doi:10.1006/jmbi.2001.5438 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 317, 375–384

JMB

Multi-resolution Contour-based Fitting of Macromolecular Structures

Pablo Chacón* and Willy Wriggers

Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla CA 92037, USA

**Corresponding author*

A novel contour-based matching criterion is presented for the quantitative docking of high-resolution structures of components into low-resolution maps of macromolecular complexes. The proposed Laplacian filter is combined with a six-dimensional search using fast Fourier transforms to rapidly scan the rigid-body degrees of freedom of a probe molecule relative to a fixed target density map. A comparison of the docking performance with the standard cross-correlation criterion demonstrates that contour matching with the Laplacian filter significantly extends the viable resolution range of correlation-based fitting to resolutions as low as 30 Å. The gain in docking precision at medium to low resolution (15-30 Å) is critical for image reconstructions from electron microscopy (EM). The new algorithm enables for the first time the reliable docking of smaller molecular components into EM densities of large biomolecular assemblies at such low resolutions. As an example of the practical effectiveness of contour-based fitting, a new pseudo-atomic model of a microtubule was constructed from a 20 Å resolution EM map and from atomic structures of α and β tubulin subunits.

Keywords: low resolution; docking; macromolecular models; electron microscopy; structure modeling

Introduction

Many biophysical techniques can usefully be applied to the understanding of large biomolecular assemblies. Over the last 10 or 20 years, low-resolution image reconstructions of large macromolecular structures from electron micrographs have evolved into a powerful method for studying the structure, assembly, and dynamics of such systems, as they yield 3D density maps under a wide range of biochemical conditions.[1,2] The resolution that can be achieved in a standard cryogenic electron microscopy (cryo-EM) reconstruction is often limited to a range of 15 to 30 Å. However, since large cellular systems are composed of individual protein and RNA subunits, it is possible to localize structure elements with high precision by docking the known atomic structures of these subunits into low-resolution EM densities. The hybrid, multi-resolution modeling of biomolecular assemblies holds much promise for the understanding of their function,[3–5] and the modeling will become increasingly attractive in the near future, when more and more low-resolution structures of large complexes and the atomic structures of their components become available.

A number of computational docking strategies have been proposed to bridge the resolution gap between low-resolution EM maps and atomic structures.[5] The most widely used method relies on the systematic maximization of the density cross-correlation of atomic models with electron density maps. In early studies,[6–10] programs for X-ray crystallographic refinement such as X-PLOR[11] and ESSENS[12] were adapted to calculate the density cross-correlation. Recently, more specific routines (COAN,[13] DOCKEM,[14] EMFIT[15]) based on density cross-correlation have been developed and applied for the docking of structures into EM maps.[16–19]

Several authors have pointed out limitations of the standard cross-correlation criterion for resolutions below 15 Å.[5,20] At low resolution, the correlation coefficient score lies within a narrow range of numerical values, making it more difficult to differentiate the correct from spurious orientations.[5,20] One concern is, therefore, that false positives may arise in the fitting, i.e. high-scoring results that deviate significantly from a known structure of the complex, in particular when the

Abbreviations used: EM, electron microscope(y); *SNR*, signal-to-noise ratio.

E-mail address of the corresponding author: pchacon@scripps.edu

0022-2836/02/030375–10 $35.00/0

low-resolution map includes extraneous densities that are not accounted for by the docked component. Also, at low resolution, densities show little variation inside of the molecules. By maximizing the density cross-correlation, one may inadvertently displace the template to regions of higher density in the interior of the target map.[14] This effect is observed when the component to be docked is significantly smaller than the low-resolution target. In this case, the translational placement is not constrained, and a number of viable positions exhibit corresponding optimal orientations that would accommodate the template in the target density. This complex docking ''landscape'' is the common scenario that one encounters when working with low-resolution EM data of large assemblies. In certain situations it is possible to ''rescue'' the performance of the density correlation criterion, e.g. by using structural constraints from biochemical data,[21] or by successive application of discrepancy mapping to simulate the molecular boundaries between individually docked domains.[18] In this work, however, we demonstrate that the problems are intrinsic to the density correlation measure itself and we seek to modify it conceptually to provide a more general solution.

We present a simple approach to overcome the limitations of cross-correlation by adding molecular contour or ''surface'' information to the fitting criterion. The consideration of the molecular contour by itself yields a more discriminative scoring function to localize small components in macromolecular structures. This method is not limited to single molecules; it can be applied to any kind of shape, and neither symmetry nor any other structural restriction is needed. As described in Materials and Methods, the contour information is extracted by filtering with a Laplacian operator. The Laplacian-based docking maximizes both density and contour overlap. Thereby, the filtering enhances the numerical scoring contrast among potential solutions, and provides more reliable and accurate docking results. The filtering is combined with a state-of-the-art 6D search algorithm based on Fourier correlation theory.

In the following, we demonstrate the superiority of the contour-matching criterion over the traditional density cross-correlation by using simulated low-resolution densities. We validate the performance of the algorithm described in detail in Materials and Methods. Finally, we present new docking results of the atomic structure of tubulin into an experimental map of the microtubule that sheds light on the interactions of the subunits in this helical assembly.

Results

Density correlation *versus* Laplacian correlation

The performance of the traditional density correlation criterion and the proposed contour-based Laplacian correlation criterion was evaluated using simulated low-resolution density maps that were calculated from known atomic structures by resolution lowering. In the first example (Figure 1), we sought to reconstruct the original atomic structure of RecA helicase[22] (PDB entry 2REC) by the docking of a monomer structure into the simulated low-resolution map of the entire hexamer. The simulated map was generated by interpolation of the mass-weighted atomic coordinates to a 3D lattice, and, subsequently, by convolving each lattice point with a Gaussian function to lower the resolution to 15 Å.[23] Correlation values were computed in an exhaustive 6D rigid-body search (see Materials and Methods). During the search, we kept track of the optimal (highest scoring) orientations for each discrete 3D position. The orientational sampling was sufficiently dense (9°, 30,481 sets of Euler angles) to provide a smooth scoring landscape for the fitting.

Figure 1 presents 2D cross-sections of the correlation values as a function of position in the plane of the hexamer, for density (Figure 1(a)) and Laplacian (Figure 1(d)) correlation. The density correlation maxima, on the one hand, are located on a ring that corresponds to a range of numerically degenerate docking possibilities of an isolated monomeric component within the simulated target density. Due to the lack of discriminative contrast, the six correct positions are indistinguishable from other spurious fits on the ring. Laplacian correlation, on the other hand, identifies the original positions unambiguously; six prominent peaks clearly mark the correct symmetry-related positions in the hexameric density.

