Nucleic Acids Research, 2005, Vol. 33, Web Server issue ***W233–W238***
doi:10.1093/nar/gki586

MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm

Darby Tien-Hau Chang[1], Yen-Jen Oyang[1,2] and Jung-Hsin Lin[3,4,*]

[1]Department of Computer Science and Information Engineering and [2]Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei 106, Taiwan, ROC, [3]School of Pharmacy, National Taiwan University, Taipei 100, Taiwan, ROC and [4]Institute of Biomedical Sciences, Academia Sinica, Taipei, 115, Taiwan, ROC

Received February 15, 2005; Revised April 7, 2005; Accepted May 2, 2005

ABSTRACT

The prediction of ligand binding sites is an essential part of the drug discovery process. Knowing the location of binding sites greatly facilitates the search for hits, the lead optimization process, the design of site-directed mutagenesis experiments and the hunt for structural features that influence the selectivity of binding in order to minimize the drug's adverse effects. However, docking is still the rate-limiting step for such predictions; consequently, much more efficient algorithms are required. In this article, the design of the MEDock web server is described. The goal of this sever is to provide an efficient utility for predicting ligand binding sites. The MEDock web server incorporates a global search strategy that exploits the maximum entropy property of the Gaussian probability distribution in the context of information theory. As a result of the global search strategy, the optimization algorithm incorporated in MEDock is significantly superior when dealing with very rugged energy landscapes, which usually have insurmountable barriers. This article describes four different benchmark cases that span a diverse set of different types of ligand binding interactions. These benchmarks were compared with the use of the Lamarckian genetic algorithm (LGA), which is the major workhorse of the well-known AutoDock program. These results demonstrate that MEDock consistently converged to the correct binding modes with significantly smaller numbers of energy evaluations than the LGA required. When judged by a threshold of the number of energy evaluations consumed in the docking simulation, MEDock also greatly elevates the rate of accurate predictions for all benchmark cases. MEDock is available at http://medock.csie.ntu.edu.tw/ and http://bioinfo.mc.ntu.edu.tw/medock/.

INTRODUCTION

Successful virtual screening of chemical libraries in the drug discovery process requires (i) a sufficiently large and chemically diverse compound library, (ii) a very accurate scoring function and (iii) an efficient search algorithm for predicting the correct binding conformation, location and orientation of ligands. These three components are actually highly entangled. For example, if the search algorithm is not sufficiently efficient, then even when the scoring function has reached satisfactory accuracy, the prediction of correct binding modes will still remain a question of serendipity. On the other hand, if the scoring function has only very limited accuracy, then no matter how efficient the search algorithm is, finding the correct binding modes will still be highly unlikely. Also, if the chemical space under investigation is not sufficiently large, then finding the most potent compounds or new chemical entities is not likely. The docking simulations aim to mimic the biochemical process of a ligand approaching the active site of its receptor using computational methodologies. In practice the structures of target receptors at the atomic resolution, either from X-ray crystallography or from NMR spectroscopy (or even from homology modeling), are used for the docking simulations, which rely heavily on the second and third components of the aforementioned process. Recently, there have been several excellent reviews of the issues related to docking and virtual screening for drug discovery (1–7). Consideration of the induced-fit effect (8–10) and the potential for allosteric interactions (11), which could both be included in the more general sense of the extent to which chemical space is sampled (12,13), will also enhance the power and accuracy of computational drug design.

*To whom correspondence should be addressed. Tel: +886 2 2312 3456, ext. 8404; Fax: +886 2 2391 9098; Email: jlin@rx.mc.ntu.edu.tw

This article describes the design of the MEDock (*M*aximum *E*ntropy based *Dock*ing) web server, which is aimed at providing an efficient utility for the prediction of ligand binding sites. In particular, a comparison is made of how well the optimization algorithm incorporated in MEDock performs with respect to the Lamarckian genetic algorithm (LGA) (14), which is the major workhorse of the well-known AutoDock program (14). The MEDock server was given that name because it incorporates a global search strategy that exploits the maximum entropy (ME) property of the Gaussian probability distribution, in the context of information theory (15). Four benchmark cases were selected to represent various scenarios of ligand–receptor interactions, and their docking simulations were conducted to demonstrate the efficiency and the reliability of the new algorithm.

