

Materials and Methods

Error distribution

We use the lognormal distribution

$$p(I_i|X, \gamma, \sigma, I) = \frac{1}{\sqrt{2\pi\sigma^2} I_i} \exp \left\{ -\frac{1}{2\sigma^2} \log^2(I_i/\gamma d_i^{-6}(X)) \right\} \quad (S1)$$

to model deviations of the observed intensities I_i from those predicted by the ISPA (*S1*). The intensities are subject to noise of unknown magnitude σ , arising from various sources (data acquisition, processing, shortcomings in theoretical models). The lognormal distribution is defined for non-negative values and represents a natural choice for modeling positive data such as intensities. We assume logical independence of the observation of NOESY (*S2*) cross-peaks, thus the total likelihood function is $p(D|X, \gamma, \sigma, I) = \prod_i p(I_i|X, \gamma, \sigma, I)$.

Conformational prior density

In order to describe our prior knowledge about the conformations of the SH3 domain, we use an empirical force field E_{phys} that encodes physical interactions within the molecule. Solvent interactions are difficult to describe and will be neglected here. At room temperature, bond lengths, bond angles and ring planarities show little variance. In good approximation we keep these parameters fixed and use the ECEPP/2 force field (*S3, S4*) to describe the covalent geometry of the molecule. In this case, 275 torsion angles are the only degrees of freedom (*S5*) and parametrize the polypeptide chain. A repulsive potential describes non-bonded forces acting between atom k and l and we obtain

$$E_{\text{phys}}(X) = \frac{1}{2} \sum_{k < l} k_{kl} \left\{ \begin{array}{ll} [d_{kl} - d_{kl}(X)]^4, & d_{kl}(X) < d_{kl} \\ 0, & d_{kl}(X) \geq d_{kl} \end{array} \right\}, \quad (S2)$$

where the sum runs over all atoms. Values for the force constants and minimum distances k_{kl} and d_{kl} , respectively, were taken from the PROLSQ (*S6*) X-ray refinement program. Assuming

that experiments are carried out at a constant temperature T and following the principle of Maximum Entropy (S7), our conformational prior density is the canonical ensemble

$$p(X|I) = \frac{1}{Z(T)} \exp \{ -E_{\text{phys}}(X)/(k_B T) \} \quad (\text{S3})$$

where k_B denotes Boltzmann's constant.

Posterior density

We use Jeffreys' prior (S8) for the two nuisance parameters γ and σ . According to Bayes' theorem the joint posterior for all unknown parameters is

$$\begin{aligned} p(X, \gamma, \sigma | D, I) &\propto \sigma^{-(n+1)} \gamma^{-1} \exp \{ -E_{\text{phys}}(X)/(k_B T) \} \\ &\times \exp \left\{ -\frac{1}{2\sigma^2} \sum_i \log^2 \left[\gamma d_i^{-6}(X)/I_i \right] \right\}. \end{aligned} \quad (\text{S4})$$

Posterior sampling

A Markov Chain Monte Carlo algorithm (S9) is used to simulate the posterior distribution. A stochastic process generates random samples $(X^{(k)}, \gamma^{(k)}, \sigma^{(k)})$ from $p(X, \gamma, \sigma | D, I)$. We use Gibbs sampling (S10) to decompose the simulation into three steps which are then iterated:

$$\begin{aligned} \gamma^{(k+1)} &\sim p(\gamma | X^{(k)}, \sigma^{(k)}, D, I), \\ \sigma^{(k+1)} &\sim p(\sigma | X^{(k)}, \gamma^{(k+1)}, D, I), \\ X^{(k+1)} &\sim p(X | \gamma^{(k+1)}, \sigma^{(k+1)}, D, I), \end{aligned} \quad (\text{S5})$$

where “ \sim ” means “drawn from”. The conditional posterior densities are obtained by fixing the parameters listed right of the conditioning stroke in the joint posterior density (S4). Thus,

$$p(\gamma | X, \sigma, D, I) = \sqrt{\frac{n}{2\pi\sigma^2}} \exp \left\{ -\frac{n}{2\sigma^2} \left[\log \gamma - \sum_i \log(d_i^6(X)I_i) \right]^2 \right\} \quad (\text{S6})$$

$$p(\sigma^{-2} | X, \gamma, D, I) = \frac{(\chi^2(X, \gamma)/2)^{n/2}}{\Gamma(n/2)} (\sigma^{-2})^{(n/2-1)} \exp \{ -\chi^2(X, \gamma)/2\sigma^2 \} \quad (\text{S7})$$

$$p(X | \gamma, \sigma, D, I) \propto \exp \{ -\chi^2(X, \gamma)/2\sigma^2 - \beta E_{\text{phys}}(X) \} \quad (\text{S8})$$

with $\chi^2(X, \gamma) = \sum_i \log^2(\gamma d_i^{-6}(X)/I_i)$ and $\Gamma(\cdot)$ being the gamma function. The conditional posterior density for γ (Eq. (S6)) is a lognormal distribution. The conditional posterior density for the inverse squared error σ^{-2} (Eq. (S7)) is a gamma distribution. Thus we can update the nuisance parameters using random number generators for the lognormal and the gamma distribution, respectively.

