Analytical Biochemistry 395 (2009) 1–7

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Analysis of residuals from enzyme kinetic and protein folding experiments in the presence of correlated experimental noise

Petr Kuzmič [a,*], Thorsten Lorenz [b,1], Jochen Reinstein [b]

[a] BioKin Ltd., 15 Main Street Suite 232, Watertown, MA 02472, USA
[b] Department of Biomolecular Mechanisms, Max-Planck Institute for Medical Research, Heidelberg, Germany

ARTICLE INFO

Article history:
Received 30 January 2009
Available online 12 June 2009

Keywords:
Enzyme kinetics
Protein folding
Mathematics
Statistics
Regression
Data filter
Michaelis–Menten
HIV protease
Autocorrelation function
Runs-of-signs test

ABSTRACT

Experimental data from continuous enzyme assays or protein folding experiments often contain hundreds, or even thousands, of densely spaced data points. When the sampling interval is extremely short, the experimental data points might not be statistically independent. The resulting neighborhood correlation invalidates important theoretical assumptions of nonlinear regression analysis. As a consequence, certain goodness-of-fit criteria, such as the runs-of-signs test and the autocorrelation function, might indicate a systematic lack of fit even if the experiment does agree very well with the underlying theoretical model. A solution to this problem is to analyze only a subset of the residuals of fit, such that any excessive neighborhood correlation is eliminated. Substrate kinetics of the HIV protease and the unfolding kinetics of UMP/CMP kinase, a globular protein from *Dictyostelium discoideum*, serve as two illustrative examples. A suitable data-reduction algorithm has been incorporated into software DYNAFIT [P. Kuzmič, Anal. Biochem. 237 (1996) 260–273], freely available to all academic researchers from http://www.biokin.com.

Introduction

Currently available laboratory instruments allow the digital recording of many experimental data points during each individual enzyme assay, or from a variety of biophysical experiments. For example, if a conventional kinetic experiment is allowed to progress for 15 min, while recording the absorbance or fluorescence signal every second, the complete data trace will contain 900 data points. The abundance of experimental data points is even greater when using rapid-kinetics techniques, where the sampling interval might be in the millisecond range. It is not unusual to encounter data traces containing tens of thousands of individual data points.

The particular statistical technique typically used for the interpretation of enzyme kinetic data is the nonlinear least-squares regression. Several excellent reviews of the least-squares method have been published specifically for biochemical audiences [1–3]. One crucially important assumption of the least-squares method is that each individual data point is *statistically independent* of all the other data points in the complete data set.

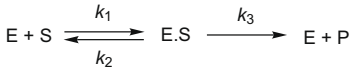
This article is concerned with those situations where the sampling interval might be too short for the experimental data points to be truly statistically independent. The loss of statistical independence may occur due to correlated fluctuations in the instrument's electronic circuitry, or short-term fluctuations in the concentrations of reactants, or a number of other possible causes. Whatever the underlying cause, it is often found in practice that densely spaced recordings from continuous enzyme assays or protein folding experiments include correlated "spikes" or "bumps" in the recorded signal (fluorescence, absorbance, and the like). The presence of such spikes inevitably distorts the results of residual analysis. As a consequence, standard statistical tests to assess the goodness-of-fit will fail, even if the experiment essentially does agree with the postulated theoretical mechanism.

This article attempts to alleviate the problem in diagnosing the goodness-of-fit due to the statistical dependence among individual data points. Our approach relies on purposely analyzing only a suitably selected subset of the residuals, such that any neighborhood correlation is removed. This is accomplished by evaluating two standard criteria of goodness-of-fit, namely, the runs-of-signs test and the autocorrelation function. Importantly, the two statistical criteria are computed not only for the original (complete) data set, but also separately for every nth residual ($n = 2$, 3, 4, and so on).

If the underlying kinetic mechanism does match the given experimental data, these two statistical criteria rapidly improve on selecting out every second (third, fourth, etc.) residual of fit. In contrast, if the underlying theoretical model deviates

* Corresponding author. Fax: +1 617 209 1616.
E-mail address: pksci01@biokin.com (P. Kuzmič).
URL: http://www.biokin.com (P. Kuzmič).
[1] Present address: Novartis AG, Basel, Switzerland.

0003-2697/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2009.05.051



Scheme 1.

systematically from the available data, then even analyzing progressively more widely spaced subsets of residuals does not improve the goodness-of-fit. In this way, we can eliminate any possible distortions in the results of standard statistical tests introduced by statistical dependence between closely spaced data points. The newly proposed method is illustrated on two unrelated examples: (1) a continuous fluorogenic assay of the HIV protease [4]; and (2) the unfolding kinetics of the globular protein UMP/CMP from *Dictyostelium discoideum* (UmpK), a nucleoside monophosphate (NMP) kinase [5].

