Perspectives Toward an Integrative Structural Biology Pipeline With Atomic Force Microscopy Topographic Images

Jean-Luc Pellequer

Journal of Molecular Recognition

WILEY

Journal of Molecular Recognition

REVIEW OPEN ACCESS

Perspectives Toward an Integrative Structural Biology Pipeline With Atomic Force Microscopy Topographic Images

Jean-Luc Pellequer

Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France

Correspondence: Jean-Luc Pellequer (jean-luc.pellequer@ibs.fr)

Received: 15 July 2024 | **Revised:** 21 August 2024 | **Accepted:** 3 September 2024

Funding: This work was supported by Agence Nationale de la Recherche (grant ANR-07-PCVI-0002-01 to J.-L.P.).

ABSTRACT

After the recent double revolutions in structural biology, which include the use of direct detectors for cryo-electron microscopy resulting in a significant improvement in the expected resolution of large macromolecule structures, and the advent of AlphaFold which allows for near-accurate prediction of any protein structures, the field of structural biology is now pursuing more ambitious targets, including several MDa assemblies. But complex target systems cannot be tackled using a single biophysical technique. The field of integrative structural biology has emerged as a global solution. The aim is to integrate data from multiple complementary techniques to produce a final three-dimensional model that cannot be obtained from any single technique. The absence of atomic force microscopy data from integrative structural biology platforms is not necessarily due to its nm resolution, as opposed to Å resolution for x-ray crystallography, nuclear magnetic resonance, or electron microscopy. Rather a significant issue was that the AFM topographic data lacked interpretability. Fortunately, with the introduction of the AFM-Assembly pipeline and other similar tools, it is now possible to integrate AFM topographic data into integrative modeling platforms. The advantages of single molecule techniques, such as AFM, include the ability to confirm experimentally any assembled molecular models or to produce alternative conformations that mimic the inherent flexibility of large proteins or complexes. The review begins with a brief overview of the historical developments of AFM data in structural biology, followed by an examination of the strengths and limitations of AFM imaging, which have hindered its integration into modern modeling platforms. This review discusses the correction and improvement of AFM topographic images, as well as the principles behind the AFM-Assembly pipeline. It also presents and discusses a series of challenges that need to be addressed in order to improve the incorporation of AFM data into integrative modeling platform.

1 | Integrative Structural Biology

The study of molecular recognition has followed the second and third precepts of Descartes [1], also known as the reductionism approach. This approach assumes that understanding the three-dimensional (3D) structures of biological macromolecules and complexes is a prerequisite for comprehending their assembly and subsequent function. Over the past 30 years, high-resolution molecular structures obtained through x-ray diffraction or nuclear magnetic resonance (NMR) techniques have provided a detailed understanding of the molecular mechanisms of these macromolecules [2]. However, these techniques have mostly been successful in studying small, globular, and soluble molecules. Despite the successes achieved in the study of some large assemblies, such as nucleosomes, ribosomes, or viruses, significant challenges remain when attempting to investigate membrane proteins, large multivalent machineries, or large highly flexible

systems. Recently, there has been a shift in scientific focus toward challenging systems, such as macromolecular complexes and dynamic networks [3]. Despite significant advances in structure determination technology, particularly in the recent context of advanced cryo-EM [4], many protein complexes remain structurally elusive due to their unique biochemical and biophysical properties. Therefore, a so-called integrative (or hybrid) approach has become a practical alternative [5–7]. The aim is to combine information from different approaches, hence the term hybrid, to analyze the structural architecture of such complexes [8–11]. This involves interpreting data from biophysical-biochemical experiments, extracting relevant structural information, and generating structural models that accurately explain the experimental data [12, 13]. Integrative modeling platforms (IMPs) are becoming increasingly popular due to the availability of data repositories [14].

The field of integrative structural biology is expanding its list of techniques [8]. However, atomic force microscopy (AFM) remains poorly integrated [9], despite its growing influence on structural biology [15] and its integration with super-resolution optical microscopy [16]. The main reason for this oversight is the challenge of translating topographic surfaces into structural constraints. As previously noted, there exists a discrepancy between the stated physical resolution of AFM imaging (sub-nm) and its practical translation into atomic information at the Å scale [17]. However, AFM boasts an exceptionally high signal-to-noise ratio (SNR), such that a single image of a molecule suffices to describe its exact contour [17]. Therefore, if all AFM imaging conditions are optimal (sample quality, sample adsorption, ideally sharp tip, best imaging mode, etc...) a single topographic image of a single molecule is sufficient to provide for the reconstruction of a complex macromolecule. Such reconstruction was presented for the first time with the AFM-Assembly pipeline [18].

The term "AFM-Assembly" design a protocol used to assemble three-dimensional units beneath a topographic surface provided by an AFM image [18]. A reverse strategy is also developed where multiple coarse-grain models of macromolecules are generated, followed by a conversion of atomic coordinates into density maps (or topographic images) [19]. A comparison between experimental AFM and computed images then provides a fitting score of the tested 3D model [19]. Several pipelines are currently being developed for the rigid-body fitting of AFM data. These developments have been reported in recent studies [20–22] including attempts to integrate structural dynamics as flexible three-dimensional coordinates using normal modes [23].

2 | AFM Data and Integrative Structural Biology

Our seminal idea behind the introduction of AFM data in structural biology emerged as a result of our participation in CASP experiments in the late 1990s [24], beginning with CASP4 [25]. At that time, comparative modeling had reached a performance plateau where easy targets were modeled correctly on a global level, while difficult targets were modeled poorly. The difficulty level was determined by identifying a hit using PSI-BLAST [26] with any known protein structure that had a significant *E*-value (<0.02) [25]. Prior to the alpha-fold revolution [27], modeling strategies required a series of tricks and personal experience. In early 2000s, we aimed to implement a significant shift in our strategy to enhance the effectiveness of our comparative models by incorporating external experimental data into the prediction pipeline. The ultimate goal of a model should not necessarily be measured by a root-mean-square metric [28], but mostly by its usefulness [29].

Although electron microscopy was initially preferred in our strategy (even before the direct detector revolution) AFM topography data was ultimately chosen, for multiple reasons, some scientific and some financial. In the 1990s and 2000s, AFM demonstrated considerable potential with the production of very high-resolution topography images by the group of Andreas Engel in Basel on two-dimensional crystal systems [30]. Inspired by these results, the concept of AFM-Assembly was announced in 2005 [31] and fully developed to construct large macromolecular systems (from large proteins to viruses) by assembling their structural components [18]. The development of AFM-Assembly was motivated by two other reasons. First, the CATH statistics on the Research Collaboratory for Structural Bioinformatics (RCSB) showed a decrease in the discovery of new folds in the Protein Data Bank (PDB) (Figure 1A), leading to the conclusion of an adequate structural coverage of small proteins [32, 33]. Second, there was an obvious discrepancy between the sequence length coverage of proteins in the PDB and the nonredundant protein sequence world (Figure 1B) [34]. Then, it was postulated that, in the event that fewer novel folds were discovered, the assembly of most folded protein structures could be achieved through the combination of known folds. It is important to note that this idea excludes completely disordered proteins or protein regions [35]. Additionally, these considerations were also made prior to the AlphaFold revolution [27], which now further enhances the potential applications of the AFM-Assembly pipeline by providing access to nearly all possible protein unit structures.

AFM-Assembly is not a strictly comparative modeling approach focused on generating three-dimensional coordinates for all residues of a protein. Instead, it is a 3D jigsaw puzzle assembly. The tool has two faces: a pipeline that can reconstruct large, flexible macromolecules and a technique that can be used as a filter to validate 3D structures of large systems using topographic data as a discriminant function (Figure 2A,B). Accordingly, AFM data can be integrated into a modeling pipeline either at its beginning or at its end. Additionally, this pipeline can address the overall fit of a large and flexible macromolecule conformation by identifying multiple alternative conformations that are compatible with single isolated molecular topographies [36, 37]. It is important to note that the same strategy can be used to determine the most likely interaction between macromolecules that bind to each other when the AFM topographic image of the complex is available.

With regard to epistemology, the AFM-Assembly pipeline offers a novel outcome in the use of single molecule imaging with scanning probe microscopy. Despite the nanometer resolution data, structural interpretation of AFM topographic images was primarily limited to qualitative comparisons or handmade overlaps [38–50]. Not surprisingly, previous studies of high-resolution AFM topography were biased toward flat 2D

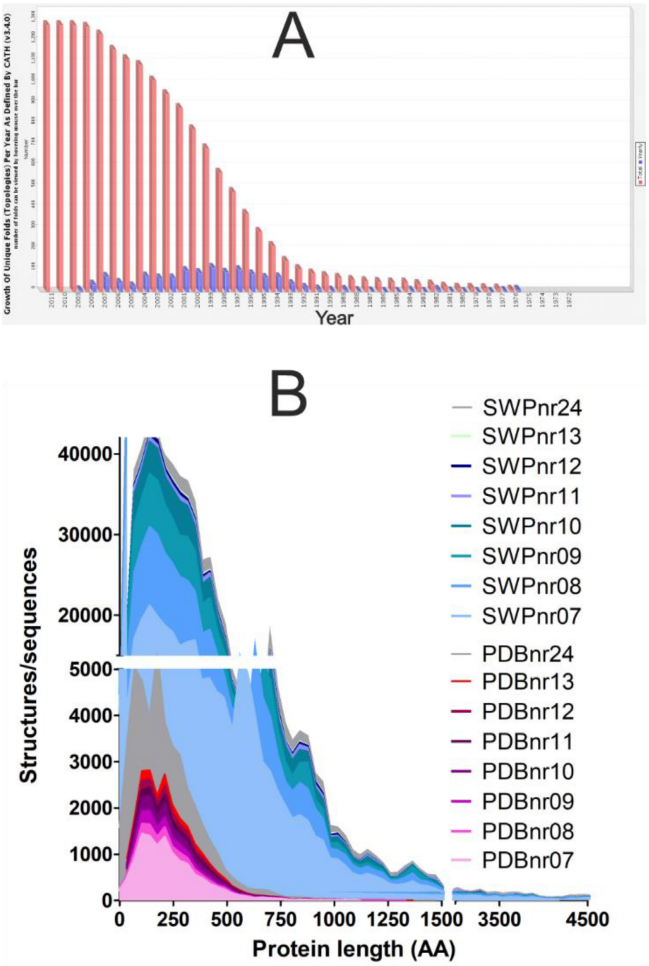

FIGURE 1 | Evolution of the distribution of protein sequence and structure databases. (A) The distribution of the growth of new folds as defined by the CATH method [138]. The plot was extracted from the statistical data section of the RCSB website, which is no longer available. The *x*-axis corresponds to years starting from 1972 (far right) to 2011 (far left). The distribution of observed folds in the Protein Data Bank (PDB) shows a plateau in the cumulative distribution in the most recent years of the graph (2009–2011), indicating that no new folds have been observed in the PDB. (B) Overlay distribution of the length of protein sequences found in a nonredundant PDB list (server Pisces [176] and the nonredundant Swiss-Prot database [177]) in function of years. Histograms were built using a bin size of 35 residues by extracting protein lengths from downloaded data files For example, in 2024, the files cullpdb_pc90.0_res0.0-3.0_len40-10000_R0.3_Xray_d2024_03_18_chains46162 and the "ID" field of uniprot_sprot.dat were used for the Pisces and Swiss-Prot databases, respectively. While there has been a notable expansion in the number of protein structures, the discrepancy in the distribution of protein sizes in sequence and structure databases is indicative of a bias in three-dimensional structures.

molecular systems [51–55]. This observation may be explained by the stability of the biological system on the AFM substrate, the low roughness of the sample, and by the possible molecular averaging provided by cross-correlation. Nevertheless, AFM can also investigate the structural features of single isolated molecules with nanometer resolution, as demonstrated by various studies [45, 50, 56–63]. The advent of AFM-Assembly has enabled the connection of nanometer-resolution AFM topographic images to the atomic resolution of protein structures [18].

In this review, we will briefly describe AFM topography data and its processing, a key step for interpreting structural biology. We then describe the AFM-Assembly pipeline to understand the critical steps of the process. Finally, we outline a development path to further integrate AFM data into integrative structural biology.

3 | AFM Imaging

Atomic force microscopy is a near-field microscopy that enables a wide range of physical measurements [64]. Initially developed for surface imaging [65], AFM evolved from the earlier invention of scanning tunneling microscopy [66]. The fundamental aspect of AFM imaging is the recording of interactions between a scanning sensor and a sample deposited on a flat substrate. The most common type of sensor is a silicon/silicon-nitride nanotip placed under a micro-sized cantilever [67]. Other materials that can be used for the tip include colloidal particles [68], carbon nanotubes [69]. The most commonly used flat substrate is a phyllosilicate mineral of aluminum and potassium, known as the muscovite mica sheet, which is atomically flat and cleavable [67]. Other substrates exist such as silicon wafers [70], highly-oriented pyrolytic graphite [71], and glass [72]. Most of these surfaces can be chemically modified to allow for the grafting of various molecular systems, such as DNA, proteins, bacteria, and cells [73].

