Detecting Local Symmetry Axis in 3-dimensional Virus Structures

Jing He[1], Desh Ranjan[1], Wen Jiang[2,3], Wah Chiu[2,3], Michael F. Schmid[2,3]

[1]Department of Computer Science, New Mexico State University, Las Cruces, NM 88003, USA
[2] National Center for Macromolecular Imaging, Baylor College of Medicine, Houston, TX 77030, USA
[3] Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA

jinghe@cs.nmsu.edu

Abstract

This paper presents an efficient computational method to identify a local symmetry axis in 3-dimensional viral structures obtained using electron cryomicroscopy. Local symmetry is frequently observed in viral structures. Many virus structures have various types of local symmetry such as 2-fold, 3-fold and 6-fold that exist in certain limited regions on the 3-dimensional structure. Locations of local symmetry axes can be used in structure averaging as well as in detecting small structural variations among different copies of the same protein. We present a computational method that uses two-dimensional local correlation to identify the local symmetry axis in 3-dimensional viral structures. Instead of enumerating all the possible orientations of the symmetry axis, this method starts with a visually identified orientation and detects the trace of the axis, from which the exact orientation of the axis can be calculated. The complexity of this algorithm is analysed, and a comparison with a naïve method is provided. This method is able to detect the symmetry axis fairly accurately if the initial orientation of the axis is within 20° from the z-–axis, the viewing axis. The application of this method to herpes simplex virus B capsid structure obtained using electron cryomicroscopy technique is also presented.

Keywords: local symmetry, algorithm, structure, electron cryomicroscopy

1 Introduction

The genetic material of viruses, such as DNA or RNA, is usually encapsulated by a layer of protein shell. Many such protein shells appear to be spherical and have icosahedral symmetry. A virus with icosahedral symmetry has twelve 5-fold, twenty 3-fold, and thirty 2-fold symmetry axes (Figure 1A). The icosahedral symmetry axes are global because the symmetry operations associated with the axes apply to all regions on the virus structure. Because of the icosahedral symmetry, the viruses have a lot of repetitive information that can be used in averaging to enhance signal to noise ratio during reconstruction of the viral structures. The icosahedral symmetry has been widely applied during the reconstruction and therefore their locations are known on viral structures. Besides icosahedral symmetry, many viruses also have local symmetry which exists within certain limited regions of the structure. In general, local symmetry is good within certain radius from the symmetry axis and gradually degrades as the distance from the axis increases. Biologists have to decide the radius for calculating symmetry based on their biological interest such as which part of the molecule they want to evaluate. In order to use local symmetry in structure averaging or structure comparison, the exact location of the axis needs to be identified.

Although the exact locations of local symmetry axes are not directly available from the structure, their approximate locations are usually evident. Figure 1 shows the approximate locations of the local symmetry axes on herpes simplex virus B capsid structure. The herpes simplex virus B capsid has been determined at 8.5Å resolution using electron cryomicroscopy and image processing (Zhou, Dougherty, Jakana, He, Rixon and Chiu 2000). The reconstructed 3-dimensional structure of herpes simplex virus B capsid is a protein density map in which a density value is associated with each pixel in the 3-dimensional space of the map. Such a structural map is displayed by Iris Explorer (NAG, Downers Grove, Illinois) after applying a threshold value (Figure 1C). The hexamerous nature of the P, C and E hexon clearly indicates the approximate locations of local 6-fold axes. In this paper, we present an efficient way of detecting the exact location of a symmetry axis using an approximate orientation to start with.

A local symmetry axis can be described by five parameters $(Cx, Cy, Cz_0, \theta, \phi)$ where (Cx, Cy, Cz_0) is the location of a point on the axis (Figure 2A) and (θ, ϕ) is the orientation of the axis (Figure 2B). If the point is chosen to be the intersection between the axis and the middle z slice of the structure map, Cz_0 is known. θ is the rotation angle around the z-axis, and ϕ is the tilt angle away from the z-axis. In such a definition, identifying the local symmetry axis means identifying four parameters (Cx, Cy, θ, ϕ) of the axis around which local structure is symmetrically located. For example, a 120° rotation around a local 3-fold symmetry axis does not change the structure within certain distance from the axis. In this paper, we consider a cylindrical region around the axis.

