Use of wavelets to compress three-dimensional protein maps

Eleanor Gardiner[1] , Valerie Gillet[1], Richard Martin[1], Stefan Senger[3] and Peter Artymiuk[2],

[1] Department of Information Studies, University of Sheffield, Sheffield, S1 4DP, UK.
[2] Department of Molecule Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK.
{e.gardiner,p.artymiuk }@sheffield.ac.uk
[3] GlaxoSmithKline Medicines Research Centre, Computational and Structural Chemistry, Stevenage, Herts. SG1 2NY, UK.

Abstract. Increasing amounts of biological data are being created and stored in the form of three-dimensional (3D) maps. Examples range from 3D EM reconstructions to calculated electrostatic potential maps. Such maps are extremely demanding of storage space. We have used a 3D discrete wavelet transform to compress examples of several different types of 3D macromolecular maps. We show that reconstruction using only 2% of the wavelet coefficients can reliably reproduce the original data. We compare the performance of a number of different wavelets using both standard and non-standard wavelet transforms and show that the non-standard transform outperforms the standard in terms of reconstruction quality. Whilst no wavelet is best overall, we show that wavelets with long support are to be preferred for EM reconstructions whilst short support wavelets perform better for calculated maps such as ligand-binding propensity maps. These findings are confirmed for ligand-binding propensity maps of small molecules.

Keywords: Discrete Wavelet Transform, GRID maps, 3D macromolecular maps, data compression.

1 Introduction

There are many bodies of biological and medical imaging data that now take the form of very large 3D grid maps in which some property of the system is sampled discretely. These may consist of experimental data from light or electron microscopy (EM), X-ray crystallographic electron density maps, or maps of molecular properties derived by calculation from atomic coordinates. In addition to this an increasingly important area is medical tomography. In all these cases the number and large size of the maps means that some form of efficient archival and storage is essential.

In the case of three-dimensional maps from electron tomography or optical microscopy [1] or from cryo-electron microscopy [2] the need to make public what is effectively the final result of the experiment prior to the authors' interpretative steps has been recognized for some time. The EMDEP database [3] has been established for the deposition of medium and high resolution cryo-electron microscopy maps. The

Cell Centred Database (CCDB) [1] has been established for the archival of mesoscale cellular data from electron microscopy and optical microscopy. For X-ray crystallographic structures the conventional route of data deposition has been the deposition of atomic coordinates in the RCSB Protein Data Bank (PDB) [4]. The value of archiving crystallographic electron density maps has also been recognized as they immediately give an impression of the quality of the information that underlies a particular atomic model, especially to non-crystallographers, and would allow re-analysis to enable checking of the suggested structure. This has been recognized by the Uppsala Electron Density Server [5] which allows downloads of available electron density maps.

In addition to such experimental maps, theoretically calculated maps have a particular requirement for efficient storage as they may be computationally very expensive to calculate, or may have been calculated using software that is not generally available to other researchers for commercial or operational reasons. In these cases it is important to make the calculated three-dimensional maps available in order to underpin published papers that make use of such information. Examples of such calculated three-dimensional maps include electrostatic potential maps [6] and ligand binding propensity maps [7]. All these maps, whether experimental or derived, are larger than the corresponding coordinate files, sampling a large structure typically on a 1Å or 0.5 Å grid, and therefore can be extremely demanding of storage space. There is therefore an urgent need for methods for compressing such maps, ideally with the capacity to search their content without complete reconstruction or decompression.

1.1 Wavelets

Wavelet analysis is a multi-resolution analysis technique pioneered in the 1980's by Mallat and by Daubechies [8, 9]. A wavelet transform is a hierarchical method for decomposing a function so that it is represented by a collection of coefficients, each of which provides some limited information about the position and frequency of the function. In addition to their hierarchical character, wavelet transforms have the following useful properties: transforming to and from a wavelet representation can be achieved in linear time; for functions typically seen in practice, many of the wavelet coefficients are either zero or negligibly small so that impressive data compression rates can be achieved; wavelets can be adapted to represent a wide variety of functions, including functions with discontinuities and functions defined on bounded domains. A wavelet transform can be seen as analogous to a Fourier transform: in Fourier space a function is represented as a summation of the basis functions, sines and cosines; in wavelet space the bases are called scaling functions and wavelets [10, 11]. Several families of wavelets, including Daubechies, Symlet and Coiflet wavelets, were discovered by Daubechies.

