A Computational Model of Binding Thermodynamics: The Design of Cyclin-dependent Kinase 2 Inhibitors

Peter A. Sims,[†] Chung F. Wong,*[,‡,§] and J. Andrew McCammon[†,‡,§]

Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, and Department of Pharmacology, University of California, San Diego, La Jolla, California 92093-0365

Received November 6, 2002

The cyclin-dependent protein kinases are important targets in drug discovery because of their role in cell cycle regulation. In this computational study, we have applied a continuum solvent model to study the interactions between cyclin-dependent kinase 2 (CDK2) and analogues of the clinically tested anticancer agent flavopiridol. The continuum solvent model uses Coulomb's law to account for direct electrostatic interactions, solves the Poisson equation to obtain the electrostatic contributions to solvation energy, and calculates scaled solvent-accessible surface area to account for hydrophobic interactions. The computed free energy of binding gauges the strength of protein−ligand interactions. Our model was first validated through a study on the binding of a number of flavopiridol derivatives to CDK2, and its ability to identify potent inhibitors was observed. The model was then used to aid in the design of novel CDK2 inhibitors with the aid of a computational sensitivity analysis. Some of these hypothetical structures could be significantly more potent than the lead compound flavopiridol. We applied two approaches to gain insights into designing selective inhibitors. One relied on the comparative analysis of the binding pocket for several hundred protein kinases to identify the parts of a lead compound whose modifications might lead to selective compounds. The other was based on building and using homology models for energy calculations. The homology models appear to be able to classify ligand potency into groups but cannot yet give reliable quantitative results.

Introduction

The role of cyclin-dependent kinases (CDKs) in cell cycle regulation has made them interesting targets for antiproliferative drug design. These protein kinases are generally categorized into G1, S, and G2 phase regulators because they are present at various checkpoints in the cell cycle.[1] As their name suggests, the CDKs are dependent on larger proteins known as cyclins for activation. Only as a complex can these proteins regulate cell growth and DNA synthesis properly. The primary CDK addressed in this project is CDK2, which combines with cyclin E at an S-phase checkpoint known as the restriction point.[1] Similarly, the completion of the S phase is subject to a complex of CDK2 and cyclin A.[1] Among other CDKs explored in this project are CDK1 (often called CDC2) and CDK4, which have important roles in G2 mitosis and G1 regulation, respectively.[1]

The cell manufactures specialized inhibitors of the CDKs known as CKIs or cyclin-dependent kinase inhibitors. These inhibitors compete with ATP for binding to the CDK active site. However, in some cancer cells it has been shown that the CKIs are underexpressed, and medicinal chemists have made numerous efforts to replace the CKIs with synthetic inhibitors.[2] Among noteworthy attempts at creating such inhibitors are a series of compounds known as flavonoids. The scaffolds of these compounds are natural products often found in plants that are actually consumed in the human diet.[3] Flavopiridol has been the most successful flavonoid as an anticancer agent in its inhibition of multiple CDKs such as CDK1, CDK2, CDK4, and CDK9.[1,4] Further experimentation with flavopiridol in clinical trials has also taken place.[1] De Azevedo et al. have determined the crystal structure of the deschloro analogue of flavopiridol bound to CDK2, which has been used in many studies of CDK inhibition, including this project.[5]

Recent computational studies of protein kinase A (PKA) by Wong et al.[6] and Gould and Wong[7] have provided some evaluation of the methods used in this project. In these studies, computational chemists tried to identify the determinants of protein−ligand recognition by carrying out computational sensitivity analysis, which is analogous to genetic or chemical modification experiments.[8] By perturbing different features of a drug lead, one can understand which features affect binding significantly and which do not. Such understanding can generate useful rules to guide the optimization of drug leads by pointing to the profitable characteristics of a compound along with parts that should be modified to improve the odds of finding better inhibitors. On the basis of the most significant features identified, one can also construct pharmacophore models for mining new drug leads with different chemical scaffolds from small-molecule libraries. Although sensitivity analysis can be carried out using more elaborate molecular dynamics simulation models in which receptor and ligand flexibility can be explicitly taken into account,[8] such calculations are expensive to use. It is therefore useful to explore less expensive models to supplement the more costly molecular dynamics simulation models. The less expensive models can be used for earlier exploration of

* To whom correspondence should be addressed. Phone: 858-534-2971. Fax: 858-534-7042. E-mail: c4wong@ucsd.edu.
[†] Department of Chemistry and Biochemistry.
[‡] Howard Hughes Medical Institute.
[§] Department of Pharmacology.

10.1021/jm0205043 CCC: $25.00
Published on Web 06/12/2003

a larger number of structures. Promising structures can then be further evaluated with more accurate but expensive models. The work by Wong et al.[6] and Gould and Wong[7] focused on exploring the use of an implicit-solvent model within the fixed-conformation approximation. This implicit-solvent model uses Coulomb's law to account for direct electrostatic interactions, solves the Poisson equation to describe the electrostatic contribution to solvation, and evaluates scaled solvent-accessible surface area to describe hydrophobic effects. Within the fixed-conformation approximation, each binding energy calculation can be done within tens of minutes for protein kinase systems by using computers such as SGI Octane2s with R12000 processors. Such efficiency is compatible with day-to-day use in computer-assisted drug discovery. In earlier applications of this model, the interactions between protein kinase A and both the balanol class of compounds[6] and the inhibitor PKI[7] were investigated. These studies demonstrated that this type of model could generate semiquantitative agreement with experiments and could be improved in a semiempirical manner by introducing several adjustable parameters to fit available experimental data. This paper further evaluates such an approach by applying it to a study of cyclin-dependent protein kinases. Here, we first evaluate whether such models can provide reasonable agreement with existing experimental data. If so, they can be used to make predictions that can be experimentally validated.

By addition of sequence information for hundreds of protein kinases and structural information obtained from X-ray crystallography, past studies also accounted for specificity in designing protein kinase inhibitors.[6,7] They tried to identify regions of the ligand-binding pocket with more variable amino acid distribution as potential targets for achieving specificity. This analysis takes little computer time to do but lacks detailed energetic information. In principle, one can build homology models for many protein kinases and use them for energy calculations. However, building homology models good enough for energy calculations is still a challenge. Here, we explore a more modest approach of building hybrid models for energy calculations. In these hydrid models, we start with the crystal structure of a protein kinase and only mutate the residues closest to the ligand-binding site into those of another kinase whose structure has not yet been determined. The longer-range effects are thus only dealt with in an approximate manner. This way, we can pay special attention to refining a smaller number of residues that are nearest the binding site and therefore likely to dominate the contribution to binding energy. We explore this method of building homology models and gauge their reliability in studying the selective binding of different inhibitors to CDK1, CDK2, and CDK4.

Methods

Binding Free Energy Model. Instead of using the CHARMM22 united-atom force field as in our earlier work with the interactions between protein kinase A and balanol or the protein kinase inhibitor (PKI),[6,7] we employed the CHARMM all-atom force field.[9] This force field provides atomic partial charges and van der Waals radii for the protein atoms. The atomic charges for all of the ligands (des-chloroflavopiridol, the flavopiridol analogues, and all designed structures) were obtained by using the Merz−Kollman method[10,11] in Gaussian 98[12] with the 6-31G* basis set.

We applied UHBD[13,14] for the actual binding free energy calculations after providing it with the CHARMM parameters.[9] The crystal structure of des-chloroflavopiridol L868276 bound to CDK2 provided the starting points for this model.[5] Hydrogens for the protein were added using CHARMM,[15] and those for L868276 were generated using Quanta.[16] Some parts of the protein were not seen in the crystal structure and were not included in our calculations. Thus, our model consisted of three noncontiguous segments: residues 1−35, residues 48−149, and residues 165−298. For the finite-difference solution of the Poisson equation, the complex, the protein, and the ligand sat in a 175 Å × 250 Å × 215 Å grid with 0.3 Å grid spacing. Since the use of an all-atom force field introduces many more atoms than a united-atom force field, we used a nonbonded cutoff of 20 Å to ease the calculation of Coulombic contributions to the electrostatic energy. Previously, no cutoff distance was used.[6,7] The internal dielectric of the solute was set at 2 and the external (solvent) dielectric was set at 78 while the van der Waals surface was used to define the dielectric boundary.

