NanoScope Software Tools for Force Spectroscopy Data Analysis

By: Ben Ohler, Ph.D., Application Scientist
Veeco Instruments Inc.

INTRODUCTION

The atomic force microscope (AFM) offers extraordinarily high resolution in force measurement applications, routinely yielding useful data down to the thermal noise floor of the cantilever, typically about 10pN. This along with the ease with which it is applied to many biological systems has made it a popular tool for studying such things as the specific interactions between biomolecules, the forces required to stretch polymeric molecules, and the forces that stabilize proteins. These sorts of applications have come to be collectively referred to as "force spectroscopy" applications.

Among these applications, those looking at the unfolding and refolding of single protein molecules are especially interesting. Significant inroads have been made in understanding how the primary structure of proteins contributes to their secondary and tertiary (folded) structures and how the proteins carry out their many functions. These measurements have complimented the many theoretical efforts to understand the details behind protein structure and function.

AFM force measurement applications vary substantially from AFM imaging applications in that the offline analysis of the resulting data is often far more involved and critical to the understanding of the experiment. So while a lot can be learned by simply looking at an image, force curve data demands quantitative analysis of the forces involved. The other major difference is that AFM force measurements are much more deeply grounded in statistical analysis than are AFM images. It is not uncommon for hundreds, or more often, thousands of force measurements to be required in a typical force spectroscopy experiment.

This Application Note will examine some of the data analysis challenges presented by force spectroscopy experiments and demonstrate how several powerful data analysis routines included in our NanoScope™ software can help you address them. For purposes of illustration, we'll use some titin pulling data acquired in our Veeco Santa Barbara force spectroscopy lab on a MultiMode™ PicoForce™ instrument. The PicoForce runs on our new, state-of-the-art NanoScope V™ controller

and represents our premier instrument for force spectroscopy applications. However, these features are also applicable to force measurements made with other Veeco AFMs like the MultiMode, the BioScope SZ™, and the new BioScope II™. The analysis examples shown here were done using NanoScope v.7r.2 software, but many of the same features are also available in v.6 and later versions, with some differences.

METHODS

A representative data set of titin unfolding data was collected as follows. A recombinant polyprotein consisting of eight tandem repeats of the I27 domain of human titin was obtained from Athena Enzyme Systems and diluted to 50µg/ml in PBS. A 50µL drop was incubated on a fresh template stripped gold surface for 10 minutes and then extensively rinsed with PBS and placed on the PicoForce scanner. Force-distance curves were collected with a ramp size of 500nm at a ramp rate of 1000nm/s using a 3nN relative trigger and a 1s surface delay. A MLCT-AUNM cantilever was used. Its spring constant was measured using the thermal tune method and found to be 35.2pN/nm. The thermal tune method is fully supported for the PicoForce and other microscopes running on v.7 NanoScope software.

SORTING THROUGH THOUSANDS OF FORCE-DISTANCE CURVES

A total of 1775 force-distance curves were collected at this ramp rate. Many of these curves contain no useful data, either because the tip failed to "catch" a molecule or because multiple interactions resulted in a complex force curve lacking the clean "fingerprint" that indicates a single protein extension. This is very typical of these experiments because the proteins are randomly distributed on the gold substrate and attachment to the tip and subsequent extension of the protein depends on non-specific physisorption to the tip. Moreover, the tip can attach to the protein anywhere along its length, so there are not always enough domains between the substrate and the tip to give a definitive "fingerprint." It is not uncommon then for only a small fraction of the total force-distance curves to ultimately be used for the final data analysis.

