Combinatorial Ligand Design Targeted at Protein Families

Nikolay P. Todorov,*,† Christoph L. Buenemann,‡ and Ian L. Alberts†

De Novo Pharmaceuticals Ltd., Vision Park, Histon, Cambridge CB4 9ZR, U.K., and Novartis Forschungsinstitut GmbH, GDC, Brunner Strasse 59, A-1235 Vienna, Austria

Received October 12, 2004

We describe a method to create ligands specific for a given protein family. The method is applied to generate ligand candidates for the cyclin-dependent kinase (CDK) family. The CDK family of proteins is involved in regulating the cell cycle by alternately activating and deactivating the cell's progression through the cycle. CDKs are activated by association with cyclin and are inhibited by complexation with small molecules. X-ray crystal stuctures are available for three of the thirteen known CDK family members: CDK2, CDK5 and CDK 6. In this work, we use novel computational approaches to design ligand candidates that are potentially inhibitory across the three CDK family members as well as more specific molecules which can potentially inhibit one or any combination of two of the three CDK family members. We define a new scoring term, SpecScore, to quantify the potential inhibitory power of the generated structures. According to a search of the World Drug Alerts, the highest scoring SpecScore molecule that is specific for the three CDK family members shows very similar chemical characteristics and functional groups to numerous molecules known to deactivate several members of the CDK family.

INTRODUCTION

Computational drug design has drawn the interest of scientists from different fields with its ability to address diverse research problems. A variety of approaches has been proposed to facilitate the development of novel drug candidates.[1−3] In general, these methods involve discerning ligands that exhibit complementarity with the binding site of the target protein. The two major strategies for discovering these ligands involve high throughput screening, in which molecules from a database of organic compounds are docked into the binding site and the fit evaluated, and *de novo* design, in which candidate molecules are generated with potentially high binding affinties. The latter method is not limited to structures within a database and, thus, explores more of the available chemical space. Both approaches depend crucially on the efficient determination of reliable ligand binding affinities. Significant advances have been made in the area of ligand binding affinity prediction by the development of free energy perturbation methodology[4] as well as faster and reliable knowledge-based approaches able to handle a large number of ligand candidates from both high-throughput screening and de novo design approaches.[5] Drug design studies are normally limited to one protein of interest and problems of specificity and cross interactions are left for consideration at a later stage during ligand optimization. There are important advantages in taking account of these issues as early as possible in order to develop ligands binding either to all members of a protein family or specifically to selected members of the family.[6−8]

Positions in the active site that are specific to each protein can be identified by analyzing the probe-interaction maps by principal component analysis.[9,10] Pairwise selectivity has been defined by using selectivity fields as the difference between the biological activities of a ligand toward related proteins.[11] Detailed analysis of electrostatic specificity has been performed to find ways to improve ligand selectivity.[12]

Methods have been proposed to relax the assumption that a ligand would retain its exact binding mode with respect to the aligned proteins.[6,13] These methods allow for adjustments of ligand position and conformation in order to adapt to the related but different protein environments. Multiple compound libraries have been docked against multiple targets.[6] The generated receptor affinity fingerprints have been used for specific ligand design and protein family identification.[13] Multiple active site corrections are also useful for improving the ranking of ligands toward a single target by improving their specificity.[14]

To our knowledge, *de novo* ligand design methodology has not been used to analyze the issue of ligand specificity. In this work we present such a method to address this problem and apply it to the CDK family. Cyclin-dependent kinases are a family of Ser/Thr protein kinases involved in the regulation of the cell cycle by activating and deactivating the cell's progression through the cycle.[15−17] CDKs are inactive in the monomeric form; they are activated by the binding of cyclins and inactivated by CDK small molecule inhibitors. CDK dysregulation is frequently associated with cancer.

Thirteen members of the CDK family are known from examination of the human genome; however, only eleven have been found to exist in complex forms. This is crucial as CDKs are only active when in combination with a regulatory component. There is high sequence similarity among the members of the CDK family, which implies that their structures also show similar features. This has been borne out by the 3-dimensional structures derived from X-ray

* Corresponding author phone: +44 1223 238858; fax: +44 1223 238888; e-mail: nikolay.todorov@denovopharma.com.
† De Novo Pharmaceuticals Ltd.
‡ Novartis Forschungsinstitut GmbH.