The energy model was composed of three terms. One was the Coulombic energy (ΔG_{COUL}) resulting from the electrostatic interactions among the atomic charges in a dielectric medium characterized by the internal dielectric constant. This was simply calculated by using Coulomb's law with a 20 Å cutoff. The solvation energy, or reaction field term (ΔG_{PB}), was obtained by solving the Poisson equation using the finite-difference method as implemented in UHBD.[13,14] The third term (ΔG_{SA}) described hydrophobic interactions, which were assumed to be proportional to the solvent-accessible surface area of the system with a proportionality constant of 25 cal/(mol·Å^2).[17] By use of this energy model, the free energy was calculated for each system (the protein, ligand, and binary complex) as

$$G_i = G_{i,COUL} + G_{i,PB} + G_{i,SA} \quad (1)$$

so that the binding free energy can be estimated as

$$\Delta G_{BIND} = G_{BINARY} - (G_{PROTEIN} + G_{LIGAND}) \quad (2)$$

Sensitivity Analysis. We dissect the determinants of molecular recognition by carrying out sensitivity analysis. This involves perturbing different parameters of an energy model to reveal their effects on binding affinity.[6−8] Perturbing important parameters can significantly affect binding, whereas perturbing unimportant parameters should affect binding negligibly. In this work, we focus on examining the role of electrostatics in binding by concentrating on the effects of perturbing atomic partial charges on molecular recognition. We describe sensitivity analysis in the way that has been used by engineers for many years,[18,19] but this concept has also been extended to molecular systems.[8] Engineers calculate derivatives $\partial O/\partial \lambda_i$ of a property O with respect to different parameters λ_i of a model to probe for the most significant λ_i values. In this case, the property we focus on is the free energy of binding. The

$\partial O/\partial\lambda_i$ derivatives measure the capacity of the model parameters to alter the system property O. They are therefore useful for pointing out the parts of a lead compound where parameter modifications may be profitable for improving binding. To identify the parts of a lead compound that are already useful or not useful in determining O, it is useful to calculate $(\partial O/\partial\lambda_i)\lambda_i$ values so that the magnitude of the parameters in the lead compound is also taken into account. Each such quantity measures the effects of turning on a parameter, as can be seen from the approximate expression

$$\Delta O(0\rightarrow\lambda_i) \approx \frac{\partial O}{\partial\lambda_i}(\lambda_i - 0) = \frac{\partial O}{\partial\lambda_i}\lambda_i \qquad (3)$$

When O represents the binding free energy ΔG_{BIND}, this equation becomes

$$\Delta\Delta G_{\mathrm{BIND}}(0\rightarrow\lambda_i) \approx \frac{\partial\Delta G_{\mathrm{BIND}}}{\partial\lambda_i}(\lambda_i - 0) = \frac{\partial\Delta G_{\mathrm{BIND}}}{\partial\lambda_i}\lambda_i \qquad (4)$$

Since a decrease in ΔG_{BIND} signifies an improvement in binding affinity, parameters that produce a negative $\Delta\Delta G_{\mathrm{BIND}}$ when turned on are considered useful for binding.

Pairwise Simulations. Double mutagenesis experiments have been used to probe the effects of two residues on free energy changes. Second-order sensitivity analysis provides similar information. Consider turning on two parameters, λ_i and λ_j, separately. Each would induce a change in free energy according to

$$\Delta\Delta G_{\mathrm{BIND}}(0\rightarrow\lambda_i) \approx \frac{\partial\Delta G_{\mathrm{BIND}}}{\partial\lambda_i}\lambda_i + \frac{1}{2}\frac{\partial^2\Delta G_{\mathrm{BIND}}}{\partial\lambda_i^{\,2}}\lambda_i^{\,2} \qquad (5)$$

and

$$\Delta\Delta G_{\mathrm{BIND}}(0\rightarrow\lambda_j) \approx \frac{\partial\Delta G_{\mathrm{BIND}}}{\partial\lambda_j}\lambda_j + \frac{1}{2}\frac{\partial^2\Delta G_{\mathrm{BIND}}}{\partial\lambda_j^{\,2}}\lambda_j^{\,2} \qquad (6)$$

where now we use a second-order, rather than a first-order, approximation to calculate the effects of changing a parameter on the binding free energy. (This is exact in the fixed-conformation Poisson model used here, although it is only an approximation in more rigorous molecular dynamics simulation models.) Likewise, when these two parameters are turned on simultaneously, the change in binding free energy induced is given by

$$\Delta\Delta G_{\mathrm{BIND}}(0\rightarrow\lambda_i,0\rightarrow\lambda_j) \approx \frac{\partial\Delta G_{\mathrm{BIND}}}{\partial\lambda_i}\lambda_i + \frac{1}{2}\frac{\partial^2\Delta G_{\mathrm{BIND}}}{\partial\lambda_i^{\,2}}\lambda_i^{\,2} + \frac{\partial\Delta G_{\mathrm{BIND}}}{\partial\lambda_j}\lambda_j + \frac{1}{2}\frac{\partial^2\Delta G_{\mathrm{BIND}}}{\partial\lambda_j^{\,2}}\lambda_j^{\,2} + \frac{\partial^2\Delta G_{\mathrm{BIND}}}{\partial\lambda_i\,\partial\lambda_j}\lambda_i\lambda_j \qquad (7)$$

Equations 5−7 imply that

$$\Delta\Delta G_{\mathrm{BIND}}(0\rightarrow\lambda_i,0\rightarrow\lambda_j) - \Delta\Delta G_{\mathrm{BIND}}(0\rightarrow\lambda_i) - \Delta\Delta G_{\mathrm{BIND}}(0\rightarrow\lambda_j) = \frac{\partial^2\Delta G_{\mathrm{BIND}}}{\partial\lambda_i\,\partial\lambda_j}\lambda_i\lambda_j \qquad (8)$$

The left-hand side corresponds to the quantity measured experimentally in a double-mutant cycle. Equation 8 therefore suggests that $(\partial^2\Delta G_{\mathrm{BIND}}/\partial\lambda_i\partial\lambda_j)\lambda_i\lambda_j$ gives similar information. A negative (positive) $[\partial^2\Delta G_{\mathrm{BIND}}/(\partial\lambda_i\,\partial\lambda j)]\lambda_i\lambda_j$ suggests favorable (unfavorable) interactions between λ_i and λ_j. In this work, $[\partial^2\Delta G_{\mathrm{BIND}}/(\partial\lambda_i\,\partial\lambda_j)]\lambda_i\lambda_j$ was calculated by using the continnum-solvent model in UHBD,[13,14] although it can also be calculated with more expensive molecular dynamics simulation models.[8]

Analogue and Designed Structure Modeling. InsightII[20] was used for the modeling of the flavopiridol analogues and the designed structures. Using its Builder module, we constructed these structures from the coordinates of des-chloroflavopiridol given by the crystal structure.[5] The initial configurations of added functional groups were chosen visually to enable good interactions with their target residues. The altered functional groups were then allowed to relax in their environment with energy minimization. During the energy minimization, we did not allow the environment to move so that cancellation of errors would be more likely to occur in the subsequent binding energy calculations in which the fixed-conformation approximation was used. In other words, we estimated the free energy of a molecular system by

$$G = W(q_{\mathrm{min}},q_{\mathrm{environment}}) \qquad (9)$$

where $W(q_{\mathrm{min}},q_{\mathrm{environment}})$ was the potential of mean force, averaged over solvent coordinates, defined by the coordinates q_{min} and $q_{\mathrm{environment}}$. $q_{\mathrm{environment}}$ defined the coordinates of the atoms common to all of the structures studied, whereas q_{min} denoted the coordinates of atoms unique to each structure. The q_{min} values for functional group atoms were obtained by adjusting them to minimize the energy of the protein−ligand complex while holding $q_{\mathrm{environment}}$ fixed in an attempt to model a more probable configuration. In cases where ligand modifications might experience steric clashes with one or more amino acid residues of the protein, q_{min} also includes hydrogens on functional groups, such as methyl and hydroxyl, of neighboring amino acid residues. However, this was not done for most structures because functional groups were usually introduced to the parts of the ligand where sufficient space is available to accommodate them. Energy minimization was run with the CVFF force field in InsightII[20] and the steepest descent algorithm for 500 iterations to obtain an rms derivative less than or equal to 0.01 kcal/(mol·Å). Additionally, a distance-dependent dielectric was used during the energy minimization. As mentioned previously, the atomic charges for these structures were assigned with Gaussian 98[12] and the thermodynamic data were generated by UHBD.[13,14] In a binding energy calculation, the energy-optimized ligand conformation obtained above was used for both the free ligand and the complex. This way, cancellation of errors is more likely in the finite-difference solution of the Poisson equation because the ligand can be positioned in exactly the same way in the grid whether it is in the bound or unbound form. We used this fixed-conformation approximation consistently in this work, and the goal is to examine how well this approximation works in ranking the binding affinity of similar ligands. Ignoring change in the intramolecular energy of the ligands may not be a bad approximation if this term is comparable for all the ligands studied and if its variation is small in comparison to other energy

