J Math Model Algor (2010) 9:195–208
DOI 10.1007/s10852-010-9128-y

Single- and Multi-Objective Cooperation for the Flexible Docking Problem

Jean-Charles Boisson · Laetitia Jourdan · El-Ghazali Talbi · Dragos Horvath

Received: 1 March 2009 / Accepted: 1 December 2009 / Published online: 19 February 2010

Abstract In this article, the impact of single-objective methods as intensification factors in a multi-objective approach is presented for the flexible docking problem. Based on a novel tri-objective model, a parallel multi-objective genetic algorithm has been designed. However, due to the high variability of the energy objective, intensification methods focused on this objective have been also included in order to improve the convergence speed of the genetic algorithm and the quality of the results. The corresponding approach, combining single- and multi-objective methods, has been proved efficient according to the tested instances and the quality criterion used.

Keywords Multi-objective optimization · Molecular docking · Genetic algorithm

1 Introduction

The molecular docking problem consists in finding how two molecules, generally a very small called ligand, and a bigger one, called receptor, are going to make together a stable complex. This problem is essential in drug design because the action of a (or more) ligand(s) may activate, inhibit or modify the behavior of the receptor. This difficult problem, as many problems in bioinformatics, has a huge associated search space. It is first due to the combination of all the possible location of the molecules from each other. Furthermore, the molecules generally modify their 3D shape during the docking process. Consequently the size of the search space explodes. Each 3D shape of a molecule is called conformation. It corresponds to the specific location of all the atoms that compound the molecule. According to the flexibility of the molecules, three versions of the docking problem exist: rigid, semi-flexible and (full)

J.-C. Boisson (✉) · L. Jourdan · E.-G. Talbi · D. Horvath
INRIA Dolphin project team (Lille), "Infochimie" laboratory, Strasbourg, France
e-mail: jean-charles.boisson@univ-reims.fr

flexible. The docking is called rigid if none of the molecules has its conformation modified during the docking process. In the semi-flexible version of the docking, one molecule, generally the ligand, is allowed to have its conformation modified. The last version of docking, the most resource and time consuming version, is the flexible one. In this version, the both molecules may have conformation modifications during the process.

However, this problem is not a new one. Since the 1990s a lot of approaches have tried to tackle this problem. Bursulaya et al. [21] proposed an interesting study concerning the different methods used to solve this problem. The most known algorithms are the ones included in the following tool suites: Autodock [18], Flex [20], DOCK [19], ...The most recent algorithms are always designed for the flexible version of the docking in order to better simulate the real biological process.

Based on a collaboration with chemists, a new multi-objective model has been designed for the flexible docking problem. It mixes energy criterion and geometric criterion. A multi-objective algorithm has been designed with this model. The aim of our work is to show the impact of single-objective intensification processes in this multi-objective approach. To present our study, the paper is divided into different sections. Firstly, our tri-objective model is detailed. Then the method implementing this model is presented step by step. Due to the cost of the evaluation criteria, a parallelized version of our approach is also proposed. After the validation of the different hypothesis linked to our model, the results of our approach according to the different configurations of our algorithm are given. Lastly, conclusions and perspectives about this work are exposed.

2 Model

In this section, our tri-objective model, mixing energy and geometric criteria, is presented.

2.1 Energy Criterion and Associated Force Field

The lower the associated energy is, the more stable a molecule is. Based on this fact, all the docking approaches (single- and multi-objective) have at least one criterion based on the energy of the ligand/receptor complex. The computation of this energy needs an associated force field. A lot of force fields exist and they are not all dedicated to work with the same types of molecule. It is very expansive to design a force field that exactly simulates all the forces and parameters of the real docking process. Due to this cost, empirical approaches for force field design, as the one presented in this section, are generally preferred. Empirical approaches are often considered as the best compromise between a realistic and an expansive simulation.

In our case, a home made force field inspired of the **C**onsistent **V**alence **F**orce **F**ield has been used. The energy criterion associated to this force field exploits the standard energy contributions and can be divided into two main terms: the bonded atom energy and the non bonded atom one.

The bonded atom energy corresponds to all the energy interactions that occurs between two atoms that are linked together (bond contribution), linked to the same atom (angle contribution) or belonging to the same torsion (torsion contribution).