The first use of wavelets for 3D data appears to be due to Muraki [12]. Data from a series of slices of MR images were decomposed using a 3D wavelet transform and then reconstituted using the inverse wavelet transform. He observed that reconstructions using fewer wavelet coefficients could acceptably reproduce the

original slices. Muraki used Battle-LeMarie wavelets [8] but concluded that they were not the best for this application.

Wavelets have been considered for many applications in bioinformatics [13]. One of the first and most interesting is the use by Carson of B-spline wavelets to represent the protein backbone [14]. The aim of Carson's work was protein structure visualization rather than data 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. To our knowledge, only one previous attempt has been made to represent the 3D structure of macromolecules using wavelets [15]. This was aimed at a molecular graphics application and used a Hierarchical Basis (HB) wavelet compression method for fast compression/decompression and interactive visualization of macromolecular maps.

In this paper we compare the use of standard wavelet transforms with non-standard wavelet transforms which combine efficient compression with the capacity to permit partial reconstruction of small 3D thumbnails of the whole map at variable levels of resolution. We apply these techniques to several different types of 3D macromolecular maps.

2 Methods

2.1 Discrete Wavelet Transform

A DWT is applied to a function, which has been represented by a vector of n values, where n is a power of 2. The result, at level 1, is n/2 'smooth' or 'averaged' values and n/2 'detail' or 'difference' values [10]. The detail values are retained and the DWT then applied to the smooth values to give n/4 smooth-smooth values and a further n/4 detail values at level 2. This process is iterated, always applied only to the remaining smooth…smooth values, until only a trivial number of smooth…smooth coefficients remain. At this stage there are still n values. However, many of these values (wavelet coefficients, denoted *wcs*) will be negligibly small and, for data compression, may be replaced by zeroes. The original function may be recreated by applying the inverse wavelet transform to either the unchanged or zero-replaced wcs.

The support of a wavelet is the region of the wavelet domain upon which the wavelet is non-zero. If this region is finite the wavelet is said to be compactly supported. Smooth functions are normally better represented by smooth wavelets which tend to have longer support. However wavelets with short support can be more efficiently used in decomposition and reconstruction [11]. There can thus be a trade-off between efficiency and effectiveness in the choice of wavelet support length. The number of coefficients of the scaling function is called the *tap*. In the case of the DWT, the short support wavelets have the smallest tap values.

2.2 2D and 3D Wavelet Transforms

In two dimensions a function, for example an image, is represented by a n×m matrix of values (where n, m are both powers of 2). The most straightforward extension of the one-dimensional DWT to two dimensions is then first to apply a 1D DWT to all rows. The output is a n×m matrix where columns 1 to m/2 are smooth coefficients and columns m/2 +1 to m are detail coefficients. Then the 1D DWT is applied to all rows of this matrix, and, as in the 1D case, this application of the DWT is iterated until finished. Thus the DWT is applied over all rows for all levels of the transform. Then, in a similar fashion, the DWT is repeatedly applied to the columns of this row-transformed matrix. This two-dimensional DWT is known as a *standard* or *linear decomposition* [11] and in effect treats the n×m matrix as a 1D vector of n×m values. The alternative *non-standard* method is more inherently two dimensional and is accomplished by alternately applying a 1D transform to the rows of the matrix and then to the columns of the transformed matrix.

