

Coarse-grained structure-based models of sequence evolution

Johan Grahnen
Department of Molecular Biology
University of Wyoming
2011-10-18

Thermodynamics of sequence evolution

$$P_{seq} \sim \Delta G$$

Inference

Phylogenetics

Fold content evolution

Modeling indel substitutions

Forward simulation

Rates post-duplication

Ease of neofunctionalization

Structure-space transitions

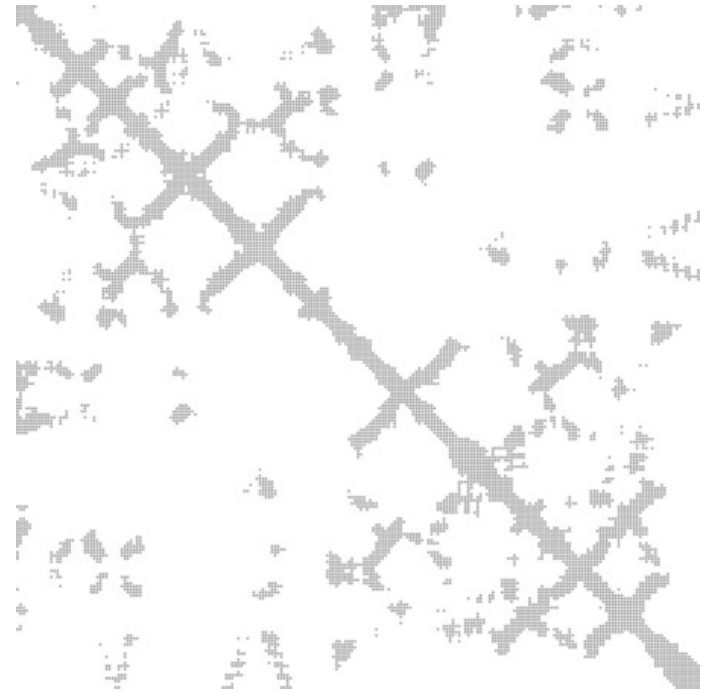
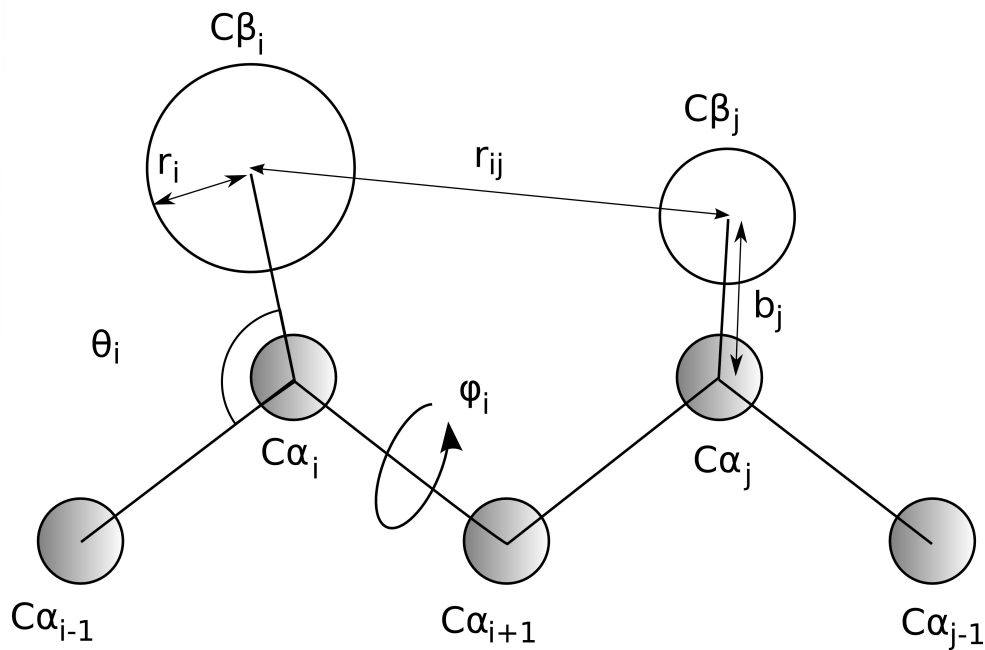
Recent work

Kleinman *et al*, 2010: “relatively weak compared with the most sophisticated site-independent models of evolution”

Nasrallah *et al*, 2010: “dependence in the data increases the amount of topological estimation error”

Lakner *et al*, 2011: “simple models are clearly unable to capture the constraints on sequence evolution”

