Use of wavelets to compress three-dimensional protein maps

Eleanor J. Gardiner*[1,2], P. Willett[2] and Peter J. Artymiuk[1]

[1] Department of Molecular Biology and Biotechnology, University of Sheffield
[2] Department of Information Studies, University of Sheffield

1 Introduction

Many properties of a protein, ranging from theoretically calculated molecular electrostatic potential fields to experimental electron density, may be represented by three-dimensional maps. Theoretical maps often require considerable computing time to calculate and therefore it is highly desirable to be able to store them rather than recalculate them. Experimental maps also require storage: the increasingly important structural techniques of cryo-electron microscopy often have insufficient resolution to permit the construction of atomic models, but instead result in low resolution density maps of important biological macromolecules. It is now recognized that it is highly desirable to able to deposit and store these 3-dimensional maps (Fuller, 2003) so that they may be available to other investigators. However because proteins are large and the maps are often relatively finely sampled, a common characteristic of these maps is that they require large amounts of computer storage.

We have examined the use of wavelets for the cost-effective storage of such data (Gardiner *et al*, in preparation). Wavelets are a mathematical tool for hierarchically decomposing functions. They describe a function at a coarse level of overall shape along with a series of details ranging from fine to coarse. By applying an inverse wavelet transform, the original function can be recovered. In two dimensions, wavelets are widely used in image compression including the JPEG2000 standard.

Wavelets have been used in many applications in bioinformatics (for a review see Lio (2003)). However, to our knowledge, only one attempt has been made to represent the 3-dimensional structure of macromolecules using wavelets (Carson, 1996). Carson used B-spline wavelets to represent the protein backbone, and his aim was protein structure visualization rather than compression. He therefore concentrated on the multiresolution aspect of the wavelet representation, which allows the protein fold to be viewed in greater or lesser detail. Additionally, protein surfaces were modelled using NURBS (Non-Uniform Rational B-spline Surfaces) and these were also the subject of a multiresolution analysis. Wavelets have also been considered as a representation for the electron density maps of small molecules (Leherte, 2001).

2 Results

In this work, we represent a variety of different 3-dimensional protein maps using a reduced number of wavelet coefficients generated by repeatedly applying a 1D wavelet transform sequentially over all 3 dimensions (Press et al., 1992) in a standard wavelet decomposition. These maps include experimental electron density, GRASP electrostatic potential maps (Nicholls *et al.*, 1993), and GRID maps (Goodford, 1985) of ligand binding propensity. We analyse the performance of several different wavelets, (including Haar, Debauchies, coiflet symmlet wavelets) for map compression. We find that using only 10% of coefficients, all wavelet types perform

very well for all maps, in that the original map can be recovered with negligible loss of data as measured by standard metrics such as a low root mean square deviation (RMSD) of the reconstructed from the original signal, and high peak signal-to-noise ratio (= 20log10(b/RMSD) where b is the maximum possible signal value.

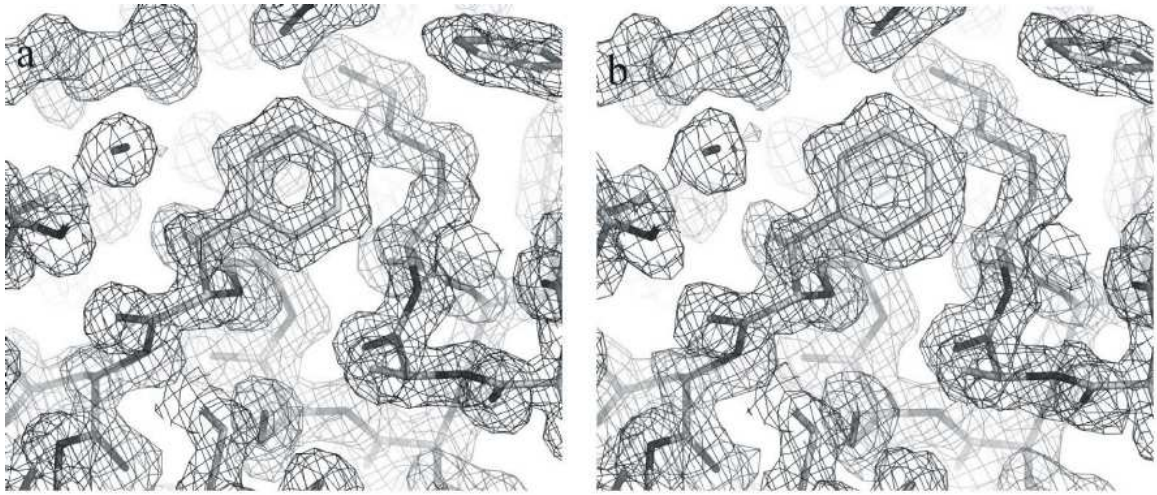

Figure 1 Electron density map. (a) shows a part of the original map, and (b) shows a wavelet reconstruction (8% coefficients, Coiflets 18-tap).