METHODS

This section elaborates the main algorithms incorporated into MEDock. The most important feature of MEDock in this regard is the use of a novel optimization algorithm that exploits the ME property of the Gaussian distribution. Although it is not a variant of the genetic algorithm (GA) (16), this novel optimization algorithm still belongs in the general category of evolutionary algorithms (17). A good analogy to the main heuristics employed in the optimization algorithm is the search for the lowest valley in a mountain range. In most complex systems, it is conceivable that valleys are clustered, and there may be several clusters of valleys within the entire mountain range. The shape of the energy landscape for docking simulations has been shown (7) to be similar to that for protein folding (18–20), which may be represented as a very rugged funnel with many insurmountable energy barriers. Accordingly, when a valley is found, one should continue to search the surrounding area for an even lower valley. Since valleys may be clustered, the likelihood of finding an even lower valley may decrease when one moves far away from the valleys that have already been identified. Therefore, the search should be governed by a bell-shaped probability distribution. The main reason why the Gaussian distribution is employed in the design of MEDock is because the Gaussian distribution has the maximum entropy, provided that the variance of the distribution is fixed (15). Since, in information theory, entropy means randomness (21), ME means maximum randomness. In other words, if the distribution of the deviation away from a valley is quantified by the variance of the distribution, a random search governed by the Gaussian distribution provides the maximum randomness for finding an even lower valley.

In each generation of the MEDock algorithm, n individuals, represented by $\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_n}$, are generated. Each $\boldsymbol{s_i}$ is in fact a vector defining the ligand's center of mass position (x_i, y_i, z_i), orientation $(\phi_i, \psi_i, \theta_i)$, and conformation $(\tau_{i1}, \tau_{i2}, \ldots, \tau_{ik})$ of the k rotatable bonds of the ligand, with τ_{ij} being the jth torsional angle. Accordingly, $\boldsymbol{s_i}$ can also be denoted as $(x_i, y_i, z_i, \phi_i, \psi_i, \theta_i, \tau_{i1}, \tau_{i2}, \ldots, \tau_{il})$, and the dimension of the vector space is $l + 6$. In order to achieve a scale-free representation, each component of $\boldsymbol{s_i}$ is linearly mapped to the numerical range of [0,1]. The fitness, i.e. the score or the free energy of binding, of each individual $f_1, f_2, \ldots, f_n$, is calculated using the AutoDock scoring function (14), which allows a fair comparison of the MEDock program with the AutoDock program. The energies are sorted in ascending order, and the one with the best score (i.e. the lowest energy) is ranked first in the whole population. The ordered individuals are denoted as $\boldsymbol{t_1}, \boldsymbol{t_2}, \ldots, \boldsymbol{t_n}$.

In the MEDock algorithm, n Gaussian distributions, denoted by $G_1, G_2, \ldots, G_n$, are generated before the new population in the next generation is created. The main idea of the MEDock algorithm is to give a bias toward the low-energy regions. The lower the energy, the more frequent the mining. Also, in order to ensure that the global minima can be located with increasing precision, the width of the Gaussian distribution should be set to diminish with the decrease of energy. Therefore, the center of each Gaussian distribution is selected randomly and independently from $\boldsymbol{t_1}, \boldsymbol{t_2}, \ldots, \boldsymbol{t_n}$, where the probability is not uniform but instead follows a discrete diminishing distribution, *n:* $n - 1 : \ldots : 1$. That is, the probability of picking up $\boldsymbol{t_1}$ is n times larger than that of selecting $\boldsymbol{t_n}$, the probability of choosing $\boldsymbol{t_2}$ is $n - 1$ times larger than that of picking up $\boldsymbol{t_n}$, and so on. Accordingly, one can expect that any distribution center $\boldsymbol{t_k}$ may be selected a number of times, and therefore some of the centers of these Gaussian distributions may be the same. After these centers are selected, one random sample from each of the Gaussian distributions will then be taken for creation of the next generation of individuals. The Gaussian distribution denoted by G_i is designated with the following probability density function:

