Protein Structure Determination: Combining Inexact Graph Matching and Deformable Templates

Kim Baxter and Janice Glasgow *
Department of Computer Science
Queen's Universit y ,Kingston, Ont. Canada K7L 3N6
email: baxter@cs.queensu.ca janice@cs.queensu.ca

Abstract

This paper presents a computational methodology for integrating techniques from ph ysics-based modeling and graph theory ,applied to the problem of protein structure determination from experimental X-ray crystallographic electron densit ymaps. Protein structure determination is an error-prone, time-consuming task, requiring considerable h uman expertise. The 3D images (electron densit ymaps) obtained from crystallographic experiments ha v e some unusual properties and manual in terpretation of these protein images is one of the main bottlenecks in structure determination. A primary goal of *molecular sc ene analysis*[1] is to automate this process. Although a topological analysis can be used to perform an initial segmentation of the image [2], a more detailed segmentation is preferred for further in terpretation [3]. Either error-correcting graph rewriting or deformable templates can be used to refine the initial segmentation, but both approaches hav e limitations when used alone.

In the proposed architecture, a topological analysis is used to segment the image. The segmentation is represen tedas an attributed graph; error-correcting subgraph isomorphism is used in conjunction with graph rewriting to improv e the segmentation and partially interpret the image. The corrected attributed graph is then used to initialize a series of deformable templates; the resulting model is used to guide further interpretation. We present the results of experiments designed to test the ability of the combined graph rewriting module and deformable templates to isolate and define individual amino acid residues. A goal of this work is to incorporate it in to a fully automated system for molecular scene analysis.

*The authors wish to thank Dr. Zongc hao Jia (Queen's University) and Dr. Marie Fraser (Univ ersity of Alberta) for providing access to the experimental EDMs for 1MSI and 3APP respectively. F unding for the researdh reported in this paper was provided b y the Institute for Robotics and Intelligent Systems, the National Science and Engineering Research Council of Canada, the Communications and Information Technology Ontario Cen ter of Excellence and the PENCE Netw ork Cen ter of Excellence.

1 Introduction

Each species of protein molecule has a unique 3D structure and the div ersity of protein structures enables these molecules to carry out thousands of different biological processes. A fundamental goal of research in molecular biology is to understand protein structure. Protein crystallography is currently the most successful method for determining protein structure; the results of experiments using X-ray diffraction can be used to create the 3D images we will be considering in this paper.

The image (called an electron density map or EDM) is a 3D grid of scalars representing the electron density probability function of a protein crystal. Because the objects depicted in an EDM hav e no real surfaces, multiple objects can occupy the same location. The image interpretation process is time-consuming and error-

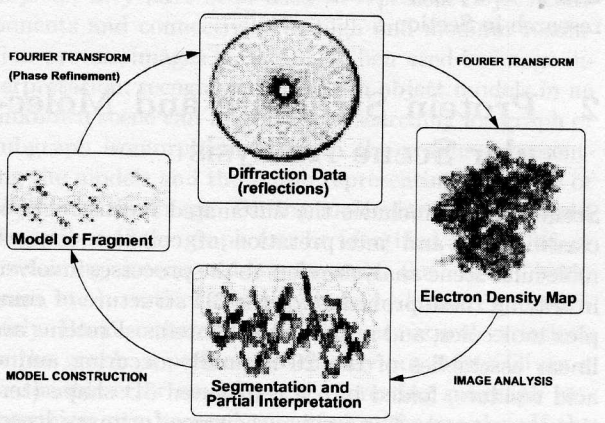

Figure 1: Protein Structure Determination From Experimental Data

prone; it requires extensive manual intervention via a graphics terminal. Most crystallographic software currently uses a combination of very low and very high level representations to model proteins [4, 5, 6]. We considered two approaches from machine vision and pattern recognition that employ models of intermediate complexity: deformable templates and attributed graphs. Deformable templates are ideally suited to represent the flexible nature of the chemical bonds in a protein, and when properly initialized, they are robust and stable. Unfortunately, initializing these representations directly from the image is problematic. Error-correcting graph rewriting can be used to refine an initial segmentation into an attributed graph suitable for further processing. Initializing the graph is not a problem, and individual residues and recurring structural motifs (secondary structure) can be identified, but the shape and adjacency information useful for interpretation is inaccurate. In our proposed system, graph rewriting is used to guide the initialization of deformable templates. The templates provide a more accurate, detailed interpretation of each object.