In practice, the 6D exhaustive search is performed with a certain positional and angular sampling step size on a discrete lattice. To improve the accuracy of the docking, we optimized the search for the highest peak by using a finer local grid in the 6D space of alignment parameters. We carried out the optimization in the neighborhood of several of the highest peaks to avoid missing the true maximum (see Materials and Methods for details). Figure 1 shows the superposition of the target structure (red ribbon) with the highest-scoring off-lattice refinement results (green ribbon) for (b) and (c) density and (e) and (f) Laplacian correlation, respectively. Only the Laplacian correlation yields the correct reconstruction. In this idealized test system, a near-perfect registration can be achieved (root-mean-square deviation (rmsd) 0.23 Å), as shown in the close-up (Figure 1(f)). In contrast, the fit with highest-scoring density correlation (Figure 1(c)) is displaced to the interior of the donut-shaped density and rotated approximately 120° about the hexameric symmetry axis.

Similar results can be obtained when comparing the two criteria as a function of rotational space coordinates: the Laplacian correlation identifies the correct orientation, whereas density correlation yields additional score maxima corresponding to orientational mismatches. We present a series of

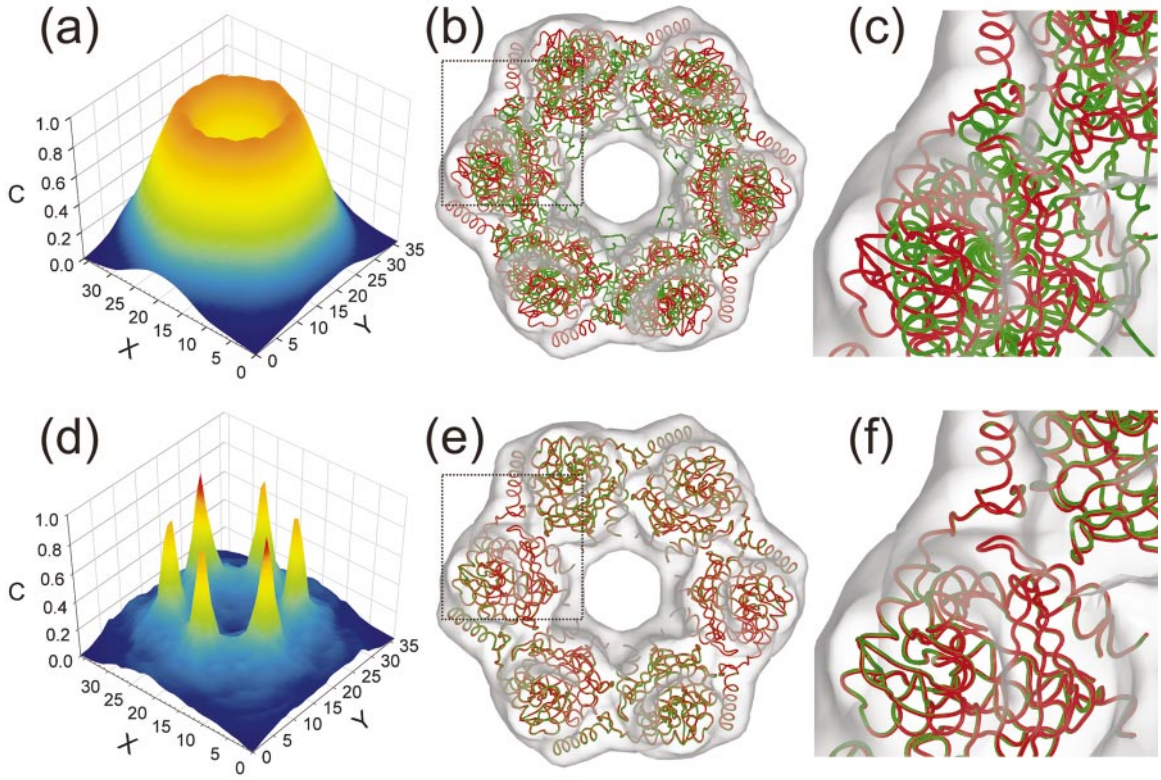

Figure 1. Docking performance of two correlation-based fitting criteria: RecA helicase. The RecA (PDB entry 2REC) monomer was docked into a simulated density map of the hexamer at 15 Å resolution (see the text) using (a)-(c) density and (d)-(f) Laplacian correlation functions. In both cases, an exhaustive 6D search with a translational and rotational step size of 5 Å and 9°, respectively, was carried out, followed by an off-lattice optimization. The Laplacian correlation function was calculated using equation (4), and the density cross-correlation was calculated similarly without the Laplacian filter. In (a) and (d), the normalized correlation values are represented on a 2D cross-section of the translational space by a color spectrum from blue ($C = 0$) to red ($C = 1$). In (b) and (e), the six symmetry-related highest-scoring fits (green ribbon) are superimposed with the original structure (red ribbon) that was the basis for the simulated target density map. A close-up view of the structures is given in (c) and (e). Plots of correlation coefficients in Figures 1-4 were created with SPSS SigmaPlot; molecular structures and densities were visualized with Situs,[20] VMD[44] and Raster3D.[45]

rotational scans that were made around the correct (Figure 1(e)) and spurious (Figure 1(b)) fits found earlier with Laplacian and density correlation, respectively. Figure 2 shows the coefficients for Laplacian (upper panels) and density (lower panels) correlation. In the Laplacian case, maximal values are achieved only for the correct fit near the correct orientation. In the density case, there are similar scores for correct and spurious fits. Clearly, the Laplacian exhibits an enhanced fitting contrast between correct and spurious fits, and thus constitutes the more discriminative criterion. This is apparent from the different numerical scale of the plots, and from the width and shape of the peaks corresponding to the local maxima.

We have tested the performance of the correlation coefficients as a function of the resolution on four oligomeric systems: helicase[22] (hexamer), catalase[24] (tetramer), oxidoreductase[25] (trimer), and thiolase[26] (dimer). Again, the task was to reproduce the correct position and orientation of individual monomeric components in the simulated oligomeric maps that were computed here at levels of resolution from 6 Å to 34 Å. Figure 3(a) presents the rmsd between the highest-scoring results and the target structures used for generating the low-resolution densities. Initially, at high resolution, both docking criteria perform very well, but eventually the fitting breaks down and large rmsd values are obtained when the resolution becomes too low. The limiting resolutions beyond which it becomes impossible to detect the correct fits unequivocally range from 8 to 15 Å for the density correlation, and from 27 to 34 Å for the Laplacian correlation in the four test systems. Variabilities in the resolution limits indicate a system-specific performance. The results indicate that Laplacian, but not density, correlation performs well in the range of low resolutions (>15 Å) relevant in EM.

