Structure, Vol. 13, 463–471, March, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.02.004

Compressed Representations of Macromolecular Structures and Properties

Chandrajit Bajaj,[1,2,*] Julio Castrillon-Candas,[1] Vinay Siddavanahalli,[1] and Zaiqing Xu[1]
[1]Computational Visualization Center
Department of Computer Sciences and
Institute for Computational and Engineering Sciences
ACES 2.128, 24th & Speedway
University of Texas
Austin, Texas 78712

Summary

We introduce a new and unified, compressed volumetric representation for macromolecular structures at varying feature resolutions, as well as for many computed associated properties. Important caveats of this compressed representation are fast random data access and decompression operations. Many computational tasks for manipulating large structures, including those requiring interactivity such as real-time visualization, are greatly enhanced by utilizing this compact representation. The compression scheme is obtained by using a custom designed hierarchical wavelet basis construction. Due to the continuity offered by these wavelets, we retain very good accuracy of molecular surfaces, at very high compression ratios, for macromolecular structures at multiple resolutions.

Introduction

Macromolecular structures (ribosomes, viruses, ion channels) form the machinery responsible for most biological processes and are relevant to our understanding of diseases such as cancer and other metabolic disorders. Today, hybrid experimental approaches utilizing X-ray and electron crystallography (Darst et al., 1998), cryo-EM (Frank et al., 2000; Heel et al., 2000; Rossmann et al., 2001), and molecular tomographic imaging (Medalia et al., 2002), together with structural proteomics (Sali et al., 2003; Zhou et al., 2001), have led to a rapid increase in the number of deposited large macromolecular structures (albeit at varying resolutions) in the Protein Data Bank (PDB) (Berman et al., 2000) and at the European Bioinformatics Institute database (EBI, http://www.ebi.ac.uk). Supercomputers across the world are also now routinely deployed for the computation of associated macromolecular properties, such as volumetric force fields based on Poisson-Boltzmann electrostatics potential estimations (Baker et al., 2001; Orlova et al., 1996).

To achieve efficient computational processing, analysis, and possibly interactive visualization of large macromolecular structures, new software techniques and tools need to be developed that manipulate the macromolecular structure representation with maximal utilization of current computer processing units (central and graphics), computer memory (main and graphics), and network bandwidth. An important solution technique used in conjunction with large data manipulation has been data compression. Data compression maximally utilizes network bandwidth for the transmission of macromolecular structures, from disk to the computer's main or graphics memory and from possibly remote data servers to compute and display clients. Furthermore, achieving interactive performance in computations (analysis and visualization) with compressed structures is strongly dependent on rapid and at times local decompression in the respective addressable memories of the computer's processing units. Hence for applications requiring interactivity, one focuses on the construction of highly compressed, feature-preserving representations, coupled with the capability to achieve fast local decompression (Bajaj et al., 1999, 2000, 2001a, 2001b).

Simple Haar wavelets with their fast and local decompression capabilities have in the past been the basis of choice for volumetric compression schemes tuned for interactive visualization activities (Bajaj et al., 2000, 2001a, 2001b). In this paper, we introduce a new wavelet compressed volumetric representation of macromolecular structures and associated properties and address the tradeoff of excellent data compression-preserving structural features while maintaining a rapid decompression capability. These wavelet function spaces are custom built for rectilinear volumetric data and are the span of a hierarchical basis (HB). HB was originally introduced in Yserentant (1992).

Numerous molecular representations have been proposed for macromolecular structures and their properties. Molecular surfaces, approximating implicit solvation interfaces and without singularities, can be generated as a smooth isosurface of a volumetric electron density function representation from atomic resolution PDB data (Blinn, 1982; Grant and Pickup, 1995; Mestres et al., 1997). The accurate computation of volumetric electron density representations from the PDB requires computations at the quantum mechanical level (Boys, 1950). One usually approximates the electron density distribution of the i^{th} atom located at $\{x_{ci}, y_{ci}, z_{ci}\}$ and VanderWaals radius R_i with a Gaussian function

$$\rho_i(\mathbf{r}) = \exp\left(\frac{Br^2}{R_i^2} - B\right),$$

where $B < 0$ is the rate of decay parameter, $\mathbf{r} = \{xy,z,\}$, and $r^2 = (x - x_{ci})^2 + (y - y_{ci})^2 + (z - z_{ci})^2$. Visualizations of molecular surfaces from the volumetric representations can be generated using isosurface and/or volume renderings (Bajaj, 1998; Baker et al., 1999). Figures 1 and 2 shows examples of isosurface and volume renderings of the large ribosomal subunit (PDB: 1JJ2), while Figure 3 shows an example of an isosurface rendering of a microtubule structure (and see Movie S1 in the Supplemental Data available online).

*Correspondence: bajaj@cs.utexas.edu
[2]Lab address: http://www.ices.utexas.edu/CCV

