Application of the stochastic tunneling method to high throughput database screening

H. Merlitz, B. Burghardt, W. Wenzel *

Forschungszentrum Karlsruhe GmbH, Institut fur Nanotechnologie, Postfach 3640, D 76021 Karlsruhe, Germany

Abstract

The stochastic tunneling technique is applied to screen a database of chemical compounds to the active site of di hydrofolate reductase for lead candidates in the receptor ligand docking problem. Using an atomistic force field we consider the ligand's internal rotational degrees of freedom. It is shown that the natural ligand (methotrexate) scores best among 10 000 randomly chosen compounds. We analyze the top scoring compounds to identify hot spots of the receptor. We mutate the amino acids that are responsible for the hot spots of the receptor and verify that its specificity is lost upon modification.

1. Introduction

Virtual screening of chemical databases to targets of known three-dimensional structure is developing into an increasingly reliable method for finding new lead candidates in drug development [1]. Both better scoring functions and novel docking strategies contribute to this trend, although no completely satisfying approach has been established yet [2]. This is not surprising since the approximations which are needed to achieve a reasonable screening rate impose significant restrictions on the virtual representation of the physical system. Relaxation of these restrictions, such as permitting ligand flexibility [3 15], often increases the reliability of the scoring process. However, the inclusion of ligand flexibility significantly increases the dimension of the configuration space, turning the docking procedure into a rather ambitious task. It is therefore mandatory to carry on the search for novel docking approaches to be prepared for the demands of the upcoming generation of virtual screening codes.

With this investigation we contribute to this effort by evaluating the stochastic tunneling method (STUN) [16], a recently developed global optimization technique, to virtual screening with flexible ligands. By virtue of construction STUN can escape from metastable states of even extremely rugged potential energy surfaces (PES), which explains its success in peptide folding [17] and a first application to ligand docking in the rigid-ligand approximation [18].

In the present investigation STUN is used to screen a database of 10 000 flexible compounds to

* Corresponding author. Fax: +49 7247 82 6434.
E mail address: wenzel@int.fzk.de (W. Wenzel).

the active site of dihydrofolate reductase (4dfr). For this target, a highly optimized inhibitor is known (methotrexate, MTX), including the crystallographic X-ray structure of its docked conformation. MTX emerges by far as the top scoring ligand in the screened database, validating both docking approach and the choice of the scoring function. It is shown that MTX and other top scoring database leads coordinate with preferred amino acids for the creation of hydrogen bonds inside the cavity. We discuss how this observation could be employed to identify hot-spots in the active site of the protein and thus support the creation of pharmacophore models.

2. Computational methods

2.1. Preparation of the three-dimensional database

The ligands were taken from the open part of the NCI database (nciopen.mol), which had been processed to generate the three-dimensional structures with the help of Corina [19,20] (nciopen3d.mol). Hydrogen atoms and partial charges were added using the InsightII software package and its esff forcefield. Rotatable bonds were identified with a simple algorithm which searched for single bonds, excluding ring-structures, trivial single-atom end-groups and atoms with sp^2 hybrid orbitals. Among 125 000 compounds available, the first 10 000 were selected which satisfied the conditions of having not more than 100 atoms and not more than 10 rotatable bonds.

2.2. Preparation of the protein coordinates and definition of the active sites

The protein coordinates were taken from the X-ray structure of *Escherichia coli* dihydrofolate reductase with MTX (pdb entry: 4dfr, monomer B) [21]. Hydrogen atoms and partial charges were attached using InsightII with esff force field. The binding site was defined with the docked ligand in the X-ray structure: A volume of 5 Å radius was defined around the center of mass position of the docked MTX. If the global minimum on the PES of any ligand was located such that either the center of mass of this ligand or more than 20 ligand atoms were inside this volume, the ligand was regarded as docked.

2.3. Scoring function

For the simulations discussed below we used the following scoring function

$$S = \sum_{\text{Protein}} \sum_{\text{Ligand}} \left(\frac{R_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\epsilon r_{ij}} \right) + \sum_{\text{H-bonds}} \cos\Theta_{ij} \left(\frac{\tilde{R}_{ij}}{r_{ij}^{12}} - \frac{\tilde{A}_{ij}}{r_{ij}^{10}} \right), \quad (1)$$

which contains the empirical Pauli repulsion (first term), the Van de Waals attraction (second term), the electrostatic Coulomb potential (third term) and the angular dependent hydrogen bond potential (term four and five). The Lennard Jones parameters R_{ij} and A_{ij} were taken from OPLSAA [22], the partial charges q_i were computed with InsightII and esff force field, and the hydrogen bond parameters $\tilde{R}_{ij}$, $\tilde{A}_{ij}$ were taken from AutoDock [10]. The electrostatic potential was precalculated on a grid. The other contributions, however, turned out to be too rapidly varying to be reliably stored on a grid and interpolated. Instead we exploited the short-range character of these interactions. We stored a list of atoms next to each grid point and computed the short-range potentials on the fly. This procedure obviates the need for interpolation of rapidly changing terms and permits a straightforward implementation of the angle dependence of the hydrogen bond potentials. With these changes the accuracy of the evaluation of the scoring function could be significantly increased over an earlier entirely grid-based implementation [18].

