

in many species not yet considered genetically tractable. SHOREmap could also be applied to mapping of quantitative trait loci, large deletions and recessive lethal or dominant mutations (Supplementary Note).

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

#### ACKNOWLEDGMENTS

We thank J. Carrington, R. Clark and N. Warthmann for useful discussions, and J. Fitz for setting up the <http://1001genomes.org/> website. This work was funded by The Danish Research Council for Technology and Production Sciences (grant 274-06-0096) (J.E.J.), the Danish Strategic Research Council (Nabiit, grant 2106-07-0013) (K.L.N.), The European Union Sixth Framework Programme IP AGRON-OMICS (contract LSHG-CT-2006-037704), a Gottfried Wilhelm Leibniz Award from Deutsche Forschungsgemeinschaft, and the Max Planck Society (D.W.).

Korbinian Schneeberger<sup>1,4</sup>, Stephan Ossowski<sup>1,4</sup>, Christa Lanz<sup>1</sup>, Trine Juul<sup>2</sup>, Annabeth Høgh Petersen<sup>3</sup>, Kåre Lehmann Nielsen<sup>3</sup>, Jan-Elo Jørgensen<sup>2</sup>, Detlef Weigel<sup>1</sup> & Stig Uggerhøj Andersen<sup>2</sup>

<sup>1</sup>Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany. <sup>2</sup>Department of Molecular Biology, University of Aarhus, Aarhus, Denmark. <sup>3</sup>Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Aalborg, Denmark. <sup>4</sup>These authors contributed equally to this work. e-mail: weigel@weigelworld.org or sua@mb.au.dk

- Ossowski, S. *et al. Genome Res.* **18**, 2024–2033 (2008).
- Lister, R., Gregory, B.D. & Ecker, J.R. *Curr. Opin. Plant Biol.* **12**, 107–118 (2009).
- Clark, R.M. *et al. Science* **317**, 338–342 (2007).

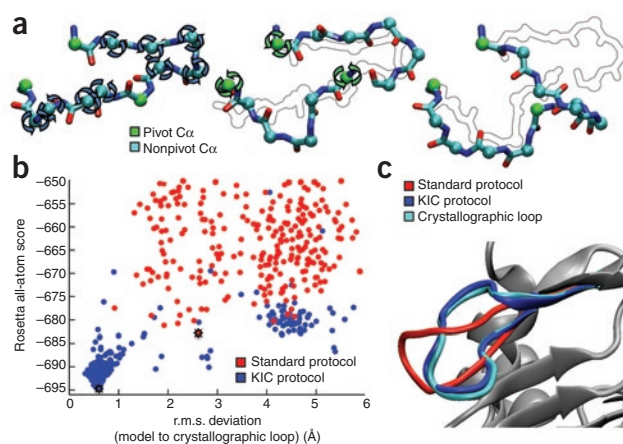
## Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling

To the Editor: Proteins exploit the conformational variability of loop regions to carry out diverse biological tasks including molecular recognition and signal transduction. New algorithms to engineer these functions by combining loop building and sequence design therefore have enormous practical applications but require high-resolution ‘loop reconstruction’: the modeling of protein loop conformations, given the amino acid sequence. Loop reconstruction in protein design may be simplified conceptually by restricting changes to the functional loop regions. However, despite progress in loop prediction methods<sup>1,2</sup>, design applications are limited by the difficulty in modeling purely local conformational moves and by the need for advances in sampling and evaluating loop conformations.