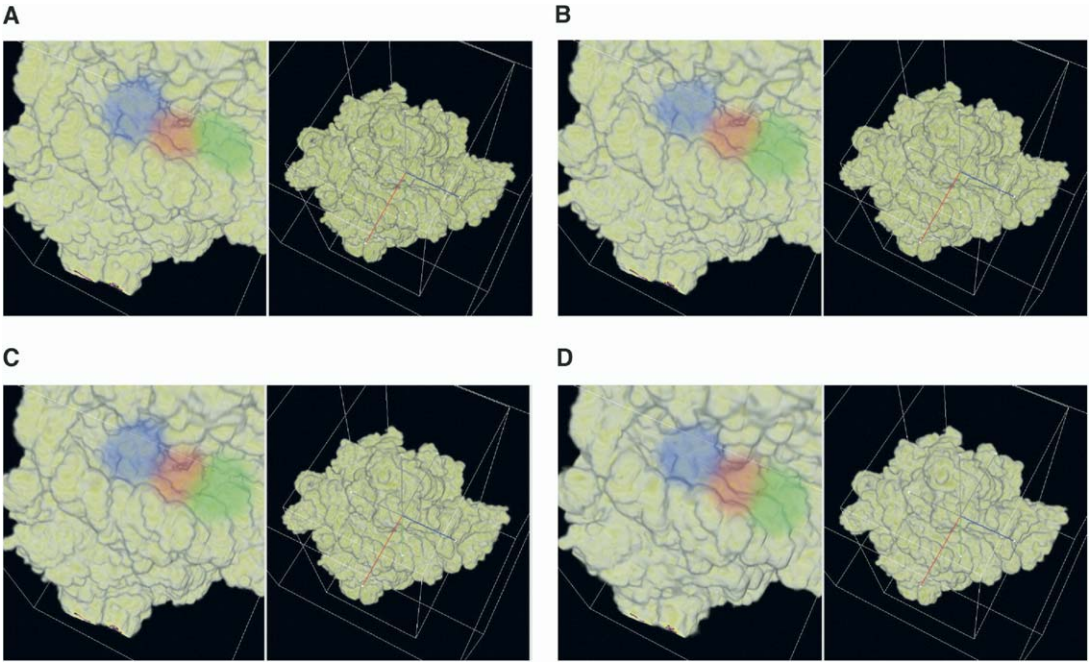

Figure 1. HB Zero-Bit Compression Example of a Volumetric Electron Density Representation of a Large Ribosomal Subunit (PDB: 1JJ2)
(Figure created using Volume Rover.)
(A) In this image we show a volume rendering of the electron density map of the large ribosomal subunit at 2.4 Å. This molecule has 98543 atoms and the electron density map has been generated using a grid of 256^3 voxels with a decay parameter of B = –1. The grid resolution is approximately 1 Å. On the right image, we show the volume rendering of the entire molecule. The left image shows a zoomed portion of the ribosomal subunit. In particular, this zoomed portion corresponds to the main active sites of the ribosome, where green is the Aminoacyl (A) site, red the Peptidyl (P) site and blue the Exit site (E).
(B) Zero-bit HB with threshold = 0.05 and compression = 11:1.
(C) Zero-bit HB with threshold = 0.15, compression = 30:1.
(D) Zero-bit HB with threshold = 0.50, compression = 65:1. For increase compression, we observe very little degradation on volume rendering of the molecular structure. Volume rendering is a smoothing operation, thus visual errors tend to be blurred out. The error in the compressed map is much more clearly seen for isosurfaces, as will be shown in Figure 2.

Volumetric electron density representations of macromolecular structures, at low or medium resolutions, are additionally available as reconstructed 3D maps from tomographic electron microscopy or single particle cryo-EM (Frank, 1996). There is a growing number of these 3D maps being deposited into the PDB and the EBI. The molecular surface of such reconstructed 3D maps of macromolecular structures can also be generated as a suitable isosurface. Again, visualizations of these maps can be generated using isosurface or volume renderings. Figure 4 shows the outer capsid layer of a Rice Dwarf Virus (RDV) map, reconstructed from single particle cryo-EM, and visualized using volume rendering.

Associated volumetric properties of high-resolution macromolecules such as electrostatic potential fields are computed by solving the Poisson-Boltzmann equation (Baker et al., 2001). Again, isosurface extraction and rendering, as well as volume rendering, provide ways of visualizing these maps. Figure 5 shows two isosurfaces from an electrostatic map of the large ribosomal subunit (PDB: 1JJ2).

The main contribution of this paper is to demonstrate that our new HB wavelets provide high feature-preserving compression balanced with a rapid and local decompression capability when applied to macromolecular structure and properties and especially in their use for interactive applications. Moreover, we also demonstrate that our new HB wavelets yield better compression tradeoffs than previously used compression schemes based on Haar wavelets.

Results and Discussion

We construct HB over 3D domains in the context of *second generation wavelets* (Sweldens, 1998). These wavelets are constructed to be adapted to the local geometry and are easy to implement. They exhibit a high degree of accuracy as well as high compression ratios. In subsection "Hierarchical Basis and Compression Scheme," we provide a discussion on wavelets, the sparsification strategy, and the encoder we have developed. Details of the mathematical construction of the HB wavelets, as well as the fast decoding scheme, are given in Bajaj et al. (2004),. In subsection "Encoding, Decoding, and Fast Random Access," we present several example applications of our HB compression scheme, including error and timing results.

Hierarchical Basis and Compression Scheme

The compression scheme can be broken in two steps. First, the data have to be sparsified using a suitable

