Orientation Refinement of Virus Structures with Unknown Symmetry

Yongchang Ji,
School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, Florida, 32816

Dan C. Marinescu,
School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, Florida, 32816

Wei Zhang, and
Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907

Timothy S. Baker
Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907

Yongchang Ji: yji@cs.ucf.edu; Dan C. Marinescu: dcm@cs.ucf.edu; Wei Zhang: leila@bilbo.bio.purdue.edu; Timothy S. Baker: tsb@bilbo.bio.purdue.edu

Abstract

Structural biology, in particular the structure determination of viruses and other large macromolecular complexes leads to data- and compute-intensive problems that require resources well beyond those available on a single system. Thus, there is an imperative need to develop parallel algorithms and programs for clusters and computational grids. We present one of the most challenging computational problems posed by the three-dimensional structure determination of viruses, the orientation refinement.

1. Introduction and Motivation

Viruses are large macromolecules that cause a variety of human, animal, and plant diseases. A virus consists of a nucleic acid genome, a protein shell or *capsid*, and sometimes a membrane or *envelope*, which can enclose or be enclosed by the capsid.

Viruses differ in terms of the genome, shape, size, molecular mass, and the number of proteins in the capsid. The genome consists of either single or double-stranded RNA or DNA. There are spherical, helical and more complex shapes of viruses. Viruses range in size from as little as 150 to 2,000 Å or more. The molecular masses of viruses differ; for example, *picornaviruses* have a mass of about 8.5×10^6 daltons about 30% of which is due to one RNA molecule of about 8,000 nucleotides. In contrast, alphaviruses like Sindbis virus, have a mass of about 50×10^6 daltons including a genome consisting of about 12,000 nucleotides [25].

Because of symmetry of the protein shell, a relatively small number of identical building blocks recognize each other and are able to assemble together spontaneously. The principle

of *genetic economy* requires that the shell be built out of multiple copies of identical units; the amount of genetic information, thus the size of the genome, is considerably smaller for a symmetric virus particle. For example, one of the smallest known viruses, the *satellite tobacco necrosis* virus has a diameter of 180 Å, a protein shell of 60 sub-units, and its RNA is very small, about 1, 120 nucleotides [3].

Viruses infect healthy cells by attaching to them, then delivering their nucleic acid into the cell. A virus has to locate a specific docking site on a host cell. The information about the structure of the capsid is critical in determining the pharmaceutical compounds that can block the virus binding site. Thus, the structure of viruses is not only of scientific interest but it has the potential to lead to discovery of antivirals for the prevention and treatment of plant, animal, and human diseases.

2. Structure Determination in Cryo-TEM

Electron microscopy is a widely used method to obtain the structure of viruses at low to moderate resolution [2], [7], [8], [9], [11], [15], [23]; X-ray crystallography has traditionally been used for high resolution structure determination at resolutions of 3 Å or better. However crystallization of large macromolecules like viruses is extremely challenging, and hence there is a desire to push the limits of electron microscopy and extend the resolutions of structure determination to the 5 Å range.

In cryo-transmission electron microscopy (cryo-TEM) a solution containing virus particles is vitrified at liquid nitrogen temperatures and the sample is placed inside the microscope where it is irradiated with an electron beam that forms an image on film or on a CCD. An entire micrograph consists of real images(2D projections) of many identical virus particles frozen in the sample in different orientations [22].

Reconstructing the 3D image of the virus from the 2D projections is conceptually similar to Computed-Aided Tomography (CAT) [13], [14], [16]. The notable difference is that the orientations and centers of the 2D images are known in CAT (the patient is in a fixed position and the images are taken at known angles of the X-ray source) while in cryo-TEM virus particles are frozen in the solution at random orientations. Moreover, we assume that all virus particles frozen in the sample are *identical* thus we have many 2D projections of a single object whose 3D electron density we wish to reconstruct. In CAT we have multiple 2D images of a single object.

The procedure for 3D structure determination in cryo-TEM consists of the following steps:

Step A Extract individual particle projections from micrographs and identify the center of each projection.

Step B Determine the orientation of each projection.

Step C Carry out the 3D reconstruction of the electron density of the macromolecule.

Step D Dock an atomic model into the 3D electron density map.

Steps B and C are executed iteratively until the 3D electron density map cannot be further improved at a given resolution; then the resolution is increased gradually. The number of iterations for these steps is in the range of hundreds and one cycle of iteration for a medium size virus may take several days. Typically it takes months to obtain a high resolution electron density map. Then Step D of the process can be pursued.

Best results in obtaining high resolution electron density maps are often obtained in the case of highly symmetrical particles such as icosahedral viruses because the high symmetry leads to redundancies in the Fourier transform data and that, in turn, aids the orientation search process.

It was estimated that approximately 2000 particle images are necessary for the reconstruction of a virus with a diameter of 1000 Å at 10 Å resolution [24], and recent results at 7–9 Å resolution for the Hepatitis B virus capsid [4], [5] have confirmed this estimate.

In this paper we are concerned only with step B of the process described above, i.e., the orientation refinement. To determine the orientation of a view of a virus particle extracted from a micrograph, we project the current reconstructed electron density at different orientations and then compare the experimental image with the calculated projections. The unknown orientation θ, φ, ω, see Figure 1, is then determined to be the orientation of the calculated projection of the best fit.

When the symmetry is known, the search process for orientation determination is restricted to a relatively small angular domain (the asymmetric unit as depicted in Figure 1(a)). We consider the more challenging case when the information regarding the symmetry of the virus particle is not available. The advantage of the method described in this paper is that one could use it not only to determine the structure of the symmetric protein shell, but also the structure of asymmetric objects. Moreover, if the virus exhibits any symmetry this method allows us to determine its symmetry group.

Previous parallel orientation refinement programs such as the one reported in [17] took advantage of the embarrassingly parallel nature of the traditional algorithm. Our algorithm is radically different. It does not make any assumption about the symmetry of the object, it performs calculations in Fourier Space, it is based upon a multi-resolution search, and a sliding window mechanism discussed for the first time in [1]. The orientation refinement algorithm is used in conjunction with our 3D reconstruction algorithm in Cartesian coordinates for objects without symmetry [18], [20], [21].

3. Problem Formulation

We use the following terms and notations:

- D is the current version of electron density map of size l^3.

- $\mathcal{D} = DFT(D)$ is the 3D Discrete Fourier Transform (DFT) of the electron density map.

- $\mathcal{C}$ is a set of 2D planes of $\mathcal{D}$ obtained by interpolation in the 3D Fourier domain.

- $\mathcal{E} = \{\mathcal{E}^1, \ldots \mathcal{E}^q, \ldots \mathcal{E}^m\}$ with $1 \leq q \leq m$ is a set of m experimental views; each view is of size $l \times l$ pixels.

