

Explicit Treatment of Active-site Waters Enhances Quantum Mechanical/Implicit Solvent Scoring: Inhibition of CDK2 by New Pyrazolo[1,5-*a*]pyrimidines

Michaela Hylsová,^{a,†} Benoît Carbain,^{a,†} Jindřich Fanfrlík,^{b,†} Lenka Musilová,^a Susanta Halder,^{b,c} Cemal Köprülüoğlu,^{b,c} Haresh Ajani,^{b,c} Pathik S. Brahmakshatriya,^b Radek Jorda,^d Vladimír Kryštof,^d Pavel Hobza,^{b,c} Aude Echalié,^e Kamil Paruch,^{a,f*} Martin Lepšík^{b*}

^a Department of Chemistry, CZ Openscreen, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic.

^b Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nám. 2, 166 10 Prague 6, Czech Republic.

^c Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, 771 46 Olomouc, Czech Republic.

^d Laboratory of Growth Regulators, Faculty of Science, Palacký University & Institute of Experimental Botany, Šlechtitelů 27, 783 71 Olomouc, Czech Republic.

^e Centre de Biochimie Structurale. CNRS UMR 5048 - UM - INSERM U 1054. 29 rue de Navacelles. 34090 Montpellier. France, Present address: Departments of Molecular and Cell Biology and of Cancer Studies. Henry Wellcome Building. University of Leicester. Lancaster Road. Leicester. LE1 9HN. United Kingdom.

^f International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic.

KEYWORDS

cyclin-dependent kinase 2; ATP-competitive type I inhibitors; pyrazolo[1,5-*a*]pyrimidine; quantum mechanical scoring; protein-ligand binding; molecular dynamics; water thermodynamics; X-ray crystal structure

ABSTRACT

We present comprehensive testing of solvent representation in quantum mechanics (QM)-based scoring of protein-ligand affinities. To this aim, we prepared 21 new inhibitors of cyclin-dependent kinase 2 (CDK2) with the pyrazolo[1,5-*a*]pyrimidine core, whose activities spanned three orders of magnitude. The crystal structure of a potent inhibitor bound to the active CDK2/cyclin A complex revealed that the biphenyl substituent at position 5 of the pyrazolo[1,5-*a*]pyrimidine scaffold was located in a previously unexplored pocket and that six water molecules resided in the active site. Using molecular dynamics, protein-ligand interactions and active-site water H-bond networks as well as thermodynamics were probed. Thereafter, all the inhibitors were scored by the QM approach utilizing the COSMO implicit solvent model. Such a standard treatment failed to produce a correlation with the experiment ($R^2=0.49$). However, the addition of the active-site waters resulted in significant improvement ($R^2 = 0.68$). The activities of the compounds could thus be interpreted by taking into account their specific noncovalent interactions with CDK2 and the active-site waters. In summary, using a combination of several experimental and theoretical approaches we demonstrate that the inclusion of explicit solvent effects enhance QM/COSMO scoring to produce a reliable structure–activity relationship with physical insights. More generally, this approach is envisioned to contribute to increased accuracy of the computational design of novel inhibitors.

1. Introduction

Reliable prediction of protein–ligand (P–L) binding affinity is the major task of computer-aided drug design. In the structure-based drug design, the 3D coordinates of the P–L complex (obtained mostly by X-ray crystallography, NMR or docking) are used for the evaluation of binding free energy using scoring or other approaches. The application of quantum mechanics (QM) in P–L scoring¹ has expanded the methodological repertoire and allowed a quantitative treatment of P–L interactions,^{2,3} metals (either in proteins^{1,4} or ligands),⁵ exotic ligands,⁶ and noncovalent interactions of quantum origin, e.g. σ -hole bonding,^{7,8} or covalent binding of ligands.⁹

The treatment of solvation in QM scoring needs to be fast and is thus usually done using continuum approaches of varying degree of accuracy.^{1,10-12} Despite the success of QM-based models (such as COSMO),¹³ some P–L complexes need explicit and dynamic treatment of solvation for a correct description.^{14,15} This approach can be further extended to evaluate enthalpic and entropic contributions of specific water molecules to ligand binding.¹⁶⁻¹⁹ Theoretical approaches crucially rely on data obtained by various experimental methods, such as X-ray crystallography, isothermal microcalorimetry or kinetic assays.^{17,20}

Cyclin-dependent kinase 2 (CDK2) is the best studied member of the CDK family of Ser/Thr protein kinases, which are essential components of cell-cycle regulation.²¹ Since the activities of CDKs are frequently deregulated in various types of cancers, these kinases have been targeted by numerous experimental anticancer therapeutics.²² The recent approval of the CDK4/CDK6 inhibitor palbociclib for the treatment of ER-positive and HER2-negative breast cancer is a culmination of these efforts.²³ CDKs phosphorylate their substrates using ATP as a phosphate donor. The ATP molecule binds to the CDK active site located in a cleft (“hinge region”) which is located between the N-terminal β -sheet and C-terminal α -helical domains. The vast majority of known CDK inhibitors are of type I, i.e. directly competing with ATP for the binding site.²⁴ Detailed understanding of the binding of small-molecule inhibitors to CDK2 has been obtained through hundreds of X-ray structures of co-crystal complexes.²⁴ However, only about one-fifth of them include the active state of CDK2 (as a ternary complex with a regulatory cyclin subunit), whose geometry is more suitable for the prediction of an interaction with small-molecule inhibitors.²⁵ In addition, the water content in the active site of CDK2 complexed with cyclin is different due to differently placed segments or amino-acid side chains (such as Lys33, Asp145 or Glu51).²⁵

One class of potent and selective CDK2 inhibitors (which includes dinaciclib, currently profiled in clinical trials²⁶) utilizes the pyrazolo[1,5-*a*]pyrimidine (PP) core (Fig. 1).²⁷⁻³⁰ The exploration of the structure–activity relationship (SAR) at positions 3 and 5 of the core has revealed that: i) the substitution of hydrogen (as R³) by bromine or ethyl can lead to a 10- to 100-fold increase in the binding affinity, but analogs containing larger substituents are significantly less active,^{28,30} and ii) compounds possessing proper unsaturated and saturated five- or six-membered rings at position 5 exhibit IC₅₀ values below 20 nM.²⁸ The significant increase in activity upon replacing R³=H by the bulkier Br or Et moieties has been ascribed to the optimal filling of the hydrophobic cavity in the gatekeeper region around Phe80.²⁸ This explanation, based on inactive binary CDK2/inhibitor co-crystal structures, also includes the expulsion of an extra water molecule (denoted here as W*), which can be accommodated in sub-series with R³=H but not with R³=Br/Et.²⁸ However, individual contributions of these factors to the overall binding affinity have not been reported to date.

Figure 1

The project reported herein is a non-trivial extension of our ongoing efforts in the area of protein kinase inhibitor design.³⁰⁻³⁴ Specifically, we report the synthesis and biological activity of a series of heretofore unknown 21 PP-based CDK2 inhibitors bearing “standard” H, Br and Et substituents at position 3 and previously unexplored biphenyl substituents at position 5. The crystal structure of the active CDK2/cyclin A ternary complex with one of the most potent inhibitors from the series serves as a starting point for molecular dynamics (MD) calculations, water thermodynamics and QM/COSMO-based scoring taking into account explicit-solvent effects. Overall, the project is a comprehensive experimental–theoretical study of challenging CDK2/inhibitor complexes and suggests some potential avenues for near-future computer-aided drug design.

2. Results and Discussion

2.1. Compound Design and Synthesis

We prepared a series of new PP-based compounds **1a–u** (Table 1) with H, Br or Et at position 3, the 3-aminomethylpyridine moiety at position 7, and biphenyl substituents at position 5. To our knowledge, no PP-based kinase inhibitors with biphenyls at position 5 have been described to date. We put several polar substituents on the distal phenyl ring, hypothesizing that binding

to the enzyme could be enhanced via noncovalent interactions to this previously unexplored part of the active site.