We present a computational method that uses two-dimensional local correlation to identify the local symmetry axis in 3-dimensional viral structures. The complexity of this algorithm is analysed, and a comparison with a naïve method is provided. Although the method was

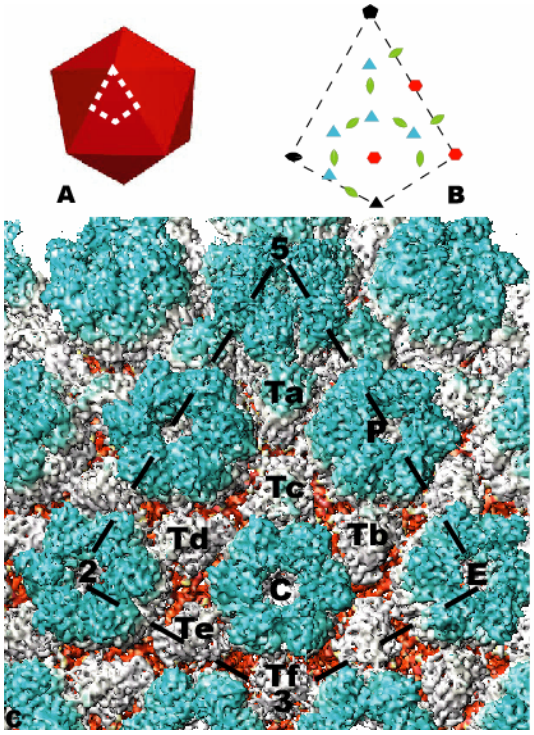

Figure 1: Local symmetry and their approximate locations on herpes simplex virus B capsid structure. (A) An icosahedron model of an icosahedral virus. An asymmetric unit, the structural unique region, on the icosahedron, and on the virus structure is indicated by dashed lines in (A), (B) and (C). (B) A diagram of the approximate locations of the unique local symmetry axes and the icosahedral symmetry axes on herpes simplex virus B capsid structure. Local 6-fold, 3-fold and 2-fold axes are indicated by red, blue and green symbols while the icosahedral 5-fold, 3-fold and 2-fold axes are indicated by black symbols. Note that the icosahedral 2-fold axis where E hexon is located is also a local 6-fold axis. (C) A portion of herpes simplex virus B capsid structure. The 8.5Å resolution structure of herpes simplex virus B capsid is obtained using electron cryomicroscopy and image processing (Zhou, Dougherty, Jakana, He, Rixon and Chiu 2000). The structure is radial colored as blue, white and red from the outer surface to the inner surface of the capsid shell. Each of the three types of hexons (P, C, and E) is composed of 6 copies of protein VP5, and each of the six types of triplexes (Ta, Tb, Tc, Td, Te, and Tf) is composed of two copies of protein VP23 and one copy of VP19C

developed primarily for structures obtained by electron cryomicroscopy, it can potentially be applied to other structural density maps. The accuracy of the method and its application on herpes simplex virus capsid structure obtained using electron cryomicroscopy technique is also presented.

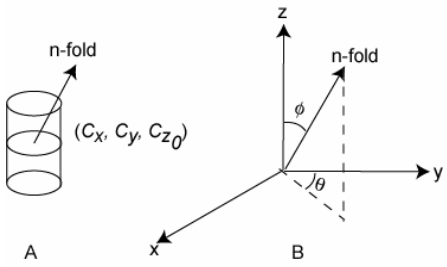

Figure 2: Definition of the symmetry axis.

