BMC Structural Biology

Methodology article

Wavelets filtering for classification of very noisy electron microscopic single particles images- application on structure determination of VP5-VP19C recombinant

Ali Samir Saad*

Address: Department of Biomedical Technology, College of Applied Medical Sciences, King Saud University, P.O. Box. 10219, Riyadh 11433, Kingdom of Saudi Arabia

Email: Ali Samir Saad* - asaad64@yahoo.com

* Corresponding author

Published: 11 December 2003

BMC Structural Biology 2003, **3**:9

Received: 30 August 2003

Accepted: 11 December 2003

This article is available from: http://www.biomedcentral.com/1472-6807/3/9

Abstract

Background: Images of frozen hydrated [vitrified] virus particles were taken close-to-focus in an electron microscope containing structural signals at high spatial frequencies. These images had very low contrast due to the high levels of noise present in the image. The low contrast made particle selection, classification and orientation determination very difficult. The final purpose of the classification is to improve the signal-to-noise ratio of the particle representing the class, which is usually the average. In this paper, the proposed method is based on wavelet filtering and multi-resolution processing for the classification and reconstruction of this very noisy data. A multivariate statistical analysis (MSA) is used for this classification.

Results: The MSA classification method is noise dependant. A set of 2600 projections from a 3D map of a herpes simplex virus -to which noise was added- was classified by MSA. The classification shows the power of wavelet filtering in enhancing the quality of class averages (used in 3D reconstruction) compared to Fourier band pass filtering. A 3D reconstruction of a recombinant virus (VP5-VP19C) is presented as an application of multi-resolution processing for classification and reconstruction.

Conclusion: The wavelet filtering and multi-resolution processing method proposed in this paper offers a new way for processing very noisy images obtained from electron cryo-microscopes. The multi-resolution and filtering improves the speed and accuracy of classification, which is vital for the 3D reconstruction of biological objects. The VP5-VP19C recombinant virus reconstruction presented here is an example, which demonstrates the power of this method. Without this processing, it is not possible to get the correct 3D map of this virus.

Background

The three-dimensional (3D) reconstruction of a virus using electron cryomicroscopy yields crucial information about the assembly of the virus, and about the mechanisms for infectivity in both humans and livestock. Structural determination begins with acquisition of images from an electron cryomicroscope, which are true projections of the virus. A major part of the data processing is aimed at determining the direction of the projection for each particle image, so that a 3D reconstruction can be computed. Figure 1 depicts a schema for such a reconstruction. The first step in this reconstruction is particle

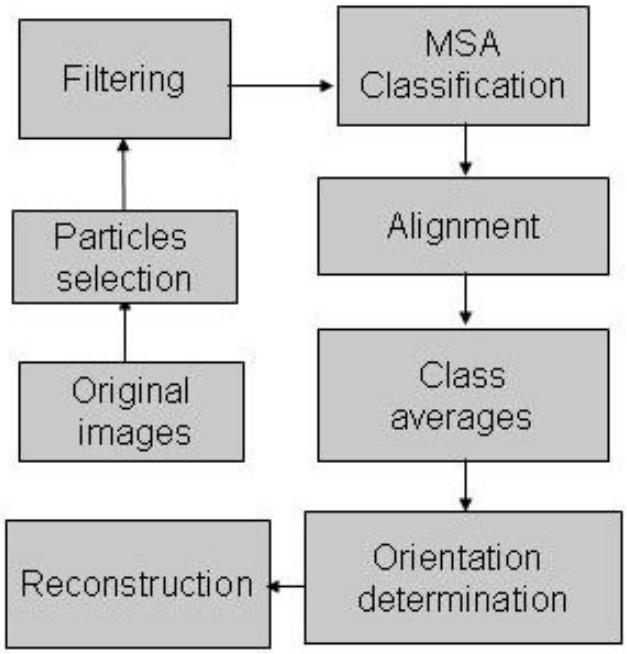

Figure 1
schematic diagram of reconstruction. A schematic diagram represents the principal operations for 3D reconstruction of a single particle. Starting by a particle selection and filtering operation, followed by classification to enhance the signal-to-noise ratio. Then an alignment is needed prior to average the particles in each class. The fourth operation is the orientation determination and the final operation is the 3D reconstruction.