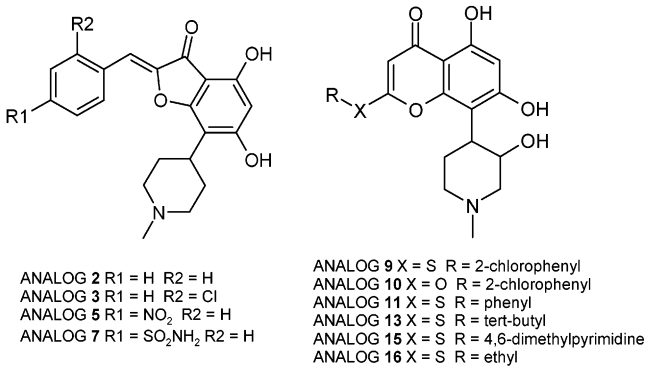

Figure 1. Structures of the flavopiridol mimics of Schoepfer et al.[28] (left) and the structures of the thio- and oxoflavopiridols of Kim et al.[29] (right).

terms that play a more important role in distinguishing binding affinity. By inclusion of bulkier groups in this study, it was feasible to identify the limits beyond which this approximation fails. We first examined whether this calculation model could rank the binding affinity of a number of similar compounds to a target by comparing simulation results with biological experiments. Once validated, we used the model to help design new inhibitors. For some of the structures used in this validation study, more than one attempt was required because of the need for further refinement. The structures of these compounds are depicted in Figure 1, and their respective numbers represent a particular modeling attempt rather than a particular compound. Unsuccessful or incomplete attempts are omitted here, and so some numbers are skipped.

The process of preparing the necessary data sets for a single compound (including energy minimization) can be completed in tens of minutes manually. This time can be reduced by writing scripts/programs to automate this process. Running UHBD on a 300 residue system (such as the complex of CDK2 with an inhibitor) takes about 20 min on an SGI Octane2/R12000. This can be further speeded up if the focusing technique is used in solving the Poisson equation. Running Gaussian 98 on the same computer to obtain ligand charges takes 35−60 min depending on the size of the ligand, but faster methods for calculating atomic partial charges can be used (e.g., the AM1-BCC method[21]).

Hybrid Structural Model Construction. To study the selective binding of these compounds to CDK1, CDK2, and CDK4, we built homology models for CDK1 and CDK4, using the crystal structure of CDK2[5] as a template. Since it is still difficult and time-consuming to build homology models well enough for energy calculations, we have explored the utility of building quick hybrid models in this work. These models only approximately describe the environment of the active site, using coordinates of the amino acids in the template structure (CDK2 in this case). Residues with most or all of their atoms within a 10 Å radius of the ligand, however, are modeled in atomic detail, using coordinates from the template structure as much as possible. This way, one can focus on refining the regions that are most important for binding while ensuring that all models are built in a consistent way. This method makes it easier to compare the relative binding energy of a number of similar inhibitors to several closely related protein kinases. Initially, the sequences of human CDK1[22] and CDK4[23] were taken from the Protein Kinase Resource at the San Diego Supercomputer Center.[24,25] Using ClustalW,[26] we aligned the two sequences individually with CDK2 to determine which residues in our 10 Å shell had to be altered in the CDK2 crystal structure in order to mimic the two protein kinases. Next, we constructed the two hybrid models using the InsightII Biopolymer module.[20] Once these preliminary models were constructed, the InsightII Discover module[20] was utilized to energy-minimize the added residues. For this process, the des-chloroflavopiridol was present in the same conformation and position as in the CDK2 crystal structure. Energy minimization with the CVFF force field took place in 7000 iterations of the steepest descent algorithm such that the square root of the average magnitude of the force was less than or equal to 0.01 kcal/(mol·Å). All atoms were fixed except for those in the newly added amino acids, and we specified a distance-dependent dielectric as in the previously described energy minimizations.

Survey of the Protein Kinase Database. Complementary to homology modeling for energy calculations, which is still difficult to apply accurately to a large number of homologous proteins, we have also applied an alternative approach to accounting for specificity in drug design. In this approach, one only attempts to construct a rough picture of the binding pocket of protein kinases, but for a large number of them.[6,7] The goal is not to obtain reliable binding energy but to generate guidelines that can improve the odds of finding selective inhibitors. By knowing which parts of a ligand-binding pocket are more variable and by designing inhibitors to target these regions, one may have a better chance of achieving specificity. In this work, a database of almost 400 protein kinase sequences from the Protein Kinase Resource[24,25] was surveyed for conservation at positions surrounding the binding site. By using the sequence alignment of these kinases in the database[27] and the crystal structure of CDK2,[5] one can construct a picture of the binding site of all the proteins in the database and carry out statistical analysis to examine which sites are conserved and which sites are variable. The conserved sites are not likely to present specificity, whereas variable sites are potentially useful targets for specificity.

Results and Discussion

Correlation with Biological Data. In this portion of the study, we attempted to validate our computational model with biological data. We modeled two groups of flavopiridol analogues that had been synthesized and assayed biologically in previous studies into the crystal structure of CDK2. The first group consisted of four compounds (Figure 1) synthesized and tested by Schoepfer et al.,[28] and we will refer to them as the flavopiridol mimics. We also addressed a group of six compounds (Figure 1) from Kim et al.[29] which we will call the thio- and oxoflavopiridols.

Using the previously described method, we computed the binding free energies of the flavopiridol mimics with UHBD.[13,14] We then graphed the resulting ΔG_{BIND} values against pIC_{50}[28] (Figure 2) and obtained a good correlation. The calculated free energies order the four

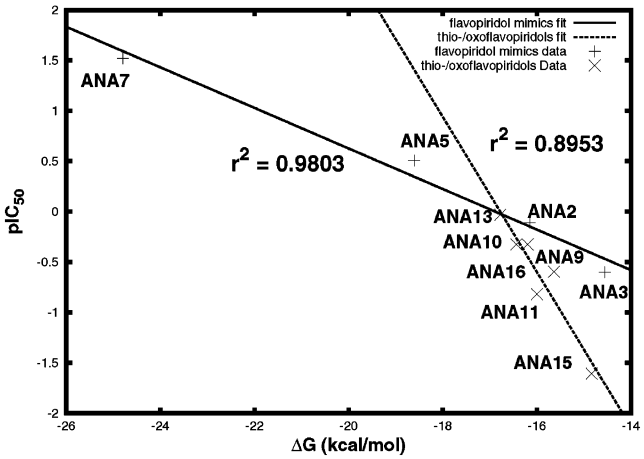

Figure 2. Combined graphs of pIC_{50} vs ΔG (kcal/mol) for the flavopiridol mimics and the thio- and oxoflavopiridols modeled in the CDK2 crystal structure.

Table 1. Thio- and Oxoflavopiridols Cross Validation Analysis

analogue	IC_{50} (μM)	predicted IC_{50} (μM)	pIC_{50}	predicted pIC_{50}
13	1.07	0.95	−0.029	0.021
9	2.10	3.01	−0.322	−0.478
10	2.11	1.74	−0.324	−0.241
16	3.93	9.15	−0.594	−.961
11	6.59	3.64	−0.819	-0.562
15	40.4	15.9	−1.610	−1.200
rmsd		10.3		0.258

Table 2. Flavopiridol Mimics Cross-Validation Analysis

analogue	IC_{50} (μM)	predicted IC_{50} (μM)	pIC_{50}	predicted pIC_{50}
7	0.03	0.0063	1.520	2.200
5	0.31	0.51	0.509	0.294
2	1.28	1.47	−0.107	−0.166
3	3.97	2.15	−0.599	−0.333
rmsd		0.92		0.381

ligands according to biological potency, and the r^2 value for the linear regression line is 0.98. (The linear regressions in this work were performed with Microsoft Excel.) An identical experiment was carried out with the thio- and oxoflavopiridols in CDK2. Although the compounds are ordered less well in this case, the overall trend expressed by the graph of ΔG_{BIND} vs pIC_{50}[29] is clear (Figure 2). The r^2 value for the linear regression line in this second series of compounds is still quite good: 0.90.

We have also carried out a leave-one-out cross-validation analysis on our correlation data (Tables 1 and 2). This analysis consists of throwing out a data point in one data set and performing a linear regression on the remaining points. We then calculate how well the new linear regression line predicts the missing point. For the thio- and oxoflavopiridols, the cross-validated r^2 becomes 0.75. The root-mean-square-deviation (rmsd) in pIC_{50} is 0.26 (about 16% of the range of pIC_{50}) and that in IC_{50} is 10 μM (about 26% of the range). This large IC_{50} rmsd is due to the outlying analogues **15** and **16**, which respectively have a significantly larger and smaller functional group than the remaining compounds. The flavopiridol mimics work somewhat better. The cross-validated r^2 are still quite good: 0.91. The rmsd in pIC_{50} was found to be 0.38 (about 18% of the range of pIC_{50}), and the rmsd in IC_{50} was found to be 0.92 μM (about 23% of the range of IC_{50}). The cross-validation analysis cautions that the fixed-conformation model should probably be used only for structures that are similar to the parent compound if quantitative results are desired. However, the model may still be used in a qualitative manner for designing structures that are less similar to the parent compound. For example, the model seems to be able to distinguish potent structures from nonpotent ones. Also, designed structures that have favorable calculated binding free energies may have higher odds to be potent inhibitors and can be selected for experimental evaluations. With this limitation of the model in mind, we have applied it to design new inhibitors.

