Review Article

The contribution of modern EPR to structural biology

Author(s):
Jeschke, Gunnar

Publication Date:
2018-02-06

Permanent Link:
https://doi.org/10.3929/ethz-b-000250152

Originally published in:
Emerging Topics in Life Sciences, http://doi.org/10.1042/etls20170143

Rights / License:
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Review Article

The contribution of modern EPR to structural biology

Gunnar Jeschke
Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 2, CH-8093 Zürich, Switzerland

Correspondence: Gunnar Jeschke (gjeschke@ethz.ch)

Electron paramagnetic resonance (EPR) spectroscopy combined with site-directed spin labelling is applicable to biomolecules and their complexes irrespective of system size and in a broad range of environments. Neither short-range nor long-range order is required to obtain structural restraints on accessibility of sites to water or oxygen, on secondary structure, and on distances between sites. Many of the experiments characterize a static ensemble obtained by shock-freezing. Compared with characterizing the dynamic ensemble at ambient temperature, analysis is simplified and information loss due to overlapping timescales of measurement and system dynamics is avoided. The necessity for labelling leads to sparse restraint sets that require integration with data from other methodologies for building models. The double electron–electron resonance experiment provides distance distributions in the nanometre range that carry information not only on the mean conformation but also on the width of the native ensemble. The distribution widths are often inconsistent with Anfinsen’s concept that a sequence encodes a single native conformation defined at atomic resolution under physiological conditions.

Introduction

During the past two decades, electron paramagnetic resonance (EPR) spectroscopy combined with site-directed spin labelling [1,2] has become an important tool in structural biology that complements higher-resolution techniques, such as X-ray crystallography (XRD), cryo-electron microscopy (cryo-EM), and NMR (nuclear magnetic resonance) [3,4]. The advent of integrative structural biology [5] has increased demand for EPR restraints on structure and requires better understanding of their accuracy and precision. Two types of EPR restraints are suitable for model building: accessibility restraints [6] and distance distribution restraints in the nanometre range [7,8]. Accessibility of a site to water or oxygen and spin label mobility are periodic in a site scan along a secondary structure element [9]. Label mobility influences the EPR line shape, allowing to study conformational equilibria and the kinetics of conformational change [10,11].

This review focuses on distance distribution restraints obtained with standard methanethiosulfonate spin labels (MTSSLs). Such distance distribution restraints are a unique contribution of EPR to structural biology. They encode information on the width of a conformational ensemble and thus on order–disorder phenomena. Evidence is growing that such phenomena underlie mechanisms of function and regulation of many proteins. To study them, methodologies for determining a single structure at atomic resolution must be complemented by methodologies that yield information on the conformational ensemble in an approximately native environment. Measurements of distance distributions by EPR techniques have already provided such information for many systems. Ongoing method development is extending the scope for such studies and will offer access to better approximations of the native environment.
EPR distance distribution restraints
Application of pulsed dipolar spectroscopy to biomolecules

Spin labels carry magnetic dipoles. The distribution of dipole–dipole couplings can be measured by pulsed dipolar spectroscopy (PDS) [12] if the exchange coupling between the two spins is negligible, as is the case at distances longer than 15 Å. Reorientation of the spin–spin vector must be slow on the timescale of the dipolar interaction, which extends from 100 ns at a distance of 20 Å to 100 μs at a distance of 160 Å. Limited excitation bandwidth sets a lower limit of the reliable distance range at 18–20 Å. The upper limit depends on the maximum observation time of the spins and thus on their decoherence time. For typical nitroxide labels, the decoherence time increases with decrease in temperature down to ~60–40 K. At such temperatures, distances up to at least 50 Å can be measured for membrane proteins and the range extends up to 160 Å for soluble proteins in favourable cases [13] if both the solvent and the protein are deuterated [14]. For ambient temperature measurements on immobilized biomolecules labelled with slowly relaxing trityl radicals [15,16], the upper distance limit reduces to 30–50 Å. Whether biomolecule conformation is more strongly perturbed by immobilization at ambient temperature or by shock-freezing in the presence of a cryoprotectant is still a matter of debate and may depend on the system under investigation. It is advisable to check consistency between EPR data obtained at cryogenic temperatures and data from other techniques obtained at ambient temperature, whenever this is possible.