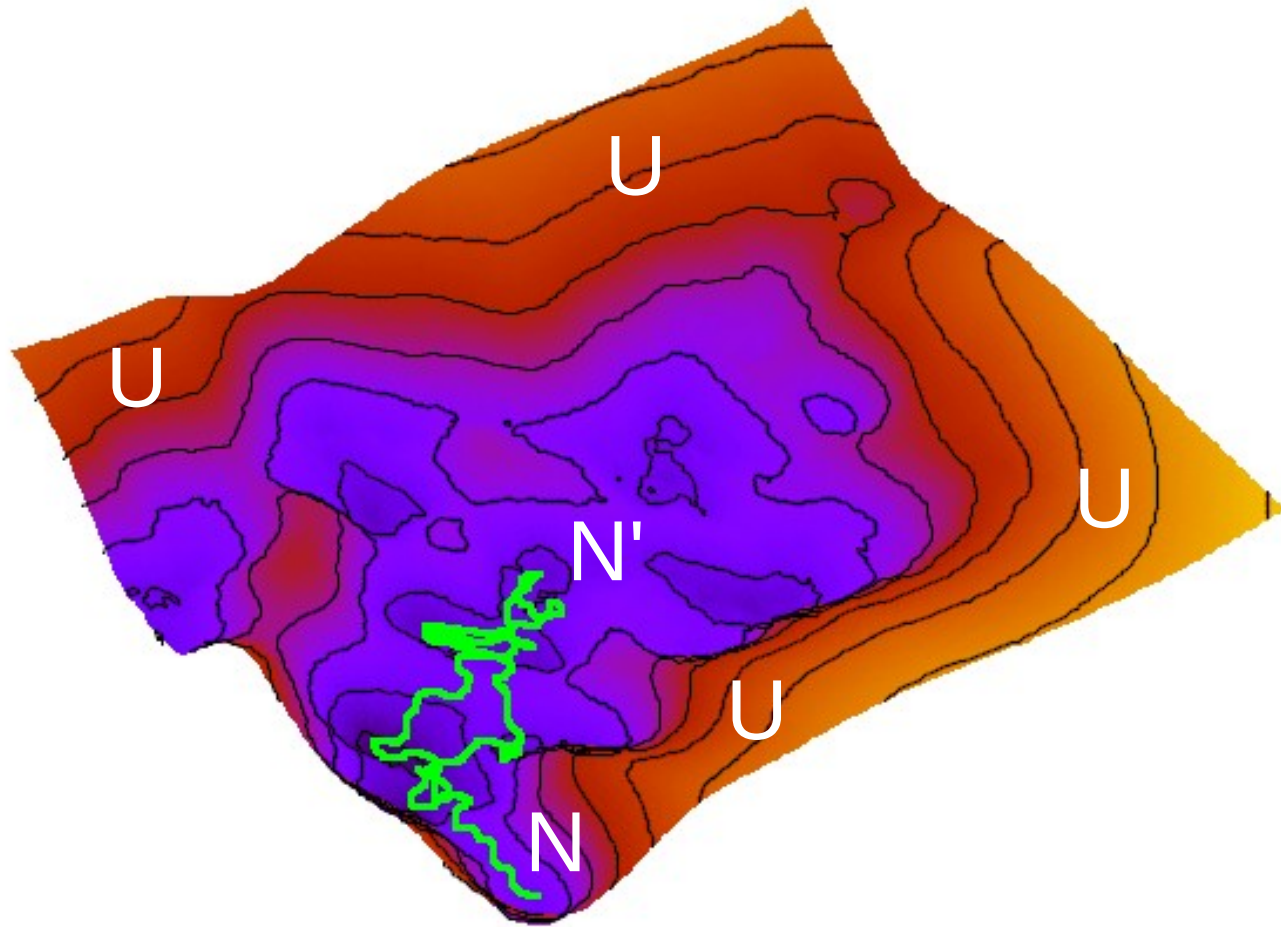
Key assumption #1: Coarse-graining



$$V(s, C) = w_{\Theta} V_{\Theta} + w_{LJ} V_{LJ} + w_{helix} V_{helix} \\ + w_{beta} V_{beta} + w_{ion} V_{ion} + w_{solv} V_{solv} + w_{S-S} V_{S-S}$$

