ProPose: Steered Virtual Screening by Simultaneous Protein−Ligand Docking and Ligand−Ligand Alignment

Markus H. J. Seifert*

4SC AG, Am Klopferspitz 19a, D-82152 Martinsried, Germany

Received December 2, 2004

The 'model-free' screening engine ProPose implements a general method for performing simultaneous protein−ligand docking, ligand−ligand alignment, pharmacophore queries—and combinations thereof—in order to incorporate a priori information into screening protocols. In this manuscript we describe a case study on herpes simplex virus thymidine kinase, an important antiviral drug target, where we evaluate different approaches for handling a specific type of a priori information, i.e., multiple target structures. We demonstrate that a simultaneous alignment on two target structures—in conjunction with logic operations on interactions and docking constraints derived from protein structure—is an effective means of (i) improving the enrichment of chemical substructures that are compatible with the a priori known ligands, (ii) ensuring the steric fit into the target protein, and (iii) handling target flexibility. The combination of ligand- and receptor-based methods steers the virtual screening by ranking molecules according to the similarity of their interaction pattern with known ligands, thereby—to some extent—outweighing the deficiencies of simple scoring functions often used in initial virtual screening.

INTRODUCTION

Prerequisites for successful virtual screening are not only algorithms to solve the computational challenges but also an efficient strategy to exploit the a priori information available for the respective target. The Protein Data Bank (PDB)[1] with its current size of nearly 30,000 deposited 3D structures of macromolecules often provides more than one structure for many targets of pharmaceutical importance. Additionally, a wealth of data is available from databases of drug molecules, for example the World Drug Index (WDI).[2] Usually, if a receptor structure is on-hand, a virtual screening based on protein−ligand docking is performed.[3,4] Where only ligand structures are known one must resort to molecular alignment or (3D) QSAR methods. A variety of approaches for molecular alignment has been proposed in the literature. The majority of these are based on a specific definition of molecular similarity including topological, quantum-chemical, or field-based measures.[5−7] A particularly successful approach for aligning a flexible molecule on a rigid template molecule focuses on the interaction possibilities of the molecules.[8] A comprehensive review of established methods has been published by Lemmen and Lengauer.[9]

The integration of all available a priori information into efficient screening protocols remains a challenge, especially when different virtual screening methodologies are under consideration for a specific target. One may wish (i) to use biological information to steer or guide the virtual screening in order to enrich molecules that are selective for one enzyme target or receptor function, (ii) to exploit chemical information to address synthetical or intellectual property aspects for example by selectively enriching specific substructures and excluding certain molecular patterns during the screening process, (iii) improve the result of screening runs on difficult targets, or (iv) simply to incorporate different 3D structures of a flexible target in a single screening run. Current approaches for 'guided' docking have been reviewed recently.[10] These approaches comprise indirect, i.e., based on knowledge-derived potentials, and direct methods, i.e., based on anchor fragments, pharmacophors, or key interaction sites. However, there exists a limitation in applicability: these methods are extensions of pure docking programs which usually implement their own interaction and scoring scheme allowing only a limited manual control by the user. In this manuscript we describe a general method for steering virtual screening based on an efficient combination of protein−ligand docking and ligand−ligand alignment. This method is implemented in the novel, 'model-free' docking and alignment program 'ProPose':[11] ProPose does not contain any hard-coded interaction geometries and energies and can therefore be easily configured for protein−ligand docking as well as for ligand−ligand alignment and different scoring schemes. We describe a new input preparation program—PrepA—which transforms ligand information into a target description file (TDF). PrepA generates a pseudoreceptor that enables the application of the docking engine to ligand−ligand alignment. The concept of pseudoreceptors has been proposed in the literature before.[12−19] We have however extended this principle by incorporating logic operations for interactions within this pseudoreceptor in order to model target flexibility.

Here we chose the Herpes Simplex virus 1 (HSV) thymidine kinase (TK) as a test case. Thymidine kinases catalyze the phosphorylation of deoxythymidine (dT) to deoxythymidine monophosphate (dTMP). HSV TK is a homodimer of subunits containing the classical mononucleotide (NMP) fold. It is well-known that NMP kinases undergo large conformational changes upon substrate bind-

* Corresponding author phone: +49-89-700763-0; fax: +49-89-700763-29; e-mail: markus.seifert@4sc.com.

10.1021/ci0496393 CCC: $30.25
Published on Web 02/23/2005

ing. However, several substrate-bound protein structures of TK exhibit only minor conformational differences in the vicinity of the substrate allowing for in silico protein−ligand docking.[20−22] In contrast to the human enzyme, HSV TK shows low substrate specificity, including deoxycytidine, pyrimidine and purine analogues.[23] This is the basis for antiviral therapeutic drugs, e.g. ganciclovir (GCV) and aciclovir (ACV). These nucleoside analogues are activated by HSV TK, and their triphosphate form blocks the viral replication by terminating the DNA elongation.[24] Additionally, suicide gene transfer using TK in combination with GCV treatment represents a widely used system for gene therapy of cancer.[25] A recent study presented inhibitors for a similar target, *Mycobacterium tuberculosis* thymidine kinase.[26]

In this paper, we show that our method—based on the docking engine ProPose[11]—for incorporating a priori information into screening protocols is capable of steering the virtual screening on the HSV TK target. More specifically, we demonstrate that the enrichment of specific, yet diverse substructures can be steered by performing a simultaneous alignment on several targets or even by combining docking into a protein structure with an alignment on ligand structures in a single screening run. This allows to improve the ranking of molecules and, in general, to integrate constraints into virtual screening that are not necessarily fulfilled by simple protein−ligand docking, e.g. biological, chemical and intellectual property constraints. As basically all published scoring functions for virtual screening predict binding affinities with only low accuracy within a large set of complexes[27] our method additionally presents an easy-to-use backdoor for coping with this problem until more reliable scoring functions have been developed. Furthermore, the application of logic operations to molecular interactions allows the consideration of different conformational states of the virtual screening target.

This manuscript is organized as follows: we begin with a brief review of the basics of our program 'ProPose'—and recent improvements thereof—and describe the procedure for ligand−ligand alignment in more detail. We then present conventional ProPose docking runs followed by a ProPose alignment on two molecules, deoxythymidine and ganciclovir. At the heart of this study is the description of a simultaneous alignment on both targets and a simultaneous alignment/docking procedure. Finally the results are discussed with respect to the advantages and disadvantages of this approach.

MATERIALS AND METHODS

ProPose. *Basic Principles*. ProPose is a docking engine based on an incremental construction algorithm.[11] A torsion angle library derived from semiempirical quantum chemical calculations provides the minimum energy torsion angles for the incremental construction of molecules within the active site of a receptor structure. In contrast to established docking programs ProPose does not contain any hard-coded interaction geometries or energies. Instead, these data are provided by user-defined configuration files. ProPose utilizes a pharmacophore-like description of molecular interaction based on discrete interaction points and centers which are transformed on-the-fly into smooth potential energy surfaces by a Lorentzian averaging procedure. This method has been described in detail before,[11] but an additional example is illustrated in Figure 1. This example (Figure 1) shows the distance dependent interaction score of two charges. The corresponding definition of this interaction is shown in Table 1. Usually, the bidirectional superposition of interaction centers and points allows for modeling angular dependencies of interaction geometries.[11] Here, modeling the radial symmetry of the charge interaction is accomplished by placing both interaction point and center on identical positions. The width of the potential can be configured by the tolerance parameter. In contrast to the physical 1/r dependence, the ProPose potential does not have a pole at zero distance and converges faster to zero with increasing distance. These features are important in order to apply the interactions acting as pharmacophors which are not necessarily linked to a real atom: for example, the interactions of displaceable cations or water molecules can be provided as pharmacophors without a real atom. This will favor ligands that interact with such a pharmacophore but does not exclude ligands that occupy the pharmacophore position thereby "expelling" the potential interaction partner. The user is allowed to place an interaction as a pharmacophore at arbitrary positions without having to worry about artifacts due to extremely high scores.

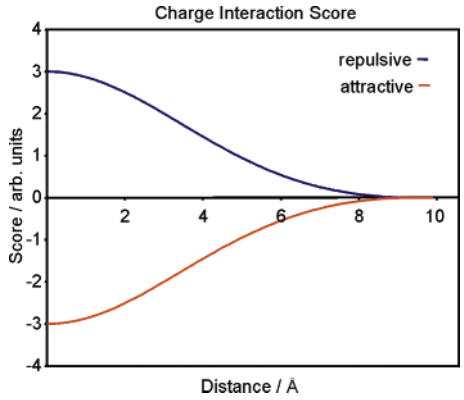

Figure 1. The ProPose interaction score of two charges of equal (blue) and opposite denominator (red) with respect to their distance is shown. The definition of the two interactions is explained in Table 1. Due to the Lorentzian averaging scheme, the score has no pole at zero distance and shows—in comparison to the physical 1/r function—a faster convergence to zero with increasing distance. Both features are essential for pharmacophore applications.