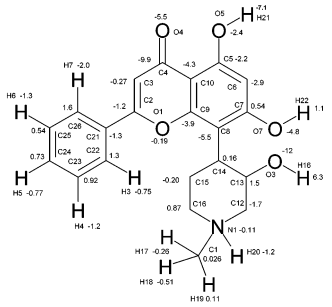

Figure 3. Structure of des-chloroflavopiridol labeled with the $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ values (kcal/mol) from sensitivity analysis where a negative number represents a useful charge.

Sensitivity Analysis of Des-chloroflavopiridol. We first carried out sensitivity analysis to gain some insights into which parts of the lead compound are profitable to keep and which portions should be improved to enhance binding. The sensitivity analysis has generated a description of charge utility in the ligand with respect to CDK2 binding. Figure 3 maps the charge utility of the ligand using $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ detailed earlier. It should be noted that the ΔG_{BIND} that was calculated for the interactions between deschloroflavopirdol and CDK2 is −14.0 kcal/mol. $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ reflects the change of binding free energy resulting from turning on different parameters λ_i, which are atomic partial charges in this study. We can draw from the figure that the most useful functional groups on the ligand are the hydroxyl group C5−O5−H21 and the ketone C4−O4, as indicated by the large negative $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ that they present. This is consistent with the previously proposed pharmacophore model that suggests that many small-molecule protein kinase inhibitors utilize two hydrogen bonds to recognize the backbone of the linker region between the N- and C-terminal lobes of the kinase domain.[30] In des-chloroflavopirdol, C5−O5−H21 serves as the hydrogen bond donor with the ketone C4−O4 as the hydrogen bond acceptor. In the phenyl ring, the polarity of one of the C−H bonds (C22−H3) hurts binding. Another one (C24−H5) is only mildly useful for binding. One may want to modify these parts of the phenyl ring to see whether more favorable inhibitors can be designed.

Another interesting result comes from the charge utility of the CH_3−N−H group (C1(H17,H18,H19)−N1−H20), which has an additive $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ of −1.9 kcal/mol. This group is the source of formal charge for the ligand, and sensitivity analysis suggests that this charge is useful for binding.

Several combinatorial chemistry studies of flavonoid CDK inhibitors have taken place in recent years. The

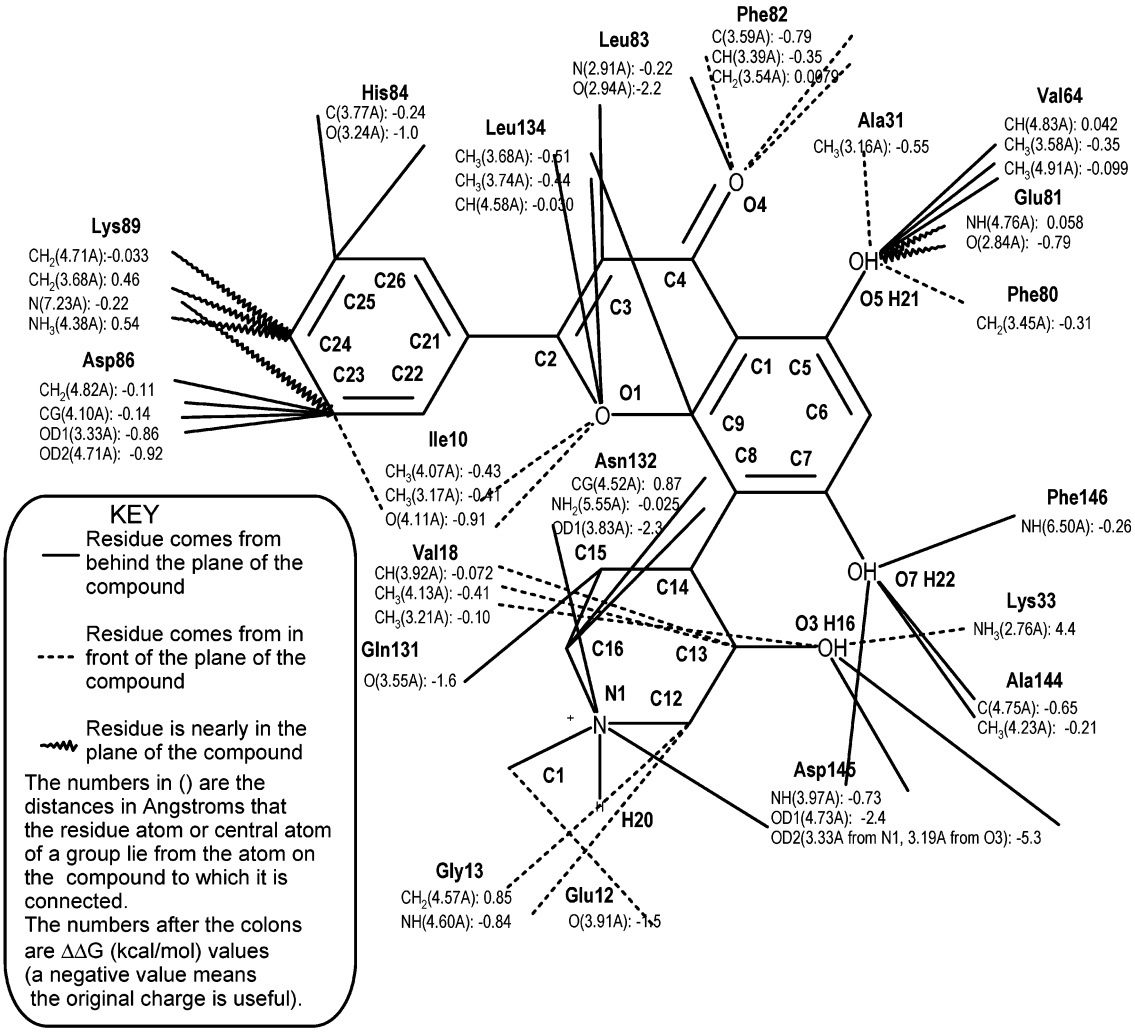

Figure 4. Active site of CDK2 with des-chloroflavopiridol labeled with the $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ values (kcal/mol) from sensitivity analysis where a negative number represents a useful charge.

results of our sensitivity analysis express the major conclusions of these studies fairly consistently. For example, Fischer and Lane[3] emphasize the importance of the two flavone hydroxyl groups C5−O5−H21 and C7−O7−H22. These two sites have additive $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ of −11.8 and −3.2 kcal/mol, respectively. A thorough study of the 3-hydroxylmethylpiperidyl ring conducted by Murthi et al.[31] adds that the third hydroxyl group (C13−O3−H16) also plays a crucial role (especially its stereochemistry), and we have measured its cumulative utility to be −4.0 kcal/mol.

Sensitivity Analysis of the CDK2 Active Site. In an identical fashion, we carried out a sensitivity analysis on 20 active site amino acids covering 103 protein atoms to determine which ones are interacting favorably with the ligand. This analysis can also identify the parts of the proteins that do not interact favorably with the ligand yet and thus can suggest modifications of the ligand toward improved interaction with these parts of the protein. The charge utility data pictured in Figure 4 can be interpreted just as that in Figure 3, with a large negative $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ indicating favorable charge utility and with a large positive $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ suggesting unfavorable charge utility. As mentioned earlier, many small-molecule inhibitors utilize two hydrogen bonds to recognize the linker region between the N-terminal and C-terminal lobes of the catalytic domain. The sensitivity analysis of the ligand presented in the previous section is consistent with this notion in which the ketone C4═O4 and the hydroxyl group C5−O5−H21 serve as the hydrogen bond acceptor and donor, respectively. The sensitivity analysis of the protein further echoes this finding by revealing which parts of the protein form hydrogen bonds with these two functional groups. The NH group of Leu83 serves as the hydrogen bond donor of the ligand ketone group and demonstrates a large negative $(\partial\Delta G_{BIND}/\partial\lambda_i)\lambda_i$ suggesting favorable electrostatic interactions with the ligand. Pairwise sensitivity analysis also confirms the interactions between this NH group and the ligand ketone group to be strong, with a large negative $[\partial^2\Delta G_{BIND}/(\partial\lambda_i\,\partial\lambda_j)]\lambda_i\lambda_j$ of −6.0 kcal/mol between this NH group and the oxygen of the ketone group and a modest $[\partial^2\Delta G_{BIND}/(\partial\lambda_i\,\partial\lambda_j)]\lambda_i\lambda_j$ of −0.2 kcal/mol between the NH group and the carbon of the ketone group. The hydroxyl group C5−O5−H21 acts as a hydrogen bond donor to interact with the carbonyl group of Glu81 with $[\partial^2\Delta G_{BIND}/(\partial\lambda_i\,\partial\lambda_j)]\lambda_i\lambda_j$ of the order of −0.8 kcal/mol. For the all-atom, as opposed to the united-atom, model used in these calculations in which small dipole moments on hydrocarbon C−H bonds are accounted for, this hydroxyl group was found to interact modestly with the dipoles of several methyl/