An advantage of the standard decomposition is that it is readily applicable to the complete decomposition of rectangular matrices whereas the non-standard decomposition only decomposes to the level allowed by the smallest dimension. However this slight disadvantage is out-weighed in some circumstances, as we will show here, by the added interpretability of the non-standard decomposition. Since the entire transform is completed at every level, the smooth wavelet coefficients at a particular level represent a smaller (and smoother) transformed version of the original image, which can still be recognisable as a 'thumbnail' image. Thus a partial reconstruction to any particular level of the transform using the inverse DWT gives a meaningful smooth subimage. In the standard decomposition a partial reconstruction is not meaningful since all the transformed columns have to be recovered before the recovery of the rows begins. This will be illustrated in the 3D case below.

A 3D wavelet decomposition is a direct extension of the 2D case. The function, for example an electron density map, is represented by a 3D matrix of values. The standard decomposition then first applies the 1D DWT to all rows for all levels of the transform, followed by the application of the transform to all columns for all levels, and finally, application of the transform to all depths for all levels of the transform. In contrast, the non-standard decomposition, at each level of the transform, applies the 1D DWT to all rows, then to all columns, then to all depths. We illustrate the non-standard 3D wavelet transform using the calculated electron density map of a hand [2]. The original hand has dimensions 98 × 98 × 98 voxels and so is placed in a cube of size 128 × 128 × 128 voxels. Figure 1 shows the original hand and its partial and full decompositions using both standard and non-standard wavelet transforms with the Daub4 wavelet. The partial non-standard transform gives a clear image of a smoothed hand (arrowed) whereas the partial standard transform has no obvious meaning.

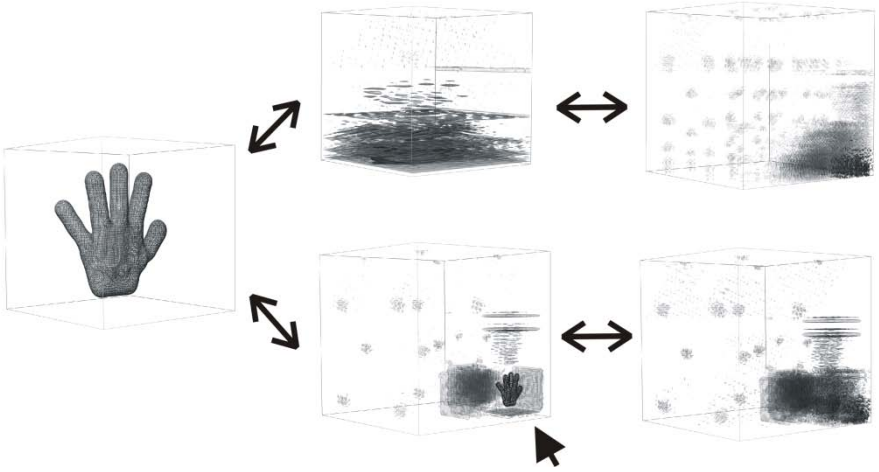

Fig. 1. Standard and non-standard wavelet decomposition of a hand. *Top* - standard decomposition after transform of all rows and final decomposition. *Bottom* – non-standard decomposition after transforming the first two levels (to a 32 × 32 × 32 grid) and final decomposition. The arrow indicates the "thumbnail" image present in the intermediate step of the non-standard decomposition.

2.3 Measures of reconstruction quality

We test the ability of the wavelet transform to compress a protein map by taking the 3D DWT of the map, replacing some number of the smallest wcs by zeroes and then applying the inverse transform to reconstruct the original map. N.B. Since reconstruction using the inverse wavelet transform requires that the position in the matrix as well as the value of the non-zero wcs is stored, this naïve approach means that, for example, using 10% of the wcs takes 20% of the space required by the original map, giving 80% compression. However there are many sophisticated indexing schemes which could be used to achieve the higher compression ratios in practice. Here we focus on the use of wavelets for compression and disregard the storage problem.

There are numerous ways to evaluate the quality of a reconstruction. We use the peak signal-to-noise ratio (PSNR) and the real space R factor. The PSNR is defined as

$$\mathrm{PSNR} = 20 * \log10(b/\mathrm{rmsd}) \quad (1)$$

where b is the largest possible value of the signal and rmsd is the root mean square difference between two images (or volumes) [43].

