General Model for Estimation of the Inhibition of Protein Kinases Using Monte Carlo Simulations

Yukio Tominaga and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Monte Carlo statistical mechanics simulations were used in combination with the extended linear response (ELR) approach to develop a model to predict the activities of kinase inhibitors. One hundred forty eight inhibitors of three protein kinases, cyclin-dependent kinase 2 (CDK2), lymphocyte-specific kinase (Lck), and p38 mitogen-activated protein kinase were considered. The inhibitor sets for the individual kinases were analyzed first, and ELR models using only three descriptors were obtained with correlation coefficients, r^2, of 0.7−0.8. Models for each pair of kinases were then developed and used to predict the activities of the inhibitors for the remaining kinase with resultant q^2 values of 0.71 (CDK2), 0.70 (Lck), and 0.54 (p38). Finally, the three datasets were combined to yield a general ELR model for kinase inhibition; with just three physically reasonable descriptors, EXX, ΔHB_{total}, and $\Delta SASA$, the r^2 and leave-one-out q^2 are 0.69 and 0.67. The optimization of the model was confirmed using a genetic algorithm. The descriptors reflect the structural requirements for strong inhibition: good steric and electrostatic complementarities between inhibitor and protein, limited loss of hydrogen bonds for the inhibitor upon binding, and increased burial of surface area of the inhibitor.

Introduction

Rapid, accurate estimation of binding affinities for protein−ligand complexes is crucial for structure-based drug design. If a valid correlation between calculated binding affinities for protein−ligand complexes and experimental binding or activity data exists, the predicted binding affinities of proposed compounds could be used in prioritizing their synthesis. Although many approaches, ranging from rapid empirical scoring methods[1−3] to computationally intensive free energy perturbation (FEP) calculations,[4,5] have been pursued, a fully satisfactory solution has yet to emerge.

Among the alternatives, the linear response (LR) approach that was originally introduced by Åqvist et al. is less computationally intensive than FEP calculations and provided promising initial results.[6] In the original method,[6] the binding free energy (ΔG_b) for a protein−ligand complex was calculated using the difference in the average interaction energies between the unbound ligand with water and the bound ligand with protein and water. In eq 1, the ensemble averages, which are normally calculated using molecular dynamics or Monte Carlo (MC) simulations, are for the difference in van der Waals and Coulombic interaction

$$\Delta G_b = \alpha \langle \Delta E_{vdW} \rangle + \tfrac{1}{2} \langle \Delta E_c \rangle \qquad (1)$$

energies in the bound and unbound states (Figure 1). The coefficient α is an empirical scaling factor determined by fitting the simulation results to experimental free energies of binding.

The method was extended to include a third parameter for the change in the solvent-accessible-surface area (SASA), and the coefficient for the electrostatic term was allowed to deviate from 0.5 (eq 2).[7−9] Expanded studies of protein−ligand binding led to further generalization,

$$\Delta G_b = \alpha \langle \Delta E_{vdW} \rangle + \beta \langle \Delta E_c \rangle + \gamma \langle \Delta SASA \rangle \qquad (2)$$

as expressed in the extended linear response (ELR) expression, eq 3.[10−14] Here, c_n represents an optimizable

$$\Delta G_b = \sum_n c_n \xi_n + \text{constant} \qquad (3)$$

coefficient for the associated descriptor ξ_n, which represents a physically reasonable quantity relevant to protein−ligand binding. Some sensible descriptors, such as EXX-LJ, ΔHB_{total}, $\Delta FOSA$, ΔE_{int}, and #RB, emerged as the most significant in studies of the inhibition of HIV-1 reverse transcriptase (RT),[10,11] thrombin,[12] cyclo-oxygenase-2 (COX2),[13] and factor Xa.[14] EXX-LJ is the ligand−protein Lennard−Jones (van der Waals) interaction energy; ΔHB_{total} represents the change in the number of hydrogen bonds for the inhibitor upon binding; $\Delta FOSA$ is the change in the hydrophobic component of the solvent accessible surface area of the inhibitor upon binding; ΔE_{int} is the change in internal energy of the inhibitor upon binding; and #RB is the number of rotatable bonds for the inhibitor. Although these target proteins belong to several protein families and their inhibitors are structurally diverse, consensus is building for the general importance of at least EXX-LJ and ΔHB_{total}, which each appear in the regression equations for three of the four cases.

In search of a general ELR model for protein−ligand binding, the diversity of the training dataset needs expansion for both the ligand structures and protein targets. Recently, the former aspect received attention in an ELR study for HIV-1 RT, which included ca. 250 inhibitors with eight different core structures.[11] In the

* To whom correspondence should be addressed. Phone: 203 432−6278. Fax: 203 432−6299. E-mail: william.jorgensen@yale.edu.
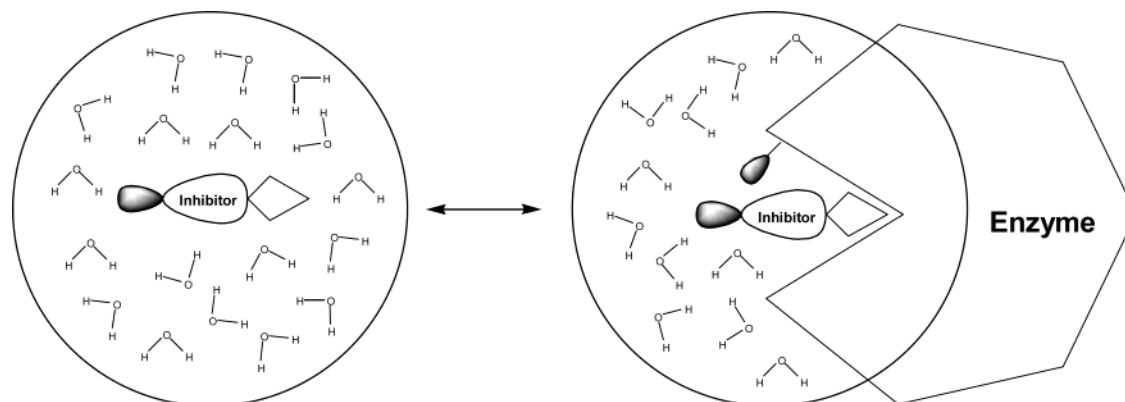
Figure 1. Schematic representation of protein–ligand binding and the environment of the ligand in unbound (left) and bound (right) states.

present study, attention is turned to increased protein diversity and the specific case of kinase selectivity. Protein kinases are important therapeutic targets, and many three-dimensional structures and data on biological activity are available.[15] Most kinase inhibitors compete with ATP for the ATP binding site. As numerous protein kinases show high sequence homology in this region,[16] selective inhibition of a single target kinase is challenging. We chose to study the following three representative systems: cyclin-dependent kinase 2 (CDK2), lymphocyte-specific kinase (Lck), and p38 mitogen-activated protein kinase (p38). The CDK2 and Lck inhibitors in this study indeed bind to the ATP-binding site. However, for additional diversity, the present p38 inhibitors were chosen because they bind to an allosteric site that is adjacent to the ATP-binding site; they interfere with the kinase activity by indirectly competing with ATP binding.

Cyclin-dependent kinases (CDKs) are members of the cdc2 family of serine-threonine protein kinases. They are important regulators that control the timing and coordination of progression of the cell cycle.[17–19] Transition between the four phases of the cell cycle is controlled by cyclins, CDKs, and CDK inhibitory proteins. Cyclin D and CDKs 4 and 6 control passage through the G1-phase; cyclin E and CDK2 control the G1- to S-phase transition; cyclin A and CDK2 regulate the passage through the S-phase; and cyclin B and CDK1 control the G2 checkpoint and regulate the entry into the M-phase. It has been shown that the aberrant regulation of CDKs and the consequential loss of the cell cycle checkpoint function directly link to the molecular pathology of cancer;[20–22] therefore, CDKs are believed to be attractive targets for the development of antitumor drugs.

Lck is a member of the nonreceptor Src family of tyrosine kinases and plays essential roles in immune response. Lck activates a number of substrates necessary for T cell antigen receptor signaling that leads to cytokine production and T cell activation.[23,24] The inhibition of Lck is potentially useful for the treatment of both chronic and acute T cell-mediated autoimmune and inflammatory disorders, such as rheumatoid arthritis, multiple sclerosis, transplant rejection, and delayed hypersensitivity reactions.[25]

p38 is a member of the family of mitogen-activated serine-threonine protein kinases. It is an important regulator of the signal transduction cascade leading to

the production of proinflammatory cytokines, such as tumor necrosis factor-α and interleukin-1β.[26] Although appropriate amounts of these cytokines play an important role in the host immune response, an excess of proinflammatory cytokines may cause chronic inflammatory diseases, such as rheumatoid arthritis, inflammatory bowel disease, and septic shock.[27,28] Therefore, the inhibition of p38 and the suppression of proinflammatory cytokines are potentially useful for the treatment of these diseases.

The goals of this study are to identify the important descriptors and optimal ELR models for reproducing kinase activities, to evaluate the transferability of the models between kinase systems, and to determine the best overall ELR model for kinase inhibition.

Methods

Dataset. The structure–activity data shown in Tables 1–10 were taken from the liturature.[29–33] Experimental IC_{50} values determined by assaying the kinase activity of CDK2[29,30] and Lck[32,33] were put on a free energy scale (ΔG_{expt}) by using eq 4;

$$\Delta G_{expt} \approx RT \ln IC_{50} \tag{4}$$

such IC_{50} data are expected to parallel relative free energies of binding in a closely related series of inhibitors.[34] For p38, experimental K_d values were reported[31] and are converted to free energies of binding via eq 5. Such data from different

$$\Delta G_{expt} = RT \ln K_d \tag{5}$$

sources cannot unequivocally be pooled owing to differences in the assays or assay conditions. This is addressed in the development of statistical models through the introduction of indicator variables that allow constant offsets for different datasets. In the following, the use of only one such indicator variable was found to be statistically significant.

Protein Structures. When modeling protein–ligand interactions quantitatively, special attention should be paid to the consistency between the protein structure and the conditions for its biological assay. To parallel the assays that yield IC_{50} data for kinase inhibition, it is important to use the structures of the activated enzymes.

The activation of CDK2 is a two-step process that requires cyclin A binding and phosphorylation of Thr160.[35–37] The structure of CDK2 phosphorylated on Thr160 is similar to that of inactive CDK2. However, upon association with cyclin A, CDK2 undergoes significant conformational changes in which active site residues are rearranged and the conformation of the ATP-binding site is influenced (Figure 2). Accordingly, the X-ray structure of the complex of cyclin A and CDK2 phos-
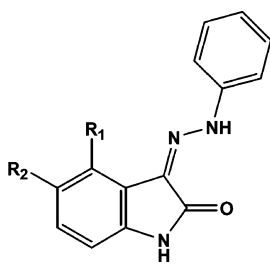
Figure 2. Structures of the inactive form of CDK2 (left) and that of the Thr160 phosphorylated CDK2-cyclinA complex (right). CDK2 is in blue, cyclin A in magenta, bound ATP in red, and Thr160 phosphate and activation segment (residues 145−172) in orange.

