3D-QSAR CoMFA on Cyclin-Dependent Kinase Inhibitors

Pierre Ducrot, Michel Legraverend,* and David S. Grierson

Institut Curie, Section de Recherche, UMR 176 CNRS, Bât. 110-112, Centre Universitaire, 91405 Orsay Cedex, France

Received April 5, 2000

Several series of cyclin-dependent kinase inhibitors previously prepared in our laboratory were compared using 3D-QSAR (CDK1) and docking (CDK2) techniques. Evaluation of our own library of 93 purine derivatives served to establish the model which was validated by evaluation of an external library of 71 compounds. The best predictions were obtained with the CoMFA standard model ($q^2 = 0.68$, $r^2 = 0.90$) and with the CoMSIA combined steric, electrostatic, and lipophilic fields ($q^2 = 0.74$, $r^2 = 0.90$). The CDK1 3D-QSAR model was then superimposed to the ATP/CDK2 binding site, giving direct contour maps of the different fields. Although too few compounds were evaluated on CDK5 to derive a 3D-QSAR model, some interesting SARs have been deduced. Comparison of the results obtained from both methods helped with understanding the specific activity of some compounds and designing new specific CDK inhibitors.

Introduction

The cyclin-dependent kinases CDK1, CDK2, CDK4, and CDK6 play an important role in cell cycle regulation.[1] Their sequential activation ensures the correct timing and ordering of events required for cell cycle progression.[2] Intense efforts in biochemistry, molecular biology, and genetics have helped to define the mechanism of action and regulation of the CDKs, as well as to identity certain of their protein substrates.[2] Small molecules which specifically inhibit CDKs through interaction with the ATP binding site[3] are potentially valuable biochemical tools in such studies[4] and may also have major application as therapeutic agents. In this context, several different families of CDK inhibitors have already been discovered,[5−10] and their binding affinities to CDK1, CDK2, and/or CDK5 (involved in neuronal cell functions[11]) have been measured.

To date, the structures of CDK1, CDK4, and CDK5 have not been reported. However, X-ray crystal structures of CDK2 complexed with ATP,[12] olomoucine,[13] roscovitine,[6] staurosporine,[14] flavopiridol,[15] purvalanol,[16] and indirubin[8] have been obtained. These studies illustrate how the ATP binding site of CDK2 can accommodate structurally diverse inhibitor types. Further, a comparison of the X-ray structures for CDK2/cyclin A/ATP,[13,17] CDK2 alone,[12] and CDK2/cyclin A/p27^{KIP1} (natural CKI)[18] reveals only limited differences in the vicinity of the ATP adenine binding site between different activation states of CDK2. This suggests that information derived from monomeric CDK2/inhibitor structures can help to design inhibitors against the active binary complex.

Recently, we described the synthesis and evaluation as inhibitors of CDK1 of several new series of purine derivatives bearing modifications at the C-6, N-9, and, in particular, C-2 positions.[19,20] In a continuation of this program, the crystal structure of CDK2 complexed to the C-2 acetylene-substituted compound **567** (Figure 1, Table 1) has been obtained.[21] Due to the high level of sequence homology between CDK1 and CDK2, this study provided a visual insight into the molecular basis for the inhibition of CDK1 by this molecule and further served as a valid starting point for molecular modeling to rationalize the structure−activity relationship (SAR) observed for the series of C-2 acetylenyl-substituted purines.

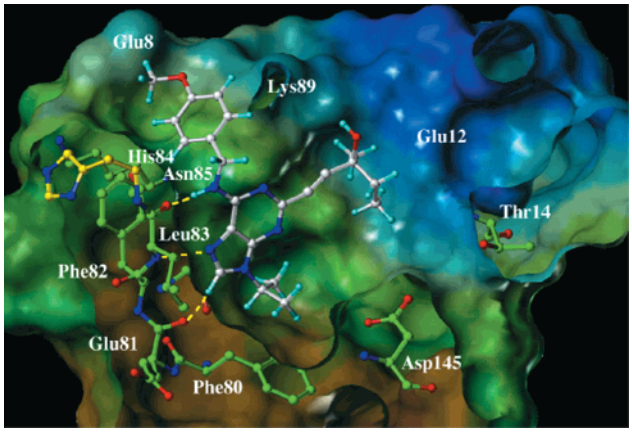


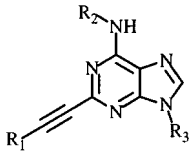
Figure 1. Compound **567** within the CDK2 binding site from an X-ray structure. Protein residues are drawn in green without hydrogens for clarity; **567** is in gray (C), blue (N), and red (O); ATP binding site surface is colored by lipophilic potential using MOLCAD (more lipophilic in brown, more hydrophilic in dark blue).

In the present paper, a quantitative SAR (QSAR) study by molecular field analysis of our complete library of 93 purine derivatives[19−22] has been carried out using SYBYL software.[23] Both CoMFA[24] (a well-documented and validated method) and CoMSIA[25] (recently available within the Sybyl QSAR package) techniques were employed. The CoMSIA technique is of particular interest since it includes a lipophilic field and two hydrogen bond fields and is less alignment-sensitive than CoMFA.[25,26] The predictiveness of each of our optimized models was evaluated using two sets of compounds. The first set was constituted of five compounds with known CDK1 activi-

* To whom correspondence should be addressed. Tel: 1-69 86 30 86. Fax: 1-69 07 53 81. E-mail: Michel.Legraverend@curie.u-psud.fr.

10.1021/jm000965t CCC: $19.00
Published on Web 10/14/2000

Table 1. CDK1 and CDK5 IC_{50} (μM) for Compounds of the Training and Testing Sets of Family 1[a]



compd	R_1	R_2	R_3	CDK1 IC_{50}	CDK1 pIC_{50}	CDK5 IC_{50}
424	$CH(OH)CH_2CH_3$	CH_2Ph	*i*-Pr	0.50	6.30	ND
567	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph-*p*-OCH_3	*i*-Pr	0.23	6.64	0.40
601	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph-*m*,*p*-$(Cl)_2$	*i*-Pr	0.43	6.37	0.50
420	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph	*i*-Pr	0.20	6.70	0.34
595	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph-*p*-Cl	*i*-Pr	0.06	7.22	0.23
556	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph-*m*,*p*-$(OCH_3)_2$	*i*-Pr	0.30	6.52	0.60
557	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph-*m*,*p*-OCH_2O	*i*-Pr	0.20	6.70	0.30
576	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph-*p*-$N(CH_3)_2$	*i*-Pr	0.15	6.82	0.26
597	$C(CH_3)(OH)CH_2CH_3$	Ph-*m*-Cl	*i*-Pr	0.06	7.22	0.50
538	$C(CH_3)(OH)CH_2CH_3$	Ph	*i*-Pr	0.40	6.40	0.30
603	$C(CH_3)(OH)CH_2CH_3$	Ph-*m*,*p*-$(Cl)_2$	*i*-Pr	0.20	6.70	0.50
564	$C(CH_3)(OH)CH_2CH_3$	Ph-*p*-OCH_3	*i*-Pr	0.20	6.70	0.30
577	$C(CH_3)(OH)CH_2CH_3$	Ph-*p*-OCH_2CH_3	*i*-Pr	0.20	6.70	0.40
575	$C(CH_3)(OH)CH_2CH_3$	CH_2Ph-*p*-OCH_2CH_3	*i*-Pr	0.18	6.74	0.40
574	$C(CH_3)(OH)CH_2CH_3$	$(CH_2)_2Ph$-*p*-OCH_3	*i*-Pr	0.40	6.40	ND
580	$C(CH_3)(OH)CH_2Ph$	*n*-Bu	*i*-Pr	0.30	6.52	ND
1417	CH_2OH	CH_2Ph	*i*-Pr	1.2	5.92	ND
421	$(CH_2)_2OH$	CH_2Ph	*i*-Pr	0.35	6.46	ND
422	$(CH_2)_3OH$	CH_2Ph	*i*-Pr	0.20	6.70	ND
501	$(CH_2)_4OH$	CH_2Ph	*i*-Pr	0.33	6.48	ND
425	$C(CH_3)(OH)Ph$	CH_2Ph	*i*-Pr	0.60	6.22	ND

[a] *i*-Pr = isopropyl; $pIC_{50} = -\log IC_{50}$; ND = not determined.

