Impact of receptor conformation on in silico screening performance

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

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

Abstract

We report on results for the in silico screening of a database of 10 000 flexible compounds against various crystal structures of the thymidine kinase enzyme complexed with 10 known inhibitors. We provide a quantitative analysis of the deviations in the ranking of the inhibitors depending on the choice of receptor conformation and imply that the inclusion of side chain degrees of freedom to the receptor would significantly improve the predictive power of the screening approach. We suggest a consensus score that, in the case of several known native structures of the receptor, enables the evaluation of scoring functions without the requirement of explicit receptor flexibility.

1. Introduction

As a result of the rapidly growing number of resolved three-dimensional target structures [1], virtual screening of large chemical databases is evolving into an increasingly important tool for drug discovery [2]. The success of in silico-screening methods critically depends on the docking algorithm [3] and the quality of the scoring function (SF). A large number of qualitatively different SFs [4] have been proposed, many of which have been used successfully to generate leads in drug discovery applications. The continued improvement of such functions will undoubtedly lead to further progress in drug discovery. However, as of yet, no clear consensus regarding even the essential ingredients to these functions has been reached. Recent studies comparing a large variety of different scoring functions remained inconclusive [5]. In many screens, the degree of database enrichment for various known ligands to a given structurally characterized receptor remains disappointing.

The limitations of presently available computational resources and the large number of possible ligands enforce severe approximations in the representation of receptor and ligand, and their interactions. Significant computational efficiency is gained, when the protein receptor is assumed to be rigid in the docking process; for this reason many tests of screening functions [5] and virtually all large scale computational screens presently rely on a rigid-receptor conformation. On the other hand, direct comparison between ligand-free and complexed crystal structures often demonstrate a significant ligand-induced alteration of the receptor structure.

At least partial receptor flexibility has been implemented in a number of docking methods [6 8], but is not yet widely available. Consensus scoring [4], i.e., the use of several scoring functions at the same time, offers an ad hoc improvement of the reliability of the screening process, but remains fundamentally unsatisfactory. To further aid the development of scoring functions and docking methods it is, therefore, useful to investigate and clarify the impact of receptor flexibility for known receptor/ligand pairs. Particularly useful for this purpose is the study of receptors for which several structurally resolved high-affinity ligands exist.

In this investigation we report screens of a database of 10 000 ligand molecules to the family of thymidine-kinase (TK) enzymes, one of which was recently used in an evaluation of the accuracy of different scoring functions [5] and of the impact of mediating crystal water molecules [9]. The TK is a very useful benchmark

* Corresponding author. Present address: Forschungszentrum Karlsruhe, Institute fuer Nanotechnologie, Hermann von Helmholtz Platz 1, Leopoldshafen, Germany. Fax: +497247826434.

E mail address: merlitz@int.fzk.de (H. Merlitz).

URL: http://www.fzk.de/biostruct/ (W. Wenzel).

system, because X-ray structures of the target are available in complex with all of the 10 inhibitors. One particular enzyme structure was generally able to dock some, but not all established inhibitors. Such a failure to dock known inhibitors into receptor conformations associated with other ligands or into a ligand-free receptor conformation may, in principle, result from (i) inadequacies in the docking algorithm, (ii) inadequacies of the scoring function (including conserved water molecules and other compounds) or (iii) lack of receptor flexibility.

In this study we focus on the errors rooted in the neglect of receptor degrees of freedom, and demonstrate that these are dominant at least in some of the investigated enzyme/inhibitor combinations. An ad hoc consensus ranking is proposed that would ameliorate this problem in those special cases when several native receptor structures are known, which then provide a suitable test platform for the evaluation and optimization of scoring functions.

2. Method

The ligands were docked using the stochastic tunneling method (STUN) [10] with flexible ligands (free rotatable bonds). This method was shown to be superior to other competing stochastic optimization methods[11] and had performed adequately in a screening of 10 000 ligands to the active site of dihydrofolate reductase (pdb code 4dfr [12]), where the known inhibitor (methotrexate) emerged as the top scoring ligand [13].