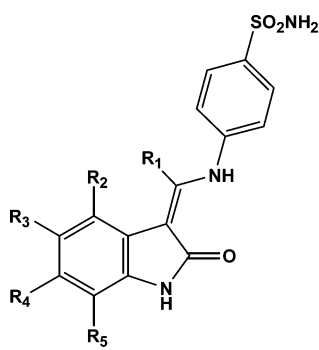**Table 1.** CDK2 Activities of Core 1 Derivatives

no.	R_1	R_2	R_3	$IC_{50}{}^a$ (nM)	$\Delta G_{expt}{}^b$
C1	H	H	H	120	−9.43
C2	I	H	H	4.6	−11.36
C3	CH_2CH_3	H	H	7.9	−11.04
C4	$CH(CH_3)_2$	H	H	2.5	−11.72
C5	$CH_2CH(CH_3)_2$	H	H	1.2	−12.16
C6	$CH{=}C(CH_3)_2$	H	H	1.5	−12.02
C7	OCH_2CH_3	H	H	93	−9.58
C8	$OCH(CH_3)_2$	H	H	3.4	−11.54
C9	OPh	H	H	13	−10.74
C10	$(CH_2)_2$-(4-pyridinyl)	H	H	21	−10.46
C11	$(CH_2)_2$-(4-phenol)	H	H	12	−10.79
C12	3-pyrazolyl	H	H	19	−10.52
C13	NO_2	H	H	2400	−7.66
C14	$CONH_2$	H	H	>1000	>−8.17
C15	H	F	H	34	−10.18
C16	H	Cl	H	43	−10.04
C17	H	Br	H	60	−9.84
C18	H	I	H	11	−10.84
C19	H	CH_3	H	46	−10.00
C20	H	OH	H	10	−10.90
C21	H	OCH_3	H	12	−10.79
C22	H	NO_2	H	15	−10.66
C23	H	NH_2	H	74	−9.72
C24	H	SO_2CH_3	H	16	−10.62
C25	H	SO_2NH_2	H	43	−10.04
C26	H	$CONH_2$	H	4.5	−11.37
C27	H	$CON(CH_3)_2$	H	17	−10.59
C28	H	$CONHCH_2$-(4-pyridinyl)	H	8.9	−10.97
C29	H	$CONHCH_2$-(3-pyridinyl)	H	2.1	−11.82
C30	H	5-oxazolyl	H	2.3	−11.77
C31	H	H	Br	43	−10.04
C32	H	H	CH_2CH_3	21	−10.46
C33	H	H	$CH(CH_3)_2$	75	−9.71
C34	H	H	OPh	>10 000	>−6.81
C35	Cl	CH_3	H	13	−10.74
C36	Cl	OCH_3	H	54	−9.90
C37	CH_3	NO_2	H	4.6	−11.36

^{a} Reference 29. ^{b} Experimental free energies from $\Delta G_{expt} \approx RT \ln(IC_{50})$ in kcal/mol.

Table 2. CDK2 Activities of Core 2 Derivatives

no.	R_1	R_2	IC$_{50}$^{a}(nM)	ΔG_{expt}^{b}
C38	H	H	>10000	>−6.81
C39	H	OCH$_3$	>10000	>−6.81
C40	H	Br	>10 000	>−6.81

^{a} Reference 29. ^{b} Experimental free energies from $\Delta G_{expt} \approx RT$ ln(IC$_{50}$) in kcal/mol.

Table 3. CDK2 Activities of Core 3 Derivatives

no.	R_1	R_2	R_3	R_4	R_5	IC$_{50}$^{a} (nM)	ΔG_{expt}^{b}
C41	H	H	H	H	H	690	−8.39
C42	CH$_3$	H	H	H	H	360	−8.78
C43	CH$_3$	H	Cl	H	H	22	−10.43
C44	H	CH$_2$OH	H	H	H	54	−9.90
C45	H	H	N(CH$_3$)$_2$	H	H	310	−8.87
C46	H	H	COOCH$_3$	H	H	2.1	−11.82
C47	H	H	H	CH$_2$OH	H	61	−9.83
C48	H	H	H	H	CH$_3$	>10000	>−6.81
C49	H	H	Cl	H	CH$_3$	>10000	>−6.81

^{a} Reference 29. ^{b} Experimental free energies from $\Delta G_{expt} \approx RT$ ln(IC$_{50}$) in kcal/mol.

Table 4. CDK2 Activities of Core 4 Derivatives

no.	R_1	R_2	IC$_{50}$^{a} (nM)	ΔG_{expt}^{b}
C50	H	H	2.8	−11.65
C51	CH$_3$	H	5.6	−11.24
C52	CH$_3$	CH$_3$	4.6	−11.36
C53	CH$_2$CH$_2$OH	H	4.7	−11.35
C54	Ph	H	4.3	−11.40
C55	pyrid-2-yl	H	9.7	−10.92
C56	CH$_2$Ph	H	5.6	−11.24
C57	COCH$_3$	H	75	−9.71

^{a} Reference 29. ^{b} Experimental free energies from $\Delta G_{expt} \approx RT$ ln(IC$_{50}$) in kcal/mol.

Table 5. CDK2 Activities of Core 5 Derivatives

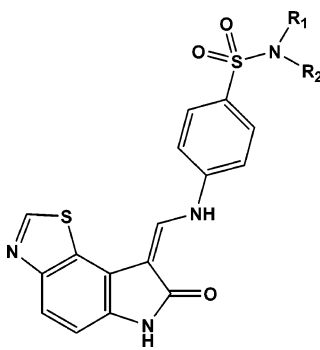| no.	R_1	R_2	IC$_{50}$^{a} (nM)	ΔG_{expt}^{b}
C58	H	H	1280	−8.36
C59	H	Cl	3970	−7.66
C60	NO$_2$	H	310	−9.23
C61	SO$_2$NH$_2$	H	30	−10.67

^{a} Reference 30. ^{b} Experimental free energies from $\Delta G_{expt} \approx RT$ ln(IC$_{50}$) in kcal/mol.

phorylated on Thr160 (1QMZ,[36] 2.2 Å) was selected as the starting point in the present studies.

Lck is activated by autophosphorylation of Tyr394 in the activation loop and is inactivated when Tyr505 near the C-terminus is phosphorylated and interacts with its own SH2 domain.[38,39] Accordingly, the X-ray structure of Lck phosphorylated on Tyr394 (1QPE,[40] 2.0 Å) was selected as the starting point in this case.
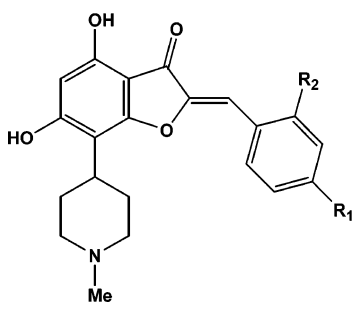
For p38, the X-ray structure (1KV1,[41] 2.5 Å) of the complex of p38 and **M1** was selected as an appropriate starting point.

System Preparation. The 10 core structures, as illustrated in Tables 1−10, were used to generate the protein−inhibitor complexes as well as the unbound inhibitors using the Biochemical and Organic Model Builder (*BOMB*) program.[42] This program was formerly called *GenMol*; the name has been changed to avoid confusion with another program with the same name. The core in Table 1 was initially positioned in the ATP-binding site of CDK2 in the manner observed in X-ray structures for oxindoles with CDK2,[29] then protein residues with all atoms beyond 15 Å from any atom of the core were removed. To avoid excessive fragmentation of the protein backbone, a few amino acids were added, and clipped residues were capped with acetyl and methylamine groups. All protein

residues beyond 10 Å from the core were kept rigid during the final MC simulations. The initial structures for the complexes with the cores for Tables 2−4 were then prepared by replacement of the first core in the ATP-binding site; the core for Table 5 was positioned as illustrated in ref 30 based on the X-ray structure for flavopiridol with CDK2. The initial complex for the Lck core (Table 10) was built in an analogous manner based on the analysis of X-ray and activity data in ref 32. As the X-ray structure of **M1** complexed with p38 was available, the structure of the **M1** complex was modified to correspond to the cores for Tables 6 to 8, and these structures were then processed in the same manner as for the CDK2 system. As there are some missing residues in the 1KV1 X-ray structure, all protein residues with all atoms beyond 12 Å from **M1** were removed, and all protein residues beyond 9 Å from it were kept rigid during the MC simulations. The final systems consisted of 115, 134, and 115 protein residues plus the inhibitors for CDK2, Lck, and p38, respectively. Specifically, for the CDK2 system, the rigid residues are 2A−6A, 15A−16A, 22A−27A, 45A−50A, 53A−54A, 56A−62A, 69A−70A, 94A, 108A−113A, 124A, 126A, 128A, 137A−141A, 149A, 166A−167A, 193A−194A, and 295A, and the flexible residues are 7A−14A, 17A−21A, 28A−35A, 51A−52A, 55A, 63A−68A, 77A−93A, 125A, 127A, 129A−136A, and 142A−148A. In the

Table 6. p38 Activities of Core 6 Derivatives
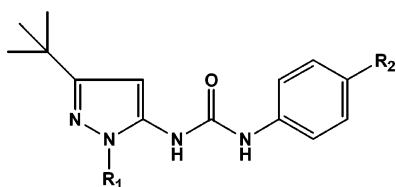
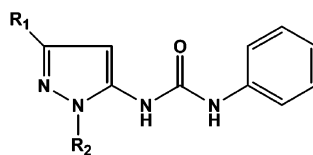| no.	R$_1$	R$_2$	$K_d{}^a$ (nM)	$\Delta G_{expt}{}^b$
M1	CH$_3$	Cl	350	−8.80
M2	phenyl	Cl	8	−11.03
M3	phenyl	H	13	−10.74
M4	cyclohexyl	H	500	−8.58
M5	2-methylphenyl	H	200	−9.13
M6	3-methylphenyl	H	2	−11.85
M7	4-methylphenyl	H	3	−11.61
M8	3,4-dimethylphenyl	H	4	−11.44
M9	2-naphthyl	H	8	−11.03
M10	3-aminophenyl	H	25	−10.36
M11	4-aminophenyl	H	7	−11.11
M12	3-methoxyphenyl	H	19	−10.52
M13	4-methoxyphenyl	H	31	−10.23
M14	4-pyridinyl	H	21	−10.46

a Reference 31. *b* Experimental free energies from $\Delta G_{expt} = RT \ln(K_d)$ in kcal/mol.

Lck system, the rigid residues are 237−248, 255−256, 263−268, 276−283, 293−299, 307−311, 328−332, 345−354, 361, 363−366, 375−376, 378, 386, and 430−431, and the flexible residues are 249−254, 257−262, 269−275, 284−292, 300−306, 312−327, 362, 367−374, 377, and 379−385. In the p38 system, the rigid residues are 19, 25−29, 32−34, 40, 56−57, 64−66, 69, 81, 89, 110−112, 135−136, 139, 143, 152−153, 184, 201−204, 206−211, 319, 321−322, 324−327, and 346−347, and the flexible residues are 30−31, 35−39, 50−55, 67−68, 70−80, 82−

88, 102−109, 134, 137−138, 140−142, 144−151, 154−157, 165−170, 205, 320, and 323. To enforce charge neutrality for the entire systems, some of the rigid Arg, Lys, Glu, and Arg residues were made neutral. The tautomeric states of His residues were assigned by visual inspection. Initial coordinates for the side chains of Gln277 and Gln309 for the Lck system, which were not determined by X-ray analysis, were generated by imposing dihedral angles from a standard rotamer library.

All protein residues and inhibitors were represented by the OPLS all-atom force field[43] with the exception that CM1A charges,[44] scaled by 1.08, were used for the atomic charges of the inhibitors. Each of the protein-core complexes for the CDK2 and Lck systems was subject to 50 steps of conjugate gradient energy minimization, using a distance-dependent dielectric constant of 4 ($\epsilon = 4r$), to relax the crystal structure and remove any arbitrariness of the manual docking. All molecular mechanics and MC calculations were performed with the *MCPRO* program.[45] Since the cores for p38 were not large enough to fill the binding site, energy minimization was not performed at this stage. Initial conformations of the bound ligands in Tables 1−10 were generated using *BOMB*, starting from the corresponding protein-core complexes. *BOMB* performs an extensive conformational search for the ligand in the binding site, optimizes the ligand's position, orientation, and dihedral angles, and saves the lowest energy complex for subsequent MC simulations. For the p38 complexes, conjugate-gradient energy minimization was carried out for each protein−ligand complex after it was built by *BOMB*. Finally, each ligand was extracted from the binding site to provide the initial structure of the free ligand.

