Crepe-ribbon representation for protein structures: Comparison of phospholipases A_2

N. Pattabiraman* and Keith B. Ward

Code 6030, Laboratory for the Structure of Matter, Naval Research Laboratory, Washington DC, USA

We describe a method to generate a novel representation for protein structures called "crepe ribbons." In our representation, each piece of the ribbon was constructed using the coordinates of the backbone atoms of each individual residue. Using the internal geometries of each residue and a helix-generating algorithm, a local origin and the direction cosines of a local orthogonal system were obtained. The locus of the local origins represents the folding of the polypeptide. A color-coded origin–origin distance plot similar to that of C_α–C_α distance plot was generated. This plot may be used to visually compare and contrast two structures. We identified linear regions in the distribution of the local origins and assigned a secondary structure description. Parameters describing the interrelation between various secondary structure segments were calculated.

We have illustrated our crepe-ribbon representation by comparing two phospholipase A_2 structures in the Brookhaven National Laboratory Protein Data Bank.

Keywords: ribbon representation, protein structure, description of secondary and tertiary structures

INTRODUCTION

The tertiary structure of a protein crucially influences its biological function. The prediction of tertiary foldings from the primary sequence is a challenging problem in protein structure research. To predict the tertiary structure, it is important to study and understand tertiary folding patterns of proteins whose structures have been solved to atomic resolution by X-ray diffraction analyses.[1] Two representations of protein structures commonly appear in the literature. They are the ribbon representation of Richardson[2] and the representation of cylinders and ribbons by Lesk and Hartman.[3] These representations present qualitatively the interrelations among various secondary structures in proteins. They quantify neither the local nor the global foldings in proteins. Also, these representations require that every residue be assigned to a certain secondary structural region. This secondary structure information may be taken from the reported crystal structure data deposited in the Brookhaven Protein Data Bank (PDB).[1] Computer programs[4,5] have been written based on Richardson's ribbon representation. Various attempts have been made to obtain an approximate axis of orientation for various secondary structures by least squares fitting of short probe (ideal) helices to all atoms in the backbone or C_α atoms of successive residues.[6–8] A *P*-curve for protein structure was developed using successive peptide planes.[9] The latter method is similar to a method, based on the base-pair planes, for describing double helical nucleic acid structures. The *P*-curve was obtained by smoothing each local coordinate system with respect to the collection of all local coordinate systems. No scheme to quantify the relationships between various secondary structures was presented with this method. Recently we made a preliminary report of a novel representation for protein structures called a "crepe ribbon."[10] In our representation, the crepe ribbon was generated using the coordinates of the backbone atoms in each residue rather than being based upon atoms in the peptide plane. We did not preassign secondary structure information to each residue. In this paper we describe the method in detail and compare the secondary and tertiary folds of two phospholipase A_2 (PLA2) structures. The two phospholipases are from bovine pancreas, represented by the PDB code 1BP2,[1,11] and from *Crotalus atrox*, represented by the PDB code 1PP2.[1,12] The 1BP2 structure is monomeric in the crystal and contains calcium whereas the 1PP2 structure is a dimer and is calcium free in the crystal. Because 1PP2 is a dimer, the *R*-chain of 1PP2 was used in the comparison of PLA2 structures.

METHOD

We divided the entire protein into residues. If there are N residues in a protein, we considered only $N-2$ fragments because the torsion angle ϕ cannot be calculated for the first residue, nor can ψ be calculated for the last residue. It has been shown[13] that for a given set of internal parameters for a monomeric unit like an amino acid residue, it is pos-

Address reprint requests to Dr. Pattabiraman at Code 6030, Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375, USA.
*Permanent address: GEO-CENTERS, Inc., 10903 Indian Head Highway, Fort Washington, MD 20744, USA.
Received 10 October 1990; accepted 15 October 1990

sible to calculate not only the helical parameters (n, the number of units per turn and h, the unit height) but also the location of the origin and the direction cosines of the orthogonal axes describing the local helical coordinate system. (A FORTRAN program was written[14] based on the helix-generating algorithm.) We obtained the local origin and the direction cosines of the local orthogonal system for each residue from the internal parameters (bond lengths, bond angles and torsion angles) of each residue. The local coordinate system was generated such that atom N of each residue lies on the x-axis and the z-axis points toward the progression of the polypeptide chain. We defined the perpendicular distance of the C_α atom in each residue to the local z-axis as the "local radius" of each residue. The locus of these local origins describes the folding of the protein structure. From the value of the local radius of each residue, it was possible to assign the residue to a particular class of secondary structure, e.g., an α-helix or β-strand.

