Articles

Structural Classification of Protein Kinases Using 3D Molecular Interaction Field Analysis of Their Ligand Binding Sites: Target Family Landscapes

Thorsten Naumann* and Hans Matter

Aventis Pharma Deutschland GmbH, DI&A Chemistry, Molecular Modeling, D-65926 Frankfurt am Main, Germany

Received July 30, 2001

Protein kinases are critical components of signaling pathways and trigger various biological events. Several members of this superfamily are interesting targets for novel therapeutic approaches. All known eukaryotic protein kinases exhibit a conserved catalytic core domain with an adenosine 5′-triphosphate (ATP) binding site, which often is targeted in drug discovery programs. However, as ATP is common to kinases and other proteins, specific protein−ligand interactions are crucial prerequisites for valuable ATP site-directed ligands. In the present study, a set of 26 X-ray structures of eukaryotic protein kinases were classified into subfamilies with similar protein−ligand interactions in the ATP binding site using a chemometrical approach based on principal component analysis (PCA) and consensus PCA. This classification does not rely on protein sequence similarities, as descriptors are derived from three-dimensional (3D) binding site information only computed using GRID molecular interaction fields. The resulting classification, which we refer to as "target family landscape", lead to the identification of common binding pattern and specific interaction sites for particular kinase subfamilies. Moreover, those findings are in good agreement with experimental selectivity profiles for a series of 2,6,9-substituted purines as CDK inhibitors. Their interpretation in structural terms unveiled favorable substitutions toward selective CDK inhibitors and thus allowed for a rational design of specific ligands with minimized side effects. Additional 3D-quantitative structure−activity relationship (QSAR) analyses of a larger set of CDK-directed purines lead to the identification of essential structural requirements for affinity in this CDK ATP binding site. The combined interpretation of 3D-QSAR and the kinase target family landscape provides a consistent view to protein−ligand interactions, which are both favorable for affinity and selectivity in this important subfamily.

1. Introduction

Large protein families such as protein kinases or proteases have been identified during the past decade containing many pharmaceutically relevant targets. The eukaryotic protein kinases form a large superfamily of homologous proteins,[1] related by virtue of their kinase domains (also known as catalytic domains), which encompass 250−300 amino acids,[2−4] and exhibit a conserved catalytic core structure, as obvious from protein X-ray structures.[5] Although there is a rich diversity of structures, regulation modes, and substrate specificities in this family, there are common features such as the catalytic core, indicating how kinases transfer phosphate to serine, threonine, or tyrosine residues of their protein substrates. The reversible regulation of protein phosphorylation by kinases and phosphatases is essential for controlling and triggering a broad variety of cellular events, e.g., cellular signaling pathways and metabolic functions.[6,7] Protein kinases themselves are regulated by a variety of mechanisms including phosphorylation and control by additional regulatory domains.

Selective inhibitors of kinases being explored as new therapeutic targets in several pharmaceutical companies may have value in a wide range of diseases, such as cancer, diabetes, and arthritis.[8] In particular, inhibition by targeting the conserved adenosine 5′-triphosphate (ATP) binding site,[3,9,10] located in the kinase domain, is seen as a valuable approach. However, there are common recognition patterns for ATP in kinases and other ATP binding protein families, which underscore the importance of specificity for ATP site-directed kinase inhibitors. More than 2000 ATP utilizing proteins are estimated in the human genome.[11,12] Small kinase inhibitors with acceptable selectivity profiles have recently been reported.[13,14] Any successful inhibitor design for the conserved ATP binding side requires highly specific interaction in order not to interfere with other kinases and pathways. While some kinases are highly specific and will only recognize a few target molecules in vivo, others show a much broader specificity profile and, once activated, are able to phosphorylate multiple target proteins. The identification and structural classification of unique protein−ligand interactions for designing selective kinase inhibitors thus is a challenging task in structural biology and drug discovery.

* To whom correspondence should be addressed. Tel.: ++49-69-305-57182. Fax: ++49-69-331399. E-mail: thorsten.naumann@aventis.com.

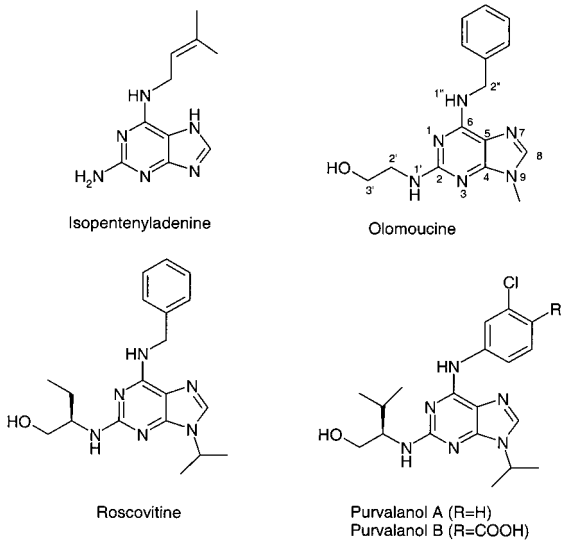
10.1021/jm011002c CCC: $22.00
Published on Web 05/07/2002

In this paper, a classification of 26 different protein kinases is described, based on three-dimensional (3D) structural information about their ATP binding site. The computational approach does not rely on protein sequence comparisons. In contrast, it is based on an adequate 3D description of potential protein−ligand interactions using the GRID force field.[15,16] This information combined with a powerful chemometrical approach can be used to extract common features among all protein kinase binding sites and highlight those areas that reveal most interaction pattern differences. Such a quantitative analysis of GRID-derived molecular interaction fields (MIFs) from kinase binding sites not only leads to a classification but also leads to an understanding of which structural features in terms of potential ligand interactions are responsible and most characteristic for a particular kinase subfamily. The availability of experimental 3D protein kinases structures allows for a quantitative description of selectivity differences using principal component analysis (PCA)[17] and consensus PCA (CPCA),[18,19] resulting in a consistent picture in good agreement with experimental data on ligand selectivity profiles and structure−activity relationships (SAR). This approach is related to previous investigations of the "protein selectivity problem" for targets such as DNA,[20] dihydrofolate reductase,[21] matrix metalloproteinases,[22,23] and cyclooxygenase,[24] while those only allowed to compare two proteins. The approach presented herein uses a different data organization and variable scaling[18] to overcome this limitation and identify specificity regions based on weaker hydrophobic interactions. This approach was first applied to understand selectivity in the factor Xa/thrombin/trypsin family.[18] We present here an application of this strategy to a broader set of proteins toward a structural classification, allowing the straightforward identification of possible modifications for selected ligands to improve their selectivity toward a chosen target kinase. Hence, this computational strategy to rationalize ligand selectivities is a valuable tool for designing kinase specific inhibitors with reduced side effects.

This approach will be illustrated in combination with ligand SAR data and 3D-quantitative SAR (QSAR) models (comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA)) for a reported series of cyclin-dependent kinase inhibitors,[25,26] where sterical, electronic, or hydrophobic effects are important to explain affinity toward CDKs. These analyses in combination with protein-derived selectivity models, which we refer to as "target family landscapes", provide valuable information toward the design of potent and selective CDK inhibitors.

The alignment rule for 3D-QSAR analyses of purine-based CDK inhibitors was derived using manual docking based on the available X-ray structure of purvalanol B (Scheme 1) in complex with CDK2.[27] The resulting superposition of all other compounds onto this template followed by energy minimization produced consistent QSAR models explaining the most relevant ligand−enzyme interactions. CoMFA[28−30] and CoMSIA[31,32] are used to derive statistical relationships between molecular property fields and biological activities. The PLS method (partial least squares)[33] is used to derive a linear relationship for highly underdetermined matrixes, while cross validation[34] is used to check for consistency and predictiveness. The contour maps from both 3D-QSAR models enhance the understanding of electrostatic, hydrophobic, and steric requirements for affinity, while the statistical protein structure analysis provides guidelines for designing selective CDK inhibitors.

Scheme 1. Chemical Structures of Selected Purine Derivatives as CDK Inhibitors



2. Experimental Section

2.1. Protein Structures. All modeling studies were done using the program Sybyl[35] on SGI workstations. Docking, QSAR analysis, and protein superposition steps were automated using scripts in PERL and SPL (Sybyl Programming Language). All energy calculations were based on the MMFF94s force field[36] using MMFF94 charges. Conformations of ligands and complexes were minimized using quasi-Newton−Raphson or conjugate gradient procedures. All 3D structures from X-ray crystallographic studies of protein kinases were retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) protein database.[37] For these files, cofactors, counterions, ligands, and structural water were removed prior to computing GRID-derived[16] MIFs. All protein structures were carefully checked for inconsistencies, missing side chains, or amino acids. Only consistent structures were retained, leading to a set of 26 kinases for statistical analysis, as summarized in Table 1.

2.2. Protein Alignment. The quality of the chemometrical results is dependent on the alignment strategy. Hence, an unbiased alignment strategy is preferable, as it allows to quickly add proteins for classification to the training set. After evaluating different approaches, the following strategy was adequate.

First, a multiple sequence alignment of all candidate proteins with known 3D structures was generated using the program Composer.[38] The identified seed residues served as the basis for 3D structural alignment, leading to the identification of structurally conserved regions (SCRs) among the entire protein ensemble. The X-ray structure of a cyclic adenosine 5′-monophosphate (cAMP)-dependent protein kinase with a peptide inhibitor 1ATP[39] was used as reference[40] for this and all following steps (resolution 2.2 Å). The SCRs were used for an initial pairwise root mean square fitting of corresponding backbone atoms of all proteins to the reference structure. In two fitting iterations, individual amino acid pairs with a deviation larger than 3 × SD from the previous alignment solution were rejected, leading to an unambiguous alignment with an acceptable superposition of important structural elements.