To investigate how well correct fits are distinguished from spurious fits, we have evaluated the fitting contrast for each criterion. We compute the fitting contrast, i.e. the ratio between the correlation coefficients of highest (correct) and second-highest (spurious) peaks as a function of the resolution. The enhanced fitting contrast obtained with

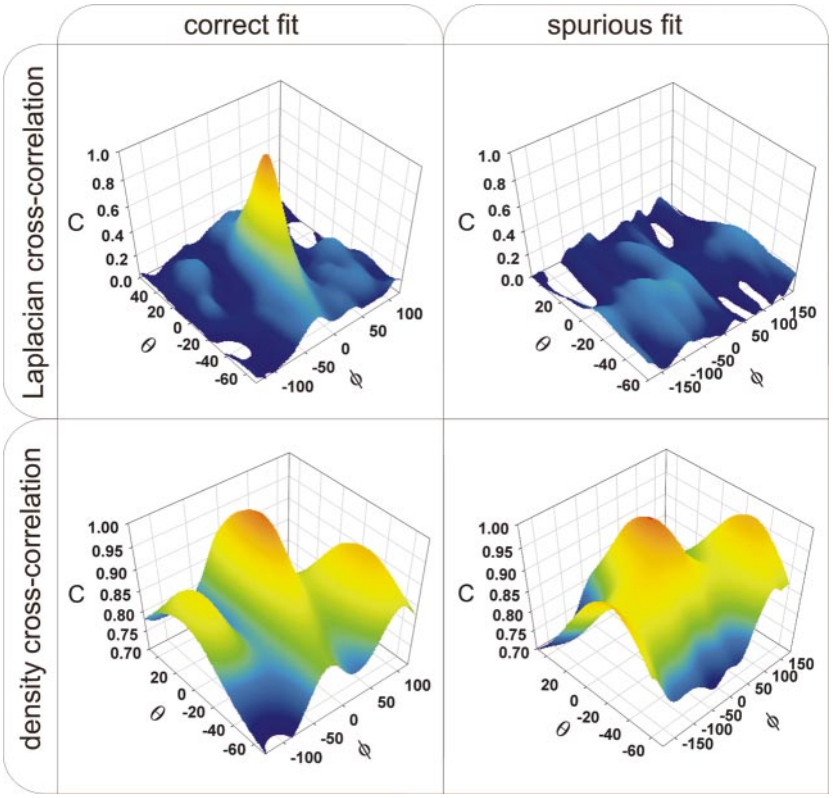

Figure 2. Rotational space dependence of the RecA helicase docking scores. Four rotational space scans of the scoring functions were performed using Laplacian (upper panels) and density cross-correlation (lower panels) starting from correct (left panels) and spurious (right panels) docking positions, shown in Figure 1(e) and (c), respectively. The normalized correlation values are presented as a function of two Euler angles (θ and ϕ), computed in sampling steps of 1°. Note the different numerical range of the correlations (0-1 for Laplacian and 0.7-1 for density correlation).

the Laplacian filter is evident from the results summarized in Figure 3(b). The average of the fitting contrast using cross-correlation varies from only 1.5 to 0.9 for resolutions between 6 Å and 12 Å, respectively, whereas the Laplacian criterion can achieve a fitting contrast from 89 to 1.5 for resolutions from 12 Å to 25 Å.

We implemented a multi-molecule refinement procedure that moves individual components independently and thereby accounts for all of the density in a multi-component map (described in Materials and Methods). This approach was used to test the effect of extraneous densities on the Laplacian-based docking. As can be seen in Figure 3(a) (dotted lines), the multi-molecule search slightly improves the Laplacian-based docking results by up to 1 Å rmsd in some cases, but the method does not extend the viable resolution range to lower resolution. The results indicate that the Laplacian criterion by itself is intrinsically robust when extraneous densities are present. The improvements achieved with multi-molecule docking in our idealized (noise-free) test systems are relatively small, so the application can be considered an optional step in the fitting.

To test the ability of the Laplacian filter under noise, we performed a systematic study of the fitting accuracy as a function of the signal to noise ratio (*SNR*; for definition, see the legend to Figure 3(c)). We degraded the *SNR* of the 20 Å resolution simulated maps (voxel size 4 Å) for all test proteins by adding Gaussian noise and examined the fitting accuracy as a function of the random shot noise amplitude. As can be seen in Figure 3(c), the Laplacian filter is a very robust choice for docking under noise. In fact, it is possible to obtain unique fits with an accuracy of <3 Å for $SNR > 0.5$. The contour information extracted by the Laplacian filter, even under noisy conditions, is sufficient to guide the docking and to successfully retrieve the correct pose.

Application to the microtubule structure

Microtubules are pseudo-helical cylindrical protein assemblies that organize cellular architecture and cytoplasmic membrane traffic by means of microtubule associated proteins. The 24 nm diameter tubes typically comprise 13 longitudinal protofilaments, each of them formed by head to tail association of α-β-tubulin dimers. The atomic structure of the tubulin dimer[27] (PDB entry 1TUB) has been placed earlier into the 20 Å resolution map to generate a near-atomic model of the microtubule.[28] In the earlier study, the conformational search was relatively limited and the relative orientation of α and β-tubulin was not a refinement parameter. Here, we present the docking results of α and β tubulin subunits into the same EM map using Laplacian correlation. The docking was performed with each subunit independently; therefore neither interactions nor relative orientations of subunits were considered *a priori*, as was the case earlier.[28]

Despite the absence of structural constraints on the global search, we found a unique high-scoring fit for each tubulin in a specific direction along the

