Journal of Magnetic Resonance 213 (2011) 182–186

Contents lists available at SciVerse ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Communication

Generative probabilistic models extend the scope of inferential structure determination

Simon Olsson [a], Wouter Boomsma [b], Jes Frellsen [a], Sandro Bottaro [b], Tim Harder [a], Jesper Ferkinghoff-Borg [b,*], Thomas Hamelryck [a,*]

[a] Bioinformatics Center, University of Copenhagen, Department of Biology, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark
[b] Biomedical Engineering, DTU Elektro, Technical University of Denmark, Ørsteds Plads, DK-2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Article history:
Received 20 June 2011
Revised 19 August 2011
Available online 6 September 2011

Keywords:
Inferential structure determination
Generative probabilistic models
Sparse data

ABSTRACT

Conventional methods for protein structure determination from NMR data rely on the *ad hoc* combination of physical forcefields and experimental data, along with heuristic determination of free parameters such as weight of experimental data relative to a physical forcefield. Recently, a theoretically rigorous approach was developed which treats structure determination as a problem of Bayesian inference. In this case, the forcefields are brought in as a prior distribution in the form of a Boltzmann factor. Due to high computational cost, the approach has been only sparsely applied in practice. Here, we demonstrate that the use of generative probabilistic models instead of physical forcefields in the Bayesian formalism is not only conceptually attractive, but also improves precision and efficiency. Our results open new vistas for the use of sophisticated probabilistic models of biomolecular structure in structure determination from experimental data.

1. Introduction

Current methods for macromolecular structure determination rely on the seminal idea of hybrid energy minimization introduced by Jack and Levitt [1]. However, the choice of model parameters, such as the weight of the experimental data with respect to a physical force field, is intrinsically problematic in this approach – a fact that was already recognized in the original study [1]. With a growing number of sources of experimental data used in protein structure determination, estimation of weights and other nuisance parameters is becoming increasingly problematic. Current methodology relies on a more or less arbitrary choice of these parameters, using heuristic approaches [2]. While a persistent concern towards the applied heuristics has been evident in the literature [3,2,4], only few quantitative methods have been described to rigorously determine these nuisance parameters [4,5]. These methods, and the underlying Bayesian approach are referred to as inferential structure determination (ISD).

Bayesian probabilistic inference has previously shown great potential in macromolecular structure determination [2,6]. However, the scope of the approach has been limited due to excessive computational demands. The current study describes a new approach to inferential structure determination which draws on the use of generative probabilistic models. Generative probabilistic models, or GPMs, are probabilistic models that allow sampling. Here, we demonstrate that the use of GPMs greatly increases efficiency, precision and scope of rigorous inferential structure determination. As these GPMs contain information about protein structure, they may supersede physical forcefields – especially in cases where data is very sparse.

2. Methods

In the ISD approach, samples are drawn from a joint posterior distribution over conformational space, X, and model parameter space, n, given experimental data, D, and prior knowledge, I:

$$p(X,n|D,I) \propto p(D|X,n,I)p(n|I)p(X|I).$$

Consequently, a natural result of posterior sampling is an ensemble of conformers representing the experimental uncertainty. That is, the Bayesian formalism accounts for uncertainty and degeneracy, a feature that is difficult to obtain when using schemes that minimize a hybrid energy consisting of a physical and a data-dependent term [7,8].

In ISD, a physical forcefield E_{phys} enters the Bayesian framework as a conformational prior through a canonical ensemble $p(X|I) \propto e^{-\beta E_{\text{phys}}}$, where $\beta = 1/kT$, k is Boltzmann's constant and T is the temperature [2]. The data enters as a likelihood function, $p(D|X,n,I)$; its product with the prior distributions, $p(n|I)p(X|I)$,

* Corresponding authors.
E-mail addresses: solsson@binf.ku.dk (S. Olsson), jfb@elektro.dtu.dk (J. Ferkinghoff-Borg), thamelry@binf.ku.dk (T. Hamelryck).

