Dynamics Characterization for Designing Functional Soft Materials

by

William Robin Lindemann

Submitted to the Department of Materials Science and Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Materials Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Author .....
Department of Materials Science and Engineering
May 8, 2020

Certified by .....
Julia H. Ortony
Associate Professor of Materials Science and Engineering
Thesis Supervisor

Accepted by .....
Frances M. Ross
Professor of Materials Science and Engineering
Chair, Department Committee on Graduate Theses

Dynamics Characterization for Designing Functional Soft Materials

by

William Robin Lindemann

Submitted to the Department of Materials Science and Engineering
on May 8, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Materials Science and Engineering

Abstract

In solutions, the dynamic behavior of soft materials is often critical to their function. In biological materials such as proteins and peptides, the edict that 'structure dictates function' has been supplanted in recent decades by recognition that features like intrinsic disorder, conformational distribution, and solvent dynamics often play a part which is equally fundamental to the binding and reactivity of these materials. The same revelation holds for many other functional soft materials, including abiotic peptides and self-assembling materials, where function is controlled by the dynamic behavior of both the compound and the substrate. In this work, I elucidate the role of dynamics in several significant functional polyamides by the synthesis and characterization of samples spin-labeled for electron paramagnetic resonance (EPR) spectroscopy. By this approach, I developed insight into several soft-materials systems, including abiotic peptide tags, combinatorially selected for bioconjugation; fibronectin mimetic peptides, designed for therapeutic purposes, biomaterials and drug delivery; and finally, novel, self-assembling polyamide materials designed for water purification and energy conservation.

Thesis Supervisor: Julia H. Ortony
Title: Associate Professor of Materials Science and Engineering

Acknowledgments

When I was in high school, attending MIT was a dream that I never thought I would realize. For my first several months here, I still had trouble believing it. The experience has been challenging – in ways I expected, and in ways I never forsaw. I am inexpressibly grateful to the friends and family who bore me through it. This work is dedicated to you.

To my parents, Bill and Ruth: I can only say that I am incredibly grateful. For as long as I can remember, you have done nothing but support me. When I was excited by a new experiment or result, I knew I could make my best days better by calling you. And on my worst days, when life reduced me to a trembling ball of nerves, I knew that talking to you would make me feel better. Every time I've jumped off a cliff, you've both been there – cheering for me when I fly, but ready to catch me when I fall. I don't know if I'll ever be able to explain how much that has meant to me. Thank you, and I love you.

To my fiance, Molly Parsons: our relationship started 5 years ago, with a leap of faith. I don't think either of us knew what we were in for when we started grad school, but through everything, I've been profoundly proud to have you on my team, and I'm glad, in turn, that you've chosen me for yours. You are my best friend, and I couldn't have gotten here without you.

To my brother, Geordie: thank you so much, for being the only person in my family who knows what I'm talking about half the time. I'm so glad to have your voice in my life, and I appreciate the many times you've let me vent about grad school.

To my labmates and collaborators, new and old: thank you all for your advice, your understanding, your kindness, and your friendship. And thank you for tolerating my incessant yammering. I can't imagine what this work would look like if I hadn't had such a supportive and brilliant team of people standing in my corner.

And finally, to my advisor, Julia: you are truly the person who most completely shaped my graduate experience at MIT, and in myriad ways, you made this work

possible.

Thank you all, for everything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Dynamic Behavior in Materials

For over 40 years, bioactive materials have played a pivotal role in the modern medicine. Bioactive materials, here defined as materials which interact constructively with cells in the human body, frequently hinge upon the incorporation of bioactive sites which provide functional value.

Few problems prove more vexing than studying and manipulating the conformational behavior of these bioactive sites. While the community now accepts the importance of dynamics to the activity of peptides and proteins[1–3], the dynamic properties of these materials remain particularly difficult to characterize. Nevertheless, a few experimental techniques, such as nuclear magnetic resonance (NMR), quasi-elastic neutron scattering (QENS), and electron paramagnetic resonance (EPR), can provide quantitative insight into the conformational dynamics of materials.

Dynamic behavior perhaps matters most to protein interactions. Proteins show tremendous potential as a tool for treating disease, healing injuries, and designing biocompatible implants and devices. By incorporating functional proteins or peptides into biomaterials, researchers have successfully facilitated bone repair[4–7], the growth of nerve and brain cells[8–11], the regeneration of dental pulp[12–14], and the formation of vasculature to speed the healing of wounds.[15–17] Their high specificity of function makes them incredibly versatile. Often, solving a biological problem simply means

finding the right protein for the job.

Although scientists recognized the importance of proteins in physiology over a century ago, we did not understand their chemical nature until the early 1950s.[18] First, in 1949, Sanger succeeded in sequencing the terminal amino acids of insulin – demonstrating the high predictability of amino acid positions in these materials.[19,20] The second great insight into protein properties came from Linus Pauling, who in 1951 predicted the helical structure of some proteins based on hydrogen bonding.[21] The revelation that a protein's structure and sequence help dictate its function gave birth to the field of molecular biology, paving the way for tremendous advances in science and medicine.

Over the next 60 years, we became more sophisticated in our attempts to characterize the structure and sequence of proteins. Today we have the protein data bank (PDB), which contains millions of crystallographic structures with atomic resolution, accessible to all, and typically confirmed using atomistic simulation and characterization techniques such as NMR and X-ray scattering.[22–31] Even in lieu of structural information, most proteins may be sequenced using mass spectrometry and database matching.[32–34]

In the 1960s, scientists began studying protein sequences in order to identify the regions involved in binding. We refer to these regions as 'epitopes'. The term, which originates in the field of immunology, initially applied only to subsequences of antigens (molecules that activate antibodies or T-cells).[35] In contemporary usage within the biomaterials community, 'epitope' refers to the active part of any protein. Epitopes have long interested immunologists, who aim to identify the active sites of antigens (sometimes called 'antigenic determinants') in order to produce vaccines, viral inhibitors, and other therapeutic agents.[36–40] Scientists then began identifying epitopes of other proteins for use in bioactive materials. By producing synthetic versions of these epitopes, known as biomimetic or epitope-mimetic peptides, researchers aim to imitate the function of the parent protein, while avoiding the drawbacks (such as toxicity or difficulty of production) associated with its use.

Some of the earliest examples of epitope identification came from Atassi, who

identified five epitopes of myoglobin and attempted to form general conclusions about functional sites in proteins.[41] In particular, Atassi advocated for the idea that all bonding subsequences of proteins: i) do not exceed six or seven amino acids, arranged consecutively; and ii) are very sensitive to mutation, losing most of their functionality with the replacement of even a single residue.

Although the scientific community still accepts some of Atassi's conclusions, scientists have reported counterexamples or exceptions to many of them. For instance, some myoglobin binding sites use several, disjointed regions of the protein sequence, rather than one continuous subsequence.[42] This led to the general classification of epitopes as continuous (sequential) or discontinuous (non-sequential). According to the work of Barlow, Edwards, and Thornton, discontinuous epitopes appear most often in globular proteins, since lengthy, continuous subsequences rarely reside at the protein surface.[43] Huang and Honda compiled a database of well-established examples of discontinuous antigen epitopes in 2006.[44]

One of the key ideas of twentieth century biology was the realization that in proteins, 'structure dictates function'. This model, called the lock-and-key mechanism for protein binding, remains fundamental to our understanding of the interactions between proteins and other molecules. It states that protein interactions occur primarily as a result of protein structure. This is particularly true in the case of proteins that target small molecules. However, in the context of protein-protein interactions, this model has proven insufficient, failing to explain the natural ubiquity of intrinsically disordered proteins in nature. For instance, hub proteins, which interact separately with large numbers of other proteins, seem to accomplish this feat through conformational changes in the chain backbone.[2] Most of these proteins display biological activity, despite the fact that they routinely diffuse through distinct structures over time.

These kinds of conformational variations occur regularly in nature, where proteins typically occupy a statistical distribution of distinct conformations, rather than one single conformation. A pair of mechanisms, known as the conformational selection model and the induced fit model, help to explain the role of conformation in protein

interactions.[45,46] The conformational selection model assumes that structure remains fundamental to the role of protein binding, and that proteins can only bind at times when they have already folded into a conformation resembling their bound state. In contrast, the induced fit mechanism assumes that the presence of a binding target induces a conformational change in the protein structure. Both of these mechanisms occur in protein binding systems, although conformational selection appears more often than induced fit.[45–49] In both of these models, binding kinetics depend upon the protein's conformational distribution. This, in turn, provides the first indication of the importance of dynamics to protein binding.

In biotechnology, we typically desire sequences with high affinity and rapid kinetics. When proteins change conformation slowly, the rate of diffusion between distinct conformations limits the protein binding kinetics, and may even modify the binding equilibrium. Importantly, this does not mean that faster-moving sequences are necessarily better. For one thing, structure still remains the dominant factor in reactivity. A slow-moving structure with a highly stable binding epitope will bind much more effectively than a structurally unstable molecule experiencing rapid conformational change. Moreover, binding in peptides likely depends on their dwell-time in favorable binding conformations. However, these models certainly demonstrate that understanding protein binding often requires a dynamic picture of these molecules, and an understanding of how complex properties like flexibility and the rate of conformational change affect functionality. Several studies support the importance of dynamics in protein binding.[1–3,50,51] However, very few of these examine the relationship between sequence, structure, and chain dynamics. Generally, these factors interconnect. For instance, a computational study demonstrated that functionally distinct proteins with similar structures possess similar dynamic properties.[52] Another study, analyzing the effect of mutation on chain dynamics, concluded that even after residue substitution, chain conformations and dynamics remain very similar to the unmutated protein.[53]

Experimental studies relating protein chain dynamics with function present a tremendous technical obstacle. Some techniques, such as NMR or QENS, enable direct observation of protein dynamics. Unfortunately, these techniques often require

isotopic labeling, a very difficult process in proteins and peptides.[1,51,54] Moreover, these techniques require very high protein concentrations, making them unrepresentative of real, biological conditions. Most other tools for measuring protein dynamics (e.g. dynamic light scattering, DLS, and X-ray photon correlation spectroscopy, XPCS) measure the diffusion of the overall molecule, rather than local chain dynamics.

One of the simplest approaches for measuring site-specific dynamics is by using electron paramagnetic resonance (EPR) – a technique that detects the motion of unpaired electrons in a material. Through the selective introduction of stable free-radicals as spin-labels, EPR allows site-specific measurement of chain motion in solution. Fitting the data with spectral simulations allows direct measurement of an unpaired electron's rotational diffusion coefficient – a quantity that relates to the rate at which a protein backbone changes conformation. This allows highly quantitative calculations of backbone dynamics, which agree with observables computed via molecular dynamics simulations.[55]

The high sensitivity and site-specificity inherent in EPR measurements makes this technique substantially more robust than NMR and QENS measurements, because it is still useful for dilute or multicomponent systems. Nonetheless, it has its own challenges, including the difficulty of synthesizing and characterizing sample, and the challenge of spectral fitting. Since I studied material dynamics using EPR, systematically overcoming these challenges became fundamental to this thesis.

Dynamic properties also relate to the behavior of self-assembling materials. For instance, scientists use EPR to study the behavior of thermodynamic phases in membranes, as well as the degradation of vesicles by antimicrobial peptides.[56–58] This approach can even distinguish distinct thermodynamic phases present within nanofibers comprised of a single, self-assembling amphiphile – a first step towards engineering the mechanical and thermal properties of these materials.[59]

In this thesis, I describe my efforts to characterize the dynamic behavior of biotic and abiotic peptides, as well as self-assembled oligamides, by developing and applying powerful new synthetic and analytical protocols to the characterization of soft

matter dynamics. With this approach I observed structural transitions and novel dynamic properties of molecules with outstanding potential for creating biological and functional materials.

Chapter 2

Experimental Techniques and Methods

2.1 Electron Paramagnetic Resonance

Historically, dynamic characterization in soft matter has been achieved in several ways. Most of these detect atoms of a particular isotopic type (e.g. hydrogen vs deuterium), and are difficult to apply to dilute systems due to their low sensitivity or high cost. This motivates the need for electron paramagnetic resonance (EPR). EPR, like NMR, measures the interaction of a spin with a magnetic field.

The fundamental principle of EPR spectroscopy is that in the presence of a magnetic field, B, the energy state of otherwise identical unpaired electrons depends upon their spin state. By absorbing photons, unpaired electrons may transition from a low-energy spin-state to a high-energy spin-state. This is called the Zeeman effect, and is described in its simplest form by the equation

$$h\nu = g_e \mu_B B \tag{2.1}$$

which equates the photon energy ($h\nu$, expressed in terms of Planck's constant, h, and photon frequency, ν) with the energy difference between the two states. This difference is written as $g_e \mu_B B$, where g_e is the electronic g-factor (a dimensionless

parameter describing the magnetic moment of the electron), μ_B is the Bohr magneton, a unit of magnetic moment, and B is the magnetic field strength.

In materials, unpaired electrons typically belong either to radical atoms, or to the d or f orbitals of metallic atoms. In order to make EPR measurements of a peptide or protein, scientists typically synthetically modify some of its constituent molecules to include a spin-label – a site-specific tag containing a stable free-radical electron. The structures are highly regular, meaning that the site-specificity of these probes implies site-specificity of EPR measurements. Using a dilute concentration of spin-labeled molecules, which we assume do not interact with each other, we measure the dynamics of the probes using a highly sensitive, bulk measurement.

EPR is customarily divided into continuous-wave (CW) measurements and pulsed, Fourier-Transform methods. In CW experiments, spectra are collected one frequency at a time, by varying either the photon energy or, more commonly, the magnetic field strength, B. By modulating the magnetic field strength sinusoidally, and collecting the first-derivative of the absorbance spectrum, rather than the absorbance spectrum itself, we drastically improve our ability to exclude noise, since noise is aperiodic. Thus, CW-EPR typically records the first derivative of absorbance. As I shall explain in Section 2.1.1, the shape of these distributions depends strongly on the dynamic behavior of the unpaired electron, and this can provide tremendous insight into the dynamic behavior of soft materials.

In contrast to this relatively straightforward method, the theoretical underpinnings of pulsed EPR experiments are fairly complex. Rather than collecting a spectrum one energy at a time, these methods make use of the Fourier transform to extract all points simultaneously. By exciting transitions using a single pulse (or series of pulses) of radiation, and by Fourier-Transforming the time-domain data into the frequency domain. By changing the number, shape or time-dependence of pulses, this approach can be used to induce different spin distributions in the sample, providing a greater degree of control than CW-EPR. Thus, pulsed EPR, unlike CW-EPR, is often used for structural determination, since experimenters have greater control over the nature of the information that they collect. Unfortunately, the time-dependence

of the signal makes dynamics measurements much more challenging than the CW case, since scientists are only now overcoming the limitations of existing theory. Thus in this work, while I make use of pulsed methods for structural analysis, I principally rely on CW-EPR methods.

The ability of EPR to make site-specific, quantitative measurements makes it invaluable for the study of soft matter. However, it suffers from two main drawbacks, which merit some discussion. First, since EPR measures unpaired electrons rather than atomic nuclei, the dynamic measurements made by EPR only relate indirectly to dynamic measurements of the molecule itself. More specifically, EPR measures the rotational diffusion (or rotational correlation times) associated with motion of the radical electron in a magnetic field. This means that EPR can only indirectly sample conformational fluctuations. However, studies have shown that the correlation times associated with electron diffusion in the slow-motion regime correspond directly to the dynamics of the spin-labeled molecule. This means that EPR measurements can still make meaningful dynamic statements, even though they don't directly quantify chain rotation. The second critique of EPR is that for most materials (i.e., for materials which do not contain free-radicals, excitons or transition metal atoms), EPR analysis requires chemically modifying target molecules. EPR spectroscopists sometimes call this problem, 'the price of peeking'.

Nonetheless, if we accept the approximation that spin-labeled molecules behave similarly to their unmodified counterparts, this technique can allow unprecedented access to the dynamic, chemical and structural properties of soft matter.

Nitroxide dynamics

The most common spin-labels are nitroxide radicals, which are prized for their high chemical stability and their multi-peak EPR spectra. In proteins, these labels are most commonly attached by chemical modification of cysteine residues with (1-Oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) Methanethiosulfonate (MTS) to produce an MTS spin-label (MTSSL).[60,61] In synthetic peptides, we can spin-label them at arbitrary positions by replacing an amino acid with 2,2,6,6-tetramethylpiperidine-*N*-oxyl-

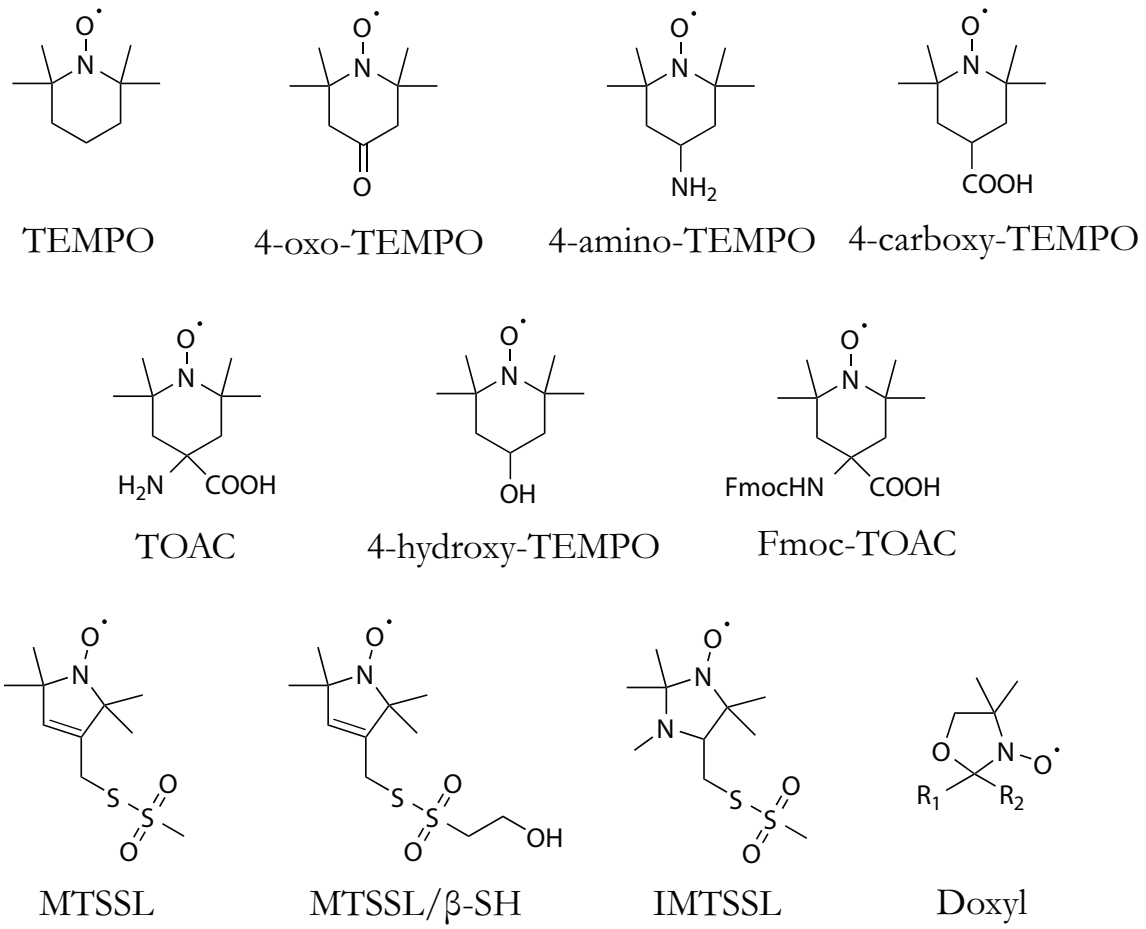


Figure 2.1: Common nitroxide probes used to study soft materials. Nitroxide radicals are highly stable in solution, and can be synthetically attached to materials as diverse as proteins, nucleotide sequences, lipids, and fatty acids.

4-amino-4- carboxylic acid (TOAC).[62,63] In addition to arbitrary placement, another advantage of TOAC is that it integrates more rigidly into the chain's backbone, resulting in more directly meaningful data. For the structures of these and several other common spin-labels, refer to Figure 2.1.

A unique advantage of using nitroxide radicals is that, unlike most organic radicals, the nitroxide radical has several well-defined peaks which can be used for fitting. This is because, in addition to the electron Zeemann (EZ) interaction that corresponds to the spin quantum number (m_s), the radical electron can transition between p orbitals of the nearby oxygen atom, and thus experiences a nuclear Zeeman (NZ) interaction corresponding to the (m_l) quantum number. The EZ interaction is described by the

Hamiltonian (H_{EZ})

$$H_{EZ} = \frac{\mu_B}{\hbar}\mathbf{B}^T\mathbf{g}\mathbf{S} \tag{2.2}$$

where $\mathbf{B}$ and $\mathbf{g}$ are tensorial forms of the magnetic field and g values found in Equation 2.1, μ_B is again the Bohr Magneton, $\hbar$ is the reduced Planck constant, and $\mathbf{S}$ is the spin operator. We use the superscript ^{T} to denote the transposition of a matrix. Similarly, the NZ interaction is described by its Hamiltonian:

$$H_{NZ} = -\frac{\mu_N}{\hbar}\sum_k g_{n,l}\mathbf{B}^T\mathbf{I}_k \tag{2.3}$$

where k denotes the k^{th} nucleus and $\mathbf{I}_k$ is the spin operator of that nucleus. The term μ_N denotes the nuclear magneton – a constant analogous to μ_B. In nitroxides, only the oxygen atom plays a significant role in splitting energy levels.

In nitroxide radicals, these terms give rise to 6 energy levels, which are further modulated by electron-nuclear (hyperfine, HF) interactions, given by the equation

$$\mathbf{H_{HF}} = \sum_k \mathbf{S}^T\mathbf{A}_k\mathbf{I}_k \tag{2.4}$$

where $\mathbf{A_k}$, the hyperfine coupling tensor, provides important information about the magnetic environment of the spin. (In general, only the nearest atom contributes significantly to this term, which can be condensed to $\mathbf{H_{HF}} = \mathbf{S}^T\mathbf{A}\mathbf{I}$). The role of these 3 interactions is summarized in Figure 2.2. Since electronic transitions must conserve momentum, only three energetic transitions are possible, resulting in a three-peak spectrum.

Nuclear quadrupole (NQ) interactions can cause small resonance shifts, but beyond that do not contribute majorly to EPR signals and are not important to this thesis. Zero-field splitting can contribute to the Hamiltonian, but only in cases where a paramagnetic species has $S > 1/2$, whereas for nitroxide radicals, $S = 1/2$.

Finally, electron-electron (EE) interactions, can be very important to EPR analysis under conditions where nitroxide radicals are placed in close proximity to one another. This may occur accidentally, when experimental conditions impose very

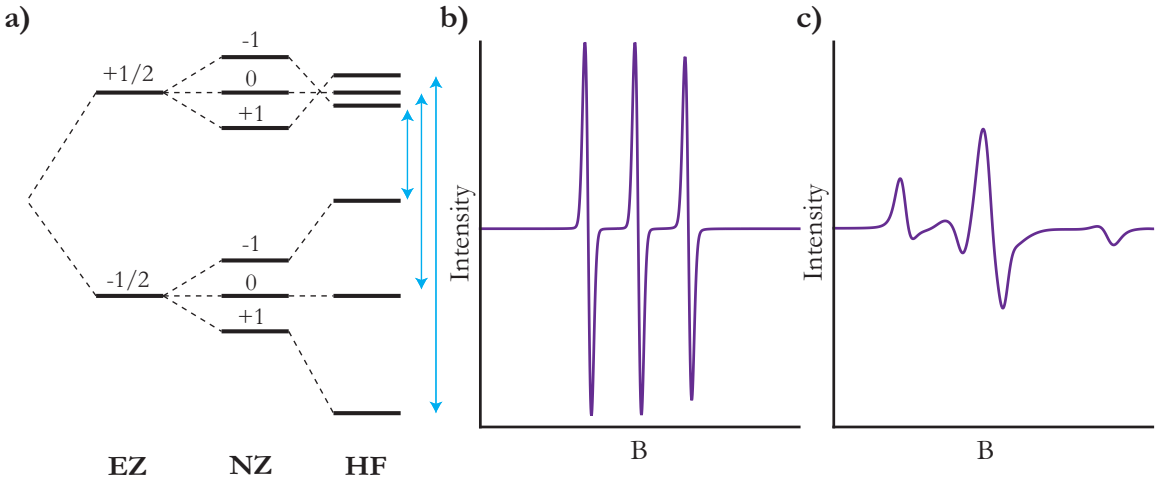


Figure 2.2: Energy levels in nitroxide radicals. a) The electron's energy is split by the electron Zeeman (EZ) and the nuclear Zeeman (NZ) interaction, giving rise to 6 energy levels. These are modulated by the hyperfine (HF) interaction, as well as several others that minorly change energy levels. Due to momentum conservation, only 3 energetic transitions occur, giving rise to the 3 peaks shown in (b). b-c) Simulated EPR spectra for nitroxide radicals near the isotropic limit (b) and the rigid limit (c). Each spectrum contains 3 peaks, though these are more easily distinguished in the isotropic limit, where peaks are narrower and do not overlap.

high concentrations of spin-labels, resulting in spectral broadening. It may also be exploited, by intentionally introducing multiple probes into the same molecule in order to deduce the distances between them. The EE Hamiltonian is given by:

$$\mathbf{H_{EE}} = \mathbf{S}_1^T \mathbf{J} \mathbf{S}_2 \tag{2.5}$$

J describes the total interaction between the two spins. By specifically studying the dipole-dipole (DD) interaction between these two probes and simplifying, we can express this Hamiltonian ($\mathbf{H_{DD}}$) in terms of the distance vector ($\mathbf{r}_{12}$) between the two spins:

$$\mathbf{H_{DD}} = \mathbf{S}_1^T \mathbf{J} \mathbf{S}_2 \tag{2.6}$$

$$= \frac{\mu_0 \mu_B^2 g_1 g_2}{4\pi\hbar} \frac{1}{r_{12}^3} \left[\mathbf{S}_1^T \mathbf{S}_2 - \frac{3}{r_{12}^2} (\mathbf{S}_1^T \mathbf{r}_{12})(\mathbf{S}_2^T \mathbf{r}_{12}) \right] \tag{2.7}$$

Because the energy varies approximately with probe-probe according to r_{12}^{-3}, experimentalists can exploit this relationship to determine the distribution of r_{12} using pulsed EPR methods like DEER or DQC. This gives data analogous to data collected from Förster resonance energy transfer (FRET) experiments, but with physically smaller probes.

2.1.1 Continuous-wave electron paramagnetic resonance theory

Today, one of the primary uses of CW-EPR is the study of dynamic behavior in spin-labeled samples. The principle is straight-forward: the shape of an observed EPR spectrum depends on the average rate of rotational diffusion of an observed probe, as well as several energetic parameters. EPR is sensitive to diffusion over a broad range of correlation times ($t_R = 10^{-12}$ s in the fast motion limit to $t_R = 10^{-6}$ s in the slow motion limit)[64], making it a very robust measurement tool. The range itself can be further subdivided, based on the type of model appropriate for spectral fitting. Dynamics for virtually all spin-labeled macromolecules fall within the slow-motional regime (with correlation times ranging from $\tau_R = 10^{-6}$ s to $\tau_R = 10^{-9}$ s), which requires the most challenging theoretical treatment.

In the slow-motion regime, which corresponds to dynamic fluctuations in chain position, continuous wave electron paramagnetic resonance (CW-EPR) spectra follow the stochastic Liouville equation:

$$I(\omega - \omega_0) = \frac{1}{\pi}\left\langle\left\langle v\middle|[(\tilde{\mathbf{\Gamma}} - i\tilde{\mathbf{H}} + i(\omega - \omega_0)I]^{-1}\middle|v\right\rangle\right\rangle \tag{2.8}$$

In this equation, I represents the spectral intensity as a function of angular microwave frequency (ω) relative to some reference frequency (ω_0), represents a starting vector that already contains the spin operator and the statistical distribution of up/down spins, $\tilde{\mathbf{\Gamma}}$ represents the symmetrized diffusion superoperator, $\tilde{\mathbf{H}}$ represents the Liouville superoperator, and i represents the imaginary unit. Given a set of parameters $\mathbf{c} = \{c_1, c_2, \cdots, c_n\}$ describing our experiment, we may solve this equation

numerically with high computational efficiency. Freed derived and implemented this approach in 1976, and his original solution remains central to most CW-EPR analysis tools.[64–68]

The precise number and definition of the parameters studied depends on the degree of detail that the user needs, but typically they include the hyperfine tensor ($\mathbf{A}$), the tensor of electron g values ($\mathbf{g}$), the rotational diffusion tensor ($\mathbf{D_R}$, which is generally expressed in terms of a tensor of rotational correlation times, $\tau_{\mathbf{R}}$, or in terms of the base 10 logarithm of tensor components, $\bar{\mathbf{R}}$). Since each of these tensors is real-symmetric, they may be expressed in terms of 3 axial components. For instance:

$$\mathbf{A} = \begin{bmatrix} A_{xx} & 0 & 0 \\ 0 & A_{yy} & 0 \\ 0 & 0 & A_{zz} \end{bmatrix} \tag{2.9}$$

Similarly, $\mathbf{g}$ may be resolved into g_{xx}, g_{yy} and g_{zz}, and $\bar{\mathbf{R}}$ may be resolved into R_{xx}, R_{yy} and R_{zz}. Because the model involves relativistic rotations between frames, a director angle (ψ), which describes spin orientation, and a set of Euler angles (α,β and γ) are also involved. Additional line-broadening parameters are also frequently included.

One assumption of this classical model is that spins are uniformly distributed in the director frame. When this assumption fails, as often happens in real systems, more complex models are needed. For these cases, researchers typically turn to two more-detailed models: the macroscopic order, microscopic disorder (MOMD) model, where spectra are generated for a variety of orientations and then composed into a single spectrum; and the slowly relaxing local structure (SRLS) model, which generalizes this idea to allow time-dependent variation in the area of locally-ordered region. Because SRLS requires higher-frequency measurements, MOMD calculations are more common.

These models make the CW-EPR spectra of radicals in the slow-motional regime computable. However, in order to truly make use of these insights, we need to reverse-engineer them, using non-linear fitting, a process where the deviation between exper-

imental data and a model is minimized as a function of important fit parameters. (For further details, refer to Section 2.1.3.) In practice, this only works well when varying a small number of parameters. Generally, if we tried to optimize over all 15+ of the parameters mentioned in this section, we could expect the fitting process to become difficult and unreliable – either because local minima of the error function would become too common, or because we would identify so many high-quality fits that it would be impossible to select the physically accurate one.

For these reasons, we make several assumptions that reduce the size of our parameter space. First, based on the experimental setup, we note that most of the angles are known so we can use trusted default values. Second, we note that the **A** and **g** tensors are independent from the diffusion rate, and can be established by freezing a sample and fitting its rigid-limit spectrum (which has the benefit of being far less computationally intensive than fits in the slow-motional regime). If further simplification is needed, we can reduce the complexity of **A** and **g** by assuming that $g_{xx} = g_{yy} = g_{\perp}$ (in this notation $g_{zz} = g_{\parallel}$) and $A_{xx} = A_{yy} = A_{\perp}$. Finally, we can (and typically do) assume that rotational diffusion occurs much more rapidly around one axis than any of the others, allowing us to rely on a single diffusion parameter $R = \log(D_R)$.[69] In practice, these assumptions allow us to reliably fit spectra using as few as 3-4 fit parameters – a much more manageable optimization problem.

Today, most researchers use Freed's NLSL package or Stoll's EasySpin package for this process.[66,67] NLSL contains a high-efficiency implementation of the Levenberg-Marquardt minimization algorithm, but uses a Fortran 77 command-line utility rather than a callable function. This makes it difficult to script. EasySpin contains a MatLab wrapper for NLSL. However, it ignores higher order contributions to spectral broadening, and it does not allow users to vary every the parameter accessible in NLSL. Both EasySpin and NLSL operate by performing nonlinear fitting of experimental spectra. The details of this procedure are described Section 2.1.3.

2.1.2 Pulsed dipolar spectroscopy

Given a molecule containing two spin-labels, pulsed EPR measurements allow us to extract the distance distribution between them (effectively, a pair-distribution function). This enables scientists to make direct distance measurements in spin-labeled molecules. When a pair of spins are close together (and thus have a strong dipolar coupling) the dipole-dipole interaction may affect a CW-EPR spectrum – however, over the longer distances (> 1.5 nm) that typically interest spectroscopists, this component is week, and can generally only be detected using Fourier-transform methods.[70–73] By using this family of techniques, generally called pulsed dipolar spectroscopy (PDS) techniques, it is possible to gain insight into the structure of proteins and other macromolecules, in much the same way that fluorescence spectroscopists use FRET experiments.

As explained in Section 2.1, pairs of spins experience an electron dipole-dipole coupling that strengthens as the distance between them shrinks. By exciting the sample with a controlled sequence of microwave pulses designed to suppress the hyperfine interaction, we become particularly sensitive to this energetic contribution, allowing us to determine the inter-spin distance of a pair of spins (see Equation 2.7). The most common technique for this is double electron-electron resonance (DEER), also known as pulsed electron-electron double resonance (PELDOR). In this experiment, we typically use 4 pulses, applied in a particular time sequence, in order to determine interprobe distances. The method works best for distances between 1.5 and 8 nm. In this work, I also employed double quantum coherence (DQC) experiments, which work similarly, but are sensitive to smaller inter-probe distance distributions.[74,75] These distributions are then fit using Tikhonov regularization, an optimization algorithm that imposes reasonable constraints (i.e. a degree of smoothness) on the shape of the inter-probe distribution.

2.1.3 Nonlinear Fitting of EPR Spectra

This approach aims to minimize the deviation of a model spectrum $I(\omega-\omega_0, \mathbf{c})$, from an experimental spectrum, $I_{exp}(\omega-\omega_0)$, as a function of the experimental parameters $\mathbf{c}$. We represent this by the χ^2 function:

$$\chi^2(\mathbf{c}) = \sum_{i=1}^{n} \frac{[I(\omega_i - \omega_0, \mathbf{c}) - I_{exp}(\omega_i - \omega_0)]^2}{\sigma_i^2} \tag{2.10}$$

In this equation, σ_i^2 represents the individual uncertainty in $I_{exp}(\omega_i - \omega_0)$ at each point i. In general, χ^2 is non-convex, meaning that it may possess many disconnected local minima. In such cases, we must rely on numerical methods to identify the optimal set of parameters, $\mathbf{c}^*$, that describe the spectrum. Below, I provide an overview of common methods for fitting.

Grid-Search Optimization

In this approach, we overlay a grid of points upon the $\mathbf{c}$ parameter space and choose the point producing the best χ^2 value. This method exists in both NLSL and EasySpin, but is very inefficient.

Monte Carlo Optimization

This approach resembles grid-search optimization, except that we choose points randomly from within the parameter space. In low-dimensional parameter spaces, this is marginally efficient than grid-search; however, it becomes more efficient when the dimension of $\mathbf{c}$ increases. This method is only implemented in EasySpin.

Levenberg-Marquardt Optimization

This algorithm uses a gradient-descent method designed specifically for curve-fitting problems. In this process, users provide an initial guess value, $\mathbf{c_0}$, and the algorithm iteratively steps along the negative gradient of χ^2 until reaching a local minimum. Adaptations exist to enable bounded optimization, but typically this algorithm is used in an unbounded fashion. The algorithm typically obtains linear convergence

(which means that the difference between χ^2 at the n^{th} iteration and χ^2 at the optimum, $\chi^2(\mathbf{c_n}) - \chi^2(\mathbf{c}^*)$, decreases at a rate proportional to e^{-n}. It is well-suited to most problems where χ^2 is convex. However, when χ^2 is not a convex function, the algorithm requires a lucky initial guess to reach the global minimum. In the case of typical EPR spectra, random initial guesses rarely reach the global minimum of χ^2. Both NLSL and EasySpin enable Levenberg-Marquardt optimization.

By establishing reasonable physical bounds, we can adapt Levenberg-Marquardt optimization for global optimization by employing a Monte Carlo or grid-search approach to selecting $\mathbf{c_0}$ values. By selecting a randomized sequence of initial guesses and starting the Levenberg-Marquardt algorithm from each of these, we become likely to identify the global optimum within the bounds. A grid-search approach to picking $\mathbf{c_0}$ will also work. Neither NLSL nor EasySpin implements these variants, despite their comparatively rapid convergence rates in CW-EPR spectral analysis.

Particle-Swarm Optimization

In particle swarm optimization, a selection of 'particles' receive initial positions and random velocities within the problem space. Then, each particle moves according to its velocity. The position of each particle updates to the best position it has found, and the particle receives a new, randomized velocity. This process iterates until reaching a convergence criterion, at which point the algorithm has reached the global minimum. The method makes no assumptions of differentiability and does not require gradient calculations, making it useful in cases where χ^2 is highly multimodal, as is often the case in EPR spectral fitting. Particle swarm optimization tends to be quite robust, generally identifying the optimum of χ^2, even in parameter spaces of fairly high dimension. For EPR fitting, particle swarm optimization is only implemented in EasySpin.

Simulated Annealing

Like particle swarm optimization, simulated annealing provides robust global fitting. In a basic iteration, the program probabilistically chooses whether to remain in its

current state or to move to a nearby state, based on the value of χ^2 in both states. Changes that reduce χ^2 are always accepted, and changes that increase χ^2 are accepted with a probability that depends upon the 'temperature' of the system. Over time, the 'temperature' of the system is slowly reduced. In this way, the algorithm explores a large fraction of parameter space before cooling to a minimum value. This helps to ensure global convergence. Neither NLSL nor EasySpin currently employs this optimization protocol, despite its popularity in the broader curve-fitting community.

2.2 Flow synthesis of peptides

In general, I synthesized the peptides presented in this thesis using the flow-synthesis methods developed in the Pentelute group.[76] The advantages of this approach over conventional solid state peptide synthesis include high speed (each coupling takes approximately 1 minute) and high synthetic yield. Briefly, a sample of H-rink amide resin is prepared in a syringe and swelled in *N*-*N* dimethylformamide (DMF). First, the resin is washed with DMF at the desired temperature generally 90 °C. Then, in turn it is: *1)* washed with a solution containing an Fluorenylmethyloxycarbonyl- (Fmoc-) protected amino acid and activating agents; *2)* washed with more DMF; *3)* washed with a solution of DMF-20% piperidine to remove the Fmoc, leaving an exposed amine; and *4)* washed again with DMF. The result is a resin that is covalently bound to the C-terminus of an amino acid. With the molecule's N-terminus free, steps *1-4* can now be run again, with a new amino acid. Downstream, an absorbance detector provides information about the quality of the synthesis. A schematic of this process, and of the expected absorbance profile, is shown in Figure 2.3.

By iteration, sequences of high length and purity can be synthesized, provided that the per-step yield remains very high. To spin-label our sequences, we incorporated TOAC (Fig. 2.1) directly into their backbone, substituting it for an amino acid of choice. While theoretically TOAC can be attached like any other amino acid (Fmoc-TOAC is commercially available) the molecule's rigid and bulky structure makes this

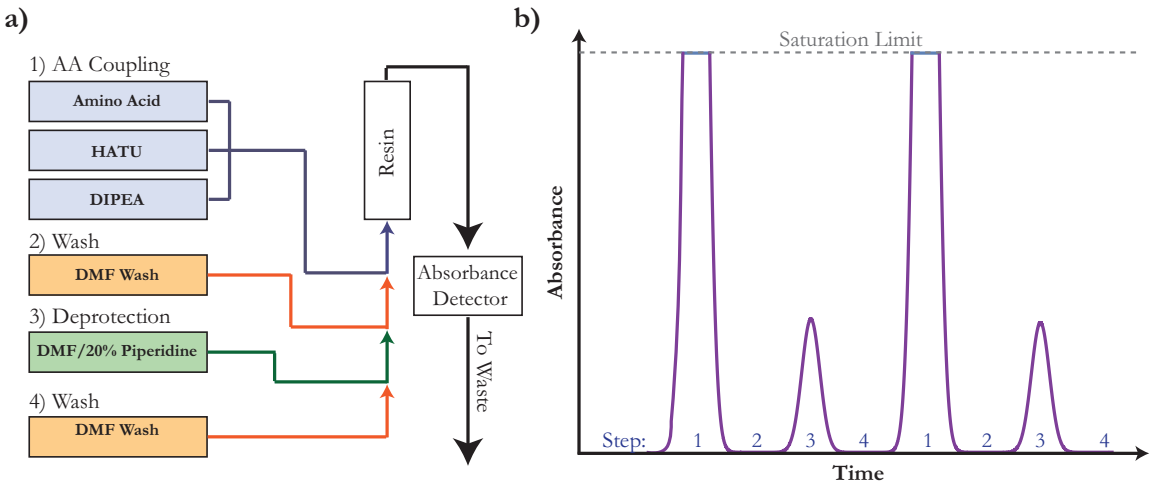


Figure 2.3: Peptide flow synthesis schematic. a) The four flow-steps that are iterated to attach each consecutive amino acid to the resin. b) A diagrammed absorbance profile of the attachment of 4 amino acids. Fmoc is UV-active, so in Step 1, the high concentration of reactants overwhelms the detector. In Step 3, the profile of Fmoc deprotection products is observed. In the wash step, signal vanishes because DMF is not active at this wavelength.

process challenging, a difficulty I overcame in Chapter 3.

2.3 Liquid Chromatography Techniques

Molecules including those synthesized according to Section 2.2 typically require characterization and purification before use. In general, peptide-based samples in this paper were characterized by reverse-phase liquid chromatography mass spectrometry (LC-MS) in order to determine their purity. In this technique, samples are suspended in a mixture of water, acetonitrile and a dilute additive (typically formic acid for LC-MS) to control pH and promote solubility. Then, a small quantity is loaded onto a column – typically a C3 or C18 column – and a water:acetonitrile gradient is used to separate distinct chemical components by their polarity. Upon reaching the end of the column, the solution is then injected into a mass-spectrometer, where the charge:mass ratio of the solute can be studied to determine its molecular weight. I used LC-MS to identify compounds of interest for subsequent purification, to verify the quality of purified products, and to quantify reaction yields.

To purify samples, I used high performance liquid chromatography (HPLC), which works very similarly. A solution of water, acetonitrile and dilute trifluoroacetic acid was used to dissolve a crude product, which was then loaded onto a preparatory HPLC column and separated by a water:acetonitrile gradient. Using UV and mass-spectrometric intensity of the eluted product, I could separate the important compounds from impurities by identifying and isolating the product-peak.

2.4 Flow cytometry

In this technique, specially prepared cell cultures are loaded onto a fluorescence detector. Fluorescently-labeled macromolecules are flowed over the cultures, so that adhesion can be detected by a fluorescence measurement. By varying the concentration of the macromolecule, we can establish the binding constant of the macromolecule. Similarly, by measuring fluorescence between a fixed concentration of the macromolecule and a varying concentration of competing analogues, we can compare the viability of those analogues quantitatively, using the half-maximal inhibitory concentration (IC50) – the competitor concentration required to inhibit 50% of the binding of the fluorescent molecule. Thus, compounds which more effectively bind to cells will have a lower IC50 than compounds which bind less effectively, since it will take a lower concentration to displace the fluorophor.

2.5 Molecular dynamics

In this work, I occasionally studied short molecular dynamics simulations of peptides in order to gain insight into their secondary structure. In order to prepare initial structures, I used the Pep-Fold 3 algorithm to predict an energetically favorable starting conformation, and ran brief simulations using the CHARMM36 force field in Gromacs. Using Gromacs algorithms, I computed hydrogen bonding maps, secondary structure maps, distance distributions, and assorted other data for comparison with EPR results.

Chapter 3

Quantifying residue-specific conformational dynamics of a highly reactive 29-mer peptide

This chapter was adapted from the publication "Quantifying residue-specific conformational dynamics of a highly reactive 29-mer peptide", originally published in Scientific Reports.[77]

3.1 Abstract

Understanding structural transitions within macromolecules remains an important challenge in biochemistry, with important implications for drug development and medicine. Insight into molecular behavior often requires residue-specific dynamics measurement at micromolar concentrations. We studied MP01-Gen4, a library peptide selected to rapidly undergo bioconjugation, by using electron paramagnetic resonance (EPR) to measure conformational dynamics. We mapped the dynamics of MP01-Gen4 with residue-specificity and identified the regions involved in a structural transformation related to the conjugation reaction. Upon reaction, the conformational dynamics of residues near the termini slow significantly more than central residues, indicating that the reaction induces a structural transition far from the reaction site.

Arrhenius analysis demonstrates a nearly threefold decrease in the activation energy of conformational diffusion upon reaction (8.0 k_BT to 3.4 k_BT), which occurs across the entire peptide, independently of residue position. This novel approach to EPR spectral analysis provides insight into the positional extent of disorder and the nature of the energy landscape of a highly reactive, intrinsically disordered library peptide before and after conjugation.

3.2 Introduction

Combinatorial, library-based strategies for discovering functional peptides have transformed biochemistry, enabling tremendous improvements in enzyme design, disease diagnosis, and drug development.[78] One prototypical example of sequences identified by combinatorial discovery is the family of MP peptides – molecules selected to undergo nucleophilic aromatic substitution (S_NAr) reactions via a single cysteine residue (Fig. 3.1).[79–81] Their mild reaction conditions make reactive MP peptides optimal for bioconjugation[82–84] while preserving orthogonality to other popular conjugation methods including click chemistry[85–87], protein-facilitated approaches (such as the biotin-streptavidin interaction)[88–92], and the use of peptide tags.[93–96] Bioconjugation tools have become essential technology, enabling controlled coupling of biomolecules for important diagnostic and therapeutic purposes. In the case of MPs, many features of their backbone dynamics and conformational behavior remain unknown because the residue-specific measurements required are difficult to achieve at low (μM) concentrations.[79–81]

Here we investigate MP01-Gen4, an abiotic 29-mer selected from among 5 x 1013 randomized peptides and subsequently optimized via experimental and computational methods.[79,81] The resulting sequence reacts rapidly with perfluoroarenes, demonstrating quantitative conversion in under five minutes (Fig. 1). Previously reported circular dichroism (CD) measurements show experimentally that MP01-Gen4 undergoes a random-coil-to-helix structural transformation upon interaction with a perfluoroarene probe.[81,97] Calculations from the PrDOS intrinsic disorder prediction tool suggested

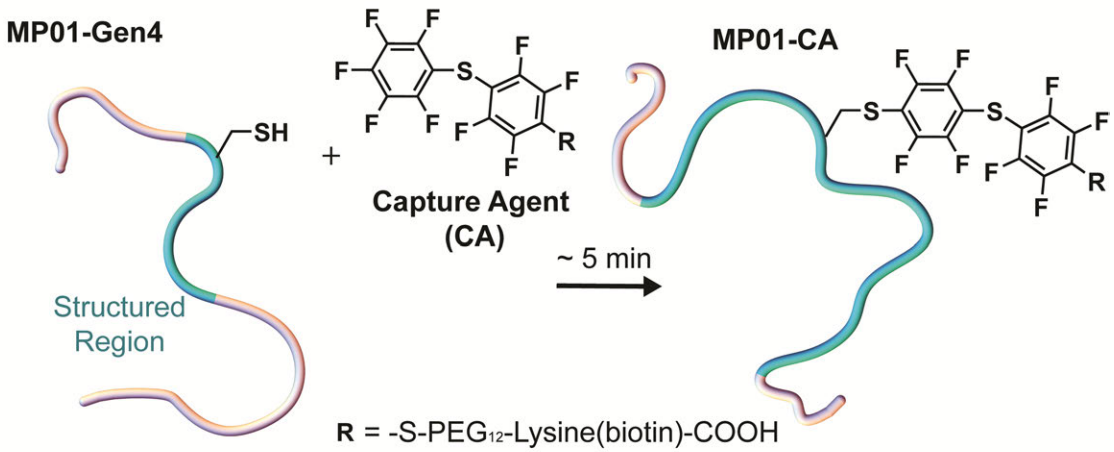


Figure 3.1: The MP01-Gen4 peptide reacts rapidly with a perfluoroarene capture agent (CA) via nucleophilic aromatic substitution (S_NAr) to form the complex MP01-CA in approx. 5 mins. The peptide backbone's dynamic structure (illustrated here as a cartoon) is related to the high reactivity of MP01-Gen4 with perfluoroarenes.

disorder in residues 1-7 and 24-29, and predictions using Rosetta software suggest the existence of transient α-helix-like order in residues near the center of the peptide, prior to S_NAr reaction.[81] Circular dichroism studies demonstrate that the peptide increases in helical content upon reaction, but neither these experiments nor PrDOS/Rosetta predictions could identify the residues involved. Thus, experimentally identifying the residues involved in this transition, and understanding the extent to which distinct regions of the sequence exhibit disorder or flexibility, is important for understanding the behavior of MP01-Gen4. Although this type of structural transition is common among natural sequences, its emergence from a library of abiotic peptides in the context of a non-biological reaction is noteworthy.[47–49] We aimed to identify the residues involved in this structural transition and to understand the relationship between the dynamic behavior of MP01-Gen4 and its structural transition.[81]

Conformational studies of peptides typically require residue-specific insight into dynamics. We acquired this information by introducing radical electron spin-labels at specific residues of MP01-Gen4 and performing electron paramagnetic resonance (EPR) spectroscopy to obtain rotational diffusion coefficients (inversely related to

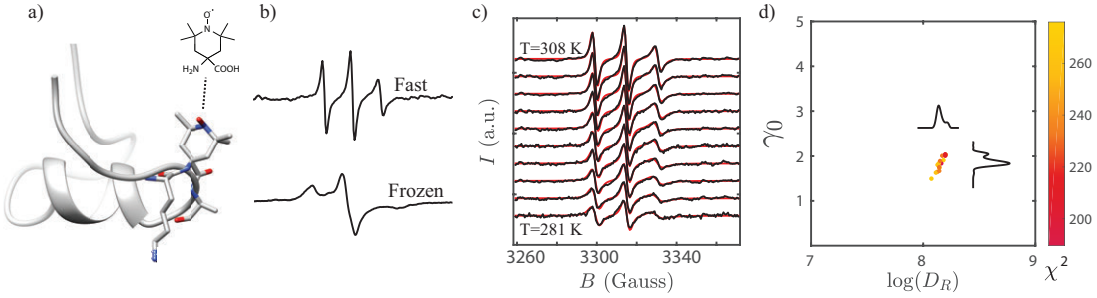


Figure 3.2: Experimental approach to dynamics measurements. a) TOAC is embedded along the MP01-Gen4 backbone. b) EPR line shapes of a TOAC-labeled MP01-Gen4 peptide at 308 K (top) and 150 K (bottom) indicate fast and slow rate of motion, respectively. c) EPR spectra are fit for rotational diffusion rate at different temperatures. d) The fitting function (χ^2) represents deviation between experimental data and a fitting model. Optimal values for fit parameters such as the log of the rotational diffusion coefficient ($\log(D_R)$) and the Gaussian line-broadening (γ_0) are extracted from clusters of good fits near the global minimum of χ^2.

rotational correlation times) of the spin-label's motion.[65–67,98] This motion primarily originates from conformational changes of the backbone, and its timescale depends on position, since more flexible regions of a peptide change conformation more rapidly.[99] We can therefore use this approach to map the flexibility of a sequence with residue-level resolution, even at micromolar concentrations.[3,58,68,100–104]

3.3 Experimental Methods

The basic methodology of our experiments is outlined in Fig. 3.2. In brief, we used TOAC (TOAC = 2,2,6,6-tetramethylpiperidine-*N*-oxyl-4-(9-fluorenyl methyloxy carbonyl-amino)-4-carboxylic acid) to spin-label each peptide (Fig. 3.2a) and measured its EPR spectrum. The line shapes of the spectra encode dynamics information (Fig. 3.2b). We fit our measurements at each probe position at ten temperatures, ranging from 280 K to 308 K (Fig. 3.2c), and measured distributions of good fits in order to quantify uncertainty (Fig. 3.2d). This strategy enabled rotational diffusion rate, D_R, measurements at each site and temperature.