PSNR is measured in decibels (dB). An increase of 20 dB corresponds to a ten-fold decrease in the rmsd between two images. Perfect reconstruction would give an infinite value for PSNR.

We have also used a real space R-factor [16] which is defined as

$$R = \Sigma \mid \rho(obs) - \rho(calc) \mid / \Sigma \mid \rho(obs) + \rho(calc) . \quad (2)$$

In the case of evaluating the fit of a model to a map: (obs) is the observed electron density, (calc) is that calculated from a model, and the summation is over all non-zero elements in (calc), often calculated over individual residues. In that application an R of < 0.2 is considered indicative of a good fit. Perfect reconstruction would give a zero value for R. The reconstruction of the electron density map of lysozyme shown in Figure 2B, created using Daub4 wavelet and 10% of the wcs, has a PSNR of 32.66 and an R factor of 0.179. (N.B. We have applied the PSNR and R factor calculations to just the non-zero values in the original maps.)

3 Results and Discussion

We have applied both the standard and non-standard 3D DWT to various Grasp[6], GRID[7], electron density and EM macromolecular maps of several different proteins, viz aconitase, lysozyme, putrescine binding protein, liver fab protein (PDB codes 1l5j, 3lzt, 1poy, 1lfo), a virus and a ribosome (EMPDEP codes 1075 and 1064). Each of the maps was first transformed into a text file. Each map was placed in the smallest 3D box where each dimension is a power of 2. So, for example, 1075_em, size 175 × 175 × 240, was placed in a box of size 256 × 256 × 256. We performed the transforms using a variety of wavelet types and tap lengths. In particular, we used: Daubechies (Daub) 4-, 12-, 20-tap wavelets; Coiflet (Coif) 6-, 12-, 18-tap wavelets; Symlet (Sym) 6-, 8-, 10-, 12-, 14-, 16-tap wavelets; and Haar wavelets. Tables 1 and 2 give the PSNR and R values for the best-performing wavelet for the standard and non-standard transforms using 10% and 1% of wcs respectively. Table 3 gives the PSNR and R values for the worst-performing wavelet for the standard and non-standard transforms using 10% of wcs.

3.1 Standard Reconstruction

In all cases, maps reconstructed with 10% of coefficients are visually almost identical to the original. In Figure 2 we show the reconstructed electron density map, 3LZT_xtal, together with the original. Fine details such as the hole in the density in the tyrosine ring are preserved. Using 2% of coefficients all GRID maps are well-reproduced using the non-standard Haar wavelet. Using only 1% of the coefficients, the results are more mixed, with the GRID and electron density maps being poorly reconstructed and showing many artefacts. However for the EM maps in particular, the reconstruction using 1% of coefficients is extremely good, and indeed, much higher levels of compression can be achieved.

There is a noticeable difference in quality between the best and worst reconstructions for some maps, e.g. the standard decomposition of the GRID, C3 probe, maps of 1POY using 10% of wcs (Tables 1 and 3). In these cases the PSNR value of the worst transform (Daub20) is only about half that of the best (Haar).

Table 1. Best Compression using 10% of coefficients. When the PSNR and R rankings were not in agreement, the highest PSNR value was deemed to belong to the best-performing wavelet. The 4-digit code is the PDB code or EMDEP code. [a] GRID maps, _C3 is the sp3 probe, _N1 the N1+ probe, _DRY the dry probe, _OH2 the water probe. Italic type indicates the better performing compression type (standard or non-standard).