1090-7807/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2011.08.039

results in the posterior distribution, $p(X,n|D,I)$. When the posterior is defined in this way, Markov Chain Monte Carlo (MCMC) sampling requires evaluation of both likelihood and priors explicitly, in each step. This can potentially lead to substantial computational costs. Conversely, using no, or a uninformative forcefield, leaves a vast conformational space [9]. Here, we use GPMs of local protein structure instead of the Boltzmann distribution of a physical forcefield. Consequently, we demonstrate that the explicit evaluation of the prior can be avoided altogether as the information of the prior enters the posterior distribution through sampling.

Recently, our group has published several GPMs of protein conformational space, describing backbone (TorusDBN) [10,11] and sidechain (Basilisk) [12] dihedral angles. These models only provide structural information on a local sequential scale, ideally complementing the long-range information obtained from NMR nuclear Overhauser enhancements experiments (NOE). As generalizations of the commonly used fragment- [13] and rotamer-libraries [14], and related potentials that involve discretization [15], these GPMs also serve to reduce the complexity of the conformational space. The particular GPMs applied here use continuous angular probability distributions to avoid the intrinsic limitations caused by discretization [16]. Furthermore, since these GPMs are probability distributions, probabilities of arbitrary conformations can be evaluated, which is not generally possible for fragment- and rotamer libraries. Consequently, the full posterior probability can be evaluated explicitly when necessary. Here, we demonstrate that the use of GPMs as conformational proposal distributions can dramatically increase convergence in MCMC sampling of protein conformers from a posterior distribution, in addition to providing an increase in precision.

The GPMs, TorusDBN and Basilisk, enter the ISD approach as $p(X|I) \propto p(b|a)p(\chi|a)$, where a denotes amino acid sequence, while b and χ denote backbone and sidechain conformations respectively. Thus, during simulation we alternate between moving in backbone and sidechain conformational space, conditioned on amino acid sequence. Following Rieping et al. we assume idealized Engh–Huber bond lengths [17] and parameterize conformations as sets of torsion angles [18]. Variations in the bond angles were allowed to facilitate conformational sampling [19].

We used a generalized ensemble Metropolis–Hastings sampling scheme to draw samples from the posterior distribution. To prioritize search in relevant regions of the conformational space we adopted the $1/k$-ensemble implemented using the generalized multi-histogram equations [20,21]. The $1/k$-ensemble allows sampling independently of temperature, thus avoiding nuisance parameters such as the number of replicas, and their temperature span. It is, however, important to stress that the statistical information provided by this sampling scheme is equivalent to the Replica Exchange Monte Carlo scheme used in the original ISD study [22]. We employ the log-normal formulation of the NOE data to evaluate $p(D|b,\chi,n,I)$, as this provides the least biased formulation of the likelihood [5].

To assess the performance of TorusDBN and Basilisk as conformational priors, and for comparison to previous results, we created a set of conformers corresponding to the lowest posterior samples, using the very sparse (154 constraints) SH3 FYN domain data [2] and the TRP-Cage data set [28]. As a model baseline, we carried out the same simulations without the models of local protein structure. This simple hard-sphere potential corresponds to the use of a prior distribution reminiscent of that of the original ISD implementation [2].

2.1. Posterior sampling

As described previously, we sample from the joint posterior distribution $p(X,n|D,I)$ [2]:

$$p(X,\mathbf{n}|D,I) \propto \sigma^{-(n+1)}\gamma^{-1}\exp\left[-\frac{1}{2\sigma^2}\chi^2(d(\chi,b),I)\right]p(\chi|a)p(b|a),$$

with the log-normal chi-square: $\chi^2(d,D)=\sum_i^n \log^2(\gamma d_i^{-\alpha}/D_i), D_i$ are experimental data and d_i calculated distances [5]. χ and b are the sidechain and backbone dihedrals, respectively. γ and σ are ISPA (isolated spin-pair approximation) equilibration parameter and experimental uncertainty, respectively. A power $\alpha = -1$ was used here as all data were derived distances.