The conditional conformational posterior density $p(X|\gamma, \sigma, D, I)$ (Eq. (S8)) is very complex; we use the Hybrid Monte Carlo (HMC) method (S11) to obtain stochastic samples from this density. HMC is an efficient method to investigate multidimensional correlated distributions. The negative logarithm of the conditional conformational posterior density is analogous to a physical energy and defines a dynamical system in conformational space. Numerical integration of Hamilton's equations of motion with the leapfrog discretization scheme (S12) yields a candidate conformation, which is accepted according to the Metropolis criterion (S13). The number of integration steps for an HMC update was 250.

The Gibbs sampling procedure (Eq. (S5)) is embedded in a Replica-exchange Monte Carlo scheme (S14) to cope with trapping in modes of the posterior distribution. The details of the algorithm are described in (S15). Two parameters control the shape of the posterior distribution: exponential weighting of the likelihood function switches the data gradually off; use of the Tsallis ensemble (S16), an extension of the Boltzmann ensemble, allows one to turn off non-bonded interactions. Multiple copies of the system are simulated at different values of the two replica parameters. After 30 Gibbs sampling steps for each copy, parameter exchanges between neighboring replicas are accepted according to the Metropolis criterion.

We simulated a replica arrangement consisting of 50 systems and calculated a total of 10000 conformations per copy. The calculation time on a 50 processor PC cluster was approximately 72 hours.

References and Notes

- S1. D. Neuhaus, M. P. Williamson, *The nuclear Overhauser effect in structural and conformational analysis*. (VCH Publishers Inc., New York, 1989).
- S2. S. Macura, R. R. Ernst, *Molecular Physics* **41**, 95 (1980).
- S3. F. A. Momany, R. F. McGuire, A. W. Burgess, H. A. Scheraga, *J. Phys. Chem.* **79**, 2361 (1975).
- S4. G. Nemethy, M. A. Pottle, H. A. Scheraga, *J. Phys. Chem.* **87**, 1883 (1983).
- S5. P. J. Flory, *Statistical Mechanics of Chain Molecules*. (Carl Hanser Verlag, Munich, 1969).
- S6. W. A. Hendrickson, *Methods in Enzymology* **115**, 252 (1985).
- S7. E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).
- S8. H. Jeffreys, *Proc. Roy. Soc. A* **186**, 453 (1946).
- S9. M. H. Chen, Q. M. Shao, J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation* (Springer Verlag, Inc., New York, 2002).
- S10. S. Geman, D. Geman, *IEEE Trans. PAMI* **6**, 721 (1984).
- S11. S. Duane, A. D. Kennedy, B. Pendleton, D. Roweth, *Phys. Lett. B* **195**, 216 (1987).
- S12. M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*. (Clarendon Press, Oxford, 1987).
- S13. N. Metropolis, M. Rosenbluth, A. Rosenbluth, A. Teller, E. Teller, *J. Chem. Phys.* **21**, 1087 (1957).

- S14. R. H. Swendsen, J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- S15. M. Habeck, M. Nilges, W. Rieping, *Phys. Rev. Lett.* **94**, 0181051 (2005).
- S16. C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988).

Fig. S1. Replica-exchange Monte Carlo algorithm. We embed the Gibbs sampling scheme (Eq. (S5)) in a Replica-exchange Monte Carlo algorithm (S14) which simulates a sequence of “heated” copies of the system. We introduce two temperature-like parameters λ and q (confer (S15) for details). λ is a weighting factor that is used to control the contribution of the data in the likelihood function. For $\lambda = 1$ the data are switched on, for $\lambda = 0$ the data are switched off. The second parameter q is the Tsallis parameter (S16) controlling the shape of the conformational prior density. For $q = 1$, the conformational prior density is identical to the Boltzmann distribution. For $q \rightarrow \infty$, physical interactions are switched off and the conformational prior density approaches a flat distribution over conformational space. By setting q to a large value, the polypeptide chain can move almost freely during HMC sampling; for our system already values $q_{\max} \approx 1.1$ are sufficiently large. We arrange the copies of the system in such a way that first the data are switched off by decreasing λ from 1.0 to $\lambda_{\min} = 0.1$ (while letting $q = 1$). In the other half of the arrangement also the physical interactions are swichted off by increasing q from 1.0 to 1.1 (while letting $\lambda = \lambda_{\min} = 0.1$, i.e. data are almost completely neglected).

Figure S1