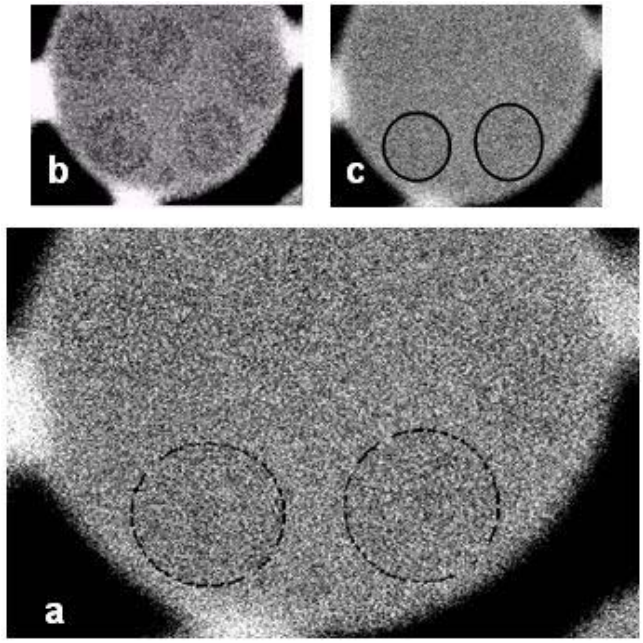

Figure 2
Wavelet multi-resolution of electron microscopic image. An electron cryomicroscopy image contains 5 particle images of the herpes simplex virus B-capsid. Two particles are circled for identification. The imaging conditions emphasize high resolution information which allows the determination of high resolution three-dimensional reconstructions. However, the effect on the image is the appearance of a very low contrast image which makes both particle selection and processing difficult. The image shown is only a small region of the image. The images currently obtained are in the order of 9000 × 13000 pixels and occupy from 100 to 300 megabytes. (b) wavelet approximation at level one of (a). (c) Fourier-filtered and sub-sampled by a factor of two in each direction of the image in (a).

selection, which involves the detection and selection of individual particles, from a large field of viruses in an electron micrograph. Traditionally, the selection of particles had been performed by hand. However, recent advances in structure biology have allowed higher resolution structures to be determined, also require a much higher number of particles to be selected [1]. This is because higher resolution requires more close-to-focus images, which contain high resolution detail but lower contrast and higher noise (Fig. 2a). The low contrast makes particle selection by hand very difficult because in close-to-focus images, particles are very difficult to see. Automated particle selection methods based on template matching to select particles, were proposed [2-5] to overcome this problem, by allowing investigators to easily select high numbers of virus images from their low contrast data.

The second step is to enhance the signal-to-noise ratio of the images to obtain accurate orientations of the particles. Enhancing the signal requires two operations. The first is filtering, which reduces the noise of the data and enhances the signal. Generally this is done by applying a band pass Gaussian filter in Fourier space [6]. The high pass filter is generally applied to eliminate the bias of alignment introduced by the very low resolution part of the signal. The next operation is grouping similar particles into classes by using a measurement of similarity. Multivariate Statistical Analysis [7,8] (MSA) is generally used in the absence of a preliminary model to perform this classification. This method is noise-dependent [7] and works better on images that are higher in contrast. Here wavelet filtering is introduced as an alternative to the Fourier filtering in order to improve the accuracy of the classification, by improving the contrast.

The third step in 3D processing is to determine the orientation of each particle. There are different criteria to determine the particle orientation. One criterion is based on the computational search of the common lines in the computed Fourier transform of individual or multiple particle images [6,9-11]. Another criterion for the particle orientation estimate is to find the correlation match between the raw images (selected particles) with many projections from a 3D model [12,13]. Regardless of the criterion used, finding the proper orientation for particle images such as in figure 2.a is difficult, because of its low contrast. For this reason pre-processing is needed in order to accurately calculate the orientation of the particle for 3D reconstruction.

In this paper, wavelet filtering for classification is described and the results obtained are compared to Fourier Gaussian filtering, which usually used for single particle classification. The 3D reconstruction of the recombinant herpes virus (VP5-VP19C) using wavelet filtering and classification is presented here.

Results

The wavelet filtering method was applied to two types of data: simulated data and real data, the following sections describe the results.

Simulated data

A set of approximately 2600 projections from the herpes 3D map [6] -covering all possible orientation of the virus- was computationally generated. First, noise was added to the projections. The noise level used [13], was equivalent to the noise generated from a JEOL 4000 operating at 400 kv and recording at 1 μm defocus (Fig. 3-b). The noisy projections are pre-processed by a band pass Fourier Gaussian filter. The low pass is a half-band Gaussian filter, and the high pass has a cut-off frequency corresponding to approximately 600 Å. The image is sub-sampled by a factor of two in each dimension, figure 3-c give a particle result of such processing. All the Fourier-filtered images are then classified by the Multivariate Statistical Analysis (MSA) classification procedure in IMAGIC-5 software [14]. The same set of projections are filtered by a biorthogonal wavelet filter and sub-sampled by a factor of two in each dimension. In other words, only the approximation component of the wavelet decomposition was used (see methods). Figure 3-d gives a particle result of such processing. Wavelet approximations for all projections were then classified by MSA.

Figures 3-c and 3-d give a qualitative assessment of the filtering methods, compared to the noise-free projection in figure 3-a. A quantitative measurement of similarity between noise-free projections and their corresponding filtered versions (for both wavelet and Fourier filtering) was made using correlation coefficient criterion. An average correlation coefficient of 0.14 is obtained for Fourier filtered projections and an average coefficient of 0.23 is obtained for the wavelet filtered ones. The result shows the superiority of wavelet filtering over Fourier one.