Table 1. Overview of 26 Experimental Kinase Structures with Cofactors and PDB Code Used for GRID/CPCA Study[a]

protein kinase	full length/domain	cofactor/ligand	PDB code	conformation
		Serine/Threonine Kinases		
cAPK	catalytic subunit, mouse 350 a.a.	adenosine	1BKX	active
	catalytic subunit, mouse 350 a.a.	PKI[b] and ATP	1ATP	acitve
	catalytic subunit, mouse 350 a.a.	PKI and ATP	1CDK	active
	catalytic subunit, mouse 350 a.a.	PKI and balanol	1BX6	active
	catalytic subunit, mouse 350 a.a.	PKI and staurostorine	1STC	active
	catalytic subunit, bovine 350 a.a.	PKI and isoquinoline	1YDT	active
	catalytic subunit, bovine 350 a.a.	PKI and isoquinoline	1YDS	active
		PKI and adenosine	1FMO	active
	catalytic subunit, bovine 350 a.a.	PKI and isoquinoline	1YDR	active
CDK2	full length 298 a.a	ATP	1HCK	inactive
	full length 298 a.a	purvalanol B	1CKP	inactive
	full length 298 a.a	ATP and cyclin	1FIN (chain A and B)	p. active
	full length 298 a.a	olomucine	1B38	inactive
in house		purine analogue	SHKB	inactive
in house		flavone	FLAV	inactive
MAP	full length, rat 364 a.a.		1ERK	inactive
	full length, rat mutation arginine a.a. 52		1GOL	inactive
	full length, 397 a.a.		1P38	inactive
phosphorylase kinase	kinase domain 298 a.a.	AMPPNP	1PHK	active
CK-1	truncated 298 a.a.	ATP	1CSN	active
		Receptor Tyrosine Kinases		
insuline receptor TK	kinase domain 306 a.a.	AMPPNP[c] and peptide substrate	1IR3	active
FGFIR	kinase domain 310 a.a.	oxindol inhibition	1FGI (chain A and B)	inactive
		Nonreceptor Tyrosine Kinases		
human c-Src	452 a.a.		2SRC	inactive
human lymph kinase	kinase domain 271 a.a.		3LCK	active

[a] For PDB structures with active or inactive kinase conformation, results from the final CPCA analysis were compared. No significant differences were identified. [b] PKI = protein kinase inhibitor peptide. [c] AMPPNP = not hydrolyzable ATP due to N linkage of β and γ phosphate.

Table 2. GRID Probes Used for GRID/PCA and GRID/CPCA Study

grid probe	chemical functionality
DRY	hydrophobic probe
O	sp2 carbonyl oxygen
N1	neutral flat NH, e.g., amide
C3[a]	sp3 methyl probe
OH2[a]	water

[a] Probe rejected for final GRID/CPCA analysis.

2.3. MIFs. A multivariate description of the binding cavities for aligned protein kinases was done using the GRID force field[15,16] with five probes (Table 2), 1 Å grid spacing, and static protein treatment (GRID directive MOVE = 0). Hydrogens were added to the aligned structures using the program GRIN. The GRID box dimensions were selected to include all relevant parts of the kinases active sites. Preliminary chemometrical analyses using CPCA[18] revealed that the GRID water and methyl probes contain similar information to other probes (CPCA superweights plot). Consequently, those probes were rejected for the final data analysis.

The data from GRID MIFs were organized as indicated in Figure 1 for the chemometrical analysis.[18] Starting from the GRID interaction field for one single probe and one protein, a vector is constructed. Subsequent interaction fields for other probes are concatenated to this vector to result in a long vector containing x GRID $\times$ n probes points. Similar vectors were derived for all protein kinases. At the end, the resulting X matrix prior to analysis has one single row for each of the 26 protein kinase 3D structures and $x \times n$ columns (i.e., GRID interaction points), as shown in Figure 1.

2.4. Chemometrical Analysis. This X matrix was analyzed using PCA[17] and CPCA,[18,19] as implemented in the program GOLPE.[41,42] A maximum cutoff value of 0 kcal/mol was used. When using only favorable interaction energies, the information related to steric repulsion is removed and the analysis is focused on attractive protein−ligand interactions. PCA studies including repulsive interactions do not qualitatively change the classifications (no data shown). However, as we were interested in extracting favorable interactions for specificity, we chose only to incorporate attractive data points for final analysis. In addition, only GRID points within a radius of 4.0 Å around the ATP binding area were considered using the GOLPE region cutout tool for appropriate variable selection. Furthermore, 2- and 3-level variables and columns with an SD below 0.03 as threshold were rejected.[41]

The chosen data organization allows to apply block unscaled weights (BUW) scaling in order to normalize the importance of probe interactions in the final PCA and CPCA model, which was not possible following the original data setup.[18,41] This method scales each single probe−protein interaction field separately, whereas the relative scales of variables within each block remain unchanged. This procedure is shown in Figure 2 with the initial GRID interaction energy distribution of the x variables for each probe and their normalized distribution after scaling.

In PCA,[17] the scaled data matrix X with interaction energies x_{ik} for i protein interactions and k grid points is decomposed to means (x_k), scores (t_{ia}), loadings (p_{ak}), and residuals (e_{ik}),

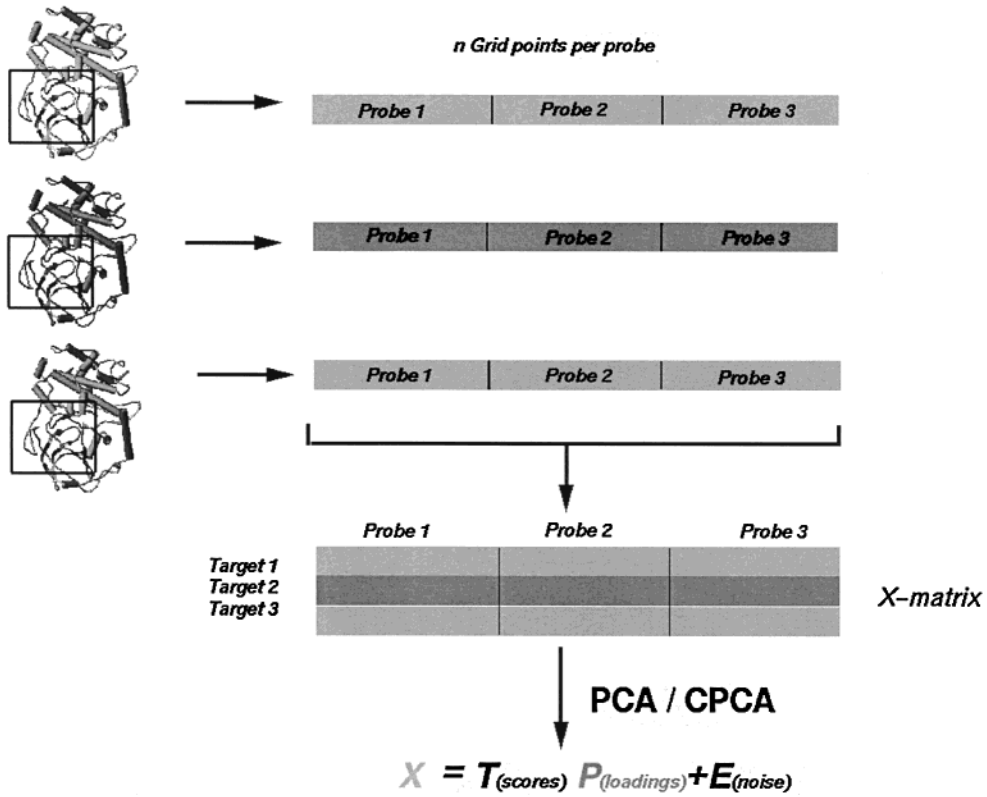


Figure 1. Data organization of GRID MIFs for PCA and CPCA.[18] Starting from GRID calculations for one probe, a vector containing all interaction energies at the *k* grid points is constructed. These vectors are compiled into one long vector containing protein−probe interaction energy points. Stacking the long vectors for every target protein results in the final X matrix for PCA.

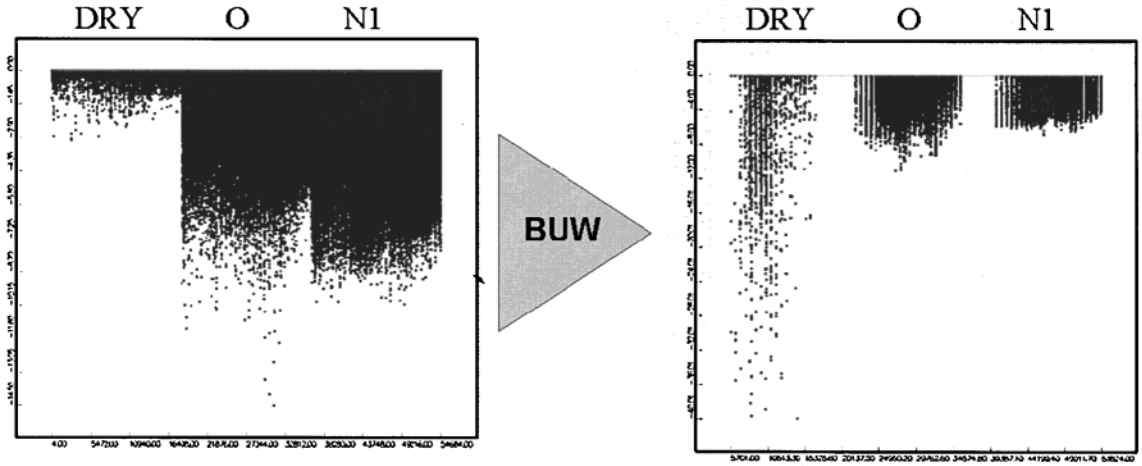


Figure 2. Raw (a) and normalized (b) GRID-derived probe interaction energies using BUW to unit variance for each probe.

with *a* denoting the number of model dimensions

$$x_{ik} = x_k + \sum_{a=1}^{A} t_{ia} \times p_{ak} + e_{ik}$$

The data matrix X is approximated by the product of two smaller matrixes, scores, and loadings. The score matrix p_{ak} gives a simplified picture of the objects (proteins interacting by multiple probes), represented by new uncorrelated variables. Scores plots can be used to visualize the grouping of protein kinases based on their protein−probe interaction pattern. The PCA of such an X matrix produces a score plot where each protein structure is represented by a single point. Protein kinases with similarities in their ATP binding sites—especially similar interaction pattern to adequate GRID probes—should be clustered. The analysis then is able to focus on differences between the common features of target subfamilies. The first new principal component (PC) describes the maximum variance among all possible directions, the second one the next largest variation among all directions orthogonal to the first one, etc. The resulting eigenvalues represent the overall variance after extraction of each successive new factor. If most of the variation of the original data can be described by the first few factors, a much simpler data structure exists. Here, the NIPALS[17a] algorithm was used, which calculates every component in a stepwise manner and is faster than diagonalization of the covariance matrix if only the first eigenvalues are desired. The PC axes were not additionally rotated.

