



Published in final edited form as:

Bergues Pupo, A. E., Göktas, M., Tunn, I., López García, P., Vila Verde, A., Blank, K., et al. (2018). Goodness of fit testing in dynamic single-molecule force spectroscopy. *The Journal of Chemical Physics*, 149: 244120. doi:10.1063/1.5055071.

Goodness of fit testing in dynamic single-molecule force spectroscopy

Ana E. Bergues-Pupo, Melis Goktas, Isabell Tunn, Patricia Lopez-Garcia, Ana Vila Verde, Kerstin G. Blank, and Angelo Valleriani

Goodness of fit testing in dynamic single-molecule force spectroscopy

Ana E. Bergues-Pupo,^{1, a)} Melis Goktas,² Isabell Tunn,² Patricia Lopez-Garcia,² Ana Vila Verde,¹ Kerstin G. Blank,² and Angelo Valleriani^{1, b)}

¹⁾Max Planck Institute of Colloids and Interfaces, Department of Theory & Bio-Systems, 14424 Potsdam, Germany

²⁾Max Planck Institute of Colloids and Interfaces, Mechano(bio)chemistry, 14424 Potsdam, Germany

(Dated: 28 November 2018)

Dynamic single-molecule force spectroscopy (SMFS) is a powerful method to characterize the mechanical stability of biomolecules. We address the problem that the standard manner of reporting the extracted energy landscape parameters does not reveal the intrinsic statistical errors associated with them. This problem becomes particularly relevant when SMFS is used to compare two or more different molecular systems. Here, we propose two methods that allow for a straightforward test of statistical significance. We illustrate the power of the methods by applying them to the experimental results obtained for three dimeric coiled coils of different length. Both methods are general and may be applied to any problem involving the fit of models with two correlated parameters.

I. INTRODUCTION

Dynamic single-molecule force spectroscopy (SMFS) is an unprecedented biophysical method to characterize the mechanical stability of biomolecular interactions¹⁻⁵. In a typical experimental setup, the molecules are pulled apart with the atomic force microscope at a constant retract speed. Force builds up between the molecule(s) until the rupture of the interaction is observed. Parameters such as the most probable rupture force F^* and the loading rate r are obtained performing hundreds of approach-retract cycles in a range of different retract speeds.

Assuming a one-dimensional energy landscape for the rupture process, the mechanical stability can be described in terms of the distance Δx to the transition state and the dissociation rate k_{off} in the absence of force. Under the assumption of a time-independent loading rate r , the parameters Δx and k_{off} can be inferred through a linear regression of the Bell-Evans equation⁶

$$F^* = \frac{k_B T}{\Delta x} \ln \left(\frac{r \Delta x}{k_{\text{off}} k_B T} \right), \quad (1)$$

that relates the most probable value F^* of the rupture force and the loading rate $r = dF/dt$ measured experimentally. Here, $k_B T$ is the thermal energy.

The values Δx and k_{off} physically characterize the molecular system. Even though these values are easily obtained from the fit, determining their statistical significance when comparing different systems is far more challenging. The following examples highlight the problems associated with the conventional way of reporting these values (see Table I). In the first two examples, the values were reported without errors. Are these values

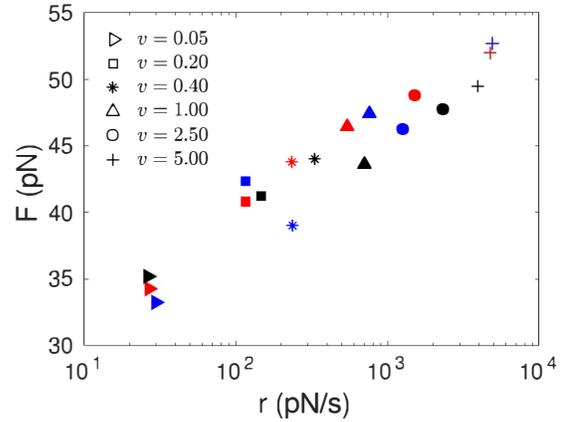


FIG. 1. Example data for the coiled coil A_4B_4 : The system was measured at six different retract speeds v ($\mu\text{m/s}$), conducting three independent experiments to obtain the most probable rupture forces and loading rates (three colors for each symbol).