The optical lever detection system is the most commonly used method for biological samples in AFM. A laser beam reflects off the back of the cantilever onto a split photodiode [67], and reviewed in detail by Allison et al. [74]. The piezoelectric tubes control the displacement of either the stage or the cantilever, depending on the AFM instrument, allowing for the surface scanning with near-angstrom resolution [75]. An additional piezoelectric tube often operates in the z-direction, providing sub-angstrom vertical displacement [76]. To obtain a topographic image of a sample, a feedback loop usually connects the photodiode detector to the z-piezo. The control of AFM imaging consists in keeping the interaction force between the sample and the cantilever constant. Therefore, if the photodiode signal increases, it indicates that the cantilever is climbing over the sample. To maintain a constant interaction force, the sample stage must lower itself proportionally. As a result, tracking the z-piezo, which controls the tip/sample movement, directly provides a sample height signal. The AFM records multiple lines to provide a topographic surface of the sample. It is important to note that the results are the same whether the z-piezo moves the sample stage or the cantilever. Recent reviews have illustrated the use of AFM for topographic surface imaging of biological samples [77, 78]. It is important to emphasize that the major advantage of AFM imaging for structural biology is the ability to perform AFM experiments in a liquid environment at ambient temperature [79].

The quality of a sample has a significant impact on the performance of the biophysical techniques to which it is subjected, including AFM. At the nanometer resolution, it can be

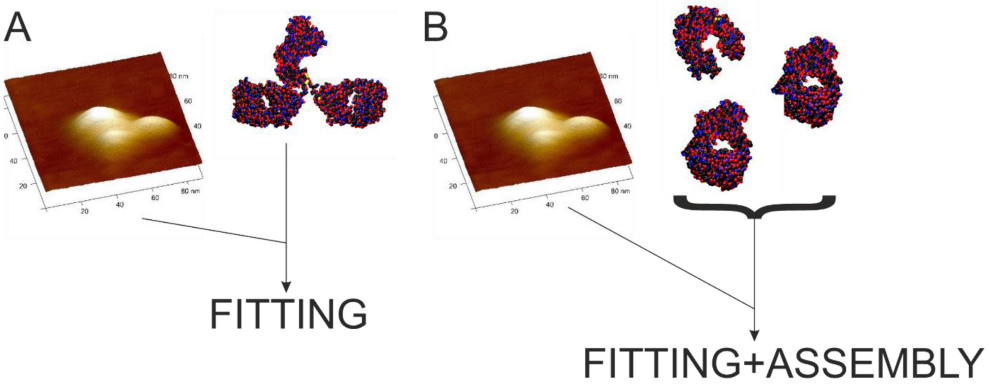

FIGURE 2 | Dual usage of the AFM-Assembly pipeline. (A) An AFM topograph of a single complete Immunoglobulin G (IgG) molecule is used to fit a complete 3D structure of an IgG (1IGT), thereby providing a best-fitting score between the atomic structure and the experimentally determined protein topography. (B) The same AFM topograph can be used to assemble the three structural units (two Fabs, 1AY1, and one Fc, 1H3T) that collectively define a complete IgG structure [18]. In this case, individual structural units are fitted first and a brute force combinatorial assembly is used to generate a list of full IgG structures. The AFM topograph is shown in 3D perspective (Nanoscope software), while the protein structures are drawn in van der Waals spheres and colored according to classical atom types. Structures are drawn using VMD [178] and rendered using Raster3D [179].

difficult to distinguish between different globular proteins. Although purity is not a physical requirement for AFM, as it does for crystallization for instance, a sample that is as free of impurities as possible is preferable. It is important to note, and too seldom mentioned in the literature, that obtaining a perfect AFM topographic image is rare. AFM imaging requires a particular alignment of parameters, including but not limited to: imaging mode (contact, tapping, or force-based), the imaging setup (feedback setpoint, scanning rate, scanning gains, and scan size), the imaging tip (large, thin, or ultrathin), and tip durability over time, as well as sample deposition on a flat substrate [80]. The operator's resilience is often tested during AFM imaging, but the effort usually pays off. It would be erroneous to assume that the infrequency of obtaining a perfect AFM image is indicative of a lack of performance or the rarity of the observed molecular system configuration. It is crucial to reiterate that AFM images capture the structural details of individual molecules. It is well established that the specific configurations of individual molecules may not align precisely with the average configuration [81]. Although the high degree of certainty displayed by the reference x-ray diffraction technique can make a dynamical view of single molecule conformations seem somewhat disturbing, it is important to note that considering multiple alternative conformations of proteins is actually a major benefit in protein engineering [82]. It should be emphasized that while the molecular conformation may vary, the short-range atomic coordinates remain relatively well conserved, even in distantly related protein structures [83]. A modeling technique should be capable of validating multiple model conformations, instead of selecting a single 'ideal' model. AFM data can contribute to this validation process in the field of integrative structural biology.

4 | Topographic Image Corrections

Scanning atomic force microscopes acquire data line by line. Each line consists of a set of data points that correspond to physical measurements such as height, phase, and signal error. These data points are often referred to as pixels in everyday language, but this term can be confusing for unfamiliar users as it suggests that AFM topographic data are images, when in fact they are true measurements represented by a series of 32-bit float data. Early on, it was recognized that the scanning direction in the raw data may not be normal to the sample surface, often requiring several corrections [84]. One of the most common issues observed is the presence of tilt and/or bow. Most instrument systems and classical free processing software such as Gwyddion [85], now include standard corrections (Figure 3A–H). This task should not be underestimated as it has consequences in image interpretation, specifically in determining height and volume. It is important to note that flattening can only be applied when the background surface is visible in the AFM image. It is not appropriate to correct the flatness of objects, such as cells or layers of nanoparticles. Only a visible background can be corrected. Automated systems have been developed to improve the task of making optimal corrections [86, 87]. Another permanent artifact in AFM images is known as the tip effect, which causes the thickening of all imaged particles due to the finite size of the AFM tip. Corrective methods have been developed for estimating the "true" size of sample images by AFM, such as for protein-DNA samples [88] or nanoparticles deposited on flat or rough substrates [89]. An optional step in reducing the scanning tip effect is to use erosion as an image-processing step of the AFM topography. A successful erosion depends on various parameters, including data resolution, tip radius, biological samples (single protein or large virus), and the structural biology question to be answered. To erode a dilated AFM image, it is essential to have an accurate understanding of the shape of the AFM tip. A number of methods can be used to determine the shape of an AFM tip, including the use of rod-shaped virus of known dimensions [90], the creation of artificial nanostructures of specified shapes [91], or the utilization of an AFM image of the tip [92]. Another methodology involves deconvolution using the localization of the true contact point between the AFM tip and the sample [93]. A shrewd implementation of this idea simultaneously increased the local sample resolution of the imaged object and reduced its apparent dilation [94]. If information about the AFM tip used to image the sample

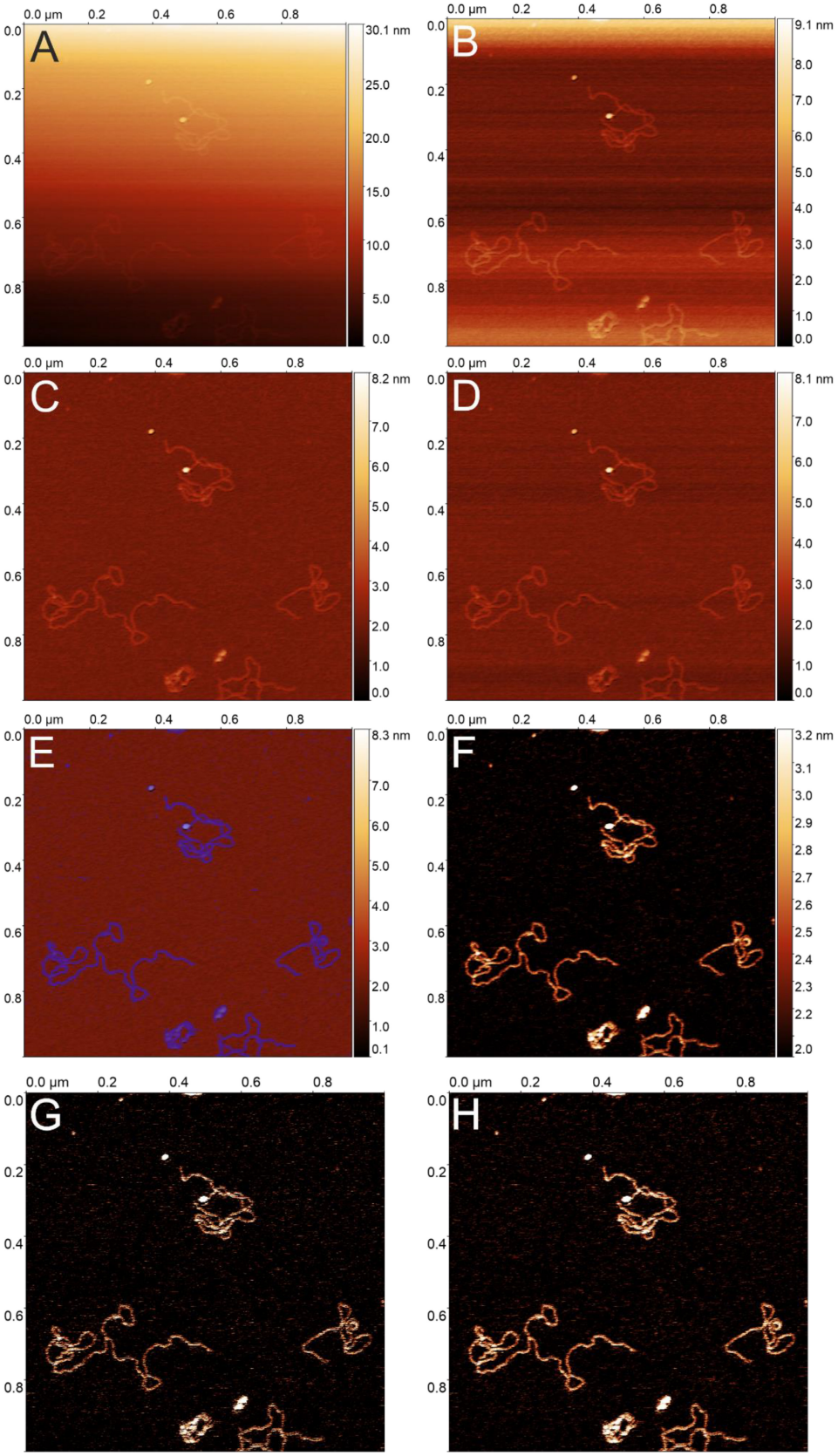

FIGURE 3 | Legend on next page

is unavailable, a model tip can be defined based on the manufacturer's specifications. The tip can be modeled as a cone with a sidewall angle and a tangent sphere with a given radius. It is important to note that an overestimation of the tip radius is preferable to an underestimation. The Villarrubia algorithm [95] is used to perform dilation and erosion operations on topographic

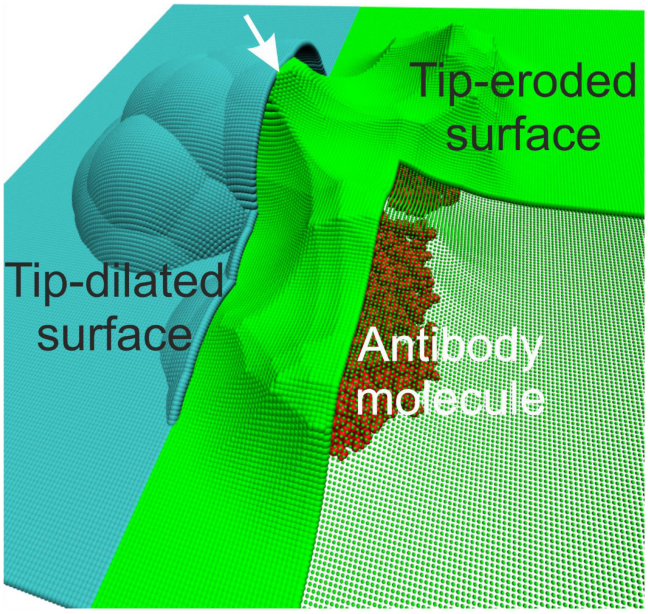

FIGURE 4 | Molecular effect of dilation/erosion on a simulated topography surface. The antibody molecule is represented in red spheres. The simulated AFM topography of the antibody molecule is represented by cyan spheres. The corresponding eroded surface is shown in green spheres/dots surface. Dilation/Erosion calculations are performed using the routines developed by Villarubia [95].

images represented as 2D matrices (Figure 4) while recent developments include the reconstruction of reentrant surfaces or undercut features [96]. More recently, neural networks have been trained using deep learning on tip-convoluted and deconvoluted image pairs [97].