	Standard				Non-Standard			
Mol Code	Wavelet	Tap	PSNR	R	Wavelet	Tap	PSNR	R
1L5J_grasp	Haar	-	73.12	0.006	*Symlet*	*8*	75.26	0.004
3LZT_xtal	*Daub*	*20*	35.38	0.13	Daub	20	35.59	0.128
3LZT_grasp	Daub	4	83.79	0.012	*Daub*	*4*	85.29	0.010
[a]1POY_C3	Haar	2	66.66	0.001	*Haar*	*2*	Inf	0.0
[a] 1POY_N1	Haar	2	68.15	0.002	*Haar*	*2*	189.7	0.0
[a] 1LFO_DRY	Haar	2	Inf	0	Haar	2	Inf	0
[a] 1LFO_OH2	Haar	2	35.84	0.052	*Haar*	*2*	59.43	0.002
1075_em	Symlet	16	43.54	0.018	*Symlet*	*16*	46.33	0.013
1064_em	Symlet	16	67.65	0.003	*Symlet*	*16*	67.77	0.003
Hand_em	Symlet	16	77.23	0.002	*Symlet*	*16*	80.95	0.001

Table 2. Best Compression using 1% of coefficients.

	Standard				Non-Standard			
Mol Code	Wavelet	Tap	PSNR	R	Wavelet	Tap	PSNR	R
1L5J_grasp	Coiflet	6	41.81	0.096	*Coiflet*	*6*	49.27	0.039
3LZT_xtal	Coiflet	18	25.72	0.449	*Coiflet*	*18*	26.11	0.428
3LZT_grasp	Haar	2	52.08	0.207	*Haar*	*2*	65.90	0.075
1POY_C3	Haar	2	20.21	0.121	*Haar*	*2*	40.67	0.010
1POY_N1	Symlet	6	28.12	0.154	*Haar*	*2*	43.90	0.026
1LFO_DRY	Haar	2	43.11	0.61	*Haar*	*2*	55.15	0.192
1LFO_OH2	Daub	4	21.40	0.258	*Haar*	*2*	24.03	0.144
1075_em	Symlet	16	33.31	0.060	*Symlet*	*16*	34.77	0.051
1064_em	Symlet	16	40.39	0.069	*Symlet*	*16*	42.53	0.051
Hand_em	Coiflet	18	56.37	0.019	*Symlet*	*14*	56.62	0.020

Table 3. Worst Compression using 10% of coefficients.

	Standard				Non-Standard			
Mol Code	Wavelet	Tap	PSNR	R	Wavelet	Tap	PSNR	R
1L5J_grasp	Daub	20	53.71	0.040	Haar	2	70.56	0.007
3LZT_xtal	Haar	2	30.62	0.230	Haar	2	31.85	0.197
3LZT_grasp	Daub	20	65.16	0.101	Daub	20	82.36	0.017
1POY_C3	Daub	20	31.66	0.040	Daub	20	88.07	0.000
1POY_N1	Daub	20	39.66	0.045	Daub	20	79.20	0.000
1LFO_DRY	Daub	20	48.92	0.530	Daub	20	60.01	0.236
1LFO_OH2	Daub	20	28.68	0.123	Daub	20	35.30	0.059
1075_em	Haar	2	40.75	0.026	Haar	2	43.13	0.019
1064_em	Haar	2	47.69	0.025	Haar	2	46.72	0.025
Hand_em	Haar	2	64.69	0.009	Haar	2	64.79	0.009

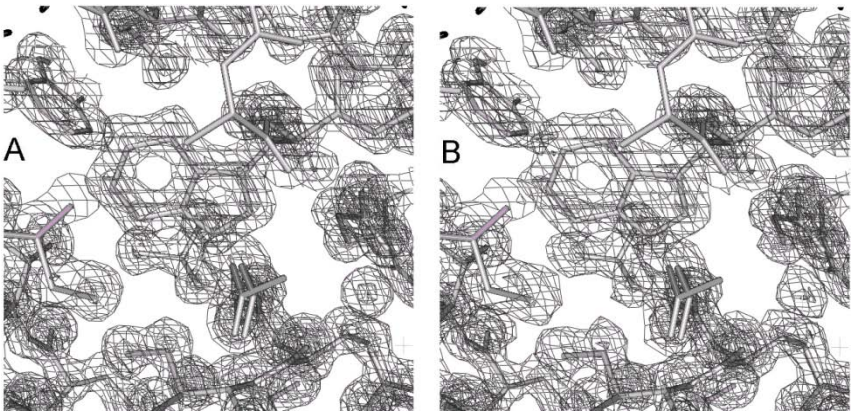

Fig. 2. Portion of the electron density map of lysozyme. A Original. B Reconstructed with 10% wavelet coefficients using the Daub4 standard wavelet transform.