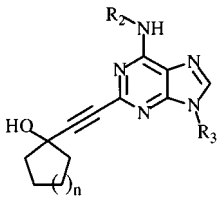
ties from our database, but which were not included in the model. The second set corresponded to a series of 79 purine derivatives recently published by Schultz et al.,[5] whose activities were measured under the same conditions as for our data set. This second set includes many purine derivatives with C-2 and N-6 modifications which are not represented in our data set. Therefore, its analysis provides a rigorous test of the predictive power and the limits of our model.

The contour maps derived from both the CoMFA and CoMSIA 3D-QSAR models permitted an understanding of the steric, electrostatic, lipophilic, and hydrogen-bonding requirements for ligand binding. As a consequence, the structural variations in the training set that gave rise to variations in the molecular fields at particular regions of the space that are correlated to biological properties served as a guide to the design of novel inhibitors. The comparison of the models obtained for CDK1 with the structure of CDK2 and the binding affinity data for CDK5 (when available) also provided information which directed the development of new selective inhibitors for the ATP binding site of the different CDKs.

Methods

All 93 of the purine derivatives in our database were evaluated for their capacity to inhibit CDK1. Depending upon the nature of the C-2 substituent, these compounds belong to five structurally different families: (i) acetylenyl (Tables 1, 2), (ii) amino (Table 3), (iii) hydroxyalkyl (Table 4), (iv) iodo (Table 5), and (v) pyrrolidino (Table 6). Of these compounds, 6 display pIC_{50} ($pIC_{50} = -\log IC_{50}$) values less than 4 (no affinity), 7 display pIC_{50} values between 4 and 5 (very low affinity), 30 exhibit pIC_{50} values between 5 and 6 (low affinity), 47 display pIC_{50} values between 6 and 7 (medium activity), and 3 show pIC_{50} values higher than 7 (good affinity). Thus, the structural diversity and the homogeneous distribution of binding affinities, which are necessary to derive meaningful

Table 2. CDK1 IC_{50} (μM) for Compounds of the Training and Testing Sets of Family 1[a]



compd	*n*	R_2	R_3	CDK1 IC_{50}	CDK1 pIC_{50}
482	1	CH_2Ph	*i*-Pr	0.50	6.30
481	2	CH_2Ph	*i*-Pr	0.18	6.74
579	3	CH_2Ph-*p*-OCH_3	*i*-Pr	0.50	6.30

[a] *i*-Pr = isopropyl; $pIC_{50} = -\log IC_{50}$.

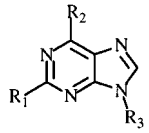
results from a 3D-QSAR analysis, are present in our set of compounds.

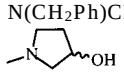
To assess the predictive power of the model, a training set containing 88 compounds from our library was used, together with a testing set of 5 molecules (**1417**, **1437**, **410**, **524**, and **577**) ramdomly selected with respect to their activities and family type.

An essential requirement for 3D-QSAR techniques using molecular field analysis is a knowledge of the active conformation of the inhibitor under study. The results of docking experiment studies employing the X-ray structure of the **567**/CDK2 complex[21] were thus used to determine the relative alignment rules. It was determined that, regardless of the substituent on N-6, the purine core always occupies the same position in the active site. Further, careful observation of the superimposed crystal data of CDK2 with different inhibitors (roscovitine and **567**) showed that the hydroxyl group on the purine C-2 side chain (along with N-1 and N-3 in the purine ring) interacts with a water molecule, rather than binding directly to the protein. Thus, a certain flexibility is possible in the positioning of this group.

The conformations and orientations of the C-6, N-9, and C-2 substituents retrieved from the olomoucine/CDK2, roscovitine/CDK2, and **567**/CDK2 crystal structures and the docking

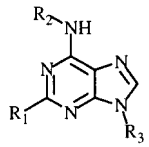
Table 3. CDK1 and CDK5 IC_{50} (μM) for Compounds of the Training and Testing Sets of Family 2[a]



compd	*R*/*S*	R_1	R_2	R_3	CDK1 IC_{50}	CDK1 pIC_{50}	CDK5 IC_{50}
1406		NH_2	Cl	cyclopentenyl	260	3.59	>100
1409		$NH(CH_2)_2OH$	$N(CH_3)CH_2CH{=}CH_2$	CH_3	470	3.33	>100
368		$NH(CH_2)_2OH$	$N(CH_3)CH_2CH{=}CH_2$	*i*-Pr	150	3.70	>100
378		$NH(CH_2)_2OH$		*i*-Pr	110	3.96	>100
olo		$NH(CH_2)_2OH$	$NHCH_2Ph$	*i*-Pr	7.0	5.15	3.0
ros	*R*	$NHCH(CH_2CH_3)CH_2OH$	$NHCH_2Ph$	*i*-Pr	0.45	6.35	ND
ros	*R*,*S*	$NHCH(CH_2CH_3)CH_2OH$	$NHCH_2Ph$	*i*-Pr	0.65	6.19	0.16
416	*R*,*S*	$NHCH((CH_2)_2CH_3)CH_2OH$	$NHCH_2Ph$	*i*-Pr	0.90	6.05	0.4
410	*R*,*S*	$NHCH((CH_2)_3CH_3)CH_2OH$	$NHCH_2Ph$	*i*-Pr	2.5	5.60	2.3
417	*R*	$NHCH(CH_3)CH_2OH$	$NHCH_2Ph$	*i*-Pr	0.85	6.07	0.7
418	*S*	$NHCH(CH_3)CH_2OH$	$NHCH_2Ph$	*i*-Pr	1.0	6.00	35
409	*R*,*S*	$NHCH_2CH(Ph)OH$	$NHCH_2Ph$	*i*-Pr	6.5	5.19	2.3
412	*S*	$NHCH(Ph)CH_2OH$	$NHCH_2Ph$	*i*-Pr	4.3	5.37	3.5
413	*R*	$NHCH(Ph)CH_2OH$	$NHCH_2Ph$	*i*-Pr	1.0	6.00	6.0
414	*S*	$NHCH(CH_2Ph)CH_2OH$	$NHCH_2Ph$	*i*-Pr	7.3	5.14	2.0
415	*R*	$NHCH(CH_2Ph)CH_2OH$	$NHCH_2Ph$	*i*-Pr	2.7	5.57	1.3
1574	*R*	$NHCH((CH(CH_3)_2)CH_2OH$	$NHCH_2Ph$-*m*,*p*-OCH_2O	*i*-Pr	0.02	7.70	ND
408		$N(CH_2Ph)CH_2CH_2OH$	$NHCH_2Ph$	*i*-Pr	2.5	5.60	8.0
419	*R*,*S*		$NHCH_2Ph$	*i*-Pr	42	4.38	2.4

[a] *i*-Pr = isopropyl; $pIC_{50} = -\log IC_{50}$; olo = olomoucine; ros = roscovitine; ND = not determined.

Table 4. CDK1 IC_{50} (μM) for Compounds of the Training and Testing Sets of Family 3[a]



compd	*R*/*S*	R_1	R_2	R_3	CDK1 IC_{50}	CDK1 pIC_{50}
459		$(CH_2)_3OH$	CH_2Ph	*i*-Pr	2.5	5.60
469		$(CH_2)_4OH$	CH_2Ph	*i*-Pr	3.0	5.52
478		$(CH_2)_5OH$	CH_2Ph	*i*-Pr	2.8	5.55
479		$CH{=}CH(CH_2)_3OH$	CH_2Ph	*i*-Pr	1.0	6.00
476	*R*,*S*	$(CH_2)_2C(CH_3)(OH)Ph$	CH_2Ph	*i*-Pr	4.5	5.35
477	*R*,*S*	$(CH_2)_2CH(OH)CH_2CH_3$	CH_2Ph	*i*-Pr	3.0	5.52
480	*R*,*S*	$(CH_2)_2C(CH_3)(OH)CH_2CH_3$	CH_2Ph	*i*-Pr	1.3	5.89

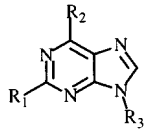
[a] *i*-Pr = isopropyl; $pIC_{50} = -\log IC_{50}$.

studies were used to build the complete set of purine derivatives using a standard fragments library. Note also that, in keeping with the X-ray data for the **567**/CDK2 complex,[21] the (*S*) stereochemistry was maintained in the acetylene series. Subsequent minimizations were performed while keeping the C-6−N-6−Ph or C-6−N-6−CH_2-Ph bonds rigid. Geometries were optimized using the Tripos force field,[27,28] without including electrostatic terms. The method of Powell, available in the Maximin2 procedure, was used for energy minimization until a gradient value smaller than 0.01 kcal/mol·Å^2 was reached. For the C-2 side chain a random search was performed to identify the lowest-energy conformations (maximum hits/conformer = 6; rms threshold = 0.2 Å; energy cuttoff = 5 kcal/mol) for those molecules in which this motif is different than in **567**, olomoucine, or roscovitine. The obtained conformers were subsequently optimized by semiempirical calculations (MOPAC 6.0 using the Hamiltonian AM1)[29] and superimposed onto **567** (extracted from the X-ray data) using the match option in Sybyl and the purine main core nitrogens as the template. Conformer selection among the low-energy conformations was achieved by comparing the overall shape with that for olomoucine, roscovitine, and **567**, whose active conformation is known. Subsequently, similar conformations were selected for related compounds.

