Predicting free energy changes using structural ensembles

To the Editor: Reliable and fast computation of protein free energy is crucial for protein-structure analysis, structure-based protein design and protein docking. Rigorous treatments based on physical effective energy functions involve computationally expensive methods such as free energy perturbation, which are time–consuming and are thus incompatible with the need to perform extensive scans. Commonly used fast methods, in turn, involve empirically derived scoring functions and usually do not include protein flexibility or are based on statistical potentials and are therefore highly dependent on the availability of case-dependent experimental training data. Hence, such methods are inherently limited in accuracy and applicability.

Here we propose a computational, structure-based method named Concoord/Poisson-Boltzmann surface area (CC/PBSA) for both fast and quantitative estimation of the folding free energy of mutants, that is, for measuring their conformational stability and for predicting the effect of mutations on protein-protein binding affinity. The first step is to rapidly generate alternative protein conformations via the program Concoord, which efficiently samples the available configurational space[1]. The crystal or nuclear magnetic resonance (NMR) input structure is translated into a geometric description of the complex, and starting from random coordinates, 300–600 structures both of the mutant and the wild type are generated by iteratively correcting the coordinates until all geometric constraints are fulfilled. Then an energy function based on physical chemistry (force field) and an efficient continuum solvent approach, the solution of the Poisson-Boltzmann equation and a term for nonpolar solvation[2], is averaged over the generated structural ensembles (**Supplementary Methods** online). The free energy is approximated by

$$\Delta G_{CC/PBSA} = \Delta G_{electrostatic} + \Delta G_{van\ der\ Waals} + \Delta G_{entropy}$$

By weighting the individual averaged energy contributions (separately for folding free energies and protein-protein binding affinities) water contributions are implicitly taken into account.

We computed free energy differences for folding free energies and binding affinities according to the respective thermodynamic cycle (**Fig. 1a**). We obtained the weighting factors by fitting to experimental data, applying fivefold cross-validation. The correlation we achieved for the folding free energies of 582 mutants of 7 proteins (**Supplementary Tables 1** and **2** online) was 0.75 (s.d. (σ) = 1.04 kcal mol^{-1}; **Fig. 1b**), comparable to FoldX[3] (R = 0.73, σ = 1.02 kcal mol^{-1}) and improved with respect to the recently developed Eris method[4] using trained parameters (R = 0.75, σ = 2.6 kcal mol^{-1}). CC/PBSA

