Structure-Based Approaches to Improve Selectivity: CDK2−GSK3β Binding Site Analysis

Anna Vulpetti,*,† Patrizia Crivori,‡ Alexander Cameron,§ Jay Bertrand,§ Maria Gabriella Brasca,|| Roberto D'Alessio,|| and Paolo Pevarello||

Nerviano Medical Sciences, Viale Pasteur 10, 20014 Nerviano (MI), Italy

Received January 26, 2005

An evaluation and comparison of two different approaches, GRID/CPCA and GRIND/CPCA (CPCA = consensus principal component analysis; GRIND = GRid-INdependent Descriptors), suitable for visualizing the structural differences between related proteins is presented. Ten crystal structures of CDK2/cyclin A and GSK3β solved in-house with different inhibitors were compared with the aim of highlighting regions that could be potential sites for gaining selectivity for CDK2 versus GSK3β. The analyses pointed out remarkable differences in the backs of the CDK2−GSK3β ATP binding pockets that guided the optimization toward a selective benzodipyrazole CDK2 inhibitor. The gain in selectivity can be associated with the two main differences in the ATP pocket between the enzymes. Phe80 of CDK2, the so-called gatekeeper residue often exploited for the design of kinase selective ligands, is replaced by a leucine in GSK3β, and Ala144 is replaced by a cysteine. As a consequence of these mutations, CDK2 has a less elongated and less flat buried region at the back of the ATP pocket.

INTRODUCTION

Protein kinases are one of the largest known families of enzymes characterized by having a well-conserved ATP binding pocket. As most of the synthetic kinase inhibitors are ATP-competitive, selectivity represents a potential problem.

Multiple or pairwise sequence comparison often provides very useful insights that can be used to rationally guide the optimization of selective kinase inhibitors.[1] An example of how sequence-based analysis can be helpful to propose qualitative models accounting for the different kinases' sensitivity for ligands is reported by Komander et al.[2]

We present, here, an evaluation and comparison of different approaches for visualizing the structural differences between related proteins. The comparative study was performed on a set of 10 kinase crystal structures of CDK2/cyclin A and GSK3β solved in-house with different inhibitors. Many CDK2 inhibitors are also quite potent on GSK3β, and thus, one of the major problems in identifying selective CDK2 kinase inhibitors is to achieve selectivity against GSK3β. The ATP binding pockets of these two kinases were compared with the intention of highlighting regions that could be potential sites for gaining selectivity for CDK2 versus GSK3β in the optimization of a novel benzodipyrazole chemical class. Although dual inhibition of CDK2 and GSK3β is not necessarily linked to unwanted side effects, GSK3β is involved in cellular events such as Wnt and hedgehog signaling,[3] which can make it difficult to assess the net effect of a CDK2 inhibitor on tumor cell proliferation (and toxicology). On the other hand, GSK3β is a target pursued by several companies in the central nervous system and metabolic disease therapeutic areas.[4]

The comparison of different targets is a difficult task. The visual inspection of the active sites alone is important but not sufficient to understand the contributions of various interactions in the binding process. One computational approach that offers a solution to this problem is GRID/CPCA (CPCA = consensus principal component analysis).[5] This is a powerful computational method for the detailed description of differences amenable to selective interactions within a family of proteins. This approach has already been applied to many protein families: serine proteases,[6] matrix metalloproteinases,[7] cytochromes P450 2C,[8] kinases,[9] and peroxisome proliferator-activated receptors.[10] Starting from the crystal structures of these proteins, a multivariate description of the binding sites is performed calculating the molecular interaction fields (MIFs). The MIFs identify in the protein binding site the regions where certain chemical groups can interact favorably. Then, the MIFs are analyzed with CPCA. The results are contour plots that highlight both the regions and the types of interactions that can be used to improve potency against the specific target of interest as well as the selectivity profile of the chemical class under study.

The result obtained with this well-known and widely used approach was compared with that obtained with the GRID-independent descriptors (GRIND).[11] Basically, the GRIND descriptors are alignment-independent variables also generated from the previously described molecular interaction fields (MIFs). The GRIND descriptors have been extensively applied for 3D−QSAR analysis,[12] ligand−pharmacophore identifications, and structure−metabolism relationships.[13] The only published study in which GRIND descriptors have been used to compare proteins is the work of Gutiérrez-de-Teràn et al.[14] In this study, the GRIND descriptors were used to

* Corresponding author. Tel. +39-0331-581544. Fax +39-0331-581360. E-mail: anna.vulpetti@nervianoms.com.
† Computational Sciences, Department of Chemistry.
‡ Prediction and Modeling, Attrition Reducing Technologies.
§ Structural Chemistry, Department of Chemistry.
|| Medicinal Chemistry, Department of Chemistry.

10.1021/ci0500280 CCC: $30.25 © 2005 American Chemical Society
Published on Web 08/12/2005

Table 1. The Ten Crystal Structures

enzyme	ID	ligand	resolution (Å)	reference (PDB code)
CDK2/A	CDK2_1	pyrrolopyrazole	2.7	
CDK2/A	CDK2_2	thiazole	2.4	27
CDK2/A	CDK2_3	benzodipyrazole	2.5	19
CDK2/A	CDK2_4	pyrazole	2.3	28 (1VYW)
GSK3β	GSK3_1	3-anilino-4-arylmaleimide	2.9	29 (1Q4L)
GSK3β	GSK3_2	alsterpaullone	2.3	29 (1Q3W)
GSK3β	GSK3_3	staurosporine	2.2	29 (1Q3D)
GSK3β	GSK3_4	benzodipyrazole	2.5	
GSK3β	GSK3_5	AMP−PNP	2.4	29 (1PYX)
GSK3β	GSK3_6	indirubin-3′-monoxime	2.0	29 (1Q41)