2 Previous Results

Although there has not been a method developed for locating local symmetry on virus structures obtained by electron cryomicroscopy, there are methods to find a non-crystallographic symmetry axis (NCS) in the crystallography community (Kleywegt and Read 1997, Vonrhein and Schulz 1999). In X-ray crystallography applications, two proteins can be related by both the rotation around the local symmetry axis and the translation along the axis. Such an axis is called an NCS. On spherical virus structures, proteins surrounding a local symmetry axis are usually related by rotation only. Therefore, the local symmetry axis discussed in this paper can be thought of as a special kind of NCS.

NCS averaging has been a powerful way to enhance the signal during the refinement process of X-ray crystallography (Kleywegt and Read 1997). Previous methods in identifying NCS consist of two major categories. One is based on the X-ray diffraction data of the protein to identify the orientation and location of the axis. The orientation of the axis can be found using a self-rotation function (Rossmann 1972). The translation parameters can be searched or occasionally figured out from the information on heavy atom binding sites and Patterson vector (Schirmer, Keller, Wang and Rosenbusch 1995, Stubbs, Nar, Löwe, Huber, Ladenstein, Sprangfort and Svensson 1996). The other is a direct method without using the knowledge of the diffraction data and heavy atoms in the structure. This method works on the protein structure directly. It uses an estimated shape mask of the protein and evaluates a large number of possible orientations of the axis (Vonrhein and Schulz 1999). 3-dimensional correlation is used to measure the similarity between two possible objects related by a possible NCS. In this paper, we present a method that does not directly evaluate any of the possible orientations of the axis. It uses 2-dimensional local correlation to calculate the orientation of the local symmetry axis if the initial structure map satisfies certain condition.

3 Method

3.1 A Naïve Method

Identifying local symmetry axis means identifying the four parameters (Cx, Cy, θ, ϕ) of the axis around which rotation with a symmetrically related angle does not change the structure map locally. A direct approach is to test all the possible orientations and locations by performing an exhaustive search. For each possible orientation and location of the axis, the local region of the structure around the axis is rotated and 3-dimensional correlation will be

used to evaluate the similarity between the rotated and un-rotated local structure. The general form of normalized correlation coefficient is defined in (1.1). For 3-dimensional correlation, A is the cylinder at the axis and B is the symmetrical rotation of A around the axis. A_i and B_i are the i^{th} pixel of A, and B respectively. A_{avg} and B_{avg} are the averaged density of the cylinder A and B respectively. S is the volume of the cylinder. The higher the correlation score (correlation coefficient), the better the similarity. Obviously, this method will have to evaluation a large number of possible orientations and locations. The number of possible evaluations also depends on how fine the search interval it uses.

$$\text{Coefficient} = \frac{\sum_{S} (A_i - A_{avg})(B_i - B_{avg})}{\sqrt{\sum_{S} (A_i - A_{avg})^2 \sum_{S} (B_i - B_{avg})^2}} \quad (1.1)$$

3.2 Our Method

Our method in this paper takes a different approach observing the fact that the rough location of the local symmetry axes can be visually identified on virus structures (Figure 1). This method requires the user to work on an initial structure where the orientation of the local symmetry axis is not too far way from the z-axis (usually within 20° from the z-axis). In this situation, the computationally intensive search step for all the possible orientations can be totally eliminated by using 2-dimensional correlation.

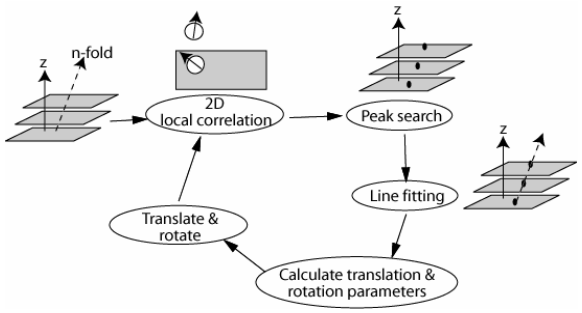

Figure 3: Local symmetry axis search process.