The crepe ribbon representation was generated as follows: The endpoints of the perpendicular vector from the C_α atom to the local z-axis were rotated and translated in intervals of θ/m degrees and h/m Å, respectively, where m is the number of intervals. (In this paper we used $m = 10$ for the generation of the crepe ribbon.) These vectors form a helical staircase about the local z-axis. The nine quadrangles formed by two adjacent vectors and the vectors joining the corresponding endpoints were divided into triangles. Eighteen triangles were generated for each residue in the protein. Each triangle was assigned a specific color, depending upon some chemical or geometrical property of the residue. We used the ray-tracing program[15] RASTER3D to generate the ray-traced crepe ribbon, based upon the collection of triangles for the two phospholipase structures.

We calculated the distances between the origins and prepared a distance plot matrix. For each distance a square of unit dimension was assumed in the plot; for example, the bottom lefthand corner of the square representing the distance between the ith and jth residues is positioned at (i, j). The squares are color coded according to the range of distances. Because the locus of the origins describes the folding of the protein, the distance plot will reflect the folding of the protein in a way that is not possible with a distance plot generated using the coordinates of the C_α atoms.[16]

We calculated a pseudobond angle between origins that is analogous to the virtual bond angle often derived between C_α atoms. We identified the sets of residues whose pseudobond angles are $\geq 165°$. A group of such consecutive residues forms a segment in which the local origins lie almost on a straight line. We performed a least-squares fit for the set of origins on each segment. The length of each segment was defined to be the distance between the perpendicular projections of the local origins of the segment's starting and ending residues onto the fitted line.

We calculated parameters between the lines representing two segments in a manner similar to the procedure described by Richards and Kundrot.[8] We labeled the projected point from the local origin of the starting residue by N and the projected point from that of the ending residue by C. Because these lines occur in three-dimensional space, they do not intersect necessarily. It is possible to identify a common

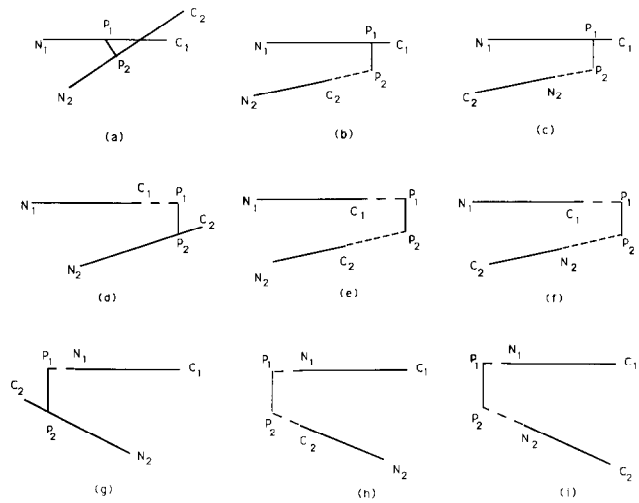

Figure 1. Nine types of intersection between two segments in space

Table 1. Approximate value of the torsion angle θ for parallel and antiparallel orientations of two segments

	Torsion angle θ (degrees)	
Type†	Parallel orientation	Antiparallel orientation
a	~0°	~180°
b	~0°	~180°
c	~180°	~0°
d	~0°	~180°
e	~0°	~180°
f	~180°	~0°
g	~180°	~0°
h	~180°	~0°
i	~0°	~180°

†Refer to Figure 1.

perpendicular to these lines. This perpendicular line will intersect the fitted lines either on the segments or on the N- or C-terminus extensions of the segments. There are nine possible ways for a perpendicular line to intersect the two fitted lines; they are represented schematically in Figure 1. The lines N_1C_1 and N_2C_2 represent two segments; P_1 and P_2 are the points of intersection of a line perpendicular to both lines. In Figures 1a–c, the point of intersection for line N_1C_1 lies on the segment, whereas the point of intersection, P_2, for line N_2C_2 lies on the segment, or the C-terminus or the N-terminus extensions respectively. Figures 1d–f are the same as the Figures 1a–c except that the intersecting point P_1 is on the C-terminus extension of line 1. Similarly, Figures 1g–i are the same as Figures 1d–f except that the point of intersection P_1 is on the N-terminus extension of line 1.

We calculated the closest-approach distance d between the two lines as the distance between the two points of intersection, and determined the torsion angle θ (N_1–P_1–

P_2–N_2) between the two lines. Table 1 lists the approximate range of torsion angles for parallel and antiparallel orientations of the two segments in the nine possible configurations of intersection. We also calculated the distances P_1N_1, P_1C_1, P_2N_2 and P_2C_2, to locate the positions of P_1 and P_2.

