# Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features

WOLFGANG KABSCH and CHRISTIAN SANDER, *Biophysics Department, Max Planck Institute of Medical Research, 6900 Heidelberg, Federal Republic of Germany*

## Synopsis

For a successful analysis of the relation between amino acid sequence and protein structure, an unambiguous and physically meaningful definition of secondary structure is essential. We have developed a set of simple and physically motivated criteria for secondary structure, programmed as a pattern-recognition process of hydrogen-bonded and geometrical features extracted from x-ray coordinates. Cooperative secondary structure is recognized as repeats of the elementary hydrogen-bonding patterns "turn" and "bridge." Repeating turns are "helices," repeating bridges are "ladders," connected ladders are "sheets." Geometric structure is defined in terms of the concepts torsion and curvature of differential geometry. Local chain "chirality" is the torsional handedness of four consecutive $C^\alpha$ positions and is positive for right-handed helices and negative for ideal twisted $\beta$-sheets. Curved pieces are defined as "bends." Solvent "exposure" is given as the number of water molecules in possible contact with a residue. The end result is a compilation of the primary structure, including SS bonds, secondary structure, and solvent exposure of 62 different globular proteins. The presentation is in linear form: strip graphs for an overall view and strip tables for the details of each of 10,925 residues. The dictionary is also available in computer-readable form for protein structure prediction work.

## INTRODUCTION

### Background

$\alpha$-Helices and pleated $\beta$-sheets were predicted in 1951 by Linus Pauling and Robert Corey[1] on the basis of hydrogen-bonding and cooperativity criteria. They were seen later, and beautifully, in the first structures shown in atomic detail by x-ray crystallography. Since then, the number of known protein structures has risen to over 100 and comprehensive analysis of secondary structure requires a computerized compilation of structure assignments, especially in the context of structure prediction methods. Existing compilations have various shortcomings. The crystallographers' assignments of secondary structure in the Brookhaven Protein Data Bank[2] are often subjective and sometimes incomplete. Objective algorithms exist, e.g., for defining turns[3–6] (reviewed in Refs. 7, 8), $\beta$-sheets,[9] and solvent accessibility,[10] but only Levitt and Greer[11] have published an extensive compilation of automatic assignments of helices and sheets. Their ap-

Biopolymers, Vol. 22, 2577–2637 (1983)

CCC 0006-3525/83/122577-61$07.10

proach has the advantage of giving assignments when only backbone $C^{\alpha}$ coordinates are known; the price paid is loss of accuracy when all-atom coordinates are known. Solvent exposure has been published for no more than a few proteins, and chirality only on microfiche.[12] We are thus motivated to make available an accurate, exhaustive, and up-to-date compilation.

## The Main Ideas

Our goal is to approximate the intuitive notion of secondary structure by an objective algorithm. An algorithm for extracting structural features from the atomic coordinates is obviously a pattern-recognition process. The elementary patterns on which this process is based should be as simple as possible yet capable of discriminating among the main types of secondary structure. To discriminate whether a pattern is present or not in a continuum of possible atomic configurations, continuous decision parameters must be fixed. Using backbone $\varphi,\psi$ angles or $C^{\alpha}$ positions requires the adjustment of several parameters, e.g., four angles for a rectangle in the $\varphi,\psi$ plane for each type of secondary structure. In contrast, the presence or absence of an H bond can be characterized by a single decision parameter, a cutoff in the bond energy. Therefore, we base our secondary structure recognition algorithm mainly on H-bonding patterns: "$n$-turns" with an H-bond between the CO of residue $i$ and the NH of residue $i + n$, where $n = 3,4,5$, and "bridges" with H bonds between residues not near each other in sequence. These two types of pattern essentially exhaust all backbone–backbone H bonds. Repeating 4-turns define $\alpha$-helices, and repeating bridges define $\beta$-structure, in good agreement with intuitive assignments. All other occurrences of the basic patterns provide an interesting survey of $3_{10}$-helices, $\pi$-helices, single turns, and single $\beta$-bridges.

The results are presented in short form as strip maps of secondary structure (Fig. A1), and in long form, together with the amino acid sequence as an easy-to-use dictionary (Table AIII). The computer program DSSP (Define Secondary Structure of Proteins) written in standard PASCAL will be available from the Protein Data Bank, Chemistry Dept., Brookhaven National Laboratory, Upton, N.Y. 11973. Publication of an update of this compilation is planned as more protein structures are solved.

## DEFINITIONS

The definitions of H-bonded features form a hierarchy: first H bonds are defined; based on them, turns and bridges; and, based on them, $\alpha$-helices and $\beta$-ladders, including common imperfections such as helical kinks and $\beta$-bulges. Features defined geometrically are bends, chirality, SS bonds, and solvent exposure. Each structural feature is defined independently of the others and structural overlaps are resolved by defining a secondary structure summary that assigns a single state to each residue. For brevity we express the pattern definitions in the form of equations. For example,

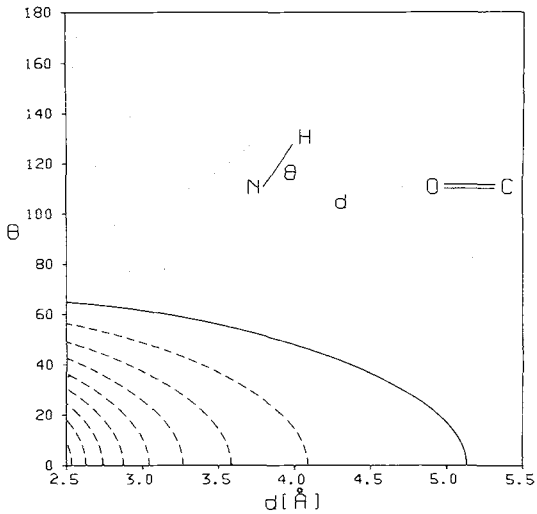



Fig. 1. H bond between peptide units is described here by the dominant electrostatic part $E$ (see text) of the H-bond energy, drawn in contours of constant $E$ at 0.5 kcal/mol intervals as a function of the distance, $d$, and the alignment angle $\theta$. Dotted lines, $E$ positive or zero; broken lines, $E$ negative. An ideal H bond has $d = 2.9$ Å, $\theta = 0$, and $E = -3.0$ kcal/mol. We assume an H bond for $E$ up to $-0.5$ kcal/mol (solid line). Thus, misalignment of up to 63° is allowed at the ideal length; an N-O distance of up to $d = 5.2$ Å is allowed for perfect alignment. This definition of H bonds is particularly simple and physically meaningful. It is more general than the historical definition of hydrogen "bond" and could be called polar interaction.

"Hbond(i,j)=: [E < −0.5kcal/mole]" means: there is an H bond $(i,j)$ if $E$ is less than −0.5 kcal/mol.

## Hydrogen-Bonded Structure

### *Hydrogen Bonds*

Hydrogen bonds in proteins have little wave-function overlap and are well described by an electrostatic model.[13] We calculate the electrostatic interaction energy between two H-bonding groups by placing partial charges on the C,O $(+q_1, -q_1)$ and N,H $(-q_2, +q_2)$ atoms, i.e.,

$$E = q_1q_2(1/r(\mathrm{ON}) + 1/r(\mathrm{CH}) - 1/r(\mathrm{OH}) - 1/r(\mathrm{CN}))*f$$

with $q_1 = 0.42e$ and $q_2 = 0.20e$, $e$ being the unit electron charge and $r$(AB) the interatomic distance from A to B. In chemical units, $r$ is in angstroms, the dimensional factor $f = 332$, and $E$ is in kcal/mol. A good H bond has about −3 kcal/mol binding energy. We choose a generous cutoff to allow for bifurcated H bonds and errors in coordinates and assign an H bond between C=O of residue $i$ and N-H of residue $j$ if $E$ is less than the cutoff, i.e., "Hbond(i,j)=: [E < −0.5kcal/mole]."

Figure 1 illustrates the relation of this one-parameter definition to the

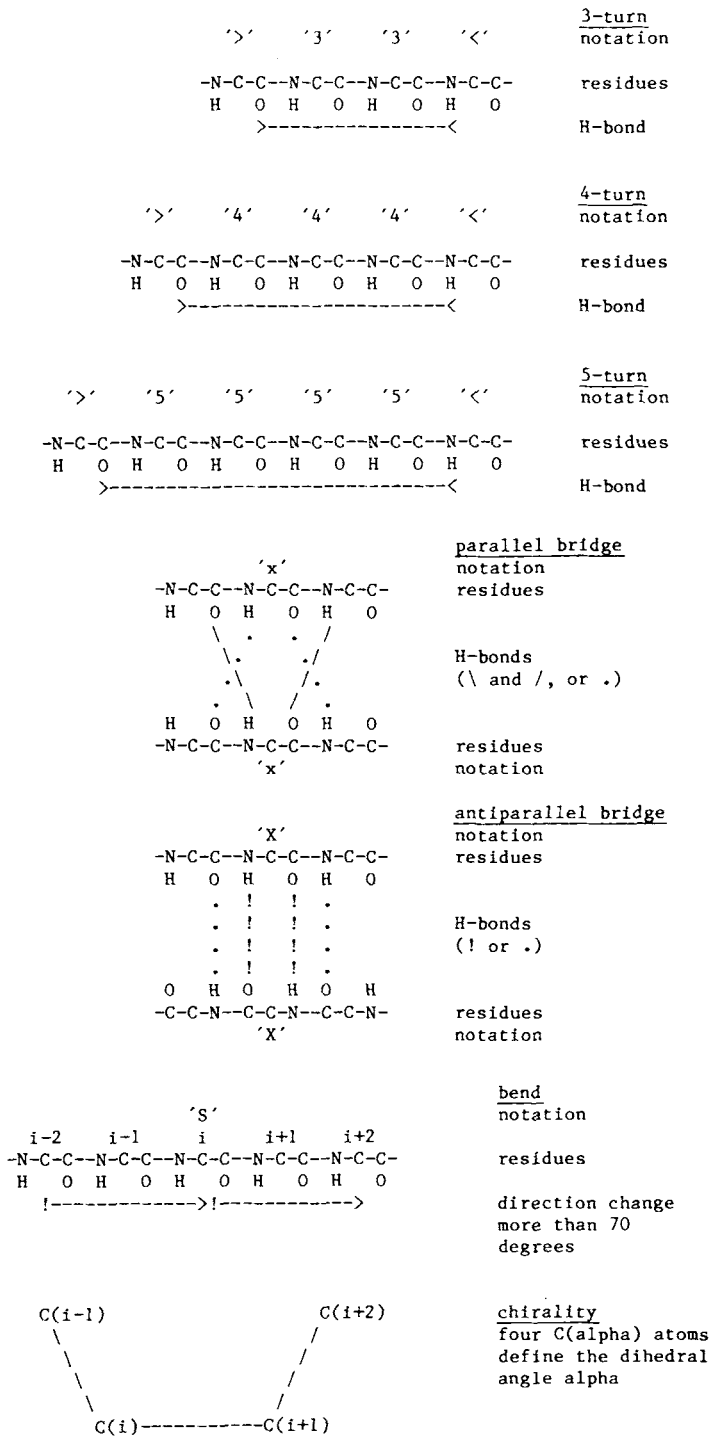



Fig. 2. Elementary patterns used in structure definition.

more complicated description of H bonds in terms of one distance and one angle. There is no generally correct H-bond definition, as there is no sharp border between the quantum-mechanical (wave-function overlap dominates at short distances) and electrostatic (electrostatic interaction dominates at larger distances) regimes and no discontinuity of the interaction energy as a function of distance or alignment. Thus, any H-bond definition is empirically tailored to a particular purpose. Our definition, well tested by trial and error, reflects a compromise suitable for the purpose of secondary structure definition. The cutoff chosen, which allows for an N-O distance up to 2.2 Å larger than the optimal value at perfect alignment or a misalignment of maximally 60° is similar to the tolerances used by Levitt and Greer[11] (1.8 Å excess and 60°) and was found to be sufficient to average over coordinate errors without leading to spurious secondary structure assignments. Were it not for historical reasons, we would use the term "polar interaction" rather than "hydrogen bond."

### *Elementary H-Bond Pattern:* n-*Turn*

The basic turn pattern (Fig. 2) is a single H bond of type $(i,i+n)$. We assign an $n$-turn at residue $i$ if there is an H bond from CO($i$) to NH($i+n$), i.e., "n-turn(i)=: Hbond(i,i + n), n = 3,4,5."

When the pattern is found, the ends of the H bond are indicated by using "⟩" at $i$ and "⟨" at $i+n$ in line 3-TURN, 4-TURN, or 5-TURN of Table AIII; the residues bracketed by the H bond are noted "3," "4," or "5" unless they are also the end points of other H bonds. Coincidence of "⟩" and "⟨" at one residue is indicated by "X." In line SUMMARY of Table AIII, residues bracketed by the hydrogen bond of an $n$-turn are marked "T," unless they are part of an $n$-helix (defined below).

### *Elementary H-Bond Pattern: Bridge*

Two nonoverlapping stretches of three residues each, $i-1,i,i+1$ and $j-1,j,j+1$, form either a parallel or antiparallel bridge, depending on which of two basic patterns (Fig. 2) is matched. We assign a bridge between residues $i$ and $j$ if there are two H bonds characteristic of $\beta$-structure; in particular,

Parallel Bridge(i,j)=: [Hbond(i − 1,j) and Hbond(j,i + 1)] or
[Hbond(j − 1,i) and Hbond(i,j + 1)]

Antiparallel Bridge(i,j)=: [Hbond(i,j) and Hbond(j,i)] or
[Hbond(i − 1,j + 1) and Hbond(j − 1,i + 1)]

Parallel bridges are marked at $i$ and $j$ by lower-case letters, antiparallel ones by upper-case letters.

*Cooperative H-Bond Pattern: Helices*

A minimal helix is defined by two consecutive $n$-turns. For example, a 4-helix, of minimal length 4 from residues $i$ to $i + 3$, requires 4-turns at residues $i - 1$ and $i$,

4-helix(i,i + 3)=: [4-turn(i − 1) and 4-turn(i)]

i.e., an H bond ($i - 1$,$i$ +3) and an H bond ($i$,$i + 4$). Note that nothing is required about the H-bond state of residues $i + 1$ and $i + 2$. Similarly, two consecutive turns are required and a 3-helix of minimal length 3 from residue $i$ to $i + 2$ and a 5-helix of minimal length 5 from residue $i$ to $i + 5$:

3-helix(i,i + 2)=: [3-turn(i − 1) and 3-turn(i)]
5-helix(i,i + 5)=: [5-turn(i − 1) and 5-turn(i)]

Longer helices are defined as overlaps of minimal helices. Conventionally, these structures are called $\alpha$-helix, $3_{10}$-helix, and $\pi$-helix. In Table AIII, a 3-helix can be recognized by the pattern ⟩⟩3⟨⟨, a 4-helix by ⟩⟩44⟨⟨, and a 5-helix by ⟩⟩555⟨⟨. In the line SUMMARY, the residues bracketed by H bonds are labeled G, H, I, e.g.,

| | | | |
|---|---|---|---|
| 5-TURN | | | ⟩⟩555⟨⟨ |
| 4-TURN | | ⟩⟩44⟨⟨ | |
| 3-TURN | ⟩⟩3⟨⟨ | | |
| SUMMARY | GGG | HHHH | IIIII |

These helices are one residue shorter at each end than they would be according to rule 6.3 of IUPAC-IUB.[14] Examples of a 3-helix and a 5-helix are shown in Fig. 3.

*Cooperative H-Bond Patterns: β-Ladders and β-Sheets*

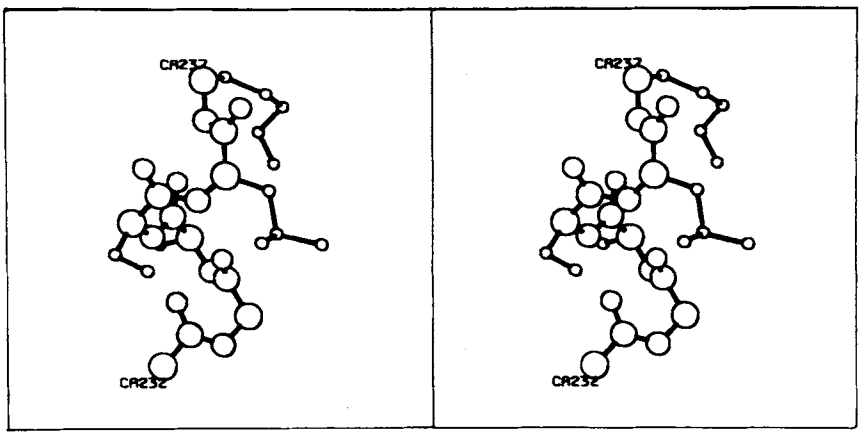
We coin the term "ladder" and define

ladder=: set of one or more consecutive bridges of identical type
sheet=: set of one or more ladders connected by shared residues

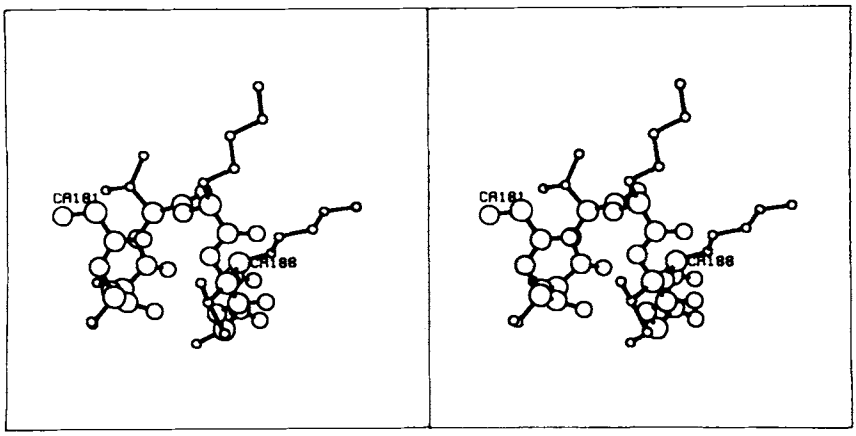
Ladders are given letter names, where a,b,c, . . . is for parallel, A,B,C . . . for antiparallel arrangement. Along the sequence, the first ladder is named "a" or "A," the second "b" or "B," etc. Sheets are also given letter names A,B,C . . . When the alphabet is exhausted, names restart at "a" or "A." In Table AIII, each residue is labeled in line SHEET by the sheet name and in lines BRIDGE by the names of the ladders in which it participates (at most two, one on each side). In line SUMMARY, residues in single bridges (ladders of length 1) are marked "B," all other ladder residues "E" (extended). Thus, continuous stretches of "E" are $\beta$-strands. The $\beta$-sheet notation is illustrated in Fig. 4.

*Secondary Structure Irregularities*

Long helices can deviate from regularity in that not all possible H bonds




(a)




(b)

Fig. 3. Stereoviews of secondary structure: (a) 3-helix ($3_{10}$-helix) and (b) 5-helix ($\pi$-helix). (a) 3-Helix Gly232–Lys237 from triose phosphate isomerase (1TIM). In Table AIII, it appears as the H-bond pattern

| | |
|---|---|
| 3-TURN | ⟩⟩⟩⟨⟨⟨ |
| SUMMARY | GGGG |
| SEQUENCE | GGASLK |

3-Helices are not uncommon, but have only two or three weak H bonds with $E$ about −1 kcal/mol and the C=O direction tilted away from the helix axis typically by 30°. (b) 5-Helix Gly181–Lys188 from alcohol dehydrogenase (4ADH), at the C-terminal end of a 4-helix. In Table AIII, it appears as the H-bond pattern

| | |
|---|---|
| 5-TURN | ⟩⟩⟩55⟨⟨⟨ |
| SEQUENCE | GSAVKVAK |

5-Helices are extremely rare; the longest one, shown here, has three H bonds. All stereoviews are by PLUTO (Sam Motherwell, unpublished). In Figs. 3 and 5, the larger atoms are backbone atoms with ¼ their hard-sphere radius ($C^{\alpha}$, 0.47; C of CO, 0.44; O, 0.35; N, 0.41 Å) and in Fig. 4 with twice these values; side-chain atoms are small, with 0.20-Å radius.

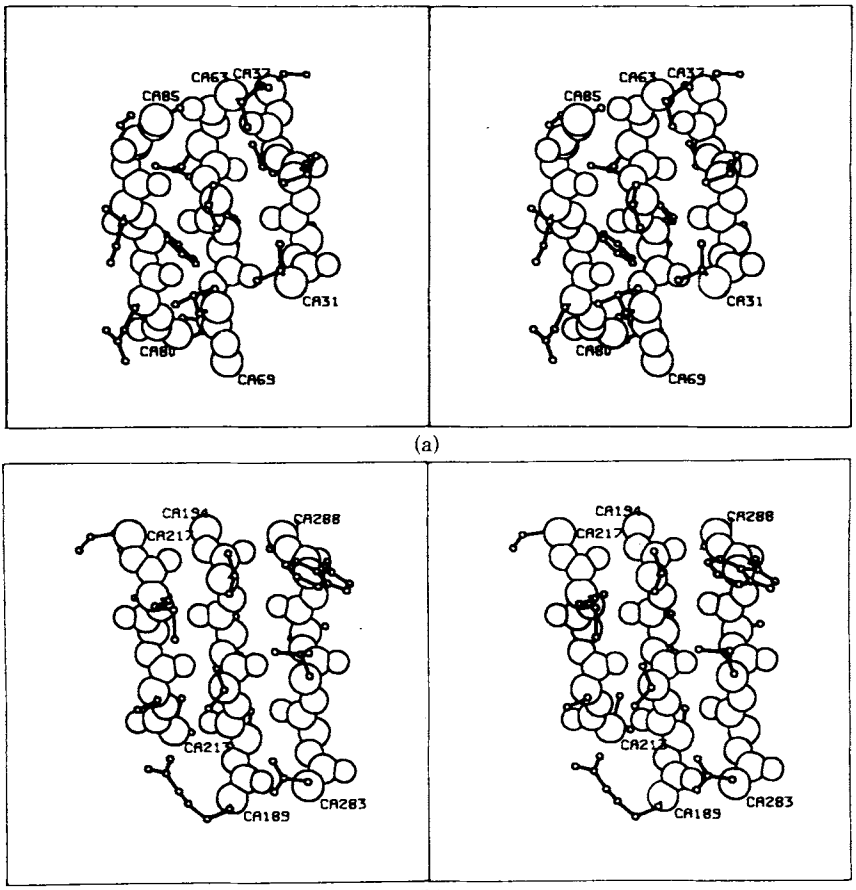



Fig. 4. Stereoviews of secondary structure: (a) antiparallel and parallel $\beta$-sheets with two ladders (three strands) each. (a) Two connected antiparallel $\beta$-ladders from trypsin (1PTN). The three participating strands are Val16(31)–Ser20(37), Ile46(63)–Gly51(69), and Glue62(80)–Ala67(85), where the first number is the sequential residue number from Table AIII and the number in parentheses the authors' residue identifier. The corresponding H-bond notation (Table AIII) is

```
SHEET..... CCC..... CCCC ..... CCCC....
BRIDGE2..... NNNN .....
BRIDGE1 ..... KKK ..... KKK..... NNNN....
SEQUENCE ..... VSLNS..... IQVRLG..... EQFISA ..
```

The middle strand participates in two ladders. Both ladders belong to sheet C. (b) Two connected parallel $\beta$-ladders, Arg172(189)–Gly177(194), Thr196(213)–Ile200(217), Asp266(283)–Ala271(288) from glutathione reductase (2GRS). The corresponding H-bond notation (Table AIII) is

```
SHEET ..... EEEE ..... EEE..... EEEE ....
BRIDGE2..... lll .....
BRIDGE1..... kkkk ..... lll..... kkkk.....
SEQUENCE..... RSVIVG..... TSLMI ..... DCLLWA ...
```

The first strand has two ladder partners. The three strands are part of sheet E.

are formed. This possibility is implicit in the above helix definition, e.g., two overlapping minimal helices offset by two or three residues are joined into one helix:

```
    ))44((                          ))44((
+     ))44((                   +      ))44((
=   ))4)X(4((   irregular       =   ))44XX44((   irregular
=    HHHHHHH                    =    HHHHHHHH
    )))))X((((   perfect            ))))XX((((   perfect
```

even though the third and/or fourth H bond is missing, compared to a perfect seven- or eight-residue helix. Such imperfections are often associated with a kink in the helix, e.g., due to a proline residue.

For $\beta$-structure, we define explicitly: a bulge-linked ladder consists of two (perfect) ladders or bridges of the same type connected by at most one extra residue on one strand and at most four extra residues on the other strand. This definition follows Richardson's[8] observation of $\beta$-bulges, a frequent lattice fault in $\beta$-sheets, but includes more general bulges than her main types. In naming ladders, a bulge-linked ladder is treated as one ladder (lines BRIDGE). In line SUMMARY, all residues in bulge-linked ladders are marked "E," including the extra residues.

## Geometrical Structure

### *Bend*

Bends are regions with high curvature. We quantify chain curvature at the central residue $i$ of five residues as the angle between the backbone direction of the first three and the last three residues. This definition of curvature is identical to that of Rose and Seltzer[5] but slightly different from that of Rackovsky and Scheraga.[15] For a bend at $i$, we require a curvature of at least 70°. The cutoff value was chosen by visual inspection of three-dimensional traces. With $\mathbf{C}^\alpha$ the position vector of $C^\alpha$, we define

$$\text{Bend(i)} =: [\text{angle}\{(\mathbf{C}^\alpha(\text{i}) - \mathbf{C}^\alpha(\text{i} - 2)),(\mathbf{C}^\alpha(\text{i} + 2) - \mathbf{C}^\alpha(\text{i}))\} > 70°]$$

and assign "S" for a bend at residue $i$.

### *Chirality*

We define chirality at each residue (except at the ends of the chain) as (Fig. 2)

$$\alpha(\text{i}) = \text{dihedral angle}(\mathbf{C}^\alpha(\text{i} - 1),\mathbf{C}^\alpha(\text{i}),\mathbf{C}^\alpha(\text{i} + 1),\mathbf{C}^\alpha(\text{i} + 2))$$

but report only the sign of $\alpha$ in Table AIII: "+" if $0° < \alpha < 180°$ and "−" if $-180° < \alpha < 0°$. Note that most helices have positive, most twisted $\beta$-ladders negative, chirality. We have found only one left-handed helix, in thermolysin. This rare specimen is shown in Fig. 5.

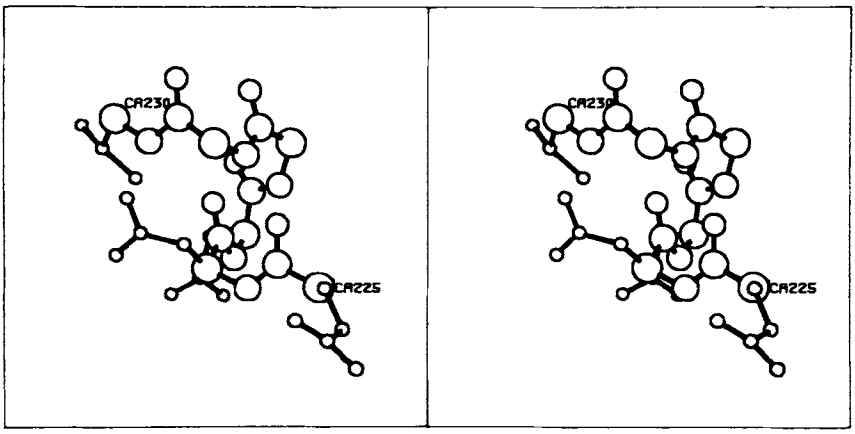



Fig. 5. Stereoviews of secondary structure: illustration of chirality. This short left-handed $\alpha$-helix, Gln225–Val230 from thermolysin (2TLN) is the only one known to us. In Table AIII (note that chirality is entered at the second residue of each quartet) it appears as:

| | |
|---|---|
| CHIRALITY | − − − |
| 4-TURN | ) ) 4 4 ( ( |
| SUMMARY | HHHH |
| SEQUENCE | QDNGGV |

## *SS Bonds*

SS bonds, i.e., covalent links between the $S^{\gamma}$ atoms of two Cys residues, are taken directly from the Data Bank SSBOND records, as they can be considered part of the amino acid sequence (primary structure). For the coordinate data sets used here, an S-S distance of less than 3.0 Å can also serve as a definition. The SS bonds are given names a,b,c . . . , and the participating residues noted by this name in the line SEQUENCE in Table AIII. Thus, Cys appears in the amino sequence either as C or as a lowercase letter.

## *Chain Breaks*

Chain breaks are assumed if the peptide bond length (distance C′-N) exceeds 2.5 Å. They are labeled "!" and counted as a break residue. Thus, "!" may reflect the absence of a chemical peptide bond, missing density in the crystallography map, or coordinate errors. The residues for which there are coordinates in the data set are numbered sequentially, including break residues. The resulting residue numbers often agree with the authors' except for proteins numbered according to sequence homology or those with missing density or chain breaks. In any case, inspection of the amino acid sequence in Table AIII always allows unambiguous identification of a residue.

## Structure Summary

To make contact with the usual notation of secondary structure and to facilitate comparison with intuitive assignments, we summarize secondary structure in a single line (SUMMARY in Table AIII). Structural overlaps are eliminated in this line by giving priority to H,B,E,G,I,T,S in this order, i.e., when several symbols coincide, the first one in this list is written. For example, a helix is also a series of bends, but the state helix is given higher priority. Pieces of 3- or 5-helix, reduced to less than minimal size due to overlaps, are labeled "T." A blank, by implication, means a piece of low curvature not in H-bonded structure.

## Static Solvent Exposure

*Physically*, we are interested in the number of water molecules in direct contact with the protein or with a particular part of the protein.

*Geometrically*, a very useful representation of a monomolecular layer of water is the surface described by all possible positions of a water molecule in touching contact with protein atoms. That was the idea of Lee and Richards'[10] water sphere rolling around the protein surface. Note that the surface associated with holes in the protein interior is very small, e.g., a hole that accommodates just one water molecule has zero area. For most of the protein exterior, however, the surface is proportional to the number of water molecules in the first hydration shell.

*Mathematically*, one calculates the surface by integrating a step function $f$ over all points $x$ on the surface of a sphere of radius $r(\text{atom}) + r(\text{water})$ around atom $i$. $f = 1$ if a water sphere centered at $x$ (by definition in contact with atom $i$) does not intersect with any other protein atom; otherwise, $f = 0$.

