J Comput Aided Mol Des (2011) 25:223–235
DOI 10.1007/s10822-011-9413-5

Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors

Petr Dobeš · Jindřich Fanfrlík · Jan Řezáč · Michal Otyepka · Pavel Hobza

Received: 12 September 2010 / Accepted: 18 January 2011 / Published online: 1 February 2011

Abstract A semiempirical quantum mechanical PM6-DH2 method accurately covering the dispersion interaction and H-bonding was used to score fifteen structurally diverse CDK2 inhibitors. The geometries of all the complexes were taken from the X-ray structures and were reoptimised by the PM6-DH2 method in continuum water. The total scoring function was constructed as an estimate of the binding free energy, i.e., as a sum of the interaction enthalpy, interaction entropy and the corrections for the inhibitor desolvation and deformation energies. The applied scoring function contains a clear thermodynamical terms and does not involve any adjustable empirical parameter. The best correlations with the experimental inhibition constants (ln K_i) were found for bare interaction enthalpy ($r^2 = 0.87$) and interaction enthalpy corrected for ligand desolvation and deformation energies ($r^2 = 0.77$); when the entropic term was considered, however, the correlation becomes worse but still acceptable ($r^2 = 0.52$). The resulting correlation based on the PM6-DH2 scoring function is better than previously published function based on various docking/scoring, SAR studies or advanced QM/MM approach, however, the robustness is limited by number of available experimental data used in the correlation. Since a very similar correlation between the experimental and theoretical results was found also for a different system of the HIV-1 protease, the suggested scoring function based on the PM6-DH2 method seems to be applicable in drug design, even if diverse protein–ligand complexes have to be ranked.

Keywords CDK2 · Semiempirical quantum mechanical method · PM6-DH2 · Non-covalent interaction · Scoring function · Drug design

Abbreviations

CDK2	Cyclin-dependent kinase 2
MM	Molecular mechanics
QM	Quantum mechanics
SQM	Semiempirical quantum mechanics
PI	Protein (P)–inhibitor (I) complex

Introduction

Cyclin-dependent kinase 2 (CDK2) is one of the prominent cell cycle regulators [1, 2], which is dominantly active during the G1 phase and G1/S transition. The deregulation of cyclin-dependent kinases (CDKs) is known to be associated with many serious diseases, such as cancer [3, 4].

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9413-5) contains supplementary material, which is available to authorized users.

M. Otyepka (✉) · P. Hobza (✉)
Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University, 17 listopadu 12, 771 46 Olomouc, Czech Republic
e-mail: michal.otyepka@upol.cz

P. Hobza
e-mail: pavel.hobza@uochb.cas.cz

P. Dobeš · J. Fanfrlík · J. Řezáč · P. Hobza
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

P. Dobeš
Center of Molecular Biology and Gene Therapy, Department of Internal Medicine–Hematooncology, University Hospital Brno, 625 00 Brno, Czech Republic

This fact has attracted attention in the long term to the development of efficient inhibitors of CDKs [5–8]. A relentless effort in this field has succeeded in bringing some CDK inhibitors to clinical trials [9, 10]. Even though the targeting of CDK2 does not have to be an optimal strategy for cancer treatment because of the redundancy of CDKs in cell cycle regulation [11, 12], the CDK2 has remained a paradigm for rational drug design, because it is the best characterised cyclin-dependent kinase in terms of structure and biochemistry [13].

Competition between an inhibitor and the native ATP substrate is an effective strategy to inhibit CDK2. An inhibitor binds to a deep cleft between two CDK2 lobes [14] (Fig. 1a) and, despite the numerous structurally-varied CDK2 inhibitors known today, some common features can be identified. The inhibitors share a flat central ring and form H-bonds to the CDK2 hinge region (Phe80-Leu83 residues), which forms the back wall of the CDK2 active site. The discovery of the CDK2–inhibitor structure [14, 15] provided a useful starting point for the rational design of CDK2 inhibitors.

Since then, many studies exploiting various rational drug-design strategies have been carried out [16–19], among which a prominent place is taken by docking and scoring studies. The main aim of the molecular docking experiment [20, 21] is to find an optimal orientation of the ligand (e.g. inhibitor) in the active site of a target (e.g. an enzyme, in our case CDK2) and to evaluate the ligand activity (e.g. inhibition constant). Empirical scoring functions based on the empirical potentials are often used to evaluate the inhibition activity [22]. Despite the fine tuning and parameterisation of the empirical scoring functions, they are in principle restricted by the known limitations of the empirical potential (also known as molecular mechanics, MM, or force field), which mainly concerns its inability to incorporate quantum effects such as charge transfer between protein and ligand and the wide variation in atomic charges between different structures of the ligand and protein [23, 24]. Another important consequence of quantum effects is the existence of halogen bond that play an important role in the binding of halogenated ligands [25]. This limitation can be overcome e.g. by employing the nonempirical quantum mechanical (QM) ab initio calculation, but this approach is seriously limited by the computer demands since protein–inhibitor complexes possess several thousand atoms. The situation is further complicated since electron correlation (or London dispersion energy) frequently plays an important role in the protein–inhibitor binding, which makes the use of less demanding methods like Hartree–Fock or density functional theory (DFT) impractical [26]. It should be added here that the present case, i.e. inhibitor binding to CDK2, is known to be

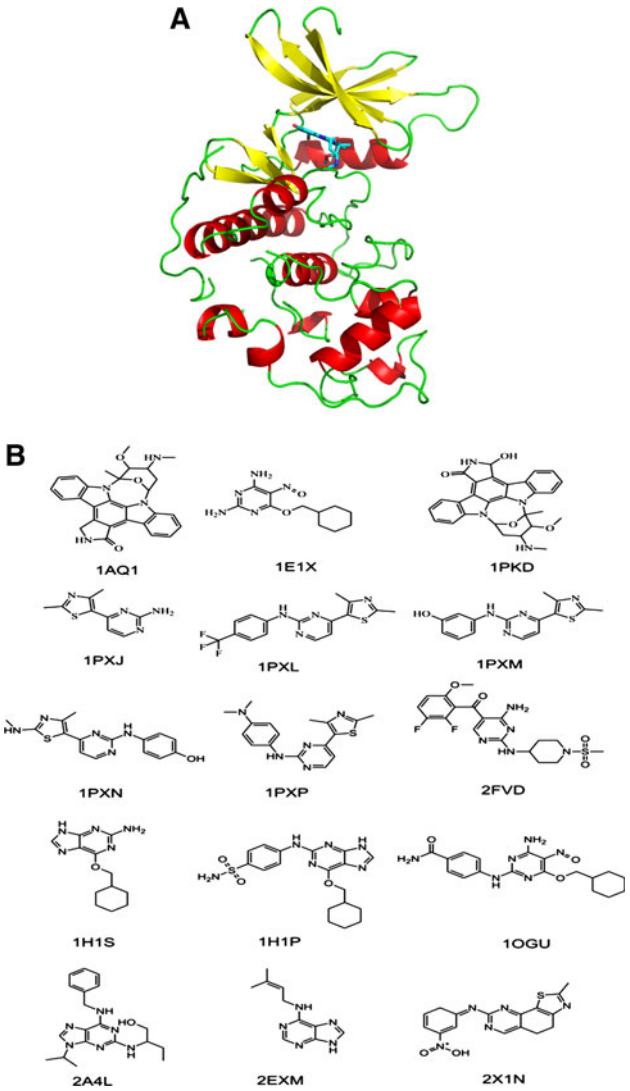

Fig. 1 **a** The ATP competitive inhibitor (here UCN-1 in sticks) binds to a deep cleft in the CDK2 structure (PDB ID code: 1PKD), which is shown in a cartoon model (the red curls represent α-helices and the yellow stripes β-strands, the N-terminal lobe is up and C-terminal lobes is down. **b** The structures of the fifteen inhibitors considered in this study

controlled by the electron correlation (dispersion) energy contributions [27].

The neglect of the quantum effects prevents the use of the computationally accessible MM methods while the size of complexes investigated prevents the use of the highly accurate nonempirical QM methods. There are two possibilities of how to solve the problem. The first is based on the use of the QM/MM approach while the other relies on the use of semiempirical QM (SQM) methods. Both of these cover quantum effects and have already been implemented in rational drug discovery [28–30]. The former enables the study of the protein active site-binding pocket at the QM level, with the rest of the protein–inhibitor complex (including water) being treated at the

MM level. The latter approach describes the whole complex directly using SQM methods. Both approaches obviously have their *pros* and *cons* [31–35], but we believe that the description of the whole complex consistently without any artificial division into QM and MM parts is better suited for drug-design purposes. The use of the SQM method for such large systems is certainly conditioned by the introduction of linear scaling techniques.