MC Simulations. Using a 22 Å TIP4P water cap,[46] a half-harmonic potential with a 1.5 kcal mol^{-1} Å^2 force constant was applied to water molecules whose oxygen atoms might drift more than 22 Å from the center of the sphere. All side-chain bond and dihedral angles of protein residues with any atoms within ca. 10 Å (CDK2 and Lck) or 9 Å (p38) from the center of the water cap and all degrees of freedom of each inhibitor

**Table 7.** p38 Activities of Core 7 Derivatives

No.	R$_1$	R$_2$	$K_d{}^a$ (nM)	$\Delta G_{expt}{}^b$
M15		Phenyl	22	-10.43
M16		Phenyl	280	-8.93
M17	Methyl	Phenyl	>7,000	>-7.02
M18	Iso-propyl	Phenyl	330	-8.83
M19		Phenyl	12	-10.79
M20		Phenyl	1,400	-7.97
M21	Cyclohexyl	Phenyl	>400	>-8.72
M22		Phenyl	760	-8.34

a Reference 31. *b* Experimental free energies from $\Delta G_{expt} = RT \ln(K_d)$ in kcal/mol.

Table 8. p38 Activities of Core 8 Derivatives
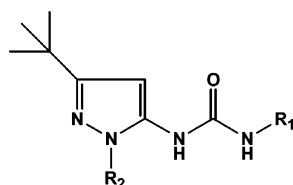
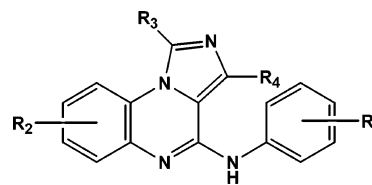| no.	R$_1$	R$_2$	K_d^a (nM)	ΔG_{expt}^b
M23	H	phenyl	>7000	>−7.02
M24	cyclohexyl	phenyl	130	−9.38
M25	2-pyridinyl	phenyl	400	−8.72
M26	3-pyridinyl	phenyl	420	−8.69
M27	4-pyridinyl	phenyl	1200	−8.07
M28	3-aminophenyl	phenyl	100	−9.54
M29	4-aminophenyl	phenyl	318	−8.85
M30	2,3-dimethylphenyl	phenyl	3	−11.61
M31	CH$_2$-phenyl	phenyl	47	−9.98
M32	CH$_2$CH$_2$-phenyl	phenyl	260	−8.97
M33	1-naphthyl	phenyl	5	−11.3
M34	2-indanyl	phenyl	14	−10.70
M35	2-fluorophenyl	4-methylphenyl	14	−10.70
M36	1-naphthyl	4-methylphenyl	1	−12.26

^{a} Reference 31. ^{b} Experimental free energies from $\Delta G_{expt} = RT \ln(K_d)$ in kcal/mol.

Table 9. Lck Activities of Core 9 Derivatives
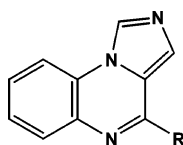
no.	R	IC_{50}^a (nM)	ΔG_{expt}^b
L1	cyclohexyl	3170	−7.49
L2	cyclopropyl	3170	−7.49
L3	*n*-propyl	2680	−7.59
L4	3-pyridinyl	820	−8.29

^{a} Reference 32. ^{b} Experimental free energies from $\Delta G_{expt} \approx RT \ln(IC_{50})$ in kcal/mol.

were sampled using the Metropolis procedure, while the protein backbone was held fixed after the initial energy minimizations. Bond lengths for the protein were not varied. A protein residue−inhibitor list, which was kept constant during the MC simulations, was determined for each complex during the initial solvent equilibration stage. A MC move of a side chain was randomly attempted every 10 configurations, while a move for the inhibitor was attempted every 56 configurations, and all remaining moves were for solvent molecules. The maximum number of internal coordinates to be varied for an attempted move was limited to 30. Residue-based cutoffs at 9 Å were used for the solvent−solvent, solute−solvent, and intrasolute nonbonded interactions. The MC simulations for the protein-inhibitor complexes of CDK2 and Lck consisted of 10 million configurations each for solvent-only equilibration, full equilibration, and averaging. Due to slower convergence of the electrostatic energy for the p38 system, the full equilibration period was doubled to 20 million configurations. As described in earlier work,[10] an annealing protocol was applied to the unbound MC simulations in order to decrease the dependency of starting coordinates to simulation results. Specifically, 10 million configurations of solvent-only equilibration were performed, followed by 5 million configurations in which only the water and dihedral angles for each inhibitor were sampled. The MC acceptance rates for the moves of the inhibitors were enhanced at this stage by increasing the temperature to 727 °C (1000 K). Then, 5 million configurations of full equilibration, followed by 10 million configurations of averaging, were performed at the normal temperature (25 or 37 °C depending on the biological assay

Table 10. Lck Activities of Core 10 Derivatives

no.	R$_1$	R$_2$	R$_3$	R$_4$	IC_{50}^a (nM)	ΔG_{expt}^b
L5	H	H	H	H	890	−8.24
L6	2-F	H	H	H	390	−8.73
L7	3-F	H	H	H	1330	−8.00
L8	2-Cl	H	H	H	60	−9.84
L9	2-Br	H	H	H	170	−9.22
L10	2-OMe	H	H	H	5520	−7.16
L11	2-Cl,4-Me	H	H	H	240	−9.02
L12	2-Cl,4,6-di-Me	H	H	H	30	−10.25
L13	2,4,6-tri-Me	H	H	H	40	−10.08
L14	2,6-di-Me	H	H	H	16	−10.62
L15	2,6-di-Br	H	H	H	50	−9.95
L16	2,6-di-Cl	H	H	H	9	−10.96
L17	2,6-di-F	H	H	H	360	−8.78
L18	2-Cl,6-Me	H	H	H	9	−10.96
L19	2,6-di-Et	H	H	H	1690	−7.86
L20	2-Cl	H	H	Me	30	−10.25
L21	2-Br	H	Me		180	−9.19
L22	2-Cl,6-Me	7,8-di-MeO	CH$_2$OH	H	6.2	−11.18
L23	2-Cl,6-Me	7,8-di-MeO	CHO	H	10	−10.90
L24	2,6-di-Me	7,8-di-MeO	H	H	2.4	−11.74
L25	2-Cl,6-Me	7,8-di-MeO	H	H	2	−11.85
L26	2,6-di-Me	6,7-di-MeO	H	H	2	−11.85
L27	2,6-di-Me	6,7-di-MeO	H	H	2.4	−11.74
L28	2-Cl,6-Me	6,7-di-OH	H	H	4	−11.44
L29	2-Cl,6-Me	6-MeO	H	H	3	−11.61
L30	2-Cl,6-Me	7-MeO	H	H	8.7	−10.98
L31	2-Cl,6-Me	8-MeO	H	H	280	−8.93
L32	2-Cl,6-Me	5-MeO	H	H	9.4	−10.94
L33	2,6-di-Me	5-NO$_2$	H	H	100	−9.54
L34	2,6-di-Me	5-NH$_2$	H	H	70	−9.75
L35	2-Cl,6-Me	6-F	H	H	26	−10.33
L36	2-Cl,6-Me	6-Br	H	H	15	−10.66
L37	2-Cl,6-Me	6-COOMe	H	H	26	−10.33
L38	2,6-di-Cl	6-NO$_2$	H	H	24	−10.38
L39	2-Cl,6-Me	6-CN	H	H	100	−9.54
L40	2-Cl,6-Me	6-NH$_2$	H	H	7	−11.11
L41	2-Cl,6-Me	6-NHAc	H	H	3	−11.61
L42	2-Cl,6-Me	7-Br	H	H	14	−10.70
L43	2-Cl,6-Me	7-NH$_2$	H	H	21	−10.46
L44	2-Cl,6-Me	7-NHAc	H	H	11	−10.84
L45	2-Cl,6-Me	7-CONH$_2$	H	H	30	−10.25
L46	2-Cl,6-Me	6-NMe$_2$	H	H	5	−11.31
L47	2-Cl,6-Me	6-NEt$_2$	H	H	2	−11.85
L48	2-Cl,6-Me	6-NHEt	H	H	10	−10.90
L49	2-Cl,6-Me	6-morpholinyl	H	H	4	−11.44
L50	2-Cl,6-Me	7-NEt$_2$	H	H	9	−10.96
L51	2-Cl,6-Me	7-morpholinyl	H	H	6	−11.20

^{a} Refs 32 and 33. ^{b} Experimental free energies from $\Delta G_{expt} \approx RT \ln(IC_{50})$ in kcal/mol.

conditions). The final three steps (heating, equilibration, and averaging) were repeated for five cycles.

ELR Descriptors. Key quantities that were averaged for the ELR analysis during the MC simulations are summarized in Table 11. The SASA was determined with a probe 1.4 Å in radius. The solute atomic radii are computed from the OPLS-AA Lennard−Jones sigmas via $r = 2^{1/6}\sigma/2$. For the hydrogen bond counts, a hydrogen bond was defined as present when the distance between an oxygen or nitrogen atom and an acidic hydrogen (O/N···HN, HO, HS) was equal to or less than 2.5 Å. Some properties of the inhibitors as estimated by the *QikProp* program[47] were also considered as possible descriptors, in particular, the octanol/water partition coefficient (QPlogPo/w) and the number of rotatable bonds (#RB).

Table 11. Principal Descriptors Considered in the ELR Analysis[a]

symbol	description
EXX-LJ	ligand−protein Lennard−Jones interaction energy
EXX-C	ligand−protein Coulombic interaction energy
ESX-LJ	ligand−water Lennard−Jones interaction energy
ESX-C	ligand−water Coulombic interaction energy
ΔE_{LJ}	EXX-LJ + ESX-LJ(bound) − ESX-LJ(unbound)
ΔE_C	EXX-C + ESX-C(bound) − ESX-C(unbound)
ΔE_{int}	change in internal energy for the ligand
ΔFOSA	change in hydrophobic surface area
ΔFISA	change in hydrophilic (N, O, H_N, H_O, H_S) surface area
ΔARSA	change in aromatic hydrocarbon surface area
ΔWPSA	change in weakly polar (P, S, halogens) surface area
ΔSASA	total change in SASA for the ligand
ΔHB_{total}	change in number of hydrogen bonds for the ligand
#RB	the number of rotatable bonds for the ligand
QPlogPo/w	octanol/water log P estimated by *QikProp*

[a] Changes Δ refer to the difference in the quantity for the ligand bound to the protein versus unbound in water.

Statistical Analysis. Multiple linear regression analysis was performed using the statistical software package JMP.[48] Descriptor sets were chosen so that the number of descriptors was minimized and the correlation coefficient, r^2, was maximized. The statistical significance of the descriptors was confirmed from the variance analysis using the F ratios (regression model mean square/error mean square), requiring that the probability of a greater F value occurring by chance (Prob > F) is less than 0.05. A cross-validated r^2 value, q^2, was obtained by the leave-one-out procedure.

Results

Binding Modes. MC simulations of 148 protein−ligand complexes and corresponding free ligands in TIP4P water were performed with *MCPRO*. Structures of the cores for Tables 1 and 10, and **M1**, and nearby binding-site residues in CDK2, Lck, and p38 are shown in Figures 3−5, respectively. Multiple images from the MC simulations are overlaid, and the structures are aligned in a consistent manner. It is recalled that the CDK2 and Lck inhibitors in this study bind in the ATP-binding site, while the p38 inhibitors bind to an allosteric site that is adjacent to the ATP site. Thus, the p38 inhibitor is shifted to the right in Figure 5. As illustrated, four, two, and three protein−ligand hydrogen bonds are present in the CDK2, Lck, and p38 systems. In addition to these normal hydrogen bonds, CH···O and OH···π interactions were observed in the Lck system (Figure 6). A substantial conformational change is observed in the X-ray structures for the conserved Asp-Phe-Gly segment (shown in cyan in Figures 3−5) for p38 compared to the other two structures.

Individual ELR Models. ELR analysis was carried out for the three individual kinase systems, and the following linear regression equations were obtained as optimal using three descriptors. Inclusion of additional descriptors was not significantly beneficial.

