

Structural bioinformatics

MESMER: minimal ensemble solutions to multiple experimental restraints

Elihu C. Ihms^{1,2,*} and Mark P. Foster^{1,2,3,*}

¹Ohio State University Biophysics Program, ²Department of Chemistry and Biochemistry, and ³Center for RNA Biology, The Ohio State University, Columbus, OH, USA

*To whom correspondence should be addressed.

Associate Editor: Burkhard Rost

Received on June 24, 2014; revised on January 27, 2015; accepted on February 3, 2015

Abstract

Motivation: Macromolecular structures and interactions are intrinsically heterogeneous, temporally adopting a range of configurations that can confound the analysis of data from bulk experiments. To obtain quantitative insights into heterogeneous systems, an ensemble-based approach can be employed, in which predicted data computed from a collection of models is compared to the observed experimental results. By simultaneously fitting orthogonal structural data (e.g. small-angle X-ray scattering, nuclear magnetic resonance residual dipolar couplings, dipolar electron-electron resonance spectra), the range and population of accessible macromolecule structures can be probed.

Results: We have developed *MESMER*, software that enables the user to identify ensembles that can recapitulate experimental data by refining thousands of component collections selected from an input pool of potential structures. The *MESMER* suite includes a powerful graphical user interface (GUI) to streamline usage of the command-line tools, calculate data from structure libraries and perform analyses of conformational and structural heterogeneity. To allow for incorporation of other data types, modular Python plugins enable users to compute and fit data from nearly any type of quantitative experimental data.

Results: Conformational heterogeneity in three macromolecular systems was analyzed with *MESMER*, demonstrating the utility of the streamlined, user-friendly software.

Availability and implementation: <https://code.google.com/p/mesmer/>

Contact: foster.281@osu.edu or ihms.2@osu.edu

Supplementary information: [Supplementary](#) data are available at *Bioinformatics* online.

1 Introduction

Changes in macromolecular structure over a range of timescales are intrinsic to many biological processes, and by definition produce structural heterogeneity within a defined timeframe (Henzler-Wildman and Kern, 2007). Such structural heterogeneity plays a role in functions as diverse as molecular recognition (Boehr *et al.*, 2009; Lange *et al.*, 2008), catalysis (Clare *et al.*, 2008; Eisenmesser *et al.*, 2005), transport (Rout and Aitchison, 2001), regulation (Tang *et al.*, 2007) and allostery (McElroy *et al.*, 2002). This heterogeneity presents unique challenges for structural characterization, as analytical tools that probe molecular structure involve measurements over defined time windows, thereby providing at the extremes

either snapshots reflecting the uniqueness of the populations, or a weighted average of the existing configurations. A further complication is that many functionally important states are short-lived or lowly populated, existing amidst a diverse background of other states and intermediates (Andersson *et al.*, 2009; Henzler-Wildman *et al.*, 2007; Krukenberg *et al.*, 2008). Therefore, in order to understand the mechanistic details of biological processes, it is often necessary to appreciate the range of structural configurations adopted by the system.

One approach to describing these configurations and their abundance is to fit experimental structural data against the predicted values obtained from *ensembles* of possible structures. Ensemble-based

techniques identify one or more pools of structures, typically generated by sampling potential conformations that are able to recapitulate experimental data (Bernado *et al.*, 2007; Bertini *et al.*, 2010; Boura *et al.*, 2011; Pelikan *et al.*, 2009). This approach has been applied to a range of structural problems using one of several publically available software packages, such as BILBOMD (Pelikan *et al.*, 2009), EOM (Bernado *et al.*, 2007) and BSS-SAXS (Yang *et al.*, 2010), all of which were developed for analyzing small-angle X-ray scattering (SAXS) or neutron scattering data. Due to the intrinsic limitations in temporal and spatial resolution of structural methods, improved accuracy in modeling can often be obtained by simultaneous fitting of data obtained from orthogonal experimental techniques. Globally fitting such data plays an important role in increasing sensitivity, removing technique-specific bias and cross validation (Beechem, 1992; Marsh and Forman-Kay, 2012). Global-fit ensemble modeling has been implemented in a number of command line-driven software packages, including EROS (Rozycki *et al.*, 2011) and SES (Berlin *et al.*, 2013), and some with user-friendly interfaces such as flexible-meccano (Ozenne *et al.*, 2012) and ENSEMBLE (Marsh and Forman-Kay, 2012), which are specifically designed for incorporation of nuclear magnetic resonance (NMR) restraints to characterize intrinsically disordered proteins.

We describe MESMER, a software package for identifying and analyzing ensembles that fit bulk-average experimental data obtained from a range of experimental techniques such as SAXS, NMR (residual dipolar couplings, RDCs; pseudocontact shifts, PCS) and dipolar electron-electron resonance (DEER) spectra. Because fitting is handled through modular plugins, MESMER can be readily extended to incorporate nearly any kind of experimental structural data, provided that a quantitative comparison can be made between the experimental and predicted data generated from components of an ensemble. The MESMER package also includes a graphical user interface (GUI) that streamlines the process of generating necessary input data, running calculations and interpreting results.