A torsion is a set of four atoms linked together thanks to three bonds. All these contributions are detailed in the following equation:

$$E_{\text{bonded\_atoms}} = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} K_\phi (1 - cos\, n(\phi - \phi_0)) \quad (1)$$

K_b, K_θ and K_ϕ are force constants linked to each contribution (bond, angle and torsion). b_0, θ_0 and ϕ_0 are empirical values for the bond size, the angle and the phase difference. b, θ and ϕ are the real value for the same characteristics. For the torsion contribution, n is the period linked to the type of the central bond of the torsion (double or triple). The bonded atom energy fluctuates during the modification of the conformation of a molecule.

The non bonded atom energy describes the energy contributions between atoms that are not near according to the topology of the molecule (the bonds, angles or torsions) but near according to the location of the atoms. The atoms concerned by the non bonded energy terms can belong to the same molecule or not. All these terms are described in the following equation:

$$E_{\text{non\_bonded\_atoms}} = \sum_{\text{Van der Waals}} \frac{K_{ij}^a}{d_{ij}^{12}} - \frac{K_{ij}^b}{d_{ij}^6} + \sum_{\text{Coulomb}} \frac{q_i q_j}{4\pi\epsilon d_{ij}} + \sum_{\text{desolvation}} \frac{K q_i^2 V_j + q_j^2 V_i}{d_{ij}^4} \quad (2)$$

In this equation, q_i is the charge of the atom i; d_{ij} is the atomic distance between the atoms i and j; V_i is a volumetric measure for the atom i; K and K_{ij}^x are constant forces linked to each contribution. The non bonded energy is the one giving the best description of the interactions between the ligand and the receptor. These interactions combine attractive and repulsive forces including the action of a solvent. Generally it is the water that is considered.

According to these two equations, our first criterion is the global energy:

$$E_{\text{total}} = E_{\text{bonded\_atoms}} + E_{\text{non\_bonded\_atoms}} \quad (3)$$

The associated force field, originally designed for protein structure prediction algorithms, has been tuned on several types of proteins and it is constantly improved.

2.2 Surface Criterion

The surface is a geometric criterion based on the hypothesis that the surface area describing both the receptor and the ligand is going to decrease because of the penetration of the ligand into the receptor. There are several ways to draw molecular surfaces. An atom is represented as a sphere with its Van der Waals radius. By summing all the spheres of a molecule, the Van der Waals surface is obtained.

However, this surface does not model the solvent action. The solvent accessible surface and the Connoly surface include the simulation of the solvent action, they are been described in [8] and [9] respectively. These surfaces model the path that follows the center of a sphere that represents the solvent. Generally it is a sphere of 1.4 Å[1] because the standard solvent is the water. The sphere center rolls on the atom spheres. The raw area draws by the sphere is the solvent accessible surface, the smoother version of the drawing is the Connoly surface.

In our model, for a good cost/quality compromise, the solvent accessible surface has been chosen as surface criterion. The algorithm used has been presented in [10] and approximate the Shrake and Rupley method [12]. A recent work using this algorithm has been presented in [11].

2.3 Stability Criterion: A Hypothesis to Validate

This criterion is based on an entropy calculus on a ligand/receptor complex. It describes the resistance of the complex against location modifications of the ligand into the receptor (rotation/translation). This calculus needs a sampling of neighbor complexes. According to the number of neighbors nbConf and their associated energy, the stability criterion is given by the Eq. 4.

$$\text{Stability} = -\frac{1}{\beta} * ln \sum_{i=1}^{\text{nbConf}} -\beta * \text{Energy}[i] \tag{4}$$

The β value is $\frac{1}{kT}$ with k being the Boltzmann constant and T being the temperature. The aim is to estimate if a complex is more probable than an other. The hypothesis contained in Eq. 4 is that a good ligand/receptor complex is rather in a valley of equivalent energies than in a narrow well surrounded by energy peaks.

3 Methods

In this section, the design of our multi-objective algorithm is detailed.

3.1 The Algorithm Choice

Three reasons have motivated the choice of genetic algorithms to solve our problem with the proposed multi-objective model:

1. A population-based method allows to propose naturally a set of solutions as a result.
2. The genetic algorithms have a good power of exploration that can be combined with intensification mutation operators. Consequently, they are able to tackle problems with a huge search space.
3. For genetic algorithms, multi-objective schemes proved to be efficient and robust exist.

[1] 1 angstrom (Å) is equivalent to 0.1 nanometer

3.2 Multi-Objective Genetic Algorithm (MOGA)

The design of genetic algorithms (GAs) is mainly based on the choice of the solution representation and evaluation. From these choices, the operators and all the other mechanisms of a GA can be defined. In our case, additional mechanisms are needed since our GA is a multi-objective one.

