Please cite this article in press as: Lindemann et al., A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.08.042

Biophysical *Journal*
Computational Tool

Biophysical Society

A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra

William R. Lindemann,[1] Ty Christoff-Tempesta,[1] and Julia H. Ortony[1,*]
[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT Electron paramagnetic resonance spectroscopy (EPR) is a uniquely powerful technique for characterizing conformational dynamics at specific sites within a broad range of molecular species in water. Computational tools for fitting EPR spectra have enabled dynamics parameters to be determined quantitatively. These tools have dramatically broadened the capabilities of EPR dynamics analysis, however, their implementation can easily lead to overfitting or problems with self-consistency. As a result, dynamics parameters and associated properties become difficult to reliably determine, particularly in the slow-motion regime. Here, we present an EPR analysis strategy and the corresponding computational tool for batch-fitting EPR spectra and cluster analysis of the χ^2 landscape in Linux. We call this tool CSCA (Chi-Squared Cluster Analysis). The CSCA tool allows us to determine self-consistent rotational diffusion rates and enables calculations of activation energies of diffusion from Arrhenius plots. We demonstrate CSCA using a model system designed for EPR analysis: a self-assembled nanoribbon with radical electron spin labels positioned at known distances off the surface. We anticipate that the CSCA tool will increase the reproducibility of EPR fitting for the characterization of dynamics in biomolecules and soft matter.

SIGNIFICANCE Electron paramagnetic resonance spectroscopy is a useful technique for measuring conformational dynamics at specific sites of biomolecules and molecular materials. This tool can provide quantitative measures of dynamics; however, spectral analysis often leads to problems with reproducibility. We developed a computational tool (CSCA, "Chi-Squared Cluster Analysis") for Linux that is designed to address this problem. CSCA allows users to visualize the χ^2 fitting process and employs statistical clustering methods to help assign best fits. We demonstrate that these capabilities reduce overfitting and provide self-consistent, quantitative measures of dynamics.

INTRODUCTION

The dynamic behavior of biomolecules and soft materials is often critical to their function (1–5). In protein biology, for instance, a dynamic picture of proteins has gradually replaced the purely structural "lock-and-key" model describing their function. More recent models, such as the induced fit and the conformational selection models, describe proteins as conformationally fluctuating molecules and relate the equilibrium conformational distribution and the rate of conformational change to the activity and interaction of these macromolecules with each other (6,7). In self-assembling materials with biological applications, dynamics measurements are used to understand thermodynamic phase behavior, and even to identify liquid or solid phases within different regions of the same self-assembled sample (2,8). Dynamic behavior is complicated and challenging to study, especially in heterogeneous systems such as membrane proteins or in systems that precipitate or aggregate at moderate concentrations.

Continuous wave electron paramagnetic resonance (CW-EPR) is a powerful tool for measuring dynamic behavior of radical electron spin labels at specific sites of molecular species in water (9–15). By synthetically introducing nitroxide (also called nitroxyl) radicals into the sample at known positions, the rotational diffusion rate (D_R) of these radicals can be determined by fitting their EPR spectra. Most commonly, EPR is used to characterize the dynamics of proteins (12,16,17), peptides (18), polymers (19,20), and small molecules in water (2,8,21). In these systems, spin labels are covalently tethered to molecules. Rotational diffusion rates of the spin labels are inversely related to rotational correlation times (τ_R), which are typically on the order of picoseconds to tens of microseconds.

Submitted May 15, 2020, and accepted for publication August 31, 2020.
*Correspondence: ortony@mit.edu
Editor: Sudha Chakrapani.
https://doi.org/10.1016/j.bpj.2020.08.042

Please cite this article in press as: Lindemann et al., A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.08.042

The stochastic Liouville equation (SLE) quantum mechanically describes the spectra associated with slowly moving electrons and is important for analyzing conformational dynamics on timescales relevant for EPR dynamics measurements (10,14,15). Several software packages that use the SLE model are available for fitting EPR spectra, most notably the NLSL software package (11), EasySpin (13), and MultiComponent, a LabVIEW wrapper for NLSL (22). Since the introduction of these software packages, the SLE model has been enhanced by incorporating the macroscopic order microscopic disorder (MOMD) model and the slowly relaxing local structure (SRLS) model (23,24). The validity of the SLE model in the slow-motion regime has been verified by molecular dynamics simulations, which can accurately reproduce EPR spectra from atomistic trajectory files and confirm that the diffusion of spin labels gives an accurate approximation of the dynamics of the local environment (25–29).

Despite the availability of fitting programs, challenges remain in the analysis of dynamics by CW-EPR. Expert analysis is typically necessary to reliably estimate fit uncertainty, as well to obtain self-consistent results in variable temperature experiments (4). Such analysis reduces the risk of overfitting and improves the reproducibility and consistency of fits. In nonlinear curve fitting, there are many cases for which the absolute best fit is less consistent with physical reality than other possible fits—even in systems with very few parameters (30,31).

In the cases in which self-consistent variable temperature experiments have been demonstrated, they have revealed Arrhenius-like dependences on temperature (4,5,32,33). Although this observation lacks a precise formal justification, it is consistent with the widely accepted belief that the motion of spin labels is connected to the diffusion of the label's parent molecule, as well as to the local viscosity of the environment. Both of these properties typically experience Arrhenius-like temperature dependence at ambient temperatures.

Here, we report an open-source MATLAB (The MathWorks, Natick, MA)-based software referred to as CSCA (Chi-Squared Cluster Analysis) that interfaces with NLSL for Linux to improve the reliability of EPR spectral fitting. The CSCA toolkit enables users to 1) automatically resample data points from spectra to incorporate the effects of noise into error analysis; 2) map and analyze the χ^2 error function, allowing visualization of clusters of local minima; 3) provide more meaningful fit values and error bars based on statistical analysis of high-quality fits; and 4) incorporate a variety of popular bounded global optimization tools, including simulated annealing, genetic algorithms, particle swarms, and Monte Carlo methods. Furthermore, CSCA allows fitting using any choice of parameters available in NLSL.