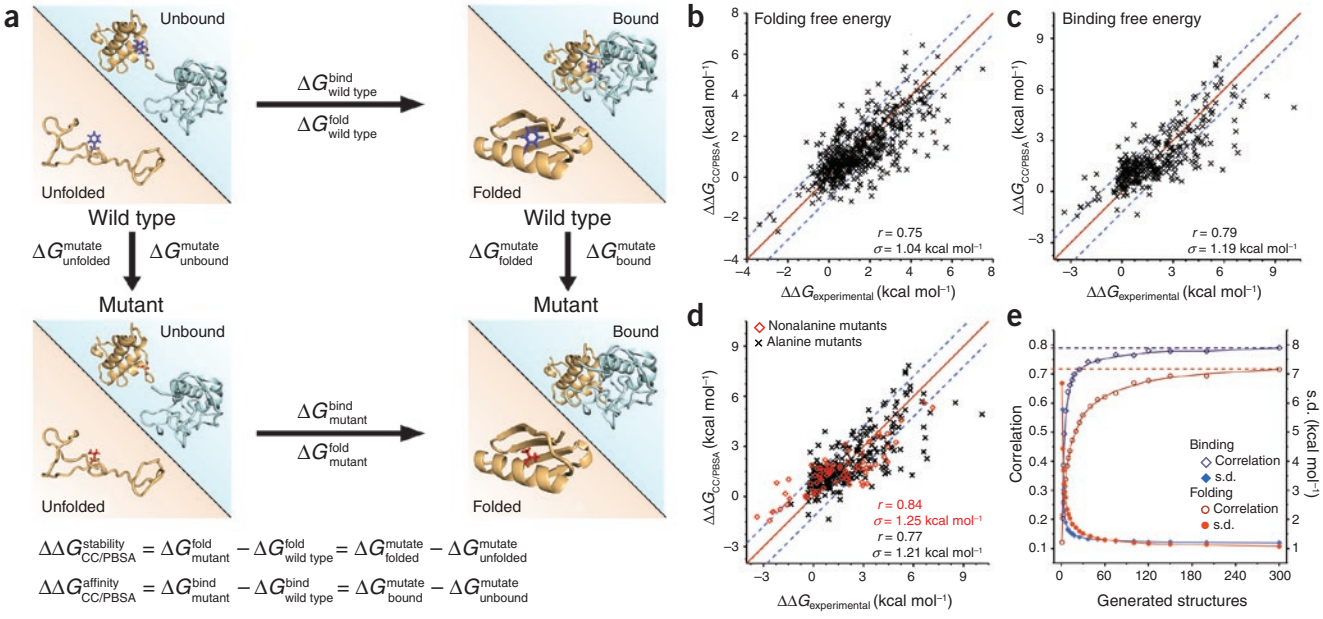

Figure 1 | Prediction of mutational free energy changes using ensembles of structures. (**a**) Thermodynamic cycle for the computation of folding free energies and binding free energies. (**b,c**) Computed values for the effect of mutations on the folding free energies (**b**) and on the binding free energies (**c**) versus experimental values (**Supplementary Tables 1–4**). Correlations between predicted and experimental values excluding outliers (>2 σ, 6% of the dataset) are R = 0.83 (σ = 0.81 kcal mol^{-1}) for the protein stability and R = 0.85 (σ = 0.94 kcal mol^{-1}) for the protein-protein binding affinity. (**d**) Computed mutational changes in protein-protein binding free energies separately for mutations to alanine and to non-alanine amino acids versus experimental values. The red line in **b**–**d** corresponds to ideal prediction, and the blue dashed lines mark a 1σ environment. (**e**) Correlation as a function of sampled structures; almost all mutations were large to small mutations.

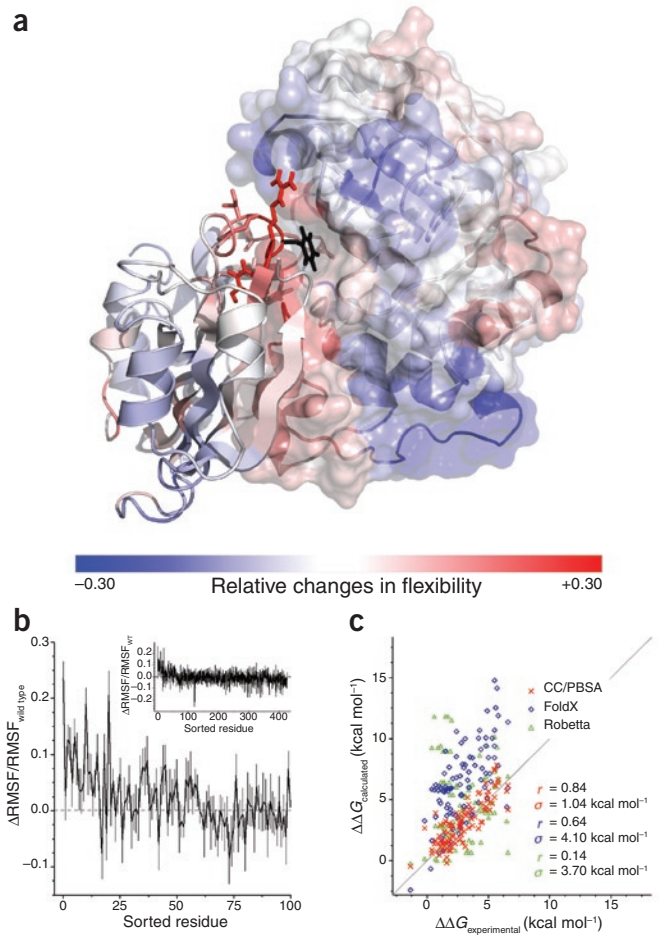

Figure 2 | Effect of alanine mutations on TEM1-BLIP complex. (**a**) Relative changes in flexibility of the TEM1-BLIP complex upon the F142A mutation (black). Side chains with large increase in flexibility are shown in stick representation. (**b**) Relative change in root mean square fluctuations (RMSF) upon the F142A mutation for all residues in the TEM1-BLIP complex, sorted according to their distance to the mutation site. (**c**) Calculated changes in binding free energy for TEM1-BLIP alanine mutants applying FoldX[3], Robetta[5] and CC/PBSA. For the comparison, parameters for CC/PBSA were fitted on the remaining dataset on other protein-protein complexes only. The diagonal line corresponds to ideal prediction.

uses only 4 weighting factors, compared to 5 for FoldX and 20 for Eris. For protein-protein binding affinities (367 mutants of 9 protein-protein complexes; **Supplementary Tables 3** and **4** online), the correlation obtained with CC/PBSA was $R = 0.79$ with $\sigma = 1.19$ kcal mol^{-1} (**Fig. 1c**). We obtained a similar accuracy for both alanine and nonalanine mutants (**Fig. 1d**), whereas we observed a slightly lower accuracy for buried and exposed mutations, considered separately (**Supplementary Fig. 1** online).

