



OPEN

End-to-end differentiable blind tip reconstruction for noisy atomic force microscopy images

Yasuhiro Matsunaga^{1✉}, Sotaro Fuchigami², Tomonori Ogane¹ & Shoji Takada²

Observing the structural dynamics of biomolecules is vital to deepening our understanding of biomolecular functions. High-speed (HS) atomic force microscopy (AFM) is a powerful method to measure biomolecular behavior at near physiological conditions. In the AFM, measured image profiles on a molecular surface are distorted by the tip shape through the interactions between the tip and molecule. Once the tip shape is known, AFM images can be approximately deconvolved to reconstruct the surface geometry of the sample molecule. Thus, knowing the correct tip shape is an important issue in the AFM image analysis. The blind tip reconstruction (BTR) method developed by Villarrubia (J Res Natl Inst Stand Technol 102:425, 1997) is an algorithm that estimates tip shape only from AFM images using mathematical morphology operators. While the BTR works perfectly for noise-free AFM images, the algorithm is susceptible to noise. To overcome this issue, we here propose an alternative BTR method, called *end-to-end differentiable BTR*, based on a modern machine learning approach. In the method, we introduce a loss function including a regularization term to prevent overfitting to noise, and the tip shape is optimized with automatic differentiation and backpropagations developed in deep learning frameworks. Using noisy pseudo-AFM images of myosin V motor domain as test cases, we show that our end-to-end differentiable BTR is robust against noise in AFM images. The method can also detect a double-tip shape and deconvolve doubled molecular images. Finally, application to real HS-AFM data of myosin V walking on an actin filament shows that the method can reconstruct the accurate surface geometry of actomyosin consistent with the structural model. Our method serves as a general post-processing for reconstructing hidden molecular surfaces from any AFM images. Codes are available at https://github.com/matsunagalab/differentiable_BTR.

Atomic force microscopy (AFM) is a unique technique for imaging structures of sample molecules bound to a surface at ambient conditions¹. Recently, high-speed AFM (HS-AFM) with dramatically faster imaging rates (up to tens of frames per second) has been developed, enabling us to directly observe biomolecules in action^{2,3}. HS-AFM can investigate detailed structure–function relationships in biomolecules that cannot be observed with other methods, and it has been established as one of indispensable techniques in modern biophysics. Example applications of HS-AFM include myosin V walking along an actin filament⁴, rotary catalysis of F₁-ATPase⁵, structural dynamics of intrinsically disordered protein⁶, and the functional dynamics of CRISPR-Cas9 in action⁷. Currently, the spatial resolutions of HS-AFM instruments are ~2 nm in the lateral direction and ~0.15 nm in the vertical direction to the AFM stage⁸.

Importantly, separately from the resolution of the image profile, the resolutions of the obtained sample surface information are further limited by the tip geometry and the tip-sample interactions. The relationship among the tip geometry, the image profile, and the sample surface is shown in Fig. 1a and b. When the tip is sufficiently thin, the obtained image profile is nearly the same as the sample surface (Fig. 1a). On the other hand, when the tip is blunt compared to the scale of samples, the image profile is blurred from the sample surface (Fig. 1b). Once the tip shape is known, algorithms, called *erosion*, have been proposed to “deconvolve” the image profile for reconstructing approximate sample surface geometry^{9–11}. Thus, to reconstruct the surface geometry of the sample molecule, it is crucial to know the tip shape accurately. Tip shape estimation is also important for inferring 3D molecular structures from AFM images. In the recent analysis of AFM images, pseudo-AFM images are emulated from 3D molecular structures (obtained with different experimental or computational techniques, e.g., X-ray crystallography or molecular dynamics simulations) and an assumed tip shape and then compared with

¹Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan. ²Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan. ✉email: ymatsunaga@mail.saitama-u.ac.jp

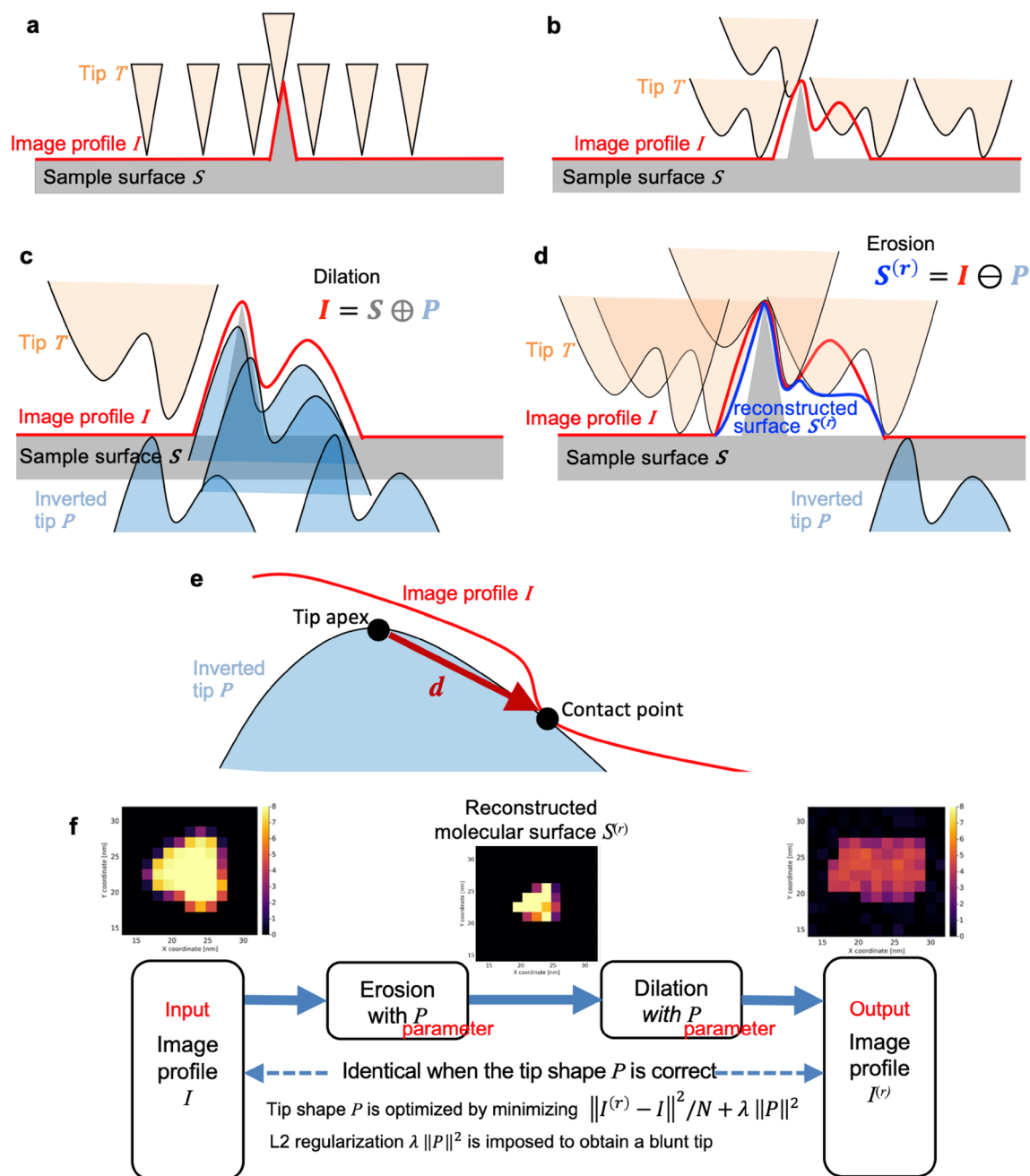


Figure 1. Morphology operators and schematic of end-to-end differentiable blind tip reconstruction (BTR). (a) Image profile obtained by scanning a surface with a thin tip. (b) Image profile obtained by scanning the same surface as (a) with a blunt tip. (c) Schematic picture illustrating a morphology operator, *dilation*. (d) Schematic picture illustrating a morphology operator, *erosion*. Here, for the sake of intuitive explanation, erosion is represented with a non-inverted tip instead of its inversion. (e) Relation of a contact point and tip apex position, explaining the idea behind the original BTR. (f) Schematic of end-to-end differentiable BTR.

the experimental AFM image^{12–21}. In the analysis, the 3D structure which generates the pseudo-AFM images most similar to the experimental AFM image is selected as the best estimate for 3D molecular structure. In this kind of analysis, the accuracy of pseudo-AFM images crucially depends on the tip shape. Thus, the tip shape should be determined as accurately as possible.

Currently, there are three possible approaches to obtain a tip shape^{22,23}. The first approach is the direct imaging of a tip using either a scanning or transmission electron microscope (SEM or TEM). However, both SEM and TEM provide only a two-dimensional projection of the sample. It is unrealistic to install special equipment for imaging the tip from various angles as a routine for determining the tip shape. Moreover, as the tip can be damaged over time during AFM measurement, determining tip shape before or after the AFM measurement is not necessarily appropriate. The second approach is to estimate the tip shape during the AFM measurement. This approach determines the tip shape from samples whose geometry is a priori known^{24–26}. With the combination

of mathematical modeling, an approximate tip shape can be reconstructed. For example, Niina et al. recently proposed a method to estimate a tip shape by comparing the pseudo-AFM images generated from 3D molecular structures with AFM images²⁰. By assuming that the tip shape is a hemisphere (radius r) combined with a circular frustum of a cone (half angle θ), they proposed to estimate these two geometric parameters. However, since real tips can be of any shape, e.g., double-tip shape with two separate acutes, the assumption of a shape limits the application of the method.

The third approach estimates the tip shape through the image analysis of AFM data without any prior knowledge of sample molecule and tip shape. The blind tip reconstruction⁹ developed by Villarrubia is a classical method to estimate arbitrary tip shapes from AFM images. The idea behind the algorithm is to recognize that sizes and depths of dents in the image profile must be smoother than the sharpness of the tip. Then, the algorithm “carves” an initially blunt tip according to the sizes and depths of dents (details will be described below). While the BTR works perfectly for noise-free AFM images, its algorithm is susceptible to noise, or it is difficult to determine a threshold parameter against noise. This is a crucial issue for the analysis of HS-AFM images because HS-AFM is generally more prone to noise than conventional AFM. Over these two decades, several improvements or new methods have been proposed for better estimation of tip shape; Dongmo et al. proposed monitoring tip volume for tuning *thresh* parameter for noisy AFM images. Tian et al. extended Villarrubia’s original BTR with the dixel representation for reconstructing general 3D tip shapes and proposed an improved regularization scheme against noise²⁷. Józwiak et al. simplified Tian’s idea to the case of standard AFM tips²⁸. Flater et al. proposed a systematic way to determine the threshold parameter against noisy AFM images²². By approximating mathematical morphology operators by linear operators, Bakucz et al. proposed a reconstruction method based on the Expectation–Maximization (EM) algorithm with the tip shape represented as a hidden variable²⁹. Despite these studies, the original BTR is not routinely utilized in the analysis of noisy AFM data, especially for the analysis of HS-AFM data, due to its susceptibility to noise and difficulty in tuning the parameter.