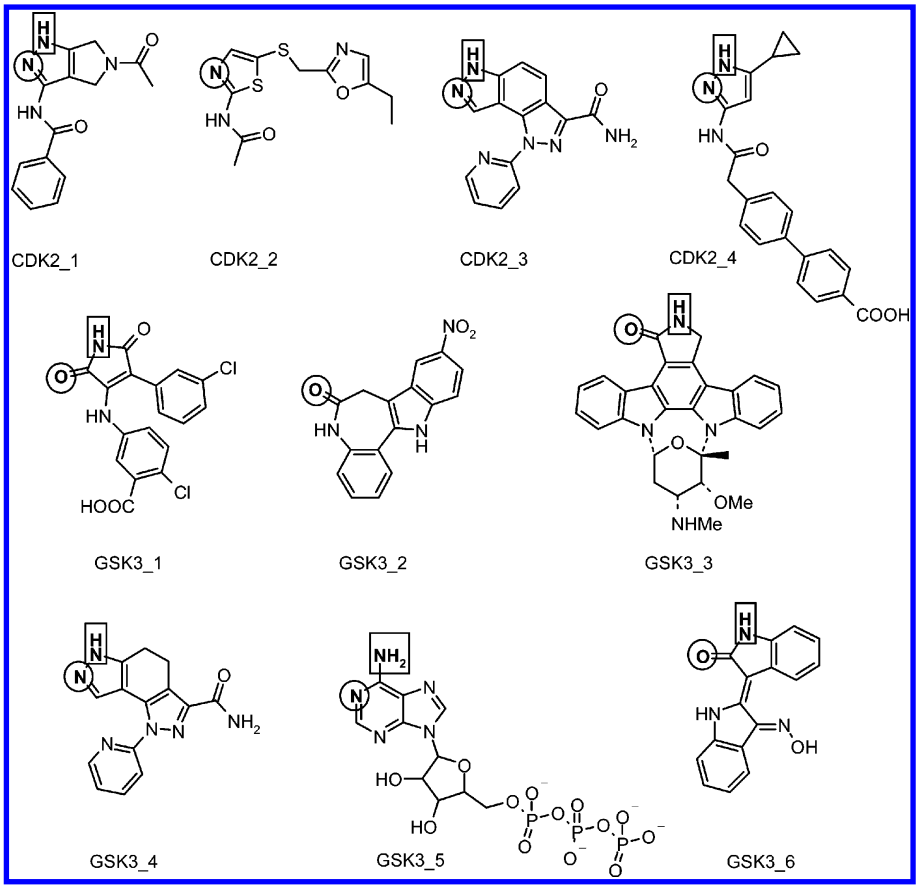


Figure 1. Structures of the ligands bound to the CDK2−GSK3β crystal structures. The atoms interacting with the backbone of CDK2−Leu83 or GSK3β−Val135 are shown circled (hydrogen acceptors); the atoms interacting with the backbone of CDK2−Glu81 or GSK3β−Asp133 are highlighted by boxes (hydrogen donors).

detect protein similarities. The goal was to assess if the postulated ribose binding site in a homology model built for the human A1 adenosine receptor was effectively similar to the existing ribose−protein complexes, for which the crystal structures are known. The structures were described by means of GRIND and a matrix of Hodgkin similarity coefficients computed. High coefficients indicate similar interaction patterns in all the analyzed ribose binding sites. In the approach described here, the GRIND descriptors analysis was performed by CPCA and visually interpreted.

The GRIND/CPCA method was investigated as an alternative approach for protein selectivity studies. This approach does not require protein alignment and allows a direct comparison of the generated variables to rapidly identify the main difference between the two protein kinases.

The application of these two different methodologies to the same problem enables us to compare the performance of the methods and enhance the reliability of the results.

METHODS AND MATERIAL

Proteins Structures. Four crystal structures of CDK2/cyclin A and six of GSK3β in a complex with different ligands were taken from in-house X-ray data and are summarized in Table 1 and Figure 1.

Crystallographic water molecules and bound ligands were removed.

GRID/CPCA. *Protein Preparation.* The crystal structures were superimposed using the carbon alpha atoms of residues that line the ATP pocket (see Table 2) excluding those located on the flexible glycine-rich loop (Gly11−Val17 of CDK2 and Gly63−Val69 of GSK3β). Figure 2 shows the 10 crystal structures aligned.

The protein hydrogens were added with the program GRIN, implemented in GRID, version 22.[15] In this program, a force-field-based mechanism is used to assign protein protonation states. That is, all the glutamic/aspartate acids

Table 2. The 37 CDK2 and GSK3β ATP Pocket Residues

CDK2	E8	I10	G11	E12	G13	T14	Y15	G16	V17	V18	K20	V29	A31
GSK3β	K60	I62	G63	N64	G65	S66	F67	G68	V69	V70	Q72	L81	A83
CDK2	K33	E51	L55	V64	L66	L78	F80	E81	F82	L83	H84	Q85	
GSK3β	K85	E97	M101	V110	L112	L130	L132	D133	Y134	V135	P136	E137	
CDK2	D86	K88	K89	K129	P130	Q131	N132	L133	L134	A144	D145	G147	
GSK3β	T138	Y140	R141	K183	P184	Q185	N186	L187	L188	C199	D200	G202	

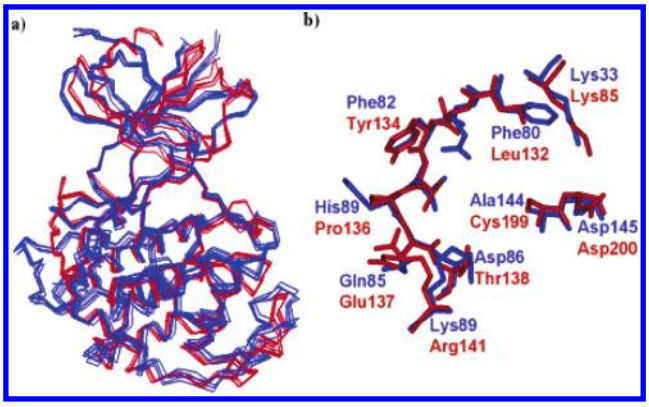


Figure 2. (a) Superimposition of CDK2 (blue) and GSK3β (red) crystal structures. Cyclin A proteins are not shown. (b) Close up of the ATP-binding pocket of the CDK2 (CDK2_4, blue) and GSK3β (GSK3_5, red) crystal structures.

and the lysine were ionized. His84 of CDK2 was protonated at Nϵ. The side chain of this residue has no direct contact with the ligands of this study.

Computing the MIFs. The MIFs were calculated in GRID. A box of 28 × 23 × 24 Å centered at the ATP active site was defined. The proteins were kept rigid (GRID directive Move=0). The number of planes (grid spacing) was set to 1 plane/Å. Default values were used for the other parameters. Table 3 lists the probes used during the generation of the MIFs in GRID.

Consensus Principal Component Analysis (CPCA). The computed MIFs were imported into GOLPE, version 4.5.12.[16] The maximum cutoff was set to zero. Block unscaled weights were used to normalize the interaction energies between the different probes. Variables with values smaller than 0.02 kcal/mol and with a standard deviation less than 0.02 kcal/mol were set to zero. The cut-out tool was used to allow focusing on the regions within 4 Å of the benzodipyrazole ligand (CDK2_3 of Figure 1). The pretreated data were used in CPCA modeling.

GRIND/CPCA. *Protein Preparation.* The GRIND are insensitive to the position and orientation of the proteins in the space; thus, the superimposition of the proteins is not required. The hydrogens were added to the proteins with the program GRIN, implemented in GRID.