Adopting the methodology we had developed previously,^{28,35} we prepared key intermediates **6a–6c** and **7a–7b**, which were used for the subsequent installation of the substituted biphenyl moiety via Suzuki coupling (Scheme 1). The syntheses of the biphenyls containing single substituents were quite straightforward and provided the target compounds **1a–1q** indicated in Table 1.

Table 1

Scheme 1

While the syntheses described above proceeded uneventfully, the preparation of brominated analogs **1r** and **1s** proved non-trivial, as the bromination of hydroxylated biphenyl derivatives was non-regioselective and proceeded on both the phenol and the pyrazole rings. Eventually, a late-stage selective bromination of derivatives containing 2- and 3-hydroxy-[1,1'-biphenyl]-4-carboxamide moieties was achieved by deactivating the phenol moiety via its protection as mesylate or pivaloate. The final cleavage of the mesylate or pivaloate under basic conditions, followed by the removal of the Boc group, afforded the target compounds **1r** and **1s** (Scheme 2).

Scheme 2

The synthesis of the analogs with nitrile-containing side chains was also relatively challenging. Direct Heck couplings of acrylonitrile with commercially available 2-bromo-4-nitrobenzoic acid or its amide were unsuccessful. However, the reaction did proceed on ester **9**, providing the acrylic nitrile derivative, which after hydrogenation afforded compound **11**. Next, the reduction of the nitro group followed by the conversion of the amine into bromide provided compound **13**, which then provided (via the Sandmeyer reaction and conversion of the ester into an amide) the desired, heretofore unknown, intermediate **15**. The subsequent Suzuki reaction, bromination and final deprotection proceeded smoothly and afforded the target compounds **1t** and **1u** (Scheme 2).

2.2. The Structure–Activity Relationship

The experimental IC₅₀ values spanned three orders of magnitude, from the weakest-binding compound **1a**, bearing unsubstituted biphenyl at position 5 and hydrogen at position 3 (12.6 μM), to the most active compound **1s** with a doubly substituted distal phenyl ring (R⁵ = *p*-CONH₂-*o*-OH-Ph) and R³ = Br (0.015 μM) (Table 1). In general, the compounds could be clearly divided into two groups based on the identity of the R³ substituent: compounds with R³ = H were weaker binders with IC₅₀ ≥ 2 μM, whereas compounds with R³ = Br or Et were significantly more potent (IC₅₀ ≤ 0.5 μM, Table 1). This structure–activity relationship (SAR), compatible with the previous observations for PP-based CDK2 inhibitors with smaller (phenylic) R⁵ substituents,²⁸ was previously explained by the filling of the hydrophobic cavity near the gatekeeper residue Phe80.^{27,28} Since this interaction profoundly influences the compounds' activity, we address it here using advanced computations (see Section 2.6).

The SAR at position 5 was independent from that at position 3. It revealed these trends:

- i. Unsubstituted biphenyls (i.e. compounds **1a** and **1b**) are less potent than analogous PP-based inhibitors with phenylic substituents at position 5.²⁸
- ii. However, installation of polar substituents on the biphenyl moiety can significantly improve the potency.
- iii. Polar substituents have to be at a proper distance from the distal phenyl ring (cf. compounds **1d** and **1i**); additional suitably oriented polar motifs can further improve the inhibitory activity (cf. compounds **1l** and **1s**).

The interaction of the polar motifs on the biphenyl part with the protein is therefore one of the key determinants of the compounds' activity; consequently, we addressed it by QM scoring (see Section 2.7).

2.3. The Crystal Structure of the Active Ternary CDK2/Cyclin A/1l Complex

In order to gain additional insight into the interaction of our inhibitors with the enzyme, we determined the crystal structure at 2.4 Å resolution of the active Thr160-phosphorylated (pThr160) CDK2/cyclin A ternary complex with one of the most potent compounds of the series, compound **1l** (PDB code: 5LMK). The data collection and refinement statistics have been compiled in the Supporting Information (Table S1). The inhibitor is bound in the active

site using the standard binding mode observed for type I inhibitors. The asymmetric unit contains two CDK2/cyclin A heterodimers. Overall, the two CDK2 molecules present in the asymmetric unit (chains A and C) assume very similar structures (the backbone RMSD of 0.47 Å excluding the flexible regions not supported by electron density: the missing loops 38–41 in the A chain and 220–251 in the C chain and the C-terminus in both chains, residues 294–298). The inhibitor bound in their ATP binding sites interacts via the same binding mode. The water-molecule patterns in their respective ATP cavities are also very similar. The description of the inhibitor bound to the CDK2 ATP binding site that follows is based on the analysis of the CDK2 chain A. Any deviation from this binding mode and the water molecule pattern in the two CDK2 chains is indicated explicitly.

Specifically, the pyrazolo[1,5-*a*]pyrimidine core binds to CDK2 via three hinge region H-bonds (two classic 7-amino N-H...O:Leu83 and N1...HN:Leu83 bonds and one weak C2-H...O:Glu81 bond) (Fig. 1 – dashed; Fig. 2 – dotted). The R³ substituents of the PP core face the gatekeeper Phe80 and the pyridyl moiety occupies the specificity surface (Figure 1) of the ATP-binding domain of CDK2. The proximal phenyl ring of the R⁵ biphenyl moiety (C19-C24, Figure 2) is connected to the PP core residue in the CDK2 ribose-binding pocket roughly in-plane with a C24-C19-C3-N4 dihedral of 33°. The distal phenyl ring (C25-C30, Figure 2) extends in the phosphate-binding pocket toward Lys129 on the surface of the enzyme (Figures 1 and 2). It is again roughly in-plane with the PP core (a C30-C25-C3-N4 dihedral of -29°), but tilted in the opposite direction as compared to the proximal phenyl ring. The distal-to-proximal phenyl dihedral C26-C25-C22-C21 is thus twisted at -63°/-87° (the A/C chain of CDK2, respectively). The terminal carboxamide group of the inhibitors is slightly rotated with respect to the distal phenyl – the O32-C31-C28-C29 dihedral angle is 42/28° (the A/C chain of CDK2, respectively, Figure 2). It is noteworthy that the electron density corresponding to the carboxamide moiety of the inhibitor bound to the CDK2 chain C is significantly weaker than in the chain A, probably due to a different crystal environment. In the surroundings of the inhibitor, there is a chain of five water molecules (W77, W206, W194, W147 and W130) occupying the region from the gatekeeper Phe80 to the phosphate-ribose pocket (Figures 1 and 2). Additionally, a water molecule with low occupancy (W224; ca 0.30) has been found between the distal phenyl ring and the CDK2 side chain of Asp145 (Figure 2). The corresponding water molecule bound to the CDK2 chain C (W171) has been modeled with full occupancy.

Figure 2

2.4. The Thermodynamics of Explicit Active-site Waters in the CDK2/11 Complex

The CDK2/11 crystal structure described above has been treated by the WaterMap program to assess the dynamics and thermodynamics of the water molecules. The clustering has confirmed the sites for all six active-site waters (W77, W206, W194, W147, W224 and W130) with full occupancy (Figure 2, Table S2, Figure S1). This computational overestimation of the structural stability of water sites (cf. the crystallographic occupancy of W224 of 0.30, see above) stems mostly from the rigid protein/rigid ligand approximation embodied in the WaterMap. Thermodynamically, the binding of all six water molecules in their protein sites is unfavorable with respect to their ΔG in the bulk solution (Table 2). This is mainly due to the entropic cost of trapping them in the protein, which is not overcome by sufficiently strong H-bond interactions (enthalpy gain) with the protein/ligand/water in the cavity. In the case of W206 and W130, even the ΔH term is unfavorable (Table 2) because of the closeness of the hydrophobic Phe80 and the proximal phenyl of **11**, respectively. As discussed below, the thermodynamic parameters of water are crucial in QM scoring.