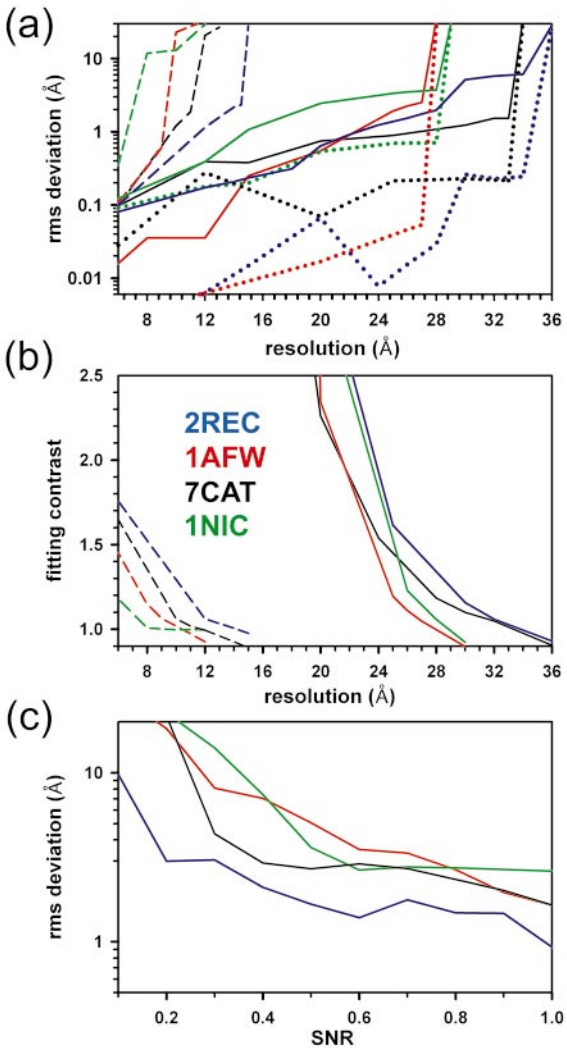

Figure 3. (a) Docking performance of two correlation-based fitting criteria: Resolution dependence. The validation experiment shown in Figure 1 was repeated with four test systems at various resolution levels. The test examples include RecA helicase[22] (blue lines; PDB entry 2REC), thiolase[26] (red lines; PDB entry 1AFW), catalase[24] (black lines; PDB entry 7CAT) and an oxidoreductase[25] (green lines; PDB entry 1NIC). The docking rmsd between the fitted component and the original oligomeric structure, which was the basis for the simulated target density map, is shown as a function of the resolution: continuous lines, Laplacian correlation; broken lines, density correlation; dotted lines, multi-docking refinement of the Laplacian-based docking results. In all cases, a rotational step size of 9° has been used for the exhaustive 6D search. The voxel size of the maps and the initial translational sampling step size was 4 Å for resolutions >12 Å, 3 Å for resolutions between 12-9 Å, and 2 Å for resolutions <8 Å. (b) Comparison of the fitting contrast obtained using Laplacian (continuous lines) and density (broken lines) fitting criteria shown as a function of the resolution. The fitting contrast is defined as the ratio between the correlation coefficients of the highest (correct fit) and the second highest (spurious fit) peaks found in the 6D searches performed in (a). The color codes are as in (a). (c) *SNR* dependence of docking precision. Equivalent restoration experiments were performed as in (a) but with the addition of Gaussian noise to the 20 Å resolution simulated maps (voxel size 4 Å) of each test protein. The *SNR* experiments were repeated five times for each protein. The docking rmsd between the atomic structure used to generate the noise-free map and the docked structures obtained with the noisy map is shown as function of the *SNR*, which is defined as: $SNR = \mathrm{var}(o)/\mathrm{var}(n)$, where $var(o)$ and $var(n)$ are the variance of the original map and the variance of the Gaussian noise, respectively.

protofilament. The correlation values for β-tubulin are shown in Figure 4(a) for a cross-section of translational space. The pointed correlation peaks that follow the characteristic twist of the microtubule helix each correspond to a unique, highly localized orientation of the subunit, as demonstrated by the rotational dependence of the correlation score in Figure 4(b).

It was not possible to distinguish the positions of α and β subunits along the protofilament at 20 Å resolution. Results equivalent to those shown in Figure 4(a) and (b) have been obtained with α tubulin (not shown). Therefore, by alternating α and β tubulin, one can form two equivalent atomic models of the protofilaments, one of which is shown in Figure 4(c) and (d). A multi-molecule refinement (see Materials and Methods) of the tubulin components against the EM data did not result in significant changes in the two models. The corresponding αβ-dimer in either model is very similar to the electron crystallographic dimer structure in Zn^{2+}-induced polymers.[27] The rmsd between the known atomic structure of the dimer and the αβ-dimer extracted from our docking results is only 1.8 Å (computed with lsqman[29]). These observed differences between the models are well within the experimental uncertainty of the underlying EM map. The intra-subunit contact maps in Figure 5(a) reflect this similarity of the relative orientations of the two dimers.

Relative to the model described by Nogales *et al.*,[28] the rmsd exhibited by the βα inter-dimer region (Figure 5(b)) in our model is only 2.3 Å. The consideration of an entire protofilament results in a slightly greater discrepancy of 3 Å. Apart from the high resemblance, it is noteworthy that our model suggests a slight relative rearrangement between subunits that opens the dimers slightly and that reduces the lateral interactions with respect to the Nogales *et al*. model (Figures 4(e) and (c), and 5(c)). The major lateral contact, i.e. the interaction between the end of helix H3[28] and loop H1-S3[28] on one dimer with loop M[28] on the apposing dimer remains intact in our model.

Discussion

The novel registration of molecular contours introduced here was inspired by the common practice of ''visual docking'' that is still in use in many EM laboratories.[5] Human vision is a powerful pattern recognition machine that outperforms many

