PALACKÝ UNIVERSITY, OLOMOUC

Faculty of Science

Department of Physical Chemistry

Mgr. Cemal KÖPRÜLÜOĞLU

Doctoral Dissertation

COMPUTER-AIDED DRUG DESIGN

Supervisor

Prof. Ing. Pavel HOBZA, DrSc., Dr. h. c., FRSC

Institute of Organic Chemistry and Biochemistry

of the Czech Academy of Sciences, Prague

Olomouc 2020

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká Fakulta

Katedra Fyzikální Chemie

Mgr. Cemal KÖPRÜLÜOĞLU

Disertační práce

COMPUTER-AIDED DRUG DESIGN

Školitel

Prof. Ing. Pavel HOBZA, DrSc., Dr. h. c., FRSC

Ústav organické chemie a biochemie, Akademie věd Českě
Republiky, v.v.i., Praha

Olomouc 2020

©Cemal KÖPRÜLÜOĞLU, 2020

All rights reserved

DECLARATION OF AUTHORSHIP

I hereby declare that the matter embodied in this thesis entitled "COMPUTER AIDED DRUG DESIGN" is the result of investigations carried out by me at the Department of Physical Chemistry, Palacky University, Olomouc under the supervision of Prof. Ing. Pavel HOBZA, DrSc., Dr. h. c., FRSC and it has not been submitted elsewhere for any degree, diploma or associateship of any university or institute. In keeping with the general practice in reporting the scientific observations and all the literature is properly cited.

Olomouc, April 2020 CEMAL KÖPRÜLÜOĞLU

Prof. Ing. PAVEL HOBZA

Dedicated to the memory of my Dad, who always believed in my ability to be successful in the academic arena. You are gone but your belief in me has made this journey possible...

ACKNOWLEDGEMENT .... **xi**
LIST OF FIGURES .... **xiii**
LIST OF TABLES .... **xv**
LIST OF ABREVIATIONS .... **xvii**
List of publications in which the author of this dissertation participated as principal / co-author .... **xix**
ABSTRACT .... **xxi**
ABSTRAKT .... **xxiii**
1. INTRODUCTION **1**
  1.1 A Historical Perspective 1
  1.2 Computer-Aided Drug Design 2
  1.3 Molecular Recognition 5
    1.3.1 Binding Free Affinity 6
  1.4 The role of quantum mechanics in structure-based drug design 7
2. STRUCTURE-BASED DRUG DESIGN (SBDD) **9**
  2.1 Target Preparation 9
  2.2 Molecular Docking 10
    2.2.1 Systematic methods 11
    2.2.2 Random and Stochastic methods 12
    2.2.3 Simulation methods 12
  2.3 Scoring Functions 13
  2.4 Non-Covalent Interactions and ΔG in molecular recognition 15
3. PROJECTS **19**
  3.1 Sampling Power and Native Binding Mode Analysis 19
  3.2 Scoring (Ranking) Power and the Role of Explicit Water 22
  3.3 Screening Power and the Virtual Screening Performance 27
4. CONCLUSION **34**
5. REFERENCES **35**
6. PRESENTATION OF THE RESULTS **55**
7. DECLARATION OF THE AUTHORSHIP **56**
PUBLICATIONS **59**

ACKNOWLEDGEMENT

This PhD thesis is written for the completion of my Ph.D. studies at Palacky University, Olomouc. At first, I would like to thank to my research supervisor Prof. Pavel Hobza for giving me the opportunity to work with him, his insightful guidance and generous support. I will always be grateful for his support for me.

I also would like to thank to my colleagues Dr. Jan Řezáč, Dr. Jindřich Fanfrlík, Dr. Martin Lepšík, Dr. Adam Pecina, Dr. Vijay Madhav Miriyala, Kristian Kříž for their amazing help and support. I would like to give my special thanks to Mrs. Helena Cerna for her kind support and constant help.

I would like to thank to Prof. Michal Otyepka, Dr. Petr Jurečka, Sylva Kaděrková and Dana Gronychová for their constant encouragement and support in the University during my studies.

I have heartfelt thanks to Saltuk Eyrilmez, Gizem Eyrilmez and Taner Eyrilmez for their amazing support during my PhD as a family, Especially Saltuk, your help and assistance during my studies that unforgettable.

I would like to also express my deep-felt gratitude to my friends outside of my work. Martin Žižka, Volkan Öztürk, Stella Skiadopoulou, Giovanni Acquaviva, Şebnem Kurhan, Agnieszka Kozub, Berna Kramer, Emre Erden, Carina Santos Hurtado, Federico Urban, Olatz Ruiz Larrabeiti, Lorenzo Reverberi, Zeynep Ergenekon Acuner, Baha Acuner, Arda Karakaya, Müge Karakaya, Mehmet Onur, Banu Sevüker Onur, Aliözgür Öztürk, Aslı Yavaş, Ozan Şahin Öztürk, Sibel Öztürk, Burak-Betül Önal, Nail-Gamze Yücel, Pavel Zedníček life would be pointless without you.

Finally, the last slot is reserved for my lovely family. My cousins; Elvan-Oktay Erdeniz, Sibel-Fevzi Özdemir, Aksel-Meltem Özbudur, Burçin Sevin, Makbule Baskıcı and my lovely aunty Neşe Özbudur, my lovely super granny Ayla Tolun, I would not make it real without your support. My cordial thanks to my only brother Ali Köprülüoğlu and his wife İlkay Köprülüoğlu. And my deepest gratitude goes to my lovely mom Tülay Tolun and my late Dad nothing would happen without you.

And of course, my lovely baby kittens Köfte and Dudo.

LIST OF FIGURES

FIGURE 1: Drug Discovery and Development Process....2
FIGURE 2 Total Crystal Structures in the RCSB PDB Database 3
FIGURE 3 Computer Aided Drug Design.....4
FIGURE 4 Some of Clinically Approved CADD Drugs.....5
FIGURE 5 Binding Free Energy Equilibrium. 6
FIGURE 6 Examples of Hydrogen Bonds Observed in P-L Complexes. 17
FIGURE 7 Influenza Neuraminidase Inhibitors 18
FIGURE 8 Synthesis of Imidazo[1,2-c]Pyrimidin-5(6H)-One Core. 20
FIGURE 9 The Binding Mode of the Imidazo[1,2-c]Pyrimidin-5(6H)-One Compounds in Active CDK2/Cyclin E Conformation....21
FIGURE 10 Three Hits for CDK2 Inhibition from the IOCB Proprietary Database 23
FIGURE 11 The Binding Mode of 18a in CDK2....24
FIGURE 12 PM6-D3H4X/COSMO Score Plotted against Experimental Binding Free Energies 25
FIGURE 13 Structural Water Molecules in the Crystallographic Complex of CDK2/1L. 26
FIGURE 14 ROC Curve. 30
FIGURE 15 PROC Curve....31
FIGURE 16 Individual Representation of RMSD vs. Energy Docking/Scoring Results....33

LIST OF TABLES

TABLE 1 List of Some Docking Software 12
TABLE 2 Thermodynamic Characteristics of Water Molecules in CDK2/1L.........................26
TABLE 3 The ROC Enrichment Factors EF, AUC and PROC AUC 29

LIST OF ABREVIATIONS

CADD – Computer Aided Drug Design

CDK2 – Cyclin-Dependent Kinase

COSMO – COnductor-like Screening MOdel

DFT – Density Functional Theory

DF – Docking Function

GA – Genetic Algorithm

HTS – High Throughput Screening

LBDD – Ligand-Based Drug Design

MC – Monte Carlo

MD – Molecular Dynamics

MM – Molecular Mechanics

NMR – Nuclear Magnetic Resonance

PDB – Protein Data Bank

PM6-D3H4X – Parametrized Model 6 – Dispersion, Hydrogen and Halogen Bonding Correction

QM – Quantum Mechanics

RMSD – Root Mean Square Deviation

SAR - Structure-Activity Relationship

SBDD – Structure-Based Drug Design

SF – Scoring Function

SQM – Semiempirical Quantum Mechanics

VDW – van der Waals

VS – Virtual Screening

List of publications in which the author of this dissertation participated as principal / co-author

A. Haresh Ajani, Josef Jansa, **Cemal Köprülüoğlu**, Pavel Hobza, Vladimír Kryštof, Antonín Lyčka, Martin Lepsik `Imidazo[1,2-c]pyrimidin-5(6H)-one as a novel core of cyclin-dependent kinase 2 inhibitors: Synthesis, activity measurement, docking, and quantum mechanical scoring` *Journal of Molecular Recognition*, **2018** (https://doi.org/10.1002/jmr.2720).

B. **Cemal Köprülüoğlu**, Milan Dejmek, Michal Šála, Haresh Ajani, Hubert Hřebabecký, Jindřich Fanfrlík, Radek Jorda, Martin Dračínský, Eliška Procházková, Pavel Šácha, Vladimír Kryštof, Pavel Hobza, Martin Lepšík and Radim Nencka `Optimization of Norbornyl-based Carbocyclic Nucleoside Analogs as Cyclin-Dependent Kinase 2 Inhibitors` *Journal of Molecular Recognition*, **2020** (https://doi.org/10.1002/jmr.2842), (Cemal Köprülüoğlu and Milan Dejmek contributed equally).

C. Michaela Hylsova, Benoit Carbain, Jindrich Fanfrlík, Lenka Musilova, Susanta Haldar, **Cemal Köprülüoğlu**, Haresh Ajani, Pathik S. Brahmkshatriya, Radek Jorda, Vladimír Krystof, Pavel Hobza, Aude Echalier, Kamil Paruch, Martin Lepsík `Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by newpyrazolo[1,5-a]pyrimidines` *European Journal of Medicinal Chemistry*, **2017**. (https://doi.org/10.1016/j.ejmech.2016.12.023).

D. Saltuk M. Eyrilmez, **Cemal Köprülüoğlu**, Jan Řezáč, and Pavel Hobza `Impressive Enrichment of Semiempirical QuantumMechanics-Based Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys` *ChemPhysChem*, **2019** (https://doi.org/10.1002/cphc.201900628), (Saltuk M. Eyrilmez and Cemal Köprülüoğlu contributed equally).

E. Adam Pecina, Saltuk M. Eyrilmez, **Cemal Köprülüoğlu**, Vijay Madhav Miriyala, Martin Lepšík, Jindřich Fanfrlík, Jan Řezáč, and Pavel Hobza `SQM/COSMO Scoring Function: Reliable Quantum-Mechanical Tool for Structure-Based Drug Design` submitted to *ChemPlusChem*, **2020.**

ABSTRACT

Protein – Ligand (P-L) interaction and accurate identification of these interactions are crucial for computer-aided drug design. To achieve success in the identification of P-L complexes, many scoring functions (SFs) and docking functions (DFs) have been developed and tested. The most important approach used in structure-based drug design and structural-based virtual screening is molecular docking. Molecular docking predicts native binding poses (sampling power), ranks the binding affinities (ranking power) and discriminates active compounds from the inactives (screening power).

Variety of different programs and tools are available for docking/scoring. The two most important components of the docking programs are the sampling algorithm and the SF. Sampling algorithm gives the possible geometric positions within the active site, while the SF predicts the binding affinity relationship between these poses. Docking programs use different sampling strategies and SFs. Docking programs have been successful so far in many protein-ligand systems. However, their SFs often fail in difficult systems such as metalloproteins, halogenated ligands, inorganic ligands etc. To overcome these problems, we have developed a novel semiempirical quantum mechanics (SQM)-based SF. The newly introduced SQM SF has achieved far better performance in sampling, ranking and screening power against the widely used classical SFs.

In this dissertation, ligand design, structure-activity relationship (SAR), native binding mode prediction, molecular docking, virtual screening, water thermodynamics and SQM-based scoring were studied associated with sampling power, ranking power and screening power.

ABSTRAKT

Interakce mezi bílkovinami a ligandy a přesná identifikace těchto interakcí jsou klíčové pro návrh léčiv pomocí výpočetních prostředků. K dosažení úspěchu v identifikaci interakcí v komplexech bílkovin a jejich inhibitorů bylo vyvinuto a testováno již mnoho skórovacích a dokovacích funkcí. Nejdůležitější z přístupů, který se používá v počítačovém návrhu léčiv a virtuálním screenování, je molekulární dokování. Molekulární dokování dokáže předpovědět možné vazebné motivy - na kterém místě a jak se ligand váže (samplovací síla, “sampling power”), dokáže tyto potenciální vazebné motivy seřadit na základě vazebné síly (skórovací síla, “scoring power”) a takto oddělit aktivní látky od neaktivních (screenovací síla, “screening power”).

Pro dokování a skórování lze dnes využít množství různých počítačových programů a výpočetních nástrojů. Obecně, dvě základní komponenty těchto programů jsou samplovací algoritmus a skórovací funkce. Samplovací algoritmus generuje možné orientace ligandu v aktivním místě proteinu a skórovací funkce dokáže předpovědět relativní sílu vazby těchto jednotlivých orientací. Různé dokovací programy využívají různé samplovací strategie a dokovací funkce. Dokovací programy byly již úspěšně použity na mnoho P-L komplexů, avšak jejich skórovací funkce často selhávají v případě obtížných systémů (tj. v případě systémů obsahujících kov v aktivním místě nebo v případě systémů s anorganickým nebo halogen-obsahujícím ligandem, apod.). Naše nová skórovací funkce založená na semiempirických kvantově-mechanických metodách (SQM) má sloužit k překonání zmíněných problémů. Tato skórovací funkce již dosáhla daleko lepších výsledků v samplování, rankování a screenování než ostatní, dnes běžně užívané skórovací funkce.

Tato dizertační práce se zabývá návrhem léčiv, vztahem molekulární geometrie a aktivity látky, předpovídáním přirozených vazebných motivů, molekulárním dokováním, virtuálním screenováním, termodynamikou vodného prostředí a SQM metodami ve vztahu k samplovací, rankovací a screenovací síle.

1. INTRODUCTION

Drugs are mostly small organic molecules that activates or inhibits the function of the biomacromolecule such as protein by covalent or non-covalent interactions [1]. Some drugs can be also proteins such as insulin. The most important advantage of small molecule drugs over the large molecules is that small molecules can be taken orally while large molecules require injection [2]. Currently, it is very important to find new therapeutic solutions for many diseases. However, these solutions can be expensive and time-consuming. From lead identification up to clinical trials, estimated procedure can take over a decade and costs beyond billion dollars [3] (cf Figure 1). The process of drug design and development is time and cost dependent, and therefore computer modelling techniques are often used to reduce the before mentioned [4]. This modelling is called computer-aided drug design (CADD). When the three-dimensional structure information is known, it is called structure-based drug design (SBDD) else it is called ligand-based drug design (LBDD) [5].