3.2.1 Solution Representation

In order to code a ligand/receptor complex, several representations are proposed in the literature. The most used is the torsion based representation. All the torsions that describes a molecule are coded as a vector of angle values. Knowing the angle of each torsion is sufficient to determine the conformation of the molecule. This representation is the smallest in memory but may generate additional calculi for the solution evaluation, the GA operators and the molecules drawing. Another representation is the Cartesian coordinates of each atom. This second representation is bigger than the torsion based one, but is very easy to apply on solution evaluation. With the atomic coordinates of each atom, the conformation of the molecule is directly known. That is why the ligand/receptor complexes are represented as a couple of atomic coordinate lists: one for the ligand and one for the receptor.

3.2.2 Multi-Objective Scheme

In order to design MOGAs, robust schemes already exist. Lot of them compare solutions using the dominance notion. A solution dominates another one if at least one of its criterion value is strictly better and the others criteria are better or equal to the other solution. Maybe the scheme the most known and used is the Non Dominated Sorting Genetic Algorithm II NSGA-II [13]. Also based on the notion of dominance, MOGA (Multi-Objective Genetic Algorithm) [15] and SPEA (Strength Pareto Evolutionary Algorithm) [16, 17] can be also cited. IBEA [14] (Indicator Based Evolutionary Algorithm) is a bit different. This algorithm is based on binary quality indicators.

Two of theses schemes have been tested thanks to the ParadisEO platform [23][2] and more precisely MOEO for Multi-Objective Evolving Object [22] : NSGA-II and IBEA. Statistical results have proved that for our approach and our instances, IBEA is better than NSGA-II [1].

3.2.3 Final Quality Evaluation

According to our home made force field, the only criterion to evaluate the quality of our approach and allowing to be comparable with other ones is the **R**oot **M**ean **S**quare **D**eviation (RMSD). According to [7], the RMSD is generally defined as follows:

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{n}(dx_i^2 + dy_i^2 + dz_i^2)}{n}} \tag{5}$$

[2] http://paradiseo.gforge.inria.fr

In the Eq. 5, n is the total number of heavy atoms (no hydrogen), dx_i, dy_i and dz_i are the deviations of the Cartesian coordinates of the atoms between the same molecules at different locations. The unity of the RMSD is the *angström* noted Å. 1 Å is equivalent to 0.1 nanometer. For our docking approach, the RMSD computation is made between the ligand in its optimal location and the locations found at the end of the algorithm. According to other approaches, a docking is considered as a good one with solution of RMSD in [1, 2] Å and as a very good one with solutions of RMSD < 1 Å.

Although the RMSD allows to easily compare different docking approach, this criterion is not a robust quality indicator. Due to its definition, several parameters may have an impact of the result of the RMSD computation: the size of the considered molecule or symmetric parts in the molecule. A small deviation in a small molecule generates a great perturbation of the RMSD but not in a bigger molecule. Symmetric parts in a conformation can generate a bad RMSD despite a good global location of the molecule.

Furthermore, using the RMSD as a quality indicator assumes that the optimal location of the ligand is known. In real studies, this location is the goal to reach and is unknown.

3.2.4 Operators

Specific recombination and mutation operators have been designed.

3.2.4.1 *Recombination:* according to the solution representation, a ligand swap between two complexes has been chosen. This type of mechanism is simple and does not need a verification step because of the criteria used in our model. If a ligand swap generates atom collisions or too narrow positions for the atoms, the associated energy will greatly increase, mainly due to the Van der Waals force. This type of recombination allows to rapidly share the good properties of the best individuals.

3.2.4.2 *Mutation:* two main types of mutations have been designed: standard ones and specific ones dedicated to our approach.

Standard mutations: during a flexible docking process, let's introduce the well known three standard modifications:

1. Conformation rotation: one member of the ligand/receptor complex (generally the smaller, so the ligand) has its conformation rotated. The center of the rotation is a virtual point in the middle of the chosen molecule and the rotation angle is randomly determined according to the X, Y and Z axis. In our case, the rotation always concerns the ligand.
2. Conformation translation: one member of the complex follows a translation according to a 3D vector randomly generated. As for the first mutation, only the ligand will be translated.
3. Torsion rotation: this modification of conformation is only feasible in the flexible case. One torsion angle is modified by rotation and consequently the conformation of the molecule changes. The impact of the rotation on the energy criterion depends of the place of the torsion in the backbone of the concerned molecule.