The use of the SQM technique in drug design was pioneered by Merz et al. (Refs. [28, 36] and the references therein), who used standard semiempirical AM1 or PM3 Hamiltonians. These methods are, however, known not to cover the London dispersion energy. Wollacott and Merz [36] solved the problem by constructing the binding free energy of a protein–inhibitor complex as a sum of the AM1 or PM3 heat of formation, solvation free energy, and the attractive term from the Lennard-Jones potential. There is a more straightforward way to solve the problem, namely by adding the missing dispersion term to a SQM method (see our recent review [26] for references).

The present SQM calculations are based on Stewart's PM6 semiempirical QM method [37]. Its implementation in the MOPAC package offered a linear scaling algorithm MOZYME [38]. MOZYME is a localized molecular orbital method which replaces the standard SCF procedure in the PM6 calculation. This method speeds us the SQM calculations significantly and unable us to consider systems with several thousands of atoms. For this SQM method, we have developed a dispersion correction as well as a correction for hydrogen bonds [39, 40]. The resulting method (using the second generation of corrections), named PM6-DH2, yields, to the best of our knowledge, the most accurate results for non-covalent interactions of all the SQM methods [26]. For small model non-covalent complexes, which allow for a comparison with the benchmark high-level QM calculations, the method reaches a chemical accuracy of $\sim$ 1 kcal/mol and the quality of the results is comparable to the much more computationally expensive wave-function theory (WFT) and DFT techniques [26].

The main limitation of present approach based on the use of SQM PM6-DH2 method is the computer time. Although the method is due to the use of the linear scaling procedure very efficient (see the end of "Discussion") it is still considerably slower than empirical potentials. In the present as well as in the previous paper we describe the PI complex fully by SQM method. In the case of larger proteins such calculations will be, however, time consuming and some acceleration procedure should be adopted. There are basically two possibilities, either the use parallelization (or even massive parallelization) or the use of QM/MM technique. Our approach will be, however, different from other QM/MM approaches in the literature. We believe that the QM part should be extended (several thousands of atoms) and, therefore, the SQM will be systematically used. This strategy was supported by Merz et al. [41] who showed that their SQM/MM method did not reproduce PL binding affinity as well as their full SQM method. However, the SQM/MM results were quite encourraging and were qualitatively competitive with full SQM results. These results show that SQM/MM approache can be considered for bigger complexes that would be too time demanding for the full SQM calculations.

The procedure described, which is based on the use PM6-DH2 scoring function, was introduced in our previous paper [42], where we studied the complexation of the HIV-1 protease with twenty-two ligands (eleven binders and eleven non-binders). The total score was constructed as an estimate of the binding free energy, i.e., as a sum of the interaction enthalpy, interaction entropy, ligand deformation energy and ligand desolvation free energy. The *in silico* predictions were finally cross-validated by the experimental data and very good agreement between the predictions and experimental data was achieved ($r^2 = 0.71$). All of the binders were either highly polar or charged, which made the desolvation free energy very large, sometimes comparable with the interaction enthalpy. In agreement with experiment, the interaction entropy plays an important role in the HIV-1 system. It was concluded that all the terms in the total score are significant and none of them can be neglected.

In the present study, we have used the same technique for another medicinally important target, CDK2. The PM6-DH2 method was used to score fifteen complexes of CDK2 with inhibitors (Fig. 1b) for which the experimental inhibition constants are known. Contrary to the previous study, most of the inhibitors are neutral, more rigid and less polar than in the case of the HIV-1 protease ligands. This point is important since it demonstrates the wide applicability of the scoring function suggested. Without any modification of the theoretical treatment for different target (CDK2 vs. HIV-1 protease in our previous paper) and different inhibitors (not only by their chemical constitution but also by their different charge, polarity and flexibility) we obtained comparable results. The theoretical predictions are again in a good agreement with the experimental inhibition constants.

In summary, the scoring function constructed as an estimate of the total binding free energy is divided into multiple separate components having clear physical interpretation. Each component is solved by the best performing method considering efficiency and today computer power. Since different procedures (based on different approximations) are adopted for each component the resulting scoring function cannot be directly equal to the change of the binding free energy. As all the components are physically clear terms the limitations stemming e.g. from approximations used in evaluation of each term are well known

which implies that the produres how to improve each term are in principle known. It is important to stress again that the assessed scoring function does not involve any adjustable empirical factor. As the scoring function does not involve any adjustable parameter it can be directly applied to a different target/ligand system without any modification and parameterization, which increases robustness and applicability of the scoring function. The promising results from both systems (HIV-1 and CDK2) with different natures of inhibitor binding document the transferability of the scoring function based on a robust and reliable PM6-DH2 method. However, more robust testing on various targets would be neccessary to find potential limitations of the scoring function.

Computational methods

The PM6-DH2 method provides accurate results for various types of noncovalent complexes, since it includes corrections to the dispersion and H-bonding energies. The dispersion correction is based on our previous work on the DFT-D methods [43] but has been adapted to the PM6 method, because it already contained a part of the dispersion term. The whole dispersion correction was scaled, and the overestimated dispersion contribution of the sp^3 hybridised carbon atoms was corrected. The H-bonding correction added a force field-like term to each pair of atoms which could possibly form an H-bond (as the proton donor and proton acceptor only oxygen and nitrogen were considered). This correction was directional—depending not only on the distance but also on the H-bond angle and other geometric parameters, which substantially improved the geometry of the hydrogen bonds. These corrections are of vital importance for predicting the correct structure of a protein–inhibitor complex.

Theoretical description of the protein–inhibitor binding

The formation of the protein (P)–inhibitor (I) complex from free (hydrated) subsystems represents a crucial step in the virtual docking scheme.

$$\mathrm{P} + \mathrm{I} \rightarrow \mathrm{P} - \mathrm{I} \tag{1}$$

The binding of a competitive inhibitor is evaluated by an inhibition constant (K_i). An inhibitor binding affinity is usually represented also by a half maximal inhibitory concentration (IC_{50}). However, IC_{50} depends on the experimental conditions more than K_i. Therefore, the inhibition (K_i) constant is a more appropriate (but not perfect, because the results from different laboratories may differ because of different experimental conditions) and thermodynamically well-defined term for measuring the inhibitor binding affinity. The inhibition constant is related to the change of the free energy of binding (binding free energy) by the following equation:

$$\Delta G_w = RT \ln(K_i). \tag{2}$$

The direct calculation of the inhibitor binding free energy using MD-based methods is nowadays possible [44–46]. However, these methods, although accurate within the limit of accuracy of empirical potential, are computationally demanding and time-consuming, which limits their application for certain stages of the drug development process, such as high-throughput screening. This quantity may be, however, determined undirectly using the thermodynamic cycle shown in our previous paper [42]. The ligand binding free energy is thus constructed as follows. First, the liquid-phase structure of the inhibitor is dehydrated and its structure is deformed to that the inhibitor possess in the PI complex. Similarly, the same is true about the protein, but the deformation term is presently omitted since we are considering the complexation of one protein with various inhibitors in this study (the desolvation energy of a protein is, however, properly taken into consideration). The deformation and often also the dehydration energies of inhibitor are positive, i.e. they oppose the binding. Second, the dehydrated and deformed inhibitor binds to the protein with an associated change of the interaction free energy. The latter term is determined as a sum of enthalpy and entropy terms. It would be desirable to evaluate all terms mentioned consistently at the same theoretical level (which is also our final goal). Unfortunately, this is unfeasible at present and several approximations have to be adopted in order to estimate the desired free energy of binding as accurately as possible (see later). Consequently, we have obtained only an estimate of the free energy of the inhibitor binding ΔG_w not involving some terms (e.g. the deformation energy of protein, protein desolvation etc.), which are expected to be constant for one target protein. This term is called the scoring function ($\Delta G_w'$) and it is calculated as follows (more detailed analysis is present in our previous paper [42]):

$$\Delta G_w' = \Delta H_w - T\Delta S_w + \Delta E_{def}(\mathrm{I}) + \Delta\Delta G_w(\mathrm{I}). \tag{3}$$

The individual terms represent the interaction enthalpy (ΔH_w), the interaction entropy (as $-T\Delta S_w$) and the corrections for inhibitor deformation energy $\Delta E_{def} = E(\mathrm{I})^{PI} - E(\mathrm{I})^{I}$ (with the upper indexes PI and I standing for the geometry of the inhibitor in a protein/inhibitor complex and the relaxed inhibitor geometry in water, respectively) and the corrections for the inhibitor hydration free energy $\Delta\Delta G_w(\mathrm{I}) = \Delta G_w^{MOPAC}(\mathrm{I})^{PI} - \Delta G_w^{G09}(\mathrm{I})^{PI}$ (where the lower index w stands for a water environment). The last term reflects the fact that structure

of the PI complex (in the water environment) is determined using the hydration model considering the electrostatic term only (MOPAC code) while accurate treatment requires the use of hydration model considering the nonelectrostatic term as well. Nonelectrostatic terms associated with formation of a cavity to accommodate the solute, and the van der Waals interaction between solute and solvent, are in the former approach neglected. The electrostatic approximation might work well for inhibitors of the similar shape but in the case of entirely different inhibitors it can cause a serious error. From this reason we adopted a more accurate calculation of the complete solvation free energy of the ligand with a SMD model based on IEFPCM calculation with radii and non-electrostatic terms by Truhlar and coworkers [47] as implemented in Gaussian 09 code.

