

**ILL Document Delivery**

REG-13641309

**NYUMSK****NLM -- W1 BI88N (GENCOLL); Film S04369 (MICROFORM)**

**Memorial Sloan-Kettering Cancer Center  
Medical Library Nathan Cummings Center (METRO #146)  
1275 York Avenue  
New York, NY 10021**

ATTN: SUBMITTED: 2005-11-25 15:33:07  
PHONE: 212-639-7441 PRINTED: 2005-11-28 14:47:09  
FAX: 646-422-2316 REQUEST NO.: REG-13641309  
E-MAIL: ill@mskcc.org SENT VIA: DOCLINE  
DOCLINE NO: 18645452

---

**REG Copy Journal BEFORE: 2005-12-01**

---

TITLE: BIOPOLYMERS  
PUBLISHER/PLACE: Wiley Interscience New York, NY :  
VOLUME/ISSUE/PAGES: 1983 Dec;22(12):2577-637 2577-637  
DATE: 1983  
AUTHOR OF ARTICLE: Kabsch W; Sander C  
TITLE OF ARTICLE: Dictionary of protein secondary structure:  
pattern  
ISSN: 0006-3525  
OTHER NUMBERS/LETTERS:  
NLM Unique ID: 0372525  
PubMed ID: 6667333  
SOURCE: PubMed  
MAX COST: \$25.00  
COPYRIGHT COMP.: Law  
CALL NUMBER: W1 BI88N (GENCOLL); Film S04369 (MICROFORM)  
NOTES: Please, we prefer as PDF or Ariel. Thank you! EFTS  
REQUESTER INFO: Gangi-Dino, Ri (Non [1864])  
DELIVERY: E-mail: ill@mskcc.org  
REPLY: Mail:

KEEP THIS RECEIPT TO RECONCILE WITH BILLING STATEMENT

For problems or questions, contact NLM at [http://wwwcf.nlm.nih.gov/ill/ill\\_web\\_form.cfm](http://wwwcf.nlm.nih.gov/ill/ill_web_form.cfm) or phone 301-496-5511.

Include LIBID and request number.

---

NOTE: THIS MATERIAL MAY BE PROTECTED BY COPYRIGHT LAW (TITLE 17, U.S. CODE)

# 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<sup>α</sup> positions and is positive for right-handed helices and negative for ideal twisted β-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

α-Helices and pleated β-sheets were predicted in 1951 by Linus Pauling and Robert Corey<sup>1</sup> 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<sup>2</sup> are often subjective and sometimes incomplete. Objective algorithms exist, e.g., for defining turns<sup>3-6</sup> (reviewed in Refs. 7, 8), β-sheets,<sup>9</sup> and solvent accessibility,<sup>10</sup> but only Levitt and Greer<sup>11</sup> have published an extensive compilation of automatic assignments of helices and sheets. Their ap-

Biopolymers, Vol. 22, 2577-2637 (1983)  
© 1983 John Wiley & Sons, Inc.

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

proach has the advantage of giving assignments when only backbone C<sup>α</sup> 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.<sup>12</sup> 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 φ,ψ angles or C<sup>α</sup> positions requires the adjustment of several parameters, e.g., four angles for a rectangle in the φ,ψ 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 α-helices, and repeating bridges define β-structure, in good agreement with intuitive assignments. All other occurrences of the basic patterns provide an interesting survey of 3<sub>10</sub>-helices, π-helices, single turns, and single β-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, α-helices and β-ladders, including common imperfections such as helical kinks and β-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 E (see text) of the H-bond as a function of the distance, *E* negative. Assume an H bond for *E* is allowed at the ideal length. This definition is general than the historical one.

"Hbond(i,j)=: |E| < *E* is less than -0.5 kJ/mol

Hydrogen bonds are well described by an interaction energy between the C,O

$$E = q_1 q_2 (1 - \cos \theta)$$

with  $q_1 = 0.42e$  and  $q_2 = 0.35e$ . The interatomic distance *E* is given in Å. The dimensional factor  $(1 - \cos \theta)$  is about -3 kcal/mol b for bifurcated H bonds between C=O of residues *i* and *j*, i.e., "Hbond(i,j)=:

Figure 1 illustrates

ments when only backbone C<sup>α</sup> atoms have been published for no more than microfiche.<sup>12</sup> We are thus most optimistic, and up-to-date compi-

as

notion of secondary structure for extracting structural features is a pattern-recognition process. This process should be as simple as the main types of secondary structure present or not in a continuous decision parameters. The C<sup>α</sup> positions requires the angles for a rectangle in the  $\varphi, \psi$  plane. In contrast, the presence or absence by a single decision parameter, we base our secondary structure using patterns: "n-turns" with an NH of residue  $i + n$ , where  $n$  can be residues not near each other. This essentially exhaust all backbone turns define  $\alpha$ -helices, and regular agreement with intuitive basic patterns provide an intercalate turns, and single  $\beta$ -bridges. As strip maps of secondary structure with the amino acid sequence

The computer program DSSP written in standard PASCAL will be available from the Chemistry Dept., Brookhaven National Laboratory. Publication of an update of this structures are solved.

NS

form a hierarchy: first H bonds; and, based on them,  $\alpha$ -helices. Connections such as helical kinks and hairpins, H bonds, chirality, SS bonds, etc. are defined independently and resolved by defining a secondary structure to each residue. For brevity we use equations. For example,

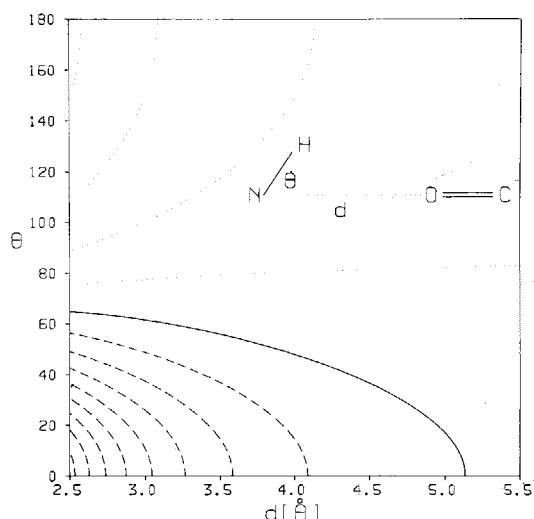


Fig. 1. H bond between peptide units is described here by the dominant electrostatic part  $E$  (see text) of the H-bond energy, drawn in contours of constant  $E$  at 0.5 kcal/mol intervals as a function of the distance,  $d$ , and the alignment angle  $\theta$ . Dotted lines,  $E$  positive or zero; broken lines,  $E$  negative. An ideal H bond has  $d = 2.9 \text{ \AA}$ ,  $\theta = 0$ , and  $E = -3.0 \text{ kcal/mol}$ . We assume an H bond for  $E$  up to  $-0.5 \text{ kcal/mol}$  (solid line). Thus, misalignment of up to  $63^\circ$  is allowed at the ideal length; an N-O distance of up to  $d = 5.2 \text{ \AA}$  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)=: [ $E < -0.5 \text{ kcal/mole}$ ]" means: there is an H bond (i,j) if  $E$  is less than  $-0.5 \text{ kcal/mol}$ .

## Hydrogen-Bonded Structure

### Hydrogen Bonds

Hydrogen bonds in proteins have little wave-function overlap and are well described by an electrostatic model.<sup>13</sup> 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.,

$$E = q_1 q_2 (1/r(ON) + 1/r(CH) - 1/r(OH) - 1/r(CN)) * f$$

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  $-3 \text{ kcal/mol}$  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)=: [ $E < -0.5 \text{ kcal/mole}$ ]."

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

'>'      '3'      '3'      '<'	<u>3-turn</u> notation
-N-C-C--N-C-C--N-C-C--N-C-C- H    O    H    O    H    O    O >-----<	residues H-bond
'>'      '4'      '4'      '4'      '<'	<u>4-turn</u> notation
-N-C-C--N-C-C--N-C-C--N-C-C--N-C-C- H    O    H    O    H    O    O >-----<	residues H-bond
'>'      '5'      '5'      '5'      '5'      '<'	<u>5-turn</u> notation
-N-C-C--N-C-C--N-C-C--N-C-C--N-C-C--N-C-C- H    O    H    O    H    O    H    O >-----<	residues H-bond
'x'	<u>parallel bridge</u>
-N-C-C--N-C-C--N-C-C- H    O    H    O    H    O \ .    . / \ . \ / . . \ / . H    O    H    O    H    O	notation residues H-bonds (\ and /, or .)
'x'	residues notation
'x'	<u>antiparallel bridge</u>
-N-C-C--N-C-C--N-C-C- H    O    H    O    H    O . !    ! . . !    ! . . !    ! . O    H    O    H    O    H -C-C-N--C-C-N--C-C-N-	notation residues H-bonds (! or .)
'x'	residues notation
's'	<u>bend</u> notation
i-2    i-1    i    i+1    i+2 -C-C--N-C-C--N-C-C--N-C-C- O    H    O    H    O    H !----->!----->	residues direction change more than 70 degrees

Fig. 2. Elementary patterns used in structure definition.

more complicated angle. There is no border between t<sub>1</sub> nates at short distance and t<sub>2</sub> nates at larger distance. The energy as a function is empirically tailored by trial and error, secondary structure at a distance up to 2.2 Å, a misalignment of and Greer<sup>11</sup> (1.8 Å) over coordinate assignments. We “polar interaction

The basic turn p  
assign an  $n$ -turn at  
i.e., "n-turn(i) =:

Two nonoverlap  $j - 1, j, j + 1$ , form which of two basic residues  $i$  and  $j$  if particular,

Parallel Brid

## Antiparallel Brid

Parallel bridges are  
by upper-case letter

'<'	<u>3-turn</u> notation
--N-C-C-	residues
H O	H-bond
--<	
'<'	<u>4-turn</u> notation
--N-C-C-	residues
H O	H-bond
--<	
'<'	<u>5-turn</u> notation
--N-C-C-	residues
H O	H-bond
--<	

parallel bridge  
notation  
residues

H-bonds  
(\ and /, or ..)

residues  
notation

antiparallel bridge  
notation  
residues

H-bonds  
(! or ..)

residues  
notation

bend  
notation

residues

direction change  
more than 70  
degrees

chirality  
four C(alpha) atoms  
define the dihedral  
angle alpha

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<sup>11</sup> (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  $i$ ,

$$4\text{-helix}(i, i+3) =: [4\text{-turn}(i-1) \text{ and } 4\text{-turn}(i)]$$

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  $i+5$ :

$$3\text{-helix}(i, i+2) =: [3\text{-turn}(i-1) \text{ and } 3\text{-turn}(i)]$$

$$5\text{-helix}(i, i+5) =: [5\text{-turn}(i-1) \text{ and } 5\text{-turn}(i)]$$

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  $\rangle\rangle 3 \langle\langle$ , a 4-helix by  $\rangle\rangle 44 \langle\langle$ , and a 5-helix by  $\rangle\rangle 555 \langle\langle$ . 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	IIII

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

### *Cooperative H-Bond Patterns: $\beta$ -Ladders and $\beta$ -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

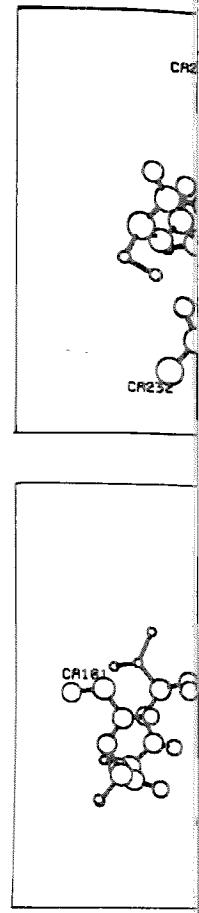


Fig. 3. Stereoview  
(a) 3-Helix Gly232-L  
as the H-bond pattern

3-Helices are not unique/mol and the C=Gly181-Lys188 from Table AIII, it appears

5-Helices are extremely rare by PLUTO (Sam M. Smith) atoms with  $\frac{1}{4}$  their value with twice these values

tern: Helices

secutive  $n$ -turns. For example, turns  $i$  to  $i + 3$ , requires 4-turns at

$-1)$  and 4-turn( $i$ )]

( $i, i + 4$ ). Note that nothing is required between  $i + 1$  and  $i + 2$ . Similarly, two 3-turns require a 3-turn of minimal length 3 from residue  $i$  to  $i + 3$ ; a 5-turn of length 5 from residue  $i$  to  $i + 5$ :

$-1)$  and 3-turn( $i$ )]

$-1)$  and 5-turn( $i$ )]

Minimal helices. Conventionally, there is the  $\alpha$ -helix,  $\beta$ -helix, and  $\pi$ -helix. In Table AIII, the H-bond patterns are given as  $\rangle\rangle\rangle 3 \langle\langle$ , a 4-helix by  $\rangle\rangle 4 4 \langle\langle$ , and  $\rangle\rangle\rangle 5 \langle\langle$ , a 5-helix by  $\rangle\rangle 5 5 \langle\langle$ , the residues bracketed by H

$\rangle\rangle 5 5 \langle\langle$

$\rangle\rangle 4 \langle\langle$

IHH IHH

at each end than they would be acceptable examples of a 3-helix and a 5-helix

### $\beta$ -Ladders and $\beta$ -Sheets

ive bridges of identical type connected by shared residues

b,c,... is for parallel, A,B,C... sequence, the first ladder is named "A". Ladders and  $\beta$ -sheets are also given letter names (e.g., "B"). Ladder names restart at "a" or "A." The  $\beta$ -SHEET is denoted by the sheet name and the strand number in which it participates (at most 2 strands). Residues in single bridges (ladder) are labeled "D", ladder residues "E" (extended). Residues in  $\beta$ -strands. The  $\beta$ -sheet notation

### Irregularities

in that not all possible H bonds

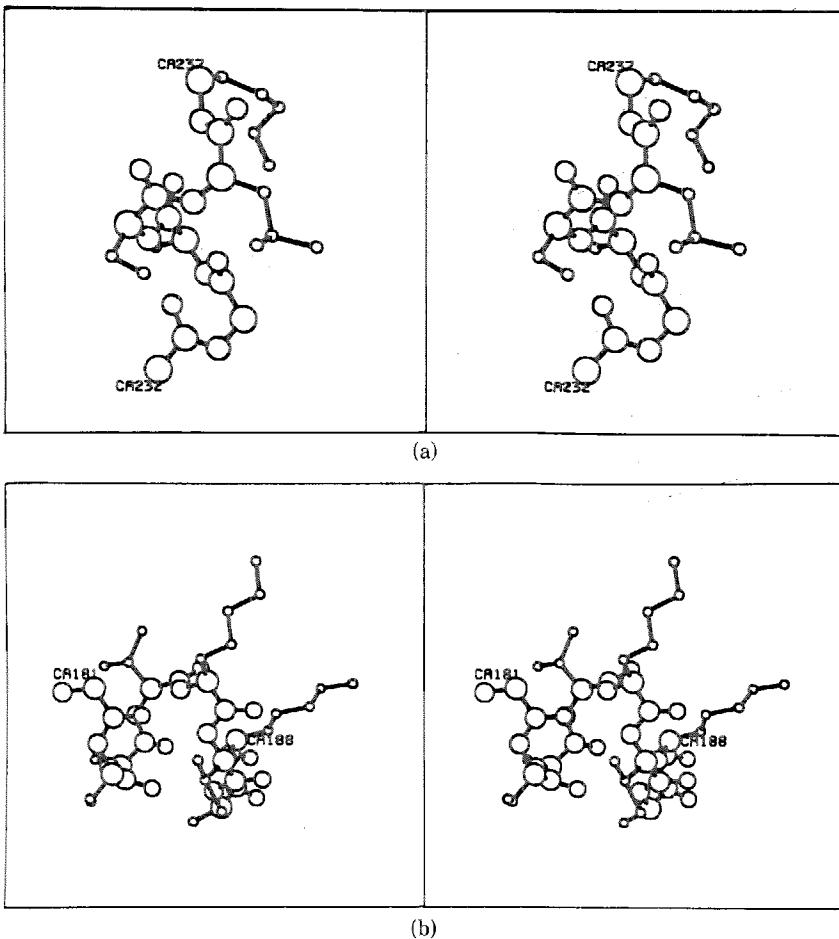


Fig. 3. Stereoviews of secondary structure: (a) 3-helix ( $3_{10}$ -helix) and (b) 5-helix ( $\pi$ -helix). (a) 3-Helix Gly232-Lys237 from triose phosphate isomerase (1TIM). In Table AIII, it appears as the H-bond pattern

3-TURN SUMMARY SEQUENCE	$\rangle\rangle\rangle \langle\langle$ GGGG 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  $30^\circ$ . (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 SEQUENCE	$\rangle\rangle\rangle 5 \langle\langle$ GSAYKVAK
--------------------	--

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  $\frac{1}{4}$  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.

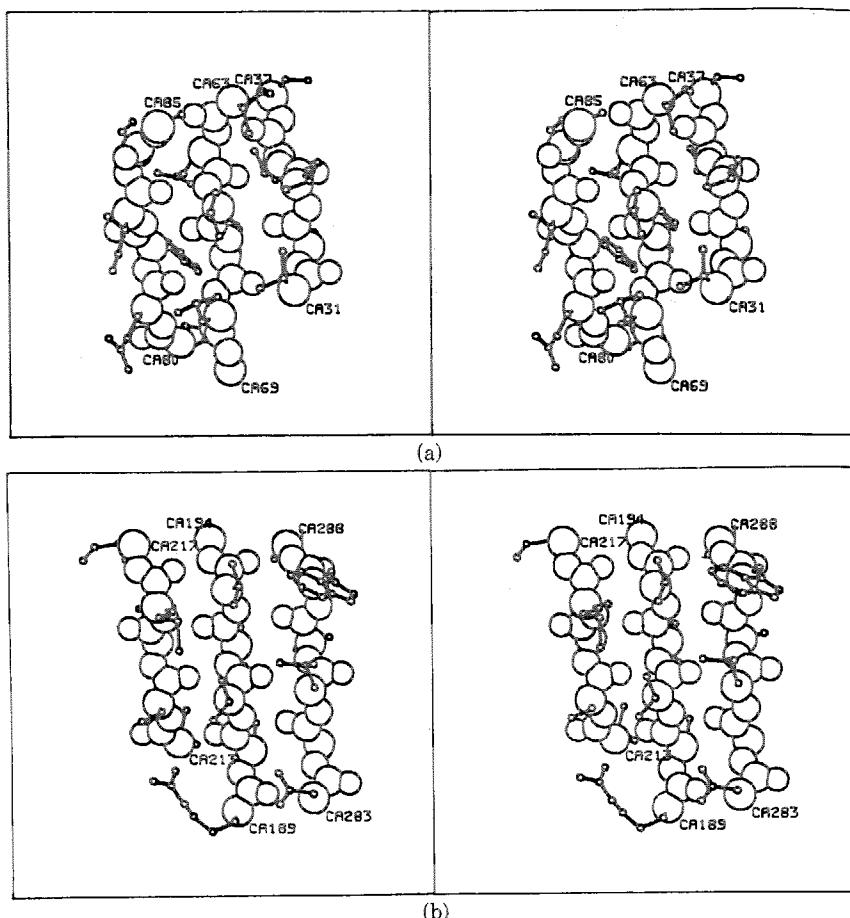


Fig. 4. Stereoviews of secondary structure: (a) antiparallel and parallel  $\beta$ -sheets with two ladders (three strands) each. (a) Two connected antiparallel  $\beta$ -ladders from trypsin (1PTN). The three participating strands are Val16(31)–Ser20(37), Ile46(63)–Gly51(69), and Glue62(80)–Ala67(85), 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.....	EEE E .....
BRIDGE2.....	111 .....		
BRIDGE1.....	kkkk .....	111.....	kkkk .....
SEQUENCE.....	RSVIVG .....	TSLMI.....	DCLLWA .....

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

are formed. This two overlapping m into one helix:

```

    ) ) 4 4 < (
+      ) ) 4 4 <
=      ) > 4 ) X ( 4 (
=      HHHHHHHH
      ) ) ) ) X < < (

```

even though the t  
perfect seven- or e  
ciated with a kink

For  $\beta$ -structure, two (perfect) ladder extra residue on one strand. This defines frequent lattice faults her main types. In ladder (lines BRID ladders are marked

Bends are regions at the central residue direction of the first curvature is identical that of Rackovsky and of at least  $70^\circ$ . The dimensional traces.

