

Automated alignment and pattern recognition of single-molecule force spectroscopy data

M. KUHN*†, H. JANOVJAK†, M. HUBAIN† & D. J. MÜLLER†

*California Institute of Technology, Division of Computer Science, Pasadena, California, 91125, USA

†Center of Biotechnology, University of Technology Dresden, Tatzberg 47–51, 01307 Dresden, Germany

Key words. Molecular interactions, protein folding, statistical analysis.

Summary

Recently, direct measurements of forces stabilizing single proteins or individual receptor–ligand bonds became possible with ultra-sensitive force probe methods like the atomic force microscope (AFM). In force spectroscopy experiments using AFM, a single molecule or receptor–ligand pair is tethered between the tip of a micromachined cantilever and a supporting surface. While the molecule is stretched, forces are measured by the deflection of the cantilever and plotted against extension, yielding a force spectrum characteristic for each biomolecular system. In order to obtain statistically relevant results, several hundred to thousand single-molecule experiments have to be performed, each resulting in a unique force spectrum. We developed software and algorithms to analyse large numbers of force spectra. Our algorithms include the fitting polymer extension models to force peaks as well as the automatic alignment of spectra. The aligned spectra allowed recognition of patterns of peaks across different spectra. We demonstrate the capabilities of our software by analysing force spectra that were recorded by unfolding single transmembrane proteins such as bacteriorhodopsin and NhaA. Different unfolding pathways were detected by classifying peak patterns. Deviant spectra, e.g. those with no attachment or erratic peaks, can be easily identified. The software is based on the programming language C++, the GNU Scientific Library (GSL), the software WaveMetrics IGOR Pro and available open-source at <http://bioinformatics.org/fskit/>.

Introduction

Force spectroscopy experiments performed using atomic force microscopy (AFM) allow the measurement of molecular forces that stabilize single proteins, form receptor–ligand bonds, control antibody–antigen binding, or mediate cellular adhesion (Dammer *et al.*, 1996; Rief *et al.*, 1997; Fritz *et al.*, 1998;

Benoit *et al.*, 2000; Oesterhelt *et al.*, 2000). To this end, single proteins, receptor–ligand complexes, or cells are tethered between the tip of the AFM cantilever and a supporting surface. The tip–sample separation is then continuously increased with a piezoelectric actuator. Recording the force as a function of tip–surface separation generates a force–distance (F–D) curve characteristic for the molecule(s) under investigation. In initial experiments, Rief and coworkers applied single-molecule force spectroscopy to the muscle protein titin consisting of repeated globular immunoglobulin and fibronectin domains (Rief *et al.*, 1997, 1998). The continuous extension of the protein construct resulted in the subsequent unfolding of the globular domains. From the F–D spectrum the force required to unfold each domain as well as unfolding pathways and unfolding kinetics were determined (Oberhauser *et al.*, 1998; Williams *et al.*, 2003). The combination of single-molecule imaging and force spectroscopy yielded surprisingly detailed insights into inter- and intramolecular interactions of membrane proteins (Oesterhelt *et al.*, 2000; Kedrov *et al.*, 2004). It has been shown that structural elements of membrane proteins unfold sequentially and that their stability depends on the physiological environment of the protein (Janovjak *et al.*, 2003; Müller *et al.*, 2002).

F–D or force vs. time curves are also obtained with alternative force measurement methods, such as optical tweezers and the biomembrane force probe (Leckband & Israelachvili, 2001). In the future, high throughput force spectroscopy assays will permit the screening of physiological relevant parameters, such as pH, electrolyte concentration, and temperature (Albrecht *et al.*, 2003; Janovjak *et al.*, 2003). These factors alter the inter- and intramolecular interactions controlling structure, function and assembly of biological macromolecules. In order to obtain statistically relevant results, many single molecules of the same species need to be studied. Thus, a large number of F–D curves have to be recorded and analysed. The latter demands a data analysis approach that offers fully automated processing of many datasets with identical analysis procedures. To discriminate F–D curves showing specific and unspecific interactions and different unfolding/

Correspondence: Daniel Müller PhD. Tel.: +49 351 463 40330; fax: +49 351 463 40324; e-mail: mueller@biotec.tu-dresden.de