While the necessity for cryogenic temperatures is a matter of concern, among all techniques in structural biology EPR spectroscopy probably poses the least restrictions on the environment of the biomolecules. Most biomolecules and small ligands are EPR-silent and, since distance distributions are measured, neither long-range nor short-range order is required. PDS can be applied to membrane proteins reconstituted into liposomes and even to cells [17–19]. The only requirement posed on the biomolecules is that they can be spin-labelled or bind a spin-labelled ligand.

Usually, the distribution of dipolar couplings is measured by the double electron–electron resonance (DEER) experiment that is also called pulsed electron–electron double resonance (PELDOR) [7]. For this experiment, the functional form of the background contribution from remote spin-labelled molecules is known and can thus be separated from the intramolecular contribution. If local concentration of spin-labelled biomolecules is known to be sufficiently low, single-frequency PDS techniques, such as the double-quantum coherence (DQC) experiment, the single-frequency technique for refocusing (SIFTER), or relaxation-induced dipolar modulation enhancement (RIDME), can also be applied and may provide higher sensitivity under some conditions.

From primary data to distance distributions

Except for orientation selection, which is often negligible or can be averaged [4], the normalized dipolar evolution function $D(t)$ can be computed from the distance distribution $P(r)$ by relying only on fundamental constants and the known $g$ value of the spin label. The normalized primary signal $V(t)$ is given by $B(t) [1 - \lambda + \lambda D(t)]$. Modulation depth $\lambda$ depends on experimental parameters and the background is usually a stretched exponential function $B(t) = \exp[-(kt)^\beta]$, with $\xi = 1$ for soluble proteins and $2/3 < \xi < 1$ for membrane proteins. The inverse problem of computing $P(r)$ from $V(t)$ is fraught with two complications. First, reliable fitting of the background parameters $k$ and $\lambda$ and possibly $\xi$ requires that $V(t)$ be observed up to 2 μs for distances up to 4 nm. The mean distance can be estimated up to 5 μs for such data, but between 4 and 5 nm the width of the distribution is unreliable. Background correction usually suppresses distances longer than 5 nm for a data trace length of 2 μs, but can also generate artefact peaks at long distances. The limits scale with the cubic root of the trace length. The most widely used software package DeerAnalysis [20] colour-codes ranges where information on the distribution shape (green), width (yellow), and mean distance (orange) is still reliable.

As a second complication, conversion of $D(t)$ into $P(r)$ is a mathematically ill-posed problem. Tikhonov regularization stabilizes the solution without imposing restraints on the distribution shape and at the expense of only weak or moderate broadening. However, if both narrow and broad peaks occur, as is the case for equilibria between ordered and disordered conformational sub-ensembles, regularization either strongly broadens the narrow peaks or cannot suppress artificial splitting of broad peaks. Fitting by a few Gaussian peaks, also implemented in DeerAnalysis and supplemented by statistical analysis in GLADD [21], may then be advantageous. An overview of further software packages is provided in ref. [21].

Ill-posedness of converting $D(t)$ into $P(r)$ and uncertainty in background separation require validation of distance distributions, as implemented in DeerAnalysis, or statistical error analysis, as implemented in GLADD.
Validation of the conversion of \( D(t) \) into \( P(r) \) by a Bayesian approach has been demonstrated [22]. The ill-posed problem can be avoided and background correction stabilized by fitting structure models to primary data \( V(t) \) [23]. Often, restraints must be supplied in terms of mean distances and standard deviations of the distance. It is then advisable to refine models against primary data or to check their consistency with primary data.

**From distance distributions to models**

For dilute biomolecules labelled at two sites, PDS provides the distribution \( P(r_{ee}) \) of the distance \( r_{ee} \) between the unpaired electrons of the two spin labels. If the unpaired electrons are delocalized on length scales comparable to \( r_{ee} \), \( P(r_{ee}) \) corresponds to an \( 1/r^3 \)-weighted average over their spatial distribution. For the commonly used nitroxide labels, this complication can be safely neglected at distances \( r > 15 \text{ Å} \), but for trityl spin labels it needs to be taken into account up to 30 Å [24].

