A Kinematics Based Evolutionary Approach for Molecular Conformational Search

Athina N. Brintaki[1], Susana K. Lai-Yuen[1] and Ioannis K. Nikolos[2]

[1]University of South Florida, abrintak@mail.usf.edu, laiyuen@eng.usf.edu
[2]Technical University of Crete, jnikolo@dpem.tuc.gr

ABSTRACT

This paper presents a new approach called Kinematics Differential Evolution (kDE) to model flexible biological molecules of different type and size through the rapid identification of low-energy molecular conformations. One of the main benefits of our proposed methodology is that a population of low-energy solutions is provided for each tested molecular structure compared with a single low-energy solution usually obtained by traditional molecular modeling approaches. This population includes a number of different molecular conformations that attain similar low energy values and hence correspond to an energetic state with high probability of occurrence. Results show that the kDE model provides a set of molecular structures comparable to those obtained by traditional molecular dynamics.

Keywords: bio-CAD, kinematics, differential evolution, energy minimization.
DOI: 10.3722/cadaps.2011.23-36

1 INTRODUCTION

Bionanotechnology research has been revolutionizing many important scientific fields ranging from engineering to medicine. This molecular scale technology has the potential of developing smaller and more efficient devices with new capabilities that include diagnosis and treatment of diseases. To understand the biological systems at the molecular level, there is a need to visualize the interactions between molecules during the design stage so that fully functional nanoscale products can be evaluated prior to actual fabrication. A main requirement for enabling the visualization of molecular interactions is the understanding and effective modeling of molecules' behavior. Molecules are very flexible and can adopt many molecular conformations (or shapes). The major challenge in modeling flexible molecules (or their conformations) lies on the exponential explosion in computational complexity as the molecular size increases and a large number of degrees of freedom (DOF) must be considered to represent the molecules' flexibility.

An appropriate optimization method is essential for directing the search towards low-energy (stable) molecular conformations. Locating the global or even local low-energy conformations for large-scale molecules such as proteins is a highly challenging task given the multitude of potential conformations. A range of stochastic search algorithms have been proposed that include: Cartesian coordinate stochastic search [27], internal coordinate Monte Carlo search [7], purely random searches [12], [16], Monte Carlo search methods that utilize biased sampling [25], molecular dynamics,

simulated annealing [18], [33], conformational space annealing, quantum annealing [11], [21], and genetic algorithms [17], [30], [32], [35]. In addition, a limited number of simplified systematic search algorithms such as tree search [15], [20], and eigenvector analysis like LMOD [19] and TORK [8] methods have been proposed in an effort to address the combinatorial explosion problem by limiting regions of the explored conformational space. However, it is difficult to identify an optimum method for the molecular conformational search problem as the criteria and objectives by which the methods were assessed vary considerably amongst the studies. Nevertheless, when stochastic methods are compared with systematic approaches, the stochastic search methods tend to be more efficient in identifying low-energy conformations early in the search but are less efficient in finding the global minimum structure. In contrast, full systematic searches of conformational space suffer from the combinatorial explosion problem that makes them unsuitable for large-scale molecules.

Stochastic methods appear to be extremely useful for identifying relatively low-energy conformation(s) of large molecules with minimal computational expense. Current stochastic molecular conformational search approaches like molecular dynamics (MD) utilize Force Fields such as CHARMM to calculate the molecular energy. Then, an optimization algorithm like Monte Carlo (MC) and simulated annealing is used to direct the search for low-energy molecular conformations in small time steps. This results in a more accurate but slow progress towards the search of a low-energy molecular conformation. MD methods require significant computer resources and are primarily used to study the behavior of conformations at the vicinity of an optimum state [32]. Traditional MD packages such as CHARMM and NAMD start from a fixed biological structure, which is later minimized to a low-energy state. At the end of the MD energy minimization, a potential low-energy molecular state is predicted for researchers to test during molecular design. This predicted molecular conformation is then incrementally heated to the desired temperature (room or body temperature), followed by an equilibration process to bring the predicted solution closer to its natural equilibrium state. The objective is to identify solutions in an energetic state with high probability of occurrence.

A few general observations amongst the stochastic methods are that evolutionary-based methods tend to outperform simulated annealing [22], [31] and MC [10] whereas the quantum annealing method outperforms both simulated annealing and genetic algorithms [21] for small molecules containing up to 100 atoms. To address the molecular conformational search problem for large-scale molecules like proteins, this research work focuses on evolutionary computation to direct the search towards low-energy minima.

The optimization of molecular geometry through molecular mechanics was one of the very first applications of evolutionary algorithms (EAs) in chemistry. EAs have shown good results in large-scale problems where other methods have struggled [32], [35]. In addition, EAs are parallel algorithms by nature, can handle both continuous and discrete variables, and are easily tailored to the problem under consideration. Many modifications to EAs have been proposed to improve the solution quality and to speed convergence. One of the most successful EAs for solving real-valued energy functions is Differential Evolution (DE), initially proposed by [26], [28]. DE is a population based stochastic function minimizer that adds the weighted difference between two individual vectors to a third vector (or donor). Recently, DE has been used for addressing the problem of flexible ligand docking where the algorithm showed robustness and remarkable performance in terms of convergence speed [9], [29], [30], [35]. Chong and Tremayne [9] presented a new DE algorithm based on Cultural Evolution concepts called CDE to study the structural search for small molecules. Bitello and Lopes [2] used a DE algorithm to solve the protein folding problem. Their approach was consistent in finding the global minimum for relatively small-sized proteins of up to 64 amino acids and performed better than a classic GA.

This paper presents a Kinematics Differential Evolution (kDE) approach to model flexible biological molecules of different type and size, and to rapidly identify a population of low-energy molecular conformations. Each molecular structure is represented as a highly articulated body to enable rapid modeling of flexibility while a differential evolution algorithm is used as an optimization tool to direct the search towards low-energy minima. The proposed kDE approach is compared against traditional MD packages such as CHARMM. The remaining of the paper is organized as follows: Section 2 introduces the basis of the differential evolution algorithm used in the proposed kDE approach. Section 3 presents the proposed kDE method. Computer implementation and results are provided in Section 4 followed by the conclusions in Section 5.