1.1 A Historical Perspective

In the antiquity, there was no way to understand the biological source of a disease. In the medieval ages, biologically active substances in nature were often interpreted more teleologically. Well known "Doctrine of Signatures" was introduced by the Swiss-Austrian physicians and scientists from the sixteenth century. The formulation of this doctrine is in perfect harmony with the philosophies at that time, and even today, it is still used in many countries. Although this doctrine is not used in modern medicine and the research of natural products today, this idea was the first approach of rational drug design. More than 100 years ago, it was noticed that the cause of a particular therapeutic response was produced by certain molecules. The effect has been first described by Emil Fischer and then more details were expounded by John Langley and Paul Ehrlich. In this concept drug and target fits perfectly such as key into its corresponding lock. However, nowadays we know the fact that receptors are highly flexible and we carry out our research with this in mind [6]. Since the late 1980s dramatic developments have been experienced for drug design & discovery in both industrial and academic studies. This period has seen the introduction of high throughput screening (HTS), combinatorial chemistry, PC farms, structure-based design (SBDD), virtual screening (VS), free-energy methods, bioinformatics, structure of ion channels, structure/activity relationships (SAR) obtained from nuclear magnetic resonance (SAR by NMR), and human genome sequences. The results thus accelerated the progression period for discovery and development process [5].

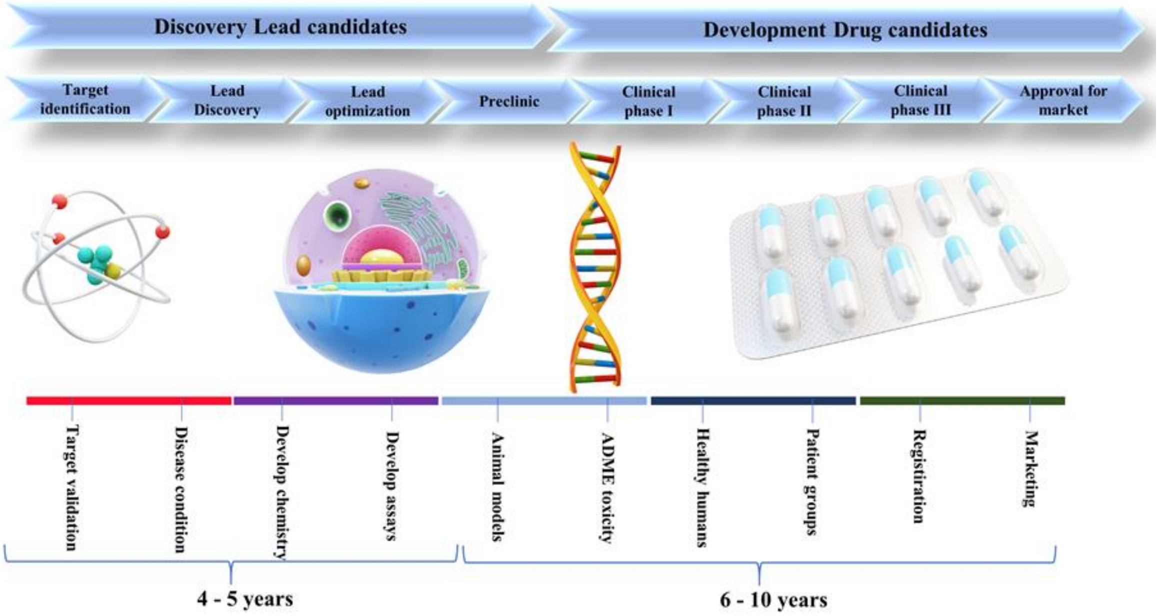

Discovery Lead candidates
Development Drug candidates
Target identification
Lead Discovery
Lead optimization
Preclinic
Clinical phase I
Clinical phase II
Clinical phase III
Approval for market
Target validation
Disease condition
Develop chemistry
Develop assays
Animal models
ADME toxicity
Healthy humans
Patient groups
Registiration
Marketing
4 - 5 years
6 - 10 years

Figure 1: Drug discovery and development process. Adapted from *Textbook of Drug Design and Discovery (p.5), by K. Stromgaard, P. Krogsgaard-Larsen, and U. Madsen, Eds. CRC Press, 2016.*

1.2 Computer-Aided Drug Design

CADD is a commonly used term for the design of compounds with interesting physicochemical properties, including the evaluation of these compounds as potential drug candidates before they are synthesized and tested [7]. The role of calculations in drug development has increased significantly since the 1960s [8]–[10]. With the rapid development of computational power and efficiency in this field, CADD has become an important tool in drug discovery process [11]. Apart from these developments, the fast increase in protein crystal structures deposited by Research Collaboratory for Structural Bioinformatics (RCSB) has increased interest in this area. The number of crystal structures, which were very few in the 1980s, has reached 140000 today. (i.e. Figure 2) [12]. Numerous crystal structure information is of great importance for the modeling of protein/ligand (P-L) complexes [5]. One of the common uses of CADD is the virtual screening of compound libraries, also known as virtual high-throughput screening (vHTS).

Since vHTS allows the removal of a large number of inactive compounds, this process saves money and time [13]. To date, CADD has had many successful applications [14]–[17]. CADD is beneficial for hit-to-lead optimisation as well as reducing the number of compounds to be synthesized and tested in vitro [13], [18]. There are two approaches to CADD, Structure-based drug design (SBDD) and Ligand-based drug design (LBDD), depending on the available structure information (illustrated Figure 3) [13]. Some of the succesful applications demonstrated in Figure 4. SBDD will be explained in detail later.

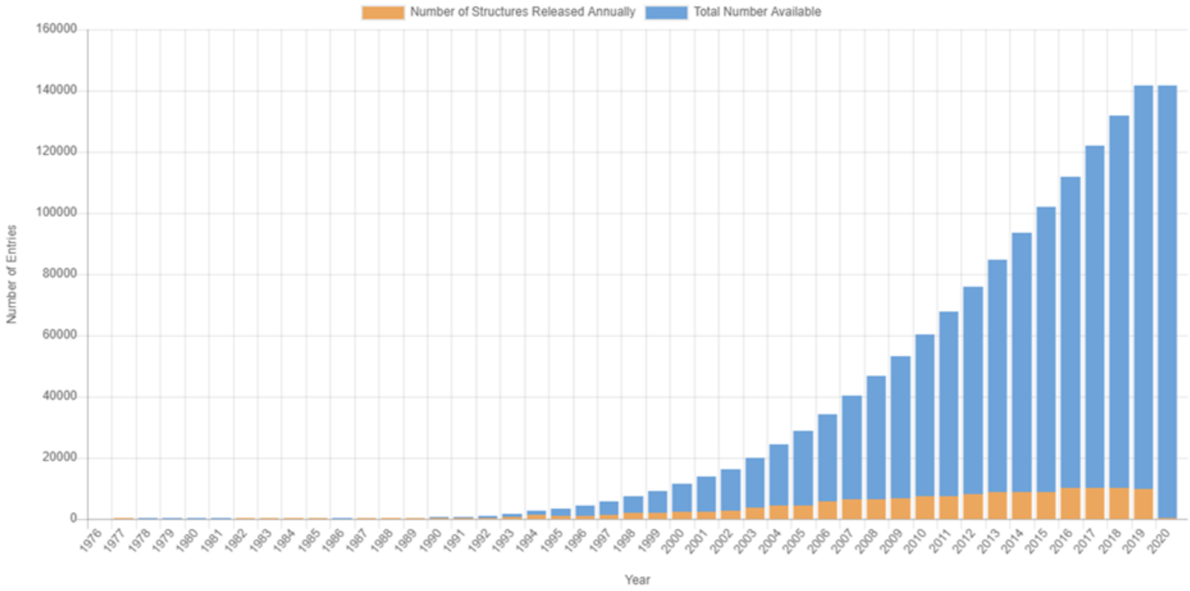


Figure 2 Total crystal structures in the RCSB PDB database as of January 11, 2020 (www.rcsb.org)

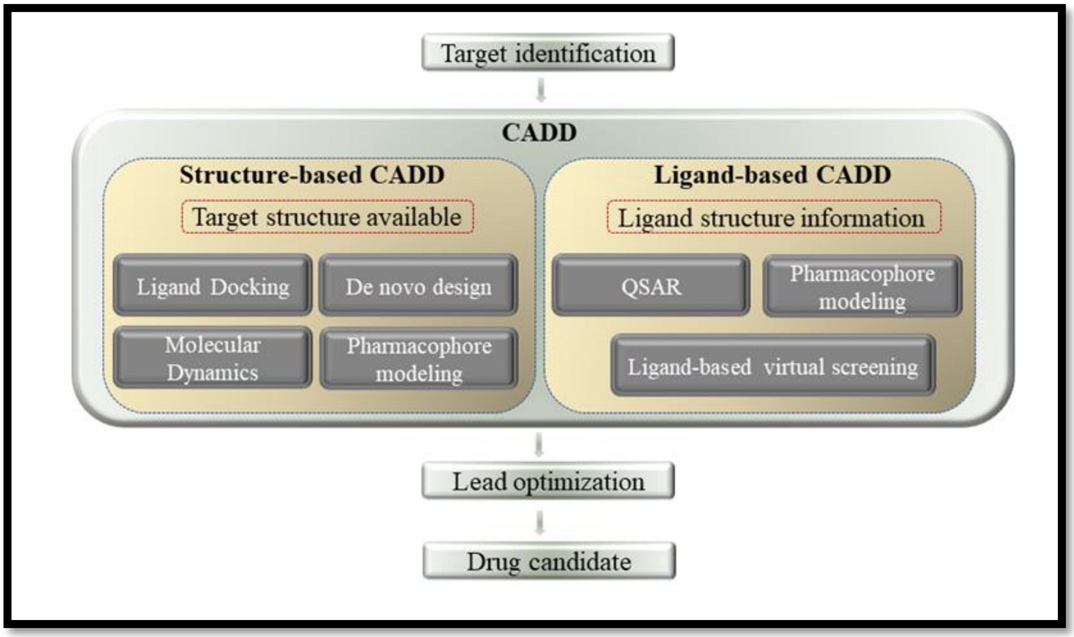

Target identification
CADD
Structure-based CADD
Target structure available
Ligand Docking
De novo design
Molecular Dynamics
Pharmacophore modeling
Ligand-based CADD
Ligand structure information
QSAR
Pharmacophore modeling
Ligand-based virtual screening
Lead optimization
Drug candidate

Figure 3 Computer Aided Drug Design. Adapted from "*Computational Methods in Drug Discovery," G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, Pharmacological Reviews, vol. 66, no. 1. pp. 334–395, 2014.*

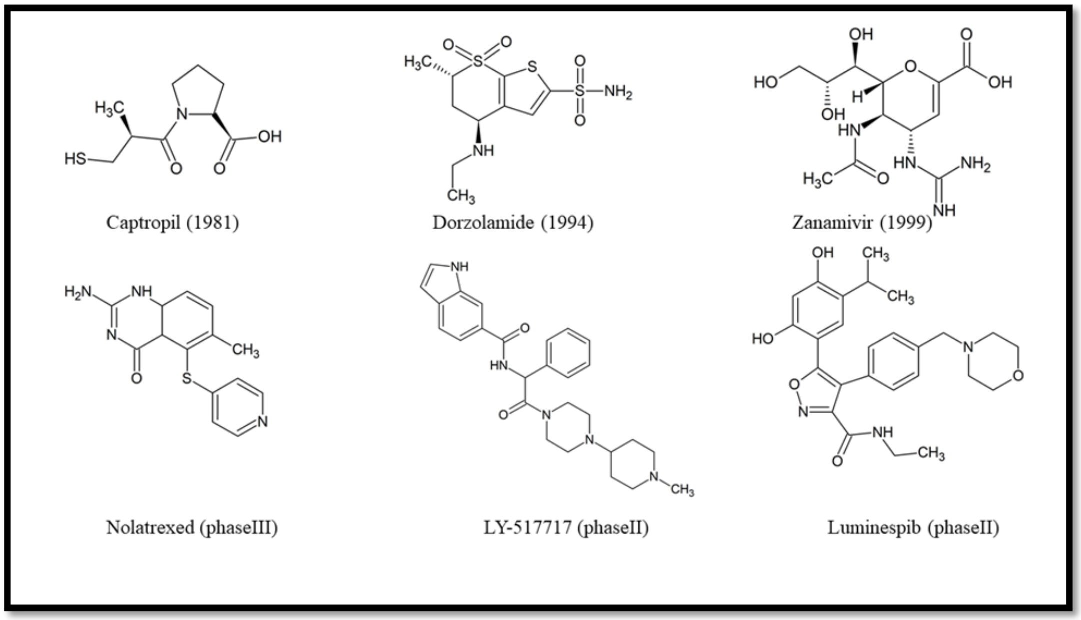


Figure 4 Some of clinically approved CADD Drugs.

1.3 Molecular Recognition

Molecular recognition is very important in understanding biological systems such as protein-ligand, antigen-antibody, sugar-lectin [19], [20] etc. In such systems the main interactions occur via hydrogen bonding, metal coordination, hydrophobic forces, van der Waals forces, π-π interactions, halogen bonding and/or electrostatic forces [21]–[26]. It depends on the favourable interaction of two or more molecules such as a neurotransmitter molecule and its protein, receptor or an enzyme and its substrate. The recognition of one molecule by another is explained by the energetics [6].

The highly selective molecular recognition was first described by Emil Fischer as "lock-and-key principle" [27]. According to this principle, the substrate and the enzyme complement each other like a key and lock. In the case of competitive inhibition, inhibitor binds to the active site in such a way that the substrate cannot bind to the active site. In the case of non-competitive inhibition, inhibitor binds the allosteric site of the protein. This changes the geometry of active site in such a way that the substrate cannot bind. Lock-and-key principle is thus unable to explain "non-competitive inhibition" mechanism [28]. In order to explain this better, the theory of "induced fit" is proposed by Daniel E. Koshland in 1958 [29]. There are a huge number of protein structures available in the Protein Databank (RCSB PDB) and these X-ray structures enable this theory to be valid today [6].