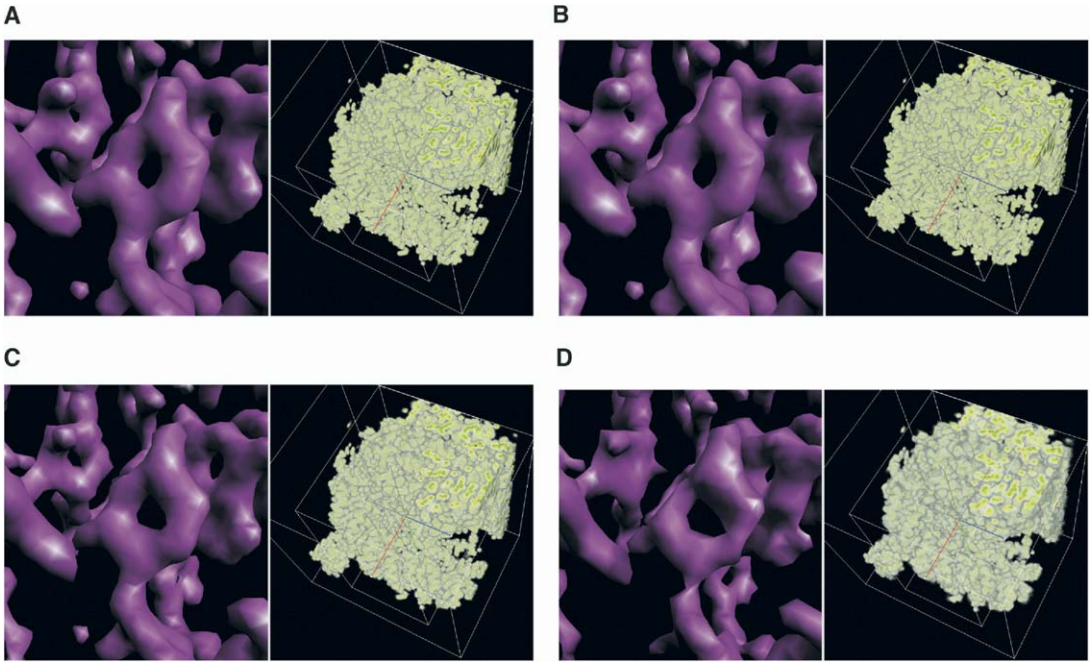

Figure 2. HB Zero-Bit Compression Example of a Volumetric Electron Density Representation of a Large Ribosomal Subunit (PDB: 1JJ2), at near Atomistic Resolution

(Figure created using Volume Rover.)

(A) In this image we show a volume rendering of the electron density map of the large ribosomal subunit at 2.4 Å. This molecule has 98543 atoms and the electron density map has been generated using a grid of 512^3 voxels with a decay parameter of $B = -5$, which shows higher molecular detail than in Figure 1. Out of this grid, we extract a subvolume of 216^3 voxels that contains the A, P, and E active sites. The grid resolution is approximately 0.5 Å. Note that this is a minimal uniform resolution needed to obtain accurate isosurfaces. A coarser resolution leads to very displeasing artifacts, where many of the atomistic details are difficult to recognize. On the right image, we have the volume rendering of the A, P, and E active sites of the ribosome. The left image shows a zoomed portion of the ribosomal structure, where several nucleotides on the active site can be seen.

(B) Zero-bit HB compression with threshold = 0.05 and compression = 8:1.

(C) Zero-bit HB compression with threshold = 0.15, compression = 18:1.

(D) Zero-bit HB compression with threshold = 0.50, compression = 42:1.

For the volume rendered image on the right, very little degradation of the ribosomal structure is visible as the compression increases. In contrast, the atomistic details shown via isosurface rendering clearly degrade.

functional representation. The second step is to produce an *encoding* and *decoding*, which tells where the nonzero terms are located. The main idea behind wavelets is to start with a coarse representation of the molecular map and do successive refinements to obtain the desired resolution. Each level of resolution will be approximated by a series of functions called the *scaling* functions. The main objective is to construct an efficient representation of a molecular map over the embedded domain at different levels of resolution. However, this is not a very efficient way to represent a function and can be very costly in terms of the number of scaling functions needed to approximate the molecular map. This motivates us to look at the differences between one level of resolution and the next. Instead of representing a molecular map at a single high-resolution level, we transform it into a very coarse approximation (i.e., level zero) and the *differences* between each level of resolution. The functions that describe these differences are called the wavelets. In comparison to the scaling functions, the coefficients associated with the wavelet exhibit fast decay for smooth data. If a molecular map contains a discontinuity, the neighborhood wavelet coefficients are in general large, but away from the discontinuity the decay is very fast. This makes wavelets very well suited for representing local and global variations. In contrast, Fourier and spherical harmonics have very slow decays due to the bias toward global representations.

A linear HB is constructed in 3D (Bajaj et al., 2004) due to the high degree of compression, ease of implementation, and fast random access. In this same paper, we devise the strategies for the compression of molecular data. The linear HB representation we construct is essentially an adaptive mesh representation. However, most adaptive meshes are discontinuous, unless a computational expensive step is used a priori to avoid discontinuities. These discontinuities lead to artifacts for the visualization of molecular volumes and surfaces; in particular, the isosurface extraction would be very poor. In contrast, the HB wavelet representation itself is continuous.

Encoding, Decoding, and Fast Random Access

As with all compression schemes, a suitable encoding of the nonzero wavelets is a necessary step. Many such schemes have been developed for efficient transmission of compressed data streams. However, very few

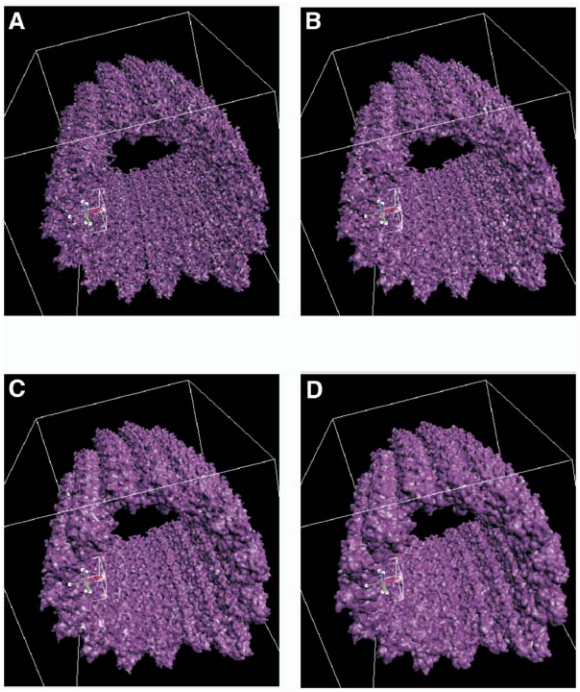

Figure 3. HB Zero-Bit Compression of a Volumetric Electron Density Representation of a Microtubule Structure

(Figure created using Volume Rover.)