Consideration of structural flexibility is crucial to reliably predict free energies: the correlation to experiment increased with the number of Concoord structures considered (**Fig. 1e** and **Supplementary Fig. 2** online). Taking only the (mutated) crystal structure as a basis for the above free energy function, we achieved a correlation of only $R = 0.57$ ($\sigma = 1.37$ kcal mol^{-1}) for the prediction of folding free energies and of $R = 0.67$ ($\sigma = 1.43$ kcal mol^{-1}) for binding affinities.

To compare CC/PBSA to other methods applying empirical effective energy functions (FoldX and Robetta[5]), we analyzed mutations to alanine in the TEM1-BLIP complex; mutations of this complex have been shown to act in a highly cooperative manner[6]. We analyzed the change in flexibility of the BLIP-F142A mutant with respect to the wild type (**Fig. 2a**). The flexibility of residues close to the mutant residue was substantially enhanced (**Fig. 2b**). CC/PBSA yielded a high correlation of $R = 0.84$ ($\sigma = 1.04$ kcal mol^{-1}; **Fig. 2c**) for this dataset; it outperformed both FoldX ($R = 0.64$, $\sigma = 4.10$ kcal mol^{-1}) and Robetta ($R = 0.14$, $\sigma = 3.70$ kcal mol^{-1}).

The main advantage of CC/PBSA is its inclusion of full protein flexibility, which dramatically improves prediction quality for protein-protein binding affinities, similar in accuracy to free energy perturbation, the linear interaction energy method[7] or to molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA)[8]. Although similar to MM/PBSA, the fast sampling of conformational space in CC/PBSA avoids explicit molecular dynamics simulations and therefore gains speed (×100) and scalability without losing accuracy.

CC/PBSA enables full mutant scanning of protein-protein interfaces (for example, the insulin dimer interface; **Supplementary Fig. 3** online) and thus identification of hot spots, the design of protein interaction surfaces and of protein-stabilizing mutations. A web-based CC/PBSA server for estimating mutational free energy changes and for generating structural ensembles is freely accessible online at http://ccpbsa.bioinformatik.uni-saarland.de/.

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

ACKNOWLEDGMENTS

We thank J. Schymkowitz for help with FoldX. A.B., C.M.B., B.L.G. and R.A.B. acknowledge support by the Deutsche Forschungsgemeinschaft (BO 2963/1-1, GR 2079/4-1, BIZ 4/1). Part of the work was supported by the Swiss National Center for Competence in Research in Structural Biology (R.A.B. and A.C.).

Alexander Benedix[1,4], Caroline M Becker[1,4], Bert L de Groot[2], Amedeo Caflisch[3] & Rainer A Böckmann[1]

[1]Theoretical & Computational Membrane Biology, Center for Bioinformatics, Saarland University, P.O. Box 15 11 50, D-66041 Saarbrücken, Germany. [2]Computational Biomolecular Dynamics, Max-Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [3]Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. [4]These authors contributed equally to this work.
e-mail: rainer@bioinformatik.uni-saarland.de

1. de Groot, B.L. *et al. Proteins Struct. Funct. Genet.* **29**, 240–251 (1997).
2. Sitkoff, D., Sharp, K.A. & Honig, B. *J. Phys. Chem.* **98**, 1978–1988 (1994).
3. Guerois, R., Nielsen, J.E. & Serrano, L. *J. Mol. Biol.* **320**, 369–387 (2002).
4. Yin, S., Ding, F. & Dokholyan, N.V. *Nat. Methods* **4**, 466–467 (2007).
5. Kortemme, T. & Baker, D. *Proc. Natl. Acad. Sci. USA* **99**, 14116–14121 (2002).
6. Reichmann, D. *et al. Proc. Natl. Acad. Sci. USA* **102**, 57–62 (2005).
7. Åqvist, J., Medina, C. & Samuelsson, J.E. *Protein Eng.* **7**, 385–391 (1994).
8. Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A. & Case, D.A. *J. Am. Chem. Soc.* **120**, 9401–9409 (1998).