*Algorithmically*, we integrate by summing over a polyhedron made of 20, 80, 320, or more approximately equal triangles. The integration points are the triangle centers, the weights are the triangle area. The polyhedron is generated starting from an icosahedron; a recursive procedure then divides each triangle into four by connecting the midpoints of the sides and projects the three new vertices onto the surface of the sphere, ready for the next level of recursion. The final polyhedron is reminiscent of the shells of certain viruses and of Buckminster Fuller's architecture of geodesic domes. Hence, we call the algorithm "geodesic sphere integration." It is similar to the algorithm of Shrake and Rupley[16] and conceptually simpler than $z$-layer integration.

With 320 integration points, the surface area of a residue is accurate to within 1 Å$^2$; with 80 points, to within 4 Å$^2$. For myoglobin, the numerical values agree with those of Lee and Richards,[10] using their parameters. The numbers given here are based on slightly different values of atomic radii: 1.40 for O, 1.65 for N, 1.87 for C$^\alpha$, 1.76 for C of CO in the backbone, 1.80 for

all side-chain atoms,[17] and 1.40 for a water molecule following observed water–protein distances (Ref. 18 as cited in Ref. 19).

In Table AIII, we report the average number, $W$, of water molecules in contact with each residue. $W$ can be estimated from the surface area by

$$W = \frac{\text{Area}}{V(\text{water molecule})^{2/3}} \approx \frac{\text{Area}}{10}$$

since the surface is proportional to the volume of the monolayer, which, in turn, is proportional to the average number of molecules in the monolayer. For a water molecule volume of 30 Å$^3$ and area in Å$^2$, the conversion factor is $9.65 \approx 10$. Note that solvent exposure differs for a monomer and a dimer: here, it is calculated in the presence of all monomers in the data set (Table AI) but omitting HETATOMs (substrates, ligands, heme, etc.). The sum over all residues is the total solvent exposure of the protein.

## RESULTS AND DISCUSSION

### Choice of Proteins

Of the more than 100 coordinate data sets in the Protein Data Bank,[2] about 75 have complete backbone coordinates and a known amino acid sequence. When two protein data sets had more than a 50% sequence homology, i.e., identical amino acids in equivalent positions, the one with higher resolution, better refinement, or more secondary structure was chosen as representative, e.g., the first one was chosen of these pairs: serine proteinase 1SGA=1SGB by 61%; lactate dehydrogenases 4LHD=1LDX by 63%; carbonic anhydrase 1CAC=1CAB by 60%; chymotrypsin 2GCH=2CHA by 98%. Both were chosen of the following pairs: sulfhydryl proteinases actinidin/papain 2ACT=8PAP by 47%; immunoglobulins 1FAB=1REI by 47%; cytochrome c550/c2 155C=1C2C by 43%; chymotrypsin/trypsin 2GHA=1PTN by 42%; elastase/trypsin 1EST=1PTN by 38%; acid protease/penicillopepsin 1APR=1APP by 43%; $\alpha/\beta$ subunit of hemoglobin 2MHB($\alpha$)=2MHB($\beta$) by 44%. The final 62 data sets thus cover essentially all known different protein structures, except those not deposited with the protein data bank (Table AI).

### H-Bonded Structure

Backbone–backbone H bonds can be simply classified by the number of residues they bracket or, in our notation, by $n$ of $(i,i+n) = (\text{CO}(i),\text{NH}(i+n))$. Let us discuss the structural role of H bonds for each $n$.

H bonds $n = 0$ and $n = 1$ are sterically disallowed. A hydrogen bond $(i,i+2)$ can be formed between two consecutive peptide units for certain $\phi,\psi$ values of residue $i+1$. This local conformation is known as $C_7$ and leads to an extended strand roughly similar to a $\beta$-strand if it repeats. When it occurs as part of a tight turn, that turn is sometimes called a $\gamma$-turn.

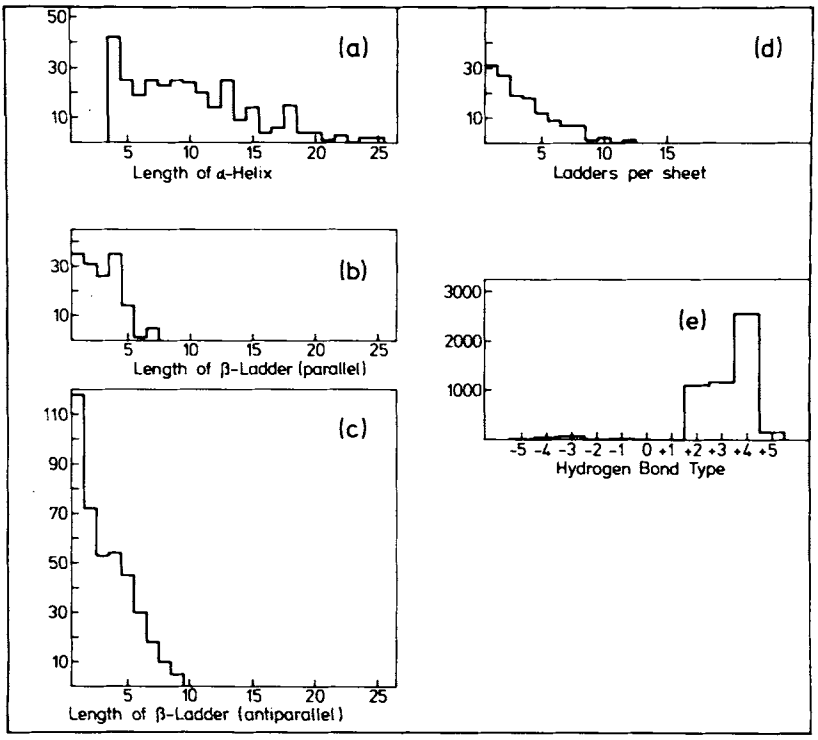



Fig. 6. The common feature of the size distribution of secondary structure segments is the gradual fall-off: larger sizes are less probable than smaller ones. Note that we give (b,c) the length of $\beta$-ladders (strand pairs) rather than the length of $\beta$-strands. A strand is often longer than the ladders in which it participates, since sheets tend to be trapezoidal rather than rectangular in shape. The number of bulge-linked ladders per sheet (d) is given as an indication of the width of the sheet. The width of a ladder is about 5 Å. In an ideal sheet, center strands take part in two ladders, edge strands in one: the number of ladders is equal to the number of strands minus one. In general, however, one strand can participate in more than one ladder on each side and the width of the sheet less than the number of ladders times 5 Å. Note: sheets consisting of a single bridge are not included in the histogram of ladders per sheet. (e) Number of H-bonds of type (CO($i$),NH($i + n$)). Due to the nature of L-amino acids, positive $n$ are heavily favored. The dominant peak at $n = 4$ represents $\alpha$-helices and 4-turns. We find that H bonds ($i$,$i + 2$) and ($i$,$i + 3$) are surprisingly common, though generally weak.

Using our H bond definition, we find that many $\beta$-strands have, in addition to the main interstrand H bonds, minor ($i$,$i + 2$) intrastrand H bonds [see peak in Fig. 6(e)]. These reflect part of the electrostatic stabilization of extended conformations due to the polar interaction of the C-O and N-H groups of adjacent peptide units, first shown by Flory's group[20] to be essential in stabilizing the $C_7$ conformation in solution. We speculate that $\beta$-strands originate as extended $C_7$ strands as the protein folds up. Outside of $\beta$-strands, we typically find one or two weak ($E < -1.0$ kcal/mol) ($i$,$i + 2$) H bonds per 100 residues, but most of them are neither repeating nor part of a tight turn.

H bonds with $n = +3,+4,+5$ are reported as turns or helices. Most $(i,i + n)$ hydrogen bonds for $n > 5$ or $n < -5$ are part of a bridge or ladder. Interestingly, H bonds $(i,i - 2)$, $(i,i - 3) \ldots (i,i - 5)$ are also rare. There is steric hindrance, e.g., in an $(i,i - 4)$ helix between the backbone oxygen and the first side-chain atom $C^{\beta}$.

3-Helices are more frequent than previously believed, although they are usually short and have mediocre hydrogen bonds. $\alpha$-Helices are rarely entirely pure: numerous H bonds in them are bifurcated, i.e., $(i,i + 4)$ and $(i,i + 3)$ or sometimes $(i,i + 5)$. The ends of $\alpha$-helices often are overwound, ending in a 3-turn or 3-helix, or underwound, ending in a 5-turn. Some of these cases were already noted and generalized by Schellman[21] and Richardson.[8] We even find a few 5-helices ($\pi$-helices)—see Fig. 3.

Tabulation of the relative number of H bonds in Table AI may be useful in calibrating spectroscopic determination (CD, laser Raman) of the percentage of secondary structure (e.g., by the algorithm of Provencher and Gloeckner[22]). In particular, we suggest that the distinction between parallel and antiparallel $\beta$-structure[23,24] in the reference spectra will improve the overall accuracy of these experiments.

## Accuracy of H-Bond and Secondary Structure Assignments

At best, secondary structure assignments can only be as accurate as the coordinates on which they are based. In using this dictionary, it is therefore very important to be aware of the state of resolution and refinement of each structure indicated in Table AI. The coordinate data sets range from refined structures at better than 1.5-Å resolution, where individual side chains can clearly be seen, to unrefined structures at a resolution just sufficient to trace the protein chain. As a test, we compare our assignments with those of the crystallographers and of Levitt and Greer[11] for three proteins of 1.5, 2.5, and 3.0 Å resolution (Table I).

For the *higher-resolution* structure of trypsin inhibitor (3PTI), Deisenhofer and Steigemann[25] assign an H bond when the N-O distance $d$ is no greater than 3.1 Å and list 18 backbone–backbone H bonds. Of these, we find all except Tyr35(CO)-Ala16(NH), which has $d = 3.1$; instead, we have Gly36(CO)-Ala16(NH), which has $E = -2.2$. In addition, we assign 11 others, due to the rather generous energy cutoff in our definition. One, Tyr35(CO)-Ile18(NH) is quite strong, with $E = -2.0$, consistent with the slow hydrogen-exchange rate of $2.6 \times 10^{-5}$ min$^{-1}$ measured by nmr.[26] Three others of type $(i,i + 3)$, with $E = -1.3, -1.7, -0.9$, form the well-known[8] 3-helix Asp3-Leu6. One $(i,i + 5)$ H bond, Asn24(CO)-Leu29(NH), is part of the $\beta$-hairpin. Six are of type $(i,i + 2)$, characteristic of the $C_7$ configuration: five weak ones and one stronger one ($E = -1.8$) in a $\gamma$-turn at Asn43. The additional H bonds assigned by us lead to identification of two unambiguous segments of secondary structure not cited by the authors but also assigned by Levitt and Greer.[11]

For the *medium-resolution* structure of cytochrome c550, Timkovich and Dickerson[27] use a conservative interpretation of hydrogen bonds and

TABLE I
Comparison of Secondary Structure Assignments for Three Proteins of Higher, Medium, and Lower Resolution

| Structure[a] | Original Authors (AU) | Levitt & Greer (LG) | This Work (KS) | |
|---|---|---|---|---|
| 3PTI | | | | |
| G1 | —[b] | 2–7 | 3–6 | Clearly $3_{10}$; LG have $\alpha$ |
| E1 | 16–25 | 14–25 | 18–24 | |
| E2 | 28–36 | 28–37 | 29–35 | |
| E3 | —[b] | 43–46 | 45–45 | $\beta$-Bridge, 2 H bonds |
| H1 | 47–56 | 47–55 | 48–55 | |
| 155C | | | | |
| H1 | 6–11 | 4–16[b] | 6–12 | 4-Turn 13–16 |
| G1 | 11–13 | — | 11–13 | Overlaps with H1 |
| E1 | — | 17–23[b] | 19–20 | AU have 2 H bonds; KS, 4 |
| E2 | — | 26–31[b] | — | Discontinuity at Asp28–Ile29 |
| E3 | — | 33–39 | 35–37 | AU have 2 H bonds; KS have 4 |
| H1 | — | 40–44[b] | — | KS have 3-Turn |
| H2 | 56–63 | 55–65 | 56–64 | |
| H3 | 73–79 | 71–80 | 73–80 | |
| H4 | — | 81–90[b] | — | Pro at 82, 84; possible helix |
| H5 | 107–118 | 106–118 | 107–117 | |
| 2ADK | | | | |
| H1 | 1–8 | 1–7 | 2–7 | |
| E1 | 10–14 | 8–15 | 10–14 | |
| H2 | 23–30 | 21–31 | 23–31 | |
| E2 | 35–38 | 34–38 | 35–38 | |
| H3 | 41–48 | 39–49 | 39–48 | |
| H4 | 53–62 | 52–61 | 52–62 | |
| H5 | 69–84 | 68–83 | 69–83 | |
| E3 | 90–94 | 88–95 | 90–93 | |
| H6 | 100–107 | 100–109 | 101–108 | |
| E4 | 114–118 | 113–120 | 114–118 | |
| H7 | 123–133 | 121–136[b] | 122–132 | $\alpha$-Helix ends in 3-turn |
| H8 | 144–158 | 141–157[b] | 143–157 | No ($i$,$i$ + 4) H bond at Asp 141 |
| H9 | 160–164 | 159–166 | 160–167[b] | Two weak H bonds at 167,168 |
| E5 | 169–173 | 169–175 | 170–173 | |
| H10 | 179–194 | 179–192 | 179–193 | |

[a] H = $\alpha$-helix, G = $3_{10}$-helix, E = $\beta$-strand. 3PTI = pancreatic trypsin inhibitor, 1.5-Å resolution, Diamond real-space refinement (Ref. 25). 155C = cytochrome c550, 2.5-Å resolution, Diamond model building to guide coordinates, assignments derived from the H-bonding diagram of Ref. 27. 2ADK = adenylate kinase, 3.0-Å resolution, unrefined (Ref. 28).

[b] Serious discrepancy (segment missing or boundary different by three or more residues).

give a minimal set of 41 backbone–backbone H bonds. We assign all of these, except Ala115(CO)-Gln119(NH), at the end of an $\alpha$-helix; instead, we see the helix end with the ($i$,$i$ + 3) H bond Ala115(CO)-Asp118(NH). We assign an additional 24 H bonds, of which 7 are the secondary partners of a bifurcated H bond, which is common in helices, and 8 others are marginal, with $E > -1.0$ kcal/mol. Of the remaining 9, four are of type ($i$,$i$ +

2) in approximate $\gamma$-turns at Glu2, Gly40, Lys 53, and Lys88; two are ($i$,$i$ + 4) H bonds at the end of $\alpha$-helices; two are ($i$,$i$ − 3) and ($i$,$i$ − 6) in the loop region Gln22-Asp28; and one is involved in forming the heme pocket by a tertiary contact between Thr80(CO) at a helix end and Met103(NH) in an extended strand. All of these have a meaningful structural interpretation. The resulting secondary structure assignments are consistent with the authors' H-bond list, except for the additional short parallel bulged $\beta$-strand pair, 19–20/35–37, which is due to two additional weak H bonds. Levitt and Greer[11] assign considerably more secondary structure (Table I), including a much longer parallel $\beta$-sheet 17–23/33–39 (probably too long), a $\beta$-strand 26–31 (roughly antiparallel to 17–23), a helix 40–44 (we assign a 3-turn), and a longer helix 81–90 (which has only two of the seven possible H bonds but looks very much like a helix in a $C^{\alpha}$ chain tracing and therefore may be seen to be a helix at higher resolution).

For the unrefined, *lower-resolution* structure of adenylate kinase (2ADK[28]), all secondary structure assignments (ours, the original authors',[28] and Levitt and Greer's[11]) are similar. Other lower-resolution coordinate data sets show more discrepancies, depending on the quality of the H bonds.

This detailed comparison shows that our H-bond energy cutoff, chosen out of necessity to allow for coordinate errors in lower-resolution data, typically leads to 50% more H bonds than conservative assignments in higher-resolution data (example, 3PTI). All these have a physical meaning in terms of electrostatic interaction energy and nearly all have an interpretation in terms of canonical secondary structure; and, most importantly, the increased number of H bonds does not give rise to spurious secondary structure assignments.

H-bond assignments become less certain for some lower-resolution data. For example, in the data sets 1APR, 3PGM, and 1ABP, Richardson[8] sees a number of $\beta$-strands, which, in Table AIII, do appear as uncurved (non-"S") strands but with relatively few H-bonded bridges between them. At least for 1APR, only partially refined at 2.5-Å resolution with tentative amino acid sequence, one may expect that more H bonds will form in the $\beta$-sheets on further refinement.

We conclude that our criteria for H-bonded secondary structure are relatively strict, in spite of a generous cutoff in the H-bonding energy. For higher-resolution data sets, our assignments are more accurate than those of Levitt and Greer,[11] and for lower-resolution data, they are conservative compared with both Levitt and Greer's program and Richardson's[8] visual processing.

## Secondary Structure Size

What is the extent of secondary structure cooperativity? Are there any preferred lengths of secondary structure segments? The length distributions [Fig. 6(a–c)] fall off almost monotonically with increasing length up

to a maximum segment length of about 30 Å, with parallel $\beta$-ladders slightly shorter. There appear to be no statistically significant peaks, either for an integral number of helical repeats or for typical domain sizes, with the possible exception of four-residue parallel $\beta$-ladders characteristic of the $\alpha/\beta/\alpha$ folding unit and, perhaps, 13- and 18-residue $\alpha$-helices. We speculate that protein folding, although cooperative, follows random polymer statistics approximately in that long segments are statistically less likely than short ones. The apparent maximum size of 30 Å perhaps reflects the maximum size of globular domains.

## OUTLOOK

The structure of influenza virus hemagglutinin,[29] with its 50-residue helix, shows that our data base certainly does not exhaust all possible variations in protein architecture. In spite of this limitation, this compilation will be used in the ongoing development of protein structure prediction methods.

## APPENDIX: DICTIONARY OF PROTEIN SECONDARY STRUCTURE

### Notes to Table AI

Proteins are ordered by function and can be found in the strip tables (Table AIII) and strip maps (Fig. AI) by their running number. % $\alpha$-helix, % $\beta$-antiparallel, % $\beta$-parallel = number of H bonds per 100 residues of type 4-turn, parallel and antiparallel bridge; these percentages can be compared with results from spectroscopy (CD, Raman, ir). Exposure = estimated number of water molecules in contact with protein surface (first hydration shell); it can also be read as the static exposed surface area in units of 10 Å$^2$. Exposure is calculated for the entire data set and then divided by the multiplicity of sequence-unique molecules, e.g., the data set 1INS has two copies each of the insulin A- and B-chain (multiplicity 2). Exposure given is that of the A- and B-chain in the tetramer. Number of residues is also for the sequence-unique molecule. Crystallographic resolution (Å) and refinement give some indication of the quality of the coordinates; both are taken from the Data Bank without further checking. In case of doubt, consult the original papers. Refinement code: D1 = Diamond model building to guide coordinates (Ref. 30); D2 = Diamond real-space refinement (Ref. 31); HK = Hendrickson-Konnert (Ref. 32); DO = Dodson, Isaacs, and Rollett (Ref. 33); JL = Jack and Levitt (Ref. 34); DS = Deisenhofer and Steigemann (Ref. 25); DF = difference Fourier; DC = difference Fourier with constraints; FD = difference Fourier and D1; LS = least squares; RL = restrained least squares; CL = constrained least squares; SD = steepest descent; LL = energy minimization of Levitt and Lifson (Ref. 35); HH = D2 and Hermans' REFINE2 and HK; DD = DS and D2; DL = DF and LS; DJ = D2 and JL; AD = Agarwal least squares (Ref. 36) and DO; DH = D2 and HK; DE = D2 and LL; MD = energy minimization of McQueen and DO; CS = constrained difference Fourier of Chambers and Stroud (Ref. 37); RE = real space and energy minimization; CC = constrained crystallographic refinement; CD = D2 and CORELS (Ref. 38).

TABLE AI

List of 62 Different Globular Proteins

| %AH | %BA | %BP | EXPO | M | LEN | RES | RF | | PROTEIN IDENTIFIER, NAME |
|---|---|---|---|---|---|---|---|---|---|
| **binding proteins** | | | | | | | | | |
| 38 | 4 | 0 | 610 | | 108 | 1.9 | DF | 1) | 1CPV CALCIUM-BINDING PARVALBUMIN B |
| 31 | 1 | 6 | 1423 | | 306 | 2.4 | -- | 2) | 1ABP L-ARABINOSE-BINDING PROTEIN |
| **electron transfer** | | | | | | | | | |
| 7 | 13 | 2 | 490 | | 85 | 2.0 | DF | 3) | 1HIP OXIDIZED HIGH POTENTIAL IRON PROTEIN (HIPIP). |
| **electron transport** | | | | | | | | | |
| 19 | 15 | 7 | 566 | | 85 | 2.8 | D2 | 4) | 2B5C CYTOCHROME B5 (OXIDIZED) |
| 57 | 0 | 0 | 665 | | 103 | 2.5 | -- | 5) | 156B CYTOCHROME B562 (E. COLI, OXIDIZED) |
| 34 | 2 | 0 | 620 | 2 | 103 | 2.0 | RL | 6) | 1CYT CYTOCHROME C (OXIDIZED). |
| 23 | 2 | 2 | 642 | | 112 | 2.0 | DC | 7) | 1C2C CYTOCHROME C2 (FERRI) |
| 23 | 0 | 3 | 781 | | 134 | 2.5 | D1 | 8) | 155C CYTOCHROME C550 |
| 38 | 2 | 0 | 482 | | 82 | 2.0 | CS | 9) | 251C CYTOCHROME C551 (OXIDIZED) |
| 9 | 9 | 0 | 316 | | 54 | 2.0 | DC | 10) | 1FDX FERREDOXIN (PEPTOCOCCUS AEROGENES) |
| 0 | 0 | 0 | 623 | | 98 | 2.8 | -- | 11) | 1FXC FERREDOXIN (SPIRULINA PLATENSIS) |
| 30 | 1 | 18 | 715 | | 138 | 1.9 | DE | 12) | 3FXN FLAVODOXIN (OXIDIZED) |
| 7 | 17 | 0 | 376 | | 54 | 1.5 | LS | 13) | 2RXN RUBREDOXIN (OXIDIZED, FE(III)) |
| 10 | 17 | 7 | 645 | | 125 | 2.7 | HK | 14) | 1AZU AZURIN |
| 2 | 21 | 10 | 513 | | 99 | 1.6 | DH | 15) | 1PCY PLASTOCYANIN |
| **hormones** | | | | | | | | | |
| 44 | 0 | 0 | 343 | | 36 | 1.4 | RL | 16) | 1PPT AVIAN PANCREATIC POLYPEPTIDE |
| 38 | 0 | 0 | 354 | | 29 | 3.0 | RE | 17) | 1GCN GLUCAGON (PH 6-7) |
| 29 | 12 | 0 | 301 | 2 | 51 | 1.5 | DL | 18) | 1INS INSULIN (A AND B CHAIN) |
| **hydrolase, phosphatide acyl** | | | | | | | | | |
| 37 | 7 | 0 | 712 | | 123 | 1.7 | AD | 19) | 1BP2 PHOSPHOLIPASE A2 |
| **hydrolases, O-glycosyl** | | | | | | | | | |
| 38 | 7 | 0 | 918 | | 164 | 2.4 | CL | 20) | 1LZM LYSOZYME (BACTERIOPHAGE T4) |
| 24 | 8 | 2 | 665 | | 129 | 2.5 | CD | 21) | 7LYZ LYSOZYME (HEN EGG WHITE, TRICLINIC) |
| **hydrolases, phosphoric diester** | | | | | | | | | |
| 19 | 20 | 3 | 842 | | 142 | <4 | -- | 22) | 1SNS STAPHYLOCOCCAL NUCLEASE (COMPLEX) |
| 14 | 28 | 2 | 709 | | 124 | 2.0 | SD | 23) | 1RNS RIBONUCLEASE-S |
| **hydrolases, proteinases** | | | | | | | | | |
| 28 | 5 | 9 | 1209 | | 308 | 2.0 | -- | 24) | 1CPA CARBOXYPEPTIDASE A |
| 3 | 13 | 3 | 1333 | | 324 | 2.5 | HK | 25) | 1APR ACID PROTEASE (RHIZOPUS CHINENSIS) |
| 7 | 32 | 6 | 1272 | | 323 | 2.8 | D2 | 26) | 1APP ACID PROTEINASE (PENICILLOPEPSIN, FUNGUS) |
| 27 | 9 | 4 | 1266 | | 316 | 2.3 | D2 | 27) | 2TLN THERMOLYSIN |
| 6 | 31 | 1 | 1033 | | 236 | 1.9 | HH | 28) | 2GCH GAMMA CHYMOTRYPSIN A |
| 3 | 37 | 3 | 821 | | 198 | 2.8 | D1 | 29) | 1ALP ALPHA LYTIC PROTEASE |
| 7 | 31 | 1 | 929 | | 223 | 1.5 | D2 | 30) | 1PTN BETA-TRYPSIN (NATIVE AT PH 8) |
| 4 | 34 | 2 | 745 | | 181 | 2.8 | D1 | 31) | 1SGA PROTEINASE A FROM STREPTOMYCES GRISEUS (SGPA) |
| 23 | 5 | 12 | 1058 | | 275 | 2.5 | FD | 32) | 1SBT SUBTILISIN BPN' |
| 4 | 35 | 0 | 1089 | | 240 | 2.5 | -- | 33) | 1EST TOSYL-ELASTASE |
| 21 | 19 | 1 | 923 | | 218 | 1.7 | LS | 34) | 2ACT ACTINIDIN |
| 19 | 15 | 1 | 968 | | 212 | 2.8 | D1 | 35) | 8PAP PAPAIN |
| **immunoglobulins** | | | | | | | | | |
| 1 | 34 | 2 | 2101 | | 428 | 2.0 | -- | 36) | 1FAB LAMBDA IMMUNOGLOBULIN FAB |
| 1 | 37 | 5 | 492 | 2 | 107 | 2.0 | CC | 37) | 1REI BENCE-JONES IMMUNOGLOBULIN (VARIABLE PORTION) |
| **isomerases** | | | | | | | | | |
| 23 | 2 | 5 | 1220 | | 230 | 2.8 | HK | 38) | 3PGM PHOSPHOGLYCERATE MUTASE (DE-PHOSPHO) |
| 35 | 1 | 15 | 1026 | 2 | 246 | 2.5 | DO | 39) | 1TIM TRIOSE PHOSPHATE ISOMERASE |
| **lectin (agglutinin)** | | | | | | | | | |
| 2 | 35 | 0 | 1125 | | 237 | 2.4 | MD | 40) | 3CNA CONCANAVALIN A |
| **lyase, carbon-oxygen** | | | | | | | | | |
| 4 | 20 | 5 | 1273 | | 256 | 2.0 | D1 | 41) | 1CAC CARBONIC ANHYDRASE FORM C |
| **oxidoreductases** | | | | | | | | | |
| 15 | 12 | 13 | 937 | | 162 | 2.5 | -- | 42) | 1DFR DIHYDROFOLATE REDUCTASE (COMPLEX) |
| 22 | 11 | 10 | 1505 | 2 | 333 | 2.9 | D1 | 43) | 1GPD D-GYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE |
| 17 | 11 | 9 | 1639 | | 374 | 2.4 | DJ | 44) | 4ADH APO-LIVER ALCOHOL DEHYDROGENASE |
| 27 | 6 | 7 | 1753 | | 329 | 2.0 | D2 | 45) | 4LDH LACTATE DEHYDROGENASE, APO ENZYME M4 |
| 22 | 12 | 7 | 2354 | | 461 | 2.0 | -- | 46) | 2GRS GLUTATHIONE REDUCTASE |
| 1 | 33 | 1 | 686 | 4 | 151 | 2.0 | HK | 47) | 2SOD CU,ZN SUPEROXIDE DISMUTASE |
| **oxygen storage** | | | | | | | | | |
| 65 | 0 | 0 | 842 | | 153 | 2.0 | D2 | 48) | 1MBN MYOGLOBIN (FERRIC IRON - METMYOGLOBIN) |
| **oxygen transport** | | | | | | | | | |
| 62 | 0 | 0 | 706 | | 136 | 1.4 | DS | 49) | 1ECD HEMOGLOBIN (ERYTHROCRUORIN DEOXY) |
| 58 | 0 | 0 | 1415 | | 287 | 2.0 | DD | 50) | 2MHB HEMOGLOBIN (HORSE, AQUO MET) |
| 47 | 0 | 0 | 864 | | 148 | 2.0 | D1 | 51) | 1LHB HEMOGLOBIN(MET)-CYANIDE V (SEA LAMPREY) |
| 62 | 0 | 0 | 824 | | 153 | 2.0 | D2 | 52) | 1HBL LEGHEMOGLOBIN (ACETATE,MET) (YELLOW LUPIN) |
| **plant seed protein** | | | | | | | | | |
| 33 | 6 | 0 | 301 | | 46 | 1.5 | HK | 53) | 1CRN CRAMBIN |
| **proteinase inhibitors** | | | | | | | | | |
| 13 | 14 | 2 | 351 | 4 | 56 | 1.9 | DJ | 54) | 1OVO OVOMUCOID THIRD DOMAIN |
| 13 | 23 | 2 | 632 | | 107 | 2.6 | DO | 55) | 2SSI STREPTOMYCES SUBTILISIN INHIBITOR |
| 12 | 17 | 0 | 412 | | 58 | 1.5 | D2 | 56) | 3PTI TRYPSIN INHIBITOR |
| **toxins** | | | | | | | | | |
| 3 | 17 | 0 | 511 | | 71 | 2.8 | -- | 57) | 1CTX ALPHA COBRATOXIN |
| 71 | 0 | 0 | 222 | 2 | 26 | 2.0 | HK | 58) | 1MLT MELITTIN |
| 0 | 29 | 0 | 406 | | 62 | 1.4 | HK | 59) | 1NXB NEUROTOXIN B (PROBABLY IDENTICAL TO ERABUTOXIN B |
| **transferases** | | | | | | | | | |
| 47 | 0 | 10 | 1251 | | 194 | 3.0 | -- | 60) | 2ADK ADENYLATE KINASE |
| 20 | 2 | 10 | 1456 | | 293 | 2.5 | D1 | 61) | 1RHD RHODANESE |
| **transport** | | | | | | | | | |
| 7 | 33 | 7 | 652 | | 114 | 1.8 | DO | 62) | 2PAB PREALBUMIN (HUMAN PLASMA) |