Here, we cannot employ the Gibbs sampling scheme applied in Rieping et al. [2], due to the inherent absence of an explicit temperature in the $1/k$ ensemble. This absence of an explicit temperature makes the implementation of the soft-sphere potential employed previously difficult without introduction of additional heuristics, and was therefore avoided [2]. Instead, we here use a Metropolis–Hastings approach, where the involved parameters are updated one at the time. The $1/k$ ensemble allows us to sample the conformational- and nuisance-space efficiently.

Low acceptance rates in the nuisance sampling was avoided by introducing a scheme exploiting the information about the current state. For the nuisance parameters, $n=\{\gamma,\sigma\}$, a log-change is proposed from a log-normal distribution with a standard deviation

$$\sigma_{n_i} = \frac{1.0}{\max\left(\left\|\frac{\partial \log p(X,n|D,I)}{\partial n_i}\right\|, 1.0\right)}.$$

This expression was derived using standard error propagation and adds a simple regularizer which ensures a maximum standard deviation of 1.0 [23]. As a result, we can draw samples efficiently from the joint posterior distribution without the temperature dependent Gibbs sampling scheme. Using the log-normal distribution in this way we can ensure being in the right domain. We avoid additional bias from the log-normal distribution in the posterior, by dividing out the bias in the Monte Carlo acceptance ratio. For completeness, the analytical expressions of the standard deviations are shown here:

$$\sigma_\sigma = \frac{1.0}{\max\left(\left\|-\frac{\chi^2(d(\chi,b),I)}{\sigma^2}+v\right\|, 1.0\right)}$$

where v is the number of datapoints, and:

$$\sigma_\gamma = \frac{1.0}{\max\left(\left\|\frac{v\log\gamma-\sum_i^v k_i}{\sigma^2}\right\|, 1.0\right)},$$

with $k_i = \ln\frac{I^i_{\text{obs}}}{I^i_{\text{calc}}}$ corresponding to the log-ratio between the observed and back-calculated experimental data.

For sampling of the conformational space, a series of MCMC moves for backbone (pivot, local [19] and semi-local [24]) and sidechain conformations were employed. All applied moves fulfill detailed balance, and were chosen with even probability with respect to backbone and sidechain conformational and nuisance space. TorusDBN was extended to account for small deviations from ideal *cis*/*trans*-angles, using a normal distribution with mean at the ideal values and a standard deviation of five degrees. In the baseline model, all angles b, χ were sampled uniformly in the interval $[0,2\pi]$. Note that Basilisk was used in a backbone independent fashion for simplicity [12]. Samples were accepted or rejected according to the generalized $1/k$ ensemble [20]. Convergence was assessed through inspection of diagnostics provided by Muninn: the multi-histogram implementation of the generalized ensemble (http://www.muninn.sourceforge.net/). It is important to stress that convergence of histograms necessarily reflect convergence of posterior samples, additional sampling allow generation of more refined ensembles.

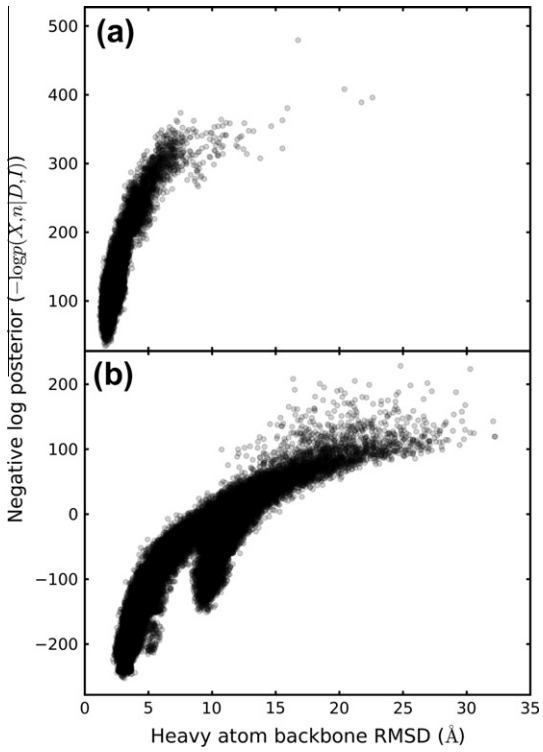

Fig. 1. Scatter plots of the RMSD of conformational samples to the crystal structure of SH3 FYN (PDB:1SHF chain A) versus $-\log p(X,n|D,I)$ (posterior density) for (a) TorusDBN and Basilisk and (b) the baseline prior after 400 million MCMC steps. Samples are from the $1/k$ ensemble.

