PROTEINS: Structure, Function, and Genetics 28:241–260 (1997)

# Structural Trees for Protein Superfamilies

**Alexander V. Efimov***
*Institute of Protein Research, Russian Academy of Sciences, Moscow Region, Russia*

***ABSTRACT*** **Structural trees for large protein superfamilies, such as β proteins with the aligned β sheet packing, β proteins with the orthogonal packing of α helices, two-layer and three-layer α/β proteins, have been constructed. The structural motifs having unique overall folds and a unique handedness are taken as root structures of the trees. The larger protein structures of each superfamily are obtained by a stepwise addition of α helices and/or β strands to the corresponding root motif, taking into account a restricted set of rules inferred from known principles of the protein structure. Among these rules, prohibition of crossing connections, attention to handedness and compactness, and a requirement for α helices to be packed in α-helical layers and β strands in β layers are the most important. Proteins and domains whose structures can be obtained by stepwise addition of α helices and/or β strands to the same root motif can be grouped into one structural class or a superfamily. Proteins and domains found within branches of a structural tree can be grouped into subclasses or subfamilies. Levels of structural similarity between different proteins can easily be observed by visual inspection. Within one branch, protein structures having a higher position in the tree include the structures located lower. Proteins and domains of different branches have the structure located in the branching point as the common fold. Proteins 28:241–260, 1997.**


**Key words: protein structure comparison; protein modeling; stepwise folding; structural motif; structural similarity**

## INTRODUCTION

Three-dimensional structures of most globular proteins can be roughly represented as definite arrangements of α helices and/or β strands. Protein molecules may differ in the number and length of α helices and/or β strands, their distribution along the polypeptide chain, the arrangement of these elements in three dimensions, and their connectivity (connection regions may also have different lengths and conformations). Prediction of secondary structure elements, including their distribution along the chain is now more or less possible[1–3] (reviewed in Ref. 4). The rules that govern the packing together of individual (not connected) α helices and β strands and that result primarily from the principle of close packing and geometric constraints have also been described previously.[5–12] However, self-assembly of α helices and β strands connected by loops into the unique protein structure remains an enormously challenging problem.

Stereochemical analysis and structural comparisons of the increasing number of known protein structures have made possible the recognition of recurring stable arrangements of secondary structural elements, the so-called "supersecondary" structures or structural motifs.[13–25] Structural motifs of a given type found in unrelated proteins may have the same overall fold despite their α helices and/or β strands being of different lengths, their connection regions differing in length and conformation, and their sequences lacking homology. While many different structural motifs have been observed to recur within globular proteins, only some of the motifs exhibit unique handedness and a unique overall fold (reviewed in Ref. 26). The high frequency of occurrence of these structural motifs in unrelated proteins may be a result of relative stability of such folds. On the other hand, these structural motifs are composed of secondary structure elements adjacent along the polypeptide chain, which can associate rapidly to form compact folds. Many small proteins and domains merely consist of the structural motifs.[27] All this taken together suggests that the structural motifs are relatively stable and can fold into unique structures per se. Thus the structural motifs having unique overall folds and unique handedness are very suitable to be taken as starting structures in protein modeling or to be considered as nuclei in protein folding. Previous analyses show that the three-dimensional arrangements of α helices and β strands in the larger proteins can be obtained by stepwise addition of secondary structure elements to the corresponding structural motifs[17,18,21,22,27,32] or to other initiating complexes,[28–31] taking into account a restricted set of simple rules.

Contract grant sponsor: Russian Foundation for Basic Research; Contract grant number: 95-04-11851a
*Correspondence to: Alexander V. Efimov, Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia.
E-mail: efimov@ipr.serpukhov.su
Received 1 August 1996; Accepted 3 January 1997

This paper describes possible pathways of stepwise growth of five structural motifs that recur within five protein superfamilies. All the intermediate structures connected by lines showing possible folding pathways are represented as structural trees with the corresponding motifs as root structures. Structural trees for α proteins, β proteins, and two-layer α/β proteins include more protein structures and show some novel folding pathways as compared to the folding schemes published previously.[17,18,27,32] Two structural trees for three-layer α/β proteins have been constructed for the first time. Application of the structural trees to protein classification is also discussed.

## General Rules Used in Construction of Structural Trees

The structural trees have been constructed taking into account a restricted set of general rules that have been derived from analysis of the structural features observed in globular proteins:

1. Overall folds of protein molecules and intermediate structures are taken into account and details of the structures are ignored. Each structure in the trees can have both the directions of the polypeptide chain but is drawn once for the space economy. For the space economy, only the pathways leading to known protein structures are shown.
2. A structural motif occurring in all the proteins of a given superfamily and having a unique overall fold, and unique handedness is taken as a starting structure in modeling or as a root structure of a tree.
3. The larger protein and intermediate structures are obtained by stepwise addition of α helices and/or β strands to a growing structure so that a structure obtained at the preceding step is maintained (it can be slightly modified). At each step, the α helix or β strand nearest to the growing structure along the chain is the first to be attached to the growing structure.[17,21,25,32]
4. The obtained structures should be compact; α helices and β sheets should be packed in accordance with the rules that govern their packing.[5–12]
5. The α helices and β strands cannot be packed into one layer because of dehydration of the free NH and CO groups of the β strands[7,32]; thus, an α helix should be packed into the α-helical layer and a β strand into the β layer of a growing structure.[7,15,32]
6. Crossing of connections and formation of knots are prohibited.[33,34]
7. All the structural motifs (not only root motifs) or folding units should have corresponding handedness and overall folds (for example, all the β-α-β units should form right-handed superhelices.[13,14])

Some of these or similar rules were used by Ptitsyn and colleagues[28–31] in modeling of folding pathways of all β and α/β proteins. One of the main differences between their folding schemes and the structural trees concerns the structure of initiating or starting complexes. For example, Ptitsyn and colleagues used β-β hairpins,[28–30] α-α hairpins,[30,31] and triple-strand β sheets[28,29] as initiating complexes in modeling. However, β-β hairpins and α-α hairpins can be right-turned or left-turned, and triple-strand β sheets can exist in two forms, as S-like or Z-like β sheets. In other words, these structures do not have unique overall folds and unique handedness. The other difference is that Ptitsyn and colleagues permit the addition of subsequent secondary structural segments to any edge of the initiating complex that results in many structures that do not occur in proteins. By contrast, the structural motifs having unique overall folds tend to be located at the edges of two- or three-layered protein structures with additional α helices and/or β strands predominantly arranged on one side of each motif.[17,18,21,25,32] Thus, to avoid crossing of connections, the motifs have to grow preferably in one direction (see the structural trees).

## Structural Tree for α-Helical Proteins Containing α-α Corners

Figure 1 represents a structural tree for α-helical proteins containing α-α corners. The α-α corner is a structural motif formed by two connected α helices packed approximately crosswise. In proteins, α-α corners occur practically always in one form in which its overall fold resembles nearly a turn of a left-handed superhelix. The α-α corners with short connections have a definite sequence pattern of the key hydrophobic, hydrophilic, and glycine residues, irrespective of whether they are found in homologous or

Fig. 1. Structural tree for proteins and domains containing α-α corners. α Helices are shown as cylinders and β strands as arrows. The structural information is taken from the following papers: SIV MA, SIV matrix antigen[35]; RNase Rh, ribonuclease Rh[36]; Reductase R1, ribonucleotide reductase protein R1[37]; endochitinase[38]; α-CTD, carboxyl-terminal domain of the RNA polymerase α subunit[39]; cro-repressor[40]; GH-5, globular domain of histone H5[41]; HNF-3, HNF-3/fork head DNA-recognition motif[42]; CAP, catabolite gene activator protein[43]; Tet repressor[44]; λ repressor[45]; 434 repressor[46]; Oct-1 POUs, Oct-1 POU-specific domain[47]; Dsb A, Dsb A protein[48]; Pol β (8-kD), 8-kD domain of rat DNA polymerase β[49]; Taq polymerase[50]; RAR-γ, ligand-binding domain of human retinoic acid receptor (RAR)-γ[51]; uteroglobin[52]; thermolysine[53]; citrate synthase[54]; p17 of HIV-1, p17 matrix protein of HIV-1[55]; Cyt c peroxidase, cytochrome c peroxidase[56]; Glu RS, glutamyl-tRNA synthetase[57]; hexokinase[58]; annexin, human annexin V[59]; HPS, hydrophobic protein from soybean[60]; ns-LTP, nonspecific lipid-transfer protein[61]; bLTP, lipid-binding protein[62]; KNT, kanamycin nucleotidyltransferase[63]; DT, diphtheria toxin[64]; globins[65]; colicin A, pore-forming domain of colicin A[66]; cyclin A-3[67]; TFIIB, transcription factor IIB[68]; chitosanase[69]; HSA, human serum albumin[70]; $Ca^{2+}$-binding domains[71]; papain[72]; G-peptidase, $Zn^{2+}$-containing D-alanyl-D-alanine peptidase[73]; E3/E1p, peripheral subunit-binding domain of dihydrolipoamide acetyltransferase[74].

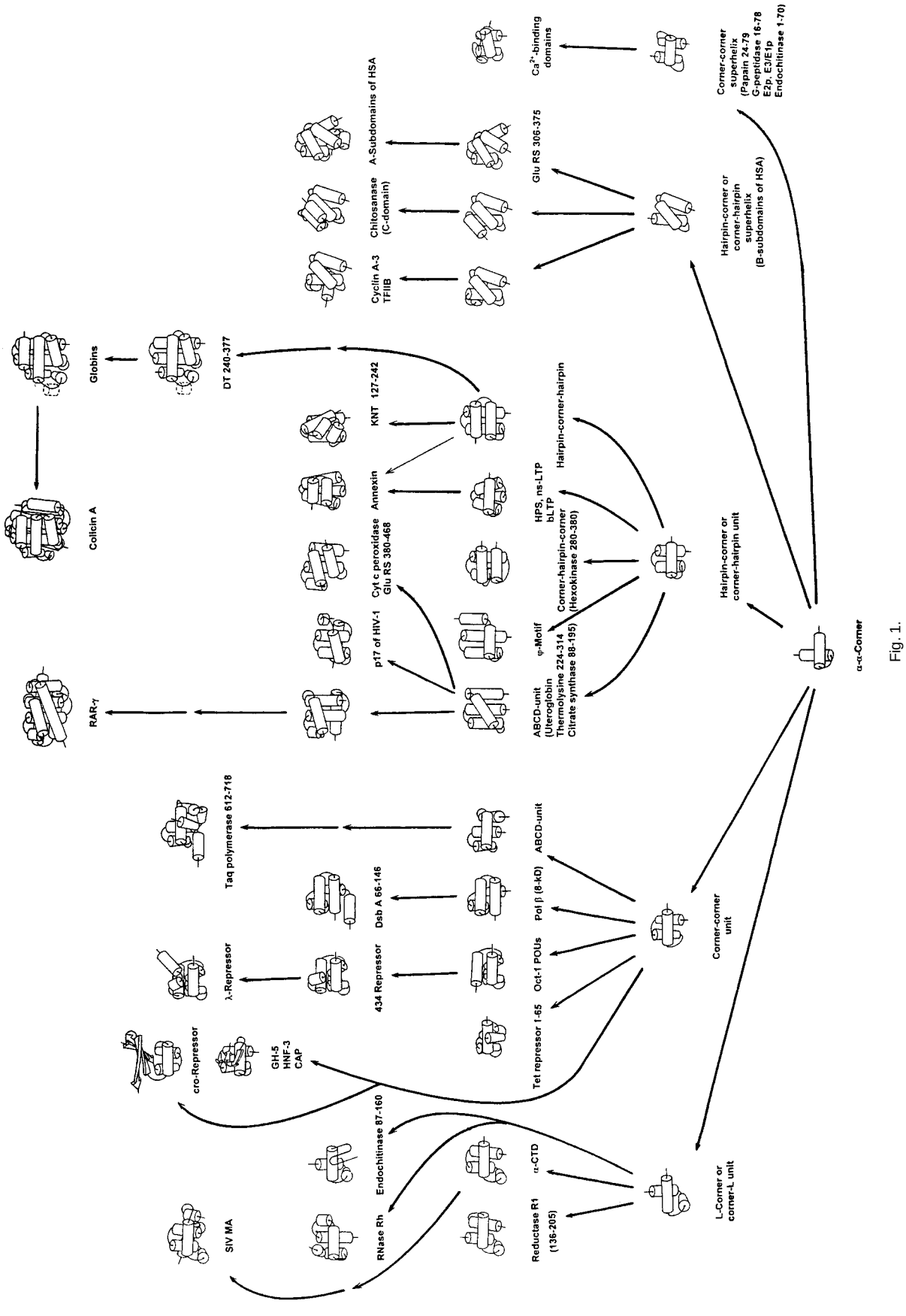
Globins
Colicin A
DT 240-377
RAR-γ
KNT 127-242
Annexin
Cyt c peroxidase Glu RS 380-468
p17 of HIV-1
Cyclin A-3 TFIIB
Chitosanase (C-domain)
A-Subdomains of HSA
Glu RS 306-375
Ca2+-binding domains
HPS, ns-LTP bLTP
Hairpin-corner-hairpin
Corner-hairpin-corner (Hexokinase 280-380)
φ-Motif
ABCD-unit (Uteroglobin Thermolysine 224-314 Citrate synthase 88-195)
Taq polymerase 612-718
λ-Repressor
Dsb A 66-146
434 Repressor
Pol β (8-kD)
Oct-1 POUs
Tet repressor 1-65
ABCD-unit
cro-Repressor
GH-5 HNF-3 CAP
SIV MA
RNase Rh
Endochitinase 67-160
α-CTD
Reductase R1 (136-205)
Hairpin-corner or corner-hairpin superhelix (B-subdomains of HSA)
Corner-corner superhelix (Papain 24-79 G-peptidase 16-78 E2p, E3/E1p Endochitinase 1-70)
Hairpin-corner or corner-hairpin unit
Corner-corner unit
L-Corner or corner-L unit
α-α-Corner

Fig. 1.

nonhomologous proteins.[18,25] Some small proteins and domains are merely composed of an α-α corner and short irregular "tails".[27] All this suggests that the α-α corner can adopt its unique structure independently of the remaining part of a protein molecule and can be a core around which the remainder of the molecule or domain is folded. So the α-α corner is used as the root structure of the structural tree shown in Figure 1.

In globular proteins, α helices pack in one of three characteristic arrangements, aligned in parallel or antiparallel, orthogonal, or slanted manner (for details, see Refs. 5–10). Taking into account these packing preferences of α helices (rule 4) and restraints on the folding imposed by the loops (rules 6 and 7), one may conclude that there is a restricted set of structures that can be formed by two α helices adjacent along the chain and connected by a short or medium-sized loop. These are α-α hairpins, α-α corners, L-shaped and V-shaped structures.[25] Thus, there is a restricted number of possible pathways of growth of the root α-α corner or intermediate structures when the next α helix is added to them (the possible pathways are shown with arrows in Fig. 1).

