

Structural investigation of single biomolecules

NMR spectroscopy and X-ray crystallography are currently the most common techniques capable of determining the structures of biological macromolecules like proteins and nucleic acids at an atomic level of resolution. The atomic force microscope (AFM) is mostly known for its imaging capabilities, but it is also a very sensitive tool for quantitatively measuring forces on a single molecule level. This capability allows the use of AFM as an alternative technique for the investigation of the structural configuration of macromolecules under native conditions. Measuring forces within and between single molecules provides another approach to the understanding of molecular energy landscapes and chemical reaction kinetics. The AFM can measure individual bonds forming and breaking, rather than being limited to statistical averages over a population.

Molecular stretching experiments

Force spectroscopy is an AFM mode where the tip is moved up and down above a single point on the surface, while the deflection or other cantilever properties are measured. This gives a profile of the forces at different heights above the surface.

In molecular stretching experiments, the tip is brought into contact with the surface, the molecules on the substrate or the tip are allowed to interact with each other, and then the cantilever is retracted. This is shown schematically in the image series in Figure 1, for the case of the tip binding to a molecule from the surface and stretching it.

As the cantilever is moved away from the surface, the deflection exerts a pulling force. If a molecule is bound between the tip and the surface, then the deflection of the cantilever reflects the force exerted on the molecule as it is stretched.

The details of the molecular stretching depend on the structure of the molecule. It may have a defined three-dimensional folded structure, as for proteins, for example, or it may be a disordered chain, such as many other long polymers. The adhesive interaction between the tip and the molecule usually takes place in the repulsive contact part of the force curve, but the features of interest for the molecular stretching are found in the retract curve.

Long chain molecules, such as DNA or dextran, can be stretched between the tip and the sample. The stiffness and persistence length can be seen from the initial stretching of the molecule. Internal molecular transitions can also be studied at higher forces, such as the melting transition in DNA as the backbone rearranges under higher tension.

Molecules with a complex 3-dimensional structure, such as many proteins, can be unfolded in a controlled way so that the structural units can be investigated. Membrane proteins can be pulled out of the membrane, and the “popping” out of individual alpha-helices has been seen. Titin and bacteriorhodopsin are examples of proteins that have been intensively studied.

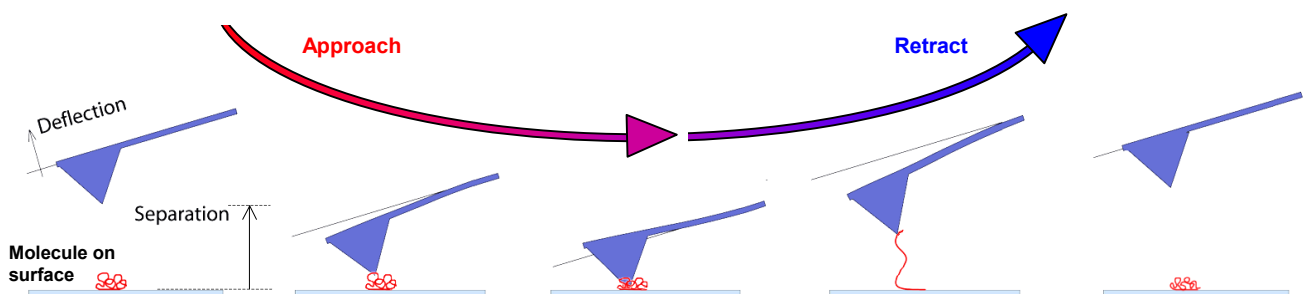


Fig. 1 Schematic diagram of a force spectroscopy experiment where the AFM tip is used to stretch a single molecule

Characteristics of macromolecules in solution

Long molecules in solution are rarely rod-like, and are usually highly flexible along the backbone. Even double-stranded DNA, a relatively "stiff" molecule because of the double-helix structure, is a random tangle for lengths over a few hundred nanometers. There is a large difference between the length of the molecule measured along the chain or backbone (contour length) and the typical dimensions in solution (e.g. the radius of gyration or the end-to-end length), which are much smaller [1].

A Low entropy conformation – unlikely in solution



B High entropy conformation – typical in solution



Fig. 2 Schematic diagram of two conformations of a long molecule.

