Using Pharmacophore Models To Gain Insight into Structural Binding and Virtual Screening: An Application Study with CDK2 and Human DHFR

Samuel Toba,* Jayashree Srinivasan, Allister J. Maynard, and Jon Sutter

Accelrys Inc., 10188 Telesis Court, Suite 100, San Diego, California 92121

Received September 18, 2005

This study provides results from two case studies involving the application of the HypoGenRefine algorithm within Catalyst for the automated generation of excluded volume from ligand information alone. A limitation of pharmacophore feature hypothesis alone is that activity prediction is based purely on the presence and arrangement of pharmacophoric features; steric effects remained unaccounted. Recently reported studies have illustrated the usefulness of combining excluded volumes to the pharmacophore models. In general, these excluded volumes attempt to penalize molecules occupying steric regions that are not occupied by active molecules. The HypoGenRefine algorithm in Catalyst accounts for steric effects on activity, based on the targeted addition of excluded volume features to the pharmacophores. The automated inclusion of excluded volumes to pharmacophore models has been applied to two systems: CDK2 and human DHFR. These studies are used as examples to illustrate how ligands could bind in the protein active site with respect to allowed and disallowed binding regions. Additionally, automated refinement of the pharmacophore with these excluded volume features provides a more selective model to reduce false positives and a better enrichment rate in virtual screening.

INTRODUCTION

The uses of structure-based and ligand-based methods with rational drug discovery have been fairly separate approaches. Structure-based methods have advantages in lead optimization; however, these require the availability of a known protein structure or a good homology model. On the other hand, ligand-based approaches are more qualitative but do not require prior knowledge of targets or an available protein structure. There have been attempts to combine methodologies such as pharmacophore docking, etc.[1−4]

In addition, it has always been a quest for medicinal and computational chemists to be able to gain insight into protein structure from ligand information alone. Pharmacophore modeling is one method that allows scientists to gain valuable information of how ligands bind in 3D space in the protein site. Several review articles[5−12] and a book[7] have been devoted to the topic of pharmacophores. Recently, several papers have further extended this approach by combining excluded volumes to pharmacophores. The use of excluded volumes can result in a more stringent pharmacophore, resulting in a more selective and predictive pharmacophore.[13−15] However, most of the studies reported have employed nonautomated methods that require the knowledge of a protein target and its structure. The manual placement of an excluded volume can also be very tedious for users relying on a trial-and-error approach, and users may miss some possible placement sites. The availability of an automated approach for adding excluded volumes to pharmacophores using solely ligand based information can ensure extensive sampling of all possible space placements and requires no prior knowledge of the target or the structure.

The use of excluded volume spheres further provides an insight to the disallowed binding regions in the protein-binding site for the ligands. In this study, we provide two case studies, CDK2 and human DHFR (hDHFR), that illustrate the application of automated excluded volume and pharmacophore model generation. It is hoped that these studies can help provide an insight into structural binding in each of the biological systems.

MATERIALS AND METHODS

All molecular modeling calculations were performed on an IBM Intellistation M Pro Intel P4 machine running RedHat Enterprise Linux WS2.1. Recent versions of the Catalyst software[16] were used to automatically generate pharmacophore models (Catalyst HypoGen module) and pharmacophore models with excluded volumes (Catalyst HypoGenRefine). The HypoGenRefine algorithm is an extension of the Catalyst HypoGen algorithm.[17]

A brief overview of the theory is as follows. The constructive phase of HypoGenRefine is analogous to the HypoGen algorithm; however, the subtractive phase is not performed. Rather, the addition of exclusion volumes is included as one of the perturbations in the simulated annealing optimization phase of the pharmacophore model generation. The best mapping of an inactive compound is used to target the addition of the excluded volumes such as to reduce that compounds predicted activity. Atoms in this best inactive mapping that are greater than a threshold distance from all atoms of the best mapping of each active compound are used as candidate locations for excluded volume addition. A random selection is made from these candidates, and the HypoGen cost function is evaluated. The perturbation will then be accepted or rejected based on the

* Corresponding author e-mail: stoba@accelrys.com.

10.1021/ci050410c CCC: $33.50
Published on Web 02/18/2006

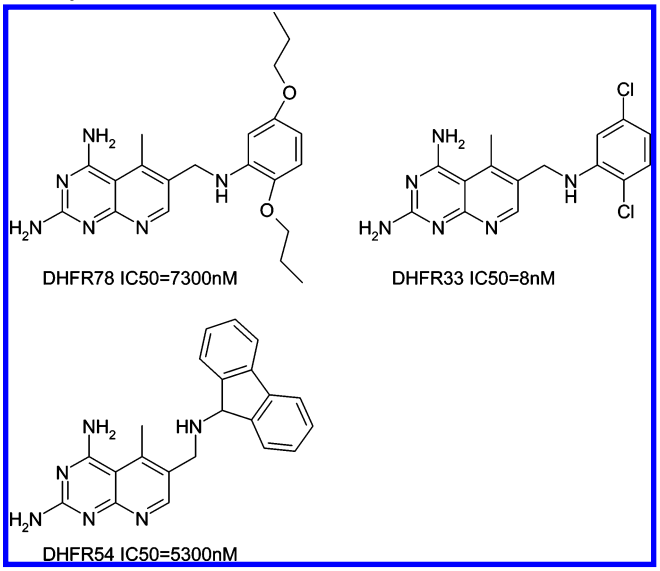
rules of the simulated annealing protocol. In this study, the parameters used to generate the pharmacophore models involved using the default settings, and any deviations are noted below.