10.1021/ci049692r CCC: $30.25
Published on Web 03/08/2005

crystallography that are currently available: unbound native CDK2, CDK2 complexes with cyclin and its derivatives as well as with small molecules, CDK5 with p25 and CDK6 with p25 and p19. Furthermore, model structures of CDK1, CDK3 and CDK7, built from their homology with CDK2, are available. These structures show distinct similarities; the general shape is reproduced, β-sheets are the main constituent of the small N-terminal lobe, the large C-terminus consists mainly of α-helices and the ATP-binding domain is found between the termini. Association of CDK with cyclin perturbs the structure into the active conformation.

The functionality of the eleven CDKs known to exist in complexes (denoted CDK1−CDK11[15]) is briefly summarized. CDK1 and CDK2 are involved with various phases of the cell division cycle; CDK7, CDK8, CDK9 and CDK11 play a crucial role in RNA transcription regulation; CDK5 has several neuronal functions; CDK2, CDK5, CDK6 and CDK9 are involved in differentiation processes; CDK1, CDK5 and CDK6 participate in cell death and CDK5 and CDK11 play roles in cell apoptosis. The CDK family also has other functions, including involvement in the regulation of retinal phosphodiesterase, insulin exocytosis and Golgi membrane traffic. More details can be found in ref 15 and references therein.

There are a number of crystal structures of proteins from the CDK family deposited in the Protein Data Bank, which make it possible to apply the methodology presented in this paper. In this work, we only utilize X-ray structures and so limit our focus to CDK2, CDK5 and CDK6. The main objective of this work is to design ligand candidates that are generic across the three CDK family members, but we will also be interested in finding ligands that are specific across any combination of one or two of the three chosen CDK family members.

METHODS

The general strategy is to first obtain ligands for each member of the family and then check if the ligands achieve good interactions with the other members. If a ligand can fit all members, it can be tested as a general binder for the whole family. In the other extreme, some ligands will only bind with good affinity to one member of the family but not to the other members. There could also be intermediate cases of ligands that are predicted to bind well to some of the members of the family but not to others. These could also be interesting as subfamily targeting agents.

Protein Superposition. All proteins in the family are aligned in the same reference frame, so that the binding sites occupy approximately the same region of space. Once ligands are created for one member, they will be roughly in the correct position for binding to the other members, subject to minor positional adjustments.

Alternatively, the alignment step can be skipped, in which case the ligands generated would be in very different reference frames for each protein. A more extensive docking effort would then be required to verify if the ligands bind to other members of the family. Therefore the first option of initially superimposing the proteins is preferable before going into ligand generation studies.

There are several methods available to superimpose related protein strucures. The program STAMP from Barton's group, for example, is a fully automated method and usually creates convincing superpositions.[18] In many cases however the position of the active site is conserved among members of a family and a reasonable superposition can be achieved based on the alignment of key active site atoms onto a base protein. The latter approach was adopted in this work.

Ligand Generation. Computational ligand generation is performed in this work by linking together molecular fragments and using evolutionary optimization methods to search for structures satisfying the constraints of the protein active site. The method has been documented in detail in refs 19−21. Novel features in comparison to related methods include the application of ScreenScore[22] to assess the receptor−ligand interactions and a library of fragments to enhance synthetic feasibility of the suggested molecules.

ScreenScore has been derived to maximize enrichment in virtual screening experiments. Enrichment is defined as the percentage of active compounds in the top 1% of a database ranked according to a particular scoring function compared to the percentage of active compounds in the whole database. ScreenScore has been found to be particularly successful in discriminating active from inactive compounds and to reduce the number of false positive and negative results. It combines features from different scoring functions, enhancing its ability to model well both the diffuse, nondirectional hydrophobic interactions as well as the more localized, directional hydrogen-bonding and ionic interactions.

The fragment library is based on retrosynthetic analysis and fragmentation of known synthetically accessible structures.[23] In the so-called RECAP procedure, several types of well-known reactions are defined. A database of ligands is searched for occurrences of the products of these reactions. Every time an example of such a product is found, the ligand is broken down into the corresponding reagents. The process continues until all possible fragmentations are performed. The derived fragments are analyzed, the most frequently occurring ones are selected and further inspected by medicinal chemists until an acceptable set is derived. The set we used in this work contained 1678 fragments. During structure generation, the fragmentation process is reversed and suitable reagents are combined together in novel ways to form ligands that are both synthetically reasonable and fitting into the active site of the protein of interest.