Advanced mutations: these mutations are divided in two families: the specific ones and the intensive ones. The first family contains mutations that follow or exploit the specific characteristics of our problem although the second one incorporates intensification mechanism dedicated to the speed-up of the stable complexe search.

– *Specific mutations:* each mutation has its own goal:

1. Mutation reverse: according to the standard definition of the RMSD, molecules with symmetric part may produce bad RMSD value despite a good position in the receptor. The reverse mutation allows not only to avoid this type of situation but also allows to restart exploration of the search when the ligand is trapped in a local optimum. This mutation makes a big rotation of the ligand, typically 180°.
2. Big and small rotations: this mutation is an extension of the standard rotation mutation. The aim is to be able to produce rotation according to two ranges of values: one with small values and the other one with high values. A priori, during the docking process, small rotations, respectively big rotations, may be effective during several steps but not during the entire search.
3. Several Mutations in One mutation (SMO): this is simply the application of the other mutations several times without evaluation. This type of mutation is dedicated to tackle the limit linked to each mutation. This mutation can be viewed as a variable neighborhood search mechanism. It allows to exit from local optimum and access to other areas of the search space. However,

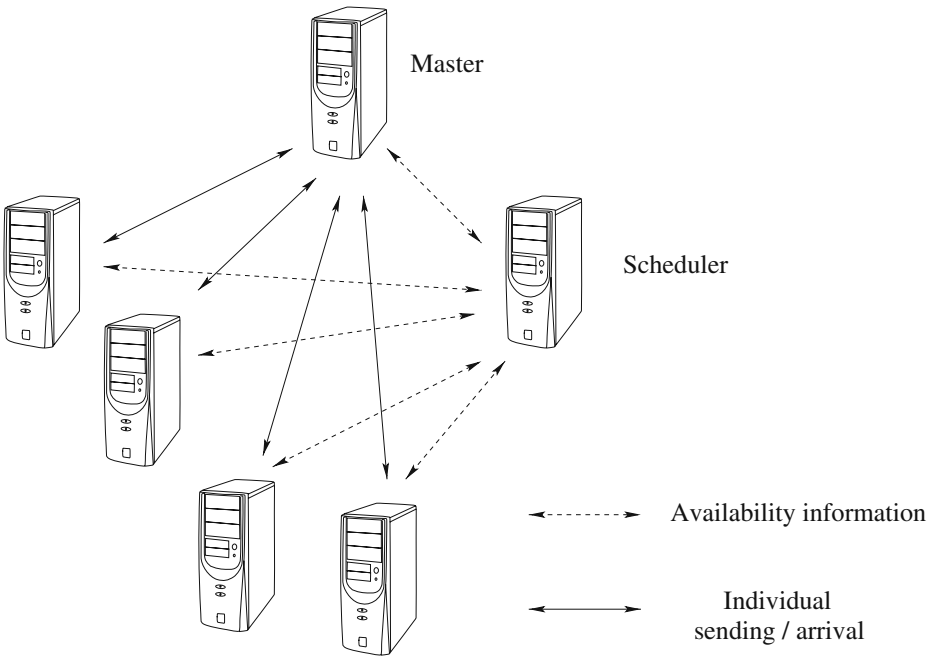

Fig. 1 Master/slave paradigm under ParadisEO-PEO. Example for four slaves (six processors are needed: four slaves, one master and one scheduler)

Table 1 Execution time in seconds (ExecTime), speed-up (S_N) and efficiency (Eff) on a cluster according to the number of slaves (Nb slaves)

Nb slaves	ExecTime	S_N	Eff
1	347	1	1
2	234	1.48	0.74
4	135	2.57	0.64
8	84	4.13	0.52
16	66	5.26	0.33
32	67	5.18	0.16
48	72	4.82	0.1
54	69	5.03	0.09

this type of mutation can be dangerous has to be applied carefully. Indeed, it can produce very deprecated solutions.