2 DIFFERENTIAL EVOLUTION (DE) ALGORITHM

A Differential Evolution (DE) algorithm based on the concepts by [13], [26], [28] is used to direct the search towards low energy molecular conformations. The DE algorithm has been presented in our previous work [6], [23], [24] to address the problem of UAV path planning. The DE algorithm is a relatively simple to implement evolutionary algorithm that has shown better convergence performance compared to other EAs. Differential Evolution embodies a type of evolutionary strategy (ES) especially formed to deal with continuous optimization problems often encountered in engineering design.

The classic DE algorithm evolves a fixed population size that consists of randomly initialized candidate problem solutions (population members or chromosomes). After initializing the population, an iterative process starts to direct the search towards more fit population members. At each iteration (or generation), a new population of candidate solutions is produced until a stopping condition is satisfied. At each generation, each element (member) of the population can be replaced with a new generated one. The new element is a linear combination between a randomly selected population member and a difference between two other randomly selected members. Below is the analytical description of the algorithm:

Given an objective function:

$$F_{objective}(X): R^{n_{param}} \rightarrow R \tag{2.1}$$

the goal is to minimize the objective function value by optimizing the values of its parameters (design variables) as follows:

$$X = (x_1, x_2, ..., x_{n_{param}}), \quad x_j \in R \tag{2.2}$$

where X denotes the vector composed of n_{param} objective function parameters (design variables). The design variables x_j take values between the specific upper and lower bounds, x_j^U and x_j^L, respectively:

$$x_j^L \le x_j \le x_j^U, j = 1, ..., n_{param} \tag{2.3}$$

The DE algorithm implements real-number encoding for the design variables. Often, the only information available are the boundaries of the parameters. Hence, to obtain a starting point for the algorithm, we initialize the population by randomly assigning values to the design variables within their boundaries as follows:

$$x_{i,j}^0 = r(x_j^U - x_j^L) + x_j^L, i = 1, ..., n_{pop}, \; j = 1, ..., n_{param} \tag{2.4}$$

where $x_{i,j}^0$ is the j^{th} design parameter of the i^{th} population member of generation 0, r is a uniformly distributed random value within the range [0, 1].

The DE mutation operator is based on a triplet of randomly selected individuals that are different from each other. A new parameter vector is generated by adding the weighted difference vector between the two members of the triplet to the third one (the donor). In this way, a perturbed individual is generated. The perturbed individual and the initial population member are then subjected to a crossover operation to generate the final candidate solution as follows:

$$x_{i,j}^{'(G+1)} = \begin{cases} x_{c_i,j}^{(G)} + F(x_{A_i,j}^{(G)} - x_{B_i,j}^{(G)}), & if \quad (r \le C_r \vee j = k) \vee j = 1, ..., n_{param} \\ x_{i,j}^{(G)} & o/w \end{cases} \tag{2.5}$$

$$\left.\begin{array}{c} i = 1, ..., n_{pop}, \quad j = 1, ..., n_{param} \\ A_i \in [1, ..., n_{pop}], \quad B_i \in [1, ..., n_{pop}], \quad C_i \in [1, ..., n_{pop}] \\ A_i \ne B_i \ne C_i \ne i \\ C_r \in [0,1], \quad F \in [0,1+], \quad r \in [0,1] \end{array}\right\} \tag{2.6}$$

where $x_{c_i,j}^{(G)}$ is called the "donor", G is the current generation, and k is a random integer within $[1, n_{param}]$, chosen once for all members of the population. The random number r is seeded for every gene of each chromosome. F and C_r are DE control parameters, which remain constant during the search process and affect the convergence behavior and robustness of the algorithm. Their values also depend on the objective function, the characteristics of the problem, and the population size.

The population for the next generation is selected between the current population and the final candidates. If each candidate vector is more fit than the corresponding current one, the new vector replaces the vector against which it was compared. The DE selection scheme for a minimization problem is described as follows:

$$X_i^{(G+1)} = \begin{cases} X_i'^{(G+1)}, & if \quad F_{obj}(X_i'^{(G+1)}) \leq F_{obj}(X_i^{(G)}) \\ X_i^{(G)} & o/w \end{cases} \tag{2.7}$$

In this research work, the improved scheme by [13] for determining the donor for the mutation operation is used to accelerate the convergence rate. In this scheme, the donor is randomly selected (with uniform distribution) from the region within the "hyper triangle", formed by the three members of the triplet. The donor comprises the local information of all the members of the triplet, providing a better starting-point for the mutation operation that result in a better distribution of the trial-vectors. One of the major characteristics of DE algorithm is the deterministic selection procedure, utilizing a comparison between each member of the current population with its candidate replacement. This deterministic selection procedure introduces elitism and results in a high selection pressure, which considerably accelerates the convergence rate of the optimization procedure. The mutation and crossover operators retain the diversity of the population, with a proper selection of their control parameters, avoiding premature convergence to local optima. DE has the advantage of requiring relatively small population sizes compared to other EAs. However, in this work, relatively large populations have been used to provide a wide range of different but equivalent optimal solutions, which correspond to similar values of the fitness function.

3 THE PROPOSED KINEMATICS DIFFERENTIAL EVOLUTION (KDE) APPROACH

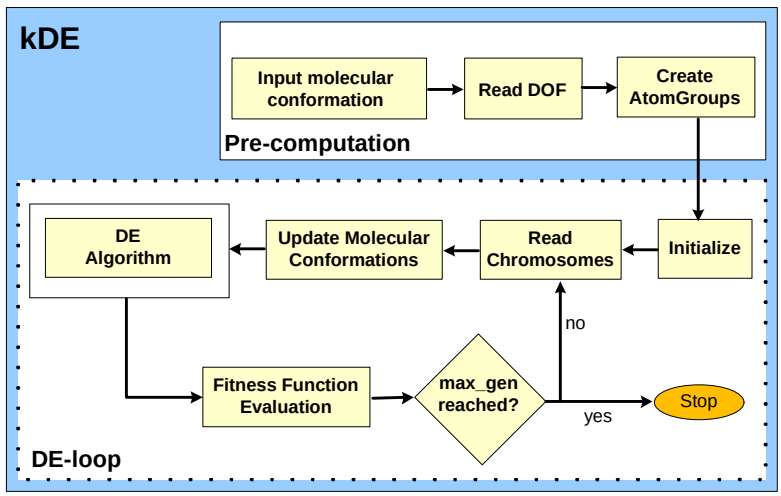


Fig. 1: Overview of the proposed kDE model.