Ligands are generated using the strategy outlined in ref 21. Several pharmacophore points are specified in order to explore the interesting regions of the active site. Pharmacophore points are associated with corresponding atoms in each family member, and their exact positions change slightly from structure to structure. During the structure generation procedure, putative ligands that do not match the pharmacophore points are penalised. In addition to that, we use ScreenScore[22] to enable the generation of strongly binding ligands. A threshold value for the ScreenScore averaged over non-hydrogen atoms in the ligand (termed ScreenScore-per-atom) is also specified to enhance the complementarity of the ligands with the receptor site. Without this complementarity term, larger ligands are favored even if they do not fit as well as smaller ones, due to the fact that the score is accumulated over pairs of receptor and ligand atoms and larger ligands have contributions from more pairs. This is confirmed by a survey of experimental data, which indicates that the free energy of the strongest-binding ligands for a

given protein increases with the number of non-hydrogen atoms with an initial slope of approximately −1.5 kcal/mol.[24]

In this study, we used more than one crystal structure for CDK2 and CDK6 from the PDB (only one X-ray structure is avaliable for CDK5) in order to model protein conformational flexibility. Each structure was used in turn to create sets of ligands. Pharmacophore points are associated with each protein conformation.

Cross-Docking Ligands to the Other Proteins in the Family. Ligands generated for one of the proteins were docked to the other members of the family. Where more than one conformation of the protein was available, the ligand was docked to all conformations, considering them one at a time. It was also possible to consider all protein conformations at the same time, but since the number of available conformations from crystal structures was quite small, we chose to explore them separately as this should yield more accurate results with better sampling.

After the initial superposition of the family members, this task is easier to carry out since the ligand is already in the relevant reference frame. Only minor adjustments of the position, orientation and the conformation of the ligand in the vicinity of the reference pose are required for the ligand to fit in the same binding mode. Docking was performed using simulated annealing methodology as previously implemented,[19−21] and ScreenScore was used to rank the poses of the docked ligands in each active site. We limited the RMSD value from the initial configuration to 3 Å and also applied pharmacophore and ScreenScore-per-atom constraints. These are instrumental in restricting the search to the relevant regions and thus vastly improving efficiency. Each ligand was docked three times to each protein conformation and the best result taken.

Scoring Ligand Specificity. We define a specificity score, SpecScore, which estimates the specificity for each ligand toward one or more proteins in the family. To illustrate how SpecScore works, let us consider an example of a protein family with four members in the family: 1, 2, 3 and 4. Let the free energy of binding for a ligand with these four proteins be ΔG_1, ΔG_2, ΔG_3, ΔG_4, respectively. If we want to find ligands binding to proteins 1 and 2, but not to proteins 3 and 4, SpecScore, in this case, is defined as

$$SpecScore(1+,2+,3-,4-) = \max(\Delta G_1,\Delta G_2) - \min(\Delta G_3,\Delta G_4)$$

The + and − signs indicate whether the ligand should or should not bind to the corresponding protein. Essentially, this score evaluates the free energy gap between the two groups of proteins with respect to ligand binding. The lower the *SpecScore* of the ligand the higher is its specificity for the selected set of proteins. Ligands can then be ordered by SpecScore in terms of their specificity properties. For example, if we want to select ligands binding to one protein only, say protein 1, we should order by *SpecScore*(1+,2−,3−,4−). If ligands binding to all proteins are required, ligands should be ordered by *SpecScore*(1+,2+,3+,4+). Another SpecScore is *SpecScore*(1−,2−,3−,4−) if we look for compounds that do not bind to any of the proteins in the family, but normally this score should not be of interest. In general, if there are n proteins in the family, there are 2^n-1 SpecScores to consider, disregarding *SpecScore*(1−, ..., n−). The values of ScreenScore for known ligands can be used as reference values for what good SpecScores should be for a particular protein family.

If there are several conformations available for a protein, taken from several different crystal structures or from an NMR structure determination study, the binding free energy entering SpecScore is given as a Boltzmann sum over all protein conformations j:

$$\Delta G = -kT \Sigma \exp(-\Delta G_j/kT)$$

In this study, we approximated ΔG by including only the protein conformation with lowest ΔG_j. Free energies of binding are difficult to determine accurately by computational methods, and various empirical schemes have been developed to facilitate this task. Here we use ScreenScore to estimate binding free energies.

To give a numerical example, let us assume that for a data set of four proteins, each of them has two conformations and the free energies of binding for a ligand are as follows: −30.0 and −31.0 for the two conformations of protein 1, −32.0 and −33.0 for protein 2, −20.0 and −21.0 for protein 3 and −22.0 and −23.0 for protein 4. The calculated SpecScore(1+, 2+, 3−, 4−) = max(−31.0, −33.0) − min(−21.0, −23.0) = −31.0 + 23.0 = −8.0.

