Euclidean Variable Neighborhood Search: A method for large computation protein design

David Allouche

Euclidean Variable Neighborhood Search: A method for large computation protein design

David Allouche[1]

1 MIA Toulouse UR-875, INRA 31320 Castanet-Tolosan

contact: david.allouche@inra.fr

Computational protein design (CPD) is an important tool for biotechnology still under development. Early applications led to proteins with novel ligand-binding functions, novel enzyme activity, and proteins that were completely "redesigned": around 2/3 of their sequence was mutated, yet their structure and stability were retained. In the last few years, CPD has allowed the creation of new protein folds, completely new enzymes, and the assembly or deassembly of multiprotein complexes. CPD methods are mainly characterized by (a) the energy function, (b) the description of the folded protein's conformational space, (c) the treatment of the unfolded state, and (d) the search method used to explore sequences and conformations.

Graphical model and cost function network have been recently introduced as new search methods for CPD[1,2] search allowing to find the global minimum energy conformation (with optimality proof). However, it can solved problems with less than 100[3] and 48[4] mutations respectively with Rosetta and Amber forcefield and Dunbrack[5] and Tuffery[6] rotamers library. In this work, we introduce a new search method for CPD dedicated to large instances. The method, based on Variable Neighborhood Search[7] (VNS), uses (partial) tree search in order exhibit the solutions with the lowest energy.

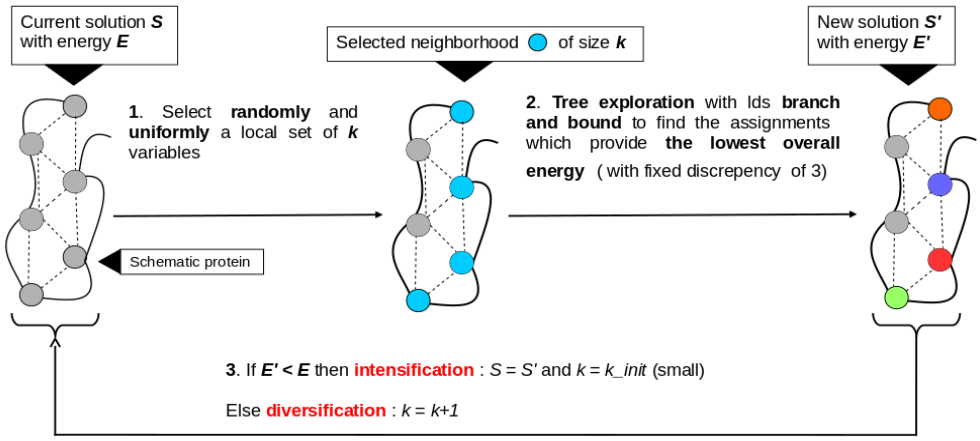

fig 1: schematic representation of variable neighborhood search algorithm

In order to improve its behavior, we implemented a new heuristic taking advantage of the euclidean information provided by the pdb structure for neighbor variable selections. This last one used in conjunction with logarithmic size incrementation of the neighborhood size improves significantly the VNS behaviors. The resulting algorithm, called euclidean VNS, is more robust. It also outperforms Replica-Exchange Method for global minimum energy search. Furthermore, the method can provide direct correlation between energy improvement and protein structure, allowing to identify energetic hot spots in the protein backbone.

References

[1] Allouche et al. Principles and Practice of Constraint Programming. 2012; pp 840–849.
[2] Allouche et al. Artificial Intelligence, 212:59–79, 2014.
[3] Simoncini et al. Journal of chemical theory and computation, 11(12):5980–9,12 2015.
[4] Mignon, D.; Simonson, T. Journal of computational chemistry 2016, 37, 1781–93.
[5] M. Shapovalov and R. Dunbrack , Structure , 19 (8):844-858 2011
[6]Tuffery, P. et al Journal of biomolecular structure & dynamics 1991, 8, 1267–89.
[7] Ouali Abdelkader et al. Uncertainty in Artificial Intelligence (UAI) 2017 Sydney Australia