We used CSCA to analyze the dynamic behavior of spin labels attached via oligoproline spacers to aramid amphiphile (AA) nanoribbons after spontaneous self-assembly in water. These amphiphilic molecules were designed to form nanoribbons of 5 nm width, 4 nm thickness, and up to 20 μm in length. AA nanoribbons exhibit negligible molecular exchange over 2 months, in addition to high Young's moduli and tensile strengths (upublished data). Fig. 1 *a* illustrates the chemical structure of the AA molecule (compound **1**) and the spin labeled analogs (compounds **2**–**5**) that are incorporated by co-assembly at low concentrations (2 mol%). The spin labeled nanoribbon is schematized in Fig. 1 *b*, and a transmission electron microscopy (TEM) image of a spin labeled assembly is presented in Fig. 1 *c*. Oligo- and polyproline spacers are often incorporated into biomaterials, in order to enhance cellular interaction with functionalized end groups (34,35). A major advantage of prolines over other spacers is their rigidity; oligo- and polyprolines are often used as "molecular rulers" because of their tendency to coil into stiff helices with known lengths (36,37).

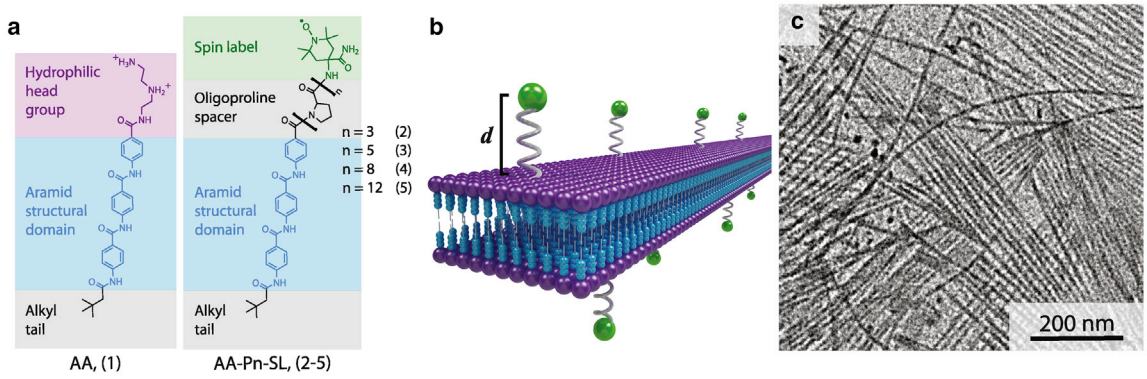

FIGURE 1 Self-assembly of nanoribbons with spin labels tethered at known distances off the surface. (*a*) A cationic aramid amphiphile (AA, compound **1**) and spin labeled analogs with oligoproline spacers (AA-Pn-SL, **2–5**) self-assemble spontaneously in water. (*b*) AA (compound **1**) is co-assembled with 2 mol % **2**, **3**, **4**, or **5** to form nanoribbons in water with integrated oligoproline-linked spin labels at distances, *d*, off the nanoribbon surface. (*c*) A representative transmission electron micrograph of compound **1** co-assembled with compound **5** shows that spin labeled oligoproline AAs do not affect the geometry or dimensions of AA nanostructures. To see this figure in color, go online.

Please cite this article in press as: Lindemann et al., A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.08.042

The dynamic structure of protein-binding peptides can also play a significant role in peptide interactions (4). Therefore, the length dependence of dynamics in oligoprolines is likely important to the behavior of binding sites attached to the end of the helix. However, these molecules are rigid, resulting in slow conformational changes (37). In this regime, self-consistent analysis of EPR results becomes challenging, making it an ideal test case for new analytical methods.

METHODS

Background on the SLE model and fitting function

The theory of EPR spectral simulation is complex and has been more thoroughly described elsewhere (10,13,14). In brief, EPR spectra of nitroxide radicals may be computed numerically by iteratively solving the SLE, which depends on several parameters. The SLE model treats electronic and nuclear spins as quantum mechanical objects while treating the reorientation of these spins classically. In particular, spectra are usually defined by an electron's gyromagnetic $\boldsymbol{g}$-tensor, its hyperfine $\boldsymbol{A}$-tensor, and its rotational diffusion $\boldsymbol{D}$-tensor, which contains elements inversely related the electron's rotational correlation times, called the $\boldsymbol{\tau}$-tensor. A set of ordering potentials, c_{20}, c_{22}, c_{40}, c_{42}, and c_{44} (listed in order of significance for describing ordering), describe the tendency of spins to become partially ordered within their local environment. Positive or negative values for these coefficients define different population distributions for nitroxide spins, which in turn can modulate the effects of orientation on EPR spectra (38). Here, we limit ourselves to the case of isotropic diffusion with an ordering potential, in wich all diagonal components of the diffusion tensor are equal to the scalar D_R and all off-diagonal components are zero.

Simply simulating EPR spectra is usually insufficient for interpreting real data sets. Instead, we perform nonlinear fitting to find simulation parameters that best describe experimental data sets; this is done using an algorithm for χ^2 minimization. χ^2 (defined in Eq. 1) is a function quantifying the difference between experimental and model-predicted values as a function of a set of fit parameters, $\boldsymbol{c}$ (11,13).

$$x^2(\mathrm{c}) = \sum_{i=1}^{n} \frac{\left[I_{\exp}(\omega_i - \omega_0) - I_{\mathrm{mod}}(\omega_i - \omega_0, \mathrm{c})\right]}{\sigma_i^2}^2. \quad (1)$$

In this equation, I_{exp} denotes the experimental spectral intensity at a frequency ω_i, which is taken with reference to frequency ω_0 (ω_0 is typically chosen to be 0). σ_i represents an uncertainty associated with each point, and it weights the fit to more closely conform to points with greater experimental significance. The set $\boldsymbol{c} = \boldsymbol{c_{min}}$, which most effectively minimizes χ^2, is the set of parameters that best explains the spectral data. Unfortunately, the χ^2 function is nonconvex, and therefore, it can have any number of local minima that often trap algorithms designed to minimize χ^2 (39). The algorithms used for this purpose range from the relatively simple family of gradient-based methods, such as gradient descent and Levenberg-Marquardt, to more sophisticated techniques such as simulated annealing and particle swarm methods (30,39).