Computing the MIFs. The MIFs were computed in GRID using three of the probes listed in Table 3: the DRY probe to represent hydrophobic interactions and the amide (N1) and carbonyl (O) probes representing, respectively, hydrogen bond donor and acceptor groups. A box of the same size as that used in GRID/CPCA was defined (28 × 23 × 24 Å centered at the ATP active site). In this case, a grid spacing of 0.5 Å was used. The total number of MIFs calculated at each grid point with each probe depends on the dimension of the box and on the grid spacing. Since the maximum number of variables allowed to be imported in GOLPE is 1 000 000, in this calculation, where a lower number of probes was used, a smaller grid spacing was possible.

Table 3. The Probes

probe	chemical group	used in GRID/CPCA	used in GRIND/CPCA
OH2	water	Yes	No
DRY	hydrophobic probe	Yes	Yes
C3	sp3 methyl probe	Yes	No
N1	Neutral flat NH, e.g., amide	Yes	Yes
N1:	sp3 N with lone pair	Yes	No
N1+	sp3 amine NH cation	Yes	No
NM3	trimethylammonium cation	Yes	No
OH	phenol or carboxy OH	Yes	No
O-	sp2 phenolate oxygen	Yes	No
O	sp2 carbonyl oxygen	Yes	Yes

As in the previously described GRID calculation, the proteins were kept rigid (GRID directive Move=0). The calculated MIFs were imported into GOLPE, and the cut-out tool was used to focus the analysis on relevant regions within 2 Å of the benzodipyrazole ligand (CDK2_3 of Figure 1). The GRIND/CPCA analysis was also performed considering 3 and 4 Å as cut-out values. Cut-out values of 2 and 3 Å gave similar results. On the other hand, a value of 4 Å generated node−node interactions not relevant for the purpose of selectivity analysis since they are located almost outside the ATP pocket.

GRIND Generation. The pretreated MIFs obtained for each protein were imported into the program ALMOND (version 3.2)[17] to generate the GRIND variables.

The generation of the GRIND descriptors involves different steps that are reported in detail by Pastor et al.[11] Briefly, the ALMOND procedure initially extracts from each protein−probe MIF a fixed number of intense favorable (negative) energies of interaction (nodes). In principle, the selection of the nodes is based on two criteria: (1) the intensity of the field at a node and (2) the mutual node−node distances between the chosen nodes. The optimization of these two criteria can be tuned by giving to each of them a different relative weight. The second step in the ALMOND procedure involves the transformation of the extracted nodes in auto- and cross-correlograms using a method known as maximum auto- and cross-correlation (MACC-2).[18] This method computes the products of normalized interaction energy for each pair of nodes extracted at the previous step. The normalization is performed as follows: Each field energy is normalized by dividing the value of the energy of the interaction by a constant representing a theoretical maximum of energy for this probe. The distance between the nodes is binned into a discrete number of categories. In each category, only the highest product of node−node interaction energy is stored, while the others are discarded. In the current calculation, the following parameters were applied: the maximum number of extracted nodes for all the fields was set to 120, 65% of the weight was assigned to the distance

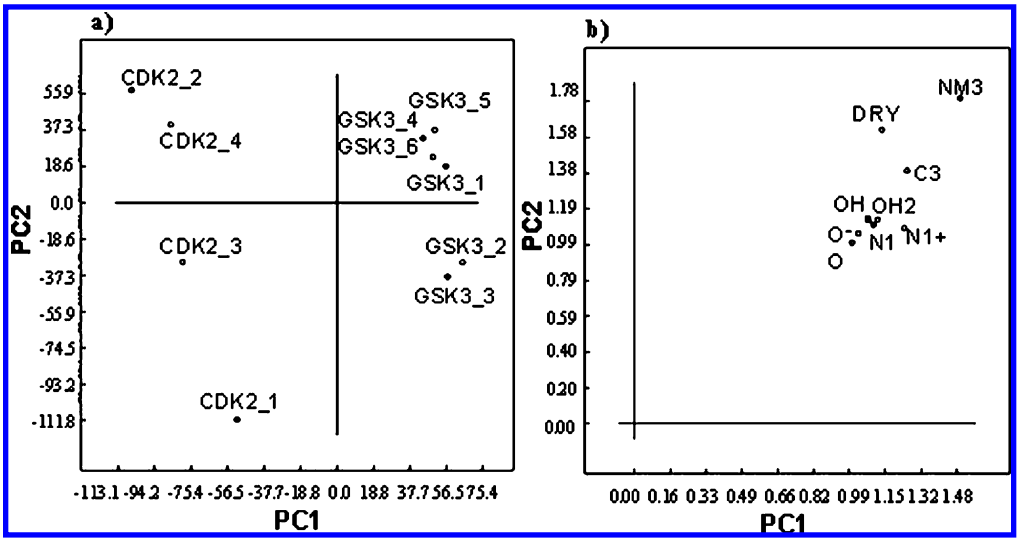


Figure 3. (a) Combined CPCA score plot for the analysis of all 10 different probes. (b) CPCA superweight obtained for the analysis with all 10 different probes.

criterion, and a smoothing window width of 0.8 grid units was used for the node−node distance range. A total of six correlograms, containing 41 variables each, were obtained for each protein. Three are autocorrelograms obtained by analyzing node−node interactions belonging to the same MIF (i.e., DRY−DRY, N1−N1, O−O). The remaining three are cross-correlograms obtained by analyzing node−node interactions belonging to different pairs of MIFs (i.e., DRY−N1, DRY−O, N1−O). Each correlogram is a compact representation of the geometrical relationship between energetically favorable regions of the MIFs.

Consensus Principal Component Analysis (CPCA). The unscaled GRIND variables generated for the CDK2 and GSK3β proteins were merged and then analyzed using CPCA modeling in GOLPE. The visual inspection of the important variables was performed in the program ALMOND.

RESULTS AND DISCUSSION

GRID/CPCA. In the GRID/CPCA approach, the MIFs data calculated by GRID with different chemical probes are organized in blocks of variables. Each block corresponds to the results of the GRID analysis with a specific chemical probe. The CPCA model is built in two levels: a block and a superblock level. As a consequence, four types of plots can be produced: the block-score plot, the superscore plot, the corresponding block-loading plot, and the superweight plot. The block-score plot represents a particular description of the analyzed proteins from the "point of view" of each block (i.e., each probe), while the superscore plot represents the integration of the information of the entire block of variables (i.e., all the probes) in a unique description. The combined CPCA superscore plot of the first two principal components (PCs), obtained analyzing the grid interaction energies between the 10 probes and all the kinase structures, is shown in Figure 3a. The two PCs cover about 38% of the variance (PC1: 28.8%, PC2: 8.9%, PC3: 7.8%, PC4: 6.5%, PC5: 6.1%). The percentage explained by the first two PCs is relatively small. This is usual for this method and might come from the noise produced in the alignment. The percentage of the variance explained by PC1 and PC2 is not an indicator of the quality of the model. The separation of the objects in the score plot and its interpretability, rather than the explained variance, is important in deciding how many PCs should be considered in the interpretation step. The distribution of the objects (i.e., proteins) in this superscore plot is almost the same for each CPCA block-score plot obtained for each single probe. The first PC discriminates between the two kinases, while the second component highlights the structural differences found within the same type of kinase. Even if we are not interested in highlighting the differences within the same enzyme, the risk of misinterpretation due to specific features present in only one structure can be avoided by increasing the number of crystallographic structures for each kinase analyzed. The main advantage of considering several distinct structures of the same enzyme is to take into account protein flexibility. As shown in Figure 2, the same enzyme can adopt different conformations. The main differences concern the conformation of the glycine-rich loop and the opening and closing of the active site cleft around the hinge.

