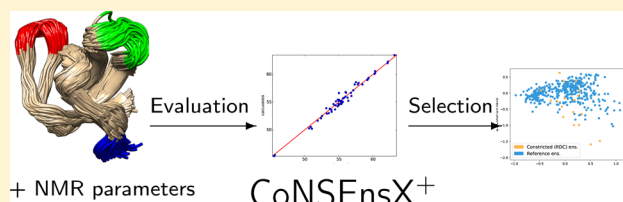


CoNSEnsX⁺ Webserver for the Analysis of Protein Structural Ensembles Reflecting Experimentally Determined Internal DynamicsDániel Dudola,¹ Bertalan Kovács, and Zoltán Gáspári^{1*}

Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, 1083 Budapest, Práter u. 50/A, Hungary

Supporting Information

ABSTRACT: Ensemble-based models of protein structure and dynamics reflecting experimental parameters are increasingly used to obtain deeper understanding of the role of dynamics in protein function. Such ensembles differ substantially from those routinely deposited in the PDB and, consequently, require specialized validation and analysis methodology. Here we describe our completely rewritten online validation tool, CoNSEnsX⁺, that offers a standardized way to assess the correspondence of such ensembles to experimental NMR parameters. The server provides a novel selection feature allowing a user-selectable set and weights of different parameters to be considered. This also offers an approximation of potential overfitting, namely, whether the number of conformers necessary to reflect experimental parameters can be reduced in the ensemble provided. The CoNSEnsX⁺ webserver is available at consensx.itk.ppke.hu. The corresponding Python source code is freely available on GitHub (github.com/PPKE-Bioinf/consensx.itk.ppke.hu).



INTRODUCTION

Besides their structure, the internal dynamics of proteins is regarded as a key determinant of their biological function. This dynamics ranges from the high conformational diversity of intrinsically disordered proteins and segments^{1,2} to well-defined structural rearrangements in globular proteins.^{2,3} While NMR spectroscopy provides unique possibilities to obtain information about conformational heterogeneity and internal motions at a diverse range of time scales, detailed structural interpretation of these parameters requires ensemble representations of protein structures that correspond to the experimental parameters.^{4,5} Such ensembles, referred to as “dynamic structural ensembles” below, differ substantially from those routinely deposited in the PDB. The reason for this is that most conventional structure calculation methods enforce the correspondence of the experimental parameters to each individual conformer in the ensemble and the diversity of the conformers stems from differences of the solutions to the same structure calculation problem. In contrast, ensemble-based methods inherently treat NMR-derived parameters as an ensemble property, and require the ensemble as a whole to correspond to these parameters. This treatment is expected to reflect reality better as with the exception of slow motions on the NMR time scale, the measured parameters represent an ensemble average and, in general, no individual conformation can be expected to fulfill all of them simultaneously.⁶ Thus, these ensembles are usually substantially more diverse than typical “NMR ensembles” and require ensemble-based validation in terms of correspondence to experiments.⁷

Structural ensembles reflecting experimental parameters can be obtained either by ensemble-restrained molecular dynamics simulations^{8,9} or by selecting a subensemble from a suitable pool of conformers.^{10,11} Examples of parameters used for ensemble

restraining include, among others, chemical shifts,³ S² order parameters,^{12,13} and residual dipolar couplings.^{14,15}

Whereas in such an ensemble each conformer should be checked for correct geometry just like for any protein structural model, validation in relation to experimental data is only meaningful at the level of the ensemble⁷ and not on a per-model basis. As usual, an important aspect is cross-validation with data sets not used for the generation of the ensemble, requiring a general tool capable of handling a number of parameter types. Another important issue for dynamic structural ensembles is overfitting. Overfitting (or under-restraining) is the phenomenon where the number of replicas is larger than minimally required to conform to all measured parameters.⁸ Larger ensembles might satisfy the experimental restrains more easily while incorporating conformations that are not realistically populated by the motions of the protein at the time scale modeled. Several aspects of overfitting can be handled during ensemble generation, e.g. the MUMO (minimal under-restraining, minimal over-restraining) calculation scheme was optimized with this respect.⁸ Specifically, the overfitting stemming from the nonlinear averaging of nuclear Overhauser effect (NOE) based distances within the ensemble is suggested to be minimized by avoiding averaging the distances between more than two replicas.^{8,16}

To our knowledge, currently available software tools for ensemble generation and analysis are related to various aspects of selection and are focusing on ensembles with high structural variability such as intrinsically disordered proteins or multi-domain proteins with flexible linkers. The downloadable packages ENSEMBLE¹¹ and ASTEROIDS¹⁰ provide robust

Received: February 3, 2017

Published: July 13, 2017

selection-based solutions primarily for intrinsically disordered proteins. The MaxOcc service¹⁷ determines the maximum possible weight of a structural model within an ensemble that can be compatible with the observed parameters.