All optimization algorithms for χ^2 are iterative, therefore we cannot be assured that any will find the global minimum with perfect confidence. Nonetheless, many effective numerical algorithms exist for global optimization, particularly over bounded intervals. In such cases, simulated annealing, genetic algorithms, particle swarms, and Levenberg-Marquardt Monte Carlo methods will all typically agree on a global optimum if run for a sufficient number of iterations. In modeling, the hardest problem is rarely identification of the global χ^2 minimum. Much more challenging is the process of understanding the extent to which an identified global minimum can be trusted to describe observed data (30,31).

The χ^2 landscape of a nonlinear fit is commonly populated by a large number of local minima, often clustered in the vicinity of the global optimum. Local minima are defined as regions where the Levenberg-Marquardt algorithm satisfies a convergence criterion—i.e., at which the χ^2 function's value approaches invariance or the algorithm's step size becomes sufficiently close to zero. In the CSCA software, the tolerances for these parameters remain as the NLSL defaults, but can be adjusted by a user to suit their needs.

The observation of a large number of clustered local minima can be referred to as "multimodality" or "ruggedness" (39). If multiple experimental repetitions are collected and analyzed, random variations will result in random deviations within this cluster, suggesting that global optima are statistically uncertain. In some cases, multiple distinct clusters of optima exist, indicating that multiple models are good fits to the data. Under these circumstances, deciding which model is more physically accurate requires an understanding of the physics of the problem, which can help constrain the fit to a particular region of parameter space (30). In rarer cases, obvious clusters of local minima are misleading, and a truly accurate model is ill-informed by χ^2 analysis alone. Therefore, true insight into the nature of the model is required for accurate fitting of data (31).

Fitting protocols and analysis of the χ^2 landscape

Here, we consider that the uncertainty associated with fitting always arises from two sources, experimental or fit uncertainty. Experimental uncertainty arises from noise, artifacts, or other error associated with data collection. Fit uncertainty arises because multiple sets of parameters describe a given data set with sufficient accuracy, and we are unable to meaningfully distinguish between them. As we demonstrate in the Results section, fit uncertainty is important, even when we analyze ideal, simulated spectra. Therefore, fit uncertainty encompasses the effects of overfitting within a rugged cluster, as well as the possibility that the SLE model or χ^2 function interacts unexpectedly with data artifacts. Our results suggest that for CW-EPR in the X-band, fit uncertainty is usually a much more significant problem than experimental error.

In most cases, the contribution of experimental error is nearly negligible for CW-EPR fitting, particularly with the development of wavelet-based methods for denoising spectral data (40). Nonetheless, we can account for the effects of any residual experimental noise in EPR fitting via a Monte Carlo approach (30). By fitting the same spectrum many times, randomly varying the experimental noise, we ensure that we accurately incorporated the effect of noise into any uncertainty estimates we generate. In our software, we refer to this option as Monte Carlo resampling. To describe fit uncertainty, we run multiple Levenberg-Marquardt optimization algorithms with a randomly varied starting point (Monte Carlo Levenberg-Marquardt (MCLM)). The iteration of this process within a bounded region identifies a large number of local χ^2 minima, as well as the function's global minimum. As a minor technical note, our algorithms operate via minimization of the reduced χ^2 function, defined as $\chi^2_\nu = \chi^2/\nu$, where ν is the number of degrees of freedom available to the fit. These have the same optima because ν is constant.

We select the subset of the minima whose values fall below a threshold value (χ^2_{thresh}). The threshold value is small enough that all models accurately describe the data but large enough to provide a statistically representative sampling of local χ^2_ν minima for the cluster containing the global optimum. We use this cluster of good fits to assign a univariate distribution to each parameter and treat this as the fit uncertainty. This distribution allows us to assign confidence intervals to each parameter. Because the values of parameters within a cluster are often correlated, these distributions probably overstate the uncertainty of our fitting process; however, this process allows us to assign reliable uncertainty estimates to fitted parameters. We combined the MCLM method with Monte Carlo spectral variation to account for both experimental and fit uncertainty.

Please cite this article in press as: Lindemann et al., A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.08.042

Poor reproducibility of CW-EPR fitting often occurs when the positions of global optima vary considerably within their good-fit cluster. In these cases, we observe the positions of the clusters themselves to be quite reproducible. We therefore take the approach of assigning a measure of central tendency to fit clusters to achieve self-consistency. Because these clusters tend to have unusual shapes, we compare three metrics for central tendency—the marginal median, the geometric median, and the medoid—to identify which of these describes the data accurately. A fourth method, the mean, was disregarded because it produced irregular results due to the disproportionate weight of outlying data. Precise formulas for these metrics are available in the Supporting Materials and Methods. All four metrics are included within our software package, and our comparisons between these and the global optima are presented in the results section.

RESULTS

Peptide spin labels incorporated into self-assembled nanoribbons

We synthesized four spin labeled molecules whose chemical structures are illustrated in Fig. 1 (with synthetic details described in Table 1). These spin labels were incorporated by co-assembly into AA nanoribbons. CSCA analysis of nanoribbons spin labeled with compounds 2–5 reveal room temperature rotational diffusion rates, D_R, reported in Table 1. D_R-values are smallest for the shortest proline sequences, indicating that the motion of these peptides are the most restricted. This trend, coupled with the observed lack of concentration dependence, is a good indicator that labeled samples integrate into the self-assembled nanostructure. If the oligoprolines had been dissolved in water rather than tethered to a nanostructure, the shorter peptides would ordinarily move faster than longer ones (41). If the labeled molecules had instead precipitated into a unique phase, spin-spin interactions would have resulted in a significantly different EPR spectral lineshape. Therefore, we conclude that the labels reside at positions separated in space tethered to the surface of the self-assembled nanoribbons.