As seen, addition of an α helix to the root α-α corner at the first step can be done in different ways and results in formation of the structures shown in the bottom row of Figure 1. These 3α helix structures can be represented as combinations of the α-α corner and an L-shaped structure (the L–corner or corner–L unit, depending on the polypeptide chain direction), the α-α corner and an α-α hairpin (the corner–hairpin or hairpin–corner unit, the corner–hairpin or hairpin–corner superhelix) or as combinations of two α-α corners (the corner–corner unit, the corner–corner superhelix). In some cases, an α-α hairpin can be replaced by a V-shaped structure without a dramatic change of the overall fold (for the sake of space economy, such replacements are not shown). These combinations can be considered as structural motifs themselves.

Addition of an α helix to the 3α helix structures can also be done in different ways and results in formation of the 4α helix structures shown in the next row of the tree. The 4α helix structures found in uteroglobin (as well as in thermolysin [224–314] and in citrate synthase [88–195]) and in Taq polymerase have very similar overall folds but different packing angles Ω between the α helix axes. In this fold, called the ABCD unit, helices B, C, and D form a left-handed superhelix BCD, and helix A is located in between helices B and D. For comparison, the abcd unit (a structural motif of β proteins[17,32]), composed of 4β strands, a, b, c, and d, has a right-handed superhelix bcd and strand a between strands b and d (Fig. 2). As seen, the overall folds of the ABCD unit and abcd unit are rather similar but "mirror-symmetrical" if segment conformations are ignored. A detailed analysis of these and other variants of the ABCD unit, including 22 examples from known proteins, will be described elsewhere (A.V. Efimov, in preparation). Another 4α helix structure, denoted the φ motif in Figure 1, since its overall fold is reminiscent of the Greek letter phi, occurs, for example, in ribonucleotide reductase protein R2 (66–170). Another variant of the φ motif containing an α-α corner is found, for example, in ribonuclease Rh (81–144). Note that there are variants of φ motifs without α-α corners (at least 16 examples of different φ motifs with and without α-α corners are found in known proteins [unpublished observations]).

All the pathways of stepwise growth of the root α-α corner that lead to known protein structures are shown with arrows. There are several levels (or rows) in the structural tree. Each row contains structures composed of the same number of α helices (the bottom row shows the structures obtained by addition of one α helix to the root α-α corner, the structures of the next row have two α helices added etc.). As seen, the structural tree has several branches. Within one branch, protein structures having a higher position in the tree include the structures located lower. Proteins and domains of different branches have a structure located in the branching point as a common fold. The higher a branching point is located in the tree, the higher level of structural similarity between proteins of the corresponding branches is observed.

Proteins and domains whose structures can be obtained by stepwise addition of α helices to the α-α corner can be grouped into one structural class, α-helical proteins and domains containing α-α corners. Proteins and domains found within branches of the structural tree can be considered as subclasses or subfamilies. It is possible to recognize at least five subfamilies of these proteins and domains. Each subfamily has its 3α helix motif (located in the

Fig. 2. Structural tree for β proteins containing abcd units. The structures are viewed end-on with β strands shown as rectangles. Near connections are shown by double lines and far connections by single lines. Long connections are simplified and drawn by dashed lines. Segments that can be presented or not presented in a structure are also shown by dashed lines. The structural information is taken from the following papers: prealbumin[75]; transthyretin[76]; CGT, cyclodextrin glycosyltransferase[77]; 3,4 PCD, protocatechuate 3,4-dioxygenase[78]; Cyo A, subunit II of quinol oxidase[79]; amicyanin[80]; plastocyanin[81]; azurin[82]; AO, ascorbate oxidase[83]; thaumatin I[84]; Hoe-467A, α-amylase inhibitor Hoe-467 A[85]; immunoglobulins[86]; NCD 1, N-cadherin domain1[87]; sCD2, cell adhesion molecule CD2[88]; SOD, copper, zinc superoxide dismutase[89]; Pap D, chaperone protein PapD[90]; zeamatin[91]; CAP, catabolite gene activator protein[43]; SPMV, satelite panicum mosaic virus coat protein[92]; STMV, satellite tobacco mosaic virus capsid protein[93]; PNG F, glycosyl-asparaginase[94]; phaseolin[95]; BBV, BBV capsid protein[96]; TBSV, tomato bushy stunt virus coat protein[97]; SBMV, southern bean mosaic virus coat protein[98]; Cry IA(a) toxin[99]; β-glucuronidase[100]; γ-crystallin[101]; protein S[102]; α-amylase[103]; dUTP, dUTPase[104]; TRAP, trpRNA-binding attenuation protein[105]; CBD, cellulose-binding domain[106]; p53 core domain, p53 tumor suppressor[107]; cytochrome f[108].

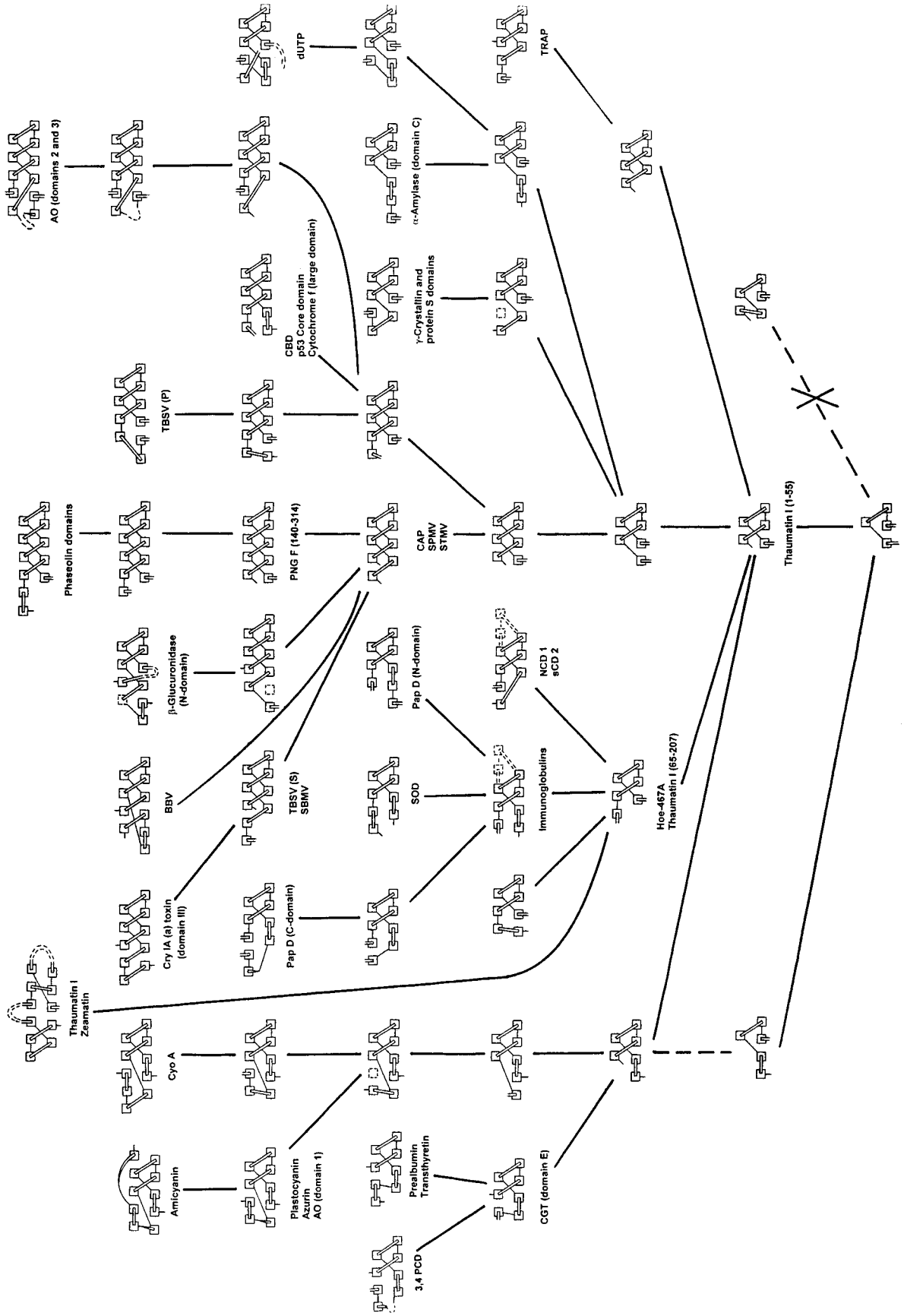
AO (domains 2 and 3)
dUTP
TRAP
α-Amylase (domain C)
γ-Crystallin and protein S domains
CBD
p53 Core domain
Cytochrome f (large domain)
TBSV (P)
Phaseolin domains
PNG F (140-314)
CAP
SPMV
STMV
β-Glucuronidase (N-domain)
Pap D (N-domain)
NCD 1
sCD 2
BBV
TBSV (S)
SBMV
SOD
Immunoglobulins
Hoe-467A
Thaumatin I (65-207)
Thaumatin I (1-55)
Cry IA (a) toxin (domain III)
Pap D (C-domain)
Thaumatin I
Zeamatin
Cyo A
Amicyanin
Plastocyanin
Azurin
AO (domain 1)
Prealbumin
Transthyretin
3,4 PCD
CGT (domain E)

Fig. 2.

branching point) as a common fold. Most proteins and domains containing the corner–corner unit have the function to bind DNA. Most proteins and domains within other subfamilies have different functions despite their structural similarity.

### Structural Tree for β Proteins Containing abcd Units

The abcd unit is a structural motif recurring in two-layer β proteins and β domains with the aligned β sheet packing.[17] The simplest variant of the abcd unit consists of four consecutive β strands, a, b, c, and d, arranged in three dimensions so that strands b, c, and d form a right-handed superhelix bcd, and strand a is always situated between strands b and d antiparallel to them (see Fig. 2). In proteins, there are more complex variants of the abcd unit.[17,25–27] For example, abcd units found in domain $V_H$ of immunoglobulins and in domain 1 of the cell adhesion molecule CD2 have triple-strand β sheets instead of strands c.

The simplified depiction of an abcd unit on a plane, using Richardson's approach,[34] results in the so-called Greek key topology. However, a long β hairpin bent in half and several other three-dimensional arrangements of β strands also have the Greek key topology[e.g.,17,25]. To avoid this ambiguity in transition from a topology to a three-dimensional structure, it is necessary to analyse and compare the three-dimensional arrangements of β strands. The abcd unit has a unique arrangement of its β strands and a unique handedness (superhelix bcd is always right-handed). Analysis of known bilayer β proteins with the aligned β sheet packing shows that the abcd unit is always located at the edge of the double layer, and the other strands are situated on the side of the d strand. So, when other β strands are added to the abcd unit step by step, it looks as if the abcd unit grows in one direction.

The strands joined to strand a of the abcd unit are labeled here as $a_1$, $a_2$, $a_3$, . . . and the strands joined to strand d as $d_1$, $d_2$, $d_3$, . . . according to their distance from strands a or d in the polypeptide chain irrespective of the chain direction. Their possible three-dimensional arrangements can be obtained by stepwise addition to the abcd unit taking into account the rules listed above. Rule 7 should be additionally explained here. In β proteins, not only strands b, c, and d but some other three consecutive β strands can form a right-handed β-β-β superhelix analogous to the β-α-β superhelix. It is permitted if the first and third β strands of such a β-β-β superhelix do not directly interact, and there is at least one additional β strand in the β layer between them (for details and exceptions, see Refs. 17 and 32). For example, the right-handed superhelix bcd of the abcd unit has strand a between strands b and d. On the one hand, the structure shown in the bottom right corner of Figure 2 (the dashed line from the root abcd unit leads to it) is forbidden, since it has superhelix $cdd_1$ in which strands c and $d_1$ directly interact between each other.

Addition of strand $a_1$ (joined to strand a) to the root abcd unit results in the structure shown in the center of the bottom row in Figure 2 (domain 1–55 of thaumatin I has this fold). Strands $a_1$ are arranged in this way in all the proteins in which they are present. Strand $a_1$ cannot be packed in the bottom layer next to strand d as crossing of loops $aa_1$ and $dd_1$ would occur (rule 6). Strand $a_1$ also cannot be packed on the other side of strand c in the upper layer or next to strand b in the bottom layer (i.e., at the edge) as loops $aa_1$ and bc would cross. Addition of strand $d_1$ to the root abcd unit results in the structure shown on the left in the bottom row of the tree. As mentioned above, strand $d_1$ cannot be packed in the upper layer next to strand c (rule 7). However, strand $d_1$ can be packed in the upper layer next to strand $a_1$ as, for example, in the structure shown on the right in the second row of the tree and found in trpRNA-binding attenuation protein (TRAP). It is possible as there is strand $a_1$ between strands c and $d_1$ in the right-handed superhelix $cdd_1$.

Similarly, arrangements of other β strands relative to the root abcd unit can be obtained if the next strands are added to it step by step. In each step, the next strand can be packed in the upper or bottom layer, depending on the structure that has been formed in the preceding step. Possible pathways of the growth that lead to the known protein structures are represented by thick lines in Figure 2. It looks very likely that domains 1–55 and 65–207 of thaumatin I fold independently of each other, and their interaction results in the completed molecule (apparently, zeamatin folds in a similar way). It should be noted that the abcd unit can grow into larger structures if α-helical or irregular segments are added to it (not shown in Fig. 2). Structures of many small proteins and domains can be obtained in this way. Many small proteins and domains are merely composed of abcd units.[27] As seen, there are several branches of the structural tree. Protein structures found within the largest branches having the corresponding six-stranded structures as common folds (see the second row of the tree) can be grouped into subfamilies.

### Structural Tree for α/β Proteins Containing abCd Units

The abCd unit is a variant of the abcd unit that has helix C instead of strand c.[17,25,32] It is also called a single split βαβ motif.[23] The α/β proteins containing abCd units and β proteins containing abcd units have very much in common. First of all, proteins of both classes have similar commonly occurring structural motifs, abcd and abCd units. In both classes these units tend to be located at the edges of molecules and most of other α helices and/or β strands

are situated on the sides of the d strands of the units. Many proteins and domains of these classes have very similar overall folds if segment conformations are ignored.[32]

Possible pathways of growth of the root abCd unit are represented in Figure 3. There are more ways of addition of the first secondary structural segment to the abCd unit as compared with the first step of the abcd unit growth. If a molecule or a domain has helix $A_1$ joined to strand a, it is packed next to helix C in the α-helical layer (rule 5) or below the β sheet (i.e., on the β sheet side that is opposite to the side where helix C is packed), giving rise to another α-helical layer. Helix $A_1$ cannot be packed on the other side of helix C in the α-helical layer as loops $aA_1$ and bC would cross. Helix $A_1$ also cannot be packed into the β layer (rule 5). If helix $A_1$ is absent from a molecule, while helix $D_1$ joined to strand d is present, it is also packed in the α-helical layer next to helix C or below the β sheet. In accordance with rules 5 and 6, there are two modes of packing of strand $d_1$. It can be packed next to strand d at the edge of the β sheet or between strands a and d as in ubiquitin and protein G. Strand $a_1$ joined to strand a can be packed below the β sheet, giving rise to another β layer as in histidine decarboxylase. The ways of addition of other α helices or β strands to the growing structures can easily be observed in Figure 3. Note that all the obtained β-α-β units form right-handed superhelices in accordance with rule 7.