Human DHFR data were obtained from the published work by A. K. Debnath,[18] and CDK2 data were obtained from the published work by Yue et al.[19] and Nugiel et al.[20] Conformer generation for all the data sets was performed using the FAST mode option (catConf module in Catalyst) with an energy threshold of 20 kcal/mol and a maximum of 250 conformers. A recent study by Langer et al.[21] has shown that Catalyst FAST conformer model generation is appropriate for finding 'bioactive' conformers most of the time.

Pharmacophores for CDK2 have been generated using a set of 21 training set compounds from a set of 57 published by Yue et al.[19] from Bristol Myers Squibb based on the novel indenopyrazole scaffold. The algorithm requires a small number of training set compounds for model generation. Ideally, this set of compounds would have maximal structural diversity and span a wide range of activity. The resultant 21 training set compounds were manually selected in an attempt to satisfy these requirements, resulting in a training set ranging in activity with IC50 values from 15 to 3700 nM. The remaining 36 compounds from the publication were retained as a validation set. An additional 11 compounds were used as another validation set based on a related publication from DuPont Pharmaceuticals Company by Nugiel et al.[20] These additional compounds were assayed using the same methodology and assay conditions as Yue et al. A third validation set was constructed from a collection of literature publications consisting of a set of nine compounds known to be actives against CDK2 as compiled by Yue et al.[19] A total of 56 compounds in the validation sets were divided into three categories of activities as follows: ***, highly active (<500 nM); **, moderately active (>500 nM and <1000 nM), and *, inactive (>1000 nM). Four features consisting of a hydrogen bond acceptor (A), a hydrogen bond donor (D), a hydrophobe (H), and a positive ionizable group (P) were selected in the generation of Catalyst HypoGen and HypoGenRefine pharmacophore models. As mentioned earlier, all Catalyst default parameters were used, with the exception of the Uncertainty value which was set to 1.1 for all compounds. Catalyst uses the uncertainty value in the algorithm to identify the most active compounds and others that can qualify as lead active compounds. Details of the uncertainty value calculations and their applications in the HypoGen algorithm have been covered in elsewhere.[7] In the use of HypoGenRefine to generate pharmacophore models with an excluded volume, a maximum of five excluded volume spheres were allowed.

Pharmacophores for hDHFR were generated using the compounds published by A. K. Debnath.[18] Catalyst HypoGen pharmacophore models were generated in accordance to the original author's selection of 20 compounds for the training set to reduce redundancy in structural features and activities. Catalyst HypoGenRefine pharmacophore models were generated using the same 20 compounds from the HypoGen training set plus an additional two inactive compounds selected from the data set that contained steric functional groups (see Chart 1). Five features consisting of ADHP and a ring aromatic (R) were selected in the generation of Catalyst HypoGen and HypoGenRefine pharmacophore models. In the use of HypoGenRefine to generate pharmacophore models with excluded volume, a maximum of five excluded volume spheres were allowed.

Chart 1: Example on Effect of a Bulky Substituent Group on HDHFR Activity[a]



[a] Compound DHFR33 is active against hDHFR, but compounds DHFR78 and DHFR54 are inactive against hDHFR due to the steric effect from side-chain substituents. Both inactive compounds DHFR54 and DHFR78 are added to the training set of HypoGenRefine for the automated pharmacophore refinement using steric information from the excluded volume.

RESULTS AND DISCUSSION

In the CDK2 system, 10 pharmacophore models were generated for each HypoGen and HypoGenRefine calculation. We selected the first pharmacophore model for each output based on the highest training set correlation (HypoGen R=0.92, HypoGenRefine R=0.93) value between the estimated and actual activity data and results from Fisher randomization validation.[22] In the Fisher randomization tests, we created a 19-scrambled training data set by randomly matching the compounds and their activities. Using this scrambled data set, we generated HypoGen and HypoGenRefine models and compared the Catalyst cost values to the selected model generated from the real (nonscrambled) training set. The Catalyst theoretical cost provides an empirical scoring which represents the successes of any pharmacophore hypothesis; the dominant contribution to the total cost include the correlation value of the predicted vs true activity value. The scope of the Catalyst cost function has been published, and readers are referred to the publication for details.[7] For a model to achieve 95% significance at a 95% confidence level, the total cost for the true model must be higher than the total cost from each of the other 19 randomized runs. We used the Fisher randomization as part of the automated statistical validation tool available in Catalyst. The selected HypoGen pharmacophore hypothesis contained three hydrophobic features and two hydrogen bond donor features. Similarly, the selected HypoGenRefine pharmacophore hypothesis contained three hydrophobic, two hydrogen bond donor, and an additional four excluded volume features (Figure 1).