In Figure 3 we present results for EM maps of the two large macromolecules. Panels A-D show the ribosome (1064_em), reconstructed using 0.1% and 0.5% of coefficients. The 0.1% reconstruction (B) uses only 2098 wcs, compared with 1.7M grid points in the original. This degree of compression gives a clearly noisy reconstruction. However the 0.5% reconstruction (from 10486 wcs) gives a very good result (C-D). Figure 3, panels E-F shows the bacteriophage T4 head (1075_em) reconstructed using only 0.5% coefficients after a standard Daub4 wavelet transform.

The original map contains approximately 7.3 million grid points and, at this level of compression, the reconstruction uses only 84,000 wcs.

3.2 Wavelet Performance

Whilst it is difficult to extrapolate from a few examples, it seems clear that, depending on the map type, either short or long support is to be preferred. There are no cases where Daub12, Symlet12 or Coiflet10 wavelets perform best. In the cases considered here, wavelets with short support perform best for GRID and Grasp maps whilst long support is better for electron density and EM maps. Both the electron density and EM maps are composed of Fourier coefficients and are inherently smooth. Thus the better performance of the longer support (and smoother) wavelets such as the Daub20 in these cases is not surprising.

In the discussion so far, we have concentrated on the standard wavelet decomposition. A comparison of the PSNR values of the non-standard decomposition with those of the standard decomposition (Tables 1 and 2) reveals that, whilst in many cases the results are very similar, overall the non-standard decomposition gives higher values for the PSNR (and correspondingly lower values for the R-factor). This is confirmed by a paired Wilcoxon test on the PSNR values giving p-values of less than 0.004 for both tables. The superior performance of the non-standard transform is especially true for the GRID maps where a standard compression gives unacceptable reconstruction at 1% for all probe types, whereas the non-standard performs well except for the water probe. The reason for the better performance of the non-standard transform is presumably that in transforming entirely in a single dimension at a time, correlations between values which are adjacent in space but not in the same dimension (i.e. not both in the same row or same column or same depth) are lost. In contrast the non-standard transform maintains correlations in all three dimensions

3.3 Small molecule examples

Although this paper is largely concerned with the compression of macromolecular maps we are also interested in the potential for compressing Goodford GRID maps of small molecules (ligands). Whilst these maps are naturally smaller than protein GRID maps it is usual to calculate them at higher resolution and potentially for much larger numbers of molecules than would be the case for proteins and compression is therefore still a worthwhile exercise. Additionally ligands are generally much more flexible than proteins. Since values in the GRID maps are dependent on the molecular conformation, a single GRID map of a ligand is not likely to capture all the available information and, once multiple maps are required, the need for compression is again strong. We have used two small molecules, PTI taken from its complex with oestrogen receptor 1 (PDB code 1uom) and DMP323 taken from its complex with HIV-1 protease (PDB code 1bve). For each we generated GRID maps using three different probes, OH2, N1+ and O- at 0.5Å resolution and performed compression experiments as described above.

The results were very similar to those of the protein GRID map experiments. In all cases the short support non-standard wavelet transform showed the least error in compression, with the Daub4 generally being marginally the best performer. Using 10% of coefficients the mean PSNR for the non-standard Daub4 transforms was 70.8dB (standard deviation 4.42) and visually the reconstructed maps are very similar to the original (Figure 4).