Microscopy data can be degraded by the presence of streaks, scratches, stripes, or curtain artifacts [98]. For instance, the origin of stripe noise can be attributed to the inadequate calibration of the relative gain, including the offset between forward and backward scans [99], environmental factors such as temperature, hygrometry, vibration, and acoustic noise, all of which are amplified when the scanning process is accelerated [100], and the truncation of the high-frequency signal [101], not to mention the ubiquitous random noise. A detailed description of common artifacts in AFM images has been provided elsewhere [75, 102]. To correct noise or artifacts from an image, filters are not always effective in completely removing noise and may also damage some important quantitative information [103]. For example, classical destriping methods identify "noisy" pixels and replace them with spatially interpolated values. However, this approach raises a data integrity issue [104]. The periodicity in stripe noise suggests that it can be removed through spectral analysis, followed by the removal of specific frequency components [99, 105]. The frequency-based method is similar to low-pass filters, which are not a perfect solution because they also eliminate some structural details that are only partially described by the removed frequencies, resulting in blurring, or ringing artifacts [99]. Furthermore, frequency domain processing may not be effective for a low density of stripes due to the absence of a strong peak in the signal data [106]. However, a recent comparison of different methods for filtering out the stripe noise shows that frequency-based methods (such as DeStripe) can perform well on conductive atomic force microscopy data [107]. Stripe noise is not the only processing available for AFM topographic images. Advanced filtering methods have also been developed to improve the visibility of images.

Because of the numerical complexity of AFM data, which is expressed as a matrix of float values, it is common to visualize such data using a color spectrum that translates the physical measurements into shades of gray/color. It is important to note that while the physical data is present, human perception fails to recognize all the quantitative information present in AFM data. This was the rationale behind the development of a set of advanced processing filters that aim to improve the visibility of an AFM topographic image [108]. The first filter combines stripe noise reduction with histogram equalization and computation of the mixed partial derivative of pixel intensity $D^2_{xy}I(x,y) \equiv \partial^2 I(x,y)/\partial x \partial y$, where $I(x, y)$ is the intensity at pixel (x, y). In short, the D^2_{xy} operation was used to reveal the magnitude of the intensity change. The visualization of protuberances of individual coat proteins of tobacco mosaic virus was made possible through this process [109]. Other image processing methods, such as the Bayesian-based iterative method for reconstructing fuzzy images, have also been utilized [110]. The compress sensing method's reconstruction capability has been tested with AFM data. This is a new type of sampling theory that allows the signal to be reconstructed from significantly fewer measurements [111] than required by the Nyquist–Shannon sampling theorem. Further development involves applying the Laplacian function of image intensity to enhance image contrast using either an additive or multiplicative factor (Figure 3G,H [108]). The effectiveness of the L-weight filter can be observed in the case of small, isolated single DNA-binding proteins [63] or helical assembly of partially disordered proteins into fibers [112]. However, this

FIGURE 3 | Various steps of AFM data correction and advanced processing with a scan of 1 μm^2 of 1024 pixels^2. (A) Raw AFM height image of a linear DNA segment of 2364 base pairs. A slope is observed in the imaged sample, with a notable variation in the color scale from the bottom to the top of the image. (B) The initial correction step is to flatten the raw data using a mean plane subtraction in Gwyddion [85]. Given that the DNA sample is deposited on a highly flat mica surface, it is evident that the image correction is not yet complete. (C) The final correction is achieved by adding a line flattening step (row alignment) using the median method. The image is fully corrected. (D) The same processing steps as in (B and C) were performed, but in a different orders. Step C was performed before Step B. The correction result is different from previous Step C. It highlights the important of the order of the processing steps. (E) In certain instances, it is possible to enhance the correction steps by excluding the sample pixels (highlighted in blue) and running another row alignment through a polynomial method (order 3) while excluding the sample pixels. (F) Final corrected AFM data where the color scale is adjusted to enhance the visibility of the DNA molecule. It should be noted that the color scale does not start at 0 nm. Advanced image processing may be employed using either the Laplacian mask filter (G) or the Laplacian weight filter (H), as described by Chen et al. [108].

filter has the disadvantage of losing the height scale, which is then expressed in arbitrary units. For optimal results, this gradient-based method should be applied to images with a large number of pixels. It was observed that a minimum of 512 or 1024 pixels gave the best results. A notable side effect of the L-weight filter is the reduction of the tip effect in AFM topography images, which opens the door to advanced fitting capabilities of AFM data in integrative structural biology, provided that the vertical data scaling is rectified.

5 | AFM-Assembly Pipeline

The AFM-Assembly pipeline is a set of scripts and programs designed to build multi-domain protein structures using a high-resolution AFM topographic images and atomic coordinates of the constituent protein domains [18]. Additionally, AFM-Assembly can simply compute a fitting score between an AFM topographic image and a single protein conformation (Figure 2A) [113]. The main idea behind AFM-Assembly is to convert AFM topographic images into volumetric data and then fit all structural units within the created volume using a six-dimensional full-space search [114]. This includes translation in three *X*, *Y*, and *Z* axes and the three rotation Euler angles [18]. To use the AFM-Assembly pipeline, an AFM topographic image (usually a crop of a larger image) and structural units of proteins of interest (usually in the form of a PDB file) are required.

Small data crops are extracted from the corrected AFM image using Gwyddion (Figure 3A–H) [85]. The final size of the crops in pixels are adjusted to fit the computing grid using Gwyddion's interpolation function (Schaum algorithm). It is not recommended to scale beyond ×4. To fit structural units under AFM topography, the quality of the surface is crucial, and therefore the highest possible raw AFM data resolution should be acquired (0.5–1 nm/pixel). The second set of inputs consists of the structural units or individual macromolecules. The Cartesian coordinates of the structural units are typically obtained from the PDB [115]. To reduce the computational cost of reconstruction, individual unit structures can be extruded (Figure 5) by removing buried atoms in the molecule using the Adepth server [116] or simply selecting carbon alpha (Cα) reduced structures. In any case, the resolution of AFM topography is insufficient for precise positioning of flexible atoms on surface proteins. Despite this limitation, full-atom fitting has been found to be successful in several systems [36, 37]. This success is usually attributed to the lower resolution of AFM topography compared with atomic resolution data of protein structures. It is also important to note that 3D coordinates obtained from the PDB represent a static view of a protein structure. Therefore, there is no guarantee that the side chains of protein structures in a 3D crystal will have a similar orientation as in a single molecule deposited on mica. However, this is not a major limitation for the fitting of atomic coordinates under the nm-resolution AFM topographic images. Recent developments suggest that using Cα protein structures as input for the AFM assembly pipeline (unpublished data) can also yield successful results.

To reconstruct a large protein structure from its constituent fragments, it is necessary to map the AFM topographic surface onto a 3D Cartesian grid. The topographic surface is interpolated to contain the same number of pixels as used in the grid (docking step). The conversion of DOT2 [117] from a docking program to a fitting program was made possible by converting the AFM topographic surface into a volume. The AFM data are then mapped onto grid points. Points above the AFM topographic surface are labeled as "forbidden", while those below are labeled as "attractive." The thickness of the attractive layer is user-defined and typically ranges from 8 to 15Å (Figure 5). Each protein structural unit is independently fitted under the AFM topographic image using an exhaustive search with the FFT-based rigid body docking software DOT2. Each unit of the protein structure is translated to every node of the grid followed by a rotation step (from 6° to 20°) and a restart of the translation search, thereby defining a unique pose. A score is computed for each pose at each grid point based on the number of atoms of the fitted structure that are within the attractive layer. To reconstruct the complete protein structure, it is necessary to assemble each constituent

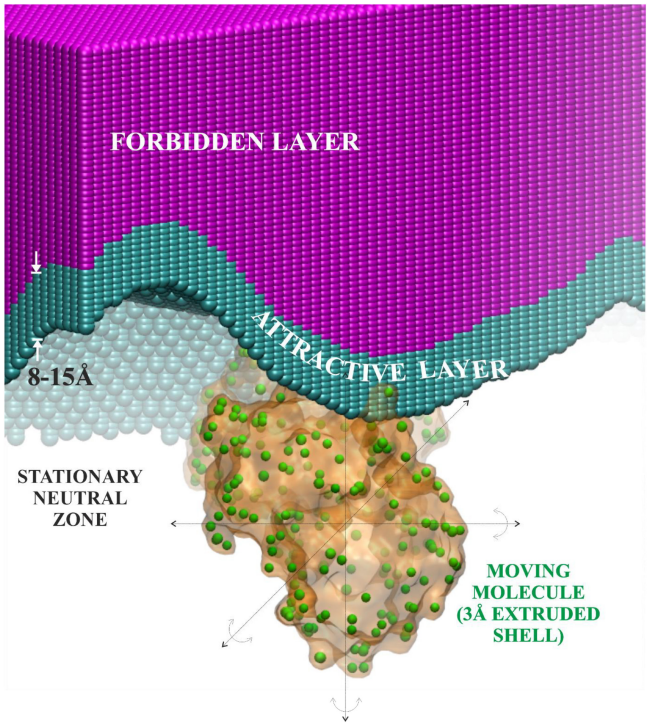

FIGURE 5 | Conversion principles of the DOT2 docking program into a fitting program. The computations are performed on a cubic grid. The grid is used to map atomic coordinates (illustrated by the extruded green spheres beneath the orange molecular surface) and the topographic data depicted here by the interface between an attractive layer in cyan (favorable) and a forbidden layer in magenta. In all other locations, the volume is considered to be a neutral zone. In the context of the DOT2 docking language, the topographic surface is regarded as the stationary system, whereas the atomic coordinates are designated as the moving system. The thickness of both layers is user-defined and depends on the shape of the molecular system (or the fitting purpose). The computation is accelerated by a FFT that was optimized for three-dimensional real transforms [114]. The orange molecule is translated to each grid point in all directions. Subsequently, the orientation of the orange molecule is altered, and the translation process is resumed. A straightforward fitting score is calculated by counting the number of green atoms that enter the attractive layer [18].

fragment. Our homemade software, combine, is used to perform the final assembly of all constituent fragments. In brief, a brute force exhaustive search is performed between the N-top fitting poses of each structural unit. The computation time increases exponentially with the number of poses used for each constituent fragment, resulting in a large number of possible combinations. Currently, we use around 2000 poses for a structure composed of three structural units to ensure reasonable computation time reasonable (a few hours). Note that the total number of structural units of a macromolecule should determine the number of poses. Distance constraints can be defined to limit the number of possible combinations and avoid testing unrealistic configurations, further reducing computation time. It is important to note that when assembling a flexible multi-domain macromolecular system or docking two macromolecular systems against each other, the same computational search is used.

6 | Perspectives

This section presents potential ways to enhance the integration of AFM data into modeling platforms. The basic concept of integrative structural biology is to integrate more data from different sources to achieve useful 3D models [6]. Integrative structural biology aims to provide understanding of structural biology, from atoms to molecules, from molecules to cells, and from cells to tissues, through high-resolution 3D structures. Illustrative works were pioneered in the realm of AFM imaging during the 1990s [70, 118–120]. The following paragraphs identify critical topics that are perceived as obstacles to the introduction of AFM in integrative structural biology.

How can an AFM-based pipeline be improved to become an IMP? According to the founding principles, an IMP should integrate data from multiple sources. Atomic coordinates can be fitted for electron microscopy data using efficient flexible fitting software [121, 122]. Can AFM-Assembly integrate data from related biophysical sources like cryo-electron microscopy (cryo-EM) or small-angle scattering techniques (such as x-ray, SAXS, or neutrons, SANS)? In principle, the strategy of fitting atomic coordinates beneath a topographic surface can handle any volumetric data, such as an external surface density, provided by cryo-EM or SAXS. If a conversion tool is available, such as the one present in the SITUS package [123], it is possible to convert a density volume into a PDB structure. The full 6D search can exhaustively place atomic coordinates within the volume by centering the volumetric density in the cubic grid of DOT2. Similar to AFM-Assembly, an attractive layer is constructed inside the volume while a forbidden layer is placed outside the volume boundary. Therefore, the AFM-Assembly pipeline should be able to run with cryo-EM or SAXS-derived data, with only a minor change in the input handling.

