Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 14 (2004) 4671–4676
Bioorganic & Medicinal Chemistry Letters

GAsDock: a new approach for rapid flexible docking based on an improved multi-population genetic algorithm

Honglin Li,[a] Chunlian Li,[a] Chunshan Gui,[b] Xiaomin Luo,[b] Kaixian Chen,[b] Jianhua Shen,[b,*] Xicheng Wang[a,*] and Hualiang Jiang[b,*]

[a]*Department of Engineering Mechanics, State Key Laboratory of Structural Analyses for Industrial Equipment, Dalian University of Technology, Dalian 116023, China*
[b]*Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201203, China*

Received 9 May 2004; revised 27 June 2004; accepted 29 June 2004
Available online 19 July 2004

Abstract—Based on an improved multi-population genetic algorithm, a new fast flexible docking program, GAsDock, was developed. The docking accuracy, screening efficiency, and docking speed of GAsDock were evaluated by the docking results of thymidine kinase (TK) and HIV-1 reverse transcriptase (RT) enzyme with 10 available inhibitors of each protein and 990 randomly selected ligands. Nine of the ten known inhibitors of TK were accurately docked into the protein active site, the root-mean-square deviation (RMSD) values between the docking and X-ray crystal structures are less than 1.7 Å; binding poses (conformation and orientation) of 9 of the 10 known inhibitors of RT were reproduced by GAsDock with RMSD values less than 2.0 Å. The docking time is approximately in proportion to the number of rotatable bonds of ligands; GAsDock can finish a docking simulation within 60 s for a ligand with no more than 20 rotatable bonds. Results indicate that GAsDock is an accurate and remarkably faster docking program in comparison with other docking programs, which is applausive in the application of virtual screening.

1. Introduction

Discovering new lead compounds through virtual screening of chemical databases targeting protein structures has shown great promise.[1,2] With the dramatic increase of pharmaceutical targets in recent years arising from the human genome project and high-throughput crystallography efforts, virtual screening will undoubtedly play an important role in identifying active ligands for the new coming protein targets.

Several virtual screening programs have been developed with primarily two parts varying: scoring functions[3–5] and searching (optimizing) methods.[6–12] Of the available docking programs, the most widely used are GOLD,[13] FlexX,[14] AutoDock,[15] and DOCK.[16] For any docking method, the primary criteria are docking accuracy (RMSD to known pose), scoring accuracy (prediction of the absolute binding free energy), screening efficiency (discrimination of active hits from random compounds), and computational speed (full conformation and orientation searching time).[17] Computational speed that strongly influenced by pose searching method is vital in virtual screening, only those able to dock a flexible ligand within a reasonable time scale (100–200 s) are suitable for virtual screening purpose. However, none of the present methods can offer a robust, accurate, and fast solution to the docking problem.[18]

In the following, we describe our newly developed docking methodology, GAsDock, which uses an entropy-based multi-population genetic algorithm to optimize the binding pose.[19] GAsDock may automatically dock a ligand into a protein binding (active) site considering the full flexibility of the ligand. Its rapid docking speed and excellent accuracy are efficient enough for virtual screening toward large-scale chemical databases.

Keywords: Docking; Genetic algorithm; Information entropy; Virtual screening.
* Corresponding authors. Tel.: +86-21-50807188/+86-411-84706223; fax: +86-21-50807088/+86-411-84708400; e-mail addresses: hljiang@mail.shcnc.ac.cn; guixum@dlut.edu.cn; jhshen@mail.shcnc.ac.cn

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.06.091

2. Materials and methods

2.1. Preparation of ligand databases

According to the strategy of Bissantz et al.[20] we first filtered the Available Chemicals Directory (ACD) database by eliminating chemical reagents, inorganic compounds and molecules with weights lower than 250 and higher than 500. Then we chose 990 molecules randomly from the database and generated their three-dimensional coordinates. For each molecule, a multi-mol2 file, which was obtained after added hydrogen atoms and optimized using Tripos force field and Gasteiger–Marsili atomic charges adopted in Sybyl6.8.[26] These molecules were appended to a TK library containing 10 available TK inhibitors (Fig. 1) and a RT library containing 10 RT known inhibitors (Fig. 2), respectively. Two final libraries, each containing 1000 molecules, were obtained accordingly.