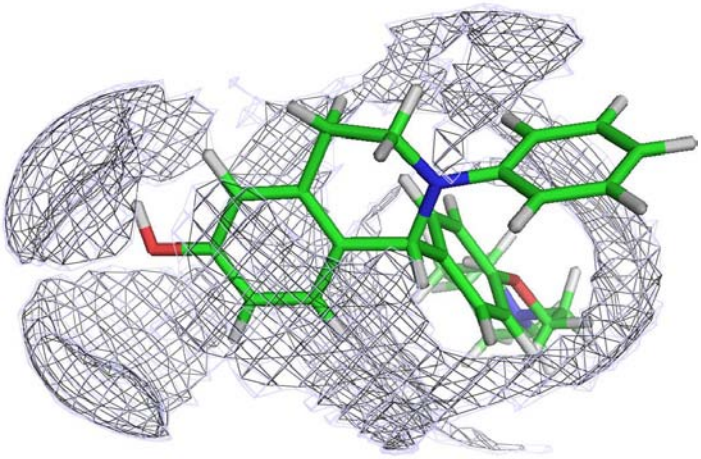

Fig. 4. Dark grey original. Light grey reconstructed with 10% wavelet coefficients using Daub4 non-standard transform.

4 Conclusions

We have shown that a 3D discrete wavelet transform can be used to compress examples of several different types of 3D macromolecular maps, and showed that reconstruction using only 2% of the wavelet coefficients can reliably reproduce the original data. This is valuable because such 3D maps require very large amounts of storage. There is increasing interest in the use of electrostatic potential or GRID maps for applications such as docking [17]. The ability to store large numbers of these maps in compressed form is therefore a useful and timely development. Using 2% of coefficients, all GRID maps (except for 1LFO using the water probe) can be reconstructed with a PSNR > 55 dB using a non-standard wavelet transform with the Haar wavelet. Similar results were found for the two small molecule GRID maps.

We also demonstrated that the non-standard transform performed better than the standard in terms of reconstruction quality. Whilst there is no clear winner in terms of wavelet type, we showed that wavelets with long support are to be preferred for electron density and electron micrographs whilst short support wavelets perform better for calculated maps such as electrostatic potential and ligand-binding propensity maps. Moreover the use of non-standard wavelet transforms also opens out new possibilities for the rapid interrogation of databases of 3D maps. This is because

the thumbnail representation generated on partial reconstruction can be regarded as a "mini" version of the full 3D image and can therefore be compared rapidly with other images at equally low resolution before proceeding to higher levels of reconstruction for more detailed analysis in appropriate cases.