### Structural Tree for α/β Proteins Containing Seven-Segment α/β Motifs

According to the structural classification of Levitt and Chothia,[15] the class of α/β proteins includes the proteins whose structures are composed of α helices and β strands that roughly alternate along the polypeptide chain. The α/β proteins can be grouped into three subclasses, depending on their overall folds, the so-called α/β barrels, three-layer α/β proteins, and two-layer α/β proteins.[e.g., 16] In the α/β barrel proteins, the parallel β sheet is twisted and coiled so that a barrel structure is formed and the α helices are packed on its outside surface. The structures of two-layer α/β proteins have been considered in the preceding section of this paper. The three-layer α/β proteins have α helices packed on both sides of a parallel β sheet. In my opinion, the three-layer α/β proteins can be subdivided in two groups, α/β proteins containing five-segment α/β motifs and those containing seven-segment α/β motifs. Structural trees for α/β proteins of these two groups are described in this and the following sections of the paper.

The seven-segment α/β motif is a structural motif composed of 4β strands and 3α helices that are folded into βαβαβ and αβ units and arranged so that the β strands form a parallel β sheet and the α helices are packed on both sides of the β sheet.[25,26] In known proteins, this motif is always located at the edges of three-layer structures. Possible ways of growth of this motif into larger protein structures are shown in Figure 4. Most of these protein structures can be obtained by stepwise addition of α helices and β strands to the root motif, taking into account rules 1–7, although there are some exceptions. 4-Epimerase, 3α,20β-hydroxysteroid dehydrogenase (HSD), 17β-HSD, mouse lung carbonyl reductase (MLCR), enoyl acyl carrier protein reductase (ENR), and dihydropteridine reductase (DHPR) contain left-handed β-α-β units. Similar to other left-handed β-α-β units found in proteins,[14] they also have long connection regions (shown with dashed lines in Fig. 4).

In most growing structures the added α helices are packed in the upper α layers. It looks like the upper α layers and β sheets grow in the first place. The growth of a structure is finished completely or in one or two steps when the next α helix is packed in the bottom α layer. This refers primarily to the structures of the central branch and the branches shown in the left half of the tree (see also domains of guanosine monophosphate (GMP) synthetase, pyruvate kinases, pyruvate oxidase, leucine dehydrogenase (Leu DH), *Vaccinia* protein (VP 39) within branches of the right half of the tree).Thus, the formation of compact and closed folds results in the completed protein structures. All proteins and domains within branches of the right half of the tree also have a third α helix in the bottom α layers. Additional compactness and closeness of the structures are achieved as a result of twisting and coiling of the β sheets.

### Structural Tree for α/β Proteins Containing Five-Segment α/β Motifs

The five-segment α/β motif is a structural motif consisting of 3β strands and 2α helices folded into two β-α-β units (with the central β strand shared) and arranged in a three-layer structure having a parallel β sheet and α helices packed on its both sides.[25,26] Similar to the other structural motifs considered here, this motif tends to be located at the edges of the three-layer structures of known proteins (for some examples of proteins where this is not so, see thioredoxinlike structures). Possible pathways of stepwise growth of the root five-segment α/β motif that lead to known protein structures are shown in Figure 5. There are left-handed β-α-β units in the large domain of phosphofructokinase (PFK), P domain of L-arabinose-binding protein (ABP), periplasmic glucose/galactose-binding protein (GBP), N domain of asparaginase II, acetylcholinesterase (AChE), *Candida rugosa* lipase (CRL), and human lipase. All these units have long connections and are formed at the last steps of the growth. The growth of most structures is finished when one or two α helices are

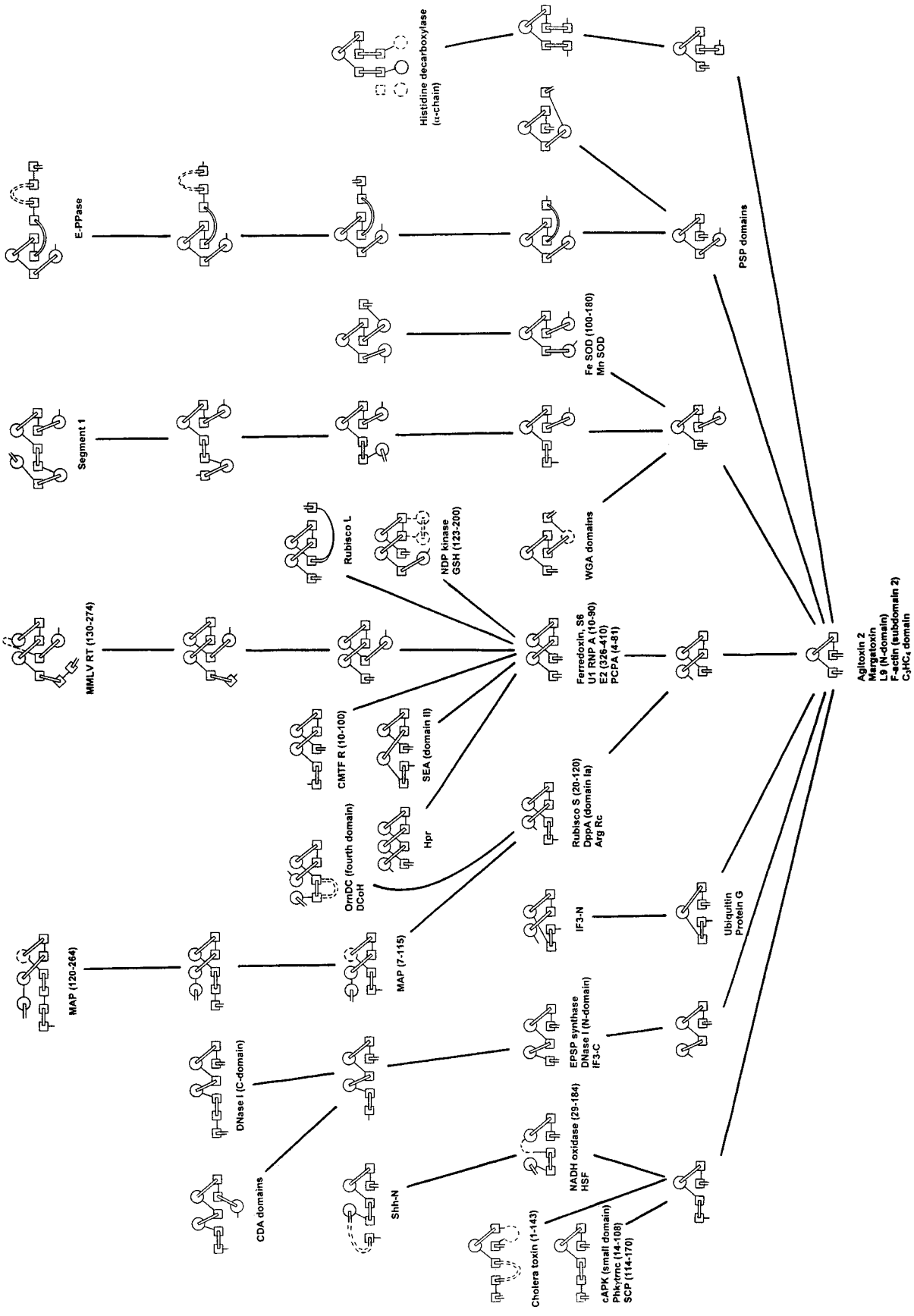
Histidine decarboxylase (α-chain)
E-PPase
PSP domains
Fe SOD (100-180)
Mn SOD
Segment 1
Rubisco L
NDP kinase
GSH (123-200)
WGA domains
MMLV RT (130-274)
Ferredoxin, S6
U1 RNP A (10-90)
E2 (326-410)
PCPA (4-81)
Agitoxin 2
Margatoxin
L9 (N-domain)
F-actin (subdomain 2)
$C_3HC_4$ domain
CMTF R (10-100)
SEA (domain II)
Hpr
OrnDC (fourth domain)
DCoH
Rubisco S (20-120)
DppA (domain Ia)
Arg Rc
MAP (120-264)
MAP (7-115)
IF3-N
Ubiquitin
Protein G
DNase I (C-domain)
EPSP synthase
DNase I (N-domain)
IF3-C
CDA domains
Shh-N
NADH oxidase (29-184)
HSF
Cholera toxin (1-143)
cAPK (small domain)
Phk-trnc (14-108)
SCP (114-170)

Fig. 3.

packed in their bottom α layers and closed structures are formed.

## DISCUSSION

This paper describes five structural trees for five protein superfamilies, although there are some more known motifs having unique overall folds that are able to grow into the larger structures in a similar way. One of them is the so-called 3β corner that recurs in β proteins with the predominantly orthogonal β sheet packing. The 3β corner can be represented as a triple-stranded β sheet folded on to itself so that its two β-β hairpins are packed approximately orthogonally in different layers and the central strand bends by ~90° in a right-handed direction when passing from one layer to the other. All the 3β corners observed in proteins, when viewed from their concave surfaces, can be considered as formed by Z-like β sheets. An initial variant of a structural tree for proteins containing 3β corners, such as serine proteases, β-lactoglobulin and its homologues, bovine neurophysin, and SH3 domains, has been published.[21] The construction of an up-to-date structural tree for such proteins is in progress now. There are also some structural motifs with unique overall folds that involve S-like β sheets. These are βSβ, βSα, αSα, Sαβ, and βαS superhelices (for details, see Refs. 22 and 26). They can be taken as root motifs in constructing the corresponding structural trees. Some examples of protein structures that can be obtained from these motifs are represented in References 22 and 32. Simple structural trees can be constructed for the α/β barrel proteins starting from a β-α-β unit and, similarly, for α/α barrel proteins and domains (such as the U domain of soluble lytic transglycosylase, helical domain of lipovitellin-phosvitin, bacterial cellulase, insecticidal δ-endotoxin, and glucoamylase) in which α helices form double-layered right-handed superhelices. Structures of parallel 4α helix bundles and other proteins with the aligned packing of α helices (e.g., bacteriorhodopsin) can be represented as combinations of right-turned and left-turned α-α hairpin. The problem is what α-α hairpin should be taken as the starting one. In the cases of short connections, it is possible to determine whether the α helices are packed into a right- or left-turned α-α hairpin, as they have different sequence patterns for the key hydrophobic, hydrophilic, and glycine residues.[25,242] So it seems reasonable to start from α-α hairpins with short connections, as they can fold into unique structures themselves and can form hydrophobic surfaces where the next α helices should be packed. If all the connections are long and flexible, it is not clear yet which α-α hairpin would be the starting one.

One of the most important results of this paper is that practically all known protein structures of the considered superfamilies can be obtained by stepwise addition of α helices and/or β strands to the corresponding structural motifs in accordance with the same set of rules. Thus, the motifs with unique overall folds may be taken as starting structures in protein modeling and prediction studies and the structural trees can be used to search for all possible protein structures.

It can be seen from analysis of the structural trees that at each step there is a restricted number of pathways of growing for the structure obtained at the preceding step. Although the structural trees represent only the pathways that lead to known protein structures, it is evident that the rules used in constructing the trees reduce drastically the number of possible structures which can be obtained from one structural motif. It seems likely that the number of structural motifs having unique overall folds and unique handedness is also limited. So it can be concluded that the number of possible overall folds that may occur in proteins is limited while the number of proteins may be very large, since there is a lot of "details" that make them different (e.g., proteins may have different lengths of α helices and β strands, different lengths and conformations of loops, their sequences may differ). This seems to be consistent with the conclusions made by other authors.[243,244]

Levels of structural similarity between different proteins and domains can easily be observed by

Fig. 3. Structural tree for α/β proteins containing abCd units. The structures are viewed end-on with α helices shown as circles and β strands as rectangles. Near connections are shown by double lines and far connections by single lines. Long loops are drawn by dashed lines. If α helices are very short or replaced by irregular regions, the circles are drawn by dashed lines (see, e.g., MAP and WGA domains). The structural information is taken from the following papers: cholera toxin, the A1 chain of cholera toxin[109]; cAPK, cAMP-dependent protein kinase[110]; Phkγtrnc, the γ-subunit of rabbit muscle phosphorylase kinase[111]; SCP, Sindbis virus core protein[112]; Shh-N, N-domain of murine Sonic hedgehog[113]; NADH oxidase[114]; HSF, heat shock transcription factor[115]; Orn DC, ornithine decarboxylase from *Lactobacillus* 30a[116]; DCoH, protein-binding transcriptional coactivator[117]; MAP, methionine aminopeptidase[118]; Rubisco, ribulose 1,5-biphosphate carboxylase/oxygenase (L and S subunits)[119]; DppA, dipeptide-binding protein[120]; Arg Rc, L-arginine binding domain of arginine repressor[121]; NDP kinase, nucleoside diphosphate kinase[122]; GSH, glutathione synthetase[123]; MMLV RT, MMLV reverse transcriptase[124]; CMTF R, regulatory chain of aspartate carbamoyltransferase[125]; SEA, staphylococcal enterotoxin A[126]; Hpr, histidine-containing phosphocarrier protein[127]; ferredoxin[128]; U1 RNP A, U1 small nuclear ribonucleoprotein A[129]; S6, ribosomal protein S6[130]; E2, bovine papilloma virus-1 E2 DNA-binding domain[131]; PCPA, procarboxypeptidase A[132]; agitoxin 2[133]; L9, N-domain of ribosomal protein L9[134]; F-actin[135]; $C_3HC_4$, the $C_3HC_4$ domain[136]; CDA, cytidine deaminase[137]; EPSP synthase, 5-enol-pyruvylshikimate-3-phosphate synthase[138]; DNase I, deoxyribonulease I[139]; IF3-C and IF3-N, C- and N-domains of translational initiation factor IF3[140]; E-PPase, *E. coli* soluble inorganic pyrophosphatase[141]; PSP, porcine pancreatic spasmolytic polypeptide[142]; segment 1, segment 1 domain of gelsolin[143]; WGA, wheat germ agglutinin[144]; Fe SOD, iron superoxide dismutase[145]; Mn SOD, manganese superoxide dismutase[146]; ubiquitin[147]; protein G, B2 domain of protein G[148]; histidine decarboxylase[149].