CDK2:

$$\Delta G_{calcd} = 0.100\langle EXX\text{-}C\rangle + 0.110\langle EXX\text{-}LJ\rangle - 0.216\langle\Delta HB_{total}\rangle - 1.350 \quad (5)$$

Lck:

$$\Delta G_{calcd} = 0.0989\langle EXX\text{-}C\rangle + 0.257\langle EXX\text{-}LJ\rangle - 0.320\langle\Delta HB_{total}\rangle + 0.623 \quad (6)$$

p38:

$$\Delta G_{calcd} = 0.0644\langle EXX\rangle + 0.00619\langle\Delta FOSA\rangle - 0.760(QPlogPo/w) - 0.636 \quad (7)$$

EXX-C, EXX-LJ, and EXX are the ligand−protein Coulomb and Lennard−Jones interaction energies and their sum, respectively. ΔHB_{total} represents the change in the total number of hydrogen bonds for the inhibitor upon binding (i.e., bound − unbound); it is normally a negative number or zero. ΔFOSA is the change in hydrophobic component of the solvent accessible surface area for the inhibitor upon binding. Using the three descriptors with the full datasets of 61, 51, and 36 compounds, the correlation coefficients, r^2, are 0.759, 0.734, and 0.678; the rms errors are 0.729, 0.676, and 0.767 kcal/mol; the mean unsigned errors are 0.602, 0.540 and 0.637 kcal/mol; and the predictive correlation coefficients from the leave-one-out procedure, q^2, are 0.722, 0.683, and 0.603, respectively. The maximum probability > F ratios for any descriptor in eqs 5, 6, and 7 is 0.0069, 0.0093, and 0.0122. Many of the ELR correlations in this paper have r^2 values near 0.7, so only a few typical plots of experimental vs calculated ΔG values are presented, e.g., Figure 7 illustrates the results from eq 5 for the CDK2 dataset.

The descriptors and the signs of their coefficients in eqs 5−7 make physical sense. All signs for EXX-C, EXX-LJ, and EXX in eqs 5−7 are positive, which reflects that good electrostatic and steric complementarities between the ligand and protein are essential for high activity (low IC_{50}, low ΔG). The coefficient of ΔHB_{total} represents the dehydration penalty and amounts to 0.216 and 0.320 kcal/mol for each hydrogen bond lost in the CDK2 and Lck systems. This coefficient has ranged from 0.22 to 1.65 in previous ELR models;[11−13] the present, relatively small values likely reflect that the loss of hydrogen bonds is also represented in the EXX-C terms here (vide infra). The second and third terms in eq 7 make the reasonable statement that removal of hydrophobic surface area upon binding and increased hydrophobicity for the inhibitor are also favorable. EXX-LJ and log $P_{o/w}$ are often well correlated until a ligand becomes too big to fit into a binding site.

If EXX-LJ and EXX-C in eqs 5 and 6 are combined into EXX and the data are refit, eqs 8 and 9 are obtained.

CDK2:
$$\Delta G_{calcd} = 0.0996\langle EXX\rangle - 0.209\langle\Delta HB_{total}\rangle - 1.78 \quad (8)$$

Lck:
$$\Delta G_{calcd} = 0.195\langle EXX\rangle - 0.319\langle\Delta HB_{total}\rangle - 0.677 \quad (9)$$

Using only two descriptors with the 61 and 51 compounds, the correlation coefficients, r^2, become 0.758 and 0.651; the rms error are 0.731 and 0.774 kcal/mol; the mean unsigned errors are 0.605 and 0.608 kcal/mol; and the leave-one-out q^2 values are 0.727 and 0.607, respectively. The maximum probability > F ratios for the descriptors in eqs 8 and 9 are 0.0074 and 0.0209. As the coefficients of EXX-LJ and EXX-C in eq 5 are similar, eq 8 shows negligible degradation in the fit, while there is modest degradation for eq 9 vs eq 6.
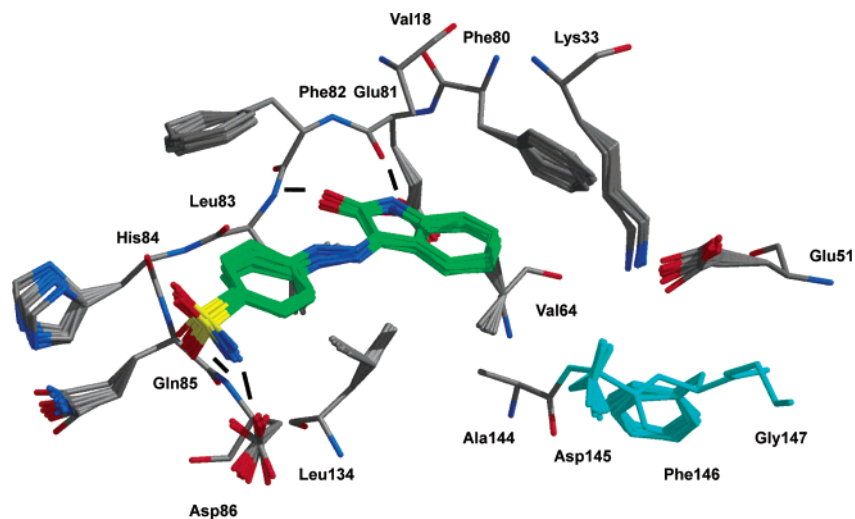
Figure 3. Representative MC configurations of the complex between CDK2 and the core from Table 1. Hydrogen bonds between the inhibitor and CDK2 are shown as black lines.
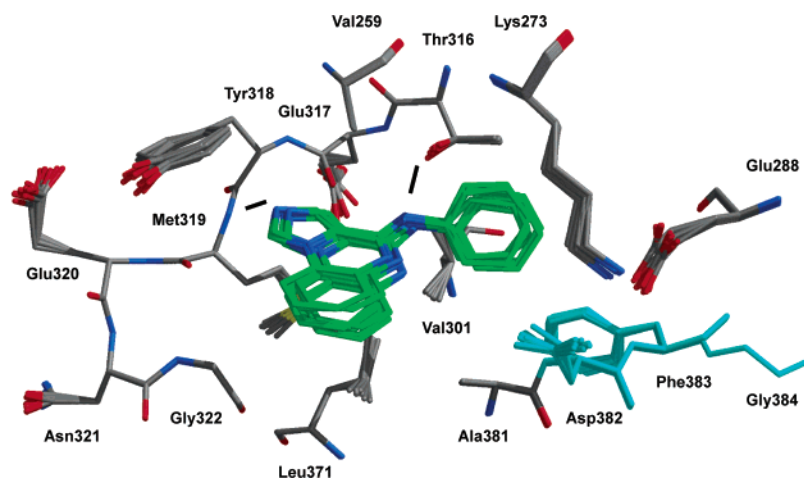

Figure 4. Representative MC configurations of the complex between Lck and the core from Table 10. Hydrogen bonds between the inhibitor and Lck are shown as black lines.
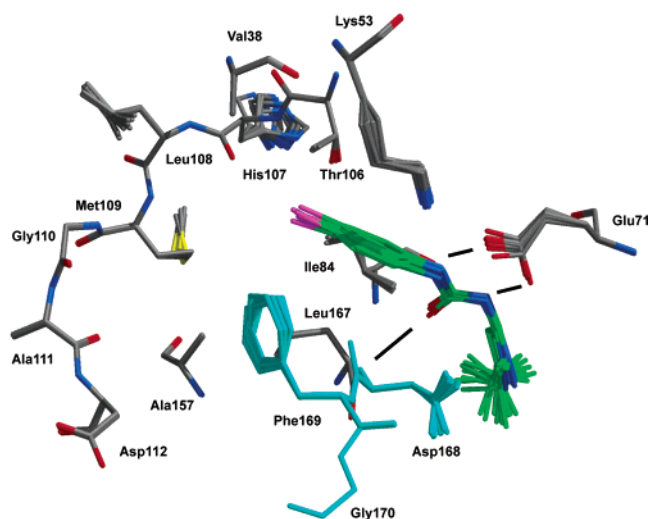

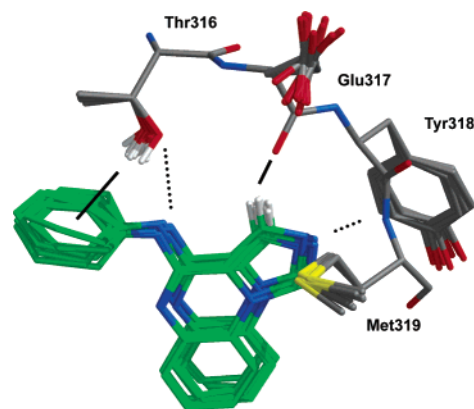**Figure 5.** Representative MC configurations of the **M1**−p38 complex. Hydrogen bonds between **M1** and p38 are shown as black lines.

Figure 6. Representative MC configurations for the complex of Lck and the core from Table 10. CH···O and OH···π interactions between the inhibitor and Lck are shown as black lines.

EXX or the combination of EXX-C and EXX-LJ is significant for all three systems, ΔHB_{total} is also important for both CDK2 and Lck, and $\Delta FOSA$ and QPlogPo/w are relevant for p38. Notably, the same descriptors are used for both CDK2 and Lck, where the target is the ATP binding site, though structurally different series of inhibitors were employed in the analyses. The hydrophobicity of the inhibitors appears to be a more distinct consideration for the p38 inhibitors, which bind to the less polar allosteric site.
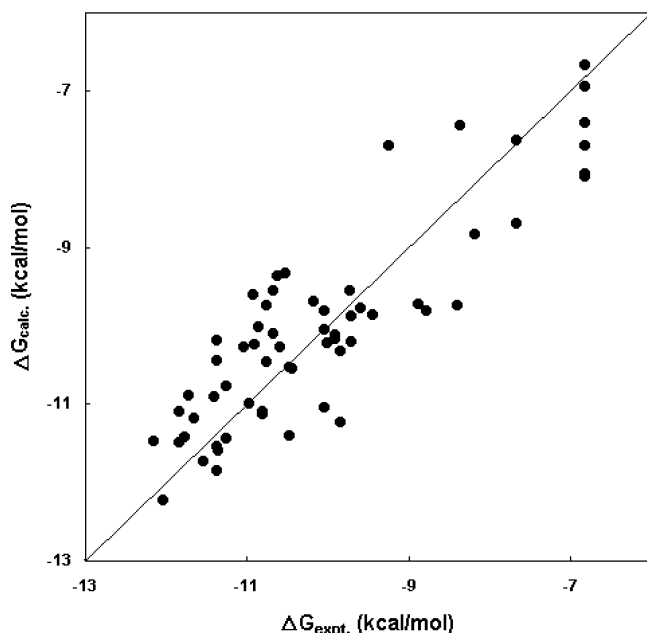
Figure 7. Plot of computed ΔG_{calcd} from eq 5 vs experimental activity data, ΔG_{expt}, for CDK2.

Figure 8. Plot of computed ΔG_{calcd} from eq 10 vs experimental activity data, ΔG_{expt}, for the combined CDK2 and Lck systems.

Using the ELR model of one system, the activities for the other two systems were then predicted. When the CDK2 model (eq 5) is applied to the Lck and p38 systems, the correlation coefficients, q^2, between the predicted and experimental ΔGs are 0.666 and 0.333. When the Lck model (eq 6) is applied to the CDK2 and p38 systems, the q^2 values are 0.538 and 0.356. When the p38 model (eq 7) is applied to the CDK2 and Lck systems, the q^2 values are 0.214 and 0.504. Thus, not surprisingly, the CDK2 and Lck models do reasonably well for either protein, but they do poorly for the p38 inhibitors. The p38 model does poorly in predictions for the CDK2 dataset, but performs better for Lck. The latter result may be fortuitous or it may reflect greater geometrical overlap of the p38 and Lck inhibitors in the binding sites (Figures 3−5).