Figure 2 shows a schematic representation of a long molecule in two different conformations. In A, the molecule is stretched out to a very extended conformation, where the end-to-end length is approximately the contour length. In B, the molecule is in a relaxed random tangle that occupies a much smaller lateral dimension than the contour length. The number of similar conformations is much higher than in A, and this high level of disorder or entropy gives these states a lower free energy.

This free energy difference translates to a force that resists the straightening of a molecule from B to A, and acts like an entropic "spring". Force is required to extend the molecule towards a straight conformation even if the bonds in the backbone are not being stretched.

Two basic models are commonly used for this entropic elasticity. In the freely jointed chain (FJC) model, the backbone is modelled as small units connected by completely "free" joints, so that adjacent units can have any angle between them without extra energy cost. The

worm-like chain (WLC) model is a continuous filament with chain bending stiffness built in, and is a better model for molecules such as double-stranded DNA. In the WLC model, the persistence length represents the length over which the initial orientation is randomised. Both models can be extended to higher force regimes by allowing for some chain elasticity, so that the length can increase beyond the initial contour length as the bonds are stretched. A brief overview of the models is given in the Appendix, more discussion can be found in reference [2].

Force-extension curves from AFM measurements with the JPK systems can be fitted automatically with these models in the data processing software. The main results are the contour length of the molecule being stretched, and the bending length, depending on the model

Proteins generally have a well defined (folded) three-dimensional structure in solution. This structure is the key to their biological function as enzymes or structural units, for example. The polypeptide chain is folded into a relatively stiff structure and held together by many bonds. If the protein is unfolded (by force, chemical or thermal changes), then the free polypeptide chain behaves like a simple linear molecule. Certain subunits of the protein structure, such as alpha-helices or globular domains, can be unfolded sequentially, and the measured "free" length of the molecule based on the stretching curves will increase with each unfolding event. The protein unfolding steps provide information about the normal folded structure and its failure mechanisms.

Xanthan – simple molecular stretching

Xanthan is a bacterial polysaccharide, with a molecular backbone identical to cellulose. The chemical properties of the side groups give the polymer many industrial applications, from food stabilization to oil recovery. When a Xanthan molecule is stretched between the tip and the surface, the behaviour is similar to a single (carboxymethylated) cellulose chain [3]. High molecular weight xanthan polymers can have lengths of several microns, and the stretching curves can be used as an example for simple molecular extension. For low forces,

the chain is stretched against the entropic spring. At some point, as the end-to-end length approaches the contour length, the elasticity of the backbone becomes important.

Figure 3 shows the retract curve of a single Xanthan molecule being stretched in a phosphate buffer solution, using the NanoWizard® AFM. The cantilever deflection has been calibrated using the thermal noise method [4]. The data is plotted such that a larger negative force corresponds to the tip being deflected more towards the surface. For fitting with WLC or FJC models, the tip-sample separation must be calculated first, to take account of the deflection of the cantilever. This is also included in the standard JPK data processing software, and the data here is shown as force against tip-sample separation.

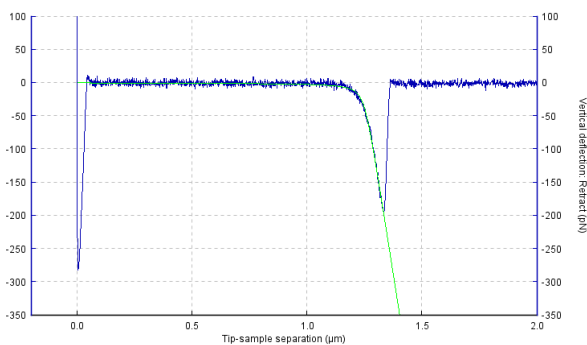


Fig. 3 Force spectroscopy retract curve of a single Xanthan molecule being stretched. A fit curve for the extended FJC model is also shown in green.

An extended FJC model fit curve is shown in Figure 3, fitted using a contour length of 1.24 microns. The fit for the simple FJC or WLC model was reasonable for small extension forces (around 50 pN in this case), but the curves diverged at higher forces. Hence the extended FJC model with the additional segment elasticity was used (see Appendix), to take account of the backbone stretching at higher forces.