CPCA[18,19] evaluates the relative importance of individual probes for the PCA model. Because the X data matrix is structured in meaningful blocks given by the GRID probe interaction fields, hierarchical PCA methods such as hierarchical PCA[43] and CPCA provide interesting information regarding the relative importance of the different blocks (i.e., probes) in

Table 3. In Vitro Antikinase Activity (IC_{50} [μM]) of Selected ATP Antagonists as CDK Inhibitors[48]

compd	CDK1 cyclin B	CDK2 cyclin A	CDK2 cyclin E	CDK4 cyclin D1	MAP kinase ERK1	MAP kinase ERK2	cAMP-dependent kinase	cGMP-dependent kinase
isopentenyl adenine	45	50		200	90	50	50	50
olomoucine	7	7	7	>1000	30	7	>2000	>2000
roscovitine	0.65	0.7	0.7	>100	34	14	>1000	1000
purvalanol A	0.004	0.07	0.035	0.85	9		9	>10
purvalanol B	0.006	0.006	0.009	>10	3.3		3.8	>100

the analysis. CPCA uses the same objective function as PCA to explain X matrix variance. The analysis is made on a block level (i.e., individual probe-derived fields) and on a consensus level combining all blocks; the latter analysis is similar to a regular PCA. For each block, individual scores and loadings are obtained together with their relative importance for the consensus block. Block scores represent a particular, constrained submodel for a particular probe (i.e., interaction type) and add information, which is not available from PCA. Those individual scores are plotted for analysis of the kinase selectivity model for the remaing individual N1, O, and DRY probes. Block scores represent a particular point of view of the model given by a certain probe and provide unique information not present in the regular PCA model. Object distances in the block scores are used in GOLPE to assess the relative importance of the different probes in their discrimination.

The analysis of PCA and CPCA loadings plot containing information about the interaction fields between a GRID probe and the target protein structure is done using active plots as novel GOLPE plot options.[18] Those active plots help to focus on those PCA or CPCA loadings, which are able to discriminate between protein kinase subfamilies. They represent isocontour plots of their GRID interaction field variables, which contribute most to explain differences between two manually specified objects in the scores plot for a multicomponent PCA or CPCA model. Hence, these active plots greatly simplify the analysis step to result in a meaningful chemical interpretation of kinase selectivity differences. On the basis of a CPCA model, those plots can be obtained for each individual probe interaction field and thus answer the question, which particular probe is responsible for an observed selectivity difference.

2.5. 3D-QSAR Studies. For manual docking, the purvalanol B/CDK2 crystal structure 1CKP[27] with a resolution of 2.0 Å was used. After analysis of key protein−ligand interactions using GRID,[16] candidate molecules were manually docked into the active site. Subsequently, the resulting protein−ligand complexes were minimized treating all ligand atoms plus protein residues within a sphere of 5 Å around the ligand as flexible, while the remaining protein was only used to compute nonbonded interactions. Other compounds taken from ref 25 were manually built, superimposed onto this template, and minimized under identical conditions. The alignment for all molecules listed in Table 4 served as basis for all 3D-QSAR studies.

For CoMFA, steric and electrostatic interaction energies between a probe atom and those molecules are calculated at predefined grid points using a volume-dependent lattice with 2 Å spacing and Sybyl default settings. For CoMSIA,[31] the same alignment was used to compute steric, electrostatic, and hydrophobic similarity index fields[44] using a probe with charge +1, a radius of +1, a hydrophobicity of +1, and 0.3 as attenuation factor α for the Gaussian type distance dependence. Cross-validated analyses after CoMFA scaling[45] were run using leave-one-out in SAMPLS[46] or two and five cross validation groups with random selection of group members. PLS analyses using two or five randomly selected cross validation groups were averaged over 100 runs. For CoMFA, columns with a variance smaller than 2.0 were excluded prior to the PLS analysis (minimum σ). The overall quality of all PLS analyses was expressed using the cross-validated r^2 value r^2(cv). For validation, all biological activities were randomized[47] 100 times, and the mean cross-validated r^2 was calculated.

3. Results and Discussion

3.1. Structural Alignment and GRID/CPCA Model. The alignment strategy produces an acceptable superposition of protein kinases based on the 1ATP fold[39] as a reference structure. Characteristic features of the protein kinase catalytic domain fold are illustrated in Figure 3 (left), where SCRs are indicated. The final alignment of 26 target proteins is shown in Figure 3 (right), where the ATP binding site is highlighted.

Only GRID points within a radius of 4.0 Å around the consensus binding region as defined by the ligand molecules in the aligned X-ray structures were selected for analysis using PCA and CPCA.[18] Using three relevant GRID probes in the final analysis (N1, O, DRY), a significant CPCA model was obtained, which highlights active site regions as potential target sites for obtaining selective interactions to particular ligands. The combined super-scores from CPCA (related to a regular PCA score plot) are plotted in Figure 4 with the first relevant component on the *x*-axis and the second on the *y*-axis. The analysis now is focused to individual contributions of particular GRID probes for analysis of kinase active site selectivity regions. Individual score plots per GRID probe from CPCA only show minor differences in the orientations of kinase subfamilies without a fundamentally different interpretation. We will name these PCA and CPCA score plots target family landscapes, as they allow to classify an important target family according to a common and unique interaction pattern between protein active sites and probe atoms as "mimic" for small molecular ligands. This classification is solely based on a 3D interaction pattern, while protein 2D similarity considerations are only taken into account during the chosen alignment strategy and not for classification.

The first PC (*x*-axis in Figure 4) in the PCA and each individual CPCA block model separate CDK and MAP/receptor kinases on the left with negative PC1 score values from the family of PKA kinases, while the interpretation of the corresponding loadings plots in the next section allows us to focus on structural features in the active sites responsible for this separation. The CDK family is represented in two distinct clusters in the target family landscape (Figure 4), formed by two different ATP binding site conformations corresponding to the activated and inactivated kinase conformations depending on the binding of cyclin in the complex. The second PC allows us to separate between MAP and other receptor kinases with positive scores in PC2 and the CDK family showing negative PC1 and PC2 scores.

3.2. Interpretation of the GRID/PCA Model in Structural Terms. Those kinase subfamily selectivity differences are plotted in Figure 5 for PC1 and PC2 for the DRY GRID probe derived from the PCA model.

Table 4. 2,6,9-Substituted Purines as CDK Inhibitors for 3D-QSAR Analysis[a]

compd	CDK1 IC_{50} (mM)	PIC_{50}	CoMFA PIC_{50}	CoMSIA PIC_{50}	compd	CDK1 IC_{50} (mM)	PIC_{50}	CoMFA PIC_{50}	CoMSIA PIC_{50}
purvalanol B	20	4.70	4.70	4.73	118	700	3.15	3.27	3.19
96	30	4.52	4.32	4.34	roscovitine	700	3.15	3.24	2.74
amino purvalanol	33	4.48	4.28	4.39	359	790	3.10	3.54	3.67
60	35	4.46	4.09	4.15	59	800	3.10	3.08	2.78
purvalanol A	35	4.46	4.01	4.11	309	900	3.05	2.91	2.99
73	40	4.40	4.29	4.26	37	1000	3.00	3.15	3.18
98	52	4.28	4.39	4.31	13	1200	2.92	3.41	3.31
356	80	4.10	4.36	4.28	209	1200	2.92	2.49	2.59
36	100	4.00	3.37	3.82	226	1200	2.92	3.09	3.01
94	100	4.00	4.15	4.16	15	1300	2.89	2.75	2.92
33	130	3.89	3.81	3.85	219	1300	2.89	2.72	2.83
220	130	3.89	3.19	3.37	223	1300	2.89	2.71	2.71
212	160	3.80	3.21	3.30	306	1400	2.85	3.09	3.04
112	170	3.77	3.68	3.52	224	1500	2.82	2.87	2.93
10	210	3.68	3.37	3.49	41	1600	2.80	2.83	3.08
304	220	3.66	3.45	3.55	61	2300	2.64	2.85	3.07
52	220	3.66	3.36	3.23	54	2700	2.57	2.60	2.47
52	220	3.66	3.32	3.21	49	2800	2.55	2.38	2.53
303	230	3.64	3.67	3.68	50	2800	2.55	2.87	2.74
75	230	3.64	3.57	3.67	222	3000	2.52	2.78	2.75
216	230	3.64	3.46	3.59	106	4000	2.40	2.18	2.76
78	240	3.62	3.96	3.96	38	4000	2.40	2.59	2.29
211	250	3.60	3.81	3.60	43	4300	2.37	2.42	2.38
45	270	3.57	3.34	3.39	318	4400	2.36	2.59	2.37
64	290	3.54	3.59	3.42	302	5000	2.30	2.35	2.43
307	300	3.52	3.78	3.68	40	5000	2.30	2.50	2.22
68	300	3.52	3.61	3.43	67	5000	2.30	2.48	2.57
26	330	3.48	3.54	3.42	69	5000	2.30	2.29	2.49
77	360	3.44	3.95	3.76	301	5000	2.30	2.27	2.12
66	400	3.40	3.27	3.47	311	5000	2.30	2.13	2.49
44	400	3.40	3.20	3.25	312	5000	2.30	3.00	2.90
65	400	3.40	3.14	3.05	313	5300	2.28	1.92	2.01
51	420	3.38	3.20	3.10	113	6000	2.22	2.38	2.13
47	430	3.37	3.54	3.61	27	6500	2.19	2.42	2.39
314	450	3.35	3.29	3.40	11	7000	2.15	2.06	2.12
58	500	3.30	3.12	3.22	310	7000	2.15	2.32	2.00
62	500	3.30	2.83	2.83	olomoucine	7000	2.15	2.53	2.94
70	500	3.30	3.44	3.09	225	9000	2.05	1.82	1.98
71	500	3.30	3.08	3.39	46	9000	2.05	2.38	2.52
308	530	3.28	3.23	3.62	316	10000	2.00	2.20	2.20
28	560	3.25	3.76	3.36	317	10000	2.00	2.19	1.86
76	600	3.22	3.27	3.26	53	10000	2.00	2.22	2.32
110	620	3.21	3.32	3.00	12	10800	1.97	1.99	1.85