Clearly it would be a daunting task to manually open, examine, and then save or delete each force-distance curve file. While graduate student labor might be cheap, carpal tunnel surgery isn't, so there are two routines in the NanoScope software to streamline this chore. The first is called "Filter Curves" and can be found under the "File" menu. This brings up the view shown in Figure 1. Here, a folder containing force-distance curves can be selected. Each force curve file can have up to three independent plots, so you must also select which plot contains the data by which you wish to sort. A "Minimum Force" and a "Minimum Distance" are then defined that are used to automatically screen every force curve file in that directory. These parameters are applied to the retract curve of each file, searching whether the minimum specified adhesion force can be found at some point at least the minimum distance from the contact point. Note that the force is measured relative to the non-contact deflection value, not the absolute deflection value, so it does not matter whether the baselines of the curves are at zero. The "Minimum Force" parameter is intended to allow you to screen for some minimum expected force while the

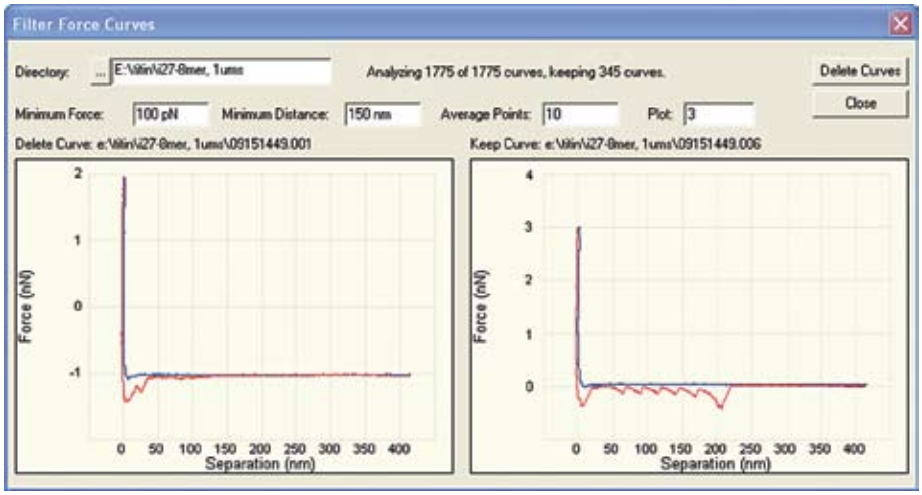

Figure 1: The Filter Curves view allows force-distance curves that contain no useful data to be rapidly identified and discarded according to user-defined criteria.

"Minimum distance" parameter allows you to avoid considering some non-specific adhesion near the contact point. As each curve is screened and sorted, they are displayed in the view with the "bad" curves on the left and the "good" curves on the right. In this way, you can quickly get a sense of whether the parameters you have chosen are appropriate. Once the routine has finished sorting through the entire folder, a dialog box is displayed requesting permission to delete the "bad" curves. These files are sent to the Windows recycle bin, but it's a good idea to back up the unsorted data in a separate folder anyway before executing this routine. In the example shown here, we were able to quickly reduce our data set of 1775 curves to just 345 curves–all without harming a graduate student.

SORTING THROUGH HUNDREDS OF FORCE-DISTANCE CURVES

Of course it would still be very unpleasant to manually open and sort through even 345 force curve files. But at some point it does become more efficient and effective to screen the force curves visually rather than depending on a software algorithm to appreciate the nuances of the curves. At this point the screened data from the Filter Curves routine will still contain curves that have multiple interactions instead of a clean pulling "fingerprint." In order to eliminate these, we include the "Review Curves" routine, which is also found under the "File" menu. This view, shown in Figure 2, allows you to rapidly page though the force curve files using the keyboard arrow keys and mark "bad" curves for deletion by hitting the keyboard "Delete" key. Like the "Filter Curves" view, you select the plot number by which you wish to sort. There are also other view options, as shown, to help make the curves easier to sort. After paging through each file, pressing the "Delete files" button will prompt to request permission to delete the files that you have tagged for deletion. Before that time, you are free to page back to previous files in case you mistakenly mark a file for deletion. Compared to individually opening every file, this greatly reduces the time required for this final sorting step. Instead of 20–30 seconds to open, view, close, and then save or delete each file, the sorting time is typically reduced to

