III. Classification Of Proteins By Patterns Of Tertiary Structure

A. Summary of the Classification System

1. Principles and Methods

Having looked at the characteristics of individual structural features and some of their local combinations, we are now in a position to sort out and classify the major structural patterns, or “folds,” that make up entire proteins. This classification will build on earlier work by Rossmann (e.g., Rossmann and Argos, 1976), Richardson (1977), and Levitt and Chothia (1976), but will attempt to combine and extend those systems, as well as including the newer structures now available.

In 2006, the major resources for classification of protein structures are SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/) (Murzin, 1995) and CATH (http://www.cathdb.info/latest/index.html) (Orengo, 1997) which use major categories similar to the ones here: a lowest “family” level from sequence homology, and distinct versions of “fold”-based categories in between. Both are organized by domains. Further notes on the many new types of folds seen since 1981 will be found in the later sections.

The most useful level at which to categorize protein structures is the domain, since there are many cases of multiple-domain proteins in which each separate domain resembles other entire smaller proteins. We have separated proteins into domains on the basis of whether the pieces could be expected to be stable as independent units or are analogous to other complete structures (see Section II,1 for a fuller explanation). Clearly demonstrated homologous families such as trypsin, chymotrypsin, elastase, and the S. griseus proteases, or the cytochromes c, c_2, c_{550}, c_{551}, and C_{555}, are treated as single examples. There are between 90 and 105 different domains represented in the current sampling of known protein structures, depending on how one counts the cases of similar domain structures within a given protein. In the schematic drawings of Figs. 72-86 such domains are illustrated separately only if they are at least as different as the range of variation common within the close homology families (and, of course, if suitable coordinates or stereos were available). The domains within each protein are distinguished by numbers in sequence order (e.g., papain d1 and papain d2), except for the immunoglobulins for which we use the standard terminology (V_L, C_H 1, etc.) for constant and variable domains.

Structural categories are assigned primarily on the basis of the type and organization of secondary-structure elements, the topology of their connections, and the number of major layers of backbone structure that are present. Since proteins fold to form a protected hydrophobic core of side chains on the interior, the simplest type of stable protein structure consists of polypeptide backbone wrapped more or less uniformly around the outside of a single hydrophobic core. We will describe such a structure as “two layer,” because a line from the solvent through the center of the protein and back out again would pass through two principal layers of backbone structure (see Fig. 71a). Over half of the known domain structures have two layers. About a third of the structures have three layers of backbone and two hydrophobic cores; the commonest such type has a central β sheet layer flanked by two helical layers (see Fig. 71b). There are three known four-layer domains (e.g., Fig. 71c) and one five-layer. Isolated loops that curl over the outside are not considered to form a distinct layer. Although there are some ambiguous cases (especially in the very small structures) they are less common than one might expect, presumably because the requirements for rapid folding rule out much tangling or recrossing of the backbone.

The approximately 100 distinctly different domains fall into four broad categories, each of which has several sub-groupings. The four broad categories are (I) antiparallel α; (II) parallel α/β; (III) antiparallel β; (IV) small SS-rich or metal-rich. The major determinant in assigning a domain to one of these categories is not just the percentage of a given secondary structure, but whether that type of secondary structure forms the central core and whether its interactions could be the dominant stabilizing ones. The two β categories are the most populous, with 30 to 35 members each; there are about 20 α-helical domains and a dozen of the small proteins. The overall classification scheme is summarized in Section III, A,2. An alphabetical index of the pro-
teins is given in Section III,A,4, with domain assignments, structure subcategories, and literature references. Obviously this classification is not the only plausible way to categorize protein structures. Indeed, for some of the individual cases there are other descriptions which would be preferable because they emphasize possible relationships to functionally similar proteins. However, the motivation here has been to achieve the most satisfactory compromise for all the structures: to fit as many examples into as few groupings as possible, while retaining enough detail to provide meaningful descriptions. Also, the prejudice has been in favor of grouping together domains whose entire structure is approximately the same rather than cases in which a relatively small portion of both structures is more exactly similar.

In order to illustrate this taxonomic system and to facilitate contrasts and comparisons among the structures, schematic backbone drawings have been made of most of the known structures. The drawings are grouped together by categories in Section III,A,3. α-Carbon coordinates

FIG. 71. Examples of protein domains with different numbers of layers of backbone structure: (a) two-layer cytochrome c`; (b) three-layer phosphoglycerate kinase domain 2; (c) four-layer triosephosphate isomerase. The arrows above each drawing point to the backbone layers.
were displayed in stereo on Richard Feldmann’s computer graphics system at NIH; a suitable view was chosen (consistent for each subcategory of structure), and plotter output was obtained at a consistent scale (approximately 20 Å per inch on the final drawings as reproduced here). The schematic was drawn on top of the plotter output for accuracy, with continual reference to the stereo for the third dimension. Loops, and to some extent β strands, were smoothed for comprehensibility, and shifts of 1 or 2 Å were sometimes necessary in order to avoid ambiguity at crossing points. A uniform set of graphical conventions was adopted (see Section III,A,3 for explanation) in which β strands are shown as arrows, helices as spiral ribbons, and nonrepetitive structure as ropes. Location and extent of β strands and helices are sometimes based on published descriptions and hydrogen-bonding diagrams, but often must be judged from the stereo view itself. Very short β interactions are shown as arrows when they form part of a larger sheet but may be left out if they are isolated. Foreshortening, overlaps, edge appearance, and relative size change are used to provide depth cues.

2. Outline of the Taxonomy
   I. Antiparallel α domains
      A. Up-and-down helix bundles
         Myohemerythrin, hemerythrin
         Cytochrome b5,62
         Cytochrome c’
         Uteroglobin
         Staphylococcal protein A fragment
         Influenza virus hemagglutinin “domain” around 3-fold
         Tobacco mosaic virus protein
         Cytochrome b3
         Tyrosyl-tRNA synthetase domain 2
         Ferritin (?) [yes]
         Purple membrane protein (?) [yes]
      B. Greek key helix bundles
         Myoglobin, hemoglobin
         Thermolysin domain 2
         T4 phage lysozyme domain 2
         Papain domain 1
         Cytochrome c peroxidase domain 1
      C. Miscellaneous antiparallel α
         Carp muscle calcium-binding protein
         Egg lysozyme
         Citrate synthase
         Catalase domain 2
         Cytochrome c peroxidase domain 2
         p-Hydroxybenzoate hydroxylase domain
   II. Parallel α/β domains
      A. Singly wound parallel β barrels
         Triosephosphate isomerase
         Pyruvate kinase domain 1
         KDPG aldolase [yes]
      B. Doubly wound parallel β sheets
         1. Classic doubly wound β sheets
            Lactate dehydrogenase domain 1
            Alcohol dehydrogenase domain 2
            Aspartate transcarbamylase catalytic domain 2
            Phosphoglycerate kinase domain 1
            Tyrosyl-tRNA synthetase domain 1(?)
            Phosphorylase domain 2, central three layers
2. Doubly wound variations
Glyceraldehyde-phosphate dehydrogenase domain 1
Phosphorylase domain 1, central three layers
Flavodoxin
Subtilisin
Arabinose-binding protein domains 1 and 2
Dihydrofolate reductase
Adenylate kinase
Rhodanese domains 1 and 2
Glutathione reductase domains 1 and 2
Phosphoglycerate mutase
Phosphoglycerate kinase domain 2
Pyruvate kinase domain 3
Hexokinase domains 1 and 2
Catalase domain 3
Aspartate aminotransferase
Aspartate transcarbamylase catalytic domain 1
Phosphofructokinase domain 1
p-Hydroxybenzoate hydroxylase domain 1
Glucosephosphate isomerase domain 1
Glutathione peroxidase