[a] Compound numbering and experimental affinities against CDK1 were taken from ref 25. Affinity predictions are based on final CoMFA or CoMSIA models.

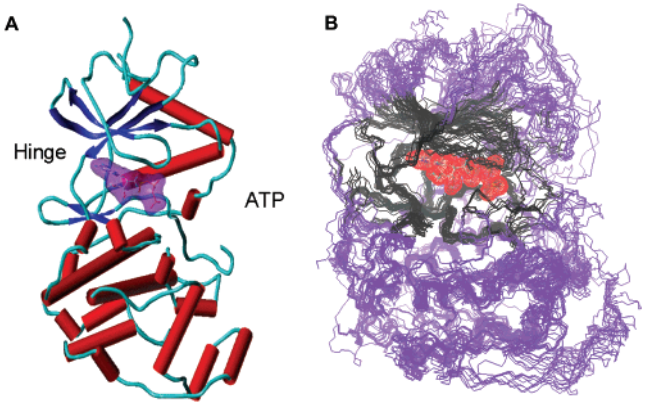


Figure 3. (a) Typical kinase fold illustrated using the protein 1ATP as reference for alignment. (b) Twenty-six protein kinases aligned using structural conserved regions (SCRs) identified using the outlined alignment strategy. For comparison, a bound molecule ATP located in the highly conserved ATP binding site is shown. Only GRID-derived probe interaction energies within 4.5 Å around this ligand are used for chemometrical analysis.

Negative contour regions are shown in cyan, and positive regions are displayed in yellow. The crystallographically determined conformation of a bound ATP molecule from the Protein Data Bank (PDB) file 1ATP (cAMP, mouse subunit[39]) is shown for comparison. Positive contours in PC1 (yellow) indicate regions with preferences for PKA for a particular chemical interaction, while negative contours (cyan) highlight favorable and selective interactions to the family of CDK/MAP kinases. In PC2, negative interactions (cyan) characterize favorable and selective regions for CDK kinases, while positive regions (yellow) are favorable for MAP kinases. PC1 is dominated by selectivity subsites in the kinase purine and hinge binding region, while only a few contour regions are located in the phosphate binding area or directed toward the substrate binding region. In contrast to the first PC, the discrimination in PC2 is also driven by structural differences in the phosphate binding area.

To illustrate the use of the target family landscape for understanding kinase selectivity profiles, a series of reported 2,6,9-substituted purines[25] as selective CDK inhibitors will be discussed in combination with GRID-derived MIFs and CPCA differential plots. Only negative contour regions will be shown, corresponding to those binding site areas that are responsible for CDK family selectivity. The compound structures are given in Scheme 1, while their kinase activities for a broader panel are summarized in Table 3.[48,49] In search for more

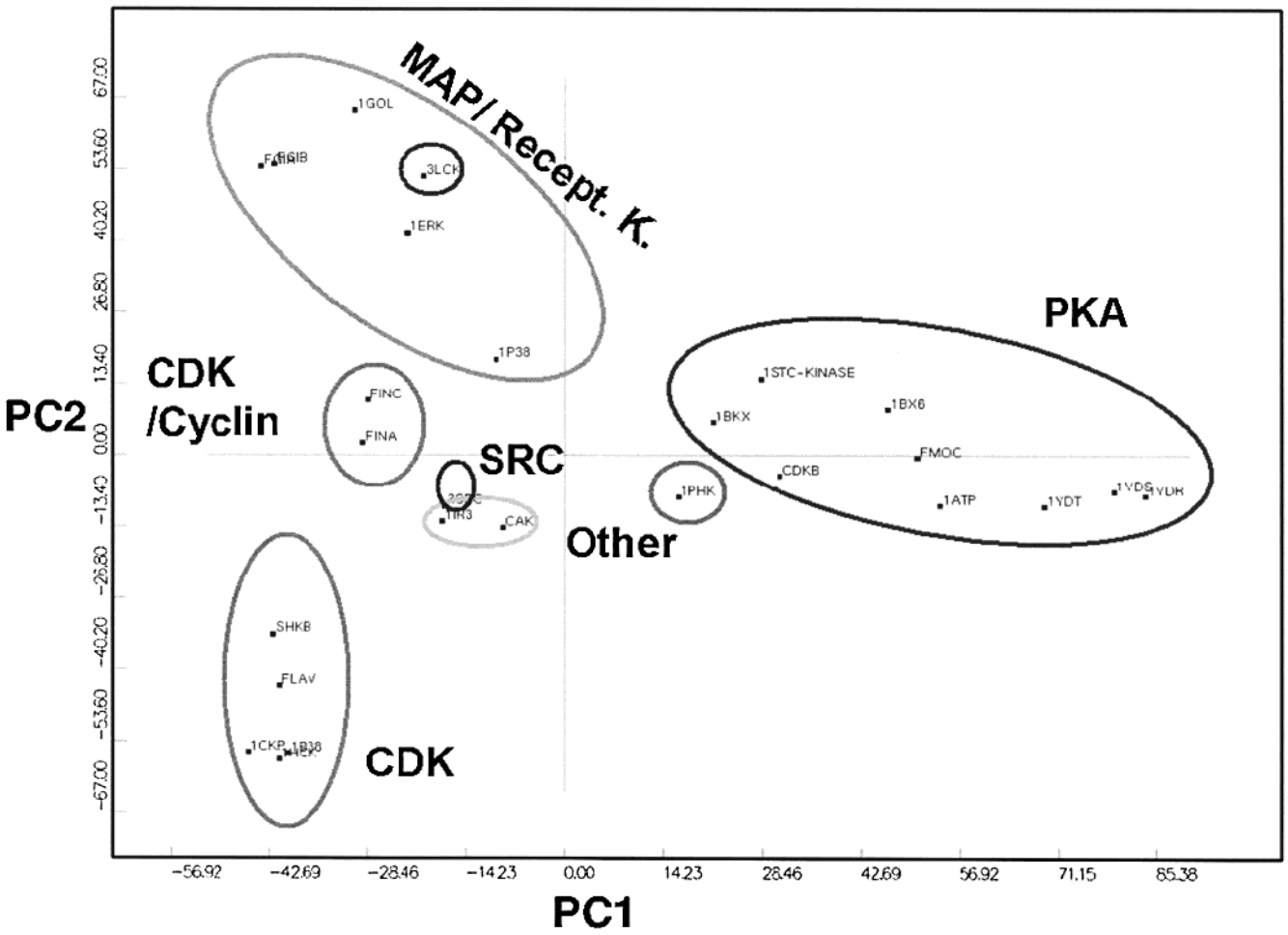


Figure 4. PCA score plot for analysis using all three different GRID probes (N1, O, and DRY) for this final analysis. The plot (target family landscape) illustrates the differences of several kinase families in the chemometrical space.

specific CDK inhibitors,[48] a growing number of scaffolds have been investigated with 6-aminopurine derivates as one interesting chemotype.[25−27,50] The inhibitors discussed herein were shown to be competitive CDK inhibitors with respect to ATP. X-ray crystal structures of representative compounds such as olomoucine, roscovitine, and purvalanol B (Table 3, Scheme 1) in complex with monomeric CDK2 reveal binding within the kinase ATP binding site, while the binding mode and H-bonding pattern significantly differs to the ATP/CDK complex.[27,51]

In Figures 6 and 7, the X-ray-derived conformation of roscovitine (IC_{50} 0.65 μM/CDK1; 0.70 μM/CDK2) from its complex with CDK2[51b] is shown in combination with contour regions from the CPCA loadings plot for three different probes to illustrate selectivity regions in both PCs. Cyan contours indicate those regions where selective interactions using the indicated probe or a chemically closely related functional group is predicted to increase selective interactions toward the CDK family (Figure 4).

