International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 03 | Mar-2016 www.irjet.net p-ISSN: 2395-0072

Metaheuristic Techniques for Conformational Search

SIEW MOOI LIM[1], MD. NASIR SULAIMAN[2], NORWATI MUSTAPHA[3], ABU BAKAR MD. SULTAN[4]

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY, UNIVERSITI PUTRA MALAYSIA, MALAYSIA

Abstract *The drawback in conformational search (CS) is in locating the most stable conformation of a molecule with the minimum potential energy based on a mathematical function. The number of local minima grows exponentially with molecular size and this makes it that more difficult to arrive at a solution. It had been confirmed that CS belongs to the category of NP-hard (non-deterministic polynomial time) problem. Such complexity requires an equally long amount of time to achieve resolution. This phenomenon is thus known as the 'combinatorial explosion'.*

Metaheuristic techniques have been constantly used in solving CS problems. These population-based probabilistic techniques explore conformational space by random perturbation of atomic Cartesian coordinates or the torsion angles of rotatable bonds. These methods focus on exploring a search space with maximum efficacy. With one or more solutions in the beginning, metaheuristic method follows with a more iterative approach to optimize the search in promising areas away from local solutions. This method is often employed in circumstances where the exact solution methods are unfeasible within a limited time frame. As such, this paper presents various past metaheuristics approaches that have been brought forth in regards to the problem of an effective exploration of the conformational states of molecular systems. Each metaheuristic method is accompanied by its advantages and disadvantages. The concepts of each approach will be explained and their respective applications are discussed.

Key Words: *Model building methods, distance geometry, smoothing methods, systematic search*

1. INTRODUCTION

Conformational search (CS) is a term familiar to those in the field of applied mathematics and computational chemistry. CS is mathematically represented as a continuous global optimization problem. In CS, the variables are the torsion angles or coordinates that are used to represent the conformation of the molecule (e.g. polypeptide chain). The objective function value is the potential energy function. By varying the values of the variables, the global minimum value of the objective function can be achieved; that is to locate the most stable conformation of a molecule with the minimum potential energy. Years of research have seen CS performing important and widely used molecular modeling applications which include flexible rings and macrocycles molecules [1], cyclic, acyclic, mixed single molecules and host-guest complexes [2], molecules under inhibited conditions, peptide and protein folding [3], simulations of protein-ligand docking [4, 5] and various drug design applications such as Quantitative Structure Activity Relationship (QSAR), virtual screening of virtual libraries and active analog approaches [6].

The exact methods are unable to solve the complexity of CS problems. Therefore, metaheuristic techniques [7] including genetic algorithm have become mandatory and has attracted considerable attention especially from evolutionary algorithm community.

We have introduced a novel real coded genetic algorithm which is capable in solving two CS application problems i.e minimizing a molecular potential energy function and finding the most stable conformation of pseudoethane through a molecular model that involves a realistic energy function [8-12]. The focus of this paper has been pivoted on the five major categories of approaches including five common metaheuristic techniques that have been applied to CS problems.

2. Application of Metaheuristic Techniques to Conformational Search Problems

Metaheuristic techniques include all stochastic algorithms with randomization and local search. Randomization allows a shift from the algorithms in a local search to the global scale. Therefore, almost all metaheuristic algorithms were designed to suit the global optimization. Metaheuristic algorithms consist of two major components namely exploration (diversification) and exploitation (intensification). GA, Tabu search and particle swarm optimization are reported to heavily study on maintaining a good balance between exploration and exploitation in preserving diversity [13, 14].
The function of exploration is such that it generates diverse solutions in order to explore the search space on the global scale. On the other hand, exploitation draws upon the local search region, in which the information of a current good solution in this region will be further

exploited. Good optimality is thus achievable through appropriate combination of these two major components.

Metaheuristic algorithms are applied to difficult problems in which deterministic or traditional operations research techniques may not be suited. Table 1.1 outlines some complexities of the problems, which are difficult to solve. Although many metaheuristics techniques have been proposed in the literature to solve CS problems, a majority of these algorithms are directed toward low-energy solutions. Furthermore, they explore the search space by applying some heuristic techniques to generate only random conformers without the generation of large number of conformations [15].