– *Intensive mutations:* due to the high variability of the energy criterion, two molecules may have a similar conformation but very different associated energies. Based on this remark, two types of mutation have been designed. These mutations extends standard mutations to decrease rapidly the energy to gain more stable complexes. Each mutation corresponds to a hill climbing algorithm based on a specific partial neighborhood. These hill climbing algorithms have been designed thanks to the Moving Object part (MO) of ParadisEO. More information about ParadisEO-MO may be found in [5]. These mutations are defined as follows:

  - ligand rotation based hill climbing: in a range of given values, new complexes are generated by ligand rotation. The search space associated to this mechanism is directly linked to the size of this interval of values (minimum and maximum angle value and step between two values in the three directions corresponding to the three dimensions). Then only a partial neighborhood is generated around an individual. The number of neighbours generated and also the cost of the evaluation function reduces the interest to make a too large neighbourhood around an individual.
  - torsion rotation based hill climbing: as the other hill climbing approach, the aim is to find a more stable conformation around an individual. For this mutation, the neighbourhood is based on the rotation of one particular torsion to find the angle that minimizes the global energy. In this case also, only a partial neighbourhood is explored to not spend too much time in applying this mutation.

Table 2 Full description and PDB identifier of the ligand/receptor complexes used

PDB means Protein Data Bank and allows to identify the instances

Ligand/receptor complexes	PDB
Ribonuclease A / Uridine-2',3'-Vanadate	6rsa
HIV-1 Protease / G26	1mbi
Thymidilate / CB3	2tsc
HIV-1 Protease / G26	1htf
Glucoamylase-471 / Alpha-d-mannose	1dog
IGA-kappa MCPC603 FAB / Phosphocholine	2mcp

Table 3 Molecular details concerning the instances used

PDB	NbAtom		DegFree	
	Ligand	Receptor	Ligand	Receptor
6rsa	31	1610	3	19
1mbi	9	2539	0	22
2tsc	56	2914	11	0
1htf	79	3127	15	4
1dog	24	3362	5	6
2mcp	25	1225	8	5

PDB means Protein Data Bank and allows to identify the instances, NbAtom corresponds to the number of atoms that compounds the instance and DegFree gives the freedom degree of the corresponding molecule. For the receptor, the DegFree value concerns the part considered as flexible (not the full molecule)

4 Parallel MOGA

The MOGA has been design under the master/slave paradigm of ParadisEO-PEO (the parallel part of ParadisEO) in order to be runnable on several processors. This paradigm, according the ParadisEO platform, is described in the Fig. 1. The main genetic algorithm runs on the master, the slaves wait for work and the scheduler knows which worker is available or not.

In our MOGA, not only the population evaluation is parallelized but also all the GA operators. An evaluation of the speed-up obtained by the parallel version of the MOGA is given in the Table 1. The speed-up and the efficiency are given by the following formulae:

$$S_N = \frac{St}{Pt} \qquad Eff = \frac{S_n}{N}$$

St is the sequential execution time and *Pt* the execution time with *N* slaves.

According to this table, it is logical that using several processors reduce the execution time of the algorithm. But a high number of processors is not necessary to keep a good efficiency. 32, 48 or 54 slaves leads to comparable results.

5 Experiments

In this section, the data preparation, the validation of the hypothesis concerning the stability criterion and the configuration of the parallel MOGA are detailed before presenting the results.

Table 4 RMSD between the seed ligand and the ligand in its optimal position for each instance

Instance	RMSD seed VS optimal (Å)
6rsa	7.15
1mbi	7.93
2tsc	13.48
1htf	14.45
1dog	10.68
2mcp	7.07