C. Miscellaneous parallel $\alpha/\beta$
Carboxypeptidase
Thioredoxin
Carbonic anhydrase
Phosphofructokinase domain 2
Glucosephosphate isomerase domain 2

III. Antiparallel $\beta$ domains
A. Up-and-down $\beta$ barrels
Papain domain 2
Soybean trypsin inhibitor
Catalase domain 1

B. Greek key $\beta$ barrels
1. Simple Greek keys
Trypsin-like serine proteases domains 1 and 2
Pyruvate kinase domain 2
Prealbumin
Plastocyanin, azurin
Immunoglobulin, variable and constant domains
Cu,Zn superoxide dismutase
Staphylococcal nuclease

2. “Jellyroll” Greek keys
Tomato bushy stunt virus protein domains 2 and 3
Southern bean mosaic virus protein
Concanavalin A
Influenza virus hemagglutinin HA1
$\gamma$-Crystallin domains 1 and 2

C. Multiple, partial, and other $\beta$ barrels
Acid proteases domains 1 and 2
Alcohol dehydrogenase domain 1
Pancreatic ribonuclease

D. Open-face $\beta$ sandwiches
T4 lysozyme domain 1
Aspartate transcarbamylase regulatory domains 1 and 2
Streptomyces subtilisin inhibitor
Glutathione reductase domain 3
Thermolysin domain 1
Glyceraldehyde-phosphate dehydrogenase domain 2
Bacteriochlorophyll protein
p-Hydroxybenzoate hydroxylase domain 2
Influenza virus hemagglutinin HA2
L7/L12 ribosomal protein

E. Miscellaneous antiparallel β
   Gene 5 protein, E. coli
   Lactate dehydrogenase domain 2
   Tomato bushy stunt virus protein “domain” 1

IV. Small disulfide-rich or metal-rich domains

A. SS-rich
   1. Toxin-agglutinin fold
      Erabutoxin, cobra neurotoxin
      Wheat germ agglutinin domains 1, 2, 3, and 4
   2. Other SS-rich
      Pancreatic trypsin inhibitor
      Insulin
      Phospholipase A2
      Crambin

B. Metal-rich
   1. Up-and-down ligand cages
      Rubredoxin
      Cytochrome b5
      Cytochrome c
      Cytochrome c3
   2. Greek key ligand cages
      Ferredoxin
      High-potential iron protein

3. Schematic Drawings of the Protein Domains by Structure Type

Detailed discussions of the categories and subgroupings are given in Sections III.B through III.E. The scale of these drawings (Figs. 72-86) is approximately 20Å to the inch. β strands are shown as arrows with thickness, helices as spiral ribbons, and nonrepetitive structure as ropes. Disulfides are shown as “lightning bolts.” Circles represent metals, and some prosthetic groups are shown as atomic skeletons, but not for all cases in which they are known to be present. A question mark in the label means that backbone connectivity is uncertain in some places. Where needed for clarity, the N-terminus of the domain is indicated by a small arrow; for a few two chain domains the C-terminus is indicated as well. [More detailed information on how the drawings were made is available in a Methods of Enzymology section (J.Richardson, 1985) and in a historical retrospective (J.Richardson, 2000 Nature Str. Biol.), and some related issues about structure representation in a review lecture (Richardson, 1992 Biophys. J.).]
Fig. 72. Antiparallel $\alpha$: up-and-down helix bundles.
Fig. 72-2. Antiparallel α: up-and-down helix bundles.
FIG. 73. Antiparallel α: Greek key helix bundles.
**Fig. 74. Antiparallel α: miscellaneous.**

Carp Muscle Calcium-binding Protein

Egg Lysozyme
**Fig. 75.** Parallel α/β: singly wound parallel β barrels.
**Fig. 76.** Parallel α/β: classic doubly wound β sheets.
Fig. 77. Parallel α/β: doubly wound parallel β sheets.
Fig. 77-2. Parallel α/β: doubly wound parallel β sheets.
Fig. 78. Parallel α/β: miscellaneous.
Fig. 79. Antiparallel β: up-and-down β barrels.
FIG. 80. Antiparallel β: Greek key β barrels.
Fig. 81. Antiparallel β: "jellyroll" Greek key β barrels.
Fig. 82. Antiparallel β: other, multiple, and partial barrels.
Fig. 83. Antiparallel β: open-face sandwich β sheets.
Fig. 83-2. Antiparallel β: open-face sandwich β sheets.
Fig. 84. Antiparallel β: miscellaneous.
Fig. 85. Small disulfide-rich.
Fig. 86. Small metal-rich.
4. Index of Proteins

[PDB (Protein Data Bank) ID codes are given here for all cases where the drawing was based on coordinates then available in the Atlas of Macromolecular Structure on Microfiche (Feldmann, 1977) or in the Protein Data Bank (Bernstein, 1977). For those drawings made from coordinates provided by the crystallographers (see acknowledgements) or for the few made from published stereo figures, the PDB code is given only where that later-deposited coordinate set is essentially the same as the one used.]

Acid proteases [1APE, 1APP], see Rhizopuspepsin

Actinidin (Baker, 1980) [2ACT], see Papain

Adenylate kinase (Schulz et al., 1974a) [2ADK]
  Doubly wound parallel β sheet (Fig. 77)

Agglutinin, wheat germ (Wright, 1977) [2WGA]
  Domains 1, 2, 3, and 4: small SS-rich (Fig. 85)

Alcohol dehydrogenase, liver (Eklund et al., 1976) [4ADH]
  Domain 1: multiple β barrel (Fig. 82)
  Domain 2: classic doubly wound β sheet (Fig. 76)

Aldolase, 2-keto-3-deoxy-6-phosphogluconate (Mavridis and Tulinsky, 1976) [1KGA; later reconnected]; (Richardson, 1979) [Allard, 2001; 1EU]
  Singly wound parallel β barrel? [yes] (Fig. 75)

Arabinose-binding protein (Quiocho et al., 1977) [1ABP]
  Domains 1 and 2: doubly wound parallel β sheet (Fig. 77)

Aspartate aminotransferase (Ford et al., 1980) [1AAT]
  Doubly wound parallel β sheet

Aspartate carbamoyltransferase, (Monaco et al., 1978) [1ATC]
  Regulatory domain 1: open-face β sandwich (Fig. 83)
  Regulatory domain 2: open-face β sandwich (Fig. 83)
  Catalytic domain 1: doubly wound parallel β sheet
  Catalytic domain 2: classic doubly wound β sheet (Fig. 76)

Aspartate transaminase: see Aspartate aminotransferase

Aspartate transcarbamylase *** (Monaco et al., 1978)***, see Aspartate carbamoyltransferase

Azurin (Adman et al., 1978) [1AZU], see Plastocyanin

Bacteriochlorophyll protein (B. W. Matthews et al., 1979) [2BCL]
  Open-face β sandwich (Fig. 83)

Bacteriorhodopsin, see Purple membrane protein

Bence-Jones protein, see Immunoglobulin

Calcium-binding protein, carp muscle (Kretsinger and Nockolds, 1973) [1CPV]
  Miscellaneous antiparallel α (Fig. 74)

Carbonate dehydratase, see Carbonic anhydrase

Carbonic anhydrase C (Lindskog et al., 1971) [1CAC]
  Miscellaneous parallel α/β (Fig. 78)

Carboxypeptidase A (Quiocho and Lipscomb, 1971) [1CPA]
  Miscellaneous parallel α/β (Fig. 78)

Catalase (Vainshtein et al., 1980) [4CAT]
  Domain 1: up-and-down β barrel
  Domain 2: miscellaneous antiparallel α
  Domain 3: doubly wound parallel β sheet