CoMFA and CoMSIA Studies. The CoMFA steric and electrostatic fields were calculated at grid lattice points using the Lennard−Jones and the Coulomb potential functions of the Tripos force field. These calculations were carried out for different values of the probe type, grid spacing, and energy cutoff. The five CoMSIA similarity index fields available within Sybyl (steric, electrostatic, lipophilic, and hydrogen bond donnor and acceptor) were calculated at grid lattice points using a common probe atom of 1 Å radius, as well as a charge, hydrophobicity, and hydrogen bond properties of +1 and an attenuation factor of 0.3, while altering the grid spacing only.

The PLS method (partial least-squares)[30−32] was used to derive a linear relationship, and cross-validation was performed using the "leave-one-out" method,[33,34] with a 2 kcal/mol column filter to check for consistency and predictiveness. The optimum number of components used to derive the nonvalidated model was defined as the number of components leading to the highest cross-validation r^2 (subsequently called q^2) and the lowest standard error of prediction (SEP). Only models with q^2 over 0.5 and the fraction of variance (*V*) greater than 0.85 for the optimum number of components were further considered.

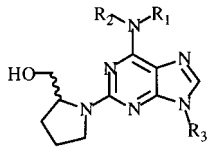
Table 5. CDK1 and CDK5 IC_{50} (μM) for Compounds of the Training and Testing Sets of Family 4[a]



compd	R_1	R_2	R_3	CDK1 IC_{50}	CDK1 pIC_{50}	CDK5 IC_{50}
362	I	Cl	*i*-Pr	150	3.82	>100
535	I	NHPh	*i*-Pr	1.0	6.00	ND
1456	I	NHPh-*m*-Br	*i*-Pr	1.3	5.89	ND
529	I	NHPh-*p*-Br	*i*-Pr	1.3	5.89	ND
371	I	$NHCH_2Ph$	*i*-Pr	8.0	5.10	13
452	I	$NHCH_2Ph$	cyclohexyl	90	4.05	>100
448	I	$NHCH_2Ph$	cyclopentyl	75	4.12	>100
442	I	$NHCH_2Ph$	CH_2Ph	100	4.00	>100
524	I	$NHCH_2Ph$-*m*-OCH_3	*i*-Pr	2.3	5.64	ND
521	I	$NHCH_2Ph$-*p*-OCH_3	*i*-Pr	0.85	6.07	ND
520	I	$NHCH_2Ph$-*o*-OCH_3	*i*-Pr	7.2	5.14	ND
522	I	$NHCH_2Ph$-*m*,*m*′-$(OCH_3)_2$	*i*-Pr	4.2	5.38	ND
519	I	$NHCH_2Ph$-*o*-CF_3	*i*-Pr	10	5.00	ND
513	I	$NHCH_2Ph$-*m*-CF_3	*i*-Pr	5.3	5.28	ND
514	I	$NHCH_2Ph$-*p*-CF_3	*i*-Pr	4.3	5.37	ND
533	I	$NHCH_2Ph$-*m*,*m*′-$(CF_3)_2$	*i*-Pr	10	5.00	ND
366	I	$N(CH_3)CH_2CH{=}CH_2$	*i*-Pr	11	4.96	14
1408	I	$N(CH_3)CH_2CH{=}CH_2$	CH_3	18	4.74	40

[a] *i*-Pr = isopropyl; $pIC_{50} = -\log IC_{50}$; ND = not determined.

Table 6. CDK1 and CDK5 IC_{50} (μM) for Compounds of the Training and Testing Sets of Family 5[a]



compd	*R*/*S*	R_1	R_2	R_3	CDK1 IC_{50}	CDK1 pIC_{50}	CDK5 IC_{50}
536	*R*	H	Ph	*i*-Pr	0.30	6.52	ND
453	*R*	H	CH_2Ph	cyclohexyl	40.0	4.40	13
515	*R*	H	CH_2Ph	*t*-Bu	10.0	5.00	ND
449	*R*	H	CH_2Ph	cyclopentyl	0.70	6.15	0.5
443	*R*	H	CH_2Ph	CH_2Ph	200	3.70	>100
496	*R*	CH_2Ph	CH_2Ph	*i*-Pr	10	5.00	ND
375	*R*	H	CH_2Ph	*i*-Pr	0.65	6.19	1.3
372	*S*	H	CH_2Ph	*i*-Pr	2.1	5.68	2.4
604	*R*	H	Ph-*m*,*p*-$(Cl)_2$	*i*-Pr	0.43	6.37	ND
596	*R*	H	Ph-*m*-Cl	*i*-Pr	0.43	6.37	ND
532	*R*	H	Ph-*p*-Br	*i*-Pr	0.30	6.52	ND
592	*R*	H	CH_2Ph-*p*-Cl	*i*-Pr	0.33	6.48	ND
1530	*R*	H	CH_2Ph-*m*-Br	*i*-Pr	0.33	6.48	ND
1437	*R*	H	CH_2Ph-*m*-I	*i*-Pr	0.45	6.35	0.16
599	*R*	H	CH_2Ph-*m*,*p*-$(Cl)_2$	*i*-Pr	0.43	6.37	ND
523	*R*	H	CH_2Ph-*o*-CF_3	*i*-Pr	4.0	5.40	ND
516	*R*	H	CH_2Ph-*m*-CF_3	*i*-Pr	0.70	6.15	ND
518	*R*	H	CH_2Ph-*p*-CF_3	*i*-Pr	0.80	6.10	ND
537	*R*	H	CH_2Ph-*m*,*m*′-$(CF_3)_2$	*i*-Pr	1.0	6.00	ND
528	*R*	H	CH_2Ph-*m*-OCH_3	*i*-Pr	0.30	6.52	ND
526	*R*	H	CH_2Ph-*p*-OCH_3	*i*-Pr	0.21	6.68	ND
525	*R*	H	CH_2Ph-*o*-OCH_3	*i*-Pr	1.8	5.74	ND
606	*R*	H	CH_2Ph-*m*,*p*-OCH_2O	*i*-Pr	0.35	6.46	ND
527	*R*	H	CH_2Ph-*m*,*m*′-$(OCH_3)_2$	*i*-Pr	0.30	6.52	ND
497	*R*	H	$CH(CH_2OH)Ph$ (*R*)	*i*-Pr	10	5.00	ND
498	*R*	H	$CH(CH_2OH)Ph$ (*S*)	*i*-Pr	0.80	6.10	ND

[a] *i*-Pr = isopropyl; $pIC_{50} = -\log IC_{50}$; ND = not determined.

Results and Discussion

The best CoMFA results were obtained using a sp^3 carbon probe carrying a charge of +1 for a grid spacing of 2 Å and when the steric and electrostatic energies were truncated to 30 kcal/mol for both the standard (STD) and hydrogen bond (HBND) fields or to 20 and 1 kcal/mol for the indicator (IND) steric and electrostatic fields, respectively (Table 7).

As seen in Table 7, the standard fields led to good q^2 and r^2 values, whereas q^2 and r^2 were lower for the hydrogen bond fields. The latter result is not surprising, as indeed, the important hydrogen bond donor or acceptor sites involving the purine ring which were identified from the X-ray data and docking experiments are invariant in our molecule set. Furthermore, in view of the structural variation that exists in the different families of analogues bearing a C-2 side chain hydroxyl

Table 7. CoMFA Results[a]

	"leave-one-out" cross-validation							influence of different fields (%)			
field	q^2	n	SEP	V	r^2	SEP	F	ster	elec	ster	elec
STD	0.683	5	0.537	0.895	0.898	0.305	143.9	57.8	42.2		
HBND	0.542	3	0.637	0.769	0.752	0.469	84.8	65.8	34.2		
IND	0.659	5	0.557	0.898	0.937	0.239	245.2	46.8	53.2		
STD+HBND	0.661	4	0.551	0.914	0.894	0.309	174.8	25.3	14.6	35.5	24.7
STD+IND	0.689	4	0.528	0.890	0.906	0.291	199.5	20.3	14.9	30.0	34.9

[a] Standard fields (STD); hydrogen bond fields (HBND); indicator fields (IND); n = number of components; SEP = standard error of prediction; V = fraction of variance; F = F ratio.