In order to assess the accuracy of the classification a comparison between noise-free projection and the class average including a filtered version of this projection has been completed. After classification of Fourier-filtered particles, and tracking the class including the particle in figure 3-c, an average particle of that class was calculated, which is shown in figure 3-f. Also, after classification of wavelet-filtered particles, the average particle of the class including the particle in figure 3-d was calculated, which is shown in figure 3-e. A quantitative measurement using correlation coefficient was also calculated for class averages. An average correlation coefficient of 0.17 was obtained for Fourier-filtered class averages and an average correlation coefficient of 0.63 is obtained for the wavelet-filtered ones.

By comparing the class average produced from the Fourier-filtering (Fig. 3-f), and the one produced from the wavelet-filtering (Fig. 3-e) to the original projection (Fig. 3-a). It is evident that the wavelet-filtered image (Fig. 3-e) produces a class average very close to the original projection. While the Fourier-filtered average (Fig. 3-f) is blurred and did not match the original projection very well. The quantitative measurement confirms clearly the visual result.

Real Data: VP5-VP19C recombinant virus

The 700 particles selected from 40 micrographs were filtered by wavelet filtering and classified by the MSA procedure. Figures 4.a,4.b,4.c show some of the class averages of VP5-VP19C obtained by using wavelet approximation components. The same data were also classified by MSA after Fourier filtering. Figures 4.d,4.e,4.f show some of the class averages obtained. These averages show that classification is not accurate, and that the averages are blurred compared to the wavelet-filtered ones.

The VP5-VP19C particle was reconstructed to 26-Å resolution (Fig. 5) using wavelet-filtered data. The reconstruction follows the scheme described in methods. The resolution of the final structures was evaluated based on the Fourier shell correlation coefficient [14], between two independents reconstructions, being larger than 0.5.

The surface of VP5-VP19C virus is displayed with a contour level of two standard deviations from the mean. It forms a T = 7 icosahedral lattice, which has an external diameter of ~880 Å and an internal diameter of ~580 Å, consisting of 60 hexons and 12 pentons. An asymmetric