Theory

Mathematical model for substrate kinetics

The model equation for each progress curve analyzed in this report is

$$f(t) = f_0 + r_P c_P(t + \Delta t), \tag{1}$$

where $f(t)$ is the fluorescence intensity at time t; f_0 is the offset on the signal axis (a property of the instrument); r_P is the molar response coefficient of the reaction product P; $c_P(t + \Delta t)$ is the concentration of P at time $t + \Delta t$; and $\Delta\, t$ is the mixing delay time. The mixing delay, in this case $\Delta t = 5$ s, is the time elapsed between the start of the enzyme reaction (by mixing the substrate and enzyme stock solution) and the start of actually recording fluorescence intensities by the instrument ($t = 0$ in the recorded data trace).

The product concentration at any arbitrary time, c_P, is computed from the initial ($t = -\Delta t$) concentrations of the enzyme, $c_E^{(0)}$, and substrate, $c_S^{(0)}$, by solving an initial-value problem defined by a system of simultaneous first-order ordinary differential equations (ODE)[2] derived from Scheme 1.

$$dc_E/dt = -k_1 c_E c_S + (k_2 + k_3) c_{ES} \tag{2}$$
$$dc_S/dt = -k_1 c_E c_S + k_2 c_{ES} \tag{3}$$
$$dc_{ES}/dt = +k_1 c_E c_S - (k_2 + k_3) c_{ES} \tag{4}$$
$$dc_P/dt = +k_3 c_{ES}. \tag{5}$$

The ODE system is integrated numerically [6].

The mathematical model defined by Eqs. (1)–(5) contains four adjustable parameters, namely, the instrument offset f_0; the molar response coefficient r_P; and the rate constants k_2 and k_3. The bimolecular association rate constant was held as a fixed parameter at $k_1 = 10^8\ M^{-1}\ s^{-1}$, which corresponds to the approximate diffusion limit in this particular class of molecules [7, p. 164]. All model equations were automatically derived by the software system DYNAFIT [6] from symbolic input listed in the Appendix.

Autocorrelation function of residuals

The autocorrelation function has been defined in slightly different ways by various investigators [8, pp. 20, 49–50; 9, p. 37]. In this work, we use the definition of Box and Jenkins [10, pp. 28–32]. For lag or step h, the autocorrelation function R_h is defined as shown in the equations

$$R_h = C_h/C_0 \tag{6}$$
$$C_h = \frac{1}{n_D}\sum_{i=1}^{n_D-h}(r_i - \bar{r})(r_{i+h} - \bar{r}) \tag{7}$$
$$C_0 = \frac{1}{n_D}\sum_{i=1}^{n_D}(r_i - \bar{r})^2, \tag{8}$$

where C_h is the autocovariance function; C_0 is the variance function; n_D is the number of data points; r_i is the residual for the ith data point; and $\bar{r}$ is the average residual.

The P values for the first autocorrelation coefficient ($h = 1$), shown in Figure 3, were computed by evaluating the probability that the standardized normal residual defined by

$$z_{1-\alpha/2} = R_1\sqrt{n_D} \tag{9}$$

could occur by random chance, given the null hypothesis that $z = 0$. A z value significantly different from zero means that the residuals are nonrandom, which indicates a lack of fit between the experimental data and the assumed fitting model.

Residual runs-of-signs test

The runs-of-signs test for the randomness of residuals has been used in the literature in various modifications [9, pp. 36–37; 11–13, pp. 259–260]. Let n_D be the number of data points (which equals the number of residuals) in a given progress curve; let n_+ be the positive residuals; and n_R the number of runs, or groupings, of equal-sign residuals within the time series. For example, if the pattern of signs in a series of residuals is +++--- +++---, we have 12 residuals ($n_D = 12$), six positive residuals ($n_+ = 6$) and four runs ($n_R = 4$).

For any given number of residuals and positive residuals—assuming that those residuals are distributed truly randomly—the average expected number of runs and the associated variance are defined by, respectively [9]

$$\mu = 2n_+(n_D - n_+)/n_D + 1 \tag{10}$$
$$\sigma^2 = (\mu - 1)(\mu - 2)/(n_D - 1). \tag{11}$$

The standardized normal deviate is then defined by [13]

$$z = (n_R - \mu + 1/2)/\sigma, \tag{12}$$

where the term 1/2 is a correction for continuity. The P values in Fig. 3 were computed as probabilities that the standardized normal deviates z could occur by random chance, given the null hypothesis that $z = 0$. As above, a z value significantly different from zero means that the residuals are suspect, and that the assumed theoretical model does not fit the experimental data.

Experimental

Cloning, expression and purification

UmpK was cloned into a modified pET27 expression vector via NcoI and BamHI restriction sites. The plasmid was transformed into *Escherichia coli* BL21 DE3 cells. Protein expression was induced with 0.1 mM IPTG at $OD_{600} = 0.6$, and the expression was allowed to proceed for approximately 18 h at 20 °C. The purification of UmpK from harvested cells was performed as described elsewhere [5]. The correct mass of 21.9 kDa was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

Kinetic analysis of UmpK unfolding

UmpK unfolding kinetics were measured with a BioLogic SFM400 stopped-flow apparatus connected to a MOS 450 optical

[2] *Abbreviation used:* ODE, ordinary differential equations.

system (BioLogic, Claix, France). Unfolding of native UmpK (20 μM) was induced by rapid 10-fold dilution into different urea concentrations (>2 M). The final UmpK concentration of each measurement was 2 μM in 50 mM Tris–HCl, pH 7.5, 100 mM Na_2SO_4, and 2 mM DTE at 25 °C. The kinetics were followed by intrinsic tryptophan fluorescence at $\lambda \geqslant 320$ nm (long-pass filter from LOT-Oriel GmbH, Darmstadt, Germany) on excitation at $\lambda = 296$ nm. The experimental data were recorded by the Bio-Kine data acquisition software (BioLogic). In the statistical analysis by DYNAFIT [6], the first 14 ms of the recorded data trace (average of three traces) was treated as the dead time of the instrument.