Fig. 1 shows the overview of the proposed Kinematics-based Differential Evolution (kDE) model that consists of two modules: the pre-computation and the DE-loop. During pre-computation, a molecule is

represented as a highly articulated body that can adopt different conformations. The kDE model starts with a random molecular conformation where the DOF of the molecular structure are defined to form groups of atoms. During the DE-loop, a differential evolution algorithm is used to direct the search towards low-energy molecular conformations and to provide a population of alternative low-energy solutions.

3.1 Pre-computation Module

At the pre-computation module, a geometric interpretation of the underlying chemical information is performed to represent the molecules' flexibility mechanism. Each molecular structure is represented as a highly articulated body able to adopt different molecular conformations while searching for a low-energy molecular state.

3.1.1 A Geometric Interpretation of Molecular Mechanics

Molecules are highly flexible bodies that tend to adopt different conformations until they reach a stable molecular state that is described by the minimum possible internal energy. This energy is a complex function composed of different energy factors that depict the interactions between the bonded and non-bonded atoms within the molecular topology. As shown by Equation (8), one of the major energy contributors are the non-bonded atoms' interactions as depicted by the van der Waals (VDW) potential and electrostatic forces, and discussed in our previous work [3],[4]:

$$E_{nb} = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \sum_{i=1}^{n}\sum_{j=1}^{n} k \cdot \frac{q_i \cdot q_j}{D \cdot r_{ij}} \tag{3.1}$$

where B_{ij} and A_{ij} are the VDW repulsion and attraction parameters, respectively; r_{ij} is the distance between every exclusive non-bonded atom pair i and j; q_i and q_j are atom's i and j charges, respectively; k is a factor to convert the electrostatic potential into kcal/mol and is defined as 322.0 kcal/mol; and D is the dielectric constant, which is defined as a function of the intra-atomic distances and is defined as in [34], [35] as $D = 4 \cdot r_{ij}$.

From a geometric point of view, a molecule can be modeled as an articulated body with at least six degrees of freedom (DOF): three translational and three rotational. Some of the chemical bonds within a molecular structure have the ability to rotate along their own axis by a torsion angle θ_i that accounts for an additional DOF as shown in Fig. 2. Changes in bond angles and bond lengths also influence the molecular conformation but will be considered constant in this work as they do not change the molecular conformation significantly. Therefore, a molecular conformation is defined in this work as changes in the torsion bonds θ_i.

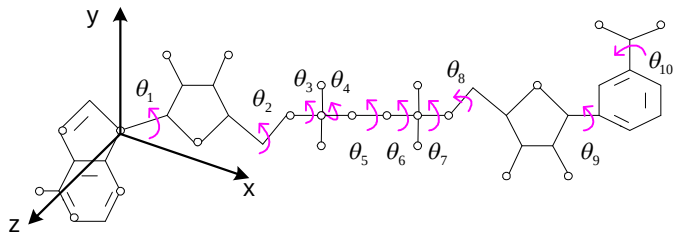
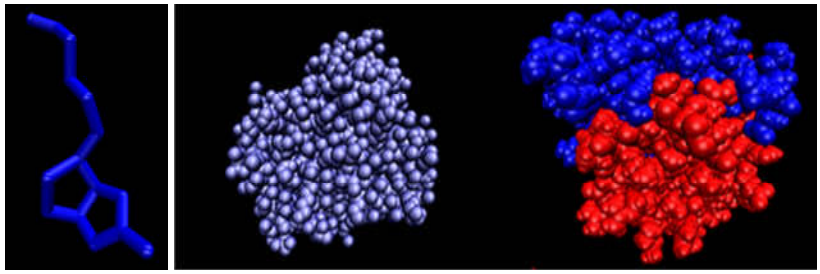


Fig. 2: Example of a small molecule as an articulated body.

3.1.2 Molecular Modeling

A ligand or drug-like molecule is a small molecular structure that usually consists of at most 50 atoms. Ligand molecules may contain rings of atoms as shown in Fig. 3(a). Rings are considered rigid during modeling as the location of the ring atoms does not change with respect to each other. Therefore,

torsion can be assumed everywhere within a ligand's topology except within the rings and within double- and triple-bonded atoms, which correspond to stronger chemical bonds.



(a) Ligand (b) Proteins with one and two chains

Fig. 3: Examples of different ligand and protein molecules.

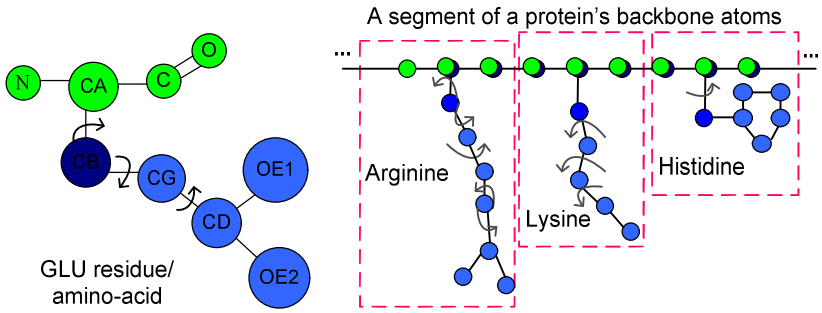
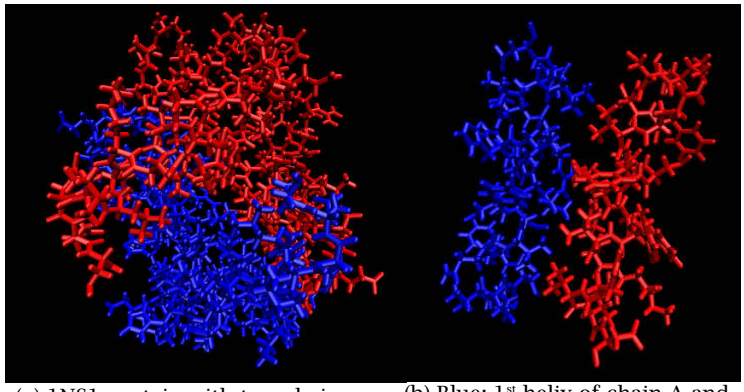


Fig. 4: Degrees of freedom within a hypothetical protein segment.