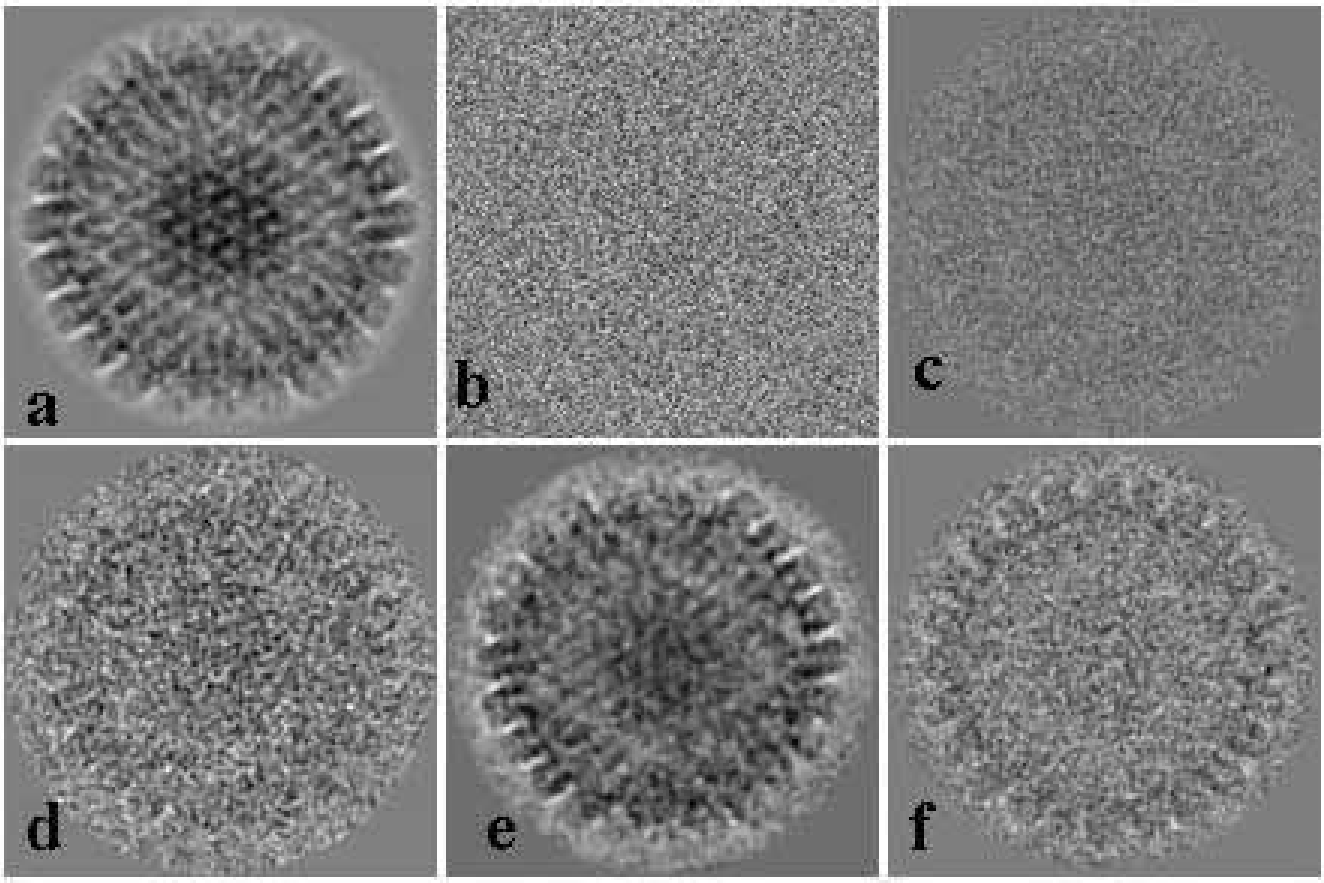

Figure 3
Classification of simulated data. This figure shows a set of images, (a) is a projection from a noise-free map of the herpes virus. (b) is a noisy version of (a), a simulated noise was added to the projection, the level of noise is equivalent to that of JEOL 4000 microscope. (c) shows the Fourier-filtered version of (b). A classification of the filtered data has been made. In order to assess this classification, it is necessary to compare the noise free projection with the class average, which includes a filtered version of this projection. After classification of Fourier-filtered data, the average particle of the class, including the particle in (c), is calculated and shown in (f). (d) is the wavelet-filtered version of (b). After classification of wavelet filtered particles, the average of the class, including the particle in (d), is shown in (e).

unit contains one penton subunit, one hexon and one-third copies of a connecting density joining neighbouring hexons. The size of 700 Å previously estimated from negatively stained images [29] was smaller, presumably due to specimen shrinkage caused by negative staining. The dimension of the hexon in the VP5-VP19C particle has an average diameter of ~145 Å and a height of ~150 Å. The importance of this reconstruction and its role in understanding the herpes-virus is discussed in detail in [16].

Discussion

The advantage of the wavelet transform over the Fourier transform is that individual wavelet functions are localized in real space. Fourier sine and cosine functions are not. This localization feature, along with wavelets localization of frequency, makes the use of wavelets more powerful for approximating data with sharp discontinuities, spikes [17]. It also helps in a number of useful applications such as data compression, features detection and noise removal in images. In wavelet analysis, the scale used to look at data plays a special role. Wavelet algorithms process data at different scales or resolutions, and the result in wavelet analysis is to see high and low resolution features.

This paper elucidates the fact that classification of simulated data gives some clear evidence of improvement of the class averages obtained by wavelet filtering in contrast