The selected models were validated against the remaining 36 compounds from the same published results using the

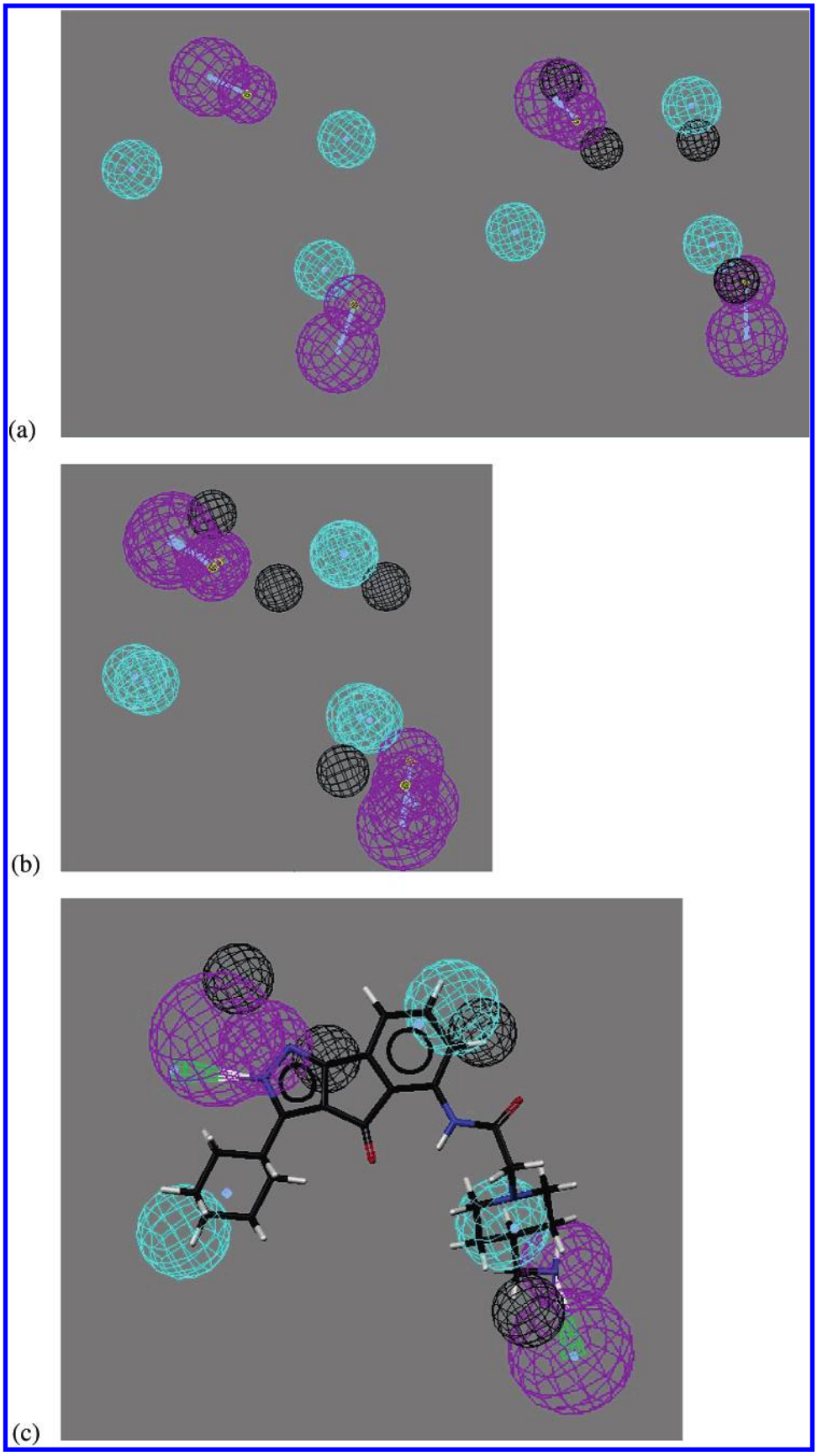


Figure 1. An overlay of the HypoGen model numbered 12M1C (top-left, a) and the HypoGenRefine model numbered 5M1C (top-right, a) for CDK2 is represented in b. In this overlaid representation (b), the hydrogen bond donor and hydrophobic groups that form the hypothesis are in the 2D plane. The excluded volume features (EV) generated by HypoGenRefine contains location and geometric constraints with respect to the other features in the hypothesis. As a result of the overlaid representation, the EV features are constrained to out-of-plane orientation with respect to the other pharmacophoric features. In this preferred mapping, a low pharmacophore overlay RMSD = 0.15 Å is obtained. The features are hydrophobic (cyan), hydrogen bond donor (magenta), and excluded volume (gray). The overlaid representation indicates a preference for placements of EV behind the molecule (compound 13q) as shown in c.