However, providing a complete review to all the metaheuristic methods is rather impossible. The next sub-sections present five widely used metaheuristic in solving CS problems found in the literature. These five algorithms are grouped into five kingdoms distilled from the broader fields of study of their own. The five kingdoms with their respective algorithms in the brackets are as follows:

1) Probabilistic algorithms (Population-based incremental learning algorithms)
2) Markov chains algorithms (Monte Carlo)
3) Physical algorithms (Simulated annealing)
4) Swarm algorithms (Bees algorithms)
5) Other metaheuristic algorithms (Tabu search)

2.1 Probabilistic Algorithms

Probabilistic algorithms model a search problem space using a probabilistic model of candidate solutions. The primary aims of these approaches are hinges upon methods that build models and estimate distributions in search domains. Pelikan et al [16] presented a comprehensive summary of the core approaches and their differences of the field of probabilistic optimization algorithms. Among the most sought after algorithms in this group are Bayesian optimization algorithm, univariate marginal distribution algorithm, compact GA and cross-entropy method. The following Section goes into detail on population-based incremental learning algorithm (PBIL) with examples based on CS problems.

2.1.1 Population-Based Incremental Learning Algorithm

PBIL was first developed by Shumeet Baluja in 1994 [17] through the combination of evolutionary optimization and hill climbing. It comes under a class of algorithms called estimation of distribution algorithms (EDAs) [18], also referred to as Population Model-Building Genetic Algorithms (PMBGA). This is a variant of the GA where the genotype of an entire population is evolved rather than individual members, thus reducing the memory required by the GA.

The information processing objective of the PBIL is conducted by lowering the population of a candidate solution to a single prototype vector of attributes so that candidate solutions are created and assessed. Updates and mutation operators are performed to the prototype vector instead of the generated candidate solutions. The basic outline of the PBIL algorithm is shown in Fig 1.1.

SM Long et al [19] conducted a CS of molecules using PBIL whereby only the changes in dihedral angels are considered to obtain the optimal value in the potential energy of the system. That being said, bond stretching and bond angle deformation are allowed in the calculation of the fitness of each conformation. As a result, PBIL was able to locate global energy minima of long alkane chains (highly flexible large molecules), and to obtain high-initial convergence rates on cycloheptadecane and drug-like molecules (rigid molecules).

2.2 The Metropolis Algorithms

This is a class of sample-generating methods that focuses on the stochastic sampling of a domain which provides good average performance and offers a low chance of the worst case performance. Using Markov chain random walk with known transition probability, it directly draws samples from various highly complex multi-dimensional distributions. Due to this characteristic, problem with huge degrees of freedom like CS, which has large, high-dimensional search spaces becomes possible. Fig 1.2 presents the Markov chain algorithm for optimization. The mechanism of Markov chain Monte Carlo method with examples given in CS area will be discussed in the subsequent section.

Start with $\zeta_0 \in S$, at $t = 0$

while (criterion)

Propose a new solution Y_{t+1};

Generate a random number $0 \le P_t \le 1$;

$$\zeta_{t+1} = \begin{cases} Y_{t+1} & \text{with probability } P_t \\ \zeta_t & \text{with probability } 1-P_t \end{cases}$$

end

Fig. 1.2: Optimization as a Markov Chain [20]

2.2.1 Monte Carlo Method

Since the 1990s, Markov chain Monte Carlo has emerged as a powerful tool for Bayesian statistical analysis and potentially optimization with high nonlinearity. In the context of CS, Monte Carlo was used as a simulated method in order to locate the minimum energy structure of a given molecular system. The conformation is generated randomly by variations on the Cartesian or internal coordinates. The CS starts by calculating the energy, E_0 for an arbitrary conformation. This is subsequent by randomly changing the dihedral angles to produce a new conformation then calculate its energy, E accordingly. If $E < E_{0,}$ then this new conformation is accepted as the starting point for the next iteration; or if the Boltzmann factor of the energy difference is larger than a random number between 0 and 1. Otherwise, the previous conformer is retained for the next iteration.