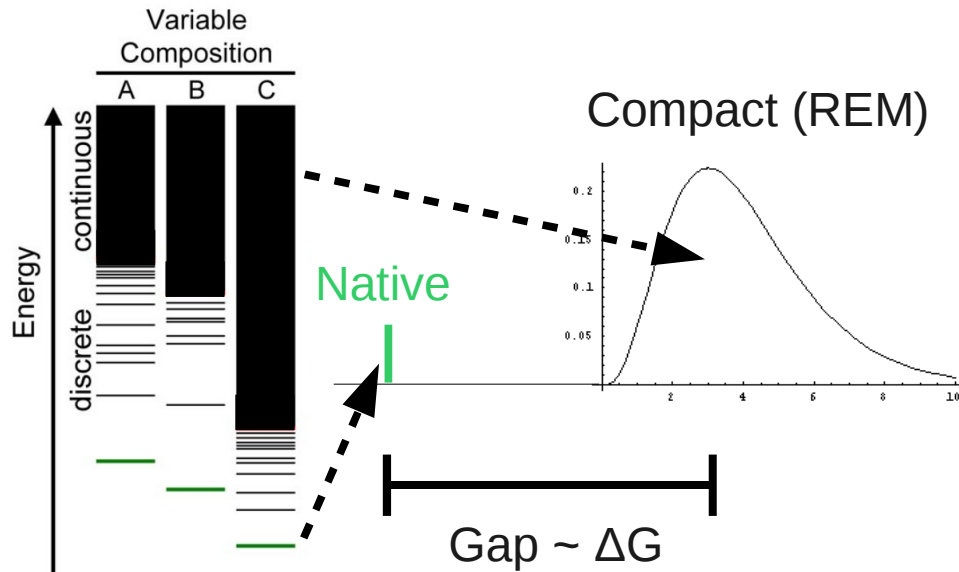
$$E(s, C) = \sum_{i,j} U(i, j) C(r_{ij}) \\ (+ E_{solv})$$