Table 1
VADAR and PROCHECK structure quality statistics for the previously published ensemble (PDB: 1ZBJ) (**1ZBJ**) [2] and current SH3 FYN (**GPMs**) ensembles and reference values presented by VADAR (**Ref**). ϕ, ψ core, allowed, generous and outside denote distinct regions of the Ramachandran plot of decreasing favoredness. ω core denotes the percentage of ω-angles in the most favored region (the three other classes are not shown here). Packing defects, free energy folding, percentage of residues 95% buried and buried charges denotes the number of packing defects, free energy of folding and bury ratios for residues and charges, respectively [25]. Percentile reference values were normalized. PROCHECK *G*-factors reflect average log-odds of (ϕ,ψ), $(\chi1,\chi2)$, $(\chi1)$ and overall dihedral angle combinations.

VADAR Dihedral prior	**1ZBJ**	**GPMs**	**Ref**
ϕ, ψ core	68.95 ± 4.19%	88.33 ± 2.85%	91.84%
ϕ, ψ allowed	27.6 ± 4.12%	9.96 ± 3.17%	7.14%
ϕ, ψ generous	1.7 ± 1.27%	1.65 ± 1.50%	1.02%
ϕ, ψ outside	0.0 ± 0.0%	0.05 ± 0.0%	0.0%
ω core	100.0 ± 0.0%	91.0 ± 2.17%	97%
ω allowed	0.0 ± 0.0%	8.0 ± 2.61%	3%
ω generous	0.0 ± 0.0%	1.0 ± 1.49%	0%
Packing defects	11.95 ±2.85	5.95 ± 2.06	4.0
Free energy fold	−40.7 ± 1.88	−46.07 ± 2.06	−42.39
Res. 95% buried	2.25 ± 1.22	4.30 ± 1.90	6.0
Buried charges	0.15 ± 0.30	0.30 ± 0.56	0.0
PROCHECK Dihedral prior	**1ZBJ**	**GPMs**	
G-factor (ϕ,ψ)	−1.41	−0.72	
G-factor $(\chi1,\chi2)$	−1.82	0.25	
G-factor $(\chi1$ only)	−0.54	0.20	
G-factor (overall)	−1.43	−0.28	

3. Results

3.1. SH3 FYN

When employing the GPMs, the sampling of the posterior distribution defined by the sparse SH3 FYN data set converges in less than 36 h of computation time on a single standard CPU core. In comparison, the previously published ISD ensemble derived from the same data set took 3 days on a 50 core computer cluster [2]. Even given the increase in average computational power since 2005, this is a substantial increase in the efficiency. We do not observe convergence within the same simulation time when applying the baseline model. This illustrates clearly how the GPMs increase efficiency of posterior sampling.

Performing posterior sampling with the baseline prior, gives rise to two distinct conformational basins (Fig. 1b). There is an excited basin corresponding to the mirror image of the native basin. The local geometry of this basin is highly unfavorable. The second basin corresponds to the correct, native fold, observed in the crystal structure. The latter of the two basins is the only one observed when using the informative GPMs as conformational priors (Fig. 1a). Evidently, the experimental data likelihood in conjunction with the baseline prior only modestly distinguishes between the two folds, resulting in slow convergence due to an excessive conformational multiplicity. The basin with the correct fold is not thoroughly explored within the given time frame, resulting in relatively inaccurate structures among the 20 highest posterior conformer ensemble (Fig. 2b). In contrast, the ensembles obtained within the same simulation time using the TorusDBN and Basilisk priors accurately capture the native state (Fig. 2a). This result illustrates the importance of prior information to resolve degeneracies in sparse experimental data. While avoidance of poor

