Structure
Article

Computational Reconstruction of Multidomain Proteins Using Atomic Force Microscopy Data

Minh-Hieu Trinh,[1] Michael Odorico,[1] Michael E. Pique,[2] Jean-Marie Teulon,[1] Victoria A. Roberts,[3] Lynn F. Ten Eyck,[3,4] Elizabeth D. Getzoff,[2] Pierre Parot,[1] Shu-wen W. Chen,[1,5] and Jean-Luc Pellequer[1,*]
[1]CEA, iBEB, Department of Biochemistry and Nuclear Toxicology, F-30207 Bagnols sur Cèze, France
[2]Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
[3]San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA
[4]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA
[5]Present address: 13 Avenue de la Mayre, 30200 Bagnols sur Cèze, France
*Correspondence: jlpellequer@cea.fr
DOI 10.1016/j.str.2011.10.023

SUMMARY

Classical structural biology techniques face a great challenge to determine the structure at the atomic level of large and flexible macromolecules. We present a novel methodology that combines high-resolution AFM topographic images with atomic coordinates of proteins to assemble very large macromolecules or particles. Our method uses a two-step protocol: atomic coordinates of individual domains are docked beneath the molecular surface of the large macromolecule, and then each domain is assembled using a combinatorial search. The protocol was validated on three test cases: a simulated system of antibody structures; and two experimentally based test cases: Tobacco mosaic virus, a rod-shaped virus; and Aquaporin Z, a bacterial membrane protein. We have shown that AFM-intermediate resolution topography and partial surface data are useful constraints for building macromolecular assemblies. The protocol is applicable to multicomponent structures connected in the polypeptide chain or as disjoint molecules. The approach effectively increases the resolution of AFM beyond topographical information down to atomic-detail structures.

INTRODUCTION

Protein structures are made of protein domains that can evolve, function, and fold independently of the rest of the protein chain (Ponting and Russell, 2002). Discovery of single-domain protein families has now saturated (Geer et al., 2002) and almost all current growth comes from multidomain architectures that are the combination of single domains. Therefore, all novelty in multidomain architecture families arises from the arrangement of known single domains within a protein. Structural genomics programs such as the second phase of the Protein Structure Initiative are aiming at further filling the gap in unknown 3D structures of domains (Dessailly et al., 2009). Traditional structural biology techniques face a great challenge to determine the structure at the atomic level of such large and flexible macromolecules (Mueller et al., 2007). Additional techniques such as cryoelectron microscopy (Cryo-EM) (Murata et al., 2010) or electron crystallography (Wisedchaisri and Gonen, 2011), or small angle X-ray scattering (SAXS) (Różycki et al., 2011) are becoming appealing alternatives. However, structural damages caused either by freeze-thawing samples or by high electron doses limit high-resolution reconstructions in EM (Zhou, 2008), whereas SAXS requires a large quantity of materials (mM range) as magnitude of the useful signal is proportional to the number of particles (Petoukhov and Svergun, 2007). More and more, computational techniques provide a contribution to the challenge.

Protein structure prediction methods differ in terms of the required input information and the quality of the output structures (Schwede et al., 2009). Increasingly, integrative methods rely on more than one type of information, especially for the structural characterization of protein assemblies (Alber et al., 2008). Even low-resolution biophysical and biochemical data can provide a rich source of structural information that can be integrated into useful representations of macromolecular assemblies (Lasker et al., 2010) as elegantly shown for the yeast nuclear pore complex whose structure was determined at about 5 nm resolution (Alber et al., 2007). The key element in integrative structure determination is that the concurrent satisfaction of all restraints derived from independent experiments drastically reduces the combinatorial possibilities of structural solutions (Schwede et al., 2009). Such a synergy is observed in crystallography toward using low-resolution data (Schröder et al., 2010), in combining NMR, X-Ray, and computation (Szymczyna et al., 2009), or in combining solid-state NMR and computation (Masica et al., 2010).

Atomic force microscopy (AFM) is a recently developed biophysical technique that has provided remarkable successes in imaging single molecules. AFM yields 2D images with an exceptional signal/noise (S/N) ratio allowing the observation of single molecules at a lateral resolution of nanometers and a vertical resolution near the Angstrom (Schabert et al., 1995). For the highest resolution AFM topographs, it has been shown that the best single-molecule images are more accurate molecular representations than ensemble averages (Fechner et al., 2009). AFM works with a very sharp tip mounted on a microcantilever that can passively sense the localized forces between the

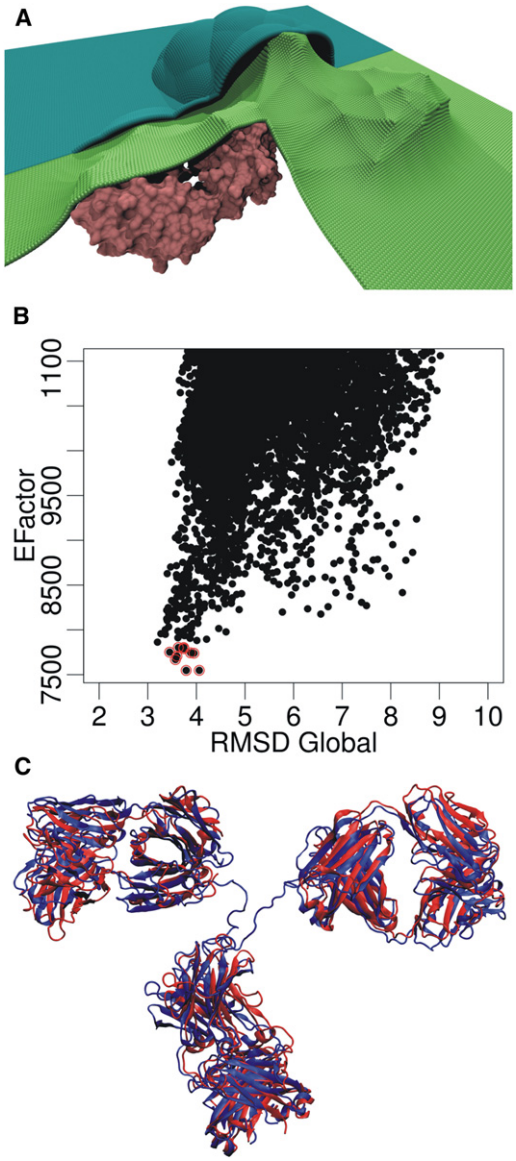

Figure 1. Reconstruction of an Antibody Model Using Simulated Data

(A) Molecular and topographic surface of a complete immunoglobulin structure. The atomic molecular surface of 1IGT is shown in magenta. The simulated topographic image of 1IGT including the tip-shape distortion is shown in blue. The eroded topographic surface that was used for assembling a complete immunoglobulin is shown in green. It shows that the simulated topographic surface used for the docking is quite different from that of the 1IGT molecule. (B) Correlation plot between the calculated $S_{EFactor}$ from the 1IGT assembly and the all-atom rmsd. The top 10 best $S_{EFactor}$ scores are shown as red dots. (C) Superposition of the target structure (1IGT, in blue) and the best assembled model (in red) built using the simulated topographic AFM image of (A) with two Fab domain structures (1AY1) and one Fc domain structure (1H3T). The all-atom rmsd is 3.79 Å.