Chymotrypsin (Birktoft and Blow, 1972) [2CHA], see Trypsin

Citrate synthase (Wiegand et al., 1979) [1CTS]
  Miscellaneous antiparallel α

Concanavalin A (Reeke et al., 1975) [2CNA]
  Jellyroll Greek key β barrel (Fig. 81)

Crambin (Hendrickson and Teeter, 1981) [1CRN]
  Small SS-rich (Fig. 85)

γ-Crystallin (Blundell et al., 1981) [1GGR]
  Domains 1 and 2: jellyroll Greek key β barrel

Cytochrome b$_5$ (Mathews et al., 1972) [2BSC]
  Small metal-rich (Fig. 86)
Cytochrome b$_{562}$ (Mathews et al., 1979) [1568]
  Up-and-down helix bundle (Fig. 72)
Cytochrome c (Swanson et al., 1977) [1CYT]
  Small metal-rich (Fig. 86)
Cytochrome c´ (Weber et al., 1980) [1CCY]
  Up-and-down helix bundle (Fig. 72)
Cytochrome c$_{5}$ (Salemme et al., 1973) [1C2G], see Cytochrome c
Cytochrome c$_{3}$ (Hase et al., 1979) [1CY3]
  Small metal-rich (Fig. 86)
Cytochrome c$_{5,50}$ (Timkovich and Dickerson, 1973) [155C], see Cytochrome c
Cytochrome c$_{5,51}$ (Almassy and Dickerson, 1978) [251C], see Cytochrome c
Cytochrome c$_{5,55}$ (Korszun and Salemme, 1977), see Cytochrome c
Cytochrome c peroxidase (Poulos et al., 1980) [2CYP]
  Domain 1: Greek key helix bundle
  Domain 2: miscellaneous antiparallel $\alpha$
Dehydrogenases, see Alcohol, Glyceraldehyde phosphate, Malate, or Lactate
Dihydrofolate reductase (Matthews et al., 1977) [2DFR]
  Doubly wound parallel $\beta$ sheet (Fig. 77)
Elastase (Sawyer et al., 1978) [1EST], see Trypsin
Erabutoxin (Low et al., 1976)
  Small SS-rich (Fig. 85)
Ferredoxin (Adman et al., 1973) [1FDX]
  Small metal-rich (Fig. 86)
Ferritin (Banyard et al., 1978; Clegg et al., 1980)
  Up-and-down helix bundle
Flavodoxin (Burnett et al., 1974) [4FXN]
  Doubly wound parallel $\beta$ sheet (Fig. 77)
Gene 5 protein, fd phage (McPherson et al., 1979) [1PGI]
  Strand connectivity changed-2GN5, 1VQB
  Miscellaneous antiparallel $\beta$
Glucosephosphate isomerase (Shaw and Muirhead, 1977) [1PGI]
  Domains 1 and 2: miscellaneous parallel $\alpha/\beta$
Glutathione peroxidase (Ladenstein et al., 1979)
  Doubly wound parallel $\beta$ sheet
Glutathione reductase (Schulz et al., 1978) [1GRS]
  Domains 1 and 2: doubly wound parallel $\beta$ sheet (Fig. 77)
  Domain 3: open-face $\beta$ sandwich (Fig. 83)
Glyceraldehyde-phosphate dehydrogenase (Buchner et al., 1974) [1GPD]
  Domain 1: doubly wound parallel $\beta$ sheet (Fig. 77)
  Domain 2: open-face $\beta$ sandwich (Fig. 83)
Glycogen phosphorylase (Sprang and Fletterick, 1979)
  Domain 1: doubly wound parallel $\beta$ sheet, five-layer
  Domain 2: classic doubly wound $\beta$ sheet, five-layer (Fig. 76)
Hemagglutinin, influenza virus (Wilson et al., 1981) [1HMG]
  HA1: jellyroll Greek key $\beta$ barrel
  HA2: open-face $\beta$ sandwich
  HA2 around 3-fold: miscellaneous helix cluster
Hemerythrin (Stenkamp et al., 1978) [1HMN], see Myohemerythrin
Hemoglobin (Ladner et al., 1977) [2MHB]
  Greek key helix bundle (Fig. 73)
Hexokinase (Steitz et al., 1976) [2YHX]
  Domains 1 and 2: doubly wound parallel $\beta$ sheet (Fig. 77)
High-potential iron protein (Carter et al., 1974) [1HIP]
  Small metal-rich (Fig. 86)
$\rho$-Hydroxybenzoate hydroxylase (4-hydroxybenzoate 3-mono-oxigenase) Wierenga et al., 1979) [1PHH]
  Domain 1: doubly wound parallel $\beta$ sheet
  Domain 2: open-face $\beta$ sandwich
Domain 3: miscellaneous antiparallel $\alpha$

Immunoglobulin (Epp et al., 1974; Silverton et al., 1977) [1REI, 2FAB]
  Variable and constant domains: Greek key $\beta$ barrel (Fig. 80)

Insulin (Blundell et al., 1972) [1INS]
  Small SS-rich (Fig. 85)

Kinases, see Adenylate kinase, Hexokinase, Phosphoglycerate kinase, Phosphofructokinase, or Pyruvate kinase

Lactate dehydrogenase (Adams et al., 1970) [4LDH]
  Domain 1: classic doubly wound $\beta$ sheet (Fig. 76)
  Domain 2: miscellaneous antiparallel $\beta$ (Fig. 84)

Lysozyme, hen egg white (Imoto et al., 1972) [1LYZ]
  Miscellaneous antiparallel $\alpha$ (Fig. 73)

Lysozyme, $T_4$ phage (Matthews and Remington, 1974) [1LZM]
  Domain 1: open-face $\beta$ sandwich (Fig. 83)
  Domain 2: Greek key helix bundle (Fig. 73)

L7/L12 ribosomal protein (Leijonmarck et al., 1980) [1CTF]
  Open-face $\beta$ sandwich

Malate dehydrogenase (Hill et al., 1972) [1MDH], see Lactate dehydrogenase

Myoglobin (Watson, 1969) [1MBN], see Hemoglobin

Myohemerythrin (Hendrickson and Ward, 1977) [1MHR]
  Up-and-down helix bundle (Fig. 72)

Neurotoxin

Cobra (Walkinshaw et al., 1980), see Erabutoxin

Sea snake (Tsernoglou and Petsko, 1977) [1NXB], see Erabutoxin

Nuclease, staphylococcal (or micrococcal) (Arnone et al., 1971) [1NSN]
  Greek key $\beta$ barrel (Fig. 80)

Papain (Drenth et al., 1971) [8PAP]
  Domain 1: Greek key helix bundle (Fig. 73)
  Domain 2: up-and-down $\beta$ barrel (Fig. 79)

Parvalbumin, see Calcium-binding protein

Pepsin (Andreeva and Gustchina, 1979) [1PEP], see Rhizopuspepsin

Phosphoglycerate kinase (Banks et al., 1979) [2PGK]
  Domain 1: doubly wound parallel $\beta$ sheet
  Domain 2: classic doubly wound $\beta$ sheet (Fig. 76)

Phosphoglycerate mutase (Campbell et al., 1974) [1PGM]
  Doubly wound parallel $\beta$ sheet (Fig. 77)

Phospholipase A2 (Dijkstra et al., 1978) [1BP2]
  Small SS-rich (Fig. 85)

Phosphorylase, see Glycogen phosphorylase

Plastocyanin (Colman et al., 1978) [1PCY]
  Greek key $\beta$ barrel (Fig. 80)