For structure modelling, the distance \( r_{aa} \) between the Ca atoms of the labelled residues or an analogous backbone–backbone distance for nucleotides are of interest. Since even for the smallest nitroxide labels, \( r_{ee} \) may differ from \( r_{aa} \) by up to 10 Å, it is mandatory to include the spin label into modelling. The possible conformations of the label side chain can be approximated by a moderate number of rotameric states [23,25]. In the statistical mechanics-based approach in the software package MMM [25,26], the rotamer distribution of the free label is estimated once for all from molecular dynamics (MD) trajectories or by Monte Carlo sampling and stored in a library. The influence of the biomolecule on this distribution is computed by adding non-bonding interaction energies with protein, nucleic acid, and cofactor atoms to the free energy. A variant estimates the rotamer distribution from a set of crystal structures of spin-labelled proteins [27], and a simpler but also less physical approach samples the accessible volume by taking into account only repulsive interactions and ignoring the torsion potentials [28]. All these approaches end up with similar accuracy of \( \sim 3 \text{ Å} \) for predicting the mean distance \( \langle r_{ee} \rangle \) from a known high-resolution structure [29]. Computationally more elaborate procedures sample the energy hypersurface of the labelled protein in a Monte Carlo approach [30] or by enhanced sampling MD [31], but do not improve strongly on that accuracy. Hence, at present, an accuracy of \( \sim 2–3 \text{ Å} \) must be accepted, which precludes building atomic-resolution models from only EPR restraints.

Forward computation of \( r_{ee} \) from a structural model requires information on only local structure around the labelling sites. This simplifies the treatment of large-scale conformational transitions, where rigid domains move with respect to each other [4], and of rigid-body docking [23,26,32]. In integrative modelling, the approximation of the local environment can be iteratively improved [33]. Rotamer library-based predictions are a computationally inexpensive approach for testing many models in ensemble refinement [34] and for simulating PDS data from MD trajectories of unlabelled biomolecules with the Python package RotamerConvolveMD [35].

**Model building with EPR restraints**

**Stand-alone modelling**

Typically, only one or two EPR restraints are obtained per expressed and purified mutated protein sample. Therefore, the number of EPR restraints is much smaller than that of residues and building of a model from only these restraints is precluded even on a coarse-grained level. In that sense, structure modelling with EPR restraints is always hybrid modelling. Stand-alone modelling denotes approaches based on experimental structures from XRD, NMR, or cryo-EM or on structures predicted from homology or de novo modelling without fitting other experimental data simultaneously. It is assumed that processing of experimental data from different methodologies is separable, in contrast with integrative modelling, where this is not the case. Stand-alone modelling presumes that the EPR data only weakly or only locally perturb the initial structure, which is often a good approximation for biomolecules containing rigid domains.

A case in point is binding of small-molecule ligands or metal ions. If the ligand can be spin-labelled or the metal ion is paramagnetic, they can be localized by a distance matrix approach [36] or by an approach akin to the global positioning system (GPS), where the position of an observer can be determined by at least three distance measurements to reference sites [37]. The distance matrix approach tests and relaxes the assumption of a rigid reference domain at the expense of additional measurements between reference points. If rigidity of the reference domain has been established by NMR or small-angle X-ray scattering (SAXS) measurements, the GPS-like approach may be advantageous. Both approaches have been implemented in MMM [26]; the GPS-like
approach is available in mtsslSuite [38]. The same approach can localize labelled sites in a flexible domain by measurements to reference sites in a rigid domain [39].

For many transporters and quite a few enzymes, ligand binding is coupled to relative movement of domains. In modelling such conformational transitions, the rigid-domain assumption can be somewhat relaxed and subjective identification of the domains can be avoided by using an elastic network model [40]. An implementation into MMM that accounts for the spin labels [41] has revealed conformational changes in the nucleotide-gated HCN ion channel upon binding of 3',5'-cyclic adenosine monophosphate [42].