statistically different? In examples 3 and 4, three sets of measurements, covering the same interval of retract speeds, were obtained experimentally. Each set was separately fitted with the Bell-Evans equation (see Eq. 1), resulting in three estimates of Δx and k_{off} for each system. Then the average values of the parameters and their standard error of the mean were determined. The same question as before holds: Are these values statistically different? For instance, a z -test in example 3 reveals that the values of Δx and k_{off} are *not* significantly different. In contrast, in example 4, the values of Δx are different at a significance level of $\alpha = 0.05$ but not at $\alpha = 0.01$, while the k_{off} values are not significantly different. Thus, are these systems physically different or not?

One aspect that the conventional way of reporting the errors does not consider properly is that the parameters Δx and k_{off} are actually correlated. Technically, the cor-

^{a)}Electronic mail: ana.bergues@mpikg.mpg.de

^{b)}Electronic mail: angelo.valleriani@mpikg.mpg.de

TABLE I. Comparison of the mechanical stability of different systems measured with SMFS. 1) protein GB1 engineered with a bi-histidine metal chelation site, characterized in the absence and presence of metal ions; 2) protein L carrying a mutation in a hydrophobic residue, which affects mechanical stability; 3) methylation of DNA at different sites can increase or decrease mechanical strength (parameters obtained from the fitting of three independent data sets with the Bell-Evans equation, Eq. 1); 4) coiled coil stability increases with the coiled coil length (parameters obtained from the fitting of three independent data sets with the Bell-Evans equation, Eq. 1).

Ex.	System	Modification	Δx (nm)	k_{off} (s^{-1})	Source
1	bi-His engineered GB1 (G4-51)	without Ni^{2+}	0.20	0.12	7
		with Ni^{2+}	0.17	0.023	
2	Protein L	wild type	0.26	0.02	8
		I60V Mutant	0.46	0.002	
3	DNA	non methylated	1.66 ± 0.35	$(8.4 \pm 35.2) \times 10^{-6}$	9
		methylated mc-3	1.44 ± 0.18	$(1.7 \pm 2.6) \times 10^{-4}$	
4	coiled coil	4 heptads	1.29 ± 0.12	$(3.2 \pm 2.1) \times 10^{-4}$	10
		3.5 heptads	0.89 ± 0.05	$(1.1 \pm 0.4) \times 10^{-1}$	
		3 heptads	1.03 ± 0.04	$(6.5 \pm 2.4) \times 10^{-3}$	

relation results from the fact that the covariance between these two parameters is not zero. Therefore, providing two independent errors for Δx and k_{off} is misleading because it gives certain combinations of these two parameters a much higher probability than it should. In addition, the statistics over only three measurements may be poor.

Here, we resolve these problems, introducing a new statistical approach tailored to the typical experimental design where at least two independent values of r and F^* are experimentally obtained at each of a given number of different retract speeds. We present this approach by applying it to example 4 (table I), for which three independent values of r and F^* were obtained at each of six different retract speeds. For the convenience of the reader, we briefly describe this experimental realization below. The use of our method to guide the formulation of statistically sound experimental designs is discussed at the end of the paper. To describe the method, we will assume that the Bell-Evans equation holds. Nevertheless, the method *per se* can also be applied to other bond rupture models that include additional physical assumptions about the rupture process, such as a change of the transition state distance with force¹¹, the effect of a non-linear loading rate dependence^{12,13} or the inclusion of rebinding events¹⁴.

II. PROPOSED STATISTICAL ANALYSIS METHOD

The systems under study in ref.¹⁰ consist of three heterodimeric coiled coils of different length. We use the following nomenclature for each coiled coil: A_4B_4 , $A_4B_{3.5}$ and A_4B_3 , indicating that helix A has a constant length of 4 heptad repeats whereas helix B consists of 4, 3.5 and 3 heptads (1 heptad = 7 amino acids). For each coiled coil, the most probable rupture forces F^* and corresponding loading rates r were determined at six differ-

ent retract speeds. To apply our proposed method to this small set of pre-defined retract speeds, two or more *independent* sets of measurements are required. Independent here means that the most probable rupture forces and loading rates have been obtained in different measurement sessions (*i.e.*, using a new cantilever and surface). We will call this method “approach 1”. In our specific example, there are three independent measurements at each speed. Thus, for each coiled coil there are 18 different values of the most probable rupture force and loading rate, organized into the six classes defined by the retract speed at which the measurements were performed (see Fig. 1).

