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^{α} 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 Corey¹ 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 Bank² are often subjective and sometimes incomplete. Objective algorithms exist, e.g., for defining turns³-6 (reviewed in Refs. 7, 8), β -sheets,³ and solvent accessibility,¹0 but only Levitt and Greer¹¹ have published an extensive compilation of automatic assignments of helices and sheets. Their ap-

proach has the advantage of giving assignments when only backbone C^{α} 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. 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^{α} 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_{10} -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,

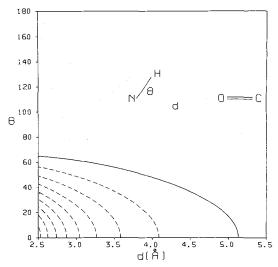


Fig. 1. H bond between peptide units is described here by the dominant electrostatic part E (see text) of the H-bond energy, drawn in contours of constant E at 0.5 kcal/mol intervals as a function of the distance, d, and the alignment angle θ . Dotted lines, E positive or zero; broken lines, E negative.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

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

Hydrogen-Bonded Structure

Hydrogen Bonds

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

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

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 kcal/mol binding energy. We choose a generous cutoff to allow for bifurcated H bonds and errors in coordinates and assign an H bond between C=O of residue i and N-H of residue j if E is less than the cutoff, i.e., "Hbond(i,j)=: [E < -0.5kcal/mole]."

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

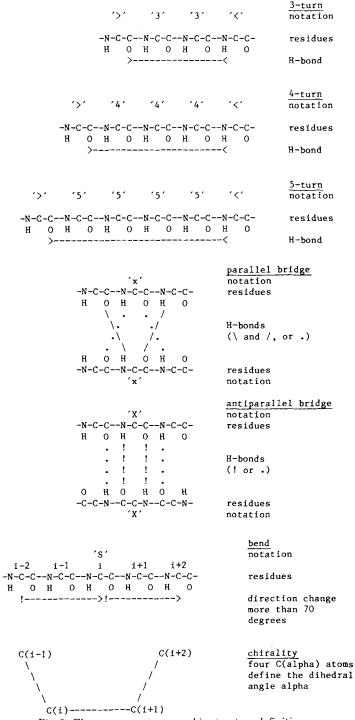


Fig. 2. Elementary patterns used in structure definition.

more complicated description of H bonds in terms of one distance and one angle. There is no generally correct H-bond definition, as there is no sharp border between the quantum-mechanical (wave-function overlap dominates at short distances) and electrostatic (electrostatic interaction dominates at larger distances) regimes and no discontinuity of the interaction energy as a function of distance or alignment. Thus, any H-bond definition is empirically tailored to a particular purpose. Our definition, well tested by trial and error, reflects a compromise suitable for the purpose of secondary structure definition. The cutoff chosen, which allows for an N-O distance up to 2.2 Å larger than the optimal value at perfect alignment or a misalignment of maximally 60° is similar to the tolerances used by Levitt and Greer¹¹ (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 β -structure; in particular,

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

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

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

Cooperative H-Bond Pattern: Helices

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

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

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

```
3-helix(i,i + 2)=: [3-turn(i-1) \text{ and } 3-turn(i)]
5-helix(i,i + 5)=: [5-turn(i-1) \text{ and } 5-turn(i)]
```

Longer helices are defined as overlaps of minimal helices. Conventionally, these structures are called α -helix, $3_{,10}$ -helix, and π -helix. In Table AIII, a 3-helix can be recognized by the pattern $\rangle\rangle3\langle\langle$, a 4-helix by $\rangle\rangle44\langle\langle$, and a 5-helix by $\rangle\rangle555\langle\langle$. In the line SUMMARY, the residues bracketed by H bonds are labeled G, H, I, e.g.,

```
5-TURN \rangle5555\langle 4-TURN \rangle44\langle 3-TURN \rangle3\langle SUMMARY GGG HHHH IIIII
```

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

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

We coin the term "ladder" and define

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

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

Secondary Structure Irregularities

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

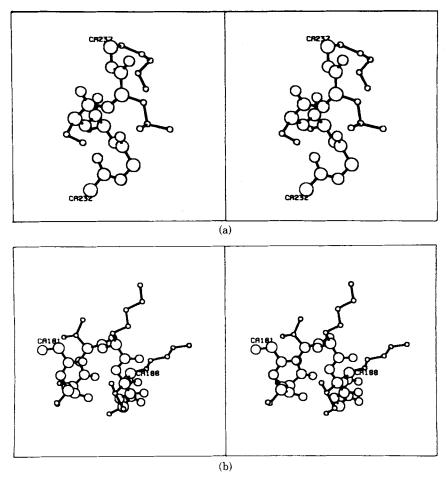


Fig. 3. Stereoviews of secondary structure: (a) 3-helix (3₁₀-helix) and (b) 5-helix (π -helix). (a) 3-Helix Gly232–Lys237 from triose phosphate isomerase (1TIM). In Table AIII, it appears as the H-bond pattern

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

5-TURN $\rangle\rangle\rangle55\langle\langle\langle$ SEQUENCE GSAVKVAK

5-Helices are extremely rare; the longest one, shown here, has three H bonds. All stereoviews are by PLUTO (Sam Motherwell, unpublished). In Figs. 3 and 5, the larger atoms are backbone atoms with $\frac{1}{4}$ their hard-sphere radius (C^{α} , 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.

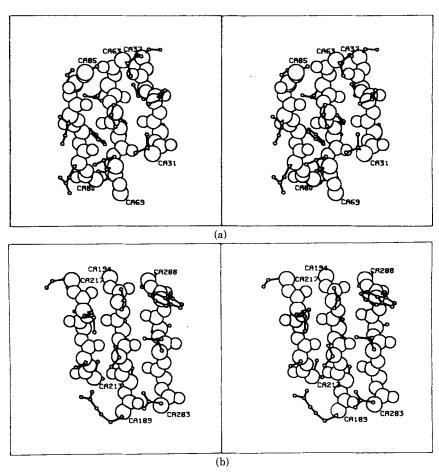


Fig. 4. Stereoviews of secondary structure: (a) antiparallel and parallel β -sheets with two ladders (three strands) each. (a) Two connected antiparallel β -ladders from trypsin (1PTN). The three participating strands are Val16(31)–Ser20(37), Ile46(63)–Gly51(69), and Glue62(80)–Ala67(85), where the first 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 β -ladders, Arg172(189)–Gly177(194), Thr196(213)–Ile200(217), Asp266(283)–Ala271(288) from glutathione reductase (2GRS). The corresponding H-bond notation (Table AIII) is

SHEET	EEEE	EEE	· EEEE · · · ·
BRIDGE2	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 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:

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 β -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 observation of β -bulges, a frequent lattice fault in β -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 Seltzer⁵ but slightly different from that of Rackovsky and Scheraga. For a bend at i, we require a curvature of at least 70°. The cutoff value was chosen by visual inspection of three-dimensional traces. With \mathbf{C}^{α} the position vector of \mathbf{C}^{α} , we define

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

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

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

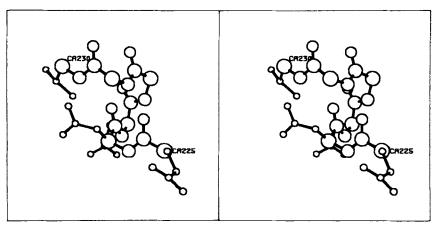


Fig. 5. Stereoviews of secondary structure: illustration of chirality. This short left-handed α -helix, Gln225-Val230 from thermolysin (2TLN) is the only one known to us. In Table AIII (note that chirality is entered at the second residue of each quartet) it appears as:

CHIRALITY	
4-TURN)) 4 4 ((
SUMMARY	нннн
SEQUENCE	QDNGGV

SS Bonds

SS bonds, i.e., covalent links between the S^{γ} 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.

Structure Summary

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

Static Solvent Exposure

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

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

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

Algorithmically, we integrate by summing over a polyhedron made of 20, 80, 320, or more approximately equal triangles. The integration points are the triangle centers, the weights are the triangle area. The polyhedron is generated starting from an icosahedron; a recursive procedure then divides each triangle into four by connecting the midpoints of the sides and projects the three new vertices onto the surface of the sphere, ready for the next level of recursion. The final polyhedron is reminiscent of the shells of certain viruses and of Buckminster Fuller's architecture of geodesic domes. Hence, we call the algorithm "geodesic sphere integration." It is similar to the algorithm of Shrake and Rupley¹⁶ and conceptually simpler than z-layer integration.

With 320 integration points, the surface area of a residue is accurate to within 1 Å²; with 80 points, to within 4 Å². For myoglobin, the numerical values agree with those of Lee and Richards, ¹⁰ using their parameters. The numbers given here are based on slightly different values of atomic radii: 1.40 for O, 1.65 for N, 1.87 for C^{α} , 1.76 for C of CO in the backbone, 1.80 for all side-chain atoms,¹⁷ and 1.40 for a water molecule following observed water-protein distances (Ref. 18 as cited in Ref. 19).

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

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

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

RESULTS AND DISCUSSION

Choice of Proteins

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

H-Bonded Structure

Backbone-backbone H bonds can be simply classified by the number of residues they bracket or, in our notation, by n of (i,i+n) = (CO(i),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 ϕ,ψ values of residue i+1. This local conformation is known as C_7 and leads to an extended strand roughly similar to a β -strand if it repeats. When it occurs as part of a tight turn, that turn is sometimes called a γ -turn.

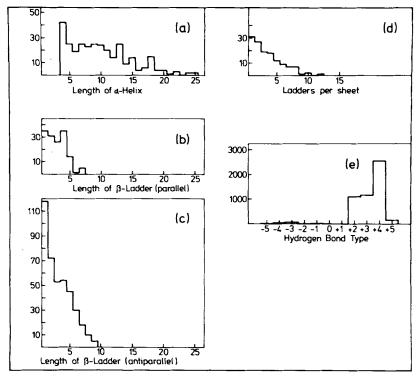


Fig. 6. The common feature of the size distribution of secondary structure segments is the gradual fall-off: larger sizes are less probable than smaller ones. Note that we give (b,c) the length of β -ladders (strand pairs) rather than the length of β -strands. A strand is often longer than the ladders in which it participates, since sheets tend to be trapezoidal rather than rectangular in shape. The number of bulge-linked ladders per sheet (d) is given as an indication of the width of the sheet. The width of a ladder is about 5 Å. In an ideal sheet, center strands take part in two ladders, edge strands in one: the number of ladders is equal to the number of strands minus one. In general, however, one strand can participate in more than one ladder on each side and the width of the sheet less than the number of ladders times 5 Å. Note: sheets consisting of a single bridge are not included in the histogram of ladders per sheet. (e) Number of H-bonds of type (CO(i), NH(i+n)). Due to the nature of L-amino acids, positive n are heavily favored. The dominant peak at n=4 represents α -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 β -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 group²⁰ to be essential in stabilizing the C_7 conformation in solution. We speculate that β -strands originate as extended C_7 strands as the protein folds up. Outside of β -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^{β} .

3-Helices are more frequent than previously believed, although they are usually short and have mediocre hydrogen bonds. α -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 α -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 Schellman²¹ and Richardson. 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 Gloeckner²²). In particular, we suggest that the distinction between parallel and antiparallel β -structure^{23,24} in the reference spectra will improve the overall accuracy of these experiments.

Accuracy of H-Bond and Secondary Structure Assignments

At best, secondary structure assignments can only be as accurate as the coordinates on which they are based. In using this dictionary, it is therefore very important to be aware of the state of resolution and refinement of each structure indicated in Table AI. The coordinate data sets range from refined structures at better than 1.5-Å resolution, where individual side chains can clearly be seen, to unrefined structures at a resolution just sufficient to trace the protein chain. As a test, we compare our assignments with those of the crystallographers and of Levitt and Greer¹¹ 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 Steigemann²⁵ 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 2.6×10^{-5} min⁻¹ measured by nmr.²⁶ Three others of type (i,i+3), with E=-1.3,-1.7,-0.9, form the well-known⁸ 3-helix Asp3-Leu6. One (i,i+5) H bond, Asn24(CO)-Leu29(NH), is part of the β -hairpin. Six are of type (i,i+2), characteristic of the C₇ configuration: five weak ones and one stronger one (E=-1.8) in a γ -turn at Asn43. The additional H bonds assigned by us lead to identification of two unambiguous segments of secondary structure not cited by the authors but also assigned by Levitt and Greer.¹¹

For the *medium-resolution* structure of cytochrome c550, Timkovich and Dickerson²⁷ use a conservative interpretation of hydrogen bonds and

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

		and Lo	wer resolution	<u> </u>
Structure ^a	Original Authors (AU)	Levitt & Greer (LG)	This Work (KS)	
3PTI				
G1	b	2-7	3–6	Clearly 3_{10} ; LG have α
E1	16-25	14-25	18-24	0.001.19 0.10, 200 110 11
E2	28-36	28–37	29-35	
E3	b	43-46	45-45	β -Bridge, 2 H bonds
H1	47-56	47-55	48-55	,
155C		2. 00		
H1	611	$4-16^{\rm b}$	6-12	4-Turn 13–16
G1	11-13	_	11-13	Overlaps with H1
E1		$17-23^{b}$	19-20	AU have 2 H bonds; KS, 4
$\mathbf{E}2$		26-31 ^b	_	Discontinuity at Asp28-Ile29
E 3		33-39	35-37	AU have 2 H bonds; KS have 4
H1	***	40-44 ^b	_	KS have 3-Turn
H2	56-63	55-65	56-64	
H 3	73-79	71-80	73-80	
H4		$81-90^{\rm b}$	_	Pro at 82, 84; possible helix
H5	107-118	106-118	107-117	
2ADK				•
H1	1-8	1-7	2-7	
E1	10-14	8-15	10-14	
H2	23-30	21-31	23-31	
$\mathbf{E}2$	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	
E 3	90-94	88–95	90-93	
H6	100-107	100-109	101-108	
E4	114–118	113-120	114–118	
H7	123~133	121-136 ^b	122-132	α -Helix ends in 3-turn
H8	144-158	$141-157^{\rm b}$	143–157	No $(i, i + 4)$ H bond at Asp 141
H9	160-164	159–166	160–167 ^b	Two weak H bonds at 167,168
$\mathbf{E}5$	169–173	169–175	170–173	
H10	179-194	179–192	179-193	

 $^{^{\}rm a}$ H = α -helix, G = 3_{10} -helix, E = β -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).

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 α -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+1)

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

2) in approximate γ -turns at Glu2, Gly40, Lys 53, and Lys88; two are (i,i+4) H bonds at the end of α -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 β -strand pair, 19–20/35–37, which is due to two additional weak H bonds. Levitt and Greer¹¹ assign considerably more secondary structure (Table I), including a much longer parallel β -sheet 17–23/33–39 (probably too long), a β -strand 26–31 (roughly antiparallel to 17–23), a helix 40–44 (we assign a 3-turn), and a longer helix 81–90 (which has only two of the seven possible H bonds but looks very much like a helix in a \mathbb{C}^{α} chain tracing and therefore may be seen to be a helix at higher resolution).

For the unrefined, *lower-resolution* structure of adenylate kinase (2ADK²⁸), all secondary structure assignments (ours, the original authors', ²⁸ and Levitt and Greer's¹¹) 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, Richardson⁸ sees a number of β -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 β -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, and for lower-resolution data, they are conservative compared with both Levitt and Greer's program and Richardon's visual processing.

Secondary Structure Size

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

to a maximum segment length of about 30 Å, with parallel β -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 β -ladders characteristic of the $\alpha/\beta/\alpha$ folding unit and, perhaps, 13- and 18-residue α -helices. We speculate that protein folding, although cooperative, follows random polymer statistics approximately in that long segments are statistically less likely than short ones. The apparent maximum size of 30 Å perhaps reflects the maximum size of globular domains.

OUTLOOK

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

APPENDIX: DICTIONARY OF PROTEIN SECONDARY **STRUCTURE**

Notes to Table AI

Proteins are ordered by function and can be found in the strip tables (Table AIII) and strip maps (Fig. AI) by their running number. $\% \alpha$ -helix, $\% \beta$ -antiparallel, $\% \beta$ -parallel = number of H bonds per 100 residues of type 4-turn, parallel and antiparallel bridge; these percentages can be compared with results from spectroscopy (CD, Raman, ir). Exposure = estimated number of water molecules in contact with protein surface (first hydration shell); it can also be read as the static exposed surface area in units of 10 Å². 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