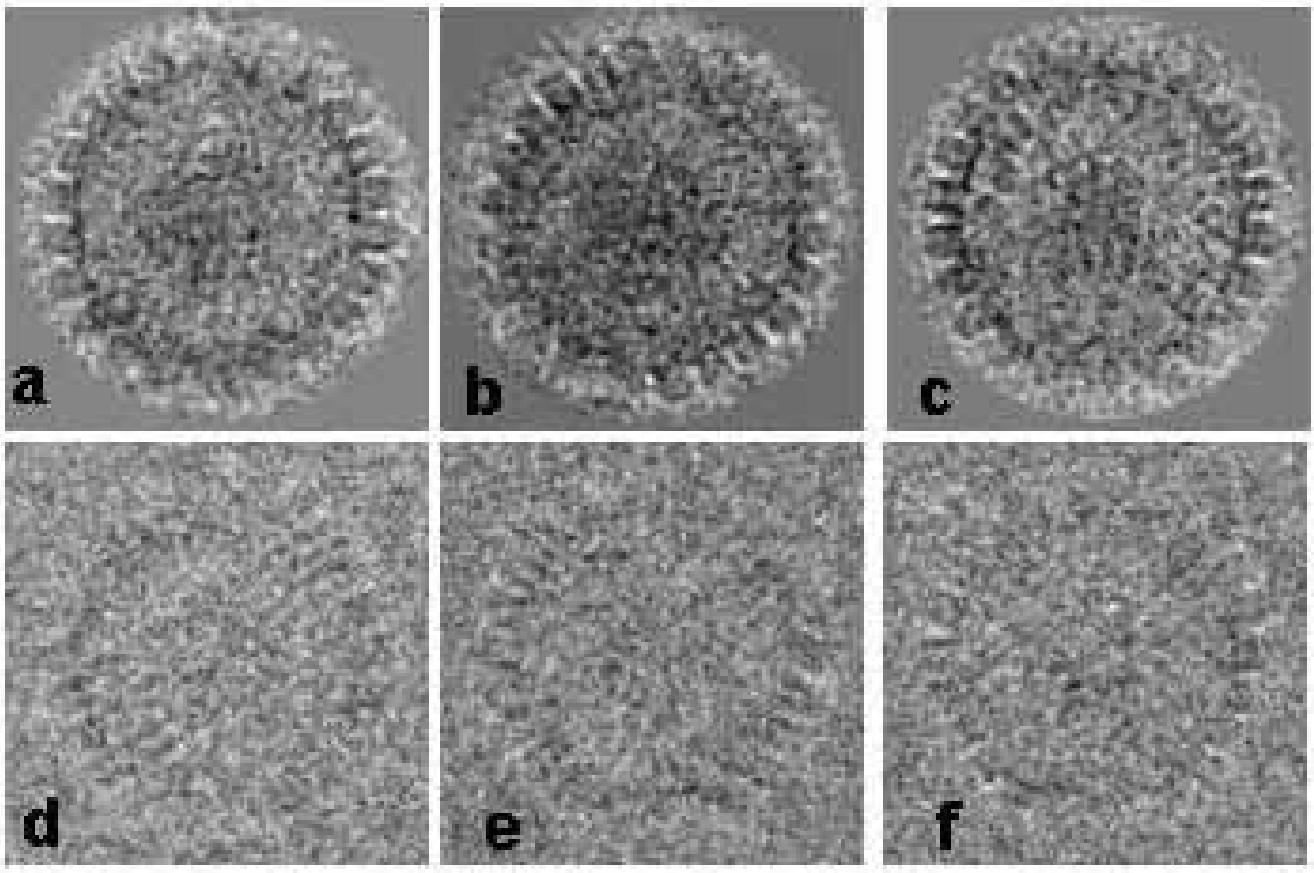

Figure 4
Classification of VP5-VP19C Data. This figure shows some class averages obtained after wavelet filtering of a set of VP5-VP19C particles (a, b, c). The corresponding class averages from the Fourier-filtered data are shown in (d, e, f).

to Fourier filtering. This improvement is due to the wavelet properties in preserving the subtle details and the sharp edges in the image. The accuracy of the MSA depend on the level of noise, it is also depend on the quality of the edges in the image, blurred edges and lack of details also disturb the classification[7,8]. Another important point is that before averaging, all particles in a class must be aligned to one another. This alignment procedure depends on the accuracy of the edges in the image. Blurred edges provide a bigger error in alignment, which provide more blurred class averages. All this arguments about misclassification and misalignment explains the blurring that exists in the class averages obtained from the Fourier-filtered data (Fig. 3-f).

The reconstruction of the VP5-VP19C particle is not possible without the wavelet filtering technique. Several attempts conducted with Fourier filtering on VP5-VP19C failed. The technique presented here provides a new way of pre-processing very noisy data. This technique has been applied to orientation determination of very noisy single particles without classification and the results obtained on herpes simplex virus are very promising [13].

Conclusions

Wavelet filtering and multi-resolution processing were proposed for very noisy data from electron cryomicroscopes. The wavelet filtering used along with MSA classification gives powerful results in class averaging compared with Fourier filtering which gives blurred class averages. The simulated data clearly proves the similarity between the original projection in figure 3-a, and the processed one with wavelet filtering technique in figure 3-e. An application of the multi-resolution processing method on real data, of recombinant virus VP5-VP19C, gives great results. The class averages obtained prove that wavelet filtering provides a more accurate classification scheme and produces much better class averages. This