$$\left(\frac{1}{\sqrt{2\pi}\cdot\sigma_i}\right)^{l+6}\exp\left(-\frac{(\boldsymbol{s_i}-\boldsymbol{t_k})^2}{2\sigma_i^2}\right),$$

where $\boldsymbol{t_k}$ is the center of G_i, and

$$\sigma_i^2 = \alpha + \frac{(\beta - \alpha)k}{n - 1}.$$

Here α and β are tuning parameters in MEDock and have been set to 0.01 and 0.5, respectively, as the default parameters in the current implementation. These default values for α and β have been determined by carrying out several docking experiments with different values of α and β. Some extreme values of α and β (such as $\alpha > 0.5$ and $\beta > 1$, or $\alpha > 0.5$ and $\beta < 0.00001$) may produce very poor docking results. Note that the default values of α and β remain the same for ligand–protein interactions of different sizes, because a scale-free unit has been applied for all components of $\boldsymbol{s_i}$. It should be stressed that exactly the same default values for α and β have been used in the four different docking benchmarks reported in the next section.

As mentioned earlier, the MEDock algorithm belongs to the evolutionary algorithm class but is quite different from the GA. The main difference is that the two key operations of the GA, i.e. crossover and mutation, have not been included in the MEDock algorithm. Figure 1 is a schematic example showing the sampling densities from different stages of the MEDock and GA evolutions. This one-dimensional energy surface is characterized by having one global minimum and a few other local minima, which are separated by normally insurmountable energy barriers. For MEDock it can be seen from Figure 1a that the initial distribution was rather uniform,

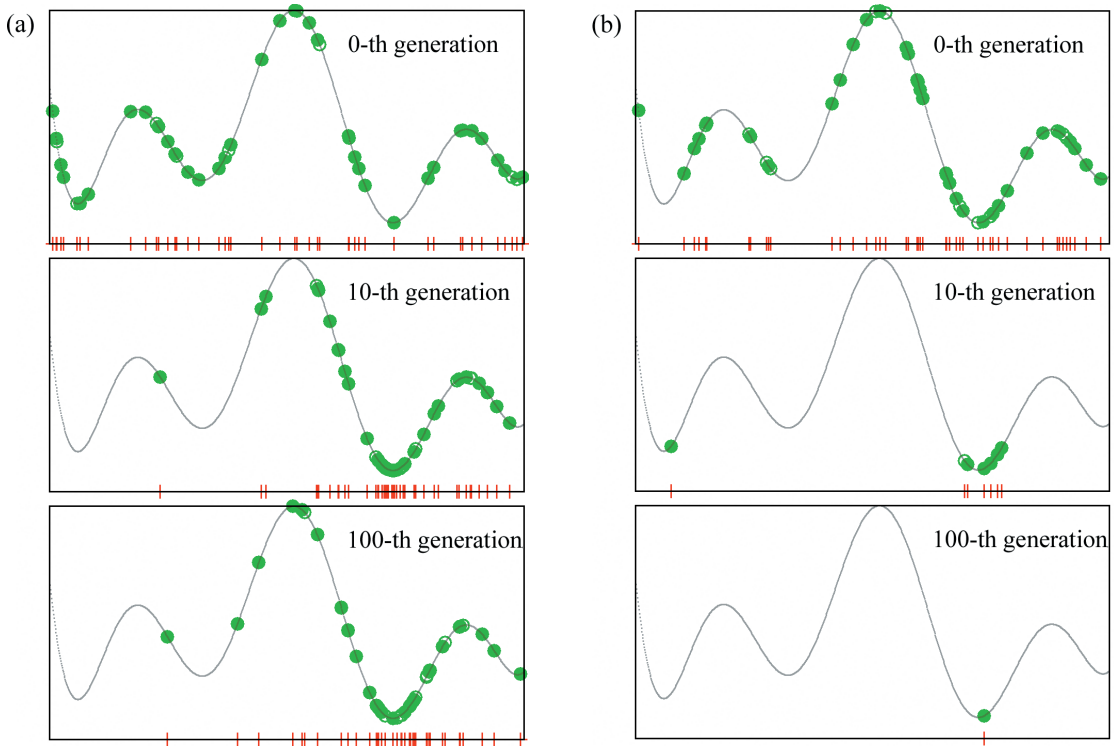

Figure 1. (**a**) A schematic example of sampling densities in different stages of the MEDock evolution. (**b**) Similar plots for the GA evolution. This one-dimensional energy surface is characterized by one global minimum with some other local minima, which are separated by usually insurmountable energy barriers.

