Imaging and three-dimensional reconstruction of chemical groups inside a protein complex using atomic force microscopy

Duckhoe Kim and Ozgur Sahin

Contents

Supplementary Sections

1. Additional control experiments: target blocking, sequence mismatch between probe and targets, and comparison of repeated images

Experimental results in Fig. 2 provide strong evidence to the specificity of the observed rupture forces (forces that exceed the threshold value). If the observed pulling forces were due to unspecific interactions, one would not expect the timing of these interactions to change dramatically with the sequence of target. In addition, the color uniformity seen in clusters of rupture events in Fig. 3 lends additional support to this notion. It is highly unlikely for unspecific interactions to result in the same rupture times within each cluster. Furthermore, data in Fig. 4 show that locations of rupture events are in agreement with predictions based on the crystal structure of the biotin-streptavidin complex.
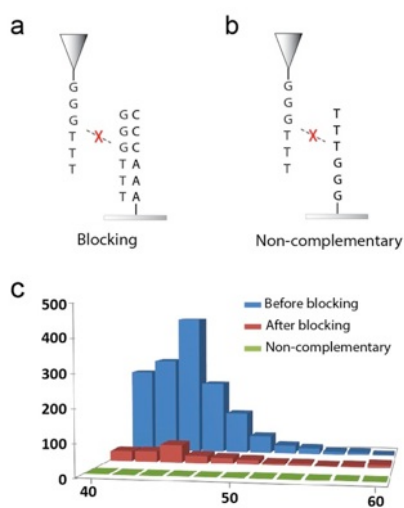

Figure S1. Blocking targets and imaging non-complementary surface dramatically reduce the frequency of events. (a) Scheme for the blocking experiment. (b) Scheme for the non-complementary experiment. (c) Force histograms obtained from complementary (blue), blocking (red), and non-complementary (green) experiments.

In addition to the above evidences, we performed separate experiments that are usually employed in force spectroscopy. First, we introduced complementary DNA to the imaging buffer and monitored the distribution of the magnitude of rupture forces. Because the targets are 6 bp long DNA strands, it is difficult to fully block them due to low melting temperature. Therefore we have used a very high concentration of complementary strands (~30 μM), which still does not ensure that all binding sites will be blocked. Nevertheless, a significant reduction in binding can be expected. The histograms in Fig.S1 show the imaging distributions before (blue) and after blocking (red). The frequency of observed rupture forces reduced substantially to 11% of the original value, in agreement with our prediction.

As a negative control, we imaged a surface having non-complementary DNA targets. In this case, no events above the noise threshold of 42 pN were observed. In the histograms in Fig. S1, we did not include force readings less than 40 pN. Because we record ~ 60K force readings, an RMS force noise of 10 pN can produce readings that reach 40 pN even without tip-sample interaction (freely vibrating cantilever).

Agreement between trace and retrace images: Repeated imaging of the same targets can provide an additional control experiment. Figure S6 shows examples of images obtained on

biotin-streptavidin complexes. With DNA surfaces, we have observed that binding probabilities are low. As a result of this, we generated a single image based on the rupture events recorded in both trace and retrace scans in Fig. 3. Two independent analyses predict that detecting a given target in one scan direction is approximately 50%. The first analysis is related to the distribution of the number of colored pixels in the ensemble of clusters, which should approximately obey Poisson's distribution. Therefore the ratio of the number of clusters with two elements to the number of isolated single rupture events ($\lambda/2 \approx 40\%$) is related to the binding probability ($1 - e^{-\lambda} \approx 55\%$). This value agrees with the estimation from the second analysis, in which we compared the locations of rupture events in the trace and retrace images. Only about 50% of the rupture events in the trace image lay within 2 nm of a rupture event with the same color in the corresponding retrace image. Considering the electrostatic interaction between immobilized DNAs, rupture events that appear in trace and retrace images that are within 2 nm are likely to belong to the same target molecule. Scanning slower can in principle increase binding probabilities, however sample drift limited how low we can reduce the scan rate. With the immobilized DNA samples, we used scan speeds of 400 nm/s and 0.4 nm/s in the fast and slow scan directions, respectively.