Table 2

2.5. Molecular Dynamics of the CDK2/11 Complex

To understand the full dynamics of the CDK2/11 complex, we performed standard molecular dynamics (MD) on the crystal structure surrounded by explicit water molecules. A correct description of the bromine atom of the inhibitors (within the limits of the MD approach) was enabled by the use of an explicit sigma hole (ESH) (see the Methods) to account for possible X... π interactions with Phe80. Two variants of the terminal *para*-carboxamide moiety were used, because the potential hydrogen bond with Lys129 represents only a weak crystallographic restraint (the **11**:N33...NZ:Lys129 distance of 3.7/3.8 Å in the A/C chain of CDK2, respectively). The original crystallographic variant is denoted as amide **A** conformation (i.e. an O32-C31-C28-C27 dihedral of 42°) and the alternative orientation, **B** conformation, related by a roughly 180° flip (due to a slight nonplanarity on C31), has a dihedral of -139°. Because the biphenyl moiety also rotates in MD, we have opted for the O32-C31-C3-N4 dihedral (i.e. the orientation of the terminal carboxamide with respect to the PP core) to monitor the orientational preferences of the terminal carboxamide group. The crystallographic values of the O32-C31-C3-N4 dihedral are -158/159°, respectively, for the A/C chain of CDK2 for the original **A**

conformation and 15/-25°, respectively, for the **B** conformation (Figure S2, blue and gray bars, respectively).

Terminal Carboxamide Rotation. First, we compared the root-mean-square deviations (RMSD) of the protein backbone and of non-hydrogen inhibitor atoms in the course of 80 ns. Upon comparing the **A** and **B** conformations, the latter was found to be stabler both for the protein (the average RMSD of 1.4 Å and a drift of 0.2 Å vs. 1.1 Å and a drift of 0.1 Å, respectively) (Figure S3A) and the inhibitor (3 plateaus with the RMSD up to 1.5 Å vs. 1 plateau of the RMSD at 0.9 Å). Since the greater stability of **B** was also confirmed by QM calculations (as described below), only this conformer was taken into account. The inhibitor RMSD oscillated with the amplitude of ca 0.4 Å in the first 40 ns and dropped down to ca 0.1 Å in the second half of the trajectory, being especially stable in the last 20 ns (Figure S3B). This was concomitant with the formation of the H-bond between the inhibitor's carboxamide in the conformation **B** and Lys129 (**11**:O32...NZ:Lys129) after 40 ns and its maintenance after 60 ns (Figure S3A). The plotting of the populations of the O32-C31-C3-N4 dihedral throughout the whole 80 ns trajectory (Figure S2, violet curve) has revealed that both conformations **A** and **B** are present and correspond to the crystallographic values (Figure S2, blue and gray bars). Figures S2, S4B (i.e. plots of isolated populations in different parts of the trajectory) demonstrate that the conformation **A** prevails in the first 40 ns while the conformation **B** is present predominantly after 40 ns. Taken together, the analyses of 80-ns MD show that both conformations **A** and **B** of the terminal carboxamide can be present in the structure of the CDK2/**11** complex.

Active-site Water Molecules. Water densities in the CDK2 active site averaged over the 80-ns MD trajectory clearly show four spherical regions corresponding to the crystallographic water molecules W77, W206, W194 and W147 (Figure S5). These water molecules are present throughout most of the simulation time, forming a hydrogen-bonded chain and linking the inhibitor N4 nitrogen (Figure 2) with the terminal side-chain amine of Lys33. In line with the crystallographic structure, W147 H-bonds with W194 (the corresponding crystallographic distance is 3.0 Å). The MD density for W130 is also present at the crystallographic position, but it is weaker when compared to the four above-mentioned water molecules. This corresponds to the lack of anchoring interactions and only a weakly directional O-H... π interaction with the proximal phenyl of **11** (Figure S5). The site for W224 is split into two densities, one closer to W130 and one farther toward the bulk solvent. Each of the MD densities is occupied by a

different water molecule, thus completing the H-bond chain extending from W130 toward the bulk solvent (Figure S5). This transient water chain could not be captured by X-ray crystallography due to its dynamic nature (the crystallographic W130...W224 inter-oxygen distance was 4.5 Å, cf. Figure 2). This was also evidenced by the low occupancy of W224 in the A chain of the crystallographic structure.

2.6. The QM/COSMO Treatment of the Effects of R³ Substituents

As described in Section 1.2., the studied inhibitors can be clearly divided into two groups based on the identity of the R³ substituent: compounds with R³=H are weaker binders with IC₅₀ ≥ 2 μM while compounds with bromine or ethyl moiety as R³ have IC₅₀ ≤ 0.5 μM. The previous qualitative structure-based explanation was based on filling the hydrophobic cavity near the gatekeeper residue Phe80.^{27,28}

Direct R³-Gatekeeper Interactions. Here, this question is approached quantitatively based on QM/COSMO interaction ‘free’ energies between three inhibitors possessing all three substituent types at R³ (i.e. compounds **1j**, **1k** and **1l**; Table 1) and the Phe80 side chain. We have found that compounds **1k** and **1l** have stronger interactions with the side chain of Phe80 (-1.7 and -1.8 kcal/mol, respectively) than **1j** (-0.8 kcal/mol), which is in agreement with the previous explanation.^{27,28} More specifically, the effect can be attributed to dispersion-driven noncovalent interactions (X...π and CH...π, respectively) of Br and Et substituents at position 3 with the phenyl ring of Phe80 and the lack thereof when R³=H. Based on this indication of the affinity difference between the R³ subseries, we proceeded to the modeling and scoring of the whole 21-compound series (Table 1), using the advanced QM/SQM scoring function.^{2,3}

2.7. The QM/COSMO Scoring of the Entire Inhibitor Series

Terminal Carboxamide Conformation. Prior to productive QM/SQM/COSMO scoring, we carefully selected the conformation of the terminal carboxamide. The preliminary scoring of the **A** and **B** conformations of inhibitors **1j–1u** (Table 1) revealed that the scores for the conformation **B** were systematically stabler by approximately 3 kcal/mol. Structurally, this was caused by the formation of the **1l**:O32...NZ:Lys129 H-bond (the O...N distance of 2.9 Å), which is similar to that found in MD after 60 ns (cf. Figure S4A). We therefore used the conformation **B** for the QM scoring of all inhibitors possessing the terminal carboxamide (and N,N-dimethyl carboxamide) motif.

Implicit Solvent QM Scoring. All complexes of CDK2 and the PP-based inhibitors, including the conformational variants of the distal phenyl substituents, were optimized using the QM/SQM/COSMO procedure. The resulting structures will be briefly described here. The terminal carboxamide formed the H-bond mentioned above. The amide methylation (**1m**, **1n**) did not change the geometry of the H-bond. Similarly, the hydroxymethyl group of **1g**, **1h** and **1i** formed the O32...H-NZ:Lys129 H-bond with the O...N distance of 2.9 Å. Wherever the carboxamide group was replaced by smaller amino or hydroxyl groups (**1c**, **1d**, **1e** or **1f**), the H-bond with Lys129 was weakened (the O...N distance increased to about 3.7 Å). The introduction of a second substituent on the distal phenyl did not change the orientation of the biphenyl in the geometry of the complex.