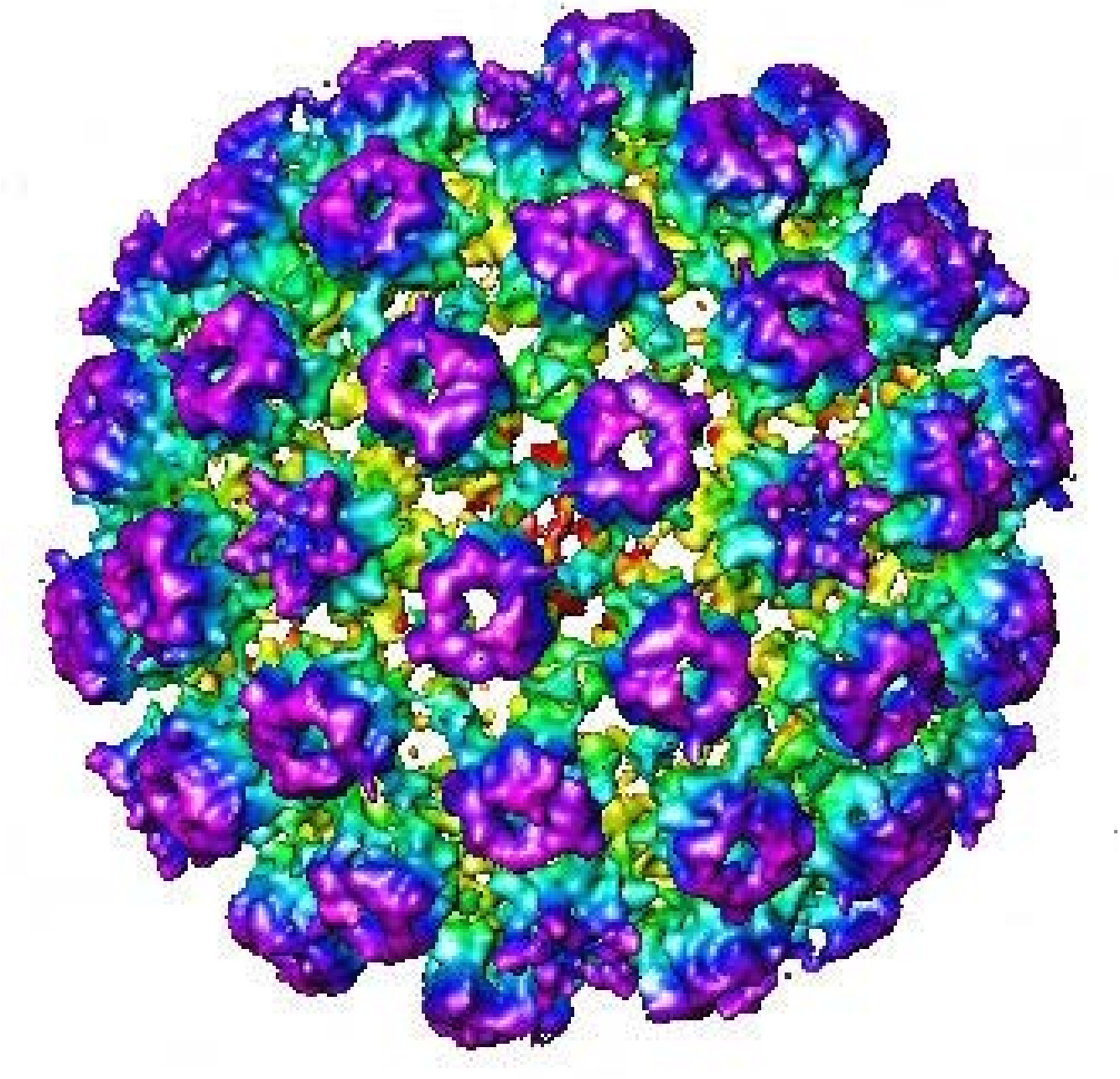

Figure 5
Surface view of VP5-VP19C 3D map. Three-dimensional reconstruction of the recombinant VP5-VP19C particle, the VP5-VP19C particle forms a T = 7 icosahedron with a diameter of 880 Å. The map is viewed along an icosahedral 2-fold axis. Iso-surface is displayed with a contour level of two standard deviations from the mean.

helps in determining more accurate orientation for each class average and consequently a better 3D structure.

Methods

Capsid preparation

VP5-VP19C particles were purified from cells infected with recombinant baculoviruses expressing only VP5 and VP19C as described previously [18].

Electron Cryomicroscopy

The ice-embedded VP5-VP19C particles were imaged with flood beam illumination. Microscope alignment, specimen assessment, and focusing were performed using a Gatan (Pleasanton, CA) 1 k × 1 k slow-scan charge-coupled device camera [19]. All micrographs were recorded at a magnification of 30,000× on Kodak SO163 film in a JEOL4000 electron cryomicroscope operating at 400 kV using a LaB_6 filament, under minimal dose conditions (~7 electrons/Å^2).

Image digitization and selection

Forty selected micrographs were digitized on a Zeiss SCAI scanner (Carl Zeiss, Englewood, Colorado) at a step size of 4.67Å/pixel. 1300 particle images (240 × 240 pixels) were selected automatically [5]. The image quality was assessed by evaluating the contrast transfer function rings visualized in the incoherently averaged Fourier transforms of particle images [20]. 700 particles from these micrographs with the first zeros of their contrast transfer functions between 1/20 - 1/24 Å^{-1} were used for further analysis.

Most of the subsequent computational steps were performed using IMAGIC-5 software [14] on an SGI (Silicon Graphics, Inc) Onyx2 supercomputer with 24 parallel processors.

Wavelet Bases Choice

An investigation to choose the best wavelet bases for electron cryomicroscopic images was performed. During this study, simulated and real electron cryomicroscopy images were used. Testing of the majority of the wavelet basis [17,21-24] existing in Matlab-5 software, has been made. The criterion used to determine the best wavelet base is one which optimizes the signal-to-noise ratio in a broad spectrum of spatial frequencies. The biorthogonal wavelets basis [25,26] especially the 3.5 basis in Matlab has yielded the best average signal-to-noise ratio in the range of the spatial frequency (1/100 - 1/8 Å^{-1}) relevant to our data analysis.