RESULTS

There are three CDKs for which crystal structures are available in the Protein Databank, namely CDK2, CDK5 and CDK6. Theoretical models are available for CDK1, CDK3 and CDK7, but these are not used in this study. The PDB codes of the structures we used are the following: CDK2-1di8 (complex with a small ligand),[25] 1di8 with a strongly bound water molecule retained, 1fin (complexed with cyclin),[26] 1hcl (uncomplexed),[27] CDK5-1h4l,[28] CDK6-1blx[29] and 1jow.[30] Missing hydrogens in the crystal structures were added and minimized using the CVFF force field and conjugate gradient optimization as implemented in Insight2.[31]

Protein Superposition. In the case of CDKs it is clear that the position of the hinge region (residues 80−86 in the CDK2 and CDK5 proteins and 98−104 for CDK6) is an important determinant of the overall quality of the fit between the members of the family, and a reasonable superposition can be achieved based on the alignment of that region (Figure 1). In this respect, CDK2 protein 1di8 was taken as the base structure, and the other receptors were superposed onto the base by aligning the atoms in the corresponding hinge regions.

The ATP binding site of CDK is flexible and can undergo some conformational changes. The buried portion of the hinge strand (Phe 80 to the backbone NH of Leu 83 in CDK2) does not move significantly. The Leu 83 carbonyl group (in CDK2 and its equivalents in CDK5 and CDK6), however, can either be oriented into the binding site in a position suitable to form hydrogen bonds or perpendicular to the adenine binding site. Here we use all available crystal structures to model protein flexibility. Our data set included both extremes of the Leu 83 backbone carbonyl orientations in CDK2. Furthermore, the two CDK6 structures in our data set showed significant side chain variation, particulary for

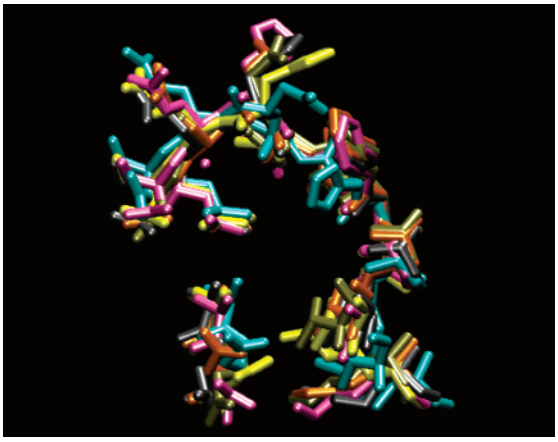

Figure 1. Superposition of the hinge regions of the seven CDK proteins. (Figures 1−4 were created using VMD[33] molecular graphics and Tachyon[34] ray-tracing).

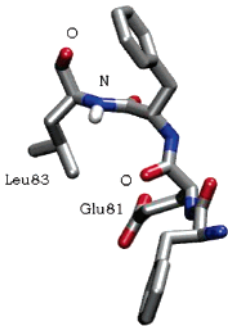

Figure 2. Pharmacophore requirements in the active site for designed ligands. Each ligand generated is required to accept a hydrogen bond from the backbone NH group of Leu 83 for CDK2 and CDK5 (Val 101 for CDK6) and donate a hydrogen bond to either the carbonyl group of Glu 81 or Glu 83 for CDK2 and CDK5 (Glu 99 or Val 101 for CDK6).

His 100 in the hinge region (for CDK6, the hinge region incorporates residues 98−104).

Ligand Generation. The pharmacophore requirements were selected based on previous work on CDK2[21] and displayed in Figure 2. In crystal structures of CDK2-small molecule complexes available in the PDB, the inhibitors bind at the ATP binding site and form one or more hydrogen bonds to the so-called hinge strand connecting the two domains. In all cases, the hydrogen bonding pattern involves donation from a central backbone NH group of residue Leu 83 to the inhibitor. In addition, most of the other known CDK2 inhibitors form a hydrogen bond to at least one of the two neighboring backbone carbonyl groups of residues Glu 81 and Leu 83. Our search constraints for CDK2 as well as CDK5 and CDK6 were constructed from this observed hydrogen bonding pattern. Generated ligands were required to accept a hydrogen from the backbone NH group of Leu 83 for CDK2 and CDK5 (Val 101 for CDK6) and donate a hydrogen to the backbone CO group of either Glu 81 or Leu 83 for CDK5 and CDK6 (Glu 99 or Val 101 for CDK6). An additional constraint was that the atom in the generated ligand structure that is involved in hydrogen bond donation to the active site is different to the one accepting a hydrogen from the receptor, thus, eliminating the possibility that a ligand hydroxyl group is involved in both hydrogen bonds. Interactions of this type have been observed in CDK2-inhibitor X-ray structures; however, we are not aware that any of these inhibitors have led to compounds of preclinical interest. The CDK2-inhibitor binding is also known to be enhanced by hydrophobic interactions between ligand aromatic, heterocyclic ring systems that incorporate the hydrogen bond donor and acceptor groups discussed above and aliphatic peptide side chains from the surrounding active site.