The Boltzmann factor (BF):

$$BF = e^{-\frac{E-E_0}{RT}}$$

This process is recurrent up to a point where a set of low conformers has been generated. Monte Carlo methods have been observed in peptides, proteins and bio-molecular systems [21, 22].

2.3 Physical Algorithms

Because it is inspired by physical processes, this group of algorithms can be easily referred to as nature inspired algorithms with mixtures of local neighborhood-based and global search techniques. Among the physical inspiring systems are music, complex dynamic systems like avalanches and the interplay between evolution and culture. The common approaches that belong to this group are memetic algorithm, extreme optimization, harmony search and cultural algorithm. The following section will uncover the details of simulated annealing (SA) from this group, inspired by metallurgy with examples given in CS problems.

2.3.1 Simulated Annealing

SA is an adaptation of the Metropolis-Hastings Monte Carlo algorithms. It is the study between statistical mechanics and liquids freeze or metals re-crystallise in the process of annealing. Fig 1.3 shows a general overview of SA.

```
s ← GenerateInitialSolution ()
T ← T0 // Temperature parameter T
While termination condition not met do
    s' ← PickAtRandom (N(s))
    If (f(s') < f(s) then
        s ← s'
    Else
       Accepts s' as a new solution
       probability p(T,s',s)
    Endif
    Update(T)
  endwhile
```

Fig. 1.3: Simulated Annealing Algorithms [7]

In an annealing process, a physical system is heated at high temperature and disordered until it melts and then slowly cooled so that the system at any time is approximately in thermodynamic equilibrium. The purpose of the entire arrangement is to upscale the size of the crystals in the material and reduce their defects, hence producing the most stable (crystalline) material. The energy of the atoms increase and the atoms move freely during the heating period. The strong suit of SA lies in its ability to avoid being trapped in local minima. In terms of CS problems, at high temperature, this method is able to overcome the energy barriers to explore different regions of the conformational space. On the slow cooling schedule, it offers a new low-energy configuration to be discovered and exploited.

Theoretically, SA can explore all conformational space. Besides torsional rotations, it can sample structural variations and thermodynamic properties can be derived from molecular dynamic movement. The drawback of this method is the infinite temperatures steps to be equilibrated, rendering the simulation impractical.
Because of this, it is often time consuming and it tends to stay in a local minimum at low temperature. This method was widely applied in the CS of molecules [23, 24].

2.4 Swarm Algorithms

Swarm intelligence refers to a field of computational systems that sprung from the collective intelligence through the interplay of various homogeneous agents in the environment. These agents include colonies of ants, flocks of birds, and schools of fish. They can be applied in searching for optimal solutions due to its adaptive features. The following section discusses bees algorithm

(BA) from this group as a dominant sub-fields of the paradigm in CS.

2.4.1 Bees Algorithm

BA is inspired by the foraging behavior of honey bees colony. The honey bees colony manages its foraging activity very well. In order to cover a wider search area, the foragers are sent in multiple directions simultaneously. Honey bees colony can fully concentrate on searching and selecting the most profitable nectar sources from the available sources with flexible adjustments in the searching pattern accurately [25]. Foraging in honey bees colony is aimed at gaining a maximum level of food by visiting the rich food sources. Fig 1.4 depicts the pseudocode of BA.

HAA Bahamish et al [26] successfully implemented the bees algorithm in protein CS. The protein conformations are normally represented as a sequence of torsion angles [27-29]. The protein CS was conducted by diversifying the values of the torsion angles randomly to locate the lowest free energy conformation. The conformation energy was subsequently calculated using ECEPP/2 force fields [30].

2.5 Other Metaheuristic Algorithms

A wide range of algorithms manage the application of an embedded neighborhood exploring (local) search process are grouped together under the category of other metaheuristic algorithm. These algorithms encompass random searches, scatter searches, hill climbing, iterated and guided local searches as well as variable neighborhood searches. These algorithms implement all sorts of strategies with different starting points issued to a neighborhood searching technique for refinement. This process is reiterated to arrive at the potential unexplored areas. The Tabu search (TS) from this group is discussed in the following section with examples given in CS.