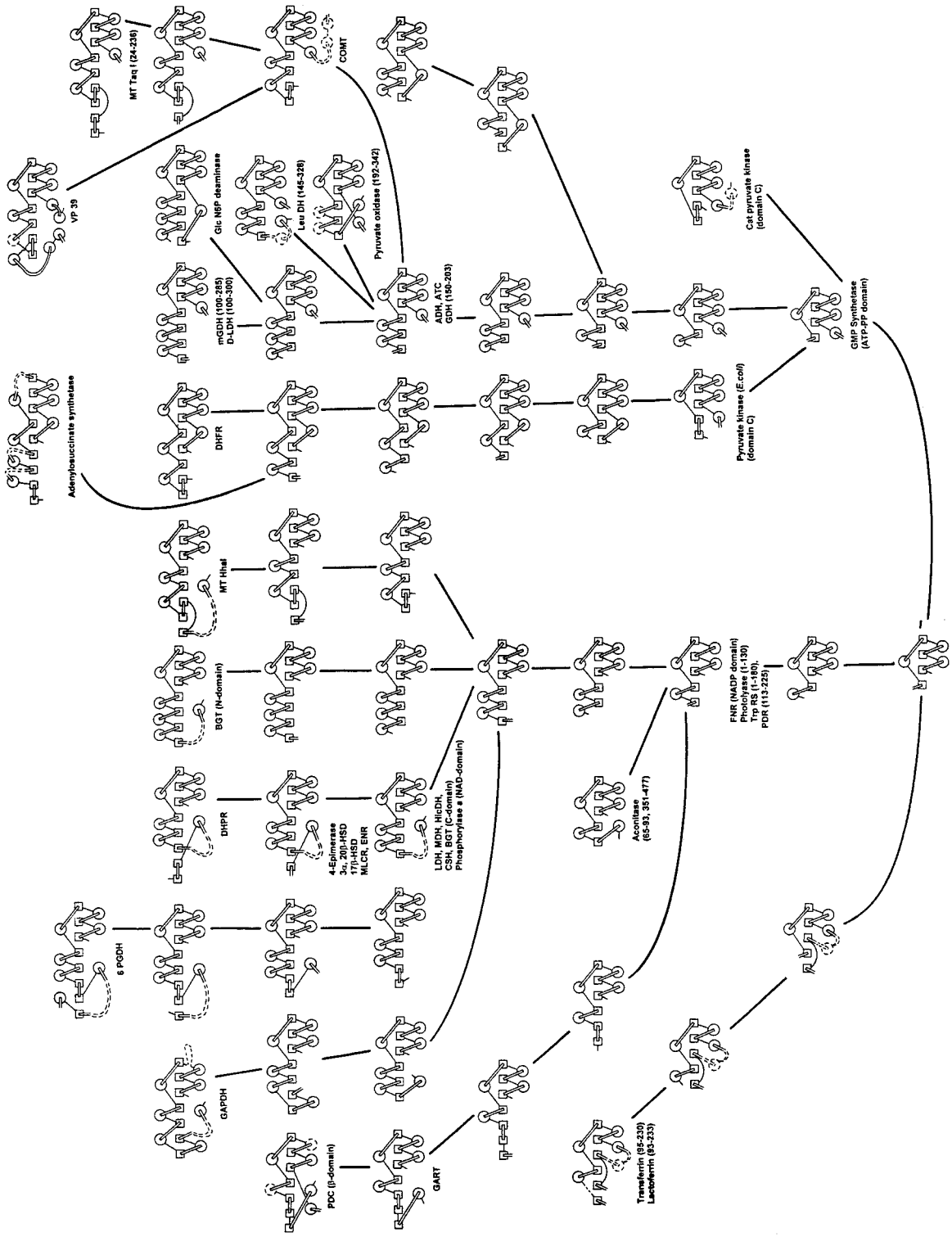
MT Taq I (24-236)
COMT
VP 39
Glc N6P deaminase
Leu DH (145-328)
Pyruvate oxidase (192-342)
mGDH (100-285)
D-LDH (100-300)
ADH, ATC
GDH (150-203)
Cat pyruvate kinase
(domain C)
GMP Synthetase
(ATP-PP domain)
Adenylosuccinate synthetase
DHFR
Pyruvate kinase (E.coli)
(domain C)
MT HhaI
BGT (N-domain)
DHPR
4-Epimerase
3α, 20β-HSD
17β-HSD
MLCR, ENR
LDH, MDH, HicDH,
CSH, BGT (C-domain)
Phosphorylase a (NAD-domain)
Aconitase
(65-93, 351-477)
FNR (NADP domain)
Pholobyase (1-130)
Trp RS (1-160),
PDR (113-225)
Aconitase (203-321)
6 PGDH
GAPDH
PDC (β-domain)
GART
Transferrin (55-239)
Lactoferrin (93-233)

Fig. 4.

visual inspection of the trees. Within one branch, protein structures having a higher position in the tree include structures located lower. Proteins and domains of different branches have a structure situated in the branching point as the common fold etc. The structural trees give a clue to understanding the reasons of structural similarity between unrelated proteins. The main reason of their similarity is that all the proteins within a superfamily have a common fold, the corresponding root structural motif, starting from which all their structures can be obtained. A number of examples of structural similarity between unrelated proteins found within one structural tree (within one structural class) and in different classes (for details, see Ref. 32) as well as the fact that the main rules used in constructing the trees were inferred from known principles of the protein structure suggest that some physical principles rather than the evolutionary divergence or functional convergence of proteins are the basis of the similarity.[29,243] There are also many examples of protein structures that have the same or very similar overall folds but different directions of the polypeptide chains. It is clear that structural similarity between such proteins cannot be explained by their evolutionary divergence as a protein molecule cannot change the direction of its polypeptide chain during its evolution.

As mentioned above, the structural trees represent only the pathways leading to the known protein structures. In principle, each pathway can be inferred from analysis of stepwise decomposition of the corresponding protein structure. However, there are some protein structures which can be obtained by at least two different pathways. The main reason of this is that many growing structures have α helices and/or β strands joined to their both N and C termini. In some cases, it is difficult to determine what segment is the first to be attached to the growing structure. For example, Figure 1 shows that the annexin structure can be obtained in two ways: (1) hairpin–corner → HPS-like structure → annexin; (2) hairpin–corner → hairpin–corner–hairpin → annexin. Figure 2 shows that protein structures of the left branch can be obtained through the thaumatin (1–55)-like structure or through the fiveβ stranded structure located in the bottom left corner of Figure 2 (for some other similar cases, see Ref. 32). As seen from visual inspection of the structural tree shown in Figure 5, the left parts of human lipase, haloalkane dehalogenase, dienelactone hydrolase (DLH), bromoperoxidase A2 (BPO-A2), AChE, CRL, human "protective protein" (HPP), and wheat serine carboxypeptidase II (CPDW-II) have the 11-segment common fold that is also found in subtilisin and located in the central branch of the tree (in the seventh row). Thus, the structures of these proteins can be obtained by stepwise addition of the rest segments to this common fold if their left parts are assumed to be formed first. However, Figure 5 is drawn so that the right parts of these proteins are formed first. It was done so in order to show some other pathways of growth of the root five-segment α/β motif. Such a representation of the tree shows rather dissimilarity than similarity of these protein structures. For classification of these proteins, it is necessary to take into account that their structures can be obtained from the above mentioned 11-segment fold, since this pathway shows a higher level of their structural similarity. It seems likely that the left and right parts of these proteins can fold independently of each other and simultaneously. It should be also noted that different motifs can coexist in the same protein molecule or domain. For example, the SH3 domain contains a right-handed β-S-β superhelix and a 3β corner (for details, see Refs. 22 and 26). Analysis of structures of the circular permutants constructed by Viguera and coworkers[245,246] shows that each of them also contains a 3β corner and a β-S-β superhelix, which are formed, however, by other β strands as compared with those of wtSH3 domain. In theory, all these structures can be obtained by stepwise addition of β strands to the 3β corner or to the S-like β sheet. This seems to be consistent with the main conclusion of Viguera and coworkers[246] that different folding transition states may result in the same native structure.

Analysis of all the data presented above has led us to a hypothesis that the structural motifs having unique overall folds can fold independently of the

Fig. 4. Structural tree for α/β proteins containing seven-segment α/β motifs. The structures are viewed end-on with α helices shown as circles and β strands as rectangles. Near connections are shown by double lines and far connections by single lines. Long loops (which may include an α helix or a β strand) are shown by dashed lines. If α helices are very short or irregular, the circles are drawn by dashed lines. The structural information is taken from the following papers: transferrin[150]; lactoferrin[151]; GART, glycinamide ribonucleotide transformylase[152]; PDC, pyruvate decarboxylase[53]; GAPDH, glyceraldehyde phosphate dehydrogenase[154]; BGT, β-glucosyltransferase[155]; 6 PGDH, 6-phosphogluconate dehydrogenase[156]; aconitase[157]; DHPR, dihydropteridine reductase[158]; 4-Epimerase[159]; 3α,20β-HSD, 3α,20β-hydroxysteroid dehydrogenase[160]; 17-β-HSD, 17-β-hydroxysteroid dehydrogenase[161]; MLCR, mouse lung carbonyl reductase[162]; ENR, enoyl acyl carrier protein reductase[163]; FNR, ferredoxin-$NADP^+$ reductase[164]; photolyase, DNA photolyase from *E. coli*[165]; Trp RS, tryptophanyl-tRNA synthetase[166]; PDR, phthalate dioxygenase reductase[167]; MT HhaI, HhaI DNA methyltransferase[168]; adenylosuccinate synthetase[169]; DHFR, dihydrofolate reductase[170]; pyruvate kinase (*E. coli*)[171]; Cat pyruvate kinase[172]; MT Taq I, DNA methyltransferase MTaq I[173]; COMT, catechol *O*-methyltransferase[174]; ADH, horse liver alcohol dehydrogenase[175]; ATC, aspartate carbamoyltransferase[176]; GDH, glucose dehydrogenase[177]; GMP synthetase[178]; mGDH, D-glycerate dehydrogenase[179]; D-LDH, D-lactate dehydrogenase[180]; VP 39, *Vaccinia* protein VP 39[181]; Glc N6P deaminase, glucosamine 6-phophate deaminase[182]; Leu DH, leucine dehydrogenase[183]; pyruvate oxidase[184]; LDH, dogfish $M_4$ apo-lactate dehydrogenase[185]; MDH, malate dehydrogenase[186]; HicDH, L-2-hydroxyisocaproate dehydrogenase[187]; CSH, *N*-carbamoylsarcosine amidohydrolase[188]; phosphorylase a, glycogen phosphorylase a[189].

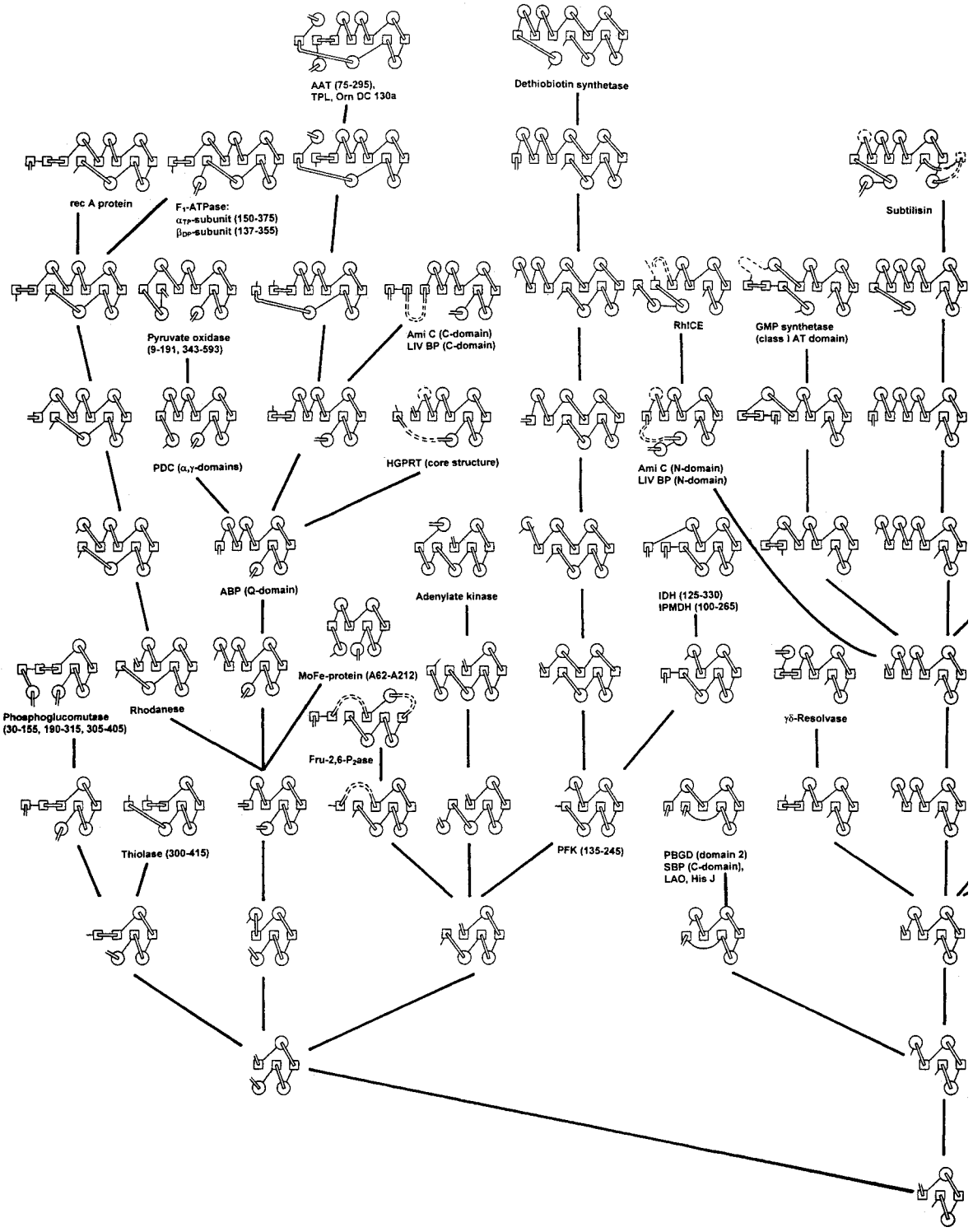


Fig. 5. Structural tree for $\alpha/\beta$ proteins containing five-segment $\alpha/\beta$ motifs. The structures are represented similar to that in Figure 4. The structural information is taken from the following papers: rec A protein[190]; phosphoglucomutase[191]; $F_1$-ATPase, bovine mitochondrial $F_1$-ATPase[192]; pyruvate oxidase[184]; PDC, pyruvate decarboxylase[153]; rhodanese[193]; MoFe-protein, nitrogenase molybdenum-iron protein[194]; thiolase, peroxisomal 3-ketoacyl-CoA thiolase[195]; AAT, aspartate aminotransferase[196]; TPL, tyrosine phenol-lyase[197]; Orn DC 130a, ornithine decarboxylase[198]; ABP, L-arabinose-binding protein[199]; Fru-2,6-$P_2$ase, rat liver fructose-2,6-biphosphatase[200]; Ami C, negative regulator of the amidase operon[201]; LIV BP, Leu/Ile/Val-binding protein[202]; HGPRT, human hypoxanthine-guanine phosphoribosyltransferase[203]; adenylate kinase[204]; PFK-,phosphofructokinase[205]; dethiobiotin synthetase[206]; PBGD, porphobilinogen deaminase[207]; SBP, sulfate-binding protein[208]; LAO, lysine-, arginine-, ornithine-binding protein[209]; His J, histidine-binding protein[210]; IDH, isocitrate dehydrogenase[211]; IPMDH, 3-iso propylmalate dehydrogenase[212]; GMP synthetase[178]; Rh ICE, interleukin-1$\beta$-converting enzyme[213]; $\gamma\delta$-resolvase[214]; subtilisin[215]; Fru-1,6-Pase, fructose-1,6-biphosphatase[216]; IM, inositol monophosphatase[217]; flavodoxin[218]; Che Y, Che Y protein[219]; ras p21, guanine-nucleotide-binding domain of p21[220]; GBP, periplasmic