For the entire inhibitor series, we used our standard QM-based scoring,⁸ i.e. no explicit water molecules were present and the solvent was modeled by the implicit COSMO model.³⁶ As shown in Figure 3A, there was a poor correlation of the QM score with the experimentally derived ΔG_{bind} (expressed as $RT \cdot \ln(\text{IC}_{50}/2)$). The coefficient of determination, R^2 , was 0.49 and the predictive index, PI, was 0.76. The PI represents a measure of the rank-order prediction: 1 stands for an always correct prediction, 0 for a random prediction and -1 for an always incorrect prediction.³⁷ Therefore, the differences in affinity between the $R^3=\text{H}$ and $R^3=\text{Br/Et}$ subseries cannot be explained solely by the small interaction with Phe80 (as calculated in Section 1.6), which corresponds to ca 1 kcal/mol.

Figure 3

Mixed Explicit/Implicit Solvent QM Scoring. We and others have recently realized that structural waters have important consequences for docking and QM scoring.^{13,16,19} Therefore, we considered the six explicit water molecules (W77, W206, W194, W147, W224 and W130) revealed by X-ray crystallography and confirmed by WaterMap simulations in the CDK2/**1l** complex (see above) in QM optimization and scoring.

In the $R^3=\text{H}$ subseries, the possibility of an additional water molecule (W^*) filling the space around the position 3 of the PP scaffold was suggested by WaterMap calculations on CDK2/**1j** (Table S2, Figure S1); therefore, we included it optionally. In specific cases, such as **1o**, the possible expulsion of W224 was also taken into account. For mixed explicit/implicit solvent QM scoring, the entropies of bridging water molecules, as obtained from WaterMap calculations (Table S2), were added to the QM score calculations (see the Methods).

Overall, the QM scoring with mixed explicit/implicit solvent treatment accounted correctly for the affinity difference between the $R^3=H$ and $R^3=Br/Et$ subseries and also made it possible to correctly describe the interactions of the substituents in *ortho*- and *meta*- positions of the biphenyl moiety. This resulted in a fair coefficient of determination (R^2 of 0.63) and the predictive index (PI of 0.80) – see Figure 3B. Upon including W^* in the $R^3=H$ subseries, the correlation of the whole inhibitor set slightly increased to R^2 of 0.68 and PI to 0.85. It is thus evident that the proper addressing of both the structural and thermodynamic effects of all the conserved active-site water molecules is indispensable in the QM scoring for reliable a description of the affinity trends. Next, we investigated the molecular determinants of affinity.

2.8. The Molecular Determinants of Potency

All the QM scoring terms for the entire inhibitor series are summarized in Table S3. In line with our previous studies,^{2,3} the computed scores were dominated by the ΔE_{int} and $\Delta\Delta G_{solv}$ terms. The $\Delta G'_{conf}(L)$ and $-T\Delta S_{solv}$ terms had similar values for all the studied inhibitors. The least active compound **1a** had the least negative ΔE_{int} term. Although the desolvation penalty ($\Delta\Delta G_{solv}$) of **1a** was also the smallest within the whole series, it did not compensate for the weak interactions of the unsubstituted phenyl ring and the missing substituent at the R^3 position. On the other hand, the most active compounds **1r** and **1s** possessed the *para*-carboxamide group and the hydroxyl group in *meta*- or *ortho*- positions, respectively, on the distal phenyl ring of the biphenyl, and a Br atom at the R^3 position. Interestingly, the hydroxyl group was crucial for the formation of two important H-bonds: with W224 and Lys129 in the former case and with W224 and W130 in the latter (Figure 4). Such interactions were responsible for the highest QM/SQM score of **1s**, due to the most negative ΔE_{int} .

Figure 4

3. Conclusions

Herein we report the design, synthesis, activity measurements, and crystallographic and computational analyses of a series of 21 new pyrazolo[1,5a]pyrimidine-based CDK2 inhibitors. The compounds contain R^5 -substituted biphenyl moieties that allowed for the exploration of the phosphate-ribose pocket of the CDK2 active site. The measured activities of all the new compounds against CDK2 span three orders of magnitude. For the molecular understanding of

the potency determinants, we have determined the X-ray structure of the active ternary complex of CDK2/cyclin A/**11**.

We have shown that active-site explicit waters and their thermodynamics need to be included in the computational description of biomolecule-inhibitor systems in order to obtain a good correlation with the experimental binding affinities. Modeling, QM scoring, MD simulations and water thermodynamics calculations have revealed that explicit-solvent effects need to be included in the scoring procedure for this system to obtain meaningful results. Using a non-trivial combination of the above-mentioned techniques, we have been able to dissect the role of interactions of R³ substituents with the Phe80 gatekeeper residue, the effect of active-site waters, and the roles of the distal phenyl substitutions.

This study has important specific consequences for the design of CDK inhibitors. We envision that our methodology (when applied more generally) will contribute to the efforts focused on increasing the prediction accuracy of computer-aided drug design.

4. Methods

4.1. Organic Synthesis

The conditions and reagents used in the synthesis of all new compounds are described in Schemes 1 and 2. Detailed experimental procedures and the characterization of the compounds are given in the Supporting Information.

4.2. Biochemical Measurements

The tested compounds were dissolved in DMSO and diluted with water (the concentration of DMSO in the reaction never exceeded 0.2%). The CDK2/Cyclin E complex was produced in Sf9 insect cells via baculoviral infection and purified on a Ni²⁺NTA column (Qiagen). Kinase (approx. 10 ng) was assayed using a mixture of the following: 1 mg/mL of histone H1, 15 μ M of ATP, 0.05 of μ Ci [γ -33P]ATP, the tested compound, and reaction buffer, in a final volume of 10 μ L. The reaction buffer consisted of: 60 mM of HEPES-NaOH, pH 7.5, 3 mM of MgCl₂, 3 mM of MnCl₂, 3 μ M of Na-orthovanadate, 1.2 mM of DTT, and 2.5 μ g / 50 μ L of PEG_{20,000}. The reactions were stopped by adding 5 μ L of 3% aqueous H₃PO₄. Aliquots were spotted onto P-81 phosphocellulose (Whatman), washed 3 times with 0.5% aqueous H₃PO₄, and finally air-dried. Kinase inhibition was quantified using a FLA-7000 digital image analyzer (Fujifilm).

The concentration of each tested compound required the decrease of the CDK activity by 50%. The IC₅₀ values were determined from the dose-response curve.

4.3. X-ray Crystallography

The expression of the recombinant human pThr160 CDK2 and Cyclin A proteins in *Escherichia coli* and the purification of the binary complex were carried out as described in.³⁸ The co-crystallization of the protein complex at 12 mg.mL⁻¹ in the presence of the inhibitor was done using published crystallization conditions, with (NH₄)₂SO₄ and KCl as the precipitant agents,³⁹ after the incubation of the protein complex with 2 mM of inhibitor **11** dissolved in DMSO for 20 min on ice. Crystals grew in a few weeks and were subjected to brief soaking in 7 M of sodium formate as a cryoprotectant step before being flash-frozen in liquid nitrogen. The diffraction data were collected on the European Synchrotron Radiation Facility (ESRF) ID29 beamline. Data processing was carried using programs of the CCP4 suite.⁴⁰ The structure of pThr160 CDK2/Cyclin A in complex with inhibitor **11** was solved by molecular replacement using the program Phaser⁴¹ and the high resolution structure of pThr160 CDK2/Cyclin A in complex with roscovitine (PDB code 3DDQ)⁴² as the search model. A clear solution emerged from the molecular-replacement procedure with two ternary complexes in the asymmetric unit. After rigid body refinement, the initial difference electron density clearly highlighted the presence of the inhibitor bound in the ATP-binding site in both copies of CDK2 found in the asymmetric unit. A standard procedure of iterative building in Coot⁴³ and refinement in Refmac⁴⁴ was used. The final complex was saved in the Protein Data Bank (PDB) under accession code 5LMK.

4.4. Computational Methodology

For the three types of calculations, i.e. molecular dynamics (MD), WaterMap and quantum chemical (QM) scoring, specific preparations of the crystal structure of the CDK2/Cyclin A/**11** ternary complex were done as described below. The CDK2 chain A was consistently used, whereas the cyclin A subunit and the magnesium ion were removed.