TABLE AII
Structure Notation Used in Table AIII

| | |
|---|---|
| First line: | running number 1–62, data set identifier (3PTI,4LDH . . .), protein name, [function], {source} |
| SHEET. . . | One-character name of $\beta$-sheet ("A," "B," "C" . . .) in which residue $i$ participates. |
| BRIDGE2 . . . | One-character name of $\beta$-ladders in which residue $i$ participates, |
| BRIDGE1 . . . | "A," "B," "C" . . . = antiparallel, "a," "b," "c" . . . = parallel. Ladders are named sequentially from N- to C-terminus. A $\beta$-strand can be part of two ladders, one to each side, so there are two lines for the possible ladder partners. Each ladder name appears twice, once for each participating strand. Partner strands can thus be easily identified by identical letters. The sheet topology can be reconstructed by starting from a $\beta$-strand and tracing all partners and their partners. |
| CHIRALITY | "+" or "−" Chirality at residue $i$ is the sign of the dihedral angle defined by $C^{\alpha}$ $i-1$ to $i+2$. Thus, a right-handed $\alpha$-helix has "+," an ideal twisted $\beta$-strand "−." |
| BEND . . . | "S" = five-residue bend centered at residue $i$. |
| 5-TURN . . . | Hydrogen-bonding pattern for turns and helices: |
| 4-TURN . . . | ")" = backbone CO of this residue makes H bond $(i,i+n)$ |
| 3-TURN . . . | "(" = backbone NH of this residue makes H bond $(i-n,i)$ "X" = both CO and NH make H bond "3," "4," "5" = residues bracketed by H bond |
| SUMMARY . . . | Structure summary: "H" = 4-helix ($\alpha$-helix) "B" = residue in isolated $\beta$-bridge "E" - extended strand, participates in $\beta$-ladder "G" = 3-helix ($3_{10}$-helix) "I" = 5-helix ($\pi$-helix) "T" = H-bonded turn "S" = bend In case of structural overlaps, priority is given to the structure first in this list. |
| EXPOSURE . . . | Solvent exposure is the estimated number of water molecules in contact with residue $i$. The scale is 0–9; "*" = more than 9 water molecules. Exposure can be read as solvated surface area in units of 10 Å$^2$. |
| SEQUENCE . . . | Amino acid sequence in one letter code: "a," "b," "c" . . . are Cys residues labeled by their SS-bond name. "!" = chain break (peptide bond length exceeds 2.5 Å). Residues including chain breaks are numbered sequentially within the coordinate data set, irrespective of the residue identifier given there. Thus, the total number of residues is equal to the total number of print positions minus the number of chain breaks. |

TABLE AIII

Strip Tables of Secondary Structure Assignment for 62 Different Proteins in the Notation of Table AII

```
1)  1CPV  CALCIUM-BINDING PARVALBUMIN B [CALCIUM BINDING PROTEINS] (CARP: CYPRINUS CARPIO).....................1CPV
    SHEET....                                                    AAA                                        AAA
    BRIDGE2..
    BRIDGE1..                                                    AAA                                        AAA
    CHIRALITY   +-----+++ +++++++++- +++--+++++ ++-++++--+ ++++++++++ -+-+-+---+ ++++++++++ -++------++ ++++++++++ +-+-+-+-++
    BEND.....    SS S SSS SSSSSSSS S SS   SSSSS SSSS SSS S SSSSSSSSSS  SS  SS  S SSSSSSSSSS SSS     SS SSSSSSSSS  SS SS   SS
    5-TURN...                              > 5555<
    4-TURN...       >>>> XXXX<<<<       >>>>X< <<<     >> >>XXXXX<<< <       >> >4<<< >444 <      >>> >XXXX<<<<         >>>
    3-TURN...  >33<        >>><<<              >33<>3 3<  >>3<<         >> 3XX3<<>33< >33<    >3 3<    >33X 33<    >3
    SUMMARY..   TT S HHH HHHHHHHT S SS   HHHHH HHTT TTS H HHHHHHHHHH  SS  SEEEH HHHHTSSTTT STT    HH HHHHHHHHT  TT SSEEEHH
    EXPOSURE.  4**5938*74 2640*81*6 7*5392*510 580225*854 *60**22471 1*9*63804* *2196*0753 *9*1740*3 12*901*701 *96830008*
  1 SEQUENCE.  APAGVLNDAD IAAALEACKA ADSFNHKAFF AKVGLTSKSA DDVKKAFAII DQDKSGFIEE DELKLFLQNF KADARALTDG ETKTFLKAGD SDGDGKIGVD

    SHEET....
    BRIDGE2..
    BRIDGE1..
    CHIRALITY  ++++++
    BEND.....  SSSSSS
    5-TURN...    >5555<
    4-TURN...  >XXX<<<<
    3-TURN...  3<>33<
    SUMMARY..  HHHHHHH
    EXPOSURE.  3136208*
101 SEQUENCE.  EFTALVKA

2)  1ABP  L-ARABINOSE-BINDING PROTEIN (BACTERIAL: ESCHERICHIA COLI).....................................................1ABP
    SHEET....        AAA                                AAA                                B                          B
    BRIDGE2..
    BRIDGE1..        aaa                                aaa                                b                          b
    CHIRALITY   +-+---+-- --+-++++++ ++++++++++ -+----+--+ +-++++++++ +++++-+--- --+--+-+++ ++++++++++ -+--+---+- -+--+-+---
    BEND.....               SSSSSSSS SSSSSSSSSS SSS         SSSSSSSSS SSSSSS          S SS  SSSSSSSSSS S       SS    SSS
    5-TURN...
    4-TURN...                >4>><>X X>XXXXX<<< <           >>>>XX<XX X>++++              > 4>>X>XXX<< <<
    3-TURN...                >33< >3 3< >33<                   >33X3 3<>>3<<                  >3 3<             >33<
    SUMMARY..        EEE    SSTTHHHH HHHHHHHHHH SSS   EEE   SHHHHHHHH HHHHHT      B  S SS  TTHHHHHHHH H     B SS     STT
    EXPOSURE.  **651@@@@2 65*240l*l@ 574@5*@2** 631*16*548 492**1491@ 67@3955190 @@@@@563*3 @64@255@*6 **291@@@l@ 40299*2*55
  1 SEQUENCE.  ENLKLGFLVK QPEEPWFQTE WKFADKAGKD LGFEVIKIAV PDGERTLNAI DSLAASGAKG FVICTPDPKL GSAIVAKARG YDMKVIAVDD QFVNAKGKPM

    SHEET....        C                                  DDD                                D D                        D
    BRIDGE2..                                           ee
    BRIDGE1..        c                                  ddd                                e e                        d
    CHIRALITY  ++---+---+ ++++++++++ ++++++-+- -+++----++ -+++-+++++ ++++++++++ +-+--+++-- ----+-+-++ ++++++++++ ++++++---
    BEND.....   SS      S SSSSSSSSSS SSSSSSS    SSS          SSSSSSS SSSSSSSSSS SSS  SSS       SSSSSS SSSSSSSSSS SS  S  S
    5-TURN...                        >55>5<55<                                >55 55<                  >5555<
    4-TURN...           > >>>XXXXXXX XXXXX<<<<                  >4>>X >XXXXXXXX< <<<                >> >>XXXXXXX< <<<
    3-TURN...             >3><3<                  >33<          >>3<<       >>3<<                    > 33<  >>3<<
    SUMMARY..   SS    B S HHHHHHHHHH HHHHHHHHT   STT  EEE     SGGGGHH HHHHHHHHH HHS  SSS E E   SSSSSH HHHHHHHHH HH  S  S E
    EXPOSURE.  *9211@1@@@ 7*1@8*@@*3 @8*21**963 86*4@1@@@@ 119*3616*4 1@4@@18217 89216**33* *292*45695 1@8*1@6612 **2**18400
101 SEQUENCE.  DTVPLVMMAA TKIGERQGQE LYKEMQKRGW DVKESAVMAI TANELDTARR RTTGSMDALK AAGFPEKQIY QVPTKSNDIP GAFDAANSML VQHPEVKHWL

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

```
    SHEET....  DD                          E E                   EE                                C    E
    BRIDGE2..                                                     G
    BRIDGE1..  dd                          f f                   ff                                c    G
    CHIRALITY  -+-+-+++++ ++++++++--+ --++++--- +-++++++++ +-+---++-- ----++++++ ++++++++++ +--+------+ ---++---+- ++++++++++
    BEND.....     SSSSSSS SSSSSSSSSS   SSS      SSSSSSSSSS SSSS  S         SSSS SSSSSSSSSS SSS     S             SSSSSSSSSS
    5-TURN...                                                                  >5 555<                            >55
    4-TURN...        >44>X >>XX<<<<               >44>X44< <              >4>4X> <>XX>XXXX< <<<                  >4>4X4<
    3-TURN...        >33X3 3<>33<       >33<      >33X>3<<                >33< > >3<X>3<<                      >33X>3<<
    SUMMARY..  EE SSSTTTH HHHHHHHSSS   STT    E ESSTTTHHHH SSSS  S    EE   TTTTH HHHHHHHHHH HHT     S    B    B   SSSTTTGGGT
    EXPOSURE.  0001102102 0008107?*4 2*33500000 0102204910 8*9*520070 0000002500 5501620684 35**6*17** 562*724410 9*42**729*
201 SEQUENCE.  IVGMNDSTVL GGVRATEGQG FKAADIIGIG INGVDAVSEL SKAQATGFYG SLLPSPDVHG YKSSEMLYNW VAKDVEPPKF TEVTDVVLIT RDNFKEELEK

    SHEET....
    BRIDGE2..
    BRIDGE1..
    CHIRALITY  -+++
    BEND.....  SS
    5-TURN...  55<
    4-TURN...  4<
    3-TURN...    >33<
    SUMMARY..  TT TT
    EXPOSURE.  55238*
301 SEQUENCE.  KGLGGK

3)  1HIP  OXIDIZED HIGH POTENTIAL IRON PROTEIN (HIPIP) [ELECTRON TRANSFER, IRON-SULFUR PROTEIN] {CHROMATIUM VINOSUM}....1HIP
    SHEET....        A             B                                C  AA          A AAA      A AA   B      C
    BRIDGE2..                                                                     C CC       C CC
    BRIDGE1..        a             D                                E  BB            BB       a      D      E
    CHIRALITY  --++---++ -++++-+--- +-+++-++++ ---+---+++ +-+++-+++- ++-++--+++ ---++-++-- --++---++- --+
    BEND.....     SSS  SS  SSSSSS    SSSSSSS SS S  SSS SSS   SSS SS   S SSS  SS     SSSSS     SS   S
    5-TURN...                >5555<
    4-TURN...              >>>><<<<         >>44 <<
    3-TURN...    >33< >33 <           >>3<<          >>3< <>>3<X33<> 33X33<       >33X33<   >33< >33<
    SUMMARY..      TTB  TT  HHHHHHT B SSGGGGSSHHH H  SSS GGG   GGGBTTEE TTSTT  SSE EEETTSTT E EETT B TT   B
    EXPOSURE.  *53*62149* 3860*83526 860**2*299 559*6553** 0514416646 7*39646*91 302*61*870 0129002440 59*99
  1 SEQUENCE.  SAPANAVAAD NATAIALKYN QDATKSERVA AARPGLPPEE QHCADCQFMQ ADAAGATDEW KGCQLFPGKL INVNGWCASW TLKAG

4)  2B5C  CYTOCHROME B5 (OXIDIZED) [ELECTRON TRANSPORT] {COW LIVER: BOS TAURUS}..........................................2B5C
    SHEET....     AA          BB  BBA AAA  AAAA                      A A                                A AAA
    BRIDGE2..                     cc   DDD                                                          D DD
    BRIDGE1..     aa          EE  EEB BBB  BBBB                      c c                                aa
    CHIRALITY  -----++++ ++++-+---++ ----+---++ ++++++-+-- +++++++++- +++++++++- --++++++++ +++------++ +++
    BEND.....        SSSS SS  SSS      SSS      SSSSSS SS  SSSSSSS SS   SSSSSSS    SSSSSSSS SS      SS SSS
    5-TURN...                                                          >5555 <
    4-TURN...       >>44< <                    > 44><44<    >>>4<<<     >>>><<<<   >>>><<<<         >44 4<
    3-TURN...         >> 3<< >33<     >33<   > 33X>3<X33< >33X>3X<3<       >>3<<       >>>X <<<    >>3 XX3<<
    SUMMARY..      EE HHHH TT EETTEEE EEETTEEEE  TTTGGG TT  SHHHHHTTTE E HHHHHHTT   HHHHHHGG GGEEEEE GG GGGG
    EXPOSURE.  *8**453*60 **3*78*302 0028*60151 4*21**3*63 8*726*5212 512*737**7 56**3**82* *37416067* 16**7
  1 SEQUENCE.  AVKYYTLEQI EKHNNSKSTW LILHYKVYDL TKFLEEHPGG EEVLREQAGG DATEDFEDVG HSTDARELSK TFIIGELHPD DRSKI

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
5)  156B  CYTOCHROME B562 (OXIDIZED) [ELECTRON TRANSPORT] (ESCHERICIA COLI).....................................156B
    SHEET....
    BRIDGE2..
    BRIDGE1..
    CHIRALITY   +++++++++ +++++++++- +-++++++++ ++++++++++ ++---+++++ ++++++++++ ++++++++++ ++++++-++ ++++++++++ ++++++++++
    BEND.....    SSSSSSSS SSSSSSSSS   SSSSSSSSS SSSSSSSSSS SSS  SSSSS   SS  SSSS SSSSSSSSSS SSSSSSSSS  SSSSSSSSSS SSSSSSSSSS
    5-TURN...                           >55>5<5 5<                                                >5555<
    4-TURN...   >>>>XXXXX XX<<X<44<     >4>>X>X XXXXXX<<<<     >444<       >>>>X XXXXXXXXXX XXXXX<<<<  >>>>XXXX<X XX><<XX44<
    3-TURN...     >>><<< >33X>3<<    >33<          >>> <<<    >33 <  >33< >> 3<< >33<    >33<          >>3XX 3<<   >33<
    SUMMARY..    HHHHHHHH HHHHHTTTS   STTTTHHHH HHHHHHHHHG GGS   TTTTT   SSTTHHHH HHHHHHHHHH HHHHHHHHT  SHHHHHHHHH HHHHHHHHHH
    EXPOSURE.  *64*716*43 **13*718*2 *698005600 8*1961Ø*6Ø 4*27Ø**2** 77**2*61*8 Ø29422*Ø06 6Ø3*31**59 4**Ø36328* 1782*564**
  1 SEQUENCE.  ADLEDDMQTL NDNLKVIEKA BBZKANDAAL VKMRAAALNA QKATPPKLED NSQPMKDFRH GFDILVEGID DALKLANEGK VKEAQAAAEQ LKTTRNAYHQ

    SHEET....
    BRIDGE2..
    BRIDGE1..
    CHIRALITY   +
    BEND.....   S
    5-TURN...
    4-TURN...   <
    3-TURN...
    SUMMARY..   S
    EXPOSURE.   93*
101 SEQUENCE.   KYR

6)  1CYT  CYTOCHROME C (OXIDIZED) [ELECTRON TRANSPORT] (ALBACORE TUNA HEART: THUNNUS ALALUNGA).....................1CYT
    SHEET....                                          A                   A
    BRIDGE2..
    BRIDGE1..                                          A                   A
    CHIRALITY   -++++++++ +++++++--+ -++--+-+-- -+++++++--+ ++-++-----+ +++++-+--+- +++++++++- ++++-++-+- +----+-+++ ++++++++++
    BEND.....    SSSSSSSS SSSSSSS    SSS    SS    SS SSS S S  SS    S SSSSS    S SSSSSSSSS  SSSSSSS           SSS SSSSSSSSSS
    5-TURN...           > 5555<
    4-TURN...   >>>>XXXXX <<<X444<                                      >> >4<<<    > >>>XXX<<<X >44<<           >>>> XXXXXXXXX<
    3-TURN...              >33<    >33<        >33X33<>3 3<>33<  >> 3<<                  >33<
    SUMMARY..    HHHHHHHH HHHTTTT    STT    SS    TT TTSBT T  TT    H HHHHS  B S HHHHHHHHH  HHHHSTT           HHH HHHHHHHHHH
    EXPOSURE.  8*65*0**21 8*923*5321 **82*89912 23932465*0 0*0*84*43* 329*924927 8*21682298 0**30972*5 *5*11***45 3920113179
  1 SEQUENCE.  GDVAKGKKTF VQKCAQCHTV ENGGKHKVGP NLWGLFGRKT GQAEGYSYTD ANKSKGIVWN NDTLMEYLEN PKKYIPGTKM IFAGIKKKGE RQDLVAYLKS

    SHEET....
    BRIDGE2..
    BRIDGE1..
    CHIRALITY   +
    BEND.....   S
    5-TURN...
    4-TURN...   <<<
    3-TURN...
    SUMMARY..   HH
    EXPOSURE.   2Ø*
101 SEQUENCE.   ATS

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

```
7)  1C2C  FERRICYTOCHROME C2 [ELECTRON TRANSPORT] {BACTERIAL: RHODOSPIRILLUM RUBRUM}.....1C2C
    SHEET....                  A  A                        B                    B
    BRIDGE2..
    BRIDGE1..                  a  a                        B                    B
    CHIRALITY  -++++++++ ++++++++--+ -++--+-+-- -++++-+--+ -+-++-+--+ +++++++--- -+-++++++ ++-++++++ +--------+ ++----++++
    BEND.....   SSSSSSSS SSSSSSS    SSS   SSS    SS SS  S S  SSS   S SSS SSSS      SSSSSSS SSSSSSSSS SS  SS   S        SSSS
    5-TURN...  >5555<                                                                                            >5
    4-TURN...   >>>>XXX<< X<4><44<                                  >4 44<           >>>><X< <4<   >444 <              >>>>X
    3-TURN...        >> ><<X33<    >33< >33<   >33X33<>3 3<>33<   > 33<                    >33<>3 3<
    SUMMARY..   THHHHHHHH HTTTBTTB   STT   TTS    TT TT BT T  TTS   T TTS SSSS   B  SHHHHHH TTSSSTTTTT TS  SS   S       HHHH
    EXPOSURE.  *6*232Ø6*1 2**2966212 **7364**12 124213*641 116**6842* 13643*6*52 5259821353 2*9Ø84116* 62588*2*1* 568*2**921
  1 SEQUENCE.  EGDAAAGEKV SKKCLACHTF DQGGANKVGP NLFGVFENTA AHKDNYAYSE SYTEMKAKGL TWTEANLAAY VKNPKAFVLE KSGDPKAKSK MTFKLTKDDE

    SHEET....
    BRIDGE2..
    BRIDGE1..
    CHIRALITY  +++++++-++
    BEND.....  SSSSSSSSSS
    5-TURN...  55><555<
    4-TURN...  XXXX<XX<4< <
    3-TURN...    >33<>33<
    SUMMARY..  HHHHHHHHHH
    EXPOSURE.  95104Ø23*6 3*
101 SEQUENCE.  IENVIAYLKT LK

8)  155C  CYTOCHROME C550 [ELECTRON TRANSPORT] {PARACOCCUS DENITRIFICANS}.....155C
    SHEET....                    AA                A A
    BRIDGE2..
    BRIDGE1..                    aa                a a
    CHIRALITY   ---++++++ +++++++++- +--++----+ --++---+++ +++---++-+ +----+++++ +++++++++- -+++++++- -+++++++- +-++---+++
    BEND.....      SSSSSS SSSSSSSS      SSS SS    SSS   SS  SSS SS SS SS   SSSSS SSSS S      SSSSSSSSS SSSSSSSSS   S    S
    5-TURN...      >5555<
    4-TURN...      >>>>X< <X<4><44<                                  >>>>XX X<<<<       >>>>XX<<< <>44><44<
    3-TURN...           > >><<X33<                          >33<>33<                        >>3< <>33< >33<
    SUMMARY..      SHHHHH HHTTTTTTEE    SSS SS    SSEEE SS  TTS TT SS SS   HHHHH HHHH S      SHHHHHHHH SSTTTTTTTS   S    S
    EXPOSURE.  **3726*Ø** 51**478623 Ø72**2764* Ø2*9122261 113680166* 85*31*1268 Ø1**699112 5762158434 6Ø95113873 *9*51*3*4*
  1 SEQUENCE.  NEGDAAKGEK EFNKCKACHM IQAPDGTDIK GGKTGPNLYG VVGRKIASEE GFKYGEGILE VAEKNPDLTW TEANLIEYVT DPKPLVKKMT DDKGAKTKMT

    SHEET....
    BRIDGE2..
    BRIDGE1..
    CHIRALITY  --+++++++ +++++++-++ --------+-+ ++
    BEND.....      S SSSS SSSSSSSS   S SS S  SS S
    5-TURN...
    4-TURN...       >>>>X XXXX<<<<
    3-TURN...         >33X33<           >3 3<
    SUMMARY..      S HHHH HHHHHHHS   S SS S  ST T
    EXPOSURE.  9*27*44222 Ø2112*31** 6795678478 4146
101 SEQUENCE.  FKMGKNQADV VAFLAQDDPD AXXXXXXXXX XXXX

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
 9)  251C  CYTOCHROME C551 (OXIDIZED) [ELECTRON TRANSPORT] {PSEUDOMONAS AERUGINOSA}.....................................251C
     SHEET....                                                          A         A
     BRIDGE2..
     BRIDGE1..                                                          A         A
     CHIRALITY   -++++++++- +++++--++- +-+---++++ ++++-+-+++ +++++++++- -++++--+-- ----+--+++ ++++++++++
     BEND.....    SSSSSSSS S SSS  SSS  SSS  SSSS SSSS S SS  SSSSSSSSSS    SSSSSS      S  SSS SSSSSSSSSS
     5-TURN...       >555 5<
     4-TURN...   >>>>X<<<<  >444<          >>>>X X<<<<   >> >>XXXXX<<<< <                >>>> XXXXXX<<<<
     3-TURN...   >33<  >33 X>3<<                    >33X33<                                        >>3< <
     SUMMARY..    HHHHHHHT TGGGT  SSS  SSS  HHHH HHHHTT TTH HHHHHHHHHH  B SSSSSS  B   S  HHH HHHHHHHHHT
     EXPOSURE.  *71*802**7 32685225*8 *8812Ø2*81 Ø4*248**41 95*2Ø*8269 23*6693**8 59*49149*2 Ø62ØØ82138 6*
   1 SEQUENCE.  EDPEVLFKNK GCVACHAIDT KMVGPAYKDV AAKFAGQAGA EAELAQRIKN GSQGVWGPIP MPPNAVSDDE AQTLAKWVLS QK

1Ø)  1FDX  FERREDOXIN [ELECTRON TRANSPORT] {PEPTOCOCCUS AERUGENES}.....................................................1FDX
     SHEET....   AA                    B      B                    A A
     BRIDGE2..
     BRIDGE1..   AA                    B      B                    A A
     CHIRALITY   ----+--+ -+++++++-+- ++++-+---- -+++--+-++ ++++-+-+++ -+
     BEND.....       S        SSSS SS     SSS     SSS       SSSS SS     S
     5-TURN...
     4-TURN...              >444<                 >444<   >> >4<<<
     3-TURN...     >33<     >3>X3<X33 <           >33<    >3 3<
     SUMMARY..    EE TT      TTGGG TT   B SSS  B  TTT     H HHHH SS  E ES
     EXPOSURE.  23625*63*5 724169*297 *1Ø*5*9552 74*6299622 14*6396401 86**
   1 SEQUENCE.  AYVINDSCIA CGACKPECPV NIIQGSIYAI DADSCIDCGS CASVCPVGAP NPED

11)  1FXC  FERREDOXIN [ELECTRON TRANSPORT] {SPIRULINA PLATENSIS}.....................................................1FXC
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   -+--++++ +-+-+-++-+ +++-+-+--+ +-+-++-+++ -+-++-++++ +--++-+--- +--++++++ -++++--+++ ---+++--++ +-+---
     BEND.....    S     S S  SSS  SS S    S  SS SS    S  S S SSSSSSS       SS    S    SSS S    S   SSS     SS  SS  SSS
     5-TURN...
     4-TURN...
     3-TURN...          >3 3< >33< >3 3<                >3 3<    >33<                              >33 <  >33<
     SUMMARY..    S     T T  STT  ST T    S  SS SS    S  T T SSSSSTT       SS    S    SSS S    S   STT     TT  SS  SSS
     EXPOSURE.  11*9***667 251699244* *5*6**5136 **74*584*9 7*61532601 Ø15574*4*2 69456*30*9 **82310322 Ø926765*** 39**3Ø57
   1 SEQUENCE.  ATYKVTLIDE AEGINETIDC DDDTYILDAA EEAGLDLPYS CRAGACSTCA GTITSGTIDQ SDQSFLDDDQ IEAGYVLTCV AYPTSDCTIK THQEEGLY

12)  3FXN  FLAVODOXIN (OXIDIZED) [ELECTRON TRANSPORT] {CLOSTRIDIUM MP}.....................................................3FXN
     SHEET....   AAAAA                          AAAA            AAA AAA  B   B                        AAAAAAAA
     BRIDGE2..   bbbbb                                          ccc cc                               dd  eeee
     BRIDGE1..   aaaa                           aaaa             bb bbb  F   F                        ccccc
     CHIRALITY   --+-+-+-+ ++++++++++ +++++-+-+- ----+++--+ +++++-+--- -+--++--+- --+++++++- +++++++++- ---++-+-++ --+++++++
     BEND.....         SSS SSSSSSSSSS SSSSSSS        SSS  S SSSSS S          SSS    SSSSSSSS SSS SS  S          SS S SSSSSSSS
     5-TURN...                      >5555<                                          >5555<
     4-TURN...          > >>>XXXXXXX XX<<<<                > 444<           >444<     >>>>XX <<<<                      >>>>XXXX
     3-TURN...                        >33<       >>3<<     >33<            >33<              >3><3X33<
     SUMMARY..   EEEEE SSS HHHHHHHHHH HHHHHTT    EEEEGGG  S TTTTT SEEE EEE  BTTTB    TTTHHHHH HHHTTT TT  EEEEEEEESS S SHHHHHHH
     EXPOSURE.  47002325*2 407*007101 *009*67**1 *44316*483 **16*3*000 0001276*45 12*9806610 **039*0954 9000000052* 2*0*01*707
   1 SEQUENCE.  MKIVYWSGTG NTEKMAELIA KGIIESGKDV NTINVSDVNI DELLNEDILI LGCSAMGDEV LEESEFEPFI EEISTKISGK KVALFGSYGW GDGKWMRDFE

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

```
     SHEET....          AA     AAAA
     BRIDGE2..
     BRIDGE1..          dd     eeee
     CHIRALITY  +++++-+--- +---+---+- ++++++++++ ++++++
     BEND.....  SSSSSSS     S     SSS  SSSSSSSSS SSSSSS
     5-TURN...    >5555<
     4-TURN...  XX<<<<                  >>>>XXX XXX<<<<
     3-TURN...    >>3<<               >>3<<         >>3<<
     SUMMARY..  HHHHHTT EE  S  EEEESS  GGGHHHHHH HHHHHHT
     EXPOSURE.  *6088763173 43*2261*87 09*29**018 106*01*9
  101 SEQUENCE.  ERMNGYGCVV VETPLIVQNE PDEAEQDCIE FGKKIANI

  13)  2RXN  RUBREDOXIN (OXIDIZED, FE(III)) [ELECTRON TRANSPORT] (CLOSTRIDIUM PATEURIANUM).............................2RXN
     SHEET....      AAA       AA                          B       B   AA A
     BRIDGE2..      BBB
     BRIDGE1..      AA        AA                          C       C   BB B
     CHIRALITY   -----++-+ ----++--++ +-+--++--+ ++--++---+ +-+--+++-- --
     BEND.....         SSS       SSS    SSS  SS  S SSS SS   S SS   SSS
     5-TURN...
     4-TURN...        >444<     >444<           >4 44<     >4 44<
     3-TURN...        >33<      >33< >> 3<< >33<>> 3<<>33<       >>3<<
     SUMMARY..      EEETTT    EE TTT  G GGS   TT  G GGS TT B  T TT  BGGGEE E
     EXPOSURE.  ***4407865 57040**126 *951*95331 **08**2605 *54533*82* 83**
    1 SEQUENCE.  MKKYTCTVCG YIYBPZBGBP BBGVBPGTBF KBIPBBWVCP LCGVGKBZFZ ZVZZ

  14)  1AZU  AZURIN [ELECTRON TRANSPORT, COPPER BINDING] (PSEUDOMONAS AERUGINOSA).....................................1AZU
     SHEET....   AAAA             BBB      AAAAAA A             C  BBBB                                  B     C     A AAAAA
     BRIDGE2..                            BBBBBB                  E
     BRIDGE1..   aaaa             ccc          aaa a             G  DDDD                                  E     G     B BBBBB
     CHIRALITY   -+----+-+ --+++----- ++-+-----+ --+-+--++- +---+-+--+ ++++++++++ ++--++--+- ----++++- --+--++-+- --+--+++--
     BEND.....         SSS   S SS     SSSS          SS  SSS S        S SSSSSSSSS SSS SSS SS  SS SS S        SS         SSS
     5-TURN...                                                                  >5555<    >5555<
     4-TURN...                                        >444 <       >4 44X>>>XXXX <<<X>44<<
     3-TURN...                                        >33<         >3 3X33<         >33<      >33<        >33<
     SUMMARY..   EEEE  SSS   S SSEEE  SSSSEEEEEE E  SS  TTT S B  EEEET TTTHHHHHHH HHH HHHHSS  SS TT SB     B TT  E EEEEESSS
     EXPOSURE.  *3*09074*5 8274*61*0* **5**13040 41629759*7 3000000026 9539103860 88227**441 9**296351* 0810037884 51*1836*1*
    1 SEQUENCE.  SVDIQGNDQM QFNTNAITVD KSCKQFTVNL SHPGNLPKNV MGHNWVLSTA ADMQGVVTDG MASGLDKDYL KPDDSRVIAH TKLIGSGEKD SVTFDVSKLK