Prealbumin (Blake et al., 1978) [2PAB]
  Greek key $\beta$ barrel (Fig. 80)

Protein A fragment, staphylococcal (Deisenhofer et al., 1978) [1FC2]
  Up-and-down helix bundle (Fig. 72)

Purple membrane protein (Henderson and Unwin, 1975) [1BRD, 1MOM]
  Either up-and-down or Greek key helix bundle [up & down]

Pyruvate kinase (Stuart et al., 1979) [1PYK]
  Domain 1: singly wound parallel $\beta$ barrel (Fig. 75)
  Domain 2: Greek key $\beta$ barrel (Fig. 80)
  Domain 3: doubly wound parallel $\beta$ sheet (Fig. 77)

Rhodanese (Ploegman et al., 1978) [1RHD]
  Domains 1 and 2: doubly wound parallel $\beta$ sheet (Fig. 77)

Rhizopuspepsin (Subramanian et al., 1977) [1APR]
  Domains 1 and 2: other $\beta$ barrel (Fig. 82)

Ribonuclease, bovine pancreatic (Wyckoff et al., 1970) [1RNS]
Partial β barrel (Fig. 82)
Rubredoxin (Watenpaugh et al., 1979) [3RXN]
  Small metal-rich (Fig. 86)
Serine proteases, see Trypsin, Chymotrypsin, Elastase, Streptomyces griseus proteases A and B, or Subtilisin
Southern bean mosaic virus protein (Abad-Zapatero et al., 1980) [1SBV]
  Jellyroll Greek key β barrel (Fig. 81)
Streptomyces griseus protease A (Brayer et al., 1978) [1SGA], see Trypsin
Streptomyces griseus protease B (Delbaere et al., 1975) [1SGB], see Trypsin
Subtilisin (Wright et al., 1969) [1SBT]
  Doubly wound parallel β sheet (Fig. 77)
Subtilisin inhibitor, Streptomyces (Mitsui et al., 1979) [2SSI]
  Open-face β sandwich (Fig. 83)
Sulfhydryl proteases, see Actinidin, Papain
Superoxide dismutase, Cu,Zn (Richardson et al., 1975) [2SOD]
  Greek key β barrel (Fig. 80)
Thermolysin (Colman et al., 1972) [2TLN]
  Domain 1: open-face β sandwich (Fig. 83)
  Domain 2: Greek key helix bundle (Fig. 83)
Thioredoxin, E. coli (Holmgren et al., 1975) [1SRX]
  Miscellaneous parallel α/β (Fig. 78)
Thiosulfate sulfurtransferase, see Rhodanese
Tobacco mosaic virus protein (Bloomer et al., 1978)
  Up-and-down helix bundle (Fig. 72)
Tomato bushy stunt virus protein (Harrison et al., 1978) [2TBV]
  "Domain" 1: miscellaneous antiparallel β (Fig. 84)
  Domains 2 and 3: jellyroll Greek key β barrel (Fig. 81)
Triosephosphate isomerase (Banner et al., 1975) [1TIM]
  Singly wound parallel β barrel (Fig. 75)
tRNA synthetase, tyrosyl (Irwin et al., 1976; D. M. Blow, personal communication)
  Domain 1: classic doubly wound β sheet
  Domain 2: up-and-down helix bundle
Trypsin (Stroud et al., 1974) [3PTP]
  Domains 1 and 2: Greek key β barrel (Fig. 80)
Trypsin inhibitor, pancreatic (Deisenhofer and Steigemann, 1975) [3PTI]
  Small SS-rich (Fig. 85)
Trypsin inhibitor, soybean (Sweet et al., 1974)
  Up-and-down β barrel (Fig. 79)
Uteroglobin (Mornon et al., 1980)
  Up-and-down helix bundle (Fig. 72)
Viral coat proteins, see Southern bean mosaic virus, Tobacco mosaic virus, or Tomato bushy stunt virus
The first major grouping of structures contains domains that are essentially all α-helical. Since there is relatively little other structure besides the helices, the simplest ways of connecting them involve predominantly antiparallel helix interactions, and that is in fact what is observed for these proteins. This category corresponds to Levitt and Chothia’s all-α category, but it has more members both because of a number of new structures and because of helical domains in proteins they classified as α + β (such as thermolysin). [This category is again now called “All-α”, both to match the “All-β” which now has parallel β folds (see below), and also because of the increased role of perpendicular helices (here seen only for the E-F hands of carp Ca-binding protein in Fig. 74) and of multichain parallel coils (here seen only for the flu haemagglutinin in Fig. 72).]

Figures 72 through 74 show schematic diagrams of the antiparallel α domains, grouped into subcategories. Almost all of them are two layer structures. The simplest and commonest subgroup looks like a bundle of sticks: usually four helices bundled in a cylinder with simple +1 connections. Most of the helices are quite close to exactly antiparallel, with typically a left-handed superhelical twist of less than 15° relative to the common axis of the bundle. These structures were first described as a group in Argos et al. (1977). Figure 87 illustrates myohemerythrin as an example of this structure type, showing an α-carbon stereo, a schematic drawing, and a topology diagram.

The simple up-and-down helix bundle structures include the hemerythrins (myohemerythrin and the hemerythrin subunits), cytochrome b562, cytochrome c’ uteroglobin, tobacco mosaic virus protein, staphylococcal protein A fragment, and probably the ferritin subunits [yes]. Tyrosine-tRNA synthetase domain 2 has quite a similar organization, but the last helix tilts away from the bundle (Blow et al., 1977). The uteroglobin subunit also has its fourth helix out to one side, but in the dimer molecule (Fig. 88) those final helices each complete a compact four-helix bundle with the rest of the opposite subunit. In cytochrome c’ there is a similar but less extreme arrangement in which the first helix lies at a greater angle to the bundle axis and forms the tightest part of the dimer contact. [The uteroglobin arrangement...]

**Fig. 87.** Myohemerythrin as an example of an up-and-down helix bundle. (a) α-Carbon stereo; (b) schematic drawing of the backbone structure, from the same viewpoint as in a.
is an early example of what is now called a domain-swapped dimer [Eisenberg]

Tobacco mosaic virus protein has a small, highly twisted antiparallel β sheet at the base of the helix bundle, with two more helices underneath the sheet (see Fig. 72). Cytochrome b₅ looks remarkably similar (see Fig. 105), but the helices are much shorter. That structure could have been classified as an up-and-down helix bundle, but we have placed it in the small metal-rich proteins because its helix bundle is very small and distorted and the heme interactions appear more important than the direct helix contacts.

All but one of the above structures have four helices in the bundle, with +1,+1,+1 connections. For the up and down topology on a cylinder, handedness can be defined by whether the chain turns to the right or to the left at the end of the first structure element (whether it is a helix or a β strand). With an even number of helices, reversing N to C direction of the chain also reverses handedness of the topology; for an odd number of helices or strands handedness is invariant to chain reversal. For +1,+1,+1 topologies in general, handedness is not a very robust criterion of similarity, since it reverses on addition or deletion of one of the structure elements at the N-terminus but not at the C-terminus, so that a given five-helix structure could have evolved from either handedness of four-helix structure. Hemerythrin, cytochrome b₅₂, cytochrome c’, uteroglobin, and tobacco mosaic virus protein are all right-handed, while cytochrome b₅, tyrosine-tRNA synthetase, and staphylococcal protein A fragment are left-handed. [The up & down 4-helix bundle is a very common superfold. But more complex topologies with long connections occur.]