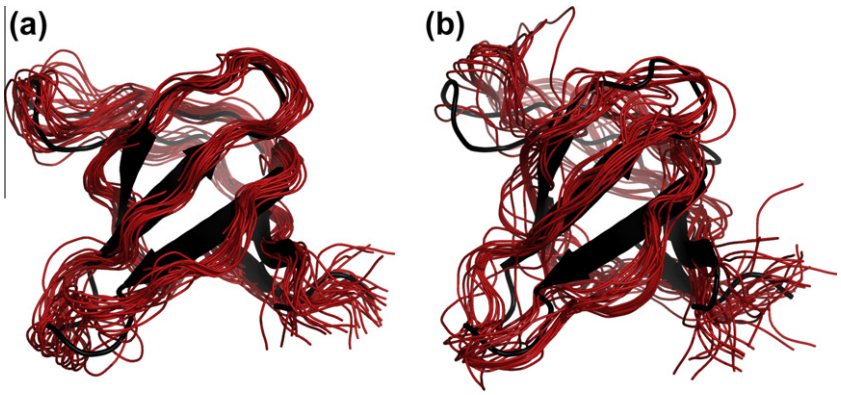

Fig. 2. Illustration of 20 of the samples with the highest posterior probability using (a) TorusDBN and Basilisk (RMSD: 1.74 ± 0.17Å) or (b) the baseline prior (RMSD: 3.12 ± 0.24 Å), after 400 million MCMC steps. Conformations are aligned to PDB: 1SHF chain A (shown in a black cartoon representation). Figure prepared using PyMOL (DeLano Scientific LLC).

stereochemistry has been pointed out previously as a feature of the ISD approach [2], degeneracy due to poor local structure has remained unaddressed.

The mean heavy-atom (C_α, C and N) root mean square deviation (RMSD) to the crystal structure from the 20 highest posterior probability structures (see Fig. 2) is comparable to the previously published ISD ensemble (1.84 ± 0.20 Å, PDB: 1ZBJ). However, statistics derived from structure validation server VADAR [25], WHATIF [26] and PROCHECK [27] were vastly improved (see Table 1 and Supplementary material) with respect to both packing quality and local structure. Importantly, clustering of (ϕ,ψ)-angle pairs in less favorable regions of the Ramachandran space is reduced dramatically (see SI). Other structure quality indicators such as number of buried charges remain unchanged. While the improvement in local