:	:-	:-	:					% BE % BE WATE MULT NUMB	R EXPOSURE IPLICITY OF DATA SET ER OF RESIDUES
:	:	:	:	:	:	:		REFT	NEMENT
ZAH	%BA	% B P	EXPO	М	LEN	RES RI	, ,	PROT	LUTION NEMENT EIN IDENTIFIER, NAME
bindin									
38	4	0	610 1423		108	1.9 DI	1)	1 CPV	CALCIUM~BINDING PARVALBUMIN B
					306	2.4	- 2)	LABP	L-ARABINOSE-BINDING PROTEIN
electr 7	on 17	2	490		8.5	2.0 DE	3)	1319	OXIDIZED HIGH POTENTIAL IRON PROTEIN (HIPIP).
electr	on tr	anspo	r t		0.6	2 0 00		3050	CALIFORNIE HE (ONIDITED)
57	15		566 665		103	2.8 02 2.5	5)		CYTOCHROME B5 (OXIDIZED) CYTOCHROME B562 (E. COLI, OXIDIZED)
34	2	Ω	620	2				ICYT	CYTOCHROME C (OXIDIZED).
23	2	3	781		112	2.0 DG 2.5 DI	. 8)	1550	CYTOCHROME C2 (FERRI) CYTOCHROME C550
38		Ô	642 781 482 316		8.2	2.0 CS	9)	251C	CYTOCHROME C551 (OXIDIZED)
9	9	0	316 623		54 98	2.0 DC	11)	1FDX	FERREDOXIN (PEPTOCOCCUS AEROGENES) FERREDOXIN (SPIRULINA PLATENSIS)
30	1	1.8	715		138	1.9 08	12)	3FXN	FLAVODOXIN (OXIDIZED)
10	17 17	0	376 645 513		125	1.5 LS 2.7 HF			RUBREDOXIN (OXIDIZED, FE(III)) AZURIN
2	21	10	513		99	1.6 DF	15)	IPCY	PLASTOCYANIN
iormon		^	2/2		36	1 / 0/	161	lnnr	AUTAM DANODCATIC DOLVDCDTIDE
38	0	0	343 354		29	1.4 RI 3.0 RE	17)	IGCN	AVIAN PANCREATIC POLYPEPTIDE GLUCAGON (PH 6-7)
29	i 2	0	354 301	2	5 (1.5 Di	18)		INSULIN (A AND B CHAIN)
ıydrol 17	ase,	phosp	hatide 712	acyl	123	1.7 AE	193	1892	PHOSPHOLIPASE A2
ydrol	ases,	0-gl	ycolsyl						
38	7	0	918 665			2.4 CI 2.5 CI			LYSOZYME (BACTERIOPHAGE T4) LYSOZYME (HEN EGG WHITE, TRICLINIC)
ydrol	ases,	phos	phoric			2.5 01	21)	7012	processe (now boo warre, ratebrate)
19	20	3	842		142	<4	22)		STAPHYLOCOCCAL NUCLEASE (COMPLEX)
	28	prot	709 einases			2.0 SI		IKNS	RIBONUCLEASE-S
28	5	9	1209 1333 1272 1266		308	2.0 2.5 HK 2.8 D2	24)		CARBOXYPEPTIDASE A
3 7	13 32	3	1333		324	2.5 HK	25)		ACID PROTEASE (RHIZOPUS CHINENSIS) ACID PROTEINASE (PENICILLOPEPSIN, FUNGUS)
27	9	4	1266		316	2.3 D2	2.7)	2 T L N	THERMOLYSIN
6	31	1	1033		236	1.9 HH 2.8 DI	28)		GAMMA CHYMOTRYPSIN A ALPHA LYTIC PROTEASE
7	31	3	821 929			1.5 02			BETA-TRYPSIN (NATIVE AT PH 8)
4	34	2	745		181	2.8 Di	31)	1 S G A	PROTEINASE A FROM STREPTOMYCES GRISEUS (SGPA)
23	35	12	1058 1089		275	2.5 Ft 2.5	32)	ISBT	SUBTILISIN BPN' TOSYL-ELASTASE
2 1	19	1	923 968		218	1.7 LS	34)	2 A C T	ACTINIDIN
19	15		968		212	2.8 DI	35)	8PAP	PAPAIN
onummi l	34	2	2101 492		428	2.0	36)	1 F A B	LAMBDA IMMUNOGLOBULIN FAB
		3	492	2	107	2.0 CC	37)	IREI	BENCE-JONES (MMUNOGLOBULIN (VARIABLE PORTION)
somer 23	2	5	1220		230	2.8 HK	38)	3 P G M	PHOSPHOGLYCERATE MUTASE (DE-PHOSPHO)
35	Į.	1.5	1026	2	246	2.5 DC	39)	1111	TRIOSE PHOSPHATE ISOMERASE
ectin	(agg 35	lutin O	in) 1125		237	2.4 ME	40)	3 C N A	CONCANAVALIN A
yase,	carb	on-ox	ygen						
	20 educt		1273		256	2.0 DI	41)	1 CAC	CARBONIC ANHYDRASE FORM C
15	12	1.3	937		162	2.5			DIHYDROFOLATE REDUCTASE (COMPLEX)
22 17		10	1505 1639	2	333 374	2.9 D: 2.4 D.			D-GYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE APO-LIVER ALCOHOL DEHYDROGENASE
27	6	7	1753		329	2.0 D2			LACTATE DEHYDROGENASE, APO ENZYME M4
22	12	7	2354		461	2.0	- 46)	2 CRS	GLUTATHIONE REDUCTASE
l axveer	33 stor		686	4	151	2.0 H	(4/)	2500	CU, ZN SUPEROXIDE DISMUTASE
65		0	842		153	2.0 D2	48)	IMBN	MYOGLOBIN (FERRIC IRON - METMYOGLOBIN)
oxyger 62	tran	sport	706		136	1 / DS	: 491	LECD	HEMOCLORIN (EDVINDOCRIORIN DROVV)
58	0		1415		287	1.4 DS 2.0 DI	50)	2 M H B	HEMOGLOBIN (ERYTHROCRUORIN DEOXY) HEMOGLOBIN (HORSE, AQUO MET)
4.7	0	0	864 824			2.0 01	. 51)	ILHB	HEMOGLOBIN(MET)-CYANIDE V (SEA LAMPREY)
62 plant	o seed				153	2.0 D2	32)	LHBL	LEGHEMOGLOBIN (ACETATE, MET) (YELLOW LUPIN)
33	6	0	301		46	1.5 88	53)	1 CRN	CRAMBIN
rotei 13			itors 351	4	56	1.9 D	541	1000	OVOMUCOID THIRD DOMAIN
13	2 3	2	632	•	107	2.6 DC	55)	2 S S I	STREPTOMYCES SUBTILISIN INHIBITOR
12 oxins		0	412		58	1.5 D2	56)	3 P T I	TRYPSIN INHIBITOR
.oxins		0	511		7 1	2.8			ALPHA COBRATOXIN
71	0	0	222	2		2.0 H	58)	IMLT	MELITTIN NEUROTOXIN B (PROBABLY IDENTICAL TO ERABUTOXIN
0 ransf	29 erase	.s	406		62	1.4 H	. 54)	TMYR	MEDICALIAN D (PRODUBLI IDENTICAL TO EXABOTOXIN
47	0	10	1251						ADENYLATE KINASE
20	2	10	1456		293	2.5 DI	61)	LRHD	RHODANESE
ransp	170								

TABLE AII Structure Notation Used in Table AIII

	Structure Notation Used in Table Ath
First line: running	number 1–62, data set identifier (3PTI,4LDH), protein name,
	[function], \source\
SHEET	One-character name of β -sheet ("A," "B," "C") in which residue i participates.
BRIDGE2	One-character name of β -ladders in which residue i participates,
BRIDGE1	"A," "B," "C" = antiparallel,
.,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	"a," "b," "c" = parallel.
	Ladders are named sequentially from N- to C-terminus.
	A β -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 β -strand and tracing all partners and
	their partners.
CHIRALITY	"+" or "-"
	Chirality at residue i is the sign of the dihedral angle defined by C^{α}
	$i-1$ to $i+2$. Thus, a right-handed α -helix has "+," an ideal twisted
	β-strand ""
BEND	"S" = five-residue bend centered at residue i .
5-TURN	Hydrogen-bonding pattern for turns and helices:
4-TURN	">" = backbone CO of this residue makes H bond $(i, i + n)$
3-TURN	"(" = backbone NH of this residue makes H bond $(i - n, i)$
	"X" = both CO and NH make H bond
	"3," "4," "5" = residues bracketed by H bond
SUMMARY	Structure summary:
	"H" = 4-helix (\alpha-helix)
	"B" = residue in isolated β -bridge
	"E" - extended strand, participates in β -ladder
	"G" = 3-helix (3 ₁₀ -helix) "I" = 5-helix (π-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
2112 0001121111	contact with residue i. The scale is $0-9$; "*" = more than 9 water
	molecules. Exposure can be read as solvated surface area in units of
	10 Å ² .
SEQUENCE	Amino acid sequence in one letter code:
	"a," "b," "c" are Cys residues labeled by their SS-bond name. "!"
	= chain break (peptide bond length exceeds 2.5 Å). Residues
	including chain breaks are numbered sequentially within the
	coordinate data set, irrespective of the residue identifier given there.
	Thus, the total number of residues is equal to the total number of
	print positions minus the number of chain breaks.

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

3	SHEET.	CAICIUM-BINDING PARVALBUMIN B [CALCIUM BINDING PROTEINS] (CARP: CYPRINUS CARPIO)	PARVALBUMIN	N B [CALCIUM	BINDING P	ROTEINS] (CA	RP: CYPRINU AAA	S CARPIO)			AAA
	BRIDGEL. CHIRALITY BEND.	**************************************	S SSSSSSS SS S S	**************************************		**************************************	AAA SS SS SS	**************************************	\$8 SSS	**************************************	AAA +-+-+-++ SS SS SS
-	4-TURN 3-TURN SUMMARY EXPOSURE.	>>>> >33< TT S HHH 4**********************************	>>>> XXXX<<<>>>3 >33 >>> XXXXX< >34 >>>> >>>> 4**5938*74 26463*81.*6 7**5392*818 APACYLNDAD IAAALEACKA ADSP NHKAPP	>>>X< SS HHHHH 7*5392*51@ ADSPNHKAPP		>>XXXXX<<< 3< >>3< >>3 нннининн е6@**22471 DDVKKAFAII	<pre>< >> >4<</pre> <pre>>> >444 >> 3xx333</pre> SS SEEH HHHTSSTTT 1*9*63864* *219660753 DQDKSGFIEE DELKLFLQNF	>> >4<< >444 < >> 3XX3<>>33< >33< EH HHHTSSTT STT 4* *219600753 *9*1 EE DELKLFLQNF KADA	7.4 G	>>> >XXXX<<<<< >>3 3< >>3 X HH HHHHHHH 4*3 12*9@1*7@1 TDG ETKTFLKAGD	>>> 33< >>> TT SSEEEHH *96838008* SDGDGKIGVD
181	SHEET BRIDGEZ BRIDGEZ CHRALITY BEND 5-TURN 3-TURN SUMMARY EXPOSURE.	+++++ SSSSS >XXXCCCC >XXXCCCCCCCCCCCCCCCCCCCC									
2)	JABP L-AR SHEET	L-ARABINOSE-BINDING PROTEIN (BACTERIAL: ESCHERICHIA COLI)	DING PROTEIN	N (BACTERIA)	L: ESCHERICE AAA	HIA COLI)		æ		JABP B	labp
	BRIDGEL CHIRALITY BEND	1 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0	\$\$\$\$\$\$\$\$\$\$\$\$\$\$	**************************************	888 SSS	**************************************	**************************************	b SS S	**+++++++ \$\$\$\$\$\$\$\$\$\$	b SS SS SS	**************************************
н	4-TURN 3-TURN SUMMARY EXPOSURE.	EEE **65198862 ENLKLGFLVK		>4>><> X XXXXXX<< < > 33< >33 < >33 < >33 < >33 < 6 <	<pre><</pre>	>>>XXXXXX > 33X3 > 33X3 SHHHHHHH 492**1491# PDGE #TLNAI	X><<<< 3<>>3< HHHHHT 6783955198 DSLAASGARG	BSSS BRBBB563* FVICTPDPP	> 4>>x>xxx<< >3 TTHHHHHHHH 3 86482558*6 L GSAIVAKARG	< <pre><< 3< ## B SS **29100010 YDM KVIA VDD</pre>	>33< STT 48299*2*55 QP VNAKGKPM
	SHEET	υ			000			Q	Ω		Q
	BRIDGEL CHIRALITY BEND 5-TURN	SS SS	+ 00 ^	++++++++-+- SSSSSSSS >55>5<5< XXXXX<<<	ppp sss	S\$\$\$\$\$\$	++++ +++++++ SSSS SSSSSSS >55	sss sss 55<	SSSSSS		S S S S
101	3-TURN SUMMARY EXPOSURE.	SS B S *921181888 DTVPLVMMAA	>3><3< HHHHHHHHHH HHHHHHHT 7*188*88*3 88*21**96: TKIGERQCQE LYKEMQKRG	HHHHHHHT B8*21**963 LYKEMQKRGW	>33< STT ERE 86*4#1#### DVKESAVMAI		>>>>< SGGGHH HHHHHHHHH HHS SSS E E SSSSSH 119*3618*4 1844818217 89216**33* *292*45695 TANELDTARR RITGSMDALK AAGFPERGIY QVPTKSNDIP	HHS SSS E E 89216**33* *; AAGFPEKQIY O	> HHS SSS E E SSSSSH 89216**33* *292*45695 AAGFPERQIY QVPTKSNDIP		33< >>3< НИННИННИН НН S S E 108*186612 **2**18489 GAPDAANSML VQHPEVKHML
SUM	IMARY	SUMMARYH-ALPHA-HELIXE-BETA-STRANDB-BETA-BRIDGEG-3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS-BEND	LIXE=BE	TA-STRAND.	8=BETA-BR	IDGE G=3-	HELIXI=	S-HELIX	T=3-,4-, O	R S-TURN	S-BEND

CG ++++++++++++++++++++++++++++++++++++		(CHROMATIUM VINOSUM)JHIP B C D E C-++	28 5C		SUMMARYH=ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND
C E C C C C C C C C C C C C C C C C C C		C C E	B 59*99 TLKAG	SS SSS >44 4< >42 44 4< >53 XX3< 67 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	5-TURN.
SSS S 555< 675 135*6*17**		.i. t∩	N >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < >33 < 33 < >33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 < 33 <	######################################	.T*3-,4-, OR
EE G C C SSS SSSSSSSS > 4>4x < > xx > xx x x x x x x x x x x x x x		SULFUR PRO	>33X33< SSE EEETTSTT E 5*91 382*61*878 TDEW KGCQLFPGKL	SSSSSSS >>>< HHHHHHHHHHHHHHHS 56**3**82*	*5-HELIX
EE G G F F F F F F F F F F F F F F F F F F		SFER, IRON-C C ++-++-++	>>3 >>3 >>3 >>3 >>3 >>3 >>3 >>3 \$25GGGSSHH H SSS GGG GGGBTTEE TYSE SE SGGGGFGSPSSHGGGFGT TYST SE SGGGTTEE TYST SE CONTROLL THE TYST SE CONTROLL TYST SECONTROLL TYST	C C C C C C C C C C C C C C C C C C C	-HELIXI
E E F F F F F F F F F F F F F F F F F F		HIPIP) (ELECTRON TRANS C AA E BB S SSS SSS SSS SS	>>3 < <>>3 < <>>3 < <>>3 < <>>3 < <>>3 < < < < <>>3 < < < < < < < < < < < < < < <	SSSSS SS SSSSSS SS 44><44 A4> A4> A4 A4> A4 A4> A4 A4 A4> A4 <	IDGEG=3
E		(HIPIP) (ELI	>>3< H SSS GGG 559*6553** AARPGLPPEE	++++++++++++++++++++++++++++++++++++++	.B=BETA-BR
		HEATTH HE SESSES SS S S S S S S S S S S S S S	>>3< SSGGGSSHHH H SSS GGG 860**2*299 559*6553** QDATKSERVA AARPGLPPEE	BBA AAA AAAA C	A-STRAND.
		ED HIGH POTENTIAL INC	>33< >33 < TTB TT HHHHHT B *53*62149* 3868*83526 SAPANAVAD NATAIALKYN HROME B5 (OXIDIZED) [E	AA BB BBA AAA AAAA AAAA AAAA AAAA AAAA	.IXE=BE1
dd -+-++++ SSSSSSS >444>X >33X3 EE SSGTTH 6081.1061.02.1VGMNDSTVL	-+++ SS 55 55 14 733 71 71 71 71 71 71 71 71 71 71 71 71 71	IZED HIGH PO	>33< >33 TTB TT *53*62149* SAPANAVAD	AA	.H=ALPHA~HEI
SHEET BRIDGE 2. BRIDGE 1. CHIRALITY BEND 5-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGE 2. BRIDGE 1. CHIRALITY BEND 4-TURN 3-TURN SUMARY SUMARY SEQUENCE.	SHEET SHEET BRIDGE2 BRIDGE1 CHIRALITY BEND 5-TURN	3-TURN SUMMARY EXPOSURE. SEQUENCE.		MARY
241	361	<u>8</u>	1	:	S CF

$\overline{}$
(continued)
AIII
ľΕ
AB.
₽

	SHEET BRIDGE2 BRIDGE1	2									
	CHIRALITY BEND	+++++++	SSSSSSS SSSSSSS	+++++++++ \$	+++++++++ \$\$\$\$\$\$\$\$\$\$\$		*************************	++++++++++++++++++++++++++++++++++++++	++-++++++	+++++++++ \$\$\$\$\$\$\$\$\$\$	\$
	4-TURN.	>>>>××××××		>4>X>X	>4>>X> XXXXXX<	>444<		>>>>XXXXX XXXXXXXXX X<<<<	XXXXXXX	>>>XXXXX<<	XX><< XX 44<
	SUMMARY.	>>><<	>33X>3<<	>33< STTTTHHHH	<<<	<< >33 < GGS TTTTT :	< >33< >	3<< >33< >33< нинининининининининининининининининини		>3XX 3<< >33X	3<< >33<
•	EXPOSURE.	*64*716*43	*64*716*43 **13*718*2	*698085600	8*19610*60	4+270++2++	77**2*61*8	4*278**2** 77**2*61*8 029422*006 603*31**59 4**036328* 1782*564**	603*31**59	4**#36328*	1782*564**
٦	SEQUENCE.	ADLEDUMOTL	ADLEDUM CTL NONLKVIEKA	BBZKANDAAL	BBZKANDAAL VKMKAAALNA QKATPPKLED NSQPMKDFRH GFDILVEGID DALKLANEGK VKEAQAAAEQ LKTTRNAYHQ	QKATPPKLED	NS OPM KOP RH	GFDILVEGID	DALKLANEGK	VKEAQAAAEO	LKTTRNAYHQ
	SHEET BRIDGE2. BRIDGE1. CHIRALITY BEND	+ w									
2	4-TURN 3-TURN SUMMARY EXPOSURE.	∧ 00 00 % 4 00 # 00 #									
(9	SHEET.	CYTOCHROME C (OXIDIZED) [ELECTRON TRANSPORT] (ALBACORE TUNA HEART: THUNNUS ALALUNGA)	(IDIZED) (EI	ECTRON TRAN	NSPORT] (ALB	ACORE TUNA	HEART: THUI	NNUS ALALUNC	3A}	•	1C YT
	BRIDGE1	+ + + + + + + + + + + + + + + + + + +	+	++++	¥ ++++	+ + + + + + + + + + + + + + + + + + + +	«	~*~*** -**** -****** ~***** ***** ********		+++++++++++++++++++++++++++++++++++++++	+ + + + + + +
	BEND.	SSSSSSSS	SSSSSSS	SSS SS	SS SS S	S SS S	SSSSS S	88888888	8888888	\$55	\$55 SSSSSSSS \$55
	4-TURN.	×××××××	<< <x444<< td=""><td></td><td></td><td></td><td>×</td><td>>>>XXX<<<x>44<<</x></td><td>>44<<</td><td>^</td><td>>>> xxxxxxxxx</td></x444<<>				×	>>>XXX<< <x>44<<</x>	>44<<	^	>>> xxxxxxxxx
	SUMMARY.	нининин	>33<	>33< STT SS	>33X33<>3 3<>33< TT TTSBT T TT	3<>33< >> T TT H	>> 3<< H HHHHS B S	в с нянинини		нин	
٦	EXPOSURE. SEQUENCE.	8*65*0**21 GDVAKGKKTF	8*923*5321 VQKCAQCHTV	**82*89912 ENGGKHKVGP	8*923*5321 **82*89912 23932465*8 8*8*84*43* 329*924927 8*21682298 8**38972*5 VQKCAÇCHTV ENGGKHKVGP NLMGLFGRKT GQABGYSYTD ANKSKGIVWN NDTLMEYLEN PKKYIPGTKM	0*0*84*43* GQAEGYSYTD	329*924927 Ankskgivwn	8*21682298 NDTLMEYLEN	0**30972*5 PKKYIPGTKM	*5*11***45 IFAGIKKKGE	3920113179 RQDLVAYLKS
101	SHEETBRIDGE1BRIDGE1BRIDGE1S-TUBNS-TUBNS-TUBNS-TUBNSUMMARYSERDGURE.	> + S + S + S + S + S + S + S + S + S +									