(A) This microtubule representation has 1,210,410 atoms and the electron density map has been embedded in a grid of 512 * 512 * 1024 voxels with a decay parameter of $B = -2.3$. The grid resolution is of approximately 0.6 Å. The image above shows the isosurface of the microtubule at atomic resolution.

(B) Zero-bit HB compression with threshold = 0.05 and compression = 32:1.

(C) Zero-bit HB compression with threshold = 0.15, compression = 77:1.

(D) Zero-bit HB compression with threshold = 0.35, compression = 132:1.

This image clearly shows the degradation of the molecular data with respect to very high compression. At a compression ratio of 132, structural feature degradation is quite apparent.

encoding schemes, like volume rendering and isocontouring, have been developed for visualization purposes. Such encoding schemes should aim for high compression ratios and fast random access. The zero-bit encoding scheme, introduced by Bajaj et al. (2001a), uses a two-stage significance map to achieve a good compromise between these two aspects. Whether wavelet coefficients are null or not is stored in "zero-bits," while the significance information of the detail coefficients of each non-null detail node is stored in additional bits. Nevertheless, since the encoding is limited to Haar wavelets, unpleasant and visual artifacts appear at moderately high compression ratios. To solve this problem, a novel encoding scheme is created on the basis of zero-bit encoding for HB linear wavelets.

The wavelet coefficients are assumed to be mostly zero, with a few localized nonzero coefficients, as shall be the case for a large number of volumetric maps. Under this assumption, a multiresolution encoding scheme can be built to optimally access the concentrated information. In this encoding scheme, the rectangular 3D volume data are subdivided into blocks of equal size (which can be predetermined by the user). Each block is further segmented into cells which contain a fixed number of voxels. Now, we run the wavelet transform on each cell separately. Since the transformed data usually has large swaths of zeros, a large number of cells or blocks can be ignored. By constructing an appropriate multiresolution indexing scheme, an efficient encoding of the nonzero coefficients is built (Bajaj et al., 2004).

Timings, Errors, and Visualizations of Molecular Maps

To gauge the effectiveness of the HB representation to molecular maps, we apply it to three different types of volumetric data, in particular to volumetric maps from X-ray crystal structures at multiple feature resolutions, electrostatic potential field data, and reconstructed 3D maps from cryo-EM. Moreover, actual random access timings for the zero-bit decoder are presented. All of the figures in this paper were created with the Volume Rover software. Volume Rover (http://ccvweb.csres.utexas.edu/ccv/projects/project.php?proID=9) is a volumetric map visualization software tool developed at the Computational Visualization Center, University of Texas at Austin.

For volumetric electron density maps created from high-resolution X-ray structural data, such as shown for the ribosome in Figure 2, the astute reader would notice that the volumetric map seems oversampled. However, such volumetric sampling resolutions are at times necessary to visualize the atomic resolution feature details because of the approximations inherent to isosurface and volume rendering. Isosurface rendering are based on surface triangle extraction from a local piecewise trilinear approximation of the volumetric data. This enables use of the triangle processing pipeline of graphics cards, and hence accelerated rendering speeds. Volume rendering is also based on a local piecewise approximation and again done so as to achieve graphics card accelerated volumetric.

Our wavelet method sparsifies the molecular map by applying a level-dependent threshold criterion denoted by ϵ. This term compares the wavelet coefficients to the peak of the molecular map and a level depend factor (Bajaj et al., 2004). As ϵ is increased, we observe that compression ratio increases, but the peak signal to noise ratio (PSNR)—which describes the quality of the compressed data—of the reconstructed image degrades.

In Figure 1 we demonstrate compression results for a synthesized 256^3 map of the large ribosomal subunit (PDB: 1JJ2) with a decay rate of $B = -1$. Notice that this is the minimum resolution needed to avoid unwanted volume rendering artifacts. Of interest in this molecule are the Aminocyl, Peptidyl, and Exit active sites on the ribosome. If one looks carefully, the left images for Figures 1A–1D show some degradation of the active sites as compression is increased. This can be explained by the smoothing operation of volume rendering algorithms. However, errors are more clearly seen for the isosurface renderings.

