

# Experiences on Memetic Computation for Locating Transition States in Biochemical Applications

Mostafa M. H. Ellabaan  
School of Computer Engineering  
Nanyang Technological University  
most0001@ntu.edu.sg;  
Mostafa.mhashim@gmail.com

Yew Soon Ong  
School of Computer Engineering  
Nanyang Technological University  
asysong@ntu.edu.sg

## ABSTRACT

Transition states constitute an essential component of the chemical reaction rate theory and are important for understanding the structural and mechanical properties of the systems, and how they react under different environmental conditions. The challenges behind the discovery of transition states however arise from the existence of a large number of rules and constraints as well as the computational complexity involved in energy calculations. In this paper, we present some recent successes of memetic computation and experiences for the discovery of first-order saddle points or transition states in biochemical systems. We show that the exploitation of *a priori* knowledge on the inherent structure of the problem in the form of memetic search operators led to enhanced search convergence and solution quality. This paper then concludes with a brief discussion on the potentials for the redefinition of memes not only as symbiosis of search operators but as building blocks for complex biochemical systems.

## Categories and Subject Descriptors

J.2 [Computer Applications]: Physical Sciences and Engineering;  
I.2.8 [Computing Methodologies]: Problem Solving, Control Methods, and Search

## Keywords

Memetic Computation, Transition states, Biochemical systems

## 1. INTRODUCTION

Unlike other stable structures, transition structures are energy maxima along the minimum energy path connecting stable states or isomers, thus posing fundamental difficulties in their discovery. Due to their fleeting nature, transition structures are almost impossible to be isolated experimentally. Therefore, the search for such structures computationally has been deemed as central but remains to pose great challenges. In the last decades, several notable optimization methods have been proposed for the identifications of first-order saddle points. In spite of the efforts, the search methodologies proposed previously for finding first-order saddle points are plagued with several pitfalls, including the *lack of sufficient precision*, *slow convergence rate*, as well as, the *limited capacity in discovery of multiple saddle points*. Memetic computation is an emerging field of computational intelligence [1, 2] that has gained increasing attention and success in the rectifications of the issues stated above. In addition to genetic evolution, memetic computation incorporates cultural biases and learning, therefore allowing faster convergence and high precision. In this paper, we thus present a recent exposition of memetic computation in the discovery of saddle points in biochemical applications.

## 2. PROBLEM DEFINITION

A transition state or first-order saddle point is defined as a stationary point with a vanished gradient and only one negative eigenvalue in the hessian matrix. These points can be mathematically expressed as:

$$\mathbf{x}_T = \left\{ \mathbf{x}_k \mid \left( \frac{\Delta f(\mathbf{x}_k)}{\Delta \mathbf{x}_k} = 0 \right) \& (\gamma_{k,i} < 0) \& (f(\mathbf{x}_k) < f_{max}) \right\} \quad (\text{Eqn. 1})$$

where  $\mathbf{x}_i \in \mathbf{R}^d$ ,  $d$ ,  $f(\mathbf{x}_i) \in \mathbf{R}$ , and  $\mathbf{x}_T$  denotes the first-order saddle points, dimension size, energy and all possible 1<sup>st</sup> order saddle points with energy below  $f_{max}$ , respectively.  $\gamma_{k,i}$  is the only negative eigenvalue of the hessian matrix  $\mathbf{H}_k$  while ‘&’ denotes a logical AND operator.

## 3. MULTI-MODAL MEMETIC FRAMEWORK

In our studies, the memetic framework depicted in Fig. 1 was introduced to tackle the problem of locating transition states, which can be formulated as a general non-linear mathematical problem of the form:

$$f_{ts}(\mathbf{x}) = \frac{f(\mathbf{x}) - f_{max}}{(|\mathbf{g}| - e)((n-1) - e)} \quad (\text{Eqn. 2})$$

where  $f(\mathbf{x})$ ,  $\mathbf{g}$ , and  $n$  is the function value, gradient and number of negative eigenvalues at the configuration  $\mathbf{x}$ , respectively.  $e$  is a small number to prevent zero division error.

The memetic framework consists mainly of three major components, namely the evolutionary, cultural and diversity operators. Evolutionary operators include those inspired by natural evolution of species, namely, reproduction operators like crossover, mutation and selection. Culturally-inspired operators include those developed based on prior knowledge about the problem domain, which are typically harnessed from human experts. This can be in the form of life-time learning procedures deemed suitable for enhancing the solution quality or search efficiency through the use of first-order saddle point local searchers or specially crafted culturally-biased procedure based on the expert-knowledge attained. Diversity operators include those inspired from the ecological term “*niche*” where individuals locally compete for the resources. Such an operator has the benefits of maintaining diversity and attaining multiple high quality saddle points in the search. In what follows, we showcase two important biochemical systems.

## 4. MOLECULAR WATER CLUSTERS

Being the most abundant and more importantly crucial substance on earth, without water, there would not be any form of life as we know. Understanding how water evaporates as well as how ice forms and melts requires profound knowledge on transition states. To locate transition states in water clusters, the optimization

algorithm needs to maintain the unity of the participating water molecules at the same time keep their constituting atoms far apart to avoid both the erroneous bonding and the steric effect, i.e. preventing atoms from overlapping or getting too close to each other. The above framework is extended in [3], incorporating culture of molecular water cluster to explore the feasible region of the search space. Water cluster-based slicing crossover operators and mutation strategies that consider the dynamics of molecular water clusters in the process of investigating transition states were introduced in [3] and depicted in Fig 2, composing Mol-MA. Mol-MA has shown not only to reproduce previously found transition states in water clusters, but also established newly discovered transition states for sizes 2-4 water molecules.