RESULTS AND DISCUSSIONS

Crepe-ribbon representation

Color Plates 1 and 2 show the crepe-ribbon representation for the 1BP2 and 1PP2 phospholipase A_2 structures, respectively. The two phospholipases were superimposed by matching the C_α atoms of the *N*-terminus helix (residues 1–12) and one of the two long helices (residues 40–54) in the two structures. The color coding is based on the local radius of each residue: red if the radius $<$ 1.0 Å; white if 1.0 Å $\leq$ radius $<$ 1.8 Å; yellow if 1.8 Å $\leq$ radius $<$ 2.9 Å; green if 2.9 Å $\leq$ radius $<$ 3.9 Å; and blue if the radius $\geq$ 3.9 Å. The darker colors are the shadows cast by other parts of the molecule. The name of this representation comes from Color Plates 1 and 2, where it appears that a long crepe ribbon has been twisted and bent to form the illustration. The various helical (yellow) and sheet (red) regions of the molecule are apparent. Unlike other representations,[2,3] the width of the ribbon in this representation is calculated from the coordinates of the backbone atoms.

Color Plate 1a shows four α-helices (yellow helical staircases): two long helices at the back of the molecule, a third, viewed down its axis, at the bottom lefthand corner of the color plate and a fourth (*N*-terminal helix) inclined at an angle to one of the two long helices at the top righthand corner of the color plate. The two β-strands (red ribbon) forming a β-sheet are at the bottom righthand corner of Color Plate 1a. In Color Plate 1b the gross folding of the structure of 1P2P is identical to that of 1BP2 except for some local differences. For example, the *N*-terminus helix is somewhat distorted from the ideal α-helix conformation, and one of the β-strands in the β-sheet is also distorted when compared to the same strand in 1BP2. We will describe these distortions in detail later in this paper. Residues 115–117 and 128–132 at the *C*-terminus form structures that are different from α-helix and β-strand.

In Color Plates 1b and 2b, an alternate color coding scheme for the crepe ribbon is based on the residue type: Red is for polar residues ASP, GLU and TYR; white, neutral residues ASN, GLN, THR, SER and CYS; blue, hydrophilic residues LYS, HIS and ARG; and green, nonpolar residues TRP, GLY, ALA, PHE, PRO, MET, VAL, LEU and ILE. It is easy to compare visually the distribution of amino acids in helices, sheets and loops. For example, the calcium binding loop has more nonpolar residues in the 1PP2 structure than in the 1BP2 structure. In Color Plate 1b, the nonpolar residues are concentrated in three regions of the 1BP2 molecule—near the *N*-terminus, in the calcium binding loop and at the end of the two long helices at the bottom of the Color Plate 1b. In Color Plate 2b, the nonpolar residues are spread over the entire molecule. Thus color coding schemes for the crepe ribbon based upon either the geometrical or chemical properties of residues provide a useful basis on which to compare similar molecules.

Assignment of secondary structures

Color Plate 3 shows the tracing of C_α atoms (yellow lines) and the tracing of origins (magenta lines) for the phospholipase A_2 structures 1BP2 (left) and 1PP2 (right). In Color Plate 3, it is clear that the magenta lines visually show the protein folding. Unlike a β-turn, if we follow the vectors joining the C_α atoms in four consecutive residues of an α-helix, the direction of the vector joining the third and fourth C_α atoms is reversed from the direction of the vector joining the first and the second C_α atoms. However, the overall direction of the progression of the polypeptide chain is along the direction of the helix axis. For a β-strand, the direction of the vector joining C_α atoms between successive residues is approximately along the direction of the progression of the polypeptide chain. The origin–origin vectors show the direction of the progression of the polypeptide chain. Because the origins were calculated from the backbone atoms, when the residues have almost identical conformations (as in α-helix and in β-strand), the origins lie on a straight line. In Color Plate 3 one can see that the locus of origin points approximates a straight line. For example, in the α-helix at the *N*-terminus of the PLA2 structures, the origins lie on a straight line. If the residues have similar conformations other than α-helix and β-strand, then the origins for these residues also lie on a line.

Table 2 lists the residues that form segments, the average origin–origin distances (calculated from the segment length divided by the number of residues minus one) and the average value of the local radius within the segment. We have also listed the assignment of secondary structures for each segment and the assignment reported in the corresponding PDB data file. We assigned α-helix or β-strand structures to the segments if the absolute value of the variation in the average origin–origin distance and the average local radius from the corresponding values for the ideal structure are within 0.15 Å. For an ideal α-helix ($\phi = -57°$ and $\psi = -47°$), the origin–origin distance is 1.53 Å and the local radius is 2.27 Å. We used the average values of the parallel and antiparallel β-sheet structures for our β-strand parameters.