In the case of Xanthan, a single long chain is stretched between the tip and the surface. Since the molecule is so long, the force to extend it is very small for much of the stretching curve, and the behaviour is dominated by the higher-force stretching where the stretched length is near the contour length. For more complex folded molecules,

such as proteins, a series of these stretching events are generally seen as different parts of the structure unfold and are extended, over shorter ranges.

Titin – sequential unfolding of subunits

Titin is a protein from muscle tissue, which consists of several repeated globular domains. Within the muscle tissue it is responsible for the mechanical properties. The AFM can be used to study the nano-mechanical properties of individual titin molecules.

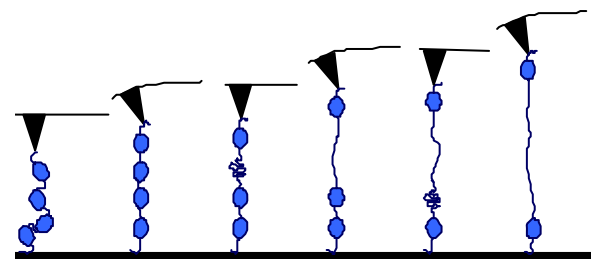


Fig. 4 Schematic diagram of titin stretching, as successive globular domains unfold.

As the molecule is stretched, at some point the bonds within a particular domain will fail. This globular domain unfolds, increasing the contour length of the molecule, and temporarily reducing the force. With further extension, one of the other domains will reach its failure point. This is illustrated schematically in Figure 4. As the protein is pulled, the domains "pop" open sequentially, producing a characteristic sequence of peaks in the deflection curves.

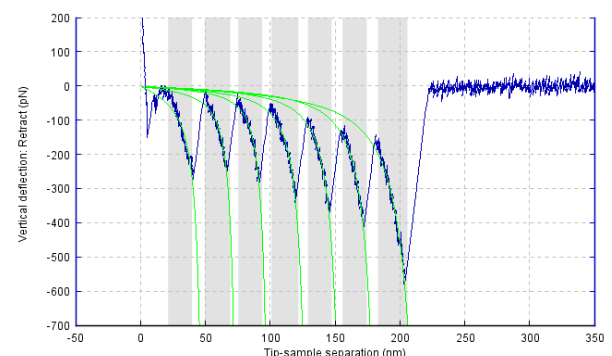


Fig. 5 Force curve of titin stretching, showing sequential unfolding of the globular domains, with fitted WLC extension curves

A typical titin extension force spectroscopy retract curve is shown in Figure 5. The characteristic sequence of unfolding events can be clearly seen. The curve of each successive extension part of the curve can be fitted with the FJC or WLC model to find the corresponding free amino acid length. The extended length of one domain (increase in contour length between successive peaks) is 28nm [5].

The JPK data processing software automatically recognises the jumps between different unfolding events, and a multi-fit routine automatically fits all the extension regions simultaneously with the selected fit model. In Figure 5, seven WLC fits are shown in green, for the seven Ig domains that have been unfolded.

Bacteriorhodopsin
– **extraction and unfolding**

Bacteriorhodopsin is an integral membrane protein from bacterial cells, which generates energy for the cells from light photons. There are seven transmembrane alpha-helices, with more disorganized peptide loops between them. When one end of the protein is grasped and pulled, then the alpha-helices are pulled out of the membrane, and they unfold as they are pulled out.

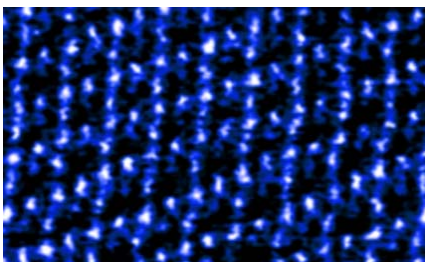


Fig. 6
Bacteriorhodopsin topography image 30 nm x 50 nm, 0.4 nm height range.

Purple membrane purified from bacteria consists of 25% lipids and 75% bacteriorhodopsin. The protein is arranged as trimers in a 2-dimensional crystal structure, which can be imaged using the AFM, as shown in Figure 6. The protein can be imaged and then individual proteins selected and pulled using the tip. The tip can be coated specially, and the molecules modified to pick up a particular part of the structure, or the tip can be pressed into the surface to pick up a molecule non-specifically [6].