Measures of central tendency achieve self-consistent descriptions of rotational diffusion rates

We used CSCA analysis to produce Arrhenius plots of the rotational diffusion rates, D_R, of each sample as a function of temperature. In almost every case, comparison of the global χ^2_ν optima showed a high degree of inconsistency of D_R and the appearance of non-Arrhenius behavior, particularly in colder, slower-moving samples. For these samples, the mean also proved inconsistent, so we chose to neglect it from further analysis. All three median-based methods, the marginal median, geometric median, and medoid, agree on the optimal value of D_R at every temperature, resulting in linear Arrhenius behavior of D_R that is typical for diffusive processes. Fig. 2, *a–d* illustrates the Arrhenius fits produced by each median method and the global optimum for compound **1** coassembled with **2**, **3**, **4**, and **5**, respectively. 50% confidence intervals based on the univariate histogram of D_R-values are overlaid, indicating that the global optimum frequently falls at an extreme position within the cluster, a fundamental weakness of the approach. Fig. 2 *e* presents the R^2 obtained for each Arrhenius fit and shows that the median methods all describe a thermally activated process, whereas the global optimum does not.

Although D_R was remains similar between AA-P8-SL and AA-P12-SL, the shorter peptides exhibit much slower rates of conformational change. Multiple explanations may justify this, but the likeliest are either 1) intermolecular interaction between the spin labels and adjacent amphiphile headgroups, which are only possible in the shortest peptides, or 2) an effect of solvent dynamics because water molecules are influenced by nearby surfaces and in particular are slowed down at supramolecular interfaces with high surface charge density (42).

Significantly, median methods reveal that for slowly moving probes, an accurate value for D_R should fall well below the value associated with the global optimum. This observation is consistent with the limitations of the SLE model, which is less reliable when rotational correlation times are longer (14). This broadening is manifest in broadening of good-fit clusters and the correspondingly larger error bars observed at low temperatures, confirming that slow-motion spectra are more prone to overfitting. Thus, our results demonstrate that cluster-based fit analyses substantially increase the range of D_R-values for which fitting can provide meaningful insights using the SLE model.

One measurement artifact may have biased our analysis is the presence of a dilute Mn^{2+} radical in the Critoseal used to close capillary tubes. Although we performed background subtraction from a reference tube, slight variations in the

TABLE 1 Spin Labeled Oligoprolines Tethered to the Surface of AA Nanoribbons

Compound	Name	Surface Sequence	Expected Molecular Weight	Observed Molecular Weight	Approximate Helix Length (Å)	Medoid D_R ($T = 299$ K)	Medoid τ_R ($T = 299$ K)
2	AA-P3-SL	-PPP**J**	961.48	961.50	9.3	5.5×10^7 s	3.2×10^{-9} s
3	AA-P5-SL	-PPPPP**J**	1155.58	1155.61	15.5	6.5×10^7 s	2.7×10^{-9} s
4	AA-P8-SL	-PPPPPPPP**J**	1446.75	1446.76	24.8	9.8×10^7 s	1.7×10^{-9} s
5	AA-P12-SL	-PPPPPPPPPPPP**J**	1834.96	1834.98	37.2	9.8×10^7 s	1.7×10^{-9} s

The spin label TOAC residue is denoted by **J**. The expected and observed molecular weights are presented, with supporting LC-MS data presented in Fig. S1. The expected oligoproline helix length is calculated based on the known left-handed poly-proline type 2 (PPII) helix length of 0.31 nm per proline residue (46). Correlation times were computed by the formula $\tau_R = 1/(6D_R)$.

Please cite this article in press as: Lindemann et al., A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.08.042

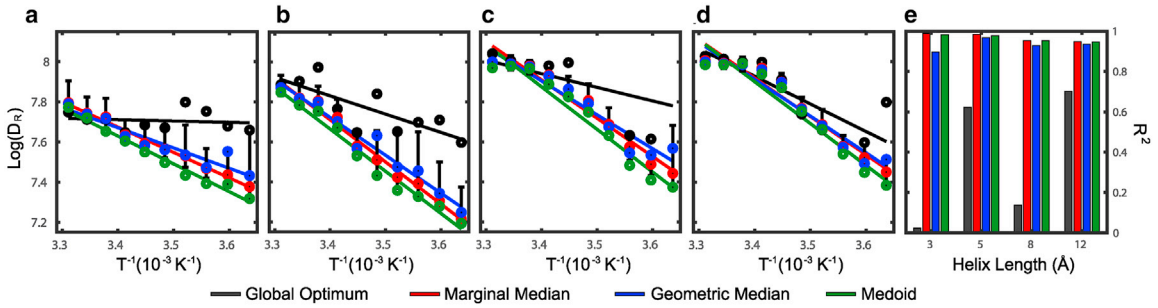

FIGURE 2 Arrhenius plots describe D_R derived from CSCA analysis of EPR fits. Arrhenius plots of co-assemblies of (*a*) compound **1** and **2**, in which the spin label is positioned close to the nanoribbon surface; (*b*) compound **1** and **3**; (*c*) compound **1** and **4**; and (*d*) compound **1** and **5**, in which the spin label is far from the nanoribbon surface, are given. Within each Arrhenius plot, D_R-values computed using the global optimum, the marginal median, the geometric median, and the medoid are included. The mean was excluded because, like the global optimum, it lacks self-consistency in D_R. A 50% confidence interval for each D_R is also presented, based on the D_R histogram. (*e*) The R^2 values for each Arrhenius-type fit are presented. Each measure of central tendency agrees on the Arrhenius behavior of D_R, with the exception of the global optimum. This figure is available for viewing in color online.

amount of sealant used resulted in a minor distortion of the spectrum in the low-field range. Fortunately, this signal only partially overlaps with the EPR signal, so we are confident that its effects are minimized during background subtraction. Furthermore, the inclusion of Monte Carlo data resampling to each fit helps ensure that our results are robust. An unsubtracted background is presented in Fig. S83.