Proteins are chains of smaller molecular entities called amino acids. A protein can be considered as a polypeptide chain characterized by the amino acid sequence along the chain in order as analyzed in our previous work [4]. A protein molecule may contain one chain or multiple chains as shown in Fig. 3(b). In contrast to ligand modeling, proteins consist of hundreds or thousands of atoms with hundreds or even thousands of DOF. Torsion changes can occur anywhere within a protein's topology, except within rings and double- and triple-bonded atoms. In this work, torsions within a protein's backbone (chain) are not considered. Changes in torsions are assumed only between the central carbon atom of a protein's backbone (CA) and a side chain atom (CB) or within the side chain atoms as shown in Fig. 4. Furthermore, torsions at the end of each side chain are neglected (i.e., the bond between CD and OE1 atoms in Fig. 4) since they do not contribute significantly to the molecular conformation.

When a protein molecule contains more than one chain as shown in Fig. 5(a), an artificial rigid bond is introduced between the closest residue-pair of the chains. This artificial bond is used to simulate the electrostatic forces that keep the chains in contact and should be created in the least flexible region of the protein to avoid the risk of breaking its structure. For example, the least flexible region in the 1NS1 protein shown in Fig. 5(a) is the area between the first helixes of the two chains as shown in Fig. 5(b). However, an artificial bond should not be placed arbitrarily between any residue-pair within the 1st helices but between the closest possible residue-pair as shown by the circle in Fig. 6. Moreover, the selected residue pair should have the same polarity. In other words, the closest residues that are both hydrophobic, polar or ionized would be good candidates for placing an artificial rigid bond.



(a) 1NS1 protein with two chains (b) Blue: 1st helix of chain A and red: 1st helix of chain B

Fig. 5: Structure representation of protein with PDB ID: 1NS1 displayed using VMD molecular graphics software [14].

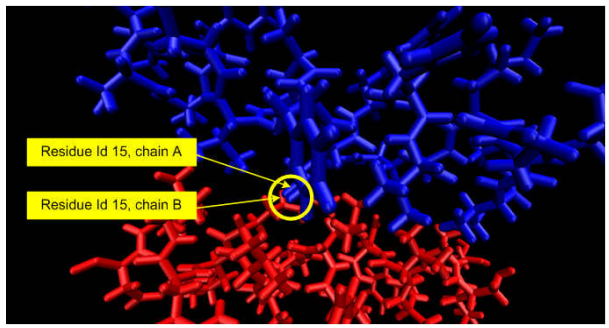
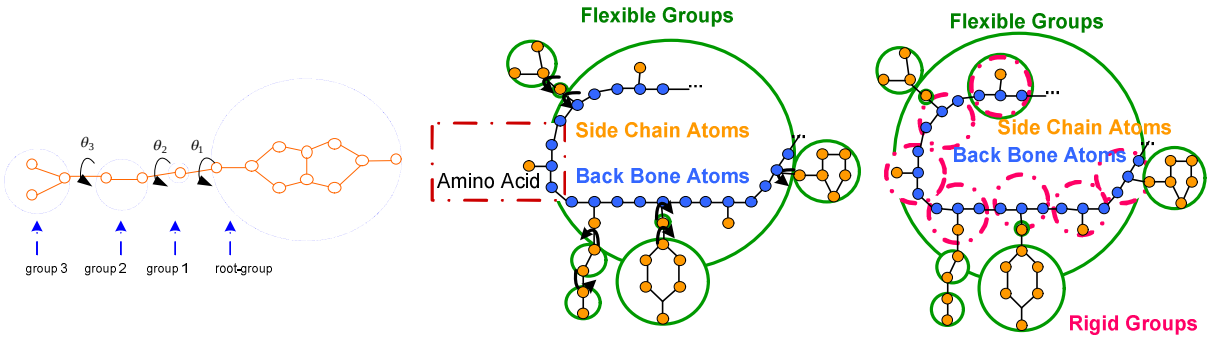


Fig. 6: Closest residue-pair between the first helices of the two chains of the 1NS1 protein.

3.1.3 Simplifying the Molecular Representation

As shown in Fig. 7, a flexible biological molecule is modeled as an articulated body with the torsion bonds $\theta_i \in [0,2\pi)$, accounting for the DOF. To reduce the computational complexity of a molecular structure, atoms of a molecule are clustered into AtomGroups based on the approach by [36]. Based on the location of the torsion bonds, atoms are clustered into AtomGroups. In other words, all the atoms within an AtomGroup are connected by rigid bonds while AtomGroups are connected by torsion bonds as shown in Fig. 7(a) and Fig. 7(b).

As shown in Fig. 7(c), if the tested molecular structure is a protein molecule, then an additional step within the kDE algorithm is performed to split the backbone atom cluster (or clusters in the case of multiple chain proteins) into smaller AtomGroups. The purpose of this additional step is to decrease the computational time for updating the atoms' positions after torsion changes have occurred. The size of these AtomGroups is based on a threshold defined by the maximum number of atoms allowed within each atom cluster. By splitting a backbone cluster, a flexible AtomGroup (i.e., the green-pink sphere in Fig. 7(c)) is obtained along with a number of rigid AtomGroups (i.e., the six pink/dashed-line spheres shown in Fig. 7(c)). The splitting of the backbone group eliminates calculations within and between the rigid groups since the atomic distances remain unchanged between and within these groups.



(a) Atom groups definition in a ligand

(b) Two types of atom clustering in a protein's segment

(c) Splitting the atom cluster of a protein's backbone

Fig. 7: Simplifying the representation for a ligand and a protein.

3.2 DE-loop Module

Once the different AtomGroups have been formed in the pre-computation module, the kDE model employs the DE algorithm presented in Section 2 to direct the search towards low-energy molecular conformations. As shown in Fig. 1, an *initialization* file is input to the DE-loop. This file denotes the total number of design variables along with their upper and lower bounds and the maximum and minimum expected fitness function values. For each molecule tested, the *initialization* file is input to the kDE only once at the beginning of the DE-loop.

3.2.1 DE Basic Components

Two steps are required in the DE loop: formulate the chromosome structure and define the fitness function. Each chromosome structure through the defined genotype represents a candidate solution to the problem under consideration, whereas the chromosome genes represent the design variables. Hence, to direct the search towards low-energy molecular conformations, the chromosome for the proposed kDE model should represent a candidate molecular conformation. The simplest possible chromosome structure for describing a molecular conformation is to consider each gene to be a degree of freedom or in our case, a torsion bond angle θ_i as shown in Fig. 8:

Fig. 8: The proposed chromosome structure.

