Designing Binding Pockets on Protein Surfaces using the A* Algorithm

Susanne Eyrisch, Volkhard Helms

Center for Bioinformatics
Saarland University
P.O. Box 151150
D-66041 Saarbruecken, Germany

eyrisch@bioinformatik.uni-saarland.de

volkhard.helms@ bioinformatik.uni-saarland.de

Abstract: The in-silico design of ligands binding to the protein surface instead of deep binding pockets is still a great challenge. Often no appropriate binding pockets are available in the apo experimental structures and standard virtual screening techniques will fail. Here, we present two new algorithms for designing tailored ligand binding pockets on the protein surface that account for protein backbone and side chain flexibility. At first, the protein surface is scanned for potential pocket positions using a program named PocketScanner. This program minimizes the protein energetically in the presence of generic pocket spheres representing the new binding pockets whose positions remain fixed. The side chains of the relaxed protein conformations are then further refined by a second program named PocketBuilder. PocketBuilder identifies all residues within a given radius of the pocket positions and searches for the best combination of side chain rotamers using the A* algorithm. Given multiple protein conformations as input, PocketBuilder identifies those that lead to the best results, namely protein conformations of low energy that possess binding pockets with desired properties. The approach was tested on the proteins BCL-X_L, IL-2, and MDM2 which are involved in protein-protein interactions and hence represent challenging drug targets. Although the native ligand binding pocket was not or only partly open in the apo crystal or NMR structures, PocketScanner and PocketBuilder successfully generated conformations with pockets into which a known inhibitor could be docked in a native-like orientation for two out of the three test systems. For BCL-X_L, the docking scores were even similar to those obtained in re-docking experiments to the inhibitor bound crystal structure.

1 Introduction

After realizing that most diseases arise from aberrant molecular interactions, it has become an important goal to identify these interactions and to modulate them, for example through the binding of additional ligands, so that the native biological processes are reestablished or unwanted interactions are inhibited. In particular, the development of computational

tools supporting the design process of such modulators has become a very interesting research area. It involves, for example, the in-silico design of ligands that should bind to concave regions on the surface of the target protein. Given the three-dimensional target structure, one may predict the binding modes of a potential ligand by scanning the protein surface for favorable binding pockets. To this end, several computational tools have been developed that use geometric (e.g. PASS [BS00], SURFNET [Las95], PocketFinder [ATA05], LigSite [HRB97]) or energetic (e.g. GRID [Goo85], MCSS [CMK93], QSiteFinder [LJ05], CS-Map [LLY$^+$07]) criteria for detecting such clefts or hot spots. An example for pocket detection tools using only geometric criteria is the PASS (Putative Active Sites with Spheres) algorithm that identifies empty volumes on the protein surface by their burial extent. An example for the tools using energetic criteria for identifying putative hot spots is MCSS that generates positions and orientations of functional groups in the field of a flexible protein. In the case of enzymes, such binding pockets often correspond to active sites that deeply extend into the protein interior and are relatively easy to identify. It is much harder to detect pockets located at flat protein surfaces that often require structural rearrangements to open and therefore may not be fully accessible in the protein conformation used. The lack of clearly shaped binding pockets at protein-protein interfaces is one of the reasons why the structure-based drug design of small molecule protein-protein interaction inhibitors (SMPPIIs) remains a great challenge [AW04]. Until today, most published SMPPIIs for this class of drug targets were identified by experimental screening methods [WM07].

We have previously presented a pocket detection protocol that provides a starting point for in silico drug design for cases in which no potential binding pocket could be identified so that standard screening methods would fail [EH07]. For the three protein systems MDM2, BCL-X_L, and interleukin-2 (IL-2), we found that large pockets not detectable in the crystal structures of the free proteins opened frequently on the protein surfaces during standard molecular dynamics (MD) simulations of 10 nanoseconds length at room temperature. The identified transient pockets represent potential binding sites for new inhibitors. At the native binding site, pockets of similar size as with a known inhibitor bound could indeed be observed for all three systems. Docking known inhibitors with AutoDock 3 [MGH$^+$98] into these transient pockets resulted in docking results with less than 2 Å root mean square deviation (RMSD) from the crystal structures. In a subsequent study, we could show that when the water solvent was replaced by methanol the transient pockets opening in the MD simulations tended to be larger and less polar (unpublished results). Moreover, the docking results improved significantly for two of the three systems. However, a limiting factor of this pocket detection protocol is the high computational demand of MD simulations on biomolecular systems and it would be desirable to achieve the opening of surface pockets by a more efficient protocol. Fortunately, in many drug design applications, the approximate location of the binding site is already known. Hence, it is sufficient to sample only the corresponding part of the protein surface. This local instead of global search allows for a more accurate and directed sampling of low-energy protein conformations with accessible pockets. These protein conformations can then be used to optimize the interaction between the protein and the ligand or for virtual screening. We will show below that the problem of finding appropriate protein conformations can be solved efficiently using an informed graph search algorithm like the A* algorithm [HNR68] that uses knowledge about