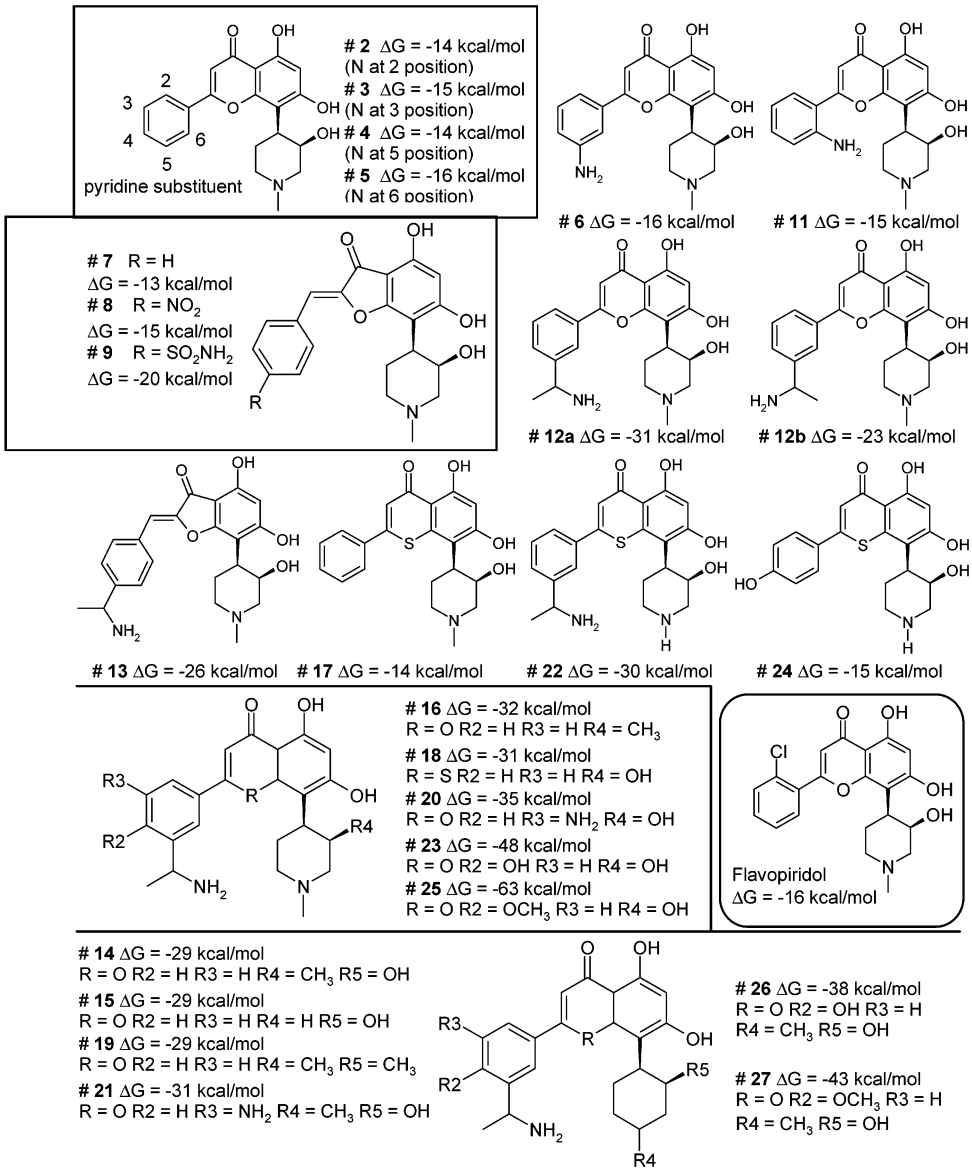

Figure 5. Structures of the computationally designed structures and their respective computed binding free energies (kcal/mol) with CDK2. Doubly protonated structures are largely grouped separately from the singly protonated ones because only the latter structures were included in the regression analysis.

methylene groups nearby with $[\partial^2\Delta G_{BIND}/(\partial\lambda_i\,\partial\lambda_j)]\lambda_i\lambda_j$ ranging from −0.3 to −0.5 kcal/mol. The aforementioned utility of the C13−O3−H16 group seems to arise from its favorable interactions with the carboxylate group of Asp145 ($[\partial^2\Delta G_{BIND}/(\partial\lambda_i\,\partial\lambda_j)]\lambda_i\lambda_j$ equals −3.5 kcal/mol with OD2 of Asp145), although this is dampened somewhat by its unfavorable interactions with the ammonium group of Lys33 ($[\partial^2\Delta G_{BIND}/(\partial\lambda_i\,\partial\lambda_j)]\lambda_i\lambda_j$ equals 0.9 kcal/mol). Additionally, the cluster of useful charges surrounding the phenyl ring from Asp86, His84, and Ile10 suggests that introducing suitable polar or charged groups to the phenyl ring to better utilize these residues may improve binding affinity. For example, a major target in future design could be the charged OD1−CG−OD2 group on Asp86, which has an additive utility of −1.9 kcal/mol. Since this group is about 4 Å away from the closest atoms in the phenyl ring of des-chloroflavopiridol, introducing suitable substituents to the ring to get closer to Asp86 could improve binding further. The NH_3^+ group of Lys89 shows an unfavorable charge utility of 0.54 kcal/mol. Thus, introducing suitable functional groups to the phenyl ring to reverse the damaging effects of this group could be useful.

In directing sensitivity analysis toward the active site atoms, we show the explanatory power of this model with respect to known experimental data. Additionally, the ability to examine charge utility at the atomic level provides an immediate set of suggested modifications to the current ligand.

Computational Design of CDK2 Inhibitors. Using the wealth of data given by previous combinatorial chemistry studies and our computational data, we turned our general method toward inhibitor design. We hope that the designed structures (Figure 5) can be evaluated with experimental measurements in the future to check whether they are indeed useful inhibitors. The most important points of comparison for considering these designed structures are the ΔG_{BIND} values that we correlated with biological data and our ΔG_{BIND} values for the flavopiridol and des-chloroflavopiridol models which are −16.5 kcal/mol (IC_{50} = 0.17−0.40 μM[28]) and −14.0 kcal/mol, respectively.

The first group of potential inhibitors (structures **2**−**5**) is based mainly on structures described in the literature for which the corresponding biological information is qualitative.[3] This experiment consisted of placing a heterocyclic nitrogen at four different positions on the phenyl ring. These structures only have moderate ΔG_{BIND} values ranging from −13.7 to −16.1 kcal/mol (somewhat less than the value calculated for flavopiridol). This result is consistent with Fischer and Lane's work[3] in which some reduction of biological potency relative to flavopiridol takes place with the substitution of the chlorophenyl group by different pyridyl groups.

The sensitivity analysis data suggest that the polarity of three of the C−H bonds (C23−H4, C22−H3, C24−H5) in the phenyl ring is not as significant as the other two in binding. Thus, we have examined the introduction of different functional groups at these three positions. Structures **6** and **11** were each obtained by replacing the hydrogen of one such bond by an amine group in an attempt at gaining binding affinity by improving interactions with Asp86. The calculated ΔG_{BIND} values were −16.1 and −15.2 kcal/mol, respectively, and so we conclude that there is some improvement over the des-chloroflavopiridol (ΔG_{BIND} = −14.0 kcal/mol). However, it appears that the amine group is not as useful as the chloro substituent in flavopiridol. Modification of the last of these three C−H bonds will be discussed below.