Combination of Two Datasets To Predict the Third. The next step was to ascertain the quality of predictions that could be made for a new kinase target using models obtained from combination of the results for two other systems. Thus, each of the three kinases was treated in turn as the new target. Since the CDK2 (eqs 5 and 8) and Lck (eqs 6 and 9) models used the same descriptors, their datasets were combined and the resultant regression equation kept the original descriptors. As the remaining two combinations, p38 (eq 7) and CDK2 (eq 8), and p38 and Lck (eq 9), had EXX in common, EXX was retained in the combined models. During the fitting process for the CDK2-p38 and Lck-p38 combinations, p38 inhibitors **M6** and **M7** consistently appeared as outliers. The experimental data do not appear to be anomalous for them, so it possible that the initial structures that were generated in these two cases are not optimal and were not relaxed adequately during the simulations. We have chosen to enforce a standard computional protocol and to not process them in any special manner. After removing them as outliers, the following optimal regression equations were obtained for the three pairs of combined datasets.
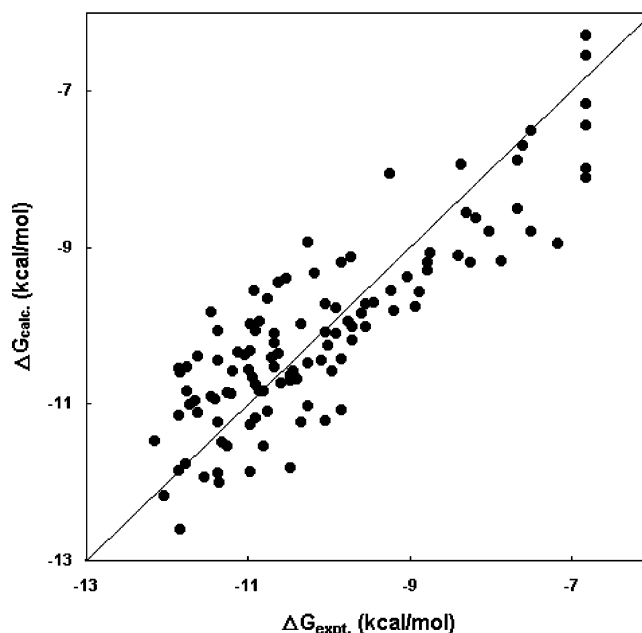
CDK2 and Lck:

$$\Delta G_{calcd} = 0.0995\langle \text{EXX} \rangle - 0.269\langle \Delta\text{HB}_{total} \rangle + 0.0103\langle \Delta\text{SASA} \rangle - 3.75 \ (L_{corr}) + 3.45 \quad (10)$$

CDK2 and p38:

$$\Delta G_{calcd} = 0.0814\langle \text{EXX} \rangle - 0.289\langle \Delta\text{HB}_{total} \rangle + 0.0110\langle \Delta\text{SASA} \rangle + 2.06 \quad (11)$$

p38 and Lck:

$$\Delta G_{calcd} = 0.0723\langle \text{EXX} \rangle - 0.312\langle \Delta\text{HB}_{total} \rangle + 0.0147\langle \Delta\text{SASA} \rangle - 3.03 \ (L_{corr}) + 3.55 \quad (12)$$

L_{corr} is an indicator variable with value 1 for an Lck inhibitor, and otherwise 0. It is presumably making a systematic adjustment for differences in the assay conditions or nature of the measurements, IC$_{50}$ or K_d. An indicator variable may have also been needed for eq 11, but was not significant. ΔSASA is the change in total solvent accessible surface area of the inhibitor upon binding. Using four, three, and four descriptors with 112, 95, and 85 compounds, the correlation coefficients, r^2, are 0.737, 0.692, and 0.649; the rms errors are 0.722, 0.795, and 0.784 kcal/mol; the mean unsigned errors are 0.598, 0.676, and 0.654 kcal/mol; and the leave-one-out q^2 values are 0.712, 0.661, and 0.602, respectively. The maximum probability $> F$ ratios for the descriptors in eqs 10, 11, and 12 are 0.0001, 0.0001, and 0.0026, respectively. The combined results for the CDK2 and Lck systems from eq 10 are illustrated in Figure 8.

All three ELR models share the same descriptors, EXX, ΔHB$_{total}$, ΔSASA, but there is some variation in the coefficients. The descriptors and signs of the coefficients are physically reasonable. Good electrostatic and steric complementarities between the inhibitor and protein are important for activity, as usual, and the penalty for the loss of each hydrogen bond is consistently near 0.3 kcal/mol. Increased burial of ligand surface
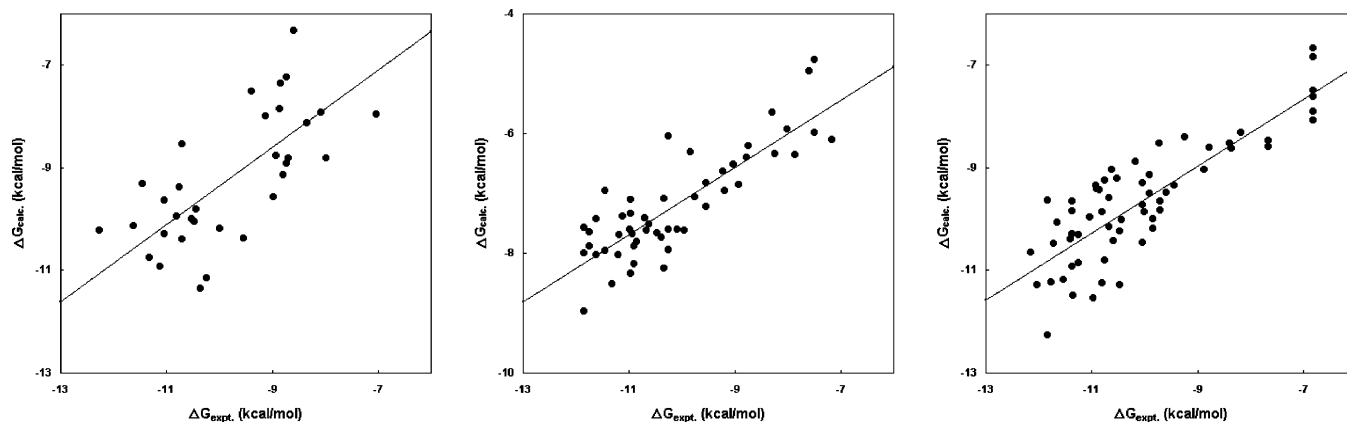
Figure 9. Data for two kinases provides a model to predict activities for the third kinase. Plots of computed ΔG_{calcd} from eqs 10−12 vs experimental activity data, ΔG_{expt}, for p38 (left), Lck (center), and CDK2 (right).

area upon binding also enhances activity. Though the contribution of the ΔSASA term is often nearly constant for a small dataset,[9] its range and significance may increase as the size and diversity of the dataset increase.

　Using the combined ELR models, the activities for the remaining system were predicted. When eqs 10, 11, and 12 are applied to the p38, Lck, and CDK2 systems, the correlation coefficient, q^2, between predicted and experimental ΔGs is 0.538, 0.704 and 0.706, respectively. The predictability of the combined ELR models improved nicely from that of the individual ELR models, which yielded an average q^2 of 0.435. Figure 9 illustrates the results and typical patterns that may be expected if the combined models are applied to predict activities of a set of proposed inhibitors for a new kinase target. All combined models predict well the activity trends for the remaining system. The predictions for the p38 inhibitors are the least good since they are based on the datasets for the two systems which feature ATP-site inhibitors. However, inclusion of the p38 dataset with either the CDK2 or Lck data still allows good prediction for the other ATP-site case. The best predictive model is expected to arise from combination of all three datasets.

　Final ELR Model Using All Three Datasets. As the three combined models used the same descriptors, a final regression equation was sought using these descriptors and all three datasets. The regression analysis was carried out for all 148 compounds, and again the two outliers from the p38 system were detected (**M6** and **M7**). With their exclusion, eq 13 was obtained.

$$\Delta G_{calcd} = 0.0848\langle EXX \rangle - 0.293\langle \Delta HB_{total} \rangle + 0.0123\langle \Delta SASA \rangle - 3.11(L_{corr}) + 3.08 \quad (13)$$

Using four descriptors and 146 compounds, the correlation coefficient, r^2, is 0.691, the rms error is 0.775 kcal/mol, the mean unsigned error is 0.647 kcal/mol, and the leave-one-out q^2 is 0.666 (Figure 10). The probability > F ratio of each descriptor is <0.0001. When EXX is separated into EXX-C and EXX-LJ, eq 14 is obtained; however, there is no change in the statistics.

$$\Delta G_{calcd} = 0.0845\langle EXX\text{-}C \rangle + 0.0771\langle EXX\text{-}LJ \rangle - 0.294\langle \Delta HB_{total} \rangle + 0.0129\langle \Delta SASA \rangle - 3.11(L_{corr}) + 3.08 \quad (14)$$
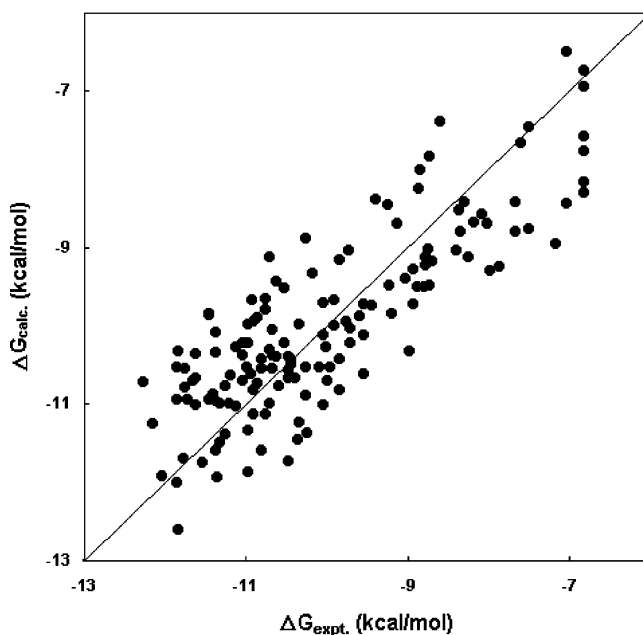

Figure 10. Plot of computed ΔG_{calcd} from eq 13 vs experimental activity data, ΔG_{expt}, for all three kinases.

　The free energies associated with EXX-C, EXX-LJ, ΔHB_{total}, $\Delta SASA$, and L_{corr} cover ranges of 6.95, 2.34, 2.32, 4.13, and 3.11 kcal/mol, respectively, so each term makes a significant contribution to the activity predictions. To evaluate the relative contribution of each descriptor more clearly, all descriptors and activities were mean centered and scaled by their standard deviation. The scaled data were then refit and yielded eq 15

$$\Delta G_{calcd} = 1.30\langle \underline{EXX\text{-}C} \rangle + 0.323\langle \underline{EXX\text{-}LJ} \rangle - 0.378\langle \underline{\Delta HB_{total}} \rangle + 0.508\langle \underline{\Delta SASA} \rangle - 1.06(\underline{L_{corr}}) \quad (15)$$

where the underlining means the descriptors are scaled. Thus, the significance of the contributions to ΔG_{calcd} decrease in the following order: EXX-C, L_{corr}, $\Delta SASA$, ΔHB_{total}, and EXX-LJ.