Instead of using 3-dimensional correlation, this method uses real space 2-dimensional correlation to identify the trace of the symmetry axis from which the orientation can be calculated directly. Four steps are involved in deriving the orientation and the location of the symmetry axis: 2-dimensional (2D) local correlation, peak search, line fitting and calculation of the translation and rotation parameters (Figure 3). Once the parameters are calculated, the structural map is translated and rotated so that the symmetry axis coincides with the z-axis. If the symmetry axis is accurately found and translated and rotated to the z-axis, the correlation peaks on all the slices should have identical x and y locations. Although one cycle is usually enough, it is possible to have multiple cycles to find the symmetry axis.

The first step in the symmetry axis search is the two-dimensional local correlation. Initially, a cubic area of interest (for example the C hexon region of herpes simplex virus B capsid) is isolated from the whole virus structure by the user. Any three-dimensional structure can be thought of as a stack of two-dimensional slices. The purpose of the two-dimensional local correlation is to find the most symmetrical point that is reflected by the maximum correlation coefficient value on each slice. The 2-dimensional correlation coefficient is defined in (1.1). For each pixel in the slice, a circular disc around it (A in 1.1) is used to calculate the correlation coefficient associated with this pixel. The radius of the disc is chosen by the user depending on the interest of biological questions. Because of the characteristic of local symmetry, symmetry is good only within a certain distance from the symmetry axis. B is the symmetrical rotation of A around the pixel, and S is the area of the circular disc. A_i and B_i are the i^{th} pixel of A and B respectively. The local correlation in this step uses a circular mask whose radius is decided by the user. In order for the correlation coefficient not to be biased by the overall brightness on different slices, the average density value of each slice (A_{avg} and B_{avg}) is used to normalize the coefficient. A general correlation coefficient formula is given in (1.1). In our application, A_{avg} and B_{avg} are the same. The result of 2-dimensional correlation of a structural map is a correlation map where each pixel is associated with a correlation coefficient calculated from a circular disc centered at that pixel. The next step, peak search, is to find the pixel with the highest correlation coefficient on each slice of the correlation map. The location of the correlation peak on each slice represents the most symmetrical point on the slice. The trace of such points from all the slices is used to do least-square fitting to find the line of the symmetry axis.

3.3 Software Implementation

We have implemented the above method in the single particle image processing package EMAN in C++ (Ludtke, Baldwin and Chiu 1999). The program *symAxisSearch* integrates all the above four steps and automatically iterates until the local symmetry axis is converged within a user-defined error threshold. The most computational intensive step in this method is the local 2D correlation step. Since the computation for different slices is independent, this step is amenable to trivial parallelization. To speed up the overall procedure, we have thus parallelized this step in the program *symAxisSearch* using OpenMP method in which the user defines the number of CPUs in the command line. The program *symAxisSearch* is freely distributed as part of EMAN package and is available at the National Center for Macromolecular Imaging, Baylor College of Medicine (http://ncmi.bcm.tmc.edu).

4 Results

4.1 Accuracy

The main contribution of this paper is to demonstrate that it is possible to eliminate the exhaustive testing step for all the possible orientations of the symmetry axis. In order to test the accuracy of this method, we used a 3-fold symmetry axis whose orientation and location is known on

a typical virus structure. A standard structure (81x81x53 pixel in size) is created where the 3-fold symmetry axis is the z-axis and passes the center of the structure. A set of

Test	*Cx*	*Cy*	θ	ϕ
1	1	1.2	30	5
2	-1.1	2.1	-20.4	8
3	1.3	-3.2	8.5	-10
4	-0.3	0.6	5.2	12.4
5	-2.2	-1.2	60.5	-15.3
6	1.4	-2	-9.2	15
7	2.5	-1.6	4.6	17
8	-0.6	0.8	2.2	-16.8
9	0	0	15.1	4.2
10	0	0	100	-3.4
11	2.5	3.5	50.6	2
12	-3.2	-3	1.2	-4
13	-1	4.2	8.6	-8
14	4.4	-5.2	-70.5	10.4