Key assumption #2:
Native state nearly optimally stable

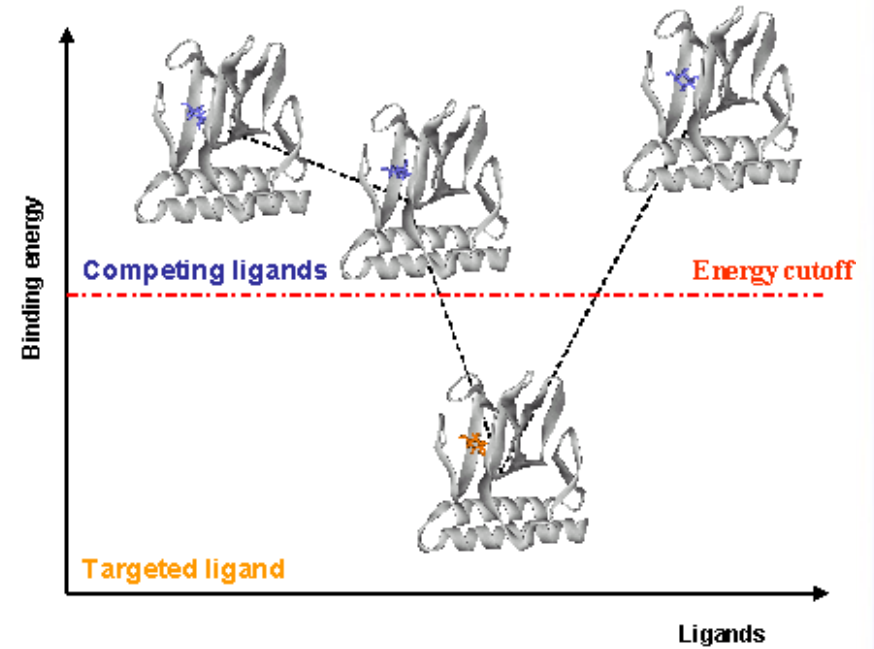


Our approach

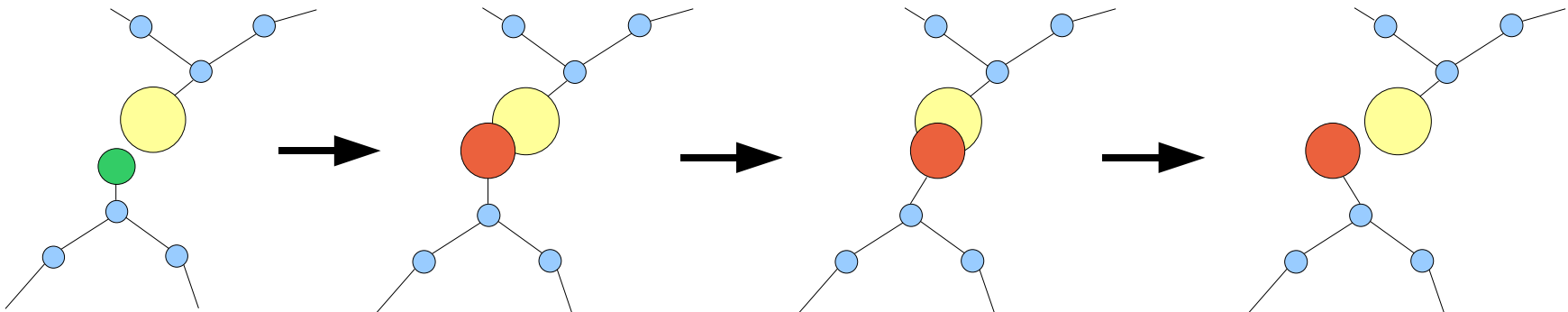
Folding specificity



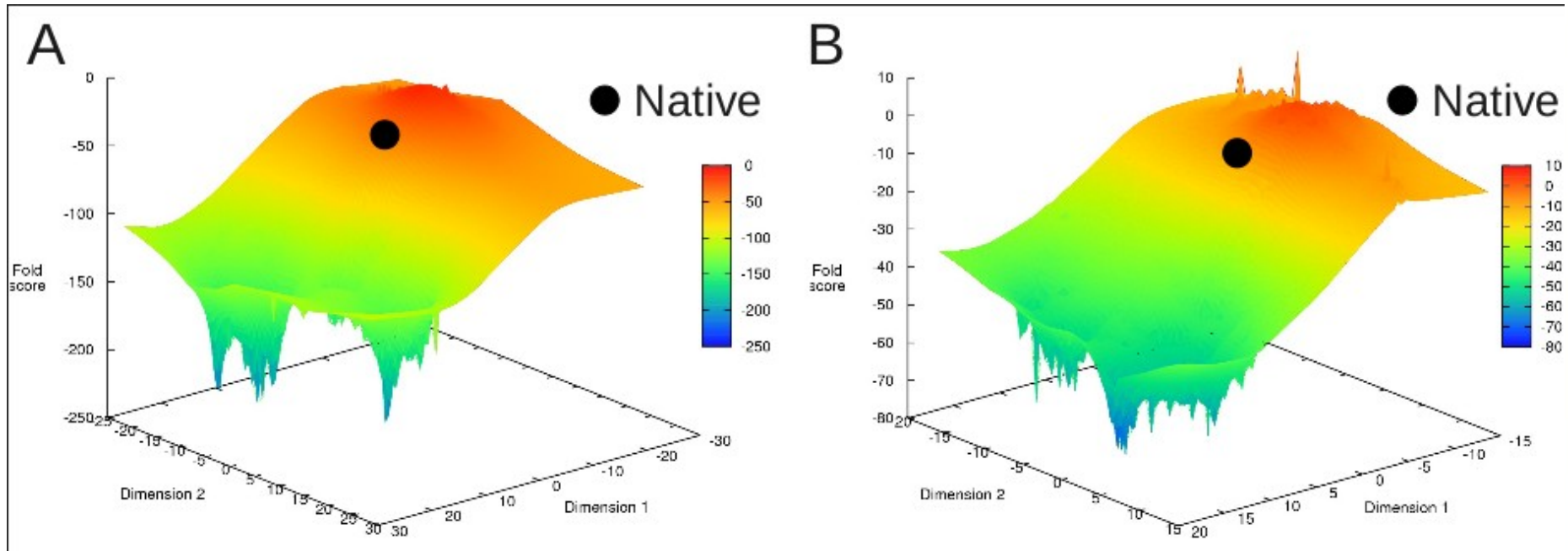