- $\mathcal{F} = \{\mathcal{F}^1, \ldots \mathcal{F}^q, \ldots \mathcal{F}^m\}$ is the set of 2D DFTs of experimental views. Here $\mathcal{F}^q = DFT(\mathcal{E}^q)$ for $1 \leq q \leq m$.

- $O_{init} = \{O^1_{init}, \ldots, O^q_{init} \ldots, O^m_{init}\}$ is the set of initial orientations, one for each view. $O^q_{init} = \{\theta^q_{init}, \phi^q_{init}, \omega^q_{init}\}$

- $O_{refined} = \{O^1_{refined}, \ldots, O^q_{refined} \ldots, O^m_{refined}\}$ is the set of refined orientations, one for each view. $O^q_{refined} = \{\theta^q_\mu, \phi^q_\mu, \omega^q_\mu\}$

- P is the number of nodes available for program execution. Parallel I/O could reduce the I/O time but in our algorithm we do not assume the existence of a parallel file system. To avoid contention, a master node typically reads an entire data file and distributes data segments to the nodes as needed.

- Given a 3D lattice D of size l^3 we define a *z-slab* of size $zslab_{size}$ to be a set of consecutive $zslab_{size}$ xy-planes. One can similarly define *x-slab* and *y-slab*.

- The resolution of the electron density map is denoted by r_{map}. In the reconstruction process we use only the Fourier coefficients up to $\frac{1}{r_{map}}$.

- The angular resolution is denoted by $r_{angular}$.

Given: (1) a set of m views and (2) the electron density map, the goal is to find the orientation of each view.

Several methods including the method of "common lines", [2] can be used to this end. Here we describe a procedure for the refinement of orientations that is less sensitive to the noise caused by experimental errors. The basic idea is to project the electron density at known angles and then to compare each experimental view with the calculated projection. Once we define the distance between an experimental view and a calculated projection, the goal of the search is to identify the calculated projection at the minimum distance from the experimental view. The procedure does not make any assumptions about the symmetry, but can detect symmetry if one exists.

We are only concerned here with the orientation refinement, in other words we are given a rough estimation of the orientation, say at 3° angular resolution and our goal is to reach a resolution of say 0.01° or better.

This paper describes an algorithm where the search is conducted in the Fourier domain. We first perform a 2D Discrete Fourier Transform (DFT) of the experimental view, $\mathcal{F}$, apply a Contrast Transfer Function (CTF) correction to it, and then compare it with a cut at a precise orientation through the 3D DFT of the electron density map, $\mathcal{C}$.

A correction of the experimental data used for the 3D structure determinations is necessary. The relationship between the electron image of a specimen and the specimen itself isin part

affected by the microscope CTF [2]. The Defocusing, which is used to enhance phase contrast and thereby enables the visualization of unstained specimens [2], must be compensated for in the reconstruction, in order to achieve a reliable representation of the structure. The CTF is an oscillatory function that produces phase reversal and attenuates amplitudes in the DFT of a TEM image. The effects of the transfer function become more pronounced at progressively higher resolutions. One can correct the transfer function by means of a variety of filtering methods [12].

In this algorithm the distance between two $l \times l$ arrays of complex numbers:

$$\mathscr{F} = [a_{j,k} + i b_{j,k}]_{1 \le j,k \le l},$$

and $\mathscr{C} = [c_{j,k} + i d_{j,k}] 1_{\ j,k \ l}$, with $i = \sqrt{-1}$. is computed as:

$$d(\mathscr{F}, \mathscr{C}) = \frac{1}{l^2} \sum_{j=1}^{l} \sum_{k=1}^{l} \sqrt{[a_{j,k} - c_{j,k}]^2 + [b_{j,k} - d_{j,k}]^2}.$$

To give more weight to higher frequency components at higher resolution (large radius in the Fourier domain) while computing the distance, we can apply a weighting function $wt(j, k)$ to $d(\mathscr{F}, \mathscr{C})$.

To determine the distance, $d(\mathscr{F}, \mathscr{C})$ at a given resolution we use only the Fourier coefficients up to $\frac{1}{r_{map}}$ thus the number of operations is reduced accordingly.

The algorithm is embarrassingly parallel, each experimental view can be processed independently by a different processor. But the 3D electron density map and its DFT can be very large; the database of calculated views could require several TBytes of data and the electron density may need several Gbytes of storage.

The size of the search space $\mathscr{P}$ is very large; if the initial orientation of an experimental view, $\mathscr{E}^q$ is given by $O^q = \{\theta^q, \varphi^q, \omega^q\}$ then the cardinality of the set $\mathscr{P}$ is:

$$|\mathscr{P}| = \frac{\theta_{max}^q - \theta_{min}^q}{r_{angular}} \times \frac{\phi_{max}^q - \phi_{min}^q}{r_{angular}} \times \frac{\omega_{max}^q - \omega_{min}^q}{r_{angular}}.$$

This step requires $\mathcal{O}(l^2 \times |\mathscr{P}|)$ arithmetic operations. For example, if $r_{angular} = 0.1°$ and the search range is from 0 to π for all three angles, then the size of the search space is very large: $|\mathscr{P}| = (1800)^3 = 5.832 \times 10^9$.

Figure 1(b) shows that the corresponding size for an icosahedral particle at 3° angular resolution consists of only 51 calculated views; at 0.1° the size of the search space is about 4, 000 calculated views, [2]. Thus, for an asymmetric particle the size of the search space increases by *six (6) orders of magnitude* compared with an icosahedral particle !! Moreover, when comparing two views of an icosahedral virus particle, a calculated and an

experimental one, we could use only a shell of thickness corresponding to the capsid, rather than the entire 2D image.

4. A SlidingWindow Multi-Resolution Parallel Orientation Refinement Algorithm

The orientation refinement is a *multi-resolution* process. Typically, we carry out several refinement steps at different angular resolutions, e.g., one at $r_{angular} = 1°$ followed by one at $r_{angular} = 0.1°$, one at $r_{angular} = 0.01°$, and finally one at $r_{angular} = 0.002°$. The advantage of this approach is clearly illustrated by an example: assume that the initial value is say $\theta = 65°$, the search domain is $60°$ to $70°$ and we require an angular resolution of $0.002°$. A one step search would require 5000 matching operations versus 35 for a multi-resolution matching. The orientation of an experimental view is given by three angles (θ, φ, ω), therefore the multi-resolution approach reduces the number of matching operations for a single experimental view by almost four orders of magnitude. As pointed out earlier, several thousand experimental views are needed for the reconstruction of the 3D electron density map at a high resolution. A matching operation consists of two steps (1) construct a cut into $\mathscr{D}$ with a given orientation and (2) compute the distance between the 2D DFT of the experimental view, $\mathscr{F}^q$, and the cut.