Table 8. CoMSIA Results[a]

	"leave-one-out" cross-validation							influence of different fields (%)				
field	q^2	n	SEP	V	r^2	SEP	F	steric	elec	lipo	H don	H acc
ST+EL	0.696	5	0.526	0.872	0.864	0.351	104.4	40.1	59.9			
LIPO	0.759	7	0.474	0.908	0.901	0.303	104.5			100		
HBND	0.499	6	0.679	0.729	0.685	0.538	29.4				36.8	63.2
ST+EL+LIPO	0.738	6	0.491	0.962	0.900	0.303	121.3	20.5	38.7	40.7		
ST+EL+HBND	0.706	6	0.520	0.894	0.884	0.327	102.7	19.7	41.0		14.1	25.1
ALL	0.736	6	0.493	0.912	0.907	0.292	132.3	13.5	28.4	27.8	10.4	20.0

[a] Steric (ST), electrostatic (EL), lipophilic (LIPO), and hydrogen bond (HBND) similarity index fields were used. SEP = standard error of prediction; V = fraction of variance; n = number of components.

Table 9. Predicted CDK1 pIC_{50} for the Testing Set Compounds[a]

compd	**1417**	**1437**	**410**	**524**	**577**	testing set predictions			Schultz set predictions		
CDK1 pIC_{50}	5.92	6.35	5.60	5.64	6.70	r^2_{pred}	mean	SD	r^2_{pred}	mean	SD
					CoMFA						
STD	6.02	6.10	6.05	5.76	6.85	0.72	0.21	0.15	0.81	0.45	0.30
IND	5.89	5.92	6.31	6.05	7.17	0.13	0.41	0.24	0.78	0.46	0.34
STD+HBND	5.99	5.90	5.90	5.49	6.74	0.74	0.20	0.17	0.81	0.44	0.30
STD+IND	6.00	5.97	5.98	6.01	6.98	0.59	0.30	0.13	0.80	0.45	0.31
					CoMSIA						
ST+EL	5.73	5.85	6.24	5.54	6.80	0.43	0.31	0.25	0.79	0.46	0.31
LIPO	5.46	6.08	6.38	5.78	6.70	0.26	0.33	0.30	0.73	0.51	0.38
ST+EL+LIPO	5.72	6.09	6.26	5.60	6.60	0.55	0.25	0.24	0.78	0.44	0.36
ST+EL+HBND	5.68	5.72	6.13	5.44	6.89	0.35	0.36	0.20	0.72	0.52	0.38
ALL	5.75	5.98	6.23	5.48	6.77	0.52	0.28	0.22	0.75	0.49	0.37

[a] Mean, maximum standard deviation (SD) of prediction standard error ($pIC_{50}^{pred} - pIC_{50}^{obsd}$).

group, the position of this otherwise crucial group has little effect on activity [for instance, compare compounds **1574** (Table 3), **597**, and **422** (Table 1)]. Indeed, steric and hydrophobic effects dominate in this region of the molecule. Therefore, the number of correlations between the activity and the hydrogen bond field is low. Further attempts to combine the hydrogen bond fields with the standard steric and electrostatic fields did not lead to any improvement. Indeed, the q^2 was still lower than that obtained with the standard fields only. Although the indicator fields (IND) led to lower q^2 than the standard fields (STD), r^2 was significantly higher. However, the latter value is probably overestimated due to introduced "noise" in the model (vide infra). Combining the standard and indicator fields (STD+IND) produced a slight increase in both q^2 and r^2, as compared to the values obtained with the standard fields alone. Therefore, for the five CoMFA models, the best results thus far were obtained for the standard fields (q^2 = 0.683, r^2 = 0.898) and the combined standard and indicator fields (q^2 = 0.689, r^2 = 0.906).

The best CoMSIA results were obtained for a grid spacing of 1 Å (Table 8). As observed in the CoMFA treatment, hydrogen bond fields (HBND) produced the lowest q^2 and r^2, whereas the combined steric and electrostatic fields (ST+EL) gave good q^2 and r^2 values (0.696 and 0.864, respectively). Although q^2 for this field is higher than in the corresponding CoMFA calculation (STD), the r^2 is slightly lower. This is probably due to an overestimated r^2 in CoMFA.

The lipophilic field alone and the combined ST+EL+LIPO fields gave the best q^2 and r^2. However, the latter combination provided more information because it contains two additional fields and therefore explains 96% of the total variance with only 6 components vs 90% for 7 components with the lipophilic model (LIPO). Attempts to combine the steric and electrostatic fields with the hydrogen bond fields (ST+EL+HBND) resulted in a reduced q^2, whereas combination of all five fields (ALL) did not lead to any significant improvement.

It is noteworthy that the steric field has less relative influence on the derived CoMSIA model (40%) than on the CoMFA model (58%). This may be due to the use of the Lennard−Jones steric potential function in CoMFA, in which the lattice points inside or close to the molecular volume are neglected.

Validation. Nine of the 11 derived CoMFA and CoMSIA models with q^2 greater than 0.5 and the fraction of the variance superior to 0.85 were used to predict the activity of the compounds in the two testing sets (Tables 9, 10). The mean and standard deviation (SD) values for the deviation between the observed and predicted activities, as well as the predictive r^2 (r^2_{pred},

Table 10. Schultz et al. Testing Set Predicted CDK1 pIC_{50} Derived from Both CoMFA and CoMSIA Models, in Decreasing Order of pIC_{50}

Schultz compd[a]	CDK1 pIC_{50}	predicted pIC_{50}		Schultz compd[a]	CDK1 pIC_{50}	predicted pIC_{50}	
		CoMFA	CoMSIA			CoMFA	CoMSIA
095	7.70	6.94	7.07	**028**	6.25	6.77	6.35
096	7.52	6.67	7.16	**076**	6.22	6.47	6.15
097	7.48	6.94	7.08	**110**	6.21	6.40	6.32
060	7.46	7.01	7.14	**118**	6.15	6.43	6.16
073	7.40	7.14	7.27	**359**	6.10	6.43	6.60
098	7.28	6.92	7.00	**059**	6.10	5.73	5.78
356	7.10	6.94	7.12	**309**	6.05	6.12	6.08
094	7.00	6.78	6.94	**037**	6.00	5.81	5.77
036	7.00	6.78	6.86	**013**	5.92	6.04	5.95
033	6.89	6.86	6.83	**209**	5.92	6.26	6.04
212	6.80	6.09	6.09	**226**	5.92	6.70	6.97
112	6.77	6.61	6.72	**015**	5.89	5.93	5.81
010	6.68	5.82	6.56	**223**	5.89	5.74	6.05
052	6.66	6.26	6.19	**219**	5.89	5.78	6.18
304	6.66	5.98	5.86	**306**	5.85	6.43	6.60
303	6.64	6.07	6.17	**224**	5.82	6.00	6.16
216	6.64	6.25	6.12	**041**	5.80	6.67	6.26
075	6.64	7.20	7.13	**061**	5.64	6.04	6.21
016	6.62	6.11	6.14	**054**	5.57	5.88	6.01
078	6.62	6.94	7.05	**050**	5.55	5.77	5.44
211	6.60	6.39	6.15	**049**	5.55	6.67	6.67
045	6.57	6.30	6.27	**222**	5.52	6.30	6.19
064	6.54	6.10	6.14	**106**	5.40	6.49	6.45
307	6.52	6.73	6.76	**038**	5.40	6.25	6.60
068	6.52	6.89	6.60	**043**	5.37	5.92	5.95
026	6.48	6.02	5.61	**042**	5.37	6.20	6.12
305	6.46	5.92	5.92	**069**	5.30	5.76	5.53
077	6.44	6.86	7.09	**302**	5.30	5.70	5.72
065	6.40	6.18	6.10	**040**	5.30	6.00	5.67
066	6.40	6.58	6.61	**067**	5.30	6.15	5.61
051	6.38	6.26	6.31	**113**	5.22	6.30	6.32
047	6.37	6.54	6.29	**027**	5.19	6.08	6.01
318	6.36	6.10	5.74	**011**	5.15	5.88	6.30
314	6.35	6.44	6.50	**310**	5.15	5.11	5.31
058	6.30	6.13	6.20	**046**	5.05	6.00	6.09
044	6.30	6.23	6.08	**225**	5.05	6.32	6.89
070	6.30	5.82	5.95	**053**	5.00	5.81	6.00
071	6.30	6.40	6.62	**227**	5.00	5.64	6.02
062	6.30	6.68	6.75	**012**	4.97	5.58	5.73
308	6.28	6.56	6.64				