The fit is done in MatLab with a linear model of F^* as a function of $\ln r$. One fit of six points gives one pair of Δx and k_{off} . We have then obtained a large number of Δx_i and $k_{\text{off},i}$ by applying the Bell-Evans fit to all possible combinations of six points out of the original 18 measurements. Here, i counts the number of possible sextuplets with $i = 1 \dots, 3^6 = 729$. The two-dimensional distributions of the Δx and k_{off} parameters obtained from each of these fits are presented in Fig. 2. This figure shows that Δx and k_{off} are correlated.

At this point, there are several ways to select a single, representative pair of values Δx and k_{off} from the entire set of values generated from the permutations. The simplest choice is the following: using all Δx_i and $k_{\text{off},i}$ values obtained in the previous step, we search for the pair of values $(\Delta x, k_{\text{off}})$ that minimizes the Residual Sum of Squares, $\text{RSS} = \sum_i ((\Delta x - \Delta x_i)^2 + (k_{\text{off}} - k_{\text{off},i})^2)$. The pair satisfying this condition will be taken as the best pair of values for the system. This step is used instead of simply calculating the averages over the values of Δx_i and $k_{\text{off},i}$ because these averages may be values that are not physically obtainable. Alternative methods, e.g. using a bivariate Gaussian fit, would also provide the 95% confidence area associated to the pair of values thus obtained. The best pairs of Δx and k_{off} values for the 3

TABLE II. Bell-Evans parameters calculated by minimizing the residual sum of squares (min(RSS); approach 1) or by fitting over the 18 points of each coiled coil system (fit(all); approach 2).

System	min(RSS)		fit(all)	
	Δx (nm)	k_{off} (s^{-1})	Δx (nm)	k_{off} (s^{-1})
A_4B_4	1.29	1.0×10^{-4}	1.28	1.4×10^{-4}
$A_4B_{3.5}$	0.88	1.1×10^{-1}	0.88	1.0×10^{-1}
A_4B_3	1.04	6.8×10^{-3}	1.03	5.8×10^{-3}

coiled coils in our example, obtained by minimizing the RSS, are shown in Table II and in Fig. 2.

Having computed the pairs of Δx and k_{off} that minimize the RSS of the three systems, the main question to be answered in the focus of this work is if these systems are significantly different. For this purpose, we test if the two-dimensional distributions of two systems, *i.e.*, $P_X(\Delta x, k_{\text{off}})$ of the Δx_i and $k_{\text{off},i}$ corresponding to system X , and $P_Y(\Delta x, k_{\text{off}})$ obtained for system Y are statistically different. To determine whether two sets of data arise from the same or different distributions, we use the Peacock test¹⁵, which is a two-dimensional extension of the Kolmogorov-Smirnov test. The method involves first computing the “maximum distance” D_{max} between the empirical cumulative distributions. Second, under the null-hypothesis that the two empirical distributions are the same, we create thousands of independent distributions P'_X and P'_Y by sampling randomly with replacement from the collection of $(\Delta x_i, k_{\text{off},i})$ of both systems X and Y and computing the maximum distance D' for each couple. This step produces a distribution of values of D' based on the assumed null-hypothesis. Finally, we compare the empirical D_{max} with the distribution of the D' to compute a p -value, *i.e.*, the probability to obtain a D' equal or larger than D_{max} . The smaller the p -value, the higher the evidence to reject the null hypothesis that the distributions are identical. Here, we fix a significance level $\alpha = 0.001$ for rejecting the null hypothesis. This test is non-parametric and it allows for correlations between the parameters. Using this procedure, we have transformed the question of whether two pairs $(\Delta x, k_{\text{off}})$ associated to two systems are significantly different into the question of whether the clouds in Fig. 2 are significantly different.