　Model Validation Using a Genetic Algorithm. The ELR model, eq 13, was developed via a build-up approach in which the dataset was gradually expanded and the fitting process was repeated. This approach may not lead to the globally optimal model. The best method to find the optimal model is an exhaustive one where

Table 12. Results of Model Optimizations Using a Genetic Algorithm

number of outliers		number of variables			
		2	3	4	5
0	q^2	0.347	0.566	0.627	0.637
	descriptors	EXX, L_{corr}	EXX, ΔSASA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, L_{corr}
	outliers	-	-	-	-
1	q^2	0.368	0.590	0.649	0.652
	descriptors	EXX, L_{corr}	EXX, ΔSASA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔE_{int}, L_{corr}
	outliers	**M6**	**M6**	**M6**	**M7**
2	q^2	0.387	0.607	0.666	0.675
	descriptors	EXX, L_{corr}	EXX, ΔSASA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔE_{int}, L_{corr}
	outliers	**M6, M7**	**M6, M7**	**M6, M7**	**M6, M7**
3	q^2	0.404	0.590	0.674	0.683
	descriptors	EXX, L_{corr}	EXX-C, ΔSASA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔE_{int}, L_{corr}
	outliers	**M6, M7, M8**	**M6, M7, L19**	**M6, M7, M35**	**M6, M7, M35**
4	q^2	0.422	0.630	0.681	0.690
	descriptors	EXX, L_{corr}	EXX, ΔSASA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔE_{int}, L_{corr}
	outliers	**M6, M7, M8, M36**	**M6, M7, M35, L19**	**M6, M7, M35, L28**	**M6, M7, M32, M35**
5	q^2	0.438	0.640	0.686	0.695
	descriptors	EXX, L_{corr}	EXX, ΔSASA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, L_{corr}
	outliers	**M6, M7, M8, M35, M36**	**M6, M7, M30, M32, M35**	**M6, M7, M10, M32, M35**	**M6, M7, M10, M13, L20**

number of outliers		number of variables		
		6	7	8
0	q^2	0.637	0.637	0.637
	descriptors	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, L_{corr}
	outliers	-	-	-
1	q^2	0.665	0.665	0.665
	descriptors	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}
	outliers	**M6**	**M6**	**M6**
2	q^2	0.682	0.682	0.682
	descriptors	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}
	outliers	**M6, M7**	**M6, M7**	**M6, M7**
3	q^2	0.690	0.690	0.690
	descriptors	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}
	outliers	**M6, M7, L20**	**M6, M7, L20**	**M6, M7, L20**
4	q^2	0.698	0.698	0.697
	descriptors	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}
	outliers	**M6, M7, M13, L28**	**M6, M7, L20, L28**	**M6, M7, M32, L20**
5	q^2	0.699	0.699	0.698
	descriptors	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, L_{corr}	EXX, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}	EXX, ΔE_{LJ}, ΔSASA, ΔHB$_{total}$, ΔARSA, ΔE_{int}, L_{corr}
	outliers	**M6, M7, M10, M13, L20**	**M6, M7, M13, L28, C12**	**M1, M6, M7, L18, L47**

all possible combinations of descriptors and outliers are considered in the fitting process. This tactic, however, is generally impractical, so we applied a genetic algoritm[49,50] (GA) to test the selection of variables and outliers and to ascertain if the final ELR model was near the global optimum. In this approach, the maximum numbers of descriptors and outliers are entered, and the best partial least-squares regression[51,52] (PLS) model is determined by optimization with the GA. The present 148 inhibitors, the independent descriptors from Table 11, and L_{corr} were considered. One thousand random combinations of variables (descriptors) and outliers were selected. Using the selected variables and eliminated outliers, PLS analysis was performed and q^2 values were evaluated. The 50 combinations of variables and outliers with the highest q^2 values were then selected as the initial population. One pair of combinations of variables and outliers was randomly selected from the initial population as parents. The parents were subjected to crossover, and each offspring was subjected to random single-point mutation. When the numbers of selected descriptors and outliers of the offspring were less than or equal to the maximum number entered initially, PLS analysis was performed and q^2 values were evaluated.

When the highest q^2 value of an offspring was superior to the lowest q^2 value of the initial population, the offspring was replaced with the parent and the initial population was explored. The selection, crossover, mutation, evaluation, and exploration processes were then repeated 5000 times.

The results are summarized in Table 12. As the number of descriptors increases from two to four, q^2 improves significantly and reaches a plateau around 0.69. When the number of descriptors is four, the descriptors used in eq 13 are also found in the best PLS models. When the number of outliers is allowed to be two or more, **M6** and **M7**, which were eliminated in the build-up approach, are always selected regardless of the number of descriptors. The descriptors and outliers used in eq 13 are also found in the best PLS model, when the maximum numbers of descriptors and outliers are set to four and two. Negligible improvement is obtained by increasing the number of descriptors or outliers. Thus, eq 13 with a q^2 of 0.666 is effectively optimal. The ratio of data points to descriptors (146/4) and the small number of outliers (2/148) are notable. The MC simulations were only carried out for these 148 inhibitors; the 148 were chosen primarily as all of the inhibitors in the
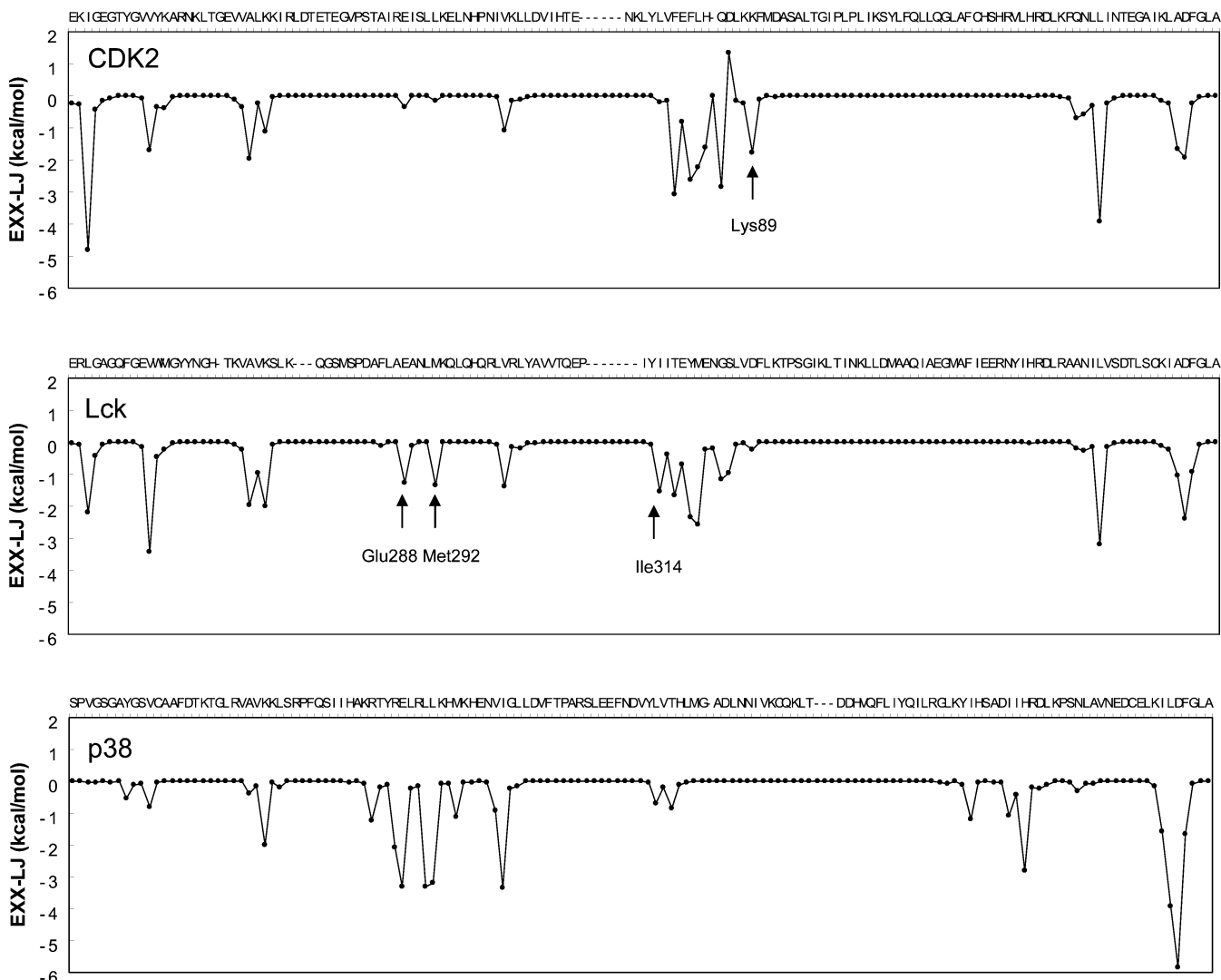
Figure 11. Residue-based EXX-LJ for CDK2, Lck, and p38.

experimental studies that could readily be built by *BOMB*, i.e., without adding additional substituents to its database.

L_{corr} **Indicator Variable.** L_{corr} was initially justified as an offset for variation in the biological assay conditions. However, the coefficient in eq 13 for L_{corr}, −3.11, seems somewhat large. An additional possibility is that it is reflecting other constant factors which contribute to improved activity and protein−ligand binding for the present Lck inhibitors. Thus, a more thorough structural analysis was carried out on configurations from the MC simulations focusing on the imidazoquinoxaline core, which the Lck inhibitors have in common (Tables 9 and 10).

As mentioned above, the imidazoquinoxaline core forms NH···N and CH···O hydrogen bonds with the backbone of Met 319 and Glu 317 (Figure 6). Such CH···O contacts are observed in many protein-kinase systems, especially in the ATP binding site.[53−55] According to ab initio HF/6-31G** calculations on model systems, the interaction energy of the CH···O hydrogen bond between the C4 proton of imidazole and oxygen of water is −2.16 kcal/mol, and the C4−O distance is 3.66 Å.[56] For comparison, at the same computational level, the OH···O interaction energy for the water dimer is −5.51 kcal/mol and the O−O separation is 2.99 Å. The

average C−O and N−N distances for the CH···O and NH···N hydrogen bonds in the 10 representative snapshots shown in Figure 6 are 3.49 and 3.02 Å, respectively. As CH···O interactions were not included in the hydrogen bond counts in this study, ΔHB_{total} may be too negative for the Lck system; however, based on the coefficient in eq 13, this would only amount to a correction of −0.293 kcal/mol. Another notable feature for the imidazoquinoxaline core is that it participates in one or two hydrogen bonds with water molecules that bridge to the side chain of Asp382. The direct interactions of the protein and inhibitor are represented in eq 13 along with a penalty for desolvation of the inhibitor. However, desolvation of the protein and related effects of water mediated protein−inhibitor hydrogen bonding are not explicitly considered. They may be relatively favorable for the present Lck inhibitors and contribute to the coefficient for L_{corr}.

Specificity for the ATP-Binding Site. Most kinase inhibitors bind to the ATP site. Successful drug candidates need to do this in a highly selective manner in order to avoid interference with kinases other than the target kinase. To gain insights into the binding-site residues that can contribute to selectivity for the present kinases, the protein−inhibitor interaction energies,
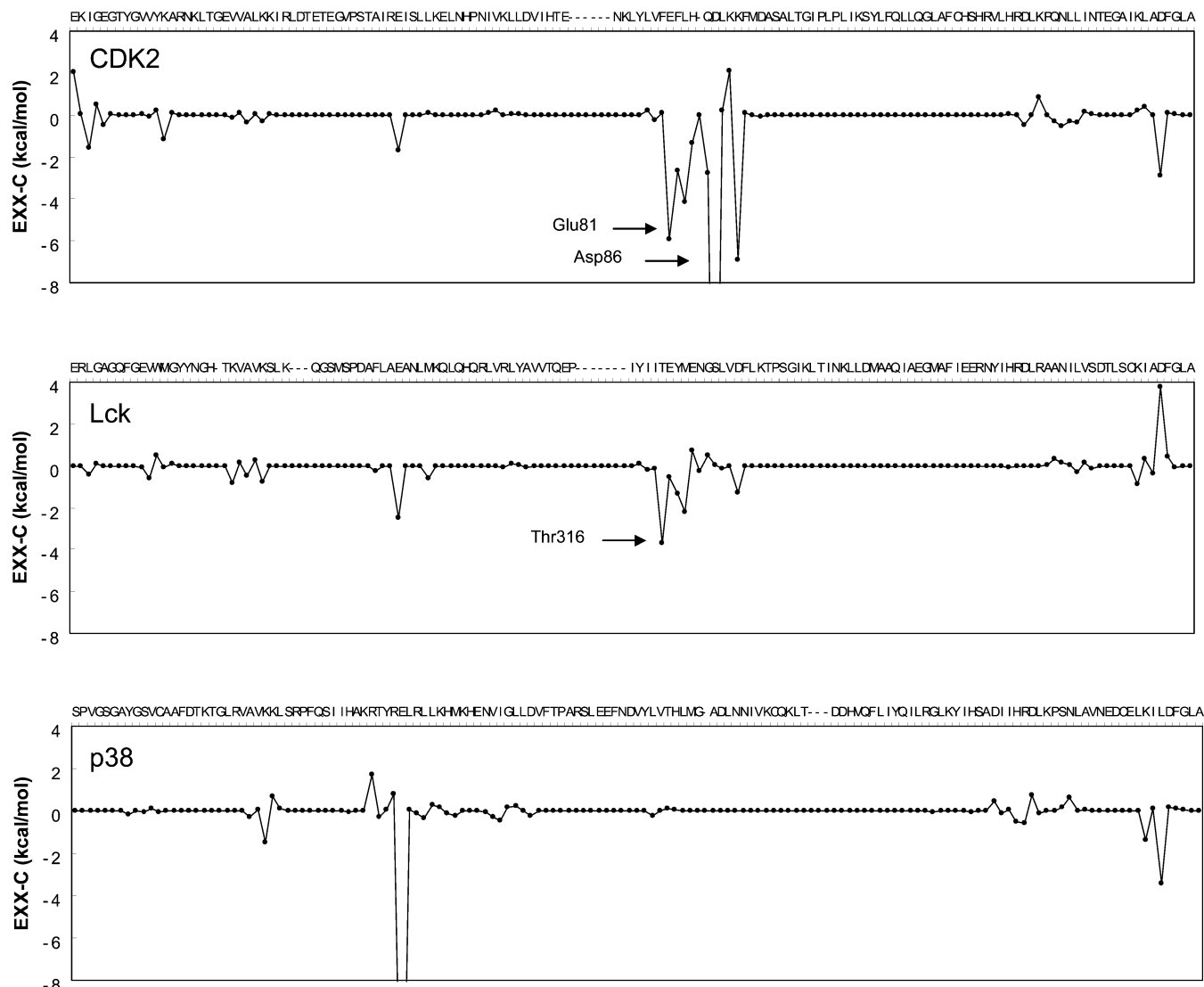
Figure 12. Residue-based EXX-C for CDK2, Lck, and p38.