the structure of the search space incorporated in heuristic functions to guide the search towards optimal solutions. During this search, a graph is built up in which each node represents a partial solution. Given an initial node representing the initial state, the algorithm searches the path to a given goal node, representing the goal state. The generated nodes are maintained in a priority queue. The priority of a partial solution *x* is given by

$$f(x) = g(x) + h(x) \tag{1}$$

where *g(x)* is the cost of this partial solution so far, i.e. from the start node to *x* and *h(x)* is the heuristic estimate of the minimal cost to reach the goal node from *x*. If the heuristic function is admissible (i.e. it never overestimates the cost of reaching the goal node) and consistent (i.e. it fulfills the triangle inequality), it will always find a path with minimal cost from a given start node to a given goal node if such a goal node exists. Leach applied the A* search to the flexible docking and the side chain placement problems [Lea94]. After placing an anchor region of the ligand into the binding site, he generated all possible ligand conformations. For each conformation that made no unfavorable interactions with the protein backbone and all rotameric states of a residue, the optimal combination of side chain rotamers was determined by an A* search. The initial node represented the structure without assigned rotamers for the residues at the binding site, while the goal nodes represented the optimal docking solutions, i.e. all residues had assigned rotamers.
In this work, we incorporated ideas from PASS, MCSS, and Leach's application of the A*-search into two new algorithms for the efficient generation of energetically favorable protein conformations with accessible binding pockets at defined locations on the surface of the BCL-X_L, IL-2, and MDM2 proteins.

2 Methods and Materials

Our method uses two programs for the construction of putative binding pockets: PocketScanner and PocketBuilder. PocketScanner scans a user-defined region of the protein surface for potential pocket positions and generates protein conformations in which the backbone has adapted to these pocket positions. PocketBuilder uses these intermediate conformations for calculating a final set of conformations that best fulfil the search criteria, namely the desired trade-off between a protein conformation with low-energy side chain rotamers and a pocket of defined volume. Both programs were implemented in C++ using the BALL library [KL00] and the CHARMM EEF1 [LK99] force field that was used to compute all energies given below. This force field treats the solvent as an implicit continuum, and including such effects is crucial for designing pockets on protein surfaces. Binding pockets are represented by generic pocket spheres that were added to the force field. In the current setup, they only interact with the protein atoms via van-der-Waals interactions (with a radius of 1, 2, or 3 Å and a well depth of 0.05 kcal/mol). The pocket volumes and polarities were calculated as described in [EH07].

2.1 Structure Preparation

The unbound (apo) and inhibitor-bound protein structures of three test systems were taken from the Protein Data Bank [BWF+00] (PBD entries 1R2D and 1YSI for BCL-X_L, 1M47 and 1PY2 for IL-2, and 1Z1M and 1T4E for MDM2). All hetero atoms (including the ligand) were manually removed. As residues 28-81 are missing in 1R2D, the two parts of the protein were modeled as two distinct chains. The missing residues in 1M47 were modelled as loops of the lowest AMBER/GBSA potential energy generated by the program RAPPER [dBDBB03]. The structure of apo MDM2 is represented by 24 NMR models that differ mainly in the loop regions. Since no model is defined as most representative, the first model was chosen. The apo structures were superimposed on the inhibitor-bound structures based on the C_α-coordinates using the VMD program [HDS96].

