SPRINT: side-chain prediction inference toolbox for multistate protein design

Menachem Fromer¹,* , Chen Yanover² , Amir Harel¹ , Ori Shachar¹ , Yair Weiss¹ and Michal Linial²

¹School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel, ²Program in Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA and ³Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Israel

1 INTRODUCTION

The objective of engineering a protein to perform a particular biological function is called protein design. The design problem is usually restricted to the search for an amino acid sequence that possesses a corresponding function. This paradigm assumes a fixed protein backbone, and the amino acid sequence to be considered is taken from a library of amino acid probabilities. Finally, SPRINT also has a module for protein side-chain prediction and single-state design.

2 METHODS

SPRINT is an open-source C++ software package that uses structural data to design functional protein sequences. The probabilistic inference core of SPRINT is based on the FastInf package (Jaimovich et al., 2010). Protein structures are cast as probabilistic graphical models, and inference is performed using the belief propagation or A* algorithms, and dead-end elimination can be applied as pre-processing. The output can either be a list of amino acid sequences simultaneously compatible with these structures, or probabilistic amino acid profiles compatible with the structures. In addition, higher order (e.g. pairwise) amino acid probabilities can also be predicted. Finally, SPRINT also has a module for protein side-chain prediction and single-state design.

3 RESULTS

SPRINT can be downloaded and installed on the user’s machine. In addition, due to their object-oriented nature, SPRINT modules can be extended to provide user-specific functionalities, such as atomic energy function calculations for input PDB structures.

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
SPRINT builds a graphical model to represent all protein structures in the multisite design by connecting corresponding positions in the structures and requiring that they choose the same amino acid. In Figure 1, SPRINT will design the PPAR interface to be optimal for binding both RXR and another PPAR monomer. Probabilistic inference is performed using either the BP (Fromer et al., 2010) or A* algorithms (Leach and Lemon, 1998), and type-dependent dead-end elimination can be applied as pre-processing for the prediction of low-energy sequences (Yanover et al., 2007). We have previously shown that BP-based approaches outperform other methods in predicting sequences with low energies and computing more accurate sequence profiles (Fromer and Yanover, 2008, 2009); see Figure 1, Supplementary Tables S1–S4, and accompanying references for details.

The user can choose to predict either multiple low-energy sequences suitable for the input protein structures, or to predict amino acid probabilities for each position (or pair of positions). Figure 1 shows both options, with the 10 interface sequences most suited to bind both RXR and PPAR (top), and positional probabilities (bottom) calculated by approximating a statistical evaluation of all possible sequences weighted by the Boltzmann distribution of their side-chain conformational free energies (Fromer and Yanover, 2008). In Fromer et al. (2010), it was found that the sequences predicted to bind both targets better match evolutionary PPAR profiles than those optimized to bind only one target.

To this date, we have successfully used SPRINT for numerous single-state design problems (7 to 92 designed positions, rotamer space 1024 to 10200, Fromer and Yanover, 2009) and hundreds of multisite design problems (20 positions, up to three states with combined rotamer space 10200, Fromer and Shifman, 2009), where in all cases all amino acids were permitted at each design position. Lastly, we note that SPRINT also has a module for protein side-chain placement and single-state design; see web site for details.

ACKNOWLEDGEMENTS

We thank Ariel Jaimovich and Ofer Mesli for help with FastInf.

Funding: Sudarsky Center for Computational Biology (to M.F.).

Conflict of Interest: none declared.

REFERENCES