The formation of a PI complex restricted motions of the ligand as well as the torsional motion of the protein. Furthermore, the vibrational entropy of a ligand was also restricted by the formation of a PI complex. The total entropy term was repulsive and, like the desolvation term, opposed the binding. The change of entropy accompanying the PI complex formation in a water environment was determined as the difference between entropies of a complex and a sum of the subsystem entropies. Ligand binding led also to the loss of the configurational entropy. The evaluation of the configuration entropy is notoriously difficult and this term was either neglected or was estimated on the basis of a reduced number of the accessible rotamers upon binding. In the present study we used the recently shown [48] evidence that vibrational entropy dominated configurational entropy. Consequently, it was recommended to evaluate the entropy term for the PI binding on the basis of the standard rigid rotor/harmonic oscillator approximation. The entropic contributions were determined using the rigid rotor/harmonic oscillator approximation based on Cornell et al. [49] empirical potential constants. The use of semiempirical Hamiltonian for the entropy calculations would be desirable but presently it is behind our possibilities. The structures of all of the systems were re-optimized using the empirical potential considering the continuous water, and the same method was used for the calculation of the second derivatives of the total energy.

Summarizing the procedure described we stress again that its main advantage is the fact that no any additional empirical parameter either for an individual component of the total score or for an individual PI complex was introduced.

Structures and inhibition constants

The structures of all the complexes (Table 1) were taken from the RSCB Protein Data Bank. Two different structural types were considered. In the first, the inhibitors are bound to the fully active form of CDK2 (Thr160-phosporylated CDK2 in complex with cyclin A3) while in the second type they are bound to the inactive form (monomeric CDK2). In the case of the Thr160-phosporylated CDK2/cyclin A3–inhibitor complexes, only one CDK2 monomer with removed cyclin A3 was used. A longer chain of CDK2 was considered for a complex with the 1OGU inhibitor. All of the crystal water molecules were removed. The hydrogen atoms were added to all of the structures considered, and their positions were minimised using AMBER [50]—ff03 [51] and gaff [52] force fields in Chimera software [53]. Partial charges of ligand atoms were calculated by RESP method from HF/6-31G* ESP charges for ff99 optimizations and from B3LYP/cc-pVTZ ESP charges for ff03 optimizations. In most cases, the crystallographically determined structures contained gaps in the chain (see Table 1), but they were situated far from the binding site of the inhibitor in complex, thus only a naturally charged group like COO^- and NH_3^+ was added to the amino acids on the gap termini to compensate for the charge of the whole structure. Staurosporines (1AQ1 and 1PKD) were considered N-protonated. The protonation of 1AQ1 ligand was deduced from the H-bonds network observed in the X-ray structure, however, the protonation state of 1PKD remained uncertain from the X-ray structure. The inhibition constants of all of the CDK2 inhibitors studied were taken from the literature and all of the references are listed in Table 1.

Strategy of calculations

The structures of all of the protein-inhibitor complexes were systematically reoptimised by the PM6-DH2 method in a continuum COSMO solvent model [54] as implemented in the MOPAC code (http://OpenMOPAC.net) according to MOPAC convergence criteria defined by exiting as soon as the gradient norm drop below 10.0 kcal/mol/Å. The respective interaction enthalpies ($\Delta H_w = \Delta H_w(\mathrm{PI}) - (\Delta H_w(\mathrm{P}) + \Delta H_w(\mathrm{I}))$) were determined at the same level ($T = 298$ K, $p = 1$ atm). In the second step, all of the complexes were reoptimised with AMBER ff99 or ff03 force fields (the missing ligand binding parameters were adopted from AMBER gaff force field) with a generalised Born solvent model [55] in the Nucleic Acid Builder (NAB, from the AMBER [50] package). The entropy term ($T = 298$ K, $p = 1$ atm.) was determined using the ideal gas/rigid rotor/harmonic oscillator approximation using the empirical potential, and the geometries of the systems were reoptimised at the empirical level with the L-BFGS TNCG algorithm to the gradient of 10^{-6} kcal/mol/Å and reoptimised by the Newton–Raphson method. This approach, assuming that the biomolecule occupies a single harmonic well, is known not to be fully adequate for biomolecules transversing many thermally accessible potential wells and

Table 1 A summary of the experimental data and calculated terms for the fifteen CDK2 inhibitors considered in this study

Structure	Res.[a]	Gaps[b]	ln(K_i)[c]	ln(err, K_i)[d]	ΔH_w^e	ΔE_{def}(I)[f]	$\Delta\Delta G_w$(I)[g]	$T\Delta S_w^h$	$\Delta G_w'$[i]
1AQ1 [73]	2.00	37–43, 150–160	−19.66 [74]	0.14	−55.63	−1.77	5.67	3.61	−55.34
1E1X [75]	1.85	37–43	−13.55 [75]	0.16	−31.66	0.75	−3.94	−23.45	−11.40
1PKD*	2.30	–	−17.32 [74]	0.24	−39.49	−0.40	7.87	−23.69	−8.33
1PXJ [76]	2.30	37–40	−11.94 [77]	0.50	−25.48	0.85	−2.87	−18.65	−8.85
1PXL [76]	2.50	37–43	−15.05 [77]	1.18	−34.35	0.55	−3.30	−17.82	−19.28
1PXM [76]	2.53	37–43	−16.63 [77]	0.55	−41.10	3.12	−1.45	−32.22	−7.21
1PXN [76]	2.50	37–40	−16.47 [77]	1.28	−48.83	3.52	−2.28	−27.61	−19.98
1PXP [76]	2.30	37–40	−15.33 [77]	1.88	−38.15	0.75	−3.99	−34.22	−7.17
2FVD [78]	1.85	38–49, 150–151	−19.62 [78]	0.27	−52.89	5.94	−7.03	−29.51	−24.47
1H1P [79]*	2.10	–	−11.33 [75]	0.26	−24.84	1.10	−1.38	−22.29	−2.83
1H1S [79]*	2.00	–	−18.93 [79]	0.08	−52.83	1.48	−2.62	−10.72	−43.25
1OGU [80]*	2.60	38–40	−17.55 [68]	0.35	−53.56	1.34	−1.87	−17.42	−36.67
2A4L [14]	2.40	36–47	−13.63 [81]	NA	−29.99	0.39	3.71	−13.97	−11.92
2EXM [82]	1.80	–	−9.46 [83]	NA	−22.10	0.19	−2.08	−15.97	−8.02
2X1N [84]*	2.75	–	−17.59 [84]	NA	−38.55	0.55	−4.52	−13.82	−28.70

NA not available

[a] Res. (in Å) is the experimental mean resolution of the respective X-ray structure, [b] missing residues in the X-ray structure, [c] ln(K_i) is the logarithm of inhibition constant (K_i in M), [d] ln(err. K_i) is the error of inhibition constant taken from literature, [e] ΔH_w is the interaction enthalpy, [f] ΔE_{def}(I) is correction for inhibitor deformation energy, [g] $\Delta\Delta G_w(I)$ is correction for ligand desolvation free energy, [h] $T\Delta S_w$ is the entropic contribution calculated by ff03, [i] $\Delta G_w'$ is the total score; all energies are in kcal/mol

* These structures contain the fully active form of CDK2, i.e. pT160-CDK2/Cyclin A3 system; only CDK2 with an inhibitor was considered here

may introduce some error. The interaction enthalpies (ΔH_w^{AMBER}) were also determined using the AMBER force fields ($T = 298$ K, $p = 1$ atm.) and generalised Born solvent model (Table 2).

The correction for the deformation energy of the inhibitor (ΔE_{def}(I)) was calculated as the difference between the energy of the inhibitor in the geometry taken from the protein/inhibitor complex ($E(I)^{PI}$) and the energy of the fully optimised inhibitor in water ($E(I)^{w}$). The PM6-DH2 optimizations with the COSMO solvation model were used for this calculation. The correction for the inhibitor deformation energy (ΔE_{def}^{AMBER}(I)) was also calculated at the empirical level using AMBER force fields with a generalised Born solvent model (Table 2).