1.3.1 Binding Free Affinity

Binding free energy (called also binding affinity) of Protein – Ligand (P-L) complex can be defined as the equilibrium between bound P-L complex and “free unbound ligand and protein as demonstrated in Figure 5.

The relation between free energy difference and the equilibrium constant K can be given as,

$$\Delta G = -RT \ln \mathrm{K} \tag{1}$$

where R is the gas constant (8.315 J/K/mol) and T is the absolute temperature.

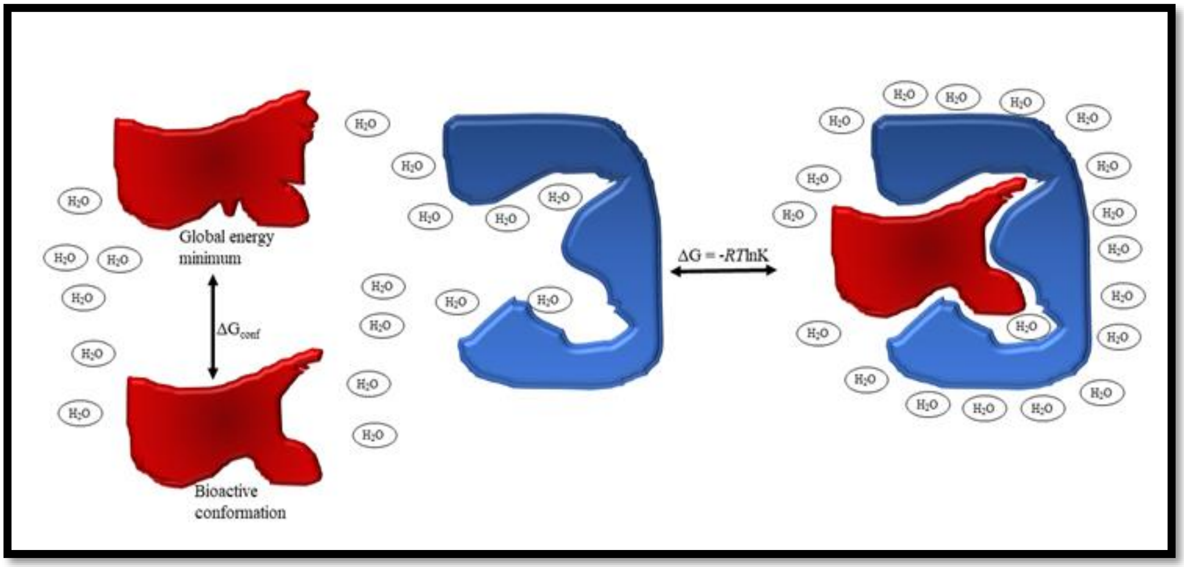


Figure 5 Binding Free Energy Equilibrium. Adapted from *Textbook of Drug Design and Discovery (p.17), by K. Stromgaard, P. Krogsgaard-Larsen, and U. Madsen, Eds. CRC Press, 2016.*

A higher affinity is the result of the equilibrium being pushed to the bound complex on the right side in Figure 5. Thus, K value becomes more positive and ΔG more negative. In medicinal chemistry the affinity is mostly given either by inhibition constant (K_i) or the half maximal inhibitory concentration (IC_{50}). Since $K = 1/K_i$ the equation 1 can be written as

$$\Delta \mathrm{G} = RT \ln \mathrm{K_i} \tag{2}$$

The ΔG can be also defined using the enthalpic (ΔH) and entropic (ΔS) components as

$$\Delta G = \Delta H - T\Delta S \tag{3}$$

In some cases, IC_{50} can be used instead of K_i values, in which case the IC_{50} values is converted to an inhibition constant K_i by the Cheng-Prusoff equation[30]:

$$K_i = \frac{IC_{50}}{\left(1 + \frac{[L]}{K_D}\right)} \tag{4}$$

where [L] is the concentration of the radioligand used in the assay, K_D is the affinity of the radioligand for the receptor.

It should be noted that IC_{50} value is related to activity and K_i is related to binding. In addition, IC_{50} values are concentration dependent, while K_i is independent of radioligand and its concentration. Radioligand is a radiolabelled biochemical substance that can be used for diagnosis or research orientated studies. Thus, K_i values can be compared across the assays [6].

1.4 The role of quantum mechanics in structure-based drug design

Using quantum mechanics (QM) in all phases of in silico drug design represents the next step in this field. However, the application of quantum mechanics to large systems is limited [5]. Development of advanced computational facilities and computers in the recent times, made the application of quantum mechanics possible in many fields [31], [32]. Molecular dynamic (MD) simulation, Molecular Mechanics (MM), QM, DFT, and semi-empirical QM calculations can be used as a key tool in the area. MM can be used for very large systems whereas the semi-empirical quantum mechanics methods or DFT methods can be applied for smaller systems [33]. MM fails many times, especially in difficult systems such as metalloproteins, halogenated ligands and inorganic ligands etc. [34]. QM/MM hybrid algorithms have been applied in our laboratory successfully for many systems [35]–[44]. The use of QM in *In Silico* drug design can be divided into two main categories: Structure-based drug design (SBDD) and ligand-based drug design (LBDD). SBDD requires the knowledge of x-ray or nuclear magnetic resonance (NMR) structures of the complex while LBDD rely on the knowledge of ligand structures [5]. Molecular docking is one of the most important tools and is widely used in structure-based drug design

(SBDD) and structure-based virtual screening (SBVS) [45]–[48]. Molecular docking can predict the native binding poses (sampling power), rank the binding affinities (scoring power) and discriminate active compounds from inactives (screening power) [49]. Semi-empirical QM-based scoring function (SQM) developed in our laboratory. P-L complexes exist in a solvent environment so the conductor-like screening implicit solvent model (COSMO) has been used [50]. SQM/COSMO SF was successfully used in cases where the standard scoring functions of the docking programs have failed. Moreover it has achieved far more sampling power performance against the widely used classical SFs for four and seventeen difficult P-L systems [39], [44], [51]. Another difficult task represents correlation between calculated score and binding affinities. Ranking power (scoring power) simply describes the ability of SF to rank different ligands in the same target. For this purpose, we applied SQM/COSMO SF to ten known carbonic anhydrase II – ligand complexes. We achieved much better results than ten different classical SFs used [41]. The last step is the virtual screening performance (screening power) of the SF. In case of screening power the success of SQM/COSMO SF is ***impressive*** compared to other SFs [52].

2. STRUCTURE-BASED DRUG DESIGN (SBDD)

The SBDD relies on the knowledge of the three-dimensional (3D) structure of the biomolecular target. 3D structure may be obtained by x-ray crystallography, NMR spectroscopy [13], [53] or neutron scattering spectroscopy [54]. The large increase in the number of 3-D protein identified in Protein Databank since 1980s (cf Figure 2) as well as the progress in genomic and proteomic studies [55], [56] led to the discovery of a large number of candidate drug targets. The resolution of the crystal structure to be used in the SBDD is very important, and should preferably be under 2.5Å. [57], [58].

In drug design, computational methods such as virtual screening (VS) allow rapid screening for hit identification and enable the analysis of large libraries to find potential candidates [59]. VS is more direct, cheaper and more effective compared to traditional experimental high-throughput screening (HTS) [60]–[62]. Molecular Docking aims to predict P-L complex structures by investigating conformational changes of the ligand within the binding site of the protein. A scoring function is applied to evaluate binding free energy between the protein and ligand in each docking pose [54].

2.1 Target Preparation

As mentioned in the previous section, the target structure is determined experimentally by X-ray crystallography or NMR techniques. Structures are deposited in the PDB and this is the ideal starting point for molecular docking [13]. Some virtual screening approaches have been reported such as comparative modelling which can be used in the absence of an experimentally determined structures [63]–[65]. Comparative modelling is also known as Homology Modelling. Comparative modelling involves five main steps [13]:

1. Identification of related structures
2. Sequence alignment
3. Copying coordinates for confidently aligned regions
4. Model generation and loop modeling
5. Model assessment and evaluation

In PDB structure file, the structure has generally no information on topologies, formal atomic charges and bond orders. To tackle structural issues such as terminal amide groups, asparagine residues, ionization and tautomeric states or steric clashes, protein preparation should be applied [54], [66]–[68]. Some popular free software are available which determine the protonation state of the amino acids in the protein, such as PROPKA [69], H++ [70], SPORES [71]. Other important steps are hydrogen assignment, partial charge assignment, capping of residues, treating metals, filling in missing loops and minimization the protein structure [54]. Protein Preparation Wizard might be used for these tasks [67], [72]. In the projects we have carried out in this thesis, all proteins are generally prepared using Maestro [73].

2.2 Molecular Docking

Molecular docking is the most popular method in SBDD [74]. Fundamentally, the main purpose of molecular docking is to determine the P-L structures using computational methods. A large number of docking software have been developed [75] such as AutoDock [76], Glide [77], FlexX [78], Surflex [79], GOLD [80], AutoDock Vina [81], SMINA [82], MOE-Dock etc. [83]. Two interrelated steps provide docking results: the first step is the sampling of ligand conformations in the active site of the protein; the other is the scoring of the individual conformations. Experimental binding modes should be reproduced by sampling algorithms and generated conformations should be ranked by scoring function [59]. In another word docking predicts the binding modes of P-L complexes and ranks in terms of their score values in SBVS [54]. For successful modelling the system must contain protein, ligand and solvent molecules. The number of degrees of freedom subsumed in the conformational space is the main part for the searching efficiency [84]. In the biological systems the P-L complexes are surrounded by solvent molecules. Solvent molecules are associated with the large number of rotational, translational and dihedral(torsional) degrees of freedom. For that reason solvent molecules either can be excluded or modelled implicitly [85]. Docking can be performed in three different ways: a) rigid docking, where the target and ligand are rigid; b) flexible ligand docking, where the ligand is flexible and the target is rigid; c) flexible docking, such as induced fit docking where both ligand and target are flexible [48]. Docking programs use many different search methods/algorithms to explore the ligand conformational space. These search methods can be examined in three sub-categories: i) Systematic methods; ii) Random and Stochastic methods; iii) Simulation methods [86].

2.2.1 Systematic methods

P-L complexes have huge number of possible binding modes due to translational, rotational as well as conformational degrees of freedom for ligand and protein. To generate all of them computationally would be so expensive [59]. Systematic algorithms try to explore all the degrees of freedom in a molecule [85] where the current state determines the next state during the search [13]. Systematic methods can be categorized into (1) exhaustive search algorithms (2) fragmentation algorithms and (3) database methods [13], [59], [85].

In the exhaustive search ligand conformations can be obtained by systematically rotating all possible rotatable bonds at a given interval. Even though many docking programs provide systematic methods, many of them are not able to provide exhaustive search of the conformational space [87]. Programs such as Glide [77] and eHiTS [87] offer exhaustive search algorithms and contain good scoring for ligand poses.

Matching algorithms (MA) [88]–[90] available in DOCK [91] and FLOG [92] programs might be used for the large libraries due to the speed [59]. In the incremental construction (IC) a ligand is divided into several fragments by breaking the rotatable bonds and one of these fragments is docked into the active site first and the remaining fragments can be added incrementally [59], [85]. The IC method is available in DOCK 4.0 [93], FlexX [78] and Hammerhead [94]. Multiple copy simultaneous search (MCSS) [95], [96] and LUDI [97] belongs also among fragment based methods. MCSS makes 1,000 to 5,000 copies of functional group placed in the binding site randomly while LUDI focuses on the hydrogen bond and hydrophobic contacts [59].

2.2.2 Random and Stochastic methods

The stochastic method strategy increases the probability of finding the global minimum while avoids final local energy minimum [98]. In this method all degrees of freedom of ligand is changing randomly at each state while sampling the conformational space [85]. Monte Carlo (MC) [99], [100], genetic algorithm (GA) [80], [101], [102] and tabu search methods [103]–[105] represents typical algorithms for random and stochastic methods [59], [85].

2.2.3 Simulation methods

Simulation methods are based on the solution of Newton’s equations of motion [85]. Molecular dynamics (MD) [106]–[108] is a simulation method most preferred in molecular modelling in many fields. MD treats both protein and ligand as flexible. MD generally proceeds in small steps, which can be considered as a disadvantage. Consequently it might be difficult to overcome high energy conformational barriers resulting in inadequate sampling [59].

Table 1 List of some docking software

Method	Feature	Licence
Glide	Exhaustive search-based docking program	Commercial
Gold	Genetic algorithm-based docking program	Commercial
AutoDock	Genetic Algorithm and Empirical Scoring Function	Open Source
AutoDock Vina	New generation of AutoDock	Open Source
DOCK	Based on Geometric Matching Algorithm	Freeware
SMINA	A customized fork of AutoDock Vina	Open Source

2.3 Scoring Functions

Correct prediction of P-L conformations using sampling algorithms is essential and important, but equally important is to rank these structures by scoring them. The ideal scoring function should be able to distinguish correct binding mode from other alternative modes. In other words, the scoring function should be able to distinguish the correct binding pose from the incorrect ones or binder from nonbinders in an affordable computational time. Scoring functions that give accurate results can be computationally very expensive and time consuming. Various approaches have been used to solve or reduce these complexities. Scoring functions can be considered as force-field based, empirical, knowledge-based and physic-based scoring functions [59], [85].

Force-field-based scoring functions [109]–[111] determine the binding free energy by calculating the sum of electrostatic and van der Waals interactions. It is generally calculated as the sum of the interaction energy and the internal energy of ligand. Coulombic law is used to calculate the electrostatic terms and distance-dependent dielectric function modulates the charge-charge interactions in the solvent. The van der Waals term is defined by Lennard-Jones potential function. Computational efficiency is one of the main problem of force-field-based scoring functions [47], [59], [85], [112], [113]. Docking software programs such as DOCK [91], [93], GOLD [114] and AutoDock [101] have some differences in the treatment of hydrogen bonds. In addition to this, free-energy perturbation methods (FEP) predict the binding energies accurately, however can only be used for similar ligands [109], [115]. Complete molecular mechanical force-fields programs such as AMBER (Assisted Model Building and Energy Refinement) [106], [116], [117], CHARMM (Chemistry at HARvard Macromolecular Mechanics) [108], GROMOS (Groningen Molecular Simulation System) [118] and OPLS (Optimized Potentials for Liquid Simulations) can also be considered [119].