The search domain at a given angular resolution is extended whenever the best fit occurs near the edge of the search domain. This *sliding-window* approach increases the number of matching operations, but at the same time improves the quality of the solution. A similar strategy is used for refining the centers of the experimental images.

The orientation refinement algorithm consists of the following steps:

Step a. Construct $\mathscr{D}$, the 3D Discrete Fourier Transform, DFT, of the electron density map.

> **Step** (a.1) The master node reads all *z-slabs* of the entire electron density map D.
>
> **Step** (a.2) The master node sends to each other node a *z-slab* of the electron density map D of size $t_{slab} = \lceil \frac{l}{P} \rceil$.
>
> **Step** (a.3) Each node carries out a 2D DFT calculation along the x- and y-directions on its *z-slab*.
>
> **Step** (a.4) A global exchange takes place after the 2D DFT calculation and each node ends up with a *y-slab* of size t_{slab}.
>
> **Step** (a.5) Each node carries out a 1D DFT along the z-direction in its *y-slab*.
>
> **Step** (a.6) Each node broadcasts its *y-slab*.

After the `all-gather` operation each node has a copy of the entire $\mathscr{D}$. To perform calculations at resolution r_{map} we only keep a subset of the $\mathscr{D}$, within a sphere of radius $\frac{1}{r_{map}}$.

This step requires a total of $\mathscr{O}(l^3 \times log_2 l)$ arithmetic operations and $\mathscr{O}(l^3)$ words of memory in each node.

Step b. Read in groups of $m' = \lceil \frac{m}{P} \rceil$ views, $\mathscr{E}$, from the file containing the 2D views of the virus, and distribute them to all the processors. The amount of space required to store the experimental views on each processor is: $m' \times (b \times l^2)$ with b the number of bytes per pixel. In our experiments $b = 2$.

Step c. Read orientation file containing the initial orientations of each view, $O_{init} = \{O^1_{init}, \ldots, O^q_{init} \ldots, O^m_{init}\}$. Distribute the orientations to processors such that a view $\mathscr{E}^q$ and its orientation O^q_{init}, $1 \quad q \quad m$, are together.

At a given angular resolution we perform the following operations for each experimental view $\mathscr{E}^q$, $1 \quad q \quad m$:

Step d. Compute $\mathscr{F}^q$. Each processor carries out the transformation of the m' views assigned to it.

This step requires $\mathscr{O}(l^2 \times log_2 l)$ arithmetic operations for each experimental view and $\mathscr{O}(l^2)$ words of memory for the data.

Step e. Perform the CTF-correction of the DFT of each view $\mathscr{F}^q$, $1 \quad q \quad m$.

Note that the views originated from the same micrograph have the same CTF. This step requires $\mathscr{O}(l^2)$ operations for each experimental view.

Step f. Given:

i. the experimental view $\mathscr{E}^q$ with the initial orientation $O^q_{init} = \{\theta^q_{init}, \phi^q_{init}, \omega^q_{init},$

ii. the search domain,

$$0 \leq \theta^q_{min} \leq \theta^q_s \leq \theta^q_{max} \leq \pi, \; 0 \leq \phi^q_{min} \leq \phi^q_s \leq \phi^q_{max} \leq \pi, \; 0 \leq \omega^q_{min} \leq \omega^q_s \leq \omega^q_{max} \leq \pi$$

Construct a set of 2D-cuts of $\mathscr{D}$, the 3D-DFT of electron density map by interpolation in the 3D Fourier domain. Call $\mathscr{C}^q = \{\mathscr{C}^q_1, \mathscr{C}^q_2, \ldots \mathscr{C}^q_s, \ldots \mathscr{C}^q_w, \}$ the set of planes spanning the search domain for $\mathscr{E}^q$. Call $O^q_s = \{\theta^q_s, \phi^q_s, \omega^q_s\}$ the orientation of the cut $\mathscr{C}^q_s$.

Call w the number of calculated cuts in $\mathscr{D}$ for a given angular search range and angular resolution. $w = w_\theta \times w_\varphi \times w_\omega$. Typical values are $w_\theta = w_\varphi = w_\omega = 10$, thus $w = 1000$.

This step requires $\mathscr{O}(w \times l^2)$ arithmetic operations for each experimental view.

Step g. Determine the distance of $\mathscr{F}^q$ to every $\mathscr{C}^q_s$,

$$d^q_s = d(\mathscr{F}^q, \mathscr{C}^q_s) \text{for} 1 \leq s \leq w.$$

This step requires $\mathcal{O}(w \times l^2)$ arithmetic operations for each experimental view.

Step h. Compute the minimum distance:

$$d_\mu^q = min\{d_1^q, d_2^q, \ldots, d_w^q\}.$$

Call O_μ^q the orientation of the cut $\mathscr{C}_\mu^q$,

$$O_\mu^q = \{\theta_\mu^q, \phi_\mu^q, \omega_\mu^q\}.$$

This step requires $\mathcal{O}(w)$ arithmetic operations for each experimental view.

Step i. If any of the three angles corresponding to this minimum distance cut, $O_\mu^q = (\theta_\mu^q, \phi_\mu^q, \omega_\mu^q)$ is near the edge of the original search domain defined in step (f), redefine the search domain. Make O_μ^q the center of the new search domain, and repeat steps (f), (g), and (h).

Call n_{window} the number that we slide the window. Then the total number of operations required for each view in steps (f), (g), and (h) is: $\mathcal{O}(n_{window} \times w \times l^2)$.

Step j. Assign to experimental view $\mathscr{E}^q$ the orientation of this minimum distance cut, O_μ^q.

Step k. Refine the position of the center of the 2D DFT. Move the center of $\mathscr{E}^q$, $(x_{center}^q, y_{center}^q)$ within a box of size $2\delta_{center}$ using the current orientation and determine the best fit with the minimum distance cut $\mathscr{C}_\mu^q$.

For each new value of the center $(x_{center,i}^q, y_{center,i}^q)$, determine the distance to $\mathscr{C}_\mu^q$:

$$d_{\mu,i}^q = d(\mathscr{E}_i^q, \mathscr{C}_\mu^q)$$

Find the minimum distance:

$$d_{\mu,opt}^q = min\{d_{\mu,1}^q, d_{\mu,2}^q, \ldots d_{\mu,n_{center}}^q\}$$

where n_{center} is the number of center locations considered. For example, if we use a 3×3 box $n_{center} = 9$.

