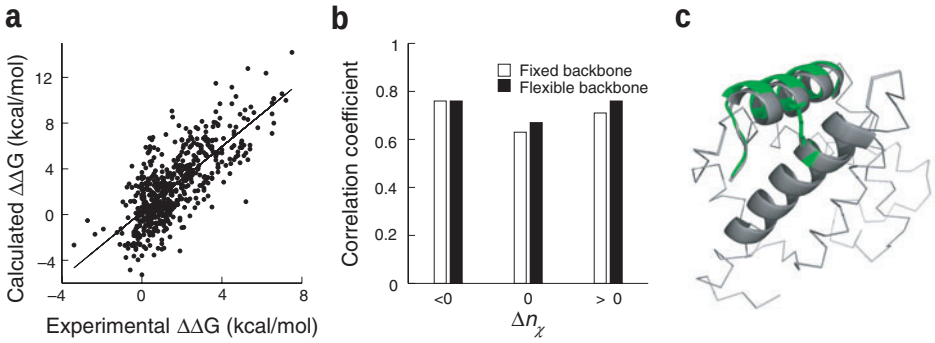

Figure 1 | Performance of Eris. (**a**) Scatter plot of $\Delta\Delta G$ calculations using Eris. The $\Delta\Delta G$ of 595 mutants were calculated and compared with experimental measurements. The Pearson correlation coefficient was 0.75 ($P = \sim 10^{-108}$) and the r.m.s. deviation between the experimental and computed $\Delta\Delta G$ values was 2.4 kcal/mol. The solid line corresponds to linear regression fit to the data points. (**b**) Correlation coefficients between the calculated and experimental $\Delta\Delta G$s for three different classes of mutations based on the change in the number of side-chain χ angles (Δn_χ). The mutations with $\Delta n_\chi < 0$ are associated with large-to-small mutations and those with $\Delta n_\chi \geq 0$ correspond to mutation to residues of the same or larger sizes. The flexible- and fixed-backbone methods have the same prediction accuracy for $\Delta n_\chi < 0$. However, the flexible-backbone $\Delta\Delta G$ prediction correlates better with experiments for $\Delta n_\chi \geq 0$ cases, owing to its ability to resolve possible side-chain clashes. (**c**) The backbone structures of wild-type and A130K mutant apomyoglobin proteins. The mutant structure is obtained from a flexible-backbone calculation. The N-terminal helix of the A130K apomyoglobin bends ~0.2 Å outward to accommodate the larger lysine side chain (green).

calculations are computationally costly (that is, efficiency is low). Modern large-scale $\Delta\Delta G$ prediction methods use heuristic algorithms with effective force fields and empirical parameters to estimate the stability changes caused by mutations in agreement with experimental data[2–5]. There are, however, two considerable drawbacks pertinent to the heuristic methods. First, most of these prediction methods rely on parameter training using available experimental $\Delta\Delta G$ data. Such training is usually biased toward mutations that feature large-to-small residue substitutions, such as alanine-scanning experiments (that is, poor transferability). Second, protein backbone flexibility, which is crucial for resolving atomic clashes and backbone strains in mutant proteins, is not considered in these methods, thereby reducing accuracy and limiting the application of heuristic methods (that is, limited applicability).

To address the issues of efficiency, transferability and applicability, we developed the Eris method, which uses a physical force field with atomic modeling as well as fast side-chain packing and backbone relaxation algorithms. The free energy is expressed as a weighted sum of van der Waals forces, solvation, hydrogen bonding and backbone-dependent statistical energies[6] (**Supplementary Methods** online). The weighting parameters are independently trained to recapitulate the native amino acid sequences for 34 proteins using high-resolution X-ray structures[6]. Additionally, an integral step of Eris is backbone relaxation when severe atom clashes or backbone strains are detected during calculation.

We tested Eris on 595 mutants from five proteins, for which the $\Delta\Delta G$ values were documented (**Fig. 1a**). We found significant agreement between the predicted and measured $\Delta\Delta G$ values with a correlation coefficient of 0.75 ($P = 2 \times 10^{-108}$). The correlation between the predictions and experiments is comparable to that reported using other methods[2–5]. Unlike previous methods, Eris also has high predictive power for small-to-large[3] side-chain-size mutations (**Fig. 1b**,**c**), owing to its ability to effectively relax backbone structures and resolve clashes introduced by mutations. As a direct comparison with other methods, we computed the stability changes of the small-to-large mutations using Eris and other web-based stability prediction servers. We found that Eris outperformed other available servers (**Supplementary Discussion** and **Supplementary Tables 1** and **2** online). Additionally, Eris features a protein structure pre-relaxation option, which remarkably improves the prediction accuracy when a high-resolution protein structure is not available (**Supplementary Discussion** and **Supplementary Fig. 1** online).