Ten MP01 variants were selected, with TOAC positions chosen to provide approximately regular spacing, by a systematic scan of alanine substitutions, which we used

Table 3.1: Names and sequences of the MP01-Gen4 variants employed in this study. J is the amino acid spin-label TOAC; C is the reactive cysteine.

Designation	Sequence
MP01-Gen4	MNQKYKMAKACFFAFLEHLKKRKLYPMSG
MP01-J3	MNJKYKMAKACFFAFLEHLKKRKLYPMSG
MP01-J5	MNQKJKMAKACFFAFLEHLKKRKLYPMSG
MP01-J7	MNQKYKJAKACFFAFLEHLKKRKLYPMSG
MP01-J13	MNQKYKMAKACFJAFLEHLKKRKLYPMSG
MP01-J16	MNQKYKMAKACFFAFJEHLKKRKLYPMSG
MP01-J18	MNQKYKMAKACFFAFLEJLKKRKLYPMSG
MP01-J20	MNQKYKMAKACFFAFLEHLJKRKLYPMSG
MP01-J23	MNQKYKMAKACFFAFLEHLKKRJLYPMSG
MP01-J27	MNQKYKMAKACFFAFLEHLKKRKLYPJSG
MP01-J29	MNQKYKMAKACFFAFLEHLKKRKLYPMSJ

to identify locations where modifications would minimally perturb the reactivity of the peptide. In a two positions (5 and 7) we were willing to replace residues known to be important for reactivity, on the basis that we didn't want to replace nearby charged residues. Sequences and designations are reported in Table 3.1.

3.3.1 Materials

1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), Fmoc-L-Ala-OH, Fmoc-L-Cys (trt)-OH, Fmoc-L-Glu (tBu)-OH, Fmoc-L-Phe-OH, Fmoc-Gly-OH, Fmoc-L-His (Boc)-OH Fmoc-L-Lys (Boc)-OH, Fmoc-L-Leu-OH, Fmoc-L-Met-OH, Fmoc-L-Asn (Trt)-OH, Fmoc-L-Pro-OH, Fmoc-L-Gln (Trt)-OH, Fmoc-L-Arg (Pbf)-OH, Fmoc-L-Ser (tBu)-OH, Fmoc-L-Tyr (tBu)-OH and Fmoc-TOAC-OH were purchased from Chem-Impex International. H-rink-amide ChemMatrix Hyr resin was obtained from PCAS BioMatrix, Inc. (7-Azabenzotriazol-1-yloxy)tripyrrolidino phosphonium hexafluorophosphate (PyAOP) was purchased from P3 BioSystems. N,N-dimethylformamide (DMF), acetonitrile (ACN) and diethyl ether were purchased from VWR (Radnor, PA). N,N-diisopropylethylamine (DIPEA), formic acid (FA), 10x phosphate buffered saline (PBS), trifluoroacetic acid (TFA) and triisopropylsilane (TIPS) were obtained from Sigma-Aldrich. Potassium hexaferrocyanate (III) ($K_3Fe(CN)_6$) was purchased from Alfa-Aesar.

Alltech low pressure polytetrafluoroethane (PTFE) tubing and Leica BioSystems Crytoseal capillary tube sealant were purchased from Fisher-Scientific. Liquid nitrogen was purchased from Airgas. Wilmad 4x250 mm quartz glass EPR tubes were purchased from Cambridge Isotope Laboratories. Water (18.2 MΩ) was purified using a Milli-Q Direct 8 system.

3.3.2 TOAC Peptide Synthesis

Peptides and the perfluoroarene capture agent (CA) were synthesized according to literature, using ChemMatrix H-rink amide resin (0.49 meq/g) on the 0.1 mmol scale.[76,79] Flow-synthesis of standard amino acids uses a DMF solution of 0.2 M amino acid, 0.17 M activating agent and 5% (v/v) DIPEA flushed over the sample at 80 mL/min for 15 s, followed by DMF washing and deprotection using DMF 20% piperidine. TOAC was coupled using 50 mM Fmoc-TOAC, 47.5 mM HATU, and 10% DIPEA at a rate of 40 mL/min for 15 s, followed by the usual washing and Fmoc-deprotection steps. Due to steric limitations of the TOAC, the kinetics of coupling natural amino acids to resin-bound TOAC proved to be exceptionally slow. To bypass this problem, we couple the sterically-hindered post-TOAC residue using 0.2 M amino acid, 0.14 M activating agent and 10% DIPEA pumped at 10 mL/min for 10 min, followed by the usual wash and deprotection steps. All subsequent residues were coupled normally. Completed peptides were cleaved for 2 h at RT using (90% TFA, 5% water, 5% TIPS v/v), a cleavage cocktail for TOAC peptides.40 The resulting peptides were then precipitated and washed 3x in diethyl ether (-80 °C), before drying under vacuum. The dried product was dissolved and purified by reverse phase high performance liquid chromatography (HPLC). Synthetic yields for each peptide, calculated from the crude mass collected, are reported in Table 3.2.

3.3.3 LC-MS Analysis

The purity of all peptides was analyzed by liquid chromatography mass spectrometry (LC-MS) using an Agilent 6520 ESI-Q-TOF mass spectrometer. For convenience,

Table 3.2: Crude yields and LC-MS data for MP01-Gen4 variants. Crude yields (mg) are reported for each peptide (column 2), along with masses expected/*observed* for the peaks present in the LC-MS traces in Figures 3.3 and 3.4. In most cases, hydroxylamine signal is dominant, due to reduction of the nitroxide in acidic conditions before/during the LC-MS scan. A minor peak, corresponding to the nitroxyl version of the C-terminal glycine deletion product, appeared in a few spectra – however, this is always a minority product and probably had a negligible impact on EPR analysis. Observed masses are calculated using the $[M+3H]^{3+}$ charge state for the unlabeled peptides, and the $[M+4H]^{4+}$ charge state for the labeled peptides. MS data are reported in Appendix B.

Name	**Crude Yield (mg)**	**Peak**	**Exp./*Obs.* MW**
MP01-J3	131 (69%)	Hydroxylamine	3576.89/*3576.87**
		Nitroxide	–
		Hydroxylamine (Labeled)	4891.35/*4891.35**
		Nitroxide (Labeled)	–
		Glycine Deletion	–
MP01-J5	50 (26%)	Hydroxylamine	3541.88/*3541.88**
		Nitroxide	3540.88/*3540.88*
		Hydroxylamine (Labeled)	4856.35/*4856.35**
		Nitroxide (Labeled)	4855.34/*4855.34***
		Glycine Deletion	–
MP01-J7	64 (33%)	Hydroxylamine	3573.91/*3573.92**
		Nitroxide	3572.90/*3572.90*
		Hydroxylamine (Labeled)	4888.37/*4888.34**
		Nitroxide (Labeled)	4887.36/*4887.39***
		Glycine Deletion	4830.34/*4830.41*
MP01-J13	110 (59%)	Hydroxylamine	3557.88/*3557.91**
		Nitroxide	3556.87/*3556.87*
		Hydroxylamine (Labeled)	4872.34/*4872.36**
		Nitroxide (Labeled)	–
		Glycine Deletion	4814.31/*4814.29*
MP01-J16	82 (41%)	Hydroxylamine	3591.86/*3591.86**
		Nitroxide	3590.85/*3590.85*
		Hydroxylamine (Labeled)	–
		Nitroxide (Labeled)	4905.32/*4905.35*†*
		Glycine Deletion	–
MP01-J18	56 (30%)	Hydroxylamine	3567.89/*3567.88**

Continued on next page

Table 3.2 – *Continued from previous page*

Name	Crude Yield (mg)	Peak	Exp./*Obs.* MW
		Nitroxide	–
		Hydroxylamine (Labeled)	4882.35/*4882.31**
		Nitroxide (Labeled)	4881.34/*4881.30*
		Glycine Deletion	–
MP01-J20	81 (42%)	Hydroxylamine	3576.85/*3576.85**
		Nitroxide	3575.84/*3575.84*
		Hydroxylamine (Labeled)	4891.32/*4891.18**
		Nitroxide (Labeled)	4890.31/*4890.17*
		Glycine Deletion	–
MP01-J23	62 (32%)	Hydroxylamine	3576.85/*576.83**
		Nitroxide	–
		Hydroxylamine (Labeled)	4891.32/*4891.22**
		Nitroxide (Labeled)	4890.31/*4890.21*
		Glycine Deletion	–
MP01-J27	100 (53%)	Hydroxylamine	3573.91/*3576.92**
		Nitroxide	–
		Hydroxylamine (Labeled)	4888.37/*4888.31**
		Nitroxide (Labeled)	–
		Glycine Deletion	4830.34/*4830.22*
MP01-J29	20 (10%)	Hydroxylamine	3647.92/*3647.94**
		Nitroxide	–
		Hydroxylamine (Labeled)	4962.39/*4962.36**
		Nitroxide (Labeled)	–
		Glycine Deletion	–

*** This is the principle peak observed in the LC-MS traces shown in Figures 3.3 and 3.4**

**** The nitroxide LC-MS trace overlaps the hydroxylamine trace, but appears to be the minor product.**

† **In all but this case, the primary product contains the hydroxylaminated version of the TOAC residue, due to reducing conditions prior to/during loading onto the LC-MS column. In this case, the true nitroxide form (which differs by the mass of an H1 atom) dominated – either because the sample was loaded relatively quickly or because proximity to the labeled cysteine more effectively protected this nitroxyl radical from reduction.**

solutions A and B are defined as follows: A – water, 0.1% formic acid; D – acetonitrile, 0.1% formic acid. LC-MS was carried out according to the following steps: in the range of 0-2 min, a 95% A - 5% B wash; in the range of 2-11 min, a 5-65% B linear ramp; and in the range of 11-12 min, a 65% B. We used a flow rate of 0.8 mL/min on a Zorbax 300SB C3 column (2.1 x 150 mm, 5 μm), at 40 °C. MS was performed by positive electrospray ionization (ESI). Observed masses were reported by averaging the major peak in the total ion current (TIC).

3.3.4 Preparative HPLC

Crude peptides were purified by reverse phase high performance liquid chromatography (HPLC). Solutions C and D are defined as follows: C – water, 0.1% trifluoroacetic acid; D – acetonitrile, 0.1% trifluoroacetic acid. Peptides were dissolved in 50% C, 50% D and loaded onto an Agilent Zorbax C3 column (21.2 x 250 mm, 7 μm). HPLC was carried out at a flow rate of 5 mL/min according to the following steps: in the range of 0-5 minutes, a 95% C – 5% D wash; in the range of 5-80 min, a 5-45% C linear ramp; and in the range of 80-85 min, a 45% C wash.

3.3.5 EPR Sample Preparation

All EPR samples were prepared by injecting 10 μL solutions of peptide in 1x phosphate buffer solution (PBS) at a concentration of 45 μM into a PTFE capillary tubes, sealed with Crytoseal resin. S_NAr reactions were performed at 45 μM peptide concentration in PBS at RT for 15 min, with CA in 5x molar excess. Potassium hexacyanoferrate(III) ($K_3Fe(CN)_6$) was added to all samples before EPR analysis to reverse the reduction of nitroxides by TFA. The maximum $K_3Fe(CN)_6$ concentration that did not result in detectable peptide degradation was used in each case, and no subsequent purification efforts were made since neither unreacted hydroxylamines nor $K_3Fe(CN)_6$ interfere with the nitroxide EPR signal. The elution profiles are reported in Fig. 3.3 (before conjugation) and Fig. 3.4 (after conjugation), with all peaks labeled. By comparing integrated intensity of the unreacted peptide elution peak in

each profile, we computed the reaction S_NAr yields reported in Fig. 3.5. In the case of unreacted peptides, 0.2 equiv. $K_3Fe(CN)_6$ was used for all analysis. In the case of the reacted peptides, 1 equiv. $K_3Fe(CN)_6$ was used for all analysis. The exception in both cases was the sequence MP01-J29. This peptide is less stable in the presence of $K_3Fe(CN)_6$, so none was added to unreacted MP01-J29 and 0.2 equiv. were used for EPR analysis of the reacted peptide. The reduction of the nitroxide in MP01-J29 increased the uncertainty of the fit of unreacted MP01-J29. After EPR, each sample was recovered and analyzed by LC-MS.

3.3.6 EPR Experiments

Continuous wave electron paramagnetic resonance (CW-EPR) spectra were collected at X-band (9.43 GHz) using a Bruker EMX+ with a variable temperature unit. Spectra were collected over 150 G field sweep with center field at $B = 3315$ G, with attenuation of 15 dB and modulation amplitude of 1.5 G. EPR spectra of a background sample containing only PBS and $K_3Fe(CN)_6$ were subtracted from each peptide spectrum. Variable temperature spectra of each sample were collected in the range of 275-310 K, in increments of 5 K. We verified by LCMS that each peptide was undamaged by the heating process and that they reacted completely with the perfluoroarene target, demonstrating that their functionality was retained.

3.3.7 EPR Fitting

Initial fitting of each sample at 150 °C was carried out to determine hyperfine **A** and electron **g** tensors using the pepper function in Easyspin.[67] Since frozen spectra were identical under scaling, regardless of the position of TOAC, the fitted tensor components of $g_{xx} = 2.0081$, $g_{xx} = 2.0051$, $g_{xx} = 2.0020$, $A_\perp = 5.13$ G and $A_\parallel = 37.6$ G were assigned to all samples during EPR fitting at higher temperatures.

Analyses of higher-temperature EPR data were carried out using non-linear least squares analysis via the NLSL program to perform Levenberg-Marquardt curve-fitting.[66,67] We fit the data for the base 10 logarithm of rotational diffusion rate,

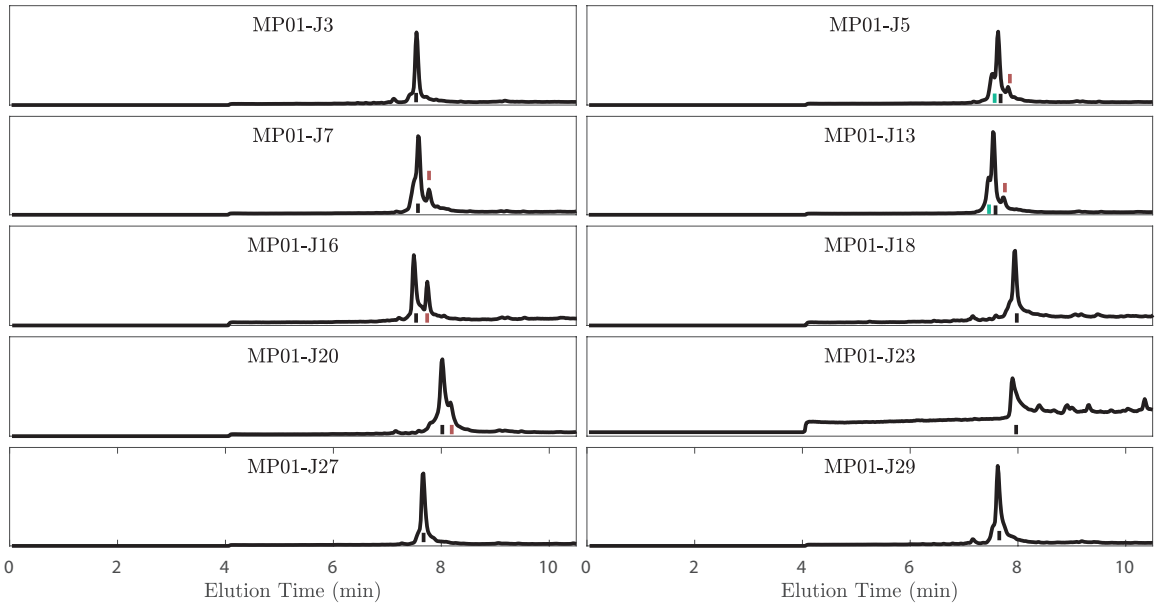


Figure 3.3: LC-MS elution profiles of MP01-Gen4 variants before S_NAr reaction and after EPR analysis. Since the samples were diluted in a solution containing TFA, and because the LC-MS column contains FA, the product often appears in both its nitroxyl and its hydroxylamine forms, and these are marked by red and black lines, respectively. The presence of a small dimer peak was usually noted. This is marked by a green line when separate from the principle peak. Reaction yields were estimated by comparing the integrated peak intensities of the unreacted species shown here (via integration/addition of both hydroxylamine and nitroxyl peaks) and after undergoing reaction (Fig. 3.4).

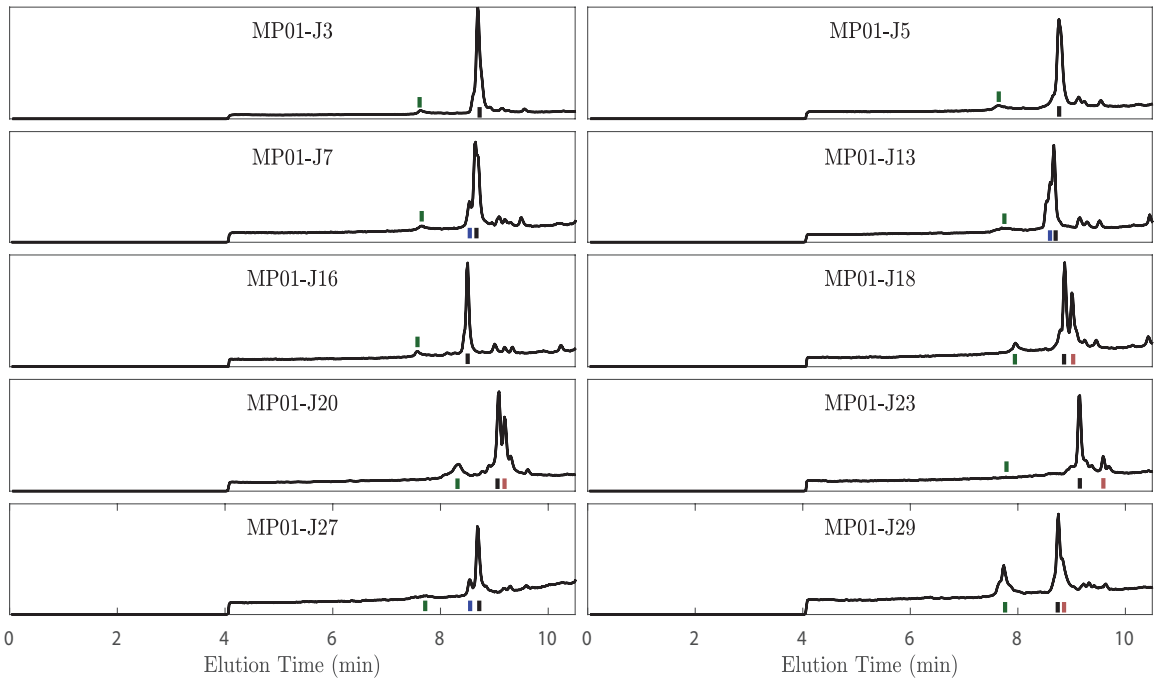


Figure 3.4: LC-MS elution profiles of the MP01-Gen4 variants following S_NAr reaction and EPR analysis. Since the samples were diluted in a solution containing TFA, and because the LC-MS column contains FA, the product often appears in both its nitroxyl and its hydroxylamine forms, and these are marked by red and black lines, respectively. The unreacted peptide peak is marked with a green line. Reaction yields were estimated by comparing the integrated peak intensities of the unreacted species shown here (via integration/addition of both hydroxylamine and nitroxyl peaks) and the same peak prior to reaction (Fig. 3.3). In MP01-J7, -J13 and -J27, a minor glycine deletion product was noted (blue) that is invisible in the elution profiles prior to S_NAr reaction, due to overlap with the principle product.

$\log(D_R)$, the Gaussian line-broadening, γ_0, and the c_{20} ordering parameter , using the microscopic order, macroscopic disorder (MOMD) model with 50 orientations. Monte-Carlo variation of initial fit parameters was carried out with an in-house MatLab script within a reasonable physical range, adding a random Gaussian noise to each data point during each iteration. The mean of this Gaussian noise distribution was zero, and the standard deviation was taken from a region of the spectrum containing no TOAC intensity (where all signal arose from instrument noise). We performed 500 fits for each of the 200 spectra, choosing parameters in the range $\log(D_R) \in [6.5, 9]$, $\gamma_0 \in [10^{-3}, 8]$, and $c_{20} \in [-5, 8]$. Fits converging outside this range were discarded. $\log(D_R)$ values were computed as the median value of good fits, inversely weighted by the fitting error χ^2, and 95% confidence intervals were computed using the distribution of fits with $\chi^2 > 1.5 * \chi_{min}$, where χ_{min} is the χ^2 value of the global best fit. This approach incorporates the error associated with overfitting into the confidence interval calculation. Overfitting is more common in slower-moving spectra, where a broader range of parameters can give rise to similar spectra. A representative example of EPR spectral fits is presented in Fig. 3.2c and remaining data are included in Appendix A.

Activation energy (Q) was calculated by linear fitting of Arrhenius plots, using the equation:

$$\log D_R = \log D_0 - \frac{Q \log e}{RT} \tag{3.1}$$

where D_0 is a constant, R is the universal gas constant, e is Euler's number and T is temperature.

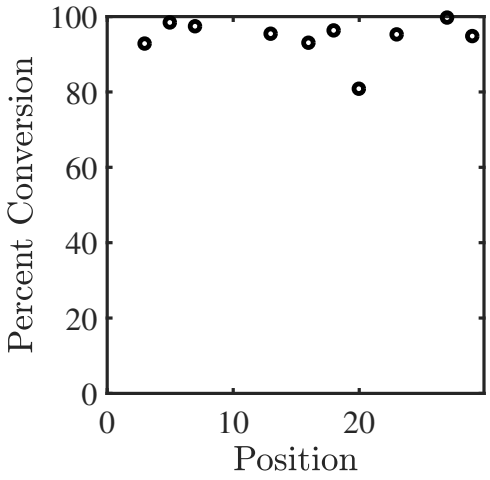


Figure 3.5: Estimated conversion yield of the S_NAr reaction, computed by the peak-integration of the LC-MS traces shown in Figure 3.3-3.4. Precise values are reported in Table 3.2.

3.4 Results and Discussion

3.4.1 Rapid flow peptide synthesis enables incorporation of amino acid spin labels

We synthesized spin-labeled MP01-Gen4 using an Fmoc-protected amino acid whose R-group contains a nitroxide radical spin-label, TOAC.[63,105] TOAC-containing peptides are desirable as EPR probes because they integrate intimately into the peptide backbone, providing an accurate measure of local dynamics. However, their synthesis remains challenging. Overcoming TOAC's steric limitations requires long coupling times and multiple couplings.[62] The speed and reliability of most amino acid couplings is improved by rapid-flow synthesis at elevated temperatures.[76,106] We adapted rapid-flow peptide synthesis to the preparation of TOAC peptides to enable reliable incorporation at arbitrarily chosen sites. In general, our results were good, producing high yields and products which could be easily purified. (Table 3.2)

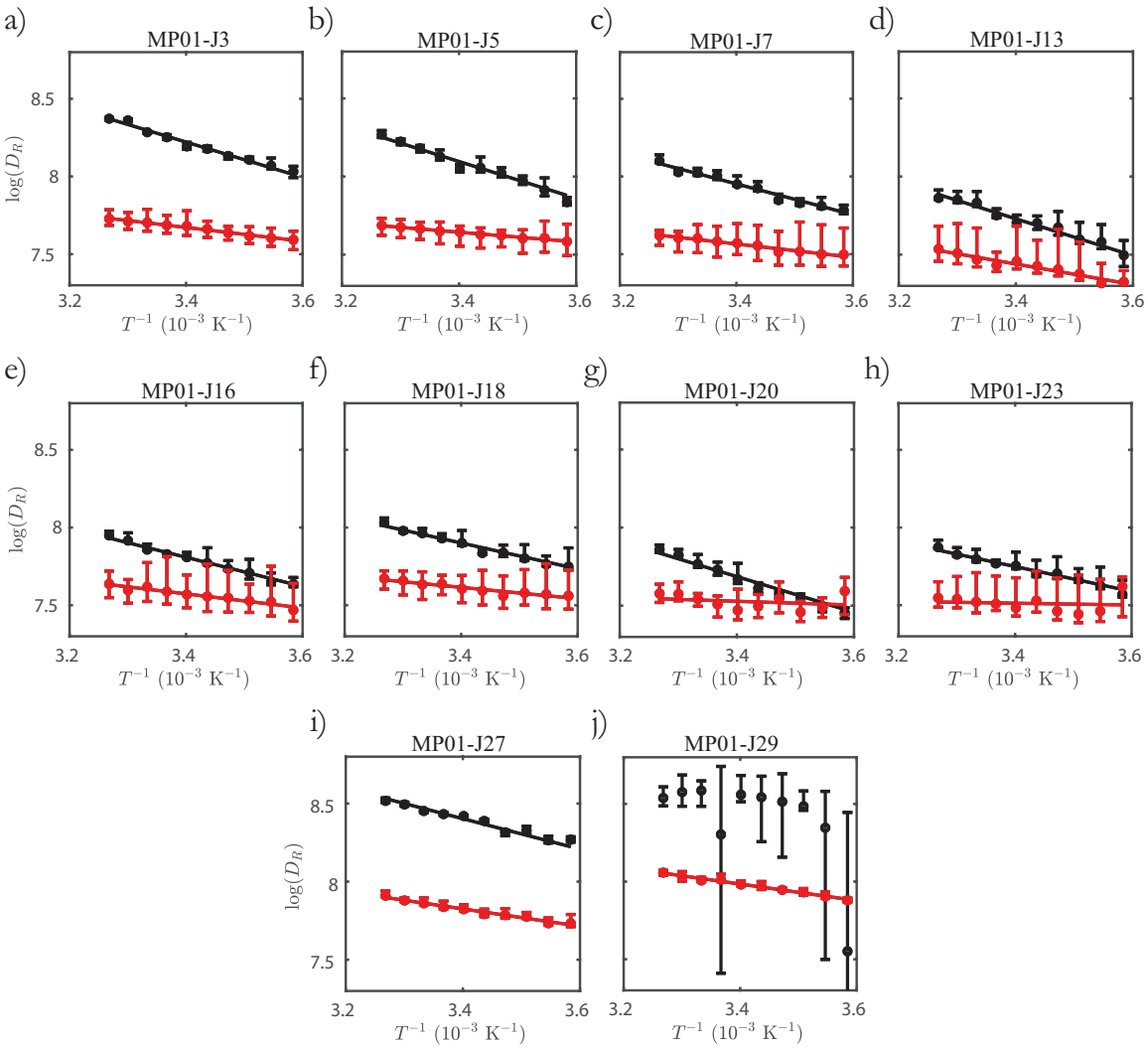


Figure 3.6: Arrhenius plots of residue-specific dynamics of MP01-Gen4 before (black) and after (red) reaction with MP01-Gen4. y-axes reflect conformational dynamics, indicated by rotational diffusion rates, determined by EPR analysis of residue-specific spin labels. Unreacted peptides diffuse more rapidly than the reacted peptides, especially near the termini. Due to the high error associated with MP01-J29, arising from low spectral intensity, a fit is not provided for these data.

3.4.2 Conformational stabilization of the peptide's termini

We introduced TOAC-substituted MP01-Gen4 to the perfluoroarene capture agent (CA) and observed rapid, near-quantitative conversion of MP01-Gen4 in almost every case (Table 3.5). These yields suggest that the incorporation of TOAC conserved the important features of MP01-Gen4 for enhanced reactivity. The only damaging substitution came at the lysine at position 20, which showed a reaction yield of 81% after replacement with TOAC. Interestingly, literature shows that substituting alanine into position 20 of a closely-related peptide enhances reaction rate.[81] Nonetheless, we kept this substitution to maintain approximately uniform TOAC spacings across the peptide.

We measured EPR spectra of TOAC peptides after conjugation with the CA. In Fig. 3.6, we report the rotational diffusion coefficients ($log(D_R)$) of each spin-label site of MP01-Gen4, both before (black) and after (red) conjugation. MP01-Gen4 experiences a sharp change in dynamics upon reaction with its target, behaving more like a rigid, structured molecule.

MP01-Gen4's rate of conformational change varies strongly with position (Fig. 3.6). For instance, dynamics of the unreacted peptide at residue 27 (Fig. 3.6i) are greater than at residue 23 (Fig. 3.6h), as shown by the overall higher rotational diffusion rates across the temperature sweep. Residues near the termini of MP01-Gen4 change conformation more rapidly than the central residues (Fig. 3.7). The five TOAC positions located within the central region show similar rates of dynamic motion at any given temperature ($\log(D_R) \approx 8$ at 310 K). Dynamics at the other five positions are faster - especially residues 27 and 29. Upon reaction with CA, the rate of dynamic motion slows dramatically throughout MP01-Gen4 (Fig. 3.7). This change is most pronounced in non-central residues, and is almost constant in the central region. The most drastic decline in dynamics upon binding occurs in the terminal residues (3, 5, 7, 27 and 29), suggesting that these undergo the greatest structural change.

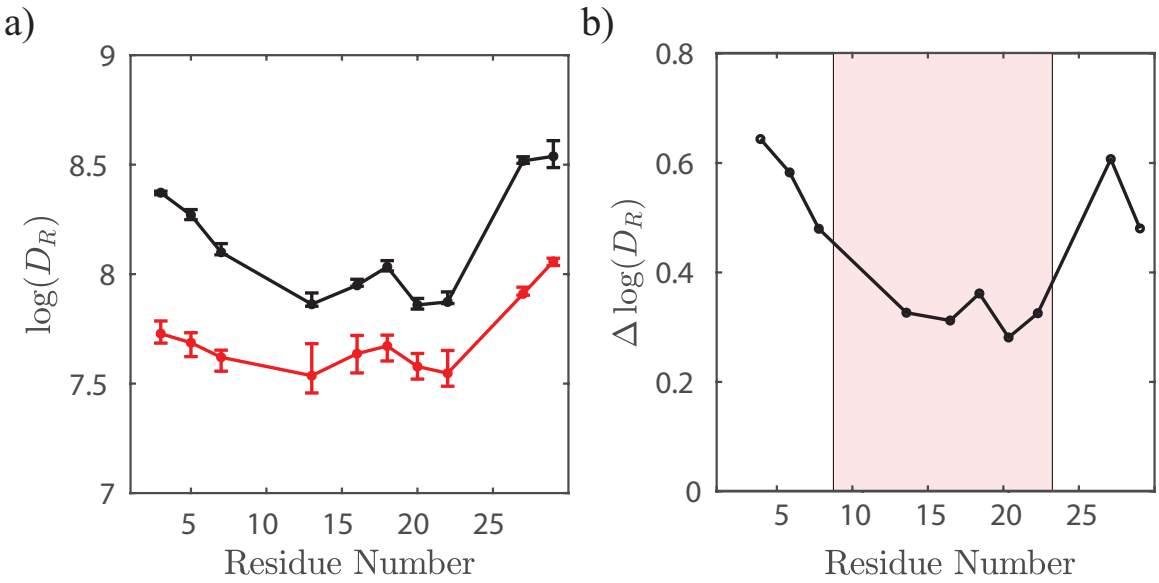


Figure 3.7: The initially disordered region of MP01-Gen4 experiences a greater change in dynamics upon reaction. a) Rotational diffusion at each probe position, collected at 35 °C for unlabeled (black) and labeled (red) MP01. b) Change in rotational diffusion upon reaction. Resides predicted to be ordered before the reaction (shaded region), experience a smaller and consistent $\Delta \log(D_R)$ of 0.32±0.03, compared to other residues, located in initially disordered regions.

3.4.3 Connecting the structural transition with the activation energy of diffusion

The activation energy (Q) of rotational diffusion represents the energetic barrier to conformational change of the peptide backbone. Using the Arrhenius plots (Fig. 3.6), we extracted activation energies of rotational diffusion of each peptide, which we plotted as a function of residue number (Fig. 3.8). Upon reaction, MP01-Gen4 exhibits a global >60% drop in Q (from an average of 8.0 k_BT to 3.4 k_BT at $T = 298$ K). The observed positional independence of Q implies that dynamic motion occurs because of global changes in conformation, rather than local effects.

At a given temperature and in the absence of significant structural changes, we would expect D_R and Q to scale inversely, according to the Arrhenius equation $D_R = D_0 \exp(-\frac{Q}{RT})$. However, despite the positional-independence of Q, we observed substantial variation in D_R as a function of position within a peptide (Fig. 4). This occurs because the Arrhenius prefactor, D_0, depends upon the degree to which each

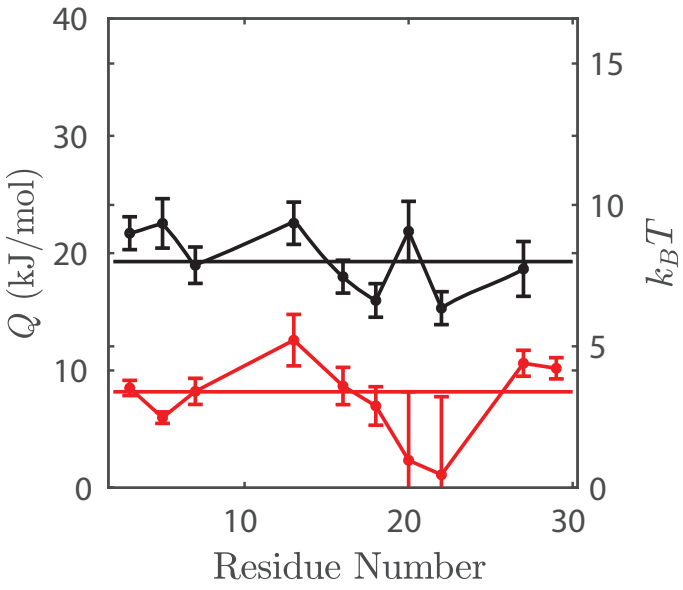


Figure 3.8: S_NAr Reaction is accompanied by a significant drop in diffusional activation energy. Activation energy (Q) of rotational diffusion vs. TOAC position in unreacted (black) and reacted (red) MP01. Average Q (8.0 k_BT unreacted, 3.4 k_BT reacted) is plotted as a line. Notably, Q is relatively independent of residue number but drops by >60% upon conjugation. Due to the high error associated with MP01-J29, arising from low spectral intensity, Q is not provided for this peptide.

conformational change displaces the nitroxide. This explains why, within a given peptide, D_R is greater in regions which are more flexible – despite the similarity of their Q –values. Changes in D_0 also explains why, upon reaction with CA, MP01-Gen4 experiences a decrease in D_R despite also experiencing a decrease in Q. In this case, the molecule becomes not only more helical, but also larger, reducing the amplitude of vibrations that displace the nitroxide radical.

In spin-labeled EPR experiments, rotational diffusion is known to occur because the peptide diffuses through distinct conformations. Therefore, temperature dependence of conformational motion is related to the peptide's conformational free energy landscape.[107–109] For peptides and proteins, this energy landscape is rough, populated by small kinetic traps.[110–113] As the peptide diffuses through its conformations, it must hop between these traps. Therefore, by measuring $\log(D_R)$ we sample the subset of those conformational changes of the backbone which move the nitroxide probe. We hypothesize that in our peptides, the activation energy of rotational diffusion must scale with the average energy barrier between distinct peptide conformations.

Typically, short proteins demonstrate energy landscape roughness values in the range of 0-5 k_BT. Our observed activation energies of 8.0 k_BT before conjugation and 3.4 k_BT after conjugation fall above this range – though still well within the diffusional regime. Other EPR experiments demonstrate similar activation behavior in proteins, noting that these values suggest H-bond formation between the nitroxide probe and the hydration shell.[108] Therefore, our observed Q values suggest that we are sampling the subset of conformational changes with sufficient energy to break this H-bond.

The decrease in Q upon MP01-Gen4 conjugation is significant. In the absence of other factors, the activation energy for diffusive motion of short polymers typically increases with molecular weight, due to increased internal friction.[114,115] Higher activation energies correspond to rougher energy landscapes, which occur when no single configuration adequately satisfies all of the intramolecular interactions necessary to fully stabilize it.[116] Rougher energy landscapes tend to correspond to greater intrinsic disorder, since disordered peptides experience greater structural change between

quasi-stable states. Therefore, this decrease in Q demonstrates that reaction with a perfluoroarene stabilizes multiple chemical sites on the MP01-Gen4 peptide that cannot otherwise be simultaneously satisfied.[116] These stabilizing interactions could form directly with the perfluoroarene, or because structural reorientation enables stronger intra-chain interactions elsewhere within the peptide. Energy landscape roughness cannot be used as a proxy for depth of the energy basin, and does not describe a peptide's overall stability or absolute state of disorder. However, in the case of MP01-Gen4, the decrease in roughness likely occurs because the system finds a more stable secondary structure. This methodology provides a novel approach for describing conformational changes within a peptide.

3.4.4 Potential reasons for positional variation in activation energy

Residue-specific variation in Q originates from three sources: site-specific variations in dynamic behavior, the uncertainty associated with spectral fitting, and the replacement of other amino acids with TOAC. We addressed fit uncertainty through Monte Carlo analysis of fit parameters, providing meaningful uncertainty estimates and avoid overfitting. Despite the care with which we selected positions for TOAC substitution, we did observe minor, sequence-specific differences in peptide reactivity. These substitutional differences likely account for most of our variation in activation energy. However, the self-consistency of both reactivity and spectral measurements indicates that the TOAC did not seriously perturb the behavior of the native MP01-Gen4 sequence (Figure 3.5).

Finally, variations arise because although conformation is a global variable, only conformational changes causing TOAC reorientation will affect an EPR spectrum. In other words, the apparent activation energy only corresponds to the subset of conformational changes that cause spin-label motion, and these changes may require traversal of an average energy barrier different than the energy barrier for overall conformational change. This means that, unlike every other technique to measure

energy landscape roughness, EPR may sample a specific subspace of the conformational energy landscape, rather than sampling the entire landscape. According to this hypothesis, a protein containing both an unstructured, flexible region and a highly structured, immobile region could experience differences in activation energy of rotational diffusion, depending on whether TOAC appears in the rigid part or the flexible part. While this effect is likely minor in small peptides like MP01-Gen4, it nonetheless explains some of the variability observed between peptides with differently positioned labels.

3.5 Conclusions

We performed the first flow-synthesis of TOAC peptides in order to study the residue-specific dynamic behavior of MP01-Gen4, a peptide designed to react with perfluoroarenes for bioconjugation chemistry. Through EPR analysis, we found that while native MP01-Gen4 is flexible and largely disordered, upon reaction with the perfluoroarene the peptide becomes significantly more rigid. Further, we identified the residues involved in the structural change, and designate the expansion of the central helical region towards the termini as its origin. Based on new physical insights, we demonstrated that a >60% decrease in the activation energy of diffusion upon reaction of MP01-Gen4 with a perfluoroarene capture agent suggests a decrease in the conformational energy landscape roughness. Thus, we conclude that MP01-Gen4 experiences a structural change upon reaction, especially in the initially-unstructured region near the N-terminus, suggesting a disorder-to-order transition upon reaction. Our results identify frustration and disorder of unreacted chains as a potentially important parameter in designing reactive peptides, and demonstrates the broad potential of EPR spectral simulations and Arrhenius analysis for studying the relationship between peptide structural transitions and reactivity. These insights could be used to design more effective screening libraries for bioconjugation.

Chapter 4

Conformational dynamics in extended-RGD binding peptide sequences

4.1 Abstract

Proteins offer a vast array of potent binding and signaling motifs with high selectivity for particular targets. The ability to reproduce the function of these signaling domains in small synthetic peptides is a central goal for advanced biomaterials design. RGD is a prolific example of a tripeptide used in biomaterials for cell adhesion, but the potency of free or surface-bound RGD tripeptide is orders-of-magnitude less than the RGD domain within natural proteins. We designed a set of peptides with varying lengths, composed of fragments of fibronectin protein whose central three residues are RGD. With these peptides, we measure conformational dynamics and transient structure of the active site. Our studies reveal how flanking residues affect conformational behavior and integrin binding. We find that disorder of the binding site is important to the potency of RGD peptides, and that transient hydrogen bonding

near the RGD site affects both the energy landscape roughness of the peptides and peptide binding. This phenomenon is independent of longer-range folding interactions, and helps explain why short binding sequences, including RGD itself, do not fully replicate the integrin-targeting properties of extracellular matrix proteins. Our studies reinforce that peptide binding is a holistic event and fragments larger than those directly involved in binding should be considered in design of peptide epitopes for functional biomaterials.

4.2 Introduction

Proteins offer a rich and bountiful array of functions including binding, signaling, catalysis, and transport. An important target is to create smart biomaterials that are capable of performing diverse functions analogous to proteins themselves.[118] One promising route towards this goal is to identify the amino acid sequence in a protein that is directly involved in a biochemical event, and covalently tether this short sequence to surfaces.[119–122] This approach has shown initial success, however, little is known about the properties and behavior of short peptide sequences when removed from their natural protein environment.

Few peptide motifs garner more attention than those containing the RGD (arginine-glycine-aspartic acid) tripeptide.[118–123] This sequence is found in extracellular matrix proteins and is known to bind to integrins – transmembrane proteins important for cell survival and adhesion. RGD is often tethered to hydrogel surfaces to promote cell adhesion, boost biocompatibility, or prevent apoptosis. RGD is also important to cancer therapeutics, where it facilitates drug delivery and inhibits angiogenesis.[122] Similar to other peptide sequences taken out of the context of natural proteins, RGD is orders-of-magnitude less effective than its parent proteins[123], but its simplicity and ease of synthesis have anyway led to its use in a vast number of technologies.[120,123]

Recent improvements in peptide synthesis techniques have enabled high throughput peptide couplings with dramatic improvements in reaction times, yields, and versatility of peptide sequence.[76,106] As a result, the prospect of producing long RGD-

containing peptides that more closely resemble the active domain of their parent protein is now a viable route to scalable integrin-targeting materials with enhanced potencies.

Many RGD peptides work by imposing a particular conformation on their binding site, as is demonstrated by molecular dynamics (MD) and nuclear magnetic resonance (NMR) studies of conformation in RGD-integrin complexes.[124–126] Very short, integrin-bound RGD peptides have been studied by X-ray crystallography, and their conformations are shown to vary depending on the specific nature of the integrin in question.[123] Only a few studies report structural information for complexes between integrins and longer fibronectin-mimetic RGD-containing sequences.[127] In solution, MD simulations demonstrate that the RGD site of fibronectin is flexible, changing conformation regularly[128], and is therefore considered to be intrinsically disordered.[129] This assessment is consistent with our predictions of disorder in the fibronectin sequence[130] using the PrDOS[97] disorder prediction tool (Fig. 4.1). The disorder of RGD is difficult to confirm directly, however disorder has been shown to play an important role in integrin-binding proteins[131–134] and other protein-binding peptides systems.[135,136]

Conformational disorder of the RGD site suggests that the conventional lock-and-key model, in which function is purely dictated by structure, does not adequately describe RGD binding interactions. More recently, two other models – the induced fit and the conformational selection models – have been used to explain the role of conformational distributions in protein binding.[45,48,137,138] In the induced fit model, the binding conformation of one binding site is induced by the presence of the other, facilitating an interaction.[137] In the conformational selection model, some subset of the unbound protein already exists in its binding conformation, and this subset is preferentially selected for binding.[138] Both cases appear in nature, and in both cases, the rate of conformational change and the range of conformations adopted by a protein binding site are central to the binding interactions of the protein.[45,139–141] Therefore, a detailed understanding of the conformational dynamics of RGD-containing fibronectin fragments is important for designing effective integrin-targeting materials.

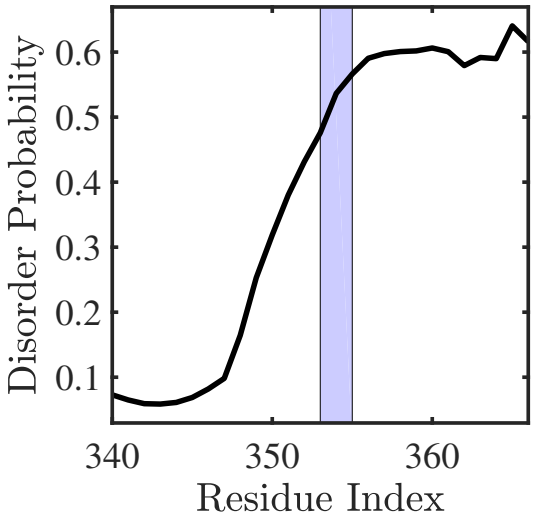


Figure 4.1: Predicted disorder probability for each amino-acid residue in the 1FNF sequence, centered on the RGD site (blue), computed by PrDOS.[97] Disorder probability ranges from 0 to 1, so the algorithm predicts that the RGD region is likely disordered.