The purpose of the research presented in this paper is to develop a computational procedure that segments and partially interprets protein images at medium resolution. Section 2 provides relevant background about protein structure in the context of molecular scene analysis. Section 3 of the paper presents our computational framework, which integrates techniques for segmentation, image interpretation and model fitting to assist in the process of structure determination. Segmentation and recognition of secondary structure motifs using error-correcting subgraph isomorphism are described. This segmentation is used to initialize a series of deformable templates used to isolate individual residues and to determine side-chain orientation. Preliminary results of the integrated approach are reported in Section 4. The paper concludes with a discussion of future research in Section 5.

2 Protein Structure and Molecular Scene Analysis

Scene analysis includes the automated reconstruction, classification and interpretation of complex images; molecular scene analysis refers to the processes involved in solving these problems for the 3D structure of complex molecules, and in particular proteins. Proteins are linear assemblies of the 20 naturally occuring amino acid residues, folded into a convoluted 3D shape (tertiary structure). The residue sequence (primary structure) is known; each residue has one of the 20 different side-chains and a common basic structure. The

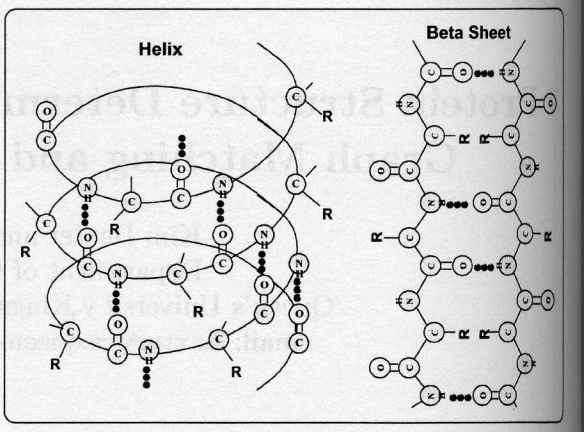

Figure 2: Examples of Secondary Structure

main chain (backbone) is composed of these common structures linked together. Chemical constraints can be used to help solve the protein's structure. Although the residue side-chains are flexible, they have a limited range of possible conformations (rotamers). In addition, interactions between the residues impose some structure (known as secondary structure) and limit the number of possibilities. Helices (coiled segments) and beta sheets (several segments forming pleated sheets) are the most common patterns (see Fig. 2).

X-rays passing through a protein crystal are diffracted by the electrons in the molecule. The diffraction pattern is recorded on a screen; the pattern of dots (reflections) in this 2D image is related to the EDM by a Fourier transform. Unfortunately, X-ray diffraction experiments provide only the amplitudes of the Fourier transform. Extracting phase information from this data is the classic phase problem described in [7]. Estimates of the lower resolution phases are used to generate an initial EDM. At each resolution, any portion of this image that can be interpreted is used (via inverse Fourier transform) to improve the phase estimates, and a new EDM is calculated. This cyclic process of crystallographic refinement and image interpretation results in a gradual improvement of the protein model (see Fig 1.). When the model is sufficiently detailed, more phase estimates are included, and the cycle is repeated at a higher resolution. Locating and defining the protein structure in each EDM is one of the bottlenecks in protein structure determination. In a low (5 Å) resolution EDM, the protein can be represented as a simple object against a low density solvent background. At medium (3 Å) resolution, objects correspond to groups of atoms. The backbone and larger side-chains are discernible. At high (1 Å) resolution, individual atoms can be located. In molecular scene analysis at medium resolution, part of the image interpretation process involves locating the

main-chain trace of the backbone, partitioning the trace into individual residues, matching the sequence to the trace, and determining the orientation and conformation of the residues.

3 System Description

Our system is composed of two interacting subsystems:

- an error-correcting graph construction module that takes as input a segmented protein image and produces an attributed graph. (Error-correction and some interpretation are performed while the graph is being built.)
- a residue fitting module that takes a partially interpreted attributed graph representation and uses it to initialize a set of deformable templates corresponding to individual residues in a backbone. The templates are used to resegment the image, and produce a reasonable correspondence between the protein image and each individual residue in the attributed graph.

These processes are intertwined, in the sense that the graph construction module uses (global) knowledge about residue connectivity and secondary structure, to initialize the deformable templates, which in turn exploit (local) knowledge about residue and backbone shapes. The processing can be iterative; the output from each module can be repeatedly used to initialize portions of the input to the other module, until a fully interpreted image emerges. In the remainder of this section, we describe the two modules.