2.2 The PocketScanner Algorithm

PocketScanner creates a grid around a given center with suitable dimensions and edge length and scans the protein surface for potential positions of pockets with a given radius. The z-axis of this grid is the solvent vector defined by the initial pocket position and the center of gravity of the 10 nearest solvent exposed atoms. The generic pocket sphere representing the pocket center is then placed on each grid point and its burial count (number of protein atoms within 8 Å) is calculated. Only those positions with a burial count above a given threshold (default: 65) are accepted, otherwise the resulting cavity may be too flat. As this criterion allows for pocket positions that are deeply buried inside the protein, we additionally require that the minimal distance to any solvent exposed atom must be smaller than 2 Å. The protein is then energy minimized in the presence of the generic pocket sphere using 500 steps of L-BFGS or until the RMS gradient is smaller than 0.01 kcal $\text{mol}^{-1}\text{Å}^{-1}$. During this energy minimization, the position of the generic pocket sphere is fixed, so that the protein relaxes its conformation. If the burial count is still high enough after the energy minimization, this protein conformation in combination with this pocket position is written to an output file and can be used as a starting conformation for PocketBuilder.

2.3 The PocketBuilder Algorithm

For calculating multiple protein conformations with putative binding pockets, PocketBuilder needs the following input data and parameters: starting conformations with putative pocket positions (either generated by PocketScanner or manually selected pocket positions), the radius of these pockets, a search radius for defining the flexible residues (default: 8 Å), a rotamer library, weights for scoring the internal protein energy, w_{energy}, and the van-der-Waals interaction energy with the pockets of the generated conformations, w_{pocket}, and the number of conformations to be generated. The algorithm consists of the initialization stage and the A*-search. The initialization is performed separately for each

starting conformation. It starts with determining all *N* residues (except for Ala and Gly) within a given distance from the generic pocket sphere and defines them as flexible. For the rigid part of the protein including all other residues and the backbone and C_β atoms of the flexible residues, the energy E_{rigid} and the van-der-Waals interaction energy with the pocket (i.e. the generic pocket spheres) $E_{rigid,pocket}$ are calculated. For each of the flexible residues *i* all rotamers *j* defined by the Dunbrack backbone independent rotamer library from 2002 [DC97] (including the original side chain conformation) are tested and their van-der-Waals interaction energy with the pocket $E_{i_j,pocket}$ and the energy change ΔE_{i_j} resulting from including this side chain rotamer in the calculation of E_{rigid} are determined. After calculating $E_{i_j}^{weighted}$ for each rotamer as

$$E_{i_j}^{weighted} = w_{energy} \cdot \Delta E_{i_j} + w_{pocket} \cdot E_{i_j,pocket} \tag{2}$$

the number of allowed rotamers for this residue is reduced by deleting all rotamers *j* with $E_{i_j}^{weighted} \geq 100$ kcal/mol. The pairwise non-bonded interaction energies E_{i_j,k_l} between the remaining rotamers *j* and *l* of each pair of residues *i* and *k* are calculated and stored in a hash table.

After the initialization stage, the algorithm builds up a tree with one subtree per starting conformation. The nodes in this tree represent partial solutions of the search problem, or more precisely conformations in which rotameric states have only been assigned to a part of the flexible residues. The order in which the flexible residues get defined side chain conformations is fixed, so all nodes of the same level in a certain subtree have the same residues already assigned. (The order in which side chains are added has no effect on the final result.) Note that the levels of the leave nodes are identical within a subtree, but may differ within different subtrees depending on the number of flexible residues defined for this starting conformation. The buildup of the tree is controlled by the A* algorithm. The algorithm assigns each node *x* a priority *f(x)* (see equation 1) that evaluates the true costs *g(x)* of this partial conformation so far and the estimated minimal cost *h(x)* for reaching a leaf node, where

$$\begin{aligned} g(x) \quad = \quad & w_{pocket} \cdot E_{rigid,pocket} + w_{energy} \cdot E_{rigid} + \\ & \sum_{i=1}^{x} \left(w_{pocket} \cdot E_{i_r,pocket} + w_{energy} \cdot \left(\Delta E_{i_r} + \sum_{k=1}^{i-1} E_{i_r,k_r} \right) \right) \end{aligned} \tag{3}$$

$$\begin{aligned} h(x) \quad = \quad & \sum_{k=x+1}^{N} \min_{l} \left(w_{energy} \cdot \Delta E_{k_l} + w_{pocket} \cdot E_{k_l,pocket} \right) + \\ & \sum_{k=x+1}^{N} \left(\left(\sum_{i=1}^{x} \min_{l} E_{i_r,k_l} \right) + \left(\sum_{n=x+2}^{N} \min_{l,m} E_{k_l,n_m} \right) \right) \end{aligned} \tag{4}$$

In the summations, *i* runs over all flexible residues with already assigned rotamers *r* (i.e. E_{i_r} indicates that side chain *i* has been locked into rotamer *r*), *k* and *n* run over the remaining ones, and *l* and *m* run over different rotamers of a side chain. In each step, the node

representing the partial conformation that seems most promising (i.e. with lowest *f(x)*) is selected. If this node is not a leaf node, a new node is added for each possible rotamer of the succeeding residue and the priorities of these new partial solutions are determined. Otherwise the corresponding conformation is written to an output file. The algorithm terminates as soon as the total number of output conformations is reached.