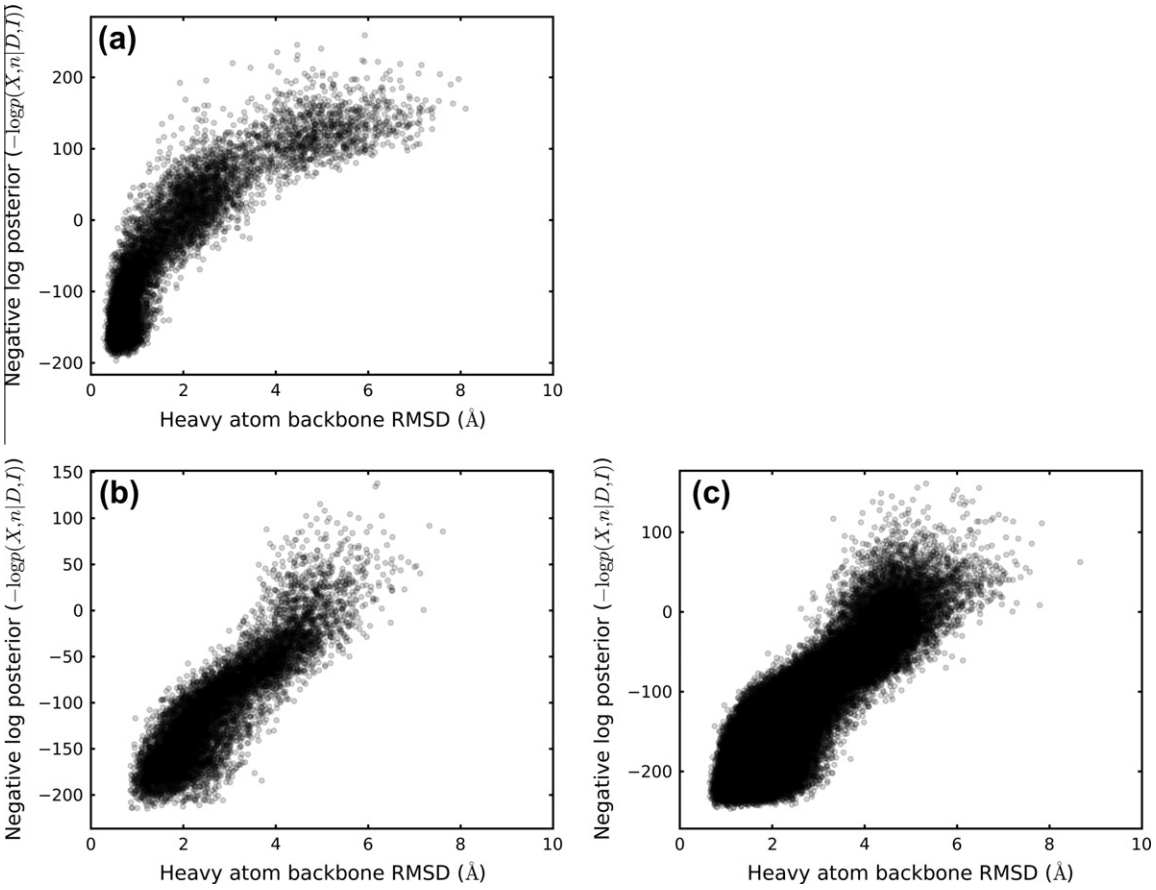

Fig. 3. Scatter plots of RMSD of conformational samples to the previously published NMR structure of TRP-Cage (PDB:1L2Y) versus $-\log p(X,n|D,I)$ (posterior density) for (a) TorusDBN and Basilisk and (b) baseline prior after 50 million MCMC steps; (c) baseline prior after 500 million MCMC steps. Samples are from the $1/k$ ensemble.

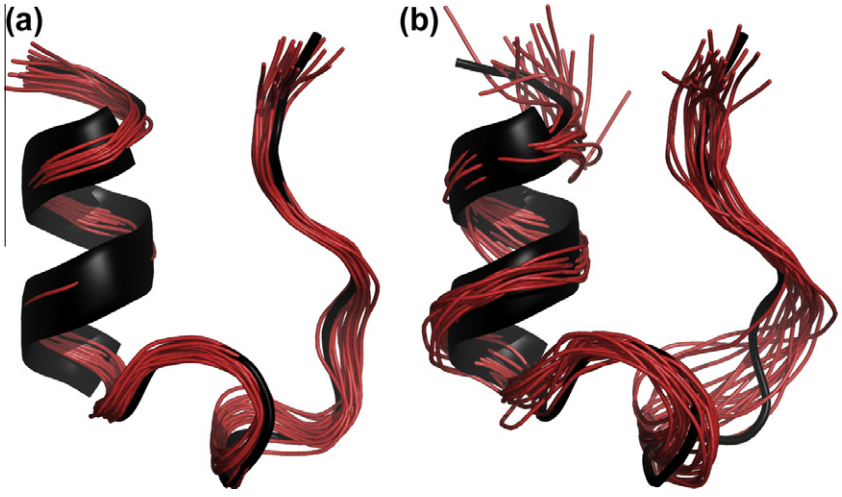

Fig. 4. Illustration of 20 of the samples with the highest posterior probability using (a) TorusDBN and Basilisk (RMSD: 0.63 ± 0.12 Å) or (b) baseline prior (RMSD: 1.41 ± 0.39 Å), after 50 million MCMC samples. Conformations are aligned to PDB: 1L2Y (shown in a black cartoon representation). Figure prepared using PyMOL (DeLano Scientific LLC).

structure is an expected consequence of the information contained in TorusDBN and Basilisk, non-local structure quality parameters such as packing defects hint increase in accuracy. The nuisance parameters σ, γ were estimated to be 0.11 ± 0.01 and 1.00 ± 0.01, respectively. These values deviate somewhat from the estimates obtained previously. The discrepancy may be linked to a different conformational prior distribution [2].