The reason why Villarrubia’s BTR is susceptible to noise is that it is difficult to correctly determine whether the tip shape geometry or noise causes an individual dent in the image profile. In the algorithm of the BTR, once a dent caused by noise is misinterpreted as the cause of the geometry of the tip, the tip is “carved” to be fitted to the noise. This leads to a reconstruction of very thin tip shape. In terms of machine learning theory, this can be regarded as overfitting the tip shape to noise.

To prevent such overfitting, we here introduce an appropriate loss function including a regularization term considering noise statistics, which are absent in the original BTR. As the morphology operators (Fig. 1c,d) used in the loss function are complicated nonlinear functions, minimizing such a complicated loss function over tip shape is challenging. The technologies developed by recent advances in deep learning studies³⁰ can potentially overcome this problem. Automatic differentiation and smart optimizers are recently being applied to optimize not only the parameters of neural networks but also the parameters of physical models, by implementing physical functions as differentiable functions^{31,32}. For example, Zhou et al. recently modeled the Lorentz TEM observation process as a differentiable neural network layer and successfully solved the inverse problem of phase retrieval stably by backpropagation³³. In the current study, we propose to use these technologies to optimize the loss function over the tip shape. Our method (called the end-to-end differentiable BTR) implements morphology operators as differentiable functions and optimizes the loss function in the same way as neural networks under a deep learning framework. Using pseudo-AFM images generated from a known molecular structure (myosin V motor domain) as a test case, we show that the differentiable BTR is robust against noise in AFM images, as well as lower parameter dependence compared to the original BTR. Furthermore, the differentiable BTR can correctly detect double tips, one of the artifacts that frequently occur in AFM measurements. Finally, we applied the method to real noisy HS-AFM images, myosin V walking along an actin filament.

Results

We briefly review the original BTR by Villarrubia⁹, then introduce the end-to-end differentiable BTR. First, let the AFM stage be the xy -plane, and the height perpendicular to the stage be z -coordinate, where the stage position corresponds to $z = 0$. Then, the height of the sample surface at coordinates (x, y) is denoted by $s(x, y)$ (Fig. 1a,b). The image profile obtained by scanning the sample surface with a tip is denoted by $i(x, y)$. Then, the coordinate system for tip shape is introduced. We define the uv -plane parallel to the xy -plane, and its origin is located at the top of the tip apex, i.e., the tip apex corresponds to $z = 0$ in the uv -coordinate system. The uv -coordinate system is a moving coordinate that moves with the tip. The height of the tip surface is represented by $t(u, v) \geq 0$, which determines the tip shape. For convenience, let us consider an inversion of the tip shape through the origin and define the surface of the inverted tip as $p(u, v) = -t(-u, -v) \leq 0$. Then, the image profile $i(x, y)$ can be obtained using *dilation*, which is an operator in the mathematical morphology⁹,

$$i(x, y) = \max_{u,v} [s(x - u, y - v) + p(u, v)]$$

Conceptually, this would be viewed as a “convolution” of the molecular surface $s(x, y)$ with a kernel function $p(u, v)$ (the inverted tip shape) (Fig. 1c). But note that the dilation is not the same as convolution because convolution is a linear transformation while dilation is a nonlinear transformation, representing the physical interaction between tip and sample. Following the terminology of mathematical morphology, let us denote the sets of all points on or below the surfaces as *umbras*, and three umbras for surface $s(x, y)$, $i(x, y)$, and $p(x, y)$ defined by $S = \{(x, y, z) | z \leq s(x, y)\}$, $I = \{(x, y, z) | z \leq i(x, y)\}$, and $P = \{(u, v, z) | z \leq p(u, v)\}$, respectively. Then the dilation can be simply written as $I = S \oplus P$. Given the umbras I of a measured image profile $i(x, y)$ and an inverted tip shape P , the sample surface S can be approximately reconstructed by *erosion*, written as $S^{(r)} = I \ominus P$, which is another operator in the mathematical morphology defined by,

$$s^{(r)}(x, y) = \min_{u, v} [i(x + u, y + v) + t(u, v)] = \min_{u, v} [i(x + u, y + v) - p(u, v)]$$

Erosion “carves” dents in the image profile $i(x, y)$ that are smaller than the tip shape $p(u, v)$. Conceptually, this would be viewed as a “deconvolution” that “carves” the image profile with a kernel function $p(u, v)$ (Fig. 1d). But note again that the erosion is a nonlinear transformation. From the properties of the erosion, we can say that $S^{(r)} \oplus P = (I \ominus P) \oplus P = I$ and $S^{(r)} \supseteq S$, which means that $S^{(r)}$ is the (least) upper bound on sample surfaces that reproduces the image profile I with the tip shape P . The dilation of the erosion of A by B is called *opening*, and is written as $(A \ominus B) \oplus B = A \circ B$, so the above relation can be written simply as $I \circ P = I$. Note that the relation $I \circ P = I$ does not always hold for any P . Instead, the relation holds only for the correct tip shape P that generated I or very thin shapes like the δ -function. In general, opening removes dents smaller than tip shape P because erosion “carves” such minor dents while the subsequent dilation cannot reconstruct them. Thus, $I \circ P = I$ does not hold, e.g., for extremely blunt tips.

The algorithm of Villarrubia’s BTR (hereafter called the *original* BTR) estimates an upper bound on the tip shape from AFM images. The original BTR starts from the relation, $I \circ P = I$, and constructs an upper bound condition on P from this relation. An important property of opening is $I \circ P = \bigcup \{P + \mathbf{a} | P + \mathbf{a} \subset I\}$ where \mathbf{a} is a three-dimensional translation³⁴. Combined with the relation $I \circ P = I$, we can say $I = \bigcup \{P + \mathbf{a} | P + \mathbf{a} \subset I\}$. This means that every point of I is contained in one or more translates of P while the translates are limited by the requirement that no part of the translated inverted tip extends above the surface of I . In particular, every point on the surface of I must be touched by the surface of one or more of these translates of P . Suppose that the translated inverted tip touches the surface of I at $\mathbf{c} \in I$, and the vector from the origin (tip apex) of the moving uv -coordinates to \mathbf{c} is defined as \mathbf{d} (Fig. 1e). Then, the translated inverted tip can be written as $P + \mathbf{c} - \mathbf{d}$, and the upper bound condition on the inverted tip shape is that the translated inverted tip is entirely on or below the surface of I , that is,

$$P \subseteq I - \mathbf{c} + \mathbf{d}$$

This condition holds for any \mathbf{c} on the surface or inside of I where \mathbf{d} exists.

The original BTR finds P that satisfies this upper bound condition. The algorithm starts from a thick tip shape (typically, a square pillar with a flat top), as an initial condition, that includes the true P as a subset. When the tip touches on the surface of image profile umbra I at point $i(c_x, c_y)$, the height of the image profile in terms of the moving uv -coordinate is $i(c_x, c_y) = p(d_u, d_v)$ at \mathbf{d} , and the heights of the image profile at other (u, v) locations are $i(c_x + u - d_u, c_y + y - d_v)$. By adding the height of the tip at the touch point (d_u, d_v) to the difference between the two heights above, we get $dill = i(c_x + u - d_u, c_y + y - d_v) - i(c_x, c_y) + p(d_u, d_v)$. By calculating $dill$ for all possible \mathbf{d} , we find the minimum value of $dill$, written as $dill_{min}$. If $dill_{min} + thresh < p(u, v)$, then the tip shape is updated (or “carved”) with a smaller height, $p(u, v) = dill_{min} + thresh$. Here, $thresh > 0$ is a parameter that determines a tolerance of inconsistency between the image profile and the tip estimate. The above calculations are performed for all (c_x, c_y) positions on the image profile $i(c_x, c_y)$. All the computational steps are repeated until the tip shape is converged. In the end, the upper bound on the true shape P , $P^{(r)} \supseteq P$, is reconstructed. In this algorithm, $thresh$ is a critical parameter for the accuracy of the tip reconstruction. Indeed, if noise is contained in the image profile, the reconstructed tip shape is strongly influenced by the geometries or dents of noise in the image profile. To choose the best $thresh$ value, however, it is necessary to find a scale that well discriminates the heights of the sample surface geometry from the heights of the noise geometry, which is a difficult task, especially for novice users.

In the following, the end-to-end differentiable BTR is introduced (Fig. 1f). The fact that the tip reconstruction is affected by noise in the original BTR can be regarded as an overfitting problem in machine learning. In machine learning, a typical approach to prevent overfitting is to use an appropriate loss function including a regularization. The point of our idea is to statistically determine whether or not to carve the tip from the entire AFM data according to a loss function that takes noise into account, rather than individually determining whether or not to carve the tip from each dent in the image profile, as in the original BTR. Following the original BTR, our differentiable BTR is based on the relation $I \circ P = I$, but the condition is relaxed because the equality does not hold in the presence of noise. Specifically, assuming that the noise is spatially independent Gaussian noise, the loss function to be minimized would be the mean square error,

$$MSE(p) = \frac{1}{N} \|i^{(r)} - i\|^2 = \frac{1}{N} \sum_{x, y} \left(i^{(r)}(x, y) - i(x, y) \right)^2$$

where N is the number of pixels, and $\|\cdot\|^2$ is the L2 norm. $i^{(r)}$ is an image created by the opening $I^{(r)} = I \circ P = (I \ominus P) \oplus P$ with some tip shape P . Expanding $i^{(r)}(x, y)$ by morphology operators, it can be written as

$$i^{(r)}(x, y) = \max_{u, v} \left[\min_{u', v'} \left[i(x - u + u', y - v + v') - p(u', v') \right] + p(u, v) \right]$$

The differentiable BTR finds the optimal tip shape $p(u, v)$ by minimizing the mean square error over tip shape through its gradient. Because max and min functions appeared in dilation and erosion seem to be non-differentiable at first glance, we treat this problem following the implementations of Max Pooling layer in convolutional neural networks (CNNs). For example, consider a one-dimensional tip shape p_u with u discretized by the index of pixels. The gradient of $h = \max[p_1, p_2, p_3]$ is, if p_3 is the maximum, $\partial h / \partial p_1 = 0$, $\partial h / \partial p_2 = 0$, $\partial h / \partial p_3 = 1$, which is equivalent to argmax function. In image processing, some studies try to apply morphology operators to