There are a number of tools available for the validation of NMR structures, some of the capable of ensemble averaging, such as RPF¹⁸ or PROSESS.¹⁹ These servers are not specifically designed to handle dynamic structural ensembles, handle NMR chemical shift and NOE distance data only and do not offer a selection option.

Our aim was to develop a freely accessible web service that is capable of performing general ensemble-based validation step in a standardized and user-friendly way and also offering a simple and straightforward, multipurpose selection option. Therefore, updating our previous service CoNSEnsX,²⁰ we introduce CoNSEnsX⁺ with enhanced data handling and visualization features and, most importantly, a subensemble selection option that can both be used to generate dynamic ensembles and to check the presence of possible overfitting.

RESULTS AND DISCUSSION

Features of the CoNSEnsX⁺ Server. The CoNSEnsX⁺ server is a complete reimplement of our previous CoNSEnsX service and is available at consensx.itk.ppke.hu. The server takes three files as input:

- a multimodel PDB file containing the ensemble
- an NMR-Star file with NOE-based distance restraints (accessible as “V2 NMR Restraints” at rcsb.org)
- a BMRB-style NMR-Star format file with NMR parameters (chemical shifts, scalar and residual dipolar couplings, order parameters; see in detail below)

While submitting a PDB-format structural ensemble is mandatory, it is sufficient to provide only one of the BMRB/distance restraint files. The server is capable of evaluating backbone and side-chain S^2 order parameters, amide N, H, $H\alpha$, C α , and C β chemical shifts, $^3J_{H_N H_\alpha}$, $^3J_{H_N C_\alpha}$, $^3J_{H_N C_\beta}$ and $^3J_{H_N C}$ scalar couplings as well as residual dipolar couplings. The server reports RMSD and Pearson correlation between all measured and ensemble-averaged back-calculated parameters, as well as the Q-factor for each RDC set. A plot is also presented showing the relationship between the correlation values obtained for the ensemble average and those calculated for the individual conformers.

For the back-calculation of RDCs, the server offers two options:

- By default, PALES is invoked for each individual conformer with the SVD flag, resulting the alignment that provides the best correspondence to experimental RDCs. While this approach probably artificially improves the ensemble-level correspondence to RDCs, it accounts for the fact that individual conformers might align differently^{21–23} in the medium and does not suffer from the limitations inherent in approximating the alignment based on the structures alone.
- Optionally, the steric alignment of each conformer can be estimated using PALES. This approach seems less biased than the previous one and might also account for the diversity of alignments between conformers but alignment estimations yield worse results than SVD for individual molecules and can not be expected to perform equally well in all of the highly diverse medium types available for RDC measurements.

The server does not provide a single “measure of goodness” because the availability and quality of different kinds of

NMR-derived parameters may greatly vary from case to case and thus no single measure is expected to be directly comparable between different proteins and would be both misleading and of limited practical use.

Key novel features of the CoNSEnsX⁺ server relative to the original version are

- an option for structure superposition using a specified range of residues
- recognition and analysis of side-chain methyl S^2 order parameters
- three Karplus equation parameter sets^{24,24,25} can be selected for the calculation of 3J -couplings from the ϕ torsion
- recognition of RDC data obtained in different experiments and invoking PALES separately for each RDC set measured under different conditions
- improved graphical representation of the parameter comparisons

The most important novel feature of the server is the incorporation of a greedy selection approach for selecting a subensemble best corresponding to a user-defined set of parameters. The selection can be invoked after a round of initial evaluation and

- can optimize the RMSD or Pearson correlation between the experimental and back-calculated parameters (if only RDCs are included in the target function, the Q-value can also be chosen)
- uses parameter sets selected by the user to be included in the target function with a weight ranging from 0 to 10
- is capable of overcoming the addition of conformers which cause temporary deterioration of the target value if later additions cause further improvements

Evaluation of the Ensemble Selection Feature. It should be noted here that in a dynamic structural ensemble it is not necessarily expected that any single member exactly corresponds to a conformer that is indeed occurring with a non-negligible probability in reality, but the conformational space represented by the ensemble as a whole should match that of the modeled real protein as closely as possible. Therefore, here we will focus on comparing the conformational space sampled by different ensembles to estimate the presence and extent of overfitting.

To evaluate and demonstrate the ensemble selection feature, we have selected three example cases including an SH3 domain, the peptidyl-prolyl cis–trans isomerase Pin1, and the disordered α -synuclein.

