A robust algorithm for the reconstruction of helical filaments using single-particle methods

Edward H. Egelman*

Department of Biochemistry and Molecular Genetics, University of Virginia, Health Sciences Center, Box 800733, Charlottesville, VA 22908-0733, USA

Received 18 May 2000; received in revised form 24 July 2000

Abstract

Some of the earliest methods for three-dimensional reconstruction from electron microscopic images were developed for helical objects. Single-particle methods have been used with great success for the three-dimensional reconstruction of macromolecular assemblies that have no internal symmetry or closed point group symmetries. An approach is presented for the application of single-particle methods to helical filaments that surmounts many of the difficulties of helical image analysis, including indexing, unbending and the need to find long helically symmetric filament segments. It is shown using both human Rad51 and E. coli RecA nucleoprotein filaments that this approach converges without user intervention to a stable solution, and that it has the potential to overcome many of the problems associated with image analysis of disordered helical polymers. The method can be applied transparently to structures where Bessel overlap would greatly complicate helical analysis. In addition, the procedure allows for the ab initio determination of helical symmetry, when no prior knowledge exists. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

Many biological macromolecular assemblies exist as helical polymers. These include, among many other complexes: F-actin, microtubules, myosin thick filaments, phage tails, and bacterial pili and flagella. The bacterial RecA protein, which can be the most abundant protein in the cell after DNA damage, forms a helical nucleoprotein filament with DNA [1]. The eukaryotic Rad51 protein forms a very similar helical filament [2]. The centrality of helical polymers to biology makes methods for structural studies of such polymers important. It is thus not coincidental that the first application of electron microscopic three-dimensional reconstruction techniques was to a helical polymer [3]. In the years since, there have been many refinements and advances, and it has been possible to use such methods to image at high-resolution (better than 10 Å) specimens such as tobacco mosaic virus [4], the acetylcholine receptor [5], the sarcoplasmic reticulum calcium pump [6] and bacterial flagella [7].

However, the application of helical methods is not simple, particularly when the specimens are flexible and disordered. One initial approach to dealing with flexibility is to computationally straighten filaments [8], but this has the potential for introducing artifacts due to the assumptions...
that need to be made (e.g., that there is no coupling between bending and twisting, that the filament undergoes a purely elastic normal mode of bending, etc.). Another difficulty with helical analysis is that the indexing of a pattern can be ambiguous, and the wrong symmetry can be chosen [9]. Further complications exist when the filament does not have a precisely defined helical symmetry, such as F-actin that has a random, variable twist [10]. A real-space approach to dealing with the variable pitch of sickle-cell hemoglobin helical fibers has already been used [11], and other groups have also employed single-particle methods to reconstruct helical filaments [12,13]. Real-space methods allow many advantages over helical approaches, particularly in surmounting the problem of disorder and flexibility. For highly ordered specimens, such as tobacco mosaic virus [4] or helical tubes of membrane proteins [5,6], the advantages may be less obvious.

A method is presented here for iterative real-space reconstruction of helical filaments. The method has been successfully applied to a number of different specimens, but results are only shown here for one state each of the human Rad51 protein and the *E. coli* RecA protein.

2. Results

Fig. 1 shows an electron micrograph of human Rad51 (hRad51) protein bound to single-stranded DNA in the presence of ADP and aluminum fluoride, which together serve as a stable, non-hydrolyzable analog for ATP. The striations arising from the ∼98 Å pitch helix can be seen. Attempts to use conventional helical methods were limited by the paucity of long, suitably straight segments. The indexing of those segments that were found was complicated and ambiguous, as very few filament sections yielded reasonable transforms. We therefore used a three-dimensional reconstruction of the homologous bacterial RecA protein [14] as a starting point for a real space procedure outlined in Fig. 2.