2 Methodology

MESMER facilitates the analysis of heterogeneous experimental data and interpretation by:

- Calculation of all predicted data from input structures without significant user intervention
- Simultaneous fitting of multiple data types and experimental conditions
- Selection of multiple ensemble solutions
- Visualization of structural attributes and relationships

MESMER takes as input a collection of protein data bank (PDB) files from which predicted experimental data are computed and saved as 'component' files, one per candidate structure. These precomputed data are rapidly accessed from memory as necessary during fitting. The overall workflow is illustrated in Figure 1. MESMER simultaneously fits experimental data obtained from one or more experimental conditions by collecting the datasets into 'target' files that are to be recapitulated from the provided component structures. This enables users to investigate the effects of experimental variables, such as temperature or ligand concentration, on the prevalence of ensemble components. During fitting, the predicted observable of the ensemble $\langle Y_{pred} \rangle$ for each type of data is computed from the predicted observations $\langle Y_i \rangle$ of each component present in the ensemble using the

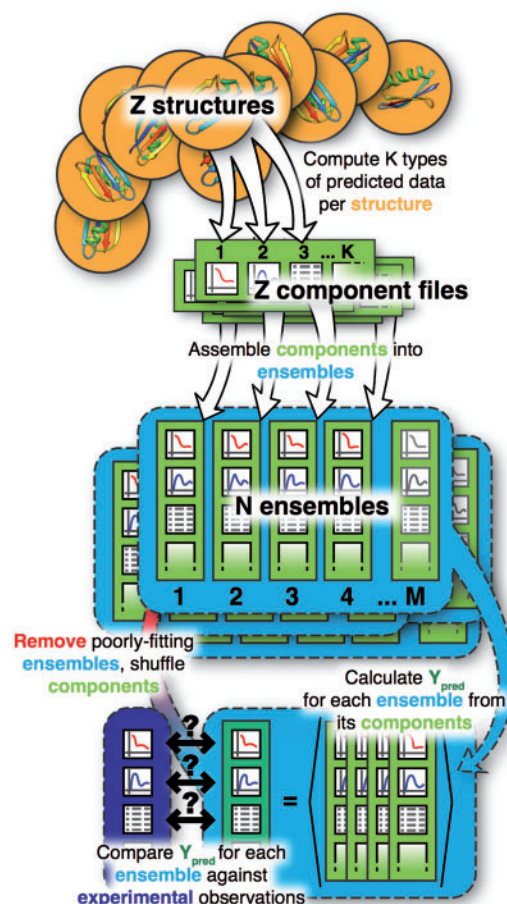


Fig. 1. Schematic overview of MESMER workflow. From each of a starting pool of Z structural candidates, the predicted counterpart data for available experiments is calculated and compiled into component files as input. A total of N ensembles, each containing M components are then assembled, initially randomly, from the pool of available components. The ability of each ensemble to recapitulate the available experimental data dictates its relative fitness for the cyclic genetic algorithm, ensuring components that contribute to accurate fits persist for future iterations (Color version of this figure is available at *Bioinformatics* online.)

coefficient c_i to describe the relative contribution (i.e. concentration) of each component (1):

$$\langle Y_{pred} \rangle = \sum_i^M C_i \langle Y_i \rangle \quad (1)$$

In practice, the number of independent degrees of freedom is increased by expanding the ensemble size M until additional components no longer result in a significant improvement to fit quality. An implicit assumption of the minimal ensemble method is that exchange between states that contribute to the observed signal occurs at a rate slower than the time resolution of the experimental measurement; otherwise, a single population-weighted average structure would generate acceptable fits to the experimental data.

If more than one target (i.e. collection of datasets corresponding to a unique set of conditions) is provided, these are fit using the same ensembles, but the relative concentrations of components in each ensemble are optimized separately. During development and testing, we found that a Newtonian conjugate-gradient type optimization (Nash, 1984) of c_i rapidly converges to acceptable solutions for small ensembles ($M < 10$), while stochastic optimizers

(Gentle *et al.*, 2004) provide rapid yet acceptable results for larger ensembles. These and several additional minimization algorithms are available in MESMER. The overall score of each ensemble, S_{total} , is obtained by summing the goodness-of-fit metrics that quantify the agreement between the predicted and experimental data acquired at T different experimental conditions for each of the K data types, for example (2):

$$S_{\text{total}} = \sum_j^T \sum_k^K w_k |\langle Y_{\text{exp}} \rangle_{jk} - \langle Y_{\text{pred}} \rangle_{jk}| \quad (2)$$

The weighting factor w_k is used to avoid overemphasis of one data type at the expense of the others, except for cross-validation purposes. Because various structural techniques have different intrinsic resolving power, weights are best assigned according to *a priori* knowledge of the precision of the measurement or rigorous cross validation. In practice, we have found it valuable to initially fit each dataset individually, thereby determining the quality of fits to the unweighted experimental data, as well inspect the resulting features of the selected ensemble components.

3 Algorithms

MESMER implements a genetic algorithm in order to iteratively refine ensembles via their predicted agreement with experimental data. Before fitting begins, MESMER generates N ensembles of M components drawn at random from the available pool of components (Fig. 1). Selecting an appropriate total number of ensembles (N) is important, as a small ensemble population may converge too quickly, while an excessively large population will demand excessive computer resources. In practice, this parameter is adjusted according to the population heterogeneity upon convergence, which is described later. Once the component concentrations of each ensemble in the ‘parent’ generation have been optimized and the resulting fitness score for each ensemble has been obtained, the pool is duplicated to form a ‘child’ generation. The child generation is diversified through two mechanisms: Crossing, in which a percentage of an ensemble’s components are exchanged with the components of another ensemble and mutation, in which a single component of an ensemble is replaced with a different component, either from components existing in the ensemble pool, or from the pool of unused components. This process is illustrated in Supplementary Figure S1. The frequency of mutation and crossing are handled stochastically, with a user-specifiable rate for each process. After crossing and mutation are complete, the component concentrations for each of the child ensembles are optimized to minimize their respective scores, and the N best fitting ensembles from the $2N$ combined parent and child pools are used as the parent population for the next generation (Supplementary Fig. S1).