The stochastic tunneling technique was proposed as a generic global optimization method for complex rugged potential energy surfaces(PES). In STUN the dynamical process explores not the original, but a transformed PES, which dynamically adapts and simplifies during the simulation. For the simulations reported here we replace the original transformation [10] with

$$E_{\mathrm{STUN}} = \ln\left(x + \sqrt{x^2 + 1}\right), \tag{1}$$

where $x = \gamma(E - E_0)$, E is the energy of the present conformation and E_0 the best energy found so far. The problem-dependent transformation parameter γ controls the steepness of the transformation [10]. The general idea of this approach is to flatten the potential energy surface in all regions that lie significantly above the best estimate for the minimal energy (E_0). Even at low temperatures the dynamics of the system becomes diffusive at energies $E \gg E_0$ independent of the relative energy differences of the high-energy conformations involved. The dynamics of the conformation on the untransformed PES then appears to 'tunnel' through energy barriers of arbitrary height, while low metastable conformations are still well resolved. Applied to receptor ligand docking this mechanism ensures that the ligand can reorient through sterically forbidden regions in the receptor pocket.

We employed the following simple, first-principle scoring function

$$S = \sum_{\text{Protein}} \sum_{\text{Ligand}} \left(\frac{R_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}} \right) + \sum_{\text{h-bonds}} \cos\Theta_{ij} \left(\frac{\tilde{R}_{ij}}{r_{ij}^{12}} - \frac{\tilde{A}_{ij}}{r_{ij}^{10}} \right), \tag{2}$$

which contains the empirical Pauli repulsion (first term), the Van de Waals attraction (second term), the electrostatic potential (third term) and an angular dependent hydrogen bond potential (term four and five). The Lennard Jones parameters R_{ij} and A_{ij} were taken from OPLSAA [14], 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 [15]. This force field lacks solvation terms to model entropic or hydrophobic contributions. The omission of such terms has been argued to be appropriate for constricted receptor pockets in which all ligands with high affinity displace essentially all water molecules.

None of the stochastic optimization methods is able, within a finite number of simulation steps, to find a global minimum with certainty. Instead, there exists only a probability to enter a given energy interval around this minimum, and a repeated docking of the same ligand, therefore, leads to a distribution of binding energies. To keep the fluctuations within a tolerable range, each screen was repeated six times and only the lowest energy was used for ranking the ligands. Repeated test simulations with some of the inhibitors and randomly chosen database ligands have shown that in this way the root-mean-square fluctuations of the binding energies were of the order of 10 15 kJ/mol, so that the screen was able to differentiate sufficiently well between ligands with high and low affinity.

3. Results

Under investigation was the degree of database enrichment of 10 000 compounds, randomly chosen from the nciopen3D database [16], and 10 known inhibitors when docked to the X-ray TK enzyme structure, which was experimentally determined in complex with one of the inhibitors, dt (deoxythymidine, pdb entry 1kim [17]). In this screen 5353 ligands attained a stable conformation with negative binding energy within the receptor pocket. Fig. 1 shows the number of ligands as a function of affinity and highlights the positions of the known TK inhibitors in the screen. Three structurally similar inhibitors, including the ligand associated with the receptor conformation, were ranked with very high

affinity. This result suggests that docking method and scoring function were adequate to approximate the affinity of these ligands to the enzyme. However, four other ligands (idu, acv, gvc, pcv, for a detailed description of TK and its inhibitors we refer to [5]) docked badly, and three further ligands did not dock at all according to the criteria above.

The resulting ranks of this screen are summarized in Table 1 (second column), which displays the rankings of the 10 inhibitors. Three were ranked within the first 1%, 6 were ranked among the first 10% of the database, respectively. This enrichment rate is comparable to the results of other scoring functions that were previously investigated for this system [5]. One one hand, this was surprising, given the simplicity of the scoring function Eq. (2), on the other hand, its complete failure to select some of the inhibitors seemed to indicate shortcomings which were to be traced.

Inspection of the crystal structures of the different receptor ligand complexes revealed differences in the conformation of some side groups inside the receptor pocket, depending on the docked inhibitor. This is a well known fact, but it was often assumed that the impact of these conformational variations on the ranking accuracy was moderate. We then repeated the screening with the X-ray structure of TK in complex with the inhibitor gcv (ganciclovir, pdb entry: 1ki2 [17]), which had scored particularly bad in the original screen. The results are shown in Table 1 (third column). Now, gcv was ranked within the leading 1% of the database, but dt, formerly ranked on position 5, dropped to 1310.