The geometric median and medoid provide the most physically representative estimates of other parameters

Despite their agreement on D_R, the measures of central tendency often fail to agree on appropriate values for other parameters, namely the c_{20} orienting potential and the Gaussian line broadening, γ_0. This shortcoming results from the unusual curvature of good-fit clusters in parameter space. Fig. 3 *a* represents the χ^2_ν landscape of these parameters for a representative AA-P3-SL sample collected at 275 K. This analysis contains χ^2_ν-values calculated during every iteration of an MCLM analysis, which started from 500 initial positions. Fig. 3 *a* shows the positions of χ^2_ν minima, depicted by purple points, which form a cluster containing the global optimum. This optimum and each metric of central tendency are labeled. For these spectra, we used a three-parameter model, fitting for c_{20}, γ_0, and D_R. The two-dimensional scatter plots in Fig. 3 *a* flatten the χ^2_ν function to show its dependence on each pair of

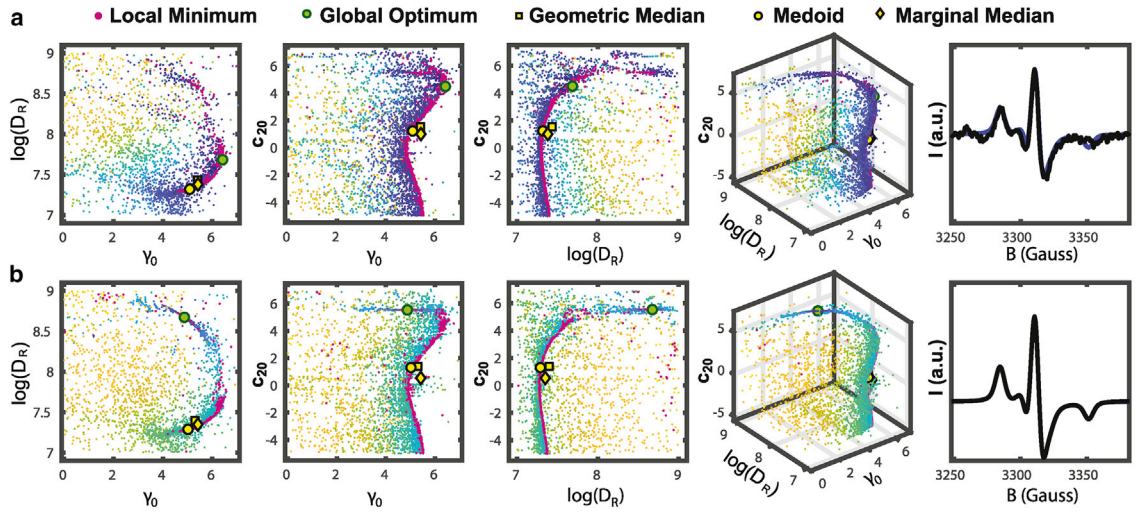

FIGURE 3 CSCA representations of the χ^2_ν landscape for a spectral fit. (*a*) Scatter plots of the χ^2_ν landscape associated with fitting the EPR spectrum of AA-P3-SL at 275 K are given. The first three plots flatten the χ^2_ν landscape to a function of two parameters, whereas the fourth preserves its full, three-dimensional shape. Color indicates the value of χ^2_ν. Smaller values are rendered in blue, and larger values are rendered in yellow and intermediate values in shades of green. Local minima are overlaid in purple, and the global optimum and the various cluster medians are plotted using large symbols in each figure. The final plot shows the raw EPR spectrum (*black*) as well as the best fit (*blue*). (*b*) This row is identical to the row above it, except that instead of fitting experimental data, we analyzed the best-fit spectrum taken from the line above. The position of the new global optimum shifts significantly, indicating that the initial data set was overfitted. In contrast, the median methods remained highly self-consistent. This figure is available for viewing in color online.

Please cite this article in press as: Lindemann et al., A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.08.042

variables, and the complete dependence is presented in the three-dimensional plot.

A three-parameter approach was chosen as the simplest model capable of describing the data to reduce the risk of overfitting. Necessarily, our assumptions—e.g., the isotropy of D_R and the thermal stability and symmetry of the hyperfine and $\boldsymbol{g}$-tensors—bias our results in favor of simple and symmetric views of the data. However, these assumptions are physically reasonable in the case of spin labeled peptides, and, in our experience, including more parameters often results in overfitting.

In Fig. 3 *b*, we fit the simulated EPR spectrum that best describes the original data shown in Fig. 3 *a*. The same randomized starting values were used to analyze both the experimental and the simulated data sets, with no synthetic noise added. In principle, fitting a simulated spectrum should identify $\chi^2_{min} = 0$ precisely at the position of the parameters used for simulation. However, we observed that the position of the global χ^2_ν optimum deviates significantly from the best-fit parameters used to generate this spectrum. In such cases, when we attempted to assign χ^2_{thresh} for 95% confidence intervals via a classical approach (using the χ^2 distribution), we found that no other observed points fell below χ^2_{thresh} (43). Yet, we know that a point exists in that neighborhood where χ^2_{min} is zero (to within round-off error), indicating that the χ^2 function for this kind of analysis has achieved such pathological roughness that our algorithms could not sufficiently sample the neighborhood of the minimum. When started at the correct coordinates, the algorithm converges to the correct minimum.