The number and identity of the blocks more relevant for the overall CPCA model can be appreciated in the superweight plot, shown in Figure 3b. This plot is useful to understand the relevance of the different blocks in the discrimination between the two clusters of objects of Figure 3a. The block variable NM3, which has the greatest loading value on the *x* axis (PC1), contributes most to the discrimination between the CDK2 and GSK3β structures. The NM3 probe has mainly a hydrophobic character, and although it is positively charged, it does not behave as a charged probe such as N1+. The latter can be ascribed to the fact that the charge is partially shielded by the three methyl groups. This also explains why the analysis of the C3 and DRY probes leads to a similar structural interpretation in terms of the identified relevant selectivity regions. For simplicity, only the analysis of the NM3 probe is reported. In Figure 3b, the N1+ probe has similar *x* coordinates to the C3 probe but is lower in the *y* axis. The analysis of the score plot and the corresponding loading plot built for N1+ did not lead to any clear MIFs-based ligand design suggestion.

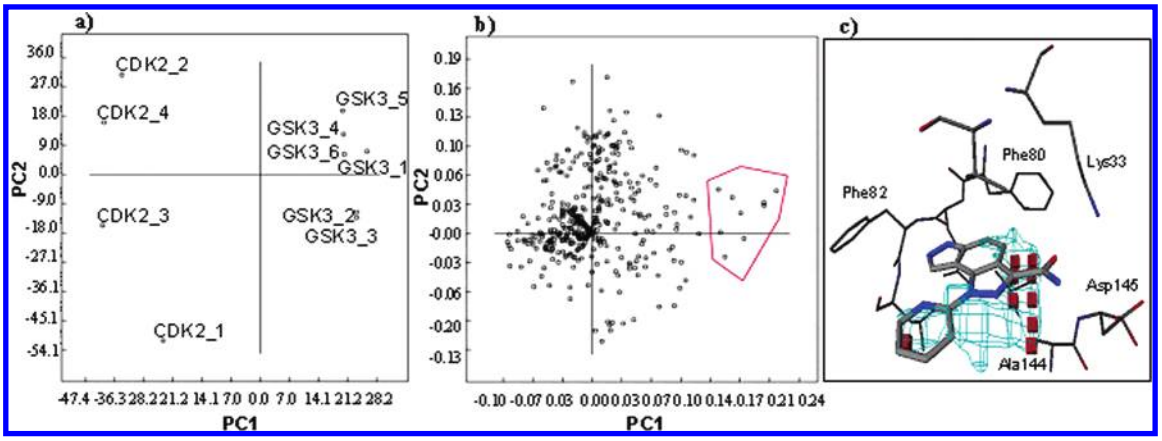


Figure 4. (a) CPCA NM3 probe score plot. (b) CPCA NM3 probe loading plot. (c) Field plot for the analysis of the NM3 probe.

Figure 4 reports the score and loading plots for NM3. In the score plot (Figure 4a), the dots are the objects of the study, that is, the proteins. In the loading plot (Figure 4b), the dots correspond to the coefficients of the variables used to characterize the proteins, that is, the MIF calculated at each point of the GRID box. The coefficients with the greatest absolute values on the first PC identify the variables responsible for the discrimination between the two sets of proteins. These variables were selected (shown circled in Figure 4b) and visualized interactively in GOLPE as red cubes in another graphical window shown in Figure 4c together with one of the CDK2 crystal structures (CDK2_3 of Table 1, referring to benzodipyrazole ligand). The variables selected in Figure 4b by the red polyhedra define a precise spatial position (grid location, represented in Figure 4c as red cubes). The visualization of these grid points allows the user to identify the areas within the protein active sites that contribute most to distinguishing between two families of proteins.

The red cubes are located in the back of the ATP pocket and identify a region (from here on called the NM3 region) made up of the side chains of Phe80, Val18, Val64, and Ala144 of CDK2. The corresponding residues in GSK3β are Leu132, Val70, Val110, and Cys199 (see Table 2). The NM3 region is not occupied by the benzodipyrazole ligand, suggesting that it could be exploited for selectivity purposes. Therefore, it was supposed that an improvement of selectivity in favor of CDK2 could be achieved by placing small hydrophobic groups, such a two methyls, at position 4 of the benzodipyrazole chemical template.

As a first approach, the 4,4-gem-dimethyl derivative was manually docked into the CDK2 and GSK3β crystal structures (CDK2_3 and GSK3_4 of Table 1) as shown in Figure 5. The larger size of GSK3−Cys199 as compared to CDK2−Ala144 and the unplanar nature of the GSK3β−Leu132 as compared to CDK2−Phe80 are translated into unfavorable steric hindrance between the 4,4-gem-dimethyl and the NM3 region of GSK3β.

Then, the 4,4-gem-dimethyl derivative (compound **3** of Table 4) was synthesized and tested against CDK2/cyclin A and GSK3β in comparison to the original compound and its aromatic counterpart (respectively compounds **1** and **2** of Table 4). As shown in Table 4, the gem-dimethyl substitution at position 4 turned out to be crucial for the selectivity. IC_{50} data clearly indicate that R═CH═CH− is preferred by CDK2 with respect to R═CH_2CH_2−. This trend is general

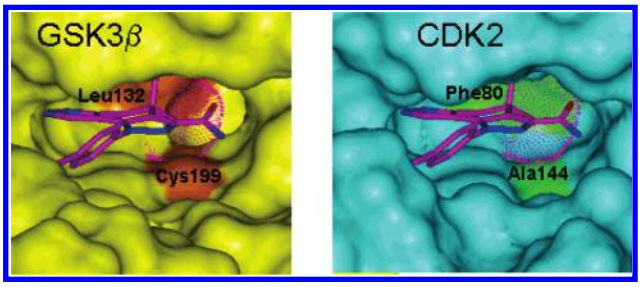
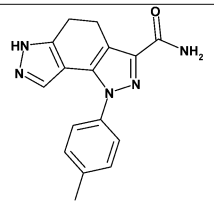
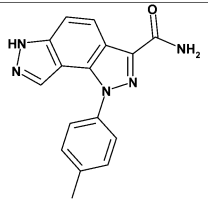
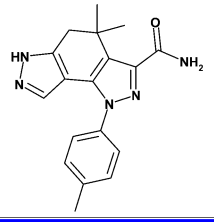


Figure 5. 4,4-Gem-dimethyl benzodipyrazole derivative docked into GSK3β and CDK2/cyclin A.