Electron paramagnetic resonance (EPR) spectroscopy is routinely used to study dynamic behavior in peptides and proteins. EPR detects interactions between a magnetic field and radical electron spin labels. In proteins and peptides, spin labels are nitroxide radical moieties incorporated by site-directed spin labeling (SDSL) or through the direct synthesis of molecules containing an amino acid with a nitroxide spin label R-group, TOAC (2,2,6,6-tetramethylpiperidine-*N*-oxide-4-amino-4-carboxylic acid).[61,63] Where possible, synthetic introduction of TOAC is preferable to SDSL because, unlike most spin label probes, TOAC is directly and rigidly integrated into the backbone of the protein or peptide.[63]

Continuous-wave EPR (CW-EPR) enables measurement of the rotational diffusion coefficient (D_R) of the probe according to the stochastic Liouville model.[98] This parameter provides information about the local dynamic behavior of the molecule at the site of the probe and has been widely used to study binding and folding behavior in peptides and other molecules.[3,58,60,100,103,104,142–145] EPR is sensitive to dynamics of analytes at micromolar concentrations, enabling analysis of samples that are prone to aggregation. Such dynamics measurements describe the rate of conformational change

of a molecule. Consequently, EPR spectral measurements are related to the conformational energy landscape.[77,107,108] In particular, the activation energy of diffusion represents the characteristic energy barrier to conformational change.[77]

4.3 Experimental Methods

4.3.1 Materials and Measurements

Amino acid couplings were performed according to literature[76] using 50 mg of ChemMatrix H-Rink amide resin (0.49 meq./g). For the TOAC coupling, 3 eq. (0.075 mmol) 2,2,6,6-tetramethylpiperidine-*N*-oxyl-4-(9-fluorenylmethyloxycarbonyl-amino)-4-carboxylic acid (Fmoc-TOAC) was coupled overnight (RT) with 2.62 eq. HATU in 2.53 mL DMF 8% (*v*/*v*) DIPEA. Fmoc was deprotected for 20 min in DMF 20% piperidine (RT). The subsequent residue was coupled in flow by a 10 min, 10 mL/min coupling (70 °C) on the peptide synthesizer, and all remaining residues were coupled normally. Peptides were cleaved using a protocol described elsewhere[62], and precipitated by centrifugation in diethyl ether (-70 °C). The identity of each peptide was verified by high-res liquid chromatography/mass spectrometry (LC-MS) analysis on an Agilent 6520 ESI-Q-TOF LC-MS system. Each peptide was then purified by high performance liquid chromatography (HPLC) and the isolated product was tested by LC-MS and lyophilized. LC-MS was performed on a Zorbax 300SB C3 column (2.1 x 150 mm, 5 μm, 0.8 mL/min) using solvents A (water 0.1% (v/v) formic acid) and B (acetonitrile 0.1% (*v*/*v*) formic acid) by the following steps: 0-2 min, a 95% A, 5% B wash; 2-11 min, a 5-65% B linear ramp; and 11-12 min, a 65% B wash. Preparative HPLC was performed on an Agilent Zorbax C3 column (21.2 x 250 mm, 7 μm, 5 mL/min) using solvents C (water 0.1% (*v*/*v*) trifluoroacetic acid) and D (acetonitrile 0.1% (*v*/*v*) trifluoroacetic acid) by the following steps: 0-5 min, a 95%C, 5%D wash; 5-80 min, a 5-45% C linear ramp; 80-85 min, a 45% C wash.

4.3.2 PDS Experiments

Thirty minutes prior to each experiment, peptides (130 uM, final concentration) were dissolved (water, 30% (v/v) glycerol, 9 mM tris buffer (p.H 7.4), 0.07 M NaCl, 1.5% (v/v) hydrogen peroxide, 0.1 mM sodium tungstate). Hydrogen peroxide and sodium tungstate helped re-oxidize the spin labels from hydroxylamines to nitroxides, and glycerol kept the samples vitreous upon freezing. Samples were centrifuged (3 min., 15,000 RPM) to separate any aggregated peptides, and the supernatant was studied. Deuterated glycerol and water were used for the longest peptide to improve measurement accuracy.[146] Samples were loaded into quartz capillaries and quenched in liquid nitrogen.

All PDS measurements were conducted at 17.3 GHz and 60 K using a home-built Ku-band pulse EPR spectrometer.[146] DQC experiments used the 6-pulse sequence[147] having $\pi/2$- and π-pulses of 3 and 6 ns, respectively and employing a 64-step phase cycling.[147] Four-pulse DEER experiments[147] utilized for echo detection a $\pi/2$-t_1-π -t_2-π pulse sequence with $\pi/2$- and π-pulses having widths of 16 and 32 ns, which was applied at the low-field side of the nitroxide spectrum. A 16 ns π-pulse pumped at a 70 MHz lower frequency corresponding to the central maximum of the EPR spectrum. In all PDS measurements four data records were obtained, advancing each time initial inter-pulse distances by a quarter period of proton ESEEM frequency (26.2 MHz), and summed up to suppress proton modulation. The data collection time was 1-2 h per sample. The recorded data were subjected to background subtraction. The latter points (about half of the record) in the logarithm of DEER data were fit to linear background which was then subtracted out. The resulting linear-scale DEER data for distance reconstruction, $V(t)$, were modified as $V'(t) = (V(t) - 1)/V(0)$ to give the amplitude at zero equal to DEER signal modulation depth and the asymptotic value of zero[148], DQC data required just to have fitted and subtracted out small linear backgrounds. Based on DEER modulation depth, spin label re-oxidation efficiency was 0.8±0.1. All data were processed into distance distributions using either L-curve Tikhonov regularization[149] followed by the MEM refinement[150]

or de-noising/SVD methods[151,152] with the outcomes being close enough, so just the Tiknonov distributions are presented.

4.3.3 CW-EPR

Measurements were performed using a Bruker EMX+ X-band EPR (9.43 GHz), with a variable temperature unit. Stock peptide solutions were prepared at a concentration of 0.75 mg/mL and were centrifuged through 10,000 MWCO Thermo-Fisher protein concentrator tubes (30 s, 15,000 RPM) to remove contaminants and larger aggregates. Peptides were then diluted to a concentration of 150 μM in a aqueous solutions containing either 0% (v/v) or 25% (v/v) DMSO, as well as 1x phosphate buffered saline (PBS, pH 7.4), 1.5% (v/v) hydrogen peroxide (aq.) and 0.1 mM sodium tungstate. Five uL of the supernatant was loaded into a PTFE capillary tube, and sealed using Crytoseal resin. We prepared 2x, 4x, 8x 16x and 20x dilutions of these samples in the same media, in order to ensure that no concentration effects (such as peptide aggregation) changed dynamics.

EPR spectra for each sample were collected in the range of 275-325 K in increments of 5 K. Spectra were collected over a 150 G range centered at $B = 3315$ G, with attenuation of 15 dB and a modulation amplitude of 1.5 G. EPR spectra of a background sample containing only water, PBS, hydrogen peroxide and sodium tungstate were subtracted from each peptide spectrum. Finally, samples were frozen to 150 K and frozen spectra were collected. Subsequent LC-MS analysis of the scanned samples confirmed the integrity of each peptide. Using the pepper function in Easyspin[67], hyperfine ($\mathbf{A}$) and electron $\mathbf{g}$ values were collected from frozen spectra. Then, NLSL[66] was used to analyze room-temperature data, employing the MOMD[99] model, and using an in-house MatLab software package to analyze error. For each spectrum, 1000 fits were started from a randomized initial guess within constraint rectangle. In every case, a randomized Gaussian noise (based on background noise from the EPR spectrum) was added to each intensity value in the spectrum to account for experimental error. Error bars were calculated based on the final coordinates of all 'good fits', defined as those with a final χ^2 parameter within 50% of the global minimum

of χ^2. Arrhenius plots were fitted using the linearized diffusion equation (Equation 3.1).

4.3.4 Molecular Dynamics

MD simulations were performed using Gromacs v5.1.4. Initial peptide inputs were predicted using PEP-FOLD 3, an in silico conformational prediction tool.[153,154] The output of these calculations was used as an input for 200 ns simulations of each peptide, using the CHARMM36 force field and explicit water under constant NPT conditions. The temperature was maintained at 310 K. Simulations were performed in dodecahedral unit cells with periodic boundary conditions, and charges were neutralized using sodium/chloride ions, as necessary. An additional 0.154 M sodium chloride concentration was added. The dodecahedral cell was designed to allow a 1.2 nm minimum space between the peptide and the boundary. Simulations were conducted run on a pair of HP Z240 Tower Workstations over the course of 14 days. Trajectories were analyzed using either built-in Gromacs functions or using Visual Molecular Dynamics (VMD). For comparison with PDS results, we corrected for TOAC probes by assuming a rigid, TOAC structure and at each step adding the Îś-carbon-nitroxide distance vectors to the inter-carbon distance vector.

4.3.5 Cell Lines and Cell Culture

The human erythroleukemia cell line K562 stably expressing recombinant $\alpha V\beta 3$ (K562-$\alpha V\beta 3$) was described previously.[127] Cells were maintained at 37 °C, 5% CO_2 in Iscove's modified Dulbecco's medium supplemented with 1 mg/ml G418, 10% fetal bovine serum, penicillin and streptomycin.

4.3.6 Fluorescence labeling of hFN10

hFN10 was labeled with *N*-hydroxysuccinimidyl ester derivative of Fluor 647 (Alexa Fluor 647) from Invitrogen according to the manufacturer instructions. Excess dye

was removed using dialysis and buffer exchange into PBS. The final protein concentration and dye/protein molar ratio were measured spectrophotometrically.

4.3.7 Ligand Binding and Flow Cytometry

K562-αVβ3 cells were harvested and washed three times in HEPES-buffered saline (20 mM HEPES and 150 mM NaCl, pH 7.4) containing 0.1% (w/v) BSA (binding buffer, BB). 0.5x106 cells were suspended in 100 ml BB containing 1mM each Ca^{2+} plus Mg^{2+} or 1mM Mn^{2+} and washed again in the right metal ion-containing BB. For competition studies, K562-αVβ3 were incubated first with serially diluted concentrations of the unlabeled RGD-peptides in Mn^{2+}-containing BB and subsequently with 10 nM of the reporter Alexa647-conjugated hFN10 ligand for an additional 20 minutes at room temperature in the dark. Cells were washed with 4 mL metal ion-containing BB to remove the unbound reporter, centrifuged for 5 minutes at 525xg, re-suspended, and fixed in 1% paraformaldehyde. They were finally analyzed using a BD-LSRII flow cytometer (BD Biosciences) and processed using FlowJo software. The binding of soluble hFN10 to K562-αVβ3 cells was expressed as mean fluorescence intensity units (MFI), and the mean and standard deviation from three independent experiments were calculated and compared using student's t-test. The binding profile of each peptide was fit in MatLab using a Hill curve model, in order to determine half-maximal inhibitory concentrations (IC50s) and associated uncertainties.

4.4 Results

We aimed to understand the relationship between dynamic behavior and peptide structure, and to identify what, if any, connection exists between these properties and peptide binding. We used CW-EPR to analyze the conformational behavior of a series of four peptides, shown in Figure 4.2. These peptides mimic the sequence of human fibronectin (Fn) in the region surrounding an RGD site from the 1FNF fragment reported in the protein data bank.[130] We designate these sequences as fibronectin-mimetic peptides (FMPs). Each peptide sequence corresponds to a fragment of the

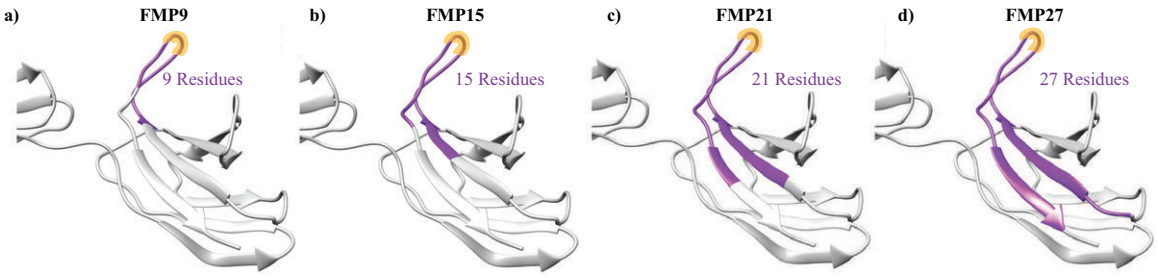


Figure 4.2: The structure of RGD-containing fibronectin fragments. RGD sites of each peptide fragment are highlighted in yellow. a) FMP9 includes the RGD site and three flanking residues on either side. b) FMP15 includes the RGD site and six flanking residues on either side. c) FMP21 includes the RGD site and nine flanking residues on either side. d) FMP27 includes the RGD site and twelve flanking residues on either side.

human fibronectin protein centered at residue 1494, the glycine of RGD. We chose four sequences for RGD conformational dynamics and binding analyses: a 9-residue sequence composed of RGD with three flanking residues on either side (FMP9, Fig. 4.2a), a 15-residue sequence with six flanking residues on either side of RGD (FMP15, Fig. 4.2b), a 21-residue sequence with nine residues flanking RGD (FMP21, Fig. 4.2c), and a 27-residue sequence with twelve residues flanking RGD (FMP27, Fig. 4.2d).

We used CW-EPR spectroscopy and spectral analysis of each peptide to measure conformational dynamics of the FMPs as a function of temperature and peptide length. To accomplish this, we substituted an amino acid spin label (TOAC) for the glycine of the RGD sequence of each FMP peptide. For simplicity, we refer to the TOAC residue as J. The amino acid sequences, designations, and expected and observed molecular weights of each compound are reported in Table 1. We measured CW-EPR of the TOAC-substituted FMPs at temperatures between -3 °C and 53 °C and fit each EPR spectrum to find the probe's rotational diffusion coefficient (D_R) (See Supplemental Information, Fig. 4.3). Lower rotational diffusion rates correspond to slower conformational dynamics at the spin label site.

Figure 4.4a shows the Arrhenius plots of each FMP in phosphate-buffered saline (PBS) solution. The rotational diffusion coefficients in the Arrhenius plots are derived from EPR fitting shown in the Supplemental Information (Fig. 4.3). Figure

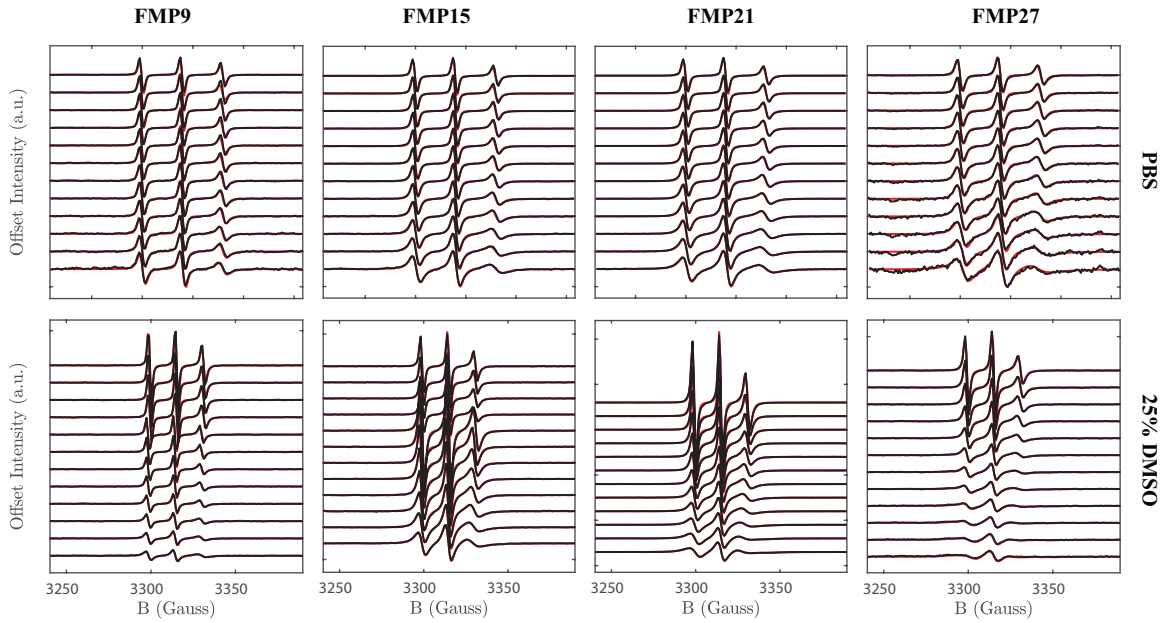


Figure 4.3: CW-EPR data (black) superimposed over spectral best-fits (red) for each FMP sample. Spectra are vertically offset based on their temperature, so that spectra collected at higher temperatures have a higher vertical offset. For associated dynamic data, refer to Figure 4.4.

4.4b presents Arrhenius plots of the same FMPs, except in a solvent mixture of 3:1 PBS:DMSO, where DMSO is a denaturant. The overall slower diffusion of the spin labeled FMPs in the presence of DMSO results from the greater viscosity of the DMSO/PBS solvent mixture compared to PBS alone. We calculated the activation energy of diffusion (Q) of each peptide as a function of length from the slopes of the Arrhenius plots (Fig. 4.4c) and found Q = 11.3 kJ/mol (FMP9), 19.4 kJ/mol (FMP15), 20.8 kJ/mol (FMP21) and 22.4 kJ/mol (FMP27) in PBS. As expected, the slope of activation energy of diffusion versus peptide length is linear when DMSO in present, since no secondary structure is possible. In contrast, a discontinuity is observed in PBS buffer, where the activation energies of diffusion of the shortest FMP (FMP9) is significantly less than that of FMP15, deviating from linear behavior. This discontinuity suggests that a structural change occurs when the number of flanking residues on each side of the RGD increases from three to six.

Few previous EPR studies describe activation energy of diffusion of polymers, peptides, or proteins. One study indicates that polymers activation energies of diffusion

Table 4.1: Fibronectin-mimetic peptide (FMP) designations, sequences, and exact masses. Each FMP sequence (named in bold) has two associated TOAC peptides – a singly labeled sequence for EPR dynamics measurement, and a twice labeled sequence used for PDS experiments. The sequences are given with the RGD site in bold, and the position of TOAC spin label residues indicated by a red J. Exact masses were computed from LC-MS spectra in Appendix C.

Name	Sequence	Exp. MW	Obs. MW
FMP9	VTGRGDSPA	857.41	857.43
FMP9-J5	VTGRJDSPA	998.53	998.51
FMP9-J0J5	JVTGRJDSPA	1196.66	1196.65
FMP15	VYAVTGRGDSPASSK	1492.74	1492.77
FMP15-J8	VYAVTGRJDSPASSK	1633.85	1633.83
FMP15-J0J8	JVYAVTGRJDSPASSK	1831.99	1831.99
FMP21	TITVYAVTGRGDSPASSKPIS	2105.09	2105.01
FMP21-J11	TITVYAVTGRJDSPASSKPIS	2246.20	2246.15
FMP21-J0J11	JTITVYAVTGRJDSPASSKPIS	2444.34	2444.36
FMP27	VDYTITVYAVTGRGDSPASSKPISINY	2872.44	2872.43
FMP27-J14	VDYTITVYAVTGRJDSPASSKPISINY	3013.55	3013.52
FMP27-J0J14	JVDYTITVYAVTGRJDSPASSKPISINY	3211.69	3211.69

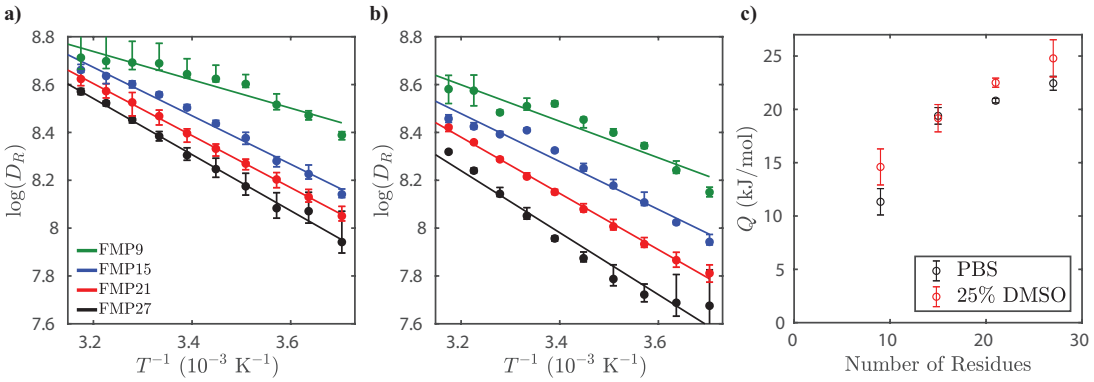


Figure 4.4: Dynamic behavior at the RGD site changes discontinuously with length. a) Arrhenius plots of rotational diffusion of TOAC peptides in buffer, generated from EPR spectral analysis. b) Arrhenius plots of rotational diffusion determined by EPR in 25% DMSO/75% buffer. The DMSO denaturant prevents secondary structure formation. c) Activation energy of rotational diffusion in buffer (black data points) and 25% DMSO as a denaturant (red data points).

exhibit no length-dependence, but in that case, the polymers were much larger than FMPs (33-500 kDa), and were not composed of amino-acids. Another report describes the diffusive motion of shorter biopolymers in terms of internal friction effects, which also applies to short peptides such as FMPs – but these results are system-specific and do not indicate the physicality of a linear relationship in this case.[114]

The activation energies of diffusion demonstrated in Figure 4.4c arise from the characteristic energy barriers associated with conformational changes that displace the nitroxide radical.[77] These barriers are influenced partially by hydrogen bonding between the peptide and its hydration shell, and partially by the roughness of the conformational free energy landscape.[77] The roughness of free-energy landscapes is important for describing protein folding.[111,113] The lower the conformational energy barrier, the more smoothly the molecule changes conformation, and the lower the conformational dwell-time.[77,111] Figure 4.4c indicates that the longer three peptides are substantially more frustrated than the shortest peptide, and are therefore more likely to exhibit strong intramolecular interactions in multiple configurations.

To understand the structural origin of dynamic behavior of FMPs, we carried out 200 nm all-atom molecular dynamics (MD) simulations and mapped the probabilities of hydrogen bond formation between residue pairs of the free FMP peptides. Figure 4.5 shows intramolecular hydrogen bonding within each FMP by indicating the probability of each pair of residues participating in a hydrogen bond. These results show that while transient hydrogen bonding is possible, the RGD site remains disordered, exhibiting several stable conformations, even in the longest peptide. This conclusion is supported by Ramachandran plots calculated for the five central residues (GRGDS) of each sequence (Fig. 4.6). Further, the disorder of FMPs is consistent with the EPR results shown in Figure 4.4, which demonstrate a relatively rapid overall rate of diffusion in all four peptides.

Figure 4.5 indicates that, near the RGD site, the longer three peptides exhibit transient hydrogen-bonding that is less pronounced in the shortest peptide (Fig. 4.5). These hydrogen bonds form intermittently between pairs of nearby residues, and do not constitute permanent secondary structure. A hydrogen bonding pattern begins

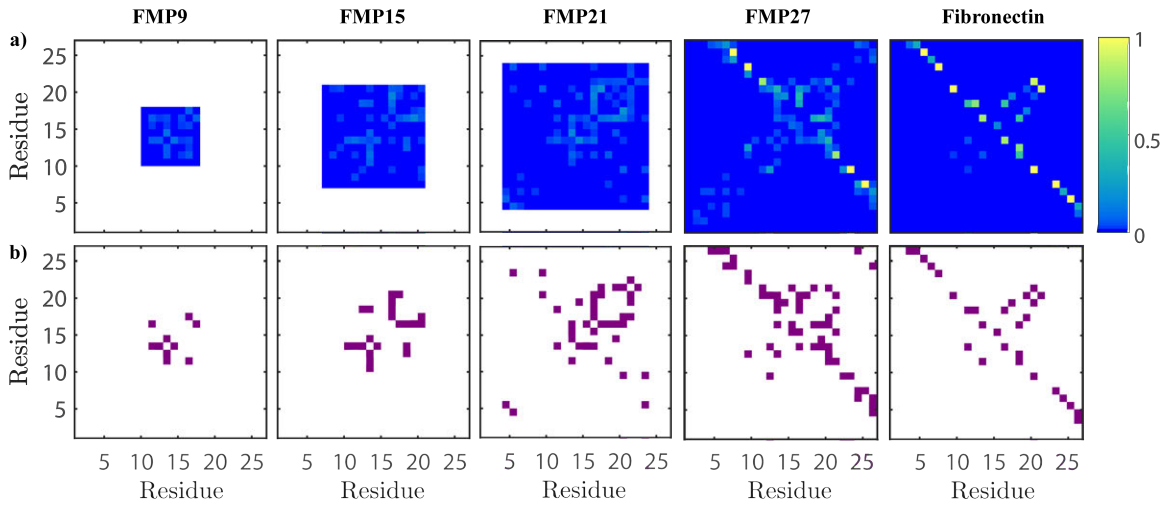


Figure 4.5: Molecular dynamics simulations illustrate intra-chain hydrogen bond formation in FMP peptides and fibronectin protein. a) Hydrogen bond probability maps of FMP peptides and the corresponding region of fibronectin, determined by MD simulations. Weak hydrogen bonding between residues near to the RGD site (appearing near the $y = x$ line) is observed. Strong hydrogen bonding between flanking chains is observed in FMP27 and fibronectin, but the residues involved in these intramolecular hydrogen bonds differ between FMP27 and fibronectin. b) Intramolecular hydrogen bonding occurring more than 4% of the time is depicted in purple. A transient hydrogen bonding pattern appears along the $y = x$ line for FMP15, FMP21, and FMP27 which has not fully evolved in FMP9.

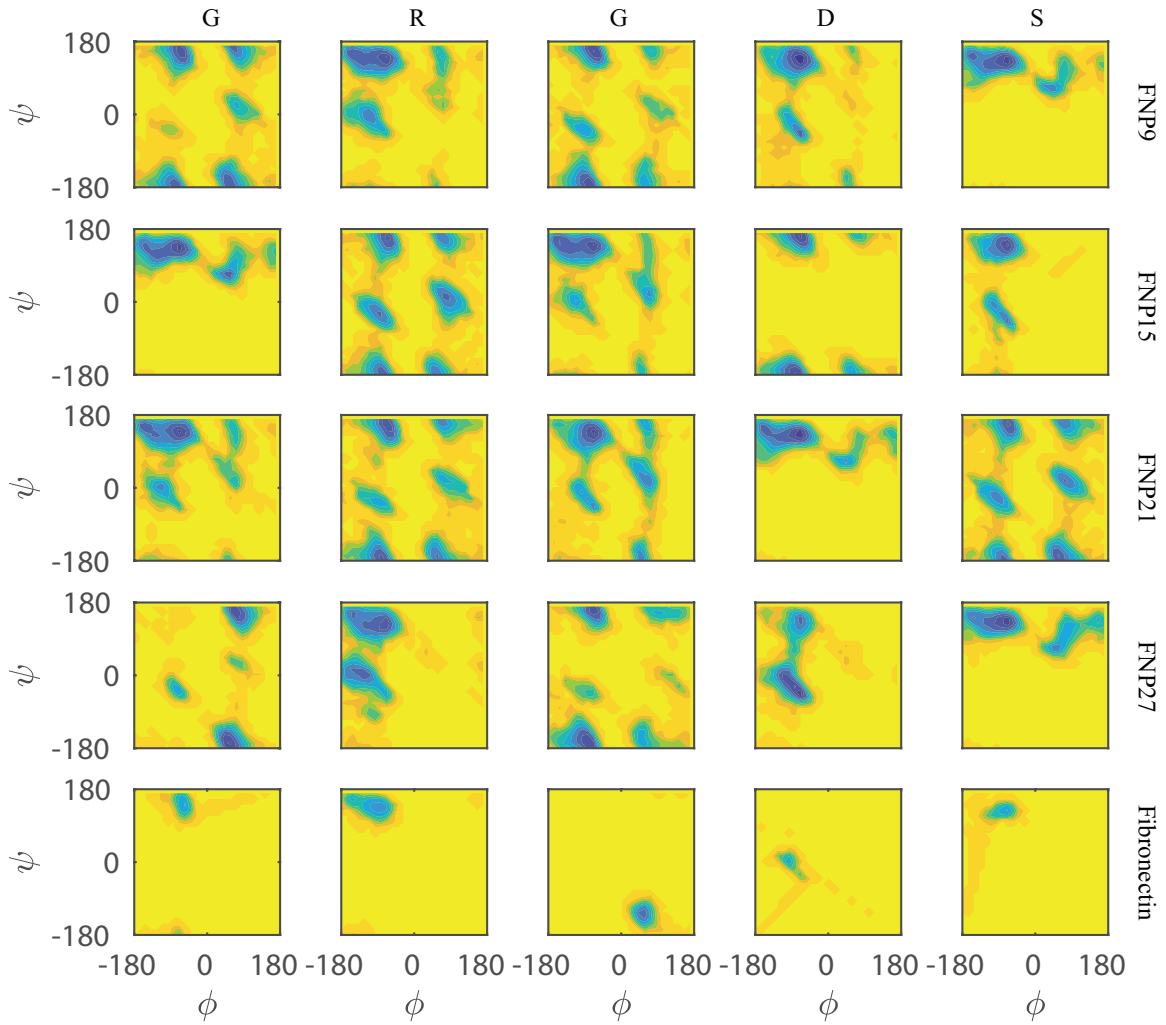


Figure 4.6: Ramachandran plots of the RGD site in 200 ns peptide simulations and a 25 ns (brief) simulation of Fibronectin, labeled by amino acid (top) and by sequence name (right). The x- and y-axes represent the dihedral angles, while the color axis represents the energy associated with a given conformation within a residue. These plots demonstrate the conformational flexibility of the RGD-loop in all four of the peptides (since many conformations are stable), but affirm that the correct binding conformation (observed in the fibronectin row) is likely accessed by all four of the shorter peptides. Fibronectin likely has other stable conformations, but these are not reflected in the 1FNF crystal structure that we studied.

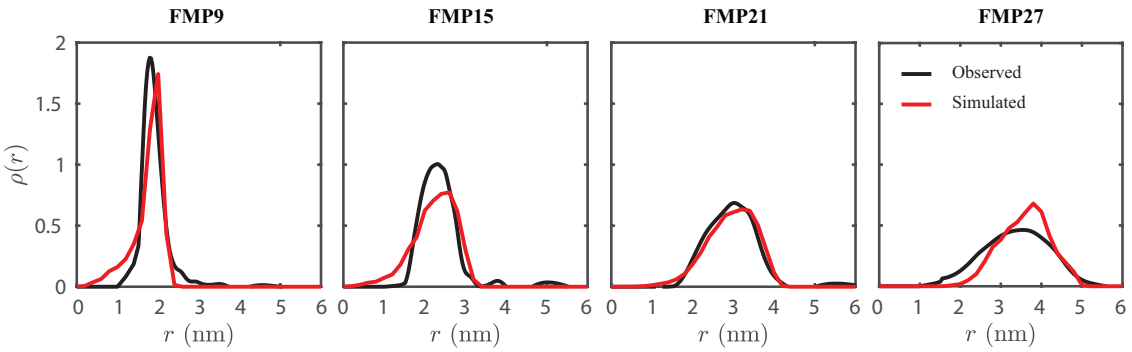


Figure 4.7: Distributions determined by pulsed dipolar spectroscopy and MD simulations describe the distances between the glycine of RGD and the N-terminus of each FMP peptide. FMP9 shows a narrow distribution centered at 1.9 nm, FMP15, FMP 21, and FMP27 each show broader distance distributions centered at 2.3 nm, 2.9 nm, and 3.4 nm, respectively.

to emerge as RGD is flanked by a greater number of residues. As seen in FMP15, FMP21, and FMP27, the hydrogen bonding probability profile remains similar despite the emergence of more permanent secondary structure in FMP27. This structure is absent in our simulation of a fibronectin protein fragment (labeled as Fibronectin in Figure 4.5), due to limitations in simulation time given the large number of atoms.

These simulations demonstrated a second change in FMP behavior as a function of length. FMP27 folded permanently due to the formation of strong hydrogen bonds between residues 9-26 and 11-24, whereas the shorter three peptides did not. Although this structure is similar to the hydrogen bonding interactions in the analogous region of fibronectin, there is a slight misalignment between the interactions of FMP27 and the analogous region in fibronectin (which has bonds between residue-pairs 6-26, 8-24, and 10-21). This mismatch preferentially distorts the RGD site into a non-ideal conformation for binding. Formation of hydrogen bonds does not appear to affect the activation energy of diffusion of the longest peptide. We attribute this phenomenon to the high stability of these hydrogen bonds, since they are nearly permanent in FMP21. While these interactions modify the overall energy landscape, they appear to have a minimal impact on its roughness.

To support these simulations, pulsed dipolar spectroscopy (PDS), an EPR technique, was used to analyze the distance between the N-terminus and the RGD binding

site. Table 1 shows the twice-labeled FMPs used for this study. With PDS, we measured inter-probe distance distributions (between 1.1-10 nm) based on the interference between radical electrons within a molecule. We used double quantum coherence[147] (DQC) and double electron-electron resonance[155] (DEER) experiments to compute each distance distribution, and compared with distributions extracted from our MD simulations. Figure 4.7 shows the distance distributions between the RGD site. The distance distributions determined by PDS are consistent with MD simulations, showing that the short peptide is generally extended, with a uniform distance distribution. As the peptide length increases, the distributions increase. This observation is consistent with the formation of hydrogen bonds in a largely disordered peptide. The general agreement between these MD and PDS distance distributions is high – the average inter-probe distance differs by no more than 2 Åin any simulation. Minor deviations between MD and PDS measurements may result from a change in sequence behavior due to the introduction of TOAC, as well as the possibility that MD simulations did not survey the entire conformational space of each peptide.

We measured binding of FMPs to $\alpha V\beta 3$ integrins on cell surfaces via displacement of fluorophore-labeled hFN10 (a high affinity ligand for the $\alpha V\beta 3$ integrin) bound to K562-$\alpha V\beta 3$ cells[127]. Cell binding measurements for each peptide are shown in Figure 4.8 along with the numerical fits used to calculate the IC50 for each FMP. Using scrambled analogs of FMPs and RGA (arginine-glycine-alanine) mutations as controls, we demonstrated that specific, rather than non-specific, binding occurs. All four peptides bound to the activated integrins, however, there were significant differences between the potency of each sequence. FMP15 and FMP21 exhibited the strongest binding, followed by FMP9 and finally FMP27. These results suggest that the dynamic transition, associated with hydrogen bonding along the peptide backbone, corresponds to increased energy landscape roughness and stronger peptide-integrin binding. We observe weaker binding between integrins and the short FMP that does not form intramolecular hydrogen bonds. We observe the weakest binding between integrins and the FMP that misfolds (FMP27), highlighting the importance of localized secondary structure in a largely disordered peptide.

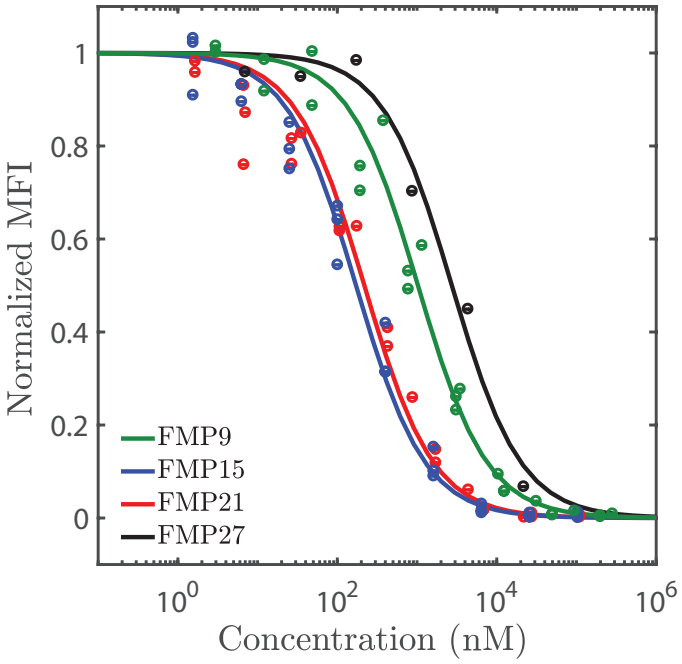


Figure 4.8: Displacement of fluorescently labeled fibronectin, bound to cellular αVβ3, by FMP peptides. Normalized mean fluorescence intensity (MFI) of integrin-bound Alexa647-labeled hFN10 protein is presented as a function FMP peptide concentration. The longer peptides, relative to FMP9, exhibit a $\tilde{5}$x improvement in their ability to bind to activated integrins. In the longest peptide (FMP27), this improvement is eliminated completely due to misfolding. The half-maximal inhibitory concentrations (IC50s) observed are: FMP9: 1030±208 nM; FMP15: 172±30 nM; FMP21: 218±52 nM; and FMP27: 2740±1070 nM.

4.5 Discussion

Here we present a strategy for improving the potencies of short peptide signals that are derived from protein active sites. We introduce four peptides that contain the RGD integrin-binding motif. Each peptide contains a different number of residues flanking RGD, with the residue sequences adapted from a fibronectin protein. We show that binding efficacy of short fibronectin fragments, with RGD as the central three residues, increases with length between nine and 21 residues, but that misfolding occurs in the 27-residue sequence. We use conformational dynamics measurements and molecular dynamics simulations to understand this behavior in the context of free energy landscapes.

The geometric and chemical structure of proteins or peptides has historically been regarded as the key feature that dictates binding efficiencies. In the case of RGD-integrin binding, our results show that dynamic behavior also plays an important role. We performed dynamics experiments on a set of short fibronectin fragments and find that rough conformational energy landscapes are correlated with significant improvements in the potency of RGD peptides. From these results, we suggest energy landscape roughness as a new design parameter for generating more potent bioactive peptides. We observed two notable features that affect binding of FMP peptides. The first feature is the emergence of transient hydrogen bonding structure. This transient hydrogen bonding increases the energy landscape roughness discontinuously with peptide length and corresponds to an increase in binding affinity of FMPs to integrins in cells. As a control, we observe that in the presence of a denaturant (DMSO), the energy landscape roughness of FMPs varies continuously with peptide length. The second feature corresponds to the formation of permanent intramolecular hydrogen bonding structure within the longest FMP (FMP27), which detrimentally affects binding.

A likely explanation for this phenomenon is that rough energy landscapes correspond to increased dwell-times of peptides in configurations appropriate for binding. These longer dwell-times stabilize a favorable conformation of the RGD site, fitting

with our understanding of conformation as the primary mechanism by which protein binding occurs. This explanation suggests that intrinsic disorder at the RGD site promotes fibronectin binding, and could reveal why this site is disordered in fibronectin. Notably, permanent folding of FMP27 via hydrogen bonding had little or no bearing on the energy landscape roughness or dynamic behavior at the RGD site. Further studies are needed to confirm that this behavior plays a role in integrin binding by fibronectin proteins.

In summary, we used EPR spectroscopy, quantitative spectral analysis and molecular dynamics simulations to understand the role of flanking residues in controlling the conformational dynamics and the free energy landscapes of RGD peptides. We observed the emergence of transient secondary structure, which significantly improved RGD binding to integrins. We conclude that energy landscape roughness may contribute to binding of native fibronectin proteins, and should be considered in future analyses of peptide binding. Finally, we suggest that dynamic and transient structure is important for peptide binding, and that binding interactions of peptide-functionalized biomaterials could benefit from consideration of dynamics and incorporation of additional residue, beyond the active binding site.

Chapter 5

A global minimization toolkit for batch-fitting and χ^2 cluster analysis of CW-EPR spectra

5.1 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. While these tools have dramatically broadened the capabilities of EPR dynamics analysis, their implementation can easily lead to overfitting, user errors, or insufficient self-consistency. As a result, dynamics parameters and associated properties become difficult to reliably determine, particularly in the slow-motion regime. Here, we present a novel EPR analysis strategy, and the corresponding computational tool, for batch-fitting EPR spectra and cluster analysis of the chi-squared landscape. We call this tool CSCA (Chi-Squared Cluster Analysis). The CSCA tool allows us to determine self-consistent rotational diffusion coefficients 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 planar nanofiber with radical electron spin labels positioned at known distances off the surface. We anticipate that this tool will increase the reproducibility of EPR fitting for the characterization of dynamics in biomolecules and soft matter.

5.2 Introduction

The dynamic behavior of biomolecules and soft materials is often critical to their function.[59,77,115,156] 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.[45,46] In self-assembling materials with biological applications, dynamics measurements are used to understand thermodynamic phase behavior, and even to identify liquid/solid phases within different regions of the same self-assembled sample.[59,68] Dynamic behavior is complicated and challenging to study, especially in heterogeneous systems such as membrane proteins, or in systems that precipitate or aggregate at high concentrations.

Continuous wave electron paramagnetic resonance (CW-EPR) is a powerful tool for measuring dynamic behavior[64–67,98,157,158] of site-specific radical electron spin labels. 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[60,61,157], peptides[63,77], polymers[159,160], and small-molecules in water.[59,68,161] In these systems, spin-labels are covalently tethered to molecules that tumble slowly. 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.

The stochastic Liouville equation (SLE) quantum-mechanically describes the spec-

tra associated with slowly-moving electrons and is important for analyzing conformational dynamics on timescales relevant for EPR dynamics measurements.[64,98,158] Several software packages that use the SLE model are available for fitting EPR spectra: most notably, the NLSL software package[66], EasySpin[67], and MultiComponent, a LabVIEW wrapper for NLSL.[162] 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.[99,163] The validity of the SLE model in the slow-motion regime has been verified by MD simulations, which can accurately reproduce EPR spectra from atomistic trajectory files, and which confirm that the diffusion of spin labels gives an accurate approximation of the dynamics of the local environment.[55,164–167]

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.[77] Such analysis reduces the risk of overfitting and improves the reproducibility and consistency of fits. In non-linear curve fitting, there are many cases where the absolute best-fit is less consistent with physical reality than other possible fits – even in systems with very few parameters.[168,169]

Here we report an open-source MatLab-based software referred to as CSCA (Chi-Squared Cluster Analysis) that interfaces with NLSL to improve the quality 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.

We used CSCA to analyze the dynamic behavior of spin labels attached via oligoproline spacers to aramid amphiphile (AA) planar nanofibers after spontaneous self-assembly in water. These amphiphilic molecules were designed to form planar

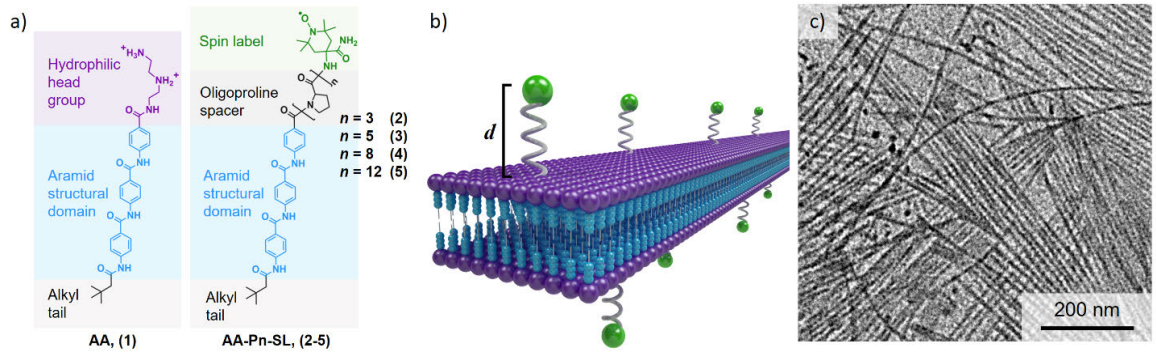


Figure 5.1: Self-assembly of planar nanofibers with spin labels tethered at known distances off the surface. a) A cationic aramid amphiphile (AA, compound (**1**)) and spin labeled analogues with oligoproline spacers (AA-Pn-SL, (**2-5**) self-assemble spontaneously in water. b) AA (compound **1**) is co-assembled with 2 mol % (**2**), (**3**), (**4**), or (**5**) to form planar nanofibers in water with integrated oligoproline-linked spin labels at distances, d, off the nanofiber surface. c) A representative TEM micrograph of compound (**1**) co-assembled with compound (**5**) show that spin labeled oligoproline AAs do not affect the geometry or dimensions of AA nanostructures.

nanofibers of 5 nm width, 4 nm thickness, and up to 20 Îijm in length. AA nanofibers exhibit negligible molecular exchange over two months, in addition to high Young's moduli and tensile strengths.[170] Figure 5.1a illustrates the chemical structure of the AA molecule (**1**) and the spin labeled analogues (**2-5**) that are incorporated by co-assembly at low concentrations (2 mol %). The spin labeled planar nanofiber is schematized in Figure 5.1b, and a transmission electron microscopy (TEM) image of a spin-labeled assembly is presented in Figure 5.1c. Oligo- and polyproline spacers are often incorporated into biomaterials, where they are used to enhance cellular interaction with functionalized end-groups.[171,172] A major advantage of prolines over other spacers is their rigidity – oligo- and polyprolines are often used as 'molecular rulers', due to their tendency to coil into stiff helices with known lengths.[173,174]

The dynamic structure of protein-binding peptides can also play a significant role in peptide interactions.[77] 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 very rigid, resulting in slow conformational changes.[174] In this regime, self-consistent analysis of EPR results becomes very chal-

(6)

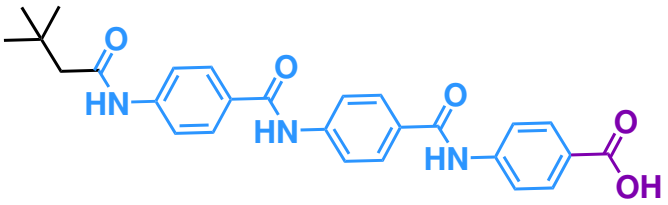


Figure 5.2: The structure of (6), the AA tail coupled to polyproline to synthesize compounds (2-5).

lenging, making it an ideal test-case for our novel analytical methods.

5.3 Experimental Methods

5.3.1 Sample synthesis

TOAC was coupled to 50 mg of ChemMatrix H-Rink amide resin (0.49 meq./g) using 3 eq. 2,2,6,6-tetramethylpiperidine-*N*-oxyl-4-(9-fluorenylmethyloxycarbonyl-amino)-4-carboxylic acid (Fmoc-TOAC) and 2.62 eq. 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) in 2.53 mL *N*,*N*-dimethylformamide (DMF) 8% (*v*/*v*) *N*,*N*-Diisopropylethylamine (DIPEA). Fmoc was cleaved using DMF 20% piperidine and the resin was washed with DMF. The subsequent proline was attached via an overnight reaction with 7 eq. Fmoc-proline and 6.65 eq. HATU in 4.80 mL DMF 8% (*v*/*v*) DIPEA. Literature protocols were used for the coupling of all subsequent proline residues.[76] Compounds (**1**) (Fig. 5.1) and (**6**) (Fig. 5.2) were synthesized according to methods which we report elsewhere.[170] After Fmoc deprotection of the N-terminal proline, (**6**) was coupled to the N-terminus of the resin-bound peptide via an overnight coupling with 5 eq. (**6**) and 4.5 eq HATU in 800 μL DMF 8% (*v*/*v*) DIPEA at 60 °C. The result was washed under DMF, dried, cleaved via a protocol described elsewhere[62], and precipitated by centrifugation in diethyl ether (-70 °C). The identity of each compound was verified by high-resolution liquid chromatography/mass spectrometry (LC-MS) analysis on an Agilent 6550 ESI-Q-TOF LC-MS system. Each peptide was then purified by high performance liquid chromatography (HPLC) and the isolated product was tested by

LC-MS and lyophilized. LC-MS was performed on a Zorbax 300 SB-C3 column (2.1 x 150 mm, 5 μm, 0.8 mL/min) using solvents A (water 0.1% (v/v) formic acid) and B (acetonitrile 0.1% (v/v) formic acid) by the following steps: 0-2 min., a 1% B wash; 2-12 min., a 1-91% B linear ramp; and 12-13 min, a 1% B wash. Preparative HPLC was performed on an Agilent Zorbax C3 column (21.2 x 250 mm, 7 μm, 5 mL/min) using solvents C (water 0.1% (v/v) trifluoroacetic acid) and D (acetonitrile 0.1% (v/v) trifluoroacetic acid) by the following steps: 0-5 min., a 95%C wash; 5-80 min., a 5-45% C linear ramp; 80-85 min., a 45% C wash.

5.3.2 EPR sample protection

Compound (**1**), AA-P3-SL, AA-P5-SL, AA-P8-SL, and AA-P12-SL were dissolved separate stock solutions of 50% C and 50% D, and the concentration of each was estimated based on the UV absorbance at 260, 280 and 320 nm, with reference to a standard absorbance curve determined for compound (**1**). Aliquots containing 98% (**1**, 2% linker (mol/mol) were prepared, and these were lyophilized. We calculated the mass of the mixture from the initial molar concentrations of each species, dissolving the dry mixture in ultrapure water to reach a final concentration of 5 mg/mL (label concentration was 200 μM). The resulting aqueous aliquots were then bench sonicated for 30 minutes, to facilitate assembly of nanofibers containing labeled oligoproline linkers. Nanofiber assembly was verified by negative-stain TEM, and potassium hexacyanoferrate was added (at a final concentration of 0.94 mg/mL) to reoxidize spin-labels that were reduced by trifluoroacetic acid (TFA) during synthesis. Ten μL of this product was then added to a PTFE capillary tube, which was sealed with crytoseal resin.