The correction of the inhibitor desolvation free energy $\Delta\Delta G_w$(I) ($T = 298$ K, $p = 1$ atm.) was calculated on an inhibitor taken from the optimised CDK2–inhibitor complex. It was determined as the difference between the free energy of the inhibitor solvation calculated by the PM6-DH2/COSMO method ($\Delta G_w^{MOPAC}(I)^{PI}$) and the free energy of the inhibitor solvation calculated by the SMD model [47] (involving all of the non-electrostatic terms, i.e. the cavitation, dispersion and repulsion terms) at the HF/6-31G* level as implemented in Gaussian 09 [56] ($\Delta G_w^{G09}(I)^{PI}$) on PM6-DH2/COSMO geometry. We are aware that this energy difference is calculated with different hamiltonians (PM6 and DFT). However, better description of just the solvation of the ligand itself improves the final result. We believe that this procedure is also theoretically justified: the contribution from the protein is the difference between empty and filled active site, in our approach modelled consistently at PM6 level. In the case of the active site occupied by the ligand, only very small part of the surface of the ligand is exposed to the solvent, and therefore the contribution of the ligand to the $\Delta\Delta G_w$ of the protein is small. This separation of the second contribution, the solvation of the free ligand, allows us to use different method to calculate it more accurately.

We are aware of one important limitation of the model used which concerns the use of optimized crystal structure instead of averaged ones generated by MD simulations. The proteins are not rigid but moving in the solvent and, thus, the the single terms in the total score (Eq. 3) should be considered as an average quantity. In our papers as well as in majority of scoring function based studies this effect is neglected. The preliminary results obtained in our laboratory for different PI complexes indicate that effect of averaging is not critical. Nevertheless, these effects are under investigation in our laboratory and will be published as soon as possible.

Results

PM6-DH2 scoring of CDK2–inhibitor complexes

This study used fifteen structurally diverse CDK2 inhibitors with known inhibition constants (K_i) and the X-ray

Table 2 A summary of the experimental data and calculated terms for the fifteen CDK2 inhibitors by AMBER force fields ff99 and ff03

Structure	$\ln(K_i)$	ff99				ff03			
		ΔH_w^{MM} [a]	ΔE_{def}^{MM} [b]	$T\Delta S_w^c$	$\Delta Gw'^{MM}$ [d]	ΔH_w^{MM} [a]	ΔE_{def}^{MM} [b]	$T\Delta S_w^c$	$\Delta G_w'^{MM}$ [d]
1AQ1	−19.66	−39.61	6.58	−20.07	−12.97	−13.97	5.62	3.61	−11.96
1E1X	−13.55	−43.17	3.51	−29.00	−10.66	−36.55	3.08	−23.45	−10.02
1PKD	−17.32	−46.39	1.54	−20.42	−24.43	−41.14	2.12	−23.69	−15.33
1PXJ	−11.94	−12.48	1.75	−13.75	3.02	−17.56	3.64	−18.65	4.74
1PXL	−15.05	−41.09	1.24	−16.67	−23.18	−12.00	1.43	−17.82	7.24
1PXM	−16.63	−42.57	2.52	−41.19	1.13	−32.10	2.75	−32.22	2.87
1PXN	−16.47	−25.97	1.76	−20.22	−3.98	−39.33	3.11	−27.61	−8.61
1PXP	−15.33	−13.43	1.69	−7.75	−3.98	−53.10	1.62	−34.22	−17.26
2FVD	−19.62	−25.38	6.90	−28.47	10.00	−34.06	5.00	−29.51	0.44
1H1P	−11.33	−14.32	1.10	−13.70	0.48	−18.43	2.81	−22.29	6.66
1H1S	−18.93	−33.49	5.93	−20.95	−6.62	−31.72	5.12	−10.72	−15.87
1OGU	−17.55	−19.09	3.78	−10.97	−4.34	−29.97	2.84	−17.42	−9.71
2A4L	−13.63	−17.03	2.44	−22.15	7.55	−42.67	1.86	−13.97	−26.84
2EXM	−9.46	−19.58	1.67	−21.03	3.12	−36.92	1.26	−15.97	−19.69
2X1N	−17.59	−28.42	1.80	−22.20	−4.41	−20.62	2.14	−13.82	−4.65

[a] ΔH_w^{MM} stands for the interaction enthalpy calculated by AMBER ff99 or ff03 force fields, [b] ΔE_{def}^{MM} (I) corrections for inhibitor deformation energy, [c] $T\Delta S_w$ for the entropic contribution, [d] $\Delta G_w'^{MM}$ the total score; all energies are in kcal/mol

structures of CDK2–inhibitor complexes. The CDK2–inhibitor complexes were fully geometrically optimized (Table 3) by the PM6-DH2 method using the COSMO continuum solvent (involving the electrostatic terms only). The interaction enthalpy ΔH_w of the inhibitor with CDK2 was calculated at the same level as described in the Computational Details section. Figure 2a shows the correlation between the interaction enthalpy (ΔH_w) and the experimental inhibition constant ($\ln K_i$) values, where a very good correlation ($r^2 = 0.87$, $\ln K_i = (0.25 \pm 0.06)$ $\Delta H_w - (5.6 \pm 2.4)$, RMSE = 1.17, $n = 15$) was found. If the 1PKD is removed from the dataset (due to uncertain protonation state) the correlation slightly improves ($r^2 = 0.89$, $\ln K_i = (0.25 \pm 0.06)$ $\Delta H_w - (5.5 \pm 2.3)$, RMSE = 1.04, $n = 14$). In our previous study on HIV-1 protease, we demonstrated that the best correlation between the experimental binding free energy and theoretical data was not for the interaction enthalpy but for the total score containing also the interaction entropy, ligand deformation energy and ligand desolvation free energy. The good correlation obtained here between the interaction enthalpy and the inhibition constants indicates some compensation between the remaining terms. Our final goal is to estimate the free energy of the inhibitor binding (which is related to K_i, see Eq. 2), and, in order to achieve this, the interaction enthalpy ΔH_w has to be augmented by the interaction entropy term ($-T\Delta S_w$), inhibitor deformation energy (ΔE_{def}(I)) and correction for inhibitor desolvation. The solvation energy calculated by the COSMO model (implemented in MOPAC), which is involved in the ΔH_w, does not account for the non-electrostatic terms [57], and the estimate of the free energy of binding should be corrected also for this shortcoming. We have corrected the scoring function for the non-electrostatic terms of inhibitor solvation ($\Delta\Delta G_w$(I)). The correlation between the interaction enthalpy and inhibition constant is slightly worse when the corrections for inhibitor deformation ($\Delta E_{def}(I)$) and desolvation energies ($\Delta\Delta G_w(I)$) are also considered (Fig. 2b) with $r^2 = 0.77$ ($\ln K_i = (0.24 \pm 0.08)$ ($\Delta H_w + \Delta E_{def}(I) + \Delta\Delta G_w(I)$) $- (5.9 \pm 3.2)$, RMSE = 2.02, $n = 15$). In a multivariable linear fit, coefficients of both added variables ($\Delta E_{def}(I)$ and $\Delta\Delta G_w(I)$) to ΔH_w are statistically insignificant (at $\alpha = 0.05$) from zero ($\ln K_i = (0.25 \pm 0.07)$ $\Delta H_w + (0.00 \pm 0.56)$ $\Delta E_{def}(I) - (0.04 \pm 0.24)$ $\Delta\Delta G_w(I) - (5.7 \pm 2.7)$, $r^2 = 0.87$, RMSE = 1.15, $n = 15$). If the 1PKD is removed from the data set the correlation is significantly higher with $r^2 = 0.87$ ($\ln K_i = (0.26 \pm 0.06)$ ($\Delta H_w + \Delta E_{def}(I) + \Delta\Delta G_w(I)$) $- (5.0 \pm 2.6)$, RMSE = 1.18, $n = 14$). This indicates that corrections for ligand deformation and desolvation are not important in the case of CDK2. On the other hand, these corrections might become signifant in other systems as shown in the case of HIV-1 protease (see below). When, however, the interaction enthalpy is augmented only by the interaction entropy term ($-T\Delta S_w$, Fig. 2c) calculated at the empirical level (using AMBER ff03 force field), the correlation becomes worse ($r^2 = 0.56$, $\ln K_i = (0.14 \pm 0.08)$ ($\Delta H_w - T\Delta S_w$) $- (12.8 \pm 1.9)$, RMSE = 3.95, $n = 15$) but this value is still acceptable (the liner model and both variables are statistically significant at $\alpha = 0.05$, if the 1PKD structure is

Table 3 Summary of root-mean-square deviations (RMSD) of backbone Cα atoms and ligand atoms in PM6-DH (SQM) and AMBER (MM) optimized structures versus the X-ray structure