3.2. TRP-cage

In addition to increasing efficiency and precision, GPMs can account for the information derived from ambiguous NOE constraints. We demonstrate this point on the TRP-cage data set [28]. Of the reported 169 restraints, 37 involve pseudo atoms, which strictly speaking yields them ambiguous. In these particular calculations, the restraints were therefore not included. The resulting set of unambiguous NOE restraints are insufficiently informative to distinguish native-like structures from conformers with an RMSD of up to 3 Å from the previously published NMR structure. However, when we use the GPMs as structural priors, we obtain an ensemble of high resemblance with the previously published structure.

The simulations of TRP-cage were performed identically to those of SH3 FYN using 50 million MCMC steps. Both simulations complete within a few hours (see Fig. 3a and b). The pattern observed for SH3 FYN emerges again: when using GPMs convergence was reached within the simulation time, whereas convergence was not reached using the baseline model. Extending the simulation time with the baseline model to 500 million MCMC steps results in convergence (Fig. 3c). However, the resulting 20 highest posterior ensemble is of significantly lower quality (RMSD: 1.24 ± 0.39 Å) than the ensemble obtained using the GPMs running for 50 million MCMC steps, Fig. 4a. With these results we again demonstrate how efficiency is gained when employing GPMs in the ISD approach. In addition the results illustrate, how the unambiguous constraints [28] can be complemented by the local information contained in the GPMs.

4. Conclusions

In both examples presented here, the difference in accuracy of the selected ensembles is modest, with mean RMSD differences of at most 1 Å. However, the highest probability (or lowest energy) criterion for selection of conformation for these ensembles may not only underestimate the spread of the ensemble [29,30], but also ignore severe degeneracies (see Figs. 1 and 3). This points to the importance of using appropriate prior information when analyzing sparse data and suggests extra caution be taken when selecting these ensembles.

This communication describes how generative probabilistic models can be applied to significantly increase efficiency and precision of inferential structure determination. As a natural extension, we propose the development of more specialized GPMs, drawing on additional prior information such as protein family membership or chemical shifts. Such models would presumably resolve degeneracies to an even greater extent, further increasing the scope, efficiency and precision of the inferential structure determination approach.

Acknowledgments

We thank F.M. Poulsen, K. Lindorff-Larsen and J.H. Jensen for critically reading the manuscript. T. Harder is funded by the Danish Council for Strategic Research (NaBiIT, 2106-06-009). W. Boomsma is funded by the Danish Council for Independent Research (FNU, 272-08-0315). J. Frellsen and S. Olsson are funded by the Danish Council for Independent Research (FTP, 09-066546). S. Bottaro acknowledge funding from Radiometer, DTU.Elektro.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2011.08.039.