3.1 Convergence

For an oversampled and structurally diverse pool of starting components, there may be hundreds or even thousands of component combinations that provide fits of comparable quality to the experimental data. Rather than allowing the genetic algorithm in MESMER to drive to a single solution that is not justified by the limited precision of the input data, a pool of comparable ensembles, each containing M components, are retained by stopping the algorithm when either the poorest-fitting ensemble has a total score S_{total} within the certainty in the experimental data (e.g. if *a priori* knowledge is available about what value constitutes an accurate fit), or when the residual standard deviation (RSD) of scores for the ensembles in the pool is below a specified value. We have found

a 1% RSD to provide ensembles with practically indistinguishable fits to the experimental data. In fact, the number of unique solutions may provide valuable information about the sampling of conformation space, as poor diversity in the final ensemble pool can be diagnostic for poor diversity in the starting structural library. In practice, oversampling of conformation space may be difficult, especially in the case of molecules with multiple flexible domains or large unstructured loops, thus, assessment of the diversity of the input structural pool should be assessed before interpreting the results of any ensemble refinement.

Although the genetic algorithm can rapidly identify solutions from a large number of possibilities, it can also lead to an underestimation of structural diversity, as the prevalence of ensembles that fit the data well, and the components that comprise those ensembles, will become greater as the algorithm progresses. This may be addressed by either increasing the total number of ensembles being improved by the algorithm, or by executing MESMER repeatedly and comparing the most prevalent components upon convergence.

3.2 Analysis

The MESMER GUI includes tools to analyze the input and resulting ensembles, by generating plots of ensemble scores, component correlations, shared attributes and fits to experimental data (Supplementary Fig. S2). Optionally, MESMER can also perform bootstrapped error analysis to estimate the confidence intervals of the best-fitting ensemble component concentrations; this is typically performed for only the best-fit ensemble, due to its computational expense (Efron and Tibshirani, 1986). Components can be selected from the original coordinate library and saved to a multi-model PDB along with attribute lists for UCSF Chimera (Pettersen *et al.*, 2004) or scripts for PyMol (<http://www.pymol.org/>) that provide options for visualizing average component concentrations and their prevalence in the ensemble pool. More advanced options for these analytical tools and others are provided by a collection of command-line utilities.

4 Results

To test and validate MESMER, we analyzed a range of experimental data types obtained from three biological systems selected for their known or suspected structural heterogeneity. For each system, MESMER analysis indicated that ensemble representation is appropriate, identifying ensembles with greatly improved agreement to the experimental data over that of any single structure or configuration.

4.1 ESCRT-I complex

Endosomal Sorting Complexes Required for Transport-I (ESCRT-I) is a multiprotein complex required for sorting and trafficking of cell receptors and other proteins (McDonald and Martin-Serrano, 2009). The binding domains of the various proteins comprising this complex are connected to the central ‘stalk’ and ‘head-piece’ of the complex core by flexible linkers ranging from 60 to 13 residues in length (Fig. 2A). Because these linkers permit structural variability that could allow for concerted opening or closing of the complex by the ubiquitin E2 variant (UEV), C-terminal (CTD) or N-terminal predicted helix domains, ESCRT-I is an attractive test system for ensemble-based analysis of macromolecules that are conformationally restricted, but structurally variable. Previously (Boura *et al.*, 2011), SAXS and double electron-electron resonance (DEER) experiments between spin-probe pairs were used to restrain describe the variable

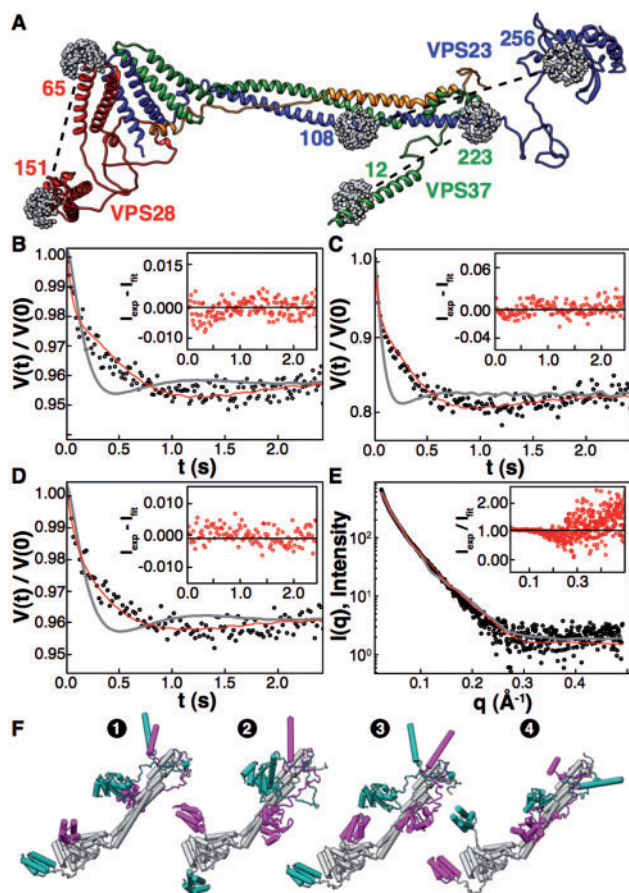


Fig. 2. Two-component ensembles can recapitulate structural data on ESCRT-I. (A) Ribbon diagram of a single ESCRT-I complex comprising three proteins: VPS23 in blue, VPS28 in red and VPS37 in green. The nitroxide rotamers (gray) used to obtain interdomain distance distributions are numbered according to the amino acid residue to which the spin labels are attached. (B–D) Experimental DEER data (black circles) and line representing the best-fitting (yet clearly inadequate) single structure (gray) and the best-fitting two-component ensemble (red line). Insets: residuals for the ensemble fits. B, DEER data from the VPS23 108–256 spin probe; C, VPS28 65–151; D, VPS37 12 to VPS23 223. (E) Corresponding SAXS data and fits. (F) Cartoon diagram of the four most structurally dissimilar two-component ensembles (out of 19) selected from a converged fit to all experimental data. The variable domains of one component in each ensemble is magenta, while the other is cyan. The invariant central core is gray (Color version of this figure is available at *Bioinformatics* online.)

structure of this complex, and in this work formed the basis for testing the utility of MESMER.