SH3 Domain with Synthetic Parameters. We have performed molecular dynamics calculations on the N-terminal SH3 domain of the Drk protein (PDB ID 2A36).²⁶ A 100 ns simulation was run with a time step of 1 fs at 300 K in implicit solvent using the GBSA algorithm.²⁷ The first model in PDB 2A36 was used as an input structure. The final 20 ns of the obtained trajectory was used as a reference ensemble for which various types of parameters (chemical shifts, five independent RDC sets and backbone S^2 order parameters) were calculated. Using the chemical shifts and RDCs, a constricted ensemble was generated from the reference ensemble to test whether the observed parameters can be represented by a smaller set of conformers. As input for an independent selection, an initial pool covering a larger conformational space was generated with the accelerated molecular dynamics (AMD) scheme²⁸ as implemented by our group previously in GROMACS.²⁹ Four parallel 20 ns AMD simulations were carried out from the same starting

structure but with different initial atomic velocities with an AMD boost potential of $E = 2400$ kJ/mol and an acceleration factor $\alpha = 0.5$. The resulting set of conformers, denoted AMD pool below, is composed of the last 18.75 ns of the four AMD trajectories (3044 conformers). From this pool a number of subensembles were selected using various combinations of the “experimental” parameters calculated for the reference ensemble (Table S1). The conformational space covered by the different ensembles was approximated using principal component analysis.

In the reference ensemble, the first motional mode corresponds to a bending motion of the loop connecting the $\beta 2$ – $\beta 3$ strands, denoted as the N-src loop in the literature, which is one of the loops flanking the ligand binding pocket.³⁰ The second mode describes the heterogeneity of the C-terminal 3 residues of the protein, which is not considered functionally relevant (Figure 1A–C, the position of a bound peptide is shown using the structure 4LNP³¹).

All selected ensembles reflect the first motional mode of the reference ensemble while achieving reasonable correspondence to all parameters considered (Table S1 and Figure S1). In the case of the constricted ensemble (Figure 1C), this reveals that the most important features of the reference ensemble can be well represented by a smaller set of conformers, providing a model

case for assessing the extent of overfitting. Selections using the AMD pool reveal that similar ensembles can be selected from a pool that covers the conformational space of the reference ensemble (Figure 1E–H). Our deterministic greedy approach is not expected to be sensitive to the exact distribution of the conformers in the initial pool as long as it sufficiently covers the reference ensemble, as the probability of selecting a particular conformer to extend the ensemble depends only on its ability to optimize the correspondence to the parameters and not on the density of conformers in specific regions of the conformational space. Importantly, this kind of selection does not guarantee the selection of the best subensemble in terms of correspondence to the parameters and a stochastic selection might yield better results. Exact optimum could only be achieved by enumeration of all conformer combinations. Nevertheless, our approach is capable of selecting a subensemble corresponding reasonably well to the parameters and can be used to assess the extent of potential overfitting.

Pin1 Peptidyl-Prolyl Cis–Trans Isomerase. The Pin1 peptidyl-prolyl cis–trans isomerase was chosen as an example of a protein where a motional mode affecting the subdomains surrounding the substrate binding site has been suggested to play a role in catalysis and its regulation.^{32,33} We have generated three

