an extension of strand E (designated E′) and an additional insert of two antiparallel α-helices between strands E and F. The second domain has an E-type Ig fold that provides the dimerization contacts. The NF-κB dimer/DNA complex has been described as resembling a butterfly with a DNA torso and protein wings (Fig. 17.1.19F; Ghosh et al., 1995; Muller et al., 1995).

NFAT

The NFAT family of transcription factors is similar to NF-κB in that its members possess a DNA-binding N-terminal C2-type Ig-like fold and a C-terminal E-type Ig fold. Furthermore, the DNA-binding domain has an additional helix inserted between strands E and F, two extra β-strands between C and C′, and an additional β-strand after strand E (E′). It is not surprising that the TonEBP (NFAT5) transcription factor forms the same type of dimeric butterfly complex with DNA as the NF-κB proteins with the added feature that its DNA-binding domains make dimer contacts as well, thus completely encircling the DNA (Stroud et al., 2002); however, other NFAT proteins bind DNA as monomers (Zhou et al., 1998) or bind DNA cooperatively with AP-1 transcription factors. The structure of a tetrameric complex of NFAT1/Fos/Jun/DNA (Chen et al., 1998b)

shows the N-terminal domain interacting with DNA through the AB and EF loops and interdomain linker, as observed in other Ig-like transcription factors (Fig. 17.1.19G). However, the C-terminal domain forms an acute angle with the N-terminal domain and maintains only minimal contact with the N-terminal domain, the bound DNA, and the neighboring Fos helix. The Fos and Jun helices interact extensively with the NFAT1 N-terminal domain resulting in a 20° DNA bend and a 15° bend in the middle of the Fos/Jun heterodimer (relative to the Fos/Jun/DNA complex (see Basic Region Leucine Zippers; Chen et al., 1998b).

CBF

The core binding factors (CBF) comprise a family of heterodimeric transcription factors. The α subunit (CBFα) has the C2-type Ig-like fold (Berardi et al., 1999; Nagata et al., 1999) for binding DNA and is encoded by several genes. The invariant β subunit (CBFβ) has an α/β fold consisting of a partly open six-stranded β-barrel with four α-helices packed against the top and bottom (Fig. 17.1.19H; Goger et al., 1999; Huang et al., 1999). Although CBFβ does not interact with DNA itself, CBFα-DNA binding is apparently enhanced significantly by dimerization with CBFβ. Unlike the two Ig-like domains in the NFAT1/DNA complex, the CBF α and β subunits interact extensively with each other either alone or in complex with DNA (Fig. 17.1.19H; Warren et al., 2000; Bravo et al., 2001; Tahirov et al., 2001).

T-domain

The T-domain (also known as T-box) transcription factors comprise a single C2-type Ig-like domain with an additional two C-terminal helices (H3 and H4) that are capped at the DNA-binding end by several additional small twisted β-strands. The crystal structure of a *Brachyury* T-domain/DNA complex reveals a triangle-shaped dimer with the two T-domains forming the sides and the DNA forming the bottom of the triangle (Fig. 17.1.19I; Muller and Herrmann, 1997). The C-terminal α-helix (H4) has a rather unusual DNA interaction, being embedded deeply in the minor groove.

STAT proteins

The signal transducers and activators of transcription (STAT) family of eukaryotic transcription factors includes large proteins (750 to 850 residues) that dimerize (homo- and heterodimers) upon phosphorylation and translocate to the nucleus. Crystal structures of STAT-1 and -3β homodimers bound to DNA reveal elongated, four-domain monomers forming an arc-shaped ternary complex with DNA (Fig. 17.1.19J; Becker et al., 1998; Chen et al., 1998c). The N-terminal domain is a coiled-coil bundle of four antiparallel α helices. These helices project away from the DNA, have no interactions with DNA or the other monomer, and possess a primarily hydrophilic surface. The second domain is the DNA-binding domain. Like the other members of the Ig-like transcription factor superfamily, this domain is a modified C2-type Ig fold. The AB loop has an extra β-strand (A′), the CC′ loop contains an inserted α-helix (α5) and β-strand (X), the EF loop has an extra β-strand (E′), and β-strand G ends with an additional β-strand (G′). All three of the modified loops and the C-terminal tail participate in DNA binding. Following the Ig domain and prior to the SH2 domain is a helical connector or linker domain consisting of a bundle of four α helices and a small two-stranded β-sheet. The SH2 domain (see Modular Domains Involved in Signal Transduction) binds a phosphorylated tyrosine from the C-terminal tail of the other monomer in the complex, effectively clamping the two monomers together.

RNA-BINDING STRUCTURAL MOTIFS AND DOMAINS

Relative to DNA, RNA is capable of folding into a much more diverse array of structural motifs. In addition to the standard A-form double helix, irregularities such as bulges, hairpin loops, helix unwinding, noncanonical base-pairing, and unstacked bases are all common in RNA structures. This results in RNA-binding protein structural motifs designed to interact with various perturbations of the RNA A-helix rather than a focus on major- and minor-groove interactions as observed in DNA-binding proteins. Although numerous three-dimensional structures have been determined for RNA-binding proteins, many of these use either unique or less common folds to recognize RNA; however, a small group of RNA-binding modules with known structures are noted for their more widespread use in RNA-protein interactions. These include RNP, OB-type domains, dsRBP, KH domains, and zinc fingers. This group of folds, along with a handful of larger, more unique structures, are briefly described below and in Table 17.1.7. For a review of some of these folds, see Perez-Canadillas and Varani, (2001).

The RNP Domain

The RNP domain, also known as the RNA recognition motif (RRM) or consensus sequence RNA-binding domain (cs-RBD), is one of the most widely observed and well characterized RNA-binding domains (Burd and Dreyfuss, 1994). It is identified in amino acid sequences by two highly conserved consensus sequences referred to as RNP1 and -2, within a more weakly conserved region of ~80 residues. The SMART database lists over 3400 proteins containing RNP (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*). The three-dimensional structure of RNP domains determined in the early 1990s (U1A splicosomal protein; Oubridge et al., 1994) and the RNA-binding domain of hnRNP C (Wittekind et al., 1992) revealed a four-stranded antiparallel β-sheet with two α-helices laying across one face with a βαββαβ topology. The crystal structure of the RNA-binding fragment of the U1A splicosomal protein complexed with a 21-nucleotide RNA hairpin showed that the RNA binds as an open structure to the exposed β-sheet surface opposite the helices (Fig. 17.1.20A; Oubridge et al., 1994). Two loops from one end of the β-sheet protrude through the middle of the RNA loop, positioning and stabilizing the RNA structure on the sheet, and preventing base pairing. This enables sequence-specific interactions to occur between the bases and both main- and side-chain atoms on the β-sheet. Although it has no sequence homology with RNP proteins, the ribosomal protein S6 has been observed to have the same fold with a topology identical to the U1A protein (Fig. 17.1.20B; Lindahl et al., 1994). Similar α/β folding patterns have also been observed in the structures of other ribosomal proteins such as L6, L7/L12, and L30 (Lindahl et al., 1994). The bacteriophage T4 regA translational regulator protein also contains two regions that have sequence homology to the RNP1- and -2 sequence motifs. This protein specifically represses the translation of 35 early T4 mRNAs. The 1.9-Å crystal structure of T4 regA consists of two antiparallel β-sheets (three and four strands, respectively) and four helices (Fig. 17.1.20C; Kang et al., 1995). The larger β-sheet contains the RNP1 and -2 sequence motifs in two adjacent β-strands exposed to solvent, and superficially resembles the RNA-binding β-sheet in the U1A protein.

OB-Fold

The cold-shock proteins CspA and -B, from *E. coli* and *Bacillus subtilis,* respectively, are implicated in binding both RNA and single-stranded DNA, and each contain a region that has sequence homology with a canonical RNP domain. The structures of these proteins (Schindelin et al., 1993, 1994; Schnuchel et al., 1993; Newkirk et al., 1994), however, consist of a five-stranded antiparallel β-barrel known as the oligonucleotide/oligosaccharide (OB) fold (Fig. 17.1.21A and B; Murzin, 1993). The RNP sequence motifs reside in two adjacent β-strands suggesting a possible RNA-binding site. A similar β-barrel structure is also observed in the ribosomal proteins S17 (Golden et al., 1993) and L14, and in the N-terminal anticodon-binding domains of both lysyl- and aspartyl-tRNA synthetases (Fig. 17.1.21C-E; Cavarelli et al., 1993). Since the early 1990s, nearly 100 unique protein structures with OB folds have been deposited in the PDB, revealing a remarkably versatile fold used to bind not only RNA, but also ssDNA, oligosaccarides, and proteins (Arcus, 2002).

KH and dsRBP Domains

Two other RNA-binding domains, the K homology (KH) and double-stranded RNA-binding domains (dsRBD), have structures that are similar in terms of general orientation of secondary structures (although not in topology) to the structure of the RNP domain. As with the RNP domain, these domains can be identified by a short consensus sequence. The three-dimensional structures for two dsRBD-containing proteins, *E. coli* RNase III (Kharrat et al., 1995) and *Drosophila* staufen protein (Fig. 17.1.22A; Bycroft et al., 1995), and one KH motif-containing protein, human vigilin (Fig. 17.1.22C; Castiglone Morelli et al., 1995), each contain a three-stranded antiparallel β-sheet with two or three helices packed against one face to form an α/β structure, albeit with different topology. Furthermore, a topological variant of the canonical KH domain has been observed in the ribosomal protein S3 from *T. thermophilus* (Carter et al., 2001; Grishin, 2001). Despite the superficial similarities in structure, the RNP, dsRBD, and KH domain RNA-binding structural motifs all appear to bind RNA by distinctly different modes. The dsRBD motif is proposed to interact with double-stranded RNA through the end loops of the β-sheet, the N-terminal end of the last helix, and a portion of one face of the β-sheet (Bycroft et al., 1995; Kharrat et al., 1995). The KH

domain has been suggested to bind single-stranded RNA via the helical side of the domain (Gibson et al., 1993). The N-terminal domain of ribosomal protein S5 (Fig. 17.1.22B) from *Bacillus stearothermophilus* shares a number of conserved residues with dsRBD of staufen protein in addition to a similar topology, suggesting it may bind RNA in the same way as a dsRBD motif (Bycroft et al., 1995).

Zinc Finger Motif

The zinc finger structural motif already described in this unit as a DNA-binding motif (see Zinc-Containing DNA-Binding Motifs), has also been implicated in binding RNA. A well characterized example is the TFIIIA protein which binds both 5S ribosomal RNA and the DNA encoding the 5S rRNA gene (Burd and Dreyfuss, 1994).

Four-Helix Bundle RNA-Binding Protein

The RNA-binding protein Rop (repressor of primer) from an *E. coli* plasmid helps regulate plasmid copy number by facilitating sense-antisense RNA pairing. This protein functions as a homodimeric four-helix bundle structure comprised of two helix-turn-helix monomers (Fig. 17.1.23A; see above). Rop has no structural or sequence homology to any other known RNA-binding proteins (Predki et al., 1995). The results of mutation studies combined with binding assays provide strong evidence that RNA binding occurs along one face of the bundle formed by helices 1 and 1′ (Predki et al., 1995).

Large RNA-Binding Structures

MS2 phage coat protein

One hundred and eighty copies of MS2 phage coat protein assemble to form the icosahedral viral capsid of the RNA bacteriophage MS2. Although comprised of the same amino acid sequence, the 129-residue MS2 subunits are present in the viral capsid in three structurally similar, but not identical, conformations referred to as A, B, and C (Valegard et al., 1994). Icosahedral symmetry results in the formation of AB and CC dimers arranged in the capsid as trimers of dimers (Fig. 17.1.23B). The dimers each consist of a ten-stranded antiparallel β-sheet with two small β-hairpins and four helices packed against one face of the sheet. In addition to its structural role, this protein also prevents translation of viral replicase by binding to replicase mRNA. A nineteen-nucleotide RNA hairpin of the Shine-Dalgarno sequence for viral replicase diffused into crystals of empty phage capsid was observed crystallographically to bind specifically to AB dimers of MS2 coat protein (Valegard et al., 1994). Crystal symmetry precluded crystallographic analysis of any RNA bound to the CC dimer. The RNA bound to AB interacts as a hairpin structure laying as a rigid crescent against the face of the β-sheet, opposite the helices, causing very little change in the protein structure relative to the unbound form of MS2.

Trp RNA-binding attenuation protein

Trp RNA-binding attenuation protein (TRAP) from *Bacillus Subtilis* binds to mRNA for the tryptophan (*trpEDCFBA*) operon in the presence of tryptophan, inducing the formation of a transcription terminator structure, thereby halting expression of the operon as tryptophan concentrations increase (Antson et al., 1995). The 1.8-Å crystal structure of TRAP in its activated form bound to tryptophan reveals a doughnut-shaped eleven-subunit oligomer containing 56% β-sheet (Fig. 17.1.23C). Each 75-residue subunit is composed of both a three- and four-stranded β-sheet packed face to face against each other. Tryptophan binds at one end of the interface between the two sheets in each subunit towards the center of the wheel. Adjacent three- and four-stranded β-sheets from neighboring subunits hydrogen bond to each other to form eleven seven-stranded β-sheets, each tilted in a way that resembles the blades of a turbine. Although the RNA binding site is not known, mutation analysis of both RNA and TRAP suggest that RNA containing eleven (G/U)AG elements forms a matching circle that binds to one side of the β-sheet wheel (Antson et al., 1995).

Elongation factor Tu

Elongation factor Tu (EF-Tu) is one of several elongation factors that participate in the translation of mRNA into a protein sequence on the ribosome. The crystal structure of EF-Tu from *Thermus aquaticus* in a ternary complex with Phe-tRNA^{Phe} and a GTP analogue shows a three-domain structure in which all three domains interact with tRNA (Nissen et al., 1995). The N-terminal domain 1 is a structurally conserved G-protein domain which binds guanine nucleotide. Domains 2 and 3 are seven- and six-stranded β-barrels respectively. The 3′ CCA-Phe end of the tRNA binds in a cleft at the interface between domains 1 and 2 whereas the 5′ end of the tRNA binds at the interface of

all three domains. The T-stem of the tRNA interacts primarily with the β-barrel in domain 3. Except for the terminal Phe side chain and terminal adenine, both at the 3′ end, EF-Tu interacts primarily with the ribose moieties and the phosphate backbone of the tRNA.

Aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetases (aaRS) are ATP-requiring enzymes that ligate amino acids to their cognate tRNAs during translation of RNA. The aaRS family of enzymes is remarkably diverse in terms of sequence, size, and three-dimensional structure. On the basis of sequence, structural analysis, and catalytic properties, aaRS can be divided into two classes, each comprising ten members (Eriani et al., 1990; Carter, 1993). As of early 2003, there are over 100 three-dimensional structures of aaRS and various complexes in the PDB (Table 17.1.8). These include bacterial and yeast aaRS for nineteen different amino acids (excluding alanine).

Class I aaRS all share a classical dinucleotide-binding or Rossman fold in the catalytic domain (Fig. 17.1.24A) and generally acylate the 2′ hydroxyl of the terminal ribose (see The Classical Dinucleotide-Binding Fold for a discussion of the Rossman fold). The two sequence motifs, His-Ile-Gly-His and Lys-Met-Ser-Lys-Ser (more generally known as Met-Ser-Lys) are both present in all class I enzymes in regions that bind ATP. Class I enzymes are generally single-chain polypeptides, except for MetRS, TyrRS, and TrpRS which are homodimers. The ten class I aaRS are further divided into three subclasses: Ia, -b, and -c (Cusack, 1995; Ibba and Soll, 2000).

Class II aaRS have a seven-stranded antiparallel β-sheet instead of a dinucleotide-binding fold in the catalytic domain (Fig. 17.1.24B) and primarily acylate the 3′ hydroxyl of the terminal ribose (except for PheRS which acylates the 2′ hydroxyl). The amino acid sequences of class II enzymes are identified by three different sequence motifs within the catalytic domain (Fig. 17.1.24B). The first sequence motif is involved in homodimerization, while the other two participate in catalysis. Eight class II enzymes are homodimers, two are $\alpha_2\beta_2$ heterotetramers (GlyRS and PheRS) and one is a homotetramer (AlaRS). Note that GlyRS comes in two varieties, a homodimer (*E. coli*) and a heterodimer. It is also noteworthy that there are two types of LysRS: class I and II. Like their class I counterparts, the class II enzymes are also divided into three subclasses: IIa, -b, and -c (Table 17.1.8; Cusack, 1995; Ibba and Soll, 2000).

Despite the strong conservation of catalytic domains observed within each aaRS class, the structures and arrangements of the other domains involved in RNA binding and dimerization are remarkably variable, even within certain subclasses. AaRS domains include α-helical bundles, β-barrels, α/β folds, and β-coiled-coils, among others. The sizes of single chains range from 300 to over 950 residues. Due to this variability, it is beyond the scope of this unit to describe each one or to develop structural classifications.

Signal Recognition Particle

Nascent polypeptide chains synthesized by ribosome machinery are targeted to different cellular compartments. All proteins with signal peptides are sorted for secretion or membrane location via signal recognition particles (SRPs), a ubiquitous machinery throughout all kingdoms of life. Although the structure of a holo-SRP is yet to be determined, fragments of SRPs with or without their bound RNA, and a fragment of SRP receptor have been solved (Table 17.1.7).

SRP domains

Eukaryotic SRPs consist of six proteins organized into two domains, the Alu domain and the S domain. The function of the Alu domain is to retard the ribosome elongation of a nascent polypeptide chain for targeting to endoplasmic reticulum. The S domain forms the core of the SRP by binding to the signal peptide and targets it to the SRP receptor-mediated translocation machinery. Both domains are organized around SRP RNA and are connected by RNA. The Alu domain contains a heterodimer of SRP9 and -14 bound to the 5′ and 3′ terminal SRP RNA. The crystal structure of a mammalian SRP9/14 dimer in complex with the 5′ RNA domain, and a construct containing both the 5′ and 3′ RNA domains has been solved (Fig. 17.1.24C; Weichenrieder et al., 2000). SRP9 and -14 share a common fold of (αβββα topology with the two helices packed on the convex side of the three-stranded β-sheet. The SRP9/14 heterodimer is formed with the β-sheets aligning against each other forming a six-stranded antiparallel β-sheet with the four helices packed on one side of the β-sheet. The RNA domain that is recognized by SRP9/14 consists of the first 47 nucleotides of the 5′ RNA and 39 nucleotides from the 3′ RNA domain. The RNA assumes a helical conformation with a 180° re-

verse bend in the middle at the SRP9/14 binding site.

The S domain is formed by an SRP19, an SRP54 that contains a GTPase domain, an SRP68/72 heterodimer, and the core part of SRP RNA. The structure of a human SRP19 in complex with its cognate RNA has been determined (Fig. 17.1.24D; Wild et al., 2001). SRP19 folds into a KH-like (K-homology) fold (see KH and dsRBP Domains) with a central three-stranded antiparallel β-sheet packed against two helices on one side of the sheet. It is different from the fold observed in SRP9 and -14, even though both have two α-helices packed against a three-stranded β-sheet. The core SRP54 comprises three domains, the amino-terminal N domain, the GTPase domain G, and the methionine-rich M domain that contains binding sites for signal peptide and SRP RNA. The structure of a *Thermus aquaticus* Ffh, a homolog of mammalian SRP54, showed that the N domain forms a four-helical bundle, and the G domain assumes a Ras-like GTPase fold with an insertion that is unique to the SRP family of GTPases (Keenan et al., 1998). The M domain of Ffh in complex with domain IV of 4.5S RNA revealed that the five α-helices containing M domain recognizes the minor groove of SRP RNA by a helix-turn-helix motif (Fig. 17.1.24E; Batey et al., 2000).

SRP receptor

SRP receptor mediates the ER translocation of signal peptide containing proteins through binding to the signal peptide complexed SRP. The translocation depends on the hydrolysis of GTP through a GTPase domain of the receptor, the so-called NG domain. The structure of the GTPase-containing portion of FtsY, a functional homologue of the SRP receptor of *E. coli*, revealed a GTPase fold that is unique to SRP-type GTPases, including Ffh of SRP. Unlike the Ras-related GTPases, the SRP-type GTPases contain a small insertion domain of ~29 amino acids called the I box, within the H-Ras fold. The I box comprises one β-strand flanked by two α-helices. Preceding the GTPase domain is a four helical bundle domain, called the N domain (Fig. 17.1.24F; Montoya et al., 1997).

CARBOHYDRATE-BINDING PROTEINS

Lectins represent a structurally diverse group of proteins that bind carbohydrate specifically. They are widely distributed throughout plants, animals, and even viruses, as well as bacterial toxins. Although often weak in binding monosaccharides, many lectins function as oligomeric proteins to increase the valency and thus, the affinity of binding toward carbohydrate ligands. Despite their functional similarity, neither the mode of oligomerization (dimers, tetramers, and pentamers have all been observed) nor the mode of carbohydrate recognition are strictly conserved in members of this family. Some members require Ca^{2+} or Mn^{2+} for saccharide binding, while others are thiodependent (Sharon, 1993; Drickamer, 1995; Rini, 1995a,b). Lectins can be divided into subtypes according to their structural homology. For the purpose of this overview, four more frequently observed subtypes are discussed: the legume and cereal lectins in plants, and the S-type and C-type lectins in animals (Table 17.1.9). β-trefoil lectins are discussed above (see Secondary Structure). Generally, members within each subtype of lectin share much higher degrees of structural conservation and conservation in carbohydrate recognition than the members of different subtypes.

Legume Lectins

Lectins from various leguminous plants have very homologous three-dimensional structures and share similar locations for the carbohydrate recognition site. They often form dimers or tetramers and require both Ca^{2+} and Mn^{2+} ligated to each domain for carbohydrate binding. The general fold consists of a thirteen-stranded β-sandwich formed from a six-stranded flat β-sheet packed against the convex face of a seven-stranded curled β-sheet. The carbohydrate-binding site is located on the outside concave portion of the seven-stranded β-sheet (Fig. 17.1.25A; Dessen et al., 1995). Although the general topology (i.e., connectivity and strand order) is preserved in legume lectins, some members of the family do have deleted or inserted β-strands, or permutated termini. Lectins usually have their N and C termini close to each other in structure. This makes termini permutation particularly simple without disrupting the overall structure. For example, the N- and C-termini for ConA from jack bean are located at the respective N-terminal end of strand F and the C-terminal end of strand E of soybean agglutinin (Fig. 17.1.25A).

Cereal Lectins

Wheat germ agglutinin (WGA) represents a member of plant chitin-binding lectins that are specific for both *N*-acetylglucosamine and *N*-acetylneuraminic acid (sialic acid). It functions as a homodimer of 17-kDa subunits. The X-ray

crystal structure of WGA shows that each monomer contains four domains, each with four disulfides, and has a tertiary structure consisting mostly of loops and a few one-turn α-helices with no regular β-structure (Fig. 17.1.25B; Wright and Jaeger, 1993). Other structurally homologous cereal lectins include hevein (Rodriguez-Romero et al., 1991) and *Urtica dioica* agglutinin (UDA; Saul et al., 2000).

C-Type Lectins

This family of animal lectins is characterized by a conserved ~120 amino acid Ca^{2+}-dependent carbohydrate-recognition domain (CRD) that is often associated with other functional domains to mediate cell adhesion and other cellular signaling events. Members of this group include endocytic glycoprotein receptors, macrophage receptors, selectins, and the soluble collectins. Each CRD consists of a set of five to seven short β-strands forming two abutting β-sheets that are sandwiched between two α-helices. The carbohydrate-binding site is located in loop regions at one end of the β-sheet (I-G-H; Fig. 17.1.25C; Weis et al., 1992; Graves et al., 1994; Boyington et al., 1999). Calcium ions are directly involved in the binding of saccharide. A much larger superfamily of proteins referred to as C-type lectin-like proteins is now being structurally characterized. Many of these proteins do not have calcium or carbohydrate binding sites even though they share the same fold. These include numerous cell surface receptors, particularly in the immune system.

S-Type Lectins

S-lectins are a family of thio-dependent, β-galactoside-binding proteins that have been proposed to play important roles in modulating cell-cell and cell-matrix interactions (Hirabayashi and Kasai, 1993). Galectins are the best known examples of this family. They often form dimers or associate with other accessory domains. Crystal structures of galectin-1 and -2 show a striking structural homology with legume lectins with the exception of a few inserted and deleted strands (Liao et al., 1994; Lobsanov et al., 1993). Each monomer folds into a curled β-sandwich with the flat sheet consisting of five β-strands and the curled sheet consisting of six β-strands (Fig. 17.1.25D). Although the carbohydrate binds on the concave portion of the curled sheet, similar to those in legume lectins, the specific mode of saccharide recognition is not conserved between legume lectins and galectins.

CALCIUM-BINDING PROTEINS

EF-Hand Calcium-Binding Motif

The EF-hand is the most commonly observed Ca^{2+}-binding motif in proteins (Satyshur et al., 1988; Nakayama and Kretsinger, 1994). As of early 2003, the SMART database listed over 2000 proteins with this domain (see Web-Based Structural Bioinformatics and *http://smart.embl-heidelberg.de/browse.shtml*) . Examples of EF-hand-containing proteins include calmodulin, troponin C, essential light-chain and regulatory light-chain myosin, parvalbumin, intestinal and sarcoplasm calcium-binding proteins, calcineurin B, recoverin, and extracellular matrix protein BM-40 (Table 17.1.10). There are two kinds of Ca^{2+}-binding sites among the EF-hand motif proteins: high and low affinity. The high affinity Ca^{2+}-binding sites appear to primarily provide structural stability to proteins, whereas the low affinity sites often function to regulate the active protein conformations, and hence are also termed regulatory sites. For example, calmodulin undergoes a conformational change upon binding of Ca^{2+} ions to the regulatory sites (Meador et al., 1992). Both crystal structures and NMR structures of a number of EF-hand proteins have been solved (Herzberg and James, 1988; Satyshur et al., 1988; Meador et al., 1992; Svensson et al., 1992; Flaherty et al., 1993; Rayment et al., 1993; Findlay et al., 1994; Xie et al., 1994; Hohenester et al., 1996). A typical EF-hand consists of a helix-loop-helix arrangement with the two helices oriented ~90° to each other and the Ca^{2+}-binding site located in the loop region (Fig. 17.1.26A). In most situations, they form pairs with the two EF-hands in a pair related by approximately two-fold symmetry. The Ca^{2+} coordination is usually described as a pentagon bipyramid geometry with a total of six to seven ligands which are from the carboxylate groups of Asp, Glu, and main chain carbonyl oxygens in the loop region. Water molecules are often found to provide one of the apical coordinations.

Annexins

Annexins are a family of calcium-dependent phospholipid-binding proteins. Unlike the classical EF-hand calcium-binding proteins, the annexins chelate calcium in an entirely different way (Huber et al., 1990a,b, 1992; Bewley et al., 1993). Each calcium-binding

domain comprises five alpha helices, A through E (Fig. 17.1.26B). The coordination geometry is that of a pentagonal bipyramid with three ligands from the backbone carbonyl oxygen groups of the residues in the AB loop (the loop between helices A and B). The three ligands are each separated by one amino acid in the sequence. Two other ligands are provided by the side chain carboxylate of a Asp or Glu residue, located at the end of helix D, 44 amino acids downstream from the first carbonyl ligand of the AB loop. The six and seventh ligands are less conserved and can be replaced with water molecules.

NUCLEOTIDE-BINDING DOMAINS

Nucleotide-binding folds first began to be characterized in the early 1970s as the structures of several NAD cofactor-containing dehydrogenases were elucidated (Rossmann et al., 1975). It was observed that single domains with conserved folding patterns were responsible for nucleotide-binding in many proteins. With the structural determination and analysis of many more nucleotide-binding proteins, this particular fold has been divided into three primary classes: dinucleotide-binding domains, the mononucleotide-binding fold, and the actin fold; however, not all nucleotide-binding proteins have these folding patterns. Proteins such as the class II aaRS (containing an antiparallel β-sheet for binding ATP) and some FMN-binding proteins (TIM-barrel structures), among others, have very different folds for interacting with single nucleotides. To avoid confusion, the di- and mononucleotide-binding folds discussed in this section (Table 17.1.11) are usually referred to in the literature as classical or canonical nucleotide-binding folds and the dinucleotide-binding fold is sometimes referred to simply as the Rossman fold.

The Classical Dinucleotide-Binding Fold

In general, classical (canonical) dinucleotide-binding proteins fall into one of three groups: NAD(P)-dependent dehydrogenases and oxidoreductases, FAD-linked oxidoreductases, and FAD/NAD-linked oxidoreductases. Although many of these proteins are multidomain α/β structures, their dinucleotide-binding domains typically have a common three-dimensional fold. These binding domains contain a central five- or six-stranded parallel sheet, strand order 3-2-1-4-5-6, connected to helices on each face through right-handed loops (Table 17.1.11; Fig. 17.1.27A). The conserved topology for this α/β/α sandwich motif as compiled by G.E. Schultz (1995) and A.M. Lesk (1995) is illustrated in Figure 17.1.27B. In each case, the dinucleotide binds across the C-terminal end of the sheet with the ADP portion interacting specifically with the central βαβ unit of this fold (β1-α1-β2). The core sequence motif (also called the fingerprint sequence) for this βαβ fold is Gly-X-Gly-X_2-Gly (where X is any amino acid) and is located in a loop between β-strand 1 and helix 1. This sequence permits unusual main chain dihedral angles in the binding loop and close contacts with ADP. The pyrophosphate moiety sits at the N-terminal end of the helix following the loop, a favorable environment for the negative charges (Wierenga et al., 1985). Another conserved feature is an Asp or Glu residue at the C-terminal end of the second β-strand of the βαβ unit for binding the ribose moieties of NAD and FAD, respectively. NADP-binding domains contain an Arg at this position for binding the 2′-phosphate (Schultz, 1992). One additional eleven-residue sequence motif has been located in β-strand 5 of the canonical dinucleotide binding fold in FAD-binding domains which includes a C-terminal aspartic acid that interacts with the ribitol moiety of FAD (Eggink et al., 1990).

The Classical Mononucleotide-Binding Fold

First noted in the structure of adenylate kinase (Fig. 17.1.27C; Schulz et al., 1974), the mononucleotide-binding fold is superficially very similar to the dinucleotide-binding fold (above), but has a different topology. This fold also consists of an α/β/α sandwich with a central five-stranded parallel β-sheet and right-handed connections to helices (Table 17.1.11, Fig. 17.1.27C and D; Schultz, 1992). The strand order is 2-3-1-4-5. As in the case of the dinucleotide-binding fold, mononucleotide binds to a critical glycine-rich loop at the C-terminal end of the β-sheet between strand 1 and helix 1. This loop (also referred to as the P-loop) is much longer than the shorter dinucleotide-binding loop, enabling it to form a giant positively charged anion hole for interacting with the mononucleotide triphosphate moiety (Dreusicke and Schulz, 1986). The loop's core sequence motif, Gly-X_2-Gly-X-Gly-Lys (where X is any amino acid) has been used to identify many more mononucleotide-binding proteins including some from the more distantly related GTP-binding proteins (also known as G proteins) which also have a P-loop (Saraste et al., 1990; Schultz, 1992; Traut,

1994). Slight variations of the mononucleotide-binding fold include the flavodoxin-like and the resolvase-like proteins. In these families, the first β-strand is displaced to create a strand order of 2-1-3-4-5 instead of 2-3-1-4-5. Furthermore, in the resolvase-like fold, the fifth β-strand is antiparallel to the rest. G proteins have an extra sixth strand in the β-sheet with strand order 2-3-1-4-5-6 and the second β-strand is antiparallel to the rest. Other variations include protein structures that do not contain the canonical P-loop fingerprint sequence—e.g., CheY chemotactic protein; Volz and Matsumura, 1991), γδ-resolvase (Sanderson et al., 1990), flavodoxin (Watenpaugh et al., 1972). Further perturbations of the mononucleotide-binding fold are listed under the heading Fold: P-loop Containing Nucleotide Triphosphate Hydrolases in the SCOP database (see Web-Based Structural Bioinformatics and *http://scop.mrc-lmb.cam.ac.uk/scop/index.html*). It should also be noted that some mononucleotide-binding proteins, such as adenylate kinase, also bind monophosphate nucleotides in a distinctly different binding site, albeit close in proximity to the classical site discussed in this section.

ENZYMES INVOLVED IN DNA/RNA REPLICATION AND DNA TRANSCRIPTION

DNA/RNA Polymerases

Polymerases are fundamentally important enzymes responsible for replication, repair, and transcription of DNA and RNA in all living organisms. The structures of these enzymes vary widely from small single-subunit proteins of only a few tens of kilodaltons in size to large multisubunit complexes approaching 1000 kDa (Joyce and Steitz, 1994).

DNA polymerases and small RNA polymerases

DNA polymerase and some small RNA polymerases have a central catalytic domain responsible for DNA binding, nucleotide binding, and phosphoryl transfer. The three-dimensional structures of the single-chain catalytic domains of several different nucleic acid polymerases from eukaryotic, prokaryotic, and viral organisms have been solved (Table 17.1.12). These include DNA polymerase structures from the Pol I (A family), Pol α (B family), Pol β (X family), lesion-bypass polymerase (Pol Y family), and reverse transcriptase (RT) families. In addition, there are structurally homologous eukaryotic poly(A) RNA polymerases and viral RNA polymerase structures from the T7 RNA polymerase family, the viral RNA-dependent RNA polymerase family, and the double-stranded (ds)phage RNA–dependent RNA polymerase family. Including mutants and structures of polymerases complexed with DNA, nucleotides, drugs, and various salts, the number of PDB entries is now in the hundreds (see Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*).

Despite numerous structural differences, the single-chain catalytic domains of the above-mentioned polymerases each possess a common arrangement of subdomains centered around a large nucleotide-binding cleft which can be loosely described as a U-shaped structure resembling a person's right hand. Using nomenclature introduced by Ollis et al. (1985), the central base of the structure is referred to as the palm subdomain, and the two vertical walls of the U are called the fingers and thumb subdomains (Fig. 17.1.28A). It is noteworthy that bovine and yeast poly(A) RNA polymerases do not have a thumb subdomain (Bard et al., 2000; Martin et al., 2000). The palm catalyzes the phosphoryl transfer reaction, the fingers are involved in template positioning, and the thumb may be important in DNA binding and processivity (Hubscher et al., 2002). Structures of polymerase/DNA and polymerase/DNA/dNTP complexes suggest that the thumb and finger subdomains in some cases make dramatic movements during different stages of the polymerization reaction (Fig. 17.1.28B; Beese et al., 1993; Pelletier et al., 1994; Franklin et al., 2001). Associated catalytic functions distinct from polymerase activity, such as the 3′-5′ exonuclease activity in the Klenow fragment (KF) of *E. coli* DNA polymerase I, RNase activity in RT, 5′-nuclease activity of pol-β and *Taq* polymerase, and nascent RNA binding by T7 RNA polymerase, are located in separate N- or C-terminal domains. Still more related functions reside in associated subunits. The central palm is the most structurally conserved subdomain among polymerases. It consists of a three- to six-stranded antiparallel β-sheet with two helices packed against one face of the sheet (Fig 17.1.28C), a structure similar to the RNP motif found in many RNA-binding proteins (see KH and dsRBP Domains). The exposed side of the sheet displays the conserved active site Asp residues. With regard to this subdomain, however, the eukaryotic β polymerase and eukaryotic poly(A) RNA polymerases are unique (Davies et al., 1994; Sawaya et al., 1994; Bard

et al., 2000; Martin et al., 2000). Although the palm subdomain has a similar α/β sandwich fold, the overall topology is quite different (Fig 17.1.28D). In contrast to the palm subdomain, the fingers and thumb subdomains appear to be functionally related rather than structurally analogous between polymerase domains. For example, the thumb subdomains of KF, *Taq* polymerase, and T7 RNA polymerase all consist of a helical bundle with one or two long helices; however, the thumb subdomains of pol-β and RT are α/β structures each having three α-helices and one or two β-strands, respectively. Finger subdomains are similarly divergent. The Y family polymerases have an additional conserved α/β domain sometimes referred to as the little finger (Ling et al., 2001). The eukaryotic poly(A) RNA polymerases have an additional C-terminal domain with an RRM-like fold (see The RNP Domain) that is known to interact with the RNA primer (Bard et al., 2000; Martin et al., 2000). Not only are the folds of the subdomains variable, but the order of the subdomains within the primary sequence differs as well in various polymerases (Sawaya et al., 1994). Moreover, RT functions as a heterodimer (p66/p51) while other homologous polymerase domains function as monomers. Interestingly, the p51 subunit of RT results from the cleavage of the C-terminal RNase H domain from p66. Thus, both p51 and p66 have the structurally equivalent thumb, fingers, and palm subdomains. In addition, both RT subunits also contain another α/β subdomain termed the connection subdomain, which bridges the RNase H and polymerase domains in the larger subunit (Jacobo-Molina et al., 1993). P66 and p51 differ dramatically in the spatial arrangement of their subdomains (Fig. 17.1.28E), resulting in two morphologically and functionally asymmetric subunits, with the polymerase activity residing only in the larger p66 subunit (Jacobo-Molina et al., 1993; Joyce and Steitz, 1994).

RNAP

The bacterial, archeal, and eukaryotic DNA-dependent RNA polymerases (RNAPs) are responsible for the transcription of DNA into RNA during gene expression. All known RNAPs including eukaryotic RNAP I, II, and III are members of a conserved protein family also known as the multisubunit RNAP family (Ebright, 2000). Structurally, RNAPs superficially resemble the DNA and viral RNA polymerases described above in that they possess a U-shaped structure; however, RNAPs are much larger, more complicated multisubunit enzymes that have a rather different architecture. The crystal structures of the prokaryotic core RNAP and holoenzyme RNAP from *Thermos aquaticus*, and eukaryotic RNAP II from *Saccharomyces cerevisiae* each reveal a large crab-claw-shaped molecule with a molecular weight of ~400 to 500 kDa (Fig. 17.1.28F and H; Zhang et al., 1999; Cramer et al., 2000, 2001; Murakami et al., 2002a; Vassylyev et al., 2002). The active site is found in the bottom of a deep channel between the two pincers of the claw. For clarity, most of following description of RNAPs is based on the simpler structure of the bacterial form rather than the structure from yeast.

The bacterial core RNAP is composed of five subunits ($\alpha_2\beta\beta'\omega$). The holoenzyme additionally incorporates a σ initiation factor that is essential for transcription initiation. In general terms, the β and β′ββ′ω subunits make up the two arms of the claw, respectively, and the two identical α subunits cap the end of the molecule opposite the claw opening (Fig. 17.1.28F). The ω subunit sits on the outer surface of the β′ subunit and the σ^{70} subunit straddles the ends of both pincers on one side of the claw, interacting with β and β′.

The α subunit N-terminal domain (NTD) consists of two α/β subdomains (Fig. 17.1.28G; Zhang and Darst, 1998; Zhang et al., 1999). Subdomain 1 is a four-stranded antiparallel β-sheet packed against two nearly orthogonal α-helices along one face. The second subdomain comprises a three-stranded antiparallel β-sheet flanked on one face by two two-stranded β-sheets and one α-helix. The two α-subunits dimerize through antiparallel helix interactions of the two helices in subdomain 1. The C-terminal domain (CTD) of the α subunit is not visible in the RNAP complex, but NMR studies have revealed a compact fold consisting of four small helices. The first three helices are nearly perpendicular to each other, while the fourth is antiparallel to the third (Jeon et al., 1995).

The ω subunit consists of a small four-helix bundle and a β-hairpin. The σ^{70} initiation factor of the *T. aquaticus* holoenzyme is an elongated all-helical protein comprising three domains (Fig. 17.1.28F, dark blue). The second and third domains are separated by a long linker region (Murakami et al., 2002a; Vassylyev et al., 2002). Crystal structures of *E. coli* σ^{70} and *T. aquaticus* σ^{A} factors reveal similar all-helical subunits (Malhotra et al., 1996; Campbell et al., 2002). A helix-turn helix motif (see Helix-

Turn-Helix Motif) in σ^A is observed to make critical interactions with the promoter DNA (Campbell et al., 2002).

The β and β′ subunits are rather complicated α/β polypeptides with eight and six structural domains, respectively. Due to their complexity, these two subunits are not described in this overview. For a detailed analysis of these two subunits, see Cramer et al. (2001) for the structurally homologous yeast RNAP II in which subunits RBP1 and -2 are equivalent to β′ and β subunits, respectively.

The RNAP II complex from yeast consists of ten subunits (RBP1 to -10; Fig. 17.1.28H). Five of the yeast subunits (RBP1, -2, -3, -11, and -6) correspond in structure and function to the five bacterial subunits (β′, β, α^I, α^{II}, and σ, respectively). In fact, R. Ebright (2000) has detailed remarkably extensive structural similarity between the bacterial and yeast RNAP complexes ranging from the general subunit locations to the folding topologies of individual domains. The yeast RNAP II structure has an additional five subunits, most of which decorate the outer surface of the 'claw', away from the internal channel (Fig. 17.1.28H; Cramer et al., 2001).

DNA Polymerase Processivity Factors

Certain DNA polymerases, such as those involved in replication, require a highly processive polymerase complex. The characteristic of processivity is conferred upon polymerases by associated processivity factors that tether the polymerase to DNA in an ATP-dependent manner in order to prevent dissociation (Kong et al., 1992). One type of processivity factor known as a sliding clamp is a simple ring-shaped structure that both encircles DNA and interacts with the respective polymerase. The three-dimensional structures of several of these sliding clamps, including the β-subunit homodimer of pol III, the RB69 and T4 bacteriophage processivity factors, and PCNA, are each star-shaped six-fold symmetric closed circular rings consisting of a circle of twelve α-helices perpendicular to the ring plane surrounded by six antiparallel β-sheets (Kong et al., 1992; Krishna et al., 1994; Shamoo and Steitz, 1999; Moarefi et al., 2000). The inside diameter of both rings is 30 to 35 Å and the outer diameters are each ~80 Å. Each ring is composed of six nearly identical domains, each of which is two-fold symmetric, consisting of two four-stranded antiparallel β-sheets packed edge-to-edge at a ~45° angle with two antiparallel helices packed against the concave side of the domain (Fig. 17.1.29A). The β-sheets of adjacent domains hydrogen bond to each other forming six curved eight-stranded antiparallel β-sheets that line the outer surface of the ring. The complete ring structure results from the 12-fold repeat of the simple topological structural motif βαβββ round the entire circle (Fig. 17.1.29B). Despite the striking structural similarity, there is only 15% sequence identity when individual domains from each protein are aligned (Krishna et al., 1994). Sequence alignments of individual domains within the same protein give similar results. A head-to-tail dimer of the β-subunit of pol III forms the entire six-fold circle; each subunit comprises three domains. The bacteriophage RB69 and T4 rings, and the PCNA ring consist of trimers with two domains per subunit. ATP and auxiliary proteins are required for assembly around duplex DNA and dissociation does not occur easily, unless the DNA is linearized. In each case the ring has a substantial net negative charge and positively charged side chains are concentrated on the twelve α-helices lining the central channel, which should allow favorable interaction with the phosphate backbone of DNA passing through the channel (Kong et al., 1992; Krishna et al., 1994). In addition, the orientation of helices facing the channel is such that they are nearly perpendicular to the phosphate backbone of B-form DNA, precluding specific interaction with the grooves. These physical characteristics, as well as the size of the inner diameter (30 to 35 Å), make both proteins well suited for encircling DNA in a manner that allows for unrestricted movement along a duplex DNA strand during polymerization. Finally, direct interaction has been observed between the T4 and RB69 sliding clamps and the corresponding DNA polymerases (Shamoo and Steitz, 1999). In other systems, intermediate proteins facilitate this polymerase linkage.

Topoisomerases

Topoisomerases are essential enzymes of eukaryotes, prokaryotes, and viruses that alter the topology of double-stranded DNA. They function by transiently breaking one or two strands of DNA, passing DNA through the opening, and then resealing the break (for reviews, see Wigley, 1995; Champoux, 2001). Topoisomerases can be divided into two primary classes: type I enzymes which transiently cleave only one strand of DNA, and type II enzymes which cleave both DNA strands during catalysis.

Type I topoisomerases

Type I enzymes are monomeric proteins which relax supercoiled DNA in an ATP-independent manner. They can also catenate and decatenate nicked double-stranded circles and form a double-stranded DNA circle from two complementary single-stranded circles. Subtype IA enzymes form a link to the 5′ phosphate of the DNA and subtype IB enzymes link to the 3′ phosphate.

Subtype IA. The three-dimensional structures of the N-terminal catalytic fragments of two subtype IA enzymes from *E. coli*, topoisomerases I and III, reveal a common structural architecture (Lima et al., 1994; Mondragon and DiGate, 1999). The N-terminal fragment is made of four distinct domains arranged around a large central channel with a diameter of 27.5 Å (Fig. 17.1.30A). Domain I is a four-stranded parallel β-sheet sandwiched between two pairs of α-helices, a fold similar to the classical dinucleotide-binding fold (see above). Domain II is an unusual looking β-sandwich composed of two long-curving three-stranded antiparallel β-sheets that cross each other orthogonally at one end of each sheet. A single helix lies between the sheets at one end of the sandwich. The β-strands of this domain line the inside of the top half of the central channel, bearing a striking resemblance to the TATA-box binding protein (see DNA-Binding Proteins that use β-Sheet Motifs; Burley, 1996). Domains III and IV are each primarily helical except for a small β-hairpin in domain IV. Domain III contains the active-site tyrosine, Tyr319, and closes one side of the central channel through an extensive interface with domains I and IV (Fig. 17.1.30A). This interface effectively buries Tyr319, suggesting that large conformational changes are required during catalysis. It has been proposed that domains II and III may move together as a rigid body away from domains I and IV, using a protease-sensitive region between domains II and IV as a hinge (Lima et al., 1994). Such a movement would open a gate to the central channel and expose the active site Tyr319. The size of the closed channel is large enough to accommodate double-stranded B-form DNA without steric hindrance. The inner surface of the channel has a strong positive potential suitable for interaction with DNA.

Subtype IB. Eukaryotic subtype IB topoisomerases share no structural or sequence similarity with prokaryotic subtype IA enzymes. Structural studies of an N-terminally truncated form of human topoisomerase IB (excluding an unstructured, highly charged ~200 residue N-terminal domain) in complex with a 22-base pair DNA duplex reveal a five-domain protein structure that wraps around the DNA (Redinbo et al., 1998; Stewart et al., 1998; Fig 17.1.30B and C). Core subdomains I, II, and III are each α/β folds, the linker domain is a two-helix antiparallel coiled-coil, and the C-terminal domain is composed of five short helices. Subdomain II is noted to resemble the fold of the homeodomain family despite only 11% sequence identity and no evidence from the topoisomerase structure that it interacts with DNA (Redinbo et al., 1998). Additionally, subdomain III shows striking structural similarity to bacteriophage HP1 integrase (Redinbo et al., 1998). Finally, the coiled-coil of the linker domain superimposes with the Rop RNA-binding protein (see Four-Helix Bundle RNA-Binding Protein) with a root-mean square (rms) distance of 0.8 Å for 60 C^{α} atoms (Stewart et al., 1998). The catalytic Tyr723 resides in the C-terminal domain within the positively charged channel region. Like prokaryotic topoisomerase IA, the central pore is lined with positively charged side chains for interaction with DNA, and concerted movement of subdomains I and II is proposed to take place during the catalytic mechanism, allowing DNA to enter and exit the complex; however the proposed mechanisms for the two enzymes are quite different (Redinbo et al., 1998; Stewart et al., 1998; Lima et al., 1994).

Type II topoismerases

Type II enzymes are multimeric proteins with homologous amino acid sequences. The eukaryotic type II topoisomerases are homodimers, while the prokaryotic enzymes are A_2B_2 tetramers where the A and B subunits correspond to the N- and C-terminal halves of the eukaryotic monomer, respectively; viral enzymes can have either subunit composition. Type II topoisomerases require ATP for catalysis and are capable of either relaxing DNA or introducing supercoils. Like the type I enzymes, they can also catenate or decatenate circular DNA. Crystal structures of the C-terminal half of yeast topoisomerase II, the homologous *E. coli* gyrA (gyrase subunit A), and the ATPase-containing subunit (gyrB) of *E. coli* gyrase complement each other and together provide a model for the structure of an intact type IIA topoisomerase. A crystal structure of the A subunit from *Methanococcus jannaschii* DNA topoisomerase VI provides insight regarding type IIB topoisomerases.

Subtype IIA. The monomer structure of the C-terminal 92-kDa yeast topoisomerse II fragment reveals a modular crescent-shaped protein with five domains (Fig. 17.1.30D). Berger et al (1996) divided the polypeptide chain into two subfragments, A′ and B′, based on their relation to the A and B subunits of the homologous bacterial enzymes. The N-terminal subfragment B′ has two domains, a Rossman fold–like five-stranded mostly parallel β-sheet (see The Classical Dinucleotide-Binding Fold for a discussion of the Rossman fold) sandwiched between four helices and a three-stranded antiparallel β-sheet packed against one helix. The C-terminal subfragment A′ has three domains. Domain I consists of a winged-helix motif (see above) of three helices and a three-stranded antiparallel β-sheet embedded between an N-terminal α-helix, two short C-terminal α-helices, and a β-hairpin. The active site tyrosine residue, Tyr783 (Fig. 17.1.30E), is located in a loop of the winged-helix at the bottom of a short tunnel leading to the protein surface. Domain II comprises a four-stranded antiparallel β-sheet packed against two helices on one face and an abutting three-stranded antiparallel β-sheet. The third domain of subfragment A′ consists of a six-helix bundle and a small β-hairpin. Three of the helices are unusually long, causing this domain to project out from the rest of the protein. In the initial crystal structure, the two monomers come together to form a flat heart-shaped dimer with a large pear-shaped central cavity 60 Å high and 55 Å wide towards the base (Fig. 17.1.30E). Although the primary dimer contacts occur at the A′-A′ interface, the B′-B′ interface involves extensive overlap of domains such that the B′ subfragment from one monomer interacts with the A′ subfragment of the other. At the interface between the first two domains in subfragment A′, a 20- to 25-Å groove runs across the protein surface including the region containing Tyr783. B-form DNA can be modeled into this groove with the recognition helix of the winged-helix motif lying in the DNA major groove (see Helix-Turn-Helix Motifs). A second crystal form of yeast topoisomerase II (Fass et al., 1999) and the structure of *E. coli* GyrA (Morais Cabral et al., 1997) reveal similar shaped dimers, but significant alterations in subdomain orientations between the three structures suggest a flexible complex that undergoes conformational changes during catalysis. The crystal structures of yeast topoisomerase II lack the N-terminal ATPase domain that is proposed to lie on top of the B′ subfragments; however, two crystal structures of the ATPase domain of the homologous *E. coli* gyrase B have been determined, enabling the modeling of the entire topoisomerase II structure (Wigley et al., 1991; Brino et al., 2000). The 43-kDa fragment of DNA gyrase B is an elongated molecule consisting of two α/β domains each with a unique fold (Fig. 17.1.30F). The first domain, which contains the ATPase site, is an eight-stranded mostly antiparallel β-sheet with five α-helices packed against the convex face. It binds ADPNP at the center of one face of the β-sheet with the phosphates coordinated by a glycine-rich loop at the N-terminus of a helix. The second domain (also referred to as the DNA capture domain) is a four-stranded mixed β-sheet surrounded by four helices, including one long C-terminal helix that protrudes away from the protein. In the gyrase B dimer, the most extensive interaction is through the N-terminal domains (Fig. 17.1.30F). A long 15-residue N-terminal arm from each monomer wraps around the other monomer contributing to the ATP-binding site. It is interesting to note that the binding of ATP is reported to induce dimerization (Berger et al., 1996). The C-terminal domain forms dimer contacts only through one end of the protruding helix from each monomer. This dimer structure creates a large 20 Å-wide hole between the two C-terminal domains (Fig. 17.1.30F). The C-terminal half of this hole is observed to be rich in Arg residues, suggesting possible interaction with DNA.

Subtype IIB. The crystal structure of a truncated form (missing residues 1 to 68) of the A subunit from *Methanococcus jannaschii* DNA topoisomerase VI reveals a U-shaped homodimer (Nichols et al., 1999; Fig. 17.1.30G). Each monomer consists of two domains linked by a short tether. The N-terminal domain has five α helices packed against a two-stranded β-sheet with a fold similar to the winged helix motif observed in catabolite activator protein (CAP; see Helix-Turn-Helix Motifs). The second domain consists of three subdomains. Subdomain I is a five-stranded antiparallel β-sandwich. Subdomain II has a Rossman fold consisting of a four-stranded parallel β-sheet flanked on each face by a pair of α helices ribose (see The Classical Dinucleotide-Binding Fold for a discussion of the Rossman fold). The last subdomain is an α/β structure comprising five α helices and a small two-stranded β-sheet. Apart from the winged helix motif, the Rossman fold, and a large central channel, this dimer is quite different from the known topoisomerase type I

and II structures (see above). The only noted similarity is in the Rossman fold.

Polynucleotidyl Transferase RNase H–Like Fold

The structures of a number of enzymes involved in polynucleotide transfer have been observed to have the same basic fold as ribonuclease H (RNaseH; Katayanagi et al., 1990, 1992; Yang et al., 1990). Members of this family include RNase H, retroviral integrase, Holiday junction RuvC resolvase, TN5 transposase, the RNase H domain of retroviral reverse transcriptase, a core fragment of bacteriaphage Mu transposase, and the 3′-5′ exonuclease domains from several DNA polymerase structures, including the Klenow fragment (Ollis et al., 1985; Davies et al., 1991, 2000; Ariyoshi et al., 1994; Dyda et al., 1994; Rice and Mizuuchi, 1995). The sequence identity among family members is generally <15%. At the center of the structure is a five-stranded β-sheet, aligned with a strand order of 3-2-1-4-5. The second strand is antiparallel to the other four strands (Fig. 17.1.31). The number and location of the helices are less conserved than the central β-sheet. Helices A and E are topologically more conserved than the other helices. They are located between strands 3 and 4, and at the C-terminal of strand 5 in sequence, primarily interacting with the central β-sheet. The number of helices between strands 4 and 5 varies, as do the helices after helix E (Fig. 17.1.31). The family can be further divided into RNase- (e.g., HIV-1 reverse transcriptase RNase H, *E. coli* RNase H, Archeon Class II RNase H) and integrase-like structures (e.g., integrase, Mu A transposase, RuvC) with greater structure resemblance observed between the members of each subclass. Limited structural homology, particularly among the central five-stranded β-sheet, is also observed with Mut S domain II, actin-like ATPase domains, and ribosomal proteins L18 and S11.

PROTEINS INVOLVED IN CYTOSKELETON AND MUSCLE MOVEMENTS

The Actin Fold

The three-dimensional structures of several ATP-binding proteins, including actin (Kabsch et al., 1990), the N-terminal domain of heat-shock cognate protein (hsc70; Flaherty et al., 1990), hexokinase (Anderson et al., 1978), glycerol kinase (Hurley et al., 1993), and acetate kinase (Buss et al., 2001), have been found to contain a conserved structural motif referred to as the actin fold, despite the absence of any significant sequence homology (Table 17.1.13). This fold consists of two α/β domains connected by a hinge region with the ATP binding site between them (Fig. 17.1.32A-D; Kabsch and Holmes, 1995). Each domain can be further divided into two subdomains. Two of the four subdomains (subdomains 1 and 3 according to actin notation) share a conserved core, a five-stranded β-sheet (with only one antiparallel strand) that is assembled as a β-meander surrounded by three helices. ATP binds at the interface between the β-sheets of subdomains 1 and 3. These two β-sheets are connected by two conserved helices and have identical topology, suggesting a gene duplication event. The other subdomains (2 and 4) are not as well conserved among these proteins, and may carry out specialized functions in different proteins sharing this fold. In actin, hsc70, and glycerol kinase, subdomains 2 and 4 are each α/β regions positioned at the C-terminal end of the sheets of the actin fold. Subdomain 4 has identical topology in the above three proteins, but a different tertiary structure in glycerol kinase (Hurley et al., 1993). Hexokinase only has subdomain 4 (subdomain 2 is absent), which is an α-helical bundle.

Actin-Binding Proteins

The assembly of actin to form filament, where it is the major component, and cortex is regulated by several actin-binding proteins, including profilin, gelsolin, severin, and villin. These molecules, through their association with actin, not only regulate cellular locomotion and shape, but also function as signaling molecules and keep actin filament in a dynamic balance between the free monomeric and polymerized forms. For example, profilin inhibits the assembly of F-actin by sequestering monomeric actin molecules, whereas gelsolin controls the actin cortex formation through fragmentation of F-actin filaments. The three-dimensional structures of several of these actin-binding proteins has revealed their folds and helped to elucidate their functions in filament formation.

Profilin

Both the X-ray crystal and NMR solution structure of profilin determined to date show a single domain fold with a central five-stranded antiparallel β-sheet pack against four α-helices, two on each side of the sheet (Fig. 17.1.32E; Schutt et al., 1993). The β-sheet is curled toward the N-terminal α-helices on one side and

the five strands are connected in the order of 1-2-3-4-5. Profilin, when complexed with β-actin, forms two major contacts with its third helix, as well as the amino-terminal portion of the fourth helix and strands 4, 5, and 6.

Gelsolin

The members of the gelsolin superfamily include gelsolin, villin, and severin. The structure of the segment 1 domain of gelsolin shows the basic fold, which consists of a central five-stranded β-sheet flanked by two α-helices, one on each side of the sheet (Fig. 17.1.32F; McLaughlin et al., 1993). The central β-sheet contains mostly antiparallel strands with the exception of the last strand of the sheet, which is in parallel configuration.

Spectrin repeats

Spectrins are actin cross-linking proteins. Their associations with actin generate meshworks to support the plasma membrane and facilitate cellular interactions. The so-called spectrin repeat defines a segment of 100 to 120 amino acid and folds into a three-helix bundle (Fig. 17.1.32G; Yan et al., 1993). The observed repeats form a stable domain-swapped dimer in which the third helix of each monomer is exchanged in the helix bundle.

ENZYMES

The α/β Hydrolase Fold

A distinct protein fold, observed among several hydrolytic enzymes, has emerged in the last few years. These enzymes, including serine carboxypeptidases, microbial lipases, esterases, dehalogenases, peroxidases, and various small molecule-hydrolyzing enzymes, such as dienelactone hydrolase, share little or no detectable sequence homologies among them and differ widely in their catalytic functions (Table 17.1.14; Ollis et al., 1992; Nardini and Dijkstra, 1999). Nonetheless, they share the same structural topology and furthermore a conserved catalytic triad consisting of a nucleophile (Ser, Cys, or Asp) at the end of β-strand 5, an absolutely conserved His after β-strand 8, and an Asp or Glu residue at the end of β-strand 7 (Liao and Remington, 1990; Franken et al., 1991; Sussman et al., 1991; Noble et al., 1993; Nardini and Dijkstra, 1999). The prototype of this fold consists of an eight-stranded mostly parallel β-sheet connected by one or more helices between adjacent strands (Fig. 17.1.33A and B). The strands are in the order 1-2-4-3-5-6-7-8 with strand 2 antiparallel to the rest. While the central eight-stranded β-sheet is well conserved among these enzymes, the length and number of helices spaced between the adjacent strands are less conserved and their positions vary significantly from enzyme to enzyme. The length and location of the loops connecting the secondary structure elements are also highly variable among members of this family.

Sialidases

Sialidases and influenza neuraminidases form a family of glycohydrolases that cleave terminal α-ketosidically-linked sialic acids of many glycoproteins, glycolipids, and oligosaccharides. Structurally, they define a prototype with a β-propeller fold composed of six four-stranded antiparallel β-sheets arranged similar to a six-blade propeller (Fig. 17.1.33C; Varghese and Colman, 1991; Burmeister et al., 1992; Crennell et al., 1993; White et al., 1995). There is a six-fold pseudo-symmetry through the center of the subunit relating each of the six blades. The topological connections between the strands within each sheet and among the adjacent sheets are identical from blade to blade. Each β-sheet displays a typical right-handed twist with strands 1 to 4 running antiparallel from the center of the propeller to the edge. The connection between each adjacent blade is always through the fourth strand of the preceding sheet to the first strand of the connecting sheet. The active site is located near the center of the propeller (see Frequently Observed Secondary Structure Assemblies of Structural Motifs for further discussion).

The TIM-Barrel Fold

The TIM-barrel fold, also referred to as an α/β-barrel or $(\beta/\alpha)_8$ fold, is one of the most frequently observed catalytic folds in protein structures. This fold is used by a multitude of different enzymes with very diverse functions and received its name from triose phosphate isomerase (TIM), the protein in which it was first discovered (Banner et al., 1975). The three-dimensional structure of the TIM-barrel is a rather elegant looking eight-stranded parallel β-barrel surrounded by a ring of helices that are essentially antiparallel to the β-strands (Fig. 17.1.34A and B). Both the β-strands and the α-helices are inclined to the axis of the barrel. The classical topology of this structure is an array of eight tandem up-down β-α motifs which are aligned side-by-side in a parallel manner with consecutive strand order, eventually wrapping around to form a β-barrel (Fig.

17.1.34C). In some TIM-barrel structures, this array is not completely regular, containing instead small insertions in loop regions or, as in the cases of enolase and muconate lactonizing enzyme, missing a helix between the last two β-strands. To date all of the crystal structures in which the TIM-barrel has been found have been enzymes and in almost every case the active site is located within the loop regions at the C-terminal end of the barrel.

Protease Folds

The various proteases that have been examined structurally can be grouped into four primary classes based on characteristics of the active sites. First, serine proteases use an essential Ser as the active site nucleophile. Second, cysteine proteases require an active site Cys nucleophile. Third, aspartic proteases work primarily at low pH, utilizing two essential Asp side chains for catalysis. Finally, metalloproteases all contain a critical zinc atom coordinated at the active site. These four classes can be further subdivided into two or more different protein folds despite the fact that members of a particular class share nearly the same reaction mechanism. A fifth class that has been observed in proteasome complexes and which uses a Thr as the active site residue is also discussed in this section.

Serine proteases

Nearly all serine proteases contain the well known catalytic triad Ser-His-Asp in their active sites. The three side chains are positioned close together through tertiary structure and are hydrogen bonded to each other in a manner which causes the Ser hydroxyl oxygen to be a strong catalytic nucleophile. An exception is the cytomegalovirus protease which uses a similar Ser-His-His catalytic triad for the same purpose. Four completely different protein folds have been observed to date in this class of proteases. Each fold is described below.

Trypsin-like serine proteases. Trypsin-like proteases are folded into two nearly identical domains, each consisting of a six-stranded antiparallel β-barrel with a Greek key topology (Fig. 17.1.34D; Fischer et al., 1994). In many cases, an α-helix is associated with each domain. The active site is formed from loop regions in a cleft between the two domains. Numerous eukaryotic, prokaryotic, and viral serine proteases have been observed to utilize this fold. Some of the more well known trypsin-like proteases include thrombin, elastase, chymotrypsin, plasmin(ogen), urokinase, and complement C1R and C1S proteases. At least three types of viral cysteine proteases also have the trypsin-like fold (see below).

Subtilisns. The three-dimensional structures of subtilisns from prokaryotic and fungal organisms (Fig. 17.1.34E) show a common core folding pattern made up of a seven-stranded parallel β-sheet (strand order 2-3-1-4-5-6-7) sandwiched between two layers of α-helices (Murzin et al., 1995). The active site is located at the C-terminal ends of the central β-sheet strands with the essential Ser residue situated at the N-terminus of an α-helix. Additions to this core fold in some subtilisns include a small two- to four-stranded antiparallel sheet positioned at the C-terminal edge of the large β-sheet and a C-terminal β-ribbon hydrogen bonded to the main β-sheet (extending it by two strands).

Serine carboxypeptidases. Serine carboxypeptidases are members of a larger family of enzymes known as the α/β-hydrolases (see The α/β Hydrolase Fold). These serine proteases contain a central, eleven-stranded, mixed, mostly parallel β-sheet (three more strands than the canonical α/β-hydrolase fold) surrounded by fifteen helices on either side of the sheet (Fig. 17.1.34F). The β-sheet has a pronounced twist of almost 180°. As in the subtilisn class of proteases (see above), the active site is located at the C-terminal end of the β-sheet with the catalytic serine at the N-terminal end of a helix (Liao and Remington, 1990).

Herpes virus–like serine proteases. First observed in the structure of a serine protease from human cytomegalovirus (hCMV), the three-dimensional structures of these single-domain proteases consist of a seven-stranded mostly antiparallel β-barrel with one Greek key motif and seven helices surrounding approximately one half of the circumference of the barrel as well as both ends of the barrel (Qiu et al., 1996; Shieh et al., 1996; Tong et al., 1996). This fold has subsequently been observed in Herpes simplex virus type II serine protease (Hoog et al., 1997), Karposi's sarcoma-associated herpes virus protease (Reiling et al., 2000), and Varicella-Zoster viral protease (Qiu et al., 1997). The active site Ser-His-His triad is positioned on the outside of the barrel on the face without any helices. This group of proteases is unique among serine proteases in that the tertiary structure of its catalytic triad contains a conserved His residue in the same position that the Asp residue is normally found in the classical Ser-His-Asp triad. These proteins func-

tion physiologically as dimers with the active sites situated distally to each other.

Cysteine proteases

To date, three different subfamilies of cysteine proteases have been characterized, each of which utilize the sulfhydryl of a Cys residue as the active site nucleophile. The most well known and studied are the papain-like cysteine proteases which are sometimes confused with the entire cysteine protease family.

Papain-like cysteine proteases. The proteases of this subfamily can be recognized by the conserved catalytic triad Cys-His-Asn, in addition to a unique fold. Analogous to the Ser-His-Asp triad of serine proteases, this hydrogen bonded triad is also formed through tertiary interactions resulting in a nucleophilic sulfhydryl on the Cys. The basic papain-like fold consists of two distinct domains designated L and R (papain nomenclature; Fig. 17.1.35A). The L domain is a small three- to four-helix bundle. The R domain is comprised of a highly curved and twisted six-stranded antiparallel β-sheet flanked by one or two helices (Musil et al., 1991). The active site is in a V-shaped crevice between the two domains. Similar to the subtilisns and serine carboxypeptidases, the nucleophilic cysteine residue is located at the N-terminal end of an α-helix.

Although small structural variations of the described fold are observed in several papain-like proteases, bleomycin hydrolase from yeast has extensive structural modifications (Joshua-Tor et al., 1995): a small helical domain is inserted into loops in the L domain and an additional N-terminal helical domain is also present. This cysteine protease, which cleaves the anticancer peptide bleomycin, functions as a hexamer with a large central channel containing the six active sites. It also binds to double- and single-stranded DNA, and represses genes of the GAL system.

Caspases. Cysteine-dependent aspartate-specific proteases (caspases) are a family of structurally homologous cysteine endopeptidases that are involved in processing proinflammatory cytokines and in apoptotic signaling (Alnemri et al., 1996). The prototypic caspase structure, that of interleukin-1β-converting enzyme (ICE or caspase-1) is a tetramer made from two heterodimers related by a two-fold rotation axis. Each heterodimer (p10/p20) contains an independent active site and is derived from a 45-kDa precursor that has been cleaved in three locations. The fold of the heterodimer consists of a six-stranded mostly parallel β-sheet sandwiched between two helices on one face and four helices on the other (Fig. 17.1.35B; Walker et al., 1994; Wilson et al., 1994). The active site formed by loops at the C-terminal end of the β-sheet contains a catalytic Cys-His diad. The active site Cys is situated at the end of a β-strand instead of the N-terminus of an α-helix as in the papain subfamily. The crystal structures of several other caspases (i.e., caspase-3, -7, -8, and -9) have confirmed this unique fold and quaternary structure except for differences in substrate binding sites (see Wei et al., 2000; Chai et al., 2001; Renatus et al., 2001, and references therein). Although caspases are mammalian enzymes, the crystal structure of the catalytic domain of a bacterial cysteine protease, gingipain R from *Porphyromonas gingivalis*, has been observed to have a very similar fold and a Cys-His diad at the active site (Eichinger et al., 1999). The two subdomains within the single chain catalytic domain correspond to the small and large subunits of the caspase fold.

Viral trypsin-like cysteine proteases. Another set of cysteine proteases that have been characterized structurally include the 3C and 2A viral cysteine proteases (Allaire et al., 1994; Matthews et al., 1994; Petersen et al., 1999). They use a Cys-His-Glu/Asp catalytic triad and surprisingly have the same fold as the trypsin-like serine proteases (see above). The coronavirus cysteine protease also shares this fold but has an extra helical domain and a Val instead of Asp or Glu in the third position of the canonical catalytic triad (Anand et al., 2002).

Aspartic proteases

The aspartic proteases (also known as acid proteases) contain primarily β-sheets and utilize two Asp residues to catalyze proteolysis at low pH. They are divided into two groups, the two-domain pepsin-like proteases found in eukaryotes and the dimeric retroviral proteases. Both groups contain a conserved catalytic triad sequence Asp-Thr/Ser-Gly at the active site and limited regions of sequence conservation are shared between them.

Pepsin-like proteases. The pepsin-like proteases are comprised of two pseudosymmetric domains, each containing a partially opened, six-stranded, antiparallel, distorted β-barrel structure with a few short helices within the loop regions (Fig. 17.1.35C; Davies, 1990). Between the domains and away from the center of the structure, a twisted six-stranded antiparallel β-sheet lays against the surface of the protein as if to form a base on which to set the

molecule. Three strands from each domain contribute to the sheet. On the side of the protein opposite the six-stranded sheet, a large cleft runs across the domain interface forming the site for the polypeptide substrate to bind. The two catalytic Asp residues (one from each domain) are located at the bottom of the active site cleft (Fig. 17.1.35C). A long hairpin loop from the N-terminal domain forms a flap which projects over the top of the active site shielding it from solvent. When pepsin-like proteases bind an inhibitor, the flap also participates in binding by clamping down over the inhibitor.

Retroviral proteases. Retroviral proteases are homodimers that cleave newly synthesized viral polyproteins into individual functional units during viral maturation. The two identical subunits are structurally very analogous to the two domains of the pepsin-like proteases. Each is comprised of a six-stranded antiparallel β-barrel, a single helix located at one end of the barrel, and a two-stranded β-sheet at the C-terminal end of the helix (Fig. 17.1.35D; Navia et al., 1989; Davies, 1990). In the dimer, the two small β-sheets hydrogen bond to form a four-stranded β-sheet analogous to the six-stranded sheet observed in pepsin-like proteases. Two conserved, catalytic Asp residues (one from each monomer) are positioned at the bottom of the active site cleft at the subunits interface. A flap from each subunit extends over the top of the active site in a manner structurally and functionally resembling the flap of the pepsin-like enzymes.

Metalloproteases

There are two main superfamilies of zinc-containing metalloproteases, each of which extend across both prokaryotes and eukaryotes: the zinc-dependent endopeptidases, also referred to as zincins (Bode et al., 1993), and the zinc-dependent exopeptidases.

Zincins. The conserved tertiary structure of the catalytic domain of zincins consists of a five-stranded mostly parallel β-sheet with two helices packed against the concave side Fig.17.1.35E; Bode et al., 1993; Gomis-Ruth et al., 1994). This fold is generally observed in the N-terminal domains of zincins. The conserved His-Glu-X_2-His sequence motif identifies three critical active site residues. These three residues are in a helix situated at the edge of the β-sheet with the two histidines coordinating the catalytic zinc and Glu, polarizing a zinc-bound water that acts as a nucleophile during catalysis. Several zincins also bind multiple Ca^{2+} ions.

Zincins comprise two distinct families known as the thermolysin-like proteases and the metzincins. The known tertiary structures of thermolysin-like proteases have a small irregular twisted β-sheet structure at one end of the central β-sheet and a C-terminal helical domain with the active site in a crevice between the two domains. A conserved glutamate residue provides the fourth zinc ligand. The metzincins, which include astacins, adamalysins, matrix metalloproteases, and serralysins have an extended consensus sequence (His-Glu-X_2-His-X_2-Gly-X_2-His), including a third zinc-coordinated histidine. The name metzincins comes in part from a unique structural characteristic of this family referred to as a Met-turn, which is a methionine-containing β-turn located just beneath the zinc atom (Bode et al., 1993). It is noteworthy that one metzincin, an alkaline protease of the serralysins, has a particularly unusual C-terminal domain consisting of a calcium-binding 21-strand β-sandwich, also referred to as a right-handed parallel β-roll (Fig. 17.1.35G; Baumann et al., 1993).

Zinc-dependent exopeptidases. Zinc-dependent exopeptidases share a common catalytic fold consisting of an α/β/α sandwich. The central eight-stranded β-sheet (strand order 1-2-4-3-8-5-7-6) is mostly parallel (only strands 2 and 8 are antiparallel) and hosts the active site at the middle of its C-terminal end (Fig. 17.1.35F; Burley et al., 1992). Carboxypeptidases A, B, and T each have one active site zinc atom pentacoordinated to two His, a Glu (both oxygens are ligands), and a water molecule. The aminopeptidases in this family, however, each have two zinc atoms coordinated less than 3.5 Å apart at the active site. As with endopeptidases, zinc-bound water in this exopeptidase family is considered a possible nucleophilic catalyst (Strater et al., 1995).

Ntn hydrolase fold

The N-terminal nucleophile (Ntn) hydrolase fold is used by eukaryotic and archael proteasomes or proteasome homologues in prokaryotes (e.g., Hs1V; Table 17.1.14; Bochtler et al., 1999). This fold has an αββα core structure consisting of a ten-stranded antiparallel β-sandwich flanked on each face by helices (Fig. 17.1.35H; Brannigan et al., 1995). The strand order for one sheet is 3-4-5-6-7 and the other is 8-1-2-9-10. A catalytic triad of Thr, Glu/Asp, and Lys forms at least part of the active site at one end of the β-sandwich and is highlighted by an essential N-terminal threonine nucleophile (Groll et al., 1997). Variations

of this fold (α- and β-subunits) are used to make large cylinder-shaped multisubunit complexes known as proteasomes or proteasome homologues that are responsible for protein degradation (see Proteasome for further discussion). Only certain β-subunits are catalytic and none of the α-subunits are. This same fold is also found in several other enzymes, such as glutamine PRPP aminotransferase, glucosamine 6-phosphate synthase, and various penicillin acylases, to name a few (Oinonen and Rouvinen, 2000). Sequence homology is generally not conserved between these enzymes, nor is the identity of the catalytic nucleophile.

Enzymes Involved in Ubiquitination

Ubiquitin-targeted protein degradation plays an important role in protein degradation by proteasomes, in DNA repairs, and in selected signal transduction pathways (Hershko and Ciechanover, 1998). A ubiquitin pathway involves three distinct steps. (1) E1-catalyzed ubiquitin activation which attaches the C-terminal Gly residue of ubiquitin to a Cys residue on E1 through a thioester linkage. (2) E2s-mediated ubiquitin conjugation that transfers the activated ubiquitin from E1 to the active site Cys residue of E2s. (3) Ligation of ubiquitin to substrate proteins by E3 ubiquitin ligases. Ubiquitin, a conserved protein of 76 amino acids, and all ubiquitin-like proteins fold into small αα/β structure with a five-stranded antiparallel β-sheet wrapped around an α-helix (sometimes referred to as a β-grasp; Fig. 17.1.36A; Vijay-Kumar et al., 1987; Rao-Naik et al., 1998). While there is usually one E1, there are multiple and diverse E2 and E3 enzymes, each specific for different protein substrates.