Images are drawn using VMD (Humphrey et al., 1996).

atoms of the scanning tip and the specimen surface (Binnig et al., 1986). An AFM scanner consists of piezoelectric ceramic elements with a subnanometer resolution in displacements. A topographical image of a deposited specimen is obtained using a raster scanning of the tip in the x-y plane and monitoring the vertical position (z) of the tip.

Macromolecular reconstruction using experimental constraints is mostly performed either by docking atomic coordinates of protein structures into TEM densities (Kawabata, 2008; Lasker et al., 2010; Zheng, 2011), by mapping subunits in multiprotein complexes (Flemming et al., 2010), by combining SAXS intensity profiles with molecular simulation (Różycki et al., 2011) or FRET data (Rochel et al., 2011), or from information-driven docking (Karaca et al., 2010). Atomic force microscopy, due to its very high S/N ratio, is a promising alternative technique. Establishing a relationship between atomic coordinates and AFM topographic images has already been illustrated (Asakawa et al., 2011; Buzhynskyy et al., 2009; Czajkowsky et al., 2004; Czajkowsky and Shao, 2009; Davies et al., 2005; Philippsen et al., 2002; Scheuring et al., 2007; Scheuring et al., 2005). In this work we use, for the first time, AFM topography as low-resolution envelopes into which we place atomic coordinates using the rigid body protein docking program DOT (Mandell et al., 2001; Ten Eyck et al., 1995). This method provides a systematic search of translation and rotation. The flexibility of DOT allowed straightforward development and implementation of a potential representing the AFM image. We describe the global protocol and show a proof of concept using antibody molecules as a model system. Then, we apply this protocol to a large virus particle (Tobacco Mosaic Virus) using experimental AFM images, demonstrating the stability of the protocol in the presence of experimental noise. Finally, we model the tetrameric complex of a membrane protein (Aquaporin Z) using a single monomer and experimentally determined high-resolution AFM topography (Scheuring et al., 1999).

RESULTS

Molecular Assembly Using a Simulated AFM Topographic Image

To validate the computational assembly protocol, we performed a simulation experiment. First, we created a simulated molecular topographic surface of a multidomain protein whose 3D structure was known. Antibodies appeared as a good template since they are made of three large domains (two Fabs and one Fc), easily accessible, and several 3D structures of complete molecules have been determined. A topographic surface was generated as explained in Experimental Procedures (Figure 1A). To present a more realistic setup, a dilated-eroded processed surface of an antibody was used and unrelated Fab and Fc domains were selected for the reconstruction.

A total of 100,000 assemblies of two Fabs and one Fc domains were generated and their rmsds were computed. The all-atom rmsd of the best assembled structure (best $S_{EFactor}$) was 3.79 Å (Figures 1B and 1C) and the top ten assembled structures had an all-atom rmsd less than 5 Å (Figure 1B). Among all the assembled structures, 108 had an all-atom rmsd lower than 3.79 Å but with a higher $S_{EFactor}$ score; the smallest all-atom rmsd (3.2 Å) was ranked 11 according to the $S_{EFactor}$. When we analyzed the relationship between rmsds and errors in translation and/or rotation, we found that a rmsd of 4 Å between the

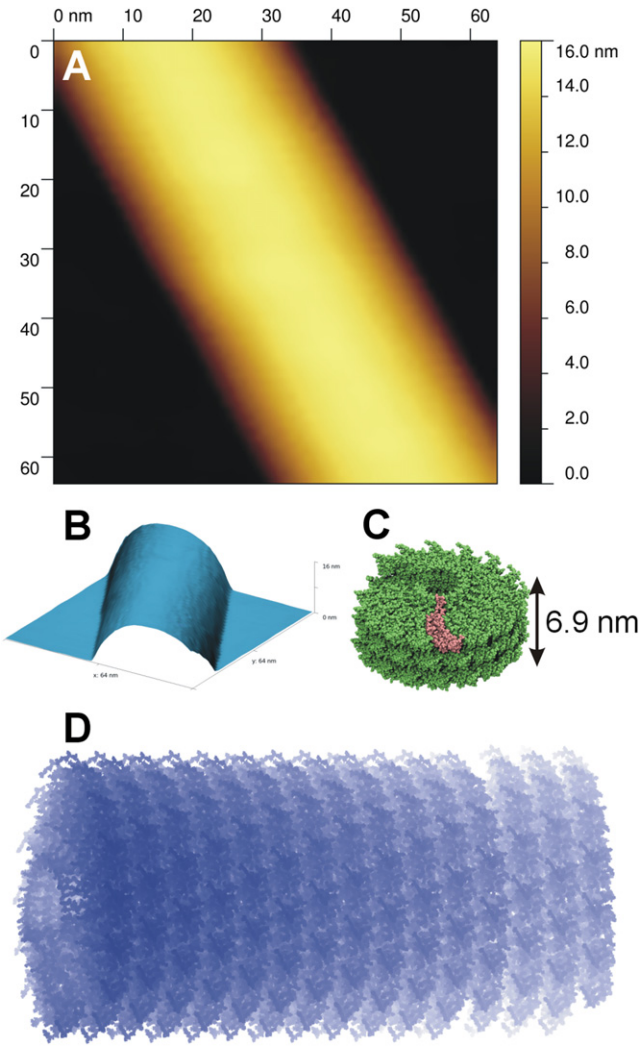

Figure 2. Reconstruction of a TMV-Pentadisk Assembly Using Experimental AFM Data

(A) AFM topographic images of a TMV particle adsorbed on mica using tapping mode in air. The image was interpolated using BSpline algorithm from Gwyddion to reach the resolution of 2.5 Å/px.

(B) 3D representation of the docking-ready AFM topographic image used for assembling multiple disks of TMV.

(C) Molecular surface (in green) of a single TMV disk represented 49 subunits (in magenta: a single subunit). The thickness of the disk will be used as a distance constraint during the final assembly stage.