In the reported structure for 1BP2, there are four α-helices and two β-strands. The first residue of the *N*-terminus helix (residues 1–13) was not assigned a secondary structure because there is no torsion angle ϕ for the first residue; hence it was not included in our calculations. We found the last residue of this helix to be distorted from an ideal α-helix conformation. Likewise our system did not assign an α-helix structure to residues 17–22, which was reported to be an α-helix by the authors in their PDB entry.[1,11] In comparing the torsion angles ϕ and ψ for these residues to that of an ideal α-helix conformation, we found that these residues deviate significantly from the ideal α-helix. For the third long helix (residues 39–58) the helix is distorted at residues 54 and 55. Therefore the third helix is made up of two separate α-helices (one from residues 40–54 and the other from 55–57). Even though the structure of residues 39–58 was reported to be an α-helix from the crystal structure, residues 54–57 were also reported to be a Type III turn, which is a distortion of an α-helix. In our assignment neither the *N*- nor the C-terminal residues are in α-helical conformation. For the fourth α-helix (residues 90–108) only

Table 2. Assignments of various secondary structures for the bovine phospholipase A_2 (1BP2) structure

Residues	Average origin–origin distance (Å)	Average local radius (Å)	Assignment from this work[a]	Reported with the structure[b]
2–12	1.52	2.30	α	α (1–13)
33–35	3.24	1.00	β	
40–54	1.52	2.27	α	α (38–57)
55–57	1.71	2.06		Type III (54–57)
59–62	1.55	2.31	α	α (58–66)
66–68	3.41	1.39		
74–78	3.41	0.82	β	β (74–78)
81–84	3.24	0.99	β	β (81–85)
90–100	1.51	2.32	α	α (89–108)
101–104	1.57	2.29	α	α (89–108)
109–112	3.14	1.10		
113–115	1.79	1.92		Type I (113–116)
120–122	1.85	1.87		Type I (120–123)

[a]For an ideal α-helix ($\phi = -57°$, $\psi = -47°$) the origin–origin distance is 1.53 Å and the local radius is 2.27 Å. For the parallel (phi = −130°, psi = 120°) and the antiparallel ($\phi = -120°$, $\psi = 140°$) β-strand, the average origin–origin distance is 0.87 Å and the local radius is 3.35 Å
[b]The values in brackets are the starting and ending residue numbers

Table 3. Assignments of various secondary structures for the *C. atrox* phospholipase A_2 (1PP2) structure

Residues	Average origin–origin distance (Å)	Average local radius (Å)	Assignment from this work[a]	Reported with the structure[b]
4–10	1.48	2.34	α	α (1–12)
16–18	3.43	1.40		
20–22	1.05	2.57		
22–24	2.11	2.08		
36–38	3.50	1.64		
42–45	1.60	2.16	α	α (41–54)
47–51	1.63	2.32	α	α (41–54)
52–54	1.74	2.09		
73–77	3.35	0.90	β	
78–80	1.91	1.98		
81–85	3.31	1.00	β	
91–93	1.63	2.30	α	α (90–109)
95–100	1.62	2.27	α	α (90–109)
100–103	1.45	2.36	α	α (90–109)
115–117	1.75	1.94		
128–132	3.00	1.42		

[a,b]See Table 2

residues 90–100 and 101–104 have α-helical conformation. One of the strands in the β-sheet region (residues 74–78) agrees with our assignment. But in the other strand (residues 81–85) our assignment shows that residue 85 does not have a β-strand conformation. It is interesting to note that residues 33–35 were assigned to be a β-strand. In Table 2 we have also included segments for which the origins lie near a straight line. They are residues 55–57, 66–68, 109–112, 113–115 and 120–122. It may be noted that residues 54–57, 112–115 and 120–122 were reported to be Type I β-turn in the PDB data file. As we analyze more protein structures we will be able to identify conformational features other than α-helix and β-strand.