Bend(i) =: [ang] and assign “S” for a

We define chirality (Fig. 2)

but report only the s if  $-180^\circ < \alpha < 0^\circ$ .  $\beta$ -ladders negative, in thermolysin. The

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:

$\rangle \rangle 4\ 4\ \langle \langle$	$\rangle \rangle 4\ 4\ \langle \langle$
$+$	$\rangle \rangle 4\ 4\ \langle \langle$
$=$	$\rangle \rangle 4\ X\ \langle\ 4\ \langle\ \text{irregular}$
$=$	$\rangle \rangle 4\ 4\ X\ X\ 4\ 4\ \langle\ \langle\ \text{irregular}$
$\rangle \rangle \rangle\ X\ \langle\ \langle\ \langle\ \text{perfect}$	$\rangle \rangle \rangle\ X\ X\ \langle\ \langle\ \langle\ \text{perfect}$

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<sup>8</sup> 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<sup>5</sup> but slightly different from that of Rackovsky and Scheraga.<sup>15</sup> For a bend at  $i$ , we require a curvature of at least  $70^\circ$ . The cutoff value was chosen by visual inspection of three-dimensional traces. With  $C^\alpha$  the position vector of  $C^\alpha$ , we define

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

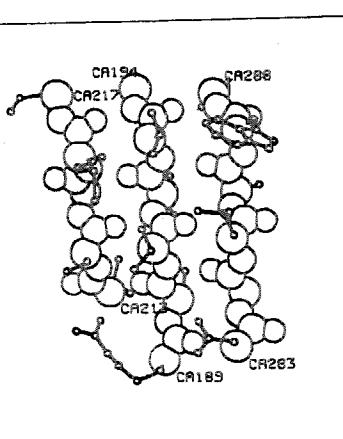
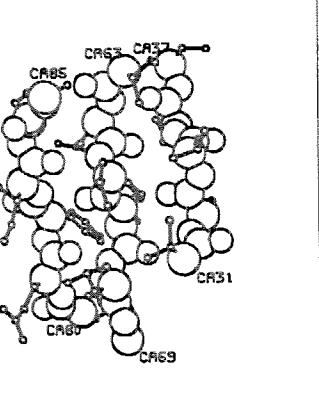
and assign "S" for a bend at residue  $i$ .

#### Chirality

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

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

but report only the sign of  $\alpha$  in Table AIII: "+" if  $0^\circ < \alpha < 180^\circ$  and "-" if  $-180^\circ < \alpha < 0^\circ$ . 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.



tiparallel and parallel  $\beta$ -sheets with two parallel  $\beta$ -ladders from trypsin (1PTN). Ser20(37), Ile46(63)-Gly51(69), and sequential residue number from Table residue identifier. The corresponding H-

.... CCCC ..... CCCC.....  
.... NNNN .....  
.... KKK ..... NNNN.....  
.... IQVRLG ..... EQFISA ..

both ladders belong to sheet C. (b) Two y177(194), Thr196(213)-Ile200(217), se (2GRS). The corresponding H-bond

.... EEE ..... EEEE ....  
.... .....  
.... 111 ..... kkkk .....  
.... TSLMI ..... DCLLWA ..

ee strands are part of sheet E.

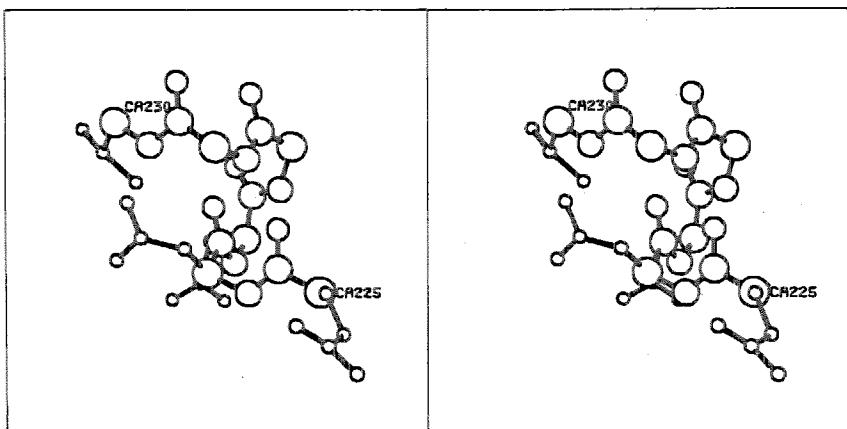


Fig. 5. Stereoviews of secondary structure: illustration of chirality. This short left-handed  $\alpha$ -helix, Gln225–Val230 from thermolysin (2TLN) is the only one known to us. In Table AIII (note that chirality is entered at the second residue of each quartet) it appears as:

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 lower-case 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.

To make contact w facilitate comparison v structure in a single li are eliminated in this l i.e., when several symb example, a helix is als priority. Pieces of 3- overlaps, are labeled " curvature not in H-bo

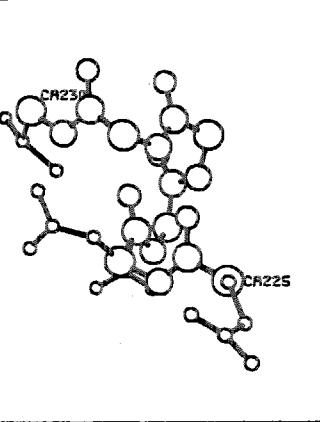
Physically, we are in contact with the prote

Geometrically, a ve of water is the surface o in touching contact w Richards<sup>10</sup> water sph surface associated with that accommodates jus protein exterior, howev molecules in the first h

Mathematically, one f over all points  $x$  on t around atom  $i$ .  $f = 1$  in tact with atom  $i$ ) does n  $f = 0$ .

Algorithmically, we 20, 80, 320, or more app are the triangle centers is generated starting fr provides each triangle into projects the three new next level of recursion, of certain viruses and domes. Hence, we cal is similar to the algorith than z-layer integratio

With 320 integration within 1 Å<sup>2</sup>; with 80 po values agree with those numbers given here ar 1.40 for O, 1.65 for N, 1



on of chirality. This short left-handed e only one known to us. In Table AIII each quartet) it appears as:

---  
))44((  
HHHH  
QDNGGV

$S^\gamma$  atoms of two Cys residues, BOND records, as they can be (primary structure). For the nce of less than 3.0 Å can also given names a,b,c . . . , and the in the line SEQUENCE in Table ence either as C or as a lower-

e bond length (distance C'-N) nted as a break residue. Thus, ptide bond, missing density in s. The residues for which there d sequentially, including break often agree with the authors' ence homology or those with se, inspection of the amino acid ambigous identification of a

### 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<sup>10</sup> 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 Shrike and Rupley<sup>16</sup> and conceptually simpler than z-layer integration.

With 320 integration points, the surface area of a residue is accurate to within 1 Å<sup>2</sup>; with 80 points, to within 4 Å<sup>2</sup>. For myoglobin, the numerical values agree with those of Lee and Richards,<sup>10</sup> 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<sup>α</sup>, 1.76 for C of CO in the backbone, 1.80 for

all side-chain atoms,<sup>17</sup> 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  $30 \text{ \AA}^3$  and area in  $\text{\AA}^2$ , 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,<sup>2</sup> 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: sulphydryl proteinases actininidin/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 A1).

### 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<sub>7</sub> 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.

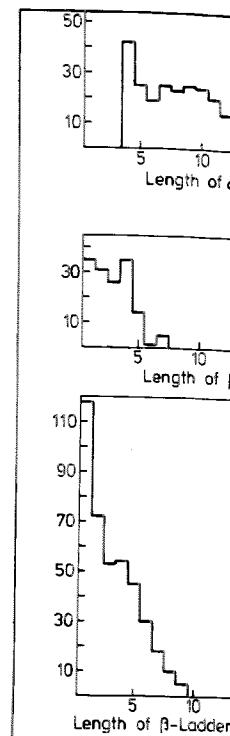


Fig. 6. The common feature of gradual fall-off: larger sizes of length of  $\beta$ -ladders (strand pairs) than the ladders in which it is tangential in shape. The number of the width of the sheet. These take part in two ladders, each of strands minus one. In general, on each side and the width of sheets consisting of a single ladder. Number of H-bonds of type  $n$  are heavily favored. The authors find that H bonds ( $i, i+2$ ) are

Using our H bond definition to the main interstrand peak in Fig. 6(e)]. The extended conformational groups of adjacent peptide segments in stabilizing the  $\beta$ -strands originate as each of  $\beta$ -strands, we typically find 2) H bonds per 100 residues part of a tight turn.

er molecule following observed in Ref. 19).

mber,  $W$ , of water molecules estimated from the surface area by

$$\frac{W}{e^{2/3}} \approx \frac{\text{Area}}{10}$$

olume of the monolayer, which, number of molecules in the monolayer in  $\text{Å}^3$  and area in  $\text{Å}^2$ , the conversion exposure differs for a monomer andence of all monomers in the data substrates, ligands, heme, etc.). ent exposure of the protein.

## SCUSSION

### teins

sets in the Protein Data Bank,<sup>2</sup> inates and a known amino acid had more than a 50% sequence uivalent positions, the one with more secondary structure was chosen of these pairs: serine dehydrogenases 4LHD=1LDX 1CAB by 60%; chymotrypsin of the following pairs: sulfhy-8PAP by 47%; immunoglobulins 2 155C=1C2C by 43%; chymoastase/trypsin 1EST=1PTN by =1APP by 43%;  $\alpha/\beta$  subunit of %. The final 62 data sets thus ein structures, except those not able AI).

### ecture

imply classified by the number, by  $n$  of  $(i,i+n) = (\text{CO}(i),\text{NH}(i+n))$  of H bonds for each  $n$ . isallowed. A hydrogen bond  $(i,i+2)$  five peptide units for certain  $\phi,\psi$  formation is known as C<sub>7</sub> and leads a  $\beta$ -strand if it repeats. When n is sometimes called a  $\gamma$ -turn.

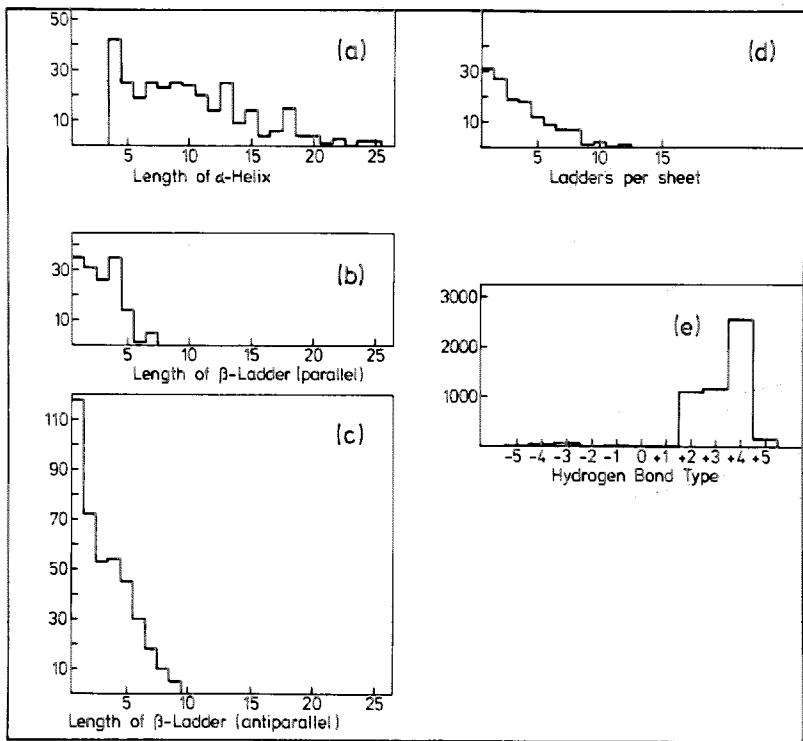


Fig. 6. The common feature of the size distribution of secondary structure segments is the gradual fall-off: larger sizes are less probable than smaller ones. Note that we give (b,c) the length of  $\beta$ -ladders (strand pairs) rather than the length of  $\beta$ -strands. A strand is often longer than 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  $(\text{CO}(i),\text{NH}(i+n))$ . 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<sup>20</sup> to be essential in stabilizing the C<sub>7</sub> conformation in solution. We speculate that  $\beta$ -strands originate as extended C<sub>7</sub> 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  $(i, i - 2), (i, i - 3) \dots (i, i - 5)$  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<sup>21</sup> and Richardson.<sup>8</sup> 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<sup>22</sup>). In particular, we suggest that the distinction between parallel and antiparallel  $\beta$ -structure<sup>23,24</sup> 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<sup>11</sup> 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<sup>25</sup> 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  $2.6 \times 10^{-5} \text{ min}^{-1}$  measured by nmr.<sup>26</sup> Three others of type  $(i, i + 3)$ , with  $E = -1.3, -1.7, -0.9$ , form the well-known<sup>8</sup> 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<sub>7</sub> 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.<sup>11</sup>

For the *medium-resolution* structure of cytochrome c550, Timkovich and Dickerson<sup>27</sup> use a conservative interpretation of hydrogen bonds and

Comparison of Secondary S

Structure <sup>a</sup>	Original Authors (AU)
3PTI	
G1	— <sup>b</sup>
E1	16-25
E2	28-36
E3	— <sup>b</sup>
H1	47-56
155C	
H1	6-11
G1	11-13
E1	—
E2	—
E3	—
H1	—
H2	56-63
H3	73-79
H4	—
H5	107-118
2ADK	
H1	1-8
E1	10-14
H2	23-30
E2	35-38
H3	41-48
H4	53-62
H5	69-84
E3	90-94
H6	100-107
E4	114-118
H7	123-133
H8	144-158
H9	160-164
E5	169-173
H10	179-194

<sup>a</sup> H =  $\alpha$ -helix, G = 3<sub>10</sub>-helix resolution, Diamond real-space solution, Diamond model building diagram of Ref. 27. 2ADK =

<sup>b</sup> Serious discrepancy (segregates).

give a minimal set of 41 these, except Ala115(CO) we see the helix end with We assign an additional of a bifurcated H bond, ginal, with  $E > -1.0 \text{ kca}$

ed as turns or helices. Most  $(i,i+5)$  are part of a bridge or ladder.  $\dots(i,i+5)$  are also rare. There is between the backbone oxygen

usly believed, although they are H bonds.  $\alpha$ -Helices are rarely bifurcated, i.e.,  $(i,i+4)$  and of  $\alpha$ -helices often are overwound, and, ending in a 5-turn. Some generalized by Schellman<sup>21</sup> and es ( $\pi$ -helices)—see Fig. 3. bonds in Table AI may be useful in (CD, laser Raman) of the per-algorithm of Provencher and that the distinction between par-e reference spectra will improve

### Secondary Structure Assignments

ts can only be as accurate as the using this dictionary, it is therefore resolution and refinement of each ordinate data sets range from resolution, where individual side chains es at a resolution just sufficient compare our assignments with tt and Greer<sup>11</sup> for three proteins

of trypsin inhibitor (3PTI), H bond when the N-O distance backbone-backbone H bonds. Of NH), which has  $d = 3.1$ ; instead,  $E = -2.2$ . In addition, we assign energy cutoff in our definition. ; with  $E = -2.0$ , consistent with  $10^{-5}$  min<sup>-1</sup> measured by nmr.<sup>26</sup>  $-1.3, -1.7, -0.9$ , form the well-H bond, Asn24(CO)-Leu29(NH),  $i,i+2$ , characteristic of the C<sub>7</sub> longer one ( $E = -1.8$ ) in a  $\gamma$ -turn found by us lead to identification secondary structure not cited by the au-er.<sup>11</sup>

of cytochrome c550, Timkovich et al. report retention of hydrogen bonds and

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

Structure <sup>a</sup>	Original Authors (AU)	Levitt & Greer (LG)	This Work (KS)	
3PTI				
G1	— <sup>b</sup>	2-7	3-6	Clearly 3 <sub>10</sub> ; LG have $\alpha$
E1	16-25	14-25	18-24	
E2	28-36	28-37	29-35	
E3	— <sup>b</sup>	43-46	45-45	$\beta$ -Bridge, 2 H bonds
H1	47-56	47-55	48-55	
155C				
H1	6-11	4-16 <sup>b</sup>	6-12	4-Turn 13-16
G1	11-13	—	11-13	Overlaps with H1
E1	—	17-23 <sup>b</sup>	19-20	AU have 2 H bonds; KS, 4
E2	—	26-31 <sup>b</sup>	—	Discontinuity at Asp28-Ile29
E3	—	33-39	35-37	AU have 2 H bonds; KS have 4
H1	—	40-44 <sup>b</sup>	—	KS have 3-Turn
H2	56-63	55-65	56-64	
H3	73-79	71-80	73-80	
H4	—	81-90 <sup>b</sup>	—	
H5	107-118	106-118	107-117	Pro at 82, 84; possible helix
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 <sup>b</sup>	122-132	$\alpha$ -Helix ends in 3-turn
H8	144-158	141-157 <sup>b</sup>	143-157	No $(i,i+4)$ H bond at Asp 141
H9	160-164	159-166	160-167 <sup>b</sup>	Two weak H bonds at 167,168
E5	169-173	169-175	170-173	
H10	179-194	179-192	179-193	

<sup>a</sup> H =  $\alpha$ -helix, G = 3<sub>10</sub>-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).

<sup>b</sup> 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  $(i,i+2)$

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<sup>11</sup> 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 17-28), 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<sup>28</sup>), all secondary structure assignments (ours, the original authors',<sup>28</sup> and Levitt and Greer's<sup>11</sup>) 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<sup>8</sup> 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,<sup>11</sup> and for lower-resolution data, they are conservative compared with both Levitt and Greer's program and Richardson's<sup>8</sup> 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 shorter. There are an integral number of possible exceptions to  $\alpha/\beta/\alpha$  folding units that protein statistics approach than short ones. The maximum size of

The structure of a helix, shows that variations in protein conformation will be used in prediction methods.

### APPENDIX:

Proteins are ordered by their maps (Fig. A1) by their number of H bonds per 100 residues. This can be compared with the number of water molecules per residue. The entire data set and the data set 1INS has two given is that of the A-chain unique molecule of the quality of the coordinates. In case of doubt, consider building to guide coordinates = Hendrickson-Konnerup and Levitt (Ref. 34); D = difference Fourier; RL = restrained least squares = energy minimization; HK; DD = DS and D2; 36) and DO; DH = D2 and DO; CS = constrained space and energy minimization; CORELS (Ref. 38).

0, Lys 53, and Lys88; two are  $(i,i+1)$  or are  $(i,i-3)$  and  $(i,i-6)$  in the involved in forming the heme pocket (0) at a helix end and Met103(NH) have a meaningful structural interaction assignments are consistent with the additional short parallel bulged due to two additional weak H bonds. more secondary structure (Table sheet 17-23/33-39 (probably too parallel to 17-23), a helix 40-44 (which has only two of the seven take a helix in a C<sup>a</sup> chain tracing and higher resolution).