Table 1: Fourteen sets of the translation and rotation parameters in the accuracy test. The unit for *Cx* and *Cy* is pixel. The unit for θ and ϕ is °.

randomly generated orientation and translation parameters (ϕ, θ, *Cx, Cy)* is used to rotate and then translate the standard structure so that the 3-fold axis is away from the z-axis. Table 1 shows the fourteen sets of translation and rotation parameters (ϕ, θ, *Cx, Cy)* used in the accuracy test. Our method is used to identify the four parameters of the axis. The absolute difference (*ECx, ECy,* $E\theta$, $E\phi$) between the true and the identified value of (ϕ, θ, *Cx, Cy)* is calculated (Table 2). In both Table 1 and Table 2, the translation parameters are in the unit of pixel and the

Test	*ECx*	*ECy*	$E\theta$	$E\phi$
1	0.0041	0.0119	0.0219	0.0475
2	0.0083	0.0175	0.3235	0.0602
3	0.0006	0.0161	0.4874	0.1155
4	0.0056	0.0038	0.561	0.0326
5	0.0033	0.0051	0.884	0.1543
6	0.0111	0.0179	1.0926	0.2037
7	0.0091	0.0223	1.2252	0.2722
8	0.0193	0.0015	0.4551	0.2521
9	0.0077	0.0057	0.1721	0.0169
10	0.0031	0.0086	0.6279	0.0123
11	0.0165	0.0085	1.834	0.0337
12	0.0092	0.0078	0.7612	0.0269
13	0.0176	0.014	0.2128	0.0307
14	0.0027	0.0214	0.4912	0.0226
Avg	**0.0084**	**0.0116**	**0.6534**	**0.0915**

Table 2: Accuracy of the identified symmetry axis. *ECx, ECy,*$E\theta$ and $E\phi$ are the absolute difference between the true and the identified value for *Cx, Cy,* θ and ϕ respectively. The unit for *ECx* and *ECy* is pixel and the unit for $E\theta$ and $E\phi$ is °.

rotation parameters are in the unit of degree. The averaged error is less than 0.02 pixel for *Cx* and *Cy,* less than 0.1° forϕ, and less than 0.7° for θ (Table 2). We found that when the initial orientation of the symmetry axis is less than 20°, our method can identify the axis fairly accurately. Note that the accuracy of the identified symmetry axis is determined by the quality of the symmetrical point on each slice of the structure. When the initial orientation of the axis is far from the z-axis, the symmetrical point on each slice is less accurate because the symmetrical information on each slice is weak. However, the initial location (*Cx, Cy*) of the symmetry axis does not affect the accuracy of the method, as long as the initial structure includes the symmetrical region of interest. This is because that the quality of symmetrical information on each slice does not depend on where the symmetrical information is located on the structure map. Therefore, only small translations were randomly generated in the tests.

We found that the accuracy is closely related to the quality of the symmetrical point on each slice. Even though two such points are needed theoretically to determine a line, more are usually needed in practice. The least-square line fitting step is used for this purpose to achieve good accuracy. In order to produce high accuracy, we also used a non-integer representation for the location of the symmetrical points. The location of each peak (symmetrical point) is represented by its center of gravity in the neighborhood. The above two strategies are important for our method to achieve good accuracy.

