Segmentation and Interpretation of 3D Protein Images

Laurence Leherte *
Laboratoire de Physico-Chimie Informatique
Facultes Universitaires Notre-Dame de la Paix
Rue de Bruxelles, 61, B-5000 Namur, Belgium
leherte@scf.fundp.ac.be
fax: 32-81-72.45.30

Kim Baxter Janice Glasgow Suzanne Fortier
Departments of Computing and Information Science and Chemistry
Queen's University
Kingston, Ontario, Canada K7L 3N6
baxter@qucis.queensu.ca; janice@qucis.queensu.ca;
fortiers@qucdn.queensu.ca
fax: (613) 545-6513

Abstract

The segmentation and interpretation of three-dimensional images of proteins is considered. A topological approach is used to represent a protein structure as a spanning tree of critical points, where each critical point corresponds to a residue or the connectivity between residues. The critical points are subsequently analyzed to recognize secondary structure motifs within the protein. Results of applying the approach to ideal and experimental images of proteins at medium resolution are presented.

Introduction

Modern crystallographic studies are at the forefront of current efforts to characterize and understand molecular structures and molecular recognition processes. The information derived from such studies provides a precise and detailed depiction of a molecular scene, an essential starting point for unraveling the complex rules of structural organization and molecular interactions in biological systems. However, despite recent technological advances, protein structure determination remains a lengthy and complex task. As a result, only a small fraction of the currently known proteins have been fully characterized.

The determination of crystal structures from their diffraction data belongs to the general class of image reconstruction exercises from incomplete and/or noisy data. In the case of protein structures, a major hurdle in the image reconstruction process is the so-called "phase problem", i.e., the extraction of phase information from the measured experimental data. Current solutions to this problem rely on gathering extensive experimental data and on considerable input from experts during the image interpretation process.

The goal of the research described in this paper is to facilitate the image reconstruction processes for protein crystals. Towards this goal, techniques from artificial intelligence, machine vision and crystallography are being integrated in a computational approach to the interpretation of electron density maps of proteins. Crucial to this interpretation process is the ability to locate and identify meaningful features of a protein structure at multiple levels of resolution. This requires a simplified representation of a structure, one that preserves shape, connectivity and distance information. In the proposed approach, molecular scenes are represented as three-dimensional (3D) spanning trees, where nodes of the tree correspond to critical points (peaks and passes) in the image data. The methodology is currently being applied to electron density maps of pro-

*The authors thank C.K. Johnson for sharing the ORCRIT program and for many helpful discussions. They also thank M. Fraser for providing experimental data. The research described in this paper has been supported by the Natural Science and Engineering Council of Canada (NSERC), the National Belgiam Foundation for Scientific Research (FNRS), IBM Belgium, Facultes Universitaires Notre-Dame de la Paix, the NATO Scientific Division and the Cambridge Crystallographic Data Center.

teins at medium ($\sim$ 3 Å) resolution. For such images, the critical points correspond to amino acid residues (peaks) and their adjacency in the primary sequence (passes). Initial results suggest that at medium resolution the electron density maps can successfully be segmented into protein and solvent regions, main and side chains and into individual residues along the main chain. Furthermore, algorithms have been developed to analyze the spanning trees so as to determine secondary structure motifs in the molecule.

The paper presents an overview of the protein structure determination problem in the context of scene analysis in machine vision. The processes of segmentation and recognition of secondary structure motifs in spanning tree representations of proteins are also described, along with some preliminary experimental results. The paper concludes with a discussion of ongoing research in the area.

Analysis of Visual Scenes

Research in machine vision has long been concerned with the problems involved in automatic image interpretation. Marr (1982) defines computational vision as "the process of discovering what is present in the world, and where it is". Similar to visual scene analysis, molecular scene analysis is concerned with the processes of reconstruction, classification and understanding of complex images. Such analyses rely on the availability of *a priori* information, in the form of structural templates and in the form of rules and heuristics, to locate and identify features in a scene. This section presents the problem of molecular scene analysis in the context of related research in machine vision.

Early vision systems consist of a set of processes that determine physical properties of three-dimensional surfaces from two-dimensional (2D) arrays. These arrays contain pixel values that denote properties such as light intensity, tissue density, depth, etc. Unlike input for the vision problem, the crystallographic experiment can yield 3D data which allow for the construction of a 3D array of voxels (volume elements). Each voxel contains a value representing the height of the electron density distribution function at the given location.