2.5.1 Tabu Search

TS is based on "steepest descent-modest ascent" strategy. The procedure involves the search for the next local minimum, followed by modest ascent path to escape from that local minimum to find the next local minimum. TS employs a *tabu list* to circumvent reverse modes and cycles. The list memorizes the moves previously done by keeping a history of all recently considered candidate solutions. Whenever a new solution is adopted, it is riveted into the list. As the list gets longer, the older solutions will be removed and it is no longer taboo to reconsider. The main components of TS consist of a short term memory which is used to avoid retracting cycles and to escape from local minima. An intermediate-term memory structure is used to guide the search to different and more promising regions of the search space. A longer-term memory structure is used as a learning process to promote a general intensity and diversity strategies. Fig 2.5 gives a simple description of TS.

```
s ← GenerateInitialSolution
TabuList ← 0
While termination conditions not met do
   s ← ChooseBestOf(N(s) | Tabulist
   Update(TabuList)
Endwhile
```

Fig. 1.5: Simple Tabu Search Algorithms [7]

Search strategies for CS based on TS are reported in [31]. The study uses steepest descent-modest ascent strategy to determine the global minimum of a function. In the minimization process, variations in dihedral angles are accounted for with minute adjustments in angles and bond distances. The exploitation process was conducted by applying non-redundant internal coordinates and varied only the dihedral angles in order that the search would not be confined in local minimum. The approach was also successfully tested on molecules with various sizes.

3. Conclusion and Future Works

This paper provides various metaheuristic approaches that have been applied to various common CS problems over the last few decades. It covers an overview of five popularly used metaheuristic approaches with their respective mechanisms on solving all kind of CS problems. Future reviews can look into other algorithmic methods to solve the CS problems.

REFERENCES

[1] Setzer WN. Conformational analysis of thioether musks using density functional theory. International journal of molecular sciences 2009; 10: 3488-3501

[2] Chang C, Gilson MK. Tork: Conformational analysis method for molecules and complexes. Journal of computational chemistry 2003; 24: 1987-1998

[3] Dorn M, e Silva MB, Buriol LS, Lamb LC. Three-Dimensional Protein Structure Prediction: Methods and Computational Strategies. Computational Biology and Chemistry 2014

[4] Shashi KD, Katiyar V, Katiyar C. A state of art review on application of nature inspired optimization algorithms in protein-ligand docking. Indian Journal of Biomechanics: Special Issue 2009; 3: 7-8

[5] Lee K, Czaplewski C, Kim S, Lee J. An efficient molecular docking using conformational space annealing. Journal of computational chemistry 2005; 26: 78-87

[6] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery 2004; 3: 935-949

[7] Blum C, Roli A. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. ACM Computing Surveys (CSUR) 2003; 35: 268-308

[8] Lim SM, Sulaiman MN, Md Sultan AB, Mustapha N, Ario Tejo B. Real coded genetic algorithm (RCGA): a new RCGA mutator called Scale Truncated Pareto Mutation. Journal of Theoretical and Applied Information Technology 2014; 60 (2): 245-253

[9] Lim SM, Sulaiman MN, Md Sultan AB, Mustapha N, Ario Tejo B. A new real coded genetic algorithm crossover: Rayleigh Crossover. Journal of Theoretical and Applied Information Technology 2014; 62 (1): 262-268

[10] Lim SM, Sulaiman MN, Md Sultan AB, Mustapha N, Ario Tejo B. New Real Coded Genetic Algorithm Operators for Minimization of Molecular Potential Energy Function. Applied Artificial Intelligence 2015 (ISI-indexed, Impact Factor: 0.563); 29(10): 979-991

[11] Lim SM, Sulaiman MN, Mustapha N, Md Sultan AB. Parameter settings for new generational genetic algorithms for solving globl optimization problems. Journal of Computer Science 2015; DOI: 10.3844/jcssp.2015

[12] Lim SM, Sulaiman MN, Md Sultan AB, Mustapha N. New Crossover and Mutation Operators of Real Coded Genetic Algorithms for Optimization of Molecular Structures. Annals of Mathematics and Artificial Intelligence 2016; Submitted