Empirical scoring functions approximate binding free energy [120]–[124] by a sum of hydrogen bond, ionic interaction, hydrophobic effect and binding entropy. For the final score each component is multiplied by a coefficient and then summed up to give a final score. The main advantage of empirical scoring functions is their computer efficiency. On the other hand, their disadvantage is dependence of the experimental data. LUDI [97] , PLP [121], [122], [125] and ChemScore [126] codes involve empirical scoring functions.

Knowledge-based scoring functions [127]–[132] use available data of ligand-protein crystal structures. PMF [127], DrugScore [133] and Bleep [128] are well known examples of knowledge-based scoring functions.

Consensus Scoring [134], [135] is a recently used strategy to assess the docking conformation by combination of different scoring functions to improve the ability of finding the correct solution. Although the prediction of binding energies might be inaccurate, this scoring function substantially improves the enrichment in virtual screening [136].

Regarding the Physic-based scoring functions, molecular mechanic methods combined with solvation free energy term can be used to estimate the binding affinities [137]. Quantum mechanical effects in P-L interactions such as charge transfer, polarization, σ-hole are crucial for computer-based drug design. Although QM calculations are demanding they provide accurate description of these effects [138], [139]. SQM-based scoring functions were introduced by Kenneth Merz group [140]. SQM-based scoring function showed superior performance in the case of metalloprotein [140], [141] but further corrections describing hydrogen bonding and dispersion were necessary [142], [143]. This SF provides fast, reliable and the accurate description of non-covalent interactions such as hydrogen bond, halogen bond and dispersion [144]–[146]. In our laboratory we have developed PM6-D3H4X method, which does not need any specific parametrization. It was successfully used for various P-L complexes [37], [39], [41], [44], [51], [52], [147]–[149]. One of the most important components of docking represents scoring function. Most docking programs predicts known P-L structures with the success rate in the range of 70 – 80% [150]–[152].

2.4 Non-Covalent Interactions and ΔG in molecular recognition

The first step in drug action represents molecular recognition. To understand its medicinal chemistry, the physicochemistry should be examined in detail. ΔG express the free energy difference and it is the sum of three main components.

- Electrostatic interactions (ion-ion, ion-dipole, and dipole-dipole interactions)
- Lipophilicity/hydrophobicity
- Shape complementarity

In addition, entropy and conformational changes affect molecular recognition as well. Several partitioning have been suggested in the literature but the eqution 5 proposed by Williams et al. [153] will be used.

$$\Delta G = \Delta G_{\mathrm{transl+rot}} + \Delta G_{\mathrm{conf}} + \Delta G_{\mathrm{polar}} + \Delta G_{\mathrm{hydrophob}} + \Delta G_{\mathrm{vdW}} \qquad (5)$$

2.4.1 Overall molecular motion

Both protein and ligand in the unbound state rotate and translate. When they create a complex, ligand loses its ability to move freely and this loss of freedom is thermodynamically unfavourable. This restriction of motion causes a decrease in entropy, which makes ΔS more negative. According to the equation 3, more negative ΔS makes more positive ΔG. The calculations show that $\Delta G_{\mathrm{transl+rot}}$ mainly depends on the "tightness" of P-L complex [6].

2.4.2 Conformational Changes

The $\Delta G_{\mathrm{transl+rot}}$ that we mentioned in the previous section represents the relative "overall" motion of the molecule and its surroundings. The ligand conformation in the solvent is different from the ligand conformation when it is complexed with the protein. The structures obtained from Xray, shows that the ligand does not preserve the most stable conformation determined by state-of-the-art computational methods. In general, rigid molecules are more advantageous in terms of ΔG_{conf} than flexible molecules. The loss of conformational entropy for the rigid molecules is low. Conformational changes of the protein are as important as the ligand conformational changes according to the binding of the different ligands. Structurally diverse ligands may change not only the side chain conformations but also backbone conformation. Thus, protein flexibility is one of the major challenges in SBDD.

2.4.3 Electrostatic interactions and hydrogen bonding

Changes in ΔG_{polar} are generally caused by interactions of amino acids containing polar functional groups with polar groups of ligands. C=O and NH backbone groups in the active site of the protein are also responsible for these changes. In many cases bridging water molecules in the binding cavity mediates indirect protein-ligand interactions. Electrostatic interactions occur between positively and negatively charged molecules. This interaction is attractive when it is caused by opposite charges, and repulsive when the same charges are considered. The strength of electrostatic interactions is indicated by the following formulation:

$$E_{polar} = \frac{q_i q_j}{\varepsilon r_{ij}} \tag{6}$$

where q_i and q_j partial atomic charges, ε is the dielectric constant, r_{ij} is the distance between the charges

While the dielectric constant equals to 1 (no shielding) in vacuum, it is 78.4 (25 °C) in water. However, quantifying the E_{polar} value for protein is a difficult task and depends on the microenvironment of the protein. Ion-ion interactions are independent of relative orientation, while ion-dipole and dipole-dipole strongly dependent on geometry.

2.4.3.1 Hydrogen Bonds

X-H…Y hydrogen bond is defined as an attractive interaction between a hydrogen bound to an electronegative atom X (such as nitrogen or oxygen) and an electronegative hydrogen bond acceptor where electrostatic energy plays a dominant role. The distance between the heavy atoms X and Y, is between 2.5 and 3.0 Å. The hydrogen bond is strongly orientation dependent and the X-H-Y angle is 180^o in the optimal conditions. Three types of hydrogen bonds are shown in Figure 6.

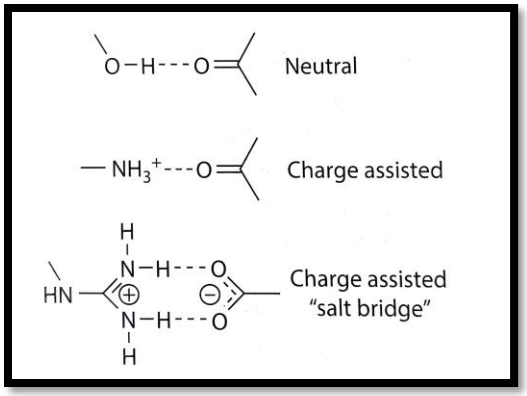


Figure 6 Examples of hydrogen bonds observed in P-L complexes. Reprinted from *Textbook of Drug Design and Discovery (p.22), by K. Stromgaard, P. Krogsgaard-Larsen, and U. Madsen, Eds. CRC Press, 2016.*

In order to understand the contribution of the hydrogen bond to the system, the equilibrium process shown in Figure 5 should be analysed. Before forming the complex polar groups of ligand and protein were surrounded by water molecules. After forming complex solvent molecules are replaced by hydrogen bonds between protein and ligand molecules. As a result, this exchange process determines the hydrogen bond contribution to the system.

2.4.3.2 Polar interactions in the Aromatic ring systems

Other polar interactions such as π-π and cation-π can be observed in the P-L complexes. Two benzene rings can be considered as example of π-π interactions. Other aromatic rings such as phenol and indole have similar electrostatic potentials. Thus, the aromatic rings of protein such as tyrosine, tryptophan and phenylalanine interact with the positively charged functional groups of ligands.

2.4.4 The hydrophobic effect

The hydrophobic effect is generally the change of the hydrogen bond dynamic network in the liquid after the interaction of a lipophilic compound with water. Water molecules are more ordered around the lipophilic compound than water molecules in a bulk water, and this change leads to an increase in entropy, resulting in more negative ΔG. The first thing to focus on when designing ligands is hydrogen bonds and electrostatic interactions. However, in many cases hydrophobic interactions may be more favourable than strong hydrogen bond interactions (cf Figure 7).

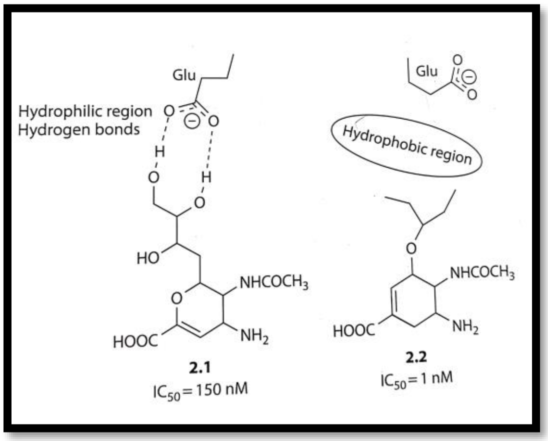


Figure 7 Influenza neuraminidase inhibitors. Reprinted from *Textbook of Drug Design and Discovery (p.25), by K. Stromgaard, P. Krogsgaard-Larsen, and U. Madsen, Eds. CRC Press, 2016.*

2.4.5 van der Waals interactions

Nonpolar interactions are also called van der Waals interactions and might be attractive or repulsive. They entirely depend on the distance of the interacting groups (in Equation 7)

$$\mathrm{EvdW} = \frac{A}{r^{12}} - \frac{B}{r^{6}} \tag{7}$$

Although atom-atom interaction is weak, they play an important role in P-L interactions when considered collectively. Correct and proper identification of dispersive interactions is crucial for success in drug design [6].

3. PROJECTS

In this thesis, projects are organized as follows: First, we discuss native ligand binding pose prediction (sampling power) using SQM-based SF and inhibitor design of Kinase target (see attachment **A, E**). Next, binding affinity estimated by SQM-based SF (scoring power) will be discussed and the importance of water upon binding will be described. (Attachment **B**, **C** and **E**). Finally, we describe the impressive virtual screening performance (screening power) of SQM-based scoring function on the HSP90 protein (Attachment **D** and **E**).

3.1 Sampling Power and Native Binding Mode Analysis

Molecular Docking represents an important step in the structure-based drug design and structure-based virtual screening. Docking predicts the binding poses (sampling power) and rank the binding affinities (scoring power) of a ligand within the binding side of the specific target [34], [154]–[159]. The main task is thus prediction of native ligand pose in the protein-ligand (P-L) complexes [51]. The success of classic SFs has been reported in many studies so far [34], [133], [160], [161]. However, classical SFs have some problems to find the native ligand pose in the challenging P-L complexes (such as metalloproteins, halogenated ligands, inorganic ligands etc.) [34]. Identifying native pose is still challenging for a single SF [34], [162], [163]. Each SF has a different approach and faces different problems [164]. To find native pose correctly, quantum mechanics QM-based approaches can be used [165]. QM approaches were pioneered first by Merz et al. [140]. The main advantage of QM methods represents the fact that they cover quantum effects (e.g. charge transfer and σ-hole binding) which play an important role in P-L interactions. Due to the size of P-L complexes, mostly semi-empirical (SQM) methods were applied (such as AM1[166], PM3[167], PM6[168], PM7[169] and DFTB3[170]). None of these methods are directly suitable for the investigation of noncovalent complexes[165]. Therefore, dispersions, electrostatic and σ-hole corrections must be included. The advanced D3H4X [171] correction terms have been used in combination with the PM6 method (PM6-D3H4X). PM6 method can be combined with the MOZYME linear scaling algorithm implemented in MOPAC [172] and possible to be apply to extended P-L complexes up to several thousand atoms. P-L complexes exist in a solvent under the physiologic conditions, which affects their structure and properties. Therefore, we used the conductor-like screening implicit solvent model (COSMO) [50]. The application of SQM/COSMO SF for a native pose identification on datasets containing diverse classes of P-L complexes was validated. SQM/COSMO SF was applied to four difficult systems [51] and 17 pharmaceutically relevant diverse complexes [44] which outperformed all classical SFs.

In this project the binding mode of the novel Imidazo[1,2-c]pyrimidin-5(6H)-one inhibitors for cyclin-dependent kinase 2 (CDK2) target have been studied. The synthesis of novel imidazo[1,2-c]pyrimidin-5(6H)-one scaffolds for CDK2 inhibition combined with docking, scoring and activity testing was studied. 26 substituents with aromatic moieties in position 8 (cf Figure 8) has been tested in vitro and they show inhibitions in CDK2. 2D structure-activity relationships (SAR) have confirmed that small substituents generally lead to single-digit micromolar IC_{50} values while showing decreased activities for bigger substituents. For the binding modes of the compounds Glide docking has been used. SQM/COSMO SF identified the favourable binding modes (See the details in the **Publication A**).

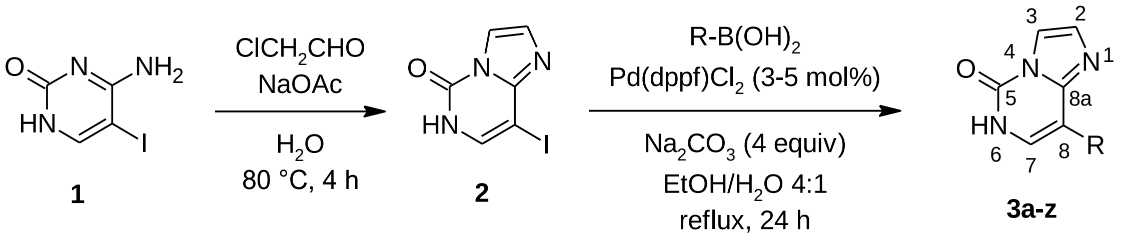


Figure 8 Synthesis of imidazo[1,2-c]pyrimidin-5(6H)-one core. *Reprinted from Publication A.*

Figure 9 demonstrates the binding modes of the novel Imidazo[1,2-c]pyrimidin-5(6H)-one inhibitors for CDK2/Cyclin E. Structure-activity relationships were rationalized using structure-based drug design approach. Glide SP method of Schrödinger [77] was used for docking the active CDK2 structure for (PDB code:3DDQ) [173]. Structures were rescored by SQM/COSMO setup (see methods in **Publication A**). Several binding modes (bm1-bm3) were detected. 20 docking poses for each compound was re-scored by SQM/COSMO SF. The SQM/COSMO results identified the preferred binding mode for further analysis. It is important to mention that docking followed by SQM/COSMO rescoring provide some insights into binding modes. The most active compounds had the Type I inhibitor binding mode (bm1a and bm1b). Standard (bm1) Type I binding mode involves 2 hinge region hydrogen bonds (E81:O…**3**:N(6)H and L83:NH…**3**:O(5)) (Figure 3.2A). Two sub-variants occurred: (1) bm1a, 8-aryl substituent is oriented towards the Lys33…Glu51 salt bridge (Figure 9A), and (2) bm1b, the 8-substituent is oriented towards the side chain of Lys33 and Asp145 (Figure 9B). In contrast, bm2 has reverse binding mode that keeps only one hinge-region hydrogen bonds being formed with Leu83 (L83:NH…**3**:O5 and L83:O…N(6)H) (Figure 9C). The last binding mode (bm3) keeps only one hinge-region hydrogen bond (Figure 9D). As a conclusion, the smaller substituents at position 8 (Figure 8) have micromolar potency. Docking results provided several binding modes and semi-empirical quantum mechanical rescoring identified the favourable binding modes. These results will be useful for further structure-based drug design and ligand optimization for the future.