To show the effect of compression on the isosurface,

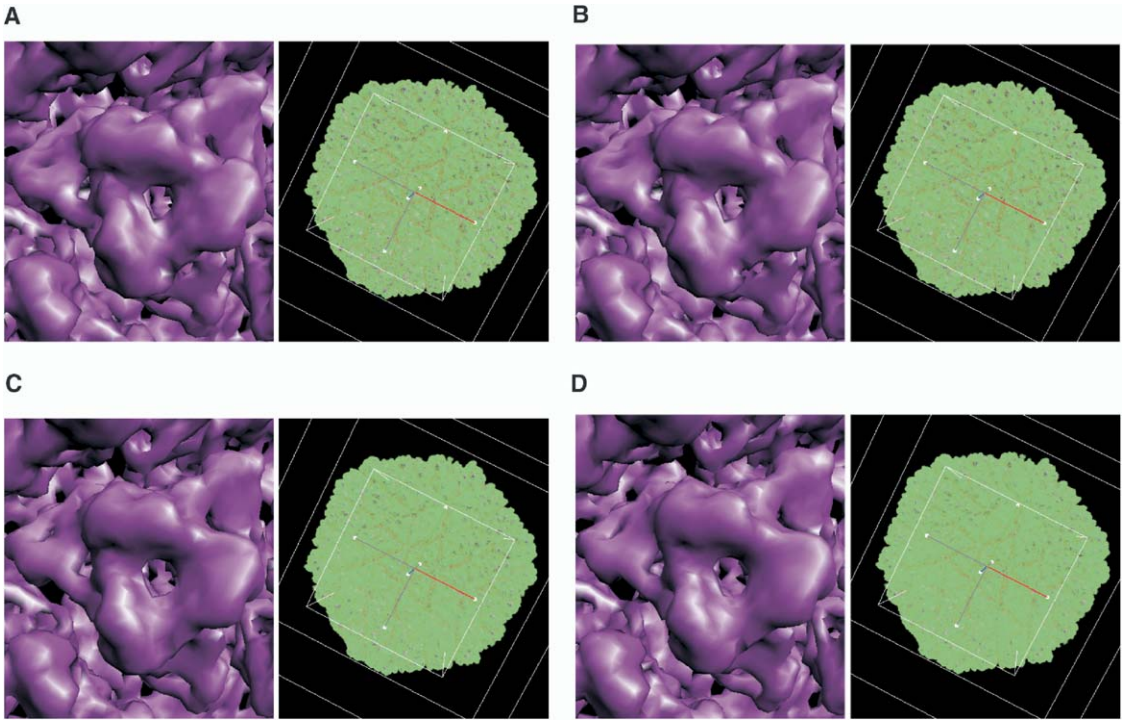

Figure 4. HB Compression Example of a Reconstructed 3D Map of an RDV Structure

(Figure created using Volume Rover.)

(A) Volume rendering and isosurface of the RDV map at 6.8 Å resolution in a grid of $241 * 481^2$ voxels. On the right image, the outer capsid layer is rendered in green. The left image shows a zoomed portion of the capsid, displaying one of the trimeric subunits (purple).

(B) Threshold = 0.15, compression = 22:1.

(C) Threshold = 0.35, compression = 37:1.

(D) Threshold = 0.50, compression = 57:1. As the compression increases, we observe a slight degradation in the trimeric subunit of the virus capsid.

we synthesize a 512^3 electron density map from the large ribosomal subunit (PDB: 1JJ2) with a rate of decay of $B = -5$. For this rate of decay, the atomistic details can be distinguished. A smaller grid resolution would lead to highly degraded isosurface. In Figures 2A–2D, we show the volume rendering of the A, P, and E active sits and an isosurface of a nucleotide on these sites. As the compression increases, it is clear that the atomistic features degrade. However, even at high compression many of the atomistic features can be distinguished.

Another interesting feature to observe is the compression capabilities of our scheme with respect to maps of varying resolutions. In Tables 1A–1C, compression results for the large ribosomal subunit (PDB: 1JJ2) electron density map at rates of decay B = –5, –2, –1. For a fixed threshold level, we observe that both the PSNR and CR increase with a decrease in resolution (by changing *B*).

In Figure 3, compression results for a 1.2 million atom microtubule structure are shown. Much of the overall structure is preserved at high compression ratios of around 130:1. However, many of the atomistic details are lost and the whole structure seems to loose resolution.

Our HB scheme is applied on a 3D map of the Rice Dwarf Virus, reconstructed from single-particle cryo-EM. The dataset dimension is 241×481^2 with a resolution of 6.8 Å. In Figure 4A, on the right image we show a volume rendering of the original RDV cryo-EM map. In addition, on the left image, an isosurface shows the trimeric unit of the virus. In Figures 4B–4D, compression results for three different thresholds are shown. High visual accuracy of the structure of the virus can be observed even at a compression ratio of 57:1. In Table 2A, we present numeric data of our compression algorithm for more threshold levels.

We finally test the HB representation with zero-bit encoding on the potential field of the large ribosomal subunit (PDB: 1JJ2). This molecular map was created from a Poisson-Boltzmann solver (Baker et al., 2001). In Table 2B, we observe that the CR increases as the PSNR falls. However, even at high compressions ratios the image quality degrades very slowly, as shown in Figures 5A–5D.

For sake of completeness, we compare our continuous zero-bit HB scheme with the zero-bit Haar compression method that to our knowledge is the only other compression-based interactive algorithm available. We choose the same data that were produced for Figure 2. It is clear from Figure 6 that at high compression, our method preserves the structure well. On the other hand, blocky artifacts are seen in the Haar case as the wavelets are discontinuous.

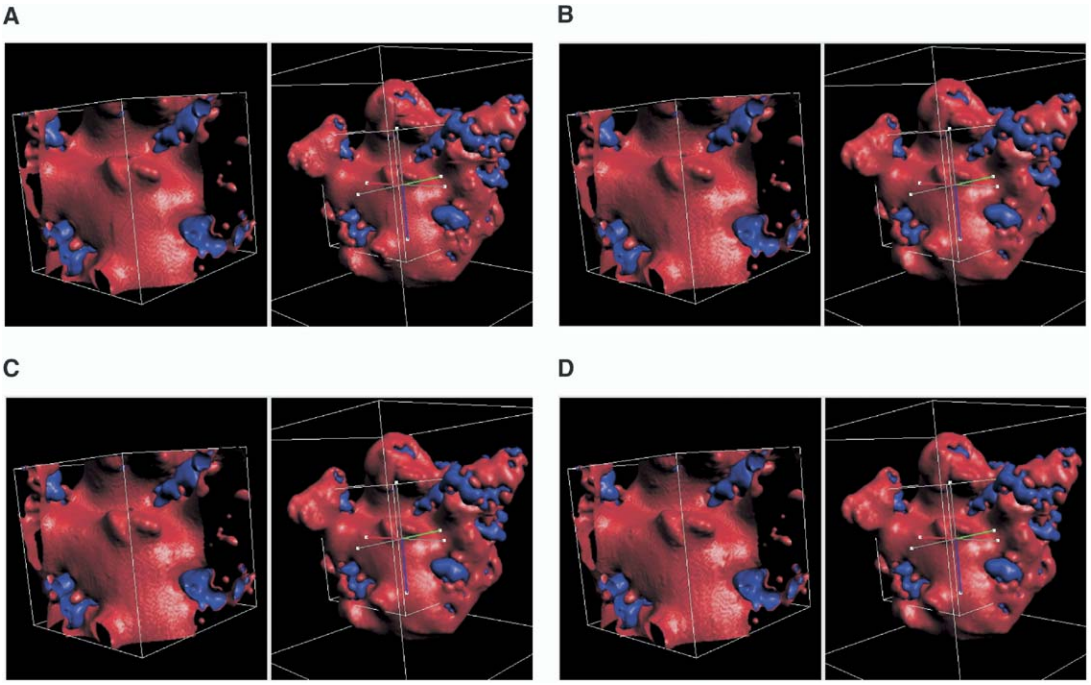

Figure 5. HB Zero-Bit Compression Example on a Computed Electrostatics Potential Field Property of a Large Ribosomal Subunit

(Figure created using Volume Rover.)