2.2. Preparation of protein structures

We obtained the X-ray structures of TK in complex with deoxythymidine (PDB entry 1KIM)[21] and RT in complex with dmp-266 (efavirenz) (PDB entry 1JKH)[22] from the Protein Database Bank (PDB).[23] Residues around the bound ligand at a radius of 6.5 Å were isolated from the protein to define the active site. Then the energy-grid file was generated by adding hydrogen atoms and Kollman charge using Sybyl6.8.[26]

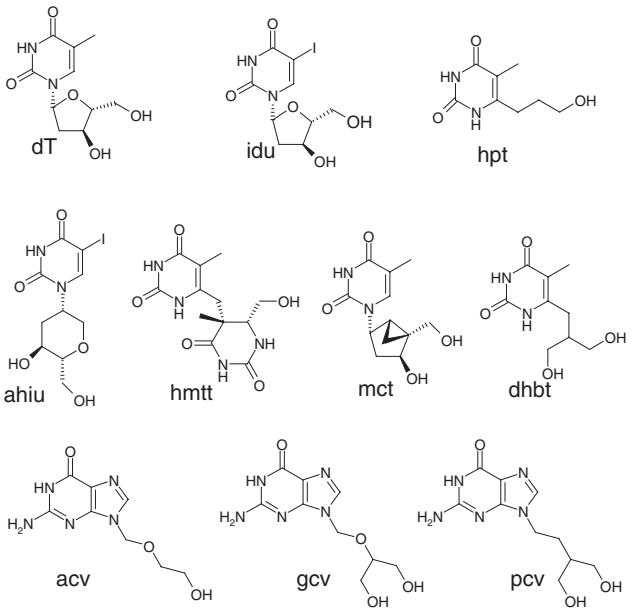

Figure 1. HSV-1 thymidine kinase inhibitors. The abbreviations are as follows: dT, deoxythymidine; idu, 5-iododeoxyuridine; hpt, 6-(3-hydrody-propyl-thymine); ahiu, 5-iodouracilanhydrohexitolnucleoside; mct, (North)-methanocarba-thymidine; hmtt, (6-[6-hydroxymethy-5-methyl-2,4-dioxo-hexahydro-pyrimidin-5-yl-methyl]-5-methyl-1*H*-pyrimidin-2,4-dione; acv, aciclovir; gcv, ganciclovir; pcv, penciclovir.

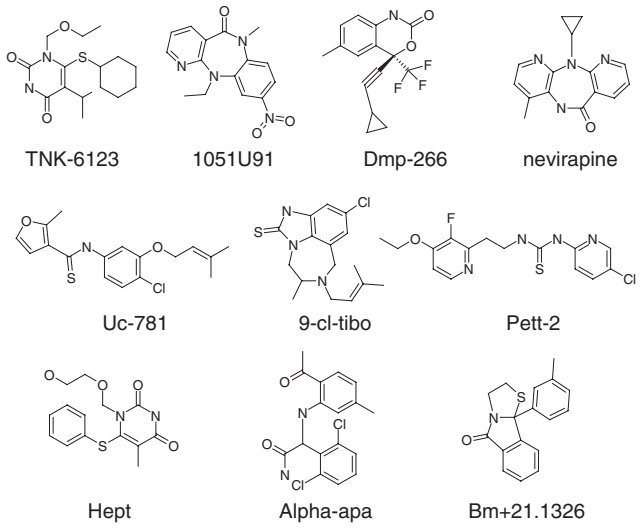

Figure 2. HIV-1 reverse transcriptase enzyme inhibitors.