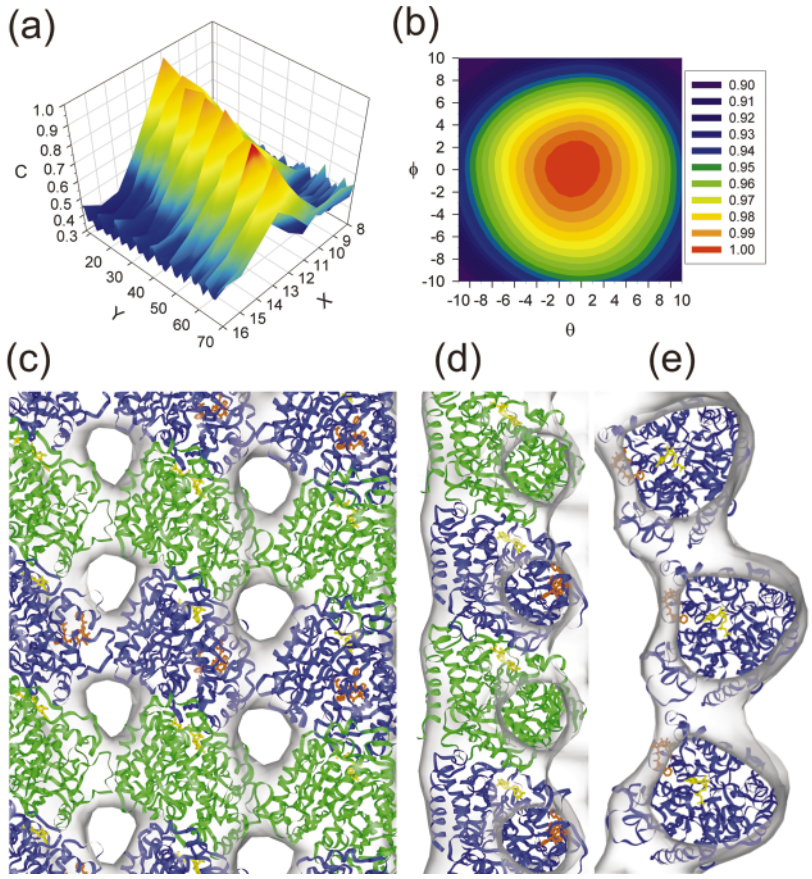

Figure 4. Construction of a microtubule model. The αβ tubulin[27] subunits (PDB entry 1TUB) were docked independently into an experimental 20 Å resolution microtubule map[46] (5 Å voxel spacing). For simplicity, the docking was performed only on three of the 15 protofilaments. (a) Translational space-dependence of the Laplacian correlation. This 2D subset of the correlation coefficients corresponds to the maximum rotational values found with an angular sampling step of 6°. Eight prominent peaks correspond to the fit of individual β subunits into the protofilament. (b) Rotational space-dependence of the Laplacian correlation. A rotational scan with sampling width 1° was performed around the optimal orientation of one fitted β-tubulin, as a function of two Euler angles. Equivalent results can be obtained by rotating the other two possible pairs of Euler angles (data not shown). (c)-(e) Final fitting results: (c) outside view, (b) lateral view, (e) cross-section of the microtubule seen from the plus end. The experimental map is represented as a transparent surface, taxotere in orange, nucleotide in yellow, and the α and β subunits are shown in green and blue, respectively. The model coordinates in PDB format can be requested from the corresponding author.

computer algorithms in complex 3D recognition tasks. In visual docking, researchers visualize the isocontour of a low-resolution object at a certain threshold level on the computer screen, and subsequently place the template structure into the isosurface envelope to bring discernible features into register. Despite the subjectivity of the manual placement, this approach is often surprisingly successful. In fact, it is a challenging task to devise computational criteria that perform equally well or better than the human eye. For instance, it has been pointed out that extraneous densities may act as noise that worsen a computational fit with the standard density correlation[14] as judged by the visual agreement of discernible surface features.

Here, we sought to emulate *in silico* the simultaneous registration of surface and volume features that is at the heart of visual docking as performed by an experienced electron microscopist. The distinction of the contour (positive values) and the interior (negative values) of a low-resolution object by the sign of the Laplacian-filtered density is well suited for the correlation criterion, since both the overlap of similarly shaped positive contours and the overlap of similarly distributed negative volumes will give positive contributions to the Laplacian correlation coefficient (Materials and Methods; equation (3); Figure 6).

In the particularly challenging situation when the structure to be docked represents only a portion of the EM density, our evaluations exposed limitations of the traditional density correlation criterion and demonstrated the practical advantages of Laplacian correlation. The emergence of false positives (Figures 1 and 2) that correspond to a drifting of the template structure to the highest-density regions limits the effectiveness of the density correlation criterion at low resolution. In the four oligomeric test systems studied, Laplacian correlation extends the limiting resolution above which it is possible to restore the original oligomeric structure unequivocally, from ~10 Å to ~30 Å (Figure 3A) relative to density correlation. These results were obtained with idealized (noise-free) systems. In practice, we foresee two possible limitations of Laplacian fitting. (i) When the interior features of a given map are sufficiently large to guide the volumetric docking. In this case, a reliable docking can be obtained with both cross-correlation and Laplacian correlation. This typically occurs at higher resolutions (approx <10 Å). (ii) When the contour information is compromised. This occurs if the map lacks a sufficiently large sur-

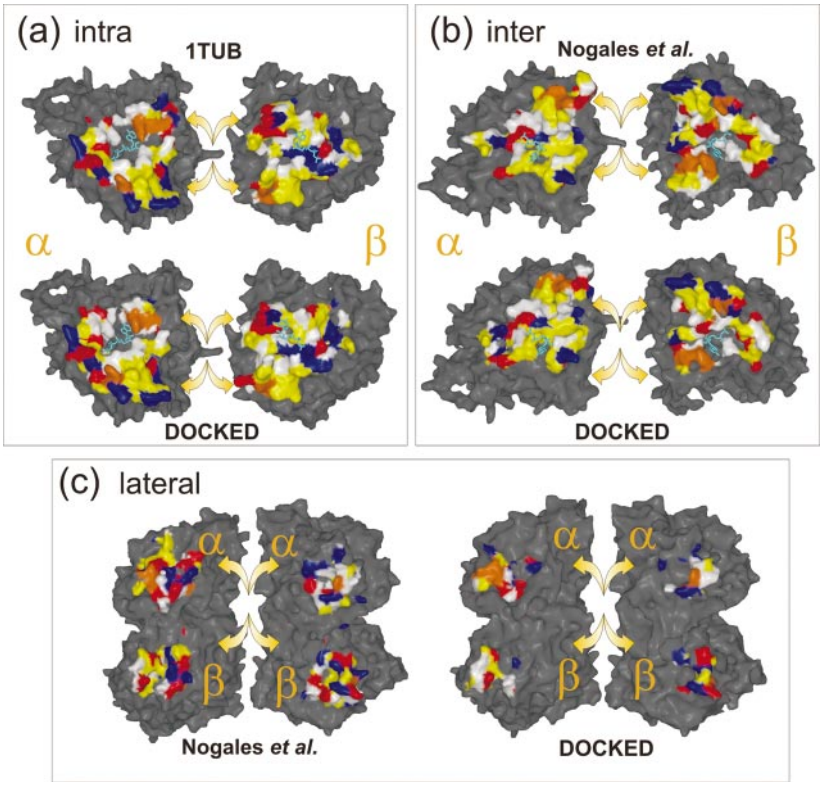

Figure 5. Comparison of model contact surfaces. The contact surfaces are colored only for interface atoms that lie within 4 Å of each other. Non-polar residues (Ala, Gly, Leu, Val, Ile, Pro, Met, Phe) are yellow, polar residues (Ser, Thr, Cys, His, Asn, Gln) are white, acidic residues (Asp, Glu) are red, basic residues (Arg, Lys) are blue, and Tyr as well as Trp are orange. (a) Intra-dimer (longitudinal) contact: comparison of the surface of the Zn sheet tubulin dimer[27] (top) and docking model (bottom). (b) Inter-dimer (longitudinal) contact: comparison of the Nogales *et al.* model[28] (top) with our docking model (bottom). (c) Lateral contact: comparison of the Nogales *et al.* model[28] (left) with our docking model (right). The contact surfaces were generated with the program Spock (http://quorum.tamu.edu/spock/) and displayed with Raster3D.[45] Detailed listings of the contact surfaces can be requested from the corresponding author.

face with which to define a reliable guiding contour. In this situation, typically at lower resolution (>30 Å), volumetric docking will likewise be limited.