Binding specificity



Residue replacement



Problem: Native stability

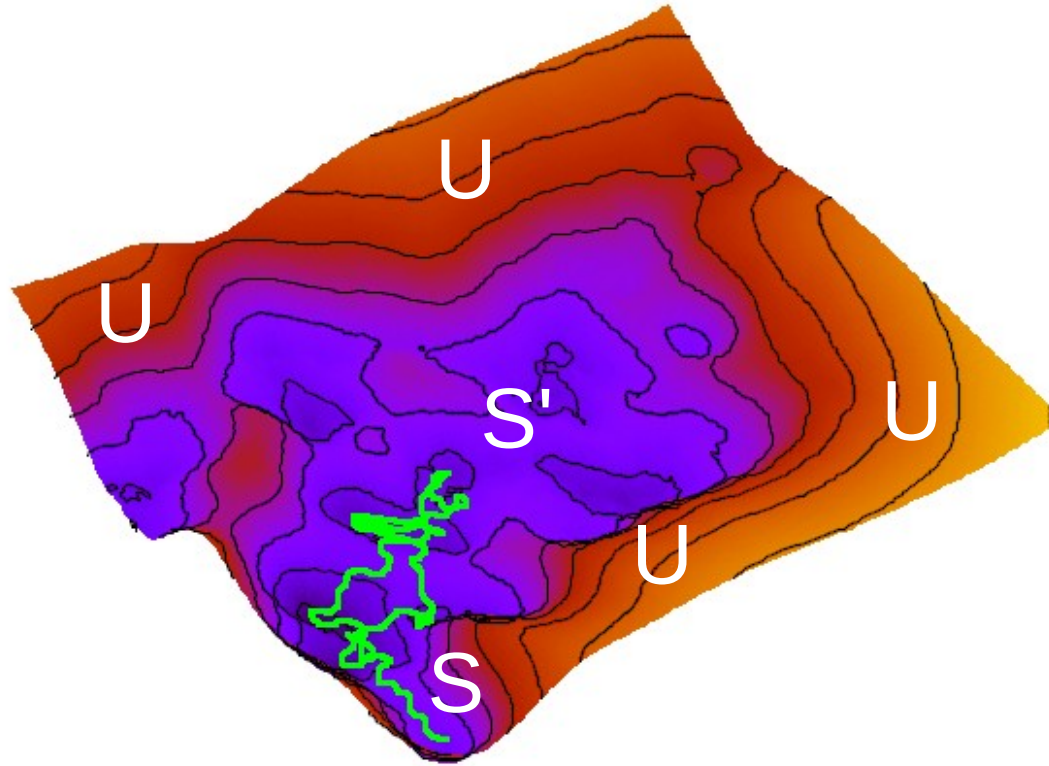


Informational model

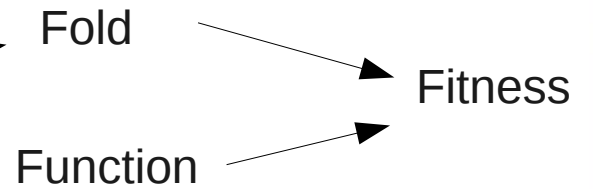
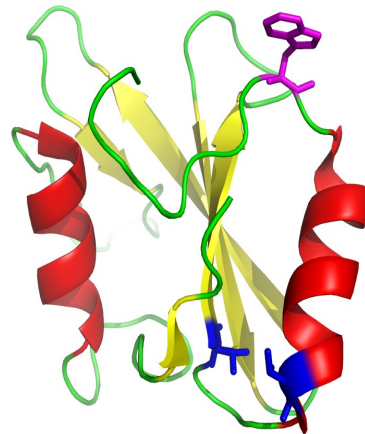
Physics-based model