Table 1: Definition of Interactions

comment line	# ammonium cation	# hydroxyl anion
	@SUBSTRUCTURE	@SUBSTRUCTURE
SMARTS string	[#7&H4]	[#8&H1&X1]
	@PRIORITY	@PRIORITY
priority	1	1
	@ATOM	@ATOM
mol2-like 3D structure	1 N1 0 0 0 N.4 | N.4	1 O1 0 0 0 O.3 | O.3
	@BOND	@BOND
	@INTERACTION	@INTERACTION
assigned interaction	1 pos_charge yes	1 neg_charge yes
-coordinates	0 0 0	0 0 0
-for graphical output	0 0 0 180	0 0 0 180
-tolerance	5.0	5.0
logical expression	@LOGIC	@LOGIC
	1	1
interaction points	@IAPOINT	@IAPOINT
-index, interaction index	1 1	1 1
-coordinates	0 0 0	0 0 0
-weight	4.94	4.94
volumes of exclusion (optional)	@RECEPTORATOM	@RECEPTORATOM

The convergence to zero is important in providing enough spatial selectivity and to speed-up score calculations by a distance cutoff (default: 10 Å).

Recent Optimizations. Recent optimizations of ProPose include modifications of the fragment cleavage scheme, of the torsion angle library and of the interaction design: (i) Substituents without internal degrees of freedom, which are not connected to any further molecular fragments, e.g. methyl and nitril, are not considered as separate fragments anymore. This speeds up incremental construction due to a smaller total number of fragments. Accordingly these fragments have been removed from the torsion angle library. (ii) A thione fragment has been added to the torsion angle library. (iii) The parameters for the pose clustering have been optimized by a grid search to speed up computation while keeping the success rate on the CCDC−Astex data set constant (for a detailed description of the validation procedure see ref 11). (iv) The configuration files for Böhm's scoring approach[28] have been revised to reduce the number of interaction points. (v) The rigid-body optimization of ligand poses has been restricted to clashing poses during the first and second iteration of incremental construction. An evaluation showed that a rigid-body optimization at later iterations does not improve the poses significantly.

Pose Clustering. Pose clustering in ProPose is accomplished by (i) removing duplicate poses with an RMSD < 0.25 Å, (ii) merging all poses which share 60% of their interaction with the target into a cluster, and (iii) calculating the number of representatives (N_i) from each cluster i according to the following formula

$$N_i = N_{av} \exp\left(T\frac{|S^i_{best} - S_{worst}|}{\Delta S}\right)$$

with N_{av} being the projected average number per cluster, the best score S^i_{best} within cluster i, the absolute worst score S_{worst}, and ΔS the difference between the absolute minimum and maximum score. T represents a variable scaling factor (default: 3.0). N_{av} is defined by $N_{av} = N_T \cdot n^{-0.8}$ with N_T being the total number of poses needed for incremental construction and n the number of clusters. The parameters in this formula were optimized to extract a suitable number of representatives in order to continue the incremental construction with the configured number of poses. (iv) Finally, a MaxMin algorithm extracts N_i diverse representatives from the corresponding cluster.

ProPose Alignment. The workflow for ProPose docking or alignment is split into two parts (see Figure 2): first the corresponding input information—either receptor or ligand data—is fed into the corresponding preparation tool PrepD and PrepA, respectively. The preparation tools convert the input into a target description file (TDF) which ProPose uses as its template for docking. The TDF contains a list of interactions including their centers, points, and a list of receptor atoms (see Table 2). The weight of the interactions can be set by a corresponding factor for the interactions points. The receptor atoms are used for the clash test and to evaluate surface area dependent properties. For docking, these atoms correspond to the protein atoms. For alignment, a different methodology is applied.

The ProPose ligand−ligand alignment procedure basically docks the ligand into a pseudoreceptor that has been constructed in order to reflect the interaction possibilities of the template ligand. No other additional similarity measure is used. PrepA reads the template ligand and generates a pseudoreceptor consisting of interactions and bounding box atoms: first, the interaction points and centers—as defined in a specific configuration file—are transformed into the local coordinate system of the corresponding ligand substructure. Compared to Böhm's docking interactions,[11,28] the alignment interactions are 'inverted' and mark the potential locations of interaction partners in the real receptor.

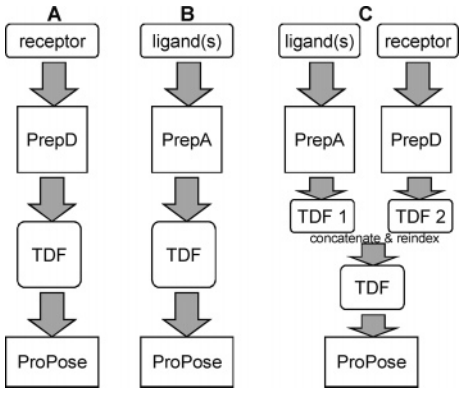

Figure 2. The workflow of ProPose docking and alignment comprises two steps: first, PrepD (docking, A) or PrepA (alignment, B) generate a target description file (TDF) that contains an abstract representation of the input information (see Table 2). TDFs can be merged and used for combined applications of docking and alignment (C). Second, ProPose reads the TDF and performs protein−ligand docking or ligand−ligand alignment.

Table 2: Format of Target Description File

group definition	@GROUP
-group index, indices of members	1 100 200 ... 2 300 400 ...
logic operation	@LOGIC 1 OR 2
definition of interaction	@INTERACTION
-index, type, flag	100 pos_charge yes
-transformed coordinates	49.3923 80.3562 53.4078
-for graphical output	0.0 0.0 0.0 180
-tolerance	5.0
definition of IA point	@IAPOINT
-index, coordinates, weight, flag	1 100 49.3923 80.3562 53.4078 4.94 no ...
bounding box atoms	@RECEPTORATOM
-index, name, coordinates, atom type, group index	1 N1 42.276 59.344 28.432 N.4 1 2 C2 43.120 60.595 28.493 C.3 1 3 C3 44.147 60.418 29.615 C.2 2 ...

An example is shown in Figure 3: four hydrogen bond acceptors are assigned to a hydrogen bond donor (Figure 3A) and positioned at locations where the real interaction partners in the receptor interaction are most likely located. Each interaction consists of an interaction center (depicted as spheres) and interaction points (depicted as crosses). An equivalent procedure is applied to hydrogen bond acceptors (Figure 3B) and all other interactions including both aromatic and hydrophobic. The interaction scores are taken from the corresponding docking interactions. Second, the bounding box atoms are created on dodecahedrons with a radius of 5 Å around the ligand atoms and are associated with a repulsive piece-wise linear potential in order to focus the alignment to the center of the pseudoreceptor (default repulsive interaction score: 3.0). This results in a penalty for ligands

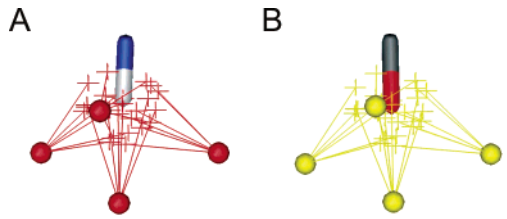

Figure 3. Examples of interaction geometries for ligand−ligand alignment are shown. A hydrogen bond donor of the template ligand is surrounded by four potential hydrogen bond acceptors (A). The interaction centers and points are depicted by spheres and crosses, respectively. A similar geometry is used for hydrogen bond acceptors of the template (B) as well as for hydrophobic and aromatic interactions (data not shown).

larger than the template ligand. Ligands much larger than the template are discarded due to clashes with the bounding box atoms.

Logic Operations and Target Flexibility. TDFs are plain text files that can be concatenated or otherwise combined (see Table 2). Using a text editor, the interactions—including centers and points—and the receptor atoms can be copied from one TDF into another. This process is not unambiguous and strongly depends on the source of data and the needs of the user. For standard applications, one would combine all interactions and receptor atoms from each source TDF into a single TDF, thereby specifying a number of discrete targets or conformational states. Before running ProPose, the interaction and atom indices have to be renumbered uniquely. To apply logic operations, the interactions and their corresponding receptor atoms are grouped according to their respective sources which may be e.g. different target structures or different conformational states of a single target. This allows to resolve possible incompatibilities between the atoms and interactions of these states. When using a logic AND of the interactions contained in separate groups the score is simply the sum of all interaction scores. Other logic operations such as OR, which considers only the best-scoring possibility of different sets of interactions, provide additional opportunities to improve virtual screening.

The concept of logic operations is ideally suited to extend a greedy, fragment-based algorithm like the incremental construction algorithm: Suppose a ligand consists of 10 fragments and each fragments has 10 incompatible interaction possibilities then 10^{10} possibilities have to be considered. In contrast, a greedy algorithm which decides to use only the minimum energy possibility for each fragment would consider just 10^2 possibilities, resulting in a speeding-up of several orders of magnitude. One, however, has to be sure that the simplification to local decisions is appropriate. Otherwise a sufficient number of ligand conformations has to be traced in parallel during incremental construction. For the molecular docking and alignment problem, important issues can be addressed by such a local, i.e., fragment-based, algorithm in conjunction with logic operations. For example, the flexibility of functional groups of amino acids usually affects only the fragments directly in contact with such a group. Therefore, one has to calculate the scores of the fragment interacting with the different receptor conformations and the greedy algorithm will continue with the minimum energy possibility.