(A) Noncompressed map of the potential field of the large ribosomal subunit. This 512^3 gridded field data was created using the Adaptive Poisson-Boltzmann Solver (Baker et al., 2001). The right image shows two isosurfaces of the field, where the blue isosurface is positive and the red one is negative potential. The left image shows a zoomed portion of the field, as captured by the smaller cube shown in the right image. The pictures are created using our center's VolRover software.

(B) Zero-bit HB with threshold= 0.05, compression = 41:1.

(C) Zero-bit HB with threshold = 0.15, compression = 55:1.

(D) Zero-bit HB with threshold = 0.35, compression = 62:1. We see very little degradation of the field as we increase the compression. However, as shown in Table 2, the errors clearly increase with higher compression.

In Table 3, we show random access timing test cases for our method tested on the electron density map of the large ribosomal subunit (PDB: 1JJ2), with varying rates of decay. The four test cases are as follows:

- t_1: The average time of 1000 random accesses of a voxel from the compressed format.
- t_2: The average time of 1000 random accesses of a full cell from the compressed format.
- t_3: The average time of 1000 random accesses of a voxel inside a cell, where the cell was chosen at random from the uncompressed format.
- t_4: The average time to decode the entire compressed volume (512^3) for 10 runs.

These timings were run on an SGI ONYX2 system using a single processor. As observed, to decode a single voxel, the slowest case is t_1. However, for most volume rending and isocontouring applications it is more efficient to decode an entire cell since many neighboring points are used.

Conclusions

We have demonstrated the value of HB wavelet compressed volumetric representations for macromolecular structures and associated properties, especially in the age of rapidly increasing structures and medium resolution maps in both the PDB and EBI. The added value of fast decompression time yields the necessary ingredient for enabling interactive visualization performance on current desktop computers. Our future work is predominantly in the development of computational analysis algorithms that work directly from HB compressed volumes of bio-molecules and associated properties, with fast decompression. We are also interested in the development of interactive visualization techniques which optimally utilize the faster bandwidth of programmable graphics hardware, commonplace on most desktop computers.

Supplemental Data

Supplemental Data include five movies and can be found with this article online at http://www.structure.org/cgi/content/full/13/3/463/DC1/.

Acknowledgments

We sincerely thank Professor Wah Chiu (Baylor School of Medicine) for the 3D map of the Rice Dwarf Virus, Professor Nathan Baker (Washington University at St. Louis), Professor Andrew McCammon (University of California, San Diego) for the electrostatics potential computed data for the 50S large subunit of the ribosome, and Professor Dave Sept (Washington University at St. Louis) for the 1.2 million atom microtubule data. This work was supported in part by

Table 1. Compression and Comparison Results for the Large Ribosomal Subunit (512^3) with Different Rate of Decay Parameter B

A. $B = -5$

ϵ	PSNR	% 0's after TX	% 0's Original	TCC	NZC	e_∞	e_2	Size (Bytes)	CR
0.50	30.15	99.10	77.67	111.73	24.94	54.13	27.84	6376816	84.19
0.35	31.87	98.81	77.67	84.07	18.77	43.37	22.82	8025803	66.89
0.15	36.13	97.73	77.67	44.13	9.85	19.09	13.97	14473038	37.09
0.05	42.52	94.71	77.67	18.91	4.22	6.63	6.63	32247352	16.64
0.01	58.49	87.39	77.67	7.93	1.77	1.30	1.04	72738554	7.38
0.00								536870980	

B. $B = -2$

ϵ	PSNR	% 0's after TX	% 0's Original	TCC	NZC	e_∞	e_2	Size (Bytes)	CR
0.50	30.82	99.19	74.77	123.72	31.20	29.80	19.76	5807819	92.43
0.35	32.76	98.92	74.77	93.06	23.47	25.71	15.73	7315547	73.38
0.15	37.35	97.99	74.77	49.80	12.56	17.18	9.31	12469751	43.05
0.05	42.15	95.76	74.77	23.63	5.96	6.53	5.31	25731023	20.86
0.01	56.88	86.17	74.77	7.23	1.82	1.31	0.96	79575816	6.74
0.00								536870980	

C. $B = -1$

ϵ	PSNR	% 0's after TX	% 0's Original	TCC	NZC	e_∞	e_2	Size (Bytes)	CR
0.50	33.16	99.35	71.58	153.85	43.72	22.24	11.93	4586097	117.06
0.35	34.52	99.22	71.58	128.86	36.62	16.91	10.23	5460302	98.32
0.15	39.18	98.54	71.58	68.56	19.48	9.27	5.92	9432914	56.91
0.05	44.58	97.06	71.58	34.07	9.68	5.77	3.18	17706006	30.32
0.01	54.79	89.12	71.58	9.19	2.61	1.31	0.96	63375575	8.47
0.00								536870980	

(A–C) Compression results with HB basis.
Legend: threshold criterion ϵ, PSNR = peak signal to noise ratio, TX: wavelet transform, TCC: total coefficient compression ratio, NZC: nonzero compression ratio, e_∞: maximum (l^∞) percent error between reconstructed compressed version and the original, e_2: energy (l^2) percent error between reconstructed compressed version and the original. CR: file compression ratio.