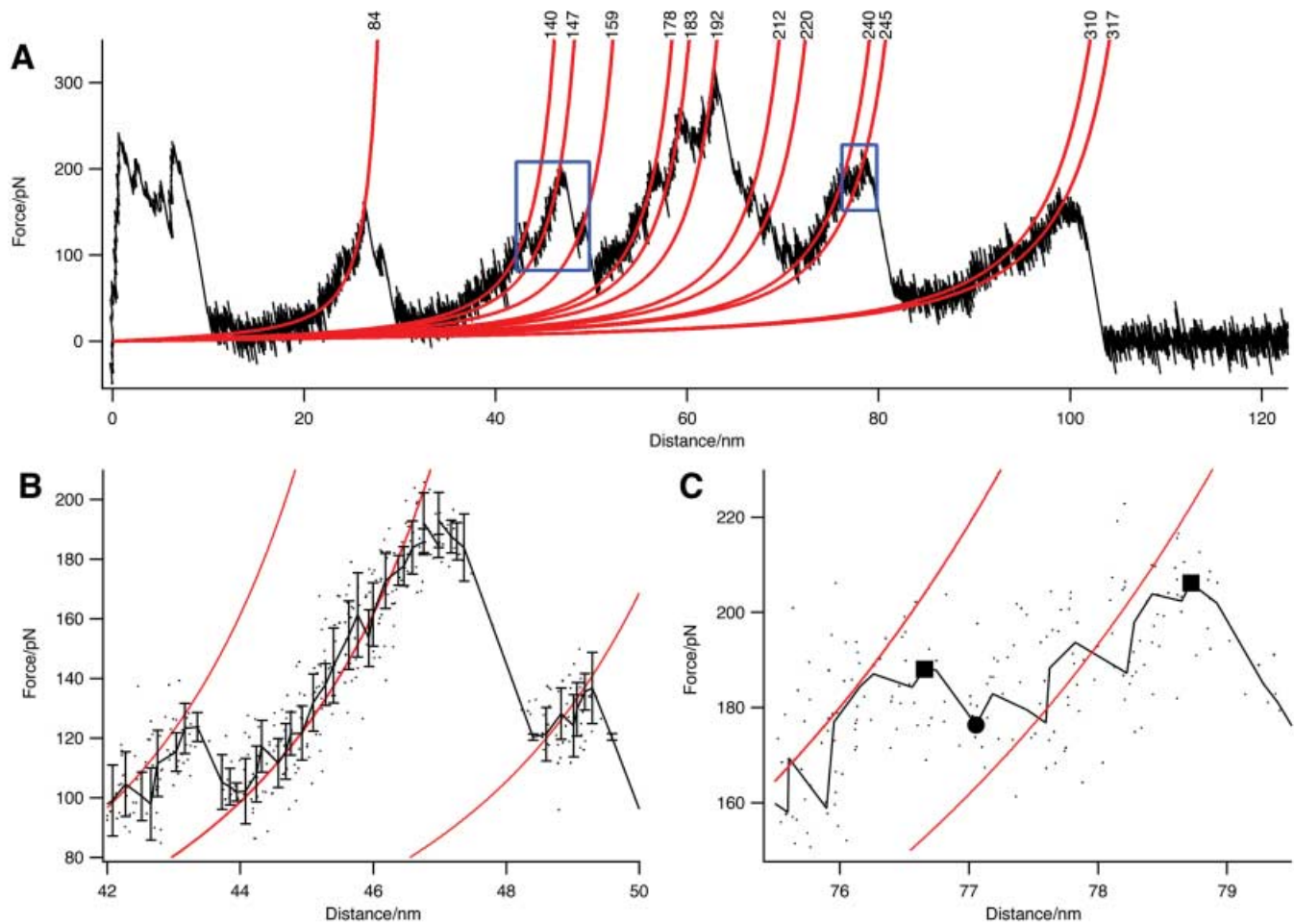


Fig. 1. Automated fitting of a single molecule force spectrum of the membrane protein NhaA. (A) Thirteen unfolding events were detected and fitted using the WLC model (red curves, numbers represent contour lengths in amino acids). Selected regions were enlarged in (B) and (C). (B, C) Raw data points were shown as dots. The smoothed spectra (black curves) were used for fitting. In (B), the error bars indicated the SD of the measured data points around the smoothed (i.e. averaged) spectrum. In (C), the squares denoted local force maxima and the circle a local force minimum, as determined by the peak detection routine.

unbinding pathways, classification and pattern recognition algorithms have to be applied. Although various off-line analysis software methods have been proposed (Saxton, 1996; Kasas *et al.*, 2000; Carl *et al.*, 2001; Gergely *et al.*, 2001; Carl & Dalhaimer, 2004), suitable pattern recognition algorithms are currently not available for the analysis of the experimental data. Therefore, we developed a software solution allowing high-throughput classification and statistical analysis of force spectra. The program applies alignment and pattern recognition procedures to reveal force statistics and reaction pathways.

Methods

Determining zero-force base line and contact point

The non-contact part of the force spectrum was used to determine and correct the slope of the baseline and the force offset.