	SQM RMSD Cα	MM RMSD Cα	SQM RMSD lig	MM RMSD lig	X-ray E81 (backbone CO)	SQM	MM	X-ray L83 (backbone NH)	SQM	MM	X-ray L83 (backbone CO)	SQM	MM
1AQ1	0.63	1.63	0.24	0.27	2.75	2.89	2.80	2.63	2.84	2.84			
1E1X	0.52	1.14	0.39	0.35	2.64	2.85	2.75	2.97	3.17	3.09	2.49	2.71	2.82
1PKD	0.58	1.03	0.59	0.64	2.81	2.96	2.86	2.74	3.10	3.09			
1PXJ	0.57	1.46	0.23	0.62	2.86	2.95	2.81	3.30	3.62	3.23			
1PXL	0.49	1.20	0.46	0.79				2.92	3.25	3.02	2.58	2.79	3.04
1PXM	0.56	1.25	0.48	0.49				3.08	3.24	3.04	2.81	2.88	2.95
1PXN	0.59	1.35	0.39	0.82				2.85	3.35	2.93	2.58	2.94	2.94
1PXP	0.51	1.65	0.91	0.59				2.95	3.27	2.98	2.70	2.91	3.03
2FVD	0.49	1.23	0.49	0.46	2.74	2.79	2.82	3.68	3.32	3.19	2.76	2.87	2.86
1H1P	0.57	1.11	0.58	0.86	2.79	2.85	2.79	3.06	3.20	3.24	2.72	2.93	2.78
1H1S	0.60	1.04	0.28	0.32	2.83	2.85	2.79	3.35	3.24	3.37	2.82	2.81	2.82
1OGU	0.53	1.04	0.64	0.47	2.72	2.77	2.81	3.50	3.23	3.30	2.74	2.80	2.82
2A4L	0.58	1.18	0.61	0.80				3.38	3.36	3.37	2.82	2.89	2.88
2EXM	0.57	1.39	0.70	1.35	2.80	2.85	2.78	3.39	3.25	3.24			
2X1N	0.61	1.04	0.46	0.42				3.13	3.13	3.18	2.62	2.83	2.82

Values are given in Angstrom unit

Distances of key H-bonds between ligands and backbone atoms of CDK2 hinge residues (E81-L83) in X-ray, PM6-DH (SQM) and AMBER (MM) optimized structures

removed the correlation equals to $r^2 = 0.59$, ln $K_i = (0.15 \pm 0.08)$ $(\Delta H_w - T\Delta S_w) - (12.6 \pm 1.9)$, RMSE = 3.84, $n = 14$). In our previous study the inclusion of entropy term improved the correlation with experiment. This was fully in agreement with experimental finding showing that in the case of HIV-1 protease the entropy term is important. It is thus evident that entropy plays an important role (in case of HIV-1 protease) and it should not be neglected. Finally, when considering the complete scoring function $\Delta G_w'$, i.e. the sum of the interaction enthalpy, interaction entropy and both corrections for inhibitor deformation and solvation energies, the correlation with the experimental inhibition constants becomes $r^2 = 0.52$ (ln $K_i = (0.14 \pm 0.08)$ $\Delta G_w' - (12.8 \pm 2.0)$, RMSE = 4.31, $n = 15$), Fig. 2d; the removal of 1PKD does not improve the correlation significantly) and it is significantly worse than in the case of the interaction enthalpy (ΔH_w) itself. The decrease in the correlation between the calculated terms involving the entropy contribution and the inhibition constant is not surprising and documents that a reliable estimation of the entropic term cannot be based on the harmonic approximation, since a biomolecule can sample multiple thermally accessible minima, and also the empirical force field. On the other hand, the calculated entropy terms can play a significant role, as in the case of HIV-1 protease. For the sake of completeness we provide here also r^2 values for HIV-1 protease with 11 inhibitors: When all energy terms were included it equals to 0.52 and this correlation is comparable with r^2 value for HIV-1 protease [compare: $r^2 = 0.71$, ΔG_i (HIV-1 PR) = (0.11 ± 0.06) $\Delta G_w' - (14.5 \pm 0.9)$ vs. $r^2 = 0.52$, ln K_i (CDK2) = (0.17 ± 0.08) $\Delta G_w' - (11.8 \pm 2.2)$]. However, when only enthalpy term was included in the case of HIV-1 protease the correlation become statistically insignificant ($r^2 = 0.09$). The ligands of the HIV-1 protease are more flexible, therefore correction for ligand deformation and entropy play more important role in the case of HIV-1 protease.

The AMBER scoring of CDK2–inhibitor complexes

As in the previous case, the structures of all fifteen complexes of CDK2 with inhibitors were scored by AMBER force fields (both ff99 and ff03) with a generalised Born solvent model. Contrary to the previous case, where some single contributions to the total score were determined at different levels, in the present case all of the components of the total Amber score $\Delta G_w'^{AMBER}$, i.e. the sum of the interaction enthalpy ΔH_w^{AMBER}, interaction entropy $-T\Delta S_w$ and inhibitor deformation energy correction $\Delta E_{def}^{AMBER}(I)$, were determined consistently at the same empirical level. From Fig. 2e, it is evident that no statistically significant

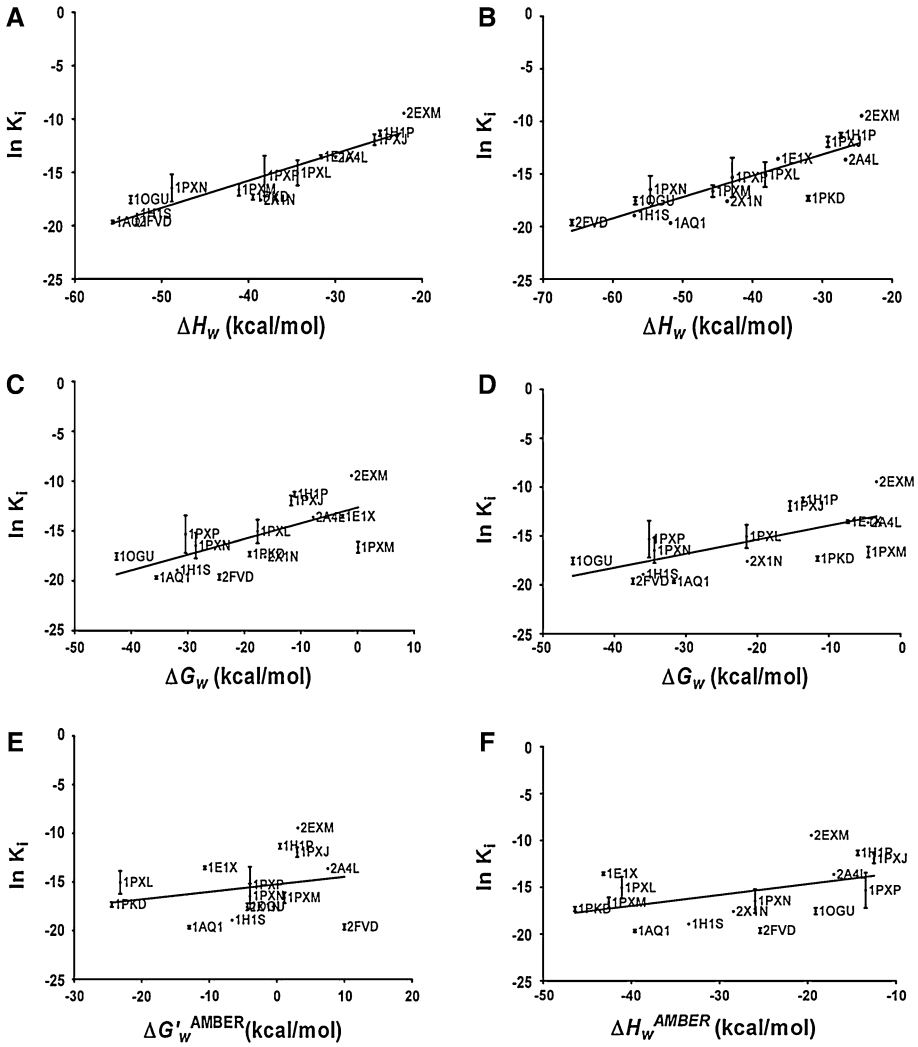

Fig. 2 The correlations between the inhibition constants (ln K_i) and **a** the interaction enthalpy ΔH_w (in kcal/mol) with $r^2 = 0.87$, **b** the sum of interaction enthalpy ΔH_w and the corrections for the inhibitor deformation ($\Delta E_{def}(I)$) and desolvation ($\Delta\Delta G_w(I)$) energies (in kcal/mol) with $r^2 = 0.77$, **c** the sum of the interaction enthalpy ΔH_w and interaction entropy $-T\Delta S_w$ (in kcal/mol) with $r^2 = 0.56$, **d** the total score $\Delta G_w'$ calculated as the sum of the interaction enthalpy (ΔH_w), the interaction entropy ($-T\Delta S_w$) and the corrections for the inhibitor deformation ($\Delta E_{def}(I)$) and desolvation ($\Delta\Delta G_w(I)$) energies (in kcal/mol) with $r^2 = 0.52$, **e** the total score $\Delta G_w'^{AMBER}$, where all of the terms were determined consistently using AMBER ff99 force field with $r^2 = 0.06$, **f** the interaction enthalpy ΔH_w^{AMBER} calculated using AMBER ff99 force field with $r^2 = 0.21$

correlation ($r^2 = 0.06$ for ff99 and $r^2 = 0.00$ for ff03) exists between the experimental inhibition constant (ln K_i) and the total AMBER score $\Delta G_w'^{AMBER}$, and the same applies also for the correlation ($r^2 = 0.21$ for ff99 and $r^2 = 0.00$ for ff03) with the interaction enthalpy ΔH_w^{AMBER} (Fig. 2f). This finding supports the idea that the interaction enthalpy should be determined as accurately as possible and that the empirical level used is clearly not adequate, at least in the case of CDK2.