Structures **7**−**9** tested the utility of the hydroxyl group that is absent in the flavopiridol mimics of Schoepfer et al.[28] (the one attached to the piperidine ring) using the scaffolds of three structures in that series. In all three cases, there is a reduction of binding affinity by about 4 kcal/mol. This result is curious because the majority of our sensitivity analysis data and the biological data from other studies agree that the polarity of this group is somewhat useful. However, the flavopiridol mimics do not use the same molecular scaffold as more direct derivatives of flavopiridol (i.e., the flavone ring is replaced by a benzofuran-3-one ring).[28]

Our model suggests that the functional group tested in this next series of structures exhibits a very favorable influence on binding affinity. Structures **12a**, **12b**, and **13**−**16** all possess a chiral substituent to the phenyl ring where the stereocenter is attached to a hydrogen, a methyl group, and a charged ammonium group. In structure **13**, this new group is attached to C24 on the phenyl ring (but uses the flavopiridol mimics scaffold), and in the remaining structures of this series, it is attached to C23. In each case, there is a dramatic increase in the computed binding affinity compared with flavopiridol. Structures **12a** and **12b** test two stereochemical configurations of the new substituent. Since the original purpose of constructing this series was to establish hydrogen bonding between Asp86's two carboxyl oxygens, it is not surprising that the *R*-configuration of structure **12a** had greater binding affinity (−31.1 kcal/mol) than structure **12b**'s *S*-configuration (−23.4 kcal/mol). This arrangement allows the ammonium group to interact with Asp86 and the less polar methyl group to interact with the nonpolar residue Ile10. These two amino acids are located about 4 Å from C23. Such an addition to the structure may allow for greater interaction between the residues and the ligand without creating excessive repulsion. Structures **13**−**16** all adopt this *R*-configuration and test the utility of other functional groups in the original ligand. As was suggested by the sensitivity analysis, replacement of the formally charged nitrogen in the piperidinyl ring by carbon in structures **14** and **15** results in a small cost to binding affinity (about −2.2 kcal/mol) relative to structure **12a**. One may sacrifice this group to improve membrane permeability if the new positively charged ammonium group is added to the phenyl ring. However, structure **16** gives a curious result when a slight increase in binding affinity (just under 1 kcal/mol) occurs upon replacement of the piperidinyl ring's hydroxyl group with a methyl group. Despite this surprising result, it is not necessarily contradictory of our earlier conclusions. The utility of the hydroxyl group is not a direct implication that the methyl group will lack utility.

In structures **17** and **18**, the ether oxygen in the flavone portion of the molecule is replaced by a sulfur. Structure **18** includes this thioether and the lead substituent from the previous series, which results in a very slight decrease in binding affinity relative to structure **12a**. Similarly, in structure **17** (where the lead substituent is absent), only a slight increase in binding affinity relative to des-chloroflavopiridol takes place. Structure **19** combines the two experiments with the piperidinyl ring in replacing the nitrogen with carbon and the hydroxyl group with a methyl group. The resulting ΔG_{BIND} is −29.4 kcal/mol.

Structures **20** and **21** represent another attempt at design based on the sensitivity analysis and pairwise data. Following our scheme of improvements to sites that are far away from the very useful charges but that can still reach important residues, an amine group was added meta to the lead substituent. The primary reason for this addition was to increase ligand contact with HIS84, which already has two useful charges that we knew of from the sensitivity analysis. The added amine group perturbs the charge distribution of the phenyl ring without adding formal charge to the ligand because of the aromatic system. Structure **20** has a more favorable calculated ΔG_{BIND} of −34.5 kcal/mol relative to structure **16**, suggesting the addition of this amine does improve binding. Structure **21** again examines the effects of replacing the piperidinyl nitrogen with carbon to lower formal charge to increase membrane permeability when needed; only 3.5 kcal/mol is lost with this change. A few pairwise simulations were run for these two structures to further investigate this interaction. The amine group nitrogen in structure **20** had a much improved second-order value of −12.4 kcal/mol, while structure **21** had a second-order value of −12.1 kcal/mol for the interaction with the carbonyl oxygen on His84 that was tested in the original des-chloroflavopiridol simulations.

In structures **23**−**27**, we experimented with the site on the phenyl group (C24−H5) that was demonstrated by sensitivity analysis to have only modest charge utility. This carbon is also located about 4−7 Å from Lys89, which was demonstrated by sensitivity analysis to have unfavorable utility for the ammonium group. This suggests that improvements at this site could be

useful. It should also be mentioned that the nature of the active site allows for more freedom in modeling at this position than the C25 position, which made experimentation with both hydroxyl and methoxy groups possible. For structure **23**, we added a hydroxyl group to C24 along with the lead chiral substituent in the same *R*-configuration. A ΔG_{BIND} of −47.9 kcal/mol was computed for this structure. According to our model, this thermodynamic binding affinity is nearly 3 times that of the lead compound flavopiridol. A sensitivity analysis was conducted on structure **23** to probe the charge utility of the added functional groups. The additive utility of the newly added C−O−H group was −3.2 kcal/mol, while that of the $C-CH_3-NH_3^+$ group was −27.2 kcal/mol. These two groups also had a significant effect on the charge distribution of the phenyl ring. The total charge on C26−H7 in des-chloroflavopiridol (as assigned by Gaussian 98[12]) was −0.05*e*, while the same C−H group in structure **23** had a total charge of 0.33*e*. The additive charge utility of the group also changed from −0.4 to −6.0 kcal/mol. However, these effects are not nearly as pronounced when the hydroxyl and flavone groups are the only substituents to the phenyl ring as in structure **24**. This ligand has a binding affinity of only −14.8 kcal/mol, which is less favorable than flavopiridol but slightly more favorable than des-chloroflavopirdol by 0.8 kcal/mol. Fischer and Lane[12] comment on flavopiridol derivatives such as structure **24**, citing a reduction in biological potency associated with such functionality that is consistent with our calculations. Structure **26**, when compared to structure **23**, shows that about 10 kcal/mol of binding affinity must be sacrificed to reduce the formal charge of the molecule, although this structure may be more membrane-permeable. Structures **25** (ΔG_{BIND} = −63.4 kcal/mol) and **27** (ΔG_{BIND} = −43.4 kcal/mol) take advantage of the freedom associated with modeling at this site. The methoxy substituent replaces the hydroxyl group in these two structures to give another large increase in computed binding affinity. The methoxy groups allow contacts with both the hydrophilic and hydrophobic parts of Lys89. From the results of energy minimization on the methoxy groups (in structures **25** and **27**), there also appears to be a contact with the aromatic ring on Phe82.

Although we have no biological data to confirm the correlation between our computational model and the actual potency of these particular structures, we have presented data suggesting such a correlation for other inhibitors. The molecules with the most potential according to our model are structures **25**, **23**, **27**, **26**, and **20**. We believe that on the basis of our results from modeling, the flavopiridol mimics and the thio- and oxoflavopiridols, the designed structures presented in this section, merit laboratory study.

Hybrid Homology Modeling for Selectivity Data. Using the procedure mentioned earlier, we built the flavopiridol mimics (for which IC_{50} values in CDK1 and CDK4 were available[28]) into our hybrid models of CDK1 and CDK4. Unfortunately, our original modeling attempts with these compounds and the thio- and oxoflavopiridols did not produce good results. One aspect of binding thermodynamics that our model did not explicitly take into account is the rotational entropy lost by

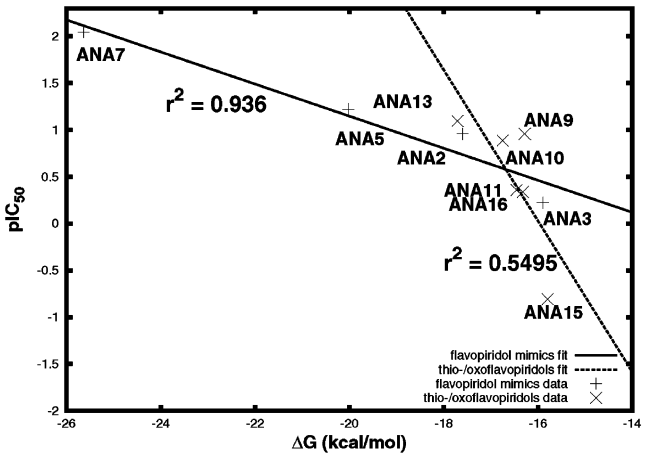

Figure 6. Combined graphs of pIC_{50} vs ΔG (kcal/mol) for the flavopiridol mimics and the thio- and oxoflavopiridols modeled in the CDK1 hybrid model with an adjustment for rotational entropy.

rotatable bonds of the ligands upon binding. Past computational studies such as by Böhm,[32] Morris et al.,[33] and Gidofalvi et al.[34] showed that these effects are not negligible for rotatable bonds between sp^3-hybridized atoms and for rotatable functional groups attached to aromatic rings. Böhm[32] and Morris et al.[33] used a value of 0.3 kcal/mol per bond in estimating these entropy contributions to binding free energies. We also adjusted the ΔG_{BIND} values of the compounds built into these hybrid models accordingly and observed a significantly more reasonable relationship between computational and biological data.