Table 3 lists the assignments of various secondary structures for the *C. atrox* phospholipase A2 (1PP2). The *N*-terminal α-helix (residues 1–12) assigned by our method is between residues 4 and 10. Residues 19–22 were reported to be an α-helix. However, from the average origin–origin distance (1.05 Å) and the average local radius (2.57 Å), which deviate significantly from the values of an ideal α-helix (1.54 Å for the origin–origin distance and 2.27 Å for the local radius), this region is very much distorted from an ideal α-helix. In one of the two long helices (residues 41–54) we observed three kinks at residues 45, 46 and 50. In the other long helix (residues 90–109) only residues 91–93, 95–100 and 100–103 had good α-helix parameters. Even

though a β-sheet similar to one observed in 1BP2 was not reported, our algorithm assigned β-strand structures to residues 74–78 and 81–85. Other segments (residues 16–18, 22–24, 36–38, 52–54, 78–80, 115–117 and 128–132) were not assigned any particular conformational features. Some of these segments may be β-turns. As we pointed out earlier we will be able to better understand these segments as we analyze more of the protein structures in PDB.

Origin–origin distance plot

Color Plate 4a shows the origin–origin distance plot for the phospholipase A_2, 1BP2. Because the distance between the *i*th and *j*th residues is the same as the distance between the *j*th and *i*th residues, we have shown only the upper diagonal of the distance plot. We have excluded also the zero-distance diagonal elements. The residues are numbered in an ascending order from left to right and from top to bottom. The squares representing each distance in the plot are color coded according to the range in which the value falls. The color coding is shown as a color bar at the bottom of the plot. For a residue in an α-helix, the color of the squares changes from cyan to light blue, dark blue, magenta, red and so forth along the horizontal. Thus for residues forming an α-helix, the sequences of similarly colored squares are parallel to the diagonal, as shown by the *N*-terminus helix (residues 2–12) in Color Plate 4a.

For β-strand residues, the squares start with blue instead of cyan and change progressively to magenta, red, yellow and green. For the parallel orientation of segments, the changing colors will form patterns parallel to the diagonal. If two segments are oriented in an antiparallel sense then one will see patterns perpendicular to the diagonal. The *N*-terminus helix is parallel to one of the long helices (40–54), and thus one can see a red pattern parallel to the diagonal surrounded by yellow squares from residues 2–12. Because the segments of the two long α-helices (residues 40–51 and residues 90–100) are antiparallel, one can see patterns of colors changing from blue to yellow along a direction perpendicular to the diagonal.

A clustering of colored squares near and perpendicular to the diagonal of the plot indicates that a turn occurs in that region of the polypeptide chain. For example, in Color Plate 4a, there is a clustering of blue squares perpendicular to the diagonal near residue 15, around which a turn is observed in the structure. As one scans along a row corresponding to a particular residue, or group of residues forming a secondary structure, it is possible to follow the folding of the polypeptide chain from changes in the color of the squares. For the *N*-terminal helix, it is clear from the changes in color that it is closer to helix 1 (90–100) than to helix 2 (40–54) because the colors of the squares representing the distances between the *N*-terminus helix and helices 1 and 2 are blue and red, respectively.

Color Plate 4b shows the origin–origin distance plot for the phospholipase A_2, 1PP2. The color coding is the same as described in Color Plate 4a. It should be noted that the sequence numbering of the primary sequence for 1PP2 is based on the sequence numbering of 1BP2. Three regions have been deleted in the primary sequence of 1PP2 (residue 15, residues 55–63 and residue 87) and residues 123–132 have been added. For this plot, however, we have renumbered the sequence starting from the beginning. Because the topology of the 1PP2 structure is identical to that of 1BP2, the gross features of the origin–origin distance plots are the same. However, there are some significant differences near the β-sheet region. One of the strands in the 1BP2 plot (residues 81–84) is closer to the *N*-terminus helix, as illustrated by the blue color of the squares representing

Table 4. Parameters describing the interrelationships between various segments in 1BP2 structure

Segment *I*				Segment *J*						
Residues	Length (Å)	P_1N_1 (Å)	P_1C_1 (Å)	Residues	Length (Å)	P_2N_2 (Å)	P_2C_2 (Å)	d (Å) (For Segments *I* and *J*)	θ (°) (For Segments *I* and *J*)	Type†
2–12	15.2	3.1	12.1	40–54	21.2	15.6	5.6	15.3	145.4	*a*
2–12	15.2	11.5	3.7	74–78	13.6	9.7	3.9	9.2	17.2	*a*
2–12	15.2	27.4	12.2	81–84	9.7	11.9	21.6	2.2	32.2	*h*
2–12	15.2	11.3	3.9	90–100	15.1	19.4	4.3	9.3	318.1	*c*
2–12	15.2	14.1	1.1	101–104	4.7	5.9	1.2	8.6	320.8	*c*
40–54	21.2	43.9	22.7	74–78	13.6	26.2	39.8	9.4	326.9	*h*
40–54	21.2	96.8	75.6	81–84	9.7	91.1	81.4	7.4	11.9	*i*
40–54	21.2	42.9	64.1	90–100	15.1	72.7	57.6	1.2	8.9	*f*
40–54	21.2	89.5	110.7	101–104	4.7	102.5	97.8	5.0	4.7	*f*
74–78	13.6	9.9	3.7	81–84	9.7	5.3	4.4	4.4	155.3	*a*
74–78	13.6	6.7	20.3	90–100	15.1	1.3	13.8	6.7	138.1	*d*
74–78	13.6	6.4	20.0	101–104	4.7	14.6	19.3	7.6	322.5	*e*
81–84	9.7	34.6	24.9	90–100	15.1	13.0	28.1	1.7	20.2	*h*
81–84	9.7	41.9	32.2	101–104	4.7	36.3	40.9	0.9	15.1	*h*
90–100	15.1	8.5	6.6	101–104	4.7	7.9	12.6	0.6	185.1	*h*