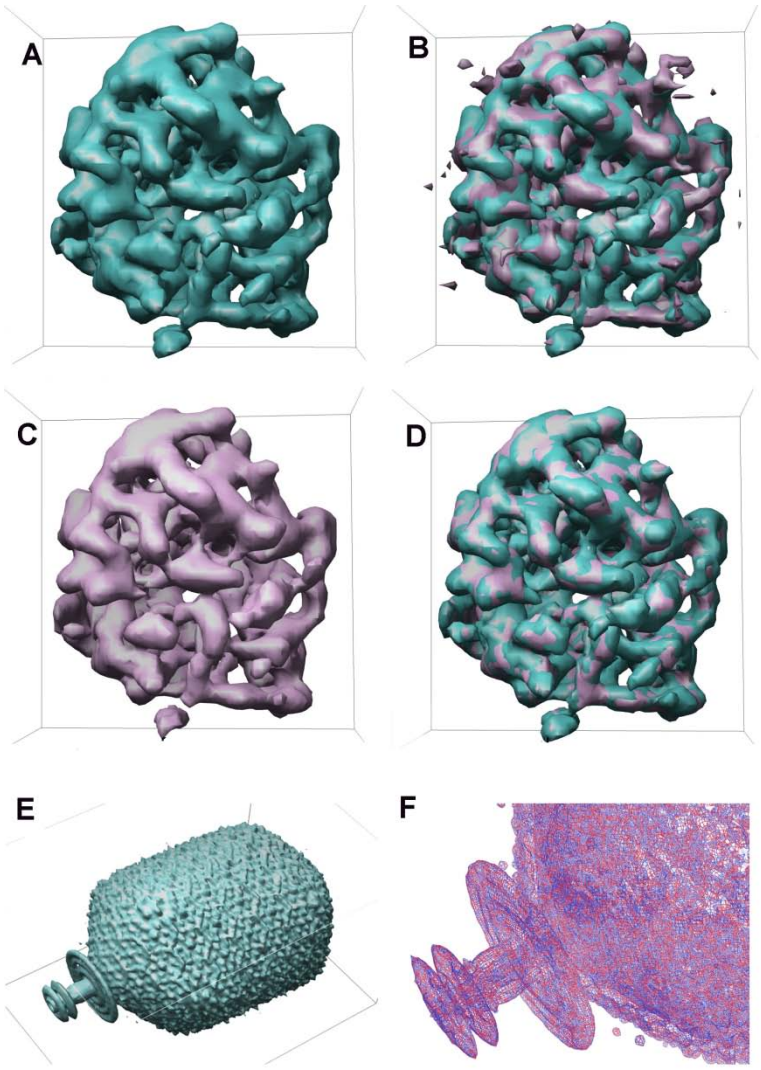

Fig. 3. A-D 1064_em. A. Original; B. Original plus reconstruction using 0.1% coefficients; C. Reconstruction using 0.5% coefficients; D. Original plus reconstruction using 0.5% coefficients. E-F 1075_em E. Rendered virus reconstruction; F. Detail of superposed original (blue) and reconstructed (red) map.

References

1. Martone, M.E., Gupta, A., Wong, M., Qian, X., Sosinsky, G., Ludascher, B., Ellisman, M.H.: A cell-centered database for electron tomographic data. Journal of Structural Biology **138** (2002) 145-155
2. Fuller, S.D.: Depositing electron microscopy maps. Structure **11** (2003) 11-12
3. Tagari, M., Newman, R., Chagoyen, M., Carazo, J.M., Henrick, K.: New electron microscopy database and deposition system. Trends in Biochemical Sciences **27** (2002) 589
4. Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S.: The Protein Data Bank. Acta Crystallographica Section D Biological Crystallography **58** (2002) 899-907
5. Kleywegt, G.J., Harris, M.R., Zou, J., Taylor, T.C., Wahlby, A., Jones, T.A.: The Uppsala Electron-Density Server. Acta Crystallographica Section D Biological Crystallography **60** (2004) 2240-2249
6. Rocchia, W., Alexov, E., Honig, B.: Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. Journal of Physical Chemistry B **105** (2001) 6507-6514
7. Goodford, P.J.: A computational-procedure for determining energetically favorable binding-sites on biologically important macromolecules. Journal of Medicinal Chemistry **28** (1985) 849-857
8. Mallat, S.G.: A theory for multiresolution signal decomposition - the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **11** (1989) 674-693
9. Daubechies, I.: Ten lectures on wavelets. SIAM, Montpelier (1992)
10. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in Fortran. Cambridge University Press, New York (1992)
11. Stollnitz, E.J., Derose, T.D., Salesin, D.H.: Wavelets for computer graphics. Morgan Kaufmann Publishers, Inc, San Francisco (1996)
12. Muraki, S.: Volume data and wavelet transforms. IEEE Computer Graphics and Applications **13** (1993) 50-56
13. Lio, P.: Wavelets in bioinformatics and computational biology: state of art and perspectives. Bioinformatics **19** (2003) 2-9
14. Carson, M.: Wavelets and molecular structure. Journal of Computer-Aided Molecular Design **10** (1996) 273-283
15. Bajaj, C., Castrillon-Candas, J., Siddavanahalli, V., Xu, Z.Q.: Compressed representations of macromolecular structures and properties. Structure **13** (2005) 463-471
16. Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M.: Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallographica Section A **47** (1991) 110-119
17. Collini, M., D'Alfonso, L., Molinari, H., Ragona, L., Catalano, M., Baldini, G.: Competitive binding of fatty acids and the fluorescent probe 1-8-anilinonaphthalene sulfonate to bovine {beta}-lactoglobulin. Protein Science **12** (2003) 1596-1603