The assignment of logic operations to interactions proceeds as follows (see Table 2): (i) Groups of interactions are defined (@GROUP) with the members of each group being linked with logical AND. (ii) The different groups are linked by a logical OR (@LOGIC). (iii) All interactions that are not assigned to a group explicitly are considered as common interactions (so-called group 0 interactions) that can be formed at the same time with the grouped interaction. The total score is calculated as follows:

$$Score = Score(\text{group } 0) + \min[Score(\text{group } 1), ..., Score(\text{group n})]$$

with

$$Score(\text{group i}) = \Sigma\, Score(\text{interactions} \in \text{i})$$

Accordingly, the interaction triangle superpositioning during base fragment placement considers only triangles whose vertices that either belong to the same group or are derived from a combination of interactions of one group with group 0 interactions.

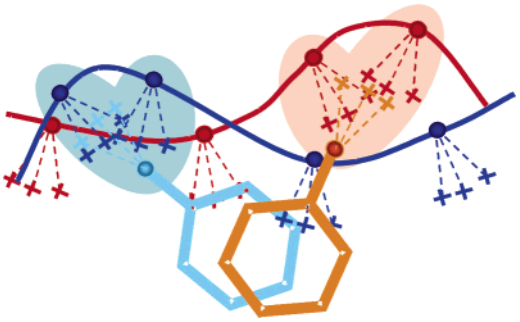

Figure 4. ProPose is able to use two types of information about conformational changes. Either the interactions originating from different ligand binding modes (orange and cyan) or interactions of different protein conformations (red and blue) are merged into interaction groups, which are subsequently linked by a logical OR. ProPose then considers only the best-scoring conformation (curved line) with respect to the docked ligand (hexagon). ProPose models interactions between chemical moieties by a bidirectional overlap of interaction points and centers which is depicted here for both conformations, i.e., groups, by blue and orange ellipsoids. Interaction points and centers are depicted as crosses and spheres, respectively.

This method allows conformational flexibility modeling for alignment as well as for docking (see Figure 4). The information about different conformations can be provided in two ways: Either the interactions of different ligand conformations are assigned to different groups (orange and cyan interactions) or interactions originating from different protein conformations are used (red and blue interactions). Both types of information—and combinations thereof—can be used to model flexibility. The influence of changes in atomic positions due to conformational changes is represented by receptor atoms (@RECEPTORATOM, see Table 1) which are linked to the interaction groups as well. The amino acid cooperativity often present in the active site of proteins is considered implicitly by providing only a fixed number of consistent conformational states of the target. To demonstrate the functionality, this method was applied to model the flexibility of the active site glutamine in the structure of HSV thymidine kinase which is described in the following section.

Virtual Screening Setup. *Screening Library*. 50,896 molecules were selected from an in-house database of

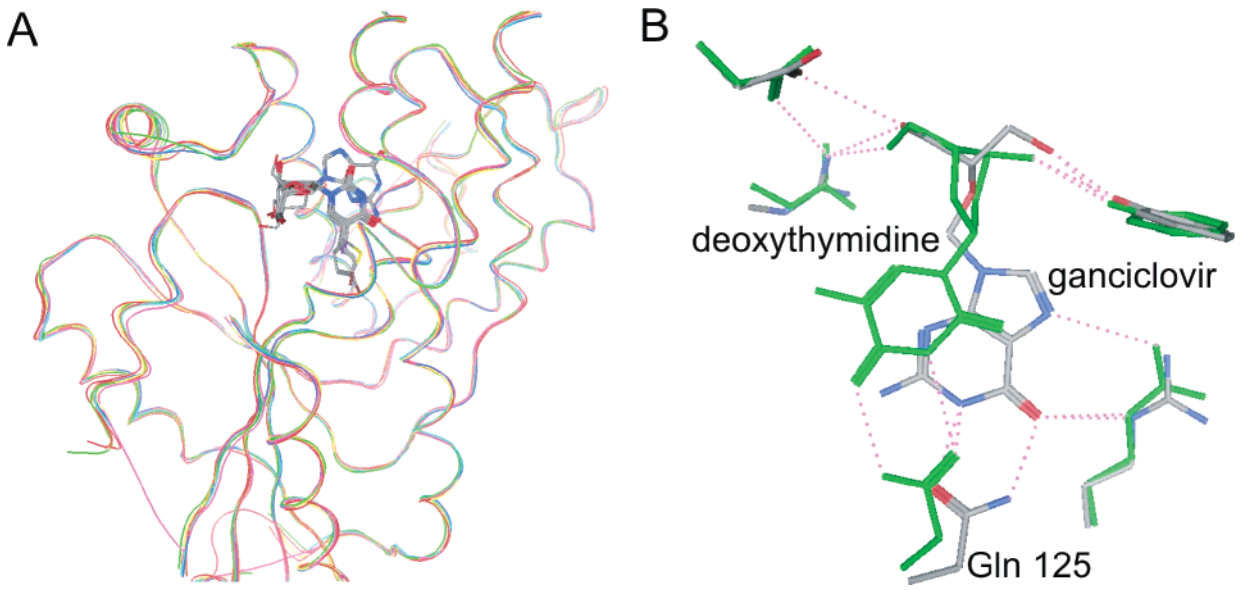

Figure 5. Seven protein structures of HSV thymidine kinase were aligned (A). This revealed a different binding mode for deoxythymidine (green) and ganciclovir (element colors) (B). The change in binding mode is accompanied by a flip of glutamine 125. Additionally, ganciclovir expels a water molecule (not shown) bound to arginine 176 from the active site. Hydrogen bonds are depicted as dotted lines.

5,060,804 commercially available, small organic compounds according to the following criteria: 1−10 H-bond acceptor atoms, 1−6 H-bond donor atoms, 1−14 rotatable single bonds, no phosphorus atoms, no triple bonds, no reactive groups, −4.0−3.0 SlogP (Log Octanol/Water Partition Coefficient),[29] 80−180 Å^2 TPSA (topological polar surface area),[30] and 100−400 D molecular weight. This selection includes molecules with properties similar to the ligands present in the PDB structures: 4−5 H-bond acceptor atoms, 2−3 H-bond donor atoms, 4−8 rotatable single bonds, no phosphorus atoms, no triple bonds, no reactive groups, −3.7−0.87 SlogP, 99−136 Å^2 TPSA, and 242−389 D molecular weight. All descriptors were calculated by MOE.[31] The selected molecules were protonated according to a neutral pH. The resulting screening library contained 4861 and 209 molecules with uracil and guanine substructures, respectively.

Protein Structures. PDB structures 1KI2, 1KI3, 1KI4, 1KI6, 1KI7, 1KI8, and 1KIM[32] including their cocrystallized ligands were aligned by MOE[31] (see Figure 5A). All water molecules were deleted from the structures. The alignment revealed a different binding mode for deoxythymidine (and analogues) and ganciclovir (see Figure 5B): glutamine 125 exists in two conformations resulting in a different interaction pattern for both ligands. Additionally, deoxythymidine is bound to arginine 176 by two water molecules, whereas ganciclovir binds directly to arginine 176. The binding modes of the other deoxythymidine analogues (PDB 1KI3, 1KI4, 1KI6, 1KI7, and 1KI8) are very similar to the binding mode of deoxythymidine. Therefore, deoxythymidine (PDB 1KIM) and ganciclovir (PDB 1KI2) have been selected as examples for different binding modes.

Docking in the Thymidine Kinase. All molecules of the screening library were docked into the active sites of PDB structures 1KIM and 1KI2. ProPose was configured for the scoring scheme according to Böhm,[28] 345 placements per base fragment, and 790 poses after pose clustering. Only the top scoring pose was stored for each molecule. In the case of 1KIM, three different active site definitions were tested: (i) removing all water molecules, (ii) keeping the two water molecules bound to Arg176, and (iii) removing the atoms of these two water molecules but keeping their interactions. This is equivalent to a water pharmacophore that causes no steric interactions.

Docking with Flexible Receptor. Based on the 1KIM protein structure a model of the active site was generated that incorporates the flexibility of Gln125: the two relevant conformations of Gln125 were extracted from PDB 1KIM and 1KI2 and assigned to two different interaction groups that comprise the atoms and the interactions of the side chain amide moiety of Gln125. These interaction groups were linked with a logical OR. All other receptor interaction were taken from the 1KIM structure and defined as common interactions that can be formed simultaneously with the interactions of the best scoring interaction group. This model captures the essential flexibility of the 1KIM and 1KI2 structures. All other configuration parameters were identical to the docking setup for rigid receptors.

Alignment on Single Template Structures. Deoxythymidine and ganciclovir were used as templates for molecular alignment. The ligand geometries resulting from the protein alignment (see Figure 5) were used as input for PrepA. A TDF was generated by PrepA in conjunction with a customized set of interaction geometries including hydrogen bond donors, hydrogen bond acceptors, aromatic ring interactions and hydrophobic interactions (see section "ProPose alignment"). All molecules in the screening library were aligned, and the 2000 best-scoring molecules were recorded.