Table 4. In Vitro Activity (IC_{50} μM) of the Three Benzopyrazole Derivatives[a]

Compound	Structure	CDK2/A	GSK3β
1		0.140	0.46
2		0.015	0.023
3		0.087	>10

[a] Values are means of at least three experiments.

within this chemical class and can be explained considering the favorable T-shaped type of arrangement between the aromatic ring of the Phe80 side chain and the middle aromatic ring of the benzodipyrazole chemical template.[19] This favorable interaction is commonly present in protein−ligand complexes.[20−22] Such a T-shaped arrangement brings the $\delta(+)$ hydrogen atoms of one aromatic group into close contact with the $\delta(-)$ π-electron cloud of the other aromatic

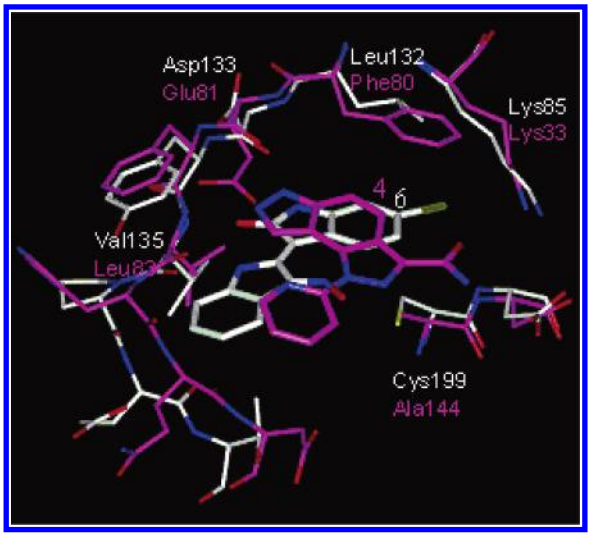


Figure 6. Overlay of benzopyrazole ligand bound into the CDK2 crystal structure (i.e., CDK2_3, magenta carbon atoms for the protein and for the compound) on that of the structure of GSK3β complexed with 6-bromoindirubin-3′-oxime (1UV5 PDB code, white carbon atoms for the protein and the ligand).

ring and makes an estimated enthalpic contribution of between −1 and −2 kcal/mol.

Meijer et al.[23,24] have provided valuable information on the CDK2−GSK3β active site differences along with specific interactions between the two kinases and indirubin analogues. They reported that a bromine substituent at position 6 of indirubin increases the selectivity for GSK3β over CDK2. The superimposition of the crystal structure of 6-bromoindirubin-3′-oxime bound into the GSK3β and the benzopyrazole ligand bound into the CDK2_3 crystal structure is shown in Figure 6. The attachment points of the bromine and the gem-dimethyl almost overlap. The two moieties occupy the same area of the ATP pocket (which correspond to the NM3 region, previously described). In the indirubin analogue, the bromine substituent takes advantage of the increased width of this pocket to achieve selectivity for GSK3β. In our benzodipyrazole class, the gem-dimethyl substituent takes advantage of the increased depth of the NM3 pocket in CDK2 to gain selectivity toward it.

Since Meijer's findings were not published at the time, we did not include bromine as a GRID probe in the GRID/CPCA study here described. To investigate whether the GRID calculation would have detected the favorable interaction between GSK3β and a bromine atom at position 6 of indirubin-3′-oxime, we carried out a retrospective GRID/CPCA calculation using all the CDK2−GSK3β crystal structures of the current study. The bromine probe identified a selectivity region for GSK3β (see the area where the red cubes are located in Figure 7); that area corresponds perfectly to the area occupied by the bromine of the 6-bromoindirubin-3′-oxime (shown as a reference in magenta in Figure 7).

Other ligands showing selectivity for GSK3β over CDK2 are reported in Table 5: the aminothiazole[25] from AstraZeneca and the *N*-phenyl-4-pyrazolo[1,5-*b*]pyridazin-3-ylpyrimidin-2-amines[26] from GlaxoSmithKline. In both cases, the authors attribute the structural basis for the selectivity to the differences highlighted in this paper (GSK3β−Leu132 vs CDK2−Phe80 and GSK3β−Cys199 vs CDK2−Ala144).

GRIND/CPCA. The six correlograms calculated for the 10 crystal structures are shown in Figure 8. In the correlograms, each point represents a pair of intense interactions separated by a certain distance. The position of the point in the *x* axis shows the distance at which this interaction was found. The product of their respective field energies is represented as the height of the point in the *y* axis.

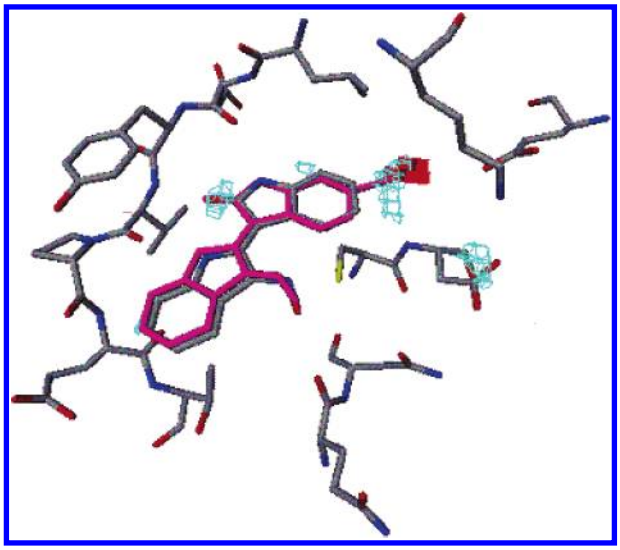
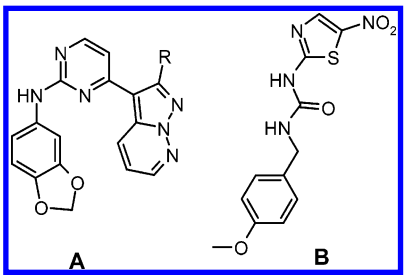


Figure 7. Overlay of indirubin-3′-oxime bound into the GSK3β crystal structure (i.e., GSK3_6, atom type colored) on that of the structure of 6-bromoindirubin-3′-oxime (magenta) extracted from 1UV5 (PDB code).