[a] The structure numbers corresponding to the compounds can be found in Schultz's paper.[5]

eq 1), were evaluated (Table 9):

$$r^2_{pred} = 1 - \Sigma(pIC_{50}^{obsd} - pIC_{50}^{pred})^2 / \Sigma(pIC_{50}^{obsd} - pIC_{50}^{mean})^2 \quad (1)$$

The best predictions using CoMFA were obtained with the STD fields and the combined steric, electrostatic, and hydrogen bond fields (STD+HBND), whereas the combined standard and indicator fields (STD+IND) gave slightly less accurate predictions. It is noteworthy that the indicator fields alone (IND) produced the least accurate predictions, even though they gave the highest r^2. This confirms our initial assumption that r^2 is overestimated in this case. The best predictions using CoMSIA were obtained with the combined steric, electrostatic, and lipophilic fields (ST+EL+LIPO). In contrast, the worst predictions were obtained for the combined steric, electrostatic, and hydrogen bond fields (ST+EL+HBND). Interestingly, combining the five fields (ALL) resulted in slightly lower predictions when compared to the ST, EL, and LIPO combined fields. This demonstrates that, although the hydrogen bond field contributes essentially nothing, it does not produce any

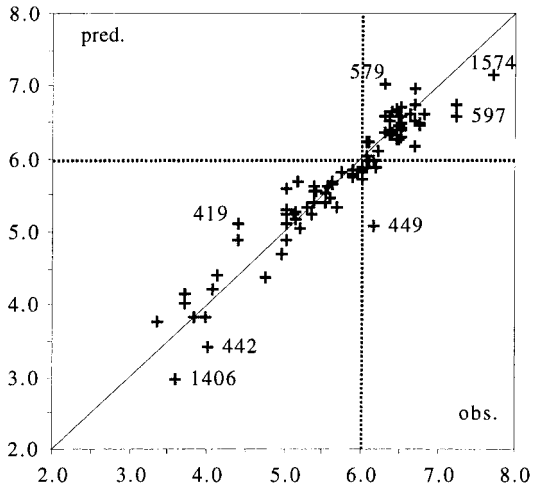


Figure 2. CDK1 pIC_{50} predicted values vs observed ones for the training set (+) derived from the CoMFA model.

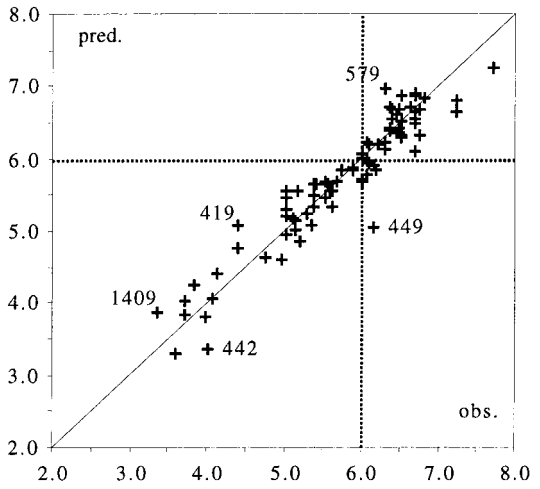


Figure 3. CDK1 pIC_{50} predicted values vs observed ones for the training set (+) derived from the CoMSIA model.

perturbation. From this point on, CoMFA will refer to the standard (STD) model, whereas the CoMSIA will refer to the combined steric, electrostatic, and lipophilic fields model (ST+EL+LIPO).

A direct plot of predicted vs observed activities using CoMFA (Figure 2) and CoMSIA (Figure 3) confirmed that the level of predictiveness for the two models is similar. In the CoMFA model, two compounds (**419** and **579**) are overpredicted by more than 0.5 log unit (pIC_{50}) and five compounds (**1406**, **442**, **449**, **597**, and **1574**) are underpredicted by at least 0.5 log unit. For **449** this can be rationalized by the fact that it is the only active compound carrying a group at N-9 which is bulkier than *i*-Pr. Compound **1574** is poorly predicted since it is the only C-2 valinol compound in the training set. In the testing set plot, it is striking that the CoMFA predictiveness is slightly higher than that for CoMSIA, particularly for compound **410** (Figure 4). Similarly, the direct plot of predicted vs observed activities for the Schultz series (Table 10) clearly illustrates the high level of predictiveness for both the CoMFA (Figure 5) and CoMSIA (Figure 6) models. Indeed, although the predicted activities of several compounds are significantly different from the observed pIC_{50}'s, almost all the molecules, and in particular the most active ones, are well-predicted (Table 10). Furthermore, as deduced from

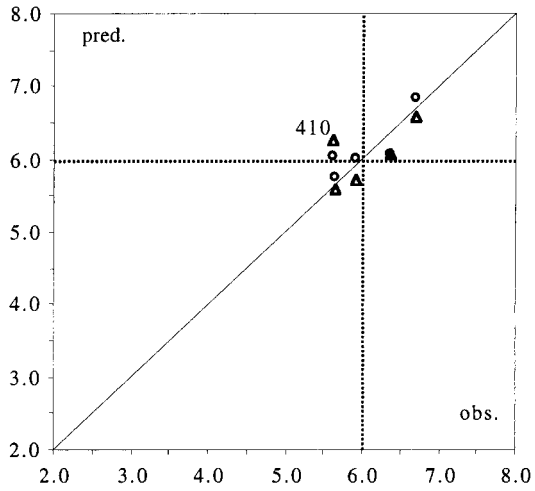


Figure 4. CDK1 pIC_{50} predicted values vs observed ones for the testing set using CoMFA (△) and CoMSIA (○) derived models.

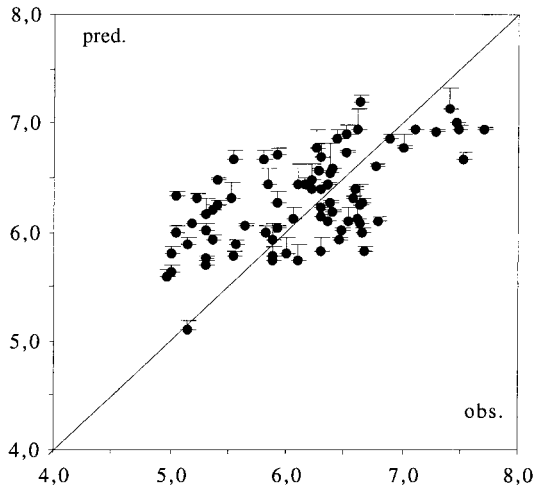


Figure 5. CDK1 pIC_{50} predicted values vs observed ones for the Schultz et al. testing set using CoMFA (●) derived model. Error margin represents the sum of extrapolated descriptor values.

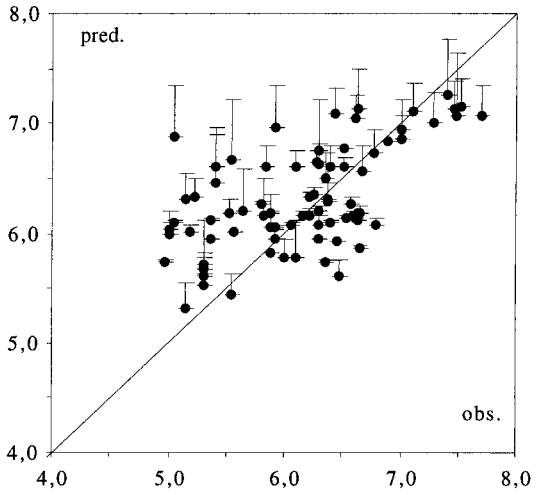


Figure 6. CDK1 pIC_{50} predicted values vs observed ones for the Schultz et al. testing set using CoMSIA (●) derived model. Error margin represents the sum of extrapolated descriptor values.