Alignment on Dual Template with Logical AND. The TDFs for deoxythymidine and ganciclovir were concatenated, and the interaction index was adjusted to allow for simultaneous alignment on both molecules. The combined TDF contained all interactions of both target molecules. The bounding box atoms were constructed on dodecahedrons with a minimum distance of 5 Å around each atom of both molecules. All interactions were linked with a logical AND, i.e., the score

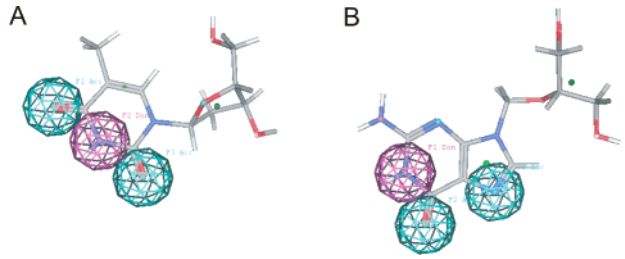

Figure 6. The pharmacophors of deoxythymidine (A) and ganciclovir (B) were defined according to their main interactions with glutamine 125 and arginine 176. Hydrogen-bond acceptors and donors are colored cyan and violet, respectively. The radii of the pharmacophors were 1.2 Å. These pharmacophors were used to calculate the enrichment rate of the virtual screens.

was simply the sum over all interaction possibilities including a combination of interactions from both sets. Again, all molecules in the screening library were aligned, and the 2000 best-scoring molecules were recorded.

Alignment on Dual Template with Logical OR. Two interactions groups were defined for the concatenated TDF (see previous paragraph) of deoxythymidine and ganciclovir comprising the interactions of the respective molecule. Both groups were linked with logical OR, i.e., a molecule can either interact with one set or the other but not with both simultaneously. No common interaction was defined. All molecules in the screening library were aligned, and the 2000 best-scoring molecules were recorded.

Alignment on Dual Template with Logical OR in Combination with Docking Constraints. The bounding box atoms were deleted from the TDF described in the previous paragraph and replaced by the 1KIM receptor protein atoms. This represents a simultaneous alignment and docking resulting in molecules with similarity to either deoxythymidine or ganciclovir as well as a good steric fit in the active site of TK. The two sets of interactions of dT and GCV were linked by a logical OR. The hydrogen-bonding interactions originating from TK were deleted; all other protein interactions were combined to group 0 interactions, i.e., they can be formed simultaneously with either group 1 (dT) or group 2 (GCV) interactions. Böhm's scoring method strongly favors hydrophilic interactions (score −4.7) in contrast to hydrophobic interactions (−0.7 ... −0.1). Therefore, hydrophilic interactions derived from the alignment templates dominate the placement of the molecular fragments, which is, however, modulated by protein specific hydrophobic interactions. Protein interactions can be considered as a perturbation to the template interactions.

Analysis. The results of the screening runs were evaluated by analyzing the corresponding pharmacophore enrichment within the 2000 best-scoring molecules. Pharmacophors were defined for deoxythymidine (Figure 6A) and ganciclovir (Figure 6B), and the pharmacophore search was conducted using the PH4[33] method of MOE.[31] The absolute positions of the pharmacophors with respect to the ligand coordinates resulting from protein alignment were used in order to avoid artifacts. The rank-dependent enrichment was calculated with a bin width of 200. Molecules that did not fit well into the pseudoreceptor generated by PrepA received a penalty due to close contacts with the bounding box atoms and are therefore not counted as hits if their total score was greater than zero.

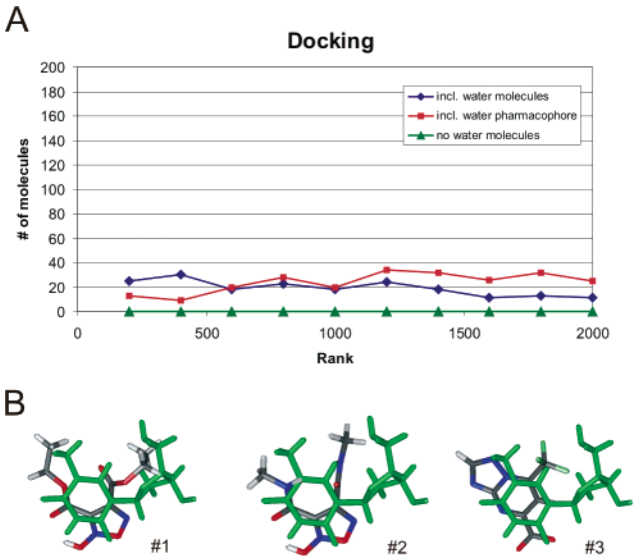

Figure 7. The enrichment (number of hit molecules vs rank) of the dT pharmacophore in various docking-based virtual screenings of a library of ∼50,000 small molecules using the PDB 1KIM structure is shown. (A) Two water molecules bound to Arg176 have to be included in the active site definition in order to find at least some hits. An ideal enrichment would result in 200 molecules found in the top-ranking bins (bin width 200); however, the enrichment here is generally low. Most the molecules found in the screening runs do not show the properties expected for biological activity (see text). For example, virtual screening without water molecules suggested substituted furoxanes (at rank #1 and #2) and acidic molecules (rank #3) as potential hits (B). The reference ligand dT is shown in green.

RESULTS AND DISCUSSION

Docking into Thymidine Kinase. Virtual screening of the ∼50,000 molecules based on protein−ligand docking revealed a generally low enrichment of hit molecules, i.e., dT and GCV-like binding molecules, even if the two water molecules bound to Arg176 in the 1KIM structure are included in the active site definition (Figure 7A). Due to the pharmacophore definition for screening analysis, a dT (GCV) hit indicates the presence of a dT-like (GCV-like) interaction pattern between the hit molecule and TK within a RMSD < 1.2 Å (see Figure 6). In contrast to a simple atom-based RMSD this hit definition captures the importance of the correct interaction types at the correct position within the active site. The examination of poses and their pharmacophoric properties in order to define hits is advantageous for the TK system due to the difficulties in scoring and predicting binding free energies, as explained later. In summary, the docking without water molecules found no hits within the 2000 (i.e. 4% of the library size) top scoring molecules, docking with two water molecules included detected 191 hits, and docking with two water pharmacophors revealed 239 hits. The 1KI2 screening identified 29 GCV-like hits and 1 dT-like hit. Additionally, a virtual screening was performed based on docking into a receptor model that incorporates the flexibility of Gln125 (see Material and Methods). Again, the enrichment was low with 39 GCV-pharmacophore hits and no dT-like hits within the 2000 top-scoring molecules (data not shown). The low enrichment rate for docking into TK is consistent with literature: Kellenberger et al. tested 8 docking programs with a library of 1000 molecules (990 random molecules, 10 TK inhibitors) and found a hit rate between 0% and 16% for a

hit list containing the Top 5% of the library.[22] Merlitz et al. found dT (GCV) at ranks between 16445 and 21206 (93−421) for a library of ∼180,000 molecules.[21] These results can be explained by the difficulties to integrate water molecules into the screening protocols and the nonoptimal binding mode of GCV with respect to the hydrogen-bonding to Arg176.[32]

In addition to the low enrichment, most of the top scoring molecules are highly undesirable for further investigations. For example, docking without water molecules revealed furoxane analogues as top scoring compounds (Figure 7B). The binding mode of these molecules was reasonable in general but tilted toward Arg176 compared to the substrate dT, thereby violating the hit criterion. Most of the following compounds contained an acidic group, e.g. a carboxylate, bound to Arg 176 by a salt bridge but lacking a reasonable interaction to Gln 125 or vice versa. Most of the hits did not show any similarity to known ligands of thymidine kinase. For several reasons they are unsuitable for further development: (i) It is well-known that dT binding to TK leads to a large conformational change that induces the formation of an ATP binding site.[23] Every drug acting on TK has to be able to induce a comparable conformational transition, otherwise the mode of drug action is undermined and the model of the active site is invalid making the virtual screening obsolete. (ii) A certain degree of similarity to nucleosides is essential for blocking viral DNA elongation, either by inhibition of viral DNA polymerase or by inducing a nonnatural DNA conformation. (iii) The furoxane analogues have adverse chemical properties. Therefore, most of the docking hits will exhibit a biological activity only by pure chance, which is highly undesirable as a reliable enrichment of real hits is the basis for efficient drug design.