If the $(x_{center,opt}^q, y_{center,opt}^q)$ is near the edge of the search box redefine the search box. Make $(x_{center,opt}^q, y_{center,opt}^q)$ the center of a new search box, and repeat step k. Then, the total number of operations required for each view in this step is: $\mathcal{O}(n_{center} \times \delta_{center}^2 \times l^2)$.

Step l. Correct $\mathscr{E}^q$ to account for the new center.

Step m. Wait for all nodes to finish processing at a given angular resolution.

Step n. Repeat the computation for the next angular resolution until the final angular resolution is obtained. Then: $O_{refined} = \{O^1_{refined}, \ldots, O^q_{refined} \cdots, O^m_{refined}\}$ is the set of refined orientations, one for each view. $O^q_{refined} = \{\theta^q_\mu, \phi^q_\mu, \omega^q_\mu, x^q_{center,opt}, y^q_{center,opt}\}$

Step o. Write the refined orientation file.

As pointed in Section 3 the structure determination is an iterative process. Given a resolution r_{map} we use the refined orientations and the new centers to reconstruct the electron density map. The new map is used again for another step of orientation refinement and the process continues until we cannot further refine the structure at that particular resolution. Then we increase the resolution and repeat the entire procedure.

5. Experimental Results

Our objectives were threefold: (a) to verify the correctness of our results, (b) to determine our ability to increase the resolution of the structure determination using the new algorithms, and (c) to obtain some indication about the performance of our programs.

We present results regarding two virus structures, those of Sindbis and reo viruses [10]. The experimental data for the Sindbis virus consist of 7,917 views of 221×221 pixels each; the highest resolution obtained using existing algorithms is 11.2 Å. The experimental data for the reo virus consist of 4,422 views of 511×511 pixels each; the highest resolution obtained using existing algorithms is 8.6 Å.

We used our new algorithm to refine the orientation of these data for angular resolutions $r_{angular} = 1°$, $0.1°$, $0.01°$ and $0.002°$, and with center resolution $\delta_{center} = 1$ pixel, 0.1 pixel, 0.01 pixel and 0.002 pixel. In our calculations we reached a resolution of 10.0 Å for the Sindbis virus and 8.0 Å for the reo virus, see Figures 5 and 6.

The first objective was achieved by comparing our results with those using existing algorithms. Figure 2 shows the cross sections 102–103 of the 3D electron density map of Sindbis virus, reconstructed with its old orientation and with our new refined orientation respectively. Figure 3 shows the 3D density map of Sindbis virus, reconstructed with its old orientation and with our new refined orientation respectively. Though differences in the two maps are difficult to visualize directly in low magnification views, high magnification views do reveal more details in the new density map.

To test the resolution achieved, we use the procedure illustrated in Figure 4. After the last step of the orientation refinement at a given resolution we compute two 3D reconstructions, one using only odd numbered experimental views and the other, even numbered views. Then we determine the correlation between the two maps. Figure 5 shows a plot of the correlation coefficients for the new reconstructed maps compared with the one based upon previously determined orientations for the Sindbis virus. Figure 6 shows a plot of the correlation coefficients for the new reconstructed maps compared with the one based upon previously determined orientations for the reo virus.

Figures 5 and 6 indicate that the new orientation refinement method gives higher correlation coefficients and hence enables us to reconstruct electron density maps at higher resolution.

To determine the highest resolution we examine the plot of the correlation coefficient versus resolution and determine the crossing point of the graph and the 0.5 line. A correlation coefficient higher than 0.5 gives a conservative estimate of the final resolution of the entire density map. Figure 5 shows that the graph corresponding to the new method crosses the 0.5 line at 10.0 Å versus 11.2 for the old method. Figure 6 shows that the correlation coefficient for new method goes below 0.5 at 8.0 Å while the one for the old method does the same at 8.7 Å.

Finally, we evaluated the performance of our new algorithms. We tested our program on a 64-node IBM SP2 system, with each node having four processors and 2 GB of memory. The four processors in each node share the node's main memory and communicate using MPI.

Our parallel orientation refinement and parallel 3D reconstruction program run on 16 processors for these two data sets. Tables 1 and 2 show the time for different steps of the orientation refinement process for one iteration of structure determination for the Sindbis and the reo virus. Tables 1 and 2 show that 99% of the time for orientation refinement, is spent in matching the experimental views with the 2*D*-cuts of $\mathscr{D}$. The execution time for 3D reconstruction for the Sindbis virus is 4,575 seconds and for the reo virus is 7,933 seconds. The 3D reconstruction time represents less than 5% of the total time per cycle. Moreover we see in some instances the sliding window mechanism activated, for example, at 0.01° instead of 9 matchings (search range) we needed 15 for the Sindbis virus and 11 for the reo virus.

An interesting question is: How fine the angular resolution should be used and what is its effect on the center location; does it make any sense to refine the angles beyond 0.01°?

6. Conclusions

In this paper we discuss a novel algorithm for the refinement of orientations of individual virus particle projections for asymmetric particles and analyze its computational complexity. The algorithm can be used to determine the symmetry group of a symmetric particle and for the 3D reconstruction of particles exhibiting no symmetry or any symmetry.

The algorithm presented in this paper was designed for a distributed memory parallel architecture. On a shared memory system we would need one copy of the electron density map and of its 3D DFT. On a distributed memory system we choose to replicate the electron density map and its 3D DFT on every node because we wanted to reduce the communication costs. The alternative is to implement a shared virtual memory where 3D "bricks" of the electron density or its DFT are brought on demand in each node when they are needed, a strategy presented in [6].

We tested our programs using experimental data gathered for symmetric virus particles because we wanted to compare the quality of our solution with the one produced by programs that exploit the known symmetry of the protein shell of a virus particle. Such programs have been used for many years by the structural biology community. Our results

indicate that our algorithm provides better quality solutions than they currently obtained. In addition to enabling structural biologists to study asymmetric systems, we were able to refine two structures to 10.0 Å and 8.0 Å versus 11.2 Å and 8.6 Å resolution respectively. We are now running additional cycles of refinement on both data sets in the hope of extending the resolution of both structures.

Acknowledgments

References

1. Baker, TS.; Martin, IMB.; Marinescu, DC. CSD-TR #97–055. Department of Computer Sciences, Purdue University; 1997. A parallel algorithm for determining orientations of biological macromolecules imaged by electron microscopy.