2.4 Docking into Designed Pockets

Docking experiments were performed with AutoDock 3.0.5 [MGH+98] as described before [EH07]. The ligands were extracted from the complex crystal structure and rotatable bonds were assigned with AutoTors. The grid maps were calculated with AutoGrid. The grid centers were chosen to coincide with the pocket positions. The default grid spacing of 0.375 Å between the grid points and the default grid dimension of 60 x 60 x 60 points was used and the standard Lamarckian Genetic Algorithm protocol with the default values. 10 independent docking runs were carried out for each PocketBuilder conformation.

3 Results and Discussion

PocketScanner and PocketBuilder were tested using the proteins BCL-X_L, MDM2, and IL-2. IL-2 is an important component of the immune response and BCL-X_L and MDM2 belong to the apoptosis pathway. The binding pockets targeted by the small molecule ligands are not or not fully open in the apo protein structures and thus cannot be used for structure-based drug design. Docking the known inhibitors into these apo structures gave poor results with lowest RMSDs of 2.9 - 3.4 Å as shown in Table 1.

	Re-Docking		Apo-Docking	
System	RMSD [Å]	Score [kcal/mol]	RMSD [Å]	Score [kcal/mol]
BCL-X_L - N3B	0.9	-10.5	3.3	-6.2
IL-2 - FRH	1.1	-10.8	2.9	-6.2
MDM2 - DIZ	1.1	-13.1	3.4	-6.7

Table 1: Best docking results for docking the inhibitor into its bound and the apo structure using AutoDock3

The crystal or NMR structures of the apo proteins were scanned for positions of inducible pockets using PocketScanner. The grid center was placed at the ligand center of mass, the dimension was 11, and the edge length 2 Å. Running PocketScanner took about 1 hour on a single CPU of an Intel Core 2 Duo processor which mainly resulted from the large number of energy minimizations. Out of the 11^3 possible positions, 67 (66) were accepted for BCL-X_L, 25 (18) for IL-2, and 29 (20) for MDM2 when using a pocket radius of 2 Å (3 Å respectively). Note that the pocket positions do not have to be located at the inhibitor binding site. The grid and the accepted positions of BCL-X_L are shown in Fig. 1.

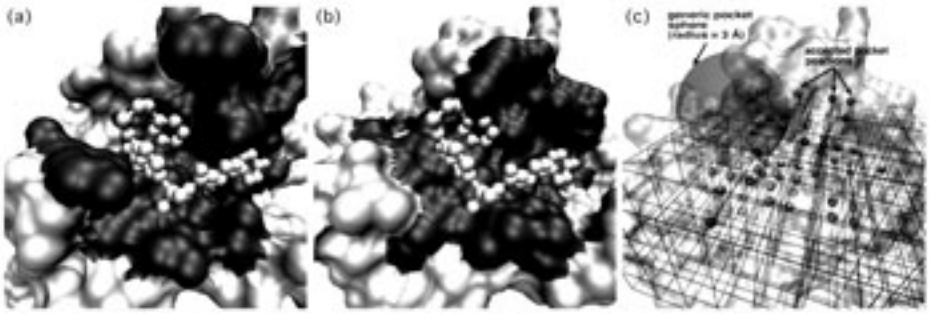

Figure 1: (a) BCL-X_L bound to its small molecule inhibitor N3B (shown as white balls and sticks), (b) BCL-X_L in its apo conformation with the inhibitor (shown to illustrate clashes with protein residues), and (c) with the grid generated by PocketScanner (accepted pocket positions are shown as black spheres) and an example for a generic pocket sphere.

Running PASS revealed that more pockets were detected in the PocketScanner conformations when the larger pocket radius was used. An overview of the properties of the pockets generated using PocketScanner is shown in Table 2. One would expect that the mean pocket volume would increase when using a larger pocket radius, but this is not the case for MDM2. A larger pocket radius may also cause a flat cavity and thus a pocket of reduced volume.