This inconsistency reveals the uncertainty of the global χ^2_ν optimum as a representation of spin label dynamics. In contrast, all median methods were self-consistent in comparisons of simulated versus real data. This consistency can be verified by inspection of Figs. S3–S42, which are analyses of every spectrum collected in this work. Therefore, using CSCA, we find that clusters of good fits are self-consistent, even if positions of the global optima are not. Furthermore, the Arrhenius behavior of all of these metrics suggests that cluster position depend logically on temperature, further supporting this method of analysis.

As illustrated in Figs. 3 and S3–S42, we find that not every median-based analysis performs equally well. Although they all agree on the value of D_R, the U-shaped nature of good-fit clusters means that the marginal median frequently selects points that do not fall within the cluster itself. In contrast, the geometric median tends to conform quite closely to the cluster, and the medoid must fall within the cluster by its mathematical construction. In general, we prefer the medoid because this metric necessarily falls within the best-fit cluster. One critique of this method is that reporting cluster centers in place of global optima has no statistical advantage because, unlike the global optimum, cluster centers cannot serve as maximal likelihood estimators for the fit parameters. However, the pathological nature of our χ^2_ν function suggests that we cannot be confident in the positions we have identified as global optima. Therefore, relying on this more heuristic approach for identifying optimal parameters is likely, in many cases, to give more physical and reproducible fitting results. In our experience, this approach is particularly well suited for fitting CW-EPR spectra in the slow-motion regime, but its applicability to other types of data fitting relies on the particulars of those analyses.

From these figures, as well as the error bars reported in Fig. 2, we observe that overfitting occurs most frequently in the rigid limit, at which the SLE model begins to break down. Physically, this phenomenon suggests that as rotational diffusion slows beyond the resolution of the SLE model, our ability to precisely define D_R is reduced. Nonetheless, we observe that the cluster positions are relatively self-consistent and predictable, a key advantage of this approach.

Classical methods exist for predicting confidence intervals for fit parameters, however these methods are only capable of providing realistic estimates when the data adequately sample the island of confidence. Therefore, we take a heuristic approach, artificially increasing χ^2_{thresh} to exceed the roughness of χ^2_ν near the minimum and giving each valid local minimum equal statistical weight. In this way, we establish univariate distributions for the uncertainty in each parameter. Therefore, our error bars do not depend on the precise position of the global best fit. In our experience, the error bars greatly overstate error because good-fit clusters tend to follow predictable patterns rather than falling randomly within the uncertainty window. However, we view this as a benefit because we have greater confidence that the accurate value falls within our prescribed range. Notably, the global best fit rarely falls within the range containing the central 50% of D_R-values. We acknowledge that the accuracy of this calculation could be biased by the nonuniform distribution of local minima, and as a result, the confidence interval boundaries should be treated with a degree of caution.

Finally, we wish to specifically address the Mn^{2+} subtraction artifact noted in Measures of Central Tendency Achieve Self-Consistent Descriptions of Rotational Diffusion Rates. This artifact may have influenced our observed values of D_R, and we have indicated our approach toward correcting the role of this distortion. However, we are confident that these distortions have negligible bearing on the results described in this section, which describe the nature of EPR spectral fitting in general. The χ^2_ν landscape for fits of simulated spectra still includes the broad "islands of uncertainty" containing local χ^2_{min}. Because simulated spectra do not contain Mn^{2+} subtraction artifacts, we are confident that the properties observed in the χ^2_ν landscape apply generally to CW-EPR fitting of nitroxides in the X-band.

Please cite this article in press as: Lindemann et al., A Global Minimization Toolkit for Batch-Fitting and χ^2 Cluster Analysis of CW-EPR Spectra, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.08.042

Observed activation energies closely correspond to predictions based on the energy landscape model

The activation energies were computed for each peptide as a function of approximate distance from the nanoribbon surface (Fig. 4). We estimated this distance based on the helix length presented in Table 1 but acknowledge that the true distance could be affected by the absence of a charged headgroup in AA-Pn-SL compounds, the added length of TOAC, or the angle between the oligoproline helix and the assembly's surface. For peptides long enough to form oligoproline helices (longer than three residues), the activation energy of diffusion is nearly constant, ~40 kJ/mol. The calculated values of activation energy of diffusion exceeded those observed for other peptides, which typically fall below ~10 kJ/mol (4). However, the peptides with activation energies of diffusion less than 10 kJ/mol are intrinsically disordered, whereas oligoprolines exhibit a stable conformational structure, especially when the oligoproline contains four or more residues, in which case stable helix formation occurs (36,37).

In previous works, the activation energy of peptide diffusion was determined by EPR and was shown to correspond to the characteristic energy barrier associated with diffusion through a rough energy landscape (4,5). In the case of oligoproline helices, these landscapes have been described by molecular dynamics simulations, making our oligoproline-based samples an ideal test of this conclusion (37). These studies indicated that the conformational energy landscape of oligoproline contains a significant number of energy wells, each corresponding to a stable point between pure polyproline type 1 (PPI) and pure polyproline type 2 (PPII) secondary structure. The observed energy barrier for transitions between these wells is ~40 kJ/mol, agreeing closely with our observations (37). Thus, our observed activation behavior is consistent with the picture of rigid oligoproline helices, whose motions are associated with surveying of the oligoproline energy landscape.