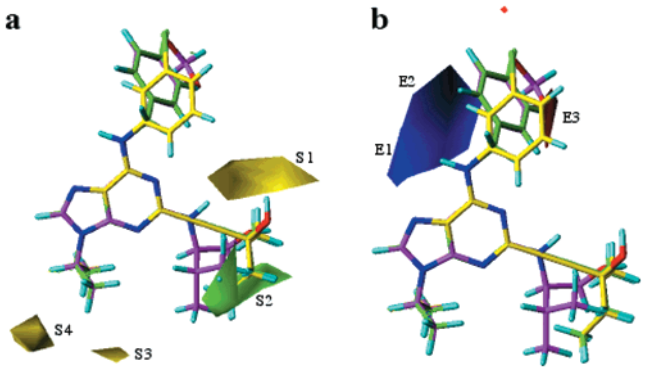


Figure 7. (a) CoMFA steric field. Green indicates regions where bulky groups increase activity; yellow indicates regions where bulky groups decrease activity. (b) CoMFA electrostatic field. Blue indicates regions where positive charges increase activity or negative charges decrease it; red indicates regions where negative charges increase activity or positive charges decrease it.

the sum of the extrapolated descriptor values (not shown), the weakly predicted compounds display features in some regions that are unexplored by our model. These derivatives are consequently outliers. Moreover, the prediction is more difficult for the racemic compounds in this data set (undefined C-2 side chain sterochemistry) since the measured activity is the mean of (*S*) and (*R*) enantiomer activities. This results in a lower predictiveness for both the CoMFA ($r^2_{pred} = 0.81$), and CoMSIA ($r^2_{pred} = 0.78$) models. However, considering the wide diversity of the C-2 substituents, and the problems associated with testing racemic compounds, the predictiveness remains good.

The 3D-QSAR contour maps from both CoMFA (Figure 7) and CoMSIA (Figure 8) treatments illustrate clearly the steric, electrostatic, and lipophilic requirements for ligand binding. In the CoMFA analysis, sterically unfavorable regions are depicted in yellow, whereas favorable regions are in green. Electrostatic positive favorable (or negative unfavorable) regions correspond to the blue areas, and the negative favorable (or positive unfavorable) regions correspond to the red areas. Lipophilic (CoMSIA only, Figure 8c) favorable regions are colored in orange, whereas hydrophilic favorable regions are cyan-colored.

Analysis of the CoMFA steric 3D-map (Figure 7a) shows that bulky groups at N-9 (S3−S4 regions), including cyclohexyl (**453**), *tert*-butyl (**515**), and benzyl (**443**), and bulky groups on the upperside of the C-2 side chain (S1) lead to a decrease in activity. In contrast, medium-sized substituents on the lower part of the C-2 motif (S2), such as ethyl, isopropyl, or pyrrolidine, lead to an enhancement in activity. This analysis provides, among other things, an explanation for the low activity of the iodo derivatives (C-2 I) which overlap the S1 yellow area and do not fill the green zone (S2). In the CoMSIA treatment the conterparts to the zones S1−S4 are S2, S3, S5, and S6, respectively. However, in the CoMSIA steric 3D-map, there are also two additional regions: the first being a favorable zone for small substituents on the *meta* and *para* positions of the N-6 benzyl ring (S1) and the second being an unfavorable zone for substituents on the lower part of the C-2 side chain (S4).

The CoMFA electrostatic fields provided relatively little information (Figure 7b), as was expected from the

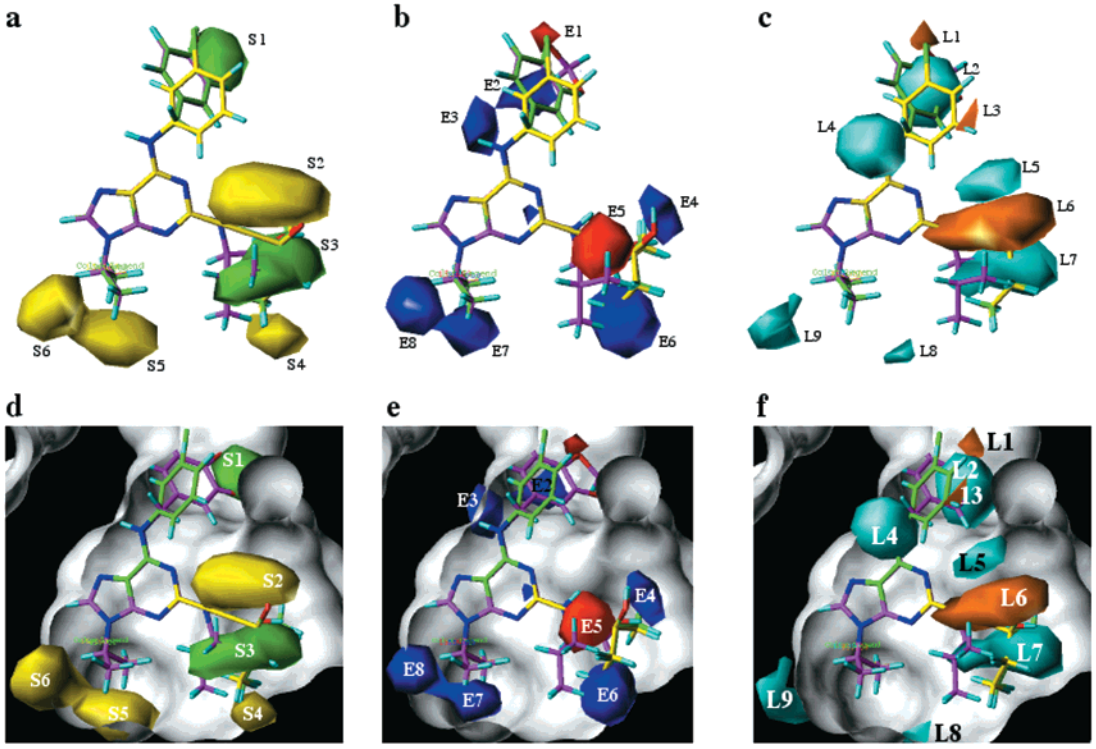


Figure 8. (a) CoMSIA steric fields. Green indicates regions where bulky groups increase activity; yellow indicates regions where bulky groups lead to decreasing activity. (b) CoMSIA electrostatic fields. Blue indicates regions where positive charges increase activity or negative charges decrease it; red indicates region where negative charges increase activity or positive charges decrease it. (c) CoMSIA lipophilic field. Orange indicates regions where hydrophobic substituents enhance activity; cyan indicates regions where hydrophilic substituents increase activity. (d) CoMSIA steric favorable (green) and unfavorable (yellow) fields projected inside the ATP binding site of CDK2. (e) CoMSIA electrostatic favorable positive (blue) and negative (red) fields projected inside the ATP binding site of CDK2. (f) CoMSIA hydrophobic favorable (orange) and unfavorable (cyan) fields projected inside the ATP binding site of CDK2.

reduced influence of electrostatics on the total variance. The blue region close to N-6 can be attributed to a favorable positively charged hydrogen on N-6 (E1) and to unfavorable negatively charged substituents on the aromatic ring of the benzyl motif (upper part, E2). Similarly, the very small red region close to the benzyl *para* position results from the favorable influence of a negative charge in this area. However, the role of the blue region between the phenyl and benzyl groups (E3) is less clear.

In the CoMSIA 3D-plot (Figure 8b), the red region over the C-2 side chain (E5) illustrates the importance of the negative charge in this area, originating either from the nitrogen atom in the pyrrolidino derivatives in family 5 or from the π-electrons of the acetylene function in derivatives in family 1. Similarly, it also shows the importance of the hydroxyl group in this side chain (upper blue region, E4). The lower blue zone (E6) can be interpreted to be an unfavorable negative region. Similarly, the blue region below the N-9 substituent (E7−E8) may be attributed to the unfavorable negative charge from the pentenyl (**1406**) and benzyl (**442** and **443**) substituents.

The hydrophobic field (Figure 8c) in CoMSIA brings to light the importance of the hydrophilic N-6 H and the C-2 side chain hydroxyl substituents (regions L4 and L5, respectively), as well as the importance of the hydrophobic component of the C-2 side chain (L6) and the presence of polar substituents on the *meta* and *para* positions of the N-6 benzyl ring (L1, L3). In addition, a hydrophobic unfavorable region occurs close to N-9 (L8−L9), originating from the benzyl group in **442** and **443** and the cyclohexyl group in **453.** These analogues are all weakly active. It also shows two favorable hydrophobic regions, one of which is close to the *meta* position of the aromatic ring. The first zone was revealed by the low activity of **497** bearing a hydroxyl group in this area, whereas the other is close to the benzyl *para* positions and more difficult to interpret.