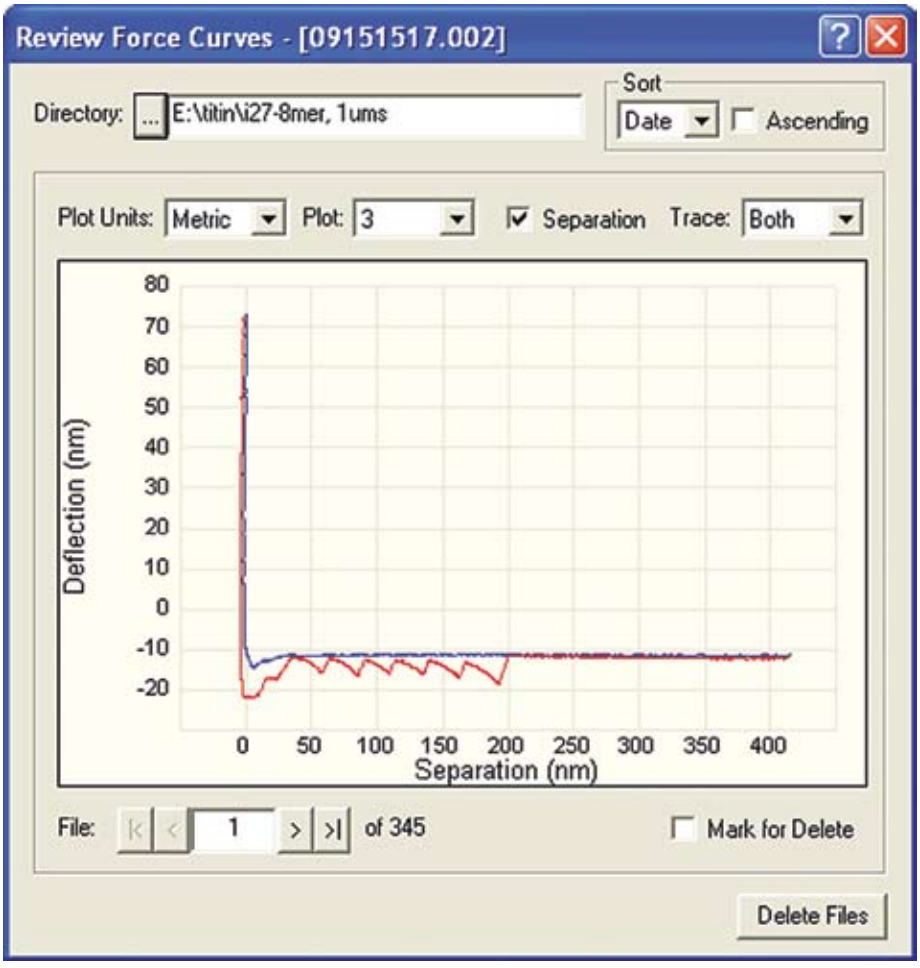

Figure 2: The Review Curves view allows the user to rapidly sort through force-distance curves without individually opening the files.

only 1–2 seconds per curve. While not fully automated, it certainly helps to quickly reduce the data set to one containing only useful data. In the case shown here, we started with 345 files after the initial screening and end up with only 25 files that contained good, clean titin pulling curves. Clearly this "good"-to-"bad" ratio will depend on the initial quality of the data as well as how stringently you sort out imperfect curves. Some groups choose to reduce the surface concentration of their proteins to maximize the chance that only single proteins will be stretched when an interaction does occur. Other groups prefer to increase the chance of interactions by increasing the surface concentration, though this also tends to increase the chance of multiple interactions. A suitable balance between the two cases must be made. We suggest that the first case– fewer but cleaner interactions– may be preferable since the Filter Curves routine is very efficient at screening out curves containing no interactions (or only a non-specific interaction), while curves containing multiple interactions must be removed using the more user-intensive Review Curves routine.

VISUALIZING MANY FORCE CURVES AT ONE TIME

Once you are finished sorting your data set, it is nice to be able to look at the remaining curves in a manner that gives a good overview of the data. The "Browse" view, found under the "View" menu, is convenient for this purpose. Its thumbnail view, shown below in Figure 3, does not display the data quantitatively or in high resolution, but it is useful for spotting any obviously bad curves that you might have missed or finding that one really nice curve at which you would like to take a closer look. Of course you can also load individual files or delete, move, or copy them. We will discuss next how you can analyze multiple curves from this view and then later how you can batch export force-distance data from this view.