3.1 Segmentation and Error-Correcting Subgraph Isomorphism

A topological analysis [1, 8] can be used to segment an EDM into disjoint regions and define connections between the regions. After using a low density threshold to label the solvent regions, the remaining EDM is partitionned using an inverted watershed algorithm. Each peak (local maximum in the EDM) has a region associated with it; at medium resolution (3 Å), larger (higher density) regions will contain from one atom to several residues. Ideally, each peak corresponds to one residue and each connection corresponds to a chemical bond. In practice, connections may be added or missing. We represent the output of the topological analysis of an EDM as an attributed graph. Each region is a vertex; region attributes include peak density, region volume, length of principal axis, moments of inertia and ellipticity. Regions which share a boundary are considered to be connected and each connection is an edge; edge attributes include density and boundary area. As the graph is interpreted, attributes identifying backbone versus side-chain regions, secondary structure, bond type, etc. will be added. Distinctive subgraphs corresponding to a variety of protein substructures will be present in the graph (see Fig. 3).

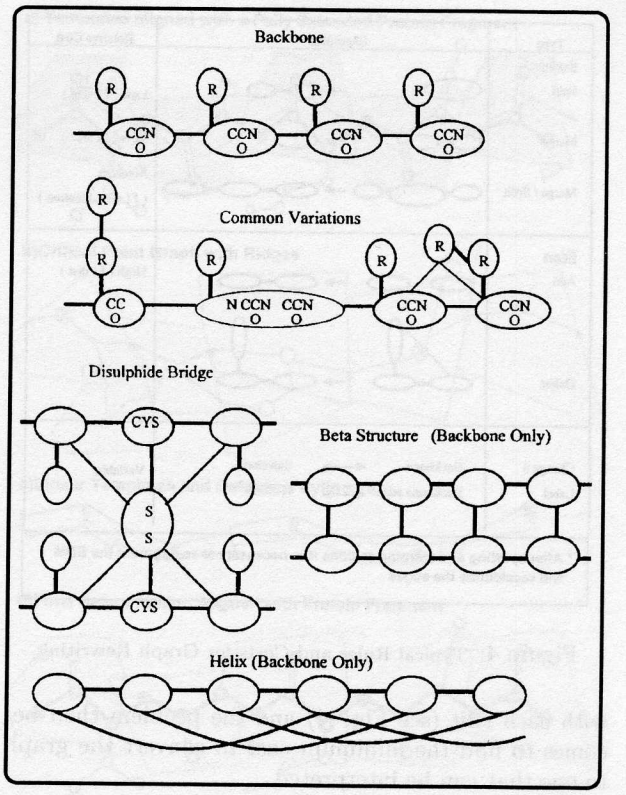

Figure 3: Common Subgraphs

Graphs, and in particular, labeled or attributed graphs provide a powerful representation for complex objects; they have been used to represent object components and connectivity in high and medium resolution protein images [1, 9, 10]. When used in image interpretation, recognition of known object models in an unknown scene can be solved by searching for graph or subgraph isomorphism between the graphs representing the models and the graph representing the scene or objects in the image. During molecular scene analysis, if portions of the graph can be identified as part of, or containing, interpretable subgraphs (i.e. helix, backbone, etc.), vertices and edges can be labeled appropriately. Because there may be noise or natural variations in the graph, a perfect match may not exist. Given a reasonable error model, error-correcting subgraph isomorphism can be used to compare the graphs. If the graph contains errors, graph rewriting can be used to correct it. The error model is used to associate a cost

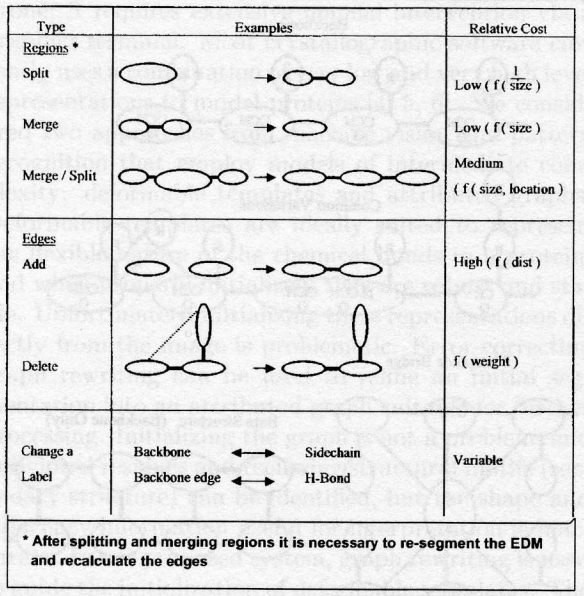

Figure 4: Typical Rules and Costs for Graph Rewriting

with each edit (see Fig. 4) and the problem then becomes to find the minimum cost to convert the graph to one that can be interpreted.