1020		++++	2555	^ ^2	*****	нинн	58*2**921 FF KLTKDDE				1550		+++++-+	S		v.	#51#3#4# OKGAKTKMT					BEND
IC2C FERRICYTOCHROME C2 [ELECTRON TRANSPORT] {BACTERIAL: RHODOSPIRILLUM RUBRUM}		4 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		,	, <u>~</u>	TS SS S	**7364**12 124213*641 116**6842* 13643*6*52 5259821353 2*9084116* 62588*2*1* 568*2**921 DQGGANXVGP NLFGVFENTA AHKDNYAYSE SYTEMKAKGL TWTEANLAAY VKNPKAFVLE KSGDPKAKSK MTFKLTKDDE				CYTOCHROME C550 [ELECTRON TRANSPORT] {PARACOCCUS DENITRIFICANS}			SSSSSSSSS	<>44><44<	<>33< >33<	6095113873 *9*51*3*4* DPKPLVKKMT DDKGAKTKMT					SUMMARYH-ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND
		* *******			>33<>3	TTSSSTT	2*9084116* 6 VKNPKAFVLE						+++++++	55555555		> >E<<						T=3-, 4-, OR
3UM)	c	-4-44444	SSSSSS	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	в ѕинини	5259821353 TWTEANLAAY						++++++	SSSS S	>>>>x	у нини	85*31*1268 Ø1**699112 GFKYGEGILE VAEKNPDLTW					=5~HELIX
IRILLUM RUBF			SSS SSS	744	33<	TTS SSSS	13643*6*52 SYTEMKAKGL				ANS }		+	SS SSSSS SSS	>>>>	нинин 55	85*31*1268 GFKYGEGILE					-HELIXI=
AL: RHODOSP			S SSS S	7		T TTS T	116**6842* AHKDNYAYSE				DENITRIFIC		_	888 88 88		>33<>33<						IDGEG=3
T) (BACTERI. B	c	4 1 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4			>33X33C>3	TT TT BT TT	124213*641 NLF GVF ENTA				(PARACOCCUS	¢	e e ++	888 88		20.00	02*9122261 GGKTGPNLYG		t os	;	>3 3< ST T 78 4146 XX XXXX	B=BETA-BR
ON TRANSPOR					>33< >33<	STT TTS					TRANS PORT]		1	SSS SS		888	072**2764* IQAPDGTDIK		S	•	>3 3< S SS S ST T 6795678478 4146 AXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	TA-STRAND
C2 [ELECTR		444444	SSSSSSS	747777	><< X 3 3 <		*6*23206*1 2**2966212 EGDAAAGEKV SKKCLACHTF		~	3. L.K	(ELECTRON '	É		88888888		> >>>> > 333 ЗНИНИН НИПТОТОТОВ	51**478623 EFNKCKACHM		SSSS SSSSSS	>>>> xxxx x<<	>33x33< S HHHH HHHHHHHS S S S T T 9*27*4422 01112*31** 6795678478 4146 FKMCRNQADV VAFLAQDDDD AXXXXXXXXX XXXX	LIXE=BE
ICYTOCHROME		*******	SSSSSSS SSSSSSS	>5555<	********	ТНИННИНН	*6*232Ø6*1 EGDAAAGEKV	+ 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	55><555< XXXX <xx<4<< td=""><td>HHHHHHHHH 95184823*6 IENVIAYLKT</td><td>CHROME C550</td><td></td><td>+++++</td><td>SSSSSS</td><td>>>>>×</td><td><</td><td>**3726*8** NEGDAAKGEK</td><td></td><td>8 8888</td><td>×<<<</td><td>S HHHH 9*27*44222 FKMGKNQADV</td><td>. н-агрна-не</td></xx<4<<>	HHHHHHHHH 95184823*6 IENVIAYLKT	CHROME C550		+++++	SSSSSS	>>>>×	<	**3726*8** NEGDAAKGEK		8 8888	×<<<	S HHHH 9*27*44222 FKMGKNQADV	. н-агрна-не
IC2C FERR SHEET	BRIDGE 2.	CHIDALITA	BEND	5-TURN.	3-TURN.	SUMMARY.	EXPOSURE. SEQUENCE.	SHEET BRIDGE2 BRIDGE1 CHIRALITY	5-TURN.		155c CYTO	BRIDGE 2.	BRIDGE1. CHIRALITY	BEND	4-TURN	SUMMARY	EXPOSURE. SEQUENCE.	SHEET BRIDGE2 BRIDGE1	BEND	4-TURN.	3-TURN SUMMARY EXPOSURE. SEQUENCE.	MARY
5							-			101	8						٦				161	SUR

TABLE AIII (continued)

BRIDGEL, SSSSSSS SSS SSS SSS SSS SSSSSSSS SSSSSS	A A A A A A A A A A A A A A A A A A A
SUMMARY., EEEEE SSS HHHHHHHHHH HHHHHTT EEEEGGG S TTTTT SEEE EEE BTTTB TTTHHHHH EXPOSUNE. 47082325*2 407*807101 *089*67**1 *444316*483 **16*3*600 0801276*45 12*9806610 1 SEQUENCE. MKIVYWSGTG NTEKMAELIA KGIIESGKDV NTINVSDVNI DELLNEDILI LGCSAMGDEV LEESEFEPFI	æ ≠ ω

	RUBREDOXIN (OXIDIZED, FE(III)) [ELECTRON TRANSPORT] [CLOSTRIDIUM PATEURIANUM]	COPPER BINDING] (PSEUDOMONAS AERUGINOSA)	SHEET BBBB BBBBBBBBBBBBBBBBBBBBBBBBBBBB
++++++ SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	DOXIN (OXIDIZED, FE(III)) [ELECTRON TRANSPORT] [CLOSTRIDIUM PATEURIAN BBB BBB AA A B B A A A B B A A A A A	COPPER BINDING] [PSEUDOMONAS AERUGINOSA]	B*BETA-BRIDGEG=3-HELIXI=5-
AA AAAA d	AA AA AA AA44 A444 A444	TRANSPORT, COPPER BIND:	### ### ##############################
BRIDGE 2 BRIDGE 2 GHIRALITY +++++	SHEET AAA BRIDGE 1 BBB BRIDGE 1 AA BRIDGE 1 AA BRIDGE 1 AA BRIDGE 1 AA FURNALITY+++ SENUM >444< 3-TURN >33< 3-TURN SBS SECOURD SECOURD SECOURD SECOURD SECOURD SECOURD SECOURD ***4497865 SECOURD ***4497865 SECOURD ***4497865 SECOURD ***4497865 SECOURD ****4497865 SECOURD ****4497865 SECOURD ****4497865 SECOURD ****4497865 SECOURD *****4497865 SECOURD **********************************	AZURIN ELECTRON TRANSPORT, BRB	BRIDGE1 BBBB BRIDGE1 DDDD BRIDGE1 SS 5-TURN SS 4-TURN SS EEEE SEMPLAN SS EEEE SEMPLAN SS EEEE SEROURD SS EEEE SEROURD SS EEEE
SHE BRH BRH BRH 11 14 14 14 14 15 16 16 16 16 16 16 16 16 16 16 16 16 16		14) 1AZ SHE BRI BRI CHI 5-T 4-T 3-T SUM SUM SEXP	SHE BRI BRI CHI CHI 5-T 3-T 3-T 5 SUM 101 SEQ SUM ARY

_
(continued
_
AIII
E AIII
ABLE AIII

15)	SHEET.	IN (ELECTRON TRANSPORT, COPPER BINDING) (POPLAR LEAVES: POPULUS NIGRA VARIANT ITALICA)
	BRIDGEL CHIRALITY BEND	aaaa BB ddd d aaa a G E E dddd
	4-TURN 3-TURN SUMMARY EXPOSURE.	>>44 \$33 \$33 \$33 \$33 \$34 \$34 \$34 \$34 \$34 \$35 \$35 \$35 \$35 \$35 \$36 <p< td=""></p<>
16)		POLYPEPTIDE [HORMONE]
	3-TURN SUMMARY EXPOSURE.	>3 3<>>> 3>
17)		6 - PH 7 FORM) +++ ++++++++ 5SS SSSSSSSS
	4-TURN 3-TURN SUMMARY EXPOSURE. 1 SEQUENCE.	S
18)		IINS INSULIN (A AND B CHAIN) [HORMONE] [PIG: SUS SCROFA]
	4-TURN 3-TURN SUMMARY EXPOSURE.	>>>>x<<< >>44<< >>>>xxxxx<< << >>4 HHHHHHB B HHHHGGB B T HHHHHHHHHHHHHHHHHH
S	UMMARY	SUMMARYH=ALPHA-HELIX,E=BETA-STRANDB=BETA-BRIDGE,G=3-HELIXI=5-HELIX,T=3-,4-, OR 5-TURNS=BEND

BBBB BBBB	++++++++++++++++++++++++++++++++++++++		Color		SUMMARYH=ALFHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS×BEND
BBBB	CCCC		+++++++ \$\$\$\$\$\$\$\$\$\$ X>>XXXX<<< >>3X3X3X HHHHHHHHHHT		.T=3-,4-,
	++++ SSS SSS >>555< +<<<>>4<<>>444 >33<>>344 >33<>>34 P#HTT TTT ##TT TTT		C	5 SS	.5-HELIX
BOS TAURUS	SS 5SS 5SSSSSS 5S 5SSS 5S 5SS 5SS 5SS 5		B	++++-+++++++++++++++++++++++++++++++++	HELIXI"
W PANCREAS:	SS 5SS SSSSS SSSSS SSSSS SSSSS SS SSSSSS		A B B +++++++++++++++++++++++++++++++++	**************************************	DGEG=3-
ROLASE) (CO	S SS SS >> >> >> S SS SS >> S SS SSH *5383*352* LGGSGTPVDD		AA A BB B+++++ S S SS K EEEES S SS HITHWOODIN	SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	.B=BETA-BRI
DE ACYL-HYD A A B	++++++++++++++++++++++++++++++++++++++	S 3<-< WNb WNb	YME [O-GLYCOLSYL HYDROLASE] {BACTERIOPHAGE T4}. AAAA AAA AA AB B A C AA AA BB B -+++++++++	+-++++++ +++++++++++++++++++++++++++++	A-STRAND
[PHOSPHATI	+++++++ +++++++ ++++-++++ \$5585858 55 55 55858 5858 >>>>XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	A +++++++++++++ + + + + + + + + +	OLSYL HYDRO A B AA A C AA S SS 4 C SS T EEEEEE	######################################	IXE=BET
18P2 PHOSPHOLIPASE A2 [PHOSPHATIDE ACYL-HYDROLASE] (COM PANCREAS: BOS TAURUS) A A A BERT B A A BERT	+++++ 5558588 58 58 58 >>>>XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	A +++++++ SSSSSSSSSSSSSSSSSSSSSSSSSSSSS	ABA	######################################	H=ALPHA~HEL
18P2 PHOSP SHEET	BRIDGEL. CHIRALITY BEND 5-TURN 4-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGE 1. CHIRALITY BRUD 4-TIRN 3-TIRN SUMMARY EXPOSURE	1LZM LYSOZ BREET BRIDGEL. BRIDGEL. CHRALITY BEND 5-TURN 3-TURN SUMMARY.	SHEET BRIDGE: BRIDGE: CHIRALITY BEDD 5-TURN 3-TURN 3-TURN	MMARY
19)	.,	101	20)	161	SU

,	$\overline{}$	
•	7*	
	≍	
	×	
	-2	
	nne q	
•	=	
	contr	
	€,	
	~	
	\sim	
	Ų,	
	_	
	_	
•	=	
,	=	
,	=======================================	
,	AIII	
	TABLE AIII	

7LYZ 7LYZ ************************************		AA AA AA BG GG GG GG GG SS SS SS SS SS SS SS SS SS	•
**************************************		AAA AAA AA AAA AAAA AA ddd GG GC	ON#8#8
SSSSSS SS SSSSSSSSSSSSSSSSSSSSSSSSSSSS			S-T-URN
ALLUS)		AAAAA P AAA BAAAAAAAAAAAAAAAAAAAAAAAAAA	## 3.00 - 1.4 - 1.00 OR
N EGG WHITE: GALLUS GALLUS) CC		ROLASE, (STAPHYLOCO +++++++++++++++++++++++++++++++++++	
EN EGG WHIT CC CC DD C DD SSS S >5555<>> >444<>> >3 EETTTEET 8228081.038		STER (DNA) HYDROLASE STAPHYLOCC	HELIXI
CC CC CC +		STER (DNA) SS SSS >33< SS STTS B153*****	<pre></pre> <pre><</pre>
IM) [O-GLYCOLSYL HY A ++++++++ SSSSSS SSSSSSS SSSSSS SSSSSS SSSSSS SSSS		SPHORIC DIE AAAAA BB Gdd CCCC HH +	++++++ \$
FORM) [O-GLYCOLS B	+++++- SSSSS 3<>33< 3<>72 436472*9	AAAAA CCCCC B B ++++- S SS 3< 111787446* DTVKLMYKGO	SS SSSSSSSS SSSS SSSS SSSS SSS SSS SSS
A B B B A A WRRHGLDNYC CRYSTAL FORM) [0-GLYCOLSYL HY B B B B A A A A A A B B B B B B B B B	-+++++++++++++++++++++++++++++++++++	YLOCOCCAL NUCLEASE (COMPLEX) [PHOSPHORIC DIESTER (DNA) HYDROLASE] [STAPHYLOCOCCUS AUREUS]	# H ++++++++ SSSSSSSSSSSSSSSSSSSSSSSSSS
145 OZ YME (TRICLINIC CRYSTAL FORM) [O-GLYCOLSYL HYDROLASE] [HEN EGG WHITE: GALLUS GALLUS]	SSS SSSS (STAPHYLOCOCCAL NUCLEASE (COMPLEX) [PHOSPHORIC DIESTER (DNA) HYDROLASE] [STAPHYLOCOCCUS AURBUS] A AAAA A AAAAA BB CCCC A AABB B CCCCC A AABB B CCCCC A AABB B CCCCC H H	SHEFT B B B B B B B B B B B B B B B B B B
7LYZ LYSOZ SHEET. BRIDGEJ. BRIDGEJ. CHIRALITY GHIRALITY 4-TURN. 3-TURN. SUMMARY. EXPOSURE.	SHEET BRIDGE 2 BRIDGE 1 CHIRALITY BEND 5-TURN 3-TURN SEROSURE.	SSNET BRIDGEL BRIDGEL CHIRALITY BRIDGEL CHIRALITY 4-TURN S-TURN SUMMARY EXPOSURE.	SHEET BRIDGE 2. BRIDGE 1. CHERALITY BEND 5-TURN 4-TURN SUMMARY EXPOSURE. SEQUENCE.
21)	101	22)	101 SUMP

C AAA	၁၃၁ ၅	**************************************	j	C>33C BTTB EEE	2**7*241*5 SKYPNbAYKT									Ad 71			D	S SSSSSSS S	***************************************	>33X33	твининит	*594035007 Ondsftaild	AAAAA A	eeeee G	ţ	SS		6 0 0 0	666667322 FLSIHSYSOL	SEND
AAAAAA C	BBBBB G	SS SS			3113##6889 2** SITDARETGS SK										8		u u			**************************************			A A AA	କ୍ଟେକ୍ଟ ପ୍ରସ୍ତୁ ପ୍ରସ୍ତୁ	1+++	SSSSS S	>>>>X	33<	A1**36518# 001 FVKNHGNFKA FLS	SUMMARYH=ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND
BBB A AV		+++ SS SS															* *********			XXXXXXXX XXXXXXXXXXXXXXXXXXXXXXXXXXXXX	ннинн	0400010014 0014018286 QMITQATGVW FAKKFTENYG				SSSSSSSSS S	>>4>XX>XX	<>33X33< >	SEVEVESTOR	T=3-, 4-, OR
TAURUS)	ם ם םםם	SSS SS	>444<	SEEEETTTE	0876588*9* 2495218825 376987588* 4*9478**69 *48458*761 MKSRNLTKDR bKPVNTFVHE SLADVQAVCS QKNVAČKNGQ TNďYQSYSTM	•								maine is	A AAAA	dddd	2222			>33 X33<	E EEEE SSTT		υ		+-	888 8888		>33 X33< >33		-5-HELIX
NCREAS: BOS		**************************************	>>>4<<<	>>X<<	376987588* SLADVQAVcS									SOR SORBOOKED MODE	A A	TO	υ «	8 88			ES SS E	**28384885 82535*8548 8888888883 YEGRPIYULK FSTGGSNRPA IWIDLGIHSR	υ	***	++-++-+	SSSSS S SS		>3 3< >33		-HELIXI
RIBONUCLEASE-S [PHOSPHORIC DIESTER (RNA) HYDROLASE] {COM PANCREAS: BOS AAAAA BBBBB	•	SSS		SEEEEE	0876500*9* 2495210025 MKSRNLTKDR DKPVNTFVHE											88	AAAAAA A	555		33<					ı	88888888		>>3<< >3		IDGEG=3
NA) HYDROLA			^ ~											TO GUAD UT OF	AAAAA AA			S		* ^		3876**22*1 LVSKLQIGRS				SS S		,		B-BETA-BR
DIESTER (R		*++++ \$8888	xx<<<		0**45801*3	B BB		7.4 F.F	‡				33248 FDASV	CHINE I WILLIAM				\$55555555		33< >33		895837748* MDLLVAQHPE			-++++++		🗸	>33<	SHHHHHHHH T TT S 0000124079 *659330051 NPNGPAFTHS ENRLWRKTRS	TA-STRAND
[PHOSPHORIC A	æ	++++++ SS SS	š		3114667*6* QHMDSSTSAA	80	· 54	Eu Eu	‡	222		E SSS EEE	6*38668388 8*5*9*3126 33248 TQANKHIIVA CEGNPYVPVH FDASV	10 10 10 10 10 10 10 10 10 10 10 10 10 1	- National Contraction			8888888		×××××××		4351*803*0 YHTLDEIYDF		•		S SSS	>>>4 <xx44<< td=""><td></td><td></td><td>LIXE=BE</td></xx44<<>			LIXE=BE
NUCLEASELS		\$\$\$\$\$\$\$\$\$	>>>XXX	нинини	*584348988 KETAAAKFER	AAAAA BBBB	લે કે કે કે	CCCCC REE	! -++-			EEEEE EEEE	6*38660300 TQANKHIIVA	TO STORY OF THE SENTENCES OF STREETS OF STREETS				SS SS		>3.3<	STT SS	*36**38355 RSTNTF NYAT	AAAA	00000 BBB	ŧ	SS S SS	4 <	\ \ \	SMDIFLEIVT	. н-агрна-не
IRNS RIBO SHEET	BRIDGE1	CHIRALITY BEND	5-TURN	3-TURN	EXPOSURE.	SHEET	BRIDGE 2.	BRIDGE1	CHIRALITY	BEND.	A-TURN.	SUMMARY	EXPOSURE.	40.71	SHEET	BRIDGE 2	BRIDGE 1.	BEND.	5-TURN.	3-TURN.	SUMMARY	EXPOSURE. SEQUENCE.	SHEET	BRIDGE 2.	CHIRALITY	BEND.	4-TURN	3-TURN.	EXPOSURE. SEQUENCE.	MARY
23)					-								101	34)	ŕ							٦							161	SUM