Sequences rapidly diverge from native state

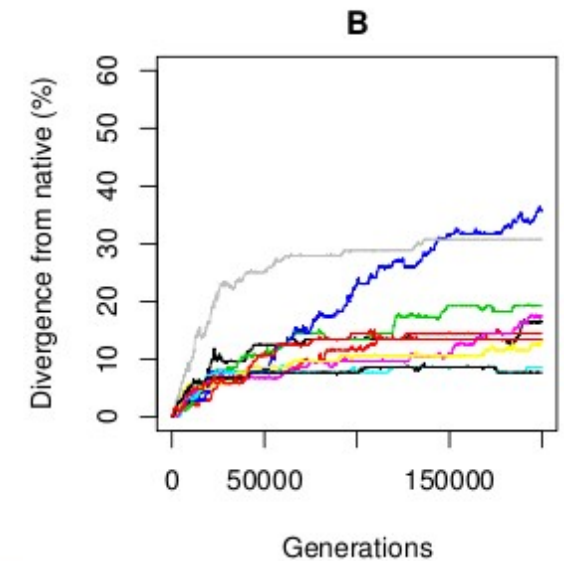
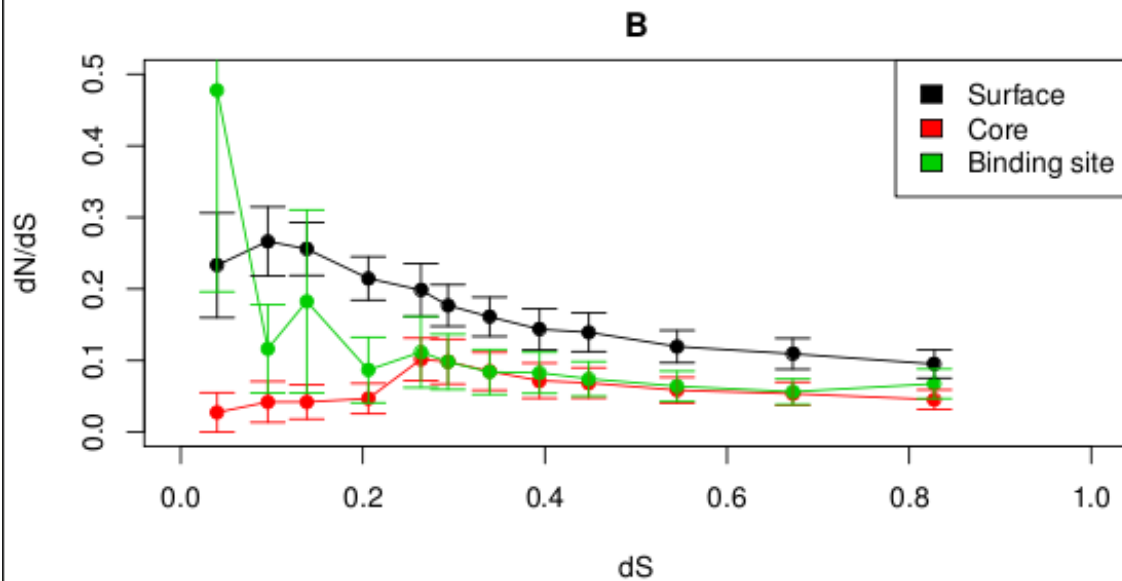
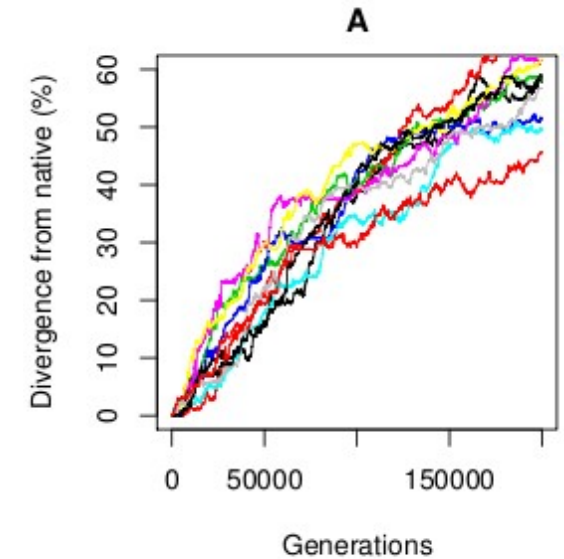
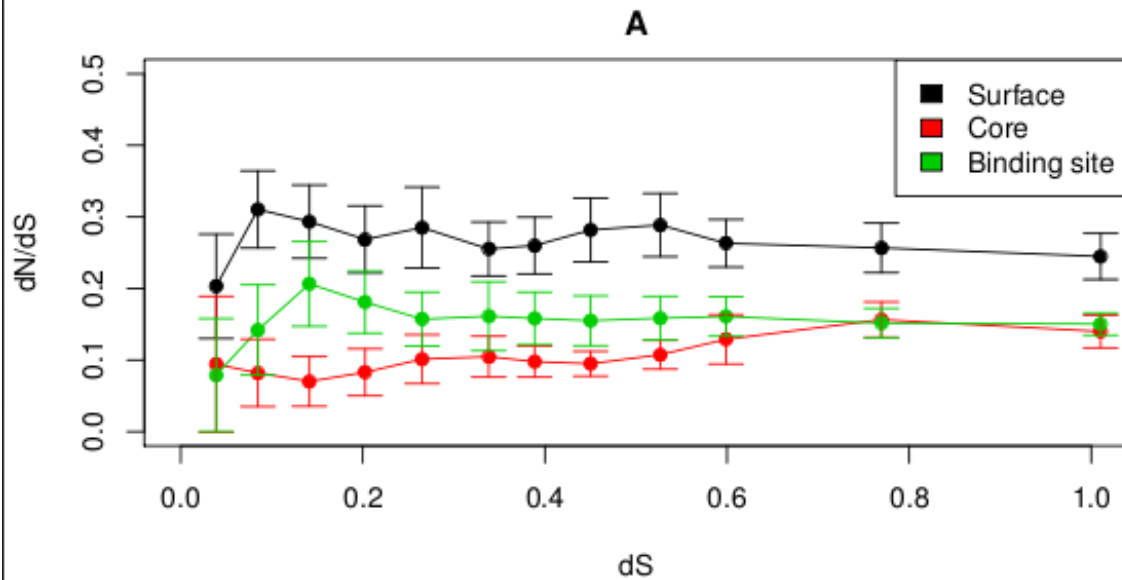
Simulate from stable state