Ubiquitin-conjugating enzymes

The structures of ubiquitin-conjugating E2 enzymes solved to date include UbcH7, Ubc9, and a Mms2-Ubc13 heterodimer complex (Bernier-Villamor et al., 2002; Moraes et al., 2001; VanDemark et al., 2001; Zheng et al., 2000). The catalytic core of E2 enzymes consists of ~160 amino acids that fold into an α/β structure with its central four- to seven-stranded antiparallel β-sheet flanked by four α-helices (Fig. 17.1.36B and C). The active site cysteine, which catalyzes the transfer of ubiquitin from E1 through a thioester linkage to E2, is located in a loop between the fourth β-strand and the second α-helix.

Ubiquitin ligases

The ubiquitin ligases (E3s) are a very diverse family of proteins whose function is to bring protein substrates into proximity with ubiquitin-conjugated E2 enzymes. The structural diversity of ubiquitin-ligases presumably reflects the need for specificity among diverse protein substrates. Common among E3s is their ability to bind E2 and specific protein substrates. There are two types of E3 enzymes, which differ in their ubiquitin transfer mechanisms. The HECT-type (homology to E6AP binding protein C terminus) ubiquitin ligases catalyze the transfer of ubiquitin by first forming an E3-ubiquitin thioester intermediate. The E6AP-Ubc7 complex is an example of an HECT-type E3 enzyme. In contrast, the RING-family ubiquitin ligases do not form an E3-ubiquitin intermediate. They all have a RING domain that interacts with the E2-ubiquitin conjugate. Examples of RING-type ubiquitin ligases include c-Cbl (Fig. 17.1.36B; Zheng et al., 2000), members of the seven in absentia homolog (Siah) family (Fig. 17.1.36D; Polekhina et al., 2002), and the SCF (Skp1-Cullin-F-box) family of E3 ligases (Fig. 17.1.36E; Zheng et al., 2002). The SCF members constitute the largest and most complex family of protein ligases and ubiquitinate a rather diverse range of proteins. The SCF complex consists of a protein-substrate-interacting subunit, Skp2, (Schulman et al., 2000), an elongated two-domain cullin subunit (Cul1), a connecting unit, Skp1, that links the Skp2 F-box–containing protein with cullin, and a RING domain nested on cullin (Fig. 17.1.36E). Human Skp2 has an ~100 residue N-terminal domain of unknown structure, followed by an F-box, ten leucine-rich repeats (LRR; see Leucine-Rich Repeats for a discussion of the LRR fold), and a 30-residue tail with little secondary structure (Schulman et al., 2000). The F-box is an ~40-residue domain comprising three helices, with one helix packing orthogonally against a pair of antiparallel helices. The N-terminal domain of Skp1 has an α/β fold consisting of a three-stranded β-sheet packed on one face against a bundle of α-helices (also known as a BTB/POZ fold) and an additional two-helix extension (Schulman et al., 2000). The N-terminal domain of Cul1 consists of three cullin repeats, each comprising a bundle of five α-helices (Fig. 17.1.36F). The C-terminal domain of Cul1 is made of two winged helix motifs (see Helix-Turn-Helix Motifs), a four-helix bundle, and an α/β subdomain. The zinc-containing Rbx1 RING domain interacts closely with the C-ter-

minal domain of Cul1 and contributes one β-strand to the five-stranded β-sheet of the α/β subdomain.

Small ubiquitin-like modifiers

In addition to the more well known E2-E3 system, there is a family of small ubiquitin-like modifiers (SUMO), that regulate nuclear transport, stress response, and signal transduction controlling cell-cycle progression. The E2 enzymes for SUMO, as shown in the structure of a Ubc9-RanGAP1 complex (Bernier-Villamor et al., 2002), can directly recognize and modify Lys residues on their protein targets without additional E3s.

ELECTRON TRANSFER PROTEINS

There are several superfamilies of proteins involved in processes of electron transfer. Most of them have a heme prosthetic group or porphyrin ring central to electron transfer coordinated in a hydrophobic pocket; however, their tertiary folds vary from the four-helix-bundle fold of ferritin to the mostly helical (multiple α-helices and some β-strands) structure of cytochrome P450, to the typical cytochrome c fold, with the polypeptide chain organized into a series of α-helices and reverse turns. Below are summaries of a few commonly observed folds in this category (Table 17.1.15). Another group of electron transfer proteins, integral membrane respiratory proteins, is discussed below (see Respiratory Enzymes).

Cytochrome P450

Cytochrome P450 catalyzes the mono-oxygenation of a number of organic molecules in a heme-dependent reaction mechanism. There are now at least eight different types of P450 that have been examined structurally. These include $P450_{cam}$, $P450_{eryF}$, $P450_{BM-3}$, $P450_{terp}$, $P450_{NOR}$, $P450_{2C5}$, P450 sterol 14α-demethylase, and CYP119, each utilizing very different substrates. The sequence homology remains relatively high within each group but relatively weak between them. Despite low sequence homologies (<20%) and very diverse substrate specificities, a common fold is maintained among all the P450 cytochromes. The molecule folds into two domains, α and β (Fig. 17.1.37A). The α domain consists of a right-handed four-helix bundle sandwiched between two clusters of helices, a top cluster of five helices, and a bottom cluster of three helices with two β-strands. The β domain consists of a β-barrel surrounded by three helices. The β-barrel is formed by two β-sheets, a two-stranded β-sheet, and a five-stranded β-sheet, folded against each other (Raag et al., 1993). The heme molecule is embedded between the two domains.

Cytochrome c Domain

Cytochrome c is often observed as an electron transfer domain in many redox enzymes. Members of this family include cytochromes c, c2, c5, c551, c553, and c555. Other proteins, such as the N-terminal domain of cytochrome cd1-nitrite reductase, flavocytochrome c sulfide dehydrogenase (FCSD), and di-haem cytochrome c peroxidase, also contain cytochrome c–type domains in their structures. Most cytochrome c proteins contain one cytochrome c domain, but some enzymes, including FCSD and di-haem cytochrome c peroxidase, contain two. The typical cytochrome c fold consists of three long α-helices and a pair of β-strands packed in the core region, with two additional short helices surrounding the core (Fig. 17.1.37B). The length of the helices, strands, and loop conformations vary considerably among the members of the family. The heme group is coordinated by residues from both helices and reverse turns in the core region (Louie and Brayer, 1990).

Cytochrome c3

The structure of cytochrome c3 from *Desulfovibrio desulfuricans* consists of three to four helices (often one long helix amidst other shorter helices), a pair of β-hairpins, and three to four long irregular loops (Fig. 17.1.37C). There are four heme molecules ligated to the ~110 residues of cytochrome c3. The heme coordination frame is formed by packing the long helix against the β-hairpins and the loop regions.

Cytochrome b5

In addition to cytochrome b5, the N-terminal domain of flavocytochrome b2 also belongs to this fold family. The structure folds into a α/β/α sandwich with a central five-stranded mixed β-sheet surrounded by six short α helices, two on the concave side of the β-sheet and four on the other (Fig. 17.1.37D; Mathews et al., 1972). The heme is coordinated by the four helices and the central two strands of the β-sheet. The five strands are hydrogen-bonded in a 1-5-3-2-4 order in cytochrome b5 and 5-3-1-2-4 order in the cytochrome b5–type domain of flavocytochrome b2.

Four-Helix Bundle Cytochromes

There are at least three families of cytochromes known to assume the fold of four-helix bundles: cytochrome b562, cytochrome c′, and ferritin. Among them, cytochrome b562 and cytochrome c′ share a close resemblance, both forming up-down-up-down helical bundles with a left-handed twist (Fig. 17.1.37E). The heme is inserted between helices 1 and 4, with helix 3 forming the floor. The structure of ferritin, on the other hand, forms an up-down-down-up four-helix bundle with left-handed twist. There is a short one-turn helix at the C-terminal of the four-helix bundle. Unlike cytochrome c′ or b562, ferritin forms a dimer with only one heme molecule inserted at the interface of the two helical bundle subunits (Fig. 17.1.37F; Frolow et al., 1994).

GLOBIN-LIKE PROTEINS

The helical globin-like proteins encompass three different families: globins, phycocyanins, and colicin. The core fold for all of these proteins consists of an α-helical sandwich which is simply two layers of α-helices stacked against each other. Using the globin nomenclature, helices A, E, and F make up one side of the sandwich, while helices B, G, and H make up the other, with helices on each side essentially orthogonal to those on the other (Table 17.1.16, Fig. 17.1.38A). Various members of these families are all completely α-helical (with the exception of flavohemoglobin), with minor differences in helix packing due to the binding of associated cofactors and insertions of helices to the basic fold (Holm and Sander, 1993; Ermler et al., 1995).

Globins

The globin family is made up of several heme-containing helical proteins used specifically for oxygen storage and transport within bacteria, insects, vertebrates, and plants. Myoglobin is the smallest of these proteins, consisting of the basic globin fold with the addition of two more short helices (C and D; Fig. 17.1.38A; Phillips, 1980). This monomeric protein binds oxygen for storage in muscle using a heme group located in a cleft between helices E and F. Mammalian hemoglobin is composed of four separate globin chains ($\alpha_2\beta_2$) in an approximately tetrahedral arrangement (Fig. 17.1.38B; Fermi et al., 1984). Each chain binds a heme group in the same location and orientation as in myoglobin. The binding of oxygen induces changes in the quaternary structure of hemoglobin such that four oxygen molecules are bound in a cooperative manner, making it well suited for oxygen transport. Other nonmammalian hemoglobins are found to have very different oligomeric states. Pig roundworm hemoglobin, for example, is an octamer of a two-domain globin fold (Kloek et al., 1993), while sea cucumber hemoglobin can be present in both monomeric and dimeric states under physiological conditions. Flavohemoglobin is a rather unusual globin. Unlike the other known globin structures, flavohemoglobin is an α/β three-domain protein containing an FAD-binding domain and an NAD-binding domain as well as an N-terminal globin domain (Ermler et al., 1995). The middle FAD-binding domain is a six-stranded antiparallel β-barrel with Greek key topology and the C-terminal NAD-binding domain is a classical dinucleotide binding domain with a five-stranded parallel β-sheet. Both redox domains share significant structural homology with both domains of ferridoxin reductase (Karplus et al., 1991). The physiological function of flavohemoglobin remains unknown.

Phycocyanins

Phycocyanins make up part of the light-harvesting antennae in cyanobacteria and red algae. The three-dimensional structure of this protein has the six helices of the globin fold with two additional antiparallel α-helices protruding from the N-terminus (Fig. 17.1.38C; Schirmer et al., 1987). A phycocyanobilin cofactor is located in the same place as the heme group of the globin structures (between helices E and F).

Colicins

Colicins are a class of multidomain antibacterial toxins produced by *E. coli* that participate in killing target cells by binding to surface receptors and forming pores in the cytoplasmic membrane. The crystal structure of a C-terminal fragment of colicin A is observed to have a canonical globin fold with three additional N-terminal helices and an extra C-terminal helix (Fig. 17.1.38D; Parker et al., 1989).

TOXINS

The structures of toxins represent a collection of a diverse family of folds. Some toxins, such as scorpion and snake toxins, are relatively small, contain few well defined secondary structure elements, and are rich in disulfides, while others, such as diphtheria toxin and heat-labile toxin, have well defined secondary structures and often form complicated mutidomain

oligomeric organizations. Numerous toxin structures have been determined by X-ray crystallography or NMR, and most belong to one of several distinct structural folds summarized in this section (Table 17.1.17).

Ribosome-Inactivating Toxins

The A chains of ricin and abrin as well as the single-chain toxins gelonin, α-momorcharin, α-trichosanthin, and shiga toxin are members of the ribosome-inactivating protein (RIP) fold family, named for their ability to block protein synthesis. Figure 17.1.39D illustrates the structure of the ricin A chain (Katzin et al., 1991). It contains mixed β-sheets and several α-helices. The B chain of ricin and abrin each form a β-trefoil type structure (see β-Trefoils and Table 17.1.9). There are two domains of such trefoils in the B chain of each toxin (Rutenber and Robertus, 1991).

ADP Ribosylation Toxins

Diphtheria toxin (DT), *Pseudomonas aeruginosa* exotoxin A (ETA), cholera toxin, heat-labile enterotoxin, and pertussis toxin all contain ADP ribosylation domains which enable these toxins to inhibit protein synthesis by catalyzing the ADP-ribosylation (and thereby the inactivation) of essential cellular proteins. The basic fold of an ADP ribosylation domain, shown in Fig. 17.1.39A, consists of a mixed structure of seven or eight β-strands and several short helices. The β-strands are grouped into two β-sheets that are packed against each other forming the core of this domain.

Diphtheria toxin

Diphtheria toxin (DT) and *Pseudomonas* exotoxin A (ETA) structures are each organized into three domains: catalytic ADP ribosylation, transmembrane, and receptor-binding (Allured et al., 1986; Choe et al., 1992). The sequential order of the three domains is from N- to C-terminal in DT and from C- to N- terminal in ETA. The transmembrane domain in both DT and ETA is all α-helical. The receptor-binding domains are both immunoglobulin-like; however, the receptor-binding domain of DT resembles an immunoglobulin V domain (see above) whereas that of ETA resembles a Con A lectin domain (see above).

Bacterial AB5 toxins

Cholera toxin, pertussis toxin, and heat-labile enterotoxin belong to a family of bacterial AB5 toxins. All of them contain an ADP ribosylation domain (domain A; see above) and a pentameric receptor-binding B domain (Sixma et al., 1991; Swaminathan et al., 1992; Stein et al., 1994). The topology of the B domain, shown in Fig. 17.1.39B, contains a pentameric ring structure of layered helices-sheets-helices. The pentameric repeat unit is an α/β/α sandwich that consists of a one-turn α-helix, six-stranded antiparallel β-sheet and long α-helix.

Superantigen toxins

The three-dimensional structures of staphylococcal enterotoxins A (SEA), B (SEB), C2 (SEC2), and a toxic shock syndrome toxin 1 (TSST-1) define a family of toxins that function as superantigens to T cells (Swaminathan et al., 1992, 1995; Prasad et al., 1993). The common fold of these toxins is a two-domain structure (Fig. 17.1.39C). The N-terminal domain folds with a Greek key topology and consists primarily of a five-stranded β barrel with strands in the order 1-2-3-5-4. It is also a member of the oligonucleotide/oligosaccaride binding (OB) fold (see The OB Fold; Murzin, 1993). The C-terminal domain of these toxins folds into a modified β-grasp motif. The primary feature of this domain is a long α-helix packed against a five-stranded mixed β-sheet.

Small Single-Subunit Toxins

Scorpion toxins

Members of scorpion toxins are usually small proteins of ~50 amino acids strengthened by disulfide bonds. They include scorpion margatoxin, noxiustoxin, charybdotoxin, scyllatoxin, agitoxin, chlorootoxin, toxin I5a, kaliotoxin, and others. The basic topology consists of a N-terminal α-helix packed against a three-stranded antiparallel β-sheet (Fig. 17.1.39E; Johnson et al., 1994). The fold is stabilized by three sets of disulfide bonds connected in the sequential order of Cys I–IV, II–V, and III–VI. Cys II and III usually follow the CxxxC pattern whereas Cys V and VI usually follow the CxC pattern. Some members of the family, referred to as long-chain scorpion toxins, have an additional loop structure at the C-terminal end and an additional disulfide bond (Housset et al., 1994). The rest of the family is known as the short-chain scorpion toxins.

Snake toxins

Most snake toxins have similar structures, known as the snake-toxin fold. Members with known structures include sea snake erabutoxin A and B, neurotoxin I and II, cobratoxin I and

II, cardiotoxin I, II, IIB, III and V, and bungarotoxin. Their structure has two antiparallel β-sheets, an N-terminal two-stranded β-sheet, and a C-terminal three-stranded β-sheet (Fig. 17.1.39F; Tsernoglou et al., 1978). There are four disulfides with the sequential bonding of Cys I–III, II–IV, V–VI, and VII–VIII in each structure.

Spider toxins

The family of spider toxins include ω-conotoxins from sea snail and agatoxins from web spider. These are very small proteins, <40 residues in sequence, and are stabilized by Cys bridges. The structure consists of largely irregular loops with only two short β-strands forming an antiparallel β-sheet at the center of the fold (Fig. 17.1.39G; Reily et al., 1995). The three common disulfides are connected with the sequence order of Cys I–IV, II–V, and III–VI. Members of agatoxins have one more disulfide near the C-terminal end.

Anthrax Toxin

Anthrax toxin is the major virulent factor of *Bacillus anthracis*. Owing to its high mortality when the infection is untreated, and its ability to exist in an aerosol form, it poses a threat to public health as a potential reagent for biological warfare and terrorism. The toxin consists of three proteins, lethal factor (LF),edema factor (EF), and protective antigen (PA). Among them, the lethal factor appears to be a protease that targets the mitogen-activated protein kinase kinase (MAPKK) family of proteins.edema factor is an adenylate cyclase and PA forms a membrane pore-like transporter once activated by furin-like cellular proteases. The crystal structures of all three factors have been solved (Table 17.1.17; Petosa et al., 1997; Pannifer et al., 2001; Drum et al., 2002; Shen et al., 2002).

Lethal factor

Lethal factor (LF), with a molecular weight of 90 kDa, consists of four domains. Domains I and IV are structurally similar, each with a set of nine to twelve helices packed against a four-stranded β-sheet (Fig. 17.1.40A). The function of domain I appears to be to bind protective antigen (PA; see below) whereas domain IV appears to be the catalytic domain of LF. The α/β part of domain IV loosely resembles a thermolysin-like zincin (see Protease Folds) with a Zn^{2+} coordinated within a His-Glu-X2-His motif. Domain I has no zinc-binding site. Domain II has an α/β fold similar to ADP ribosylation toxins (see ADP Ribosylation Toxins), however, a critical Glu is not conserved, suggesting this domain does not have ADP-ribosylating activity (Pannifer et al., 2001). Domain III is a small α-helical bundle. The N-terminal peptide of MAPKK is bound at the interface between domains III and IV.

Oedema factor

Oedema factor (EF) is a calmodulin (CaM)-dependent adenylate cyclase, whose activity increases 1000-fold upon activation by CaM. The structures of both the catalytic portion of EF alone and EF in complex with CaM have been determined (Drum et al., 2002; Shen et al., 2002). The fold of EF consists of a largely α-helical domain, three switch regions, and two α/β catalytic domains, C_A and C_B (Fig. 17.1.40B). Domains C_A and C_B have folds similar to the palm domains of DNA polymerase β and RNA poly(A) polymerase (see DNA Polymerases and Small RNA Polymerases). The binding of CaM induces large conformational changes in the association of the helical domain with respect to the switch and catalytic domains (Fig. 17.1.40C).

Protective antigen

Protective antigen (PA), an 83-kDa protein, consists of four domains, I to IV (Fig. 17.1.40D). Domain I contains a jelly-roll-like β-sandwich and two calcium ions coordinated by residues in an EF-hand motif. Domain I has the furin cleavage site and, once cleaved, PA forms a heptameric membrane pore transporter. Domain II folds into a modified Greek-key-like β-barrel. Domain III adopts the same topology as ferredoxins and the domain A of the toxic-shock-syndrome toxin-1 (TSST-1). Domain IV contains an immunoglobulin-like fold (see above) and is presumably responsible for host cell binding of PA.

LIPID-BINDING PROTEINS

There are numerous different protein folds involved in binding to lipids or lipid-like molecules, most with β-sheet structures. Three well known families of lipid-binding proteins are discussed in this section: lipocalins, fatty acid-binding proteins, and serum albumins. See Modular Domains Involved in Signal Transduction for a discussion of other phospholipid-binding proteins.

Lipocalins

Lipocalins (or retinol-binding protein-like proteins) comprise a family of mostly extracel-

lular proteins 160 to 180 residues in length, typically involved in the transport of small hydrophobic molecules (Flower et al., 2000). Examples include retinol-binding protein, β-lactoglobulin, bilin-binding protein, and ordorant-binding protein (Table 17.1.18). The crystal structure of retinol binding protein (Cowan et al., 1990) and subsequent structures reveals an eight-stranded antiparallel β-barrel with a β-meander topology (Fig. 17.1.41A; also see Frequently Observed Secondary Strucutre Assemblies or Structural Motifs). A short helix is sometimes located before the first β-strand and after the last β-strand of the β-barrel. Ligands bind inside the barrel.

Fatty Acid–Binding Proteins

Fatty acid binding proteins (FABPs) constitute a family of small, predominantly cytoplasmic lipid-binding proteins. They include myelin P2 protein, adipocytes lipid-binding protein, cellular retinol-binding protein (CRBP), and numerous tissue-specific forms of FABP. The fold comprises a ten-stranded antiparallel β-sheet with a pair of α-helices between the first and second β-strands. Like lipocalins, the topology is that of a β-meander. The β-sheet folds in half to form what is often referred to as a β-clam (Fig 17.1.41B). In the resulting structure the β-strands in each half of the clam shell are nearly orthogonal to each other and a ligand-binding cavity is created between them.

Serum Albumin

Serum albumin is a remarkably abundant fatty acid transport protein in the circulatory system. In humans, concentrations of 40 to 50 mg/ml are maintained in the blood. Despite extensive characterization, the atomic three-dimensional structure of serum albumin remained elusive until the early 1990s. Crystal structures of human and horse forms reveal a mostly helical heart-shaped protein composed of three structurally homologous domains (I to III; Fig. 17.1.41C; He and Carter, 1992; Ho et al., 1993; Curry et al., 1998; Sugio et al., 1999). Each domain has ten helices and is divided into two subdomains (a and b) with six and four helices, respectively. Furthermore, the first four helices in subdomain a are structurally homologous to the four helices in subdomain b. It is noteworthy that serum albumin contains 35 cysteine residues resulting in 17 disulfide bridges and one free cysteine. Despite the striking symmetry within the structure of serum albumin, its five fatty acid binding sites are located asymmetrically around the molecule (Curry et al., 1998): domain I has two binding sites, domain II has one binding site (shared with domain I), and β domain III has three binding sites (Fig. 17.1.41C).

LARGE MULTISUBUNIT PROTEINS

Chaperonin Proteins: GroES and GroEL

Chaperonins are a class of proteins that assist in catalyzing the folding of other proteins. The GroEL/GroES complex represents one of the best characterized chaperonin systems and both components are required for proper protein folding in bacteria (Xu et al., 1997b). Crystal structures of GroEL, GroES, an asymmetric GroEL/GroES complex, and other homologous chaperonin complexes have been determined by several groups (Table 17.1.19). GroEL (L for large) forms the cylindrical core of the GroEL/GroES complex and consists of fourteen identical subunits arranged as two heptameric rings stacked back to back (Fig 17.1.42A). Each individual protomer is comprised of three distinct domains: equatorial, intermediate, and apical (Fig 17.1.42B and C). The equatorial domain consists primarily of six long α-helices, two short helices, and three pairs of short β-strands. This is also the ATPase domain of GroEL. The intermediate domain contains a three-stranded β-sheet and three α-helices. The apical domain consists of a mixture of five α-helices and a seven-stranded β-sandwich. Structurally, the equatorial domain provides most of the intersubunit contacts within each heptameric ring and all the contacts between the two back-to-back stacked heptameric rings. The intermediate domain appears to bridge between the equatorial and the apical domains. Most of the contacts between GroEL and GroES are mediated through the apical domain.

GroES (S for small; also known as chaperonin-10) is a single ring of seven identical subunits that forms a dome over one end of the GroEL cylinder (Fig 17.1.42A). Each individual subunit folds into a five-stranded antiparallel β-barrel with four more associated β-strands and a long mobile loop (Fig. 17.1.42D). The end of the barrel near the long loop interacts with the GroEL subunits.

In the GroEL complex alone, both heptameric rings are essentially identical; however, when complexed with GroES and ADP at one end, the *cis* ring (i.e., the one closest to GroES and bound to ADP) of GroEL has dramatically

different domain orientations than the *trans* ring (see Fig. 17.1.42B,C). During the cycles of protein folding, the GroEL/GroES complex is proposed to undergo significant conformational changes, including the dissociation and reassociation of the GroES cap (Keskin et al., 2002).

Myosin Subfragment-1

The crystal structure of myosin subfragment-1 (Table 17.1.19) illustrates the overall arrangement of myosin heavy chain, regulatory light chain, and essential light chain (Houdusse et al., 1999; Rayment et al., 1993). The heavy chain forms much of the myosin head as well as an ~85-Å-long coiled-coil α-helix (Fig. 17.1.42E). The myosin head consists of mostly α-helices with the actin-binding site located near the tip. Both the regulatory and essential light chain subunits bind to the long coiled-coil helical tail of the heavy chain. The structure of the regulatory light chain resembles that of calmodulin with a divalent cation binding site in its first EF-hand region (see above). The essential light chain consists mostly helical elements that wrap around the heavy chain helical tail.

G Proteins and Regulators of G-Proteins

Small G proteins and their activators

Ras and RasGAP. As a prototype of small G proteins, the structure of p21Ras (Pai et al., 1990) reveals a variation of the mononucleotide binding fold (Fig. 17.1.43A; Table 17.1.19; also see The Classical Mononucleotide-Binding Fold). Like many small G proteins, the intrinsic GTPase activity of Ras is quite low and requires GTPase activation protein (GAP) to increase the enzymatic rate of GTP hydrolysis. The structure of p120GAP is all α-helical (Fig. 17.1.43B; Scheffzek et al., 1996). The structure of a complex between Ras and the Ras guanine-nucleotide-exchange-factor region of the son of sevenless (Sos) reveals a tight association between them that buries 3600 $Å^2$ of surface area. This results in the dissociation of the nucleotide from Ras due to alteration of the conformation of the switch 1 and 2 regions (Boriack-Sjodin et al., 1998). The Ras-exchange-factor region of Sos consists of N- and C-terminal α-helical domains, with the C-terminal α-helical domain interacting with Ras (Fig. 17.1.43C). Another group of Ras-like GTP-binding proteins is the Ran family that are involved in nuclear transport. The structure of Ran-GppNHp has also been determined (Vetter et al., 1999).

Rho and RhoGAP. The Rho family of small G proteins includes Rho, Rac, and Cdc42Hs. They regulate the phosphorylation rate of a variety of proteins involved in cell proliferation and cytoskeleton formation. The intrinsic GTPase activity of the Rho family of G proteins (i.e., the rate of hydrolysis of GTP to GDP) is slow but can be increased by 10^5-fold upon interacting with Rho family of G protein activators. Several crystal structures of Rho family G proteins and their complexes with RhoGAPs have been determined, thus revealing the molecular mechanism of the activation of Rho by its activator proteins (Fig. 17.1.43D). Among them, the structures of Rac1 (Hirshberg et al., 1997) and p50rhoGAP (Barrett et al., 1997), as well as the complexes between RhoA and RhoGAP (Rittinger et al., 1997b), Cdc42Hs and p50rhoGAP, and Rac1 and Tiam 1 have been published (Rittinger et al., 1997a; Worthylake et al., 2000; Table 17.1.19). The structure of Rho assumes a typical Ras-like G protein fold and RhoGAP adopts an α-helical fold similar to the p120GAP. In the structure of the complexes, an Arg residue on RhoGAP interacts with the P-loop of Rho, presumably to stabilize the transition state of the GTP hydrolysis.

Large GTP-binding proteins

Unlike the Ras-like small G-proteins, the family of large GTP-binding proteins is characterized by higher molecular weight and higher intrinsic GTPase activity. Members of this family include guanylate-binding proteins 1 and 2 (GBP1 and 2), Mx, and dynamin. The crystal structure of an intact human GBP1 with a molecular weight of 67,000 reveals a two-domain fold (Fig. 17.1.43E; Prakash et al., 2000). The N-terminal domain resembles a modified Ras-like G-protein fold with several insertions compared to the canonical Ras structure. The C-terminal domain is elongated and consists of seven α-helices including one that is 118-Å long. GBP1 lacks the conserved nucleotide binding motif (Asn/Thr)-Lys-X-Glu as observed in all Ras-like G-proteins. GBP1 is stable without the bound nucleotide, whereas Ras is not.

Heterotrimeric G proteins and their regulators

G protein trimer. G protein-coupled transmembrane receptor signaling cascades are ubiquitous throughout eukaryotic cells and util-

ize heterotrimeric G protein complexes. Crystal structures of two heterotrimeric $G_{t\alpha}/i_{\alpha}GDP$ $G_{t\beta\gamma}$ complexes reveal the molecular architecture and provide mechanistic insight (Table 17.1.19; Wall et al., 1995; Lambright et al., 1996). The $G_{t\alpha}$ subunit (Fig. 17.1.43F and G) is organized into three structural components: (1) a Ras-like GTPase domain that contains a mononucleotide-binding fold (see The Classical Mononucleotide-Binding Fold), (2) an all α-helical domain with a central long helix surrounded by five shorter helices, and (3) an N-terminal helix that projects away from the remainder of the $G_{t\alpha}$. The $G_{t\beta}$ subunit has an N-terminal helix followed by seven repeating WD motifs that are organized into a β-propeller structure (see Frequently Observed Secondary Structure Assemblies or Structural Motifs). Each WD-structural motif (also known as the Trp-Asp or WD40 motifs) consists of four antiparallel β-strands that form one propeller blade; however, the so-called WD repeat sequence corresponds to the three inner strands of a blade together with the outside (fourth) strand of the preceding blade. The $G_{t\gamma}$ subunit consists of two helices that pack against the β subunit.

The RGS fold. Regulators of G-protein signaling (RGS) modulate the GTPase activity of heterotrimeric G proteins. Their function is analogous to the GAPs that activate the GTPase activity of Ras-like G proteins (see above). The mechanism of RGS modulation is proposed to be stabilization of the transition state of catalytic G_{α} subunits, since RGS possesses no affinity for the G_{α}-GDP complex, modest affinity for the G_{α}-GTP state, and high affinity for the transition analog G_{α}-GDP-AlF_4^- complex. The crystal structures for members of RGS include the structure of RGS4 in complex with $G_{i\alpha}$-GDP-AlF_4^- (Tesmer et al., 1997), and the structure of the rgRGS domain of p115RhoGEF (Chen et al., 2001). The RGS domain consists of ~130 amino acids that fold into an all α-helical structure with nine helices divided into two subdomains (Fig. 17.1.43H; Tesmer, 1997). The larger domain, comprised of helices α4 to α7, assumes a classical right-handed antiparallel four-helix bundle fold, while the small domain consists of helices α1 to α3, α8, and α9. Instead of binding to the catalytic residues of the GTPase domain, RGS4 interacts with the three switch regions of $G_{i\alpha}$, thereby stabilizing their conformation.

The GoLoco motif. The nineteen–amino acid GoLoco motif regulates the function of heterotrimeric G proteins by binding to the GDP-bound form of the G_{α} subunit, thereby preventing reassociation of $G_{\beta\gamma}$ with the GDP-bound G_{α}. By doing so, GoLoco enables a continued presence of $G_{\beta\gamma}$ which can interact with effector proteins without activating G_{α}. This motif is present in RGS12, RGS14, LOCO, Purkinje-cell protein-2 (Pcp2), and Rap1GAP isoforms. The crystal structure of a $G_{\alpha i}$-GDP bound to the GoLoco region of RGS14 revealed key interactions between the GoLoco motif and the $G_{\alpha i1}$ subunit (Fig. 17.1.43I; Kimple et al., 2002). In particular, the conserved (N/E)QR triad of GoLoco interacts with the nucleotide binding pocket of G_{α} to make direct contacts with the α- and β-phosphates of GDP. The binding site of GoLoco on the $G_{\alpha i1}$ partially overlap with that of $G_{\beta\gamma}$, thus preventing reassociation of the G protein heterotrimer.

The F_1-ATPase (ATP synthase)

The F_1F_0ATPase found in the membranes of mitochondria, chloroplasts, and bacteria is responsible for synthesizing ATP from ADP and inorganic phosphate by utilizing a proton gradient that has been generated across the membrane. The large multisubunit ATPase complex is composed of an F_1 globular head catalytic domain linked by a central stalk to an integral membrane F_0 proton translocating domain. The F_1 domain, including the stalk, consists of five subunits with the stoichiometry $\alpha_3\beta_3\gamma\delta\varepsilon$. The F_0 domain is composed of a, b, and c subunits with varying stoichiometry. Various species have additional F_0 subunits.

Crystal structures of the F_1 domain and stalk regions from mitochondria, chloroplasts, and bacteria (Table 17.1.19) reveal that the α and β subunits are arranged in an alternating fashion around a pseudosymmetric three-fold axis to make a structure shaped like a slightly squashed pear (Fig. 17.1.44A and B). The α and β subunits are structurally very homologous even though there is marginal sequence identity (20%) between them. They each consist of three domains: an N-terminal six-stranded antiparallel β-barrel, a central nine-stranded mostly parallel β-sheet surrounded on both faces by nine helices, and a C-terminal six- to seven-helix bundle (Fig. 17.1.44B). The central domain of both α and β subunits bind mononucleotides, although only the β subunit is catalytic. Both subunits have a P-loop with a conserved mononucleotide-binding sequence motif at the C-terminal end of the sheet. The γ, δ, and ε subunits fold together intimately to form an α/β foot located just below the $\alpha_3\beta_3$ complex, con-

nected to it by a long 90-Å α-helix from the γ subunit that runs roughly parallel to the F_1 pseudo-three-fold axis (Fig. 17.1.44B; Gibbons et al., 2000). The γ subunit comprises a Rossman-like fold (see The Classical Dinucleotide-Binding Fold) with a five-stranded mostly parallel β-sheet (strand order 3-2-1-4-5) surrounded by six helices. The long, curving C-terminal helix and shorter N-terminal helix of this fold form the stalk region. The eukaryotic δ subunit (bacterial ε subunit) is a ten-stranded β-sandwich followed by a pair of antiparallel helices that project away from one end of the sandwich. The 50-residue mitochondrial ε subunit comprising a helix-loop-helix structure apparently has no counterpart in chloroplast or bacterial ATP synthases (Gibbons et al., 2000).

The structure of the entire complex has been more elusive. A 3.9-Å resolution unrefined C^{α} model of F_1F_0ATPase from yeast shows ten c subunits from the F_0 domain arranged in a circle (Stock et al., 1999). Each c subunit is a pair of antiparallel helices 47- and 58-Å long, respectively. This results in two concentric rings of helices with their helical axes running approximately perpendicular to the plane of the circle (Fig. 17.1.44C). The outer helices are kinked in the middle, giving the barrel a slight hourglass appearance.

Proteasome

The proteasome is the central nonlysosomal protein degradation enzyme observed in archeabacterium and eukaryotes. In most prokaryotes, a simpler proteasome homolog such as the ATP-dependent protease Hs1V plays a similar role. In addition to a host of cellular functions, including the degradation of abnormal or misfolded proteins, and the removal of various transient regulatory proteins, the proteasome also plays a major role in generating antigenic peptides for presentation by MHC class I molecules in mammals. The proteasome is composed of a 20S core proteolytic chamber and a 11S or 19S regulatory particle at the ends of the chamber.

The crystal structures of archeabacterium *Thermoplasma acidophilum* (Lowe et al., 1995), yeast (Groll et al., 1997), and bovine (Unno et al., 2002) 20S proteasomes reveal structures consisting of four heptameric rings, $\alpha_7\beta_7\beta_7\alpha_7$, stacked on top of each other to form a barrel-shaped complex with an approximate length of 150 Å and a diameter of 110 Å (Fig. 17.1.45, red and green regions). The α and β subunits share the same fold (known as the Ntn hydrolase fold; see Ntn Hydrolase Fold, consisting of a central β-sandwich flanked by two α-helices on one side, and two or three on the other (Fig. 17.1.35H). The eukaryotic forms consist of seven distinct types of β-subunits and seven distinct α-subunits, whereas the archael proteasome has one type of α- and one type of β-subunit. Only three of the eukaryotic β-subunits (β1, β2, and β5) are fully post-translationally processed to form mature catalytic domains with an N-terminal threonine nucleophile. None of the α-subunits are catalytic. Jawed vertebrates express an additional set of three catalytic β-subunits (β1i, β2i, and β5i) that are incorporated into the immunoproteasome that is responsible for generating MHC ligands for antigen presentation (Unno et al., 2002). The *Escherichia coli* ATP-dependent protease HslVU forms a dimer of two hexamers (Bochtler et al., 1997). The single HslV subunit is homologous to the β-subunit of proteasomes and has an N-terminal threonine nucleophile.

Crystal structures of the 11S regulator alone (human Regα) and in complex with the 20S proteasome (*T. brucei* 11S and yeast 20S) reveal a barrel-shaped heptamer with outer dimensions of ~90 Å in diameter and 70 Å in length (Knowlton et al., 1997; Whitby et al., 2000). Each monomer is a bundle of four long α-helices. This barrel is situated at both ends of the 20S particle essentially extending the proteasome cylinder (Fig 17.1.45). The inner diameter is 33 Å at the proteasome-binding end and 24 Å at the opposite end.

Viral Coat Proteins of Spherical Viruses

The capsid proteins of a virus provide a protective shell around the enclosed RNA or DNA. These viral coats are made from an assembly of identical or nearly identical small protein subunits. The protein capsids of all known spherical viruses are built using icosahedral symmetry. In its simplest form, an icosahedron is a 20-sided polygon with two-, three-, and five-fold symmetry, made from an arrangement of identical equilateral triangles (Fig. 17.1.46A). Because constructing a protein subunit with equilateral three-fold symmetry is difficult, the simplest known viruses contain three asymmetric subunits per triangle or 60 subunits per virus. Larger and more complex viral coats can be made by further subdividing these asymmetric units and using more protein subunits, all the while maintaining icosahedral symmetry. Viruses can be classified by their triangulation (T) number which indicates the

number of capsid subunits in multiples of 60. For example, a T = 1 virus has 60 subunits and a T = 3 virus has 3 × 60 = 180 subunits.

The three-dimensional structures of numerous capsid proteins for spherical viruses that infect bacteria, plants, insects, and animals have revealed a common fold for most of these structures, namely an eight-stranded jelly-roll β sandwich (see Elements of Protein Structure). Strands 1-8-3-6 form one sheet and strands 2-7-4-5 form the other sheet of the β sandwich (Fig. 17.1.46B). Despite the lack of sequence homology between various classes of viral-coat proteins, the jelly-roll β sandwich topology is well conserved in >50 known capsid structures according to SCOP classification of the PDB (Murzin et al., 1995). It forms a wedge-shaped β sandwich, generally with short loops at one end and longer loops at the other. A striking exception to this fold is observed in the coat protein of bacteriophage MS2, which is a T = 3 virus (see MS2 Phage Coat Protein).

INTEGRAL MEMBRANE PROTEINS

High-resolution structural information is available for only ~50 integral membrane proteins (Table 17.1.20), compared with over 10,000 soluble proteins, although an estimated 15% to 30% of genes encode membrane proteins. Difficulties in protein expression, solubilization, and the growth of well-ordered crystals have plagued structural studies. Nonetheless, remarkable progress has been made in solving structures of integral membrane proteins owing to advances in overexpression, availability of detergents, and the use of antibody fragments in crystallization. Complementary methodological approaches, including X-ray diffraction, cryo-electron microscopy, NMR spectroscopy, and electron paramagnetic resonance spectroscopy promise even faster future progress. Updated compendiums of structures of membrane proteins can be found at *http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html* and *http://www.mpibp-frankfurt.mpg.de/michel/public/memprotstruct.html*.

The diversity of integral membrane protein folds appears to be much less than soluble proteins (Schulz, 2002). Due to the thermodynamic cost of burying charges in the hydrophobic environment of the membrane, integral membrane structures must minimize exposed charges and unsatisfied hydrogen bonds. Two classes of structures are observed satisfying these requirements: the α-helix and the closed β-barrel. The α-helical structures are most prevalent and are found in cytoplasmic membrane proteins. They include both transmembrane α-helices and α-helical structures in monotopic integral membrane proteins which do not cross the bilayer. β-barrel structures are characteristic of proteins of the outer membranes of Gram-negative bacteria.

Photosynthetic Proteins

The first high-resolution structure determination of an integral membrane protein was a photosynthetic reaction center from the purple bacterium *Rhodopseudomonas virdis,* for which Johann Deisenhofer, Robert Huber, and Hartmut Michel were awarded the Nobel Prize in Chemistry in 1988 (Deisenhofer et al., 1984, 1985, 1995; Deisenhofer and Michel, 1989). This large complex includes four protein subunits and fourteen cofactors. The transmembrane core of the complex has pseudo two-fold symmetry, formed by the structurally related chains L and M; each of which contains five membrane-spanning α-helices. The chlorophyll, quinone, carotenoid, and nonheme iron cofactors associate with the L and M subunits. The H subunit forms an additional transmembrane α-helix and also contains β structures on the cytoplasmic face. The eleven transmembrane helices are arranged approximately parallel to each other (Fig. 17.1.47A). The cytochrome subunit, bound to four heme groups, is on the periplasmic face and does not span the membrane. More recently, additional photosynthetic complexes have been solved, including the reaction center of *Rhodobacter sphaeroides*, the two photosystems of chloroplasts, and antenna light-harvesting complexes (Axelrod et al., 2002; Chirino et al., 1994; Jordan et al., 2001; Koepke et al., 1996; Prince et al., 1997; Zouni et al., 2001).

Seven-Helix Proteins

A number of families of integral membrane proteins contain seven membrane-spanning α-helices. Among them, the microbial rhodopsins are the best studied. The crystallographic studies of bacteriorhodopsin from *Halobacterium salinarum* have provided insight into the mechanism of this light-driven proton pump (Lanyi, 1999; Subramaniam, 1999). The helices pack around a covalently linked retinal chromophore, forming an internal channel lined by hydrophilic residues (Fig. 17.1.47B). Conformational changes in the transmembrane regions have been observed during the course of the photocycle (Kataoka and Kamikubo,

2000). More recently, structures of two related haloarchaeon proteins, halorhodopsin (Kolbe et al., 2000) and sensory rhodopsin II (Luecke et al., 2001; Royant et al., 2001; Gordeliy et al., 2002), have also been determined. The versatility of the seven transmembrane helix bundle is apparent. For example, bacteriorhodopsin and halorhodopsin are ion transporters, while sensory rhodopsin participates in photosensory signaling and interacts with transducer proteins. Structural differences, particularly in the retinal binding pocket and conformation of bound retinal, reflect the distinct functional role of sensory rhodopsin (Luecke et al., 2001; Royant et al., 2001).

In higher organisms, the seven-helix fold is seen in receptors coupled to G proteins, a large family of integral membrane proteins. The crystal structure of the mammalian retinal photoreceptor rhodopsin represents the first high-resolution structure of a G protein-coupled receptor (Palczewski et al., 2000). Although the overall topology of the seven-helix bundle of bovine rhodopsin resembles that of microbial rhodopsins, its structure appears unique. Unlike the parallel arrangement of the helices in bacteriorhodopsin, some of the bovine rhodopsin helices are tilted or strongly bent at proline residues (Fig. 17.1.47C). Moreover, bovine rhodopsin contains more extensive secondary structures in the intra- and extracellular regions, consistent with the disparate receptor function. There is also an additional short helix that lies parallel to the membrane surface on the cytoplasmic face. This amphipathic eighth helix is of particular interest because of its location in the G protein binding region. On the extracellular face, the three loops associate with portions of the N-terminus to form a compact structure centered around two twisted β hairpins. The second extracellular loop, which contains the inner hairpin, is linked through a conserved disulfide bond to transmembrane helix 3, and forms a plug covering the chromophore binding pocket (Bourne and Meng, 2000).

Ion Channels

Ion channels allow movement of selected ions along electrochemical gradients, and can open and close (i.e., gate) in response to various stimuli. A successful approach to their structural study has involved use of bacterial homologs. Two bacterial potassium channel structures have been determined: the KcsA H^+-gated channel and the Ca^{2+}-gated MthK channel (Doyle et al., 1998; Jiang et al., 2002a,b). The bacterial mechanosensitive channels MscL and MscS, which respond to mechanical stretching of the cell membrane, have also been solved (Bass et al., 2002; Chang et al., 1998).

Structures of two ClC chloride channels were recently determined from *E. coli* and *Salmonella typhimurium* (Dutzler et al., 2002). The ClC chloride channel is dimeric, with each subunit forming its own pore to make up a double-barreled structure, as predicted by biophysical studies (Fig. 17.1.47D). The secondary structure is all α-helical, with each subunit containing seventeen membrane-inserted α-helices. Many of the membrane-inserted helices are tilted relative to the plane of the membrane, and not all pass entirely through the bilayer. The crystal structure revealed an internal structural repeat, wherein the N-terminal portion of each subunit is structurally related to the C-terminal portion. This unanticipated internal symmetry results in an antiparallel arrangement of the subunit halves. A selectivity filter is formed by helix dipoles and main chain atoms that interact with bound Cl^-. The partial positive charge generated by the helix dipoles likely supports electrostatic stabilization of Cl^- but does not result in tight binding (and thus does not restrict Cl^- movement through the pore). The carboxylate of a conserved glutamic acid near the chloride-binding site protrudes into the pore, suggesting a gating function to open and close the channel through electrostatic repulsion.

ABC Transporters

The large family of ATP binding cassette (ABC) transporters are involved in ATP-driven import and export of a variety of substrates. The functional transporters contain two transmembrane domains and two cytoplasmic nucleotide-binding domains. The structures of MsbA lipid transporter (flippase) and BtuCD vitamin B_{12} transporter are two members of this family solved to date (Chang and Roth, 2001; Locher et al., 2002). The membrane-spanning portion of the *E. coli* BtuCD transporter, made up of two BtuC subunits, has a total of 20 transmembrane α-helices that display a complex packing arrangement (Fig. 17.1.47E). The translocation channel seems to be located at the BtuC dimer interface. The two BtuD cassettes interact with each other and are attached to the BtuC module on the cytoplasmic face. A large, water-filled vestibule located in the cytoplasm between the four subunits may allow vitamin B_{12} to exit the transporter; two cytoplasmic loops of BtuC appear to serve as the channel gate. Comparison

of BtuCD with MsbA shows that the arrangement and packing of the transmembrane domains differ significantly. While in BtuCD, both the cytoplasmic and the transmembrane domains of each subunit remain in close contact, in MsbA they are oriented away from each other.

P-type ATPases

The structure of a calcium ATPase from skeletal-muscle sarcoplasmic reticulum has been determined in both E1 (Ca^{2+}-bound) and E2 (Ca^{2+}-free) states, providing details of the mechanism of this P-type ATPase (Toyoshima et al., 2000; Toyoshima and Nomura, 2002). The Ca^{2+} ATPase has ten transmembrane α-helices and three cytoplasmic domains. Two calcium ions are bound at the center of four transmembrane helices in the E1 state, and two of these helices are in unwound conformations to efficiently coordinate the ions. Four transmembrane helices, one of which is 41-residues long, appear to extend through the bilayer into the cytoplasm, forming a stalk region that connects the transmembrane portion to the cytoplasmic domains. The crystal structure of the E2 state revealed that large movements of the cytoplasmic domains and rearrangements of the transmembrane helices take place during the reaction cycle (Fig. 17.1.47F; Toyoshima and Nomura, 2002).

Other Channels and Transporters

The structure of *E. coli* efflux transporter AcrB, a major multidrug exporter which cooperates with the β-barrel membrane protein TolC (see below), was recently determined by X-ray crystallography (Murakami et al., 2002b). The first structure of a mammalian channel protein was that of a bovine aquaporin water channel, AQP1, which was studied by electron crystallography, electron microscopy, and X-ray crystallography (Murata et al., 2000; Ren et al., 2001; Sui et al., 2001). AQP1 shares the same overall topology as its bacterial homolog GlpF, a glycerol channel (Fu et al., 2000). AQP1 contains a quadruple-barreled channel, with individual pores formed by each of its four subunits. Each monomer has six transmembrane and two membrane-inserted α-helices (Fig. 17.1.47G). Severe tilting of the transmembrane helices relative to the membrane normal, as was seen in the ClC chloride channel, allows for formation of the dumbbell-shaped pore in AQP1. The structure of the pore reveals that water selectivity is based on a narrow constriction region that imposes a 2.8-Å diameter steric limit on the channel.

Respiratory Enzymes

Several structures of respiratory enzymes have been solved, including both bacterial and mammalian cytochrome *c* oxidase (Iwata et al., 1995; Tsukihara et al., 1996; Ostermeier et al., 1997; Soulimane et al., 2000), fumarate reductase (Iverson et al., 1999; Lancaster et al., 1999), and formate reductase-N (Jormakka et al., 2002). Crystal structures of the bovine and chicken cytochrome bc_1 complex (complex III) as well as its *Saccharomyces cerevisiae* homolog provide intriguing insights into the function of the respiratory chain (Xia et al., 1997; Iwata et al., 1998; Zhang et al., 1998; Hunte et al., 2000). The complex functions as a dimer of multisubunit monomers. In the mammalian bc_1 complex, seven of the eleven subunits have membrane-spanning components, forming a total of 26 transmembrane helices in the dimer (Fig. 17.1.47H). Conformations of the Rieske iron-sulfur protein in different crystal forms have led to the suggestion that the soluble portion of the subunit may move during catalysis to facilitate the transfer of electrons from ubiquinol to cytochrome c_1 (Iwata et al., 1998; Zhang et al., 1998).

β-Barrel Membrane Proteins

A distinct class of membrane proteins contain membrane-spanning β-strands arranged into closed barrels. Transmembrane β-barrel structures are composed of 8 to 22 antiparallel strands (Schulz, 2002). This class is best illustrated by porins, which are channels of the outer membrane of Gram-negative bacteria (Fig. 17.1.1K). Related to the porins are the iron transporters FepA and FhuA, which form 22-stranded barrels housing globular plug domains (Fig. 17.1.47I; Buchanan et al., 1999; Ferguson et al., 2002). Both porins and maltoporins appear as trimers with each subunit forming one β-barrel channel. A more unusual structure is TolC, an *E. coli* outer membrane protein involved in export of a range of molecules. Three protein subunits each contribute a four-stranded β sheets to form a single twelve-stranded outer membrane-spanning β-barrel (Fig. 17.1.47J). In addition, TolC also has a large, left-handed α-helical barrel stabilized by coiled-coil interactions. The α-helical barrel is contiguous with the β-barrel domain, and extends 100 Å into the periplasm. This results in the formation of a 140-Å long channel-tunnel that spans both the outer membrane and the

periplasmic space, allowing interaction with inner membrane proteins.

ACKNOWLEDGEMENTS

The authors would like to thank L. Birnbaumer, T. Darden, and T. Transue for the insightful comments and suggestions they contributed to this manuscript.

LITERATURE CITED

Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E. 1994. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370:621-628.

Adams, M.J., Ford, G.C., Koekoek, R., Lentz, P.J., McPherson, A. Jr., Rossmann, M.G., Smiley, I.E., Schevitz, R.W., and Wonacott, A.J. 1970. Structure of lactate dehydrogenase at 2-8 Å resolution. *Nature* 227:1098-1103.

Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M., and Harrison, S.C. 1988. Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. *Science* 242:899-907.

Albert, A., Yenush, L., Gil-Mascarell, M.R., Rodriguez, P.L., Patel, S., Martinez-Ripoll, M., Blundell, T.L., and Serrano, R. 2000. X-ray structure of yeast Hal2p, a major target of lithium and sodium toxicity, and identification of framework interactions determining cation sensitivity. *J. Mol. Biol.* 295:927-938.

Alden, R.A., Birktoft, J.J., Kraut, J., Robertus, J.D., and Wright, C.S. 1971. Atomic coordinates for subtilisin BPN′ (or Novo). *Biochem. Biophys. Res. Commun.* 45:337-344.

Allaire, M., Chernaia, M.M., Malcolm, B.A., and James, M.N. 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* 369:72-76.

Allured, V.S., Collier, R.J., Carroll, S.F., and McKay, D.B. 1986. Structure of exotoxin A of *Pseudomonas aeruginosa* at 3.0-Angstrom resolution. *Proc. Natl. Acad. Sci. U.S.A.* 83:1320-1324.

Alnemri, E.S., Livingston, D.J., Nicholson, D.W., Salvesen, G., Thornberry, N.A., Wong, W.W., and Yuan, J. 1996. Human ICE/CED-3 protease nomenclature. *Cell* 87:171.

Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J., and Hilgenfeld, R. 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* 21:3213-3224.

Anderson, C.M., McDonald, R.C., and Steitz, T.A. 1978. Sequencing a protein by x-ray crystallography. I. Interpretation of yeast hexokinase B at 2.5 Å resolution by model building. *J. Mol. Biol.* 123:1-13.

Anderson, W.F., Ohlendorf, D.H., Takeda, Y., and Matthews, B.W. 1981. Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature* 290:754-758.

Andrade, M.A. and Bork, P. 1995. HEAT repeats in the Huntington's disease protein. *Nat. Genet.* 11:115-116.

Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W., and Bork, P. 2001. Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* 309:1-18.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* 181:223-230.

Antson, A.A., Otridge, J., Brzozowski, A.M., Dodson, E.J., Dodson, G.G., Wilson, K.S., Smith, T.M., Yang, M., Kurecki, T., and Gollnick, P. 1995. The structure of trp RNA-binding attenuation protein. *Nature* 374:693-700.

Appel, R.D., Bairoch, A., and Hochstrasser, D.F. 1994. A new generation of information retrieval tools for biologists: The example of the ExPASy WWW server. *Trends Biochem. Sci.* 19:258-260.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob, H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M., Servant, F., Sigrist, C.J., and Zdobnov, E.M. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29:37-40.

Arcus, V. 2002. OB-fold domains: A snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* 12:794-801.

Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad.Sci. U.S.A.* 88:10148-10152.

Aritomi, M., Kunishima, N., Okamoto, T., Kuroki, R., Ota, Y., and Morikawa, K. 1999. Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition scheme. *Nature* 401:713-717.

Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. 1994. Atomic structure of the RuvC resolvase: A holliday junction-specific endonuclease from *E. coli*. *Cell* 78:1063-1072.

Armstrong, N., Sun, Y., Chen, G.Q., and Gouaux, E. 1998. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* 395:913-917.

Arora, A., Abildgaard, F., Bushweller, J.H., and Tamm, L.K. 2001. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat. Struct. Biol.* 8:334-338.

Arvai, A.S., Bourne, Y., Hickey, M.J., and Tainer, J.A. 1995. Crystal structure of the human cell cycle protein CksHs1: Single domain fold with similarity to kinase N-lobe domain. *J. Mol. Biol.* 249:835-842.

Axelrod, H.L., Abresch, E.C., Okamura, M.Y., Yeh, A.P., Rees, D.C., and Feher, G. 2002. X-ray structure determination of the cytochrome c2: Reaction center electron transfer complex from *Rhodobacter sphaeroides*. *J. Mol. Biol.* 319:501-515.

Bairoch, A. and Bucher, P. 1994. PROSITE: Recent developments. *Nucleic Acids Res.* 22:3583-3589.

Baldwin, E.T., Franklin, K.A., Appella, E., Yamada, M., Matsushima, K., Wlodawer, A., and Weber, I.T. 1990. Crystallization of human interleukin-8. A protein chemotactic for neutrophils and T-lymphocytes. *J. Biol. Chem.* 265:6851-6853.

Ball, L.J., Jarchau, T., Oschkinat, H., and Walter, U. 2002. EVH1 domains: Structure, function and interactions. *FEBS Lett.* 513:45-52.

Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D., and Waley, S.G. 1975. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature* 255:609-614.

Banner, D.W., D'Arcy, A., Janes, W., Gentz, R., Schoenfeld, H.J., Broger, C., Loetscher, H., and Lesslauer, W. 1993. Crystal structure of the soluble human 55 kd TNF receptor-human TNF-β complex: Implications for TNF receptor activation. *Cell* 73:431-445.

Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L., and Bohm, A. 2000. Structure of yeast poly(A) polymerase alone and in complex with 3′-dATP. *Science* 289:1346-1349.

Barford, D., Flint, A.J., and Tonks, N.K. 1994. Crystal structure of human protein tyrosine phosphatase 1B. *Science* 263:1397-1404.

Barford, D., Das, A.K., and Egloff, M.P. 1998. The structure and mechanism of protein phosphatases: Insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.* 27:133-164.

Barlow, P.N., Norman, D.G., Steinkasserer, A., Horne, T.J., Pearce, J., Driscoll, P.C., Sim, R.B., and Campbell, I.D. 1992. Solution structure of the fifth repeat of factor H: A second example of the complement control protein module. *Biochemistry* 31:3626-3634.

Barlow, P.N., Steinkasserer, A., Norman, D.G., Kieffer, B., Wiles, A.P., Sim, R.B., and Campbell, I.D. 1993. Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol.* 232:268-284.

Barrett, T., Xiao, B., Dodson, E.J., Dodson, G., Ludbrook, S.B., Nurmahomed, K., Gamblin, S.J., Musacchio, A., Smerdon, S.J., and Eccleston, J.F. 1997. The structure of the GTPase-activating domain from p50rhoGAP. *Nature* 385:458-461.

Bass, R.B., Strop, P., Barclay, M., and Rees, D.C. 2002. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* 298:1582-1587.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. 2002. The Pfam protein families database. *Nucleic Acids Res.* 30:276-280.

Batey, R.T., Rambo, R.P., Lucast, L., Rha, B., and Doudna, J.A. 2000. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* 287:1232-1239.

Baumann, U., Wu, S., Flaherty, K.M., and McKay, D.B. 1993. Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa: A two-domain protein with a calcium binding parallel beta roll motif. *EMBO J.* 12:3357-3364.

Baxevanis, A., Davison, D.B., Page, R.D.M., Petsko, G.A., Stein, L.D., and Stormo, G.D. (eds.) 2004. Current Protocols in Bioinformatics. John Wiley & Sons, Hoboken, N.J.

Becker, S., Groner, B., and Muller, C.W. 1998. Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature* 394:145-151.

Beese, L.S., Derbyshire, V., and Steitz, T.A. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* 260:352-355.

Berardi, M.J., Sun, C., Zehr, M., Abildgaard, F., Peng, J., Speck, N.A., and Bushweller, J.H. 1999. The Ig fold of the core binding factor alpha Runt domain is a member of a family of structurally and functionally related Ig-fold DNA-binding domains. *Structure. Fold. Des* 7:1247-1256.

Berg, J.M. and Shi, Y. 1996. The galvanization of biology: A growing appreciation for the roles of zinc. *Science* 271:1081-1085.

Berger, J.,M., Gamblin, S.J., Harrison, S.C., and Wang, J.C. 1996. Structure and mechanism of DNA topoisomerase II. *Nature* 379:225-232. [published erratum appears in *Nature* 380:179]

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235-242.

Bernier-Villamor, V., Sampson, D.A., Matunis, M.J., and Lima, C.D. 2002. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* 108:345-356.

Bernstein, H.J. 2000. Recent changes to RasMol, recombining the variants. *Trends Biochem.Sci.* 25:453-455.

Bewley, M.C., Boustead, C.M., Walker, J.H., Waller, D.A., and Huber, R. 1993. Structure of chicken annexin V at 2.25-Å resolution. *Biochemistry* 32:3923-3929.

Bezprozvanny, I. and Maximov, A. 2001. Classification of PDZ domains. *FEBS Lett.* 509:457-462.

Bianchet, M.A., Bains, G., Pelosi, P., Pevsner, J., Snyder, S.H., Monaco, H.L., and Amzel, L.M. 1996. The three-dimensional structure of bovine odorant binding protein and its mechanism of odor recognition. *Nat. Struct. Biol.* 3:934-939.

Bianchet, M.A., Hullihen, J., Pedersen, P.L., and Amzel, L.M. 1998. The 2.8-Å structure of rat liver F1-ATPase: Configuration of a critical intermediate in ATP synthesis/hydrolysis. *Proc. Natl. Acad.Sci. U.S.A* 95:11065-11070.

Binda, C., Newton-Vinson, P., Hubalek, F., Edmondson, D.E., and Mattevi, A. 2002. Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders. *Nat. Struct. Biol.* 9:22-26.

Birck, C., Poch, O., Romier, C., Ruff, M., Mengus, G., Lavigne, A.C., Davidson, I., and Moras, D. 1998. Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* 94:239-249.

Birktoft, J.J. and Blow, D.M. 1972. Structure of crystalline -chymotrypsin. V. The atomic structure of tosyl-chymotrypsin at 2 Å resolution. *J. Mol. Biol.* 68:187-240.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329:506-512.

Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., and Margolis, B. 1994. A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. *J. Biol. Chem.* 269:32031-32034.

Blomberg, N., Baraldi, E., Nilges, M., and Saraste, M. 1999. The PH superfold: A structural scaffold for multiple functions. *Trends Biochem. Sci.* 24:441-445.

Bochtler, M., Ditzel, L., Groll, M., Hartmann, C., and Huber, R. 1999. The proteasome. *Annu. Rev. Biophys. Biomol. Struct.* 28:295-317.

Bochtler, M., Ditzel, L., Groll, M., and Huber, R. 1997. Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 94:6070-6074.

Bock, J.B., Matern, H.T., Peden, A.A., and Scheller, R.H. 2001. A genomic perspective on membrane compartment organization. *Nature* 409:839-841.

Bode, W., Gomis-Ruth, F.X., Huber, R., Zwilling, R., and Stocker, W. 1992. Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* 358:164-167.

Bode, W., Gomis-Ruth, F.X., and Stockler, W. 1993. Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett* 331:134-140.

Bone, R., Springer, J.P., and Atack, J.R. 1992. Structure of inositol monophosphatase, the putative target of lithium therapy. *Proc. Natl. Acad. Sci. U.S.A* 89:10031-10035.

Booker, G.W., Breeze, A.L., Downing, A.K., Panayotou, G., Gout, I., Waterfield, M.D., and Campbell, I.D. 1992. Structure of an SH2 domain of the p85 alpha subunit of phosphatidylinositol-3-OH kinase. *Nature* 358:684-687.

Booker, G.W., Gout, I., Downing, A.K., Driscoll, P.C., Boyd, J., Waterfield, M.D., and Campbell, I.D. 1993. Solution structure and ligand-binding site of the SH3 domain of the p85 alpha subunit of phosphatidylinositol 3-kinase. *Cell* 73:813-822.

Boriack-Sjodin, P.A., Margarit, S.M., Bar-Sagi, D., and Kuriyan, J. 1998. The structural basis of the activation of Ras by Sos. *Nature* 394:337-343.

Bourne, H.R. and Meng, E.C. 2000. Structure. Rhodopsin sees the light. *Science* 289:733-734.

Boyington, J.C. and Sun, P.D. 2002. A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. *Mol. Immunol.* 38:1007-1021.

Boyington, J.C., Riaz, A.N., Patamawenu, A., Coligan, J.E., Brooks, A.G., and Sun, P.D. 1999. Structure of CD94 reveals a novel C-type lectin fold: Implications for the NK cell-associated CD94/NKG2 receptors. *Immunity* 10:75-82.

Boyington, J.C., Motyka, S.A., Schuck, P., Brooks, A.G., and Sun, P.D. 2000. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* 405:537-543.

Bracey, M.H., Hanson, M.A., Masuda, K.R., Stevens, R.C., and Cravatt, B.F. 2002. Structural adaptations in a membrane enzyme that terminates endocannabinoid signaling. *Science* 298:1793-1796.

Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L., and Sigler, P.B. 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 371:578-586.

Braig, K., Adams, P.D., and Brunger, A.T. 1995. Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nat. Struct. Biol.* 2:1083-1094.

Brannigan, J.A., Dodson, G., Duggleby, H.J., Moody, P.C., Smith, J.L., Tomchick, D.R., and Murzin, A.G. 1995. A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* 378:416-419. [published erratum appears in *Nature* 378:644]

Bravo, J., Li, Z., Speck, N.A., and Warren, A.J. 2001. The leukemia-associated AML1 (Runx1)—CBF beta complex functions as a DNA-induced molecular clamp. *Nat. Struct. Biol.* 8:371-378.

Breeden, L. and Nasmyth, K. 1987. Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of *Drosophila*. *Nature* 329:651-654.

Brino, L., Urzhumtsev, A., Mousli, M., Bronner, C., Mitschler, A., Oudet, P., and Moras, D. 2000. Dimerization of *Escherichia coli* DNA-gyrase B provides a structural mechanism for activating the ATPase catalytic center. *J. Biol. Chem.* 275:9468-9475.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., and Wiley, D.C. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364:33-39.

Buchanan, S.K., Smith, B.S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der, H.D., and Deisenhofer, J. 1999. Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. *Nat. Struct. Biol.* 6:56-63.

Burd, C.G. and Dreyfuss, G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265:615-621.

Burgering, M.J., Boelens, R., Gilbert, D.E., Breg, J.N., Knight, K.L., Sauer, R.T., and Kaptein, R. 1994. Solution structure of dimeric Mnt repressor (1-76). *Biochemistry* 33:15036-15045.

Burley, S.K. 1996. The TATA box binding protein. *Curr. Opin. Struct. Biol.* 6:69-75.

Burley, S.K., David, P.R., Taylor, A., and Lipscomb, W.N. 1990. Molecular structure of leucine aminopeptidase at 2.7-Å resolution. *Proc. Natl. Acad.. Sci. U.S.A.* 87:6878-6882.

Burley, S.K., David, P.R., Sweet, R.M., Taylor, A., and Lipscomb, W.N. 1992. Structure determination and refinement of bovine lens leucine aminopeptidase and its complex with bestatin. *J. Mol. Biol.* 224:113-140.

Burmeister, W.P., Ruigrok, R.W., and Cusack, S. 1992. The 2.2 Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. *EMBO J.* 11:49-56.

Burmeister, W.P., Gastinel, L.N., Simister, N.E., Blum, M.L., and Bjorkman, P.J. 1994. Crystal structure at 2.2 Å resolution of the MHC-related neonatal Fc receptor. *Nature* 372:336-343.

Buss, K.A., Cooper, D.R., Ingram-Smith, C., Ferry, J.G., Sanders, D.A., and Hasson, M.S. 2001. Urkinase: Structure of acetate kinase, a member of the ASKHA superfamily of phosphotransferases. *J. Bacteriol.* 183:680-686.

Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., and Stuart, D.I. 2001. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* 410:235-240.

Bycroft, M., Grunert, S., Murzin, A.G., Proctor, M., and St Johnston, D. 1995. NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J.* 14:3563-3571. [published erratum appears in *EMBO J.* 14:4385]

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. 2001. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 104:901-912.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell* 9:527-539.

Carter, C.W. Jr. 1993. Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. *Annu. Rev. Biochem.* 62:715-748.

Carter, A.P., Clemons, W.M. Jr., Brodersen, D.E., Morgan-Warren, R.J., Hartsch, T., Wimberly, B.T., and Ramakrishnan, V. 2001. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* 291:498-501.

Casasnovas, J.M., Larvie, M., and Stehle, T. 1999. Crystal structure of two CD46 domains reveals an extended measles virus-binding surface. *EMBO J.* 18:2911-2922.

Castiglone Morelli, M.A., Stier, G., Gibson, T., Joseph, C., Musco, G., Pastore, A., and Trave, G. 1995. The KH module has an alpha beta fold. *FEBS Lett.* 358:193-198.

Cavarelli, J., Rees, B., Ruff, M., Thierry, J.C., and Moras, D. 1993. Yeast tRNA(Asp) recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature* 362:181-184.

Cesareni, G., Panni, S., Nardelli, G., and Castagnoli, L. 2002. Can we infer peptide recognition specificity mediated by SH3 domains? *FEBS Lett.* 513:38-44.

Cha, S.S., Kim, M.S., Choi, Y.H., Sung, B.J., Shin, N.K., Shin, H.C., Sung, Y.C., and Oh, B.H. 1999. 2.8 Å resolution crystal structure of human TRAIL, a cytokine with selective antitumor activity. *Immunity.* 11:253-261.

Chai, J., Shiozaki, E., Srinivasula, S.M., Wu, Q., Datta, P., Alnemri, E.S., Shi, Y., and Dataa, P. 2001. Structural basis of caspase-7 inhibition by XIAP. *Cell* 104:769-780.

Champoux, J.J. 2001. DNA topoisomerases: Structure, function, and mechanism. *Annu. Rev. Biochem.* 70:369-413.

Chang, G. and Roth, C.B. 2001. Structure of MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* 293:1793-1800.

Chang, G., Spencer, R.H., Lee, A.T., Barclay, M.T., and Rees, D.C. 1998. Structure of the MscL homolog from Mycobacterium tuberculosis: A gated mechanosensitive ion channel. *Science* 282:2220-2226.

Chen, F.E., Huang, D.B., Chen, Y.Q., and Ghosh, G. 1998a. Crystal structure of p50/p65 heterodimer of transcription factor NF- κB bound to DNA. *Nature* 391:410-413.

Chen, L., Glover, J.N., Hogan, P.G., Rao, A., and Harrison, S.C. 1998b. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* 392:42-48.

Chen, X., Vinkemeier, U., Zhao, Y., Jeruzalmi, D., Darnell, J.E. Jr., and Kuriyan, J. 1998c. Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* 93:827-839.

Chen, Y.Q., Ghosh, S., and Ghosh, G. 1998d. A novel DNA recognition mode by the NF-kappa B p65 homodimer. *Nat. Struct. Biol.* 5:67-73.

Chen, Z., Wells, C.D., Sternweis, P.C., and Sprang, S.R. 2001. Structure of the rgRGS domain of p115RhoGEF. *Nat. Struct. Biol.* 8:805-809.

Chirino, A.J., Lous, E.J., Huber, M., Allen, J.P., Schenck, C.C., Paddock, M.L., Feher, G., and Rees, D.C. 1994. Crystallographic analyses of site-directed mutants of the photosynthetic reaction center from *Rhodobacter sphaeroides*. *Biochemistry* 33:4584-4593.

Cho, W. 2001. Membrane targeting by C1 and C2 domains. *J. Biol. Chem.* 276:32407-32410.

Cho, Y., Gorina, S., Jeffrey, P.D., and Pavletich, N.P. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* 265:346-355.

Choe, S., Bennett, M.J., Fujii, G., Curmi, P.M., Kantardjieff, K.A., Collier, R.J., and Eisenberg, D. 1992. The crystal structure of diphtheria toxin. *Nature* 357:216-222.

Chook, Y.M. and Blobel, G. 1999. Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature* 399:230-237.

Chow, D., He, X., Snow, A.L., Rose-John, S., and Garcia, K.C. 2001. Structure of an extracellular gp130 cytokine receptor signaling complex. *Science* 291:2150-2155.

Cingolani, G., Petosa, C., Weis, K., and Muller, C.W. 1999. Structure of importin-β bound to the IBB domain of importin-alpha. *Nature* 399:221-229.

Clark, K.L., Halay, E.D., Lai, E., and Burley, S.K. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364:412-420.

Clore, G.M., Appella, E., Yamada, M., Matsushima, K., and Gronenborn, A.M. 1990. Three-dimensional structure of interleukin 8 in solution. *Biochemistry* 29:1689-1696.

Coll, M., Seidman, J.G., and Muller, C.W. 2002. Structure of the DNA-bound T-box domain of human TBX3, a transcription factor responsible for ulnar-mammary syndrome. *Structure.(Camb.)* 10:343-356.

Collins, E.J., Garboczi, D.N., and Wiley, D.C. 1994. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature* 371:626-629.

Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. 1998. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell* 94:193-204.

Cowan, S.W., Newcomer, M.E., and Jones, T.A. 1990. Crystallographic refinement of human serum retinol binding protein at 2Å resolution. *Proteins* 8:44-61.

Cowan, S.W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R.A., Jansonius, J.N., and Rosenbusch, J.P. 1992. Crystal structures explain functional properties of two *E. coli* porins. *Nature* 358:727-733.

Cowan, S.W., Newcomer, M.E., and Jones, T.A. 1993. Crystallographic studies on a family of cellular lipophilic transport proteins. Refinement of P2 myelin protein and the structure determination and refinement of cellular retinol-binding protein in complex with all-*trans*-retinol. *J. Mol. Biol.* 230:1225-1246.

Cramer, P., Larson, C.J., Verdine, G.L., and Muller, C.W. 1997. Structure of the human NF-kappaB p52 homodimer-DNA complex at 2.1 A resolution. *EMBO J.* 16:7078-7090.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288:640-649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863-1876.

Crennell, S.J., Garman, E.F., Laver, W.G., Vimr, E.R., and Taylor, G.L. 1993. Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. *Proc. Natl. Acad. Sci. U.S.A.* 90:9852-9856.

Curry, S., Mandelkow, H., Brick, P., and Franks, N. 1998. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nat. Struct. Biol.* 5:827-835.

Cusack, S. 1995. Eleven down and nine to go. *Nat. Struct. Biol.* 2:824-831.

Czworkowski, J., Wang, J., Steitz, T.A., and Moore, P.B. 1994. The crystal structure of elongation factor G complexed with GDP, at 2.7 Å resolution. *EMBO J.* 13:3661-3668.

Daopin, S., Piez, K.A., Ogawa, Y., and Davies, D.R. 1992. Crystal structure of transforming growth factor-beta 2: An unusual fold for the superfamily. *Science* 257:369-373.

Daopin, S., Li, M., and Davies, D.R. 1993. Crystal structure of TGF-β2 refined at 1.8 Å resolution. *Proteins* 17:176-192.

Darimont, B.D., Wagner, R.L., Apriletti, J.W., Stallcup, M.R., Kushner, P.J., Baxter, J.D., Fletterick, R.J., and Yamamoto, K.R. 1998. Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev.* 12:3343-3356.

Das, A.K., Helps, N.R., Cohen, P.T., and Barford, D. 1996. Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* 15:6798-6809.

Davies, C., White, S.W., and Ramakrishnan, V. 1996. The crystal structure of ribosomal protein L14 reveals an important organizational component of the translational apparatus. *Structure.* 4:55-66.

Davies, D.R. 1990. The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* 19:189-215.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., and Rayment, I. 2000. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289:77-85.

Davies, J.F., Hostomska, Z., Hostomsky, Z., Jordan, S.R., and Matthews, D.A. 1991. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* 252:88-95.

Davies, J.F., Almassy, R.J., Hostomska, Z., Ferre, R.A., and Hostomsky, Z. 1994. 2.3 A crystal structure of the catalytic domain of DNA polymerase beta. *Cell* 76:1123-1133.

De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O., and Kim, S.H. 1993. Crystal structure of cyclin-dependent kinase 2. *Nature* 363:595-602.

de Vos, A.M., Ultsch, M., and Kossiakoff, A.A. 1992. Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science* 255:306-312.

Decanniere, K., Babu, A.M., Sandman, K., Reeve, J.N., and Heinemann, U. 2000. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J. Mol. Biol.* 303:35-47.

Deisenhofer, J. and Michel, H. 1989. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis*. *Biosci. Rep.* 9:383-419.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. 1984. X-ray structure analysis of a membrane protein complex. Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodopseudomonas viridis*. *J. Mol. Biol.* 180:385-398.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. 1985. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution. *Nature* 318:618-624.

Deisenhofer, J., Epp, O., Sinning, I., and Michel, H. 1995. Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. *J. Mol. Biol.* 246:429-457.

Demene, H., Jullian, N., Morellet, N., de Rocquigny, H., Cornille, F., Maigret, B., and Roques, B.P. 1994. Three-dimensional ^{1}H NMR structure of the nucleocapsid protein NCp10 of Moloney murine leukemia virus. *J. Biomol. NMR.* 4:153-170.

Dessen, A., Gupta, D., Sabesan, S., Brewer, C.F., and Sacchettini, J.C. 1995. X-ray crystal structure of the soybean agglutinin cross-linked with a biantennary analog of the blood group I carbohydrate antigen. *Biochemistry* 34:4933-4942.

Dietmann, S. and Holm, L. 2001. Identification of homology in protein structure classification. *Nat. Struct. Biol.* 8:953-957.

Dominguez, R., Freyzon, Y., Trybus, K.M., and Cohen, C. 1998. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: Visualization of the pre-power stroke state. *Cell* 94:559-571.