$\overline{}$	
$\boldsymbol{\sigma}$	
ā	
-7	
7	
7	
2	
~	
00	
٠,	
_	
_	
=	
-	
⋖	
~	
⋖	
LE A	
LE A	
LE A	
BLE A	

						2000	, man				
	SHEET	AAA			AAA	«		AAAA			
	BRIDGE 2	ပ္ပ						999			
	BRIDGE1	fff			fff			eeee			
	CHIRALITY BEND	-+-++++ S SS	ł	+ ++++++++ +++++++++++++++++++++++++	\$ SSS	**************************************	**************************************	++++++ S S	**************************************	********* 888888888	+++++++++
	S-TURN.						>555	5<			
	4-TURN.		XXX<<<	>>>>XXXXX XXXX<<<	,	>>44<<	>>>XX<<<	.	4	XXX<<	×
	S-TUKN.	22 2444	>3.3< TTHHHHH	>5<<	200	>33< FHUHE	>><<		>33<	>>3<<	>33<
	EXPOSURE.	60162644**		00+502630+	98*4182*90		0000020278 6160000000	6160000000	31*5**14*1		
261	SEQUENCE.	LLYPYGYTTQ		SIPDKTELNQ VAKSAVAALK SLYGTSYKYG	SLYGTSYKYG		SIITTIYQAS GGSIDWSYNQ GIKYSFTFEL RDTGRYGFLL	GIKYSFTFEL	RDTGRYGF LL	PASQLIPTAQ	ETWLGVLTIM
	SHEET										
	BRIDGE 2										
	CHIRALITY	*									
	BEND	5555									
	S-TURN	***									
	2 TIONN	>>>×									
	SUMMARK	HHHT									
į	EXPOSURE.	7684**842									
301	SEQUENCE.	EHTVNNIGY									
25).	IAPR ACID			PROTEINASE]	SOLO	CHINENSIS)					1
	SHEET.	⋖	AB C	ည ၁	BBB B	DO E	ω		Œ,	GGF	z: z:
	BRIDGEL	*	AB E	E	2 6	££ a	ō		;:	111	F.
	CHIRALITY		-+++	+	+-+++	++++1	+-+++	111+++-++	-+-++	+ 1 + 1 - 1 - 1	+
	BEND	ໝ	SSSS	SSS	SS	SSS	SS SS	SS SS	SS		SSS
	5-TURN.						,	,,,,,			
	3-TURN.	>33<	33<				^ ^	334			
	SUMMARY.	TT	TTSS BB	E SSS EE	EEESS B	EE BS SS	SS SS B	TTT SS	8 55	EEB	BSSS B
-	EXPOSURE,	**6489273*		5*04512090 4012*56715	0111022011	#111#22#11 11##681#55	00*8628715			9**175584* 2615150114	
4	SEQUENCE.	GVGTVPMTDY		GNDVEYYGQV TIGTRGKSFN		LNF DTGSSNL WVGSVQAQAS	Ga KGGR DKF N	PSDGS TF KAT	GYDASIGYGD	GSASGVLGYD	TVQVGGIDVT
	SHEET	5 500		888			II	I III	מ	×	LL
	BRIDGE 2.	,		ρį					•	;	;
	BRIDGEL	7 7 2 2		200	:		X XX	KKK	.	E,	
	BEND	SS	88888888	88 8	88888	+++++++	288	++	-+++		+
	5-TURN.	2		2			2	2		o c	9
	4-TURN				^			>444<			
	3-TURN	>33<		:		>3<<		>33<	^		
3	EXPOSURE.	4015121049	1244624631			1041068688	HHHHHHHSS SSS EE TTT EEE 1041068688 1**3000044 023**85130	EE E TTT EEE 140 023**85130	SS TTSB 2132*5**43	SS TTSB S B 2132*5**43 37*2591726	SSSS EE 21*1701070
TAT	SEQUENCE.	GGPULULAUR	LGGGGF PGDN	Derrereror		AFDQVSAQGK	LSITFÇESTN AKDQVSAÇÇK VIQPVRVVYL AASNISDGDR	AASNISDGDP	TM PGW I DNKY	GGTLENTNID	AGEGYWALNV
SUM	ARY	.H*ALPHA-HE	LIXE=BE	SUMMARYH*ALPHA-HELIXE*BETA-STRANDB*BBTA-BRIDGEG*3-HELIXI*5-HELIXT*3-,4-, OR 5-TURNS*BEND	B=BETA-BR	IDGEG=3-	HELIXI=	5-HELIX	T=3~, 4~, OF	5-TURN	S=BEND

E.	p p p	SSSSBEE 9668020 SLGEAIL		C CCC C PPP PPP PPP PPP PPP PPP PPP PPP	SEEEE 78*58 TAHGO	WWWW W	EEEEEE 194240 VDSYTA	
	\$888 8888	SSSSBEE 0049668020 GASSLGEAIL		CC	>33< EETTEEEEEE 4026070*50 TVGGVTAHGQ	μ. ω υ 3 ε	B 628* WSFN	S=BEN
z	R SSS	SSS B ***9314001 EDDSGbTSGI		CCCCCCC C OO PPPP P MMMM MM I	>33< PEEEEEEEE EETTEEEEE 9312014140 4826879*50 ASGNVFTDSV TVGGVTAHGQ	EEEE +++-+- \$\$\$\$ >5555	>>3<< >>3<<	S-TURN
	SSSS	SSSS Ø7148ØØ354 FEIITALGNA		THINELLUM)	4<< >33< >33< HH EREEEEE EEEE TTS E 782**5*768 Ø82*6**515 ATGRESGYT WSISYGDGSS	E B + R S S S S S S S S S S S S S S S S S S		.T=3-,4-, 0
0	888	>33< B STTS BSSS B SSSS 6155*4773* 301156*1* 0714980354 VIADTSAGFK SIPWSIYSAI FEITTALGNA		PROTEINASE] {FUNGUS: PENICILLIUM JANTHINELLUM} CC C D C C C C C C C C C C C C C C C C	<pre><< >>4 4<</pre> >33 >33 >33 >33 HTTS B HH EEEEEE EEEE TTS EEEEEEEEE EETTEEEEEE 495353870* 782**5*768 082*6**515 0312014148 4826678*50 0SGHSVXNPS ATGKELSGYT WSISYCDGSS ASGNVFTDSV TVGGYTAHGQ	BBBBB DDDD BBBBB +++S	>4 <x<44< 33<>>33<>33< 33<335 83406-615* 410000048* *812000074 FDTVKSSLAQ PLFAVALKHQ QPGVYDFGFI</x<44< 	=5-HELIX
z	R			UNGUS: PENIC D +++++	<< >33< HTTS 495353 QSGHSV	BBBBB DDDD CCCC +++	SEEEEE SS 410000048* PLFAVALKHQ	-HELIXI
	8888	SSSS 3897*73210 GAQDAALGGF		reinase] (Fu C D 1 1 K q ++-+++	>>44 << >33< E BSSS HHH HTTS 00048077*3 4953 VF STELPASQ QSGH	N A +++++++++++++++++++++++++++++++++++	>> >4 <x<44< >> 34<x>44< + HHHTTPBSS SEEEEE # # # # # # # # # # # # # # # # #</x></x<44< 	IDGEG=3
	**************************************	3< TSSSEE SS TTTGGGTTS 443404027* 1032816618 GTSLLILPDE AAVGNLVGFA			EEESS EE 2020110000 NFDTGSADLW	N N I I N N N N N N N N N N N N N N N N	BSSS 31***755 TVQPQSQT	.B=BETA-BR
¥	00 +++-+-+ 8888 88	3< TSSSEE SS 443404027* GTSLLILPDE	PJ TL TL BB BB APVA	CCC CCCC CCC HHH jjj I HHHHH H	>33< EEEE EEETTEEEEE EEESS 8286 8382768868 282811 :XITP VTIGGTTLNL NFDTGS	CCC KK jjj	>>3<< EEE S GGG 1000622952 LLGLAFSSIN	A-STRAND
יני	000	>3 S EE T 26*2*01011 GAIFQAILDT	KKK KKK NNN NNN ++-++ SSS >4444 EEETTTEEE B025*71050 VF DRDNGIRL	A BBBBBCC CCCC CCCC CCC CCC CCC CCC CCC	>33< TTS 278171 TANDEE	F ++++++++++++++++++++++++++++++++++++	>> >4<<< >>32< >>3< EEE SE H HHH TT SE EEE S GGG 181887*134 837*478118 1888822952 AVQAAQQISA QEQQUINNDG LLGLAFSSIN	CIXE=BE7
	**************************************	SSSS 510506*5*0 TGATADSTYL	P T ++++		B EEEEEE *472618862 AASGVATNTP	CCCC CC PPP 11 00	>> EEEE SEE H 101007*134 AVQAAQQISA	. H=Alpha-He
SHEET	BRIDGEI BRIDGEI CHIRALITY BEND S-TURN	3-TURN SUMMARY EXPOSURE.	SHEET. BAIDGE 2. BAIDGE 1. CHIRALITY BEND. 5-TURN. 3-TURN. SUMMARY. EXPOSURE.	26) JAPP ACID SHEET BRIDGEZ BRIDGEI CHIRALITY BEND	S-TURN 4-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGE2 BRIDGE1 CHIRALITY BEND	4-TURN 3-TURN SUMMARY EXPOSURE.	SUMMARYH=ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND

_
\boldsymbol{z}
6
-3
nued
٠
7
cont
\sim
AIII
ABLE AIII

(bu :	tt <	S SS S	>> >3 S SSEEE H *685802000 SGIGFLIFGD		2TLN B	SSSSSS >55555<	STTTTS 802444220 LSYDGNNAAI	J +++++	SSB SS SS 507277**64 DVYTPGISGD
HHHH FFF	BBBB AA uuu	S S	33 334 335 336 337 337 338 338 338 339 339 339 330 331 331 331 332 332 333 333 333 333 333 334 334 334 334 335 336 336 337 336 337 337 		THERMOPROTECLYTICUS}BB AA AA AA GG DD bb	++++++++ \$\$\$\$\$\$\$\$ >\$5\$\$\$ X <x<<4<< td=""><td>HHHTTTS 131Ø7*72*3 YDYYKNVHNR</td><td>DD KK ++-++ SS S SS SS</td><td>CGC HHHH HHRHHHRHHH HTT SSHH HHHHHHHHHHHHHH</td></x<<4<<>	HHHTTTS 131Ø7*72*3 YDYYKNVHNR	DD KK ++-++ SS S SS	CGC HHHH HHRHHHRHHH HTT SSHH HHHHHHHHHHHHHH
н нин	8 88	SS SSS >>555	>44 4 >33 >> 4 >> A >> A > > CREED TYPERESER ### ### ### ### ### ### ### ### ### ##			SSSSSS SSSSSSS SSSSSSS	SST	>>>XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	SSHH HHHHHHHH HHHHHHHHHHHHHHHHHHHHHHHH
99999	XXXX	SS	33< >>4 4 TTEEEEE GG 6748060608 SGYTATVPGS		OTEOLYTICUS AA bb	ŧ	987 087		GGG HHHH HHHHHHHHHHHHHHHHHHHHHHHHHHHHH
. •	_	S.	S EEEEE 3*22862621 SSVPDFSVSI		US THERMOPR BB AA GG DD	888	EEE TT S S EE EESSS 03204**8*6 6353252692 TYDAKYRTTL PGSLWADADN	++ SSS S 5556	HTT SSHH *4121998*6 YTAGLIYQNE
ннн ннн			/// / / / / / / / / / / / / / / / / /		NASE] (BACILL BB BBB ff fff E GG	8 8 8	EEE TT S S 0320***8*6 TYDAKYRTTL	**************************************	13.7.33. НИНННИННИ 01.10610007 АНЕЦТНА VTD
ннн	zzzz n	SS SS SS	<pre></pre>		CO-PROTEINAS B BB ff	5888	EEE SSSEE B SSTT EE 400*4**550 106853*002	**************************************	
Cu.	n	5555555	5)>>> << E HHHHHHH TT #598#66#1 952 LDDS VVSQYY SQV	IE CR CR BB BB PQA	JTRAL METALI AAA AA C DD AAA C	888	EEE SSSEE 400*4**550 NTTYSTYYYL	n h h	EEE SSSB 8003195*61 VYGDGDGQTF
E4 4	א מת ה צ	55 55	>33< E EE TT SSEE Ø 1020324303 G IADTGTTLL	I BBB B BBBB C CCC EEEE C ACC SEEE +++++++++-+ 5SSS 5SSS 5SSS 5SSS 5SSS 7SSSS 888118888337178888 IF LKSGYVVR DSDGPQLGFA	DROLASE: NEUTRAI A AAAA AAA C A AAAA AAA	888	EEEEEE E SSS EEEE EEE SSSEE 9448749138 *03*55*3*8 468*4**555 ITGTSTVGVG RGVLGDORNI NTTYSTYYY	BB BB II III Ph	EEEEEETT EE SSSE EEE SSSB PØØ134*36 57Ø6373810 ØØØ3195*61 RSSVHYSGGY NNAFWNGSEM VYGDGDGQTF
9 555 F	xxx	55	33< TTEEE 3*791*4268 GSQSGDGFSG	I BBB F + + + + + + + + + + + + + + + + + + +	THERMOLYSIN [HYDROLASE: NEUTRAL METALLO-PROTEINASE] (BACILLUS 12 AAAAAA A AAAA AAA C D E ff fff 11 AAAAAA A AAAA AAA C E GG	+ + + + + + + + + + + + + + + + + + + +	EEEEEE 9*48749138 ITGTSTVGVG	BBBB B hh h ffff +++ S SSS	EEEEEETTT *000134*36 RSSVHYSQGY
SHEET	BRIDGE1.	BEND		SHEET BRIDGE2 CHIRALITY BEND 2-TURN 3-TURN EXPOSURE	SHEET BRIDGE2	CHIRALITY BEND 5-TURN 4-TURN	SUMMARY EXPOSURE.	SHEET BRIDGE2 BRIDGE1 CHIRALITY BEND 5-TURN	
			201	361	27)		1		101

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

3134 st		M M M M M M M M M M M M M M M M M M M	BBBB HHH HHH S C C S C C C C C C C C C C C C C C C C	
######################################		2.20 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	##### #### #### #### #### #### #### ####	
			######################################	
++++++++ SSSSSSSSS >>>XXXXXXXX HHHHHHHHHHHHHHHHH		CCC D 0000 P	######################################	
++++ SSSS: >XXX >XXX HHHH		CCC D 0000 F	A + + + + + + + S S S S S S S S S S S S	
+ SS + S		•	A 12 8 2 12 24 55 54 56 56 56 56 56 56 56 56 56 56 56 56 56	;
ss \$S \ \ \ \ \ \ \ \ \ \ \ \ \		CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	BBBBB GGGG F-+++++-+ S SS 3< BEEEE SS HIGH003632 MICAGASGV	
++-+-+ 555S SS 555S S 6555C >> 177 7 33C 177 7 17 174 17 17 174 17 17 174 17 17 17 17 17 17 17 17 17 17 17 17 17		CCCCC CCCCC SSS EEEEE	######################################	
			BSSSSS S S S S S S S S S S S S S S S S	
++++++++++++++++++++++++++++++++++++++		TAURUS) CCC NNN KKK+++-++ S S S S S S EEES S TT 8682223464	######################################	1
		BOS TAURUS C CCC N NNN K KKK ++++++-++ 5.55 S S S S 33< TTSE EEES S TT TTSE EEES S TT TTSD	**************************************	
7423 VGIGI		: BOS:	SSSS >>>444 SSSS >>>44X >>>44X >> HHHH >> HHHHH >>>	
EE EE LL LL SSSSSSSS SSSSS S S S S 44<>>>><<<>>>><<<>>>><>>>><>>>><>>>><>>>>>>> HUNGILNA AYLISGGTH YGVSVVGIGR		E] {COM PANCREAS: BOS TAURUS} C C CCC L L MMMM LLL	BBBB DDD F F F F F F F F F F F F F F F F F F	
LLL S S S S S S S S S S S S S S S S S S		ON A CO	888 888 8 88 8 8 8 8 8 8 8 8 8 8 8 8 8	,
EE +++++++++++++++++++++++++++++++++++		C C CCC L L MMMM LLL LLL SSSS >334 > EETTEEEE	888 DDD BB SS SS 8929*1381 YITPDRLOQA YITPDRLOQA	
++++ SSSSS >>5555 XX< >>>555 3< HHHHHT #60018		SE) (C C C C C C C C C C C C C C C C C C C	+ 699 X1T	!
SSSSS SSSSS SSSSS SSSSS SSSSS SSSSS SSSS		C CCC C C L L L L L L L L L L L L L L L	SS S S S S S S S S S S S S S S S S S S	
++++++++ SSSSSSSSS 44<>>>XXX >>><<>>> TTTHHHHHHHHHHHHHHHHHHHHHHHHHHH		E PROTEINAS C CCCC J JJJJ SS +-++-+	BBB BBB BBB EE EE	i
		ROMD!	8	
++++-+ SSSSSSSS >>>44XX SSSHHHHT 7*3**\$51,00		LASE: SERIN CCCCC KKKK JJJJJ +++++ SSS CYBEEEE SSTTEEEE	# + + + + + + + + + + + + + + + +	
‡		C C C C C C C C C C C C C C C C C C C	33< TT 73 5\$51	
-++++- SS SSS 55< (>>3 77222369** YGDPDHYSKR		A BB A BB +++++ SS5 >336 >336 >337 SBT EF TT SY SW OF PROVING PAYON	+++	
S	++-+ SSS 5555< <<<< ><<< HHHT DAVGVK	A BB A B	68318 VSAV	
		PPSIN A P P P P P P P P P P P P P P P P P P	CCC B OOO SS SS BS CKEESS SS B CKLSTAASFS QTVSAVELPS H=ALPHA-HELIXE=BET	1
RK KK S SS SS S SS SS S S SS SS SS SS SS SS S	++++++++ SSSSSSSS >>>XXXXXXX >334 > 34 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	CHYMOTRYPE ++ SS +8#854+29#	CCC CCC OOO NM NA SS SS SS SS SS SS SS SS SS SS SS SS SS	
S S S S S S S S S S S S S S S S S S S	++++ SSSS >>>X >>>X >>	CHY + + + + + + + + + + + + + + + + + + +	CCC 000 000 MM = ++	
		GAMMA CHYMOTRYPEIN A [HYDROLASE: SERINE PROTEINASE] [COM PANCREAS: BOS CCCC C CCCC C CCCC C CCC C CCC C CCC C		
SHEET BRIDGEJ CHIRALITY CHIRALITY EEND 5-TURN 4-TURN SUMMARY EXPOSURE.	SHEET BRIDGE 2 CHIRALITY CHIRALITY BEND 5-TURN 4-TURN SUMMARY EXPOSURE.	SHEET SHEET BRIDGE2 BRIDGE1 CHIRALITY BEND 4-TURN SUMMARY	SHEET BRIDGE2 CHIRALITX CHIRALITX F-TURN 4-TURN 3-TURN STUMMARY	
O E E O E E O E E O		-	SHEET CCC B B BBBB BBBB BBBB BBBB BBBB BB	
201	381	28)	1.6 Su	,