The adjusted binding free energies for the CDK1 hybrid model ordered the four compounds correctly according to biological potency just as in the CDK2 model. The linear regression line for the graph of pIC_{50} vs ΔG_{BIND} (Figure 6) also revealed a good linear trend with an r^2 value of 0.94. The similarities in accuracy between the CDK2 crystal structure model and the hybrid model of CDK1 are not surprising because the two protein kinases have nearly 66% sequence identity overall. Additionally, we found that if the entropy adjustment was applied to the original flavopiridol mimics model in CDK2 and the resulting values were plotted on the same graph as the new CDK1 data (Figure 7), a good r^2 value of 0.87 still results. This information suggests that our CDK1 hybrid model combined with the CDK2 crystal structure model could provide some degree of selectivity information for structures similar to the flavopiridol mimics.

We were met with much less success in modeling the flavopiridol mimics in the CDK4 hybrid model. A nearly straight line results from modeling the first three compounds, but the final inhibitor appears far out of alignment, resulting in an r^2 value of about 0.5 for the linear regression (Figure 8). (However, r^2 improved to 0.68 when the problematic point was removed.) Nonetheless, the CDK4 hybrid model does rank the flavopiridol mimics in the correct order of biological potency. It should be noted that in creating the CDK4 hybrid model many more residues had to be mutated because CDK4 has only about 45% sequence identity with CDK2.

The same two hybrid models were applied to the thio- and oxoflavopiridols.[29] This group of compounds followed

Table 3. Degree of Amino Acid Conservation at Different Sites near the Ligand-Binding Pocket

	degree of conversion (%)														
residue	A144	A31	N132	D145	D86	E12	G13	I10	L134	K33	K89	F80	F82	V18	V64
G	12.2	0	0	0	0	1.8	99.0	0	0	0	1.0	0.3	0	0	0
A	32.0	93.2	0	0.3	1.8	9.1	0	0.3	0	0	4.2	0	1.0	1.6	1.3
V	2.9	4.7	0	0	0.3	2.3	0	8.0	1.8	0	1.0	3.4	0.3	93.2	51.6
L	0.8	0.5	0	0	0	1.0	0	55.7	81.8	0	1.3	16.9	22.1	0.5	9.1
I	8.6	0.5	0	0	0	1.0	0	34.1	1.3	0	0.8	2.3	2.3	1.6	24.7
S	15.4	0	0	0	25.5	12.0	0.5	0.3	0.3	0	10.7	1.3	0.3	0.3	0.5
T	13.8	0	0	0	3.6	5.0	0.3	0	0	0	7.6	20.8	0	1.0	4.7
D	0	0.3	0	98.7	38.5	1.0	0.3	0	0	0.3	24.2	0	0	0	0.3
N	0	0	100	0	7.8	2.6	0	0.3	0	0	5.0	0.3	0	0	0
K	0	0	0	0	0.3	17.4	0	0.8	0	100	8.3	0	0.5	0	0.5
E	0	0	0	0.8	17.2	20.3	0	0	0	0	9.6	0.3	0.5	0	0.3
Q	0	0	0	0	0.8	7.0	0	0	0	0	7.8	2.9	0	0	0.5
R	0	0	0	0	0	13.0	0	0	0	0.3	5.7	0.3	0.5	0	0
H	0.3	0	0	0	0	2.1	0	0	0	0	3.1	0	3.1	0	0.3
F	0	0.3	0	0	0.3	0.5	0	0.5	5.2	0	4.2	17.2	21.1	0	0
C	14.1	0.8	0	0	2.1	2.3	0.3	0	0	0	0	0	1.8	2.1	1.6
W	0	0	0	0	0	0.3	0	0	0	0	1.0	0	1.8	0	0
Y	0	0	0	0.3	0	0.5	0	0	0	0	3.4	4.7	43.5	0	0.3
M	0	0.3	0	0	0.3	0.8	0	0.3	9.6	0	1.0	29.4	1.0	0	0.8
P	0	0	0	0	1.3	0	0	0	0	0	0	0	0	0	3.6
−		0	0	0	0.3	0.3	0.3	0.3	0	0	0	0	0	0.3	0

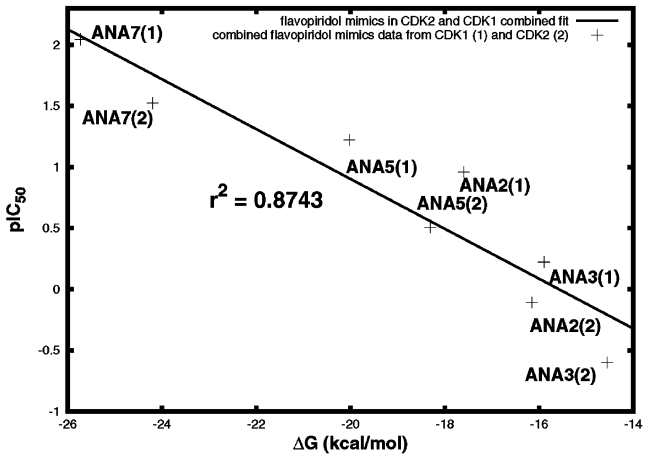

Figure 7. Combined plot of pIC_{50} vs ΔG (kcal/mol) for the flavopiridol mimics in the CDK1 hybrid model and the CDK2 crystal structure with an adjustment for rotational entropy.

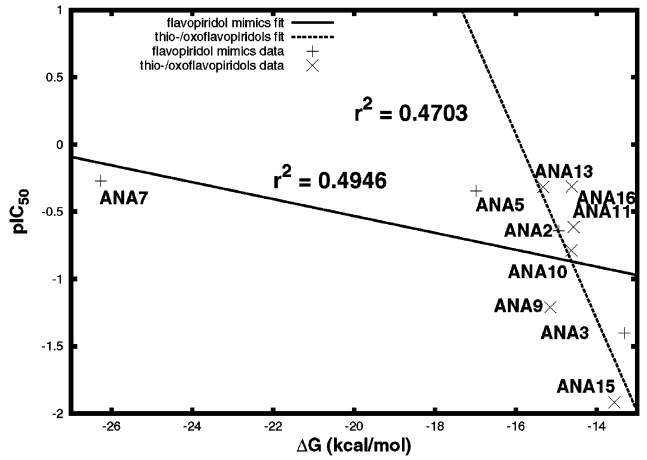

Figure 8. Combined graphs of pIC_{50} vs ΔG (kcal/mol) for the flavopiridol mimics and the thio- and oxoflavopiridols modeled in the CDK4 hybrid model with an adjustment for rotational entropy.

the linear trend rather effectively for five of the six compounds in the CDK1 hybrid model (Figure 7), until the most favorable one complicated the model. The resulting linear regression line had an r^2 value of 0.55 (improved to 0.78 when the problematic compound was removed). Interestingly, the most potent inhibitor in the CDK4 model of the flavopiridol mimics also destroyed the linear trend. The same group of six compounds exhibited a poor correlation ($r^2 = 0.47$, improved to 0.83 when the problematic compound was removed) in the CDK4 hybrid model just as with the flavopiridol mimics (Figure 8).

It was because of our success in modeling the flavopiridol mimics and the thio- and oxoflavopiridols in the CDK2 crystal structure that we attempted to model selectivity as described above. Rather than concluding that the thermodynamic model is incapable of gauging selectivity, we believe that these less reasonable results stem more from our hybrid homology modeling technique. Without a detailed force-field model accounting for solvation effects realistically, we took a conservative approach to refining the homology structures after they were built from the template structure. Relatively minor conformational adjustments were allowed. The idea was not to let the homology structures deviate too much from the experimental template structure because refining the homology structures using a less accurate force field model can generate structures that might even be more unreliable. Long-range effects were also only approximated by the homologous CDK2 structure, and conformational fluctuations were ignored. Improved technology in the area of fast homology modeling could expand the capabilities of this thermodynamic model toward selectivity determination. Nevertheless, it is gratifying to see that the hybrid homology techniques employed here could at least order ligand potency correctly.

Protein Kinase Database Survey. Since building homology models well enough for energy calculations is still very difficult, we also employed a quick and simple method that can still generate useful qualitative insights into the design of selective inhibitors. The 15 active site amino acids in CDK2 that have side chains facing the interior of the binding pocket were tested for conservation at their respective positions against almost 400 aligned protein kinases (Table 3) from the Protein

Kinase Resource.[24,25] Seven of these residues are more than 80% conserved at their positions, but seven are less than 40% conserved, giving hope for the design of selective inhibitors. The four least conserved amino acids in the CDK2 active site are Lys89 (8.3%), Phe80 (17.2%), Glu12 (20.3%), and Phe82 (21.1%). As discussed earlier in sensitivity analysis and ligand design, there are also affinity benefits in targeting some of the less conserved residues. The shape of the CDK2 active site allowed us to model functional groups in the designed structures to increase contact with residues such as Lys89. Although such a database survey does not quantify selectivity on an energetic basis, it provides an efficient way to guide ligand design by elucidating obvious targets for selectivity.