Another possible limitation of the Laplacian is the effect of noise in experimental EM maps. One can show that a Laplacian convolution in real space corresponds to the filtering with a harmonic function in Fourier space.[30] Hence, the Laplacian acts as a high-pass filter that suppresses low frequencies. This property is highly desirable, since the frequency filtering reduces distracting low-resolution terms in real space that originate from extraneous densities beyond the spatial extent of the template. However, the harmonic function also amplifies the high-resolution frequency noise present in experimental EM maps. We argue that in real space, the high-frequency noise is distributed at random, as opposed to any localized extraneous densities, and will therefore have less impact on the global correlation due to a cancellation of errors than a localized low-frequency perturbation (Materials and Methods, equation (3)). The tests carried out on degrading the *SNR* of the 20 Å simulated maps (Figure 3(c)) demonstrate this point and the robustness of the proposed approach. As other authors suggest,[14] to validate the proposed approach further, tests should ideally be performed with real EM data. Unfortunately, experimental EM maps for which atomic structures are known are rare. Future development of a database of EM structures will facilitate access to this valuable data,[31] and will allow the complete validation of any fitting method. It will also permit the construction of an adequate framework for future improvements.

To illustrate the performance of the Laplacian criterion, we tested the fitting of tubulin components to an experimental 20 Å resolution microtubule map. Without any *a priori* consideration of the relative orientations of α and β tubulins, we were able to reconstruct the atomic structure of the αβ-tubulin dimer (within the constraints of the microtubule EM density) to within 2 Å of the known dimer crystal structure. Our atomic model of the microtubule is in good agreement with the previously reported docking of the dimer by Nogales *et al.*[28] and with a more recent microtubule model.[28,32] Minor differences in our model relative to that of Nogales *et al.*[28] suggest a slight opening of tubulin subunits through a reduction of the lateral protofilament interactions. The resulting more pronounced fenestration of the microtubule surface (Figure 4(c) and (d)) would be consistent with the experimentally observed fast access of taxol molecules to the microtubule lumen.[33,34]

In summary, the novel rigid-body registration algorithm introduced here offers a robust docking method for the fitting of atomic subunits into low-resolution densities of macromolecular structures. The inclusion of contour information in the docking enhances the contrast among potential solutions and, thereby, yields a discriminative scoring function for resolutions as low as 30 Å. We hope

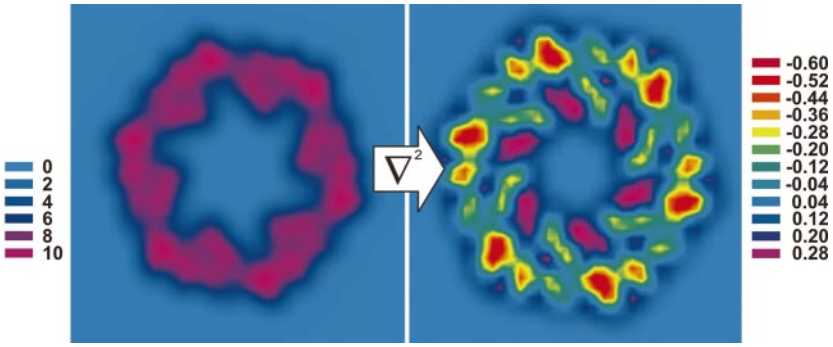

Figure 6. Laplacian filter. On the left, the contour plot represents a cross-section of the 15 Å simulated density map of the RecA hexameric structure. On the right, the same cross-section is represented after application of the Laplacian filter. Note that positive values correspond to a range of colors from blue to magenta, and negative values to colors ranging from red to blue.

that the developed methodology will become a valuable tool in the study of the architecture and dynamics of large macromolecular assemblies.

Materials and Methods

6D rigid-body search using Fourier template convolution

The geometric match between two molecules A and B can be measured by the linear cross-correlation defined as:

$$C_{x,y,z} = \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} a_{l,m,n} \times b_{l+x,m+y,n+z} \qquad (1)$$

where a and b are the corresponding shape functions on a 3D lattice formed by N^3 points (with lattice indexes l, m, n), for a given translation vector (x, y, z). The use of this criterion for multi-resolution requires that the data sets be compared at like resolution. To this end, the atomic structure, B, is first projected to the cubic lattice of the EM data A by tri-linear interpolation, and subsequently, each lattice point $b_{l,m,n}$ is convolved with a Gaussian function g in real space.[23] Therefore, the density correlation can be expressed as:

$$C_{x,y,z} = \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} a_{l,m,n} \times (g \otimes b_{l+x,m+y,n+z}) \qquad (2)$$

The major innovation of this work is the inclusion of ''surface'' information in the volume docking procedure. In the absence of hard boundaries, the contour of a low-resolution object is contained in the 3D edge information instead of a 2D surface. In principle, any 3D edge enhancement filter can be used to extract this information;[35] here, we selected the Laplacian filter (∇^2) for its simplicity and low computational cost. High-pass filters like the Laplacian operator are widely used in general image processing.[2] On a discrete lattice, the Laplacian filter kernel is defined by the finite difference approximation:[35]

$$\nabla^2 a = -6a_{l,m,n} + a_{l+1,m,n} + a_{l-1,m,n} + a_{l,m+1,n} + a_{l,m-1,n} + a_{l,m,n+1} + a_{l,m,n-1}$$

The convolution of 3D maps with this kernel gives an approximation of the second derivative of the scalar density, typically taking positive values at the molecular contour and negative values in the interior volumes.[35] Figure 6 demonstrates the effect of the Laplacian filter.

The modified Laplacian correlation coefficient is defined as:

$$C_{x,y,z} = \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}(\nabla^2 \otimes a_{l,m,n}) \times (\nabla^2 \otimes g \otimes b_{l+x,m+y,n+z}) \qquad (3)$$

Since convolution is associative, both resolution lowering ($g \otimes \ldots$) and Laplacian filtering were performed in one step by convolving b with a Mexican Hat operator, $M = \nabla^2 \otimes g$.