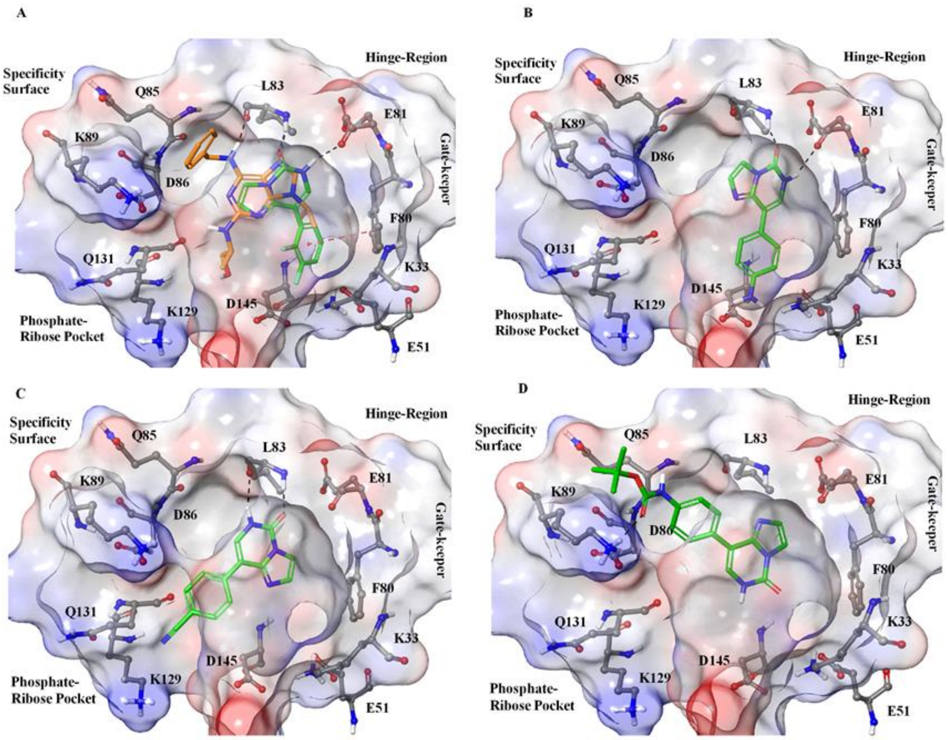
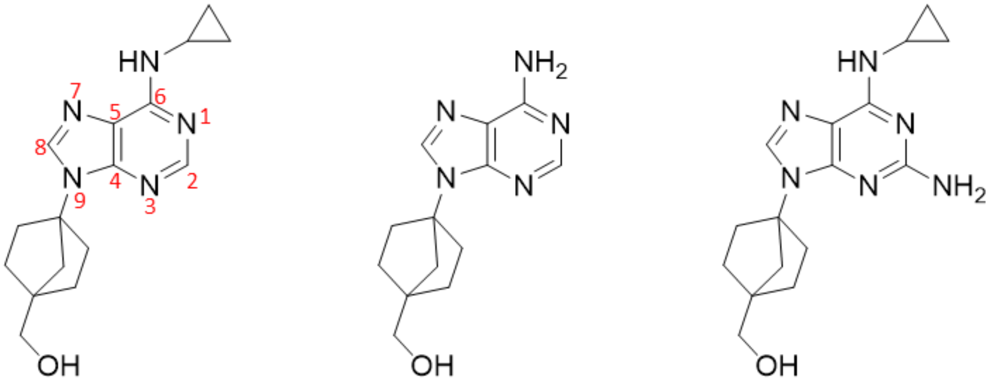

A
Specificity Surface
Q85
L83
Hinge-Region
E81
Gate-keeper
K89
D86
F80
Q131
K33
D145
K129
Phosphate-Ribose Pocket
E51
B
Specificity Surface
Q85
L83
Hinge-Region
E81
Gate-keeper
K89
D86
F80
Q131
K33
D145
K129
Phosphate-Ribose Pocket
E51
C
Specificity Surface
Q85
L83
Hinge-Region
E81
Gate-keeper
K89
D86
F80
Q131
K33
D145
K129
Phosphate-Ribose Pocket
E51
D
Specificity Surface
Q85
L83
Hinge-Region
E81
Gate-keeper
K89
D86
F80
Q131
K33
D145
K129
Phosphate-Ribose Pocket
E51

Figure 9 The binding mode of the imidazo[1,2-c]pyrimidin-5(6H)-one compounds in active CDK2/Cyclin E conformation. (A) The active compound **3j** (green, bm1a) compared with the crystal structure of roscovitine (PDB code:3DDQ). (B) The active compound **3m** (bm1b). (C) The least active compound **3n** (bm2). (D) The inactive compound **3s** (bm3). The ligands are shown in sticks, CDK2 as surface and important amino acid residues are shown in ball and sticks. Colouring (C:green-ligand/grey-CDK2, N: blue, O:red, F: cyan, H: white). The figure was prepared by Maestro (Schrodinger). *Reprinted from Publication A.*

3.2 Scoring (Ranking) Power and the Role of Explicit Water

In the previous part we demonstrated the ability of SQM/COSMO SF to predict native binding pose successfully. In this section, ranking power (scoring power) of SQM/COSMO will be examined. Accurate estimation of binding affinity is the most important step for virtual screening and hit-to-lead optimization. As mentioned before, proper description of noncovalent interactions represents the crucial part of the step. The ranking power (scoring power) of SQM/COSMO was investigated for a set of 10 inhibitors to carbonic anhydrase II. Comparing results that obtained with other standard scoring functions, SQM/COSMO showed better correlation. It should be noted that several standard SFs did not show any correlation [41]. Finally, two projects related to the ranking power of SQM/COSMO scoring function will be presented. The first one concerns the optimization of norbornyl-based carbocyclic nucleoside analogues for CDK2 and the second one describes the importance of explicit water in the active site and inhibition of CDK2 by new pyrazolo[1,5-α]pyrimidines.

In the first project the discovery of norbornyl moiety as a novel structural motif for CDK2 inhibitors has been reported. The unique collection of compounds in the Institute of Organic Chemistry and Biochemistry (IOCB) proprietary library enabled us to explore new novel inhibitors. Over 1,000 compounds from IOCB database have been used for further analysis of virtual screening against CDK2. Three norbornane based compounds have been identified as hits for CDK2 inhibition (cf Figure 10). These compounds occupy ATP binding site of CDK2 as type I inhibitors.



1
IC_{50} = 5.7 μM

2
IC_{50} = 4.7 μM

3
IC_{50} = 12.6 μM

Figure 10 Three hits for CDK2 inhibition from the IOCB proprietary database. *Reprinted from Publication B.*

For further treatment we have added 36 substituents in position 2 and 37 substituents in position 6 combined them with the hydroxynorbornyl moiety in position 9 on a purine scaffold. We built the modifications on the purine core in the CDK2 active site and scored using SQM/COSMO approach. Finally, eight compounds with high scores were selected for synthesis. Furthermore, another eight compounds with the cyclohexyl substituent in position 9 were suggested to explore the importance of the hydroxynorbornyl moiety. The aim was not only to discover a novel structural motif but also to understand the importance of the hydroxynorbornyl moiety (see the details in the **Publication B**)

The ligands with the norbornyl substituents in position 9 had micro-sub micromolar activities. The ligands with the cyclohexyl substituents in position 9 were active only in case of trans-1,4-cyclohexanediamine in position 2. The strongest affinity of 0.19 μM was observed for the compound 18a (cf. Figure 11a) having trans-1,4-cyclohexanediamine at position 2, phenylsulfonamidepyrrolidin at position 6 and norbornyl at position 9. For all the compounds the standard binding mode was identified in the purine core with 2 hinge-region hydrogen bonds (L83: NH…N(7) and L83:O…N(6)H) similar to roscovitine. The docking results suggested four types of orientations of the phenylsulfonamidepyrrolidine substituent in position 6 across the compound series. We built all the possibilities into the strongest active compound 18a to yield quantitative agreement with the activity of the compounds using SQM/COSMO SF. The interactions can be seen in Figure 11a with Lys89 i) via sulphonamide oxygen (Figure 11b) ii) via pyrrolidine nitrogen and Lys89 (Figure 11c), iii) shifted without interaction (Figure 11d) and iv) without interaction with 5-member ring slightly rotated (Figure 11e). (See the details in the **Publication B**).

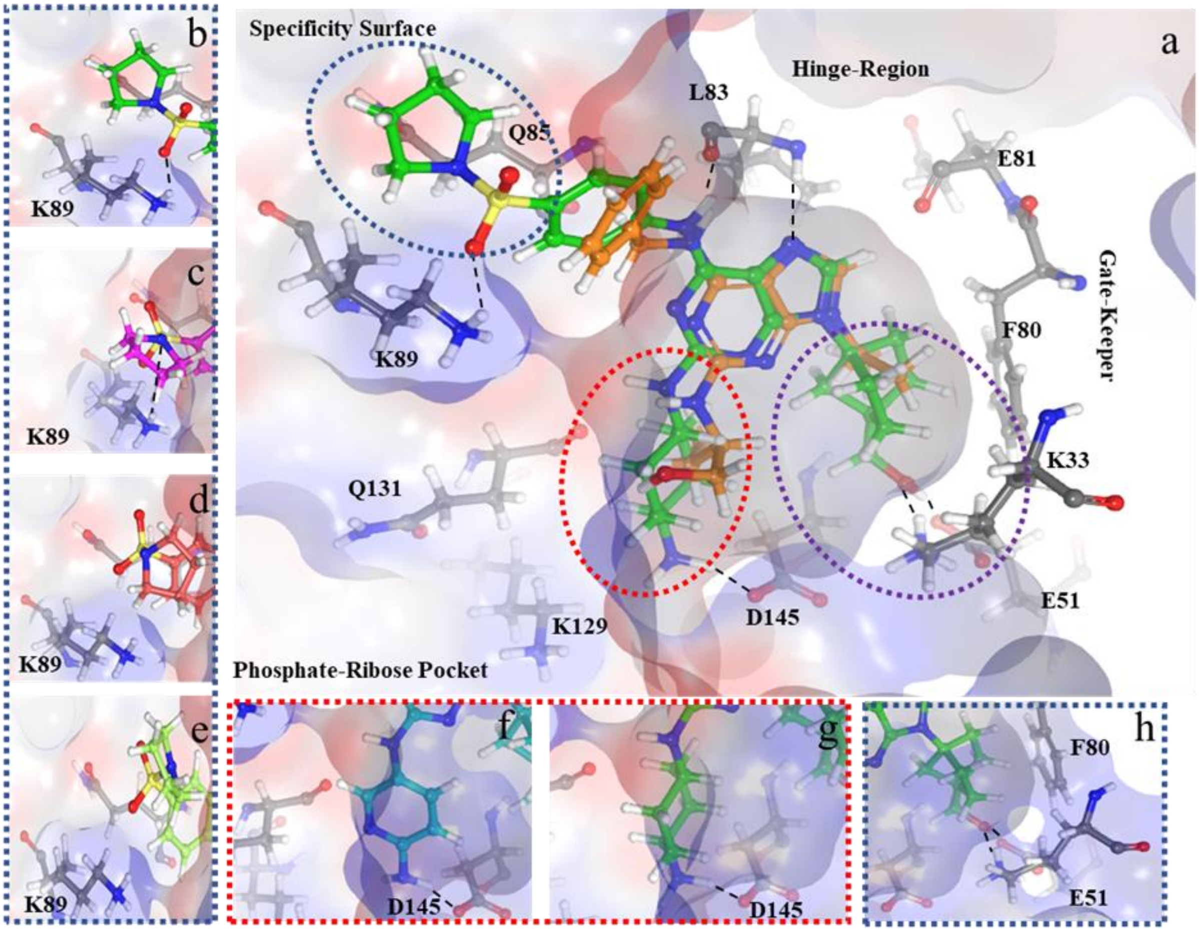

b
K89
c
K89
d
K89
e
K89
Specificity Surface
a
Hinge-Region
L83
Q85
E81
Gate-Keeper
F80
K89
K33
Q131
E51
D145
K129
Phosphate-Ribose Pocket
f
D145
g
D145
h
F80
E51

Figure 11 a) The binding mode of 18a (green sticks of carbon atoms) in CDK2 compared with the crystal structure of roscovitine with hydrogens added (orange sticks for carbon atoms; PDB code: 3DDQ). Important CDK2 residues are presented as ball and sticks. Colours of atoms (C:green/orange for the ligand and grey for CDK2 residues, N:blue, O:red, S:yellow, and H:white). Zoom into position 6 for 18a with different orientations (see text above): **b)** i), **c)** ii), **d)** iii), **e)** iv), **f)** Zoom into position 2 for 27a, **g)** Zoom into position 2 for 18a, **h)** Zoom into position of 9 for 18a. The figure was prepared with Maestro (Schrodinger). *Reprinted from Publication B.*

As a result, all compounds having hydroxymethylnorbornyl substituent in position 9 had sub-micromolar potencies. The compounds with the cyclohexyl in position 9 were active in case of trans-1,4-cyclohexanediamine in position 2 while other substituents in position 2 were inactive. Using docking and SQM/COSMO SF, we can separate the actives from the inactives (cf. Figure 12). Overall, the hydroxymethyl norbornyl substituent opens a new perspective for the future structure-based drug design progress.