The design of novel kinase inhibitors usually involves identifying suitably substituted heterocycle scaffolds that can satisfy the hydrogen bonding requirements. Our previous work on CDK2 *de novo* design[21] focused on discerning fragments that fulfill specific hydrogen bonding interactions with the hinge region. This approach was particularly successful and produced many interesting scaffolds that could be tested as a first step in a research program aimed at identifying weakly, but specifically binding compounds.

Here, our aim is also to explore the space away from the hinge region. To achieve this, as well as deploying the pharmacophore constraints described above, we set minimum requirements for ScreenScore and the ScreenScore-per-atom values. Satisfying the latter constraints will require structure generation over a greater portion of the active site. We performed five structure generation runs for each protein varying the threshold of the required ScreenScore (from −20 to −40 in steps of 5). Also, the most active compounds for one protein may not be active for another protein in the family and, therefore, less active alternatives for the first protein should be explored. In addition, we set a minimum ScreenScore per atom value of −1.5 and incorporated the pharmacophore constraints. Ten simulations were conducted for each run, and a structure was output in each simulation if the ScreenScore (and ScreenScore-per-atom) and pharmacophoric constraints were satisfied. Each run can, therefore, lead up to a maximum of 10 structures.

A total of 185 ligands were generated: 124 for CDK2, 29 for CDK5 and 32 for CDK6. More ligands are produced for CDK2, because four protein structures were used for design. CDK5 and CDK6 produced approximately equal numbers of ligands although there were two protein structures used for CDK6 subfamily and only one for CDK5. One of the CDK6 protein structures (1blx) did not produce ligands as the ScreenScore requirements were tightened. The reason for this behavior was traced to the conformational changes of residues His100 and Asp163 that close the active site and do not allow for ligands to be grown.

Ligand Cross-Docking. Each ligand generated for each of the CDK proteins was docked to all members of the family. A total of 1295 docking attempts were performed: 185 ligands were docked in all seven proteins. Docking was performed using simulated annealing methodology as previously implemented,[19−21] and ScreenScore was used to rank the poses of the docked ligands in each active site. A maximum RMSD deviation of 3 Å was allowed for the docked pose relative to the initial configuration and pharmacophore and ScreenScore-per-atom constraints were invoked.

Small adjustments are observed in the position, orientation and conformation of the ligands in order for them to fit optimally in the slightly modified binding site of a different or even the same protein they were designed for. Figure 3 shows an example illustrating this point. The structure in Figure 3a is designed in the binding site of CDK2 1di8 (accessible surface representation). Figure 3b shows how it docks in the active site of CDK6 1jow (accessible surface

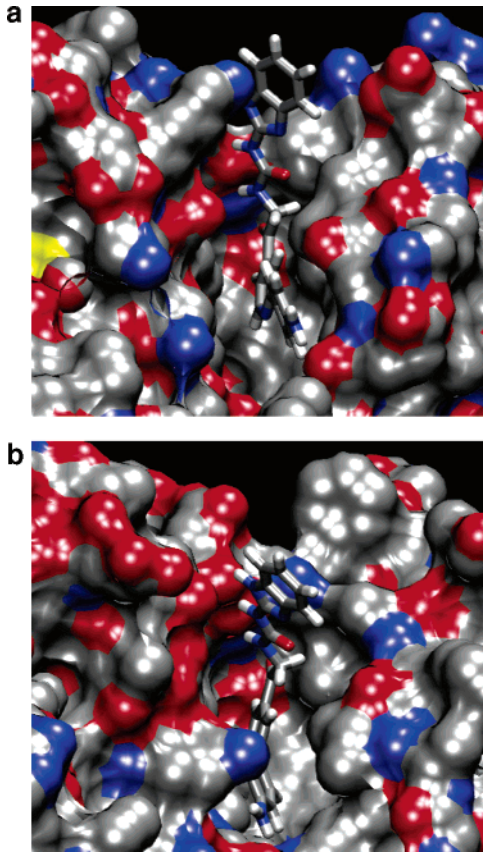

Figure 3. a) A structure generated in the 1di8 site (ScreenScore −40.0). b) The same structure docked in the active site of 1jow (ScreenScore −40.6).