4.4.1. Molecular Dynamics Simulations.

System setup. The missing loop between the residues 38–41 was modeled using the MODELLER program⁴⁵ via the UCSF Chimera package.⁴⁶ Hydrogens were added by the LEaP program of the AMBER14 suite.⁴⁷ Subsequently, the loop with its flanking residues (residues 37–44) was minimized using the Sander program of AMBER14 package⁴⁷ using 500 cycles of

the steepest descent, followed by 300 cycles of the conjugate gradient method. The CDK2/**11** complex was prepared by a procedure described previously (see below).⁸ Explicit TIP3P water molecules were added to fill the dodecahedron box repeating in space using periodic boundary conditions (PBC). The system was neutralized by adding two Cl⁻ ions. For the protein, the AMBER ff03.r1 force field⁴⁸ was used. For the ligands, the general AMBER force field (GAFF)⁴⁹ with HF/6-31G* RESP charges⁵⁰ and fragment-based approach⁵¹ were used. Explicit sigma-hole (ESH) using the “all-fit” (AF) approach⁵² was utilized for the ligands with R³=Br to describe potential X... π interactions.

MD simulations. All the MD simulations and most of the analyses were run in GROMACS 4.5.4⁵³ on CPUs. The starting relaxation steps included minimizations and heating MD steps, followed by the MD production run. First, only the explicit waters were minimized and then heated slowly in MD to 300 K in the canonical NVT ensemble, i.e. with a fixed number of particles (N), volume (V) and temperature (T). Next, the whole system was minimized except for the backbone of the protein. Then, the whole system was heated to 300 K. The equilibration simulation in an isothermal–isobaric (NPT) ensemble, i.e. with a fixed number of particles (N), pressure (P), and temperature (T) consisted of a 5-ns run with a velocity-rescale thermostat and a Berendsen barostat, respectively. Finally, the production run of 80 ns was carried out with a Nose–Hoover thermostat (the reference temperature of 300 K, the coupling coefficient $\tau_T=0.5$ ps) and a Parrinello–Rahman barostat (the reference pressure of 1 bar and the coupling coefficient of τ_P 0.5 ps).

The Newton’s equations of motion were integrated using a 1.0-fs time step with the leap-frog algorithm. The 10 Å non-bonded cut-off was used for long-range electrostatics by the particle-mesh Ewald (PME) method as well as for the Lennard–Jones (LJ) interactions. All the bonds were constrained with the LINCS algorithm.

The visualization of the MD trajectories was done in VMD 1.9.1.⁵⁴ The water density was calculated by the VolMap tool⁵⁵ in VMD 1.9.1.

4.4.2. WaterMap Calculations.

The CDK2/**11** and CDK2/**1j** complexes were prepared with the Protein Preparation Wizard in Maestro⁵⁶ using the default options. The crystallographic water molecules were retained in CDK2/**11** and transferred to CDK2/**1j**. The standard settings for the WaterMap^{57,58} calculations were used except for the following: the protein residues outside the 15 Å shell around the inhibitor were removed and the MD of explicit waters was run for 10 ns.

4.4.3. Quantum Mechanical (QM) Calculations.

System setup. Residues 38 and 41 were capped by N-methyl amide and acetamide, respectively, at their C- and N-termini. All the crystallographic waters were initially deleted following the standard procedure.³ The mixed explicit–implicit solvent procedure (see the Results) included six structural water molecules (W77, W206, W194, W147, W130 and W224) in their crystallographic positions (Figure 2). For the ligands with R³=H, one additional water molecule (W*) was optionally added to the position defined by WaterMap. The compound **1o** was additionally scored without W224.

The REDUCE program of AMBER14⁴⁷ was used to determine potential flips of Asn, Gln and His residues. Histidines were modeled as neutral or monoprotonated, based on the visual inspection of the surroundings: His121 was protonated on N δ and other His residues on N ϵ . All the Lys, Arg, Asp, Glu residues as well as the N- and C-termini of CDK2 were modeled in their charged state. Hydrogen atoms were added by the LEaP module of AMBER14⁴⁷ and relaxed by a short high-temperature molecular dynamics run (for 2 ps at 2400 K and then cooled down to 0 K for 8 ps).

Modeling. All 21 inhibitors in complex with CDK2 were modeled from the crystallographic CDK2/**1l** complex in PyMol⁵⁹ by deleting/building atoms. The compounds with substituents in the *ortho*- and *meta*- positions on the distal biphenyl ring were considered in two orientations, related by a 180° flip of the distal phenyl, to account for both possible binding orientations with respect to the protein. Similarly, the terminal amide in the *para*- position of the biphenyl was considered in both **A** and **B** conformations. Hydrogen atoms were added to the ligands with the UCSF Chimera program.⁴⁶ All the added atoms were relaxed by simulated annealing from 600 to 0 K. The cooling runs were 2 ps long using the Berendsen thermostat, 1-fs time step, and the generalized Born solvent model (Bondi radii and igb=7 sander option).⁶⁰ Such a preparation protocol has been reported to be optimal for QM scoring.³⁰

QM/SQM optimization. We used an improved multi-layer setup⁸ by adopting the QM/SQM/COSMO methodology. The rationale of this approach was that COSMO had been shown to be more reliable than GB for the solvation of neutral ligands.⁶¹ The procedure is identical to that reported in Ref.⁸ with the following exceptions: we considered only residues within 10 Å of the inhibitor for optimization while all the residues farther than 8 Å from the inhibitor were frozen during the optimization. All the complexes were optimized by the QM/SQM/COSMO method and the FIRE optimization algorithm prior to scoring.

QM scoring. For QM scoring (single-point calculations), the whole CDK2/inhibitor complexes were used. The QM part comprised the inhibitors, all explicit water molecules and residues Ala33, Val64, Phe80, Asp127, Lys129, Pro130, Gln131 and Asn132. The small QM region was treated at the DFT-D3 (the TPSS functional with the TZVPP basis set) level of theory using the Turbomole 6.5 program.⁶² The rest of the system (SQM region) was calculated by PM6-D3H4X³ using MOPAC2009.⁶³

The binding free energy was approximated by the total score expressed by Eq. 1.² with the new consistent notation described in Ref.³:

$$\text{Score} = \Delta E_{\text{int}} + \Delta \Delta G_{\text{solv}} + \Delta G'_{\text{conf}}{}^{\text{w}}(\text{L}) - T\Delta S_{\text{solv}} \quad [\text{Eq. 1}].$$

The individual terms describe the gas-phase interaction energy (ΔE_{int}), the interaction solvation/desolvation free energy ($\Delta \Delta G_{\text{solv}}$), the change of the conformational “free” energy of the ligand ($\Delta G'_{\text{conf}}{}^{\text{w}}(\text{L})$), and optionally the entropy of the explicit water molecules ($T\Delta S_{\text{solv}}$). It should be mentioned, however, that another portion of the entropy contribution is already included in the $\Delta \Delta G_{\text{solv}}$ term via the parameterization of the implicit solvent models. ΔE_{int} was calculated using the QM/SQM method described above. The solvation free energy was determined using two implicit solvent models: COSMO³⁶ at the PM6 level and SMD⁶⁴ at the HF/6-31G** level (the latter only for the ligands to increase the accuracy).⁶¹ For the evaluation of the change of the conformational “free” energy of the ligand ($\Delta G'_{\text{conf}}{}^{\text{w}}(\text{L})$), the gas-phase DFT-D3 (TPSS/TZVPP single-point energy at the B-LYP/SVP optimized geometry) energy is combined with the SMD solvation free energy.