Results

HIV protease: Complete set of residuals

The experimental data from a typical fluorogenic assay of the HIV protease [4,14] are shown in Fig. 1. The enzyme reaction was followed for 5 min, with sampling interval 0.5 s, resulting in 601 data points. Only the first 4 min (481 data points) of the assay was statistically analyzed. The raw experimental data are shown as the thin jagged curve; the thick solid curve in Fig. 1 represents the best least-squares fit to the theoretical model, represented by the system of Eqs. (1)–(5). The best-fit values of the model parameters and the corresponding formal standard errors were $f_0 = 5.92 \pm 0.02$ RFU (relative fluorescence units); $r_P = 15.9 \pm 0.03$ RFU/μM; $k_2 = 193 \pm 5\ s^{-1}$; and $k_3 = 8.7 \pm 0.2\ s^{-1}$. The regression analysis was performed using the software package DYNAFIT [6].

The inset to Fig. 1 shows the residuals of fit. The expanded detail plot shows that the residuals apparently are not statistically independent. Within each cluster of residuals, the fluorescence signal seems to be rising and falling in a correlated fashion (an increase in fluorescence seems more likely to occur if the fluorescence intensity also increased at the immediately preceding data point, and vice versa). The mutually correlated clusters seem to occur in groups of approximately three to five data points.

The open circles in Fig. 2 display the autocorrelation function for all residuals plotted in Fig. 1. The lag (h) on the horizontal axis

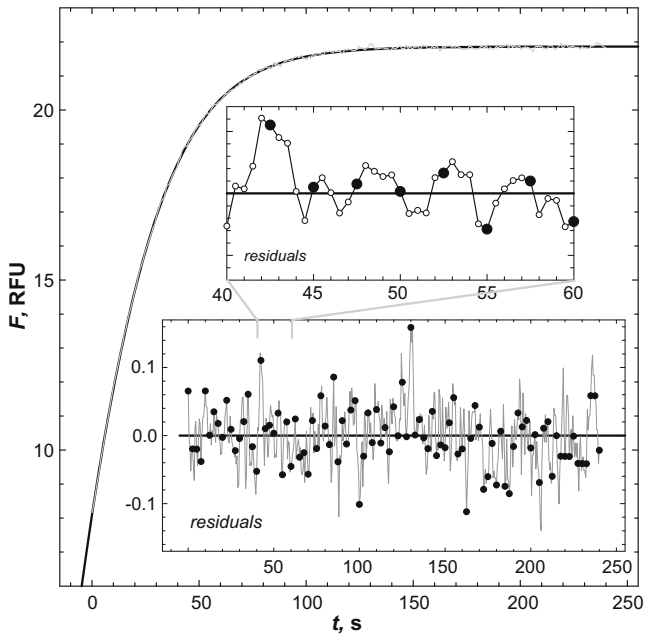


Fig. 1. Experimental data from the fluorogenic assay of the HIV protease. Experimental conditions: [E] = 10 nM; [S] = 1.0 μM. For complete experimental details, see Ref. [4]. Inset: Residual plot. The solid dots represent every fifth residual of the total 481 residuals.

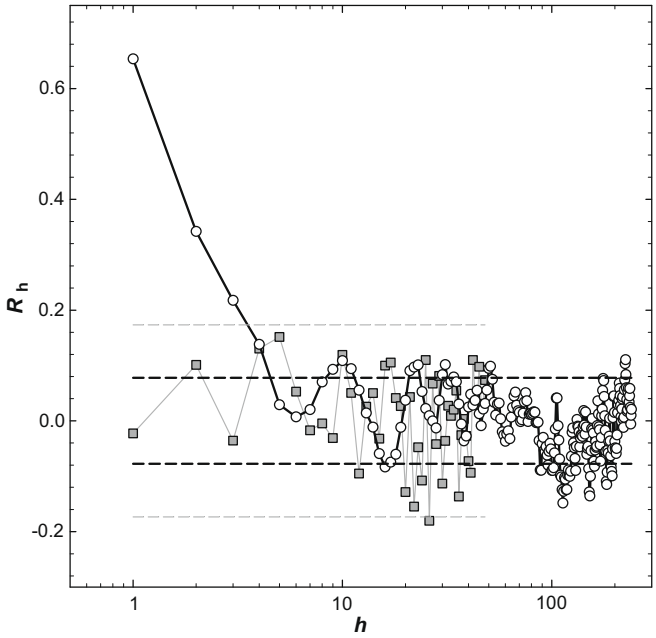


Fig. 2. Autocorrelation function for residuals of fit shown in *Fig. 1*. Open circles: All 481 residuals were analyzed. Gray squares: A subset of every fifth residual was analyzed. The dashed lines represent the critical values of the neighborhood autocorrelation coefficient, R_1, at the 95% confidence level.

signifies the distance between data points. For example $h = 1$, or unit spacing, corresponds to the neighborhood correlation coefficient R_1 on the vertical axis (i.e., the autocorrelation between two immediately neighboring residuals). Similarly, $h = 4$ and R_4, or four-point spacing, represent the correlation between pairs or residuals separated by three other data points (residual numbers 1, 5, 9, 13, etc.).