The structure of adenylate kinase domains (ours, the original authors'<sup>28</sup>) Other lower-resolution coordinate pending on the quality of the H

our H-bond energy cutoff, chosen to errors in lower-resolution data, than conservative assignments in

All these have a physical meaning energy and nearly all have an inter- y structure; and, most importantly, not give rise to spurious secondary

ain for some lower-resolution data. GM, and 1ABP, Richardson<sup>8</sup> sees AIII, do appear as uncurved (non- bonded bridges between them. At 2.5-Å resolution with tentative that more H bonds will form in the

-bonded secondary structure are off in the H-bonding energy. For ments are more accurate than those olution data, they are conservative s program and Richardson's<sup>8</sup> visual

#### Structure Size

ture cooperativity? Are there any e segments? The length distributionally 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,<sup>29</sup> 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<sup>2</sup>. 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

ZAH	ZBA	Z3P	EXPO	M	LEN	RES	RF	Z ALPHA HELICAL AND 4-TURN HYDROGEN BONDS
								Z BETA ANTI-PARALLEL HYDROGEN BONDS
								Z BETA PARALLEL HYDROGEN BONDS
								WATER EXPOSURE
								MULTIPLICITY OF DATA SET
								NUMBER OF RESIDUES
								RESOLUTION
								REFINEMENT
								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) 1A3P L-ARABINOSE-BINDING PROTEIN
electron transfer								
7	13	2	490		85	2.0	DF	3) 1X1P OXIDIZED HIGH POTENTIAL IRON PROTEIN (HPIP).
electron transport								
19	15	7	566		85	2.8	D2	4) 2B5G 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	DI	8) 155C CYTOCHROME C550
38	2	0	682		82	2.0	GS	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) 3FKN FLAVODOXIN (OXIDIZED)
7	17	0	376		54	1.5	LS	13) 2RXN RUBREDOXIN (OXIDIZED, FE(III))
10	17	7	645		125	2.7	NH	14) 1AZU AZURIN
2	21	10	513		99	1.6	DH	15) 1PCY PLASTOCYANIN
hormones								
44	0	0	343		36	1.4	RL	16) 1PET 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) 1IRN RIBONUCLEASE-S
hydrolases, proteinases								
28	5	9	1209		308	2.0	--	24) 1CPA CARBOXYPEPTIDASE A
3	13	3	1333		324	2.5	NH	25) 1APR ACID PROTRASE (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	NH	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) 1SST 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) 3PAP 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	NK	38) 3PGM PHOSPHOGLYCERATE MUTASE (DE-PHOSPHO)
35	1	15	1026	2	246	2.5	DO	39) 1TIM TRIOSE PHOSPHATE ISOMERASE
lectrin (agglutinin)								
2	35	0	1125		237	2.4	MD	40) 3CNA CONCANAVALIN A
lyase, carbon-oxygen								
4	20	5	1273		256	2.0	DI	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	DI	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	NK	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 (ERYTHROCYTIN DEOXY)
58	0	0	1415		287	2.0	DD	50) 2MBH HEMOGLOBIN (HORSE, AQUO MET)
47	0	0	864		148	2.0	DI	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	NK	53) 1CKN 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) 2SS1 STREPTOMYCES SUBTILISIN INHIBITOR
12	17	0	412		58	1.5	D2	56) 3PTI TRYPSIN INHIBITOR
toxins								
3	17	0	511		71	2.8	--	57) 1OTX ALPHA COBRA TOXIN
71	0	0	222	2	26	2.0	NK	58) 1MLT MELLITIN
0	29	0	406		62	1.4	NK	59) 1NXB NEUROTOXIN B (PROBABLY IDENTICAL TO ERABUTOXIN B)
transfases								
47	0	10	1251		194	3.0	--	60) 2ADK ADENYLATE KINASE
20	2	10	1456		293	2.5	DI	61) 1RHD RHODANESE
transport								
7	33	7	652		114	1.8	DO	62) 2PAB PREALBUMIN (HUMAN PLASMA)

First line: running number  
[func]  
SHEET... One-part  
BRIDGE2... One-part  
BRIDGE1... "A,"  
"a,"  
Ladd  
A  $\beta$ -s  
two li  
twice  
easily  
recon  
their  
"+"  
Chira  
i - 1  
 $\beta$ -stra  
"S" =  
BEND...  
5-TURN...  
4-TURN...  
3-TURN...  
"X"  
"S,"  
SUMMARY...  
Struc  
"H" =  
"B" =  
"E" -  
"G" =  
"T" =  
"T" =  
"S" =  
In cas  
this li  
EXPOSURE...  
Solve  
contact  
moled  
10 Å  
SEQUENCE...  
Amin  
"a,"  
= cha  
includ  
coord  
Thus  
print

## ular Proteins

HELICAL AND 4-TURN HYDROGEN BONDS  
NTIPARALLEL HYDROGEN BONDS  
PARALLEL HYDROGEN BONDS

POSURE  
CITY OF DATA SET  
F RESIDUES  
ON  
NT  
IDENTIFIER, NAME

---

CIUM-BINDING PARVALBUMIN B  
RIBONOSE-BINDING PROTEIN

DIZED HIGH POTENTIAL IRON PROTEIN (HIFIP),

OCHROME B5 (OXIDIZED)  
OCHROME B562 (E. COLI, OXIDIZED)  
OCHROME C (OXIDIZED)  
OCHROME C2 (FERRI)  
OCHROME C550  
OCHROME C551 (OXIDIZED)  
REDOXIN (PEPTOCOCCUS AEROGENES)  
REDOXIN (SPIRULINA PLATENSIS)  
VODOXIN (OXIDIZED)  
REDOXIN (OXIDIZED, FE(III))  
RIN  
STOCYANIN

AN PANCREATIC POLYPEPTIDE  
CACON (PH 6-7)  
ULIN (A AND B CHAIN)

SPHOLIPASE A2

ZOZYME (BACTERIOPHAGE T4)  
ZOZYME (HEN EGG WHITE, TRICLINIC)

PHYLOCOCCAL NUCLEASE (COMPLEX)  
IONUCLEASE-S

PROXYPEPTIDASE A  
D PROTEASE (RHIZOPUS CHINENSIS)  
D PROTINASE (PENICILLIOPEPSIN, FUNGUS)

EMOLYSIN  
IMA CHYMOTRYPSIN A  
PA LYtic PROTEASE  
TA-TRYPSIN (NATIVE AT PH 8)  
DEINASE A FROM STREPTOMYCES GRISEUS (SGPA)

STYLISIN BPN'  
SYL-Elastase  
TINIDIN  
PAIN

IBDA IMMUNOGLOBULIN FAB  
NCE-JONES IMMUNOGLOBULIN (VARIABLE PORTION)

OSPHOGLYCERATE MUTASE (DE-PHOSPHO)  
ISOSE PHOSPHATE ISOMERASE

NCANALVIN A

RBNIC ANHYDRASE FORM C

HYDROFOLATE REDUCTASE (COMPLEX)  
GYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE  
O-LIVER ALCOHOL DEHYDROGENASE  
CTATE DEHYDROGENASE, APO ENZYME M4  
UTATHIONE REDUCTASE  
,ZN SUPEROXIDE DISMUTASE

OGLOBIN (FERRIC IRON - METMYOGLOBIN)  
MOGLOBIN (ERYTHROCUORIN DEOXY)  
MOGLOBIN (HORSE, AQUO MET)  
MOGLOBIN(MET)-CYANIDE V (SEA LAMPREY)  
GHEMOGLOBIN (ACETATE,MET) (YELLOW LUPIN)

AMBIN

OMUCOID THIRD DOMAIN  
REPTONYCES SUBTILISIN INHIBITOR  
LYPSIN INHIBITOR

PHA COBRA TOXIN  
LITTTIN  
UROTOXIN B (PROBABLY IDENTICAL TO ERABUTOXIN B)

ENYLETIC KINASE  
IODANESE

REALBUMIN (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>i</i> participates.
BRIDGE2...	One-character name of $\beta$ -ladders in which residue <i>i</i> participates, "A," "B," "C" ... = antiparallel,
BRIDGE1...	"a," "b," "c" ... = parallel.
CHIRALITY	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. "+" or "-" Chirality at residue <i>i</i> is the sign of the dihedral angle defined by $C^{\alpha}$ <i>i</i> to <i>i</i> + 2. Thus, a right-handed $\alpha$ -helix has "+," an ideal twisted $\beta$ -strand "-."
BEND...	"S" = five-residue bend centered at residue <i>i</i> . Hydrogen-bonding pattern for turns and helices: "/" = backbone CO of this residue makes H bond ( <i>i</i> , <i>i</i> + <i>n</i> ) "\\" = backbone NH of this residue makes H bond ( <i>i</i> - <i>n</i> , <i>i</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 <sub>10</sub> -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>i</i> . The scale is 0-9; "*" = more than 9 water molecules. Exposure can be read as solvated surface area in units of 10 $\text{\AA}^2$ .
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 $\text{\AA}$ ). 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 (continued)

5) 156B CYTOCHROME B562 (OXIDIZED) [ELECTRON TRANSPORT] {ESCHERICHIA COLI}.....156B
SHEET....
BRIDGE1..
BRIDGE2..
CHIRALITY +
BEND.... S
5-TURN... <
4-TURN... <
3-TURN... <
SUMMARY.. S
EXPOSURE. *6*716*43 **13*60 8*70*42* 77*2*61*8 02422*006 60**31*59 4**036*28* 1792*54**
1 SEQUENCE. ADLEDDMOTL NDNLKVIEKA BBZKANDAAL VKMRAALINA QKATPPPLED NSQEMKKFRH GDPDLVCGID DALKANEKG VKEAQAQAEQ LKTRNAYHQ
SHEET....
BRIDGE1..
BRIDGE2..
CHIRALITY +
BEND.... S
5-TURN... <
4-TURN... <
3-TURN... <
SUMMARY.. S
EXPOSURE. 93*
101 SEQUENCE. KIR
SHEET....
BRIDGE1..
BRIDGE2..
CHIRALITY +
BEND.... S
5-TURN... <
4-TURN... <
3-TURN... <
SUMMARY.. S
EXPOSURE. 93*
101 SEQUENCE. ADLEDDMOTL NDNLKVIEKA BBZKANDAAL VKMRAALINA QKATPPPLED NSQEMKKFRH GDPDLVCGID DALKANEKG VKEAQAQAEQ LKTRNAYHQ
1 CYT CYTOCHROME C (OXIDIZED) [ELECTRON TRANSPORT] {ALBACORE TUNA HEART: THUNNUS ALALUNGA}.....1.CYT
A
SHEET....
BRIDGE1..
BRIDGE2..
CHIRALITY +
BEND.... S
5-TURN... <
4-TURN... <
3-TURN... <
SUMMARY.. S
EXPOSURE. 8*65*9*21 892*35*321 *892*891912 329*9*43* 329*9*4927 8*16382*98 0**3092*5 *5*11***45 392*0113179
1 SEQUENCE. GDVAKKTKF VQRCAQCHV ENGGKHKVG P NMGIFGRKT GQAEGYSYTD ANKSKGIVWN NDTLMETYLEN PKYVPISTK IFAGIKKGE RDQLVWLS
SHEET....
BRIDGE1..
BRIDGE2..
CHIRALITY +
BEND.... S
5-TURN... <
4-TURN... <
3-TURN... <
SUMMARY.. HH
EXPOSURE. 20*
101 SEQUENCE. ATS
SUMMARY..... H=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-SERYLX... T=5-HELIIX... . T=3-, 4-... OR 5-TURN... S=BEND... .

7) 1C2C FERRICYTOCHROME C2 [ELECTRON TRANSPORT] [BACTERIAL: RHODOSPIRILLUM RUBrum].....1C2C
SHEET....
BRIDGE1..
BRIDGE2..
CHIRALITY +
BEND.... S
5-TURN... <
4-TURN... <
3-TURN... <
SUMMARY.. HH
EXPOSURE. *6*2328*6*1 2**956612 **7164*12 124213*41 116*6842* 5258921153 2**91884116* 62558*21* 568*4*921
1 SEQUENCE. EGDAAGKV SKKCLACHTE DOGSAKNGVP NIFGWENUTA ANKONVAYSE SYENPKAGL TWIRANLAD YVNPRAVLE KLGDRKASKE MTE KLGDRK
A
a
B
B
SHEET....
BRIDGE1..
BRIDGE2..
CHIRALITY +
BEND.... S
5-TURN... <
4-TURN... <
3-TURN... <
SUMMARY.. HH
EXPOSURE. *6*2328*6*1 2**956612 **7164*12 124213*41 116*6842* 5258921153 2**91884116* 62558*21* 568*4*921
1 SEQUENCE. EGDAAGKV SKKCLACHTE DOGSAKNGVP NIFGWENUTA ANKONVAYSE SYENPKAGL TWIRANLAD YVNPRAVLE KLGDRKASKE MTE KLGDRK

A

SHEET...:

BRIDGE1...  
 BRIDGE2...  
 CHIRALITY +  
 BEND... S  
 5-TURN...  
 4-TURN... <<<  
 3-TURN... >>>XXX<< <<XX444<  
 SUMMARY... HHHHHHHH HHHTTT STT SS TT TSBT T TT H BHHS B S BHHHHHHH HHHHSTT  
 EXPOSURE. 8\*65\*6\*21 8\*92\*35\*321 \*8\*2\*8\*91\*2 239\*32\*15\*8 0\*0\*8\*43\* 3\*9\*9\*4927 8\*216\*22\*98 6\*3\*3\*9\*7\*2\*5 \*5\*11\*45 3\*2\*11\*31\*79  
 1 SEQUENCE. GDVAKGKTE VORCAQHTV ENGGKHKGVP NLNGLFGKRT GOAEGTSYTD ANKSNGIWNW NDTLMFYLEN PKYIPTKRN IFAGIKRKGE RQDVLVAKS

SHEET...:  
 BRIDGE2...  
 BRIDGE1...  
 CHIRALITY +  
 BEND... S  
 5-TURN...  
 4-TURN... <<<  
 3-TURN...  
 SUMMARY... HH  
 EXPOSURE. 20\*  
 101 SEQUENCE. ATS

SUMMARY. ....H=ALPHA-HELIX....E=BETA-STRAND....B=BETA-BRIDGE....C=3-HELI...  
 SHEET...:  
 BRIDGE2...  
 BRIDGE1...  
 CHIRALITY +  
 BEND... S  
 5-TURN...  
 4-TURN... <<<  
 3-TURN...  
 SUMMARY... HH  
 EXPOSURE. 20\*  
 101 SEQUENCE. ATS

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

7) 1C2C FERRICYTOCHROME C2 [ELECTRON TRANSPORT] [BACTERIAL: RHODOSPIRILLUM RUBrum].  
 SHEET...:  
 BRIDGE2...  
 BRIDGE1...  
 CHIRALITY -  
 BEND... SSSSSSS SSSSSSS SSSSSSS SSSSSSS SSSSSSS SSSSSSS SSSSSSS SSSSSSS SSSSSSS  
 5-TURN... >555\*  
 4-TURN... >>>XXX<< X<4>44<  
 3-TURN... >>>XXX<< 3>>3<  
 SUMMARY... THHHHHHHH HTTVBTTB STT TTS TT BT T TTS SSSS B SHHHHHH TTSSSTTTT TS SS S HHHH  
 EXPOSURE. \*6\*220\*61 2\*\*206\*612 \*\*736\*4\*\*12 12421\*3\*61 116\*6842\* 1364\*6\*52 525\*921\*353 2\*9\*841\*16\* 625\*88\*2\*1\* 568\*4\*\*2\*1  
 1 SEQUENCE. EGDAAGEKV SKKCCLACHTF DQGGANKVKGPNLGCFVFNTHA KHDKNYATE SYTEKAKGL TWTEANLAY VKNPRAF VLE KSGDFKNSK MTFLKLKD

SHEET...:  
 BRIDGE2...  
 BRIDGE1...  
 CHIRALITY ++++++-++  
 BEND... SSSSSSSSS  
 5-TURN... 55>55<  
 4-TURN... XXXXXXXX<  
 3-TURN... >3<>3<  
 SUMMARY... HHHHHHHHHH  
 EXPOSURE. 9510\*#02\*2\*6  
 101 SEQUENCE. IENVAYLK T K

8) 155C CYTOCHROME C550 [ELECTRON TRANSPORT] [PARACOCCUS DENTRIFRICANS].  
 SHEET...:  
 BRIDGE2...  
 BRIDGE1...  
 CHIRALITY ---+----+----+----+  
 BEND... SSSSS SSSSSSS SSS S S S S S S S S S S S S S S S S S S  
 5-TURN... >555\*  
 4-TURN... >>>X<<44<  
 3-TURN... >>>X<<33<  
 SUMMARY... SHBH HHTTTTBTB SSS SS SSEEE SS TTS TT SS S S BHNN HHHH S SHHHHHHH SSTTTTTT S S  
 EXPOSURE. \*#376\*4\*\* 51\*78523 072\*8\*274\* 02\*8122\*61 113680\*166\* 85\*31 01\*69112 57621\*583\*34 60951\*138\*73 9\*11\*4\*  
 1 SEQUENCE. NEGDAAKGEK EFNKCKACHM IQAPPGTIDK GGKTPGNLG VVGRKIASEE GFKYSEGIE VAERNFDLTW TEANLIVETV DPKPVLVKKMT DDKGAKTKMT

SHEET...:  
 BRIDGE2...  
 BRIDGE1...  
 CHIRALITY ---+----+----+----+----+----+----+----+  
 BEND... S SSSSS SSSSSSS S SS S S S S S  
 5-TURN...  
 4-TURN...  
 3-TURN... >3333<  
 SUMMARY... S HHHH HHHHHHHH S SS S ST T  
 EXPOSURE. 9\*2\*744\*22 02112\*31\* 6795\*78\*78 4146  
 101 SEQUENCE. FNGGNNQDDV VAFLAQDDPD AXXXXXXXX XXXX

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

TABLE AIII (*continued*)

9) 251C CYTOCHROME C551 (OXIDIZED) [ELECTRON TRANSPORT] [PSEUDOMONAS AERUGINOSA].  
SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY...  
BEND...  
5-TURN...  
4-TURN...  
3-TURN...  
SUMMARY...  
EXPOSURE...  
1 SEQUENCE.

10) 1PDX FERREDOXIN [ELECTRON TRANSPORT] {PEPTOCOCCUS AEROGENES}.  
SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY...  
BEND...  
5-TURN...  
4-TURN...  
3-TURN...  
SUMMARY...  
EXPOSURE...  
1 SEQUENCE.

11) 1PXC PERREDOXIN [ELECTRON TRANSPORT] [SPIRULINA PLATENSIS].  
SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY...  
BEND...  
5-TURN...  
4-TURN...  
3-TURN...  
SUMMARY...  
EXPOSURE...  
1 SEQUENCE.

12) 3FXN FLAVOREDOXIN (OXIDIZED) [ELECTRON TRANSPORT] [CLOSTRIDIUM NP].  
SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY...  
BEND...  
5-TURN...  
4-TURN...  
3-TURN...  
SUMMARY...  
EXPOSURE...  
1 SEQUENCE.

13) 2RNM RUBREDOXIN (OXIDIZED, FE(III)) [ELECTRON TRANSPORT] [CLOSTRIDIUM PATURIANUM].  
SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY...  
BEND...  
5-TURN...  
4-TURN...  
3-TURN...  
SUMMARY...  
EXPOSURE...  
1 SEQUENCE.

(111) 1PCX FERREDOXIN [ELECTRON TRANSPORT] {SPIRULINA PLATENSIS} . . . . . 1PCX SHEET

SHEET...  
 BRIDGE 2...  
 BRIDGE 1...  
 CHIRALITY  
 BEND...  
 5-TURN...  
 4-TURN...  
 3-TURN...  
 SUMMARY...  
 EXPOSURE...  
 101 SEQUENCE.  
 13) RXN RUBRETOKIN (OXIDIZED, FE(II)) [ELECTRON TRANSPORT]  
 SHEET...  
 AA  
 AAA  
 dd  
 eeee  
 +++++-+---+  
 SSSSSS  
 S  
 SSS  
 SSSSSSSS  
 SSSSSS  
 >555<  
>XX<<<  
>>>>><<<  
>3<<  
>>3<<  
HHHHHTT EE  
S EEEESS  
GGHHHHHHH  
HHHHHTH  
\*6#08763.173  
\*13\*261-87  
#9 29\*#018  
#10\*#119  
ERMSYGCCV VETPLIVQE PDEAQCDCIE FGKKKANI  
{CLOSTRIDIUM PATEURIANUM}  
B AA A  
..... 2RN

SUMMARY.....H-ALPHA-HELIX.....E=BETA-STRAND....B=BETA-BRIDGES....G=J-HELIX....I=5-HELIX....T=3-A-, OR 5-TURN....S=BEND.....

TABLE III (*continued*)



TABLE III (*continued*)

SHEET...  
 BRIDGE...  
 BRIDGE...  
 CHIRALITY  
 BEND....  
 5- TURN...  
 4- TURN...  
 3- TURN...  
 SUMMARY...  
 EXPONENTIAL  
 101 SEQUENCE.  
 22) 1SNS STAPHYLOCOCCAL NUCLEASE (COMPLEX) [PHOSPHORIC DIESTER (DNA) HYDROLASE]  
 SHEET... A AAAA A (COMPLEX)  
 BRIDGE2...  
 BRIDGE...  
 CHIRALITY  
 BEND....  
 5- TURN...  
 4- TURN...  
 3- TURN...  
 < >><<<  
 >>><<<  
 >>>>>>  
 SSIGGGGSHH HRRHTWTS G  
 \*711055171 24\*\*\*697\*1  
 DGDGMNAWA WNRBKGTDV  
 QWIRGARL  
 1SNS STAPHYLOCCUS AUREUS] [STAPHYLOCCUS AUREUS] ISNS

SHEET, ...  
 BRIDGE 2, ...  
 BRIDGE 1, ...  
 CHIRALITY, ...  
 BEND, ...  
 5-TURN, ...  
 4-TURN, ...  
 3-TURN, ...  
 SUMMARY, ...  
 EXCUSE, ...  
 101 SEQUENCE,  
 SUMMARY, ...  
 B=ALPHA-HELIX, ... E=BETA-STRAND, ... S=BETA-BRIDGE, ... G=3-HELIX, ... T=5-HELIX, ... I=5-HELI  
 X, ... V=VITRIN, ... Q=KERATIN, ... H=PROTEIN, ... R=REPTILE, ... D=DEUTERIUM, ... C=CARBON, ...  
 OR 5-TURN, ... S=BEND, ...