Decomposition of the total score

Figure 3 shows the single contributions to the total PM6-DH2 score for four selected inhibitors. Following Eq. 3 the total score (the last bar) is constructed as a sum of the interaction enthalpy in solution (the first bar), the deformation energy of an inhibitor (the second bar), the correction for the inhibitor hydration free energy (the third bar) and the interaction entropy (the fourth bar). Evidently, the interaction enthalpy in solution (which varies from −34 to −52 kcal/mol) represents the dominant attractive term while deformation energy and entropy terms are repulsive and oppose the binding.

Optimized structures

Table 3 summarizes geometrical parameters of the optimized CDK2-inhibitor complexes (see Supplementary data for the optimized geometries). On average, the RMSD between X-ray and optimized structure is smaller for the PM6-DH method than for the empiricial calculation with ff99. According to a visual inspection, the PM6-DH and ff99 optimized structures are very similar, but 1PXP structure (and to some extent also 2FVD) optimized at the empirical level deviates from both X-ray and PM6-DH optimized structures. The key H-bond distances between inhibitor and CDK2 hinge residues are shorter in X-ray than in the optimized structures.

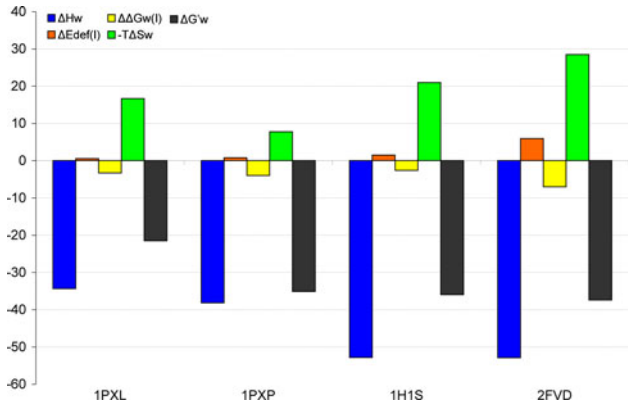

Fig. 3 The decomposition of the total score $\Delta G_w'$ into the interaction enthalpy in solution (ΔH_w), the deformation energy of an inhibitor (ΔE_{def}(I)), the correction for the inhibitor hydration free energy ($\Delta\Delta G_w$(I)) and the interaction entropy ($-T\Delta S_w$). All energies are in kcal/mol. The decomposition is shown for four inhibitors; **1PXP**: *N*-[4-(2,4-dimethyl-thiazol-5-yl)-pyrimidin-2-yl]-*n*′,*n*′-dimethyl-benzene-1,4-diamine, **1PXL**: 4-(2,4-dimethyl-1,3-thiazol-5-yl)-*n*-[4-(trifluoromethyl)phenyl]pyrimidin-2-amine[4-(2,4-dimethyl-thiazol-5-yl)-pyrimidin-2-yl]-(4-trifluoromethyl-phenyl)-amine], **2FVD**: (4-amino-2-([1-(methylsulfonyl)piperidin-4-yl]amino)pyrimidin-5-yl)(2,3-difluoro-6-methoxyphenyl)methanone, and **1H1S**: 4-[[6-(cyclohexylmethoxy)-9H-purin-2-yl]amino]benzenesulfonamide

Discussion

A sufficiently robust and reliable scoring function capable of scoring CDK2 inhibitors is an attractive but also challenging task. The attractiveness of this tasks arises from the fact that CDK2 is a member of a large family of protein kinases [58], which are involved in cellular signal transduction and the regulation of many cellular processes. Mutations and deregulations of protein kinases play also causal roles in many human diseases, and therefore great effort is invested in the development of protein kinase inhibitors (see Ref. [58] and the references therein). Despite this appeal, finding a reliable scoring function which could be highly predictive in the *in silico* drug design of CDK2 inhibitors still represents a challenging task. The first docking study, which was published 10 years ago, involved eight inhibitors, and the authors concluded that they had not found any statistically significant relationship between the activity (IC_{50}) and the interaction energy (using the DOCK score [59]) because of the low correlation coefficient ($r^2 = 0.15$) [60]. Later, many authors tried to find a more reliable function using the Gold score ($r^2 = 0.50$ between the Gold score and IC_{50}) [61], the Glide score ($r^2 = 0.61$ to IC_{50}) [62], the FlexX and LigandFit (the best score $r^2 = 0.60$ with IC_{50}) [63], GlideXP rescored by a molecular mechanics-generalised Born with a surface area (MM-GBSA) method ($r^2 = 0.69$ with IC_{50}) [64], scoring based on molecular dynamics simulations with a quantum-refined force field ($r^2 = 0.55$ with IC_{50}) [65], and the CDOCKER docked ligands rescored by the MM-GBSA method ($r^2 = 0.36$ after the removal of the outliers $r^2 = 0.63$, with pK_i) [66]. In general, the pharmacophore model named HypoRefire ($r^2 = 0.83$ between the score and IC_{50} for 302 molecules) [19] and 3D-QSAR CoMFA ($r^2 = 0.75$ to IC_{50}) [67] and CoMSIA ($r^2 = 0.81$ to IC_{50}) [67] gave better correlations than the scoring functions used by the docking software.

It is obvious that the best scoring function is able to explain only about 70% of the data variability and that the available scoring functions are not capable of describing all the physics behind the inhibitor binding to CDK2. Many authors have analysed the nature of CDK2–inhibitor binding using various methods and concluded that the dispersion interactions (part of van der Waals interactions) is important for inhibitor binding [17, 18, 27, 68] and may outweigh electrostatics [23]. Evidently, however, not only the dispersion but also the other terms should be properly covered. This is demonstrated by the fact that the total scoring with AMBER, which evaluates dispersion energy satisfactory [27, 69, 70], does not correlate well with the experimental data (see above). This further implies that also quantum chemical methods have to be chosen with caution, because any methods unable to cover the dispersion interaction (the HF method and the vast majority of the popular DFT functionals, e.g. B3LYP, and semiempirical methods, e.g. AM1 and PM6) will fail [26, 27]. In this respect, it is not surprising that the hybrid QM/MM approach did not provide any better correlation ($r^2 = 0.64$ with IC_{50} for 75 compounds) [71] than the empirical scoring functions, because it was based on the ONIOM model, where all of the layers (B3LYP/6-31G(d):HF/3-21G:PM3) failed in the description of the dispersion interaction. A large compensation of the gas-phase interaction energy by the solvation energy also has to be considered [18].

A reliable scoring function for CDK2 inhibitors has to involve a correct enough description of the dispersion and polarization interactions, but also the other physical processes have to be described carefully, as they might play an important role. Among them, the inhibitor deformation energy, changes in the inhibitor solvation (inhibitor desolvation) and the entropic contributions are likely candidates to be considered [64]. Here, we present a scoring function whose core is based on the semiempirical QM method accurately covering H-bonding and dispersion interaction (PM6-DH2) and further extended by corrections for the inhibitor deformation energy, the inhibitor desolvation and the entropic term. The last three terms are, however, based on empirical force field calculations. It must be stressed here that no empirical, adjustable parameter either for the single terms of the total score or for any individual inhibitor was introduced. This makes the