[13] Paenke I, Branke J, Jin Y. On the influence of phenotype plasticity on genotype diversity 2007: 33-40

[14] Wilke DN, Kok S, Groenwold AA. Comparison of linear and classical velocity update rules in particle swarm optimization: notes on diversity. International Journal for Numerical Methods in Engineering 2007; 70: 962-984

[15] Watts KS, Dalal P, Murphy RB, Sherman W, Friesner RA, Shelley JC. ConfGen: a conformational search method for efficient generation of bioactive conformers. Journal of chemical information and modeling 2010; 50: 534-546

[16] Pelikan M, Goldberg DE, Lobo FG. A survey of optimization by building and using probabilistic models. Computational optimization and applications 2002; 21: 5-20

[17] Baluja S. . Population-based incremental learning.a method for integrating genetic search based function optimization and competitive learning 1994

[18] Larranaga P. A review on estimation of distribution algorithms. In: Estimation of distribution algorithms: Springer. 2002: 57-100

[19] Long SM, Tran TT, Adams P, Darwen P, Smythe ML. Conformational searching using a population-based incremental learning algorithm. Journal of computational chemistry 2011; 32: 1541-1549

[20] Yang X. Nature-inspired metaheuristic algorithms: Luniver press. 2010

[21] Christen M, Van Gunsteren WF. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review. Journal of computational chemistry 2008; 29: 157-166

[22] Ozkan SB, Meirovitch H. Conformational search of peptides and proteins: Monte Carlo minimization with an adaptive bias method applied to the heptapeptide deltorphin. Journal of computational chemistry 2004; 25: 565-572

[23] Frausto-Solis J, Soberon-Mainero X, Liñán-García E. MultiQuenching annealing algorithm for protein folding problem. In: MICAI 2009: Advances in Artificial Intelligence: Springer. 2009: 578-589

[24] Frausto-Solis J, Román E, Romero D, Soberon X, Liñán-García E. Analytically tuned simulated annealing applied to the protein folding problem. In: Computational Science–ICCS 2007: Springer. 2007: 370-377

[25] Lučić P. . Modeling transportation problems using concepts of swarm intelligence and soft computing 2002

[26] Bahamish HAA, Abdullah R, Salam RA. Protein conformational search using bees algorithm 2008: 911-916

[27] Vengadesan K, Gautham N. A new conformational search technique and its applications. Curr Sci 2005; 88: 1759-1770

[28] Zhan L, Chen JZ, Liu W. Conformational study of met-enkephalin based on the ECEPP force fields. Biophysical journal 2006; 91: 2399-2404

[29] Morales LB, Garduño–Juárez R, Aguilar–Alvarado J, Riveros–Castro F. A parallel tabu search for conformational energy optimization of oligopeptides. Journal of Computational Chemistry 2000; 21: 147-156

[30] Eisenmenger F, Hansmann UH, Hayryan S, Hu C. An enhanced version of SMMP—open-source software package for simulation of proteins. Computer Physics Communications 2006; 174: 422-429

[31] Stepanenko S, Engels B. Tabu Search Based Strategies for Conformational Search†. The Journal of Physical Chemistry A 2009; 113: 11699-11705

[32] Michalewicz Z, Fogel DB. Why Are Some Problems Difficult to Solve?. In: How to Solve It: Modern Heuristics: Springer. 2004: 11-30

[33] Pham D, Ghanbarzadeh A, Koc E, Otri S, Rahim S, Zaidi M. The bees algorithm-a novel tool for complex optimisation problems 2006: 454-459

Table 1.1: Why are Some Problems Difficult to Solve? [32]

Items	Descriptions
1	A function, *f* may be noisy or varies with time; therefore, an entire series of solutions are required instead of a single solution.
2	It is unfeasible to perform an exhaustive search for the best answer from a large pool of possible solutions.
3	It is a challenging task to locate an optimum solution due to the possible solutions that are so heavily constrained.
4	Only the solution to a *simplification model* of the problem is solved. For example, Travelling Salesman Problem (TSP) is modeled as a graph; the nodes correspond to cities and the edges are the distances between the cities. Several important parameters are being omitted such as petrol prices, time of the day, weather, traffic etc.