The first autocorrelation coefficient (R_1 according to Eq. 6) is significantly larger than the 95% critical value represented by the thick horizontal dashed lines (the thinner, gray horizontal lines will be explained in the next section). Even the second, third, and fourth autocorrelation coefficients (R_2–R_4, shown as the second through fourth points from the left plotted in Fig. 2) are significantly large. Thus, the residuals are strongly nonrandom. This normally indicates a lack of fit between the experimental data and the presumed fitting model.

Similar results were obtained for P values derived from the runs-of-signs test (Eqs. (10)–(12)). In particular, the inset to Fig. 1 shows 481 residuals, 237 of which are positive, grouped in 138 runs. Based on the total number of residuals and the number of positive residuals, the statistical theory predicts that the average expected number of runs, and the corresponding standard deviation, is 241 ± 11. The actual number of runs ($n_R = 138$) is very much smaller than the average expected number ($\mu = 241 \pm 11$); the corresponding P value is practically indistinguishable from zero. Again, the residuals appear strongly nonrandom, which usually indicates a lack of fit between the data and the model.

The statistical analysis up to this point revealed an obvious contradiction. On the one hand, the experimental data and the presumed theoretical model agree very well: on visual examination of the main panel in Fig. 1, it is difficult even to distinguish the raw data trace overlaid on the best-fit model curve. Even more importantly, a visual examination of the residual plot also seems satisfactory, because we do not see any obvious systematic pattern in the residuals. Contradicting these observations, two independent statistical tests for the randomness of residuals (namely, the runs-of-signs test and the serial correlation function) seem to indicate a serious lack of fit.

The following section describes a possible remedy to these contradictions.

HIV protease: Analysis of various subsets of the residuals

Both the autocorrelation function and the runs-of-signs test indicate that our experimental data occur in statistically correlated clusters. This violates one of the basic assumption of nonlinear least-squares regression, namely, the assumption of statistical independence between the experimental data points [1–3]. One possible way to avoid this neighborhood correlation problem is to analyze a suitably selected subset of the original residuals of fit.

An important insight for the development of this procedure is provided by the serial correlation function, represented by the open circles in Fig. 2. The first point plotted in Fig. 2, corresponding to the neighborhood correlation coefficient R_1 (lag $h = 1$), represents an autocorrelation value significantly higher than the critical value. However, the fifth correlation coefficient R_5 (lag $h = 5$) is essentially equal to zero. In other words, there is virtually no correlation within a subset of residuals composed of every fifth original residual. This particular subset of residuals is shown by solid filled circles in the residual plot (inset to Fig. 1).

The autocorrelation function for the 1/5 subset of the original residuals is shown as gray squares in Fig. 2. Clearly, within this particular subset, the residuals show no autocorrelation, which indicates statistical independence between individual data points.

Similar results are found on examining the runs-of-signs test for the 1/5 subset of the residuals. The solid filled circles in Fig. 1 represent $n_D = 97$ residuals, of which $n_+ = 45$ have positive value. Within the subset, there are 52 runs of equal sign residuals. Statistical theory predicts that if the residuals were purely random and normally distributed, with $n_D = 97$ and $n_+ = 45$ we should expect on average 49.2 ± 4.9 runs. The observed number of runs (52) is even higher than the average expected number (49), which means that the 1/5 subset of residuals does appear completely random.

The goodness-of-fit criteria (autocorrelation function and runs-of-signs test) were computed for all various equally spaced $1/N$ subsets of the residuals, that is, every other residual ($N = 2$), every third residual ($N = 3$), and so on. For each subset, we computed the corresponding P values. The results are summarized in Fig. 3, in which the solid circles represent the P values for the neighborhood correlation coefficient R_1, and the open circles represent the P values for runs-of-signs test. Both statistical criteria produce extremely nonrandom values ($P \approx 0$) when all residuals are analyzed ($N = 1$). As the spacing interval increases, the P values also increase, until essentially perfect randomness is achieved at $N = 5$ (every fifth residual analyzed).

Unfolding kinetics of UMP/CMP kinase

The results of fit from a typical protein unfolding kinetic experiment are shown in Fig. 4. Native UmpK was diluted into 3.6 M urea, and intrinsic tryptophan fluorescence intensity was monitored for 5 s with 1 ms sampling interval (5000 data points). We have established that the unfolding of UmpK follows the two-step mechanism $A \rightarrow B \rightarrow C$ [5]. The data were fit to the corresponding theoretical model using the DYNAFIT [6] notation `A --> B: k1; B --> C: k2`. The DYNAFIT software has automatically generated a suitable system of simultaneous differential equations.

