Finding α-helices in skeletons

Nina Amenta Sunghee Choi Maria E. Jump
Ravi Krishna Kolluri
Thomas Wahl

Technical Report TR-02-27
Computer Sciences Department
University of Texas at Austin †
Keywords: Crystallography, Computational Geometry, Object Recognition

October 16, 2002

Abstract

We consider a problem which is part of the process of determining the three-dimensional structure of a protein molecule using X-ray crystallography: given an estimated map of the electron density of the molecule as a function on three-dimensional space, we identify regions which are likely to belong to α-helices. Our approach is to compute a new kind of skeleton - the *power shape* - and then identify the helical substructures within the power shape with a variant of geometric hashing.

1 Introduction

X-ray crystallography is one of the main techniques for determining three-dimensional protein structure. Experimental diffraction data provides the amplitudes of some of the Fourier coefficients of a three-dimensional map of electron density in a crystal of the protein. The phases of the Fourier coefficients are estimated using a variety of experimental and computational techniques. When there is high-resolution diffraction data and the phases are well-estimated, individual atoms are visible in the electron density map and determining the three-dimensional structure is easy. Often, however, only a noisy low-resolution map is available.

† Computer Sciences Dept., Austin, TX 78712, USA. Fax: 512-471-8885. Supported by NSF/CCR-9731977 and an Alfred P. Sloan Foundation Research Fellowship. amenta@cs.utexas.edu

At this point, a chemist will spend days or weeks at a computer graphics terminal, manually aligning a stick-figure molecular model containing thousands of atoms to the density map. Most of the really time consuming, difficult but decipherable, maps are at between 3 and 4 Åresolution. Finding secondary structures, especially the α-helices, is one of the first steps a human expert takes when aligning the model with the map, and hence it is one of the first steps we should attempt to automate.

Our work: Given a density map represented by a three-dimensional grid of function values as input, we compute an isosurface. We then compute a skeletal representation of the solid bounded by the isosurface, known as the *power shape*, composed of triangles. For each triangle in the power shape, we examine a set S' of nearby power shape vertices and find the helix that best agrees with S' by geometric hashing. If there is sufficient agreement, we report the points as part of a helix. Since there is a direct mapping between the power shape and the isosurface, this corresponds to labeling a section of isosurface as belonging to a helix, as in Figure 1.

We have tested the method successfully on two density maps, one at 3.0 Angstroms and the other at 3.5 Angstroms. At these resolutions α–helices are visible as twisted shapes in the isosurface. See Figure 1.

Importance of the problem: There are at least three ways in which automatically locating α-helices can be useful. First, it can be used as a domain-specific visualization tool. Highlighting helical portions of the isosurface can make things easier for the chemist during manual model building. Second, finding helices is used as part of a density map refinement algorithm. Information about the three-dimensional structure of the molecule is used to improve the estimated phases, thus improving the quality of the map itself. Often reconstruction is an iterative process in which model building alternates with phase improvement. This would be most useful for noisier, lower resolutions maps than those we have considered so far, but our technique might be applicable. Finally, it might be possible to combine automatic geometric interpretation of the density map with AI methods for predicting secondary structure from sequence data to automatically form tentative matches of portions of sequence data to the map.

2 Related work

There is an excellent existing tool for finding structural fragments such as α–helices in electron density maps. The most recent version of Kevin Cowtan's `fffear` program [19] can find helices in *very* low quality low resolution maps (6-8 Angstroms, larger than a single turn of a helix). It searches a discrete set of possible orientations of the fragment. For each orientation, it convolves the map with a filter resembling the fragment, by multiplication in the frequency domain. This is quite efficient, and independent of the fragment size. It takes advantage of the fact that the frequency domain representation is