The fitness function (*ff*) plays the role of the evaluation criterion for each candidate solution. In this work, to evaluate the fitness of each candidate chromosome (molecular conformation), we use the total non-bonded atoms' potential as described by the VDW and electrostatic forces as follows:

$$ff = E_{nb} = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \sum_{i=1}^{n}\sum_{j=1}^{n} 322 \cdot \frac{q_i \cdot q_j}{4 \cdot r_{ij}^{2}} \tag{3.2}$$

where B_{ij}, A_{ij}, r_{ij}, q_i and q_j are defined as in Section 3.1.1.

3.2.2 DE-loop Description

At the beginning of the DE-loop, the chromosome files are input to the kDE model to define a

population of candidate problem solutions (molecular conformations). Each population member, corresponding to a candidate molecular conformation, is described as a function of the torsion bond angles θ_i as shown in Fig. 8. The DE algorithm evolves a fixed population size (*popsize*) composed of candidate problem solutions (population members or chromosomes), randomly initialized. After initializing the population, an iterative process starts to direct the search towards better fitted population members or stable molecular solutions. At each generation (iteration), a new population of candidate solutions (conformations) is produced until a stopping criterion is satisfied. In this work, the termination criterion is the maximum allowed number of generations performed (*maxgen*). At each generation, each population member (candidate molecular conformation) can be replaced with a new generated one. The new member is a linear combination between a randomly selected member (the donor) and the difference between two other randomly selected members. Genetic operators (mutation, crossover, and selection) are applied to provide the next generation of better fitted candidate problem solutions (molecular conformations).

4 COMPUTER IMPLEMENTATION AND RESULTS

The presented method and algorithms have been implemented on a dual 3.0 GHz CPU workstation using Visual C++ and Visual Basic programming languages. Different molecules with different number of atoms, chains, and DOF have been tested with the proposed kDE approach. The example molecules were obtained from the Protein Data Bank (PDB) [1] with PDB IDs as follows: 1DO3 and 1NS1. Fig. 9 shows some of the tested molecules, graphically displayed using the VMD package [14].

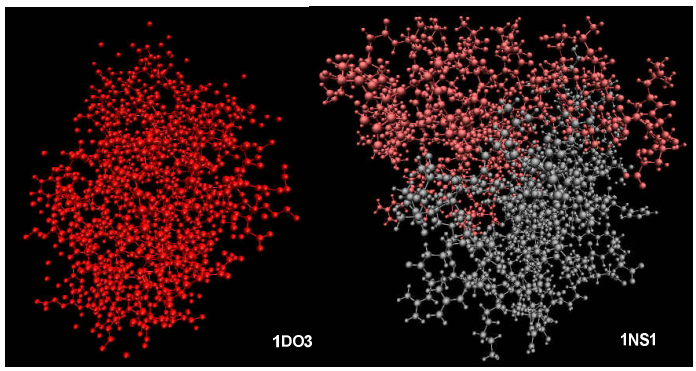


Fig. 9: Example of protein molecules tested with the proposed kDE approach.

The termination criterion for the 1DO3 protein was set to *maxgen* = 500 generations performed and *popsize* = 100 candidate molecular conformations (population members) considered in each generation. For the 1NS1 protein, given the large number of DOF considered in each experimental scenario, *maxgen* was set to 600 generations and 300 population members used as *popsize*. Finally, the DE's control parameters used in all experiments were F = 0.6 for the mutation parameter, and Cr = 0.45 for the crossover probability. Tab. 1 shows a representative list of the performance analysis for the proposed kDE approach on two example protein molecules. As shown in Tab. 1, the first column indicates the PDB IDs for the tested molecules. The second column specifies the number of atoms within each molecular structure and the third column shows the DOF considered in each scenario. The pre-selected DOF for each experiment are the chemically-allowed DOF for the protein except those in the protein's backbone, which are not considered in this work.

Proteins	Number Atoms	DOF	E_nmr (kcal/mol)	E_kDE (kcal/mol)	Conv. Gener.	T (s)
1DO3	2,466	9	-876.33	-2,401.89	189/500	2.55
		22		-2,407.06	182/500	2.62
		36		-2,404.90	211/500	2.58
1NS1	2,342	42	-1126.35	-4,363.91	156/600	2.3
		58		-4,364.77	168/600	2.27
		98		-4,362.78	138/600	2.26

Tab. 1: Performance results for the proposed kDE model.

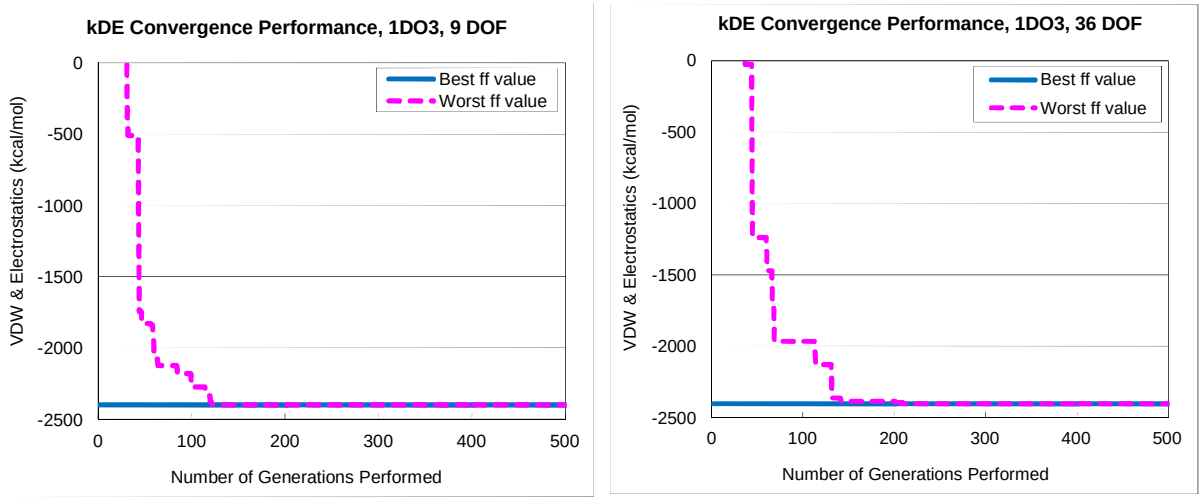
To evaluate the kDE approach, we compared our results with the NMR structures published in the Protein Data Bank [1]. Using Eqn. (3.1), the non-bonded intra-molecular energy for each NMR structure (*E_nmr*) was calculated and compared against the predicted non-bonded internal energy (*E_kDE*) by the proposed kDE algorithm. As shown in Tab. 1, the kDE approach succeeded to converge to a much smaller energy value compared to the corresponding energy of the NMR structure for all the performed experiments. It was observed that when the electrostatic potential is included into the energy calculation, the predicted energy values are significantly smaller compared with those obtained by using only the VDW interactions [5]. This phenomenon occurs since all the incorporated intra-molecular energy terms (i.e., VDW and electrostatic potential) are an approximation of a molecule's potential energy and not the molecule's free energy, which requires entropy calculations, among other.