In our specific example, the p -values obtained from the Peacock test analysis over the $P(\Delta x, k_{\text{off}})$ distributions are less than 0.001 in all cases (see table III), allowing us to conclude that the sets of parameters describing each system are statistically different. The power of our approach is clear: both the Peacock test and the inspection of Fig. 2 leave no doubts that the systems are significantly different, whereas the original statistical treatment done in¹⁰ could not firmly establish this difference.

From an experimental point of view, the question arises if the proposed method can also be applied if only one

TABLE III. p -values of the Peacock test. Approach 1: between the distributions $P(\Delta x, k_{\text{off}})$ of the coiled coil systems. Approach 2: between the distributions $P(r, F^*)$ of the coiled coil systems.

Pairs	p -value	
	approach 1	approach 2
A_4B_4 - A_4B_3	$p < 0.001$	0.006
$A_4B_{3.5}$ - A_4B_3	$p < 0.001$	$p < 0.001$
A_4B_4 - $A_4B_{3.5}$	$p < 0.001$	$p < 0.001$

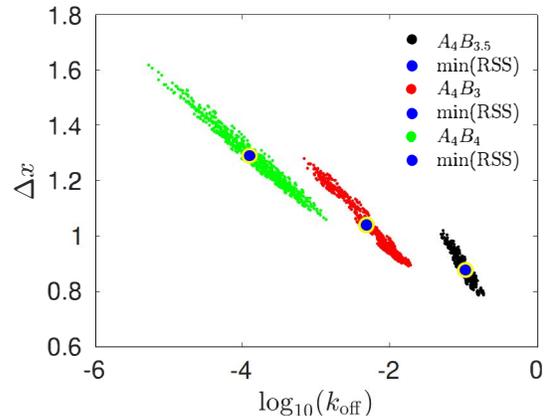


FIG. 2. Bell-Evans parameter clouds for each coiled coil system obtained from the fitting of the permuted data, together with the Δx and k_{off} pairs that minimize the Residual Sum of Squares (min(RSS); blue circles) in each system.

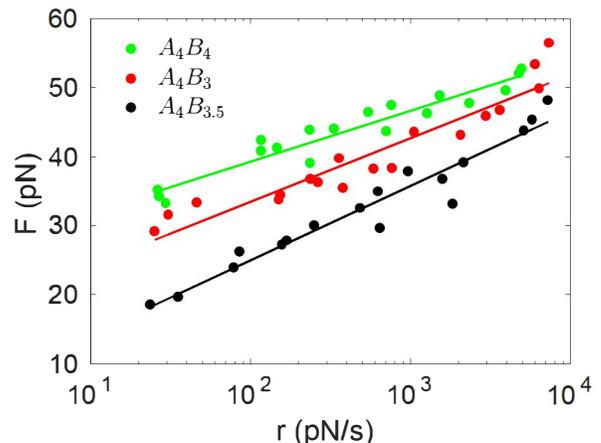


FIG. 3. Most probable rupture force *vs.* most probable loading rate for the three coiled coil systems, and corresponding Bell-Evans fit (lines) over the 18 points of each coiled coil system. The fit parameters are in Table II (fit(all)).

set of (r, F^*) pairs is available for each retract speed used. With only one (r, F^*) pair, the classification and the permutations described above are not possible. For

such cases, we propose an alternative approach (“approach 2”) where the Peacock test is used *directly* over the (r, F^*) pairs of each system, and not over the parameters of the fit. The two-dimensional distributions $P(r, F^*)$ will have considerably fewer points than the distributions $P(\Delta x, k_{\text{off}})$, which will decrease the accuracy of the test. Despite this shortcoming, this alternative method still enables the comparison of experimental results with far more statistical confidence than usual. The first step is to perform the Peacock test for the distributions $P_X(r, F^*)$ and $P_Y(r, F^*)$ of the systems X and Y . Second, if the distributions are statistically different according to the Peacock test, one can conclude that the parameters $\Delta x, k_{\text{off}}$ calculated from the fitting of the *entire* data set of each system with the equation (Eq. 1), are statistically different.