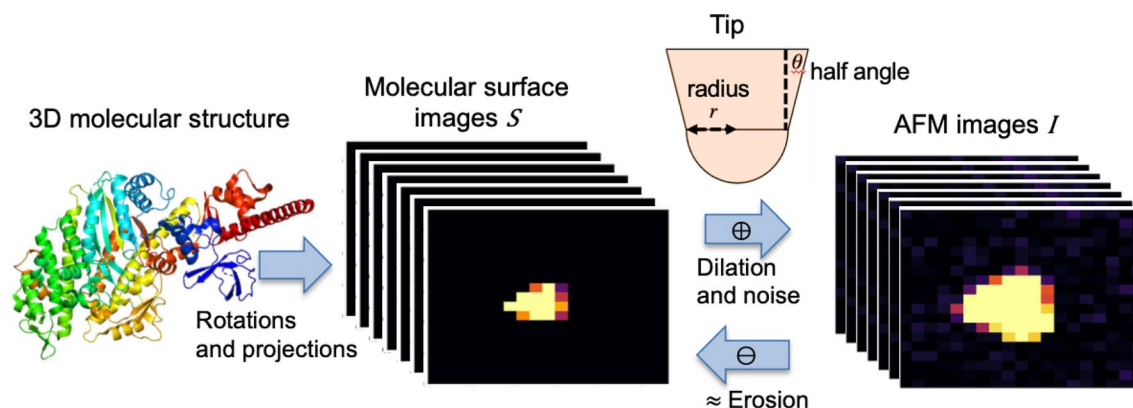


Figure 2. Schematic of twin experiments. From the structural model, the images of molecular surfaces are generated. Then, the molecular surfaces are converted to pseudo atomic force microscopy (AFM) images by dilation with a given tip shape. These molecular surfaces and tip shape are used as ground truths in the experiments. The 3D molecular structure is drawn with PyMOL³⁹.

image processing by using argmax and argmin functions for differentiations³⁵. In this study, we also use argmax and argmin functions for the differentiations of max and min functions, respectively.

Unfortunately, minimizing the above loss function is an ill-posed problem³⁶. Other than the correct tip shape, for example, a very thin tip like the δ -function can make the value of the loss function almost zero (resulting in an overfit to noise for noisy AFM images). Therefore, regularization is introduced to find the upper bound on tip shape without an overfit while keeping the loss function small. In this study, we add the L2 norm of tip shape P as a regularization term. Finally, the loss function used in the differentiable BTR becomes

$$\text{MSE}(p) + \lambda \|p\|^2 = \frac{1}{N} \sum_{x,y} \left(i^{(r)}(x,y) - i(x,y) \right)^2 + \lambda \sum_{u,v} p^2(u,v)$$

The purpose of introducing the regularization term is to find a blunt tip because unreasonably thin tips may be ascribed to the overfitting to noise. We minimize the sum of $\text{MSE}(p)$ and the regularization term (Fig. 1e). By applying the regularization term, a blunter tip shape is expected to be obtained because thin tips have large $\|p\|^2$ values. Here, the L2 norm was chosen for the availability of optimizers. To find an appropriate value of λ , it is typical to start with $\lambda = 0$ or very small λ , then gradually increase λ to find a blunt tip keeping $\text{MSE}(p)$ small. To systematically find an appropriate λ value, this study propose to use the so-called *the one standard error rule*³⁷ (discussed below).

Moreover, we used a square pillar ($p(u,v) = 0$ for all (u,v) over the defined area for representing the tip) as an initial condition for the tip reconstruction, which is the same strategy as the original BTR. This choice of initial condition contributes to finding a blunt tip or the upper bound on tip shape. The optimization algorithm was implemented in a neural network/machine learning framework (the Flux.jl package³⁸ in Julia programming language). Specifically, we implemented the dilation and the erosion as differentiable parts of neural network layers and optimized the tip shape in the same way in the training of neural networks (see “Methods” for details).

Twin experiment: noise-free AFM images. By performing twin experiments, we compared the accuracies of two blind tip reconstruction algorithms, the original and end-to-end differentiable BTRs (Fig. 2). In the experiments, 20 frames of artificial pseudo-AFM images were generated from the structure of myosin V motor domain (PDB ID: 1OE9⁴⁰) using a typical tip shape (a hemisphere combined with a circular frustum of a cone²⁰, which serves as the ground truth tip). The reconstructed tip shape and molecular surface were compared to those of the ground truth (see Methods for details). For each pseudo-AFM image, the structure was randomly oriented. To investigate the relationship between the parameters (*thresh* and λ) of the algorithms and the identity $I \circ P = I$, we applied wide range of parameter values and monitored the mean square error $\text{MSE}(p)$ and reconstructed tip shapes.

Figure 3 shows the results of BTRs from noise-free AFM images. Note that some of the pixels around the sample molecule in the image are cropped and enlarged for visual clarity. The full-size images are shown in Supplementary Fig. 1. Under the noise-free condition, the original BTR can perfectly reconstruct the ground truth tip with *thresh* $\ll 1$ (Fig. 3d, left). $\text{MSE}(p)$ increases as *thresh* increases (Fig. 3b), correlated with the thickness of the reconstructed tip shape (Fig. 3d, right). The rate of the increase is slow for a range of small *thresh*, but is gradually accelerated later. The same is true for the differentiable BTR. The differentiable BTR also works perfectly to reconstruct the true tip shape with $\lambda \ll 1$ (Fig. 3e, left). $\text{MSE}(p)$ increases as λ increases (Fig. 3c), reconstructing thicker tips (Fig. 3e, right). Notably, the slope of $\text{MSE}(p)$ with respect to λ increases suddenly.

Figure 3h shows the molecular surfaces reconstructed by deconvolving the 1st frame of the pseudo-AFM images (Fig. 3a) of the data set with the reconstructed tips. Here, the deconvolution was performed by using the erosion. The results show the shape outlines of myosin V motor domain recovered by using both BTRs. Although the reconstructed molecular surfaces $S^{(r)}$ by the erosion have a property of $S^{(r)} \supseteq S$, making the reconstructed surface thicker than the ground truth, the reconstructed molecular surfaces look qualitatively similar to that of

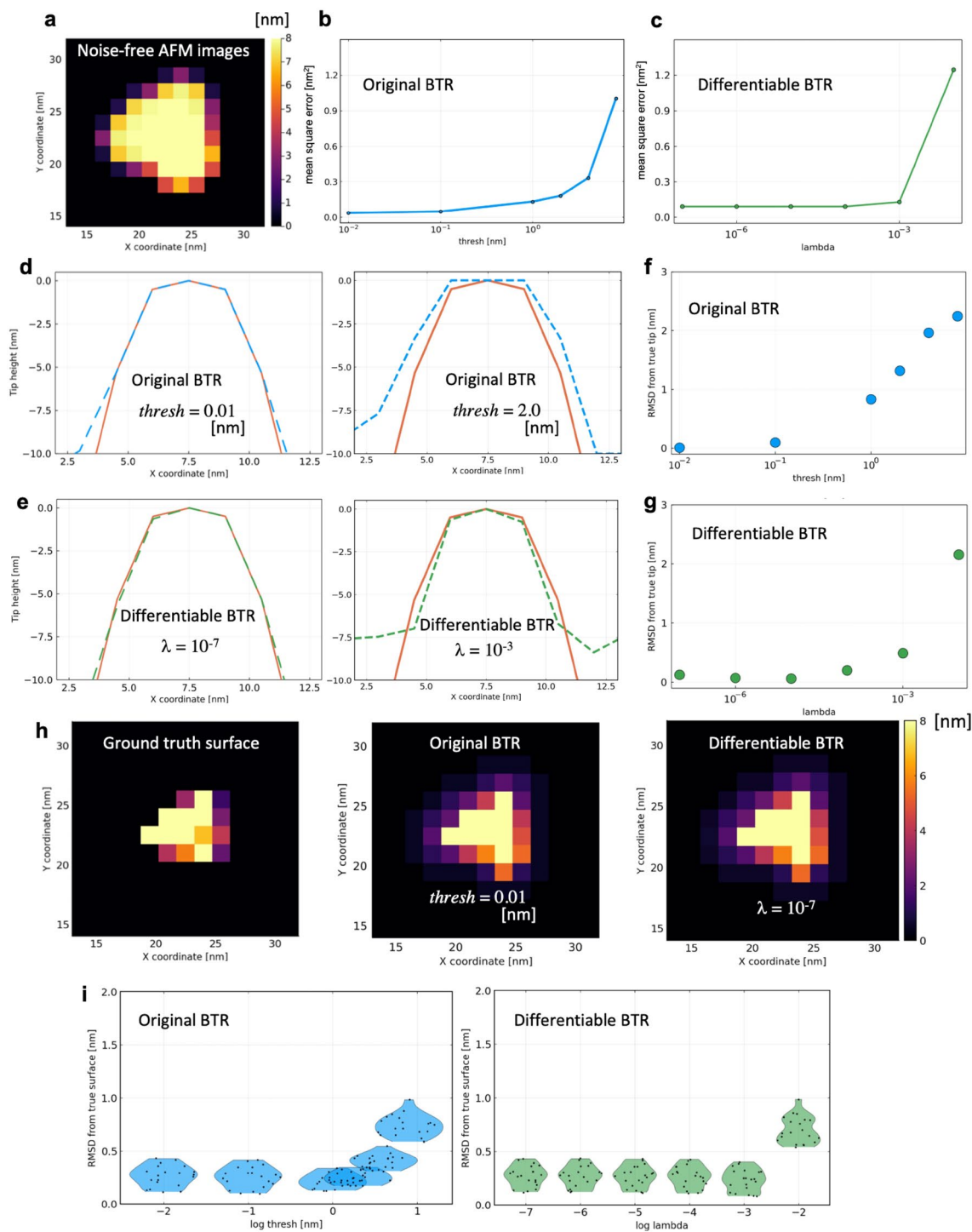


Figure 3. Results of twin experiment in noise-free condition. **(a)** 1st frame of 20 images used for the twin experiment. **(b, c)** Mean square errors at various parameter values. **(d, e)** Cross sections of reconstructed tip shapes along the x -axis with the original and differentiable blind tip reconstructions (indicated by dashed blue lines and dashed green lines, respectively), compared with the ground truth (red line). **(f, g)** Root mean square deviations (RMSDs) of the reconstructed tips from the ground truth (with the same coloring scheme). **(h)** Reconstructed molecular surfaces by the deconvolutions with the reconstructed tips. **(i)** RMSDs of the deconvoluted molecular surfaces of all 20 frames from the ground truth visualized by violin plots.

myosin V motor domain for both BTRs. The accuracies of the reconstructed molecular surfaces are quantitatively evaluated by root mean square deviations (RMSDs) from the ground truth surfaces (Fig. 3i). In the calculation, deconvolutions were performed for all 20 frames of the images and compared with the ground truth. The RMSDs show that both BTRs have good accuracies for reconstructing molecular surfaces from noise-free AFM images.