†Refer to Table 1 and Figure 1

the distance. But the same strand in the 1PP2 plot (residues 72–76) is farther from the *N*-terminus helix, as illustrated by the redder color of squares representing the distance.

Description of tertiary folds

Table 4 lists the parameters describing the interrelationships between various segments in the 1BP2 structure. We have selected only the segments that were assigned to either an α-helix or β-strand in Table 2. There are six such segments for the 1BP2 molecule. Table 4 shows the parameters calculated for the 15 possible unique relationships between these segments, taking them in segment *I*–segment *J* pairs. We have listed the residues in each segment, its length and the *PN* and *PC* distances, as shown in Figure 1. From the *PN* and *PC* distances for each segment it is possible to determine whether the perpendicular line is intersecting on the segment, or on the *N*- or *C*-terminal extension. If the sum of the *PN* and *PC* distances is equal to the length of the segment then the intersecting point of the perpendicular line is on the segment. For example, for the pairing of the segment described by residues 2–12 with the segment described by residues 74–78, both intersecting points lie on the segments because the sum of *PC*s and *PN*s is equal to the length of the segment. If $PN < PC$, then the intersecting point is on the *N*-terminal extension of the segment; if $PN > PC$, then the intersecting point is on the *C*-terminal extension of the segment. For example, for the pair, (residues 2–12 with residues 40–54), P_1 is nearer to the *N*-terminus because $P_1N_1 < P_1C_1$. However, for the second segment P_2 is nearer to the *C*-terminus, because $P_2N_2 > P_2C_2$. Using the parameters listed in Table 4 one may generate the coordinates of the *N*- and *C*-terminuses of the two segments in space. Using databases of local origins, the endpoints of the segments and the PDB coordinates of the PDB structures, it is possible to generate the coordinates of the backbone and side-chain atoms of the residues involved in the segments. From Table 1, the type of orientation is *a* for the first row in Table 4, because the value of θ is 145.4°; segments *I* and *J* are oriented in an antiparallel way. For the entire structure, *d* ranges from 0.6 to 15.3 Å and θ clusters around 0° and 180°.

Table 5 lists the parameters describing the interrelationships between various segments for the 1PP2 structure. In this case there are eight segments. The distance *d* ranges from 0.1 to 15.0 Å. As observed in the case of 1BP2, θ falls into two groups, one clustered around 0° and the other around 180°. In both cases the type *h* orientation is much more common than the other eight types of orientations.

Comparison of two phospholipase structures

Secondary structure comparison For the 1BP2 structure we assigned four α-helices and two β-strands. We assigned three α-helices and two β-strands for the 1PP2 structure. We assigned only three helices for 1PP2 because the residues involved in one of the α-helices are deleted in 1PP2 structure. In 1BP2, the *N*-terminus α-helix starts at residue 2 and ends at residue 12, but this helix in 1PP2 occurs between residues 4 and 10. When we compared the sequences from 2 to 12 between these two structures we discovered a replacement at residue 3 of a bulky aromatic amino acid, TRP, by a small hydrophobic amino acid, VAL. Because of this substitution residues 2 and 3 are distorted from an ideal α-helix. For these two phospholipase structures, residues 4, 5 and 9 are conserved, and the replacements for residues 6–8 and 10 are such that the side chains are comparable in size.