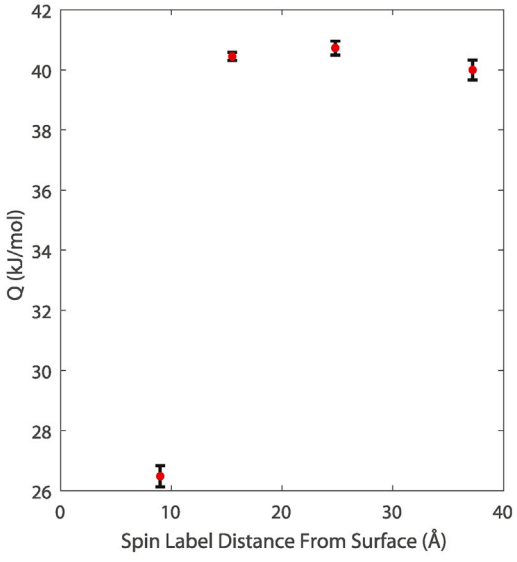

FIGURE 4 Activation energies of diffusion, Q, of spin labels tethered at known distances off a nanoribbon surface. Values of Q are computed from the medoid values for D_R as a function of spin label height, with error bars based on the standard error of each linear fit. The oligoproline spacer of compound **2** is too short to form a complete helix (<4 residues). The activation energies of diffusion of compounds **3**–**5** co-assembled with compound **1** are similar—~40 kJ/mol— and comparable to the energy barrier associated with conformational change in oligoproline helices. This figure is available for viewing in color online.

DISCUSSION

Here we present a software, CSCA, for batch-fitting and χ-squared cluster analysis of CW-EPR spectra. This software allows one to self-consistently place an optimum within a rugged cluster of local minima. Furthermore, we identify CW-EPR as an experimental area in which statistical methods may provide valuable insights into the physics of a system. We use cluster-based methods that select the medoid, the geometric median, or the marginal median to demonstrate that the diffusional motion of peptides tethered to a nanostructure is a thermally activated (Arrhenius) process, and we find that the activation energy matches the characteristic energy barrier associated with conformational change in oligoproline helices (37). In concert with rapid protocols for peptide synthesis (4,44), this methodology provides a robust and high-throughput approach for the experimental analysis of conformational energy landscapes in peptides. We employ a medoid-based approach to assure that good fits are fully descriptive of experimental data (i.e., they fall within an obvious cluster of local minima) and self-consistent enough to allow Arrhenius analysis of D_R. By making these capabilities available publicly through the CSCA software toolkit, this approach may be adapted to suit the needs of a variety of experimenters.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2020.08.042.

AUTHOR CONTRIBUTIONS

W.R.L. and J.H.O. conceived and designed the research, which is adapted from work published in W.R.L.'s doctoral thesis (45). W.R.L. and J.H.O. composed the manuscript. W.R.L. and T.C.T. synthesized and chemically characterized the compounds. W.R.L. developed the computational tools, carried out EPR experiments, and analyzed data. J.H.O. provided project administration, funding acquisition, and supervision.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of Bradley L. Pentelute, who provided instrumentation used for sample synthesis and purity determination.

This material is based upon work supported by the National Science Foundation under Grant No. CHE-194550. W.R.L. and T.C.-T. acknowledge the support of the National Science Foundation Graduate Research Fellowship Program (Grant No. 1122374).