Table 1. Best SpecScores for the CDK Family from 185 Ligands

SpecScore type	Best SpecScore
cdk2+, cdk5+, cdk6+	−40.0
cdk2+, cdk5+, cdk6−	−35.6
cdk2+, cdk5−, cdk6+	−42.9
cdk2−, cdk5+, cdk6+	−4.9
cdk2+, cdk5-, cdk6−	−44.0
cdk2−, cdk5+, cdk6−	−41.4
cdk2−, cdk5−, cdk6+	−35.2

representation). Differences can be noticed in the position, orientation and conformation of the ligand in the two figures. Movements are important to avoid repulsive interactions with the modified active site in CDK6 compared to CDK2.

Similar movements are also observed when the ligand is designed for the same protein but in a different conformation. Even when designed to the same active site in the same protein conformation, a ligand may be able to a achieve better ScreenScore when docked while keeping a similar binding mode by utilizing small displacements.

Specificity Score Analysis. There are seven SpecScores available for a family with three proteins. The best SpecScore for each of the seven classes over the 185 ligands is given in Table 1. ScreenScore threshold values in the simulations ranged from −20 to −40, and, thus, we may expect good SpecScores to be of similar magnitude as borne out by the values in Table 1. It appears that fairly specific ligands can be found for all classes but one (cdk2−, cdk5+, cdk6+).

In this work, we were mainly interested in ligands binding to all CDK proteins and therefore we focused further analysis to that class (cdk2+, cdk5+, cdk6+). Figure 4 shows the

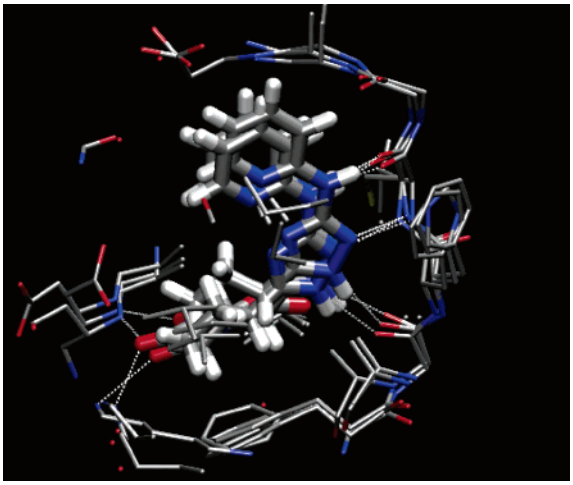

Figure 4. The structure with the best SpecScore (cdk2+,cdk5+, cdk6+) docked in the active sites of cdk2 (1fin) with a ScreenScore of −40.1, cdk5 (1h4l) with a ScreenScore of −40.0 and cdk6 (1jow) with a ScreenScore of −45.4. Hydrogen bonding interactions with the corresponding active are depicted with dotted lines.

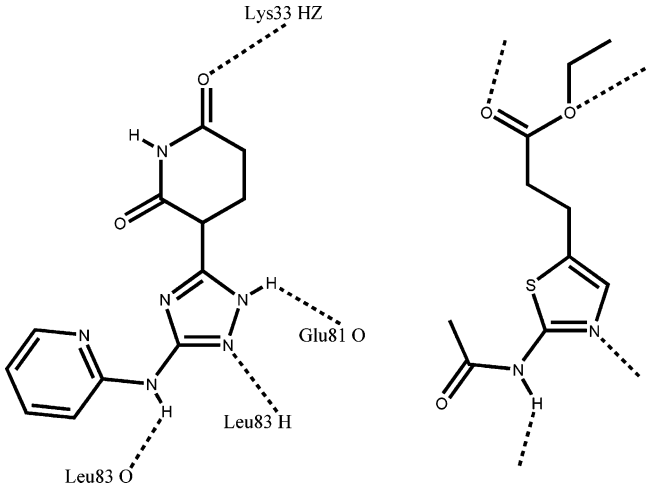

Figure 5. a) 2-Dimensional representation of the highest SpecScore ranked ligand that is specific for CDK2, CDK5 and CDK6, including H-bonding interactions with the corresponding active site. b) A hit from the WDA search (patent number WO9965884A1), including potential H-bonding interactions, that shows similarity to the molecule in Panel a and is known to be inhibitory for CDK1 through CDK8.