and that, as the evolution progressed, the distribution became denser in the proximity of the global minimum. However, owing to the design of the Gaussian function in the MEDock algorithm, there will always be a finite probability of sampling the regions surrounding the previously identified local minima. It should be noted that this is distinct from the GA or its derivatives, e.g. the LGA, where the whole population is 'purified' after certain generations, as shown in Figure 1b. That is, all the individuals in the whole population will possess the same 'chromosome' (or set of 'genes'), which represents the orientation, location and conformation of the ligand, after the GA purifies the population.

One may argue that it is possible to preserve some of the sampling around the other local minima in a GA by enlarging the number of elites or by increasing the mutation rate. However, adjusting the GA parameters in this manner can not solve the problem completely and will always reduce the searching efficiency. It should be emphasized that this shortcoming automatically arises owing to the design of the GA, which uses crossover (also called recombination) and mutation as basic operations to generate new populations. 'Winner-takes-all' is always an unavoidable consequence as long as the GA's evolution continues for a sufficient number of generations.

Another central feature of MEDock is its incorporation of a new algorithm that aims to generate good-quality initial seeds for the optimization process. This algorithm is based on the novel kernel density estimation algorithm that was recently developed in-house and then incorporated into the ProteMiner web server, which identifies the cavities on the surface of a protein's tertiary structure (22). The geometric center of the alpha carbons of the residues that form one cavity then defines the center of one of the Gaussian distributions, from which the initial set of individuals is generated.

In order to enhance the searching efficiency, the Solis–Wets local search method (14) has also been employed in MEDock. This local search method basically facilitates random Monte Carlo moves around the currently found minima in order to find lower minima. For a very rugged, even fractal-like, energy landscape, such a non-derivative-based local search method is especially suitable for finding the true local minima. This also makes the benchmark cases a truly fair comparison between MEDock and the LGA in the AutoDock program (14).

INPUT, OUTPUT AND OPTIONS

The input file format is in the PDBQ format, which is an extension of the PDB format. The PDBQ format for ligands can be generated by many chemical software suites or web servers. For example, Dundee's PRODRG server (23) (http://davapc1.bioch.dundee.ac.uk/programs/prodrg/) provides a convenient visual interface to generate this file format from the PDB file (or from other file formats) of a ligand. Although computationally more demanding, the quantum chemical calculation procedure used in e.g. the relaxed complex scheme (9) may be invoked for a more accurate assignment of partial charges on each atom of the ligand molecule. It should be emphasized that the accurate assignment of the ligand's partial charges is critically important when dealing with ligand–receptor interactions that are dominated by electrostatics. The ligand files for these four benchmarks were prepared

using the relaxed complex procedures (9). The PDBQ file for proteins can be derived from the PDB2PQR server (24) (http://agave.wustl.edu/pdb2pqr/) and a simple awk or perl script. The PDB2PQR server also provides a prediction of the protonation states of the ionizable residues in a protein, which is an important issue for the correct description of ligand–receptor interactions. The MEDock web server also includes these two web servers for automatically converting the PDB format to the PDBQ format, just in case users do not have other preferred procedures for the required calculations and conversions.

RESULTS AND DISCUSSION

To demonstrate the efficiency of the new algorithm, four benchmark cases were selected and compared with the results of the AutoDock program. These four benchmarks were selected to represent various scenarios of ligand binding interactions. The graphical renderings of these benchmarks are shown in Figure 2. HIV protease is a classic example for virtual screening, in which a well-defined tunnel sits at the interface of the two monomers of this homodimeric protein. The second case is the binding of FK506, an immunosuppressant, to its target protein, FKBP. In contrast to the first case, the binding site for FK506 on the FKBP surface is rather shallow, and there is another, smaller binding site close to the main binding site which can serve as a decoy. The third case is the interaction of phospholipase A2 and aspirin, which, compared with the other cases, involves a relatively small ligand binding to a target protein that could contain many possible decoy binding sites. The fourth case considers a DNA–protein interaction in which a segment of DNA called the TATA-box selectively targets its binding partner, the TATA-box binding protein (TBP). This DNA ligand is substantially larger than the ligands used in the other three benchmark cases (e.g. the molecular weight of the DNA ligand was 8608.91, whereas aspirin has a molecular weight of only 180.04). There is also no well-defined cavity or distinct crevice that would obviously be recognized as the DNA binding site.