The connectivity is not known for the seven-helix bundle of purple membrane protein (Henderson and Unwin, 1975), but on the basis of its resemblance to other antiparallel α proteins the most likely topologies would be either up-and-down or Greek key (see below). An analysis based on the sequence and the relative electron-densities of the helices (Engelman et al., 1980) considers a left-handed up-and-down topology as the most probable model. [The up & down 7 helix bundle structure was confirmed by higher resolution electron diffraction and x-ray structures, both for purple membrane protein (bacteriorhodopsin) and for rhodopsin.]

Many of the up-and-down helix bundle proteins form large multisubunit arrays. Hemerythrin is an octomer, with the end of one helix bundle butting against the side of the next one around the 4-fold axis (Ward et al., 1975). The 24 ferritin subunits form a hollow spherical shell with the helix bundles approximately tangential to the shell and the subunit interactions around the 3-fold and 4-fold axes rather like the interactions in hemerythrin. Tobacco mosaic virus protein, on the other hand, forms a tightly packed long helix of subunits; the α-helical bundles are aligned radially, with RNA bound at their inner ends. Purple membrane protein spans the membrane, forming a two dimensional crystalline array with the helix bundles perpendicular to the membrane and parallel to each other around the 3-fold axis.

One of the most important and interesting antiparallel α structures is the globin fold, which has been found in the three-dimensional structures of a large group of related proteins including myoglobin and the hemoglobins of various mammals, glycera, lamprey, insect and even legume root nodules. The globin fold is a good example of how there may be several alternative useful ways of describing a given structure. To someone studying hemoglobin function the relevant level of description includes all the structural detail that can be made comprehensible, or perhaps generalized to include what is common to all the globin structures. On the other hand, if one is concerned, as we are here, with obtaining a memorably simple description of the whole structure and relating it to other protein structures, then the issue is deciding which features are most important to include in the simplification and with which if any other proteins it can meaningfully be compared. Classifying the globins as all-α proteins is obviously true and useful, but Levitt and Chothia’s (1976) scheme of representing the globin topology does not suggest similarities to any of the other all-α proteins, even when the more recent structures are included. Argos and Rossmann (1979) have suggested an interesting similarity of structure around the heme pocket for the globins, cytochrome b₅, and cytochrome c₅₅₁. Their description is probably the most relevant one for trying to understand how heme-binding pockets are organized, but it does not seem suitable as a general structural description since the omitted halves of the three structures are all extremely
Fig. 89. Hemoglobin (β subunit) as an example of a Greek key helix bundle. (a) α-Carbon stereo, (b) schematic drawing of the backbone as two perpendicular layers of α-helices (shown here as cylinders); (c) schematic drawing of the backbone as a Greek key helix bundle (from the same viewpoint as in a); (d) schematic end-on view of the hemoglobin helix bundle, to show that it is a slightly flattened cylinder in cross section (the C-D loop is shown dashed because it would cover part of the cylinder).

different and do not form separable domains.

Figure 89 illustrates two different tries at simplified representation of the globin structure. For reference, Fig. 89a shows the hemoglobin β chain in stereo. Figure 89b shows the globin structure schematically as two layers of helices with the elements in one layer approximately perpendicular to those in the other layer; this can be contrasted with a possible description of the up-and-down helix bundles as two layers with their elements approximately parallel to each other. The perpendicular layers provide a rather successful simple schema for the globin structure, but unfortunately there are no other proteins that can be adequately described as two perpendicular layers of helices. Also, specification of the topology in this scheme is cumbersome, since the chain skips back and forth between layers.

Figure 89c schematizes the globin structure as a twisted cylinder of helices, analogous to the antiparallel β barrels to be discussed in Section III,D. The up-and-down helix bundle structures are of course also readily described as cylinders, so that this schema makes the majority of antiparallel α structures directly analogous to the majority of antiparallel β structures. Their topologies can be conveniently specified by the simple nomenclature listing connection types (see Section II,B). The major irregularity of the globin fold when considered as a cylinder is that one element (the A and B helices) bends sharply to close the cylinder; this feature is also seen in five- and six-stranded β barrels such as trypsin. But perhaps the most satisfying
feature of schematizing the globin fold on a cylinder is that it can then be grouped with other structures (thermolysin d2, T4 phage lysozyme d2, papain d1, and cytochrome c peroxidase d2) which also show the “Greek key” (see Richardson, 1977) topology of + 3, - 1, - 1. Papain domain 1 also shows the diagnostic feature of Greek key structures by containing a non-nearest-neighbor connection which skips across the end of the cylinder, however, most of its helices are short and they form a rather irregular bundle. Papain domain 1 contains two disulfides; we will find repeatedly that increasing disulfide content goes along with decreasing regularity of both secondary and tertiary structure.

These four structures then form the second major subgrouping of antiparallel α domains, which we will call Greek key helix bundles (see Fig. 73). The helix elements lie on an approximate cylinder (see Fig. 89d for an end view), with 0 to 45˚ right-handed twist relative to the cylinder axis; they are connected with a Greek key topology which can have either a counterclockwise (globins) or a clockwise (thermolysin d2 and T4 lysozyme d2) swirl when viewed from the outside.

The remaining structures in this category (carp Ca-binding protein, egg lysozyme, citrate synthase, catalase d2, and ρ-hydroxybenzoate hydroxylase d3) are miscellaneous helical domains. However, there is good evidence from sequences and from functional resemblances (Kretsinger, 1976) that carp Ca-binding protein exemplifies a whole group of proteins that are constructed of “E-F hands” (see Section II,F) and that regulate or are regulated by changes in Ca2+ concentration. [This is now a large group typified by calmodulin. The “T-shaped” helix-turn-helix motifs of many transcription factors are superimposable onto E-F hands in the reverse sequence direction (Richardson).] Citrate synthetase may be the first example of a group of larger helical domains with three layers. [The peroxidases (e.g. 2CYP) are examples, but the layers are usually not distinct.] Irregular helical structures with a moderate number of disulfides can be classified either here or as small SS-rich. We have classified egg lysozyme here (with only 4 disulfides in 129 residues), while phospholipase A2 (with 7 disulfides in 123 residues) is classified with the small SS-rich proteins.

[The most distinct new all-α folds are the cylindrical or crescent-shaped repeats of helix hairpins, both of which are exemplified by the two chains of farnesyl transferase.]
C. Parallel α/β Domains

The largest grouping of structures contains domains organized around a parallel or mixed β sheet, the connections for which form structure (usually helical) protecting both sides of the sheet, with the helices also predominantly parallel to each other. Of course, each helix and its neighboring strand are antiparallel to one another, but this structure category is called parallel α/β because both the β sheet interactions and the α-helix interactions are internally parallel. The parallel α/β category is the same as Levitt and Chothia’s α/β proteins. Figures 75-78 show schematic drawings of this group of structures. It is interesting to note that there seems no a priori reason not to have parallel all-α structures or parallel all-β structures formed of two helix layers or of two parallel β sheets, yet such structures are not found. [There are now indeed examples of both these types, described in the annotations to sections III.B for helical horse-shoes and III.D. for β-helices.] All of the domains with parallel organization have both a β sheet of at least four or five strands and at least three or four α-helices. Almost all have at least three layers.

The first subgrouping under the parallel α/β category contains two of the largest but simplest domain structures that have yet been found. They are the eight-stranded parallel β barrels of triosephosphate isomerase (see Fig. 90) and pyruvate kinase domain d1, both of which are connected in +1x,+1x topology all the way around. (In structures with both β sheet and also several helices it is convenient to use just the β strands for designating the topology.) The connections are α-helices, which form a larger cylinder of parallel helices concentric with the β barrel. The structural elements of both α and β cylinders have a pronounced right-handed twist around the cylinder axis. Connections between the parallel β strands must lie on the outside of the barrel since the interior is filled by the packed hydrophobic side chains. If all of the crossover connections must be right-handed and no knots are allowed, then the chain must wind consistently around the barrel in one direction, and the +1x, +1x, +1x topology is not only the simplest but essentially the only possible topology for such a structure (Richardson, 1977), since all other topologies are knotted and unfoldable. We call this structure the singly wound parallel β barrel, since successive crossover connections are wound on the barrel progressing in a single direction with no reversal or backtracking. Figure 91a is a highly schematized representation of the “singly wound” structure, viewed from one end of the barrel.