TABLE AIII (continued)

SHEET...	AAA	AAA A	AAA A	AAA
BRIDGE 2..	GG	GG	GCG	
BRIDGE 1..	fff	fff	Gehee	
CHIRALITY	+--++++-+--	-+++-+--+--	+--++++-+--	+--++++-+--
BEND....	S SS	SSSSSSSSSS SSS	SSSSSSSS S S	SSSSSSSS SSSSSSSSS
4-TURN...			>555 5<	
3-TURN...	>33<	>>>XXX XXXXX<<<	>>>XX<<	>>>XXX XXXXXXXXX
SUMMARY..	EES SS	TTHHHHHH HHHHHHHH SSS 4	S HHHHHHH T S EEEE	S SSSST 33<
EXPOSURE..	00102044* 91948*569 0052650*	98*1182*90 005057*610	0000020278 6160000000 31***1*1*1	43*053005 002400200
201 SEQUENCE..	LILPYGYMTQ SIPDKTEINQ VAKSAVALK SLYCTSYG SITIYQAS GGSIDWSYNO GIYTSFTEL RDTGRYCFLL PASQIIPAQ ETWLGVLTM			
SHEET...				
BRIDGE 2..				
BRIDGE 1..				
CHIRALITY	+--++++-+--			
BEND....	SSSS			
5-TURN...	X<<<			
4-TURN...	>>>XX<<			
3-TURN...	HRRHRT			
SUMMARY..				
EXPOSURE..	7604**042			
1 SEQUENCE..	EHTNNNIGY			
25) LAPR ACID PROTEASE [HYDROLASE: PROTEINASE] {RHIZOPUS CHINENSIS} .....				
SHEET...	AB C C	BBB B DD E	E	F
BRIDGE 2..		ccc		GGF H H
BRIDGE 1..	A	AB E E	D ff g	
CHIRALITY	-+-----	-+----+--+--	-+--+--+--	IIH J J
BEND....	S	SSSS	SS SS SS	SS
5-TURN...				
4-TURN...				
3-TURN...	>33<	>33<	>444<	
SUMMARY..	TTT B TTSS BB E E SSS E	EESS B EE BS	SS	EEB BSSS B
EXPOSURE..	*0409273* 5*051290 4012*5615 0111022011 110681*55 008628715 480*66*75 9**175584* 2615150114 106126105	TTT B TTSS SS SS	B	
1 SEQUENCE..	GVGTVPHTDY GNDVEYYGVQ TIGTPGKSPN LNFDTGSSLN WVGSVQAS GaKGDRDFN PSDSTETAT CYDASICYGD GSASGVUGYD TVQYGGDVR			
SHEET...				
BRIDGE 2..				
BRIDGE 1..				
CHIRALITY	-+-----	-+----+--+--	-+--+--+--	L L
BEND....	S S	SSSSSSSS SS S S SSSS	SSSSSSSS SSS	SS SSS S
5-TURN...				
4-TURN...				
3-TURN...	>33<	>444<	>444<	
SUMMARY..	S TEE SSSSSSS SSEE S SS SSSSS	HHHHHHSSSS SSS	EE E TTT EEE	SS "TSB S SSSS B
EXPOSURE..	4015121049 1244624631 0020000149 73152**170 1041066688 1*30660044 024**85130 2112*5*43 3742591726 21*171070	LSITRQSTN AFQNSAAQK VIFPVEVNL ASNEDGDF TMFWLDNKY GGTLTNTD AGCYWALV		
101 SEQUENCE..	GGPIQLAQR LGGGGEFGDN DGLIGLCFD			
SUMMARY..				
..... H=ALPHA-HELIX. *.=BETA-STRAND. .*=BETA-BRIDGE. ...=G-HELIX. **=I=5-HELIX. ***=T=3-, --, OR 5-TURN. .... S-BEND. ....				
SHEET...				
BRIDGE 2..				
BRIDGE 1..				
CHIRALITY	+--++++-+--	-+----+--+--	-+--+--+--	R R
BEND....	SSSS	S SSSS	SSSS	SS SSS S
5-TURN...				
4-TURN...				
3-TURN...	>33<	>33<	>33<	
SUMMARY..	S EE 2612*1011 44340027* 1032816618 3897*3210 6155*473* 30115***1 9134001 004668054	B S TTS BSSS SSS B	SSSS B	SSSSBEE
EXPOSURE..	TGATADSTYL GAIFQILDT GTSLLTLPDE AVGNLVGF AODAALGGF VIABTSAGFK SIPWSYSA 1 FELITALGA EDISGBTGTI GASSGEAL			
201 SEQUENCE..				
SHEET...	P KKK	KKK	P	

3- TURN . . . > 33< . . . . .  
 3- TURN . . . > 33< . . . . .  
 SUMMARY . . . TT B TISS BB E EESSS EE SS B EEB SS B EEB SS B  
 \* 0469273 . . . 5\* 045120808 401256715 011022011 110668155 00\*628715 2TT SS B  
 EXPOSURE . . . GVGTVYNTY GNDVVEYYGV TIGTPGKSFN LNF DTGSSNL WVGSVQaQAS GaKGDRDFN PSDGSTKAT GYDASIGYD GSASVGLGYD TVQNGGDVT  
 1 SEQUENCE.  
 SHEET . . . DDG G BBB  
 BRIDGE 2 . . . fff I DCC  
 BRIDGE 1 . . . S S SSSSSSSS SS S SSSSSS SSSSSSSS SSS  
 CHIRALITY . . . S S SSSSSSSS SS S SSSSSS SSSSSSSS SSS  
 BEND . . . . .  
 5- TURN . . . 4-TURN . . .  
 3- TURN . . . > 33< . . . . .  
 SUMMARY . . . S T TEEE S SSSSSS SSEE S S SSSSS SSS  
 \* 026806149 124426431 401512089 73112\*170 1041068688 1\*30008040 E TTT EEE  
 EXPOSURE . . . GGPQFLQAR LGGGSPCDN DGLLGFDT LSTPFGSTN AFQDSVAGK VIQQFWVVYL AASNSIDGDF  
 1 SEQUENCE.  
 SUMMARY . . . H=ALPHA-HELIX. . . . B=BETA-BRIDGE. . . . G=3-HELIX. . . . I=5-RELIX. . . . T=3-4-. . . . R=5-TURN. . . . S=BEND. . . .

TABLE AIII (continued)

SHEET...:	GGG	FFF	FFF	FFP	V	HHHH	HHHH	GGGG	GGGG	H	HHH	HHHH	FFP	FFF
BRIDGE2..:	t tt	S	uu u	--t+t-t+--	-t+t+t+t+t+	zzzz	AA	YYYY	YYYY	BBB	BBB	v	ttt	
BRIDGE1..:	XXX	SSS	SSS	SSSSSSSS	SSSSSSSS	SSSSSS	SSS	SS	SS	AA	uuu		ttt	
CHIRALITY..:	+-----+----+	+-----+-----+	+-----+-----+	+-----+-----+	+-----+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	
BEND...::	SS	SS	SS	SS	SS	SS	SS	SS	SS	AA	uuu		ttt	
5- TURN...::										S	ss			
4-TURN...::										S	ss			
3-TURN...::										S	ss			
3-TURN.::										S	ss			
SUMMARY..:	TTTTFF	EE	EE	TTT	SSEE	EE	BBBBHHHH	TTT	SSSEEE	TTT	EEEEE	TT	SSSEEE	
EXPOSURE..:	3*91*426	3*923*303	3*923*6501	3*93*524*	3*2886*021	3*2886*76*	6748866608	82016000131	*6580000	8	3*3192*28	6748866608	82016000131	*6580000
201 SEQUENCE..:	GSQSGDGFSQ IADGTLL LDRSVVSYQQ SYQSGAQQDS NAGGYWFDS SSVYDFESVSI SGYATVPGS LINYGPSCNG STALGGIQS N SIGGFLIGD													
SHEET...:	I	BBB	B	BBB	B	E								
BRIDGE2..:		C	CCC	C	EEB	CR								
CHIRALITY..:	+++++-----+-----+	-+-----+-----+	-+-----+-----+	-+-----+-----+	-+-----+-----+	-+-----+-----+								
BEND...::	SSSS	SSS												
5- TURN...::														
4-TURN...::														
3-TURN...::														
SUMMARY..:	RRHHTB	EEE	ETTEEE	BB	BBB	BBB	BB	BBB	BB	BB	AA	AA	AA	AA
EXPOSURE..:	60011000008	337171000	548	IPLKSOVVF DSDQFOLGEA PQA										
27) 2TIN THERMOLYSIN [HYDROLASE: NEUTRAL METALLO-PROTEINASE] {BACILLUS THERMO-PROTEOLYTICUS}.....:														
SHEET...:	AAAAAA	A	AAA	AAA	AA	AA	BB	BBB	BBB	BB	BB		BB	
BRIDGE1..:	bb	c	dd	ee	ff	ff	ee	ee	ee	ee	ee			
CHIRALITY..:	AAAAA A	AAA	AAA	C	E	E	G	G	G	D	DD	bb		
BEND...::	+-+-----+-----+	-----+-----+	-----+-----+	-+-----+	-+-----+	-+-----+	-+-----+	-+-----+	-+-----+	-+-----+	-+-----+			
5- TURN...::														
4-TURN...::														
3-TURN...::														
SUMMARY..:	EEEEEE	E	SSS	EEE	EEE	EEE	EE	EEE	EE	EEE	EE	EEEEE	SSPTTTHHH HHHHTHTS	STTTTS
EXPOSURE..:	94*8749130 * 60*553*6 32*4**50 32*8*8*6 63*325692 9075*2110 6000*10018 13107*72*3 8044*4226													
1 SEQUENCE..:	ITGTSIVGV RGVLGDQNI NTIYTSTYL QDNTRDGF TYDAKTRTTL PSLSLWADQN QFFASYDPA VDIAHYTGT YDVKVHNRLS TDGHNAI													
SHEET...:	BBBBB	B	BB	B	B	B	C						DD	C
BRIDGE2..:	hh	h	h	h	h	h	h	J	J	SS	SSS	SSSSSSSSSS	SSSSSSSSSS	J
CHIRALITY..:	----+---+--+--+	++-+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	KK
BEND...::	S	SSS	SSS	SSS	SSS	SSS	SSS	SSS	SSS	SSS	SSS	SSSSSSSSSS	SSSSSSSSSS	S
5- TURN...::														
4-TURN...::														
3-TURN...::														
SUMMARY..:	EE	SSSE	EEE	SSSS	EEE	SSSS	EEE	SSSS	EEE	SSSS	EEE	SSSS	SSSS	SSSS
EXPOSURE..:	*80013*36 5786313810 6003100010 0110010007 *4100000000 0100010000 6001810007 *4121998*6 1600000000 1600010000 6001810007 *41331003 507277**64													
101 SEQUENCE..:	RSSVHQSQQY NNRFNWGSEW VIGDGCGCIP FLSGSDIVV ANELHATND YIFAGLYIN EGAIAEHAIS IEGTLETFYVA NWNPMEIGE DWYTPGICGD													
SUMMARY....:	H=ALPHA-HELIX...: E=BETA-STRAND...: E=BETA-STRAND...: I=5-HELIX...: G=3-HELIX...: E=HELIX...: T=3-4-, OR 5-TURN...: S=BEND...: .													
SHEET...:	DD													
BRIDGE2..:	KK													
BRIDGE1..:	S	SS	SS	SS	SS	SS	SS	SS	SS	SS	SS	SSSSSSSSSS	SSSSSSSSSS	S
CHIRALITY..:	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	-----+-----+	
BEND...::	S	SS	SS	SS	SS	SS	SS	SS	SS	SS	SS	SSSSSSSSSS	SSSSSSSSSS	S
5- TURN...::														
4-TURN...::														
3-TURN...::														
SUMMARY..:	SESS	GQQ	TT	SCGC	SSHHHHHH	TTTHHHHHHH	HHHHHH	EES	EE	SESS	SS	SS	SSSSSSSSSS	S
EXPOSURE..:	#5304805* 77222368* 7*3*59100 541000000 660180000 73811803* 678174239 *60090370 0149 001568391 DKLGRKFYRA LGQTYLTFNSN FSQRLAAVQ SATDLSQGS													
201 SEQUENCE..:	SLRSMSDPAK YGDPPDRYSKR YTGTQDRNGV HINGLINKA AYLISQSGTH YGFSVVGICGR DKLGRKFYRA LGQTYLTFNSN FSQRLAAVQ SATDLSQGS													
SHEET...:	EE													
BRIDGE2..:														
BRIDGE1..:														

EXPOSURE. \*94879130 \*83\*5\*3\*4 488\*4\*550 11685\*#02 0320\*\*\*8\*6 63322692 9875\*#210 63322692 9875\*#02 0320\*\*\*8\*6 QDNTRGDIF TIDAKRTPL POSILWADDN QFASIDAPKA VDARYAGT YDVKVHNK ISIDGNNAI  
 1 SEQUENCE.  
 SHEET... BBBB B BB B B C DD C  
 BRIDGE1.. hh h I I J KK J  
 CHIRALITY SSSS ->33< SSS S S S S SS  
 BEND... ffff SSSS SSSS SSS SSS SSSS SSS SSS SSS SSSS  
 5-TURN... >44< SSSS SSSS SSS SSS SSS SSS SSS SSS SSS  
 4-TURN... >33<>33< SSS SSS SSS SSS SSS SSS SSS SSS SSS  
 3-TURN... >44< >33<>33< SSS SSS SSS SSS SSS SSS SSS SSS SSS  
 SUMMARY. PEEFTTTT EE SSEE EEE EEE EEE EEE EEE SSS  
 EXPOSURE. #6014\*36 5796373810 6093195\*61 42000007 0.11010007 11219986 10000000 10000000 10000000 10000000 10000000 10000000 552777\*64  
 101 SEQUENCE.  
 RESVHTQY NHFWGSEW VIGDGGPFI IPGLGIDVV AELTHLTVDV YTAGLIVQE SCAINAISD IFTGLVEFA KNPMEWEIG DVYTPCISGQ  
 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 DD EE EE LL LL  
 BRIDGE2.. RK S SS SSS  
 CHIRALITY ->33< SSS  
 BEND... >555< SSS  
 5-TURN... >555< SSS  
 4-TURN... >33<>33< SSS  
 3-TURN... >33< >33<>33< SSS  
 SUMMARY. SEEES GGG TT SSGG SSSHHHHH TTPTTHHHHH HHHHHHT EES SS EE S S TTHHHHHHH HTT TT HHHHHHH HHHT TT  
 EXPOSURE. #53064005\* 7122239\* 7\*3\*5100 5410000020 66801813\*2 758174239 60800370 049167\*14 0596341049 04566839\*1  
 201 SEQUENCE. SLRSMSDAK YGDRDHTSKR YGTQDNGV HINGGINKA AYLISQGQGH YGVSVNGIGR DLKQKPYRA LTOYLPFSN FSQRLRAAVQ SATLYGTS  
 SHEET...  
 BRIDGE2..  
 BRIDGE1..  
 CHIRALITY +++++++ ++-+  
 BEND...<---- SSSSSSSSS SSS SSS  
 5-TURN... >555<  
 4-TURN... >>>>>>>>>>>>  
 3-TURN... >33</>33<  
 SUMMARY. HHHHHHHH HHHT  
 EXPOSURE. \*26000500 800200\*  
 301 SEQUENCE. QEVASVQAF DAVVK  
 28) 2GCH GAMMA CHYMOTRYPSIN A [HYDROLASE: SERINE PROTEINASE] {COW PANCREAS: BOS TAURUS}. CCCCC CCC C CCCC CCC D D CC  
 BRIDGE... A BB KKKK JJJJ L L MMM M KK NNN N NN  
 CHIRALITY +---+----+  
 BEND...<---- SS SSS S  
 5-TURN... >33< >33< >33< >33< >33< >33< >33< >33<  
 4-TURN...<---- SS SSS S  
 3-TURN...<---- SS SSS S  
 2-TURN...<---- SS SSS S  
 1 SEQUENCE. aGVPAIQVVL SYNEQAEVPG SWEQVQLD KTGFHFbGGS LINENMVTAA ABQGVTSDV VVAGEPDQGS SSEKIQIKL AKEVNSKVN SELINNDV  
 SHEET... CCC B BBBBB BBB BBBB  
 BRIDGE2.. 000 FE BBBB BBBB  
 BRIDGE1.. MN D DDDDD DDD DDD  
 CHIRALITY ->33< SSS  
 BEND...<---- SSS S  
 5-TURN... >33< >33< >33< >33< >33< >33< >33< >33< >33< >33< >33< >33< >33<  
 4-TURN...<---- SSS S  
 3-TURN...<---- SSS S  
 2-TURN...<---- SSS S  
 1 SEQUENCE. BESS SS B TT EE EEEESS S SS EEE EEE EEE EEE EEE S  
 SUMMARY. BESS SS B TT EE EEEESS S SS EEE EEE EEE EEE EEE S  
 EXPOSURE. #667651877 \*531312415 575\*316559 00600046\* \*929\*1301 504\*0233\*67 024961000 00929\*637\*60 \*\*5165273 16529\*19723 000123464 \*544\*5\*57 875531019  
 101 SEQUENCE. LKLSTASAS QRVASAVLPS ASDFDFAGTT CTWTGNGFLV VITDPLQQA SULPLSNINA KRYWGTQIKRD ARIKdGASGV SseMGDGGP LVCKRKGANT  
 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 III (continued)

SHEET...	B BBB B	BBBB BB	B BBBB BB	B BBB	A	B BBB BB BB B B D
BRIDGE1..	H HH	II II	EEEE S SSS	EEE EBB	A	H HHH HH HH
BRIDGE1..	III I	GGG G	DDDD D	G GGG	A	H HHH HH HH
CHIRALITY	+--+ +--+	--+ +--+	DODDD D	F	A	C II I I P
BEND....	S S	SSS	--+ +--+	--+ +--+	-+ +--+	-+ +--+
5-TURN...			S SSS	SS	SS SSSS	SS SSSS S
4-TURN...						
3-TURN...						
SUMMARY...	EEEEEE S SSSS	EGGGTHHH HHHHHH	>44>XX XXX<<	>33<		
EXPOSURE...	00010512*	703830100	050160150 *	1380*		
1 SEQUENCE.	LGVISNSGS	TESTSTRCCV ARVALYNNW QCTLAAN				
201						
29) JALP ALPHA LYtic PROTEASE [HYDROLASE: SERINE PROTEINASE] {MYXOBACTER495: LYSOBACTER ENZYMOGENES} .....JALP						
SHEET...	A BBB B	BBBB BBB FF G GG	BBB BB HHHH IIII IIII	C BBBB AAAA D		
BRIDGE2..	DD D	FF G GG	BBB BB HHHH IIII IIII	BBB B		
BRIDGE1..	a CCC	EEE EEE E	DDD HH HH j GGC	BBBB q		
CHIRALITY	-+ +--	--+ +--	-+ +--	-+ +--	-+ +--	-+ +--
BEND....	SS	SSSS	SSSS S S	SS S	SS	SS
5-TURN...						
4-TURN...						
3-TURN...						
SUMMARY...	BT EEE ESSSSEE EEEETEE EE SSSSE	EEEEE S B EEEE EEE EEE	>33<	>33<		
EXPOSURE...	6*0*017702 18*188000 0105*89640 000004296 *006143668 006143668 006143668	006143668 006143668 006143668	587*140*1 3973675370			
1 SEQUENCE.	ANIVGIBYS INNALSAGVG FSVTRGATKG FVPAAGHAGV NATARIGAV VGTSFARVFP GNDRAWVLT SAQTLIPRVA NGSSFTVTRG STEAAVCAV					
SHEET...	CCCC C CC C CC	CCCC E	CCC D	CCC CCC	BB	
BRIDGE2..	LL MM	MM MM	OO	PPP PPP	PP	
BRIDGE1..	KKK K M MM M	M J NNN R	LL q	R	FF	
CHIRALITY	-+ +--	--+ +--	-+ +--	-+ +--	-+ +--	-+ +--
BEND....	SSS S	SSS S	S S SS	SSSSSS S SSS	SSSSSS S	SSSSSS S
5-TURN...						
4-TURN...						
3-TURN...						
SUMMARY...	E EEEEEE EEEEEE EEEEEE BT TT EEEEEE T B EEEEEE EEEEEE S SSSSS	EEEEE EEEEEE S SSSSS	>33<	>33<		
EXPOSURE...	0001*5455 62*0537*47 1**360*2 000202010* 1020000003* 703010001 3367**0207 6313**5401 1000*3422* 24172395	000202010* 1020000003* 703010001 3367**0207 6313**5401 1000*3422* 24172395				
101	SEQUENCE.	BRSGRTGQ YO bGT TAKNT ANTAEGAVRG LTQDNGACMGR GDGGSMWTS AGQAQGMWSG GNVOQNGNC GIPASQDSL FERQPLISQ YGLSLVNG				
30) IPTN BETA-TRIPSIN (NATIVE AT PH 8) [HYDROLASE: SERINE PROTEINASE] {COW PANCREAS: BOS TAURUS} .....IPTN						
SHEET...	A BB	C CCC C CCC	CCC CCC	CCCC CCC	CCCC	
BRIDGE2..	KKK L L	MMM MNNN	NNNN	NNNN	NNNN	
BRIDGE1..	A BB	J JJJJ	L L	KKK	KKK	
CHIRALITY	+ +--	+ +--	+ +--	+ +--	+ +--	+ +--
BEND....	S SS	SS S	SS S	SS S	SS S	SS S
5-TURN...						
4-TURN...						
3-TURN...						
SUMMARY...	BS EEE TT SSSSEEEEEE EEEEEE EEEEEE EEEEEE EEEEEE EEEEEE S SSSSS	EEEEE EEEEEE EEEEEE S SSSSS	>33<	>33<		
EXPOSURE...	0113*83164 912100023 743100002 541000007 0*5.615050 11333* *5 5338242764 452*7678*7 342*000040 *811887*70	541000007 0*5.615050 11333* *5 5338242764 452*7678*7 342*000040 *811887*70				
1 SEQUENCE.	IYGGTGAN TPYQVQLNS GYFBGSLI NSGWVSAH BYRSQIOLR GEININVEG NEQIASKS IVPSVNTS LANDIMILKL KSASINSRV					
SUMMARY.....	H=ALPHA-HELIX... B=BETA-STRAND... G=3-HELIX... T=5-HELIX... B=BETA-BRIDGE...	I=5-HELIX... T=3- HELIX... B=BEND... S=S-END...				