EXTRACTING QUANTITATIVE DATA FROM MANY FORCE-DISTANCE CURVES

Once the force curves have been sorted so that only those containing clean, distinct "fingerprints" remain, the researcher is still left with the task of measuring and tabulating the values for each unfolding peak. In principle, this can be done

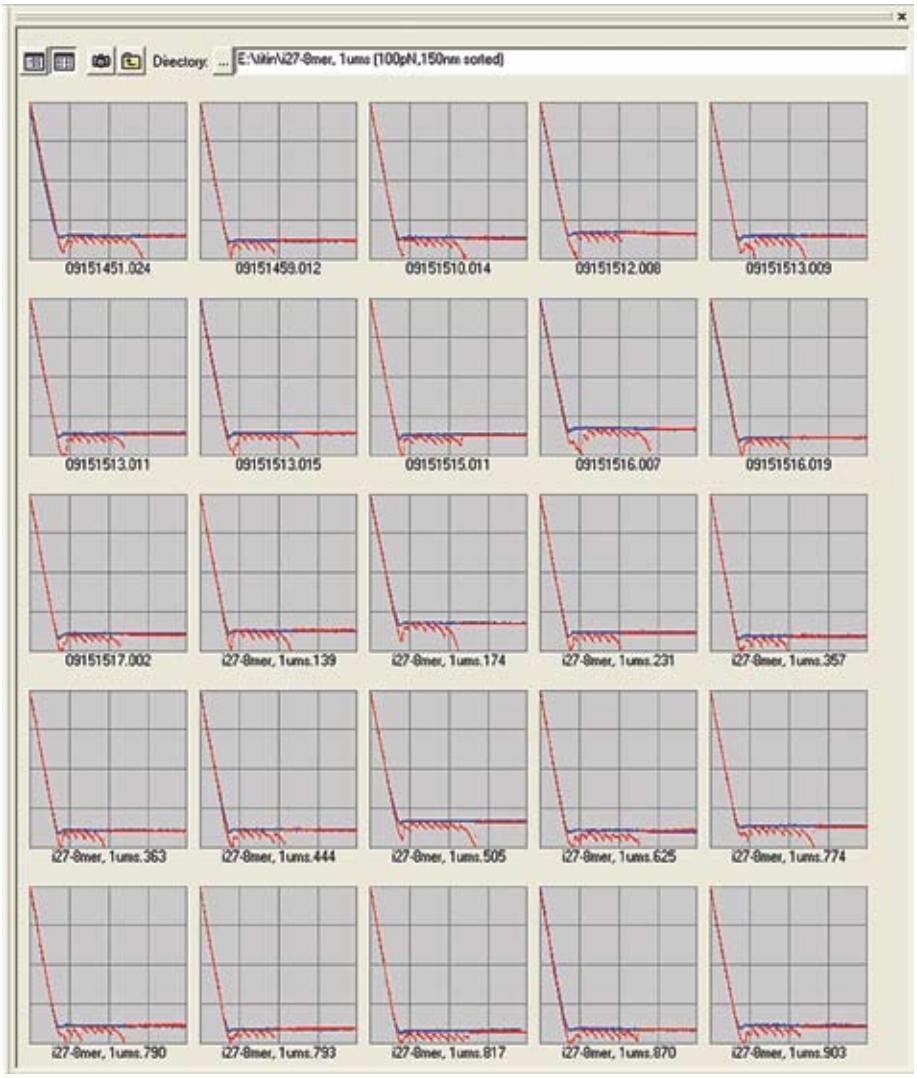

Figure 3: The Browse view lets you visualize many force-distance curves at once and carry out many normal file management tasks.

by individually loading each file and measuring the forces in the normal offline force curve view. But as with the sorting tasks, this is a tedious job. The "Multiple Curve Analysis" routine makes this analysis much simpler.