2. Additional multicolor images with different tip and sample combinations.

To examine the robustness of the multicolor imaging, we imaged samples with different tip and sample combinations. We first imaged a surface with A_6 and G_6 targets using a probe complementary to these targets. We also generated additional multicolor images of samples with T_6 and C_6 targets. Fig. S2b and Fig. S2c are generated with different tips. Fig. 3d of the main text and Fig. S2c are generated with the same tip on different regions of the same sample. The low density of targets in Fig.S2a is because of the relatively low concentration of DNA used for the surface preparation.

Statistical analysis of the clusters in Fig. S2c gave similar resolution estimate (0.49 nm, 0.58 nm) and Fig. S2b gave slightly higher values (0.8 nm, 0.96 nm). The color contents also exhibited uniformity within each cluster. 83% of the clusters in Fig. S2b and 85% of the clusters in Fig. S2c have single color.
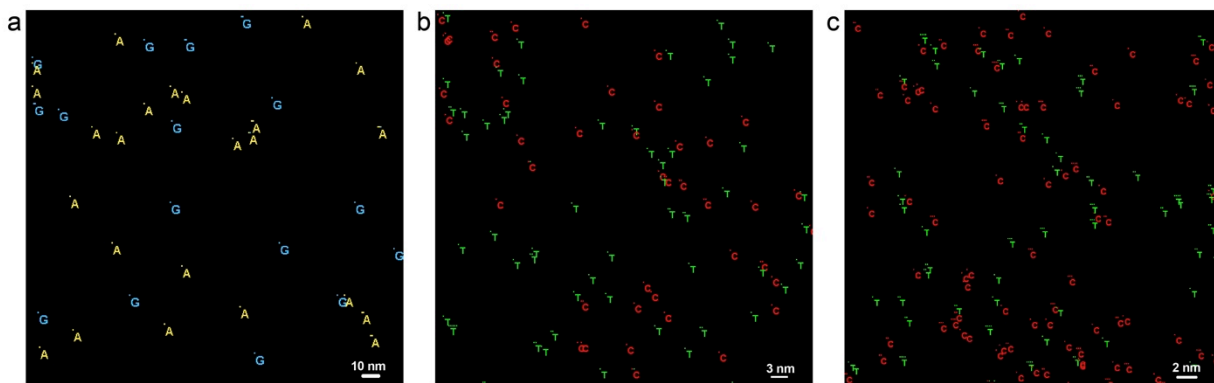

Figure S2. Multicolor imaging of DNA surfaces with different tip and sample sequences. (a) An Image of a sample with A_6 and G_6 targets. (b,c) Images of samples with T_6 and C_6 targets. Each target is color coded and marked with a letter, A, G, T, or C.

3. Entire dataset for biotin-streptavidin complexes

To analyze the distances between biotins incorporated into a streptavidin molecule, we collected images from 34 biotin-streptavidin complexes that showed more than two separate locations of biotins. The distance information from this collection is presented in Fig. 4f in the main text. Some of the images below exhibit three locations for targets. We obtained distance values from each pair of targets in these images.
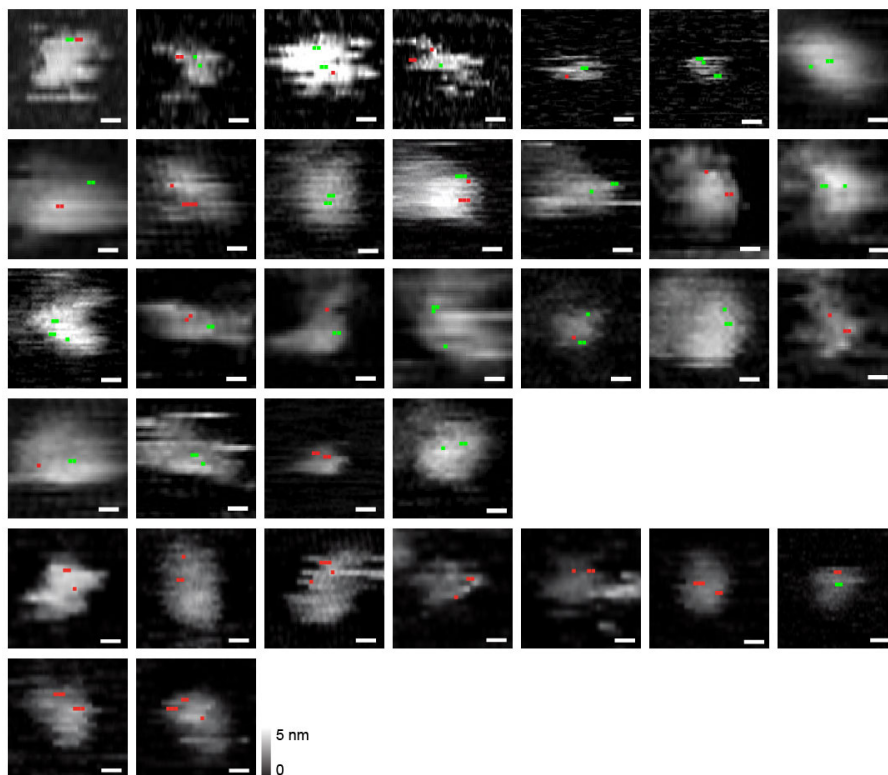