Images of segments of the hRad51 filaments were collected using an automated procedure based upon cross-correlation with the projection of the RecA reference. In this procedure, cross-correlation peaks were found between a projection of the RecA reference filament (∼400 Å long, or 100 pixels) and the electron micrograph. The reference projection was then rotated in the plane by 4°, and new cross-correlation peaks were found. This rotation was repeated 44 times, so that projections of the reference between 0 and 180° were used for the cross-correlations. This procedure ignores the polarity of the filaments, which would require references between 0 and 360°. However, polarity is best determined in subsequent stages, and ignoring it at this point expedites the selection process. Since the pitch of the hRad51 filaments was observed to be ∼98 Å, the RecA reference reconstruction was generated

![Fig. 1. An electron micrograph showing negatively stained hRad51 protein bound to single-stranded DNA in the presence of ADP and aluminum fluoride. Micrographs were recorded at a nominal magnification of 30,000 × at an accelerating voltage of 80 kV using a JEOL 1200 EXII microscope. The scale bar is 2000 Å.](image-url)
using Fourier–Bessel methods [3] for a helix with a symmetry of 37 subunits in six turns of a 98 Å pitch helix (as opposed to the average 37/6 symmetry of a 93 Å pitch helix for the native RecA). This change in structure only involved a change in the helical repeat, c, that is used during the helical synthesis.

The cross-correlation method yielded 7620 segments, of size 60 x 60 pixels (with a sampling of 4 Å/pixel), each containing slightly more than two turns of the Rad51 helix. Most of the segments did not overlap, but subsequent work with this procedure has suggested that it is most efficient if there is substantial overlap. This is because two overlapping segments, related by an axial displacement, can provide important information about the helical symmetry. The size of the boxes was chosen not to be too long so as to encounter problems with filament bending and internal disorder, but not be too short so that cross-correlations would be ambiguous. This was determined empirically, and might need to be optimized for different structures depending upon these different factors (such as the mass per unit length, the signal-to-noise ratio, the flexibility, etc.). The SPIDER software package [15] was used for most of the operations described. The initial part of the cycle was the same as has been applied to objects such as the ribosome [16]: a multi-reference alignment was used between 90
projections of the reference RecA filament (at 4°
angular increments about the filament axis) and
the raw images. This yielded five parameters for
each filament: a coefficient of correlation with the
most similar reference, an in-plane rotation, an x-
shift, a y-shift, and the azimuthal angular orienta-
tion of the segment (from the known azimuthal
orientation of the reference image that yielded the
highest cross-correlation against the raw image).
Images that had poor correlation coefficients, or
large shifts, were excluded. Back projection was
then used on the resulting aligned images, in what
is essentially a single-axis tilt series with uniform
sampling of all angles and no missing information.
This generated a 3D volume in which no assump-
tions have been made about internal symmetry.

It was apparent that the reconstructed object
was helical, but it was also clear that large edge
effects existed at the “top” and “bottom” of the
filament segment reconstructed. This was ex-
pected, as there was a direct relationship between
the axial length of the search used for the two-
dimensional projection alignments and the length
of the helically symmetric portion of the recon-
structed volume. For example, in Fig. 2 the multi-
reference alignment search is conducted over a
region containing ~12 subunits, and large edge
effects in the reconstruction only appear near the
ends of a filament segment which contains ~14
subunits. Thus, the reconstruction could be
treated as reasonable within a central spherical
shell. To be conservative, only the central 9
subunits were treated as “correct” within this
volume, and the subunits suffering from potential
dge effects were therefore ignored in the sub-
sequent step.

Helical symmetry was then imposed upon this
central volume (containing 9 subunits) by first
determining the two helical parameters (the
azimuthal rotation, \( \Delta \phi \), and the axial translation,
\( \Delta z \)) that related each subunit to its neighbors. A
starting guess for \( \Delta z \) was used in a search for \( \Delta \phi \),
calculating the mean-squared deviation \( \langle d^2 \rangle \) in
density between voxels of density and the density
at eight symmetry-related positions in the central
volume as \( \Delta \phi \) was varied. This was done by first
converting the density array \( \rho(x, y, z) \) to cylind-
rical coordinates, \( \sigma(r, \phi, z) \), where the helical axis
was placed along the line \( r = 0 \) in the new
coordinate system. For such cylindrical coordi-
nates, symmetry related positions are given by
\( \sigma(r, \phi + n\Delta\phi, z + n\Delta z) \) for \( n = \cdot -2, -1, 0, 1, 2, \ldots \). The mean squared deviation in density
as a function of the symmetry operators \( \Delta \phi \) and
\( \Delta z \) was calculated by examining a cylindrical slab,