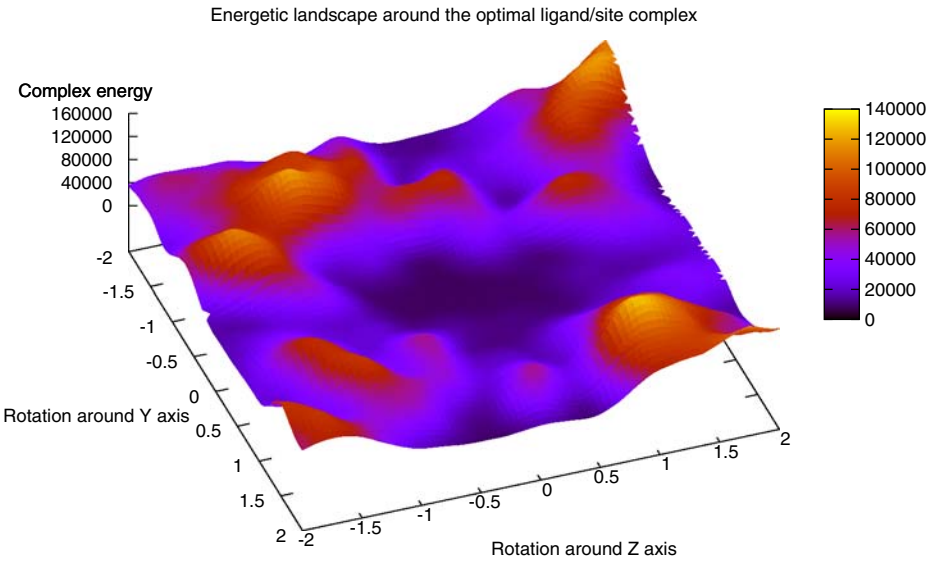

Fig. 2 Energy landscape around the crystallographic complex for the *6rsa* instance. This landscape is based of rotational modification of the ligand location according to the Y and Z axis

5.1 Data

Our simulation data are based on crystal structures taken from the RCSB Protein Data Bank (PDB). These data and the associated details are summarized in Tables 2 and 3. These complexes are taken from the clean list of the CCDC/Astex dataset. The work presented in this article is a part of global study about the multi-objectivization of the docking problem. Then, a reduce number of instances has been used. In [6], a lot of models are tested to complete the algorithm configuration.

Table 5 The eleven configurations (config) tested for the genetic algorithm: six for rigid docking (CX_R) and six for flexible docking (CX)

Config	SR	TorsRot	Reverse	SBRot	SMO	LRHC	TRHC
C1_R	X						
C1	X	X					
C2_R	X		X				
C2	X	X	X				
C3_R	X			X			
C3	X	X		X			
C4_R	X				X		
C4	X	X			X		
C5_R	X					X	
C5	X	X				X	
C6	X	X					X

SR standard rotation and translation of the ligand, *TorsRot* rotation of one torsion, *SBRot* rotation with small and big values for the rotation angle, *SMO* single mutations in one, *LRHC* ligand rotation driven by a hill climbing approach, *TRHC* torsion rotation driven by a hill climbing approach

Table 6 Statistic results for each instance according to the configuration of the algorithm

	C1	C1_R	C2	C2_R	C3	C3_R	C4	C4_R	C5	C5_R	C6
Minimum	2	2.1	2.4	2.3	2	2.1	*1.9*	2	2.2	*1.9*	2.4
First quartile	*3.2*	3.4	3.3	3.425	3.4	3.225	3.3	3.5	3.825	3.6	3.25
Median	*4.05*	4	4.5	5.75	*4.1*	*4*	4.35	*4.1*	*4.1*	4.05	4.6
Mean	*4.343*	4.4	4.729	5.293	5.126	4.686	5.169	5.129	4.733	4.964	4.776
Third quartile	5.075	4.875	6	6.75	6.3	5.875	6.325	6.15	*4.775*	7	6
Maximum	10.5	*7.7*	10.8	10	11	11	12.6	12	14	11.8	8.5
Standard deviation	1.651	*1.556*	1.937	2.078	2.524	2.183	2.865	2.576	2.29	2.42	1.794

From these complexes, two tool suites have been used in order to prepare our instances: the Vega ZZ tool[3] [4] and the UCSF Chimera tool suite[4] from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco funded by the NIH P41 RR-01081 [2, 3].

In Table 3, the degree of freedom of the receptor corresponds to the part considered as an actor in the flexible docking i.e. the torsions considered as flexible in the process. Only lateral chains of the amino acids that are nearer than 3 Å of one amino acid of the ligand in its optimal configuration are considered. Of course, this specific limitation of the freedom degree of the receptor is applied to reduce the size of the associated search space. It is only possible with instances where the best configuration (the crystallographic one) is already known.

In order be used in our approach, the instances coming from the PDB have been prepared as follows:

- For each instance, the receptor and the ligand are separated into two different files in a specific file format: Tripos mol2. A complete description of this format can be found in the document located at this address http://tripos.com/tripos_resources/fileroot/pdfs/mol2_format2.pdf. If necessary missing hydrogen atoms are added.
- The initial position of the ligand is saved as the optimal position (to reach). Then the ligand is manually extracted from this location, thanks to the Chimera tool suite, but its position is kept in front of the receptor pocket. This new position is considered as the ligand seed for the initial population of the genetic algorithm. According to the instance, the distance, in term of RMSD, between the seed position and the optimal position is given by the Table 4. This distance is useful to compare the other methods that start with a ligand nearer of its optimal location.