Twin experiment: noisy AFM images. We then compared the robustness of the two BTRs against noise by adding spatially independent Gaussian noise with a standard deviation of $\sigma = 0.3$ nm to the 20 frames of pseudo-AFM images used in the previous noise-free condition. Here, $\sigma = 0.3$ nm is a typical noise size in the current HS-AFM measurements²⁰. 100 sets of noisy pseudo-AFM data (each containing 20 frames) were created using 100 different noise realizations. Again, to investigate the relationship between the parameters (*thresh* and λ) and the identity $I \circ P = I$, we applied parameter values in a wide range and monitored $\text{MSE}(p)$ (Fig. 4b,c). The mean and standard deviation of $\text{MSE}(p)$ was calculated only from the single set of 20-frame images by using the cross validation (see “Methods” for details). The accuracies of reconstructed tip shapes and surfaces were evaluated by 100 sets of different noise realizations.

In both BTRs, $\text{MSE}(p)$ sharply increases. When the tip shapes are plotted for the original BTR, reconstructed tip shapes seem to depend not only on parameters but also on noise realization (Fig. 4d). Indeed, even with the same *thresh* parameter 4.0 nm, the original BTR reconstructed the tip shape thinner or thicker than the ground truth, depending on the realization of the noise (Fig. 4d, right). The tip shape tends to be thin for *thresh* $\ll 1$ (Fig. 4d, left). This means that the original BTR algorithm overfitted the tip shape to the noise in the pseudo-AFM images. The RMSD from the ground truth tip shows that *thresh* ≈ 4.0 nm is the best parameter value for accurate tip reconstruction (Fig. 4f). Parameter tuning of *thresh* has been discussed in previous studies. Villarrubia proposed, from the analysis of a single-frame pseudo-AFM image, to use $2-4\sigma$ for *thresh*⁹. Although Villarrubia analyzed only a single-frame image, this value may depend on the number of frames used because the chance for noise level exceeding *thresh* changes by the number of samples. In this sense, *thresh* = 4.0 nm would be reasonable for the current 20-frame AFM images. Dongmo et al.²⁶ proposed to monitor the change of tip volume by changing *thresh* parameter from low to high values. They discussed that the parameter where tip volume change becomes large would be regarded as the optimal one²⁶. In the current context, that parameter value might correspond to the value where the mean square error $\text{MSE}(p)$ starts to increase sharply (*thresh* ≈ 4.0 nm in Fig. 4b).

On the other hand, the reconstructed tip shapes by the differentiable BTR show less dependence than the original BTR, and similar tip shapes (close to the ground truth) are reconstructed more stably for most noise realizations (Fig. 4e). Some tip shapes look distorted in the reconstructed tips of $\lambda = 10^{-3}$ (Fig. 4e, right), partially due to the boundary conditions of the tip image. Interestingly, while the tip shape became thicker as λ increases due to the regularization penalty (Fig. 4e, right), the reconstructed tip shape did not fall into thin tip shapes even when the λ becomes small ($\lambda \ll 1$, Fig. 4e, left). This robustness in the differentiable BTR enables novice users to reproduce accurate reconstructions using a parameter value from a wide range. The average RMSD of the reconstructed tips from the ground truth is 2.0 nm for the original BTR with *thresh* = 4.0 nm, and 0.4 nm for the differentiable BTR with $\lambda = 10^{-7}$ (Fig. 4f,g).

Figure 4h shows the molecular surfaces obtained by deconvolving the 1st frame of the pseudo-AFM images (Fig. 4a) using ones of the reconstructed tips. In Fig. 4h, the reconstructed molecular surface by the original BTR looks somewhat blurred because the erosion “carves” the shape edges in the image profile by the reconstructed tip shape with a larger half angle (the black line in Fig. 4d) than the ground truth. On the other hand, the differentiable BTR successfully deconvolved the shape outlines of myosin V motor domain. The accuracies of the reconstructed surfaces were quantitatively evaluated by RMSDs from the ground truth (Fig. 4i). The surface reconstructions by the differentiable BTR outperform the original BTR in this noisy condition. The average RMSD of the reconstructed molecular surfaces from the ground truth is 2.1 nm for the original BTR with *thresh* = 4.0 nm, and 1.1 nm for the differentiable BTR with $\lambda = 10^{-4}$.

Thus far, we have not found the lower limit of λ even in the above noisy AFM images, suggesting the noise level was small enough for the current loss function to avoid the overfit. To explore the lower limit of λ , we intentionally increased the noise level and conducted another twin experiment. In this twin experiment, we have applied a large noise level with a standard deviation of $\sigma = 1.2$ (Supplementary Fig. 2). From the results, we found the estimation accuracy becomes worse for small values of λ , suggesting that the tips with lower values of λ tend to be overfitted to the noise.

For selecting an appropriate value of λ in this situation, we here propose to use the so-called *the one standard error rule*³⁷. In this rule, we first perform the cross validation and calculate the mean and standard deviation of $\text{MSE}(p)$. Then, we find the λ value with the minimum $\text{MSE}(p)$, and select the largest λ whose mean falls within or around the one standard deviation of the minimum. The idea behind this rule is to choose the bluntest tip shape from a set of tip shapes appearing to be more or less equally good in terms of $\text{MSE}(p)$. In the case of this twin experiment, we found that $\lambda = 10^{-3}$, slightly above the one standard error bound, yields the most accurate tip reconstruction (indicated by the red arrow in Supplementary Fig. 2c).

Twin experiment: double-tip effect. The double-tip effect is one of the most frequent artifacts in AFM measurements, which occurs when the tip is damaged or contaminated during the measurement⁴¹. Since AFM images measured by damaged or contaminated tips are often discarded without further analysis, it would be helpful if the double-shaped tip could be reconstructed only from AFM images and the doubled images were deconvolved to obtain true molecular surfaces. In this twin experiment, we mimicked the double tip effect with a double-peaked tip shape created by aligning the single tips used in the previous twin experiments along the *x*-axis. As in the previous twin experiment, we added spatially independent Gaussian noise with a standard deviation of $\sigma = 0.3$ nm to the pseudo-AFM images emulated by the double tips (Fig. 5a). Again, to investigate

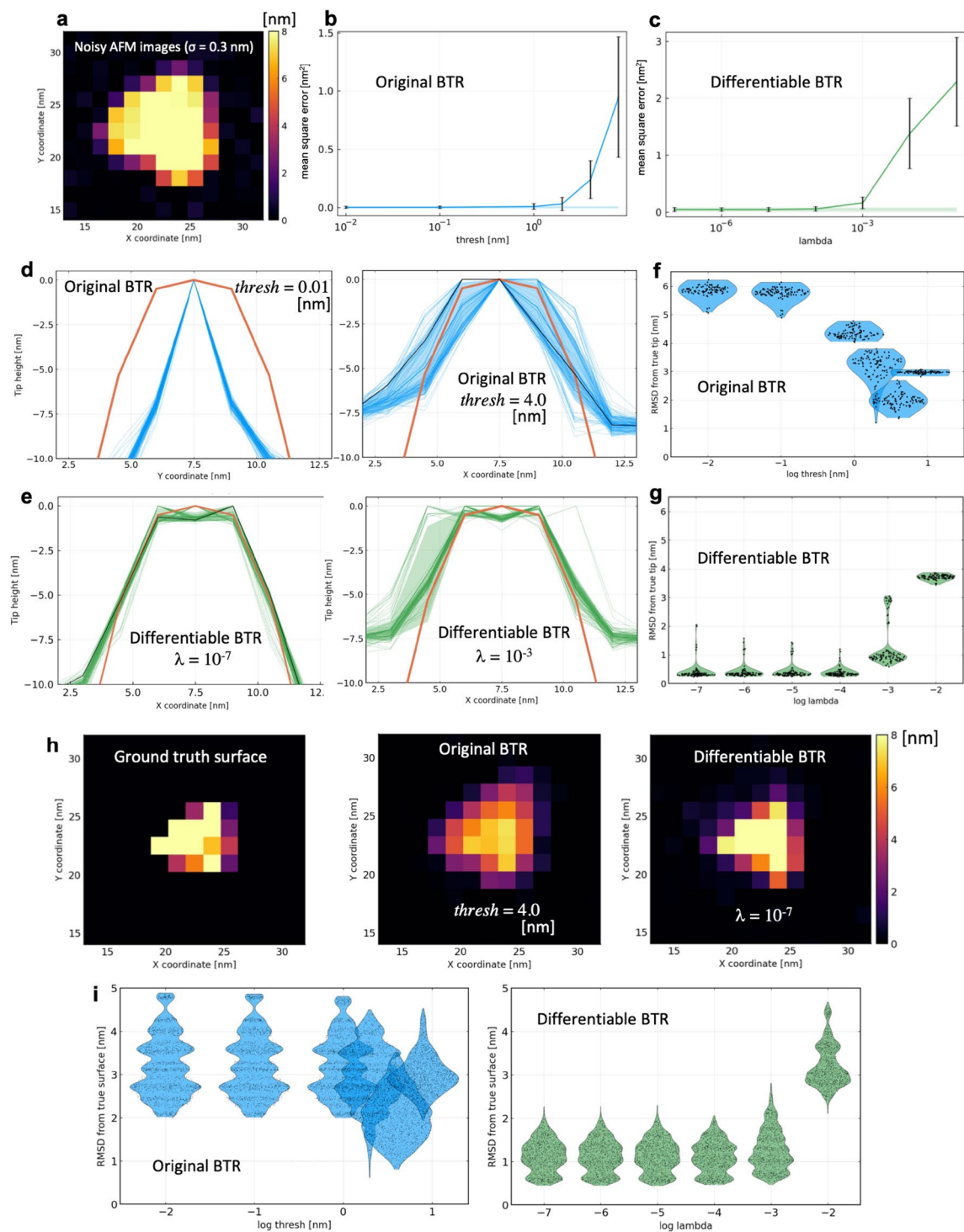


Figure 4. Results of twin experiment in noisy condition. (a) 1st frame of 20 images used for the twin experiment. (b, c) Mean square errors (MSEs) optimized at various parameter values. The mean and standard deviation of MSE was calculated only from the single set of 20-frame images using the cross validation. Shaded area indicates the one standard error bound. (d, e) Cross sections of reconstructed tip shapes along the x -axis with the original and differentiable blind tip reconstructions (indicated by dashed blue lines and dashed green lines, respectively), compared with the ground truth (red line). Shaded area represents the standard deviation. The tip shapes used for deconvolution are indicated by the black lines. (f, g) Root mean square deviations (RMSDs) of the reconstructed tips from the ground truth visualized by violin plots (with the same coloring scheme). (h) Reconstructed molecular surfaces by the deconvolutions with the reconstructed tips. (i) RMSDs of the deconvoluted molecular surfaces of all frames (20 images) from the ground truths visualized by violin plots.

