Predicting protein 3D structure from evolutionary sequence variation

Thomas A. Hopf
Protein Prediction I, 06/18/2015
Outline

Prologue: Correlated mutations

Local vs. global models for 2D prediction

Application to 3D structure prediction
The structure prediction problem

genotype

ACTGTGCACG
TAATGGGCATC

phenotype
Structure from sequence alone?

Christian Anfinsen, Nobel Prize for Chemistry 1972
Sequence-structure gap is not closing!
A protein
Evolutionary selection leaves residue covariation signature
Folding proteins from evolutionary couplings
Outline

Prologue: Correlated mutations

Local vs. global models for 2D prediction

Application to 3D structure prediction
Try something simple: correlation between two columns

single column frequencies: \( f_i(A_i) \) \( f_j(A_j) \)

column pair frequencies: \( f_{ij}(A_i, A_j) \)

\[ f_{ij}(A_i, A_j) - f_i(A_i)f_j(A_j) \]

To what extent do we see a pair of amino acids more/less often than expected by chance?
Mutual information measures correlation between two columns

\[
MI_{ij} = \sum_{A_i, A_j=1}^{q} f_{ij}(A_i, A_j) \ln \left( \frac{f_{ij}(A_i, A_j)}{f_i(A_i) f_j(A_j)} \right)
\]

- **Sum all possible amino acid combinations**
- **Weight**
- **Deviation from statistical independence**
Bad news: doesn‘t work.

Local model main problem: transitivity

PDB structure residue contacts
residue pairs with 100 highest MI values
Solution: use a global model!

global probability model explains observed correlations by causative pair interactions

Inverse Potts inference (undirected graphical model)

observed correlations, causal, direct coevolution
Probability model connects correlations to direct interactions

\[
P(\sigma) = \frac{1}{Z} \exp \left( \sum_{i=1}^{N} h_i(\sigma_i) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} J_{ij}(\sigma_i, \sigma_j) \right)
\]

observed data (sequences) \[\rightarrow\] approximate maximum likelihood inference \[\rightarrow\] direct pair interactions
From sequences to pair scores

Infer parameters

\[ P(\text{data} | \text{parameters}) = \prod_{\sigma \in \text{alignment}} P(\sigma | \mathbf{h}, \mathbf{J}) \]

Calculate evolutionary couplings

\[ FN_{ij} = \|\mathbf{J}_{ij}\|_2 = \sqrt{\sum_{k=1}^{21} \sum_{l=1}^{21} J_{ij}(k, l)^2} \]

+ some other technical details

References: Marks et al. (2011), Ekeberg et al. (2013)
Most global model pairs are close in 3D

local model (mutual information)

global model (MaxEnt, Marks et al., 2011)
He solved it before everyone else did...

Alan Lapedes
Outline

Prologue: Correlated mutations

Local vs. global models for 2D prediction

Application to 3D structure prediction
What could we predict?

3D structure

protein complexes
A brief reminder
The breakthrough: 15 proteins folded from sequences alone

Marks et al., PLoS ONE (2011)
Also works for membrane proteins!

predicted  experimental

Hopf et al., Cell (2012)
What could we predict?

3D structure

protein complexes
Interacting proteins co-evolve to maintain interaction
Complex interactions from the evolutionary sequence record
Accurate prediction of the ABC transporter MetNI
De novo prediction of unsolved complex interactions
Try folding yourself!

www.evfold.org
References for further reading

Protein 3D Structure Computed from Evolutionary Sequence Variation

Debora S. Marks¹ *, Lucy J. Colwell²*, Robert Sheridan³, Thomas A. Hopf¹, Andrea Pagnani⁴, Riccardo Zecchina⁵, Chris Sander²

¹Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, United States of America, ²MRC Laboratory of Molecular Biology, Hills Road, Cambridge, United Kingdom, ³Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York, United States of America, ⁴Human Genetics Foundation, Torino, Italy, ⁵Politecnico di Torino, Torino, Italy

Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing

Thomas A. Hopf¹,² Lucy J. Colwell³, Robert Sheridan⁴, Burkhard Rost⁵, Chris Sander⁶, and Debora S. Marks¹ *

¹Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
²Department of Informatics, Technische Universität München, 80788 Garching, Germany
³MRC Laboratory of Molecular Biology, Hills Road, CB2 0QH Cambridge, UK
⁴Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York City, NY 10065, USA
⁵Correspondence: avl@lmb.cam.ac.uk
DOI 10.1016/j.cel.2012.04.012

Protein structure prediction from sequence variation

Debora S Marks¹, Thomas A Hopf² & Chris Sander²

Sequence co-evolution gives 3D contacts and structures of protein complexes

Thomas A Hopf¹,², Charlotte P I Schärfe¹,³,⁴, Joao P G L M Rodrigues⁵, Anna G Green¹, Oliver Kohlbacher²,³, Chris Sander⁴, Alexandre M J J Bonvin⁵, Debora S Marks¹
Acknowledgments

Debora Marks
(Harvard Medical School)

Chris Sander
(MSKCC)

Burkhard Rost
(TU München)

Charlotta Schärfe
(HMS & Uni Tübingen)