Wavelet filtering

In diverse fields from planetary science to molecular spectroscopy, scientists are faced with the problem of recovering a true signal from incomplete or noisy data. Wavelets help to solve this problem through a technique called wavelet shrinkage and thresholding methods [27] and multi-resolution by contrast modification filtering method [28].

Wavelet decomposition has the power to separate the signal into low-resolution information (approximation) and high resolution information (details). The high-resolution information contains the details of the image and the noise. The characteristic of the noise is randomly distributed in the image and does not have any structured information like (edges, contours, segment in horizontal, vertical or diagonal directions), the wavelets filters strengthen the structured information, all other components are down weighted, mainly the noise which has a random distribution in the object. This means that the high amplitude of gray values in the histogram of each component are mainly "details" [27,28] of the object, and the values close to zero are mainly due to the contribution of the noise. In other words, if the details are small they might be omitted without substantially affecting the main feature of the data set. The idea of thresholding then is to set to zero all coefficients less than a particular threshold. The histograms of the details components are used to set the threshold in each component in order to reduce that noise. Once those values are removed it means that those pixels will be dumped to zeros and will not contribute to the reconstructed image. In a multi-resolution processing [5] method the approximation component is used as the filtered image. The approximated image is smaller in size, which helps to accelerate the processing, but lower in resolution because all the details and noise have been removed from the image. The level of reduction of the original image, depends on the application, and on the resolution required. The following section describes the use of the multi-resolution processing for the 3D reconstruction of VP5-VP19C.

3D Reconstruction of VP5-VP19C

Wavelets filtering and classification

In order to determine the orientation of the particles accurately, two pre-processing steps are commonly used in order to enhance the signal-to-noise ratio of the particle. First is filtering of individual particle. The second step is the classification, which consists on grouping the particles having similar orientation in one class, and then averages them, to get a further noise reduction and restoration of some flexible part of the particle. This classification is obtained by MSA procedure in IMAGIC-5 software [14]. This step generally gives much better improvement of the signal-to-noise ratio of the particle, which helps the orientation determination algorithms to be more accurate. The accuracy of the classification process is noise dependent. For close-to-focus data (highly noisy data) a better filtering is needed before classification of particles

into different groups. A wavelet filtering is proposed as an alternative to the Fourier Gaussian filter, to be used before classification when the noise level is very high.

Low-resolution model

During this step, the average particles of the approximation components of wavelet decomposition (using biorthogonal filters) were used. Figure 2-b shows a wavelet approximation of figure 2-a. For comparison purpose, the Fourier-filtered version having the same sub-sampling (factor two in each direction) is shown in figure 2-c. The angular reconstitution method [29] was used to assign Euler angles to each class average. Then a 3D reconstruction was calculated using the exact-filter back-projection algorithm [30]. The orientations of individual particles were initially assigned as those of the class averages to which they belonged. They were then iteratively refined using projections computed from the preliminary reconstruction. The resolution of the model obtained from the wavelet approximation images is very low (37 Å).

Final reconstruction

The low-resolution model of the VP5-VP19C particles did not make use of all the information inherent in our raw image data. Therefore, the final map was reconstructed from the original VP5-VP19C particle images without wavelet approximation and class averaging [16]. First, the low-resolution model was scaled up to the same dimension as the original image (240 × 240). A wavelet filtering using the thresholding technique was applied to the original data, a threshold of one standard deviation was chosen as trade-off between detail retention, and noise suppression. Initial orientation for original size images were assigned from the approximation components. Projections were computed from the scaled model to refine the Euler angles for each of the particle images using the angular reconstitution technique. Consequently, an improved 3D reconstruction was computed and used for further angles and centre refinement. This reconstruction-refinement procedure was iterated for several rounds until no further significant improvement was obtained.

Acknowledgements

I specially thank Wah Chiu and Hong Zhou for interesting discussion; I thank also Frazer Rixon, Joanita Jakana for providing the VP5-VP19C and HSV-1capsid data, Angel Paredes and Kelechi Ogbuehi for improving the English of the manuscript. This work was supported by the National Institutes of Health (P41RR002250, R01AI38469) and Human Frontier Science Program.