(D) The best assembly of five TMV disks using the experimental topographic surface in (A).

reconstructed and native antibodies could be due to a rotation error of 10° with no error in translation or a rotation error of 7° with a translation error of 2 Å. It is important to notice that the grid step of docking was 2 Å and the minimal rotation variation was about 6°.

Molecular Assembly Using Experimental AFM Topographic Image of TMV

The crystallographic structure of TMV is determined by fiber diffraction data (Pattanayek and Stubbs, 1992). TMV is made of a unique coat protein which is replicated helicoidally 2,130 times around its RNA, forming a cylinder 300 nm in length and 18 nm diameter. The AFM-assembly protocol was challenged against real experimental topographic AFM images of TMV (Figures 2A and 2B) deposited on mica and imaged using the tapping mode in air. A segment of TMV cylindrical structure of length 40 nm was built using five disk-like fragments (see Experimental Procedures, Figure 2C). The goal of this test is not to show that our protocol can build helicoidal assembly of a known virus but to test the robustness of the reconstruction algorithm with real topographic images and five molecular constituents.

Table 1. Top-10 Assemblies of TMV Pentadisks

Rank	Average[a] Distances between Disk to Disk (Å)					Cα Rmsd[b] (in Å)
	Disks 1-2	Disks 2-3	Disks 3-4	Disks 4-5	Average	
1	76.7	70.0	70.0	70.0	71.7	6.5
2	76.7	70.0	70.0	70.0	71.7	6.1
3	76.7	70.0	70.0	70.0	71.7	6.1
4	77.4	70.0	70.0	70.0	71.9	8.0
5	77.8	70.0	70.0	70.0	72.0	5.1
6	79.2	70.0	70.0	70.0	72.3	6.9
7	76.7	70.0	72.5	70.0	72.3	6.7
8	76.7	72.5	70.0	70.0	72.3	6.9
9	76.7	70.0	70.0	72.5	72.3	6.5
10	76.7	70.0	70.0	72.5	72.3	6.5

[a] The distance between two disks are computed by averaging the top, the middle, and the bottom distances.

[b] Rmsd between 38,710 Cα atoms from 245 subunits of modeled and X-ray pentadisks after global superposition.

Before the reconstruction, AFM topographic surfaces were eroded as previously described (Trinh et al., 2011). The rmsds for the top 10 assembly of pentadisks were from 5 to 8 Å using 38,713 Cα (245 TMV coat protein subunits). Ranking of the top 10 was obtained by averaging distances between each of the five disks (Table 1). The $S_{EFactor}$ could not be computed in this case because the assembly of docking disks did not cover the entire experimental surface. The average distance for the top 10 assembly was about 7.1 nm compared to the expected 6.9 nm (Figure 2C). This can be explained by the fact that we used a discrete docking system (grid step of 2.5 Å and a rotation step of 6°), which can induce translation errors and misalignments between two successive disks. Because we used a rigid-body assembly approach and we restricted the presence of steric clashes, some assembled structures were eliminated due to small steric overlaps. We evaluated that for an 18 nm diameter disk, with a rotation around its geometric center, a rotation of about half of a rotation step used during the docking process (3°) will cause a translation of about 5 Å at the top/bottom of the disk. It follows that our assembly protocol is well adapted for large system since the average distance error in our models was less than 1 nm for a structure assembly of 40 nm long (Figure 2D). Larger assemblies of TMV particles were attempted (up to seven disks) by selecting docking conformations in which the disk axis was collinear with the axis of the topographic image (data not shown).

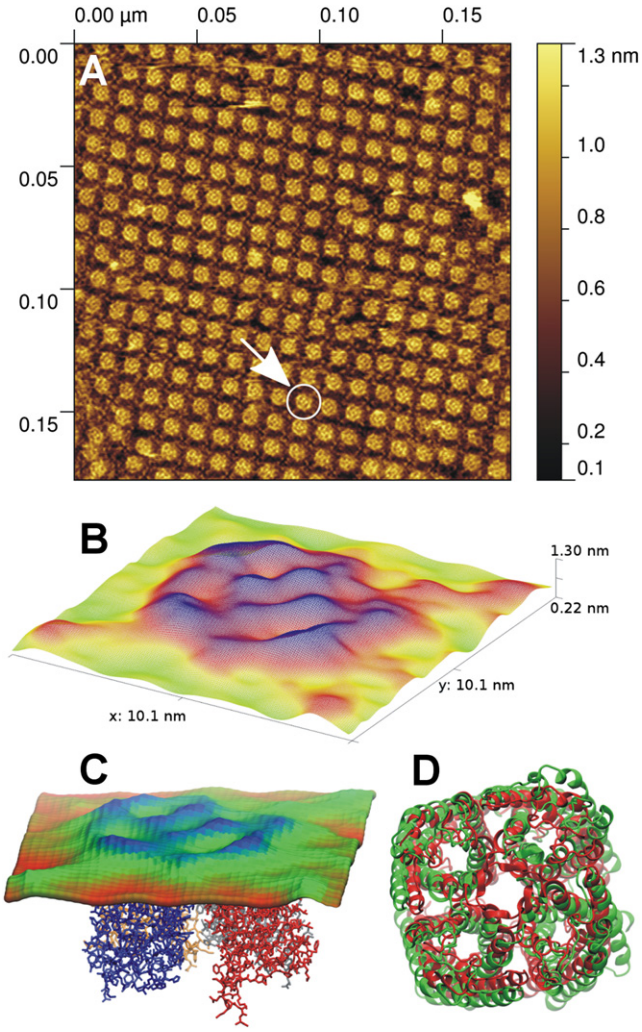

Figure 3. Reconstruction of a Tetrameric Membrane Protein Aquaporin Z Using Experimental AFM Data

(A) AFM topographic image of a 2D crystal of aquaporin Z (AqpZ) adsorbed on mica using contact mode in liquid (Scheuring et al., 1999) at a resolution of 3.5 Å/px. A portion containing one AqpZ (marked by the white circle) was extracted and interpolated by a BSpline to the resolution of 0.5 Å/px.

(B) 3D representation of the docking-ready AFM topographic image used for the assembly of AqpZ tetramer.

(C) Best assembled model of AqpZ built from a single monomer. Each monomer is colored differently.

(D) Superposition of the best assembled AqpZ tetramer (in green) on the target crystallographic structure (1RC2 in red). Only the secondary structures are displayed. The all-atom rmsd between the two structures is 4.46 Å.