A 3D image of the atomic arrangement in a crystal is readily accessible for small molecules from data generated using X-ray diffraction techniques. Given the magnitudes of the diffracted waves and prior knowledge about the physical behavior of electron density distributions, probability theory is applied to retrieve phase information. Once magnitudes and phases are known, the spatial arrangement of the atoms within the crystal can be obtained using a Fourier transform procedure. The function that is obtained, $\rho(\mathbf{r})$, is a scalar field visualized as a 3D grid of real values (electron density map) in which high density centers are associated with atoms.

For proteins, the construction of a 3D image from the diffraction data is much more complex and time-consuming. It usually involves many iterations of calculation, map interpretation and model building, which rely extensively on input from an expert. It has been suggested, however, that the process could be significantly enhanced by combining mathematical and AI strategies, and rephrased as a hierarchical and iterative scene analysis exercise (Fortier *et al.* 1993). The goal of the exercise would be to reconstruct and interpret images of progressively higher resolution. Thus, in an initial low resolution map, where the protein appears as a simple object outlined by its molecular envelope, the goal would be to locate and identify protein and solvent regions. At medium resolution, where the protein appears as a more complex object, the goal would be to locate and identify main and side chains, recognize secondary structure motifs and possibly locate individual residues along the main chain. At higher resolution, the analysis would attend to the identification of amino acid residues and, possibly, the location and identification of individual atoms.

A primary step in low level scene analysis is to automatically partition (segment) an image into disjoint regions that can be given a symbolic description. Ideally, each region will correspond to a semantically meaningful component or object of the scene. These parts can be used as input to a high level recognition task. The nature of the partition and symbolic description depend on the type of processing to be applied. When model-based recognition is the next step in the analysis, the description should be in a form that is easily comparable with models in the database. The quality of the final output is dependent on the quality of the segmentation. Although these processes may appear sequential – first segmentation then recognition – in practice they are often interdependent. General purpose, domain independent segmentation techniques may be a necessary first step, but domain knowledge, in the form of a partial interpretation, is often useful for assessing and guiding further segmentation.

Several approaches to image segmentation have been considered in the vision literature.[1] *Thresholding* has proven effective for separating a small number of objects from a contrasting background, while *edge detection* has been used to separate regions by locating differences between the regions. One operator used in the latter approach is the zero-crossing of the second derivative (Marr 1982). In *region extraction*, segmentation is carried out by determining similarity within a region. Typically, a seed region is chosen, and then expanded by adding adjoining similar regions. *Topological* approaches have been used to provide initial estimates for segmentation in range images and in some medical applications. Besl and Jain (1986) apply a topological approach which evaluates the surface curvature and sign of the Gaussian for each point on the surface of range images, and uses the derived primi-

[1] See (Arman & Aggarwal 1993; Pal & Pal 1993) for a detailed overview of these approaches.

tives (peak, pit, ridge, saddle ridge,...) to perform the initial splitting. Gauch and Pizer (1993) identify ridges and valley bottoms in 2D images (a ridge is defined as a point where intensity falls off sharply in two directions, a valley bottom is a point where the intensity increases sharply in two directions) and follow their behaviour through scale space. As the resolution is reduced with Gaussian blurring, ridges and valleys are annihilated; the resulting hierarchy can be used for several analysis tasks including segmentation.

As will be discussed in the next section, a topological approach is being used in the segmentation and recognition of molecular scenes. Similar to the approach of Gauch and Pizer, critical points are used to delineate a skeletal image of a protein and segment it into meaningful parts (secondary structure, residues, atoms, etc.). These critical points are analyzed (using domain rules) to aid in the recognition of the segmented parts. This approach has some similarity with the *skeletonization* method which has been described by Hilditch (1969) , and applied in protein crystallography by Greer (1974) . However, unlike Greer's algorithm, which "thins" an electron density map to a set of connected points that trace the main and secondary chains of the molecule, the proposed representation preserves the original volumetric shape information by retaining the curvatures of electron density at the critical points. A methodology for outlining the envelope of a protein molecule in its crystallographic environment has previously been proposed by Wang (1985) , while Jones *et al.* (1991) have achieved significant advances in approaches for the interpretation of medium to high resolution protein maps.