4DER

CHIRALITY --+--+---+-- 4-TURN...  
 BEND... SSS S SSS S S S SS S S S  
 5-TURN...>44<  
 3-TURN... EEEETEEE E EEEEEE EEEEEE EEEEEE S SSS S S S S S S  
 SUMMARY. 0001\*54455 62\*053\*47 1\*6\*160\*2 00\*02801\*07 1026003 70601001 3377\*0207 6332\*5401 1000\*392\*\* 24172395  
 101 SEQUENCE. DRGRTYQ DGTTAKNT ANYAEGVRG LQSNAGMR GDSGGWITS AGAQGMG GRVQSGNC GIPASORSL FERQPLSQ YGISLVIG

30) IPTN BETA-TRYPSIN (NATIVE AT PH 8) [HYDROLASE: SERINE PROTEINASE] {COW PANCREAS: BOS TAURUS}...  
 SHEET... A BB CCCC C CCC C  
 BRIDGE1.. RRRK L LLL J JJJJ  
 BRIDGE2.. A BB MMM KKK  
 CHIRALITY +--+---+-- 4-TURN...  
 BEND... S SS SSSS S SS S SS S S S  
 5-TURN...>3 <33<  
 3-TURN...>3 <33<  
 SUMMARY. BS EE TT SSSBBBBB EEEEEE GG SS EEEEEE S SSS S S S S S S  
 EXPOSURE. 0153\*2364 911210023 7\*4310002 547100007 0\*5\*615050 1.12337\*\*5 53382\*2764 452\*7678\*7 342000040 \*851887\*70  
 1 SEQUENCE. IVCGYTAAN TVPYQVSLNS GYHEFGSLLI NSQWVSAAH BYKSGIQVRL GEININVVEG NEFTIASKS IVRPSINSNT LNNDIMIKL 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....

SHEET... B BBBBBBB  
 BRIDGE2.. B DDDDD  
 BRIDGE1.. C DDDDD  
 CHIRALITY -+---+-- 4-TURN...  
 BEND... SS S SSS S S S S S S S  
 5-TURN...>3 <33<  
 3-TURN...>3 <33<  
 SUMMARY. B SS TT EEEEEE S SSS EEEEEE EB HHHHHH HSTTT TTE EEEES TWSS B TT TT E EEEEEE EEE SSSBT  
 EXPOSURE. 351851\*763 8664\*1.000 0002\*65\*6 \*2730401\*0 4125992098 1059924831 20007714\* 111\*2\*0000 010\*490100 11375118\*  
 101 SEQUENCE. ASISLPSCA SACTQQLIG WGNTKSSGTS YPDVYLKLA PILNSSESKS AYPCQTSNM FeAGYLGKX DSFGQSGGP VNGSGIAQK

SHEET... D BBBB  
 BRIDGE2.. IIII  
 BRIDGE1.. P GGG  
 CHIRALITY +--+---+-- 4-TURN...  
 BEND... S SS SSSSSSSSS S  
 5-TURN...>3 <33<  
 3-TURN...>>>>><<<  
 SUMMARY. TB EEEPEGG GHIIHHHHHH HH  
 EXPOSURE. B70000602 713950\*14 889  
 201 SEQUENCE. NKGIVVTKC NYVSVIQT1 ASN

31) ISGA PROTEINASE A (SGPA) [HYDROLASE: SERINE PROTEINASE] {STREPTOMYCIS CRIBUS}...  
 SHEET... AA PR BB BBBB  
 BRIDGE2.. AA C DDDDD  
 BRIDGE1.. AA BB BB BBBB  
 CHIRALITY -+---+-- 4-TURN...  
 BEND... S S SSS S S S S S S S  
 5-TURN...>3 <33<  
 3-TURN...>3 <33<  
 SUMMARY. 400054545 95\*1000111 426420000 0000066349 1621774153 4762000\* 184932528 0442\*7\*96 281700251  
 1 SEQUENCE. IAGGEATTG GSRCGLFNV SYNGVHLA AGCTNISAS WSIGHTGTG FPNNDGILR HSMPAAANGR VLYNNSYQD ITTGANAFVG QAVQRGSTV

SHEET... EEEEC C CC CCCC  
 BRIDGE2.. MMMMH H H HH 9 III  
 CHIRALITY -+---+-- 4-TURN...  
 BEND... S S SS S S S S S S S  
 5-TURN...>3 <33<  
 3-TURN...>3 <33<  
 SUMMARY. EEEEEEEE EEEEEE GGG EEEEEE TT BT E EEEEEE EEEESBTWT B EEE BH BRRHHHTT  
 EXPOSURE. 158\*1501619 \*222\*6\*91 70330107071 118\*002000 005530000 11259611\* 2350300305 907965093 \*  
 101 SEQUENCE. GLRSGSVTGL NAIVNVSQSG IVVGNQTNV CAPGSDGGS LPAGSFALGL TSGSGNCRT GTGTF QPVT EAISAGATV L  
 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 III (*continued*)

```

SHEET ... E E B BBBB B F F D BBBB BB BBBBB BB
BRIDGE 2... P P DDD D DDD Q 0 BB EE FFFF GGGG
BRIDGE 1... CCCC 0 EE SSSSSS SSS SSS SSS SSS
CHIRALITY ... ---+---+---+---+---+---+---+---+---+---+
BEND ... S SSS SS SSSSSS SSS SSS SSS SSS
5- TURN ... >555< >555<
4-TURN ... >>4< >33<
3- TURN ... >>33< >>33< >>33< >>33< >>33<
SUMMARY... S S BEEE EEEE FS EEEE B SB EEEE
EXPOSURE... *705968099 61651* 634* 476-250600 00022*554* 71*2133000 22001729 9001*1*3000
1 SEQUENCE... QSVPLNSVQ LGVLPRASTI LANNSPBYT QWGLTRNGQ LAQTQAVL PIVDIAICSS SMWAGGDCV RSDSGCG DGDSCG PLBLVNGQY

```

	SHEET...	E E	B BBBB B F	F D BBBB BB	B BBBB BBBB	A BBBB	BBB BBBB
	BRIDGE1..	P P	D DDD D	D DDD D	F FFF FFF	G GGGG	G GGGG
	CHIRALITY	--+-+---+	--+-+---+	CCC 0	EE EE	CCC	CCC
	BEND...	S SSS	--+-+---+-- SS	0 B B	--+-+---+-- SS	A A	--+-+---+-- SS
	5- TURN...			--+-+---+-- SS	--+-+---+-- SSSSSS	SS SS	--+-+---+-- SS
3- TURN...		>33<	>33<	>33<	>33<	>33<	>33<
SUMMARY.		S BTB	61651*TT	EEEEE	EE HHRRHH	SB TT TT	EEEEEETT
EXPOSURE.	*709668999	1651*654*	TT	EEEEE	T TEEBE	SS SB	EEEEEETT
SEQUENCE.	(DQVTLNSYQ	025600	2125*6045*	71*21310680	2125*6045*	9001*11300	010647**
101		LQVLPRAGT	LANNSPDYT	GNGLTRTNGQ	LACTLQOAYL	PTVDYACSS	RSGdQGDSGG PLBLVNGQY
	SHEET...	BB BBB B G	G H H BBBB				
	BRIDGE1..	GG GG	R S S	HHHH			
	CHIRALITY	--+-+---+	--+-+---+	FFFF			
	BEND...			--+-+---+-- SSSSSS			
	5- TURN...						
3- TURN...		> 33<	>33<	>33<	>33<	>33<	>33<
SUMMARY.		PEEEPEE B	TTBSEGTB	EEEEPGGS	HIIHRIHHT		
EXPOSURE.	1610010321	*10548*8	0000300512	8A07*	2899		
SEQUENCE.	AVHGVTSE	VS RIGDNTRKP	TFTVRSAYI	SWINNVIASN			
34)	2ACT- ACTINIDIN [HYDROLASE: SULPHYDRYL PROTEINASE] (KIWFUIT: ACTINIDIA CHINENSIS) . . . . .	C	D	D	D	D	2ACT
	SHEET...	AA	B				E C
	BRIDGE1..	AA	I	J	K	L	J
	CHIRALITY	--+-+---+	--+-+---+	--+-+---+-- SSSSSS	--+-+---+-- SSSSSS	--+-+---+-- SSSSSS	--+-+---+-- SSSSS
	BEND...	S SSSS	SS	SSSSSSSS	SS	SSSSSSSS	SS
	5- TURN...	>555	5<	>55	>555	>555	>555
4- TURN...		>><<	>><<	>><<	>><<	>><<	>><<
3- TURN...		>><<	>3 <	>3 <	>3 <	>3 <	>3 <
SUMMARY.	S BEGGG T	B T	HIIHHHHHHHHH	HHS B H	HIIHHHHHHHHH	T B BTTTS	H
EXPOSURE.	*6*224*5	4822*23515	1*3020*100	0100*01167	6746754000	0000*022*1*6*	57373*489
1	SEQUENCE.	LPSYDWRSA	GAVIDIKSQG	EaGGCWAFSA	IATVEGINKI	TGSCLSLSE	QELIBDGRQT NTRGADGYYI TDGFQFLIND GGINTENYP YTADQGDDBDV
	SHEET...	E AAAA	AAAAA AA	AAAAAA AA	AAAAAA AA	AAAAAA AA	AAAAA
	BRIDGE1..	L BBBB	CCCCC	SSSSSSSS	SS	SSSSSS	GGGG
	CHIRALITY	--+-+---+	--+-+---+	--+-+---+-- SSSSSS	--+-+---+-- SSSSSS	--+-+---+-- SSSSSS	--+-+---+-- SSSSSS
	BEND...	SSSS	SS	SSSSSSSS	SS	SSSSSS	SS
	5- TURN...	><<	>	>><<	<	>><<	>4 <
3- TURN...		>33<	>33<	>33<	>33<	>33<	>33<
SUMMARY.	HIIHH B	EEEEE TT	HIIHHHHHHHH	SS EEEEEE	EEEEE	EEEEE	EEEEE
EXPOSURE.	744*6*061	8559*18*67	021014103*	31409*2793	4292*43964	9110100000	024640187
101	SEQUENCE.	ALDQKVVNT DTYENPYNN	EWALQTAVTY	QPVSVALDA	GDFAKQYASG	IFNGPCTAV	TDGVDFWIV KNSTWDTWGE EGWMRLRN
	SUMMARY. ....H-ALPHA-HELIX....B-BETA-STRAND....B-BETA-BRIDGE....G=3-HELIX....J=5-HELIX....T=3-4... OR 5-TURN....S=BEND....						

TABLE AIII (continued)

35) 8PAP PAPAIN [HYDROLASE: SULFHYDRYL PROTEINASE] [PAPAYA FRUIT LATEX: CARICA PAPAYA] . . . . .	B	C B
SHEET . . .	AAA	
BRIDGE1..	D	
BRIDGE2..	BBB	
CHIRALITY.	+----+---+---+ +---+---+	
BEND . . .	SS SSSSSS S	
5-TURN . . .		
4-TURN . . .		
3-TURN . . .	>3X>3X<3<	
SUMMARY..	TT GGGTTS S FEEF	
EXPOSURE..	5521.00084 8005034*	
SEQUENCE..	CGAGTCGAT NPSYPVKY	
1		
36) IFAB LAMBDA IMMUNOGLOBULIN FAB (NEW) [HUMAN] . . . . .		
SHEET . . .	A BB BB AA AAA	BB
BRIDGE1..	BB BBB	BBB
BRIDGE2..	GG GFF	BBB
CHIRALITY.	A dd dd A	EEEEE
BEND . . .	-+---+---+ +---+---+ +---+---+ +---+---+	FF
5-TURN . . .	X<3<	
4-TURN . . .	X<3<	
3-TURN . . .	X<3<	
SUMMARY..	GTTSE EEEE	
EXPOSURE..	00130112063 5*	
SEQUENCE..	GLYTSF PV KN	
1		
SHEET . . .	AAAA	
BRIDGE2..	D	
BRIDGE1..	BBBB	
CHIRALITY.	+---+---+---+	
BEND . . .	S SS	
5-TURN . . .		
4-TURN . . .		
3-TURN . . .		
SUMMARY..	GTTSE EEEE	
EXPOSURE..	00130112063 5*	
SEQUENCE..	GLYTSF PV KN	
1		
SHEET . . .	H=ALPHA-HELIX. . . . .	
E=BETA-STRAND. . . . .	B=BETA-BRIDGE. . . . .	
Q=3-HELIX. . . . .	I=5-HELIX. . . . .	
T=3-, 4-, OR 5-TURN. . . . .	S=BEND. . . . .	
37) 1FAB LAMBDA IMMUNOGLOBULIN FAB (NEW) [HUMAN] . . . . .		
SHEET . . .	A BB BB AA AAA	BB
BRIDGE1..	BB BBB	BBB
BRIDGE2..	GG GGG	BBB
CHIRALITY.	A EEEE	CCCC
BEND . . .	-+---+---+ +---+---+ +---+---+ +---+---+	CC
5-TURN . . .	X<3<	
4-TURN . . .	X<3<	
3-TURN . . .	X<3<	
SUMMARY..	GTTSE EEEE	
EXPOSURE..	00130112063 5*	
SEQUENCE..	GLYTSF PV KN	
1		
SHEET . . .	S B SEE EE EEE TTTT SS EEEEEE	SS SSEEPE EEEEEE
EXPOSURE..	93*1237*72 102778*170 5052*6803 5676150*681 886390*3*3	6*5*5*2676 8654*02010 530**1211 0101064*41 101026*04
SEQUENCE..	XSVLTOQPSV SGACPORVNI SATGSSSNAG AGNHRKWKIQQ LPGPAPKLLI FHNNRFRSVS KSGSATLAI TGLOEDAD YIAQYDTSL RYFGGTRLT	
SHEET . . .	B C CC	CCCC CCC
BRIDGE2..	III III	KK JJJ

CHIRALITY +---+  
 BEND... S S S SS SSSSSSSS +---+  
 BRIDGE2.. S S S SS SSSSSSSS +---+  
 BRIDGE1.. BB BB BB BB BB  
 CHIRALITY +---+  
 BEND... S SS  
 4-TURN... < >>> XXXX<<<  
 3-TURN... < >>> S EEE S HH HHHHHS E EEEE S SS TTT SSSE S S S EE FFFEE SSE EEEE SS SSS  
 SUMMARY.. 3\*91\*0925\*\*\*5\*7552 12\*1029210 00202071\*7 0\*\*2\*33372\* 2\*4\*+5921 000001065 0102016198 216\*026807 45\*777910 >3  
 EXPOSURE.. SEQUENCE. 0013412@03 5\*  
 201 SEQUENCE. GLYSFSFPPV 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...  
  
 SHEET... AAAA  
 BRIDGE2.. D  
 BRIDGE1.. BBBB  
 CHIRALITY +---+  
 BEND...  
 4-TURN...  
 5-TURN... X<3<  
 3-TURN...  
 SUMMARY.. GTTS EERE  
 EXPOSURE.. 0013412@03 5\*  
 201 SEQUENCE. GLYSFSFPPV 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...

36) IFAB LAMBDA IMMUNOGLOBULIN FAB (NEW) {HUMAN}.....  
 BRIDGE2.. A BB BB AA AAA BBBBBB BB ..... BB .....

BRIDGE1.. A dd dd BB BBB EEREE FF ..... FF .....

CHIRALITY -+-+  
 BEND... S S SSS SS SSSSS SS SSS  
 5-TURN... 5555<  
 4-TURN...>3<  
 3-TURN...>3<  
 SUMMARY.. S B SEE BE TIS EEE FEE TTTT SS EEEEEE SSSS EE SS SEEERE BETTERERE S STT EE EEEEEE EEEEEE  
 EXPOSURE.. 93\*1\*23\*72 702778\*70 5052\*6\*603 8869\*6\*43\* 6\*6\*8654\*6\*276 \*8654\*2010 530\*4\*1211 01020441 10102004  
 1 SEQUENCE.. XSVLTQPSV SGACPQRVVI S@GSSNSN AGNWKWVHQI LPGTAPEKLI FHNNRARSYS KSGSATLAI TGLOREAD Y@QSYDSL RVFGGTKL  
 SHEET... B C CC CCCC CCC DDDDD D CCC CCC CCC CCC CCC  
 BRIDGE2.. d H HH IIII IIII M JJJ KK JJJ  
 BRIDGE1.. -+-+-- SSS SSSSS SSS SSS  
 CHIRALITY +--+  
 BEND... S  
 5-TURN...>4> X4<4<  
 4-TURN...>3>3<  
 3-TURN...>3<  
 SUMMARY.. ES EEE TTTT EEEEEE SSS SSSSSS  
 EXPOSURE.. 17\*39\*3616 0221451\*70 7\*8\*11100 002505173 \*33161879 4\*\*12563\*4 5570\*34410 001605729\* \*\*66\*24104 041848\*6  
 101 SEQUENCE.. VLRQFKAPS VTLFPSSSE LOQNATLVD LISOFYPCAV TWAHKDSSP VKAGUETTP SKQSNNKFAA SSYLSLTPQ WKSNSYSDQ VTHEGSTVE  
 SHEET... DD EEEEEE F F PPP P PPP P PPP P  
 BRIDGE2.. NN 00000 r r 0 0000 SSSSSS SSS SSS  
 BRIDGE1.. -+-+--+ SSS SSS SSS SSS  
 CHIRALITY +--+  
 BEND... S  
 5-TURN...>44 <  
 4-TURN...>3<  
 3-TURN...>3<  
 SUMMARY.. FE SS EEEEEE EEEEEE TTT EEEEEE EEEEEE TTT EEEEEE EEEEEE EEEEEE EEEEEE  
 EXPOSURE.. 6347\*4\*6\* 7\*2\*57214 2\*58\*40093 04156293\* 26080224\* 8\*31610010 3\*\*5754\*7 91\*437485 6\*\*67024\* 046168830  
 201 SEQUENCE.. TVAPPIEcSix VQLESGPQL VRPSRSLT dTVGSGTSFN DYTWVRQPP GRGLEWIGV FYHGSIDDT PLRSNVTMLV N'STRNQFSSR LSSTVAA  
 SHEET... GGGGGG GG G GGFF H I J JJJJ JH KKKK  
 BRIDGE2.. WV WV ZZZZZZ  
 BRIDGE1.. TTTTTT VV V VV VV Z  
 CHIRALITY -+-+-- S S S S SSS SSS S  
 BEND... SS S  
 5-TURN...  
 4-TURN...>44 <  
 3-TURN...>3<  
 SUMMARY.. EEEEEE SS S EEEEEE SS B B B SS SSS EEEEEEBS S EEEEEEBS S EEEEEEBS S EEEEEEBS S EEEEEEBS S EEEEEEBS S  
 EXPOSURE.. 301003174 3310821913 4028\*6\*59 2252\*30123 2\*\*396112 100000330 64292413\*6 85\*\*33\*223 30\*3\*\*5640 100293888  
 301 SEQUENCE.. VYGRDLIA GCIWNGQGS LVTVSASTK GPSVFPLAS SKTSVGVA LGELWDKDFP EPVTVSWNSG ALTSGVHFFP AVLGSGLYS LSSTVTPVS  
 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*)

101 SEQUENCE. GTKLOQT

38) 3PGM PHOSPHOGLYCERATE MUTASE (DE-PHOSPHO) [ISOMERASE] {YEAST: SACCHAROMYCES CEREVISIAE}.....