As can be seen from the PC1 contour plots, which highlight regions discriminating the CDK/MAP families from PKA, hydrophobic interactions close to position N9 of the purine scaffold and C2′ at the purine N1′ substituent are favorable for selectivity toward the CDK family (atom numbering, see Scheme 1). Especially, the CPCA PC1 loadings derived using the DRY probe interactions (Figure 6a, upper left panel) correspond to

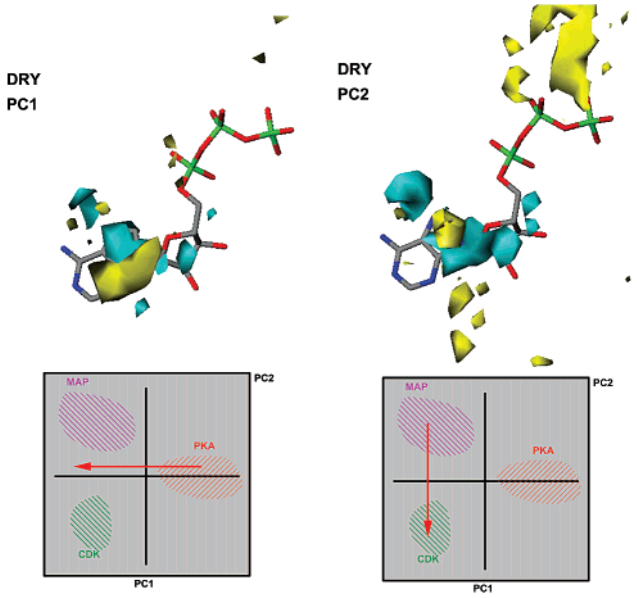


Figure 5. PCA differential plots for the first (a) and second (b) PCA component for all three different probes to highlight field differences for discrimination between various kinase families. The separation between kinase families could be rationalized by different molecular field interaction areas within the ATP binding pocket of 1ATP. In PC1, main differences are almost in the ATP-purin pocket, while in PC2 significant differences occur also in the phosphate binding area and in additional pockets not occupied by ATP.

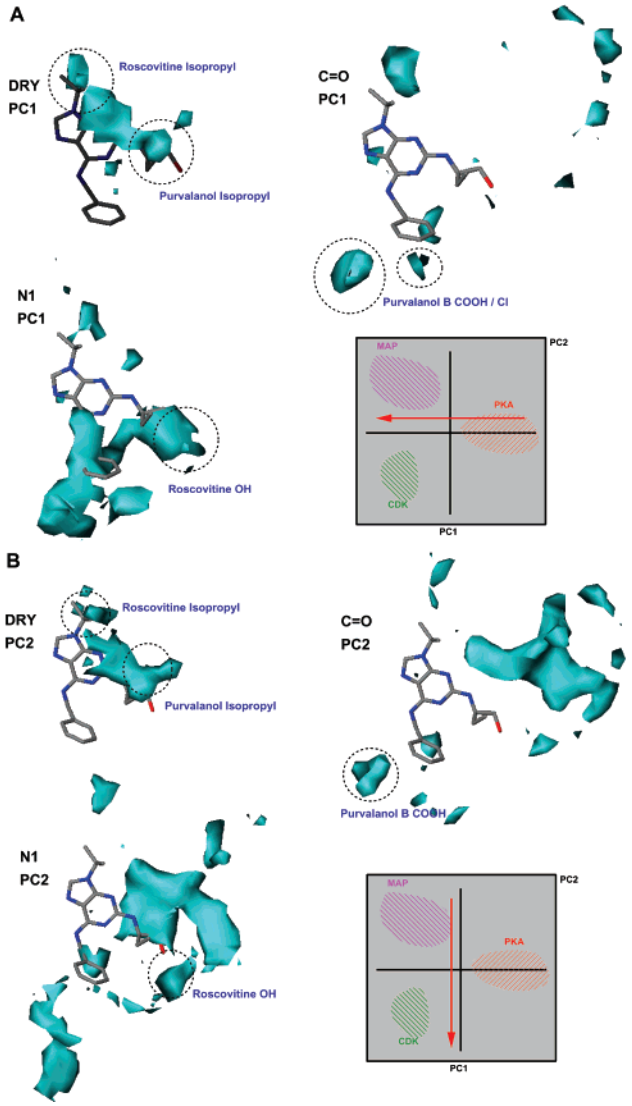


Figure 6. CPCA differential plots for the first (a) and second (b) CPCA component for all three individual probes to highlight differences between CDK/MAP und PKA kinases in their protein−probe interaction pattern. The X-ray structure of roscovitine taken from its complex with CDK2 is shown for comparison, while hydrogens are omitted for clarity. Please note different binding modes for roscovitine and ATP in Figure 5. Characteristic interactions from the GRID DRY probe are shown in the upper left part for both differential plots, to the carbonyl probe in the upper right panel and to the N1 probe in the lower left panel.

the experimental finding that olomoucine is a more selective CDK inhibitor than the unsubstituted isopentenyladenine,[51a] as shown in Table 3. In agreement with these DRY probe-derived contour maps, the addition of a methyl group at N9 and a hydroxyethyl moiety at N1′ are correlated with this enhanced selectivity profile. CDK selectivity is further improved in the more potent inhibitors roscovitine, purvalanol A, purvalanol B, and amino-purvalanol. Here, N9 carries the larger, hydrophobic isopropyl group instead of methyl (Table 3), which is better accommodated in the hydrophobic back-pocket formed by the amino acid side chains of Val64, Phe80, and Ala144 in CDK2. Moreover, the lipophilic amino acid Phe80, which is the amino acid at the beginning of the kinase hinge region limits this hydrophobic back-pocket. When inspecting the 3D structures of PKA kinases, this phenylalanine is replaced against a methionine with a different lipophilic character. This residue also closes the back-pocket in the PKA family, while it shows increased possibilities for side chain flexibility. In contrast, MAP kinases have a polar residue at this key position in the ATP binding pocket (Thr in p38 kinase, Gln in ERK), which opens or closes the back-pocket depending on size and key interactions.

Furthermore, the CDK selective molecules roscovitine, purvalanol A, and purvalanol B are substituted at carbon atom C2′ of the hydroxyethyl chain with a hydrophobic ethyl or isopropyl substituent, respectively. This favorable substitution is directed toward another hydrophobic binding site area close to residues Ile10 and Val18, while the primary hydroxy functionality is located in a polar ATP binding site region. These modifications at the isopentenyladenine scaffold consistently increase affinity and selectivity of the resulting molecules toward CDK1 and 2, while the basal activity against other kinase families is not significantly affected, which is in perfect agreement with the target family landscape-derived kinase selectivity map.

The favorable effect of the hydroxyethyl addition at position N1′ is in agreement with the N1 probe-derived CPCA loadings plot (Figure 6a, lower left panel), where the contour maps clearly indicate favorable polar interactions in this area close to the primary hydroxy group. This binding site area is solvent-exposed and located in the vicinity of polar amino acid side chains. The polar probes N1 and O (Figure 6, upper right panel) further indicate that particular substituents introduced into the meta and/or para position of the roscovitine N-benzyl ring should also improve selectivity. The N-phenyl ring of purvalanol B, for example, is accommodated by the side chains of Ile10 and the more hydrophobic Cβ-Cγ chain of Gln85. Both proposed substitutions have been realized in purvalanol A on the N-phenyl ring (*meta*-chlorine) and purvalanol B (*meta*-chlorine and *para*-carboxylic acid), resulting in highly CDK selective kinase inhibitors. The aromatic chlorine atom is directed toward the Asp86 carboxylate in the corresponding X-ray structure 1CKP, while the purvalanol B carboxylate is within an interacting distance to Lys89-Nϵ. Purvalanol B, for example, has a CDK1/CDK2A affinity of 6 nM (9 nM for CDK2B), while it shows a lower in vitro activity against representative members of other kinase families.

The second PC explains selectivity differences between the CDK family vs MAP kinases. Again, substitution at the positions N9 and C2′ at the purine scaffold is favorable to increase selectivity toward CDK, as can be seen from inspection of Figure 6b with the corresponding CPCA loading plots for all three GRID-derived MIFs. The GRID carbonyl oxygen (O) probe indicates an interaction at the roscovitine N-benzyl ring, which could be used in further derivatives to increase selectivity. The N1 probe indicates a favorable interaction toward the CDK family in the vicinity of the roscovitine primary hydroxy functionality, which agrees to the improved selectivity, when comparing isopentenyladenine to all other molecules in Table 3. In addition, the inspection of CPCA loading plots indicates that the kinases phosphate binding region offers additional opportunities to achieve selective interactions, while this

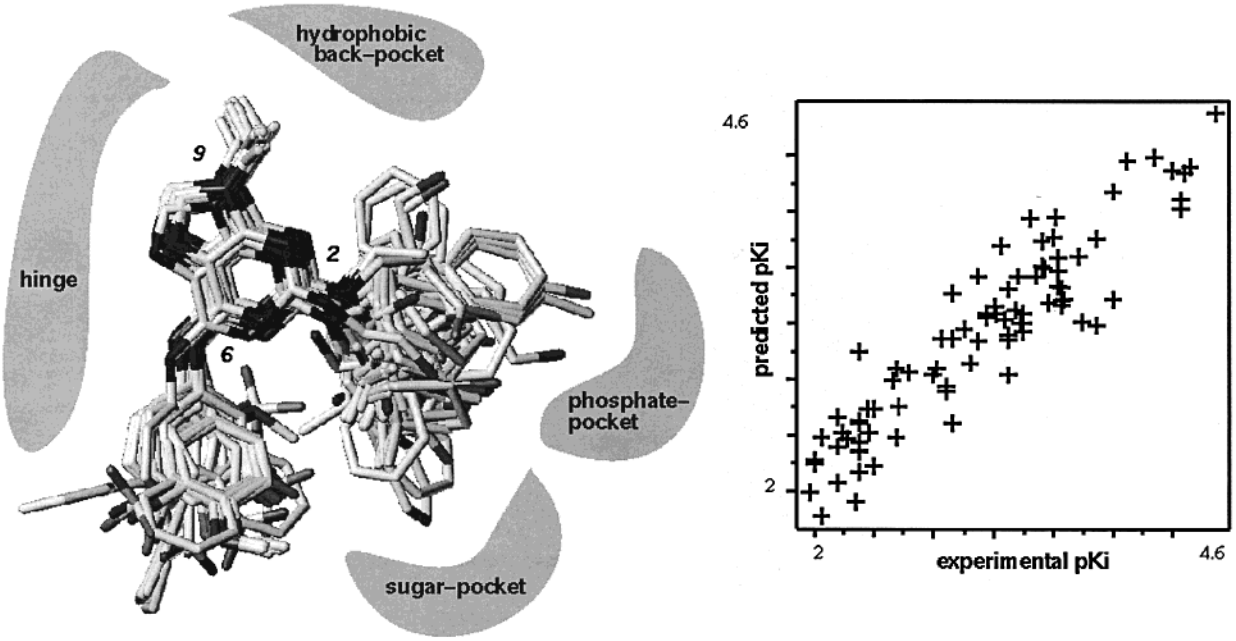


Figure 7. (a) Superposition of 86 2,6,9-substituted purines as CDK1 inhibitors taken from ref 25 for 3D-QSAR studies (Table 4), built on the basis of the docked conformation of purvalanol B in CDK2 and fitted into the protein cavity. Hydrogens are omitted for clarity; characteristic regions in the CDK2 active site are schematically indicated. (b) Graph of observed vs fitted biological activity for the CoMFA model.

region has not been explored by any of the shown compounds.