In order to determine the length of the non-contact part the standard deviation (SD) of a linear fit to the last 5 nm of the force trace was computed. (The user can adjust this value and all other threshold values and ratios.) Subsequently, the fitting range was extended towards the force curve origin until the SD rose above a threshold of 1.5 times the minimum encountered SD. The rise in SD indicated that the last force peak was reached and the non-contact part ended. To find the contact point (i.e. the point at which the undeflected cantilever was in contact with the sample surface) the first intersection of the fitted baseline with the spectrum was determined.

Smoothing and peak detection

In order to even out noise, the spectrum was smoothed combining data points from equally spaced intervals of 0.1 nm. For example, in the spectrum of Fig. 1 the noise around a

linear fit to the baseline initially had a SD of 12.0 pN. After smoothing, the relevant part of the spectrum (extensions > 15 nm) was reduced from 5300 points to 830 points, yielding a SD of 4.4 pN (Fig. 1B,C). Peaks were detected by monitoring force differences between local minima and maxima. Any peak above 30 pN was defined as a force maximum if it was preceded and followed by force minima, which had a force difference of at least twice the SD of the baseline to the maximum (e.g. 8.8 pN in Fig. 1). By default, peaks occurring within the first 15 nm of the spectrum (measured from the contact point) were not fitted to avoid contributions of non-specific interactions. All peaks were fitted with the given polymer extension model using the Levenberg–Marquardt algorithm as implemented in the GNU Scientific Library. We implemented the worm-like chain (WLC) model but other models can be easily added by creating a C++ class with information about the F–D function and its derivatives. If the fits of two or more peaks had contour lengths that were separated by less than 1.5 nm, they were merged and fitted again. After fitting each peak, the program stored contour length and rupture force in a text file.

Alignment of spectra

Different pick-up points on the polypeptide chain and other factors can lead to a misalignment of the F–D curves. The alignment procedure that we used followed an approach similar to hierarchical tree clustering used in cluster analysis (Yeung *et al.*, 2003). The result was a binary tree, i.e. each node either had two or zero nodes as children. At first, an alignment score was calculated for each pair of spectra. Secondly, a hierarchical tree of spectra was built and used to determine the offsets for alignment. To calculate the alignment score, each spectrum was approximated by a sequence of normal distributions representing the density of data points within an interval of 0.1 nm. Each interval was described by two parameters: mean (μ) and SD (σ) in units of force. Arithmetic mean and SD were calculated without assigning weights to any points. Thus, for each extension interval, a probability density function p depending on the force y was defined as:

$$p(y, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (1)$$

Given two distributions, their overlap f was calculated as the integral of the product of the probability density functions:

$$\begin{aligned} f(\mu, \sigma, \mu', \sigma') &= \int_{-\infty}^{\infty} p(y, \mu, \sigma) \cdot p(y, \mu', \sigma') dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma'\sqrt{2\pi}} e^{-\frac{(y-\mu')^2}{2\sigma'^2}} dy \quad (2) \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma^2 + \sigma'^2}} e^{-\frac{(\mu-\mu')^2}{2(\sigma^2 + \sigma'^2)}} \end{aligned}$$

The quality of an alignment $a_{i,j}$ with (discrete) distance offset δ represented the sum over the overlaps calculated for all data points x . It was weighted by the geometric mean of the spectral lengths l and l' to normalize the score for spectra showing different pulling lengths.

$$a_{i,j}(\delta) = \sqrt{\frac{2\pi}{l \cdot l'}} \sum_x f(\mu_i(x), \sigma_j(x), \mu_j(x + \delta), \sigma_j(x + \delta)) \quad (3)$$

The constant factor $\sqrt{2\pi}$ was dropped, because it did not influence the relative scores. The alignment score $s_{i,j}$ represented the maximum score over all possible offsets between the spectra i and j . Therefore, the corresponding δ yielded the best alignment.

$$s_{i,j} = \max_{\delta} a_{i,j}(\delta) \quad (4)$$

After the alignment score had been calculated for all spectral pairs, hierarchical clustering was applied to determine the relative offset of the spectra. The two closest spectra, i.e. those with the highest alignment score, were combined to form a node. All alignment scores relative to these two spectra were mapped onto the node. Each of the remaining spectra had two alignment scores to the pair that just had been formed. Of these, the higher score was retained as the alignment score between that spectrum and the new node. This corresponds to single or nearest-neighbour linkage in cluster analysis. While the hierarchical tree was built by repeating the combination step, the distance offsets yielding the alignment scores of each spectrum were applied to the spectra. Lastly, all spectra were shifted along the distance axis such that the average contact point coincided with the origin. Thus, the average contact point is the starting point for WLC curves that are drawn to the aligned spectra.