Table 2. Zero-Bit Encoded HB Compression Results for the Potential Field of the Large Ribosomal Subunit (PDB: 1JJ2) and a Cryo-EM map of the Rice Dwarf Virus

A. Zero-bit HB Compression results for the 6.8 Å resolution Rice Dwarf Virus Cryo-EM map embedded in a grid of $481 * 481 * 241$ voxels.

ϵ	PSNR	% 0's after TX	% 0's Original	TCC	NZC	e_∞	e_2	Size (Bytes)	CR
0.50	31.10	99.15	72.61	117.83	32.27	41.11	25.61	4771349	46.74
0.35	32.97	98.88	72.61	89.49	24.51	36.45	20.64	6042206	36.91
0.150	36.03	98.08	72.61	52.29	14.32	18.95	14.53	10031643	22.23
0.05	42.69	94.73	72.61	19.01	5.20	6.68	6.74	26590966	8.38
0.01	55.69	86.26	72.61	7.27	1.99	1.31	1.48	67075200	3.32
0.00								223032072	

B. Zero-bit HB compression results for the potential field of the large ribosomal subunit (PDB: 1JJ2), 2.4 Å Resolution), with a grid resolution of 512^3 voxels (Baker et al., 2001).

ϵ	PSNR	% 0's after TX	% 0's Original	TCC	NZC	e_∞	e_2	Size (Bytes)	CR
0.500	38.50	98.84	0.58	85.91	85.40	63.01	3.48	8263922	64.96
0.35	40.24	98.78	0.58	82.39	81.91	45.67	2.85	8572631	62.62
0.15	44.18	98.60	0.58	71.85	71.43	19.75	1.81	9675507	55.48
0.05	50.29	98.06	0.58	51.56	51.26	6.28	0.90	12972064	41.38
0.01	60.36	96.13	0.58	25.85	25.70	1.29	0.28	24344104	22.05
0.00								536870980	

(A) The cryo-EM of the Rice Dwarf Virus at 6.8 Angstroms resolution is embedded in $481 * 481 * 241$ a grid.
(B) The potential field of the ribosome with 2.4 Å resolution is embedded in a grid of 512^3 voxels (≈ 0.5 Å resolution). This data set was created with the Adaptive Poisson-Boltzmann Solver (Baker et al., 2001).
Legend: threshold criterion ϵ, PSNR = Peak Signal to noise ratio, TX: wavelet transform, TCC: total coefficient compression ratio, NZC: nonzero compression ratio, e_∞ is equal to the maximum (l^∞) percent error between reconstructed compressed version and the original, e_2 is the energy (l^2) percent error between reconstructed compressed version and the original. CR: file compression ratio.

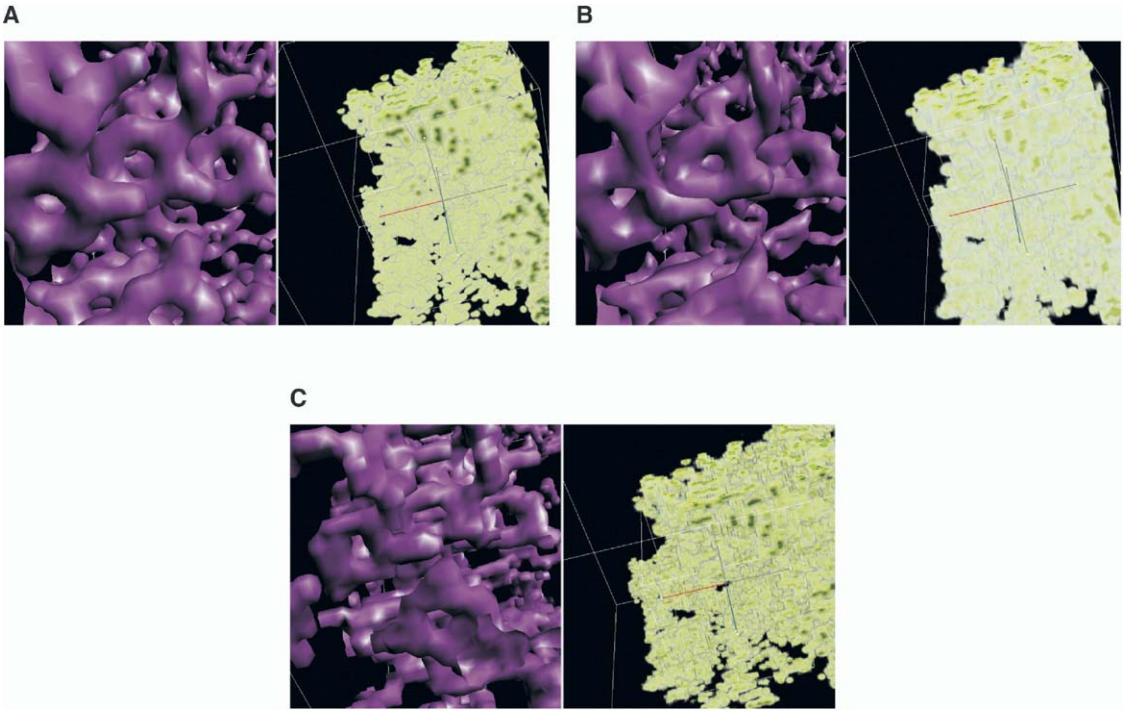

Figure 6. Compression Comparison between Zero-Bit HB and Haar Schemes

(Figure created using Volume Rover.)

(A) Rendering example of the large ribosomal subunit data set (See also Figure 2). In the right image, a portion of the electron density of the ribosome is shown. This portion corresponds to the active P site (Peptidyl transferase activity). The left image shows a zoomed part of the active site. This figure contains a colored purple isosurface showing the hexagonal shape of a purine nucleotide near the active site.