Our test validates the unbiased force field, side-chain packing and backbone relaxation algorithms in Eris. We anticipate Eris will be applicable to examining a much larger variety of mutations during protein engineering. We built a web-based Eris server for $\Delta\Delta G$ estimation. The server is freely accessible online (http://eris.dokhlab.org).

Note: Supplementary information is available on the Nature Methods website.

COMPETING INTERESTS STATEMENT
The authors declare no competing financial interests.

Shuangye Yin, Feng Ding & Nikolay V Dokholyan

Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.
e-mail: dokh@med.unc.edu

1. Kollman, P. *Chem. Rev.* **93**, 2395–2417 (1993).
2. Gilis, D. & Rooman, M. *J. Mol. Biol.* **272**, 276–290 (1997).
3. Guerois, R., Nielsen, J.E. & Serrano, L. *J. Mol. Biol.* **320**, 369–387 (2002).
4. Bordner, A.J. & Abagyan, R.A. *Proteins* **57**, 400–413 (2004).
5. Saraboji, K. *et al. Biopolymers* **82**, 80–92 (2006).
6. Ding, F. & Dokholyan, N.V. *PLoS Comput. Biol.* **2**, e85 (2006).

An automated tool for maximum entropy reconstruction of biomolecular NMR spectra

To the editor: High resolution is essential for successful application of NMR spectroscopy to biomolecules, but involves a classic 'catch-22'. High magnetic fields increase chemical shift dispersion, thus increasing resolution and reducing spectral overlap, but the required increase in sampling rate (to avoid aliasing) means longer acquisition times in the indirect dimensions of multidimensional experiments (indirect dimensions are sampled by iteration, whereas the lone 'direct' dimension is sampled in real time). Consequently the potential resolution afforded by high magnetic fields is rarely realized in the indirect dimensions. There is a growing realization that this is a consequence of

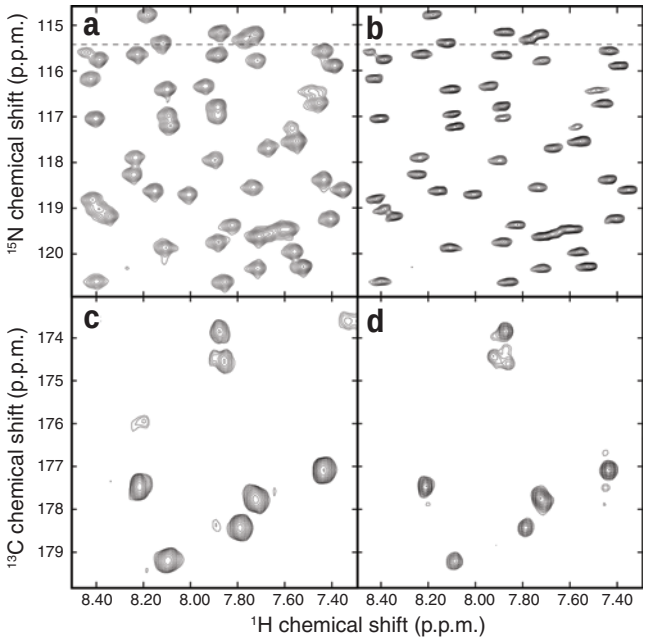

Figure 1 | Automated MaxEnt reconstruction can dramatically reduce data collection time or improve resolution. (**a**–**d**) ^{15}N heteronuclear single quantum correlation (HSQC; **a**,**b**) and HNCO (**c**,**d**) spectra are shown for DNA polymerase X obtained using conventional processing (linear-prediction extrapolation and sinebell apodization; **a**,**c**), and automated MaxEnt reconstruction using linewidth deconvolution to improve resolution without the sensitivity losses characteristic of apodization (**b**), and nonuniform sampling to achieve a sevenfold reduction in data acquisition time (**d**). Two-dimensional cross-sections of the three-dimensional spectrum in **c** and **d** correspond to the ^{15}N frequency indicated by the dashed lines in **a** and **b**.

uniform sampling required by conventional Fourier methods of spectrum analysis.

A variety of approaches for overcoming this sampling problem have been introduced, including Bayesian and maximum likelihood (MLM)[1,2], maximum entropy (MaxEnt)[3–6], reduced dimensionality (RD)[7] and G-matrix Fourier transform (GFT)[8], back-projection reconstruction (BPR)[9], multidimensional decomposition (MDD)[10], and nonuniform discrete Fourier transformation[11–13]. Each realizes higher resolution along indirect dimensions by collecting samples at long evolution times without collecting samples at every integer multiple of the sampling interval. The artifacts that can occur with these methods tend to reflect the particular strategy used for nonuniform sampling[14].