Table 5. In Vitro Activity (IC_{50} μM) of GSK3β Selective Inhibitors



template	R	CDK2	GSK3β
A	H	0.019	0.002
A	phenyl	>19.95	0.002
B		>100.000	0.104

A visual inspection of the correlograms showed that all of them, with the exception of O−O, exhibit a different trend depending on the kinase class membership, coded with different colors in Figure 8. For example, considering the DRY−DRY correlogram at short distances (left-hand part of the DRY−DRY correlogram), the GRIND variables calculated for the CDK2 structures show greater values compared to those of GSK3β. At higher distances (right-hand part of the DRY−DRY correlogram), weaker hydrophobic−hydrophobic interactions are present only in GSK3β.

The correlograms underline that there is a different geometrical relationship between the optimal interaction sites in the CDK2 and GSK3β active sites.

All the alignment-free array of GRIND descriptors calculated on the whole set of crystal structures were then used to perform a CPCA chemometric analysis. A significant CPCA model with two PCs explaining 54% of the total variance was obtained (explained variance for the first five PCs: PC1, 41.71%; PC2, 11.9%; PC3, 9.7%; PC4, 6.7%;

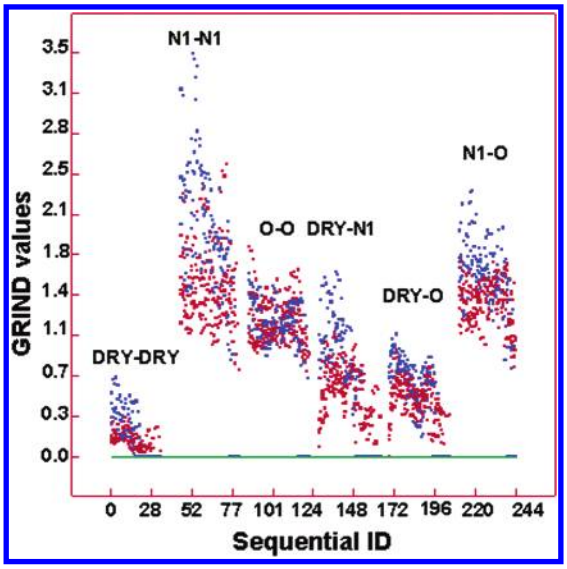


Figure 8. Auto- and cross-correlograms obtained from the ALMOND analysis of the six GSK3β and four CDK2 active sites. Points are color coded according to kinase family: red GSK3β and blue CDK2.

PC5, 5.7%). The first PC distinguishes the two classes of kinases relatively well. The superscore and superweight plots for the first two PCs are reported in Figure 9a and b, respectively.

Then, the single CPCA block score and loadings plots were generated and analyzed to identify the GRIND descriptors that mostly contribute to kinase active site selectivity regions. The O−O and the N1−N1 GRIND variables were not investigated further because of their poor ability to separate CDK2 and GSK3β crystal structures in two distinct clusters. This evidence suggests a primary importance of the hydrophobic probe.

The relevant DRY−DRY, DRY−N1, DRY−O, and N1−O GRIND variables, responsible for the CDK2−GSK3β differences, were selected from the CPCA loadings plots and visualized graphically to allow a structural interpretation. Both the nodes involved in the GRIND variables and the distance between them can be represented together with the 3D structure of the proteins.

One of the relevant DRY−DRY GRIND variables, able to discriminate between CDK2 and GSK3β, is shown in Figure 10. This GRIND variable identifies two hydrophobic sites of interaction at a distance of about 3 Å, located in the NM3 region discussed in previous paragraphs. Since these two hydrophobic nodes are relatively close, they can be considered as a single hydrophobic region that is quite extended in CDK2 and also has a strong intensity. GSK3β does not have a DRY−DRY interaction of such strength at such a short distance anywhere in the pocket. Thus, this hydrophobic region is highlighted by the GRIND/CPCA analysis as a main interaction area for the CDK2 receptor.

Thus, the analysis performed using the GRIND descriptors confirmed the presence of favorable hydrophobic areas near the Phe80 of CDK2 that can be fitted optimally by adding two hydrophobic moieties above and below the plane identified by the benzodipyrazole scaffold (nodes A and B of Figure 10).

The visual inspection of the other CDK2-specific DRY−N1 and DRY−O GRIND variables indicates that the N1 or O probes can also occupy an area near the node B area of Figure 10. That means that other groups might be added at equatorial position 4 of the benzodipyrazole scaffold. The favorable N1 region arises from a mixed hydrophobic−polar interaction with the phenyl ring of the Phe80 residue. The neutral and flat sp^2 amidic NH probe (N1) can form both a favorable pseudo hydrogen bond and a $\pi-\pi$ interaction with the phenyl of Phe80 of CDK2. The favorable O region, instead, arises from the interaction of the NH of the backbone of Asp145 and the side chain of Lys33 of CDK2. These indications were not explored because of synthetic difficulties.

As already remarked upon by looking at the correlograms, GSK3β has node−node interactions for DRY−DRY, DRY−N1, and DRY−O of a lower intensity at a higher node−node distance. Those are not present in CDK2. These interactions involve two areas about 10−15 Å apart. In every case, the node identified by the DRY probe is located in the back of the GSK3β pocket (in the area occupied by the middle ring of the benzodipyrazole), and the other node area is at the entrance of the ATP pocket. The second hydrophobic

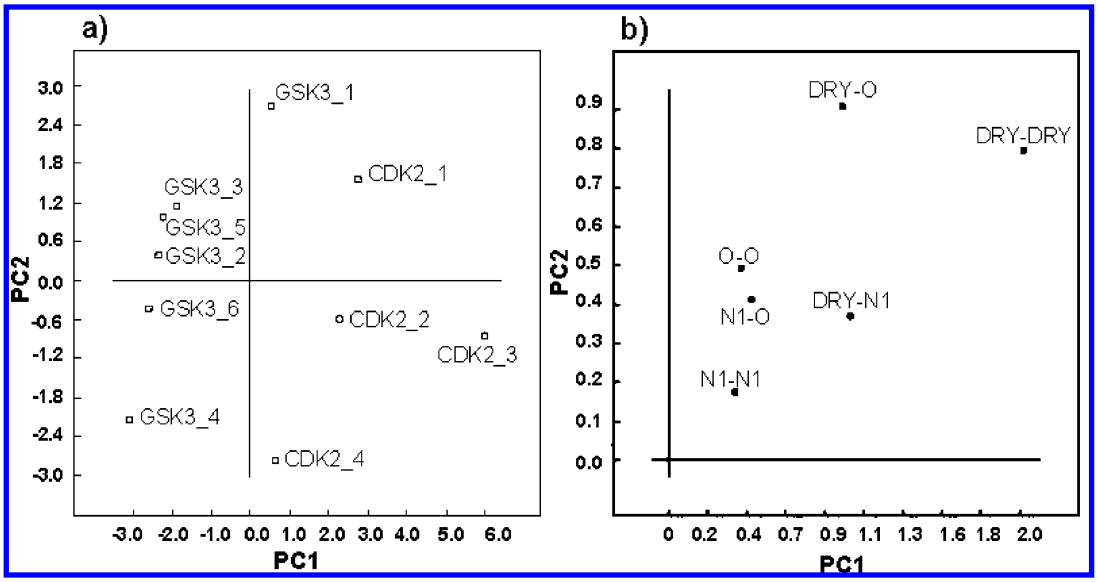


Figure 9. (a) Combined CPCA score plot for the analysis of all six correlograms. (b) CPCA superweight obtained for the analysis with all six correlograms.