![Fig. 3. A least-squares search of the volume generated by back projection is used to determine the two helical symmetry parameters: the axial rise (\( \Delta z \)) and the azimuthal rotation (\( \Delta \phi \)) between two adjacent subunits. Since the two parameters are coupled, an initial “guess” for \( \Delta z \) is used to determine \( \Delta \phi \), and this value of \( \Delta \phi \) is then used for a new determination of \( \Delta z \). This needs to be iterated only once, using the updated values of \( \Delta z \) to determine a new \( \Delta \phi \) and the new value of \( \Delta \phi \) to determine \( \Delta z \). The data shown are the mean-square deviations in density as a function of changes in \( \Delta z \) and \( \Delta \phi \) from a typical search.](image-url)
\( \sigma(0 \leq r \leq r_{\text{max}}, 0^\circ \leq \phi < 360^\circ, z_0 \leq z < z_0 + \Delta z) \), and determining the eight symmetry-related densities for every element (voxel). From these nine densities we have a mean density for this element, and a mean-squared deviation. By summing over every element in this volume, we obtain the mean-squared deviation for the entire volume as a function of the symmetry operator. Fig. 3 shows the results of such a search for these two parameters. For the angular component of the symmetry, \( \Delta \phi \), the resulting values were fit with a non-linear algorithm to a quadratic equation of the form

\[ \langle d^2 \rangle = a + b \Delta \phi + c (\Delta \phi)^2. \]

Minimization of this function occurs when

\[ \Delta \phi = -b/2c \]

and this value of \( \Delta \phi \) was then used to find \( \Delta z \). The same quadratic fit and minimization was applied to the search for \( \Delta z \). The procedure was then iterated, since the determination of the best value for \( \Delta \phi \) requires a value for \( \Delta z \), and vice versa. In practice, the change of one parameter was small with respect to a change of the other parameter, and both parameters remained close to the initial estimates. The test that these values were the best estimates came from the fact that there was no further change in the values of \( \Delta \phi \) and \( \Delta z \) with further iterations. The parameters \( \Delta \phi \), \( \Delta z \) were then imposed upon the three-dimensional volume, generating a helically symmetric reconstruction. This completes a full cycle, and the new helically symmetric reconstruction can be used as the reference for projections within a new cycle (Fig. 2).

Several concerns might exist about the method. Is the final reconstruction dependent upon the starting model? Will the true helical symmetry be found, or will the symmetry of the starting model be imposed upon the results? Fig. 4 presents some
statistics from 123 cycles of this procedure. The helical symmetry converges to ~6.4 subunits in a 99 Å pitch helix after ~80 cycles. This was after starting with a model having 6.17 subunits per turn of a 98 Å pitch helix. It can be seen that the goodness-of-fit between the projections of the reconstruction and the raw images reaches a nearly asymptotic value after about 60 cycles.

Fig. 5 shows surfaces of the helically symmetrized volumes as a function of the number of cycles. Fig. 5a is the initial reference model, while Figs. 5b–e are actual reconstructions from the images. Typically, ~4200 images (out of the starting 7620 images) were included in each reconstruction, with the remainder being excluded due to poor coefficients of correlation, large shifts, or large rotations from either 0 or 180°. The boxed segments were extracted after cross-correlation with a vertically oriented reference filament, so all of the segments should have in-plane rotations of ~0 or ~180°. It can be seen that there is virtually no change in structure between cycles 80 and 123, consistent with a lack of change in the helical parameters over these cycles. It can also be seen that the resulting structure deviates quite significantly from the starting model. Determination of resolution is complicated by the fact that helical averaging is imposed upon the three-dimensional reconstruction. However, for the asymmetric reconstruction, the resolution is judged to be 20 Å using a conservative Fourier shell correlation threshold of 0.5 between two different reconstructions generated from two halves of the data set.