5.3.3 EPR data collection

EPR was performed on a Bruker EMX+ X-band EPR (9.43 GHz), with a variable temperature unit. Dilutions down to 1 mg/mL, 0.5 % (mol %) labeled oligoproline were prepared, but noting no deviation in the shape of the EPR spectra, we chose

the comparatively higher label-concentrations noted in the previous subsection, since these provided the greatest signal-to-noise ratio. Spectra were collected over a 150 G field sweep centered at $B = 3315$ G, with an attenuation of 15 dB and a modulation amplitude of 1.5. Variable temperature experiments were performed in the range of 275-302 K, in increments of 3 K.

5.4 Computational Methods

5.4.1 Background on the SLE model and fitting function

The theory of EPR spectral simulation is complex, and has been more thoroughly described elsewhere.[64,67,98] In brief, EPR spectra of nitroxide radicals may be computed numerically by iteratively solving the SLE equation, which depends on several parameters. In particular, spectra are usually defined by an electron's gyromagnetic **g**-tensor, its hyperfine **A**-tensor, and its rotational diffusion **D**-tensor, which contains elements inversely related the electron's rotational correlation times, called the τ-tensor. In the literature, **D** and τ appear with relatively equal frequency. Here, we limited ourselves to the case of isotropic diffusion, where all diagonal components of the diffusion tensor are equal to the scalar D_R, and all off-diagonal components are zero (because the tensor is defined relative to the principle axes of diffusion).

Simply simulating EPR spectra is usually insufficient for interpreting real data sets. Instead, we perform nonlinear fitting in order to find simulation parameters which best describe experimental data sets – this is done using an algorithm for χ^2 minimization. χ^2 (defined in Eq. 2.10) is a function quantifying the difference between experimental and model-predicted values as a function of a set of fit-parameters, **c** (10, 12).[64,66]

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 $\mathbf{c} = \mathbf{c}_{min}$,

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.[175] The algorithms used for this purpose range in sophistication 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 (29, 36).[168,175]

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.[168,169]

The χ^2 landscape of a non-linear fit is commonly populated by a large number of local minima, often clustered in the vicinity of the global optimum. This phenomenon can be referred to as multimodality or ruggedness.[175] 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.[168] In rarer cases, obvious clusters of local minima are misleading, and a truly accurate model is ill-informed by χ^2 analysis alone. This means that true insight into the nature of the model is required for accurate fitting of data.[169] Continuous-wave EPR analysis is subject to these confusions, and thus global optimization alone is not sufficient for the analysis of experimental spectra.

5.4.2 Fitting protocols and analysis of the χ^2 landscape

Here, we consider that the uncertainty associated with fitting always arises from two sources. 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 will 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 a much more significant problem than experimental error.

We can easily account for the effects of experimental noise in EPR fitting via a Monte-Carlo approach.[168] By fitting the same spectrum many times, randomly varying the experimental noise, we ensured that we accurately incorporated the effect of noise into any uncertainty estimates we generated. In our software, we refer to this option as Monte Carlo resampling. To describe fit uncertainty, we ran 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 these minima whose values fall below some threshold value (χ^2_{thresh}). The threshold value is small enough that all models accurately describe the data, but is 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 reliable confidence intervals to each parameter. Since the values of parameters within a cluster are often correlated,

these distributions certainly overstate the uncertainty of our fitting process – however, this process allows us to assign uncertainty estimates to fitted parameters. We combined the MCLM method with Monte Carlo spectral variation to account for both experimental and fit uncertainty.

In our view, the reproducibility problem of continuous wave EPR fitting results from the fact that positions of global optima vary considerably within their good-fit cluster as a consequence of random variations in data. In contrast, we observed the positions of the clusters themselves to be quite reproducible. We therefore took the approach of assigning a measure of central tendency to fit clusters, in order to achieve self-consistency. Since these clusters tended to have unusual shapes, we compared four metrics for central tendency – the mean, the marginal median, the geometric median and the medoid – to identify which of these could describe the data accurately. All four metrics are included within our software package, and our comparisons between these and the global optima are presented in the results section.

5.5 Results

5.5.1 Peptide spin labels incorporated into self-assembled nanofibers

We synthesized four samples, whose sequences are outlined in Table 5.1, for co-assembly into AA nanofibers. D_R values were smallest for the shortest proline sequences, indicating that the motion of these peptides was the most restricted. This, coupled with the observed lack of concentration dependence, is a good indicator that labeled samples integrate into the self-assembled nanostructure. If the samples had dissolved in water, shorter peptides would ordinarily move faster than longer ones.[114] If the labeled molecules had instead precipitated into a unique phase, spin-spin interactions would have resulted in a significantly different EPR spectral shape. Therefore, we conclude that the labels are present in all four systems, as a dilute species dissolved in the self-assembled nanofibers.

As evidence of synthetic purity, we submit LC-MS elution profiles and mass-

Table 5.1: Spin-labeled oligoprolines tethered to the surface of AA nanofibers. The spin label TOAC residue is denote by J. The expected and observed molecular weights are presented, with supporting LC-MS data presented in Figure 5.3. The expected oligoproline helix length is calculated based on the known PPII helix length of 0.31 nm per proline residue.[176]

Compound	Name	Surface sequence	Exp. MW	Obs. MW	Appx. helix length (Å)
(2)	AA-P3-SL	-PPPJ	961.48	961.50	9.3
(3)	AA-P5-SL	-PPPPPJ	1155.58	1155.61	15.5
(4)	AA-P8-SL	-PPPPPPPPJ	1446.75	1446.76	24.8
(5)	AA-P12-SL	-PPPPPPPPPPPPJ	1834.96	1834.98	37.2

spectrometry data for each spin-labeled amphiphile in Figure 5.3. A small, secondary elution peak was observed slightly prior to the elution of the principle peak in all four peptides. This resulted from an impurity on the LC-MS column, which was present in the absence of any injection. It features more prominently in the elution profiles of shorter peptides because these were loaded at smaller concentrations.

5.5.2 Measures of central tendency achieve self-consistent descriptions of D_R

We produced Arrhenius plots of the rotational diffusion coefficient, D_R, of each sample as a function of temperature. In almost every case, using the global optimum of D_R resulted in a high degree of inconsistency 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. However, all three median-based methods agreed on the value of D_R at every temperature, resulting in the linear, Arrhenius behavior for D_R that is typical for diffusive processes. Figure 5.4 presents these results, with Figure 5.4(a-d) demonstrating the Arrhenius fits produced by each method for compound (**1**) co-assembled 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. Figure 5.4(e) presents the R^2 obtained for each Arrhenius fit. This figure indicates that all three

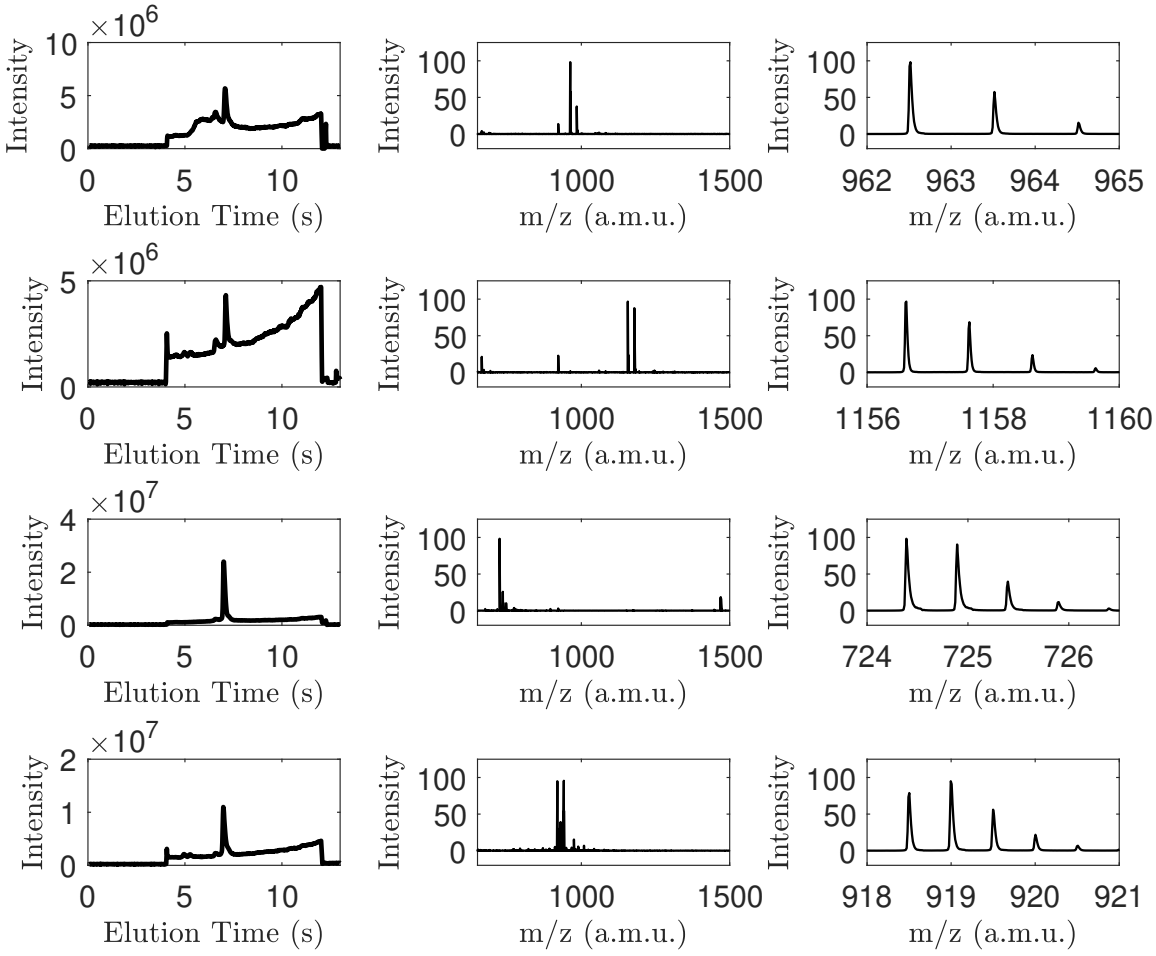


Figure 5.3: Elution profiles (left), and mass-spectrometry (m/z) profiles (middle, right) of HPLC-purified a) AA-P3-SL, b) AA-P5-SL, c) AA-P8-SL, and d) AA-P12-SL. In every case, the principle peak corresponds to the mass expected for each compound (Table 1). For AA-P3-SL and AA-P5-SL the $[M+1H]^+$ charge state was used to compute molecular weight, and for AA-P8-SL and AA-P12-SL, the $[M+2H]^{2+}$ state was used. We observed high relative signal from sodium adducts, explaining the multiple peak-sets observed in the central mass spectrometry profiles.

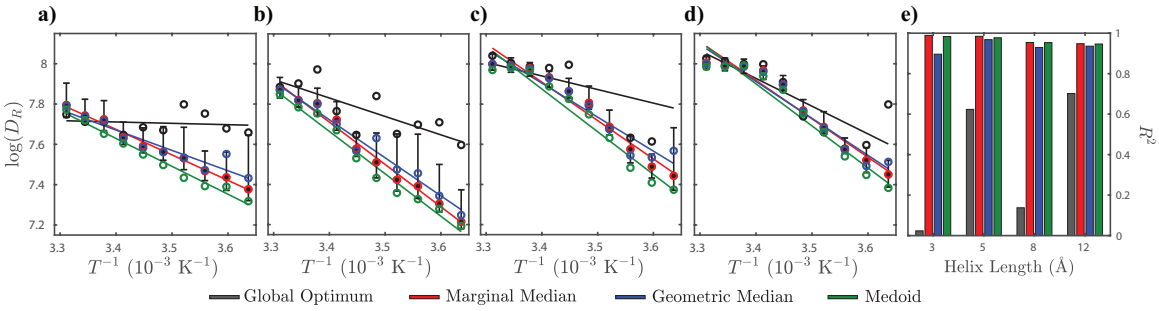


Figure 5.4: Arrhenius plots describe D_R derived from EPR fits. Arrhenius plots of co-assemblies of a) compound (**1**) and (**2**), where the spin label is positioned close to the nanofiber surface, b) compound (**1**) and (**3**), c) compound (**1**) and (**4**), and d) compound (**1**) and (**5**), where the spin label is far from the nanofiber surface. 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.

median methods agree that D_R is thermally activated, whereas the global optimum method does not.

Although D_R was effectively unchanged between AA-P8-SL and AA-P12-SL, the shorter peptides exhibited much slower rates of conformational change. Multiple explanations may justify this, but the likeliest are either i) interaction between the helices, labels, and/or amphiphile head-groups, which are only possible in the shortest peptides; or ii) an effect of solvent dynamics, since it is well-established that the dynamics of water molecules slow down near interfaces.

Significantly, median-methods agree that for slowly-moving probes, an accurate value for D_R should fall well below the value associated with the global optimum. In the EPR community, it is known that the SLE tends to perform worse when rotational correlation times are longer.[64] Our observation strongly supports this fact, indicating that overfitting becomes more prevalent at lower temperatures due to broadening of good-fit clusters. This broadening is manifest in the larger error-bars observed at low temperatures. Thus, a particular strength of cluster-based fit analyses appears to be a substantial increase in the range of D_R values for which fitting can provide

meaningful insights using the SLE model.

5.5.3 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 failed to agree on appropriate values for other parameters. This resulted from the unusual curvature of good fit clusters in parameter space. Figure 5.5(a) represents the χ^2_ν landscape of these parameters for a representative AA-P3-SL sample collected at 275 K, based on values calculated during every iteration of an MCLM analysis which started from 500 initial positions. Purple points are overlaid to represent the positions of χ^2 minima, which clearly form a cluster containing the global optimum (labeled). Each metric of central tendency is also indicated. For these spectra we used a 3-parameter model, fitting for the c_{20} ordering parameter, the Gaussian line broadening, γ_0, and D_R. The two-dimensional scatter plots in Figure 5.5(a) flatten the χ^2_ν function to show its dependence on each pair of these variables, while the complete dependence is presented in the three-dimensional plot. The positions of the global optimum and the various median positions are marked on the plots, and the final figure compares the global optimum fit with the original data.

Figure 5.5(b) exists to demonstrate the overfitting problem. It is identical to Figure 5.5(a), except that instead of fitting the original data set, we fit the simulated, global optimum spectrum that best describes the original data. The same randomized starting values were used to analyze both the experimental and the simulated data sets. In a self-consistent system, these two spectra should have the same global optimum. However, we observed that the position of the global χ^2 optimum deviates significantly from the best-fit parameters used to generate this spectrum. This demonstrates the inconsistency and uncertainty of the global χ^2 optimum as a representation of reality.

In contrast, all median methods were self-consistent in comparisons of simulated vs. real data. This can be verified by comparing inspection of the figures in Ap-

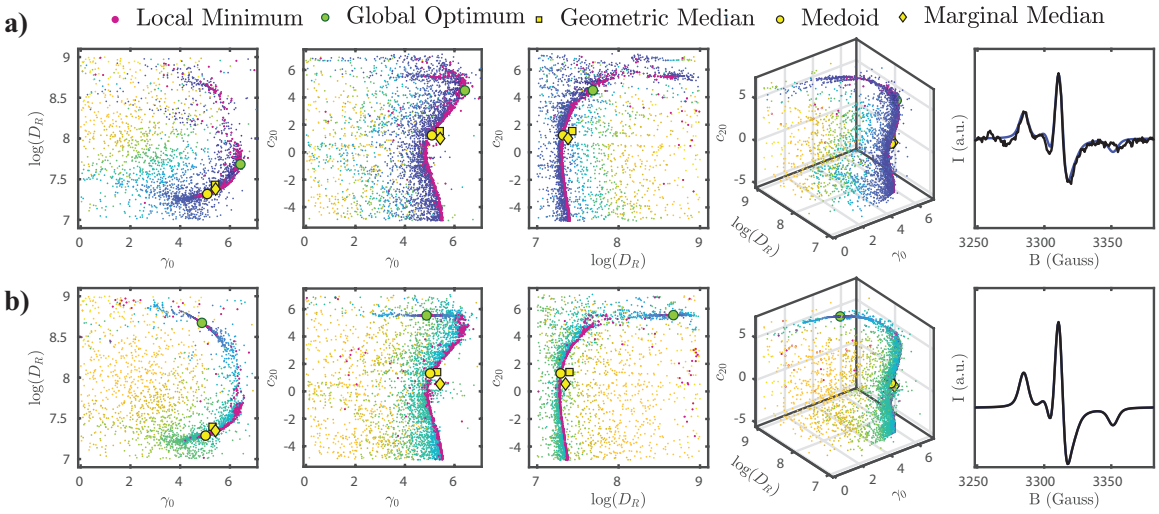


Figure 5.5: Representations of the χ^2_ν landscape for a spectral fit. a) Scatter plots of the χ^2_ν landscape associated with fitting the EPR spectrum of AA-P3-SL at 275 K. The first three plots flatten the χ^2_ν landscape to a function of 2 parameters, while the fourth preserves its full, three-dimensional shape. Local minima, the global optimum, and the various cluster medians are plotted 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 over-fit. In contrast, the median-methods remained highly self-consistent.

pendix D, which are identical to Figure 5.5, except that they describe every other sample/temperature spectrum we collected. This indicates that clusters of good fits are self-consistent, even if positions of the global optimum are not. Therefore, we argue that all four metrics describe reality more accurately than the global optimum of χ^2, since they reference self-consistent clusters of good fits. Furthermore, the Arrhenius behavior of all of these metrics suggests that cluster position seems to depend logically on temperature, further supporting this method of analysis.

However, not every median-measurement performed equally. While they all agreed on the value of D_R, the U-shaped nature of good-fit clusters meant that the marginal median frequently selected points that did not fall within the cluster itself. In contrast, the geometric median tended to conform quite closely to the cluster, and the medoid must fall within the cluster by its own mathematical construction. In general, we preferred the medoid, since this produced a highly self-consistent and physically representative description of all parameters used to fit a spectrum.

Since we established univariate distributions for each parameter, our error-bars do not depend on the precise position of the global best fit. In our experience, they greatly overstated error (since good fit clusters tended to follow predictable patterns, and appropriate analysis of these could produce a significantly more certain estimate of uncertainty in each parameter). 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.

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

The activation energies computed for each peptide are presented in Figure 5.6, as a function of estimated distance from the fiber surface. We estimated this distance based on the helix length presented in Table 5.1, but acknowledge that the absence of a head-group in AA-Pn-SL compounds, the added length of the TOAC, and the

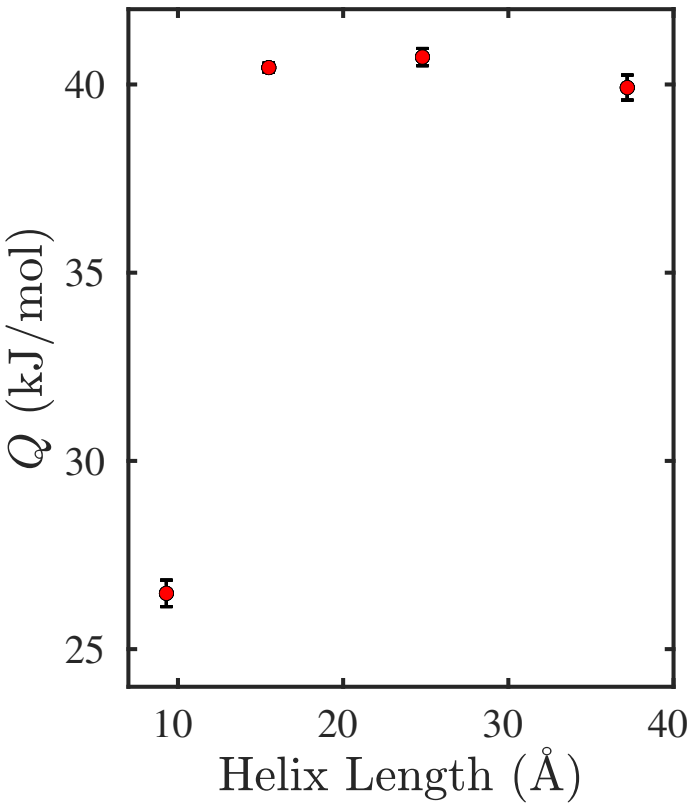


Figure 5.6: Activation energies of diffusion, Q, of spin labels tethered at known distances off a nanofiber surface. Values of Q computed from the medoid values for D_R, as a function of spin label height with error bars based on the standard error of each linear fit. The oligoproline spacer of compound (**2**) is too short to form a complete helix (<4 residues). The activation energies of diffusion of compounds (**3-5**) integrated into compound (**1**) nanofibers are consistent – approximately 40 kJ/mol – comparable to the energy barrier associated with conformational change in oligoproline helices.

angle between the helix and the assembly's surface may affect this length. For peptides long enough to form oligoproline helices (longer than 3 residues), the activation energy of diffusion was effectively constant, 40.4 ± 0.4 kJ/mol. These values far exceed the values observed for other peptides, which typically fall below 10 kJ/mol.[77] However, the peptides with activation energies of diffusion < 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.[173,174]

In other works, we have concluded that the activation energy of peptide diffusion corresponds to the characteristic energy barrier associated with diffusion through a rough energy landscape. In oligoproline helices, these landscapes have been thoroughly examined by molecular dynamics simulations, making our oligoproline-based samples an ideal test of this conclusion.[174] These studies indicate that the conformational energy landscape of oligoproline contains a significant number of energy wells, each corresponding to a stable point between pure PPI and pure PPII secondary structure. The observed energy barrier for transitions between these wells was approximately 40 kJ/mol, indicating very close agreement with our observations.[174] Thus, our observed activation behavior is consistent with the picture of rigid oligoproline helices, whose only motions are associated with the gain or loss of PPII structural character. The strong agreement between our data and simulations also supports the connection between the activation energy of D_R and the characteristic roughness of the energy landscape.

5.6 Discussion

While serious consideration has been dedicated to pinpointing global optima in rugged χ^2 landscapes, we have seen surprisingly few works that prioritize the self-consistency of these fits over the location of an absolute optimum. In this manuscript, we implement methods to self-consistently place an optimum within a rugged cluster of local minima. Furthermore we identify continuous-wave EPR as an experimental area where these methods can provide significant and valuable insights into the physics of a system, which would elude analyses based on the global best fit. We used cluster-based methods that select the medoid, the geometric median or the marginal median to demonstrate that diffusional motion within oligoproline linkers is a thermally activated (Arrhenius) process, and we demonstrated that, as we had previously hypothesized[77], the activation energy matches the characteristic energy barrier associated with conformational change in oligoproline helices.[174] Naturally, these techniques will not apply for all curve-fitting analyses, but in our experience based on EPR analysis

of these and more than 20 other spin-labeled peptides, these methods are extremely reliable for fitting continuous-wave EPR spectra collected in the X-band. In concert with rapid protocols for peptide synthesis[76,77], this methodology provides a robust new approach for the experimental analysis of conformational energy landscapes in peptides. By employing a medoid-based approach, we can assure that good fits are both 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 CSCA software toolkit available publically, we hope that this approach can be adapted to suit the needs of a variety of experimenters. Moreover, we hope that our analysis will contribute to the growing body of evidence supporting the need for nuance and open-mindedness with respect to non-linear fitting of data.[169] In a mathematical sense, the goal of optimization is clear – identification of the point within a defined region which minimizes or maximizes a function. However, in a scientific context, we should never forget to prioritize reproducibility and meaning over mathematical idealism. By developing methods to balance these priorities, we achieved new insights into the physics of an important biomaterial system. We hope that our approach will allow others to achieve the same.

Chapter 6

Perspectives on the Role of Dynamics in Biomaterials

For the past 50 years, the concept that 'structure dictates function' has informed our understanding of molecular biology. This idea is central to our understanding of how life evolved on the earth, and is important for developing potent drugs, therapeutics and biomaterials. However, if I were to summarize my graduate work in one, simple phrase, it would be this: structure alone isn't enough. Although structure always plays a major role in the behavior of biopolymers in vivo, we must remember that evolution operates without rational principles of any kind. Life simply chooses methods that work.

In this work, I provide evidence for a connection between conformational dynamics and function within intrinsically disordered peptide sequences. I further demonstrate that this connection can be decoupled from permanent structural changes within the material, suggesting transient configurations also play a key role in the function of disordered peptides.

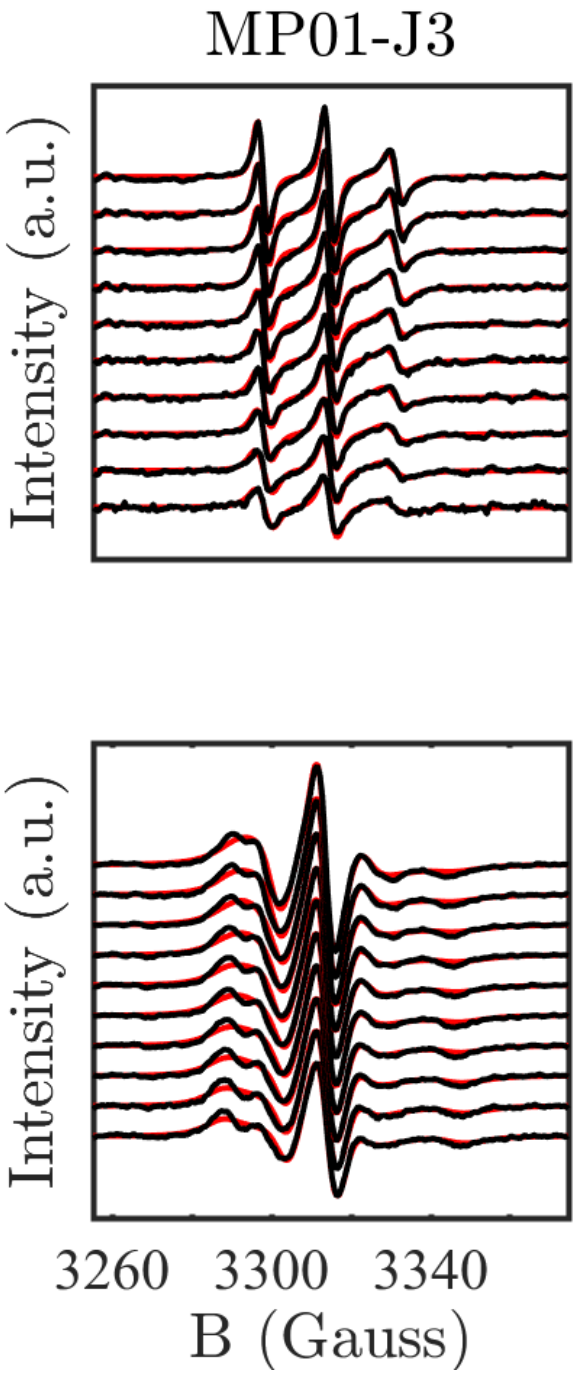
Conformationally unstable biopolymers are ubiquitous within the natural world, where they play many important, functional roles.[2,47] In my view, it is probable that many such proteins, the rate of structural transformation is an important parameter governing protein activity.

Thus, the continued study of conformational dynamics remains an important ob-

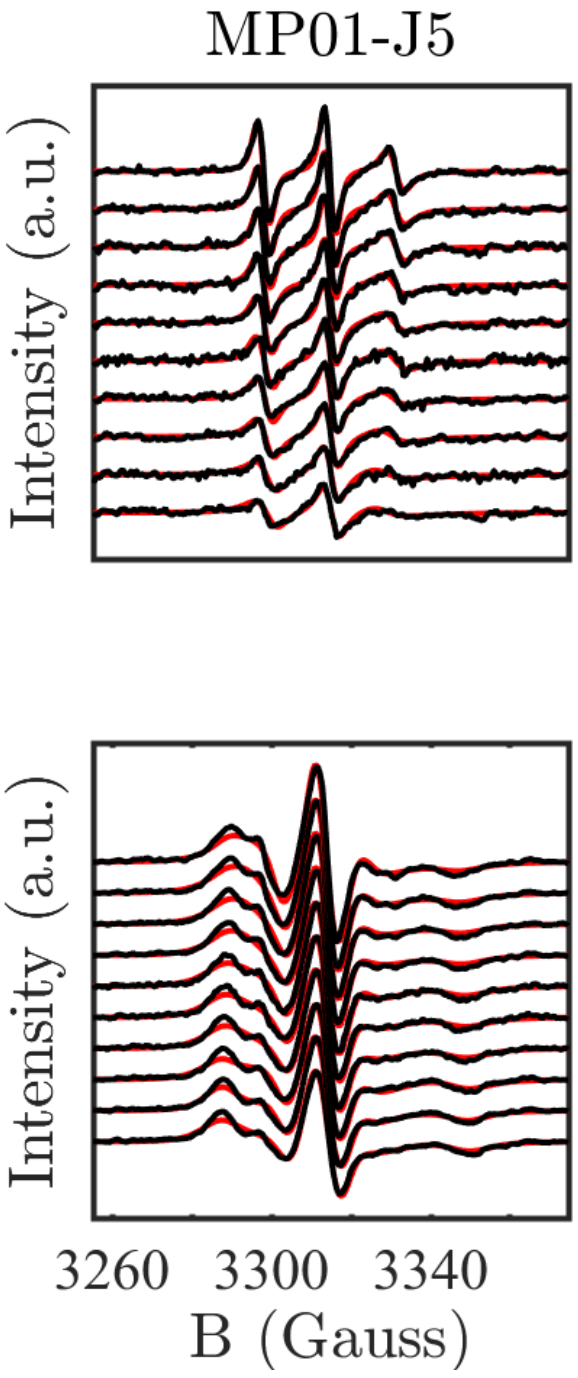
jective for the design of better biomaterials. For medically important biopolymers such as antimicrobial peptides[58], peptide-based vaccines[177], cell-penetrating peptides[178], and nuclear localization signaling peptides[179], the role of conformational dynamics remains a signficant unknown. However, the generality of our approch for synthesizing and characterizing spin-labeled peptides provides a path to answer these questions, and to expand our capacity to rationally engineer biomedical technologies.

Appendix A

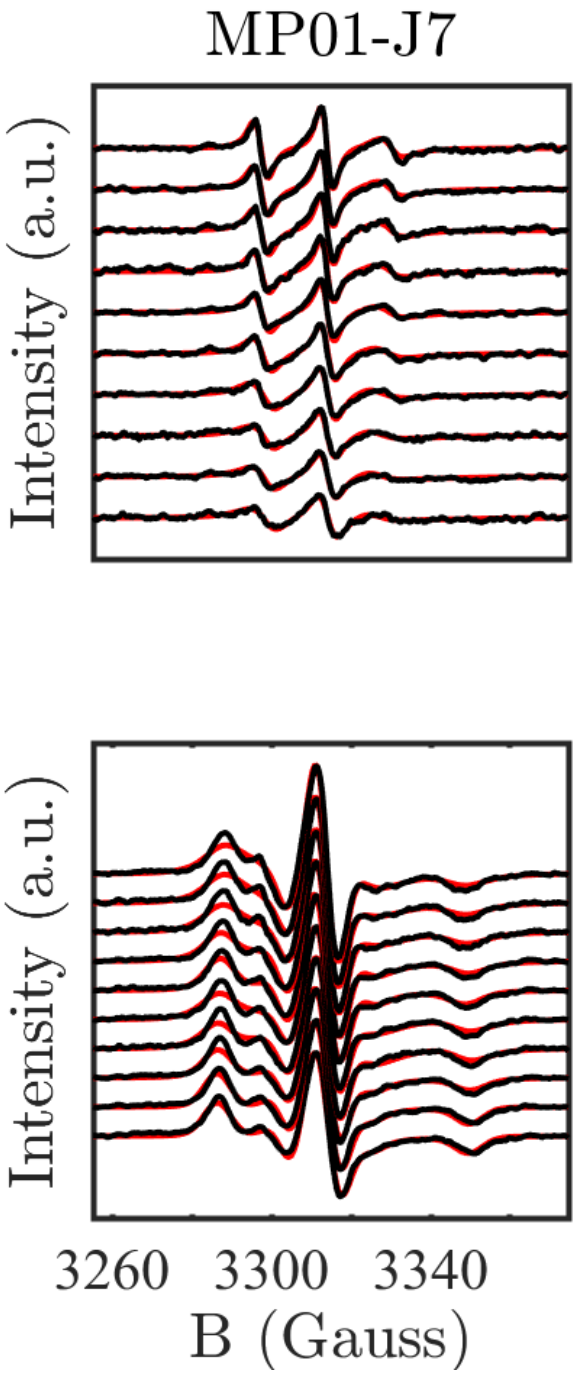
EPR fits for MP01-Gen4 spectra



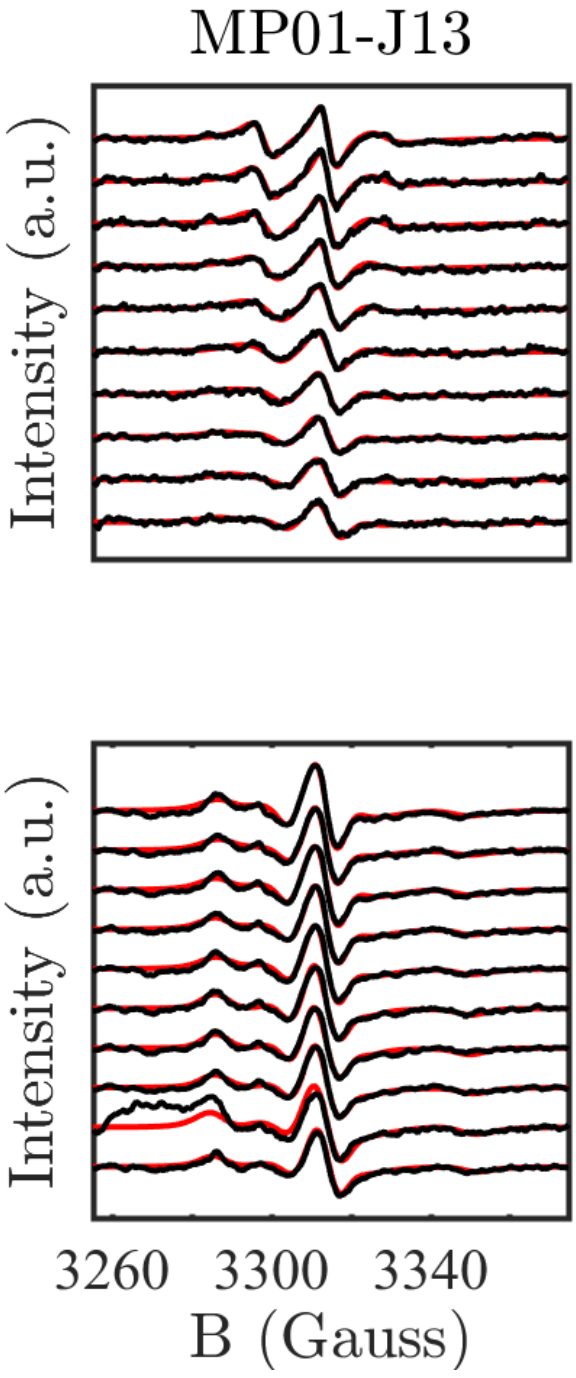
EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J3



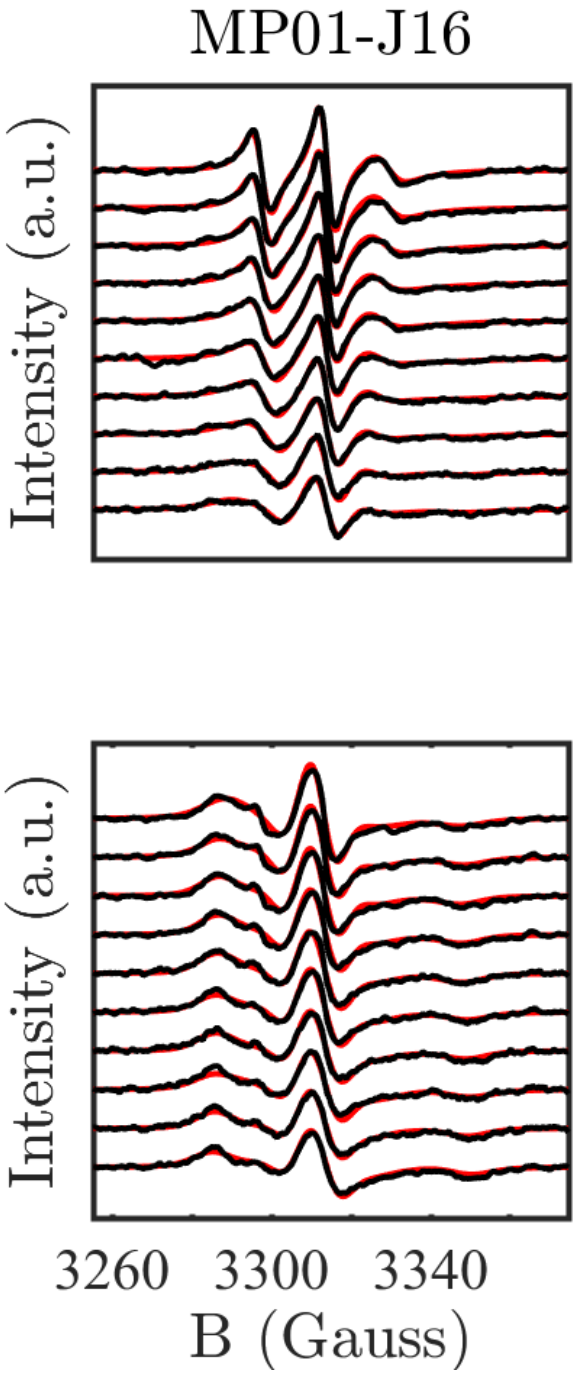
EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J5



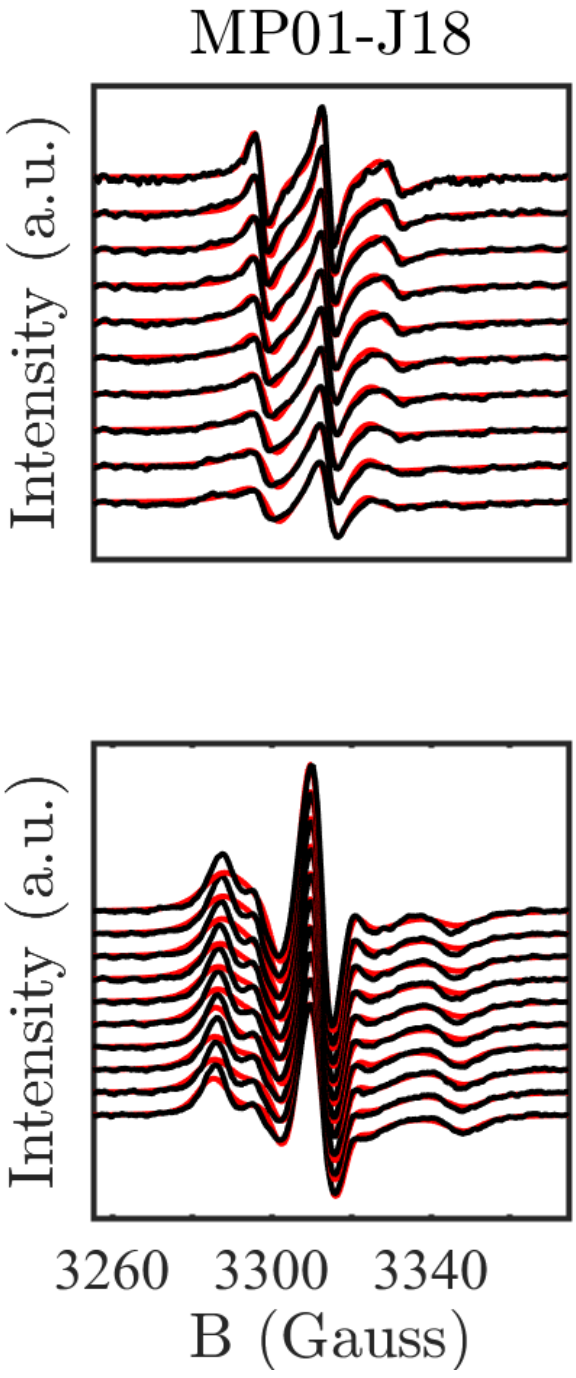
EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J7



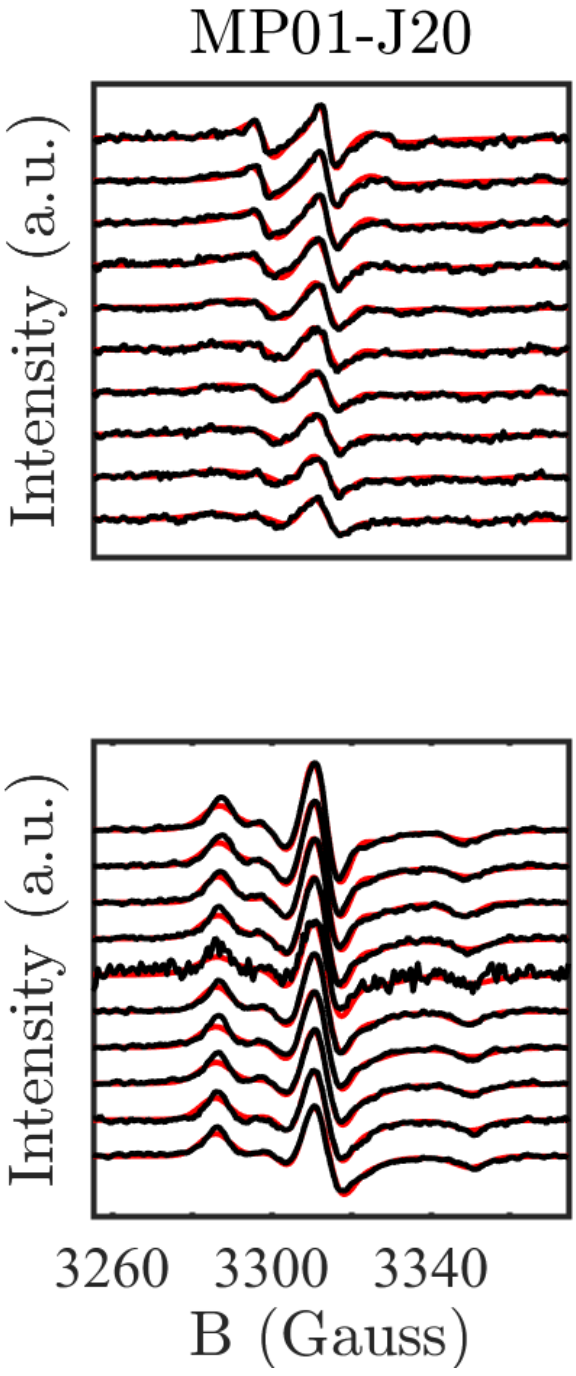
EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J13



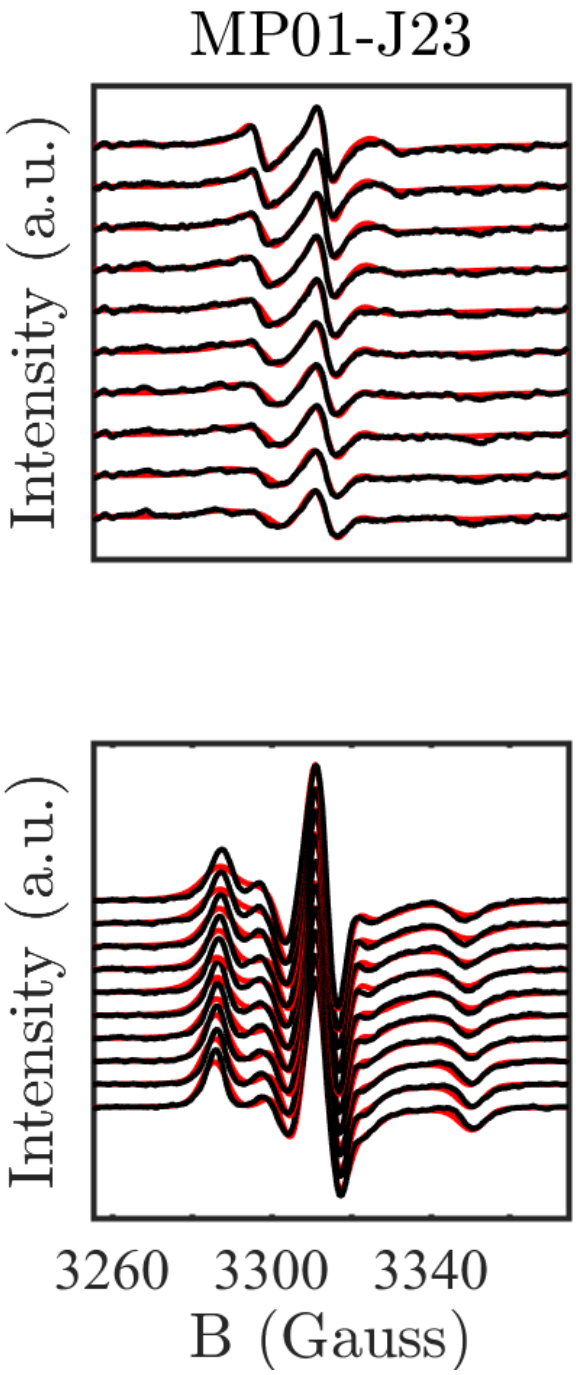
EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J16



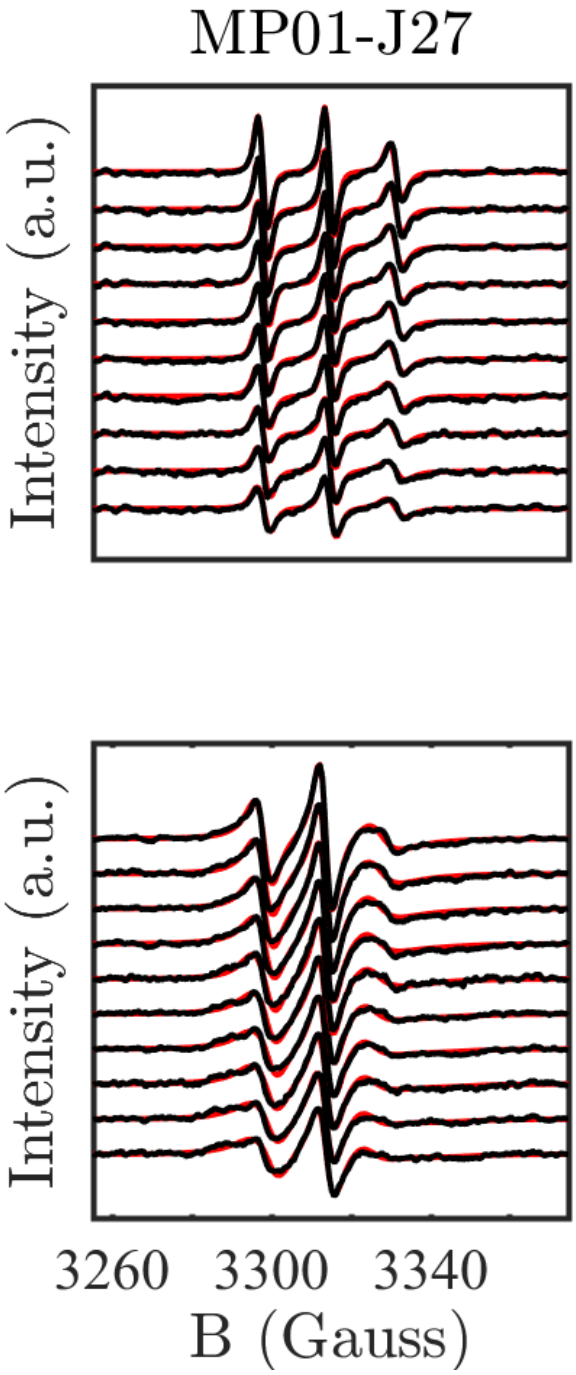
EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J18



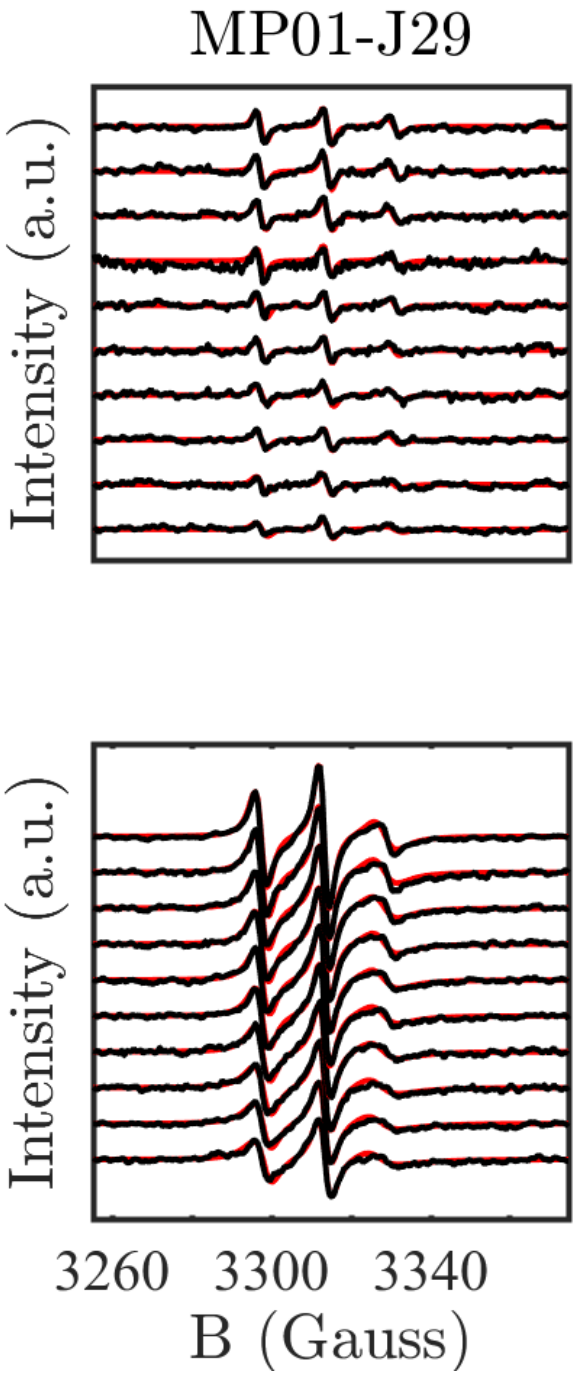
EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J20



EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J23



EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J27

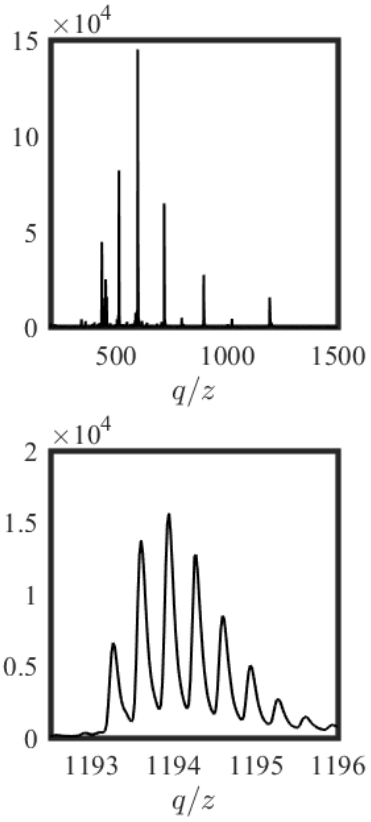


EPR spectra and best fits for unreacted (top) and reacted (bottom) MP01-J29. The unreacted peptides have very low spectral intensity because they were not reacted with $K_3Fe(CN)_6$, since LC-MS demonstrated that this damaged unreacted MP01-J29. Consequently, the bulk of the MP01-J29 existed in the spectrally inert hydroxlyaminated form, giving very low-intensity spectra. Consequently, noise heavily affected these fits, resulting in a wide range of fitting error.