A critical aspect of a complex computer pipeline is to ensure the delivery of data in a standard format. In other words, standardization is essential for computational pipelines. This was accomplished successfully for 3D atomic structures with the construction of the Protein Data Bank [124], its extension including a pipeline for data collection, processing, archiving, and query [115], and further refinement through the wwPDB [125]. The PDB format provided a structured file format for macromolecular structures, including proteins, DNA, sugars, lipids, and small molecules. However, due to the amount of data and the speed of its acquisition, the PDB format has become obsolete and has been replaced by the newer mmCIF file format [126]. Unfortunately, there is currently no databank for AFM data, yet. One of the main challenges in AFM is the variety of data types, including images, force-distance curves, and force-volume maps, as well as the different recording modes such as height, phase, friction, or signal error. However, despite previous standardization efforts in AFM [127–131], file format standardization remains a major obstacle. It is important to recognize that standardization efforts should not be underestimated, as they require significant resources and efforts both on the metadata dictionary definition and on the computer pipelines. There is a trade-off between flexibility and reliability in AFM data standardization. Excessive flexibility creates a significant barrier to mastering the pipeline, while strong reliability can limit its applicability. Therefore, it is crucial to standardize AFM data to achieve the same level of accessibility [132] as other structural techniques such as PDB [124], EMDB for cryo-EM data [133], or SASBDB for small angle scattering data [134].

The development of advanced systems requires well-controlled pipeline automation. As previously mentioned, AFM image processing has attempting automation, as demonstrated by the TopoStats program [87]. The objective is to correct raw AFM image data using a robust procedure. Subsequently, with a corrected AFM image, the next step is to automatically extract topography of isolated single molecules. The challenge lies not in the technical aspect, but in the selectivity of this step. Classical image segmentation tools such as thresholding, watershed, and similar techniques [135], provide the basis extracting topograms. One challenge is to screen for "unspecific" topograms that do not accurately describe the surface of the molecule of interest. This step is accessible to human vision, but requires additional knowledge for an automated program. This includes the expected molecular height, size, and the effect of the AFM imaging mode. Additionally, it is important to consider the expected contamination of the sample, as this can greatly impact the efficiency of segmentation. Segmentation of an AFM image is straightforward for a pure sample and a recently developed object detector gave interesting results on single DNA molecules [136]. However, in the presence of denatured particles, contaminating particles, or fuzzy topographic edges, this step can be compromised, and manual picking may be necessary. Therefore, even though image segmentation or picking is commonly used for EM data, there is a need to develop automated picking features for AFM images.

The AFM-Assembly pipeline requires multiple sets of 3D coordinates for the investigated molecule or coordinates of the interacting molecules. The definition of structural units in protein structures is rather fuzzy and does not correspond to the "standard" topological definition. Structural units in a multimeric system may include monomers, functional units such as catalytic subunits, or regions of low flexibility (lacking flexible linkers). Practical definitions, such as the use of tobacco mosaic virus disks to assemble a larger section of TMV, are also possible [18]. Manual and automated computational approaches have been developed to detect domains in protein structures [137–139] and evolutionary-based domain identification has also been

developed in Ecod [140]. Recent attempts to delineate domains in protein structures include Chainsaw, a novel neural network supervised learning approach [141], and Spectraldom that applies a graph coding inter-atomic fluctuations derived from an elastic network model to delineate protein domains [142]. Spectraldom is built upon Spectrus that is efficient in providing multiple subdivisions of a protein [143].

However, current sequence- and structure-based methodologies do not fully capture the visual classification of structural units necessary for assembling large macromolecular systems. The identification of flexible linkers in proteins alone is insufficient to meet this requirement. For instance, a standard type G immunoglobulin (IgG) is composed of three units: two Fabs and one Fc. Each Fab is composed of two chains, each chain consisting of two domains, and each domain containing a single Ig-like fold. In this case, Pfam separates the two Fv domains due to a short structural link between the two consecutive Ig folds [144]. However, the structural flexibility of two consecutive Ig folds is rather limited compared to the flexibility between a Fab and an Fc unit, essentially due to the dimerization of the light and heavy chains of a Fab. Additionally, the Spectrus approach has a limitation in the input as it only accepts proteins based on their PDB chains [143]. This is a limitation for antibody structures. Accurately defining the structural unit at the start of the project, which is currently a manual step, is crucial. For instance, the use of metafolds is not a promising strategy, despite their potential to simplify the fold space [145], similarly to use the definitions employed in the CASP experiments, as they were intended to classify targets based on prediction difficulty [146]. Manual curation of protein folds is ongoing [139], with recent focus on defining evolutionary domain [140]. A graph-spectral method [147] may be a promising approach for identifying structural units. Are the use of elastic network models or the identification of deeply buried structural cores be promising avenues for research? Therefore, although there are structural domain predictors, further development is still needed to describe protein structures hierarchically and simplify three-dimensional coordinates, such as expected by the definition of units within AFM-Assembly.

The AFM-Assembly pipeline can also perform simple shape fitting by limiting its activity to a 6D search of the atomic coordinates beneath the topography image [113]. To take full advantage of this mode, alternative protein structures should be provided, and thus avoid the assembly (combine) part of the pipeline (Figure 2B). The question is whether one protein conformation fits better than another one under the experimental topography image. The use of AlphaFold to provide complete atomic structures of any protein is undoubtedly advantageous. However, to determine the best fit, it is preferable to test hundreds or thousands of putative structures. There are several approaches to provide alternative protein structures, such as full-atom simulations (molecular dynamics [148] or Monte Carlo search [149]) or normal mode analysis. The advantage of MD simulation is the physical-based quality of the atomic structures, while the MC simulation offers larger conformational space sampling. However, both methods have a major drawback of intensive computational costs with a rather low diversity of generated protein structures. A simple method (Concoord) has been developed to obtain most prominent collective structure variations [150], however, limited to the concept of structural fluctuations. To study the collective motion of atoms in a protein structure, normal mode analysis is a computational method that can be used [151, 152]. Normal mode analysis can provide insight into the flexibility and dynamics of a protein by calculating its vibrational modes. These vibrational modes represent the different ways in which a protein can move without breaking covalent bonds. The normal modes with the lowest frequency are associated with the significant amplitude conformational changes in proteins. To speed up the calculation of normal modes of large proteins, various methods have been employed [153], such as the nonlinear rigid block method [154]. From these simulations, it is possible to generate of a set of pseudo-random molecular structures with a constant RMSD from a reference molecule [155]. Additionally, the Gaussian mixture model, an extension of normal modes, has been used to produce diverse protein conformations from simulated AFM images [19]. Recent developments in elastic network models are enhancing the ability to predict the functional states of proteins [156]. Further progress in characterizing conformational variability is promising [157]. Therefore, there is a need to develop techniques that can generate alternative protein structure conformations, including the possibility of generating an "equidistant beyond fluctuation" set of protein structures, at least from an atomic coordinate perspective rather than an energetic one.

An alternative approach to generating alternative protein structure conformations is to directly image single molecules in real time. This is currently accomplished through the use of high-speed atomic force microscopy (HS-AFM). HS-AFM is a technique that allows for rapid imaging of surfaces [158]. The instrument was designed to capture dynamic processes in real-time with high spatial resolution. The principles of HS-AFM are based on a meticulous optimization of the tip-sample interactions, an enhanced feedback control system, and the implementation of advanced imaging modes [159]. In particular, HS-AFM benefits from the improvement of small cantilevers [160], which allows for excitation frequencies in the MHz range. Furthermore, the development of sample surface preparation techniques, such as streptavidin 2D crystals, which permit the binding of biotinylated single proteins [161] has been instrumental in enabling high-speed imaging. To cite a few examples, HS-AFM has been used to visualize intrinsically disordered regions of proteins [162], to image "walking" myosin V [163], to image the growth and assembly of amyloid-like fibers [164], to image rotary catalysis of F1-ATPase [165], to characterize the motion of membrane proteins [166], to study membrane deformation [167], or to image various systems as transmembrane channels and transporters [168]. Additionally, high-speed AFM is capable of providing information in high-speed force spectroscopy [169], as demonstrated by Rico et al. [170] or of studying single protein unfolding [171]. The connection established between AFM topography and macromolecular structures with the AFM-Assembly pipeline has been already extended to HS-AFM data [172] where the authors survey various methodologies, including simulations, rigid-body fitting, and flexible fitting, which are used to construct 3D atomic structures from AFM images. Recently, an extension of an existing frame-by-frame rigid-body fitting analysis

has been adapted to multiple frames in HS-AFM data [21]. Similarly to AFM data, a reverse strategy has also been applied to HS-AFM data with 3D coordinates obtained from molecular dynamics simulations [173]. In such a reverse strategy, the conformational space of the molecule of interest is extensively covered. A transformation of such conformational poses into simulated AFM images allows for a direct comparison with AFM and HS-AFM experimental data. Once a fit between the experimental and simulated images has been identified, the corresponding conformational pose can be readily extracted [22]. The integration of these flexible fitting methods into a comprehensive modeling pipeline, along with the incorporation of HS-AFM data, remains the most exciting challenge currently facing the field.

The most demanding pipeline step is undoubtedly the fitting scoring function. This is because the ranking of conformational poses critically influences the assembly of structural units into complete molecules [6]. Additionally, the ranking is also essential for selecting alternative conformations, as explained above. Alternative conformations are not necessarily uncertainties but instead the expression of the dynamics in protein function. AFM data present a challenge due to the difference in scale between topography (in the nm range) and atomic coordinates (in the Å range). Recent developments using localization image reconstruction algorithms allow the improvement of AFM topographic image resolution beyond the limits imposed by the tip radius [174]. Similar to super-resolution fluorescence microscopy, localization atomic force microscopy (LAFM) takes advantage of the superposition of AFM images (trace, retrace, and averaging) and the extraction of local maxima. It is then possible to assign a probability function to these local maxima and produce improved LAFM maps. This new family of AFM image processing is likely to improve the integration of AFM data into structural biology applications [175].

Smoothing and interpolation may also be applied, but as demonstrated in the image processing section, these techniques can introduce artifacts that negatively impact the scoring function. To simplify matters, atomic structures can be represented as coarse grains or simply a Cα trace. But, the conversion to topography (simulated AFM images), or vice versa becomes more arduous. In AFM-Assembly, scores differ between fitting and assembly. The fitting score is proportional to the number of atoms from the tested structure, that are within the expanded volume under the topographic surface. Consequently, this scoring function produces different scores for the same molecule depending on whether a full atom or Cα trace representation is used. To address this issue, a normalization procedure should be developed. In the "combine" step, the score is determined by the agreement with distance constraints as well as the height of pixel (i, j) on the experimental image and the theoretical topographic image generated from the evaluated assembled model [18]. Studying the impact of different scoring functions, such as those used in simulated AFM topographic images with known protein 3D structures, can indicate the appropriate scaling to apply between the surface and the atomic data. Therefore, there is a need for the development of new scoring functions in the field of AFM integrative structural biology. Developments apply to both the reconstruction process (combine) and the global fitting score to validate the agreement between a predicted molecular conformation and an AFM topographic image of the molecule.

A crucial aspect to consider when designing advanced integrative structural biology computational pipelines is their accessibility. While open source is the preferred option, it may not be sufficient. The strategy for using the pipeline is critical. Will it be run on a remote server or locally? Does it contain third-party software that taints the open source principle and makes portability difficult? Can the pipeline be easily packaged for installation on multiple computer architectures? How can future pipeline developments be ensured by depositing the source code on open-source platforms such as the GitHub or GitLab repositories? Who will be responsible for maintaining the package? These are often overlooked details when starting a project, but they can become limitations to the FAIR usage of integrative modeling pipelines. From a long experience in laboratory software development (as opposed to company software), I have found that the user interface is constantly the most critical aspect in terms of both design and development time. It is essential to decouple the computational routines from the graphical interface, especially for young scientists who are not computer experts and are interested in developing laboratory software. Graphical interfaces often rely on libraries that are constantly evolving, sometimes resulting in a lack of backward compatibility. Additionally, the graphical interface may become an easy step for future AI-based programming. Therefore, it may be more efficient to focus on computation under the command

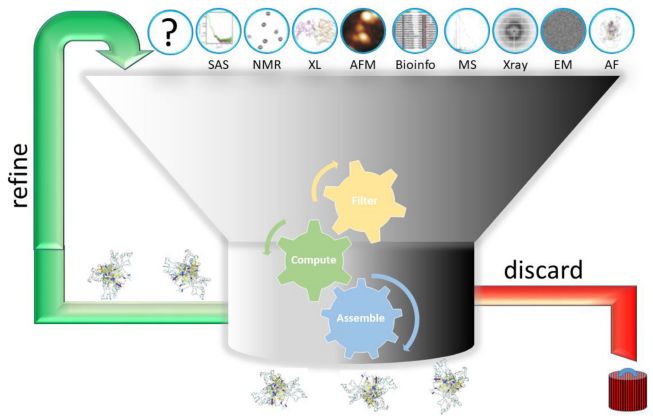

FIGURE 6 | Data workflow of an optimal integrative modeling pipeline (IMP). The experimental data are symbolized by a medallion above the funnel. A nonexhaustive list of data sources includes small angle scattering (SAS), chemical cross-linking (XL), atomic force microscopy (AFM), mass spectrometry (MS), x-ray crystallography (X-ray), nuclear magnetic resonance spectroscopy (NMR), electron microscopy (EM), alphafold server (AF), bioinformatics (bioinfo), and any additional putative technique, which is listed under "?" An optimal IMP would filter input data (reformat, correct, sort, rank, reject, etc.) prior to computing theoretical models and finally assembling the desired complex output. An optimal IMP will prioritize the ranking of intermediate and final constructions, reintroduce these data sets into the input, reject any inconsistent initial input data, and perform a virtue cycle by reprocessing models anew. It is important to note that the naming of experimental data is inclusive and does not restrict them to a particular set of experiments or computer versions.

line of the Linux environment before spending time choosing a graphical interface.