This approach is particularly suitable for systems with one measurement at each velocity. For illustration purposes, we apply it to the three coiled coil systems we consider here. Table III shows the p -values of each coiled coil system, obtained when performing the Peacock test over the 18 pairs (r, F^*) (right column). Because the p -values are very small we can conclude that the systems are statistically different, allowing us to fit each of them individually. The linear fits, using all 18 pairs for each coiled coil, are shown in Fig. 3 and the values of the parameters are presented in Table II (“approach 2”). The Δx and k_{off} values obtained using the two methods show good agreement (Table II), highlighting that both approaches can be used to compare different systems.

III. DISCUSSION

We present two approaches to test for statistical differences in the fit parameters extracted from dynamic single-molecule force spectroscopy. As mentioned in the introduction, these two approaches are general and can straightforwardly be applied to other models that relate the mean/maximum ruptures forces with the loading rates^{11–14}. For models with more than two parameters, the only necessary change is the generalization of the statistical test to higher dimensional distributions for approach 1.

The value of our two approaches goes beyond their power to determine statistical differences in the energy landscape parameters of pairs of systems after having the “final” results of SMFS experiments. Rather, they should be used iteratively throughout a study until a satisfactory experimental design is reached, i.e., to determine whether a given set of results is sufficient to answer the question “do systems X and Y differ in their free energy landscape” to the desired level of confidence, or whether further independent measurements are necessary. Equivalently, it is not possible to define a minimum number of data points necessary to apply our approaches; the appropriate number of data points depends on the systems being compared and the desired level of confidence

in the statistical test. After an initial set of (r, F^*) data are collected for each system, approach 2 (or approach 1, if at least two (r, F^*) pairs at each retract speed already exist) can be used to compare different experimental systems. We illustrate this point in Table IV, which shows the p -values obtained for approach 2 when fewer pairs are considered in the analysis of the systems in our example. With only 6 pairs of (r, F^*) , we can no longer discard the hypothesis that the systems A_4B_4 and A_4B_3 are different because the p -value is large. Even though statistically sound conclusions may be obtained in some cases with a single (r, F^*) at each retract speed (see the comparison between $A_4B_4 - A_4B_{3.5}$ and $A_4B_{3.5} - A_4B_3$ in Table IV), there are clear advantages to obtaining at least two independent (r, F^*) pairs. This experimental design not only gives more accurate estimates of the parameters associated with the free energy landscape of each system. Additionally, examination of the clouds of points obtained with approach 1 provides useful insight into the quality of experimental design: e.g., clouds with two highly populated areas separated by a low density region might suggest two types of physical events taking place, or a problem in the setup of one of the measurements.

The suitability of the method based on the Peacock test relies on some assumptions that should be tested for each particular system. For instance, in the permutation approach, it is assumed that there are no correlations between different (r, F^*) pairs. One possible source of correlation could arise from the use of the same cantilever to cover the whole range of retract speeds, as was the case for the coiled coil data analyzed here. Nevertheless, in this particular case, we have confirmed that no significant correlation exists even though each set of six retract speeds was measured with the same cantilever. This suggests that the random error associated with each F^* and r value is larger than the systematic error introduced by the cantilever calibration.

In this study we focused on the statistical significance of the parameters obtained from the Bell-Evans model. It is known that the parameters obtained from this model can be affected by wrong calibration of AFM cantilevers¹⁶ or the use polymeric handles^{12,13}, because they strongly affect the estimation of the loading rate values. The methods proposed here can be applied independently of the presence (and correction) of those bias since the corrections of biases and systematic errors have to be applied before the analysis tools proposed here. As mentioned in the introduction, the two analysis methods are general and could be in principle extended to others models that relate the mean/maximum ruptures forces with the loading rates^{11–14}. Because some of these models have more than two parameters, the statistical test should be generalized to higher dimensional distributions for the approach 1.

TABLE IV. p -values of the Peacock test for the approach 2 when some pairs (r, F^*) are removed from the original data set. For each comparison, 10 random removals of the given number of points was performed and for each of them a p -value was computed. The 10 p -values were combined in a single p -value following the Fisher's method¹⁷.