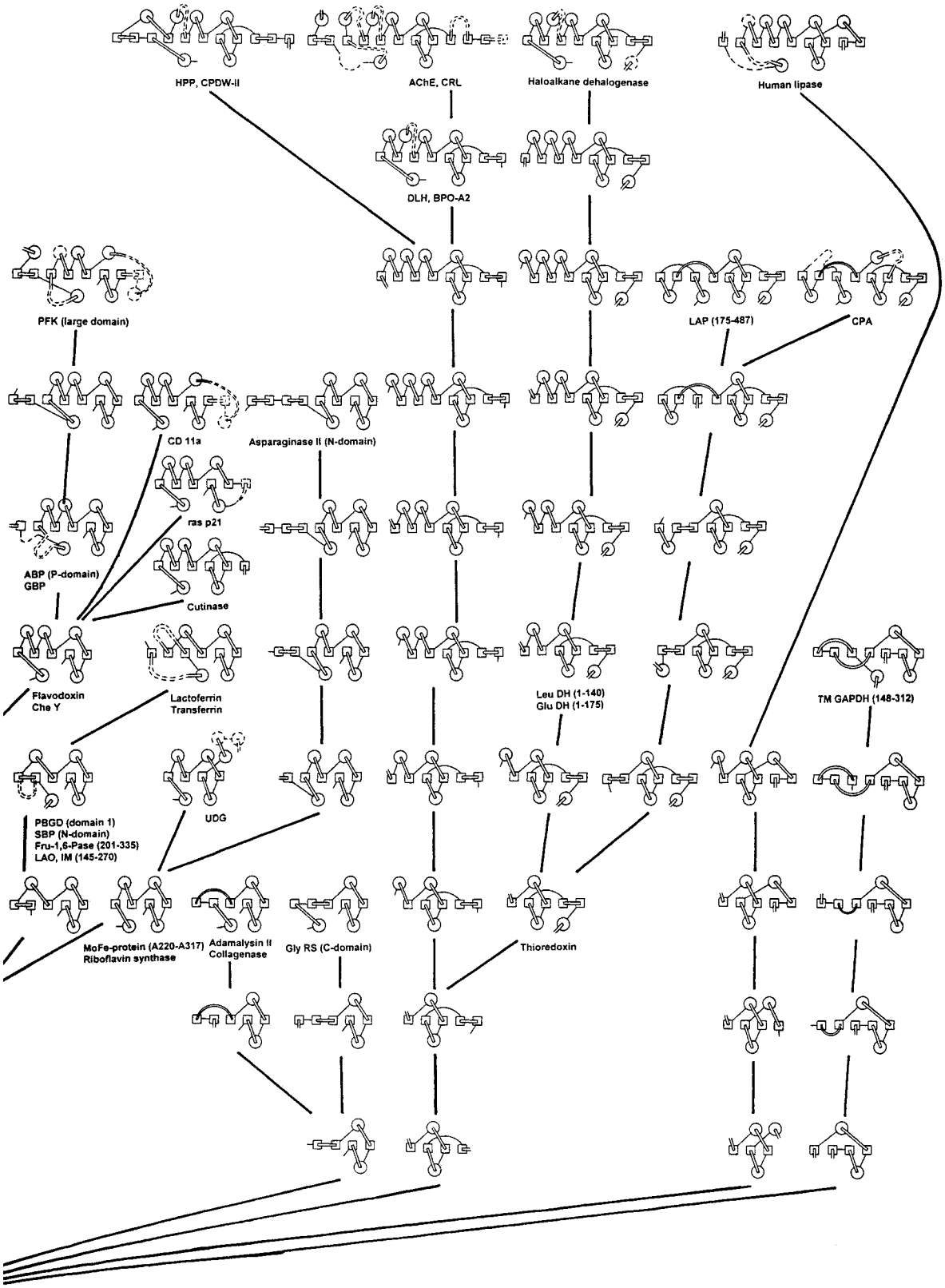


Fig. 5. (Continued) glucose/galactose-binding protein[221]; riboflavin synthase[222]; UDG, uracil-DNA glycosylase[223]; lactoferrin[151]; transferrin[150]; CD11a, CD11a I-domain of integrin[224]; cutinase[225]; adamalysin II[226]; collagenase[226]; Gly RS, glycyl-tRNA synthetase[227]; asparaginase II[228]; thioredoxin[229]; Leu DH, leucine dehydrogenase[183]; Glu DH, glutamate dehydrogenase[230]; haloalkane dehalogenase[231]; ACh E, acetylcholinesterase[232]; CRL, *Candida rugosa* lipase[233]; DLH, dienelactone hydrolase[234]; BPO-A2, bromoperoxidase A2[235]; HPP, human ''protective protein''[236]; CPDW-II, wheat serine carboxypeptidase II[237]; LAP, leucine aminopeptidase[238]; CPA, carboxypeptidase A[239]; Human lipase[240]; TM GAPDH, D-glyceraldehyde-3-phosphate dehydrogenase[241].

remaining parts of molecules and can act as nuclei or "ready-made" building blocks in protein folding and the pathways of their stepwise growth can be considered as possible folding pathways of the proteins. The high frequency of occurrence of the structural motifs in unrelated proteins (see, e.g., Figs. 1–5) and the fact that many small proteins and domains merely consist of the motifs[27] suggest that they are relatively stable and can fold into unique structures per se. Nuclear magnetic resonance (NMR) studies of partially folded proteins can give insight into early events in protein folding. For example, Barbar and colleagues[247] have shown that there is a stable core of the antiparallel β sheet in partially folded BPTI, which is proposed to be an intermediate in BPTI folding. Intermediate structure II, shown in Figure 1 of Reference 247, is essentially a variant of the abcd unit. The fact that protein structures of the considered superfamilies can be obtained from the corresponding structural motifs in accordance with the same set of rules also supports the hypothesis, although the order of events in protein folding may be different from that presented in the trees.

The structural trees can be used for structural classification of proteins. Proteins and domains whose structures can be obtained by stepwise addition of α helices and/or β strands to the same root motif can be grouped into one structural class or superfamily. Proteins and domains found within branches of a structural tree can be considered as subclasses or subfamilies. Note that some protein structures can be obtained in at least two pathways. As seen, this classification is different from those suggested by other authors[248–251] (reviewed in Refs. 252–254) in some aspects. First of all, amino acid sequences of proteins are not taken into account in this classification. This classification is based on similarity of overall folds and possible folding pathways (strongly speaking, on modeled pathways of stepwise growth of the motifs). Details of protein structures are ignored. Such an approach permits the structural classification of unrelated proteins but, on the other hand, makes this classification difficult to automate. Note that two structures having the same or similar overall folds can have a rather large value of the root-mean-square deviation, since their α helices and/or β strands may be of different lengths, and their connection regions may differ in length and conformation. Two or more protein structures can have the same or similar overall folds, but different conformations of some regions (e.g., an α helix can be replaced by an irregular or β-structural region). Long loops may contain α helices and/or β strands themselves. Nevertheless, I hope this approach offers an unusual and stimulating perspective regarding the structural classification of proteins and their folding pathways.

## ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, grant 95-04-11851a.

## REFERENCES

1. Lim, V.I. Algorithm for prediction of α-helical and β-structural regions in globular proteins. J. Mol. Biol. 88:873–894, 1974.
2. Chou, P.Y., Fasman, G.D. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. 47:45–148, 1978.
3. Ptitsyn, O.B., Finkelstein, A.V. Theory of protein secondary structure and algorithm of its prediction. Biopolymers 22:15–25, 1983.
4. Sternberg, M.J.E. Secondary structure prediction. Curr. Opin. Struct. Biol. 2:237–241, 1992.
5. Crick, F.H.C. The packing of α helices: Simple coiled-coils. Acta Crystallogr. 6:689–697, 1953.
6. Chothia, C., Levitt, M., Richardson, D. Structure of proteins: Packing of α helices and pleated sheets. Proc. Natl. Acad. Sci. U.S.A. 74:4130–4134, 1977.
7. Efimov, A.V. Stereochemistry of packing of α helices and β-structure in a compact globule. Dokl. Akad. Nauk S.S.S.R. 235:699–702, 1977.
8. Richmond, T.J., Richards, F.M. Packing of α helices: Geometrical constraints and contact areas. J. Mol. Biol. 119:537–555, 1978.
9. Efimov, A.V. Packing of α helices in globular proteins. Layer-structure of globin hydrophobic cores. J. Mol. Biol. 134:23–40, 1979.
10. Chothia, C., Levitt, M., Richardson, D. Helix to helix packing in proteins. J. Mol. Biol. 145:215–250, 1981.
11. Chothia, C., Janin, J. Orthogonal packing of β-pleated sheets in proteins. Biochemistry 21:3955–3965, 1982.
12. Murzin, A.G., Finkelstein, A.V. General architecture of the α-helical globule. J. Mol. Biol. 204:749–769, 1988.
13. Rao, S.T., Rossmann, M.G. Comparison of super-secondary structures in proteins. J. Mol. Biol. 76:241–256, 1973.
14. Sternberg, M.J.E., Thornton, J. M. On the conformation of proteins: The handedness of the β strand-α helix-β strand unit. J. Mol. Biol. 105:367–382, 1976.
15. Levitt, M., Chothia, C. Structural patterns in globular proteins. Nature 261:552–558, 1976.
16. Richardson, J. S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:167–339, 1981.
17. Efimov, A.V. Super-secondary structure of β proteins. Mol. Biol. 16:799–806, 1982.
18. Efimov, A.V. A novel super-secondary structure of proteins and the relation between the structure and the amino acid sequence. FEBS Lett. 166:33–38, 1984.
19. Presnell, S.R., Cohen, F.E. Topological distribution of four-α helix bundles. Proc. Natl. Acad. Sci. U.S.A. 86:6592–6596, 1989.
20. Efimov, A.V. Structure of coiled β-β hairpins and β-β corners. FEBS Lett. 284:288–292, 1991.
21. Efimov, A.V. A novel super-secondary structure of β proteins: A triple-strand corner. FEBS Lett. 298:261–265, 1992.
22. Efimov, A.V. Super-secondary structures involving triple-strand β sheets. FEBS Lett. 334:253–256, 1993.
23. Orengo, C.A., Thornton, J.M. Alpha plus beta folds revisited: Some favoured motifs. Structure 1:105–120, 1993.
24. Murzin, A.G. OB (oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. EMBO J. 12:861–867, 1993.
25. Efimov, A.V. Standard structures in proteins. Prog. Biophys. Mol. Biol. 60:201–239, 1993.
26. Efimov, A.V. Favoured structural motifs in globular proteins. Structure 2:999–1002, 1994.
27. Efimov, A.V. Common structural motifs in small proteins and domains. FEBS Lett. 355:213–219, 1994.
28. Ptitsyn, O.B., Finkelstein, A.V., Falk (Bendzko), P. Principal folding pathway and topology of all-β proteins. FEBS Lett. 101:1–5, 1979.