Those CPCA loading plots derived from particular GRID MIFs are highly influenced by the particular conformation of a protein binding site, as geometrical changes cause different protein−ligand interaction patterns. In the CDK family, the observed selectivity differences within the phosphate binding region are mainly caused by so-called inactive CDK conformation without bound cyclin. As kinases are similar to molecular switches in a cellular signaling cascade, they adopt either an active or an inactive conformation, which significantly differ in their local conformation in the ATP binding site. A stabilization of the inactive conformations would be a valuable approach for drug design, also in agreement with the target family landscape.

3.3. CoMFA and CoMSIA Statistical Models. To get additional information on key protein−ligand interactions favorable for affinity within the active site for CDKs and compare those with the obtained target family landscape for selectivity, a 3D-QSAR analysis of a series of 86 2,6,9-substituted purines[25] was carried out. This data set includes molecules from the previous section to highlight selectivity regions (except isopentenyladenine), while broader affinity data for all compounds were not available. Although this data set was used for external validation of a previously reported CDK1 3D-QSAR model,[26] it has never been used to our knowledge as a training data set for QSAR. Our intention was to use 3D-QSAR results only to compare key determinants for CDK1 affinity with those required for selectivity unveiled in the previous section. The resulting 3D-QSAR models to explain CDK1 affinity based on the alignment rule derived using the purvalanol B/CDK2 active site structure showed a high degree of internal consistency. This section summarizes statistical results for both CoMFA and CoMSIA models and the model validation studies. A CoMFA model with an r^2(cv) value of 0.63 for five relevant PLS components and a conventional r^2 of 0.86 was obtained (Table 3). The alignment and the graph of observed vs fitted biological activities[52] are displayed in Figure 7. The steric field descriptors (1716 grid-based variables) explain 49% of the variance, while the electrostatic field accounts for 51%.

The effect of the alignment relative to the grid position was evaluated by consistently moving all compounds in increments of 0.5 Å in all three dimensions *x*, *y*, and *z*. Here, the relative alignment is not changed, but the absolute orientation with respect to the grid is changed. The r^2(cv) values for each orientation range from 0.63 to 0.56 (mean r^2(cv), 0.59; SD, 0.03) for five component PLS models, suggesting only a minor dependence of the final model on the absolute orientation of the grid box. The effect of different choices of the origin of the grid was investigated using 14 atom types in addition to carbon C3 as probe atoms with a 2 Å grid spacing. The r^2(cv) values for each probe atom range from 0.63 to 0.57 for five components PLS models with C3 and other carbon-based atom types showing the highest r^2(cv), showing only a slight dependence on the chosen probe (mean r^2(cv), 0.59; SD, 0.03).

The mean r^2(cv) for 100 randomizations of the biological activity is −0.09 (SD, 0.10; high, 0.04; low, −0.38), showing that the final CoMFA model is significantly better than a random model. PLS analyses were run 100 times each with two and five randomly chosen cross validation groups, respectively. The mean r^2(cv) value of 0.60 for five PLS components and five cross validation groups (SD, 0.04; high, 0.68; low, 0.51) is slightly lower than using the leave-one-out method. Using only two cross validation groups, a lower mean r^2(cv) value of 0.53 for five PLS components was observed (SD, 0.07; high,

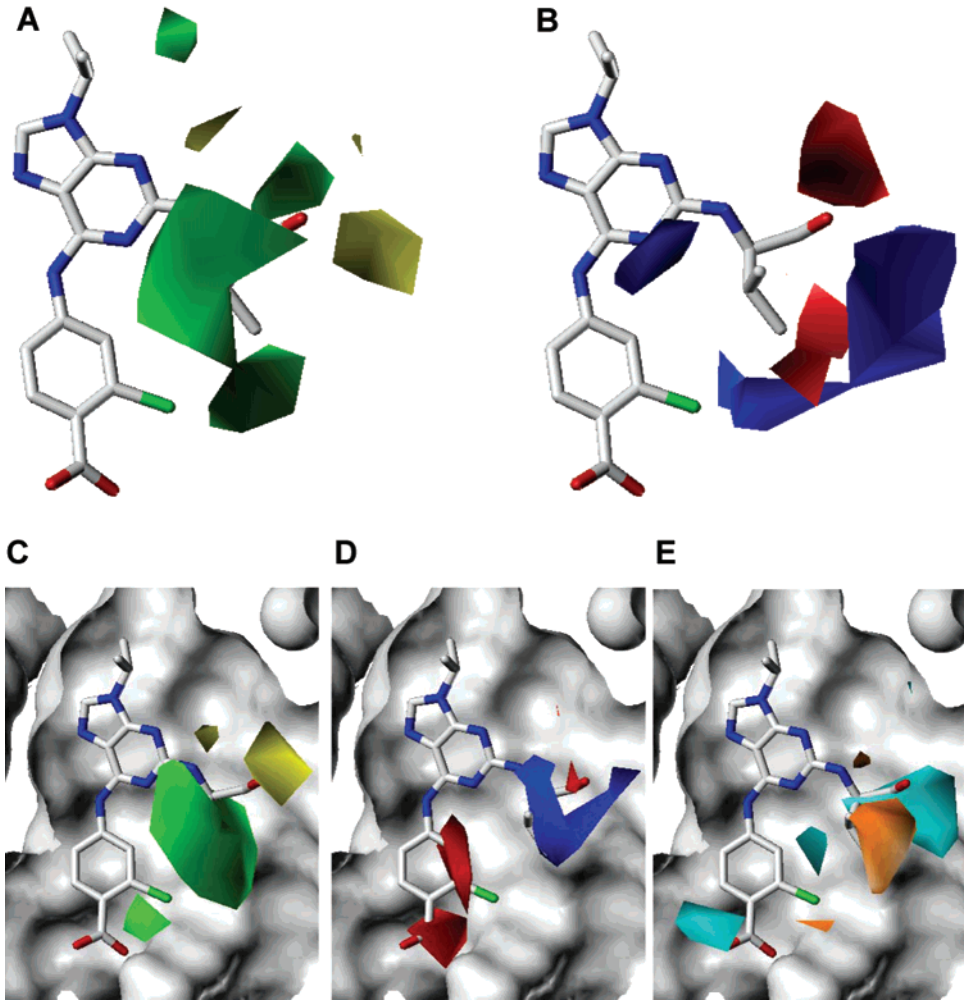


Figure 8. Contour maps from the final CoMFA and CoMSIA analyses in combination with the inhibitor purvalanol B; hydrogens are omitted for clarity. (a) CoMFA steric *std*coeff* contour map. Green contours (>80% contribution) refer to sterically favored regions; yellow contours (<20% contribution) indicate disfavored areas. (b) CoMFA electrostatic *std*coeff* contour map. Blue contours (>80% contribution) refer to regions where negatively charged substituents are disfavorable; red contours (<20% contribution) indicate regions where negatively charged substituents are favorable. (c) CoMSIA steric *std*coeff* contour map with CDK2 solvent accessible surface. Green contours (>80% contribution) refer to sterically favored regions; yellow contours (<20% contribution) indicate disfavored areas. (d) CoMSIA electrostatic *std*coeff* contour map with CDK2 solvent accessible surface. Blue contours (>80% contribution) refer to regions where negatively charged substituents are disfavorable; red contours (<20% contribution) indicate regions where negatively charged substituents are favorable. (e) CoMSIA hydrophobic *std*coeff* contour map with CDK2 solvent accessible surface. Cyan contours (>80% contribution) refer to regions where hydrophilic substituents are favorable; orange contours (<20% contribution) indicate regions where hydrophobic substituents are favorable.

0.69; low, 0.27). Both investigations support the finding of a stable, significant, and predictive model.

Similar results were obtained for CoMSIA, a model with a slightly lower r^2(cv) value of 0.56 for six PLS components, and a conventional r^2 of 0.87 was obtained (Table 3). The steric field descriptors (1716 variables) explain only 20% of the variance, while the proportion of the electrostatic descriptors remains in a similar range with 40%. The additional hydrophobic field explains the remaining 40% of the variance. When randomizing biological activities, a mean r^2(cv) of −0.11 (SD, 0.14; high, 0.18; low, −0.59) is observed, revealing the significance of this model. When averaging 100 PLS analyses with two random cross validation groups, a mean r^2(cv) of 0.42 (SD, 0.09; high, 0.63; low, 0.12) results for six component models. This value is even increased, when using five cross validation groups. Here, a mean r^2(cv) of 0.52 (SD, 0.05; high, 0.60; low, 0.34) results for six PLS components and five cross validation groups, which again is only slightly lower than the r^2(cv) value obtained using the leave-one-out procedure.