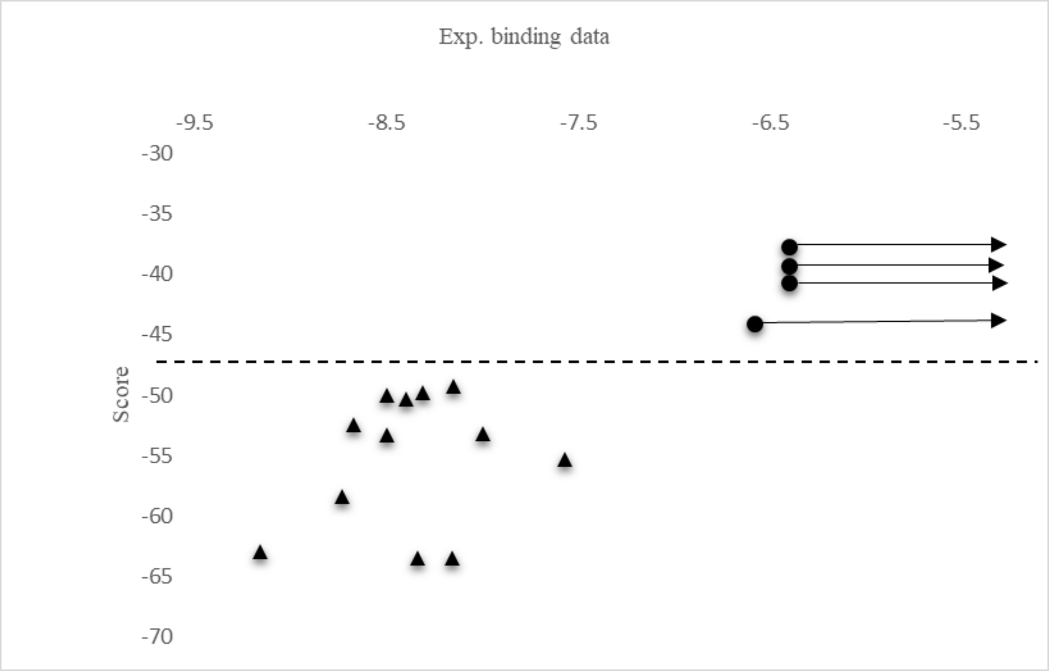


Figure 12 PM6-D3H4X/COSMO score plotted against experimental binding free energies expressed as RT*ln(IC_{50}/2), all in kcal/mol. Triangles are active compounds while circles are inactives. *Reprinted from Publication B.*

In the second project we present an explicit treatment of active site water and inhibition of CDK2 by new pyrazolo[1,5-α]pyrimidines. 21 new inhibitors of CDK2 with the pyrazolo[1,5-α]pyrimidine cores have been prepared. All inhibitors were scored by SQM/COSMO SF. COSMO implicit solvent model failed here to produce a correlation with the experimental binding affinities (R^2 = 0.49). After considering the explicit active-site waters in conjunction with SQM/COSMO significant improvement was observed (R^2 = 0.68). These results show that thermodynamic properties and solvation patterns around the ligands play an important role in drug design. Three types of calculations, i.e. molecular dynamics (MD), WaterMap [174] and SQM/COSMO scoring has been carried out.

The crystal structure of CDK2/1l was investigated by the WaterMap program to assess the dynamics and thermodynamics of water molecules. Six active-site waters (W77, W206, W194, W147, W224 and W130) were confirmed. Location of six water molecules in CDK2 active site is shown in Figure 13. The binding free energies of all six water molecules are thermodynamically 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 (see details in the **Publication C**).

Table 2 Thermodynamic characteristics of water molecules in CDK2/1l (ΔG: free energy, ΔH: enthalpy, -TΔS: entropy; all in kcal/mol) calculated by WaterMap. *Adapted from Publication C.*

Water	ΔG	ΔH	-TΔS
W77	1.2	-3.2	4.4
W206	4.4	0.5	3.9
W194	2.4	-1.6	4.0
W147	2.4	-2.0	4.5
W224	0.8	-4.4	5.1
W130	6.2	2.1	4.1

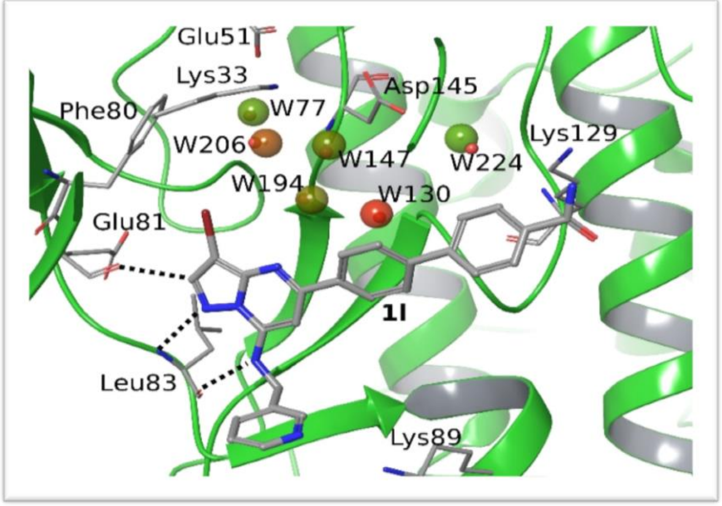


Figure 13 Structural water molecules in the crystallographic complex of CDK2/1l. *Reprinted from Publication C.*

3.3 Screening Power and the Virtual Screening Performance

Structure-based virtual screening can be considered successful if the scoring function yields reliable P-L complex geometry and determines reliable affinity. All these requirements are difficult to meet with a single SF. The SQM/COSMO SF was, however successful in both scenarios (see below) [52]. For this purpose, HSP90 protein and around 5000 ligands and decoys obtained from DUD-E [175] database were considered. The HSP90 (heat shock protein 90) is a protein that stabilizes various growth receptors[176] and some signalling molecules[177] required for the survival of cancer cells. Since the experimental structure data was not available, a different approach was selected for generating binding poses. There is a high number of the local minimum for P-L complexes and it would be impractical to search the whole landscape at the SQM level. DFs were used to generate a large number of poses to increase the possibility of finding the native binding pose. To obtain reliable native pose with a single DF is impractical in many cases. For this reason we have used nine different docking functions : Glide(SP,XP) [77], AutoDock4 [76], AutoDock Vina [81], Smina[82], and GOLD software[80], [178], [179] using ASP[180], GoldScore[178][80], ChemPLP[181] and ChemScore[180]. As mentioned earlier, our priority is to find the native binding poses. The generated poses were re-scored in the next step by means of SQM/COSMO SF. However, it is impractical to apply SQM/COSMO calculations for hundreds of thousands of structures. SQM SF can be applied up to a thousand P-L complexes and therefore we should use the method effectively. To be effective we have introduced SQM-based virtual screening frame. In this frame the complexes obtained by DFs were optimized at the molecular mechanics level. This optimization eliminated redundant poses and produced high-quality structures for further SQM/COSMO calculations. Two different AMBER optimizations were performed. The first one is the classical AMBER geometry optimization using the AMBER biomolecular simulation package (MM_N) and the other is the advanced AMBER optimization where bond lengths and bond angles are taken from the PM6 optimization calculations (MM_A) (see details in **Publication D** and **E**).

In order to understand the success of the method used, the capacity to distinguish the actives from the inactives was examined. Enrichment Factor, pROC and ROC AUC values have been used to determine the power of each SF. The performance of nine standardly used SFs, AMBER SF and SQM SFs were compared [52].

Docking results are visualised using a novel Post Dock tool[182] implemented in MOE [183] software in the Figure 16. For these results of 1UYF[184] crystal ligand have been used for the comparison. The reference ligand localization in the binding site is shown in green with a stick model. The poses from the DFs are shown in the surface model. Docking energies were represented by a transparency scale. When the transparency increases the score gets worse. The colours from the tinted yellow to the faint blue represent the RMSD values. As the colour goes blue, the RMSD gets worse. According to the results, some DFs couldn't find any native binding pose (such as B, H and I) whereas, some find native poses but couldn't score them correctly. It can be clearly seen that Glide SP provided binding modes very close to the crystal pose (i.e. Figure 16 E), however, provided the correct pose with worse score due to the transparency. In addition to all these, SQM SF provided less transparent colour which means the binding mode totally matches with the crystal pose with better score. All these results show that the SQM SF provides better and more reliable results than the other standard SFs.

Table 3 clearly shows that standardly used SFs have a poor performance. Consider that random value is 50, many SFs perform worse than random AUC ROC (%). However, when the SQM_2 SF was applied, the performance increases tremendously. Six out of nine SFs reach 90 AUC ROC (%) values and above. Considering the huge leap between the single SFs and SQM, standard SFs provide native poses but score poorly. SQM includes the quantum effects and score with a high performance. The most important thing to remember here, using the PM6 parameters in the MM_A treatment improves the geometry of the ligand, which leads to better geometries of the P-L complex and, consequently, to higher enrichment. The consideration of all poses provided impressive enrichment and AUC performance. In other words, SQM_2 SF shows three times better pROC performance and inserts 7 times more actives into selected dataset than the standard SF (i.e. Glide SP).

Table 3 The ROC Enrichment Factors EF, AUC and pROC AUC obtained for single-docking-function (DF) and SQM_2//DF (combination of scoring and docking; the P-L structures were optimized with the MM_A method) and SQM_2//ALL (P-L geometries from all SFs were optimized with MM_A method) docking functions. *Adapted from Publication D, E.*

DF	EF1	AUC ROC (%)	AUC pROC
AD4	1	49	0.383
SQM_2//AD4	40	91	2.104
VINA	0	30	0.192
SQM_2//VINA	42	93	2.052
SMINA	0	34	0.224
SQM_2//SMINA	37	93	1.997
GlideSP	7	75	0.880
SQM_2//SP	34	81	1.670
GlideXP	4	71	0.730
SQM_2//XP	32	85	1.710
ASP	3	76	0.787
SQM_2//ASP	31	93	1.877
Gscr	0	60	0.488
SQM_2//Gscr	44	97	2.329
Cscr	0	34	0.270
SQM_2//Cscr	29	89	1.757
PLP	1	51	0.383
SQM_2//PLP	31	95	2.096
SQM_2//ALL	47	98	2.477

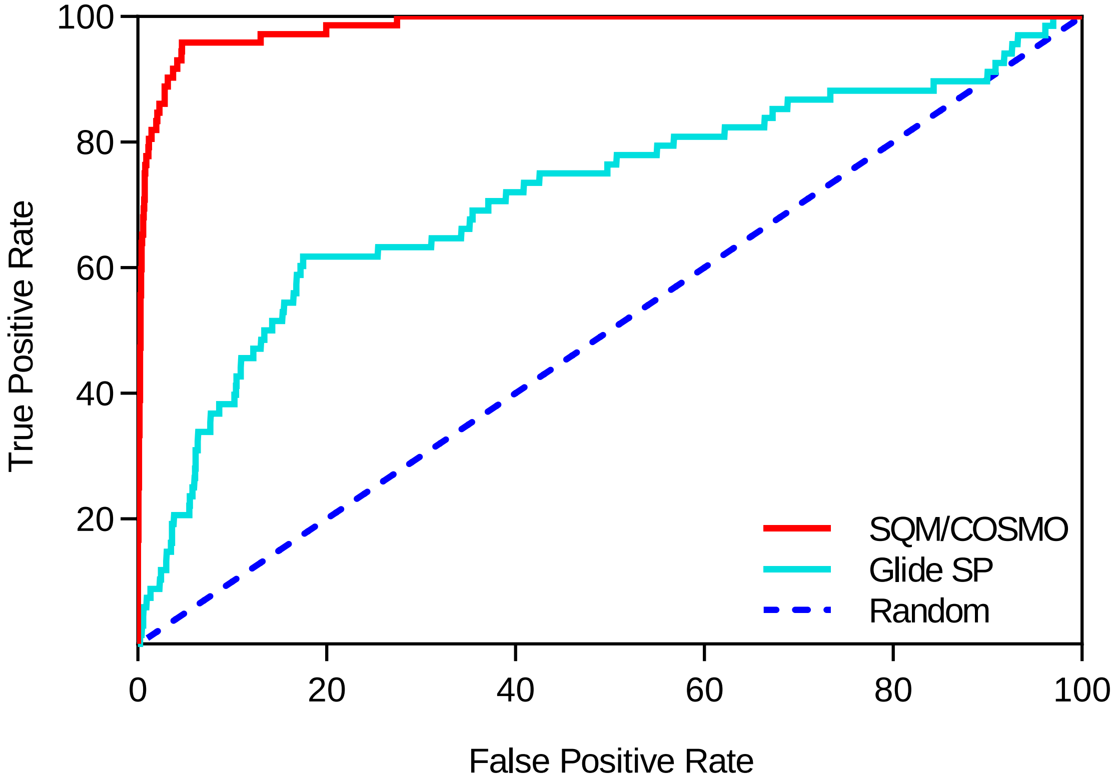


Figure 14 ROC curve for the comparison of the overall performance of the best standard SF (Glide SP) in light blue, SQM SF (P-L complexes from MM$_A$) in red and random value in dark blue. *Reprinted from Publication E.*

Figure 14 represents the ROC (receiver operator characteristics) curve of the results. In the virtual screening mainly, a huge number of compounds need to be considered. The screening performance of SFs may be defined as distinguish known actives from inactives (decoys). Enrichment Factor and receiver operator characteristic (ROC) plots [185] was used for the evaluation. In another word, ROC tells us how good our model is to distinguish between active and inactive. The score of all ligand poses binding to the respective target protein was calculated, ranked and plotted. EF defined as

$$\mathrm{EF_{subset}} = (\mathrm{ligand_{selected}}/\mathrm{N_{subset}})/(\mathrm{ligand_{total}}/\mathrm{N_{total}}) \qquad (8)$$

EF_{subset} provides information about the number of the actives among the inactives in the given subset in comparison with a random selection [186]. The ROC curves were plotted as (100% − Sp%) (i.e. % of selected decoys) versus Se% (i.e. % of selected active compounds) [187]. The AUC is the area under the ROC curve. This result gives us an idea of how well the model performs. The AUC (%) value is high when the SF performs sufficiently.