In summary, the analysis of molecular scenes can be considered in the general class of scene analysis problems. However, the representation, segmentation and recognition of molecular images differ from vision applications in a number of ways. Most significantly, diffraction data are often 3D in nature, which simplifies or eliminates many of the problems faced in low level vision (e.g., occlusion, shading). The complexity that does exist in the crystallographic domain relates to the incompleteness of data due to the phase problem.

Analysis of Electron Density Maps

In the development of a computational methodology for the analysis of protein structures, methods from machine vision and crystallography were considered. Among the methods studied, the topological approach seemed the most natural way to catch the fluctuations of the density function $\rho(\mathbf{r})$ in the molecular image. In this section we overview a methodology that transforms a three-dimensional electron density map into a spanning tree of critical points that trace the main chain of the protein structure. Experimental results from applying the approach to the segmentation and interpretation of medium resolution maps are also presented.

Representation of Protein Structures

The topological approach to protein image interpretation is based on the representation of a scene in terms of the critical points of the electron density function, i.e., the points where the gradient of $\rho(\mathbf{r})$ vanishes. At such points, maxima and minima are defined by computing second derivatives which adopt negative or positive values respectively. For a 3D function, three principal second derivatives, or eigenvalues, are computed at each position vector $\mathbf{r}$. Four possible cases are considered depending upon the number of negative eigenvalues, n_E. When $n_E = 3$, the critical point corresponds to a local maximum or *peak*; a point where $n_E = 2$ is a saddle point or *pass*. $n_E = 1$ corresponds to a saddle point or *pale*, while $n_E = 0$ characterizes a *pit*. The use of critical point mapping as a method for analyzing protein electron density maps was first proposed by Johnson (1977b) and later used in Crysalis (Terry 1983), an expert system designed for the automated interpretation of high resolution protein electron density maps. Within the framework of the Molecular Scene Analysis project (Fortier *et al.* 1993; Glasgow, Fortier, & Allen 1993), the topological approach is being extended for the analysis of medium and low resolution maps of proteins.

Topological analysis has been implemented by Johnson in the computer program ORCRIT (Johnson 1977a). By first locating and then connecting the critical points, this program generates a graph representation for an electron density map of a protein. The occurrence probability of a connection between two critical points i and j is determined by following the density gradient vector $\nabla\rho(\mathbf{r})$. For each pair of critical points, the program calculates a weight w_{ij}, which is inversely proportional to the occurrence probability of the connection. The collection of critical points and their linkage is represented as a set of minimal spanning trees (connected acyclic graphs of minimal weight). In the earlier Crysalis project, the ORCRIT program was used to segment a high resolution electron density map into critical points where peaks correspond to the location of atomic parts and passes correspond to the bonds between atoms. More recently, we have determined that at medium resolution peaks correspond to amino acid residues along the main chain of the protein and passes to the connectivity determined by the primary structure of the protein (Leherte *et al.* 1994). As illustrated in Figure 1, the topological approach produces a skeleton of a protein backbone as a sequence of alternating peaks (solid circles) and passes (open circles), where each peak is associated with one residue of the protein. For larger residues, the side chains are also included in the tree.

It should be noted that the electron density distribution function is a smooth function with no sudden changes. Its zero-crossings are detected by characterizing points where the gradient, or the first-derivative, of the function vanishes. The second derivatives provide