Subgraph isomorphism is in general an NP-hard problem and most algorithms only compare two graphs at a time. Often there is a large database of object model graphs and it is necessary to see if any of them are in the scene, or to see if part of the scene occurs in any of them. Messmer and Bunke [11, 12] have developed two algorithms to simultaneously compare a large number of object model graphs with the unknown input graph in polynomial time. Both exact and inexact matching are possible. In both cases, it is necessary to preprocess the database of model graphs into a data structure. This step is exponential, but it only needs to be done once. We have applied their algorithm to interpret the graph and to guide the resegmentation of 'problematic' areas of the EDM.

An incremental graph growing approach is used to further reduce complexity. Initially, the graph contains all the vertices and no edges. Edges are added in order of decreasing density[1]. When a connected subgraph excedes a certain size, exact subgraph isomorphism is used to determine if there is part of any recurring motif present and the graph labels are tentatively updated accordingly. Once a subgraph excedes a larger threshold, inexact subgraph isomorphism is used to locate any inexact matches to the target subgraphs. If the cost for the match is low, the graph is rewritten and relabeled[2].

[1]Locally, edges corresponding to backbone connections and disulphide bridges have the highest density. Side-chain connections are next, followed by secondary structure and then by noise.

3.2 Deformable Models and Templates

Protein images are noisy, contain objects which are flexible, have a range of possible appearances and have no clearly definable edges. Deformable models and templates are ideally suited to extract the variety of possible shapes from a noisy image. Deformable (active) models are energy minimizing representations that iteratively deform from an initial state until the resulting models lie on or near structures in the image. The type of structure and the final model shape are dictated by the nature of the energy equation. One of the simplest (and the earliest) forms of active models is the snake [14], an energy minimizing spline. Given an initial placement, it is pushed by image forces to the nearest salient image contour. Specifically, Kass' snake has an energy function, defined as:

$$E_{snake} = \int_0^1 (E_{int}v(s) + E_{image}v(s) + E_{constraint}v(s)).ds$$

where

E_{int} represents the internal energy of the spline due to stretching and bending,

E_{image} is a measure of the attraction of image features such as contours, and

$E_{constraint}$ is a measure of external constraint forces such as those due to higher level knowledge.

Cohen [15] developed surface-seeking 'balloon' models to extract surfaces in 3D images. A balloon is initialized inside the object and an internal pressure is used to inflate the balloon model. A variety of deformable models have been used to analyze 3D images [16]. More generally, $E_{internal}$ is an implicit description of the preferred shape; E_{image} encourages the model to find desired features in the image and to deform the model to the best compromise between the preferred form and the image data. $E_{constraint}$ can be used to incorporate miscellaneous higher-level knowledge which may not be related to either the model shape or image features. Many early implementations of deformable models were susceptible to falling into local minima; careful initialization or more global, time-consuming optimization techniques such as simulated annealing have been used to avoid this problem.

Snakes and balloons are useful when there are no strong assumptions about the expected shape of the model. When there are hard constraints associated with the possible shapes, the best features of rigid templates and deformable models can be combined by using a parametric shape-model with comparatively few degrees of freedom. The shape-model is a geometric model (similar to a rigid template), with parameters that have a probability distribution, or range. This prunes the search space to those shapes that are possible solutions. The template is matched to the image by

[2]Currently the decision to rewrite is interactive.

finding the parameters that minimize some energy function. $E_{external}$ is a function of the match to the image, and $E_{internal}$ is a function of the probability distribution (or range) of the shapes. Because the objects in an EDM merge with each other, there are no clear surfaces. Snakes and balloons could be used to outline a backbone tube and side-chain attachments; deformable templates of individual linked residues would be more suitable for a final resegmentation.

For the purposes of our molecular scene analysis, a variety of increasingly complex deformable templates are used to further define the residues and side-chain orientation. The final output of each stage is used to initialize the next set of models. First, the output of the attributed graph is used to determine the number of residues and to initialize the first template. Two types of control points are used to define the templates: "peak" points and "ridge" points. Peak points are considered to represent the centre of the residue model; each backbone region has one peak point initialized to its topological peak (if the region was not split or merged during graph rewriting) or to its centroid. Some number of ridge points are initialized on the ridge or the medial axis between neighbouring peak points.