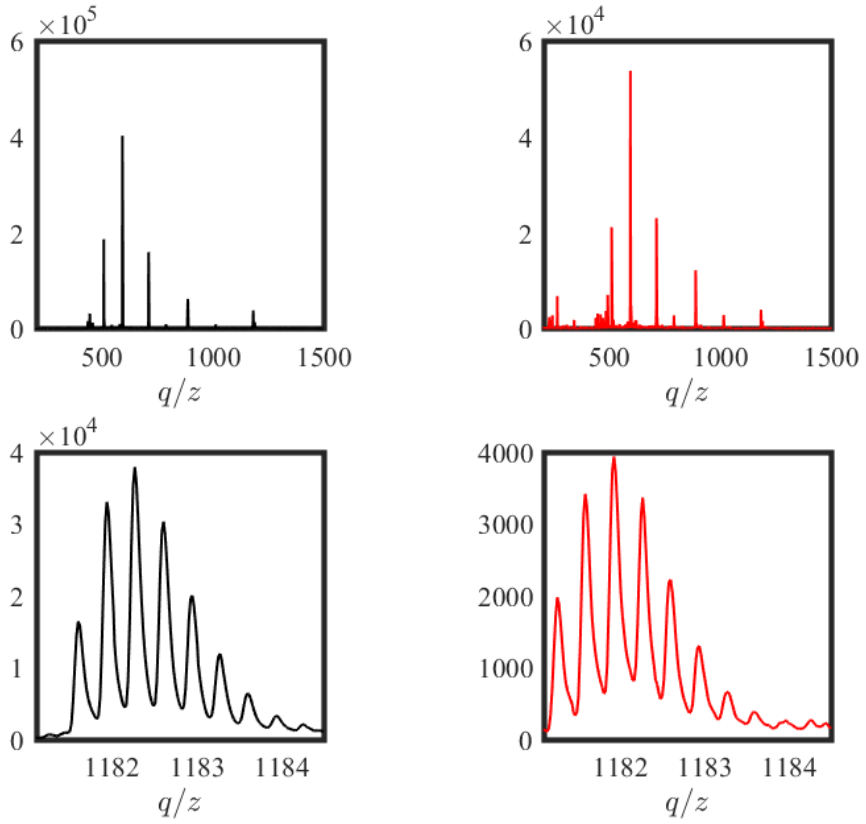
Appendix B

LC-MS spectra for MP01-Gen4 peptides

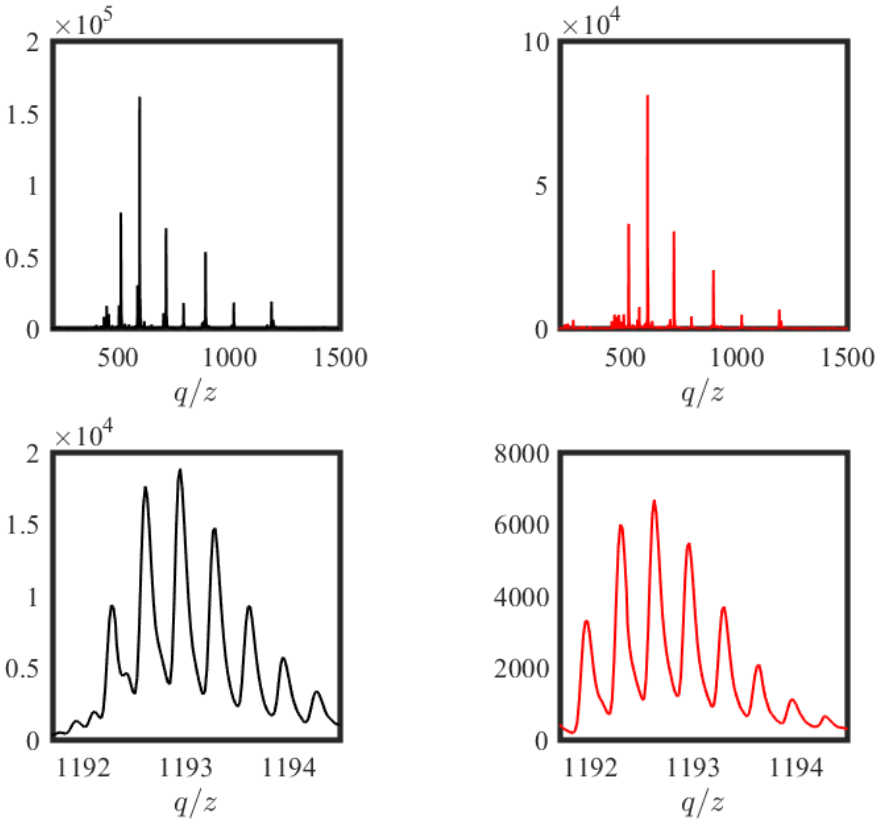
B.1 Before S_NAr labeling



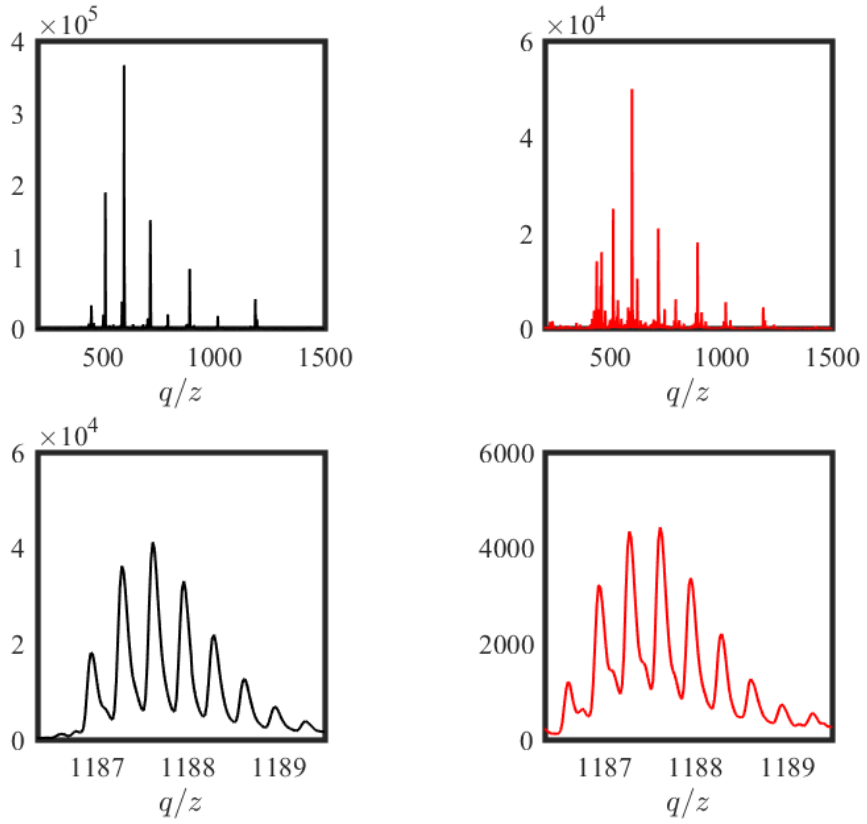
The LC-MS spectrum of major peaks of the MP01-J3 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. For predicted and calculated masses, refer to Table 3.2.



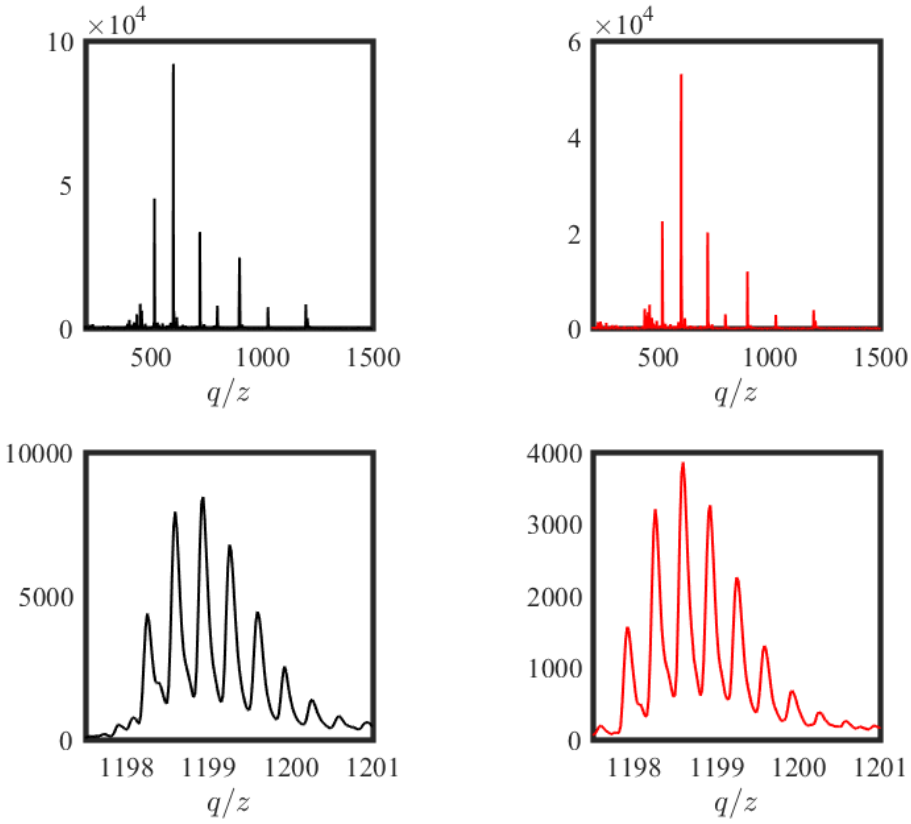
The LC-MS spectrum of major peaks of the MP01-J5 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2.



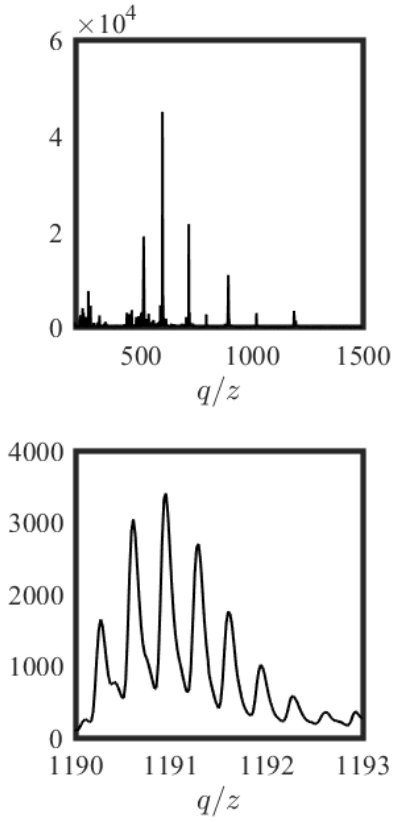
The LC-MS spectrum of major peaks of the MP01-J7 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2.



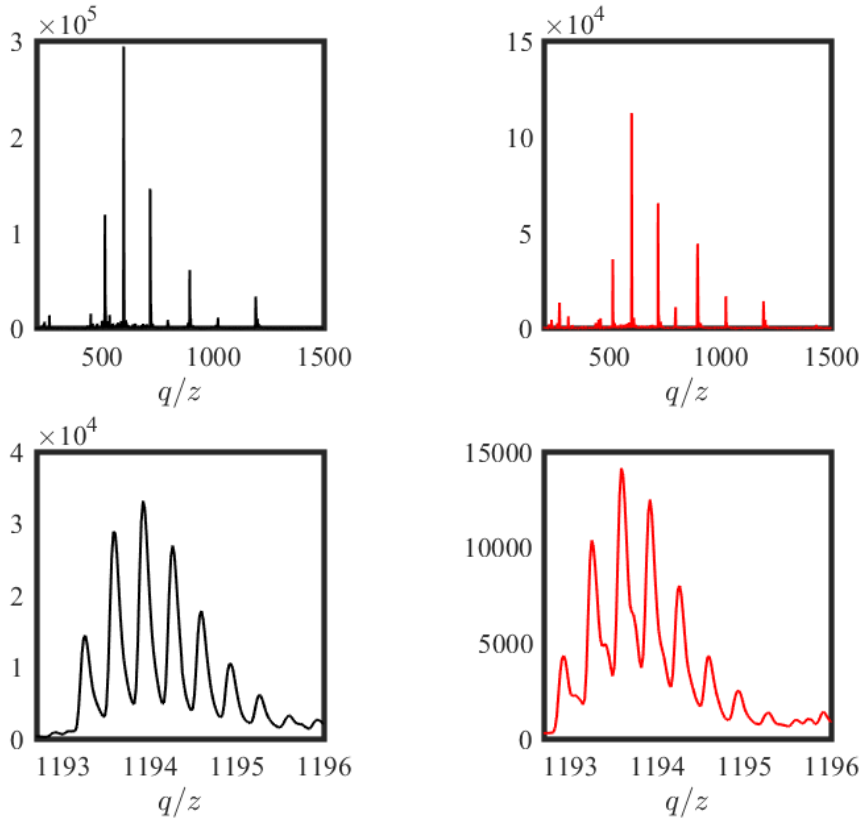
The LC-MS spectrum of major peaks of the MP01-J13 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2.



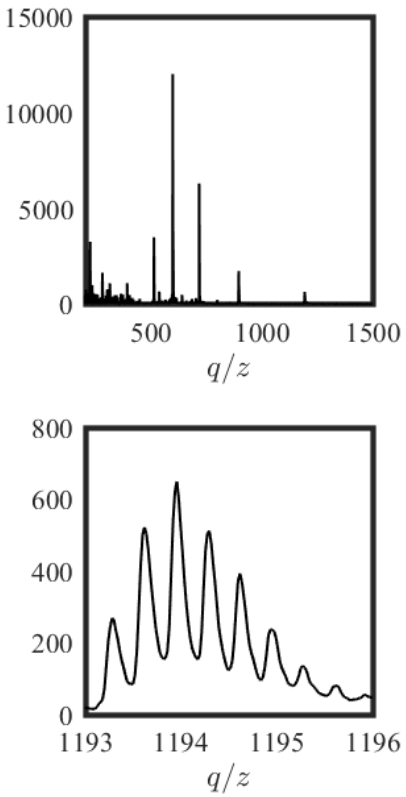
The LC-MS spectrum of major peaks of the MP01-J16 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2. Smaller peaks in these spectra originate from a minor, dimerized peptide.



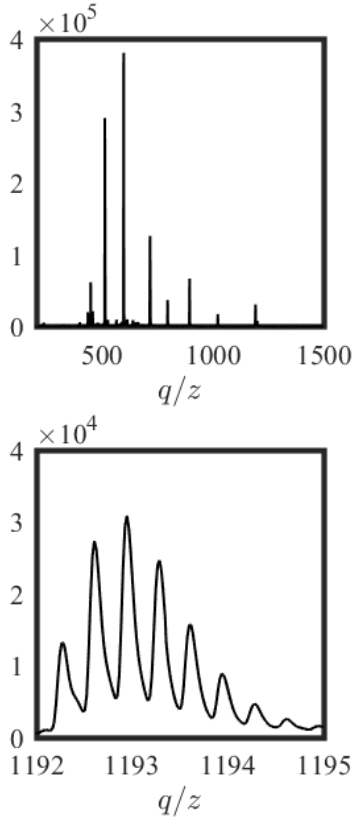
The LC-MS spectrum of major peaks of the MP01-J18 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. For predicted and calculated masses, refer to Table 3.2.



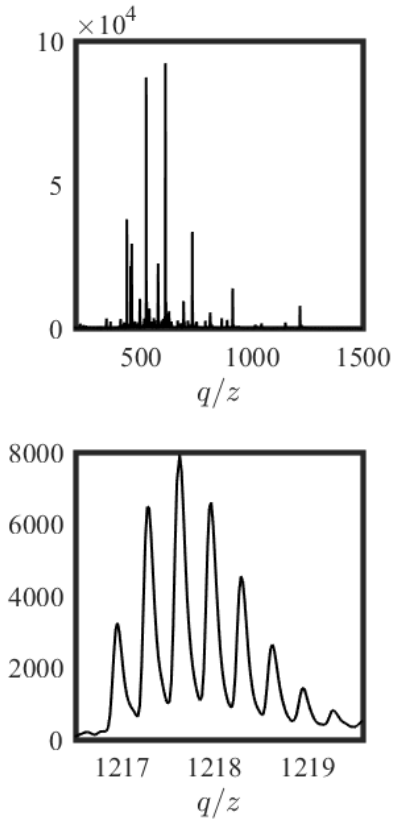
The LC-MS spectrum of major peaks of the MP01-J20 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2. Smaller peaks in the hydroxylamine signal originate from a minor, dimerized peptide.



The LC-MS spectrum of major peaks of the MP01-J23 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. For predicted and calculated masses, refer to Table 3.2.

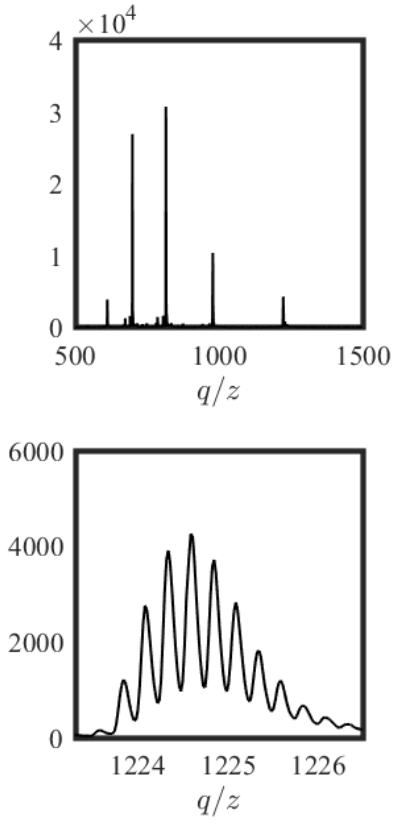


The LC-MS spectrum of major peaks of the MP01-J27 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. For predicted and calculated masses, refer to Table 3.2. Smaller peaks in the hydroxylamine spectrum originate from a minor, dimerized peptide.

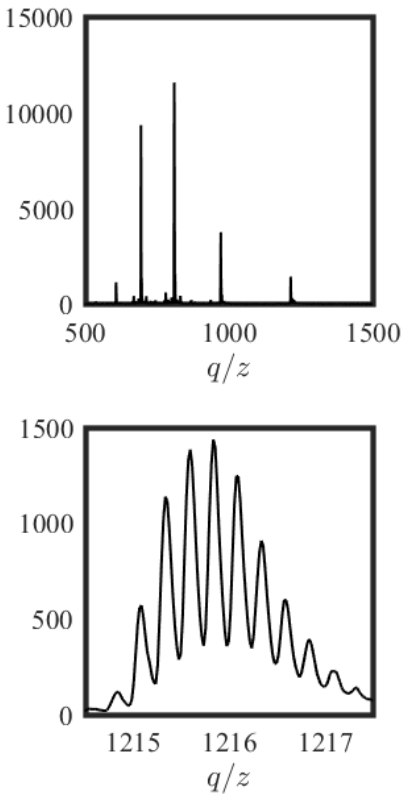


The LC-MS spectrum of major peaks of the MP01-J29 chromatography curve. A feature of the $[M+3H]^{3+}$ charge state used for mass calculation is shown beneath. For predicted and calculated masses, refer to Table 3.2. The dominant set of impurity peaks in the hydroxylamine spectrum originate from a TOAC deletion product, which is EPR-invisible.

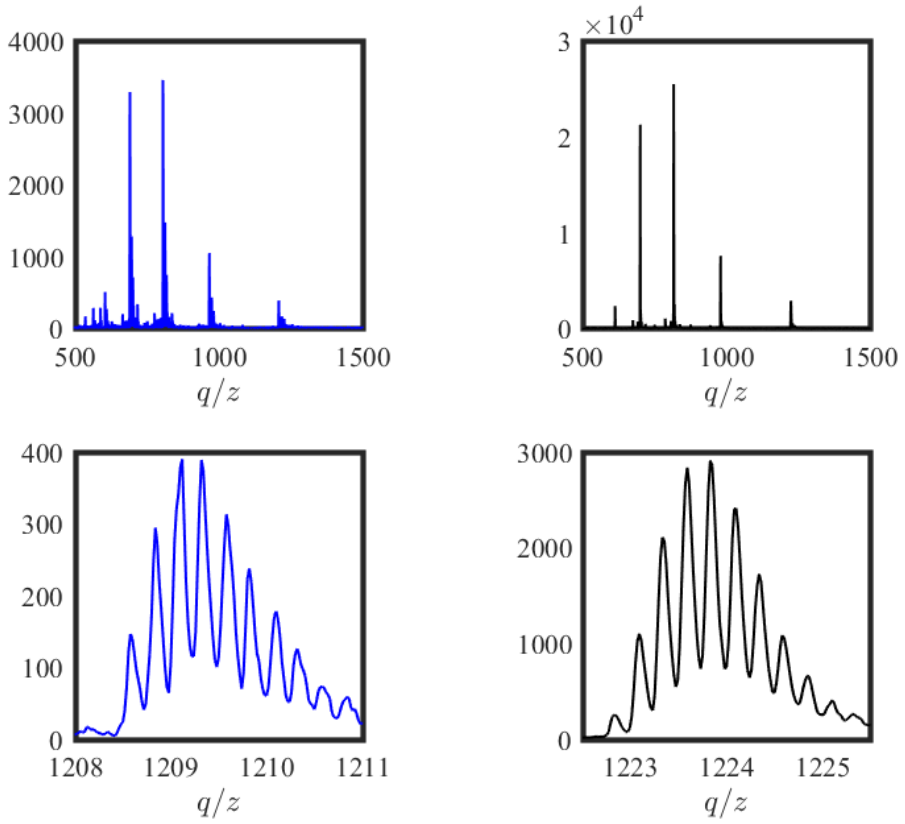
B.2 After S_NAr labeling



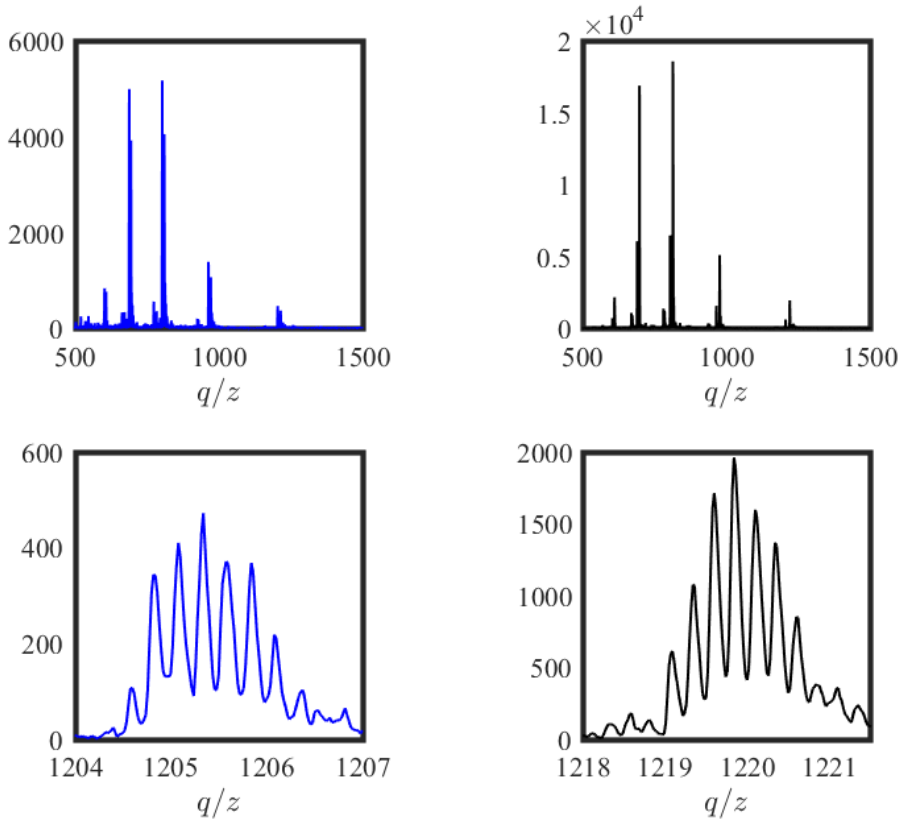
The LC-MS spectrum of major peaks of the labeled MP01-J3 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. For predicted and calculated masses, refer to Table 3.2.



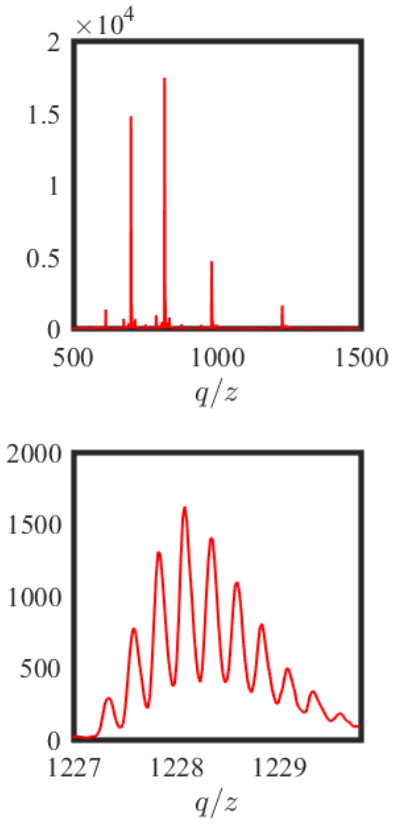
The LC-MS spectrum of major peaks of the labeled MP01-J3 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. For predicted and calculated masses, refer to Table 3.2.



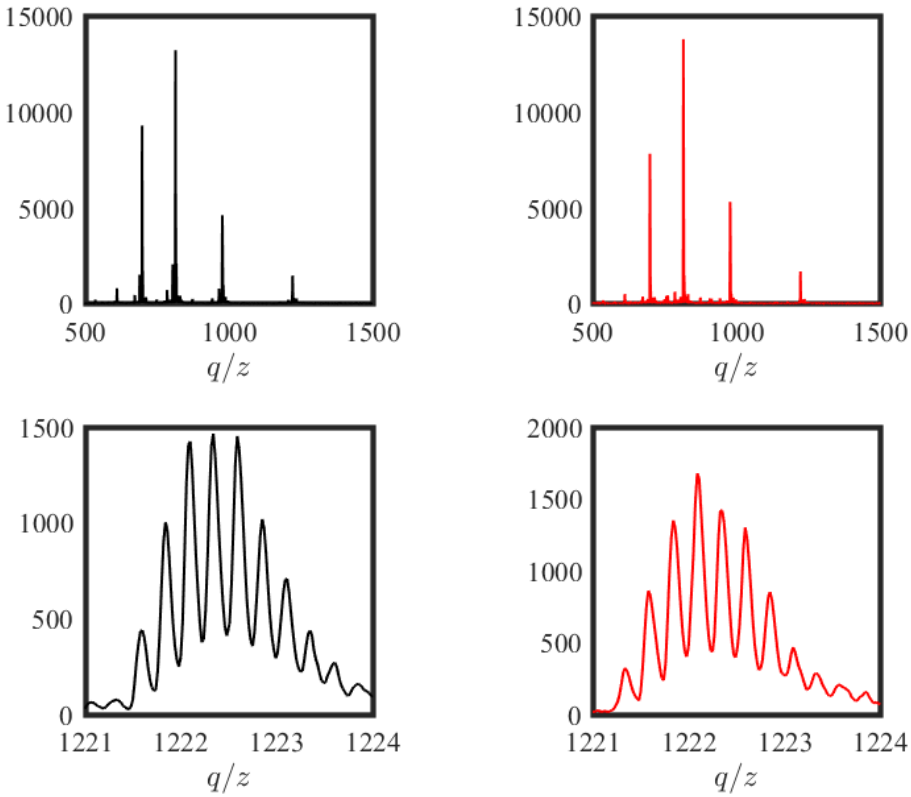
The LC-MS spectrum of major peaks of the labeled MP01-J7 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black and the (minor) nitroxylated glycine deletion product is shown in blue. For predicted and calculated masses, refer to Table 3.2.



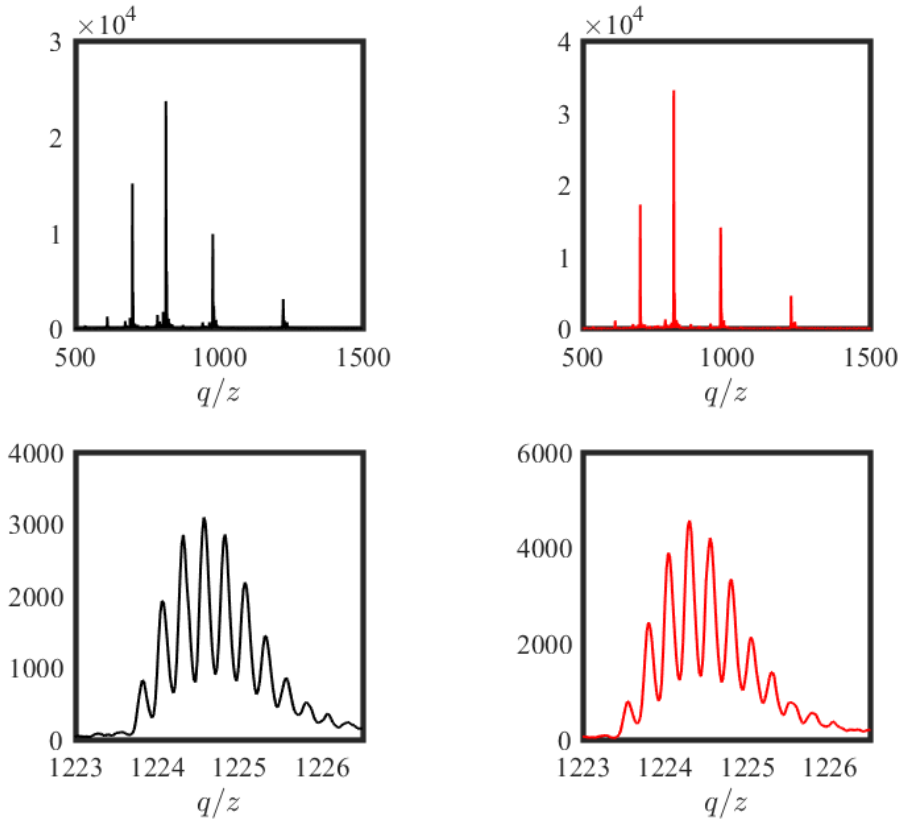
The LC-MS spectrum of major peaks of the labeled MP01-J13 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black and the (minor) nitroxylated glycine deletion product is shown in blue. For predicted and calculated masses, refer to Table 3.2.



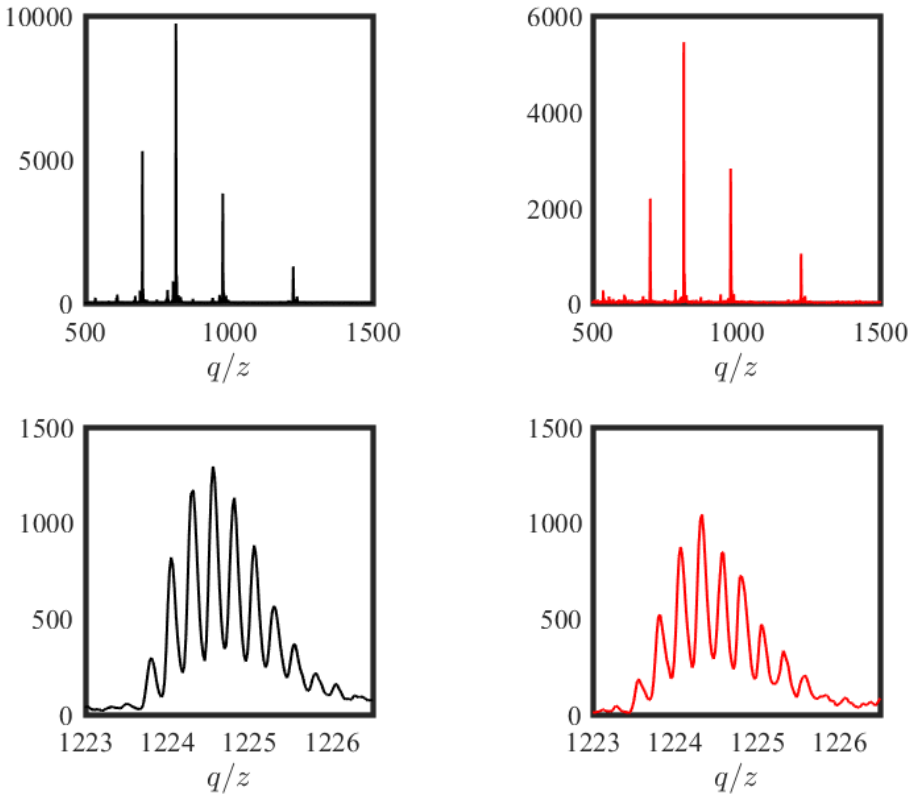
The LC-MS spectrum of major peaks of the labeled MP01-J13 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black and the (minor) nitroxylated glycine deletion product is shown in blue. For predicted and calculated masses, refer to Table 3.2.



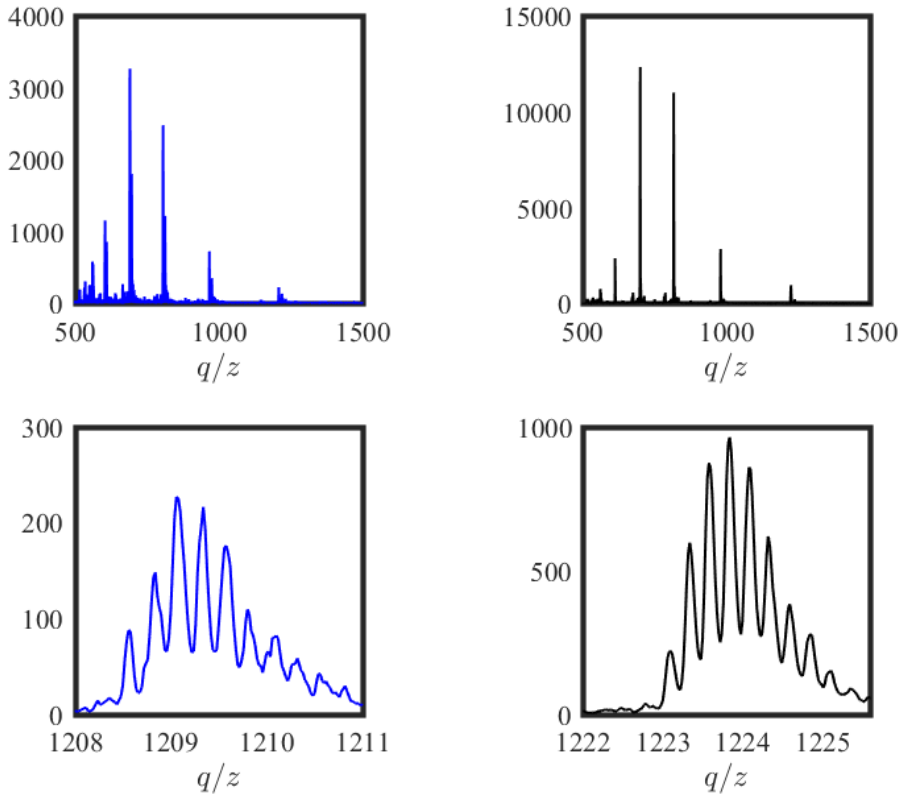
The LC-MS spectrum of major peaks of the labeled MP01-J18 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2.



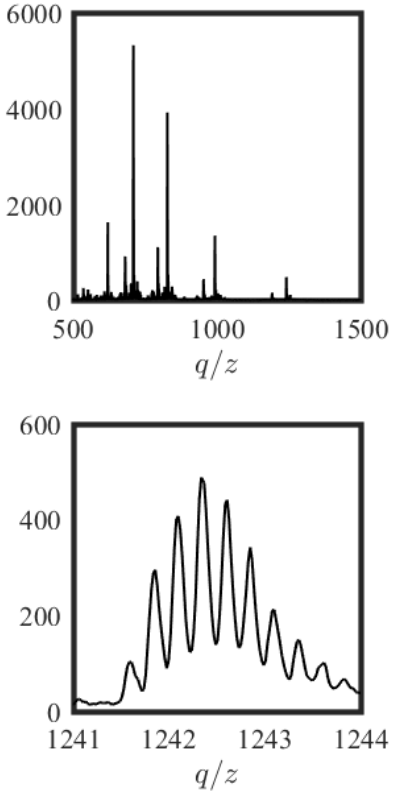
The LC-MS spectrum of major peaks of the labeled MP01-J20 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2.



The LC-MS spectrum of major peaks of the labeled MP01-J23 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black, and the (minor) nitroxide product is shown in red. For predicted and calculated masses, refer to Table 3.2.



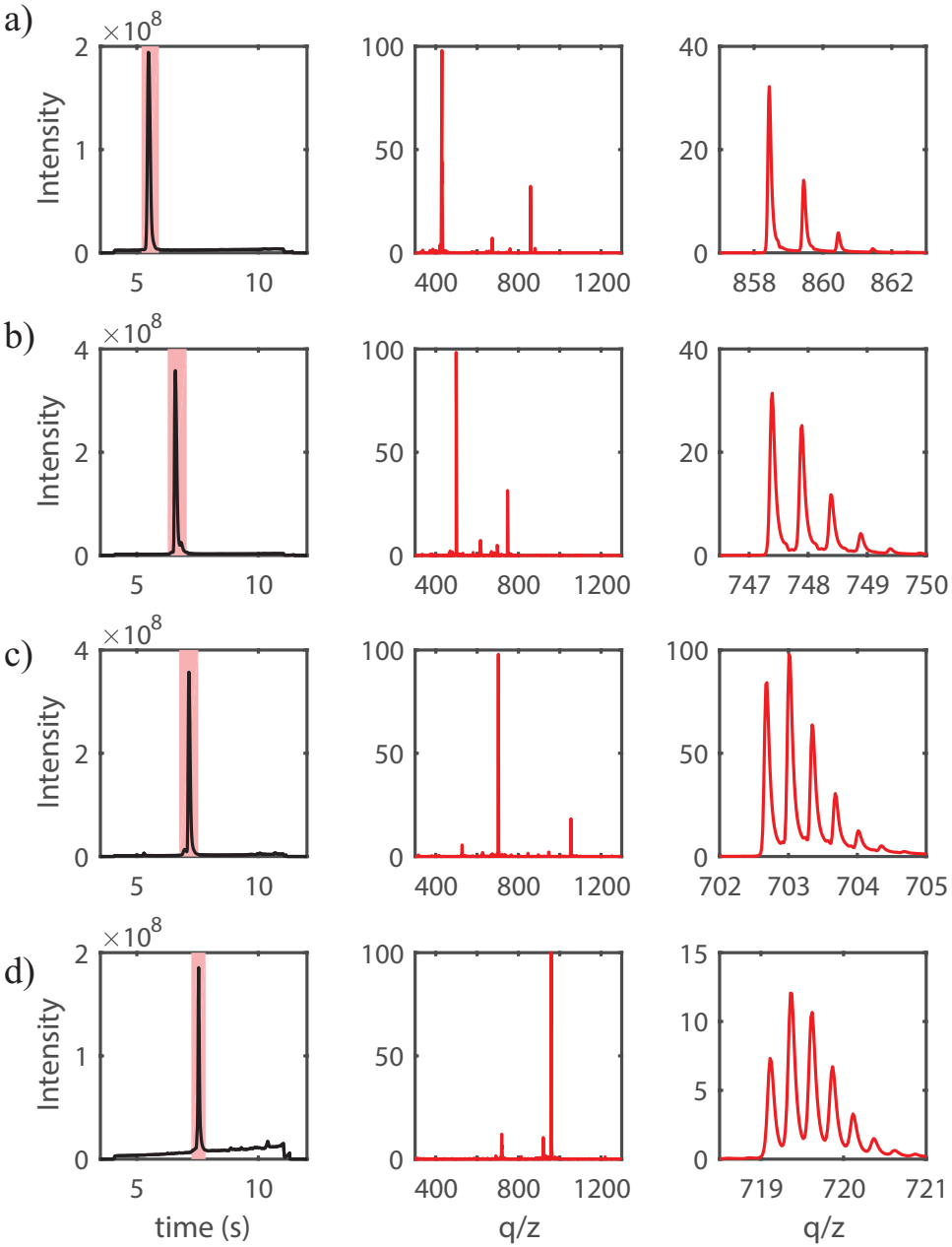
The LC-MS spectrum of major peaks of the labeled MP01-J27 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The (major) hydroxylamine product is shown in black and the (minor) nitroxylated glycine deletion product is shown in blue. For predicted and calculated masses, refer to Table 3.2.



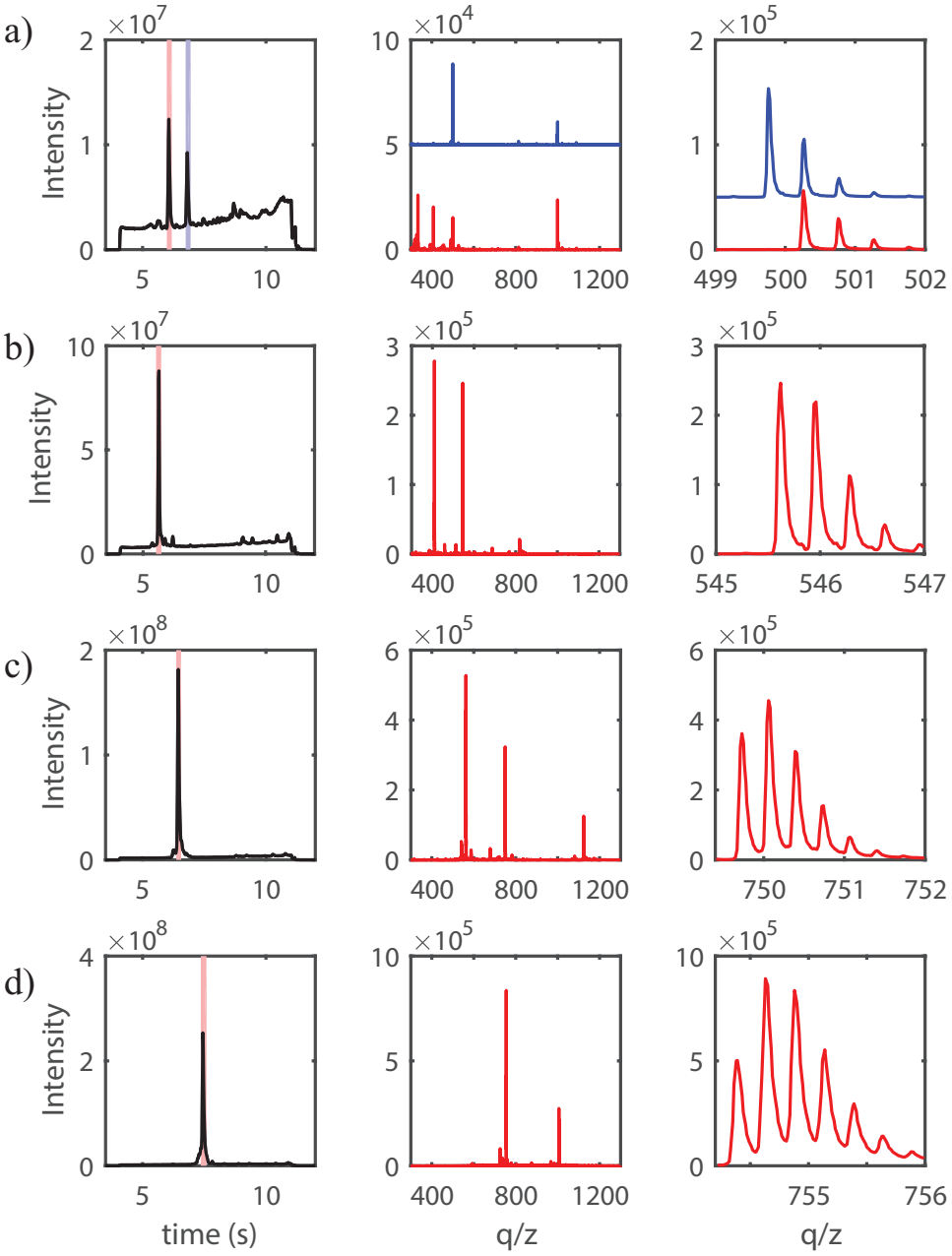
The LC-MS spectrum of major peaks of the labeled MP01-J29 chromatography curve. A feature of the $[M+4H]^{4+}$ charge state used for mass calculation is shown beneath. The smaller set of peaks visible in the signal originate from the TOAC deletion product, to which EPR measurement is insensitive. For predicted and calculated masses, refer to Table 3.2.