RMSD	Molecules					
	1DO3			1NS1		
min	**0.024**	**0.065**	**0.159**	**0.035**	**0.036**	**0.023**
max	9.000	22.000	36.000	42.000	58.000	98.000
average	0.451	0.868	1.414	1.111	1.481	2.384
DOF	9	22	36	42	58	98

Tab. 2: RMSD performance of the proposed kDE model.

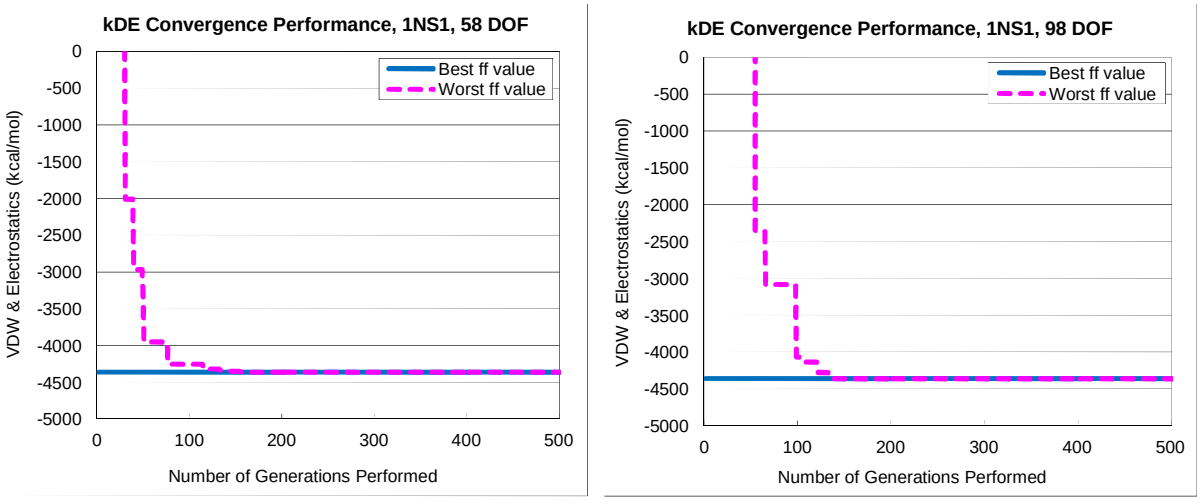
To verify the structure of the predicted molecular solutions obtained with the kDE approach, the solutions were evaluated with the corresponding NMR structures using the Root Mean Square Deviation (RMSD) in Angstroms. Generally, RMSD values below or near 3.0 Å indicate closer resemblance between observed and predicted structures. As shown in Tab. 2, the kDE algorithm succeeded to identify structures with average RMSD values in the range of 0.451 to 2.384 for the tested proteins. This indicates that the kDE approach predicted solutions that are structurally very similar to the corresponding NMR structures.

The sixth column (*Conv.Gener.*) in Tab. 1 indicates the generation where the proposed kDE algorithm converged. Fig. 10 and Fig. 11 show the convergence performance of the kDE approach considering different DOF for 1DO3 and 1NS1 proteins, respectively. As shown in these figures, the proposed algorithm demonstrates a good convergence performance. This is very important given that one of the main drawbacks in an evolutionary-based algorithm is the convergence uncertainty.



(a) 1DO3 with 9 DOF (b) 1DO3 with 36 DOF

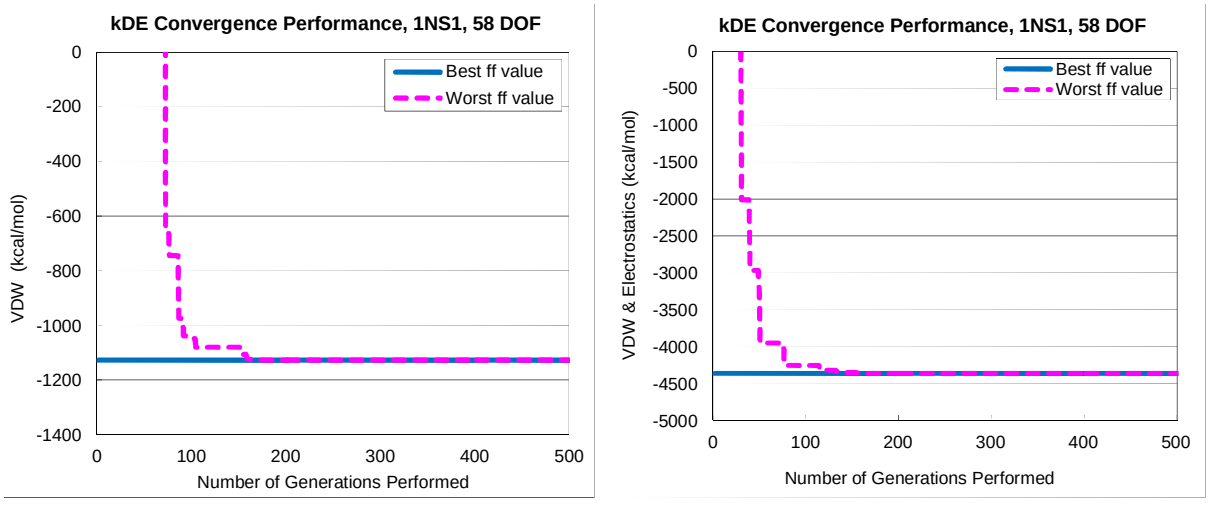
Fig. 10: Convergence performance of the proposed kDE algorithm for 1DO3 protein considering different DOF.