This view is hidden away in the Browse view that was described earlier. To access it, select a block of force curve files from the Browse view, then right-click and select "Curve Analysis." The curves will display, overlaying on one another. Because the non-contact deflection value is arbitrary, there will probably be considerable offsets between them. Therefore, one of the first things you will want to do is to turn on "Auto-align curves" for both the X- and Y- axes. Click "Refresh Plot" to update the view. The individual force curves should align with one another according to their contact points, as shown in Figure 4. Up to fifty curves can be displayed at one time.

If you set "Select Mode" to "On," the force curves will all become the same color. But if you select an individual file in the list below then that curve will turn a different color. If you wish to display just a few curves, you can make a graph like that shown in Figure 5 by setting "Y axis Auto-Zero" to "Cascade" and setting the desired "Cascade Offset." This is convenient for comparing a small subset of the curves. The most powerful feature in the Multiple Curve Analysis view, however, is the ability to automatically find, measure, and tabulate the size and position of the adhesion peaks. This is done by setting a "Minimum size" and "Minimum

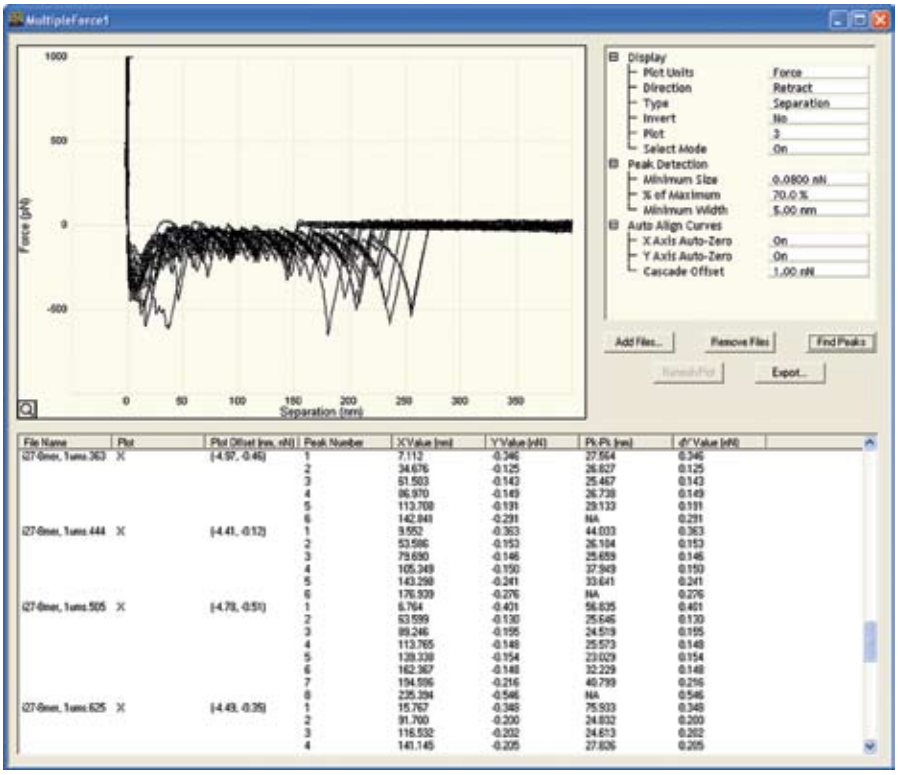

Figure 4: The Multiple Curve Analysis view offers functions that automatically align force curves by their contact points, finds adhesion peaks according to user-defined criteria, and can create a report of the tabulated peak forces and peak-to-peak distances.