Peak classification

Force peaks were classified into peak classes according to their contour lengths. The classification routine started by grouping the closest pair of single peaks (i.e. the one with the smallest contour length difference across all spectra) into a new peak class. The contour length of a peak class represented the average of all single peak contour lengths contained in that class. Then, the next closest pair of peaks was merged. Either two single peaks were combined to a new class, or a single peak joined an existing peak class, or two peak classes were merged. This process of combining peaks was repeated until no further classes could be found. This was the case when all single peaks and peak classes were separated by more than 1.8 nm (corresponding to five amino acid residues of 0.36 nm length). Only single peaks from different spectra could be grouped into the same peak class. When a combination step combined two single peaks of the same spectrum into the same class, both peaks were temporarily removed from the new class.

Subsequently, the peak lying closer to the centre of the class joined it, while the other peak was not allowed to join this class anymore.

User interface

A graphical user interface for Microsoft Windows and Mac OS X is provided to operate the program. Additionally, a command line version can be compiled with a standard C compiler on most platforms. A plug-in for WaveMetrics IGOR Pro is available with the following functionality: importing and exporting spectra, loading alignment information, plotting force curves, and selecting spectra by common peaks. The software is based on the programming language C++, the GSL, WaveMetrics IGOR Pro and available for download at <http://bioinformatics.org/fskit/>.

Results and discussion

Three-step analysis of individual force traces

The software program developed allowed an automated analysis of large numbers of F–D traces. Single spectra were analysed for their individual force peaks as well as aligned against and compared with each other. To demonstrate the capabilities of our algorithms, we have chosen single-molecule unfolding curves recorded on the membrane proteins bacteriorhodopsin (BR) from *Halobacterium salinarum* (Oesterhelt *et al.*, 2000; Müller *et al.*, 2002) and the sodium-proton antiporter NhaA from *Escherichia coli* (Kedrov *et al.*, 2004). In contrast to most globular proteins, membrane proteins follow many different unfolding pathways yielding complicated force spectra exhibiting main and shoulder peaks (Müller *et al.*, 2002).

In a first step each single trace was automatically analysed according to the following procedure: (1) finding the zero-force baseline and the contact point (the point where the undeflected cantilever touches the surface); (2) smoothing the spectrum; (3) detecting and fitting force peaks using the WLC model. In the example given (Fig. 1A), 13 peaks were detected and fitted. The details of the smoothing and peak detection process were presented for two regions of the same force trace (Fig. 1B,C). In (B) and (C), black lines represented the smoothed curve derived from the original data (black points). In (C), the force difference between the local force maximum of the first peak (square) and the following local minimum (circle) was 11.7 pN. This force spectrum was typical for single-molecule experiments and emphasized the need for a low peak detection threshold (8.8 pN in this case) to detect main and minor force peaks (see Methods).

Force curve alignment

Among several approaches suitable to align spectra, three were tested. At first, force spectra were aligned using the contour

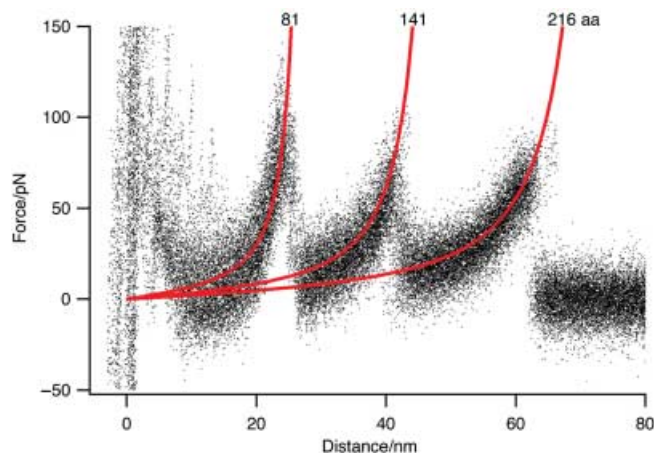


Fig. 2. Automated force curve alignment. Nineteen force spectra recorded on single BR molecules were superimposed using the X-offsets obtained by clustering and aligning similar spectra. The superimposition enhanced common unfolding patterns observed among the individual molecules. The contour lengths of WLC fits (red curves) and the contact point represented averages of individual contour lengths and contact points, respectively. While the absolute contour lengths varied depending on the choice of the contact point (Müller *et al.*, 2002, and Fig. 4), the relative contour lengths were more preserved.

lengths of the force peaks. This approach failed because the peaks of typical F–D curves showed high variations in absolute and relative contour lengths as the point of contact between protein and AFM tip was variable (Rief *et al.*, 1997; Oesterhelt *et al.*, 2000). Secondly, an alignment procedure using a two-dimensional histogram containing a pixel image of the force spectrum was developed. Two spectra were aligned by shifting their pixel images against each other and calculating an alignment score for each position, similar to cross-correlation analysis. Although this approach yielded promising results, a higher density of data points that was frequently observed at the base of peaks had a negative impact on the alignment quality.