All four of these benchmarks were selected to satisfy the following criterion: the correct binding modes (i.e. the conformation, location, and orientation of the ligand in the crystallographic structure) should be reproducible by the AutoDock program, which implies that the use of the AutoDock scoring function is adequate in these cases. This greatly simplified the goal of the current work, which was to design a more efficient and reliable searching algorithm. To enable an extensive comparison between MEDock and AutoDock, docking simulations with fixed ligand conformations were conducted to reduce the dimensionality and the computational costs. It should also be noted that, although flexible ligands are permissible in many docking programs, very few of them implement an accurate molecular mechanics force field in their search for the correct ligand conformations. Balancing the ability to calculate both the ligand–receptor interactions and the ligand's internal conformational energy accurately and efficiently is a subtle issue and requires further, extensive investigation. In order to avoid introducing another source of error, these four benchmarks were performed with fixed

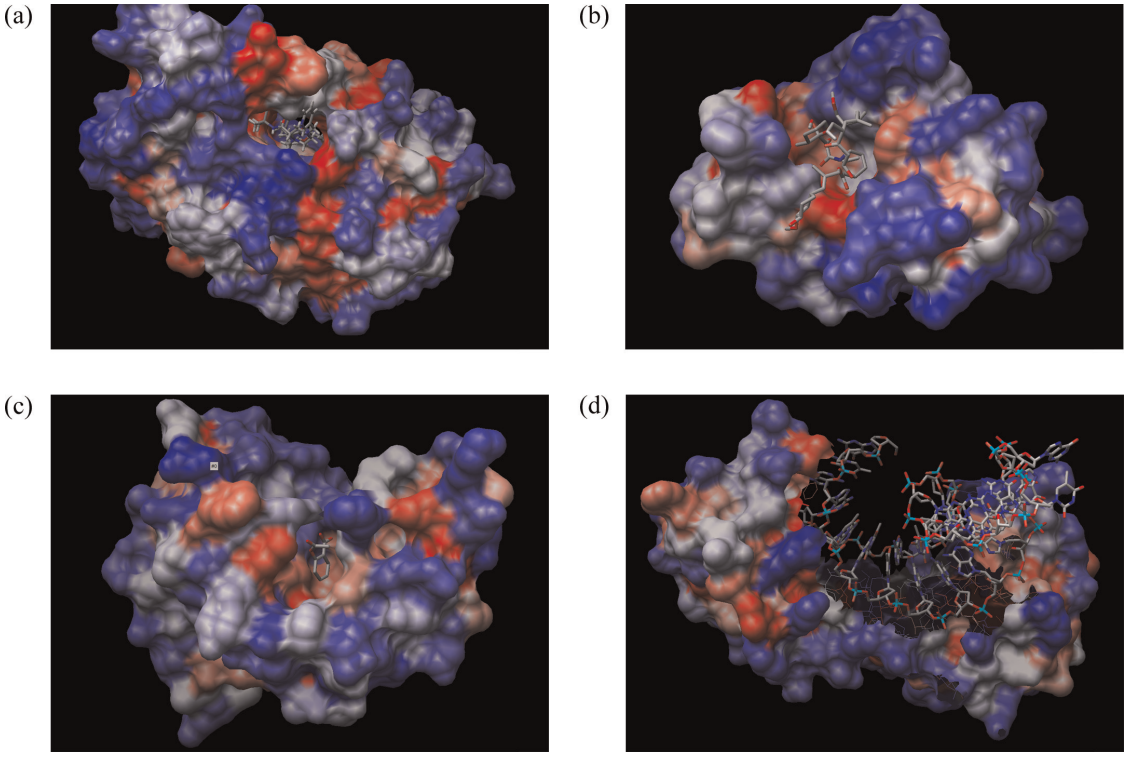

Figure 2. Molecular graphics display of the four benchmark cases: (**a**) HIV-II protease complexed with its inhibitor L-735,524 (PDB ID: 1HSH); (**b**) FKBP-FK506, an immunophilin-immunosuppressant complex (PDB ID: 1FKF); (**c**) Complex formed between phospholipase A2 and aspirin (PDB ID: 1OXR); and (**d**) TATA-box binding protein (YTBP) complexed with DNA containing a TATA-box (PDB ID: 1YTB).