The 6D rigid-body search procedure used in this study was adapted from the protein-protein shape complementarity algorithm described by Katchalski-Katzir *et al.*,[36] which is widely used in ''exterior'' (ligand-receptor) docking programs, such as FTDOCK,[37] DOT[38] and GRAMM.[39] The algorithm takes advantage of Fourier correlation theory and the fast Fourier transform (*FFT*) to scan rapidly the translation of a probe molecule relative to a fixed reference molecule. By virtue of the Fourier correlation theorem, the Laplacian correlation (as a function of translational displacements) can be computed very efficiently as the reverse Fourier transform of the product of two Fourier transforms: one that corresponds to the Mexican Hat-filtered atomic structure, and one that corresponds to the complex conjugate of the Laplacian-filtered low resolution map:

$$C = FFT^{-1}[FFT(\nabla^2 \otimes a_{l,m,n})^* \times FFT(M \otimes b_{l,m,n})] \qquad (4)$$

Note that C is automatically evaluated for all translations on the 3D lattice. The computational cost (excluding the application of the filters) is $N^3 \log N^3$ for each *FFT*, much less than the N^6 operations required in the direct evaluation of equation (3) for all translations (x, y, z). The *FFT* corresponding to the fixed molecule A can be omitted from all but the first calculation.

An orientational search complemented the Fourier-based evaluation of C. To this end, the target A was fixed and the template B was rotated systematically using a homogeneous distribution of Euler rotation angles.[40] For each set of Euler angles, C was evaluated using equation (4). A number of algorithms for computing homogeneous distributions of Euler angles were implemented[37,41,42] that essentially performed equally well for the small angular step sizes (6-12°) needed to detect suitable high-scoring orientations. After each iteration of the orientational search, a 3D lattice of the maximal coefficients C found so far was updated, and the optimal set of Euler angles was saved for each (x, y, z).

Off-lattice refinement

Following the global exploration of the 6D search space at discrete sampling points, we implemented a local, off-lattice gradient ascent maximization of the correlation coefficient (as a function of the six rigid-body degrees of freedom), to enable sub-lattice accuracy in our fitting. A simple, automatic peak detection filter was applied to the correlation values on the translational lattice, yielding a ranked list of potential docking solutions. The peak detection involved (i) the smoothing of the correlation values over nearest-neighbor voxels to ignore any single-voxel noise spikes, (ii) a convolution with a Laplacian filter for the detection of local maxima, and (iii) the sorting of the peaks according to their correlation score. Subsequently, Powell's quadratically convergent optimization method[43] was applied to refine a user-defined number of maxima to within sub-lattice precision. In the presence of unique maxima, finding the highest-scoring result is sufficient, but additional maxima may be relevant if additional symmetry-related fits are feasible (Figures 1(d) and 4(a)).

Multi-molecule docking

We implemented an additional refinement step that enables multiple components to rotate and to translate independently. This approach adds, for each independent component, six dimensions to the complexity of the search space. Therefore, an exhaustive exploration becomes prohibitively expensive for more than one independent subunit, but a local optimization is feasible once a start model has been constructed, e.g. from symmetry-related units in oligomeric structures. The advantage of multi-molecule docking is that all of the density in a multi-component map can be accounted for by atomic structures, thereby reducing the effect of extraneous densities on the docking of a given subunit. A simulated annealing method[43] was used to optimize the input structure. In this technique, an artificial ''temperature'' is introduced to stochastically move the molecules. The system was cooled gradually to settle in a new optimum fit, similar to the annealing technique used in metallurgy when a molten metal reaches its crystalline state. In our case, the temperature was initialized to provide perturbations according to angular step size of the rotational search and the grid size of the lattice. The stochastic sampling of the high-dimensional search space was thereby restricted to the local vicinity of the start model. After the annealing, a local Powell optimization (see above) was performed.

Performance and availability

The computational cost of the Laplacian-based search depends on both the size of the density map and on the angular sampling. For the RecA helicase example, approximately six CPU hours were required to carry out an exhaustive search with 9° angular steps on a $60 \times 60 \times 60$ lattice (850 Mhz Pentium III PC/Linux). The algorithms were implemented in a novel utility termed CoLoRes (contour low resolution docking) that will be distributed with the low-resolution docking package Situs.[20] CoLoRes is written in C and supported by a fast FFT library, FFTW 2.3. (http://www.fftw.org). The source code of CoLoRes will be made available in Situs version 2.0 at http://situs.scripps.edu.

Acknowledgments

Support by the National Institutes of Health (R01 GM62968 and P41 RR 12255) and the LJIS Interdisciplinary Training Program/Burroughs Wellcome Fund are gratefully acknowledged. We thank Ron Milligan and Francisco Asturias, and their co-workers for helpful comments and suggestions, Essam Metwally and Julio Kovacs for reading the manuscript, and finally Eva Nogales for kindly providing the microtubule model used in our structure comparison.

References

1. DeRosier, D. J. & Harrison, S. C. (1997). Macromolecular assemblages. Sizing things up. *Curr. Opin. Struct. Biol.* **7**, 237-238.
2. Frank, J. (1996). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, San Diego.
3. Baker, T. S. & Johnson, J. E. (1996). Low resolution meets high: towards a resolution continuum from cells to atoms. *Curr. Opin. Struct. Biol.* **6**, 585-594.
4. Baumeister, W. & Steven, A. C. (2000). Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem. Sci.* **25**, 624-631.
5. Wriggers, W. & Chacón, P. (2001). Modeling tricks and fitting techniques for multi-resolution structures. *Structure,* **9**, 779-788.
6. Wikoff, W. R., Wang, G., Parrish, C. R., Cheng, R. H., Strassheim, M. L., Baker, T. S. & Rossmann, M. G. (1994). The structure of a neutralized virus: canine parvovirus complexed with neutralizing antibody fragment. *Structure,* **2**, 595-607.
7. Cheng, R. H., Kuhn, R. J., Olson, N. H., Rossmann, M. G., Choi, H. K., Smith, T. J. & Baker, T. S. (1995). Nucleocapsid and glycoprotein organization in an enveloped virus. *Cell,* **80**, 621-630.
8. Grimes, J. M., Jakana, J., Ghosh, M., Basak, A. K., Roy, P., Chiu, W. *et al*. (1997). An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron cryomicroscopy. *Structure,* **5**, 885-893.
9. Hewat, E. A. & Blaas, D. (1996). Structure of a neutralizing antibody bound bivalently to human rhinovirus 2. *EMBO J.* **15**, 1515-1523.
10. Kolatkar, P. R., Bella, J., Olson, N. H., Bator, C. M., Baker, T. S. & Rossmann, M. G. (1999). Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor. *EMBO J.* **18**, 6249-6259.
11. Brünger, A. T. (1992). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR*, Yale University, Connecticut.
12. Kleywegt, G. J. & Jones, T. A. (1997). Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallog. sect. D,* **53**, 179-185.
13. Volkmann, N. & Hanein, D. (1999). Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J. Struct. Biol.* **125**, 176-184.
14. Roseman, A. M. (2000). Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallog. sect. D,* **56**, 1332-1340.