The results were also supported by the AUC pROC (cf table 3) values, which focus on early enrichment [188], [189]. AUC pROC values for random enrichment were determined as follows:

$$\lim_{a\to 0}\int_a^1 (-\log_{10} X)dX = \frac{-1}{\log 10}\lim_{a\to 0}\int_a^1 (\log X)dX$$
$$= 0.434 \lim_{a\to 0}\{X - X\log X\}\Big|_a^1 = 0.434 \qquad (9)$$

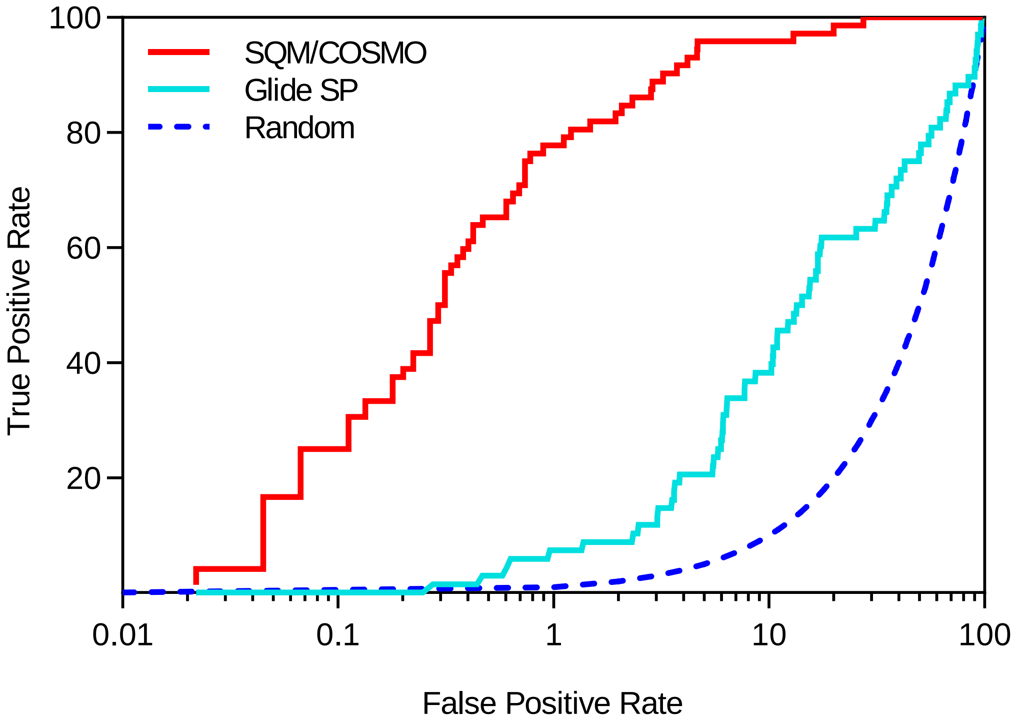


Figure 15 pROC curve for the comparison their early enrichment performance of the best standard SF (Glide SP) in light blue, SQM SF (P-L complexes from MM_A) in red and random value in dark blue. *Reprinted from Publication E.*

Figure 15 demonstrates the pROC curve which provides information about the early stage performance. SQM SF has performed impressively in the early stage as well. Almost half of the active structures were found in 1% of whole dataset which means SQM SF shows three times better pROC performance than the standard SF (i.e. Glide SP).

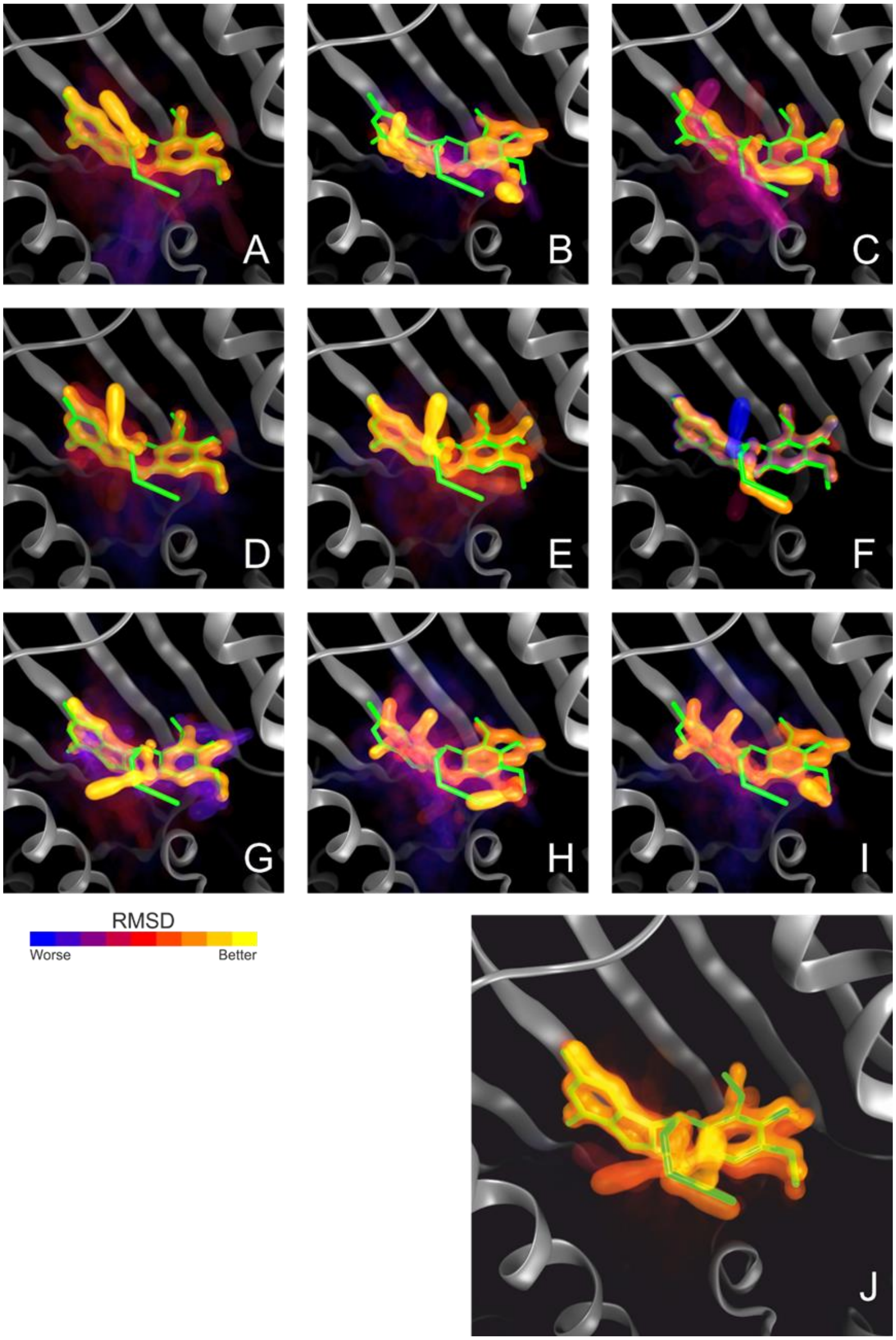

A
B
C
D
E
F
G
H
I
RMSD
Worse
Better
J

Figure 16 Individual representation of RMSD vs Energy docking/scoring results of 9 different DFs and the scoring result of SQM SF: **A** AD4 **B** ASP **C** CSCR **D** GSCR **E** Glide SP **F** Glide XP **G** PLP **H** SMINA **I** VINA and **J** represents SQM SF. *Reprinted from Publication D.*

4. CONCLUSION

In this thesis, the superior performance of SQM SF developed in our laboratory has been demonstrated. Sampling is the first and one of the most important steps in structure-based drug design. Several studies carried out in our laboratory demonstrated the superior sampling power of the SQM SF. Standardly used docking functions frequently provide correct native binding pose, but fail in scoring. According to our first project (see attached A), results show that docking software provide several binding modes but only SQM SF identifies favorable binding modes. For Kinases, it is crucial to identify the type of binding mode for further structure-based studies. The other important factor is the ranking success of the SF. Accurate estimation of binding affinity is related with the ranking performance of the SF. Better ranking performance enables better hit-to-lead optimization. The ranking success of SQM SF is demonstrated in attachments B and C. In the former case, Norbornyl-based carbocyclic nucleoside analogs were studied as CDK2 inhibitors. The hydroxymethyl norbornyl substituent was identified as the most efficient inhibitor and opens a new perspective for the future structure-based drug design progress for CDK2 target. In the latter case, the importance of explicit water in the active site and inhibition of CDK2 by new pyrazolo[1,5-α]pyrimidines was studied. The most important result of this study is to consider crystal waters explicitly and include these results in the score. After considering the active-site waters in conjunction with SQM/COSMO, significant improvement for ranking was observed ($R^2 = 0.68$). SQM SF performed successfully in both sampling and ranking. Finally, the success of virtual screening (screening power) has been shown in Attachment D. HSP90 protein and around 5000 compounds, both actives and inactives, were obtained from DUD-E database. Our investigations found that a single SF fails to achieve both docking/scoring and ranking. If SQM SF is used on top of classical SF, impressive results were obtained close to the perfect case. As a final conclusion, the SQM/COSMO is proposed as an efficient tool in structure-based drug design and details can be found in attachment E.

5. REFERENCES

[1] M. M. R. Arkin and J. A. Wells, “Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream,” *Nature Reviews Drug Discovery*, vol. 3, no. 4. Nature Publishing Group, pp. 301–317, 2004.

[2] J. Samanen, “Similarities and differences in the discovery and use of biopharmaceuticals and small-molecule chemotherapeutics,” in *Introduction to Biological and Small Molecule Drug Research and Development: Theory and Case Studies*, Elsevier Inc., 2013, pp. 161–203.

[3] S. Myers and A. Baker, “Drug discovery - An operating model for a new era. Despite the advent of new science and technologies, drug developers will need to make radical changes in their operations if they are to remain competitive and innovative,” *Nature Biotechnology*, vol. 19, no. 8. pp. 727–730, 2001.

[4] F. Ghasemi, A. Fassihi, H. Pérez-Sánchez, and A. Mehri Dehnavi, “The role of different sampling methods in improving biological activity prediction using deep belief network,” *J. Comput. Chem.*, vol. 38, no. 4, pp. 195–203, Feb. 2017.

[5] K. M. Merz, D. Ringe, and C. H. Reynolds, *Drug Design: Structure- and Ligand-Based Approaches*. Cambridge University Press, 2010.

[6] K. Stromgaard, P. Krogsgaard-Larsen, and U. Madsen, Eds., *Textbook of Drug Design and Discovery, Fifth Edition*. CRC Press, 2016.

[7] C. M. Song, S. J. Lim, and J. C. Tong, “Recent advances in computer-aided drug design,” *Briefings in Bioinformatics*, vol. 10, no. 5. pp. 579–591, 2009.

[8] D. E. Clark, “What has computer-aided molecular design ever done for drug discovery?,” *Expert Opinion on Drug Discovery*, vol. 1, no. 2. pp. 103–110, Jul-2006.

[9] W. L. Jorgensen, “The Many Roles of Computation in Drug Discovery,” *Science*, vol. 303, no. 5665. pp. 1813–1818, 19-Mar-2004.

[10] A. B. Richon, “An early history of the molecular modeling industry,” *Drug Discovery Today*, vol. 13, no. 15–16. pp. 659–664, Aug-2008.

[11] M. Zheng, X. Liu, Y. Xu, H. Li, C. Luo, and H. Jiang, “Computational methods for drug design and discovery: Focus on China,” *Trends in Pharmacological Sciences*, vol. 34, no. 10. pp. 549–559, Oct-2013.

[12] “RCSB PDB: Homepage.” [Online]. Available: https://www.rcsb.org/. [Accessed: 11-Jan-2020].

[13] G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, “Computational methods in drug discovery,” *Pharmacological Reviews*, vol. 66, no. 1. pp. 334–395, 2014.

[14] J. H. Drie, “Computer-aided drug design: The next 20 years,” *J. Comput. Aided. Mol. Des.*, vol. 21, no. 10–11, pp. 591–601, 2007.

[15] R. Vijayakrishnan, “Structure-based drug design and modern medicine,” *Journal of Postgraduate Medicine*, vol. 55, no. 4. pp. 301–304, 2009.

[16] T. Talele, S. Khedkar, and A. Rigby, “Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic,” *Curr. Top. Med. Chem.*, vol. 10, no. 1, pp. 127–141, Jan. 2010.

[17] N. Claude Cohen, “Structure-based drug design and the discovery of Aliskiren (Tekturna®): Perseverance and creativity to overcome a R&D pipeline challenge,” *Chem. Biol. Drug Des.*, vol. 70, no. 6, pp. 557–565, Nov. 2007.

[18] I. J. Enyedy and W. J. Egan, “Can we use docking and scoring for hit-to-lead optimization?,” *J. Comput. Aided. Mol. Des.*, vol. 22, no. 3–4, pp. 161–168, 2008.

[19] M. R. Lockett *et al.*, “The binding of benzoarylsulfonamide ligands to human carbonic anhydrase is insensitive to formal fluorination of the ligand,” *Angew. Chemie - Int. Ed.*, vol. 52, no. 30, pp. 7714–7717, Jul. 2013.

[20] B. Breiten *et al.*, “Water networks contribute to enthalpy/entropy compensation in protein-ligand binding,” *J. Am. Chem. Soc.*, vol. 135, no. 41, pp. 15579–15584, 2013.

[21] G. C. Pimentel and A. L. McClellan, “HYDROGEN BONDING,” *Annu. Rev. Phys. Chem.*, vol. 22, pp. 347–385, 1971.

[22] S. K. Burley and G. A. Petsko, "Aromatic-aromatic interaction: A mechanism of protein structure stabilization," *Science (80-.).*, vol. 229, no. 4708, pp. 23–28, 1985.

[23] C. A. Hunter, J. Singh, and J. M. Thornton, "π-π interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins," *J. Mol. Biol.*, vol. 218, no. 4, pp. 837–846, Apr. 1991.

[24] C. Bissantz, B. Kuhn, and M. Stahl, "A medicinal chemist's guide to molecular interactions," *Journal of Medicinal Chemistry*, vol. 53, no. 14. pp. 5061–5084, 2010.

[25] K. E. Riley and P. Hobza, "On the importance and origin of aromatic interactions in chemistry and biodisciplines," *Acc. Chem. Res.*, vol. 46, no. 4, pp. 927–936, 2013.

[26] R. Wilcken, M. O. Zimmermann, A. Lange, A. C. Joerger, and F. M. Boeckler, "Principles and applications of halogen bonding in medicinal chemistry and chemical biology," *Journal of Medicinal Chemistry*, vol. 56, no. 4. pp. 1363–1388, 2013.

[27] E. Fischer, "Einfluss der Configuration auf die Wirkung der Enzyme," *Berichte der Dtsch. Chem. Gesellschaft*, vol. 27, no. 3, pp. 2985–2993, Oct. 1894.