Donaldson, L.W., Petersen, J.M., Graves, B.J., and McIntosh, L.P. 1996. Solution structure of the ETS domain from murine Ets-1: A winged helix-turn-helix DNA binding motif. *EMBO J.* 15:125-134.

Downing, A.K., Driscoll, P.C., Gout, I., Salim, K., Zvelebil, M.J., and Waterfield, M.D. 1994. Three-dimensional solution structure of the pleckstrin homology domain from dynamin. *Curr. Biol.* 4:884-891.

Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. 1996. Crystal structures of a complexed and peptide-free membrane protein- binding domain: Molecular basis of peptide recognition by PDZ. *Cell* 85:1067-1076.

Doyle, D.A., Morais, C.J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R. 1998. The structure of the potassium channel: Molecular basis of K^{+} conduction and selectivity. *Science* 280:69-77.

Drennan, C.L., Huang, S., Drummond, J.T., Matthews, R.G., and Lidwig, M.L. 1994. How a protein binds B12: A 3.0 Å X-ray structure of B12-binding domains of methionine synthase. *Science* 266:1669-1674.

Dreusicke, D. and Schulz, G.E. 1986. The glycine rich loop of adenylate kinase forms a giant anion hole. *FEBS Lett* 208:301-304.

Drickamer, K. 1995. Increasing diversity of animal lectin structures. *Curr. Opin. Struct. Biol.* 5:612-616.

Drum, C.L., Yan, S.Z., Bard, J., Shen, Y.Q., Lu, D., Soelaiman, S., Grabarek, Z., Bohm, A., and Tang, W.J. 2002. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. *Nature* 415:396-402.

Dumas, J.J., Merithew, E., Sudharshan, E., Rajamani, D., Hayes, S., Lawe, D., Corvera, S., and Lambright, D.G. 2001. Multivalent endosome targeting by homodimeric EEA1. *Mol. Cell* 8:947-958.

Durocher, D. and Jackson, S.P. 2002. The FHA domain. *FEBS Lett.* 513:58-66.

Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon, S.J., and Yaffe, M.B. 2000. The molecular basis of FHA domain: Phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell* 6:1169-1182.

Dutzler, R., Campbell, E.B., Cadene, M., Chait, B.T., and MacKinnon, R. 2002. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* 415:287-294.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R., and Davies, D.R. 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 266:1981-1986.

Ealick, S.E., Cook, W.J., Vijay-Kumar, S., Carson, M., Nagabhushan, T.L., Trotta, P.P., and Bugg, C.E. 1991. Three-dimensional structure of recombinant human interferon-γ. *Science* 252:698-702.

Ebright, R.H. 2000. RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* 304:687-698.

Eck, M.J., Dhe-Paganon, S., Trub, T., Nolte, R.T., and Shoelson, S.E. 1996. Structure of the IRS-1 PTB domain bound to the juxtamembrane region of the insulin receptor. *Cell* 85:695-705.

Eggink, G., Engel, H., Wriend, G., Terpstra, P., and Witholt, B. 1990. Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases bases on one NAD and two FAD fingerprints. *J. Mol. Biol.* 212:135-142.

Eichinger, A., Beisel, H.G., Jacob, U., Huber, R., Medrano, F.J., Banbula, A., Potempa, J., Travis, J., and Bode, W. 1999. Crystal structure of gingipain R: An Arg-specific bacterial cysteine proteinase with a caspase-like fold. *EMBO J.* 18:5453-5462.

Eklund, H., Samma, J.P., Wallen, L., Branden, C.I., Akeson, A., and Jones, T.A. 1981. Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 Å resolution. *J. Mol. Biol.* 146:561-587.

Ellenberger, T. 1994. Getting a grip on DNA recognition: Structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. *Current Opinion in Structural Biology* 4:12-21.

Ellenberger, T.E., Brandl, C.J., Struhl, K., and Harrison, S.C. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: Crystal structure of the protein-DNA complex. *Cell* 71:1223-1237.

Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S.C. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev.* 8:970-980.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 347:203-206.

Eriksson, A.E., Cousens, L.S., Weaver, L.H., and Matthews, B.W. 1991. Three-dimensional structure of human basic fibroblast growth factor. *Proc. Natl. Acad. Sci. U.S.A* 88:3441-3445.

Ermler, U., Siddiqui, R.A., Cramm, R., and Friedrich, B. 1995. Crystal structure of the flavohemoglobin from *Alcaligenes eutrophus* at 1.75 Å resolution. *EMBO J.* 14:6067-6077.

Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366:483-487.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K., and Bairoch, A. 2002. The PROSITE database, its status in 2002. *Nucleic Acids Res.* 30:235-238.

Fan, Q.R., Long, E.O., and Wiley, D.C. 2001. Crystal structure of the human natural killer cell inhibitory receptor KIR2DL1-HLA-Cw4 complex. *Nat. Immunol.* 2:452-460.

Farooq, A., Chaturvedi, G., Mujtaba, S., Plotnikova, O., Zeng, L., Dhalluin, C., Ashton, R., and Zhou, M.M. 2001. Solution structure of ERK2 binding domain of MAPK phosphatase MKP-3: Structural insights into MKP-3 activation by ERK2. *Mol. Cell* 7:387-399.

Fass, D., Bogden, C.E., and Berger, J.M. 1999. Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands. *Nat. Struct. Biol.* 6:322-326.

Fauman, E.B., Cogswell, J.P., Lovejoy, B., Rocque, W.J., Holmes, W., Montana, V.G., Piwnica-Worms, H., Rink, M.J., and Saper, M.A. 1998. Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell* 93:617-625.

Favelyukis, S., Till, J.H., Hubbard, S.R., and Miller, W.T. 2001. Structure and autoregulation of the insulin-like growth factor 1 receptor kinase. *Nat. Struct. Biol.* 8:1058-1063.

Fedorov, A.A., Fedorov, E., Gertler, F., and Almo, S.C. 1999. Structure of EVH1, a novel proline-rich ligand-binding module involved in cytoskeletal dynamics and neural function. *Nat. Struct. Biol.* 6:661-665.

Feinberg, H., Mitchell, D.A., Drickamer, K., and Weis, W.I. 2001. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* 294:2163-2166.

Feng, J.A., Johnson, R.C., and Dickerson, R.E. 1994. Hin recombinase bound to DNA: The origin of specificity in major and minor groove interactions. *Science* 263:348-355.

Ferguson, A.D., Hofmann, E., Coulton, J.W., Diederichs, K., and Welte, W. 1998. Siderophore-mediated iron transport: Crystal structure of FhuA with bound lipopolysaccharide. *Science* 282:2215-2220.

Ferguson, A.D., Chakraborty, R., Smith, B.S., Esser, L., van der, H.D., and Deisenhofer, J. 2002. Structural basis of gating by the outer membrane transporter FecA. *Science* 295:1715-1719.

Ferguson, K.M., Lemmon, M.A., Schlessinger, J., and Sigler, P.B. 1994. Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin. *Cell* 79:199-209.

Fermi, G., Perutz, M.F., Shaanan, B., and Fourme, R. 1984. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* 175:159-174.

Fernandez, I., Ubach, J., Dulubova, I., Zhang, X., Sudhof, T.C., and Rizo, J. 1998. Three-dimensional structure of an evolutionarily conserved N-terminal domain of syntaxin 1A. *Cell* 94:841-849.

Ferre-D'Amare, A.R., Prendergast, G.C., Ziff, E.B., and Burley, S.K. 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363:38-45.

Ferre-D'Amare, A.R., Pognonec, P., Roeder, R.G., and Burley, S.K. 1994. Structure and function of the b/HLH/Z domain of USF. *EMBO J.* 13:180-189.

Findlay, W.A., Sonnichsen, F.D., and Sykes, B.D. 1994. Solution structure of the TR1C fragment of skeletal muscle troponin-C. *J. Biol. Chem.* 269:6773-6778.

Fischer, D., Wolfson, H., Lin, S.L., and Nussinov, R. 1994. Three-dimensional, sequence order-independent structural comparison of a serine protease against the crystallographic database reveals active site similarities: Potential implications to evolution and to protein folding. *Protein Sci.* 3:769-778.

Flaherty, K.M., DeLuca-Flaherty, C., and McKay, D.B. 1990. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* 346:623-628.

Flaherty, K.M., Zozulya, S., Stryer, L., and McKay, D.B. 1993. Three-dimensional structure of recoverin, a calcium sensor in vision. *Cell* 75:709-716.

Fletcher, C.M., Harrison, R.A., Lachmann, P.J., and Neuhaus, D. 1994. Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure.* 2:185-199.

Flower, D.R., North, A.C., and Sansom, C.E. 2000. The lipocalin protein family: Structural and sequence overview. *Biochim. Biophys. Acta* 1482:9-24.

Foster, M.P., Wuttke, D.S., Radhakrishnan, I., Case, D.A., Gottesfeld, J.M., and Wright, P.E. 1997. Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nat. Struct. Biol.* 4:605-608.

Franken, S.M., Rozeboom, H.J., Kalk, K.H., and Dijkstra, B.W. 1991. Crystal structure of haloalkane dehalogenase: An enzyme to detoxify halogenated alkanes. *EMBO J.* 10:1297-1302.

Franklin, M.C., Wang, J., and Steitz, T.A. 2001. Structure of the replicating complex of a polα family DNA polymerase. *Cell* 105:657-667.

Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A., and Wilson, I.A. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* 257:919-927.

Fremont, D.H., Crawford, F., Marrack, P., Hendrickson, W.A., and Kappler, J. 1998. Crystal structure of mouse H2-M. *Immunity.* 9:385-393.

Freund, C., Dotsch, V., Nishizawa, K., Reinherz, E.L., and Wagner, G. 1999. The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nat. Struct. Biol.* 6:656-660.

Frolow, F., Kalb, A.J., and Yariv, J. 1994. Structure of a unique twofold symmetric haem-binding site. *Nat. Struct. Biol.* 1:453-460.

Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J., and Stroud, R.M. 2000a. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290:481-486.

Fu, H., Subramanian, R.R., and Masters, S.C. 2000b. 14-3-3 proteins: Structure, function, and regulation. *Annu. Rev. Pharmacol. Toxicol.* 40:617-647.

Fukami, T.A., Yohda, M., Taguchi, H., Yoshida, M., and Miki, K. 2001. Crystal structure of chaperonin-60 from *Paracoccus denitrificans*. *J. Mol. Biol.* 312:501-509.

Fulop, V. and Jones, D.T. 1999. Beta propellers: Structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* 9:715-721.

Gajiwala, K.S. and Burley, S.K. 2000. Winged helix proteins. *Curr. Opin. Struct. Biol.* 10:110-116.

Gans, J.D. and Shalloway, D. 2001. Qmol: A program for molecular visualization on Windows-based PCs. *J. Mol. Graph. Model.* 19:557-9, 609.

Gao, G.F., Tormo, J., Gerth, U.C., Wyer, J.R., McMichael, A.J., Stuart, D.I., Bell, J.I., Jones, E.Y., and Jakobsen, B.K. 1997. Crystal structure of the complex between human CD8α(α) and HLA- A2. *Nature* 387:630-634.

Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E., and Wiley, D.C. 1996. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384:134-141.

Garcia, K.C., Degano, M., Pease, L.R., Huang, M., Peterson, P.A., Teyton, L., and Wilson, I.A. 1998. Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* 279:1166-1172.

Gardner, K.H., Anderson, S.F., and Coleman, J.E. 1995. Solution structure of the Kluyveromyces lactis LAC9 Cd2 Cys6 DNA-binding domain. *Nat. Struct. Biol.* 2:898-905.

Garman, S.C., Kinet, J.P., and Jardetzky, T.S. 1998. Crystal structure of the human high-affinity IgE receptor. *Cell* 95:951-961.

Garman, S.C., Wurzburg, B.A., Tarchevskaya, S.S., Kinet, J.P., and Jardetzky, T.S. 2000. Structure of the Fc fragment of human IgE bound to its high-affinity receptor Fc epsilonRI alpha. *Nature* 406:259-266.

Garrett, T.P., Saper, M.A., Bjorkman, P.J., Strominger, J.L., and Wiley, D.C. 1989. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature* 342:692-696.

Ghosh, G., van Duyne, G., Ghosh, S., and Sigler, P.B. 1995. Structure of NF-κB p50 homodimer bound to a κB site. *Nature* 373:303-310.

Gibbons, C., Montgomery, M.G., Leslie, A.G., and Walker, J.E. 2000. The structure of the central stalk in bovine F(1)-ATPase at 2.4 A resolution. *Nat. Struct. Biol.* 7:1055-1061.

Gibrat, J-F., Madej, T., and Bryant, S.H. 1996. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* 6:377-385.

Gibson, T.J., Thompson, J.D., and Heringa, J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. *FEBS Lett.* 324:361-366.

Glover, J.N. and Harrison, S.C. 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos- c-Jun bound to DNA. *Nature* 373:257-261.

Goger, M., Gupta, V., Kim, W.Y., Shigesada, K., Ito, Y., and Werner, M.H. 1999. Molecular insights into PEBP2/CBF beta-SMMHC associated acute leukemia revealed from the structure of PEBP2/CBF beta. *Nat. Struct. Biol.* 6:620-623.

Goldberg, J., Huang, H.B., Kwon, Y.G., Greengard, P., Nairn, A.C., and Kuriyan, J. 1995. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* 376:745-753.

Goldberg, J., Nairn, A.C., and Kuriyan, J. 1996. Structural basis for the autoinhibition of calcium/calmodulin-dependent protein kinase I. *Cell* 84:875-887.

Golden, B.L., Hoffman, D.W., Ramakrishnan, V., and White, S.W. 1993. Ribosomal protein S17: Characterization of the three-dimensional structure by ^{1}H and ^{15}N NMR. *Biochemistry* 32:12812-12820.

Gomis-Ruth, F.X., Kress, L.F., Kellermann, J., Mayr, I., Lee, X., Huber, R., and Bode, W. 1994. Refined 2.0 Å X-ray crystal structure of the snake venom zinc-endopeptidase adamalysin II. Primary and tertiary structure determination, refinement, molecular structure and comparison with astacin, collagenase and thermolysin. *J. Mol. Biol.* 239:513-544.

Gordeliy, V.I., Labahn, J., Moukhametzianov, R., Efremov, R., Granzin, J., Schlesinger, R., Buldt, G., Savopol, T., Scheidig, A.J., Klare, J.P., and Engelhard, M. 2002. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* 419:484-487.

Gorina, S. and Pavletich, N.P. 1996. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science* 274:1001-1005.

Graham, T.A., Weaver, C., Mao, F., Kimelman, D., and Xu, W. 2000. Crystal structure of a β-catenin/Tcf complex. *Cell* 103:885-896.

Graves, B.J., Hatada, M.H., Hendrickson, W.A., Miller, J.K., Madison, V.S., and Satow, Y. 1990. Structure of interleukin 1 alpha at 2.7-Å resolution. *Biochemistry* 29:2679-2684.

Graves, B.J., Crowther, R.L., Chandran, C., Rumberger, J.M., Li, S., Huang, K.S., Presky, D.H., Familletti, P.C., Wolitzky, B.A., and Burns, D.K. 1994. Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains. *Nature* 367:532-538.

Greenwald, J., Fischer, W.H., Vale, W.W., and Choe, S. 1999. Three-finger toxin fold for the extracellular ligand-binding domain of the type II activin receptor serine kinase. *Nat. Struct. Biol.* 6:18-22.

Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M., and Henderson, R. 1996. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* 259:393-421.

Grishin, N.V. 2001. KH domain: one motif, two folds. *Nucleic Acids Res.* 29:638-643.

Grobler, J.A. and Hurley, J.H. 1997. Similarity between C2 domain jaws and immunoglobulin CDRs. *Nat. Struct. Biol.* 4:261-262.

Grobler, J.A., Essen, L.O., Williams, R.L., and Hurley, J.H. 1996. C2 domain conformational changes in phospholipase C-delta 1. *Nat. Struct. Biol.* 3:788-795.

Groenen, L.C., Nice, E.C., and Burgess, A.W. 1994. Structure-function relationships for the EGF/TGF-alpha family of mitogens. *Growth Factors* 11:235-257.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. 1997. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386:463-471.

Groth, G. and Pohl, E. 2001. The structure of the chloroplast F1-ATPase at 3.2 Å resolution. *J. Biol. Chem.* 276:1345-1352.

Groves, M.R. and Barford, D. 1999. Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* 9:383-389.

Groves, M.R., Hanlon, N., Turowski, P., Hemmings, B.A., and Barford, D. 1999. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell* 96:99-110.

Gustafson, T.A., He, W., Craparo, A., Schaub, C.D., and O'Neill, T.J. 1995. Phosphotyrosine-dependent interaction of SHC and insulin receptor substrate 1 with the NPEY motif of the insulin receptor via a novel non-SH2 domain. *Mol. Cell Biol.* 15:2500-2508.

Hage, T., Sebald, W., and Reinemer, P. 1999. Crystal structure of the interleukin-4/receptor alpha chain complex reveals a mosaic binding interface. *Cell* 97:271-281.

Handel, T.M. and Domaille, P.J. 1996. Heteronuclear (^{1}H, ^{13}C, ^{15}N) NMR assignments and solution structure of the monocyte chemoattractant protein-1 (MCP-1) dimer. *Biochemistry* 35:6569-6584.

Hanks, S.K. and Hunter, T. 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J.* 9:576-596.

Hansen, J.L., Long, A.M., and Schultz, S.C. 1997. Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure.* 5:1109-1122.

Harris, L.J., Larson, S.B., Hasel, K.W., and McPherson, A. 1997. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry* 36:1581-1597.

Hart, P.J., Deep, S., Taylor, A.B., Shu, Z., Hinck, C.S., and Hinck, A.P. 2002. Crystal structure of the human TβR2 ectodomain—TGF-β3 complex. *Nat. Struct. Biol.* 9:203-208.

Hay, J.C. 2001. SNARE complex structure and function. *Exp. Cell Res.* 271:10-21.

He, X.M. and Carter, D.C. 1992. Atomic structure and chemistry of human serum albumin. *Nature* 358:209-215.

Henrick, K. and Thornton, J.M. 1998. PQS: A protein quaternary structure file server. *Trends Biochem. Sci.* 23:358-361.

Hershko, A. and Ciechanover, A. 1998. The ubiquitin system. *Annu. Rev. Biochem.* 67:425-479.

Herzberg, O. and James, M.N. 1988. Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 Å resolution. *J. Mol. Biol.* 203:761-779.

Hill, C.P., Osslund, T.D., and Eisenberg, D. 1993. The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. *Proc. Natl. Acad. Sci. U.S.A.* 90:5167-5171.

Hillier, B.J., Christopherson, K.S., Prehoda, K.E., Bredt, D.S., and Lim, W.A. 1999. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science* 284:812-815.

Hillig, R.C., Renault, L., Vetter, I.R., Drell, T., Wittinghofer, A., and Becker, J. 1999. The crystal structure of rna1p: A new fold for a GTPase-activating protein. *Mol.Cell* 3:781-791.

Hirabayashi, J. and Kasai, K. 1993. The family of metazoan metal-independent beta-galactoside-binding lectins: Structure, function and molecular evolution. *Glycobiology* 3:297-304.

Hirshberg, M., Stockley, R.W., Dodson, G., and Webb, M.R. 1997. The crystal structure of human rac1, a member of the rho-family complexed with a GTP analogue. *Nat. Struct. Biol.* 4:147-152.

Ho, J.X., Holowachuk, E.W., Norton, E.J., Twigg, P.D., and Carter, D.C. 1993. X-ray and primary structure of horse serum albumin (Equus caballus) at 0.27-nm resolution. *Eur. J. Biochem.* 215:205-212.

Ho, Y.S., Swenson, L., Derewenda, U., Serre, L., Wei, Y., Dauter, Z., Hattori, M., Adachi, T., Aoki, J., Arai, H., Inoue, K., and Derewenda, Z.S. 1997. Brain acetylhydrolase that inactivates platelet-activating factor is a G-protein-like trimer. *Nature* 385:89-93.

Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M.J., and Shoelson, S.E. 1998. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* 92:441-450.

Hofmann, K. and Bucher, P. 1995. The FHA domain: A putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem.Sci.* 20:347-349.

Hohenester, E., Maurer, P., Hohenadl, C., Timpl, R., Jansonius, J.N., and Engel, J. 1996. Structure of a novel extracellular Ca(2+)-binding module in BM-40. *Nat. Struct. Biol.* 3:67-73.

Hohenester, E., Sasaki, T., and Timpl, R. 1999. Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. *Nat. Struct. Biol.* 6:228-232.

Holden, H.M., Ito, M., Hartshorne, D.J., and Rayment, I. 1992. X-ray structure determination of telokin, the C-terminal domain of myosin light chain kinase, at 2.8 Å resolution. *J. Mol. Biol.* 227:840-851.

Holm, L. and Sander, C. 1993. Structural alignment of globins, phycocyanins and colicin A. *FEBS Lett* 315:301-306.

Holm, L. and Sander, C. 1996. Mapping the protein universe. *Science* 273:595-603.

Holm, L. and Sander, C. 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* 26:316-319.

Holmes, M.A. and Matthews, B.W. 1982. Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol.* 160:623-639.

Holmgren, A. and Branden, C.I. 1989. Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature* 342:248-251.

Hommel, U., Zurini, M., and Luyten, M. 1994. Solution structure of a cysteine rich domain of rat protein kinase C. *Nat. Struct. Biol.* 1:383-387.

Hoog, S.S., Smith, W.W., Qiu, X., Janson, C.A., Hellmig, B., McQueney, M.S., O'Donnell, K., O'Shannessy, D., DiLella, A.G., Debouck, C., and Abdel-Meguid, S.S. 1997. Active site cavity of herpesvirus proteases revealed by the crystal structure of herpes simplex virus protease/inhibitor complex. *Biochemistry* 36:14023-14029.

Hopfner, K.P., Karcher, A., Craig, L., Woo, T.T., Carney, J.P., and Tainer, J.A. 2001. Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* 105:473-485.

Houdusse, A., Kalabokis, V.N., Himmel, D., Szent-Gyorgyi, A.G., and Cohen, C. 1999. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: A novel conformation of the myosin head. *Cell* 97:459-470.

Housset, D., Habersetzer-Rochat, C., Astier, J.P., and Fontecilla-Camps, J.C. 1994. Crystal structure of toxin II from the scorpion Androctonus australis Hector refined at 1.3 Å resolution. *J. Mol. Biol.* 238:88-103.

Hu, S.H., Parker, M.W., Lei, J.Y., Wilce, M.C., Benian, G.M., and Kemp, B.E. 1994. Insights into autoregulation from the crystal structure of twitchin kinase. *Nature* 369:581-584.

Huang, X., Peng, J.W., Speck, N.A., and Bushweller, J.H. 1999. Solution structure of core binding factor beta and map of the CBF alpha binding site. *Nat. Struct. Biol.* 6:624-627.

Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., and Eck, M.J. 2000. Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nat. Struct. Biol.* 7:634-638.

Hubbard, S.R., Wei, L., Ellis, L., and Hendrickson, W.A. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* 372:746-754.

Huber, A.H. and Weis, W.I. 2001. The structure of the β-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by β-catenin. *Cell* 105:391-402.

Huber, A.H., Nelson, W.J., and Weis, W.I. 1997. Three-dimensional structure of the armadillo repeat region of β-catenin. *Cell* 90:871-882.

Huber, R., Schneider, M., Mayr, I., Muller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H., and Kayser, H. 1987. Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J. Mol. Biol.* 198:499-513.

Huber, R., Romisch, J., and Paques, E.P. 1990a. The crystal and molecular structure of human annexin V, an anticoagulant protein that binds to calcium and membranes. *EMBO J.* 9:3867-3874.

Huber, R., Schneider, M., Mayr, I., Romisch, J., and Paques, E.P. 1990b. The calcium binding sites in human annexin V by crystal structure analysis at 2.0 Å resolution. Implications for membrane binding and calcium channel activity. *FEBS Lett.* 275:15-21.

Huber, R., Berendes, R., Burger, A., Schneider, M., Karshikov, A., Luecke, H., Romisch, J., and Paques, E. 1992. Crystal and molecular structure of human annexin V after refinement. Implications for structure, membrane binding and ion channel formation of the annexin family of proteins. *J. Mol. Biol.* 223:683-704.

Hubscher, U., Maga, G., and Spadari, S. 2002. Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 71:133-163.

Hung, A.Y. and Sheng, M. 2002. PDZ domains: Structural modules for protein complex assembly. *J. Biol. Chem.* 277:5699-5702.

Hunt, J.F., Weaver, A.J., Landry, S.J., Gierasch, L., and Deisenhofer, J. 1996. The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature* 379:37-45.

Hunt, J.F., van der Vies, S.M., Henry, L., and Deisenhofer, J. 1997. Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. *Cell* 90:361-371.

Hunte, C., Koepke, J., Lange, C., Rossmanith, T., and Michel, H. 2000. Structure at 2.3 Å resolution of the cytochrome bc(1) complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Structure. Fold. Des* 8:669-684.

Hurley, J.H. and Misra, S. 2000. Signaling and subcellular targeting by membrane-binding domains. *Annu. Rev. Biophys. Biomol. Struct.* 29:49-79.

Hurley, J.H., Faber, H.R., Worthylake, D., Meadow, N.D., Roseman, S., Pettigrew, D.W., and Remington, S.J. 1993. Structure of the regulatory complex of *Escherichia coli* IIIGlc with glycerol kinase. *Science* 259:673-677.

Huxford, T., Huang, D.B., Malek, S., and Ghosh, G. 1998. The crystal structure of the IκBα/NF-κB complex reveals mechanisms of NF-κB inactivation. *Cell* 95:759-770.

Ibba, M. and Soll, D. 2000. Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* 69:617-650.

Irwin, M.J., Nyborg, J., Reid, B.R., and Blow, D.M. 1976. The crystal structure of tyrosyl-transfer RNA synthetase at 2-7 Å resolution. *J. Mol. Biol.* 105:577-586.

Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., and Knowles, P.F. 1991. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* 350:87-90.

Iverson, T.M., Luna-Chavez, C., Cecchini, G., and Rees, D.C. 1999. Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science* 284:1961-1966.

Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H. 1995. Structure at 2.8 Å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature* 376:660-669.

Iwata, S., Lee, J.W., Okada, K., Lee, J.K., Iwata, M., Rasmussen, B., Link, T.A., Ramaswamy, S., and Jap, B.K. 1998. Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. *Science* 281:64-71.

Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D. Jr., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., Hizi, A., Hughes, S.H., and Arnold, E. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U.S.A.* 90:6320-6324.

Jawad, Z. and Paoli, M. 2002. Novel sequences propel familiar folds. *Structure.(Camb.)* 10:447-454.

Jedrzejas, M.J., Chander, M., Setlow, P., and Krishnasamy, G. 2000. Mechanism of catalysis of the cofactor-independent phosphoglycerate mutase from *Bacillus stearothermophilus*. Crystal structure of the complex with 2-phosphoglycerate. *J. Biol. Chem.* 275:23146-23153.

Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N.P. 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* 376:313-320.

Jeon, Y.H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A., and Kyogoku, Y. 1995. Solution structure of the activator contact domain of the RNA polymerase α subunit. *Science* 270:1495-1497.

Jia, X., Grove, A., Ivancic, M., Hsu, V.L., Geiduscheck, E.P., and Kearns, D.R. 1996. Structure of the *Bacillus subtilis* phage SPO1-encoded type II DNA-binding protein TF1 in solution. *J. Mol. Biol.* 263:259-268.

Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B.T., and MacKinnon, R. 2002a. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* 417:515-522.

Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B.T., and MacKinnon, R. 2002b. The open pore conformation of potassium channels. *Nature* 417:523-526.

Johansson, K., Ramaswamy, S., Ljungcrantz, C., Knecht, W., Piskur, J., Munch-Petersen, B., Eriksson, S., and Eklund, H. 2001. Structural basis for substrate specificities of cellular deoxyribonucleoside kinases. *Nat. Struct. Biol.* 8:616-620.

Johnson, B.A., Stevens, S.P., and Williamson, J.M. 1994. Determination of the three-dimensional structure of margatoxin by ^{1}H, ^{13}C, ^{15}N triple-resonance nuclear magnetic resonance spectroscopy. *Biochemistry* 33:15061-15070.

Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K., and Stuart, D.I. 1992. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature* 360:232-239.

Jones, E.Y., Harlos, K., Bottomley, M.J., Robinson, R.C., Driscoll, P.C., Edwards, R.M., Clements, J.M., Dudgeon, T.J., and Stuart, D.I. 1995. Crystal structure of an integrin-binding fragment of vascular cell adhesion molecule-1 at 1.8 Å resolution. *Nature* 373:539-544.

Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., and Krauss, N. 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* 411:909-917.

Jordan, S.R. and Pabo, C.O. 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. *Science* 242:893-899.

Jormakka, M., Tornroth, S., Byrne, B., and Iwata, S. 2002. Molecular basis of proton motive force generation: Structure of formate dehydrogenase-N. *Science* 295:1863-1868.

Josephson, K., Logsdon, N.J., and Walter, M.R. 2001. Crystal structure of the IL-10/IL-10R1 complex reveals a shared receptor binding site. *Immunity* 15:35-46.

Joshua-Tor, L., Xu, H.E., Johnston, S.A., and Rees, D.C. 1995. Crystal structure of a conserved protease that binds DNA: The bleomycin hydrolase, Gal6. *Science* 269:945-950.

Joyce, C.M. and Steitz, T.A. 1994. Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63:777-822.

Kabsch, W. and Holmes, K.C. 1995. The actin fold. *FASEB J.* 9:167-174.

Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F., and Holmes, K.C. 1990. Atomic structure of the actin:DNase I complex. *Nature* 347:37-44.

Kajava, A.V. 1998. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* 277:519-527.

Kakuta, Y., Pedersen, L.G., Carter, C.W., Negishi, M., and Pedersen, L.C. 1997. Crystal structure of estrogen sulphotransferase. *Nat.Struct.Biol.* 4:904-908.

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., and Burley, S.K. 2001. Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell* 106:71-81.

Kamphuis, I.G., Kalk, K.H., Swarte, M.B., and Drenth, J. 1984. Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.* 179:233-256.

Kang, C., Chan, R., Berger, I., Lockshin, C., Green, L., Gold, L., and Rich, A. 1995. Crystal structure of the T4 regA translational regulator protein at 1.9 Å resolution. *Science* 268:1170-1173.

Karplus, P.A., Daniels, M.J., and Herriott, J.R. 1991. Atomic structure of ferredoxin-$NADP^{+}$ reductase: prototype for a structurally novel flavoenzyme family. *Science* 251:60-66.

Kataoka, M. and Kamikubo, H. 2000. Structures of photointermediates and their implications for the proton pump mechanism. *Biochim. Biophys. Acta* 1460:166-176.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., and Morikawa, K. 1990. Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* 347:306-309.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., and Morikawa, K. 1992. Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* 223:1029-1052.

Katzin, B.J., Collins, E.J., and Robertus, J.D. 1991. Structure of ricin A-chain at 2.5 Å. *Proteins* 10:251-259.

Kavanaugh, W.M. and Williams, L.T. 1994. An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science* 266:1862-1865.

Ke, H.M., Zhang, Y.P., Liang, J.Y., and Lipscomb, W.N. 1991. Crystal structure of the neutral form of fructose-1,6-bisphosphatase complexed with the product fructose 6-phosphate at 2.1-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 88:2989-2993.

Keenan, R.J., Freymann, D.M., Walter, P., and Stroud, R.M. 1998. Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* 94:181-191.

Keskin, O., Bahar, I., Flatow, D., Covell, D.G., and Jernigan, R.L. 2002. Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry* 41:491-501.

Kharrat, A., Macias, M.J., Gibson, T.J., Nilges, M., and Pastore, A. 1995. Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J.* 14:3572-3584.

Kieffer, B., Driscoll, P.C., Campbell, I.D., Willis, A.C., van der Merwe, P.A., and Davis, S.J. 1994. Three-dimensional solution structure of the extracellular region of the complement regulatory protein CD59, a new cell-surface protein domain related to snake venom neurotoxins. *Biochemistry* 33:4471-4482.

Kim, C.A., Gingery, M., Pilpa, R.M., and Bowie, J.U. 2002. The SAM domain of polyhomeotic forms a helical polymer. *Nat. Struct. Biol.* 9:453-457.

Kim, E.E. and Wyckoff, H.W. 1991. Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J. Mol. Biol.* 218:449-464.

Kim, J.L., Nikolov, D.B., and Burley, S.K. 1993a. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365:520-527.

Kim, S., Narayana, S.V., and Volanakis, J.E. 1995a. Crystal structure of a complement factor D mutant expressing enhanced catalytic activity. *J. Biol. Chem.* 270:24399-24405. [published erratum appears in *J. Biol. Chem.* 270:31414]

Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. 1993b. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365:512-520.

Kim, Y., Eom, S.H., Wang, J., Lee, D.S., Suh, S.W., and Steitz, T.A. 1995b. Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* 376:612-616.

Kimple, R.J., Kimple, M.E., Betts, L., Sondek, J., and Siderovski, D.P. 2002. Structural determinants for GoLoco-induced inhibition of nucleotide release by Galpha subunits. *Nature* 416:878-881.

Kimura, Y., Vassylyev, D.G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T., and Fujiyoshi, Y. 1997. Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* 389:206-211.

Kirsch, T., Sebald, W., and Dreyer, M.K. 2000. Crystal structure of the BMP-2-BRIA ectodomain complex. *Nat. Struct. Biol.* 7:492-496.

Kissinger, C.R., Liu, B.S., Martin-Blanco, E., Kornberg, T.B., and Pabo, C.O. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: A framework for understanding homeodomain-DNA interactions. *Cell* 63:579-590.

Kissinger, C.R., Parge, H.E., Knighton, D.R., Lewis, C.T., Pelletier, L.A., Tempczyk, A., Kalish, V.J., Tucker, K.D., Showalter, R.E., and Moomaw, E.W. 1995. Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* 378:641-644.

Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., and Pabo, C.O. 1994. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* 77:21-32.

Kleywegt, G.J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K., and Jones, T.A. 1994. Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure.* 2:1241-1258.

Kloek, A.P., Yang, J., Mathews, F.S., and Goldberg, D.E. 1993. Expression, characterization, and crystallization of oxygen-avid Ascaris hemoglobin domains. *J. Biol. Chem.* 268:17669-17671.

Knighton, D.R., Zheng, J.H., Ten Eyck, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S., and Sowadski, J.M. 1991. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253:407-414.

Knofel, T. and Strater, N. 1999. X-ray structure of the *Escherichia coli* periplasmic 5′-nucleotidase containing a dimetal catalytic site. *Nat. Struct. Biol.* 6:448-453.

Knowlton, J.R., Johnston, S.C., Whitby, F.G., Realini, C., Zhang, Z., Rechsteiner, M., and Hill, C.P. 1997. Structure of the proteasome activator REGalpha (PA28alpha). *Nature* 390:639-643.

Kobe, B. 1999. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. *Nat. Struct. Biol.* 6:388-397.

Kobe, B. and Deisenhofer, J. 1993. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* 366:751-756.

Kobe, B. and Deisenhofer, J. 1995. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* 374:183-186.

Kobe, B., Gleichmann, T., Horne, J., Jennings, I.G., Scotney, P.D., and Teh, T. 1999. Turn up the HEAT. *Structure. Fold. Des.* 7:R91-R97.

Kobe, B. and Kajava, A.V. 2001. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* 11:725-732.

Koepke, J., Hu, X., Muenke, C., Schulten, K., and Michel, H. 1996. The crystal structure of the light-harvesting complex II (B800-850) from Rhodospirillum molischianum. *Structure.* 4:581-597.

Kohda, D. and Inagaki, F. 1992. Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions. *Biochemistry* 31:11928-11939.

Kolbe, M., Besir, H., Essen, L.O., and Oesterhelt, D. 2000. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science* 288:1390-1396.

Kollmar, M., Durrwang, U., Kliche, W., Manstein, D.J., and Kull, F.J. 2002. Crystal structure of the motor domain of a class-I myosin. *EMBO J.* 21:2517-2525.

Kong, X.P., Onrust, R., O'Donnell, M., and Kuriyan, J. 1992. Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell* 69:425-437.

Koradi, R., Billeter, M., and Wuthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14:51-32.

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. 2000. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* 405:914-919.

Kraulis, P.J. 1991. MolScript v2.1. *J. Appl. Cryst.* 24:946-950.

Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M., and Kuriyan, J. 1994. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79:1233-1243.

Kunishima, N., Shimada, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., Nakanishi, S., Jingami, H., and Morikawa, K. 2000. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* 407:971-977.

Kuriyan, J. and Cowburn, D. 1997. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* 26:259-288.

Kurumbail, R.G., Stevens, A.M., Gierse, J.K., McDonald, J.J., Stegeman, R.A., Pak, J.Y., Gildehaus, D., Miyashiro, J.M., Penning, T.D., Seibert, K., Isakson, P.C., and Stallings, W.C. 1996. Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* 384:644-648.

Kutateladze, T. and Overduin, M. 2001. Structural mechanism of endosome docking by the FYVE domain. *Science* 291:1793-1796.

Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E., and Sigler, P.B. 1996. The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature* 379:311-319.

Lamers, M.B., Antson, A.A., Hubbard, R.E., Scott, R.K., and Williams, D.H. 1999. Structure of the protein tyrosine kinase domain of C-terminal Src kinase (CSK) in complex with staurosporine. *J. Mol. Biol.* 285:713-725.

Lancaster, C.R., Kroger, A., Auer, M., and Michel, H. 1999. Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. *Nature* 402:377-385.

Lanyi, J.K. 1999. Progress toward an explicit mechanistic model for the light-driven pump, bacteriorhodopsin. *FEBS Lett.* 464:103-107.

Lapthorn, A.J., Harris, D.C., Littlejohn, A., Lustbader, J.W., Canfield, R.E., Machin, K.J., Morgan, F.J., and Isaacs, N.W. 1994. Crystal structure of human chorionic gonadotropin. *Nature* 369:455-461.

Laudet, V., Stehelin, D., and Clevers, H. 1993. Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res.* 21:2493-2501.

Lawrence, C.M., Ray, S., Babyonyshev, M., Galluser, R., Borhani, D.W., and Harrison, S.C. 1999. Crystal structure of the ectodomain of human transferrin receptor. *Science* 286:779-782.

Lawson, C.L., van Montfort, R., Strokopytov, B., Rozeboom, H.J., Kalk, K.H., de Vries, G.E., Penninga, D., Dijkhuizen, L., and Dijkstra, B.W. 1994. Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from Bacillus circulans strain 251 in a maltose- dependent crystal form. *J. Mol. Biol.* 236:590-600.

Leahy, D.J., Axel, R., and Hendrickson, W.A. 1992a. Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution. *Cell* 68:1145-1162.

Leahy, D.J., Hendrickson, W.A., Aukhil, I., and Erickson, H.P. 1992b. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* 258:987-991.

Lee, J.O., Rieu, P., Arnaout, M.A., and Liddington, R. 1995. Crystal structure of the A domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell* 80:631-638.

Lee, J.O., Russo, A.A., and Pavletich, N.P. 1998. Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature* 391:859-865.

Lee, J.O., Yang, H., Georgescu, M.M., Di Cristofano, A., Maehama, T., Shi, Y., Dixon, J.E., Pandolfi, P., and Pavletich, N.P. 1999. Crystal structure of the PTEN tumor suppressor: Implications for its phosphoinositide phosphatase activity and membrane association. *Cell* 99:323-334.

Lesk, A.M. 1995. NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* 5:775-783.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., and Bork, P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* 30:242-244.

Li, J., Lee, G.I., Van Doren, S.R., and Walker, J.C. 2000. The FHA domain mediates phosphoprotein interactions. *J. Cell Sci.* 113 Pt 23:4143-4149.

Li, J., Williams, B.L., Haire, L.F., Goldberg, M., Wilker, E., Durocher, D., Yaffe, M.B., Jackson, S.P., and Smerdon, S.J. 2002a. Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. *Mol. Cell* 9:1045-1054.

Li, P., Willie, S.T., Bauer, S., Morris, D.L., Spies, T., and Strong, R.K. 1999. Crystal structure of the MHC class I homolog MIC-A, a gammadelta T cell ligand. *Immunity.* 10:577-584.

Li, P., Morris, D.L., Willcox, B.E., Steinle, A., Spies, T., and Strong, R.K. 2001. Complex structure of the activating immunoreceptor NKG2D and its MHC class I-like ligand MICA. *Nat. Immunol.* 2:443-451.

Li, P., McDermott, G., and Strong, R.K. 2002b. Crystal structures of RAE-1β and its complex with the activating immunoreceptor NKG2D. *Immunity.* 16:77-86.

Li, T., Stark, M.R., Johnson, A.D., and Wolberger, C. 1995. Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science* 270:262-269 [published erratum appears in *Science* 270:1105].

Liao, D.I. and Remington, S.J. 1990. Structure of wheat serine carboxypeptidase II at 3.5-A resolution. A new class of serine proteinase. *J. Biol. Chem.* 265:6528-6531.

Liao, D.I., Kapadia, G., Ahmed, H., Vasta, G.R., and Herzberg, O. 1994. Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* 91:1428-1432.

Liao, H., Byeon, I.J., and Tsai, M.D. 1999. Structure and function of a new phosphopeptide-binding domain containing the FHA2 of Rad53. *J. Mol. Biol.* 294:1041-1049.

Lima, C.D., Wang, J.C., and Mondragon, A. 1994. Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* 367:138-146.

Lindahl, M., Svensson, L.A., Liljas, A., Sedelnikova, S.E., Eliseikina, I.A., Fomenkova, N.P., Nevskaya, N., Nikonov, S.V., Garber, M.B., and Muranova, T.A. 1994. Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* 13:1249-1254.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. 2001. Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication. *Cell* 107:91-102.

Liu, D., Bienkowska, J., Petosa, C., Collier, R.J., Fu, H., and Liddington, R. 1995. Crystal structure of the zeta isoform of the 14-3-3 protein. *Nature* 376:191-194.

Livnah, O., Stura, E.A., Middleton, S.A., Johnson, D.L., Jolliffe, L.K., and Wilson, I.A. 1999. Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* 283:987-990.

Lo, C.L., Brenner, S.E., Hubbard, T.J., Chothia, C., and Murzin, A.G. 2002. SCOP database in 2002: Refinements accommodate structural genomics. *Nucleic Acids Res.* 30:264-267.

Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H., and Rini, J.M. 1993. X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-Å resolution. *J. Biol. Chem.* 268:27034-27038.

Locher, K.P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J.P., and Moras, D. 1998. Transmembrane signaling across the ligand-gated FhuA receptor: Crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell* 95:771-778.

Locher, K.P., Lee, A.T., and Rees, D.C. 2002. The *E. coli* BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science* 296:1091-1098.

Lodi, P.J., Garrett, D.S., Kuszewski, J., Tsang, M.L., Weatherbee, J.A., Leonard, W.J., Gronenborn, A.M., and Clore, G.M. 1994. High-resolution solution structure of the β chemokine hMIP-1β by multidimensional NMR. *Science* 263:1762-1767.

Lohi, O., Poussu, A., Mao, Y., Quiocho, F., and Lehto, V.P. 2002. VHS domain — a longshoreman of vesicle lines. *FEBS Lett.* 513:19-23.

Louie, G.V. and Brayer, G.D. 1990. High-resolution refinement of yeast iso-1-cytochrome c and comparisons with other eukaryotic cytochromes c. *J. Mol. Biol.* 214:527-555.

Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R., and Wright, P.E. 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* 376:791-795.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. 1995. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268:533-539.

Lubkowski, J., Bujacz, G., Boque, L., Domaille, P.J., Handel, T.M., and Wlodawer, A. 1997. The structure of MCP-1 in two crystal forms provides a rare example of variable quaternary interactions. *Nat. Struct. Biol.* 4:64-69.

Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P., and Lanyi, J.K. 1999. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* 291:899-911.

Luecke, H., Schobert, B., Lanyi, J.K., Spudich, E.N., and Spudich, J.L. 2001. Crystal structure of sensory rhodopsin II at 2.4 angstroms: Insights into color tuning and transducer interaction. *Science* 293:1499-1503.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352:497-505.

Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., and Saenger, W. 1998. Crystal structure of human arylsulfatase A: The aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* 37:3654-3664.

Lux, S.E., John, K.M., and Bennett, V. 1990. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* 344:36-42.

Ma, P.C., Rould, M.A., Weintraub, H., and Pabo, C.O. 1994. Crystal structure of MyoD bHLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77:451-459.

Macias, M.J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M., and Oschkinat, H. 1994. Structure of the pleckstrin homology domain from beta-spectrin. *Nature* 369:675-677.

Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. 1996. Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* 382:646-649.

Macias, M.J., Wiesner, S., and Sudol, M. 2002. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* 513:30-37.

Madden, D.R., Gorga, J.C., Strominger, J.L., and Wiley, D.C. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* 70:1035-1048.

Madej, T., Gibrat, J-F., and Bryant, S.H. 1995. Threading a database of protein cores. *Protein Struct. Funct. Genet.* 23:356-369.

Maffucci, T. and Falasca, M. 2001. Specificity in pleckstrin homology (PH) domain membrane targeting: a role for a phosphoinositide-protein co-operative mechanism. *FEBS Lett.* 506:173-179.

Main, A.L., Harvey, T.S., Baron, M., Boyd, J., and Campbell, I.D. 1992. The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions. *Cell* 71:671-678.

Malhotra, A., Severinova, E., and Darst, S.A. 1996. Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell* 87:127-136.

Malkowski, M.G., Wu, J.Y., Lazar, J.B., Johnson, P.H., and Edwards, B.F. 1995. The crystal structure of recombinant human neutrophil-activating peptide-2 (M6L) at 1.9-Å resolution. *J.Biol.Chem.* 270:7077-7087.

Mande, S.C., Mehra, V., Bloom, B.R., and Hol, W.G. 1996. Structure of the heat shock protein chaperonin-10 of *Mycobacterium leprae*. *Science* 271:203-207. [published erratum appears in *Science* 271:1655]

Manning, G., Plowman, G.D., Hunter, T., and Sudarsanam, S. 2002a. Evolution of protein kinase signaling from yeast to man. *Trends Biochem.Sci.* 27:514-520.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. 2002b. The protein kinase complement of the human genome. *Science* 298:1912-1934.

Mao, Y., Nickitenko, A., Duan, X., Lloyd, T.E., Wu, M.N., Bellen, H., and Quiocho, F.A. 2000. Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell* 100:447-456.

Marino, M., Braun, L., Cossart, P., and Ghosh, P. 1999. Structure of the lnlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. *Mol. Cell* 4:1063-1072.

Marmorstein, R. and Harrison, S.C. 1994. Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a Zn_2Cys_6 binuclear cluster. *Genes Dev.* 8:2504-2512.

Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C. 1992. DNA recognition by GAL4: Structure of a protein-DNA complex. *Nature* 356:408-414.

Marsden, I., Jin, C., and Liao, X. 1998. Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. *J. Mol. Biol.* 278:293-299.

Martin, G., Keller, W., and Doublie, S. 2000. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J.* 19:4193-4203.

Mathews, F.S., Argos, P., and Levine, M. 1972. The structure of cytochrome b5 at 2.0 Angstrom resolution. *Cold Spring Harb. Symp.Quant. Biol.* 36:387-395.

Matthews, D.A., Smith, W.W., Ferre, R.A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J.E., Janson, C.A., McElroy, H.E., and Gribskov, C.L. 1994. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* 77:761-771.

Maxwell, K.F., Powell, M.S., Hulett, M.D., Barton, P.A., McKenzie, I.F., Garrett, T.P., and Hogarth, P.M. 1999. Crystal structure of the human leukocyte Fc receptor, Fc gammaRIIa. *Nat. Struct. Biol.* 6:437-442.

Mayer, B.J. 2001. SH3 domains: Complexity in moderation. *J. Cell Sci.* 114:1253-1263.

McDonald, N.Q., Lapatto, R., Murray-Rust, J., Gunning, J., Wlodawer, A., and Blundell, T.L. 1991. New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor. *Nature* 354:411-414.

McLaughlin, P.J., Gooch, J.T., Mannherz, H.G., and Weeds, A.G. 1993. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. *Nature* 364:685-692.

McTigue, M.A., Wickersham, J.A., Pinko, C., Showalter, R.E., Parast, C.V., Tempczyk-Russell, A., Gehring, M.R., Mroczkowski, B., Kan, C.C., Villafranca, J.E., and Appelt, K. 1999. Crystal structure of the kinase domain of human vascular endothelial growth factor receptor 2: A key enzyme in angiogenesis. *Structure. Fold. Des* 7:319-330.

Meador, W.E., Means, A.R., and Quiocho, F.A. 1992. Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science* 257:1251-1255.

Meinke, G. and Sigler, P.B. 1999. DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B. *Nat. Struct. Biol.* 6:471-477.

Merritt, E.A. and Bacon, D.J. 1997. Raster3D: Photorealistic molecular graphics. *Meth. Enzymol.* 277:505-524.

Merritt, E.A., Sarfaty, S., Chang, T.T., Palmer, L.M., Jobling, M.G., Holmes, R.K., and Hol, W.G. 1995. Surprising leads for a cholera toxin receptor-binding antagonist: Crystallographic studies of CTB mutants. *Structure.* 3:561-570.

Meyer, J.E., Hofnung, M., and Schulz, G.E. 1997. Structure of maltoporin from *Salmonella typhimurium* ligated with a nitrophenyl-maltotrioside. *J. Mol. Biol.* 266:761-775.

Misra, S. and Hurley, J.H. 1999. Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. *Cell* 97:657-666.

Misra, S., Beach, B.M., and Hurley, J.H. 2000. Structure of the VHS domain of human Tom1 (target of myb 1): Insights into interactions with proteins and membranes. *Biochemistry* 39:11282-11290.

Misra, S., Puertollano, R., Kato, Y., Bonifacino, J.S., and Hurley, J.H. 2002. Structural basis for acidic-cluster-dileucine sorting-signal recognition by VHS domains. *Nature* 415:933-937.

Misura, K.M., May, A.P., and Weis, W.I. 2000a. Protein-protein interactions in intracellular membrane fusion. *Curr. Opin. Struct. Biol.* 10:662-671.

Misura, K.M., Scheller, R.H., and Weis, W.I. 2000b. Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex. *Nature* 404:355-362.

Mitchell, D.T., Kitto, G.B., and Hackert, M.L. 1995. Structural analysis of monomeric hemichrome and dimeric cyanomet hemoglobins from Caudina arenicola. *J. Mol. Biol.* 251:421-431.

Mizuguchi, K., Deane, C.M., Blundell, T.L., and Overington, J.P. 1998. HOMSTRAD: A database of protein structure alignments for homologous families. *Protein Sci.* 7:2469-2471.

Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M., and Kuriyan, J. 2000. Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *J. Mol. Biol.* 296:1215-1223.

Mohammadi, M., Schlessinger, J., and Hubbard, S.R. 1996. Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell* 86:577-587.

Mondragon, A. and DiGate, R. 1999. The structure of *Escherichia coli* DNA topoisomerase III. *Structure. Fold. Des* 7:1373-1383.

Mongkolsapaya, J., Grimes, J.M., Chen, N., Xu, X.N., Stuart, D.I., Jones, E.Y., and Screaton, G.R. 1999. Structure of the TRAIL-DR5 complex reveals mechanisms conferring specificity in apoptotic initiation. *Nat. Struct. Biol.* 6:1048-1053.

Montoya, G., Svensson, C., Luirink, J., and Sinning, I. 1997. Crystal structure of the NG domain from the signal-recognition particle receptor FtsY. *Nature* 385:365-368.

Moore, B.E. and Perez, V.J. 1967. Specific acidic proteins of the nervous system. *In* Physiological and Biochemical Aspects of Nervous Integration (F.D. Carlson, ed.), pp. 343-359. Prentice Hall, Englewood Cliffs, N.J.

Moraes, T.F., Edwards, R.A., McKenna, S., Pastushok, L., Xiao, W., Glover, J.N., and Ellison, M.J. 2001. Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13. *Nat. Struct. Biol.* 8:669-673.

Morais Cabral, J.H., Petosa, C., Sutcliffe, M.J., Raza, S., Byron, O., Poy, F., Marfatia, S.M., Chishti, A.H., and Liddington, R.C. 1996. Crystal structure of a PDZ domain. *Nature* 382:649-652.

Morais Cabral, J.H., Jackson, A.P., Smith, C.V., Shikotra, N., Maxwell, A., and Liddington, R.C. 1997. Crystal structure of the breakage-reunion domain of DNA gyrase. *Nature* 388:903-906.

Moras, D., Olsen, K.W., Sabesan, M.N., Buehner, M., Ford, G.C., and Rossmann, M.G. 1975. Studies of asymmetry in the three-dimensional structure of lobster D-glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.* 250:9137-9162.

Muller, C.W. and Herrmann, B.G. 1997. Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* 389:884-888.

Muller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L., and Harrison, S.C. 1995. Structure of the NF-κB p50 homodimer bound to DNA. *Nature* 373:311-317.

Muller, Y.A., Ultsch, M.H., and de Vos, A.M. 1996. The crystal structure of the extracellular domain of human tissue factor refined to 1.7 Å resolution. *J. Mol. Biol.* 256:144-159.

Muller-Dieckmann, H.J. and Schulz, G.E. 1994. The structure of uridylate kinase with its substrates, showing the transition state geometry. *J. Mol. Biol.* 236:361-367.

Munson, M., Chen, X., Cocina, A.E., Schultz, S.M., and Hughson, F.M. 2000. Interactions within the yeast t-SNARE Sso1p that control SNARE complex assembly. *Nat. Struct. Biol.* 7:894-902.

Murakami, K.S., Masuda, S., and Darst, S.A. 2002a. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296:1280-1284.

Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. 2002b. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* 419:587-593.

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., and Fujiyoshi, Y. 2000. Structural determinants of water permeation through aquaporin-1. *Nature* 407:599-605.

Murray-Rust, J., McDonald, N.Q., Blundell, T.L., Hosang, M., Oefner, C., Winkler, F., and Bradshaw, R.A. 1993. Topological similarities in TGF-β2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure.* 1:153-159.

Murzin, A.G. 1993. OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. *EMBO J.* 12:861-867.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536-540.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. 1992. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* 359:851-855.

Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., and Katunuma, N. 1991. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: The structural basis for its specificity. *EMBO J.* 10:2321-2330.

Nagar, B., Jones, R.G., Diefenbach, R.J., Isenman, D.E., and Rini, J.M. 1998. X-ray crystal structure of C3d: A C3 fragment and ligand for complement receptor 2. *Science* 280:1277-1281.

Nagata, T., Gupta, V., Sorce, D., Kim, W.Y., Sali, A., Chait, B.T., Shigesada, K., Ito, Y., and Werner, M.H. 1999. Immunoglobulin motif DNA recognition and heterodimerization of the PEBP2/CBF Runt domain. *Nat. Struct. Biol.* 6:615-619.

Nakayama, S. and Kretsinger, R.H. 1994. Evolution of the EF-hand family of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 23:473-507.

Nam, H.J., Poy, F., Krueger, N.X., Saito, H., and Frederick, C.A. 1999. Crystal structure of the tandem phosphatase domains of RPTP LAR. *Cell* 97:449-457.

Narayana, S.V., Carson, M., el Kabbani, O., Kilpatrick, J.M., Moore, D., Chen, X., Bugg, C.E., Volanakis, J.E., and DeLucas, L.J. 1994. Structure of human factor D. A complement system protein at 2.0 Å resolution. *J. Mol. Biol.* 235:695-708.

Nardini, M. and Dijkstra, B.W. 1999. α/β hydrolase fold enzymes: The family keeps growing. *Curr. Opin. Struct. Biol.* 9:732-737.

Nassar, N., Hoffman, G.R., Manor, D., Clardy, J.C., and Cerione, R.A. 1998. Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP. *Nat. Struct. Biol.* 5:1047-1052.

Navia, M.A., Fitzgerald, P.M., McKeever, B.M., Leu, C.T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., and Springer, J.P. 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 337:615-620.

Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S.D., Inouye, M., and Montelione, G.T. 1994. Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: Identification of a binding epitope for DNA. *Proc.Natl.Acad.Sci.U.S.A.* 91:5114-5118.

Nichols, M.D., DeAngelis, K., Keck, J.L., and Berger, J.M. 1999. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J.* 18:6177-6188.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377:119-128.

Nishizawa, K., Freund, C., Li, J., Wagner, G., and Reinherz, E.L. 1998. Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proc. Natl. Acad. Sci. U.S.A.* 95:14897-14902.

Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B.F., and Nyborg, J. 1995. Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* 270:1464-1472.

Noble, M.E., Cleasby, A., Johnson, L.N., Egmond, M.R., and Frenken, L.G. 1993. The crystal structure of triacylglycerol lipase from *Pseudomonas glumae* reveals a partially redundant catalytic aspartate. *FEBS Lett.* 331:123-128.

Noble, M.E., Endicott, J.A., Brown, N.R., and Johnson, L.N. 1997. The cyclin box fold: Protein recognition in cell-cycle and transcription control. *Trends Biochem.Sci.* 22:482-487.

Noel, J.P., Hamm, H.E., and Sigler, P.B. 1993. The 2.2 A crystal structure of transducin-α complexed with GTPγS. *Nature* 366:654-663.

Norman, D.G., Barlow, P.N., Baron, M., Day, A.J., Sim, R.B., and Campbell, I.D. 1991. Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.* 219:717-725.

O'Callaghan, C.A., Tormo, J., Willcox, B.E., Braud, V.M., Jakobsen, B.K., Stuart, D.I., McMichael, A.J., Bell, J.I., and Jones, E.Y. 1998. Structural features impose tight peptide binding specificity in the nonclassical MHC molecule HLA-E. *Mol. Cell* 1:531-541.

Obsil, T., Ghirlando, R., Klein, D.C., Ganguly, S., and Dyda, F. 2001. Crystal structure of the 14-3-3ζ:serotonin N-acetyltransferase complex. A role for scaffolding in enzyme regulation. *Cell* 105:257-267.

Oefner, C., D'Arcy, A., Winkler, F.K., Eggimann, B., and Hosang, M. 1992. Crystal structure of human platelet-derived growth factor BB. *EMBO J.* 11:3921-3926.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. 1994. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* 79:639-648.

Oinonen, C. and Rouvinen, J. 2000. Structural comparison of Ntn-hydrolases. *Protein Sci.* 9:2329-2337.

Olah, G.A., Mitchell, R.D., Sosnick, T.R., Walsh, D.A., and Trewhella, J. 1993. Solution structure of the cAMP-dependent protein kinase catalytic subunit and its contraction upon binding the protein kinase inhibitor peptide. *Biochemistry* 32:3649-3657.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., and Steitz, T.A. 1985. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* 313:762-766.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., and Schrag, J. 1992. The α/β hydrolase fold. *Protein Eng.* 5:197-211.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J., and Gronenborn, A.M. 1993. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* 261:438-446.

Onesti, S., Miller, A.D., and Brick, P. 1995. The crystal structure of the lysyl-tRNA synthetase (LysU) from *Escherichia coli*. *Structure.* 3:163-176.

Orengo, C.A., Flores, T.P., Taylor, W.R., and Thornton, J.M. 1993. Identification and classification of protein fold families. *Protein Eng.* 6:485-500.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. 1997. CATH—a hierarchic classification of protein domain structures. *Structure* 5:1093-1108.

Orengo, C.A., Bray, J.E., Buchan, D.W., Harrison, A., Lee, D., Pearl, F.M., Sillitoe, I., Todd, A.E., and Thornton, J.M. 2002. The CATH protein family database: a resource for structural and functional annotation of genomes. *Proteomics.* 2:11-21.

Ostermeier, C., Harrenga, A., Ermler, U., and Michel, H. 1997. Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome c oxidase complexed with an antibody FV fragment. *Proc. Natl. Acad. Sci. U.S.A.* 94:10547-10553.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372:432-438.

Overduin, M., Rios, C.B., Mayer, B.J., Baltimore, D., and Cowburn, D. 1992. Three-dimensional solution structure of the src homology 2 domain of c-abl. *Cell* 70:697-704.

Owen, D.J., Noble, M.E., Garman, E.F., Papageorgiou, A.C., and Johnson, L.N. 1995. Two structures of the catalytic domain of phosphorylase kinase: an active protein kinase complexed with substrate analogue and product. *Structure.* 3:467-482.

Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W., and Wittinghofer, A. 1990. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: Implications for the mechanism of GTP hydrolysis. *EMBO J.* 9:2351-2359.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le, T., I, Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., and Miyano, M. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289:739-745.

Pan, H. and Wigley, D.B. 2000. Structure of the zinc-binding domain of *Bacillus stearothermophilus* DNA primase. *Structure. Fold. Des* 8:231-239.

Pannifer, A.D., Wong, T.Y., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska, J., Lacy, D.B., Collier, R.J., Park, S., Leppla, S.H., Hanna, P., and Liddington, R.C. 2001. Crystal structure of the anthrax lethal factor. *Nature* 414:229-233.

Park, Y.C., Burkitt, V., Villa, A.R., Tong, L., and Wu, H. 1999. Structural basis for self-association and receptor recognition of human TRAF2. *Nature* 398:533-538.

Parker, M.W., Pattus, F., Tucker, A.D., and Tsernoglou, D. 1989. Structure of the membrane-pore-forming fragment of colicin A. *Nature* 337:93-96.

Patel, S., Yenush, L., Rodriguez, P.L., Serrano, R., and Blundell, T.L. 2002. Crystal structure of an enzyme displaying both inositol-polyphosphate-1-phosphatase and 3′-phosphoadenosine-5′-phosphate phosphatase activities: A novel target of lithium therapy. *J. Mol. Biol.* 315:677-685.

Pautsch, A. and Schulz, G.E. 1998. Structure of the outer membrane protein A transmembrane domain. *Nat. Struct. Biol.* 5:1013-1017.

Pautsch, A. and Schulz, G.E. 2000. High-resolution structure of the OmpA membrane domain. *J. Mol. Biol.* 298:273-282.

Pavletich, N.P. and Pabo, C.O. 1991. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809-817.

Pavletich, N.P. and Pabo, C.O. 1993. Crystal structure of a five-finger GLI-DNA complex: New perspectives on zinc fingers. *Science* 261:1701-1707.

Pawson, T., Gish, G.D., and Nash, P. 2001. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol.* 11:504-511.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J.P., and Landau, E.M. 1997. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* 277:1676-1681.

Pellegrini, L., Tan, S., and Richmond, T.J. 1995. Structure of serum response factor core bound to DNA. *Nature* 376:490-498.

Pellegrini, L., Burke, D.F., von Delft, F., Mulloy, B., and Blundell, T.L. 2000. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* 407:1029-1034.

Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H., and Kraut, J. 1994. Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. *Science* 264:1891-1903.

Perez-Canadillas, J.M. and Varani, G. 2001. Recent advances in RNA-protein recognition. *Curr. Opin. Struct. Biol.* 11:53-58.

Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A.B., Chet, I., Wilson, K.S., and Vorgias, C.E. 1994. Crystal structure of a bacterial chitinase at 2.3 Å resolution. *Structure* 2:1169-1180.

Petersen, J.F., Cherney, M.M., Liebig, H.D., Skern, T., Kuechler, E., and James, M.N. 1999. The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J.* 18:5463-5475.

Petosa, C., Collier, R.J., Klimpel, K.R., Leppla, S.H., and Liddington, R.C. 1997. Crystal structure of the anthrax toxin protective antigen. *Nature* 385:833-838.

Pfuhl, M. and Pastore, A. 1995. Tertiary structure of an immunoglobulin-like domain from the giant muscle protein titin: A new member of the I set. *Structure.* 3:391-401.

Phillips, S.E. 1980. Structure and refinement of oxymyoglobin at 1.6 Å resolution. *J. Mol. Biol.* 142:531-554.

Picot, D., Loll, P.J., and Garavito, R.M. 1994. The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1. *Nature* 367:243-249.

Polekhina, G., House, C.M., Traficante, N., Mackay, J.P., Relaix, F., Sassoon, D.A., Parker, M.W., and Bowtell, D.D. 2002. Siah ubiquitin ligase is structurally related to TRAF and modulates TNF- α signaling. *Nat. Struct. Biol.* 9:68-75.

Ponting, C.P., Phillips, C., Davies, K.E., and Blake, D.J. 1997. PDZ domains: targeting signalling molecules to sub-membranous sites. *Bioessays* 19:469-479.

Prakash, B., Praefcke, G.J., Renault, L., Wittinghofer, A., and Herrmann, C. 2000. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* 403:567-571.

Prasad, G.S., Earhart, C.A., Murray, D.L., Novick, R.P., Schlievert, P.M., and Ohlendorf, D.H. 1993. Structure of toxic shock syndrome toxin 1. *Biochemistry* 32:13761-13766.

Prasad, G.S., Radhakrishnan, R., Mitchell, D.T., Earhart, C.A., Dinges, M.M., Cook, W.J., Schlievert, P.M., and Ohlendorf, D.H. 1997. Refined structures of three crystal forms of toxic shock syndrome toxin-1 and of a tetramutant with reduced activity. *Protein Sci.* 6:1220-1227.

Predki, P.F., Nayak, L.M., Gottlieb, M.B., and Regan, L. 1995. Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. *Cell* 80:41-50.

Prehoda, K.E., Lee, D.J., and Lim, W.A. 1999. Structure of the enabled/VASP homology 1 domain-peptide complex: A key component in the spatial control of actin assembly. *Cell* 97:471-480.

Priestle, J.P., Schar, H.P., and Grutter, M.G. 1989. Crystallographic refinement of interleukin 1 β at 2.0 Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 86:9667-9671.

Prince, S.M., Papiz, M.Z., Freer, A.A., McDermott, G., Hawthornthwaite-Lawless, A.M., Cogdell, R.J., and Isaacs, N.W. 1997. Apoprotein structure in the LH2 complex from *Rhodopseudomonas acidophila* strain 10050: Modular assembly and protein pigment interactions. *J. Mol. Biol.* 268:412-423.

Qian, X., Jeon, C., Yoon, H., Agarwal, K., and Weiss, M.A. 1993. Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS. *Nature* 365:277-279. [published erratum appears in *Nature* 376:279]

Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J., and Wuthrich, K. 1989. The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell* 59:573-580. [published erratum appears in *Cell* 61:548]

Qin, B.Y., Bewley, M.C., Creamer, L.K., Baker, H.M., Baker, E.N., and Jameson, G.B. 1998. Structural basis of the Tanford transition of bovine β-lactoglobulin. *Biochemistry* 37:14014-14023.

Qin, B.Y., Chacko, B.M., Lam, S.S., de Caestecker, M.P., Correia, J.J., and Lin, K. 2001. Structural basis of Smad1 activation by receptor kinase phosphorylation. *Mol. Cell* 8:1303-1312.

Qin, B.Y., Lam, S.S., Correia, J.J., and Lin, K. 2002. Smad3 allostery links TGF-β receptor kinase activation to transcriptional control. *Genes Dev.* 16:1950-1963.

Qiu, X., Culp, J.S., DiLella, A.G., Hellmig, B., Hoog, S.S., Janson, C.A., Smith, W.W., and Abdel-Meguid, S.S. 1996. Unique fold and active site in cytomegalovirus protease. *Nature* 383:275-279.

Qiu, X., Janson, C.A., Culp, J.S., Richardson, S.B., Debouck, C., Smith, W.W., and Abdel-Meguid, S.S. 1997. Crystal structure of varicella-zoster virus protease. *Proc. Natl. Acad. Sci. U.S.A.* 94:2874-2879.

Qu, A. and Leahy, D.J. 1995. Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, αL β2) integrin. *Proc. Natl. Acad. Sci. U.S.A.* 92:10277-10281.

Raag, R., Li, H., Jones, B.C., and Poulos, T.L. 1993. Inhibitor-induced conformational change in cytochrome P-450CAM. *Biochemistry* 32:4571-4578.

Radaev, S., Motyka, S., Fridman, W.H., Sautes-Fridman, C., and Sun, P.D. 2001a. The structure of a human type III Fcgamma receptor in complex with Fc. *J. Biol. Chem.* 276:16469-16477.

Radaev, S., Rostro, B., Brooks, A.G., Colonna, M., and Sun, P.D. 2001b. Conformational plasticity revealed by the cocrystal structure of NKG2D and its class I MHC-like ligand ULBP3. *Immunity.* 15:1039-1049.

Ramakrishnan, V. and White, S.W. 1992. The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature* 358:768-771.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* 362:219-223.

Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G.G., and Stuart, D. 1995. The structure of a Ca(2+)-binding epidermal growth factor-like domain: its role in protein-protein interactions. *Cell* 82:131-141.

Rao-Naik, C., delaCruz, W., Laplaza, J.M., Tan, S., Callis, J., and Fisher, A.J. 1998. The rub family of ubiquitin-like proteins. Crystal structure of *Arabidopsis* rub1 and expression of multiple rubs in Arabidopsis. *J. Biol. Chem.* 273:34976-34982.

Rastinejad, F., Perlmann, T., Evans, R.M., and Sigler, P.B. 1995. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* 375:203-211.

Raumann, B.E., Rould, M.A., Pabo, C.O., and Sauer, R.T. 1994. DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature* 367:754-757.

Rayment, I., Rypniewski, W.R., Schmidt-Base, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G., and Holden, H.M. 1993. Three-dimensional structure of myosin subfragment-1: A molecular motor. *Science* 261:50-58.

Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G. 1998. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* 279:1504-1513.

Rees, D.C., Lewis, M., and Lipscomb, W.N. 1983. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J. Mol. Biol.* 168:367-387.

Reiling, K.K., Pray, T.R., Craik, C.S., and Stroud, R.M. 2000. Functional consequences of the Kaposi's sarcoma-associated herpesvirus protease structure: Regulation of activity and dimerization by conserved structural elements. *Biochemistry* 39:12796-12803.

Reily, M.D., Thanabal, V., and Adams, M.E. 1995. The solution structure of omega-Aga-IVB, a P-type calcium channel antagonist from venom of the funnel web spider, *Agelenopsis aperta. J. Biomol. NMR.* 5:122-132.

Reinherz, E.L., Tan, K., Tang, L., Kern, P., Liu, J., Xiong, Y., Hussey, R.E., Smolyar, A., Hare, B., Zhang, R., Joachimiak, A., Chang, H.C., Wagner, G., and Wang, J. 1999. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science* 286:1913-1921.

Ren, G., Reddy, V.S., Cheng, A., Melnyk, P., and Mitra, A.K. 2001. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc. Natl. Acad. Sci. U.S.A.* 98:1398-1403.

Renatus, M., Stennicke, H.R., Scott, F.L., Liddington, R.C., and Salvesen, G.S. 2001. Dimer formation drives the activation of the cell death protease caspase 9. *Proc. Natl. Acad. Sci. U.S.A.* 98:14250-14255.

Renaud, J.P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., and Moras, D. 1995. Crystal structure of the RAR-γ ligand-binding domain bound to all-trans retinoic acid. *Nature* 378:681-689.

Rice, P. and Mizuuchi, K. 1995. Structure of the bacteriophage Mu transposase core: A common structural motif for DNA transposition and retroviral integration. *Cell* 82:209-220.

Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87:1295-1306.

Richardson, D.C. and Richardson, J.S. 1992. The kinemage: A tool for scientific communication. *Protein Sci.* 1:3-9.

Richardson, J.S., Richardson, D.C., Thomas, K.A., Silverton, E.W., and Davies, D.R. 1976. Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc superoxide dismutase subunit. *J. Mol. Biol.* 102:221-235.

Rini, J.M. 1995a. Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* 24:551-577.

Rini, J.M. 1995b. X-ray crystal structures of animal lectins. *Curr. Opin. Struct. Biol.* 5:617-621.

Rittinger, K., Walker, P.A., Eccleston, J.F., Nurmahomed, K., Owen, D., Laue, E., Gamblin, S.J., and Smerdon, S.J. 1997a. Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP. *Nature* 388:693-697.

Rittinger, K., Walker, P.A., Eccleston, J.F., Smerdon, S.J., and Gamblin, S.J. 1997b. Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature* 389:758-762.

Rodgers, A.J. and Wilce, M.C. 2000. Structure of the γ-ε complex of ATP synthase. *Nat. Struct. Biol.* 7:1051-1054.

Rodriguez-Romero, A., Ravichandran, K.G., and Soriano-Garcia, M. 1991. Crystal structure of hevein at 2.8 Å resolution. *FEBS Lett.* 291:307-309.

Rossmann, M.G., Liljas, A., Bränden, C.-I., and Banaszak, L.J. 1975. Evolutionary and structural relationships among dehydrogenases. *The Enzymes* 11:61-102.

Rould, M.A., Perona, J.J., and Steitz, T.A. 1991. Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* 352:213-218.

Royant, A., Nollert, P., Edman, K., Neutze, R., Landau, E.M., Pebay-Peyroula, E., and Navarro, J. 2001. X-ray structure of sensory rhodopsin II at 2.1-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 98:10131-10136.

Rozwarski, D.A., Gronenborn, A.M., Clore, G.M., Bazan, J.F., Bohm, A., Wlodawer, A., Hatada, M., and Karplus, P.A. 1994. Structural comparisons among the short-chain helical cytokines. *Structure* 2:159-173.

Rudolph, M.J. and Gergen, J.P. 2001. DNA-binding by Ig-fold proteins. *Nat. Struct. Biol.* 8:384-386.

Russo, A.A., Tong, L., Lee, J.O., Jeffrey, P.D., and Pavletich, N.P. 1998. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* 395:237-243.

Rutenber, E. and Robertus, J.D. 1991. Structure of ricin B-chain at 2.5 Å resolution. *Proteins* 10:260-269.

Ryu, S.E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X.P., Xuong, N.H., Axel, R., Sweet, R.W., et al. 1990. Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* 348:419-426.

Sacchettini, J.C., Gordon, J.I., and Banaszak, L.J. 1989. Crystal structure of rat intestinal fatty-acid-binding protein. Refinement and analysis of the Escherichia coli-derived protein with bound palmitate. *J. Mol. Biol.* 208:327-339.

Sanchez, L.M., Chirino, A.J., and Bjorkman, P. 1999. Crystal structure of human ZAG, a fat-depleting factor related to MHC molecules. *Science* 283:1914-1919.

Sanderson, M.R., Freemont, P.S., Rice, P.A., Goldman, A., Hatfull, G.F., Grindley, N.D., and Steitz, T.A. 1990. The crystal structure of the catalytic domain of the site-specific recombination enzyme γδ resolvase at 2.7 Å resolution. *Cell* 63:1323-1329.

Santelli, E. and Richmond, T.J. 2000. Crystal structure of MEF2A core bound to DNA at 1.5 Å resolution. *J. Mol. Biol.* 297:437-449.

Saraste, M., Sibbald, P.R., and Wittinghofer, A. 1990. The P-loop - a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* 15:430-434.

Satyshur, K.A., Rao, S.T., Pyzalska, D., Drendel, W., Greaser, M., and Sundaralingam, M. 1988. Refined structure of chicken skeletal muscle troponin C in the two-calcium state at 2-Å resolution. *J. Biol. Chem.* 263:1628-1647.

Saul, F.A., Rovira, P., Boulot, G., Damme, E.J., Peumans, W.J., Truffa-Bachi, P., and Bentley, G.A. 2000. Crystal structure of Urtica dioica agglutinin, a superantigen presented by MHC molecules of class I and class II. *Structure. Fold. Des* 8:593-603.

Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H., and Kraut, J. 1994. Crystal structure of rat DNA polymerase β: Evidence for a common polymerase mechanism. *Science* 264:1930-1935.

Sayle, R.A. and Milner-White, E.J. 1995. RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* 20:374.

Scheffzek, K., Lautwein, A., Kabsch, W., Ahmadian, M.R., and Wittinghofer, A. 1996. Crystal structure of the GTPase-activating domain of human p120GAP and implications for the interaction with Ras. *Nature* 384:591-596.