(a) 1NS1 with 58 DOF (b) 1NS1 with 98 DOF

Fig. 11: Convergence performance of the proposed kDE algorithm for 1NS1 protein considering different DOF.

The convergence performance of the kDE algorithm was tested when both the VDW and electrostatic interactions were considered and when only the VDW forces were considered as shown in Fig. 12. As shown in this figure, when the electrostatic potential becomes part of the energy minimization process, the kDE algorithm identifies the low-energy minima earlier in the search and converges slightly faster to a much smaller energy value. In addition, when both the VDW and electrostatic terms are included into the energy function, the RMSD values for the predicted structures are much smaller and hence, closer to their corresponding NMR structure.



(a) 1NS1, 58 DOF with VDW energy only

(b) 1NS1, 58 DOF with both VDW and electrostatic energy

Fig. 12: Convergence performance of the proposed kDE algorithm for 1NS1 protein with different energy terms.

One of the main benefits of the kDE algorithm is that it outputs a final population of low-energy solutions for each tested molecule. This final population contains a large number of different molecular conformations that attain similar low-energy values for researchers to test during molecular design. The more solutions identified in a predicted energetic state, the higher probability of occurrence this energetic state has. Hence, the kDE succeeded in predicting energetic states with high occurrence probability. The low-energy solutions identified by the kDE algorithm can then be structurally clustered to identify those closest to their corresponding NMR structure.

	CHARMM				**kDE**			
Proteins	1NS1		1DO3		1NS1		1DO3	
Structure	**NMR**	**Minimized**	**NMR**	**Minimized**	**NMR**	**Minimized**	**NMR**	**Minimized**
Enb	-2,996.07	-6,494.56	-2,315.11	-4,616.84	-1,126.35	-4,362.78	-876.33	-2,404.89
ΔEnb	**-3,498.49**		**-2,301.73**		**-3,236.43**		**-1,528.56**	
RMSD	**3.50**		**1.46**		**0.02**		**0.16**	

Tab. 3: Comparison of the proposed kDE approach with a traditional MD package.

To evaluate and validate the kDE approach, the predicted results were compared against the CHARMM MD package. For comparison, CHARMM was run under the same conditions as the kDE model with no solvent environment and only considering torsion angles as the pre-selected degrees of freedom. As shown in Tab. 3, although CHARMM and kDE started from a different energetic state and ended-up in a different minimized energetic state, both models have approximately the same change in non-bonded energy. For example, for 1NS1 protein, $\Delta E_{nb,CHARMM} = -3498.49 \cong -3236.43 = \Delta E_{nb,kDE}$. However, the kDE approach outputs solutions that are significantly closer to their corresponding NMR structure compared with the solution predicted by CHARMM (i.e., for 1NS1 protein $RMSD_{kDE} = 0.02 < \ RMSD_{CHARMM} = 3.5$). Additionally, the kDE approach is able to predict a population of alternative low-energy molecular conformations for researchers to test during molecular design, whereas the CHARMM MD package outputs a single molecular conformation.

5 CONCLUSIONS

This paper presented a new Kinematics-based Differential Evolution (kDE) model to effectively identify a population of low-energy molecular conformations for molecular design. The proposed model consists of two modules: the pre-computation and the DE-loop. At the pre-computation module, a molecule is represented as a highly articulated body able to adopt different molecular conformations. At the DE-loop, a differential evolution algorithm is used as a direct search technique towards low-energy molecular conformations. Computer implementation and results demonstrate that the proposed kDE approach rapidly and accurately finds a population of low-energy molecular conformations for structures of different type and size. Results also show that the kDE algorithm attains very good convergence performance while it outputs a population of alternative low-energy solutions for researchers to test during molecular design. A comparison of our proposed kDE approach with CHARMM MD package demonstrates that both methods predict approximately the same change in non-bonded energy, whereas the kDE approach provides structures that are closer to the corresponding NMR structure. The proposed kDE approach can be used in the modeling of flexible molecules for molecular docking and assembly applications.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF) under Grant (CMMI-0841451). The authors would also like to thank Ashley Mullen and Christi Young from the Department of Chemistry at USF for the insightful discussions and their assistance.

REFERENCES

[1] Berman, H. M; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; et al.: The Protein Data Bank, Nucleic Acids Research, 28, 2000, 235-242. doi:10.1093/nar/28.1.235

[2] Bitello, R.; Lopes, H. S.: A Differential Evolution Approach for Protein Folding, Proceedings of the IEEE Symp. on Comp. Int. in Bioinformatics and Comp. Biology, 1-5, 2006.

[3] Brintaki, A. N.; Lai-Yuen, S. K.: BioGeoFilter: A Tool for Identifying Geometrically Feasible Molecular Conformations in Real-Time for Bionanomanufacturing, Transactions of the North American Manufacturing Research Institution of SME, 36, 2008, 153-160.

[4] Brintaki, A. N.; Lai-Yuen, S. K.: eBGF: An Enhanced Geometric Hierarchical Representation for Protein Modeling and Rapid Self-Collision Detection, Computer-Aided Design and Applications, 6(5), 2009, 625-638. doi:10.3722/cadaps.2009.625-638

[5] Brintaki, A. N.; Lai-Yuen, S. K.; Nikolos, I. K.: A Kinematics and Evolutionary Approach for Modeling Molecular Conformations, Proceedings of the Institute of Industrial Engineers (IIE) Annual Conference & Expo, Cancun, Mexico, June 5-9, 2010.

[6] Brintaki, A. N.; Nikolos, I. K.: Coordinated UAV Path Planning Using Differential Evolution, Proc. of the 13th Mediterranean Conference on Control and Automation, IEEE, Cyprus, 2005, 549-556.

[7] Chang, G.; Guida, W. E.; Still, W. C.: An Internal Coordinate Monte Carlo Method for Searching Conformational Space, J. Am. Chem., 44, 1992, 277-290.