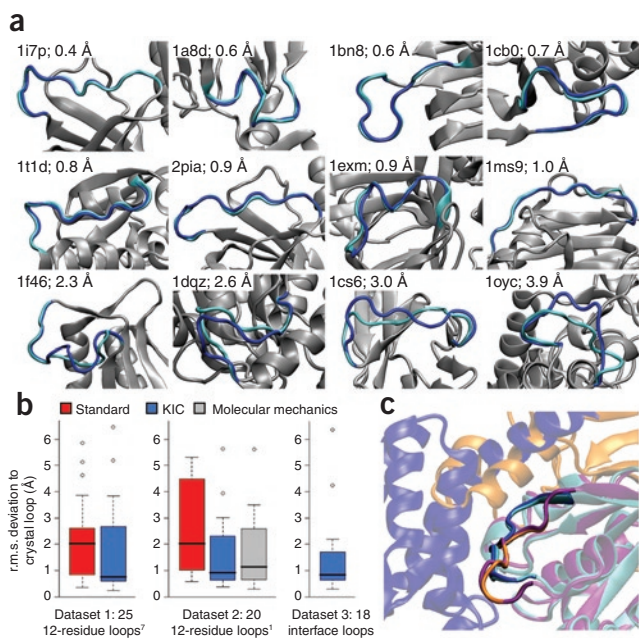
Here we address these challenges with a robotics-inspired local loop reconstruction method for peptide chains, called kinematic closure (KIC). Calculating the accessible conformations of objects subject to constraints, such as determining the possible positions of the interior joints of a robot arm given fixed positions for the shoulder and fingertips, has been well-studied in inverse kinematics, a subfield of robotics. Building on the first<sup>3</sup> and subsequent applications (Supplementary Methods) of kinematics to protein modeling, the KIC method presented here analytically determines all mechanically accessible conformations for 6 torsions of a peptide chain of any length, while simultaneously sampling the remaining torsions and N-C $\alpha$ -C bond angles using polynomial resultants<sup>4</sup> (Fig. 1a, Supplementary Methods and Supplementary Fig. 1). To enable a range of applications, we coupled KIC to the Rosetta

method for protein structure modeling<sup>5</sup>. Our loop reconstruction protocol iterates KIC calculations as Monte Carlo moves first with loop backbone minimization in a low-resolution stage, in which side-chains are represented as centroids, and then in a high-resolution all-atom stage with minimization of the loop backbone and all side chains in the loop environment (Supplementary Fig. 2 and Supplementary Methods). At the beginning of each KIC simulation, we discard all native loop bond lengths, bond angles and torsions. In addition, we perform reconstructions without knowledge of native side-chain conformations in both the loop and the protein scaffold (Supplementary Methods), which makes prediction substantially more challenging but broadens the range of applications to designing new loop conformations that may interact differently with neighboring side chains.

We found that KIC substantially improves model accuracy over the standard loop building method in Rosetta, which combines insertion of torsion segments from homologous proteins and a numerical closure technique<sup>6</sup>. We generated 1,000 models by the KIC method and compared its performance to the standard Rosetta method with the same number of Monte Carlo steps on twenty-five 12-residue protein loops (dataset 1; ref. 7). For each protein, we computed the root mean squared (r.m.s.) deviation of the backbone atoms of the best scoring loop model to the crystallographic loop, after superimposing the non-loop regions of the model onto the crystal structure. The KIC protocol frequently sampled regions of conformational space that were <1.0 Å from the crystallographic loop, which were not sampled by the standard Rosetta method (Fig. 1b). In the majority of cases (15/25), the best-scoring models were very close to the crystallographic loop conformation (Fig. 1b,c).



**Figure 1** | Loop reconstruction with KIC. (a) In the KIC move, 3 C $\alpha$  atoms of an  $N$ -residue chain are designated as pivots (green spheres); the remaining  $N - 3$  are non-pivot C $\alpha$  atoms (cyan spheres; left). In a 12-residue loop, 24 torsions are modeled. Nonpivot torsions are sampled from a residue type-specific Ramachandran map, opening the chain (middle). KIC then finds all values for the pivot torsions that close the loop, if any exist, keeping the endpoints fixed (right). The previous state is shown in outline. (b) Performance of the Rosetta KIC protocol and standard protocols on a 12-residue loop (Protein Data Bank (PDB): 1spr). Only KIC densely sampled regions < 1.0 Å r.m.s. deviation from the crystallographic loop. Asterisks mark the lowest-scoring reconstructions from the two methods. The Rosetta all-atom score includes the enthalpy plus the solvation contribution to the entropy but not the configurational entropy. (c) The lowest scoring reconstructions from b are shown. KIC improved reconstruction accuracy to 0.6 Å from 2.6 Å using the standard protocol.