5.2 Stability Criterion Validation

In order to validate the hypothesis that ligand/receptor complexes are robust to modifications of the location of the ligand into the receptor, the energy landscape around each instance has been drawn according to rotational and translational modification of the ligand position. Figure 2 shows one of these landscapes.

[3]Drug Design Lab : http://www.ddl.unimi.it

[4]http://www.cgl.ucsf.edu/chimera/

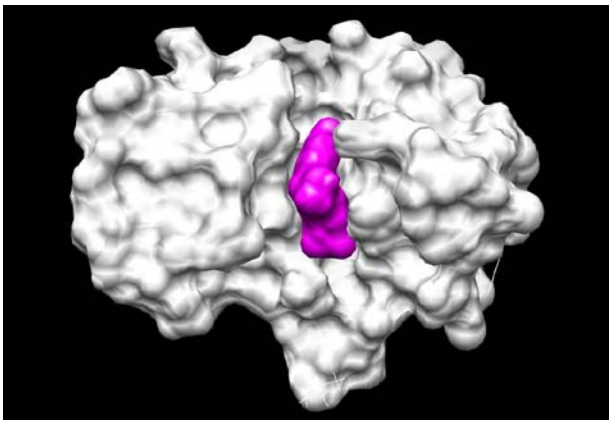

Fig. 3 Optimal complex for the 6rsa instance

According to all the obtained landscapes,[5] the optimal complex always belongs in valleys of equivalent energies for small modifications of the ligand location.[6]

5.3 Algorithm Configuration

The aim of our approach is to evaluate the impact of advanced mutations in our multi-objective model. In order to compare the results, several configurations of the algorithm have been tested. All these configurations are summarized in the Table 5. It can be noticed that the C6_R configuration does not exist. Torsion rotation based mutation can not be used in the rigid docking. The impact of each mutation has been independently evaluated with the standard mutations always activated.

Concerning the parameters of the MOGA, the population of solutions has been set to 100, the maximum number of generations was 10,000, but the algorithm may stop earlier if 500 consecutive generations without any new non-dominated solution discovered are observed. The probability of recombination and mutation has been set to 0.9 and 0.5 respectively.

5.4 Results

The Table 6 summarizes the results obtained for the all the instances according to each configuration of our genetic algorithm. All the configurations globally allows to find good docking solutions (with RMDS < 2 Å).

According to this table, the use of advanced mutations operators allows to get better results than with only standard mutations. It can be remarked that also the configuration C6 using a mutation based on a torsion rotation driven by a hill climbing does not give as good results as the C5 approach. This mutation is to restrictive in the area of the search space that is analyzed. It is difficult to find a good time/quality compromise for this specific mutation. In the C5 configuration,

[5]The graphics are available on demand

[6]The valleys are larger for translational modification based landscape

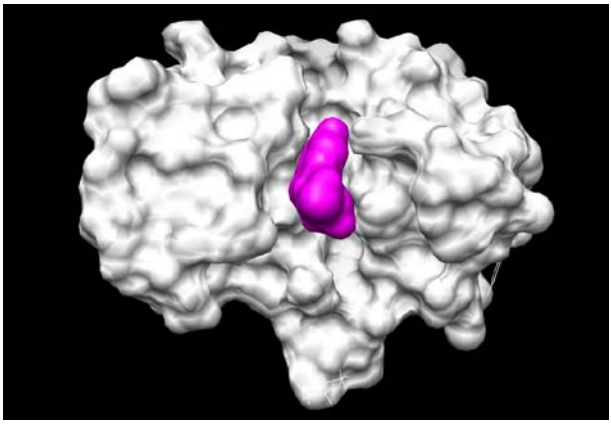

Fig. 4 Best complex found for the 6rsa instance

the hill climbing using ligand rotation explores enough search space to improve the quality of the results. Finally, the C4 configuration based on a mutation that acts like a variable neighborhood approach gives the best result. This configuration is able to obtain efficient solutions more rapidly than other mutations. This process may be made cleverer in order to improve the results.

The Figs. 3 and 4 show respectively the ligand/receptor complex with the optimal location for the ligand and best found location. These figures has been obtained thanks to the Chimera tool suite.

Comparing to the optimal location, the ligand is well enter into the receptor. However, the conformation has not been completely modified to respect the optimal one.