To overcome this limitation of different data point densities we converted the data points to an equally spaced representation in the distance axis: the original distribution of data points in a short segment of the spectrum was approximated by a normal distribution with the two parameters arithmetic mean and SD (see Methods). This representation made it possible to determine the best alignment of two spectra by shifting two vectors against each other. The distance offset yielding the best alignment was stored together with the corresponding alignment score for all pairs of spectra. Then a superposition of the spectra was obtained by clustering the spectra according to their alignment score (i.e. similarity). At first, the two most similar spectra were aligned using their determined distance offset, thereby forming a cluster. Depending on their similarity, the remaining spectra either created new clusters or joined

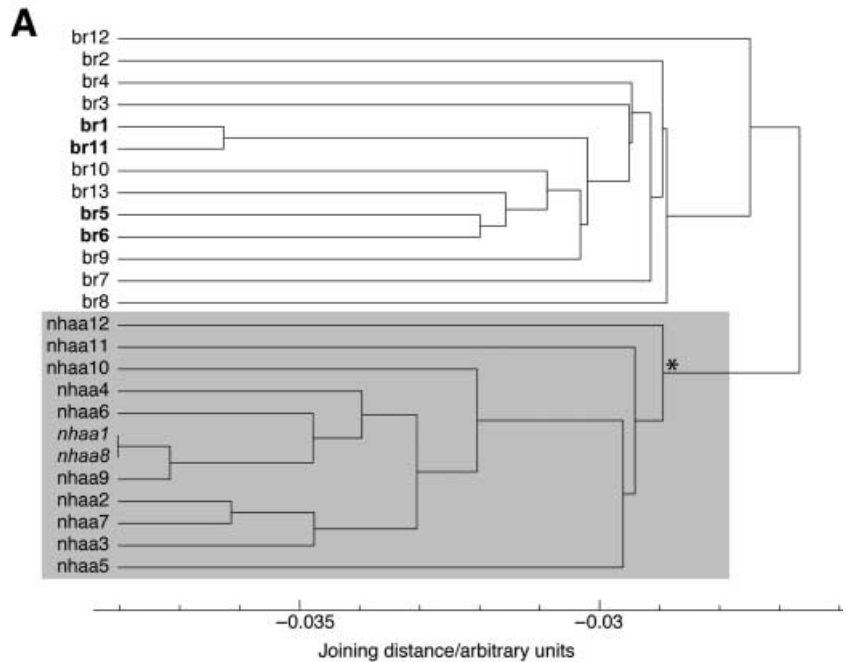
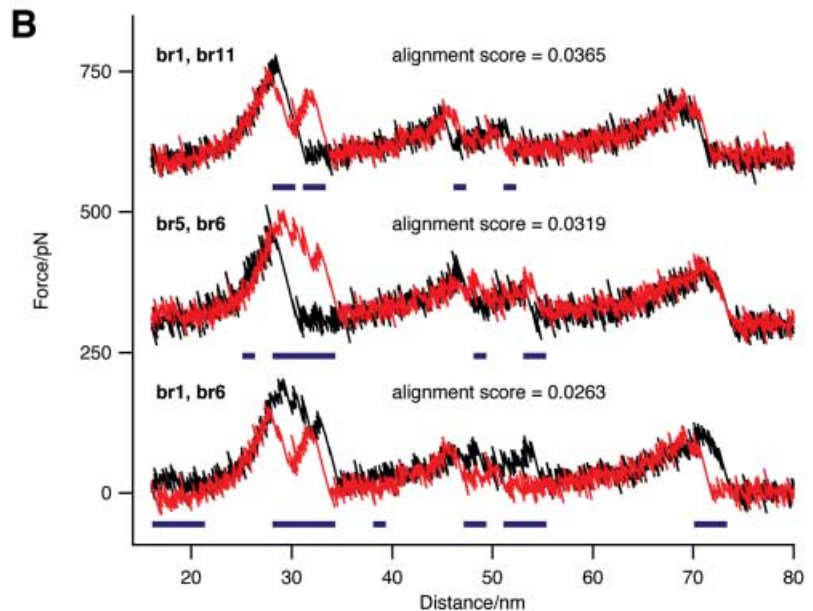