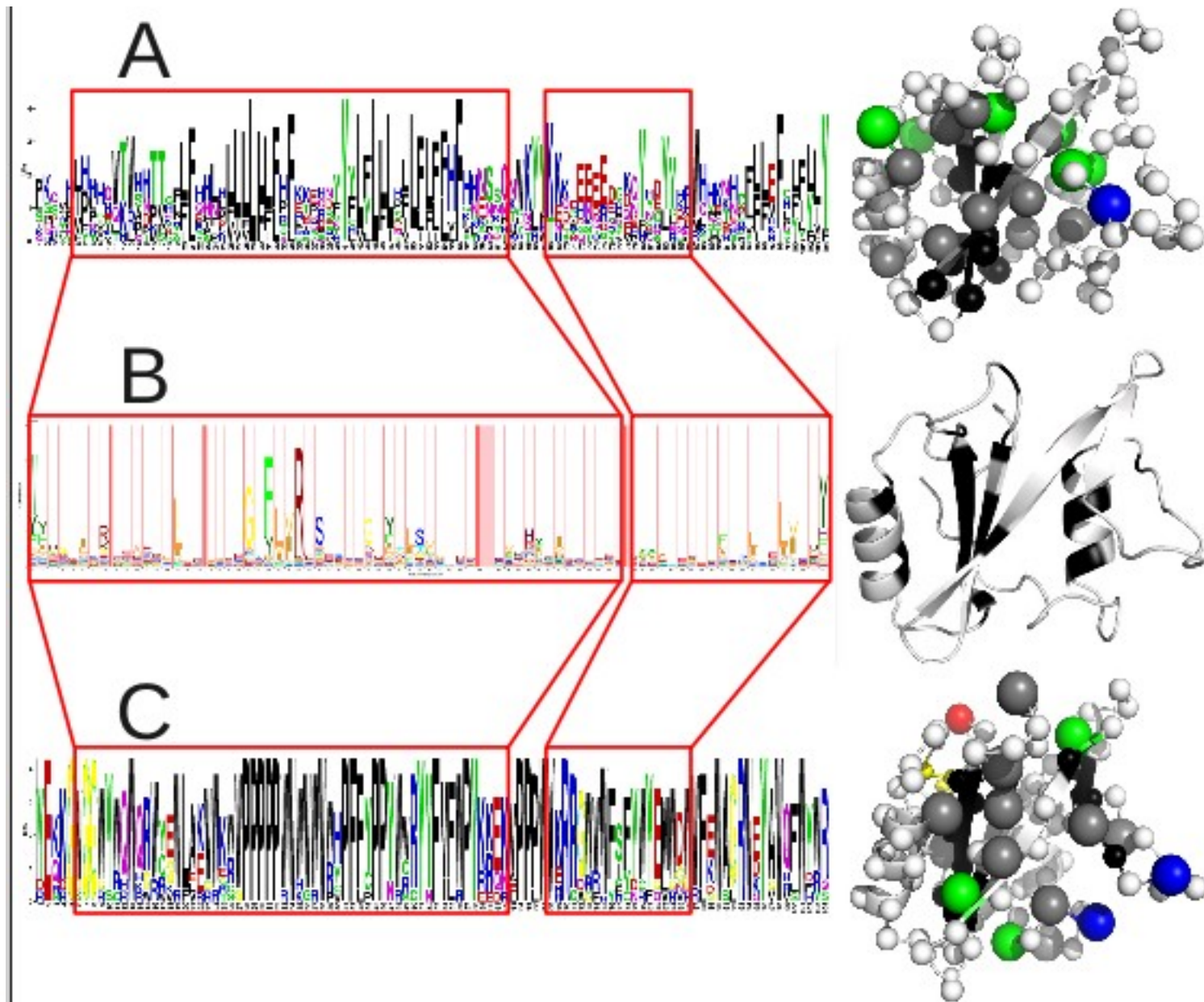
ATGAACGCAGTGGCT....