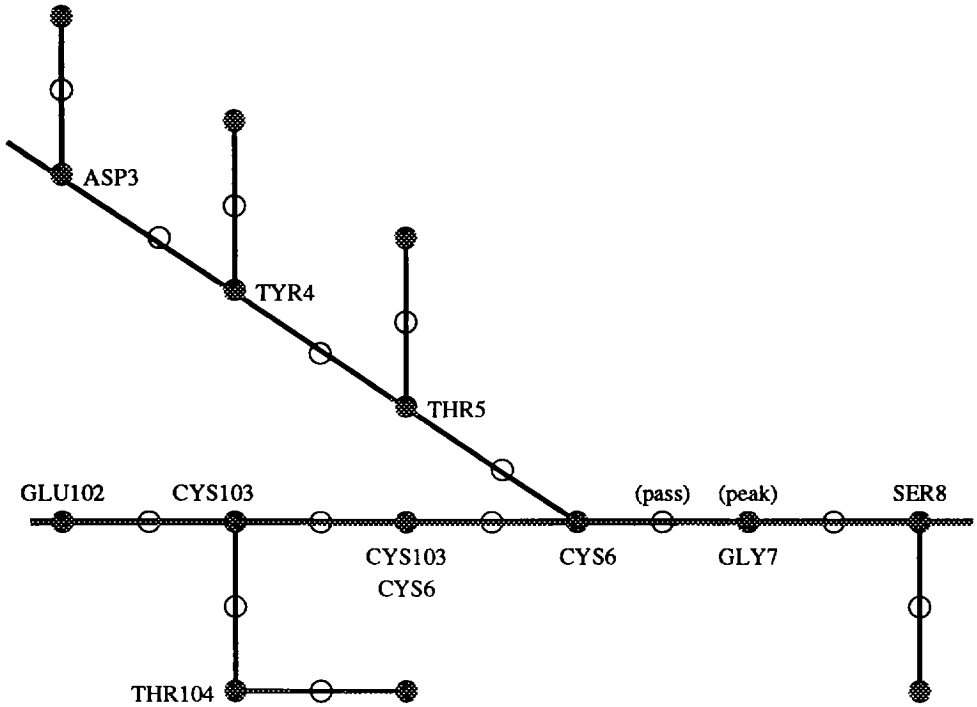

Figure 1: Planar representation of the critical point spanning tree for protein structure 1RNT.

information on the characteristics of the zero-crossings and, in particular, identify whether they are peaks, passes, pales or pits. In 2D images, such as those considered by Marr (1982), sudden changes in intensity are present. They give rise to a peak or a pit in the first derivative and, therefore, the contours of images are detected at points where the second derivative vanishes.

Segmentation of Electron Density Maps

This section presents experimental studies that have been carried out on electron density maps at 3 Å resolution. Computations were first performed on calculated maps reconstructed from available structural data in order to generate a procedure for the further analysis of experimental maps. Three protein structures, Phospholipase A2 (1BP2), Ribonuclease T1 complex (1RNT) and Trypsin inhibitor (4PTI), retrieved from the Brookhaven Protein Databank (PDB) (Bernstein *et al.* 1977), were considered. These structures are composed of 123, 104 and 53 residues, respectively. The electron density maps for the proteins were constructed using the XTAL program package (Hall & Stewart 1990), and were then analyzed using ORCRIT.

High density peaks and passes were the only critical points considered in this study. Low density critical points are less significant since the electron density distribution is modulated by either experimental noise and/or errors due to the fast Fourier transform process. In addition, the analysis levels some low density points (those with negative values) to zero. High density peaks and passes were considered by imposing a cut-off value below which the critical point search procedure is not applied.[2]

The results obtained from the analysis of the three calculated density maps led to the following observations:

- The main branch of the spanning tree traces out the backbone of the protein molecule.
- Each peak of the main branch of the tree is associated with a single residue of the primary sequence for the protein. Furthermore, the peaks are located close to the $C_\alpha CO$ centres of charge for the residue.
- Side chains are often observable, particularly for the larger residues. These chains are represented as side branches that link to the main branch of the spanning tree.

The result of applying the ORCRIT program is thus a partitioning of the electron density map into two main regions: the protein region represented by a chain of connected critical points, and a solvent region which is characterized by low density values and non-connected critical points.

As was illustrated in Figure 1, the ideal critical point represention of a protein at medium (3 Å) resolution can be depicted as a tree composed of a long principal branch built on alternating peaks and passes with

[2]Further details of the experimental process are reported in (Leherte *et al.* 1994).

small side branches jutting out of it. In practice, however, such a representation may include some errors originating from the presence of connections between critical points associated with non-adjacent residues. Figure 2 presents a comparison between the backbone of a protein structure and its critical point representation. In the constructed main chain of the spanning tree, jumps or bridges occur because of the presence of disulfide bridges (S-S), heteroatoms (CA++), or H-bonds between close residues. These connections can often be detected by applying further analysis to the critical points.

Neglecting the passes located between peaks, geometrical parameters were computed for short fragments composed of four adjacent peaks in the main branch of the spanning trees. Before achieving this geometrical analysis, some preprocessing work was done in order to fit the spanning trees to the ideal model described above. Symmetry coincident critical points were removed. Distances were computed for sets of adjacent peaks, and peaks separated by a distance smaller than 1.95 Å were merged into a single point. The critical point linkage was then checked: if two adjacent peaks were separated by a distance of ≤ 7 Å then the peaks were assumed to be connected. Considering three peaks at a time, if the distance between the first and third peak was larger than 4 Å, then the middle peak was considered to be a side chain peak.