29. Ptitsyn, O.B., Finkelstein, A.V. Similarities of protein topologies: Evolutionary divergence, functional convergence or principles of folding? Q. Rev. Biophys. 13:339–386, 1980.
30. Ptitsyn, O.B. Protein folding: General physical model. FEBS Lett. 131:197–202, 1981.
31. Ptitsyn, O.B., Rashin, A.A. A model of myoglobin self-organization. Biophys. Chem. 3:1–20, 1975.
32. Efimov, A.V. Structural similarity between two-layer α/β and β proteins. J. Mol. Biol. 245:402–415, 1995.
33. Lim, V.I., Mazanov, A.L., Efimov, A.V. A stereochemical theory of globular protein tertiary structure. I. Highly helical intermediate structures. Mol. Biol. 12:206–213. 1978.
34. Richardson, J.S. β-Sheet topology and the relatedness of proteins. Nature 268:495–500, 1977.
35. Rao, Z., Belyaev, A.S., Fry, E., Roy, P., Jones, I.M., Stuart, D.I. Crystal structural of SIV matrix antigen and implications for virus assembly. Nature 378:743–747, 1995.
36. Kurihara, H., Nonaka, T., Mitsui, Y., Ohgi, K., Irie, M., Nakamura, K.T. The crystal structure of ribonuclease Rh from *Rhizopus niveus* at 2.0Å resolution. J. Mol. Biol. 255:310–320, 1996.
37. Uhlin, U., Eklund, H. Structure of ribonucleotide reductase protein R1. Nature 370:533–539, 1994.
38. Hart, P.J., Pfluger, H.D., Monzingo, A.F., Hollis, T., Robertus, J.D. The refined crystal structure of an endochitinase from *Hordeum vulgare L.* seeds at 1.8Å resolution. J. Mol. Biol. 248:402–413, 1995.
39. Jeon, Y.H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A., Kyogoku, Y. Solution structure of the activator contact domain of the RNA polymerase α subunit. Science 270:1495–1497, 1995.
40. Anderson, W.F., Ohlendorf, D.H., Takeda, Y., Matthews, B.W. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. Nature 290:754–758, 1981.
41. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., Sweet, R.M. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature 362:219–223, 1993.
42. Clark, K.L., Halay, E.D., Lai, E., Burley, S.K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature 364:412–420, 1993.
43. McKay, D.B., Weber, I.T., Steitz, T.A. Structure of catabolite gene activator protein at 2.9−Å resolution. J. Biol. Chem. 257:9518–9524, 1982.
44. Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., Saenger, W. The complex formed between Tet repressor and tetracycline-$Mg^{2+}$ reveals mechanism of antibiotic resistance. J. Mol. Biol. 247:260–280, 1995.
45. Pabo, C.O., Lewis, M. The operator-binding domain of λ repressor: Structure and DNA recognition. Nature 298: 443–447, 1982.
46. Mondragon, A., Subbiah, S., Almo, S.C., Drottar, M., Harrison, S.C. Structure of the amino-terminal domain of phage 434 repressor at 2.0Å resolution. J. Mol. Biol. 205:189–200, 1989.
47. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W., Pabo, C.O. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. Cell 77:21–32, 1994.
48. Martin, J.L., Bardwell, J.C.A., Kuriyan, J. Crystal structure of the DsbA protein required for disulphide bond formation in vivo. Nature 365:464–468, 1993.
49. Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H., Kraut, J. Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. Science 264:1891–1898, 1994.
50. Kim, Y., Eom, S.H., Wang, J., Lee, D.-S., Suh, S.W., Steitz, T.A. Crystal structure of *Thermus aquaticus* DNA polymerase. Nature 376:612–616, 1995.
51. Renaud, J-P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H., Moras, D. Crystal structure of the RAR-γ ligand-binding domain bound to all-trans retinoic acid. Nature 378:681–689, 1995.
52. Morize, I., Surcouf, E., Vaney, M.C., Epelboin, Y., Buehner, M., Fridlansky, F., Milgrom, E., Mornon, J.P. Refinement of the $C222_1$ crystal form of oxidized uteroglobin at 1.34 Å resolution. J. Mol. Biol. 194:725–739, 1987.
53. Holmes, M.A., Matthews, B.W. Structure of thermolysin refined at 1.6 Å resolution. J. Mol. Biol. 160:623–639, 1982.
54. Remington, S., Wiegand, G., Huber, R. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. J. Mol. Biol. 158:111–152, 1982.
55. Matthews, S., Barlow, P., Boyd, J., Barton, G., Russell, R., Mills, H., Cunningham, M., Meyers, N., Burns, N., Clark, N., Kingsman, S., Kingsman, A., Campbell, I. Structural similarity between the p17 matrix protein of HIV-1 and interferon-γ. Nature 370:666–668, 1994.
56. Finzel, B.C., Poulos, T.L., Kraut, J. Crystal structure of yeast cytochrome c peroxidase refined at 1.7-Å resolution. J. Biol. Chem. 259:13027–13036, 1984.
57. Nureki, O., Vassylyev, D.G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazawa, T., Yokoyama, S., Morikawa, K. Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. Science 267:1958–1965, 1995.
58. Bennet, W.S., Jr., Steitz, T.A. Glucose-induced conformational change in yeast hexokinase. Proc. Natl. Acad. Sci. USA 75:4848–4852, 1978.
59. Huber, R., Romisch, J., Paques, E.-P. The crystal and molecular structure of human annexin V, an anticoagulant protein that binds to calcium and membranes. EMBO J. 9:3867–3874, 1990.
60. Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S., Lehmann, M.S. Crystal structure of hydrophobic protein from soybean: A member of a new cysteine-rich family. J. Mol. Biol. 231:877–887, 1993.
61. Shin, D.H., Lee, J.Y., Hwang, K.Y., Kim, K.K., Suh, S.W. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. Structure 3:189–199, 1995.
62. Heinemann, B., Andersen, K.V., Nielesen, P.R., Bech, L.M., Poulsen, F.M. Structure in solution of a four-helix lipid binding protein. Protein Sci. 5:13–23, 1996.
63. Pedersen, L.C., Benning, M.M., Holden, H.M. Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. Biochemistry 34: 13305–13311, 1995.
64. Choe, S., Bennett, M.J., Fujii, G., Curmi, P.M.G., Kantardjieff, K.A., Collier, R.J., Eisenberg, D. The crystal structure of diphtheria toxin. Nature 357:216–222, 1992.
65. Ladner, R.C., Heidner, E.J., Perutz, M.F. The structure of horse methaemoglobin at 2.0Å resolution. J. Mol. Biol. 114:385–414, 1977.
66. Parker, M.W., Postma, J.P.M., Pattus, F., Tucker, A.D., Tsernoglou, D. Refined structure of the pore-forming domain of colicin A at 2.4 Å resolution. J. Mol. Biol. 224:639–657, 1992.
67. Brown, N.R., Noble, M.E.M., Endicott, J.A., Garman, E.F., Wakatsuki, S., Mitchell, E., Rasmussen, B., Hunt, T., Johnson, L.N. The crystal structure of cyclin A. Structure 3:1235–1247, 1995.
68. Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., Burley, S.K. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature 377:119–128, 1995.
69. Marcotte, E.M., Monzingo, A.F., Ernst, S.R., Brzezinski, R., Robertus, J.D. X-ray structure of an anti-fungal chitosanase from streptomyces N 174. Nature Struct. Biol. 3:155–161, 1996.
70. He, X.M., Carter, D.C. Atomic structure and chemistry of human serum albumin. Nature 358:209–215, 1992.
71. Kretsinger, R.H., Nockolds, C.E. Carp muscle calcium-binding protein. II. Structure determination and general description. J. Biol. Chem. 248:3313–3326, 1973.
72. Drenth, J., Jansonius, J.N., Koekoek, R., Wolthers, B.G. The structure of papain. Adv. Protein Chem. 25:79–115, 1971.
73. Dideberg, O., Charlier, P., Dive, G., Joris, B., Frre, J.M., Ghuysen, J.M. Structure of a $Zn^{2+}$-containing D-alanyl-D-

alanine-cleaving carboxypeptidase at 2.5 Å resolution. Nature 299:469–470, 1982.
74. Kalia, Y.N., Brocklehurst, S.M., Hipps, D.S., Appella, E., Sakaguchi, K., Perham, R.N. The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. J. Mol. Biol. 230:323–341, 1993.
75. Blake, C.C.F., Geisow, M.J., Oatley, S.J., Rérat, B., Rérat, C. Structure of prealbumin: Secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. J. Mol. Biol. 121:339–356, 1978.
76. Cyszak, E., Cody, V., Luft, J.R. Crystal structure determination at 2.3-Å resolution of human transthyretin-3′,5′-dibromo-2′,4,4′,6-tetrahydroxyaurone complex. Proc. Natl. Acad. Sci. U.S.A. 89:6644–6648, 1992.
77. Klein, C., Schulz, G.E. Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. J. Mol. Biol. 217:737–750, 1991.
78. Ohlendorf, D.H., Orville, A.M., Lipscomb, J.D. Structure of protocatechuate 3,4-dioxygenase from *Pseudomonas aeruginosa* at 2.15 Å resolution. J. Mol. Biol. 244:586–608, 1994.
79. Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E., Saraste, M. Crystal structure of the membrane-exposed domain from a respiratory quinol oxidase complex with an engineered dinuclear copper center. Proc. Natl. Acad. Sci. U.S.A. 92:11955–11959, 1995.
80. Romero, A., Nar, H., Huber, R., Messerschmidt, A., Kalverda, A.P., Canters, G.W., Durley, R., Mathews, F.S. Crystal structure analysis and refinement at 2.15 Å resolution of amicyanin, a type I blue copper protein, from *Thiobacillus versutus*. J. Mol. Biol. 236:1196–1211, 1994.
81. Guss, J.M., Freeman, H.C. Structure of oxidized poplar plastocyanin at 1.6 Å resolution. J. Mol. Biol. 169:521–563, 1983.
82. Norris, G.E., Anderson, B.F., Baker, E.N. Structure of azurin from *Alcaligenes denitrificans* at 2.5 Å resolution. J. Mol. Biol. 165:501–521, 1983.
83. Messerschmidt, A., Ladenstein, R., Huber, R., Bolognesi, M., Avigliano, L., Petruzzelli, R., Rossi, A., Finazzi-Agro, A. Refined crystal structure of ascorbate oxidase at 1.9 Å resolution. J. Mol. Biol. 224:179–205, 1992.
84. Ogata, C.M., Gordon, P.F., de Vos, A.M., Kim, S.-H. Crystal structure of a sweet tasting protein thaumatin I, at 1.65 Å resolution. J. Mol. Biol. 228:893–908, 1992.
85. Pflugrath, J. W., Wiegand, G., Huber, R., Vértesy, L. Crystal structure determination, refinement and the molecular model of the α-amylase inhibitor Hoe-467 A. J. Mol. Biol. 189:383–386, 1986.
86. Saul, F.A., Amzel, L.M., Poljak, R.J. Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.0 Å resolution. J. Biol. Chem. 253:585–597, 1978.
87. Shapiro, L., Kwong, P.D., Fannon, A.M., Colman, D.R., Hendrickson, W.A. Considerations on the folding topology and evolutionary origin of cadherin domains. Proc. Natl. Acad. Sci. U.S.A. 92:6793–6797, 1995.
88. Bodian, D.L., Jones, E.Y., Harlos, K., Stuart, D.I., Davis, S.J. Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5 Å resolution. Structure 2:755–766, 1994.
89. Tainer, J.A., Getzoff, E.D., Beem, K.M., Richardson, J.S., Richardson, D.C. Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. J. Mol. Biol. 160:181–217, 1982.
90. Holmgren, A., Brändén, C.-I. Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. Nature 342:248–251, 1989.
91. Batalia, M.A., Monzingo, A.F., Ernst, S., Roberts, W., Robertus, J.D. The crystal structure of the antifungal protein zeamatin, a member of the thaumatin-like, PR-5 protein family. Nature Struct. Biol. 3:19–22, 1996.
92. Ban, N., McPherson, A. The structure of satellite panicum mosaic virus at 1.9 Å resolution. Nature Struct. Biol. 2:882–890, 1995.
93. Larson, S.B., Koszelak, S., Day, J., Greenwood, A., Dodds, J.A., McPherson, A. Three-dimensional structure of satellite tobacco mosaic virus at 2.9 Å resolution. J. Mol. Biol. 231:375–391, 1993.
94. Norris, G.E., Stillman, T.J., Anderson, B.F., Baker, E.N. The three-dimensional structure of PNGase F, a glycosyl asparaginase from *Flavobacterium meningosepticum*. Structure 2:1049–1059, 1994.
95. Lawrence, M.C., Suzuki, E., Varghese, J.N., Davies, P.C., Van Donkelaar, A., Tulloch, P.A., Colman, P.M. The three-dimensional structure of the seed storage protein phaseolin at 3 Å resolution. EMBO J. 9:9–15, 1990.
96. Wery, J.-P., Reddy, V.S., Hosur, M.V., Johnson, J.E. The refined three-dimensional structure of an insect virus at 2.8 Å resolution. J. Mol. Biol. 235:565–568, 1994.
97. Harrison, S.P., Olson, A.J., Schutt, C.E., Winkler, F.K., Bricogne, G. Tomato bushy stunt virus at 2.9 Å resolution. Nature 276:368–373, 1978.
98. Silva, A.M., Rossmann, M.G. Refined structure of southern bean mosaic virus at 2.9 Å resolution. J. Mol. Biol. 197:69–87, 1987.
99. Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J.-L., Brousseau, R., Cygler, M. *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. J. Mol. Biol. 254:447–464, 1995.
100. Jain, S., Drendel, W.B., Chen, Z., Mathews, F.S., Sly, W.S., Grubb, J.H. Structure of human β-glucuronidase reveals candidate lysosomal targeting and active-site motifs. Nature Struct. Biol. 3:375–381, 1996.
101. Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P., Blundell T. X-ray analysis of the eye lens protein-γ-II crystallin at 1.9 Å resolution. J. Mol. Biol. 170:175–202, 1983.
102. Bagby, S., Harvey, T.S., Eagle, S.G., Inouye, S., Ikura, M. NMR-derived three-dimensional solution structure of protein S complexed with calcium. Structure 2:107–122, 1994.
103. Qian, M., Haser, R., Payan, F. Structure and molecular model refinement of pig pancreatic α-amylase at 2.1 Å resolution. J. Mol. Biol. 231:785–799, 1993.
104. Cedergren-Zeppezauer, E.S., Larsson, G., Nyman, P.O., Gauter, Z., Wilson, K.S. Crystal structure of a dUTPase. Nature 355:740–743, 1992.
105. Antson, A.A., Otridge, J., Brzozowski, A.M., Dodson, E.J., Dodson, G.G., Wilson, K.S., Smith, T.M., Yang M., Kurecki, T., Gollnick, P. The structure of trp RNA-binding attenuation protein. Nature 374:693–700, 1995.
106. Xu, G.-Y., Ong, E., Gilkes, N.R., Kilburn, D.G., Muhandiram, D.R., Harris-Brandts, M., Carver, J.P., Kay, L.E., Harvey, T.S. Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. Biochemistry 34:6993–7009, 1995.
107. Cho, Y., Gorina, S., Jeffrey, P.D., Pavletich, N.P. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. Science 265:346–355, 1994.
108. Martinez, S.E., Huang, D., Szczepaniak, A., Cramer, W.A., Smith, J.L. Crystal structure of chloroplast cytochrome f reveals a novel cytochrome fold and unexpected heme ligation. Structure 2:95–105, 1994.
109. Zhang, R.-G., Scott, D.L., Westbrook, M.L., Nance, S., Sprangler, B.D., Shipley, G.G., Westbrook, E.M. The three-dimensional crystal structure of cholera toxin. J. Mol. Biol. 251:563–573, 1995.
110. Zheng, J., Knighton, D.R., Ten Eyck , L.F., Karlsson, R., Xuong, N., Taylor, S.S., Sowadski, J.M. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with Mg ATP and peptide inhibitor. Biochemistry 32:2154–2161, 1993.
111. Owen, D.J., Noble, M.E.M., Garman, E.F., Papageorgiou, A.C., Johnson, L.N. Two structures of the catalytic domain of phosphorylase kinase: An active protein kinase complexed with substrate analogue and product. Structure 3:467–482, 1995.
112. Lee, S., Owen, K.E., Choi, H.-K., Lee, H., Lu, G., Wengler,

