doi:10.1017/S1431927610057429
Microsc. Microanal. 16 (Suppl 2), 2010

Parallel Genetic Algorithm with Biased Initial Population For The Simultaneous Hybrid Modeling of Molecular Assemblies

M. Rusu, S. Birmanns

University of Texas Health Science Center at Houston, School of Health Information Sciences, Houston, TX 77030

Multi-component macromolecular assemblies are responsible for fundamental biological processes taking place inside organisms. Although little is known about the mechanisms that govern these processes, valuable insights may be gained by studying the molecular architecture of the system relative to the different component biomolecules. Structural information is often determined by various biophysical sources and could be integrated using modeling approaches. For example, an atomic interpretation of low-resolution cryo-electron microscopy (cryo-EM) maps can be achieved by docking each component into the envelope of the entire assembly.

Some multi-resolution modeling techniques examine each biomolecule independently of the other components and identify the optimal placement by exhaustively investigating all possible transformations [2]. Additional spatial constraints can be introduced indirectly if all constituents are considered in a simultaneous registration. Although beneficial, such an approach increases the dimensionality of the search, which in turn prohibits an exhaustive exploration of the score landscape.

Recently, we proposed an evolutionary tabu-search strategy for the simultaneous registration of multiple atomic constituents into the cryo-EM envelope of their assemblies [3]. MOSAEC (Multi-Object Simultaneous Alignment using Evolutionary Computing) is a population-based technique derived from genetic algorithms and enhanced with tabu-search strategies to prevent the unnecessary exploration of local optima. Moreover, MOSAEC uses parallel computing strategies to enable an efficient search through the concurrent exploration of multiple paths in the search space (Figure 1). Our approach successfully identified the spatial organization of multiple components inside the assembly, with accuracies within one order of magnitude of the nominal resolution of the maps, 35-40 Å (Figure 2).

With the goal of further improving the efficiency of our method, we introduce a novel initial population strategy to guide the registration towards meaningful solutions early in the evolution. Classic genetic algorithms start the optimization with a population of randomly distributed candidate solutions. Here, we take advantage of *a priori* knowledge regarding the problem and initialize the starting population with individuals that satisfy spatial constraints. These candidate solutions are created by placing the atomic structures inside the cryo-EM envelope, at the center of the structural features obtained from a Voronoi tessellation [4]. This strategy generates, at least, partial good solutions in the first generations, creating the building blocks for the identification of the global optima. Multiple synthetic and experimental data sets were used to investigate the efficiency of the new method.

References

[1] B. Alberts, *Cell* 92 (1998) 291—294

[2] W.Wriggers and S.Birmanns, *Journal of Structural Biology*, 133 (2001)
[3] M. Rusu W. and S. Birmanns, *Journal of Structural Biology*, in press
[4] W. Wriggers, R.A. Milligan, K. Schulten, *J.A. McCammon, Journal of Molecular Biology,* 284 (1998) 1247—1254.
[5] The present work was supported by NIH grant R01GM62968, a grant from the Gillson Longenbaugh Foundation, and startup funds from the University of Texas Houston (to S.B.)

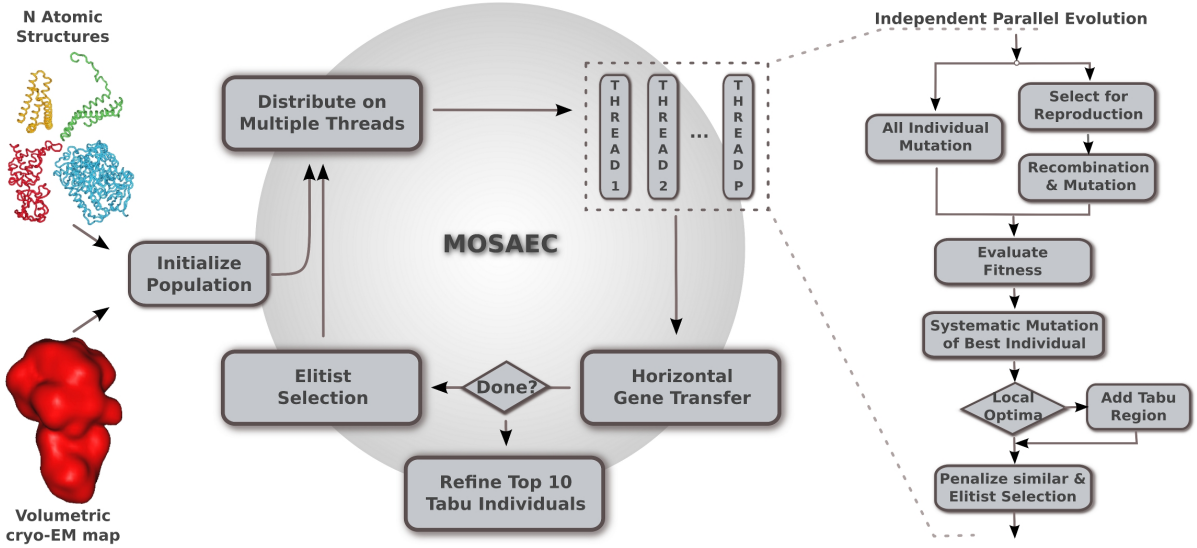

FIG. 1. Flowchart of MOSAEC: the classic genetic algorithm was enhanced with biased initial sampling, tabu search and parallel computing strategies

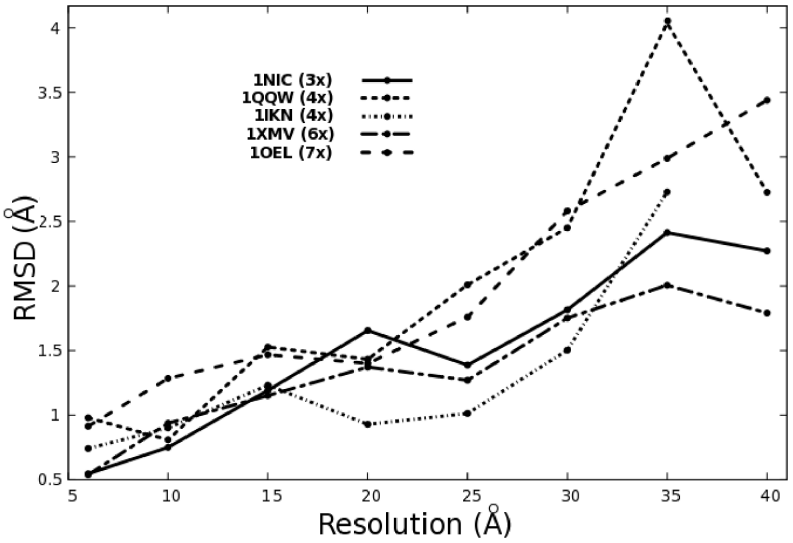

FIG. 2. The accuracy of MOSAEC as a function of the resolutions. The legend shows the PBD ID followed by the number of components simultaneously docked.

https://doi.org/10.1017/S1431927610057429 Published online by Cambridge University Press