A statistical analysis of the geometry of critical points sequences further showed that the most useful parameters for the identification of helices and extended motifs (β-sheet segments) were the torsion angles and the distances between peaks p_i and p_{i+3}, while bond angle values were less discriminating. In the next section we discuss how these criteria were used to determine secondary structure motifs in a protein.

Secondary Structure Recognition

From our experiments on ideal electron density maps of proteins it was concluded that the topological approach was able to segment the protein structure into main and side chains and capture the conformation of its main chain. The recognition of secondary structure features from medium resolution electron density maps could thus be based on pattern matching of the critical point networks with templates of critical point networks for idealized secondary structure motifs. A set of *IF-THEN* rules were derived that compare the angles and distances for an uninterpreted spanning tree of critical points with those derived from the previously determined maps. Table 1 summarizes the geometrical parameters that form the basis of these rules applied for the classification of protein segments.

The parameters in Table 1 provide a basis for the calculation of measures that represent the quality of fit between a critical point segment and a helical or sheet motif. Two degrees of belief for a critical point c can be calculated: the degree of belief that critical point c belongs to a β sheet ($dbs(c)$), and the degree of belief that c belongs to a helix ($dbh(c)$). A degree of belief falls between the values 0 and 100, where the larger values denote greater confidence in the classification. An ideal sequence of critical points (either depicting a helix or a β-sheet segment) would be characterized by a sequence of large belief measures (*dbh* or *dbs* respectively). For example, a sequence *<26,53,80,100,...,100,80,53,26>* of *dbh* values would denote an idealized helix.[3]

Geometrical parameter	Helix	β sheet
Torsion angle (degrees)	30-90	|110-180|
Distance 1-4 (Å)	4.4-6.3	≥ 6.9
Bond angle (degrees)	60-110	≥ 90

Table 1. Ranges of angle and distance values considered for the identification of secondary structure motifs in critical point spanning trees.

An additional test, which considers the environment of the extended system, can be applied to the recognition of β-sheets. Each pair of segments is considered in this test. When a pair of segments is characterized by at least three pairs of peaks having interdistances ranging between 4 and 7 Å, then the level of belief that they are parts of a β-sheet is increased.

The results of the application of the above rules to the protein 1BP2 are illustrated in Figure 3. In this figure, the horizontal axis represents the peak sequences aligned next to each other, while the degree of belief values resulting from the application of the helix (*dbh*) and β-sheet (*dbs*) segment recognition rules are reported along the vertical axis. Figure 3 demonstrates that, when a peak sequence effectively corresponds to an existing helix (H), the rules yield degrees of belief larger than 46. The locations corresponding to turns (T) have lower confidence levels. All β-sheet segments found to be parallel (S) were effectively associated with a β-sheet.

Application to an Experimental Map of Penicillopepsin

The proposed approach has also been applied to an experimental map of penicillopepsin, which was calculated using the Groningen BIOMOL crystallographic program package. In a previous paper (Leherte *et al.* 1994) we reported results that were derived by considering the peaks as independent objects, i.e., the success rate of the recognition method was estimated by considering percentages of correctly identified peak. The results indicated that 82% of the identified peaks were correctly recognized. However, secondary structure motifs are built on sequences of residues (or peaks). It is thus important to also consider the degree of be-

[3]Note that the maximum degree of belief for end points of a segment is less than 100. This is because there are fewer torsion angles that can be measured to raise the confidence.