[8] Chang, C.-e.; Gilson, M. K.: TORK: Conformational Analysis Method for Molecules and Complexes, J. Comput. Chem., 24, 2003, 1987-1998. doi:10.1002/jcc.10325

[9] Chong, S. Y.; Tremayne, M.: Combined Optimization Using Cultural and Differential Evolution: Application to Crystal Structure Solution From Powder Diffraction Data, Chemistry Communication, Royal Society of Chemistry Journal, 2006, 4078-4080.

[10] Cvijovic, D.; Klinowski, J.: Taboo Search- an Approach to the Multiple Minima Problem, Science, 267, 1995, 664-666. doi:10.1126/science.267.5198.664

[11] Finnila, A.; Gomez, M.; Sebenik, C.; Stenson, C.; Doll, J.: Quantum Annealing: A New Method for Minimizing Multidimensional Functions, Chem. Phys. Lett., 219, 1994, 343.

[12] Foreman, K. W.; Phillips, A. T.; Rosen, J. B.; Gutierrez, D.: Comparing search strategies for Finding Global Optima on Energy landscapes, J. Comput. Chem., 20, 1999, 1527-1532.

[13] Hui-Yuan; F., Lampinen, J.; Dulikravich, G. S.: Improvements to Mutation Donor Formulation of Differential Evolution, Proceedings of EUROGEN Conference on Evolutionary Methods for Design, Optimization and Control, Appl. to Industrial and Societal Problems, CIMNE, Barcelona, 2003.

[14] Humphrey, W.; Dalke, A.; Schulten, K.: VMD-Visual Molecular Dynamics, Journal of Molecular graphics, 14, 1999, 33-38. doi:10.1016/0263-7855(96)00018-5
[15] Izgorodina, E. I.; Lin, C. Y.; Coote, M. L.: Energy-Directed Tree search: An Efficient Systematic Algorithm for Finding the Lowest Energy Conformation of Molecules, J. Chem. Phys., 9, 2007, 2507-2516.
[16] Judson, R. S.; Colvin, M. E.; Meza, J. C., Huffer, A.; Gutierrez, D.: Do Intelligent Configuration Search Techniques Outperform Random Search for Large Molecules? J. Quantum Chem., 44, 1992, 277-290. doi:10.1002/qua.560440214
[17] Judson, R. S.; Jaeger, E. P.; Tresurywala, A. M.; Peterson, M. L.: Genetic Algorithms in Molecular Recognition and Design, J. Comput. Chem., 14, 1993, 1407-1414. doi:10.1002/jcc.540141117
[18] Kirkpatrick, S.; Gelatt, C.D. Jr. and Vecchi, M.P.: Optimization by Simulated Annealing, Science, 220, 1983, 671. doi:10.1126/science.220.4598.671
[19] Kolossvary, I.; Guida, W. C.: Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acylic Alcanes and Cyclic Peptides, J. Am. Chem. Soc., 118, 1996, 5011-5019. doi:10.1021/ja952478m
[20] Lipton, M.; Still, W. C.: The Multiple Minimum Problem in Molecular Modeling. Tree Searching Internal Coordinate Conformational Space, J. Comput. Chem., 9, 1988, 343-355.
[21] Liu, P.; Berne, B. J.: Quantum Path Minimization: An Efficient Method for Global Optimization, J. Chem. Phys., 118, 2003, 2999-3005. doi:10.1063/1.1527919
[22] Meza, J. C.; Martinez, M. L.: Direct search Methods for the Molecular Conformation problem, J. Comput. Chem., 15, 1994, 627-632. doi:10.1002/jcc.540150606
[23] Nikolos, I. K; Brintaki, A. N.; Zografos, E. S: Coordinated UAV Path Planning Using an ANN Assisted Differential Evolution Algorithm, Proceedings of the EUROGEN Conference (Sixth Conference on Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems), Munich, Sep. 12-14, 2005.
[24] Nikolos, I. K; Zografos, E. S.; Brintaki, A. N: UAV Path Planning Using Evolutionary Algorithms, Springer Publications, Studies in Computational Intelligence Book Series, Innovations in Intelligent Machines Book-1, ch. 20, 70, 2007, 77-111.
[25] Ozkan, S. B.; Meirovitch, H.: Efficient Conformational Search Method for Peptides and proteins: Monte Carlo Minimization with an Adaptive Bias, J. Phys. Chem. B., 107, 2003, 9128-9131.
[26] Price, K. V.; Storn, R. M.; Lampinen, J. A.: Differential Evolution, a Practical Approach to Global Optimization, Springer-Verlag, Berlin Heidelberg, 2005.
[27] Saunders, M.: Stochastic Exploration of Molecular Mechanics Energy Surfaces. Hunting for the Global Minimum, J. Am. Chem. Soc., 109, 1987, 3150-3152. doi:10.1021/ja00244a051
[28] Storn, R.; Price, K.: DE - a Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Space, ICSI, Technical Report TR-95-012, 1995.
[29] Thomsen, R.: Flexible Ligand Docking Using Differential Evolution, Proceedings of the 2003 Congress on Evolutionary Computation, IEEE Press, 4, 2003, 2354-2361.
[30] Thomsen, R.; Christensen, M. H.: MolDock: A New Technique for High-Accuracy Molecular Docking, American Chemical Society, Med. Chem. Journal, 49, 2006, 3315-3321.
[31] Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R.: A Critical Comparison of Search Algorithms Applied to the Optimization of Protein-Side-Chain Conformations, J. Comput. Chem., 14, 1993, 790-798. doi:10.1002/jcc.540140705
[32] Wehrens, R.: Small-Molecule Geometry Optimization and Conformational Search, Evolutionary Algorithms in Molecular Design, 2000.
[33] Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E.: Conformational Analysis, Simulated Annealing, Tetrahedron Lett., 1988, 4343.
[34] Yang, J. M.; Kao, C. Y.: Flexible Ligand Docking Using a Robust Evolutionary Algorithm, Journal of Computational Chemistry, 21(11), 2000, 988-998.
[35] Yang, J. M.; Horng, J. T.; Kao, C. Y.: Integrating Adaptive Mutations and Family Competition with Differential Evolution for Flexible Ligand Docking, IEEE, 2001.
[36] Zhang, M.; Kavraki, L. E.: A new method for fast and accurate derivation of molecular conformations, Chemical Information and Computing Sciences Journal, 42, 2004, 64-70. doi:10.1021/ci010327z