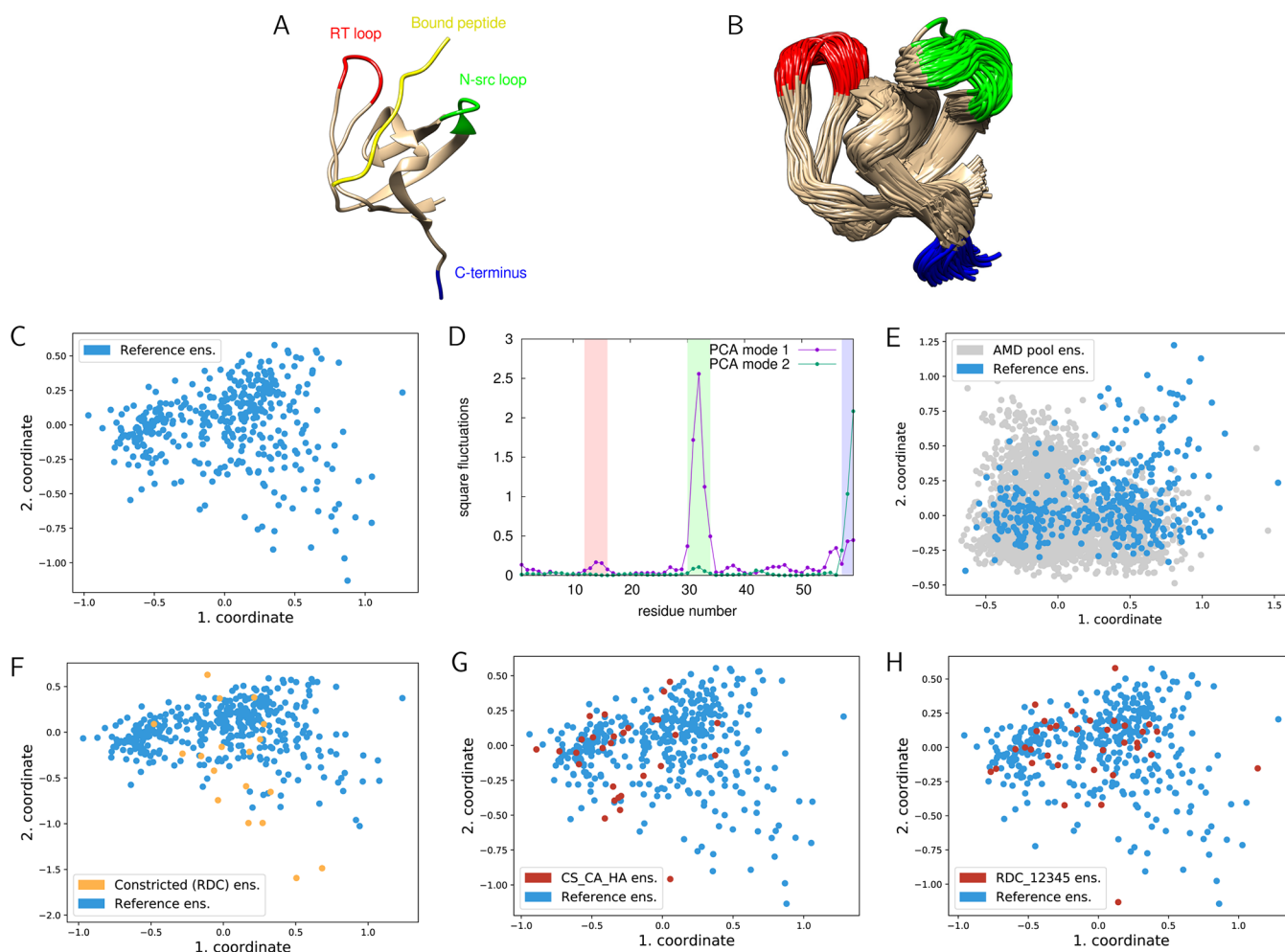


Figure 1. (A) Cartoon representation of a ligand-bound SH3 structure (PDB ID 4LNP); (B) superimposed reference ensemble; and (C) its PCA plot with (D) squared fluctuations of residues in the first two PCA modes. (E) PCA plot of the reference ensemble and the AMD pool. (F) PCA plot of the reference and the constricted ensemble. (G and H) PCA plot of the reference ensemble and the CS and RDCS ensembles, respectively (see text for details).

initial ensembles with different AMD parametrization from the starting structure 2RUC.³⁴ We have then evaluated these ensembles using the experimental chemical shifts deposited for this structure (BMRB ID 11559) and performed selections based on the chemical shifts. Correspondence to chemical shifts was generally high in all ensembles investigated and selected. However, all ensembles generated with AMD performed better than the original 2RUC ensemble available from the PDB, compatible with the general notion that a diverse ensemble fulfills the parameters more easily. This correspondence is still slightly improved for the selected subensembles which typically have between 5 and 15 members depending on the types of chemical shifts and measure (correlation or RMSD) used in the selection (see Table S2 for data on some highlighted ensembles). Although the differences are subtle, it is apparent that the subensembles selected from the AMD4800 pool exhibit the best correspondence to the experimental chemical shifts (Table S3). Principal component analysis reveals that the AMD4800 ensemble, generated with the longest simulation time, is the most diverse, and the subensembles selected from it contain conformers from the conformational space covered only by this ensemble (Figure S2A and B). To assess the significance of this observation, we checked whether the selected ensembles reflect a biologically relevant motion. To this end, we used the motional modes identified in our previous work using the consensus of three different dynamic structural ensembles and other parvulin-type PPIase (Rotamase) domains.³³ We focused on a subset of 53 CA atoms identified previously as common to different PPIase domains based on a structure alignment. Results of a principal component analysis of these atomic positions were compared to our previously published PCA results.³³ It is clear the first motional mode of the ensembles selected from the AMD4800 pool shows a moderate but clearly detectable overlap with the motional mode believed to be functionally most relevant (Figure S2C and D), corresponding to the opening–closing motion around the ligand-binding site. In contrast, the full AMD5000 and AMD5200 pools and the subensembles selected from them show no such overlap. This analysis, besides highlighting the well-known importance of sufficient conformational sampling, reveals that even subtle improvements in the correspondence to experimental parameters can be relevant. Thus, at least in special cases, even this admittedly simple selection scheme is capable of producing a small ensemble with good overall correspondence to chemical shifts and—given a suitable initial pool of conformers—reflecting a biologically relevant dynamical feature.