4.2 An Analysis of the Running Time

Algorithm NAÏVE (*Structure*, *R*, *X*, *Z*, $\Delta\theta$, $\Delta\phi$, *n)*

//This algorithm locates a local *n*-fold symmetry axis for the input structure *Structure*.
//The input structure is given as density map on the cuboid [0…*X*,0…*X*,0…*Z*]
//The algorithm uses the parameters $\Delta\phi$ and $\Delta\theta$ as incremental steps in conducting a naïve search for the *n*-fold axis
//*R* is the radius of locally symmetric cylindrical region.

max ← 0;
ϕ ← 0;
θ ← 0;
for *i* ←1 to 180/$\Delta\phi$
  ϕ ← ϕ +$\Delta\phi$;
  for *j* ←1 to 360/$\Delta\theta$
    θ ← θ + $\Delta\theta$;
    for *Cx* ← *R* to *X*-*R*
      for *Cy* ← *R* to *X*-*R*
        Let S_1 be the cylinder with axis (*Cx, Cy,*θ,ϕ), radius *R* and height *Z*;
        Let S_2 be the cylinder obtained by rotating S_1 around the axis by 360/*n* degrees;
        cc ← correlation coefficient of S_1 and S_2;
        if (*cc* > *max*)
          max ← *cc*;
          opt ← (*Cx, Cy,*θ,ϕ);
return *opt*;

In order to analyse the running time, the pseudo code of both the naïve method and our method (Algorithm TRACE) are shown. In the pseudo code, *Structure* is the 3-dimensional structure with length and width *X* and height *Z*. *R* is the radius of local symmetry selected by the

user, $\Delta\theta$ and $\Delta\phi$ are the step size for θ and ϕ respectively. n is the type of symmetry. For example, when n=2, the methods will look for the 2-fold symmetry axis.

In the naïve method, the correlation involves two 3-dimensional cylinders of radius R and height Z. One is the cylinder at the potential axis, and the other is the same cylinder rotated around the potential axis by 360/n degrees. The time needed for each correlation is proportional to the volume of the cylinder that is $2\pi R^2 Z$. Since the number of correlations performed is $(360/\Delta\theta)(180/\Delta\phi)(X\text{-}2R)^2$, the total time is proportional to $(360/\Delta\theta)(180/\Delta\phi)(X\text{-}2R)^2 2\pi R^2 Z$.

In the TRACE method, the time for calculating the correlation coefficient of two circular discs of radius R is proportional to the size of the disc $2\pi R^2$. The number of correlation performed is $(X\text{-}2R)^2 Z$. The time for least-square line-fitting is proportion to Z, and calculating (Cx, Cy, θ, ϕ) takes constant time. The last two are negligible compared to the time required for correlation computation. So the total running time for TRACE is proportional to $(X\text{-}2R)^2 2\pi R^2 Z$.

The running time of the naïve method is inversely proportional to $\Delta\theta$ and $\Delta\phi$, where as the running time of algorithm TRACE is independent of these parameters. This makes a vast difference in the running time of the two methods. As an example, if $\Delta\theta$ and $\Delta\phi$ are both chosen to be 1°, the naïve method will be roughly 360*180=64800 times slower than the TRACE method.

Algorithm TRACE (*Structure, R, X, Z, n)*

//This algorithm locates a local *n*-fold symmetry axis for the input structure *Structure*.
//The input structure is given as density map on the cuboid [0…*X*,0…*X*,0…*Z*]
//*R* is the radius of locally symmetrical cylindrical region.

```
for z ← 0 to Z
    max ← 0;
    for Cx ← R to X-R
        for Cy ← R to X-R
            Let S1 be the disc of radius R, centered at
            (Cx, Cy);
        Let S2 be the disc obtained by rotating S1
            around (Cx, Cy) on the same z slice by
            360/n degrees;
        cc ← correlation coefficient of S1 and S2;
        if (cc > max)
            max ← cc;
            peak[z] ← (Cx, Cy);
Fit a straight line through the points in peak using
least-square fitting;
Calculate (Cx, Cy, θ, φ) for this line;
opt ← (Cx, Cy, θ, φ);
return opt;
```

4.3 Local 6-fold Symmetry of Herpes Virus Capsid Structure

The method developed in this paper is used to evaluate the symmetry in the local 6-fold regions on herpes simplex virus B capsid structure (Zhou, Dougherty, Jakana, He, Rixon and Chiu 2000). The local 6-fold region is composed of a hexon which has 6 copies of the major protein VP5. Depending on the location in an asymmetric unit of the structure, there are three types of hexons: P, C, and E (Figure 1). Visual inspection shows that the three types of hexons are almost identical. We first located the local 6-fold symmetry axis and then calculated the 6-fold symmetry to see if the three types of hexons have similar quality of symmetry.