In 1BP2, residues 40–54 form the second α-helix. This long helix is split into two short helices in the 1PP2 structure, one running between residues 42 and 45 and the other between residues 47 and 51. On comparing the sequences in this region, residues 42–44, 48, 49 and 51 are conserved between these two structures. The distortion at residue 46 is due to the replacement of GLN by an aromatic residue, PHE, in 1PP2. Thus, the favorable interactions between ARG 43 and GLN 46 are lost in 1PP2. The α-helix between residues 59 and 62 does not occur in 1PP2, due to a deletion in this region. The other long helical region in 1BP2 consists of two helices, one comprising residues 90–100 and the other residues 101–104. The same helical region in 1PP2 is split into three helices, one from residue 91 to 93, the second from 95 to 100 and the third from 100 to 103. The residue ALA 90 is replaced by PRO in 1PP2. The conserved residues between the two structures in this region are 91, 95, 96, 98, 99, 102 and 103. The distortion at residue 94 is due to the replacement of PHE in 1BP2 by GLN in 1PP2.

The two β-strands assigned by our method occur in the same regions in both structures.

Tertiary structure The *N*-terminal helix, (residues 2–12) in 1BP2, is the same distance from the long helix (residues 40–54) as the *N*-terminal helix (residues 4–10) is from the long helical region (residues 42–45 and residues 47–51) in 1PP2, but the relative orientation of these two regions differs in the two structures. However, the *N*-terminal helices are disposed in a similar way with respect to their β-stranded regions (81–84 in 1BP2 and 81–85 in 1PP2). Likewise, the *N*-terminal helices are located and oriented similarly with respect to the other helical regions (90–100 in 1BP2 and 95–100 in 1PP2). The long helical region (42–45 and 47–51) in 1PP2 has the same type of orientation with respect to the β-strands as the orientation of the long helix (40–54) in 1BP2 with respect to the β-strands. But the values of *d* and θ are slightly different between these structures. The relative orientation between the two β-strands are the same for these two structures. For the rest of the segments in Tables 4 and 5, there are some differences between the two structures.

From the assignment of secondary structures for segments (Tables 2 and 3) and the parameters calculated between them (Tables 3 and 4), it is clear that residues 4–10 and residues 91–93, 95–100 and 101–103 are very similar in both structures in their secondary structure and tertiary folding. It is of interest to note that the sequences in these regions are not exact homologues. Once these regions of similar secondary and tertiary structure are identified, they can be used to superimpose the two structures; that is, one can use the backbone atoms of these regions only instead of using the local folding information, such as *N*-terminal helices or other secondary structure segments.[17] As discussed earlier, even though the secondary structures are much the same, there are some differences in their mutual orientation within the tertiary folding. For example, the interrelationship be-

Table 5. Parameters describing the interrelationships between various segments in the 1PP2 structure

Segment *I*				Segment *J*						
Residues*	Length (Å)	P_1N_1 (Å)	P_1C_1 (Å)	Residues*	Length (Å)	P_2N_2 (Å)	P_2C_2 (Å)	d (Å) (For Segments *I* and *J*)	θ (°) (For Segments *I* and *J*)	Type†
4–10	8.9	1.7	10.6	42–45	4.8	14.4	9.6	14.3	322.0	*f*
4–10	8.9	1.8	10.7	47–51	6.5	6.2	0.3	15.0	327.7	*d*
4–10	8.9	31.3	40.2	73–77	13.4	28.9	42.3	3.2	16.6	*e*
4–10	8.9	55.4	46.5	81–85	13.3	42.1	55.4	1.1	14.2	*h*
4–10	8.9	8.4	0.5	91–93	3.3	18.4	15.1	9.6	324.8	*c*
4–10	8.9	7.8	1.1	95–100	8.1	11.9	3.8	9.3	323.5	*c*
4–10	8.9	11.2	2.3	100–103	4.3	7.6	3.3	8.6	326.7	*i*
42–45	4.8	5.5	0.7	47–51	6.5	2.3	8.8	0.5	185.9	*h*
42–45	4.8	26.4	21.6	73–77	13.4	8.7	22.1	12.8	305.9	*h*
42–45	4.8	49.1	44.3	81–85	13.3	45.9	32.6	6.6	334.4	*i*
42–45	4.8	9.4	4.6	91–93	3.3	16.2	12.9	8.6	174.8	*i*
42–45	4.8	20.3	25.1	95–100	8.1	39.7	31.6	8.1	356.1	*f*
42–45	4.8	97.2	92.4	100–103	4.3	86.5	90.8	1.0	354.4	*h*
47–51	6.5	19.8	13.3	73–77	13.4	11.1	24.5	13.8	311.7	*h*
47–51	6.5	53.1	46.6	81–85	13.3	58.8	45.5	7.5	340.3	*i*
47–51	6.5	68.6	75.1	91–93	3.3	86.0	82.7	4.4	353.4	*f*
47–51	6.5	72.5	79.0	95–100	8.1	83.7	75.6	2.0	353.3	*f*
47–51	6.5	22.5	16.0	100–103	4.3	20.0	24.3	9.2	1.5	*h*
73–77	13.4	10.8	2.6	81–85	13.3	7.7	5.6	4.1	151.1	*a*
73–77	13.4	0.7	14.1	91–93	3.3	3.2	0.1	7.6	128.3	*d*
73–77	13.4	0.8	14.2	95–100	8.1	3.0	11.1	7.6	307.1	*e*
73–77	13.4	0.6	14.0	100–103	4.3	10.1	14.4	8.7	310.1	*e*
81–85	13.3	29.9	16.6	91–93	3.3	9.2	12.5	1.3	335.8	*h*
81–85	13.3	29.5	16.2	95–100	8.1	15.1	23.2	1.5	334.8	*h*
81–85	13.3	33.0	19.7	100–103	4.3	26.2	30.5	3.8	339.7	*h*
91–93	3.3	3.5	6.8	95–100	8.1	9.7	17.8	0.1	358.5	*e*
91–93	3.3	13.3	10.0	100–103	4.3	1.0	5.3	0.3	172.6	*h*
95–100	8.1	7.4	0.7	100–103	4.3	0.7	5.0	0.1	187.3	*b*