The elastic network model approach performs well only for simple hinge motion or combination of a hinge motion with rotation of one domain [41]. For more complex conformational transitions, one may resort to comparative modelling by satisfaction of spatial restraints as implemented in MODELLER [43]. Conformational changes in the multidrug transporter EmrE with respect to the crystal structure and upon protonation were assessed using the MMM interface to MODELLER that allows for imposing distance restraints between spin-labelled sites [44]. By the same approach, additional restraints on secondary structure can be included, as shown for the external loop eL4 in proline/sodium symporter PutP [45]. A model study on T4 lysozyme demonstrated structure refinement by restraining an MD ensemble simulation by label-to-label distances [31].

If several structure models are available, the problem may boil down to discriminating between them at a sample composition not amenable to other techniques and to testing whether the model is consistent. For this approach, optimal labelling sites can be selected by analysing the distance matrix [46].

Rigid-body docking for homo-oligomers, protein complexes, or rigid domains connected by flexible linkers requires only few restraints. The two-body problem has three translational and three rotational degrees of freedom, and each additional body adds six more degrees of freedom. For homo-oligomers with known symmetry, only two translations and two rotations are free irrespective of the number of protomers. Such problems can be overdetermined by distance distribution restraints. The two-body problem can be solved by a genetic algorithm implemented in mtsslSuite [32], by a complete grid search [23,26], or by an Xplor–NIH [47]-based protocol [48]. For more than two bodies, the RigiFlex approach has been proposed [26].

In the absence of any related experimental structure, EPR restraints can be used to steer de novo modelling by the Rosetta approach [49]. Favourable labelling sites are selected by maximizing sequence separation under the constraint that predicted secondary structure elements are pairwise connected [50]. With a moderately sized set of 25 EPR distance distribution restraints, such de novo modelling provided useful results for the soluble form of Bax with 192 residues [51].

Distance distributions are particularly valuable for recognizing disorder, as illustrated in Figure 1, and for restraining the ensemble of internal or terminal intrinsically disordered domains [52]. They can be complemented by restraints on bilayer immersion depth from accessibility measurements and on secondary structure from spin labelling site scans [26]. An ensemble model for N-terminal residues 3–13 of the light-harvesting complex LHCII of green plants was generated by this approach [53].

**Integrative modelling**

The Xplor–NIH-based protocol for determining relative domain orientation and translation in proteins with two domains [48], developed in a pioneering study on a tandem of polypeptide transport-associated domains in Omp85 [54], integrates EPR distances with NMR nuclear Overhauser Effect (NOE) restraints. The structure was still ill-defined with 2907 short-range NOE restraints and could be fixed by only three additional DEER restraints. Distance distributions showed that the linker is not flexible as had been assumed before. A similar protocol supplements medium-range paramagnetic relaxation enhancement (PRE) restraints from NMR by DEER constraints [55]. Since PRE restraints are more informative on domain–domain placement than NOE restraints, the improvement is less impressive, but still the backbone root mean square deviation of an ensemble of 10 models could be improved from 4.2 to 3 Å by adding only three DEER restraints. Integration of PRE and DEER restraints had been demonstrated before [56] using the CYANA software [57].

More complex models can also be built by integrating solution-state NMR and DEER data. The structure of the 70 kDa protein/RNA complex RmsE/RmsZ was modelled with a CYANA-based protocol [58] from 4664 intramolecular NOE restraints for the protein, ~850 intramolecular NOE restraints for the RNA, and 21 distance distribution restraints between RNA stem loops [59]. Underlying assumptions and technical issues have been discussed in detail [60]. The structure consists of three rigid protein domains and a 72 nucleotide long
non-coding RNA that binds to the protein via four stem loops. Bimodal distance distributions showed that the complex adopts two major conformations.

For discoidal high-density lipoprotein particles, six DEER distance distribution restraints fixed the shape of the apolipoprotein ring enclosing the lipid bilayer sheet, which was ill-defined when using only PRE and NOE restraints [33]. Integration of EPR distance restraints with sparse pseudo-contact shift data from solution-state NMR reduces the number of free parameters in fitting the susceptibility tensor of the lanthanide probe from 8 to 5 [61].