$E_{internal}$ for each peak point includes a function of the distance between it and adjacent peak points. Any distance less than 3.0 Å or greater than 4.5 Å adds an energy term proportional to the difference. There is a similar function for the distance between each peak point, and the peak points two and three regions away. These encourage realistic residue spacing, and realistic angles and dihedral angles, respectively. Ridge points only have a constraint on the distance to the neighbouring control points. E_{image} is inversely proportional to electron density, and proportional to the distance from the ridge between the peaks and passes. $E_{external}$ constrains the end of the template from moving.

To minimize the total energy of the deformable templates(s), a simple gradient descent algorithm is used. Each energy term can be converted to a force acting on the control points. At each time step, the control point moves a distance proportional to the force, in the direction of the force; this requires minute time steps and good initialization. The relative weights in the force calculation can be altered to balance robustness and accuracy.

Once the energy is minimized, the resulting model is used to initialize the next deformable template; each linear segment becomes the axis of a narrow cylinder (see Fig. 5c). The cylinder surface is represented as a grid around an axis, and an energy term is calculated for each grid point. In addition, the cylinder exerts a force on the control points at its ends. E_{image} for

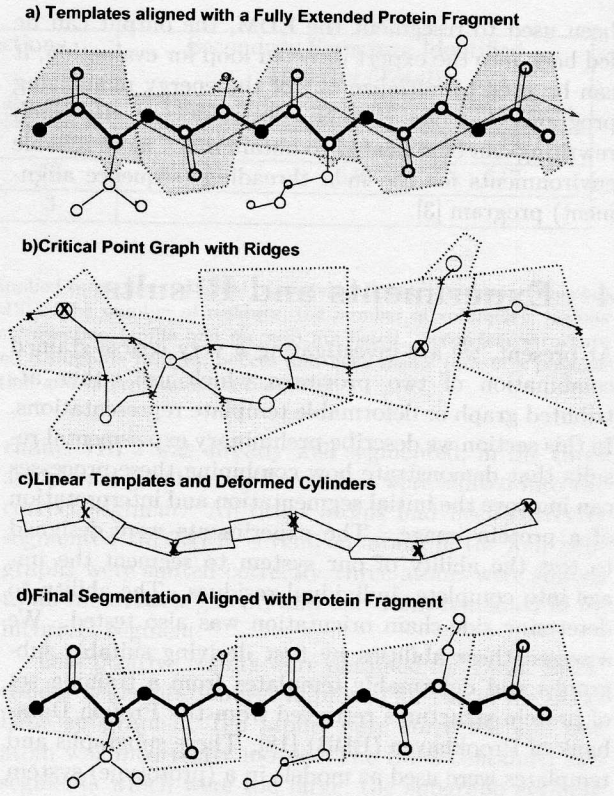

Figure 5: An idealized illustration of applying our system to the interpretation of an extended protein fragment. (a) shows a 2D view of the non-hydrogen atoms and bonds of a fully extended protein fragment comprising five residues, aligned with the matching linear and volumetric deformable templates developed using the database. Each volumetric template model contains the nitrogen (N), carboxyl (C=O) and carbon alpha (C) atoms and a protrusion where the side-chain would be. (b) shows the regions, critical points and ridges obtained using topological analysis of the EDM for the protein fragment. This segmentation results in one region per residue and three of the regions contain the expected atoms. The middle region is oversized: it contains the C alpha and side-chain from the (small) region to its left. (c) Graphical analysis would identify the backbone regions and edges; the initial, piecewise-linear models correspond (approximately) to the (backbone) ridges between the peaks. (c) shows a cylindrical model initialized on the final linear model. (d) gives the alignment between the final model and the protein. The atoms that were incorrectly located in the centre region are now associated with the correct region/residue.

the grid points includes an inflation energy related to the density and the gradient along the direction from the axis. A term limiting the moment of inertia along the entire axis prevents overinflation. $E_{internal}$ for the grid points initially encourages symmetry along the axis. $E_{internal}$ for the control points includes the same terms as the equation for the linear models. Later, the pseudo-cylinders are encouraged to define side-chain direction.

After the deformable models and templates have

been used to resegment the EDM, the output can be fed back into the expert directed loop for evaluation, it can be used to initialize one of the energy minimizing programs that uses chemial constraints, or, the graph rewriting can be repeated to obtain much more reliable environments for use in a threading (sequence alignment) program [3].