Figure S3. A collection of biotin-streptavidin complexes showing two or more locations for the labels. Samples in the top four rows are labeled with a mixture of green and red labels (sequences). The remaining samples are labeled with the red label only. Scale bars indicate 5 nm.

The greyscale topography images in Fig. S3 show relatively low resolution compared to the chemical force maps, which exhibit clustering of the detected rupture sites to regions approximately one nanometer in size. The differing resolutions in topography and chemical maps can be explained by the different spatial dependencies of the forces responsible for the two image channels. Similar behavior in simultaneously obtained images has also been observed in combined STM-AFM setups and in higher-harmonic AFM. For example, in combined STM-AFM, the very same tip records both the tunneling and the force images but these images exhibit

different resolutions, due to faster decay of tunneling current with distance[1] compared to tip-sample forces. In higher-harmonic AFM, the higher harmonic force components correspond to derivatives of the tip-sample force field, which also decay faster with distance when compared to the force at the fundamental frequency[2]. The situation with probe and target DNAs is similar, in that, the hybridization probability decays exponentially with Boltzmann's factor $e^{-x^2/2\sigma^2}$ (the square in x comes from the energy of bending). In addition, the rupture force can be expected to drop with distance approximately with the same Boltzmann factor since the energy barrier against rupture is lowered as a result of the added bending energy to the bound molecule. For example, a target DNA that is away from the probe is not only unlikely to bind, but it is also unlikely to produce a rupture force that passes the threshold. In contrast, the resolution in the topography image is determined by how the repulsive forces change with distance. In this case, the probe DNA is not the only contributor to the repulsive forces and therefore its resolution can be lower than the chemical map.

4. Images from biotin-streptavidin complexes labeled with non-complementary DNAs

To determine the contribution of noise and non-specific interactions to the rupture events detected on biotin-streptavidin complexes, we imaged biotin-streptavidin complexes in which biotins were labeled with non complementary DNA strands. For these measurements, we employed a cantilever that has detected recognition events on streptavidin molecules labeled with complementary sequences. We imaged more than 20 complexes and did not observe any rupture events.
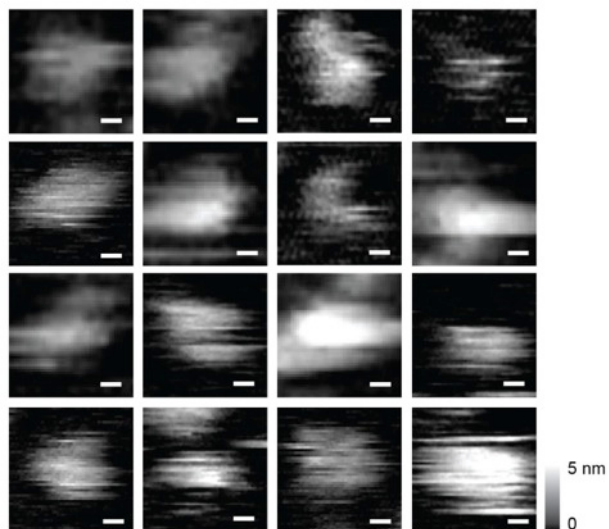

Figure S4. A selection of images from biotin-streptavidin complexes labeled with non-complementary DNAs. No rupture events are detected. Scale bars indicate 5 nm.