REFERENCES

1. Cheng, C.-Y., and S. Han. 2013. Dynamic nuclear polarization methods in solids and solutions to explore membrane proteins and membrane systems. *Annu. Rev. Phys. Chem.* 64:507–532.
2. Ortony, J. H., C. J. Newcomb, …, S. I. Stupp. 2014. Internal dynamics of a supramolecular nanofibre. *Nat. Mater.* 13:812–816.
3. Zheng, W., D. de Sancho, and R. B. Best. 2016. Modulation of folding internal friction by local and global barrier heights. *J. Phys. Chem. Lett.* 7:1028–1034.
4. Lindemann, W. R., E. D. Evans, …, J. H. Ortony. 2020. Quantifying residue-specific conformational dynamics of a highly reactive 29-mer peptide. *Sci. Rep.* 10:2597.
5. Lindemann, W. R., A. J. Mijalis, …, J. H. Ortony. 2020. Conformational dynamics in extended RGD-containing peptides. *Biomacromolecules.* 21:2786–2794.
6. Vogt, A. D., and E. Di Cera. 2012. Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry.* 51:5894–5902.
7. Paul, F., and T. R. Weikl. 2016. How to distinguish conformational selection and induced fit based on chemical relaxation rates. *PLoS Comput. Biol.* 12:e1005067.
8. Lindemann, W. R., and J. H. Ortony. 2018. Probing local molecular properties of self-assembled systems with electron paramagnetic resonance (EPR). *In* Self-Assembling Biomaterials: Molecular Design, Characterization and Application in Biology and Medicine. S. H. Azevedo and R. M. P. da Silva, eds. Woodhead Publishing.
9. Freed, J. H. 1976. Theory of slow tumbling ESR spectra for nitroxides. *In* Spin Labeling: Theory and Applications. L. J. Berliner, ed. Academic Press, pp. 53–112.
10. Schneider, D. J., and J. H. Freed. 1989. Calculating slow motional magnetic resonance spectra: a user's guide. *Biol. Magn. Reson.* 8:1–76.
11. Budil, D. E., S. Lee, …, J. H. Freed. 1996. Nonlinear-least-squares analysis of slow-motion EPR spectra in one and two dimensions using a modified levenberg–marquardt algorithm. *J. Magn. Reson.* 120:155–189.
12. Hustedt, E. J., and A. H. Beth. 1999. Nitroxide spin-spin interactions: applications to protein structure and dynamics. *Annu. Rev. Biophys. Biomol. Struct.* 28:129–153.
13. Stoll, S., and A. Schweiger. 2006. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* 178:42–55.
14. Stoll, S., and A. Schweiger. 2007. EasySpin: simulating cw ESR spectra. *Biol. Magn. Reson.* 27:299–321.
15. Mitra, S. K., and J. H. Freed. 2011. Molecular motions. *In* Multifrequency Electron Paramagnetic Resonance. S. K. Misra, ed. Wiley, pp. 497–543.
16. López, C. J., S. Oga, and W. L. Hubbell. 2012. Mapping molecular flexibility of proteins with site-directed spin labeling: a case study of myoglobin. *Biochemistry.* 51:6568–6583.
17. Hubbell, W. L., C. J. López, …, Z. Yang. 2013. Technological advances in site-directed spin labeling of proteins. *Curr. Opin. Struct. Biol.* 23:725–733.
18. Schreier, S., J. C. Bozelli, …, C. R. Nakaie. 2012. The spin label amino acid TOAC and its uses in studies of peptides: chemical, physicochemical, spectroscopic, and conformational aspects. *Biophys. Rev.* 4:45–66.
19. Xia, Y., Y. Li, …, R. H. Grubbs. 2011. EPR study of spin labeled brush polymers in organic solvents. *J. Am. Chem. Soc.* 133:19953–19959.
20. Ortony, J. H., D. S. Hwang, …, S. Han. 2013. Asymmetric collapse in biomimetic complex coacervates revealed by local polymer and water dynamics. *Biomacromolecules.* 14:1395–1402.
21. Newcomb, C. J., S. Sur, …, S. I. Stupp. 2014. Cell death versus cell survival instructed by supramolecular cohesion of nanostructures. *Nat. Commun.* 5:3321.
22. Altenbach, C. 2010. Research Website of Wayne Hubbell. http://www.biochemistry.ucla.edu/Faculty/Hubbell/.
23. Meirovitch, E., A. Nayeem, and J. H. Freed. 1984. Analysis of protein-lipid interactions based on model simulations of electron spin resonance spectra. *J. Phys. Chem. B.* 88:3454–3465.
24. Liang, Z. C., and J. H. Freed. 1999. An assessment of the applicability of multifrequency ESR to study the complex dynamics of biomolecules. *J. Phys. Chem. B.* 103:6384–6396.
25. Sezer, D., J. H. Freed, and B. Roux. 2008. Simulating electron spin resonance spectra of nitroxide spin labels from molecular dynamics and stochastic trajectories. *J. Chem. Phys.* 128:165106.
26. Sezer, D., J. H. Freed, and B. Roux. 2009. Multifrequency electron spin resonance spectra of a spin-labeled protein calculated from molecular dynamics simulations. *J. Am. Chem. Soc.* 131:2597–2605.
27. Oganesyan, V. S. 2011. A general approach for prediction of motional EPR spectra from Molecular Dynamics (MD) simulations: application to spin labelled protein. *Phys. Chem. Chem. Phys.* 13:4724–4737.
28. Zerbetto, M., D. Licari, …, A. Polimeno. 2013. Computational tools for the interpretation of electron spin resonance spectra in solution. *Mol. Phys.* 111:2746–2756.
29. Martin, P., S. Stoll, and D. Thomas. 2017. Simulating electron paramagnetic resonance spectra of slow-motion systems in the time domain. *Biophys. J.* 112:445a–446a.
30. Motulsky, H., and A. Christopoulos. 2003. Fitting Models to Biological Data using Linear and Nonlinear Regression., Second Edition. GraphPad Software, Inc., San Diego, CA.
31. Fernández Slezak, D., C. Suárez, …, G. Stolovitzky. 2010. When the optimal is not the best: parameter estimation in complex biological models. *PLoS One.* 5:e13283.
32. Lappan, U., B. Wiesner, and U. Scheler. 2015. Rotational dynamics of spin-labeled polyacid chain segments in polyelectrolyte complexes studied by CW EPR spectroscopy. *Macromolecules.* 48:3577–3581.
33. Lappan, U., B. Wiesner, and U. Scheler. 2016. Segmental dynamics of poly(acrylic acid) in polyelectrolyte complex coacervates studied by spin-label EPR spectroscopy. *Macromolecules.* 49:8616–8621.
34. Valsesia-Wittmann, S. 2001. Role of chimeric murine leukemia virus env β-turn polyproline spacers in receptor cooperation. *J. Virol.* 75:8478–8486.
35. Pallarola, D., A. Bochen, …, H. Kessler. 2014. Interface immobilization chemistry of *c*RGD-based peptides regulates integrin mediated cell adhesion. *Adv. Funct. Mater.* 24:943–956.
36. Doose, S., H. Neuweiler, …, M. Sauer. 2007. Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc. Natl. Acad. Sci. USA.* 104:17400–17405.
37. Moradi, M., V. Babin, …, C. Sagui. 2009. Conformations and free energy landscapes of polyproline peptides. *Proc. Natl. Acad. Sci. USA.* 106:20746–20751.
38. Earle, K. A., and D. E. Budil. 2006. Calculating slow-motion ESR spectra of spin-labeled polymers. In: Advanced ESR Methods in Polymer Research, Schlick S, ed. (Hoboken, NJ), pp. 53–84.
39. Weise, T. 2009. Global optimization algorithms–theory and application. *Self-Published Thomas Weise*.
40. Srivastava, M., C. L. Anderson, and J. H. Freed. 2016. A new wavelet denoising method for selecting decomposition levels and noise thresholds. *IEEE Access.* 4:3862–3877.

41. de Sancho, D., A. Sirur, and R. B. Best. 2015. Molecular origins of internal friction effects on protein folding rates. *Nat. Commun.* 79:211–227.

42. Ortony, J. H., B. Qiao, …, S. I. Stupp. 2017. Water dynamics from the surface to the interior of a supramolecular nanostructure. *J. Am. Chem. Soc.* 139:8915–8921.

43. Bevington, P. R., and D. K. Robinson. 2003. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill Education, New York, pp. 194–212.

44. Mijalis, A. J., D. A. Thomas, III, …, B. L. Pentelute. 2017. A fully automated flow-based approach for accelerated peptide synthesis. *Nat. Chem. Biol.* 13:464–466.

45. Lindemann, W. R. 2020. Dynamics characterization for designing functional soft materials. PhD thesis. Massachusetts Institute of Technology.

46. Adzhubei, A. A., M. J. E. Sternberg, and A. A. Makarov. 2013. Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* 425:2100–2132.