The largest subgrouping within the parallel α/β category contains structures with a central twisted wall of parallel or mixed β sheet, protected on both sides by its crossover connections (most of which are helical). This is called the doubly wound parallel β sheet, because with right-handed crossovers the simplest way of protecting both sides of the sheet is to start near the middle and wind toward one edge, then return to the middle and wind to the other edge. Figure 91b is a highly schematized representation of the “doubly wound” structure, viewed from one end of the sheet (compare with Fig. 91a). The singly wound barrel has four major layers of backbone structure and the doubly wound sheet has three major layers (with two separate hydrophobic cores); most other domain structures have only two major backbone layers with a single hydrophobic core, and are on the average considerably smaller.

The doubly wound structures were first recognized as a category by Rossmann in comparing flavodoxin with lactate dehydrogenase d1. As more and more protein structures were solved which fell into this category, the relationships between them have been described and debated at considerable length. The initial descriptions were in terms of the β–α–β–α–β unit as a supersecondary structure (Rao and Rossmann, 1973). Quite soon the emphasis shifted to the functional properties of the nucleotide-binding site which most of them share, and to the probable evolutionary relationships between these “nucleotide-binding domains” (Schulz and Schirmer, 1974; Rossmann et al., 1974). By now the consensus appears to be that some of the most similar of these structures must certainly be related to each other, while at least some of the most dissimilar examples surely cannot be related (Rossmann and Argos, 1976; Matthews et al., 1977; Levine et al., 1978; McLachlan, 1979a).

We will group these domains into five gradually loosening levels of topological similarity, without attempting to make any definite decision as to where the dividing line lies between divergent and convergent examples.

The darkest inner box in Fig. 93 includes those “classic” doubly wound parallel sheets that exactly match the topology of lactate dehydrogenase d1. Phosphorylase domain 2 is a five-layer structure in which the central three layers are a classic doubly wound sheet and the outer helical layers are formed by the two ends of the chain. The next box includes examples in which deleting one or two strands either at an end of the chain or at an edge of the β sheet will produce a five- or six-strand doubly wound sheet, while in the next box...
Fig. 90. Triosephosphate isomerase as an example of a singly wound parallel β barrel. (a) α-Carbon stereo, viewed from one end of the barrel; (b) backbone schematic, viewed as in a; (c) α-carbon stereo, viewed from the side of the barrel; (d) backbone schematic, viewed as in c; (e) topology diagram showing the +1x right-handed connections between the β strands.
such deletions yield four doubly wound strands. In the outermost solid box it is necessary to omit strands interior to both the sheet and the sequence in order to get four strands of doubly wound β sheet. The structures inside the dotted box can yield no more than three such strands and cannot really be described as doubly wound, they share with the rest of this large subgrouping only the general organization of a central wall of parallel or mixed β sheet protected on both sides by its connections (see Fig. 94 for carboxypeptidase as an example). As one progresses outward from the classic to the most peripheral cases, the number of antiparallel strand pairs mixed in with the parallel gradually increases. Aside from the “classic” examples in the inner box, there are several other exact duplicates of doubly wound topologies between different proteins: phosphorylase d1 versus glycer-aldehyde-phosphate dehydrogenase d1; aspartate transcar-bamylase catalytic d1 versus rhodanese d1, d2; catalase d3 versus flavodoxin; and ρ-hydroxybenzoate hydroxylase d1 versus glutathione reductase d1.

As one progresses from classic to peripheral doubly wound sheets, the number of domains that bind nucleotides also decreases. A favorable site for binding dinucleotides (or in a few cases, mononucleotides) is associated with this general category of structure and to a large extent with the doubly wound topology. The dinucleotides are all bound in approximately equivalent positions at the C-terminal end of the β sheet strands, within one strand of the central position where the winding switches direction (see Fig. 91b). [The importance of these “topological switch points” for active sites in doubly-wound domains is well explained in Branden and Tooze (1991).] Nucleotides are also bound at the C-termini of β strands in the singly wound barrels. In most of these cases, each nucleotide is associated with a “mononucleotide-binding fold” of three β strands and two helices with +1x, +1x topology; combination of two of these folds around a local 2-fold axis produces the classic doubly wound sheet. In some cases, however (such as hexokinase or dihydrofolate reductase), the topology is quite significantly different. Also there seems to be another quite different type of nucleotide-binding site such as the active site in staphylococcal nuclease (Arnone et al., 1971) or the AMP site in phosphorylase (Sygusch et al., 1977); both of these sites rely mainly on arginines for binding the nucleotide phosphates.

One quite surprising and intriguing feature of this group of structures is that it contains extremely few examples of the “reverse doubly wound” topology (see Fig. 91c), a different but equally plausible pattern related to the doubly wound sheet by reversing the N- to C-terminal direction.

**Fig. 91.** Highly simplified sketches (viewed from the C-terminal end of the β strands) of (a) a singly wound parallel β barrel; (b) a classic doubly wound β sheet; (c) a reverse doubly wound β sheet. Thin arrows next to the diagrams show the direction in which the chain is progressing from strand to strand in the sheet.
FIG. 92. Lactate dehydrogenase domain 1 as an example of a classic doubly wound parallel β sheet. (a) α-Carbon stereo, viewed from one edge of the sheet; (b) backbone schematic, viewed as in (a); (c) backbone schematic, viewed from one face of the sheet.

of the chain or by switching relative positions of the two halves of the β sheet. Of the four reverse examples (found in PGK d2, GPDH d1, glucosephosphate isomerase d1, and phosphorylase d1) none forms a nucleotide-binding site, all belong to a sheet that also has a normal doubly wound section, and none includes more than four strands. Those cases do demonstrate that such a topology is stable and can fold, but there must be some strong reason why it is so rare. Some simple explanations of this regularity would be either that most of the nucleotide-binding domains are related, or that they must fold strictly from the N-terminus, or that the requirements for forming a nucleotide-binding site are restrictive enough to constrain the usual doubly wound topology. None of these explanations is completely satisfying, however, because a number of domains are known that cannot fold strictly from the N-terminus, because the relative placement of features forming the nucleotide sites is only rather approximate (e.g., see D. A. Matthews et al., 1979), and because the rearrangement necessary to produce a reverse doubly wound sheet seems much less drastic than many of the rearrangements that must be proposed if all of these proteins are related.
Fig. 93. Topology diagrams for the doubly wound and miscellaneous α/β domains illustrated in Figs. 76 through 78. Arrows represent the β strands; thin connections lie behind the β sheet and fat ones above it. The darkest upper box surrounds the classic doubly wound sheets; successively lighter solid boxes include domains that are progressively less like the classic topology; the dotted box encloses the miscellaneous α/β structures. K = kinase; P = phospho; DH = dehydrogenase; ATCase = aspartate transcarbamylase.
Fig. 94. Carboxypeptidase A as an example of a miscellaneous α/β structure. (a) α-Carbon stereo, viewed from one edge of the mixed β sheet; (b) backbone schematic, viewed as in a.
The next major grouping consists of domains that are organized around an antiparallel β sheet. They are as numerous as the parallel α/β structures, and their topology and classification have been discussed before (see Levitt and Chothia, 1976; Sternberg and Thornton, 1977a,b; Richardson, 1977). This category is the most varied in terms of size and organizational patterns. Figures 79 through 84 show backbone schematics for the antiparallel β domains, grouped into subcategories.