What would the ideal integrative modeling platform look like? In brief, it would be a self-deciding pipeline that recognizes the quality of input data, automatically selects which data to include, how to include it, and which technique to use when assembling output computational models. The platform should enable continuous improvement of a model by adding new data, rejecting some unreliable previous data, and reformulating a new set of models (Figure 6). This is a long wish list, but all these steps are critical in producing useful models. Keep in mind that each step listed is composed of a set of multiple tasks that each deserves a tight cooperation. Also, remember that the goal of computational models is to enlighten the function of biological molecules in the context of the cell or tissue. Nice-looking pictures or movies do not suffice. A main difficulty is the limited experience available in a single laboratory. The greater challenge is the need to shift the protein structure paradigm to accept that multiple conformations are not a lack of decision (selection criteria) but the best representation of functional molecules. Achieving this requires a concerted effort by a conglomerate that agrees to develop a part of such an ideal platform. This effort will require a high level of standardization and a smooth communication between the participating teams. Perhaps it is time for a young explorer to take up the pilgrim's staff and explore the possibility of forming such a conglomerate.

Author Contributions

J.-L.P. wrote this review.

Acknowledgments

IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA). This work acknowledges the AFM platform at the IBS. The initial part of this work was supported by the French ANR agency (grant ANR-07-PCVI-0002-01 to J.-L.P.). J.-L.P. extends his gratitude to Shu-wen W. Chen for her invaluable assistance and expertise in advanced AFM image processing and computational structural biology. J.-L.P. thanks Minh-Hieu Trinh for bringing AFM-Assembly to life and for creating some original images of the AFM-Assembly principles, and Rui C. Chaves for building the AFM-Assembly pipeline. J.-L.P. thanks Elisabetta Boeri-Erba (IBS-EPIGEN) for providing a MS spectrum, Florine Dupeux (IBS-VRM) for providing an x-ray diffraction pattern, Grégory Effantin (IBS-MEM) for providing a Cryo-EM image, Cédric Laguri (IBS-MP) for providing a NMR spectrum. The contribution of DeepL/write to improve the quality of the manuscript is acknowledged. J.-L.P. thanks Sergei Grudinin and Felix Rico for their constructive criticisms of the manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

Data sharing are not applicable to this article as no new data were created or analyzed in this study.

References

1. R. Descartes, *Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences* (Grand'Salle du Palais: Girard Théodore, 1668), https://gallica.bnf.fr/ark:/12148/bpt6k56982285.

2. M. Egli, "Diffraction Techniques in Structural Biology," *Current Protocols in Nucleic Acid Chemistry* 65 (2016): 7.13.1–7.13.41, https://doi.org/10.1002/cpnc.4.

3. R. Nussinov, C. J. Tsai, A. Shehu, and H. Jang, "Computational Structural Biology: Successes, Future Directions, and Challenges," *Molecules* 24 (2019): 637, https://doi.org/10.3390/molecules24030637.

4. W. Kuhlbrandt, "Biochemistry. The Resolution Revolution," *Science* 343 (2014): 1443–1444.

5. F. Alber, S. Dokudovskaya, L. M. Veenhoff, et al., "Determining the Architectures of Macromolecular Assemblies," *Nature* 450 (2007): 683–694, https://doi.org/10.1038/nature06404.

6. A. Sali, "From Integrative Structural Biology to Cell Biology," *Journal of Biological Chemistry* 296 (2021): 100743, https://doi.org/10.1016/j.jbc.2021.100743.

7. D. S. Ziemianowicz and J. Kosinski, "New Opportunities in Integrative Structural Modeling," *Current Opinion in Structural Biology* 77 (2022): 102488, https://doi.org/10.1016/j.sbi.2022.102488.

8. A. B. Ward, A. Sali, and I. A. Wilson, "Integrative Structural Biology," *Science* 339 (2013): 913–915, https://doi.org/10.1126/science.1228565.

9. M. P. Rout and A. Sali, "Principles for Integrative Structural Biology Studies," *Cell* 177 (2019): 1384–1403, https://doi.org/10.1016/j.cell.2019.05.016.

10. L. Cerofolini, M. Fragai, E. Ravera, C. A. Diebolder, L. Renault, and V. Calderone, "Integrative Approaches in Structural Biology: A More Complete Picture From the Combination of Individual Techniques," *Biomolecules* 9 (2019): 370, https://doi.org/10.3390/biom9080370.

11. S. J. Ziegler, S. J. B. Mallinson, P. C. St John, and Y. J. Bomble, "Advances in Integrative Structural Biology: Towards Understanding Protein Complexes in Their Cellular Context," *Computational and Structural Biotechnology Journal* 19 (2021): 214–225, https://doi.org/10.1016/j.csbj.2020.11.052.

12. C. Dominguez, R. Boelens, and A. M. Bonvin, "HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information," *Journal of the American Chemical Society* 125 (2003): 1731–1737, https://doi.org/10.1021/ja026939x.

13. D. Russel, K. Lasker, B. Webb, et al., "Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies," *PLoS Biology* 10 (2012): e1001244, https://doi.org/10.1371/journal.pbio.1001244.

14. B. Vallat, B. Webb, J. D. Westbrook, A. Sali, and H. M. Berman, "Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules," *Structure* 26, no. 6 (2018): e892, https://doi.org/10.1016/j.str.2018.03.011.

15. A. C. Dumitru and M. Koehler, "Recent Advances in the Application of Atomic Force Microscopy to Structural Biology," *Journal of Structural Biology* 215 (2023): 107963, https://doi.org/10.1016/j.jsb.2023.107963.

16. R. K. G. Tank, V. A. Lund, S. Kumar, et al., "Correlative Super-Resolution Optical and Atomic Force Microscopy Reveals Relationships Between Bacterial Cell Wall Architecture and Synthesis in *Bacillus subtilis*," *ACS Nano* 15 (2021): 16011–16018, https://doi.org/10.1021/acsnano.1c04375.

17. P. Fechner, T. Boudier, S. Mangenot, S. Jaroslawski, J. N. Sturgis, and S. Scheuring, "Structural Information, Resolution, and Noise in High-Resolution Atomic Force Microscopy Topographs," *Biophysical Journal* 96 (2009): 3822–3831, https://doi.org/10.1016/j.bpj.2009.02.011.

18. M.-H. Trinh, M. Odorico, M. E. Pique, et al., "Computational Reconstruction of Multidomain Proteins Using Atomic Force Microscopy Data," *Structure* 20 (2012): 113–120, https://doi.org/10.1016/j.str.2011.10.023.

19. B. Dasgupta, O. Miyashita, and F. Tama, "Reconstruction of Low-Resolution Molecular Structures From Simulated Atomic Force Microscopy Images," *Biochimica et Biophysica Acta* 1864 (2020): 129420, https://doi.org/10.1016/j.bbagen.2019.129420.

20. T. Niina, Y. Matsunaga, and S. Takada, "Rigid-Body Fitting to Atomic Force Microscopy Images for Inferring Probe Shape and Biomolecular Structure," *PLoS Computational Biology* 17 (2021): e1009215, https://doi.org/10.1371/journal.pcbi.1009215.

21. T. Ogane, D. Noshiro, T. Ando, A. Yamashita, Y. Sugita, and Y. Matsunaga, "Development of Hidden Markov Modeling Method for Molecular Orientations and Structure Estimation From High-Speed Atomic Force Microscopy Time-Series Images," *PLoS Computational Biology* 18 (2022): e1010384, https://doi.org/10.1371/journal.pcbi.1010384.

22. R. Amyot, N. Kodera, and H. Flechsig, "BioAFMviewer Software for Simulation Atomic Force Microscopy of Molecular Structures and Conformational Dynamics," *Journal of Structural Biology: X* 7 (2023): 100086, https://doi.org/10.1016/j.yjsbx.2023.100086.

23. R. Vuillemot, J.-L. Pellequer, and S. Grudinin, "AFMfit: Conformational Dynamics from AFM Data Using Fast Nonlinear NMA and FFT-Based Search," *bioRxiv* (2024), https://doi.org/10.1101/2024.06.03.597083.

24. J. Moult, J. T. Pedersen, R. Judson, and K. Fidelis, "A Large-Scale Experiment to Assess Protein Structure Prediction Methods," *Proteins* 23 (1995): 2–4, https://doi.org/10.1002/prot.340230303.

25. A. Tramontano, R. Leplae, and V. Morea, "Analysis and Assessment of Comparative Modeling Predictions in CASP4," *Proteins* S5 (2001): S22–S38, https://doi.org/10.1002/prot.10015.

26. S. F. Altschul, T. L. Madden, A. A. Schaffer, et al., "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," *Nucleic Acids Research* 25 (1997): 3389–3402, https://doi.org/10.1093/nar/25.17.3389.

27. J. Jumper, R. Evans, A. Pritzel, et al., "Highly Accurate Protein Structure Prediction With AlphaFold," *Nature* 596 (2021): 583–589, https://doi.org/10.1038/s41586-021-03819-2.

28. J. L. Pellequer, A. J. Gale, and E. D. Getzoff, "Blood Coagulation: The Outstanding Hydrophobic Residues," *Current Biology* 10 (2000): R237–R240, https://doi.org/10.1016/S0960-9822(00)00373-0.

29. S.-w W. Chen and J.-L. Pellequer, "Identification of Functionally Important Residues in Proteins Using Comparative Models," *Current Medicinal Chemistry* 11 (2004): 595–605, https://doi.org/10.2174/092986704345589 1.

30. A. Engel and D. J. Muller, "Observing Single Biomolecules at Work With the Atomic Force Microscope," *Nature Structural Biology* 7 (2000): 715–718, https://doi.org/10.1038/78929.

31. J.-L. Pellequer, J.-M. Teulon, L. Bellanger, C. Vidaud, M. Odorico, and P. Parot, "Introduction of Morphological and Dynamical Constraints in the Modeling of Protein Super-Complexes. 15th IUPAB & 5th EBSA International Biophysics Congress, Montpellier, France," *European Biophysics Journal* 34, no. 6 (2005): 743.

32. D. Kihara and J. Skolnick, "The PDB Is a Covering Set of Small Protein Structures," *Journal of Molecular Biology* 334 (2003): 793–802, https://doi.org/10.1016/j.jmb.2003.10.027.

33. Y. Zhang and J. Skolnick, "The Protein Structure Prediction Problem Could Be Solved Using the Current PDB Library," *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005): 1029–1034, https://doi.org/10.1073/pnas.0407152101.

34. M. Gerstein, "How Representative Are the Known Structures of the Proteins in a Complete Genome? A Comprehensive Structural Census," *Folding & Design* 3 (1998): 497–512, https://doi.org/10.1016/S1359-0278(98)00066-2.

35. A. K. Dunker, J. D. Lawson, C. J. Brown, et al., "Intrinsically Disordered Protein," *Journal of Molecular Graphics & Modelling* 19 (2001): 26–59, https://doi.org/10.1016/S1093-3263(00)00138-8.

36. R. C. Chaves, J.-M. Teulon, M. Odorico, P. Parot, S.-w W. Chen, and J.-L. Pellequer, "Conformational Dynamics of Individual Antibodies Using Computational Docking and AFM," *Journal of Molecular Recognition* 26 (2013): 596–604, https://doi.org/10.1002/jmr.2310.

37. R. C. Chaves, S. Dahmane, M. Odorico, G. A. F. Nicolaes, and J.-L. Pellequer, "Factor Va Alternative Conformation Reconstruction Using Atomic Force Microscopy," *Thrombosis and Haemostasis* 112 (2014): 1167–1173, https://doi.org/10.1160/TH14-06-0481.

38. A. Philippsen, W. Im, A. Engel, T. Schirmer, B. Roux, and D. J. Muller, "Imaging the Electrostatic Potential of Transmembrane Channels: Atomic Probe Microscopy of OmpF Porin," *Biophysical Journal* 82 (2002): 1667–1676, https://doi.org/10.1016/S0006-3495(02)75517-3.

39. D. M. Czajkowsky, E. M. Hotze, Z. Shao, and R. K. Tweten, "Vertical Collapse of a Cytolysin Prepore Moves Its Transmembrane Beta-Hairpins to the Membrane," *EMBO Journal* 23 (2004): 3206–3215, https://doi.org/10.1038/sj.emboj.7600350.