**Figure 2** | Performance of the KIC loop reconstruction protocol. **(a)** Representative set of 12-residue loop reconstructions (blue) on dataset 2. PDB identifiers and r.m.s. deviation to the crystallographic loop (cyan) are shown. **(b)** Box-plot comparison of the standard Rosetta and KIC Rosetta protocols on dataset 1 (left), both Rosetta protocols with the molecular mechanics method on dataset 2 (middle), and the KIC Rosetta protocol on dataset 3 (right). Boxes span the interquartile range (IQR, 25th–75th percentiles), black lines represent the median, whiskers extend to furthest values within 0.8 times the IQR, and open circles are outliers. **(c)** KIC reconstruction of conformational changes in the Rac switch I loop when bound to ExoS toxin (blue reconstruction on cyan crystal structure, blue partner; PDB 1he1) or Rho guanine dissociation inhibitor (orange reconstruction on purple crystal structure, orange partner; PDB 1hh4).

Over the entire 25-loop set, KIC improved the median accuracy to 0.8 Å r.m.s. deviation from 2.0 Å r.m.s. deviation when we applied the standard Rosetta method (Supplementary Table 1). As both methods use the same scoring function, these results suggest that KIC increased accuracy by improved conformational sampling (although sampling and scoring errors cannot be considered entirely independently as scoring guides the simulation trajectories; see Supplementary Discussion, Supplementary Tables 1–8, and Supplementary Figs. 3,4 for additional analysis of method performance and error sources). The standard method required ~280 central processing unit hours per protein, and KIC required ~320.

To compare KIC loop reconstruction directly to the state-of-the-art molecular mechanics method<sup>1</sup>, we applied the Rosetta KIC

method and standard Rosetta method to the same twenty 12-residue starting structures with perturbed loops and side-chain environments used to assess the molecular mechanics method (dataset 2 (ref 1); Fig. 2a). The Rosetta KIC protocol improved the median accuracy to 0.9 Å from 1.2 Å using the molecular mechanics method, and from 2.0 Å using the standard Rosetta method (Fig. 2b and Supplementary Tables 2 and 5).

Functional loops in signaling proteins in complex with their partners exhibit conformational plasticity against a relatively structured core. To assess the ability of KIC to model such regions, we applied the method to interface loops from 4 proteins crystallized with 18 different partners (dataset 3; Supplementary Methods). KIC reconstructed the loops to 0.8 Å median r.m.s. deviation (Fig. 2b). Notably, the KIC protocol produced high-accuracy reconstructions of the same switch protein loop adopting different conformations when bound to different partners (Fig. 2c, Supplementary Table 3 and Supplementary Discussion). This result highlights the potential of KIC for modeling functional conformational changes. Sub-angstrom loop reconstructions by the local robotics-inspired sampling protocol described here could be coupled with the Rosetta design method<sup>5</sup> to model and engineer protein loops precisely matching a particular binding partner, creating highly selective protein interfaces.

The described state-of-the-art loop reconstruction method is available free of charge as a module of the academic release version 3.1 of the Rosetta program for protein modeling and design at <http://www.rosettacommons.org/>.

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

#### ACKNOWLEDGMENTS

We thank D. Baker and A. Sali for valuable comments, and B. Sellers and M. Jacobson for sharing data and for helpful discussions. This work was supported by grants from the US National Institutes of Health to T.K. (PN2-EY016525) and E.A.C. (R01-GM08171), the University of California Lab Research Program (T.K.), and a PhRMA Foundation Predoctoral Fellowship (D.J.M.).

**Daniel J Mandell<sup>1,2</sup>, Evangelos A Coutsias<sup>3</sup> & Tanja Kortemme<sup>1,2,4</sup>**

<sup>1</sup>Graduate Program in Bioinformatics and Computational Biology, and <sup>2</sup>California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, California, USA. <sup>3</sup>Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, USA. <sup>4</sup>Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, USA.  
e-mail: kortemme@cgl.ucsf.edu

1. Sellers, B.D. *et al. Proteins* **72**, 959–971 (2008).
2. Felts, A.K. *et al. J. Chem. Theory Comput.* **4**, 855–868 (2008).
3. Go, N. & Scheraga, H.A. *Macromolecules* **3**, 178–187 (1970).
4. Coutsias, E.A. *et al. Int. J. Quantum Chem.* **106**, 176–189 (2006).
5. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. *Science* **310**, 638–642 (2005).
6. Canutescu, A.A. & Dunbrack, R.L. Jr. *Protein Sci.* **12**, 963–972 (2003).
7. Wang, C., Bradley, P. & Baker, D. *J. Mol. Biol.* **373**, 503–519 (2007).