A schematic diagram of the unfolding process is shown in Figure 7. The seven alpha-helices are shown with the AFM tip picking up one end (in the actual protein, the alpha-helices are grouped together in a bundle, not a line). There are many possible unfolding pathways, but the most likely ones involve the alpha-helices being pulled out in pairs. A progression of three partially unfolded states is shown Figure 7, with pairs of alpha-helices unfolding as they are extracted from the membrane.

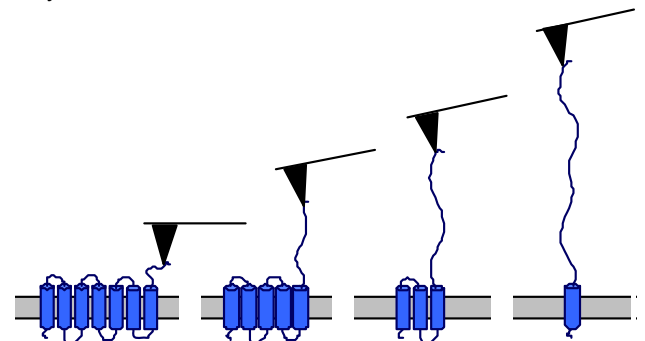


Fig. 7 Schematic diagram of bacteriorhodopsin unfolding, with intermediate states as the alpha-helices are pulled out in pairs.

Figure 8 shows a bacteriorhodopsin unfolding curve, with the force plotted against tip-sample separation. There is some short-range adhesion at the surface, followed by three main stretching and unfolding events, corresponding to the three states shown in Figure 7. The main peaks are around 20, 40 and 60 nm, as expected from the literature [6]. The cantilever was calibrated using the thermal noise method [4], and the experiments carried out under buffer (10mM TRIS, 150mM KCl, pH 7.6).

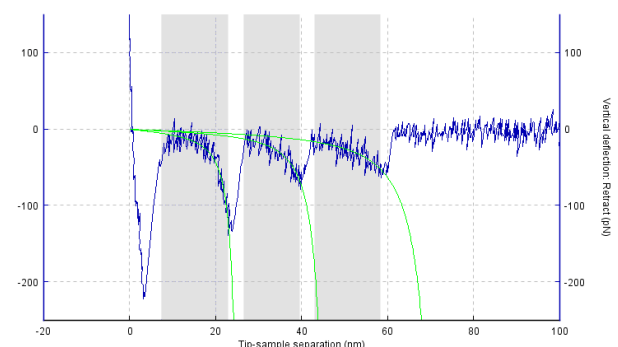


Fig. 6 Unfolding curve of bacteriorhodopsin being pulled out of purple membrane. Three worm-like chain (WLC) fits are shown in green superimposed on the data.

The exact position of the bond failure events varies from curve to curve, since the bonds have an energy in the range of the thermal energy at room temperature. For a bond with energy near the thermal energy, there is some probability of it failing within a certain time, even with no applied force. This corresponds to the normal "off-rate" seen for the binding reaction free in solution. For bonds with energies further above the thermal energy, which are normally stable, the probability of them failing spontaneously increases as a force is applied [7].

One way to think of a stretching event, therefore, is that the bonds are strained until the energy barrier to unfolding is within the thermal range. The actual moment when the bonds suddenly fail depends on a random thermal fluctuation, which takes the molecule over the unfolding energy barrier. The time dependence means that the measured unfolding rate will depend on the loading rate, or the speed with which the tip is pulling the free end of the molecule [7,8].

The unfolding curves also show different pathways for the molecule to unfold and pull out of the membrane. From a knowledge of the protein structure and amino acid sequence, the extensional curves can be fitted to show clearly which unfolding path is taken in a particular force curve [9,10].

Applications

A wide variety of proteins can be stretched and unfolded using the AFM, to gain information about both the normal protein structure and its failure modes. In the case of bacteriorhodopsin, the protein forms very highly packed structures in the bacterial cell wall, and so is one of the few membrane proteins that can be crystallized for high-resolution structural determination.

This method can be more generally applied, however. For most membrane proteins, the amino acid sequence is much easier to determine than the folded structure, since most membrane proteins are not suitable for crystallization. Unfolding events measured with the AFM, which show the change in amino acid length as different structural units unfold, can be used with the known sequence to interpret

the tertiary structure of membrane proteins. Specific coating of the probe, and modification of the protein to have specific "pick-up" points can lead to pulling the protein at different locations, and hence further information about how the structural units are connected together in the 3-D structure [6,9,10].