Hits from a reliable virtual screening should at least reproduce the consensus interactions of nucleoside analogues, i.e., hydrogen bonds to Gln 125 and Arg 176 and hydrogen bonds to the residues near the sugar pocket. Established methods to improve the results include the application of more sophisticated scoring methods and pre/postfiltering approaches. The preference of Böhm's scoring approach for salt bridges is well-known. Using more elaborate scoring methods may avoid this particular scoring artifact. However, recent studies on large numbers of protein−ligand complexes using various established scoring functions revealed a poor predictive power for binding affinities in general.[27,34] In particular, the binding affinity to TK is difficult to estimate due to a large entropy contribution (dT + TK → TK-dT: $\Delta G = -30.1$ kJ/mol, $\Delta H = -79.9$ kJ/mol, $T\Delta S = -49.8$ kJ/mol),[23] which involves protein dimerization, conformational changes and accompanying changes in water structure. Generic scoring functions used in virtual screening do not reflect these protein-specific thermodynamic properties. Methods for calculating this entropy—for example free energy perturbation—are very demanding in terms of computational power and therefore unsuitable for high-throughput virtual screening, today. Even if scoring functions for virtual screening fail to predict correct binding affinities for the TK system, they, however, provide a means of ranking binding modes with respect to steric fit and interaction geometries.[27,34] Therefore modifications in the receptor definition are expected to represent a suitable method for steering virtual screening even with standard scoring methods.

As mentioned before, an additional problem arises from the mode of drug action on TK: active compounds not only have to be phosphorylated by TK but also have to bind to DNA polymerase as well. Such target-specific constraints that are not implemented in scoring functions can be applied by pre- and postfiltering of the molecules. However, substructure queries often fail to retrieve all suitable molecules due to the vast number of different heterocycles where a simple 2D structure alignment fails to predict the correct interaction-based similarity. A textbook example of this fact would be methotrexate and dihydrofolate.[35] Another approach would use docking into DNA polymerase as a filter before or after docking into TK. The binding site of DNA polymerase (see for example PDB 1T3N or 1JX4) is dominated by a triphosphate pocket which is assisted by an aromatic cluster and the required hydrogen bonding pattern to the complementary DNA nucleotide. However, the steric constraints are much less demanding compared to the active site of TK; therefore, the efficiency of docking into DNA polymerase as a pre- or postfilter is expected to be low. Additionally, a variety of DNA polymerase types exist, each of which would have to be examined for its usefulness. An approach based on substrate alignment is able to avoid these difficulties. Some established docking programs are capable of applying pharmacophore constraints during docking.[36] However, enforcing the presence of a single, rigid pharmacophore in all molecules does not consider the flexibility of such a pharmacophore due to receptor flexibility and limits the diversity of the results significantly.

This underlines the necessity of applying methods that outweigh the deficiencies of conventional docking methods and established scoring functions for virtual screening. A straightforward and efficient approach is the application of a priori knowledge in order to steer the virtual screening. Using ProPose, a priori information on binding modes and compounds with proven activity can be easily incorporated into the virtual screening protocol, which is described in the following paragraphs.

Alignment on Single Template Structures. Most of the known TK ligands are dT-analogues which differ only in the substituents at the 5-position of the uracil ring and the sugar type. These dT analogues exhibit very similar binding modes in the crystal structures (see Figure 5A). We therefore focused on two molecules with different core structures, namely dT itself—representing 5-substituted uracil analogues such as 5-iodo-2′-deoxyuridine (IDU), (E)-5-(2-bromovinyl)-2′-deoxyuridine (BVDU), 5-(5-bromothien-2-yl)-2′-deoxyuridine (BTDU), or 1,5-anhydro-2,3,dideoxy-2-(5-iodouracil-1-yl)-D-arabino-hexitol (AHIU)—and GCV—representing deoxyguanosine analogues such as penciclovir and aciclovir. Alignment of the screening library on dT as well as on GCV is a very efficient and selective means for enriching the corresponding pharmacophore: Figure 8 shows the pseudoreceptor for dT created by PrepA (Figure 8A) and the top scoring compounds from the virtual screening of the screening library (Figure 8B). As expected, dT itself was found at rank 1. An equivalent result was seen in an alignment screen on ganciclovir (data not shown). The pharmacophore enrichment within the 2000 top scoring molecules is shown in Figure 9A,B. Alignment on dT found 778 molecules with correct pharmacophore and one with a GCV pharmacophore. Alignment on GCV revealed 215 compounds with the

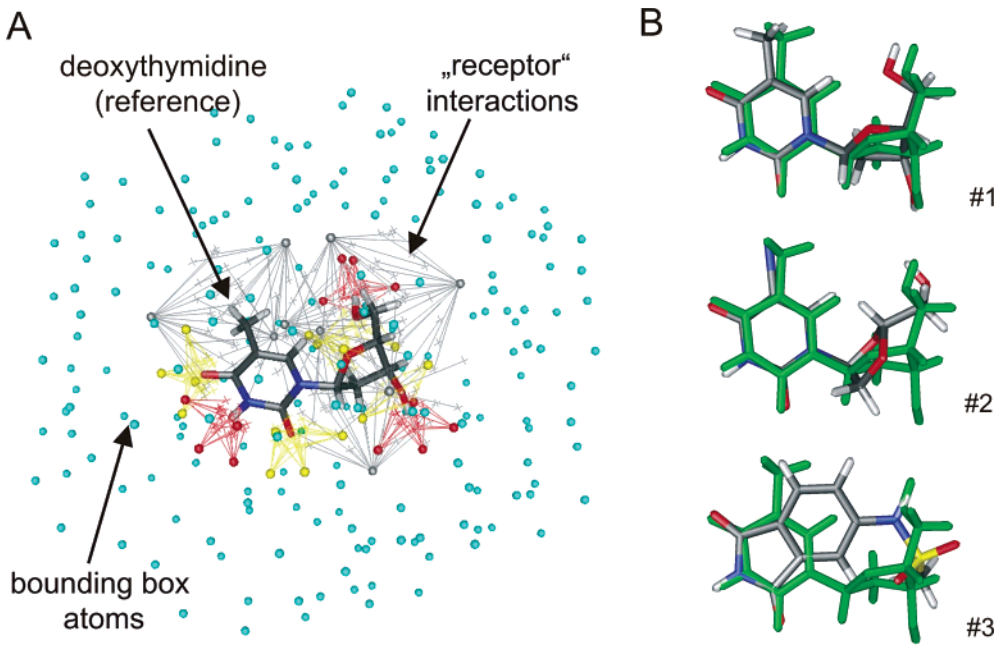

Figure 8. PrepA generates a pseudoreceptor reflecting the interaction possibilities of the template deoxythymidine (A). Potential receptor interactions are shown in gray (hydrophobic), red (hydrogen bond acceptors), and yellow (hydrogen bond donors). Bounding box atoms (cyan) focus the alignment to molecules of similar size compared to the reference ligand. Subsequently, ligand−ligand alignment is accomplished by docking into this pseudoreceptor. This revealed thymidine analogues as top scoring molecules (B). The three best-ranked molecules and the reference molecule (dT in green) are shown.

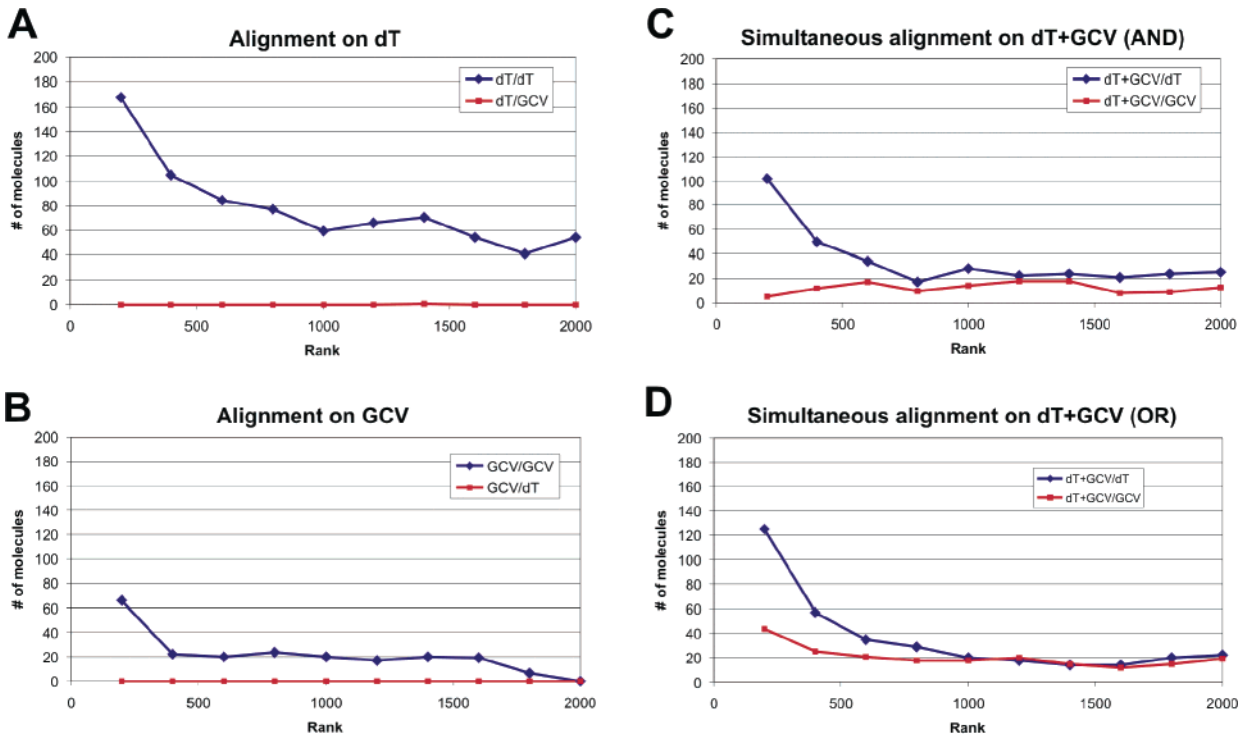

Figure 9. Alignment on deoxythymidine (A) and ganciclovir (B) selectively enriches molecules that exhibit a corresponding pharmacophore. The legend first specifies the alignment template and then—separated by a slash—the query pharmacophore. Here, the number of hit molecules vs rank is plotted, i.e., the curve represents the number of molecules, which contain the query pharmacophore, per bin (bin width 200) in dependence on the rank within the list of the 2000 top-scoring molecules. Simultaneous alignment on both targets with the interactions being linked by an logical AND (C) and OR (D) is able to enrich both binding modes. The amplitude of the curve for dT and GCV differs due to the total number of compounds within the screening library that contain the corresponding pharmacophore.