Integration of solid-state NMR and DEER data may require diamagnetic dilution of the spin-labelled protein. This was demonstrated for the oligomeric structure of sensory rhodopsin reconstituted into lipid bilayers [62]. A combination of solid-state PRE and DEER restraints ruled out dimer formation, whereas analysis of DEER modulation depth ruled out higher oligomers than trimers. By a CNS-based [63] protocol, the trimer structure could be modelled from short-range distance restraints, torsion angle restraints, and PRE restraints. The addition of the DEER restraints led to significant refinement, especially on the cytoplasmic side.

Distance distribution restraints are a nice complement to small-angle scattering curves, which provide information on the global shape of a macromolecule on the same length scales where PDS provides a few sequence-assigned distances. The tetramer structure of histone chaperones in solution was modelled from an XRD-based dimer structure and DEER restraints, whereas SAXS curve fitting was used for testing for the correct global shape [64]. A similar approach was applied to a chromatin-remodelling enzyme [65]. Integration of SAXS and DEER data in modelling a minimal ensemble of conformations, combined with tests against single-molecule fluorescence resonance energy transfer data, was performed by unbiased replica exchange Monte Carlo simulations [66]. Structural integrity of spin-labelled mutants of the FnIII-3,4 domains of integrin α6β4 was checked by monitoring their SAXS curves in solution [67]. A combination of DEER and SAXS data to determine relative orientation and translation of the two domains led to lower uncertainty than achieved with only the 13 DEER distance distributions.

PDS-based restraints have proved valuable in validating and even refining XRD and cryo-EM structures. Even in the age of high-resolution cryo-EM, not all systems yield high-resolution images and DEER data may be useful for resolving ambiguities in data interpretation [68]. Ambiguity was also encountered in solving the structure of arrestin bound to rhodopsin by XRD on micrometre-sized crystals using a free electron laser [69]. The crystals were twinned and DEER could validate the structure. Except for small single-domain globular proteins, one may need to test whether a crystal structure is representative of the solution structure. PDS-based distance distributions can provide this information [70], which is particularly useful when different techniques result in different predictions of solution structure [71] or when a shortened construct had to be used for crystallization [46,72].

Based on DEER distance distributions and coarse models derived from them [39,73], the first all-atom model was proposed for the mitochondrial pore formed by the pro-apoptotic protein Bax [74]. Together with

---

**Figure 1. Label-to-label distance distributions reveal backbone disorder.**

(a) Distribution of the Cα–Cα backbone distance \( r_{\alpha\alpha} \) (blue) and the label-to-label distance \( r_{ee} \) (red) for residue pair 96/143 in the rigid core of plant light-harvesting complex LHCII. The width of \( P(r_{ee}) \) is well below 10 Å. (b) Distributions for residue pair 3/34 in the flexible N-terminal domain. Label-induced broadening is enhanced by the variation of side chain orientations of disordered residue 3. The width of \( P(r_{ee}) \) is well above 10 Å. The distributions were simulated in MMM [26] based on PDB structure 2BHW and the restraints reported in [53].
MD simulations of pathogenic mutants, the model could be used for assessing the influence of the mutations on pore formation.

**A fuzzy relation of structure to function**

Atomic-resolution models not only are aesthetically pleasing, but also provide a wealth of information for generating hypotheses on function. To obtain them, it has become quite usual to truncate or otherwise mutate constructs or rely on constructs augmented with an easy-crystallizing domain. Based on the Anfinsen dogma, it became customary to explain function in terms of a single conformation or of well-defined transitions between a few conformations defined at atomic resolution. While this is certainly a reasonable approximation in some cases [75–77], availability of distance distributions demonstrates that rather often conformation transitions are coupled to order–disorder transitions or are shifts in disorder equilibria [39,78–94]. Among the systems addressed by PDS to date, the fraction where at least one state is genuinely disordered is surprisingly large. Although systems not fully amenable to structure determination at atomic resolution may preferably end up in EPR laboratories, this body of data suggests that the hitherto explored part of biomolecular structure space is biased towards the most strongly ordered systems. Even in these cases, conformational uncertainty may be underestimated compared with the one in a native environment. The available distance distributions hint at a structure–function relation that transcends not only the Anfinsen dogma of a single native conformation, but also the dichotomy between Anfinsen-type ordered and intrinsically disordered proteins. In the following, I discuss a few cases where this is apparent.