α-Synuclein. An *α*-synuclein ensemble³⁶ available in pE-DB⁵ (ID 9AAC) has been chosen to test the server on an intrinsically disordered protein. This ensemble was calculated based on paramagnetic relaxation enhancement (PRE) data,³⁷ the analysis of which is currently not implemented in our server. Therefore, we performed the calculations and selection using chemical shift data only, using BMRB entry 16300.³⁸ Thus, in this case the ensemble has been calculated based on experimental data different from those used for our selection. The selected ensemble consists of 37 members. PCA analysis reveals that the selected ensemble still covers the conformational variability described in the first four PCA modes of the original ensemble (Figure S3). Analysis of the secondary structure of the original and the selected ensembles shows that the positions of the helical segments are well reproduced in the selected ensemble, although the propensities are usually higher. It is important to stress here that the chemical shift data set does not exactly correspond to the

experiments used for the calculation of the original ensemble, thus the slightly different secondary structure propensities might reflect relevant dependence on the exact conditions. In summary, we can conclude that the most relevant features of the conformational diversity of the disordered *α*-synuclein are still reasonably well captured with the selected ensemble, although the accuracy of the ensemble can not be expected to be comparable to that calculated based on PRE restraints.

CONCLUSIONS

Protein structural ensembles reflecting internal dynamics require specialized validation and analysis tools. The CoNSEnsX⁺ web server described here provides a convenient and reproducible way to assess the correspondence of ensembles to experimental NMR parameters. The selection feature allows both the generation of ensembles consistent with the data and to assess the extent of possible overfitting. Our selection approach does not guarantee to produce the smallest possible ensemble with a given level of correspondence to experimental data, but still sets an upper limit on the size of such well-performing ensembles. As our deterministic greedy approach starts with the single conformer best reflecting the selected data, the structural diversity of the selected ensemble is expected to be required to account for the experimental results and is not expected to be a result of arbitrary choice of the ensemble members. This consideration is in line with our results on various sample cases of selection described above.

Potential further uses of the CoNSEnsX⁺ tool could be in the field of protein structure prediction, because ensemble-based evaluation of protein models is an important issue in assessing the accuracy of the predicted conformations.³⁹ In addition, the selection feature can aid the choice of suitable low-dimensional projections to direct effective conformational sampling for proteins.⁴⁰

METHODS

Implementation of the CoNSEnsX⁺ Server. The CoNSEnsX⁺ method was implemented in Python, and the webserver interface has been built using the Django framework. The web application can be used with any webserver which implements the WSGI (webserver gateway interface) specification, like Nginx or Apache.

The following external libraries and programs have been incorporated:

- NMRPystar library for parsing BMRB parameter files
- ProDy protein structure analysis library for handling atomic coordinates including structure superposition⁴¹
- SHIFTX program for the calculation of NMR chemical shifts from the structures,⁴² chosen over the more accurate SHIFTX2⁴³ for its higher speed
- PALES program to provide a standard way of calculating residual dipolar couplings from the structures⁴⁴
- PRIDE-NMR program to assess the general correspondence of the conformers to NOE-based distance restraints⁴⁵

Uploaded data are stored in a SQL database on the server, as well as the pages with the results of the calculations. Since no registration is required to use the service, the result pages of the calculations ran by the user can be accessed using the permalink shown on the corresponding page. All calculated data are also stored in the SQL database to allow easy reuse. Loading the permalink gets the source of the generated result page in a single SQL query.

Ensemble-averaged values of all back-calculated parameters except NOE distances are calculated by simply taking the arithmetic mean of the values obtained for the individual conformers. NOE distances are averaged using the r^{-6} scheme for intramolecular distances in the case of restraints belonging to more than a single atom pair. For the full ensemble, r^{-6} or r^{-3} averaging can be chosen by the user. All back-calculated data except S^2 order parameters are stored for each conformer for easy reuse in the selection algorithm.

The implemented selection algorithm is a version of a deterministic greedy approach. The process starts from a single conformer (for most experimental data types) or an ensemble of two conformers (for S^2 order parameters) best corresponding to the included parameters according to the chosen measure. The approach iteratively adds further conformers resulting in the best possible correspondence for the enlarged ensemble at each step. It is generally required that the addition of the new conformers enhances the correspondence, whenever possible. To allow the generation of well-performing ensembles not necessarily achievable with the requirement of monotonic increase of the correspondence with each addition step, a so-called overdrive parameter can be set to allow decrease of the correspondence at several consecutive steps. Further extending the ensemble with additional conformers might yield an overall better correspondence even in such cases. The minimum and maximum size of the final subensemble to be returned can be specified as well as the measure by which the subensembles will be evaluated (Pearson correlation, Q-factor, or RMSD). If the calculation contained RDC sets and single value decomposition (SVD) was enabled, all the RDC parameters in a given set must have the same weight, if included in the target function. This policy is enforced by the weighting of the target function and can not be overridden. The output is the correspondence of the parameters to the full ensemble and the selected subensemble as well as the subensemble as a downloadable PDB file, in which the models are superimposed and their number in the original ensemble is marked.