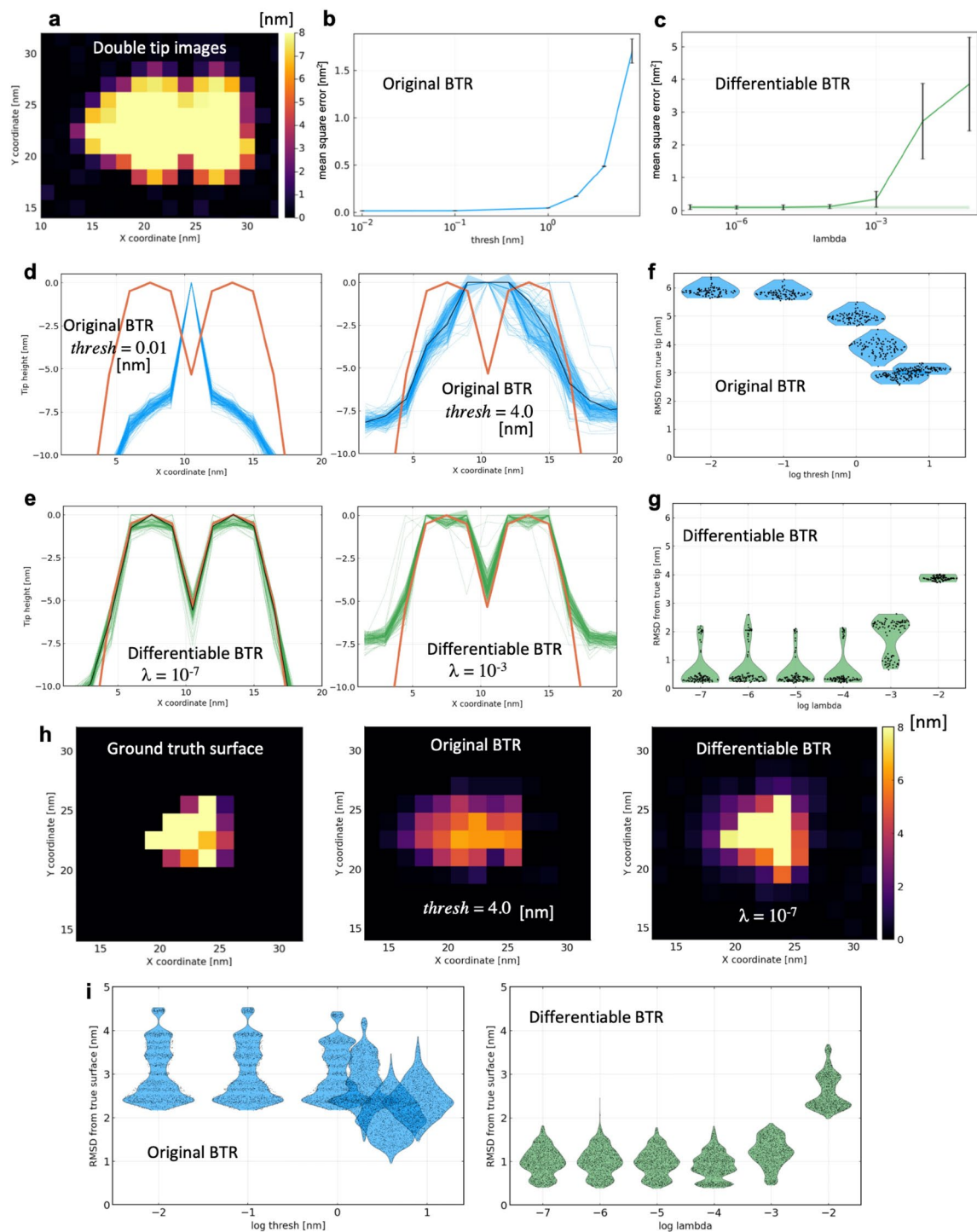


Figure 5. Results of twin experiment in a double-tip case. **(a)** 1st frame of 20 images used for the twin experiment. **(b, c)** Mean square errors at various parameter values. The mean and standard deviation of MSE was calculated only from the single set of 20-frame images using the cross validation. Shaded area indicates the one standard error bound. **(d, e)** Cross sections of reconstructed tip shapes along the x -axis with the original and differentiable blind tip reconstructions (indicated by dashed blue lines and dashed green lines, respectively), compared with the ground truth (red line). The tip shapes used for deconvolution are indicated by the black lines. **(f–g)** Root mean square deviations (RMSDs) of the reconstructed tips from the ground truth visualized by violin plots (with the same coloring scheme). **(h)** Reconstructed molecular surfaces of all frames (20 images) from the ground truths visualized by violin plots. **(i)** RMSDs of the deconvoluted molecular surfaces from the ground truths visualized by violin plots.

the relationship between the parameters ($thresh$ and λ) and the identity $I \circ P = I$, we applied parameter values in a wide range and monitored $MSE(p)$ (Fig. 5b,c). The mean and standard deviation of $MSE(p)$ was calculated only from the single set of 20-frame images by using the cross validation (see “Methods” for details). The accuracies of reconstructed tip shapes and surfaces were evaluated by 100 sets of different noise realizations.

Figure 5d shows tip shapes reconstructed with the original BTR. The original BTR failed to detect double tips, and the reconstruction resulted in single-peak tips for most parameter values. On the other hand, the differentiable BTR successfully detected double tips, and their shapes are roughly consistent with the ground truth over a wide parameter range (Fig. 5e). Again, translational shifts in some of the reconstructed tips are observed due to the boundary effect of the tip image. The result does not mean that the original BTR algorithm inherently cannot detect double tips. Indeed, under noise-free condition, the original BTR is able to reconstruct the double-tip shape with high accuracy. In the original BTR algorithm, the tip apex is set to be at the origin of the uv -coordinates, and the apex should be fixed at the origin during the reconstruction without any modifications. In noisy conditions, while one of two tip apexes is fixed at the origin, the other tip apex is irreversibly “carved” by chance due to noise, making it difficult to detect double tips. On the other hand, since the algorithm of the differentiable BTR can reconstruct the tip without imposing any constraints on the tip apex, it can stably detect double tips. In the differentiable BTR, the origin of the uv -coordinates is set to the center of mass of the tip (see “Methods” for details). The average RMSD of the reconstructed tips from the ground truth is 2.9 nm for the original BTR with $thresh = 4.0$ nm, and 0.5 nm for the differentiable BTR with $\lambda = 10^{-5}$ (Fig. 5f,g).

Figure 5h shows the molecular surfaces obtained by deconvolving the 1st frame of the pseudo-AFM images (Fig. 5a) using one of the reconstructed tips (Fig. 5d,e). With a single tip reconstructed by the original BTR, the molecular surface looks blurred because the erosion “carves” the shape edges in the image profile by the reconstructed tip shape with a larger half angle (the black line in Fig. 5d) than the ground truth double-tip. Note also that the maximum height of the reconstructed surface decreases because the reconstructed tip with a large half angle can erode smaller edges, including the top edge of the image profile. On the other hand, double tips reconstructed with the differentiable BTR can remove the doubled molecular shapes and reconstruct a molecular surface much closer to the ground truth. The accuracies of the reconstructed surfaces were quantitatively evaluated by RMSDs from the ground truth ones (Fig. 5i). The average RMSD of the reconstructed molecular surfaces from the ground truth is 2.0 nm for the original BTR with $thresh = 4.0$ nm, and 0.9 nm for the differentiable BTR with $\lambda = 10^{-4}$.

Real experimental data: Myosin V walking along actin filament. Finally, we compared the two BTR algorithms by analyzing real AFM data. We analyzed the HS-AFM data of myosin V measured by Kodera et al.⁴. Myosin V is a homo-dimeric molecular motor that moves linearly along the actin filament driven by ATP hydrolysis free energy. Myosin V serves to transport cargos attached to the center of the dimer. The HS-AFM movies by Kodera et al. successfully capture the events that myosin V walks on the actin filament coupled with its ATP-dependent conformational change. It is important to deconvolve the AFM images to obtain the detailed shape of the Myosin V’s molecular surface during the walking step. Here, we analyzed 30 frames of a HS-AFM movie that capture a single walking step of Myosin V.

As illustrated in Fig. 6a, the HS-AFM images contain numerous horizontal scars emanating from the molecules to the right direction along the x -axis, which corresponds to the scanning direction of the tip. These scars are often called the parachuting artifact⁴². The “parachuting” means that, when the tip scan velocity is high and the target molecular height suddenly decreases along the tip scanning line, the tip can completely detach from the molecular surface. It takes some time for the tip to land on the stage or the sample surface at a different point, creating scars in the direction of scanning⁴². The appearance of scars is stochastic because the parachuting caused by thermal fluctuations of the sample molecule that lift the tip from the stage. On the other hand, after a parachuting happened, the decreasing pattern of the scar’s height is deterministic and could be modeled as a part of dilation. Thus, we here recognize the decreasing pattern of the scars as part of the signal and try to reflect them in the effective tip shape.

Figure 6b and c show the means and standard deviations of $MSE(p)$ calculated from the 30-frame HS-AFM images by using the cross validation at various $thresh$ and λ parameters with the original and differentiable BTRs, respectively. Figure 6d and e show the x - and y -axes cross sections of the tip shapes reconstructed from the real HS-AFM images by both BTRs. The tip shapes reconstructed with the original BTR tend to be thicker along the x -axis (Fig. 6d, left) while very thin along the y -axis (Fig. 6d, right). This is largely due to the scars of the parachuting artifact. After the parachuting occur in a stochastic manner, the decreasing pattern of the scar’s height would be modeled as a part of dilation along the x -axis. However, along the y -axis, whether the parachuting appears or not are quite stochastic: The parachuting appears only when the fluctuating molecular surface happens to be high (large z -coordinates). Therefore, along the y -axis, dents caused by the differences between parachuting and non-parachuting lines should be recognized as stochastic noise. The original BTR misinterpreted that the large dents along the y -axis can be ascribed to the tip shape and excessively “carved” the tip shape. While the tip shapes reconstructed by the differentiable BTR well capture a smooth attenuation by parachuting on the right side of the x -axis (Fig. 6e, left), those along the y -axis do not become thin as in the original BTR but maintain a certain degree of thickness with less parameter dependence (except for a very small weight, $\lambda = 0.0010$, Fig. 6e, right).