4 Experiments and Results

At present, we are investigating a very coarse-grained combination of two processes which use either attributed graph or deformable template representations. In this section we describe preliminary experimental results that demonstrate how combining these processes can improve the initial segmentation and interpretation of a protein image. The experiments were designed to test the ability of our system to segment the image into complete, individual residues. The ability to determine side-chain orientation was also tested. We assessed these abilities by first deriving suitable subgraphs and deformable templates from a training set of protein structures retrieved from the Protein Databank of Brookhaven (PDB) [18]. These subgraphs and templates were used as models in a (prototype) system that was tested on various proteins.

We used a database of twelve structurally-diverse proteins (providing a total of 2136 residues) to derive the structural subgraphs and templates as follows. Using a crystallographic software package, XTAL [17], EDMs at medium (3 Å) resolution were generated from data in the PDB for each of the twelve proteins. Complete proteins were used to generate the EDMs for subgraph discovery. Topological analysis of each reconstructed map provided an initial segmentation for each protein. The set of subgraphs associated with each type of structure was compiled. Common subgraphs which reflected chemical bonds were used to generate models (see Fig. 3). The frequency of each variation from the model was used to estimate an error cost for correction.

Only the data for the proteins' backbone atoms were used to generate the EDMs used to recover deformable templates corresponding to backbone segments. The first atom of each side-chain was included in the EDMs used to recover templates with side-chain orientation. After the topological segmentation of these EDMs, knowledge of each atom's location was used to guide resegmentation of the EDM into residue-sized segments (see Fig. 5c). Values for length, volume, and geometric parameters were compiled. The distribution of these values was used to estimate an acceptable range, and a crude energy function for each parameter[3].

[3]Because the database used to generate our models is small, the models and their associated functions are quite simplistic. A more extensive database is available for future work.

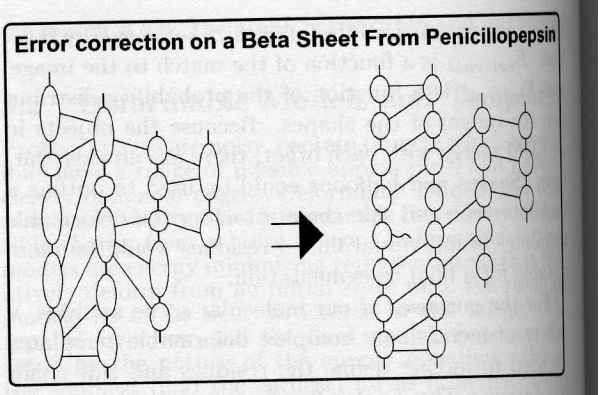

Figure 6: Results of Graph Rewriting Using Error-Correction: The initial segmentation of this beta sheet does not clearly show the 'ladder-like' subgraph associated with beta sheets. While identifying the backbone subgraph, several (five) large regions were split. With the new edges resulting from this resegmentation, the beta sheet subgraph was clear enough to guide further splits, merges, and split/merges. The right side of the diagram shows the final graph - the beta sheet structure is quite obvious at this point.

The system was tested on EDMs from three proteins not included in the database used to generate the models. The graph rewriting and the deformable template modules were applied to both ideal and experimental EDMs at 3 Å resolution. An ideal map for the protein *bovine phospholipase* (1BP2) was reconstructed from data in the PDB. Experimental data (reflections and structure factors) were available for two solved proteins, *penicillopepsin* (3APP) and a *fish antifreeze* (1MSI). EDMs for these two proteins were reconstructed from this data. Topological analysis followed by graph rewriting were used to segment and partially interpret each protein image. Several medium length (nine residue) backbone traces were chosen from the corrected graph for each protein. The traces included a variety of secondary structures; good and bad segmentations were represented. A deformable template for nine residues was initialized on each trace. For each protein image, the output of the system is an attributed graph and the residue models recovered using deformable templates.