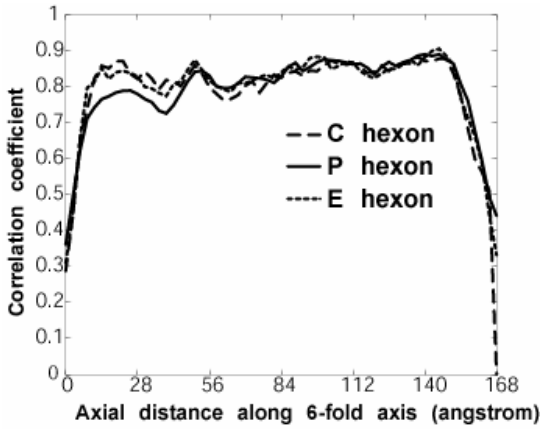

Figure 4: Local 6-fold symmetry of P, C and E hexons of herpes simplex virus B capsid. The local symmetry is calculated using a cylindrical mask of 64.4 Å in radius. The quality of the local 6-fold symmetry of the P, C, and E hexons is shown for all the slices perpendicular to the axis.

The 6-fold symmetry of the P, C and E hexons are shown in figure 4. The quality of the hexon symmetry is reflected by the correlation coefficients between the original hexon and a 60°-rotated hexon. In order to know the symmetry change along the axis, the correlation coefficient was calculated for each slice perpendicular to the symmetry axis. The 6-fold symmetry of the hexons was calculated using a cylindrical mask of 64.4 Å in radius. Theoretically, the correlation coefficient of symmetry can range from –1 to 1 with 1 being the perfect symmetry. However, it is usually not possible to have a perfect local symmetry in a structure from real data because of the breakdown of the local symmetry or the presence of noise in the structure. The overall 6-fold symmetry of P, C and E hexons are good with the majority of the slices having correlation coefficient over 0.8. This result coincides well with the visual inspection of the structure. However, in the region of about 34 Å-thick at the left of the symmetry curve, the correlation coefficient of P hexon is about 0.05 to 0.1 lower than that of the C and E hexons. This 34 Å-thick region is located at the inner surface of the capsid shell. Since the lower symmetry value in the P hexon is exclusive to this region while the outer $^3/_4$ of the hexon has close matching symmetry value, the reduced symmetry is likely to suggest a genuine conformational difference in this portion of the P hexon. Further investigation of this

region shows that two of the 6 copies of the VP5 molecules moved slightly away from the 6-fold symmetry in this region of the P-hexon (He, Schmid, Zhou, Rixon and Chiu 2001).

5 Discussion

The naïve method does not have the requirement for the initial orientation of the axis as opposed to our method which requires the initial ϕ to be less than 20° to achieve good accuracy. There is no requirement for the initial value for *Cx, Cy* and θ. The naïve method is a general method. However, given the fact that rough locations of the local symmetry axes are often easy to be identified visually on virus structures (Figure 1), the requirement of 20° for ϕ is usually satisfied for virus structures. This paper provides an efficient algorithm to find the exact location of a local symmetry axis starting with a visually identified axis. It does not screen any of the possible orientations of the symmetry axis. Rather, it calculates the orientation using a trace of the symmetry axis from 2-dimensional local correlation. The most time consuming step of both methods is the calculation of the correlation coefficient which is proportional to the volume of the cylinder $2\pi R^2 Z$. Even when the naïve method is applied on the same initial structure (ϕ<20°), the number of correlations is 20*360=7200 times more than the 2-dimensional correlation method assuming using step size of 1° for accuracy.