15. Rossmann, M. G. (2000). Fitting atomic models into electron-microscopy maps. *Acta. Crystallog. sect. D,* **56**, 1341-1349.
16. Pletnev, S. V., Zhang, W., Mukhopadhyay, S., Fisher, B. R., Hernandez, R., Brown, D. T. *et al*. (2001). Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. *Cell,* **105**, 127-136.
17. Rye, H. S., Roseman, A. M., Chen, S., Furtak, K., Fenton, W. A., Saibil, H. R. & Horwich, A. L. (1999). GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell,* **97**, 325-338.
18. Volkmann, N., Hanein, D., Ouyang, G., Trybus, K. M., DeRosier, D. J. & Lowey, S. (2000). Evidence for cleft closure in actomyosin upon ADP release. *Nature Struct. Biol.* **7**, 1147-1155.
19. Volkmann, N., DeRosier, D., Matsudaira, P. & Hanein, D. (2001). An atomic model of actin filaments cross-linked by fimbrin and its implications for bundle assembly and function. *J. Cell. Biol.* **153**, 947-956.
20. Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185-195.
21. Hanein, D., Volkmann, N., Goldsmith, S., Michon, A. M., Lehman, W., Craig, R. *et al*. (1998). An atomic model of fimbrin binding to F-actin and its implications for filament crosslinking and regulation. *Nature Struct. Biol.* **5**, 787-792.
22. Yu, X. & Egelman, E. H. (1997). The RecA hexamer is a structural homologue of ring helicases. *Nature Struct. Biol.* **4**, 101-104.
23. Wriggers, W. & Birmanns, S. (2001). Using Situs for flexible and rigid-body fitting of multi-resolution single molecule data. *J. Struct. Biol.* **133**, 193-202.
24. Fita, I. & Rossmann, M. G. (1985). The NADPH binding site on beef liver catalase. *Proc. Natl Acad. Sci. USA,* **82**, 1604-1608.
25. Adman, E. T., Godden, J. W. & Turley, S. (1995). The structure of copper-nitrite reductase from *Achromobacter cycloclastes* at five pH values, with NO_2-bound and with type II copper depleted. *J. Biol. Chem.* **270**, 27458-27474.
26. Mathieu, M., Modis, Y., Zeelen, J. P., Engel, C. K., Abagyan, R. A., Ahlberg, A. *et al*. (1997). The 1.8 Å crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: implications for substrate binding and reaction mechanism. *J. Mol. Biol.* **273**, 714-728.
27. Nogales, E., Wolf, S. G. & Downing, K. H. (1998). Structure of the alpha beta tubulin dimer by electron crystallography. *Nature,* **391**, 199-203.
28. Nogales, E., Whittaker, M., Milligan, R. A. & Downing, K. H. (1999). High-resolution model of the microtubule. *Cell,* **96**, 79-88.
29. Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallog. sect. D,* **52**, 842-857.
30. Bracewell, R. N. (1986). *The Fourier Transform and Its Applications*, 2nd edit., McGraw-Hill, New York.
31. Carazo, J. M. & Stelzer, E. H. (1999). The BioImage Database Project: organizing multidimensional biological images in an object-relational database. *J. Struct. Biol.* **125**, 97-102.
32. Meurer-Grob, P., Kasparian, J. & Wade, R. H. (2001). Microtubule structure at improved resolution. *Biochemistry,* **40**, 8000-8008.
33. Díaz, J. F., Strobe, R., Engelborghs, Y., Souto, A. A. & Andreu, J. M. (2000). Molecular recognition of taxol by microtubules. Kinetics and thermodynamics of binding of fluorescent taxol derivatives to an exposed site. *J. Biol. Chem.* **275**, 26265-26276.
34. Díaz, J. F., Valpuesta, J. M., Chacón, P., Diakun, G. & Andreu, J. M. (1998). Changes in microtubule protofilament number induced by Taxol binding to an easily accessible site. Internal microtubule dynamics. *J. Biol. Chem.* **273**, 33803-33810.
35. Russ, J. C. (1995). *The Image Processing Handbook,* 2nd edit., CRC Press, Florida.
36. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. & Vakser, I. A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl Acad. Sci. USA,* **89**, 2195-2199.
37. Gabb, H. A., Jackson, R. M. & Sternberg, M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **272**, 106-120.
38. Mandell, J. G., Roberts, V. A., Pique, M. E., Kotlovyi, V., Mitchell, J. C., Nelson, E. *et al*. (2001). Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* **14**, 105-113.
39. Vakser, I. A., Matar, O. G. & Lam, C. F. (1999). A systematic study of low-resolution recognition in protein-protein complexes. *Proc. Natl Acad. Sci. USA,* **96**, 8477-8482.
40. Goldstein, H. (1980). *Classical Mechanics*, 2nd edit., Addison-Wesley, New York.
41. O'Rourke, J. (1998). *Computational Geometry in C*, 2nd edit., Cambridge University Press, New York.
42. Roberts, V. A. & Pique, M. E. (1999). Definition of the interaction domain for cytochrome c on cytochrome c oxidase. III. Prediction of the docked complex by a complete, systematic search. *J. Biol. Chem.* **274**, 38051-38060.
43. Press, H., Teukolsky, S. A., Flannery, B. P. & Vetteerling, W. T. (1992). *Numerical Recipes in C. The Art of Scientific Computing*, 2nd edit., Cambridge University Press, New York.
44. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD - Visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38.
45. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.
46. Sosa, H., Dias, D. P., Hoenger, A., Whittaker, M., Wilson-Kubalek, E. & Sablin, E. *et al*. (1997). A model for the microtubule-ncd motor protein complex obtained by cryo-electron microscopy and image analysis. *Cell,* **90**, 217-224.

Edited by W. Baumeister

(Received 28 August 2001; received in revised form 15 January 2002; accepted 16 January 2002)