We measure the success of the system using several criteria: the number of backbone segments correctly identified in the graph; the number of secondary structures correctly identified in the graph; the number of atoms correctly/incorrectly segmented by the deformable models; the angle between each actual side-chain and its corresponding modelled side-chain. Table 1 illustrates the results obtained by applying error-correcting graph rewriting to the segmentation ob-

Protein	#residues	Backbone Regions		Corrections		Secondary Structure Identified (Correct/Total)		
		# before	# after	#split	#merged	# Disulphide	# Helix	# Beta Sheet
1BP2	123	113	124	13	2	7/7	5/5	1/2
3APP	323	305	336	46	15	1/1	6/6	6/7
1MSI	65	53	60	10	3	-	-	-

Table 1: Results of Error-Correction with Graph Rewriting when applied to segmented EDM of proteins *Bovine Pancreas Phospholipase A2* (1BP2), *Penicillopepsin* (3APP) and a *fish antifreeze* (1MSI). The number of residues, the number of volumetric regions before and after correction, and the number of each type of region correction (splits and merges) are listed. Secondary structure identified (#correct/total) denotes the number of correctly identified disulphide bridges, helices and beta sheets (out of a possible total) that have been assigned the correct attributes when construction and interpretation is complete.

tained with topological analysis. 1BP2 was an ideal map. Error-correction resulted in one backbone segment too many. All disulphide bridges and helices were conclusively identified. Residues at the ends of the helices, and in one of the two beta sheets, were tentatively identified. The other beta sheet was not identified. In 3APP, error-correction resulted in thirteen backbone segments too many. The disulphide bridge and helices were conclusively identified. Six of the seven beta sheets were tentatively identified (see Fig. 6); the seventh was not identified. (The individual beta strands were quite short.) We believe that applying subgraph isomorphism after fitting the deformable templates will improve the identification rate for beta sheets. Results for 1MSI were not as good; there was no secondary structure to assist the resegmentation. In several instances three residues were represented by two segments, resulting in too few backbone segments. This adversely affected subsequent processing. Using deformable templates corresponding to two or three residues to guide error-correction might improve resegmentation of ambiguous regions.

Table 2 illustrates the results obtained when (nine residue) deformable templates are fitted to the resegmented images output by the graph rewriting module. Subgraphs of the attributed graph, comprised of nine backbone segments, were used to initialize the deformable templates. Ten of these subgraphs were resegmented: two were in helices, two were in beta sheets, and six were in loops. A single residue template (fitted to residue i) contains the carbon alpha and carboxyl atoms for residue i, and the nitrogen atom for residue i+1 (see Fig. 5c). The quality of inter-residue segmentation was evaluated as follows. After resegmentation, we determined how many backbone atoms were shifted into the correct/incorrect segment. We also determined how many backbone atoms remained in the incorrect segment. For residues with a side-chain, we determined the difference in orientation between the template's 'side-chain' attachment and the actual side-chain. 1BP2 was already well segmented; in the three sample subgraphs only five atoms were shifted into the correct segment. All other atoms had been correctly segmented. In 3APP, seventeen atoms in the four subgraphs were shifted correctly, three atoms were shifted to an incorrect segment, and one atom remained in an incorrect segment.

Less positive results were obtained for 1MSI. Inter-residue segmentation was improved for two of the backbone subgraphs. Ten atoms were corrected and one atom was incorrectly moved. The third contained two segments which were too large; the subgraph actually corresponded to ten residues. The deformable templates were not able to overcome the errors in initialization. Although three atoms (far from the over-sized segments) were correctly shifted, the status of many other atoms were ambiguous. Correct/incorrect inter-residue segmentation near the error was difficult to define. Nine atoms were clearly incorrect. Excess distortion in the fitted model could be used to identify such situations.

Except for the problematic subgraph in 1MSI, side-chain orientation was determined within 45^o for all side-chains with more than three atoms, except for proline[4]. Side-chains with three or fewer atoms were consistently located within 60^o of the correct orientation. The exceptions which occurred in 1MSI were due to the extreme distortion of the template.

5 Discussion and Conclusions

The initial results are promising; our approach can significantly improve the quality of the segmentation for parts of the EDM. The resulting models should be sufficiently accurate to generate input to a threading module. We are currently identifying suitable attributes to

[4] Because prolines bend back to the backbone, the side-chain is difficult to locate. It may be possible to recognize this structure in the graph rewriting module and create a different template for such structures.