_
$\overline{}$
~
\sim
02
===
(continued)
=
~
43
~
~
_
0
•
Ο.
$\overline{}$
_
~
_
•
AIII
Ą
_
7
_
~
_
_
$^{\mathrm{TABL}}$
,~
_
,

	A AAAA D BBBB q++++ +++-+- SS S33< ETTEEE B T T T T T T T T T T T T T T T T	FF -+	SS SSS *851887*70 KSAASLNSRV S=BBND
		CC CCC BB PPPPN NN N FF ++++	## BETA-TRYPSIN (NATIVE AT PH 8) [HYDROLASE: SERINE PROTEINASE] [COM PANCREAS: BOS TAURUS]
			SE: SERINE PROTEINASE] (COM PANCREAS: BOS TAURUS) L MMMM L MLL THL THL TKK THL TKK THL THL T
	HYDROLASE: SERINE PROTEINASE (MYXOBACTER495: LYSOBACTER ENZYMOGENES)	C P S 3367	SE: SERINE PROPEINASE] (COM PANCREAS: BOS TAURIC CCC CC CC CCC CC CC CC CC CC CC CC CC
	E PROTEINASE] (MYXOBACTER495; LYS BB BBB BB BBB BBB GG HHHH III III E	CCCC E CCC D CCC C NMMN R LL Q OO PPP P NNN R LL Q OO PPP P	PROTEINASE] (COM PANCE CCCC NKK +++++ +++++++++++++++++++++++++++++
	(MYXOBAC BBBB BB HHH HHH DDD HH S S S SSSS T T EEEETTE T T EEETTE S S S S S S S S S S S S S S S S S S S	CCC DO CO	PROTEINASE CCCC CCCC WNNN KNN ++-+++ S SS G SS EEEE G SS EEEE G SS EEEE
######################################	BBB BB G G G G G G G G G G G G G G G G	CCC CCC C C C C C C C C C C C C C C C	ASE: SERINE PR CCCCC LLL LLL -+++-++ SSS SSS SS >334 >>3 << ETTEEEE GG G SATIBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
BB C C C C C C C C C C C C C C C C C C	OLASE: SERIN BBBB BBB FF G EEEE EEE -++-SSS >33< EEEETTEEE	C C C C C C C C C C C C C C C C C C C	### PRYPRIN (NATIVE AT PH 8) [HYDROLA B B CCCCC CCCCCC CCCCCCCC KK LL LJJJJJ J JJJJJ J JJJJJ J JJJJJ J SS SSS
H HH II GGGG +++++++ S SSS EEEEEEE 73 3<>33<>34>34 N 7 TYSEEEE N 7 TYSEEEE LVGIVSWGSS 76STSTPGVY	ESSSS IN 18 18 18 18 18 18 18 18 18 18 18 18 18	CCC CC C C C C C C C C C C C C C C C C	A BB CCCCC A BB JJJJJ ++ 5 SS SSS SS S
# ## ## ## ## ## ## ## ## ## ## ## ## #		CCCC CCC C LL KKK KKK -++-+-+-+- SSS S > 444< EEBETTTEEE E 0001*55455 62	PER BETA-TRYPSIN (NATIVE AT PH SHEDGE)
SHEET BRIDGEL. CHIRALITY BEND 5-TURN 4-TURN 3-TURN EXPOSURE	SHEET BRIDGEZ BRIDGEZ BRIDGEZ BRIDGEZ CHIRALITY -++++ 4-TURN 3-TURN 3-TURN SEMMARY EXPOSURE. 6*82817	SHEET BRIDGE2. BRIDGE2. BRIDGE2. CHIRALITY ETURN 4-TURN SUMMARY. EXPOSURE.	1PTN BETA BRIDGE: BRIDGE: BRIDGE: CHIRALITY BEND. 5-TURN. 3-TURN. 3-TURN. SUMMARY. SEROSOURE. SEROSOURE.
201	29)	101	3¢)

0 4-1 × × × × × × × × × × × × × × × × × × ×		4	18 871		
I I I P P +-+-+ P SSSS S SEE SSSSBT 1153751184 VSWGSGEAQR		EEEEE NNN MMMM MMMMM	SSS EEEEE SSS 2817000251 QAVQRSGSTT		:
B B + + + + + S S S 11537. VSWGS		E E E E E E E E E E E E E E E E E E E	EEE VOR		BEN
		E E			ŝ
BB HH III III III III III III III III II			++ SS >33 24*2*96 AGNAFVG	•	:
1B BB C C C C C C C C C C C C C C C C C			4+2		URN
BBB BB BB HHH HH HH EE c II ++++ SSS >33< EEFTTEEEEE B18449198 W44546018			+++ SS >33 M8724*2*96 ITTAGNAFVG	•	5-1
0 H 1		מממ		第 U	g
A		ים יו	5SS EE SSS EEE 9447*7776*	! V . 18-4	4-,
SS SS >33)			SSS 47*7	++++-+- SSSSS >5555 >X<4< X3X3X3< HHHHHTT 90296558	3-,
A A B B DS					Ė.
BBB GGG F -++-+++ S SSSS SEES 733 <pre>>33 EEES 775</pre> <pre>PEAGYLOGG #</pre>			\$5555 \$5 \$5555 \$5 \$33< \$55775 \$E 18*9832528	F CCC KKK P III +++ SS C >> 33 B EEE HH 235633035 GGTTF YQPVT	;
S SSS > 33< > 175 677*1			SSSSS SSSSS >33< SSTTS 8*9832 SNPAAA	CCC KKK III III H++ EEE	(IT)
BBB GGG F -++-++ 5 SSSS >33< EEES TTSS 200077014 FeAGYLQGG		្នែង	SSSSS SSSSS >33< ESSTTS 18*983 HSNPAAA	F CCC KKK P III ++ C C C C C C C C C C C C C C C C C C C	2-HE
		5			H.
B G +-++++++		BB	: ~×	C CC F K KK	:
85 85 444 444 1111			555 SSS SBS 7828 PNND	KX KX H+++ S 259	LIX
SSSSS SSS >>>X< X444 + HHHHH HSTT 5922091 189992		Σ			HE
BB P F ++++++ SSSSSS >>>>X <br EB HHHHH 4125992098 PILSNSSERS		SPTC B F	SS >33< BTTEEEE 1621779153 WSIGTRTGTS	EEE EE 00 NN 00 ++++++++++++++++++++++++++++	G#3
358 358 358 369 369 388 388		STRE BBB FFF	28E	SS >33< SSS >33< SETTEE 325336	:
BB D P F C C C C C C C C C C C C C C C C C C			SS >33< BTTEEEE 16217791	EEE SEE SEE SEE SEE SEE SEE SEE SEE SEE	DGE.
		ASE	585858 8 88 >>44< >+++++		BRI
BBBBB DDDDD BB ++- SS EEEEE 730401*0		BIN	SS SS SS NIS	C ++++	TA-
\$5 738 738		ROT	SSSSS S >>44< HHHHSS S MG40068349 AGHCTNISAS	C E + + + + - + + + + - + + + + - + + + +	#B E
BBBBB		8			
SSS SS		SERINI BBBBBB DDDD BBBB	A COSE	CCCC H HH III 	ē.
1 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8		SER BBBBI DDI BBBB	SY SY SY A I	CCCC H HH III S S S S S S S S S S S S S S S S S S	FRA
S S S S S S S S S S S S S S S S S S S	+ SS	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	SSS >33< EETTEEEEE 42*6422000 SVNGVAHALT	CC CCCC HH H HH 9 III -+++ S S EEEEEEEES 70*3017071	15-A
		ROLASI BB BB C C BB BB			B E T/
BEBBBB EE DDDDDD +- EEEEEE *14000	SSS X X X X X X X X X X X X X X X X X X	CK H	EE EE (CSLGFNV		Ŧ.
BBBBBB EE DDDDDD -++++- SS 733 TT EEEEEE 8664*14809	++++++++ SSSSSSSSS >>>>XXXX (<<<>>>3 713950**14 NYVSWIKQTI	A A E	S EE EE 95*1008111 GSRCSLGFNV	C CC H HH ++ SS SS(EEEE GGG *282*63*91	:
		GPA			ELIX
B (C +	D BEBBB IIII P GGGG ++++ S SS 3< >>>> TB EEEEEGG 87000000000	S)	SSS SS SSS SS SSSEE SS 4000550457	O 표! 월경일	4-HE
C B 8511	D BBBBB LIII P GGG +	SE A AA AA	SSS SSSEE SSSEE 1005504	EEEEEC ++ EEEEEEEE 58*1506 LRSGSVI	CPH/
3518	S +++S +++S +++	INA	SS 24.00 SI IA GC	BE NAME OF THE NAM	H≖A
	• • • • • • • • • • • • • • • • • • • •	PROTEINASE A (SGPA) [HYDROLASE: SERINE PROTEINASE] AA BB BB BBBBBB BE2 C DDDD C DDDD E2 AA AA BB BB BBBBBBBBBBBBBBBBBBBBBBBB	× · · · · · ·		SUMMARYH-ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND
SHEET BRIDGE2 BRIDGE1 CHIRALITY CHIRALITY CHIRALITY A-TURN 3-TURN SUMMARY SUMMARY	SHEET BRIDGE2 GRIDGE1 CHIRALITY BEND 5-TURN 4-TURN SUMMARY	1SGA PROSHEET BRIDGE2 BRIDGEI	CHIRALITY BEND 5-TURN 4-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGE J CHIRALITY BEND S-TURN 3-TURN SUMMARY EXPOSURE.	:
ALIBO ALIBO ALIBO ALIBO ALIBO ALIBO ALIBO ALIBO	ALEET ALIDE ALINA ALINA TUR TUR TOS	1SGA SHEET BRIDG SRIDG	TIRE TIUR TIUR TIUR TIUR TOOS	ALEET TING TING TING TING TING TING TING TIN	ζ
•		STREET	SE 34-5-1	•	IMAF
191	281	31)		161	S

_
ă
ā
3
nne
٠.
ب
2
0
ಿ
AIII
ĹΕ
긎
щ
ABĬ

AA AAA BB CCCC aa aaaa JJ	AAA	CCCCC NUNN LLLL S SS S
SS S SSTT >33< SSTT 12000232	C AAAAAA A AAAAAAA A AAAAAAAAAAAAAAAAA	++++- SS SS SSS 33<>33<>33<>34 TT TGGG **3867423
NS} +-+-+-+ SS SSSS 4<	AAAAAA	SEST TOSYL-ELASTASE [HYDROLASE: SERINE PROFEINASE] [PIG PANCREAS: SUS SCROPA]
SUBTILISIN BPN' [HYDROLASE: SERINE PROTEINASE] [PROBABLY BACILLUS AMYLOLIQUEFACIENS] AAAA AA AAAA AA AAAA AA AAAA AA 1 Aaaa aa 1 SSSSS S S S S S S S S S S S S S S S	C K K K K K K K K K K K K K K K K K K K	OF A) OF A C C C C C S S S S S S S S S S S S S S
CILLUS AMYLO SS SS 733 734 1 1****51992 1 VPSETPNFQD	AAA AA e ee ddd d	AS: SUS SCROPA) MANH JJJJJ +++ S S S EREEES S BT 8915808121 13 E. LTP RVVVGEH NI
(PROBABLY BAC		HYDROLASE: SERINE PROTEINASE PIG PANCREAS: CCCCC C CC
SERINE PROTEINASE] {PAAA AA	SSSSSSS >>>>XXXXX HHHHHHHH 2594885866 GSAALKAAUD ++++++++ 555555558 (XXXX44	ROTEINASE] (PC CCC C CC C CC C CC C CC C CC C CC
AAA AA	AAAA C dddd bbbb K bbbb K 5858585 S S S S S S S S S S S S S S S S	HYDROLASE: SERINE PRC CCCCC C CCC C JJJJJ II IIII I H+++
[HYDROLASE: ++++++++ S SSSSSSS 55< >55< >555 <x>4<!--</--> HTHHHHTTT BG64233*8 I KAPALHSQG</x>		-ELASTASE [HYDROLASE: SI A BB CCCCC C C C C C C C C C C C C C C
TLISIN BPN' SSSS >5555 >>>>< >>> >>> >>> >>> >>> >>> >>> > >> >> >> >> >> >> >> >> >> >> >> >> >> ><	1JJ +++++++ SSSSSS < >>>>XXXXX	TOSYL-ELASTASE A BB 11. BB 11
SHET SUBT SHEET BRIDGEL CHIRALITY BEND 5-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGEL BRIDGEL F-TURN S-TURN 3-TURN 3-TURN SUMMARY SEACOURE SHEEL BRIDGEL SHERNTURNTURNTURNTURNTURNTURNTURNTURNTURNTURNTURNTURNTURNTURNTURNTURN	SHETTOSS SHETTOSS SHETTOSS BRIDGELTOSS CHIRALITY BRND 5-TURN 3-TURN 3-TURN SUMMARY EXPOSURE
32)	161	33)

BBBBBB BB GGGGGG CCC +-+ SS BBJB64+7** PLHDLVNGQY		+ SS S >> S53-73-489	AAAAA GGGGG H ee+
BBBBBB GGGGG CCC CCC +		SS SS S73*	AAAAA GGGGG H ee +
A ++++ S SS SS SS SS TT TT TT TT B001*10300 RSGG QCDSGG		E C	A B A B A F I H H L L L L L L L L L L L L L L L L L
FFFF A ++++++- S SS S SS S		++++++++++++++++++++++++++++++++++++++	FF CGGG FFFFF SSS >33< EETTEEEEEE 7*87652888 TEGGUDYMIV T=3-,4-, OR
+++ SSS SSS S 5555 33<>>>3<<>>>3<<>>>		. + o ^ \ = a H	AAAAA AAA FF FFF CCCC AA +-+++ EEEEEEEEE 911818888 DHAIVINGYG
		ACTINIDIA CHINENSIS] J	AA AAAA AAA AAA AAA AAA AAA AAA AAA AA
B F P BBBB D DDDD		JIPRUIT: ACT C J +-+-+++ SSS 55< 6746754080 TSGSLISLSE	AA
B F D Q Q = +++++++++++++++++++++++++++++++	+++++++ SS SSSSSSS >> >>XXXCCCC	TDRYL PROTEINASE KIW 	AAAAA D CCCC S CCCC S EEEEE S EEEEE QPVSVALDAA B=BETA-BRI
BBBB DDD CCC SS SS TT EEEE 476*256609 LANNSPDYIT	BBBBB HHHH FFFF SSS >>3< EEEEEGGG	4++++ SS	AAAA AAAAA BBBB BBBB CCCC
55 >33< TT 61651*654*	G H H R S S SS SS SS TTBSSBTTB **10548*50	DLASE: SULFH B ++ 5 5 7 4022*23515 GAVVDIKSQG	AAAA BBBB +- SS EEEE TT 8559*18*67 DTYENVPYNN
E E SSS > 33<	BB BB B G G H H HHH H G G G R R S S -++	A SHERT A SHERET A SHERET A SHIDGE 2. A THAN THAN THAN THAN THAN THAN THAN THA	BRIDGEL. L BBB CCCC A FFFFF F GGG G GGGG GGGGGGGGGGGGG
SHEET BRIDGE1 CHIRLITE BEND STUBN 4-TUBN 3-TUBN SEMMARY EXPOSURE.	SHEET BRIDGE 2 BRIDGE 1 CHRALITY BEND 5-TURN 4-TURN SUMMARY. EXPOSURE.	2ACT ACTI SHEET BRIDGE2 BRIDGE2 CHIRALITY BEND 5-TURN 4-TURN SUMMARY. EXPOSURE.	SHEET BRIDGEL. BRIDGEL. CHIRALITY BRUD 5-TURN 4-TURN 3-TURN SUMMARY. EXPOSURE. SEQUENCE.
191	291	34)	101 SUM

_
$\overline{}$
α
ã
3
=
~
7
\bar{z}
8
υ.
_
_
=
⊸
ď
4
3
\mathbf{x}
A H
ч,
_
٠,

	[HYDROLASE: SULFHYDRYL PROTEINASE] [PAPAYA FRUIT LATEX: CARICA PAPAYA] P C B C B C B C C B C C C C C C C C C C	REF GGG GGGG GG	
	ICA PAPAYA). - +++++++ S SSS SSS 5 5 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 8 8 6 4 4 3 2 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	AA AAA AA AA FE F G C C C C C AA + - + S S E E E E E E E E E S E S E S E	
	T LATEX: CARIC +++++++++++++++++++++++++++++++++++	AA CC CC -++-+ S S S EE	
	AYA FRUIT B H +-+++ SS <<<>>> H H B H *85*636808 R TGNLNQYSE	AA +++++++++++++++++++++++++++++++++++	
		A AAA AA AAA AAA AAA AAA AAAAAAAAAAAAA	
	YDRYL PROTE ++++++ SS SSSS SSSSS SSSSSS	AAA A A BBBB C SS SSSSSSS	
AAAA D BBBB +++ S EEEE 8005034*	AA AA		% NX **
AAAA BBBB SS SSSSS S >33X>3X<3X<3X >521188884 8685334* GGAGTCGIAT MPSYPVKY	S +-+ A P S F Y P E Y V	C A 1	AAAA BBBB ++++++
SHEET BRIDGE1. CHIRALITY CHIRALITY B-TURN 4-TURN 3-TURN SEPOSURE. SEQUENCE.	BRPAP PAPAIN BRIEGEZ BRIEGEZ BRIDGEZ CHIRALITY EEND. 5-TURN. 3-TURN. 3-TURN. SUMMARY. SECOUNCE.	SHEET BRIDGE2 CHINDE1 CHINDE1 FINDR 5-TURN 4-TURN SUMMARY EXPOSURE.	SHEET BRIDGE2 CHIRALITY BEND 5-TURN 3-TURN SUMMARY. EXPOSURE.
261	35)	101	201