Appendix C

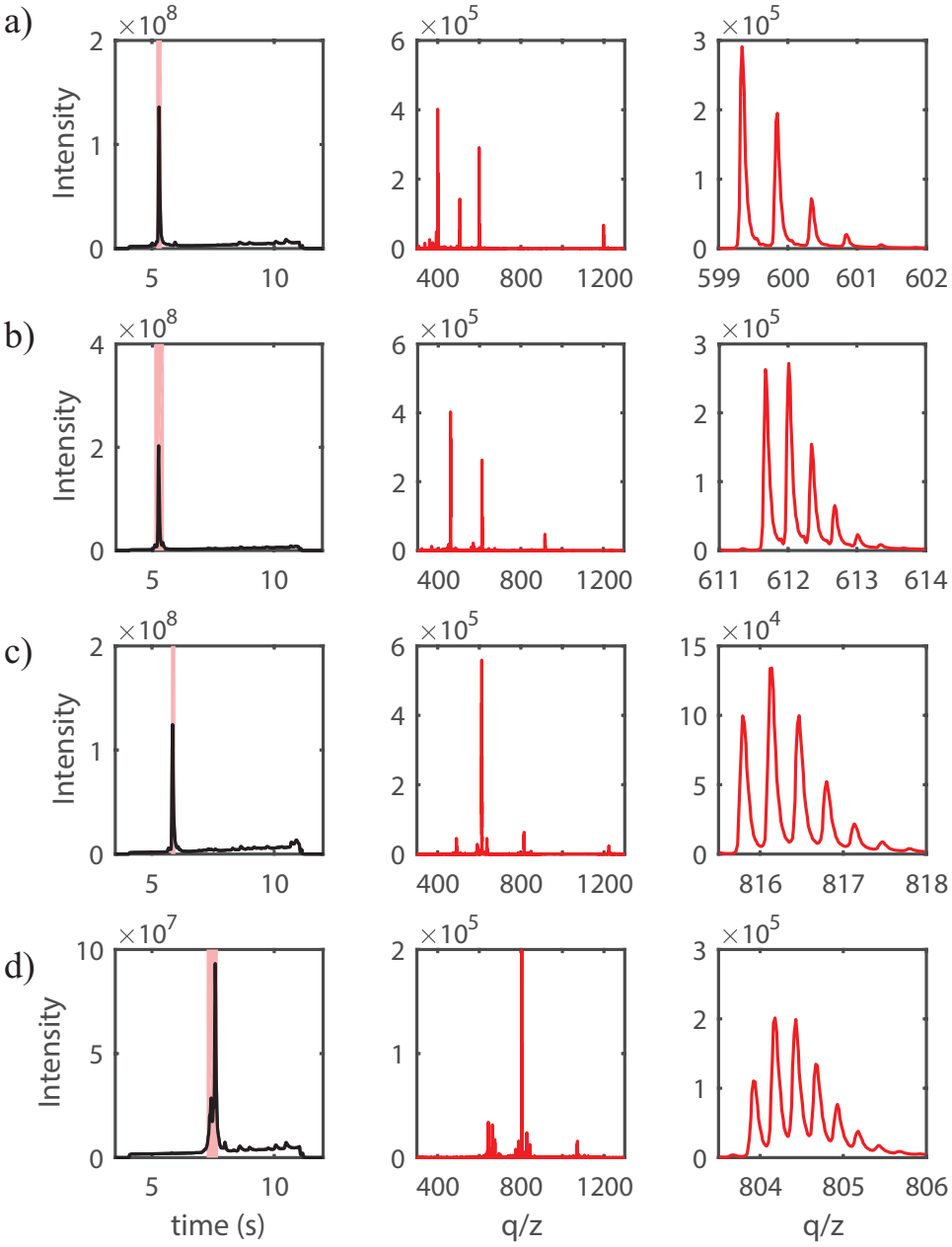
LC-MS spectra for FMP peptides



LC-MS elution profiles (left) and mass-to-charge (m/z) data integrated over the highlighted region (middle, right) for a) FMP9, b) FMP15, c) FMP21, and d) FMP27. The specific peaks used for mass determination were the FMP9 [M+H]1+ peak-set, the FMP15 $[M+2H]^{2+}$ peak-set, the FMP21 [M+3H]3+ peak-set and the FMP27 $[M+4H]^{4+}$ peak-set. These are shown in high-resolution on the right.



LC-MS elution profiles (left) and mass-to-charge (m/z) data integrated over the highlighted region (middle, right) for a) FMP9-J5, b) FMP15-J8, c) FMP21-J11, and d) FMP27-J14. The specific peaks used for mass determination were the FMP9-J5 $[M+2H]^{2+}$ peak-set, the FMP15-J8 $[M+3H]^{3+}$ peak-set, the FMP21-J11 $[M+3H]^{3+}$ peak-set and the FMP27-J14 $[M+4H]^{4+}$ peak-set. These are shown in high-resolution on the right. The observed masses typically deviate by 1 hydrogen from the mass of the TOAC peptides because TOACs reduce to hydroxylamines (+1 H atom) under the acidic conditions on the column. They were reoxidized during experiments, so this did not affect other measurements. In FMP9-J5 both forms were observed as separate peaks. The nitroxide peak is plotted/highlighted in blue, while the hydroxylamine peak is plotted/highlighted in red.

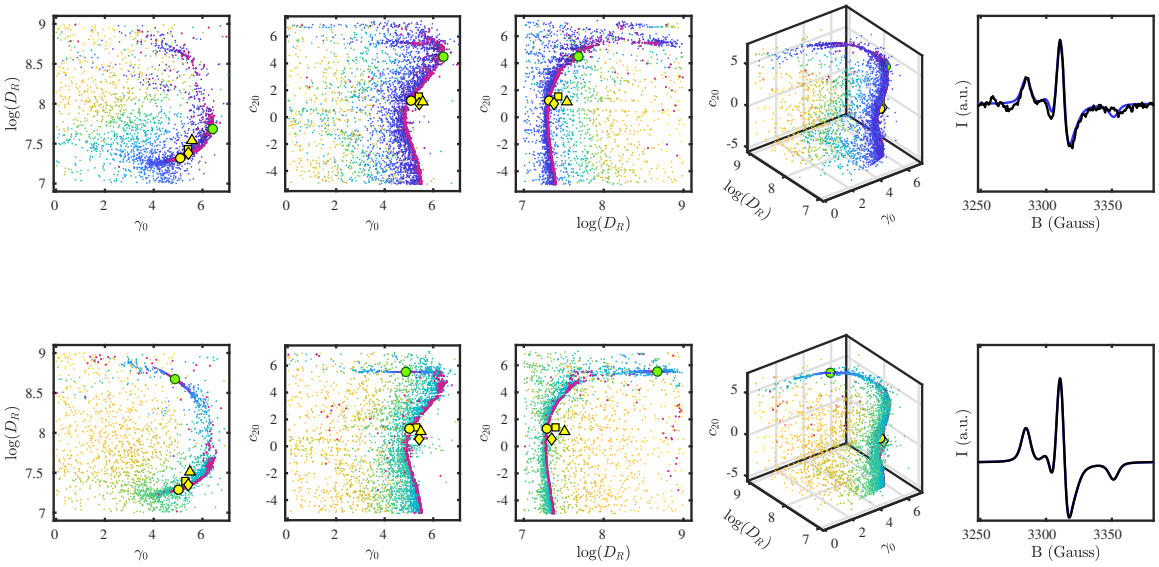


LC-MS elution profiles (left) and mass-to-charge (m/z) data integrated over the highlighted region (middle, right) for a) FMP9-J0J5, b) FMP15-J0J8, c) FMP21-J0J11, and d) FMP27-J0J14. The specific peaks used for mass determination were the FMP9-J0J5 $[M+2H]^{2+}$ peak-set, the FMP15-J0J8 $[M+3H]^{3+}$ peak-set, the FMP21-J0J11 $[M+3H]^{3+}$ peak-set and the FMP27-J0J14 $[M+4H]^{4+}$ peak-set. These are shown in high-resolution on the right. The observed masses deviate by 2 hydrogens from the mass of the TOAC peptides because TOACs reduce to hydroxylamines (+1 H atom) under the acidic conditions on the column. They were reoxidized during experiments, so this did not affect other measurements.

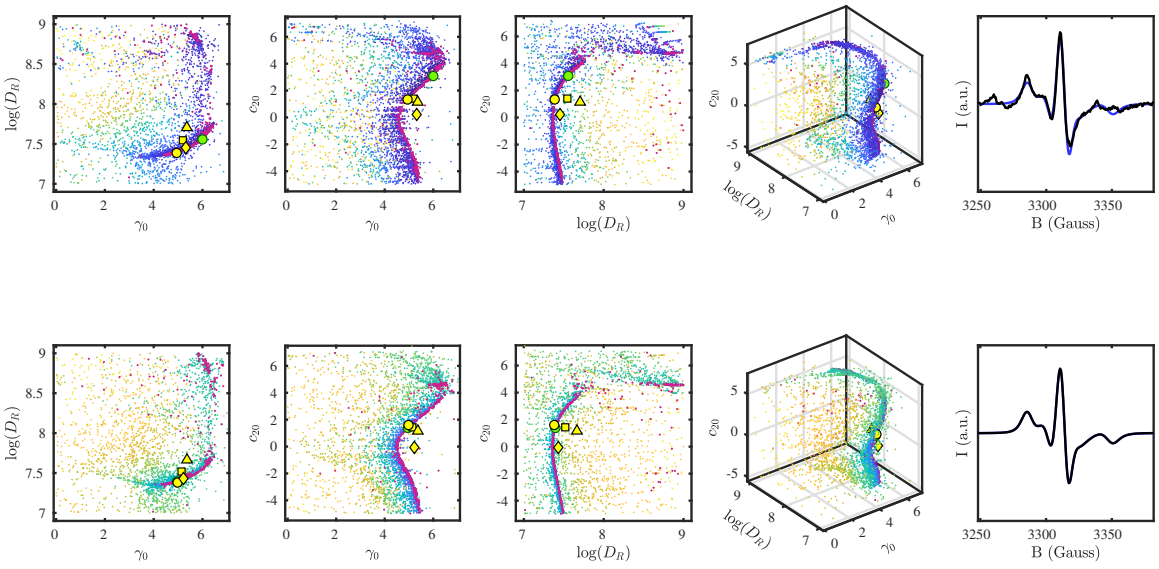
Appendix D

Calculated χ^2_{min} plots for AA-Pn-SL spectra

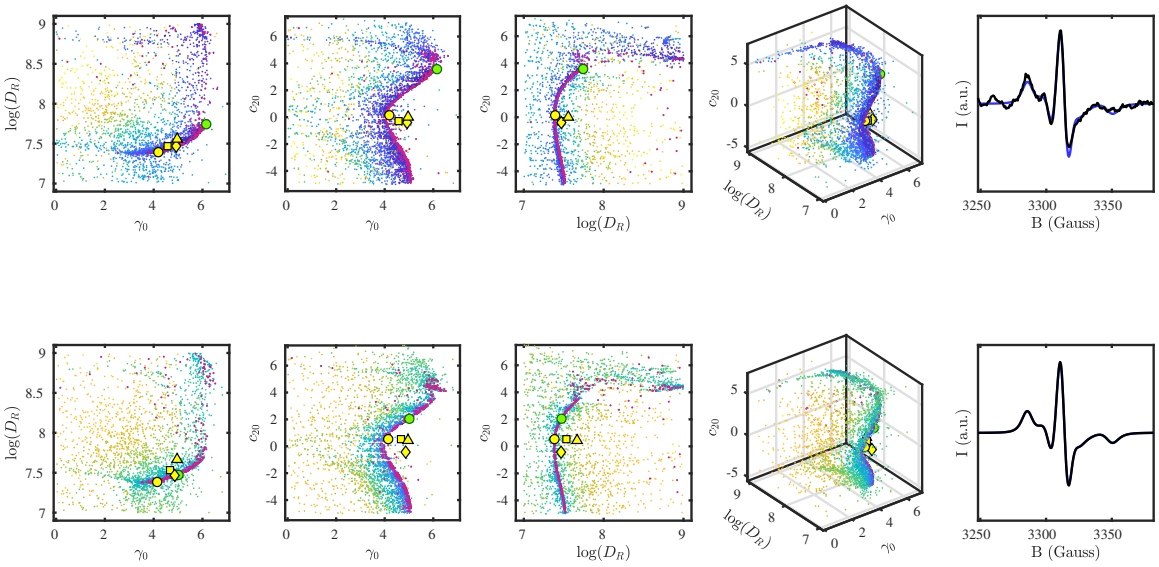
Below are the calculated χ^2 landscapes for each EPR spectrum we computed, along with fits to the simulated best-fit spectrum. These figures are analogous to Figure 5.6 of Chapter 5. They record the position of the global optimum, as well as the marginal median, the geometric median, the medoid and the mean the cluster of local minima with $\chi^2 < 2\chi^2_{min}$, but were observed to be broadly invariant when larger χ^2 thresholds were chosen. Fits to simulated spectra were analyzed within a different threshold ($\chi^2 < 2\chi^2_{min}$) because they had a different y-scaling and because the perfect quality of the data reduced the value of χ^2_{min} to near-zero. Thus, a larger threshold was needed to identify clusters large enough that measures of central tendency became invariant. Although means are reported for each plot, we broadly observed that they were inconsistent metrics, and we disregarded them for the purpose of Arrhenius analysis.



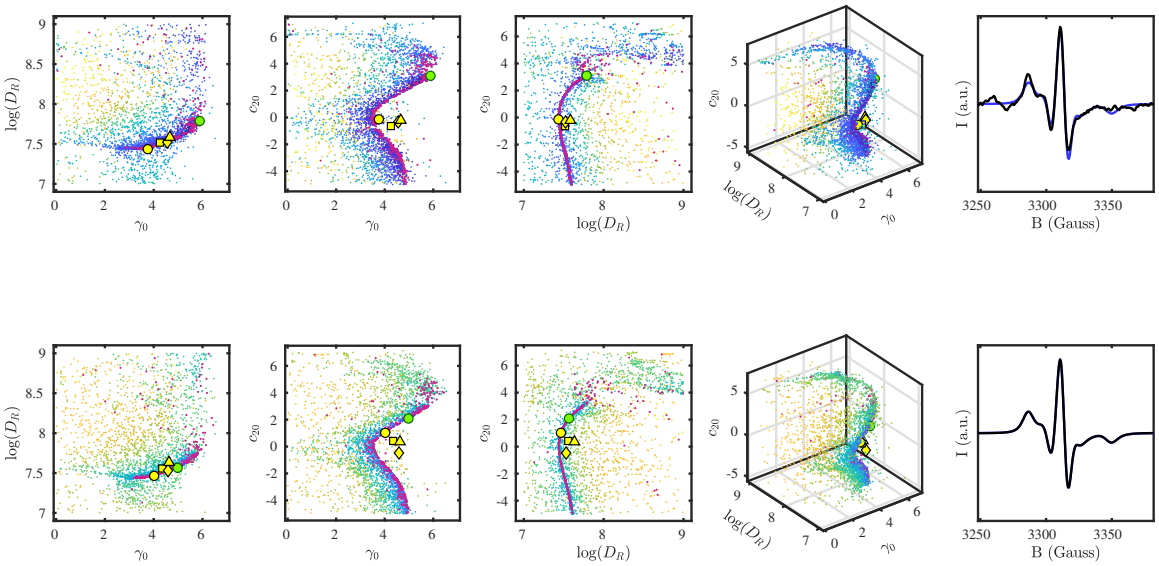
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 275K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



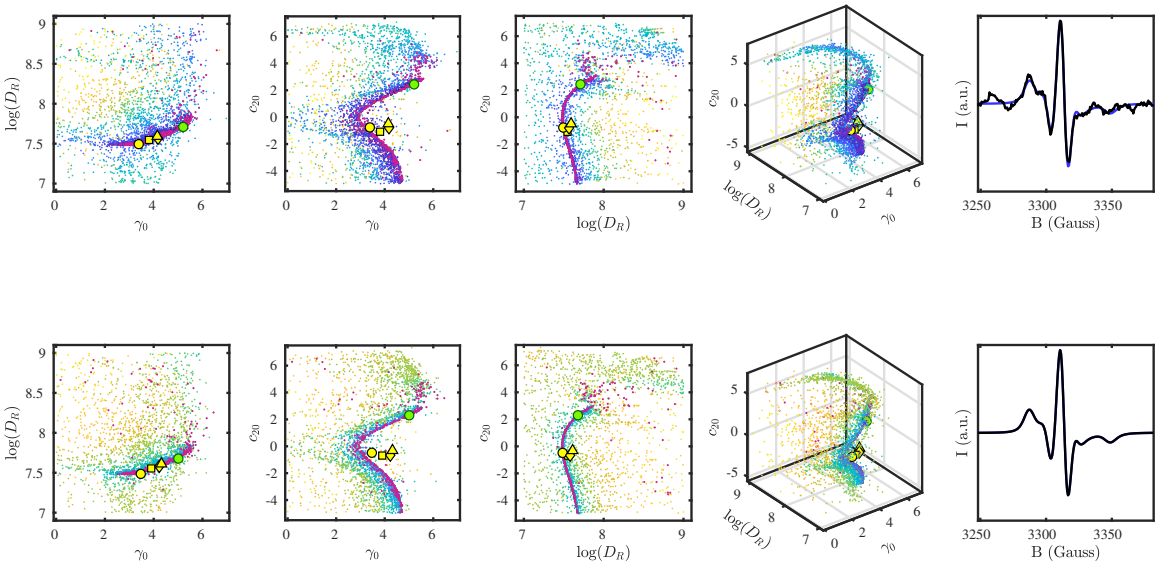
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 278K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



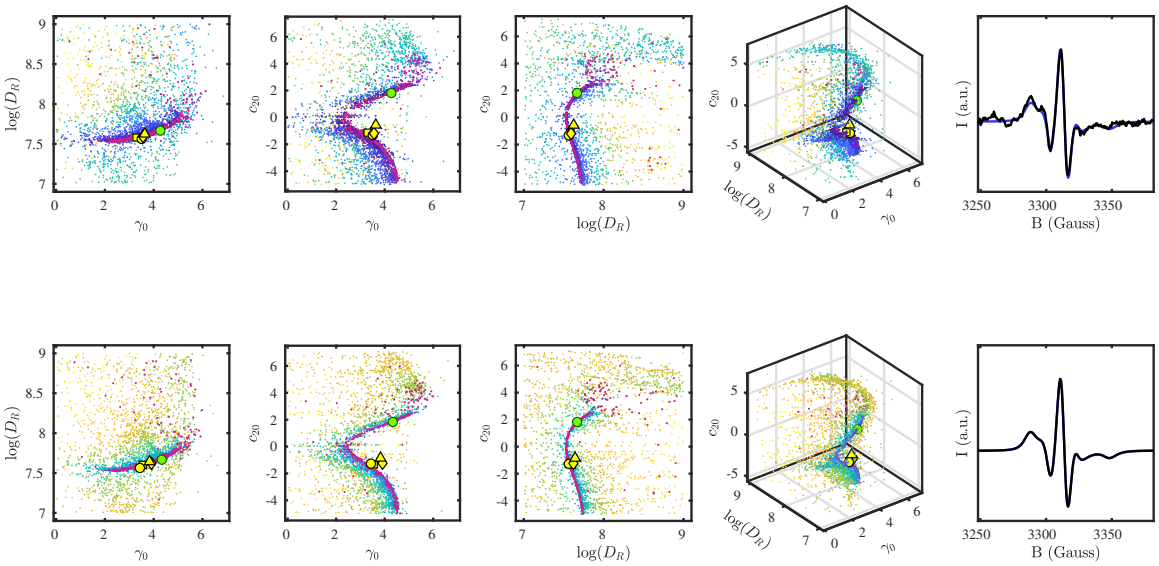
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 281K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



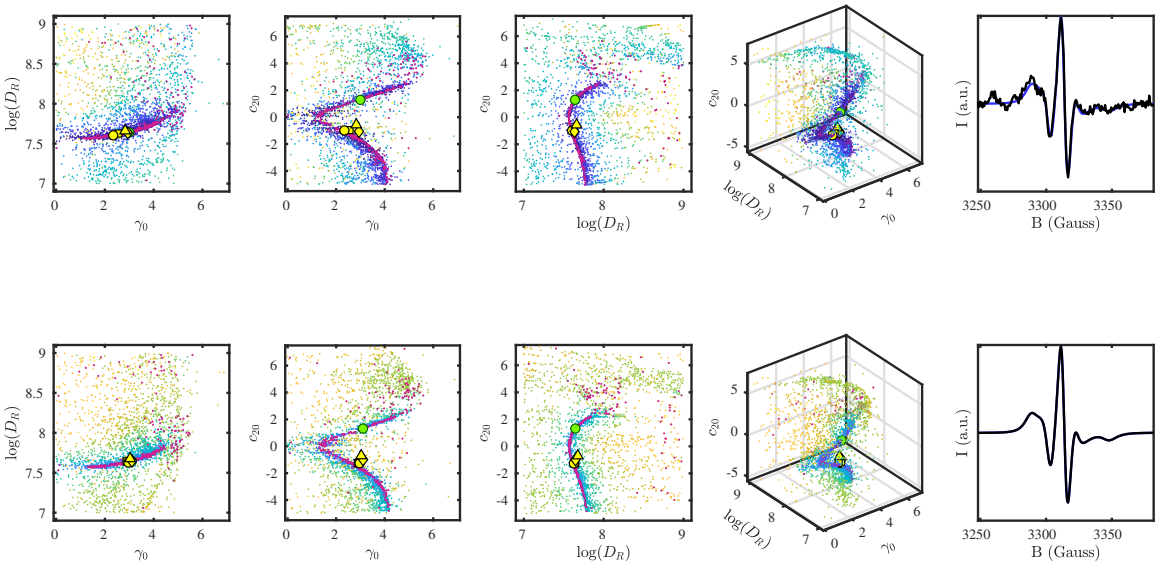
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 284K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



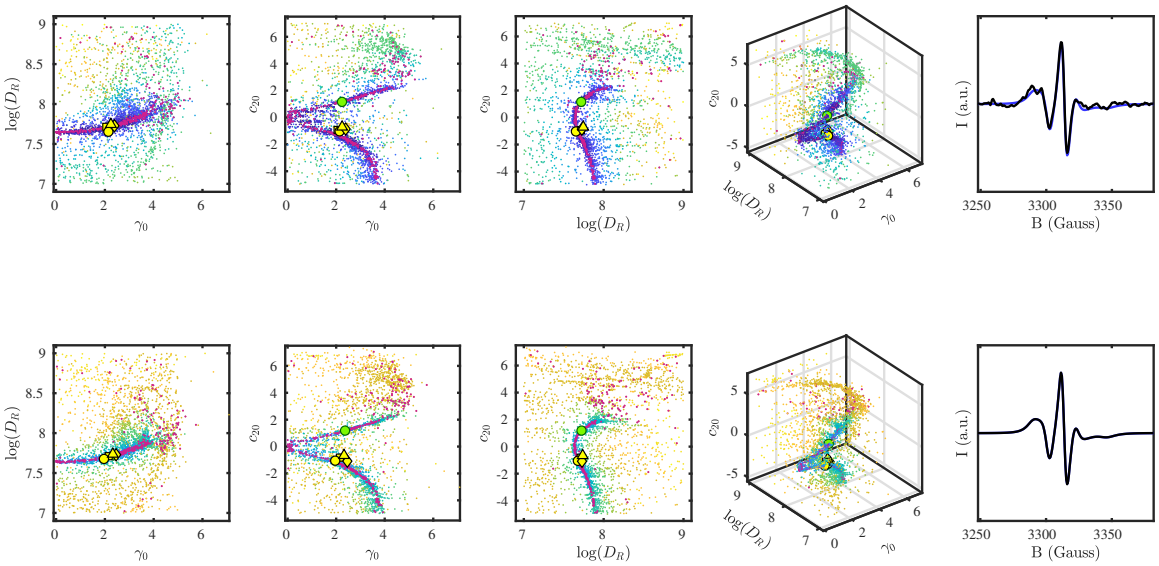
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 287K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



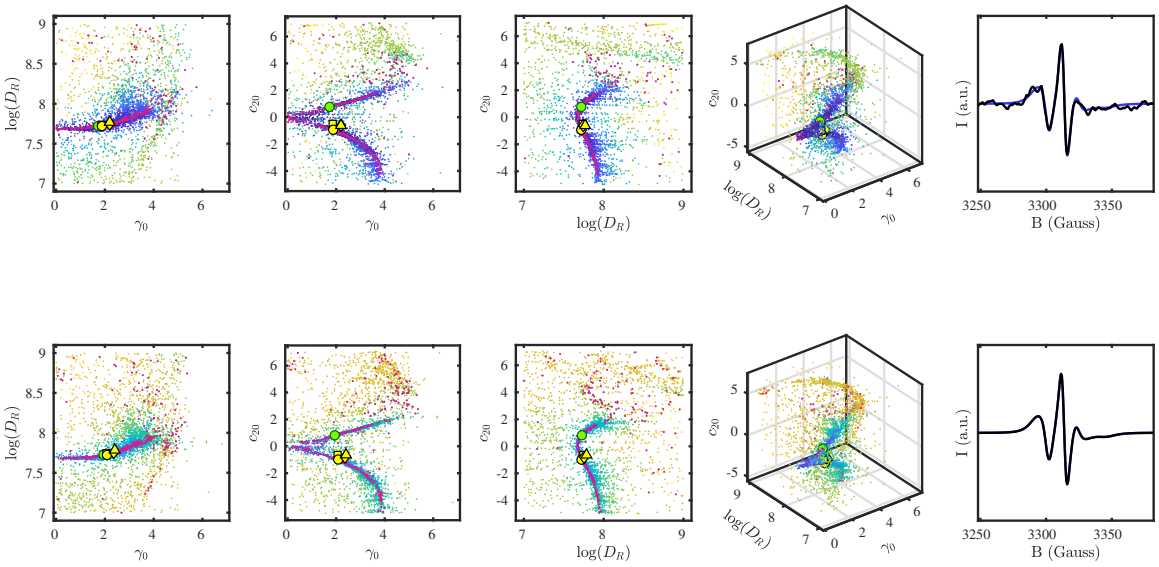
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 290K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



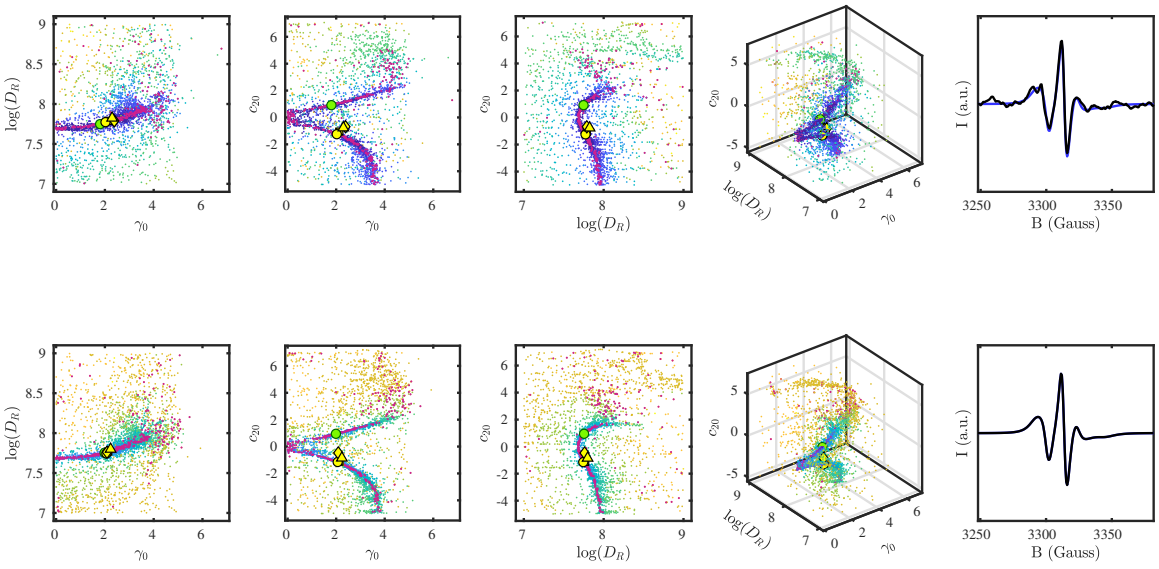
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 293K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



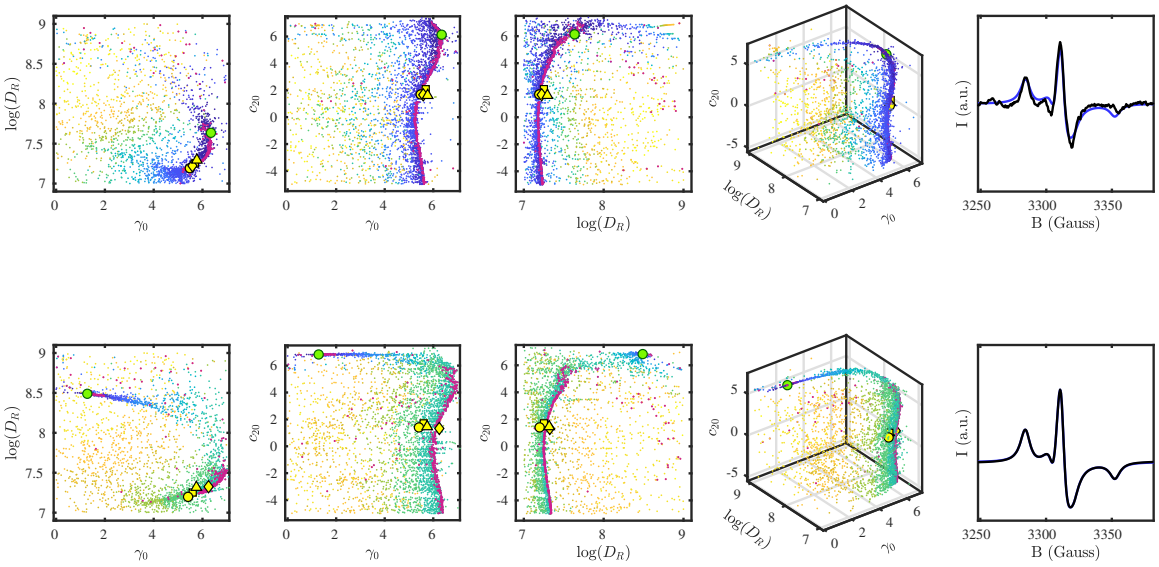
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 296K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



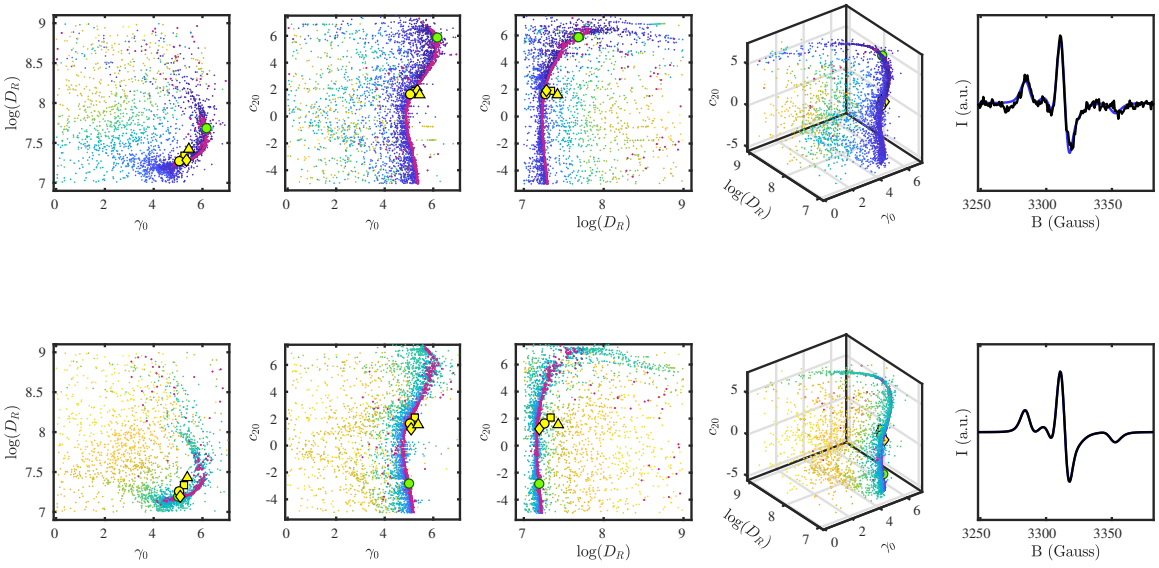
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 299K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



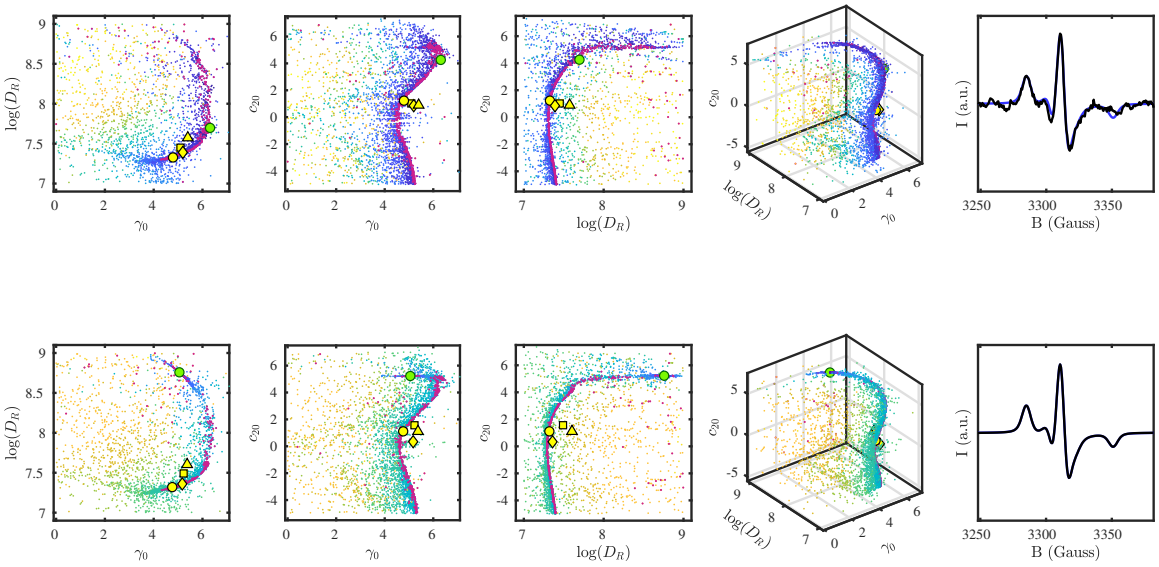
χ^2_ν maps for fitting of the experimental AA-P3-SL spectrum (top), collected at 302K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



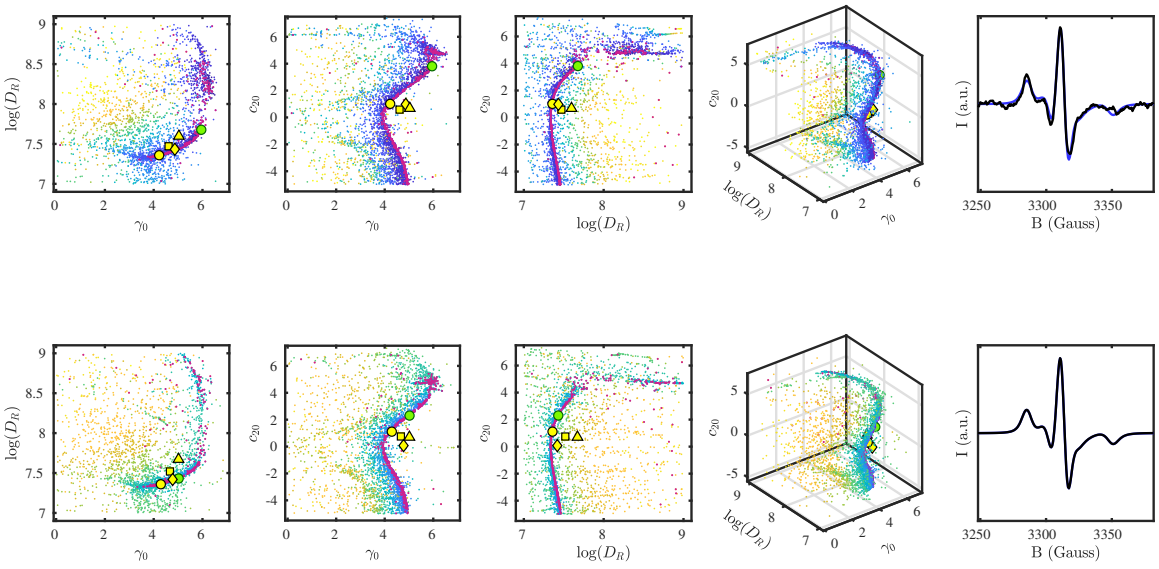
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 275K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



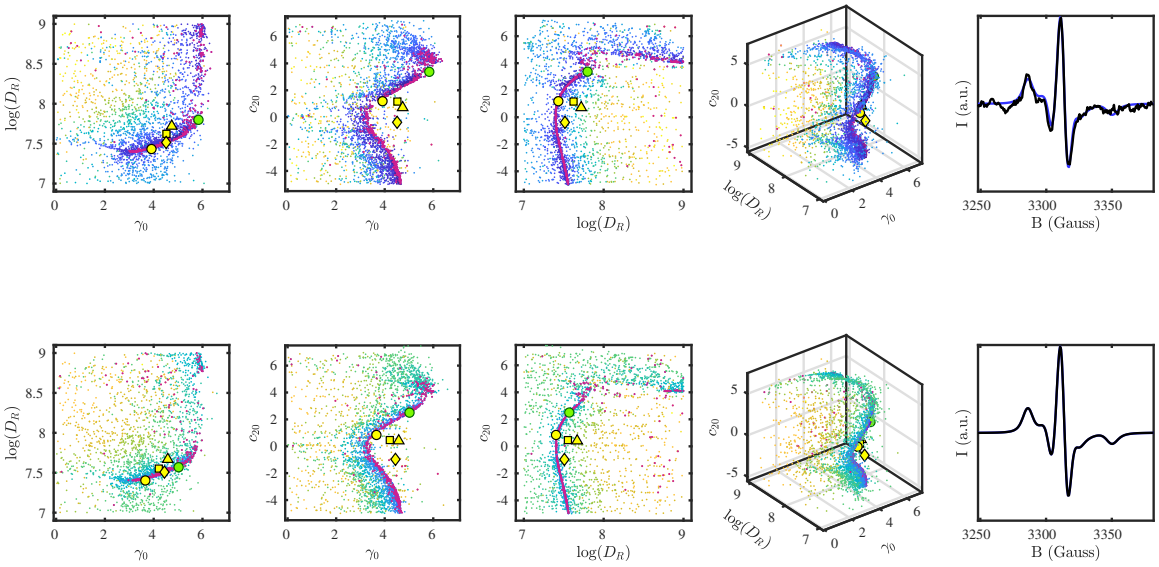
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 278K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



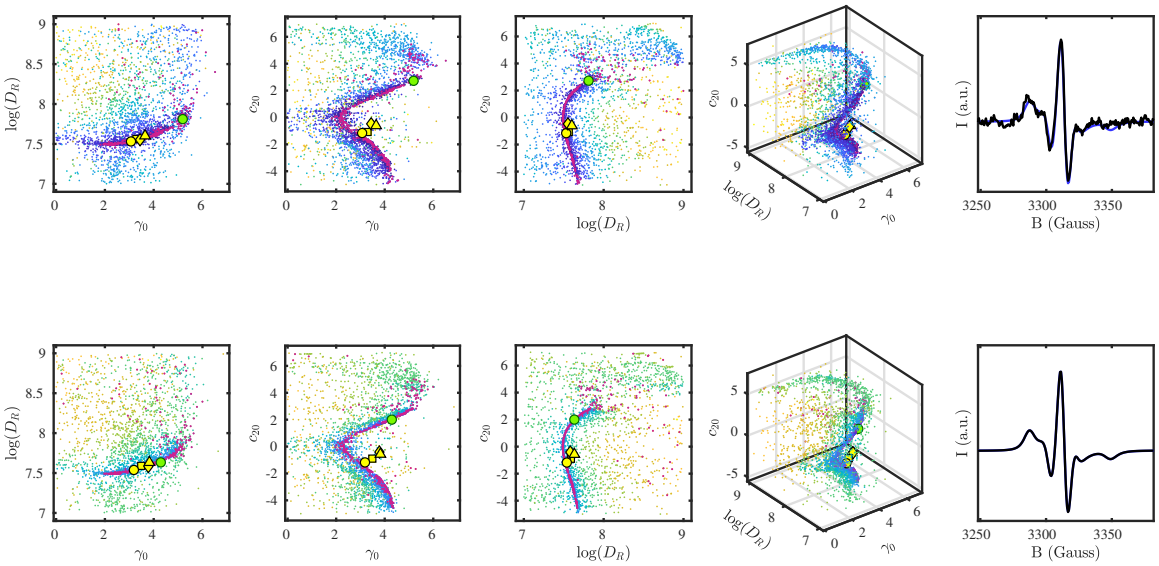
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 281K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



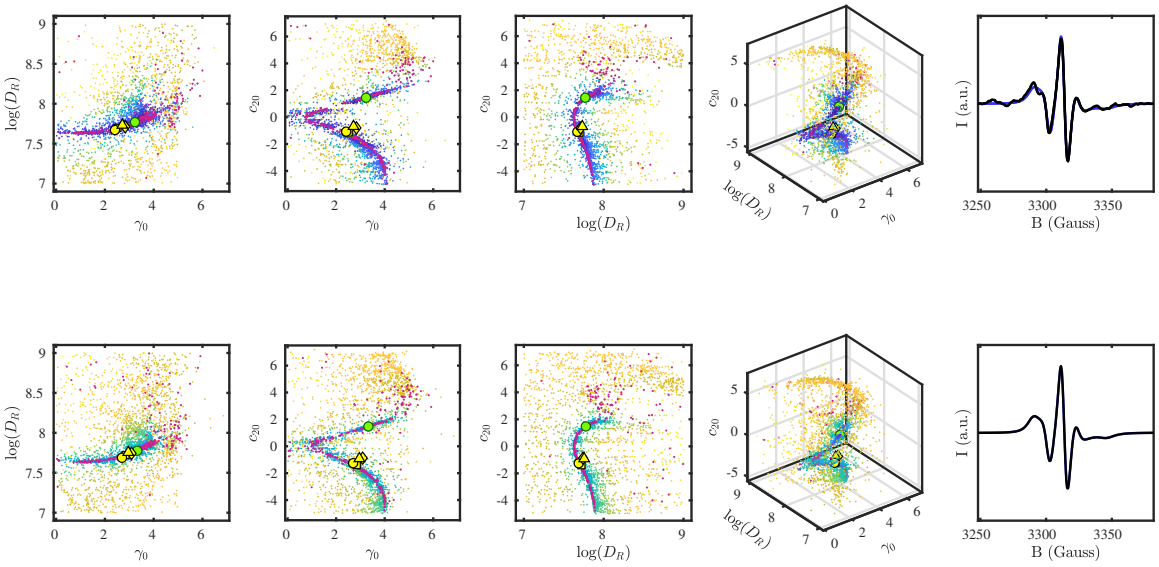
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 284K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



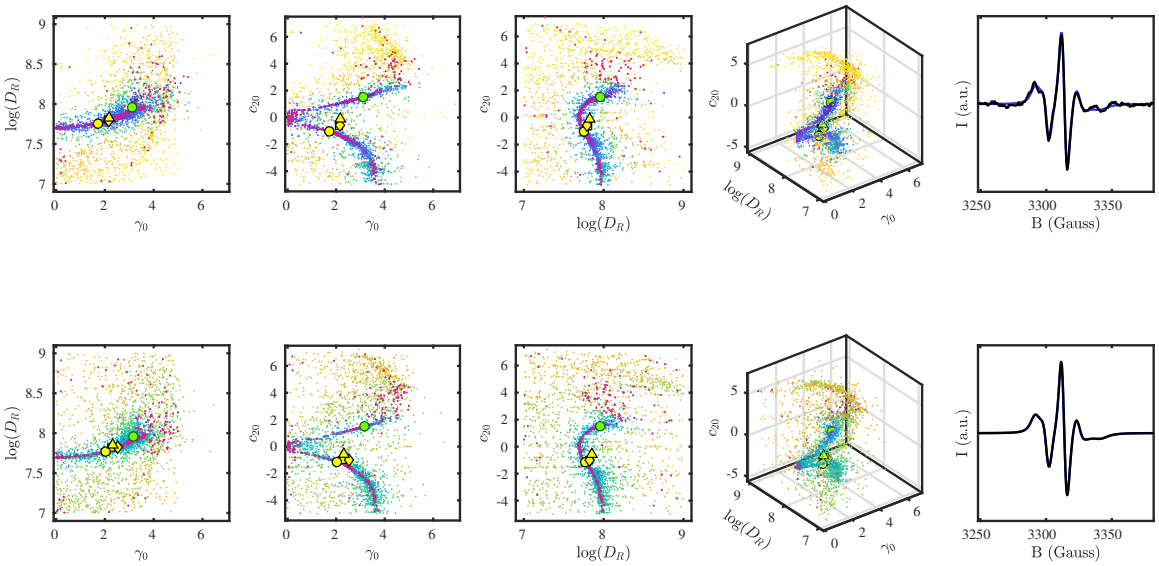
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 287K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



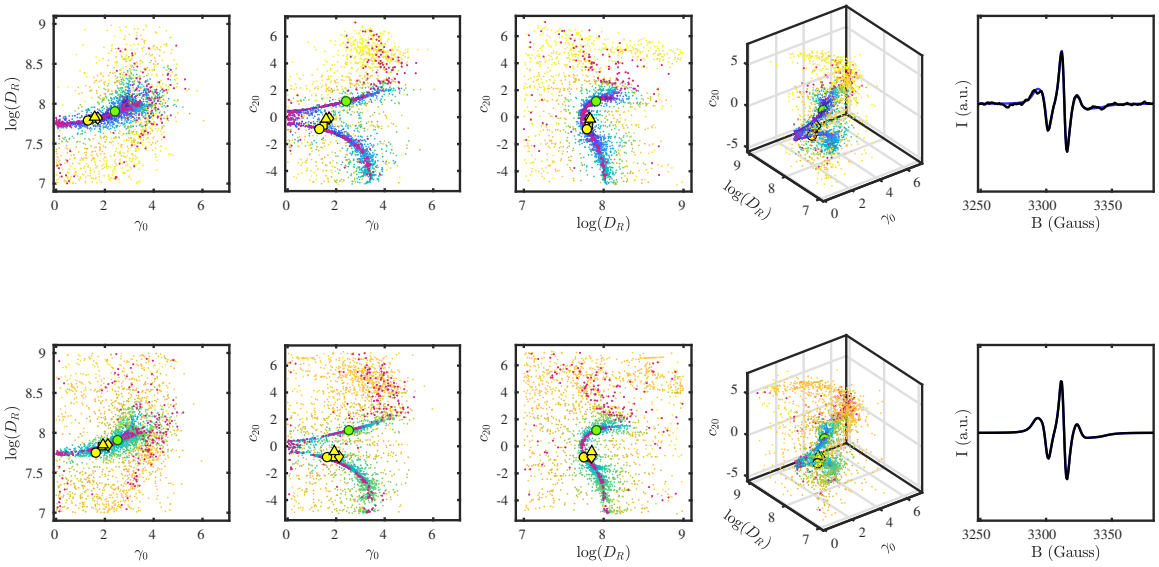
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 290K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



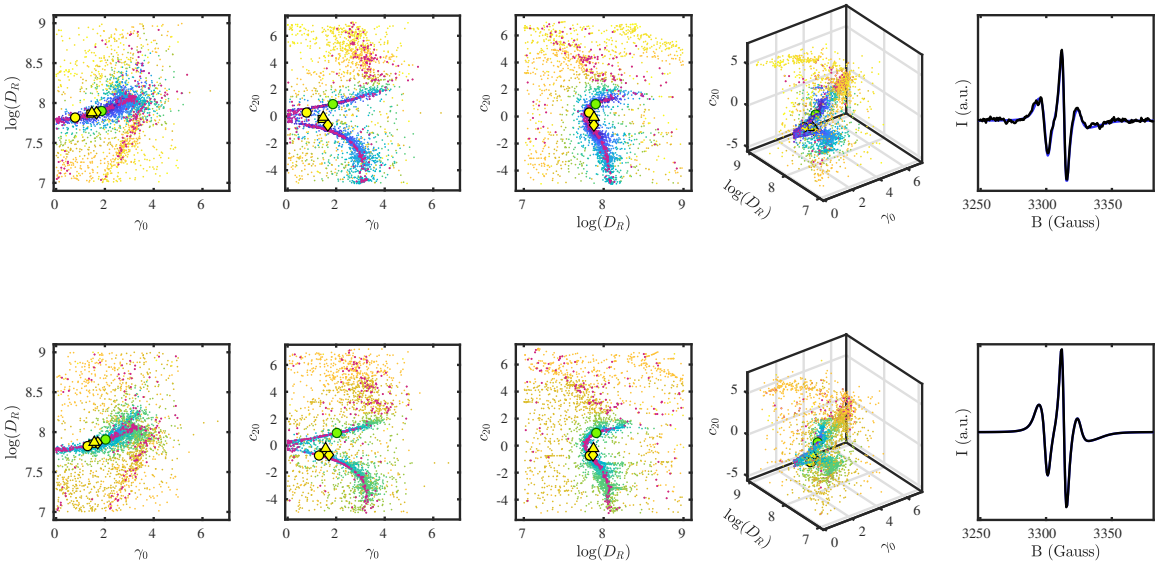
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 293K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