System	Pocket Radius [Å]	Detected Pockets [%]	Ø Pocket Volume [Å^3]	Ø Pocket Polarity
BCL-X_L	2	43	381.3 ± 82.3	0.33 ± 0.03
	3	86	394.9 ± 109.7	0.29 ± 0.04
IL-2	2	52	311.8 ± 59.1	0.31 ± 0.04
	3	78	328.8 ± 58.9	0.29 ± 0.03
MDM2	2	31	376.8 ± 91.7	0.33 ± 0.02
	3	75	315.7 ± 98.6	0.31 ± 0.03

Table 2: Properties of the pockets detected in the conformations generated using PocketScanner

The conformations and the corresponding pocket positions generated by PocketScanner with a pocket radius of 2 and 3 Å were then used as starting conformations for PocketBuilder. As the weighting of the internal protein energy and the protein-pocket interaction energy crucially influences the A* search, we calculated 500 final conformations using three different weightings for the two pocket radii: (1) internal protein energy and protein-pocket interaction energy weighted equally (0.5 and 0.5), (2) a strong emphasis on the pocket (0.1 and 0.9), and (3) a dominance of the pocket (0.01 and 0.99). The bottleneck for the run time of PocketBuilder is the initialization stage. This stage takes 6-10 minutes per starting conformation depending on the number of flexible residues (here, 8-18 flexible residues) and the number of accepted rotamers. For this purpose we added a greedy preselection of the starting conformations: For each conformation, the weighted sum of the

internal protein energy and protein-pocket interaction energy is calculated and the 20 starting conformations with lowest score are retained. We are aware that this preselection may delete conformations that would later on score better with altered side chain rotamers, but running the algorithm with too many starting conformations is nearly infeasible. The A* search took between 40 minutes and 4 hours depending on the number of possible nodes in the search tree and on how similar the scores of these nodes are. The ratio between the number of possible nodes and the number of generated nodes gives a measure for the efficiency of the algorithm. As listed in Table 3, the number of possible conformations increases with augmenting w_{pocket}. At the same time, the algorithm generally finds the 500 leaf nodes with lowest score more efficiently, suggesting that the interaction energy between the protein and the pocket is more diverse in the generated nodes than the internal protein energy. This is not surprising as the absolute value of the internal protein energy is about 4 orders of magnitude larger than the interaction energy with the pocket. No trend is apparent for the influence of the weighting and the pocket radius on the mean pocket volume and polarity. These mean volumes even seem to suggest that PocketBuilder reduces the volume of most pockets to snugly fit around the generic pocket spheres.

System	Pocket Radius [Å]	w_{pocket}	# Conformations	Efficiency	Ø Pocket Volume [Å^3]	Ø Pocket Polarity
BCL-X_L	2	0.5	$1.0 \cdot 10^{12}$	$8.3 \cdot 10^{6}$	715.3 ± 21.9	0.36
	2	0.9	$1.9 \cdot 10^{12}$	$1.7 \cdot 10^{7}$	343.6 ± 31.7	0.27 ± 0.01
	2	0.99	$3.4 \cdot 10^{12}$	$1.6 \cdot 10^{9}$	337.4 ± 37.2	0.27 ± 0.01
	3	0.5	$1.7 \cdot 10^{11}$	$2.4 \cdot 10^{6}$	282.6 ± 34.2	0.30 ± 0.01
	3	0.9	$5.6 \cdot 10^{11}$	$7.1 \cdot 10^{6}$	276.1 ± 55.0	0.31 ± 0.01
	3	0.99	$4.5 \cdot 10^{14}$	$1.2 \cdot 10^{9}$	485.2 ± 92.7	0.37 ± 0.01
IL-2	2	0.5	$2.0 \cdot 10^{15}$	$1.0 \cdot 10^{11}$	291.7 ± 3.8	0.27
	2	0.9	$2.7 \cdot 10^{16}$	$9.3 \cdot 10^{11}$	290.3 ± 4.8	0.27
	2	0.99	$1.9 \cdot 10^{18}$	$4.1 \cdot 10^{13}$	359.6 ± 36.3	0.33 ± 0.01
	3	0.5	$1.2 \cdot 10^{14}$	$5.9 \cdot 10^{9}$	450.9 ± 80.2	0.31 ± 0.01
	3	0.9	$2.2 \cdot 10^{15}$	$6.9 \cdot 10^{10}$	507.4 ± 90.7	0.30 ± 0.01
	3	0.99	$4.6 \cdot 10^{16}$	$1.2 \cdot 10^{12}$	344.4 ± 23.3	0.33 ± 0.01
MDM2	2	0.5	$1.5 \cdot 10^{14}$	$1.4 \cdot 10^{10}$	314.0 ± 56.8	0.31 ± 0.02
	2	0.9	$1.4 \cdot 10^{15}$	$1.4 \cdot 10^{10}$	420.0 ± 50.2	0.33 ± 0.02
	2	0.99	$2.1 \cdot 10^{16}$	$2.4 \cdot 10^{7}$	277.9 ± 19.6	0.32 ± 0.01
	3	0.5	$2.6 \cdot 10^{12}$	$7.0 \cdot 10^{8}$	233.8 ± 26.3	0.32 ± 0.01
	3	0.9	$8.8 \cdot 10^{13}$	$7.6 \cdot 10^{9}$	235.3 ± 27.1	0.32 ± 0.01
	3	0.99	$2.0 \cdot 10^{15}$	$1.4 \cdot 10^{10}$	339.1 ± 89.1	0.31 ± 0.02