Schindelin, H., Marahiel, M.A., and Heinemann, U. 1993. Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature* 364:164-168.

Schindelin, H., Jiang, W., Inouye, M., and Heinemann, U. 1994. Crystal structure of CspA, the major cold shock protein of *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.* 91:5119-5123.

Schirmer, T., Bode, W., and Huber, R. 1987. Refined three-dimensional structures of two cyanobacterial C-phycocyanins at 2.1 and 2.5 Å resolution. A common principle of phycobilin-protein interaction. *J. Mol. Biol.* 196:677-695.

Schirmer, T., Keller, T.A., Wang, Y.F., and Rosenbusch, J.P. 1995. Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science* 267:512-514.

Schlunegger, M.P. and Grutter, M.G. 1993. Refined crystal structure of human transforming growth factor β2 at 1.95 Å resolution. *J. Mol. Biol.* 231:445-458.

Schnuchel, A., Wiltscheck, R., Czisch, M., Herrler, M., Willimsky, G., Graumann, P., Marahiel, M.A., and Holak, T.A. 1993. Structure in solution of the major cold-shock protein from *Bacillus subtilis*. *Nature* 364:169-171.

Schulman, B.A., Carrano, A.C., Jeffrey, P.D., Bowen, Z., Kinnucan, E.R., Finnin, M.S., Elledge, S.J., Harper, J.W., Pagano, M., and Pavletich, N.P. 2000. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* 408:381-386.

Schultz, G.E. 1992. Binding of nucleotides by proteins. *Curr. Opin. Struc. Bio.* 2:61-67.

Schulz, G.E. 2002. The structure of bacterial outer membrane proteins. *Biochim. Biophys. Acta* 1565:308-317.

Schulz, G.E., Elzinga, M., Marx, F., and Schirmer, R.H. 1974. Three-dimensional structure of adenylate kinase. *Nature* 250:120-123.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* 95:5857-5864.

Schultz, S.C., Shields, G.C., and Steitz, T.A. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science* 253:1001-1007.

Schumacher, M.A., Choi, K.Y., Zalkin, H., and Brennan, R.G. 1994. Crystal structure of LacI member, PurR, bound to DNA: Minor groove binding by α-helices. *Science* 266:763-770.

Schutt, C.E., Myslik, J.C., Rozycki, M.D., Goonesekere, N.C., and Lindberg, U. 1993. The structure of crystalline profilin-β-actin. *Nature* 365:810-816.

Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. 1993. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* 75:567-578.

Sedgwick, S.G. and Smerdon, S.J. 1999. The ankyrin repeat: A diversity of interactions on a common structural framework. *Trends Biochem. Sci.* 24:311-316.

Shamoo, Y. and Steitz, T.A. 1999. Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99:155-166.

Shapiro, L., Fannon, A.M., Kwong, P.D., Thompson, A., Lehmann, M.S., Grubel, G., Legrand, J.F., Als-Nielsen, J., Colman, D.R., and Hendrickson, W.A. 1995. Structural basis of cell-cell adhesion by cadherins. *Nature* 374:327-337.

Sharon, N. 1993. Lectin-carbohydrate complexes of plants and animals: An atomic view. *Trends Biochem. Sci.* 18:221-226.

Shen, Y., Lee, Y.S., Soelaiman, S., Bergson, P., Lu, D., Chen, A., Beckingham, K., Grabarek, Z., Mrksich, M., and Tang, W.J. 2002. Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylyl cyclase exotoxins. *EMBO J.* 21:6721-6732.

Shewchuk, L.M., Hassell, A.M., Ellis, B., Holmes, W.D., Davis, R., Horne, E.L., Kadwell, S.H., McKee, D.D., and Moore, J.T. 2000. Structure of the Tie2 RTK domain: Self-inhibition by the nucleotide binding loop, activation loop, and C-terminal tail. *Structure. Fold. Des* 8:1105-1113.

Shi, Y., Hata, A., Lo, R.S., Massague, J., and Pavletich, N.P. 1997. A structural basis for mutational inactivation of the tumour suppressor Smad4. *Nature* 388:87-93.

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., and Greene, G.L. 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95:927-937.

Shiba, T., Takatsu, H., Nogi, T., Matsugaki, N., Kawasaki, M., Igarashi, N., Suzuki, M., Kato, R., Earnest, T., Nakayama, K., and Wakatsuki, S. 2002. Structural basis for recognition of acidic-cluster dileucine sequence by GGA1. *Nature* 415:937-941.

Shieh, H.S., Kurumbail, R.G., Stevens, A.M., Stegeman, R.A., Sturman, E.J., Pak, J.Y., Wittwer, A.J., Palmier, M.O., Wiegand, R.C., Holwerda, B.C., and Stallings, W.C. 1996. Three-dimensional structure of human cytomegalovirus protease. *Nature* 383:279-282. [published erratum appears in *Nature* 384:288]

Shindyalov, I.N. and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11:739-747.

Shirakihara, Y., Leslie, A.G., Abrahams, J.P., Walker, J.E., Ueda, T., Sekimoto, Y., Kambara, M., Saika, K., Kagawa, Y., and Yoshida, M. 1997. The crystal structure of the nucleotide-free α3β3 subcomplex of F1-ATPase from the thermophilic Bacillus PS3 is a symmetric trimer. *Structure.* 5:825-836.

Sicheri, F., Moarefi, I., and Kuriyan, J. 1997. Crystal structure of the Src family tyrosine kinase Hck. *Nature* 385:602-609.

Sielecki, A.R., Hayakawa, K., Fujinaga, M., Murphy, M.E., Fraser, M., Muir, A.K., Carilli, C.T., Lewicki, J.A., Baxter, J.D., and James, M.N. 1989. Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5 Å resolution. *Science* 243:1346-1351.

Sixma, T.K., Pronk, S.E., Kalk, K.H., Wartna, E.S., van Zanten, B.A., Witholt, B., and Hol, W.G. 1991. Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature* 351:371-377.

Skelton, N.J., Aspiras, F., Ogez, J., and Schall, T.J. 1995. Proton NMR assignments and solution conformation of RANTES, a chemokine of the C-C type. *Biochemistry* 34:5329-5342.

Smalla, M., Schmieder, P., Kelly, M., Ter Laak, A., Krause, G., Ball, L., Wahl, M., Bork, P., and Oschkinat, H. 1999. Solution structure of the receptor tyrosine kinase EphB2 SAM domain and identification of two distinct homotypic interaction sites. *Protein Sci.* 8:1954-1961.

Smith, L.J., Redfield, C., Smith, R.A., Dobson, C.M., Clore, G.M., Gronenborn, A.M., Walter, M.R., Naganbushan, T.L., and Wlodawer, A. 1994. Comparison of four independently determined structures of human recombinant interleukin-4. *Nat. Struct. Biol.* 1:301-310.

Smith, T.J. 1995. MolView: A program for analyzing and displaying atomic structures on the Macintosh personal computer. *J. Mol. Graph.* 13:122-5, 115.

Snijder, H.J., Ubarretxena-Belandia, I., Blaauw, M., Kalk, K.H., Verheij, H.M., Egmond, M.R., Dekker, N., and Dijkstra, B.W. 1999. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* 401:717-721.

Somers, W.S. and Phillips, S.E. 1992. Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature* 359:387-393.

Somers, W., Ultsch, M., de Vos, A.M., and Kossiakoff, A.A. 1994. The X-ray structure of a growth hormone-prolactin receptor complex. *Nature* 372:478-481.

Sondermann, P., Huber, R., Oosthuizen, V., and Jacob, U. 2000. The 3.2-Å crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. *Nature* 406:267-273.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J.E. 1996. Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* 274:1859-1866.

Song, H., Hanlon, N., Brown, N.R., Noble, M.E., Johnson, L.N., and Barford, D. 2001. Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. *Mol. Cell* 7:615-626.

Soulimane, T., Buse, G., Bourenkov, G.P., Bartunik, H.D., Huber, R., and Than, M.E. 2000. Structure and mechanism of the aberrant ba(3)-cytochrome c oxidase from *thermus thermophilus*. *EMBO J* 19:1766-1776.

Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364:593-599.

South, T.L. and Summers, M.F. 1993. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* 2:3-19.

Sprang, S.R. and Bazan, J.F. 1993. Cytokine structural taxonomy and mechanisms of receptor engagement. *Curr. Opin. in Struc. Biol.* 3:815-827.

Stahelin, R.V. and Cho, W. 2001. Roles of calcium ions in the membrane binding of C2 domains. *Biochem. J.* 359:679-685.

Stamper, C.C., Zhang, Y., Tobin, J.F., Erbe, D.V., Ikemizu, S., Davis, S.J., Stahl, M.L., Seehra, J., Somers, W.S., and Mosyak, L. 2001. Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses. *Nature* 410:608-611.

Stapleton, D., Balan, I., Pawson, T., and Sicheri, F. 1999. The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. *Nat. Struct. Biol.* 6:44-49.

Stebbins, C.E. and Galan, J.E. 2000. Modulation of host signaling by a bacterial mimic: Structure of the Salmonella effector SptP bound to Rac1. *Mol. Cell* 6:1449-1460.

Stehle, T. and Schulz, G.E. 1990. Three-dimensional structure of the complex of guanylate kinase from yeast with its substrate GMP. *J. Mol. Biol.* 211:249-254.

Stein, P.E., Boodhoo, A., Armstrong, G.D., Cockle, S.A., Klein, M.H., and Read, R.J. 1994. The crystal structure of pertussis toxin. *Structure.* 2:45-57.

Steinbacher, S., Hof, P., Eichinger, L., Schleicher, M., Gettemans, J., Vandekerckhove, J., Huber, R., and Benz, J. 1999. The crystal structure of the Physarum polycephalum actin-fragmin kinase: An atypical protein kinase with a specialized substrate-binding domain. *EMBO J.* 18:2923-2929.

Stenmark, H., Aasland, R., and Driscoll, P.C. 2002. The phosphatidylinositol 3-phosphate-binding FYVE finger. *FEBS Lett.* 513:77-84.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L., and Wiley, D.C. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 368:215-221.

Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G., and Champoux, J.J. 1998. A model for the mechanism of human topoisomerase I. *Science* 279:1534-1541.

Stewart, A.E., Dowd, S., Keyse, S.M., and McDonald, N.Q. 1999. Crystal structure of the MAPK phosphatase Pyst1 catalytic domain and implications for regulated activation. *Nat. Struct. Biol.* 6:174-181.

Stock, D., Leslie, A.G., and Walker, J.E. 1999. Molecular architecture of the rotary motor in ATP synthase. *Science* 286:1700-1705.

Strater, N., Klabunde, T., Tucker, P., Witzel, H., and Krebs, B. 1995. Crystal structure of a purple acid phosphatase containing a dinuclear Fe(III)-Zn(II) active site. *Science* 268:1489-1492.

Stroud, J.C., Lopez-Rodriguez, C., Rao, A., and Chen, L. 2002. Structure of a TonEBP-DNA complex reveals DNA encircled by a transcription factor. *Nat. Struct. Biol.* 9:90-94.

Stuckey, J.A., Schubert, H.L., Fauman, E.B., Zhang, Z.Y., Dixon, J.E., and Saper, M.A. 1994. Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 Å and the complex with tungstate. *Nature* 370:571-575.

Su, X.D., Taddei, N., Stefani, M., Ramponi, G., and Nordlund, P. 1994. The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. *Nature* 370:575-578.

Subramaniam, S. 1999. The structure of bacteriorhodopsin: An emerging consensus. *Curr. Opin. Struct. Biol.* 9:462-468.

Sudol, M. and Hunter, T. 2000. NeW wrinkles for an old domain. *Cell* 103:1001-1004.

Sugio, S., Kashima, A., Mochizuki, S., Noda, M., and Kobayashi, K. 1999. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng* 12:439-446.

Sui, H., Han, B.G., Lee, J.K., Walian, P., and Jap, B.K. 2001. Structural basis of water-specific transport through the AQP1 water channel. *Nature* 414:872-878.

Sun, P.D. and Davies, D.R. 1995. The cystine-knot growth-factor superfamily. *Annu.Rev.Biophys.Biomol.Struct.* 24:269-291.

Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., and Silman, I. 1991. Atomic structure of acetylcholinesterase from Torpedo californica: A prototypic acetylcholine-binding protein. *Science* 253:872-879.

Sutton, R.B., Davletov, B.A., Berghuis, A.M., Sudhof, T.C., and Sprang, S.R. 1995. Structure of the first C2 domain of synaptotagmin I: A novel Ca^{2+}/phospholipid-binding fold. *Cell* 80:929-938.

Sutton, R.B., Fasshauer, D., Jahn, R., and Brunger, A.T. 1998. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* 395:347-353.

Svensson, L.A., Thulin, E., and Forsen, S. 1992. Proline *cis*-trans isomers in calbindin D9k observed by X-ray crystallography. *J. Mol. Biol.* 223:601-606.

Swaminathan, K., Flynn, P., Reece, R.J., and Marmorstein, R. 1997. Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DNA recognition by a protein containing a Zn_2Cys_6 binuclear cluster. *Nat. Struct. Biol.* 4:751-759.

Swaminathan, S., Furey, W., Pletcher, J., and Sax, M. 1992. Crystal structure of staphylococcal enterotoxin B, a superantigen. *Nature* 359:801-806.

Swaminathan, S., Furey, W., Pletcher, J., and Sax, M. 1995. Residues defining V beta specificity in staphylococcal enterotoxins. *Nat. Struct. Biol.* 2:680-686.

Tahirov, T.H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., Sato, K., Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. 2001. Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFβ. *Cell* 104:755-767.

Takahashi, N., Takahashi, Y., and Putnam, F.W. 1985. Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α2-glycoprotein of human serum. *Proc. Natl. Acad. Sci. U.S.A.* 82:1906-1910.

Tan, S. and Richmond, T.J. 1998. Crystal structure of the yeast MATα2/MCM1/DNA ternary complex. *Nature* 391:660-666.

Tanaka, I., Appelt, K., Dijk, J., White, S.W., and Wilson, K.S. 1984. 3-Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature* 310:376-381.

Tanaka, N., Nonaka, T., Nakanishi, M., Deyashiki, Y., Hara, A., and Mitsui, Y. 1996. Crystal structure of the ternary complex of mouse lung carbonyl reductase at 1.8 Å resolution: The structural origin of coenzyme specificity in the short-chain dehydrogenase/reductase family. *Structure.* 4:33-45.

Taneja, B. and Mande, S.C. 2002. Structure of *Mycobacterium tuberculosis* chaperonin-10 at 3.5 Å resolution. *Acta Crystallogr. D. Biol. Crystallogr.* 58:260-266.

Tereshko, V., Teplova, M., Brunzelle, J., Watterson, D.M., and Egli, M. 2001. Crystal structures of the catalytic domain of human protein kinase associated with apoptosis and tumor suppression. *Nat. Struct. Biol.* 8:899-907.

Tesmer, J.J., Berman, D.M., Gilman, A.G., and Sprang, S.R. 1997. Structure of RGS4 bound to AlF4—activated G(i α1): Stabilization of the transition state for GTP hydrolysis. *Cell* 89:251-261.

Thanos, C.D., Goodwill, K.E., and Bowie, J.U. 1999. Oligomeric structure of the human EphB2 receptor SAM domain. *Science* 283:833-836.

Thompson, J., Winter, N., Terwey, D., Bratt, J., and Banaszak, L. 1997. The crystal structure of the liver fatty acid-binding protein. A complex with two bound oleates. *J. Biol. Chem.* 272:7140-7150.

Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27:527-533.

Tong, L., Qian, C., Massariol, M.J., Bonneau, P.R., Cordingley, M.G., and Lagace, L. 1996. A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature* 383:272-275.

Tormo, J., Natarajan, K., Margulies, D.H., and Mariuzza, R.A. 1999. Crystal structure of a lectin-like natural killer cell receptor bound to its MHC class I ligand. *Nature* 402:623-631.

Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405:647-655.

Toyoshima, C. and Nomura, H. 2002. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* 418:605-611.

Transue, T.R., Smith, A.K., Mo, H., Goldstein, I.J., and Saper, M.A. 1997. Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nat. Struct. Biol.* 4:779-783.

Traut, T. 1994. The functions and consensus motifs of nine type of peptide segments that form different types of nucleotide-binding sites. *Eur. J. Biochem.* 222:9-19.

Travers, A. 2000. Recognition of distorted DNA structures by HMG domains. *Curr. Opin. Struct. Biol.* 10:102-109.

Tsernoglou, D., Petsko, G.A., and Hudson, R.A. 1978. Structure and function of snake venom curarimimetic neurotoxins. *Mol. Pharmacol.* 14:710-716.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. 1996. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272:1136-1144.

Unno, M., Mizushima, T., Morimoto, Y., Tomisugi, Y., Tanaka, K., Yasuoka, N., and Tsukihara, T. 2002. The structure of the mammalian 20S proteasome at 2.75 Å resolution. *Structure.(Camb.)* 10:609-618.

Valegard, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J., and Liljas, L. 1994. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* 371:623-626.

van der Greer, P., Wiley, S., Lai, V.K., Olivier, J.P., Gish, G.D., Stephens, R., Kaplan, D., Shoelson, S., and Pawson, T. 1995. A conserved amino-terminal Shc domain binds to phosphotyrosine motifs in activated receptors and phosphopeptides. *Curr. Biol.* 5:404-412.

VanDemark, A.P., Hofmann, R.M., Tsui, C., Pickart, C.M., and Wolberger, C. 2001. Molecular insights into polyubiquitin chain assembly: Crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105:711-720.

Varghese, J.N. and Colman, P.M. 1991. Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 Å resolution. *J. Mol. Biol.* 221:473-486.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417:712-719.

Verdaguer, N., Corbalan-Garcia, S., Ochoa, W.F., Fita, I., and Gomez-Fernandez, J.C. 1999. Ca(2+) bridges the C2 membrane-binding domain of protein kinase Cα directly to phosphatidylserine. *EMBO J.* 18:6329-6338.

Vetter, I.R., Arndt, A., Kutay, U., Gorlich, D., and Wittinghofer, A. 1999. Structural view of the Ran-Importin beta interaction at 2.3 Å resolution. *Cell* 97:635-646.

Vijay-Kumar, S., Bugg, C.E., and Cook, W.J. 1987. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* 194:531-544.

Vis, H., Mariani, M., Vorgias, C.E., Wilson, K.S., Kaptein, R., and Boelens, R. 1995. Solution structure of the HU protein from *Bacillus stearothermophilus*. *J. Mol. Biol.* 254:692-703.

Voegtli, W.C., White, D.J., Reiter, N.J., Rusnak, F., and Rosenzweig, A.C. 2000. Structure of the bacteriophage lambda Ser/Thr protein phosphatase with sulfate ion bound in two coordination modes. *Biochemistry* 39:15365-15374.

Vogelstein, B. 1990. Cancer. A deadly inheritance. *Nature* 348:681-682.

Vogt, J. and Schulz, G.E. 1999. The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Structure.Fold.Des* 7:1301-1309.

Volz, K. and Matsumura, P. 1991. Crystal structure of *Escherichia coli* CheY refined at 1.7-Å resolution. *J.Biol.Chem.* 266:15511-15519.

Waksman, G., Kominos, D., Robertson, S.C., Pant, N., Baltimore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., and . 1992. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* 358:646-653.

Walker, N.P., Talanian, R.V., Brady, K.D., Dang, L.C., Bump, N.J., Ferenz, C.R., Franklin, S., Ghayur, T., Hackett, M.C., Hammill, L.D., and . 1994. Crystal structure of the cysteine protease interleukin-1 β-converting enzyme: A (p20/p10)2 homodimer. *Cell* 78:343-352.

Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluhi, J.A., Posner, B.A., Gilman, A.G., and Sprang, S.R. 1995. The structure of the G protein heterotrimer Gi α1 β1 β2. *Cell* 83:1047-1058.

Walter, M.R., Windsor, W.T., Nagabhushan, T.L., Lundell, D.J., Lunn, C.A., Zauodny, P.J., and Narula, S.K. 1995. Crystal structure of a complex between interferon-γ and its soluble high-affinity receptor. *Nature* 376:230-235.

Walther, D. 1997. WebMol—a Java-based PDB viewer. *Trends Biochem. Sci.* 22:274-275.

Wang, J.H., Yan, Y.W., Garrett, T.P., Liu, J.H., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L., and Harrison, S.C. 1990. Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* 348:411-418.

Wang, Z., Harkins, P.C., Ulevitch, R.J., Han, J., Cobb, M.H., and Goldsmith, E.J. 1997. The structure of mitogen-activated protein kinase p38 at 2.1-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 94:2327-2332.

Wang, B., Jones, D.N., Kaine, B.P., and Weiss, M.A. 1998. High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. *Structure.* 6:555-569.

Wang, P., Byeon, I.J., Liao, H., Beebe, K.D., Yongkiettrakul, S., Pei, D., and Tsai, M.D. 2000a. II. Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphotyrosyl peptides. *J. Mol. Biol.* 302:927-940.

Wang, Y., Geer, L.Y., Chappey, C., Kans, J.A., and Bryant, S.H. 2000b. Cn3D: Sequence and structure views for Entrez. *Trends Biochem. Sci.* 25:300-302.

Wang, J.H., Meijers, R., Xiong, Y., Liu, J.H., Sakihama, T., Zhang, R., Joachimiak, A., and Reinherz, E.L. 2001. Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc. Natl. Acad. Sci. U.S.A.* 98:10799-10804.

Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M. 2002. Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110:501-512.

Warren, A.J., Bravo, J., Williams, R.L., and Rabbitts, T.H. 2000. Structural basis for the heterodimeric interaction between the acute leukaemia-associated transcription factors AML1 and CBFbeta. *EMBO J.* 19:3004-3015.

Watenpaugh, K.D., Sieker, L.C., Jensen, L.H., Legall, J., and Dubourdieu, M. 1972. Structure of the oxidized form of a flavodoxin at 2.5-Angstrom resolution: Resolution of the phase ambiguity by anomalous scattering. *Proc. Natl. Acad. Sci. U.S.A.* 69:3185-3188.

Weatherman, R.V., Fletterick, R.J., and Scanlan, T.S. 1999. Nuclear-receptor ligands and ligand-binding domains. *Annu. Rev. Biochem.* 68:559-581.

Wedekind, J.E., Trame, C.B., Dorywalska, M., Koehl, P., Raschke, T.M., McKee, M., FitzGerald, D., Collier, R.J., and McKay, D.B. 2001. Refined crystallographic structure of *Pseudomonas aeruginosa* exotoxin A and its implications for the molecular mechanism of toxicity. *J. Mol. Biol.* 314:823-837.

Wei, Y., Fox, T., Chambers, S.P., Sintchak, J., Coll, J.T., Golec, J.M., Swenson, L., Wilson, K.P., and Charifson, P.S. 2000. The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. *Chem. Biol.* 7:423-432.

Weichenrieder, O., Wild, K., Strub, K., and Cusack, S. 2000. Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* 408:167-173.

Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R., Laue, E.D., and Thomas, J.O. 1993. Structure of the HMG box motif in the B-domain of HMG1. *EMBO J.* 12:1311-1319.

Weiss, M.S. and Schulz, G.E. 1992. Structure of porin refined at 1.8 Å resolution. *J. Mol. Biol.* 227:493-509.

Weis, W.I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W.A. 1991. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* 254:1608-1615.

Weis, W.I., Drickamer, K., and Hendrickson, W.A. 1992. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* 360:127-134.

Wendt, K.U., Lenhart, A., and Schulz, G.E. 1999. The structure of the membrane protein squalene-hopene cyclase at 2.0 Å resolution. *J. Mol. Biol.* 286:175-187.

Werner, M.H., Huth, J.R., Gronenborn, A.M., and Clore, G.M. 1995. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* 81:705-714.

Whitby, F.G., Masters, E.I., Kramer, L., Knowlton, J.R., Yao, Y., Wang, C.C., and Hill, C.P. 2000. Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* 408:115-120.

White, C.L., Janakiraman, M.N., Laver, W.G., Philippon, C., Vasella, A., Air, G.M., and Luo, M. 1995. A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. *J. Mol. Biol.* 245:623-634.

Wierenga, R.K., De Maeyer, M.C.H., and Hol, W.G.J. 1985. Interaction of pyrophosphate moieties with α-helices in dinucleotide binding proteins. *Biochemistry* 24:1346-1357.

Wiesmann, C., Fuh, G., Christinger, H.W., Eigenbrot, C., Wells, J.A., and de Vos, A.M. 1997. Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. *Cell* 91:695-704.

Wiesmann, C., Ultsch, M.H., Bass, S.H., and de Vos, A.M. 1999. Crystal structure of nerve growth factor in complex with the ligand- binding domain of the TrkA receptor. *Nature* 401:184-188.

Wigley, D.B. 1995. Structure and mechanism of DNA topoisomerases. *Annu. Rev. Biophys. Biomol. Struct.* 24:185-208.

Wigley, D.B., Davies, G.J., Dodson, E.J., Maxwell, A., and Dodson, G. 1991. Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature* 351:624-629.

Wild, K., Sinning, I., and Cusack, S. 2001. Crystal structure of an early protein-RNA assembly complex of the signal recognition particle. *Science* 294:598-601.

Wiles, A.P., Shaw, G., Bright, J., Perczel, A., Campbell, I.D., and Barlow, P.N. 1997. NMR studies of a viral protein that mimics the regulators of complement activation. *J. Mol. Biol.* 272:253-265.

Williams, S.P. and Sigler, P.B. 1998. Atomic structure of progesterone complexed with its receptor. *Nature* 393:392-396.

Wilson, K.P., Shewchuk, L.M., Brennan, R.G., Otsuka, A.J., and Matthews, B.W. 1992. *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc. Natl. Acad. Sci. U.S.A.* 89:9257-9261.

Wilson, K.P., Black, J.A., Thomson, J.A., Kim, E.E., Griffith, J.P., Navia, M.A., Murcko, M.A., Chambers, S.P., Aldape, R.A., and Raybuck, S.A. 1994. Structure and mechanism of interleukin-1 beta converting enzyme. *Nature* 370:270-275.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. 2000. Structure of the 30S ribosomal subunit. *Nature* 407:327-339.

Wingren, C., Crowley, M.P., Degano, M., Chien, Y., and Wilson, I.A. 2000. Crystal structure of a γδ T cell receptor ligand T22: A truncated MHC-like fold. *Science* 287:310-314.

Winter, N.S., Bratt, J.M., and Banaszak, L.J. 1993. Crystal structures of holo and apo-cellular retinol-binding protein II. *J. Mol. Biol.* 230:1247-1259.

Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G., and Mueller, L. 1992. ^{1}H, ^{13}C, and ^{15}N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins. *Biochemistry* 31:6254-6265.

Wlodawer, A., Pavlovsky, A., and Gustchina, A. 1993. Hematopoietic cytokines: Similarities and differences in the structures, with implications for receptor binding. *Protein Sci.* 2:1373-1382.

Wolan, D.W., Teyton, L., Rudolph, M.G., Villmow, B., Bauer, S., Busch, D.H., and Wilson, I.A. 2001. Crystal structure of the murine NK cell-activating receptor NKG2D at 1.95 Å. *Nat. Immunol.* 2:248-254.

Wolberger, C., Dong, Y.C., Ptashne, M., and Harrison, S.C. 1988. Structure of a phage 434 Cro/DNA complex. *Nature* 335:789-795.

Worthylake, D.K., Rossman, K.L., and Sondek, J. 2000. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* 408:682-688.

Wright, C.S. and Jaeger, J. 1993. Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J. Mol. Biol.* 232:620-638.

Wright, P.E. and Dyson, H.J. 1999. Intrinsically unstructured proteins: Reassessing the protein structure-function paradigm. *J. Mol. Biol.* 293:321-331.

Wu, H., Lustbader, J.W., Liu, Y., Canfield, R.E., and Hendrickson, W.A. 1994. Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein. *Structure.* 2:545-558.

Wu, G., Chen, Y.G., Ozdamar, B., Gyuricza, C.A., Chong, P.A., Wrana, J.L., Massague, J., and Shi, Y. 2000. Structural basis of Smad2 recognition by the Smad anchor for receptor activation. *Science* 287:92-97.

Wu, J.W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D.J., Kyin, S., Muir, T.W., Fairman, R., Massague, J., and Shi, Y. 2001. Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-β signaling. *Mol. Cell* 8:1277-1289.

Wybenga-Groot, L.E., Baskin, B., Ong, S.H., Tong, J., Pawson, T., and Sicheri, F. 2001. Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell* 106:745-757.

Xia, D., Yu, C.A., Kim, H., Xia, J.Z., Kachurin, A.M., Zhang, L., Yu, L., and Deisenhofer, J. 1997. Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria. *Science* 277:60-66.

Xiao, B., Smerdon, S.J., Jones, D.H., Dodson, G.G., Soneji, Y., Aitken, A., and Gamblin, S.J. 1995. Structure of a 14-3-3 protein and implications for coordination of multiple signalling pathways. *Nature* 376:188-191.

Xie, X., Harrison, D.H., Schlichting, I., Sweet, R.M., Kalabokis, V.N., Szent-Gyorgyi, A.G., and Cohen, C. 1994. Structure of the regulatory domain of scallop myosin at 2.8 Å resolution. *Nature* 368:306-312.

Xie, X., Kokubo, T., Cohen, S.L., Mirza, U.A., Hoffmann, A., Chait, B.T., Roeder, R.G., Nakatani, Y., and Burley, S.K. 1996. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* 380:316-322.

Xiong, J.P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D.L., Joachimiak, A., Goodman, S.L., and Arnaout, M.A. 2001. Crystal structure of the extracellular segment of integrin αVβ3. *Science* 294:339-345.

Xu, H.E., Stanley, T.B., Montana, V.G., Lambert, M.H., Shearer, B.G., Cobb, J.E., McKee, D.D., Galardi, C.M., Plunket, K.D., Nolte, R.T., Parks, D.J., Moore, J.T., Kliewer, S.A., Willson, T.M., and Stimmel, J.B. 2002. Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARα. *Nature* 415:813-817.

Xu, R.M., Carmel, G., Sweet, R.M., Kuret, J., and Cheng, X. 1995. Crystal structure of casein kinase-1, a phosphate-directed protein kinase. *EMBO J.* 14:1015-1023.

Xu, W., Harrison, S.C., and Eck, M.J. 1997a. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 385:595-602.

Xu, Z., Horwich, A.L., and Sigler, P.B. 1997b. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* 388:741-750.

Xu, Z., Bernlohr, D.A., and Banaszak, L.J. 1993. The adipocyte lipid-binding protein at 1.6-Å resolution. Crystal structures of the apoprotein and with bound saturated and unsaturated fatty acids. *J. Biol. Chem.* 268:7874-7884.

Yaffe, M.B. 2002a. How do 14-3-3 proteins work?—Gatekeeper phosphorylation and the molecular anvil hypothesis. *FEBS Lett.* 513:53-57.

Yaffe, M.B. 2002b. Phosphotyrosine-binding domains in signal transduction. *Nat. Rev. Mol. Cell Biol.* 3:177-186.

Yamaguchi, H. and Hendrickson, W.A. 1996. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature* 384:484-489.

Yamaguchi, H., Matsushita, M., Nairn, A.C., and Kuriyan, J. 2001. Crystal structure of the atypical protein kinase domain of a TRP channel with phosphotransferase activity. *Mol. Cell* 7:1047-1057.

Yan, K.S., Kuti, M., and Zhou, M.M. 2002. PTB or not PTB — that is the question. *FEBS Lett.* 513:67-70.

Yan, Y., Winograd, E., Viel, A., Cronin, T., Harrison, S.C., and Branton, D. 1993. Crystal structure of the repetitive segments of spectrin. *Science* 262:2027-2030.

Yang, J., Liang, X., Niu, T., Meng, W., Zhao, Z., and Zhou, G.W. 1998. Crystal structure of the catalytic domain of protein-tyrosine phosphatase SHP-1. *J. Biol. Chem.* 273:28199-28207.

Yang, W., Hendrickson, W.A., Crouch, R.J., and Satow, Y. 1990. Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* 249:1398-1405.

Yoon, H.S., Hajduk, P.J., Petros, A.M., Olejniczak, E.T., Meadows, R.P., and Fesik, S.W. 1994. Solution structure of a pleckstrin-homology domain. *Nature* 369:672-675.

York, J.D., Ponder, J.W., Chen, Z.W., Mathews, F.S., and Majerus, P.W. 1994. Crystal structure of inositol polyphosphate 1-phosphatase at 2.3-Å resolution. *Biochemistry* 33:13164-13171.

Yu, H., Rosen, M.K., Shin, T.B., Seidel-Dugan, C., Brugge, J.S., and Schreiber, S.L. 1992. Solution structure of the SH3 domain of Src and identification of its ligand-binding site. *Science* 258:1665-1668.

Yuvaniyama, J., Denu, J.M., Dixon, J.E., and Saper, M.A. 1996. Crystal structure of the dual specificity protein phosphatase VHR. *Science* 272:1328-1331.

Zanotti, G., Scapin, G., Spadon, P., Veerkamp, J.H., and Sacchettini, J.C. 1992. Three-dimensional structure of recombinant human muscle fatty acid-binding protein. *J. Biol. Chem.* 267:18541-18550.

Zdanov, A., Schalk-Hihi, C., Gustchina, A., Tsang, M., Weatherbee, J., and Wlodawer, A. 1995. Crystal structure of interleukin-10 reveals the functional dimer with an unexpected topological similarity to interferon gamma. *Structure* 3:591-601.

Zeng, Z., Castano, A.R., Segelke, B.W., Stura, E.A., Peterson, P.A., and Wilson, I.A. 1997. Crystal structure of mouse CD1: An MHC-like fold with a large hydrophobic binding groove. *Science* 277:339-345.

Zhang, F., Strand, A., Robbins, D., Cobb, M.H., and Goldsmith, E.J. 1994. Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* 367:704-711.

Zhang, G. and Darst, S.A. 1998. Structure of the Escherichia coli RNA polymerase α subunit amino-terminal domain. *Science* 281:262-266.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98:811-824.

Zhang, G., Kazanietz, M.G., Blumberg, P.M., and Hurley, J.H. 1995. Crystal structure of the cys2 activator-binding domain of protein kinase Cδ in complex with phorbol ester. *Cell* 81:917-924.

Zhang, J.D., Cousens, L.S., Barr, P.J., and Sprang, S.R. 1991. Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin 1β. *Proc. Natl. Acad. Sci. U.S.A.* 88:3446-3450.

Zhang, W., Young, A.C., Imarai, M., Nathenson, S.G., and Sacchettini, J.C. 1992. Crystal structure of the major histocompatibility complex class I H-2Kb molecule containing a single viral peptide: Implications for peptide binding and T-cell receptor recognition. *Proc. Natl. Acad. Sci. U.S.A.* 89:8403-8407.

Zhang, X., Boyar, W., Toth, M.J., Wennogle, L., and Gonnella, N.C. 1997. Structural definition of the C5a C terminus by two-dimensional nuclear magnetic resonance spectroscopy. *Proteins* 28:261-267.

Zhang, X., Schwartz, J.C., Nathenson, S.G., and Almo, S.C. 2001. Crystallization and preliminary X-ray analysis of the complex between human CTLA-4 and B7-2. *Acta Crystallogr. D. Biol. Crystallogr.* 57:898-899.

Zhang, Y., Boesen, C.C., Radaev, S., Brooks, A.G., Fridman, W.H., Sautes-Fridman, C., and Sun, P.D. 2000. Crystal structure of the extracellular domain of a human Fc gamma RIII. *Immunity* 13:387-395.

Zhang, Z., Huang, L., Shulmeister, V.M., Chi, Y.I., Kim, K.K., Hung, L.W., Crofts, A.R., Berry, E.A., and Kim, S.H. 1998. Electron transfer by domain movement in cytochrome bc1. *Nature* 392:677-684.

Zhao, K., Chai, X., Johnston, K., Clements, A., and Marmorstein, R. 2001. Crystal structure of the mouse p53 core DNA-binding domain at 2.7 Å resolution. *J. Biol. Chem.* 276:12120-12127.

Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. 2000. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin- protein ligases. *Cell* 102:533-539.

Zheng, N., Schulman, B.A., Song, L., Miller, J.J., Jeffrey, P.D., Wang, P., Chu, C., Koepp, D.M., Elledge, S.J., Pagano, M., Conaway, R.C., Conaway, J.W., Harper, J.W., and Pavletich, N.P. 2002. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* 416:703-709.

Zhou, M.M., Ravichandran, K.S., Olejniczak, E.F., Petros, A.M., Meadows, R.P., Sattler, M., Harlan, J.E., Wade, W.S., Burakoff, S.J., and Fesik, S.W. 1995. Structure and ligand recognition of the phosphotyrosine binding domain of Shc. *Nature* 378:584-592.

Zhou, M.M., Huang, B., Olejniczak, E.T., Meadows, R.P., Shuker, S.B., Miyazaki, M., Trub, T., Shoelson, S.E., and Fesik, S.W. 1996. Structural basis for IL-4 receptor phosphopeptide recognition by the IRS-1 PTB domain. *Nat. Struct. Biol.* 3:388-393.

Zhou, P., Sun, L.J., Dotsch, V., Wagner, G., and Verdine, G.L. 1998. Solution structure of the core NFATC1/DNA complex. *Cell* 92:687-696.

Zhu, W., Zeng, Q., Colangelo, C.M., Lewis, M., Summers, M.F., and Scott, R.A. 1996. The N-terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nat. Struct. Biol.* 3:122-124.

Zhu, X., Komiya, H., Chirino, A., Faham, S., Fox, G.M., Arakawa, T., Hsu, B.T., and Rees, D.C. 1991. Three-dimensional structures of acidic and basic fibroblast growth factors. *Science* 251:90-93.

Zouni, A., Witt, H.T., Kern, J., Fromme, P., Krauss, N., Saenger, W., and Orth, P. 2001. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* 409:739-743.

Contributed by Peter D. Sun and Christine E. Foster
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Rockville, Maryland

Jeffrey C. Boyington
National Institute of Environmental Health Science
National Institutes of Health
Research Triangle Park, North Carolina

APPENDIX

The following section lists all of the Tables and Figures mentioned throughout the text (figures follow tables).

All ball-and-stick models and ribbon diagrams appearing in this unit were created from Protein Data Bank (PDB) coordinates (Berman et al., 2000; also see Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*) using the program MolScript (Kraulis, 1991) or Raster3D (Merritt and Bacon, 1997). Unless otherwise noted, α-helices and β-strands are colored red and green, respectively. For reasons of clarity, certain proteins are colored by domain or by subunit as designated in the figure legends. Ball-and-stick models of amino acid side chains or ligands are colored by atom type: carbon, black; oxygen, red; nitrogen, blue; and sulfur, yellow. DNA and RNA are represented by blue ball-and-stick models. Metal ions are shown as colored spheres.

Note that the black and white facsimile of the figures throughout the printed version of this unit are intended only as placeholders; for the full-color version of figures go to the Current Protocols website at Wiley Interscience (*http://www.interscience.wiley.com/c_p/colorfigures.htm*).

Table 17.1.1 Immunoglobulin and Other Immunoreceptor Folds

Protein/complex	Domain/region	Fold[a]	PDB[b]	References
Immunoglobulins				
Antibodies			1igt	Harris et al. (1997)
	Variable region	V- type		
	Constant region	C1-type		
Major histocompatibility complexes				
MHC class I			2clr	(Bjorkman et al., 1987; Collins et al., 1994b)
	Heavy chain α1+α2 domain	MHC		
	Heavy chain α3 domain	C1-type		
	β2-microglobulin	C1-type		
MHC class II			1aqd	Brown et al. (1993)
	Heavy chain α1+β1 domain	MHC		
	α chain, C-terminal domain	C1-type		
	β chain, C-terminal domain	C1-type		
Cell surface immunoglobulin-like receptors				
T-cell receptor, αβ type			1ao7	Garboczi et al. (1996)
	Vα, Vβ domains	V-type		
	Cβ domain	C1-type		
CD8, T-cell coreceptor		V-type	1cd8	Leahy et al. (1992a)
Neonatal Fcγ receptor			1fru	Burmeister et al. (1994)
	Heavy chain α3 domain	C1-type		
	β chain	C1-type		
Growth hormone receptor				
	Domains 1 and 2	C2(Fn)	3hhr	de Vos et al. (1992)
EPO receptor		C2(Fn)	1ern	Livnah et al. (1999)
Prolactin receptor				
	Domains 1 and 2	C2(Fn)	1bp3	Somers et al. (1994)
IFN-γ receptor		C2(Fn)	1fg9	Walter et al. (1995)
FGFR2		I-C2	1e0o	Pellegrini et al. (2000)
GCSFR		C2(Fn)	1cd9	Aritomi et al. (1999)
IL-10R		C2(Fn)	1j7v	Josephson et al. (2001)
IL-4R		C2(Fn)	1iar	Hage et al. (1999)
Gp130 receptor		C2(Fn)	1i1r	Chow et al. (2001)
CD4, T-cell surface glycoprotein			3cd4	Ryu et al. (1990), Wang et al. (1990)
	Domains 1 and 3	V-type[c]		
	Domains 2 and 4	C2-like		
Fcγ receptor, type III		C2-like	1fnl	Zhang et al. (2000)
Fcγ receptor, types IIA and IIB		C2-like	1fcg	Maxwell et al. (1999)
Fcε receptor, type I		C2-like	1f2q	Garman et al. (1998)

continued

Table 17.1.1 Immunoglobulin and Other Immunoreceptor Folds, *continued*

Protein/complex	Domain/region	Fold[a]	PDB[b]	References
Coagulation factor				
Tissue factor	Ectodomain	C2(Fn)	2hft	Muller et al. (1996)
Cell matrix and adhesion proteins				
Fibronectin	Type III domain	C2(Fn)	1fna	Main et al. (1992)
Tenascin, repeats		C2(Fn)	1ten	Leahy et al. (1992b)
CD2			1hnf	Jones et al. (1992)
	Adhesion domain	V-type		
	C-terminal domain	C2 like		
VCAM-1			1vca	Jones et al. (1995)
	Domain 1	I-type		
	Domain 2	C2 like		
Cadherin[d]		C1 type	1nci	Shapiro et al. (1995)
Muscle proteins				
Telokin		I-type	1tlk	Holden et al. (1992)
Titin		I-type	1tnm	Pfuhl and Pastore (1995)
Transcription factors[e]				
NF-κB, p50 subunit			1nfk	Ghosh et al. (1995), Muller et al. (1995)
	First Rel homology domain[f]	E-type	1svc	Muller et al. (1995)
	Second Rel homology domain	E-type		
Enzymes with immunoglobulin-like domains				
Chaperone protein, papD			3dpa	Holmgren and Branden (1989)
	Domain 1[g]	C2		
	Domain 2[h]	C2		
Oxidoreductase				
Cu, Zn superoxide dismutase[i]		C2	2sod	Richardson et al. (1976)
Galactose oxidase	C domain	E-type	1gof	Ito et al. (1991)
Transferase				
Cyclodextrin glycosyltransferase		E-type	1cdg	Lawson et al. (1994)
Lyase				
Chitinase A	N-terminal domain	E-type	1ctn	Perrakis et al. (1994)
C-type lectin-like receptors				
Mannose binding protein (MBP)		CTLR	1msb	Weis et al. (1991)
CD94		CTLR	1b6e	Boyington et al. (1999)
Ly48A		CTLR	1qo3	Tormo et al. (1999)
NKG2D		CTLR	1hq8	Wolan et al. (2001)
DC-SIGN		CTLR	1k9j	Feinberg et al. (2001)

continued

Table 17.1.1 Immunoglobulin and Other Immunoreceptor Folds, *continued*

Protein/complex	Domain/region	Fold[a]	PDB[b]	References
Immunoreceptor-ligand complexes				
A6 TCR/HLA-A2(Tax)		TCR-MHC	1ao7	Garboczi et al. (1996)
2C TCR/H-2Kb(dEV8)		TCR-MHC	2ckb	Garcia et al. (1998)
CD8/HLA-A2		CD8/MHC-I	1akj	Gao et al. (1997)
CD4/I-A(k)		CD4/MHC-II	1jl4	Wang et al. (2001)
B7-1/CTLA-4		B7/CTLA-4	1i84	Stamper et al. (2001)
KIR2DL2/HLA-Cw3		KIR/HLA	1efx	Boyington et al. (2000)
KIR2DL1/HLA-Cw4		KIR/HLA	1im9	Fan et al. (2001)
NKG2D/MICA		NKG2D/ligand	1hyr	Li et al. (2001)
NKG2D/ULBP3		NKG2D/ligand	1kcg	Radaev et al. (2001b)
NKG2D/Rae-1b		NKG2D/ligand	1jsk	Li et al. (2002b)
FcεRI/Fc		FcR/Fc	1f6a	Garman et al. (2000)
FcγRIII/Fc		FcR/Fc	1iis,1iix, 1e4k	Radaev et al. (2001a), Sondermann et al. (2000)
MHC-like molecules				
HLA-A2		MHC	2clr	Bjorkman et al. (1987)
HLA-Aw6B		MHC	1hsb	Garrett et al. (1989)
HLA-B27		MHC	1hsa	Madden et al. (1992)
H-2K^b		MHC	1vac	Fremont et al. (1992), Zhang et al. (1992)
HLA-DR1		MHC	1aqd	Brown et al. (1993), Stern et al. (1994)
Neonatal Fc receptor		MHC	1fru	Burmeister et al. (1994)
HLA-DM		MHC	1k8i	Fremont et al. (1998)
HLA-E		MHC	1mhe	O'Callaghan et al. (1998)
CD1		MHC	1cd1	Zeng et al. (1997)
Zinc α-2-glycoprotein (Zag)		MHC	1zag	Sanchez et al. (1999)
MIC-A		MHC	1b3j	Li et al. (1999)
T22		MHC	1c16	Wingren et al. (2000)

[a]See Proteins Involved in the Function of Immune Systems for explanations of folds.
[b]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.
[c]Strand A pairs with strand G only.
[d]Strand A pairs with strand G instead of B.
[e]See also Table 17.1.6.
[f]Insertion strands between strands C and C′, and E and F.
[g]Strand A pairs with both B and G.
[h]One additional strand at C terminus.
[i]One additional strand at N terminus.

Table 17.1.2 Complement System Protein Folds

Protein	PDB entry[a]	Reference
C3 fragment C3d	1c3d	Nagar et al. (1998)
C5a anaphylatoxin	1kjs	Zhang et al. (1997)
Complement factor D	1dsu	Narayana et al. (1994)
CD59 (1-70)	1cdq,1erg	Fletcher et al. (1994), Kieffer et al. (1994)
Factor H CCP module 16	1hcc	Norman et al. (1991), Barlow et al. (1992)
Factor H CCP modules 15 and 16	1hfh	Barlow et al. (1993)
CCP modules 3 and 4[b]	1vvc	Wiles et al. (1997)
SCR modules 1 and 2[c]	1ckl	Casasnovas et al. (1999)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

[b]From Vaccinia virus complement control protein mimic.

[c]From CD46.

Table 17.1.3 Cytokines and Receptors

Protein	PDB entry[a]	References
Four-helix bundle cytokines		
Long chain, G-CSF	1rhg	Hill et al. (1993)
Short chain, IL-4	1rcb	Rozwarski et al. (1994), Smith et al. (1994)
Interferon-γ-like	1rfb	Ealick et al. (1991)
IL-10	1ilk	Zdanov et al. (1995)
β-*trefoil cytokines*		
IL-1α	2ila	Graves et al. (1990)
IL-1β	1hib	Clore and Gronenborn (1991), Priestle et al. (1989)
Fibroblast growth factor	4fgf	Eriksson et al. (1991)
	2fgf	Zhang et al. (1991)
	1bas, 1bar	Zhu et al. (1991)
Cystine-knot cytokines		
TGF-β	2tgi	Daopin et al. (1992, 1993), Schlunegger and Grutter (1993)
Nerve growth factor	1bet	McDonald et al. (1991)
Platelet-derived growth factor Bb	1pdg	Oefner et al. (1992)
Human chorionic gonadotropin	1hrp, 1hcn	Lapthorn et al. (1994), Wu et al. (1994)
Chemokines		
IL-8	1il8	Clore et al. (1990)
MIP-1β	1hum	Lodi et al., 1994
RANTES	1rto	Skelton et al. (1995)
Neutrophil activating peptide-2	1nap	Malkowski et al. (1995)
TNF receptor superfamily		
TNFR	1tnr	Banner et al. (1993)
TRAIL	1d4v	Mongkolsapaya et al. (1999)
Receptors for cystine-knot growth factors		
ActRII	1bte	Greenwald et al. (1999)
BRIA	1es7	Kirsch et al. (2000)
TBRII	1ktz	Hart et al. (2002)
TrkA	1www	Wiesmann et al. (1999)
Domain 2 of Flt-1	1flt	Wiesmann et al. (1997)
Nuclear receptors		
TR	1bsx	Darimont et al. (1998)
RAR	2lbd	Renaud et al. (1995)
ER	2ert	Shiau et al. (1998)
PR	1a28	Williams and Sigler (1998)
PPAR	1kkq	Xu et al. (2002)
Importin-α	1ial	Kobe (1999)
Importin-β	1qgr	Cingolani et al. (1999)
Adhesion receptors		
Integrin I domain (LFA-1)	1lfa	Qu and Leahy (1995)
Integrin I domain (CR3)	1ido	Lee et al. (1995)
Integrin (intact)	1jv2	Xiong et al. (2001)
β-catenin/E-cadherin	1i7x	Huber and Weis (2001)
β-catenin/Tcf	1g3j	Graham et al. (2000)
Cysteine-rich scavenger receptor		
SRCR domain of Mac-2 binding protein	1by2	Hohenester et al. (1999)
Glutamate receptors		
iGluR	1gr2	Armstrong et al. (1998)
mGluR	1ewt, 1ewv	Kunishima et al. (2000)
Transferrin receptor		
Human transferrin receptor	1cx8	Lawrence et al. (1999)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.4 Modular Domains Involved In Signal Transduction

Protein	PDB entry[a]	References
Domains that bind phosphotyrosine		
SH2 domains		
v-src tyrosine kinase SH2 domain	1sha	Waksman et al. (1992)
p85 α subunit of PI 3-kinase SH2	2pnb	Booker et al. (1992)
C-abl SH2 domain	1ab2	Overduin et al. (1992)
PTB domains		
Shc PTB domain	1shc	Zhou et al. (1995)
IRS-1 PTB domain	1irs	Eck et al. (1996), Zhou et al. (1996)
Phosphoserine and phosphothreonine binding domains		
FHA domains		
RAD53 FHA domain	1g6g, 1dmz	Liao et al. (1999), Durocher et al. (2000)
14-3-3 domains		
14-3-3 ζ	1a40	Liu et al. (1995)
14-3-3 τ	—	Xiao et al. (1995)
Polyproline binding domains		
SH3 domains		
α Spectrin SH3 domain	1shg	Musacchio et al. (1992)
Chicken c-src SH3 domain	1rlp	Yu et al. (1992)
p85 α subunit of PI 3-kinase SH3	2pni	Booker et al. (1993)
ww domains		
Kinase associated protein WW	—	Macias et al. (1996)
Distrophin WW domain	1eg4	Huang et al. (2000)
EVH1 domains		
Murine enabled EVH1 domain	1evh	Prehoda et al. (1999)
Murine ena/vasp-like EVH1 domain	1qc6	Fedorov et al. (1999)
GYF domains		
Drosophila CD2BP2 GYF domain	1gyf	Freund et al. (1999)
Phospholipid binding domains		
PH domains		
Dynamin PH domain	1dyn	Downing et al. (1994), Ferguson et al. (1994)
β-spectrin PH domain	1btn	Macias et al. (1994)
Pleckstrin PH domain	1pls	Yoon et al. (1994)
C1 domains		
Rat protein kinase C-γ C1 domain	—	Hommel et al. (1994)
Rat protein kinase C-δ C1 domain	1ptq	Zhang et al. (1995)
C2 domains		
Synaptotagmin I C2 domain	I1rsy	Sutton et al. (1995)
Phospholipase C-δ C2 domain	1qas	Grobler et al. (1996)
Protein kinase C-α C2 domain	1dsy	Verdaguer et al. (1999)
FYVE domains		
Hrs FYVE domain	1dvp	Mao et al. (2000)
Vps27p FYVE domain	1vfy	Misra and Hurley (1999)
Eea1 FYVE domain	1joc	Dumas et al. (2001)
Protein interaction domains		
PDZ domains		
PSD-95 third PDZ domain	1be9, 1bfe	Doyle et al. (1996)
Dlg third PDZ domain	1pdr	Morais Cabral et al. (1996)
VHS domains		
Hrs VHS domain	1dvp	Mao et al. (2000)
Tom1 VHS domain	1elk	Misra et al. (2000)
GGA1 VHS domain	1jwg, 1jwf	Shiba et al. (2002)

continued

Table 17.1.4 Modular Domains Involved In Signal Transduction, *continued*

Protein	PDB entry[a]	References
GGA3 VHS domain	1juq, 1jpl	Misra et al. (2002)
SNARES		
Syntaxin-1A/synaptobrevin-II/SNAP-25B	1sfc	Sutton et al. (1998)
SAM domains		
EphB2 receptor SAM domain	1b4f, 1sgg	Smalla et al. (1999), Thanos et al. (1999)
EphA4 receptor SAM domain	1b0x	Stapleton et al. (1999)
Polyhomeotic SAM domain	1kw4	Kim et al. (2002)
MH2 domains		
Smad2 MH2 domain	1dev	Wu et al. (2000)
Smad3 MH2 domain	1mk2	Qin et al. (2002)
Structural repeat motifs		
Leucine rich repeats (LRRs)		
Ribonuclease inhibitor	2bnh	Kobe and Deisenhofer (1993)
Ribonuclease inhibitor/RNAse A	1dfj	Kobe and Deisenhofer (1995)
Internalin B	1d0b	Marino et al. (1999)
RanGAP protein rna1p	1yrg	Hillig et al. (1999)
HEAT repeats		
Protein phosphatase 2A PR65/A subunit	1b3u	Groves et al. (1999)
Karyopherin 2/Ran-GppNHp complex	1qbk	Chook and Blobel (1999)
Armadillo (ARM) repeats		
Karyopherin α	1bk5, 1bk6	Conti et al. (1998)
β-catenin	3bct	Huber et al. (1997)
β-catenin bound to E-cadherin	1i7w	Huber and Weis (2001)
Ankyrin (ANK) repeats		
GABPβ	1awc	Gorina and Pavletich (1996)
p16INK4a bound to Cdk6	1bi7	Russo et al. (1998)
IκBα/NF-κB complex	1ikn	Huxford et al. (1998)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*. For proteins with no entry listed, coordinates have not been deposited in the Protein Data Bank.

Table 17.1.5 Kinases and Phosphatases

Protein	PDB entry[a]	References
Protein kinases		
Protein tyrosine kinases		
Tyrosine kinase C-src	1fmk	Xu et al. (1997a)
Lymphocyte cell kinase (lck)	3lck	Yamaguchi and Hendrickson (1996)
Haematopoietic cell kinase (hck)	1ad5	Sicheri et al. (1997)
C-terminal src kinase	1byg	Lamers et al. (1999)
Tie2 protein tyrosine kinase	1fvr	Shewchuk et al. (2000)
Human insulin receptor kinase	1irk	Hubbard et al. (1994)
Insulin-like growth factor-1 receptor	1k3a	Favelyukis et al. (2001)
FGF receptor-1	1fgk	Mohammadi et al. (1996)
EPHB2 receptor	1jpa	Wybenga-Groot et al. (2001)
Vascular growth factor receptor-2	1vr2	McTigue et al. (1999)
Protein serine/threonine kinases		
Cyclic AMP kinases	1atp	Knighton et al. (1991), Olah et al. (1993)
Cyclin-dependent kinase 2	1bi7	De Bondt et al. (1993), Jeffrey et al. (1995)
Cyclin A-CDK	1fin	Jeffrey et al. (1995)
Cell cycle protein CksHs1	1dks	Arvai et al. (1995)
MAP kinase ERK2	1gol	Zhang et al. (1994)
MAP kinase p38	1p38	Wang et al. (1997)
Phosphorylase kinase	1phk	Owen et al. (1995)
Calcium/calmodulin-dependent kinase	1ao6	Goldberg et al. (1996)
Twitchin kinase	1koa	Hu et al. (1994)
Casein kinase I	1cki	Xu et al. (1995)
Death associated protein kinase	1jks	Tereshko et al. (2001)
Phosphatases		
Protein tyrosine phosphatases		
Subtype I		
Low molecular weight	1phr	Su et al. (1994)
Subtype II		
High molecular weight		
PTB1	2hnq	Barford et al. (1994)
SHP-1	1gwz	Yang et al. (1998)
SHP-2	2shp	Hof et al. (1998)
Yersinia PTP	1ypt	Stuckey et al. (1994)
SptP tyrosine phosphatase domain	1g4u	Stebbins and Galan (2000)
RPTP LAR tandem phosphatase	1lar	Nam et al. (1999)
Dual specificity		
VHR	1vhr	Yuvaniyama et al. (1996)
Kinase associated phosphatase (kap)	1fpz	Song et al. (2001)
MAP kinase phosphatase	1mkp	Stewart et al. (1999)
PTEN tumor suppressor	1d5r	Lee et al. (1999)
Third subtype		
CDC25A	1c25	Fauman et al. (1998)
ERK2 binding domain of MKP-3	1hzm	Farooq et al. (2001)
Protein serine/threonine phosphatases		
PPP family enzymes		
PP1	1fjm	Goldberg et al. (1995)
Calcineurin	1aui	Kissinger et al. (1995)
Bacteriophage λ phosphatase	1g5b	Voegtli et al. (2000)
Purple acid phosphatase	1kbp	Strater et al. (1995)

continued

Table 17.1.5 Kinases and Phosphatases, continued

Protein	PDB entry[a]	References
UDP sugar hydrolase	1ush	Knofel and Strater (1999)
MRE11 nuclease	1ii7	Hopfner et al. (2001)
PPM family		
Protein phosphatase 2C	1a6q	Das et al. (1996)
Alkaline phosphatase fold		
Alkaline phosphatase	1ali	Kim and Wyckoff (1991)
Arylsulfatase	1auk	Lukatela et al. (1998)
Prokaryotic cofactor-independent phosphoglycerate mutase	1eqj	Jedrzejas et al. (2000)
Sugar phosphatases		
Fructose 1,6-biosphastase	5fbp	Ke et al. (1991)
Inositol monophosphatase	2hhm	Bone et al. (1992)
Inositol polyphosphate 1-phosphatase	1inp	York et al. (1994)
3,5-adenosine bisphosphatase	1qgx	Albert et al. (2000)
PIPase	1jp4	Patel et al. (2002)
The cyclin fold		
Cyclin A	1fin	Jeffrey et al. (1995)
Transcription factor IIB	1vol	Nikolov et al. (1995)
Retinoblastoma tumor suppressor protein	1gux	Lee et al. (1998)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.6 DNA-Binding Proteins

Protein	PDB entry[a]	Reference
Helix-turn-helix motifs		
Prokaryotic and phage repressors		
λ Cro protein	1cro	Anderson et al. (1981)
λ-repressor	1lmb	Jordan and Pabo (1988)
434 repressor	2or1	Aggarwal et al. (1988)
434 cro protein	2cro	Wolberger et al. (1988)
Homeodomains or homeodomain-like		
Engrailed	1hdd	Kissinger et al. (1990)
Antennapedia	1hom	Qian et al. (1989)
Oct-1 POU	1oct	Klemm et al. (1994)
Proto onco gene product Myb	1mse	Ogata et al. (1994)
Hin-related prokaryotic DNA recombinase	1hcr	Feng et al. (1994)
Winged helix		
HNF-3/fork head transcription factor	2hfh	Clark et al. (1993), Marsden et al. (1998)
Globular domain of histone H5	1hst	Ramakrishnan et al. (1993)
ETS transcription factor	1etc	Donaldson et al. (1996)
E. coli CAP protein	1cgp	Schultz et al. (1991)
Biotin repressor	1bia	Wilson et al. (1992)
PurR repressor		
PurR repressor	1pru	Schumacher et al. (1994)
Zinc-containing DNA-binding motifs		
Cys_2 His_2 zinc fingers		
ZIF286	1zaa	Pavletich and Pabo (1991)
GLI oncogene product	2gli	Pavletich and Pabo (1993)
Tramtrack	2drp	Fairall et al. (1993)
Zinc ribbon motif		
Transcription elongation factor TFIIS	1tfi	Qian et al. (1993)
Transcription initiation factor TFIIB	1pft	Zhu et al. (1996)
RNA Polymerase II RPB9 fragment	1qyp	Wang et al. (1998)
DNA primase fragment	1d0q	Pan and Wigley (2000)
GATA-binding motif		
GATA-1	1gat	Omichinski et al. (1993)
Zn_2Cys_6 binuclear cluster		
GAL-4	1d66	Marmorstein et al. (1992)
Pyrimidine pathway regulator 1(PPR1)	1pyi	Marmorstein and Harrison (1994)
PUT3	1zme	Swaminathan et al. (1997)
CD2-LAC9	1cld	Gardner et al. (1995)
Nuclear hormone receptor DNA-binding domains		
Glucocorticoid receptor	1gdc	Luisi et al. (1991)
Estrogen receptor	1hcq	Schwabe et al. (1993)
Orphan nuclear receptor NGFI-B	1cit	Meinke and Sigler (1999)
Thyroid hormone receptor	2nll	Rastinejad et al. (1995)
MADS box domain		
Serum response factor	1srs	Pellegrini et al. (1995)
MCM1 transcriptional regulator	1mnm	Tan and Richmond (1998)
MEF2A core	1egw	Santelli and Richmond (2000)
Basic leucine zipper		
GCN4 homodimer	1ysa	Ellenberger et al. (1992)
c-Fos:c-Jun heterodimer	1fcs	Glover and Harrison (1995)

continued

Table 17.1.6 DNA-Binding Proteins, *continued*

Protein	PDB entry[a]	Reference
Helix-loop-helix		
MAX transcription factor	2an2	Ferre-D'Amare et al. (1993)
MyoD transcription activator	1mdy	Ma et al. (1994)
Upstream stimulatory factor	1an4	Ferre-D'Amare et al. (1994)
Transcription factor E47	—	Ellenberger et al. (1994)
HMG box domain		
Human SRY	1hry	Werner et al. (1995)
LEF-1 transcription factor	1lef	Love et al. (1995)
HMG1 B domain	1hme	Weir et al. (1993)
Histone-fold proteins		
Nucleosomal core octamer (H2A, H2B, H3, H4)	1hio	Arents et al. (1991)
Drosophila $TAF_{II}42$ and $TAF_{II}62$	1taf	Xie et al. (1996)
Human $TAF_{II}28$ and $TAF_{II}18$	1bh9	Birck et al. (1998)
Negative cofactor 2	1jfi	Kamada et al. (2001)
HMFA and HMFB Archeon histones	1b67	Decanniere et al. (2000)
β-*sheet motifs*		
Ribbon-helix-helix		
Arc repressor	1par	Raumann et al. (1994)
Met repressor[b]	1cma	Somers and Phillips (1992)
Mnt repressor	1mnt	Burgering et al. (1994)
Histone-like HU family		
Histone-like protein (HU)	1hue	Tanaka et al. (1984), Vis et al. (1995)
Integration host factor (IHF)/DNA complex	1ihf	Rice et al. (1996)
DNA-binding protein TF1	1wtu	Jia et al. (1996)
TBP family		
TATA box–binding protein/DNA complex	1ytb/1tgh	Kim et al. (1993a,b)
Ig-like transcription factors		
p53 tumor supressors		
Human p53/DNA complex	1tsr, 1tup	Cho et al. (1994)
Mouse p53 DNA-binding domain	1hu8	Zhao et al. (2001)
NF-κB		
p50/p50/DNA complex	1nfk, 1svc	Ghosh et al. (1995), Muller et al. (1995)
p52/p52/DNA complex	1a3q	Cramer et al. (1997)
p65/p65/DNA complex	1ram	Chen et al. (1998d)
p65/p50/DNA complex	1vkx	Chen et al. (1998a)
NFAT		
NFAT1/Fos/Jun/DNA complex	1a02	Chen et al. (1998b)
NFATC1/DNA complex	1a66	Zhou et al. (1998)
TonEBP/DNA complex	1imh	Stroud et al. (2002)
CBF		
Human PEBP2(CBFα) runt domain	1cmo	Nagata et al. (1999)
Human CBFα runt domain	1co1	Berardi et al. (1999)
Human Runx-1(CBFα)/CBFβ/DNA complex	1h9d	Bravo et al. (2001)
Mouse Runx-1(CBFα)/CBFβ/DNA	1hjb	Tahirov et al. (2001)
Mouse Runx-1(CBFα)/CBFβ/c-EBPβ/DNA	1io4	Tahirov et al. (2001)
T-domain		
Xenopus laevis Brachyuri T-domain/DNA complex	1xbr	Muller and Herrmann (1997)
Human TBX3 T-domain/DNA complex	1h6f	Coll et al. (2002)
STAT proteins		
STAT-1/STAT-1/DNA complex	1bf5	Chen et al. (1998c)
STAT3β/STAT3β/DNA complex	1bg1	Becker et al. (1998)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*. For proteins with no entry listed, coodinates have not been deposited in the Protein Data Bank.
[b]This protein is also known as MetJ.

Table 17.1.7 RNA-Binding Proteins

Protein	PDB entry[a]	Reference
RNP domain		
U1A splicosome	1urn	Oubridge et al. (1994)
hnRNP C	1sxl	Wittekind et al. (1992)
Ribosomal protein S6	1ris	Lindahl et al. (1994)
T4 regA translational regulator	1reg	Kang et al. (1995)
OB-fold		
B. subtilis major cold-shock protein (CspB)	1csp, 1nmg	Schindelin et al. (1993)
E. coli major cold-shock protein (CspA)	1mef, 1mjc	Newkirk et al. (1994), Schindelin et al. (1994)
Ribosomal protein L14	1whi	Davies et al. (1996)
Ribosomal protein S17	1rip	Golden et al. (1993)
N-terminal domain of lysyl and aspartyl tRNA synthetases	1asz, 1lyl	Cavarelli et al. (1993), Onesti et al. (1995)
KH domain		
Vigilin	1vih	Castiglone Morelli et al. (1995)
dsRNA-binding domain		
E. coli RNase III		Kharrat et al. (1995)
Drosophila staufen protein	1stu	Bycroft et al. (1995)
N-terminal domain of ribosomal protein s5	1pkp	Ramakrishnan and White (1992)
Retroviral CCHC zinc fingers		
NCp7 complex with dACGCC	1aaf	South and Summers (1993)
Ncp10	1a6b	Demene et al. (1994)
Four-helix bundle RNA-binding protein		
E. coli ROP	1rop	Predki et al. (1995)
Other RNA-binding proteins		
MS2 phage coat protein	1mst	Valegard et al. (1994)
Trp RNA-binding attenuation protein (TRAP)	1wap	Antson et al. (1995)
Elongation factor-Tu (EF-Tu)	1ttt	Nissen et al. (1995)
SRP domains		
SRP9/14 in complex with RNA	1e8o, 1e8s	Weichenrieder et al. (2000)
SRP19 in complex with RNA	1jid	Wild et al. (2001)
Ffh	2ffh	Keenan et al. (1998)
M domain and RNA complex	1dul	Batey et al. (2000)
SRP receptor NG domain	1fts	Montoya et al. (1997)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.8 Aminoacyl-tRNA Synthetases

Protein	Subunits	PDB entries[a]
Class I		
Subclass Ia		
ArgRS	α	1bs2, 1f7u, 1f7v, 1iq0
CysRS	α	li5, 1li7
IleRS	α	1ile, 1ffy, 1jzq, 1jzs, 1qu2, 1qu3
LeuRS	α	1h3n
MetRS	α, α_2	1a8h, 1f4l, 1qqt
ValRS	α	1gax
LysRS I[b]	α	1irx
Subclass Ib		
GlnRS	α	1euq, 1euy, 1exd, 1gsg, 1gtr, 1gts, 1qrs, 1qrt, 1qru, 1qtq
GluRS	α	1g59, 1gln
Subclass Ic		
TrpRS	α_2	1d2r, 1i6k, 1i6l, 1i6m, 1m83, 1mau, 1maw, 1mb2
TyrRS	α_2	1tya, 1tyb, 1tyc, 1tyd, 1h3f, 1jii, 1jij, 1jik, 1jil, 2ts1, 3ts1, 4ts1, 1h3e
Class II		
Subclass IIa		
GlyRS[c]	α_2	1ati, 1b76, 1ggm
HisRS	α_2	1h4v, 1qe0, 1adj, 1ady, 1htt, 1kmm, 1kmn
ProRS	α_2	1hc7, 1h4q, 1h4s, 1h4t
SerRS	α_2	1sry, 1ser, 1ses, 1set
ThrRS	α_2	1evk, 1evl, 1fyf, 1kog, 1qf6
Subclass IIb		
AspRS	α_2	1l0w, 1b8a, 1eov, 1eqr, 1g51, 1asy, 1asz, 1coa, 1efw, 1il2
AsnRS	α_2	11as, 12as
LysRS II[b]	α_2	1bbw, 1bbu, 1e1o, 1e1t, 1e22, 1e24, 1lyl
Subclass IIc		
PheRS	$(\alpha\beta)_2$	1ehz, 1pys, 1b70, 1b7y, 1eiy, 1jjc
AlaRS[c]	α_4, α	–[c]
GlyRS[d]	$(\alpha\beta)_2$	–[c]

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.
[b]There are both class I and class II LysRS
[c]As of 2003, structures have not been determined.
[d]GlyRS is found in two subclasses: IIa and IIc.

Table 17.1.9 Carbohydrate-Binding Proteins

Protein	PDB entry[a]	Reference
Legume lectins		
Soybean agglutinin	2sba	Dessen et al. (1995)
Cereal lectins		
Wheat germ agglutinin	2wgc	Wright and Jaeger (1993)
C-type		
Mannose-binding	2msb	Weis et al. (1992)
E-selectin	1esl	Graves et al. (1994)
S-type		
S-lectin	1slt	Liao et al. (1994)
S-lac lectin	1hlc	Lobsanov et al. (1993)
β-*trefoil lectins*		
Ricin B	1aai	Rutenber and Robertus (1991)
Amaranthus agglutinin	1jlx, 1jly	Transue et al. (1997)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.10 Calcium-Binding Proteins

Protein	PDB entry[a]	Reference
EF-hand motif		
Calmodulin	1cdl	Meador et al. (1992)
Chick skeletal muscle troponin C	1top	Satyshur et al. (1988)
Turkey skeletal muscle troponin C	5tnc	Herzberg and James (1988)
Calbindin	4icb	Svensson et al. (1992)
Myosin subfragment	1mys	Rayment et al. (1993)
Scallop myosin	1scm	Xie et al. (1994)
Recoverin	1rec	Flaherty et al. (1993)
BM-40	1bmo	Hohenester et al. (1996)
Tr_{1c} fragment of sketetal muscle troponin C	1trf	Findlay et al. (1994)
Annexins		
Annexin V	1ala	Huber et al. (1990a,b), Bewley et al. (1993)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.11 Nucleotide-Binding Proteins

Protein	PDB entry[a]	Reference
Dinucleotide-binding fold		
Lactate dehydrogenase	8ldh	Adams et al. (1970)
Alcohol dehydrogenase	6adh	Eklund et al. (1981)
Carbonyl reductase	1cyd	Tanaka et al. (1996)
Glyceraldehyde 3-phosphate dehydrogenase	1gpd	Moras et al. (1975)
Glutaminyl tRNA synthetase	1gtr	Rould et al. (1991)
Tyrosyl-tRNA synthetase	1tyb	Irwin et al. (1976)
Mononucleotide-binding fold		
Adenylate kinase	3adk	Schulz et al. (1974)
Guanylate kinase	1gky	Stehle and Schulz (1990)
Uridylate kinase	1uky	Muller-Dieckmann and Schulz (1994)
Deoxyribonucleoside kinase	1j90	Johansson et al. (2001)
Estrogen sulfotransferase	1aqu	Kakuta et al. (1997)
Flavodoxin-like mononucleotide-binding fold		
Flavodoxin	1fxl	Watenpaugh et al. (1972)
CheY chemotactic protein	3chy	Volz and Matsumura (1991)
Acetylhydrolase	1es9	Ho et al. (1997)
Methionine synthase	1bmt	Drennan et al. (1994)
Ribosomal protein S2	1j5e	Wimberly et al. (2000)
Resolvase-like mononucleotide-binding fold		
γδ-resolvase	2rsl	Sanderson et al. (1990)
5′-3′ exonuclase from Taq DNA polymerase	1taq	Kim et al. (1995b)
G-protein mononucleotide-binding fold		
H-ras p21	121p	Pai et al. (1990)
Elongation factor Tu (EF-Tu)	1ttt	Nissen et al. (1995)
Elongation factor G (EF-G)	1efg	Czworkowski et al. (1994)
CDC42	1grn	Nassar et al. (1998)
Transducin α	1tnd	Noel et al. (1993)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.12 Replication and Transcription Enzymes

Protein	PDB entry[a,b]	Reference
DNA polymerases		
Pol I family (A family)		
Klenow fragment from *E. coli*	1kln	Beese et al. (1993)
Pol α family (B family)		
Bacteriophage RB69 DNA polymerase	1ih7, 1ig9	Franklin et al. (2001)
Pol β family (X family)		
Polymerase β from rat	1bpd	Sawaya et al. (1994)
Lesion bypass polymerase family (Y family)		
DNA polymerase IV from *S. solfataricus*	1jx4, 1jxl	Ling et al. (2001)
Reverse transcriptase family		
HIV-1 reverse transcriptase	2hmi	Jacobo-Molina et al. (1993)
Small RNA polymerases		
T7 polymerase family		
Bacteriophage T7 polymerase	4rnp	Sousa et al. (1993)
Viral RNA-dependent RNA polymerase family		
Poliovirus 3D RNA polymerase	1rdr	Hansen et al. (1997)
dsRNA phage RNA–dependent RNA polymerase family		
Bacteriophage Φ6 RNA polymerase	1hhs	Butcher et al. (2001)
Eukaryotic poly(A) RNA polymerases		
Yeast poly(A) RNA polymerase	1fa0	Bard et al. (2000)
Bovine poly(A) RNA polymerase	1f5a	Martin et al. (2000)
DNA-dependent RNA polymerases (RNAPs)		
T. aquaticus core RNAP ($\alpha_2\beta\beta'\omega$)	1hqm, 1i6v	Zhang et al. (1999), Campbell et al. (2001)
T. aquaticus holoenzyme RNAP ($\alpha_2\beta\beta'\omega\sigma^{70}$)	1l9u	Murakami et al. (2002a)
T. thermophilus holoenzyme RNAP ($\alpha_2\beta\beta'\omega\sigma^{70}$)	1iw7	Vassylyev et al. (2002)
E. coli RNAP α subunit N-terminal domain	1bdf	Zhang and Darst (1998)
E. coli RNAP α subunit C-terminal domain	1coo	Jeon et al. (1995)
E. coli RNAP σ^{70} initiation factor	1sig	Malhotra et al. (1996)
T. aquaticus RNAP σ^{A} factor	1ku2, 1ku3, 1ku7	Campbell et al. (2002)
Saccharomyces cerevisiae RNAP II (10 subunits)	1en0, 1I50	Cramer et al. (2000, 2001)
DNA polymerase processivity factors		
Dimeric processivity factors		
E. coli β subunit of polymerase III	2pol	Kong et al. (1992)
Trimeric processivity factors		
Proliferating cell nuclear antigen (PCNA)	1plq	Krishna et al. (1994)
Topoisomereases		
Subtype IA		
E. coli topoisomerase I 67-kDa fragment	1ecl	Lima et al. (1994)
Subtype IB		
Human topoisomerase I	1a36	Redinbo et al. (1998), Stewart et al. (1998)
Subtype IIA		
Yeast topoisomerase II 92-kDa fragment	1bgw	Berger et al. (1996)

continued

Table 17.1.12 Replication and Transcription Enzymes, *continued*

Protein	PDB entry[a,b]	Reference
E coli GyrB 43-kDa fragment	1ei1	Wigley et al. (1991), Brino et al. (2000)
Subtype IIB		
M. jannaschii topoisomerase VI-A′	1d3y	Nichols et al. (1999)
RNAse H-like polynucleotide transferases		
E. coli RNase H	2m2	Katayanagi et al. (1990), Yang et al. (1990), Katayanagi et al. (1992)
HIV-I reverse transcriptase RNase H	1hrh	Davies et al. (1991)
RuvC resolvase	1hjr	Ariyoshi et al. (1994)
HIV-I integrase	1itg	Dyda et al. (1994)
Mu transposase	1beo	Rice and Mizuuchi (1995)
TN5 transposase	1f3i	Davies et al. (2000)
Klenow fragment 3′-5′ exonuclease	1kln	Beese et al. (1993)

Table 17.1.13 Proteins Involved in Cytoskeleton and Muscle Movements

Protein	PDB entry[a]	Reference
Actin fold		
Actin	1atn	Kabsch et al. (1990)
70-kDa heat-shock cognate protein	1nga	Flaherty et al. (1990)
Hexokinase	2yhx	Anderson et al. (1978)
Glycerol kinase	2gla	Hurley et al. (1993)
Actin-binding protein		
Profilin	2btf	Schutt et al. (1993)
Actin depolymerizing proteins		
Gelsolin segment I	—	McLaughlin et al. (1993)
Actin cross-linking proteins		
Spectrin	1spc	Yan et al. (1993)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*. For proteins with no entry listed, coordinates have not been deposited in the Protein Data Bank.