5. Imaging biotin-streptavidin complexes with different cantilever preparations

We observed that different preparations of cantilevers could reproduce rupture events on biotin-streptavidin complexes. We scanned the complexes with thirteen different cantilevers, and seven among them succeeded in the imaging of biotins. Selections of images that belong to different biotin-streptavidin complexes obtained with each of the five cantilevers are given in Fig. S5.
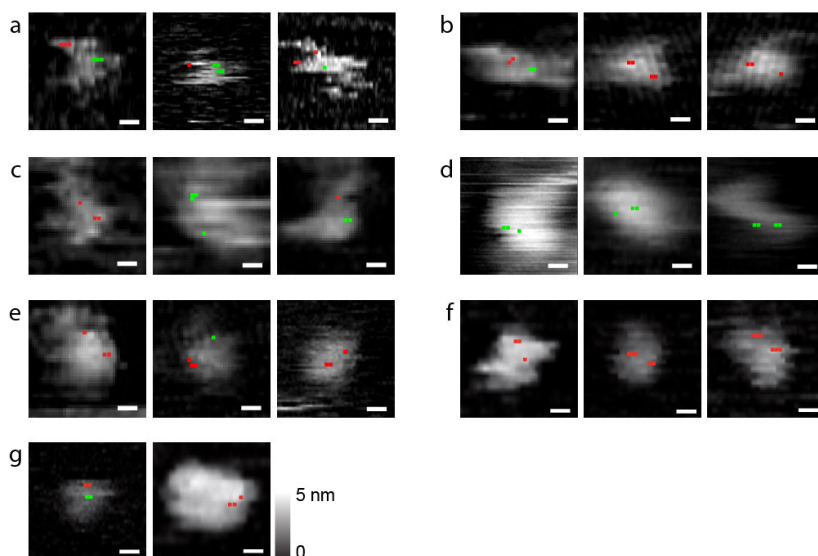

Figure S5. Seven out of thirteen cantilevers succeeded in imaging of biotins bound to streptavidin molecules. Sets of images in (a-g) represent different complexes obtained with each of the seven cantilevers. Scale bars indicate 5 nm.

We note the differences in greyscale topography images. While patterns show differences from one cantilever to the other, there is a noticeable consistency among images obtained with the same cantilever. These differences are likely due to variations in the location of the probe DNA relative to the tip apex, as well as the tip shape. Because we keep peak tapping forces below 10 pN, the probe DNA may serve as a very sharp tip for topography images, especially when it is situated at the apex.

6. Repeated imaging of the same biotin-streptavidin complexes

We imaged the same biotin-streptavidin complex repeatedly to observe reproducibility of the arrangements and color assignments. These images are given in Fig.S6. Each subset in Fig.S6 contains images of the same streptavidin. As seen from the data repeated imaging reproduced the colors and approximate locations of rupture events. Slight shifts in the locations (see for example Fig.S6h and Fig.S6i) may be due to movements of streptavidin molecules.
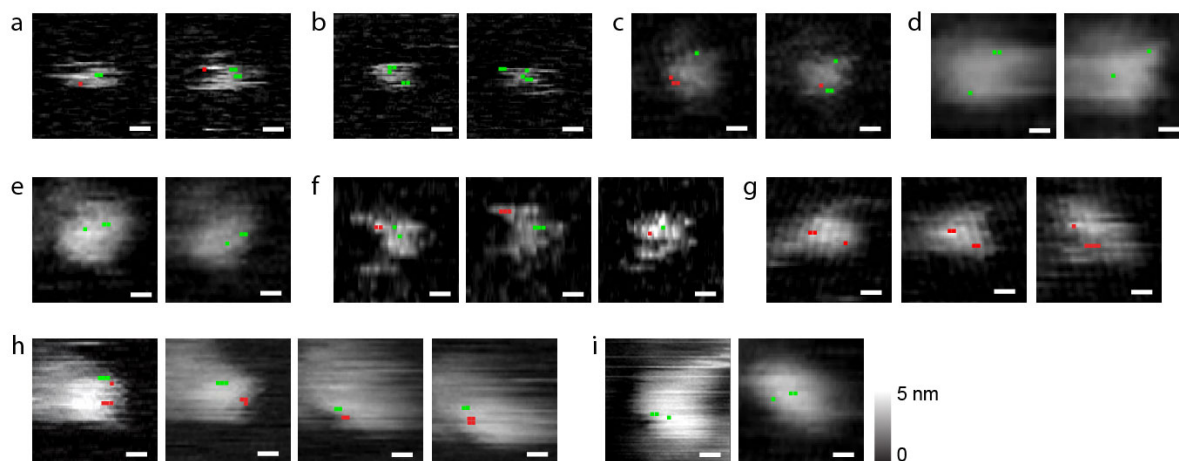

Figure S6. Repeated imaging of biotin-streptavidin complexes. (a-e) Trace (left) and subsequent retrace (right) images. (f,g) Trace, retrace and a secondary scan images given from left to right. (h,i) These images are recorded from different rounds of imaging of the same streptavidin molecules. Scale bars are 5 nm.