†Refer to Table 1 and Figure 1
*Following the assigned residue numbers as reported in the crystal structure

tween the two β-strands is the same in both structures, but their relationships with one of the long helical segments are different.

CONCLUSIONS

In this paper we have described a method for generating a novel representation of protein structures, called the "crepe ribbon." Using this approach we were able to quantify both local and global folding. In our representation we did not preassign secondary structures to primary structures. Instead, it was possible to assign objectively a secondary structure to segments based upon our calculations. It was also possible to prepare visual illustrations of the folding of protein structures by joining local origins defined by this method. We used an origin–origin distance plot to provide a rapid visual comparison of two structures. We also calculated parameters that describe the spatial relationships between secondary structural elements. These parameters can be used to compare the orientation of similar segments in other proteins. We are currently using this method to process all of the protein structures in PDB to better understand and classify various types of intermolecular folding. A database containing the parameters describing the tertiary folds for all proteins will be created so that researchers will be able to search for segments having similar local and global folding patterns.

ACKNOWLEDGMENT

This work was supported in part by the United States Army Medical Research and Development Command.

REFERENCES

1 Berstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyers, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. *J. Mol. Biol.* 1977, **112**, 535
2 Richardson, J.S. *Adv. Protein Chem.* 1982, **34**, 167; Richardson, J.S. *Methods Enzymol.* 1985, **115**, 341

3 Lesk, A.M. and Hardman, K.D. *Science* 1982, **216**, 539
4 Carson, M. and Bugg, C.E. *J. Mol. Graphics* 1986, **4**, 121; Carson, M. *J. Mol. Graphics* 1987, **5**, 103
5 Priestle, J.P. *J. Appl. Cryst.* 1988, **21**, 572
6 Louie, A.H. and Somarjai, R.L. *J. Mol. Biol.* 1983, **168**, 143
7 Barlow, D.J. and Thornton, J.M. *J. Mol. Biol.* 1988, **201**, 601
8 Richards, F.M. and Kundrot, C.E. *Proteins: Struct. Function Genetics* 1988, **3**, 71
9 Sklenar, H., Etchebest, C. and Lavery, R. *Proteins: Struct. Function Genetics* 1989, **6**, 40
10 Pattabiraman, N. and Ward, K.B. in *Proceedings of the Expanding Frontiers in Polypeptide and Protein Structural Research* (V. Renugopalakrishnan, P.R. Carey, S.G. Huang, A. Storer and I.C.P. Smith, Eds.) in press
11 Dijkstra, B.W., Kalk, K.H., Hol, W.G.J. and Drenth, J. *J. Mol. Biol.* 1981, **147**, 97
12 Brunie, S., Bolin, J., Gewirth, D. and Sigler, P.B. *J. Biol. Chem.* 1985, **260**, 9742
13 Sugeta, H. and Miyazawa, T. *Biopolymers* 1967, **5**, 673
14 Pattabiraman, N. PhD thesis, 1979, Indian Institute of Science, Bangalore, India
15 Bacon, D. and Anderson, W.F. *J. Mol. Graph.* 1988, **6**, 219
16 Phillips, D.C. *Biochem. Soc. Symp.* 1970, **313**, 11
17 Renetseder, R., Brunie, S., Dijkstra, B.W., Drenth, J. and Sigler, P.B. *J. Biol. Chem.* 1985, **260**, 11627