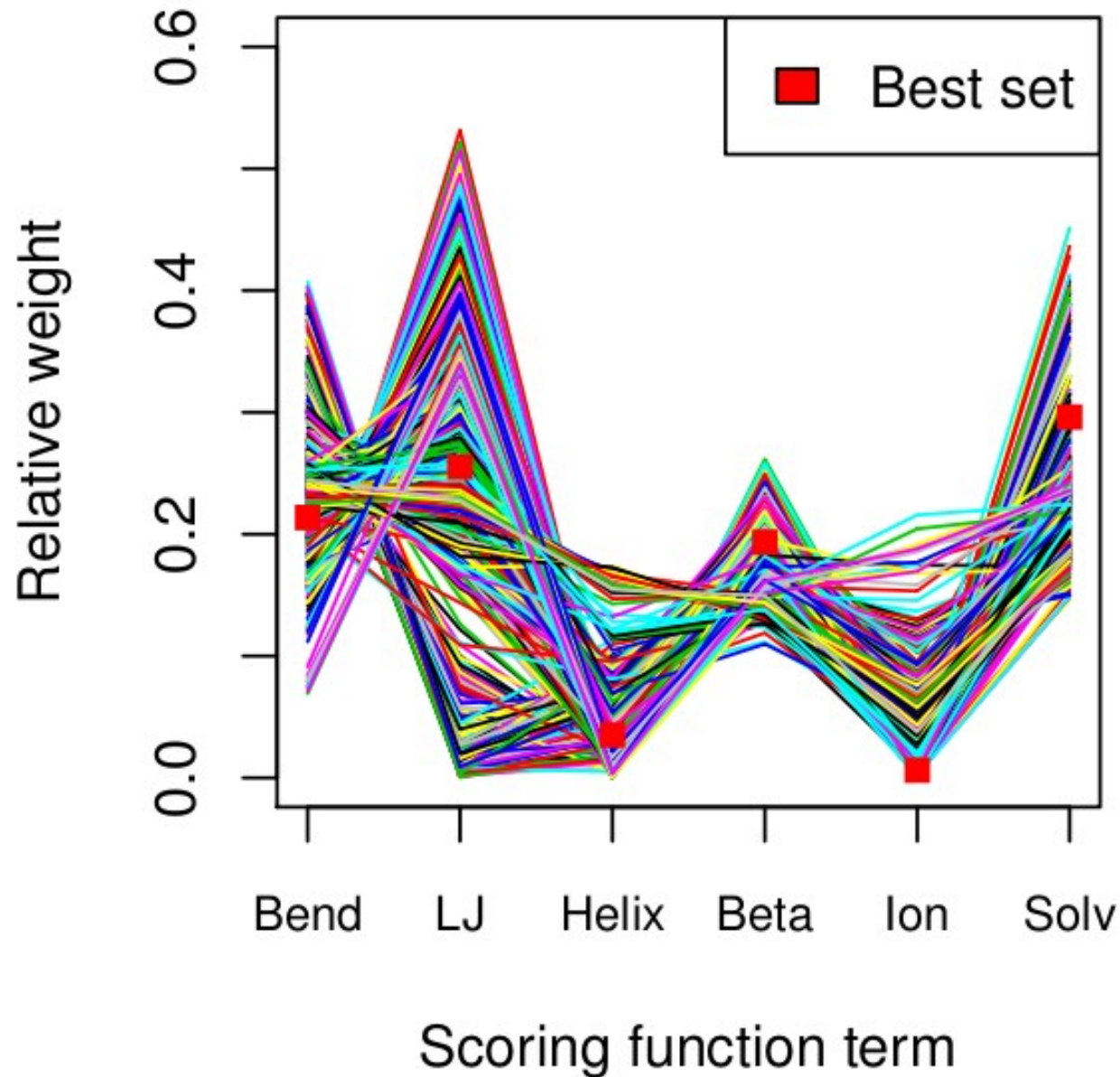
Simulated populations



Simulated vs real sequences



Parameter choice in physical model



Overall model properties

Both: space-filling, hydrophobic residues

Informational: conservative but not fold-specific substitutions

Physics-based: fold-specific, but not native, interactions preserved

Take-home message

Some properties of real sequences reproduced

Structural model(s) too coarse-grained

Informational model non-specific

Physics-based model promising

Acknowledgments

David Liberles

Jan Kubelka

Priyanka Nandakumar

NIH-INBRE

Thank you for listening!