2.3. Docking method with multi-population genetic algorithm based on information entropy

During the docking simulation, we assume that the ligand is flexible and the receptor is rigid, the orientation and conformation of a ligand are enclosed in a chromosome as state variables. One key component of docking is to find the best pose of a ligand to fit a protein binding pocket that is to minimize the binding affinity between the ligand and the protein. Therefore, we design this optimization problem as Eq. 1

$$\begin{aligned} &\min \quad f(\mathbf{d}) \\ &\text{s.t.} \quad g_j(\mathbf{d}) \leqslant 0 \quad j = 1, 2, \ldots, q \end{aligned} \tag{1}$$

where $\boldsymbol{d} = \{T_x, T_y, T_z, R_x, R_y, R_z, T_{b1}, \ldots, T_{bn}\}^T$ is a vector of design variables, in which (T_x, T_y, T_z) and (R_x, R_y, R_z) are respectively, the state variables of translation and rotation of the entire ligand for the orientation search; and $T_{b1}, \ldots, T_{bn}$ are the torsion angles of the n rotatable bonds of the ligand for the conformation search. The objective function $f(\boldsymbol{d})$ is the intermolecular interaction energy between the ligand and protein:

$$f(\mathbf{d}) = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332.0 \frac{q_i q_j}{D r_{ij}^a} \right) \tag{2}$$

where each term is a double sum over ligand atoms i and receptor atoms j; r_{ij} is the distance between atom i in ligand and atom j in receptor; A_{ij} and B_{ij} are van der Waals repulsion and attraction parameters, respectively; a and b are respectively, van der Waals repulsion and attraction exponents, q_i and q_j are point charges on atoms i and j; and D is dielectric function, and 332.0 is the factor that converts the electrostatic energy into kilocalories per mole.

For genetic algorithm (GA), each chromosome consists of the above three design variables that represent a ligand in a particular orientation and conformation. The

design space of (T_x,T_y,T_z) must be limited in the circumcuboid boundary of the receptor binding pocket. Accordingly, the constraints for the design variables ($g(\mathbf{d})$s) can be represented as

$$\begin{cases} \underline{X} \leqslant T_x \leqslant \overline{X} \\ \underline{Y} \leqslant T_y \leqslant \overline{Y} \\ \underline{Z} \leqslant T_z \leqslant \overline{Z} \\ -\pi \leqslant R_x, R_y, R_z, T_{\mathrm{b}1}, \ldots, T_{\mathrm{b}n} \leqslant \pi \end{cases} \tag{3}$$

where $(\underline{X},\underline{Y},\underline{Z})$ and $(\overline{X},\overline{Y},\overline{Z})$ are the minimum and maximum coordinates of the boundary space, respectively. The rest variables are allowed to vary between $-\pi$ and π rad.

The constraint optimization problem 1 can be transferred into an unconstraint optimization problem using quasi-exactness penalty function:

$$\varphi_{\psi}(\mathbf{d}) = f(\mathbf{d}) + (\alpha/\psi)\ln\left\{1 + \sum_{i=1}^{q} \exp(\psi g_i(\mathbf{d}))\right\} \tag{4}$$

where the ψ is a parameter chosen in the range of 10^3–10^5, and α is a positive penalty factor. Accordingly, the fitness function of GA, Eq. 1, can be written as

$$\max \quad F(\mathbf{d},\alpha) = C - \varphi_{\psi}(\mathbf{d},\alpha) \tag{5}$$

where C is a large positive number to ensure $F>0$. Therefore, Eq. 5 becomes the genetic evolution model for molecular docking optimization.