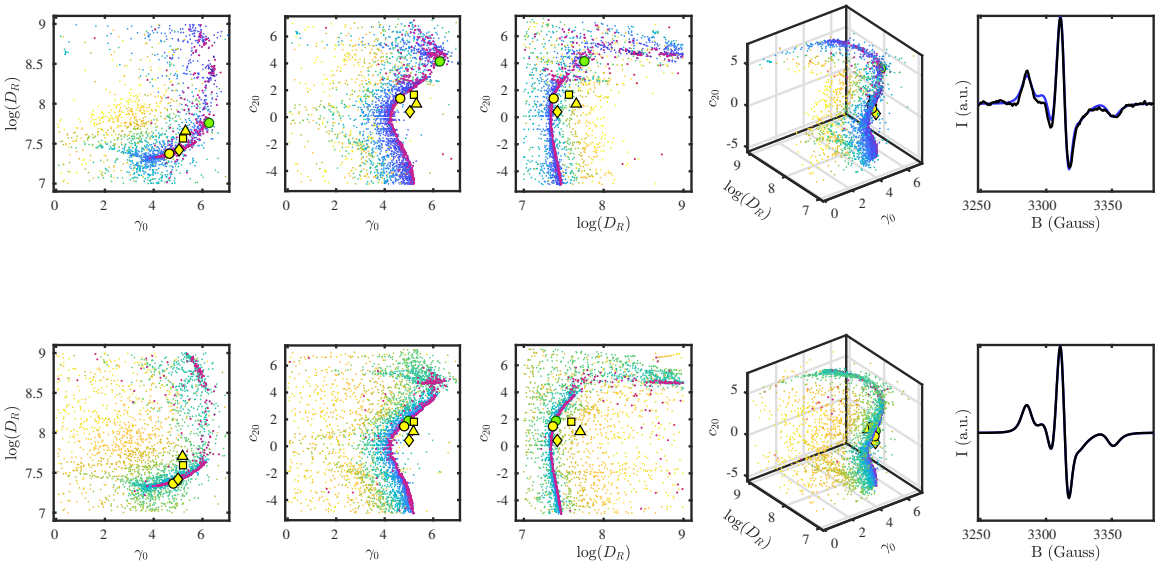
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 296K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



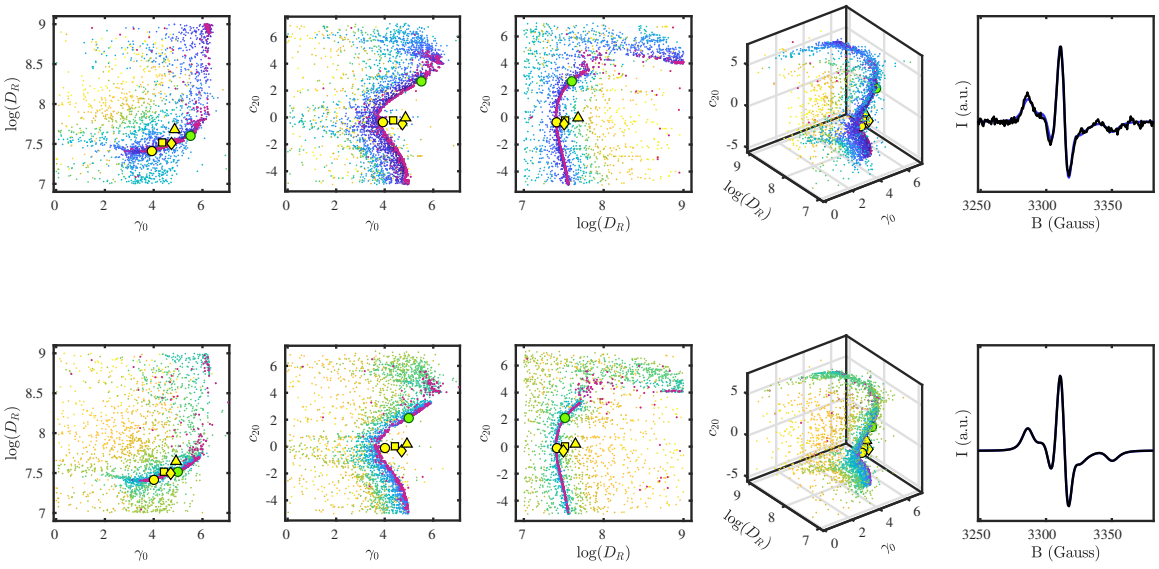
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 299K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



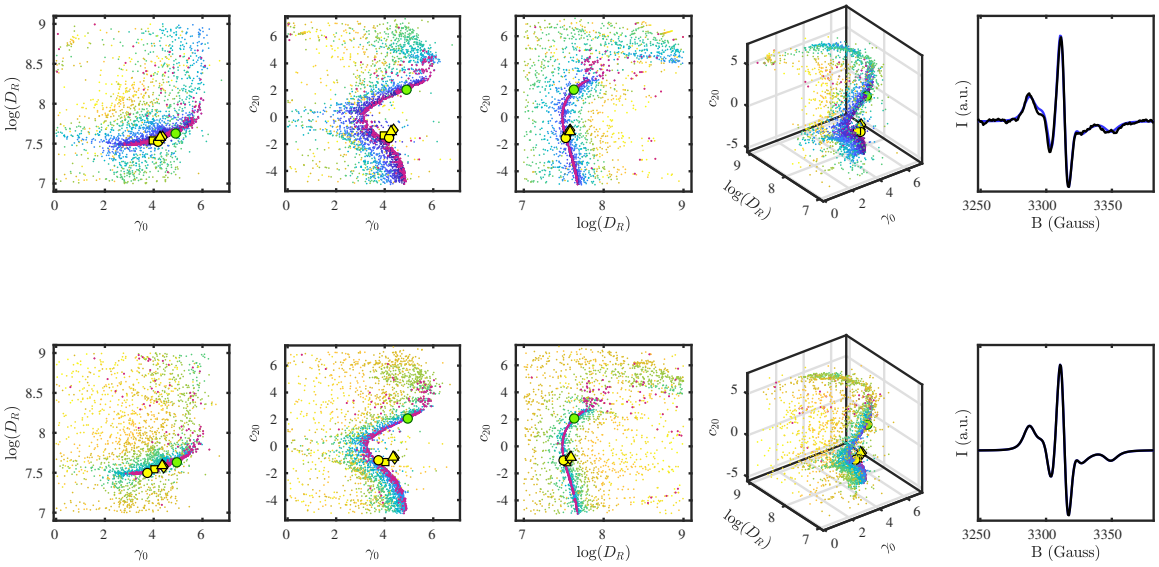
χ^2_ν maps for fitting of the experimental AA-P5-SL spectrum (top), collected at 302K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



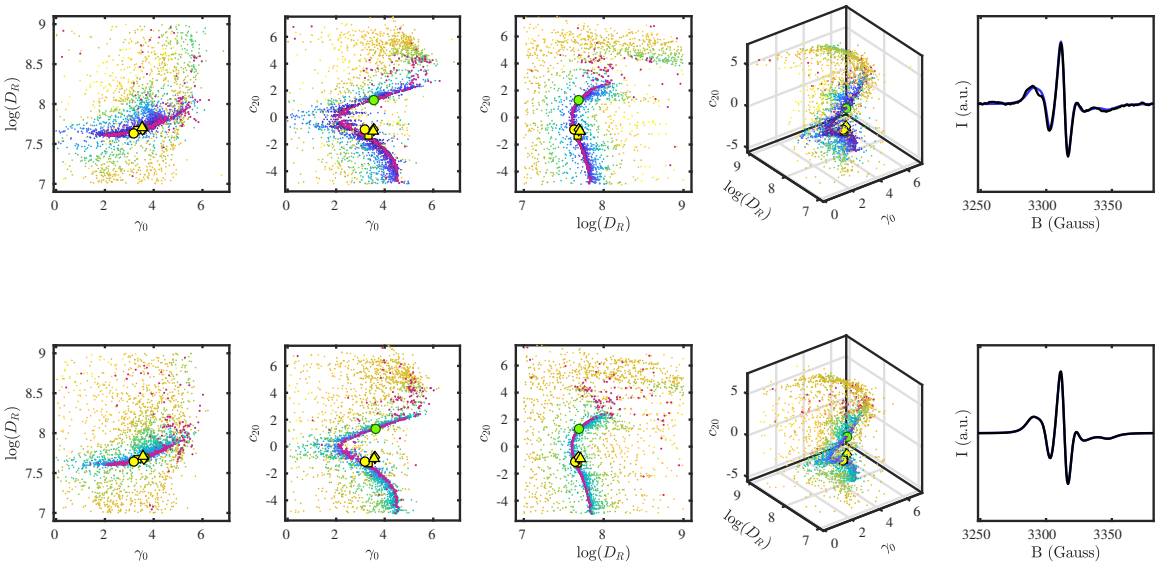
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 275K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



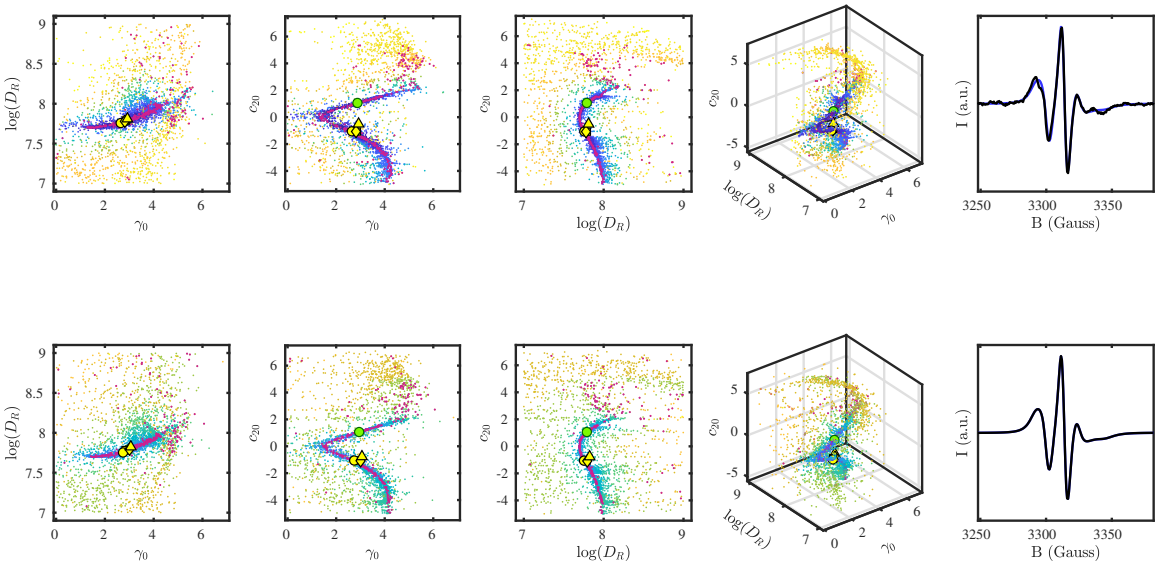
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 278K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



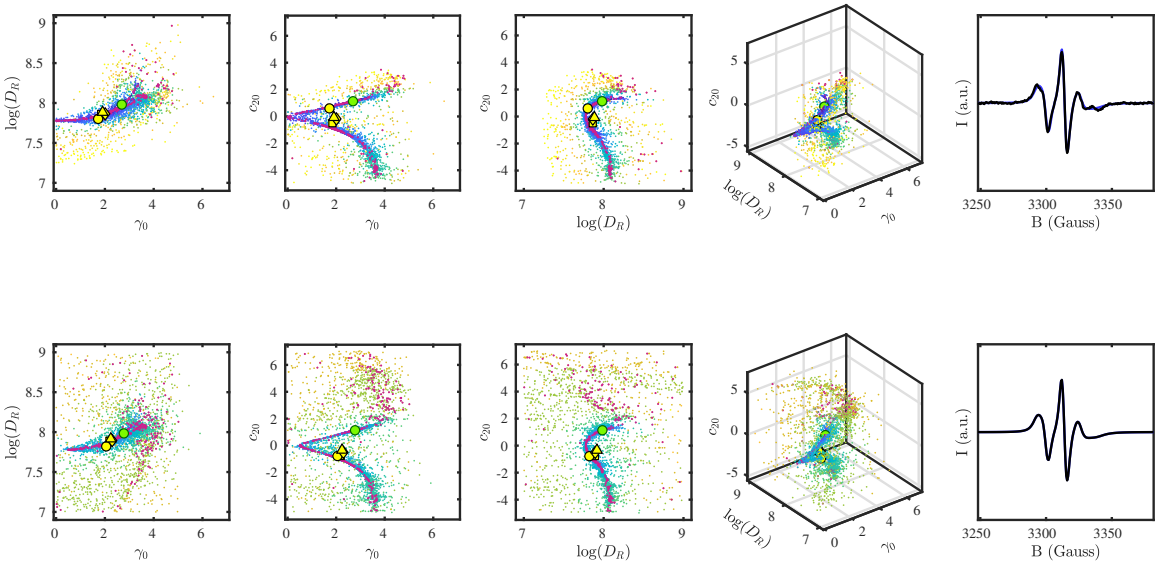
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 281K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



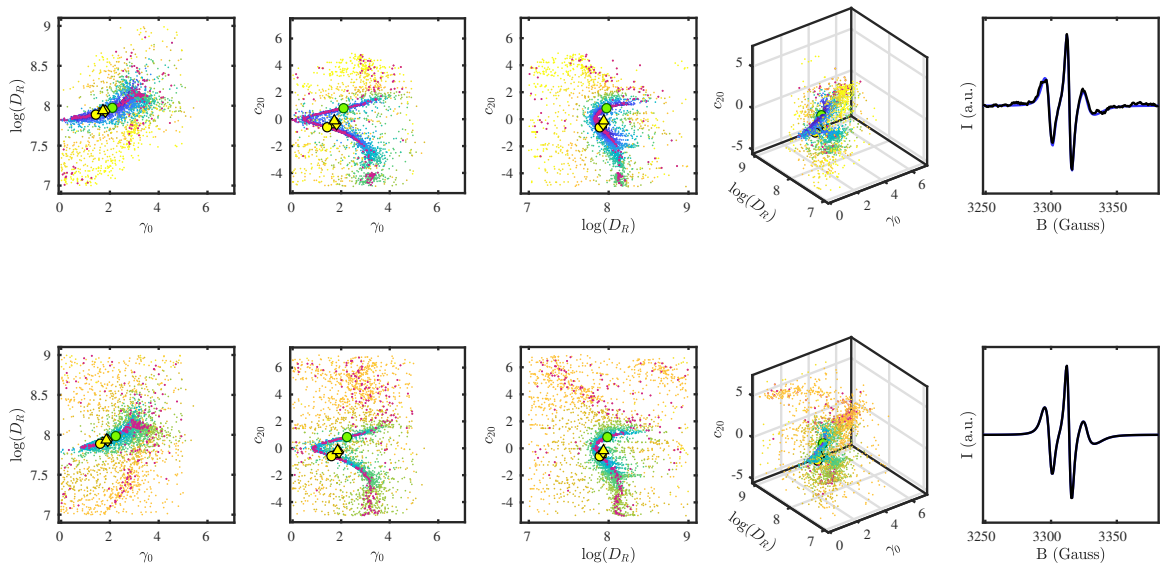
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 284K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



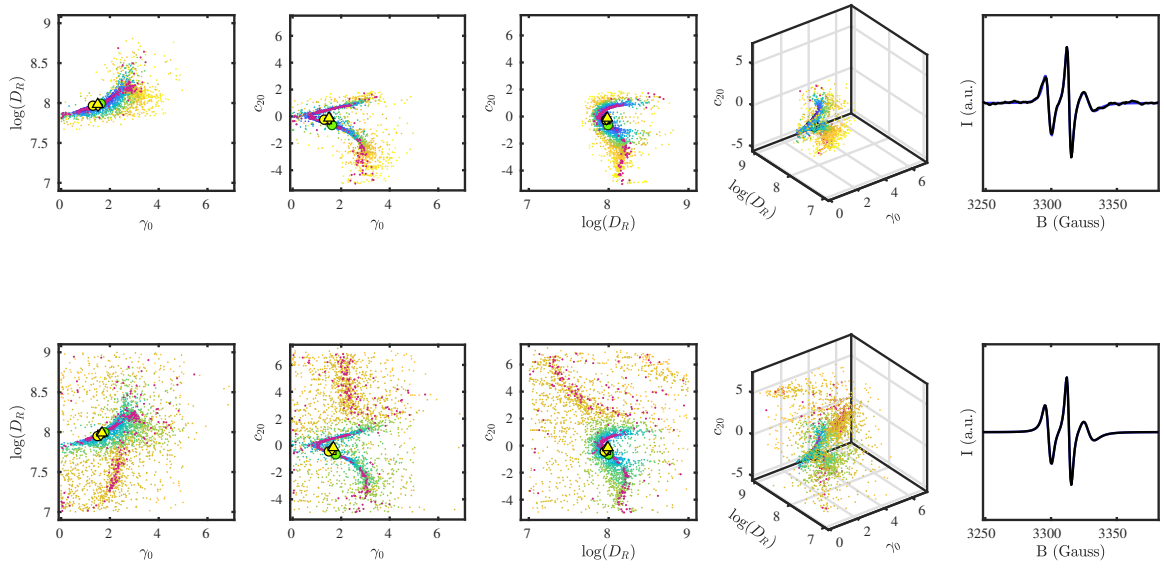
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 287K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



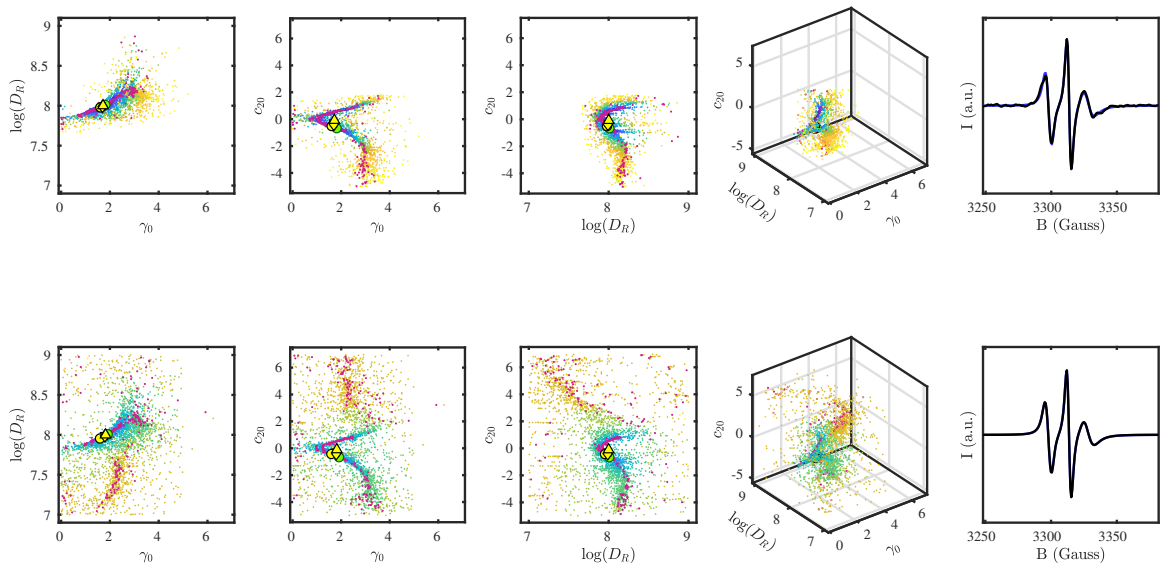
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 290K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



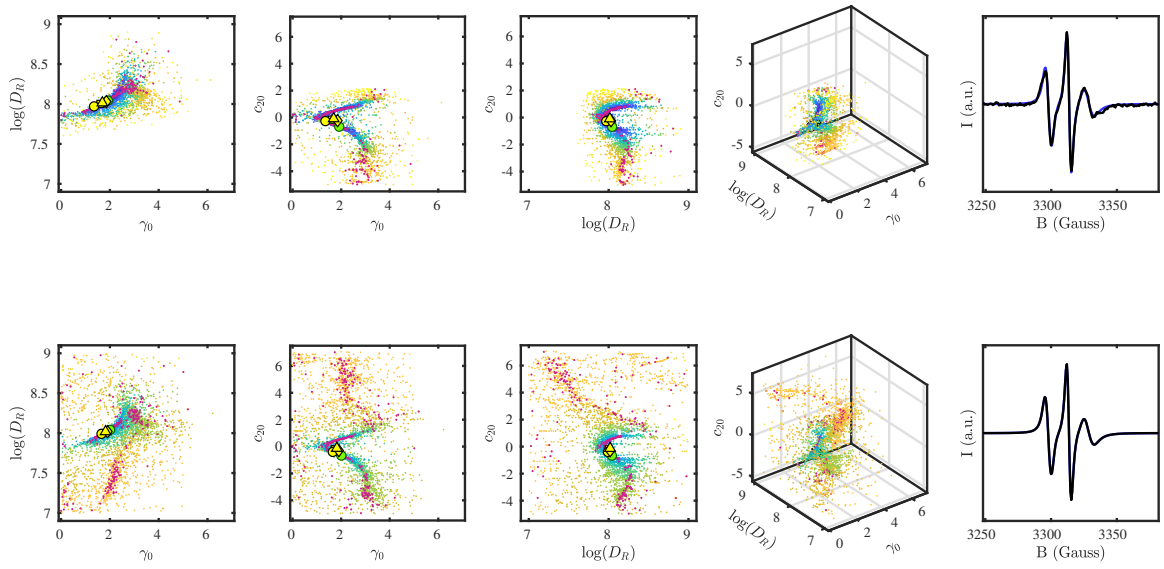
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 293K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



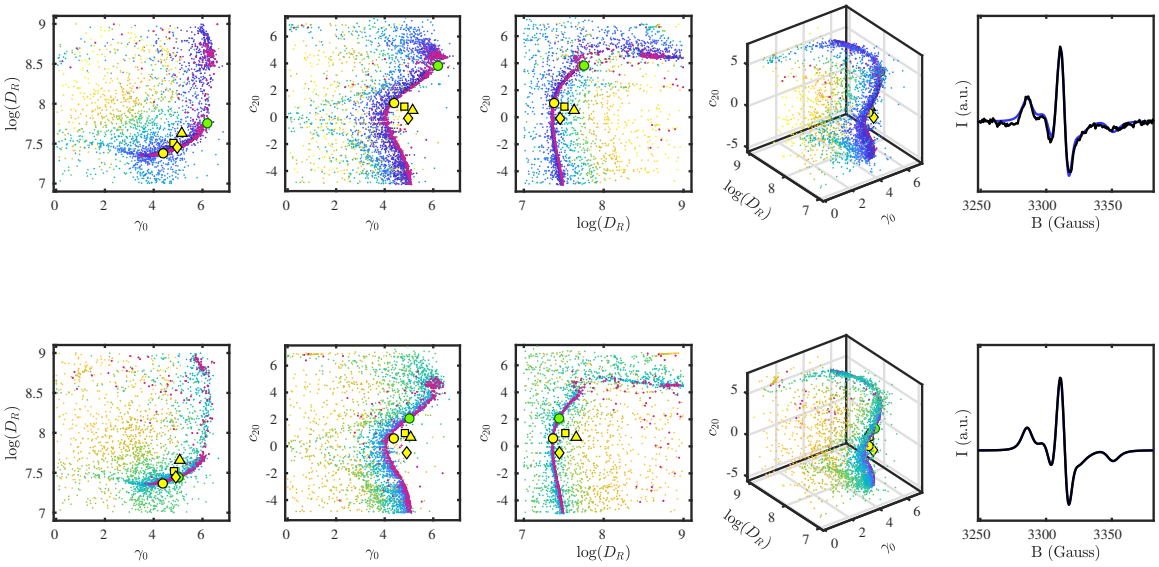
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 296K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



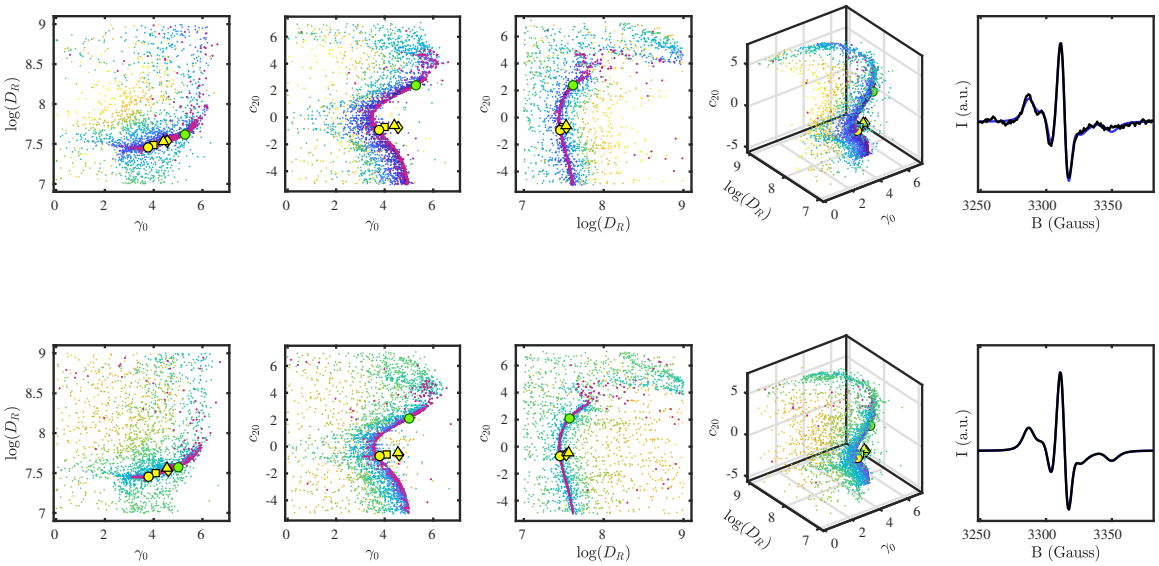
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 299K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



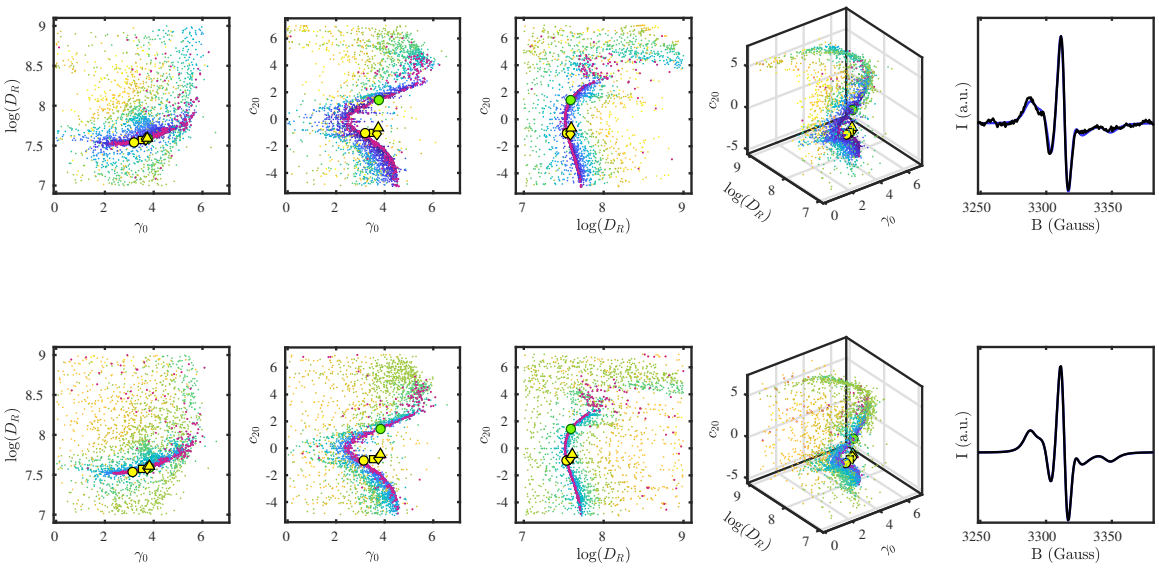
χ^2_ν maps for fitting of the experimental AA-P8-SL spectrum (top), collected at 302K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



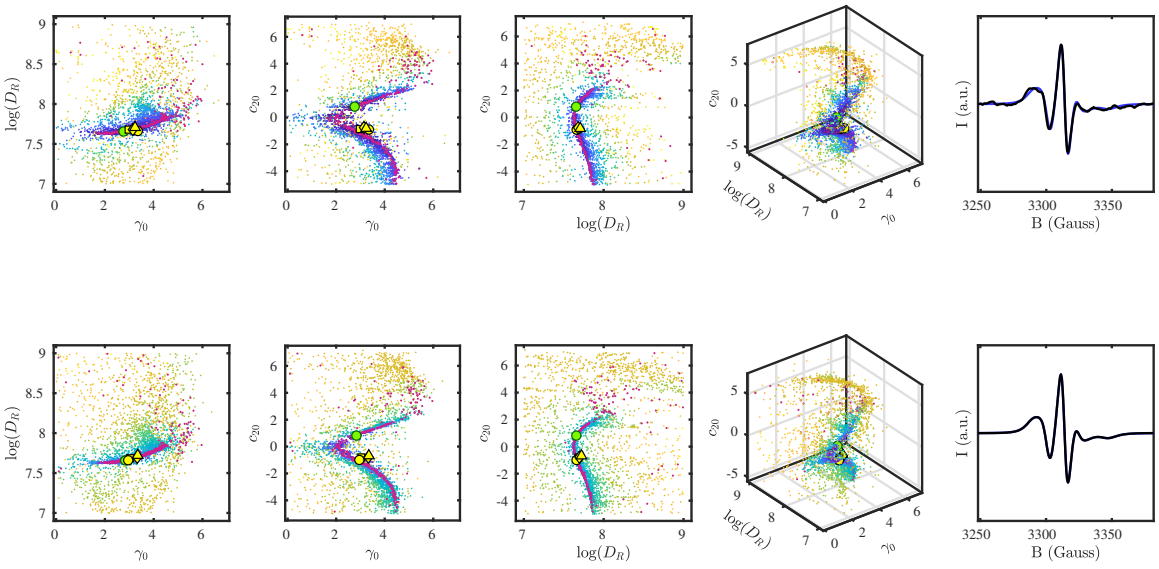
χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 275K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



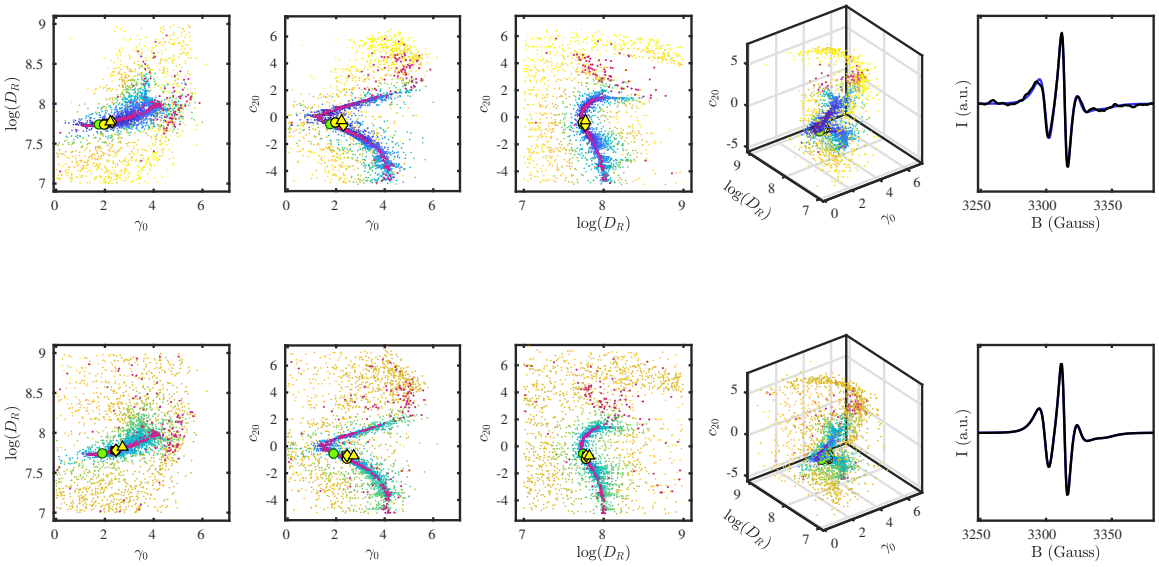
χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 278K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



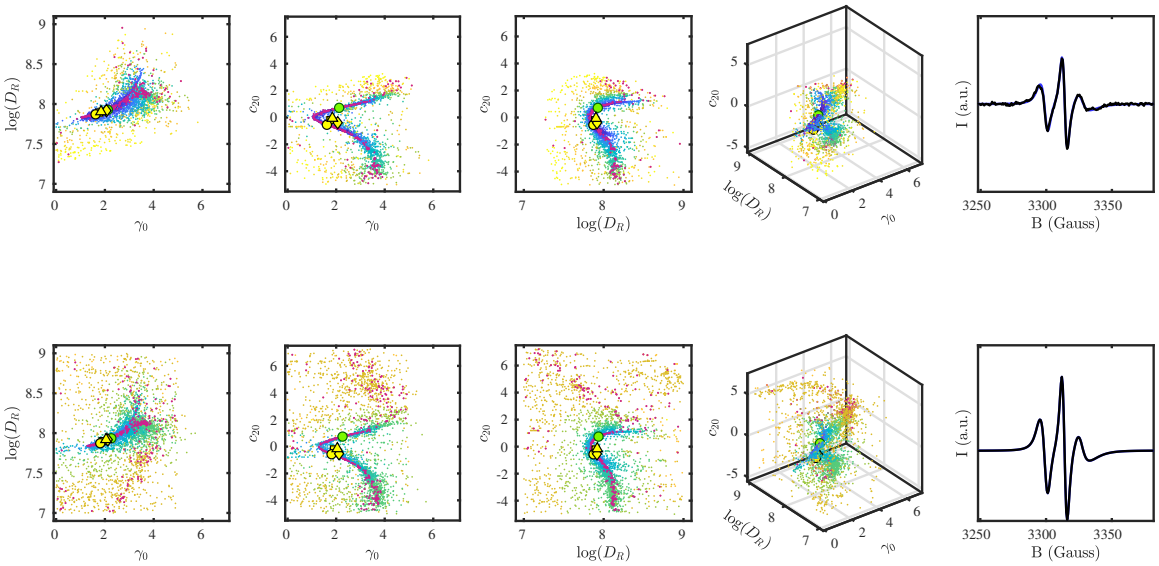
χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 281K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



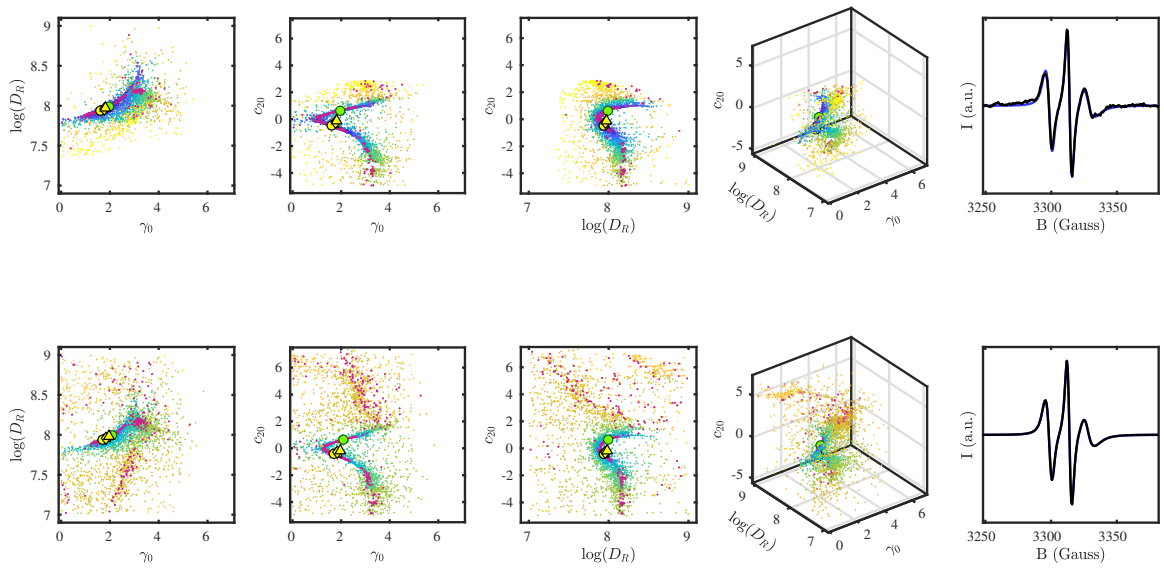
χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 284K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



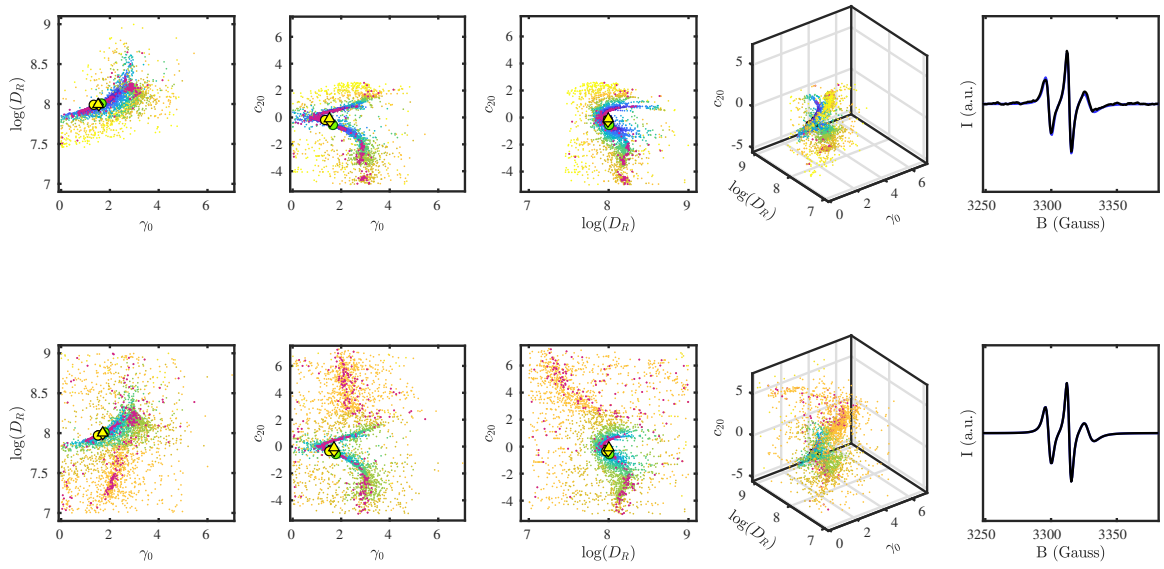
χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 287K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



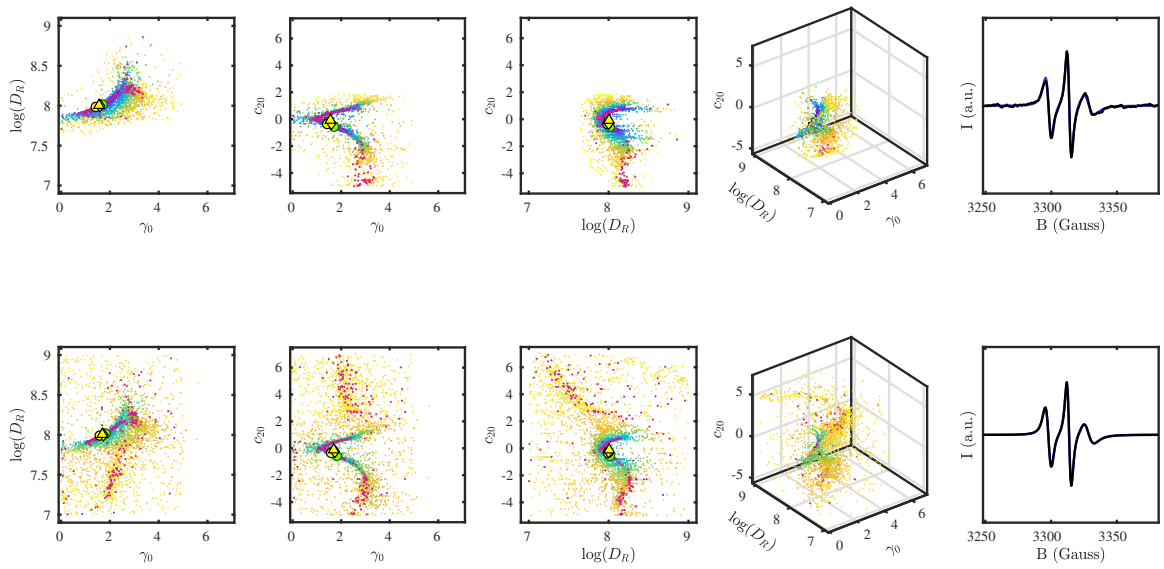
χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 290K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



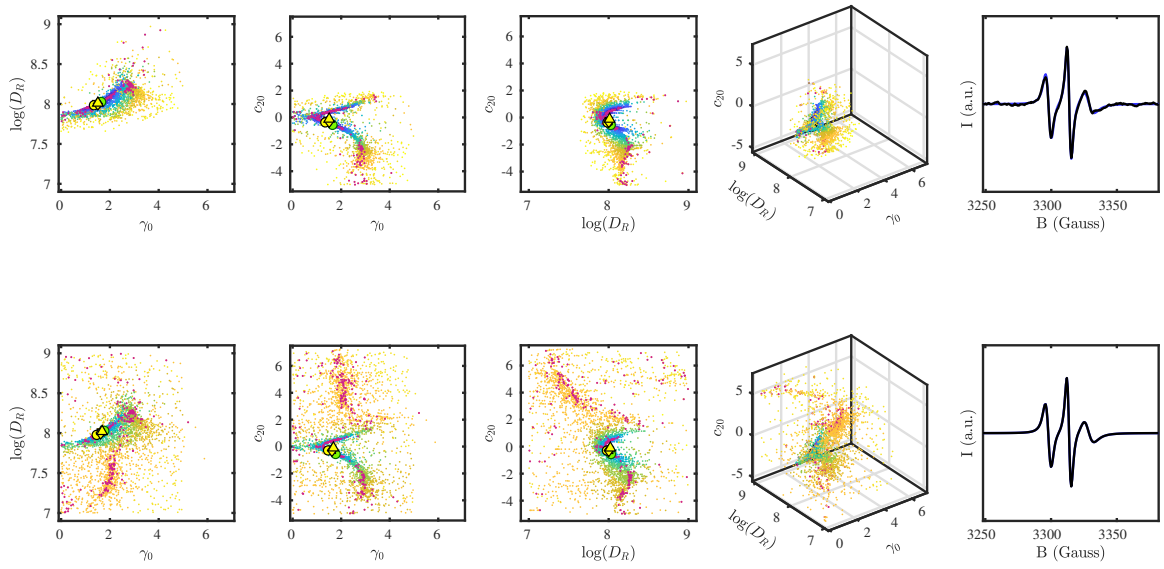
χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 293K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 296K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 299K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.



χ^2_ν maps for fitting of the experimental AA-P12-SL spectrum (top), collected at 302K, and fitting of the simulated spectrum (bottom) that best describes the experimental data. Local minima are plotted in purple, the large green marker corresponds to the global minimum, and the yellow square, circle and diamond corresponds to the geometric median, the medoid and the marginal median.

Appendix E

Bibliography

[1] Lewis E Kay. Protein dynamics from NMR 1. *Cell*, 76(Palmer 1993):145–152, 1998.

[2] Tanja Mittag, Lewis E. Kay, and Julie D. Forman-Kaya. Protein dynamics and conformational disorder in molecular recognition. *Journal of Molecular Recognition*, 23(2):105–116, 2010. ISSN 09523499. doi: 10.1002/jmr.961.

[3] Yuri E. Nesmelov and David D. Thomas. Protein structural dynamics revealed by site-directed spin labeling and multifrequency EPR. *Biophysical Reviews*, 2 (2):91–99, 2010. ISSN 18672469. doi: 10.1007/s12551-010-0032-5.

[4] Ganesan Balasundaram, Chang Yao, and Thomas J. Webster. TiO2 nanotubes functionalized with regions of bone morphogenetic protein-2 increases osteoblast adhesion Ganesan. *Journal of Biomedical Materials Research Part A*, 84A(2): 447–453, 2007. ISSN 1593098X. doi: 10.1002/jbm.a.

[5] P. C. Bessa, M. Casal, and R. L. Reis. Bone morphogenetic proteins in tissue engineering: the road from the laboratory to the clinic, part I (basic concepts). *Journal of tissue engineering and regenerative medicine*, 2(1):1–13, 2008. ISSN 1932-7005. doi: 10.1002/term.

[6] Chunmei Li, Charu Vepari, Hyoung-Joon Jin, Hyeon Joo Kim, and David L Kaplan. Electrospun silk-BMP-2 scaffolds for bone tissue engineering. *Biomaterials*, 27(16):3115–24, 2006. ISSN 0142-9612. doi: 10.1016/j.biomaterials.2006.01.022.

[7] Sungsoo S. Lee, Brian J. Huang, Stuart R. Kaltz, Shantanu Sur, Christina J. Newcomb, Stuart R. Stock, Ramille N. Shah, and Samuel I. Stupp. Bone regeneration with low dose BMP-2 amplified by biomimetic supramolecular nanofibers within collagen scaffolds. *Biomaterials*, 34(2):452–459, 2013. ISSN 01429612. doi: 10.1016/j.biomaterials.2012.10.005.

[8] M J Mahoney and W M Saltzman. Transplantation of brain cells assembled around a programmable synthetic microenvironment. *Nature biotechnology*, 19 (October):934–939, 2001. ISSN 1087-0156. doi: 10.1038/nbt1001-934.

[9] Gregory R.D Evans, Keith Brandt, Steven Katz, Priscilla Chauvin, Lisa Otto, Melissa Bogle, Bao Wang, Rudolph K Meszlenyi, Lichun Lu, Antonios G Mikos, and Charles W Patrick. Bioactive poly(l-lactic acid) conduits seeded with Schwann cells for peripheral nerve regeneration. *Biomaterials*, 23(3):841–848, 2002. ISSN 01429612. doi: 10.1016/S0142-9612(01)00190-9.

[10] Marcin Jurga, Maria B. Dainiak, Anna Sarnowska, Anna Jablonska, Anuj Tripathi, Fatima M. Plieva, Irina N. Savina, Lukasz Strojek, Hans Jungvid, Ashok Kumar, Barbara Lukomska, Krystyna Domanska-Janik, Nico Forraz, and Colin P. McGuckin. The performance of laminin-containing cryogel scaffolds in neural tissue regeneration. *Biomaterials*, 32(13):3423–3434, 2011. ISSN 01429612. doi: 10.1016/j.biomaterials.2011.01.049.

[11] Natalia Gomez and Christine Schmidt. Nerve growth factor-immobilized polypyrrole: Bioactive electrically conducting polymer for enhanced neurite extension. *Journal of Biomedical Materials Research Part A*, 81A(1):135–149, 2007. ISSN 1593098X. doi: 10.1002/jbm.a.

[12] K M Galler, R N D'Souza, J D Hartgerink, and G Schmalz. Scaffolds for dental pulp tissue engineering. *Advances in dental research*, 23(3):333–9, 2011. ISSN 1544-0737. doi: 10.1177/0022034511405326.