Most of the antiparallel β domains have their β sheets wrapped around into a cylinder, or barrel, shape. None of the antiparallel barrels has as symmetrical or as continuously hydrogen-bonded a cylindrical sheet as the singly wound parallel β barrels of triosephosphate isomerase and pyruvate kinase d1; however, antiparallel barrels are very much more common. Because of gaps in the hydrogen bonding, some of these structures have been described as two β sheets facing each other (e.g., Schiffer et al., 1973; Blake et al., 1978; Harrison et al., 1978). Our reasons for treating them all as barrels are that the gap positions are sometimes different in domains that are probably related, and that the barrel description yields very much simpler and more unified topologies.

Barrels seem to prefer pure parallel or antiparallel β structure even more strongly than does β sheet in general. All the known singly wound barrels are pure parallel. An antiparallel barrel with an odd number of strands is constrained to have one parallel interaction, but no other parallel strand pairs occurs within antiparallel barrels except in the acid proteases. [Of course, there are now more absolute numbers of such exceptions, but they are still quite rare.] Also, even-stranded barrels are much more common than odd-stranded ones.

The first type of antiparallel β barrel, in analogy with the first type of helix bundle, has simple up and down +1, +1, +1 connections all around. Although it is relatively unusual for a barrel to be composed entirely of up-and-down strands, many of the larger barrels and sheets have four- to six-stranded sections of simple up-and-down topology embedded within them. There are three examples of pure up-and-down β barrels: soybean trypsin inhibitor, papain d1, and catalase d1. [Two interesting later examples are retinol-binding protein (1KT7) and the large, membrane-spanning, 16-strand or more up-and-down barrel of porins (e.g. 1HXX).] Figure 95 shows a stereo and a schematic drawing of papain d1. Soybean trypsin inhibitor has long excursions at the

---

**Fig. 95. Papain domain 2 as an example of an up-and-down antiparallel β barrel. (a) α-Carbon stereo, viewed from one side of the barrel; (b) backbone schematic, viewed as in a.**
ends of three of the β strand pairs, forming separate, twisted β ribbons; there is a strong internal 3-fold symmetry which includes these ribbons as well as the strand pairs in the barrel (McLachlan, 1979C). Catalase d1 is an eight-stranded up-and-down barrel with less extreme loop excursions. Rubredoxin could be considered as a very irregular and incomplete up-and-down barrel in which β-type hydrogen bonds are formed between only about half of the strand pairs (see Fig. 76). It is very small and compact, and is presumably stabilized partly by the network of Cys ligands to the iron; therefore we have placed it in the small metal-rich category.

Soybean trypsin inhibitor, papain d1, and rubredoxin have identical topologies: six strands of +1, +1, +1, ... proceeding to the left around the barrel if the chain termini are at the bottom. However handedness is not nearly as meaningful a property for up-and-down topologies as it is for Greek keys, since up-and-down handedness can change on addition or defection of a single strand.

The commonest subgroup of antiparallel β barrel structures has a Greek key topology, with -3, +1, +1, -3 connections or a close variant. The first Greek key barrel structures were compared in Richardson et al. (1976), and they and the up-and-down barrels were described as categories in Richardson (1977). Figure 96 illustrates Cu,Zn superoxide dismutase as an example of a Greek key β barrel. There are 13 Greek key barrels in our sample, and 12 of them (all except staphylococcal nuclease) have the same handedness: viewed from the outside, the Greek key pattern forms a counterclockwise swirl (see Fig. 97). The four barrels shown in Figure 81 have a more complicated “jellyroll” topology with an extra swirl in the Greek key (this pattern was also common on Greek vases); the “jellyroll” Greek key topologies are shown in Fig. 98. The jellyroll pattern is produced by having a pair of connections, rather than just one connection, crossing each end of the barrel. The Greek key barrels have between 5 and 13 strands, but in all cases they enclose approximately the same cross-sectional area (see Section II,B). The cross sections are somewhat elliptical, with more flattening the more strands there are. For 8- to 10-stranded barrels, it is noticeable that the direction of the long axis of the cross-section twists from one end of the barrel to the other by close to 90° (see Fig. 99).

±3 connections are not particularly common outside of the barrels so that the prevalence of Greek key topologies is not due simply to chance combination of the connection

FIG. 96. Cu,Zn superoxide dismutase as an example of a Greek key antiparallel β barrel. (a) α-Carbon stereo, viewed from one side of the barrel; (b) backbone schematic viewed as in a.
Fig. 97. Topology diagrams of the Greek key antiparallel β barrels. The dashes on either side of a topology diagram indicate that the barrel was opened up at that point and laid out flat; all barrels are shown viewed from the outside.

Fig. 98. Topology diagrams of the “jellyroll” Greek key β barrels.
types that make it up. There are two different ways of analyzing the Greek key which could perhaps explain both its frequent occurrence and its strongly preferred handedness. The first approach is to consider the stability of the final barrel, given its size, shape, and twist. Figure 99 shows that the Greek key pattern provides neat, efficient connections across the top and bottom of the barrel, lying next to the ±1 connections. In tomato bushy stunt virus d3 there is actually some β-type hydrogen bonding between the +1, -3, and +5 connecting loops. In combination with the twist of the strands and of the barrel cross section, a counterclockwise Greek key (as shown) produces ±3 connections that are approximately perpendicular to each other on opposite ends of the barrel and that can both cross along a short axis of the cross section. A clockwise Greek key would place the ±3 connections in a weaker position approximately parallel to each other, and one of them would be along the long axis. This argument could not account for the handedness of the partial Greek keys with -3, +1, +1 topology (such as staphylococcal nuclease and chymotrypsin) where there is a ±3 connection at only one end of the barrel.

The other possible explanation hypothesizes an effect during the protein folding process, very similar to the one proposed to explain crossover handedness (see Section II,B). All Greek keys, even the “jellyrolls,” necessarily have a folding point halfway along the chain from which two paired strands can be followed back next to each other as they curl around the structure. Given the prevalence of Greek key patterns in the known structures, it seems very likely that the polypeptide chain can fold up by first folding in half and forming a long, two-stranded β ribbon, and then curling up the ribbon to produce the further β sheet interactions. This sort of process is illustrated in Fig. 100.

Since the initial ribbon would presumably have a strong right-handed twist (see Section II,B), it would impart a right-handed twist to the curling direction and always end up with a counterclockwise Greek key. Besides the β barrels, there are other pieces of protein structure that suggest this sort of process, such as the long β ribbons in lactate dehydrogenase d2 (see Fig. 74). This kind of folding hypothesis has been utilized by Ptitsyn and Finkelstein (1980) to obtain rather successful predictions of β strand contacts and topologies. [Ray Salemme (1983) explained the curling up of a long, 2-stranded ribbon by the fact that ϕ,ψ preferences are slightly different for β-hairpin residues between a narrow pair of H-bonds and those between a wide pair, such that the former prefer the concave side of a curl and tend to end up on the inside.]