Multi-population strategy was used in our improved GA, that is m populations, each containing n' individuals, are randomly created in the searching space. Selection and mutation for each population are performed separately. However, cross-over is carried out between two different populations to keep the diversity of the population. Because of the randomness of genetic algorithms, the result of each population is not always the same during the procedure of each generation, which reflects different information for judging the optimal solution. This process is similar to the communication process of information theory[27,28] that is entropy decreases during the genetic procedure. Therefore, information entropy was adopted in the optimization process to rapidly narrow down the searching space. Thus, the optimization model of information entropy is constructed as

$$\begin{cases} \min\left(-\sum_{j=1}^{m} p_j F(\mathbf{x})\right) \\ \min H = -\sum_{j=1}^{m} p_j \ln(p_j) \\ \text{s.t.} \sum_{j=1}^{m} p_j = 1 \quad p_j \in [0,1] \end{cases} \tag{6}$$

where H is the information entropy, p_j is defined as the probability that the optimal solution of Eq. 3 occurs in the population j. It can be proved that the optimization problems 3 and 6 have the same optimal solution. p_j of Eq. 6 can be obtained easily using Eq. 7,

$$p_j = \exp(\gamma F_j(\mathbf{d}))/\sum_{j=1}^{m} \exp(\gamma F_j(\mathbf{d})) \tag{7}$$

where γ is a quasi-weight coefficient.

In practice, we use contracted space as the convergence criterion. The lower limit ($\underline{d}_i(K)$) and upper limit ($\bar{d}_i(K)$) of the contracted space are defined as follows:

$$\begin{aligned} \underline{d}_i(K) &= \max\left\{[d_i^*(K) - 0.5(1-p_j)D(K)], \underline{d}_i(0)\right\} \\ \bar{d}_i(K) &= \min\left\{[d_i^*(K) - 0.5(1-p_j)D(K)], \bar{d}_i(0)\right\} \end{aligned} \tag{8}$$

where $D(K)$ is the searching space of Kth iteration, which can be calculated by the value of the $(K-1)$th searching space, $D(K-1)$,

$$D(K) = (1-p_j)D(K-1) \tag{9}$$

Note that the design space is defined as the initial searching space $D(0)$. $(1-p_j)$ can be taken as the contracted coefficient of the searching space. The optimization is preformed by genetic iteration till the convergence criterion reaches to a given precision, then the obtained result of evolution is the optimal solution.

3. Results and discussion

GAsDock was evaluated by docking accuracy, screening efficiency and docking speed. Docking accuracy and screening efficiency were assessed by two protein targets, TK and HIV-1 RT, with 10 known ligands and 990 random ligands for each protein, respectively. The GAsDock results of TK were compared with the results derived by GOLD,[20] FlexX,[20] DOCK,[20] Surflex,[17] and Glide.[18]

3.1. Result of thymidine kinase

Thymidine kinase (TK) was chosen to test the availability of GAsDock because its binding cavity was accessible to water and some side chains of the active site adopt rotameric states for different bound ligands. In addition, this protein has been used to test the effectiveness of several other docking programs,[17,18,20] thus there are enough data for comparison.

Ten TK inhibitors (Fig. 1) were docked into the active site of TK by GAsDock, and the resultant binding poses were superposed with their binding poses in the X-ray crystal structures of TK-inhibitor complexes. The calculation results are listed in Table 1 and Fig. 3, which indicate that 7 of the 10 TK inhibitors were docked accurately with the active site of TK, the RMSD values between the GAsDock poses and crystal poses are less than 1.35 Å. Of the three purines, two compounds (acv and gcv) may dock into the TK active site in an acceptable accuracy, their RMSD values are less than 1.7 Å. Only one purine compound (pcv in Fig. 1) did not find its true binding pose, its RMSD is more than 5 Å. The reason of this is that, in the X-ray crystal structure of purine–TK complexes, the side chain of Gln125 flipped

Table 1. Docking accuracy of TK inhibitors by GAsDock and other docking programs[a]