ligand conformations. In general, the ligand's conformation will not be known in advance when attempting to predict the ligand's binding site; therefore, the docking of flexible ligands is still required. Consequently, flexible docking simulations were also performed for the benchmark cases, and the binding sites of all these cases were correctly predicted. However, the conformations, orientations and locations of ligands did deviate from the crystallographic binding modes.

Although there have been some comparisons of the different search algorithms (25–27), most of them simply use the default docking protocols and run parameters from the tested programs when performing their benchmarks. It should be emphasized that the default docking parameters may not be optimal for different systems, and in general they should be optimized for each system in order to perform a fair comparison of the different searching algorithms. This study did not intend to compare all the existing search algorithms with the MEDock algorithm; rather, its goal was to provide a solid basis for a fair comparison with the LGA of the AutoDock program. Therefore, all the GA parameters have been optimized for each of our four benchmark cases, and the same local search parameters were used for both the GA and MEDock. Figure 3 shows how the global search algorithm incorporated in MEDock performed in comparison with the LGA algorithm incorporated into AutoDock in terms of the number of energy evaluations it took to find the globally optimal docking state of the given ligand–protein pair.

At the start of these benchmarks, excess docking simulations were performed in order to determine the energies of the global minima, which were equivalent to the crystallographic binding modes for these four systems. The energies of these global minima were then used to define the stop criteria, so that statistics could be calculated regarding the number of energy evaluations that were consumed/run in order to reach that global minimum. For each of the four test cases, 100 independent runs of the same experiment were conducted with randomly generated seeds. The *y*-axis of Figure 3 represents the number of runs (out of the 100 independent runs) that converged to the correct binding state when the number of energy evaluations was limited to the corresponding value on the *x*-axis. In these experiments, the number of energy evaluations it took to reach convergence was used to measure the amount of time consumed by the software packages, because energy evaluations dominate the execution time of docking software. In these experiments the following most system-dependent run parameters used in AutoDock's LGA: the population size, the Solis-Wet iteration number, and the local search frequency, were tuned to improve its convergence rate, and then exactly the same set of optimized run parameters were used in MEDock.

As Figure 3 shows, for two out of the four test cases, AutoDock failed to guarantee that the globally optimal docking state will be found within a reasonable amount of time. In contrast, all the MEDock simulations were able to converge to the global minima within a much shorter time scale. This

