

Pairwise Decomposition of Many-body Energies using Generic Side-chains and Extreme Optimization in Computational Protein Design

Naigong Zhang,¹ Chen Zeng,² Ned S. Wingreen³

Keywords: protein design, pairwise surface area, pairwise Poisson-Boltzmann energy, combinatorial optimization, simulated annealing, extreme optimization

1 Introduction.

Computational protein design refers to the use of computers to select amino acid sequences that can fold into a given structure. It is the inverse of protein folding, and it is desirable because structure plays an important role in function. Despite some noted successes (see [1] and [2]), fully automated de novo design is still a difficult task. Two main challenges are the finding of appropriate force fields for design and efficient optimization algorithms for searching through an astronomical number of conformation choices. Here we present our recent work in these two areas.

2 Pairwise decomposition of many-body energies.

To capture the effect of hydrophobicity we include a solvation term proportional to surface area to encourage the burial of nonpolar atoms and exposure of polar atoms. For fast graph optimization algorithms such as dead-end elimination (DEE) and the extreme optimization method we will introduce in this work, energy is required to be written in single- or pair-residue terms only; however, surface area is a many-residue quantity. We have improved on a method of pairwise decomposition of surface area,[3] by introducing generic side-chains to be placed at positions where the identity of the residues are not known.[4] We have optimized the parameters for the generic side-chains, and our method is fast and accurate for both total surface area (see Fig. 1) and residue-by-residue area. We have also used generic side-chains to pairwise decompose Poisson-Boltzmann reaction field energy (as calculated using DelPhi) and have achieved similarly good results.

3 Extreme optimization.

Using a backbone-dependent rotamer library, for a 33-residue protein we are currently designing, we have in total about 4000 rotamers and 10^{68} conformations to search through. The deterministic DEE can be applied but it requires the storage of energy and often leaves too many states to enumerate. We recently adapted a combinatorial optimization algorithm, extreme optimization, inspired by the non-equilibrium physics of self-organized criticality. Starting from a random conformation and after defining a fitness parameter for each position, it strives to improve on positions with bad fitnesses (using a power-law distribution to select the site for improvement). Compared to simulated annealing, extreme optimization reaches

¹Department of Physics, George Washington Univ., Washington, DC. E-mail: ngzhang@gwu.edu

²Department of Physics, George Washington Univ., Washington, DC. E-mail: chenz@gwu.edu

³Department of Molecular Biology, Princeton University, Princeton, NJ. E-mail: wingreen@molbio.princeton.edu

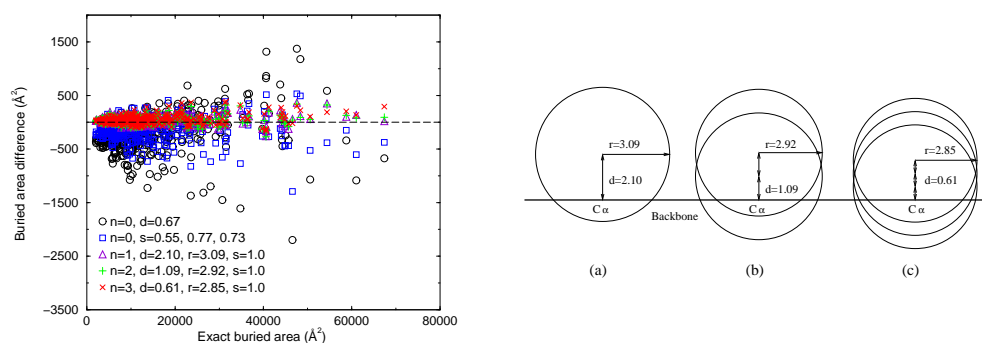


Figure 1: Left: difference of pairwise buried area and exact buried area vs exact buried area for 377 single-domain proteins in the CATH protein set. The five methods are: no generic side-chain ($n = 0$) with one or three scaling parameters and generic side-chains with one, two, or three spheres ($n = 1, 2, 3$). Right: schematic view of optimal 1,2,3-sphere generic side-chains drawn to scale (in Å). d is distance from C_{α} atom and separation among the spheres. r is sphere radius.

low-energy states much more quickly and maintains large fluctuations to continuously search for the global minimum-energy conformation (see Fig. 2).

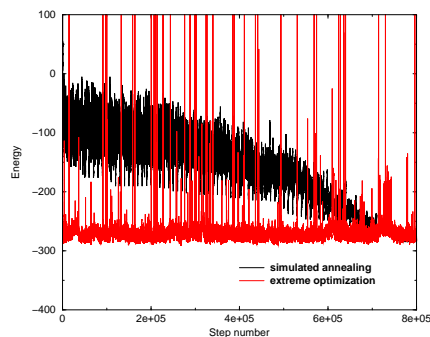


Figure 2: Comparison of simulated annealing and extreme optimization for the design of a 33-residue protein. Note that extreme optimization reaches low-energy states much more quickly and maintains large fluctuations to search for global minimum energy.

References

- [1] Dahiyat, B. I. and Mayo, S. L. 1997. De novo protein design: fully automated sequence selection. *Science* 278:82–87.
- [2] Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. and Baker, D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–8.
- [3] Street A. G. and Mayo S. L. 1998. Pairwise calculation of protein solvent accessible surface areas. *Fold. Des.* 3:253–258.
- [4] Zhang, N. G., Zeng, C. and Wingreen, N. 2004. Fast Accurate Evaluation of Protein Solvent Exposure. *Proteins* 57:565.