ganciclovir pharmacophore and none with a dT pharmacophore. In both cases, only the template pharmacophors were enriched within the top-scoring molecules, resulting in a low level of chemical novelty of the hits. However, in a drug discovery project it is desirable not only to enrich known compounds but additionally to uncover novel structures with similar binding modes.

Alignment on Dual Templates. An alternative to alignment on a single template structure is using templates derived from more than one ligand structure. This approach is easy

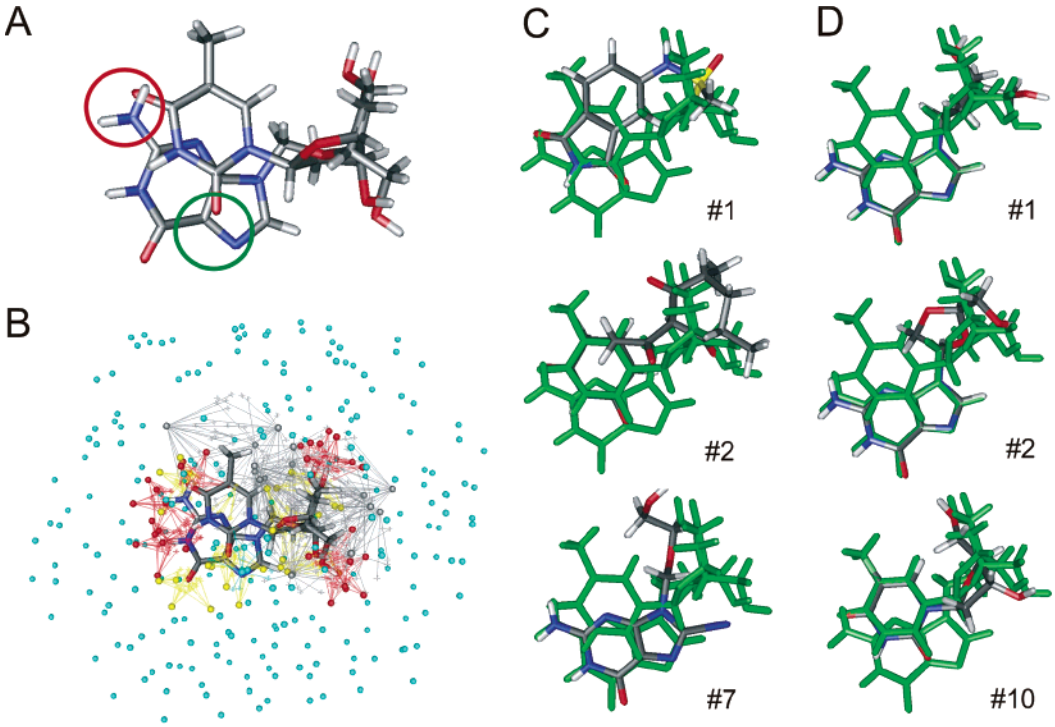

Figure 10. The template for simultaneous alignment on deoxythymidine and ganciclovir is depicted (A, B). Compatible and incompatible interactions are marked with green and red circles (see text). Linking interaction by a logical AND mainly enriches deoxythymidine-like molecules (C), whereas a logical OR enriches both binding modes (D). The numbers indicate the rank within the ranking list. The dual template is shown in green.

to implement with ProPose simply by concatenating the corresponding TDFs, adjusting the interaction indices, and generating a bounding box that covers both ligands. ProPose allows one to link sets of interactions—usually originating from different molecules—by logic operations (see Table 2, @LOGIC): the interaction sets are used either simultaneously (logical AND) or exclusively (logical OR). Both logic operations were tested on a TDF containing the interactions of dT and GCV (Figure 10A,B).

A logic AND of the interactions during alignment resulted in 347 dT and 124 GCV pharmacophors among the 2000 top scoring molecules from the screening library. The ranking list now contains molecules of both binding modes (Figure 10C). This approach found 44.6% and 57.6% of the dT and GCV pharmacophors, respectively, compared to the hits from single template alignment. The pharmacophore enrichment is shown in Figure 9C. While molecules with a dT pharmacophore are clearly enriched, the enrichment of GCV-like binding molecules is not so obvious. This is due to a complex interference between the two sets of interactions. Depending on the particular spatial arrangement of the interaction points some interaction pairs originating from different sets are easily formed by molecules simultaneously, while others due to their incompatibility are not. This leads to a preference for some of the interactions which receive an artificially high score. For example, the amine and the carbonyl which are marked by a red circle in Figure 10A are incompatible, while the hydrogen bond acceptor functionality of the carbonyl and the nitrogen—marked by a green circle—are compatible. Therefore a single interaction partner for the later interaction will score twice in contrast to a potential partner of the amine/carbonyl interactions which will score only once but will additionally receive a penalty for being close to the incompatible interaction. This may lead to scoring artifacts, as can be seen from the poor enrichment of the GCV-like binding molecules. However, the logic AND can be very helpful when the interference of interactions is negligible. For example, to promote the binding of charged fragments to an oxanion hole a single positive charge interaction can be positioned in the center of such a hole in order to act as a pharmacophore. Such an interaction does not have to be associated with a real atom. Due to the Lorentzian averaging scheme in ProPose no scoring artifacts will arise.

Linking the interaction sets by a logical OR is the recommended method to deal with different sets of interactions. This is proven by aligning the screening library on the dual template consisting of dT and GCV with the interactions of both molecules linked by an "OR": now 207 (96.3%) and 354 (45.5%) of the GCV-like and dT-like binding molecules, respectively, were found among the 2000 top scoring molecules (Figure 10D). The enrichment is clearly visible for both pharmacophors (Figure 9D). However, the method does not reveal all of the hits from alignment on single targets. But the decreased number of hits is paid-off by an increased diversity of the hits compared to conventional alignment. Furthermore, the hit list still contains much more potential hits than revealed by pure docking approaches (see Figure 7). The better enrichment rate for GCV-like molecules can be attributed to their higher number of possible hydrogen bonds (two more than dT). Measured binding affinities (K_m) of dT and GCV were 0.2 and 47 μM, respectively.[24] However, the essential parameter for successful virtual screening is the enrichment, not the absolute scoring.

Alignment on Dual Templates with Docking Constraints. The protein structure of TK may be combined with the dual template described in the previous paragraph, leading to simultaneous docking and alignment: the generic bound-

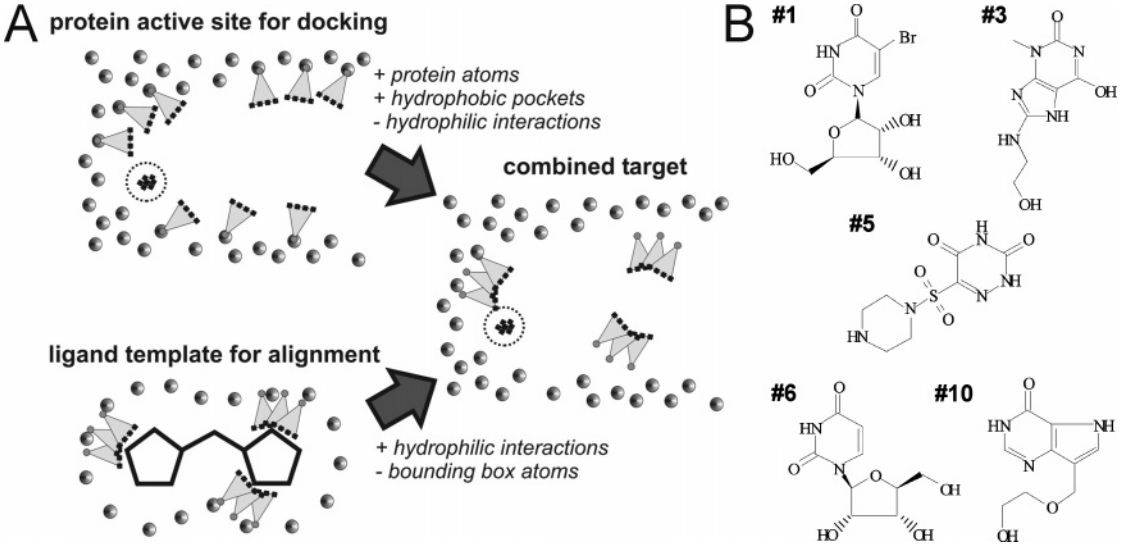

Figure 11. The methodology and the results of a simultaneous alignment on deoxythymidine and ganciclovir in combination with docking into the active site of thymidine kinase are shown. (A) The combined alignment/docking target is generated by merging the hydrophilic interactions (gray triangles) of the alignment template with the protein atoms (spheres) and their hydrophobic interactions (marked by a dashed circle). Hydrophilic interactions of the protein and the generic bounding box atoms of the alignment template are removed. If more than one protein or template conformation has to be considered, additional interactions are added and assigned to interaction groups. (B) This method detects molecules within the screening library that are structurally related to nucleosides and additionally fit sterically into the protein's active site. The numbers indicate the ranks within the 2000 top scoring molecules resulting from virtual screening.