Protein	Sec. Struct.	Atoms - Changes # corr.	# incorr.	# miss.	Side-chain Orientation $\leq 45^o$	$\leq 60^o$
1BP2	Helix	1	0	0	8/8	0/8
	Loop	2	0	0	7/9	1/9
	Loop	2	0	0	7/8	1/8
3APP	Helix	3	1	1	7/8	0/8
	Sheet	4	1	0	5/6	1/6
	Sheet	5	0	0	9/9	0/9
	Loop	5	1	0	7/9	1/9
1MSI	Loop	7	1	0	8/9	0/9
	Loop	3	0	0	7/8	0/8
	Loop	3	NA	9	2/9	4/9

Table 2: Results of resegmenting the image using deformable templates. For each protein, three or four traces (representing a variety of secondary structures) comprising nine regions each, were evaluated. The number of atoms that were associated with a different region after resegmentation are listed: (#correct) were shifted to the correct region, (#incorrect) were shifted from the correct region to a different region. (#missed) is the number of atoms that should have been shifted to a different region, and weren't. Side-chain orientation measures the angle between the model's side-chain "attachment" and the first two/three atoms of the actual side-chain. No results are listed for glycine.

be used for threading [3].

Automatically identifying incorrect segmentation by the graph rewriting module is a problem that will need to be addressed. One approach involving a more fine-grained iteration, where smaller groups of deformable templates are used to guide the splitting/merging process, is being implemented. Deformable templates for each side-chain rotamer are also being generated.

More data is needed to fine tune the parameters used in energy minimization; future work also includes finding more sophisticated energy functions and more robust energy minimization methods.

References

[1] S.Fortier,I.Castleden,J.Glasgow,D.Conklin, C.Walmsley, L.Leherte, and F.H.Allen. Molecular Scene Analysis: The integration of direct methods and artificial intelligence strategies for solving protein crystal structures. Acta Crystallographica, D49:168-178, 1993

[2] L.Leherte,J.Glasgow,K.Baxter, E.Steeg, and S.Fortier. Analysis of three-dimensional protein images. Journal of Artificial Intelligence Research (JAIR), pp. 125-159, 1997

[3] K.Baxter,E.Steeg,R.Lathrop,J.Glasgow,S.Fortier. From Electron Density and Sequence to Structure: Integrating Protein Image Analysis and Threading for Structure Determination. In Proc. of the 4th Int. Conf.on Intelligent Systems for Molecular Biology (ISMB96). AAAI/MIT Press, 1996

[4] T.A.Jones, M.Kjeldgaard. Electron-Density Map Interpretation. Methods in Enzymology, Vol.277. Academic Press, 1997, pp. 173-208

[5] G.J.Kleywegt, T.A.Jones. Efficient Rebuilding of Protein Structures. Acta Cryst. D. 1998. in press.

[6] G.J.Kleywegt, T.A.Jones. Taking the fun out of map interpretation. News from the Uppsala Software Factory, 7. http://alpha2.bmc.uu.se/ gerard/manuals/ , last updated Feb. 1998.

[7] G.Stout, L.Hensen. X-Ray Structure Determination. Wiley, New York, 1989

[8] K.Baxter. Protein Structure Determination (Improving Analysis of Crystallographic Data Using Inexact Graph Matching). Poster session, 6th Int. Conf. on Intelligent Systems for Molecular Biology (ISMB98). AAAI/MIT Press, 1998

[9] J.Greer, Three-dimensional Pattern Recognition: An Approach to Automated Interpretation of Electron Density Maps of Proteins. J.Mol.Biol. 82. 1974, p.279-301

[10] C. K. Johnson. ORCRIT. The Oak Ridge Critical Point Network Program. Chemistry Division, Oak Ridge National Laboratory, USA, 1977.

[11] B.T.Messmer, H.Bunke. Subgraph isomorphism detection in polynomial time on preprocessed model graphs. Proc. of the Asian Conference on Computer Vision ACCV, pg. 151-155, 1995

[12] B.T.Messmer, H.Bunke. Fast error-correcting graph isomorphism based on model procompilation. Tech Report IAM-96-012 University of Bern, 1996

[13] B.T.Messmer, Efficient Graph Matching Algoritms for Preprocessed Model Graphs. PhD Thesis 1995, University of Bern.

[14] M.Kass, A.Whitten, D.Terzopoulos. Snakes: Active Contour Models. Int.J.Comp.Vision. 1988, pp. 321-331

[15] L.D.Cohen, I.Cohen. Finite-Element Methods for Active Contour Models and Balloons for 2-D and 3-D Images. IEEE PAMI. Vol.15.No.11. Nov 1993, pp.1131-1147

[16] T.McInerney, D.Terzopoulos. Deformable Models in Medical Image Analysis: A Survey. , Medical Image Analysis, 1(2), 1996

[17] S. R. Hall and J. M. Stewart, eds. XTAL 3.0 User's Manual. Universities of Western Australia and Maryland, 1990.

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