     SHEET....      BBBB           BB BBBB
     BRIDGE2..      FFFF           FF FF
     BRIDGE1..      DDDD              ccc
     CHIRALITY  ++----+--+ --++++++-- -+-
     BEND.....  SS         SSS SSSS
     5-TURN...
     4-TURN...             >444<
     3-TURN...            >3><3<
     SUMMARY..  SS  EEEE   STTTTTTSEE EEEE
     EXPOSURE.  *89*140000 264128*040 5063*
  101 SEQUENCE.  EGEQYMFFCT FPGHSALMKG TLTLK

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
15)  1PCY  PLASTOCYANIN [ELECTRON TRANSPORT, COPPER BINDING] {POPLAR LEAVES: POPULUS NIGRA VARIANT ITALICA}.............1PCY
     SHEET....   AAAA          AA  BBB B    AAAAA A     C  B B                        C      AA AAA    BBB BBB          BBBBBB
     BRIDGE2..     BB                       CCCCC                                                    FFF FFF          FFFFFF
     BRIDGE1..   aaaa          BB  ddd d      aaa a     G  E E                        G     CC CCC       EE            dddd
     CHIRALITY   ---+--+-+ -+--+++++-+ --++------ +-+-++--+- --+++--++- -++++---++ --++-++--- +---++-+--- +---+++++- +++----
     BEND.....      SS SS   S  SSSS     SS         SS         SSSS SS   SSSSS  SS     SSS          S S       SSSSSS S
     5-TURN...                                                                                              >555 5<
     4-TURN...                                                          >>44<<
     3-TURN...         >33<             >33<                  >33< >33< >>3<<  >33 <   >33<                  >>3<X33 <
     SUMMARY..   EEEES TT   S EESSEEE E TT EEEEE E SS  B  E E TTSS TT   HHHHS  TT   B STT EE EEE  S EEE EEE GGGTTT T EEEEEE
     EXPOSURE.  6905002**4 6630655*27 0665890306 0324750001 15**522*93 829*1224** 871*5*4765 *14089*3*0 3050474997 40715030*
   1 SEQUENCE.  IDVLLGADDG SLAFVPSEFS ISPGEKIVFK NNAGFPHNIV FDEDSIPSGV DASKISMSEE DLLNAKGETF EVALSNKGEY SFYCSPHQGA GMVGKVTVN

16)  1PPT  AVIAN PANCREATIC POLYPEPTIDE [HORMONE] {TURKEY PANCREAS: MELEAGRIS GALLOPAVO}.............................1PPT
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   ---------+ +--+++++++ ++++++++++ +-++
     BEND.....           S SS SSSSSSS SSSSSSSSSS SSS
     5-TURN...                        >5 555<
     4-TURN...              >>>>XXXX XXXXXXXX<< <<
     3-TURN...          >3 3<>33<              > 33<
     SUMMARY..           T TS HHHHHHH HHHHHHHHHH HTT
     EXPOSURE.  *6**3**63* *27**669*4 *8*69*6932 99****
   1 SEQUENCE.  GPSQPTYPGD DAPVEDLIRF YDNLQQYLNV VTRHRY

17)  1GCN  GLUCAGON (PH 6 - PH 7 FORM) [HORMONE] {PIG PANCREAS: SUS SCROFA}.....................................1GCN
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   ++-++++++ ++++++++++ +++++++
     BEND.....    S S SSSS SSSSSSSSSS SSSSSS
     5-TURN...
     4-TURN...       >>44X <4>X4>XX>X <X<<4<
     3-TURN...       >>3<X 3>X3<<     >>3<<
     SUMMARY..    S S HHHH TTTHHHHHHH HHHTTS
     EXPOSURE.  ***4***8** 9**7*8**5* 8*9******
   1 SEQUENCE.  HSQGTFTSDY SKYLDSRRAQ DFVQWLMNT

18)  1INS  INSULIN (A AND B CHAIN) [HORMONE] {PIG: SUS SCROFA}.....................................................1INS
     SHEET....              A        B      A                      BBB
     BRIDGE2..                                                     CCC
     BRIDGE1..              A        B      A                      B
     CHIRALITY   +++++++--- --++++++++     ---++-- ++++++++++ +++-----++-
     BEND.....    SSSSSSS    SSSSSSS           S SSSSSSSSSS SSSS
     5-TURN...
     4-TURN...  >>>>X<<<<   >>44<<             > >>>XXXXX<< <<
     3-TURN...              >>><<<          >> 3<<        >>3<<
     SUMMARY..   HHHHHHHS  B HHHHGGGB       B   T HHHHHHHHHH HGGG EEE
     EXPOSURE.  6125705*78 025*709*72 *0***9*143 3*005106*4 12**008173 **
   1 SEQUENCE.  GIVEQabTSI aSLYQLENYc N1FVNQHLbG SHLVEALYLV cGERGFFYTP KA

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

```
19)  1BP2  PHOSPHOLIPASE A2 [PHOSPHATIDE ACYL-HYDROLASE] (COW PANCREAS: BOS TAURUS).....1BP2
     SHEET....                          A   A                                                BBBB   BBBB
     BRIDGE2..                          B
     BRIDGE1..                          A   A                                                CCCC   CCCC
     CHIRALITY   +++++++++ +++++--+++ ++++-+++- --++-----+ ++++++++++ +++++++-++ +++-+--+++ -+-+--+--+ -----++--+ ++++++++++
     BEND.....    SSSSSSSS SS SS  SSS SSSS SSSSS S   SS SSS SSSSSSSSSS SSSSSSS SS SSSS   SSS        SSS       SS  S SSSSSSSSSS
     5-TURN...                    >55 55<                                          >5555<
     4-TURN...  >>>>XXXXX< <<<   >>44 <<  >444<          >> >>XXXXXXXX XX<<<< >>> 4<<<  >444 <                  >> >>XXXXXXXX
     3-TURN...                 >33<>33<  >33<                                 >>3<<    >33< >33<       >33 <   >33<
     SUMMARY..   HHHHHHHHH HH TT  HHH HTTTBTTTBS S   SS SSH HHHHHHHHHH HHHHHTT HH HHHTT  TTT     EEEETT EEEE TT  H HHHHHHHHHH
     EXPOSURE.  1**0242017 1339*1*0** 6276000303 *5383*352* 00*0147247 32*90**4*4 0**9*580*9 0*066667*8 *0844**277 0546007007
   1 SEQUENCE.  ALWQFNGMIK aKIPSSEPLL DFNNYGbYcG LGGSGTPVDD LDRdcQTHDN eYKQAKKLDS fKVLVDNPYT NNYSYSaSNN EITgSSENNA fEAFIgNeDR

     SHEET....                A
     BRIDGE2..
     BRIDGE1..                B
     CHIRALITY  +++++++-- --++++--+ +
     BEND.....  SSSSSSSSS    SSSSS  S S
     5-TURN...
     4-TURN...  XXXX<<<<
     3-TURN...      >>3<<   >>3X<3<>> 3<<
     SUMMARY..  HHHHHHHTS    GGGBT  G GG
     EXPOSURE.  8004414*48 95**46*397 **6
  101 SEQUENCE.  NAAIdFSKVP YNKEHKNLDK KNb

20)  1LZM  LYSOZYME [O-GLYCOLSYL HYDROLASE] (BACTERIOPHAGE T4}.....1LZM
     SHEET....                A B AA      AAAA   AA A                      B
     BRIDGE2..                            BBB
     BRIDGE1..                A C AA      AA A   BB B                      C
     CHIRALITY   -++++++++ ----+-+--- +-+-+-+--+ ----+-+-++ ++++++++-- +-+--++--+ ++++++++++ ++++++++++ -+++++++++ --+++++++
     BEND.....    SSSSSSSS S   SS     SSS    SSS     S S SS SSSSSSSSSS SS  SS   S SSSSSSSSSS SSSSSSSSSS SSSSSSSSS S SSSSSSS
     5-TURN...                                                    >>555 <<
     4-TURN...    >>44<X44 4<              >444 <        > >4>XX>XXXX<<X X>>XXX<<<< >44>X>><<< <>>>>XXXXX
     3-TURN...    >>3<<                > 33<                    >33<     >33< >3 3X33< >33<    >33X>3< X33X>3<X33 < >33<  >>
     SUMMARY..    SHHHHSTT T  EEEEEE  TTS EEEETT EEEES S SS HHHHHHHHHT TS  TTB  H HHHHHHHHHH HHHHHHHHHT STTTHHHHHH S HHHHHHHH
     EXPOSURE.  8636900*52 424*7*25*4 **4930l012 3812*4*68* 3037603*55 6**1*0726* *309*119*4 17414*009* 3**0*504*2 1572381001
   1 SEQUENCE.  MNIFEMLRID EGLRLKIYKD TEGYYTIGIG HLLTKSPSLN AAKSELDKAI GRNCNGVITK DEAEKLFNQD VDAAVRGILR NAKLKPVYDS LDAVRRCALI

     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  +++++--+++ ++++++++++ ++-++++++ +++++-++++ ++++++++++ ++++-+-+++ +-
     BEND.....  SSSSSS SSS SSS SSSSSS SSS   SSSS SSSSSSSSSS SSSSSSSSS  SSSSSSS  S SS
     5-TURN...                      > 5555<
     4-TURN...  XXX<<<X>44 << >>>4XX< 4<<  >444<      >>>4< <X>>>XX<XX X><<<<
     3-TURN...  3<<   >>3< X33<   >>3 X<3< >33<    >33<        >3><3<   >33<
     SUMMARY..  HHHHHH HHH HTT HHHHHH HHTT  TTTS SSSTTSHHHH HSHHHHHHHH HHHHHSS  S SS
     EXPOSURE.  00136636*4 01639801*4 19***5*801 8503*3*17* *37*209501 500*7255*2 1***
  101 SEQUENCE.  NMVFQMGETG VAGFTNSLRM LQQKRWDEAA VNLAKSRWYN QTPNRAKRVI TTFRTGTWDA YKNL

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
21)  7LYZ  LYSOZYME (TRICLINIC CRYSTAL FORM) (O-GLYCOLSYL HYDROLASE) (HEN EGG WHITE: GALLUS GALLUS).....................7LYZ
     SHEET....   A                 B   B                A    CC        CC    CC    DD             DD
     BRIDGE2..                                                            DD
     BRIDGE1..   A                 B   B                A    CC        CC    DD    ee             ee
     CHIRALITY   ---++++++ ++++-++-+- -+--++++++ ++++---+-+ +-----+-+- ----+----+ +++-+-++-+ ++-++-+--+ ++++-+-+++ ++++++++++
     BEND.....      SSSSSS SSSSSS SSS SS  SSSSSS SSSSSSSS S S     SSS    S SSS   SS S S   SS   S SSSSSS SS SSSSSSSSS
     5-TURN...            >5555<              >5555<                    >5555<>5 555<
     4-TURN...    >>>4XXX ><<<<          >>>>XX XX<<<<                    >444< >4 44<                        >>> >XXXXXX<<<
     3-TURN...              >33<>3>X 3<<>>3<<          >3 3<                  >3 3<      >3 3< >33< >> >X<<<   >3 3<     >33
     SUMMARY..   B  HHHHHH HHHHTT TTB TTB THHHHH HHHHHTSSBT T EE  SSS    EETTTTEET TTTEE SS T T   TT EEG GGGGSSS HH HHHHHHHHHH
     EXPOSURE.  9*1473*003 108*6506*7 *654002000 0056529350 *287*7*720 0220001030 6*30746*19 65*672*470 8304669152 109005*113
   1 SEQUENCE.  KVFGRaELAA AMKRHGLDNY RGYSLGNWVb AAKFESNFNT QATNRNTDGS TDYGILQINS RWWcNDGRTP GSRNLdNIPc SALLSSDITA SVNdAKKIVS

     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   -+--+++-++ +++++-+-++ ++++++-
     BEND.....   SSS S SSSS SSSSSSSS   SSSSSS
     5-TURN...              >5555<
     4-TURN...   <      >>> >****<
     3-TURN...   < >>><<<       >33<>> 3<<>33<
     SUMMARY..   SSSGGGGSHH HHHHTTTS G GGSSTT
     EXPOSURE.   *7*1055171 2***0*97*1 *364*72*9
 101 SEQUENCE.   DGDGMNAWVA WRNRbKGTDV QAWIRGaRL

22)  1SNS  STAPHYLOCOCCAL NUCLEASE (COMPLEX) (PHOSPHORIC DIESTER (DNA) HYDROLASE) (STAPHYLOCOCCUS AUREUS).................1SNS
     SHEET....           A AAAA  A     AAAAA     AAAAAA  BB                                    AAAAA      A     AAA AAAA  AA
     BRIDGE2..                         CCCCC       ddd                                       EEEEE            F E EEEE
     BRIDGE1..           A AABB  B      B BB     CCCCC   HH                                    AAA        F     ddd   GG  GG
     CHIRALITY    +++-+---- +--+--+--+ ++-+-----+- +------+-- ---++-+--+ --++++++++ +++++++-+ -------++- +--+-+--+- -----+--++
     BEND.....    SSSS             SS S      SS       SS      SS SSSS S SSSSSSSS SSSSSSSS S        SS     SSS   S    SSS  SS
     5-TURN...
     4-TURN...                                                    >>>>XXX XXXX<<<<                                    >>>
     3-TURN...                     >3 3<             >33<       >33<  >33<       >33X33<              >33<        >33<
     SUMMARY..     SSSS   E EEEEEEE ST TEEEEE SS  EEEEEETTEE    SS STTS S TTHHHHHHH HHHHHHHT S  EEEEE SS   B TTS EEE EEEETTEEHH
     EXPOSURE.  **7***4*64 *1*39*2710 0117074*6* *5303111103 0153*****3 *6*807*036 519*219*2* *140130*29 *92**47030 002075*210
   1 SEQUENCE.  ATSTKKLHKE PATLIKAIDG DTVKLMYKGQ PMTFRLLLVD TPETKHPKKG VEKYGPEASA FTKKMVENAK KIEVEFNKGQ RTDKYGRGLA YIYADGKMVN

     SHEET....           B B
     BRIDGE2..
     BRIDGE1..           H H
     CHIRALITY   +++++-++-- ++---+++++ ++++++++++ ++++---+++
     BEND.....   SSSSSSSS         SS  S SSSSSSSSSS SSSSS  SSS
     5-TURN...     >5555<                      >5555<
     4-TURN...   >X<<<<                 >>>>XXXXXX X<<<<
     3-TURN...   >33X33<          >33<>3 3<          >33<>>><  <<
     SUMMARY..   HHHHHTTS E E    TT  T THHHHHHHHH HHHHTT GGG G
     EXPOSURE.   42008*0109 01*3**4146 34*909*125 707*9*4444 5*
 101 SEQUENCE.   EALVRQGLAK VAYVYKPNNT HEQHLRKSEA QAKKEKLNIW SE

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

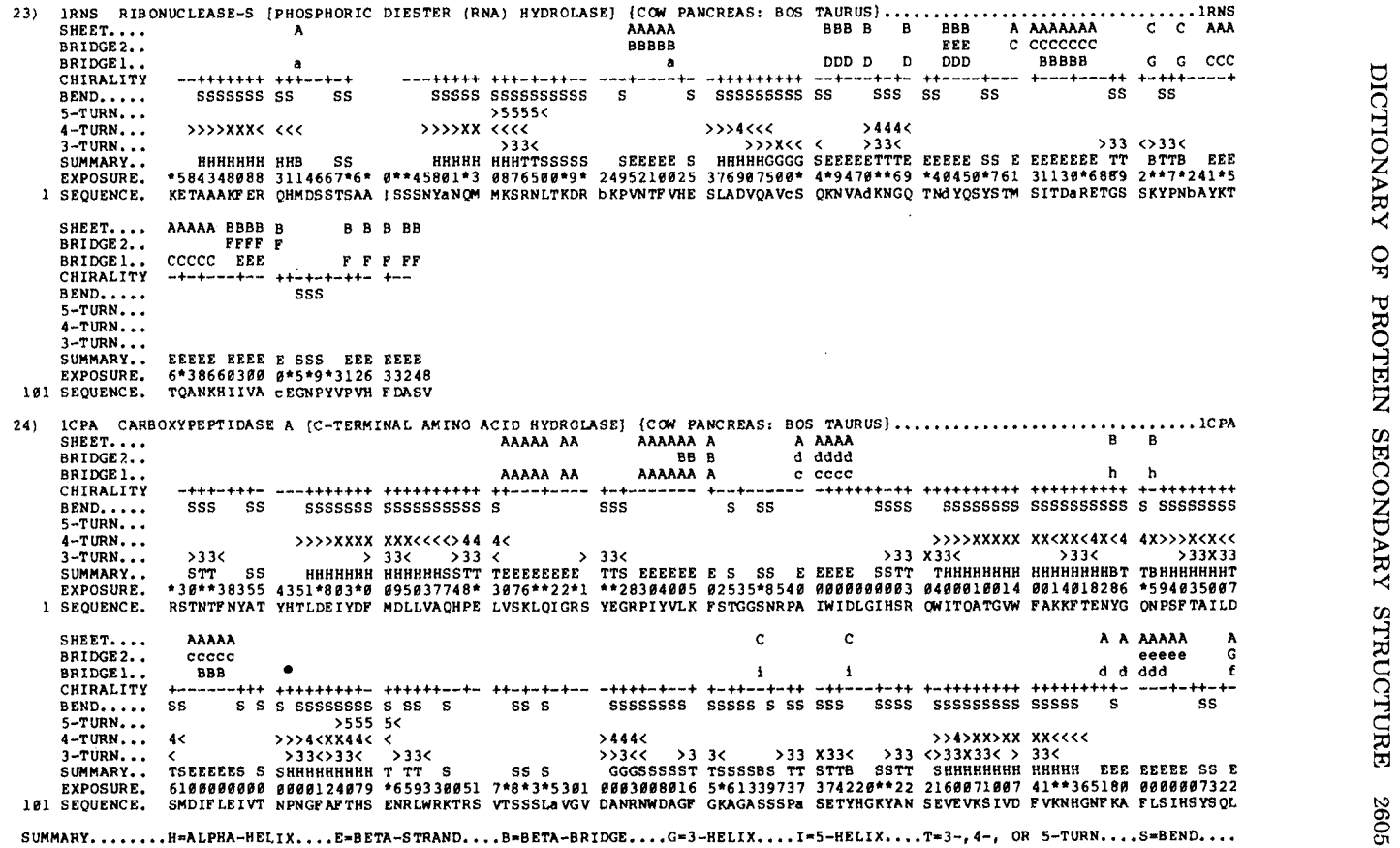


TABLE AIII (*continued*)

```
     SHEET....  AAA                                    AAA A                    AAAAA
     BRIDGE2..  GG                                                               GGG
     BRIDGE1..  fff                                    fff f                    eeeee
     CHIRALITY  +--++++-+- --+++++++++ ++++++++++ +-------+- -++++-----+ +-++++++++- +++++------ +--+-+-++- -+++++++++ ++++++++++
     BEND.....    S SS      SS SSSSS SSSSSSSSSS SSS      S  SSSSS     S SSSSSSSS S S         S SSSSSS    SSSSSSSSS SSSSSSSSSS
     5-TURN...                                                          >555 5<
     4-TURN...               >>>>XXX XXXXX<<<<                 >>44<<      >>>>XX<<<< <                        >>>>XXX XXXXXXXXXX
     3-TURN...              >33<            >>3< <             >33<           >>><< <                 >33<  >>3<<       >33<
     SUMMARY..  EEES SS     TTHHHHHH HHHHHHHHTT SSS    EEE EHHHHS     S HHHHHHHH T S EEEEE  S SSSSTT    GGGHHHHHH HHHHHHHHH
     EXPOSURE.  00102044** 91948*5069 00*502630* 98*4182*90 205857*610 0000020278 6160000000 31*5**14*1 43*4053005 0024002200
201  SEQUENCE.  LLYPYGYTTQ SIPDKTELNQ VAKSAVAALK SLYGTSYKYG SIITTIYQAS GGSIDWSYNQ GIKYSFTFEL RDTGRYGFLL PASQIIPTAQ ETWLGVLTIM

     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  ++++
     BEND.....  SSSS
     5-TURN...
     4-TURN...  X<<<<
     3-TURN...   >>3<<
     SUMMARY..  HHHHT
     EXPOSURE.  7604**042
301  SEQUENCE.  EHTVNNIGY

25)  1APR  ACID PROTEASE (HYDROLASE: PROTEINASE) (RHIZOPUS CHINENSIS).....1APR
     SHEET....          A       AB  C C      CC   BBB     B DD E              E              F           GGF       H     H
     BRIDGE2..                                    ccc
     BRIDGE1..          A       AB  E E      EE   B       D ff g              g              H           IIH       J     J
     CHIRALITY  -+------- --++----+- --+--+-+-- ----++-+-+ ----++--++ +++-+------ +++-+++--- +--+--+-+- ---+--++-+ +----+---+
     BEND.....   S        SSSS         SSS          SS         S   SS SS SS      SSS SS             SS                SSS
     5-TURN...
     4-TURN...                                                                  > 444<
     3-TURN...  >33<     > 33<                                                 > 33<
     SUMMARY..   TT     B  TTSS BB  E E SSS  EE   EEESS   B EE BS   SS SS SS   B  TTT SS       B     SS    EEB       BSSS B
     EXPOSURE.  **0409273* 5*04512090 4012*56715 0111022011 1100681*55 00*8628715 4*80*66*75 9**175584* 2615150114 1061261705
   1 SEQUENCE.  GVGTVPMTDY GNDVEYYGQV TIGTPGKSFN LNFDTGSSNL WVGSVQaQAS GaKGGRDKFN PSDGSTFKAT GYDASIGYGD GSASGVLGYD TVQVGGIDVT

     SHEET....        DDG G             BBB                                  II I     III          J       K           LL
     BRIDGE2..                           D
     BRIDGE1..        ffI I             ccc                                  KK K     KKK          L       M           OO
     CHIRALITY  --------++- ---++++-+- ----++---+- ++-+-++--- +++++++-++ -+-++--++- --++-+----+ -+----+++- +----+---- ++----+---
     BEND.....  S S     SS  SSSSSSSS  SS    S SS S SSSSS    SSSSSSSSSS  SSS        SSS        SS   SSSS  S          SSSS
     5-TURN...
     4-TURN...                                          > >4>X<4<<              >444<
     3-TURN...   >33<                                   > >3<<                   >33<          >33<
     SUMMARY..  S TT EEEEE  SSSSSSSS  SSEEE S SS S SSSSS    HHHHHHHSSS  SSS    EE E TTT EEE   SS   TTSB  S    B     SSSS  EE
     EXPOSURE.  4015121049 1244624631 0020000149 73152**170 1041068688 1**3000040 023**85130 2132*5**43 37*2591726 21*1701070
101  SEQUENCE.  GGPQIQLAQR LGGGGFPGDN DGLLGLGFDT LSITPQSSTN AFDQVSAQGK VIQPVFVVYL AASNISDGDF TMPGWIDNKY GGTLLNTNID AGEGYWALNV

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

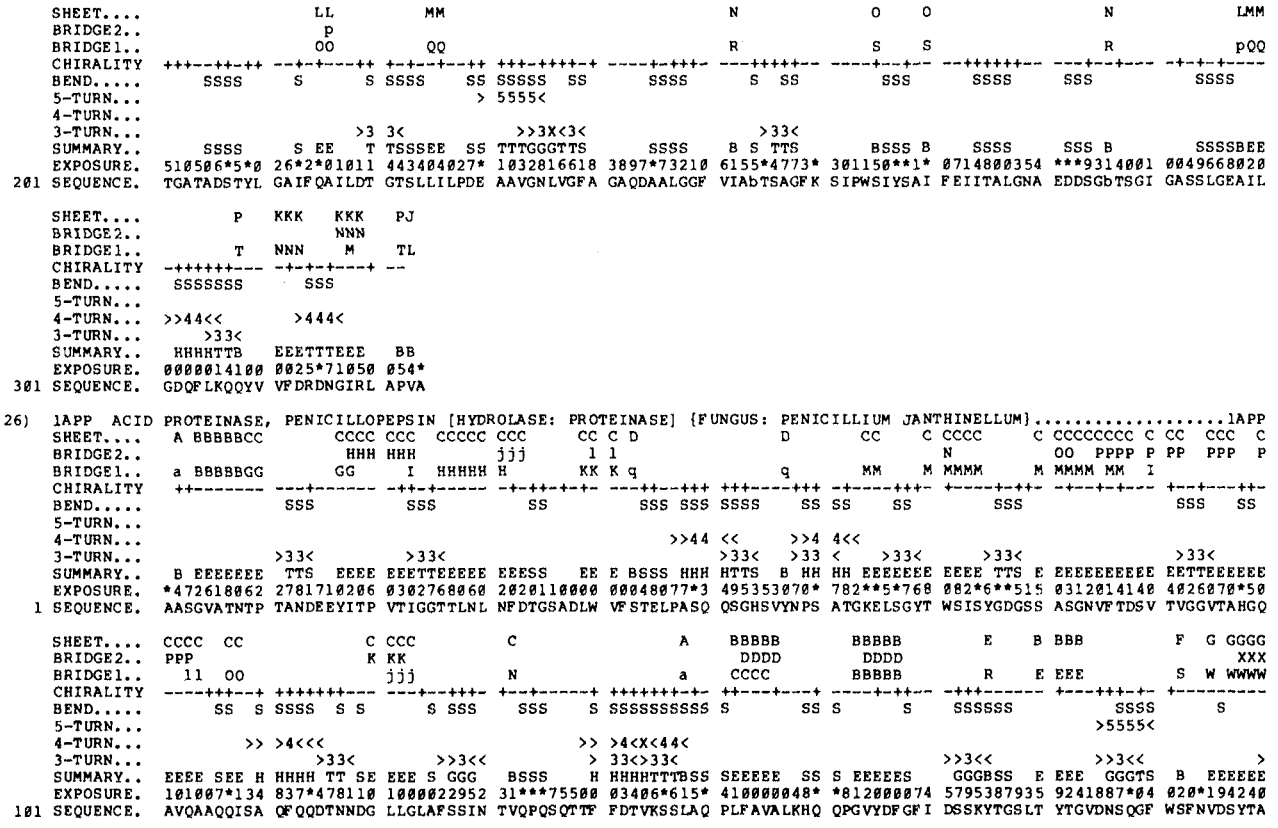


TABLE AIII (*continued*)

```
     SHEET....    GGG    F FF       FF F               HHHH      HHHH          GGGGG   GGGGG       H HHH      HHHH FFF         FFF
     BRIDGE2..           t tt        V                   AA        AA           YYYYY                BBBB                  V
     BRIDGE1..    XXX     S        uu u               ZZZZ      ZZZZ          WWWWW   YYYYY       B BBB       AA  uuu         ttt
     CHIRALITY  ++-------+- ---++-+--+ --++++++++ ++-+++---+ +-++++--++ -----+---- -+-------++ ++---++--+ -----++--- --+-+---+-+
     BEND.....  SS             SS SS     SSSSSSSS SS SS    S SSSS    SS S          SS       SS SS     SSS S    S      S SS    S
     5-TURN...                                             >5 555<                                  >555 5<
     4-TURN...                           >>>>XX<<< <       >4 44<                          >44 4<                          >>
     3-TURN...  33<          >33<             >>> <<<     >3 3<     >33 <        > 33<    >>> <<<    >33 <                    >3
     SUMMARY..  TTEEE    E EE TT SSEE E HHHHHHHH TT SSEEEEET TTTEEEE TT S    EEEEE TTEEEEE GG GGEEEEETTT TEEEESEEE   S SSEEE H
     EXPOSURE.  3*791*4260 1020324303 0598006601 952*52**4* *35150076* 3*22802021 6748060608 30331922*8 8201000131 *685802000
201  SEQUENCE.  GSQSGDGFSG IADTGTTLLL LDDSVVSQYY SQVSGAQQDS NAGGYVFDaS SSVPDFSVSI SGYTATVPGS LINYGPSGNG STaLGGIQSN SGIGFLIFGD

     SHEET....       I BBB B    BBBB  IE
     BRIDGE2..           F F    FF
     BRIDGE1..       C CCC C    EEEE  CR
     CHIRALITY  +++++-+--- -+------++- -
     BEND.....  SSSSS        SSS
     5-TURN...              >5555<
     4-TURN...  4><<4<
     3-TURN...  3X>3<<
     SUMMARY..  HHHTTB EEE ETTTTEEEE  BB
     EXPOSURE.  0001100000 33*7170000 548
301  SEQUENCE.  IFLKSQYVVF DSDGPQLGFA PQA

27)  2TLN  THERMOLYSIN [HYDROLASE: NEUTRAL METALLO-PROTEINASE] {BACILLUS THERMOPROTEOLYTICUS}.....................2TLN
     SHEET....      AAAAAA A     AAAA AAA    AA  B       BB BBB          BB AA    AA                                          B
     BRIDGE2..        bb               C    DD          ff fff
     BRIDGE1..      AAAAAA A     AAAA AAA     C  E        E  GG          GG DD    bb                                          f
     CHIRALITY   +--+-+-+-- --+-+--+--- +--+-+---+ -+++------ -----+-+-+ -----+-+-+ -++-++++++ ++++++++++ ++++++++-+- --+-+-----
     BEND.....              SSS          SSS       SSSS         S S S S      SSS    SSSSSSS SSSSSSSSSS SSSSSSSS   SSSSSS
     5-TURN...                                                                                    >5555<   >5555<
     4-TURN...                                                                        >444 <>44><>>X> X<X<<4<
     3-TURN...                                        >33<      >33<                    >>3X<3<     >33<
     SUMMARY..      EEEEEE E SSS EEEE EEE SSSEE  B  SSTT EE EEE TT S S S EE EESSS EE SGGGTTT SSTTTTTHHH HHHHTTTS   STTTTS   E
     EXPOSURE.  9*48749130 *03*55*3*0 400*4**550 106853*002 0320***8*6 6353252692 9075*92110 0000*10010 13107*72*3 80244*4220
   1 SEQUENCE.  ITGTSTVGVG RGVLGDQKNI NTTYSTYYYL QDNTRGDGIF TYDAKYRTTL PGSLWADADN QFFASYDAPA VDAHYYAGVT YDYYKNVHNR LSYDGNNAAI

     SHEET....  BBBB B         BB     B B B       C                                                          DD     C
     BRIDGE2..   hh  h               I I
     BRIDGE1..  ffff           II     h h h       J                                                          KK     J
     CHIRALITY  ---++---+- +-+---++-- ------++-- --+++-++++ ++++++++++ +-----++-+ ++++++++++ +++++++++- +--+-+--++ ++--++-+++
     BEND.....      S SSS         SSS         SSS    SSS SSSS SSSSSSSSSS SSS   SSSS SSSSSSSSSS SSSSSSSSSS  SS  S SS  SSS SS SS
     5-TURN...                                                  >5 555<                    >555 5<
     4-TURN...       >444<                              >>>>X XXXXXXXX<< <<     >>> >XXXXXXXXX XXXXXXX<<< <
     3-TURN...                                        >>3<<   >  33<>33<
     SUMMARY..  EEEEEETTT      EE SSSE EEE    SSSB   GGG HHHH HHHHHHHHHH HTT   SSHH HHHHHHHHHH HHHHHHHHHH TSS  SEES  SSB SS SS
     EXPOSURE.  *000134*36 5706373810 0003195*61 4200000010 0110010007 *4121998*6 1000000000 1000100394 *4*4333003 507277**64
101  SEQUENCE.  RSSVHYSQGY NNAFWNGSEM VYGDGDGQTF IPLSGGIDVV AHELTHAVTD YTAGLIYQNE SGAINEAISD IFGTLVEFYA NKNPDWEIGE DVYTPGISGD

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

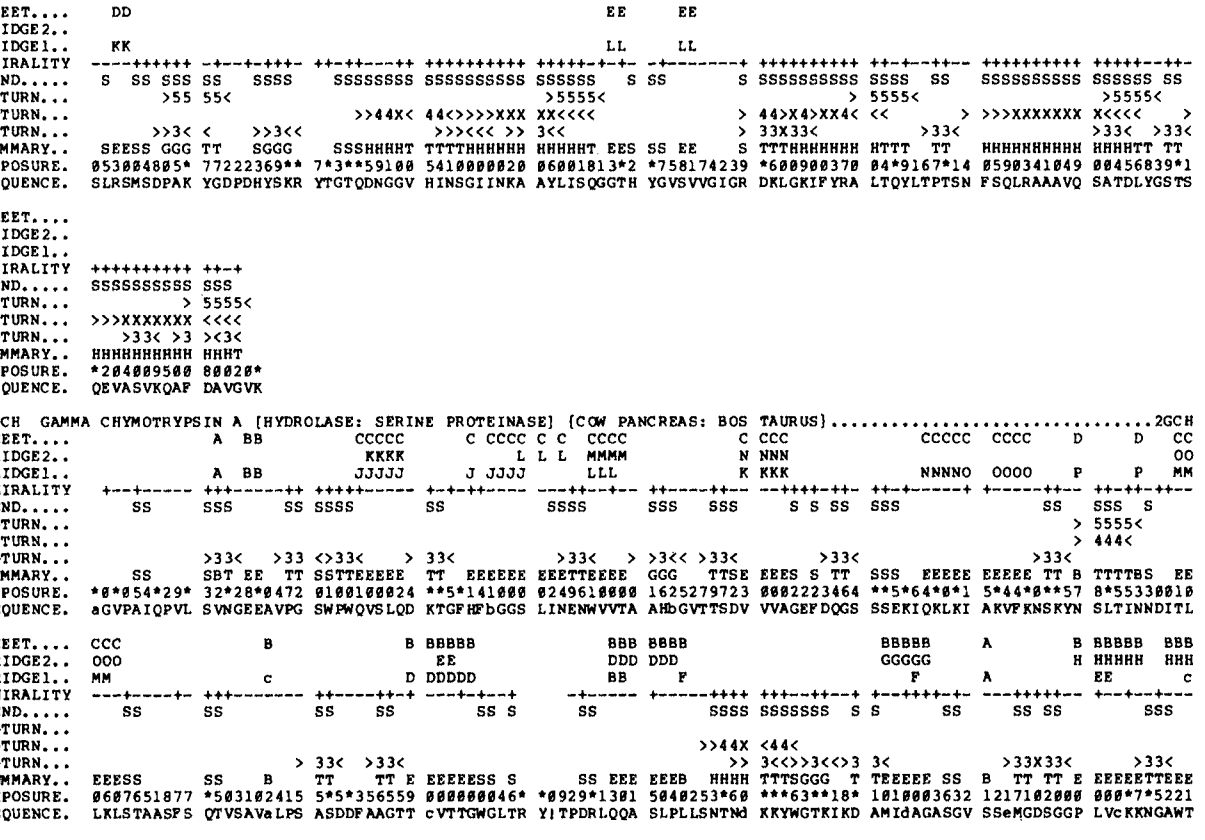


TABLE AIII (*continued*)