2.4. Optimization method

Many stochastic optimization methods get routinely trapped in metastable minima of the potential energy surface. This problem is particularly acute in receptor ligand docking, where there is little specificity between ligand and receptor outside the receptor pocket and little room for

reorientation inside. The degree of affinity between receptor and ligand is discernable only where close to the ideal docking position, in particular because of the short-range nature of stabilizing hydrogen bonds. In the proximity of typical docked positions, however, the ligand cannot be significantly rotated or distorted without clashing with receptor atoms.

The STUN method [16] avoids this problem by letting the particle in the minimization process 'tunnel' forbidden regions of the PES. As in simulated annealing [23] we retain the idea of a biased random walk, but apply a non-linear transformation to the potential energy surface

$$E_{\mathrm{STUN}}(x) = 1 - \exp[-\gamma(E(x) - E_0)], \qquad (2)$$

where E_0 is the lowest minimum encountered by the dynamical process so far. Alternately a suitable upper bound for the global minimum can be used for E_0. This effective potential preserves the locations of all minima, but maps the entire energy space from E_0 to the maximum of the potential onto the interval [0,1]. At a given finite temperature of O(1), the dynamical process can therefore pass through energy barriers of arbitrary height, while the low energy-region is resolved even better than in the original potential. The degree of steepness of the cutoff is controlled by the tunneling parameter γ.

2.5. Computation

The electrostatic potentials were stored on a $30 \times 30 \times 30$ $\mathrm{\AA}^3$ grid with grid spacing of 0.5 Å, which was positioned around the center of mass of the X-ray structure of MTX. Each of the 10 000 ligands was shifted into the cavity and, if required, moved so that heavy clashes of the initial configuration were avoided. We then performed 10 STUN simulations, comprising 3×10^5 energy evaluations each. Only the repetition of the docking procedure ensured that weakly bound ligands are reliably docked. The (non-physical) simulation temperature was fixed to 5 K, the STUN parameter γ was chosen as 0.01 for all ligands. A drift term was applied in order to increase the sampling rate inside the cavity with respect to peripheral regions [18]. The size of the displacements in the conformational updates was automatically adjusted during the simulation to keep the acceptance rate around 0.5. In average, each ligand consumed about 20 seconds of CPU time for 10^5 updates on a PC (AMD Athlon 1.2 GHz).

3. Results

6100 ligands, out of 10 000 compounds, reached configurations where the ligand was embedded in the cavity and the external binding energy, scored with Eq. (1), was below −50 kJ/mol. It is well known that scoring functions as the one used here are too inaccurate to yield quantitatively approximate the natural affinity of the ligand. Nevertheless they often provide a useful relative ranking of leads when docked under identical conditions. Fig. 1 shows the external binding energies of the docked compounds. Among them, the natural ligand MTX clearly scored best, with a minimum energy conformation which differed only 1.4 Å from the X-ray crystal structure (Fig. 2). This figure shows that both conformations essentially differ in the positions of only one of their carboxylate groups. In the minimal conformation of the scoring function this group is turned around to

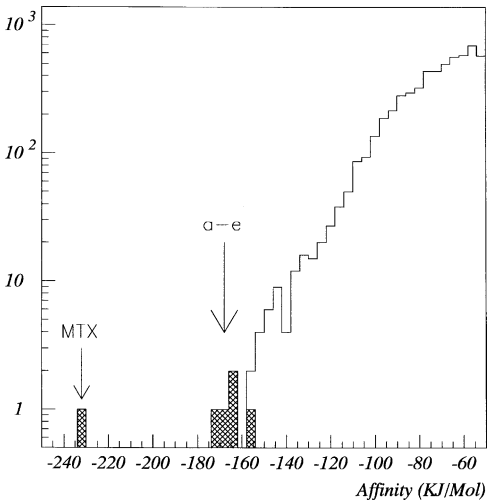

Fig. 1. The affinities of 6100 ligands docked to 4dfr. The natural ligand (MTX) was scoring best. The five top ranking database leads are denoted as $a, \ldots, e$.