7. Biotin-streptavidin complexes displaying a single label DNA

We observed numerous biotin-streptavidin complexes that exhibit only one label. This could be due to several reasons. There may be one biotinylated DNA bound to the streptavidin or due to the orientation of the complex on the surface. Some of the label DNAs may have an orientation that makes it difficult to bind to the probe. It is also possible that a rupture event is not detected despite the presence of more than one label. A selection of these images is given in Fig.S8.
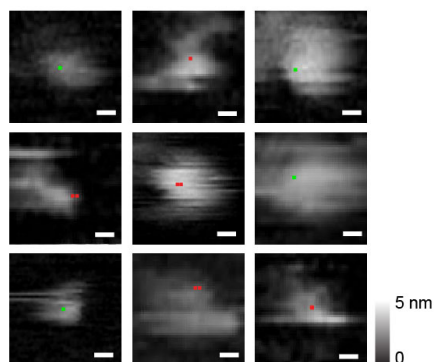

Figure S7. Biotin-streptavidin complexes showing only one label DNA. Scale bars indicate 5 nm.

8. Topography of immobilized streptavidin molecules

We mostly find a mixture of aggregates and isolated individual streptavidin molecules on the sample surface. To find isolated streptavidin molecules, we first scan a large area and then reduce the scan size to image individual streptavidin molecules. Figure S8 shows a typical topographic image of immobilized streptavidin molecules.
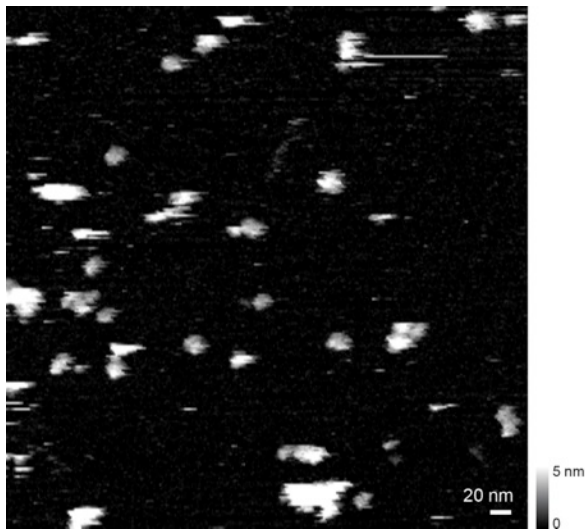

Figure S8. Topography of the sample surface displaying several streptavidin molecules.

9. The influence of the orientations of label-DNAs to the probability of detection

We introduced label-DNAs after deposition of streptavidin molecules. This makes it difficult for the label-DNAs to access to the binding sites facing the surface as shown in Figure S9a. Other orientations can allow additional binding sites as represented in Figure S9b. In these cases, the label-DNAs oriented towards the surface will not bind to the probe DNA easily, because it will require significant bending of the probe and label DNAs as depicted in Figure S9b. The energy required to bend short strands will reduce the binding probability and the rupture force. Due to these reasons it is unlikely to detect all four binding sites on a given streptavidin. This is in agreement with the observation that most of our images in Fig. S3 show two rupture locations and a small group of them show three rupture locations.

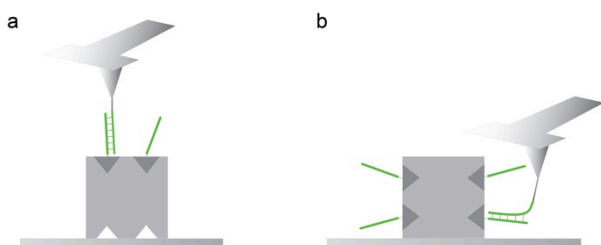**Figure S9.** Scheme for possible orientations of biotin-streptavidin complexes. (a) Two label-DNAs facing the top are both detectable to the probe DNA. Binding sites in the bottom are blocked by the surface. (b) A complex on its side. More than two label-DNAs can bind to the complex. However, detecting label-DNAs closer to the surface needs to overcome a significant mechanical bending energy.