QM fragmentation. The interaction energies between the inhibitor and protein fragments (side chains and peptide bonds) were determined at the DFT-D3 (TPSS/TZVPP)⁶² level combined with the COSMO implicit solvent model.³⁶

ASSOCIATED CONTENT

The X-ray data collection and the refinement statistics, analyses of MD trajectories, water site location and thermodynamics from WaterMap, energy terms of the QM score and NMR, IR and MS spectra of all the compounds are presented in the Supporting Information.

AUTHOR INFORMATION

Author Contributions

‡ these authors contributed equally to this work

Corresponding Authors

*E-mail: paruch@chemi.muni.cz

*E-mail: lepsik@uochb.cas.cz

ACKNOWLEDGEMENT

We are grateful to Dr. F. Hoh for diffraction data collection. We thank Dr. Michaela Nekardová for helpful discussions. This work has been supported by the project “Employment of Best Young Scientists for International Cooperation Empowerment” (CZ.1.07/2.3.00/30.0037) co-financed from the European Social Fund and the state budget of the Czech Republic, and by the European Regional Development Fund under grant FNUSA-ICRC (No. CZ.1.05/1.1.00/02.0123). This work has been supported by the Ministry of Education, Youth and Sports of the Czech Republic (project No. L01305) and also from the Large Infrastructures for Research, Experimental Development and Innovations project “IT4Innovations National Supercomputing Center – LM2015070”. RJ and VK thank the Czech Science Foundation (15-15264S). JF, PH and ML thank the Gilead Sciences & IOCB Research Centre. JF, SH, CK, HA, PH and ML thank the research project RVO 61388963, awarded by the Academy of Sciences of the Czech Republic and the Czech Science Foundation (grant number P208/12/G016). KP thanks the project CZ-OPENSOURCE: National Infrastructure for Chemical Biology (LM2015063).

REFERENCES

- (1) K. Raha, K.M. Merz A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding, *J. Am. Chem. Soc.* 126 (2004), 1020–1021.
- (2) J. Fanfrlík, A.K. Bronowska, J. Rezac, O. Prenosil, J. Konvalinka, P. Hobza, A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands, *J. Phys. Chem. B.* 114 (2010), 12666–12678.
- (3) M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, J. Fanfrlík, The semiempirical quantum mechanical scoring function for in silico drug design, *ChemPlusChem.* 78 (2013), 921–931.
- (4) A. Pecina, M. Lepšík, J. Řezáč, J. Brynda, P. Mader, P. Řezáčová, P. Hobza J. Fanfrlík, QM/MM calculations reveal the different nature of the interaction of two carborane-based sulfamide inhibitors of human carbonic anhydrase II, *J. Phys. Chem. B* 117 (2013), 16096–16104.
- (5) A. Ciancetta, S. Genheden, U. Ryde, A QM/MM study of the binding of RAPTA ligands to cathepsin B, *J. Comput. Aided. Mol. Des.* 25 (2011), 729–742.
- (6) J. Fanfrlík, J. Brynda, J. Rezac, P. Hobza, M. Lepsík, Interpretation of protein/ligand crystal structure using QM/MM calculations: case of HIV-1 protease/metallacarborane complex, *J. Phys. Chem. B* 112 (2008), 15094–15102.
- (7) J. Fanfrlík, M. Kolář, M. Kamlar, D. Hurný, F.X. Ruiz, A. Cousido-Siah, M. Mitschler, J. Rezac, E. Munusamy, M. Lepšík, P. Matějčíček, J. Veselý, A. Podjarny, P. Hobza, Modulation of aldose reductase inhibition by halogen bond tuning, *ACS Chem. Biol.* 8 (2013), 2484–2492.
- (8) J. Fanfrlík, F.X. Ruiz, A. Kadlčíková, J. Řezáč, A. Cousido-Siah, A. Mitschler, S. Haldar, M. Lepšík, M.H. Kolář, P. Majer, A.D. Podjarny, P. Hobza, The Effect of Halogen-to-Hydrogen Bond Substitution on Human Aldose Reductase Inhibition, *ACS Chem. Biol.* 10 (2015), 1637–1642.

- (9) J. Fanfrlík, P.S. Brahmshatriya, J. Řezáč, A. Jílková, M. Horn, M. Mareš, P. Hobza, M. Lepšík, Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic *Schistosoma mansoni* cysteine peptidase by vinyl sulfone inhibitors, *J. Phys. Chem. B* 117 (2013), 14973–14982.
- (10) M. Kolář, J. Fanfrlík, M. Lepšík, F. Forti, J.F. Luque, P. Hobza, Assessing the Accuracy and Performance of Implicit Solvent Models for Drug Molecules: Conformational Ensemble Approaches, *J. Phys. Chem. B* 117 (2013), 5950-5962.
- (11) P. Soderhjem, J. Kongsted, U. Ryde, Ligand Affinities Estimated by Quantum Chemical Calculations, *J. Chem. Theory Comput.* 6 (2010), 1726–1737.
- (12) S. Decherchi, M. Masetti, I. Vyalov, W. Rocchia, Implicit solvent methods for free energy estimation, *Eur. J. Med. Chem.* 91 (2015), 27–42.
- (13) A. Pecina, R. Meier, J. Fanfrlík, M. Lepšík, J. Řezáč, P. Hobza, C. Baldauf, The SQM/COSMO filter: reliable native pose identification based on the quantum-mechanical description of protein–ligand interactions and implicit COSMO solvation, *Chem. Commun.* 52 (2016), 3312–3315.
- (14) S. Genheden, P. Mikulskis, L. Hu, J. Kongsted, P. Söderhjelm, U. Ryde, Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration, *J. Am. Chem. Soc.* 133 (2011), 13081–13092.
- (15) P. Setny, R. Baron, P.M. Kekenus-Huskey, A. McCammon, J. Dzubiella, Solvent fluctuations in hydrophobic cavity–ligand binding kinetics, *Proc. Natl. Acad. Sci. USA* 110 (2013), 1197–1202.
- (16) B. Vorlová, D. Nachtigallová, J. Jirásková-Vaníčková, H. Ajani, P. Jansa, J. Řezáč, J. Fanfrlík, M. Otyepka, P. Hobza, J. Konvalinka, M. Lepšík, Malonate-based inhibitors of mammalian serine racemase: kinetic characterization and structure-based computational study. *Eur. J. Med. Chem.* 89 (2015), 189–197.

- (17) B. Breiten, M. R. Lockett, W. Sherman, S. Fujita, M. Al-Sayah, H. Lange, C.M. Bowers, A. Heroux, G. Krilov, G.M. Whitesides, Water networks contribute to enthalpy/entropy compensation in protein–ligand binding, *J. Am. Chem. Soc.* 135 (2013), 15579–15584.
- (18) D. Robinson, T. Bertrand, J.C. Carry, F. Halley, A. Karlsson, M. Mathieu, H. Minoux M.A. Perrin, B. Robert, L. Schio W. Sherman, Differential Water Thermodynamics Determine PI3K-Beta/Delta Selectivity for Solvent-Exposed Ligand Modifications, *J. Chem. Inf. Model.* 56 (2016), 886–894.
- (19) A. Cousido-Siah, F.X. Ruiz, J. Fanfrlik, J. Giménez-Dejoz, A. Mitschler, M. Kamlar, J. Vesely, H. Ajani, X. Pares, J. Farrés, P. Hobza, A.D. Podjarny, IDD388 polyhalogenated derivatives as probes for an improved structure-based selectivity of AKR1B10 inhibitors, *ACS Chem. Biol.* 11 (2016), 2693–2705.
- (20) A. Biela, M. Betz, H. Heine, G. Klebe, Water makes the difference: rearrangement of water solvation layer triggers non-additivity of functional group contributions in protein–ligand binding, *Chem. Med. Chem.* 7 (2012), 1423–1434.
- (21) M. Malumbres, Cyclin-dependent kinases. *Genome Biol.* 15 (2014), 122.
- (22) U. Asghar, A.K. Witkiewicz, N.C. Turner, E.S. Knudsen, The history and future of targeting cyclin-dependent kinases in cancer therapy, *Nat. Rev. Drug Discov.* 14 (2015), 130–46.
- (23) J.A. Beaver, L. Amiri-Kordestani, R., Charlab, W. Chen, T. Palmby, A. Tilley, J.F. Zirkelbach, J. Yu, Q. Liu, L. Zhao, J. Crich, X.H. Chen, M. Hughes, E. Bloomquist, S. Tang, R. Sridhara, P.G. Kluetz, G. Kim, A. Ibrahim, R. Pazdur, P.C. Cortazar, FDA Approval: Palbociclib for the Treatment of Postmenopausal Patients with Estrogen Receptor-Positive, HER2-Negative Metastatic Breast Cancer, *Cancer Res.* 21 (2015), 4760–4766.
- (24) A. Huwe, R. Mazitschek, A. Giannis, Small molecules as inhibitors of cyclin-dependent kinases. *Angew. Chem. Int. Ed. Engl.* 42 (2003), 2122–2138.