most specific ligand (thick lines) for all three CDK proteins as docked in the corresponding active sites (thin lines). It is clear that some adjustments take place in order for the ligand to fit the slightly different requirements of the three sites. The WDA was searched for similar ligands using Daylight tools,[32] and several molecules were identified which are known to have activity across several members of the CDK family. One example is shown in Figure 5. Two-dimensional representations of the above-mentioned highest scoring (cdk2+, cdk5+, cdk6+) SpecScore ligand, including H bonding interactions to the corresponding active site, as well as one of the hits from the WDA search (patent number WO9965884A1) are depicted in Figures 5a and 5b, respectively. The two ligands have potentially similar H-bonding characteristics. Three of the four H-bonding interactions depicted for the top-ranked SpecScore ligand in Figure 5a could also potentially be formed by the WDA ligand (Figure 5b): the heterocyclic ring N (acceptor), exocyclic NH group

(donor) located 2 bonds away from the N acceptor and a carbonyl acceptor located 5 or 6 bonds from the N acceptor. The WDA molecule shown is known to be inhibitory for CDK1 through CDK8. Furthermore, seven of the top ten ranked SpecScore ligands that are specific for the three proteins have a similar chemotype to the top ranked molecule shown in Figure 5a and are also similar to the WDA molecule shown in Figure 5b.

DISCUSSION

We presented a method facilitating the development of novel ligands binding specifically to selected members of a protein family. Normally drug design studies are limited to one protein of interest and problems of specificity and cross interactions are left for consideration at a later stage during ligand optimization. Our method offers important advantages by taking into account these issues early in the design process.

The method consists of three stages. During the first stage, putative ligands are created for each protein in the family using advanced computational structure generation methodology.[19−21] Pharmacophore-point, interaction energy and complementarity constraints are used to direct the process and utilize previous knowledge.

In the second stage, the ligands are docked to each protein in the family. Ligands usually require small adjustments in their position, orientation and conformation in order to fit optimally in the slightly modified binding site of a different protein in the family. We take advantage of preliminary superposition of all proteins to minimize the amount of search during docking. The search starts from the ligand configuration obtained at the structure generation stage, which is already in the same reference frame.

During the third stage, the data from stage two are analyzed. As a measure of specificity, we define a new specificity score, SpecScore, which maximizes the free energy gap between the worst binder in the target set of proteins and the best binder in the non-target set of proteins to a given ligand. SpecScore is calculated for each ligand, for all combinations of proteins, and the most specific ligands are selected as those with the largest energy gaps.

This procedure and the SpecScore ranking are likely to be helpful in directing the search for novel and specific ligands. We have applied the technique to the family of cyclin-dependent kinases—important pharmacological targets involved in the regulation of the cell cycle.

We found that reagents can be identified that are specific for all combinations (except one) of the three CDK family members. Using our methodology, we generated several molecules that are similar to compounds with known activity across several members of the CDK family. This methodology is general and can be applied to a variety of protein families emerging from the Human Genome Project.

REFERENCES AND NOTES

(1) Gadek, T. R.; Nicholas, J. B. Small molecule antagonists of proteins. *Biochem. Pharmacol.* **2003**, *65*, 1−8.

(2) Waszkowycz, B. Structure-based approaches to drug design and virtual screening. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 407−413.

(3) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.

(4) Marrone, T. J.; Briggs, J. M.; McCammon, J. A. Structure-based drug design: Computational advances. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71−90.

(5) Bissantz, C.; Folkers, G.; Roganan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4763.

(6) Lamb, M. L.; Burdick, K. W.; Toba, S.; Young, M. M.; Skillman, A. G.; Zou X.; Arnold, J. R.; Kuntz, I. D. Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* **2001**, *42*, 296−318.

(7) Janin, J. Quantifying biological specificity: the statistical mechanics of molecular recognition. *Proteins* **1996**, *25*, 438−445.

(8) Sims, P. A.; Wong, C. F.; McCammon, J. A. A. A computational model of binding thermodynamics: the design of cyclin-dependent kinase 2 inhibitors. *J. Med. Chem.* **2003**, *46*, 3314−3325.

(9) Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma E. E. J.; Fox, T. GRID/CPCA: A new computational tool to design selective ligands. *J. Med. Chem.* **2000**, *43*, 3033−3044.

(10) Ridderstrom, M.; Zamora, I.; Fjellstrom, O.; Andersson, T. B. Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18 and 2C19 homology models using GRID/CPCA. *J. Med. Chem.* **2001**, *44*, 4072−4081.