References

[1] A. Jack, M. Levitt, Refinement of large structures by simultaneous minimization of energy and R factor, Acta Crystallogr. A 34 (1978) 931–935.
[2] W. Rieping, M. Habeck, M. Nilges, Inferential structure determination, Science 309 (2005) 303–306.
[3] M. Williamson, C. Craven, Automated protein structure calculation from NMR data, J. Biomol. NMR 43 (2009) 131–143. doi:10.1007/s10858-008-9295-6.
[4] M. Habeck, W. Rieping, M. Nilges, Weighting of experimental evidence in macromolecular structure determination, Proc. Natl. Acad. Sci. USA 103 (2006) 1756–1761.
[5] W. Rieping, M. Habeck, M. Nilges, Modeling errors in NOE data with a log-normal distribution improves the quality of NMR structures, J. Am. Chem. Soc. 127 (2005) 16026–16027.
[6] C.K. Fisher, A. Huang, C.M. Stultz, Modeling intrinsically disordered proteins with bayesian statistics, J. Am. Chem. Soc. 132 (2010) 14919–14927.
[7] C. Schwieters, J. Kuszewski, N. Tjandra, G. Clore, The Xplor-NIH NMR molecular structure determination package, J. Magn. Reson. 160 (2003) 66–74.
[8] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. Malliavin, M. Nilges, Aria2: automated NOE assignment and data integration in NMR structure calculation, Bioinformatics 23 (2007) 381–382.
[9] M. Habeck, Statistical mechanics analysis of sparse data, J. Struct. Biol. (2010).
[10] T. Hamelryck, J.T. Kent, A. Krogh, Sampling realistic protein conformations using local structural bias, PLoS Comput. Biol. 2 (2006) e131.
[11] W. Boomsma, K.V. Mardia, C.C. Taylor, J. Ferkinghoff-Borg, A. Krogh, T. Hamelryck, A generative, probabilistic model of local protein structure, Proc. Natl. Acad. Sci. USA 105 (2008) 8932–8937.
[12] T. Harder, W. Boomsma, M. Paluszewski, J. Frellsen, K.E. Johansson, T. Hamelryck, Beyond rotamers: a generative, probabilistic model of side chains in proteins, BMC Bioinf. 11 (2010) 306.
[13] K.T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, J. Mol. Biol. 268 (1997) 209–225.
[14] S.C. Lovell, J.M. Word, J.S. Richardson, D.C. Richardson, The penultimate rotamer library, Proteins 40 (2000) 389–408.
[15] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases, Prot. Sci. 5 (1996) 1067–1080.
[16] G.L. Butterfoss, B. Kuhlman, Computer-based design of novel protein structures, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 49–65.
[17] R.A. Engh, R. Huber, Accurate bond and angle parameters for X-ray protein structure refinement, Acta Cryst. A47 (1991) 392–400.
[18] L.M. Rice, A.T. Brünger, Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement, Proteins 19 (1994) 277–290.
[19] S. Bottaro, W. Boomsma, K.E. Johansson, C. Andreetta, T. Hamelryck, J. Ferkinghoff-Borg, Realizing the potential of Monte Carlo: subtle kinetics in dense protein systems, unpublished results.
[20] B. Hesselbo, R.B. Stinchcombe, Monte Carlo simulation and global optimization without parameters, Phys. Rev. Lett. 74 (1995) 2151–2155.
[21] J. Ferkinghoff-Borg, Optimized Monte Carlo analysis for generalized ensembles, Eur. Phys. J. B 29 (2002) 481–484.
[22] M. Habeck, M. Nilges, W. Rieping, Replica-exchange Monte Carlo scheme for Bayesian data analysis, Phys. Rev. Lett. 94 (2005) 018105-1–018105-4.
[23] S. Meyer, Data Analysis for Scientists and Engineers, Wiley, 1975.
[24] G. Favrin, A. Irbäck, F. Sjunnesson, Monte Carlo update for chain molecules: biased Gaussian steps in torsional space, J. Chem. Phys. 114 (2001) 8154–8158.
[25] L. Willard, A. Ranjan, H. Zhang, H. Monzavi, R.F. Boyko, B.D. Sykes, D.S. Wishart, Vadar: a web server for quantitative evaluation of protein structure quality, Nucleic Acids Res. 31 (2003) 3316–3319.
[26] G. Vriend, WHAT IF: a molecular modeling and drug design program, J. Mol. Graph. 8 (1990). 52–56, 29.
[27] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, Procheck: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[28] J.W. Neidigh, R.M. Fesinmeyer, N.H. Andersen, Designing a 20-residue protein, Nat. Struct. Biol. 9 (2002) 425–430.
[29] D. Zhao, O. Jardetzky, An assessment of the precision and accuracy of protein structures determined by NMR: dependence on distance errors, J. Mol. Biol. 239 (1994) 601–607.
[30] C.A. Spronk, S.B. Nabuurs, A.M. Bonvin, E. Krieger, G.W. Vuister, G. Vriend, The precision of NMR structure ensembles revisited, J. Biomol. NMR 25 (2003) 225–234. 10.1023/A:1022819716110.