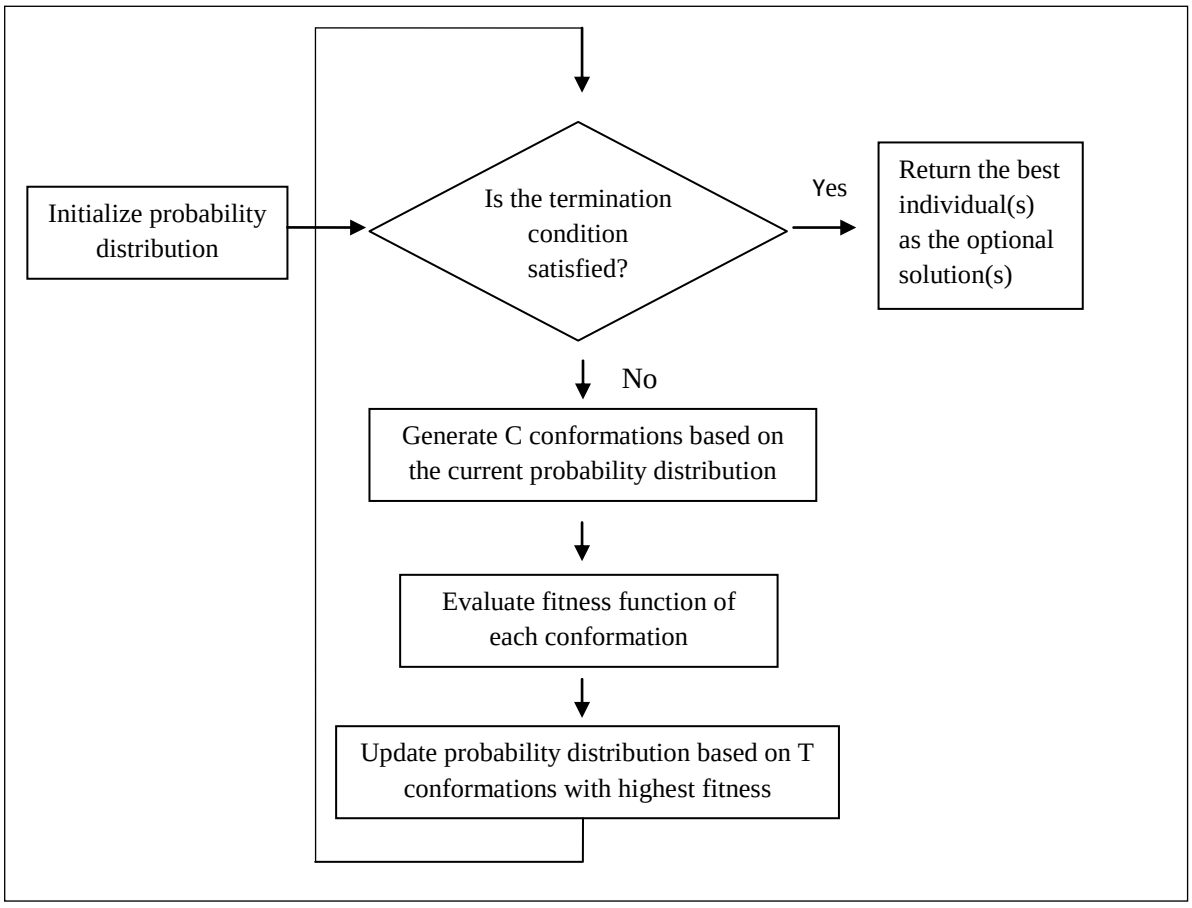

Fig 1.1: Population-Based Incremental Learning Algorithms [19]

1. Initialize population with random solutions.
2. Evaluate fitness of the population.
3. While (stopping criterion not met)
   // forming new bee population.
4. Select elite bees.
5. Select sites for neighborhood search.
6. Recruit bees around selected sites and evaluate fitness.
7. Select the fittest bee from each site.
8. Assign remaining bees to search randomly and evaluate their fitness.
9. End while.

Fig 1.4: Bees Algorithms [33]