References

1. Saad A, Ludtke S, Jakana J, Rixon F, Tsuruta H, Chiu W: **Fourier Amplitude decay of Electron cryomicroscopic images of single particles and effects on structure determination.** *J Struct Biol* 2001, **133:**32-42.
2. Thuman-Commike PA, Chiu W: **Automatic detection of spherical particles from spot-scan electron microscopy images.** *J Microsc Soc Amer* 1995, **1:**191-20.
3. Harauz G, Fong-Lochovsky A: **Automatic selection of macromolecules from electron micrographs by component labeling and symbolic processing.** *Ultramicroscopy* 1989, **31:**333-344.
4. Lata KR, Penczek P, Frank J: **Automatic particles picking from electron micrographs.** *Ultramicroscopy* 1995, **58:**381-391.
5. Saad A, Chiu W, Thuman-Commike P: **Multiresolution approach to automatic detection of spherical particles from electron cryomicroscopy images.** *Chicago, USA, IEEE SP Soc ICIP* 1998.
6. Zhou ZH: **High resolution three-dimensional electron cryomicroscopy and reconstruction of herpes simplex virus capsids.** *Houston, Baylor College of Medicine, Biochemistry* 1995.
7. van Heel M, Frank J: **Use of multivariate statistics in analysing the images of biological macromolecules.** *Ultramicroscopy* 1981, **6:**187-194.
8. van Heel M: **Multivariate statistical classification of noisy images(Randomly oriented biological macromolecules).** *Ultramicroscopy* 1983, **13:**165-184.
9. Crowther RA: **Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs.** *Phil Trans Roy Soc Lond B* 1971, **261:**221-230.
10. Thuman-Commike PA, Chiu W: **Improved common-line based icosahedral virus particle image orientation estimation algorithms.** *Ultramicroscopy* 1997, **68:**231-256.
11. Zhou ZH, Chiu W, Haskell K, Spears HJ, Jakana J, Rixon FJ, Scott LR: **Refinement of herpesvirus B-capsid structure on parallel supercomputers.** *Biophys J* 1998, **74:**576-588.
12. Baker TS, Cheng RH: **A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy.** *J Struct Biol* 1996, **116:**120-130.
13. Saad A, Chiu W: **Hierarchical wavelet projection matching for orientation determination of low contrast electron cryomicroscopic images of icosahedral virus particles.** *Istanbul Turky, ICASSP* 2000.
14. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M: **A new generation of the IMAGIC image processing system.** *J Struct Biol* 1996, **116:**17-24.
15. Rixon FJ, Addison C, McGregor A, Macnab SJ, Nicholson P, Preston VG, Tatman JD: **Multiple interactions control the intracellular localization of the herpes simplex virus type 1 capsid proteins.** *J Gen Virol* 1996, **77:**2251-2260.
16. Saad A, Zhou H, Jakana J, Chiu W, Rixon F: **Role of triplex and scaffolding proteines in Herpes simplex virus type 1 capsid formation suggested by structures of recombinant particles.** *J Virol* 1999, **73:**6821-6830.
17. Vetterli M, Herely C: **wavelets and filter banks: theory and design.** *IEEE Trans SP* 1992, **40:**2207-2232.
18. Rixon FJ, Addison C, McGregor A, Macnab SJ, Nicholson P, Preston VG, Tatman JD: **Multiple interactions control the intracellular localization of the herpes simplex virus type 1 capsid proteins.** *J Gen Virol* 1996, **77:**2251-2260.
19. Sherman MB, Brink J, Chiu W: **Performance of a slow-scan CCD camera for macromolecular imaging in a 400 keV electron cryomicroscope.** *Micron* 1996, **27:**129-139.
20. Zhou ZH, Hardt S, Wang B, Sherman MB, Jakana J, Chiu W: **CTF determination of images of ice-embedded single particles using a graphics interface.** *J Struct Biol* 1996, **116:**216-222.
21. Daubechies I: ***Ten lectures on wavelets.*** *SIAM* 1992.
22. Strang G, Nguyen T: **Wavelets and Filter Banks.** *Wellesley-Cambridge Press* 1997.
23. Mallat S: **Multi-frequency Channel Decompositions of images and wavelet Models.** *IEEE on ASSP* 1989, **37:**2091-2110.
24. Beylkin G, Coifman R, Rokhlin V: **Fast Wavelet transforms and numerical algorithms.** *Comm Pure Appl Math* 1991, **44:**141-183.
25. Cohen A, Daubechies I, Feauveau JC: **Biorthogonal Bases of Compactly supported Wavelets.** *Comm Pure Appl Math* 1992, **45:**485-560.
26. Phoong SM, Kim CW, Vaidyanathan , Ansari R: **A new class of two-channel biorthogonal filter P. P. banks and wavelets bases.** *IEEE Trans SP* 1997, **43:**649-665.
27. Donoho D: **De-noising by soft thresholding.** *IEEE Trans Inf Th* 1995, **41:**613-627.
28. Saad A, El-Assad S, Barba D: **Speckle filtering in SAR images by contrast modification, comparison with a large number of filters.** *Anal Telecom* 1996, **51:**233-244.

29. van Heel M: **Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction.** *Ultramicroscopy* 1987, **21:**111-124.
30. Harauz G, van Heel M: **Exact filters for general geometry three dimensional reconstruction.** *Optik* 1986, **73:**146-156.