As in the fluorogenic assays of the HIV protease (see Fig. 1 above), the fluorescence intensities in the UmpK unfolding assay display statistically significant serial correlation spanning approximately five or six consecutive data points. The inset to Fig. 4 displays in greater detail these quasi-periodic oscillations in the experimental signal, and makes clear that the experimental data points are not statistically independent but instead occur in correlated clusters. Correspondingly, the autocorrelation function computed from all residuals of fit in Fig. 5 (top curve labeled "$N = 1$") shows very high values for lag $h = 1$ (neighborhood correlation coefficient, R_1) through $h = 8$. If the residuals of fit were truly uncorrelated, the plot of the entire correlation function (for all values of lag h) would fit inside the bounds corresponding to the critical values (horizontal straight lines in Fig. 5) at the given significance level (here, 95%).

When the residuals of fit in Fig. 4 were sampled with progressively wider spacing, the serial correlation function quickly

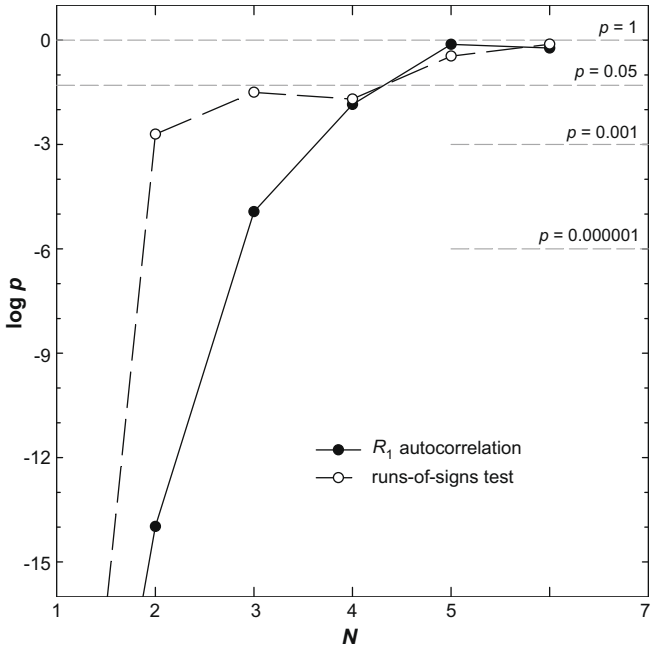


Fig. 3. Probabilities (P values) for randomness of residuals as measured by the R_1 neighborhood correlation coefficient (solid circles) and the runs-of-signs test (open circles) for variously spaced subsets of residuals shown in Fig. 1.

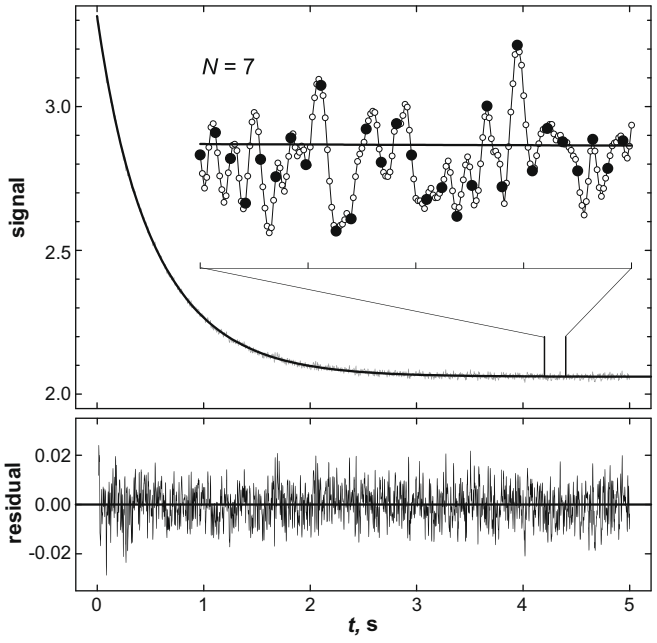


Fig. 4. Protein unfolding kinetics for UmpK at 3.6 M urea: least-squares fit to a two-step unfolding mechanism $A \rightarrow B \rightarrow C$. For details, see Experimental.

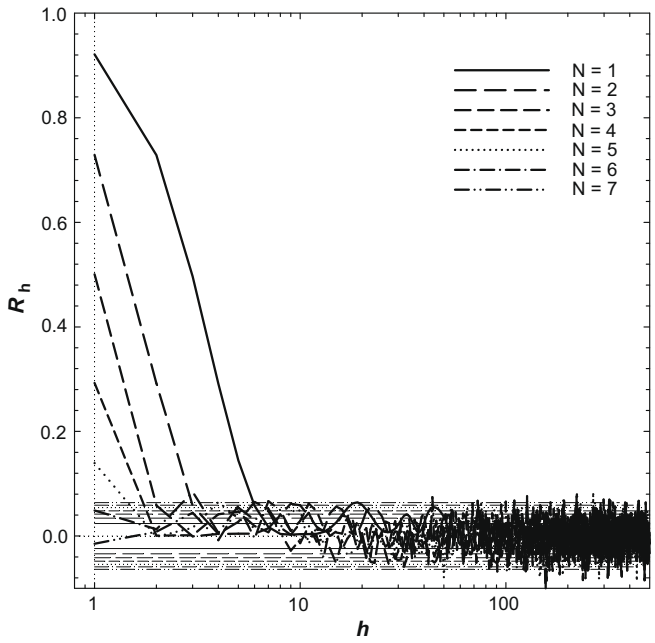


Fig. 5. Autocorrelation function computed from the residuals displayed in Fig. 4 (two-step unfolding mechanism $A \rightarrow B \rightarrow C$) with progressively wider spacing (N) of residuals.