Assembly of Tetrameric Aquaporin Z Using a Single Monomer and High-Resolution AFM Topographs

Before the reconstruction, AFM topographic surfaces were eroded as described in Experimental Procedures. A tetrameric reconstruction of membrane protein Aquaporin Z was assembled using experimentally determined high-resolution AFM topography (Figures 3A and 3B) and a single crystallographic monomeric structure (see methods). The top 10 tetrameric assemblies were ordered according to their $S_{EFactor}$ in Table 2.

Table 2. Top-10 Assemblies of AqpZ Tetramer

	Monomer Center to Center Distance (Å)					All-Atoms
Rank	1-2	2-3	3-4	4-1	Average	Rmsd (Å)
1	29.9	26.9	27.8	28.0	28.2	4.5
2	28.4	29.9	27.7	26.9	28.2	4.5
3	27.8	28.0	29.5	27.5	28.2	4.5
4	27.7	28.4	27.5	29.5	28.3	4.5
5	27.9	29.6	27.7	27.5	28.2	4.4
6	27.6	27.9	26.9	29.9	28.1	4.3
7	27.6	27.9	27.5	29.5	28.1	4.3
8	28.4	29.9	27.7	26.9	28.2	4.4
9	27.5	27.7	29.5	28.4	28.3	4.4
10	28.0	29.1	28.0	27.8	28.3	4.5

The average distance separating the centers of each monomer was about 28 Å compared to 24 Å in the crystallographic structure (the distance constraint was a range from 20 to 30 Å). The all-atoms rmsds between assembled models and the crystallographic tetramer was less than 5 Å. (Table 2). By optimizing the best model (Figure 3C) with a rigid-body minimization process in X-PLOR using a single NOE restraint between the Cα of Phe62 of each monomer (30 ± 0.5 Å), the final all-atom rmsd reached 3.5 Å (Figure 3D). Errors highlighted by the rmsd come from the discrete grid-based docking similarly to what was found with antibodies and TMV. It should be noted that the ranking for the tetrameric assembly of Aquaporin Z (AqpZ) was entirely based on the $S_{EFactor}$ which is the resemblance of topographic surfaces between the assembly and AFM image. Similar rmsds were obtained when using another crystallographic structure of AqpZ monomer (2ABM, data not shown).

DISCUSSION

The low abundance of 3D structures of large proteins combined with the saturation in the coverage of single protein domain structure creates a gap between low-resolution images of single particles and the atomic structures of individual domains. In this work, we present a novel methodology that combines high-resolution AFM topographic images with atomic coordinates of proteins to assemble very large macromolecules or particles. Applications of AFM in structural biology covers large and flexible proteins, which are difficult to study by traditional techniques (Schröder et al., 2010), as well as proteins that interact with solid/flat surfaces such as those adsorbed on biominerals or interacting with membranes. AFM also complements transmission electron microscopy (Conroy et al., 2004; Lin and Goh, 2002; Ubbink and Schär-Zammaretti, 2005); however, AFM does not rely on symmetry averaging as in EM or on physical averaging as in crystallized sample. This strategy is well suited for modeling flexible macromolecules by treating them as a set of domains that can be fitted independently as rigid bodies. Our modeling protocol is performed in real space rather than in Fourier space and presents the advantage of making it convenient to introduce any additional constraints. Our approach will have its limitations but the ability to include sparse long-range distance constraints will enhance the quality of assembly.

Limits and Expected Resolution

The term "resolution" is used to cover multiple aspects of our protocol. First, the docking resolution is related to the computing grid spacing. Currently, it is possible to use a cubic grid of 256 elements; thus, a 25.6 nm long target can be mapped on a grid at 1 Å per grid point. Second, the structural resolution is more difficult to evaluate for isolated single particles. For 2D crystals, a lateral resolution of 1 nm or less has been measured (Fechner et al., 2009). By observing distinguishable details on the topographic image of the TMV, we could estimate the topographic resolution to 3–5 nm. It is likely that AFM image processing could enhance this resolution (Chen and Pellequer, 2011). Finally, although it is not strictly a true "resolution," it is necessary to consider the topographic image sampling that is expressed in number of angstroms per pixel. With current AFM controllers, it is possible to obtain images of 1,024 pixels or more on a field of 256 nm providing a sampling of 2.5 Å per pixel; scanning in a smaller field or increasing the numbers of pixels will increase this sampling. But more important is the distortion of the molecule being imaged by the tip. There are two types of distortions: one due to the finite size of the tip and can be treated using mathematical morphology tools (Villarrubia, 1997), and one due to the pressure of the tip onto the surface of the protein. For the latter, it has been shown that by imaging with gentle care the tip crushed a little more that surface-exposed side chains (Trinh et al., 2011). It might have been expected that the docking precision is dependent on the detailed coordinates of selected domains. Indeed, there is no reason to think that proteins deposited on a flat surface will possess the exact same conformation as found in X-ray crystallography data. It turns out that at the current structural resolution of AFM images (3–5 nm) used in this work, there is little impact on docking precision.

AFM and AFM Assembly in Structural Biology

Despites limitations in structural resolution, AFM has multiple advantages. Among the most important ones is the amount of material required. Because AFM works at the single molecule level, a concentration of <1 μg/ml is enough for imaging a population of molecules compared to <1 mg/ml for X-ray diffraction or NMR. Besides, AFM does not require ultrapure protein sample as long as contaminants have difference shapes and sizes that make them distinguishable. Another advantage is the capability of imaging single molecules with AFM in buffer solution compared to the frozen and crystallized state in TEM and X-ray diffraction, respectively.

It is obvious that a hard and highly charged mineral substrate (such as mica) creates difficulties during the adsorption of molecules on which they tend to partially denature (Umemura et al., 1996). However, softer systems for depositing molecules such as functionalized self-assembled monolayers are emerging (Lv et al., 2010). Moreover, these systems may allow the orientation of covalently attachment molecules.

The integrative assembly protocol presented in this work takes advantage of AFM strengths. The key parameter is the experimental data provides by the topographical surface of the whole-length protein. Because AFM images are obtained in the real space, it is possible to include in the protocol several types of restraints such as excluded volume (steric clashes), distances (a min-max range from any set of atoms), symmetry (not necessary but easy to implement as distance constraints). The ranking of macromolecular assemblies is performed using a quality-of-fit parameter ($S_{EFactor}$) that determines the agreement between the experimental AFM topographic surface and the surface of the assembled macromolecule.