If one only focuses on comparing CDK1 and CDK2, their binding pockets are quite similar. The amino acids that have side chains facing the binding pocket are the same in both proteins. Selective binding to these two proteins will need to come from longer-range interactions or/and differences in other parts of the proteins that propagate into the binding pockets. CDK2 and CDK4 are more different. Among the amino acids that have side chains facing the binding pocket, three of them are different between the two proteins. They are Glu12, Phe82, and Lys89 in CDK2 and are respectively replaced by Val, His, and Thr in CDK4. These three sites are expected to play an important role in accounting for the selective binding of the two classes of compounds to CDK2 and CDK4.

Conclusions

As a test of our computational model of protein−ligand binding, the data generated in this project suggest good potential for implementation in drug design. When an actual crystal structure serves as the basis for input, useful correlation between biological and computational data results. Additionally, our data show that the hybrid homology modeling technique that we applied is only feasible for generating qualitative ranking of inhibitors. However, coupled with comparative analysis of the binding pockets of a large number of protein kinases, useful rules and guidelines can be derived to improve the odds of finding selective inhibitors.

Throughout this project, a great deal of consideration was given to the limitations of the computational method. In particular, the nature of a fixed conformation model can have a number of consequences, especially in ligand modeling. Without taking into account an ensemble of structures accessible at the relevant temperatures, energy minimizations were performed only on relevant groups to improve the chance of cancellation of errors in comparing energy differences. On a related note, one may also notice that we were more successful in modeling the flavopiridol mimics than the thio- and oxoflavopiridols. This result was not surprising because there is greater ambiguity in the exact coordinates of the latter group. Although energy minimization and a clear picture of the active site dimensions relieve this uncertainty to some degree, we acknowledge that this group of compounds presents a bigger challenge to our model. Nonetheless, the correlation with biological data is still encouraging, especially considering the resolution of potency demanded by the IC_{50} values (in both groups of compounds). The benefits of these approximations with respect to efficiency are also significant. The major computational tasks here include minimizing the energy of small groups (a matter of 1−2 min), calculating atomic charges with Gaussian 98 (35−60 min, can be reduced by using faster methods such as AM1-BCC[21]), and running UHBD (12−15 min for the ligand, 20 min for the binary system). Manual preparation of the necessary input data and files along with establishment of the initial coordinates for a ligand can be accomlished in tens of minutes but can be automated by suitable scripting/programming. In addition, sensitivity analysis helps to suggest where and how a lead compound should be modified so that one can come up with promising structures with just a small number of trials.

The encouraging results of the evaluative aspects of this project led us to attempt inhibitor design. We have come up with several structures that may bind significantly more strongly to CDK2 than their parent compounds. It will be useful to evaluate these compounds in the laboratory.

Acknowledgment. This work has been supported in part by the NIH, NSF, Howard Hughes Medical Institute, Accelrys, Inc., National Biomedical Computation Resource, Center for Theoretical Biological Physics at UCSD, and the W. M. Keck Foundation. We thank Professor Sung Hou Kim for providing the coordinates of the CDK2-L868276 complex.

References

(1) Sielecki, T. M.; Boylan, J. F.; Benfield, P. A.; Trainor, G. L. Cyclin-dependent kinase inhibitors: useful targets in cell cycle regulation. *J. Med. Chem.* **2000**, *43*, 1−18.

(2) Losiewicz, M. D.; Carlson, B. A.; Kaur, G.; Sausville, E. A.; Worland, P. J. Potent inhibition of CDC2 kinase activity by the flavonoid L86-8275. *Biochem. Biophys. Res. Commun.* **1994**, *201*, 589−595.

(3) Fischer, P. M.; Lane, D. P. Inhibitors of cyclin-dependent kinases as anti-cancer therapeutics. *Curr. Med. Chem.* **2000**, *7*, 1213−1245.

(4) de Azevedo, W. F., Jr.; Canduri, F.; da Silveira, N. J. Structural basis for inhibition of cyclin-dependent kinase 9 by flavopiridol. *Biochem. Biophys. Res. Commun.* **2002**, *293*, 566−571.

(5) De Azevedo, W. F., Jr.; Mueller-Dieckmann, H. J.; Schulze-Gahmen, U.; Worland, P. J.; Sausville, E.; Kim, S. H. Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2735−2740.

(6) Wong, C. F.; Hünenberger, P. H.; Akamine, P.; Narayana, N.; Diller, T.; McCammon, J. A.; Taylor, S.; Xuong, N. H. Computational analysis of PKA−balanol interactions. *J. Med. Chem.* **2001**, *44*, 1530−1539.

(7) Gould, C.; Wong, C. F. Designing specific protein kinase inhibitors: insights from computer simulations and comparative sequence/structure analysis. *Pharmacol. Ther.* **2002**, *93*, 169−178.

(8) Wong, C. F.; Thacher, T.; Rabitz, H. Sensitivity analysis in biomolecular simulation. *Reviews in Computational Chemistry*; Wiley-VCH: New York, 1998; pp 281−326.

(9) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; JosephMcCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; WiorkiewiczKuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phy. Chem. B* **1998**, *102*, 3586−3616.

(10) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(11) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(12) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(13) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. Electrostatics and diffusion of molecules in solution—simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **1995**, *91*, 57−95.

(14) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Electrostatics and diffusion of molecules in solution—simulations with the University-of-Houston-Brownian Dynamics Program. *Comput. Phys. Commun.* **1991**, *62*, 187−197.

(15) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(16) *Quanta*; Accelrys Inc.: San Diego, CA, 2000.

(17) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. Affinities of amino acid side chains for solvent water. *Biochemistry* **1981**, *20*, 849−855.

(18) Franck, P. *Introduction to system sensitivity theory*; Academic Press: New York, 1987.

(19) Tomovick, R.; Vukobratovic, M. *General sensitivity theory*; American Elsevier: New York, 1972.

(20) *InsightII*; Accelrys Inc.: San Diego, CA, 2000.

(21) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(22) Lee, M. G.; Nurse, P. Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature* **1987**, *327*, 31−35.

(23) Hanks, S. K. Homology probing: identification of cDNA clones encoding members of the protein−serine kinase family. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 388−392.

(24) Smith, C. M. The protein kinase resource and other bioinformation resources. *Prog. Biophys. Mol. Biol.* **1999**, *71*, 525−533.

(25) Smith, C. M.; Shindyalov, I. N.; Veretnik, S.; Gribskov, M.; Taylor, S. S.; Ten Eyck, L. F.; Bourne, P. E. The protein kinase resource. *Trends Biochem. Sci.* **1997**, *22*, 444−446.

(26) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.

(27) Hanks, S.; Quinn, A. M. Protein kinase catalytic domain sequence database: Identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **1991**, *200*, 38−62.

(28) Schoepfer, J.; Fretz, H.; Chaudhuri, B.; Muller, L.; Seeber, E.; Meijer, L.; Lozach, O.; Vangrevelinghe, E.; Furet, P. Structure-based design and synthesis of 2-benzylidene-benzofuran-3-ones as flavopiridol mimics. *J. Med. Chem.* **2002**, *45*, 1741−1747.

(29) Kim, K. S.; Sack, J. S.; Tokarski, J. S.; Qian, L.; Chao, S. T.; Leith, L.; Kelly, Y. F.; Misra, R. N.; Hunt, J. T.; Kimball, S. D.; Humphreys, W. G.; Wautlet, B. S.; Mulheron, J. G.; Webster, K. R. Thio- and oxoflavopiridols, cyclin-dependent kinase 1-selective inhibitors: synthesis and biological effects. *J. Med. Chem.* **2000**, *43*, 4126−4134.

(30) Traxler, P. M. Protein tyrosine kinase inhibitors in cancer treatment. *Expert Opin. Ther. Pat.* **1997**, *7*, 571−588.

(31) Murthi, K. K.; Dubay, M.; McClure, C.; Brizuela, L.; Boisclair, M. D.; Worland, P. J.; Mansuri, M. M.; Pal, K. Structure−activity relationship studies of flavopiridol analogues. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1037−1041.

(32) 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*, 243−256.

(33) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(34) Gidofalvi, G.; Wong, C. F.; McCammon, J. A. Entropy loss of hydroxyl groups of balanol upon binding to protein kinase A. *J. Chem. Educ.* **2002**, *79*, 1122−1126.

JM0205043