G., Brown, D.T., Rossmann, M.G., Kuhn, R.J. Identification of a protein binding site on the surface of the alphavirus nucleocapsid and its implication in virus assembly. Structure 4:531–541, 1996.
113. Tanaka Hall, T.M., Porter, J.A., Beachy, P.A., Leahy, D.J. A potential catalytic site revealed by the 1.7-Å crystal structure of the amino-terminal signalling domain of Sonic hedgehog. Nature 378:212–216, 1995.
114. Hecht, H.J., Erdmann, H., Park, H.J., Sprinzl, M., Schmid, R.D. Crystal structure of NADH oxidase from *Thermus thermophilus*. Nature Struct. Biol. 2:1109–1115, 1995.
115. Harrison, C.J., Bohm, A.A., Nelson, H.C.M. Crystal structure of the DNA binding domain of the heat shock transcription factor. Science 263:224–227, 1994.
116. Momany, C., Ernst, S., Ghosh, R., Chang, N.-L., Hackert, M.L. Crystallographic structure of a PLP-dependent ornithine decarboxylase from *Lactobacillus* 30a to 3.0 Å resolution. J. Mol. Biol. 252:643–655, 1995.
117. Endrizzi, J.A., Cronk, J.D., Wang, W., Crabtree, G.R., Alber, T. Crystal structure of DCoH, a bifunctional, protein-binding transcriptional coactivator. Science 268, 556–559, 1995.
118. Roderick, S.L., Matthews, B.W. Structure of the cobalt-dependent methionine aminopeptidase from *Escherichia coli*: A new type of proteolytic enzyme. Biochemistry 32:3907–3912, 1993.
119. Knight, S., Andersson, I., Brändén, C.-I. Crystallographic analysis of ribulose 1,5-biphosphate carboxylase from spinach at 2.4 Å resolution. Subunit interactions and active site. J. Mol. Biol. 215:113–160, 1990.
120. Nickitenko, A.V., Trakhanov, S., Quiocho, F.A. 2Å resolution structure of DppA, a periplasmic dipeptide transport/chemosensory receptor. Biochemistry 34:16585–16595, 1995.
121. Van Duyne, G.D., Ghosh, G., Maas, W.K., Sigler, P.B. Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*. J. Mol. Biol. 256:377–391, 1996.
122. Dumas, C., Lascu, I., Moréra, S., Glaser, P., Fourme, R., Wallet, V., Lacombe, M.-L., Véron, M., Janin, J. X-ray structure of nucleoside diphosphate kinase. EMBO J. 11:3203–3208, 1992.
123. Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J., Katsube, Y. Three-dimensional structure of the glutathione synthetase from *Escherichia coli* B at 2.0Å resolution. J. Mol. Biol. 229:1083–1100, 1993.
124. Georgiadis, M.M., Jessen, S.M., Ogata, C.M., Telesnitsky, A., Goff, S.P., Hendrickson, W.A. Mechanistic implications from the structure of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase. Structure 3:879–892, 1995.
125. Kim, K.H., Pan, Z., Honzatko, R.B., Ke, H., Lipscomb, W.N. Structural asymmetry in the CTP-liganded form of aspartate carbamoyltransferase from *Escherichia coli*. J. Mol. Biol. 196:853–875, 1987.
126. Schad, E.M., Zaitseva, I., Zaitsev, V.N., Dohlsten, M., Kalland, T., Schlievert, P.M., Ohlendorf, D.H., Svensson, L.A. Crystal structure of the superantigen staphylococcal enterotoxin type A. EMBO J. 14:3292–3301, 1995.
127. Herzberg, O., Reddy, P., Sutrina, S., Saier, M.H., Reizer, J., Kapadia, G. Structure of the histidine-containing phosphocarrier protein Hor from *Bacillus subtilis* at 2.0-Å resolution. Proc. Natl. Acad. Sci. U.S.A. 89:2499–2503, 1992.
128. Adman, E.T., Sieker, L.C., Jensen, L.H. The structure of a bacterial ferredoxin. J. Biol. Chem. 248:3987–3996, 1973.
129. Nagai, K., Oubridge, C., Jessen, T.H., Li, J., Evans, P.R. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature 348:515–520, 1990.
130. Lindahl, M., Svensson, L.A., Liljas, A., Sedelnikova, S.E., Eliseikina, I.A., Fomenkova, N.P., Nevskaya, N.A., Nikonov, S.V., Garber, M.B., Muranova, T.A., Rykunova, A.I., Amons, R. Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. EMBO J. 13:1249–1254, 1994.
131. Hegde, R.S., Grossman, S.R., Laimins, L.A., Sigler, P.B. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. Nature 359:505–512, 1992.
132. Guasch, A., Coll, M., Avilés, F.X., Huber, R. Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. J. Mol. Biol. 224:141–157, 1992.
133. Krezel, A.M., Kasibhatla, C., Hidalgo, P., MacKinnon, R., Wagner, G. Solution structure of the potassium channel inhibitor agitoxin 2: Caliper for probing channel geometry. Protein Sci. 4:1478–1489, 1995.
134. Hoffman, D.W., Davies, C., Gerchman, S.E., Kycia, J.H., Porter, S.J., White, S.W., Ramakrishnan, V. Crystal structure of prokaryotic ribosomal protein L9: A bi-lobed RNA-binding protein. EMBO J. 13:205–212, 1994.
135. Lorenz, M., Popp, D., Holmes, K.C. Refinement of the F-actin model against x-ray fiber diffraction data by the use of a directed mutation algorithm. J. Mol. Biol. 234:826–836, 1993.
136. Barlow, P.N., Luisi, B., Milner, A., Elliott, M., Everett, R. Structure of the $C_3HC_4$ domain by $^1$H-nuclear magnetic resonance spectroscopy: A new structural class of Zinc-finger. J. Mol. Biol. 237:201–211, 1994.
137. Betts, L., Xiang, S., Short, S.A., Wolfenden, R., Carter, C.W., Jr. Cytidine deaminase: The 2.3 Å crystal structure of an enzyme: Transition-state analog complex. J. Mol. Biol. 235:635–656, 1994.
138. Stallings, W.C., Abdel-Meguid, S.S., Lim, L.W., Shieh, H.-S., Dayringer, H.E., Leimgruber, N.K., Stegeman, R.A., Anderson, K.S., Sikorski, J.A., Padgette, S.R., Kishore, G.M. Structure and topological symmetry of the glyphosate target 5-enol-pyruvylshikimate-3-phosphate synthase: A distinctive protein fold. Proc. Natl. Acad. Sci. U.S.A. 88:5046–5050, 1991.
139. Oefner, C., Suck, D. Crystallographic refinement and structure of DNase I at 2 Å resolution. J. Mol. Biol. 192:605–632, 1986.
140. Biou, V., Shu, F., Ramakrishnan, V. X-ray crystallography shows that translational initiation factor IF3 consists of two compact $\alpha/\beta$ domains linked by an $\alpha$ helix. EMBO J. 14:4056–4064, 1995.
141. Kankare, J., Neal, G.S., Salminen, T., Glumhoff, T., Cooperman, B.S., Lahti, R., Goldman, A. The structure of *E.coli* soluble inorganic pyrophosphatase at 2.7 Å resolution. Protein Eng. 7:823–830, 1994.
142. De, A., Brown, D.G., Gorman, M.A., Carr, M., Sanderson, M.R., Freemont, P.S. Crystal structure of a disulfide-linked "trefoil" motif found in a large family of putative growth factors. Proc. Natl. Acad. Sci. U.S.A. 91:1084–1088, 1994.
143. McLaughlin, P.J., Gooch, J.T., Mannherz, H.-G., Weeds, A.G. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. Nature 364:685–692, 1993.
144. Wright, C.S. Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8 Å resolution. J. Mol. Biol. 194:501–529, 1987.
145. Carlioz, A., Ludwig, M.L., Stallings, W.C., Fee, J.A., Steinman, H.M., Touati, D. Iron superoxide dismutase: Nucleotide sequence of the gene from *Escherichia coli* K12 and correlations with crystal structures. J. Biol. Chem. 263:1555–1562, 1988.
146. Parker, M.W., Blake, C.C.F. Crystal structure of manganese superoxide dismutase from *Bacillus stearothermophilus* at 2.4 Å resolution. J. Mol. Biol. 199:649–661, 1988.
147. Vijay-Kumar, S., Bugg, C.E., Cook, W.J. Structure of ubiquitin refined at 1.8 Å resolution. J. Mol. Biol. 194:531–544, 1987.
148. Achari, A., Hale, S.P., Howard, A.J., Clore, G.M., Gronenborn, A.M., Hardman, K.D., Whitlow, M. 1.67-Å x-ray structure of the B2 immunoglobulin-binding domain of streptococcal protein G and comparison to the NMR structure of the B1 domain. Biochemistry 31:10449–10457, 1992.
149. Gallagher, T., Rozwarski, D.A., Ernst, S.R., Hackert, M.L. Refined structure of the pyruvoyl-dependent histidine

decarboxylase from *Lactobacillus* 30a. J. Mol. Biol. 230: 516–528, 1993.
150. Bailey, S., Evans, R.W., Garratt, R.C., Gorinsky, B., Hasnain, S., Horsburgh, C., Jhoti, H., Lindley, P.F., Mydin, A., Sarra, R., Watson, J.L. Molecular structure of serum transferrin at 3.3-Å resolution. Biochemistry 27: 5804–5812, 1988.
151. Day, C.L., Anderson, B.F., Tweedie, J.W., Baker, E.N. Structure of the recombinant N-terminal lobe of human lactoferrin at 2.0 Å resolution. J. Mol. Biol. 232:1084–1100, 1993.
152. Klein, C., Chen, P., Arevalo, J.H., Stura, E.A., Marolewski, A., Warren, M.S., Benkovic, S.J., Wilson, I.A. Towards structure-based drug design: Crystal structure of a multisubstrate adduct complex of glycinamide ribonucleotide transformylase at 1.96 Å resolution. J. Mol. Biol. 249:153–175, 1995.
153. Arjunan, P., Umland, T., Dyda, F., Swaminathan, S., Furey, W., Sax, M., Farrenkopf, B., Gao, Y., Zhang, D., Jordan, F. Crystal structure of the thiamin diphosphate-dependent enzyme pyruvate decarboxylase from the yeast *Saccharomyces cerevisiae* at 2.3 Å resolution. J. Mol. Biol. 256:590–600, 1996.
154. Skarzysnki, T., Moody, P.C.E., Wonacott, A.J. Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. J. Mol. Biol. 193:171–187, 1987.
155. Vrielink, A., Ruger, W., Driessen, H.P.C., Freemont, P.S. Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. EMBO J. 13:3413–3422, 1994.
156. Adams, M.J., Ellis, G.H., Gover, S., Naylor, C.E., Phillips, C. Crystallographic study of coenzyme, coenzyme analogue and substrate binding in 6-phosphogluconate dehydrogenase: Implications for NADP specificity and the enzyme mechanism. Structure 2:651–668, 1994.
157. Robbins, A.H., Stout, C.D. The structure of aconitase. Proteins 5:289–312, 1989.
158. Varughese, K.I., Skinner, M.M., Whiteley, J.M., Matthews, D.A., Xuong, N.H. Crystal structure of rat liver dihydropteridine reductase. Proc. Natl. Acad. Sci. U.S.A. 89:6080–6084, 1992.
159. Bauer, A.J., Rayment, I., Frey, P.A., Holden, H.M. The molecular structure of UDP-galactose 4-epimerase from *Escherichia coli* determined at 2.5 Å resolution. Proteins 12:372–381, 1992.
160. Ghosh, D., Wawrzak, Z., Weeks, C.M., Duax, W.L., Erman, M. The refined three-dimensional structure of 3α,20β-hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases. Structure 2:629–640, 1994.
161. Ghosh, D., Pletnev, V.Z., Zhu, D.-W., Wawrzak, Z., Duax, W.L., Pangborn, W., Labrie, F., Lin, S.-X. Structure of human esterogenic 17-β-hydroxysteroid dehydrogenase at 2.20 Å resolution. Structure 3:503–513, 1995.
162. Tanaka, N., Nonaka, T., Nakanishi, M., Deyashiki, Y., Hara, A., Mitsui, Y. 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, 1996.
163. Rafferty, J.B., Simon, J.W., Baldock, C., Artymiuk, P.J., Baker, P.J., Stuitje, A.R., Slabas, A.R., Rice, D.W. Common themes in redox chemistry emerge from the x-ray structure of oilseed rape (*Brassica napus*) enoyl acyl carrier protein reductase. Structure 3:927–938, 1995.
164. Karplus, P.A., Daniels, M.J., Herriott, J.R. Atomic structure of ferredoxin-NADP$^+$ reductase: Prototype for a structurally novel flavoenzyme family. Science 251:60–66, 1991.
165. Park, H.-W., Kim, S.-T., Sancar, A., Deisenhofer, J. Crystal structure of DNA photolyase from *Escherichia coli*. Science 268:1866–1672, 1995.
166. Doublie, S., Bricogne, G., Gilmore, C., Carter, C.W., Jr. Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. Structure 3:17–31, 1995.
167. Correll, C.C., Batie, C.J., Ballou, D.P., Ludwig, M.L. Phthalate dioxygenase reductase: A modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. Science 258:1604–1610, 1992.
168. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., Roberts, R.J. Crystal structure of the Hhal DNA methyltransferase complexed with *S*-adenosyl-L-methionine. Cell 74:299–307, 1993.
169. Poland, B.W., Silva, M.M., Serra, M.A., Cho, Y., Kim, K.H., Harris, E.M.S., Honzatko, R.B. Crystal structure of adenylosuccinate synthetase from *Escherichia coli*: Evidence for convergent evolution of GTP-binding domains. J. Biol. Chem. 268:25334–25342, 1993.
170. Oefner, C., D'Arcy, A., Winkler, F.K. Crystal structure of human dihydrofolate reductase complexed with folate. Eur. J. Biochem. 174:377–385, 1988.
171. Mattevi, A., Valentini, G., Rizzi, M., Speranza, M.L., Bolognesi, M., Coda, A. Crystal structure of *Escherichia coli* pyruvate kinase type I: Molecular basis of the allosteric transition. Structure 3:729–741, 1995.
172. Muirhead, H., Clayden, D.A., Barford, D., Lorimer, C.G., Fothergill Gilmore, L.A., Schiltz, E., Schmitt, W. The structure of cat muscle pyruvate kinase. EMBO J. 5:475–481, 1986.
173. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I. Three-dimensional structure of the adenine-specific DNA methyltransferase MTaq I in complex with the cofactor *S*-adenosylmethionine. Proc. Natl. Acad. Sci. U.S.A. 91:10957–10961, 1994.
174. Vidgren, J., Svensson, L.A., Liljas, A. Crystal structure of catechol *O*-methyltransferase. Nature 368:354–358, 1994.
175. Eklund, H., Samama, J.-P., Wallen, L., Brändén, C.-I., Åkerson, Å., Jones, T.A. Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 Å resolution. J. Mol. Biol. 146:561–587, 1981.
176. Ke, H.-M., Honzatko, R.B., Lipscomb, W.N. Structure of unligated aspartate carbamoyltransferase of *Escherichia coli* at 2.6-Å resolution. Proc. Natl. Acad. Sci. U.S.A. 81:4037–4040, 1984.
177. John, J., Crennell, S.J., Hough, D.W., Danson, M.J., Taylor, G.L. The crystal structure of glucose dehydrogenase from *Thermoplasma acidophilum*. Structure 2:385–393, 1994.
178. Tesmer, J.J.G., Klem, T.J., Deras, M.L., Davisson, V.J., Smith, J.L. The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families. Nature Struct. Biol. 3:74–86, 1996.
179. Goldberg, J.D., Yoshida, T., Brick, P. Crystal structure of a NAD-dependent D-glycerate dehydrogenase at 2.4 Å resolution. J. Mol. Biol. 236:1123–1140, 1994.
180. Stoll, V.S., Kimber, M.S., Pai, E.F. Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of *Lactobacillus pentosus* D-lactate dehydrogenase. Structure 4:437–447, 1996.
181. Hodel, A.E., Gershon, P.D., Shi, X., Quiocho, F.A. The 1.85 Å structure of vaccinia protein VP39: A bifunctional enzyme that participates in the modification of both mRNA ends. Cell 85:247–256, 1996.
182. Oliva, G., Fontes, M.R.M., Garratt, R.C., Altamirano, M.M., Calcagno, M.L., Horjales, E. Structure and catalytic mechanism of glucosamine 6-phosphate deaminase from *Escherichia coli* at 2.1 Å resolution. Structure 3:1323–1332, 1995.
183. Baker, P.J., Turnbull, A.P., Sedelnikova, S.E., Stillman, T.J., Rice, D.W. A role for quaternary structure in the substrate specificity of leucine dehydrogenase. Structure 3:693–705, 1995.
184. Muller, Y.A., Schulz, G.E. Structure of the thiamine- and flavin-dependent enzyme pyruvate oxidase. Science 259: 965–967, 1993.
185. Abad-Zapatero, C., Griffith, J.P., Sussman, J.L., Rossmann, M.G. Refined crystal structure of dogfish $M_4$ apolactate dehydrogenase. J. Mol. Biol. 198:445–467, 1987.