Thus AFM is able to give insight into the configuration of a single molecule and allow an independent measurement of molecular properties to refine molecular modelling techniques. AFM can also be combined with single-molecule fluorescence techniques such as TIRF or FRET.

This report concentrates on the use of the AFM to manipulate single molecules to extract information about the molecular structure or conformation, and intramolecular binding forces. It is also possible to study time-dependent phenomena, such as intramolecular dynamics in macromolecules, molecular recognition or protein folding.

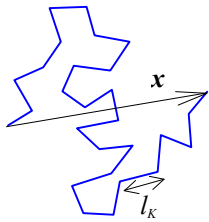
Literature

- [1] Israelachvili J. "Intermolecular and surface forces", Academic Press, London ISBN 0-12-375181-0.
- [2] Janshoff A., Neitzert M., Obersörfer Y., and Fuchs H. "Force Spectroscopy of molecular systems – single molecule spectroscopy of polymers and biomolecules". *Angew. Chem. Int. Ed.* 39:3212-3237 (2000).
- [3] Li H., Rief M., Oesterhelt F. and Gaub H.E. "Single-molecule force spectroscopy on Xanthan by AFM". *Advanced Materials* 3(4): 316-319 (1998).
- [4] Hutter J.L. and Bechhoefer J. "Calibration of atomic-force microscope tips". *Rev. Sci. Instrum.* 64:1868-1873 (1993).
- [5] Rief M., Gautel M., Österhelt F., Fernandez J.M. and Gaub H.E. "Reversible unfolding of individual titin immunoglobulin domains by AFM" *Science* 276:1109-1111 (1997).
- [6] Oesterhelt F., Oesterhelt D., Pfeiffer M., Engel A, Gaub H.E. and Müller J. "Unfolding pathways of individual bacteriorhodopsins". *Science* 288:143-145 (2000).
- [7] Evans E. "Probing the relationship between force – lifetime – and chemistry in single molecular bonds" *Annu. Rev. Biophys. Biomol. Struct.* 30:105-128 (2001).
- [8] Evans E. and Ritchie K. "Dynamic strength of molecular adhesion bonds" *Biophys. J.* 72(4):1541-1555 (1997).
- [9] Kedrov A., Ziegler C. and Müller D.J. "Differentiating ligand and inhibitor interactions of a single antiporter" *J. Mol. Biol.* 362:925–932 (2006).
- [10] Kedrov A., Appel M., Baumann H., Ziegler C. and Müller D.J. "Examining the dynamic energy landscape of an antiporter upon inhibitor binding" *J. Mol. Biol.* 375:1258–1266 (2008).

Appendix: Theory

For more details, see the review by Janshoff et al. [1].

Freely jointed chain (FJC):



Chain made from rigid links, with free joints between them.

Chain link size (Kuhn length) l_k
Contour length $L = n l_k$

End-to-end length x

$$F(x) = \frac{k_B T}{l_p} \left[\frac{1}{4} \left(1 - \frac{x}{L} \right)^{-2} + \frac{x}{L} - \frac{1}{4} \right]$$

Fit parameter: Persistence length, l_p

This model can also be extended to include the elasticity of the polymer backbone.

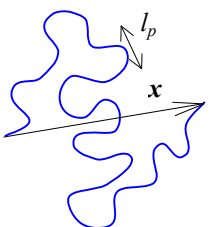
$$x(F) = L \left[\coth \left(\frac{F l_k}{k_B T} \right) - \frac{k_B T}{F l_k} \right]$$

Fit parameter: Kuhn length, l_k

The model can be extended to cover elastic deformation of the backbone, so that it is possible to have an end-to-end length $x > L$. The equation is modified with a segment elasticity, κ , to become:

$$x(F) = L \left[\coth \left(\frac{F l_k}{k_B T} \right) - \frac{k_B T}{F l_k} \right] \left(1 + \frac{F}{\kappa L} \right)$$

Wormlike chain (WLC):



Continuous filament, with resistance to bending.

Average length over which the direction becomes random, (persistence length) l_p

Contour length L
End-to-end length x