First, we generated 50 000 structural models (conformers) of the ESCRT-I complex using Xplor-NIH (Schwieters *et al.*, 2003) by treating the UEV, CTD, NTH and core domains as rigid bodies, while the linkers and unstructured C-terminal residues of the Mvb12 protein in the complex core were subjected to Monte Carlo randomization of backbone torsion angles (Supplementary Fig. S4B). Linkers were defined as previously described (Boura *et al.*, 2011): residues 159–217 of VPS23, residues 118–147 of VPS28 and residues 33–46 of VPS37. Geometric violations and steric clashes were removed through a brief energy minimization. SAXS data for the structural models were calculated with FoXS (Schneidman-Duhovny *et al.*, 2013), which in comparison with profiles calculated by CRY SOL, provided more accurate fits to the experimental data.

Because the MTSL spin labels used to collect the DEER data experience a significant amount of motion about their tethers (Galiano *et al.*, 2009), the program MMM (Polyhach *et al.*, 2011) was used to sample conformation space around the tether for each labeled domain. The resulting spin-probe labeled domains were then superimposed onto the matching unlabeled domains for each of the Monte Carlo models, resulting in structures of the complex decorated with possible configurations of the spin probe (Fig. 2A). The predicted DEER spectrum for each label-label pair was then calculated as previously described (Boura *et al.*, 2011). The modulation depth λ was optimized for DEER data and a consistent scaling factor was applied to the SAXS Y_{pred} to minimize the reduced χ^2 score for each ensemble to their respective experimental observations (3):

$$S_{\text{SAXS,DEER}} = \chi^2 = \frac{1}{Q} \sum_q \left(\frac{Y(q)_{\text{exp}} - Y(q)_{\text{pred}}}{\sigma(q)_{\text{exp}}} \right)^2 \quad (3)$$

Relative weighting between the SAXS and each of the three DEER restraints was obtained by fitting each restraint separately, and assigning a weight to the restraint that would provide a final best-fit fitness score of 1.0 upon convergence (Supplementary Table S2). When each set of DEER data was fit independently, ensembles with two components contributing roughly equal proportion were necessary in order to fit each set of experimental data. This result is consistent with the bimodal distribution of interlabel distances for each pair previously observed, possessing interlabel distance distributions centered at 25.9 and 48.4 Å for VPS23, 21.9 and 40.7 Å for VPS28, 24.7 and 42.7 Å for VPS37 (Boura *et al.*, 2011) (Supplementary Fig. S4C, Table S1). However, when the three sets of DEER data were fit simultaneously, two-component ensembles could no longer accurately recapitulate the experimental data, requiring that the ensemble size be increased to six—the same number reported by Boura *et al.* (2011). This result implied that our library was deficient in pairs of models in which all four sets of constraints could be simultaneously satisfied. This could in principle reflect the requisite diversity of the structures, or alternatively, due to inadequate sampling of the variable domain positions. Indeed, inspection of the conformational properties of the randomly generated library revealed that the majority of conformations in the library possessed interprobe distances far from those consistent with the experimental DEER data (Supplementary Fig. S4B,C).

To better sample conformations with the potential to satisfy the experimental data, a second library of 10 000 ESCRT-I structures were generated by Monte-Carlo randomization of linker torsions, but with the added condition that VPS23, VPS28 and VPS37 inter-label distances were independently restrained to those observed in the individual DEER fits (Supplementary Table S1, Fig. S4D). Using this enhanced library, a converged fit (total fitness score RSD <1%) identified 19 unique two-component ensembles, which exhibited satisfactory global fits to the combined DEER and SAXS data (Fig. 2B–E, Supplementary Table S2). A modest deviation of the ensemble fit to the SAXS data at high scattering angles may be due to incomplete modeling of solvation around unstructured regions (Schneidman-Duhovny *et al.*, 2013) or because of the partial structuring of one or more linkers; ensembles of three or more components resulted in minor improvement to SAXS fit quality (Supplementary Fig. S5), but did not improve fits to the DEER data or change conclusions about positioning of the variable domains. The structural attributes of model components present in the selected ensembles were consistent with previous findings (Boura *et al.*, 2011) that featured the UEV domain of VPS23 close to the core, despite its 59-residue long linker (Fig. 2F; Supplementary Figs S3, S4E). The configuration of the

VPS28 CTD domain displayed a slightly greater variability, but it was typically localized around the core's headpiece domain. Such variability implies a larger extent of motion for this domain, possibly reflecting equilibrium between an open and a closed state. We found a preponderance of structures with the NTH domain of VPS37 adopting two roughly collinear poses, which are typically oriented away from the VPS23 UEV. While the hypothesis of cooperative or concerted positioning of the ECRT-I domains is consistent with the finding that the experimental data can be fit with minimal ensembles containing only two, equally-contributing poses of each domain, the data cannot rule out a more complex model. Moreover, the variability in the orientations of the VPS23 and VPS28 flexible domains with respect to the complex's core indicate that the SAXS data lack the precision required to unambiguously position these domains (Fig. 2F; Supplementary Figs S3, S4E). These findings demonstrate the utility of MESMER for diagnosing and revealing structural features from highly variable systems restrained by a collection of experimental data. The iterative approach described here also demonstrates how the selective fitting of individual restraints in MESMER can be used to rationally design component libraries, enhancing sampling of reasonable conformations.