40. E. Davies, K. S. Teng, R. S. Conlan, and S. P. Wilks, "Ultra-High Resolution Imaging of DNA and Nucleosomes Using Non-Contact Atomic Force Microscopy," *FEBS Letters* 579 (2005): 1702–1706, https://doi.org/10.1016/j.febslet.2005.02.028.

41. S. Scheuring, J. Busselez, and D. Levy, "Structure of the Dimeric PufX-Containing Core Complex of *Rhodobacter blasticus* by In Situ Atomic Force Microscopy," *Journal of Biological Chemistry* 280 (2005): 1426–1431, https://doi.org/10.1074/jbc.M411334200.

42. S. Scheuring, N. Buzhynskyy, S. Jaroslawski, R. P. Goncalves, R. K. Hite, and T. Walz, "Structural Models of the Supramolecular Organization of AQP0 and Connexons in Junctional Microdomains," *Journal of Structural Biology* 160 (2007): 385–394, https://doi.org/10.1016/j.jsb.2007.07.009.

43. S. Scheuring, T. Boudier, and J. N. Sturgis, "From High-Resolution AFM Topographs to Atomic Models of Supramolecular Assemblies," *Journal of Structural Biology* 159 (2007): 268–276, https://doi.org/10.1016/j.jsb.2007.01.021.

44. N. Buzhynskyy, M. Golczak, J. Lai-Kee-Him, et al., "Annexin-A6 Presents Two Modes of Association With Phospholipid Membranes. A Combined QCM-D, AFM and Cryo-TEM Study," *Journal of Structural Biology* 168 (2009): 107–116, https://doi.org/10.1016/j.jsb.2009.03.007.

45. D. M. Czajkowsky and Z. Shao, "The Human IgM Pentamer Is a Mushroom-Shaped Molecule With a Flexural Bias," *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009): 14960–14965, https://doi.org/10.1073/pnas.0903805106.

46. H. Asakawa, K. Ikegami, M. Setou, N. Watanabe, M. Tsukada, and T. Fukuma, "Submolecular-Scale Imaging of Alpha-Helices and C-Terminal Domains of Tubulins by Frequency Modulation Atomic Force Microscopy in Liquid," *Biophysical Journal* 101 (2011): 1270–1276, https://doi.org/10.1016/j.bpj.2011.07.020.

47. S. A. Mari, J. Pessoa, S. Altieri, et al., "Gating of the MlotiK1 Potassium Channel Involves Large Rearrangements of the Cyclic Nucleotide-Binding Domains," *Proceedings of the National Academy of Sciences of the United States of America* 108 (2011): 20802–20807, https://doi.org/10.1073/pnas.1111149108.

48. E. T. Herruzo, H. Asakawa, T. Fukuma, and R. Garcia, "Three-Dimensional Quantitative Force Maps in Liquid With 10 Piconewton, Angstrom and Sub-Minute Resolutions," *Nanoscale* 5 (2013): 2678–2685, https://doi.org/10.1039/c2nr33051b.

49. J. G. Vilhena, A. C. Dumitru, E. T. Herruzo, et al., "Adsorption Orientations and Immunological Recognition of Antibodies on Graphene," *Nanoscale* 8 (2016): 13463–13475, https://doi.org/10.1039/c5nr07612a.

50. A. L. B. Pyne, A. Noy, K. H. S. Main, et al., "Base-Pair Resolution Analysis of the Effect of Supercoiling on DNA Flexibility and Major Groove Recognition by Triplex-Forming Oligonucleotides," *Nature Communications* 12 (2021): 1053, https://doi.org/10.1038/s41467-021-21243-y.

51. A. A. Baker, W. Helbert, J. Sugiyama, and M. J. Miles, "High-Resolution Atomic Force Microscopy of Native Valonia Cellulose I Microcrystals," *Journal of Structural Biology* 119 (1997): 129–138, https://doi.org/10.1006/jsbi.1997.3866.

52. S. Scheuring, D. J. Muller, P. Ringler, J. B. Heymann, and A. Engel, "Imaging Streptavidin 2D Crystals on Biotinylated Lipid Monolayers at High Resolution With the Atomic Force Microscope," *Journal of Microscopy* 193 (1999): 28–35, https://doi.org/10.1046/j.1365-2818.1999.00434.x.

53. C. Moller, M. Allen, V. Elings, A. Engel, and D. J. Muller, "Tapping-Mode Atomic Force Microscopy Produces Faithful High-Resolution Images of Protein Surfaces," *Biophysical Journal* 77 (1999): 1150–1158, https://doi.org/10.1016/S0006-3495(99)76966-3.

54. N. Persike, M. Pfeiffer, R. Guckenberger, M. Radmacher, and M. Fritz, "Direct Observation of Different Surface Structures on High-Resolution Images of Native Halorhodopsin1," *Journal of Molecular Biology* 310 (2001): 773–780, https://doi.org/10.1006/jmbi.2001.4782.

55. S. Scheuring, J. Seguin, S. Marco, et al., "AFM Characterization of Tilt and Intrinsic Flexibility of *Rhodobacter sphaeroides* Light Harvesting Complex 2 (LH2)," *Journal of Molecular Biology* 325 (2003): 569–580, https://doi.org/10.1016/s0022-2836(02)01241-x.

56. J. Mou, D. M. Czajkowsky, Y. Zhang, and Z. Shao, "High-Resolution Atomic-Force Microscopy of DNA: The Pitch of the Double Helix," *FEBS Letters* 371 (1995): 279–282, https://doi.org/10.1016/0014-5793(95)00906-p.

57. Z. Shao, "Probing Nanometer Structures With Atomic Force Microscopy," *News in Physiological Sciences* 14 (1999): 142–149, https://doi.org/10.1152/physiologyonline.1999.14.4.142.

58. A. San Paulo and R. Garcia, "High-Resolution Imaging of Antibodies by Tapping-Mode Atomic Force Microscopy: Attractive and Repulsive Tip-Sample Interaction Regimes," *Biophysical Journal* 78 (2000): 1599–1605, https://doi.org/10.1016/S0006-3495(00)76712-9.

59. I. A. Mastrangelo, M. Ahmed, T. Sato, et al., "High-Resolution Atomic Force Microscopy of Soluble Abeta42 Oligomers," *Journal of Molecular Biology* 358 (2006): 106–119, https://doi.org/10.1016/j.jmb.2006.01.042.

60. L. Hamon, D. Pastre, P. Dupaigne, C. Le Breton, E. Le Cam, and O. Pietrement, "High-Resolution AFM Imaging of Single-Stranded DNA-Binding (SSB) Protein – DNA Complexes," *Nucleic Acids Research* 35 (2007): e58, https://doi.org/10.1093/nar/gkm147.

61. C. Leung, A. Bestembayeva, R. Thorogate, et al., "Atomic Force Microscopy With Nanoscale Cantilevers Resolves Different Structural Conformations of the DNA Double Helix," *Nano Letters* 12 (2012): 3846–3850, https://doi.org/10.1021/nl301857p.

62. S. Ido, K. Kimura, N. Oyabu, et al., "Beyond the Helix Pitch: Direct Visualization of Native DNA in Aqueous Solution," *ACS Nano* 7 (2013): 1817–1822, https://doi.org/10.1021/nn400071n.

63. S. W. Chen, A. S. Banneville, J. M. Teulon, J. Timmins, and J. L. Pellequer, "Nanoscale Surface Structures of DNA Bound to *Deinococcus radiodurans* HU Unveiled by Atomic Force Microscopy," *Nanoscale* 12 (2020): 22628–22638, https://doi.org/10.1039/d0nr05320a.

64. H. P. Lang, M. Hegner, E. Meyer, and C. Gerber, "Nanomechanics From Atomic Resolution to Molecular Recognition Based on Atomic Force Microscopy Technology," *Nanotechnology* 13 (2002): R29–R36, https://doi.org/10.1088/0957-4484/13/5/202.

65. G. Binnig, C. F. Quate, and C. Gerber, "Atomic Force Microscope," *Physical Review Letters* 56 (1986): 930, https://doi.org/10.1103/PhysRevLett.56.930.

66. G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, "Surface Studies by Scanning Tunneling Microscopy," *Physical Review Letters* 49 (1982): 57, https://doi.org/10.1103/PhysRevLett.49.57.

67. B. Drake, C. B. Prater, A. L. Weisenhorn, et al., "Imaging Crystals, Polymers, and Processes in Water With the Atomic Force Microscope," *Science* 243 (1989): 1586–1589, https://doi.org/10.1126/science.2928794.

68. J. Vesenka, S. Manne, R. Giberson, T. Marsh, and E. Henderson, "Colloidal Gold Particles as an Incompressible Atomic Force Microscope Imaging Standard for Assessing the Compressibility of Biomolecules," *Biophysical Journal* 65 (1993): 992–997, https://doi.org/10.1016/S0006-3495(93)81171-8.

69. J. H. Hafner, C. L. Cheung, A. T. Woolley, and C. M. Lieber, "Structural and Functional Imaging With Carbon Nanotube AFM Probes," *Progress in Biophysics and Molecular Biology* 77 (2001): 73–110, https://doi.org/10.1016/s0079-6107(01)00011-6.

70. W. F. Kolbe, D. F. Ogletree, and M. B. Salmeron, "Atomic Force Microscopy Imaging of T4 Bacteriophages on Silicon Substrates," *Ultramicroscopy* 42–44 (1992): 1113–1117, https://doi.org/10.1016/0304-3991(92)90411-c.

71. J. V. Zoval, P. R. Biernacki, and R. M. Penner, "Implementation of Electrochemically Synthesized Silver Nanocrystallites for the Preferential SERS Enhancement of Defect Modes on Thermally Etched Graphite Surfaces," *Analytical Chemistry* 68 (1996): 1585–1592, https://doi.org/10.1021/ac951114+.

72. H. J. Butt, K. H. Downing, and P. K. Hansma, "Imaging the Membrane Protein Bacteriorhodopsin With the Atomic Force Microscope," *Biophysical Journal* 58 (1990): 1473–1480, https://doi.org/10.1016/S0006-3495(90)82492-9.

73. Y. L. Lyubchenko, A. A. Gall, L. S. Shlyakhtenko, et al., "Atomic Force Microscopy Imaging of Double-Stranded DNA and RNA," *Journal of Biomolecular Structure & Dynamics* 10 (1992): 589–606, https://doi.org/10.1080/07391102.1992.10508670.

74. D. P. Allison, N. P. Mortensen, C. J. Sullivan, and M. J. Doktycz, "Atomic Force Microscopy of Biological Samples," *WIREs Nanomedicine and Nanobiotechnology* 2 (2010): 618–634, https://doi.org/10.1002/wnan.104.

75. P. Eaton and P. West, *Atomic Force Microscopy* (Oxford: Oxford University Press, 2010).

76. S. Karrasch, R. Hegerl, J. H. Hoh, W. Baumeister, and A. Engel, "Atomic Force Microscopy Produces Faithful High-Resolution Images of Protein Surfaces in an Aqueous Environment," *Proceedings of the National Academy of Sciences of the United States of America* 91 (1994): 836, https://doi.org/10.1073/pnas.91.3.836.

77. A. Stylianou, S. V. Kontomaris, C. Grant, and E. Alexandratou, "Atomic Force Microscopy on Biological Materials Related to Pathological Conditions," *Scanning* 2019 (2019): 8452851, https://doi.org/10.1155/2019/8452851.

78. Z. Bednarikova, Z. Gazova, F. Valle, and E. Bystrenova, "Atomic Force Microscopy as an Imaging Tool to Study the Bio/Nonbio Complexes," *Journal of Microscopy* 280 (2020): 241–251, https://doi.org/10.1111/jmi.12936.

79. H. G. Hansma, J. Vesenka, C. Siegerist, et al., "Reproducible Imaging and Dissection of Plasmid DNA Under Liquid With the Atomic Force Microscope," *Science* 256 (1992): 1180, https://doi.org/10.1126/science.256.5060.1180.

80. C. Godon, J.-M. Teulon, M. Odorico, et al., "Conditions to Minimize Soft Single Biomolecule Deformation When Imaging With Atomic Force Microscopy," *Journal of Structural Biology* 197 (2017): 322–329, https://doi.org/10.1016/j.jsb.2016.12.011.

81. E. Schrödinger, *What Is Life?* (Cambridge, UK: Cambridge Univ. Press, 1944), https://doi.org/10.1017/CBO9781139644129.

82. J.-L. Pellequer and S.-w W. Chen, "Multi-Template Approach to Modeling Engineered Disulfide Bonds," *Proteins* 65 (2006): 192–202, https://doi.org/10.1002/prot.21059.

83. J. L. Pellequer, R. Brudler, and E. D. Getzoff, "Biological Sensors: More Than One Way to Sense Oxygen," *Current Biology* 9 (1999): R416–R418, https://doi.org/10.1016/S0960-9822(99)80257-7.