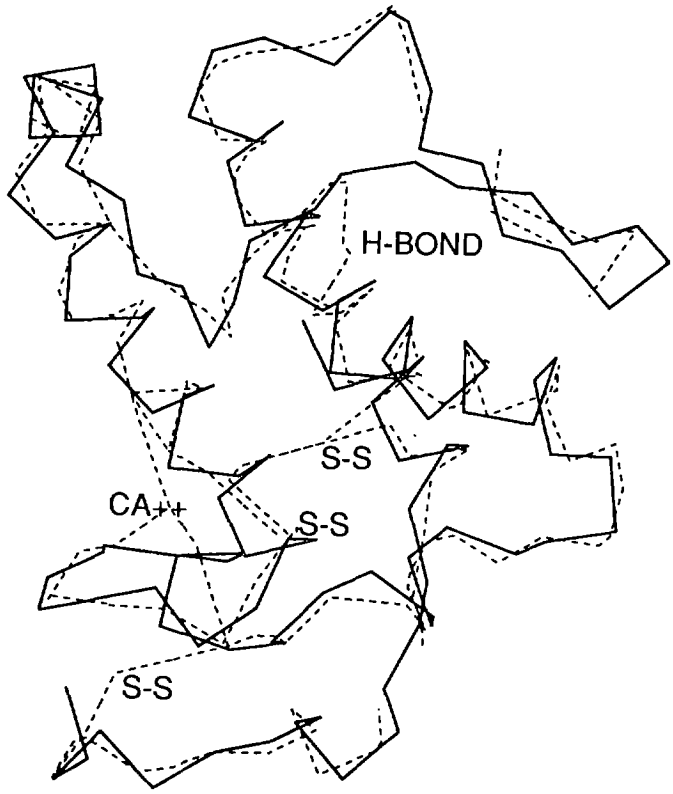

Figure 2: Perspective view depicting the superimposition of the Cα chain for protein structure 1BP2 (solid line) to the main chain of the corresponding spanning tree (dashed line).

lief associated with the neighbouring peaks. Using the previously described geometry-based rules, 59 probable β-sheet segments were obtained. The results were grouped into five classes, based on various combinations of the following conditions:

- $c1$: the segment under study has at least one *dbs* value equal to 100;
- $c2$: the segment is parallel to another one; and
- $c3$: the sequence of the dbs values follows the values reported in the previous section.

Table 2 summarizes the experimental results for the five classes considered.

The first class ($c1 \bigwedge c2 \bigwedge c3$) reported on in Table 2 yielded accurate results. All the segments are effectively associated with a real β-sheet. However, it is observed that the percentage of correctly identified peak (success rate) is not 100%. This is due to the fact that most of the recognized β-sheet segments are usually shifted by one residue with respect to the definition given in the PDB file.[5] It is thus concluded that at medium resolution the results of a fully successful recognition procedure lead to a success rate of about 90% when compared to secondary structure assignment results obtained at high resolution.

[5]The shift of the segment by one residue is due to the ambiguity in recognizing the extremity points: they may be part of two different possible secondary structure motifs.

In the second class, only one segment does not correspond to a real β-sheet segment. Its maximum *dbs* value is equal to 53. The success rate of this class is still impressive (86%), but is characterized by a large variation. Not surprisingly, the first two classes, which correspond to the highest degrees of belief, also involve the longest segments. The results worsen when either condition $c2$ or condition $c3$ is not observed. The most important or crucial condition appears to be the parallelism criteria. Effectively, in the third class where parallelism alone is considered, 67% of the peaks are correctly identified. Unfortunately, only three occurrences were observed for this class, so the statistics are not as reliable as for the previous classes. For the fourth class, in which condition $c3$ alone is considered, only 57% of the peaks are correctly identified. In the last class, the three segments which are really parts of a β-sheet contain only two peaks. They are at the origin of the 30% success rate. Such small segments do not occur in any of the other four classes.

Due to their low occurrence frequency and size in the penicillopepsin structure, the number of recognized helical motifs was quite low. The application of the recognition procedure led to the identification of 8 helical segments, all characterized by *dbh* values less than 100. However, as concluded from the analysis of reconstructed maps, the 6 segments having a *dbh* value larger than 46 reflected the presence of a real helix.