4.2 Calmodulin

The calmodulin calcium-binding family of signal messenger proteins consists of two rigid domains separated by a flexible linker of seven amino acids, which in some crystallographic studies has been shown to form a long, uninterrupted alpha helix (Chattopadhyaya *et al.*, 1992; Grabarek 2005). NMR experiments have indicated that this linker is largely unstructured in solution (Zhang *et al.*, 1995), permitting significant reorientation of the two domains with respect to one another (Anthis *et al.*, 2011; Bertini *et al.*, 2004) (Fig. 3A). This dynamic behavior, and the potential for populating one or more favored conformational states makes Calmodulin an attractive candidate for ensemble-based analysis of conformational heterogeneity. The N60D mutant of calmodulin has been shown to preferentially bind lanthanide metal ions in the N-terminal domain, allowing the collection of paramagnetic pseudocontact shifts (PCS) and RDCs by NMR (Bertini *et al.*, 2004). These observations provide a bulk-average measure of the relative orientation between the two domains. In addition to paramagnetic restraints, SAXS data for the protein is also available (Bertini *et al.*, 2010).

The program RanCh (Bernado *et al.*, 2007), which treats two domains as rigid bodies and randomizes the connecting linker backbone Phi and Psi angles, was used to generate a library of 50 000 model structures from the atomic coordinates of the N- and C-terminal domains in the Maximum Occurrence example dataset (Bertini *et al.*, 2010); we used RanCh in this case instead of XPLOR because of lower computational cost and enhanced sampling. As CRYSOLO resulted in superior fits to the SAXS data than FoXs for calmodulin, it was used to calculate the predicted SAXS profiles for each structure, with 128 points per simulated profile to a maximum scattering angle of $0.35 \cdot 4\pi \sin(\theta) \text{ (1/\AA)}$. Solvent electron density was set to 0.334 e/\AA^3 , and a hydration shell contrast of 0.03 e/\AA^3 was used. The magnetic susceptibility tensor orientation and anisotropies for terbium and thulium were obtained by fitting the N-terminal domain PCS values using the program Numbat (Schmitz *et al.*, 2008) and matched previously published values (Bertini *et al.*, 2004). These tensors were subsequently used to calculate predicted PCS and RDCs for the model structures using a MESMER plugin based on the PyParaTools python library (M. Stanton-Cook, X.-C. Su, G. Otting, T. Huber, <http://comp-bio.anu.edu.au/mscook/PPT/>).

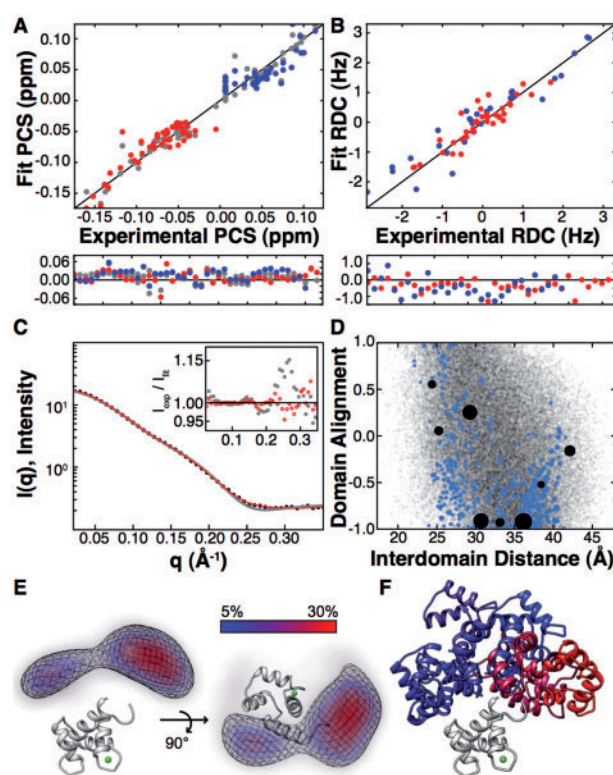


Fig. 3. Analysis of calmodulin structural variability. (A, B) Correlation between experimental and fitted pseudocontact shifts (PCS) and RDCs for C-terminal domain atoms in the global fit of SAXS, PCS and RDC values from an eight-component ensemble. Red dots are for Tb lanthanide, blue for Tm. Gray dots are best-fit values for a six-component fit to PCS data only. Residuals are in the lower panel. (C) Fit to experimental SAXS data (black circles) of the globally best-fitting ensemble (red line), or best-fitting single component (gray). Residuals for both fits are inset. (D) Structural attribute plot (interdomain distance versus alignment of primary inertial tensors) of all 50 000 components in the original pool (small grey dots), and those present in converged ensembles to fits of SAXS and PCS data (blue); the point size is proportional to the component's contribution to its containing ensemble. Black circles highlight the eight components structures from the best fitting ensemble in panel F. (E) Density map of the C-terminal domain of structures selected in a converged fit to SAXS and PCS data. The N-terminal domain is held fixed (in gray, paramagnetic center in green). Density map color is ramped blue-red in proportion to the concentration of the component in the ensemble. (F) Superimposition of the eight components of the best-fitting ensemble to SAXS, PCS and RDC data. The C-terminal domain cartoon is colored according to its contribution to the fitted ensemble (Color version of this figure is available at *Bioinformatics* online.)