Table 1: CDK2 Validation from Yue et al.[19] [a]

name	exptl	pharmacophore only	with excluded volume
IP 24i	***	***	***
IP 24e	***	***	***
IP 24g	***	***	***
IP 24j	***	***	***
IP 24f	***	***	***
IP 13k	***	***	***
IP 24m	***	***	***
IP 13l	***	***	***
IP 8b	***	***	***
IP 13m	***	***	***
IP 8h	***	***	***
IP 13p	***	***	***
IP 24k	***	***	***
IP 13s	***	*	*
IP 8f	***	***	***
IP 24l	***	***	***
IP 5m	***	***	***
IP 24h	***	***	***
IP 24c	***	***	***
IP 5b	***	***	***
IP 13t	***	*	*
IP 13n	***		
IP 13u	***	*	*
IP 5c	***	***	***
IP 5j	***	***	***
IP-5h	***	***	***
IP 5d	***	***	***
IP 5e	**	***	***
IP 13r	**	*	*
IP 5k	**	***	***
IP 13i	*	*	*
ip-13w	*	*	*
IP 5i	*	***	**
IP 13v	*	**	**
IP 13f	*	*	*
IP 13a	*	*	*

[a] 80% of validation correctly predicted.

Table 2: CDK2 Additional Validation from Nugiel et al.[a]

name	exptl	pharmacophore only	with excluded volume
IP 1a	***	***	***
IP 3d	***	***	***
IP 3f	***	***	***
IP 4a	***	***	***
IP 3h	***	***	***
IP 35a	***	***	***
IP 33a	***	***	***
IP 36a	***	***	***
IP 37a	***	***	***
IP 38a	***	***	***
IP 39a	***	***	***

[a] All highly actives compounds were correctly predicted.

three binned categories: highly actives, moderately actives, and inactive. We found that the model accurately predicted 80% of the validation set, see Table 1. A validation was also carried out on the 11 additional compounds classified as highly actives against CDK2 using the same biological assay from Nugiel et al. publications. Both Catalyst HypoGen and HypoGenRefine accurately predicted these compounds as highly actives, Table 2. When we compared the models generated with a set of nine compounds known to be actives against CDK2 from literature publications, we found that the HypoGen model underpredicted two compounds (compound id: PD172803 and AG12286) and the HypoGenRefine

Table 3: Known CDK2 Compounds[a]

name	exptl	pharmacophore only	with excluded volume
olomoucine	***	***	***
flavopiridol	***	***	***
butyrolactone	***	***	***
GW8510	***	***	***
PD172803	***	**	***
AG12286	***	**	*
roscovitine	***	***	***
purvalanol A	***	***	***
purvalanol B	***	***	***

[a] PD172803 and AG12286 are reported actives against CDK2/cyclinA, but the model was generated based on activity information for CDK2/cyclinE. We view the difference in activity prediction as promising. One possibility is that our Catalyst models contain the selectivity necessary to capture this slight difference in the biological assay condition. Nevertheless, we cannot fully confirm the model selectivity and predicted activity differences until these two compounds are subjected to the same biological assay used in the training set compounds.

underpredicted one compound (compound id: AG12286), as seen in Table 3. However, upon further investigation, we discovered that both PD172803 and AG12286 activities were reported for CDK2/CyclinA. This is different from our Catalyst models that have been generated from compounds reporting actives against CDK2/CyclinE.

A figure of the bound structure for one of the compounds (compound 13q) in the CDK2 ATP binding pocket had been published by Yue et al.[19] Although the actual protein CDK2 structure coordinates bound to compound 13q are not publicly available, we can infer binding information from the figure and information provided from the publication and apply them toward other publicly available CDK2 structures. Hence we chose compound 13q to further our structural binding analysis and compare it against other CDK2 structures. For the protein, we have selected two CDK2 structures from the PDB with good resolution, 1FIN and 1QMZ with 2.3 Å and 2.2 Å resolution, respectively. To gain structural insight into ligand-protein binding from ligand-derived pharmacophores hypothesis models, we evaluated the different possible mappings of compound 13q to the selected HypoGen and HypoGenRefine pharmacophore hypothesis.

Compound 13q is able to map in two different orientations to the CDK2 HypoGen hypothesis models (shown in Figure 2). Examples of the two different mappings are represented by the pharmacophore models generated numbered as 12M1C and 20M1C. However, a clear preference for the mapping shown in 12M1C is noticed which occurs 85% of the time. With the CDK2 HypoGenRefine hypothesis, compound 13q also showed two different mappings as shown by the model numbers 5M1C and 11M1C (Figure 1). A clear preference is for the mapping shown in 5M1C, which occurs 75% of the time. In this (5M1C) mapped orientation, the excluded volume features are located behind the compound 13q. This HypoGenRefine model 5M1C mapping is also consistent with the more prevalent HypoGen 12M1C mapped orientation where the carbonyl group in the indenopyrazole core is able to form better intramolecular hydrogen bond interactions to the exocyclic amide. Attempts to overlay the HypoGen and HypoGenRefine hypothesis further indicate a preference for the excluded volume features to be placed