Figure 6h and i show the molecular surfaces of the 27th frame of the HS-AFM movie deconvolved with the reconstructed tip shapes by the original and differentiable BTRs, respectively. The 27th frame was chosen because flexible fitting molecular dynamics simulations were performed for this frame in our previous study⁴³. Following the result of the twin experiment with a large noise level ($\sigma = 1.2$ nm), we here used the one standard error rule for selecting a λ value of the differentiable BTR. Specifically, we selected $\lambda = 0.0051$, which is slightly above the one standard error bound (Fig. 6c), for the deconvolution. For the original BTR, we used the tip of $thresh = 1.95$ nm

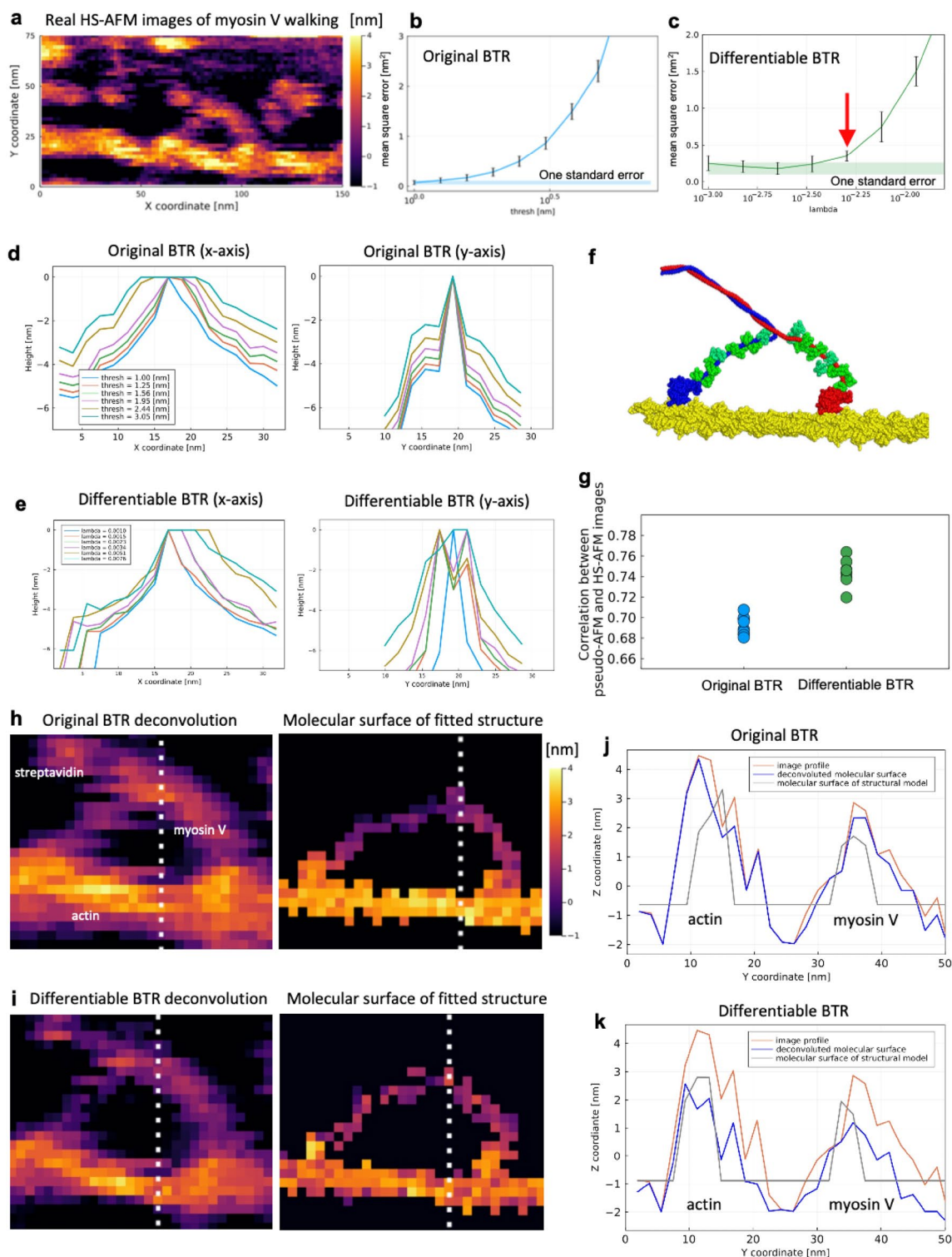


Figure 6. Blind tip reconstruction (BTR) from the high-speed atomic force microscopy (HS-AFM) data of myosin V walking. **(a)** 27th frame of 30 images used for the analysis. **(b, c)** Mean square errors at various parameter values. The mean and standard deviation of MSE was calculated using the cross validation. Shaded area indicates the one standard error bound. The red arrow indicates the selected λ for deconvolution. **(d, e)** Cross sections of reconstructed tip shapes along the x -axis with the original and differentiable BTR. **(f)** Representative structure of 10 structures taken from previous flexible fitting molecular dynamics simulations. **(g)** Maximum correlation coefficients between the 27th HS-AFM image and pseudo-AFM images generated from each of 10 structures. **(h)** Reconstructed molecular surfaces created by deconvolutions with the BTR tips compared with the molecular surfaces of structural models with the maximum correlation coefficients. The white dashed lines indicate the cross sections used in **(j)** and **(k)**. **(j, k)** Cross sections of the molecular surfaces along the y -axis. Red lines are the image profiles, blue lines are the reconstructed molecular surfaces, and gray lines are the molecular surfaces created from the structural models.

because the tip apex radius of this parameter value is comparable to (or slightly larger than) that of the differentiable BTR with the chosen $\lambda = 0.0051$ (see Fig. 6d,e) along the x -axis. In the molecular surface deconvolved by the tip of the original BTR (Fig. 6h), the legs of Myosin V are thinner than in the image profile, but there are still horizontal scars due to the parachuting artifact. On the other hand, the molecular surface deconvolved by the tip shape of the differentiable BTR has fewer parachuting lines (Fig. 6i). This is because the differentiable BTR has a thicker tip shape along the y -axis direction, which contributes to more extensive “carvings” by the erosion of the image profiles along the y -axis. The same trend can be seen in all the frames (Supplemental Movie 1).

Are the molecular surfaces obtained by the deconvolutions accurate? To verify this, we compared these surfaces with structural models. Here we used 10 structures taken from our previous flexible fitting molecular dynamics simulation study of the 27th frame of the same data⁴³. In the study, the structures of myosin V and actin filament were modeled using the crystal structure by homology modeling, and then coarse-grained molecular dynamics simulations were performed, imposing restraints on the structure fitted to the 27th frame of the experimental HS-AFM movie. Since the tip shape used in the flexible fitting is not the same as the current study, we selected 10 plausible structures as templates to make a comparison with the deconvoluted surfaces (Fig. 6f). Each structure was exhaustively translated and rotated to create pseudo-AFM images, and the correlation coefficients of those images with the target frame of the HS-AFM data were calculated. We recorded the maximum correlation coefficients and corresponding poses (translation and rotation) for the 10 structures. Figure 6g compares the maximum correlation coefficients of the 10 structures for both BTRs. Although the improvement in the coefficients by the differentiable BTR is marginal, the distributions of both BTRs are non-overlapping with each other implying that the improvement is statistically meaningful. This further means that the tip shape is consistent with structural models and the HS-AFM data. Figure 6j and k show the cross sections of the deconvoluted molecular surfaces and the surfaces obtained by the structural models that have the maximum correlation coefficients for two BTRs, respectively. In the figure, actin filament is located at around $y = 13$ nm. The reconstructed tip of the original BTR barely “carves” the image profile with erosion (the red and blue lines are almost same in Fig. 6j) due to its thin shape along the y -axis while the tip of the differentiable BTR “carves” the image profile. In particular, the deconvoluted surface of the differentiable BTR is consistent with the structural model. Here, the height of the molecular surface of a structural model is modeled by $\bar{\alpha}z + \bar{\beta}$, where $\bar{\alpha}$ and $\bar{\beta}$ were determined by the least squares fitting to each deconvoluted surface (see “Methods” for details). Again, this result implies the improved accuracy of the reconstructed tip by the differentiable BTR.

Discussion

In this study, we have proposed the end-to-end differentiable BTR method based on a loss function including a regularization term considering noise in AFM image to avoid overfitting. The loss function is systematically optimized following the framework developed by the recent advances in deep learning studies. The results of the twin experiments showed that the differentiable BTR is more robust to noise than the original BTR. Finally, demonstration of real HS-AFM data shows that the differentiable BTR reconstructs tip shapes and the deconvoluted molecular surface is consistent with the structural model. Since the method is a quite general, we expect that the method would become a routine method for analyzing noisy AFM images as well as HS-AFM data.

The twin experiments have shown that reconstructed tip shapes with the differentiable BTR are less dependent on the parameter λ . For relatively small noise ($\sigma = 0.3$ nm), any small λ values ($\lambda \ll 1$) are expected to reproduce tip shapes close to the ground truth one. For relatively large noise ($\sigma = 1.2$ nm), small λ may reconstruct unreasonably thin tip shapes even with the differentiable BTR. In such cases, the one standard error rule is recommended to reconstruct reasonably thick tip shapes. As with the case of neural network frameworks, there are several hyperparameters in the optimization process in the differentiable BTR. For example, if the number of epochs is too large, that increases the chance to overfit to noise, while too small epochs result in insufficient learning. Using techniques such as early stopping, or hyperparameter optimization tools⁴⁴, may help to choose appropriate hyperparameters.

As mentioned in Introduction, Tian et al. proposed an improved regularization scheme against noise in the framework of original BTR²⁷, and Józwiak et al. simplified Tian’s idea to the case of standard AFM tips²⁸. In this scheme, the update procedure of tip shape becomes more conservative, and it is thus expected to be robust against noise, as shown by Józwiak et al.²⁸. Supplementary Fig. 3 shows the results of the original BTR using this regularization scheme for noisy single-tip and double-tip twin experiments. As expected, the regularization scheme improves the estimation and reconstructs blunter tip shapes compared to the original BTR. However, the RMSDs from the ground truth tip shape and molecular surface show that the differentiable BTR still outperforms the original BTR with the regularization scheme.

Since this study mainly targets the analysis of HS-AFM data, pseudo and real AFM data consisting of 20–30 frames were mainly analyzed. To check the dependence of reconstruction accuracy on the number of frames, we conducted noisy single-tip twin experiments using 1, 10, and 100 pseudo-AFM frames (Supplementary Fig. 4). Although the accuracies of tip and surface reconstructions are comparable in both BTRs for a single frame, the differentiable BTR outperforms the original BTR for 10 and 100 frames. An only drawback of the differentiable BTR here is that the computation time is much longer than the original BTR, especially for the analysis of 100 frames. For example, in our twin experiment (the single-tip case, 20 frames), the original BTR took 0.02 s to estimate the tip shape of a specific *thresh* while the differentiable BTR took 5.7 s for a specific λ with Intel Xeon Gold 6330 CPU (without multi-threads). Implementing GPU kernels for dilation and erosion, or using smooth functions instead of max and min, would accelerate the optimization.