- (25) G. Kontopidis, C. McInnes, S.R. Pandalaneni, I. McNae, D. Gibson, M. Mezna, M. Thomas, G. Wood, S. Wang, M.D. Walkinshaw, P.M. Fischer, Differential binding of inhibitors to active and inactive CDK2 provides insights for drug design, *Chem Biol.* 13 (2006), 201–211.
- (26) K. Paruch, M.P. Dwyer, C. Alvarez, C. Brown, T.Y. Chan, R.J. Doll, K. Keertikar, C. Knutson, B. McKittrick, J. Rivera, R. Rossman, G. Tucker, T. Fischmann, A., Hruza, V. Madison, A.A. Nomeir, Y. Wang, P. Kirschmeier, E. Lees, D. Parry, N. Sgambellone, W. Seghezzi, L. Schultz, F. Shanahan, D. Wiswell, X. Xu, Q. Zhou, R.A. James, V.M. Paradkar, H. Park, L.R. Rokosz, T.M. Stauffer, T.J. Guzi, Discovery of Dinaciclib (SCH 727965): A Potent and Selective Inhibitor of Cyclin-Dependent Kinases, *ACS Med. Chem. Lett.* 1 (2010), 204–208.
- (27) P.M. Dwyer, K. Paruch, C. Alvarez, R.J. Doll, K. Keertikar, J. Duca, T.O. Fischmann, A. Hruza, V. Madison, E. Lees, D. Parry, W. Seghezzi, N. Sgambellone, F. Shanahan, D. Wiswell, T.J. Guzi, Versatile templates for the development of novel kinase inhibitors: Discovery of novel CDK inhibitors, *Bioorg. Med. Chem. Lett.* 17 (2007), 6216–6219.
- (28) K. Paruch, P.M. Dwyer, C. Alvarez, C. Brown, T.Y. Chan, R.J. Doll, K. Keertikar, C. Knutson, B. McKittrick, J. Rivera, R. Rossman, G. Tucker, T.O. Fischmann, A. Hruza, V. Madison, A.A., Nomeir, Y. Wang, E. Lees, D. Parry, N. Sgambellone, W. Seghezzi, L. Schultz, F. Shanahan, D. Wiswell, X. Xu, Q. Zhou, R.A. James, V.M. Paradkar, H. Park, L.R. Rokosz, T.M. Stauffer, T.J. Guzi, Pyrazolo[1,5-a]pyrimidines as orally available inhibitors of cyclin-dependent kinase 2. *Bioorg. Med. Chem. Lett.* 17 (2007), 6220–6223.
- (29) R. Jorda, K. Paruch, V. Krystof, Cyclin-dependent kinase inhibitors inspired by roscovitine: purine bioisosteres, *Curr Pharm Des.* 18 (2012), 2974–2980.
- (30) P.S. Brahmshatriya, P. Dobeš, J. Fanfrlik, J. Rezáč, K. Paruch, A. Bronowska, M. Lepšík, P. Hobza, Quantum mechanical scoring: structural and energetic insights into cyclin-dependent

kinase 2 inhibition by pyrazolo[1,5-a]pyrimidines, *Curr. Comput. Aided Drug Des.* 9 (2013), 118–129.

(31) T. Gucký, R. Jorda, M. Zatloukal, V. Bazgier, K. Berka, E. Řezníčková, T. Béres, M. Strnad, V. Kryštof, A novel series of highly potent 2,6,9-trisubstituted purine cyclin-dependent kinase inhibitors. *J. Med. Chem.* 56 (2013), 6234–6247.

(32) R. Jorda, L. Havlíček, I.W. McNae, M.D. Walkinshaw, J. Voller, A. Sturc, J. Navrátilová, M., Kuzma, M. Mistrík, J. Bártek, M. Strnad, V. Krystof, Pyrazolo[4,3-d]pyrimidine bioisostere of roscovitine: evaluation of a novel selective inhibitor of cyclin-dependent kinases with antiproliferative activity, *J. Med. Chem.* 54 (2011), 2980–2993.

(33) P. Dobeš, J. Fanfrlík, J. Rezáč, M. Otyepka, P. Hobza, Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors, *J. Comput. Aided Mol. Des.* 25 (2011), 223–235.

(34) P. Dobes, J. Rezáč, J. Fanfrlík, M. Otyepka, P. Hobza, Semiempirical quantum mechanical method PM6-DH2X describes the geometry and energetics of CK2-inhibitor complexes involving halogen bonds well, while the empirical potential fails, *J Phys Chem B.* 115 (2011), 8581–8589.

(35) T.J. Guzi, K. Paruch, M.P. Dwyer, M. Labroli, F. Shanahan, N. Davis, L. Taricani, D. Wiswell, W. Seghezzi, E. Penaflo, B. Bhagwat, W. Wang, D. Gu, Y. Hsieh, S. Lee, M. Liu, D. Parry, Targeting the replication checkpoint using SCH 900776, a potent and functionally selective CHK1 inhibitor identified via high content screening, *Mol. Cancer Ther.* 10 (2011), 591–602.

(36) A. Klamt, G. Schüürmann. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perkin Trans. 2* (1993), 799–805.

- (37) D.A. Pearlman, P.S. Charifson, Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system, *J. Med. Chem.* 44 (2001), 3417–3423.
- (38) N.R. Brown, M.E. Noble, J.A. Endicott, L.N. Johnson, The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases, *Nat. Cell Biol.* 1 (1999), 438–443.
- (39) T.G. Davies, J. Bentley, C.E. Arris, F.T. Boyle, N.J. Curtin, J.A. Endicott, A.E. Gibson, B.T. Golding, R.J. Griffin, I.R. Hardcastle, Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor, *Nat. Struct. Mol. Biol.* 9 (2002), 745–749.
- (40) Collaborative Computational Project, Number 4, The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* 50 (1994), 760–763.
- (41) A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, L. C., R.J.X. Read, Phaser crystallographic software, *J. Appl. Cryst.* 40 (2007), 658–674.
- (42) K. Bettayeb, N. Oumata, A. Echalié, Y. Ferandin, J.A. Endicott, H. Galons, L. Meijer, CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases, *Oncogene* 27 (2008), 5797–807.
- (43) P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, *Acta Cryst. D* 66 (2010), 486–501.
- (44) G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Cryst. D* 53 (1997), 240–255.
- (45) A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, *Protein Science* 9 (2000), 1753–1773.
- (46) E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25 (2004), 1605–1612.