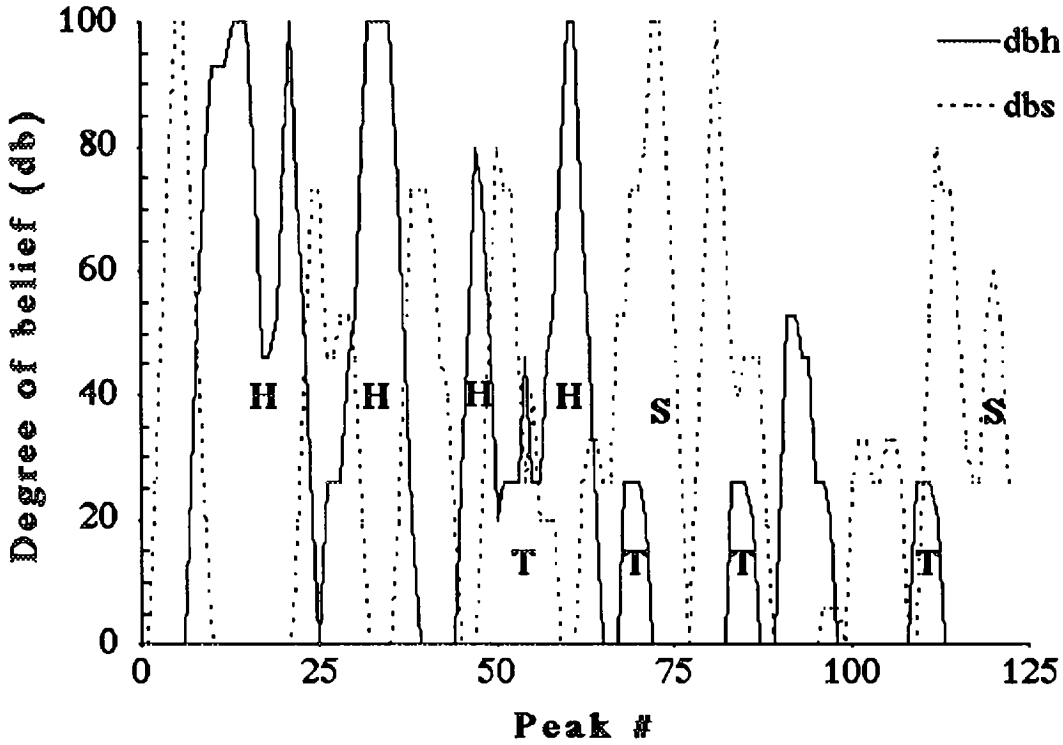

Figure 3: Helix degree of belief (*dbh*) and β-sheet degree of belief (*dbs*) calculated from the application of the secondary structure recognition rules to the peak sequences obtained from the topological analysis of the reconstructed maps of protein 1BP2 at 3 Å resolution using ORCRIT.

In this experiment, 5 of the 32 β-sheet segments and 5 of the 10 helices were not discovered. The presence of jumps and breaks in the critical point sequences were responsible for the non-detection of 4 motifs (2 sheet segments and 2 helices), and 6 motifs (3 sheet segments and 3 helices), respectively.

Discussion

It was reported in this paper that the topological approach can effectively segment medium resolution electron density maps of proteins. Furthermore, it was shown that secondary structure motifs could be recognized in the map through the use of simple geometry-based rules. The application of these rules yields a measure of the degree of confidence in the recognition of a given motif. This is important since the proposed methodology can serve, not only as an aid to expert crystallographers in their interpretation and model building tasks, but also more actively in the structure determination process. Required levels of confidence would clearly depend on the use that is made of the results of the topological analysis.

The work described here is now being extended for applications in both lower (5 Å) and higher (2.7 Å) resolution maps. In particular, experiments are being conducted at low resolution to assess the usefulness of the topological approach to the definition of the protein envelope. At higher resolution, the goal is to determine the direction of the main chain and to attend to the identification of individual residues.

In addition, the tree construction algorithm of ORCRIT is being altered to output multiple plausible skeletons. Additional methods for evaluating these skeletons are also being considered. One promising approach borrows from research in protein structure prediction and, in particular, from its formulation as an inverse folding problem (Lathrop & Smith 1994). Given an amino acid sequence and a set of core segments (pieces of secondary structure forming the tightly packed internal protein core), this approach evaluates each possible alignment (threading) of the sequence onto possible core templates. The problem of identifying individual residues in a critical point map constructed at medium to high resolution can be addressed in a similar manner, i.e., by threading a sequence onto a core structure. However, the problem is simpler than in protein structure prediction since it is reduced to threading a sequence onto its own experimentally determined structure, rather than onto templates retrieved from a library of possible models. In the threading approach proposed by Lathrop and Smith, a scoring function is used. This function considers the sum of singleton terms, which depends only on the threading of single core segments, and the sum