Fig. 3. Hierarchical tree built during alignment of force spectra. (A) Each curve name corresponded to a single NhaA or BR trace. The alignment routine grouped spectra for each of the two proteins together. The abscissa represented the distance between two spectra in arbitrary units, corresponding to the negative alignment score calculated by superimposing pairs of spectra. Each vertical line corresponded to a node (e.g. as marked by the asterisk). The horizontal location of each vertical line showed the distance between the spectra of a node. For example, the two closest spectra are *nhaa1* and *nhaa8* with a distance of about -0.038 , while the most distant NhaA spectra are joined at a distance of about -0.029 (marked by the asterisk). NhaA spectra were highlighted by the grey box. (B) Three pairs of BR spectra are shown, with decreasing similarity from top to bottom. Their trace names are set in bold type in the (A). The blue line was drawn for intervals of 1 nm if the average difference of the two spectra was greater than 20 pN in that range.



existing clusters, e.g. the one formed by the two most similar spectra. This procedure was iterated until all spectra were grouped in one cluster. In the course of clustering, the previously calculated offsets were applied to shift spectra in the distance direction, yielding the optimal superimposition. When two clusters were merged, the offset of the two closest members of either cluster was used. Figure 2 shows the result of the automatic alignment of 19 force curves revealed while unfolding BR. The common unfolding pattern of the individual traces became visible and highlighted the good quality of the superimposition.

Generation of a hierarchical tree and selection of spectra

In the course of the alignment, a hierarchical tree was built that contains a coarse estimate of similarity between spectra (Fig. 3A). The distance between two spectra in the tree was taken to be the negative of the alignment score. Therefore, groups of similar spectra had a low joining distance, whereas deviant spectra had a higher distance. The estimate of similarity was sufficiently precise to sort out highly deviant spectra and significantly differing clusters (see below). As expected, applying the tree generation to an ensemble of typical spectra