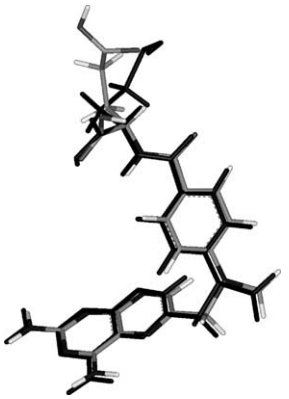

Fig. 2. MTX in its crystallographic X ray conformation (black) and the global minimum found with STUN (color). Both conformations differ in the positions of their carboxylate groups.

create an extra hydrogen bond to Lys-32 which is absent in the crystallographic conformation. In the experimental configuration a conserved water molecule (Wat-672) mediates competing hydrogen bonds in the natural environment, an effect which cannot be accounted for with the present scoring function.

The results of the screening process were encouraging not only because the natural ligand was selected with high specificity, but also because other high ranking ligands showed significant similarities in the binding mechanism. This observation, more than the exact ranking in the binding affinity, suggests that virtual screening methods can aid rational drug design. In Table 1

Table 1
Properties of ligand MTX and the 5 top ranking leads of the database, denoted as *a e*

Lead	#Atoms	#Rot.	Affinity (kJ/mol)	Hydrogen bonds to
MTX	54	9	233	Arg 52, Asp 27, Arg 57, Lys 32, Ile 5
a	46	10	173	Arg 52, Ile 14, Ala 7
b	28	8	169	Asp 27, Trp 22, Leu 24
c	38	7	164	Arg 52, Ile 50, Ala 7, Ile 94
d	29	4	163	Arg 52, Asp 27, Trp 22
e	37	9	156	Arg 52, Asp 27, Trp 22

#Rot. means the number of rotatable bonds, i.e. the number of internal degrees of freedom in the simulation.

we compare some of the relevant features of MTX with the five ligands which scored next. It is obvious that no other ligand could create as many hydrogen bonds with the protein as MTX. On the other hand, it appears that there exist certain residues inside the cavity which serve as preferred binding sites, in particular Arg-52 and Asp-27. One may hypothesize that these residues, along with the general shape of the receptor pocket, are responsible for the specificity of the receptor ligand interaction. We therefore replaced both amino acids with valine, which cannot act as a hydrogen bond donor/acceptor. The resulting affinities of a repeated screening process are shown in Fig. 3. Not surprisingly, the binding energies are now generally reduced and those ligands which had originally scored well have lost their top ranking positions. This exercise demonstrates how virtual screening can be employed to identify hotspots inside the active site of the protein to support the creation of pharmacophore models.

Finally, the reliability of docking MTX and the top ranking leads was investigated. The docking cycle was defined as successful if the lowest minimum led to a ligand conformation which was closer than 2 Å to the best known minimum which

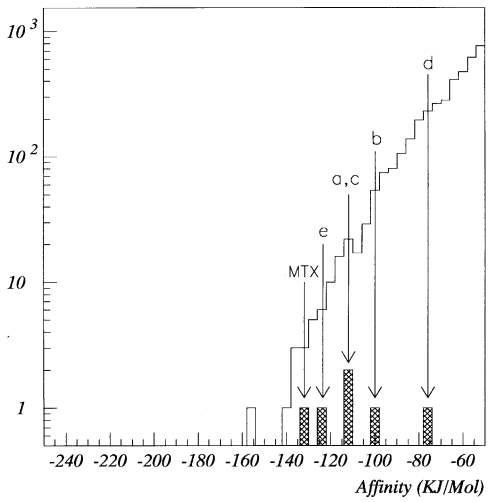

Fig. 3. The affinities of the ligands after docking to a modified 4dfr, in which Arg 52 and Asp 27 are replaced by Valine. Ligands which originally scored best have now lost their top ranking positions, including MTX.

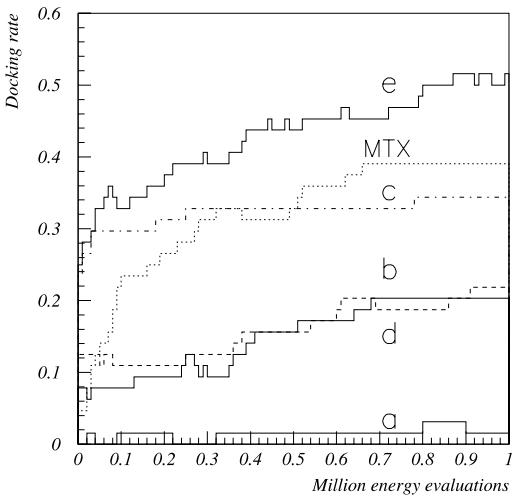

Fig. 4. Docking rates of MTX and the five top ranking leads as a function of energy evaluations.