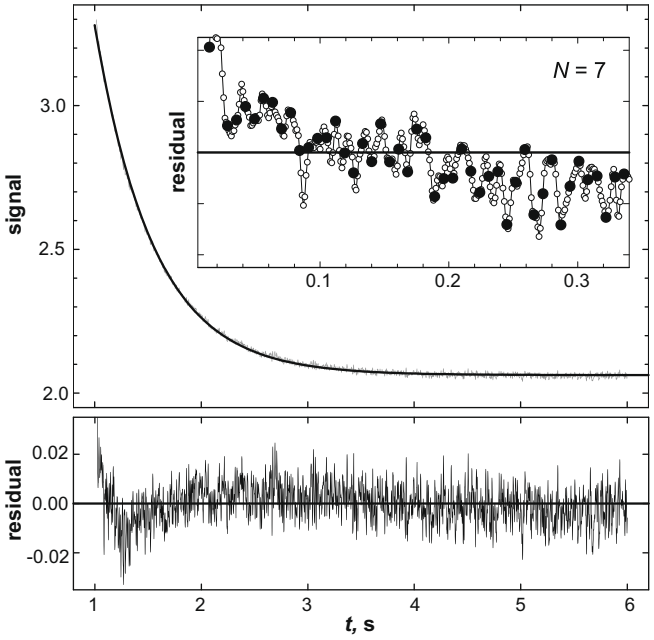


Fig. 6. Protein unfolding kinetics for UmpK at 3.6 M urea – least-squares fit to a single-step unfolding mechanism $A \rightarrow C$. For details, see Experimental.

flattened out and, finally, at $N = 7$ (every seventh residual analyzed) the neighborhood correlation coefficient R_1 (plotted at $h = 1$) became lower in absolute magnitude than the corresponding critical value. The same was true for the remaining values of the autocorrelation function ($h = 2,3, \ldots, n_D/2$, where n_D is the total number of data points).

The results of fit from the same unfolding experiment to an overly simple kinetic model, $A \rightarrow B$, are shown in Fig. 6. The residuals of fit (lower panel in Fig. 6) show an easily detectable nonrandom pattern. The autocorrelation function computed from *all* residuals of fit in Fig. 7 (top curve labeled "$N = 1$") again shows a strongly nonrandom pattern, well outside the bounds depicted by the critical values (horizontal straight lines in Fig. 7). Importantly, the goodness-of-fit does *not* improve as we start sampling the residuals of fit in order to compute the autocorrelation function from every second residual ($N = 2$), every third ($N = 3$), every fourth ($N = 4$), and so on. No matter how much we increase the spacing of the residuals, the first few values of the autocorrelation function ($h = 1,2,\ldots$) always remain outside the critical values delineated by the horizontal lines in Fig. 7.

The runs-of-signs tests for the protein unfolding kinetics showed the same pattern as was previously seen for the HIV protease substrate assays. Briefly, the runs-of-signs test according to Eqs. (10)–(12) continued to show extremely low P values (too few runs for the given number of data points) in the case of the one-step mechanism $A \rightarrow B$, irrespective of the sampling interval.

In the case of the two-step mechanism $A \rightarrow B \rightarrow C$, the runs-of-signs test did indicate a serious "lack of fit" when *all* residuals in Fig. 4 were tested. Specifically, with the total number of data points $n_D = 4987$ and $n_+ = 2527$ positive residuals, the statistical theory predicts that, if the residuals were truly random, we would observe on average 2494 ± 35 runs of equal signs. The residual plot in the bottom panel of Fig. 4 contains only 632 runs of signs, and thus the corresponding P value is essentially indistinguishable from zero. However, when every seventh residual was analyzed ($N = 7$), the total number of residuals was $n_D = 713$, the number of positive residuals was $n_+ = 360$, and the number of runs was $n_R = 372$. The statistical theory predicts that if the residuals were truly random, for $n_D = 713$ and $n_+ = 360$ there should be, on average, 357 ± 14 runs of equal signs. The observed number of runs ($n_R = 372$) was even higher than that prediction, and so it is random.