Although the hydrogen bond fields were not taken into account in the CoMSIA model, the comparison of the electrostatic and lipophilic fields reveals the importance of the N-6 hydrogen (Figure 8b,c). However, from the steric fields 3D-map it is clear that the N-2 hydrogen is not involved in a direct hydrogen bond with the protein. Indeed small aminoalkyl substituents are tolerated on this position, as demonstrated by the activities observed for derivatives in the pyrrolidino series (e.g. **526**). By analogy to the results for the **567**/CDK2 crystal structure, this region of CDK1 may be occupied by a water molecule bound to Asp86 and the C-2 side chain hydroxyl group.

To summarize, the hydroxyl group on the C-2 side chain is seen to be important for activity, and although it occupies different positions in the different analogues, it should remain proximal to carbon C-2. Furthermore, although the presence of a proton on the nitrogen atom in this side chain seems to enhance activity, neither the NH proton nor the nitrogen atom itself is necessary for activity. This is demonstrated by the activities observed for the acetylene family of molecules. A hydrophobic substituent on the C-2 side chain greatly increases the

activity (*i*-Pr > Et ≫ Me > H). Concerning the N-9 position, there is a preference for medium-sized alkyl substituents (*i*-Pr > Me). Furthermore, considering the weak activity for the cyclohexyl-substituted analogue **452**, it appears that a cyclopentyl ring, as found in **449**, probably represents the upper limit in steric volume.

Comparison with the CDK2 Binding Site. Because our alignment rules are based on CDK2 docking studies and on X-ray crystallography structures of CDK2 complexed to several ligands, we compared the CDK1 QSAR model with the CDK2 binding site by superposing them using the ligand purine core. Indeed, CDK1 differs from CDK2 by changes in two essential amino acids that interact directly with olomoucine, roscovitine, and **567** in the ATP binding site: Ser84 for His and Met85 for Gln. From the CDK2 X-ray data it appears that His84 plays a crucial role in the positioning of the phenyl ring in Phe82. Therefore one might expect the Phe82−Ser86 loop conformation to be affected, as well as the Lys20 and Glu8 side chains. The results of this analysis suggest that analogues with a 3′-pyrrolidine, 3-hydroxycyclopent-1-yl, or alaninol substituent at N-9 may display activity, since these substituents could take the place of the water molecule which is otherwise located in this vicinity. This suggestion agrees with what is observed in both the CoMSIA steric and electrostatic 3D-maps.

In Figure 8d−f it is noteworthy that the steric unfavorable regions (yellow areas, such as S4−S6) penetrate beyond the available binding pocket into the interior of the protein, whereas the two main favorable regions (green areas) are located in the pocket. The electrostatic regions also fit nicely in the CDK2 structure. For instance, the positive region E3 interacts with the Leu83 carbonyl group, the zone E7 with the carboxylate group of Asp145 (via a water molecule), and E8 with the π-electrons of Phe80. One can also speculate that Glu81 (conserved in all families of CDK) orients into the solvent-accessible pocket next to Phe80, thus accounting for the predicited E8 site.[35] Similarly, lipophilic and hydrophilic favorable areas derived from the CoMSIA model clearly accommodate the corresponding region in the binding site (Figure 8f, such as L2, L4, and L5, where water molecules are found in the CDK2 crystal structure). This validates the chosen alignment rules, especially for substituents at C-2.

Comparison of CDK1 and CDK5 Binding Activities. Compared to CDK1, there are three important amino acid changes in the ATP binding site in CDK5 that are involved in inhibitor binding (Cys83 vs Leu, Asp84 vs Ser, Gln85 vs Met), and relative to CDK2 there are two principal modifications (Figure 1) in the same region (Cys83 vs Leu, Asp84 vs His). As for CDK1, the presence of Asp84 (vs His84 in CDK2) probably affects the shape of the binding site close to the *ortho*′ and *meta*′ positions in the benzyl substituent in **567** and the *ortho* phenyl position in **597**. Unfortunately, the lack of IC_{50} values for 50 of the 93 purine-based analogues prevented us from deriving a 3D-QSAR model for CDK5. Thus, to vizualize in 3D the CDK1/CDK5 selectivity, we employed the intersection and union volume method.

The direct plot of CDK5 vs CDK1 activities (Figure 9) shows that most of our compounds have similar IC_{50}'s for both enzymes. However, two trends do emerge. First,

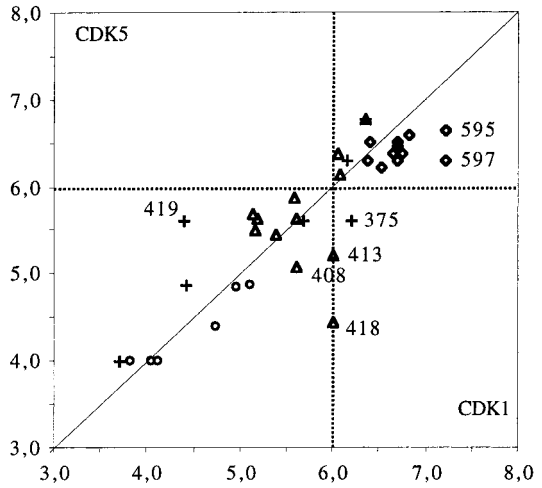


Figure 9. pIC_{50} observed values for CDK5 vs CDK1 plot: (○) iodo series; (◇) acetylenic series; (+) amino series; (△) pyrrolidino series.

except for compounds **375**, **408**, **413**, and **418**, the amino (△) and pyrrolidino (+) series compounds inhibit CDK5 more strongly than CDK1. Note that the marked difference in activities for the latter two molecules reflects the capacity of CDK5 to differentiate between their two enantiomers. Second, the acetylene (◇) based compounds (e.g. **597** and **595**) are both more potent and more selective inhibitors of CDK1. The iodo family of compounds was excluded from this analysis since very few of these compounds were tested against CDK5, and those that were (○) displayed no activity.

It was further found for the compounds bearing the 2-amino side chain that increasing steric bulk of the substituent next to the nitrogen correlates with a stronger inhibition of CDK5 relative to CDK1. This suggests that these molecules adopt a somewhat different orientation in the protein due to a wider hydrophobic pocket in the region about the benzyl group. This hypothesis correlates with the increased affinity of **453** and **515** where the bulky N-9 substituent pushes these molecules toward the entrance of the cleft.

For the acetylene-containing analogues, the contrastingly lower activity for CDK5 vs CDK1 (the largest differences being observed for **597** which is 8.3-fold less active on CDK5) may result from a slightly different conformation for the Phe82−Asp86 loop. This may be especially true for the *meta*- vs *para*-substituted aniline compounds, but less so for the different N-6 benzyl derivatives, for which the difference in activities remains essentially constant.

Conclusion

From a technical viewpoint, it is noteworthy from our study that CoMSIA region shapes (Figure 8) are smoother and closer to the involved atoms or functionalities than the corresponding regions in the CoMFA treatment (Figure 7). The CoMSIA maps are for this reason easier to interpret. However, although CoMSIA is less alignment-sensitive than CoMFA and produces smoother 3D-QSAR contour maps, its predictiveness is significantly less accurate for compounds with substituents outside the regions explored by the model. This results from the use of lattice points closer to the molecule area as compared to CoMFA. In contrast, the

lack of such points in CoMFA enlarges its explored region and therefore reduces the number of extrapolated points. Furthermore, CoMSIA required a smaller grid spacing, thereby improving both q^2 and r^2. However, there is a correspondingly large increase in the calculation time. CoMSIA is thus a versatile visual tool which is highly complementary to the CoMFA method.

The CoMFA and CoMSIA methods were successfully used in our study to build 3D-QSAR models that define accurately the molecular basis for the inhibition of CDK1 by purine derivatives. The comparison of these models with the structure of CDK2, which on the basis of homology is very similar to CDK1,[12] validated the alignment rules that we employed. Furthermore, the use of the external set of compounds described by Schultz et al.[5] both validated and defined the limits of our model.

The CoMFA and CoMSIA 3D-maps can thus be used to design new inhibitors and, in an interactive fashion, to predict their affinity and assess their selectivity. From both our work[19−22] and the efforts of others,[5,7,36] a considerable body of biological data has been accumulated concerning the 6-anilino series, essentially all of which is consistent with our model. Less effort has been addressed to SAR studies in the 6-benzylamino series, and in this context, the QSAR models suggest interesting avenues for future exploration.