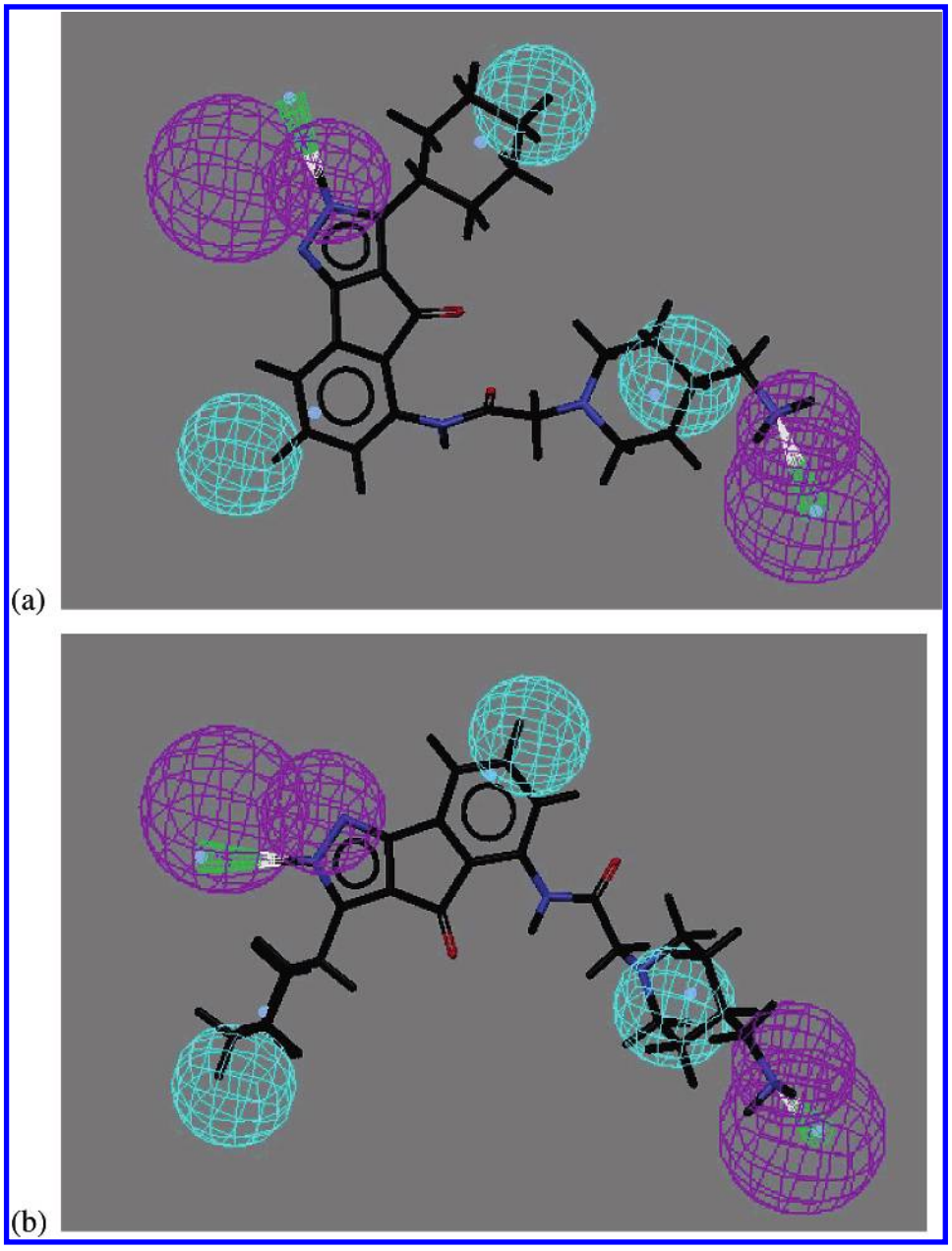


Figure 2. Two mappings of CDK2 compound 13q against the HypoGen model. a: Model number 12M1C − indenopyrazole-scaffold mapped ~85% of all mappings. b: Model number 20M1C − indenopyrazole-scaffold mapped inverse position ~15% of all mappings. The features are hydrophobic (cyan) and hydrogen bond donor (magenta).

behind the molecules. This overlay yields the lowest RMSD of the hypothesis features with a value of 0.15 Å (Figure 1). Taking the mapped orientation of the pharmacophore with excluded volumes as shown in 5M1C, and manually placing it within the ATP active site of CDK2 for 1FIN and 1QMZ, it is immediately apparent that the excluded volume is positioned on top of the active site residues with low flexibility (see Table 4). With the exception of EV1 (excluded volume number 1) which is positioned on Glu-12, a solvent exposed residue near the N-terminus, EV2 to EV4 are all located on residues with significantly low-temperature factors (B-factor) when compared to the protein average and range for the B-factor. This identification of protein active site 'disallowed region' through the use of ligand derived pharmacophores with excluded volume can greatly help scientists obtain insight into target active sites when the protein or target may be unknown.

Table 4: EV Positioned at Regions of Low Mobility[a]

	excluded volume (EV) positions	B-factor 1FIN	B-factor 1QMZ	notes
EV1	β-sheet backbone (CA@Glu12)	80	70	residue is solvent exposed and near N-terminus
EV2	β-sheet backbone (NH@Asp145)	35	26	
EV3	β-sheet backbone (CG@Leu134)	30	28	
EV4	β-sheet backbone (NH@Leu83)	35	28	

[a] 1FIN: resolution 2.3 Å, average B-factor 49 with range 17−80, 1QMZ: resolution 2.2 Å, average B-factor 36 with range 18−87.