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

BBBBB BBB GGGG GG GGGG GG L++ EEEEEEEEEE	NNN LL SSS EEEE RESSS EEEE 841*8485*6	E C C V V V V V V V V V V V V V V V V V	13333 22222 22222 22222 22222 22223 22223 22223 22223 22223 22223 22223 22223 22223 22223 22223 22223 22223 22223 22223 22223 22222 22222 22222 22222 22222 22222 2222
BBBBBB B GGGGG G EEEEE +	CCCC	SSSS EEEE 6**67020********************************	33 + 537 + 548 AVLQSSGLYS
BB GG E ++++++- S SS S ST E S ST E S ST E TGLQAEDEAD		EEEE 0000 1111111111111111111111111111	JJJJJJ JH KKKK J ZZZZ Z Y X W CCC Y Y X W CCC
AA AAAAAA CCCCCC CCCCCC C BBBBB SS ++ SS EETTEEEEEE *865482818 KSGSSATLAI	CC	G GGG S UUU -+++ SSS S S S S S S S S S S S S S S S S S	CCCC +++ SSSS SSSS >444 >444 S EEEETTT 84282413*6 EPVTVSWNSG
AAAA AA AAAA AA CCCC CC +++++	CCC	GG GGG GGG SS	33222 2 2 2 2 4 4 4 4 4 4 4 4 4 4 4 4 4
BB -++	DDDDD D LLLLL M ++	GGGGGGG TTTTTTT TTTTTTT SSSSS C +	SS SSS > 33 < SS STTS 2**3896312 SKSTSGGTAA IDGEG=3
**************************************	CCC HHH THH THH SSS EEE SSS BØ25851373 LISDFYPGAV	P P 0000++++ SSS SSS SSS SSS SSS SSS SSS SSS SSS S	GGFF H I J + SS
AAA BB BB +++++ SSSS SSSS SSSS SSSS SSSS	CCCC IIII ++++ SSSS X4<4< ><3< TTTT EEEE 7*8*711100	F EEEE PPPP	GGFF H SS SEEE SS B 40205*6*59 LVTVSSASTK
BULIN FAB BB BB dd +++ SSS >33 SE TTS EE 7027781706 SGAPGORVII	HH HH 5SS >4>4 932145177 VTLFPPSSEE	EEEEE F 00000 r +-+++ S EEEEE SE 7*2*372714	GG G VV V S + S EEEEE S S318521913 GCIDVWGGGS IXE=BET
i	6	DD NN ++-+++ SS EE SS 8347**4*0* TVAPTECSIX	GGGGGG GG GGFF H I J ZZZZZ Z Z CCC
JFAB LAMBDA SHET BRIDGE2. BRIDGE2. CHIRALITY CHIRALITY 	SHEET BRIDGES. BRIDGES. CHERALITY BEND 4-TURN 5-TURN SUMMARY. EXPOSURE.	SHERT BRIDGE2. BRIDGE1. CHIRALITY BEND 5-TURN 3-TURN SUMMARY. EXPOSURE.	SHEET BRIDGEZ. BRIDGEZ. CHIRACILIY BEND 5-TURN 3-TURN SUMMARY. EXPOSURE. SEQUENCE.
36)	181	201	361 SUM1

_
$\overline{}$
B
o,
-3
une
\simeq
~
+>
(continue
\bar{a}
\simeq
٠,
_
- 4
_
=
A
Ξ
٦.
IABI
Q,
- :

i		H #0		PPGM +++ 15555	:
		1REI H+++ SSSS SSSS B 474834328+ YOSLPYTFGO		**************************************	ND.
		SSSS 47483		s s s s s s s s s s s s s s s s s s s	S≖BE
		BBBBB B BBBBB B BBBBB B B		A	:
		BBB HH EEE 5 SEEE 61301		A d d d d d d d d d d d d d d d d d d d	-TUR
		\$ \$ \$ \ \$ \$ \$ \$ \ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$		AA A A A A A A A A A A A A A A A A A A	OR 5
		\$2 \ \$3.000		A d d d d d d d d d d d d d d d d d d d	4-,
		AAAAA BBBBBB CCCC HH BBBB + FFEEEE SS S S EEEEES G GG SEFEEEE >> 3		-++++ SS 555 1155**ADRLMADRLMADRLMADRLMADRLMADRLMADRLMADRLM	ſ=3-,
				## HOGLYCERATE MUTASE (DE-PHOSPHO) [ISOMERASE] [YEAST: SACCHAROMYCES CEREVISIAE] ### AA A A A A A A A A A A A A A A A A	
		AAAAAA A CCCCCC C C C C C C C C C C C C		EVIS ++++ 5555 XXXX HHHH 6821	ELIX
		**************************************		CER +++ SSSS SSS	.=5-H
		BB AA GC CC CC SS		CHAROMYCES CEREVISIAE] AAAAA ddd ccc+ ++++++++++ S SSS SSSSSSSSSSSSSS	:
				AAAAA AAAAA ddd ddd ddd dcc ccc	ELIX
		VARIABLE PORTION (REI) (HUMAN)		SACC S S S S S S S S S S S S S S S S S S S	3=3-5
		UMAN) BB BB GG FF F -+++ SS SS A4 EEEEET PKLLIYE		(YEAST: ++++++ 5SSSSS >555<< XX<<<< HHHHHS 118**36 LLKEKGW	:
		I) (H S SS 33 3 3 3 3 4 5 1 T T S 8 8 8 3 3 4 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S 5 1 T S S S 5 1 T S S S 5 1 T S S S S S S S S S S S S S S S S S S		-PHOSPHO) [ISOMERASE] [YEAST: S++ ++++++++++++++++++++++++++++	IDGE
		NN (REI) BBBB FFF EEEE S S IEEEE 73 FFF FFF NYOOTP		RASE] ++++ SSSSS XXXXX HHHHH 1900*	TA-BI
		BBBBBBB FFF FFF FFF FFF FFF FFF FFF FFF		1SOMERASE] +++++++ ++++++ 5SSSSSSSS >XXXXXXXXXXXXXXXXXXXXXXXXXXXX	8≖8
	* U	AAAAA BBBBB BBBBB BBBBB BBBBB BBBBB BBBBB BBBB		SS S S S S S S S S S S S S S S S S S S	:::
	KKK	VARIABLE PO AAAAA BBBB AAAA +++ SS SEEEESS T R40918*90B		OS PHC	TRANI
	KKK KKK DDD DDD C	AAAAAAAAAAABBBBBBBBBBBBBBBBBBBBBBBBBBB		SS >33< TT TT 1134 GWVD'S	TA-S
	DDD DDD + 55< C TEEE	BB		TASE (D +++-++ SSSS >44<< >33< HHHHS 25*8552	E=BE
	KKK KKK DDD CC CC DDD -+-+-+ SSS >5555< >>33 >>3 WHKPSNTKV	NOGLOBULIN BBB BB ddd -+-++ SS >33		MUTASE (D) -++++-++ SSSSSS >>44<< >>33< SSHHHS Ø75258852	: ×
	KKK KKI DDD DDD CC CC CC CC	IMMUNOGI B BBB d ddd -++ -+- S SE EEE SE EEE SS LSA(SA)		. ATE	HELI
	1 8 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	-JONES IMMUNOCLOBULIN AAAA B BBB AA AAAA d ddd BB AAAA d ddd SS++++++ S S SS EEEE SE EEE TT EE AP#1686467 3814857*15 D10MT0SPSS LISAVODRVT	BBBBB HH dddd eeree	HOCLYCERATE MUTASE (DE-PHOSPHO) [ISOMERASE] (YEAST: SACCHAROMYCES ABABA BB	LPHA-
	++++ 5SS 33< 17555 *39**	10 10 10 10 10 10 10 10 10 10 10 10 10 1	BBH HH dc 60996 GTMI	PH OG	. H.
	CHITY CHITY CALL	BERET. BENCE-JONES IMMUNOCLOBULIN VARIABLE PORTION (REI) (HUMAN)	EBSK	SARAN PHOGENCERATE MUTASE DE-PHOSPHO [ISOMERASE] (YEAST: SACCHAROMYCES CEREVISIAE] AA A A A A A A A A A A A A A A A A A	SUMMARYH=ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND
	SHEET BRIDGE2 CHIRALITY CHIRALITY 5-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGE2 BRIDGE2 BRIDGE1 CHIRALITY BENDOCT 5-TURN 3-TURN	SHEET BRIDGE2 BRIDGE1 CHIRALITY BEND 5-TURN 4-TURN SUMMARY EXPOSURE.	SHEET BRIDGE2. BRIDGE1. BRIDGE1. BRIDGE1. 5-TURN 5-TURN 3-TURN 5-TURN	RY.
-	4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	38) S S S S S S S S S S S S S S S S S S S	SUMMA
	•	m		m	

-++++++ SS SSSSSS <		AA. AA. Cee d d d ++-++++++++++++++++++++++++++++	SSSS SSSS XX<< <x>>>>< HHHHH HHHH #### ##### ################</x>
S S S S S S S S S S S S S S S S S S S		S GALL AA	S SSS. X XX X
++++++++ -+ SSSSSSSSS SS >>>>XXX<<< 37 THHHHHHHH SS INSLR64*1 95:		USCLE: GALLUS GALLUS AA AA e e dd d -++ ++++++++++++++++++++++++++++++	#####################################
AAA ccc a +-+++++++++++++++++++++++++++++		C	SSS SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
		(CHICKEN B AAA ddd ddd ++-+ S SSSS B11201657* AAQNCYKVPK	AA AAAAA 9 999 f fffff ++
######################################		COSPHATE-KETOL-ISOMASE COSPHATE-KETOL-ISOMASE COSPHATE-KETOL-ISOMASE COSPHATE-KETOL-K	AA AAAAA 9 999 ff ffff ffff ffff ffff
+++++++ ++++++++++++++++++++++++++++++		AAAAEAETEL-ISOMASE AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	**************************************
		DEHYDE-3-PHO AAAA A CCC A aaaa a ++++ SS SS EEEE E *96837119 8 KLSADTEVVC G	######################################
++++++++++++++++++++++++++++++++++++++	A S S S S S B S TT S B S S TT S B S TT	ERASE [D-GLYCERAL -+++ ++++++++++ -SSS SSSSSSS >>>> XXXXXX<<<< >>>> XXXXXX<<<<88-41 1-7-10488-41 1-7-10488-41 1-7-105	AAAAAAA ffffff ++ S S EEEEEEE H 1010001094 GVIACIGEKL
++++++++++++++++++++++++++++++++++++++	### ##################################	E PHOSPHATE ISOMERASE	+++++++++++++++++++++++++++++++++++
SSSS SSSSS 5SS SSSS SSSSS 5S55 > 555 5 4<<>>>>4<<>>>>4 HHT SHHH HSTSS LKKFGEEKFN TYRRSF L	B AA	TRIOSE PHOSPHATE AAAA 3.2. aaaaa 11.TY N N N N SEEEEE N N N.	BRIDGE: ### ################################
SHEET BRIDGE1 CHIRALITY BEND 5-TURN 4-TURN 3-TURN EXPOSURE.	SHEET BRIDGE BRIDGEI CHIRALITY BEND 5-TURN 4-TURN SUMMARY EXPOSURE	SHEET SHEET BRIDGEL BRIDGEL CHIRALITY BEND 5-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGE2. BRIDGE2. CHINGE1 CHINGE1 5-TURN 3-TURN SUMMARY. EXPOSURE.
181	201	39)	101 SUMJ

(continued)
AIII
TABLE

	BBBBBB BBBBBB CCA AAAAA IIII	BB BBBB BBBBBB A AAAA BB BB B	
	AAAAAA AAAAAA AAAAAA AAA BBB BBBB BBBB	A AAAA BB BB R R R PEPEF MM	
	BBBBB BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB	A + + + + + 5 S S S S S S S S S S S S S S S S S	
	ANA VALLA ENSIFORMIS) BBBB BBBB B B B BBBB B B B BBBB B B B B	BB NN JJ SSS SSS SSSE S *128360*56 DGNLELTRVS	
++++ SSSS <<<< >33< HHHT *0161**	NAVALIN A [LECTIN, AGGLUTININ] (JACK BEAN: CANAVALIA EN AAAAAA BBBBB CCC GGGG -++ ++++++- ++++	A AAAA BB F FFFF NN SS SS 1 4774418147 L ILQCDATTGT	
AAAA hada hada hada hada hada hada hada	(JACK BEAN: AAA BB CCC -+ + -+ SS SS SS ESS SS SS ESS SS SS	S SS S	SSS SS S >4 44
AAAA bbbb SSSSSSS SSS >4>XX<<< < THHHHH TSTT EEEE TTHHHHH TSTT EEEE A45451816 6*518828	AAAAAA CCC BBBBBB +-+	BBBBBB LLLLLLL ++++ + \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	SS SSS SS SSS SS SSST BS SSST 1 PSGS TGRLL
+-++++++ + -++ SSSSSSSS SSS >4>>x<< < < >	LECTIN, AGG	BB BBBBB BBBBBBB BBBBBBBBBBBBBBB	A AAAAA DDD A AAAAA
AAAA	SHEFT AAAAAA (AAAAAAAAAAAAAAAAAAAAAAAAAAA	BB BB K MM JJ LL +-+++	SSSS
SHEET BRIDGEZ. BRIDGEZ. CHIRACII. CHIRALITY BEND 3-TURN 3-TURN EXPOSURE.	3CNA CONC SHEET BRIDGEZ. BRIDGEZ. CHIRALITY BEND 5-TURN 3-TURN SEQUENCE.	SHEET BRIDGEZ BRIDGEJ CHIRALITY BEND 5-TURN 4-TURN SUMMARY. EXPOSURE.	SHEET BRIDGE1 CHIRALITY BEND 5-TURN 3-TURN SUMMARY EXPOSURE.
201	48)	101	201

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

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

BBBB HHHH F F + + + + + + + + + + + + + + + + + +	BB NN ++-+-+ SSS EE SSS F0813705 PGSLTTPPLL		AAA AAA eee eee eee eee eee eee eee eee
BBB BB HHH HH GG FF SS SS EEEEE SS EEEEEE 2*53051431 LDGTKHLIQF	BB MM L L S S S S S S EE 52*2860212 LPESLDWTY		
+ S EES4	D D ++++++ S SS SS TTTT F TTTT F 28*692*51 DF TWF DPRGL		f t LACTOBACILLUS CAS f t ++++++++++++++++++++++++++++++++++
	_		***OLATE/ACCPTR] (BACTERI
IEN LYASE, CARBONNTE DEHYDRATASE BBBB	+++++++++ SSSSSSSS >4x<4<< 33x3x3 <x> HHHHHHGGG 92.14*18*51</x>	BB MM MM	AAA AAA AAAAABBEERHHYTSS SSS SSSEERERHHYTSS SSS SSEERERHHYTSS SS SSEERERHAABBAAABBAAABBAAABBAAABBAAABBAAABB
CARBONATE DEH BBB CCCC SCCC S SS SS SEEE 71 4 4 7 3 4 5 6 1 7 3 4 5 6 1 7 7 3 4 5 6 1 7 7 8 1 7 8 1 8 1 8 1 8 1 8 1 8 1 8 1	BBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB	C C BB ++++++++++++- SSS SS SS S S S S S S S S S S S S S S	E: NADPH/DONR, DIHYDROF AAA eee Gdd H+++++++++++++ SSSSSS SSSSSSS X<<<<>>>4 X<
SSS S >444< >444< >33 TTTS TT 85*917*3** DTHTAKYDPS	######################################	C ++++-++++ S S 3< T S B 667*2*0350 GEPEELMVDN	SE: NADPH/DC AAA dee edee +++++++ SSSSSS X<<<<
AB FORM C [CARBON-OXY AA +++++++++++++++++++++++++++++++++++	B H -++++ ++-+++ SSSSSSS SSSSS >>< 4< < ` ' ' ' ' ' ' ' ' ' ' ' ' ' ' ' ' '	KKKK 0 + +++++ + -++ SSSS SSS SS S S S S S S	**IDOREDUCTAS -++++++++++++++++++++++++++++++++++++
NIC ANHYDRASE FORM C [CARBON-OXY AA +-+++++	BBBBBBBB IIIIIII HHHHHHHH SEEEEEEE S010011102	1 4.0	AAAAAA A B B B B B B B B B B B B B B B
CARBONIC ANHYDRASE FORM C [CARBON-OXYGEN LYASE, CARBONATE DEHYDRATASE] BEBB BBBB BBBB BBBB BBBB BBBB BBBB	AA A B B B B +++	BBBBB NN L 1)1)1) ++-++-S S EEEEE SS 50130010*7 EC VTW I VLKE	AAAAAA AAAAAA babbbbb C SS EEEEEETT 32002121**
SHEET SHEET BRIDGEL. BRIDGEL. CHIRALITY BEND 5-TURN 4-TURN SUMMARY. EXPOSURE.	SHEET BRIDGE1. CHIRALITY CHIRALITY BEND 5-TURN 4-TURN SUMMARY. EXPOSURE.	SHEET BRIDGE 2. CHIRALITY CHIRALITY CHIRALITY A-TURN 4-TURN SUMMARY EXPOSURE.	1DFR DIHYD SHEET. BRIDGE2. BRIDGE1. CHIRALITY BEND. 5-TURN. 4-TURN. SUMMARY. EXPOSURE.
41)	101	201	42)

ed)
continued)
_
AIII
Ξ
ABLE
~

		IGPD + C S S S S S S S S S S S S S S S S S S	E P + + + SSS SSS SSS SSS SSS SSS SSS SSS	000 000 000 000 000 000 000
		AMERICALGPD A AAA d ddd a aaa s aaa	E P +++++++ S S S S S S S S S S S S S S	G G G OOO G G G OOO SS SSS SSS SSS SSS SSS SSS SSS SSS
		:: HOMARUS A d d d -+++++++ S SSS > 5555< > 4446 < > 3346 < > 3348	E ++++++ S -+++ S S S S 120536*	F ++++ ++ 4 SSS SS
		FER: HOWAR + -++++-+ 5 S S S S S S S S S S S S S S S S S S S	SS SS SS TT SB TT SB TT SB TAT OKTVD	5 58 58 58 58 58 58 58 58 58 58 58 58 58
		AA	LLLL KKKKK KKKKK SS SS SS SS SS EEEEE TT TT BSSS MITUHANTAT OKTUDGPSAK DWRGGRGAAQ	9 9 9 S 5 5 S 73 3319***493
	* :	HYDE/DONR, NAD/ACCPT CCC CCC CCC TIX HHHH HHHH III S SS. SSS SS. SSS SS. SSS SS. SSS SS. SSS SS. SSS	D DDDDD E E E E E E E T S S S S S S S S S S S S	F+++++++++++++++++++++++++++++++++++++
(par	AAAAAA B A AAA AAA AAAAAAAAAAAAAAAAAAA	3-PHOSPHATE DEHYDROGENASE [OXIDOREDUCTARE: ALDEHYDRODON,NAD/ACCPTR] [LOBSTER: HOWARUS AMERICAIGPD A AAA C C CCC CC AA		DDDDD DDDDD DDDDD DDD DDDD F F F F F DDD DDDDD F F F F
ABLE AIII (continued)	A AAAA H HHHH + ++++ ++ SSSS S S S S S S S S S S S S S S	SE [OXIDOREDUCTASE: A B B B B B C C C C C C C C C C C C C C	AAAA f eeee ++++ ++++++++++++++++++++++++	DDD DDDD NNNN LLL L
IADLE A	A A H H H H H + + + + + + + + + + + + +	ASE [OXIDOREDUCT] ++++++++++++++++++++++++++++++++++	++++++++++++++++++++++++++++++++++++++	DDD N THE LLL S S S S S S S S S S S S S S S S S
	B	: DEHYDROGENASE [OXIDOR AA C C C C C C C C C C C C C C C C C C	AAAAA	DDDDD MMMMM KKKKK SSS SSS EEEEE 66*3826065
	AAAAAAA GGGGGG bbbbbb +	-+ ++++++++ ++ ++++++++ +	AAAAA eeee dddd +-+-+ S S SS < C SEEEEES 21*3001226 GAKKVVISAP	DDDDD HWMMM MMMMM KKKK +++ ++++++++++++++++++++++++
	++++++++ \$\$\$\$\$\$\$\$\$ >>XX<<< 3<< >33 *198528881	D-GYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE [OXIDOREDUCTASE: ALDEHYDE/DONN,ND/ACCPTR] [LOBSTER: HOWARUS AMERICAIGPD A AAA A AAA A AAA C C CCC CC AA d ddd ddd	+++++++ +++++++++++++++++++++++++++++	DDDDD MMMM
	SHEET BRIDGE2 BRIDGE1 CHIRALITY BEND 5-TURN 3-TURN SUMMARY	1GPD SHEET BRIDG BRIDG BRIDG BRIDG BRIDG BALDA 3-TURI SUMMAN	SHEET BRIDGEZ BRIDGEZ CHIRALITY BEND 5-TURN 3-TURN SUMMARY EXPOSURE.	SHEET BRIDGE1. CHIRALITY BEND 5-TURN 5-TURN 3-TURN SUMMARY EXPOSURE.
	ç	43)	191	201