- (47) D.A. Case, R.M. Betz, W. Botello-Smith, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H., Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, P.A. Kollman, AMBER 2014 (2014), University of California, San Francisco.
- (48) Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem.* 24 (2003), 1999–2012.
- (49) J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and Testing of a General Amber Force Field, *J. Comput. Chem.* 25 (2004), 1157–1174.
- (50) C.I. Bayly, P. Cieplak, W. Cornell P.A. Kollman, A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Determining Atom-Centered Charges: The RESP Model, *J. Phys. Chem.* 97 (1993), 10269–10280.
- (51) M. Lepsík, Z. Kríz, Z. Havlas, Efficiency of a second-generation HIV-1 protease inhibitor studied by molecular dynamics and absolute binding free energy calculations, *Proteins* 57 (2004), 279–293.
- (52) M. Kolar, P. Hobza, On Extension of the Current Biomolecular Empirical Force Field for the Description of Halogen Bonds, *J. Chem. Theory Comput.* 8 (2012), 1325–1333.
- (53) D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen H.J.C. Berendsen, *Gromacs User Manual*, Version 4.5.4 (2010), www.gromacs.org.
- (54) W. Humphrey, A. Dalke, K. Schulten, VMD – Visual Molecular Dynamics, *J. Mol. Graph.*, 14 (1996), 33–38.

- (55) J. Cohen, A. Arkhipov, R. Braun K. Schulten, Imaging the migration pathways for O₂, CO, NO, and Xe inside myoglobin. *Biophysical Journal* 91 (2006), 1844–1857.
- (56) Schrödinger Release 2014-4: Maestro, Version 10.0, Schrödinger, LLC, New York, NY, 2014.
- (57) Schrödinger Release 2014-4: WaterMap, Version 2.1, Schrödinger, LLC, New York, NY, 2014.
- (58) R. Abel, T. Young, R. Farid B.J. Berne R.A. Friesner, Role of the active-site solvent in the thermodynamics of factor Xa ligand binding, *J. Am. Chem. Soc.* 130 (2008), 2817–2831.
- (59) The PyMOL Molecular Graphics System, Version 1.7.6.3., Schrödinger, LLC.
- (60) J. Mongan, C. Simmerling, A.J. McCammon, D. Case, A. Onufriev, Generalized Born with a simple, robust molecular volume correction, *J. Chem. Theory Comput.* 3 (2007), 156–169.
- (61) M. Kolář, J. Fanfrlík, M. Lepšík, F. Forti, F.J. Luque, P. Hobza, Assessing the accuracy and performance of implicit solvent models for drug molecules: conformational ensemble approaches, *J Phys Chem B.* 117 (2013), 5950–5962.
- (62) R. Ahlrichs, M. Bar, M. Haser, H. Horn, C. Kolmel, Electronic Structure Calculations on Workstation Computers: The Program System Turbomole, *Chem. Phys. Lett.* 162 (1989), 165–169.
- (63) J.J.P. Stewart, MOPAC2012 (2012), Stewart Computational Chemistry, Colorado Springs, CO; <http://OpenMOPAC.net>.
- (64) A.V. Marenich, C.J. Cramer, D.G. Truhlar, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions, *J. Phys. Chem. B* 113 (2009), 6378–6396.

FIGURE CAPTIONS

Figure 1. Schematical representation of interaction of pyrazolo[1,5-*a*]pyrimidines with the CDK2 active site with indicated key amino acids. Three hinge region hydrogen bonds are dashed.

Scheme 1. Synthesis of intermediates **6** and **7** leading to targets **1a** to **1q**.

Reagents and conditions: **i.** LDA, THF, -78 °C, then 4-iodobenzoyl chloride or 4-bromobenzoyl chloride, THF, -78 °C to r.t., (**2a**: 81%), (**2b**: 92%); **ii.** (a) 3-aminopyrazole or (b) 3-amino-4-ethylpyrazole, AcOH, reflux, (**3a**: 61%), (**3b**: 71%), (**3c**: 65%); **iii.** POCl₃, pyr., r.t., (**4a**: 85%), (**4b**: 42%), (**4c**: 95%); **iv.** 3-picolylamine, MeCN, reflux, (**5a**: 93%), (**5b**: 79%), (**5c**: 97%); **v.** Boc₂O, DMAP, CH₂Cl₂, r.t., (**6a**: 92%), (**6b**: 90%), (**6c**: 94%); **vi.** B₂Pin₂, K₃PO₄, PdCl₂dppf, DME, H₂O, reflux, (**7a**: 72% from **6a**, 85% from **6c**), (**7b**: crude used as such in next step); **vii.** K₃PO₄, PdCl₂dppf, DME, H₂O, reflux; **viii.** AcCl, CH₂Cl₂, NEt₃, 0 °C to r.t., (**8d**: 82%); **ix.** NaBH₄, MeOH, 0 °C to r.t., (**8i**: 88%), (**8l**: 88%); **x.** NBS, MeCN, r.t.; **xi.** 3M aq. HCl, EtOH, 60 °C; **xii.** TFA, CH₂Cl₂, r.t.

Scheme 2. Synthesis of compounds **1p**, **1r**, **1s**, **1t** and **1u**.

Reagents and conditions: **i.** SOCl₂, MeOH, reflux, 95%; **ii.** acrylonitrile, Pd(P(t-Bu)₃)₂, DIPEA, dioxane, 100 °C, 49%; **iii.** H₂, Pd(OAc)₂, PPh₃, toluene, 80 °C, 70%; **iv.** SnCl₂, EtOAc, H₂O, reflux, 55%; **v.** isoamylnitrite, CuBr₂, CH₃CN, 60 °C, 77%; **vi.** NaOH, THF, r.t., quant.; **vii.** SOCl₂, DMF, THF, r.t., then NH₃ in dioxane, THF, 0 °C, 72%; **viii.** Compound **7a**, PdCl₂dppf, K₃PO₄, DME, H₂O, reflux, (**20a**: 77%); **ix.** SOCl₂, THF, reflux then NH₃ in MeOH, THF, 0 °C, 66%; **x.** PivCl, NEt₃, CH₂Cl₂, THF, 75%; **xi.** Compound **7a**, PdCl₂dppf, K₃PO₄, DME, H₂O, reflux, (**20b**: 40%); **xii.** Na₂CO₃, SOCl₂, NH₃ in MeOH, 80 °C, 68%; **xiii.** BCl₃, CH₂Cl₂, r.t., 90%; **xiv.** Compound **7a**, K₃PO₄, PdCl₂dppf, DME, H₂O, reflux, (**20c**: 80%); **xv.** MsCl, NEt₃, DMF, 0 °C, (**20d**: 87%); **xvi.** NBS, CH₂Cl₂, 0 °C, (**21a**: 91%), (**21b**: 86%), (**21c**: 83 %); **xvii.** NaOH, THF, reflux, (22: 90%); **xviii.** K₂CO₃, MeOH, r.t., (**23**: 86%); **xix.** HCl, EtOH, 60 °C, (**1r**: 85%), (**1s**: quant.), (**1u**: 90%) ; **xx.** HCl, EtOH, 60 °C, (**1p**: 90%), (**1t**: 78%).

Figure 2. Inhibitor binding mode, atom numbering and structural water molecules in the crystallographic complex of CDK2/**1l**. Three hinge-region hydrogen bonds between the protein and the inhibitor are shown as dotted lines. Color coding of sticks: grey – carbon, blue – nitrogen, red – oxygen, brown – bromine. Spheres represent water sites as obtained from

WaterMap. Their coloring ranges from green (the smallest value of $\Delta\Delta G$, Tab. 2) via brown to red (the largest value of $\Delta\Delta G$, Tab. 2).

Figure 3. Calculated QM/SQM/COSMO scores plotted against experimental $RT \cdot \ln(IC_{50}/2)$, all in kcal/mol, A) pure implicit solvent model, B) mixed explicit/implicit solvent model, i.e. six explicit water molecules added on top of the implicit solvent treatment.

Figure 4. Interactions of the R^5 substituents of the strongest inhibitor **1s** with CDK2 and two active-site waters. The geometry was obtained by modeling and QM/SQM optimization. Color coding as in Fig. 2.