Using the same starting model but images of an unusual state of the *E. coli* RecA protein on double-stranded DNA, it can be seen in Fig. 6 that the procedure now converges to a very different solution from that shown in Fig. 5. While normal RecA-DNA-ATP-γ-S filaments have a pitch of ~93 Å [9,14], we have used filaments with a pitch of 86 Å. These filaments would have been discarded as candidates for helical image analysis, as additional layer lines beyond the equator and
the 1/(86 Å) lines were not seen. The single-particle method, applied to 6964 segments, converges to a stable solution for both the symmetry (Fig. 6b and c) and the structure (Fig. 6a) after ~40 cycles. The symmetry that is revealed, 6.05 subunits per turn of an 86 Å pitch helix, explains the lack of other visible layer lines using helical diffraction. A computed transform (Fig. 6d) of the projection of the reconstruction can be seen to have a small amount of layer-line splitting, but to predominantly have only the equator and two orders of the 1/(86 Å) layer line. The projection was from a helically symmetrized density containing 125 subunits, and transforms of shorter segments would not display any splitting due to layer line broadening. In the presence of noise the amount of splitting shown would also be nearly impossible to detect. Thus, in practice one would rarely observe more than orders of the 1/(86 Å) layer line.
from such filaments, and reconstructing these filaments using Fourier–Bessel analysis would be complicated. While the separation of Bessel functions from such overlapping layer lines has been solved for other specimens [17], the method is much more complicated and requires more user-intervention than the fully automated procedure described here.

Since the method requires both an initial structure and an initial estimate of the helical symmetry, it is interesting to explore how far these can be from the actual structure for proper convergence to still occur. In Fig. 7 it can be seen that the radius of convergence is very large. In fact, the method allows for the ab initio determination of the helical symmetry of hRad51 filaments.

Fig. 7. A continuous 1-start helical density (a) was used as the initial reference for aligning 7620 segments of hRad51 filaments. A 2₁ screw symmetry was imposed upon the first reconstruction from these aligned segments, and the symmetrized volume is shown in (b). After only three more cycles of iterative alignment, reconstruction and symmetrization, a strong modulation by subunits can be seen (c). The projection down the filament axis for one turn of this structure is shown (e), and it can be seen that the asymmetric unit contains three subunits (labeled 1, 2 and 3). The 2₁ symmetry generates the subunits labeled 1*, 2* and 3*. By taking the structure in (c) after the fifth cycle and iterating only five more times using ~6 subunits per turn (rather than ~2 subunits per turn), a filament is reconstructed (d) that is very similar to what was achieved when one started with a known approximate symmetry (Fig. 5). The change in the symmetry operator $\Delta \phi$ as a function of the number of iterations is shown (f), and it can be seen that it relaxes from the initial value of 180° to a value of $\sim 173°$. 
Initially, the continuous density along RecA’s 1-start helix (Fig. 7a) was used as a reference for the alignment of the 7620 hRad51 segments. This continuous density distribution has polarity, but no modulation due to subunits. It was generated by using a helical reconstruction with only the equator and the strong layer line corresponding to the 1-start helix. After alignment, a reconstruction was generated by back projection. For an infinite number of images, all modulation along the helix should cancel since the reference was featureless. However, for a finite number of images the reconstruction will contain noise. To aid in convergence, the simplest helical symmetry was assumed, which is a $2_1$ screw. This symmetry has $\Delta \phi = 180^\circ$ and $\Delta z = 49 \text{ Å}$, and provides a means for internal averaging of the reconstructed density.