Table 3: Influence of the pocket radius and the weighting on the performance of PocketBuilder and the properties of the induced pockets

The main goal of this study is to design pockets on the protein surface that are suitable for ligand binding. Therefore, the known inhibitors were now docked into the generated conformations. The main questions are: (1) Can docking into the designed pockets reproduce

the native ligand binding mode? (2) Which weighting and pocket radius requires the lowest number of generated conformations? Table 4 lists the best scored docking results with RMSD $\leq$ 2 Å (or the docking result with lowest RMSD) for each weighting and pocket radius.

System	Pocket Radius [Å]	w_{pocket}	RMSD [Å]	Score [kcal/mol]	Relative Score Rank [%]	PocketBuilder Conformation
BCL-X_L	2	0.5	1.9	-10.0	42.0	213
	2	0.9	2.0	-10.1	34.4	169
-	2	0.99	2.0	-10.2	33.6	241
	3	0.5	1.7	-10.2	55.4	82
N3B	3	0.9	1.5	-10.4	58.2	376
	3	0.99	2.0	-11.3	6.2	29
IL-2	2	0.5	1.8	-6.5	6.4	75
	2	0.9	1.8	-7.3	1.7	226
-	2	0.99	2.0	-4.3	54.4	428
	3	0.5	2.0	-5.6	44.1	167
FRH	3	0.9	2.0	-6.6	29.6	285
	3	0.99	2.0	-4.4	60.6	430
MDM2	2	0.5	2.6	-7.9	83.1	193
	2	0.9	2.6	-7.8	90.5	225
-	2	0.99	2.9	-9.1	4.9	113
	3	0.5	3.2	-9.7	5.9	436
DIZ	3	0.9	3.1	-8.8	27.4	345
	3	0.99	2.2	-9.1	88.3	41

Table 4: Influence of the pocket radius and the weighting on the docking results (shown are the best scored docking results with RMSD $\leq$ 2 Å or the docking result with lowest RMSD)

Interestingly, for all setups the native ligand binding mode was found for BCL-X_L and IL-2. This indicates that PocketBuilder was successful in inducing the opening of native-like binding pockets on the surface of the BCL-X_L and the IL-2 proteins. An example of how PocketScanner and PocketBuilder change the apo conformation is shown in Figure 2. For BCL-X_L, the docking scores were even quite similar to that obtained in the re-docking experiment. The unsatisfying docking scores for IL-2 may be due to the fact that this binding site consists of two subpockets, that lie about 15 Å apart and with this setup, only one of these subpockets can be induced. Here, using more than one generic pocket sphere would most probably improve the docking score. However, the large relative rank of most docking results shown in Table 4 indicates that our setup does not only lead to the generation of pockets similar to those seen in the bound structure, but also to alternative pocket conformations that possess the desired properties as well. Moreover, most setups seem to prefer such alternative pockets because the best suited protein conformation is often generated quite late during the A* search. For MDM2, our method was not able to completely reproduce the native binding mode of the ligand. But when comparing the docking results listed in Table 4 to the results when docking into the apo structure (listed