ing box atoms are replaced with the protein atoms. Additionally, the nonpolar interactions of TK are added to the TDF as group 0 interactions (see Methods and Figure 11A). This strategy allows to combine the specificity of alignment with the diversity gained by the dual template approach while ensuring a proper steric fit of the molecules into the active site of the target protein: Using Böhm's scoring method hydrophilic interactions receive a much better score (−4.7) compared to hydrophobic interactions (−0.7 ... −0.1). Therefore the placement of the molecule in the active site is dominated by the hydrophilic interactions of the alignment templates but still modulated by protein specific hydrophobic interactions. From a theoretical point of view, the protein interactions are a perturbation to the template interactions. However, the protein atoms act as a specific bounding box and therefore enforce a correct steric fit into the active site.

The single-scan docking and alignment now reveals molecules with the appropriate similarity to nucleosides and a good steric fit into the active site of TK. Figure 11B shows a selection of top-scoring molecules. All molecules are nucleoside analogues including bromouracil (**1**), purine (**3**), triazine (**5**), uracil (**6**) and pyrrolopyrimidine (**10**). Molecules **1**, **3**, **6**, and **10** contain a hydroxy moiety that may be potentially linked to phosphate, whereas molecule **5** is more likely to be a TK inhibitor. To summarize, these molecules are chemically diverse but still similar to nucleosides and therefore suitable for the next step—a biological assay—which is, however, beyond the scope of this manuscript.

Note that a conventional pre-/postfiltering approach would miss several of the depicted molecules. For example, a 2D substructure search for uracil would discard molecules **3**, **5**, and **10**. A pure 3D pharmacophore search for uracil's most important interaction pattern—acceptor, donor, acceptor—will find **5** additionally but still neglects **3** and **10**. A query using the SMARTS pattern "[*]1~[#6](~[#8])~[#7]~[#6](~[#8])~[*]~[*]1" returns molecules **1**, **3**, **5**, and **6** but still misses **10**. Therefore, our method is clearly superior to the conventional methods due to its configurable selectivity—diversity tradeoff: the user can steer the virtual screening by providing a diverse set of target structures, and ProPose will enrich molecules selectively that are compatible with these templates while still providing a diverse set of possible hits. Additionally, the alignment-based approach considers the conformational flexibility which is difficult to integrate into substructure or pharmacophore queries.

The alignment on a dual template with docking constraints implicitly considers the effect of the different Gln 125 conformation in the dT and GCV-bound states. In general, it is possible to provide not only a number of ligand binding modes but also equivalently a number of different protein conformations with their interactions linked by a logical OR, which directly incorporates protein flexibility into the docking protocol (see section "Docking into Thymidine Kinase"). For each incremental construction step, the scoring routine decides on-the-fly which protein (for docking) or template (for alignment) conformation leads to the best score for the respective ligand. During the construction a number of poses (typically 100−1000) are traced, each of which may interact with a protein/template conformation that receives an optimal score for interacting with the current pose. After completion of the construction, a defined number of ligand poses is stored. This allows one to consider protein and template flexibility in a straightforward manner by providing the appropriate interactions in the TDF. As the number of interactions influences computing times, one should focus on a limited number of conformations or reduce the number of interactions per conformation.

Computational Efficiency. While screening the library of ~50,000 molecules on Intel Pentium (Intel Corporation) III 1.3 GHz computers, the average computing times for docking, alignment, multiple alignment, and combined alignment and docking were 6.2, 4.2, 7.2, and 10.3 min per

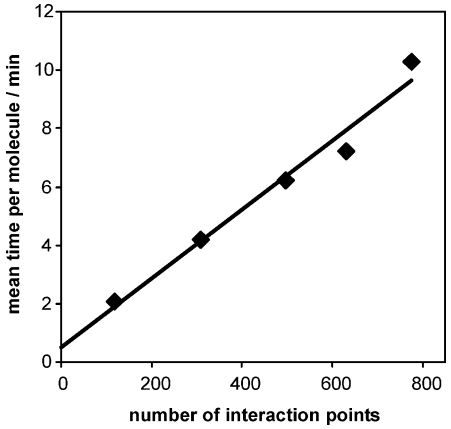

Figure 12. The mean computing time per molecule strongly correlates with the number of interaction points defined in the target description file. Therefore the computing time can be optimized by reducing the number of interaction points.

molecule, respectively. A strong correlation of the computing time with the number of interactions points in the TDF was found (Figure 12).

The prolonged computing times originate solely from the scoring routine with only a minimal administration overhead. The evaluation of the logical expression has no significant additional effect on computational efficiency. The scoring routine itself utilizes hash data structures to efficiently cope with large numbers of interactions.[11] But a strong impact on the computing time results from the potentially large number of poses before pose clustering, which may reach $\sim 10^4$ poses. For example, adding 10 conformations of an asymmetrically substituted aliphatic ring fragment with 6 different torsion angles to 790 previously generated poses would result in a total number of 47,400 new poses. Presumably, most of these can be discarded before scoring due to clashes with other atoms, but still a significant number of poses has to be scored. Therefore a compromise between conformational sampling and speed/memory requirements has to be found. ProPose allows for a user-defined tradeoff by intermediary clustering and garbage-collection stages and by a manual control of the torsion angle library.

To summarize, the computing time can be scaled by adding or removing interaction points from the TDF. A method for reducing the number of interactions has been proposed recently.[37] As a default, PrepD clusters hydrophobic interaction points and thereby reduces their number. A similar algorithm can be applied for all other interaction types. For example, using a minimal representation (1 point per interaction) of the combined alignment/docking target dT+GCV(OR), now comprising 119 interaction points, the mean time per molecule drops to 2.1 min. Additionally, the concatenation of two TDFs does not have to double the number of interaction points if a significant number of shared interactions is present which can be merged to common (group 0) interactions.

In conclusion, an overnight run on a subset of 500 CPUs on an in-house Linux cluster is sufficient to screen the library of ~50,000 molecules explicitly, even if the mean time per molecule is 6−7 min. Upon integration into the intelligent database screening system 4SCan,[38] these methods can be used to screen effectively our complete in-house database of ~10 millions of molecules by iteratively docking only a subset of 10×5000 molecules. Therefore the main design aspect of ProPose was to provide a generally applicable and flexible platform for data mining in molecular databases with adequate speed optimization for specific applications. However, efforts are ongoing in order to optimize the computing times while preserving the general concept.

SUMMARY

We have shown that combinations of molecular alignment and docking in conjunction with logic operations can be used to steer the virtual screening for ligands of HSV thymidine kinase. This method has several advantages compared to conventional docking and alignment approaches: (i) it allows the integration of a priori knowledge into screening protocols in an efficient manner and significantly increases (ii) the enrichment of hit molecules for this difficult docking target and (iii) the diversity of the molecules compared to conventional alignment approaches. Our method does not improve, in general, the prediction of binding affinity but does however provide a target-specific, empirical ranking based on a priori information.

Our method presents a novel concept for virtual screening: different types of a priori information—including protein structures, ligand data, and pharmacophors from medicinal chemistry—are transformed into a pharmacophore-based potential for docking. This concept allows one to perform virtual screening based on protein−ligand docking, ligand−ligand alignment, and pharmacophore queries—and combinations thereof—with a single program. In contrast to established software, this general concept is easy to implement in our docking program ProPose due to its fully configurable interaction and scoring scheme. This allows for the incorporation of various important effects into screening protocols. For example, protein flexibility—either derived from different ligand binding modes or directly from different protein structures—is considered implicitly by the help of logic operations on sets of interactions.

Today, structure-based drug design is able to identify potent inhibitors for most of the target proteins of pharmaceutical interest. Enzyme inhibition, however, is only one part of the game: a high attrition rate of drug candidates at the (pre)clinical stage is usually caused by negative chemical and biological properties of the compounds.[39,40] Besides ADMET properties the physiological pathway of drug action has to be considered. A simple example was discussed in this manuscript: the mode of action of HSV TK targeting drugs requires that molecules not only bind to TK but additionally are nucleotide-mimics. Our ligand-based extension to conventional docking allows one to consider such chemical and biological factors in the early stages of virtual screening and may therefore provide a better starting point for lead identification and finally drug development.