84. J. P. P. Starink and T. M. Jovin, "Background Correction in Scanning Probe Microscope Recordings of Macromolecules," *Surface Science* 359 (1996): 291–305, https://doi.org/10.1016/0039-6028(96)00367-6.

85. D. Nečas and P. Klapetek, "Gwyddion: An Open-Source Software for SPM Data Analysis," *Central European Journal of Physics* 10 (2012): 181–188, https://doi.org/10.2478/s11534-011-0096-2.

86. B. W. Erickson, S. Coquoz, J. D. Adams, D. J. Burns, and G. E. Fantner, "Large-Scale Analysis of High-Speed Atomic Force Microscopy Data Sets Using Adaptive Image Processing," *Beilstein Journal of Nanotechnology* 3 (2012): 747–758, https://doi.org/10.3762/bjnano.3.84.

87. J. G. Beton, R. Moorehead, L. Helfmann, et al., "TopoStats - A Program for Automated Tracing of Biomolecules From AFM Images," *Methods* 193 (2021): 68–79, https://doi.org/10.1016/j.ymeth.2021.01.008.

88. A. T. Winzer, C. Kraft, S. Bhushan, V. Stepanenko, and I. Tessmer, "Correcting for AFM Tip Induced Topography Convolutions in Protein-DNA Samples," *Ultramicroscopy* 121 (2012): 8–15, https://doi.org/10.1016/j.ultramic.2012.07.002.

89. P. Klapetek, M. Valtr, D. Necas, O. Salyk, and P. Dzik, "Atomic Force Microscopy Analysis of Nanoparticles in Non-Ideal Conditions," *Nanoscale Research Letters* 6 (2011): 514, https://doi.org/10.1186/1556-276X-6-514.

90. M.-H. Trinh, M. Odorico, L. Bellanger, M. Jacquemond, P. Parot, and J.-L. Pellequer, "Tobacco Mosaic Virus as an AFM Tip Calibrator," *Journal of Molecular Recognition* 24 (2011): 503–510, https://doi.org/10.1002/jmr.1118.

91. K. Onishi and D. Fujita, "Novel Tip Shape Reconstruction Method for Restoration of AFM Topography Images Using Nano-Structures With Given Shapes," *Analytical Sciences* 27 (2011): 157, https://doi.org/10.2116/analsci.27.157.

92. C. Fleischmann, K. Paredis, D. Melkonyan, and W. Vandervorst, "Revealing the 3-Dimensional Shape of Atom Probe Tips by Atomic Force Microscopy," *Ultramicroscopy* 194 (2018): 221–226, https://doi.org/10.1016/j.ultramic.2018.08.010.

93. Z. Chen, J. Luo, I. Doudevski, S. Erten, and S. H. Kim, "Atomic Force Microscopy (AFM) Analysis of an Object Larger and Sharper Than the AFM Tip," *Microscopy and Microanalysis* 25 (2019): 1106–1111, https://doi.org/10.1017/S1431927619014697.

94. L. Lutter, C. J. Serpell, M. F. Tuite, L. C. Serpell, and W. F. Xue, "Three-Dimensional Reconstruction of Individual Helical Nano-Filament Structures From Atomic Force Microscopy Topographs," *Biomolecular Concepts* 11 (2020): 102–115, https://doi.org/10.1515/bmc-2020-0009.

95. J. S. Villarrubia, "Algorithms for Scanned Probe Microscope Image Simulation, Surface Reconstruction, and Tip Estimation," *Journal of Research of the National Institute of Standards and Technology* 102 (1997): 425–454, https://doi.org/10.6028/jres.102.030.

96. X. Qian and J. S. Villarrubia, "General Three-Dimensional Image Simulation and Surface Reconstruction in Scanning Probe Microscopy Using a Dexel Representation," *Ultramicroscopy* 108 (2007): 29–42, https://doi.org/10.1016/j.ultramic.2007.02.031.

97. L. K. S. Bonagiri, Z. Wang, S. Zhou, and Y. Zhang, "Precise Surface Profiling at the Nanoscale Enabled by Deep Learning," *Nano Letters* 24 (2024): 2589–2595, https://doi.org/10.1021/acs.nanolett.3c04712.

98. J. Schwartz, Y. Jiang, Y. Wang, et al., "Removing Stripes, Scratches, and Curtaining With Nonrecoverable Compressed Sensing," *Microscopy and Microanalysis* 25 (2019): 705–710, https://doi.org/10.1017/S1431927619000254.

99. M. Bouali and S. Ladjal, "Toward Optimal Destriping of MODIS Data Using a Unidirectional Variational Model," *IEEE Transactions on Geoscience and Remote Sensing* 49 (2011): 2924–2935, https://doi.org/10.1109/Tgrs.2011.2119399.

100. M. Schimmack and P. Mercorelli, "A Wavelet Packet Tree Denoising Algorithm for Images of Atomic-Force Microscopy," *Asian Journal of Control* 20 (2018): 1367–1378, https://doi.org/10.1002/asjc.1718.

101. L. Massimi, F. Brun, M. Fratini, I. Bukreeva, and A. Cedola, "An Improved Ring Removal Procedure for In-Line X-Ray Phase Contrast Tomography," *Physics in Medicine and Biology* 63 (2018): 045007, https://doi.org/10.1088/1361-6560/aaa706.

102. F. Golek, P. Mazur, Z. Ryszka, and S. Zuber, "AFM Image Artifacts," *Applied Surface Science* 304 (2014): 11–19, https://doi.org/10.1016/j.apsusc.2014.01.149.

103. B. Münch, P. Trtik, F. Marone, and M. Stampanoni, "Stripe and Ring Artifact Removal With Combined Wavelet – Fourier Filtering," *Optics Express* 17 (2009): 8567–8591, https://doi.org/10.1364/Oe.17.008567.

104. H. Wendt, N. Dobigeon, J. Y. Tourneret, M. Albinet, C. Goldstein, and N. Karouche, "Detection and Correction of Glitches in a Multiplexed Multichannel Data Stream-Application to the MADRAS Instrument," *IEEE Transactions on Geoscience and Remote Sensing* 54 (2016): 2803–2811, https://doi.org/10.1109/Tgrs.2015.2505902.

105. S.-w W. Chen and J.-L. Pellequer, "DeStripe: Frequency-Based Algorithm for Removing Stripe Noises From AFM Images," *BMC Structural Biology* 11 (2011): 7, https://doi.org/10.1186/1472-6807-11-7.

106. A. Pollatou, "An Automated Method for Removal of Striping Artifacts in Fluorescent Whole-Slide Microscopy," *Journal of Neuroscience Methods* 341 (2020): 108781, https://doi.org/10.1016/j.jneumeth.2020.108781.

107. M. Li, J. Rieck, B. Noheda, J. Roerdink, and M. H. F. Wilkinson, "Stripe Noise Removal in Conductive Atomic Force Microscopy," *Scientific Reports* 14 (2024): 3931, https://doi.org/10.1038/s41598-024-54094-w.

108. S.-w W. Chen, J.-M. Teulon, C. Godon, and J.-L. Pellequer, "Atomic Force Microscope, Molecular Imaging, and Analysis," *Journal of Molecular Recognition* 29 (2016): 51–55, https://doi.org/10.1002/jmr.2491.

109. S.-w W. Chen, M. Odorico, M. Meillan, et al., "Nanoscale Structural Features Determined by AFM for Single Virus Particles," *Nanoscale* 5 (2013): 10877–10886, https://doi.org/10.1039/C3NR02706F.

110. S. F. Konrad, W. Vanderlinden, and J. Lipfert, "Quantifying Epigenetic Modulation of Nucleosome Breathing by High-Throughput AFM Imaging," *Biophysical Journal* 121 (2022): 841–851, https://doi.org/10.1016/j.bpj.2022.01.014.

111. G. Q. Han and B. Lin, "Optimal Sampling and Reconstruction of Undersampled Atomic Force Microscope Images Using Compressive Sensing," *Ultramicroscopy* 189 (2018): 85–94, https://doi.org/10.1016/j.ultramic.2018.03.019.

112. A. Malki, J.-M. Teulon, A. Camacho Zarco, et al., "Intrinsically Disordered Tardigrade Proteins Self-Assemble Into Fibrous Gels in Response to Environmental Stress," *Angewandte Chemie, International Edition* 61 (2022): e202109961, https://doi.org/10.1002/anie.202109961.

113. R. C. Chaves and J.-L. Pellequer, "DockAFM: Benchmarking Protein Structures by Docking Under AFM Topographs," *Bioinformatics* 29 (2013): 3230–3231, https://doi.org/10.1093/bioinformatics/btt561.

114. J. G. Mandell, V. A. Roberts, M. E. Pique, et al., "Protein Docking Using Continuum Electrostatics and Geometric Fit," *Protein Engineering* 14 (2001): 105, https://doi.org/10.1093/protein/14.2.105.

115. H. M. Berman, J. Westbrook, Z. Feng, et al., "The Protein Data Bank," *Nucleic Acids Research* 28 (2000): 235, https://doi.org/10.1093/nar/28.1.235.

116. S.-w W. Chen and J.-L. Pellequer, "Adepth: New Representation and Its Implications for Atomic Depths of Macromolecules," *Nucleic Acids Research* 41 (2013): W412–W416, https://doi.org/10.1093/nar/gkt299.

117. V. A. Roberts, E. E. Thompson, M. E. Pique, M. S. Perez, and L. F. Ten Eyck, "DOT2: Macromolecular Docking With Improved Biophysical Models," *Journal of Computational Chemistry* 34 (2013): 1743–1758, https://doi.org/10.1002/jcc.23304.

118. Y. Kunioka and T. Ando, "Innocuous Labeling of the Subfragment-2 Region of Skeletal Muscle Heavy Meromyosin With a Fluorescent Polyacrylamide Nanobead and Visualization of Individual Heavy Meromyosin Molecules," *Journal of Biochemistry* 119 (1996): 1024–1032, https://doi.org/10.1093/oxfordjournals.jbchem.a021343.

119. H. J. Butt, E. K. Wolff, S. A. C. Gould, B. D. Northern, C. M. Peterson, and P. K. Hansma, "Imaging Cells With the Atomic Force Microscope," *Journal of Structural Biology* 105 (1990): 54–61, https://doi.org/10.1016/1047-8477(90)90098-W.

120. T. Ushiki, M. Shigeno, and K. Abe, "Atomic Force Microscopy of Embedment-Free Sections of Cells and Tissues," *Archives of Histology and Cytology* 57 (1994): 427, https://doi.org/10.1679/aohc.57.427.

121. J. R. Lopéz-Blanco and P. Chacón, "iMODFIT: Efficient and Robust Flexible Fitting Based on Vibrational Analysis in Internal Coordinates," *Journal of Structural Biology* 184 (2013): 261–270, https://doi.org/10.1016/j.jsb.2013.08.010.

122. J. R. López-Blanco and P. Chacón, "Structural Modeling From Electron Microscopy Data," *WIREs Computational Molecular Science* 5 (2015): 62–81, https://doi.org/10.1002/wcms.1199.

123. W. Wriggers, R. A. Milligan, and J. A. McCammon, "Situs: A Package for Docking Crystal Structures Into Low-Resolution Maps From Electron Microscopy," *Journal of Structural Biology* 125 (1999): 185–195, https://doi.org/10.1006/jsbi.1998.4080.

124. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, et al., "The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures," *Journal of Molecular Biology* 112 (1977): 535–542, https://doi.org/10.1016/S0022-2836(77)80200-3.

125. H. Berman, K. Henrick, and H. Nakamura, "Announcing the Worldwide Protein Data Bank," *Nature Structural Biology* 10 (2003): 980, https://doi.org/10.1038/nsb1203-980.

126. P. E. Bourne, H. M. Berman, K. Watenpaugh, J. D. Westbrook, and P. M. D. Fitzgerald, "The Macromolecular Crystallographic Information File (mmCIF)," *Methods in Enzymology* 277 (1997): 571–590, https://doi.org/10.1016/S0076-6879(97)77032-0.

127. M. Sandal, F. Benedetti, M. Brucale, A. Gomez-Casado, and B. Samorì, "Hooke: An Open Software Platform for Force Spectroscopy," *Bioinformatics* 25 (2009): 1428–1430, https://doi.org/10.1093/bioinformatics/btp180.

128. J. te Riet, A. J. Katan, C. Rankl, et al., "Interlaboratory Round Robin on Cantilever Calibration for AFM Force Spectroscopy," *Ultramicroscopy* 111 (2011): 1659–1669, https://doi.org/10.1016/j.ultramic.2011.09.012.

129. J. E. Sader, R. Borgani, C. T. Gibson, et al., "A Virtual Instrument to Standardise the Calibration of Atomic Force Microscope Cantilevers," *Review of Scientific Instruments* 87 (2016): 093711, https://doi.org/10.1063/1.4962866.