10. Increased signal to noise ratio after cluster based filtering of the distance data

Because our method relies on a force threshold, there is a small but non-zero probability for noise to exceed the threshold at each pixel of a given image. Consequently, some of the colored pixels could be due to noise. However, these locations are unlikely to result in a cluster because the probability of force noise to exceed the threshold in two adjacent pixels is extremely low. Therefore, clusters of colored pixels are more reliable data points. If both of the target locations in a given image come from isolated colored pixels, then reliability of that particular distance measurement is substantially lower (i.e. if any one of the two locations is due to noise, then that data point raises the noise floor). By eliminating these images from our statistical analysis, we observed that the three peaks in the data became more significant. The density values in Fig. S10 correspond to the unfiltered (a) and filtered (b) data sets, respectively.
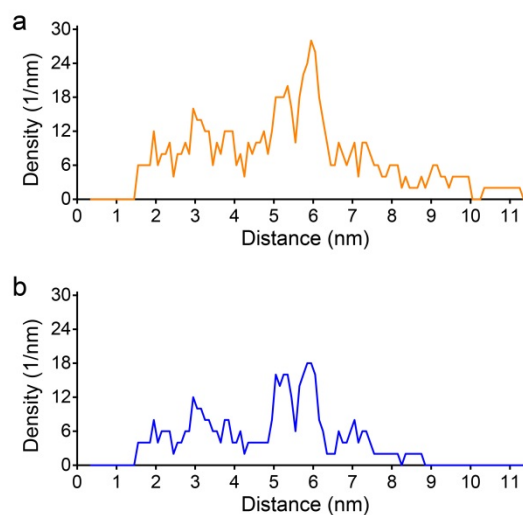

Figure S10. The density values calculated for the unfiltered (a) and filtered (b) data sets.

11. Discussion of the three-dimensional reconstruction process

The biotin-streptavidin complex has three two-fold symmetries in x-, y-, and z-axes. As a result, the locations of carboxylic acid groups of biotins form a tetrahedron with opposite edges having equal length. This means that the pairwise distances between the carboxylic acid groups of biotins can assume three values. Our rupture force maps allow determining distance values rather than the actual three-dimensional configuration. However, using simple algebraic equations one can determine the shape of the tetrahedron from the three distance values. The distances between the corners x_i, y_i, z_i of the tetrahedron would satisfy the following equations:

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 = r_{12}^2$$

$$(x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2 = r_{13}^2$$

$$(x_1 - x_4)^2 + (y_1 - y_4)^2 + (z_1 - z_4)^2 = r_{14}^2$$

Due to the translational degree of freedom, we set one of the corners as the origin:

$$x_1, y_1, z_1 = 0.$$

Due to the rotational degrees of freedom, we set one edge of the tetrahedron to be on the x-axis and another edge to lie in the first quadrant of the xy plane using the following equations.

$$x_2 = r_{12}, \quad y_2 = 0, \quad z_2 = 0.$$

$$z_3 = 0, \ x_3 > 0, \ y_3 > 0.$$

The set of equations and inequalities above would be satisfied by two mirror symmetric geometries. We have chosen the one that matches the symmetry of the known crystal structure.

Note that because we use label DNAs to detect biotins, the locations of rupture events correspond to these labels rather than the locations of the carboxylic acids. The detected locations would still form a tetrahedron with opposite edges having equal length. This new tetrahedron can be related to the one formed by the carboxylic acids, as we describe below.