use of present technique for diverse protein–ligand complexes very promising. We have chosen fifteen structurally-diverse inhibitors of CDK2 with known X-ray structures and inhibition constants K_i. The correlation between pK_i and ΔH_w amounts to $r^2 = 0.87$, which is the best correlation found in the set of the above-discussed empirical scoring functions, pharmacological and 3D-QSAR models. When the corrections for the inhibitor deformation energy and inhibitor desolvation are involved, the correlation worsens to $r^2 = 0.77$. When full score is considered (i.e. also the entropic term is added), the correlation is significantly lower $r^2 = 0.52$. The decrease caused by the entropy term can be explained by the fact that a reliable calculation of the entropic changes based on the harmonic approximation and evaluated at empirical MM level is not adequate for biomolecules. The scoring function presented here yielded the best correlation with in K_i considering the empirical scoring function used in the docking experiments. There is also some room for improvement, as the correlation is not perfect, and specifically a better estimation of the entropic term would be useful. On the other hand, the presented scoring function represents significant progress in the field not only because it performs well but also because it does not require any additional empirical parameters, and therefore the availability of the protein–inhibitor complex (at least protein) structure is the only, albeit unavoidable, condition. The method is also capable of treating the complete CDK2–inhibitor complex, which typically contains about 5,000 atoms. The only known drawback is the computer time, because a single-point calculation (evaluation of one drug-receptor pose) typically takes $\sim$20 min (one core of Intel Core2 Quad 2.40 GHz processor) and the optimisation of the CDK2–inhibitor complex takes up to 3 weeks (1 core of Intel Core2 Quad 2.40 GHz processor). On the other hand, there is a large scope for acceleration, e.g. parallelisation, the usage of hybrid QM/MM models etc., and the increasing power of computers (Moor's law) works also for us. Considering all the *pros* and *cons*, the presented scoring function may represent a promising step in a development of a reliable scoring function for future drug design. This fact was supported by a very recent work by Zhou and Caflish, who showed that scoring functions based on quantum mechanics can be applied for high-throughput virtual screening [72].

Acknowledgments This work was a part of research project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and was supported by Grants No. LC512 and MSM6198959216 from the Ministry of Education, Youth and Sports of the Czech Republic. The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007, is also acknowledged. It was also supported by The Czech Science Foundation (P208/11/0295). This work was supported by the Operational Program Research and Development for Innovations—European Social Fund (CZ.1.05/2.1.00/03.0058).

References

1. Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol 13:261–291
2. Malumbres M, Barbacid M (2005) Mammalian cyclin-dependent kinases. Trends Biochem Sci 30(11):630–641
3. Malumbres M, Barbacid M (2009) Cell cycle, cdks and cancer: a changing paradigm. Nat Rev Cancer 9(3):153–166
4. Child ES, Hendrychova T, McCague K, Futreal A, Otyepka M, Mann DJ (2010) A cancer-derived mutation in the pstaire helix of cyclin-dependent kinase 2 alters the stability of cyclin binding. Biochim Biophys Acta Mol Cell Res 1803(7):858–864
5. Cohen P (1999) The development and therapeutic potential of protein kinase inhibitors. Curr Opin Chem Biol 3(4):459–465
6. Besson A, Dowdy SF, Roberts JM (2008) Cdk inhibitors: cell cycle regulators and beyond. Dev Cell 14(2):159–169
7. Fischer PM, Lane DP (2000) Inhibitors of cyclin-dependent kinases as anticancer therapeutics. Curr Med Chem 7:1213–1245
8. Knockaert M, Greengard P, Meijer L (2002) Pharmacological inhibitors of cyclin-dependent kinases. Trends Pharmacol Sci 23(9):417–425
9. Meijer L, Raymond E (2003) Roscovitine and other purines as kinase inhibitors. From starfish oocytes to clinical trials. Acc Chem Res 36(6):417–425
10. Johnson LN (2009) Protein kinase inhibitors: contributions from structure to clinical compounds. Q Rev Biophys 42(1):1–40
11. Tetsu O, McCormick F (2003) Proliferation of cancer cells despite cdk2 inhibition. Cancer Cell 3(3):233–245
12. Murray AW (2004) Revisiting the cell cycle: cyclins revisited. Cell 116:221–234
13. Echalier A, Endicott JA, Noble MEM (2010) Recent developments in cyclin-dependent kinase biochemical and structural studies. Biochim Biophys Acta Proteins Proteomics 1804(3):511–519
14. De Azevedo WF, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim SH (1997) Inhibition of cyclin-dependent kinases by purine analogues—crystal structure of human cdk2 complexed with roscovitine. Eur J Biochem 243(1–2):518–526
15. De Azevedo WF, Mueller-Dieckman HJ, Schulze-Gahmen U, Worland PJ, Sausville EA, Kim SH (1996) Structural basis for specificity and potency of a flavonoid inhibitor of human cdk2, a cell cycle kinase. Proc Natl Acad Sci USA 93:2735–2740
16. Ducrot P, Legraverend M, Grierson DS (2000) 3d-qsar comfa on cyclin-dependent kinase inhibitors. J Med Chem 43:4098–4108
17. Otyepka M, Krystof V, Havlicek L, Siglerova V, Strnad M, Koca J (2000) Docking-based development of purine-like inhibitors of cyclin-dependent kinase-2. J Med Chem 43:2506–2513
18. Otyepka M, Kriz Z, Koca J (2002) Dynamics and binding modes of free cdk2 and its two complexes with inhibitors studied by computer simulations. J Biomol Struct Dyn 20(2):141–154
19. Vadivelan S, Sinha BN, Irudayam SJ, Jagarlapudi SARP (2007) Virtual screening studies to design potent cdk2-cyclin a inhibitors. J Chem Inf Model 47(4):1526–1535
20. Halperin I, Ma BY, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins Struct Funct Genet 47(4):409–443
21. Brooijmans N, Kuntz ID (2003) Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 32:335–373
22. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 49(20):5851–5855

23. Zhou T, Huang D, Caflisch A (2008) Is quantum mechanics necessary for predicting binding free energy? J Med Chem 51(14):4280–4288
24. Raha K, Merz KM (2004) A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. J Am Chem Soc 126(4):1020–1021
25. Lu YX, Shi T, Wang Y, Yang HY, Yan XH, Luo XM, Jiang HL, Zhu WL (2009) Halogen bonding-a novel interaction for rational drug design? J Med Chem 52(9):2854–2862
26. Riley KE, Jurecka P, Pitonak M, Hobza P (2010) Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. Chem Rev 110:5023–5063
27. Dobes P, Otyepka M, Strnad M, Hobza P (2006) Interaction energies for the purine inhibitor roscovitine with cyclin-dependent kinase 2: correlated ab initio quantum-chemical, dft and empirical calculations. Chem Eur J 12(16):4297–4304
28. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz KM (2007) The role of quantum mechanics in structure-based drug design. Drug Dis Today 12(17–18):725–731
29. Zhou T, Huang DZ, Caflisch A (2010) Quantum mechanical methods for drug design. Curr Top Med Chem 10(1):33–45
30. Menikarachchi LC, Gascon JA (2010) Qm/mm approaches in medicinal chemistry research. Curr Top Med Chem 10(1):46–54
31. Cavalli A, Carloni P, Recanatini M (2006) Target-related applications of first principles quantum chemical methods in drug design. Chem Rev 106(9):3497–3519
32. Peters MB, Raha K, Merz KM (2006) Quantum mechanics in structure-based drug design. Curr Opin Drug Discovery Dev 9(3):370–379
33. Raha K, Merz KM (2005) Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. J Med Chem 48(14):4558–4575
34. Khandelwal A, Lukacova V, Comez D, Kroll DM, Raha S, Balaz S (2005) A combination of docking, qm/mm methods, and md simulation for binding affinity estimation of metalloprotein ligands. J Med Chem 48(17):5437–5447
35. Merz KM (2010) Limits of free energy computation for protein-ligand interactions. J Chem Theory Comput 6(5):1769–1776
36. Wollacott AM, Merz KM (2007) Assessment of semiempirical quantum mechanical methods for the evaluation of protein structures. J Chem Theory Comput 3(4):1609–1619
37. Stewart JJP (2007) Optimization of parameters for semiempirical methods v: modification of nddo approximations and application to 70 elements. J Mol Model 13(12):1173–1213
38. Stewart JJP (2009) Application of the pm6 method to modeling proteins. J Mol Model 15(7):765–805
39. Rezac J, Fanfrlik J, Salahub D, Hobza P (2009) Semiempirical quantum chemical pm6 method augmented by dispersion and h-bonding correction terms reliably describes various types of noncovalent complexes. J Chem Theory Comput 5(7):1749–1760
40. Korth M, Pitonak M, Rezac J, Hobza P (2010) A transferable h-bonding correction for semiempirical quantum-chemical methods. J Chem Theory Comput 6(1):344–352
41. Hayik SA, Dunbrack R, Merz KM (2010) Mixed quantum mechanics/molecular mechanics scoring function to predict protein-ligand binding affinity. J Chem Theory Comput 6(10): 3079–3091
42. Fanfrlik J, Bronowska AK, Rezac J, Prenosil O, Konvalinka J, Hobza P (2010) 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 114:12666–12678
43. Jurecka P, Cerny J, Hobza P, Salahub DR (2007) Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. J Comput Chem 28(2):555–569
44. Deng YQ, Roux B (2006) Calculation of standard binding free energies: aromatic molecules in the t4 lysozyme l99a mutant. J Chem Theory Comput 2(5):1255–1273
45. Deng YQ, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B 113(8):2234–2246
46. Wang JY, Deng YQ, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. Biophys J 91(8):2798–2814
47. Marenich AV, Cramer CJ, Truhlar DG (2009) 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(18):6378–6396
48. Chang CEA, Chen W, Gilson MK (2007) Ligand configurational entropy and protein binding. Proc Natl Acad Sci USA 104(5):1534–1539
49. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz JKM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A 2nd generation force-field for simulation of proteins, nucleic-acids and organic-molecules. J Am Chem Soc 117:5179–5197
50. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688
51. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong GM, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang JM, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24(16): 1999–2012
52. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174
53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) Ucsf chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
54. Klamt A, Schuurmann G (1993) 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(5):799–805
55. Jayaram B, Sprous D, Beveridge DL (1998) Solvation free energy of biomacromolecules: parameters for a modified generalized born model consistent with the amber force field. J Phys Chem B 102(47):9571–9576
56. Frisch MJ (2009) Gaussian 09, revision a.02. Gaussian Inc., Wallingford
57. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105(8):2999–3093
58. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298:1912–1934
59. Ewing TJA, Kuntz ID (1997) Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 18(9):1175–1189
60. Otyepka M, Bartova I, Kriz Z, Koca J (2006) Different mechanisms of cdk5 and cdk2 activation as revealed by cdk5/p25 and cdk2/cyclin a dynamics. J Biol Chem 281:7271–7281
61. Ibrahim DA, El-Metwally AM (2010) Design, synthesis, and biological evaluation of novel pyrimidine derivatives as cdk2 inhibitors. Eur J Med Chem 45(3):1158–1166