It is often assumed that intrinsically disordered proteins order upon binding to other proteins or to small-molecule ligands. The multidrug transporter EmrE is ordered in its substrate-bound form, but flexible when protonated at a pH of 5 [87]. Likewise, the homodimeric multidrug ABC transporter LmrA undergoes a disorder–order transition from a very broad conformational ensemble to a rather well-defined conformation upon nucleotide binding [80]. However, in the Toc34 GTPase homodimer, the GDP-bound state is tight and ordered and the GTB-bound state disordered [85]. In helicase PcrA from superfamily 1, the two motor domains are flexible with respect to each other in the apo- and ADP-bound states, while ATP-analogue binding brings them closer together and tightens the conformational bundle [94]. DNA binding further rigidifies the motor domains. In contrast, only slight additional narrowing of the already narrow ensemble and no significant mean distance changes are observed in helicase XPD from superfamily 2. The secondary sodium/aspartate transporter GltPh samples inward- and outward-facing conformations with almost equal probability and remains conformationally heterogeneous even in the presence of substrate and sodium ions [81,82]. Synaptotagmin 1 remains heterogeneous after binding to SNAREs [78]. In cardiac myosin-binding protein C, phosphorylation causes compaction and reduces disorder of the Pro/Ala-rich linker between two immunoglobulin domains [88]. The human and teleost fish secretory components, which could be crystallized, exhibit well-defined conformations in the absence of their immunoglobulin ligands [95], whereas the unliganded avian secretory component is flexible, pointing to a distinct mechanism of ligand binding of secretory components among vertebrates [91]. Pro-apoptotic Bax exhibits disorder in the piercing domain that anchors Bax in the mitochondrial membrane, which may be related to its ability to form pores of different sizes by varying its degree of oligomerization [39]. The diversity of these examples indicates that disorder is a broadly used evolutionary strategy for optimizing protein function. Figure 2 depicts a possible strategy for tuning binding affinity by coupling an order–disorder transition to ligand binding.

![Figure 2](https://doi.org/10.1042/ETLS20170143)
Conclusion
EPR spectroscopy in conjunction with site-directed spin labelling provides structure restraints irrespective of system size and in a broad range of environments. Distance distributions from PDS match dimensions of protein domains or whole proteins and are a unique source of information on the width of conformation ensembles. Although modelling approaches that integrate EPR restraints are still in their infancy, they have already revealed several cases of conformational heterogeneity related to function. This relation is probably much more common than is suggested by the current body of structural data on biomolecules. The Anfinsen era of structural biology may be nearing its end.

Summary
- EPR-based accessibility and distance distribution restraints complement information from X-ray crystallography, cryo-EM, and NMR.
- Structures with atomic resolution can be built by integrating EPR with NMR restraints, and informative models can be built by integrating them with independently determined domain structures and small-angle scattering data.
- The broad distance distributions observed in diverse systems cast doubt on the Anfinsen dogma that a sequence usually encodes a unique native conformation.

Abbreviations
CNS, crystallography and NMR system; Cryo-EM, cryo-electron microscopy; DEER, double electron–electron resonance; EPR, electron paramagnetic resonance; GPS, global positioning system; MD, molecular dynamics; MTSSLs, methanethiosulfonate spin labels; NMR, nuclear magnetic resonance; NOE, Nuclear Overhauser Effect; PDS, pulsed dipolar spectroscopy; PRE, paramagnetic relaxation enhancement; SAXS, small-angle X-ray scattering; XRD, X-ray diffraction.

Funding
This work was supported by SNSF grant [200020_169057].

Acknowledgement
I thank Enrica Bordignon, Olivier Duss, Ines García Rubio, Christoph Gmeiner, Daniel Hilger, Benesh Joseph, Daniel Klose, Yevhen Polyhach, and Maxim Yulikov for helpful discussions during the past few years.

Competing Interests
The Author declares that there are no competing interests associated with this manuscript.

References