186. Hall, M.D., Levitt, D.G., Banaszak, L.J. Crystal structure of *Escherichia coli* malate dehydrogenase: A complex of the apoenzyme and citrate at 1.87 Å resolution. J. Mol. Biol. 226:867–882, 1992.
187. Niefind, K., Hecht, H.-J., Schomburg, D. Crystal structure of L-2-hydroxyisocaproate dehydrogenase from *Lactobacillus confusus* at 2.2 Å resolution: An example of strong asymmetry between subunits. J. Mol. Biol. 251:256–281, 1995.
188. Romao, M.J., Turk, D., Gomis-Röth, F.-X., Huber, R., Schumacher, G., Müllering, H., Rüssmann, L. Crystal structure analysis, refinement and enzymatic reaction mechanism of *N*-carbamoylsarcosine amidohydrolase from *Arthrobacter* sp. at 2.0 Å resolution. J. Mol. Biol. 226:1111–1130, 1992.
189. Fletterick, R.J., Sygusch, J., Semple, H., Madsen, N.B. Structure of glycogen phosphorylase a at 3.0 Å resolution and its ligand binding sites at 6 Å. J. Biol. Chem. 251:6142–6146, 1976.
190. Story, R.M., Weber, I.T., Steitz, T.A. The structure of the *E. coli* rec A protein monomer and polymer. Nature 355:318–325, 1992.
191. Lin, Z., Konno, M., Abad-Zapatero, C., Wierenga, R., Murthy, M.R.N., Ray, W.J., Jr., Rossmann, M.G. The structure of rabbit muscle phosphoglucomutase at intermediate resolution. J. Biol. Chem. 261:264–274, 1986.
192. Abrahams, J.P., Leslie, A.G.W., Lutter, R., Walker, J.E. Structure at 2.8 Å resolution of $F_1$-ATPase from bovine heart mitochondria. Nature 370:621–628, 1994.
193. Ploegman, J.H., Drent, G., Kalk, K.H., Hol, W.G.J. Structure of bovine liver rhodanese. I. Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. J. Mol. Biol. 123:557–594, 1978.
194. Kim, J., Rees, D.C. Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. Nature 360:553–560, 1992.
195. Mathieu, M., Zeelen, J.Ph., Pauptit, R.A., Erdmann, R., Kunau, W.-H., Wierenga, R.K. The 2.8Å crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: A five-layered αβαβα structure constructed from two core domains of identical topology. Structure 2:797–808, 1994.
196. Hohenester, E., Jansonius, J.N. Crystalline mitochondrial aspartate aminotransferase exists in only two conformations. J. Mol. Biol. 236:963–968, 1994.
197. Antson, A.A., Demidkina, T.V., Gollnick, P., Dauter, Z., Von Tersch, R.L., Long, J., Berezhnoy, S.N., Phillips, R.S., Harutyunyan, E.H., Wilson, K.S. Three-dimensional structure of tyrosine phenol-lyase. Biochemistry 32:4195–4206, 1993.
198. Momany, C., Ghosh, R., Hackert, M.L. Structural motifs for pyridoxal-5-phosphate binding in decarboxylases: An analysis based on the crystal structure of the *Lactobacillus* 30a ornithine decarboxylase. Protein Sci. 4:849–854, 1995.
199. Gilliland, G.L., Quiocho, F.A. Structure of the L-arabinose-binding protein from *Escherichia coli* at 2.4 Å resolution. J. Mol. Biol. 146:341–362, 1981.
200. Lee, Y.-H., Ogata, C., Pflugrath, J.W., Levitt, D.G., Sarma, R., Banaszak, L.J., Pilkis, S.J. Crystal structure of the rat liver fructose-2,6-biphosphatase based on selenomethionine multiwavelength anomalous dispersion phases. Biochemistry 35:6010–6019, 1996.
201. Pearl, L., O'Hara, B., Drew, R., Wilson, S. Crystal structure of Ami C: The controller of transcription antitermination in the amidase operon of *Pseudomonas aeruginosa*. EMBO J. 13:5810–5817, 1994.
202. Sack, J.S., Saper, M.A., Quiocho, F.A. Periplasmic binding protein structure and function: Refined x-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. J. Mol. Biol. 206:171–191, 1989.
203. Eads, J.C., Scapin, G., Xu, Y., Grubmeyer, C., Sacchettini, J.C. The crystal structure of human hypoxanthine-guanine phosphoribosyltransferase with bound GMP. Cell 78:325–334, 1994.
204. Müller, C.W., Schulz, G.E. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor $Ap_5A$ refined at 1.9 Å resolution: A model for a catalytic transition state. J. Mol. Biol. 224:159–177, 1992.
205. Shirakihara, Y., Evans, P.R. Crystal structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction products. J. Mol. Biol. 204:973–994, 1988.
206. Huang, W., Lindqvist, Y., Schneider, G., Gibson, K.J., Flint, D., Lorimer, G. Crystal structure of an ATP-dependent carboxylase, dethiobiotin synthetase, at 1.65 Å resolution. Structure 2:407–414, 1994.
207. Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Warren, M.J., Woodcock, S.C., Jordan, P.M. Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. Nature 359:33–39, 1992.
208. Pflugrath, J.W., Quiocho, F.A. The 2 Å resolution structure of the sulfate-binding protein involved in active transport in *Salmonella typhimurium*. J. Mol. Biol. 200: 163–180, 1988.
209. Kang, C.-H., Shin, W.-C., Yamagata, Y., Gokcen, S., Ames, G.F.-L., Kim, S.-H. Crystal structure of the lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium* at 2.7 Å resolution. J. Biol. Chem. 266: 23893–23899, 1991.
210. Oh, B.-H., Kang, C.-H., De Bondt, H., Kim, S.-H., Nikaido, K., Joshi, A.K., Ames, G.F.-L. The bacterial periplasmic histidine-binding protein. J. Biol. Chem. 269:4135–4143, 1994.
211. Hurley, J.H., Thorsness, P.E., Ramalingam, V., Helmers, N.H., Koshland, D.E., Jr., Stroud, R.M. Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. Proc. Natl. Acad. Sci. U.S.A. 86:8635–8639, 1989.
212. Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y., Oshima, T. Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2 Å resolution. J. Mol. Biol. 222:725–738, 1991.
213. Walker, N.P.C., 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., Herzog, L., Hugunin, M., Houy, W., Mankovich, J.A., McGuiness, L., Orlewicz, E., Paskind, M., Pratt, C.A., Reis, P., Sumani, A., Terranova, M., Welch, J.P., Xiong, L., Müller, A., Tracey, D.E., Kamen, R., Wong, W.W. Crystal structure of the cysteine protease interleukin-1-converting enzyme: $A(p20/p10)_2$ homodimer. Cell 78:343–352, 1994.
214. Rice, P.A., Steitz, T.A. Refinement of αδ resolvase reveals a strikingly flexible molecule. Structure 2:371–384, 1994.
215. Drenth, J., Hol, W.G.J., Jansonius, J.N., Koekoek, R. Comparison of the three-dimensional structures of subtilisin BPN′ and subtilisin Novo. Cold Spring Harbor Symp. Quant. Biol. 36:107–116, 1972.
216. Ke, H., Zhang, Y., Lipscomb, W.N. Crystal structure of fructose-1,6-biphosphatase complexed with fructose 6-phosphate, AMP, and magnesium. Proc. Natl. Acad. Sci. U.S.A. 87:5243–5247, 1990.
217. Bone, R., Springer, J.P., Atack, J.R. Structure of inositol monophosphatase, the putative target of lithium therapy. Proc. Natl. Acad. Sci. U.S.A. 89:10031–10035, 1992.
218. Smith, W.W., Burnett, R.M., Darling, G.D., Ludwig, M.L. Structure of the semiquinone form of flavodoxin from *Clostridium* MP: Extension of 1.8 Å resolution and some comparisons with the oxidized state. J. Mol. Biol. 117:195–225, 1977.
219. Stock, A.M., Mottonen, J.M., Stock, J.B., Schutt, C.E. Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. Nature 337:745–749, 1989.
220. Pai, E.F., Kabsch, W., Krengel, U., Holmes, K.C., John, J., Wittinghofer, A. Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. Nature 341:209–214, 1989.
221. Zou, J., Flocco, M.M., Mowbray, S.L. The 1.7 Å refined

x-ray structure of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. J. Mol. Biol. 233:739–752, 1993.
222. Ladenstein, R., Schneider, M., Huber, R., Bartunik, H.-D., Wilson, K., Schott, K., Bacher, A. Heavy riboflavin synthase from *Bacillus subtilis*. Crystal structure analysis of the icosahedral $\beta_{60}$ capsid at 3.3 Å resolution. J. Mol. Biol. 203:1045–1070, 1988.
223. Savva, R., McAuley-Hecht, K., Brown, T., Pearl, L. The structural basis of specific base-excision repair by uracil-DNA glucosylase. Nature 373:487–493, 1995.
224. Qu, A., Leahy, D.J. Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, $\alpha_L\beta 2$) integrin. Proc. Natl. Acad. Sci. U.S.A. 92:10277–10281, 1995.
225. Martinez, C., Nicolas, A., van Tilbeurgh, H., Egloff, M.-P., Cudrey, C., Verger, R., Cambillau, C. Cutinase, a lipolytic enzyme with a preformed oxyanion hole. Biochemistry 33:83–89, 1994.
226. Stocker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F.-X., McKay, D.B., Bode, W. The metzincins: topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. Protein Sci. 4:824–840, 1995.
227. Logan, D.T., Mazauric, M.-H., Kern, D., Moras, D. Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. EMBO J. 14:4156–4167, 1995.
228. Swain, A.L., Jaskólski, M., Housset, D., Rao, J.K.M., Wlodawer, A. Crystal structure of *Escherichia coli* L-asparaginase, an enzyme used in cancer therapy. Proc. Natl. Acad. Sci. U.S.A. 90:1474–1478, 1993.
229. Katti, S.K., LeMaster, D.M., Eklund, H. Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. J. Mol. Biol. 212:167–184, 1990.
230. Yip, K.S.P., Stillman, T.J., Britton, K.L., Artymiuk, P.J., Baker, P.J., Sedelnikova, S.E., Engel, P.C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R., Rice, D.W. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure 3:1147–1158, 1995.
231. Verschueren, K.H.G., Franken, S.M., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W. Refined x-ray structures of haloalkane dehalogenase at pH 6.2 and pH 8.2 and implications for the reaction mechanism. J. Mol. Biol. 232:856–872, 1993.
232. Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., Silman, I. Atomic structure of acetylcholinesterase from *Torpedo californica*: A prototypic acetylcholine-binding protein. Science 253:872–879, 1991.
233. Grochulski, P., Li, Y., Schrag, D., Bouthillier, F., Smith, P., Harrison, D., Rubin, B., Cygler, M. Insights into interfacial activation from an open structure of *Candida rugosa* lipase. J. Biol. Chem. 268:12843–12847, 1993.
234. Pathak, D., Ollis, D. Refined structure of dienelactone hydrolase at 1.8 Å. J. Mol. Biol. 214:497–525, 1990.
235. Hecht, H.J., Sobek, H., Haag, T., Pfeifer, O., van Peie, K.-H. The metal-ion-free oxidoreductase from *Streptomyces aureofaciens* has an $\alpha/\beta$ hydrolase fold. Nature Struct. Biol. 1:532–537, 1994.
236. Rudenko, G., Bonten, E., d'Azzo, A., Hol, W.G.J. Three-dimensional structure of the human 'protective protein': Structure of the precursor form suggests a complex activation mechanism. Structure 3:1249–1259, 1995.
237. Liao, D.-I., Breddam, K., Sweet, R.M., Bullock, T., Remington, S.J. Refined atomic model of wheat serine carboxypeptidase II at 2.2-Å resolution. Biochemistry 31:9796–9812, 1992.
238. Burley, S.K., David, P.R., Sweet, R.M., Taylor, A., Lipscomb, W.N. Structure determination and refinement of bovine lens leucine aminopeptidase and its complex with bestatin. J. Mol. Biol. 224:113–140, 1992.
239. Rees, D.C., Lewis, M., Lipscomb, W.N. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. J. Mol. Biol. 168:367–387, 1983.
240. Derewenda, Z.S., Cambillau, C. Effects of gene mutations in lipoprotein and hepatic lipases as interpreted by a molecular model of the pancreatic triglyceride lipase. J. Biol. Chem. 266:23112–23119, 1991.
241. Korndörfer, I., Steipe, B., Huber, R., Tomschy, A., Jaenicke, R. The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. J. Mol. Biol. 246:511–521, 1995.
242. Efimov, A.V. Structure of $\alpha$-$\alpha$ hairpins with short connections. Protein Eng. 4:245–250, 1991.
243. Finkelstein, A.V., Ptitsyn, O.B. Why do globular proteins fit the limited set of folding patterns? Prog. Biophys. Mol. Biol. 50:171–190, 1987.
244. Chothia, C., Proteins. One thousand families for the molecular biologist. Nature 387:543–544, 1992.
245. Viguera, A.R., Blanco, F.J., Serrano, L. The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. J. Mol. Biol. 247:670–681, 1995.
246. Viguera, A.R., Serrano, L., Wilmanns, M. Different folding transition states may result in the same native structure. Nature Struct. Bio.3:874–880, 1996.
247. Barbar, E., Barany, G., Woodward, C. Unfolded BPTI variants with a single disulfide bond have diminished non-native structure distant from the crosslink. Folding Design 1:65–76, 1996.
248. Holm, L., Ouzounis, C., Sander, C., Tuparev, G., Vriend, G. A database of protein structure families with common folding motifs. Protein Sci. 1:1691–1698, 1992.
249. Orengo, C.A., Flores, T.P., Taylor, W.R., Thornton, J.M. Identification and classification of protein fold families. Protein Eng. 6:485–500, 1993.
250. Rufino, C.D., Blundell, T.L. Structure-based identification and clustering of protein families and superfamilies. J. Comp. Aided Mol. Design 8:5–27, 1994.
251. Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C. SCOP: A structural classification of protein database for the investigation of sequences and structures. J. Mol. Biol. 247:536–540, 1995.
252. Orengo, C. Classification of protein folds. Curr. Opin. Struct. Biol. 4:429–440, 1994.
253. Holm, L., Sander, C. Mapping the protein universe. Science 273:595–602, 1996.
254. Wodak, S.J. Extending molecular systematics to the third dimension. Nature Struct. Biol. 3:575–578, 1996.