62. Duca JS, Madison VS, Voigt JH (2008) Cross-docking of inhibitors into cdk2 structures. 1. J Chem Inf Model 48(3): 659–668
63. Sato H, Shewchuk LM, Tang J (2006) Prediction of multiple binding modes of the cdk2 inhibitors, anilinopyrazoles, using the automated docking programs gold, flexx, and ligandfit: an evaluation of performance. J Chem Inf Model 46(6):2552–2562
64. Guimaraes CRW, Cardozo M (2008) Mm-gb/sa rescoring of docking poses in structure-based lead optimization. J Chem Inf Model 48(5):958–970
65. Ferrara P, Curioni A, Vangrevelinghe E, Meyer T, Mordasini T, Andreoni W, Acklin P, Jacoby E (2006) New scoring functions for virtual screening from molecular dynamics simulations with a quantum-refined force-field (qrff-md). Application to cyclin-dependent kinase 2. J Chem Inf Model 46(1):254–263
66. Pearce BC, Langley DR, Kang J, Huang HW, Kulkarni A (2009) E-novo: an automated workflow for efficient structure-based lead optimization. J Chem Inf Model 49(7):1797–1809
67. Dessalew N, Singh SK (2008) 3d-qsar comfa and comsia study on benzodipyrazoles as cyclin dependent kinase 2 inhibitors. Med Chem 4(4):313–321
68. Heady L, Fernandez-Serra M, Mancera RL, Joyce S, Venkitaraman AR, Artacho E, Skylaris CK, Ciacchi LC, Payne MC (2006) Novel structural features of cdk inhibition revealed by an ab initio computational method combined with dynamic simulations. J Med Chem 49(17):5141–5153
69. Hobza P, Muller-Dethlefs K (2010) Non-covalent interactions. RSC Publishing, Cambridge
70. Zgarbova M, Otyepka M, Sponer J, Hobza P, Jurecka P (2010) Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of amber intermolecular terms with rigorous dft-sapt calculations. Phys Chem Chem Phys 12: 9611–9614
71. Alzate-Morales JH, Caballero J, Jague AV, Nilo FDG (2009) Insights into the structural basis of n2 and o6 substituted guanine derivatives as cyclin-dependent kinase 2 (cdk2) inhibitors: prediction of the binding modes and potency of the inhibitors by docking and oniom calculations. J Chem Inf Model 49(4): 886–899
72. Zhou T, Caflisch A (2010) High-throughput virtual screening using quantum mechanical probes: discovery of selective kinase inhibitors. Chem Med Chem 5(7):1007–1014
73. Lawrie AM, Noble MEM, Tunnah P, Brown NR, Johnson LN, Endicott JA (1997) Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with cdk2. Nat Struct Biol 4:796–801
74. Zhao B, Bower MJ, McDevitt PJ, Zhao HZ, Davis ST, Johanson KO, Green SM, Concha NO, Zhou BBS (2002) Structural basis for chk1 inhibition by ucn-01. J Biol Chem 277(48):46609–46615
75. Arris CE, Boyle TF, Calvert AH, Curtin NJ, Endicott JA, Garman EF, Gibson AE, Golding BT, Grant S, Griffin RJ, Jewsbury P, Johnson LN, Lawrie AM, Newell DR, Noble MEM, Sausville EA, Schultz R, Yu W (2000) Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. J Med Chem 43:2797–2804
76. Wu SY, McNae I, Kontopidis G, McClue SJ, McInnes C, Stewart KJ, Wang SD, Zheleva DI, Marriage H, Lane DP, Taylor P, Fischer PM, Walkinshaw MD (2003) Discovery of a novel family of cdk inhibitors with the program lidaeus: structural basis for ligand-induced disordering of the activation loop. Structure 11(4):399–410
77. Wang SD, Meades C, Wood G, Osnowski A, Anderson S, Yuill R, Thomas M, Mezna M, Jackson W, Midgley C, Griffiths G, Fleming I, Green S, McNae I, Wu SY, McInnes C, Zheleva D, Walkinshaw MD, Fischer PM (2004) 2-anilino-4-(thiazol-5-yl)pyrimidine cdk inhibitors: synthesis, sar analysis, x-ray crystallography, and biological activity. J Med Chem 47(7):1662–1675
78. Chu XJ, DePinto W, Bartkovitz D, So SS, Vu BT, Packman K, Lukacs C, Ding QJ, Jiang N, Wang K, Goelzer P, Yin XF, Smith MA, Higgins BX, Chen YS, Xiang Q, Moliterni J, Kaplan G, Graves B, Lovey A, Fotouhi N (2006) Discovery of [4-amino-2-(1-methanesulfonylpiperidin-4-ylamino) pyrimidin-5-yl](2, 3-difluoro-6-methoxyphenyl)methanone (r547), a potent and selective cyclin-dependent kinase inhibitor with significant in vivo antitumor activity. J Med Chem 49(22):6549–6560
79. Davies TG, Bentley J, Arris CE, Boyle TF, Curtin NJ, Endicott JA, Gibson AE, Golding BT, Griffin RJ, Hardcastle IR, Jewsbury P, Johnson LN, Mesguiche V, Newell DR, Noble MEM, Tucker JA, Wang L, Whitfield HJ (2002) Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. Nat Struct Biol 9(10):745–749
80. Sayle KL, Bentley J, Boyle FT, Calvert AH, Cheng YZ, Curtin NJ, Endicott JA, Golding BT, Hardcastle IR, Jewsbury P, Mesguiche V, Newell DR, Noble MEM, Parsons RJ, Pratt DJ, Wang LZ, Griffin RJ (2003) Structure-based design of 2-arylamino-4-cyclohexylmethyl-5-nitroso-6-aminopyrimidine inhibitors of cyclin-dependent kinases 1 and 2. Bioorg Med Chem Lett 13(18):3079–3082
81. Meijer L, Bisagni E, Legraverend M (1998) Purine derivatives having, in particular, antiproliferative properties, and their biological uses USA Patent
82. Schulze-Gahmen U, Brandsen J, Jones HD, Morgan D, Meijer L, Veselý J, Kim S-H (1995) Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with atp and two inhibitors, olomoucine and isopenthenyladenine. Proteins Struct Funct Genet 22:378–391
83. Veselý J, Havlíček L, Strnad M, Blow JJ, Donella-Deana A, Pinna L, Letham DS, Kato J-Y, Detivaud L, Leclerc S, Meijer L (1994) Inhibition of cyclin dependent kinases by purine analogues. Eur J Biochem 224:771–786
84. McIntyre NA, McInnes C, Griffiths G, Barnett AL, Kontopidis G, Slawin AMZ, Jackson W, Thomas M, Zheleva DI, Wang SD, Blake DG, Westwood NJ, Fischer PM (2010) Design, synthesis, and evaluation of 2-methyl- and 2-amino-n-aryl-4, 5-dihydrothiazolo[4, 5-h]quinazolin-8-amines as ring-constrained 2-anilino-4-(thiazol-5-yl)pyrimidine cyclin-dependent kinase inhibitors. J Med Chem 53(5):2136–2145