Fig. 2 displays identical sections of the docking sites of 1kim (with ligand dt, top panel) and 1ki2 (with ligand gcv, central panel). It is clearly discernible how the amide group of Gln125 had flipped up to allow for an optimized interaction with gcv. Glu225, originally creating an h-bond with dt, had moved aside to give space for the bulkier gcv ligand. His58 had turned around to create a new h-bond with gcv. The modifications in the receptor were significant enough to make 1kim incompatible with gcv and 1ki2 to dt. Next the screen was repeated with the X-ray structure of TK in complex with pcv (penciclovir, pdb entry: 1ki3 [17]), which in the former screens was ranked on positions 4845 and 952. The results are shown in column 4 of the Table 1: pcv now appeared among the top scoring ligands at position 4 of the list, but dt, dhbt and hpt dropped significantly in rank (compared to 1kim). In all of these screens the natural ligands ranked well in their their natural receptor conformation, but badly in others. It appears impossible to optimize a scoring function in a way that these modifications inside the receptor could be compensated for. For such a system, no single receptor conformation can be used to identify all of the inhibitors in a single screen.

Common workarounds are the application of less steep than 6 12 potentials or the scaling down of the vdW radii in the force field. These approaches in fact enable additional inhibitors to enter a rigid receptor, but also increase the chance of false positives. If a key doesn't enter a lock, it is not sufficient to simply scale down the size of the key or to widen the lock. However, it has been demonstrated how a united protein description generated from a superimposed structure of the receptor ensemble [7] or a careful choice of the most 'promiscuous' receptor [18] led to promising results for systems which exhibit small side chain movements.

One example of a notoriously unsuccessful candidate was hmtt (6-[6-hydroxymehty-5-methyl-2,4-dioxo-hexahydro-pyrimidin-5-yl-methyl]-5-methyl-1H-pyrimidin-2,4-dione, in complex with pdb entry 1e2n [19]). hmtt is an especially bulky ligand which docked neither to 1kim nor to 1ki2 and 1ki3, because it was sterically prohibited to fit into the pockets. We, therefore, re-screened the database using its natural receptor conformation. The resulting ranking is displayed in column 5 of Table 1. In this screen hmtt was classified as docking, as sterical clashes with the

Table 1
Ranking of the TK inhibitors in a screen of 10 000 randomly chosen ligands of the nciopen3D database

Inhibitor	1kim	1ki2	1ki3	1e2n[a]	1e2n[b]	1e2h	All
acv	719	9	22	168	270	2048	41
ahiu	nd[c]	nd	nd	4561	4345	nd	5968
dhbt	4	104	118	315	464	38	18
dt	5	1310	2576	835	801	2779	23
gcv	3351	78	15	18	276	4516	63
hmtt	nd	nd	nd	5153	2126	nd	3605
hpt	6	152	266	302	493	36	28
idu	515	2436	3272	2474	1877	2913	1247
mct	nd	6074	nd	2927	2178	nd	3669
pcv	4845	952	4	234	517	4739	19
Score	3751	3705	4575	4128	3179	1926	5808

The top row designates the crystal structure of the receptor that was used in the screen, the last column is the consensus rank described in the text.
[a] Receptor with sulfate ion removed.
[b] Receptor including (fixed) sulfate ion.
[c] nd, Not docked.

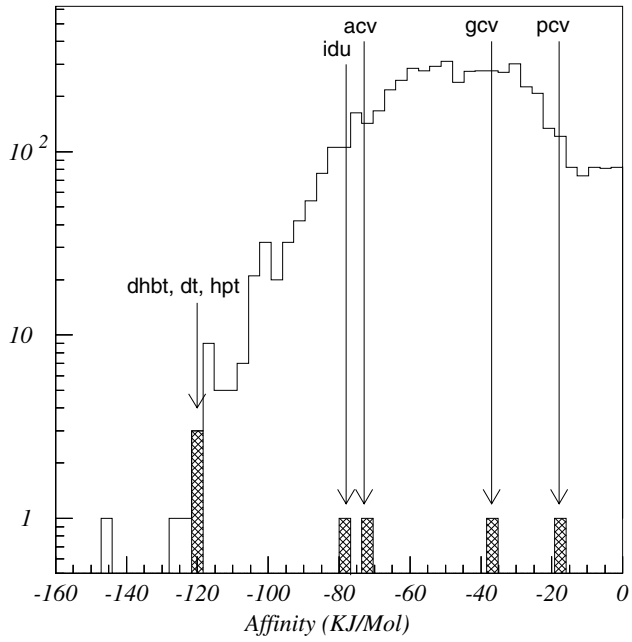

Fig. 1. Binding energies of the 5353 docked ligands (see text) to the enzyme conformation complexed with deoxythymidine (pdb code: 1kim).