In the DHFR system, we selected the HypoGen and the HypoGenRefine models based on the training correlation value and then further subjected these models to Fisher randomization. The HypoGen and HypoGenRefine models

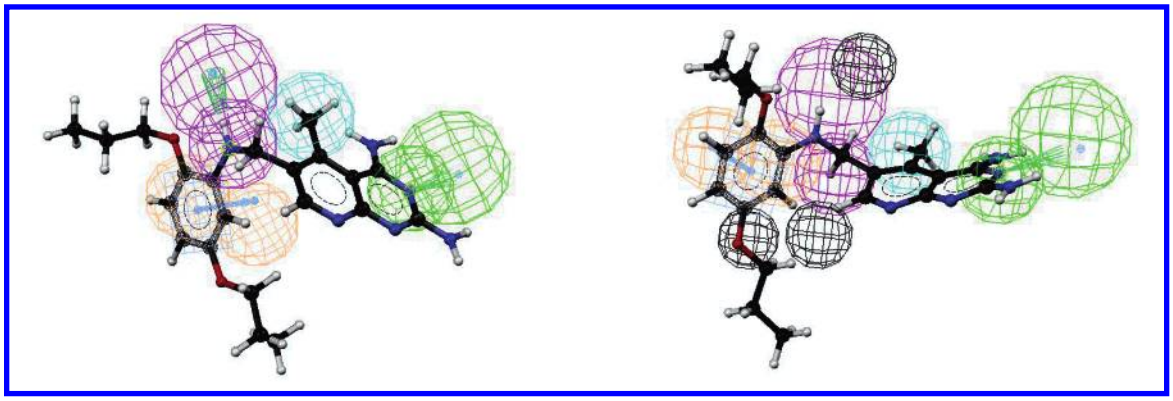


Figure 3. DHFR78 ligand mapped to all of its features in the Hypogen model (left) and the HypoRefine model (right). The features are hydrophobic (cyan), aromatic ring (orange), hydrogen bond acceptor (green), hydrogen bond donor (magenta), and excluded volume (gray).

Table 5: Compounds from the Derwent WDI (World Drug Index, 1999 Release) Compound Collection that Were Predicted To Be DHFR Inhibitors Based on the Pharmacophore Query Generated by Hypogen and HypoRefine Methods[a]

	total	known actives	% yield	enrichment
Derwent WDI 1999	48405	43	0.09%	1X
HypoGen	8246	15	0.18%	2X
HypoRefine (EV=3)	4594	14	0.30%	3X

[a] EV = excluded volume.

selected have a training set correlation $R = 0.95$ and $R = 0.97$, respectively. The HypoGenRefine model generated has three excluded volumes. The inactive compounds (such as DHFR78) that were able to map to the HypoGen model are no longer able to map to the HypoGenRefine models due to the inclusion of steric effects as represented by the excluded volumes in the HypoGenRefine hypothesis. A graphical representation of the top pharmacophore model from both HypoGen and HypoGenRefine are illustrated in Figure 3.

The robust nature of both HypoGen and HypoGenRefine hDHFR pharmacophore models was put to the test by using them to screen the WDI version 2003 (Derwent compound collection)[23] for DHFR compounds. In this database, we have searched using the 1D text string 'DIHYDROFOLATE REDUCTASE' and considered all 43 hits retrieved from the 1D text search as potential hDHFR inhibitors. An important point to note here is that the 43 hits retrieved from the WDI database encompass inhibitors to potentially different isoforms of DHFR as well as different species of DHFR, while the authors would have liked to reduce this subset to those relevant to only the human DHFR protein. Right away, results presented in Table 5 suggest that the number of active molecules retrieved by using either of the HypoGen or HypoGenRefine pharmacophore models is nearly identical. In addition, the use of excluded volume features with pharmacophores does not significantly impact the number of active molecules, also referred to as true positives, retrieved. Looking at the results in this light, the benefits of using excluded volumes with pharmacophore models in virtual screening is not immediately apparent when compared to the actives retrieved from the simple pharmacophore model.

Addressing the total number of molecules predicted as active molecules by the two models highlights the impact of the combination of excluded volumes to the simple pharmacophore model. Thus while the HypoGen model (the simple pharmacophore model) retrieved a total of 8246 hits, the HypoGenRefine model (with the excluded volumes added to the pharmacophore) retrieved a total of 4594 hits. The ability of HypoGenRefine models to reduce the number of false positives (or hits incorrectly identified as DHFR active compounds) significantly helped increase the enrichment factor to 3X when compared to a random search. These results suggest that the placement of excluded volume features on the HypoGenRefine pharmacophore models can serve to reduce false positives and provide a higher enrichment rate in virtual screening. Finally the authors wish to point out that the combination of excluded volumes to the description of the pharmacophore does not slow database search times providing an added incentive to incorporate the extra description to the pharmacophore models. In this study, the search times for both the HypoGen and HypoGenRefine models with the WDI database remains nearly identical (<3 min). We speculate that the percent yield obtained in our Catalyst search may potentially be much higher had we been able to differentiate and consider just the known hDHFR inhibitors from the WDI database.