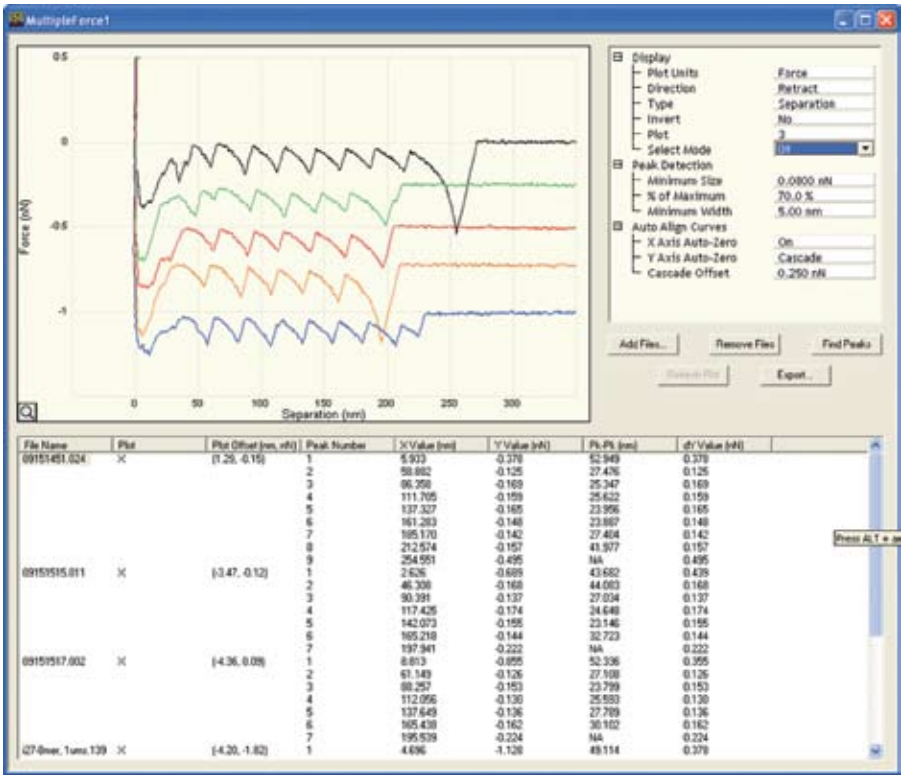

Figure 5: The Multiple Curve Analysis view can also display force curves in a "cascade" style view, which makes comparing several curves more convenient.

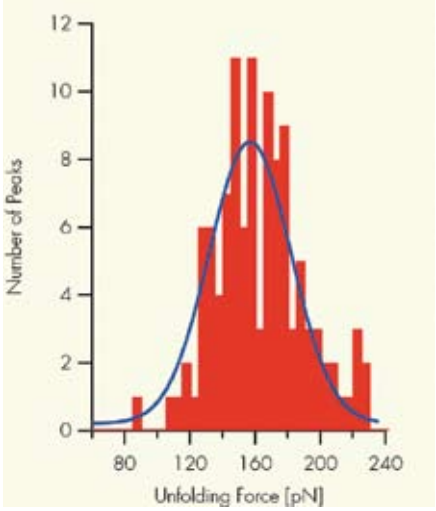

Figure 6: A Histogram of the peak force data is easily created from data file exported from the Multiple Curve Analysis view.

Width" for a peak to be counted. The minimum width is measured at a user-defined point defined by a percentage of the peak maximum. One these are set, click "Find Peaks" and the peak information fills the bottom file list, as shown in Figure 4. This list can then be exported to a delimited text file by clicking the "Export..." button. This file is then easily loaded into an analysis package like Igor Pro or even Excel to create a histogram of the unfolding forces likes that shown here in Figure 6.

The parameter values needed to most reliably find the peaks will vary. It is helpful to try different values with just a few curves before analyzing a large group so that you can more easily see which peaks are found. Note that the final large pull-off peak will also be tabulated. These values will appear in the histogram at higher forces and can simply be discarded.

ADVANCED ANALYSIS OF INDIVIDUAL FORCE CURVES

There are a number of analysis options available when you load a single force curve from the Browse view. For example, by dragging markers in onto the graphs you can read values from the curves for the X and Y positions, delta X and Y, and slopes. Be sure that the "Active Curve" parameter is set to the correct curve, Extend or Retract. Besides these simple features, it is also possible to fit a function to the curve between the two markers. You can choose between either a line fit or a worm-like chain fit. The fit parameters are output to the parameter list on the right side of the display. The fit function itself is overlaid on the actual data in the plot, as shown in Figure 7. The worm-like chain equation is:

$$f(x) = \frac{kT}{4b}\left[\frac{1}{\left(1-\frac{x}{L}\right)^2} + \frac{4x}{L} - 1\right]$$

where b is the persistence length (treated as a user-defined value, not a fit variable) and L is the contour length.