Pairs	p -value	
	6 pairs removed	12 pairs removed
$A_4B_4-A_4B_3$	$p < 0.001$	0.85
$A_4B_{3.5}-A_4B_3$	$p < 0.001$	$p < 0.001$
$A_4B_4-A_4B_{3.5}$	$p < 0.001$	$p < 0.001$

IV. CONCLUDING REMARKS

We have developed two methods to determine statistical differences in the energy landscape parameters obtained in dynamic single-molecule force spectroscopy. Applying these methods to the published data of the coiled coil systems in¹⁰, we conclude that the mechanical stability of these coiled coils is indeed different. We show that the statistical analysis is in principle possible for experiments where only a pair (r, F^*) at each retract speed is available. It is clearly of advantage, however, to obtain more than one independent pair over a wide range of retract speeds; in fact, such an experimental design should be the standard for conducting this type of experiments. This allows to perform the permutation approach to generate a reasonable amount of data without increasing the number of experiments. This, in turn, provides an excellent starting point for the statistical test.

V. SUPPLEMENTARY MATERIAL

A Matlab code to perform the analysis is provided. The code requires as input the most probable rupture forces and loading rates and outputs the clouds of the

parameters Δx and k_{off} with their min(RSS) and p -values for the approach 1 and Δx and k_{off} from the fitting over all points and the p -values for the approach 2.

VI. ACKNOWLEDGMENTS

This work has been funded by the Max Planck Society. MG, IT, PL also thank the IMPRS on Multiscale Bio-Systems for financial support.

- ¹E. Florin, V. Moy, and H. Gaub, *Science* **264**, 415 (1994).
- ²E. Evans and K. Ritchie, *Biophysical Journal* **72**, 1541 (1997).
- ³B. Heymann and H. Grubmüller, *Physical Review Letters* **84**, 6126 (2000).
- ⁴F. Ritort, *Journal of Physics: Condensed Matter* **18**, R531 (2006).
- ⁵A. Noy, *Current opinion in Chemical Biology* **15**, 710 (2011).
- ⁶E. Evans, *Annual Review of Biophysics and Biomolecular Structure* **30**, 105 (2001).
- ⁷Y. Cao, T. Yoo, and H. Li, *Proceedings of the National Academy of Sciences* **105**, 11152 (2008).
- ⁸D. P. Sadler, E. Petrik, Y. Taniguchi, J. R. Pullen, M. Kawakami, S. E. Radford, and D. J. Brockwell, *Journal of Molecular Biology* **393**, 237 (2009).
- ⁹P. M. Severin, X. Zou, H. E. Gaub, and K. Schulten, *Nucleic Acids Research* **39**, 8740 (2011).
- ¹⁰M. Goktas, C. Luo, R. M. A. Sullan, A. E. Bergues-Pupo, R. Lipowsky, A. V. Verde, and K. Blank, *Chemical Science* **9**, 4610 (2018).
- ¹¹O. K. Dudko, G. Hummer, and A. Szabo, *Phys. Rev. Lett.* **96**, 108101 (2006).
- ¹²C. Friedsam, A. K. Wehle, F. Kühner, and H. E. Gaub, *Journal of Physics: Condensed Matter* **15**, S1709 (2003).
- ¹³C. Ray, J. R. Brown, and B. B. Akhremitchev, *The Journal of Physical Chemistry B* **111**, 1963 (2007).
- ¹⁴R. W. Friddle, A. Noy, and J. J. De Yoreo, *Proceedings of the National Academy of Sciences* **109**, 13573 (2012).
- ¹⁵J. A. Peacock, *Monthly Notices of the Royal Astronomical Society* **202**, 615 (1983).
- ¹⁶J. Te Riet, A. J. Katan, C. Rankl, S. W. Stahl, A. M. van Buul, I. Y. Phang, A. Gomez-Casado, P. Schön, J. W. Gerritsen, A. Cambi, *et al.*, *Ultramicroscopy* **111**, 1659 (2011).
- ¹⁷R. A. Fisher, *Statistical methods for research workers* (Genesis Publishing Pvt Ltd, 2006).