6 Conclusion

In this paper, a parallel multi-objective genetic algorithm has been designed for the flexible docking problem. Standard and advanced mutation operators including single-objective intensification processes based on the energy evaluation has been proposed. According to the results, the designed model is valid and the advanced mutations allows to get better results. Currently, improvement of the methods are tested on new instances by combining the best advanced mutations. Furthermore, GPU versions of our algorithm are currently designed. A multi-objectivization study of the molecular docking problem is also in progress.

Acknowledgements The model design of the docking problem has been supported by the French National Research Agency through the "Dock" project.

References

1. Boisson, J.-C., Jourdan, L., Talbi, E.-G., Horvath, D.: Parallel multi-objective algorithms for the molecular docking problem. In: IEEE 2008 Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Sun Valley Resort, Idaho, USA, 15–17 September 2008

2. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E.: UCSF Chimera - a visualization system for exploratory research and analysis. J. Comput. Chem. **25**(13), 1605–1612 (2004)
3. Sanner, M.F., Olson, A.J., Spehner, J.C.: Reduced surface: an efficient way to compute molecular surfaces. Biopolymers **38**(3), 305–320 (1996)
4. Pedretti, A.A., Villa, L., Vistoli, G.: VEGA - an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. J.C.A.M.D. **18**, 167–173 (2004)
5. Boisson, J-C., Jourdan, L., Talbi, E-G.: ParadisEO-MO: A Framework to Design Single Solution Metaheuristics. INRIA Technical report, INRIA Lille Nord Europe (2008)
6. Boisson, J-C.: Modélisation et résolution par métaheuristiques coopératives : de l'atome à la séquence protéique PhD Thesis, Université des Sciences et Technologie de Lille (2008)
7. Thomsen, R.: Protein-Ligand Docking with Evolutionary Algorithms Chapter in the book entitled Computational Intelligence in Bioinformatics (2008)
8. Lee, B., Richard, F.M.: The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. **55**, 379–400 (1971)
9. Richard, F.M.: Areas, volumes, packing and protein structure. Annu. Rev. Biophys. Bioeng. **6**, 151–176 (1977)
10. Le Grand, S.M., Merz, K.M., Jr.: Rapid approximation to molecular surface area via the use of boolean logic and look-up tables. J. Comput. Chem. **14**, 349–352 (1993)
11. Leaver-Fay, A., Butterfoss, G.L., Snoeyink, J., Kuhlman, B.: Maintaining solvent accessible surface area under rotamer substitution for protein design. J. Comput. Chem. **28**(8), 1336–1341 (2007)
12. Shrake, A., Rupley, J.A.: Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J. Mol. Biol. **79**(2), 351–364 (1973)
13. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 181–197 (2002)
14. Zitzler, E., Kunzli, S.: Indicator-based selection in multiobjective search. Parallel Problem Solving from Nature, PPSN VIII, vol. 3242, pp. 382–842. Springer, Germany (2004)
15. Fonseca, C.M., Fleming, P.J.: Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In: Proceedings of the 5th International Conference on Genetic Algorithms, pp. 416–423. San Mateo, California, USA, (1993)
16. Zitzler, E., Thiele, L.: Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans. Evol. Comput. **3**(4), 257–271 (1999)
17. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: improving the strength pareto evolutionary algorithm for multiobjective optimization, In: Proceedings of the Evolutionary Methods for Design, Optimisation and Control with Application to Industrial Problems (EUROGEN 2001), pp. 95–100 (2001)
18. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Har, W.E., Belew, R.K., Olson, A.J.: Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. **19**, 1639–1662 (1998)
19. Ewing, T.J.A., Kuntz, I.D.: Critical evaluation of search algorithms for automated molecular docking and database screening. J. Comput. Chem. **18**(9), 1175–1189 (1997)
20. Rarey, M., Kramer, B., Lengauer, T., Klebe, G.: A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. **261**(3), 470–489 (1994)
21. Bursulaya, B.D., Totrov, M., Abagyan, R., Brooks, C.L.: Comparative study of several algorithms for flexible ligand docking. J. Comput.-Aided Mol. Des. **17**(11), 755–763 (2003)
22. Liefooghe, A., Basseur, M., Jourdan, L., Talbi, E.-G.: ParadisEO-MOEO: a framework for multiobjective optimization. In: The Proceedings of EMO'2007, pp. 457–471 (2007)
23. Cahon, S., Melab, N., Talbi, E.-G.: ParadisEO: a framework for the reusable design of parallel and distributed metaheuristics. Journal of Heuristics **10**(3), 357–380 (2004)