Expected Range of Improvements

As a nascent technique, our approach can benefit from several types of improvements. Regarding the structural resolution in AFM, technological developments are required to image protein samples with the smallest possible forces (tens of piconewton) compared to the hundreds of piconewton currently. Even in tapping mode, deformation of imaged objects is unavoidable. To reduce this effect on the reconstruction protocol, it is possible to allow bumps between constituent fragments and the docking forbidden zone. Currently, no overlapping between atomic coordinates and the forbidden zone is allowed. Furthermore, it is possible to "shave" molecular constituent structures to remove for instance side chains that could bias the docking stage; preliminary results are encouraging. New imaging bimodal mode is among these recent technical developments (Dietz et al., 2011). Improvements are also expected regarding the AFM tip, which is usually a sharp or ultrasharp silicon tip. Single-wall carbon nanotubes (SWCNT) offer great promises toward the improvement of imaging quality and not necessarily for reducing the tip-shape distortion. Diameter of SWCNT is very small (1–4 nm) and more important it is constant through time, which is not the case of silicon-based tips; besides SWCNT offers little contamination during imaging. The current challenge with SWCNT is the production of correctly oriented SWCNT on AFM cantilever (Marty et al., 2006). Another road for improvement is the capacity to correct image distortion due to the finite tip size. Although mathematical tools are available to correct for tip-dilated AFM images, the weak point remains the characterization of the AFM-tip shape. Although blind estimate algorithms have been developed (Villarrubia, 1997) and successfully applied in cryo-AFM (Sheng et al., 1999), this approach is very sensitive to noise levels (Trinh et al., 2011) and could not be applied here. Most straightforward techniques use hard self-imaging sample but since they are highly destructive of AFM tips they can only be used a posteriori; thus it is not possible to know the shape of the tip during imaging. Alternative options included the addition of a known biological sample on the substrate which can be used as a calibrator (Trinh et al., 2011) but it remains to be demonstrated on mixed deposited samples.

Regarding the computational part of the assembly protocol, it may benefit from a reduction in the search space (such as with the membrane protein AqpZ) in order to reduce unproductive orientations during docking. As common to all integrative methods, the availability of additional knowledge on the protein to study should be taken advantage. Scoring functions recently attract a great deal of attention due to the relative ease in producing multiple docking solutions for a single assembly. The conformational space being well covered implies that the near-native solution exists in the pool of answers; thus, the key is to identify that best solution, a non-trivial problem (Feliu and Oliva, 2010). Regarding the present protocol, future efforts will turn toward the development of a scoring function so that currently low ranking solution become reachable for assembly

testing. Finally, improvement will be pursued regarding the efficient combinatorial optimizers (Lasker et al., 2009) to reduce CPU time when assembling multiple fragments.

Conclusions

Current developments in structural biology make it timely to develop approaches that can contribute to obtaining large macromolecular structures on the basis of the structural information that is available for individual components and to couple this approach with available low-resolution structural information such as those from AFM or electron microscopy.

We now understand that many proteins in the cell do not function in isolated manner and are frequently found in large macromolecular complexes. The assembling protocol presented here is completely adapted to the assembly of intermolecular complexes since the only difference between the reconstruction of a multidomain protein and a multiprotein complex is the presence of the polypeptide junction in the former. The greatest promise in this approach is the possibility to combine additional various structural data such as those from EM and SAXS experiments.

EXPERIMENTAL PROCEDURES

Modeling with AFM constraints

The AFM assembly concept is a two-stage procedure: first, the molecular constituents of the target system are docked beneath the experimentally obtained topographic surface; second, all the docked molecular constituents are assembled into a final structure. A detailed description of the algorithm will be presented elsewhere (M.-H.T., M.O., J.-M.T., P.P., S.-w.W.C., and J.-L.P., unpublished data).

In the docking step an exhaustive search is carried out by FFT-based rigid docking software DOT, testing all translations and rotations for optimum docking solution. The search space below the AFM topographic surface is mapped to a cubic grid of 256^3 nodes. The grid step is adjusted so that the search space covers the entire AFM topographic surface. Each molecular constituents of the target protein is translated to every node of the grid. A rotation step of 6°, representing a total of 54,000 different rotations, is applied on each molecular constituent. The AFM topographic image is partitioned into forbidden, favorable, and neutral zones. The forbidden zone is all points above the AFM surface. The favorable zone is the layer immediately beneath the AFM surface (8–15 Å thick). The remaining points are assigned a neutral value such that they do not contribute to the docking score. The moving object is constructed from the atomic coordinates extracted from PDB file, keeping only a thin skin (~3 Å) of surface atoms (Chen et al., 2009). Each placement of the moving molecule that does not overlap the forbidden zone is scored by counting the number of atoms inside the favorable zone.

After the docking stage, the final assembly of all the molecular constituents is performed using custom software called *combine*. Target protein models are built by exhaustively combining docked constituents. As the constituents are docked independently, collisions between them may occur. The program *combine* eliminates models with excessive steric clashes. However, distance constraints between constituents are used to prevent the unnecessary calculation of steric clashes since it is a CPU-intensive stage of the protocol. We developed a quality-of-fit parameter called the $S_{EFactor}$ score to estimate the agreement between the surface of the assembled multidomain protein and that of the experimental AFM topographic image. It is defined as:

$$S_{EFactor} = w_E \times EFactor$$

$$EFactor = \sum \left| P_{exp}(i,j) - P_{mod}(i,j) \right|$$

$$\text{With } (i,j) \in M \mid M(i,j) = 1$$

where w_E is a user-defined weight set to 1.0, $P_{exp}(i,j)$ and $P_{mod}(i,j)$ are the height of pixel (i,j) on the eroded experimental image and that of the eroded theoretical topographic image built from the assembled model, respectively, and $M(i,j)$ is a binary mask (values of 0 or 1) identifying only pixels of interest. The total number of retained docking solution varies in function of the number of domains to assemble. For example, it takes about 13 hr on 24 CPUs for assembling three domains using 2,000 docking solutions of each domain. The most CPU-consuming task in the reconstruction is the computation of EFactor. Timing is related to the size of the modeled tip used during erosion/dilation of images. As the number of retained docking solutions is limited, we applied a non-similarity filter on the docking solution pool. In order to refine the final molecular assembly, each fragment that composes best models is used as seeds to select a new pool of docking solutions around each fragment. Those newly selected docking solutions are used in turn to run a second assembly with the same *combine* protocol.