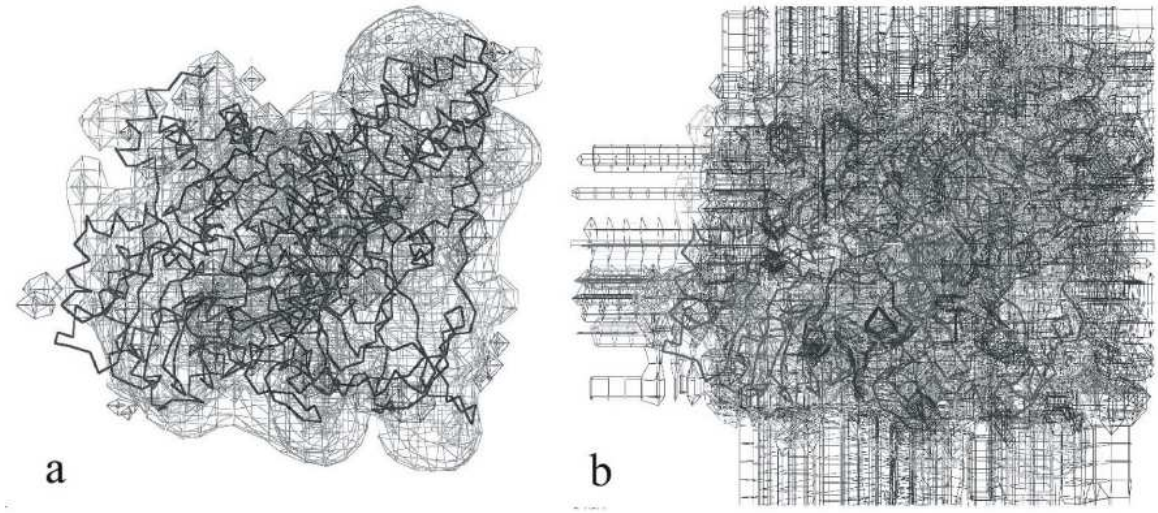

Figure 2. Grasp potential map (Nicholls et al, 1993) of aconitase. **(a)** shows the original map; reconstructions using 4% of coefficients are essentially identical. **(b)** shows a reconstruction of the same Grasp map, using only 1% coefficients (Haar wavelet) which shows very poor reconstruction with obvious artefacts.

This can be seen in Figure 1(a), which shows a portion of an experimental X-ray crystallographic electron density map of lysozyme, and Figure 1(b) which shows the same map reconstituted from 8% of the wavelet coeffi cients. As can be seen, the reconstruction retains all the essential details of the original map to a very acceptable degree.

However, using 1% of coeffi cients, we fi nd that most wavelet representations deteriorate very badly. Thus Figure 2(a) shows the GRASP potential map of an enzyme (aconitase B), whilst Figure 2(b) shows the map reconstituted from 1% of the Haar wavelet coeffi cients. It is immediately apparent that the reconstruction is extremely poor with obvious artefacts. In this

and other examples, we find that when the number of coefficients used is within the 1-10% range, performance is map- and wavelet-type dependent.

We conclude that wavelets provide a cost-effective method for the compression of a variety of protein maps.

Acknowledgements

This work was supported by a BBSRC project grant. Computational facilities were provided by the BBSRC.

References

Carson M. (1996). Wavelets and molecular structure. *Journal of Computer-Aided Molecular Design*, **10**, 273-283.

Fuller, S.D. (2003). Depositing electron microscopy maps. *Structure*, **11**, 11-12.

Goodford P.J. (1985). A computational procedure for determining energetically favourable binding sites on biologically important biomolecules. *Journal of Medicinal Chemistry*, **28**, 849-857.

Leherte. (2001). Application of multiresolution analyses to electron density maps of small molecules: Critical point representations for molecular superposition. *Journal of Mathematical Chemistry*, **29**, 47-83.

Lio, P. (2003). Wavelets in bioinformatics and computational biology: state of art and perspectives. *Bioinformatics*, **19**, 2-9.

Nicholls A. Bharadwaj, R. and Honig, B. (1993). GRASP- Graphical Representation and Analysis of Surface Properties. *Biophysical Journal*, **64**, 166.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). *Numerical Recipes in Fortran*. New York. Cambridge University Press.