Class	Total number of segments (Ns)	Mean max. dbs	Mean length	# of segments associated with a real β-sheet$^+$	% of correctly identified peaks (calc. over Ns)
$c1 \bigwedge c2 \bigwedge c3$	11	100	9±2	11	90±8
$\sim c1 \bigwedge c2 \bigwedge c3$	14	64±15	6±1	13	86±21
$\sim c1 \bigwedge c2 \bigwedge \sim c3$	4	55±4	5±1	3	67±27
$\sim c1 \bigwedge \sim c2 \bigwedge c3$	11	62±26	5±2	7	57±43*
$\sim c1 \bigwedge \sim c2 \bigwedge \sim c3$	19	50±23	4±2	3	30±36*

Table 2. Classification of the recognized β-sheet segments of penicillipepsin.[4]

of pairwise interactions between neighboring core elements. (These functions represent the amino acids statistical preference for certain environments.) Additions to the scoring function, such as statistical bulk properties, are being considered to take full advantage of the information provided by ORCRIT.

A long-term goal of our research in molecular scene analysis is to develop a computational methodology that can aid in the reconstruction of protein structures from their initial low resolution electron density maps so as to resolve the map until a high-resolution fully interpreted image emerges. The topological approach presented here is an important component of this methodology. Further research is required, however, to extend it to low and high resolution maps, and to incorporate more domain knowledge into the analyses.

References

Arman, F., and Aggarwal, J. 1993. Model-based object recognition in dense-range images - a review. *ACM Computing Survery* 25(1):5–43.

Bernstein, F. C.; Koetzle, T. F.; Williams, J. B.; Meyer Jr., E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; and Tasumi, M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535-542.

Besl, P., and Jain, R. 1986. Invariant surface characteristics for 3d object recognition in range images. *CVGIP* 33:33–80.

Fortier, S.; Castleden, I.; Glasgow, J.; Conklin, D.; Walmsley, C.; Leherte, L.; and Allen, F. 1993. Molecular scene analysis: The integration of direct methods and artificial intelligence strategies for solving protein crystal structures. *Acta Crystallographica* D1.

Gauch, J., and Pizer, S. 1993. Multiresolution analysis of ridges and valleys in grey-scale images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-15(6):635–646.

Glasgow, J.; Fortier, S.; and Allen, F. 1993. Molecular scene analysis: crystal structure determination through imagery. In Hunter, L., ed., *Artificial Intelligence and Molecular Biology.* AAAI Press.

Greer, J. 1974. Three-dimensional pattern recognition: an approach to automated interpretation of electron density maps of proteins. *Journal of Molecular Biology* 82:279–301.

Hall, S. R., and Stewart, J. M., eds. 1990. *XTAL 3.0 User's Manual.*

Hilditch, C. 1969. Linear skeletons from square cupboards. *Machine Intelligence* 4:403–420.

Johnson, C. K. 1977a. ORCRIT. the Oak Ridge critical point network program. Technical report, Chemistry Division, Oak Ridge National Laboratory, USA.

Johnson, C. 1977b. Peaks, passes, pales and pits: a tour through the critical points of interest in density maps. In *Proceedings of the American Crystallographic Association Meeting.* Abstract JQ6.

Jones, T.; Zou, J.; Cowan, S.; and Kjeldgaard, M. 1991. Improved methods for building protein models in electron-density maps and the location of errors in those models. *Acta Crystallographica* A47:110–119.

Lathrop, R., and Smith, T. 1994. A branch-and-bound algorithm for optimal protein threading with pairwise (contact potential) amino acid interactions. In *Proceedings of the 27th Hawaii International Conference on System Science.*

Leherte, L.; Fortier, S.; Glasgow, J.; and Allen, F. 1994. Molecular scene analysis: A topological approach to the automated interpretation of protein electron density maps. *Acta Crystallographica D* D50:155–166.

Marr, D. 1982. *Vision.* W.H. Freeman and Company: San Francisco.

Pal, N., and Pal, S. 1993. A review on image segmentation techniques. *Pattern Recognition* 26(9):1277–1294.

Terry, A. 1983. *The Crysalis Project: Hierarchical Control of Production Systems.* Ph.D. Dissertation, Stanford Heuristic Programming Project, Stanford University, California, USA.

Wang, B. 1985. Resolution of phase ambiguity in macromolecular crystallography. In Wyckoff, H.; Hirs, C.; and Timasheff, S., eds., *Diffraction Methods for Biological Macromolecules.* Academic Press, New York.