EXX-LJ and EXX-C, were broken down into their contributions from individual residues.

The sequences of the ATP-binding domains for the three kinases were aligned with ClustalX.[57] The average residue-based interaction energies, EXX-LJ and EXX-C, for the three proteins are illustrated in Figures 11 and 12, respectively. These results have been averaged over all of the inhibitors and during the entire averaging segments of the MC simulations for the complexes. The patterns for the energy components for p38 are clearly different from those for CDK2 and Lck owing to the difference in binding sites. The present analysis then focuses on the differences for the CDK2 and Lck inhibitors as competitors for ATP-binding sites. For CDK2, residues Ile10, Val18, Ala31, Lys33, Val 64, Phe80, Phe82, Leu83, Gln85, Leu134, Ala144, and Asp145 have notable favorable Lennard−Jones (van der Waals) interactions with the inhibitor as do the corresponding residues of Lck, Leu251, Val259, Ala271, Lys273, Val301, Thr316, Tyr318, Met319, Gly322, Leu371, Ala381, and Asp382. Residues with highly favorable Lennard−Jones interactions with the ligand in only one system are Lys89 in CDK2, and Glu288, Met292, and Ile314 in Lck. These specific residues with the cores from Tables 1 and 10 are overlaid in Figure

13. For CDK2, the side chain of Phe80 blocks the right side of the ATP-binding site and the CDK2 ligands cannot contact the residues corresponding to Glu288, Met292, and Ile314 in Lck. Thus, inhibitor designs that feature interactions with Glu288, Met292, and Ile314 can improve specificity for Lck.

For the Coulombic interactions, residues Glu51, Leu83, and Lys89 of CDK2 and their Lck analogues, Glu288, Met319, and Asp326, show particularly favorable interactions with the inhibitors. Asp145 of CDK2 also has notably attractive electrostatic interactions, but the corresponding Asp382 of Lck has a net repulsive electrostatic interaction with the inhibitor. As mentioned above, the imidazoquinoxaline core of the Lck inhibitors forms water-mediated hydrogen bonds with Asp382, and this may compensate for the unfavorable interaction. The residues with favorable electrostatic interactions for just one kinase are Glu81 and Asp86 in CDK2 and Thr316 in Lck. These common and unique residues are overlaid with the example inhibitors in Figure 14. The Coulombic interaction between Asp86 in CDK2 and the sulfonamide group of the inhibitor is particularly favorable and it is reinforced by the interaction with Lys89. Removal of the sulfonamide eliminates CDK2 inhibition (Table 2). Thus, maximizing favorable interactions on
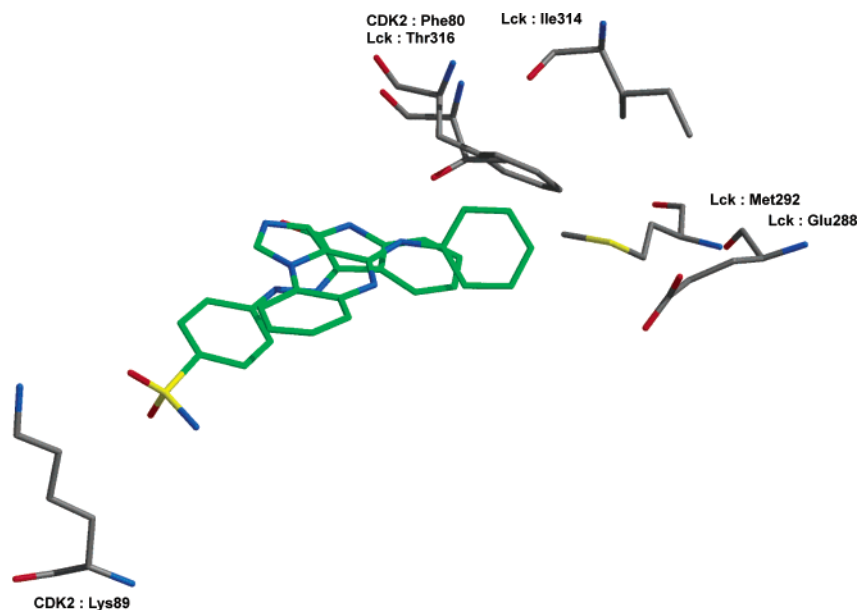
Figure 13. Key residues showing specificity for CDK2 and Lck based on Lennard−Jones interactions.
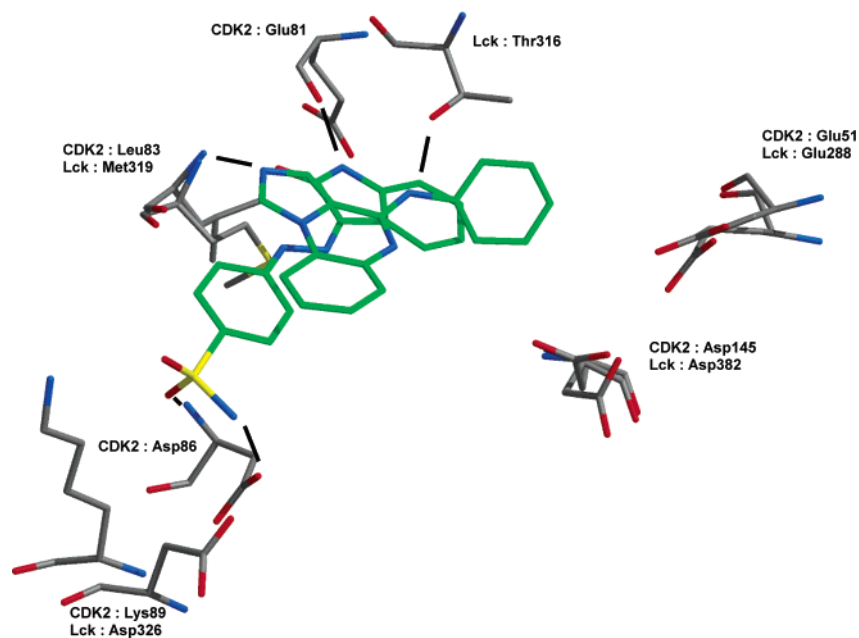

Figure 14. Key residues showing specificity for CDK2 and Lck based on Coulombic interactions.

this West end of the ATP-binding for CDK2 inhibitors and maximizing them on the East end for Lck inhibitors (Figure 13) should lead to high selectivity between this kinase pair.

Discussion

As noted above, in our previous ELR studies,[11−14] the descriptors that emerged most often as significant are EXX-LJ, ΔHB_{total}, $\Delta FOSA$, ΔE_{int}, and #RB. ΔHB_{total} and EXX-LJ are also applicable to the kinase systems. In addition, $\Delta SASA$ and EXX-C arose as significant in this study. Though the descriptors are all reasonable and the number in each ELR model is small, the selected set and their coefficients are variable. Therefore, a universal ELR equation has not been obtained. In particular, the constant terms have ranged widely: 0.3 for factor Xa,[14] −1.3 for HIV-1 RT,[11] 3.1 for kinases (eq 13), −7.9 for COX-2,[13] and −13.3 kcal/mol for throm-

bin.[12] Though some of the variation may reflect differences in assay conditions, it is most likely that the constants are absorbing some contributions from descriptors that are relatively invariant for a specific protein or set of inhibitors.

The lack of EXX-C in the prior ELR models is particularly striking, though it has some overlap with ΔHB_{total}. Notably, the thrombin inhibitors[12] all have *a* +1 charge and feature strong electrostatic interactions with the protein including the hallmark salt bridge between an amidinium group and the side chain of Asp189. Only 20 inhibitors were covered in the thrombin study; it is expected that the strong protein−ligand electrostatic attraction is relatively constant for them and it is embedded in the large constant term, −13.3 kcal/mol. Given the much greater size and diversity of the dataset for the present study of kinase inhibition and the polarity of the ATP binding site, it is reasonable

and reassuring that EXX-C has now emerged as significant. Nevertheless, there is still a substantial constant in eq 13, which along with the ΔSASA term may be masking other relatively invariant contributors that might be revealed by analysis of still larger datasets.

Another striking point is that ΔHB_{total} appears in the regression equations for the kinases, HIV-1 RT, thrombin, and COX-2, and the term was largely invariant for factor Xa. It is derived from eq 16,

$$\Delta HB_{total} = HB_{total}^{P-L} - (HB_{total}^{L-S(free)} - HB_{total}^{L-S(complex)}) \quad (16)$$

where HB_{total}^{P-L} is the total number of hydrogen bonds between protein and ligand, and $HB_{total}^{L-S(complex)}$ and $HB_{total}^{L-S(free)}$ are the total number of hydrogen bonds between ligand and water in the protein−ligand complex and unbound state. The sign of its coefficient is always negative, but the magnitude has varied from 0.22 to 1.65. The magnitude here, 0.3, is consistent with the contribution of typical hydrogen bonds to the EXX-C term in eq 13. For example, as mentioned above, the cores from Tables 1 and 10 form hydrogen bonds with Leu83 in CDK2 and Met319 in Lck. EXX-C was broken down into its contributions from the individual residues. The Coulombic interaction energy between core 1 and Leu83 is −3.43 kcal/mol and it is −3.49 kcal/mol between core 10 and Met319. The contributions of these Coulombic interactions to ΔG_{calcd} are then both −0.3 kcal/mol using eq 14. This value is the same as the coefficient of ΔHB_{total}. In the present model then, if the ligand forms a hydrogen bond with the protein, there is a net favorable contribution of ca. 0.3 kcal/mol to ΔG_{calcd} from the EXX-C term because ΔHB is zero. If the ligand forms a hydrogen bond to water in both the unbound and bound state, there is no net contribution to ΔG_{calcd} because ΔHB is again zero. No explicit term for desolvation of the protein is evident in eq 13, though such effects are reflected in EXX. If the protein and ligand have a good hydrogen-bonding match, EXX will be more favorable.