Table 17.1.14 Enzymes

Protein	PDB entry[a]	Reference
α/β ***Hydrolases***		
Haloalkane dehaologenase	2had	Franken et al. (1991)
Torpedo californica acetylcholinesterase	1ack	Sussman et al. (1991)
Pseudomonas glumae triacylglycerol lipase	1tah	Noble et al. (1993)
Sialidases		
Influenza A neuraminidase	1nn2	Varghese and Colman (1991)
Influenza B neuraminidase	1nsb	Burmeister et al. (1992)
Salmonella typhimurium LT2 sialidase	1dil	Crennell et al. (1993)
TIM barrel fold		
Chicken muscle triose phosphate isomerase	1tim	Banner et al. (1975)
Serine proteases		
Trypsin-like		
α-Chymotrypsin	2cha	Birktoft and Blow (1972)
Subtilisins		
Subtilisin BPN′	1sbt	Alden et al. (1971)
Serine carboxypeptidases		
Wheat carboxypeptidase II	1wht	Liao and Remington (1990)
Herpes-like serine protease		
Cytomegalovirus protease	1cmv, 1wpo, 1lay	Qiu et al. (1996), Shieh et al. (1996), Tong et al. (1996)
Cysteine proteases		
Papain-like		
Papain	9pap	Kamphuis et al. (1984)
Caspases		
ICE (Caspase-1)	1ice	Wilson et al. (1994)
Viral trypsin-like		
Hepatitus A virus 3C proteinase	1hav	Allaire et al. (1994)
Aspartic proteases		
Pepsin-like		
Human rennin	2ren	Sielecki et al. (1989)
Retroviral		
HIV-1 Aspartyl protease	2hvp	Navia et al. (1989)
Metalloproteases		
Zincins		
Thermolysin	4tln	Holmes and Matthews (1982)
Astacin	1ast	Bode et al. (1992)
Zinc-dependent exopeptidases		
Leucine aminopeptidase	1lap	Burley et al. (1990)
Bovine carboxypeptidase A	5cpa	Rees et al. (1983)
Proteasome (Ntn hydrolase fold)		
Archeon *T. acidophilum* 20S proteasome	1pma	Lowe et al. (1995)
Yeast 20S proteasome	1ryp	Groll et al. (1997)
Bovine 20S proteasome	1iru	Unno et al. (2002)
E. coli HslV	1ned	Bochtler et al. (1997)
Proteins in the ubiquitin pathway		
Ubiquitin	1ubq	Vijay-Kumar et al. (1987)
c-Cbl-UbcH7	1fbv	Zheng et al. (2000)
Ubc9-RanGAP1	1kps	Bernier-Villamor et al. (2002)
Siah	1k2f	Polekhina et al. (2002)
hMms2-Ubc13	1j74,1j7d	Moraes et al. (2001)
Mms2-Ubc13	1jat	VanDemark et al. (2001)
Skp1-Skp2	1fs2,1fqv	Schulman et al. (2000)
Cul1-Rbx1-Skp1-F-box-Skp2	1ldd,1ldj,1ldk	Zheng et al. (2002)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.15 Electron-Transfer Proteins

Protein	PDB entry[a]	Reference
Cytochrome $P450_{CAM}$	1phc	Raag et al. (1993)
Yeast iso-I-cytochrome c	1ycc	Louie and Brayer (1990)
Cytochrome b5	1cyo	Mathews et al. (1972)
Ferritin	1bcf	Frolow et al. (1994)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.16 Globin-Like Proteins

Protein	PDB[a] entry	Reference
Globins		
Oxymyoglobin	1mbd	Phillips (1980)
Human deoxyhemoglobin	1hhb	Fermi et al. (1984)
Sea cucumber hemoglobin	1hlm, 1hlb	Mitchell et al. (1995)
Phycocyanins		
Cyanobacterial C-phycocyanin	1cpc	Schirmer et al. (1987)
Colicins		
Colicin A	1col	Parker et al. (1989)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.17 Toxins

Protein	PDB entry[a]	Reference
Small single-subunit toxins		
Margatoxin	1mtx	Johnson et al. (1994)
Androctonus australis hector toxin II	1ptx	Housset et al. (1994)
Snake venom neurotoxin	1nxb	Tsernoglou et al. (1978)
Agelenopsis aperta venom	1agg	Reily et al. (1995)
Ribosome-inactivating toxins		
Ricin A	1rtc	Katzin et al. (1991)
ADP ribosylation toxins		
Diphtheria toxin	1ddt	Choe et al. (1992)
Pseudomonas aeruginosa exotoxin A	1ikq	Wedekind et al. (2001)
E. coli enterotoxin	1ltt	Sixma et al. (1991)
Pertussis toxin	1prt	Stein et al. (1994)
Cholera toxin	1chp	Merritt et al. (1995)
Superantigen toxins		
Staphylococcus enterotoxin C2	1se2	Swaminathan et al. (1995)
Staphylococcus enterotoxin B	—	Swaminathan et al. (1992)
Toxic shock syndrome toxin I	3tss	Prasad et al. (1997)
Anthrax toxin		
Leathal factor (LF)	1j7n,1jky	Pannifer et al. (2001)
Oedema factor (EF)	1k8t,1k93,1k90	Drum et al. (2002)
Edema factor (EF)	1lvc	Shen et al. (2002)
Protective antigen (PA)	1acc	Petosa et al. (1997)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*. For proteins with no entry listed, coodinates have not been deposited in the Protein Data Bank.

Table 17.1.18 Lipid-Binding Proteins

Protein	PDB entry[a]	Reference
Lipocalins		
Pieris brassicae bilin-binding protein	1bbp	Huber et al. (1987)
Human retinol binding protein	1rbp	Cowan et al. (1990)
Bovine odorant-binding protein	1obp	Bianchet et al. (1996)
Bovine β-lactoglobulin	3blg, 2blg	Qin et al. (1998)
Fatty acid-binding protein like		
Rat intestinal FABP	2ifb	Sacchettini et al. (1989)
Human muscle FABP	2hmb	Zanotti et al. (1992)
Rat liver FABP	1lfo	Thompson et al. (1997)
Rat cellular retinol-binding protein II	1opa, 1opb	Winter et al. (1993)
Human retinoic acid-binding protein II	1cbs	Kleywegt et al. (1994)
Bovine P2 myelin protein	1pmp	Cowan et al. (1993)
Murine adipocyte lipid-binding protein	1lid	Xu et al. (1993)
Serum albumin		
Human serum albumin	1uor, 1ao6	He and Carter (1992), Sugio et al. (1999)
Human serum albumin liganded	1bj5, 1bke	Curry et al. (1998)
Horse serum albumin	—	Ho et al. (1993)

Table 17.1.19 Large Multisubunit Proteins

Protein	PDB entry[a]	Reference
Chaperonin complexes		
GroEL-like		
E. coli GroEL	1oel, 1aon	Braig et al. (1994), Braig et al. (1995), Xu et al. (1997b)
P. denitrificans chaperonin-60	1iok	Fukami et al. (2001)
GroES-like		
E. coli GroES	1aon	Xu et al. (1997b)
M. leprae chaperonin-10	1lep	Hunt et al. (1996), Mande et al. (1996)
M. tuberculosis chaperonin-10	1hx5	Taneja and Mande (2002)
T4 co-chaperonin gp31	1g31	Hunt et al. (1997)
Myosin		
Scallop myosin subfragment S1	1b7t	Houdusse et al. (1999)
Chicken myosin S1	2mys	Rayment et al. (1993)
Chicken myosin motor domain	1br2	Dominguez et al. (1998)
Slime mold myosin motor domain	1lkx	Kollmar et al. (2002)
G-proteins and associated regulators		
G-protein $\alpha\beta\gamma$ heterotrimer	1gp2, 1got	Lambright et al. (1996), Wall et al. (1995)
p21Ras	1gnr	Pai et al. (1990)
p120GAP	1wer	Scheffzek et al. (1996)
Ras-Sos complex	1bkd	Boriack-Sjodin et al. (1998)
Ran	1k5d,1ibr	Vetter et al. (1999)
RhoA-RhoGAP	1tx4	Rittinger et al. (1997a)
GBP1	1f5n	Prakash et al. (2000)
RGS4	1agr	Tesmer et al. (1997)
GoLoco	1kjy	Kimple et al. (2002)
F_1 ATPases		
Bovine mitochondrial $\alpha_3\beta_3\gamma\delta\varepsilon$	1cow, 1e79	Abrahams et al. (1994), Gibbons et al. (2000)
Rat liver mitochodrial $\alpha_3\beta_3\gamma$	1mab	Bianchet et al. (1998)
Spinach chloroplast $\alpha_3\beta_3\gamma\varepsilon$	1fx0	Groth and Pohl (2001)
Thermophilic bacillus $\alpha_3\beta_3$	1sky	Shirakihara et al. (1997)
E. coli γ-ε	1fs0	Rodgers and Wilce (2000)
F_1F_0 ATPase		
Yeast mitochondrial $\alpha_3\beta_3\gamma\delta c_{10}$	1qo1	Stock et al. (1999)
Proteasome		
Archeon *T. acidophilum* 20S proteasome	1pma	Lowe et al. (1995)
Yeast 20S proteasome	1ryp	Groll et al. (1997)
Yeast 20S proteasome and *T. brucei* 11S regulator	1fnt	Whitby et al. (2000)
Bovine 20S proteasome	1iru	Unno et al. (2002)
E. coli HslV	1ned	Bochtler et al. (1997)
Human Regα 11S regulator	1avo	Knowlton et al. (1997)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

Table 17.1.20 Integral Membrane Proteins of Known Structure

Protein	PDB entry[a]	Reference
Photosynthetic proteins		
Photosynthetic reaction center	1prc	Deisenhofer et al. (1995)
	1pss	Chirino et al. (1994)
	1l9j	Axelrod et al. (2002)
Photosystem I	1jb0	Jordan et al. (2001)
Photosystem II	1fe1	Zouni et al. (2001)
Light-harvesting complex	1kzu	Prince et al. (1997)
	1lgh	Koepke et al. (1996)
Seven-helix proteins		
Microbial rhodopsins		
Bacteriorhodopsin	1c3w	Luecke et al. (1999)
	1ap9	Pebay-Peyroula et al. (1997)
	2brd	Grigorieff et al. (1996)
	1at9	Kimura et al. (1997)
Halorhodopsin	1e12	Kolbe et al. (2000)
Sensory rhodopsin II	1jgj	Luecke et al. (2001)
	1h68	Royant et al. (2001)
G protein-coupled receptors		
Rhodopsin	1f88	Palczewski et al. (2000)
Transporters and channels		
AcrB multidrug efflux transporter	1iwg	Murakami et al. (2002b)
Aquaporin water channel	1fqy	Murata et al. (2000)
	1ih5	Ren et al. (2001)
	1j4n	Sui et al. (2001)
GlpF glycerol channel	1fx8	Fu et al. (2000)
Ion channels		
KcsA potassium channel	1bl8	Doyle et al. (1998)
MthK potassium channel	1lnq	Jiang et al. (2002a,b)
MscL mechanosensitive channel	1msl	Chang et al. (1998)
MscS voltage-modulated mechanosensitive channel	1mxm	Bass et al. (2002)
ClC chloride channel	1kpk, 1kpl	Dutzler et al. (2002)
P-type ATPases		
Calcium ATPase	1eul	Toyoshima et al. (2000)
	1iwo	Toyoshima and Nomura (2002)
ATP-binding-cassette transporters		
MsbA lipid flippase	1jsq	Chang and Roth (2001)
BtuCD vitamin B_{12} transporter	1l7v	Locher et al. (2002)
Respiratory enzymes		
Fumarate reductase	1l0v	Iverson et al. (1999)
	1qla, 1qlb	Lancaster et al. (1999)
Cytochrome *c* oxidases	1occ	Tsukihara et al. (1996)
	1ar1	Ostermeier et al. (1997)
	1ehk	Soulimane et al. (2000)
Cytochrome bc_1 complex	1qcr	Xia et al. (1997)
	1bcc	Zhang et al. (1998)
	1bgy	Iwata et al. (1998)
	1ezv	Hunte et al. (2000)
Formate dehydrogenase-N	1kqf, 1kqg	Jormakka et al. (2002)
Monotopic membrane proteins		
Cyclooxygenase-1	1prh	Picot et al. (1994)

continued

Table 17.1.20 Integral Membrane Proteins of Known Structure, continued

Protein	PDB Entry[a]	Reference
Cyclooxygenase-2	1cx2	Kurumbail et al. (1996)
Squalene-hopene cyclase	2sqc, 3sqc	Wendt et al. (1999)
Fatty acid amine hydrolase	1mt5	Bracey et al. (2002)
Monoamine oxidase B	1gos	Binda et al. (2002)
β-barrel membrane proteins		
Porin	2por	Weiss and Schulz (1992)
OmpF porin	2omf	Cowan et al. (1992)
PhoE phosphoporin	1pho	Cowan et al. (1992)
Maltoporin	1mal	Schirmer et al. (1995)
	2mpr	Meyer et al. (1997)
OmpA	1bxw	Pautsch and Schulz (1998)
	1qjp	Pautsch and Schulz (2000)
	1g90	Arora et al. (2001)
OmpX	1qj8	Vogt and Schulz (1999)
OmpLA	1qd5, 1qd6	Snijder et al. (1999)
TolC	1ek9	Koronakis et al. (2000)
FhuA	1fcp, 2fcp	Ferguson et al. (1998)
	1by3, 1by5	Locher et al. (1998)
FepA	1fep	Buchanan et al. (1999)
FecA	1kmo, 1kmp	Ferguson et al. (2002)
α-hemolysin	7ahl	Song et al. (1996)

[a]See Web-Based Structural Bioinformatics and *http://www.rcsb.org/pdb*.

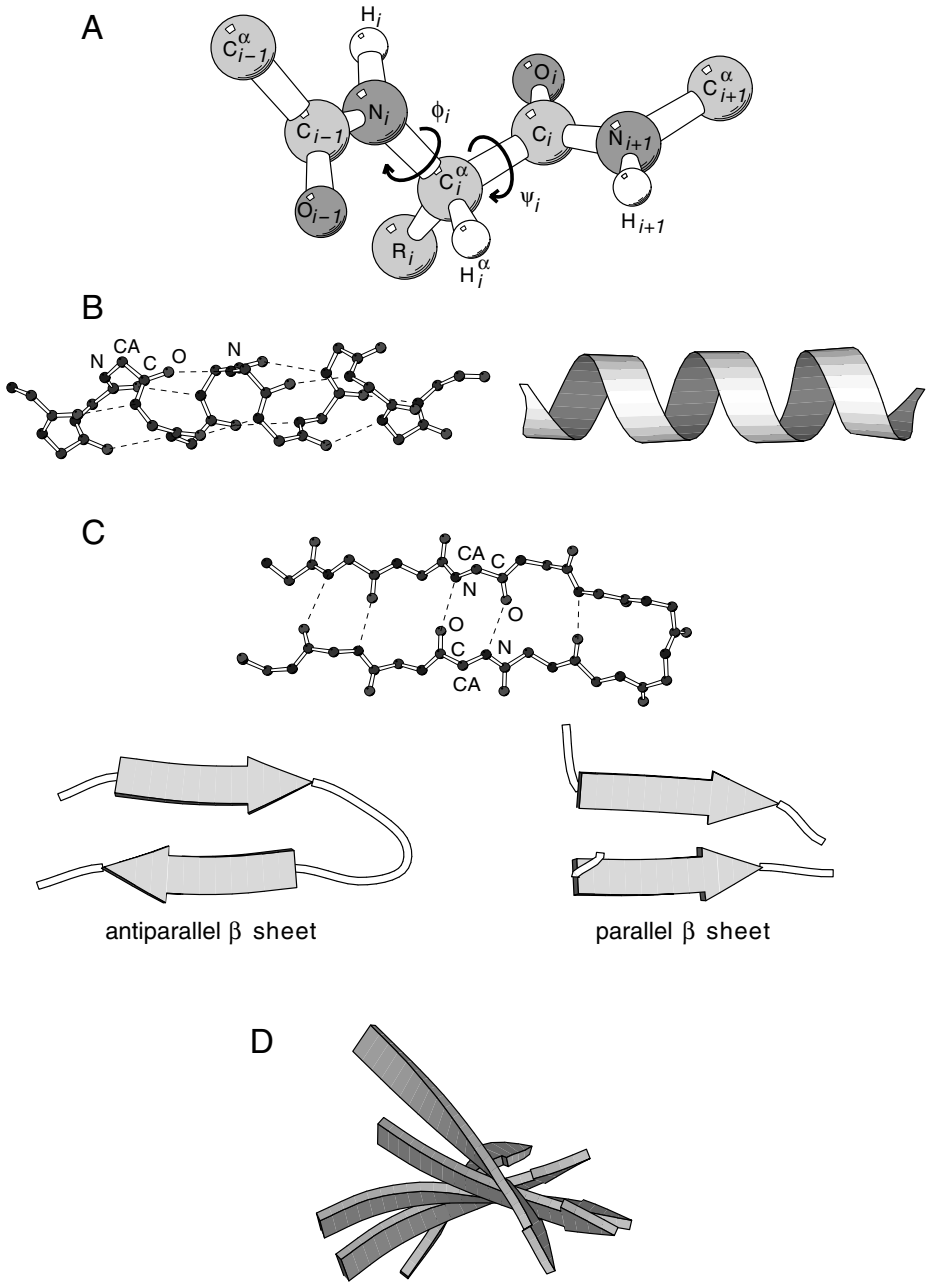


Figure 17.1.1 *(continues on next three pages)* (**A**) Drawing of an L-polypeptide chain using a ball-and-stick model to illustrate torsion angles ϕ and ψ for residue *i*. Torsion angle ϕ defines the angle between the planes specified by atoms C_{i-1}–N_i–C_i^{α} and N_i–C_i^{α}–C_i, respectively. Torsion angle ψ defines the angle between the plane specified by atoms N_i–C_i^{α}–C_i and C_i^{α}–C_i–N_{i+1} respectively. Also shown are both ball-and-stick and ribbon representations of an (**B**) α-helix and (**C**) β-sheet. The latter is shown in both anti- and parallel orientations. (**D**) Illustration of the characteristic right-handed twist of a β-sheet as observed in flavodoxin (PDB entry 1flv). (**E**) Types I and II tight turns. Examples of commonly observed secondary structure assemblies: (**F**) four-helix bundle (top and side view; PDB entry 1bcf), (**G**) β-hairpin structure (PDB entry 1bpi), (**H**) β-sheet with Greek key topology (topology diagram), (**I**) jelly-roll motif (PDB entry 1pgs); (**J**) β-sandwich (PDB entry 4gcr), (**K**) 16-stranded β-barrel (PDB entry 2por), (**L**) α/β-barrel (PDB entry 1btm), and (**M**) seven-bladed β-propeller (PDB entry 1got).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

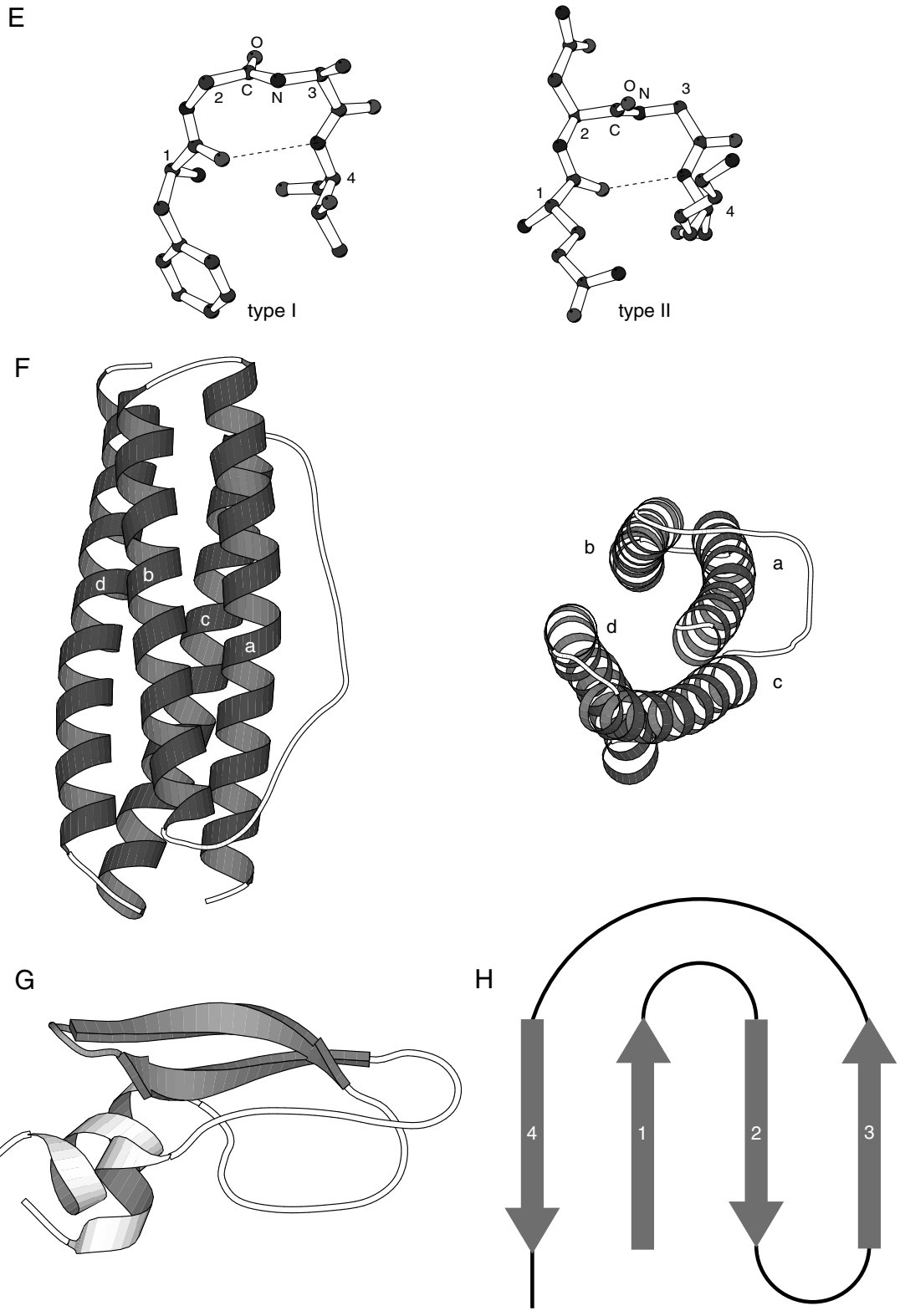


Figure 17.1.1 *(continued)*

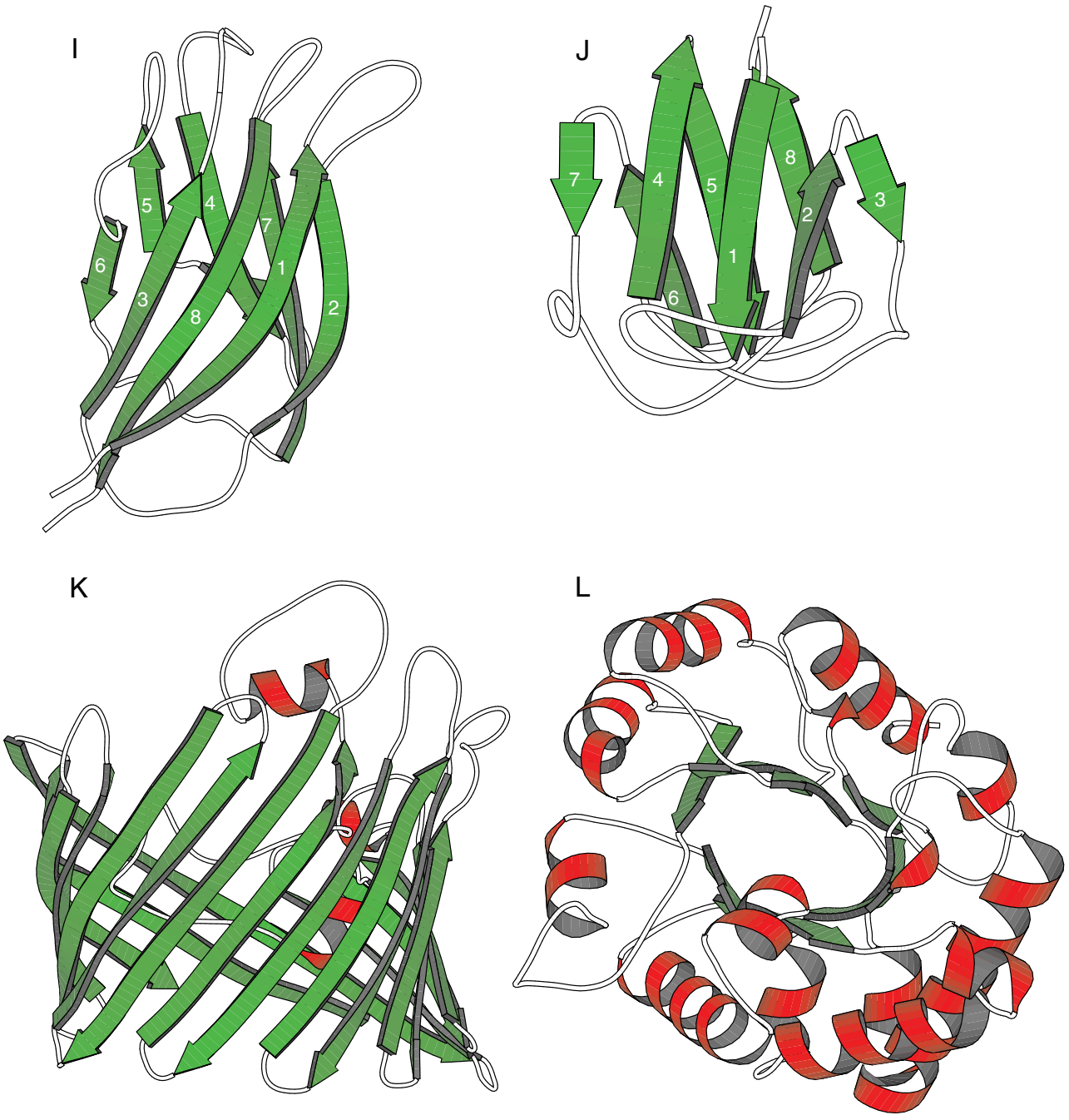


Figure 17.1.1 *(continued)*

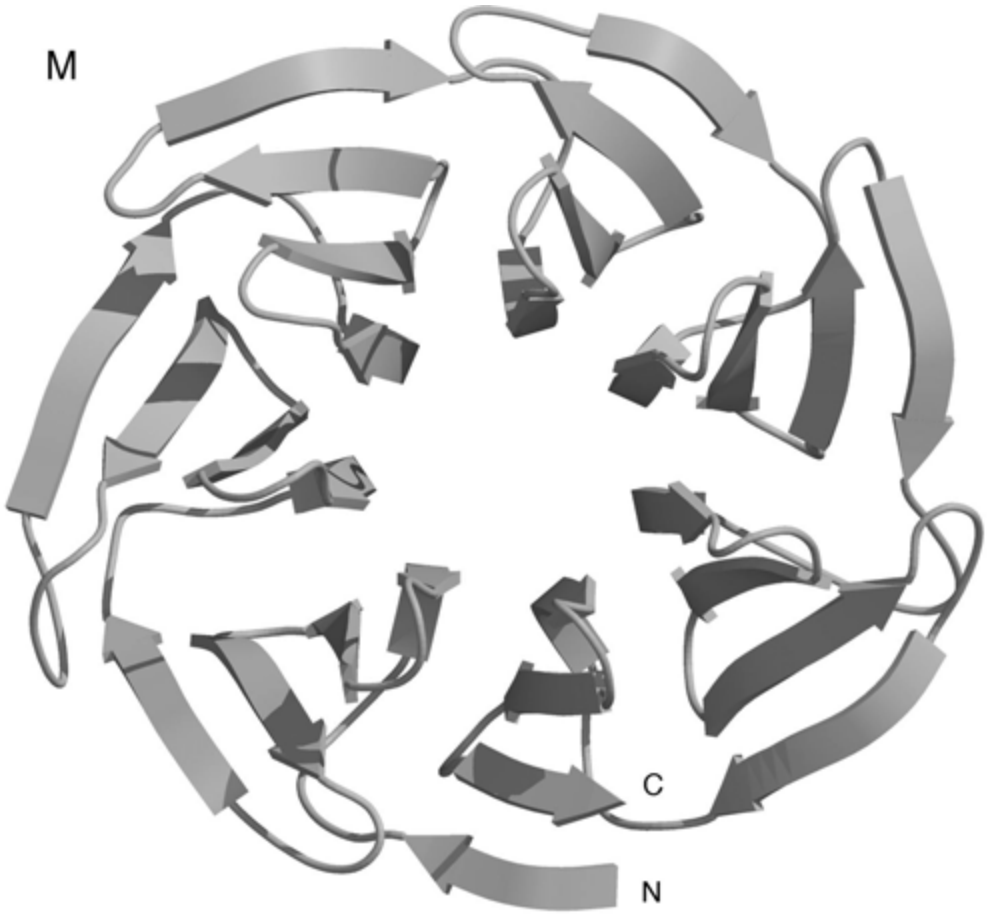


Figure 17.1.1 *(continued)*

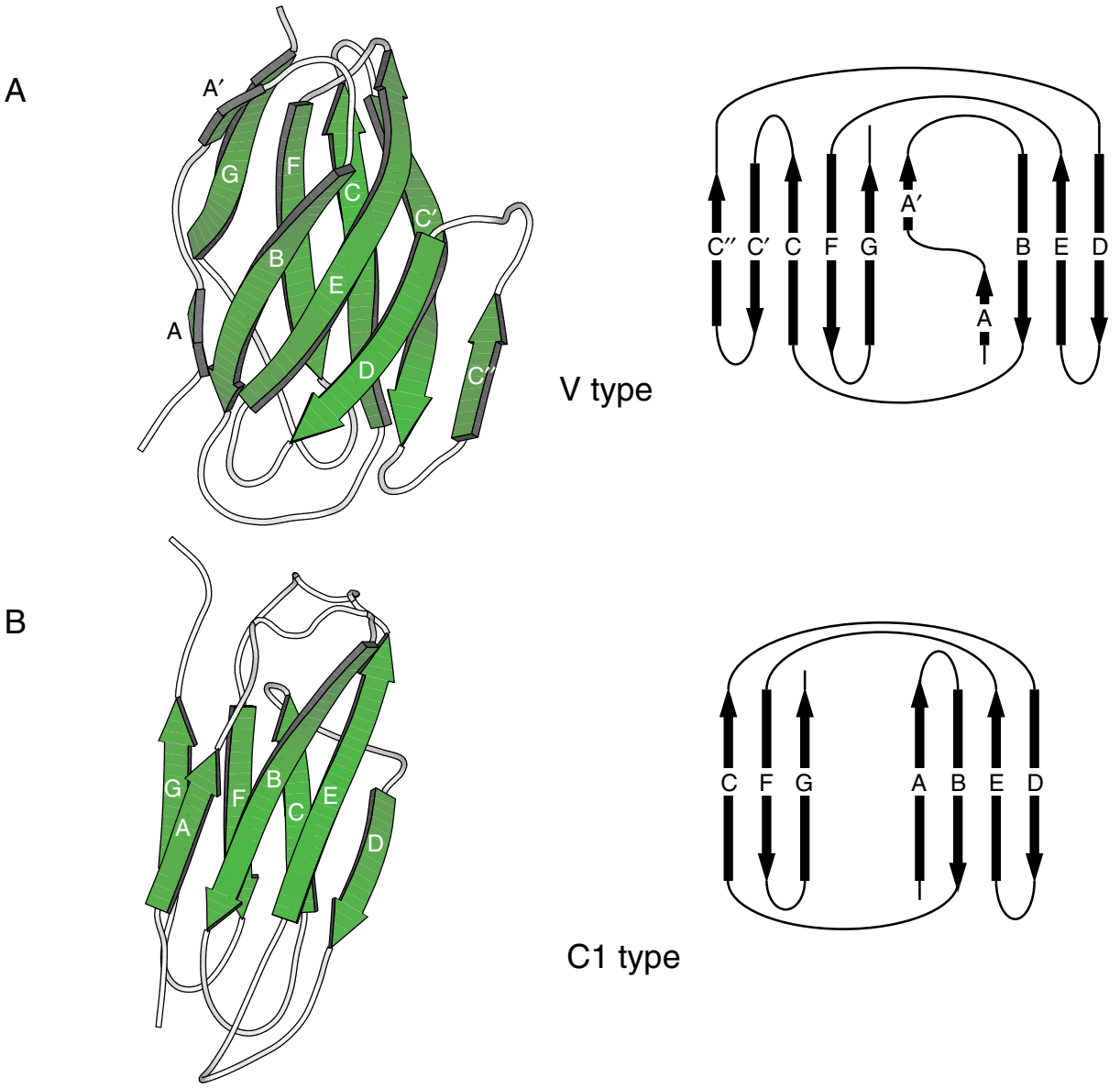


Figure 17.1.2 *(continues on next page)* Tertiary and secondary structures of immunoglobulin fold. The coordinates used for the ribbon diagrams are taken from the PDB entries (**A**) 3hfl (V type), (**B**) 1hnf (C1 type), (**C**) 1fna (C2 type), (**D**) 1 + tlk (I type), and (**E**) 1gof (E type).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

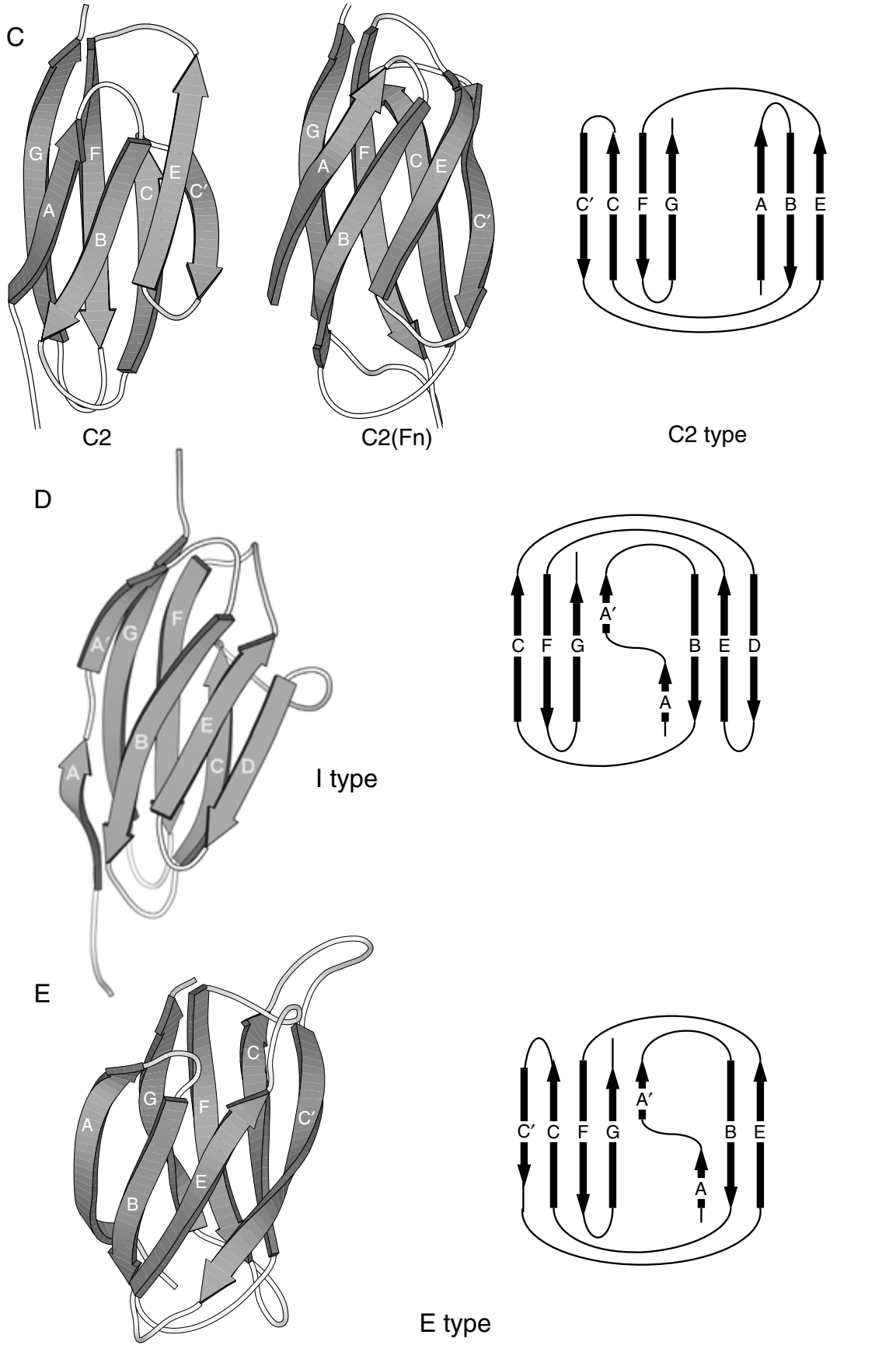


Figure 17.1.2 *(continued)*

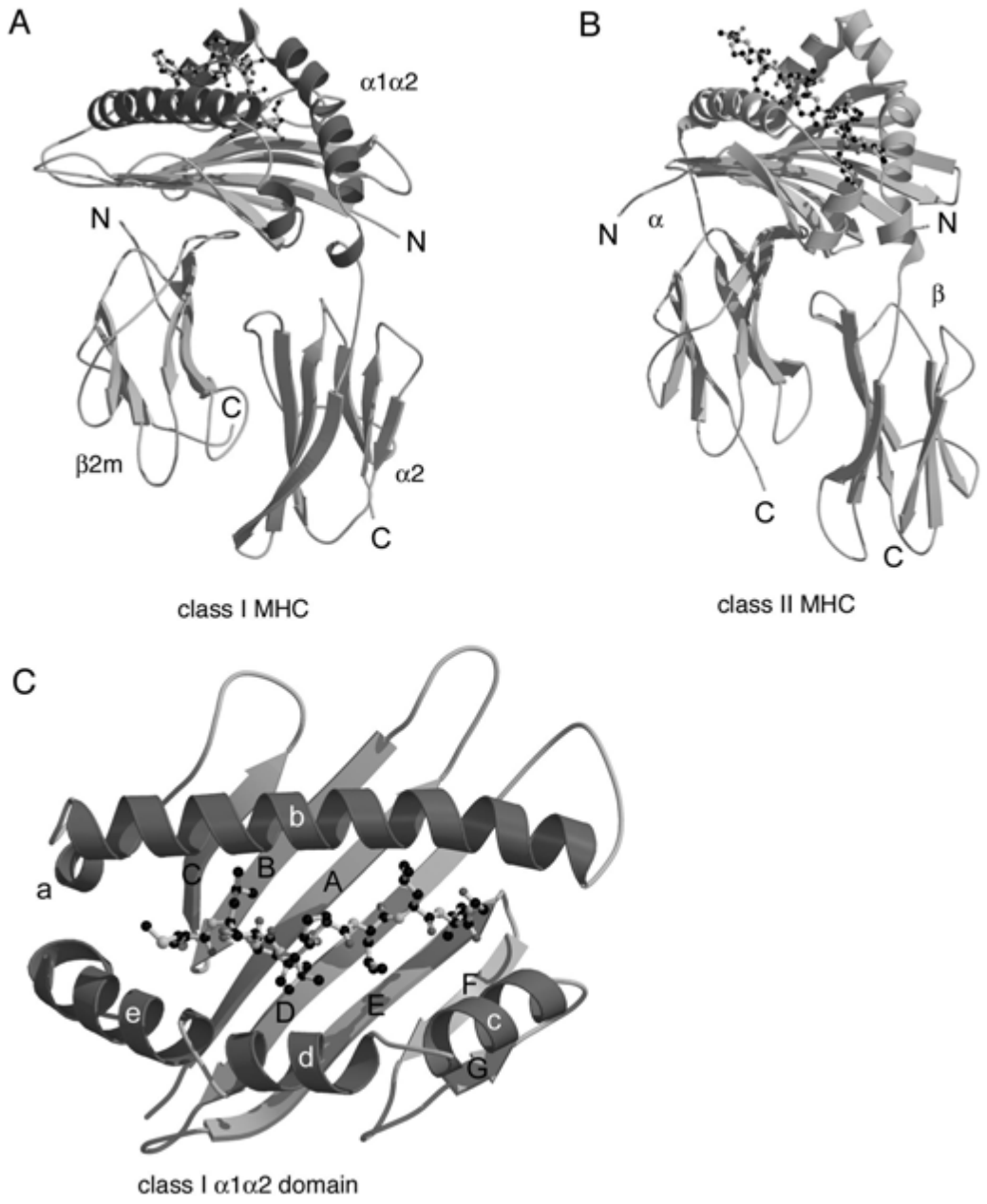


Figure 17.1.3 MHC structures. (**A**) Class I MHC HLA-A2 complex (PDB entry 2clr) and (**B**) class II MHC complex HLA-DR1 (PDB entry 1dlh), each with an antigenic peptide. The α and β chains in (**B**) are colored blue and green respectively. The peptide is shown as a ball-and-stick model. (**C**) Close-up view of the α1α2 peptide-binding domain of HLA-A2 (PDB entry 2clr).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

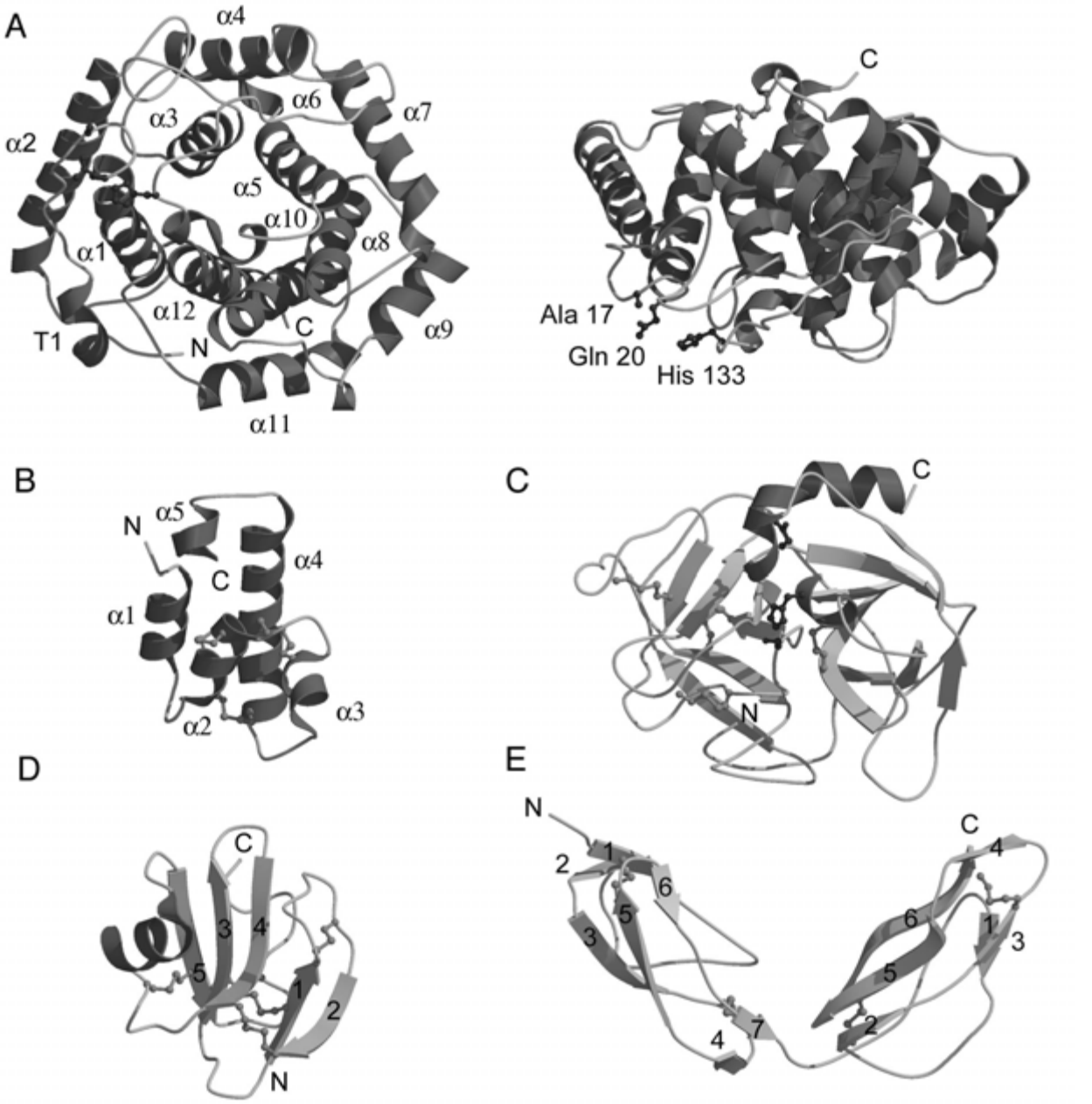


Figure 17.1.4 Protein folds in the complement system. (**A**) C3d (PDB entry 1c3d; residues 996 to 1287). Left side shows the view down the barrel axis; the right side shows the side view of the barrel. The α helices are numbered 1 to 12 and the N-terminal 3_{10} helix is labeled T1. The residues critical for covalent attachment to the pathogen surface, His 133, Gln 20, and Ala 17 (Cys residue in the wild type protein), are represented by a ball-and-stick model. (**B**) C5a (PDB entry 1kjs).The α helices are numbered from 1 to 5 (**C**) Complement factor D serine protease (PDB entry 1dsu). Catalytic residues His 57, Asp 102, and Ser 195 are represented by a ball-and-stick model. (**D**) Complement regulatory protein CD59 (PDB entry 1cdq; residues 1 to 70). The β strands are numbered according to their order in the protein sequence. (**E**) CCP modules 15 and 16 from complement factor H (PDB entry 1hfh). The β strands for each separate CCP module are numbered according to their order in the protein sequence.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

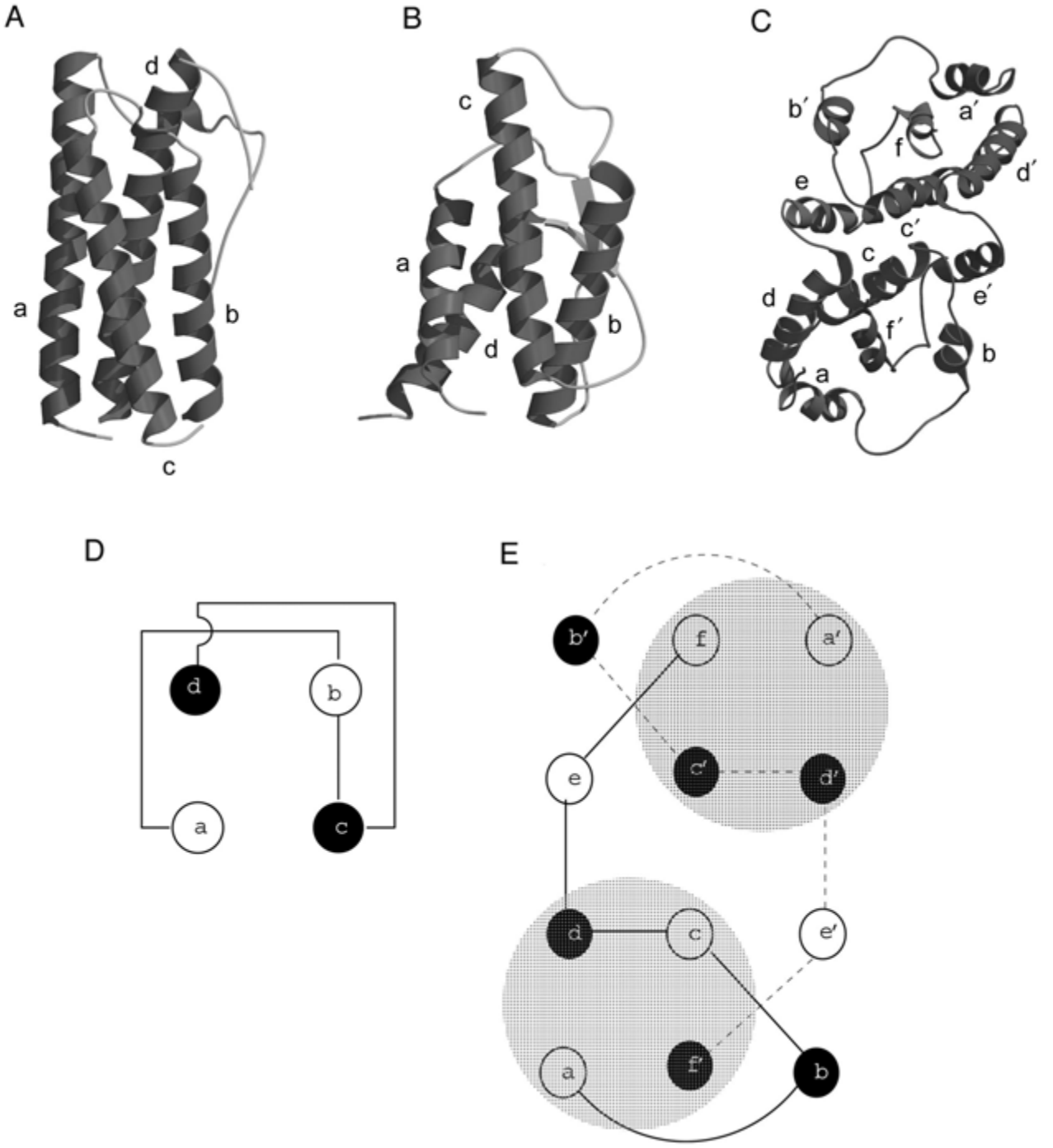


Figure 17.1.5 Long and short-chain cytokines. The structure of (**A**) a long-chain helical cytokine, GCSF (PDB entry 1rhg), (**B**) a short-chain helical cytokine, IL-4 (PDB entry 1rcb), and (**C**) interferon-γ (PDB entry 1rfb). The two monomers of interferon-γ are shown in red and purple. (**D**) Connectivity between helices in four-helix bundle cytokines. The up helices are drawn in white and the down helices are in black. (**E**) Connectivity between helices of IFN-γ. The block shaded regions correspond to the four-helix bundles.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

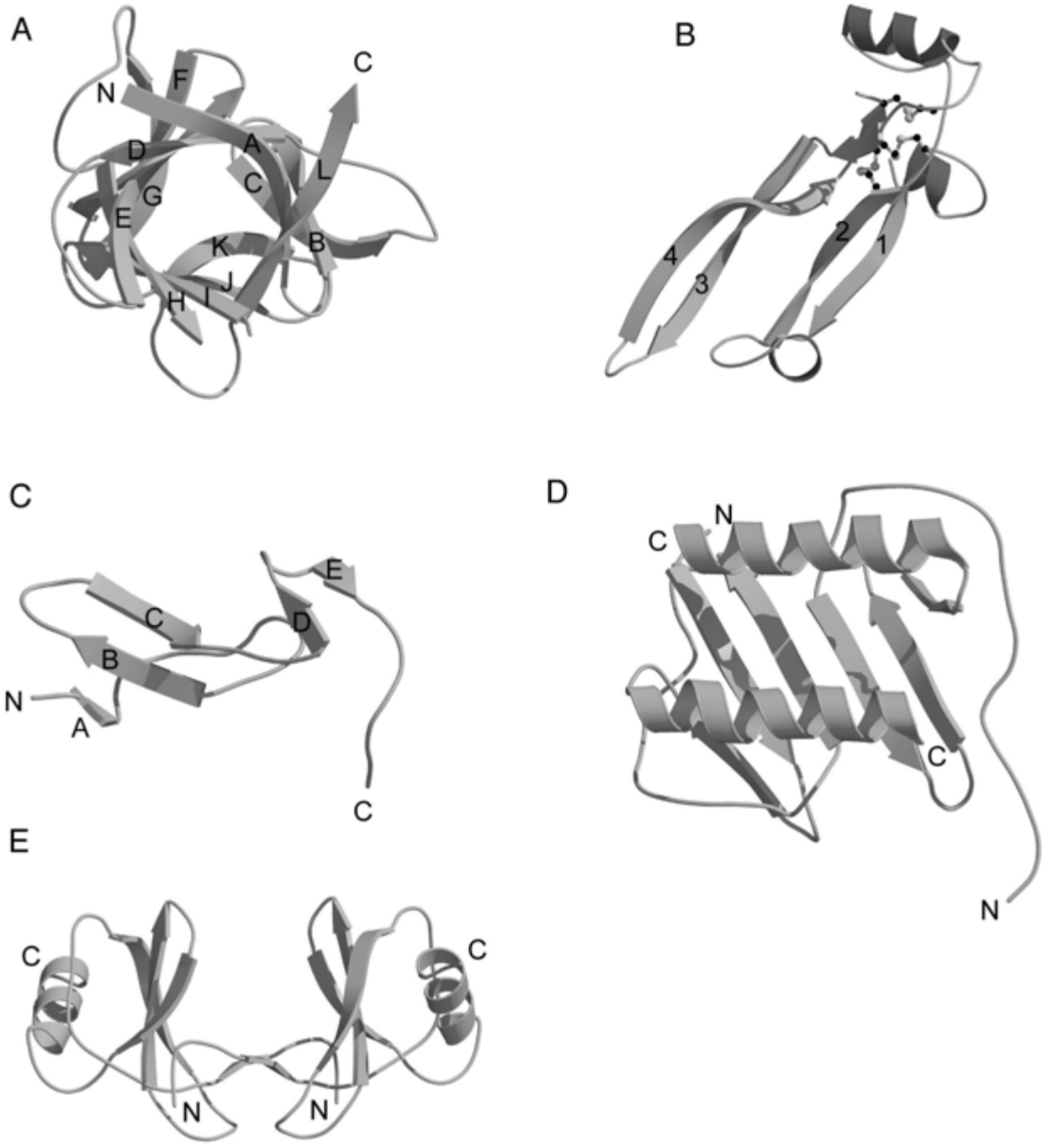


Figure 17.1.6 Cytokines and chemokines. (**A**) Human IL-1β, a member of the β-trefoil fold. (PDB entry 1hib). (**B**) Human transforming growth factor-β2 (PDB entry 2tgi). The four strands that define the cysteine-knot fold are shown in green (Anderson et al., 1978). The six knotted cysteines are shown in ball-and-stick model (with yellow sulfur atoms). (**C**) Murine EGF (PDB entry 1epj). (**D**) CXC chemokine (PDB entry 1il8), (**E**) CC chemokine (PDB entry 1rto).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

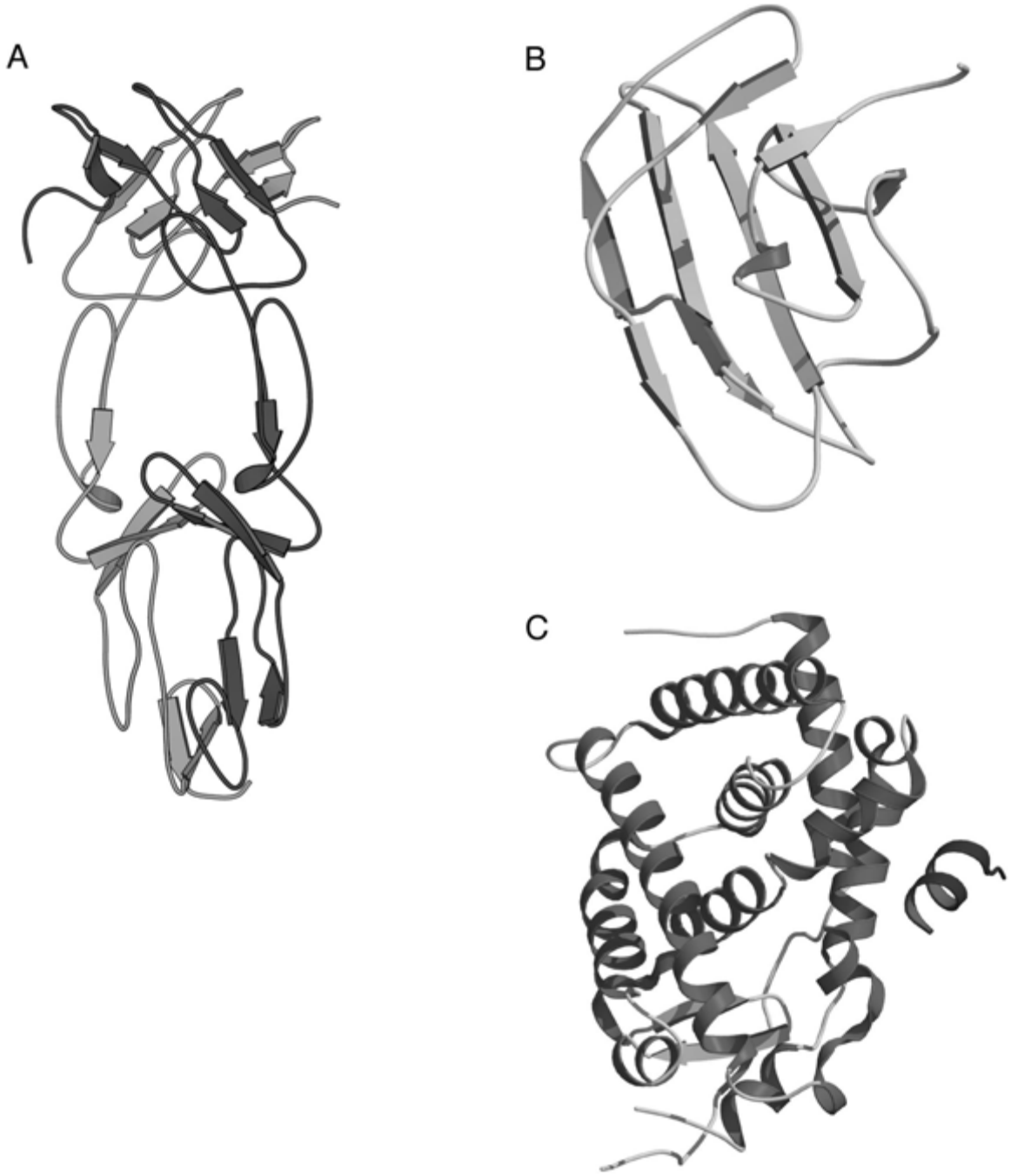


Figure 17.1.7 *(continues on next page)* The structures of *(***A**) tumor necrosis factor receptor TNFR (PDB entry 1tnr), (**B**) type II TGF-β receptor (PDB entry 1ktz), (**C**) thyroid hormone receptor (PDB entry 1bsx), (**D**) integrin I domain (PDB entry 1lfa), (**E**) scavenger receptor (PDB entry 1by2), (**F**) glutamate receptor (PDB entry 1gr2) bound to the neurotoxin kainate (ball-and-stick model), and (**G**) transferrin receptor (PDB entry 1cx8).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

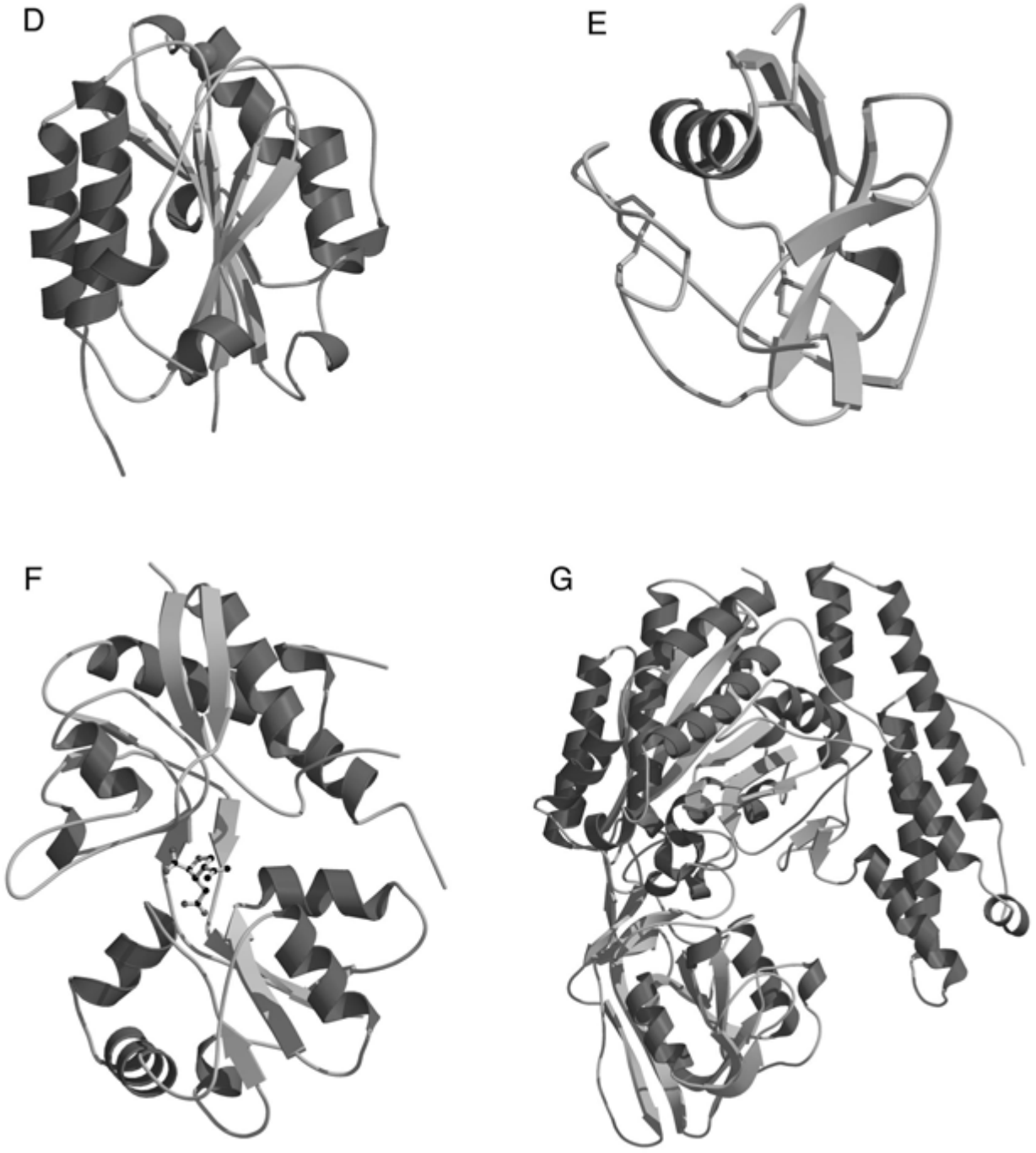


Figure 17.1.7 *(continued)*

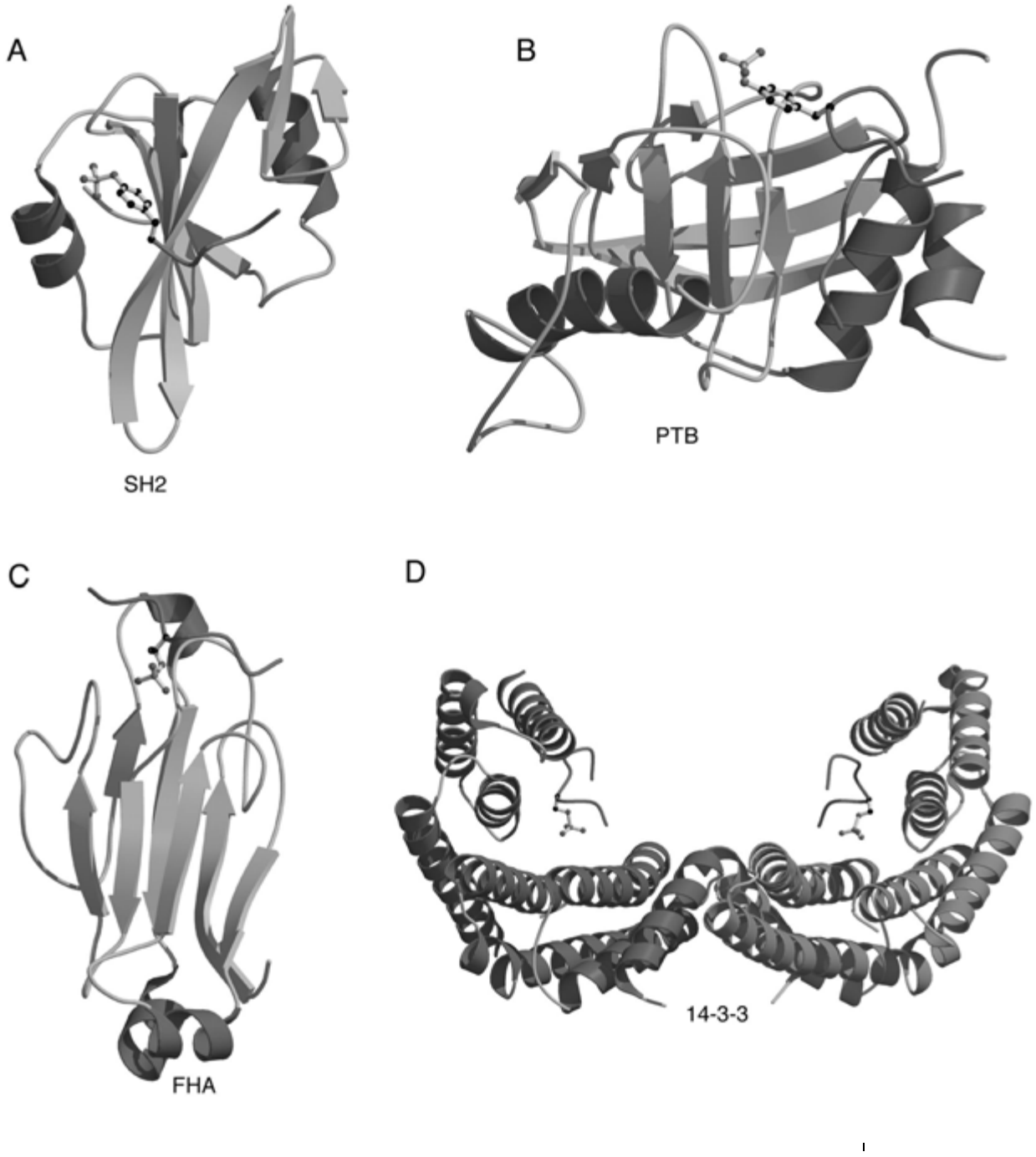


Figure 17.1.8 Folds that bind phosphopeptides. The phosphopeptides are shown as magenta colored worms. The phosphorylated amino acid is represented by a ball-and-stick model. (**A**) The SH2 domain from v-src tyrosine kinase bound to a five-residue phosphotyrosine peptide (PDB entry 1sha). (**B**) PTB domain from shc complexed with a twelve-residue phosphotyrosine peptide (PDB entry 1shc). (**C**) FHA domain from protein kinase RAD53 complexed to a twelve-residue phosphothreonine peptide (PDB entry 1g6g). (**D**) Homodimer of 14-3-3 protein ζ bound to eight-residue phosphoserine peptides (PDB entry 1qja).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

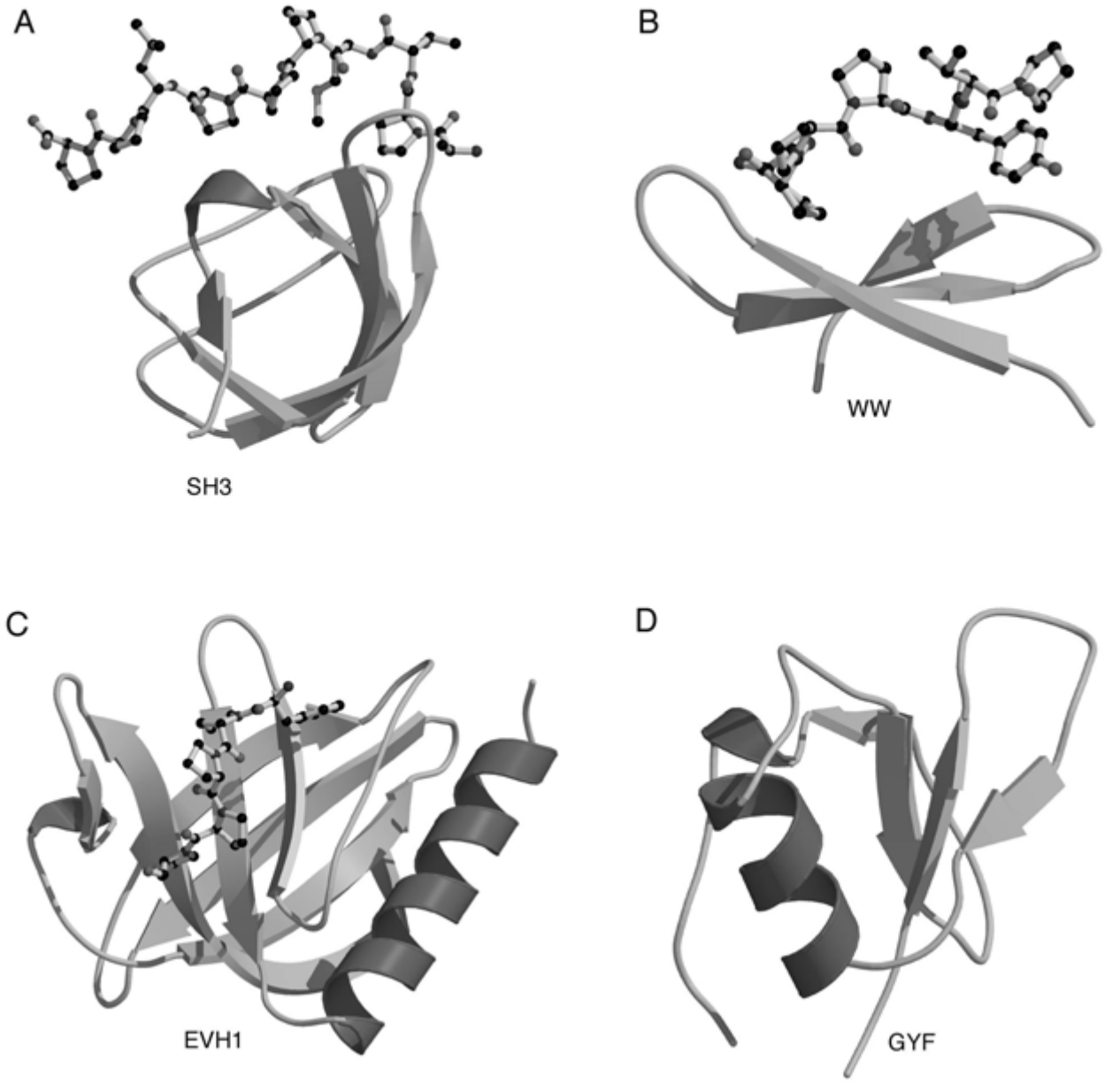


Figure 17.1.9 Folds that bind to polyproline peptides. Bound polyproline peptides are represented by ball-and-stick models. (**A**) An SH3 domain from the Abl tyrosine kinase complexed with the ten-residue synthetic peptide 3Bp-1 (PDB entry 1abo). (**B**) A WW domain from dystrophin in complex with a seven-residue β-dystroglycan peptide (PDB entry 1eg4). (**C**) An EVH1 domain from Enabled, bound to the Acta peptide (PDB entry 1evh). (**D**) GYF domain from CD2Bp2 (PDB entry 1gyf). *(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)*