3.4. Interpretation of 3D-QSAR Models for Affinity and Selectivity. The steric and electrostatic *std*coeff* fields for the final CoMFA analysis A of 86 molecules based on the alignment from the CDK ATP binding site topology are displayed in Figure 8 with the potent inhibitor purvalanol B (Table 3). In Figure 8a, steric field contributions correlated with biological activity changes are displayed. Green contours (>80% contribution) indicate regions where steric bulk is favorable for new inhibitors, while yellow contours (<20% contribution) highlight regions where bulky substituents are detrimental for biological activity. A similar analysis is given for the electrostatic *SD*coeff* field (Figure 8b). Blue contours (>80% contribution) refer to regions where an increase of positive charge (or a decrease of negative charge) is favored for new ligands to enhance

affinity, while red contours (<20% contribution) indicate those regions where an increase of negative charge is favorable for biological activity. Those 3D-QSAR results were derived taking only ligand information into account, which was generated by docking and minimization of protein−ligand complexes. However, the PLS contour maps are interpreted in combination with the ATP binding site topology to underscore consistency with steric, electrostatic, and hydrophobic requirements of this kinase ATP binding site and to identify those regions that favorably both influence affinity and selectivity.

The interpretation of the steric CoMFA contour maps shows steric bulky substituents to be favorable at positions N9 directed toward the CDK hydrophobic back-pocket and in some positions close to the hydroxyethyl substitution at N1′. Those contour maps indicate that adding an additional ethyl or isopropyl group to the hydroxyethyl substituent increases affinity toward CDK1. As both regions have consistently been identified to be important for achieving selective interactions toward the CDK family, they are of high relevance for further CDK-directed inhibitor design. Furthermore, steric bulk is favorable close to the *m*-chlorine substituent of purvalanol B, demonstrating its positive influence on both affinity and selectivity. The interpretation of the CoMSIA model is consistent to CoMFA. The corresponding contour maps are displayed in Figures 8c−e in combination with the solvent accessible surface derived from the purvalanol B binding site of its complex with CDK2. Again, green contours from the PLS-derived *std*coeff* field (>80% contribution) indicate regions where steric bulk is favorable, while yellow contours (<20% contribution) highlight regions where steric bulk is detrimental. The sterically favorable region at N9 is only visible in CoMSIA at lower contour regions.

The inspection of the CoMFA and CoMSIA electrostatic contour maps in Figure 8b,d indicates that more positive charge at the phosphate binding pocket at the entrance of the ATP binding site is favorable, which is typically occupied by the hydroxyethyl substituent in this series. This preference is reflected in the target family landscape CPCA maps for polar N1 and C═O probes indicating that polar substitutions in this area increase selectivity toward CDK. More negative charge is favorable close to the primary hydroxyl group of the substituent attached to N1′ and close to the purvalanol B carboxylate, which agrees to the selectivity model.

To estimate entropic contributions to ligand binding, hydrophobic CoMSIA fields were incorporated into the PLS analysis. Cyan contour regions in Figure 8e indicate regions where hydrophilic interactions are favorable for biological activity (>80% contribution). Those regions are located close to the purvalanol B carboxylate and primary hydroxy group, showing the importance of polar substituents in those areas for active and selective CDK inhibitors. The main favorable hydrophobic interaction, as indicated by orange contour regions (<20% contribution), is close to the isopropyl substituent attached to C2′ in purvalanol B. As mentioned, hydrophobic interactions characterized using the GRID DRY probe were important to achieve selective interactions toward CDK2 and thus provide a favorable area for chemical optimization toward this important family.

Hence, the analysis of MIFs plus an informative data set with 86 purines as CDK1 inhibitors show that it is possible to identify regions in the CDK ATP binding site, which are on one hand related to favorable protein−ligand interactions toward potent inhibitors, while some of these regions were identified in the target family landscape contour maps to increase selective interactions toward this kinase subfamily.

4. Conclusions

Using an efficient method to compute potentially favorable protein−ligand interactions combined with modern chemometrical tools, it was possible to identify subfamilies in the kinase superfamily derived from differences in favorable protein−ligand interactions in the ATP binding site. A plot indicating similarities and differences in this target family derived from PCA and/or CPCA score plots was obtained, allowing us to identify selectivity areas. This model was able to rationalize selectivity differences for a small series of 2,6,9-substituted purines tested in a broader panel of kinases and gave useful insights into key protein−ligand interactions, which are unique to a particular subfamily. The classification of kinase subfamilies could also be used to guide the search and identification for related kinases for a particular target to be assayed for selectivity. After the qualitative agreement between selectivity profiles for substituted purines as CDK inhibitors, a 3D-QSAR analysis of a larger series of 2,6,9-substituted purines led to the extraction of key protein−ligand interactions, which are favorable for affinity in this series. The resulting 3D-QSAR models were consistent for explaining CDK1 affinity and provide valuable insights for further drug design programs. Finally, both analyses were combined toward the identification of binding site areas and favorable protein−ligand interactions, which allow us to design potent and selective kinase inhibitors.

This target family landscape-derived classification is also in good agreement with internal selectivity data on ligand interactions with various kinases and provides valuable information for designing novel, selective inhibitors. The employed data set of 26 X-ray structures offers a solid basis to investigate selectivities within this important protein family, while those structures do not always represent different proteins but also mutants and proteins from different species. Unfortunately, many important kinases have not yet been crystallized, while the rapidly growing protein structure database will enable us to apply this chemometrical analysis technique to larger and thus more informative protein data sets, allowing to obtain a better view from protein binding sites to discriminating interactions responsible for selectivity. The presented work was carried out to illustrate the potential in analyzing a protein family by means of target family landscapes. It now becomes possible to highlight main discriminating features between distinct protein families, while it also offers the opportunity to differentiate between related proteins for fine-tuning of selectivities. Hence, a better view of selectivities on a molecular level driven by protein ligand interactions emerges.

Acknowledgment. We thank K.H. Baringhaus (Aventis) and G. Cruciani (University of Perugia, Italy)

for many interesting discussions. Furthermore, we thank F. S. Jørgensen for a preprint of his work on matrix metalloproteinase selectivity studies.

References

(1) Hanks, S. K.; Hunter, T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **1995**, *9*, 576−596.

(2) Hanks, S. K.; Quinn, A. M.; Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **1988**, *241*, 42−52.

(3) Hanks, S. K. Eukaryotic protein kinases. *Curr. Opin. Struct. Biol.* **1991**, *1*, 369−383.

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

(5) Johnson, L. N.; Lowe, E. D.; Noble, M. E. M.; Owen, D. J. The structural basis for substrate recognition and control byprotein kinases. *FEBS Lett.* **1998**, *430*, 1−11.

(6) *Protein Kinases*; Woodgett, J. R., Ed.; Oxford University Press: New York, 1994.

(7) (a) *The Protein Kinase Factsbook: Protein-Serine Kinases*; Hardie, G., Hanks, S., Eds.; Academic Press Inc.: San Diego, CA, 1995. (b) *The Protein Kinase Factsbook: Protein-Tyrosine Kinases*; Hardie, G., Hanks, S., Eds.; Academic Press Inc.: San Diego, CA, 1995.

(8) (a) Fry, D. W.; Kraker, A. J.; Connors, R. C.; Elliot, W. L.; Nelson, J. M.; Showalter, H. D.; Leopold, W. R. Strategies for the discovery of novel tyrosine kinase inhibitors with anticancer activity. *Anticancer Drug Des.* **1994**, *9*, 331−351. (b) Groundwater, P. M.; Solomons, K. R. H.; Drewe, J. A.; Munawar, M. A. Protein kinase inhibitors. *Prog. Med. Chem.* **1996**, *33*, 233−329. (c) Lawrence, D. S.; Niu, J. Protein kinase inhibitors: The tyrosine-specific protein kinases. *Pharmacol. Ther.* **1998**, *77*, 81−114. (d) Lee, J. C.; Adams, J. L. Inhibitors of serine/threonine kinases. *Curr. Opin. Biotechnol.* **1995**, *6*, 657−661. (e) Patrick, D. R.; Heimbrook, D. C. Protein kinase inhibitors for the treatment of cancer. *Drug Discovery Today* **1996**, *1*, 325−330. (f) Adams, J. L.; Lee, D. Recent progress towards the identification of selective inhibitors of serine/threonine kinases. *Curr. Opin. Drug Discovery Dev.* **1999**, *2*, 96−109.

(9) Wei, L.; Hubbard, S. R.; Smith, R. F.; Ellis, L. Protein kinase superfamily−comparisons of sequence data with three-dimensional structures. *Curr. Opin. Struct. Biol.* **1994**, *4*, 450−455.

(10) Toledo, L. M.; Lydon, N. B.; Elbaum, D. The structure-based design of ATP site-directed protein kinase inhibitors. *Curr. Med. Chem.* **1999**, *6*, 775−805.

(11) Denessiouk, K. A.; Johnson, M. S. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins* **2000**, *38*, 310−326.

(12) Cohen, P. The development and therapeutic potential of protein kinase inhibitors. *Curr. Opin. Chem. Biol.* **1999**, *39*, 459−465.

(13) McMahon, G.; Sun, L.; Liang, C.; Tang, C. Protein kinase inhibitors: structural determinants for target specificity. *Curr. Opin. Drug Discovery Dev.* **1998**, *1*, 131−146.

(14) Garcia-Echeverria, C.; Traxler, P.; Evans, D. B. ATP site-directed competitive and irreversible inhibitors of protein kinases. *Med. Res. Rev.* **2000**, *20*, 28−57.

(15) (a) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857. (b) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083−1094. (c) Wade, R. C.; Clerk, K. J.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140−147. (d) Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148−156.