When the initial orientation (ϕ) is less than 20° from the z-axis, the TRACE method is able to find the symmetry axis fairly accurately and efficiently because all of the fourteen tests finishes in one or two iterations. When the initial ϕ is more than 20° from the z-axis, the symmetrical information on each slice is weak. We did not perform a rigorous test in this situation. However, our experience suggests that for initial ϕ ranging from 20° to 45°, the TRACE method is usually able to find the axis through more iteration. For initial ϕ less than 1°, our method may have problem finding a more accurate axis depending on the number of z-slices on the structure. In general, the more number of z-slices exists on the structure, the more accurate the identified symmetry axis is. This is because that there are more symmetrical points for line fitting. The best way to use the TRACE method is to rotate the structure map by visual inspection so that the initial ϕ is within 20° from the z-axis.

Both of the methods discussed in this paper calculate correlation in real space, the structure space. In general, a real space correlation can be computed more efficiently in Fourier space, the frequency space, by using Fast Fourier Transform (FFT) (Gonzalez and Woods 1992). For example, if the problem is to locate a template in an image, the typical way of using correlation is as the following. Apply FFT on both the image and the template, and then multiply the Fourier transform of the image and the complex conjugate of the Fourier transform of the template. The correlation coefficients are obtained by applying inverse FFT to the resulting Fourier transform. However, this method does not directly apply to the problem of locating local symmetry axis. In order to find the local similarity, the template in the correlation should reflect the local structural information. At different locations of the structure, the correlation should be done with different templates that are the local structures at those locations.

6 Acknowledgements

We would like to thank Hong Z. Zhou for his support and providing the herpes structural map. We would also like to thank Steve J. Ludtke for his generous help during program development. The work is supported by grants from NIH (P41RR02250 and RO1AI38469).

7 References

Gonzalez, R.C. and Woods, R.C. (1992): *Digital Image Processing*. Adison-Wesley.

He, J., Schmid, M.F., Zhou, Z.H., Rixon, F. and Chiu, W. (2001): Finding and using local symmetry in identifying lower domain movements in hexon subunits of the herpes simplex virus type 1 b capsid. *J Mol Biol* **309**(4)**:**903-14.

Kleywegt, G.J. and Read, R.J. (1997): Not your average density. *Structure* **5**(12)**:**1557-69.

Ludtke, S.J., Baldwin, P.R. and Chiu, W. (1999): EMAN: Semi-automated software for high resolution single particle reconstructions. *J Struct Biol* **128**(1)**:**82-97.

Rossmann, M.G. (1972): The locked rotation function. *J Mol Biol* **64**(1)**:**246-9.

Schirmer, T., Keller, T.A., Wang, Y.F. and Rosenbusch, J.P. (1995): Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science* **267**(5197)**:**512-4.

Stubbs, M.T., Nar, H., Löwe, J., Huber, R., Ladenstein, R., Sprangfort, M.D. and Svensson, L.A. (1996): Locating a local symmetry axis from Patterson map cross vectors: application to crystal data from GroEl, GTP Cyclohydrolase I and the Proteosome. *Acta Crystallogr D Biol Crystallogr* **52**(Pt 3)**:**447-52.

Vonrhein, C. and Schulz, G.E. (1999): Locating proper non-crystallographic symmetry in low-resolution electron-density maps with the program GETAX. *Acta Crystallogr D Biol Crystallogr* **55**(Pt 1)**:**225-9.

Zhou, Z.H., Dougherty, M., Jakana, J., He, J., Rixon, F.J. and Chiu, W. (2000): Seeing the herpesvirus capsid at 8.5 Å. *Science* **288**(5467)**:**877-880.