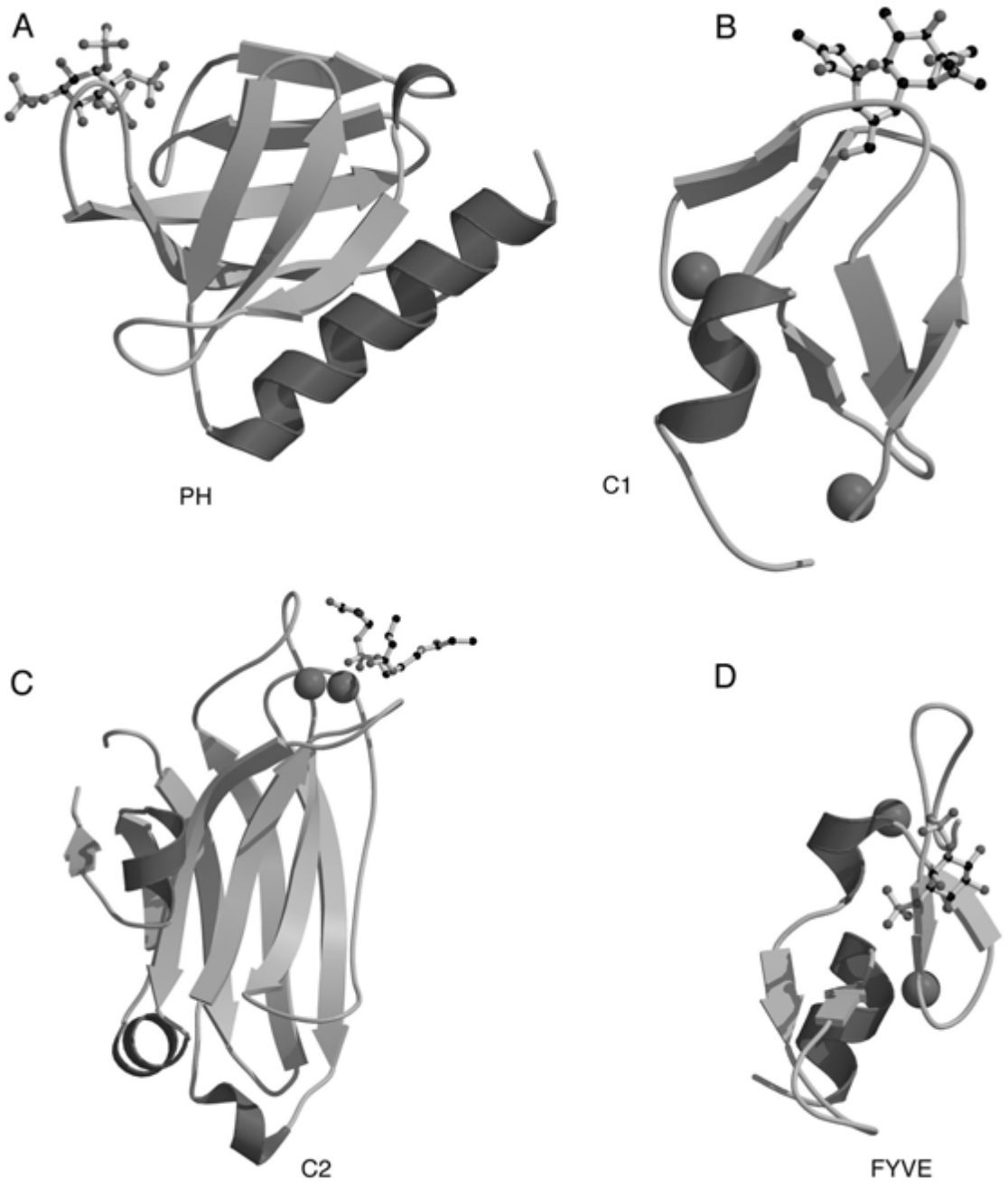


Figure 17.1.10 Phospholipid-binding domains. Lipid ligands are displayed as ball-and-stick models and metal cations are represented by magenta spheres. (**A**) Pleckstrin homology (PH) domain from Dappl/Phish complexed with inositol 1,3,4,5-tetrakisphosphate (PDB entry 1fao). (**B**) C1 domain from protein kinase Cδ complexed with phorbol-13-acetate (PDB entry 1ptr). (**C**) C2 domain from protein kinase C(α) complexed with Ca^{2+} and phosphatidylserine (PDB entry 1dsy). (**D**) A FYVE domain from EEA1 bound to inositol 1,3-diphosphate (PDB entry 1joc).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

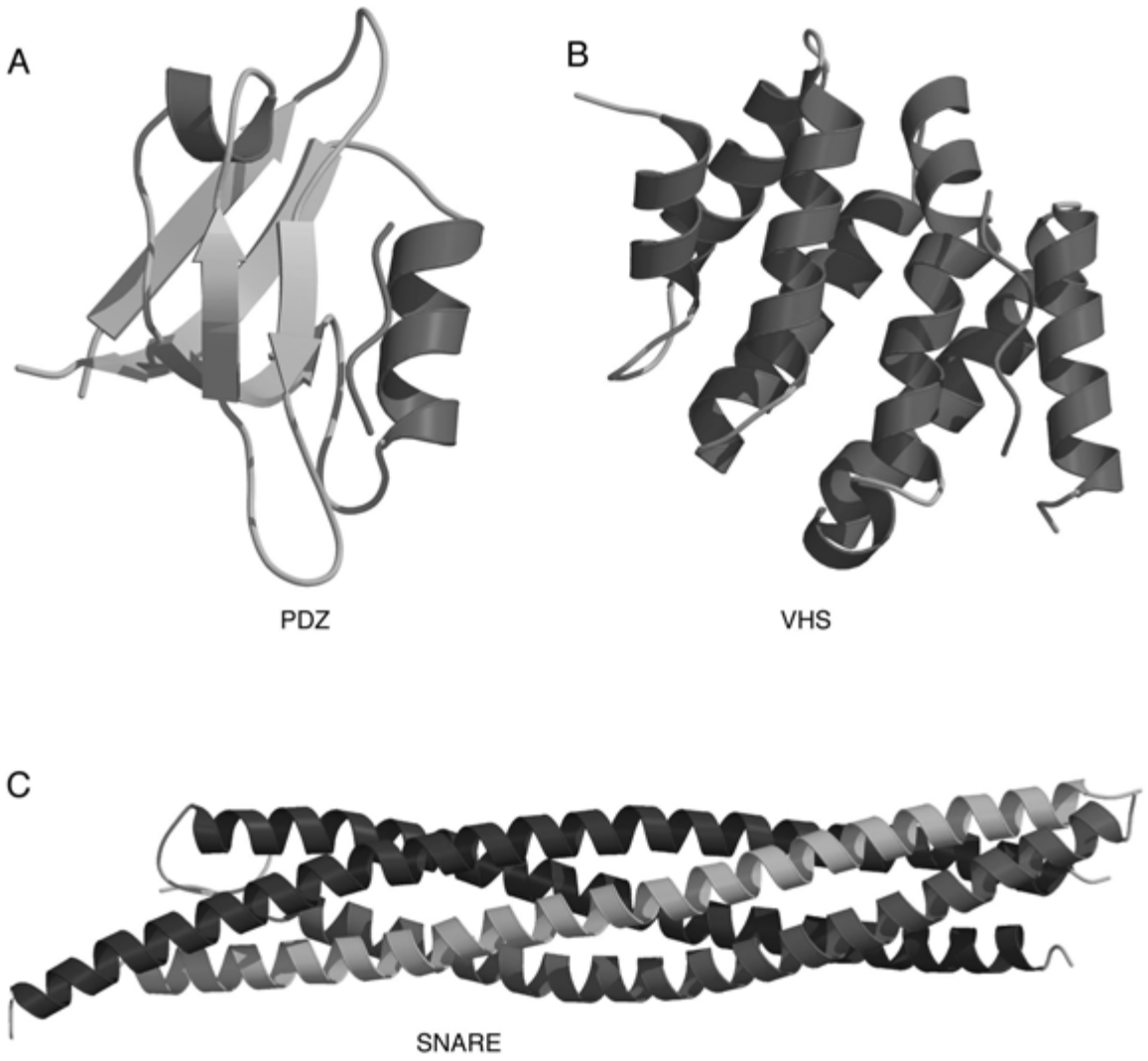


Figure 17.1.11 *(continues on next page)* Protein interaction domains. Bound peptides are represented by magenta worms. (**A**) The syntrophin PDZ domain bound to the peptide GVKESLV (PDB entry 2pdz). (**B**) VHS domain of GGA1 complexed with cation-independent mannose-6-phosphate receptor C-terminal peptide (PDB entry 1jwg). (**C**) SNARE fusion complex containing syntaxin-1A (green), synaptobrevin-II (red), and SNAP-25B (blue; PDB entry 1sfc). (**D**) SAM domain from human EPHB2 receptor (PDB entry 1b4f). (**E**) MH2 domain from human Smad2 (PDB entry 1khx). (**F**) Homotrimer of a phosphorylated MH2 domain from human Smad2 (PDB entry 1khx). Phosphorylated Ser465 and -467 are represented by ball-and-stick models. The phosphoserine binding loop L3 is shown in magenta. (**G**) Complex of the Smad2 MH2 domain (cyan) with the SARA complex (magenta; PDB entry 1dev).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

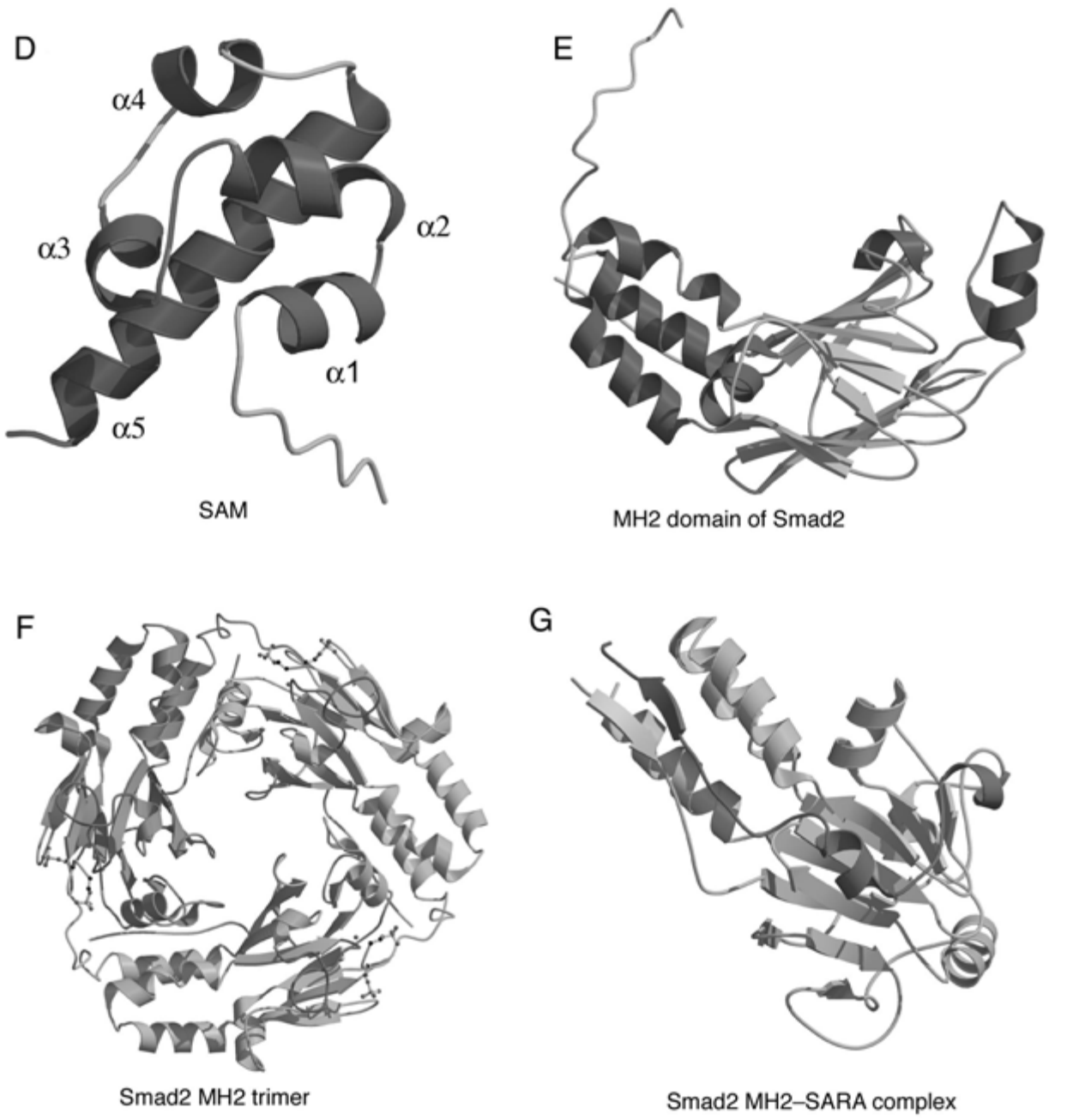


Figure 17.1.11 *(continued)*

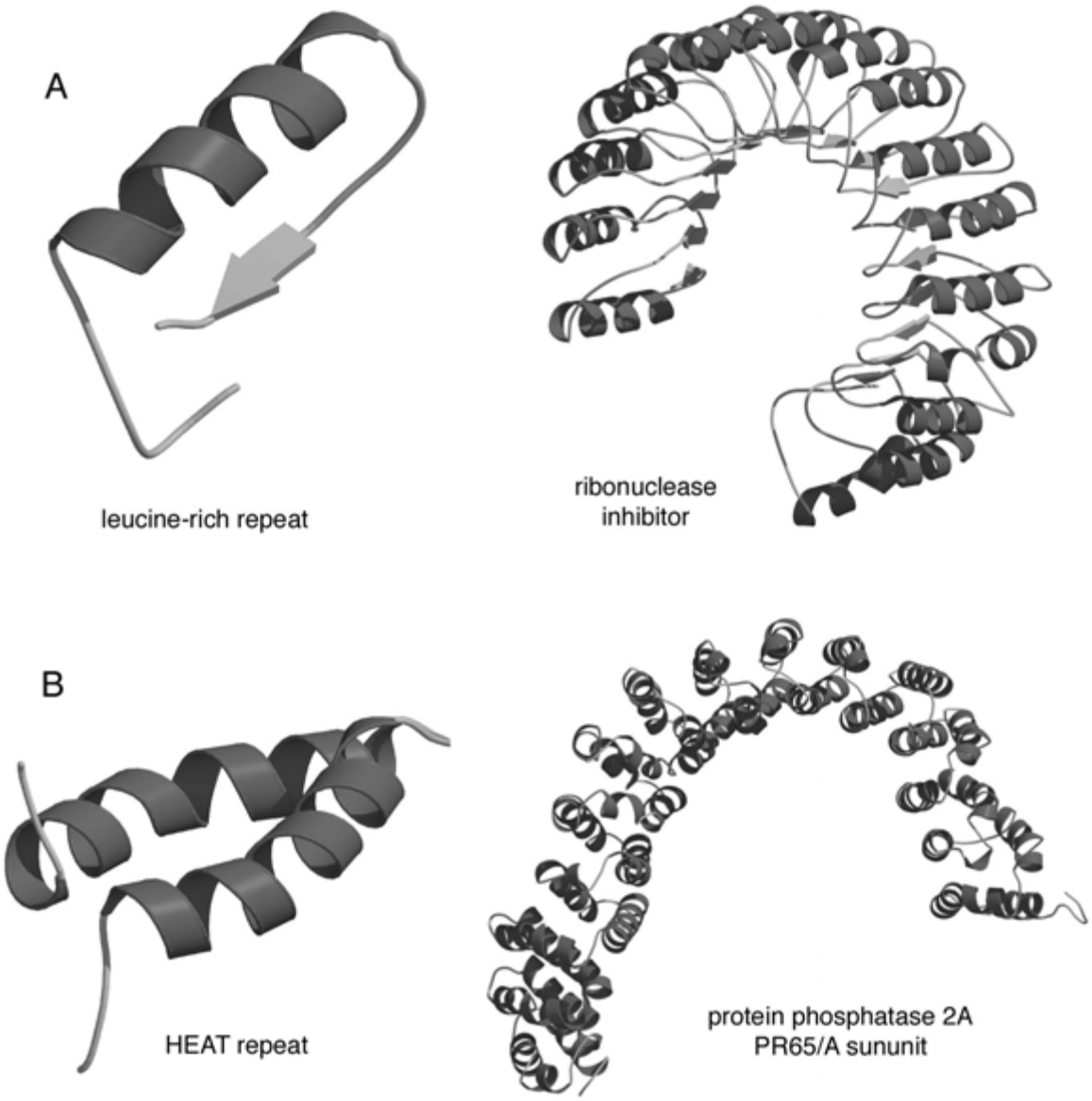


Figure 17.1.12 *(continues on next page)* Structural repeat motifs. *(***A**) A single Leu-rich repeat (left) from ribonuclease inhibitor (right; PDB entry 2bnh). (**B**) A HEAT repeat (left) from the PR65/A subunit of protein phosphatase 2A (right; PDB entry 1b3u). (**C**) An ARM repeat (left) from β-catenin (right; PDB entry 2bct). (**D**) An ankyrin repeat (left) from GABPβ (right; PDB entry 1awc).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

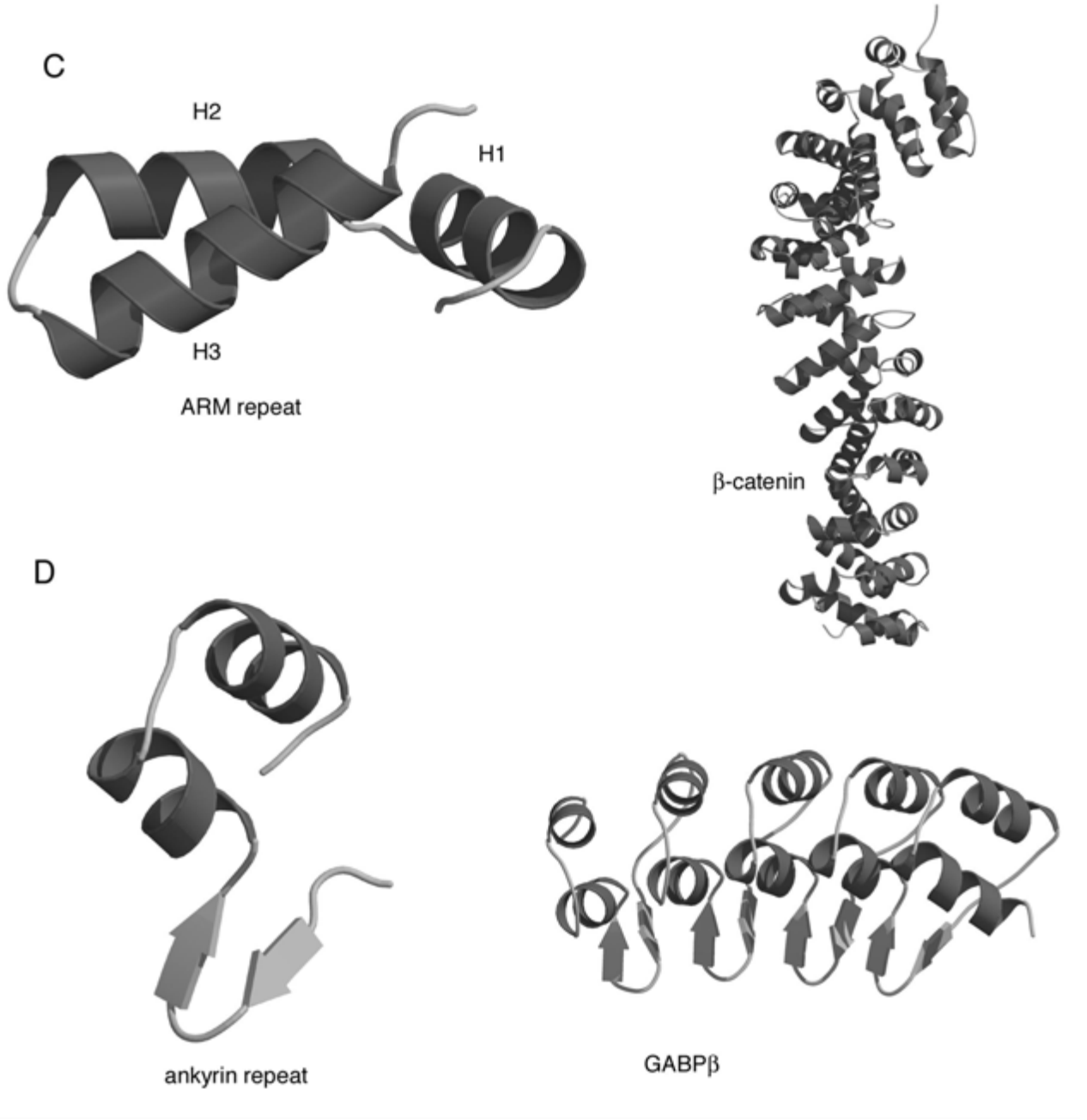


Figure 17.1.12 *(continued)*

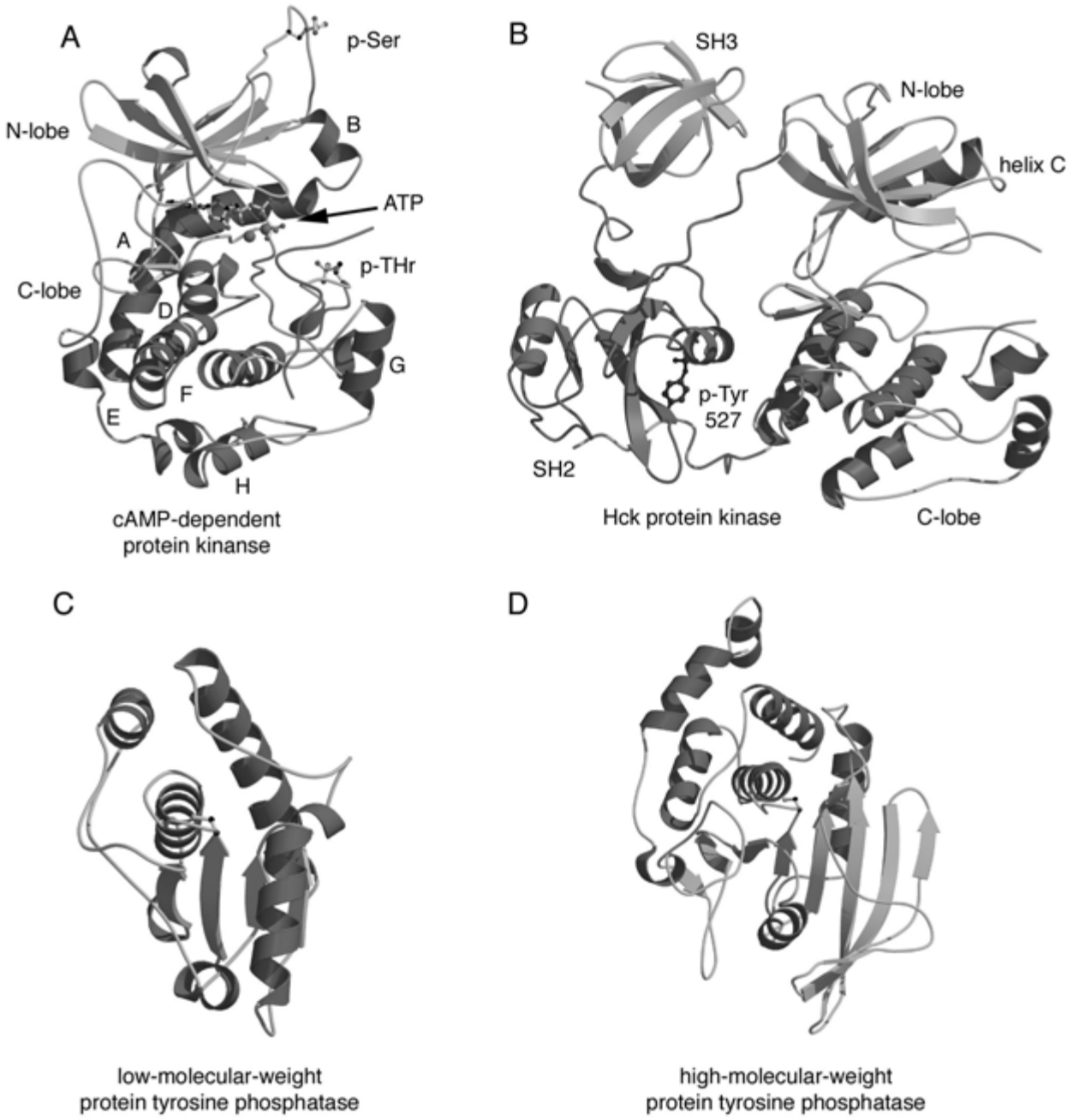


Figure 17.1.13 *(continues on next page)* Protein kinase and phosphatase structures. *(***A**) The structure of phosphorylated cyclic AMP-dependent protein kinase complexed with ATP, Mn, and inhibitor (PDB entry 1atp). The peptide inhibitor is represented by a magenta worm. The ATP, Mn, and phosphorylated serine and threonine are shown as ball-and-stick models. (**B**) The structure of the inactive form of hematopoietic cell kinase of the Src family of protein kinases (PDB entry 1ad5). The SH3 and SH2 domains are colored blue-green and magenta, respectively, and the catalytic domain is colored green and red. The phosphorylated tyrosine 572 is represented by a blue ball-and-stick model. The following protein phosphatases are all shown in approximately the same orientation with the catalytic cysteine displayed as a ball-and-stick model. (**C**) Low-molecular-weight protein tyrosine phosphatase (PDB entry 1phr). (**D**) High-molecular-weight protein tyrosine phosphatase-1B (PDB entry 2hnq). (**E**) Dual-specificity protein phosphatase CDC25A (PDB entry 1c25). (**F**) Protein serine/threonine phosphatase-1 of the PPP family (PDB entry 1fjm).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

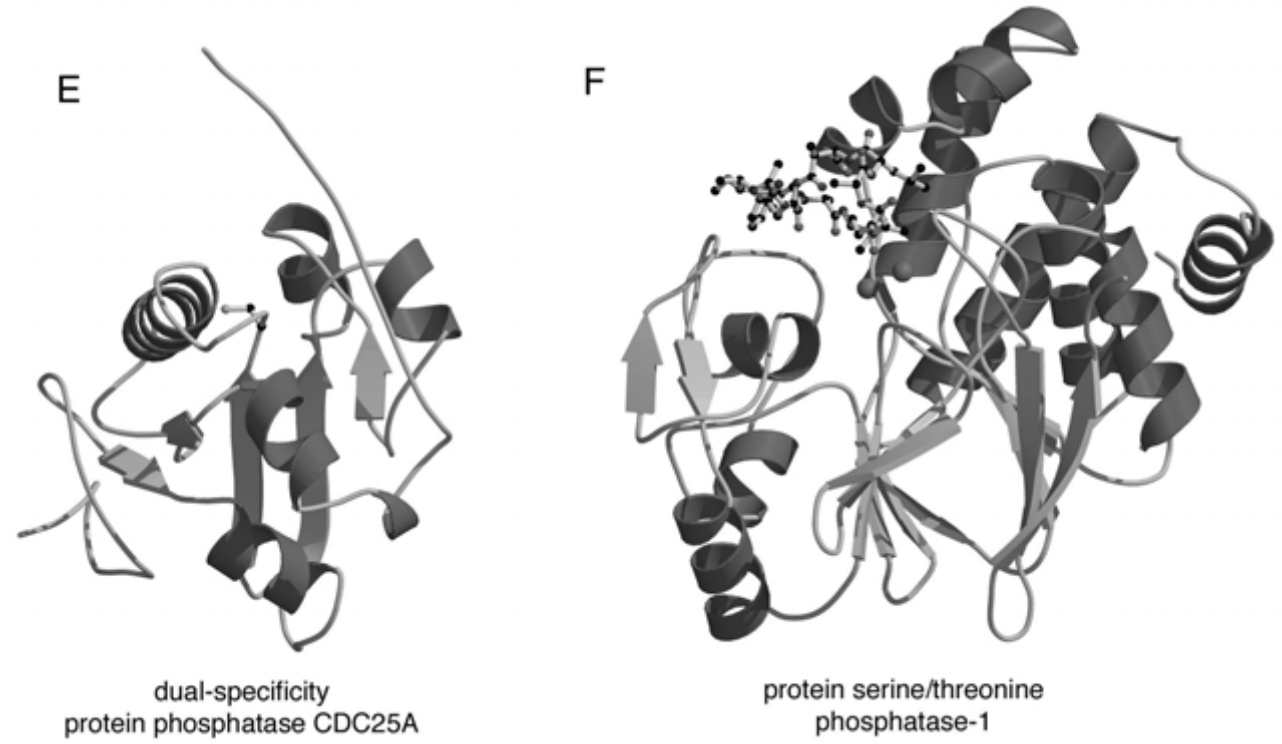


Figure 17.1.13 *(continued)*

Figure 17.1.14 Phosphatase, kinase, and related structures. (**A**) Alkaline phosphatase (PDB entry 1ali). (**B**) Fructose-1,6-bisphosphatase (PDB entry 5fbp) as a representative of the sugar phosphatase fold. (**C**) Cyclin A (PDB entry 1fin). The N-terminal cyclin box domain is colored magenta.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

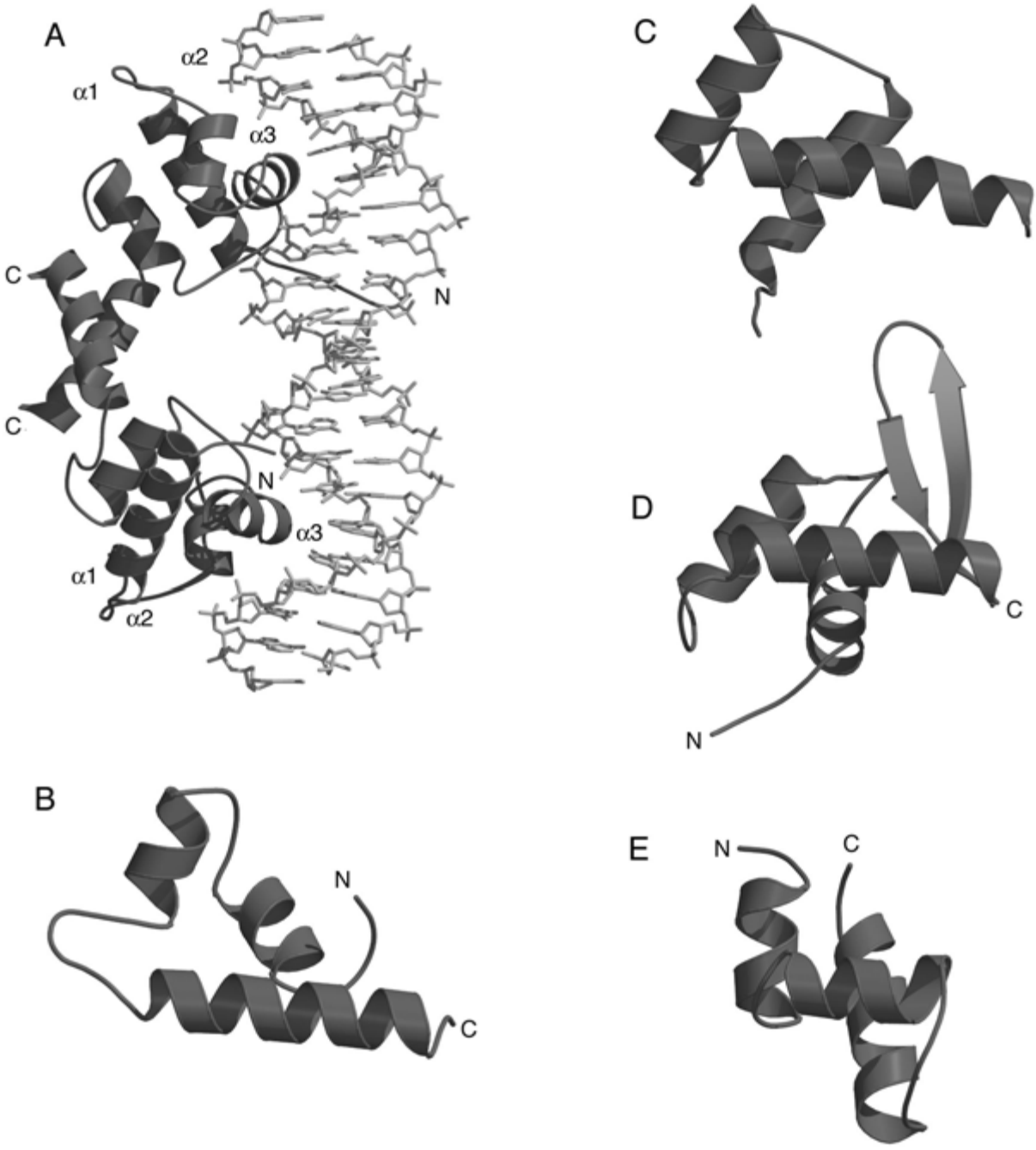


Figure 17.1.15 Variations of the helix-turn-helix (HTH) structural motif and the MADS box. The recognition helix is oriented horizontally in all but panel A. (**A**) Homodimer of the DNA-binding domain of λ repressor bound to DNA (PDB entry 1lmb). The three helices of the HTH motif are labeled α1, α2, and α3 in each monomer. Both recognition helices (α3) sit in the major groove. (**B**) λ repressor (prokaryotic HTH; PDB entry 1lmb). (**C**) Engrailed homeodomain (PDB entry 1hdd). (**D**) Globular domain of histone H5 (PDB entry 1hst). This example of the winged helix motif has only one wing (the β hairpin). (**E**) Purine repressor (purR; PDB entry 1pru).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

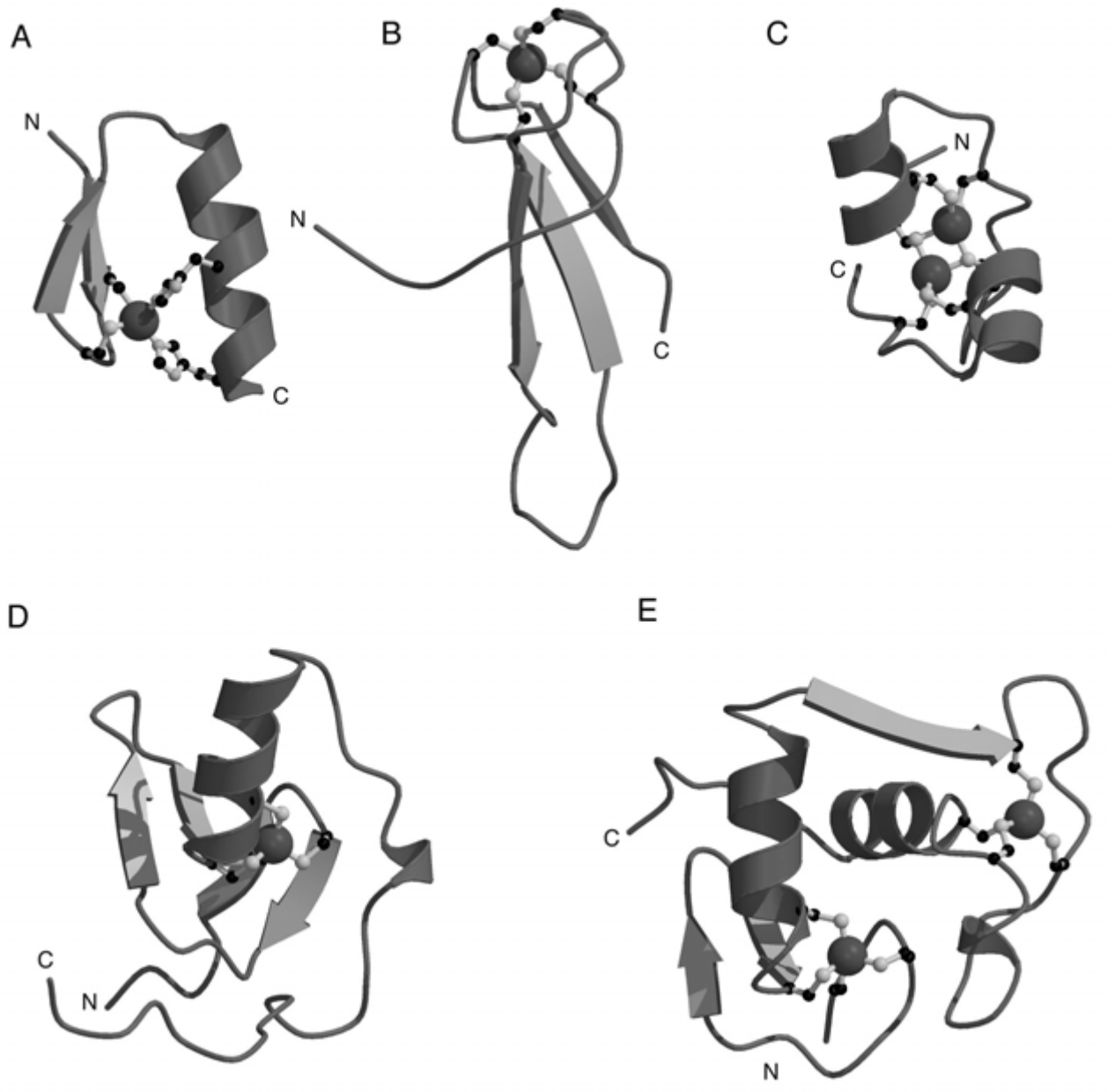


Figure 17.1.16 Zinc-binding motifs within DNA-binding domains. The zinc atom and side-chain ligands to the zinc are represented by ball-and-stick models. (**A**) Second zinc finger from ZIF268 mouse intermediate protein (PDB entry 1zaa). (**B**) Elongation factor TFIIS (PDB entry 1tfi) (**C**) Zn_2Cys_6 binuclear cluster from GAL4 (PDB entry 1d66). (**D**) GATA-1 chicken erythroid transcription factor (PDB entry 1gat). (**E**) DNA-binding module from the glucocorticoid receptor (PDB entry 1gdc). *(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)*

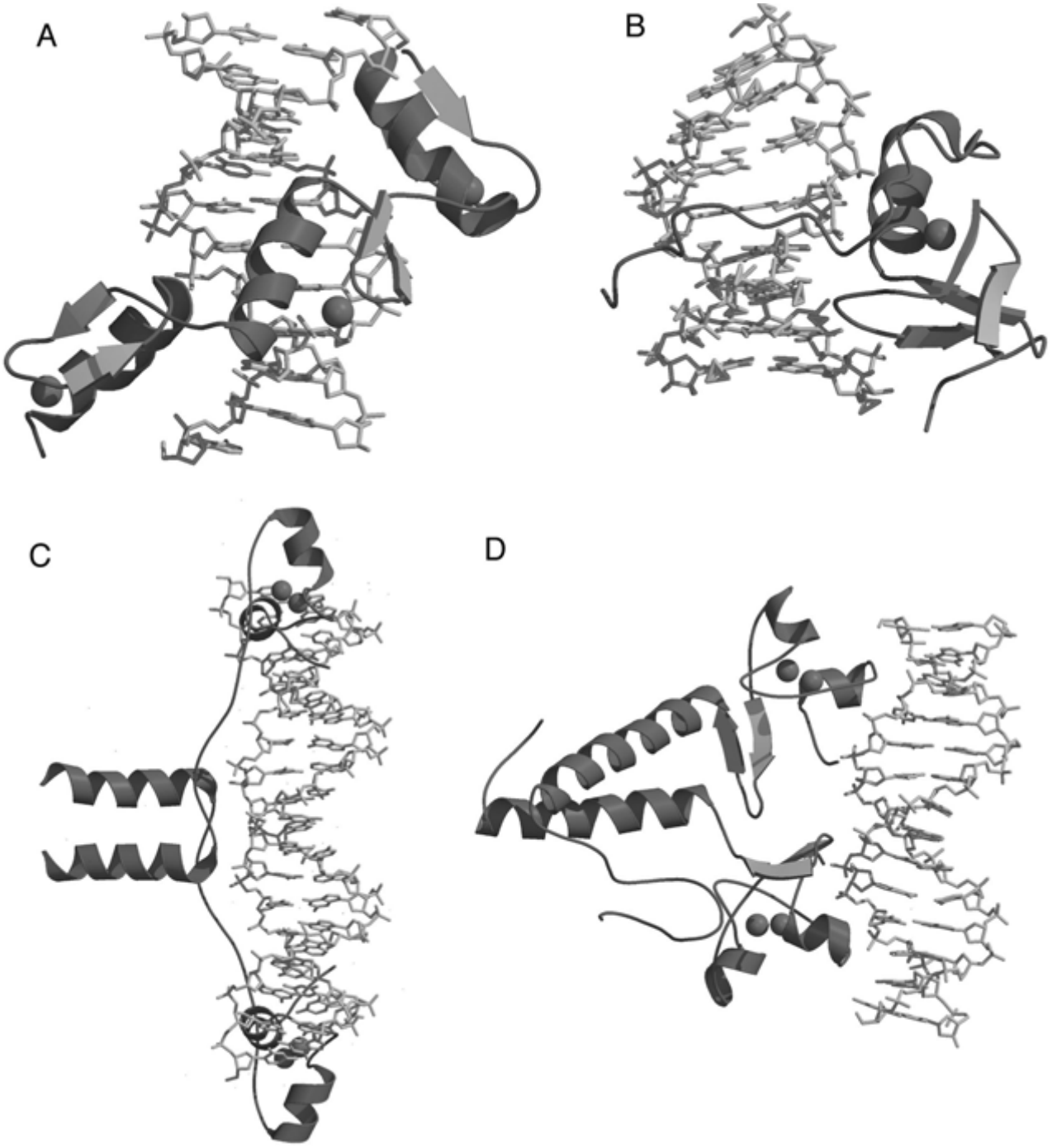


Figure 17.1.17 *(continues on next page)* Binding of zinc-containing modules to DNA. The zinc atoms are represented by spheres. (**A**) The three zinc fingers of ZIF268 (PDB entry 1zaa). (**B**) GATA-1 transcription factor (PDB entry 1gat). Zn_2Cys_6 binuclear clusters (**C**) Gal4 (PDB entry 1d66) and (**D**) pyrimidine pathway regulator 1 (PPR1) DNA-binding fragment (PDB entry 1pyi) bound to DNA. (**E**) Glucocorticoid receptor complexed with DNA (PDB entry 1glu).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

Figure 17.1.17 *(continued)*

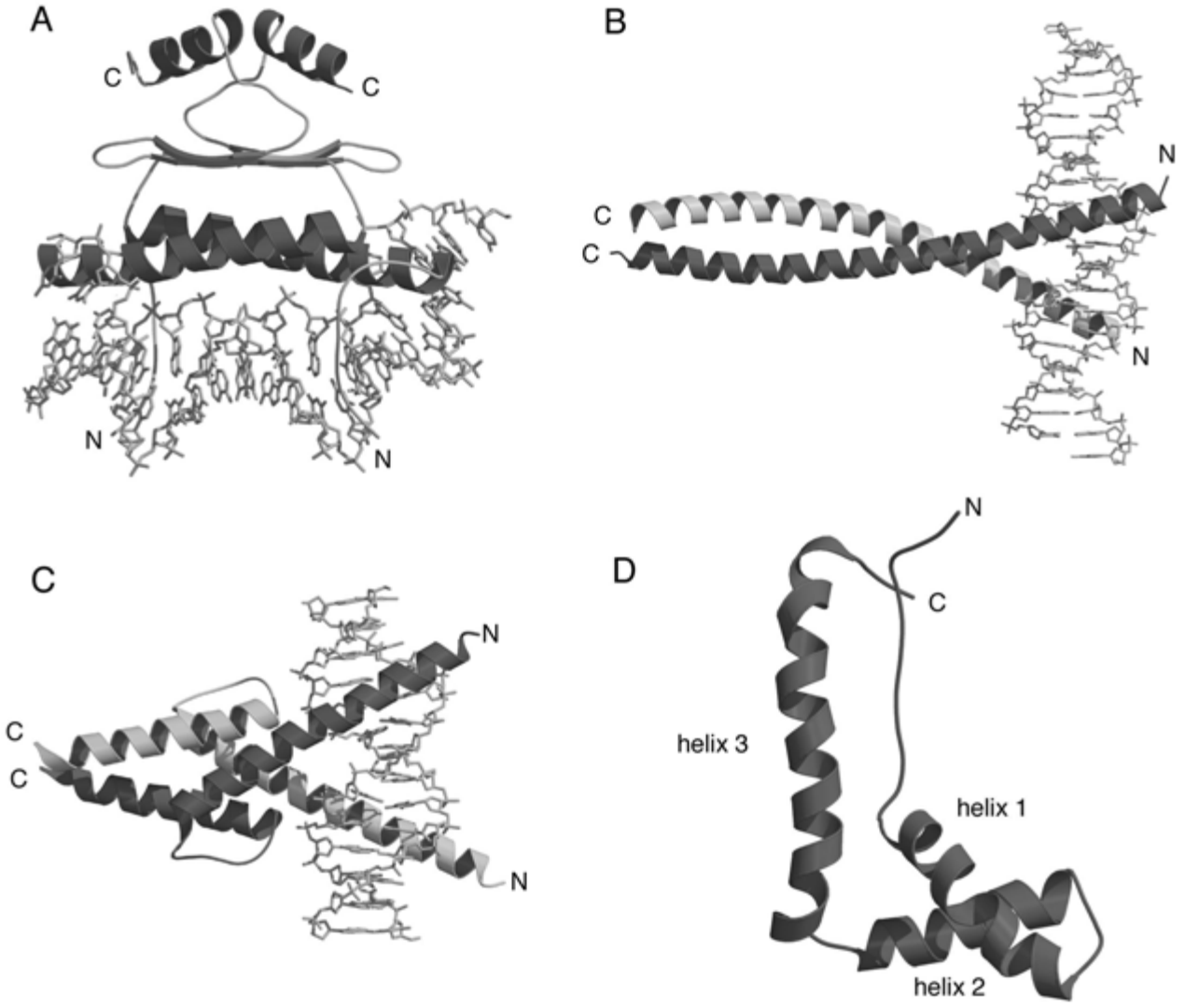


Figure 17.1.18 *(continues on next page)* Helical DNA binding domains (**A**) MADS box of serum response factor bound to DNA (PDB entry 1srs). (**B**) Basic-region leucine-zipper (bZIP) c-Fos/c-Jun heterodimer complexed with DNA (PDB entry 1fos). Monomers within the dimer are shaded differently. (**C**) Basic helix-loop-helix (bHLH) MyoD homodimeric transcription activator complexed with DNA (PDB entry 1mdy). Monomers within the dimer are colored differently. (**D**) High-mobility group (HMG) fragment B from rat (PDB entry 1hme). (**E**) Structure of a heterodimer of dTAF42 and dTAF62 (PDB entry 1taf). Each monomer contains the histone fold.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

Figure 17.1.18 *(continued)*

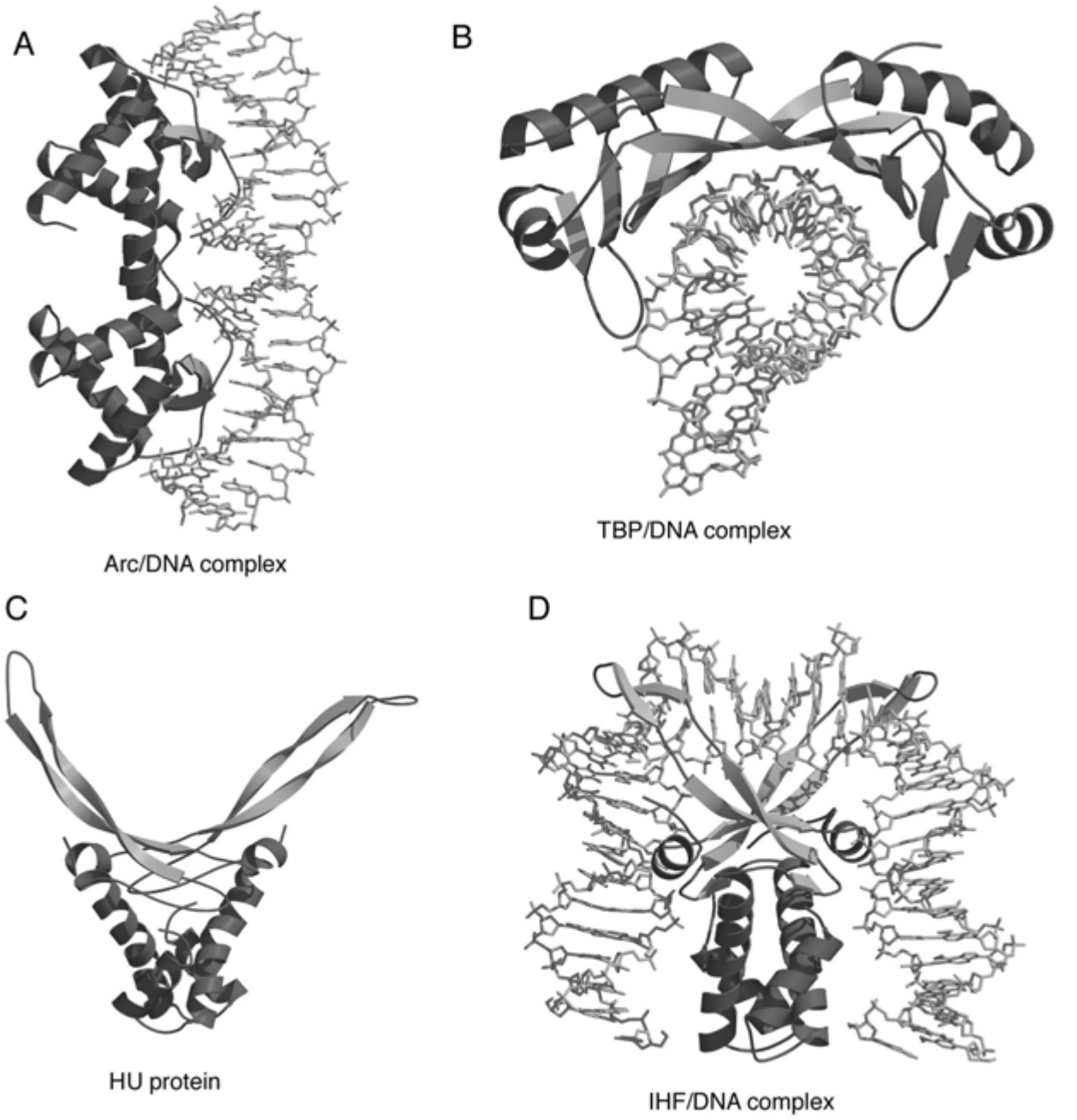


Figure 17.1.19 *(continues on next two pages)* β-sheet DNA-binding motifs. (**A**) Arc-repressor tetramer complexed with DNA (PDB entry 1par). (**B**) TATA-box-binding protein (TBP) complexed with DNA (PDB entry 1ytb). (C) Histone-like HU protein (PDB entry 1hue). (**D**) HU-like IHF complexed with DNA (PDB entry 1ihf). (**E**) The p53 tumor-suppressor monomer bound to DNA (PDB entry 1tsr). Regions involved in DNA binding are labeled and colored red. The zinc atoms and side-chain ligands to the zinc are represented by ball-and-stick models. (**F**) Structure of the p50/p50 homodimer complexed with DNA (PDB entry 1svc). The two insertion regions (magenta) may play a role in binding other transcription factors. (**G**) Tetrameric complex of NFAT1 (magenta), Fos/Jun (AP-1), and DNA (PDB entry 1a02). (**H**) Ternary complex of CBFα (green), CBFβ (magenta), and DNA (blue) (PDB entry 1h9d). (**I**) Brachyuri T-domain homodimer bound to DNA (PDB entry 1h6f). (**J**) STAT-1 homodimer bound to DNA (PDB entry 1bf5). The coiled-coil domain, DNA-binding domain, linker domain, and SH2 domain are colored blue, red, green, and magenta, respectively.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

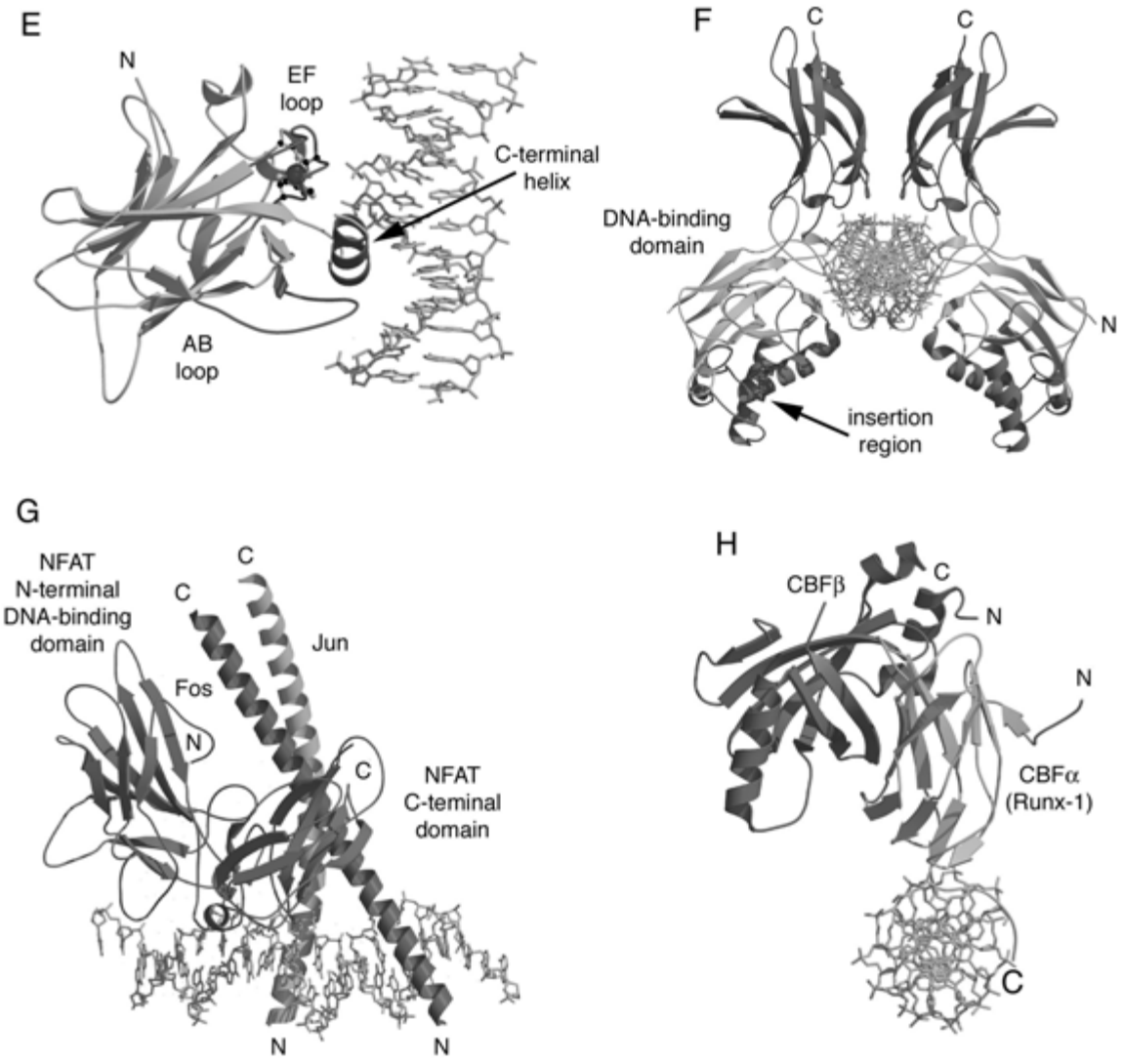


Figure 17.1.19 *(continued)*

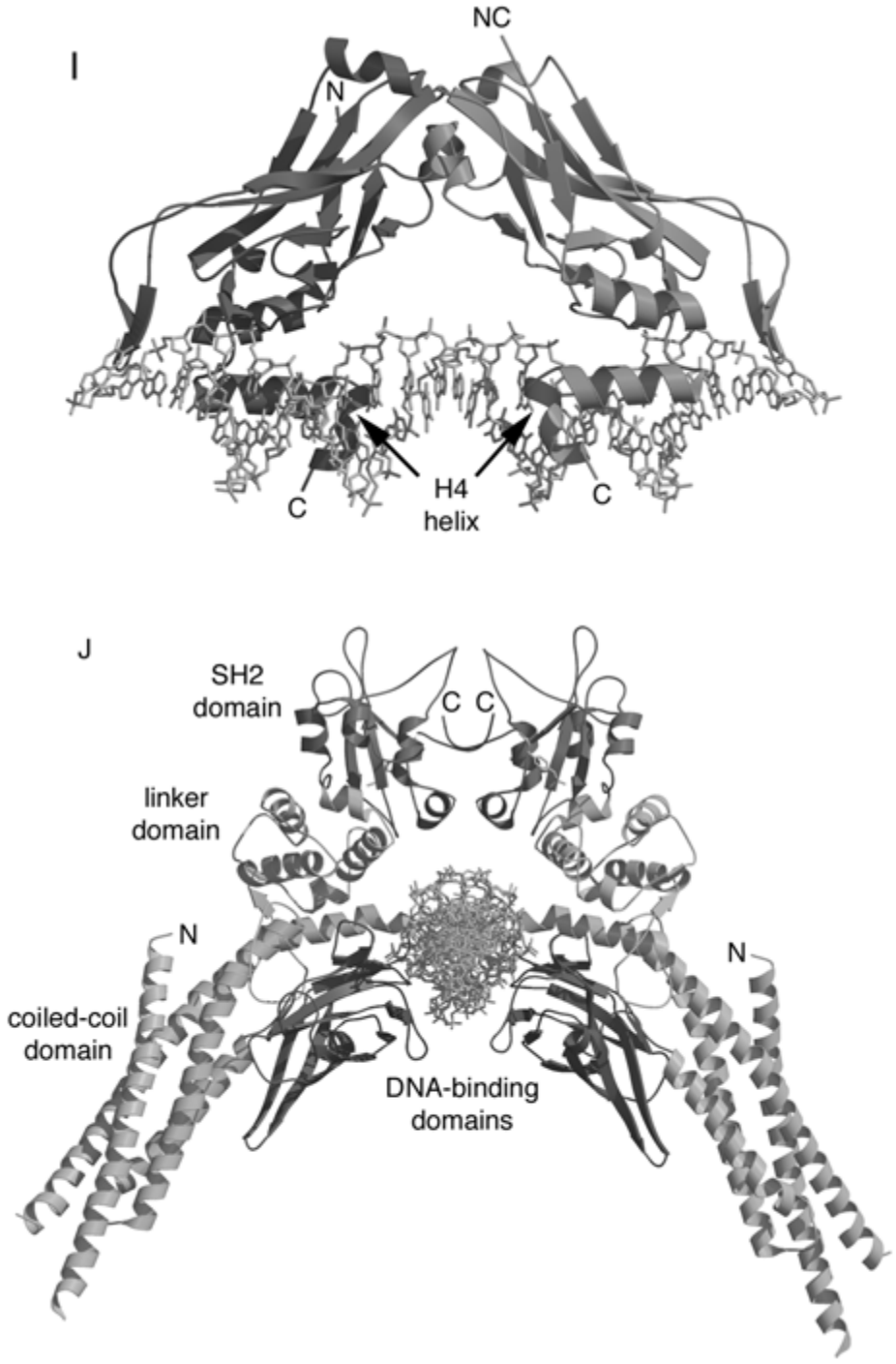


Figure 17.1.19 *(continued)*

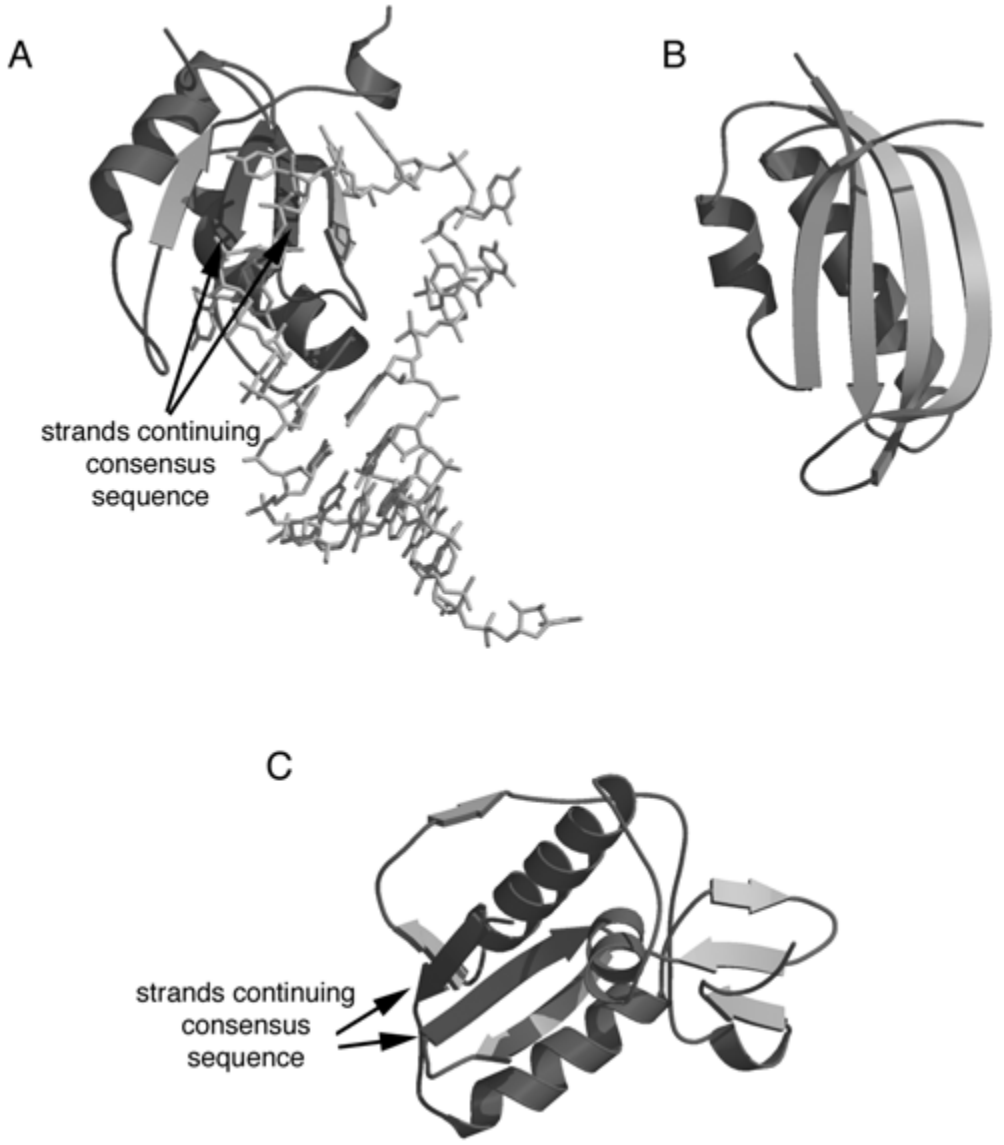


Figure 17.1.20 RNP domains. β strands containing the RNP1 or -2 consensus sequences are labeled (arrows). (**A**) U1A splicosomal protein (PDB entry 1urn). (**B**) Ribosomal protein S6 (PDB entry 1ris). (**C**) Bacteriophage T4 regA protein (PDB entry 1reg).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

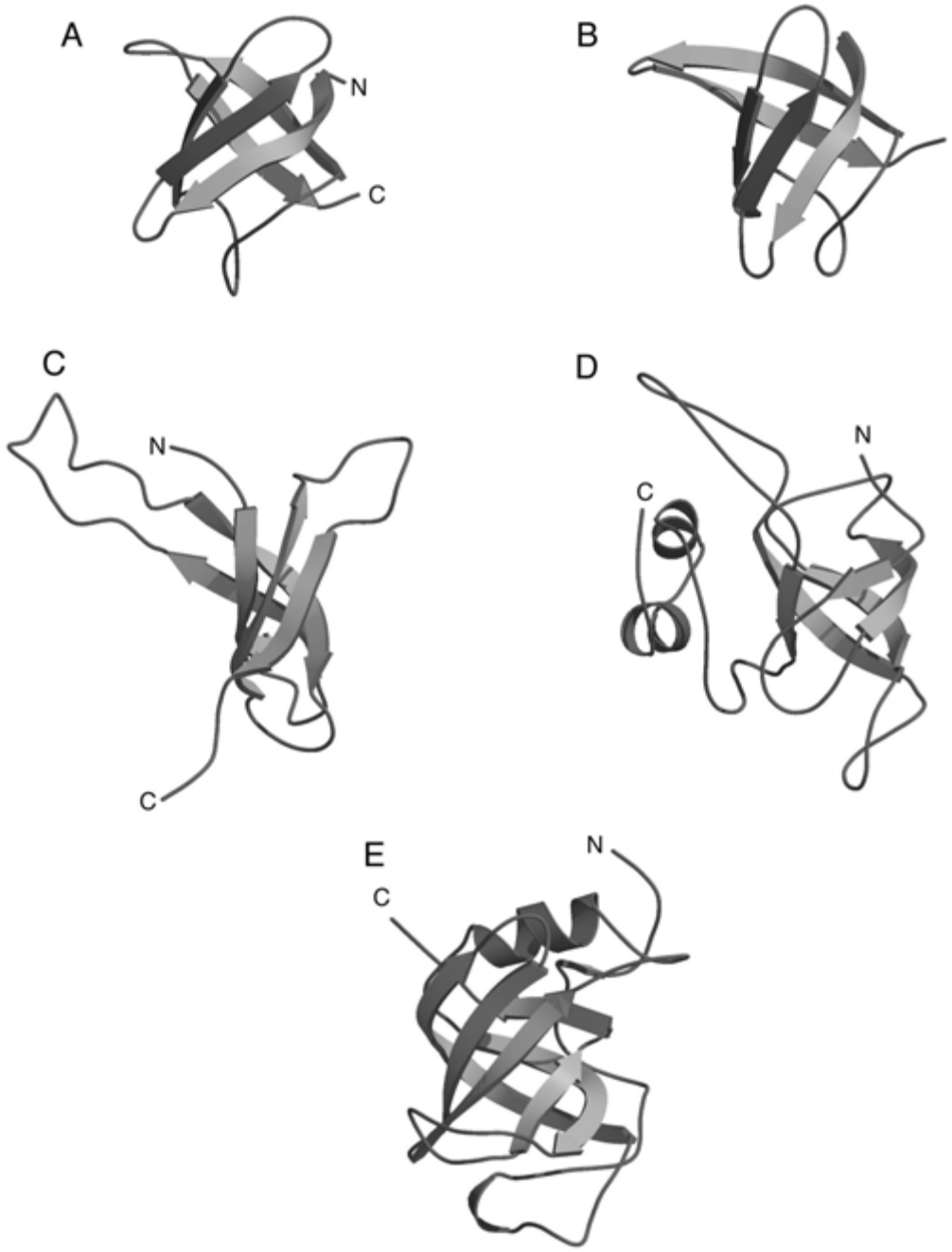


Figure 17.1.21 The OB fold. (**A**) Cold-shock protein A (PDB entry 1mjc). β strands containing RNP1 or -2 consensus sequences are colored purple. (**B**) Cold-shock protein B (PDB entry 1csp). β strands containing RNP1 or -2 consensus sequences are labeled. (**C**) Ribosomal protein S17 (PDB entry 1rip). (**D**) Ribosomal protein L14 (PDB entry 1whi). (**E**) Anticodon-binding domain of aspartyl-tRNA synthetase (PDB entry 1asz).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

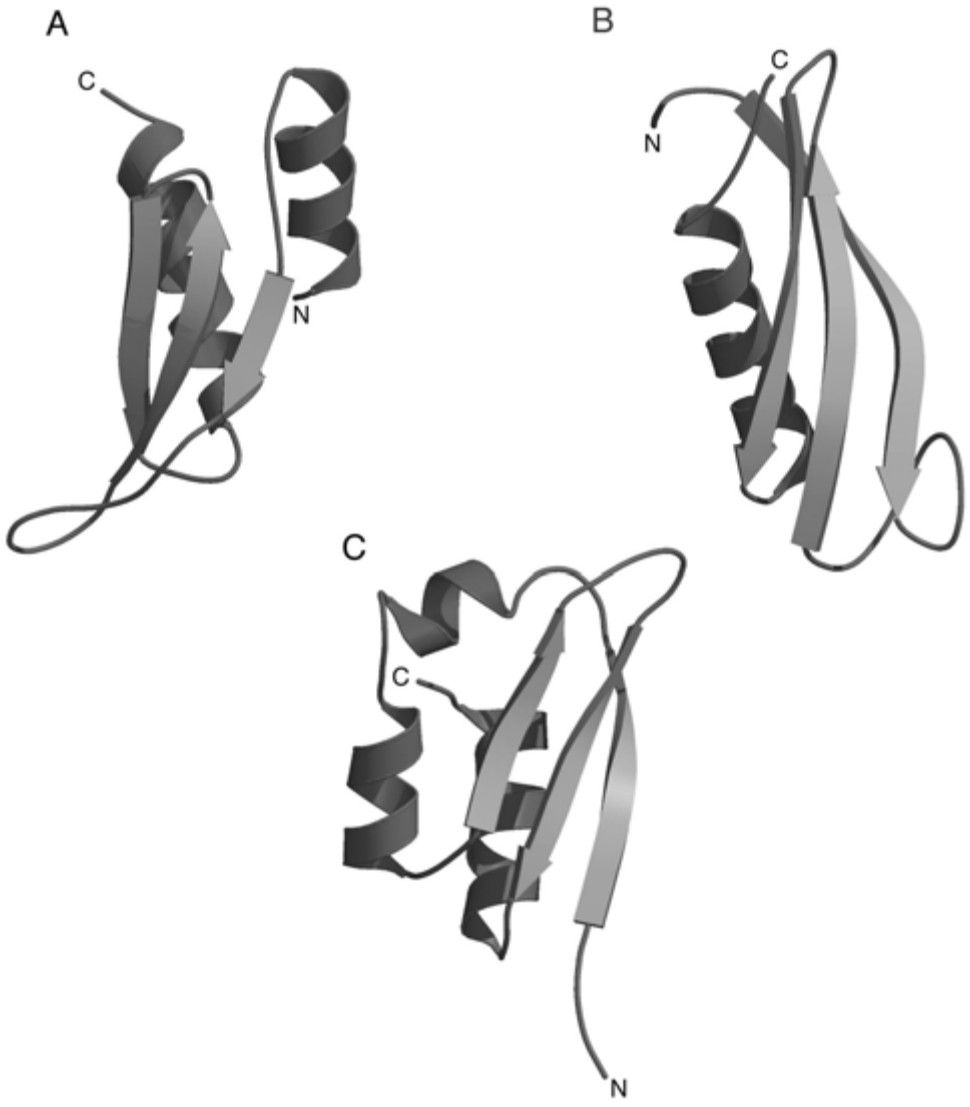


Figure 17.1.22 Double-stranded RNA-binding domain (dsRBD) and KH domain. (**A**) *Drosophila* staufen protein (PDB entry 1stu), an example of dsRBD fold. (**B**) N-terminal domain of ribosomal protein S5 (PDB entry 1pkp). (**C**) Human vigilin (PDB entry 1vih), an example of a KH domain. *(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)*

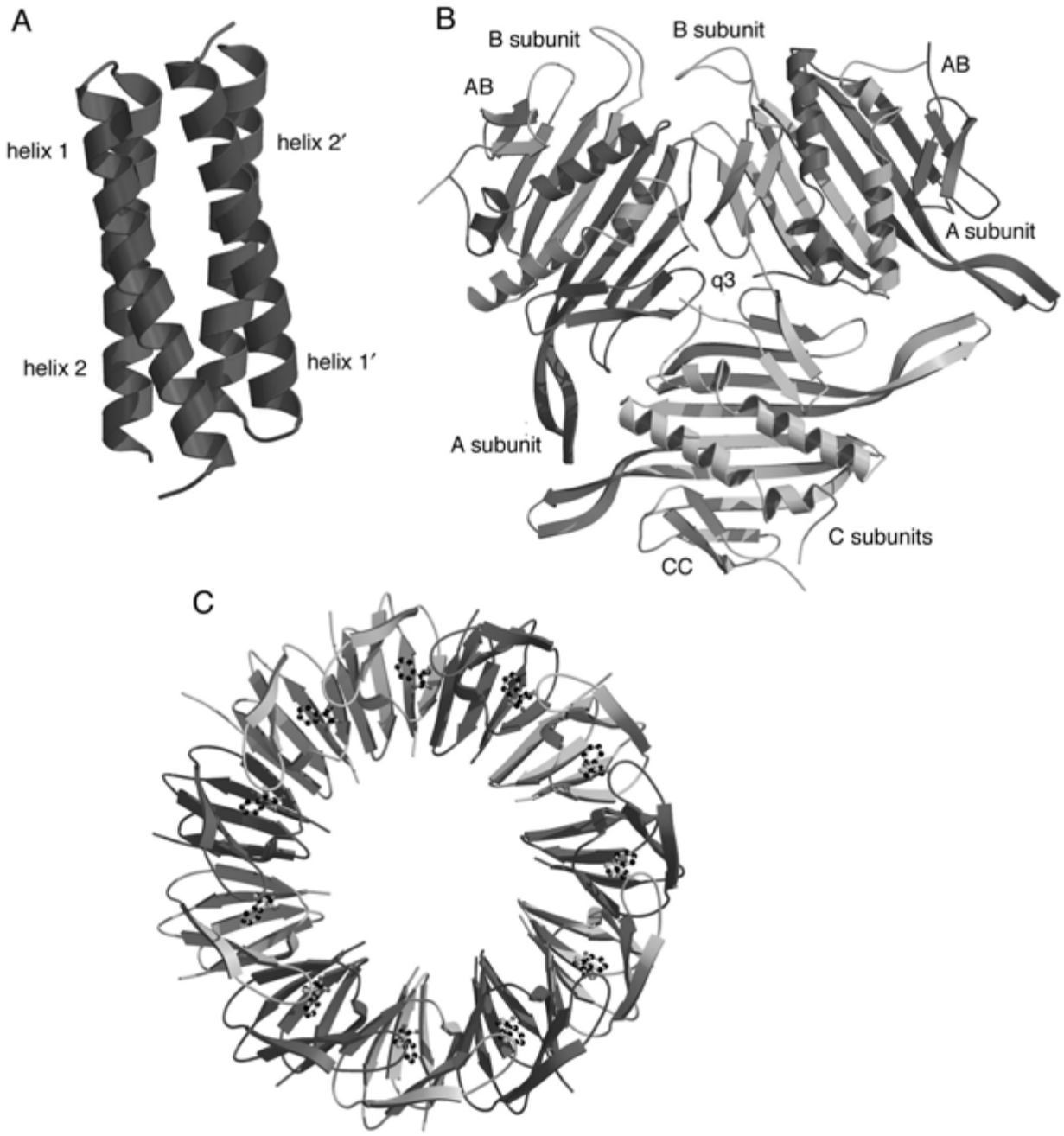


Figure 17.1.23 (**A**) The RNA-binding protein Rop from *E. coli*. (PDB entry 1rop). Helices 1 and 1′, postulated to bind to RNA, are labeled. (**B**) Hexamer of the MS2 phage-coat protein (PDB entry 1mst). Two AB dimers and a CC dimer are arranged around a quasi-3-fold rotation axis located in the center of the figure (labeled q3). (**C**) Eleven-subunit oligomer of Trp RNA-binding attenuation protein (TRAP; PDB entry 1wap). Bound tryptophan molecules are shown by a ball-and-stick model. Alternating monomers are shaded differently for clarity.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

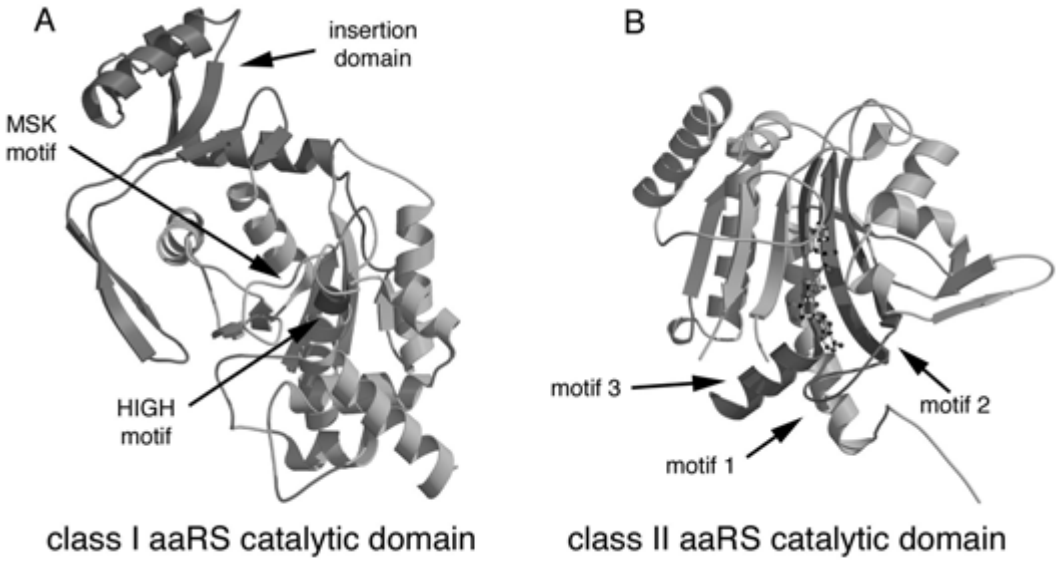


Figure 17.1.24 *(continues on next page)* Aminoacyl tRNA transferase catalytic domains *(***A** to **B**) and signal recognition particle (SRP) domains (**C** to **E**). (A) Example of a class I aaRS catalytic domain (GluRS, PDB entry 1gln). The MSK and HIGH motifs are highlighted by green and red respectively. An insertion domain (common in class I catalytic domains) is colored magenta. (B) Example of a class II catalytic domain (HisRS, PDB entry 1htt). Motifs 1 to 3 are colored green, red, and magenta, respectively. (C) Structure of SRP9/14 heterodimer of SRP *Alu* domain bound to RNA (red coil). SRP9 and -14 are colored in green and cyan, respectively (PDB entry 1e8o). (D) Structure of SRP19 in complex with RNA (cyan; PDB entry 1jid). (E) Structure of ffh-M domain (green) in complex with RNA (red; PDB entry 1dul). (F) Structure of a SRP receptor (PDB entry 1fts). The N domain and I box are both colored magenta.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