Simulating AFM Topographic Images of Antibodies

Because the AFM tip apex has a finite size, the topographic surface obtained during the imaging process is not equal to the real surface of the object imaged, but is "distorted" due to the tip shape (Villarrubia, 1997). Because our protocol works in the real space, it is always necessary to process raw AFM height images to reduce the tip-shape distortion in the topographic surface using mathematical morphology tools (Villarrubia, 1997). AFM topographic images of antibodies were simulated by taking into account this tip-effect artifact. An AFM topographic image (128 × 128 pixels for a field of 25.6 × 25.6 nm) was obtained by dilating the atomic surface of the structure 1IGT (Harris et al., 1997) with a model tip with an apex radius of 5 nm and side wall angles of 28° (average values from commercially available AFM tips). As performed for experimental AFM topographic images, simulated surfaces were eroded using the same tip dimension. It is important to note that erosion/dilation is not reversible (Trinh et al., 2011), and thus the eroded surface is slightly different from the original atomic surface of 1IGT (Figure 1A). To present a more realistic case for assembling the domains of the antibody 1IGT, atomic coordinates of 1IGT were not used but instead those from 1AY1 (Murali et al., 1998) and 1H3T (Krapp et al., 2003) for the Fab and Fc domains, respectively. Missing coordinates in the 1H3T Fc domain were modeled using symmetry-related coordinates (Gale et al., 2007).

Assembly of Complete Antibody with Simulated Topographic Images

The structure 1AY1 and 1H3T were docked independently, under the simulated AFM topographic image with a favorable layer of 12 Å, a grid step of 2 Å, and a rotation step of 6°. For the assembly, we used a single distance-constraint range of 35–42 Å between Fab and Fc domains. This range was obtained by measuring the average distances between Cα of Pro 227 on a Fab and Ser 252 on the Fc domain using three available complete crystallographic structures of antibodies (PDB codes 1IGT, 1IGY, and 1HZH).

The assembled Fab-Fc antibody models were compared to the complete structure of 1IGT using all-atom rmsds. However, due to a slight difference in the number of residues in Fabs and Fc of 1IGT and that in the structures 1AY1 and 1H3T, a rmsd cannot be straightforwardly computed. To work around this problem, a reference structure was built by superimposing two Fab domains 1AY1 and one Fc domain 1H3T onto the structure 1IGT using sup3d (Chen and Pellequer, 2004). This reference structure is then used to compute rmsd with computational assembled models.

AFM Height Images of Tobacco Mosaic Virus

A drop of 70 μl of $NiCl_2$ 20 mM was deposited on a freshly cleaved mica disk (muscovite, Veeco, Santa Barbara, CA) and 10 μl of TMV (38 nM in 1 mM EDTA) was injected in the drop. After 10 min, the sample was washed several times using MilliQ water to eliminate salts and nonadsorbed particles. The sample was dried using a Laboport vacuum pump (KNF Neuberger, Trenton, NJ). The sample was conserved at room temperature and imaged using a commercial Dimension D3100 (Veeco) equipped with a Hybrid XYZ scanner head (closed loop). Silicon nitride tip (Multi40, Veeco) with a spring constant k = 0.9 N/m and nominal apex radius 8 nm were used. The height image was made using the tapping mode in air at a scan rate of 1 Hz; the tip was oscillating around 40 kHz with free amplitude (A_0) of about 30 nm. The amplitude set point was set 2 nm less than the free amplitude, thus the ratio A/A0 was less than

10%. The image resolution was 1 nm per pixel. The image was flattened and analyzed using Gwyddion software (http://gwyddion.net).

Assembly of Tobacco Mosaic Virus

The second reconstruction test fits five disk-like fragments of TMV into the experimental topographic image of a single TMV particle about 40 nm long. Each disk-like was made of 49 coat proteins extracted from 1VTM (Pattanayek and Stubbs, 1992), a total of 63,651 atoms (Figure 2C). Topographic AFM images were eroded using an estimated tip radius of 6.5 nm and sidewall angle of 15° (Trinh et al., 2011). For the docking stage, a cubic grid of 256 points was used with a favorable layer 15 Å thick beneath the AFM topographic surface, a grid step of 2.5 Å, and a rotation step of 6°. A set of three distance constraints of 69 ± 20 Å (top, middle, and bottom of disks) were defined for the final assembly. A maximum number of 20 steric clashes are allowed in the reconstruction; a steric clash is defined when the distance between atoms from two different fragments is less than 2 Å.

A total of 100,000 docking solutions of TMV disk-like fragments beneath the experimental AFM topographic surface (Figure 2B) were retained. From these, the top 2,000 nonsimilar docking solutions, 5 Å translation and 13° rotation, were selected for the reconstruction of a TMV pentadisk. A set of three distance constraints of 69 ± 20 Å (top, middle, and bottom of disks) were used for the assembly. Then, fragments from the top 500 reconstructed pentadisks were used to search initial docking solutions present in the top 100,000 list by selecting those within 5 Å translation and 13° rotation. The top 3,500 near-docking solutions were retained to reconstruct a refined pentadisk.

Docking and Assembly of AqpZ

Using AFM in liquid environment, Scheuring et al. have successfully imaged an artificial membrane containing a 2D crystal of several aquaporins (Scheuring et al., 1999). We extracted the topographic surface of a single AqpZ molecule. The selected surface was interpolated with a b-spline algorithm to a final resolution of 0.5 Å per pixel using Gwyddion. The extracellular face of AqpZ is shown in Figure 3A. To reduce the tip-shape distortion, this image was eroded with a theoretical tip of 7 Å radius and 30° side wall. These dimensions were estimated based on the depth of the central channel of the AqpZ topography (Figure 3B).

The monomeric structure of AqpZ was extracted from the chain A of the PDB code 1RC2 (Savage et al., 2003). This structure contains 1,661 atoms for 231 residues. The starting pose of AqpZ was that of the chain A, but because the reconstruction was performed only with chain A docking solutions, the effect of the starting pose is negligible regarding the final assembly. For the docking we used a cubic grid of 256 points, a favorable layer of 8 Å, a grid step of 0.5 Å and a rotation step of 4°. Knowing that the AqpZ is a membrane protein, we limited among all the 54,000 possible orientations only those allowing a docking of AqpZ where its major elongated axis was oriented 90 ± 20° relative to the membrane plane. The docking of the AqpZ monomer under the experimental AFM topographs was performed using a set of 13,860 rotations on 24 Xeon CPUs in 42 min. Among the top 100,000 docking solutions, the top 5,000 was selected such that similar docking solutions of monomers within 1 Å in translation and 8° in rotation were removed and only the best ranked conformations were kept. To assemble the tetrameric conformation of AqpZ, we used a single distance constraint range from 20 to 30 Å between the centers of each monomer pair. The actual center-to-center distance in the crystallographic tetrameric structure is 24 Å. During the assembly, models were ranked solely based on the $S_{EFactor}$, the fit between the assembled-model topographic surface and the experimental topographic surface. The first run of *combine* generated 2,609 assemblies (19 min on 24 CPUs). A second run of *combine* was performed using near-docking solutions pool within 1 Å in translation and 8° in rotation generated from all the assemblies of the first run.