(16) *GRID*, version 18; Molecular Discovery Ltd., West Way House: Elms Parade, Oxford, U.K., 2000.

(17) (a) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52. (b) Dillon, W. R.; Goldstein, M. *Multivariate Analysis: Methods and Applications*; Wiley: New York, 1984. (c) Malinowski, E. R.; Howery, D. G. *Factor Analysis in Chemistry*; Wiley: New York, 1980. (d) Stahle, L.; Wold, S. Multivariate Data Analysis and Experimental Design in Biomedical Research. In *Progress in Medicinal Chemistry*; Ellis, G. P., West, G. B., Eds.; Elsevier: Amsterdam, 1988; pp 292−338. (e) Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johanson, E.; Lindberg, W.; Sjöström, M. Multivariate Data Analysis in Chemistry. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B. R., Ed.; NATO, ISI Series C 138; D. Reidel Publishing Co.: Dordrecht, Holland, 1984; pp 17−96.

(18) Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E. J.; Fox, T. GRID/CPCA: A New Computational Tool to Design Selective Ligands. *J. Med. Chem.* **2000**, *43*, 3033−3044.

(19) Westerhuis, J. A.; Kourti, T.; Macgregor, J. F. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemom.* **1998**, *12*, 301−321.

(20) Cruciani, G.; Goodford, P. J. A search for specificity in DNA-drug interactions. *J. Mol. Graphics* **1994**, *12*, 116−129.

(21) Pastor, M.; Cruciani, G. A novel strategy for improving ligand selectivity in receptor-based drug design. *J. Med. Chem.* **1995**, *38*, 4637−4647.

(22) Matter, H.; Schwab, W. Affinity and Selectivity of Metrix Metalloproteinase Inhibitors: A Chemometrical Study from the Perspective of Ligands and Proteins. *J. Med. Chem.* **1999**, *42*, 4506−4523.

(23) Matter, H.; Schwab, W. A view on affinity and selectivity of nonpeptidic matrix metalloproteinase inhibitors from the perspective of ligands and target. In *Molecular Modeling and Prediction of Bioactivity*; Gundertofte, K., Jørgensen, F. S., Eds.; Kluwer: New York, 2000; pp 123−128.

(24) Filipponi, E.; Cecchetti, V.; Tabarrini, O.; Bonelli, D.; Fravolini, A. Chemometric rationalization of the structural and physicochemical basis for selective cyclooxygenase-2 inhibition: toward more specific ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 277−291.

(25) 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.

(26) Ducrut, P.; Legraverend, M.; Grierson, D. S. 3D-QSAR CoMFA on Cyclin-Dependent Kinase Inhibitors. *J. Med. Chem.* **2000**, *43*, 4098−4108.

(27) 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.

(28) Cramer, R. D.; Patterson, D. E.; Bunce, J. E. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(29) Clark, M.; Cramer, R. D.; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative Molecular Field Analysis (CoMFA). 2. Towards its use with 3D-Structural Databases. *Tetrahedron Comput. Methodol.* **1990**, *3*, 47−59.

(30) *3D-QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM, Leiden (NL), 1993.

(31) 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.

(32) (a) Klebe, G.; Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1−10. (b) Böhm, M.; Stürzebecher, J.; Klebe, G. Three-Dimensional Quantitative Structure−Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem.* **1999**, *42*, 458−477. (c) Klebe, G. Comparative molecular similarity indices analysis. CoMSIA. *Perspect. Drug Discovery Des.* **1998**, *12/13/14*, 87−104.

(33) (a) 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 Methodol. *Quant. Struct.-Act. Relat.* **1984**, *3*, 131−137. (b) Geladi, P. Notes on the history and nature of partial least squares (PLS) modelling. *J. Chemom.* **1988**, *2*, 231−246.

(34) (a) Wold, S. Cross-Validatory Estimation of the Number of Component in Factor and Principal Component Models. *Technometrics* **1978**, *4*, 397−405. (b) Diaconis, P.; Efron, B. Computer-Intensive Methods for Statistics. *Sci. Am.* **1984**, *116*, 96−117. (c) Cramer, R. D.; 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) *SYBYL Molecular Modelling Package*, version. 6.6; Tripos: St. Louis, MO, 1999.

(36) (a) Halgren, T. Maximally diagonal force constants in dependent angle-bending coordinates. II. Implications for the design of empirical force fields. *J. Am. Chem. Soc.* **1990**, *112*, 4710−4723. (b) Halgren, T. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720−729.

(37) (a) RCSB Protein Data Bank, from the Research Collaboratory for Structural Bioinformatics; http://www.rcsb.org/pdb/index.html. (b) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(38) (a) Blundell, T. L.; Carney, D. P.; Gardner, S.; Hayes, F. R. F.; Howlin, B.; Hubbard, T. J. P.; Overington, J. P.; Singh, D. A.; Sibanda, B. L.; Sutcliffe, M. J. Knowledge-based protein modeling and design. *Eur. J. Biochem.* **1988**, *172*, 513−520. (b) Sutcliffe, M. J.; Haneef, I.; Carney, D. P.; Blundell T. L. Knowledge-based modeling of homologous proteins. Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.* **1987**, *1*, 377−384. (c) Sutcliffe, M. J.; Hayes, F. R. F.; Blundell, T. L. Knowledge based modeling of homologous proteins. Part II: Rules for the conformations of substituted side chains. *Protein Eng.* **1987**, *1*, 385−392.

(39) Zheng, J.; Trafny, E. A.; Knighton, D. R.; Nguyen, H. X.; Taylor, S. S.; Ten Eyck, L. F.; Sowadski, J. M. 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with manganese ATP and a peptide inhibitor. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1993**, *D49*, 362−365.

(40) Taylor, S. S.; Radzio-Andzelm, E.; Knighton, D. R.; Ten Eyck, L. F.; Sowadski, J. M.; Herberg, F. W.; Yonemoto, W.; Zheng, J. Crystal structures of the catalytic subunit of cAMP-dependent protein kinase reveal general features of the protein kinase family. *Receptor* **1993**, *3*, 165−172.

(41) *GOLPE 4.5, Multivariate Infometric Analysis*; Srl: Viale die Castagni, 16 Perugia, Italy, 1999.

(42) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9−20.

(43) Wold, S.; Kettaneh, N.; Tjessem, K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *J. Chemom.* **1996**, *10*, 463−482.

(44) Kearsley, S. K.; Smith, G. M. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.

(45) (a) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A.; Rognan, D. Recommendations for CoMFA studies and 3D QSAR publications. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 711−717. (b) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and limitations. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 583−616. (c) Cramer, R. D.; DePriest, S. A.; Patterson, D. E.; Hecht, P. The developing practice of comparative molecular field analysis. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 443−485.

(46) Sheridan, R. P.; Nachbar, R. B.; Bush, B. L. Extending the trend vector: The trend matrix and sample-based partial least squares. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 323−340.

(47) Clark, M.; Cramer, R. D. The probability of chance correlation using partial least squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137−145.

(48) Fischer, P. M.; Lane, D. P. Inhibitors of Cyclin-Dependent Kinases as Anti-Cancer Therapeutics. *Curr. Med. Chem.* **2000**, *7*, 1213−1245.

(49) Please note that slightly different binding affinities have been reported in the literature for purvalanol A and B CDK affinities in refs 25 and 48. For consistency, we use the data from ref 25 for 3D-QSAR, while to compare the selectivity profiles against a panel of kinases, the data reported in ref 48 were used.

(50) (a) Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E. M.; Sausville, E. A.; Schultz, R.; Yu, W. Identification of Novel Purine and Pyrimidine Cyclin-Dependent Kinase Inhibitors with Distinct Molecular Interactions and Tumor Cell Growth Inhibition Profiles. *J. Med. Chem.* **2000**, *43*, 2797−2804. (b) Vesely, J.; Havlicek, L.; Strnad, M.; Blow, J. J.; Donella-Deana, A.; Pinna, L.; Letham, D. S.; Kato, J.-Y.; Detivaud, L.; Leclerc, S.; Meijer, L. Inhibition of cyclin-dependent kinases by purine analogues. *Eur. J. Biochem.* **1994**, *224*, 771−786. (c) Meijer, L.; Borgne, A.; Mulner, O.; Chong, J. P. J.; Blow, J. J.; Inagaki, N.; Inagaki, M.; Delcros, J.-L.; Moulinoux, J.-P. Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of cyclin-dependent kinases cdc2, cdk2 and cdk5. *Eur. J. Biochem.* **1997**, *243*, 527−536. (d) Brooks, E. E.; Gray, N. S.; Joly, A.; Kerwar, S. S.; Lum, R.; Mackman, R. L.; Norma, T. C.; Rosete, J.; Rowe, M.; Schow, S. R.; Schultz, P. G.; Wang, X.; Wick, M. M.; Schiffman, D. CVT-313, a specific and potent inhibitor of CDK2 that prevents neointimal proliferation. *J. Biol. Chem.* **1997**, *272*, 29207−29211. (e) Schow, S. R.; Mackman, R. L.; Blum, C. L.; Brooks, E.; Horsma, A. G.; Joly, A.; Kerwar, S. S.; Lee, G.; Schiffman, D.; Nelson, M. G.; Wang, X.; Wick, M. M.; Zhang, X.; Lum, R. T. Synthesis and activity of 2,6,9-trisubstituted purines. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2697−2702. (f) 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. (g) 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.

(51) Schulze-Gahmen, U.; Brandsen, J.; Jones, H. D.; Morgan, D. O.; Meijer, L.; Vesely, J.; Kim, S.-H. Multiple modes of ligand recognition: Crystal structure of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins* **1995**, *22*, 378−391. (b) de Azevedo, W. F., Jr.; 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.

(52) Biological activities are expressed as $\log(1/K_i \times 1\,000\,000)$.

JM011002C