Inhibitor	RMSD of each docking method (Å)					
	GAsDock	Glide	DOCK	FlexX	GOLD	Surflex
dT	**0.19**	0.45	0.82	0.78	0.72	0.74
ahiu	**0.42**	0.54	1.16	0.88	0.63	0.87
mct	**0.98**	0.79	7.56	1.11	1.19	0.87
dhbt	**1.14**	0.68	2.02	3.65	0.93	0.96
idu	**0.28**	0.35	9.33	1.03	0.77	1.05
hmtt	**1.35**	2.83	9.62	13.30	2.33	1.78
hpt	**0.70**	1.58	1.02	4.18	0.49	1.90
acv	**1.53**	4.22	3.08	2.71	2.74	3.51
gcv	**1.62**	3.19	3.01	6.07	3.11	3.54
pcv	**5.96**	3.53	4.10	5.96	3.01	3.84

[a] Data of DOCK, FlexX, GOLD were taken from Ref. 20 data for Surflex were taken from Ref. 17; data for Glide were taken from Ref. 18.

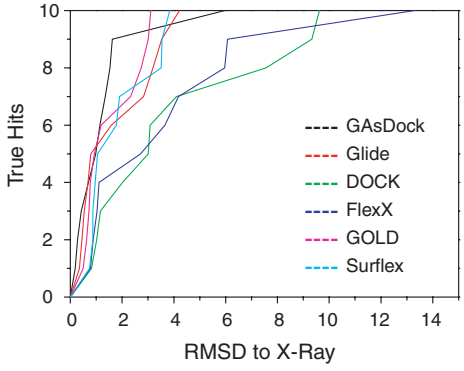

Figure 3. RMSD of docked TK true hits from their X-ray poses.

about 180° in comparison with the crystal structures of other inhibitor–TK complexes. Figure 3 presents the plots of the RMSD values derived by GAsDock and other docking programs versus corresponding compound numbers. As can be seen from Figure 3, GAsDock reproduced 90% of the X-ray crystal poses of the 10 TK inhibitors with RMSD values less than 2.0 Å. This demonstrates that the GAsDock results are much better than that of other docking programs (Table 1 and Fig. 3).

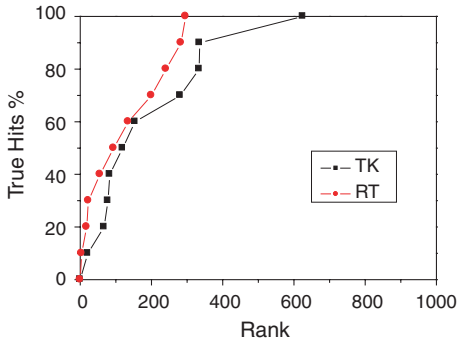

Figure 4. Ranking and true hits rate of TK and RT inhibitors.

Figure 4 shows the scoring position of the 10 TK inhibitors among the 1000 compounds (10 TK inhibitors and 990 randomly selected compounds) docked by GAsDock to TK, 50% true TK inhibitors were vested in the top 10% scorers, which is much higher than other docking programs.

3.2. Result of HIV-1 reverse transcriptase

In contrast to TK, HIV-1 RT has a different binding pocket, which is more hydrophobic, represented by a larger (900–1000 Å^2) solvent-accessible cavity.[24,25] This polymerase active-site was closely associated with the substrate binding site. Many crystal structures of inhibitor–RT complexes are available for testing docking programs. To further demonstrate the accuracy and efficiency of GAsDock, we selected 10 HIV-1 RT inhibitors and docked them into the active site of the protein (PDB entry 1JKH) using GAsDock, DOCK, and FlexX, the results are listed in Table 2.

GAsDock reproduced the binding poses of nine inhibitors in the X-ray crystal structures of inhibitor–protein complexes with RMSD values less than 2.0 Å, but DOCK and FlexX only reproduced the binding poses of four and two inhibitors, respectively. Top 10% scorers of the 1000 compounds cover 50% of the 10 true inhibitors (Fig. 4 and Table 2). This result demonstrates again the effectiveness of GAsDock.