A concern with differentiable BTR is the relationship between ground truth tip shape and erosion. To reconstruct the tip accurately with differentiable BTR, erosion must be a good approximation to the inverse function of dilation. If the ground truth tip has a cone shape, the half angle must be small for the erosion to approximate the

inverse function of the dilation well. As the half angle increases and exceeds a threshold value (e.g., 30 degrees), the accuracy of tip reconstruction deteriorates quickly because erosion fails to approximate the inverse function. However, as shown by the numerical study by Sumikama et al.⁴⁵, in the tapping mode of HS-AFM, the half angle is expected to become rarely large because the effective tip shape becomes thinner. Also, one possible direction to overcome this problem would be to employ a CNN as a substitute for erosion. If the amount of data is large enough, a CNN is expected to approximate the inverse function of dilation well.

In the analysis of the HS-AFM data of Myosin V, we reconstructed the effective tip shape regarding the scars caused by the parachuting effect as a signal. However, there may be cases where we want to estimate the real physical tip shape after removing the parachuting effect. In this case, it may be possible to divide the dilation into two layers in the differentiable BTR. The first dilation introduces a tip shape that adds lines linearly on the right side of the molecule, and the slope of the lines is parametrized. The second dilation is a standard dilation with a tip shape $p(u, v)$. In such a framework, it would be possible to estimate both parameters simultaneously from AFM images. Since dilation is implemented as a part of the neural network layers, it is straightforward to extend the current BTR algorithm in this way. From another perspective, the loss function can be changed according to the types of AFM data. For examples, if AFM data contain large noise as outliers, using the L1 norm instead of the mean squared error (used in this work) for the loss function may work better for tip reconstruction.

Recently, Heath et al. developed localization AFM (LAFM)⁴⁶, an elegant image reconstruction technique to overcome resolution limitations in AFM. In this approach, peak positions contacted with the tip apex in AFM image profiles are detected and accumulated to increase the resolution. Although the LAFM currently only uses the peak positions in the localization algorithm, it might be possible to accumulate positions contacted with the side of the tip for the localization algorithm if the tip shape is accurately reconstructed with the differentiable BTR in the future.

Methods

Original blind tip reconstruction. The C language source codes provided by Villarrubia⁹ were ported to the Julia programming language. Inside the codes, xy -coordinates and uv -coordinates are discretized by pixel indices (x, y) and (u, v) , respectively. Starting with a square pillar tip, i.e., $p_{u,v} = 0$ (for any (u, v) , zero is the maximum height for the inverted tip) as an initial condition, the tip shape was iteratively “carved” according to the dents in neighboring pixels until no “carving” events occur. For the analysis of multiple frames, the algorithm was sequentially applied to frames as proposed by Villarrubia⁹.

End-to-end differentiable blind tip reconstruction. We implemented the differentiable BTR in Julia programming language. In the following, the xy -coordinate and uv -coordinate coordinates were discretized and used as pixel indices. The following algorithm was designed to minimize the loss function $L(p) = \frac{1}{N} \sum_{x,y} \left(i_{x,y}^{(r)} - i_{x,y} \right)^2 + \lambda \sum_{u,v} p_{u,v}^2$ with respect to $p_{u,v}$ (where N is the number of pixels):

1. Initialize the tip shape to be a square pillar shape, $p_{u,v} = 0$ (for any u and v). Here, zero is the maximum height for the inverted tip.
2. Calculate the gradient of the loss function $L(p)$ for the entire frames. Here, only the gradient of the mean square error is needed because the regularization term is implicitly imposed by AdamW⁴⁷ optimizer.
3. Update $p_{u,v}$ according to the computed gradient using AdamW optimizer imposing a regularization term $\lambda \sum_{u,v} p_{u,v}^2$. Learning rate of 0.1 nm, exponential decay for the first β_1 of 0.9 and the second β_2 of 0.999 were used.
4. Trim $p_{u,v}$ values greater than zero. Positive $p_{u,v}$'s are changed to zero, $p_{u,v} = 0$.
5. Using the height of $p_{u,v}$ relative to the minimum height of $p_{u,v}$ as weights, translate $p_{u,v}$ so that its center of weight is located at the origin of the uv -coordinate. For the unknown $p_{u,v}$ near the boundary (edge) created due to the translation, the minimum value of $p_{u,v}$ was assigned.
6. Repeat step 2 to step 5 using a batch size of 1 image until 100–2000 epochs.

Dilation and erosion were implemented as layers of the neural network framework (the Flux.jl package³⁸ in Julia language), and their custom pullbacks were defined using the ChainRulesCore.jl package⁴⁸. Differentiation of opening and the backpropagation of the loss function were computed with the automatic differentiation by the package Zygote.jl⁴⁹. The gradient of each dilation and erosion was tested during the development.

Twin experiments. To evaluate the accuracy of the tip shape reconstruction, we performed twin experiments. We first generated pseudo-AFM images from a structure of myosin V motor domain without bound nucleotide (PDB ID: 1OE9⁴⁰) using a specific tip shape, which serves as a ground-truth tip shape. The light chains were removed from the original crystal structure. Using the pseudo-AFM images, we performed the BTR and compared the tip shape reconstructed from the images with the ground truth tip shape. We set the z -axis of the xy -plane system to be the direction orthogonal to the stage (i.e., the xy -plane is parallel to the stage). Generated pseudo-AFM images consist of 30×30 pixels, where the width and height of each pixel are both 1.5 nm. In the simulation, the center of mass of the motor domain was placed at the origin of the xy -coordinate. Then the motor domain was randomly rotated in three dimensions. The heights of the van der Waals (vdW) sphere surfaces at the xy -coordinates at the pixel centers were calculated for neighboring atoms of each center, and the largest height was taken as the molecular surface. After obtaining the molecular surface at the pixel centers, the tip shape was used to create a pseudo-AFM image by performing dilation. We used the pixel size of 10×10 for the single-tip, 14×10 for the double-tip, and 17×11 for the HS-AFM data of myosin V walking. Finally, spa-

tially independent Gaussian noise with a standard deviation of 0.3 nm was added to each pixel to obtain noisy pseudo-AFM images. The standard deviation of 0.3 nm is a typical noise size for the current HS-AFM²⁰. Then, we performed another experiment by increasing the noise level to a standard deviation of 1.2 nm. For the single tip twin experiment, the tip shape of a hemisphere (with radius r) combined with a circular frustum of a cone, its apex radius of 2.5 nm and half angle of 10 degrees were used as the ground truth. For the double-tip twin experiment, two single tips were aligned in the x -axis direction. By repeating the above procedure, 20 pseudo-AFM images were simulated for each setup.

At various values of parameters, $thresh$ and λ , we calculated the mean square errors, $MSE(p)$. The mean and standard deviation of $MSE(p)$ were evaluated from the single set of 20-frame images by using the leave-one-out cross-validation (LOOCV). These quantities were used for applying the one standard error rule for selecting an appropriate λ . The accuracies of reconstructed tips and surfaces were evaluated from 100 different noise realization sets of 20-frame images. In the calculation of the root mean square deviation (RMSD) from the ground truth tip shape, uv -coordinates where the ground tip $p(u, v) > -7.0$ nm were used. 7.0 nm roughly corresponds to the height of myosin V motor domain. In the calculation of the RMSD from the ground truth molecular surface, xy -coordinates where the surface height $s^{(r)}(x, y) > 1.0$ were used. 1.0 nm roughly corresponds to 3σ (0.9 nm) of noise. Considering translational shifts in the reconstructed tips, we translated the tips exhaustively and recorded the minimum RMSD value for each tip.

HS-AFM data of myosin V. HS-AFM measurements of myosin V walking along actin filaments were carried out by Kodera et al., and details for the experiments are given in the paper⁴. In the measurements, the ATP-driven translocations of Myosin V (M5-HMM) along actin filaments were observed by the HS-AFM. Actin filaments were partially biotinylated and immobilized using streptavidin on biotin-containing lipid bilayers formed on a mica surface (stage). A positively charged lipid was included in the bilayer to facilitate the weak sideways adsorption of M5-HMM onto the bilayer surface. HS-AFM measurements were conducted with the tapping mode. This study used 30 frames of HS-AFM data sets, which well captured single-step walks of M5-HMM. Each image consists of 80×40 pixels, and the resolution of each pixel is $1.875 \text{ nm} \times 1.875 \text{ nm}$. We used these 30 frames to perform original and differentiable BTRs over various parameters. The tilt of the stage in the HS-AFM data was corrected by fitting a xy -plane to the data.

Analysis of HS-AFM data. Structural modeling and flexible fitting molecular dynamics simulations of Myosin V and actin filament were performed in the same way as in the previous study by Fuchigami and Takada⁴³. First, we generated all-atom models of a tail-truncated myosin V (each of two heads is in the lever-arm down and up conformations) and an actin 31-mer filament using MODELLER version 9.25⁵⁰. Then, we constructed coarse-grained models of the actin-myosin complex in the down-down state using CafeMol version 3.2.0⁵¹. We performed ten flexible fitting simulations of the actin-myosin complex in the down-down state using structure-based AICG2 + potential⁵², AFM stage potential, and AFM image-based bias potential⁵³. See the reference⁴³ for more details.

Using ten structures with the highest cosine similarity in each of the ten flexible fitting simulations, we performed the rigid-body fitting of 3D structures to 2D HS-AFM images of Myosin V. For each structural model, we exhaustively rotated the structure around the rotation axis (roll axis) of the helix of actin filament by 3 degrees (within the range of 60 degrees), combined with the rotation around the yaw axis of the filament by 3 degrees (within the range from -30 to 30 degrees starting from the actin filament parallel to the x -axis). For all the combinations of two angles, we created pseudo-AFM images from the structure. The pseudo-AFM images were created in the same way as in the twin experiment. For the radii of coarse-grained Ca beads, we used the effective radius of amino acid residues used in CryoEM data analysis⁵⁴. The coiled coil domain of myosin V were ignored in the generation of the pseudo-AFM images because this domain was not clearly observed in the HS-AFM. Since the height scale in the real HS-AFM data roughly corresponds to half of the structural model, we doubled the height of the tips reconstructed from the real HS-AFM images. Using it, we then applied the dilation to the structure-model molecular surface. The pseudo-AFM images from the structure model were translated over possible locations on the real HS-AFM image (note that the pseudo-AFM image is smaller than the HS-AFM images because it contains only a single myosin V and a fragment of actin filament), and we computed the correlation coefficient between the overlapped regions of two images. We recorded the maximum correlation coefficient between the pseudo-AFM image and the real AFM image and its corresponding pose (translation and rotation). The structure, of which pseudo-AFM image has the maximum correlation coefficient, was compared with the molecular surface obtained by the deconvolution of the HS-AFM image. In the comparison, the height of the molecular surface of the structural model was transformed by $\bar{\alpha}z + \beta$, where $\bar{\alpha}$ and β were determined by the least squares fitting to the deconvoluted surface. This transformation was needed to adjust the calibration and the offset level.