ACKNOWLEDGMENTS

This work was supported by the French ANR agency (grant ANR-07-PCVI-0002-01 to J.-L.P.) and by the French Atomic and Alternative Energies Commission, by National Science Foundation grant DBI 99-04559 (V.A.R. and L.F.T.), and by National Institutes of Health grant GM070996 (V.A.R., L.F.T., and M.P.). Authors are grateful to Dr. Simon Scheuring for providing the AFM data regarding the Aquaporin Z. The reconstruction protocol uses a set of scripts and programs running under Linux that are freely available upon request; however, no package can be prepared at the moment as it includes third-party software.

Received: August 5, 2011
Revised: October 5, 2011
Accepted: October 10, 2011
Published: January 10, 2012

REFERENCES

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007). Determining the architectures of macromolecular assemblies. Nature *450*, 683–694.

Alber, F., Förster, F., Korkin, D., Topf, M., and Sali, A. (2008). Integrating diverse data for structure determination of macromolecular assemblies. Annu. Rev. Biochem. *77*, 443–477.

Asakawa, H., Ikegami, K., Setou, M., Watanabe, N., Tsukada, M., and Fukuma, T. (2011). Submolecular-scale imaging of α-helices and C-terminal domains of tubulins by frequency modulation atomic force microscopy in liquid. Biophys. J. *101*, 1270–1276.

Binnig, G., Quate, C.F., and Gerber, C. (1986). Atomic force microscope. Phys. Rev. Lett. *56*, 930–933.

Buzhynskyy, N., Golczak, M., Lai-Kee-Him, J., Lambert, O., Tessier, B., Gounou, C., Bérat, R., Simon, A., Granier, T., Chevalier, J.M., et al. (2009). Annexin-A6 presents two modes of association with phospholipid membranes. A combined QCM-D, AFM and cryo-TEM study. J. Struct. Biol. *168*, 107–116.

Chen, S.-W., and Pellequer, J.L. (2004). Identification of functionally important residues in proteins using comparative models. Curr. Med. Chem. *11*, 595–605.

Chen, S.-W., and Pellequer, J.L. (2011). DeStripe: frequency-based algorithm for removing stripe noises from AFM images. BMC Struct. Biol. *11*, 7.

Chen, S.-W., Van Regenmortel, M.H.V., and Pellequer, J.-L. (2009). Structure-activity relationships in peptide-antibody complexes: implications for epitope prediction and development of synthetic peptide vaccines. Curr. Med. Chem. *16*, 953–964.

Conroy, M.J., Jamieson, S.J., Blakey, D., Kaufmann, T., Engel, A., Fotiadis, D., Merrick, M., and Bullough, P.A. (2004). Electron and atomic force microscopy of the trimeric ammonium transporter AmtB. EMBO Rep. *5*, 1153–1158.

Czajkowsky, D.M., and Shao, Z. (2009). The human IgM pentamer is a mushroom-shaped molecule with a flexural bias. Proc. Natl. Acad. Sci. USA *106*, 14960–14965.

Czajkowsky, D.M., Hotze, E.M., Shao, Z., and Tweten, R.K. (2004). Vertical collapse of a cytolysin prepore moves its transmembrane beta-hairpins to the membrane. EMBO J. *23*, 3206–3215.

Davies, E., Teng, K.S., Conlan, R.S., and Wilks, S.P. (2005). Ultra-high resolution imaging of DNA and nucleosomes using non-contact atomic force microscopy. FEBS Lett. *579*, 1702–1706.

Dessailly, B.H., Nair, R., Jaroszewski, L., Fajardo, J.E., Kouranov, A., Lee, D., Fiser, A., Godzik, A., Rost, B., and Orengo, C. (2009). PSI-2: structural genomics to cover protein domain family space. Structure *17*, 869–881.

Dietz, C., Herruzo, E.T., Lozano, J.R., and Garcia, R. (2011). Nanomechanical coupling enables detection and imaging of 5 nm superparamagnetic particles in liquid. Nanotechnology *22*, 125708.

Fechner, P., Boudier, T., Mangenot, S., Jaroslawski, S., Sturgis, J.N., and Scheuring, S. (2009). Structural information, resolution, and noise in high-resolution atomic force microscopy topographs. Biophys. J. *96*, 3822–3831.

Feliu, E., and Oliva, B. (2010). How different from random are docking predictions when ranked by scoring functions? Proteins *78*, 3376–3385.

Flemming, D., Thierbach, K., Stelter, P., Böttcher, B., and Hurt, E. (2010). Precise mapping of subunits in multiprotein complexes by a versatile electron microscopy label. Nat. Struct. Mol. Biol. *17*, 775–778.

Gale, A.J., Yegneswaran, S., Xu, X., Pellequer, J.-L., and Griffin, J.H. (2007). Characterization of a factor Xa binding site on factor Va near the Arg-506 activated protein C cleavage site. J. Biol. Chem. *282*, 21848–21855.

Geer, L.Y., Domrachev, M., Lipman, D.J., and Bryant, S.H. (2002). CDART: protein homology by domain architecture. Genome Res. *12*, 1619–1623.

Harris, L.J., Larson, S.B., Hasel, K.W., and McPherson, A. (1997). Refined structure of an intact IgG2a monoclonal antibody. Biochemistry *36*, 1581–1597.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27–28.

Karaca, E., Melquiond, A.S., de Vries, S.J., Kastritis, P.L., and Bonvin, A.M. (2010). Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. Mol. Cell. Proteomics *9*, 1784–1794.

Kawabata, T. (2008). Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. Biophys. J. *95*, 4643–4658.

Krapp, S., Mimura, Y., Jefferis, R., Huber, R., and Sondermann, P. (2003). Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. J. Mol. Biol. *325*, 979–989.

Lasker, K., Topf, M., Sali, A., and Wolfson, H.J. (2009). Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. J. Mol. Biol. *388*, 180–194.

Lasker, K., Phillips, J.L., Russel, D., Velázquez-Muriel, J., Schneidman-Duhovny, D., Tjioe, E., Webb, B., Schlessinger, A., and Sali, A. (2010). Integrative structure modeling of macromolecular assemblies from proteomics data. Mol. Cell. Proteomics *9*, 1689–1702.

Lin, A.C., and Goh, M.C. (2002). A novel sample holder allowing atomic force microscopy on transmission electron microscopy specimen grids: repetitive, direct correlation between AFM and TEM images. J. Microsc. *205*, 205–208.