(B) HB method, compression = 42:1. For a 42:1 compression ratio we observe degradation of the molecular structure. However, the nucleotide structure can still be distinguished with our HB method. This in contrast to the Haar representation.

(C) Haar method, compression = 42:1. For the Haar method we observe that for the same compression as HB, the hexagonal shape of the nucleotides is highly distorted. In addition, volume rendered image on the right appears blocky.

NSF grants INT-9987409, ACI-022003, and EIA-0325550 and a grant from the NIH 0P20 RR020647. Our HB software for volumetric compression and decompression, and our Volume Rover visualization software can be downloaded at Computational Visualization Center's software download page (http://www.ices.utexas.edu/ccv/software). Finally, we mention that to our knowledge, we do not have any financial conflict that can influence the results or interpretation of this manuscript.

Received: December 10, 2004
Revised: February 10, 2005
Accepted: February 11, 2005
Published: March 8, 2005

References

Bajaj, C. (1998). Data Visualization Techniques (Chichester, UK: John Wiley & Sons).

Bajaj, C., Castrillon-Candas, J.E., Siddavanahalli, V., and Xu, Z. (2004). Compressed volumetric representations of macromolecular structures and properties. ICES Technical report, University of Texas, Austin.

Bajaj, C., Pascucci, V., and Zhuang, G. (1999). Progressive compression and transmission of arbitrary triangular meshes. In Proceedings of IEEE Visualization Conference (San Francisco, CA), pp. 307–316.

Bajaj, C., Ihm, I., and Park, S. (2000). Compression-based 3D texture mapping for real-time rendering. Graph. Models *62*, 391–410.

Bajaj, C., Ihm, I., and Park, S. (2001a). 3D RGB image compression for interactive applications. ACM Trans. Graph. *20*, 10–38.

Bajaj, C., Ihm, I., and Park, S. (2001b). Visualization-specific compression of large volume data. In Proc. of Pacific Graphics (Tokyo, Japan), pp. 212–222.

Baker, N., Spet, D., Joseph, S., Holst, M.J., and McCammon, J.A.

Table 3. Reconstruction Random Access Timings for the Electron Density Maps of the Large Ribosomal Subunit (PDB: 1JJ2) with Varying Rates of Decay

Rate of Decay	t_1	t_2	t_3	t_4
$B = -1$	5.049×10^{-5}	1.456×10^{-4}	2.58×10^{-6}	26.679
$B = -2$	5.379×10^{-5}	1.427×10^{-4}	2.58×10^{-6}	25.835
$B = -5$	5.404×10^{-5}	1.375×10^{-4}	2.58×10^{-6}	24.829

The timings are in seconds.

Legend: t_1= The average time of 1000 random accesses of a voxel from the compressed format. t_2= The average time of 1000 random accesses of a full cell from the compressed format. t_3= The average time of 1000 random accesses of a voxel inside a cell, where the cell was chosen at random from the uncompressed format. t_4= The average time to decode the entire compressed volume (512^3) for 10 runs.

(2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Baker, T.S., Olson, N.H., and Fuller, S.D. (1999). Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. American Soc. Microbiol. *63*, 862–922.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank (www.pdb.org). Nucleic Acids Res. *28*, 235–242.

Blinn, J.F. (1982). A generalization of algebraic surface drawing. ACM Trans. Graph. *1*, 235–256.

Boys, S. (1950). Electronic wave functions. i, a general method of calculation for the stationary states of any molecular system. Proc. R. Soc. Lond. A Math. Phys. Sci. *200*, 542–554.

Darst, S.A., Polyakov, A., Richter, C., and Zhang, G. (1998). Insights into Escherichia coli RNA polymerase structure from a combination of x-ray and electron crystallography. J. Struct. Biol. *124*, 115–122.

Frank, J. (1996). Three-Dimensional Electron Microscope of Macromolecular Assemblies (San Diego, CA: Academic Press).

Frank, J., Penczek, P., Agrawal, R.K., Grassucci, R.A., and Heagle, A.B. (2000). Three-dimensional cryoelectron microscopy of ribosomes. Methods Enzymol. *317*, 276–291.

Grant, J., and Pickup, B. (1995). A gaussian description of molecular shapes. J. Phys. Chem. *99*, 3503–3510.

Heel, M.V., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000). Single-particle electron cryo-microscopy: Towards atomic resolution. Q. Rev. Biophys. *33*, 307–369.

Medalia, O., Weber, I., Frangakis, A.S., Nicastro, D., Gerisch, G., and Baumeister, W. (2002). Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. Science *298*, 1209–1213.

Mestres, J., Rohrer, D.C., and Maggiora, G.M. (1997). Mimic: A molecular-field matching program. exploiting applicability if molecular similarity approaches. J. Comput. Chem. *18*, 934–954.

Orlova, E.V., Serysheva, I., Heel, M.V., Hamilton, S.L., and Chiu, W. (1996). Two structural configurations of the skeletal muscle calcium release channel. Nat. Struct. Biol. *3*, 547–552.

Rossmann, M.G., Bernal, R., and Pletnev, S.V. (2001). Combining electron microscopic with x-ray crystallographic structures. J. Struct. Biol. *136*, 190–200.

Sali, A., Glaeser, R., Earnest, T., and Baumeister, W. (2003). From words to literature in structural proteomics. Nature *422*, 216–225.

Sweldens, W. (1998). The lifting scheme: A construction of second generation wavelets. SIAM J. Math. Anal. *29*, 511–546.

Yserentant, H. (1992). Hierarchical bases. ICIAM *91*. RE Ed, SIAM Philadelphia *49*, 329–412.

Zhou, Z.H., Baker, M.L., Jiang, W., Dougherty, M., Jakana, J., Dong, G., Lu, G., and Chiu, W. (2001). Electron cryomicroscopy and bioinformatics suggest protein fold models for rice dwarf virus. Nat. Struct. Biol. *8*, 868–873.