SHEET 1..  
 BRIDGE 2..  
 BRIDGE1.. AAA B  
 BRIDGE2.. BB  
 CHIRALITY . a f  
 BEND... ---+  
 5-TURN... SSSSS SSSSSSSSS SS SSSSS SSSSS SSSSS SSSSS SSSSS SSSSS SSSSS SSSSS  
 4-TURN... >4<< >55<< >> XXXXXXXX XX<<<  
 3-TURN... .>33< >33<  
 SUMMARY.. EEE B  
 EXPOSURE. 54001104 0755\*552 1133\*9125\* 0\*60100\* 118\*\*36906 1110662521 35008812\*\* 1\*5\*4957\* 1371140641 8321845391  
 1 SEQUENCE. PKLVLVRHGQ SENNEKELFT GWDVQLSAK QDEAARAGE LINEKGIVNL VDVTSLRSA IOTANIALEK ADRLNIPNR SWRLINEHYG DLQKDKAQT

SUMMARY..... H=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... T=5-HELIX... I=5-HELIX... S=BEND....

SHEET 1..  
 BRIDGE 2..  
 BRIDGE1.. ++++++-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
 CHIRALITY ++++++-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
 BEND... SSSSS SSSSSSS S SSSSS  
 5-TURN... 555< 555 5 >55<< >> XXXXXXXX XX<<<  
 4-TURN... 4<< 4>< <<  
 3-TURN... >33< >33<  
 SUMMARY.. HHHHHRS TT S EEEES BHH HHHHHHHHHH HT S EE ESSS S S SS SSSHH  
 EXPOSURE. 8\*\*\*1\*\*5\* 1693\*\*640 3719\*\*4\*93 3\*\*9\*7\*\*2 \*941\*3\*211 248152264 33\*\*4\*12\*51 8930010021 11030124\*1 9517\*625\*  
 101 SEQUENCE. LKFGGEKEN TYRRFEDVPF PPIADSSPF S QKGDERKYYV DPNVLPTES LALVDRLLP YQDVAKLV GRTSMIAAHG NSDLGLVKHL EGISDAIK

SHEET 1..  
 BRIDGE 2.. B AA  
 BRIDGE1.. E BB  
 CHIRALITY +--+ +---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
 BEND... SSS S S S S S S S S  
 5-TURN...  
 4-TURN...  
 3-TURN... >444<  
 SUMMARY.. BSS EE TTTT S B S TT  
 EXPOSURE. \*\*2430.011 1416\*\*\*4\*47 \*99725\*\*6  
 201 SEQUENCE. LNIPGFTIL FELDENIKPS KSYVYDPEA

SHEET 1..  
 BRIDGE 2.. AAAA B  
 BRIDGE1.. bbb  
 CHIRALITY -+---+  
 BEND... ---+  
 5-TURN... S SSS SSSSSSS SSS SSSSS SSS SSSSS SSS SSSSS SSS SSSSS SSSSS SSSSS  
 4-TURN...  
 3-TURN... >>> X XXXXH<<<  
 SUMMARY.. EEEE B HHHH HHHHHHHHH SS EEEE E TTHHHHH HHHS TTEE EEE SSSS BS SS HH HHHHHT EE EE HHHHHH  
 EXPOSURE. \*9930001 130428\*41 1\*20\*71362 \*968\*7110 0002100130 581699020 011201657\* 0120121004 017231630 10001213\*6  
 1 SEQUENCE. APKKEFVGQN WRMNGRKSL GEIHLTDGA KLSADTEEVVC GAFLSIVLDFA RQRLDAKIGV RAQNCVYKPK GAFTGEGSPA MIRDGRAWV ILGHSHSRHV

SHEET 1..  
 BRIDGE 2.. AAAA C  
 BRIDGE1.. AAAA C  
 CHIRALITY .ccc ccc ddd J  
 BEND... aaaa  
 5-TURN... ---+  
 4-TURN... >>> X XXXXH<<<  
 3-TURN... >33< >33<  
 SUMMARY.. EEEE B HHHH HHHHHHHHH SS EEEE E TTHHHHH HHHS TTEE EEE SSSS BS SS HH HHHHHT EE EE HHHHHH  
 EXPOSURE. 5524631007 61230109661 10100001094 405\*806629\* 111\*788610 000113289 65\*23\*71 39109\*334 019\*4\*720  
 101 SEQUENCE. FGSDDBLIGO KVAKAALAEGL GVACGEKL DEPAGITEK VIFQETKALA DVKDMSKVW LAYEPWAIG TGATAPQQA QBVHELRGM LTKHVSDAV  
 SUMMARY..... H=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... T=5-HELIX... I=5-HELIX... S=BEND....

AAAAAA  
fffffff  
e e

TABLE III (*continued*)

**BRIDGE1..** JU LL LLLL FFFFF NN JJ EEEE OO G GGGGGGG  
**CHIRALITY** +-+---+--+ +---+--+ +---+--+ +---+--+ +---+--+ +---+--+ +---+--+ +---+--+  
**BEND...** S SSS  
**5-TURN...**  
**4-TURN...**  
**3-TURN...** >33<  
**SUMMARY..** EES FE EEEEEE TT TTTT LLLLLL FFFFF NN JJ EEEE OO G GGGGGGG  
**EXPOSURE.** 801005313 12190722\*6 \*\*6.97\*3\*4 8\*6\*\*3\*61 477448147 \*47230\*56 \*\*4721802 010314623 564\*67312 3\*31106\*  
**101 SEQUENCE.** KETNTLWSN PFKLKNST HTDQKDL NF NSKQQLDF DGNLELTWS NGSPPGSSV GRALYAPVH INESSAVTS FDTAPFLX

**SHEET1...** A AAAAA  
**BRIDGE2..** DDD  
**CHIRALITY** -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
**BEND...** SSSS SS SSS S S  
**5-TURN...**  
**4-TURN...**  
**3-TURN...** >33<  
**SUMMARY..** EEEEEE S SSEEEES BE S SSSSEE S SSS  
**EXPOSURE.** 15\*10106 00208\*29 41\*400930 0107967  
**101 SEQUENCE.** SDSDHADGI AFFINIDS IFSGGTGRL GLFPDAN

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

41) ICAC CARBONIC ANHYDRASE FORM C [CARBON-OXYGEN LYASE CARBONATE DEHYDRATASE] {HUMAN ERYTHROCYTES}.....1CAC  
**SHEET...**  
**BRIDGE2..**  
**BRIDGE1..**  
**CHIRALITY** +-+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+  
**BEND...** SSSSSS SSS SSS  
**5-TURN...**  
**4-TURN...**  
**3-TURN...** >33<> 3<<X3>>><<  
**SUMMARY..** SSTSIS GG TTS GCG G SS SS EE TTT S EEEE TT S EEEE S S SSSSBEEE S S EEEE EEEE S S  
**EXPOSURE.** \*5143\*433 \*51\*\*\*7913 \*3\*\*005151 65\*9\*7\*3\* 4\*73\*561\* 2606\*03171 5003450749 \*6811\*1\*22 2\*53951431 211104\*865  
**1 SEQUENCE.** HMGYCKHDCGP EHMKHDFPAA KGEROQPVII DHTAKYDS LKPILSVSYCQ ATSLRLNDG EDKAVLKGGGP LDGTYRLIGP HFHMGLDGG

**SHEET...** AA A BBBB BBBB BBB B BB B BB  
**BRIDGE2..** B IIIIII H IIIII H KK KK  
**CHIRALITY** +-+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
**BEND...** SS SSS S S SSSSSS SSSSS  
**5-TURN...**  
**4-TURN...**  
**3-TURN...** >33<  
**SUMMARY..** SSEEETBB SEEPEEEEG EGGGSSHHHH GTTSSEEE EEEEDDES RHHHHHGCG GTTSSSE E E E F F F F F F F F F F F F F  
**EXPOSURE.** 001136\*8 5010011102 1.3\*7666590 3\*\*900001 0011\*53873 9214\*1051 \*6071\*5811 \*28\*692511 52\*2860212 4004013765  
**101 SEQUENCE.** GSQHIVDKRK YAAELHLVHW NTKYKEDFGRA VQQPQGLAVL GIFLKVGSKA PGLOQKVDSL DSIKTKGNSA DFTNEDPRGL LPESLDYNTY PGSLUTTPPL

**SHEET...** BBBB BBBB C C BB  
**BRIDGE2..** NN L IIIIII O O MM  
**CHIRALITY** +-+---+--+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
**BEND...** S SS SSS SSS S S S S S S S  
**5-TURN...**  
**4-TURN...**  
**3-TURN...** >4>< 4<<  
**SUMMARY..** S EEEE SS EEEFHHHH HHTT BSS T S B S EE  
**EXPOSURE.** 501360107 42504\*128 614\*355\*6\*7\*\*0350 8144468\*4 \*36129  
**201 SEQUENCE.** ECVTVWVLE FISVSSEQWL KF RKLNDFQ GEPEULMVN WRPAQPLKNR QIKAFP

42) IDPR DIHYDROOLATE REDUCTASE [OXIDOREDUCTASE: NADPH/DONR, DIHYDROOLATE/ACCPTR] [BACTERIAL: LACTOBACILLUS CASEI].....1DPR  
**SHEET...** AAAA A AAA  
**BRIDGE2..** bbbbbb  
**BRIDGE1..** aa  
**CHIRALITY** -+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
**BEND...** S S S S SSS S S S S S S S S S S S S S S S S  
**5-TURN...**  
**4-TURN...**  
**3-TURN...** >33<  
**SUMMARY..** FEEBEBTT B SBSSSS S S HHHHHHH HHHHATSEEE HHHHATTS SSS SSS S S S S S S S S S S S  
**EXPOSURE.** 32002121\*\* 21315\*598 2\*\*06549\*\*14 \*65052052 01\*\*3075\*\*4400 041\*9\*\*3\*1 \*524548946 \*7\*202021  
**1 SEQUENCE.** TAPLAQHN GLIGRDGHF WHLPFDLHFE RAQTQYKIM VGRRTYESP KRPERTIN VLTHQEDYCA QGAVVYHVA AVFATAKQL DQELVIAAGA

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

TABLE III (continued)

SHEET. . . .	AAAAAA	B	A A AAAA	AAAAAA
BRIDGE2. . . .	GGGGGG bbbbb	I	H H HHHH	HHHHH GGGGGG
BRIDGE1. . . .	+-----+ CHIRALITY	-+---+ BEND. . . .	-+---+ 5-TURN. . . .	-+---+ 4-TURN. . . .
3-TURN. . . .	>>XX<<<	S	S S S	SSS S
SUMMARY. . . .	HHHHHHHTS	SEEFFFFF	SB	>44 4< >33 <
EXPOSURE. . . .	*20852881	631206071	68*387*22*	28992986 7**3*5*6*3 2202223* **
161 SEQUENCE. . . .	QIFAFKDNV DFLVTRLAG SFEGDKRIP LNWDFTVS SRTVEDTPA LTHYEVQK KA			
43) 1GPD D-GYCERALDEHYDE-3-PHOSPHATE DE HYDROGENASE [OXIDOREDUCTASE: ALDEHYDE/DONR, NAD/ACCPTR] [LOBSTER: HOWARUS AMERICA1GPD				
SHEET. . . .	AAA	AA	B	CCCC CCC AA
BRIDGE2. . . .	bb	bb	G	HHHH HHHH III cc
BRIDGE1. . . .	aaaa	bb	G	-+---+ +---+ CHIRALITY
BEND. . . .	S SSSSSSSS	SSSS SSSS	S S S	-+---+ 5-TURN. . . .
4-TURN. . . .	>555 5< 5555X5 55<<	>>>XX X<<<444<	SSSS SSSS S	>555< 555<
3-TURN. . . .	>3	>3<<	EE TTS	>44 <44<
SUMMARY. . . .	EEBE TTT HHRRHHHHHHH TTT	EE TTT HHRRHHHHHHH BTTB	S EEEF TEPEEEEE	>44 <44<
EXPOSURE. . . .	551011235 3011014014 *535752001 237247412 60159@3164 82*2*29*# 81131*9*1 533*6*#9* 0308*73161 000155*541	237247412 60159@3164 82*2*29*# 81131*9*1 533*6*#9* 0308*73161 000155*541	EE SSSTT S TTTT E EEE	S S S
1 SEQUENCE.	SKGUDMKFGK IGRVLMAL SGQAVWAN DPALELWV YMFEDTSRG VKSEKVRMED GALVDGKRI TVFNEMPEN IPMSKAGAY IVESTGVPTT			
SHEET. . . .	AAAAAA	A	AAA	D DDDD
BRIDGE2. . . .	eeee	f	f	LLLL
BRIDGE1. . . .	dddd	-+---+	-+---+	J
CHIRALITY	+-----+ +-----+	+-----+ +-----+	+-----+ +-----+	KKKK
BEND. . . .	SSSSSSSS S S	SS SSS	SS SSSSSSSS	P
5-TURN. . . .	>555<	>44 <44<	SSSS SSSS	P
4-TURN. . . .	>>XX<< <	>3><3<	SSSS SSSS	SS S
3-TURN. . . .	HHHHHHHH S SEEFFF	SSS B TTTT	EEE EEE	SS S
SUMMARY. . . .	4*80328*2 21*3001226 1*74331611 13**16*4 600003110 1.080111511 4**2*1*1 1.08010148* 2*3.12*53.6* *212426276	SEEFF TTTT	>33 < 3X33<	SS S
EXPOSURE. . . .	TERASAKI GAKKVISAP SADAPMVFVG UNLKYSMD TVVSNAASC TT MTTVHAVAT QKTVDFGSAK DWNGRCQAQ			
101 SEQUENCE.				
SHEET. . . .	DDDD	DDDD	F F F	F FF
BRIDGE2. . . .	NNNN	NNNN	F F	DDDD
BRIDGE1. . . .	KKKK	LJ	F F	DDDD
CHIRALITY	-+---+ -+---+	-+---+ -+---+	-+---+ -+---+	FF
BEND. . . .	S SSSSSSSS SSS	S S	SSSSSSSS SSSS	FF
5-TURN. . . .	>>><<<	>>>XXXXXX <<<	>>>XXXXXX <<<	FF
4-TURN. . . .	>>><<<	>33<	>33 < 3X33<	FF
3-TURN. . . .	EEEE HH HHRRHHSSS SSS	EEE EEE S	HHHHHHHH HHHTTTE EEE S	SS S
SUMMARY. . . .	5*59164502 *0009216*0 66*382065 3653000304 136905*23 28913321*7 31*3*2*342 319**493 92637*6110 11*52*88	TTT	SS E EESSSTT EEE	SS S
EXPOSURE. . . .	NILPSSTGAA KAKKVIFEL DGLTGAFR VP LPDVSVD LTWRIGCEKS YDDIKAMKT ASEGPOLGF GTEDDVWSS DF IGDRNSSI FDRAKGQIQLS			
201 SEQUENCE.				
SUMMARY. . . .	H=ALPHA-HELIX. . . . S=BETA-STRAND. . . . B=BETA-BRIDGE. . . . G=3-HELIX. . . . T=5-HELIX. . . . P=3-. . . . A=. . . . OR 5-TURN. . . . S=BEND. . . .			

SHEET. . . .	DDDD	DDDD	FF	FF
BRIDGE2. . . .	OOO	NNNN	FF	FF
BRIDGE1. . . .	NNNN	NNNN	FF	FF
CHIRALITY	+-----+ +-----+	+-----+ +-----+	+-----+ +-----+	FF
BEND. . . .	SS	SSSSSSSS SSSSSSSS	SSSSSSSS SSSSSSSS	FF
5-TURN. . . .	>>><<<	>>>XXXXXX <<<	>>>XXXXXX <<<	FF
4-TURN. . . .	>>><<<	>33 <	>33 < 3X33<	FF
3-TURN. . . .	TTTEEE	HHHHHHHH HHHTTTE EEE S	HHHHHHHH HHHTTTE EEE S	SS S
SUMMARY. . . .	85*1711631 012312010 0311966*3 28*	TTT	SS E EESSSTT EEE	SS S
EXPOSURE. . . .	KTFVKVSWY DNFKGFSQRV TDLKHMQRV DSA			
301 SEQUENCE.				

44) 4ADH APO-LIVER ALCOHOL DEHYDROGENASE [OXIDOREDUCTASE: CHOH/DONR, NAD/ACCPTR] [HORSE LIVER: EQUUS CABALLUS] . . . . . 4ADH SHEET. . . . AAAA AAAA CCCC CCCC

101 SEQUENCE. IERASAHFKG GAKKVISAP SADAPMVC VNLKYSKDM TVVNSASC TT NCLAPAVKL HENPTEVGL MTYHAYTAT QKTVDDGFPAK DWNGNGAAQ

SHEET...  
BRIDGE1.. DDDDD  
BRIDGE1.. NNNNN  
CHIRALITY ..  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
EXPOSURE. 5\*3916502 \*00092126\* 66138605 3653000304 120995\*23 24913371# 313\*2342 3319\*44 116\*54\*88  
201 SEQUENCE. NIPISSFGAA KAGKVIPEL DGRITGMAR VPPDTSVVD LPRLGRCES YDDIKAMKT ASQPLQGF GLYTEDVVSS DP TGDRSSTI FDAKAGIQLS  
SUMMARY..... H=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... T=5-HELIX... I=5-HELIX... F=3-HELIX... R=5-TURN... S=BEND... .

SHEET...  
BRIDGE2.. DDDDD  
BRIDGE2.. OOO  
CHIRALITY ..  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
EXPOSURE. 85\*1711631 012312610 0311963\*3 28\*  
301 SEQUENCE. RTVKVKVSWY DNPFQGQVR IDLKLKMQKV DSA

44) 4ADH APO-LIVER ALCOHOL DEHYDROGENASE (OXYDOREDUCTASE: CHOH/DONR, NAD/ACCPTR) [HORSE LIVER: EQUUS CABALLUS],..... 4ADH  
SHEET...  
BRIDGE1.. AAAA AAB  
CHIRALITY ..  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
EXPOSURE. \*\*4\*8\*09 00008\*\*\*\* 1475905135 1\*\*100650 40800\*501 301524\*14 526100000 0613234\*51 \*64\*76500 001246\*2  
1 SEQUENCE. STAGKVICK AAVWEEKK PFSIEEVAPK PFAHEVKM VANGICRSDD HVPSGTLYTP LPVIAGHEAA GIVESIGEV TVVRGDVKV PLFPGQKC

SHEET...  
BRIDGE2.. DDD  
BRIDGE1.. C C DD F  
CHIRALITY ..  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
EXPOSURE. \*32\*8\*29 092268\* 031\*5128 167\*7151# 136100050 0012310\*8 867041700 001105202 005811\*18 612000011  
101 SEQUENCE. RVCRKHPGNF CIRNDLSPMR GMQDGTSRF TCRGKPIHFF LGTSTPSQYT VDEVISVAKI DASPLEKVC LICGFSTGY CSAVKVKAT QGSVTCAVF GL

SHEET...  
BRIDGE2.. FF FF  
BRIDGE1.. P PP  
CHIRALITY ..  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
3-TURN...  
4-TURN...  
5-TURN...  
EXPOSURE. 314000020 0\*6145700 01379\*5\*6 60\*1124774 12\*8961 7710462\*4 6020000281 9510133002 001\*5\*0300 332836\*5\*\*  
201 SEQUENCE. GGVLGSVNG CRANGARII GVDINKEKA KARKEVATC EK VHQDQYKPKI QEVLTENSNG GVDSEEVIG RLIVMVTALS CCQEAKGVWV IWGVPPDSQN  
SUMMARY..... H=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... T=5-HELIX... I=5-HELIX... F=3-HELIX... R=5-TURN... S=BEND... .