The agreement score between the predicted PCS and RDC values and the experimental results was obtained from the quality score Q , as previously defined (Cornilescu *et al.*, 1998) (4):

$$S_{RDC,PCS} = Q = \frac{rms(\Delta\delta^{meas} - \Delta\delta^{pred})}{rms(\Delta\delta^{meas})} \quad (4)$$

Relative weighting of the SAXS, PCS and RDC restraints was determined in the same manner as described for the ESCRT-I system restraint types (Supplementary Table S3). As RDC values are sensitive to small structural perturbations (Zweckstetter and Bax, 2002), initial fits were made with only the SAXS and PCS data (Fig. 3A–C). An ensemble size of six components was found to provide accurate fits, with no significant improvement in fit quality obtained using larger ensembles. Components present in ensembles obtained from an RSD <1% pool of solutions exhibited C-terminal domains adopting a wide variety of poses with respect to the N-terminal

domain, consistent with previous reports (Fig. 3D and E), while certain configurations with otherwise sterically permitted geometries (gray dots in Fig. 3D) were not observed in the final pool of selected ensembles (blue dots in Fig. 3D), although they were present in the component pool and starting ensembles. Because of the single, significantly shorter linker, and the fairly uniform coverage of interdomain distances and alignments (Fig. 3D) undersampling was judged to be significantly less problematic than with the larger ESCRT-I complex.

However, when RDC data were included as a restraint, the ensemble size M necessary to generate reasonable fits increased to at least eight components (Supplementary Fig. S9), and significantly narrowed the diversity of structures that populate converged ensembles (Fig. 3D, red dots), typically resulting in a single unique ensemble of eight structures (Fig. 3F; Supplementary Fig. S10). As multiple independent MESMER runs selected different ensembles upon convergence (yet all with similar scores), this was judged to be a side effect of the number of ensembles used during the fitting, which is limited by computer memory, and not inadequate sampling of structural variability in the starting structure library. This is consistent with the increased resolution of RDC restraints compared to PCS and SAXS data, and is supported by cross-validation plots of PCS and RDC restraints indicating that structures obtained by fitting RDC data alone can be used to accurately recapitulate PCS data, but not vice versa (Supplementary Fig. S11A and B). Overall, these results are consistent with the present understanding of the structural plasticity of calmodulin, including the presence of favored configurations and the absence of disfavored ones (Anthis *et al.*, 2011; Barbato *et al.*, 1992; Bertini *et al.*, 2004). Such observations confirm the utility of MESMER in the ensemble analysis of protein structural variability.

4.3 TRAP and AT hetero-oligomerization

The homo-oligomeric TRAP and Anti-TRAP (AT) proteins participate in a polydentate condensation process that leads to the formation of heterogeneous chains or clusters of the components (Ihms *et al.*, 2014). Thermodynamic and structural analysis of the chaining process indicates that multiple AT trimers (AT_3) bind to sites on the surface of the ring-shaped TRAP undecamer ($TRAP_{11}$). Thus, at limiting AT_3 ratios, clusters are formed as increasing numbers of $TRAP_{11}$ rings bind to AT_3 . However, as the $AT_3:TRAP_{11}$ ratio increases, additional AT_3 s bind to $TRAP_{11}$, resulting in smaller, AT-saturated clusters. This condition-dependent modulation of hetero-complex abundance offers a unique opportunity to examine the utility of MESMER to detect the configuration and relative abundance of species present in a heterogeneous solution.

As previously described (Ihms *et al.*, 2014), bead models of $AT_3 + TRAP_{11}$ clusters were assembled from models of the components obtained from *ab initio* shape reconstruction. The SAXS profiles for the models were calculated by CRYSOLOG with 256 points per simulated profile to a maximum scattering angle of $0.32 \cdot 4\pi \sin(\theta) \cdot (1/\text{\AA})$. Solvent electron density was set to 0.334 e/\AA^3 , and a hydration shell contrast of 0.03 e/\AA^3 was used. Nine separate SAXS curves from samples at three $AT_3:TRAP_{11}$ ratios (0.5, 1.0 and 1.5) and three TRAP concentrations (3, 5 and 10 mg/ml) were obtained at pH 8 (Ihms *et al.*, 2014). Under basic conditions, complexes exist in rapid equilibrium with each other and the free components, resulting in an average complex size smaller than those present at neutral pH (E. Ihms, unpublished results). The SAXS curves for each condition were simultaneously constrained with an additional, equally weighted stoichiometry restraint S_{ratio} to ensure agreement between the $AT_3:TRAP_{11}$ ratios in the ensembles and in the experimental sample (5):