```
     SHEET....  B BBB B          BBBB BB
     BRIDGE2..  H HH               II II
     BRIDGE1..    III I          GGGG G
     CHIRALITY  ++---+-++-+ ---++---+- --++++++++ +++++
     BEND.....           S S  SSS       SSSSSSSS SSSSS
     5-TURN...
     4-TURN...                         >44>X>>X XXX<<<<
     3-TURN...          >3 3<>33<      >>3<<
     SUMMARY..  EEEEEEE  T T  TTSEEEE EEGGGTHHHH HHHHHH
     EXPOSURE.  000010532* 7034830100 0503601*50 **1388*
  201 SEQUENCE.  LVGIVSWGSS TeSTSTPGVY ARVTALVNWV QQTLAAN

  29)  1ALP  ALPHA LYTIC PROTEASE [HYDROLASE: SERINE PROTEINASE] (MYXOBACTER495: LYSOBACTER ENZYMOGENES).................1ALP
     SHEET....      A  BBB B    BBB    BBBB  BBB BB           BBBB  BB  BBB BBB   C  BBBBBB         AAA A  AAAA  D
     BRIDGE2..         DD D             FF     G GG           HHHH      III III      IIIIII         BBB B
     BRIDGE1..      a  CCC      CCC    EEEE  EEE E             DDD  HH  HH       j    GGG           a     BBBB  q
     CHIRALITY  -+-++---- --++-+---- -+----+-+- -+--+++--+ +----+-+-+ -----+--++ --++------ ++----+--+ --+---+++ +------+-+-
     BEND.....      SS     SSSSS          SSS        SSSS S S    SSSS          S S          SSS         SS            SS
     5-TURN...
     4-TURN...
     3-TURN...     >33<                   >33<          >3 3<   >33<                                  >33<          >33<
     SUMMARY..      BT EEE ESSSSEEE    EEEETTEEE EE  SSSS T T EEEETTEE EEEEEEEE S B  EEEEEE  SSS    EEE ETTEEEE  B      TT
     EXPOSURE.  6*02017702 18*7880000 0105*89640 0000042096 *050618868 0061439624 3404020326 993*443*03 58*7*140*1 3973675370
   1 SEQUENCE.  ANIVGGIEYS INNASLaSVG FSVTRGATKG FVTAGHaGTV NATARIGGAV VGTFAARVFP GNDRAWVSLT SAQTLLPRVA NGSSFVTVRG STEAAVGAAV

     SHEET....  CCCC   CCC C  C CC C  CC    CC   CCCC    E       CCC     D  CCC C C      E          CC CCC           BB
     BRIDGE2..  LL                    MM   MMMM            OO        PPP P P                      PP PPP
     BRIDGE1..  KKKK   KKK K  M MM M  MM    j    NNN    R        LL    q  OO          R            NN N             FF
     CHIRALITY  --+-++---+- +--++-++-- -+-+------+ --+-+----+ ++++--+--+ -+-+---+-+ ----+-+--++ +--+++-+-- ---++++++ -+----
     BEND.....      SSS S        S       SSS          S   S S SS     S S              SSSSSS SS SSS        SSSSSSS S
     5-TURN...                                                                                              >55 55<
     4-TURN...     >444<                                                                  >444<       >>4>X<4 <<
     3-TURN...                                        >3 3X33<   >3 3<           >33<      >>3<<       >33<
     SUMMARY..  EEEETTTEEE E  EEEEEEE EE SSSEEEE EEEE S  BT T TT EEE T T B  EEEEE E   TTSBSS SS GGG  EE EEESHHHHHH HT EE
     EXPOSURE.  0001*55455 6240537*47 1*6**360*2 0040200*10* 102000003* 7060100001 3367**0207 6383**5401 1080*302** 24172395
  101 SEQUENCE.  bRSGRTTGYQ bGTITAKNVT ANYAEGAVRG LTQGNAcMGR GDSGGSWITS AGQAQGVMSG GNVQSNGNNc GIPASQRSSL FERLQPILSQ YGLSLVTG

  30)  1PTN  BETA-TRYPSIN (NATIVE AT PH 8) [HYDROLASE: SERINE PROTEINASE] (COW PANCREAS: BOS TAURUS).....................1PTN
     SHEET....    A  BB          CCCCC    C CCCCC  C  CCCC          CCCC                CCCCC CC CC              CCCCC
     BRIDGE2..                     KKK         LL  L  MMMM          NNNN                                         OOOOO
     BRIDGE1..    A  BB          JJJJJ    J JJJJ      LLL          KKK                 NNNNO OO OO              MMMM
     CHIRALITY    ++-------+ +-++----+- +-++-+--+- -++---+--++ ++-++----+ +++-++-+-+ +--+---+-- ---++--++- +++++----- +----+-+++
     BEND.....     S     SS SSSS     S SS         SSS     SS S  SS      S SSSSS  S        S     SS  SSS   S         SS     SSS
     5-TURN...
     4-TURN...                                                                                  >444 <
     3-TURN...          >33 <>33<                  >33<   >>3 <<           >33<                 >33<     >33<
     SUMMARY..   BS EE  TT SSTTEEEEES SSEEEEEEEE ETTEEEE GG G  SS EEEE S SSTTS  S   EEEEEEEE EE TT  TTT  TT  EEEEE SS     SSS
     EXPOSURE.  0153*82364 9132100023 7*43100002 5471000007 0*5*615050 11233*7**5 5338242764 452*7678*7 3420000040 *851887*70
   1 SEQUENCE.  IVGGYTaGAN TVPYQVSLNS GYHFbGGSLI NSQWVVSAAH bYKSGIQVRL GEDNINVVEG NEQFISASKS IVHPSYNSNT LNNDIMLIKL KSAASLNSRV

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

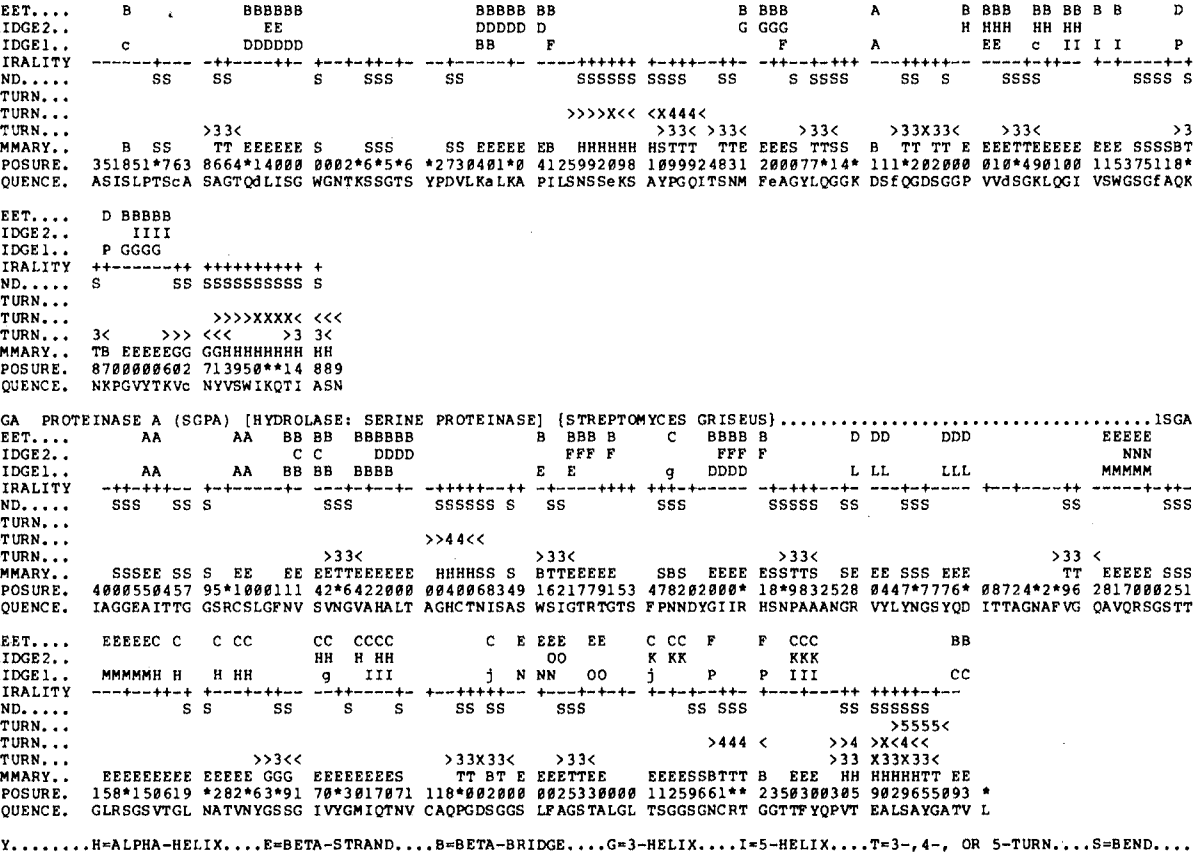


TABLE AIII (*continued*)

```
32)  1SBT  SUBTILISIN BPN' [HYDROLASE: SERINE PROTEINASE] {PROBABLY BACILLUS AMYLOLIQUEFACIENS}.....................1SBT
     SHEET....                              AAAA AA              AAAA                                        AA AAAA BB
     BRIDGE2..                               bbb b                                                           cccc
     BRIDGE1..                               aaaa aa              cccc                                       aa aaaa JJ
     CHIRALITY  ----+++++ -++++++++-+ +-+++-+--- +-----++-+ +----+---+ -++-+-++-+ +-+++++++ +-+--+-+-- ---++++--- -+-++--+--
     BEND.....      SSSSS S SSSSSSSS  S SS       SS   SS S S          SS   SS   SSSSSSSSSS SS   SSSSS    SS S           SS
     5-TURN...        >55 55<  >5555 <
     4-TURN...      >>>><< <X>>4<<<                                                >>>4<XX> 4<<<                    >444
     3-TURN...      >33X33 <     >33<   >33<          >33X3 3<         >33<           >>3< <          >33<          >33<
     SUMMARY..       HHHHH HTHHHHHTTT  S TT EEEE EES    TT T T    EEEE   TT   SS   SSSHHHHHHH HHH  SSSSS    SSTT EE EEEE EETTT
     EXPOSURE.  *779380074 0604203*85 8207*16000 0000036907 02*3522334 1**5*51992 9535000000 1101*9961 1200023220 40000409*4
   1 SEQUENCE.  AQSVPYGVSQ IKAPALHSQG YTGSNVKVAV IDSGIDSSHP DLKVAGGASM VPSETPNFQD DNSHGTHVAG TVAALNNSIG VLGVAPSSAL YAVKVLGDAG

     SHEET....  BB                  AAAA   C                    AAA AA              C        AAAAAA      A            AAA
     BRIDGE2..                       dddd                          e ee                        ffff                     h
     BRIDGE1..  JJ                   bbbb   K                     ddd d               K       eee  G      G            fff
     CHIRALITY  -+-++++++ ++++++---+ --+-++--+- --++++++++ ++++-+---+ ----+--+-- +-++---++++ -++-----+- -+-+---++- -+-+++--+-
     BEND.....     SSSSSSS SSSSSSS  S S       S    SSSSSSSS SSSSSS        S   SS SS    SSSS SSSS        SS    SSS    SS
     5-TURN...                >5555<                        >5555<
     4-TURN...  < >>>>XXXX XXX<<<<                 >>>>XXXXX XX<<<<                  >444<
     3-TURN...    >33<     >>3<X33<                        > >3<<              >3 3<     >33 X33<       >33<  >33<   >33<
     SUMMARY..  EE HHHHHHH HHHHHHTT S EEEE   BS    HHHHHHHH HHHHHT EEE EE  S   ST TS    BTTT STTSEEEEEE  TT  B TTS    TT  EEE
     EXPOSURE.  46*6551390 0660477*16 00000351*8 2594085006 *039*31000 01017*18*6 **32116005 2*30100000 5785*51*60 0009603000
 101 SEQUENCE.  SGQYSWIING IEWAIANNMD VINMSLGGPS GSAALKAAVD KAVASGVVVV AAAGNEGSTG SSSTVGYPGK YPSVIAVGAV DSSNQRASFS SVGPELDVMA

     SHEET....  A   DDDDD    DDDDD                                      A           A
     BRIDGE2..                                                                       I
     BRIDGE1..  f   LLLLL    LLLLL                                      I           h
     CHIRALITY  ++-+-+---+ -+-+---+-+ ++++++++++ ++++++++++ --++++++++ +++--+---+ ++---+---+ +++
     BEND.....    SS     S SS       S SSSSSSSSSS SSSSSSSSSS S SSSSSSSS SSSS   S S SSSSS    S SSS
     5-TURN...
     4-TURN...          >4 44<     >> 4>XX>XXXXX X<XX<4<<    >>>><XX<4 <<      >> 44<<    >> >4<<<
     3-TURN...             >3 3<>33X33<  >33X33<>33 <  >33X33X >3<<    >3 3<>33<
     SUMMARY..  E SSEEEEEET TTEEEEE  H HHHHHHHHHH HHHHHHHSTT S HHHHHHHH HTTSB  S H HHHTT B   H HHHH
     EXPOSURE.  0036180033 5**616*900 1000001000 000023439* 4828905531 2*437*43*9 5941*01030 9400*
 201 SEQUENCE.  PGVSIQSTLP GNKYGAYNGT SMASPHVAGA AALILSKHPN WTNTQVRSSL QNTTTKLGDS FYYGKGLINV QAAAQ

33)  1EST  TOSYL-ELASTASE  [HYDROLASE: SERINE PROTEINASE] {PIG PANCREAS: SUS SCROFA}.....................................1EST
     SHEET....   A  BB         CCCCCC CC  CCCC C CCCC C  CC CC           CCCCC   D        C CCCC CCCC                  CCCCC
     BRIDGE2..                   JJJJJ              KK K  LL LL           MMMM                                          NNNNN
     BRIDGE1..   A  BB          IIIIII II  IIII I III     KK K          JJJJJ    O        M MMMN NNNN                  LLLL
     CHIRALITY  ++------++ ++++------ --++---+-- +--++-++-- +--++++-+ ------++++ -++----++-- +--++------ +--++-++++- ++++------+
     BEND.....    S     SS S  S        SSS          SSS      SSSS S         S S  SSS  S                  SS  SS SSS           S
     5-TURN...
     4-TURN...         >44 4<
     3-TURN...  >33<   >33 <>33<       >33<          >33<     >>3<<             >33<                > 33<>33X>3< <
     SUMMARY..   BT EE  TT T TTEEEEEE EETTEEEEEE EEEEEETTEE EE SGGG S    EEEEES S BTTS  S  E EEEEEEEEE  TT  TT GGG     EEEEES
     EXPOSURE.  01438*1**9 4331100017 *66*763241 00000*9310 000008125** 8915000121 13***26335 *233*97362 **386*7423 0800000706
   1 SEQUENCE.  VVGGTEAQRN SWPSQISLQY RSGSSWAHTa GGTLIRQNWV MTAAHaVDRE LTFRVVVGEH NLNQNNGTEQ YVGVQKIVVH PYWNTDDVAA GYDIALLRLA

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

```
    SHEET....       E  E                  BBBB B  F       F  D BBBB   BB                   BBBB       A         BBBBBB  BB
    BRIDGE2..                              DDD D                DDDD                     FFFF                 GGGGGG
    BRIDGE1..       P  P                  CCC     Q       Q  O BB     EE                   EE         A          CCC    GG
    CHIRALITY  ----+-+++- -------++-- --++-+---+ -+--+-++-- --+-----+- ----++++-- ++---++--+ +--+-++-++ +---+++++- ----+--+-+
    BEND.....  S    SSS         SS     SS         S  SSSS     SS           SSSSSS SSS  SSS S S      SSS S   SS SS        SS
    5-TURN...                                                              > 5555<
    4-TURN...                                                           >>44<<
    3-TURN...       >33<       >33<   >33<           >33<                  >33X 33<>>3<<>3 3<           >33X33<       >33<
    SUMMARY..  S    BTTB        TT     TT  EEEE ES BSSTT   B SB EEEE   EE HHHHTS TTTTGGGS T TEEEE  SSS SB   TT TT  EEEEEETTEE
    EXPOSURE.  *709668909 61651*654* 476*250600 00026*5*4* 71*2133080 2125*76045 *7632*5079 3220021729 9001*10300 001064*7**
  101 SEQUENCE.  QSVTLNSYVQ LGVLPRAGTI LANNSPbYIT GWGLTRTNGQ LAQTLQQAYL PTVDYAIcSS SSYWGSTVKN SMVcAGGDGV RSGdQGDSGG PLHbLVNGQY

    SHEET....  BB BBB B G   G  H  H  BBBBB
    BRIDGE2..     HHH H              HHHH
    BRIDGE1..  GG GG    R   R  S  S  FFFF
    CHIRALITY  -++--+---- +--++-++-- -----+++++ ++++++++
    BEND.....             SS SS SS       SSSSS SSSSSSSS
    5-TURN...
    4-TURN...                           >> >>XXX<<<<
    3-TURN...          > 33<  >33<      >>3<<       >>3<<
    SUMMARY..  EEEEEEEE B TTBSSBTTB  EEEEEGGGSH HHHHHHHHT
    EXPOSURE.  1010010321 **10548*50 0000300512 8407*12899
  201 SEQUENCE.  AVHGVTSFVS RLGdNVTRKP TVFTRVSAYI SWINNVIASN

 34)  2ACT  ACTINIDIN [HYDROLASE: SULFHYDRYL PROTEINASE] (KIWIFRUIT: ACTINIDIA CHINENSIS).............................2ACT
    SHEET....      AA           B                                C          D   D                    E C
    BRIDGE2..
    BRIDGE1..      AA           I                                J          K   K                    L J
    CHIRALITY   -+++-+++- ++-------- +-+++++++ ++++++++++ +-+-+---++ +++++++--+ ++-++-+--+ +++++++++- -----+++-- +-+-+----+
    BEND.....    S   SSSS S          SS   SSSSS SSSSSSSSSS SSS      S SSSSSS   S S  SSS   S SSSSSSSSSS S    SSSS    SS     S
    5-TURN...       >555 5<                          >55 55<         >5555<                 >555 5<
    4-TURN...                        >>>>XXX XXXXXXXXX< <<<     >> >>X<<<<           >> >>XXXXX<<< <   >444<          >>
    3-TURN...       >>><< <       >3 3< >33<                      >3 3<>>3<< >> 3<<
    SUMMARY..    S EEGGGG T      B T TS  HHHHHH HHHHHHHHHH HHS    B H HHHHHH  BT TB GGG   H HHHHHHHHHH T B BTTTS    SS     H
    EXPOSURE.  *6**224**5 4022*23515 *130200100 0100001167 6746754000 0000221*6* *0*03*23*1 33006204*2 4000239617 573*73*489
      1 SEQUENCE.  LPSYVDWRSA GAVVDIKSQG EaGGCWAFSA IATVEGINKI TSGSLISLSE QELIDbGRTQ NTRGaDGGYI TDGFQFIIND GGINTEENYP YTAQDGDbDV

    SHEET....        E    AAAA                 AAAAA              AA          AAAAA AAA AA  AAAAAA A  B     A   AAAAA
    BRIDGE2..                                   D                    FF FFF FF   GGGG G             GGGGG
    BRIDGE1..        L    BBBB                 CCCCC              ee          CCCCC AA      FFFFFF F  I     H   H ee
    CHIRALITY  ++++-++--- --------++- +++++++++- ---+-+-+++ -++++++-+- -+--+--+-+- +-+-+--+-+- ---+-+----- --+--++---- -+-----++-
    BEND.....  SSSS              SS  SSSSSSSSSS S          SSSSSS  SS        SS               SSS          SS SSSSS SS
    5-TURN...
    4-TURN...  >4<<<              > >>>XXXX<<< <          >>>4<<<                                          >44 4<
    3-TURN...                    >33<       >33<             >33<                          >33<          >33< > 33<
    SUMMARY..  HHHH    B     EEEE  TT  HHHHHHHHHH S EEEEE    SHHHHH  SS EE     SS   EEEEEEEEE EETTEEEEEE E SB TTSTB TTEEEEE
    EXPOSURE.  744*6**061 8559*18*67 1*2014103* 2000010215 3*409*2793 4292*43984 9110100000 7*87652000 111766*226 *020404187
  101 SEQUENCE.  ALQDQKYVTI DTYENVPYNN EWALQTAVTY QPVSVALDAA GDAFKQYASG IFTGPcGTAV DHAIVIVGYG TEGGVDYWIV KNSWDTTWGE EGYMRILRNV

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
     SHEET....              AAAA
     BRIDGE2..              D
     BRIDGE1..              BBBB
     CHIRALITY  +-----++++ ++---+
     BEND.....   SS SSSSSS S
     5-TURN...
     4-TURN...
     3-TURN...  >33X>3X<3<
     SUMMARY..   TT GGGTTS S EEEE
     EXPOSURE.  5521100004 8005034*
 201 SEQUENCE.  GGAGTcGIAT MPSYPVKY

 35)  8PAP  PAPAIN    [HYDROLASE: SULFHYDRYL PROTEINASE] {PAPAYA FRUIT LATEX: CARICA PAPAYA}.............................8PAP
     SHEET....      AA                                                B                                C  B
     BRIDGE2..
     BRIDGE1..      AA                                                H                                I  H
     CHIRALITY   -+-+-+++- ++-------- +-++++++++ ++++++++++ +-+------++ +++++++++- -++++--+++ ++++++---- --+++--+-+ -+--++++-
     BEND.....    S   SSSS S          SS  SSSSSS SSSSSSSSSS SS       S SSSSSS SSS  SSS   SSS SSSSSSSS     SSSS   S S      SSSS
     5-TURN...       >555 5<                                          >555 5<
     4-TURN...                     >>>>XXX XXXXXXXXX< <<<     >> >>X<<<<          >>>> XXXX<<<<                    >>44<
     3-TURN...        >33 <          >33<         >33<                          >33< >33<                            >33
     SUMMARY..    S EESTTT T         SS  HHHHHH HHHHHHHHHH HH     B H HHHHHH TTT TSTT   HHH HHHHHHHS B  BSSSS   S S      HHHH
     EXPOSURE.  *7**134*** 3022932714 *030210100 0100000155 *85*636000 0000031**0 *04*3281*0 005202**00 194*606491 86**150**9
   1 SEQUENCE.  IPEYVDWRQK GAVTPVKNQG SaGSCWAFSA VVTIEGIIKI RTGNLNQYSE QELLDbDRRS YGaNGGYPWS ALQLVAQYGI HYRNTYPYEG VQRYbRSREK

     SHEET....      C    A AAA                A AAAA              AA          AA AAA AA   A AAAA            AAAAA
     BRIDGE2..                                  D                                FF  F   G GGGG            GGGGG
     BRIDGE1..      I    B BBB                C CCCC              ee          CC CCC AA     F FF              ee
     CHIRALITY  +--++----- ----++-+++ ++++++---- +-+-+++-++ ++++-+--+- -+--+-++-+ --+-----++- -----+--++ ----+-----  ++-+-+--+-
     BEND.....   S S   S       SS SSS SSSSSSSS         S SS SSS SSS     S  S              SS       SS SS SSSSS        SS SS SS
     5-TURN...
     4-TURN...  <                >>>> XXXX<<<<           > 444<                                          >444<
     3-TURN...  <                  >33<                  > 33<                                                    >>3
     SUMMARY..   S SB  S E EEE SS HHH HHHHHHHS E EEEE  S SS TTT SSSEE   S  S   EE EEEEEE SSE EEEE SS SS STTTSEEEEE   SS SS GG
     EXPOSURE.  3*951*0925 **1*5*7552 12*1028210 0020207*7 0**2*3372* 2*42**5921 0000010565 0102016398 216*020807 45*777*010
 101 SEQUENCE.  GPYAAKTDGV RQVQPYNQGA LLYSIANQPV SVVLQAAGKD FQLYRGGIFV GPcGNKVDHA VAAVGYGPNY ILIKNSWGTG WGENGYIRIK RGTGNSYGVc

     SHEET....       AAAA
     BRIDGE2..        D
     BRIDGE1..       BBBB
     CHIRALITY  +++-+-+---
     BEND.....  S SS
     5-TURN...
     4-TURN...
     3-TURN...  X<3<
     SUMMARY..  GTTS EEEE
     EXPOSURE.  0013412403 5*
 201 SEQUENCE.  GLYTSSFYPV KN