Molecular Dynamics Calculations and Analysis of the Test Cases. For the generation of SH3 and Pin1 ensembles, molecular dynamics simulations were run with GROMACS 4.5.5,⁴⁶ when stated, using the accelerated molecular dynamics (AMD) scheme²⁸ as implemented previously by our group in this package.²⁹ PCA analysis in all cases was performed with ProDy.⁴¹

For the SH3 test case, a 100 ns simulation was run with a time step of 1 fs at 300 K in implicit solvent using the GBSA algorithm.²⁷ The first model in the PDB ensemble 2A36²⁶ was used as an input structure. The final 20 ns of the obtained trajectory was used as a reference ensemble for which various types of parameters (chemical shifts, five independent RDC sets, and backbone S^2 order parameters) were calculated. As input for an independent selection, an initial pool covering a larger conformational space was generated with the AMD scheme. Four parallel 20 ns AMD simulations were carried out from the same starting structure but with different initial atomic velocities with an AMD boost potential of $E = 2400$ kJ/mol and an acceleration factor $\alpha = 0.5$. The resulting set of conformers, denoted the AMD pool, is composed of the last 18.75 ns of the four AMD trajectories (3044 conformers). From this pool a number of subensembles were selected using various combinations of the “experimental” parameters calculated for the reference ensemble.

For the Pin1 test case, the first conformer of the PDB ensemble 2RUC³⁴ as a starting structure of three different AMD simulations (Table 1). Other parameters were the same as described for the SH3 test case.

Table 1. AMD Parameters Used for the Calculations of Different Pin1 Ensembles

ensemble	AMD4800	AMD5000	AMD5200
boost energy (kJ/mol)	4800	5000	5200
α	1.0	0.8	0.5
simulation time (ns)	20	5	5

Visual inspection of the ensembles was performed with MOLMOL.³⁵

Secondary structure content evaluation of the α -synuclein ensembles was performed by running DSSPcont⁴⁷ on each ensemble member and averaging the results.

■ ASSOCIATED CONTENT

● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.7b00066.

Figure S1: PCA and displacement plots of selected SH3 ensembles obtained by selection from the AMD pool. Figure S2: PCA analysis of different SH3 ensembles selected from the AMD pool. Figure S3: PCA plots of the full α -synuclein ensemble and the selected one, helical propensities of the ensembles (PDF)

Table S1: Correspondence of parameters in selected SH3 ensembles to those derived from the reference ensemble. Table S2: Correspondence of back-calculated and experimental chemical shifts in selected Pin1 ensembles. Table S3: Correspondence of back-calculated and experimental shifts in the selected α -synuclein ensembles (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gaspari.zoltan@itk.ppke.hu. Phone: +36 (1)886 4780. Fax: +36 (1)886 4724.

ORCID

Dániel Dudola: 0000-0002-5086-3213

Zoltán Gáspári: 0000-0002-8692-740X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the support of the National Research, Development and Innovation Office (NKFIH) through grant no. NF104198. The authors thank Prof. Gyula Batta from the University of Debrecen for insightful discussions regarding the manuscript.

■ REFERENCES

- (1) Kosol, S.; Contreras-Martos, S.; Cedeno, C.; Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **2013**, *18*, 10802–10828.
- (2) Jensen, M. R.; Ruigrok, R. W.; Blackledge, M. Describing Intrinsically Disordered Proteins at Atomic Resolution by NMR. *Curr. Opin. Struct. Biol.* **2013**, *23*, 426–435.
- (3) Kukic, P.; Alvin Leung, H. T.; Bemporad, F.; Aprile, F. A.; Kumita, J. R.; De Simone, A.; Camilloni, C.; Vendruscolo, M. Structure and

Dynamics of the Integrin LFA-1 I-Domain in the Inactive State Underlie its Inside-Out/Outside-In Signaling and Allosteric Mechanisms. *Structure* **2015**, *23*, 745–753.

(4) Angyan, A. F.; Gaspari, Z. Ensemble-Based Interpretations of NMR Structural Data to Describe Protein Internal Dynamics. *Molecules* **2013**, *18*, 10548–10567.