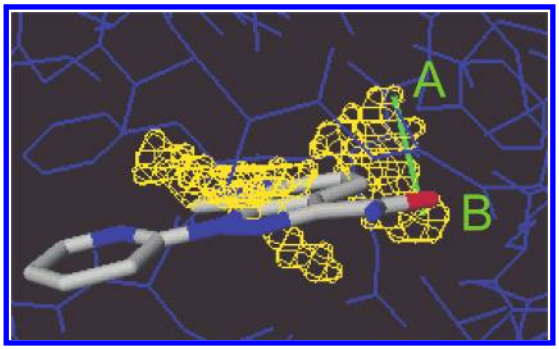


Figure 10. DRY−DRY GRIND variable able to discriminate CDK2−GSK3β proteins. In yellow, the MIFs for the DRY probe.

region at the entrance of the ATP pocket is determined by the π system of the GSK3β−Arg141 side chain. The hydrogen bond acceptor regions identified by the O probe in the DRY−O variables are also due to the GSK3β−Arg141 residue. Considering the high solvent exposure of Arg141, its high flexibility, and the fact that CDK2 also has a positive-charged highly flexible residue in that position (Lys89, see Table 2), this residue may not be optimal to be targeted to obtain selectivity toward GSK3. Another small region in GSK3β favorable to hydrogen bond acceptors is located near the OH of the Tyr134 (Phe82 in CDK2). This indication might be explored for the development of GSK3β selective inhibitors, which was not the aim of our medicinal chemistry program.

CONCLUSION

The two approaches, GRID/CPCA and GRIND/CPCA, were found to be valuable tools for elucidating the structural ATP binding pocket differences between CDK2 and GSK3β. In the present study, probes with different chemical properties were used to calculate MIFs, which were then analyzed with CPCA or used to extract the GRIND descriptors. In particular, the interpretation of the CPCA score plot for the MIF with the NM3 probe showed remarkable differences in the back of the CDK2−GSK3β ATP binding pockets. The 2D analysis of sequence differences is often useful to identify putative regions of selectivity. The GRID/CPCA approach enforces the foundation of these hypotheses by giving a more accurate 3D description of the identified selectivity regions. In particular, the subsequent ligand design benefits from such detailed and precise 3D selectivity region maps, which give indications of where proper chemical moieties should be added to gain selectivity. Moreover, this tool has the advantage of also taking protein flexibility into account.

Although both the GRID/CPCA and GRIND/CPCA approaches use the same type of information, that is, they are both MIF-based methods, they are characterized by relevant differences. The main difference is that GRIND/CPCA has the advantage that the proteins do not need to be prealigned. Thus, the preparation of the proteins requires minimal human intervention, and this approach is especially useful where it is not possible to generate a good structural alignment because of the high structural diversity of the two or more structures to be investigated. The issue of selectivity is usually encountered with very similar/homologous proteins, but it is widely accepted that totally different folds can adopt similar recognition schemes with common intermolecular protein−ligand interactions. However, the alignment-independent nature of the GRIND descriptors render them more difficult to interpret. In fact, similar interactions happening at similar node−node distances might occur in different areas of the two types of proteins.

A GRIND variable is classified as unique for one protein if it is not present anywhere in the binding site of the other protein; that is, the relative spatial orientation of the GRIND calculated for the two proteins is not considered during the comparison. On the other hand, in GRID/CPCA the spatial distribution of the MIFs derived from the aligned proteins is taken into account; that is, only the MIFs calculated for the two proteins at the same grid points are compared.

The two methods, therefore, have different strengths and weaknesses. Whenever a good alignment is possible, the authors suggest using both of the methods. The application of both of the methodologies to the same problem enhances the reliability of the results.

In the current work, the selectivity regions identified using the GRIND/CPCA approach are in good agreement with those obtained using the well-known GRID/CPCA method. Both these methods provided insights on structural features important for improving selectivity against both GSK3β and CDK2 kinases. In particular, the GRID/CPCA approach allowed us to rationally design a novel, potent, and selective CDK2 benzodipirazole inhibitor.

REFERENCES AND NOTES

(1) Vulpetti, A.; Bosotti, R. Sequence and structural analysis of kinase ATP pocket residues. *Farmaco* **2004**, *59*, 759−765.

(2) Komander, D.; Kular, G. S.; Bain, J.; Elliott, M.; Alessi, D. R.; Van Aalten, D. M. Structural basis for UCN-01 (7-hydroxystaurosporine) specificity and PDK1 (3-phosphoinositide-dependent protein kinase-1) inhibition. *Biochem J.* **2003**, *375*, 255−262.

(3) Meijer, L.; Flajolet, M.; Greengard, P. Pharmacological inhibitors of glycogen synthase 3. *Trends Pharmacol. Sci.* **2004**, *25*, 471−480.

(4) Jia, J.; Amanai, K.; Wang, G.; Tangs, J.; Wang, B.; Jiang, J. Shaggy/GSK3 antagonizes hedgehog signaling by regulating Cubitus interruptus. *Nature* **2002**, *416*, 548−552.

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

(6) Matter, H.; Defossa, E.; Heinelt, U.; Naumann, T.; Schreuder, H.; Wildgoose, P. Rational Approaches to Drug Design. *Proceedings of the 13th European Symposium on Quantitative Structure−Activity Relationships*; Prous Science: Duesseldorf, Germany, 2001; pp 177−185.

(7) Terp, G. E.; Cruciani, G.; Christensen, I. T.; Jorgensen, F. S. Structural Differences of Matrix Metalloproteinases with Potential Implications for Inhibitor Selectivity Examined by the GRID/CPCA Approach. *J. Med. Chem.* **2002**, *45*, 2675−2684.

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

(9) Naumann, T.; Matter, H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J. Med. Chem.* **2002**, *45*, 2366−2378.

(10) Pirard, B. Peroxisome Proliferator-Activated Receptors target family landscape: A chemometrical approach to ligand selectivity based on protein binding site analysis. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 785−796.