Table 2. Docking accuracy and speed of HIV-1 RT inhibitors by GAsDock, DOCK and FlexX

Ligand	PDB entry	N_{tor}[a]	GAsDock			DOCK[c]			FlexX[c]		
			CPU time (s)	RMSD	RANK[b]	CPU time (s)	RMSD	RANK[b]	CPU time (s)	RMSD	RANK[b]
Dmp-266	1JKH	3	5.54	0.39	4	19.06	0.88	21	4.90	0.97	10
Neritapine	1VRT	1	3.07	1.7	18	4.99	1.97	192	8.25	9.72	484
TNK-6123	1C1C	6	11.58	0.83	23	53.04	1.45	774	27.95	12.51	394
Bm+21.1326	1C0T	1	5.52	1.17	56	12.22	2.32	657	6.51	10.91	327
Uc-781	1JLG	7	7.61	0.78	94	48.50	5.18	569	23.22	1.53	110
hept	1RTI	7	7.21	1.30	135	54.23	5.73	721	10.61	2.19	328
Pett-2	1DTT	8	7.38	1.04	200	68.84	6.43	934	16.11	11.86	160
1051U91	1LW2	2	4.41	4.98	241	14.45	4.90	387	8.26	11.14	401
9-cl-tibo	1REV	3	5.80	1.98	283	18.59	1.78	531	8.75	10.89	583
Alpha-apa	1VRU	6	9.77	1.61	296	63.13	16.88	386	93.92	14.33	428

[a] N_{tor} is the number of rotatable bonds.
[b] The rank of the inhibitor in 1000 compounds containing 990 random compounds.
[c] Default parameters are used for DOCK and FlexX.

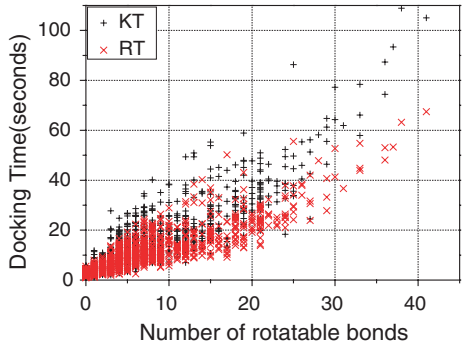

Figure 5. GAsDock time versus the number of rotatable bonds.

3.3. Docking speed

Docking speed of different programs should not be compared directly because of different computer hardware and system. However, it is a key point for the application of a docking method in virtual screening. Therefore, improving the computational speed of a docking method is of significance. Jain[17] estimated that FlexX, DOCK, and GOLD[20] need 50–100 s to dock a ligand into the active site of a protein by using a single CPU computer.[17,20] The computational time of GAsDock is approximately in proportion to (linear scale) the rotatable bond number of ligand (Fig. 5). For all the testing compounds used in this study, the number of rotatable bonds range from 0 to 41. The total docking time of GAsDock for the 1000 testing compounds to TK is 15597.2 s using one CPU on SGI Origin3800 hardware, the average docking time is about 15.2 s for docking one molecule using one CPU. As to individual molecule, the maximum and minimum docking time are 108.1 s (41 rotatable bonds) and 0.9 s (0 rotatable bond), respectively. The total docking time of GAsDock for the 1000 molecules to HIV-1 RT active site is about 11348.9 s, with an average time of 11.35 s for an individual molecule, the maximum and minimum time are 67.48 and 0.97 s, respectively. This indicates that GAsDock is fast enough for virtual screening on large-scale chemical databases.

In comparison with the optimization algorithms of other docking methods, information entropy was employed in the genetic algorithm of GAsDock and contracted space was used as the convergence criterion, which effectively controls the convergence of the algorithm, ensuring that GAsDock can converge rapidly and steadily. This is the main reason that GAsDock can bring better results in accuracy and higher speed than other programs (Tables 1 and 2).