Fig. 7b shows the result of imposing this $2_1$ symmetry on the first reconstruction. Very weak modulations of the density can now be seen along the helix. This is used as a new reference volume, a search is made (about $8^\circ$) of the density by subunits can now be seen (Fig. 7c). It is quite clear after these five cycles that there are approximately 6 subunits per turn of the helix. The symmetry operators that are found for this fifth cycle are $\Delta \phi = 179.8^\circ$ and $\Delta z = 49.1 \text{ Å}$. Since the rotation operator continues to fall from the initial $180^\circ$ (Fig. 7f), it suggests that there are slightly more than 6 subunits per turn (if it was greater than $180^\circ$, it would suggest slightly less than 6 subunits per turn). A projection down the filament axis of one turn of the helix in Fig. 7c is shown in Fig. 7e, where the subunits related by the $2_1$ screw symmetry are labeled. One could now take this density and begin a new set of cycles, starting with a symmetry of $\Delta \phi = 59.9^\circ$ ($179.9^\circ/3$) and $\Delta z = 16.4 \text{ Å}$ ($49.1 \text{ Å}/3$). After just five more cycles, the structure (Fig. 7d) is now nearly the same as was achieved using 80 cycles with an initial reference of 6.17 RecA subunits per turn (Fig. 5d), even though we began with no known symmetry. Continuing for more cycles leads to the identical structure (data not shown). If one continues to cycle using ~2 subunits per turn, the rotation operator falls to $\Delta \phi = 173^\circ$ (Fig. 7f), although it was started at $180^\circ$. This suggests an asymmetric unit containing three subunits, with a rotation of $57.67^\circ$ ($173^\circ/3$) between each. This is different from the value found in Fig. 4, ~56.3° per subunit, and reflects the limited utility of the noisy reference that results from enforcing the $\sim 2_1$ symmetry. Nevertheless, the procedure converges after an initial imposition of $2_1$ symmetry, even though there are ~6.4 subunits per turn of the helix, and not 6.0.

3. Conclusions

The iterative method of real-space reconstruction of helical filaments has been shown to converge to a stable solution that is different in both helical symmetry and structure from the starting model. Indeed, the method can be used to determine the helical symmetry of an unknown structure by imposing a limited averaging, assuming two subunits per turn. For the example shown, hRad51 filaments, the method actually converges to a final structure more rapidly when no assumptions are made about symmetry than when the structure of a homologous protein, RecA, is used as a starting model. The results presented suggest that this approach to determining symmetry may work on many structures, even if the helical symmetry is not an even, integral number of subunits per turn.

A dramatic difference between the approach presented here and traditional helical methods is that the entire procedure is largely independent of human intervention. In contrast, the straightening and selection of filaments for helical processing is laborious, requires an experienced individual, and subjective judgments are frequently required. The single-particle method is reasonably efficient, and takes ~60 min/iteration using 7620 images on an SGI Octane. Because the most time-consuming portion of the procedure, the multi-reference alignment, lends itself to coarse-grained parallelization, we have been able to reduce this time considerably using clusters of inexpensive computers. For example, using a cluster of only five 950 MHz PCs (each costing ~$1500) it has been
possible to reduce the time per cycle to 35 min. Due to the coarse-grained nature of the parallelization, this time scales nearly linearly with the number of nodes. While this may still be one or two days of computer time for a stable solution, there is no user intervention needed. Attempting to reconstruct such filaments using helical methods might take the full time attention of a trained investigator for weeks, or might be nearly impossible, as was the case for the examples shown.

It is shown that the procedure can be applied transparently to structures having a small integral or nearly integral number of subunits per helical turn, where Bessel overlap would complicate helical analysis. The method also lends itself to more detailed analysis of heterogeneity and subpopulations. By using several different reference models, it is possible to sort the images and reconstruct distinct subgroups. This has been possible for both F-actin (Galkin, Lukoyanova, Orlova and Egelman, in preparation) and RecA (Yu, Jacobs and Egelman, in preparation). Finally, the method can be easily extended to cryo-EM of filaments in ice, where significant out-of-plane tilt will be encountered, due to the fact that the set of reference projections can be expanded to allow for such tilt.

The programs to implement the procedures described in this paper will either be available within SPIDER [15], or may be obtained from the author.

Acknowledgements

I would like to thank Xiong Yu, Albina Orlova, Natalya Lukoyanova, Vitold Galkin, and Steve Jacobs, who have helped in the application of this method to Rad51, RecA and actin. I would like to thank Stephen West for generously supplying human Rad51 protein, Xiong Yu for specimen preparation and electron microscopy, and Pawel Penczek for helpful discussions. This work was supported by NIH GM35269 and AR42023.

References