While combination of excluded volumes to pharmacophore models is not a new concept introduced in this study, we wish to highlight that the ***automated*** excluded volume generation through ligand steric information within Catalyst is what is introduced here as the novel and powerful aspect of the pharmacophore generation algorithm. In addition to gaining insight into structural binding and active site without the knowledge of a protein target structure or an active site, this method presents an unbiased, less cumbersome and thorough approach to testing the placement of excluded volumes around the different pharmacophores and quantitatively assessing their statistical relevance. To investigate the relevance and meaning of the excluded volume generated from the hDHFR ligands with respect to the receptor structure; we have further performed a structural analysis of the pharmacophore with excluded volume in the active site of DHFR. A crystal structure of human DHFR bound to methotrexate and a cofactor NADPH is available from Protein Data Bank (pdb id: 1DLS, resolution: 2.3 Å). We have taken the structure of 1DLS and mapped the bound methotrexate against the Catalyst HypoGenRefine pharmacophore with an excluded volumes model using the rigid

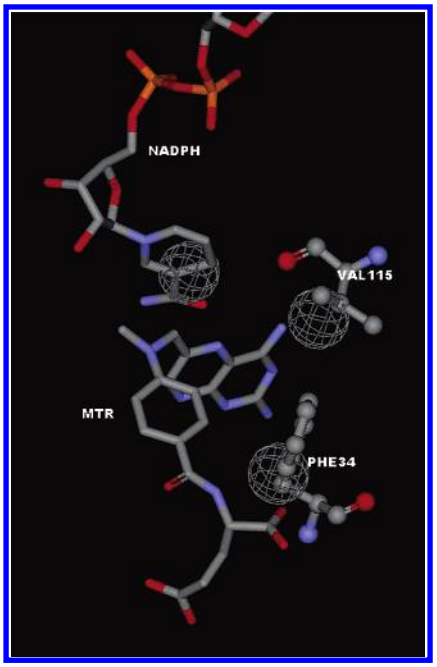


Figure 4. Excluded volume features (gray) are mapped in the disallowed region of the protein. The excluded volumes are shown with methotrexate, NADPH, and the protein residues PHE-34 and Val-115 from the hDHFR PDB structure: 1DLS.

mapping option within Catalyst. The mapped pharmacophore with excluded volume is then visualized with regards to the placement of excluded volume in the protein active site.

The three excluded volume features placements within the active site of 1DLS generated valuable insight into steric interferences encountered by the inactive ligands. All three excluded volume features were mapped to the 'disallowed' regions of the protein active site, as represented in Figure 4. Excluded volume feature number one (EV1) was mapped to protein residue Val-115, which is part of B-Sheet-5 of the protein. EV2 was mapped to Phe-34 of Helix-1. EV3 was mapped to one end of the protein cofactor NADPH, exposed to the active site cavity. The temperature factor (B-factor) for Val-115 and Phe-34 is 20 for both, which is well below the protein average temperature factor range of 34 and the min-max range of 16−81. The low-temperature factor residues of Val-115 and Phe-34 indicate the low mobility of these residues. The favored placement of excluded volume features on low mobility residues, coupled with the placement of excluded volume on the cofactor NADPH validated the Catalyst HypoGenRefine algorithm's ability to use inactive ligand information to gain valuable insight into 'disallowed' regions within the protein active site.

CONCLUSIONS

We have provided studies using ligand-based information alone to gain structural insights into pharmacophore features responsible for the active site binding and identifying regions of 'disallowed' binding for two different proteins: CDK2[24] and hDHFR. The automated excluded volume detection approach was able to correctly identify excluded volume features placement within the active site of a protein on residues with the lowest temperature factor. Understanding the location of protein residues within the active site from ligand information alone will help scientists to better design ligands that avoid the likelihood of having a steric clash with active site residues. This is especially useful when the protein structure is unknown, as is often the case for lead finding or lead optimization of antagonists against membrane proteins. The use of automated excluded volume generation in pharmacophore model building can help provide researchers valuable insight in rational drug design, get better model selectivity, reduce false positives, and increase enrichment rates in virtual screening without much sacrifice with respect to speed and ease of use.

ACKNOWLEDGMENT

We thank the following individuals for their scientific contributions and critical comments: Konstantin Poptodorov, Remy Hoffman, Tien Luu, Katalin Nadassy, Shikha Varma, and Marvin Waldman.