 SUMMARY.........H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

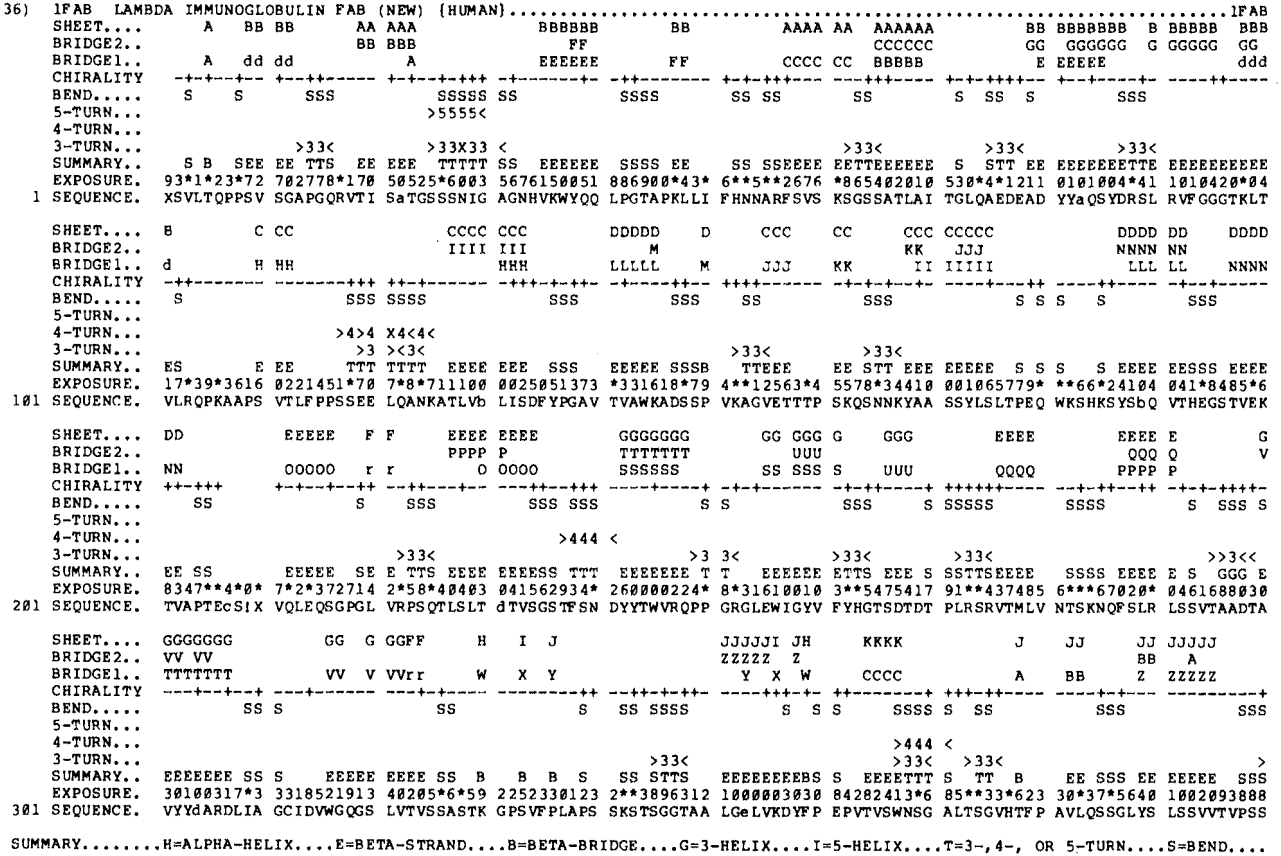


TABLE AIII (*continued*)

```
     SHEET....          KKK KKK     KKK KKK
     BRIDGE2..          DDD DDD
     BRIDGE1..           CC CC      DDD DDD
     CHIRALITY  ++++------ -+-++-+--- ---+---
     BEND.....    SSS          SSS           S
     5-TURN...                >5555<
     4-TURN...
     3-TURN...  33<            >>3<<
     SUMMARY..  TTSSS  EEE EEEGGGTEEE EEE   S
     EXPOSURE.  *39**82403 060*23*3*4 *391*688*
 401 SEQUENCE.  SLGTQTYIeN VNHKPSNTKV DKKVEPKSc

 37) 1REI  BENCE-JONES IMMUNOGLOBULIN VARIABLE PORTION (REI) (HUMAN).....1REI
     SHEET....      AAAA  B BBB      AA AAAAA        BBBBBB       BB BB    BB         AAAAAA  A AAAAA          BBBBBB        B
     BRIDGE2..                     BB BBBB            FFF           GG                      C CCCCC          HH  I
     BRIDGE1..      AAAA  d ddd      AAAA          EEEEEE        FF F     GG         CCCCCC  B BBBBB          EEEEEE        I
     CHIRALITY   -----+-++ -+--++---- -----+-+-- ++-------- +----++-+- ++----+--+ ++--+----++ -----+---+ +++------- ++-+------
     BEND.....        S S      SS          SS    S        S SS      SS SS   SS  S S     SSS       SS  S S  S        SSSS
     5-TURN...
     4-TURN...                                                        >4 44<
     3-TURN...                >33<            >3 3<      >3 3<              >33<>3 3<    >33<         >> 3<<
     SUMMARY..     EEEE SE EEE TT  EE EEEEESS  T T EEEEEE T TS  EEEEET TTEE TT  T TEEEEEETTE EEEEESS  G GG SEEEEEE  SSSS  B
     EXPOSURE.  *0*1606467 3814857*15 040918*908 870000325* 8*31*00054 0786388l5* 5061739269 04030730b8 *061301010 474834320*
   1 SEQUENCE.  DIQMTQSPSS LSASVGDRVT ITaQASQDII KYLNWYQQTP GKAPKLLIYE ASNLQAGVPS RFSGSGSGTD YTFTISSLQP EDIATYYaQQ YQSLPYTFGQ

     SHEET....   BBBBB
     BRIDGE2..   HH
     BRIDGE1..    dddd
     CHIRALITY  -----
     BEND.....
     5-TURN...
     4-TURN...
     3-TURN...
     SUMMARY..   EEEEE
     EXPOSURE.  00900*5*
 101 SEQUENCE.  GTKLQIT

 38) 3PGM  PHOSPHOGLYCERATE MUTASE (DE-PHOSPHO) [ISOMERASE] (YEAST: SACCHAROMYCES CEREVISIAE].....3PGM
     SHEET....      AA B                                                AAAAA                        AA A
     BRIDGE2..      BB                                                    ddd
     BRIDGE1..      a  f                                                ccc                        dd d
     CHIRALITY    +--+----- -++++-+++ -+-+----++ ++++++++++ ++++-+---+ -----+-+++ ++++++++++ -++++----+ -+++------ +++++--++++
     BEND.....              SSSSSSS    SS     SS SSSSSSSSSS SSSSSS   S      S SSS SSSSSSSSSS SS   S      SSS       S  S  SSSS
     5-TURN...                          >  5555<    > >555<<                      >55 55<                              >
     4-TURN...                >>44<<          >>> >XXXXXXXXX XX<<<<          >>>> XXXXXXXX<< <<                        >>>
     3-TURN...                  >33<   >33<
     SUMMARY..      EE B      SSHHHHS    TT     HH HHHHHHHHHH HHHHHS   S EEEEES HHH HHHHHHHHHH HT   S  EE ESSS       S  S  SSHH
     EXPOSURE.  5400014104 07525*8552 1134391125* 0**601900* 118**36906 1110062521 35008212** 1*5**4957* 1371140041 8321845391
   1 SEQUENCE.  PKLVLVRHGQ SEWNEKNLFT GWVDVKLSAK GQQEAARAGE LLKEKGVNVL VDYTSKLSRA IQTANIALEK ADRLWIPVNR SWRLNERHYG DLQGKDKAQT

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

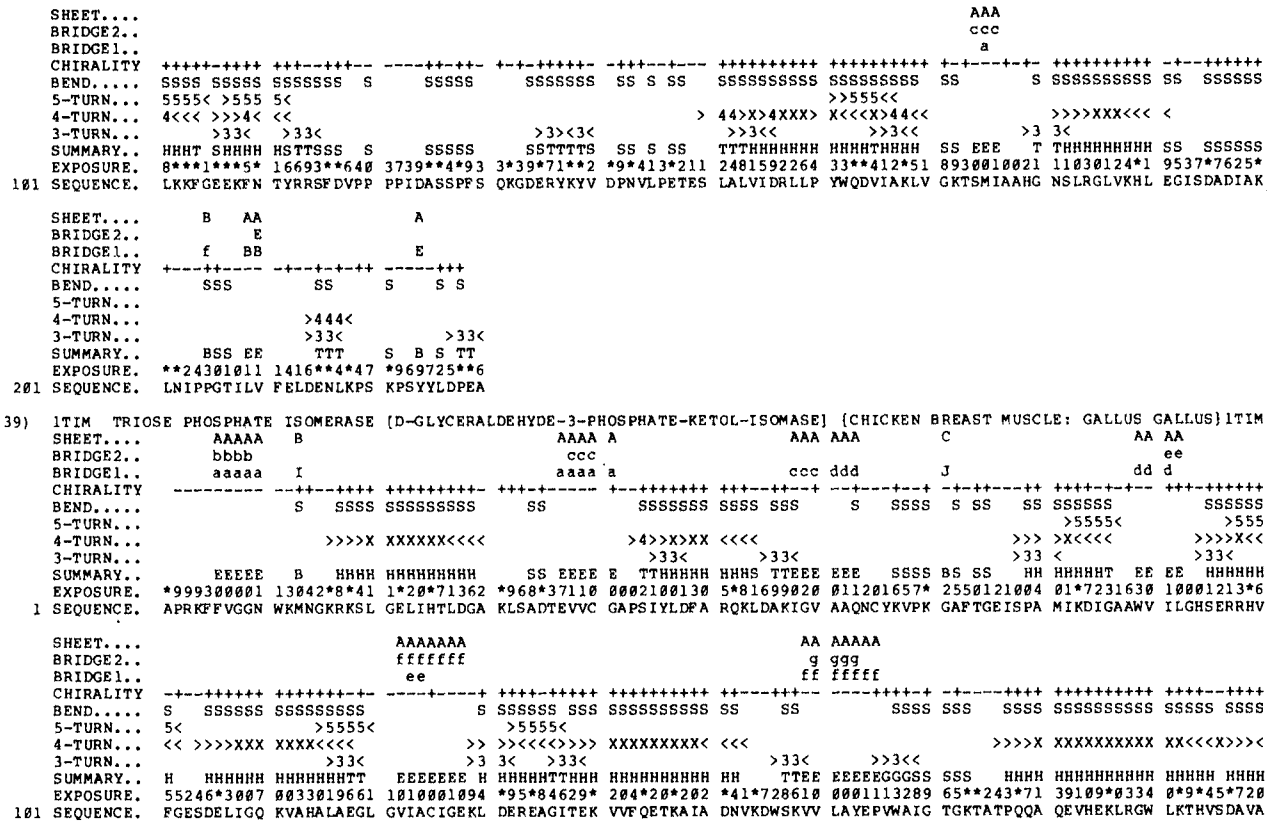


TABLE AIII (*continued*)

```
     SHEET....     AAAA                    AAAA
     BRIDGE2..       h                       h
     BRIDGE1..     gggg                    bbbb
     CHIRALITY  ++-----+-- +-+++++++ +-+++----+ +++++-++++ +++++
     BEND.....  SSS     S    SSSSSSSS SSSS       SSSSSSSSSS SSSS
     5-TURN...
     4-TURN...  <<<         >4>>X><<< <         >444< >>>> <<<<
     3-TURN...                    >3 3X33<      >>><<X33<   >33<
     SUMMARY..  HHSEEEE S    TTHHHHHH TSTT  EEEE SGGGGSTHHH HHHT
     EXPOSURE.  8802001128 045451*701 62*5100020 26207*7502 *0161**
201  SEQUENCE.  VQSRIIYGGS VTGGNCKELA SQHDVDGFLV GGASLKPEFV DIINAKH

40)  3CNA  CONCANAVALIN A [LECTIN, AGGLUTININ] (JACK BEAN: CANAVALIA ENSIFORMIS).....3CNA
     SHEET....     AAAAAAA                AAAAAA        AAA        BBBB BBBBB    B BBBBBB       BBBBBB          CCA AAAAAA
     BRIDGE2..      BBBBBB                CCC                        HH HHH      I IIIII                          EEEEE
     BRIDGE1..     AAAAAA                 BBBBBB        CCC        GGGG GGGGG      HHHHH       IIIIII          OOD DD
     CHIRALITY   -+-+-----+ +-++++-+-- +-+-+-----+ -+--+----- ---++----- -++--++-+- ++--+--++- ---+---+-- ++++--++-- +-+---+++-
     BEND.....             S   SSS   SS       SS SSS SS        SS           SSS           SSS             SSSS SS          SSS
     5-TURN...               >5555<
     4-TURN...                                                          >444<                      >444<
     3-TURN...                  >33<                        >33<                 >33<
     SUMMARY..     EEEEEEE  S  TTTT   SS EEEEEES SSS SSEEE     TT EEEE EEEEETTT E EEEEEE TTS   EEEEEE    TTTS SEEE EEEEEE SSS
     EXPOSURE.  *541001000 05419*2513 984010002* 21*29*62*1 *5773*8050 807174**9* 12261518*4 *729138*27 2**4169*04 10000002*8
   1 SEQUENCE.  ADTIVAVELD TYPNTDIGDP SYPHIGIDIK SVRSKKTAKW NMQDGKVGTA HIIYNSVDKR LSAVVSYPNA DATSVSYDVD LNDVLPEWVR VGLSASTGLY

     SHEET....      BB  BB BBBBB         BBBBBBB           A AAAA  BB      BB              A AAAAA   CC          BB BBBBBBBB
     BRIDGE2..      K   MM MMMMM                                            NN                FFFFF                 MM MMMMM  K
     BRIDGE1..      JJ  LL LLLLL         LLLLLLL           F FFFF  NN      JJ              E EEEE    OO          G GGGGGGGG
     CHIRALITY  +--+--+-++ ++--+++-+- +--+-++--+ +---+--++- -+--++-++- -+---+-+-- +-+---++-- +-+--+---+ ++-+++++++ --++----++
     BEND.....        S           SSS            S   S  SS     SS    S SSS   S    SSS   SS        SS       SS S
     5-TURN...
     4-TURN...                  >444 <
     3-TURN...                  >33<                                                                    >33<
     SUMMARY..      EES EE EEEEE  TTT    EEEEEEE S   S  SSE EEEES EE S SSSEE S    SSS    SS E EEEEESS EE    TT S EE EEEEEEEE
     EXPOSURE.  8010005313 12190722*6 **6975*3*4 8*06**3*61 4774418147 *120300*56 **47251802 01003*4623 564*647312 3*0311060*
101  SEQUENCE.  KETNTILSWS FTSKLKSNST HQTDALHFMF NQFSKDQKDL ILQGDATTGT DGNLELTRVS SNGSPEGSSV GRALFYAPVH IWESSAATVS FEATFAFLIK

     SHEET....           A AAAAA
     BRIDGE2..             DDD
     BRIDGE1..           A AAAAA
     CHIRALITY  -+-+-----+ -+-+--++-- --+++-++-- ++--+
     BEND.....   SSSS            SS     SS  SSS  SSS S
     5-TURN...
     4-TURN...                            >4 44<
     3-TURN...
     SUMMARY..   SSSS    E EEEEE SS     SS  SSST TTS S
     EXPOSURE.  1*5**10100 00020028*29 41*4050930 01007967
201  SEQUENCE.  SPDSHPADGI AFFISNIDSS IPSGSTGRLL GLFPDAN

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

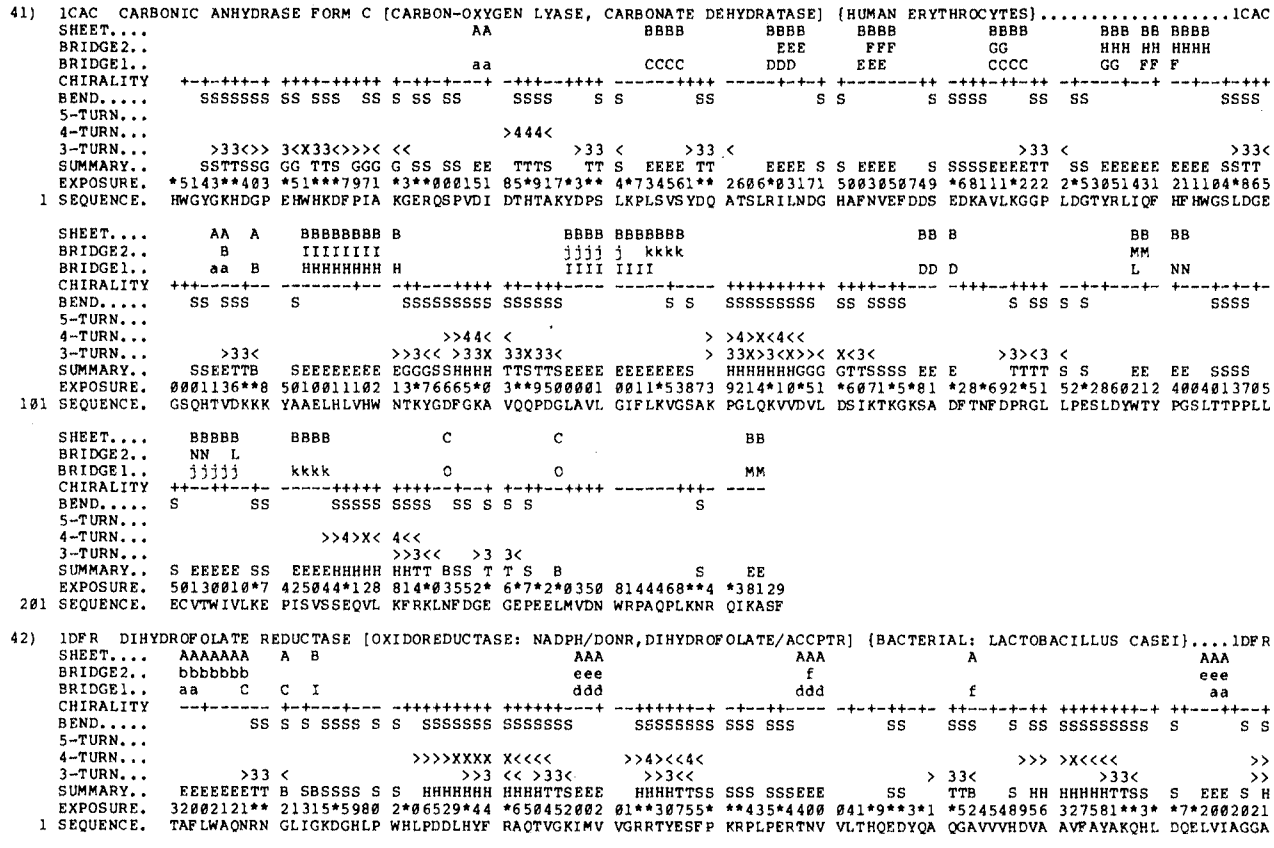


TABLE AIII (*continued*)

```
     SHEET....              AAAAAAA         B             A A AAAA          AAAAAAA
     BRIDGE2..              GGGGGGG                                         HHHHHH
     BRIDGE1..              bbbbbbb         I             H H HHHH          GGGGGGG
     CHIRALITY  ++++++++++ +-------+- --+-++---+ -++-+--+++ +------+-++ ++--+-----
     BEND.....  SSSSSSSSS  S       SS     S         S S  S        SSSS S
     5-TURN...
     4-TURN...  >>XX<<<<                                          >44 4<
     3-TURN...  3<<  >33<                                         >33 <
     SUMMARY..  HHHHHHHTS  SEEEEEEESS     SB        S S EEE EEEE  SSTT T EEEEEEE
     EXPOSURE.  *208528*81 6312006071 68*387*22* 289*929986 7**3*5*6*3 22022234** **
 101 SEQUENCE.  QIFTAFKDDV DTLLVTRLAG SFEGDTKMIP LNWDDFTKVS SRTVEDTNPA LTHTYEVWQK KA

  43)  1GPD  D-GYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE [OXIDOREDUCTASE: ALDEHYDE/DONR,NAD/ACCPTR] {LOBSTER: HOMARUS AMERICA1GPD
     SHEET....    AAAA                        AA                  B   B      CCCC   CCCC  CCC  AA                 A AAA
     BRIDGE2..      bb                        cc                                            III                    d ddd
     BRIDGE1..    aaaa                        bb                  G   G      HHHH   HHHH  III  cc                 a aaa
     CHIRALITY  +---+-+-+ ++++++++++ +--+-+---+ +++--++++ +++++-++-- --+--+-++- +-----+-+- ---+-+-+++ -++++-++++- --++-++++-
     BEND.....      S SSS SSSSSSSSSS SSS         SSS SSSSS SSSS  SSS     S     S SS  SSS         SSSSS S  SSS   S       SS  S
     5-TURN...       >555 5< >555>X5 55<<              >5 555<                                          >5555<
     4-TURN...          > >>>XXXX<<< <              >>>>XX X<<<<>444<                          >444 < >444<             >
     3-TURN...         >3 >X3<<                  >33<                        >3 3<  >33<          >33 <  >33<
     SUMMARY..    EEEE TTT HHHHHHHHHH TTT     EE  TTS HHHHH HHHH BTTTB    S EEEET TEEEETTEEE  EE  SSTTT S  TTTT  E EEE  SS  S
     EXPOSURE.  5510112135 3011014014 *535752001 23*247*412 601590З3264 82*6*2*9** 81131*59*1 533*6**09* 0308*73161 000155*541
   1 SEQUENCE.  SKIGIDGFGR IGRLVLRAAL SCGAQVVAVN DPFIALEYMV YMFKYDSTHG VFKGEVKMED GALVVDGKKI TVFNEMKPEN IPWSKAGAEY IVESTGVFTT

     SHEET....               AAAAA         A                AAAA                          D    DDDDD         E              E
     BRIDGE2..                eeee                            f                                 LLLL
     BRIDGE1..                dddd          f                 eeee                          J    kkkkk         P              P
     CHIRALITY  ++++++++++ +-+-+--+-- -++--+--+- ++-+++-+++ ++------+++ ++++++++++ +++----+-- ---------++ ----+++-++ ----+--+++
     BEND.....  SSSSSSSSSS S S    SS   SSS    SS S SSS  SS  S      SSS SSSSSSSSSS SSSS               SS   SS     S S      SSS
     5-TURN...  >5555<                                                    >>5 55<<
     4-TURN...  >>><XXX4<< <                  >44 4<                >>>> <XXX>XXX<X X<4<<
     3-TURN...     >>>X<<<                     > 3><3< >33<          >3 3<    >33<                    >33 <       >3 3X33<
     SUMMARY..  HHHHHHHHHH S SEEEEES   SSS  B TT TTTTS  TT  SEEEE  HHH HHHHHHHHHH HHHH  B    EEEEE   TT   SB     T T TT  BSSS
     EXPOSURE.  4*8Ø3205*2 21*3001226 1*74331011 13***16**4 6000003410 1100111511 4**2*14*1* 130610148* 2*3120536* **42240276
 101 SEQUENCE.  IEKASAHFKG GAKKVVISAP SADAPMFVCG VNLEKYSKDM TVVSNASCTT NCLAPVAKVL HENFEIVEGL MTTVHAVTAT QKTVDGPSAK DWRGGRGAAQ

     SHEET....    DDDDD                    DDDDD        DDD DDDD                              F F F               F FF      DDD
     BRIDGE2..                             MMMMM          N NNNN
     BRIDGE1..    MMMMM                    kkkkk        LLL L  J                              q q q               q qq      OOO
     CHIRALITY  ---+---+++ +++++++-+++ ++++------+ -+-++------ -+---+---- ++++++++++ +++-++++++ ---+------+ ++++-+----+ --++-+--+-
     BEND.....          S SSSSSSSSSS SSS         S  S          SS    SSSSSSSSSS SSSSS  SS      S    S   SSS  SS    SSSS
     5-TURN...                                                                    >5555<
     4-TURN...         >>> >XX<<<<                                 >  >>>XXXXXXX <<<<
     3-TURN...            >33<                              >33<                  >3><3<          >33 <              >33<  >
     SUMMARY..    EEEEE HH HHHHHHSSSS SSS  EEEEE  S  S  EEE EEEE TT    HHHHHHHHHH HHHTTTTTSE EEE S   TT   SSS  SSE EESSTT EEE
     EXPOSURE.  5*39164502 *0009216*0 66*3826065 3653000304 130905**23 2*913321*7 31*3*2*342 3319***493 92637*8110 110*542*88
 201 SEQUENCE.  NIIPSSTGAA KAVGKVIPEL DGKLTGMAFR VPTPDVSVVD LTVRLGKECS YDDIKAAMKT ASEGPLQGFL GYTEDDVVSS DFIGDNRSSI FDAKAGIQLS

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

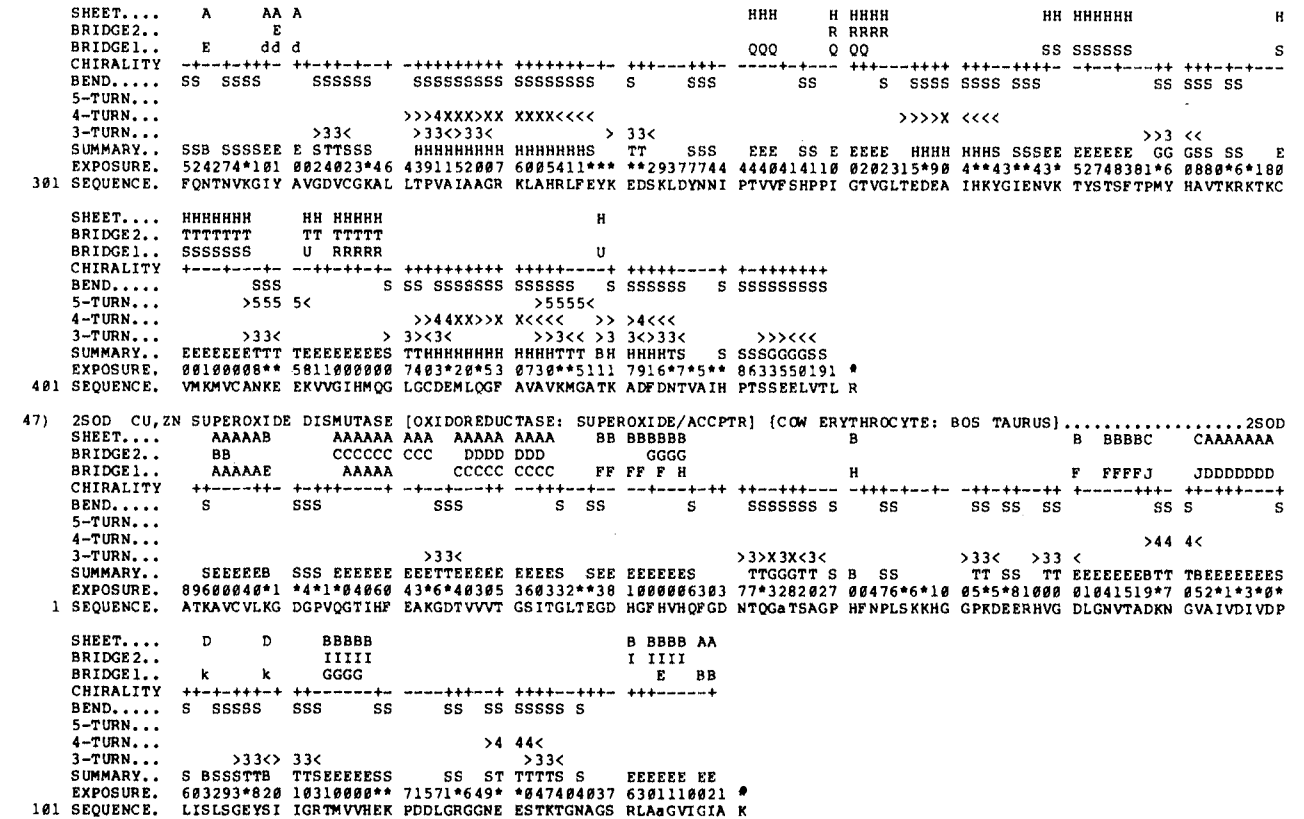
```
    SHEET....    DDDDD
    BRIDGE2..    OOO
    BRIDGE1..    NNNNN
    CHIRALITY  +++-+-+--- ++++++++++ ++++++++++ +
    BEND.....  SS           SSSSSSSS SSSSSSSSSS S
    5-TURN...                              >55 55<
    4-TURN...             >>>><XX<>X X>XXXXXX<< <<
    3-TURN...  33<          >>3XX3<<  >33<
    SUMMARY..  TTEEEEEE     HHHHHHHHH HHHHHHHHHH HT
    EXPOSURE.  85*1711031 0123120310 03119603*3 28*
301 SEQUENCE.  KTFVKVVSWY DNEFGYSQRV IDLLKHMQKV DSA

44)  4ADH  APO-LIVER ALCOHOL DEHYDROGENASE (OXIDOREDUCTASE: CHOH/DONR, NAD/ACCPTR) {HORSE LIVER: EQUUS CABALLUS}.....4ADH
    SHEET....       AAAA AAAB        AAAAAAA       CCCCC   DDD                    B     D   CC CC            CEE E
    BRIDGE2..            B             CC          HHHH    K                     f         I
    BRIDGE1..       AAAA AAAE        AAAAAAA       GGGGG   JJJ                    E     K   HH HH                IMM M
    CHIRALITY   -++---+-+ ---++-++-- --+------- --++------+ ------++++ +++-++-+-- ++--++++-+ --++---+++ ++--++-+-- --+--+-++-
    BEND.....    SSS           SSSS               SS              SSSS SSSSSSS    SS            S   SS   S  SS       SS   SSS
    5-TURN...                                                        >5555<
    4-TURN...                                                  >>>>X <<<<                                              >
    3-TURN...              >33<              >33<              >>3<<                          >33<    >33<
    SUMMARY..   SSS EEEE EEEB STTS   EEEEEEE     TTEEEEE   EEE  HHHH HHHTTSS    SSB     B   EEEEE TT   S  TT BEE E SS   SSS
    EXPOSURE.  **54*81*09 00008****9 14759051 35 1*4*100060 400001*501 3015244*14 5201000000 0613232*51 *60*762500 00201246*2
  1 SEQUENCE.  STAGKVIKCK AAVLWEEKKP FSIEEVEVAP PKAHEVRIKM VATGICRSDD HVVSGTLVTP LPVIAGHEAA GIVESIGEGV TTVRPGDKVI PLFTPQCGKC

    SHEET....                                A AA  AA B          A CC CCC   EEE                                        FFFF
    BRIDGE2..                                  DD                                                                     oooo
    BRIDGE1..                                C C   DD f          B GG GGG   MMM                                        nnnn
    CHIRALITY  ++++-++++- -++-++--+- +--+-+--++ +--+------ -+-+++-++- +--+++---- -++--++++ ++-+++++++ ++++++--- ++------+-+
    BEND.....  SSSSSSS     SS SSSS    S SSS  S   SSSS      SS   SSSS     SSS      SS  SSSSS SSSSSSSSS SSSSSSS    SS     S
    5-TURN...                                                                              >5555<     >>>55<<<
    4-TURN...  444<                                    > 444<                        >444<     >>>>XXX X<<<<
    3-TURN...      >33<                  >33<      >33<    > 33<          >>3<<       >33X>> X<<<    >3 3<       > 33<
    SUMMARY..  TTTSSTT     SS SSSS    S TTS  SE EETTEE B   TTT  SBSEE EEEGGGEEE   SS  TTTGG GGGTHHHHHH HHHHTTT    TT  EEEE
    EXPOSURE.  *32*8**229 09*2268*3* 031*756128 10*7*71510 13010000350 00323100*0 867041*700 0001100202 0005811*18 *6120000011
101 SEQUENCE.  RVCKHPEGNF CLKNDLSMPR GTMQDGTSRF TCRGKPIHHF LGTSTFSQYT VVDEISVAKI DAASPLEKVC LIGCGFSTGY GSAVKVAKVT QGSTCAVFGL