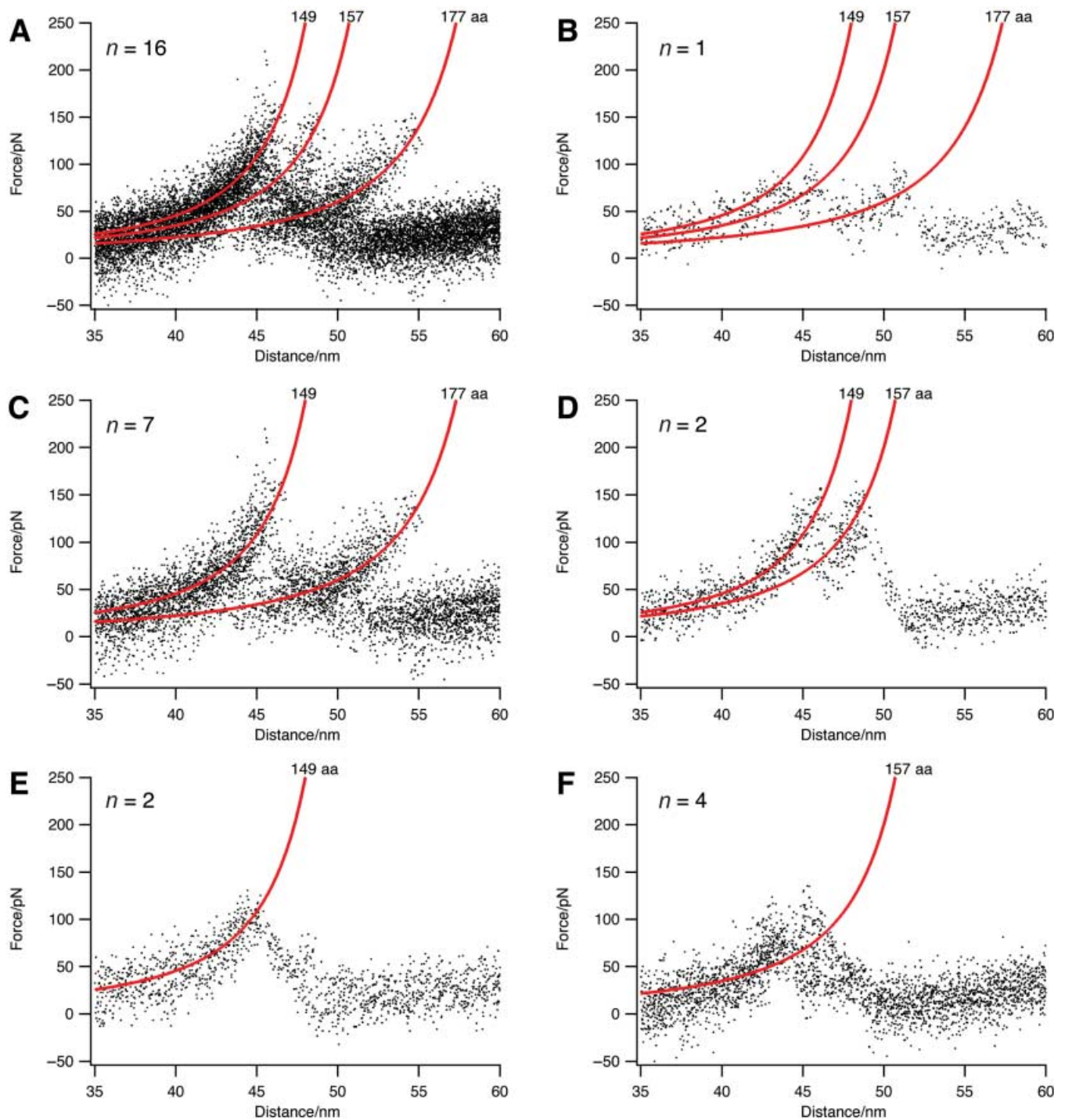


Fig. 4. Separation of different unfolding pathways based on the peak classification routine. Sixteen spectra recorded on BR have been aligned and grouped according to the occurrence of unfolding peaks. (A) contained all spectra, whereas the other frames only showed spectra with force curves indicating separate unfolding pathways. Force curves were fitted with the WLC model for each spectrum and automatically classified across spectra (red lines). Five unfolding pathways were identified (B–F) in agreement with a recent study on BR (Müller *et al.*, 2002). The counts of spectra were given in the top left corner of each frame.

recorded on BR and NhaA yielded a hierarchical tree with two different classes. The classes found represented either unfolding of BR or NhaA. Figure 3(B) shows three examples of trace pairings from the hierarchical tree. The two spectra of each pair are more similar to each other than to any other trace. In Fig. 3(B), the alignment score increases and the traces become more similar from top to bottom.

High deviations within data sets (for example, caused by contaminated force probes or thermal instabilities) introduced random peaks to the set of contour lengths. This lowered the quality of the following peak classification. Thus, it was advisable to exclude deviant spectra from the data set, or to split the data set into clusters. For this selection process, the hierarchical tree was used as it contained information about similarity between force curves. By default, the hierarchical tree was split at the average alignment score plus one SD of the score. This means that all nodes with a distance greater than this threshold were removed, while a number of clusters of more similar spectra remained. These clusters were used in the following peak classification routine, but not re-aligned. Re-alignment has no effect on the relative offsets of the spectra of such a cluster because the alignment of similar spectra is not influenced by more distant spectra. However, the average contact point of such a cluster will usually not coincide with the origin, whereas the average contact point of all aligned spectra does so. As it is evident in Fig. 3(B), unfolding pathways are not always grouped together in the hierarchical tree. This is due to the fact that spectra can follow different pathways in different parts of the protein. For example, the peaks at 30 nm distance showed different shoulder peaks although the major peaks of the BR spectra were similar. Therefore, it was desirable to distinguish different unfolding pathways based on a classification of the present peaks.

Peak classification in force curves

It has been shown previously that different unfolding pathways exist for membrane proteins such as BR (Müller *et al.*, 2002), NhaA (Kedrov *et al.*, 2004) and aquaporin-1 (Möller *et al.*, 2003). Different unfolding pathways showed characteristic unfolding events, which could be distinguished by the presence or absence of certain peaks. Therefore, it was highly desirable to group spectra that shared the same peaks. The identification of common peaks was achieved through a peak classification routine. Starting with the aligned spectra, peaks that had close contour lengths were combined as described in the Methods. The peak classification routine allowed distinguishing different unfolding classes by separating curves that shared common peaks (Fig. 4). In (A), the overlap of 16 spectra was shown. In contrast, the other frames showed spectra that only followed a given unfolding pathway (i.e. show a certain combination of peaks). This enabled the user to calculate probabilities for the occurrence of different unfolding and unbinding pathways.

Negative influences on alignment quality

The proposed algorithm worked well in clustering a large number of similar single molecule unfolding spectra. As mentioned above, deviant spectra showed low alignment scores and were sorted out. However, we found that our algorithm was susceptible to large numbers of deviant spectra. As the algorithm relied on pairwise alignments, a series of low-quality alignments seemed to lower the overall quality of the alignment. The subjective quality of the alignment was also lowered if F–D curves with a short non-contact part were recorded. The best fit to such very short baselines introduced a slight slope in the F–D curve leading to a shift in the alignment. This effect was avoided by recording F–D curves over twice the length of the entirely stretched polypeptide of the molecule.