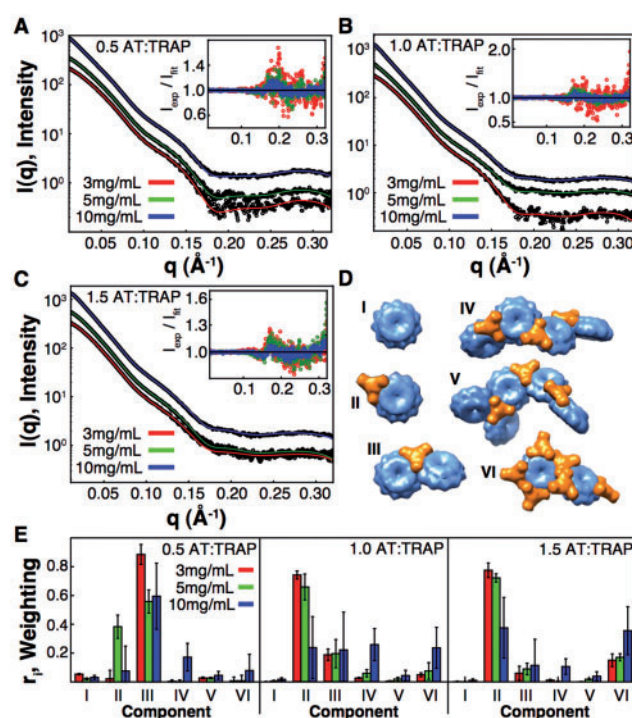


Fig. 4. Simultaneous analysis of TRAP + Anti-TRAP titration data. (A) Fits from best fitting six-component ensemble (red, green and blue lines) to experimental SAXS data (black circles) obtained from samples mixed at a ratio of 0.5 $AT_3:TRAP_{11}$ and three TRAP concentrations. Residuals from fits are inset. (B) The same six-component ensemble fit to experimental datasets obtained from 1.0 $AT_3:TRAP_{11}$ mixtures. (C) Fits to experimental datasets obtained from 1.5 $AT_3:TRAP_{11}$ mixtures. (D) The best-fitting ensemble of six TRAP and AT + TRAP components determined by MESMER. (E) A bar graph of the best-fitting component contribution to its ensemble as a function of TRAP concentration for the three AT:TRAP ratios used. Error bars are from MESMER's bootstrapped error analysis (Color version of this figure is available at Bioinformatics online.)

$$S_{ratio} = \sum_i \left| \ln \left(\frac{R_i^{exp}}{R_i^{fit}} \right) \right| \quad (5)$$

where the ratio of each component (R_i) is calculated from either the concentration of each protein present in the sample (C_i), or the number of TRAP or Anti-TRAPs present in the ensemble (n_i) (6):

$$R_i = \frac{[C_i]}{\sum_j [C_j]} = \frac{n_i}{\sum_j n_j} \quad (6)$$

Ensembles with four components were sufficient to fit each of the SAXS datasets with an average chi-square < 1.5 (Fig. 4A–C, Supplementary Fig. S12; Table S4) while increasing the ensemble size beyond six components provided little additional reduction in chi-square (Supplementary Fig. S12). A preponderance of extended configurations, with multiple $TRAP_{11}$ rings stitched together by AT_3 (Fig. 4D) were required to fit the data. The only component present in all of the selected ensembles was free $TRAP_{11}$, although its contribution to the ensemble did not exceed 5% at the lowest concentration of TRAP and AT (Fig. 4E). Examination of the ensemble component contributions as a function of the $AT_3:TRAP_{11}$ ratio reveals several interesting trends. The free $TRAP_{11}$ component did not significantly contribute in ensemble fits to either the 1 or 1.5 $AT_3:TRAP_{11}$ data, and the contribution of the $AT_3-2TRAP_{11}$

component also decreases dramatically in fits to data obtained at higher ratios. The 1AT₃-1TRAP₁₁ component only significantly contributes in fits to data with ≥ 1 AT₃:TRAP₁₁ ratios, suggesting that multiple TRAPs cooperatively bind AT, but not necessarily vice versa (Fig. 4E).

The relatively small concentration range tested has little apparent effect on the relative contributions of the components, consistent with the fact that the affinity of TRAP₁₁ for AT₃ is well above the concentrations tested. However, the contribution of the most AT-saturated complex is directly proportional to both AT₃:TRAP₁₁ and increasing component concentration, while the contribution of the less saturated AT₃-TRAP₁₁ complexes concomitantly decrease (Fig. 4E).

These results demonstrate the utility of MESMER for describing the structure and population of species present in heterogeneous solutions, and how these properties vary under experimental conditions. This utility may find value in investigating how ligands or other variables modulate conformational or oligomerization equilibria.

5 Discussion

Obtaining meaningful results from ensemble-based approaches is contingent upon various factors. These include the discriminating power of the input data, the proper sampling of conformational space and the selection of an appropriate ensemble size. If the ensemble size is too small, insufficient structural variability will typically result in poor fits, while overly large ensembles run the risk of over-fitting the experimental data and potentially including unreasonable components. Indeed, the ensemble size embodies the difference between ‘minimal ensemble’ and other ensemble-based techniques. Some methods that employ larger ensembles and thus greater computational requirements, such as EROS (Rozycki *et al.*, 2011) typically avoid over-fitting by using statistical methods, such as maximum entropy which restricts excessive optimization of ensemble component concentrations/contributions. However, even these approaches are subject to significant pitfalls, most notably that insufficient or inaccurate sampling of conformational space may result in unnecessarily large ensembles. In the other extreme, it is possible that an artificially small set of component structures may provide acceptable fits to experimental data while insufficiently describing the system’s true heterogeneity. Both concerns may be partially addressed by cross validation with different types of experimental data, and examination of the variability present in multiple converged solutions.

MESMER facilitates the process of finding structural ensembles that can recapitulate data obtained from heterogeneous solutions. We have validated MESMER by investigating a variety of case studies previously shown to exhibit structural heterogeneity. Despite this heterogeneity, the averaged experimental data available from these systems were sufficient to report on specific structural characteristics, providing new insights into the systems’ function. MESMER introduces several innovations that may find utility in similar ensemble-based methods, streamlining much of the necessary workflow while retaining flexibility. MESMER’s GUI is intended to enable a wide user base to readily and intuitively investigate complicated structural heterogeneity, while its extensible and modular nature enables advanced users to incorporate new experiments and analytical techniques.

Acknowledgements

We thank Dr Bartosz Rozycki and Dr Gerhard Hummer for the ESCRT-I data and Dr Michel Hammel for trial data. We also thank Kevin Dyer and Greg Hura of the Lawrence Berkeley National Lab and the staff at the

Advanced Light Source for assistance in collecting the SAXS data for the TRAP system, and Paul Gollnick for providing the TRAP samples and for helpful discussions.