2. Baker TS, Olson NH, Fuller SD. Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron microscopy. Microbiology and Molecular Biology Reviews. 1999; 63(4):862–922. [PubMed: 10585969]

3. Branden, C.; Tooze, J. Introduction to protein structure. New York and London: Garland Publishing; 1991.

4. Böttcher B, Wynne SA, Crowther RA. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature (London). 1997; 386:88–91. [PubMed: 9052786]

5. Conway JF, Cheng N, Zlomick A, Wingfield PT, Stahl SJ, Steven AC. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. Nature (London). 1997; 386:91–94. [PubMed: 9052787]

6. Cornea-Hasegan MAD, Marinescu C, Zhang Z, Rice JR, Lynch RE, Rossmann MG. Macromolecular electron density averaging on distributed memory MIMD systems. Concurrency: Practice and Experience. 1993; 5(8):635–657.

7. Crowther RA, DeRosier DJ, Klug A. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc. Roy. Soc. London. 1970; A 317:319–340.

8. DeRosier DJ, Klug A. Reconstruction of three-dimensional structures from electron micrographs. Nature. 1967; 217:130–134.

9. DeRosier DJ, Moore PB. Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. J. Molec. Biol. 1970; 52:355–369. [PubMed: 5485914]

10. Dryden KA, Farsetta DL, Wang G, Keegan JM, Fields BN, Baker TS, Nibert DL. Internal \ structures containing transcriptase-related proteins in top component particles of mammalian orthoreovirus. Virology. 1998; 245:33–46. [PubMed: 9614865]

11. Frank, J. Three-dimensional electron microscopy of macromolecular assemblies. Academic Press; 1996.

12. Gonzalez, RC.; Woods, RE. Digital image processing. Addison-Wesley Publishing Company, Inc; 1993.

13. Gordon R. Three-dimensional reconstruction from projections: A review of algorithms. Intern. Rev. of Cytology. 1974; 38:111–151.

14. Grangeat, P.; Amans, J-L., editors. Three-dimensional image reconstruction in Radiology and Nuclear Medicine. Kluwer Academic Publishers; 1996.

15. van Heel M, Growen B, Matadeen R, Orlova E, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A. Single-particle electron cryo-microscopy: towards atomic resolution. Quarterly Review of Biophysics. 2000; 33(4):307–369.

16. Herman, GT., editor. Image reconstruction from projections, implementation and applications. Springer-Verlag; 1979.

17. Johnson CA, Weisenfeld NI, Trus BL, Conway JF, Martino RL, Steven AC. Orientation determination in the 3D reconstruction of icosahedral viruses using a parallel computer. CS & E. 1994:555–559.

18. Lynch, RE.; Marinescu, DC.; Lin, H.; Baker, TS. Proc. IPPS/SPDP. IEEE Press; 1999. Parallel algorithms for 3D reconstruction of asymmetric objects from electron micrographs; p. 632-637.

20. Marinescu DC, Ji Y, Lynch RE. Space-Time Tradeoffs for Parallel 3D Reconstruction Algorithms for Virus Structure Determination. Concurrency and Computation: Practice and Experience. 2001; 13:1083–1106.

21. Marinescu, DC.; Ji, Y.; Marinescu, GM. Proc. Int. Symp.on Parallel and Distributed Computing. Vol. XI. Analele Stiintifice ale Universitatii Al.I.Cuza, Iasi, ISSN 1224–2268; 2002. Grid Computing and Applications to Computational Biology; p. 11-46.

22. Martin IM, Marinescu DC, Baker TS, Lynch RE. Identification of spherical particles in digitized images of entire micrographs. J. of Structural Biology. 1997; 120:146–157.

23. Ramanchandran GN, Lakshminarayanan AV. Three dimensional reconstructions from radiographs and electron micrographs: Application of convolution instead of Fourier transforms. Proc. National Academy of Sciences. 1971; 68:2236–2240.

24. Rossmann MG, Tao Y. Cryo-electron-microscopy reconstruction of partially symmetric objects. J. Structural Biology. 1999; 125:196–208.