(5) Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; Sussman, J.; Svergun, D. I.; Uversky, V. N.; Vendruscolo, M.; Wishart, D.; Wright, P. E.; Tompa, P. pE-DB: a Database of Structural Ensembles of Intrinsically Disordered and of Unfolded Proteins. *Nucleic Acids Res.* **2014**, *42*, D326–335.

(6) Best, R. B.; Vendruscolo, M. Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure* **2006**, *14*, 97–106.

(7) Vranken, W. F. NMR Structure Validation in Relation to Dynamics and Structure Determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *82*, 27–38.

(8) Richter, B.; Gsponer, J.; Varnai, P.; Salvatella, X.; Vendruscolo, M. The MUMO (Minimal Under-Restraining Minimal Over-Restraining) Method for the Determination of Native State Ensembles of Proteins. *J. Biomol. NMR* **2007**, *37*, 117–135.

(9) Fenwick, R. B.; Esteban-Martin, S.; Richter, B.; Lee, D.; Walter, K. F.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. Weak Long-range Correlated Motions in a Surface Patch of Ubiquitin Involved in Molecular Recognition. *J. Am. Chem. Soc.* **2011**, *133*, 10336–10339.

(10) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.

(11) Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W. Y.; Forman-Kay, J. D. Characterization of Disordered Proteins with ENSEMBLE. *Bioinformatics* **2013**, *29*, 398–399.

(12) Best, R. B.; Vendruscolo, M. Determination of Protein Structures Consistent with NMR Order Parameters. *J. Am. Chem. Soc.* **2004**, *126*, 8090–8091.

(13) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous Determination of Protein Structure and Dynamics. *Nature* **2005**, *433*, 128–132.

(14) Hess, B.; Scheek, R. M. Orientation Restraints in Molecular Dynamics Simulations Using Time and Ensemble Averaging. *J. Magn. Reson.* **2003**, *164*, 19–27.

(15) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. Recognition Dynamics up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **2008**, *320*, 1471–1475.

(16) Bonvin, A. M.; Brunger, A. T. Conformational Variability of Solution Nuclear Magnetic Resonance Structures. *J. Mol. Biol.* **1995**, *250*, 80–93.

(17) Bertini, I.; Ferella, L.; Luchinat, C.; Parigi, G.; Petoukhov, M. V.; Ravera, E.; Rosato, A.; Svergun, D. I. MaxOcc: a Web Portal for Maximum Occurrence Analysis. *J. Biomol. NMR* **2012**, *53*, 271–280.

(18) Huang, Y. J.; Rosato, A.; Singh, G.; Montelione, G. T. RPF: a Quality Assessment Tool for Protein NMR Structures. *Nucleic Acids Res.* **2012**, *40*, W542–546.

(19) Berjanskii, M.; Liang, Y.; Zhou, J.; Tang, P.; Stothard, P.; Zhou, Y.; Cruz, J.; MacDonell, C.; Lin, G.; Lu, P.; Wishart, D. S. PROSESS: a Protein Structure Evaluation Suite and Server. *Nucleic Acids Res.* **2010**, *38*, W633–640.

(20) Ángyán, A. F.; Szappanos, B.; Perczel, A.; Gáspári, Z. CoNSEnsX: an Ensemble View of Protein Structures and NMR-Derived Experimental Data. *BMC Struct. Biol.* **2010**, *10*, 39.

(21) Louhivuori, M.; Otten, R.; Lindorff-Larsen, K.; Annala, A. Conformational Fluctuations Affect Protein Alignment in Dilute Liquid Crystal Media. *J. Am. Chem. Soc.* **2006**, *128*, 4371–4376.

(22) Salvatella, X.; Richter, B.; Vendruscolo, M. Influence of the Fluctuations of the Alignment Tensor on the Analysis of the Structure

and Dynamics of Proteins Using Residual Dipolar Couplings. *J. Biomol. NMR* **2008**, *40*, 71–81.

(23) Montalvao, R. W.; De Simone, A.; Vendruscolo, M. Determination of Structural Fluctuations of Proteins from Structure-Based Calculations of Residual Dipolar Couplings. *J. Biomol. NMR* **2012**, *53*, 281–292.

(24) Habeck, M.; Rieping, W.; Nilges, M. Bayesian Estimation of Karplus Parameters and Torsion Angles from Three-Bond Scalar Couplings Constants. *J. Magn. Reson.* **2005**, *177*, 160–165.

(25) Wang, A. C.; Bax, A. Determination of the Backbone Dihedral Angles ϕ in Human Ubiquitin from Reparametrized Empirical Karplus Equations. *J. Am. Chem. Soc.* **1996**, *118*, 2483–2494.

(26) Bezsonova, I.; Singer, A.; Choy, W. Y.; Tollinger, M.; Forman-Kay, J. D. Structural Comparison of the Unstable drkN SH3 Domain and a Stable Mutant. *Biochemistry* **2005**, *44*, 15550–15560.