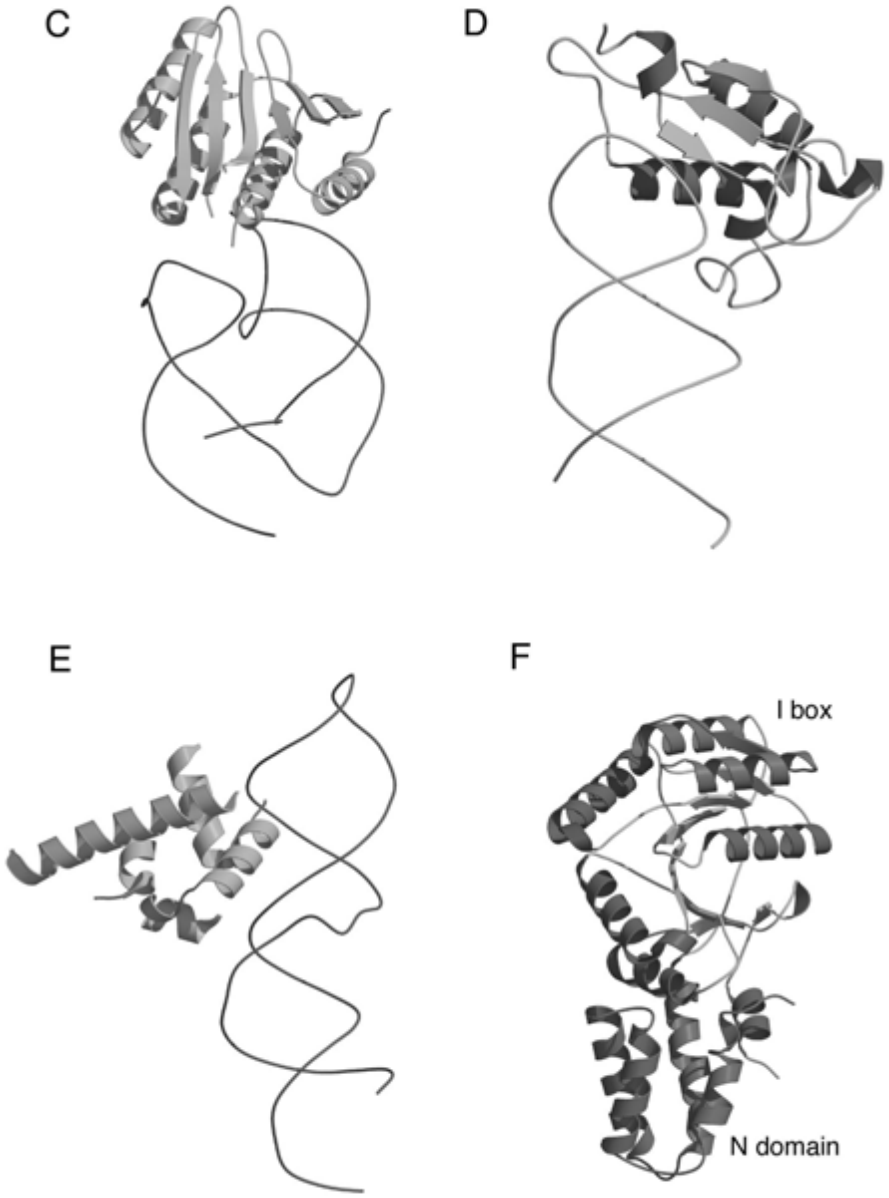


Figure 17.1.24 *(continued)*

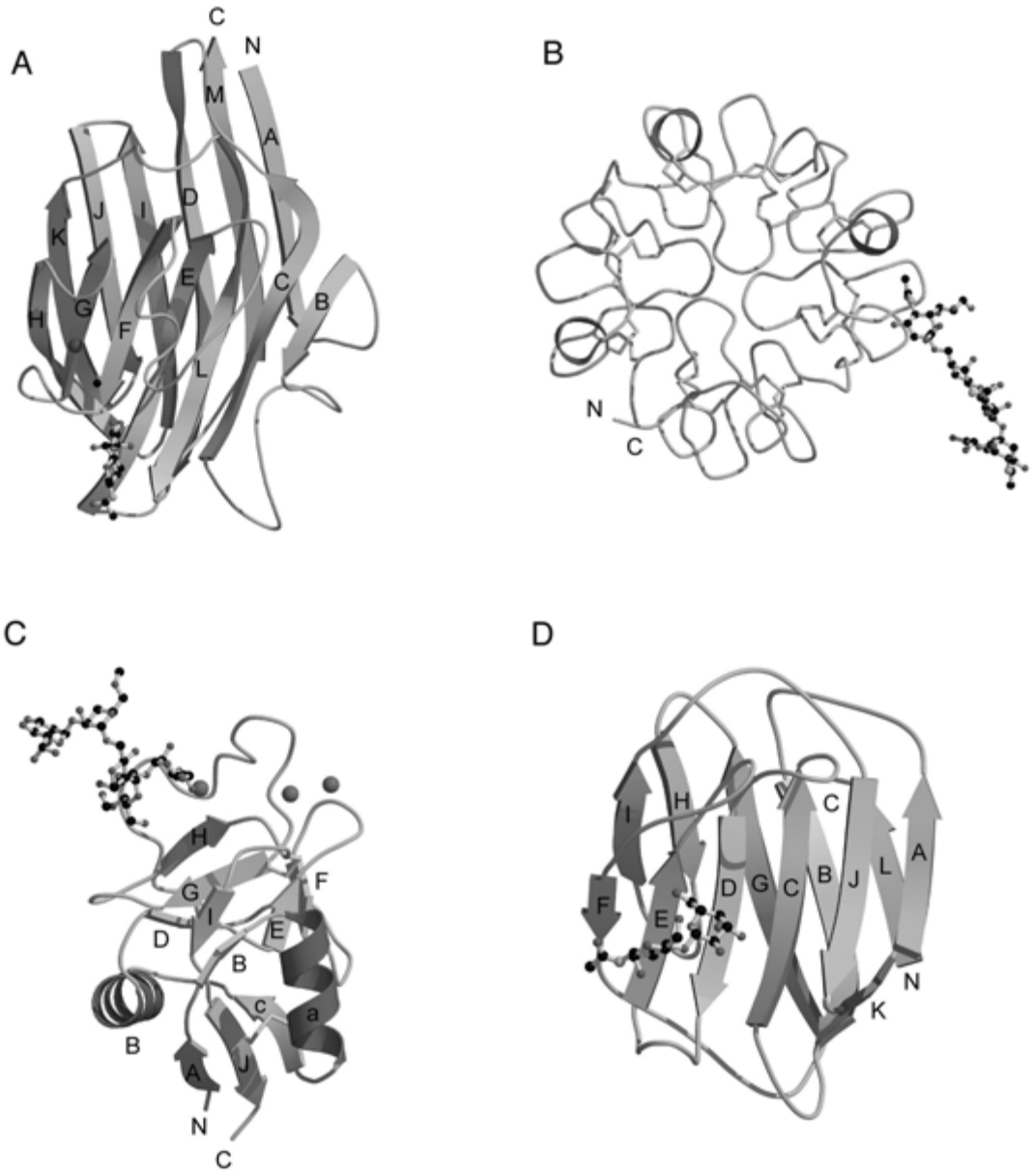


Figure 17.1.25 Lectin folds. (**A**) Legume lectin soybean agglutinin (PDB entry 1sba). The β strands are labeled A through M. The bound carbohydrate molecules are represented by ball-and-stick models. (**B**) Wheat-germ agglutinin monomer with bound carbohydrate. Disulfide bonds are shown as yellow sticks. (PDB entry 2wgc). (**C**) The C-type lectin mannose-binding protein A (PDB entry 2msb) with bound carbohydrate. Calcium ions are shown as magenta spheres. (**D**) S-lectin with bound carbohydrate (PDB entry 1slt).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

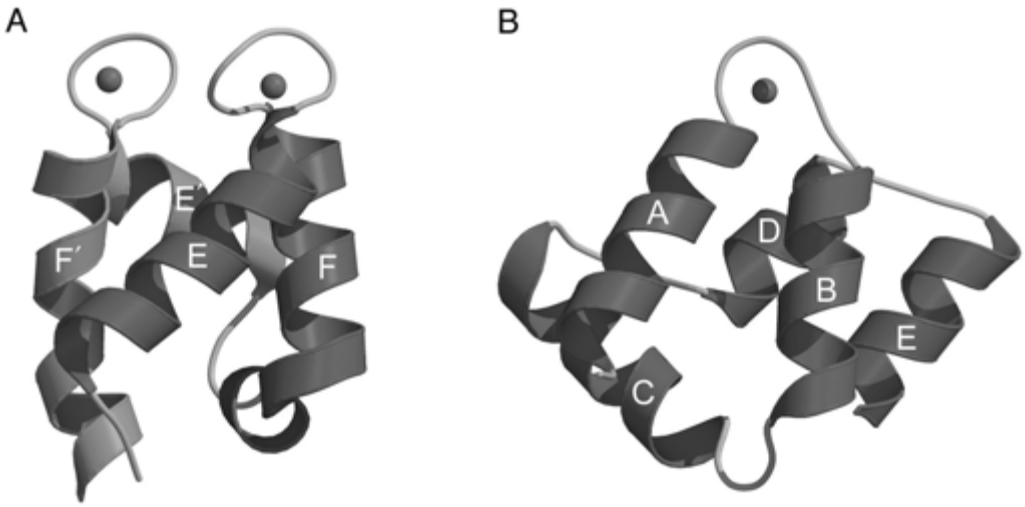


Figure 17.1.26 Calcium-binding folds. Calcium ions are drawn as magenta spheres. (**A**) A pair of calcium-binding EF-hands (red and orange respectively) complexed with Ca^{2+} from bovine calbindin D9K. The corresponding helices are labeled E, F, E′, and F′ (PDB entry 4icb). (**B**) Calcium-binding domain of annexin V (PDB entry 1ala).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

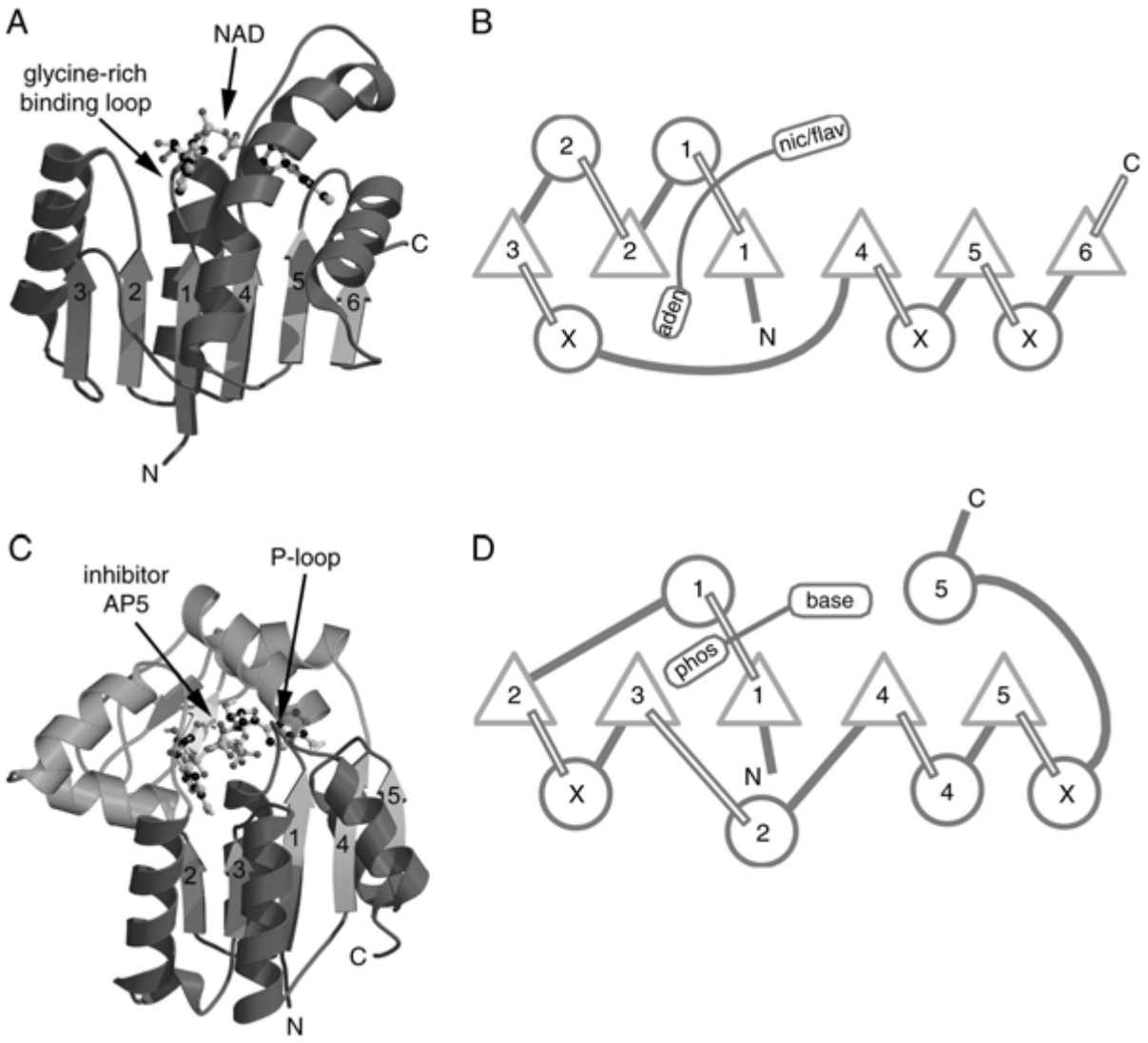


Figure 17.1.27 The classical dinucleotide- and mononucleotide-binding folds shown by ribbon drawings and topology diagrams. In the topology diagrams, the β-sheets are viewed from the C-terminal edge with β-strands represented by green triangles and α-helices by red circles. Numbers indicate strand and helix order within the structure. Circles containing an "X" represent possible domain insertions. Bound nucleotides are colored purple. (**A**) Ribbon drawing of the NAD-binding domain of dogfish lactate dehydrogenase with bound NAD represented as a ball-and-stick model (PDB entry 1ldm). (**B**) Topology diagram for the common core of the classical dinucleotide-binding fold. (**C**) Ribbon drawing of adenylate kinase as an example of a mononucleotide-binding protein (PDB entry 1aky). The portion of the structure that is not part of the nucleotide-binding fold is colored purple. The bound inhibitor *bis*(adenosine)-5′-pentaphosphate (AP5) is shown using a ball-and-stick model. (**D**) Topology diagram for the common core of the classical mononucleotide-binding fold.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

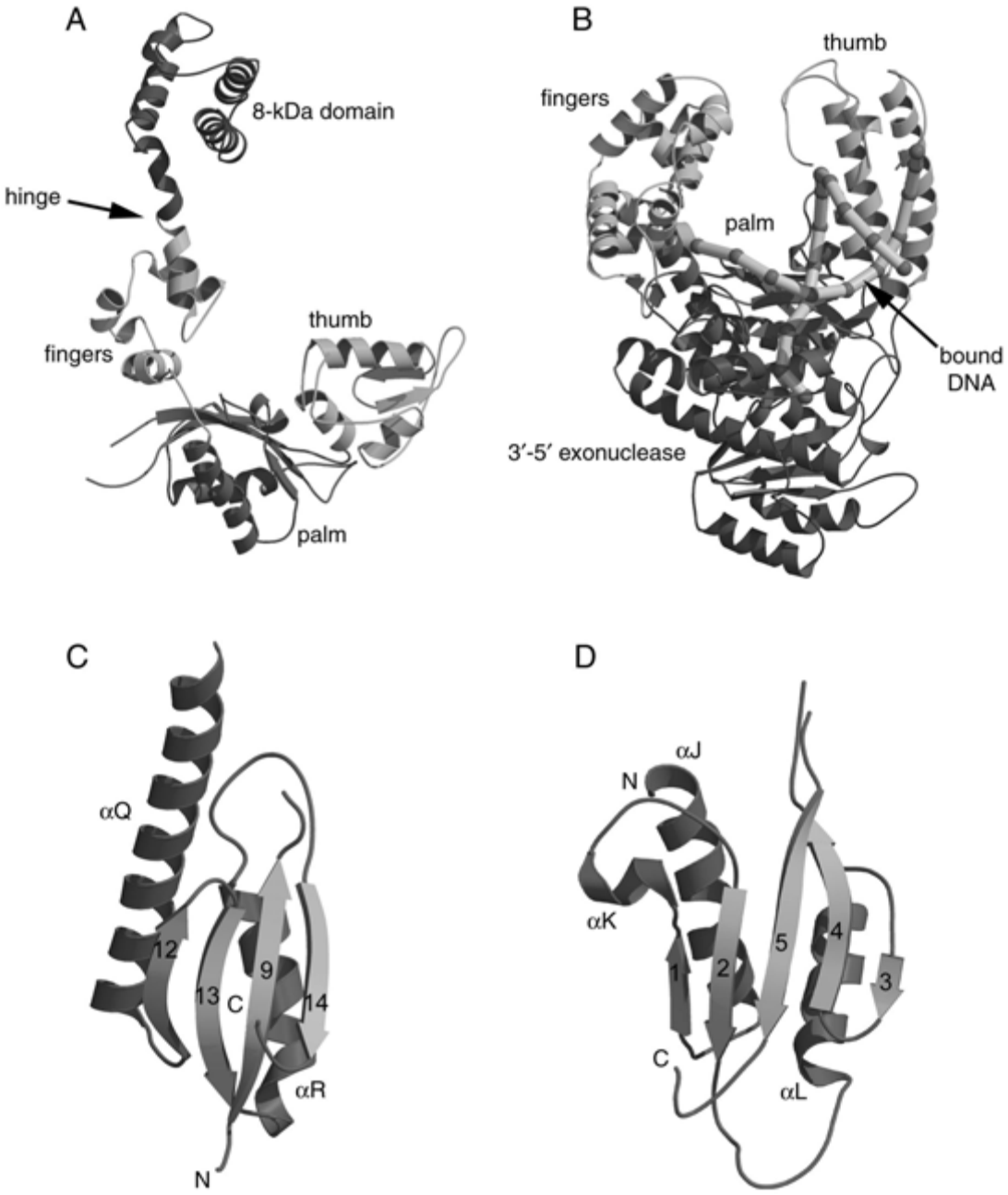


Figure 17.1.28 *(continues on next two pages)* Structures of DNA and RNA polymerases. (**A**) The 39-kDa catalytic domain of rat DNA polymerase β (PDB entry: 1bpd). (**B**) The Klenow fragment of *E. coli* DNA polymerase I bound to an 11-bp duplex DNA in an editing complex (PDB entry: 1kln). The DNA is represented by a yellow and purple phosphate backbone trace. (**C**) The conserved palm subdomain from Klenow fragment (PDB entry 1kln). (**D**) The palm subdomain from rat DNA polymerase β (PDB entry 1bpd). (**E**) Structure of the HIV-1 reverse transcriptase heterodimer complexed with a 19/18-base duplex DNA (PDB entry 1hmi). The p66 subunit is colored by subdomain and the p51 subunit is dark blue. The bound DNA is represented by a yellow and purple phosphate backbone trace. (**F**) *Thermus thermophilus* RNA polymerase holoenzyme (PDB entry 1iw7). The β′, β, α′, α″, ω, and σ^{70} subunits are labeled and colored green, cyan, red, tan, magenta, and dark blue, respectively. (**G**) The N-terminal domain of the RNAP α subunit from *E. coli* (PDB entry 1bdf). (**H**) RNAP II from yeast (PDB entry 1i50). Visible subunits are labeled and the five subunits in common with bacterial RNAP follow the same coloring scheme as in (F).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

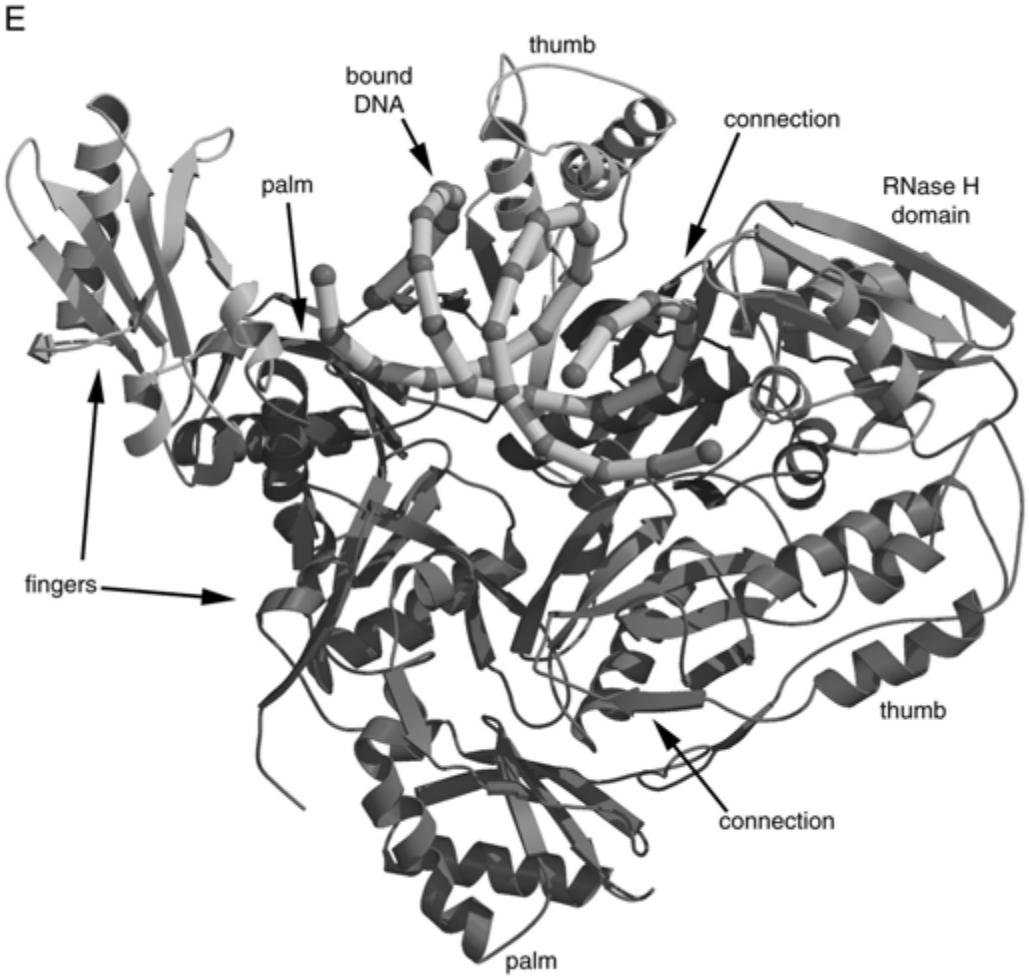


Figure 17.1.28 *(continued)*

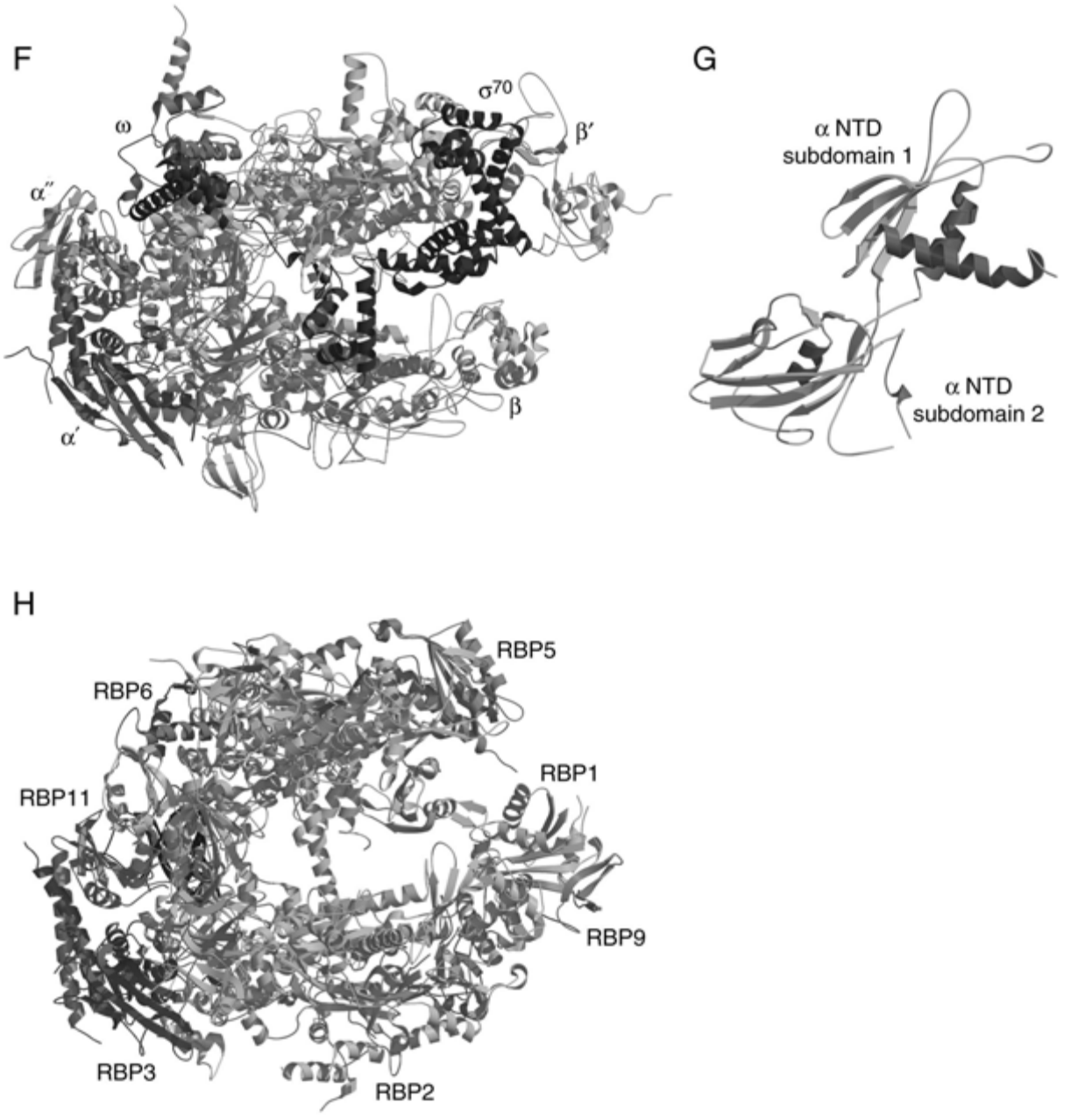


Figure 17.1.28 *(continued)*

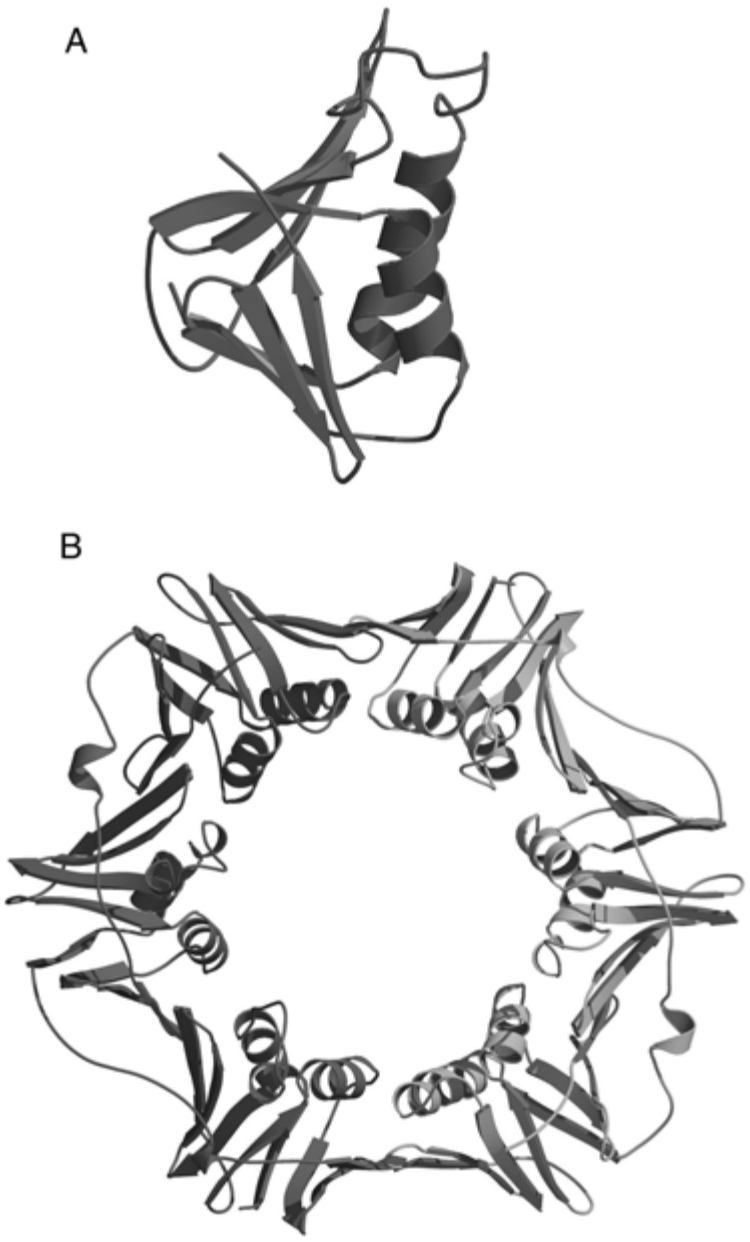


Figure 17.1.29 Structure of the β-subunit of pol III (PDB entry: 2pol), an example of a sliding-clamp DNA polymerase processivity factor. (**A**) A single domain of the β-subunit. (**B**) View of the entire ring structure of the homodimer looking down the six-fold symmetry axis. The two monomers are colored red and green, respectively.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

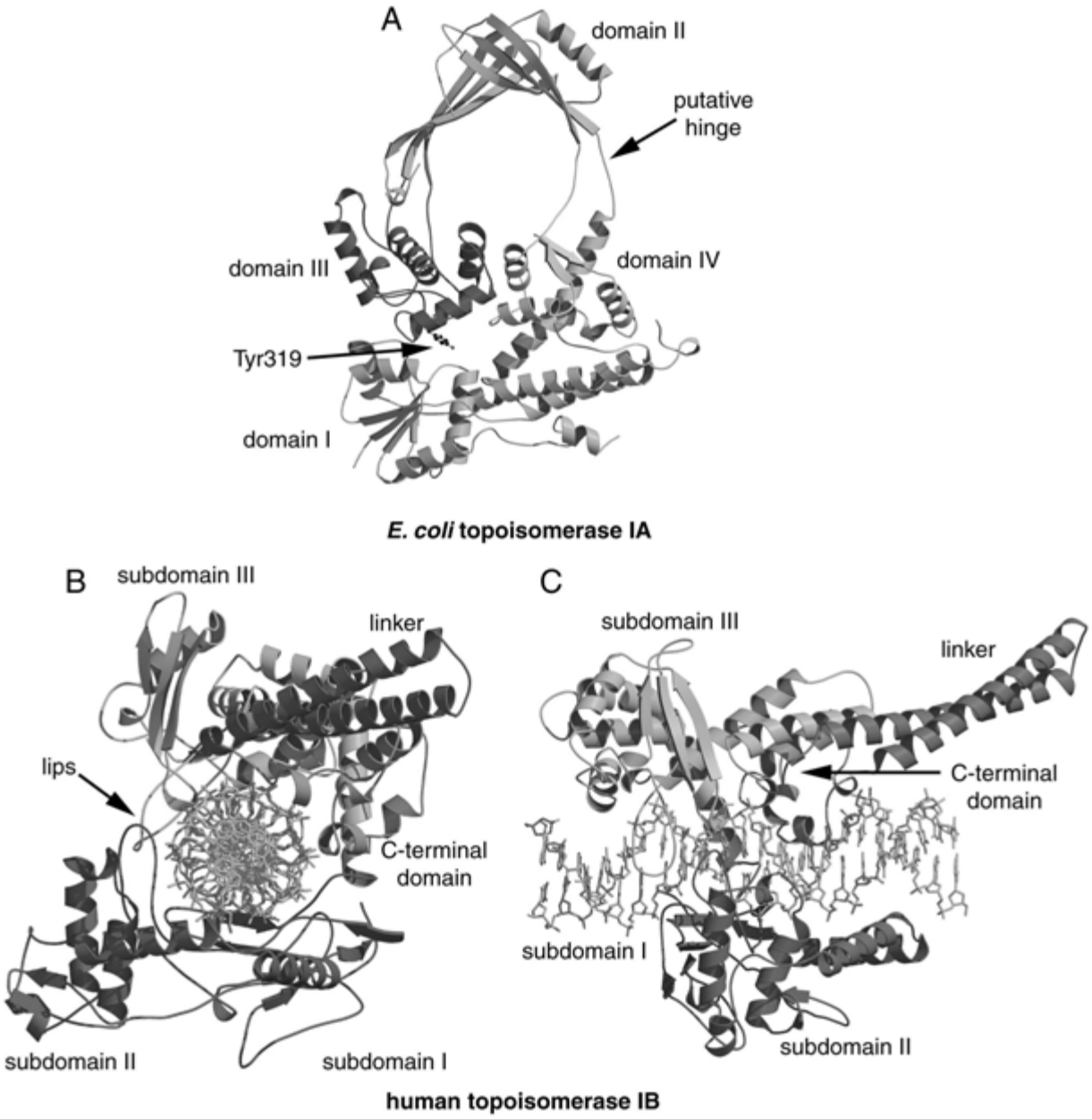


Figure 17.1.30 *(continues on next two pages)* Topoisomerase structures. (**A**) Structure of the N-terminal 67 kDa fragment of *E. coli* topoisomerase I (PDB entry 1ecl). The four domains are labeled and the side chain of the active-site Tyr319 is represented by a ball-and-stick model. (**B**) Complex of human topoisomerase IB with 22-bp DNA duplex (PDB entry 1a36). Each domain is color coded and labeled. DNA is a light-blue ball-and-stick model. (**C**) Same as (B), but rotated 90° around the vertical axis. (**D**) Structure of the 92-kDa fragment of yeast topoisomerase IIA showing the organization of domains. (PDB entry 1bgw). (**E**) Homodimeric form of the 92-kDa fragment of yeast topoisomerase IIA. For clarity, the two monomers are colored red and green, respectively. (**F**) Homodimer of the 43-kDa fragment of *E. coli* GyrB (PDB entry 1ei1). One monomer is blue-green, the other is red-orange. ADPNP bound to each monomer is represented by a ball-and-stick model. (**G**) Homodimer of *Methanococcus jannaschii* topoisomerase VI subunit A (PDB entry 1d3y). One monomer is colored according to domain and subdomain, the other is gray (program color "white").
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

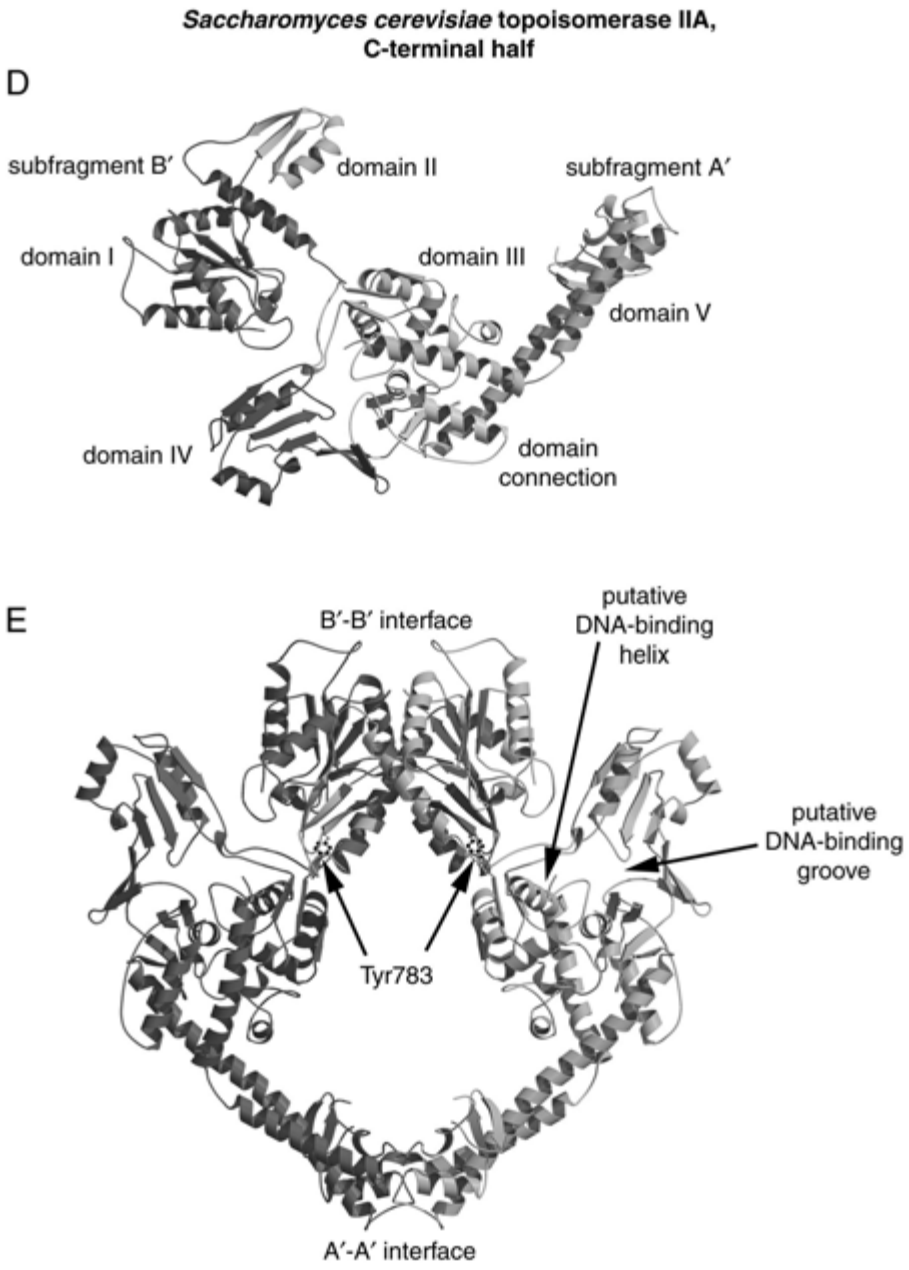
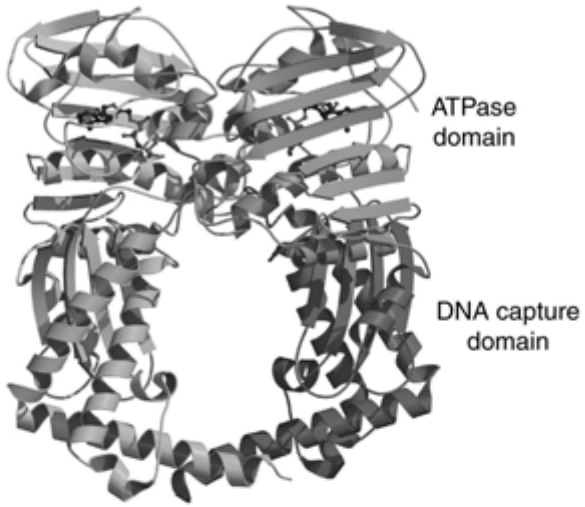


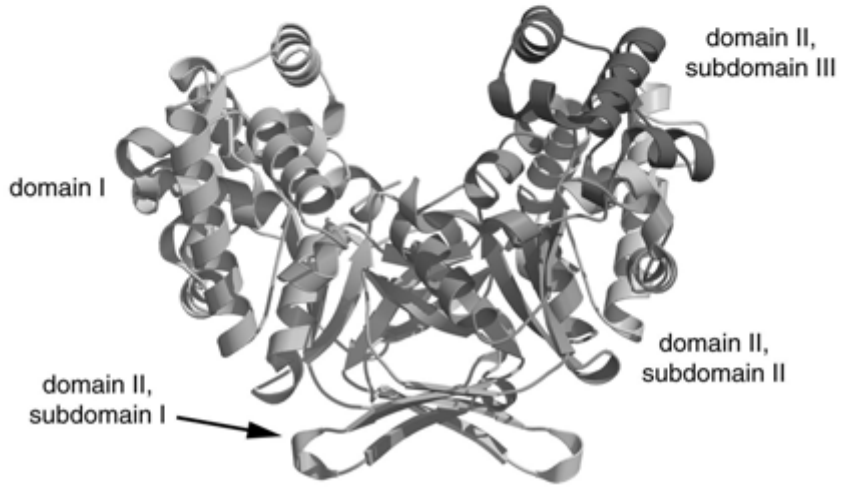
Figure 17.1.30 *(continued)*

F



43-kDa fragment of *E. coli* GyrB subunit

G



***Methanocuccus jannaschii* DNA topoisomerase VI subunit A**

Figure 17.1.30 *(continued)*

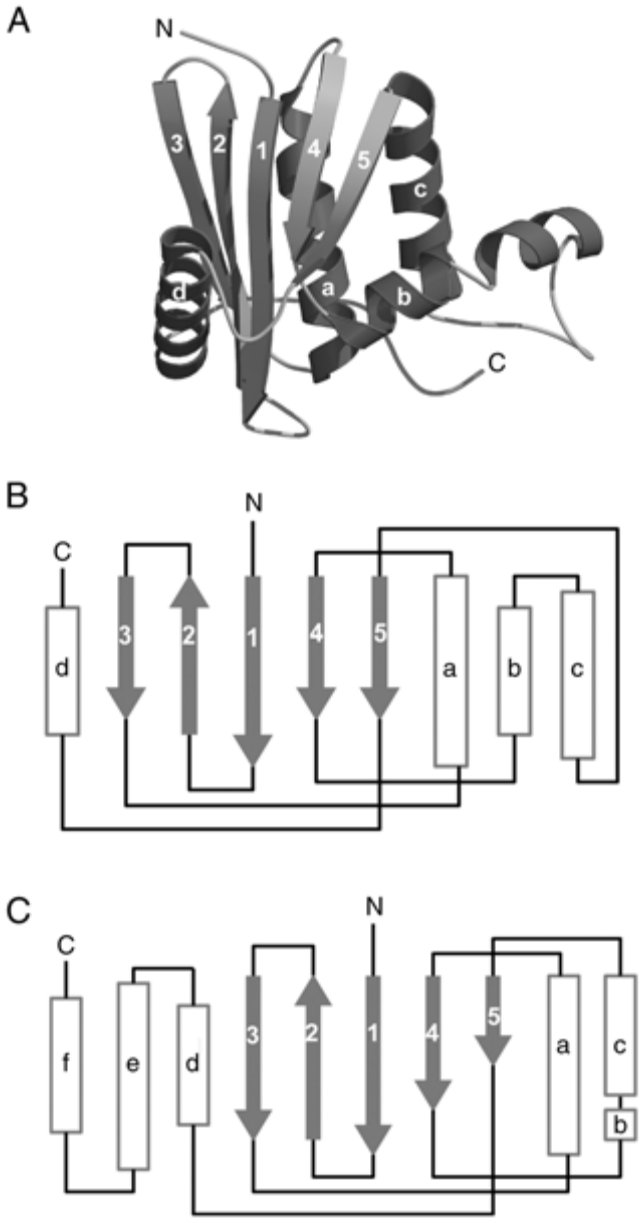


Figure 17.1.31 The polynucleotidyl transferase RNase H-like family of folds. (**A**) Structure of RNase H from *E. coli (PDB entry 2rn2) and (***B**) Corresponding secondary structure topology. (**C**) The secondary structural topology of HIV integrase.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

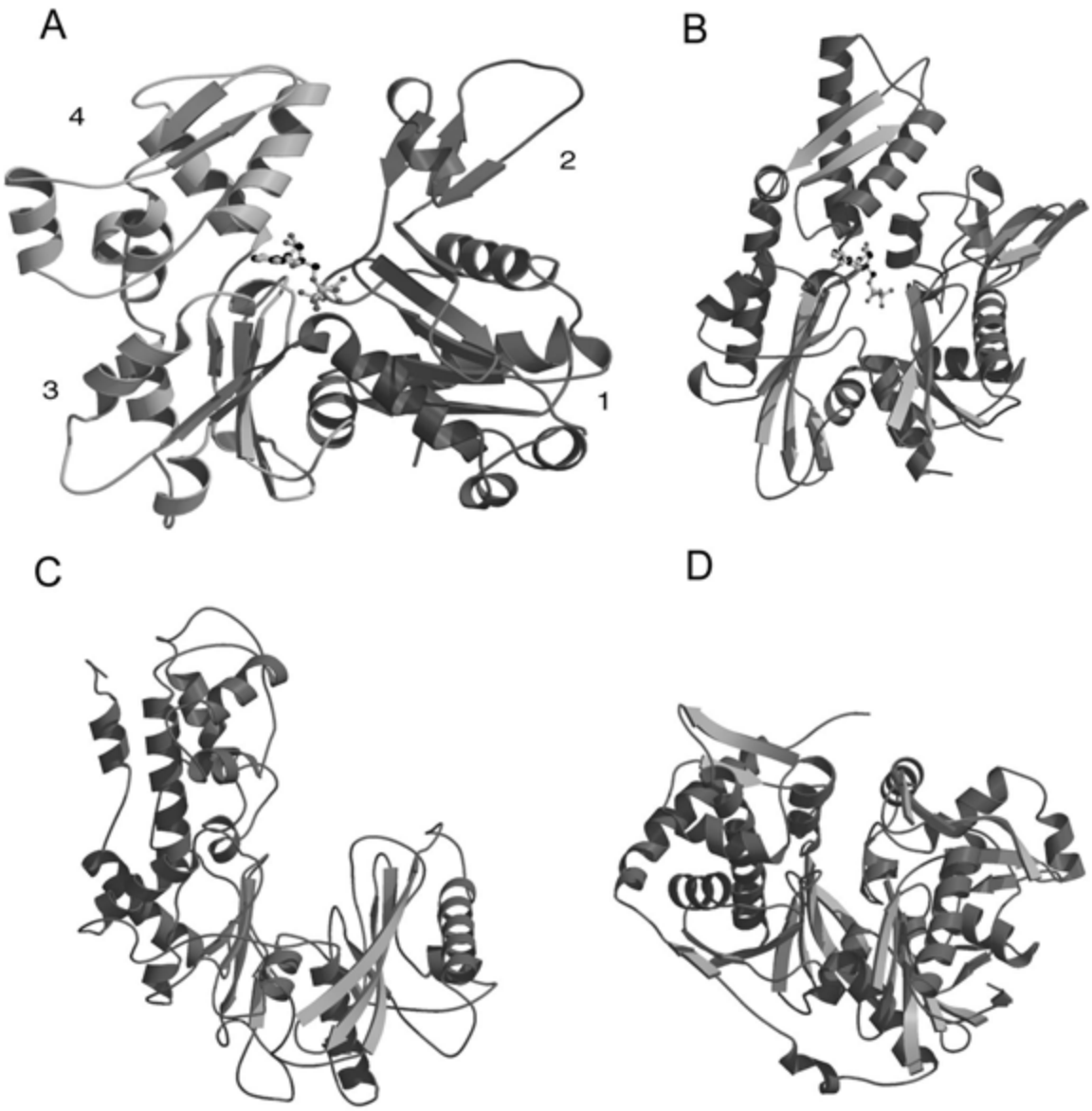


Figure 17.1.32 *(continues on next page)* Structures containing the actin fold (**A** to **D**) and structures that bind to actin (**E** to **G**). (A) Actin complexed with ATP (PDB entry 1atn). Domains 1 and 2 are colored purple and blue, respectively. Subdomains 1 to 4 are labeled. ATP is shown as a ball-and-stick model. (B) The N-terminal fragment of heat-shock cognate protein (hsc70) complexed with ADP (PDB entry 1nga). ADP is shown as a ball-and-stick model. (C) Hexokinase B in its open form (PDB entry 2yhx). Structures of (D) glycerol kinase (PDB entry 1gla), (E) profilin (PDB entry 2btf), and (F) severin (NMR structure, PDB entry 1svq). (G) The spectrin repeat is shown as a dimer (PDB entry 1spc), with the two monomers colored red and blue, respectively.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

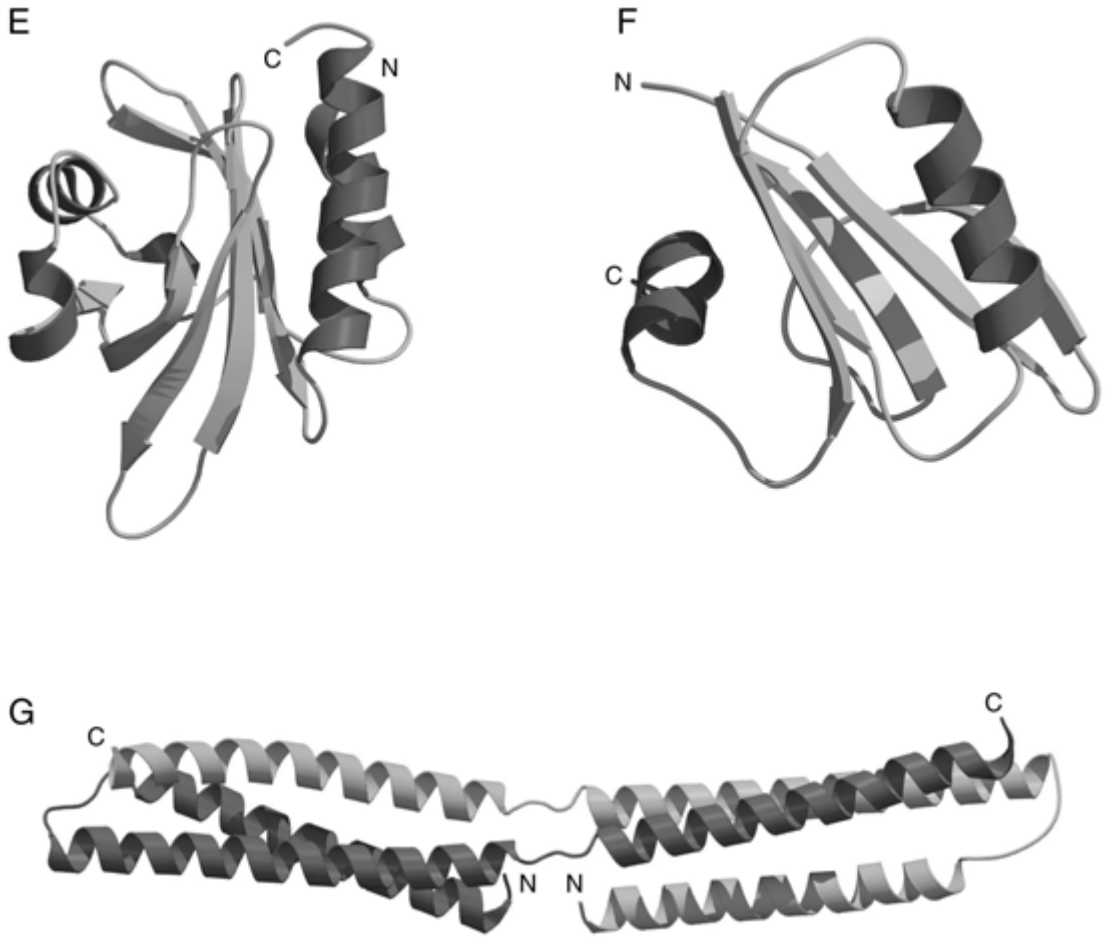


Figure 17.1.32 *(continued)*

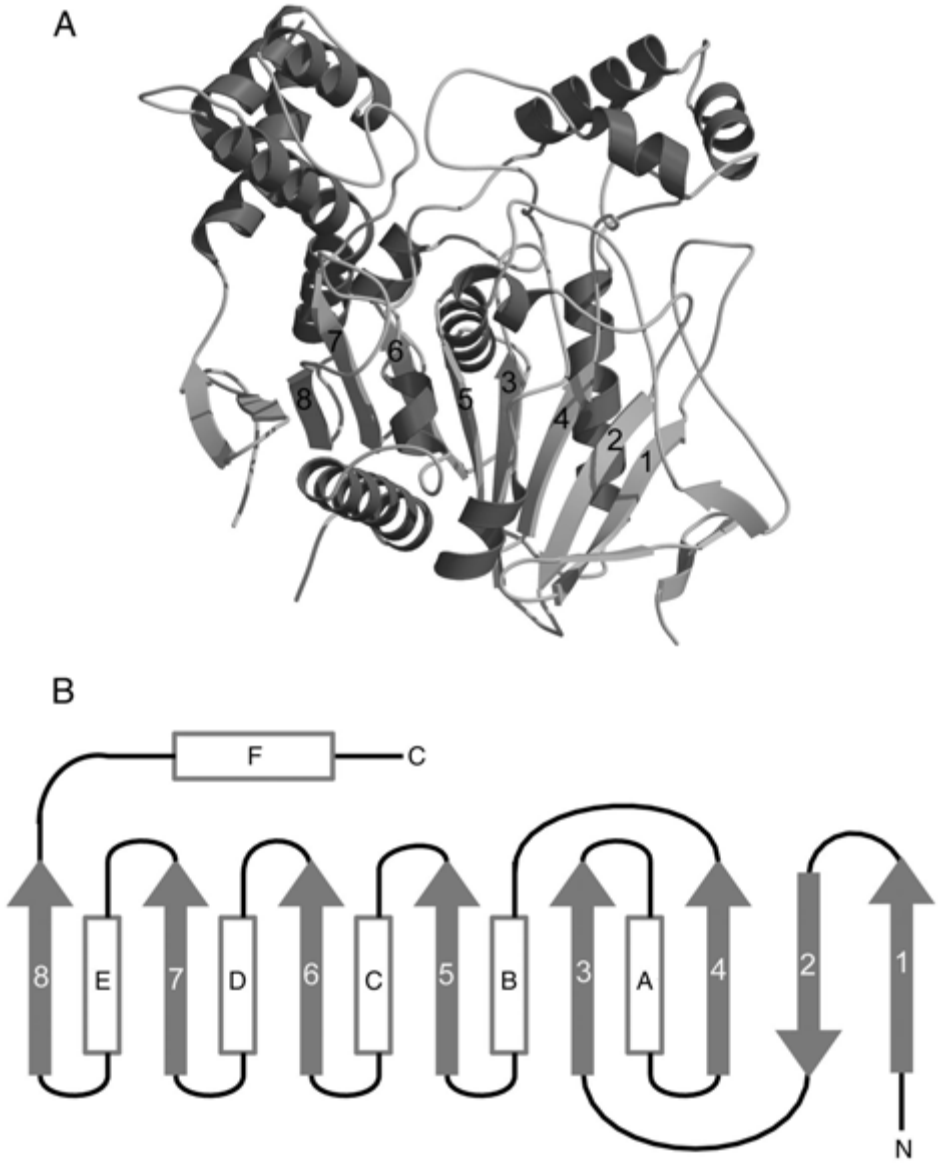


Figure 17.1.33 *(continues on next page)* (**A**) Structure of acetylcholinesterase from *Torpedo Californica*. The central eight-stranded β-sheet is shown in green. (PDB entry 1ack). (**B**) Diagram of the secondary structure topology common to all α/β hydrolases. The strands are numbered from one to eight and helices are labeled A through F. (**C**) Structure of an influenza virus neuraminidase complexed with an inhibitor (PDB entry 1nnc). The active site-bound inhibitor is shown as ball-and-stick model.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

Figure 17.1.33 *(continued)*

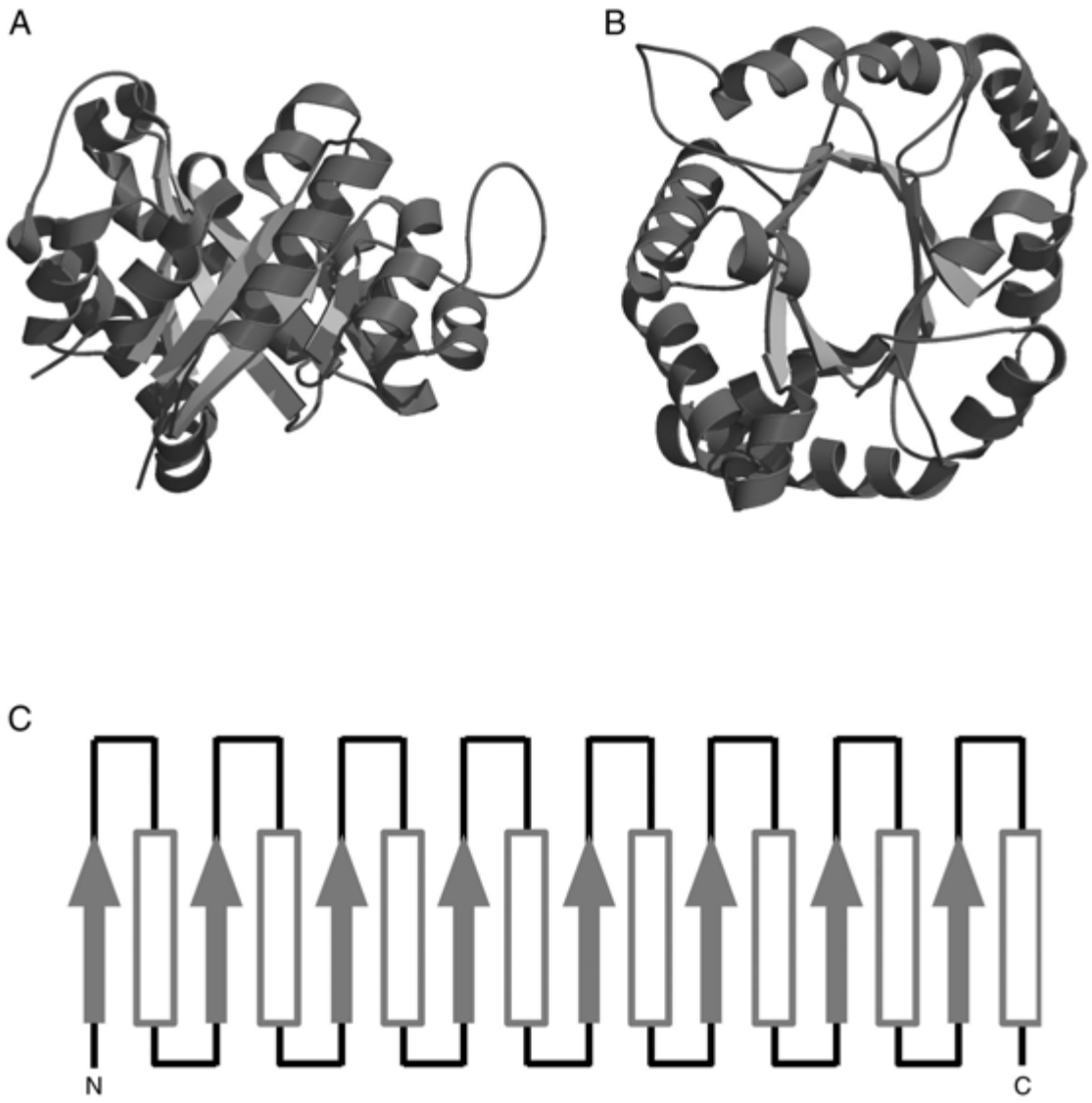


Figure 17.1.34 *(continues on next page)* The TIM-barrel *(***A** to **C**) and serine protease fold (**D** to **F**). The side chains of residues in the active site catalytic triad are shown as ball-and-stick models in panels (**D**) and (**F**). (A) Side view of a ribbon drawing of triose phosphate isomerase as an example of a TIM-barrel. (B) Ribbon drawing of triose phosphate isomerase viewed from the top. (C) Secondary structure schematic of the classical TIM-barrel fold. β-strands are represented by green arrows and α-helices by red rectangles. (D) Structure of the trypsin-like serine protease, collagenase (PDB entry 1hyl). (E) Structure of the subtilisn serine protease, subtilisn BPN′ (PDB entry 1sup). (F) The structure of the serine carboxypeptidase, wheat serine carboxypeptidase II (PDB entry 1wht).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

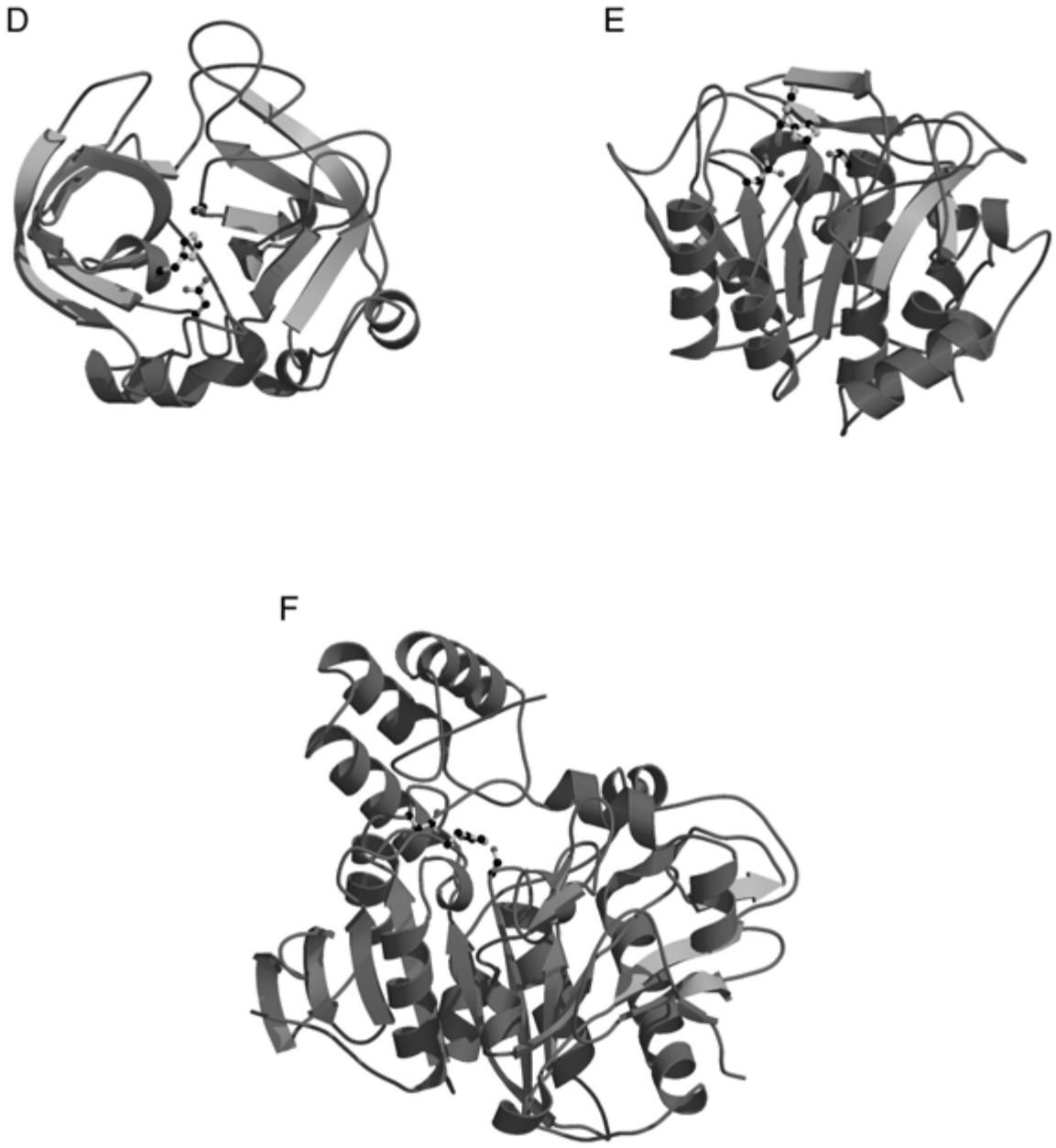


Figure 17.1.34 *(continued)*

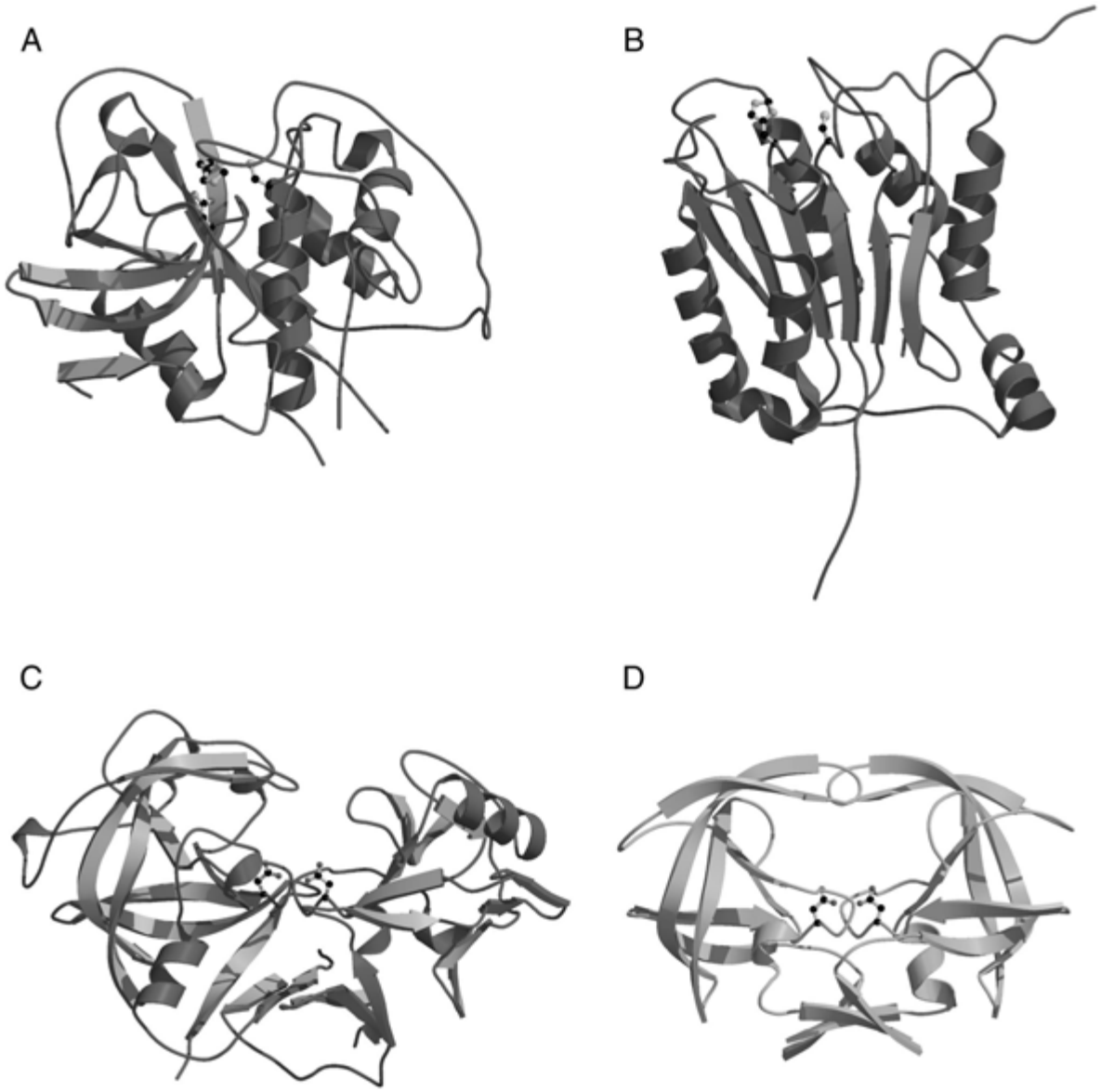


Figure 17.1.35 *(continues on next two pages)* Examples of cysteine proteases *(***A** and **B**), aspartic proteases (**C** and **D**), and metalloproteases (**E** to **G**). The side chains of critical active site residues mentioned in the text are shown as ball-and-stick models. Zinc atoms are shown as magenta balls. (A)The papain-like cysteine protease, human cathepsin B (PDB entry: 1huc). (B) Interleukin-1b converting enzyme (ICE; PDB entry: 1ice). (C) Pepsin-like protease, human renin (PDB entry 1bbs). (D) The retroviral protease, HIV-1 protease (PDB 1hpx). (E) Structure of a zinc-dependent endopeptidase, a metzincin, snake venom adamalysin II (PDB entry 1iag). (F) Structure of a zinc-dependent exopeptidase, aminopeptidase from *Aeromonas proteolytica* (PDB entry 1amp). (G) Structure of alkaline protease, a metzincin from *Pseudomonas aeruginosa* (PDB entry 1akl). The active site zinc and coordinated side chains are shown as ball-and-stick models in the N-terminal catalytic domain (left side). Bound calcium ions are shown as black balls in the C-terminal parallel β-roll domain (right side). (H) A catalytic β-subunit from the 20S yeast proteasome as a representative of the Ntn fold (PDB entry 1ryp). The nucleophilic threonine is shown as a ball-and-stick model.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

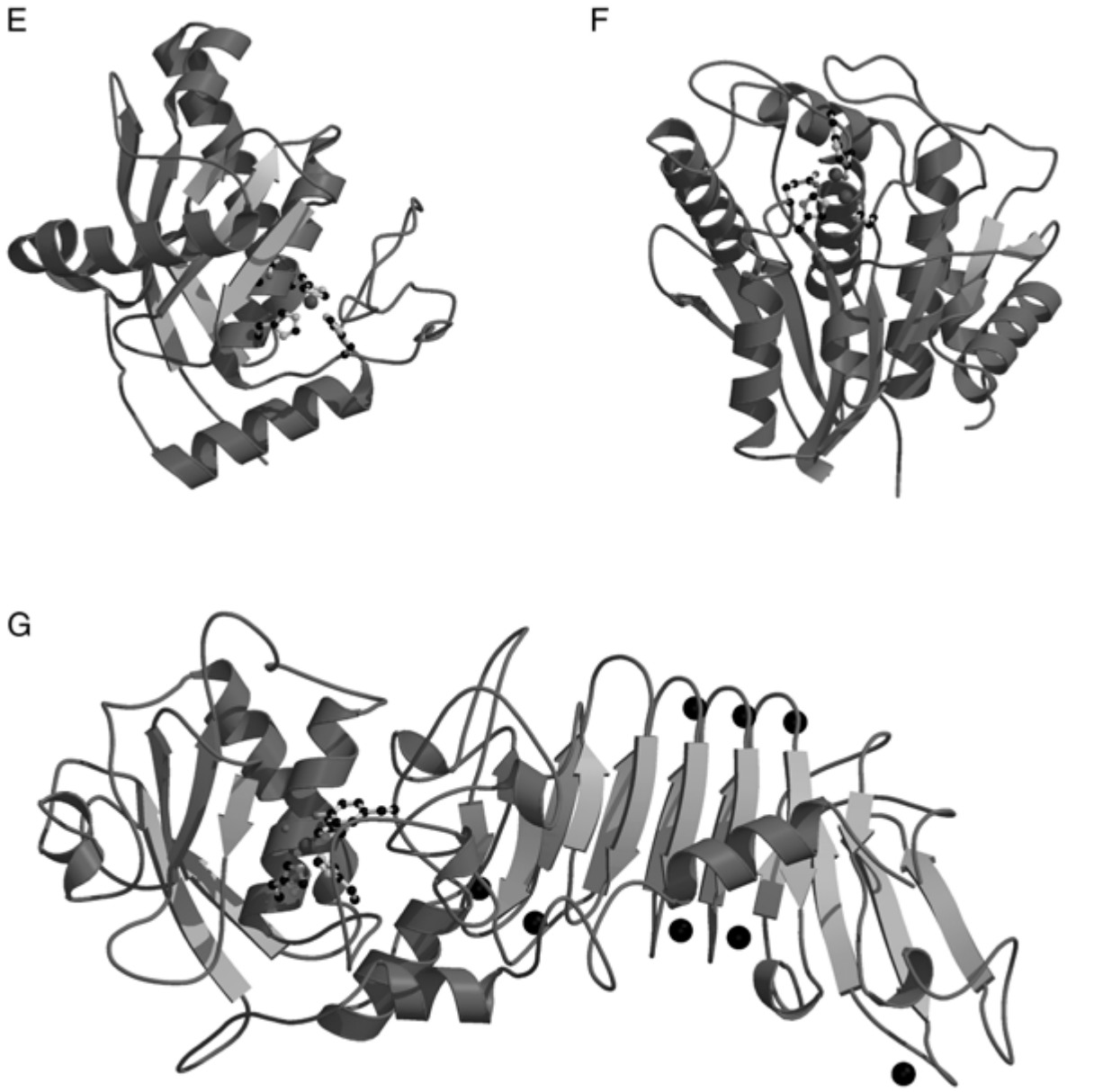


Figure 17.1.35 *(continued)*

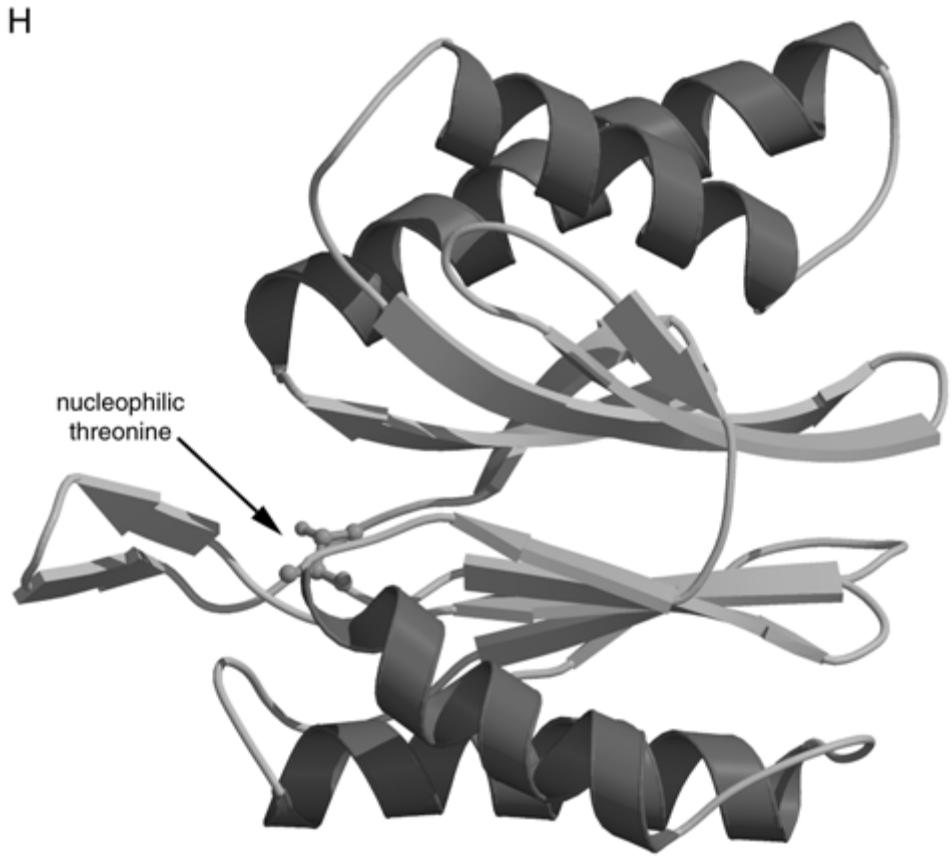


Figure 17.1.35 *(continued)*

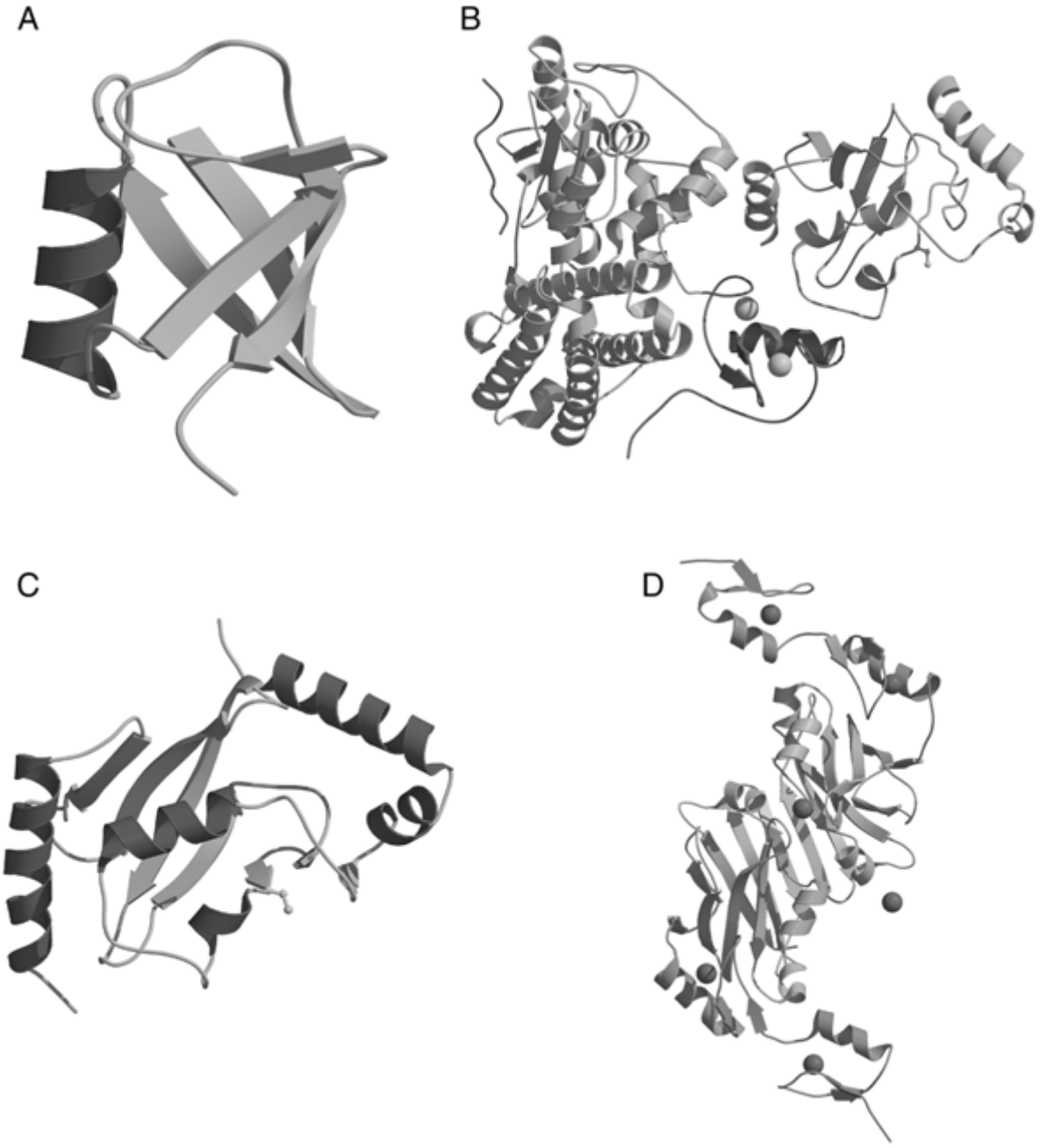


Figure 17.1.36 *(continues on next page)* Enzymes of the ubiquitin pathway. *(***A**) The structure of ubiquitin (PDB entry 1ubq). (**B**) Complex between the ubiquitin-conjugating E3 enzyme c-Cbl (cyan) and the E2 enzyme UbcH7 (green; PDB code 1fbv). The c-Cbl RING domain is colored magenta with Zn shown as yellow balls and the c-Cbl-bound ZAP-70 peptide is shown in red. The active site Cys86 of UbcH7 is shown as a yellow ball-and-stick model. (**C**) SUMO E2 enzyme Ubc9 with active site Cys93 shown as a yellow ball-and-stick model (PDB entry 1kps). (**D**) Siah homodimer (PDB entry 1k2f). The monomers are colored blue and green, respectively. The RING domains are in orange. Zinc atoms are magenta spheres. (**E**) Cul1-Rbx1-Skp1-F-box-Skp2 complex. The Cul1, Ring-Box, SKP1, and Skp2-F-box are colored in cyan, red, orange, and magenta, respectively (PDB code 1ldk). Zn ions are shown as yellow spheres. (**F**) The second cullin repeat from cul1 (PDB entry 1ldk).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