25. Strauss JH, Strauss EG. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. 1994; 58:491–562. [PubMed: 7968923]
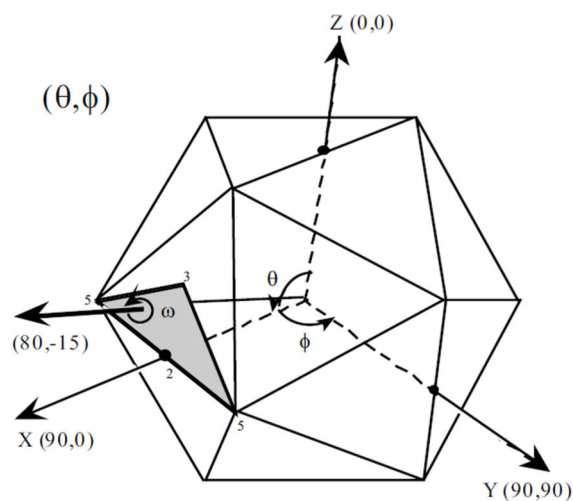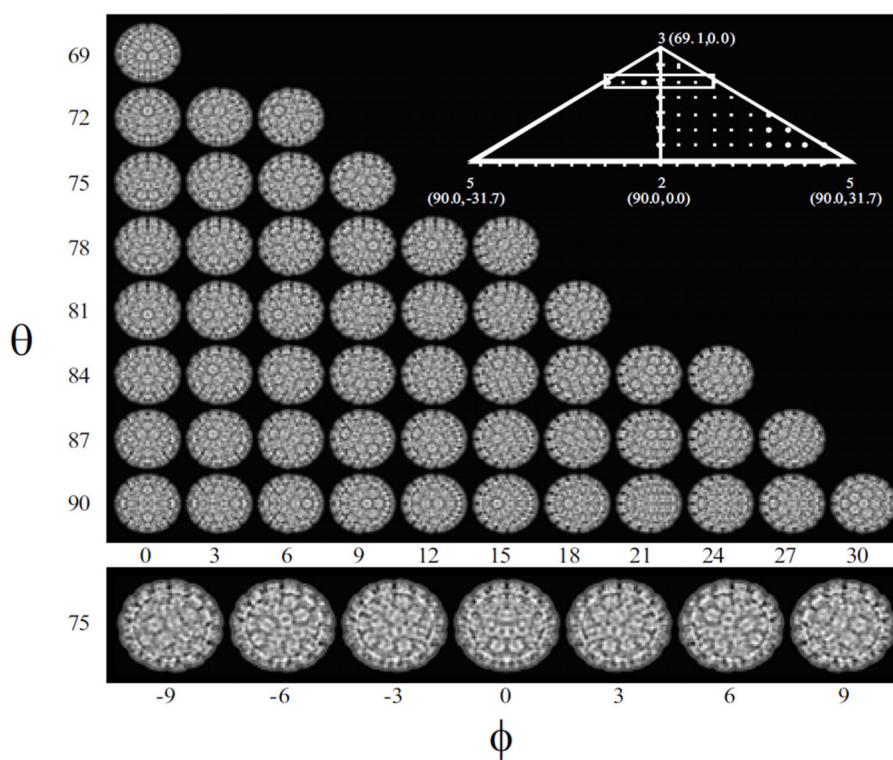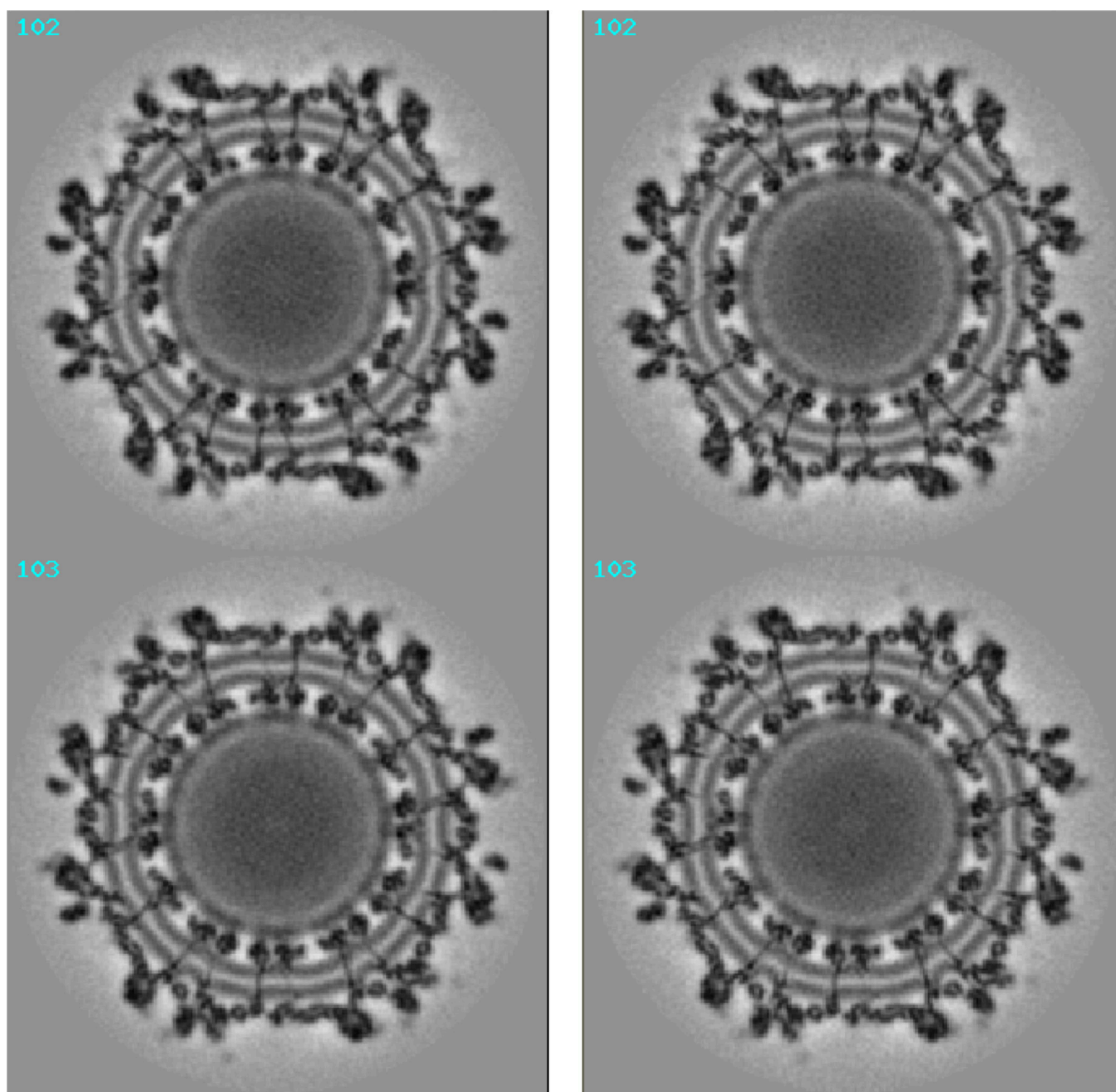
Figure 1.
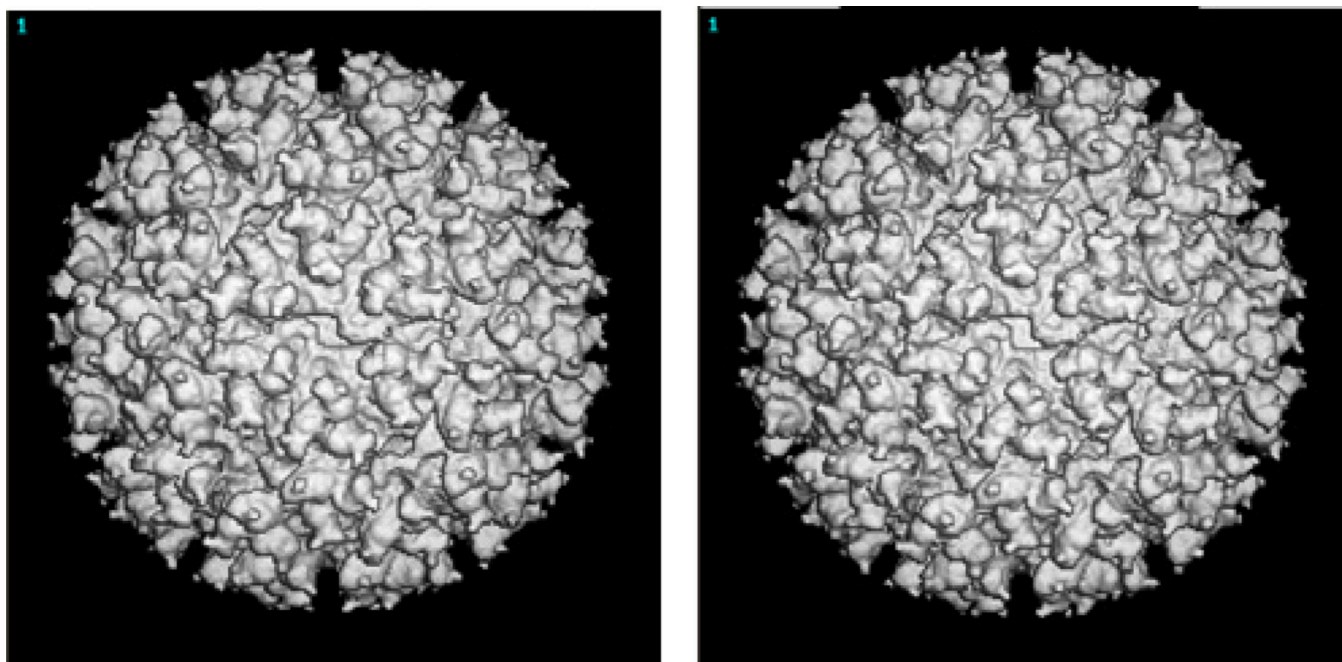(a) The three angles used to characterize the orientation of a view. (b) The set of calculated views for an icosahedral virus at a 3° angular interval.
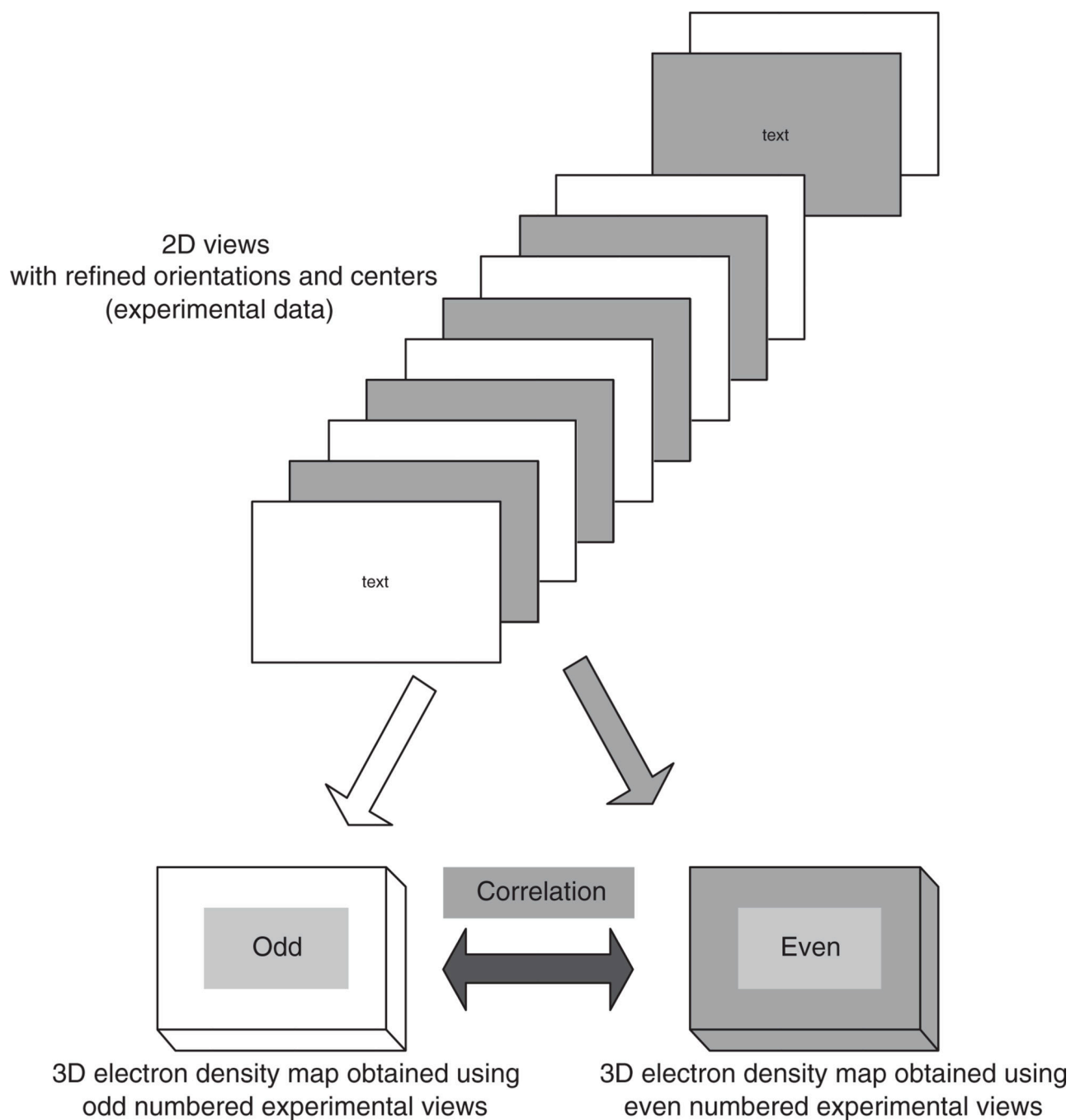
Figure 2.
Cross sections 102–103 of the 3D electron density map of Sindbis virus obtained with its old orientation (left) and the orientation produced by our new algorithm (right).