(11) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi S. GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.

(12) Benedetti, P.; Mannhold, R.; Cruciani, G.; Ottaviani G. GRIND/ALMOND investigations on CysLT1 receptor antagonists of the quinolinyl(bridged)aryl type. *Bioorg. Med. Chem.* **2004**, *12*, 3607−3617.

(13) Afzelius, L.; Masimirembwa, C. M.; Karlén, A.; Andersson, T. B.; Zamora, I. Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus noninhibitors using alignment independent GRIND descriptors. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 443−458.
(14) Gutiérrez-de-Teràn, H.; Centeno, N. B.; Pastor, M.; Sanz, F. Novel approaches for modeling of the A1 adenosine receptor and its agonist binding site. *Proteins* **2004**, *54*, 705−715.
(15) *GRID*, version 22; Molecular Discovery Ltd.: Pinner, Middlesex, U. K. (www.moldiscovery.com).
(16) *GOLPE*, version 4.5.12; Molecular Discovery Ltd.: Pinner, Middlesex, U. K. (www.moldiscovery.com).
(17) *ALMOND*, version 3.2; Molecular Discovery Ltd.: Pinner, Middlesex, U. K. (www.moldiscovery.com).
(18) Clementi, M.; Clement, S.; Clementi, S.; Cruciani, G.; Pastor, M.; Nilsson, J. E. In *Chemometric detection of binding sites of 7TM receptors*; Gundertofte, K., Jorgensen, F. S., Eds.; Kluwer Academic/Plenum Publishers: New York, 2000; pp 207−212.
(19) D'Alessio, R.; Bargiotti, A.; Metz, S.; Brasca, M. G.; Cameron, A.; Ermoli, A.; Marsiglio, A.; Polucci, P.; Roletto, F.; Tibolla, M.; Vazquez, M. L.; Vulpetti, A.; Pevarello, P. Benzodipyrazoles: a new class of potent CDK2 inhibitors" *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1315−1319.
(20) Burley, S. K.; Petsko, G. A. Aromatic−aromatic interaction: a mechanism of protein structure stabilization. *Science* **1985**, *229*, 23−28.
(21) Burley, S. K.; Petsko, G. A. Dimerization energetics of benzene and aromatic amino acid side chains. *J. Am. Chem. Soc.* **1986**, *108*, 7995−8001.
(22) Singh, J.; Thornton, J. M. The interaction between phenylalanine rings in proteins. *FEBS Lett.* **1985**, *191*, 1−6.
(23) Meijer, L.; Skaltsounis, A.-L.; Magiatis, P.; Polychronopoulos, P.; Knockaert, M.; Leost, M.; Ryan, X. P.; Vonica, A.; Brivanlou, A.; Dajani, R.; Crovace, C.; Tarricone, C.; Musacchio, A.; Roe, S. M.; Pearl, L.; Greengard, P. GSK-3-Selective Inhibitors Derived from Tyrian Purple Indirubins. *Chem. Biol.* **2003**, *10*, 1255−1266.
(24) Polychronopoulos, P.; Magiatis, P.; Skaltsounis, A. L.; Myrianthopoulos, V.; Mikros, E.; Tarricone, A.; Musacchio, A.; Roe, S. M; Pearl, L.; Leost, M.; Greengard, P.; Meijer, L. Structural Basis for the Synthesis of Indirubins as Potent and Selective Inhibitors of Glycogen Synthase Kinase-3 and Cyclin-Dependent Kinases. *J. Med. Chem.* **2004**, *47*, 935−946.
(25) Bhat, R.; Xue, Y.; Berg, S.; Hellberg, S.; Ormoe, M.; Nilsson, Y.; Radesaeter, A.-C.; Jerning, E.; Markgren, P.-O.; Borgegard, T.; Nyloef, M.; Gimenez-Cassina, A.; Hernandez, F.; Lucas, J. J.; Diaz-Nido, J.; Avila, J. Structural Insights and Biological Effects of Glycogen Synthase Kinase 3-specific Inhibitor AR-A014418. *J. Biol. Chem.* **2003**, *278*, 45937−45945.
(26) Tavares, F. X.; Boucheron, J. A.; Dickerson, S. H.; Griffin, R. J.; Preugschat, F.; Thomson, S. A.; Wang, T. Y.; Zhou, H.-Q. N−Phenyl-4-pyrazolo[1,5-b]pyridazin-3-ylpyrimidin-2-amines as Potent and Selective Inhibitors of Glycogen Synthase Kinase 3 with Good Cellular Efficacy. *J. Med. Chem.* **2004**, *47*, 4716−4730.
(27) Kim, K. S.; Kimball, S. D.; Misra, R. N.; Rawlins, D. B.; Hunt, J. T.; Xiao, H. Y.; Lu, S.; Qian, L.; Han, W. C.; Shan, W.; Mitt, T.; Cai, Z. W.; Poss, M. A.; Zhu, H.; Sack, J. S.; Tokarski, J. S.; Chang, C. Y.; Pavletich, N.; Kamath, A.; Humphreys, W. G.; Marathe, P.; Bursuker, I.; Kellar, K. A.; Roongta, U.; Batorsky, R.; Mulheron, J. G.; Bol, D.; Fairchild, C. R.; Lee, F. Y.; Webster, K. R. Discovery of Aminothiazole Inhibitors of Cyclin-Dependent Kinase 2: Synthesis, X-ray Crystallographic Analysis, and Biological Activities. *J. Med. Chem.* **2002**, *45*, 3905−3927.
(28) Pevarello, P.; Brasca, M. G.; Amici, R.; Orsini, P.; Traquandi, G.; Corti, L.; Piutti, C.; Sansonna, P.; Villa, M.; Pierce, B. S.; Pulici, M.; Giordano, P.; Martina, K.; Fritzen, E. L.; Nugent, R. A.; Casale, E.; Cameron, A.; Ciomei, M.; Roletto, F.; Isacchi, A.; Fogliatto, G.; Pesenti, E.; Pastori, W.; Marsiglio, A.; Leach, K. L.; Clare, P. M.; Fiorentini, F.; Varasi, M.; Vulpetti, A.; Warpehoski, M. A. 3-Aminopyrazole Inhibitors of CDK2/Cyclin A as Antitumor Agents. 1. Lead Finding. *J. Med. Chem.* **2004**, *47*, 3367−3380.
(29) Bertrand, J. A.; Thieffine, S.; Vulpetti, A.; Cristiani, C.; Valsasina, B.; Knapp, S.; Kalisz, H. M.; Flocco, M. Structural Characterization of the GSK-3β Active Site Using Selective and Nonselective ATP-mimetic Inhibitors. *J. Mol. Biol.* **2003**, *333*, 393−407.

CI0500280