receptor atoms were now absent, but its ranking remained very poor. A detailed inspection of the crystal structure revealed the presence of a conserved sulfate-ion that coordinated with hmtt (Fig. 2, bottom panel). This sulfate-ion appeared also in the other crystal structures, but formed no hydrogen bonds with dt, gcv or pcv. In an attempt to better model the receptor pocket, this ion was then included as a fixed additional molecule in the receptor and the screen repeated. Column 6 shows an improved, yet still unsatisfactory ranking for hmtt, which may reflect a shortcoming in the treatment of solvation/de-solvation effects in the forcefield. It was noted that a number of small compounds within the database created a strong ionic bond with the sulfate and consequently scored better than most of the inhibitors, whose overall ranking was, therefore, dropping.

For comparison purposes, we finally performed a screen of the ligand free X-ray structure of TK (pdb entry: 1e2h [20]), which would most likely be used in a screen if no inhibitor was known. In this screen the receptor was unbiased to any of the inhibitors, which resulted in a significant loss of screening performance. As shown in column 7 only two ligands scored reasonably well (within the upper 10% of the database), all others would be discarded by any rational criterion as possible lead candidates.

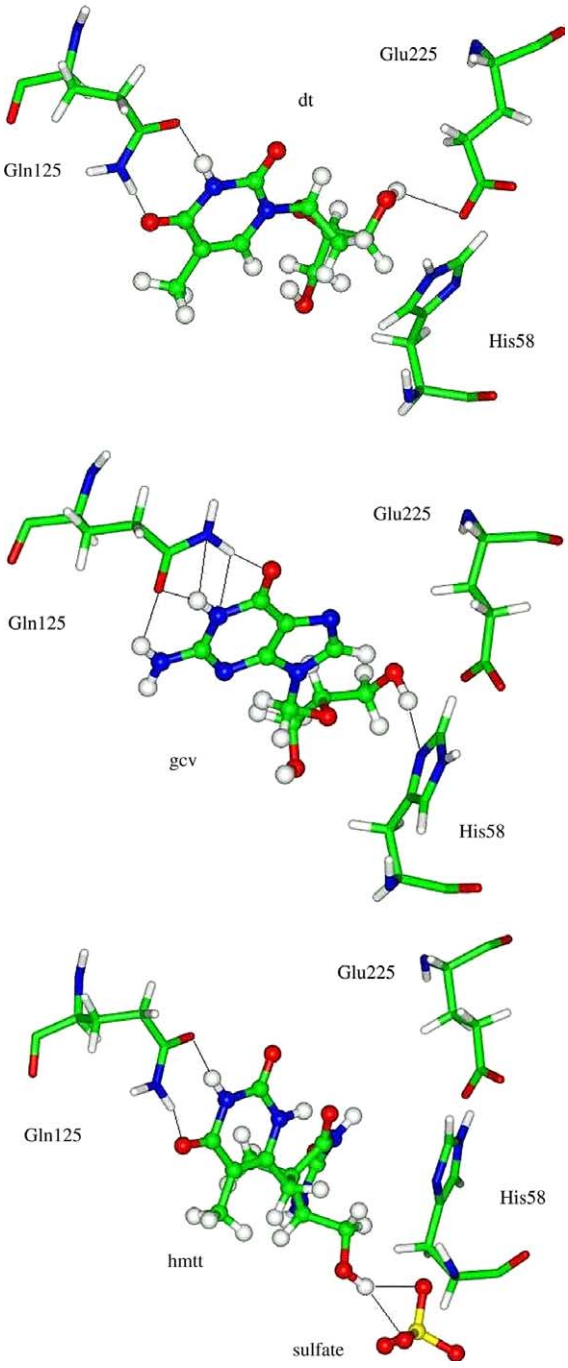

Fig. 2. Binding of dt (top) to the receptor pocket of 1kim with an indi cation of the h bonds formed in the complex and binding of gcv (center) to the receptor pocket of 1ki2. The amide group of the sidechain of Gln125 has turned up to form a hydrogen bond, the sidechain of Glu225 has moved to the right and the sidechain of His58 has turned to create another h bond. Binding of hmtt (bottom) to its natural receptor conformation (pdb entry 1e2n). Compared to gcv (center), His58 made a half turn and the amide group of Gln125 flipped. Also shown is the stabilizing sulfate ion present in the experimental crystal structure. Oxygen atoms are shown in red, nitrogen atoms in blue.