(27) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304.

(28) Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A. Implementation of Accelerated Molecular Dynamics in NAMD. *Comput. Sci. Discovery* **2011**, *4*, 015002.

(29) Fizil, A.; Gaspari, Z.; Barna, T.; Marx, F.; Batta, G. "Invisible" Conformers of an Antifungal Disulfide Protein Revealed by Constrained Cold and Heat Unfolding, CEST-NMR Experiments, and Molecular Dynamics Calculations. *Chem. - Eur. J.* **2015**, *21*, S136–S144.

(30) Borchert, T. V.; Mathieu, M.; Zeelen, J. P.; Courtneidge, S. A.; Wierenga, R. K. The Crystal Structure of Human CskSH3: Structural Diversity Near the RT-Src and n-Src Loop. *FEBS Lett.* **1994**, *341*, 79–85.

(31) Zhao, D.; Wang, X.; Peng, J.; Wang, C.; Li, F.; Sun, Q.; Zhang, Y.; Zhang, J.; Cai, G.; Zuo, X.; Wu, J.; Shi, Y.; Zhang, Z.; Gong, Q. Structural Investigation of the Interaction Between the Tandem SH3 Domains of c-Cbl-Associated Protein and Vinculin. *J. Struct. Biol.* **2014**, *187*, 194–205.

(32) Velazquez, H. A.; Hamelberg, D. Conformational Selection in the Recognition of Phosphorylated Substrates by the Catalytic Domain of Human Pin1. *Biochemistry* **2011**, *50*, 9605–9615.

(33) Czajlik, A.; Kovacs, B.; Permi, P.; Gaspari, Z. Fine-tuning the Extent and Dynamics of Binding Cleft Opening as a Potential General Regulatory Mechanism in Parvulin-Type Peptidyl Prolyl Isomerases. *Sci. Rep.* **2017**, *7*, 44504.

(34) Xu, N.; Tochio, N.; Wang, J.; Tamari, Y.; Uewaki, J.; Utsunomiya-Tate, N.; Igarashi, K.; Shiraki, T.; Kobayashi, N.; Tate, S. The C113D Mutation in Human Pin1 Causes Allosteric Structural Changes in the Phosphate Binding Pocket of the PPIase Domain Through the Tug of War in the Dual-histidine Motif. *Biochemistry* **2014**, *53*, 5568–5578.

(35) Koradi, R.; Billeter, M.; Wuthrich, K. MOLMOL: a Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graphics* **1996**, *14*, 51–55.

(36) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. Determination of the Free Energy Landscape of Alpha-Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. *J. Am. Chem. Soc.* **2009**, *131*, 18314–18326.

(37) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping Long-Range Interactions in Alpha-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127*, 476–477.

(38) Rao, J. N.; Kim, Y. E.; Park, L. S.; Ulmer, T. S. Effect of Pseudorepeat Rearrangement on Alpha-Synuclein Misfolding, Vesicle Binding, and Micelle Binding. *J. Mol. Biol.* **2009**, *390*, S16–S29.

(39) Jamroz, M.; Kolinski, A.; Kihara, D. Ensemble-Based Evaluation for Protein Structure Models. *Bioinformatics* **2016**, *32*, i314–i321.

(40) Novinskaya, A.; Devaurs, D.; Moll, M.; Kavrakli, L. E. Defining Low-Dimensional Projections to Guide Protein Conformational Sampling. *J. Comput. Biol.* **2017**, *24*, 79–89.

(41) Bakan, A.; Dutta, A.; Mao, W.; Liu, Y.; Chennubhotla, C.; Lezon, T. R.; Bahar, I. Evol and ProDy for Bridging Protein Sequence Evolution and Structural Dynamics. *Bioinformatics* **2014**, *30*, 2681–2683.

- (42) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and Accurate Calculation of Protein ^1H , ^{13}C and ^{15}N Chemical shifts. *J. Biomol. NMR* **2003**, *26*, 215–240.
- (43) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR* **2011**, *50*, 43–57.
- (44) Zweckstetter, M. NMR: Prediction of Molecular Alignment from Structure Using the PALES Software. *Nat. Protoc.* **2008**, *3*, 679–690.
- (45) Angyan, A. F.; Perczel, A.; Pongor, S.; Gaspari, Z. Fast Protein Fold Estimation from NMR-Derived Distance Restraints. *Bioinformatics* **2008**, *24*, 272–275.
- (46) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (47) Carter, P.; Andersen, C. A.; Rost, B. DSSPcont: Continuous Secondary Structure Assignments for Proteins. *Nucleic Acids Res.* **2003**, *31*, 3293–3295.