MaxEnt reconstruction has several advantages over the other methods. As it makes no assumptions regarding the signals, it is more robust (especially for low S/N) than methods that do make assumptions (such as Bayesian, MLM and MDD). In contrast to GFT and BPR, MaxEnt can use essentially arbitrary nonuniform sampling. MaxEnt can also be used for deconvolution to achieve additional resolution enhancement or virtual decoupling[15]. Efficient algorithms for MaxEnt reconstruction have been developed[16,17], and their properties have been extensively investigated. But despite more than three decades of implementation, the use of MaxEnt has been limited to several expert laboratories, in part because of the need to specify adjustable parameters. These parameters are an estimate of the noise level in the data and a scale factor related to the sensitivity of the spectrometer[18]. The latter is difficult to determine empirically, as it depends on many factors.

Fortunately the results of MaxEnt are not overly sensitive to the value of the scale factor. A useful heuristic is to choose a value larger than the noise but smaller than the weakest peak. We have implemented a web-based script generator (**Supplementary Fig. 1** online) that implements the heuristic via the Rowland NMR Toolkit[15] to automatically determine the noise level and appropriately set the MaxEnt reconstruction parameters. The results of automatic MaxEnt reconstruction of two- and three-dimensional spectra of a 20 kDa protein, DNA polymerase X (ref. 19), as compared to conventional processing, are illustrated in **Figure 1**. This automated procedure for MaxEnt reconstruction should make the method accessible to a much broader cross-section of the biomolecular NMR community.

The script generator and the Rowland NMR Toolkit are available to academic and nonprofit organizations without charge (http://sbtools.uchc.edu/nmr/).

Note: Supplementary information is available on the Nature Methods website.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods.

Mehdi Mobli[1–3], Mark W Maciejewski[1,3], Michael R Gryk[1] & Jeffrey C Hoch[1]

[1]University of Connecticut Health Center, 263 Farmington Ave., Farmington, Connecticut 06030, USA. [2]Present address: Manchester Interdisciplinary Biocenter, University of Manchester, 131 Princess St., Manchester M1 7ND, UK. [3]These authors contributed equally to this work.
e-mail: hoch@uchc.edu

1. Bretthorst, G.L. in *Maximum Entropy and Bayesian Methods in Science and Engineering* (eds. Rychert, J., Erickson, G. & Smith, C.R.) 1–28 (American Institute of Physics, Melville, New York, 2001).
2. Chylla, R.A., Volkman, B.F. & Markley, J.L. *J. Biomol. NMR* **12**, 277–297 (1998).
3. Aggarwal, K. & Delsuc, M.A. *Magn. Reson. Chem.* **35**, 593–596 (1997).
4. Barna, J.C.J., Laue, E.D., Mayger, M.R., Skilling, J. & Worrall, S.J.P. *J. Magn. Reson.* **73**, 69 (1987).
5. Jones, J.A., Hodgkinson, P., Barker, A.L. & Hore, P.J. *J. Magn. Reson.* **B113**, 25–34 (1996).
6. Schmieder, P., Stern, A.S., Wagner, G. & Hoch, J.C. *J. Biomol. NMR* **3**, 569–576 (1993).
7. Ding, K. & Gronenborn, A.M. *J. Magn. Reson.* **156**, 262–268 (2002).
8. Kim, S. & Szyperski, T. *J. Am. Chem. Soc.* **125**, 1385–1393 (2003).
9. Kupce, E. & Freeman, R. *J. Am. Chem. Soc.* **126**, 6429–6440 (2004).
10. Jaravine, V., Ibraghimov, I. & Orekhov, V.Y. *Nat. Methods* **3**, 605–607 (2006).
11. Coggins, B.E. & Zhou, P. *J. Magn. Reson.* **182**, 84–95 (2006).
12. Kazimierczuk, K., Kozminski, W. & Zhukov, I. *J. Magn. Reson.* **179**, 323–328 (2006).
13. Marion, D. *J. Biomol. NMR* **36**, 45–54 (2006).
14. Mobli, M., Stern, A.S. & Hoch, J.C. *J. Magn. Reson.* **182**, 96–105 (2006).
15. Hoch, J.C. & Stern, A.S. *NMR Data Processing* (Wiley-Liss, New York, 1996).
16. Skilling, J. & Bryan, R. *Mon. Not. R. Astron. Soc.* **211**, 111–124 (1984).
17. Schmieder, P., Stern, A.S., Wagner, G. & Hoch, J.C. *J. Magn. Reson.* **125**, 332–339 (1997).
18. Daniell, G.J. & Hore, P.J. *J. Magn. Reson.* **84**, 515–536 (1989).
19. Maciejewski, M.W. *et al. Nat. Struct. Biol.* **8**, 936–941 (2001).