TABLE III (*continued*)

```

SHEET...          SHEET...
B R I D G E 2... B R I D G E 1...
C H I R A L I T Y... C H I R A L I T Y...
B E N D ...      B E N D ...
5 - T U R N ...  5 - T U R N ...
4 - T U R N ...  4 - T U R N ...
3 - T U R N ...  3 < 3 <
S U M M A R Y... S U M M A R Y...
E X P O S U R E ... E X P O S U R E ...
S E Q U E N C E ... S E Q U E N C E ...

```



TABLE AIII (continued)

SHEET...	A	AA A	H	HHH	H HHHHHH	H
BRIDGE1..	E	dd d	E	R RRR	R RRR	
CHIRALITY	-+---+--+-	++-++-+-+--	-+-----+	++-----+	++-----+	
BEND...	SS SSSS	SSSSS	SSSSSSSS	SSSS SSSS	SSSS SSSS	S
5-TURN...						
4-TURN...			>>4XXX>XX	XXXX<<<	>>>X <<<	
3-TURN...		>33<	>33<>33<	>33<	>>>X <<<	
SUMMARY..	SSB	SSSSSE E STSSS	HHHHHHHH HHHHHHHH	TTT SSS	EEE SSS EEEE	>>3 <<
EXPOSURE.	524274*101	002423*16	4391152087	60054111**	*29377744 444014110	GG SS
SEQUENCE.	FONTNVKGIV AVGDWCGKAL LTPVIAIGR KLAHLFLFYEK EDSKLDDNNI PTYVFSHPPI GTVGILDEA IKYGIENVK TYSTSETPY HAVTAKTKC					
SHEET...	HHHHHHH	HH HHHHH	H			
BRIDGE1..	TTTTTT	TT TTTT	U			
CHIRALITY	SSSSSS	RRRRR				
BEND...	+---+--+-	----+--+-	----+--+--+	----+--+--+	----+--+--+	
5-TURN...	SSS	S	SSSSSS	SSSSS	S SSSSSSSS	
4-TURN...	>55 5<		>>4XXX>X	X <<<	>4<<	
3-TURN...	>33<		>>4XXX>X	X <<<	>4<<	
SUMMARY..	EEEEEETT TEEEEEETT	TTHHHHHHHH HHHHHHHH	>33<	>33<	>>>X <<<	
EXPOSURE.	001000088*	581100000	7403*2053	0730*5111	791657*5**	S 8633505081 *
1 SEQUENCE.	VNNNCCARE ERKVGNHQ LGCDMLQG AVAVNGTK AFEDNTWIK PTSSFEELTL R					
47) 260D CU/ZN SUPEROXIDE DISMUTASE: SUPEROXIDE/ACCPTR] [COW ERTHROCYTE: BOS TAURUS] ..... B BBBRC CAAAAAA						
BRIDGE1..	AAAAB	AAA AAA	BB BBBB	B	BBBRC	CAAAAAA
BRIDGE2..	BB	CCC CC	DDD DDD			
CHIRALITY	CCCCC CCC	CCCC CCC	FF FF F	H	F FFFFJ	JDDDDDD
BEND...	S -----+--	----+--+--+	----+--+--+	----+--+--+	----+--+--+	
5-TURN...	SSS	SSS	SS	SS	SS	S
4-TURN...						
3-TURN...						
SUMMARY..	SSSSBB SSS EEEEEE EEEEEE EEEEEE	SEE EEEEEE	TTCCGTT S B	SS	>33< X3K<	>33< <
EXPOSURE.	8960000401 *4* *04060 43*6*46305	360432*338	1000006503	TT SS	0416*6*10	TT SS
1 SEQUENCE.	ATKNCVKE DGIVQGTIE EAKSDTWVNT GSFIGLTDG HGEVHRQGD NTQGATSAFP HF NPLSKHG GPKEEERHG DGRVTADEKN GVA+VDVP					
SHEET...	D	BBBB	B BBBB AA			
BRIDGE1..	k	IIII	I III			
CHIRALITY	+--+--+--+	+--+--+	+--+--+			
BEND...	S SSSSS	SSS	SS	SS	SS	S
5-TURN...						
4-TURN...						
3-TURN...	>33<	33<	>4 4<			
SUMMARY..	S BSSSB	TTSPEEEESS	SS ST TTTS S	EEEEEE EE		
EXPOSURE.	603293820	1010000**	715716*649*	*07400037	6301110021 *	
101 SEQUENCE.	LISLGSEYS I GRTMVVHEK PDDLGRCNE ESTKGNAGS RTAaGvGIA K					
SUMMARY.....	H=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... I=5-HELIX... T=3-, 4-, OR 5-TURN... S=END... .					

48) IMBN MYOGLOBIN [OXYGEN STORAGE] (FERRIC IRON - METMYOGLOBIN) (SPERM WHALE) ..... IMBN SHEET...

BRIDGE1..

CHIRALITY

BEND...

5-TURN...

4-TURN...

3-TURN...

SUMMARY..

EXPOSURE.

1 SEQUENCE.

IMBN

BRIDGE1..

CHIRALITY

BEND...

5-TURN...

4-TURN...

3-TURN...

SUMMARY..

EXPOSURE.

1 SEQUENCE.

IMBN

1 SEQUENCE.

SHEET...  
BRIDGE1... AAAAEF SSSSSS  
CHIRALITY +---+  
BEND.... S SSSSSS  
5-TURN...  
4-TURN...  
3-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

3>33<  
>33<  
>33<  
>33<  
>33<  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

3-TURN...  
4-TURN...  
5-TURN...  
601293820 103106060\* 71571649\* \*017404037 63011021 \*  
LISLGYSI IGTIVVHEK PDLGRGNEK ESTKTNAGS RLAQVIGHA K

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

1 SEQUENCE.

SHEET...  
BRIDGE2... D D  
BRIDGE1... k k  
CHIRALITY +---+  
BEND.... S SSSSSS  
5-TURN...  
4-TURN...  
3-TURN...>33<  
4-TURN...>44<  
5-TURN...>33<  
6-TURN...>33<  
7-TURN...>33<  
8-TURN...>33<  
9-TURN...>33<  
10-TURN...>33<  
11-TURN...>33<  
12-TURN...>33<  
13-TURN...>33<  
14-TURN...>33<  
15-TURN...>33<  
16-TURN...>33<  
17-TURN...>33<  
18-TURN...>33<  
19-TURN...>33<  
20-TURN...>33<  
21-TURN...>33<  
22-TURN...>33<  
23-TURN...>33<  
24-TURN...>33<  
25-TURN...>33<  
26-TURN...>33<  
27-TURN...>33<  
28-TURN...>33<  
29-TURN...>33<  
30-TURN...>33<  
31-TURN...>33<  
32-TURN...>33<  
33-TURN...>33<  
34-TURN...>33<  
35-TURN...>33<  
36-TURN...>33<  
37-TURN...>33<  
38-TURN...>33<

SEEPEEB SSS EEEEEE EEEEEE EEEEEE SEE EEEEEE  
8966084\*1 \*4-\* \*0460 43-6\*46305 360332\*38 1000003 03 773282\*27 00476\*110 055\*81000 01001159\*7 052\*13\*0\*  
ATRAVCVLKG DGPGQGTIFH EAKDTVVNT GSITGLFGD HGFHVHQGD NTQGATSAGP HFNLRSKKG GPDEERHVG DGNVADRN GVAVIDVP

1 SEQUENCE.

SHEET...  
BRIDGE2... B BBBB A A  
BRIDGE1... I IIII  
CHIRALITY E E B B  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...>33<  
4-TURN...>44<  
5-TURN...>33<  
6-TURN...>33<  
7-TURN...>33<  
8-TURN...>33<  
9-TURN...>33<  
10-TURN...>33<  
11-TURN...>33<  
12-TURN...>33<  
13-TURN...>33<  
14-TURN...>33<  
15-TURN...>33<  
16-TURN...>33<  
17-TURN...>33<  
18-TURN...>33<  
19-TURN...>33<  
20-TURN...>33<  
21-TURN...>33<  
22-TURN...>33<  
23-TURN...>33<  
24-TURN...>33<  
25-TURN...>33<  
26-TURN...>33<  
27-TURN...>33<  
28-TURN...>33<  
29-TURN...>33<  
30-TURN...>33<  
31-TURN...>33<  
32-TURN...>33<  
33-TURN...>33<  
34-TURN...>33<  
35-TURN...>33<  
36-TURN...>33<  
37-TURN...>33<  
38-TURN...>33<

SEEPEEB SSS EEEEEE EEEEEE EEEEEE SEE EEEEEE  
601293820 103106060\* 71571649\* \*017404037 63011021 \*  
LISLGYSI IGTIVVHEK PDLGRGNEK ESTKTNAGS RLAQVIGHA K

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

1 SEQUENCE.

SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

\*15551\*5 5\*01650997 3210026111 \*11\*558912 \*85\*8011\* 68701\*33\*6 07\*054\*6 4103101082 \*0\*78\*74\*0 7458\*8\*25  
VLSSEGWL VHWKAPEAD VAGHQDQD LI RLFSHPETL EKFDRK KILK TEAMKASED LKRHGTVTLT ALGALKKGS HHEAEKLPLA QSHATHKKIP

SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

1 SEQUENCE.

SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

1 SEQUENCE.

48) IMBN MYOGLOBIN [OXYGEN STORAGE] (FERRIC IRON - METMYOGLOBIN) [SPERM WHALE].....IMBN

SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

1 SEQUENCE.

SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

1 SEQUENCE.

49) LECD HEMOGLOBIN (ERYTHROCYTOFIL, DEOXY) [OXYGEN TRANSPORT] (CHIRONOMUS THUMMI).....LECD

SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

1 SEQUENCE.

SHEET...  
BRIDGE2...  
BRIDGE1...  
CHIRALITY  
BEND....  
5-TURN...  
4-TURN...  
3-TURN...  
2-TURN...  
SUMMARY...  
EXPOSURE...  
SEQUENCE.

1 SEQUENCE.

TABLE III (*continued*)

SHEET. . . .  
BRIDGE 2. . .  
BRIDGE 1. . .  
CHIRALITY  
BEND. . . .  
5-TURN. . .  
4-TURN. . .  
3-TURN. . .

ANDER

SHEET...  
 BRIDGE2...  
 BRIDGE1...  
 CHARALITY  
 BEND...  
 5- TURN...  
 4- TURN...  
 3- TURN...  
 SUMMARY...  
 EXPOSURE.  
 101  
 SEQUENCE.  
 52)  
 1BBL LEGHOMOCLORIN (ACETATE, MET) [OXYGEN TRANSPORT] (YELLOW LUPIN ROOT NODULES: LUPINUS LUTEUS L.)  
 1BBL  
 SHEET...  
 BRIDGE2...  
 BRIDGE1...  
 CHARALITY  
 BEND...  
 5- TURN...  
 4- TURN...  
 3- TURN...  
 SUMMARY...  
 EXPOSURE.  
 101  
 SEQUENCE.  
 52)  
 1BBL  
 SHEET...  
 BRIDGE2...  
 BRIDGE1...  
 CHARALITY  
 BEND...  
 5- TURN...  
 4- TURN...  
 3- TURN...  
 SUMMARY...  
 EXPOSURE.  
 1  
 SEQUENCE.  
 52)

TABLE III (*continued*)

-+ + + +  
SSSS

EXPOSURE. Z-1103  
101 SEQUENCE. GSSVFA

56) 3PTI TRIPSIN INHIBITOR [PROTEINASE INHIBITOR] [COW PANCRAS: BOS TAURUS].....3PTI  
 SHEET...  
 BRIDGE2...  
 BRIDGE1...  
 CHIRALITY  
 BEND...  
 TURN...  
 4-TURN...  
 3-TURN...  
 SUMMARY...  
 EXPOSURE.  
 1. SEQUENCE.  
 RDPELEPYVNGPDKARIR IFYNAKAGL QTVFGGDRK KRNTGGA  
 CTRNAAWY

57)	ICTx ALPHA COBRA TOXIN [Cobra: Naja Naja siamensis]. . . . .	ICTx
SHEET. . . .	A A A A A	AAAAA
BRIDGE 2. . .	B B B B B	BBBBB
BRIDGE 1. . .	A A A A A	AAAAA
CHITRALITY . .	--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+	+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
BEND. . . .	SSSSS , SSS	SSS SSS
5-TURN. . . .		SS S S S S
4-TURN. . . .		SS S S S S
3-TURN. . . .	>33<	>4<<
SUMMARY. . . .	SSSSS	>33<
EXPOSURE. . . .	5*#253**#9 55957*1*2# 3249797*1# 94*4*56# 345483*7# 6554*33999 *38*498*	T S STT
SEQUENCE. . . .	IRATPDT SKDPSNGHVA YTWKWDPC SINGKRVDLG baaTqPTVKT GVDIQdestD NeNPFPTKR P	
58)	IMLT MELITTIN [HEMOLYTIC POLYPEPTIDE] {HONEY BEE VENOM: APIS MELLIFERA}	IMLT
MELT		

TABLE AII (continued)

SHEET...	AAAA		AAA	
BRIDGE2...	ddd	d dd	d dd	d dd
BRIDGE1...	bbb	-+-----+	-+-----+	-+-----+
CHIRALITY...	+++++++-+--	+++++++-+--	+++++++-+--	+++++++-+--
BEND...	SSSSSSSS S	SSSSSSSS S	SSSSSSSS S	SSSSSSSS S
5-TURN...	>5555<	>5555<	>5555<	>5555<
4-TURN...	>>>XXXXX	>>>XXXXX	>>>XXXXX	>>>XXXXX
3-TURN...	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<
SUMMARY...	>33<	>33<	>33<	>33<
EXPOSURE...	HIIHHHHH HIIHHHHH HIIHHHHH HIIHHHHH			
101 EXPOSURE.	48520*820 8520*6370 2*316*1* 94*46*8* 86*52*4* *63631980 12*27*365 7*2*699* 734752493 **2*	48520*820 8520*6370 2*316*1* 94*46*8* 86*52*4* *63631980 12*27*365 7*2*699* 734752493 **2*	48520*820 8520*6370 2*316*1* 94*46*8* 86*52*4* *63631980 12*27*365 7*2*699* 734752493 **2*	48520*820 8520*6370 2*316*1* 94*46*8* 86*52*4* *63631980 12*27*365 7*2*699* 734752493 **2*
61) IRHD RHODANSE [THIOSULFATE-CYANIDE SULFURTRANSFERASE] [COW LIVER: BOS TAURUS]	A AAA BB	C BB	D DD	E
SHEET...	AA			
BRIDGE2...	ccc			
BRIDGE1...	b ee	F ee		
CHIRALITY...	-+---+--+--	-+---+--+--	-+---+--+--	-+---+--+--
BEND...	SSSSSSSS S	SSS SSSSSS	SS SSSSSS	SS SSSSSS
5-TURN...	>555555<	>555555<	>555555<	>555555<
4-TURN...	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<
3-TURN...	>33<	>33<	>33<	>33<
SUMMARY...	EE HIIHHHHH HT BTTEEE EE TTTT EE T	EE HIIHHHHH HT BTTEEE EE T	EE HIIHHHHH HT BTTEEE EE T	EE HIIHHHHH HT BTTEEE EE T
EXPOSURE...	*4*4*75*151 51*48*13* 75*51*4141 0*01654*39 7*14*29** 0*0340*3051 9*78*34 5*222*7*34 1*50082004 6*0200102	*4*4*75*151 51*48*13* 75*51*4141 0*01654*39 7*14*29** 0*0340*3051 9*78*34 5*222*7*34 1*50082004 6*0200102	*4*4*75*151 51*48*13* 75*51*4141 0*01654*39 7*14*29** 0*0340*3051 9*78*34 5*222*7*34 1*50082004 6*0200102	*4*4*75*151 51*48*13* 75*51*4141 0*01654*39 7*14*29** 0*0340*3051 9*78*34 5*222*7*34 1*50082004 6*0200102
1 SEQUENCE.	VHQVLYRALV STPKWLAESVR AGKVGPGLRV LIDASWYSPT REARKEYLER HVPGASFDFI BECRDMSAPY EVMPLSEAGP ADYVGSLGTS NUTHVVNGC	VHQVLYRALV STPKWLAESVR AGKVGPGLRV LIDASWYSPT REARKEYLER HVPGASFDFI BECRDMSAPY EVMPLSEAGP ADYVGSLGTS NUTHVVNGC	VHQVLYRALV STPKWLAESVR AGKVGPGLRV LIDASWYSPT REARKEYLER HVPGASFDFI BECRDMSAPY EVMPLSEAGP ADYVGSLGTS NUTHVVNGC	VHQVLYRALV STPKWLAESVR AGKVGPGLRV LIDASWYSPT REARKEYLER HVPGASFDFI BECRDMSAPY EVMPLSEAGP ADYVGSLGTS NUTHVVNGC
SHEET...	AAA	C	DD	
BRIDGE2...	ddd			
BRIDGE1...	a	F	99	
CHIRALITY...	-+---+--+--	-+---+--+--	-+---+--+--	-+---+--+--
BEND...	SSSS SSSSS SSSSS	SS S SSSSS	SS SSSSS	S S SSS
5-TURN...	>5555<	>5555<	>5555<	>5555<
4-TURN...	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<
3-TURN...	>33<	>33<	>33<	>33<
SUMMARY...	SPHS HHIIHHHTH EEEPTTHH HHIIHHHTH B			
EXPOSURE...	2*5*0161*000 0*00114*002 3*00014*003 1*01976*1 5*3353*061* 5*7*4526*6 0*12*5471* *175*717	2*5*0161*000 0*00114*002 3*00014*003 1*01976*1 5*3353*061* 5*7*4526*6 0*12*5471* *175*717	2*5*0161*000 0*00114*002 3*00014*003 1*01976*1 5*3353*061* 5*7*4526*6 0*12*5471* *175*717	2*5*0161*000 0*00114*002 3*00014*003 1*01976*1 5*3353*061* 5*7*4526*6 0*12*5471* *175*717
161 SEQUENCE.	DQDGFYAP WMMMPVFSH RTVYUNGEF RNWLGEHV TSEPVPERA IPEKATNRSI LKTYQVLEN LSKRFQLVD SPAGSYLT QPEPAVGD	DQDGFYAP WMMMPVFSH RTVYUNGEF RNWLGEHV TSEPVPERA IPEKATNRSI LKTYQVLEN LSKRFQLVD SPAGSYLT QPEPAVGD	DQDGFYAP WMMMPVFSH RTVYUNGEF RNWLGEHV TSEPVPERA IPEKATNRSI LKTYQVLEN LSKRFQLVD SPAGSYLT QPEPAVGD	DQDGFYAP WMMMPVFSH RTVYUNGEF RNWLGEHV TSEPVPERA IPEKATNRSI LKTYQVLEN LSKRFQLVD SPAGSYLT QPEPAVGD
SHEET...	E	DD	D	E
BRIDGE2...	DD	jj	jj	
BRIDGE1...	K 11	L	hh	
CHIRALITY...	+---+--+--	-+---+--+--	-+---+--+--	-+---+--+--
BEND...	SS	SSSS SSS	SS S SSS	SS SSS S
5-TURN...	>5555<	>5555<	>5555<	>5555<
4-TURN...	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<	>>>XXX<<<
3-TURN...	>33<	>33<	>33<	>33<
SUMMARY...	B TT EE TTTB TS B HHIIHHH HHIIHHH TT S EEEPE SSS STTHHHHHH HHHT TT EE ETTHHHHHH HS GCGSB S S	B TT EE TTTB TS B HHIIHHH HHIIHHH TT S EEEPE SSS STTHHHHHH HHHT TT EE ETTHHHHHH HS GCGSB S S	B TT EE TTTB TS B HHIIHHH HHIIHHH TT S EEEPE SSS STTHHHHHH HHHT TT EE ETTHHHHHH HS GCGSB S S	B TT EE TTTB TS B HHIIHHH HHIIHHH TT S EEEPE SSS STTHHHHHH HHHT TT EE ETTHHHHHH HS GCGSB S S
EXPOSURE...	1830*615*0 2297*14*0 5253*9*05 40*8*16* *6000*0240 0*0600*0800 1551*180*00 0*00326* *1672*35*5*	1830*615*0 2297*14*0 5253*9*05 40*8*16* *6000*0240 0*0600*0800 1551*180*00 0*00326* *1672*35*5*	1830*615*0 2297*14*0 5253*9*05 40*8*16* *6000*0240 0*0600*0800 1551*180*00 0*00326* *1672*35*5*	1830*615*0 2297*14*0 5253*9*05 40*8*16* *6000*0240 0*0600*0800 1551*180*00 0*00326* *1672*35*5*
201 SEQUENCE.	SGHRSVNM FMDLTLENG FERSEPEELRA MFEAKKVDLT KPLIACTRGS VTA�IALA YLGCPDVAI YOGSPFEMFH RAPPETWVSQ GRG	SGHRSVNM FMDLTLENG FERSEPEELRA MFEAKKVDLT KPLIACTRGS VTA�IALA YLGCPDVAI YOGSPFEMFH RAPPETWVSQ GRG	SGHRSVNM FMDLTLENG FERSEPEELRA MFEAKKVDLT KPLIACTRGS VTA�IALA YLGCPDVAI YOGSPFEMFH RAPPETWVSQ GRG	SGHRSVNM FMDLTLENG FERSEPEELRA MFEAKKVDLT KPLIACTRGS VTA�IALA YLGCPDVAI YOGSPFEMFH RAPPETWVSQ GRG
SUMMARY.....	N=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... I=5-HELIX... T=3-4... OR 5-TURN... S=END... .	N=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... I=5-HELIX... T=3-4... OR 5-TURN... S=END... .	N=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... I=5-HELIX... T=3-4... OR 5-TURN... S=END... .	N=ALPHA-HELIX... E=BETA-STRAND... B=BETA-BRIDGE... G=3-HELIX... I=5-HELIX... T=3-4... OR 5-TURN... S=END... .