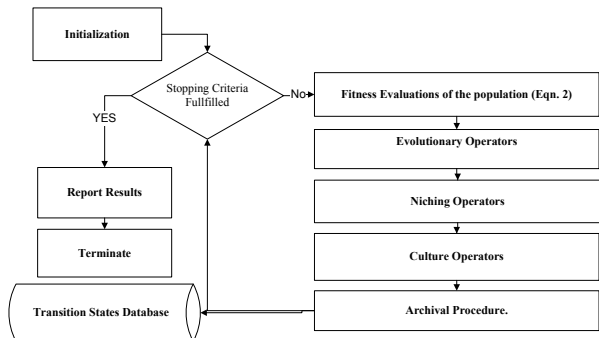


Fig 1: Multimodal Memetic Framework for Locating Transition States

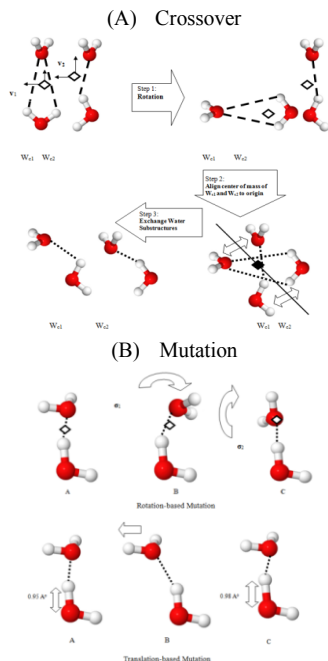


Fig 2: Water-Cluster Crossover and Mutation [2].

## 5. COVALENTLY-BONDED MOLECULAR SYSTEMS

Unlike water clusters where water molecules are bonded together with vulnerable—easy-to-break and easy-to-form—hydrogen bonds, covalently-bonded molecules are maintained by a sequence of covalent bonds that preserve the identity of the molecule. Due to the abundance of such molecules in nature and their importance in biology and chemistry, a memetic computational methodology for the discovery of low-energy transition states of covalently-bonded

molecules is introduced in [4]. To address the challenges of maintaining the same set of bonds over the course of searching for transition states and through avoiding the steric effect, three culturally-inspired, tree-based evolutionary operators (See Fig 4) were introduced in [4]. A molecular memetic algorithm (TCM-MA)—tailored specifically to deal with molecules that involve covalent bonding but contain no cyclic structures using the three novel evolutionary operators—is then proposed for the efficient search of the transition states of ring-deficient covalently-bonded molecules. The proposed TCM-MA efficiently discovered more transition states compared to other state-of-the-art algorithm [3].

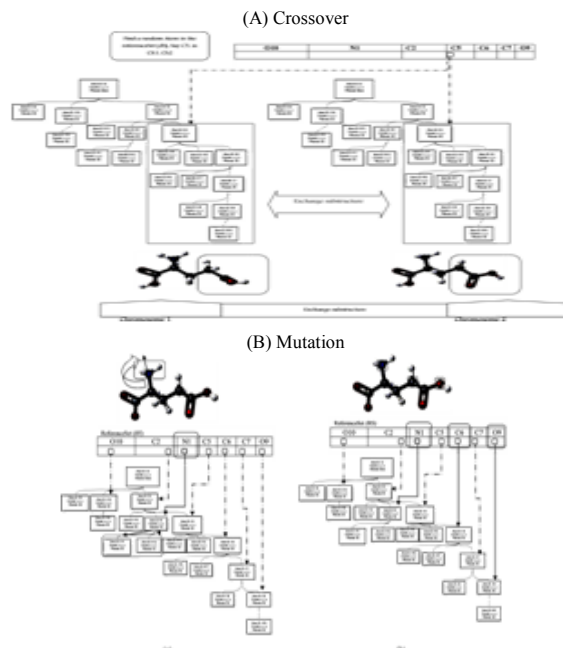


Fig 3: Culturally-inspired, Tree-based Crossover and Mutation [3].

## 6. CONCLUSION AND FUTURE WORK

Memetic computation extends the evolutionary computation framework, incorporating local searchers, domain knowledge and culture to enhance search performance. In this paper, some experiences on memetic computation for locating transition states in biochemical systems is presented. In future, the memes shall be treated not only as symbiosis of different local searchers but also as the building blocks of complex bio-molecular systems that can be mined from the database generated in our previous studies or biochemical databases currently available online, posing an interesting future research. Finding transition states in complex drug molecules, protein as well as identifying minimum energy paths that bridge non-living matters and minimal living organism shall be our next immediate interest.

## 7. REFERENCES

- [1] Y. S. Ong, M. H. Lim and X. S. Chen, "Research Frontier: Memetic Computation - Past, Present & Future", IEEE Comp. Intell. Mag., Vol. 5, No. 2, 24 -36, 2010.
- [2] Ellabaan, M., Ong, Y. S., Nguyen, Q. C. and Kuo, J.-L. Evolutionary Discovery of Transition States in Water Clusters. Journal of Theor. and Comp. Chemistry, In press 2012.
- [3] Ellabaan, M., Handoko, S. D., Ong, Y. S., Kwok, C. K., Bahnassy, S., Ellassawy, F. and Man, H. Y. A tree-structured covalent-bond-driven molecular memetic algorithm for optimization of ring-deficient molecules. Comp. and Math. with Appl., In press 2012.