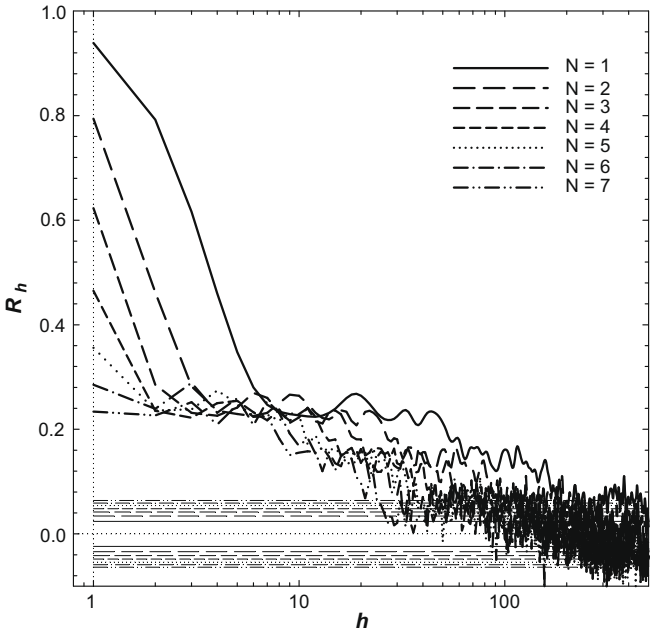


Fig. 7. Autocorrelation function computed from the residuals displayed in Fig. 6 (single-step unfolding mechanism $A \rightarrow B$) with progressively wider spacing (N) of residuals.

Discussion

Currently available instrumentation techniques can easily generate thousands of data points in a single kinetic experiment, and a number of advanced software systems are available to analyze this

digital torrent. Following an early example set by the KINSIM package [15,16], some software systems such as DYNAFIT [6] allow the investigator to specify the mathematical model by using a symbolic notation. Indeed, entering "E + S = E.S = E.P = E + P" on the keyboard is much easier for most experimentalists than deriving the corresponding system of simultaneous differential equations. This convergence of advanced digital instrumentation, increasingly powerful computer hardware, and numerical data processing algorithms has created what some researchers [17,18] fittingly describe as the "New Enzymology".

One of the challenges of this "New Enzymology" is that certain established data analysis techniques may no longer work, at least not without appropriate modifications. This article addresses one such situation, arising, specifically, in the analysis of residuals. Residual analysis is one of the classic biochemical data analysis techniques [9,11,12]. It is used to determine whether the postulated fitting model, or mechanism, does adequately describe the experimental data. If the residuals are sufficiently randomly distributed, we can conclude that the model fits well. If the residuals appear nonrandomly distributed, we must conclude that the postulated theoretical model is suspect.

Fig. 1 presents a striking example of a contradiction created, in a sense, by having "too many" closely spaced experimental data points. We do know from many independent experiments that the HIV protease follows the simple Michaelis-Menten mechanism, shown in Scheme 1. More importantly, in the specific case of the experiment shown in Fig. 1, the sum of squared deviations between this model and the experimental data is so small that it is practically indistinguishable from the random noise generated by the instrument employed in this assay [4]. And yet, two different statistical tests for the randomness of residuals had failed to confirm the validity of this independently established mechanistic model. Both the runs-of-signs test [12] and the serial correlation test [9] produced *P* values that nominally indicate a serious lack of fit.

This paradox is explained by the fact that the digitized fluorescence intensities, recorded in this case with a half-second sampling interval, are not statistically independent as is required by the statistical theory of least-squares regression. One possible solution to this problem would be to collect more widely spaced (i.e., essentially, fewer) experimental data points. In this particular example, the statistical correlation between neighboring fluorescence intensities becomes vanishingly small with data spacing $\Delta t = 2.5$ s (every fifth data point actually analyzed). With this wider data spacing, it becomes feasible again to use residual analysis to discriminate between well-fitting and poorly fitting mathematical models.

Purposely reducing the number of experimental data points may lead to undesirable loss of information, especially in the case of highly "ill-conditioned" theoretical models. A prototypical example of ill-conditioning is the multi-exponential model [19] encountered in protein folding kinetics. Thus, according to the data analysis method presented in this report, we retain the full data set for the purpose of determining the best-fit model parameters. However, in assessing the suitability of the fitting model—as measured by the degree of randomness in the residuals—we analyze various progressively smaller subsets of the experimental data points. If the theoretical model fits well, any neighborhood correlation (essentially an artifact of the experiment due to the sampling interval being too small) will quickly vanish. On the other hand, if a model exhibits a bona fide discrepancy with the experimental data, the randomness of residuals will not improve on filtering.

The results from the unfolding kinetics of UmpK illustrate that when the presumed mathematical model and the experimental data are in genuine disagreement, the filtering technique described above will not remove the nonrandom characteristics of the residual of fit. In contrast, when the model and the data fundamentally agree, the proposed residual filtering scheme will quickly improve the statistical goodness-of-fit (as measured by the runs-of-signs test and by the autocorrelation function).

One possible limitation inherent in the proposed method of assessing the randomness of residuals is that it requires a relatively large number of originally recorded data points. The runs-of-signs test as defined by (Eqs. (10)–(12)) is statistically valid only for $n_D > 20$ [9,13], although special modifications of the test have been designed for a smaller number of data points [12]. Therefore, it is required that the kinetic traces contain at least several hundred experimental data points.