    SHEET....                    FF FF                FF F                      G FFFF              G     FFF F
    BRIDGE2..                     p pp                                             qqqq                    rrr r
    BRIDGE1..                    nn nn                pp p                      s oooo                 s    qqq q
    CHIRALITY  -+++++++++ +++---+--- -+-+-+++++ ++++-+-+-- --+++-+--+ ++++++++-+- --+---++-+ -+++++++++ +--+++---- +++--++---
    BEND.....  SSSSSSSSSS SSSSS         S SSSSS SSSSSS S     SSSSSS S SSSSSSSSSS   S     S   SSSSSSSSS SS SSS          SS
    5-TURN...                                 >5555<                      >5555<
    4-TURN...  >>>>XXXXXX X<<<<              >>> >X<<<<      >444<  >> 4>X<4<<                >>>>XXX<<< < >444<
    3-TURN...                             >>3<<                >33<  >3 3<>33X33<                      >3 3<>33<          >33<
    SUMMARY..  SHHHHHHHHH HHHHS    EE EE S GGGHH HHHHHT SEE E TTTSSS H HHHHHHTTTS  BSEEEE S    HHHHHHHHH TB TTT EEE E    TT
    EXPOSURE.  3140000020 0*64145700 01379**5*6 60**120774 124**8*961 7*104*62*4 0042000281 95*1033002 003*5*0300 332836*5**
201 SEQUENCE.  GGVGLSVIMG CKAAGAARII GVDINKDKFA KAKEVGATEC VNPQDYKKPI QEVLTEMSNG GVDFSFEVIG RLDTMVTALS CCQEAYGVSV IVGVPPDSQN

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
     SHEET....              FFFF                                  D DDD D                  DD DDD
     BRIDGE2..                                                                             11 111
     BRIDGE1..              rrrr                                  1 111 1                     JJJ
     CHIRALITY  -+--++++-- +--+-+--+-- +--+++++++ ++++++-++- +++++++-+- -+++++++++ ++++-+-+-- --
     BEND.....      SSSSSS S      SSS S  SSSSSSS SSSSSSSS     SSS       SS  SSSSS SSSSS
     5-TURN...        >555 5<
     4-TURN...       >>44<<             >>4>XX>X XXX<<<<     >44<       >>>>XXX <<<<
     3-TURN...      >>3<X33 <      >>3 <<    >33<     >33<    >33<      >3><3<  >3 3X33<
     SUMMARY..       THHHHT T EEEE SGG G  HHHHHHH HHHHHHTS     TTTEEEEEE ETTTHHHHHH HHHTS   EE EEE
     EXPOSURE.  5*898*46*6 4376854610 12305*104* 1055385**1 5072034593 50**04*009 538*5*1320 0057
 301 SEQUENCE.  LSMNPMLLLS GRTWKGAIFG GFKSKDSVPK LVADFMAKKF ALDPLITHVL PFEKINEGFD LLRSGESIRT ILTF

 45) 4LDH  LACTATE DEHYDROGENASE, APO ENZYME M4 [OXIDOREDUCTASE: CHOH/DONR, NAD/ACCPTR] {DOGFISH MUSCLE:SQUALUS ACANTHIUS4LDH
     SHEET....                      AAAAA                       AAAA A                          AAAAA            AAAA
     BRIDGE2..                       bbbb                       cccc c                            dd
     BRIDGE1..                       aaaaa                      aaaa a                       ccccc               bbbb
     CHIRALITY   -+++++++-- -+++--+--+ +----+-+++ ++++++++++ +-++-+----- +-+-++++++ ++++++++++ +-+-+++--- +-+++++++-+ ---+++----
     BEND.....    SSSSSS    S       S S       SS SSSSSSSSSS SS   SS      S SSSSSS SSSSSSSS S SS  S      SSSSSS S S
     5-TURN...    >5555<                              >5 555<
     4-TURN...    >>44<<                           >> >>XXXXXX<< <<            >>>>XX< XXX4<<<
     3-TURN...   >>3<<     >33<                     >3 3X33<                     >33<   >>3<<>3>X3 <<           >>3<<
     SUMMARY..    THHHHS    TT      S SEEEEE  SH HHHHHHHHHH HTT  SEEEE E S HHHHHH HHHHHHTTGG GS  SEEEEE SSSGGG S S EEEE
     EXPOSURE.  *8**9**96* **9*****2* 6100000163 3021008101 **71061001 02*99*91*4 43**37*25* *4*1882421 *927303417 00001442*6
   1 SEQUENCE.  ATLKDKLIGH LATSQEPRSY NKITVVGCDA VGMADAISVL MKDLADEVAL VDVMEDKLKG EMMDLQHGSL FLHTAKIVSG KDYSVSAGSK LVVITAGARQ

     SHEET....                                  AA                          A                                BB  C       BB
     BRIDGE2..                                  e                                                                   
     BRIDGE1..                                  dd                          e                                FF  G       FF
     CHIRALITY  -++--+++++ ++++++++++ ++++++++++ ----+-+++ ++++++++++ -+--++++-- ++++++++++ ++++++++-+- -+++-+-+++ +-++++----
     BEND.....   SS  SSSSS SSSSSSSSSS SSSSSSS SS       SS S SSSSSSSSSS S   SSS      SSSSSSSS SSSSSSSSS   SSSS      SSSSSS
     5-TURN...      >5555<                                                                          >5555<
     4-TURN...      >>>4XX X>XX<XX<4X X>>X<<<<          >> >>XXXXXX<< <<          >4>>X4X< <><4><44<  >444<              >
     3-TURN...         > 33<                >33X33 <          >33<      >33<    >33<   >>3 <<                        >33<
     SUMMARY..   SS  HHHHH HHHHHHHHHH HHHHHHH TT    EE SS H HHHHHHHHHH H   TTS B   TTTTHHHHH TTTTTTTTT   TTTS  EE  BSSSTT  EE
     EXPOSURE.  **584*6*11 39008707*1 06503*307* 0000022*31 0200300392 2617*74000 1002211470 7*310***75 *386030500 0262*40110
 101 SEQUENCE.  QEGESRLNLV QRNVNIFKFI IPNIVKHSPD CIILVVSNPV DVLTYVAWKL SGLPMHRIIG SGCNLDSARF RYLMGERLGV HSCSGVGWVI GQHGDSVPSV

     SHEET....       D D                              E   E                      FFF   CCC             CCC  F FFG  G   F
     BRIDGE2..                                                                                          HHH  L
     BRIDGE1..       I I                              j   j                      KKK   HHH               G  K KKM  M   L
     CHIRALITY  +++++-+-+-- ++++-+++-+ --+++++++ ++++++++++ +---++++++ ++++++++++ +-+-+---+- +--++++--+- +-+-+-+---+-+-+------
     BEND.....  SSS  SSS   SSSSS   SS SSSSSSSSSS SSSS  SSSS SSS   SSSS SSSSSSSSSS SS            SSSSS   SS           SSS  S
     5-TURN...                                 >>5 55<<                        >5 555<
     4-TURN...  444<     > >>4<<<          >44>X4 4<<   >>44 <<   >>>>X X<X<X>X><< <<
     3-TURN...       >33<  >33<          >>><<<      >33<>33 <     >33< >>3<<    > 33<          >33X33<                >33<
     SUMMARY..  TTT  BTTB  HHHHH    SS SSSGGGGTHH HHTSTTSHHH HBS   HHHH HHHHTHHHHH HTT  EEE   EEE TTSTT   SS EEE  E EEBTTB  SB
     EXPOSURE.  3610756**8 248*6542** 9*947*23*6 60262**68* **6*51*420 *1006003D0 2*98696110 01106*7551 *641000010 1035*014*3
 201 SEQUENCE.  WSGMWNALKE LHPELGTNKD KQDWKKLHKD VVDSAYEVIK LKGYTSWAIG LSVADLAETI MKNLCRVHPV STMVKDFYGI KDNVFLSLPC VLNDHGISNI

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

```
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  ------++++ ++++++++++ +++++-+
     BEND.....       SSSS SSSSSSSSSS SS    S
     5-TURN...
     4-TURN...       >>4>X X>XXXXXXXX <<<<
     3-TURN...       >33X3 3<>33<     >33<
     SUMMARY..        HHHH HHHHHHHHHH HHH   S
     EXPOSURE.  6*4*29**16 *705*01962 7*28*2**9
 301 SEQUENCE.  VKMKLKPNEE QQLQKSATTL WDIQKDLKF

 46) 2GRS  GLUTATHIONE REDUCTASE [OXIDOREDUCTASE: GSSG/ACCPTR, NADPH/DONR, FLAVOENZYME] {HUMAN ERYTHROCYTE}.............2GRS
     SHEET....         AA   B                AA AAA    B
     BRIDGE2..          b                    cc ccc
     BRIDGE1..         aa   f                   aa     f
     CHIRALITY   +--++--+- +-+++++++++ ++++-+---- -+-+--+--+ ++++++++++ ++++++++++ ++-++++-+- --+----+++ ++++++++++ ++++++++++
     BEND.....      S    S  SSSSSSSSS SSSSS         SS SS S SSSSSSSSSS SSSSSSSSSS SSSS   SS          SS SSSSSSSSS SSSSSSSSS
     5-TURN...                        >5555<
     4-TURN...              >4>><>XX> X<<<<             >>> ><<<X>>>XX XXXXXXXXX< <<<              >>> >XXXXXXXXX XXXXXXXXX
     3-TURN...                       >>3<<           >3><3 <       >3 3<              >>3<<                    >33<
     SUMMARY..      S  EES  BTTHHHHHH HHHHTT  EE EEESS TBHH HHHHSHHHHH HHHHHHHHHH HHSS GGGS          HH HHHHHHHHH HHHHHHHHH
     EXPOSURE.  *553700001 0131011005 702*960*00 0036**2026 2024253037 4028304*23 *5*6045*7* **8*5*7*1* 732**05391 7*23698*86
   1 SEQUENCE.  VASYDYLVIG GGSGGLASAR RAAELGARAA VVESHKLGGT aVNVGaVPKK VMWNTAVHSE FMHDHAVYGF PSCEGFFNWR VIKEKRDAYV SRLNAIYQNN

     SHEET....       AAAAA     C        C      C     AAA     DD              EE              EEEE                      EEE
     BRIDGE2..                           H            ddd                                    111                       mm
     BRIDGE1..       ccccc     G        G      H       b     II              jj              kkkk                      111
     CHIRALITY  +++-+----+ --+---+-+- +--+--+--- +-+---++++ +------+++ -++++++--+ ++++++++-- ++---+-+-+ ++++++++++ +-+-----+-
     BEND.....  SSSS       S     SSSS S   SSS      SS             SSS SSS SS   S SSSSS SS   SS      SS SSSSSSSSS SSS
     5-TURN...  >5555<                                                                                        >5 555<
     4-TURN...  <<<<                                                    >> 44<<                 >> >>XXXXXX<< <<
     3-TURN...  >33<                        >33<                    >33 <  >33< >3 3X>3<<              >33<
     SUMMARY..  HHHTTEEEEE S   B SSSS S B STT  B   SSEEE    EE    STT SSS TT EEH HHHTT SS   SSEEEE  SH HHHHHHHHHH HTT   EEE
     EXPOSURE.  09*391961* 2*12117*9* 51081*6854 5085000150 183392**9* 4710972631 4202*5**61 3200021248 5001111003 1241900100
 101 SEQUENCE.  LTKSHIEIIR GHAAFTSDPK PTIEVSGKKY TAPHILIATG GMPSTPHESQ IPGASLGITS DGFFQLEELP GRSVIVGAGY IAVEMAGILS ALGSKTSLMI

     SHEET....                              EE   FFF FFF       FFFFFF        FFF F     EEEE    DD              G      G
     BRIDGE2..                                                   0000                    kkkk
     BRIDGE1..                              mm   NNN NNN       NNNNNN        000 0       jj    II              P      P
     CHIRALITY  +-+-++++-- ++++++++++ +++-++---- +--++---++ -+-------+ +-+------- -----+---+ +-----++++ -++++-+--- -+-+--+--+-+
     BEND.....  SSSSS  SS  SSSSSSSSSS SSSSS      S       S  SSS         SSSS         SS       S    S SS S  SSS      SSS     S
     5-TURN...                                                                                    >5555<
     4-TURN...          > >>>XXXXX<< <<                                                           >444<
     3-TURN...      >33<> 33<       >33<    >3 3<     >33 <                                  >33 <>33<      >33<
     SUMMARY..  SSSSS TTS  HHHHHHHHHH HTTSS  EET TEEEEEEETT SSS EEEEEE  SSSS  EEE E  SS EEEE  S EE S TT S TTTTT  B  TTS B   S
     EXPOSURE.  *68301**40 54016101*2 386532922* 5091*908** 4*33613143 57*6**6**8 392*952101 29555172** 141**351*8 6**240647*
 201 SEQUENCE.  RHDNVLRSFD SMISTNCTEE LENAGVEVLK FSQVKEVKKT LSGLEVSMVT AVPGRLPVMT MIPDVDCLLW AIGRVPNTKD LSLNKLGIQT DDKGHIIVDE

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)



```
48)  1MBN  MYOGLOBIN [OXYGEN STORAGE] (FERRIC IRON - METMYOGLOBIN) {SPERM WHALE}.....1MBN
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   --++++++++ ++++++++++ ++++++++++ +++++-++++ ++-+++++++ -+++++++-++ ++++++++++ ++++++++--+ ++++++++++ +++++++-+--
     BEND.....     SSSSSSS SSSSSSSSSS SSSSSSSSSS SSSSS SSSS SS SSSSS   SSSSSSS SS SSSSSSSSSS SSSSSSSSSS    SSSSSSS SSSSSSS
     5-TURN...                                                                                >5555<              >>555<<
     4-TURN...    >>>>XXXX XXXX<<<< > >>>XXXXXXX XX<<<X>>>4< <<>>44<<<   >>>>>><<X>> >XXXXXXXXX XXX<<<<     >>4>XX>XX XX<<<<   >
     3-TURN...             >3>X3<<                 >>3<X 33X33X33<            >33<          >>3X<3 <>33<                 >
     SUMMARY..     HHHHHHH HHHHHHHTTS HHHHHHHHHH HHHHH HHHH HT HHHHT   SHHHHHH HH HHHHHHHHHH HHHHHHTTTT   HHHHHHHH HHHHHTT
     EXPOSURE.  *15*551*50 5*01650997 3210026111 *11*558812 *86**08*1* 68791*83*6 07*305*408 41031*882 *0*781*740 74508*9*25
   1 SEQUENCE.  VLSEGEWQLV LHVWAKVEAD VAGHGQDILI RLFKSHPETL EKFDRFKHLK TEAEMKASED LKKHGVTVLT ALGAILKKKG HHEAELKPLA QSHATKHKIP

     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  ++++++++++ ++++++++++ +++-++++++ ++++++++++ +++++++++-+ +
     BEND.....  SSSSSSSSSS SSSSSSSS S SS SSSSSSS SSSSSSSSSS SSSSSSSSSS
     5-TURN...                                          >5555 <
     4-TURN...  >>>XXXXXXX XXXXX<<<X4 44<>>>>XXX XXXXXXXXXX XXXXXX<<<<
     3-TURN...  33<  >33X3 3<   >33<  >33<                  >33<
     SUMMARY..  HHHHHHHHHH HHHHHHHH T TTTSHHHHHH HHHHHHHHHH HHHHHHHHHT
     EXPOSURE.  4*42634091 119109892* 78048*1730 06900*223* 601621**67 *8*
 101 SEQUENCE.  IKYLEFISEA IIHVLHSRHP GNFGADAQGA MNKALELFRK DIAAKYKELG YQG

49)  1ECD  HEMOGLOBIN (ERYTHROCRUORIN, DEOXY) [OXYGEN TRANSPORT] {CHIRONOMOUS THUMMI THUMMI}.....1ECD
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   -++++++++ ++++++++++ ++++++++++ -+++++++-++ +++--+++++ +-+++++++++ +++++++++- +-+++++++ +++++++++- +--++++++++
     BEND.....    SSSSSSSS SSSSSSSS S SSSSSSSSSS  SSSSSS SS SSSS SSSSS SSSSSSSSSS SSSSSSSSSS SS S  SSSS SSSSSSSSSS    SSSSSSS
     5-TURN...                                                                                  >555 5<
     4-TURN...   >>>>XXXXX X<X<<<4< >> >>XXXXX<<< X>44<< >44 4<  >>44<<  >>>>XXXXX XXXXXXXX<<< <<   >>>>X XXXX<<<<     >>>>XXXX
     3-TURN...   >33<       >>3<X33<              >33X>3<X33 X33<>33<>3 3<                  >> 3X<3<            >>3< <
     SUMMARY..    HHHHHHHH HHHHTTTT H HHHHHHHHHH  HHHHTT TT TTTS HHHHT TSHHHHHHHH HHHHHHHHHH HTTT  HHHH HHHHHHGGG T  HHHHHHH
     EXPOSURE.  768*526507 611**1*861 31114100*5 295226*5*9 059*82*818 86750*7508 *514724*11 5428*0*620 8832755*** 4367*21754
   1 SEQUENCE.  LSADQISTVQ ASFDKVKGDP VGILYAVFKA DPSIMAKFTQ FAGKDLESIK GTAPFETHAN RIVGFFSKII GELPNIEADV NTFVASHKPR GVTHDQLNNF

     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  ++++++++++ +-+++++++++ ++++++++++ ++++
     BEND.....  SSSSSSSSSS SS  SSSSSS SSSSSSSSSS SSSS
     5-TURN...
     4-TURN...  XXXXXXXX<<< << >44>X>> XXXXXXXXXX XX<<<<
     3-TURN...               >>><<<              >33<
     SUMMARY..  HHHHHHHHHH HS GGGGHHH HHHHHHHHHH HHHHH
     EXPOSURE.  65003710*8 *3*1571981 02201*6214 406955
 101 SEQUENCE.  RAGFVSYMKA HTDFAGAEAA WGATLDTFFG MIFSKM

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  +++++-+---  -+++++++++ +++++--+-+ -+++++++++ ++++++
     BEND.....  SSSSSSSS     SSSSSSSSS SSSSSS  S   SSSSSSSSS SSSSSS
     5-TURN...  5555<
     4-TURN...  XX<<X<44<      >>>>XXX XXX<<<<      >>>>XXXXX X<<<<
     3-TURN...               >>3<<  >33 <                    >>><<<
     SUMMARY..  HHHHHTTT     GGGHHHHHH HHHHHH  S    SHHHHHHHH HHHHGGG
     EXPOSURE.  Ø3*71****2 62*5182305 221*62378* 2124811732 1421*55*
   101 SEQUENCE.  SGKHAKSFEV DPEYFKVLAA VIADTVAAGD AGFEKLLRMI CILLRSAY

  52)  1HBL  LEGHEMOGLOBIN (ACETATE,MET) [OXYGEN TRANSPORT] {YELLOW LUPIN ROOT NODULES: LUPINUS LUTEUS L}.................1HBL
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   -+-++++++ ++++++++++ ++++++++++ ++++++-+++ +++-+++-++ ----++-+++ ++++++++++ ++++++++++ -+---+-+++ ++++++++-
     BEND.....    S SSSSSS SSSSSSSSSS  SSSSSSSSS SSSSSS SSS SSS SSS SS     SS SSS SSSSSSSSSS SSSSSSSSSS SSS    SSS SSSSSSSSS
     5-TURN...                                                                          >555 5<                  >555
     4-TURN...     >>>>XXX XXXXXX<<<< >>4>XX>XXX <XX<4<<  > 444X444<         >>>> XXXXXXXXXX XXXXXXXX<< <<    >>>> XXXXXX<<<<
     3-TURN...  >33<              >>3 X<3<              >>>X X<<X>3<X33 <            >33<                        >33
     SUMMARY..   TT HHHHHH HHHHHHHHHT THHHHHHHHH HHHHHH GGG GGG GGG TT     SS HHH HHHHHHHHHH HHHHHHHHHH HSS    HHH HHHHHHHHHT
     EXPOSURE.  2*56*61156 Ø9703*7088 654*2Ø5810 3103*83981 5*65620*3* 3*23**4**Ø 477Ø2*4Ø44 14*114826* 8682578762 993Ø5555**
       1 SEQUENCE.  GALTESQAAL VKSSWEEFNA NIPKHTHRFF ILVLEIAPAA KDLFSFLKGT SEVPQNNPEL QAHAGKVFKL VYEAAIQLEV TGVVASDATL KNLGSVHVSK

     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  +--+++++++ ++++++++++ ++++++-+++ ++++++++++ ++++++++++ +
     BEND.....     SSSSSSS SSSSSSSSSS SS SS  SSS SSSSSSSSSS SSSSSSSSSS S
     5-TURN...  5<
     4-TURN...    >44>X>>X XXXXXXXXX< <<<   >>>> XXXXXXXXXX XXXXXXXX<< <<
     3-TURN...  < >>3X<3<               >>3<<                      > 33<
     SUMMARY..  T  GGGHHHH HHHHHHHHHH HHGGG  HHH HHHHHHHHHH HHHHHHHHHH HT
     EXPOSURE.  535*863734 4*ØØ2*ØØ9* 114**54*61 34ØØ35Ø268 119225*41* *4*
   101 SEQUENCE.  GVADAHFPVV KEAILKTIKE VVGAKWSEEL NSAWTIAYDE LAIVIKKEMD DAA

  53)  1CRN  CRAMBIN [PLANT SEED PROTEIN] {ABYSSINIAN CABBAGE SEED: CRAMBE ABYSSINICA}.....................................1CRN
     SHEET....   AA                                 AA
     BRIDGE2..
     BRIDGE1..   AA                                 AA
     CHIRALITY   ---+-++++ ++++++++-+ --++++++++- +----+-+-- -+++
     BEND.....      SSSSSS SSSSSSSSS    SSSSSSSS S    SSS    SSS
     5-TURN...                  >5555 <
     4-TURN...       >>>>X XXXX<<<<    >>>>XX<<< <
     3-TURN...                 >3><3<  >33<                 >>3<<
     SUMMARY..   EE SSHHHH HHHHHHHTTT   HHHHHHHH S EE SSS    GGG
     EXPOSURE.  82Ø055**16 6819*29**5 4876*Ø1**2 41583*48*5 58*778
       1 SEQUENCE.  TTabPSIVAR SNFNVcRLPG TPEAIcATYT GbIIIPGATa PGDYAN

   SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
54)  1OVO  OVOMUCOID THIRD DOMAIN [PROTEINASE INHIBITOR, KAZAL] (JAPANESE QUAIL: COTURNIX COTURNIX JAPONICA).............1OVO
     SHEET....      AAA                  BBB    B B                   B  BB
     BRIDGE2..                           CCC
     BRIDGE1..      aaa                  BB     B B                   C  CC
     CHIRALITY    +----++++ +--+---+--+ -----+-+-- ---++++++++ +++--+++-+ --+-
     BEND.....           S    SS            SSS    SSSSSSSSS SSSSSSS      SS
     5-TURN...                                                 >5555<
     4-TURN...                                   >>>>XXXX <<<<
     3-TURN...         >33 <                 >33<              >>3X<3<
     SUMMARY..      EEE  TT    SS          EEETTS E ESSHHHHHHH HHHTTTS  E EEES
     EXPOSURE.   *969715Ø2* 258*837*9* *3ØØ55**6 5518Ø6ØØ22 12*374*171 7793*8
   1 SEQUENCE.   LAAVSVDaSE YPKPAbPKDY RPVcGSDNKT YSNKbNFaNA VVESNGTLTL NHFGKc

55)  2SSI  STREPTOMYCES SUBTILISIN INHIBITOR [PROTEINASE INHIBITOR].....................................................2SSI
     SHEET....      AAAAAA  B            AAAACC      CC                                        D  B  AAAAAA   AAAAAA   B D
     BRIDGE2..        BBBB                                                                      E  CCCCCC
     BRIDGE1..      AAAAA   D            BBBBFF      FF                                        g  D  AAAAA    CCCCCC   E g
     CHIRALITY    --+++---+ -+-+-+++-- +------++- +-++-+-+++ +++++++++- -+-+++---- -+----+-+- -----++--- -+--++++-- ---+++++++-
     BEND.....             S SSSSS    S       S S   SSSS S SSSSSSSSSS  S SS S    SS               S S SSS         SSSSSSSSS
     5-TURN...                                                  >555 5<
     4-TURN...                                        >> >>XXXXX<<< <                                          >>>><X<<
     3-TURN...                >33<                      >33<      >>><< <  >33<                        > 33<          >33<
     SUMMARY..      EEEEEE  B SSTTS    SEEEEEE S S  EESSTTH HHHHHHHHHH TS STTS    SS       B  B  EEEEEE TTEEEEEE   BSBHHHHHHT
     EXPOSURE.   *482425130 Ø*Ø936*936 **85251524 *376491*42 47Ø17717*3 6261962**5 **9*49*7*9 1110208223 *5**4*2*67 Ø618360720
   1 SEQUENCE.   YAPSALVLTV GKGVSATTAA PERAVTLTaA PGPSGTHPAA GSAaADLAAV GGDLNALTRG EDVMbPMVYD PVLLTVDGVW QGKRVSYERV FSNEbEMNAH

     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY  -++++
     BEND.....   SSSS
     5-TURN...
     4-TURN...  4<
     3-TURN...
     SUMMARY..  TSSSS
     EXPOSURE.  2*11Ø55
  1Ø1 SEQUENCE. GSSVFAF

56)  3PTI  TRYPSIN INHIBITOR [PROTEINASE INHIBITOR] (COW PANCREAS: BOS TAURUS).....................................3PTI
     SHEET....                    AAA AAAA    AA AAAAA          A
     BRIDGE2..                       B
     BRIDGE1..                    AAA AAAA    AA AAAAA          B
     CHIRALITY   -++++---- --+-+-+-+- ----++-+-- -+-----++-- -+++-+-+++ +++++-
     BEND.....    SSSSS      S           SSSSS        SSS   SS  SSSSS SSSSSS
     5-TURN...                          >5555<
     4-TURN...                          >444<                    >>>> XX<<<<
     3-TURN...   >>><<<                 >33<
     SUMMARY..    GGGGS      S    EEE EEEETTTTEE EEEEE SSS   SS BSSHHH HHHHHS
     EXPOSURE.  *5*4Ø*5*59 7295*5*6*5 72145*5482 88272Ø42*5 **Ø22*33*8 Ø7*5Ø24*
   1 SEQUENCE.  RPDFaLEPPY TGPbKARIIR YFYNAKAGLc QTFVYGGbRA KRNNFKSAED cMRTaGGA

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

```
57)  1CTX  ALPHA COBRATOXIN (COBRA: NAJA NAJA SIAMENSIS).....1CTX
     SHEET....                      A AAAAA          AAAAA            AAAAA
     BRIDGE2..                        BBBBB
     BRIDGE1..                      A AAAA           BBBBB            AAAAA
     CHIRALITY   --+-++--+ -++-++-+-- -+---+-+++ ++------+- +-+----+-+ +-+----+-+ +++----+-
     BEND.....       SSSSS      SSS          SSS SSS          SS     S S        S S  S  SSS
     5-TURN...
     4-TURN...                            >> 44<<
     3-TURN...              >33<        >33X 33<                              >3 3<    >33<
     SUMMARY..       SSSSS      TTS E EEEEE  TTH HHH  EEEEE   SS     S S EEEEE  T T  S  STT
     EXPOSURE.  5*0253***9 55957*1*20 324097**1 9**44*5600 345*83*7** 6554*33999 *307*49*8* ●
   1 SEQUENCE.  IRaFITPDIT SKDbPNGHVa YTKTWcDAFc SIRGKRVDLG bAATdPTVKT GVDIQdeSTD NeNPFPTRKR P

58)  1MLT  MELITTIN (HEMOLYTIC POLYPEPTIDE) (HONEY BEE VENOM: APIS MELLIFERA).....1MLT
     SHEET....
     BRIDGE2..
     BRIDGE1..
     CHIRALITY   +++++++++ ++++++++++ ++++
     BEND.....    SSSSSSSS SSSSSSSSSS SSSS
     5-TURN...
     4-TURN...  >>>>XXX<X< <>X>>XXXXX XX<<<<
     3-TURN...             >33<
     SUMMARY..   HHHHHHHHH TTHHHHHHHH HHHHH
     EXPOSURE.  9*3671**58 *4387727*3 8*****
   1 SEQUENCE.  GIGAVLKVLT TGLPALISWI KRKRQQ

59)  1NXB  NEUROTOXIN B (PROBABLY = ERABUTOXIN B) (SEA SNAKE: LATICAUDA SEMIFASCIATA).....1NXB
     SHEET....   AAA          AAA        BBBBBBB B  BBBBBBB B          BBBB
     BRIDGE2..                           CCCC
     BRIDGE1..   AAA          AAA        BBBBBBB B  BBBBBBB B          CCCC
     CHIRALITY   +-++---+- -+------++ ++-----+-+ -+------+- -------++- ++---+--++
     BEND.....       SSS  S       SS             SS        SS     SS        SSS
     5-TURN...
     4-TURN...
     3-TURN...       >33<        >33 <          >33<                        >3>< 3<
     SUMMARY..   EEE  STT  S  EEE  TT    EEEEEEE ETTEEEEEEE ES     SS   EEEE STTT T
     EXPOSURE.  860251859* **69**17*8 7620343987 6**3*21721 4065*6**65 *4754*6933 0*
   1 SEQUENCE.  RIaFNQHSSQ PQTTKTbSPG ESSaYHKQWS DFRGTIIERG bGcPTVKPGI KLScdESEVd NN

60)  2ADK  ADENYLATE KINASE (ATP-AMP PHOSPHOTRANSFERASE) (PORCINE MUSCLE: SUS SCROFA).....2ADK
     SHEET....          A AAAA                    AAAA                                                A AAA
     BRIDGE2..            bbbb                                                                      c ccc
     BRIDGE1..          a aaa                     cccc                                              a aaa
     CHIRALITY   +++++++--+ --++--++-+ ++++++++++ +-+---+-++ ++++++++-- -+++++++++ ++-+----++ ++++++++++ +++++++--- ---++++-++
     BEND.....    SSSSSS         SSS  SSSSSSSSSS SS      SS SSSSSSSSS   SSSSSSSSS SSSS    SS SSSSSSSSS SSS SS S     SS  SSSS
     5-TURN...                          >55 55<                    >5555 <
     4-TURN...  >>>><<<<           > 4>>X>XXX<< <<     >>> >XXXX<<<<  >>>>XXXXX< <<<    >>> >XXXXXXXXX <<X<44<            >
     3-TURN...                 >33 <                  >33 <                                          >33<
     SUMMARY..   HHHHHHS E EEEE  SSTT TTHHHHHHHHH HT  EEEEHH HHHHHHHHS   HHHHHHHHH HHSS    HH HHHHHHHHHH HHHTTT S E EEES  SSSS
     EXPOSURE.  9***2**240 1032114242 *3651*926* *543362480 923*429974 3940*803*2 5**6***554* 131812*921 77*4**2*00 112421**7*
   1 SEQUENCE.  MEEKLKKSKI IFVVGGPGSG KGTQCEKIVQ KYGYTHLSTG DLLRAEVSSG SARGKMLSEI MEKGQLVPLE TVLDMLRDAM VAKVDTSKGF LIDGYPREVK

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

TABLE AIII (*continued*)