Partial, multiple, and other barrels have been grouped together as another subgroup within the antiparallel β category (see Fig. 82). Ribonuclease contains a four-stranded antiparallel β sheet that looks like a five-stranded barrel with one strand missing. Alcohol dehydrogenase d1 includes a five-stranded antiparallel barrel (with a topology of +1, +3x, -2, -1) and another partial five-stranded barrel. Back-to-back β barrels that share one wall occur in the variable half of immunoglobulin Fab structures (except for Rhe: see Wang et al., 1979), where Vl and VH are each an antiparallel β barrel and the contact between them forms an even more regular eight-stranded barrel with four strands contributed from each domain (see Fig. 101). The three barrels pack against each other with a right-handed superhelical twist, and the angle between the axes of adjoining barrels is the same as the angle between opposite strands in one of the barrels. The two domains of the acid proteases have complicated, very similar mixed β sheets that could be described either as a six-stranded barrel with side sheets or as several interlocking β sheets. When more examples are available, it will probably be possible to find patterns to the ways in which small subsidiary β sheets can interlock into the edges of larger sheets (such as in the acid proteases or thermolysin d1), but for now no attempt has been made to classify them.

The next subcategory of antiparallel β domains each has a single, more or less twisted β sheet, either pure antiparallel or predominantly so, but not closing around to form a barrel. They are shown in Fig. 83, and Fig. 102 shows glyceraldehyde-phosphate dehydrogenase as an example. Their common feature is a layer of helices and loops which covers only one side of the sheet, so that they are two-layer structures. Many β barrels have been described as “sandwiches,” with two slices of β sheet “bread” and a “filling” of hydro-
phobic side chains; based on that analogy these structures would be “open-face sandwiches,” with a single slice of β sheet “bread” and a “topping” of helices and loops. The open-face β sandwiches could rival a Danish buffet for variety on a theme: they range from 3 to 15 strands, with a wide assortment of topologies, curvatures, and placement of helices and loops. Bacterio-chlorophyll protein, the largest of them, encloses between the β sheet layer and the helical layer a core of seven bacteriochlorophyll molecules, tightly packed in an orderly but quite asymmetrical array.

The remaining three antiparallel β structures form a miscellaneous category (see Fig. 84). Lactate dehydrogenase d2 and gene 5 protein each has several two-stranded antiparallel β ribbons, but they do not coalesce into any readily described overall pattern. The N-terminal domain of tomato bushy stunt virus protein has a unique β structure in which equivalent pieces of chain from three different subunits wrap around a 3-fold axis to form what has been called a “β annulus” (Harrison et al., 1978). Each of the three chains contributes a short strand segment to each of three three-stranded, interlocking β sheets. This “domain” provides one of the subunit contacts that hold the virus shell together. However, only one-third of the 180 subunits contribute to the β annuli; for the other quasi-equivalent subunits, the N-terminal part of the chain is disordered with respect to the virus shell.

[Many new variants of all-β structures have been seen more recently, including two entirely unprecedented fold types. The β-propeller is made up of 4, 5, 6, 7, or 8 units, each a small 4-
stranded up-and-down β sheet, arranged radially like propeller or rotor blades around a center of approximate symmetry (e.g., 1TBB G protein β subunit). The β-helix is a parallel all-β structure, which winds around in a shallow spiral forming 3 parallel β-sheets in a triangular cylinder. It comes in both righthanded (1O88 pectate lyase C) and lefthanded (1LXA Lpx A acyltransferase) forms, with different cross-sectional shapes, and is the only handed protein structure that commonly occurs in either handedness.}

Fig. 101. Packing of two β barrel domains in the immunoglobulin VL. dimer (from Bence-Jones REI): (a) α-carbon stereo, viewed from the sides of the barrels; (b) simplified schematic of the barrels as cylinders, viewed as in a; (c) α-carbon stereo, viewed from one end of the barrels. The contact between the two domains forms a third barrel in the center.
Fig. 102. Glyceraldehyde-phosphate dehydrogenase domain 2 as an example of an open-face sandwich antiparallel β sheet. (a) α-Carbon stereo, viewed from the buried side of the sheet; (b) backbone schematic, viewed as in a.
E. Small Disulfide-Rich or Metal-Rich Domains

The last major category (shown in Figs. 85 and 86) consists of small (usually less than 100 residues) domains whose structures seem to be strongly influenced by their high content either of disulfide bonds (S) or of metal ligands (M). These S–M proteins often look like distorted versions of other, more regular, proteins. The disulfide-rich ones include many toxin and enzyme-inhibitor structures. For most of the disulfide-rich proteins it is known experimentally that they are completely unstable if the disulfides are broken (in contrast to larger disulfide-containing proteins, for which disulfides merely provide additional stabilization for an already-determined structure). Figure 103 shows pancreatic trypsin inhibitor as an example of a disulfide rich protein, and Fig. 104 shows cytochrome c as an example of a metal-rich protein. Most of the S–M proteins are single-domain and monomeric: only wheat germ agglutinin has multiple domains, and only insulin has multiple subunits in the molecule.

The only subgroup of similar structures within the S–M proteins is the toxin-agglutinin folds of the snake neurotoxins and the domains of wheat germ agglutinin (see Fig. 85). They are made up of extended-chain loops with an almost identical topology of -1, +3, -1, +2x rather like a series of half-hitch knots (the β structure is extremely minimal in wheat germ agglutinin) strongly linked by a core of four disulfides, three of which are equivalent (see Drenth et al., 1980). High-potential iron protein and ferredoxin share a local loop structure that binds the iron-sulfur cluster, but otherwise are different.

Most of the metal-rich proteins form approximately cylindrical two-layer structures with either an up and down (rubredoxin, cytochrome c) or a Greek key (ferredoxin) topology, but in which the elements forming the cylinder are a mixture of helices, β strands, and more or less extended portions of the backbone. Cytochrome c₃ is perhaps the ultimate example of an S–M protein, with four hemes in just over a hundred residues, and essentially no secondary structure at all except for one helix.

One way of considering these proteins is as distorted versions of the other structural types. Most S–M proteins can fairly clearly be grouped as either distorted helix clusters (phospholipase, cytochrome c, cytochrome b₅), distorted β barrels (rubredoxin, high-potential iron protein), or distorted open-face sandwiches (erabutoxin, wheat germ agglutinin, pancreatic trypsin inhibitor, or ferredoxin). Figure 105 shows an example of each of these relationships. In fact, one reasonable taxonomy would do away with this fourth major category altogether and place all the S–M proteins as irregular examples of either an α or a β category. We have not chosen that approach, however, because several of the structures (crambin, insulin, and cytochrome c₃) are rather difficult to place in one of the other categories, and also because these small proteins influenced by nonpolypeptide interactions appear to share important features, especially in terms of the probable complexity of their folding process (see Section IV,C).

Another suggestive fact is that there are no small, irregular structures related to the parallel α/β category. Per-
haps this reflects the fact that domains organized around parallel β sheet are necessarily fairly large and seem to be dependent on large, buried, and quite regular β structure for their stability. There are in fact no hemes (in spite of all the helices) or iron-sulfur clusters in parallel α/β proteins, and no disulfides except for the single active-site disulfides of thioredoxin, glutathione peroxidase, and glutathione reductase.

**Fig. 104.** Cytochrome c as an example of a small metal-rich protein. (a) α-Carbon stereo, with heme; (b) backbone schematic, viewed as in a. The backbone forms an approximate up and down cylinder with the heme tucked into the center, but the elements forming the cylinder are a mixture of helices and extended strands.
Fig. 105. Examples of small disulfide-rich or metal-rich proteins (shown on the right side) compared with their more regular counterparts in other structural categories (shown at the left). (a) Tobacco mosaic virus protein, an up-and-down helix bundle; (b) cytochrome b$_5$, a distorted up-and-down helix bundle; (c) trypsin domain 1, a Greek key antiparallel $\beta$ barrel; (d) high-potential iron protein, a distorted Greek key $\beta$ barrel; (e) glutathione reductase domain 3, an open-face sandwich $\beta$ sheet; (f) ferredoxin, a distorted open-face sandwich $\beta$ sheet.