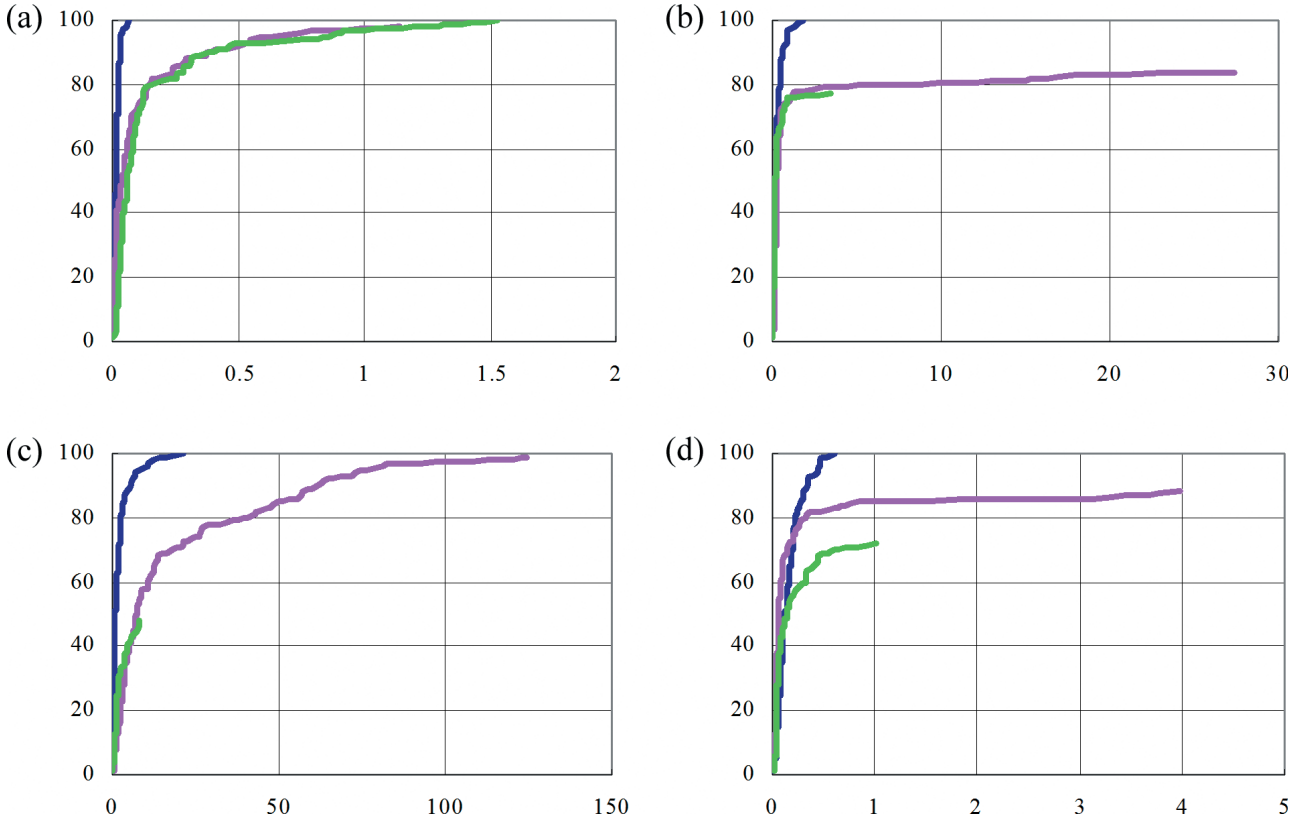

Figure 3. Number of runs to reach convergence versus the number of energy evaluations consumed (in units of 10^7): blue, MEDock results; magenta, LGA results (with parameters tuned); green, LGA results (with default parameters). (**a**) HIV-II protease complexed with its inhibitor L-735,524 (PDB ID: 1HSH); (**b**) FKBP-FK506, an immunophilin-immunosuppressant complex (PDB ID: 1FKF); (**c**) complex formed between phospholipase A2 and aspirin (PDB ID: 1OXR); and (**d**) TATA-box binding protein (YTBP) complexed with DNA containing a TATA-box (PDB ID: 1YTB).

Table 1. The number of energy evaluations taken by AutoDock and by MEDock to successfully identify the globally optimal docking state in 80 out of the 100 independent runs

	AutoDock	MEDock	Ratio
1HSH	1 345 942	224 700	6.0:1
1FKF	12 603 986	4 149 939	3.0:1
1OXR	273 082 703	25 901 232	10.5:1
1YTB	2 568 118	2 137 165	1.2:1

echoes the discussion in Methods that there will always be a certain probability for search algorithms such as the GA and its derivatives to fail to find the global minima. Table 1 compares the number of energy evaluations taken by MEDock and by AutoDock to successfully identify the optimum docking state in 80 out of the 100 independent runs for the four test cases. The data presented in Figure 3 and Table 1 together clearly demonstrate the efficiency and the reliability of the design of MEDock.

CONCLUSION

We have developed a novel methodology for docking simulations, which has been shown to be both more efficient and more reliable than the LGA. Our benchmarks consistently showed that MEDock converged to the correct binding modes while consuming significantly smaller numbers of energy evaluations. Given a threshold for the number of energy evaluations used in the docking simulation, MEDock also greatly elevated the rate of accurate prediction for all benchmark cases.

ACKNOWLEDGEMENTS

We thank Dr Alexander L. Perryman for carefully correcting the manuscript. J.-H.L. thanks the National Science Council of Taiwan for funding under contract NSC 93-2112-M-002-027. Funding to pay the Open Access publication charges for this article was provided by National Science Council of Taiwan.