```
     SHEET....                AAAAA                                                                A AAA
     BRIDGE2..                 dddd
     BRIDGE1..                 bbbb                                                                d ddd
     CHIRALITY  +++++++-++ --+-----+- -+++++++++ +++++-++-+ ++++++++++ ++++++++++ ++++++-+++ ---+++--++ ++++++++++ ++
     BEND.....  SSSSSSSSS    S         SSSSSSSSS SSSSSS       SSSSSSSS SSSSSSSSS SSSSSSSS          S SS SSSSSSSSS SS
     5-TURN...                                    >5555<     >5555<                                          >5555 <
     4-TURN...  >>>XX<<<<              >>>>XXXXX< <X<44<      >>>>XXXXX XXXX<<X<>> X>XX<<<<           >>> >XXXXXXXXX <<<<
     3-TURN...  >33<                            > >3<X33<                 >33<                        >33 <          >33<
     SUMMARY..  HHHHHHHHS    SEEEEE    HHHHHHHHH HHTTTT       HHHHHHHH HHHHHHTTH HHHHHHHS E EEE   S HH HHHHHHHHH HHH
     EXPOSURE.  405206*820 8252056371 2**316**1* *94**46*8* 86**52***4 *636*31980 12*27**365 7*2*4*699* 7347524943 **2*
 101 SEQUENCE.  QGEEFERKIG QPTLLLYVDA GPETMTKRLL KRGETSGRVD DNEETIKKRL ETYYKATEPV IAFYEKRGIV RKVNAEGSVD DVFSQVCTHL DTLK

  61)  1RHD  RHODANESE [THIOSULFATE-CYANIDE SULFURTRANSFERASE] {COW LIVER: BOS TAURUS}.....1RHD
     SHEET....          AA               A   AAA BB                  C    BB                                                AAAA
     BRIDGE2..                                ccc                                                                           ddd
     BRIDGE1..          aa                b   b   ee                  F    ee                                               ccc
     CHIRALITY   ++---++-- -++++++++++ -+++-++--+ -++++--++- -++++++++ ---+-----+ +++--++-+- -+----++++ +++++++-+-- ++---++--+
     BEND.....             SSSSSSSSS S   SSS          SSS   SSSSSSSS   SS     S SSSS S SSS S     SSSS SSSSSSSS   SSS
     5-TURN...                  >55 55X5555<
     4-TURN...             >>>>XXXX<< <<               >444 <>>>4X<<4<                                    >>>>X X<X<<4<
     3-TURN...                 >33X3 3<                >33<          >33<   >3 3<  >33<                    > 33<>>3<< > 33<
     SUMMARY..          EE  HHHHHHHHH HT BTTTEEE EE     TTT   HHHHHTTS B TT EE  T TSSS TTSSS S     HHHH HHHHGGGS   TTSEEEE
     EXPOSURE.  ****75*151 51*801*13* 75*51*4141 0101634*99 7*1**729** 0034023051 9*077*83*4 522227*730 1*50082004 6*02000102
   1 SEQUENCE.  VHQVLYRALV STKWLAESVR AGKVGPGLRV LDASWYSPGT REARKEYLER HVPGASFFDI EECRDKASPY EVMLPSEAGF ADYVGSLGIS NDTHVVVYNG

     SHEET....                          AAAA                C                   DD               DDDD
     BRIDGE2..                          ddd                                                      ii
     BRIDGE1..                           aa                 F                   gg               hhhh
     CHIRALITY  -+-++++++ +++++++-+- +++---+--+ +++++-+--- -+--+---+- ---+---+++ +--++++++ ++-++--+-- +--++++-++ ---++--+--
     BEND.....  SSSS SSSSS SSSSSSSS         SS S SSSSSS                    SSS    SSSSSSS SSS  S      S SSSSSS  S  SS SS
     5-TURN...                                    >5555<                                                >5555<
     4-TURN...       >>>> XXXX<<<<          >>> >XX<<<<                          >4>>X4XX <4<<         >>>4<<<
     3-TURN...  >33<       >>3X<3<      >33<                                 >33<                      >33X33<
     SUMMARY..  STTS SSHHH HHHHHHHT     EEEETTHH HHHHHHT    B                  TTS EE TTHHHHH HHH  SEEEE  S HHHHHT  S  SS SS
     EXPOSURE.  2*5016l000 0001030051 *300014002 8208*85576 49**2*4*84 *1*1976*71 3353*506*1 5*7**45200 012*25471* **17*5*737
 101 SEQUENCE.  DDLGSFYAPR VWWMFRVFGH RTVSVLNGGF RNWLKEGHPV TSEPSRPEPA IFKATLNRSL LKTYEQVLEN LESKRFQLVD SRAQGRYLGT QPEPDAVGLD

     SHEET....    E    DD       F       F                        DDDD                        DD D                E
     BRIDGE2..                                                     jj                          jj
     BRIDGE1..    K    ii       L       L                        hhhh                         g g                K
     CHIRALITY  +---++--+- -++++--+-+ ----++++++ ++++-+--++ -----+-+---+ ++++++++++ ++-++++--- -+--++++++ +--++++-+-+ -
     BEND.....       SS       SSSSS SSS     SSSSSS SSSSSS  SS S      SSS SSSSSSSSSS SSSS        SS SSSSSS SS SSSS  S S
     5-TURN...                                 >5555<                          > 5555<
     4-TURN...                            >>4>XX> X<<<<                 >>>>XXXX <<<<          >>>>X<< <<
     3-TURN...      >33<    >3><3<>33<        >33<  >>3<< >33 <          >3><3<   > 33< >33<   >3>X3<X33<   >>3<<
     SUMMARY..    B TT EE   TTTTB TTS  B  HHHHHH HHHHTT  TT S EEEE SSS STTHHHHHHH HHHT TT EE ETTTHHHHHH HS GGGSB S S
     EXPOSURE.  1030*61520 2297014*90 52538*80*5 40*88*164* *600000240 0100000000 1545110000 010003204* *167*2353* 5**
 201 SEQUENCE.  SGHIRGSVNM PFMDFLTENG FEKSPEELRA MFEAKKVDLT KPLIATCRKG VTACHIALAA YLCGKPDVAI YDGSWFEWFH RAPPETWVSQ GKG

SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND
```

```
62)  2PAB  PREALBUMIN [THYROXIN, RETINGL TRANSPORT] (HUMAN PLASMA).....2PAB
     SHEET....     AAAAAAA     AA    B BBBBBB      BBB BBBB      AA           BBB BBBB                  BBBBBBB        AAAAA
     BRIDGE2..       BB CC             FFFFF                                 GGG GGGG                                 DDDDD
     BRIDGE1..     aaaaaaa      CC    E EEEEEE      EEE EEEE      BB             F FFFF                  GGGGGGG        aaaaa
     CHIRALITY    ++-------+ +-+---++-- --------+-+ ---+--+--- -+-+-+++-- -++++-+--- ----++++++ ++-+---+-- ----+--+++ ----------
     BEND.....           S SSSS   S          SSS S           SSS    S   SSSS SS         SSSS SSSS   S S          S SS
     5-TURN...          >5 555<                                                  > 5555<
     4-TURN...          >4 44<                                      >444<          >>>>XX <<<<
     3-TURN...                          >33<           >33<        >33<            > >3<<
     SUMMARY..    EEEEEEET TTTEE  S E EEEEEE TTS SEEEEEEEE   TTSEE  S   TTTS SEEE EEEE HHHHH HHHT   S S  EEEEEEE S SS   EEEEE
     EXPOSURE.   62040080705 8*74406*01 0704761*** 99**4282*0 5*60418822 8***16*160 3050515630 **77693*5* 7092747037* *3*2*34040
   1 SEQUENCE.   CPLMVKVLDA VRGSPAINVA VHVFRKAADD TWEPFASGKT SESGELHGLT TEEQFVEGIY KVEIDTKSYW KALGISPFHE HAEVVFTAND SGPRRYTIAA

     SHEET....   AAA  AAAAA AAA
     BRIDGE2..   DDD
     BRIDGE1..   aa   DDDDD DDD
     CHIRALITY   -+-+++--+- --
     BEND.....      SS
     5-TURN...
     4-TURN...
     3-TURN...    >33<
     SUMMARY..   EEETTEEEEE EEE
     EXPOSURE.   7043*78668 799*
 101 SEQUENCE.   LLSPYSYSTT AVVT

 SUMMARY.....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....G=3-HELIX....I=5-HELIX....T=3-,4-, OR 5-TURN....S=BEND....
```

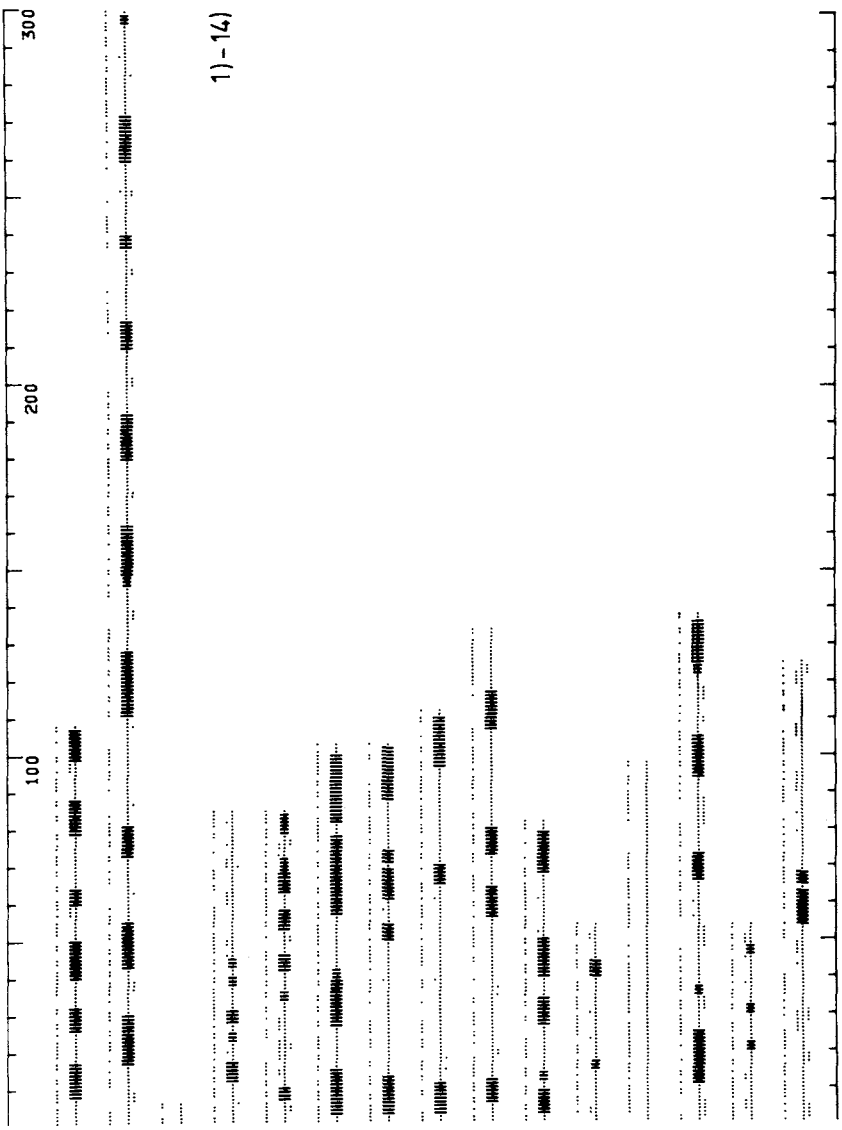



Fig. A1. Strip maps of secondary structure and solvent exposure for 62 proteins. Top line dots: residues with more than three contacting water molecules. Vertical bars: short, 3-helix; medium, 4-helix ($\alpha$-helix); long, 5-helix. Dots above baseline: residue has antiparallel $\beta$-bridge partner(s). Dots below baseline: $\beta$-strand has parallel $\beta$-bridge partner(s). The four-letter code is the Protein Data Bank data set identifier (Table AI).

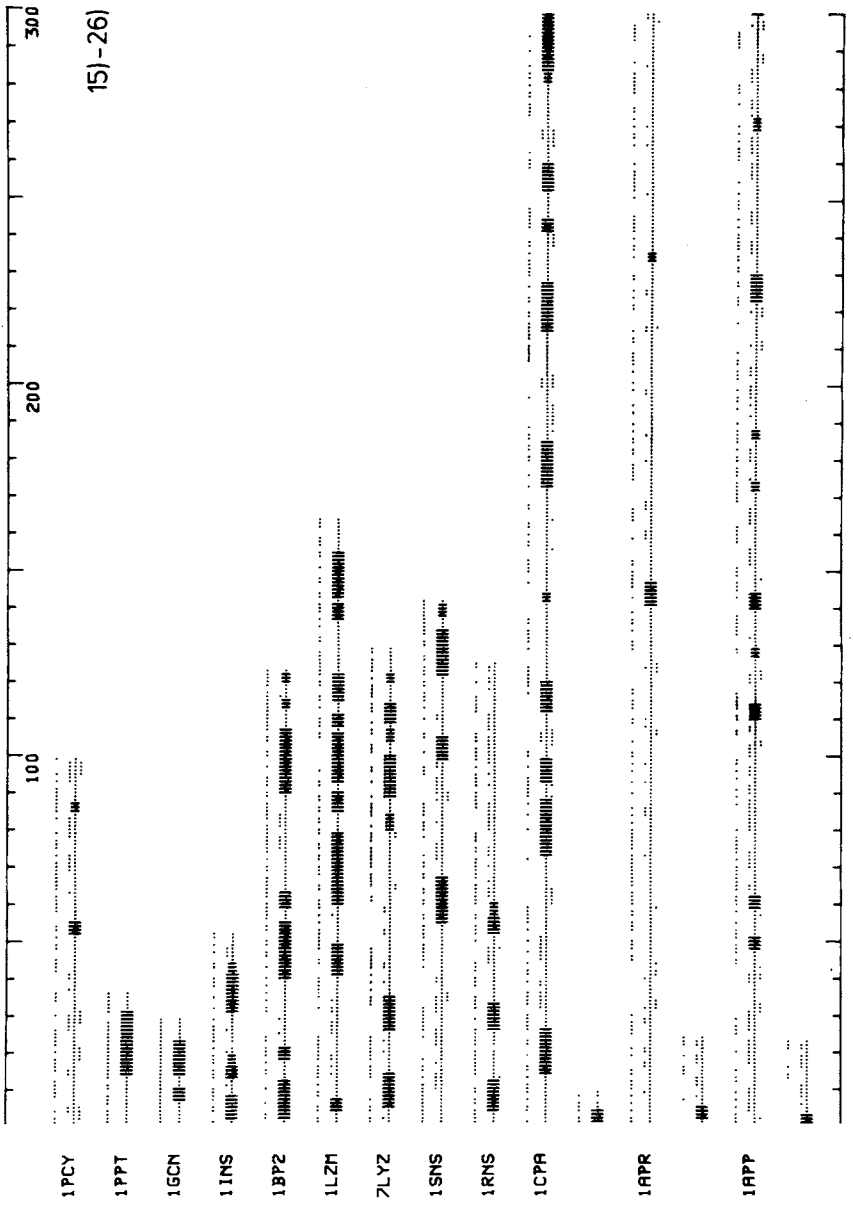



Fig. A1. (*continued from the previous page*)

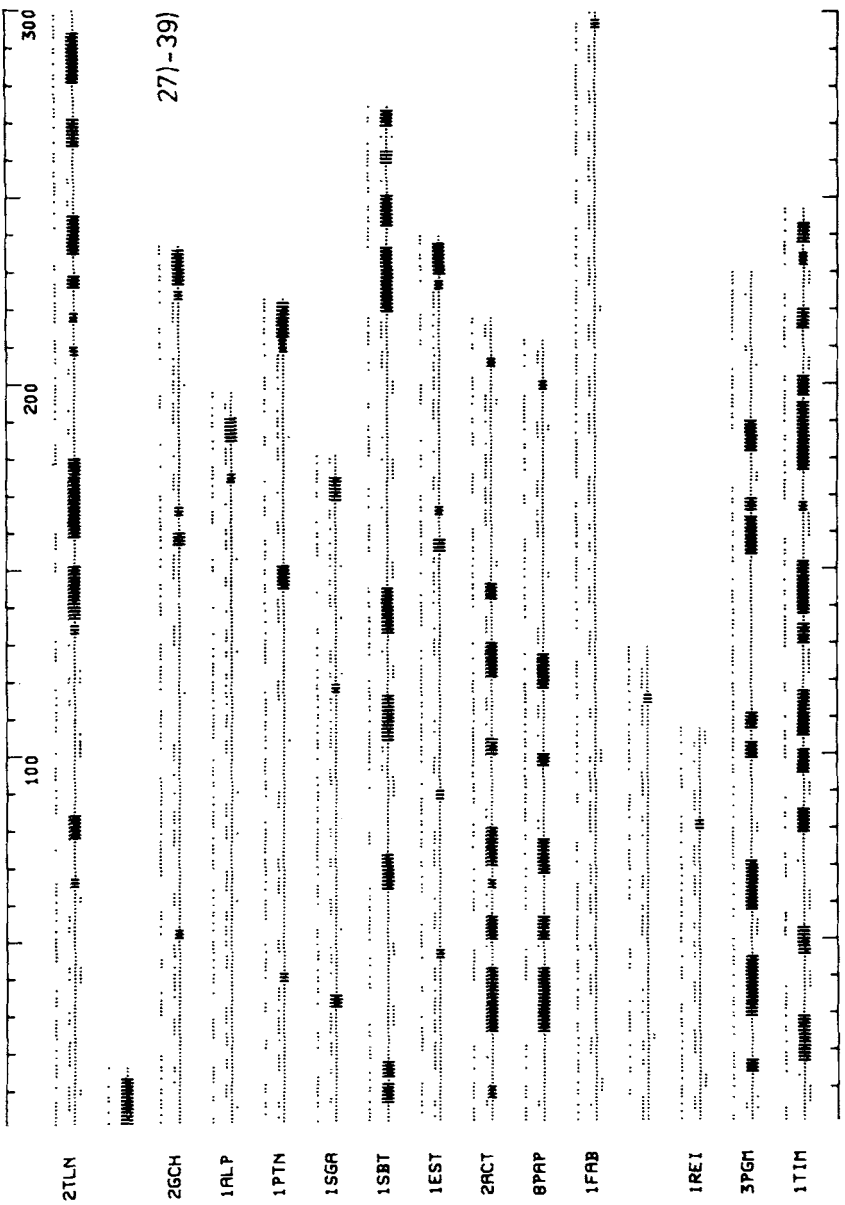



Fig. A1. (*continued from the previous page*)

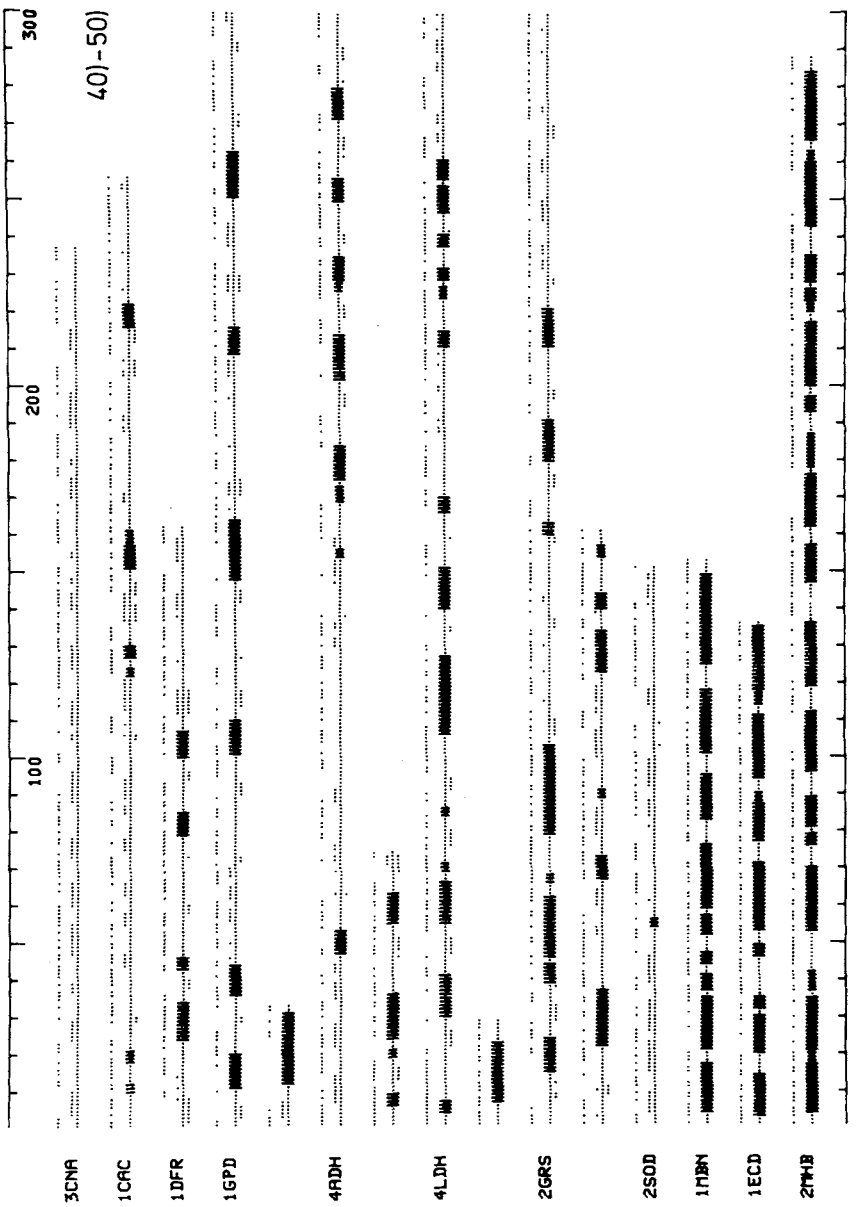



Fig. A1. (*continued from the previous page*)

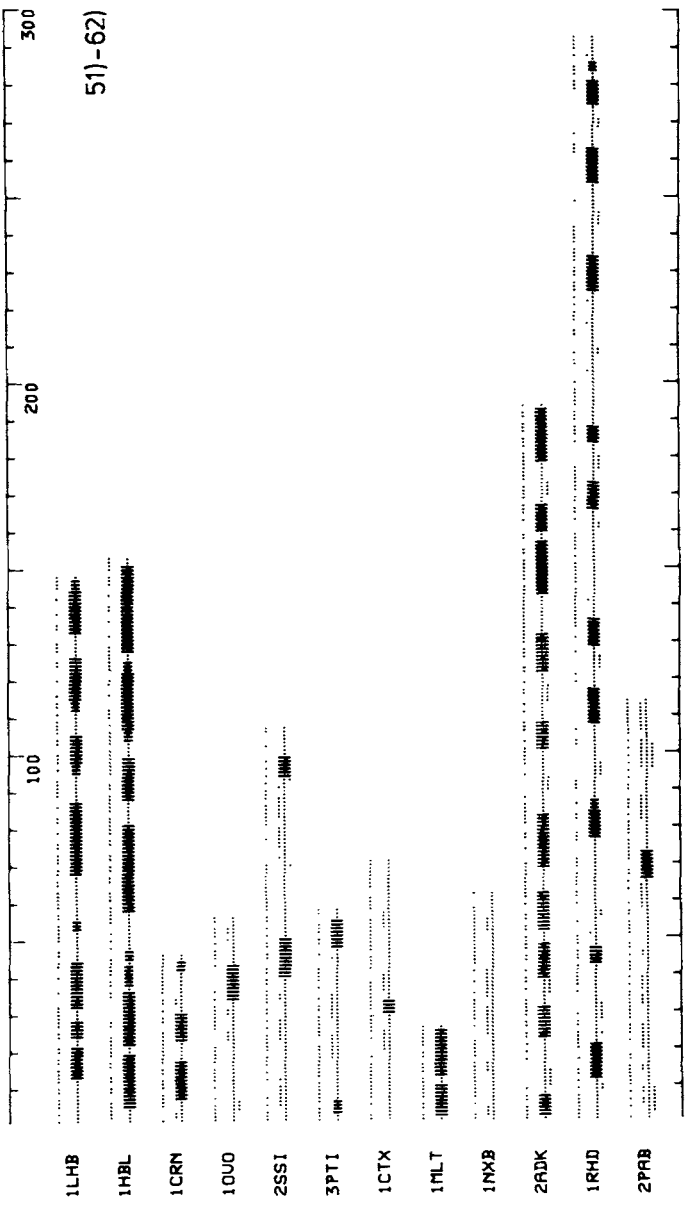



Fig. A1. (*continued from the previous page*)

For reasons of space it is impossible to cite the tremendous amount of work by the crystallographers on which this paper is based; references for each structure are in the Protein Data Bank. C. Oefner provided computer graphics software. The Deutsche Forschungsgemeinschaft gave financial support to the project "Protein Structure Theory."

## References

1. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. USA* **37,** 729–740.
2. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J Mol. Biol.* **112,** 535–542.
3. Kuntz, I. (1972) *J. Am. Chem. Soc.* **94,** 4009–4012.
4. Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1973) *Biochim. Biophys. Acta* **303,** 211–229.
5. Rose, G. D. & Seltzer, J. P. (1977) *J. Mol. Biol.* **113,** 153–164.
6. Chou, P. Y. & Fasman, G. D. (1977) *J. Mol. Biol.* **115,** 135–175.
7. Smith, J. A. & Pease, L. G. (1980) *CRC Crit. Rev. Biochem.* **8,** 315.
8. Richardson, J. S. (1981) *Adv. Prot. Chem.* **34,** 167–339.
9. Lifson, S. & Sander, C. (1980) *J. Mol. Biol.* **139,** 627–639.
10. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
11. Levitt, M. & Greer, J. (1977) *J. Mol. Biol.* **114,** 181–239.
12. Feldmann, R. J. (1976) *Atlas of Macromolecular Structure on Microfiche*, Tracor Jitco Inc., Rockville, Md.
13. Lifson, S., Hagler, A. T. & Dauber, P. (1979) *J. Am. Chem. Soc.* **101,** 5111–5121.
14. IUPAC-IUB Commission on Biochemical Nomenclature (1970) *J. Biol. Chem.* **245,** 6489–6497.
15. Rackovsky, S. & Scheraga, H. A. (1978) *Macromolecules* **11,** 1168–1174.
16. Shrake, A. & Rupley, J. A. (1973) *J. Mol. Biol.* **79,** 351–371.
17. Chothia, C. (1975) *Nature* **254,** 304–308.
18. Bode, W., Schwager, P. & Huber, R. (1975) Proceedings of the 10th FEBS Meeting, Vol. 40, Desnuelle P. & Michelson A. M., Eds., North-Holland, Amsterdam, pp. 3–20.
19. Finney, J. L. (1979) in *Water, A Comprehensive Treatise*, Vol. 6, Franks F., Ed., Plenum Press, New York, p. 93.
20. Brant, D. A., Miller, W. G, & Flory, P. J. (1967) *J. Mol. Biol.* **23,** 47–65.
21. Schellman, C. (1980) in *Protein Folding*, Jaenicke, R., Ed., Elsevier, Amsterdam, pp. 53–61.
22. Provencher, S. W. & Gloeckner, J. (1981) *Biochemistry* **20,** 33–37.
23. Williams, R. W. & Dunker, A. K. (1981) *J. Mol. Biol.* **152,** 783–813.
24. Hennessey, J. P. & Johnson, W. C. (1981) *Biochemistry* **20,** 1085–1094.
25. Deisenhofer, J. & Steigemann W. (1975) *Acta Crystallogr., Sect. B* **31,** 238–250.
26. Wuethrich, K. & Wagner, G. (1979) *J. Mol. Biol.* **130,** 1–18.
27. Timkovich, R. & Dickerson, R. E. (1976) *J. Biol. Chem.* **251,** 4033–4046.
28. Schulz, G. E., Elzinga, M., Marx, F. & Schirmer, R. H. (1974) *Nature* **250,** 120–123.
29. Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981) *Nature* **289,** 366–368.
30. Diamond, R. (1966) *Acta Crystallogr., Sect. A* **21,** 253–266.
31. Diamond, R. (1971) *Acta Crystallogr., Sect. A* **27,** 436–452.
32. Hendrickson, W. A. & Konnert, J. H. (1980) *Computing in Crystallography*, Diamond, R., Ramaseshan, S. & Venkatesan, K., Eds., Indian Academy of Sciences, Bangalore, pp. 13.01–13.23.
33. Dodson, E. J., Isaacs, N. W. & Rollett, J. S. (1976) *Acta Crystallogr., Sect. A* **32,** 311–315.
34. Jack, A. & Levitt, M. (1978) *Acta Crystallogr., Sect. A* **34,** 931–935.
35. Levitt, M. & Lifson, S. (1969) *J. Mol. Biol.* **46,** 269–279.
36. Agarwal, R. C. (1978) *Acta Crystallogr., Sect. A* **34,** 791–809.
37. Chambers, J. L. & Stroud, R. M. (1977) *Acta Crystallogr., Sect. B* **33,** 1824–1837.
38. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. (1977) *Acta Crystallogr., Sect. A* **33,** 800–804.

Received December 22, 1982
Accepted May 12, 1983