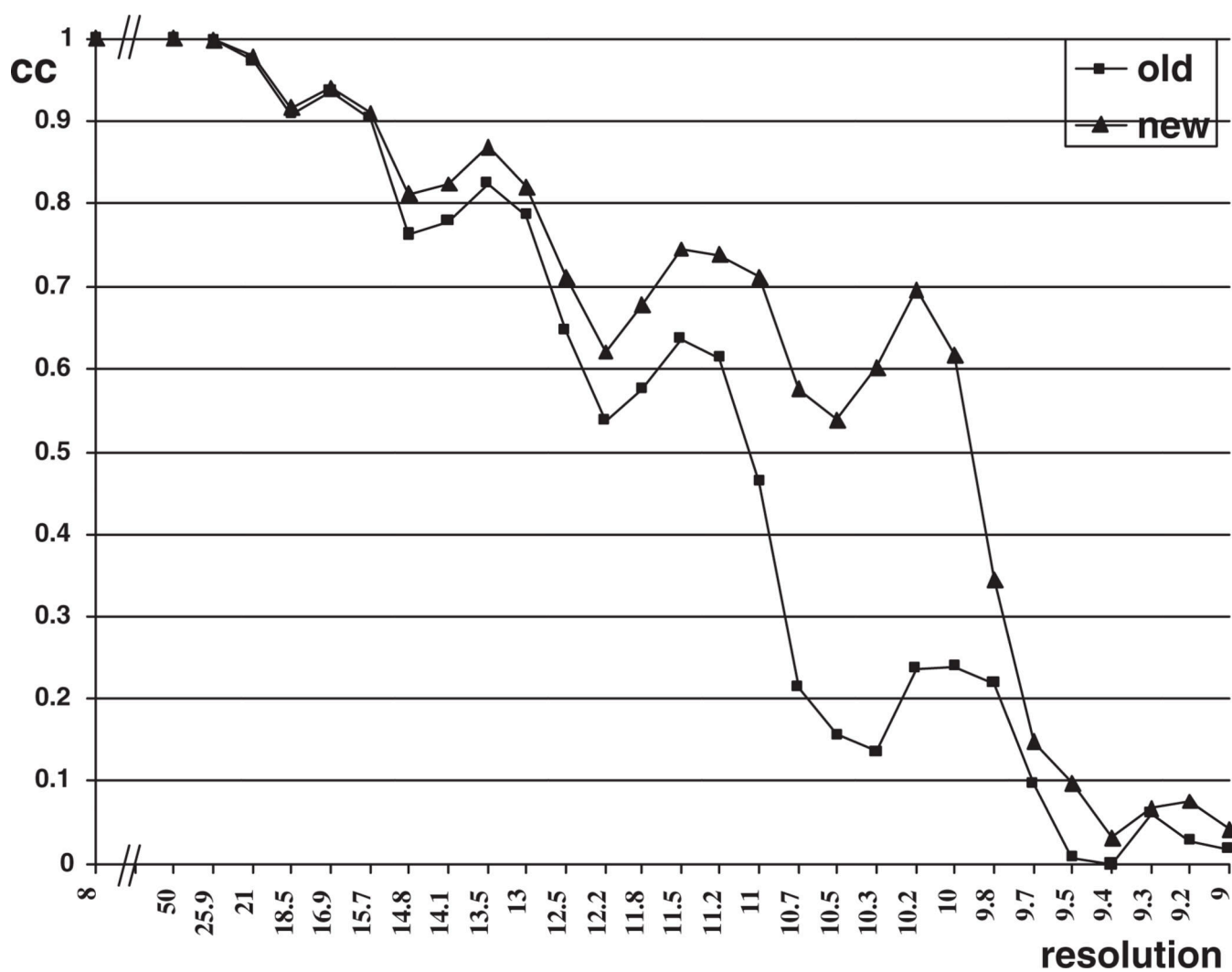
Figure 3.
3D electron density map of Sindbis virus obtained with its old orientation (left) and our new refined orientation(right).
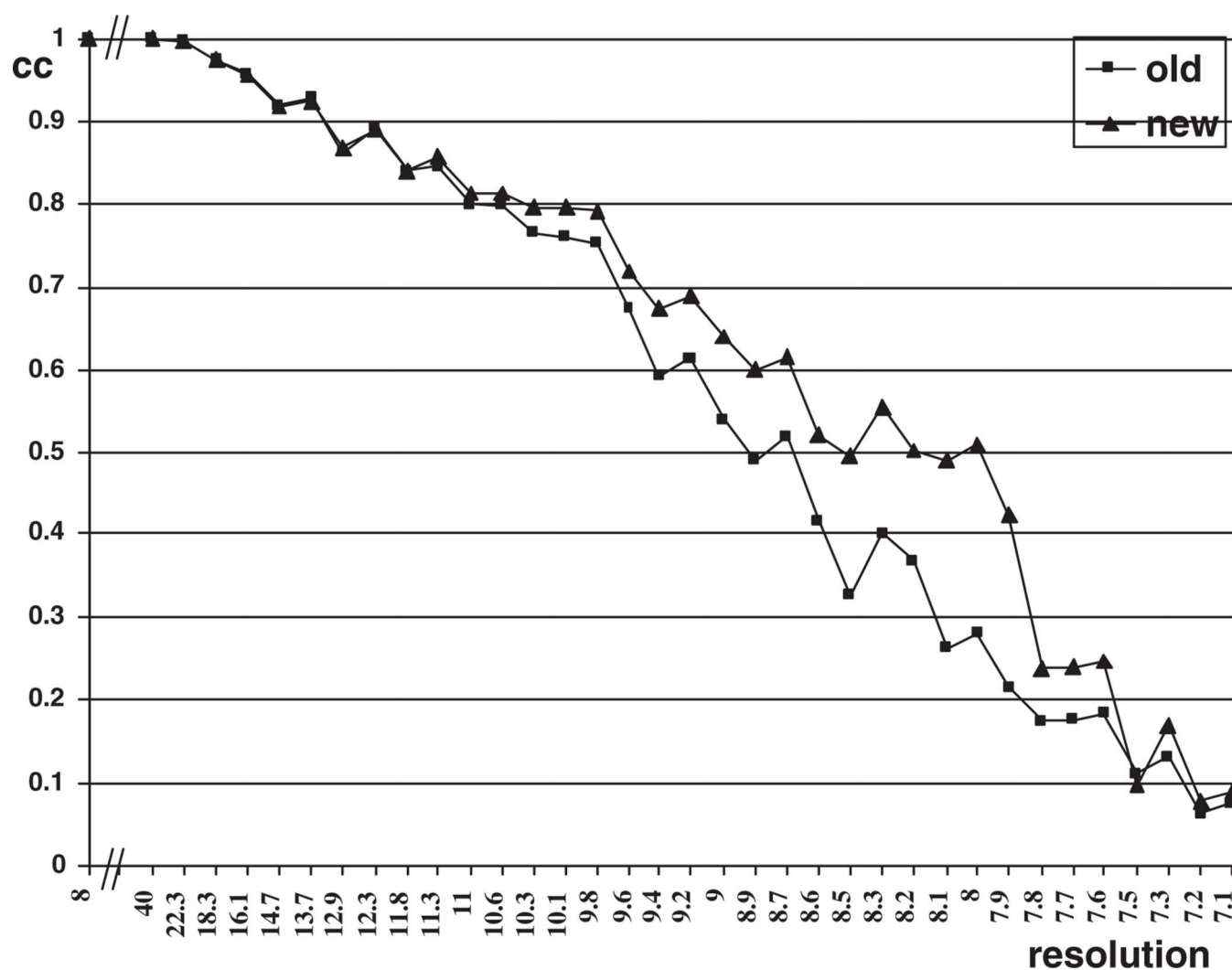
Figure 4.
The procedure used to determine the resolution of the electron density map.

Figure 5.
Correlation-coefficient plot for Sindbis virus.

Figure 6.
The correlation-coefficient plot for reo virus.

Table 1

The time for different steps of the orientation refinement process for the structure determination of Sindbis virus.

Angular resolution (°)	1	0.1	0.01	0.002
Search range	9	9	15	9
3D DFT (*s*)	19	15	13	13
Read image (*s*)	246	152	158	155
FFT analysis (*s*)	46	46	46	46
Orientation refinement (*s*)	14,053	14,109	71,065	26,901
Total time (*s*)	14,364	14,308	71,282	27,116

Table 2

The time for different steps of the orientation refinement process for the structure determination of reo virus.

Angular resolution (°)	1	0.1	0.01	0.002
Search range	9	9	11	9
3D DFT (*s*)	175	206	178	155
Read image (*s*)	550	533	573	529
FFT analysis (*s*)	138	142	137	138
Orientation refinement (*s*)	19,942	21,957	69,672	43,786
Total time (*s*)	20,805	22,839	70,561	44,608