4. Conclusions and perspectives

The results offer a good demonstration that the ranking of known inhibitors can strongly depend on the particular receptor structure used for the screen. The differences in affinity and rank of a given ligand in different receptor conformations were of the same order of magnitude as the affinity of the best ligands. Any high

affinity ligand could, therefore, rank either at the very top of the database or somewhere in its tail. With repeated docking it was easily verified that the docking algorithm could not be made responsible for these fluctuations, and the fact that several inhibitors scored on top of the database inside their native enzyme conformations indicates that the scoring function was accurate enough to identify the close fit of the receptor ligand complex. The example of hmtt, which scored poorly into its native enzyme conformation, clearly indicates that the conformation of the receptor, though responsible for sterical clashes in several instances, could not be made responsible for all deficiencies in the screening methodology. The sulfate-ion that coordinated with hmtt turned out to be an additional factor, and also certain conserved crystal water molecules have been identified to influence the docking performance in TK [9]. Both are also liable to conformational changes, of course, and hence support the necessity of methods which take into account some of the receptor degrees of freedom.

Regarding the evaluation of scoring methodologies and the validation of scoring functions the results of the screen of the unbiased ligand-free receptor structure were alarming. The poor ranking of the known ligands in this screen indicates that high enrichment rates for rigid receptor screens against a receptor conformation complexed with a known ligand were fortuitous. The high ranking obtained for some of the known inhibitors (such as in column 2 of Table 1) was essentially a result of the restriction of the search space which was particularly favorable for these inhibitors. In the absence of such a restriction (column 6 of the Table 1), the enrichment rate for the same scoring function dropped dramatically. As a consequence, a good enrichment rate in a rigid receptor screen does not necessarily validate a scoring function even for the system under consideration.

These findings suggest the importance of a flexible binding pocket to obtain a better unbiased scoring of high-affinity ligands. Ultimately, only the consideration of receptor flexibility, at least of sidechain dihedral rotations, in the docking procedure will fundamentally address this problem. However, the incorporation of such degrees of freedom will dramatically increase the dimension of the docking search space and render some of the most efficient docking methods used today inapplicable. Particularly affected are methods which are based on a deterministic discretization of the conformational space, including the class of multi-conformer approaches [21]. Even with efficient generic stochastic optimization methods (e.g., STUN [10], tabu search [22], genetic algorithms [23]), which can deal with larger numbers of degrees of freedom, the computational cost of the screening procedure is increasing substantially. The extension of the search space also increases the burden on the scoring function which must now differentiate between many more possible ligand conformations.

Before flexible receptor methods will be widely available and computationally affordable, we suggest to apply a consensus ranking of ligands in different receptor conformations to reduce the bias in rigid-receptor screens. The strategy is this: A table is set up in which each ligand (including all database ligands) is associated with the best rank obtained in any of the individual screens. The resulting table is then sorted according to their rank. The position of each ligand in this list, the 'consensus rank' is given in the last column of Table 1. Note that the score of a particular ligand in this ranking is not necessarily the optimal score obtained in the other columns, as some database ligand not shown in the table might have scored better than this particular ligand in one of the other screens. In this consensus ranking, 6 of the 10 known inhibitors were now placed within the upper one percent of the database. Because the ranking procedure for the consensus score is nonlinear, a ligand which docked well in one screen and poorly in others is preferred to other ligands which docked fairly well in all screens.

To quantitatively compare different screens against the same ligand database, which used different receptor geometries, scoring functions or docking methods, it is sensible to assign an overall score to each screen which rates its performance [24]. We computed such a 'score' for the entire screen from the ranks of the docked known inhibitors among the $N = 1000$ best ligands. This score was computed as the sum of $N - P$ (where P was the rank of the known inhibitor) and dumped into the bottom row of Table 1. Since only the best N inhibitors were evaluated, screens which docked many known inhibitors with moderate rank may have got comparable scores with screens which performed perfectly for one inhibitor, but failed for all others. For the individual screens performed here, the scores ranged from between 1926 for the screen against 1e2h, the ligand free X-ray structure of TK, to 4575 (1ki3, X-ray structure of TK in complex with pcv), which was, therefore, the best performing screen of all receptor conformations. According to this measure the consensus ranking scheme performed better than any individual screen with a score of 5808, which represents a 29% (64%) improvement over the best (average) individual screen, respectively.