Funding

The Advanced Light Source is supported by the Director, Office of Science, U.S. Department of Energy under contract DE-AC02-05CH11231. This work was supported by National Institutes of Health Grant GM077234 to M.P.F. and in part by an allocation of computing time from the Ohio Supercomputer Center.

Conflict of Interest: none declared.

References

- Andersson, M. *et al.* (2009) Structural dynamics of light-driven proton pumps. *Structure*, **17**, 1265–1275.
- Anthis, N.J. *et al.* (2011) Transient, sparsely populated compact states of apo and calcium-loaded calmodulin probed by paramagnetic relaxation enhancement: interplay of conformational selection and induced fit. *J. Am. Chem. Soc.*, **133**, 18966–18974.
- Barbato, G. *et al.* (1992) Backbone Dynamics of calmodulin studied by N-15 relaxation using inverse detected 2-dimensional Nmr-spectroscopy - the central helix is flexible. *Biochemistry*, **31**, 5269–5278.
- Beechem, J.M. (1992) Global analysis of biochemical and biophysical data. *Methods Enzymol.*, **210**, 37–54.
- Berlin, K. *et al.* (2013) Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J. Am. Chem. Soc.*, **135**, 16595–16609.
- Bernado, P. *et al.* (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, **129**, 5656–5664.
- Bertini, I. *et al.* (2004) Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc. Natl Acad. Sci. USA*, **101**, 6841–6846.
- Bertini, I. *et al.* (2010) Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.*, **132**, 13553–13558.
- Boehr, D.D. *et al.* (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, **5**, 789–796.
- Boura, E. *et al.* (2011) Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc. Natl Acad. Sci. USA*, **108**, 9437–9442.
- Chattopadhyaya, R. *et al.* (1992) Calmodulin structure refined at 1.7 Angstrom resolution. *J. Mol. Biol.*, **228**, 1177–1192.
- Clare, D.K. *et al.* (2008) Multiple states of a nucleotide-bound group 2 chaperonin. *Structure*, **16**, 528–534.
- Cornilescu, G. *et al.* (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.*, **120**, 6836–6837.
- Efron, B. and Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.*, **54**–75.
- Eisenmesser, E.Z. *et al.* (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, **438**, 117–121.
- Galiano, L. *et al.* (2009) Solute effects on spin labels at an aqueous-exposed site in the flap region of HIV-1 protease. *J. Phys. Chem. B*, **113**, 1673–1680.
- Gentle, J.E. *et al.* (2004) *Handbook of Computational Statistics: Concepts and Methods*. Springer, Berlin.
- Grabarek, Z. (2005) Structure of a trapped intermediate of calmodulin: Calcium regulation of EF-hand proteins from a new perspective. *J. Mol. Biol.*, **346**, 1351–1366.
- Henzler-Wildman, K. and Kern, D. (2007) Dynamic personalities of proteins. *Nature*, **450**, 964–972.
- Henzler-Wildman, K.A. *et al.* (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature*, **450**, 838–844.
- Ihms, E.C. *et al.* (2014) Gene regulation by sub-stoichiometric heterocomplex formation of undecameric TRAP and trimeric anti-TRAP. *Proc. Natl Acad. Sci. USA*, **111**, 3442–3447.
- Krukenberg, K.A. *et al.* (2008) Multiple conformations of E-coli Hsp90 in solution: Insights into the conformational dynamics of Hsp90. *Structure*, **16**, 755–765.

- Lange, O.F. *et al.* (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, **320**, 1471–1475.
- Marsh, J.A. and Forman-Kay, J.D. (2012) Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins*, **80**, 556–572.
- McDonald, B. and Martin-Serrano, J. (2009) No strings attached: the ESCRT machinery in viral budding and cytokinesis. *J. Cell Sci.*, **122**, 2167–2177.
- McElroy, C. *et al.* (2002) TROSY-NMR studies of the 91 kDa TRAP protein reveal allosteric control of a gene regulatory protein by ligand-altered flexibility. *J. Mol. Biol.*, **323**, 463–473.
- Nash, S.G. (1984) Newton-type minimization via the Lanczos method. *SIAM J. Num. Anal.*, **21**, 770–788.
- Ozenne, V. *et al.* (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, **28**, 1463–1470.
- Pelikan, M. *et al.* (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.*, **28**, 174–189.
- Pettersen, E.F. *et al.* (2004) UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
- Polyhach, Y. *et al.* (2011) Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.*, **13**, 2356–2366.
- Rout, M.P. and Aitchison, J.D. (2001) The nuclear pore complex as a transport machine. *J. Biol. Chem.*, **276**, 16593–16596.
- Rozycki, B. *et al.* (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure*, **19**, 109–116.
- Schmitz, C. *et al.* (2008) Numbat: an interactive software tool for fitting Delta chi-tensors to molecular coordinates using pseudocontact shifts. *J. Biomol. NMR*, **41**, 179–189.
- Schneidman-Duhovny, D. *et al.* (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.*, **105**, 962–974.
- Schwieters, C.D. *et al.* (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
- Tang, C. *et al.* (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*, **449**, 1078–1082.
- Yang, S. *et al.* (2010) Multidomain assembled states of Hck tyrosine kinase in solution. *Proc. Natl Acad. Sci. USA*, **107**, 15757–15762.
- Zhang, M. *et al.* (1995) Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.*, **2**, 758–767.
- Zweckstetter, M. and Bax, A. (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J. Biomol. NMR*, **23**, 127–137.