satisfied the docking criteria of Section 2.2. Fig. 4 plots the average docking rates as a function of energy evaluations. We observe that after 10^6 scorings none of the ligands were able to reach its minimum conformation with more than 50% probability, but that significant docking rates were reached already after 10^5 energy evaluations. Given a finite amount of computational resources it is therefore advantageous to perform several shorter docking runs instead of few long cycles to localize the global minimum. One exception is lead *a*: Here, a docking position outside the cavity exists which is 20 kJ/mol lower than the best conformation inside the cavity. As a result the rate of docking inside the cavity remained very low, independent on the length of the cycle. Interesting is the fact that lead *e* was docking even faster and more reliable than MTX. After only 10^4 scorings a docking rate of 25% was reached. This may indicate that the configuration space of this particular receptor ligand system is especially favorable for reaching the global minimum.

4. Summary

In this investigation we have performed an automated virtual screening of 10 000 compounds, which were taken from the openly available database nciopen3D, to the active site of *E. coli* dihydrofolate reductase (4dfr). The global optimization technique STUN, which was already successful in the rigid ligand approximation, proved reliable and efficient also for the more demanding application to flexible ligands. It was able to match the natural ligand MTX (which contains 9 rotatable bonds) to an accuracy of 1.4 Å with the X-ray structure. The scoring function Eq. (1) was ranking this ligand as the best among the 10 000 screened molecules.

We have further demonstrated how a virtual database screening may serve as a tool for pharmacophore modeling, because the top ranking leads are useful probes to identify hot spots inside the cavity. Despite the fact that there appear to exist leads that docked faster and more reliable than MTX, none of them reached a similarly high affinity. We hope this study contributes to the search of new powerful techniques for virtual screening especially under the perspective of future projects which ask for additional degrees of freedom by means of rotatable groups in the receptor and the inclusion of solvent molecules. Our results motivate further investigations which employ STUN as a global optimization technique in such applications.

Acknowledgements

We thank the Fond der Chemischen Industrie, the BMBF, the Deutsche Forschungsgemeinschaft (Grant WE 1863/11-1) and the Kurt Eberhard Bode Stiftung for financial support.

References

[1] W. Walters, M. Stahl, M. Murcko, Drug Discov. Today 3 (1998) 160.
[2] C. Bissantz, G. Folkerts, D. Rognan, J. Med. Chem. 43 (2000) 4759.
[3] I. Kuntz, J. Blaney, S. Oatley, R. Langridge, T. Ferrin, J. Mol. Biol. 161 (1982) 269.
[4] E. Perola, K. Xu, T. Kollmeyer, S. Kaufmann, F. Prendergast, Y. Pang, J. Med. Chem. 43 (2000) 401.
[5] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470.
[6] W. Welch, J. Ruppert, A. Jain, Chem. Biol. 3 (1996) 449.

[7] C. Baxter, C. Murray, D. Clark, D. Westhead, M. Eldridge, Proteins 33 (1998) 367.
[8] T. Hou, L. Chen, X. Xu, Protein Eng. 12 (1999) 639.
[9] G. Jones, P. Wilett, R. Glen, A. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 721.
[10] G. Morris, D. Goodsell, R. Halliday, R. Huey, W. Hart, R. Belew, A. Olson, J. Comput. Chem. 19 (1998) 1639.
[11] P. Charifson, J. Corkery, M. Murcko, W. Walters, J. Med. Chem. 42 (1999) 5100.
[12] D. Gehlhaar, G. Verkhivker, P. Rejto, C. Sherman, D. Fogel, L. Fogel, S. Freer, Chem. Biol. 2 (1995) 317.
[13] D. Goodsell, A. Olson, Proteins 8 (1990) 195.
[14] M. Liu, S. Wang, J. Comput. Aided Mol. Des. 13 (1999) 435.
[15] C. McMartin, R. Bohacek, J. Comput. Aided Mol. Des. 11 (1997) 333.
[16] W. Wenzel, K. Hamacher, Phys. Rev. Lett. 82 (1999) 3003.
[17] U.H.E. Hansmann, Eur. Phys. J. B 12 (1999) 607.
[18] H. Merlitz, W. Wenzel, Chem. Phys. Lett. 362 (2002) 271.
[19] J. Sadowski, J. Gasteiger, Chem. Rev. 93 (1993) 2567.
[20] J. Sadowski, J. Gasteiger, G. Klebe, J. Chem. Inf. Comput. Sci. 34 (1994) 1000.
[21] J. Bolin, D. Filman, A. Matthews, R. Hamlin, J. Kraut, J. Biol. Chem. 257 (1982) 13650.
[22] W. Jorgensen, N. McDonald, J. Mol. Struct. 424 (1997) 145.
[23] S. Kirkpatrick, C. Gelatt, M. Vecchi, Science 220 (1983) 671.