[13] Kerstin M. Galler, Jeffrey D. Hartgerink, Adriana C. Cavender, Gottfried Schmalz, and Rena N. D'Souza. A customized self-assembling peptide hydrogel for dental pulp tissue engineering. *Tissue engineering. Part A*, 18(1-2):176–84, 2012. ISSN 1937-335X. doi: 10.1089/ten.TEA.2011.0222.

[14] Michel Goldberg, Jean Christophe Farges, Sally Lacerda-Pinheiro, Ngampis Six, Nadège Jegat, Frank Decup, Dominique Septier, Florence Carrouel, Stéphanie Durand, Catherine Chaussain-Miller, Pamela DenBesten, Arthur Veis, and Anne Poliard. Inflammatory and immunological aspects of dental pulp repair. *Pharmacological Research*, 58(2):137–147, 2008. ISSN 10436618. doi: 10.1016/j.phrs.2008.05.013.

[15] D Seliktar, A H Zisch, M P Lutolf, J L Wrana, and J A Hubbell. MMP-2 sensitive, VEGF-bearing bioactive hydrogels for promotion of vascular healing. *J Biomed Mater Res A*, 68(4):704–716, 2004. ISSN 1549-3296. doi: 10.1002/jbm.a.20091.

[16] Hojae Bae, Amey S. Puranik, Robert Gauvin, Faramarz Edalat, Brenda Carrillo-Conde, Nicholas A. Peppas, and Ali Khademhosseini. Building Vascular Networks. *Science Translational Medicine*, 4(160):160ps23–160ps23, 2012. ISSN 1946-6234, 1946-6242. doi: 10.1126/scitranslmed.3003688.

[17] Ruth R. Chen, Eduardo A. Silva, William W. Yuen, and David J. Mooney. Spatio-temporal VEGF and PDGF delivery patterns blood vessel formation and maturation. *Pharmaceutical Research*, 24(2):258–264, 2007. ISSN 07248741. doi: 10.1007/s11095-006-9173-4.

[18] Thomas Burr Osborne. *The Vegetable Protein*, volume 1. Longman's Green & Co., 1859. ISBN 3663537137.

[19] Frederick Sanger. The terminal peptides of insulin. *Biochem. J.*, 45(5):563–574, 1949. ISSN 0264-6021.

[20] Antony O W Stretton. The first sequence: Fred Sanger and insulin. *Genetics*, 162(2):527–532, 2002. ISSN 00166731.

[21] L Pauling, R B Corey, and H R Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 37(4): 205–211, 1951. ISSN 0027-8424. doi: 10.1073/pnas.37.4.205.

[22] H M Berman, J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov, and P E Bourne. The Protein Data Bank. *Nucleic acids research*, 28(1):235–242, 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.235.

[23] Helen Berman, Kim Henrick, Haruki Nakamura, and John L. Markley. The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, 35(SUPPL. 1):2006–2008, 2007. ISSN 03051048. doi: 10.1093/nar/gkl971.

[24] K Wüthrich. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science*, 243:45, 1989.

[25] K Wüthrich. The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. *Accounts of chemical research*, (10):36–44, 1989. ISSN 0001-4842. doi: 10.1021/ar00157a006.

[26] Torsten Herrmann, Peter Güntert, and Kurt Wüthrich. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology*, 319(1):209–227, 2002. ISSN 00222836. doi: 10.1016/S0022-2836(02)00241-3.

[27] Andrea Cavalli, Xavier Salvatella, Christopher M Dobson, and Michele Vendruscolo. Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23): 9615–20, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0610313104.

[28] B W Matthews. X-ray Crystallographic Studies of Proteins. *Annual review of physical chemistry*, 27(1):493, 1976. ISSN 0066-426X. doi: 10.1146/annurev.pc.27.100176.002425.

[29] J R Helliwell. Synchrotron X-radiation protein crystallography: instrumentation, methods and applications. *Reports on Progress in Physics*, 47(11):1403–1497, 1984. ISSN 0034-4885. doi: 10.1088/0034-4885/47/11/001.

[30] Christopher D Putnam, Michal Hammel, Greg L Hura, and John a Tainer. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*, 40(3):191–285, 2007. ISSN 00063495. doi: 10.1016/j.bpj.2010.12.409.

[31] Sébastien Boutet, Lukas Lomb, Garth J Williams, Thomas R M Barends, Andrew Aquila, R Bruce Doak, Uwe Weierstall, Daniel P Deponte, Jan Steinbrener, Robert L Shoeman, Marc Messerschmidt, Anton Barty, Thomas A White, Stephan Kassemeyer, Richard A Kirian, M Marvin Seibert, Paul A Montanez, Chris Kenney, Ryan Herbst, Philip Hart, Jack Pines, Gunther Haller, Sol M Gruner, Hugh T Philipp, Mark W Tate, Marianne Hromalik, Lucas J Koerner, Niels Van Bakel, John Morse, Wilfred Ghonsalves, David Arnlund, Karol Nass, Lars Redecke, Francesco Stellato, Nicusor Timneanu, Dingjie Wang, John C H Spence, Henry N Chapman, and Ilme Schlichting. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science*, 337: 362–364, 2012.

[32] Vlado Dančík, Theresa A Addona, Karl R Clauser, James E Vath, and Pavel A Pevzner. De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol*, 6(3-4):327–342, 1999. ISSN 1066-5277 (Print). doi: 10.1089/106652799318300.

[33] K Gevaert and J Vandekerckhove. Protein identification methods in proteomics. *Electrophoresis*, 21(6):1145–1154, 2000. ISSN 0173-0835. doi: 10.1002/(SICI)1522-2683(20000401)21:6<1145::AID-ELPS1145>3.0.CO;2-Z.

[34] Hanno Steen and Matthias Mann. The ABC's (and XYZ's) of peptide sequencing. *Nature reviews. Molecular cell biology*, 5(9):699–711, 2004. ISSN 1471-0072. doi: 10.1038/nrm1468.

[35] David Franklyn Williams. The Williams Dictionary of Biomaterials. In *The Williams Dictionary of Biomaterials*, volume 1, pages 39–44. 1999. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.

[36] Hong Zhang, Guangwen Wang, Jian Li, Yuchun Nie, Xuanling Shi, Gewei Lian, Wei Wang, Xiaolei Yin, Yang Zhao, Xiuxia Qu, Mingxiao Ding, and Hongkui Deng. Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies. *Journal of virology*, 78(13):6938–45, 2004. ISSN 0022-538X. doi: 10.1128/JVI.78.13.6938-6945.2004.

[37] Bruno E Correia, John T Bates, Rebecca J Loomis, Gretchen Baneyx, Chris Carrico, Joseph G Jardine, Peter Rupert, Colin Correnti, Oleksandr Kalyuzhniy, Vinayak Vittal, Mary J Connell, Eric Stevens, Alexandria Schroeter, Man Chen, Skye Macpherson, Andreia M Serra, Yumiko Adachi, Margaret A Holmes, Yuxing Li, Rachel E Klevit, Barney S Graham, Richard T Wyatt, David Baker,

Roland K Strong, James E Crowe, Philip R Johnson, and William R Schief. Proof of principle for epitope-focused vaccine design. *Nature*, 507(7491):201–6, 2014. ISSN 1476-4687. doi: 10.1038/nature12966.

[38] Anthony W Purcell, James McCluskey, and Jamie Rossjohn. More than one reason to rethink the use of peptides in vaccine design. *Nature reviews. Drug discovery*, 6(5):404–414, 2007. ISSN 1474-1776. doi: 10.1038/nrd2224.

[39] Jessica J. Chiang, Matthew R. Gardner, Brian D. Quinlan, Tatyana Dorfman, Hyeryun Choe, and Michael Farzan. Enhanced Recognition and Neutralization of HIV-1 by Antibody-Derived CCR5-Mimetic Peptide Variants. *Journal of Virology*, 86(22):12417–12421, 2012. ISSN 0022-538X, 1098-5514. doi: 10.1128/JVI.00967-12.

[40] John A. Robinson. Max Bergmann lecture Protein epitope mimetics in the age of structural vaccinology. *Journal of Peptide Science*, 19(3):127–140, 2013. ISSN 10752617. doi: 10.1002/psc.2482.

[41] M. Z. Atassi. Antigenic structure of myoglobin: The complete immunochemical anatomy of a protein and conclusions relating to antigenic structures of proteins. *Immunochemistry*, 12(5):423–438, 1975. ISSN 00192791. doi: 10.1016/0019-2791(75)90010-5.

[42] David C. Benjamin, Jay A. Berzofsky, lain J. East, Frank R. N. Gurd, Charles Hannum, Sydney J. Leach, Emanuel Margoliash, J. Gabriel Michael, Alexander Miller, Ellen M. Prager, Morris Reichlin, Eli E. Sercarz, Sandra J. Smith-Gill, Pam E. Todd, and A.C. Wilson. The antigenic structure of proteins: a reappraisal. *Annual review of Immunology*, 2:67–101, 1984.

[43] D J D J Barlow, M S M S Edwards, and J M J M Thornton. Continuous and discontinuous protein antigenic determinants. *Nature*, 322(6081):747–748, 1986. ISSN 0028-0836. doi: 10.1038/322747a0.

[44] Jian Huang and Wataru Honda. CED: A conformational epitope database. *BMC Immunology*, 7:1–8, 2006. ISSN 14712172. doi: 10.1186/1471-2172-7-7.

[45] Austin D Vogt and Enrico Di Cera. Conformational selection or induced fit? A critical Appraisal of the kinetic mechanism. *Biochemistry*, 51:5894–5902, 2012.

[46] Fabian Paul and Thomas R Weikl. How to Distinguish Conformational Selection and Induced Fit Based on Chemical Relaxation Rates. *PLoS Computational Biology*, pages 1–17, 2016. doi: 10.1371/journal.pcbi.1005067.

[47] Rebecca B Berlow, H Jane Dyson, and Peter E Wright. Functional advantages of dynamic protein disorder. *Federation of European Biochemical Societies Letters*, 589:2433–2440, 2015.

[48] Munehito Arai, Kenji Sugase, H Jane Dyson, and Peter E Wright. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proceedings of the National Academy of Sciences*, 2015. doi: 10.1073/pnas.1512799112.

[49] H Jane Dyson and Peter E Wright. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Protein (CBP) and p300. *The Journal of biological chemistry*, 291(13):6714–6722, 2016. doi: 10.1074/jbc.R115.692020.

[50] Robert B. Rose, Charles S. Craik, and Robert M. Stroud. Domain flexibility in retroviral proteases: Structural implications for drug resistant mutations. *Biochemistry*, 37(8):2607–2621, 1998. ISSN 00062960. doi: 10.1021/bi9716074.

[51] Rieko Ishima and Dennis Torchia. Protein dynamics from NMR. *Nature structural & molecular biology*, 76(2-3):145–152, 2000. ISSN 0829-8211. doi: 10.1139/o98-024.

[52] O Keskin, R L Jernigan, and I Bahar. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophysical journal*, 78(4):2093–2106, 2000. ISSN 00063495. doi: 10.1016/S0006-3495(00)76756-7.

[53] N Sinha and R Nussinov. Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proceedings of the National Academy of Sciences of the United States of America*, 98(6):3139–3144, 2001. ISSN 00278424. doi: 10.1073/pnas.051399098.

[54] Frank Gabel, Dominique Bicout, Ursula Lehnert, Moeava Tehei, Martin Weik, and Giuseppe Zaccai. Protein dynamics studied by neutron scattering. *Quarterly Reviews of Biophysics*, 35(4):327–367, 2002. ISSN 00335835. doi: 10.1017/S0033583502003840.

[55] Vasily S Oganesyan. A general approach for prediction of motional EPR spectra from Molecular Dynamics (MD) simulations: application to spin labelled protein. *Physical chemistry chemical physics : PCCP*, 13:4724–4737, 2011. ISSN 1463-9076. doi: 10.1039/c0cp01068e.

[56] Danuta Pentak, W. W. Sułkowski, and Anna Sułkowska. Calorimetric and EPR studies of the thermotropic phase behavior of phospholipid membranes. *Journal of Thermal Analysis and Calorimetry*, 93(2):471–477, 2008. ISSN 13886150. doi: 10.1007/s10973-007-8653-z.

[57] Irina V. Ionova, Vsevolod A. Livshits, and Derek Marsh. Phase diagram of ternary cholesterol/palmitoylsphingomyelin/ palmitoyloleoylphosphatidylcholine mixtures: Spin-label EPR study of lipid-raft formation. *Biophysical Journal*, 102:1856–1865, 2012. ISSN 00063495. doi: 10.1016/j.bpj.2012.03.043.

[58] Hiromi Sato and Jimmy B. Feix. Peptide-membrane interactions and mechanisms of membrane destruction by amphipathic alpha-helical antimicrobial peptides. *Biochimica et Biophysica Acta*, 1758:1245–1256, 2006. ISSN 00052736. doi: 10.1016/j.bbamem.2006.02.021.

[59] Julia H. Ortony, Christina J. Newcomb, John B. Matson, Liam C. Palmer, Peter E. Doan, Brian M. Hoffman, and Samuel I. Stupp. Internal dynamics of a supramolecular nanofibre. *Nature Materials*, 13(8):812–816, 2014. ISSN 14764660. doi: 10.1038/nmat3979.

[60] Carlos J. López, Shirley Oga, and Wayne L. Hubbell. Mapping molecular flexibility of proteins with site-directed spin labeling: A case study of myoglobin. *Biochemistry*, 51(33):6568–6583, 2012. ISSN 00062960. doi: 10.1021/bi3005686.

[61] Wayne L. Hubbell, Carlos J. López, Christian Altenbach, and Zhongyu Yang. Technological advances in site-directed spin labeling of proteins. *Current Opinion in Structural Biology*, 23:725–733, 2013. ISSN 0959440X. doi: 10.1016/j.sbi.2013.06.008.

[62] Christine B. Karim, Zhiwen Zhang, and David D. Thomas. Synthesis of TOAC spin-labeled proteins and reconstitution in lipid membranes. *Nature Protocols*, 2(1):42–49, 2007. ISSN 17542189. doi: 10.1038/nprot.2007.2.

[63] Shirley Schreier, José Carlos Bozelli, Nélida Marín, Renata F F Vieira, and Clóvis R. Nakaie. The spin label amino acid TOAC and its uses in studies of peptides: Chemical, physicochemical, spectroscopic, and conformational aspects. *Biophysical Reviews*, 4(1):45–66, 2012. ISSN 18672450. doi: 10.1007/s12551-011-0064-5.

[64] Stefan Stoll and Arthur Schweiger. EasySpin: Simulating cw ESR spectra. *Biol. Magn. Reson.*, 27:299–321, 2007. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

[65] Jack H. Freed. Theory of Slow Tumbling ESR Spectra for Nitroxides. In Lawrence J. Berliner, editor, *Spin Labeling: Theory and Applications*, chapter 3, pages 53–112. Academic Press, 1 edition, 1976.

[66] David E Budil, Sanghyuk Lee, Sunil Saxena, and Jack H Freed. Nonlinear-Least-Squares Analysis of Slow-Motion EPR Spectra in One and Two Dimensions Using a Modified Levenberg – Marquardt Algorithm. *Journal of Magnetic Resonance*, 120:155–189, 1996.

[67] Stefan Stoll and Arthur Schweiger. EasySpin , a comprehensive software package for spectral simulation and analysis in EPR. *Journal of Magnetic Resonance*, 178:42–55, 2006. doi: 10.1016/j.jmr.2005.08.013.

[68] William R Lindemann and Julia H Ortony. Probing local molecular properties of self-assembled systems with electron paramagnetic resonance (EPR). In

Helena S Azevedo and Riccardo M P da Silva, editors, *Self-Assembling Biomaterials: Molecular Design, Characterization and Application in Biology and Medicine*, number April, chapter 13. Woodhead Publishing, 1 edition, 2018. ISBN 9780081020159. doi: 10.3233/NRE-171453.

[69] David E. Budil. *CW-EPR Spectral Simulations: Slow-Motion Regime*, volume 563. Elsevier Inc., 1 edition, 2015. ISBN 9780128028346. doi: 10.1016/bs.mie.2015.05.024.

[70] Mark D Rabenstein and Yeon-Kyun Shin. Determination of the distance between two spin labels attached to a macromolecule. *Proc Natl Acad Sci U S A*, 92(August):8239–8243, 1995.

[71] Eric J Hustedt and Albert H Beth. Structural Information from CW-EPR Spectra of Dipolar Coupled Nitroxide Spin Labels. In Lawrence J Berliner, Gareth R Eaton, and Sandra S Eaton, editors, *Distance Measurements in Biological Systems by EPR*, volume 19, chapter 3, pages 155–184. Kluwer Academic/Plenum Publishers, New York, New York, 2000. ISBN 978-1-4757-0575-1. doi: 10.1007/978-0-306-47109-4.

[72] Jorg Fritscher, Mario Beyer, and Olav Schiemann. Synthesis , crystal structure and magnetic properties of a novel nitroxide biradical . Theoretical investigation of the exchange mechanisms. *Chemical Physics Letters*, 364(October):393–401, 2002.

[73] Gunnar W Reginsson and Olav Schiemann. Pulsed electron – electron double resonance : beyond nanometre distance measurements on biomacromolecules. *Biochemical Journal*, pages 353–363, 2011. doi: 10.1042/BJ20101871.

[74] Petr P Borbat and Jack H Freed. Multiple-quantum ESR and distance measurements. *Chemical Physics Letters*, 313(November):145–154, 1999.

[75] Peter P Borbat and Jack H Freed. Pros and cons of pulse dipolar ESR: DQC and DEER. *EPR Newsletter*, 17(2):21–33, 2007.

[76] Alexander J. Mijalis, Dale A. Thomas, Mark D. Simon, Andrea Adamo, Ryan Beaumont, Klavs F. Jensen, and Bradley L. Pentelute. A fully automated flow-based approach for accelerated peptide synthesis. *Nature Chemical Biology*, 13 (5):464–466, 2017. ISSN 15524469. doi: 10.1038/nchembio.2318.

[77] William R. Lindemann, Ethan D. Evans, Alexander J. Mijalis, Olivia M. Saouaf, Bradley L. Pentelute, and Julia H. Ortony. Quantifying residue-specific conformational dynamics of a highly reactive 29-mer peptide. *Scientific reports*, 10 (1):2597, 2020. ISSN 20452322. doi: 10.1038/s41598-020-59047-7.

[78] Asier Galán, Lubos Comor, Anita Horvatić, Josipa Kuleš, Nicolas Guillemin, Vladimir Mrljak, and Mangesh Bhide. Library-based display technologies: Where do we stand? *Molecular BioSystems*, 12(8):2342–2358, 2016. ISSN 17422051. doi: 10.1039/c6mb00219f.

[79] Ethan D Evans and Bradley L Pentelute. Discovery of a 29-amino-acid reactive abiotic peptide for selective cysteine arylation. *ACS Chemical Biology*, 13(3): 527–532, 2018. doi: 10.1021/acschembio.7b00520.

[80] Ethan D Evans and Bradley L Pentelute. Studies on a landscape of perfluoroaromatic-reactive peptides. *Organic & biomolecular chemistry*, 2018. doi: 10.1039/c8ob01678j.

[81] Ethan D. Evans, Zachary P. Gates, Zhen-yu J Sun, Alexander J. Mijalis, and Bradley L. Pentelute. Conformational stabilization and rapid labeling of a 29-residue peptide by a small molecule reaction partner. *Biochemistry*, 2019. ISSN 0006-2960. doi: 10.1021/acs.biochem.8b00940.

[82] Alexander M Spokoyny, Yekui Zou, Jingjing J Ling, Hongtao Yu, Yu-shan Lin, and Bradley L Pentelute. A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *Journal of the American Chemical Society*, 135:5946–5949, 2013. doi: 10.1021/ja400119t.

[83] Chi Zhang, Matthew Welborn, Tianyu Zhu, Nicole J. Yang, Michael S. Santos, Troy Van Voorhis, and Bradley L. Pentelute. π-Clamp-mediated cysteine conjugation. *Nature Chemistry*, 8(2):120–128, 2016. ISSN 17554349. doi: 10.1038/nchem.2413.

[84] S Kalhor-Monfared, M R Jafari, J T Patterson, P I Kitov, J J Dwyer, J M Nuss, and R Derda. Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone. *Chemical Science*, 7:3785–3790, 2016. doi: 10.1039/c5sc03856a.

[85] Eliana Saxon and Carolyn R Bertozzi. Cell Surface Engineering by a Modified Staudinger Reaction. *Science*, 287(2000):2007–2010, 2000. ISSN 00368075. doi: 10.1126/science.287.5460.2007.

[86] Vsevolod V. Rostovtsev, Luke G. Green, Valery V. Fokin, and K. Barry Sharpless. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie - International Edition*, 41(14):2596–2599, 2002. ISSN 14337851. doi: 10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4.

[87] Nicholas J. Agard, Jennifer A. Prescher, and Carolyn R. Bertozzi. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society*, 126 (46):15046–15047, 2004. ISSN 00027863. doi: 10.1021/ja044996f.

[88] Irwin Chen, Mark Howarth, Weiying Lin, and Alice Y. Ting. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nature Methods*, 2(2):99–104, 2005. ISSN 15487091. doi: 10.1038/nmeth735.

[89] Maximilian W. Popp, John M. Antos, Gijsbert M. Grotenbreg, Eric Spooner, and Hidde L. Ploegh. Sortagging: A versatile method for protein labeling. *Nature Chemical Biology*, 3(11):707–708, 2007. ISSN 15524469. doi: 10.1038/nchembio.2007.31.

[90] Marta Fernández-Suárez, Hemanta Baruah, Laura Martínez-Hernández, Kathleen T. Xie, Jeremy M. Baskin, Carolyn R. Bertozzi, and Alice Y. Ting. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nature Biotechnology*, 25(12):1483–1487, 2007. ISSN 10870156. doi: 10.1038/nbt1355.

[91] Georgyi V. Los, Lance P. Encell, Mark G. McDougall, Danette D. Hartzell, Natasha Karassina, Chad Zimprich, Monika G. Wood, Randy Learish, Rachel Friedman Ohana, Marjeta Urh, Dan Simpson, Jacqui Mendez, Kris Zimmerman, Paul Otto, Gediminas Vidugiris, Ji Zhu, Aldis Darzins, Dieter H. Klaubert, Robert F. Bulleit, and Keith V. Wood. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chemical Biology*, 3(6):373–382, 2008.

[92] Zhixing Chen, Chaoran Jing, Sarah S. Gallagher, Michael P. Sheetz, and Virginia W. Cornish. Second-generation covalent TMP-tag for live cell imaging. *Journal of the American Chemical Society*, 134(33):13692–13699, 2012. ISSN 00027863. doi: 10.1021/ja303374p.

[93] B. Albert Griffin, Stephen R. Adams, and Roger Y. Tsien. Specific covalent labeling of recombinant protein molecules inside live cells. *Science*, 281(5374): 269–272, 1998. ISSN 00368075. doi: 10.1126/science.281.5374.269.

[94] Suman Lata, Annett Reichel, Roland Brock, Robert Tampé, and Jacob Piehler. High-affinity adaptors for switchable recognition of histidine-tagged proteins. *Journal of the American Chemical Society*, 127(29):10205–10215, 2005. ISSN 00027863. doi: 10.1021/ja050690c.

[95] Brian V. Popp and Zachary T. Ball. Structure-selective modification of aromatic side chains with dirhodium metallopeptide catalysts. *Journal of the American Chemical Society*, 132(19):6660–6662, 2010. ISSN 00027863. doi: 10.1021/ja101456c.

[96] Ulrike Reinhardt, Jonathan Lotze, Sarah Zernia, Karin Mörl, Annette G. Beck-Sickinger, and Oliver Seitz. Peptide-templated acyl transfer: A chemical method for the labeling of membrane proteins on live cells. *Angewandte Chemie - International Edition*, 53(38):10237–10241, 2014. ISSN 15213773. doi: 10.1002/anie.201403214.

[97] Takashi Ishida and Kengo Kinoshita. PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research*, 35(SUPPL.2):460–464, 2007. ISSN 03051048. doi: 10.1093/nar/gkm363.

[98] David J. Schneider and Jack H. Freed. Calculating slow motional magnetic resonance spectra: a user's guide, 1989.

[99] Zhichun C Liang and Jack H Freed. An assessment of the applicability of multifrequency ESR to study the complex dynamics of biomolecules. *Journal of Physical Chemistry B*, 103:6384–6396, 1999. ISSN 1520-6106. doi: Doi 10.1021/Jp9907746.

[100] Wayne L. Hubbell and Christian Altenbach. Investigation of structure and dynamics in membrane proteins using site-directed spin labeling. *Current Opinion in Structural Biology*, 4:566–573, 1994. ISSN 0959440X. doi: 10.1016/S0959-440X(94)90219-4.

[101] Julia H. Ortony, Chi Yuan Cheng, John M. Franck, Ravinath Kausik, Anna Pavlova, Jasmine Hunt, Songi Han, Subrata Pal, Prabal K Maiti, Marina Bennati, Thomas F Prisner, and Yi-qiao Song. Probing the hydration water diffusion of macromolecular surfaces and interfaces. *New Journal of Physics*, 13: 015006, 2011. ISSN 13672630. doi: 10.1088/1367-2630/13/1/015006.

[102] John M Franck, Anna Pavlova, John A Scott, and Songi Han. Quantitative cw Overhauser effect dynamic nuclear polarization for the analysis of local water dynamics. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 74:33–56, 2013. ISSN 0079-6565. doi: 10.1016/j.pnmrs.2013.06.001.

[103] Lu Yu, Wei Wang, Shenglong Ling, Sanling Liu, Liang Xiao, Yanlong Xin, Chaohua Lai, Ying Xiong, Longhua Zhang, and Changlin Tian. CW-EPR studies revealed different motional properties and oligomeric states of the integrin β1atransmembrane domain in detergent micelles or liposomes. *Scientific Reports*, 5:1–11, 2015. ISSN 20452322. doi: 10.1038/srep07848.

[104] Cholpon Tilegenova, Benjamin W Elberson, D. Marien Cortes, and Luis G. Cuello. CW-EPR Spectroscopy and Site-Directed Spin Labeling to Study the Structural Dynamics of Ion Channels. In William A Coetzee and Bernardo Rudy, editors, *Potassium Channels*, chapter 21, pages 279–288. Springer Science+Business Media, LLC, New York, 1 edition, 2018. ISBN 9781493973613. doi: 10.1038/npg.els.0005670.

[105] C. Toniolo, M. Crisma, and F. Formaggio. TOAC, a nitroxide spin-labeled, achiral C-alpha -tetrasubstituted alpha-amino acid, is an excellent tool in material science and biochemistry. *Biopolymers - Peptide Science Section*, 47:153–158, 1998. ISSN 00063525. doi: 10.1002/(SICI)1097-0282(1998)47:2<153::AID-BIP4>3.0.CO;2-T.

[106] Michael J Mitchell, Patrick L Heider, Andrea Adamo, Alexander A Vinogradov, Surin K Mong, Xiyuan Li Li, Tatiana Berger, Rocco L Policarpo, Chi Zhang, Yekui Zou, Xiaoli Liao, Alexander M Spokoyny, Klavs F Jensen, and Bradley L Pentelute. Rapid Flow-Based Peptide Synthesis. *ChemBioChem*, 15(5):713–720, 2014. ISSN 15378276. doi: 10.1088/1367-2630/15/1/015008.Fluid.

[107] Rita Guzzi, Mohammad Babavali, Rosa Bartucci, Luigi Sportelli, Mikael Esmann, and Derek Marsh. Biochimica et Biophysica Acta Spin-echo EPR of Na , K-ATPase unfolding by urea. *BBA - Biomembranes*, 1808(6):1618–1628, 2011. ISSN 0005-2736. doi: 10.1016/j.bbamem.2010.11.008.

[108] Derek Marsh, Rosa Bartucci, Rita Guzzi, Luigi Sportelli, and Mikael Esmann. Librational fluctuations in protein glasses. *Biochimica et Biophysica Acta*, 1834: 1591–1595, 2013. doi: 10.1016/j.bbapap.2013.05.001.

[109] Rita Guzzi, Rosa Bartucci, and Derek Marsh. Heterogeneity of protein substates visualized by spin-label EPR. *Biophysical Journal*, 106(3):716–722, 2014. ISSN 00063495. doi: 10.1016/j.bpj.2013.12.039.

[110] R. Zwanzig. Diffusion in a rough potential. *Proceedings of the National Academy of Sciences*, 85(7):2029–2030, 1988. ISSN 0027-8424. doi: 10.1073/pnas.85.7.2029.

[111] Ruti Kapon, Reinat Nevo, and Ziv Reich. Protein energy landscape roughness. *Biochemical Society transactions*, 36(Pt 6):1404–1408, 2008. ISSN 0300-5127. doi: 10.1042/BST0361404.

[112] L. Milanesi, J. P. Waltho, C. A. Hunter, D. J. Shaw, G. S. Beddard, G. D. Reid, S. Dev, and M. Volk. Measurement of energy landscape roughness of folded and unfolded proteins. *Proceedings of the National Academy of Sciences*, 109(48): 19563–19568, 2012. ISSN 0027-8424. doi: 10.1073/pnas.1211764109.

[113] Martin Volk, Lilia Milanesi, Jonathan P. Waltho, Christopher A. Hunter, and Godfrey S. Beddard. The roughness of the protein energy landscape results in anomalous diffusion of the polypeptide backbone. *Physical Chemistry Chemical Physics*, 17(2):762–782, 2015. ISSN 14639076. doi: 10.1039/c4cp03058c.

[114] David de Sancho, Anshul Sirur, and Robert B Best. Molecular origins of internal friction effects on protein folding rates. *Nature Communications*, 79(2):211–227, 2015. ISSN 1527-5418. doi: 10.1177/000312413519445.Are.

[115] Wenwei Zheng, David De Sancho, and Robert B. Best. Modulation of Folding Internal Friction by Local and Global Barrier Heights. *Journal of Physical Chemistry Letters*, 7(6):1028–1034, 2016. ISSN 19487185. doi: 10.1021/acs.jpclett.6b00329.

[116] Reinat Nevo, Vlad Brumfeld, Ruti Kapon, Peter Hinterdorfer, and Ziv Reich. Direct measurement of protein energy landscape roughness. *EMBO Reports*, 6 (5):6–10, 2005. doi: 10.1038/sj.embor.7400403.

[117] William R Lindemann, Alexander J Mijalis, José L Alonso, Peter P Borbat, Jack H Freed, M Amin Arnaout, Bradley L Pentelute, and Julia H Ortony. Conformational dynamics in extended-RGD binding peptides (Accepted Manuscript). *Biomacromolecules*, 2020.

[118] E F Plow, M D Pierschbacher, E Ruoslahti, G A Marguerie, and M H Ginsberg. The effect of Arg-Gly-Asp-containing peptides on fibrinogen and von Willebrand factor binding to platelets. *Proceedings of the National Academy of Sciences of the United States of America*, 82(23):8057–61, 1985. ISSN 0027-8424. doi: 10.1073/pnas.82.23.8057.

[119] Edward F. Plow, Thomas A. Haas, Li Zhang, Joseph Loftus, and Jeffrey W. Smith. Ligand binding to integrins. *Journal of Biological Chemistry*, 275(29): 21785–21788, 2000. ISSN 00219258. doi: 10.1074/jbc.R000003200.

[120] Yoshikazu Takada, Xiaojing Ye, and Scott Simon. The integrins. *Genome Biology*, 8(5), 2007. ISSN 14747596. doi: 10.1186/gb-2007-8-5-215.

[121] Jay S. Desgrosellier and David A. Cheresh. Integrins in cancer: Biological implications and therapeutic opportunities. *Nature Reviews Cancer*, 10(1):9–22, 2010. ISSN 1474175X. doi: 10.1038/nrc2748.

[122] Klaus Ley, Jesus Rivera-Nieves, William J Sandborn, and Sanford Shattil. Integrin-based therapeutics: biological basis, clinical use and new drugs. *Nature reviews. Drug discovery*, 15:173–183, 2016.

[123] Tobias G. Kapp, Florian Rechenmacher, Stefanie Neubauer, Oleg V. Maltsev, Elisabetta A. Cavalcanti-Adam, Revital Zarka, Ute Reuning, Johannes Notni, Hans Jürgen Wester, Carlos Mas-Moruno, Joachim Spatz, Benjamin Geiger, and Horst Kessler. A comprehensive evaluation of the activity and selectivity profile of ligands for RGD-binding integrins. *Scientific Reports*, 7(November 2016):1–13, 2017. ISSN 20452322. doi: 10.1038/srep39805.

[124] Yu Ping Yu, Qi Wang, Ying Chun Liu, and Ying Xie. Molecular basis for the targeted binding of RGD-containing peptide to integrin $\alpha V\beta 3$. *Biomaterials*, 35 (5):1667–1675, 2014. ISSN 01429612. doi: 10.1016/j.biomaterials.2013.10.072.

[125] Y. Gilad, E. Noy, H. Senderowitz, A. Albeck, M. A. Firer, and G. Gellerman. Synthesis, biological studies and molecular dynamics of new anticancer RGD-based peptide conjugates for targeted drug delivery. *Bioorganic and Medicinal Chemistry*, 24(2):294–303, 2016. ISSN 14643391. doi: 10.1016/j.bmc.2015.12.020.

[126] Sean M. McHugh, Julia R. Rogers, Sarah A. Solomon, Hongtao Yu, and Yu Shan Lin. Computational methods to design cyclic peptides. *Current Opinion in Chemical Biology*, 34:95–102, 2016. ISSN 18790402. doi: 10.1016/j.cbpa.2016.08.004.

[127] Johannes F Van Agthoven, Jian-Ping Xiong, José Luis Alonso, Xianliang Rui, Brian D Adair, Simon L Goodman, and M. Amin Arnaout. Structural basis for pure antagonism of integrin $\alpha V\beta 3$ by a high affinity form of fibronectin. *Nat Struct Mol Biol*, 21(4):383–388, 2014. doi: 10.1038/nsmb.2797.Structural.

[128] André Krammer, David Craig, Wendy E. Thomas, Klaus Schulten, and Viola Vogel. A structural model for force regulated integrin binding to fibronectin's RGD-synergy site. *Matrix Biology*, 21(2):139–147, 2002. ISSN 0945053X. doi: 10.1016/S0945-053X(01)00197-4.

[129] Hongbo Xie, Slobodan Vucetic, Lilia M Iakoucheva, Christopher J Oldfield, A Kieth Dunker, Vladimir N Uversky, and Zoran Obradovic. Functional anthology of intrinsic disorder. I. Biological processes and functions of proteins with long disordered regions. *Journal of Proteome Research*, 6(5):1882–1898, 2008. doi: 10.1021/pr060392u.Functional.

[130] Daniel J. Leahy, Ikramuddin Aukhil, and Harold P. Erickson. 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell*, 84(1):155–164, 1996. ISSN 00928674. doi: 10.1016/S0092-8674(00)81002-8.

[131] Daniel Monleón, M. Paz Moreno-Murciano, Helena Kovacs, Cezary Marcinkiewicz, Juan J. Calvete, and Bernardo Celda. Concerted Motions of the Integrin-binding Loop and the C-terminal Tail of the Non-RGD Disintegrin Obtustatin. *Journal of Biological Chemistry*, 278(46):45570–45576, 2003. ISSN 00219258. doi: 10.1074/jbc.M307030200.

[132] Daniel Monleón, Vicent Esteve, Helena Kovacs, Juan J Calvete, and Bernardo Celda. Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. *The Biochemical journal*, 387(Pt 1):57–66, 2005. ISSN 1470-8728. doi: 10.1042/BJ20041343.

[133] Isabella Daidone, Massimiliano Aschi, Maria Patamia, Argante Bozzi, and Raffaele Petruzzelli. Structural and dynamical properties of KTS-disintegrins: A comparison between Obtustatin and Lebestatin. *Biopolymers*, 99(1):47–54, 2013. ISSN 00063525. doi: 10.1002/bip.22138.

[134] Luisa Calvanese, Lucia Falcigno, and Gabriella D Auria. Essential Dynamics Analysis Captures the Concerted Motion of the Integrin-Binding Site in Jerdostatin, an RTS Disintegrin. *Biopolymers*, 103(3):158–166, 2014. doi: 10.1002/bip.22578.

[135] Kannan Gunasekaran, Chung Jung Tsai, Sandeep Kumar, David Zanuy, and Ruth Nussinov. Extended disordered proteins: Targeting function with less scaffold. *Trends in Biochemical Sciences*, 28(2):81–85, 2003. ISSN 09680004. doi: 10.1016/S0968-0004(03)00003-3.

[136] V. J. Hilser and E. B. Thompson. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proceedings of the National Academy of Sciences*, 104(20):8311–8315, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0700329104.

[137] D. E. Koshland. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences*, 44(2):98–104, 1958. ISSN 0027-8424. doi: 10.1073/pnas.44.2.98.

[138] A S Burgen. Conformational changes and drug action. *Federation Proceedings*, 40(13):2723–2728, 1981.

[139] G Schreiber, G Haran, and H.-X Zhou. Fundamental aspects of protein-protein association kinetics. *Chemical Reviews*, 109(3):839–860, 2009. doi: 10.1021/cr800373w.Fundamental.

[140] Peter Tompa. Intrinsically disordered proteins : a 10-year recap. *Trends in Biochemical Sciences*, 37(12):509–516, 2012. ISSN 0968-0004. doi: 10.1016/j.tibs.2012.08.004.

[141] Johannes F. Van Agthoven, Hengameh Shams, Frank V. Cochran, José L. Alonso, James R. Kintzing, Kiavash Garakani, Brian D. Adair, Jian Ping Xiong, Mohammad R.K. Mofrad, Jennifer R. Cochran, and M. Amin Arnaout. Structural Basis of the Differential Binding of Engineered Knottins to Integrins αVβ3 and α5β1. *Structure*, 27(9):1443–1451.e6, 2019. ISSN 18784186. doi: 10.1016/j.str.2019.06.011.

[142] Uwe Lappan, Brigitte Wiesner, and Ulrich Scheler. Rotational Dynamics of Spin-Labeled Polyacid Chain Segments in Polyelectrolyte Complexes Studied by CW EPR Spectroscopy. *Macromolecules*, 48(11):3577–3581, 2015. ISSN 15205835. doi: 10.1021/acs.macromol.5b00474.

[143] Uwe Lappan, Brigitte Wiesner, and Ulrich Scheler. Segmental dynamics of poly(acrylic acid) in polyelectrolyte complex coacervates studied by spin-label EPR spectroscopy. *Macromolecules*, 49(22):8616–8621, 2016. ISSN 15205835. doi: 10.1021/acs.macromol.6b01863.

[144] Agnieszka Skowronska, P. Mlotkowska, S. Okrasa, S. Nielsen, and M. T. Skowronski. Modulatory effects of steroid hormones, oxytocin, arachidonic acid, forskolin and cyclic AMP on the expression of aquaporin 1 and aquaporin 5 in the porcine uterus during placentation. *Journal of Physiology and Pharmacology*, 67(2):311–319, 2016. ISSN 18991505. doi: 10.1016/j.bbabio.2015.10.010.

[145] Vasily S. Oganesyan, Fatima Chami, Gaye F. White, and Andrew J. Thomson. A combined EPR and MD simulation study of a nitroxyl spin label with restricted internal mobility sensitive to protein dynamics. *Journal of Magnetic Resonance*, 274:24–35, 2017. ISSN 10960856. doi: 10.1016/j.jmr.2016.11.001.

[146] Peter P. Borbat, Elka R. Georgieva, and Jack H. Freed. Improved sensitivity for long-distance measurements in biomolecules: Five-pulse double electron-electron resonance. *Journal of Physical Chemistry Letters*, 4(1):170–175, 2013. ISSN 19487185. doi: 10.1021/jz301788n.

[147] Peter P. Borbat and Jack H. Freed. Dipolar spectroscopy - single-resonance methods. *eMagRes*, 6(4):465–494, 2017. ISSN 20556101. doi: 10.1002/9780470034590.emrstm1519.

[148] Elka R. Georgieva, Peter P. Borbat, Haley D. Norman, and Jack H. Freed. Mechanism of influenza A M2 transmembrane domain assembly in lipid membranes. *Scientific Reports*, 5:11757, 2015. ISSN 20452322. doi: 10.1038/srep11757.

[149] Yun Wei Chiang, Peter P. Borbat, and Jack H. Freed. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *Journal of Magnetic Resonance*, 172(2):279–295, 2005. ISSN 10907807. doi: 10.1016/j.jmr.2004.10.012.

[150] Yun Wei Chiang, Peter P. Borbat, and Jack H. Freed. Maximum entropy: A complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR. *Journal of Magnetic Resonance*, 177(2):184–196, 2005. ISSN 10907807. doi: 10.1016/j.jmr.2005.07.021.

[151] Madhur Srivastava, Elka R Georgieva, and Jack H Freed. A New Wavelet Denoising Method for Experimental Time-Domain Signals: Pulsed Dipolar Electron Spin Resonance Madhur. *J Phys Chem A*, 121(12):2452–2465, 2017. doi: 10.1021/acs.jpca.7b00183.A.

[152] Madhur Srivastava and Jack H Freed. Singular Value Decomposition Method to Determine Distance Distributions in Pulsed Dipolar Electron Spin Resonance. *J Phys Chem Letters*, 8(22):5648–5655, 2017. doi: 10.1021/acs.jpclett.7b02379.Details.

[153] Pierre Thévenet, Yimin Shen, Julien Maupetit, Frédéric Guyon, Philippe Derreumaux, and Pierre Tufféry. PEP-FOLD: An updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Research*, 40(W1):288–293, 2012. ISSN 03051048. doi: 10.1093/nar/gks419.

[154] Alexis Lamiable, Pierre Thévenet, Julien Rey, Marek Vavrusa, Philippe Derreumaux, and Pierre Tufféry. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*, 44(W1):W449–W454, 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw329.

[155] Gunnar Jeschke. DEER Distance Measurements on Proteins. *Annual Review of Physical Chemistry*, 63(1):419–446, 2012. ISSN 0066-426X. doi: 10.1146/annurev-physchem-032511-143716.

[156] Chi-Yuan Cheng and Songi Han. Dynamic Nuclear Polarization Methods in Solids and Solutions to Explore Membrane Proteins and Membrane Systems. *Annual Review of Physical Chemistry*, 64:507–532, 2013. ISSN 0066-426X. doi: 10.1146/annurev-physchem-040412-110028.

[157] E J Hustedt and A H Beth. Nitroxide spin-spin interactions: applications to protein structure and dynamics. *Annual review of biophysics and biomolecular structure*, 28:129–153, 1999. ISSN 1056-8700. doi: 10.1146/annurev.biophys.28.1.129.

[158] Sushil K Mitra and Jack H Freed. Molecular Motions. In Sushil K. Misra, editor, *Multifrequency Electron Paramagnetic Resonance*, chapter 11, pages 497–543. Wiley, Weinheim, 2011. ISBN 9783527407798.

[159] Yan Xia, Yongjun Li, Alan O. Burts, M. Francesca Ottaviani, David A. Tirrell, Jeremiah A. Johnson, Nicholas J. Turro, and Robert H. Grubbs. EPR study of spin labeled brush polymers in organic solvents. *Journal of the American Chemical Society*, 133(49):19953–19959, 2011. ISSN 00027863. doi: 10.1021/ja2085349.

[160] Julia H. Ortony, Dong Soo Hwang, John M. Franck, J. Herbert Waite, and Songi Han. Asymmetric collapse in biomimetic complex coacervates revealed by local polymer and water dynamics. *Biomacromolecules*, 14(5):1395–1402, 2013. ISSN 15264602. doi: 10.1021/bm4000579.

[161] Christina J. Newcomb, Shantanu Sur, Julia H. Ortony, One Sun Lee, John B. Matson, Job Boekhoven, Jeong Min Yu, George C. Schatz, and Samuel I. Stupp. Cell death versus cell survival instructed by supramolecular cohesion of nanostructures. *Nature Communications*, 5:1–10, 2014. ISSN 20411723. doi: 10.1038/ncomms4321.

[162] Christian Altenbach. Research Website of Wayne Hubbell, 2010.

[163] Eva Meirovitch, Akbar Nayeem, and Jack H Freed. Analysis of Protein-Lipid Interactions Based on Model Simulations of Electron Spin Resonance Spectra. *Journal of Physical Chemistry B*, 88:3454–3465, 1984. ISSN 0022-3654. doi: 10.1021/j150660a018.

[164] Heinz-Jürgen Steinhoff, Matthias Müller, Christian Beier, and Matthias Pfeiffer. Molecular dynamics simulation and EPR spectroscopy of nitroxide side chains in bacteriorhodopsin. *Journal of Molecular Liquids*, 84:17–27, 2000. ISSN 01677322. doi: 10.1016/S0167-7322(99)00107-5.

[165] Leslie E W LaConte, Vincent Voelz, Wendy Nelson, Michael Enz, and David D Thomas. Molecular dynamics simulation of site-directed spin labeling: experimental validation in muscle fibers. *Biophysical journal*, 83:1854–1866, 2002.

[166] Mirco Zerbetto, Daniele Licari, Vincenzo Barone, and Antonino Polimeno. Computational tools for the interpretation of electron spin resonance spectra in solution. *Molecular Physics*, 111:2746–2756, 2013. ISSN 0026-8976. doi: 10.1080/00268976.2013.800602.

[167] Peter Martin, Stefan Stoll, and David Thomas. Simulating Electron Paramagnetic Resonance Spectra of Slow-Motion Systems in the Time Domain. *Biophysical Journal*, 112(3):445a–446a, 2017. ISSN 00063495. doi: 10.1016/j.bpj.2016.11.2388.

[168] Harvey Motulsky and Arthur Christopoulos. *Fitting Models to Biological Data using Linear and Nonlinear Regression.* GraphPad Software, Inc, San Diego, CA, 2 edition, 2003.

[169] Diego Fernandez Slezak, Cecilia Suarez, Guillermo A Cecchi, Guillermo Marshall, and Gustavo Stolovitzky. When the Optimal Is Not the Best : Parameter Estimation in Complex Biological Models. *PLoS ONE*, 5(10):e13283, 2010. doi: 10.1371/journal.pone.0013283.

[170] Ty Christoff-tempesta, Yukio Cho, Dae-yoon Kim, Andrew J Lew, Guillaume Lamour, Xiaobing Zuo, William R Lindemann, and Julia H Ortony. Self-assembly of aramid amphiphiles into ultra-stable nanofibers and aligned nanofiber threads. 2020.

[171] Sandrine Valsesia-wittmann. Role of Chimeric Murine Leukemia Virus env β-Turn Polyproline Spacers in Receptor Cooperation. *Journal of Virology*, 75(18): 8478–8486, 2001. doi: 10.1128/JVI.75.18.8478.

[172] Diego Pallarola, Alexander Bochen, Heike Boehm, Florian Rechenmacher, Tariq R Sobahi, Joachim P Spatz, and Horst Kessler. Interface Immobilization Chemistry of c RGD-based Peptides Regulates Integrin Mediated Cell Adhesion. *Advanced Functional Materials*, 24:943–956, 2014. doi: 10.1002/adfm.201302411.

[173] Soren Doose, Hannes Neuweiler, Hannes Barsch, and Markus Sauer. Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *PNAS*, 104 (44):17400–17405, 2007.

[174] Mahmoud Moradi, Volodymyr Babin, Christopher Roland, Thomas A Darden, and Celeste Sagui. Conformations and free energy landscapes of polyproline peptides. *PNAS*, 106(49):20746–20751, 2009.

[175] Thomas Weise. *Global Optimization Algorithms – Theory and Application.* Citeseer, 2008.

[176] Alexei A. Adzhubei, Michael J.E. Sternberg, and Alexander A. Makarov. Polyproline-II helix in proteins: Structure and function. *Journal of Molecular Biology*, 425(12):2100–2132, 2013. ISSN 10898638. doi: 10.1016/j.jmb.2013.03.018.

[177] Weidang Li, Medha Joshi, Smita Singhania, Kyle Ramsey, and Ashlesh Murthy. Peptide Vaccine: Progress and Challenges. *Vaccines*, 2(3):515–536, 2014. ISSN 2076-393X. doi: 10.3390/vaccines2030515.

[178] Kelly M Stewart, Kristin L Horton, and Shana O Kelley. Cell-penetrating peptides as delivery vehicles for biology and medicine. *Organic & biomolecular chemistry*, 6(13):2242–55, 2008. ISSN 1477-0520. doi: 10.1039/b719950c.

[179] Yanping Sun, Lei Xian, Haonan Xing, Jiankun Yu, Zhen Yang, Tianzhi Yang, Li Yang, and Pingtian Ding. Factors influencing the nuclear targeting ability of nuclear localization signals. *Journal of Drug Targeting*, 24(10):927–933, 2016. ISSN 10292330. doi: 10.1080/1061186X.2016.1184273.