REFERENCES AND NOTES

(1) Daeyaert, F.; de Jonge, M.; Heeres, J.; Koymans, L.; Lewi, P.; Vinkers, M. H.; Janssen, P. A. A pharmacophore docking algorithm and its application to the cross-docking of 18 HIV-NNRTI's in their binding pockets. *Proteins* **2004**, *54*, 526−533.
(2) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: pharmacophore-based protein−ligand docking. *J. Med. Chem.* **2004**, *47*, 6804−6811.
(3) Hindle, S. A.; Rarey, M.; Buning, C.; Lengauer, T. Flexible docking under pharmacophore type constraints. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 129−149.
(4) Joseph-McCarthy, D.; Thomas IV, B. E.; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins* **2003**, *51*, 172−188.
(5) Dannhardt, G.; Laufer, S. Structural approaches to explain the selectivity of COX-2 inhibitors: is there a common pharmacophore? *Curr. Med. Chem.* **2000**, *7*, 1101−1112.
(6) Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr. Med. Chem.* **2004**, *11*, 71−90.
(7) Guner, O. F. Pharmacophore perception, development, and use in drug design. UIL: La Jolla, 2000.
(8) Guner, O. F. History and evolution of the pharmacophore conception in computer-aided drug design. *Curr. Top. Med. Chem.* **2002**, *2*, 1321−1332.
(9) Guner, O. F. The impact of pharmacophore modeling in drug design. *IDrugs* **2005**, *8*, 567−572.
(10) Marceau, F. A possible common pharmacophore in the non-peptide antagonists of the bradykinin B1 receptor. *Trends Pharmacol. Sci.* **2005**, *26*, 116−118.
(11) Malawska, B.; Scatturin, A. Application of pharmacophore models for the design and synthesis of new anticonvulsant drugs. *Mini Rev. Med. Chem.* **2003**, *3*, 341−348.
(12) Kurogi, Y.; Guner, O. F. Pharmacophore Modeling and three-dimensional database searching for drug design using Catalyst. *Curr. Med. Chem.* **2001**, *8*, 1035−1055.
(13) Palomer, A.; Cabre, F.; Pascual, J.; Campos, J.; Trujillo, M. A.; Entrena, A.; Gallo, M. A.; Garcia, L.; Mauleon, D.; Espinosa, A. Identification of novel cyclooxygenase-2 selective inhibitors using pharmacophore models. *J. Med. Chem.* **2002**, *45*, 1402−11.
(14) Norinder, U. Refinement of Catalyst hypotheses using simplex optimisation. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 545−57.
(15) Greenidge, P. A.; Carlsson, B.; Bladh, L.-G.; Gillner, M. Pharmacophores incorporating numerous excluded volumes defined by X-ray crystallographic structure in three-dimensional database searching: application to the thyroid hormone receptor. *J. Med. Chem.* **1998**, *41*, 2503−12.
(16) *Catalyst*; Accelrys, Inc.: San Diego, CA 92102, 2004. Pharmacophore models generated for the hDHFR system used an earlier version of the Catalyst software (4.9), whereas the models for CDK2 were generated using the recent 4.10 version. Since the default parameters were used in both systems, the authors feel that the hDHFR models should be the same if they were regenerated with the newer version of the software. The hDHFR models were developed just before the recent version of the software was released, and hence the difference in version between the two biological cases is addressed in this study.

(17) Maynard, A. J. et al. HypoGenRefine and HipHopRefine: Pharmacophore Refienement using Steric Information from Inactive Compounds. HypoGenRefine Algorithm. Presented at the ACS National Meeting Spring 2004. Manuscript in preparation.
(18) Debnath, A. K. Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of Mycobacterium avium complex dihydrofolate reductase. *J. Med. Chem.* **2002**, *45*, 41−53.
(19) Yue, E. W.; Higley, C. A.; DiMeo, S. V.; Carini, D. J.; Nugiel, D. A.; Benware, C.; Benfield, P. A.; Burton, C. R.; Cox, S.; Grafstrom, R. H.; Sharp, D. M.; Sisk, L. M.; Boylan, J. F.; Muckelbauer, J. K.; Smallwood, A. M.; Chen, H.; Chang, C. H.; Seitz, S. P.; Trainor, G. L. Synthesis and evaluation of indenopyrazoles as cyclin-dependent kinase inhibitors. 3. Structure activity relationships at C3(1,2). *J. Med. Chem.* **2002**, *45*, 5233−5248.
(20) Nugiel. D. A.; Etzkorn, A.-M.; Vidwans, A.; Benfield, P. A.; Boisclair, M.; Burton, C. R.; Cox, S.; Czerniak, P. M.; Doleniak, D.; Seitz, S. P. Indenopyrazoles as novel cyclin dependent kinase (CDK) inhibitors. *J. Med. Chem.* **2001**, *44*, 1334−1336.
(21) Kirchmair, J.; Laggner, C.; Wolber, G.; and Langer, T. Comparative Analysis of Protein-Bound Ligand Conformations with Respect to Catalyst's Conformational Space Subsampling Algorithms. *J. Chem. Inf. Model.* **2005**, *45*, 422.
(22) Fischer, R. *The Design of Experiments;* Hafner Publishing: New York, 1966; Chapter 2.
(23) Derwent WDI, http://scientific.thomson.com/products/wdi/. 2003, Thomson.
(24) Although we have taken precautions (external test set/Fischer randomization test) to ensure that the pharmacophore model is statistically significant, we recognize that the training set selection used in the CDK2 example could lead to some bias in the test set. With the appropriate data a further study could be performed to confirm the ability of the pharmacophore model to adequately distinguish between active and inactive compounds outside of the initial data set.

CI050410C