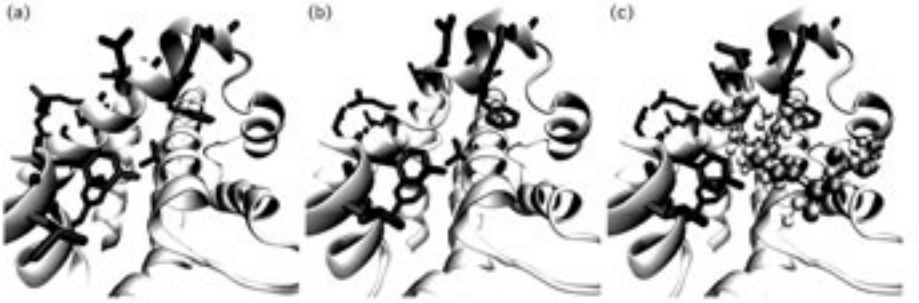

Figure 2: Change of the backbone (white cartoon representation) and the flexible residues (black licorized representation) in a subpocket of BCL-X_L (a) before running PocketScanner, (b) after running PocketScanner (pocket radius = 3 Å), and (c) after running PocketBuilder (w_{pocket} = 0.99). In (c), the ligand N3B is shown in its native conformation (thin white balls and sticks) and in its docked conformation (grey thick balls and sticks).

in Table 1), it becomes apparent that an opening of the native binding was at least partly induced. In such cases using a larger generic pocket sphere may be helpful in making the native binding pocket fully accessible.

4 Conclusion and Outlook

Accounting for protein flexibility is one of the current challenges in the protein-ligand docking field. Here, we have presented a rigorous algorithm to scan the rotameric space of residues on protein surfaces for openings of suitable ligand binding pockets. As shown for the model systems BCL-X_L, IL-2, and MDM2 a systematic scanning of the protein surface around the interface is computationally feasible. For two out of the three systems, the PocketBuilder algorithm was then able to induce pockets of suitable volumes and shapes so that the small molecule ligands may bind in a native-like orientation. By testing the algorithm on a larger number of protein-ligand complexes in the near future, we plan to tune the set of control parameters to further enhance the efficiency of this approach and the ranking of the native binding mode.

References

[ATA05] J. An, M. Totrov, and R. Abagyan. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol. Cell Proteomics*, 4:752–761, 2005.

[AW04] M.R. Arkin and J.A. Wells. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.*, 3:301–317, 2004.

[BS00] G.P. Brady and P.F. Stouten. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.*, 14:383–401, 2000.

[BWF+00] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28:235–242, 2000.

[CMK93] A. Caflisch, A. Miranker, and M. Karplus. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.*, 36:2142–2167, 1993.

[dBDBB03] P.I.W. de Bakker, M.A. DePristo, D.F. Burke, and T.L. Blundell. Ab initio construction of polypeptide fragments: Accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the Generalized Born solvation model. *Proteins*, 51(1):21–40, 2003.

[DC97] R.L. Dunbrack and F.E. Cohen. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.*, 6:1661–1681, 1997.

[EH07] S. Eyrisch and V. Helms. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.*, 50:3457–3464, 2007.

[Goo85] P.J. Goodford. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, 28:849–857, 1985.

[HDS96] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[HNR68] P.E. Hart, N.J. Nilsson, and B. Raphael. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Trans. on SSC*, 4(2):100–107, 1968.

[HRB97] M. Hendlich, F. Rippmann, and G. Barnickel. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.*, 15:359–363, 1997.

[KL00] O. Kohlbacher and H.P. Lenhof. BALL–rapid software prototyping in computational molecular biology. Biochemicals Algorithms Library. *Bioinformatics*, 16:815–824, 2000.

[Las95] R.A. Laskowski. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.*, 13:323–330, 1995.

[Lea94] A.R. Leach. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.*, 235:345–356, 1994.

[LJ05] A.T. Laurie and R.M. Jackson. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, 21:1908–1916, 2005.

[LK99] T. Lazaridis and M. Karplus. Effective energy function for proteins in solution. *Proteins*, 35:133–152, 1999.

[LLY+07] M.R. Landon, D.R. Lancia, J. Yu, S.C. Thiel, and S. Vajda. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J. Med. Chem.*, 50:1231–1240, 2007.

[MGH+98] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 14:1639–1662, 1998.

[WM07] J.A. Wells and C.L. McClendon. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, 450:1001–1009, 2007.