Conclusion

The software and algorithms presented allowed the automated analysis of a large number of force spectra. For each spectrum, single peaks were detected and fitted using polymer extension models, yielding contour lengths and rupture forces. Force curves were aligned automatically by representing the spectra as a sequence of normal distributions and maximizing the overlap. The alignment procedure used a similarity measure that was also utilized to exclude deviant spectra. Combining the procedures of fitting and aligning the force spectra allowed the identification of common peaks being across all recorded spectra. This approach enables the observation and characterization of different interaction pathways of individual (bio)molecules among statistically relevant amounts of data.

Acknowledgements

We are grateful to Alexej Kedrov for providing his force spectroscopy data revealed on NhaA and Jens Struckmeier for stimulating discussions. The Free State of Saxony, the European community and the Deutsche Volkswagenstiftung supported this work.

References

- Albrecht, C., Blank, K., Lalic-Multhaler, M., *et al.* (2003) DNA: a programmable force sensor. *Science*, **301**, 367–370.
- Benoit, M., Gabriel, D., Gerisch, G. & Gaub, H. (2000) Discrete interactions in cell adhesion measured by single-molecule force spectroscopy. *Nat Cell Biol.* **2**, 313–317.
- Carl, P. & Dalhaimer, P. (2004) <http://www.mpikg-golm.mpg.de/gf/people/fery/techniques/PUNIAS/>.
- Carl, P., Kwok, C.H., Manderson, G., Speicher, D.W. & Discher, D.E. (2001) Forced unfolding modulated by disulfide bonds in the Ig domains of a cell adhesion molecule. *Proc. Natl Acad. Sci. USA*, **98**, 1565–1570.
- Dammer, U., Hegner, M., Anselmetti, D., Wagner, P., Dreier, M., Huber, W. & Güntherodt, H. (1996) Specific antigen/antibody interactions measured by force microscopy. *Biophys. J.* **70**, 2437–2441.

- Fritz, J., Katopodis, A.G., Kolbinger, F. & Anselmetti, D. (1998) Force-mediated kinetics of single P-selectin/ligand complexes observed by atomic force microscopy. *Proc. Natl Acad. Sci. USA*, **95**, 12283–12288.
- Gergely, C., Senger, B., Voegel, J.C., Horber, J.K.H., Schaaf, P. & Hemmerle, J. (2001) Semi-automatized processing of AFM force-spectroscopy data. *Ultramicroscopy*, **87**, 67–78.
- Janovjak, H., Kessler, M., Oesterhelt, D., Gaub, H. & Müller, D.J. (2003) Unfolding pathways of native bacteriorhodopsin depend on temperature. *EMBO J.* **22**, 5220–5229.
- Kasas, S., Riederer, B.M., Catsicas, S., Cappella, B. & Dietler, G. (2000) Fuzzy logic algorithm to extract specific interaction forces from atomic force microscopy data. *Rev. Sci. Instrum.* **71**, 2082–2086.
- Kedrov, A., Ziegler, C., Janovjak, H., Kühlbrandt, W. & Müller, D.J. (2004) Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy. *J. Mol. Biol.* **340**, 1143–1152.
- Leckband, D. & Israelachvili, J. (2001) Intermolecular forces in biology. *Quart. Rev. Biophys.* **34**, 105–267.
- Möller, C., Fotiadis, D., Suda, K., Engel, A., Kessler, M. & Müller, D.J. (2003) Determining molecular forces that stabilize human aquaporin-1. *J. Struct. Biol.* **142**, 369–378.
- Müller, D.J., Kessler, M., Oesterhelt, F., Möller, C., Oesterhelt, D. & Gaub, H. (2002) Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. *Biophys. J.* **83**, 3578–3588.
- Oberhauser, A.F., Marszalek, P.E., Erickson, H.P. & Fernandez, J.M. (1998) The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, **393**, 181–185.
- Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H.E. & Müller, D.J. (2000) Unfolding pathways of individual bacteriorhodopsins. *Science*, **288**, 143–146.
- Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J.M. & Gaub, H.E. (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, **276**, 1109–1112.
- Rief, M., Gautel, M., Schemmel, A. & Gaub, H.E. (1998) The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy. *Biophys. J.* **75**, 3008–3014.
- Saxton, W.O. (1996) Semper: Distortion compensation, selective averaging, 3-D reconstruction, and transfer function correction in a highly programmable system. *J. Struct. Biol.* **116**, 230–236.
- Williams, P.M., Fowler, S.B., Best, R.B., Luis Toca-Herrera, J., Scott, K.A., Steward, A. & Clarke, J. (2003) Hidden complexity in the mechanical properties of titin. *Nature*, **422**, 446–449.
- Yeung, K., Medvedovic, M. & Bumgarner, R. (2003) Clustering gene-expression data with repeated measurements. *Genome Biol.* **4**, R34.