130. H. Schillers, C. Rianna, J. Schäpe, et al., "Standardized Nanomechanical Atomic Force Microscopy Procedure (SNAP) for Measuring Soft and Biological Samples," *Scientific Reports* 7 (2017): 5117, https://doi.org/10.1038/s41598-017-05383-0.

131. J. López-Alonso, M. Eroles, S. Janel, et al., "PyFMLab: Open-Source Software for Atomic Force Microscopy Microrheology Data Analysis," *Open Research Europe* 3 (2024): 187, https://doi.org/10.12688/openreseurope.16550.2.

132. P. E. Bourne, J. Westbrook, and H. M. Berman, "The Protein Data Bank and Lessons in Data Management," *Briefings in Bioinformatics* 5 (2004): 23–30, https://doi.org/10.1093/bib/5.1.23.

133. C. L. Lawson, M. L. Baker, C. Best, et al., "EMDataBank.org: Unified Data Resource for CryoEM," *Nucleic Acids Research* 39 (2011): D456–D464, https://doi.org/10.1093/nar/gkq880.

134. E. Valentini, A. G. Kikhney, G. Previtali, C. M. Jeffries, and D. I. Svergun, "SASBDB, a Repository for Biological Small-Angle Scattering Data," *Nucleic Acids Research* 43 (2015): D357–D363, https://doi.org/10.1093/nar/gku1047.

135. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. (New Jersey: Pearson Education, Inc., 2008).

136. J. Sotres, H. Boyd, and J. F. Gonzalez-Martinez, "Enabling Autonomous Scanning Probe Microscopy Imaging of Single Molecules With Deep Learning," *Nanoscale* 13 (2021): 9193–9203, https://doi.org/10.1039/d1nr01109j.

137. A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, "SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures," *Journal of Molecular Biology* 247 (1995): 536–540, https://doi.org/10.1006/jmbi.1995.0159.

138. C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton, "CATH – A Hierarchic Classification of Protein Domain Structures," *Structure* 5 (1997): 1093–1108, https://doi.org/10.1016/S0969-2126(97)00260-8.

139. J. M. Chandonia, N. K. Fox, and S. E. Brenner, "SCOPe: Manual Curation and Artifact Removal in the Structural Classification of Proteins – Extended Database," *Journal of Molecular Biology* 429 (2017): 348–355, https://doi.org/10.1016/j.jmb.2016.11.023.

140. H. Cheng, R. D. Schaeffer, Y. Liao, et al., "ECOD: An Evolutionary Classification of Protein Domains," *PLoS Computational Biology* 10 (2014): e1003926, https://doi.org/10.1371/journal.pcbi.1003926.

141. J. Wells, A. Hawkins-Hooker, N. Bordin, B. Paige, and C. Orengo, "Chainsaw: Protein Domain Segmentation With Fully Convolutional Neural Networks," *Bioinformatics* 40 (2024): 296, https://doi.org/10.1093/bioinformatics/btae296.

142. F. Cazals, J. Herrmann, and E. Sarti, "Simpler Protein Domain Identification Using Spectral Clustering," *bioRxiv* (2024), https://doi.org/10.1101/2024.02.10.579762.

143. L. Ponzoni, G. Polles, V. Carnevale, and C. Micheletti, "SPECTRUS: A Dimensionality Reduction Approach for Identifying Dynamical Domains in Protein Complexes From Limited Structural Datasets," *Structure* 23 (2015): 1516–1525, https://doi.org/10.1016/j.str.2015.05.022.

144. E. L. Sonnhammer, S. R. Eddy, and R. Durbin, "Pfam: A Comprehensive Database of Protein Domain Families Based on Seed Alignments," *Proteins* 28 (1997): 405–420, https://doi.org/10.1002/(sici)1097-0134(199707)28:3<405::aid-prot10>3.0.co;2-l.

145. R. Day, D. A. Beck, R. S. Armen, and V. Daggett, "A Consensus View of Fold Space: Combining SCOP, CATH, and the Dali Domain Dictionary," *Protein Science* 12 (2003): 2150–2160, https://doi.org/10.1110/ps.0306803.

146. M. L. Tress, I. Ezkurdia, and J. S. Richardson, "Target Domain Definition and Classification in CASP8," *Proteins* 77 (2009): 10–17, https://doi.org/10.1002/prot.22497.

147. R. K. Sistla, K. V. Brinda, and S. Vishveshwara, "Identification of Domains and Domain Interface Residues in Multidomain Proteins From Graph Spectral Method," *Proteins* 59 (2005): 616–626, https://doi.org/10.1002/prot.20444.

148. L. Heo, C. F. Arbour, G. Janson, and M. Feig, "Improved Sampling Strategies for Protein Model Refinement Based on Molecular Dynamics Simulation," *Journal of Chemical Theory and Computation* 17 (2021): 1931–1943, https://doi.org/10.1021/acs.jctc.0c01238.

149. P. Minary and M. Levitt, "Probing Protein Fold Space With a Simplified Model," *Journal of Molecular Biology* 375 (2008): 920–933, https://doi.org/10.1016/j.jmb.2007.10.087.

150. B. L. de Groot, D. M. van Aalten, R. M. Scheek, A. Amadei, G. Vriend, and H. J. Berendsen, "Prediction of Protein Conformational Freedom From Distance Constraints," *Proteins* 29 (1997): 240–251, https://doi.org/10.1002/(sici)1097-0134(199710)29:2<240::aid-prot11>3.0.co;2-o.

151. N. Go, T. Noguti, and T. Nishikawa, "Dynamics of a Small Globular Protein in Terms of Low-Frequency Vibrational-Modes," *Proceedings of the National Academy of Sciences of the United States of America* 80 (1983): 3696–3700, https://doi.org/10.1073/pnas.80.12.3696.

152. M. Levitt, C. Sander, and P. S. Stern, "Protein Normal-Mode Dynamics – Trypsin-Inhibitor, Crambin, Ribonuclease and Lysozyme," *Journal of Molecular Biology* 181 (1985): 423–447, https://doi.org/10.1016/0022-2836(85)90230-X.

153. F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand, "Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules," *Proteins* 41 (2000): 1–7, https://doi.org/10.1002/1097-0134(20001001)41:1<1::aid-prot10>3.0.co;2-p.

154. A. Hoffmann and S. Grudinin, "NOLB: Nonlinear Rigid Block Normal-Mode Analysis Method," *Journal of Chemical Theory and Computation* 13 (2017): 2123–2134, https://doi.org/10.1021/acs.jctc.7b00197.

155. E. Neveu, P. Popov, A. Hoffmann, et al., "RapidRMSD: Rapid Determination of RMSDs Corresponding to Motions of Flexible Molecules," *Bioinformatics* 34 (2018): 2757–2765, https://doi.org/10.1093/bioinformatics/bty160.

156. E. Laine and S. Grudinin, "HOPMA: Boosting Protein Functional Dynamics With Colored Contact Maps," *Journal of Physical Chemistry. B* 125 (2021): 2577–2588, https://doi.org/10.1021/acs.jpcb.0c11633.

157. V. Lombard, S. Grudinin, and E. Laine, "Explaining Conformational Diversity in Protein Families Through Molecular Motions," *Scientific Data* 11 (2024): 752, https://doi.org/10.1038/s41597-024-03524-5.

158. T. Ando, N. Kodera, E. Takai, D. Maruyama, K. Saito, and A. Toda, "A High-Speed Atomic Force Microscope for Studying Biological Macromolecules," *Proceedings of the National Academy of Sciences of the United States of America* 98 (2001): 12468–12472, https://doi.org/10.1073/pnas.211400898.

159. T. Ando, T. Uchihashi, N. Kodera, et al., "High-Speed Atomic Force Microscopy for Observing Dynamic Biomolecular Processes," *Journal of Molecular Recognition* 20 (2007): 448–458, https://doi.org/10.1002/jmr.843.

160. M. B. Viani, T. E. Schäffer, G. T. Paloczi, et al., "Fast Imaging and Fast Force Spectroscopy of Single Biopolymers With a New Atomic Force Microscope Designed for Small Cantilevers," *Review of Scientific Instruments* 70 (1999): 4300–4303, https://doi.org/10.1063/1.1150069.

161. D. Yamamoto, N. Nagura, S. Omote, M. Taniguchi, and T. Ando, "Streptavidin 2D Crystal Substrates for Visualizing Biomolecular Processes by Atomic Force Microscopy," *Biophysical Journal* 97 (2009): 2358–2367, https://doi.org/10.1016/j.bpj.2009.07.046.

162. A. Miyagi, Y. Tsunaka, T. Uchihashi, et al., "Visualization of Intrinsically Disordered Regions of Proteins by High-Speed Atomic Force Microscopy," *ChemPhysChem* 9 (2008): 1859–1866, https://doi.org/10.1002/cphc.200800210.

163. N. Kodera, D. Yamamoto, R. Ishikawa, and T. Ando, "Video Imaging of Walking Myosin V by High-Speed Atomic Force Microscopy," *Nature* 468 (2010): 72–76, https://doi.org/10.1038/nature09450.

164. P. E. Milhiet, D. Yamamoto, O. Berthoumieu, et al., "Deciphering the Structure, Growth and Assembly of Amyloid-Like Fibrils Using High-Speed Atomic Force Microscopy," *PLoS One* 5 (2010): e13240, https://doi.org/10.1371/journal.pone.0013240.

165. T. Uchihashi, R. Iino, T. Ando, and H. Noji, "High-Speed Atomic Force Microscopy Reveals Rotary Catalysis of Rotorless F(1)-ATPase," *Science* 333 (2011): 755–758, https://doi.org/10.1126/science.1205510.

166. I. Casuso, J. Khao, M. Chami, et al., "Characterization of the Motion of Membrane Proteins Using High-Speed Atomic Force Microscopy," *Nature Nanotechnology* 7 (2012): 525–529, https://doi.org/10.1038/nnano.2012.109.

167. N. Chiaruttini, L. Redondo-Morata, A. Colom, et al., "Relaxation of Loaded ESCRT-III Spiral Springs Drives Membrane Deformation," *Cell* 163 (2015): 866–879, https://doi.org/10.1016/j.cell.2015.10.017.

168. G. R. Heath and S. Scheuring, "Advances in High-Speed Atomic Force Microscopy (HS-AFM) Reveal Dynamics of Transmembrane Channels and Transporters," *Current Opinion in Structural Biology* 57 (2019): 93–102, https://doi.org/10.1016/j.sbi.2019.02.008.

169. I. Casuso, L. Redondo-Morata, and F. Rico, "Biological Physics by High-Speed Atomic Force Microscopy," *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 378 (2020): 20190604, https://doi.org/10.1098/rsta.2019.0604.

170. F. Rico, L. Gonzalez, I. Casuso, M. Puig-Vidal, and S. Scheuring, "High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations," *Science* 342 (2013): 741–743, https://doi.org/10.1126/science.1239764.

171. F. Sumbul, A. Marchesi, H. Takahashi, S. Scheuring, and F. Rico, "High-Speed Force Spectroscopy for Single Protein Unfolding," *Methods in Molecular Biology* 1814 (2018): 243–264, https://doi.org/10.1007/978-1-4939-8591-3_15.

172. H. Flechsig and T. Ando, "Protein Dynamics by the Combination of High-Speed AFM and Computational Modeling," *Current Opinion in Structural Biology* 80 (2023): 102591, https://doi.org/10.1016/j.sbi.2023.102591.

173. S. Kato, S. Takada, and S. Fuchigami, "Particle Smoother to Assimilate Asynchronous Movie Data of High-Speed AFM With MD Simulations," *Journal of Chemical Theory and Computation* 19 (2023): 4678–4688, https://doi.org/10.1021/acs.jctc.2c01268.

174. G. R. Heath, E. Kots, J. L. Robertson, et al., "Localization Atomic Force Microscopy," *Nature* 594 (2021): 385–390, https://doi.org/10.1038/s41586-021-03551-x.

175. G. R. Heath, E. Micklethwaite, and T. M. Storer, "NanoLocz: Image Analysis Platform for AFM, High-Speed AFM, and Localization AFM," *Small Methods* (2024): e2301766, https://doi.org/10.1002/smtd.202301766.

176. G. Wang and R. L. Dunbrack, Jr., "PISCES: A Protein Sequence Culling Server," *Bioinformatics* 19 (2003): 1589–1591, https://doi.org/10.1093/bioinformatics/btg224.

177. A. Bairoch, R. Apweiler, C. H. Wu, et al., "The Universal Protein Resource (UniProt)," *Nucleic Acids Research* 33 (2005): D154–D159, https://doi.org/10.1093/nar/gki070.

178. W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual Molecular Dynamics," *Journal of Molecular Graphics* 14 (1996): 33–38, https://doi.org/10.1016/0263-7855(96)00018-5.

179. E. A. Merritt and D. J. Bacon, "Raster3D: Photorealistic Molecular Graphics," *Methods in Enzymology* 277 (1997): 505–524, https://doi.org/10.1016/s0076-6879(97)77028-9.