Conclusions

Individual ELR models were obtained to reproduce inhibitory activities for CDK2, Lck, and p38 with r^2 values of 0.759, 0.734, and 0.678, respectively. Although structurally different series of inhibitors were employed in the analyses, all descriptors, EXX-C, EXX-LJ, and ΔHB_{total}, are the same for the CDK2 and Lck models. As the inhibitors for p38 bind to a different, allosteric binding site, alternative descriptors, EXX, ΔFOSA, and QlogPo/w, became significant. All three models shared EXX or its components, EXX-C and EXX-LJ, in common.

The datasets for two of the three kinases were then combined in turn for training. The combined ELR models yielded q^2 values of 0.649 to 0.737 for predictions on the remaining kinase dataset using one indicator variable, L_{corr}, and three common descriptors.

Finally, the complete dataset covering 146 inhibitors for the three kinases was used to yield one ELR model, which reproduces the experimental activities well ($r^2 = 0.691$) and shows good predictive ability ($q^2 = 0.666$). The selected descriptors, EXX, ΔHB_{total}, and ΔSASA, provide insight on the factors that control kinase inhibition. Specifically, good steric and electrostatic matches between inhibitor and protein are favorable, loss of hydrogen bonds for the inhibitor upon binding is unfavorable, and burial of surface area of the inhibitor is favorable. The choice of descriptors and quality of the model were confirmed with an optimization procedure that featured a genetic algorithm.

In prior work, the ability of the ELR approach to reproduce observed activities for inhibitors of a single protein using a small number of descriptors has been clearly demonstrated. Though a few consensus descriptors have emerged, the optimal models for different proteins are sufficiently varied that discovery of a universally applicable ELR equation has been elusive. Progress on this issue has been made in the present study of the inhibition of three protein kinases. It has been demonstrated that the ELR approach can not only reproduce well experimental activities for kinase inhibitors, but that there is also good transferability of the ELR models to yield useful predictions for new kinase targets and a potentially valuable tool for inhibitor design.

References

(1) 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.

(2) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791−804.

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

(4) Lamb, M. L.; Jorgensen, W. L. Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449−457.

(5) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(6) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.

(7) Jones-Hertzong, D. K.; Jorgensen, W. L. Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method. *J. Med. Chem.* **1997**, *40*, 1539−1549.

(8) Smith, R. H. J.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, M. B. K. Prediction of Binding Affinities for TIBO Inhibitors of HIV-1 Reverse Transcriptase Using Monte Carlo Simulations in a Linear Response Method. *J. Med. Chem.* **1998**, *41*, 5272−5286.

(9) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the Binding Affinities of FKBP12 Inhibitors Using a Linear Response Method. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.

(10) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 145−154.

(11) Rizzo, R. C.; Udier-Blagović, M.; Wang, D.; Watkins, E. K.; Smith, M. B. K.; Smith, R. H. J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of Activity for Nonnucleoside Inhibitors with HIV-1 Reverse Transcriptase Based on Monte Carlo Simulations. *J. Med. Chem.* **2002**, *45*, 2970−2987.

(12) Pierce, A. C.; Jorgensen, W. L. Estimation of Binding Affinities for Selective Thrombin Inhibitors via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 1043−1050.

(13) Wesolowski, S. S.; Jorgensen, W. L. Estimation of Binding Affinities for Celecoxib Analogues with COX-2 via Monte Carlo-Extended Linear Response. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 267−270.

(14) Ostrovsky, D.; Udier-Blagović, M.; Jorgensen, W. L. Analyses for Activity of Factor Xa Inhibitors Based on Monte Carlo Simulations. *J. Med. Chem.* **2003**, *46*, 5691−5699.

(15) Cohen, P. Protein kinases − the major drug targets of the twenty-first century? *Nature Rev.* **2002**, *1*, 309−315.

(16) Hardie, G.; Hanks, S. *The protein kinase facts book I*; Academic Press: London, 1995.

(17) Harper, J. W.; Adams, P. D. Cyclin-Dependent Kinases. *Chem. Rev.* **2001**, *101*, 2511−2526.

(18) Morgan, D. O. Cyclin-dependent kinases: Engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 261−291.

(19) Stein, G.; Baserga, R.; Giordano, A.; Denhardt, D. T. *The Molecular Basis of Cell Cycle and Growth Control*; John Wiley & Sons: New York, 1998.

(20) Gray, N.; Detivaud, L.; Doerig, D.; Meijer, L. ATP−Site Inhibitors of Cyclin-Dependent Kinases. *Curr. Med. Chem.* **1999**, *6*, 859−875.

(21) Walker, D. H.; Small-Molecule Inhibitors of Cyclin-Dependent Kinases: Molecular Tools and Potential Therapeutics. *Curr. Top. Microbiol. Immunol.* **1998**, *227*, 149−165.

(22) Garrett, M. D.; Fattaey, A. CDK Inhibitors and Cancer Therapy. *Curr. Opin. Genet. Dev.* **1999**, *9*, 104−111.

(23) Weiss, A.; Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **1994**, *76*, 263−274.

(24) Straus, D. B.; Weiss, A. Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor. *Cell* **1992**, *70*, 585−593.

(25) Hanke, J. H.; Pollok, B. A.; Changelian, P. S. Role of tyrosine kinases in lymphocyte activation: targets for drug intervention. *Inflammation Res.* **1995**, *44*, 357−371.

(26) Dinarello, O. A. Inflammatory cytokines: interleukin-1 and tumor nerosis factor as effector molecules in autoimmune diseases. *Curr. Opin. Immunol.* **1991**, *3*, 941−948.

(27) Foster, M. L.; Halley, F.; Souness, J. E. Potential of p38 inhibitors in the treatment of rheumatoid arthritis. *Drug News Perspect.* **2000**, *13*, 488−497.

(28) Feldmann, M.; Brennan, F. M.; Maini, R. N. Role of cytokines in rheumatoid arthritis. *Annu. Rev. Immunol.* **1996**, *14*, 397−440.

(29) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T.; Harris, J. P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shew-chuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-Based Inhibitors of Cyclin-Dependent Kinase 2 (CDK2): Design, Synthesis, Enzymatic Activities, and X-ray Crystallographic Analysis. *J. Med. Chem.* **2001**, *44*, 4339−4358.

(30) 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.

(31) Regan, J.; Breitfelder, S.; Cirillo, P.; Gilmore, T.; Graham, A. G.; Hickey, E.; Klaus, B.; Madwed, J.; Moriak, M.; Moss, N.; Pargellis, C.; Pav, S.; Proto, A.; Swinamer, A.; Tong, L.; Torcellini, C. Pyrazole Urea-Based Inhibitors of p38 MAP kinase: From Lead Compound to Clinical Candidate. *J. Med. Chem.* **2002**, *45*, 2994−3008.

(32) Chen, P.; Norris, D.; Iwanowicz, E. J.; Spergel, S. H.; Lin, J.; Gu, H. H.; Shen, Z.; Wityak, J.; Lin, T. A.; Pang, S.; Fex, H. F. D.; Pitt, S.; Shen, D. R.; Doweyko, A. M.; Bassolino, D. A.; Roberge, J. Y.; Poss, M. A.; Chen, B. C.; Schieven, G. L.; Barrish, J. C. Discovery and Initial SAR of Imidazoquinolines as Inhibitors of the Src-Family Kinase p56Lck. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1361−1364.

(33) Chen, P.; Iwanowicz, E. J.; Norris, D.; Gu, H. H.; Lin, J.; Moquin, R. V.; Das, J.; Wityak, J.; Spergel, S. H.; Fex, H. D.; Pang, S.; Pitt, S.; Shen, D. R.; Schieven, G. L.; Barrish, J. C. Synthesis and SAR of Novel Imidazoqunoxaline-Based Lck Inhibitors: Improvement of Cell Potency. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3153−3156.

(34) Cheng, Y.; Prusoff, W. H. Relationship Between Inhibition Constant (Ki) and Concentration of Inhibitor Which Causes 50 Per Cent Inhibition (I50) of Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.

(35) Schulze-Gahmen, U.; De Bondt, H. L.; Kim, S. H. High-Resolution Crystal Structures of Human Cyclin-Dependent Kinase 2 with and without ATP: Bound Waters and Natural Ligand as Guides for Inhibitor Design. *J. Med. Chem.* **1996**, *39*, 4540−4546.

(36) Brown, N. R.; Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. The Structural Basis for Specificity of Substrate and Recruitment Peptides for Cyclin-Dependent Kinases. *Nature Cell Biology* **1999**, *1*, 438−443.

(37) Johnson, L. N.; Lewis, R. J. Structural Basis for Control by Phosphorylation. *Chem. Rev.* **2001**, *101*, 2209−2242.

(38) Watts, J. D.; Wilson, G. M.; Ettenhadieh, E.; Clark-Lewis, I.; Kubanek, C. A.; Astell, C. R.; Marth, J. D.; Aebersold, R. Purification and Initial Characterization of the Lymphocyte-Specific Protein-Tyrosyl Kinase p56lck from a Baculovirus Expression System. *J. Biol. Chem.* **1992**, *267*, 901−907.

(39) Jullien, P.; Bougeret, C.; Camoin, L.; Bodeus, M.; Durand, H.; Disanto, J. P.; Fischer, S.; Benarous, R. Tyr394 and Tyr505 are Autophosphorylated in Recombinant Lck Protein-Tyrosine Kinase Expression in *Escherichia coli. Eur. J. Biochem.* **1994**, *224*, 589−596.

(40) Zhu, X.; Kim, J. L.; Newcomb, J. R.; Rose, P. E.; Stover, D. R.; Toledo, L. M.; Zhao, H.; Morgenstern, K. A. Structural Analysis of the Lymphocyte-Specific Kinase Lck in Complex with Non-Selective and Src Family Selective Kinase Inhibitors. *Structure* **1999**, *7*, 651−661.

(41) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibiton of p38 MAP Kinase by Utilizing a Novel Allosteric Binding Site. *Nature Struct. Biol.* **2002**, *9*, 268−272.

(42) Jorgensen, W. L. *BOMB*, version 2.2; Yale University: New Haven, CT, 2003.

(43) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(44) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class IV Charge Models − a New Semiempirical Approach in Quantum Chemistry. *J. Computer-Aided Mol. Des.* **1995**, *9*, 87−110.

(45) Jorgensen, W. L. *MCPRO*, version 1.68: Yale University: New Haven, CT, 2001.

(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(47) *QikProp*, version 2.0.001: Schrödinger, Inc.: New York, 2001.

(48) JMP, version 4: SAS Institute Inc.: Cary, NC, 2000.

(49) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms Part 1. Concepts, Properties and Context. *Chemom. Intell. Lab. Syst.* **1993**, *19*, 1−33.

(50) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms Part 2. Representation, Configuration and Hybridization. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 99−145.

(51) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: a Tutorial. *Anal. Chem. Acta*, **1986**, *185*, 1−17.

(52) Glen, W. G.; Dunn, W. J., III; Scott, D. R. Principal Component Analysis and Partial Least Squares Regression. *Tetrahedron Comput. Methodol.* **1989**, *6*, 349−376.

(53) Dreyer, M. K.; Borcherding, D. R.; Dumont, J. A.; Peet, N. T.; Tsay, J. T.; Wright, P. S.; Bitonti, A. J.; Shen, J.; Kim, S. H. Crystal Structure of Human Cyclin-Dependent Kinase 2 in Complex with the Adenine-Derived Inhibitor H717. *J. Med. Chem.* **2000**, *44*, 524−530.

(54) Engh, R. A.; Girod, A.; Kinzel, V.; Huber, R.; Bossemeyer, D. Crystal Structures of Catalytic Subunit of cATP-Dependent Protein Kinase in Complex with Isoquinolinesulfonyl Protein Kinase Inhibitors H7, H8, H89. Structural Implications for Selectivity. *J. Biol. Chem.* **1996**, *271*, 26157−26164.

(55) Schindler, T.; Bornmann, W.; Pellicena, W.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. *Science* **2000**, *289*, 1938−1942.

(56) Pierce, A. C.; Sandretto, K. L.; Bemis, G. W. Kinase Inhibitors and the Case for CH···O Hydrogen Bonds in Protein−Ligand Binding *Proteins* **2002**, *49*, 567−576.

(57) Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *24*, 4876−4882.