5. Summary

The investigation of the TK enzyme family furnished a clear example for the impact of receptor conformation in rigid receptor screens. Docking a database including a known ligand into the native receptor conformation of this ligand induces a bias into the screening procedure which tends to overestimate the accuracy of the screening methodology. As evidenced in the present study, the use of an unbiased receptor conformation

significantly reduced the overall screening performance of the score. If an entire set of known ligands is used to validate a scoring function, one particular member of this set has got an advantage over the rest.

This implies that the evaluation and comparison of scoring functions by screening a database plus certain known inhibitors to a single receptor conformation is not justified unless it was verified in advance that this receptor remained rigid in its natural environment. For all other cases we have suggested a consensus scoring which significantly improves the accuracy of the evaluation if several native conformations of enzyme/inhibitor complexes are known: The screen is carried out with each of the receptor conformations using the same scoring function. Here, even crystal water molecules or other conserved ions could be included in each of the structures with their respective coordinates. The final consensus score then has to be compared to the ones gained with other scoring functions, again applied to the entire set of conformations. Although the bias against the database ligands remains present, the same is eliminated between the inhibitors which are now treated on the same footing. In this case the assumption that all inhibitors have to score well is finally justified, and the comparison and optimization of different scoring functions becomes feasible.

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] J. Drews, Science 287 (2000) 1960.
[2] W. Walters, M. Stahl, M. Murcko, Drug Discov. Today 3 (1998) 160.
[3] G. Schneider, H. Boehm, Drug Discov. Today 7 (2003) 64.
[4] G. Klebe, H. Gohlke, Angew. Chem. Int. Ed. 41 (2002) 2644.
[5] C. Bissantz, G. Folkers, D. Rognan, J. Med. Chem. 43 (2000) 4759.
[6] V. Schnecke, L. Kuhn, Persp. Drug Des. Discov. 20 (2000) 171.
[7] H. Claussen, C. Buning, M. Rarey, T. Lengbauer, J. Mol. Biol. 308 (2001) 377.
[8] F. Osterberg, Proteins 46 (2002) 34.
[9] P. Pospisil, T. Kuoni, L. Scapozza, G. Folkers, J. Recept. Signal. Transduct. Res. 22 (2002) 141.
[10] W. Wenzel, K. Hamacher, Phys. Rev. Lett. 82 (1999) 3003.
[11] H. Merlitz, W. Wenzel, Chem. Phys. Lett. 362 (2002) 271.
[12] J. Bolin, D. Filman, A. Matthews, R. Hamlin, J. Kraut, J. Biol. Chem. 257 (1982) 13650.
[13] H. Merlitz, B. Burghardt, W. Wenzel, Chem. Phys. Lett. 370 (2003) 68.
[14] W. Jorgensen, N. McDonald, J. Mol. Struct. 424 (1997) 145.
[15] G. Morris, D. Goodsell, R. Halliday, R. Huey, W. Hart, R. Belew, A. Olson, J. Comput. Chem. 19 (1998) 1639.
[16] G. Milne, M. Nicklaus, J. Driscoll, S. Wang, D. Zaharevitz, J. Chem. Inf. Comput. Sci. 34 (1994) 1219.
[17] J. Champness, M. Bennett, F. Wien, R. Visse, W. Summers, P. Herdewijn, E. de Clerq, T. Ostrowski, R. Jarvest, M. Sanderson, Proteins 32 (1998) 350.
[18] L. Birch, C.W. Murray, M.J. Hartshorn, I.J. Tickle, M.L. Verdonk, J. Comput. Aid. Mol. Des. 16 (2002) 855.
[19] C. Wurth, U. Kessler, J. Vogt, G. Schulz, G. Folkers, L. Scapozza, Protein Sci. 10 (2001) 63.
[20] J. Vogt, R. Perozzo, A. Pautsch, A. Prota, P. Schelling, B. Pilger, G. Folkers, L. Scapozza, G. Schulz, Proteins 42 (2000) 545.
[21] D. Lorber, B. Shoichet, Protein Sci. 7 (1998) 938.
[22] C. Baxter, C. Murray, D. Clark, D. Westhead, M. Eldridge, Proteins 33 (1998) 367.
[23] G. Jones, P. Wilett, R. Glen, A. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 721.
[24] R. Knegtel, M. Wagener, Proteins 37 (1999) 334.