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

	HAPPINGS HORSE LIVER: EQUUS CABALLUS	FFFF 0000 0000 0000 0000 0000 0000 000	BRIDGEL. BRIDGEL. CHIRALITY BRIDGEL. CHIRALITY CHIRAL
	VER: EQUUS CC CC I H H H S SS SS EEEEE TT 0613232*51	++++++ +++++ SSSSSSSS SSSS >>5555 >>>XXXXXXXXXXXXXXXXXXXXXXXXX	SSSSSSS SS SS >>>>XXX >>>>XXXX >>>>XXX >>>>XXX 95*1033002 803*5*8 95*1033002 803*5*8 RLDTMVTALS CCQEAYC
	CHOH/DONR, NAD/ACCFTR] [HORSE LIVER: B D CC	-++++++ SS SSSS >444< >333X>> SS TTTGG 867841778 BAASPLEKUC	G FFFF 9999 50000 8888888 88888888 60042000281
	HAD/ACCPTR +++-+-+- 555555 6666 HHHTTSS 3015244*14 HVVSGTLVTP	A CC CCC EEE B GG GGG MMM SS SSS SSS 444 >	PP P +++++++++++++++++++++++++++++++++
	CHOH/DONR, DDD K JJ SSSSSSS >>>>X PSSS >>>>X PSSS >>>>X PSSS >>>>X PSSS >>>X PSSS >>X PSSS >>>X PSSS >>X PSSS >>>X PSSS >>X PSSS >>>X PSSS >>>X PSSS >>>X PSSS PSSS		P + + + + + + + + + + + + + + + + + + +
+ 555 66 66 187 187 188	AAAAAA CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		P P P P P P P P P P P P P P P P P P P
++++++++++++++++++++++++++++++++++++++	NASE [OXIDS AAAAAA CC AAAAAAA	SS SSS S SSS S SSS S SS SS S SSS S SSS S	PP +++++++ \$ \$\$\$\$\$ >>> >>> >>> BES \$ GGGHH #1379**5*6 GVDINKDKFA
	IVER ALCOHOL DEHYDROGENASE COXIDOREDUCTASE: CHOH/DONR, AAAAAA CCCCC DDD		P P P P P P P P P P P P P P P P P P P
DDDDD 000 NNNNN +++++++ SS 13< TTEEEE 85+71031 KTF VKVVSWY	## APDH APO-LIVER ALCOHOL DEHYDROGENASE [OXIDOREDUCTASE: ### AAAA AAAB #### CCCC ############################	++++-++- SSSSSS 444< >333< TTTSSTT *3288*229	SSSSSSSS SSSS >>>XXXXXX X< >>>>XXXXXX X SHHHHHHHH HHHHS 314000020 0*6414 GCVGLSVIMG CKAACA -H=ALPHA-HELIX
SHEET BRIDGEZ. BRIDGEZ. CHIRALITY CHIRALITY 5-TURN 5-TURN SUMMARY. EXPOSURE.	4ADH APO- SHEET BRIDGEZ BRIDGEZ CHIRALITY BEND 5-TURN 4-TURN SUMMARY EXPOSURE.	SHEET BRIDGE2 BRIDGE1 CHIRALITY BEND 5-TURN 4-TURN 3-TURN SUMMARY. EXPOSURE.	SHEET BRIDGE1. BRIDGE1. CHIRALITY BRUD 5-TURN 4-TURN SUMMARY. EXPOSURE. SEQUENCE.
381	1	191	201 SUM!

TABLE AIII (continued)

	S4LDH +	BB FF + S	G F S S S S S S S S S S S S S S S S S S
	AAAA dd dd bbb+++ EEEE 60001442*6	C BB C ++++ SSSSS > 33 >	F FFG G L KKM M++ SSS >333< E EEBTTB 10 103581 PC VLNDHGI S≍BEND.
	SQUALUS ACANTHIUS 4LDH	BB ++++++++	CCC HHH G G SS SS EEE 6419999 EA19999
DDD 333 	CPTR] [DGGF ISH MUSCLE: AAAAA ++++++++++ + +++++ SSSSSSS S S S S XXX4< >>3 >>4 >4 >4<td></td><td>FFF CCC KKK HHH +-++ SSS SS SSS SSS SSS</td>		FFF CCC KKK HHH +-++ SSS SS SSS SSS SSS
DD DDD 11 111 5555 6<6<6 3 3 3 3 3 4 HH75 EE EEE HH75 EE EEE 13 3 3 4 5 4 1 3 2 0 0 0 5 7 LLRSGESIRT ILTF		A SSS SSSSSSSSSS	KKK
D 1 -+++++++ SS SSSS >>>>XXX >>>>XXX FTTHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	**SE: CHOH/DONR, NAD/AC	AA	1
D DDD ++++++++ SSS >3444< >334 TTTEEEEE 5072034593 ALDPLITHUL	TASE: CHOH/DOI AAAA A C CCCC C CCC C CCCC C	AA e	######################################
++++++ +++++++++++++++++++++++++++++++	AAAA AAAAA Bbbb aaaaa 	AA e d d d d d d d d d d d d d d d d d d	++++++++++++++++++++++++++++++++++++++
	AAAAA bbbb aaaaa ++++ S SEEEE SH 61000015	######################################	######################################
rrrr +-+ + + 5SS S S S S S S S S S S S S S S S S	S S S S S S S S S S S S S S S S S S S		SSSS SS >>4
SSSSS >>44< >>344< >>34898 THHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	### CAN A CA	SSSS SSSS >>555< >>555< >>555< >>4++++	BRIDGEL. BRIDGEL. BRIDGEL. CHRALITY ++++++++++++++++++++++++++++++++++++
SHEET BRIDGE 2. BRIDGE 1. CHIRALITY BEND 4-TURN 3-TURN EXPOSURE. SEQUENCE.	SHEET SHEET BRIGGE1. BRIGGE1. CHIRALITY BEND 5-TURN 4-TURN 3-TURN SWIMARY EXPOSURE.	SHEET BRIDGE2. BRIDGE1. CHIRALITY BEND 4-TURN 3-TURN SUMAARY. EXPOSURE. SEQUENCE.	SHEET BRIDGEI. BRIDGEI. CHIRALITY BEND 5-TURN 4-TURN 3-TURN SUMMARY EXPOSURE.
301	45)	161	2.0.1 S.UMI

	+++++++++ SSSSSSSSS XXXXXXXXXX XXXXXXXXX	EEE mm 111	G G G ++++++-+ SSS SSS >5444 >3444 >334 >334 >334 -334 -334 -334 -334 -334 -334 -344 -
	######################################	++++++++ SSSSSSSS >5 >>XXXXXX >>XXXXXXX 133 14HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	G SSS SSS SSS SSS SSS SSS SSS SSS SSS S
	IAN ERYTHROCY SS >>> HH **68*5*7*1" PSCEGFFWRR	EEEE 111 KKKK +++ +++ +++ SS SSEEEE 32000211 GRSVIVGA	II II S S.SS S.
	R. NADPH/DONR, FLAVOENZYME] HUM.	S SSSS SS S SSSS SS 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	FFF F EEEE KKK 000 1 SSSS SSSS SSSS SSSS SSSS SSSS SSSS
	R, NADPH/DONR, FLAVOENZYME] [HU] ++++++++++++++++++++++++++++++++++++	++++	SSSS EEE 5746**6**8
	E f f f f f f f f f f f f f f f f f f f	DD II + EE 183392* GMPSTPH	FFFFFF 0000 NNNNN -+ SSS SSS EEEEEE 4*33613143 LGGLEVSMYT IDGE6=3-
	AA AAA B	AAA ddd b t-t+++ SS SSEEE S88588158 TAPHILIATG	######################################
++++-+ SS S <<<< >>33< T+28+2*9 WDIQKDLKF	CTASE [OXIDOREDUCTASE: GSSG/ACCPB	C C C H H C C C C C C S S S S S S S S S	######################################
++++ +++++++++++++++++++++++++++++++++	1CTASE [OXIDOREDUCT f ++++++++ SSSSSSSSS	C G C C C C C C C C C C C C C C C C C C	+++++++ +++-++ \$S\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$ >>>XXXXXX< << >>>3
SSSS	ATHIONE REDIAN AND AND AND AND AND AND AND AND AND A	AAAA cccc +++ SSS SSS (<< 3555 (<<< 931914 LTKSHEFIR	EE FFF FF
SHEET BRIDGE 1. CHIRALITY CHIRALITY 5-TURN 1-TURN STORN STORN STORN	SHEET SHEET BRIDGE1 CHIRALITY ENRO SHENO 4-TURN 3-TURN EXPOSURE. SKOUSURE.	SHEET BRIDGE2 BRIDGE1 CHIRALITY BRND 5-TURN 4-TURN SUMMARY. EXPOSURE.	SHEET BRIDGE 1 CHRALITY CHRALITY SHEND 5-TURN 4-TURN SUMMARY. EXPOSURE. SEQUENCE.
361	16)	161	261 SUM)

(continued)
_
E AII
IABL

+++++++ ++++++++ \$\\$5\$\$\\$5\$\$\\$5\$\$\\$\\$5\$5\\$\\\$\\\$\\$\\$\\$\\$		++++++ SSSSSSS 5 >>>XXXX T HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	5-TURNSEBEND
++++++++++++++++++++++++++++++++++++++		NOMOUS THUMMI THUMMI	s 5-TURN
RM WHALE}		CARENTRANSPORT CHIRONOMOUS THUMMI THUMMI] CHIRONOMOUS THUMMI THUMMI] CHIRONOMOUS THUMMI THUMMI] CHIRONOMOUS THUMMI THUMMI] CHIRONOMOUS THUMMI CHIRONOMOUS THUMMI] CHIRONOMOUS THUMMI CHIRONOMOUS THUMMI CHIRONOMOUS THUMMI TSHUHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	.T=3-,4-, 0
RM WHALE}		#MI THUMMI} ++++++++ \$\$\$\$\$\$\$\$\$ XXXXXXXX<< PHHHHHHHHHHHHHHHHHHHHHHHHHHH	=5-HELIX
	* 8 * 8 * V	ORT] (CHIRONOMOUS THUMMI THUMMI]. +++++++++++++++++++++++++++++++++++	HELIXI
N - METMYOGLOBIN) (SPE ++++++++++++++++++++++++++++++++++++	+++++++ SSSSSSSS >5555 XXXXXXXXXXXX HHHHHHHH 601621**67 DIAAKYELG	:YGEN TRANSPORT] [CHIRO	.DGEG=3
N - METMYOGLOBIN) (SPE ++++++++++++++++++++++++++++++++++++	++++++++ SSSSSSSSS XXXXXXXXXXXX HHHHHHHHHH		++++ SSSS XX<<<<>>3333496955 MIPSKM
	++++++++ +++++++++++++++++++++++++++++	DBIN (ERYTHROCRUORIN, DEOXY) [OXYGEN TRANSPORT] [CHIRONOMOUS THUMMI THUMMI] ++++++++++++++++++++++++++++++++	SSSSSSSS SS SSSSSSSSSS SSSSSSSSS SSSSSS
DBIN [OXYGEN STORAGE] ++++++ ++++++++++++++++++++++++++++	++++++++ SSSSSSS S XXXXX<<<< 4 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	++++++++ SSSSSSS S X <x<<4<>> >3XXX<10 >> > ></x<<4<>	**************************************
#YOGLOBIN [OXYGEN STORAGE] (FERRIC IRON - METMYOGLOBIN) [SPERM WHALE]	++++++++++++++++++++++++++++++++++++++	11.	SSSSSSSS XXXXXXXXX HHHHHHHHHHHH 65003710*8
SHEET SHEET BRIDGE2. BRIDGE2. CHIRALITY S-TURN 5-TURN 3-TURN SEMMARY EXPONENCE.	SHEET BRIDGE2 CHIRALITY BEND 5-TURN 4-TURN SUMMARY. EXPOSURE.	SHEET SHEET SHEET BRIDGE1. CHIRALITY CHIRALITY 6-TURN 3-TURN STUMARY EXUMARY	BRIDGE: BRIDGE: BRIDGE: BRIDGE: CHRALITY ++++++++++++++++++++++++++++++++++++
48)	101	1	101 SUMP

TABLE AIII (continued)

2M HB S \$55555 >>>>XX >>>>XX >>>XX T*186822*2 LRVDPVNFKL	SSSSSSS S >>>4<++++++ >3433 >>>4 SSHHHHT H *8653852* SNPGAVMGNP		-+++++++ S SSS >3>XX < STTGGHH 6**2393**5
	SSS SS SSSSSSS S XXXXX3<	+++++ SSSSS X<<<< HHHHS *023**5*	• 1 VE
######################################	**************************************	**************************************	+++++++++ \$\$\$\$\$\$\$\$\$ XXXXXXXXX HHHHHHHHH HHHHHHHHH
LUS}	SSSSSSSS >>> XXXXXXX HHHHHHHHH 74**001200	######################################	######################################
MET) [OXYGEN TRANSPORT] [HORSE: EQUUS CABALLUS] +++++++++++++++++++++++++++++++	######################################	######################################	OXYGEN TRANSPORT] [SEA LAMPREX: PETROMYZON MARINUS] ++++++++++++++++++++++++++++++++
EQUUS CABAL ++++++ \$S \$S\$ \$S << <t33< \$7<br="" t5="">*73**485\$* TYPPHF DLSH</t33<>	SSSS	+++++++ SSSSSSS XXSSSSS XXXCXCX HHHHHTTT HHHHTTT R#65287*2 *42283402 SELHCDKIHV DPENFRLIGN	SPORT] (SEA LAMPREY: F ++++++++++++++++++++++++++++++++++++
MET) [OXYGEN TRANSPORT] [HORSE: ++++++++++++++++++++++++++++++++++++	######################################	++++++++++++++++++++++++++++++++++++++	NSPORT] (SEA +++++++++ +++++++++ SSSSSSSSS >>x>xxxxx 7 134 134 136 136 136 136 136 136 136 136 136 136
TRANSPORT (HOILD HONOR HONOR	++++++++ ++++++++ ++++++++++++++++++++		(OXYGEN TRAN ++++++++ SSSSSSSS 5 133 132 132 132 132 63 99 34 SAWAPVXSDY
BIN (AQUO MET) [OXYG 	######################################	SSSSSSS SSSSSSS >>XXXXXXXX XXXX<<	LOBIN(MET)-CYANIDE V [
CLOBIN (AQUO SSSSSSS >>>>XXXXX HHHHHHHHHHHHHHHHHHHHH	++++++++ +++++++++++++++++++++++++++++	SSSSSSSS SSSSSSSSSSSSSSSSSSSSSSSSSSSSS	GLOBIN (MET) SSS \$342*6*379 PIVDTGSVAP
SHEBT SHEBT SHEBT BRIDGEI CHIRALITY BRIDG 5-TURN 4-TURN 3-TURN SEQUENCE.	SHEET BRIDGE 2 BRIDGE 1 CHIRALITY BEND 4-TURN 3-TURN SYMARY EXPOSURE.	SHEET BRIDGE 1 CHIRALITY BEND 5-TUN 4-TUN S-TUN S-TUN S-TUN S-TUN S-TUN	SHEET SHEET BRIDGE 2 BRIDGE 1 CHIRALITY BRIDGE 1 CHIRALITY 5-TURN 5-TURN SUMMARY. EXPOSURE.
58)	101	201	51)

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

		ICRN	S=BEND
			S-TURN
	ROOT NODULES: LUPINUS LUTEUS L}. +-++ ++++++++++++++++++++++++++++++		.T=3-,4-, OF
	T NODULES: LUPINUS ++-+++++++++++++++++++++++++++++++++	- F	5-нецх
	ROOT NODUL ++-++ SS SS NODUL SS HHH 3*23*44*8	+ + + + + + + + + + + + + + + + + + +	HELIXI=
+++++ SSSSSS X<< >>><< 1421*55* CILLRSAY	T] {YELLOW LUPIN S SSS SSS SSS SS C > 444X444 S SSS SSS SS T > 745 T > 75 T > 75	++++++++ SSSSSSSSSSSSSSSSSSSSSSSSSSSSS	DGEG=3-
	[OXYGEN TRANSPORT] {YELLOW LUPIN ++++++++++++++++++++++++++++++++++++	++++ +++++++++++++++++++++++++++++++++	. B=BETA-BRI
+ + 0	OGLOBIN (ACETATE, MET) [OXYGEN TRANSPORT] [YELLOW LUPI]+++++ +++++++++++++++++++++++++++++	======================================	A-STRAND
	CLOBIN (ACETATE, MET) [OX ++++++++++++++++++++++++++++++++++++	######################################	IXE=BET
++++++++++++++++++++++++++++++++++++++	LEGHEMOGLOBIN (ACETATE, MET) [OXYGEN TRANSPORT] {YELLOW LUPIN ROOT NODULES: LUPINUS LUTEUS L}	22. 11TY	H=ALPHA-HEL
SHEFT BRIDGE CHIRALITY CHIRALITY FFT 5-TURN 3-TURN SJATURN EXPOSURE.	1HBL LEGHES SHEET BRIDGES CHIRALITY HEND CHIRALITY A-TURN 4-TURN SUMMARY EXPOSURE.	BRIDGE 2. BRIDGE 2. BRIDGE 2. BEND 4. FURN 4-TURN 5-TURN 5-TURN 5-TURN 5-TURN 6-TURN	SUMMARYH=ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND
101	1	101	SUMM