Remarks on Nomenclature. ProPose handles protein−ligand docking, ligand−ligand alignment, and pharmacophore queries with an identical algorithm. Therefore "docking a ligand" may refer to docking as well as to alignment, which is accomplished by docking into a pseudoreceptor. A "target" may be a protein structure, a template ligand, a pharmacophore, or a combination thereof.

Abbreviations. ACV − aciclovir; ADMET − absorption, distribution, metabolism, excretion, and toxicity; ATP −

adenosine triphosphate; dT − deoxythymidine; dTMP − deoxythymidine monophosphate; GCV − ganciclovir; HSV − Herpes Simplex Virus; NMP − nucleotide monophosphate; PDB − Protein Data Bank; PrepA − preparation tool for alignment; PrepD − preparation tool for docking; QSAR − quantitative structure−activity relationship; RMSD − root-mean-square deviation; TDF − target description file; TK − thymidine kinase; TPSA − topological polar surface area; SlogP − log(octanol/water) partition coefficient; WDI − world drug index.

ACKNOWLEDGMENT

Tobias Kailich contributed to the parameter optimization. Bernd Kramer, Thomas Herz, Kristina Wolf and Jürgen Kraus promoted the development of ProPose with their ideas and bug reports. We thank Cathal Meere for proof-reading. Frank Schmitt was a great help in solving programming problems in C++. Daniel Vitt is acknowledged for his unwavering support.

REFERENCES AND NOTES

(1) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2002**, *28*, 235−242, http://www.rcsb.org/pdb/.

(2) Thomson Scientific, 3501 Market Street, Philadelphia, PA 19104, USA. http://thomsonderwent.com/products/lr/wdi/.

(3) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*(3), 151−166.

(4) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: an overview of search algorithms and guide to scoring functions. *Proteins* **2002**, *47*(4), 409−443.

(5) Girones, X.; Carbo-Dorca, R. TGSA-Flex: Extending the capabilities of the Topo-Geometrical Superposition Algorithm to handle flexible molecules. *J. Comput. Chem.* **2004**, *25*(2), 153−159.

(6) Bultinck, P.; Kuppens, T.; Girones, X.; Carbo-Dorca, R. Quantum similarity superposition algorithm (QSSA): a consistent scheme for molecular alignment and molecular similarity based on quantum chemistry. *J. Chem. Inf. Comput. Sci*. **2003**, *43*(4), 1143−115.

(7) Melani, F.; Gratteri, P.; Adamo, M.; Bonaccini, C. Field interaction and geometrical overlap: a new simplex and experimental design based computational procedure for superposing small ligand molecules. *J. Med. Chem.* **2003**, *46*(8), 1359−1371.

(8) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*(23), 4502−4520.

(9) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*(3), 215−232.

(10) Fradera, X.; Mestres, J. Guided docking approaches to structure-based design and screening. *J. Curr. Top. Med. Chem.* **2004**, *4*(7), 687−700.

(11) Seifert, M. H. J.; Schmitt, F.; Herz, T.; Kramer, B. ProPose: a docking engine based on a fully configurable protein-ligand interaction model. *J. Mol. Model. (Online)* **2004**, *10*, 342−357.

(12) Vedani, A.; Zbinden, P.; Snyder, J. P. Pseudoreceptor modeling: a new concept for the three-dimensional construction of receptor binding sites. *J. Recept. Res.* **1993**, *13*(1−4), 163−177.

(13) Zbinden, P.; Dobler, M.; Folkers, G.; Vedani, A. PrGen: Pseudoreceptor modeling using receptor-mediated ligand alignment and pharmacophore equilibration. *Quant. Struct.-Act. Relat.* **1998**, *17*, 122−130.

(14) McMartin, C.; Bohacek, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*(4), 333−344.

(15) Traversa, U.; Bombi, G.; Camaioni, E.; Macchiarulo, A.; Costantino, G.; Palmieri, C.; Caciagli, F.; Pellicciari, R. Rat brain guanosine binding site. Biological studies and pseudoreceptor construction. *Bioorg. Med. Chem*. **2003**, *11*(24), 5417−5425.

(16) Walters, D. E.; Hinds, R. M. Genetically evolved receptor models: a computational approach to construction of receptor models. *J. Med. Chem.* **1994**, *37*(16), 2527−2536

(17) C2.Receptor: Accelrys, Inc.; 10188 Telesis Court, Suite 100; San Diego, CA 92121; U.S.A. http://www.accelrys.com/cerius2/receptor_ms.html.

(18) Hahn, M. A.; Receptor surface models. 1. Definition and construction. *J. Med. Chem.* **1995**, *38*, 2080−2090.

(19) Pei, J.; Zhou, J.; Xie, G.; Chen, H.; He, X. PARM: a practical utility for drug design. *J. Mol. Graph Model*. **2001**, *19*(5), 448−454, 472−473.

(20) Pospisil, P.; Kuoni, T.; Scapozza, L.; Folkers, G. Methodology and problems of protein−ligand docking: case study of dihydroorotate dehydrogenase, thymidine kinase, and phosphodiesterase 4. *J. Recept. Signal Transduct.* **2002**, *22*(1), 141−154.

(21) Merlitz, H.; Herges, T.; Wenzel, W. Fluctuation analysis and accuracy of a large-scale in silico screen. *J. Comput. Chem.* **2004**, *25*(13), 1568−1575.

(22) Kellenberger, E.; Rodrigo, J.; Müller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening. *Proteins* **2004**, *57*(2), 225−242.

(23) Perozzo, R.; Jelesarov, I.; Bosshard, H. R.; Folkers, G.; Scapozza, L. Compulsatory Order of substrate binding to herpes simplex virus type 1 thymidine kinase. *J. Biol. Chem.* **2000**, *275*(21), 16139−16145.

(24) Wurth, C.; Kessler, U.; Vogt, J.; Schulz, G. E.; Folkers, G.; Scapozza, L. The effect of substrate binding on the conformation and structural stability of Herpes simplex virus type 1 thymidine kinase. *Protein Sci.* **2000**, *10*, 63−73

(25) Fischer, U.; Steffens, S.; Frank, S.; Rainov, N. G.; Schulze-Osthoff, K.; Kramm, C. M.; Mechanisms of thymidine kinase/ganciclovir and cytosine deaminase/5-fluorocytosine suicide gene therapy-induced cell death in glioma cells. *Oncogene*. **2005**, *24* (7), 1231−1243.

(26) Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Busson, R.; Rozenski, J.; Herdewijn, P.; Van Calenbergh, S. Discovery of Bicyclic Thymidine Analogues as Selective and High-Affinity Inhibitors of *Mycobacterium tuberculosis* Thymidine Monophosphate Kinase. *J. Med. Chem.* **2004**, *47*(25), 6187−6194.

(27) Wang, R.; Lu, Y.; Fang, X.; Wang, S. An Extensive Test of 14 Scoring Functions Using the PDBbind Refined Set of 800 Protein−Ligand Complexes. *J. Chem. Inf. Comput. Sci*. **2004**, *44*(6), 2114−2125.

(28) Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*(3), 243−256.

(29) Wildman, S. A.; Crippen, G. M. Prediction of Physiochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*(5), 868−873.

(30) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.

(31) Molecular Operating Environment. Chemical Computing Group, Inc., 1010 Sherbrooke St. West Suite 910, Montreal, Quebec H3A2R7, Canada (http://www.chemcomp.com/Corporate_Information/MOE.html).

(32) Champness, J. N.; Bennett, M. S.; Wien, F.; Visse, R.; Summers, W. C.; Herdewijn, P.; de Clerq, E.; Ostrowski, T.; Jarvest, R. L.; Sanderson, M. R. Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. *Proteins* **1998**, *32*(3), 350−361.

(33) PH4 pharmacophores: http://www.chemcomp.com/Journal_of_CCG/Articles/ph4.htm.

(34) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein−ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.

(35) Böhm, H. J.; Klebe, G.; Kubinyi, H. In *Wirkstoffdesign* [*Drug design*]; Spektrum Akademischer Verlag: Heidelberg, Germany, 1996; section 17.7, p 320.

(36) Hindle, S. A.; Rarey, M.; Buning, C.; Lengauer, T. Flexible docking under pharmacophore type constraints. *J. Comput.-Aided Mol. Des.* **2002**, *16*(2), 129−149.

(37) Schellhammer, I.; Rarey, M. FlexX-Scan: fast, structure-based virtual screening. *Proteins* **2004**, *57*(3), 504−517.

(38) Seifert, M. H. J.; Wolf, K.; Vitt, D. Virtual high-throughput in-silico screening. *DDT Biosilico* **2003**, *1*, 143−149.

(39) Ajay. Predicting drug-likeness: why and how? *Curr. Top. Med. Chem.* **2002**, *2*(12), 1273−1286.

(40) Oprea, T. I. Virtual screening in lead discovery: a viewpoint. *Molecules* **2002**, *7*, 51−62.

CI0496393