Conflict of interest statement. None declared.

REFERENCES

1. Kitchen,D.B., Decornez,H., Furr,J.R. and Bajorath,J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.*, **3**, 935–949.
2. Lengauer,T., Lemmen,C., Rarey,M. and Zimmermann,M. (2004) Novel technologies for virtual screening. *Drug Discov. Today*, **9**, 27–34.
3. Jorgensen,W.L. (2004) The many roles of computation in drug discovery. *Science*, **303**, 1813–1818.
4. Jain,A.N. (2004) Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discov. Dev.*, **7**, 396–403.
5. Alvarez,J.C. (2004) High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.*, **8**, 365–370.
6. Shoichet,B.K. (2004) Virtual screening of chemical libraries. *Nature*, **432**, 862–865.
7. Halperin,I., Ma,B.Y., Wolfson,H. and Nussinov,R. (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins*, **47**, 409–443.
8. Lin,J.H., Perryman,A.L., Schames,J.R. and McCammon,J.A. (2002) Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.*, **124**, 5632–5633.
9. Lin,J.H., Perryman,A.L., Schames,J.R. and McCammon,J.A. (2003) The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers*, **68**, 47–62.
10. Cavasotto,C.N. and Abagyan,R.A. (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.*, **337**, 209–225.
11. Perryman,A.L., Lin,J.H. and McCammon,J.A. (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.*, **13**, 1108–1123.
12. Dobson,C.M. (2004) Chemical space and biology. *Nature*, **432**, 824–828.
13. Lipinski,C. and Hopkins,A. (2004) Navigating chemical space for biology and medicine. *Nature*, **432**, 855–861.
14. Morris,G.M., Goodsell,D.S., Halliday,R.S., Huey,R., Hart,W.E., Belew,R.K. and Olson,A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, **19**, 1639–1662.
15. Hyvarinen,A., Karhunen,J. and Oja,E. (2001) *Independent Component Analysis*. J. Wiley, New York.
16. Holland,J.H. (1975) *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, Ann Arbor, MI.
17. Clark,D.E. (ed.) (2000) *Evolutionary Algorithms in Molecular Design*. Wiley-VCH, Weinheim, New York.
18. Onuchic,J.N., Wolynes,P.G., Lutheyschulten,Z. and Socci,N.D. (1995) Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl Acad. Sci. USA*, **92**, 3626–3630.
19. Bryngelson,J.D., Onuchic,J.N., Socci,N.D. and Wolynes,P.G. (1995) Funnels, pathways, and the energy landscape of protein-folding-A synthesis. *Proteins*, **21**, 167–195.
20. Leopold,P.E., Montal,M. and Onuchic,J.N. (1992) Protein folding funnels—a kinetic approach to the sequence structure relationship. *Proc. Natl Acad. Sci. USA*, **89**, 8721–8725.
21. Cover,T.M. and Thomas,J.A. (1991) *Elements of Information Theory*. Wiley, New York.
22. Chang,D.T.H., Chen,C.Y., Chung,W.C., Oyang,Y.J., Juan,H.F. and Huang,H.C. (2004) ProteMiner-SSM: a web server for efficient analysis of similar protein tertiary substructures. *Nucleic Acids Res.*, **32**, W76–W82.
23. Schuttelkopf,A.W. and van Aalten,D.M.F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 1355–1363.
24. Dolinsky,T.J., Nielsen,J.E., McCammon,J.A. and Baker,N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.
25. Yang,J.M. and Chen,C.C. (2004) GEMDOCK: a generic evolutionary method for molecular docking. *Proteins*, **55**, 288–304.
26. Friesner,R.A., Banks,J.L., Murphy,R.B., Halgren,T.A., Klicic,J.J., Mainz,D.T., Repasky,M.P., Knoll,E.H., Shelley,M., Perry,J.K. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **47**, 1739–1749.
27. Bursulaya,B.D., Totrov,M., Abagyan,R. and Brooks,C.L. (2003) Comparative study of several algorithms for flexible ligand docking. *J. Comput. Aided Mol. Des.*, **17**, 755–763.