TABLE AIII (continued)

54)	10VO OVOM SHEET.	OVOMUCOID THIRD DOMAIN [PROTEINASE INHIBITOR, KAZAL] (JAPANESE QUAIL: COTURNIX COTURNIX JAPONICA)	DOMAIN (PRO	OTEINASE INF	NHIBITOR, KAZ B B	AL] (JAPANE	SE QUAIL: C	OTURNIX COT	URNIX JAPON	ICA)	0,001
	BRIDGE 2	60 60		၁၃၁	ø	U					
	CHIRALITY	+++++	+++	+++		-++++++	3 ‡				
	BEND.	S	SS	SSS	888888888	SSSSSSS >5555<	SS				
	4-TURN.				>>>×××××						
	3-TURN.	×33	v	>33<			1				
	SUMMARY.	TT SEE TT	EEE TT SS BEETTS E ESSHHHHHHH HHHTTTS E EEES	EEETTS E	ESSHHHHHH	HHHTTTS E	EEES				
-	SEQUENCE.	LAAVSVDaSE	"309/1282" 230-03/19" "388833"-0 3310808822 12"3/4"1/1 //93"6 LAAVSVDASE YPKPAbpKDY RPVCGSDNKT YSNKBNFANA VVESNGTLTL NHFGKC	RPVCGSDNKT	YSNKbNFanA	VVESNGTLTL	NHF GKC				
55)	25SI STRE	STREPTOMYCES SUBTILISIN INHIBITOR [PROTEINASE INHIBITOR]	BTILISIN IN	HIBITOR (PRO	TEINASE IN	HIBITOR]					2881
	SHEET	AAAAA	c o	AAAACC	99			Ω		AAAAA	вр
	BRIDGEZ	8888	c	0000	Ē			•	ECCCCC		
	CHIRALITY	******	1 + + + + + + + + + + + + + + + + + + +	‡		-++++++++ +++-+-++-+	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	5 +	DANAAA	20000	5 1
	BEND.		S SSSS S			SSSSSSSSSS S SSSS	SSSS	SS	SS	SSS	SSSSSSSS
	5-TURN					>555	5<				
	4-TURN.		,		^;	>>>xxxxxx<< <<	\ \ \		•	,	>>>>×
	STURN.	9999	>5554 	0 66660	U	>334 > 334	A			33<	>335
	EXPOSURE.	*482425130	92	**85251524	*376491*42	47017717*3	6261962**5	6*46*46**	1110208223	*5**4*2*67	8618368728
-	SEQUENCE.	YAPSALVLTV	YAPSALVLTV GKGVSATTAA PERAVTLTAA PGPSGTHPAA GSAAADLAAV GGDLNALTRG EDVMbPMVYD PVLLTVDGVW	PERAVTLTAA	PGPSGTHPAA	GSAAADLAAV	GGDLNALTRG	EDVMb PMVYD	PVLLTVDGVW	OCKRVSYERV FSNEDEMNAH	FSNEDEMNAH
	SHEET										
	BRIDGEZ.										
	CHIRALITY	***									
	BEND	SSSS									
	5-TURN										
	4-TURN.	*									
	SUMMARY.	TSSSS									
		2*11055									
101	SEQUENCE.	GSSVF AF									
56)	3PTI TRYP	TRYPSIN INHIBITOR (PROTEINASE INHIBITOR) (COM PANCREAS: BOS TAURUS)	OR [PROTEIN	ASE INHIBIT	OR] (COW PA)	CREAS: BOS	TAURUS)	•			3PTI
	SHEET		AAA	AAA AAAA AA	AA AAAAA	«					
	BRIDGEL		AAA	AAA AAAA	AN ANAN	æ					
	CHIRALITY	+++1	++			-++++ +++-+	-++++				
	BEND	SSSSS	S		SSS	SS SSSSS	SSSSS SSSSS				
	5-TURN.			>5555<							
	A-TUKN.			>444<		***	>>>> xx<<<<				
	SUMMARY.	85555	SEBE	EEE EREETTTEE	EEEEE SSS	SS BSSHHH ИНННЯ	ннння				
•	EXPOSURE.	*5*40*5*59	*5*4Ø*5*59 7295*5*6*5 72145*5482	72145*5482	88272042*5	88272042#5 ###222#33#8 #7#5#24#	07*5024*				
1	1 SEQUENCE.	RPDFaleppy	RPDFaleppy TGPbkariir Yfynakaglo QTFVYGGbra krnnfksaed cartagga	YF YNA KAGLC	OTF VYGGBRA	KRNNF KSAED	CMRTaGGA				

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

ICTX ALPHA COBRATOXIN (COBRA: NAJA NAJA SIAMENSIS)AAAAA AAAAAA AAAAA	BBBBB AAAA BBBBB	*******	>> 44< >33X 33X 33 SSSSS TTS EREEF THE HH EREEF SS EREEFS STT 5*8255***9 55957*1,*2 0 3240977*1 9**4445600 345*83*7** 6554*3599 *307*49*8* 	MELITTIN (HEMOLYTIC POLYPEPTIDE) (HONEY BEE VENOM: APIS MELLIFERA)	**** **** \$282 S	XXX<< IH HHHH II S***** II KRKRQQ	ERABUTOXIN B) (SEA SNAKE: LATICAUDA SEMIFASCIATA) PREMBBB B BBBBBB B BBBBBBBBBBBBBBBBBBBBBB	B BBBBBB B	\$\$ \$\$ \$\$ \$\$ \$\$	>33<	AADR ADENYLATE KINASE [ATP-AMP PHOSPHOTRANSPERASE] (PORCINE MUSCLE: SUS SCROFA)	DDDDC GCC ++++++++++++++++++++++++++++++++	SSSSSSSSSS >55	4>>x>xxxx< <<	НИННИНЯ В ЕВЕВ 5737 ТУНИНИНИНИ ИТ ВЕВЕИН ИНИНИНИНЫ ИНИНИНИНИНЫ НИКЅ НИ НИНИНИНИНИНИНИНИНИТУТ В ВЕВБ SSSS 9**12**248 1832114242 *3651*926* *543362488 923*429974 3948*8883*2 5**6**554* 131812*921 77*4**2*88 112421**7*
ICTX ALPHA COBRATOXIN (COBRA: N SHEET	:	BEND SSSSS SSS	5-TURN 4-TURN 3-TURN 533< \$UMMARY \$SSSS TTS EXPOSURE. \$\$4.55.8**9 \$5957*1*28 1 \$EQUENCE. IRaFITPDIT \$KDDPNGHVA	IMLT MELITTIN (HEMOLYTIC POLYPE SREET	BRIDGE2. BRIDGE1. CHIRALITY +++++++ ++++++++++++++++++++++++++++	JUNN >>>>XXXXX XXXXX XXXXX XXXXX XXXXX XXXXX XXXX	INXB NEUROTOXIN B (PROBABLY = ERABUTOXIN B) SHEET AAA AAA BBBBBBB BBTAGET.	AAA	CHIRALITY +++++- SSS S S S S S S S S A-TIBN		2ADK ADENYLATE KINASE (ATP-AMP SHEET	BRIDGEL: A asa CHIRALITY +++++++++	BEND SSSSSS SSSSS SS-TURN	>>>>	

TABLE AIII (continued)

++ SS SS C< C< N 33 N 14H H H H H H H H H H H H H H H H H H H H	AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA	S SS S	N
A AAA ++++++++ +++++++++++++++++++++++	RHODANESE [THIOSULFATE-CYANIDE SULFURTRANSFERASE] (COM LIVER: BOS TAURUS)	++++-+ S SSSSS > 5555< >>>4 33x33 S BL BL	E +++-+-+- SS SSS S S
A AAA d ddd +++++++++++++++++++++++++++	\$5\$\$\$.\$	DDDD 11 99	9 +++++ SS SSSSSS >>>>X >>X <br ETTTHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
A AAA +++++++++++++++++++++++++++++++++	US}+-+	=	DD +++++ 9 55SS 55SS 5555 (<<<< HHT T EE 1545118600 YCGRPDVAI
++++++++ ++++++++ ++++++++++++++++++++	AAA BB	C F F F F F F F F F F F F F F F F F F F	DDDD hhh SSSSSS SS SSSSSSSSSSS X<<<< > > > > > XXXXX XXXXXXXXXXXXXXXXXXXXXX
	{COM LIVER -+++++++ \$S\$\$\$\$\$\$ <>>> 4x 4<<br HHHHTTS 7*1**729** REARKEYLER	P C C C C C C C C C C C C C C C C C C C	DDDD 11 ++++
++++-++ SSSSS <xsss <xx44 HHTTTT +HTTTT *94*4668*</xx44 </xsss 	PRANSFERASE] BB ee -++++- SSS > 444 > 33< RE TTT 010163499 LDASWYSFGT		++++++ >555SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
-+++++++ SSSSSSSSSSSS >>>>XXXXXX HHHHHHHHHHHHHHH	NIDE SULFURTRA AAA BB CC CC CC CC CC CC	AAAA ddd aa ++++ S EEEET *308014 RTVSVLN	L L L L L L L L L L L L L L L L L L L
AAAAA dddd bbbb ++++++++ 5\$	AA +++++	+++++-+- SSSSSSS XXXX< XXXX	F +++++ SSSSS SSS 33< 71717B TTC 2297014*90 PFMDFLTENG LIXE=BET
++++++++ \$\$\$\$\$\$\$\$\$ >>>xx<<<<< +HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	ANESE [THIOS AA ++++ EE ****75*151 UHQVLYRALV	++++++ SSSS SSSS >>>> 31TS SSHH 2*5#161##B	K 11 +++-+
SHEET BRIDGE1 CHIRALITY BEND 5-TURN 3-TURN SUMMARY. EXPOSURE.	SHET SHETGEL. BRIDGEL. CHIRALIT. CHIRALIT. 5-TURN 4-TURN SUMARY. EXPOSURE.	SHEET BRIDGE2 CHINGE1 CHIALITY BEND 5-TURN 3-TURN SUMMARY. EXPOSURE.	SHEET BRIDGE2. BRIDGE1. CHIRALITY SHEW 5-TURN 4-TURN SUMMARY. EXPOSURE. SEQUENCE.
181	61)	181	241 SUMM

AAAAA AAAAA DDDDD Baaaa +S	EEEEEEET TTTEE S EEEEEE TTSEE S TTTS SEEE EHHHH HHHT S S EEEEEE S SS EEEE 624406705 874406*01 8784761*** 99**4282*8 56941882 8***16*168 858515658 **77693*5* 789274783* *3*2*34848 CCMWYCLDA WRGSPAINVA WHVFRKAADD TWEPFASGKT SESCELHGLT TEEGFVEGIY KVEIDTKSYW KALGISPFHE HAEVVFTAND SGPRRYTIAA AAA AAAA AAAA AAA AAAA AAA AAAA AA	.S=BEND
BBBBBBB GGGGGGG	EEEEEE S SS 789274783* *3*2 HAEVVFTAND SGFFF	3 5-TURN
SSS SSSS S S X X X X X X X X X X X X X	HHHT S S **77693*5 * KALGISPFHE	T=3-,4-, 0
SAA BBB BBB BBBBBBB BBBBBBB BBBBBBB	TTTS SEE EEEE HHHH HHHT ***16*160 3658515530 **777 EEOFVEGIY KVEIDTKSYW KALG	5-HELIX
BBB GGG GGG GGG A444<	TTTS SEEE 18**16*168 TEEQF VEGIY	HELIXI*
LASMA) AA AA BB BB SSS S	775EE S 566418822 585GELHGLT	DGEG=3~
TINGL TRANSPORT) (HUMAN PLA B BBBBB BBB BBBB EFFF EFF EFF EFF EFF EF	Seeeeeee	.B=BETA-BRI
INGL TRANSPOR B BBBBBB FFFF E EEEEEE SSS	EPEEEEET TTTEE S E EEEEEE TTS SEEEEEEEE 84888785 8*74466*01 8784761*** 99**4282*(LMYKULDA WGSFAINVA VHVFRKAADD TWEPFASCK A AAAA AAA DDDD DDD DDD DDD SS 533< ETTEEEE EEE 2778 SEEEEEEEE 8787875 AVVT	A-STRAND
ROXIN, RETINAA B CC E +++++- SSSS S 555<	TTTEE S E WAA446681 WGSPAINVA AAA DDDD EEE AVVT	IXE*BET
SHEET	EEEEEEET TTTE 6284889785 88"4 CPLMVKULDA URGS DDD DDD AA AAAAA AAA AAA AAAA AAA AAA AAAA AAA AA AAAAA AAA AAA AAAAA AAA AAA AAAAA AAAA AAAAAA	SUMMARYH=ALPHA-HELIXE=BETA-STRANDB=BETA-BRIDGEG=3-HELIXI=5-HELIXT=3-,4-, OR 5-TURNS=BEND
2PAB PREAL SHEET BRIDGEL BRIDGEL CHIRALITY BEND 5-TURN		MARY
62)	191	SUM

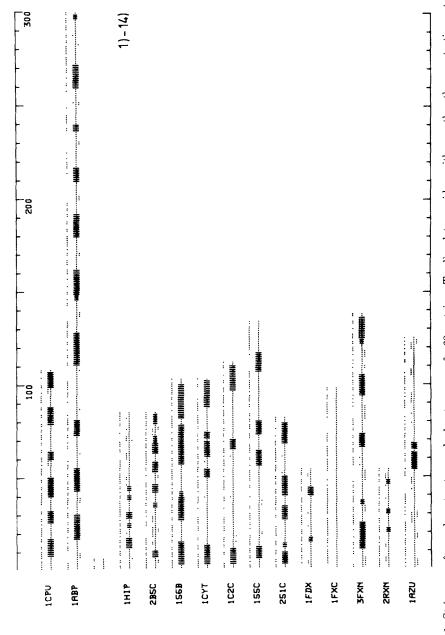
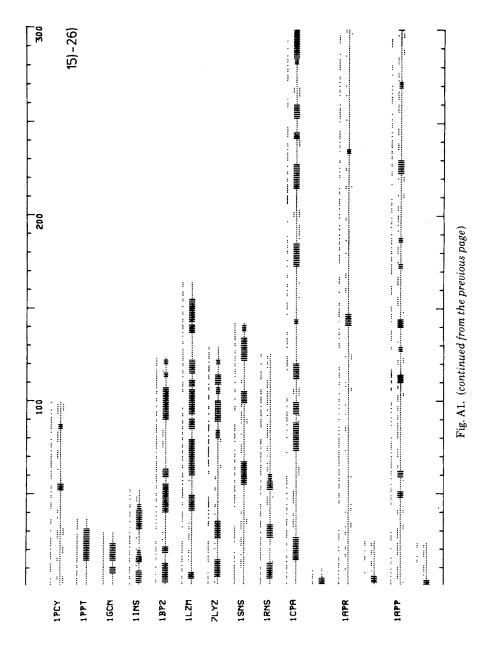
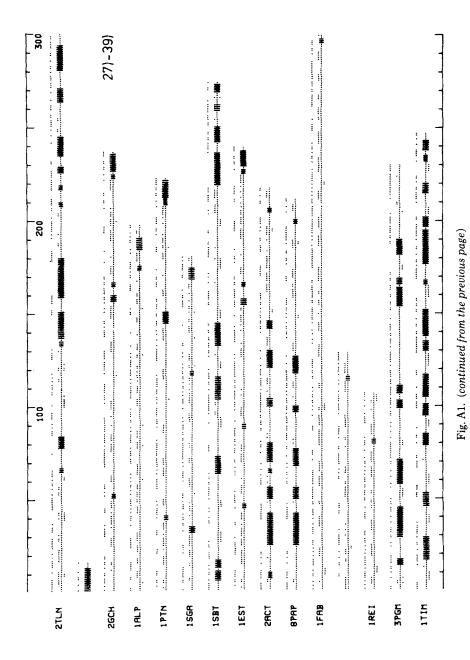
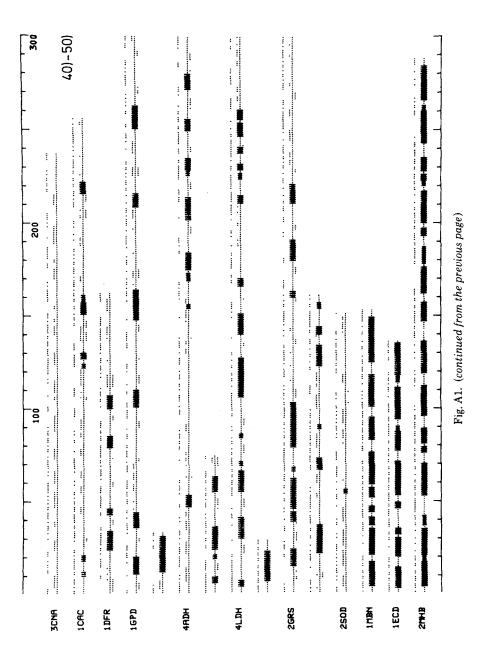
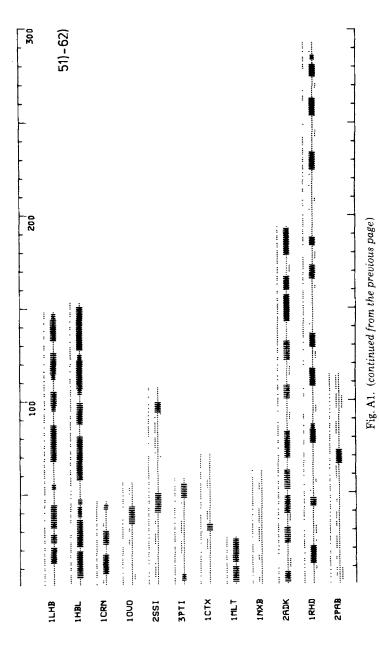


Fig. A1. Strip maps of secondary structure and solvent exposure for 62 proteins. Top line dots: residues with more than three contacting water molecules. Vertical bars: short, 3-helix, medium, 4-helix (α -helix); long, 5-helix. Dots above baseline: residue has antiparallel β -bridge partner(s). Dots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









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

Received December 22, 1982 Accepted May 12, 1983