Data availability

Jupyter notebooks to reproduce the results of this paper are publicly available at https://github.com/matsunagalab/differentiable_BTR. Some notebooks automatically download the structure of the target molecule from Protein Data Bank and calculate pseudo-AFM images used in twin experiments. The structural models of myosin V and actin filament are included in the repository. Real HS-AFM data of myosin V walking along actin filament are available upon reasonable request to corresponding author and with the permission of Kodera et al.⁴.

Received: 9 November 2022; Accepted: 23 December 2022

Published online: 04 January 2023

References

- Binnig, G., Quate, C. F. & Gerber, Ch. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
- Ando, T., Uchihashi, T. & Fukuma, T. High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. *Prog. Surf. Sci.* **83**, 337–437 (2008).
- Ando, T. *High-Speed Atomic Force Microscopy in Biology* (Springer, 2022).
- Kodera, N., Yamamoto, D., Ishikawa, R. & Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature* **468**, 72–76 (2010).
- Uchihashi, T., Iino, R., Ando, T. & Noji, H. High-speed atomic force microscopy reveals rotary catalysis of rotorless F₁-ATPase. *Science* **333**, 755–758 (2011).
- Kodera, N. *et al.* Structural and dynamics analysis of intrinsically disordered proteins by high-speed atomic force microscopy. *Nat. Nanotechnol.* **16**, 181–189 (2021).
- Shibata, M. *et al.* Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat. Commun.* **8**, 1430 (2017).
- Ando, T. High-speed atomic force microscopy and its future prospects. *Biophys. Rev.* **10**, 285–292 (2018).
- Villarrubia, J. S. Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation. *J. Res. Natl. Inst. Stand. Technol.* **102**, 425 (1997).
- Keller, D. Reconstruction of STM and AFM images distorted by finite-size tips. *Surf. Sci.* **253**, 353–364 (1991).
- Keller, D. J. & Franke, F. S. Envelope reconstruction of probe microscope images. *Surf. Sci.* **294**, 409–419 (1993).
- Ando, T. *et al.* High-speed atomic force microscopy for capturing dynamic behavior of protein molecules at work. *J. Surf. Sci. Nanotechnol.* **3**, 384–392 (2005).
- Scheuring, S., Busselez, J. & Lévy, D. Structure of the dimeric PufX-containing core complex of rhodobacter blasticus by in situ atomic force microscopy. *J. Biol. Chem.* **280**, 1426–1431 (2005).
- Scheuring, S., Boudier, T. & Sturgis, J. N. From high-resolution AFM topographs to atomic models of supramolecular assemblies. *J. Struct. Biol.* **159**, 268–276 (2007).
- Asakawa, H. *et al.* Submolecular-scale imaging of α -helices and C-terminal domains of tubulins by frequency modulation atomic force microscopy in liquid. *Biophys. J.* **101**, 1270–1276 (2011).
- Trinh, M.-H. *et al.* Computational reconstruction of multidomain proteins using atomic force microscopy data. *Structure* **20**, 113–120 (2012).
- Chaves, R. C. *et al.* Conformational dynamics of individual antibodies using computational docking and AFM: Conformational dynamics of IGG using docking and AFM. *J. Mol. Recognit.* **26**, 596–604 (2013).
- Dasgupta, B., Miyashita, O. & Tama, F. Reconstruction of low-resolution molecular structures from simulated atomic force microscopy images. *Biochim. Biophys. Acta Gen. Subj.* **1864**, 129420 (2020).
- Amyot, R. & Flechsig, H. BioAFMviewer: An interactive interface for simulated AFM scanning of biomolecular structures and dynamics. *PLoS Comput. Biol.* **16**, e1008444 (2020).
- Niina, T., Matsunaga, Y. & Takada, S. Rigid-body fitting to atomic force microscopy images for inferring probe shape and biomolecular structure. *PLoS Comput. Biol.* **17**, e1009215 (2021).
- Amyot, R., Marchesi, A., Franz, C. M., Casuso, I. & Flechsig, H. Simulation atomic force microscopy for atomic reconstruction of biomolecular structures from resolution-limited experimental images. *PLoS Comput. Biol.* **18**, e1009970 (2022).
- Flater, E. E., Zacharakis-Jutz, G. E., Dumba, B. G., White, I. A. & Clifford, C. A. Towards easy and reliable AFM tip shape determination using blind tip reconstruction. *Ultramicroscopy* **146**, 130–143 (2014).
- Ruggeri, F. S., Sneideris, T., Vendruscolo, M. & Knowles, T. P. J. Atomic force microscopy for single molecule characterisation of protein aggregation. *Arch. Biochem. Biophys.* **664**, 134–148 (2019).
- Takenaka, H. *et al.* AFM tip characterizer fabricated by Si/SiO₂ multilayers. *J. Surf. Sci. Nanotechnol.* **9**, 293–296 (2011).
- Zeng, Z. *et al.* A simple method for AFM tip characterization by polystyrene spheres. *Ultramicroscopy* **108**, 975–980 (2008).
- Dongmo, L. S. *et al.* Experimental test of blind tip reconstruction for scanning probe microscopy. *Ultramicroscopy* **85**, 141–153 (2000).
- Tian, F., Qian, X. & Villarrubia, J. S. Blind estimation of general tip shape in AFM imaging. *Ultramicroscopy* **109**, 44–53 (2008).
- Józwiak, G., Henrykowski, A., Masalska, A. & Gotszalk, T. Regularization mechanism in blind tip reconstruction procedure. *Ultramicroscopy* **118**, 1–10 (2012).
- Bakucz, P., Krüger-Sehm, R. & Koenders, L. Investigation of blind tip estimation. *Rev. Sci. Instrum.* **79**, 073703 (2008).
- Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
- Schoenholz, S. S. & Jax, M. D. A framework for differentiable physics. *J. Stat. Mech.* **22**, 1–10 (2021).
- Greener, J. G. & Jones, D. T. Differentiable molecular simulation can learn all the parameters in a coarse-grained force field for proteins. *PLoS ONE* **16**, e0256990 (2021).
- Zhou, T., Cherukara, M. & Phatak, C. Differential programming enabled functional imaging with Lorentz transmission electron microscopy. *NPJ Comput. Mater.* **7**, 141 (2021).
- Villarrubia, J. S. *Morphological Estimation of Tip Geometry for Scanned Probe Microscopy* (Springer, 2022).
- Franchi, G., Fehri, A. & Yao, A. Deep morphological networks. *Pattern Recogn.* **102**, 107246 (2020).
- Nakashizuka, M., Ashihara, Y. & Iiguni, Y. *Morphological Regularization for Adaptation of Image Opening* Vol. 5 (Springer, 2022).
- Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification And Regression Trees* (Routledge, 2017). <https://doi.org/10.1201/9781315139470>.
- Innes, M. *et al.* Fashionable modelling with Flux. *Pattern Recogn.* <https://doi.org/10.48550/ARXIV.1811.01457> (2018).
- Schrödinger, LLC. *The PyMOL Molecular Graphics System.* (2015).
- Coureux, P.-D. *et al.* A structural state of the myosin V motor without bound nucleotide. *Nature* **425**, 419–423 (2003).
- Chen, Y. Elucidation and identification of double-tip effects in atomic force microscopy studies of biological structures. *JSEMAT* **02**, 238–247 (2012).
- Kodera, N., Sakashita, M. & Ando, T. Dynamic proportional-integral-differential controller for high-speed atomic force microscopy. *Rev. Sci. Instrum.* **77**, 083704 (2006).
- Fuchigami, S. & Takada, S. Inferring conformational state of myosin motor in an atomic force microscopy image via flexible fitting molecular simulations. *Front. Mol. Biosci.* **9**, 882989 (2022).
- Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2623–2631 (ACM, 2019). <https://doi.org/10.1145/3292500.3330701>.
- Sumikama, T., Foster, A. S. & Fukuma, T. Computed atomic force microscopy images of chromosomes by calculating forces with oscillating probes. *J. Phys. Chem. C* **124**(2213), 2218 (2020).
- Heath, G. R. *et al.* Localization atomic force microscopy. *Nature* **594**, 385–390 (2021).
- Loshchilov, I. & Hutter, F. *Decoupled Weight Decay Regularization.* (2017) <https://doi.org/10.48550/ARXIV.1711.05101>.
- Frames Catherine White *et al.* *JuliaDiff/ChainRulesCore.jl: v1.15.2.* (2022) <https://doi.org/10.5281/ZENODO.6835210>.
- Innes, M. *Don't Unroll Adjoint: Differentiating SSA-Form Programs.* <http://arxiv.org/abs/1810.07951> (2019).
- Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinform.* **54**, 10 (2016).

51. Kenzaki, H. *et al.* CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* **7**, 1979–1989 (2011).
52. Li, W., Wang, W. & Takada, S. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *Proc. Natl. Acad. Sci. USA.* **111**, 10550–10555 (2014).
53. Niina, T., Fuchigami, S. & Takada, S. Flexible fitting of biomolecular structures to atomic force microscopy images via biased molecular simulations. *J. Chem. Theory Comput.* **16**, 1349–1358 (2020).
54. Cossio, P. & Hummer, G. Bayesian analysis of individual electron microscopy images: Towards structures of dynamic and heterogeneous biomolecular assemblies. *J. Struct. Biol.* **184**, 427–437 (2013).

Acknowledgements

We thank Noriyuki Kodera for giving the HS-AFM data of myosin V walking and suggesting us to apply the method to a double tip case. We thank Tohru Niina for stimulating discussion on the blind tip reconstruction. We also thank Ryuhei Oshima for his help to prepare Fig. 1. This work was supported by JST CREST (Grant Number: JPMJCR1762 to Y.M. and S.T.), MEXT as “Program for Promoting Researches on the Supercomputer Fugaku” (Biomolecular dynamics in a living cell, Grant Number: JPMXP1020200101 to Y.M. and S.T.), JSPS KAKENHI (Grant Number: 20K21380 to Y.M.), and the Cooperative Research Program of “Network Joint Research Center for Materials and Devices” (Grant Number: 20221300 to Y.M.).

Author contributions

Y.M., S.F., S.T. wrote the manuscript. Y.M. developed the code for the method. Y.M., T.O. performed the twin experiments. Y.M., S.F., S.T. analyzed the real experimental data (Myosin V walking on an actin filament). All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-27057-2>.

Correspondence and requests for materials should be addressed to Y.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023