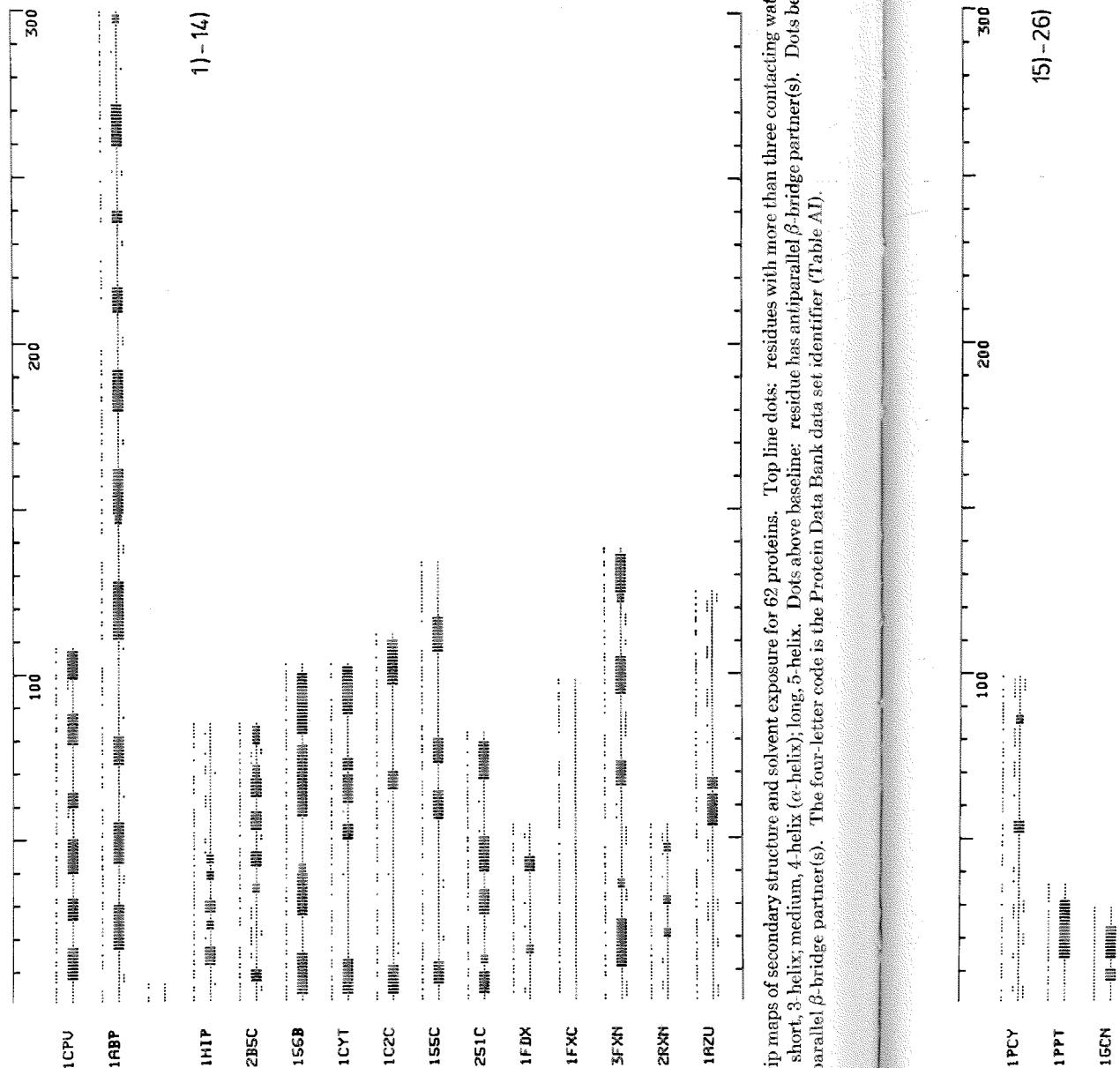


Fig. A1. Strip maps of secondary structure and solvent exposure for 62 proteins. Top line dots: residues with more than three contacting water molecules. Vertical bars: short, 3-helix; medium, 4-helix ( $\alpha$ -helix); long, 5-helix. Dots above baseline: residue has antiparallel  $\beta$ -bridge partner(s).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

1FDX
1FXC
3FXN
2RXN
1RZU

Fig. A1. Strip maps of secondary structure and solvent exposure for 62 proteins. 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 A1).

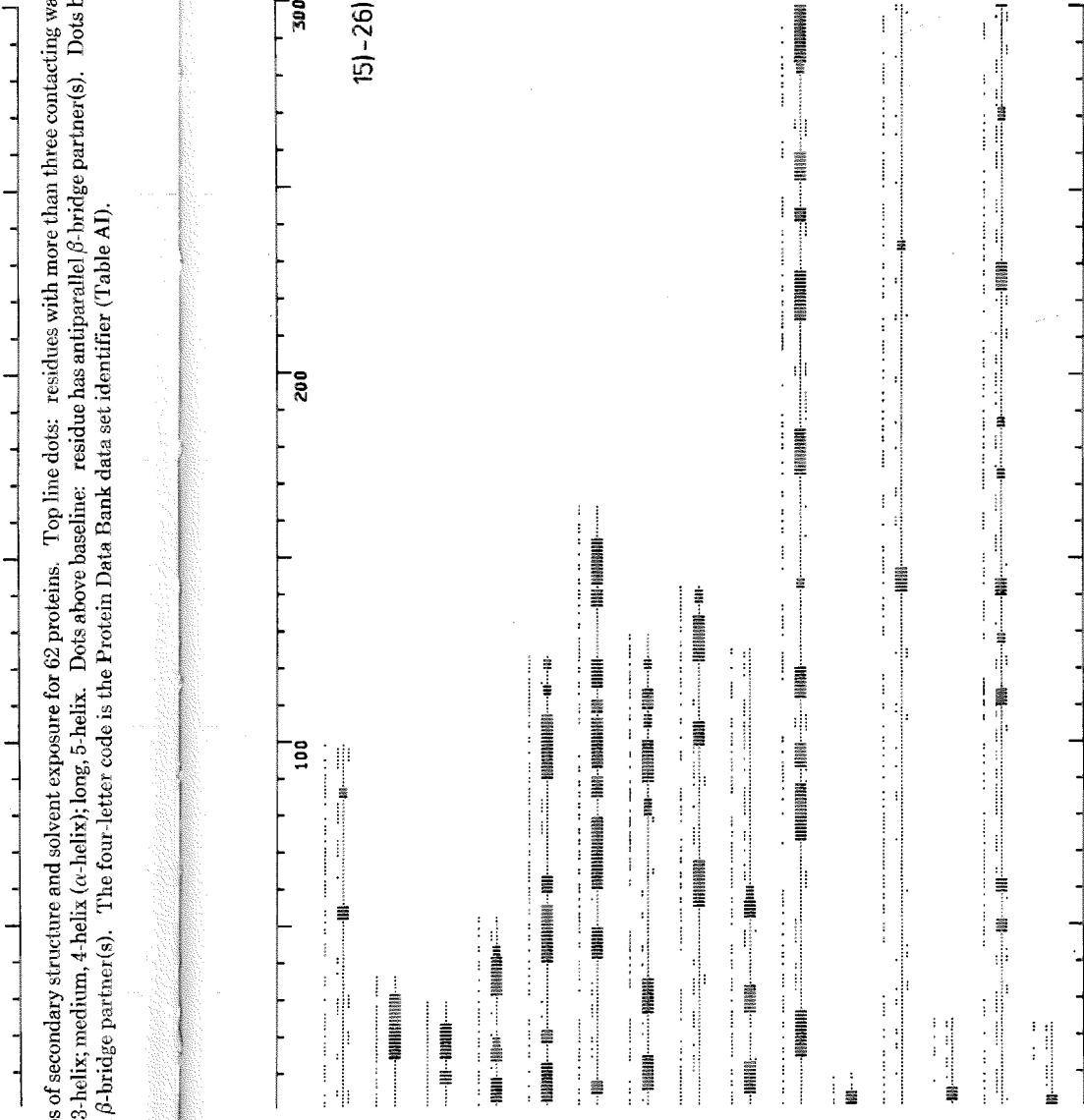


Fig. A1. (continued from the previous page)

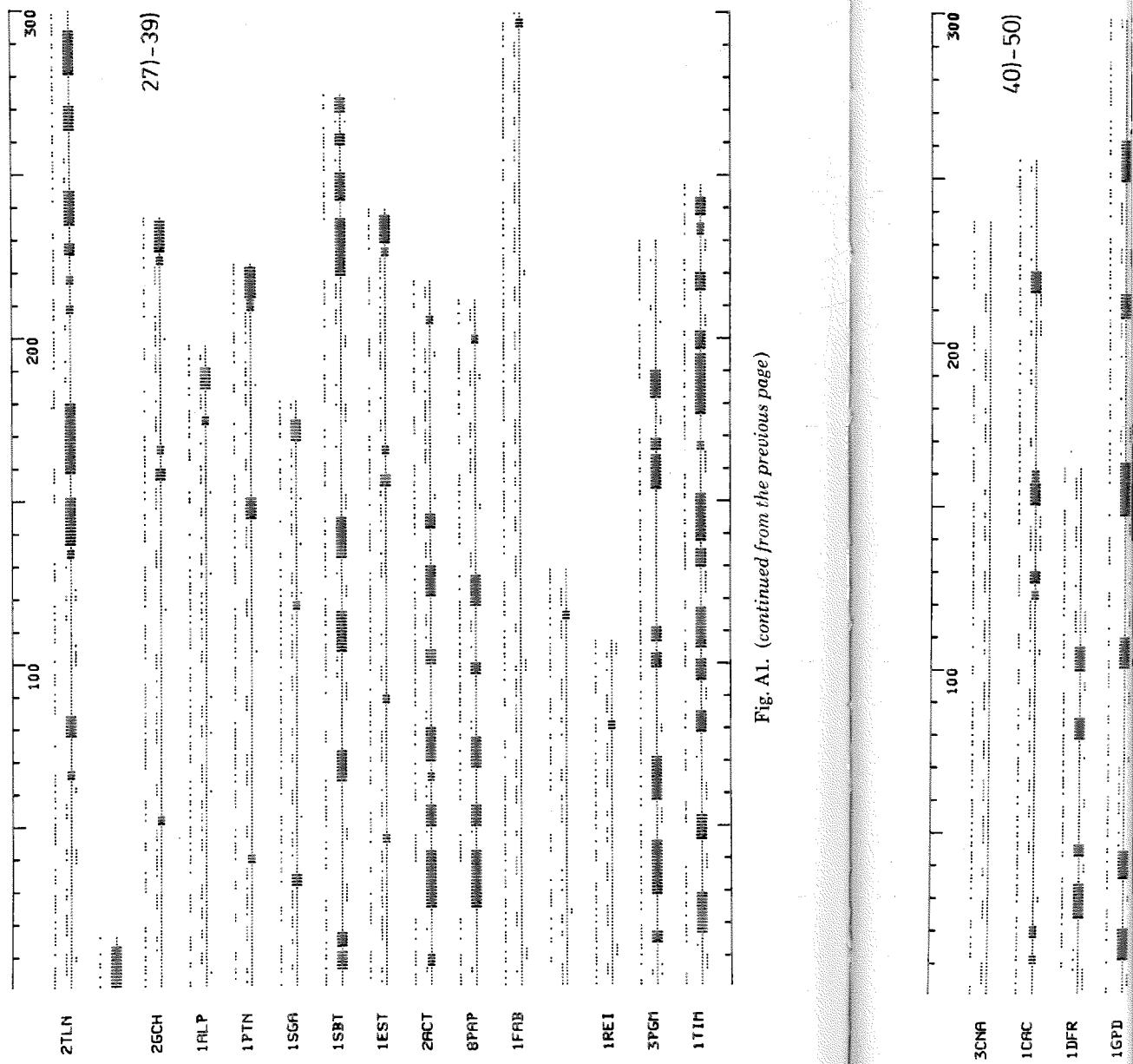


Fig. A1. (continued from the previous page)

SPAP  
1FAB  
1RE1  
3P6H  
1T1H

Fig. A1. (continued from the previous page)

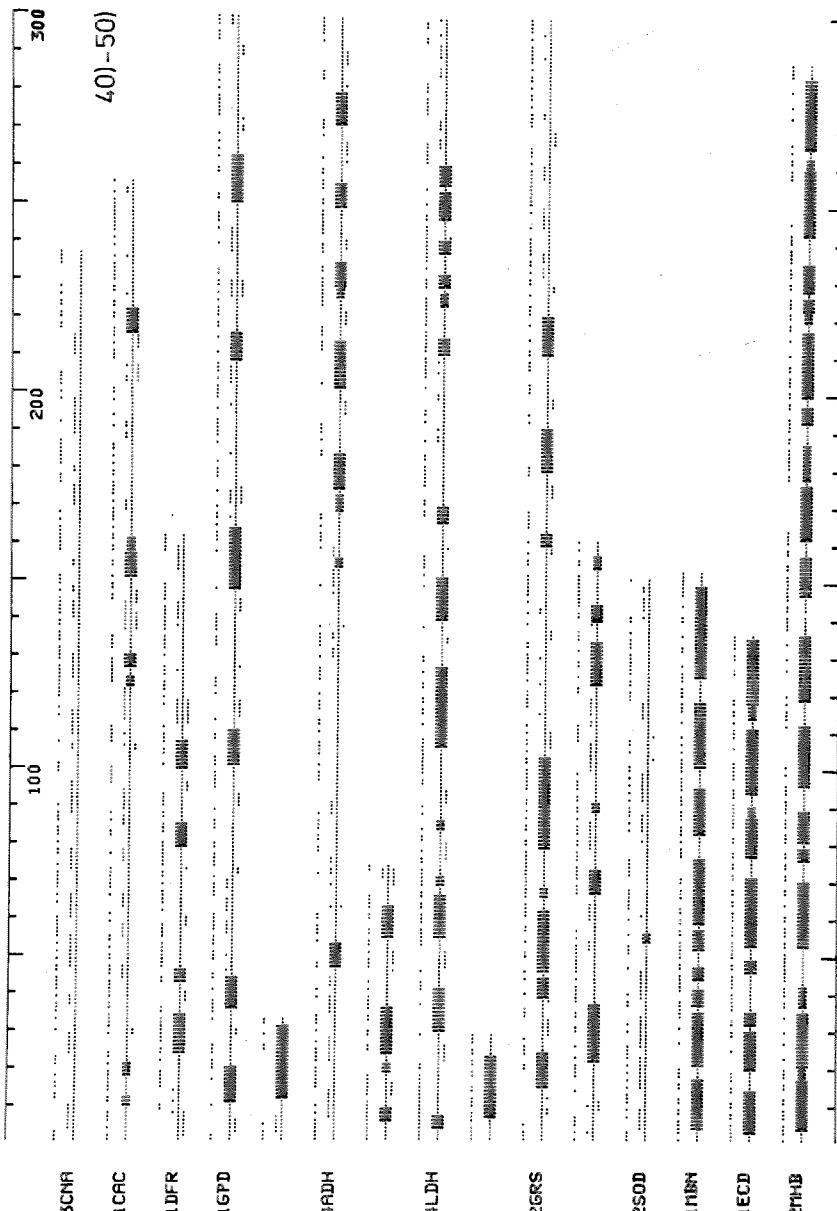


Fig. A1. (continued from the previous page)

For reasons of  
tallographers on  
Data Bank. C. O  
gemeinschaft gave

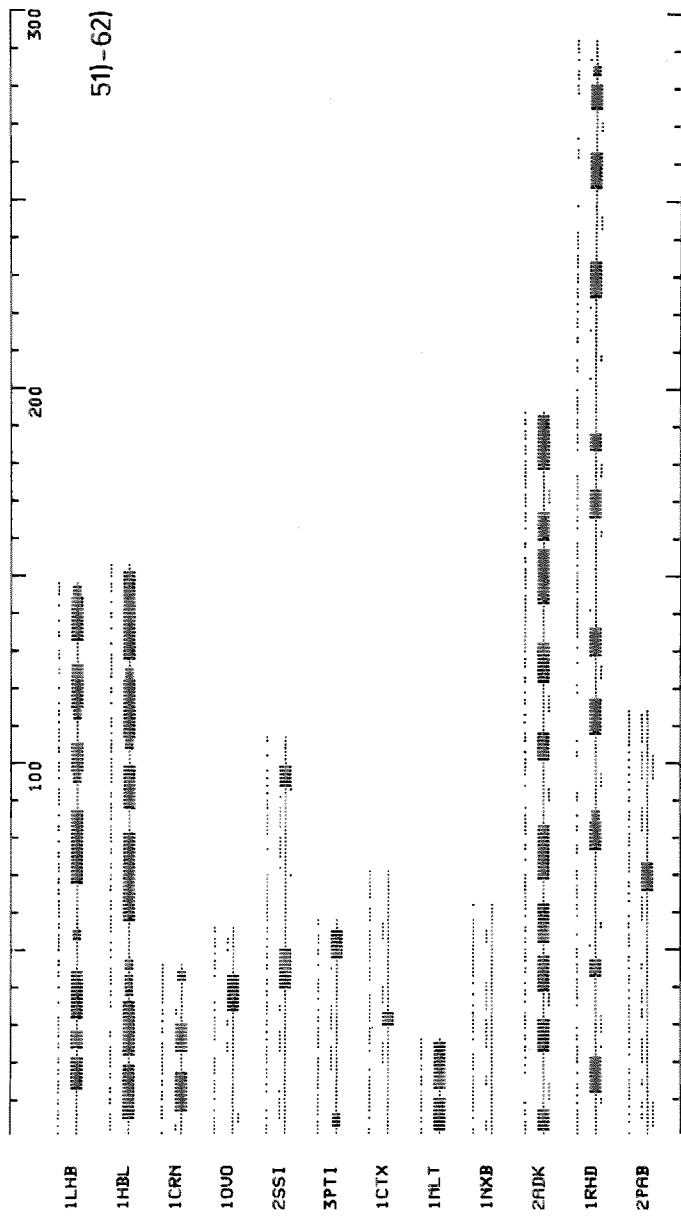


Fig. A1. (continued from the previous page)

1. Pauling, L.
2. Bernstein, E.
- J. R., Kennard, O.
3. Kuntz, I. (1)
4. Lewis, P. N.  
211-229.
5. Rose, G. D.
6. Chou, P. Y.
7. Smith, J. A.
8. Richardson,
9. Lifson, S. &
10. Lee, B. & R.
11. Levitt, M. &
12. Feldmann,  
Inc., Rockville, M.
13. Lifson, S., I.
14. IUPAC-IUP  
6489-6497.
15. Rackovsky,
16. Shrake, A.
17. Chothia, C.
18. Bode, W., S
- Vol. 40, Desnuelle
19. Finney, J. L.
- Press, New York,
20. Brant, D. A.
21. Schellman,  
53-61.
22. Provencher,
23. Williams, R.
24. Hennessey,
25. Deisenhofer
26. Wuethrich,
27. Timkovich,
28. Schulz, G. E
29. Wilson, I. A.
30. Diamond, R.
31. Diamond, R.
32. Hendrickson
- R., Ramaseshan, S
- 13.01-13.23.
33. Dodson, E.  
311-315.
34. Jack, A. & L.
35. Levitt, M. &
36. Agarwal, R.
37. Chambers, J.
38. Sussman, J.
- Sect. A 33, 800-80

Received December 1983  
Accepted May 1984

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. (1978) *J. Mol. Biol.* **129**, 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. Scheilman, 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

Fig. A1. (continued from the previous page)