(11) Baskin, I. I.; Tikhonova, I. G.; Palyulin, V. A.; Zefirov, N. S. Selectivity fields: comparative molecular field analysis (CoMFA) of the glycine/NMDA and AMPA receptors. *J. Med. Chem.* **2003**, *46*, 4063−4069.

(12) Kangas, E.; Tidor, B. Electrostatic specificity in molecular ligand design. *J. Chem. Phys.* **2000**, *112*, 9120−9131.

(13) Greenbaum, D. C.; Arnold, W. D.; Lu, F.; Hayrapetian, L.; Baruch, A.; Krumrine, J.; Toba, S.; Chehade, K.; Brömme, D.; Kuntz, I. D. Small molecule affinity fingerprinting: a tool for enzyme family subclassification, target identification, and inhibitor design. *Chem., Biol.* **2002**, *9*, 1085−1094.

(14) Vigers, G. P.; Rizzi, J. Multiple active site corrections for docking and virtual screening. *J. Med. Chem.* **2004**, *47*, 80−89.

(15) Knockaert, M.; Greengard, P.; Meijer, L. Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol. Sci.* **2002**, *23*, 417−425.

(16) Morgan, D. O. Principles of CDK regulation. *Nature* **1995**, *374*, 131−134.

(17) Pavletich, N. P. Mechanisms of cyclin-dependent kinase regulation: Structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J. Mol. Biol.* **1999**, *287*, 821−828.

(18) Russell, R. B.; Barton, G. J. Multiple protein-sequence alignment from tertiary structure comparison − asignment of global and residue confidence levels. *Proteins* **1992**, *14*, 309−323.

(19) Todorov, N. P.; Dean, P. M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 175−192.

(20) Todorov, N. P.; Dean, P. M. Computational ligand design by free energy minimization. In *Biological Phisics;* Frauenfelder, H., Hummer, G., Garcia, R., Eds.; American Institute of Physics: 1999.

(21) Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Boehm, H.-J.; Dean, P. M. A validation study on the practical use of automated de novo design. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 459−478.

(22) Stahl, M.; Rarey, M. Detailed analysis of scoring fucntions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042

(23) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP − Retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying priviledged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511−522.

(24) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997−10002.

(25) Shewchuk, L.; Hassell, A.; Wisely, B.; Rocque, W.; Holmes, W.; Veal, J.; Kuyper, L. F. Binding mode of the 4-anilinoquinazoline class of protein kinase inhibitor: X-ray crystallographic studies of 4-anilinoquinazolines bound to cyclin-dependent kinase 2 and p38 kinase. *J. Med. Chem.* **2000**, *43*, 133−138.

(26) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, A.; Hurwitz, J.; Massague, J.; Pavletich, N. P. Mechanism of CDK activation revealed by the structure of a cyclin A- CDK2 complex. *Nature* **1995**, *376*, 313−320.

(27) Schulze-Gahmen, U.; Brandsen, J.; Jones, H. J.; Morgan, D. O.; Meijer, L.; Vesely, J.; Kim, S. H. Multiple modes of ligand recognition − crystal structures of cyclin-dependent-protein-kinase-2 in complex with ATP and 2 inhibitors, olomoucine and isopentenyladenine. *Proteins* **1995**, *22*, 378−391.

(28) Tarricone, C.; Dhavan, R.; Peng, J.; Areces, L. B.; Tsai, L. H.; Musacchio, A. Structure and regulation of the CDK5-p25(nck5a) complex. *Mol. Cell* **2001**, *8*, 657−669.

(29) Brotherton, D. H.; Dhanaraj, V.; Wick, S.; Brizuela, L.; Domaille, P. J.; Volyanik, E.; Xu, X.; Parisini, E.; Smith, B. O.; Archer, S. J.; Serrano, L.; Brenner, S. L.; Blundell, T. L.; Laue, E. D. Crystal structure of the complex of the cyclin D dependent kinase Cdk6 bound to the cell-cycle inhibitor p19(INK4d). *Nature* **1998**, *395*, 244−250.
(30) Schulze-Gahmen, U.; Kim, S. H. Structural basis for CDK6 activation by a virus-encoded cyclin. *Nature Struct. Biol*. **2002**, *9*, 177−181.
(31) Insight II is aproduct of Accelrys Inc.
(32) Daylight Chemical Information System Inc., http://www.daylight.com/
(33) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph*. **1996**, *14*, 33−38.
(34) http://jedi.ks.uiuc.edu/~johns/raytracer

CI049692R