Lv, Z., Wang, J., Chen, G., and Deng, L. (2010). Imaging recognition events between human IgG and rat anti-human IgG by atomic force microscopy. Int. J. Biol. Macromol. *47*, 661–667.

Mandell, J.G., Roberts, V.A., Pique, M.E., Kotlovyi, V., Mitchell, J.C., Nelson, E., Tsigelny, I., and Ten Eyck, L.F. (2001). Protein docking using continuum electrostatics and geometric fit. Protein Eng. *14*, 105–113.

Marty, L., Iaia, A., Faucher, M., Bouchiat, V., Naud, C., Chaumont, M., Fournier, T., and Bonnot, A.M. (2006). Self-assembled single wall carbon nanotube field effect transistors and AFM tips prepared by hot filament assisted CVD. Thin Solid Films *501*, 299–302.

Masica, D.L., Ash, J.T., Ndao, M., Drobny, G.P., and Gray, J.J. (2010). Toward a structure determination method for biomineral-associated protein using combined solid- state NMR and computational structure prediction. Structure *18*, 1678–1687.

Mueller, M., Jenni, S., and Ban, N. (2007). Strategies for crystallization and structure determination of very large macromolecular assemblies. Curr. Opin. Struct. Biol. *17*, 572–579.

Murali, R., Helmer-Citterich, M., Sharkey, D.J., Scalice, E.R., Daiss, J.L., Sullivan, M.A., and Krishna Murthy, H.M. (1998). Structural studies on an inhibitory antibody against Thermus aquaticus DNA polymerase suggest mode of inhibition. Protein Eng. *11*, 79–86.

Murata, K., Liu, X., Danev, R., Jakana, J., Schmid, M.F., King, J., Nagayama, K., and Chiu, W. (2010). Zernike phase contrast cryo-electron microscopy and tomography for structure determination at nanometer and subnanometer resolutions. Structure *18*, 903–912.

Pattanayek, R., and Stubbs, G. (1992). Structure of the U2 strain of tobacco mosaic virus refined at 3.5 A resolution using X-ray fiber diffraction. J. Mol. Biol. *228*, 516–528.

Petoukhov, M.V., and Svergun, D.I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. Curr. Opin. Struct. Biol. *17*, 562–571.

Philippsen, A., Im, W., Engel, A., Schirmer, T., Roux, B., and Müller, D.J. (2002). Imaging the electrostatic potential of transmembrane channels: atomic probe microscopy of OmpF porin. Biophys. J. *82*, 1667–1676.

Ponting, C.P., and Russell, R.R. (2002). The natural history of protein domains. Annu. Rev. Biophys. Biomol. Struct. *31*, 45–71.

Rochel, N., Ciesielski, F., Godet, J., Moman, E., Roessle, M., Peluso-Iltis, C., Moulin, M., Haertlein, M., Callow, P., Mély, Y., et al. (2011). Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. Nat. Struct. Mol. Biol. *18*, 564–570.

Różycki, B., Kim, Y.C., and Hummer, G. (2011). SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. Structure *19*, 109–116.

Savage, D.F., Egea, P.F., Robles-Colmenares, Y., O'Connell, J.D., 3rd, and Stroud, R.M. (2003). Architecture and selectivity in aquaporins: 2.5 a X-ray structure of aquaporin Z. PLoS Biol. *1*, E72.

Schabert, F.A., Henn, C., and Engel, A. (1995). Native Escherichia coli OmpF porin surfaces probed by atomic force microscopy. Science *268*, 92–94.

Scheuring, S., Ringler, P., Borgnia, M., Stahlberg, H., Müller, D.J., Agre, P., and Engel, A. (1999). High resolution AFM topographs of the Escherichia coli water channel aquaporin Z. EMBO J. *18*, 4981–4987.

Scheuring, S., Busselez, J., and Lévy, D. (2005). Structure of the dimeric PufX-containing core complex of Rhodobacter blasticus by in situ atomic force microscopy. J. Biol. Chem. *280*, 1426–1431.

Scheuring, S., Boudier, T., and Sturgis, J.N. (2007). From high-resolution AFM topographs to atomic models of supramolecular assemblies. J. Struct. Biol. *159*, 268–276.

Schröder, G.F., Levitt, M., and Brunger, A.T. (2010). Super-resolution biomolecular crystallography with low-resolution data. Nature *464*, 1218–1222.

Schwede, T., Sali, A., Honig, B., Levitt, M., Berman, H.M., Jones, D., Brenner, S.E., Burley, S.K., Das, R., Dokholyan, N.V., et al. (2009). Outcome of a workshop on applications of protein models in biomedical research. Structure *17*, 151–159.

Sheng, S., Czajkowsky, D.M., and Shao, Z. (1999). AFM tips: how sharp are they? J. Microsc. *196*, 1–5.

Szymczyna, B.R., Taurog, R.E., Young, M.J., Snyder, J.C., Johnson, J.E., and Williamson, J.R. (2009). Synergy of NMR, computation, and X-ray crystallography for structural biology. Structure *17*, 499–507.

Ten Eyck, L.F., Mandell, J., Roberts, V.A., and Pique, M.E. (1995). http://www.supercomp.org/sc95/proceedings/636_LTEN/SC95.HTM. In Proceedings of the 1995 ACM/IEEE Supercomputer Conference, A. Hayes, and M. Simmons, eds. (San Diego: IEEE Computer Society Press, Los Alamitos, CA).

Trinh, M.-H., Odorico, M., Bellanger, L., Jacquemond, M., Parot, P., and Pellequer, J.-L. (2011). Tobacco mosaic virus as an AFM tip calibrator. J. Mol. Recognit. *24*, 503–510.

Ubbink, J., and Schär-Zammaretti, P. (2005). Probing bacterial interactions: integrated approaches combining atomic force microscopy, electron microscopy and biophysical techniques. Micron *36*, 293–320.

Umemura, K., Sutoh, K., Tokunaga, F., Kataoka, M., Kamikubo, H., Arakawa, H., and Ikai, A. (1996). The structure difference of proteins isolated on substrate with different techniques as studied by the atomic force microscope. Scanning *18*, 275–280.

Villarrubia, J.S. (1997). Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation. J. Res. Natl. Inst. Stand. Technol. *102*, 425–454.

Wisedchaisri, G., and Gonen, T. (2011). Fragment-based phase extension for three-dimensional structure determination of membrane proteins by electron crystallography. Structure *19*, 976–987.

Zheng, W. (2011). Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. Biophys. J. *100*, 478–488.

Zhou, Z.H. (2008). Towards atomic resolution structural determination by single-particle cryo-electron microscopy. Curr. Opin. Struct. Biol. *18*, 218–228.