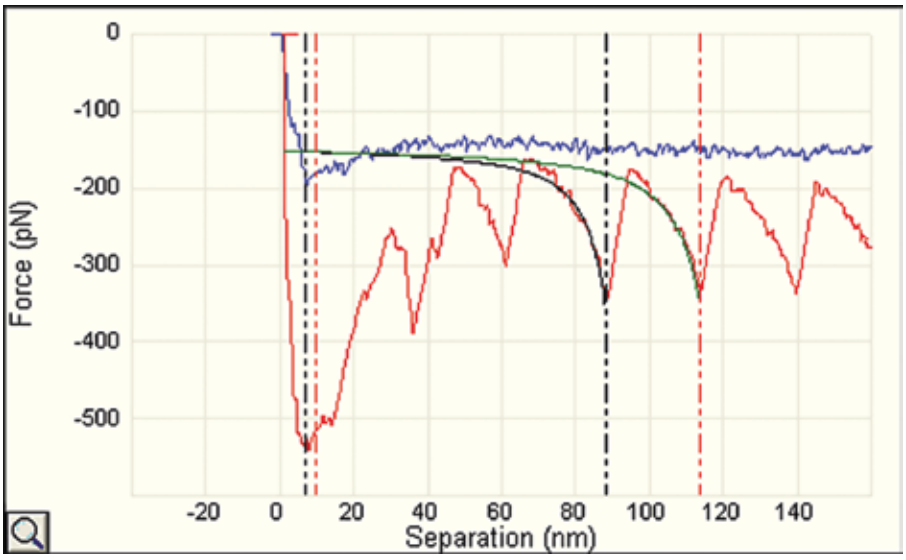

Figure 7: Worm-like chain fit of a titin extension curve

BATCH EXPORTING FORCE DISTANCE CURVE DATA

Finally, for analyses not directly supported in the NanoScope software, it is easy to batch export your force curve data. This can be done by selecting a group of files from the Browse view and then right-clicking and choosing Export...ASCII. This displays the dialog box shown in Figure 8, where one can select the exact channels and units to be exported. The full header can also be exported, although this is not normally required and is not recommended if the files are to be loaded as simple delimited text files.

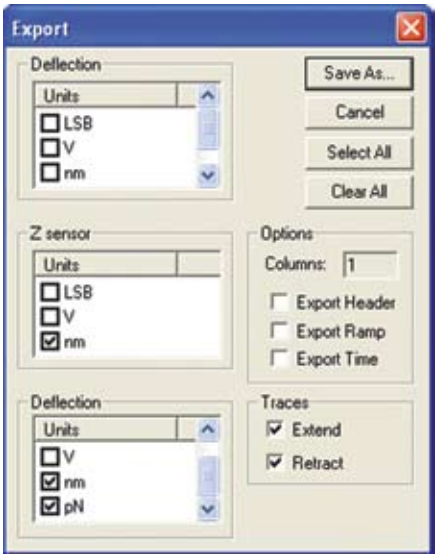

Figure 8: Batch export dialog box

SUMMARY

This Application Note has described a number of features available within the NanoScope software that make offline analysis of force spectroscopy data much easier and faster. These include features to help sort through large data sets of force-distance curves and remove those that contain no useful data. Analysis functions were described, including the Multiple Curve Analysis view, which can automatically find and tabulate peak values and positions in many force curves. We hope that you find these features helpful in making your research more productive.

WORLDWIDE CUSTOMER SUPPORT FROM THE INDUSTRY LEADER

Bruker Corporation is a leading provider of high-performance scientific instruments and solutions for molecular and materials research, as well as for industrial and applied analysis. For more information, visit www.bruker.com, email productinfo@bruker-nano.com, or call +1.805.967.1400/800.873.9750.