Figure 17.1.36 *(continued)*

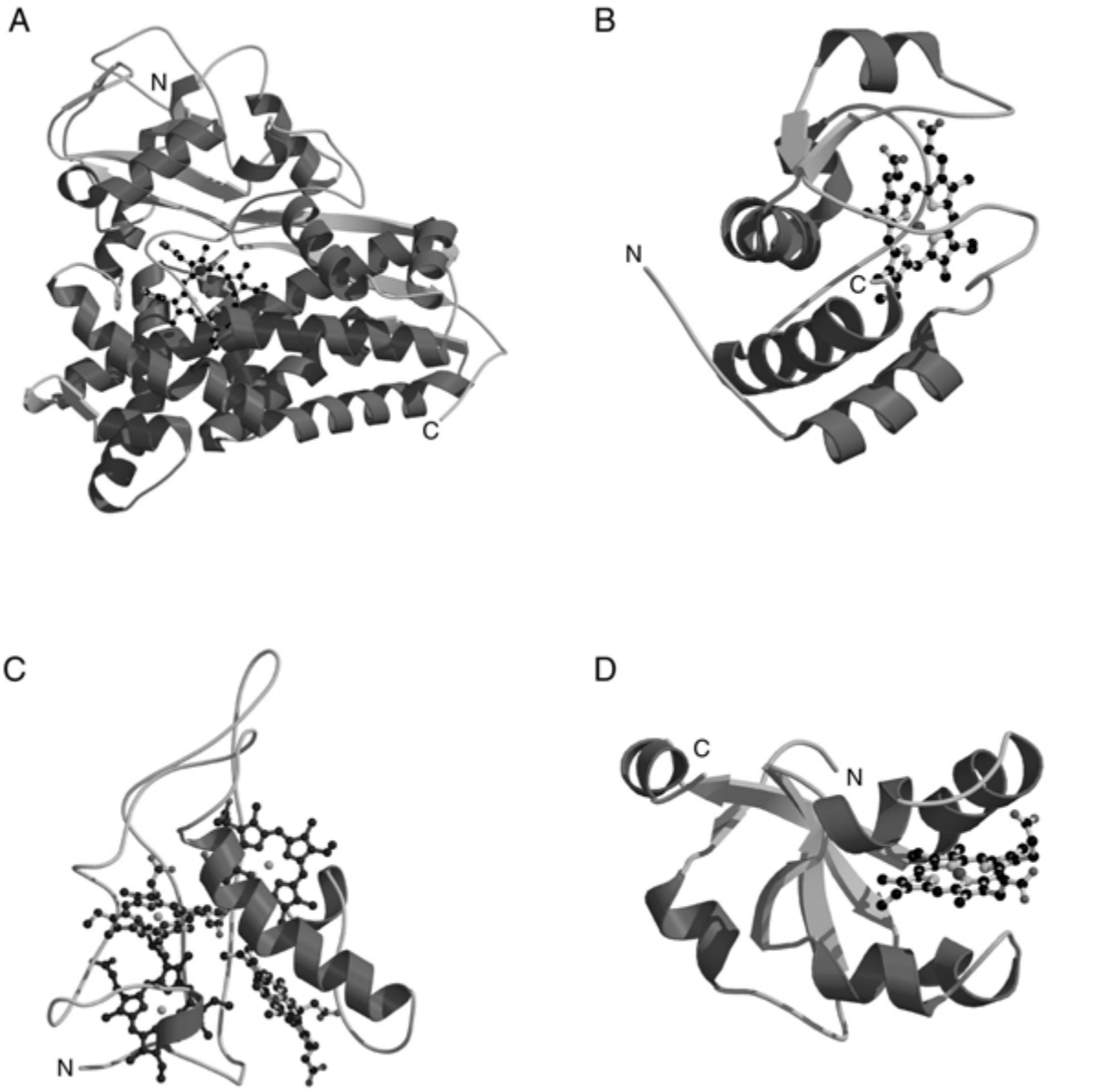


Figure 17.1.37 *(continues on next page)* Structures showing the fold of (**A**) cytochrome P450-CAM (PDB entry 1phc), (**B**) cytochrome c (PDB entry 1ycc), (**C**) cytochrome c3 (PDB entry 2cy3) with the four heme molecules colored differently, (**D**) cytochrome b5 (PDB entry 1cyo), (**E**) cytochrome b562 (PDB entry 1cgn), and (**F**) a dimer of bacterioferritin (also known as cytochrome b1; PDB entry 1bcf). The heme molecules are shown by ball-and-stick models.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

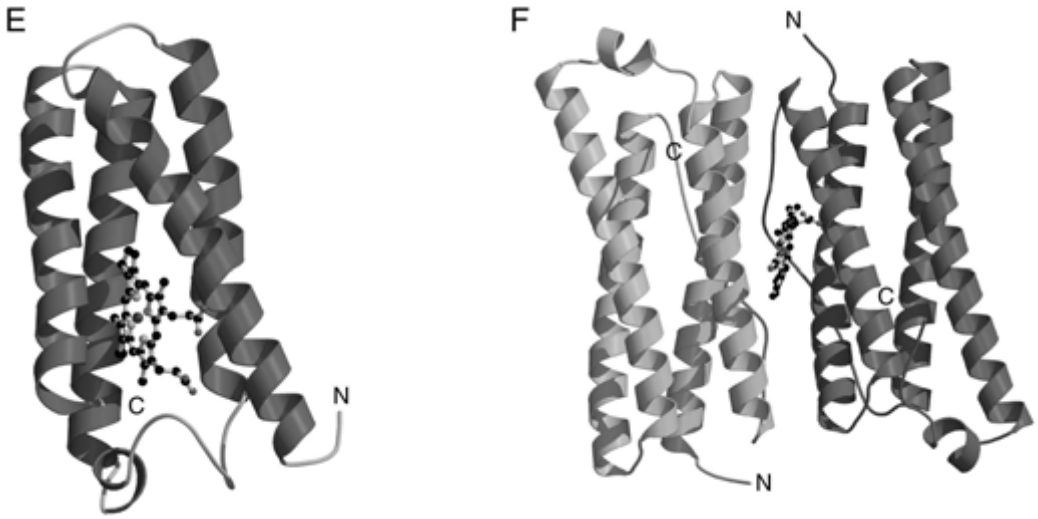


Figure 17.1.37 *(continued)*

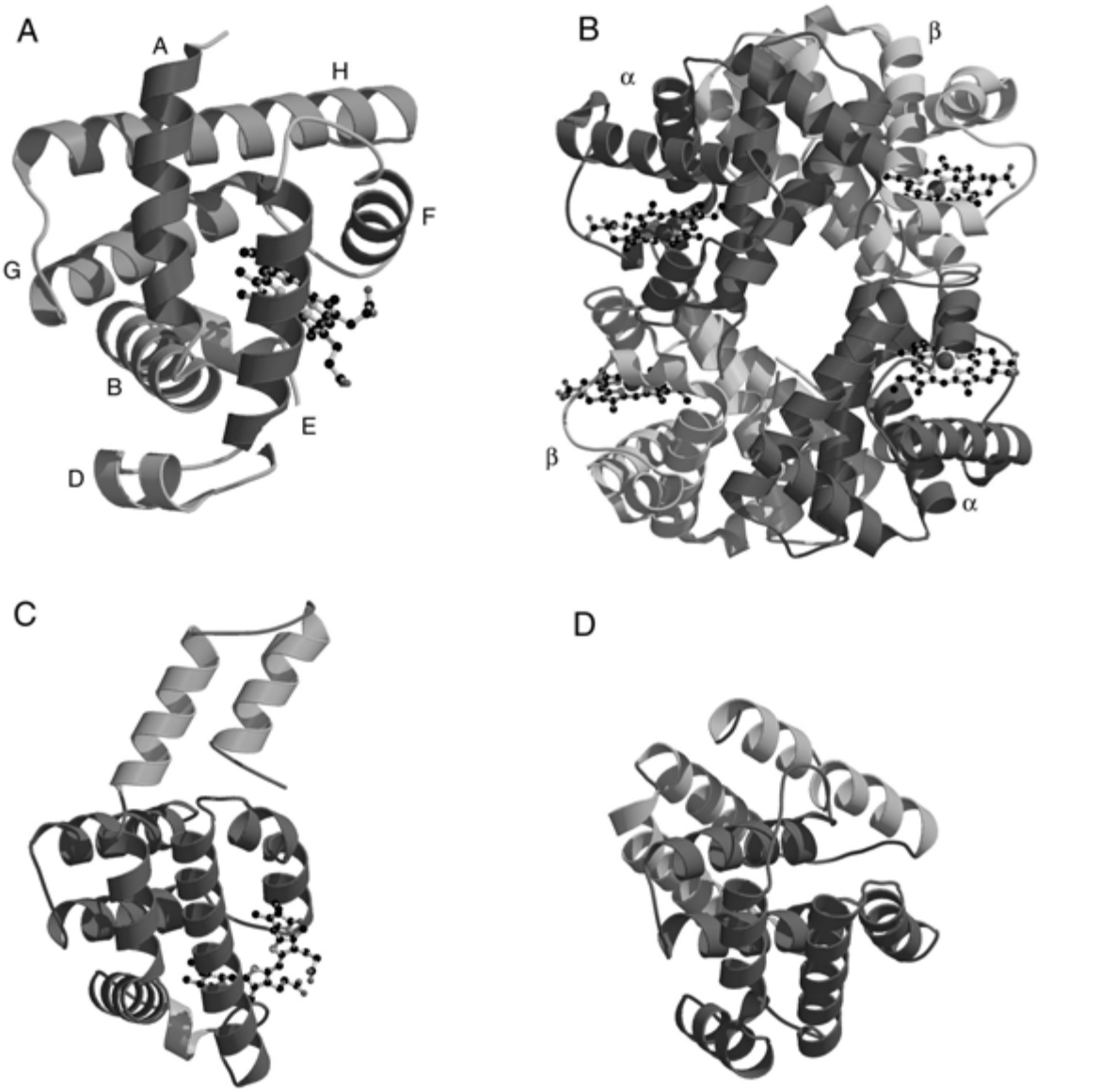


Figure 17.1.38 Structures of globin-like proteins from each of the three globin-like families. All three proteins are shown in the same orientation with respect to the globin fold. Helices that are not considered part of the core globin fold are colored green. *(***A**) Sperm whale myoglobin (PDB entry 1mbd). Helices are labeled A to H according to traditional globin nomenclature. Helices A, E, and F are colored red, while helices B, G, and H are tan. The heme group is represented by a ball-and-stick model. (**B**) Structure of human deoxyhemoglobin showing all four subunits (PDB entry 1hhb). The heme groups are represented by ball-and-stick models. The α subunits are colored red and the β subunits are colored yellow. (**C**) C-phycocyanin from cyanobacteria (PDB entry 1cpc). The phycocyanobilin cofactor is represented by a ball-and-stick model. (**D**) Colicin A from *E. coli* (PDB entry 1col).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

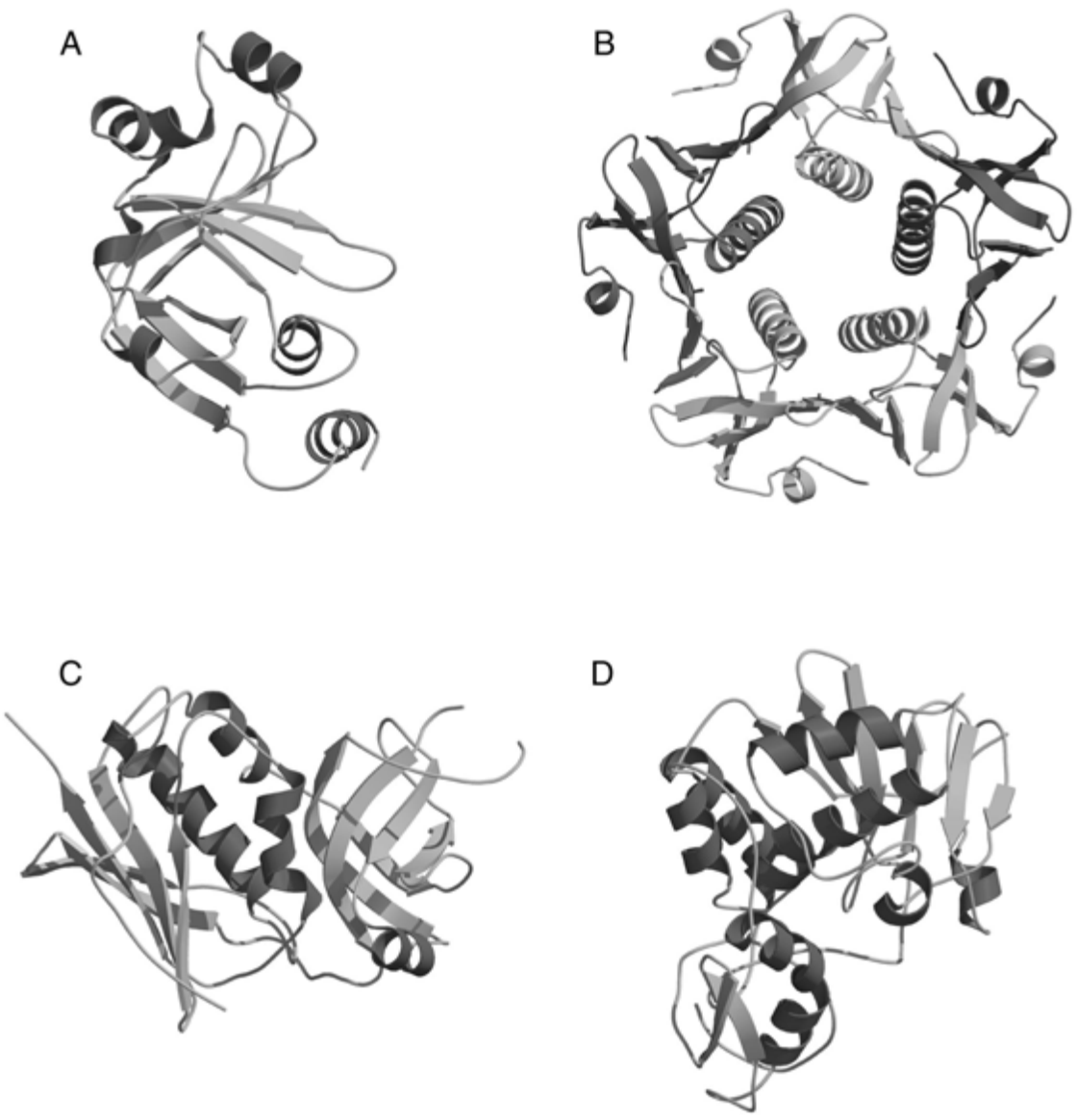


Figure 17.1.39 *(continues on next page)* Tertiary folds of toxins. *(***A**) ADP ribosylation domain of diphtheria toxin (PDB entry 1ddt). (**B**) Pentameric B domain of bacteria AB5 cholera toxin with each domain colored differently (PDB entry 1chp). (**C**) Superantigen toxin (PDB entry 1se2). (**D**) Ricin A chain (PDB entry 1rtc) The structure of (**E**) scorpion toxin (PDB entry 1mtx), (**F**) snake toxin (PDB entry 1nxb), and (**G**) spider toxin (PDB entry 1eit). Disulfide bonds are represented by yellow stick models.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

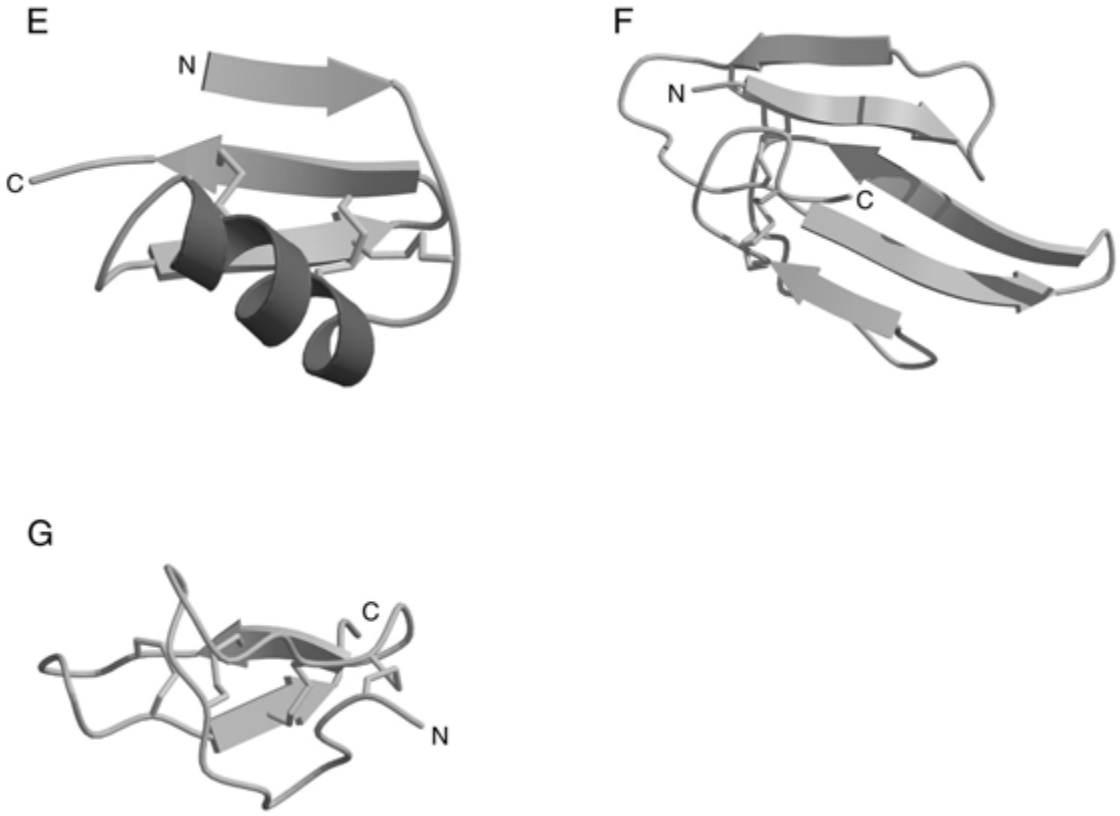


Figure 17.1.39 *(continued)*

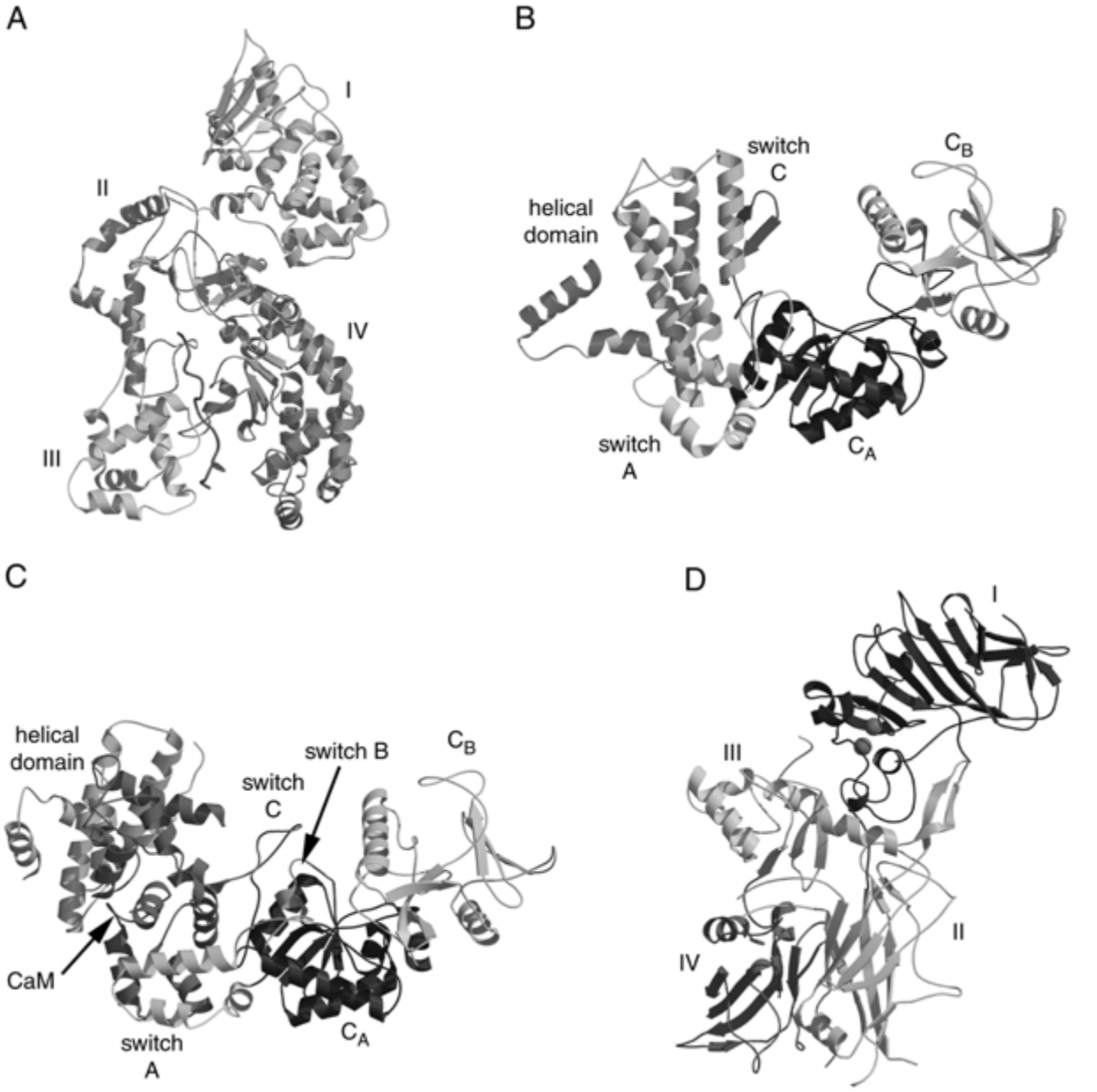


Figure 17.1.40 Structure of anthrax toxin. (**A**) Domains I through IV of the lethal factor (LF) are colored in blue, green, yellow, and orange, respectively (PDB entry 1jky). The bound MAPKK peptide is shown in red. This PDB entry does not include coordinates for the Zn atom. (**B**) The structure ofedema factor (EF; PDB entry 1k8t). The catalytic domains C_A and C_B are shown in blue and cyan, respectively. The helical domain and the switch A and C regions are shown in green, yellow, and magenta, respectively. The location of switch B is shown in the next panel. (**C**) CaM complexededema factor (PDB entry 1k90). The coloring scheme is the same as (B) with CaM in red and switch B in orange. (**D**) Domains I to IV of the protective antigen (PA) are shown in blue, green, yellow, and red, respectively (PDB entry 1acc). Ca^{2+} ions are shown as magenta spheres.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

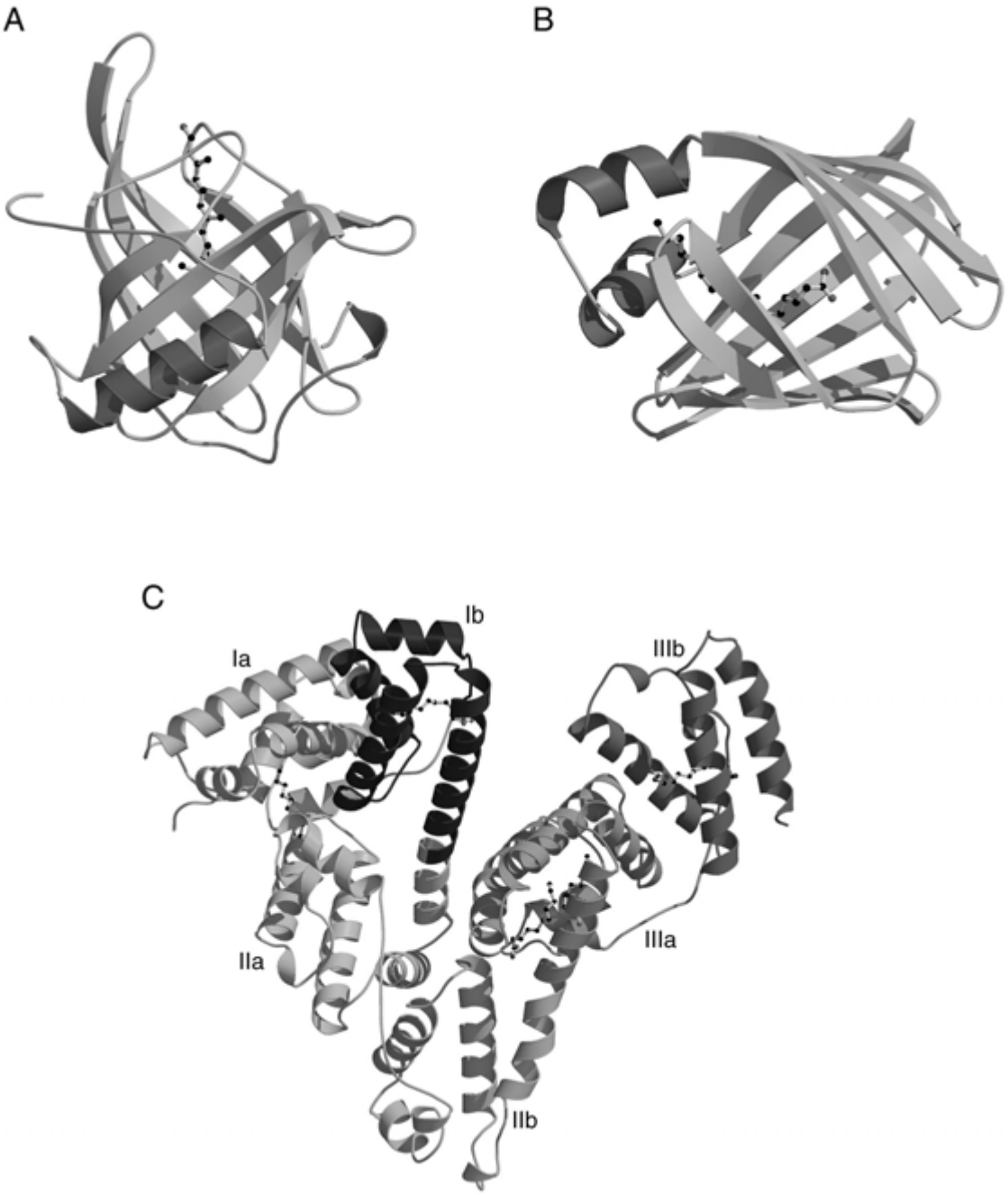


Figure 17.1.41 Lipid-binding proteins with respective ligands shown as ball-and-stick models. (**A**) Human retinol-binding protein bound to retinol (PDB entry 1rbp), an example of a lipocalin. (**B**) Rat intestinal fatty acid-binding protein bound to palmitate (PDB entry 2ifb). (**C**) Human serum albumin with five bound myristate molecules (PDB entry 1bj5). Note that a helix spans both the I/II and II/III domain boundaries. This results in 28 instead of 30 helices in the structure.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

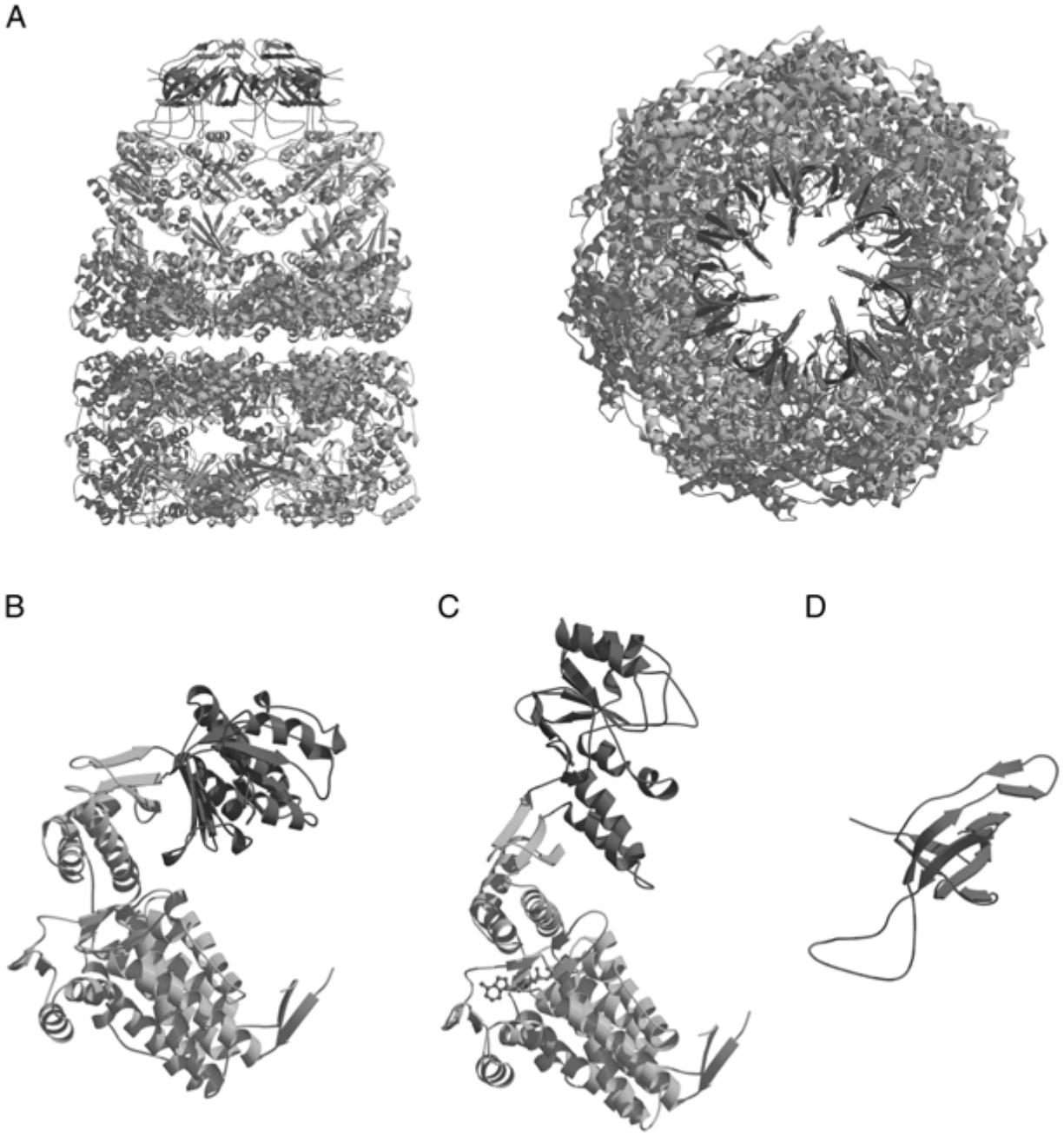


Figure 17.1.42 *(continues on next page)* Chaperonin and myosin structures. (**A**) The GroEL/GroES complex viewed from the side (left) and top (right; PDB entry 1aon). The GroEL *trans* ring is green, the *cis* ring is blue, and the GroES ring is magenta. The domain structure of GroEL in the (**B**) *trans* and (**C**) *cis* rings of the GroEL/GroES complex (PDB entry 1aon). The equatorial domains are blue, the intermediate domains are green, and the apical domains are red. ADP is shown as a ball-and-stick model. (**D**) The GroES subunit (PDB entry 1aon). (**E**) The structure of the scallop myosin subfragment S1 (PDB entry 1b7t). Bound Ca^{2+} is shown as a small purple sphere. The heavy-chain domain is shown in green. The essential and regulatory light chains are colored blue and red, respectively.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

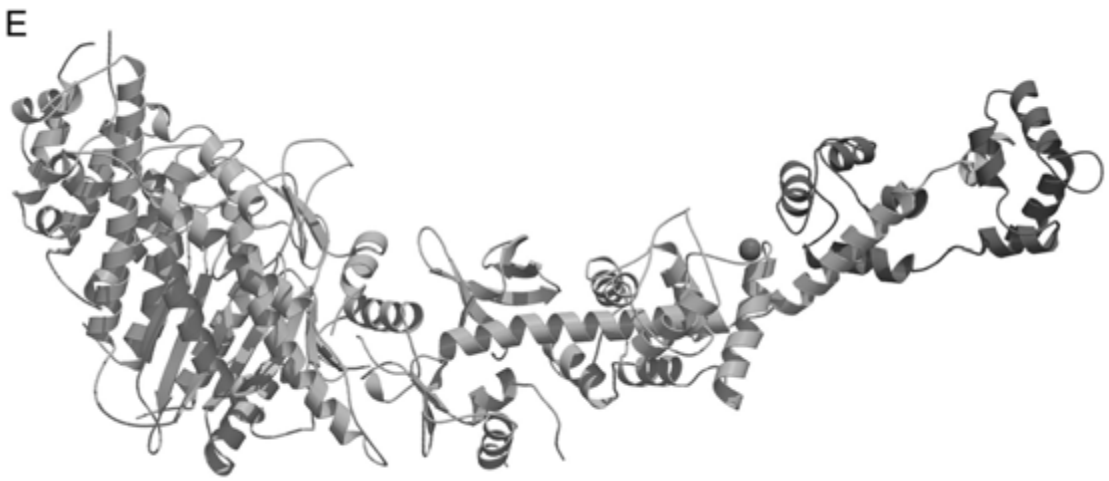


Figure 17.1.42 *(continued)*

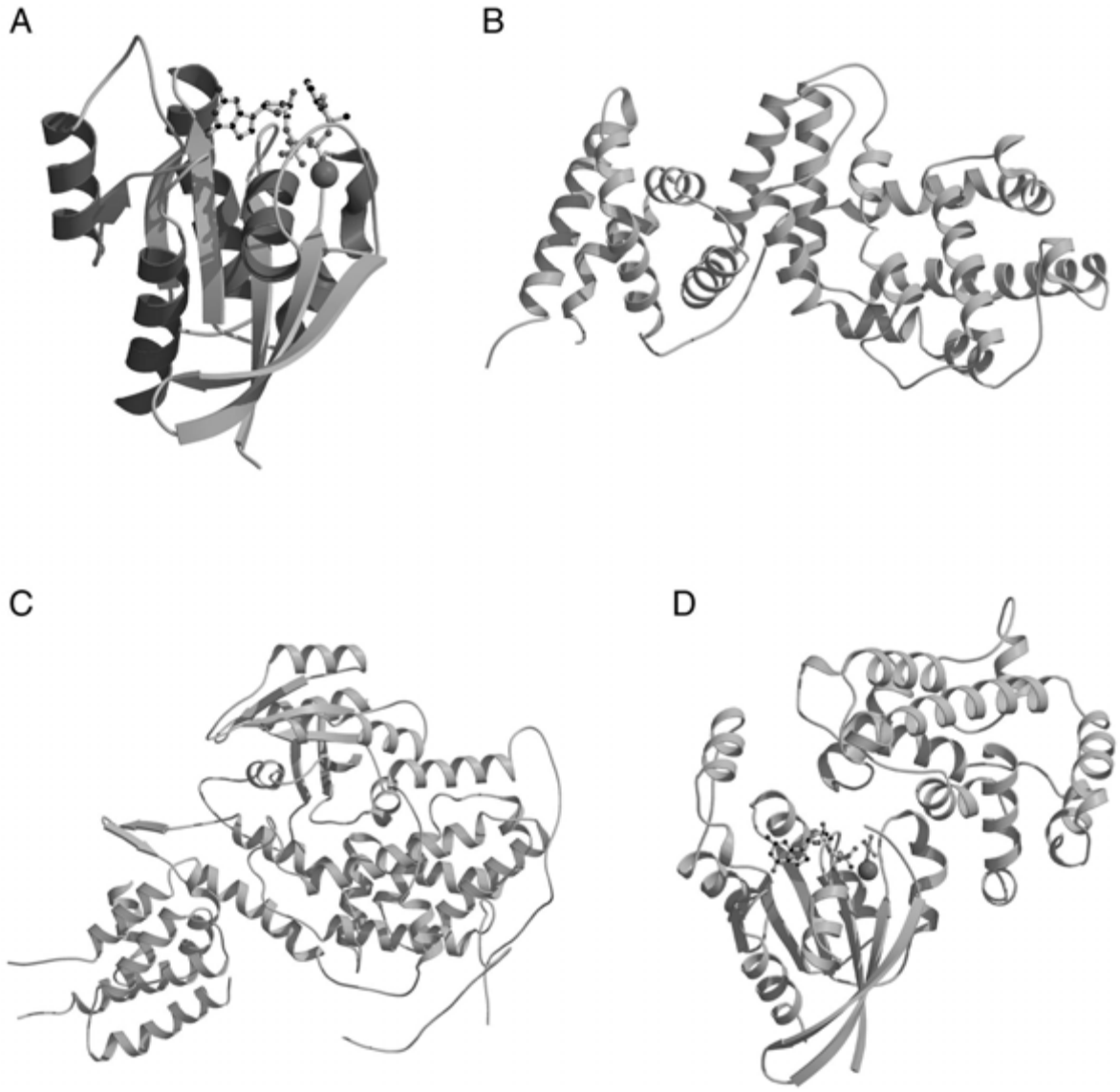


Figure 17.1.43 *(continues on next two pages)* G-protein and its regulators. The bound Mg^{2+} and GDP are shown as a magenta sphere and ball-and-stick model. (**A**) Structure of p21Ras (PDB entry 1gnr). A nonhydrolyzable GTP analog is shown as a ball-and-stick model. (**B**) The structure of the Ras GTPase activation domain of a human p120GAP (PDB entry 1wer) . (**C**) The structure of Ras in complex with the Ras guanine-nucleotide-exchange-factor region of Sos (PDB entry 1bkd). Ras and Sos are colored in green and blue, respectively. (**D**) The structure of RhoA (green) in complex with RhoGAP (blue; PDB entry 1tx4). GDP and AlF4 are shown as ball-and-stick models. (**E**) The structure of GBP1 (PDB entry 1f5n). (**F**) Side and (**G**) top view of the structure of a heterotrimeric G-protein complex (PDB entry 1got). The $G_{t\alpha}$-$G_{i\alpha}$ chimera is in red, $G_{\tau\beta}$ is green, and $G_{\tau\gamma}$ subunit is blue. (**H**) The structure of RGS4 (blue) in complex with $G_{i\alpha}$-GDP-AlF_4^- (green; PDB entry 1agr). (**I**) The crystal structure of a $G_{\alpha i}$-GDP (green and yellow) bound to the GoLoco region of RGS14 (blue; PDB entry 1kjy). GDP is shown as a ball-and-stick model.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

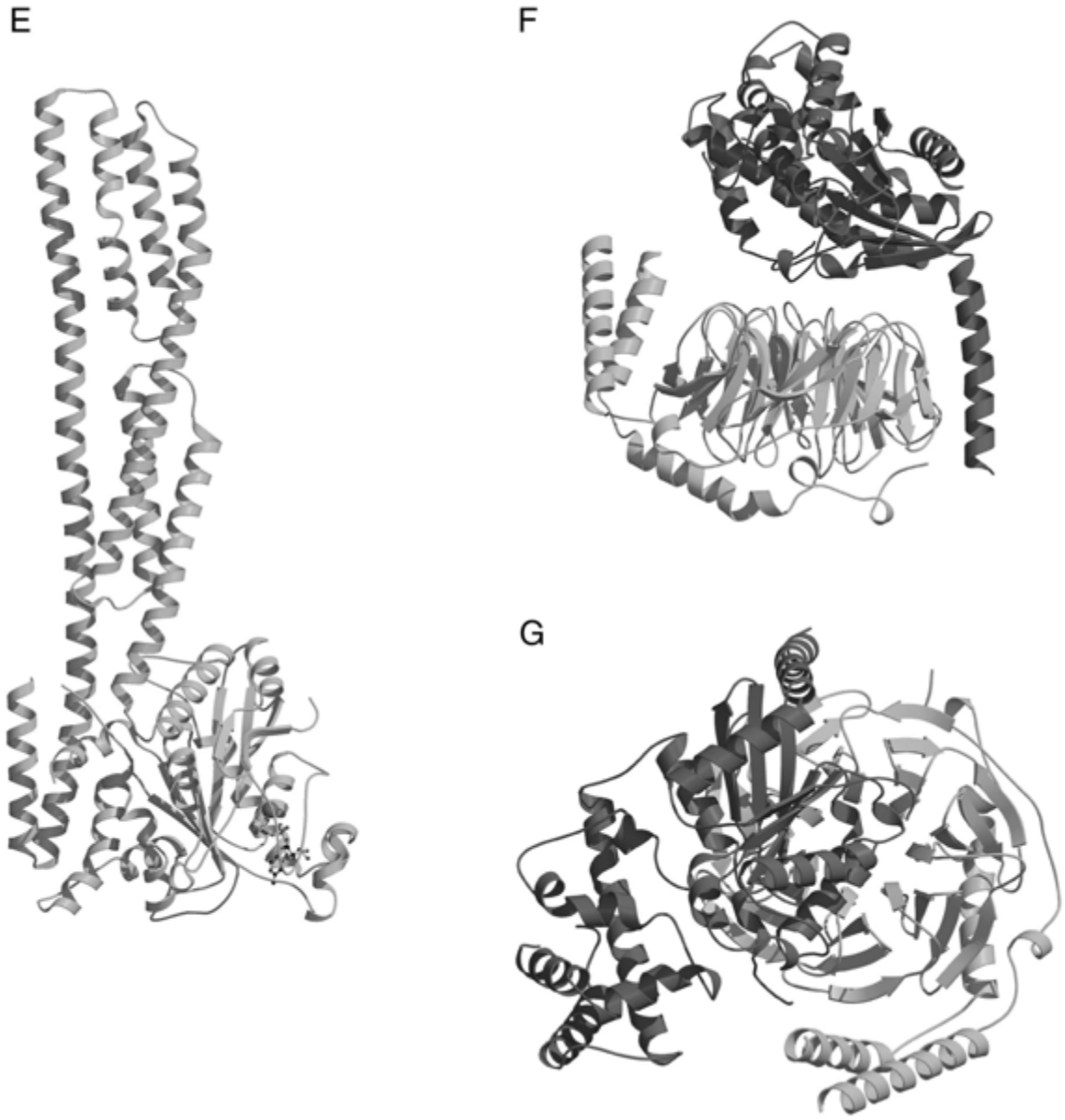


Figure 17.1.43 *(continued)*

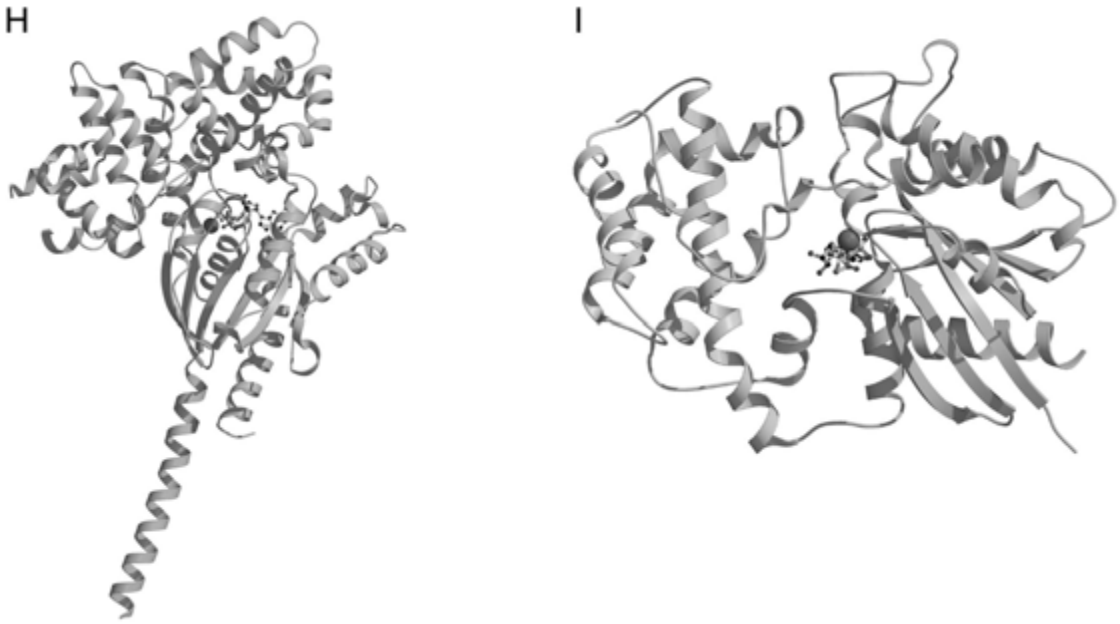


Figure 17.1.43 *(continued)*

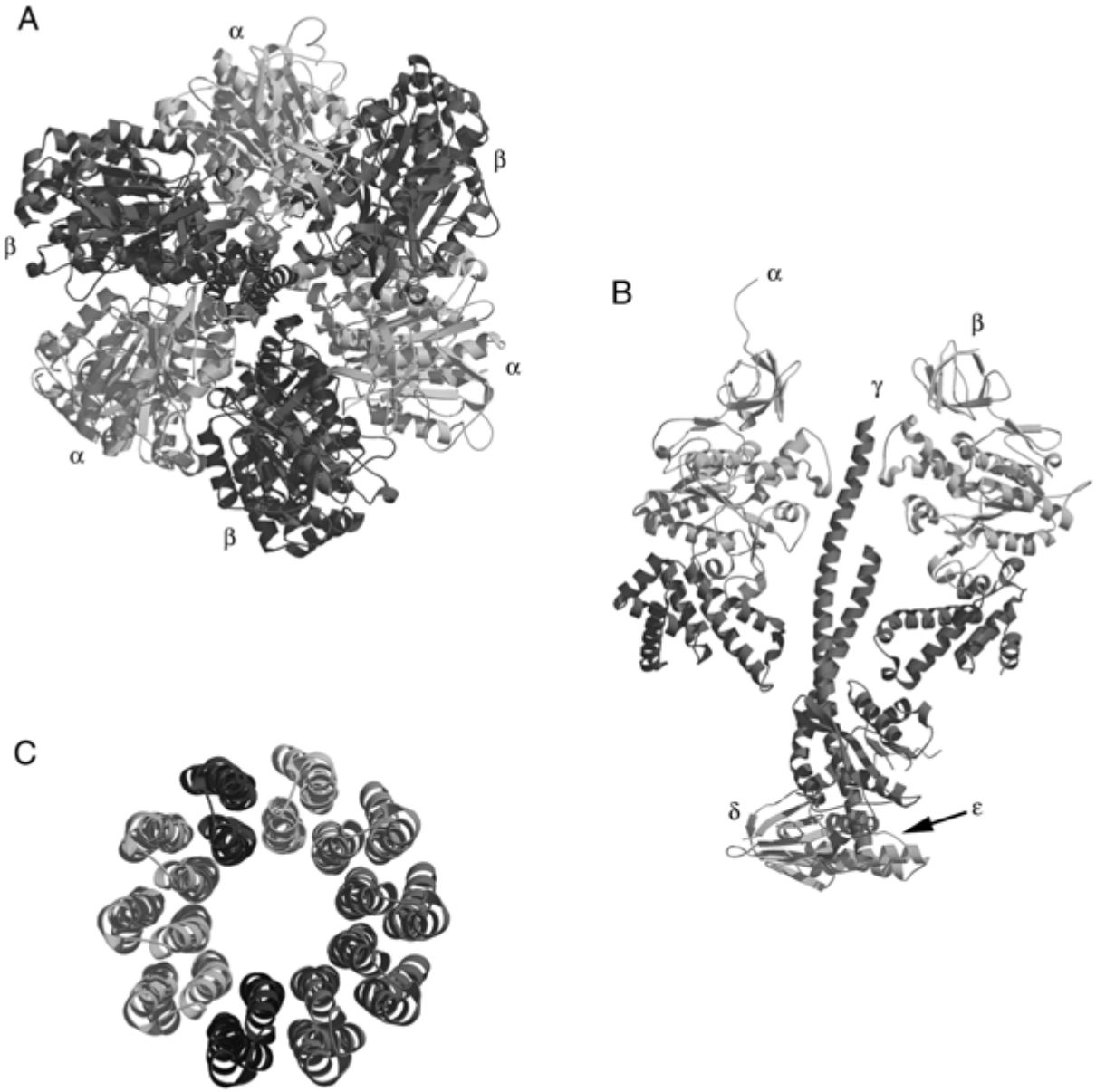


Figure 17.1.44 Structure of the F_1ATPase. (**A**) Top view of bovine F_1ATPase α (yellow), β (red), and γ (magenta, N- and C-terminal helices only) subunits (PDB entry 1e76). (**B**) Side view of bovine mitochondrial F_1ATPase showing one α subunit (left), one β subunit (right), and the γ (magenta, center), δ (cyan, bottom), and ε (green, bottom) subunits. The α and β subunits are each color coded by domain: N-terminal β-barrel is cyan, the central nucleotide-binding domain is green, and the C-terminal α-helical bundle is red. (**C**) Cα model of ten c subunits from a yeast F_0ATPase membrane domain (PDB entry 1qo1). Each subunit is a long α-helical hairpin.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

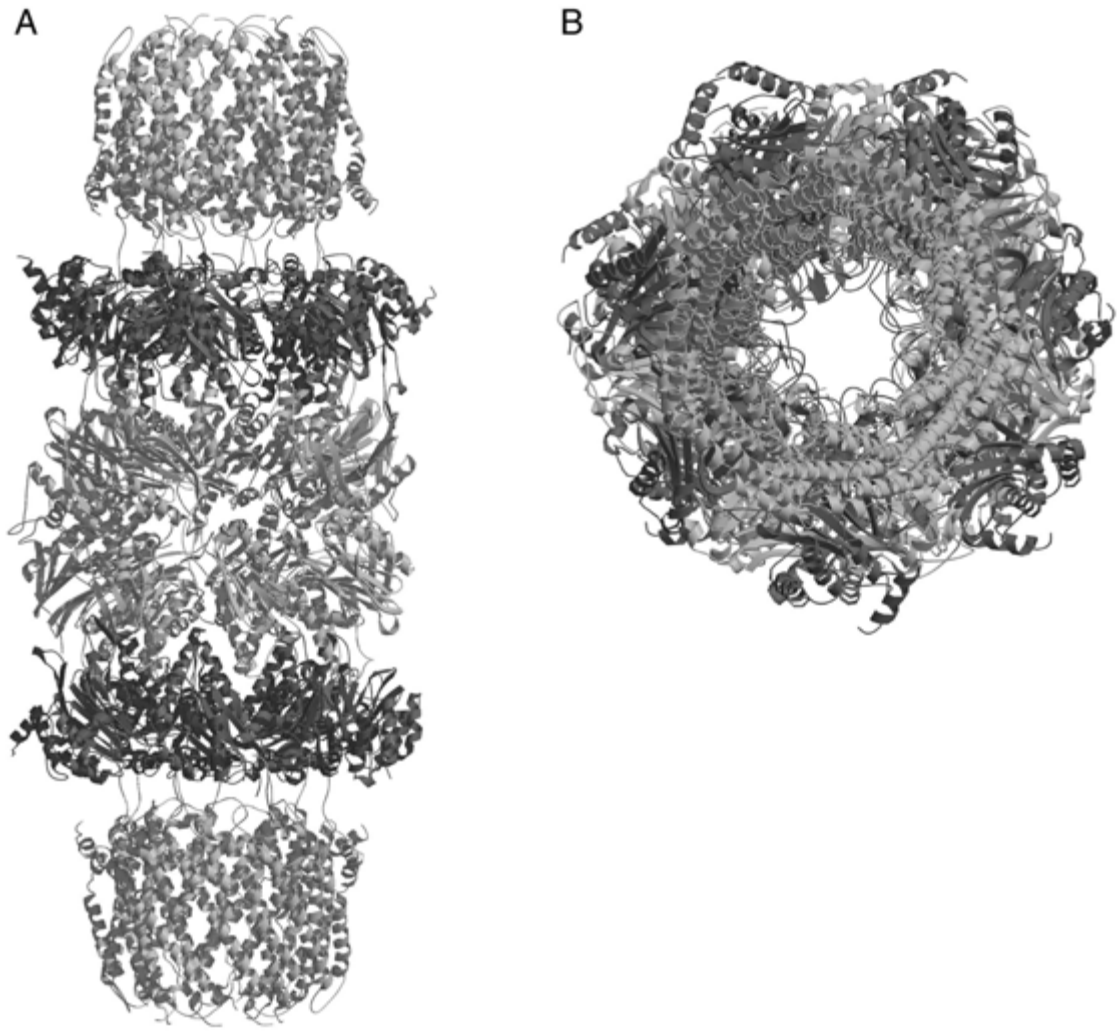


Figure 17.1.45 The 20S proteasome from yeast complexed with the 11S regulator from *T. brucei* (PDB entry 1fnt). The α, β, and regulator subunits are colored red, green, and blue, respectively. (**A**) Side view of the barrel-shaped complex. (**B**) View of the complex looking down the axis of the barrel.
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

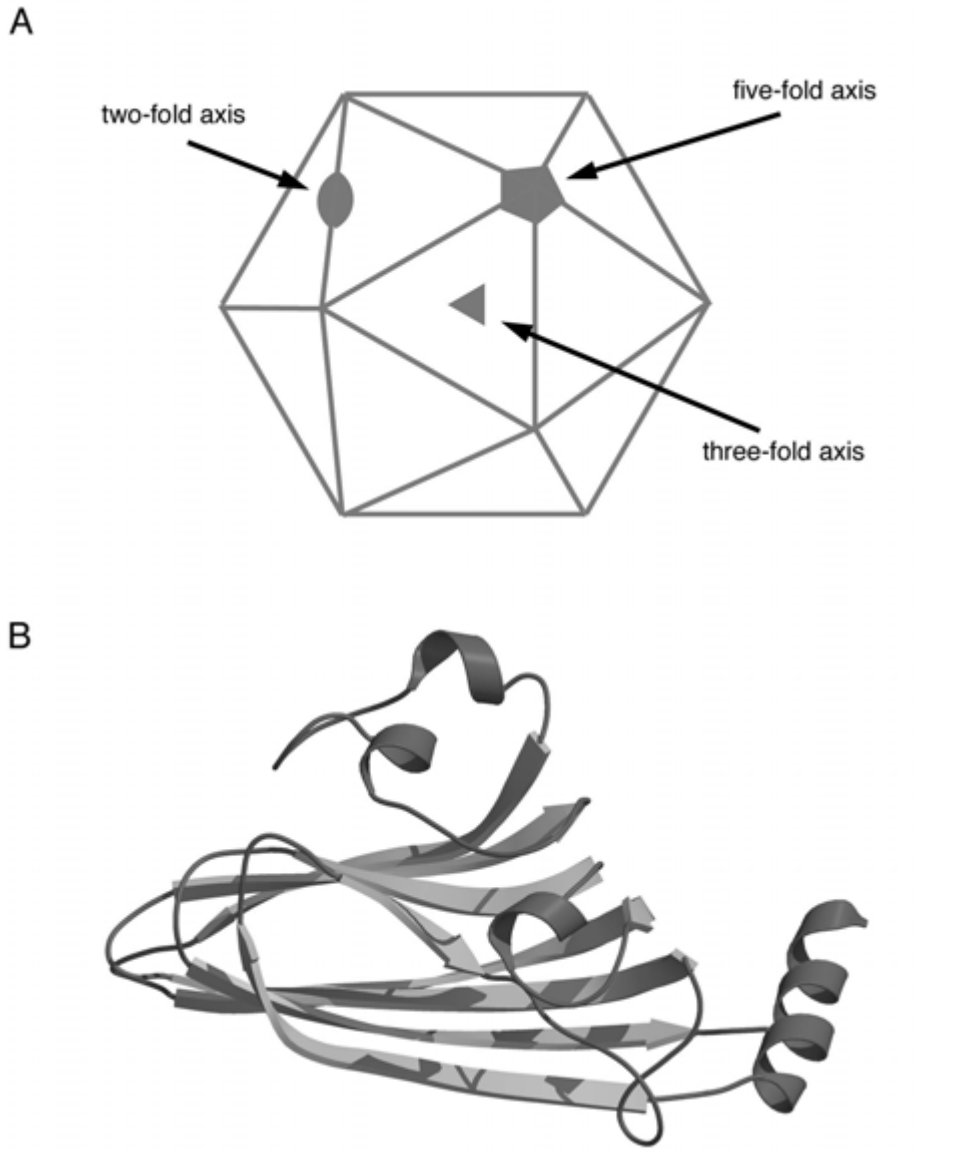


Figure 17.1.46 Viral capsid proteins. (**A**) Illustration of an icosahedron showing two-, three-, and five-fold symmetry. (**B**) Structure of the jelly-roll β-sandwich fold of satellite tobacco necrosis virus-coat protein (PDB entry 2stv).
(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

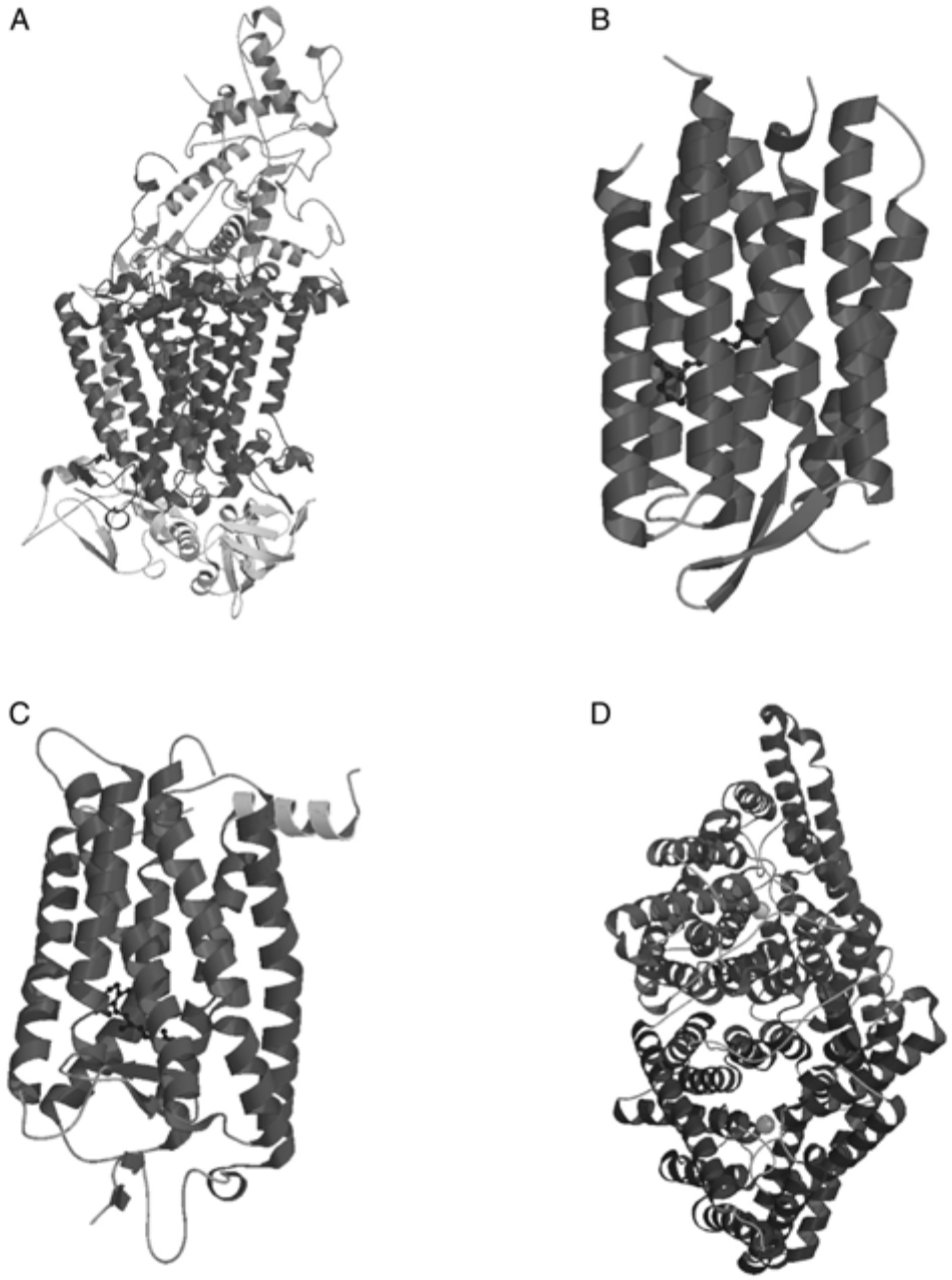


Figure 17.1.47 *(continues on next two pages)* Structures of integral membrane proteins. *(***A**) Photosynthetic reaction center (PDB entry 1prc). The four protein subunits are shown: cytochrome (gray), M (green), L (red), and H (yellow). (**B**) The structure of *Halobacterium salinarum* bacteriorhodopsin (PDB entry 1c3w), with the cytoplasmic face at the top. The retinal chromophore is shown as a blue ball-and-stick model. (**C**) Rhodopsin from the outer segment of bovine rod photoreceptor cells (PDB entry 1f88). The receptor is shown with the C-terminal (cytoplasmic) domain at the top, and the N-terminal (extracellular) domain at the bottom of the diagram. The covalently linked eleven-*cis*-retinal ligand is shown as a blue ball-and-stick model. The additional eighth helix is colored in yellow. (**D**) The ClC chloride channel from *Salmonella typhimurium* (PDB entry 1kpl). The double barrel of the homodimer is shown with Cl^- ions, visible in the pores, displayed as green spheres. View is perpendicular to the plane of the membrane with the two subunits colored red and blue. (**E**) The *E. coli* vitamin B_{12} transporter BtuCD. The dimer is depicted, with BtuC subunits in red and BtuD subunits in blue (PDB entry 1l7v). (**F**) Calcium ATPase from skeletal-muscle sarcoplasmic reticulum in the E1 state with bound calcium ions displayed as orange spheres. The three cytoplasmic domains are at the top (PDB entry 1eul). (**G**) Bovine AQP1 water channel monomer (PDB entry 1j4n). View is from the plane of the membrane with the extracellular face at the top. The two membrane-inserted helices that do not span the membrane are colored in blue. (**H**) Bovine cytochrome bc_1complex (PDB entry 1bgy). The subunits with membrane-spanning portions are colored in the following scheme. Cytochrome *b* is red, cytochrome c_1 is aqua, Rieske iron-sulfur protein is blue, subunit 7 is orange, subunit 10 is green, and subunit 11 is yellow. (**I**) The *E. coli* ferric enterobactin receptor FepA (PDB entry 1fep). The β-barrel domain is shown in green and the N-terminal plug domain is shown in red. (**J**) TolC outer membrane protein of *E. coli* (PDB entry 1ek9). The trimer is shown with the subunits colored separately.

(For full-color version of figure go to http://www.interscience.wiley.com/c_p/colorfigures.htm.)

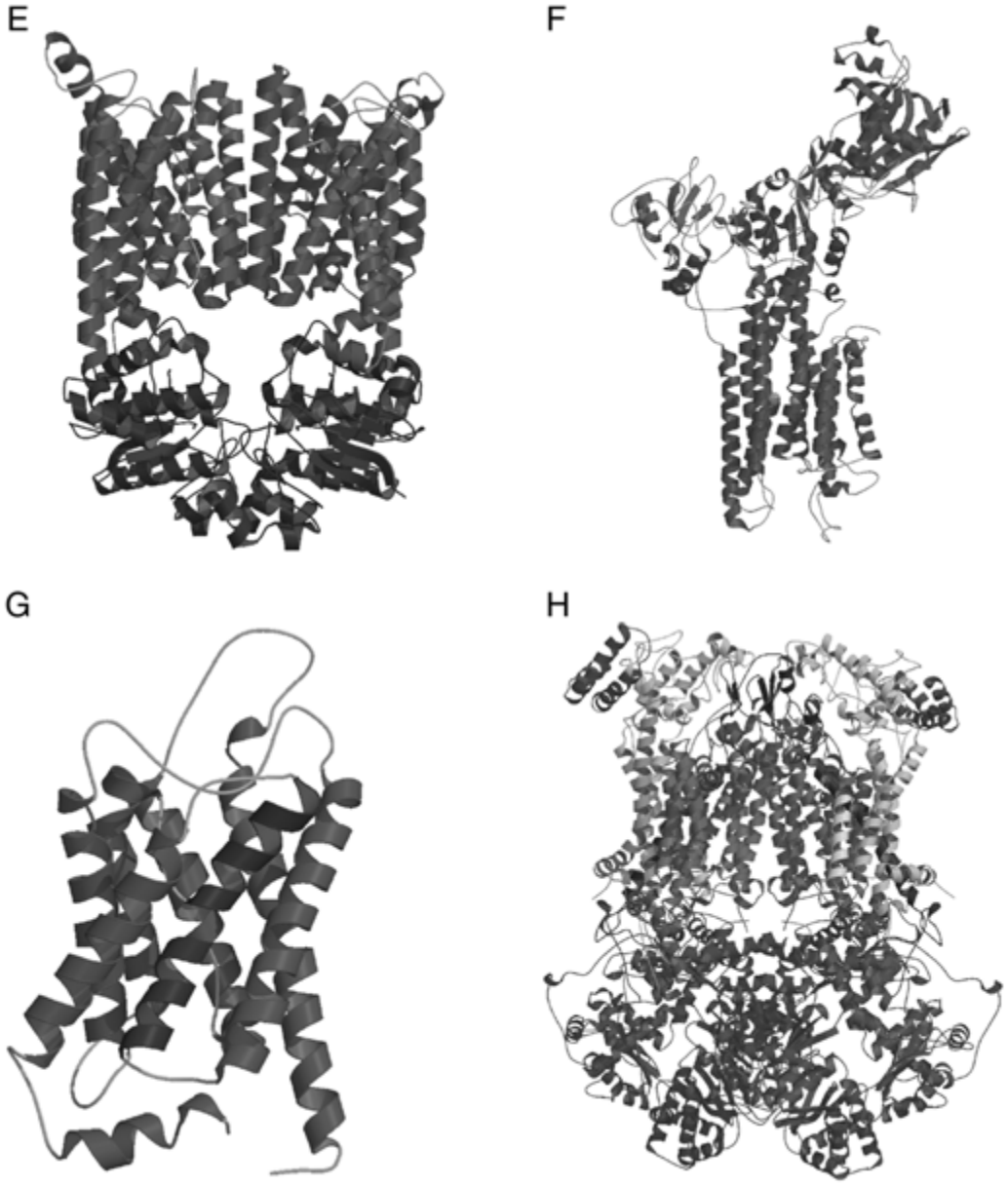


Figure 17.1.47 *(continued)*

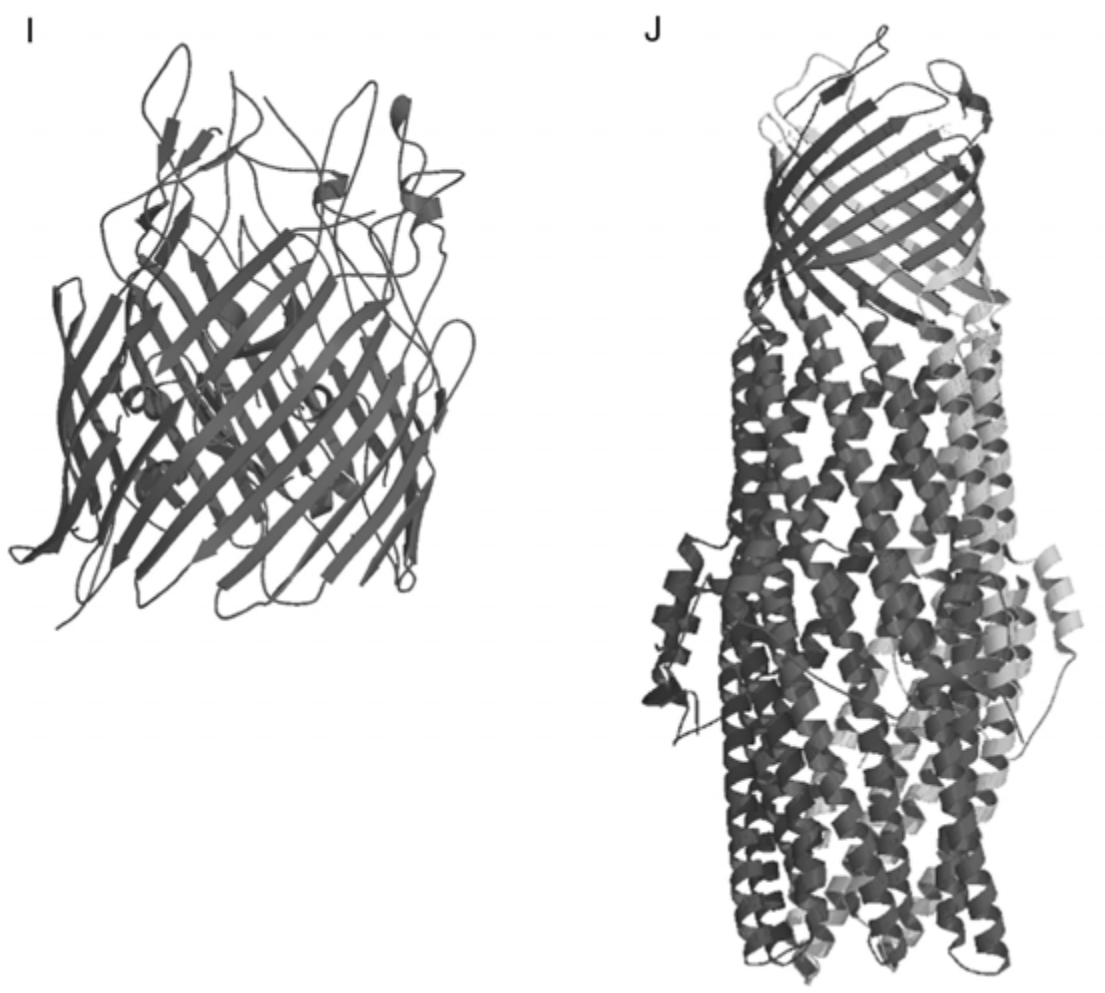


Figure 17.1.47 *(continued)*