For instance, one can infer that polar substituents such as an amino or hydroxy group in the *para* position of the 6-benzylamino substituent should enhance the inhibitory activity of the C-2 amino- and acetylene-based families of purine derivatives. Similarly, the presence of polar substituents on the *meta*/*para* positions of the aromatic ring or on the *meta* position alone should lead to an increase in activity. Concerning the anilino series, one can expect that the simultaneous presence of polar groups at the *meta* and *meta′* positions will lead to improved CDK1 inhibition.

Furthermore, the comparison of the CoMSIA-derived 3D-maps and CDK2 crystal structure revealed unexplored regions in the protein active site, suggesting the design of new molecules bearing polar substituents at N-9 (e.g. (*S*)-CH(CH_3)CH_2OH, (*S*)-3-hydroxycyclopentan-1-yl, etc.). Finally in the acetylene series, the π-electrons of the acetylene group hydrogen-bond to a water molecule (also interacting with N-1) and the hydroxyl group is located close to the magnesium ion, which is found in the CDK2/ATP X-ray crystal structure. This motif thus mimics nicely the ribose sugar component in ATP. This offers a considerable number of possibilities for the design of inhibitors bearing functionality which interacts with the triphosphate binding region, through chelation to the magnesium ion.The synthesis and evaluation of new purine-based CDK inhibitors designed using the present model will be reported in a future communication.

Acknowledgment. This work was supported in part by a grant (No. 9814) from the Association pour la Recherche sur le Cancer (ARC, France). We also thank the Centre de Ressources Informatiques de Haute Normandie (CRIHAN) for calculation facilities and support.

References

(1) Morgan, D. O. Principles of CDK regulation. *Nature* **1995**, *374*, 131−134.
(2) Morgan, D. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 261−291.
(3) Gray, N.; Detivaud, L.; Doerig, C.; Meijer, L. ATP-site directed inhibitors of cyclin-dependent kinases. *Curr. Med. Chem.* **1999**, *6*, 859−875.
(4) 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.
(5) Chang, Y.-T.; Gray, N. S.; Rosania, G. R.; Sutherlin, D. P.; Kwon, S.; Norman, T. C.; Sarohia, R.; Leost, M.; Meijer, L.; Schultz, P. G. Synthesis and application of functionally diverse 2,6,9-trisubstituted purine libraries as CDK inhibitors. *Chem. Biol.* **1999**, *6*, 361−375.
(6) De Azevedo, W. F.; Leclerc, S.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S. H. Inhibition of cyclin-dependent kinases by purine analogues − Crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.* **1997**, *243*, 518−526.
(7) Imbach, P.; Capraro, H. G.; Furet, P.; Mett, H.; Meyer, T.; Zimmermann, J. 2,6,9-trisubstituted purines: Optimization towards highly potent and selective CDK1 inhibitors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 91−96.
(8) Hoessel, R.; Leclerc, S.; Endicott, J.; Noble, M.; Lawrie, A.; Tunnah, P.; Leost, M.; Damiens, E.; Marie, D.; Marko, D.; Niederberger, E.; Tang, W.; Eisenbrand, G.; Meijer, L. Indirubin, the active constituent of a chinese antileukeamia medicine, inhibits cyclin-dependent kinases. *Nat. Cell Biol.* **1999**, *1*, 60−67.
(9) Schultz, C.; Link, A.; Leost, M.; Zaharevitz, D. W.; Gussio, R.; Sausville, E. A.; Meijer, L.; Kunick, C. Paullones, a series of cyclin-dependent kinase inhibitors: Synthesis, evaluation of CDK1/cyclin B inhibition, and in vitro antitumor activity. *J. Med. Chem.* **1999**, *42*, 2909−2919.
(10) Kent, L. L.; Hull Campbell, N. E.; Lau, T.; Wu, J. C.; Thompson, S. A.; Nori, M. Characterization of novel inhibitors of cyclin-dependent kinases. *Biochem. Biophys. Res. Commun.* **1999**, *260*, 768−774.
(11) Green, S. L.; Vulliet, P. R.; Pinter, M. J.; Cork, L. C. Alterations in cyclin-dependent protein kinase 5 (CDK5) protein levels, activity and immunocytochemistry in canine motor neuron disease. *J. Neuropathol. Exp. Neurol.* **1998**, *57*, 1070−1077.
(12) De Bondt, H. L.; Rosenblatt, J.; Jancarik, J.; Jones, H. D.; Morgan, D. O.; Kim, S.-H. Crystal structure of cyclin-dependent kinase 2. *Nature* **1993**, *363*, 595−602.
(13) Schulze-Gahmen, U.; Brandsen, J.; Jones, H. D.; Morgan, D. O.; Meijer, L.; Vesely, J.; Kim, S.-H. Multiple modes of ligand recognition: Crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins Struct. Funct. Genet.* **1995**, *22*, 378−391.
(14) Lawrie, A. M.; Noble, M. E. M.; Tunnah, P.; Brown, N. R.; Johnson, L. N.; Endicott, J. A. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. *Nat. Struct. Biol.* **1997**, *4*, 796−801.
(15) De Azevedo, W. F.; 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.
(16) Gray, N. S.; Wodicka, L.; Thunnissen, A. M. W. H.; Norman, T. C.; Kwon, S. J.; Espinoza, F. H.; Morgan, D. O.; Barnes, G.; LeClerc, S.; Meijer, L.; Kim, S. H.; Lockhart, D. J.; Schultz, P. G. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **1998**, *281*, 533−538.
(17) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massagué, J.; Pavletich, N. P. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **1995**, *376*, 313−320.
(18) Russo, A. A.; Jeffrey, P. D.; Patten, A. K.; Massagué, J.; Pavletich, N. P. Crystal structure of the $p27^{Kip1}$ cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **1996**, *382*, 325−331.
(19) Legraverend, M.; Ludwig, O.; Bisagni, E.; Leclerc, S.; Meijer, L. Synthesis of C2 alkynylated purines, a new family of potent inhibitors of cyclin-dependent kinases. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 793−798.
(20) Legraverend, M.; Ludwig, O.; Bisagni, E.; Leclerc, S.; Meijer, L.; Giocanti, N.; Sadri, R.; Favaudon, V. Synthesis and in vitro evaluation of novel 2,6,9-trisubstituted purines as cyclin-dependent kinases inhibitors. *Bioorg. Med. Chem.* **1999**, *7*, 1281−1293.
(21) Legraverend, M.; Tunnah, P.; Noble, M.; Ducrot, P.; Ludwig, O.; Grierson, D. S.; Leost, M.; Meijer, L.; Endicott, J. Cyclin-dependent kinase inhibition by C-2 alkynylated purine derivatives and molecular structure of a CDK2-inhibitor complex. *J. Med. Chem.* **2000**, *43*, 1282−1292.

(22) Legraverend, M.; Ludwig, O.; Leclerc, S.; Meijer, L. Synthesis of a new series of purine derivatives and their anti-cyclin-dependent kinase activities. *J. Heterocycl. Chem.* **2000**, in press.
(23) *Sybyl*, 6.3 ed.; SYBYL molecular modeling software, Tripos Associates Ltd.: St. Louis, MO, 1992.
(24) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(25) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.
(26) Klebe, G.; Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput.-Aid. Mol. Des.* **1999**, *13*, 1−10.
(27) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982−1012.
(28) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative molecular field analysis (CoMFA). 2. Towards its use with 3D-structural databases. *Tetrahedron Comput. Methods* **1990**, *3*, 47−59.
(29) Stewart, J. J. P. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aid. Mol. Des.* **1990**, *4*, 1−103.
(30) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindberg, W.; Sjöström, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dortrecht, The Netherlands, 1984; pp 17−95.
(31) Dunn, W. J.; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. Multivariate structure−activity relationship between data from a battery of biological tests and an ensemble of structure descriptors: The PLS methodology. *Quant. Struct.-Act. Relat. Chem. Biol.* **1984**, *3*, 31−137.
(32) Geladi, P. Notes on the history and nature of partial least squares (PLS) modelling. *J. Chemom.* **1988**, *2*, 231−246.
(33) Wold, S. Cross-validatory estimation of the number of component in factor and principal component models. *Technometrics* **1978**, *4*, 397−405.
(34) (a) Diaconis, P.; Efron, B. Computer-intensive methods for statistics. *Sci. Am.* **1984**, *116*, 96−117. (b) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. Crossvalidation, bootstrapping and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.−Act. Relat.* **1988**, *7*, 18−25.
(35) Relevent argument forwarded by the referees.
(36) Zimmerman, J.; Capraro, H. G.; Peterli, P.; Furet, P. Purine derivatives and processes for their preparation. Patent WO 97/716452, 1997.

JM000965T