In Figure S11a, green dots represent the corners of the tetrahedron defined by the three predominant distances seen in multicolor chemical force maps. Assuming that the DNA labels orient approximately normal to the surface of streptavidin (likely due to electrostatic repulsion) and further assuming that this direction can be approximated by the radial line pointing away from the geometric center of the streptavidin (due to the globular shape of this protein complex), one can translate the corners to the geometric center of the tetrahedron (the black dot in the middle) by a distance that would account for the length of the label DNA. To estimate this length, we consider that the binding between probe and target DNA would occur with the highest probability when their midpoints align. This is because bending of the probe and target would create an additional energy barrier against their binding. The configuration with the least amount
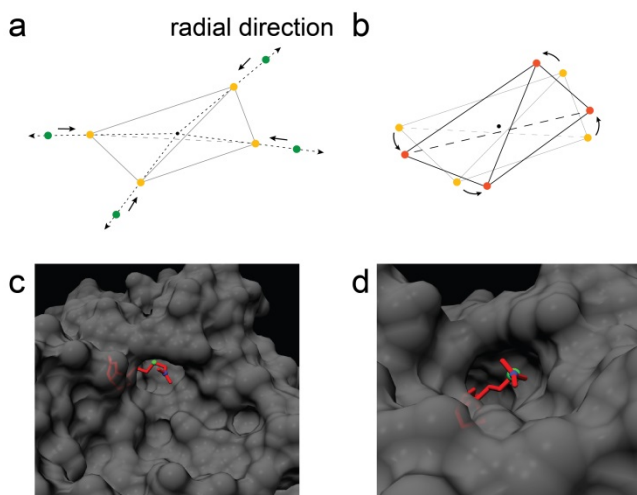
Figure S11. (a) Illustration of the tetrahedron corresponding to the location of DNA labels (green dots) and predicted locations of carboxylic acids (orange dots). Geometric center of the tetrahedrons are marked with the black dot in the middle. (b) A portion of the error in the prediction corresponds to the rotation of the structure. (c) Side view of the biotin within the binding pocket of streptavidin. The carbon (C1) in the carboxylic acid is highlighted with blue, and the second nearest carbon (C3) to C1 is highlighted with green. (d) Biotin is visualized from a viewpoint that aligns C1 and C3.

of bending is approximately when midpoints of the probe and target align (see the illustration in Fig. 4g). To account for the distance from the base of the label DNA to its midpoint, we used the contour length of 3-bp long double helical DNA in S-form (~11 Angstroms). The resulting coordinates would correspond to the predicted locations of the carboxylic acids.

In the above prediction, we neglected the role of the aliphatic linker that connects the DNA label to the carboxylic acid group of the biotins. Because the overall linker has a small size (contour length ~ 8.7 Angstroms), the error due to the linker length would not be prohibitive. In fact, the result in Fig. 4h shows that the predicted locations of the carboxylic acids deviate from the locations in the crystal structure by approximately 2 Angstroms, smaller than the length of the linker. This is likely due to the orientation of the linkers, which minimize the error in our prediction. Figure S10c shows the streptavidin surface defined by the van der Waals radii of individual atoms, along with the biotin shown in red. The linker extends from the carboxylic acid of the biotin, which will approximately be in the direction of carbons C1 and C3 (highlighted in blue and green colors in Figure S11c, respectively). We skipped C2 to account for the zigzag shape of the carbon chain. Figure S11d shows a closer view of the same protein region from a viewpoint that aligns C1 and C3 carbons. As seen from this picture, a linker oriented in this direction will not be interfered by the protein surface in a significant way. Because this orientation is approximately perpendicular to the radial line from the geometric center of streptavidin, the linker will not affect the radial position of the predicted locations of carboxylic acids significantly. Furthermore, part of the error due to the linkers is equivalent to a rotational transformation (illustrated in Figure S11b). Because two structures that can be related to each other with a rotational transformation are essentially identical, part of the error caused by the linkers is nullified in our comparison of the predicted locations with the crystal structure.

Possible influence of molecular deformations caused by the AFM tip: The heights and overall shapes of the biotin-streptavidin complexes (Fig. S3-S7) exhibit some variations. If these are due to deformations caused by the force of the AFM tip, then the influence of these deformations to the 3-D reconstruction process have to be considered. We neglected this influence for two reasons. First, the peak tapping forces were kept at approximately 10 pN. Indentation of streptavidin is likely to be negligibly at this force level. Second, regardless of the origin of the observed variations in the shape and heights of streptavidin molecules, these variations would have little effect on the reconstruction process. Our distance calculations depend solely on the rupture force maps. We neglect any possible height differences between the detected binding sites. This is because even if there is a height difference of 1 nm between two binding sites that are 3 nm apart in the image plane, the absolute distance between them will be 3.16 nm instead of 3 nm. The error will be even smaller for binding sites that are 5 nm apart (5.1 nm vs. 5.0 nm).

References

1. Giessibl, F. J. AFM's path to atomic resolution. *Materials Today* **8**, 32-41 (2005).
2. Giessibl, F. J. Higher-harmonic atomic force microscopy. *Surf.Interface Anal.* **38**, 1696-1701 (2006).