4. Conclusions

We have demonstrated the accuracy and efficiency of our recently developed docking program GAsDock, which uses an effective optimization method, entropy-based multi-population genetic algorithm. Testing calculations of 20 inhibitors docking into the active sites of TK and HIV-1 RT (10 inhibitors for each protein) indicate that GAsDock is more accurate than other docking programs like GOLD,[20] FlexX,[20] DOCK,[20] Surflex,[17] and Glide.[18] The computational time of GAsDock is linearly in proportion to the rotatable bonds of compounds, and is fast enough for searching large chemical databases targeting special proteins. Although GAsDock shows applausive prospect in virtual screening, there are still several respects needed to be improved, such as the scoring function and the flexibility of targeting proteins. We will fulfil these tasks in the future version of GAsDock.

Acknowledgements

We gratefully acknowledge financial support from National Natural Science Foundation (10272030), the Special Funds for Major State Basic Research Project (G1999032805) of China, the 863 Hi-Tech Program (Nos. 2002AA233011, 2002AA104270), and the State Key Program of Basic Research of China (No. 002CB512802).

References and notes

1. Chen, K. X.; Jiang, H. L.; Ji, R. Y. In *Computer aided drug design—principle, method and application,* 1st ed.; Shanghai Scientific and Technical Publishers: Shanghai, 2000; Vol. 1, pp 1–466.
2. Shen, J. H.; Xu, X. Y.; Cheng, Y.; Liu, H.; Luo, X. M.; Shen, J. K.; Chen, K. X.; Zhao, W. M.; Shen, X.; Jiang, H. L. *Curr. Med. Chem.* **2003**, *10*, 2327.
3. Bohm, H. J. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243.
4. Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791.
5. Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337.
6. Welch, W.; Ruppert, J.; Jain, A. N. *Chem. Biol.* **1996**, *3*, 449.
7. Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1.
8. Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y. P. *J. Med. Chem.* **2000**, *43*, 401.
9. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317.
10. McMartin, C.; Bohacek, R. *J. Comput. Aided Mol. Des.* **1997**, *11*, 333.
11. Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins* **1998**, *33*, 367.
12. Liu, M.; Wang, S. *J. Comput. Aided Mol. Des.* **1999**, *13*, 435.
13. Jones, G.; Wilett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
14. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
15. Morris, G. M.; Goodsell, D. S.; Halliday, R.; Huey, R.; Hart, W. E.; Belew, R. K.; Arthur, J. *J. Comput. Chem.* **1998**, *19*, 1639.
16. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
17. Jain, A. N *J. Med. Chem.* **2003**, *46*, 499.
18. 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.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739.

19. Li, C. L.; Wang, X. C.; Zhao, J. C. *Comput. Eng.* **2003**, *39*, 31.
20. Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759.
21. Champness, J. N.; Bennett, M. S.; Wien, F.; Visse, R.; Summers, W. C.; Herdewijn, P.; Clerq, E.; Ostrowski, T.; Jarvest, R. L.; Sanderson, M. R. *Proteins* **1998**, *32*, 350.
22. Prabu, J. M.; Nalivaika, E. A.; Schiffer, C. A. *Structure* **2002**, *10*, 369.
23. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids Res.* **2000**, *28*, 235.
24. Campiani, G.; Ramunno, A.; Maga, G.; Nacci, V.; Fattorusso, C.; Catalanotti, B.; Morelli, E.; Novellino, E. *Curr. Pharm. Des.* **2002**, *8*, 615.
25. Chamberlain, P. P.; Ren, J.; Nichols, C. E.; Douglas, L.; Lennerstrand, J.; Larder, B. A.; Stuart, D. I.; Stammers, D. K. *J. Virol.* **2002**, *8*, 615.
26. Sybyl (molecular modeling package), version 6.8. Tripos Associates: St. Louis, MO, 2000.
27. Mackay, D. J. C. Information Theory, Inference, and Learning Algorithms, version 6; Cambridge university press: London, 2003; pp 3–269.
28. Adami, C. *Phys. Life Rev.* **2004**, *1*, 3.