As a matter of course, a data analyst should never place too much significance on any one statistical test, or perhaps even on two separate tests, such as the autocorrelation test and the runs-of-signs test employed in this report. We have not discussed other known statistical tests for normality, such as the Durbin–Watson test for serial correlation [20,21], or other related statistical tests [22,23].

All statistical tests described in this report have been incorporated into an updated version of the DYNAFIT software package [6], which continues to be freely available to all academic researchers from http://www.biokin.com.

Acknowledgment

P.K. is indebted to Domenico Gatti (Wayne State University, Detroit, MI, USA) for bringing to his attention the problem discussed in this article and for many useful discussions. We thank Sarah McCord for making helpful comments on the article.

Appendix A

The following DYNAFIT [6] script file will fit the experimental data shown in Fig. 1 so the system of Eq. (1)-(5). The entire dataset is used in the regression, but only every fifth residual is used to compute the goodness-of-fit tests (`AnalyzeEveryNthPoint` = 5).

```
[task]
  task = fit | data = progress
[mechanism]
  E + S <===> E.S  :  k1   k2
  E.S ---> E + P  :  k3
[constants] ; units = μM, s
  k1 = 100, k2 = 100 ??, k3 = 10 ??
[concentrations] ; units = μM
[responses]
  P = 2.57 ?
[data] ; units = s
  directory ./hiv-protease/data/05-06-1994
  extension txt | delay 5
  file ex15 | offset auto ?
  conc E = 0.010, S = 1
[output]
  directory ./hiv-protease/output/fit-ex15
[settings]
  Residuals | AnalyzeEveryNthPoint = 5
  Filter | TimeMax = 240
  Output | Autocorrelations = 1
[end]
```

References

[1] M.L. Johnson, Why, when, and how biochemists should use least squares, Anal. Biochem. 206 (1992) 215–225.
[2] M.L. Johnson, S.G. Frasier, Nonlinear least-squares analysis, Methods Enzymol. 117 (1985) 301–342.

[3] M.L. Johnson, Use of least-squares techniques in biochemistry, Methods Enzymol. 240 (1994) 1–22.
[4] A.G. Peranteau, P. Kuzmič, Y. Angell, C. García-Echeverría, D.H. Rich, Increase in fluorescence upon the hydrolysis of tyrosine peptides: application to proteinase assays, Anal. Biochem. 227 (1995) 242–245.
[5] T. Lorenz, J. Reinstein, The influence of proline isomerization and off-pathway intermediates on the folding mechanism of eukaryotic UMP/CMP kinase, J. Mol. Biol. 381 (2008) 443–455.
[6] P. Kuzmič, Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase, Anal. Biochem. 237 (1996) 260–273.
[7] A. Fersht, Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding, third ed., Macmillan, New York, 1998.
[8] C. Chatfield, The Analysis of Time Series: An Introduction, fourth ed., Chapman & Hall, New York, 1989.
[9] J.G. Reich, Curve Fitting and Modelling for Scientists and Engineers, McGraw-Hill, New York, 1992.
[10] G.E.P. Box, G. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1976.
[11] B. Mannervik, Design and analysis of kinetic experiments for discrimination between rival models, in: L. Endrényi (Ed.), Kinetic Data Analysis: Design and Analysis of Enzyme and Pharmacokinetic Experiments, Plenum, New York, 1981, pp. 235–270.
[12] B. Mannervik, Regression analysis, experimental error, Regression analysis, experimental error, and statistical criteria in the design and analysis of experiments for discrimination between rival kinetic models, Methods Enzymol. 87 (1982) 370–390.
[13] J.O. Rawlings, Applied Regression Analysis: A Research Tool, Wadsworth, Belmont, CA, 1988.
[14] P. Kuzmič, A.G. Peranteau, C. García-Echeverría, D.H. Rich, Mechanical effects on the kinetics of the HIV proteinase deactivations, Biochem. Biophys. Res. Commun. 221 (1996) 313–317.
[15] B.A. Barshop, R.W. Wrenn, C. Frieden, Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM—a flexible, portable system, Anal. Biochem. 130 (1983) 134–145.
[16] C.T. Zimmerle, C. Frieden, Analysis of progress curves by simulations generated by numerical integration, Biochem J. 258 (1989) 381–387.
[17] K.A. Johnson, Z.B. Simpson, T. Blom, Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data, Anal. Biochem. 387 (2009) 20–29.
[18] K.A. Johnson, Z.B. Simpson, T. Blom, Fitspace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data, Anal. Biochem. 387 (2009) 30–41.
[19] A. Cornish-Bowden, Fundamentals of Enzyme Kinetics, Butterworth, London, 1979.
[20] J. Durbin, G.S. Watson, Testing for serial correlation in least squares regression. I, Biometrika 37 (1950) 409–428.
[21] J. Durbin, G.S. Watson, Testing for serial correlation in least squares regression. II, Biometrika 38 (1951) 159–179.
[22] R. D'Agostino, M. Stephens, Goodness-of-Fit Techniques, Marcel Dekker, New York, 1986.
[23] J.C.W. Rayner, D.J. Best, Smooth Tests of Goodness of Fit, Oxford Univ. Press, Oxford, 1989.