[28] D. E. Koshland, "The Key–Lock Theory and the Induced Fit Theory," *Angewandte Chemie International Edition in English*, vol. 33, no. 23–24. pp. 2375–2378, 03-Jan-1995.

[29] D. E. Koshland, "Application of a Theory of Enzyme Specificity to Protein Synthesis," *Proc. Natl. Acad. Sci.*, vol. 44, no. 2, pp. 98–104, Feb. 1958.

[30] C. Yung-Chi and W. H. Prusoff, "Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction," *Biochem. Pharmacol.*, vol. 22, no. 23, pp. 3099–3108, Dec. 1973.

[31] D. G. Fedorov and K. Kitaura, "Extending the power of quantum chemistry to large systems with the fragment molecular orbital method," *Journal of Physical Chemistry A*, vol. 111, no. 30. pp. 6904–6914, 2007.

[32] A. Cavalli, P. Carloni, and M. Recanatini, "Target-related applications of first principles quantum chemical methods in drug design," *Chemical Reviews*, vol. 106, no. 9. pp. 3497–

3519, 2006.

[33] O. A. Arodola and M. E. S. Soliman, “Quantum mechanics implementation in drug-design workflows: Does it really help?,” *Drug Design, Development and Therapy*, vol. 11. pp. 2551–2564, 2017.

[34] Z. Wang *et al.*, “Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power,” *Phys. Chem. Chem. Phys.*, vol. 18, no. 18, pp. 12964–12975, 2016.

[35] M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, and J. Fanfrlík, “The semiempirical quantum mechanical scoring function for in silico drug Design,” *ChemPlusChem*, vol. 78, no. 9. pp. 921–931, Sep-2013.

[36] P. S. Brahmkshatriya *et al.*, “Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo[1,5-a]pyrimidines,” *Curr. Comput. Aided-Drug Des.*, vol. 9, no. 1, pp. 118–129, Feb. 2013.

[37] J. Fanfrlík *et al.*, “The Effect of Halogen-to-Hydrogen Bond Substitution on Human Aldose Reductase Inhibition,” *ACS Chem. Biol.*, vol. 10, no. 7, pp. 1637–1642, Jul. 2015.

[38] J. Řezáč, “Cuby: An integrative framework for computational chemistry,” *J. Comput. Chem.*, vol. 37, no. 13, pp. 1230–1237, May 2016.

[39] A. Pecina *et al.*, “The SQM/COSMO filter: Reliable native pose identification based on the quantum-mechanical description of protein-ligand interactions and implicit COSMO solvation,” *Chem. Commun.*, vol. 52, no. 16, pp. 3312–3315, Feb. 2016.

[40] M. Hylsová *et al.*, “Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by new pyrazolo[1,5-a]pyrimidines,” *Eur. J. Med. Chem.*, vol. 126, pp. 1118–1128, 2017.

[41] A. Pecina *et al.*, “Ranking Power of the SQM/COSMO Scoring Function on Carbonic Anhydrase II–Inhibitor Complexes,” *ChemPhysChem*, vol. 19, no. 7, pp. 873–879, Apr. 2018.

[42] H. Ajani *et al.*, “Imidazo[1,2-c]pyrimidin-5(6H)-one as a novel core of cyclin-dependent

kinase 2 inhibitors: Synthesis, activity measurement, docking, and quantum mechanical scoring," *J. Mol. Recognit.*, vol. 31, no. 9, p. e2720, Sep. 2018.

[43] M. Nekardová *et al.*, "Structural Basis of the Interaction of Cyclin-Dependent Kinase 2 with Roscovitine and Its Analogues Having Bioisosteric Central Heterocycles," *ChemPhysChem*, vol. 18, no. 7, pp. 785–795, Apr. 2017.

[44] H. Ajani *et al.*, "Superior Performance of the SQM/COSMO Scoring Functions in Native Pose Recognition of Diverse Protein-Ligand Complexes in Cognate Docking," *ACS Omega*, vol. 2, no. 7, pp. 4022–4029, Jul. 2017.

[45] T. Lengauer and M. Rarey, "Computational methods for biomolecular docking," *Curr. Opin. Struct. Biol.*, vol. 6, no. 3, pp. 402–406, 1996.

[46] D. A. Gschwend, A. C. Good, and I. D. Kuntz, "Molecular docking towards drug discovery," in *Journal of Molecular Recognition*, 1996, vol. 9, no. 2, pp. 175–186.

[47] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: Methods and applications," *Nature Reviews Drug Discovery*, vol. 3, no. 11. pp. 935–949, Nov-2004.

[48] V. Mohan, A. Gibbs, M. Cummings, E. Jaeger, and R. DesJarlais, "Docking: Successes and Challenges," *Curr. Pharm. Des.*, vol. 11, no. 3, pp. 323–333, Feb. 2005.

[49] C. Shen *et al.*, "Comprehensive assessment of nine docking programs on type II kinase inhibitors: prediction accuracy of sampling power, scoring power and screening power," *Brief. Bioinform.*, vol. 00, no. October, pp. 1–16, 2018.

[50] A. Klamt and 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*, vol. 12, no. 5, pp. 799–805, 1993.

[51] A. Pecina *et al.*, "SQM/COSMO Scoring Function at the DFTB3-D3H4 Level: Unique Identification of Native Protein-Ligand Poses," *J. Chem. Inf. Model.*, vol. 57, no. 2, pp. 127–132, Feb. 2017.

[52] S. M. Eyrilmez, C. Köprülüoğlu, J. Řezáč, and P. Hobza, "Impressive Enrichment of

Semiempirical Quantum Mechanics-Based Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys," *ChemPhysChem*, vol. 20, no. 21, pp. 2759–2766, Nov. 2019.

[53] M. Congreve, C. W. Murray, and T. L. Blundell, "Keynote review: Structural biology and drug discovery," *Drug Discovery Today*, vol. 10, no. 13. pp. 895–907, 01-Jul-2005.

[54] E. Lionta, G. Spyrou, D. Vassilatis, and Z. Cournia, "Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances," *Curr. Top. Med. Chem.*, vol. 14, no. 16, pp. 1923–1938, 2014.

[55] K. Lundstrom, "Genomics and drug discovery," *Future Medicinal Chemistry*, vol. 3, no. 15. pp. 1855–1858, Nov-2011.

[56] S. Bambini and R. Rappuoli, "The use of genomics in microbial vaccine development," *Drug Discovery Today*, vol. 14, no. 5–6. pp. 252–260, Mar-2009.

[57] J. Kirchmair, G. Wolber, C. Laggner, and T. Langer, "Comparative performance assessment of the conformational model generators omega and catalyst: A large-scale survey on the retrieval of protein-bound ligand conformations," *J. Chem. Inf. Model.*, vol. 46, no. 4, pp. 1848–1861, 2006.

[58] G. L. Warren, T. D. Do, B. P. Kelley, A. Nicholls, and S. D. Warren, "Essential considerations for using protein-ligand structures in drug discovery," *Drug Discovery Today*, vol. 17, no. 23–24. pp. 1270–1281, Dec-2012.

[59] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery," *Curr. Comput. Aided-Drug Des.*, vol. 7, no. 2, pp. 146–157, 2012.

[60] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. R. Corbeil, "Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go," in *British Journal of Pharmacology*, 2008, vol. 153, no. SUPPL. 1, pp. 7–26.

[61] S. L. Mcgovern, B. W. Genentech, and J. J. Irwin, "Hits, leads and artifacts from virtual and high throughput screening," *Mol. Informatics "Confronting Complexity,"* no. July 2015, pp. 1–13, 2002.

[62] D. Bailey and D. Brown, “High-throughput chemistry and structure-based design: Survival of the smartest,” *Drug Discovery Today*, vol. 6, no. 2. pp. 57–59, 01-Jan-2001.

[63] O. M. Becker *et al.*, “An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression,” *J. Med. Chem.*, vol. 49, no. 11, pp. 3116–3135, Jun. 2006.

[64] S. L. Warner *et al.*, “Identification of a lead small-molecule inhibitor of the Aurora kinases using a structure-assisted, fragment-based approach,” *Mol. Cancer Ther.*, vol. 5, no. 7, pp. 1764–1773, Jul. 2006.

[65] B. Budzik *et al.*, “Novel N-substituted benzimidazolones as potent, selective, CNS-penetrant, and orally active M1 mAChR agonists,” *ACS Med. Chem. Lett.*, vol. 1, no. 6, pp. 244–248, 2010.

[66] T. Cheng, Q. Li, Z. Zhou, Y. Wang, and S. H. Bryant, “Structure-based virtual screening for drug discovery: A problem-centric review,” *AAPS Journal*, vol. 14, no. 1. pp. 133–141, 2012.

[67] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, “Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments,” *J. Comput. Aided. Mol. Des.*, vol. 27, no. 3, pp. 221–234, 2013.

[68] W. R. Pitt, M. D. Calmiano, B. Kroeplien, R. D. Taylor, J. P. Turner, and M. A. King, “Structure-based virtual screening for novel ligands,” *Methods Mol. Biol.*, vol. 1008, pp. 501–519, 2013.

[69] H. Li, A. D. Robertson, and J. H. Jensen, “Very fast empirical prediction and rationalization of protein pK a values,” *Proteins Struct. Funct. Genet.*, vol. 61, no. 4, pp. 704–721, Oct. 2005.

[70] R. Anandakrishnan, B. Aguilar, and A. V Onufriev, “H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations,” *Nucleic Acids Res.*, vol. 40, no. W1, pp. W537-541, 2012.

[71] T. Ten Brink and T. E. Exner, “PKa based protonation states and microspecies for

protein-ligand docking," *J. Comput. Aided. Mol. Des.*, vol. 24, no. 11, pp. 935–942, 2010.

[72] "Protein Preparation Wizard, Maestro, Schtodinger LLC." .

[73] Schrödinger, "Maestro | Schrödinger," *Schrödinger Release 2018-1*. 2018.

[74] J. de Ruyck, G. Brysbaert, R. Blossey, and M. F. Lensink, "Molecular docking as a popular tool in drug design, an in silico travel," *Adv. Appl. Bioinforma. Chem.*, vol. 9, no. 1, pp. 1–11, 2016.

[75] N. S. Pagadala, K. Syed, and J. Tuszynski, "Software for molecular docking: a review," *Biophysical Reviews*, vol. 9, no. 2. pp. 91–102, 2017.

[76] G. M. Morris *et al.*, "Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *J. Comput. Chem.*, vol. 30, no. 16, pp. 2785–2791, Dec. 2009.

[77] R. A. Friesner *et al.*, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, vol. 47, no. 7, pp. 1739–1749, 2004.

[78] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," *J. Mol. Biol.*, vol. 261, no. 3, pp. 470–489, Aug. 1996.

[79] A. N. Jain, "Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine," *J. Med. Chem.*, vol. 46, no. 4, pp. 499–511, 2003.

[80] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *J. Mol. Biol.*, vol. 267, no. 3, pp. 727–748, Apr. 1997.

[81] O. Trott and A. J. Olson, "Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2010.

[82] D. R. Koes, M. P. Baumgartner, and C. J. Camacho, "Lessons learned in empirical scoring

with smina from the CSAR 2011 benchmarking exercise," *J. Chem. Inf. Model.*, vol. 53, no. 8, pp. 1893–1904, 2013.

[83] C. R. Corbeil, C. I. Williams, and P. Labute, "Variability in docking success rates due to dataset preparation," *Journal of Computer-Aided Molecular Design*, vol. 26, no. 6. pp. 775–786, 2012.

[84] I. Muegge and M. Rarey, "Small Molecule Docking and Scoring," in *Reviews in Computational Chemistry*, vol. 17, John Wiley & Sons, Ltd, 2001, pp. 1–60.

[85] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, "Protein-ligand docking: Current status and future challenges," *Proteins: Structure, Function and Genetics*, vol. 65, no. 1. pp. 15–26, 21-Jul-2006.

[86] A. Andricopulo, R. Guido, and G. Oliva, "Virtual Screening and Its Integration with Modern Drug Design Technologies," *Curr. Med. Chem.*, vol. 15, no. 1, pp. 37–46, 2008.

[87] Z. Zsoldos, D. Reid, A. Simon, S. B. Sadjad, and A. P. Johnson, "eHiTS: A new fast, exhaustive flexible ligand docking system," *J. Mol. Graph. Model.*, vol. 26, no. 1, pp. 198–212, Jul. 2007.

[88] A. T. Brint and P. Willett, "Algorithms for the Identification of Three-Dimensional Maximal Common Substructures," *J. Chem. Inf. Comput. Sci.*, vol. 27, no. 4, pp. 152–158, 1987.

[89] D. Fischer, R. Norel, H. Wolfson, and R. Nussinov, "Surface motifs by a computer vision technique: Searches, detection, and implications for protein–ligand recognition," *Proteins Struct. Funct. Bioinforma.*, vol. 16, no. 3, pp. 278–292, Jul. 1993.

[90] R. Norel, D. Fischer, H. J. Wolfson, and R. Nussinov, "Molecular surface recognition by a computer vision-based technique," *Protein Eng. Des. Sel.*, vol. 7, no. 1, pp. 39–46, Jan. 1994.

[91] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, "A geometric approach to macromolecule-ligand interactions," *J. Mol. Biol.*, vol. 161, no. 2, pp. 269–288, Oct. 1982.

[92] M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan, “FLOG: A system to select ‘quasi-flexible’ ligands complementary to a receptor of known three-dimensional structure,” *J. Comput. Aided. Mol. Des.*, vol. 8, no. 2, pp. 153–174, Apr. 1994.

[93] T. J. A. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz, “DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases,” *J. Comput. Aided. Mol. Des.*, vol. 15, no. 5, pp. 411–428, 2001.

[94] W. Welch, J. Ruppert, and A. N. Jain, “Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites,” *Chem. Biol.*, vol. 3, no. 6, pp. 449–462, 1996.

[95] A. Miranker and M. Karplus, “Functionality maps of binding sites: A multiple copy simultaneous search method,” *Proteins Struct. Funct. Bioinforma.*, vol. 11, no. 1, pp. 29–34, 1991.

[96] M. B. Eisen, D. C. Wiley, M. Karplus, and R. E. Hubbard, “HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site,” *Proteins Struct. Funct. Bioinforma.*, vol. 19, no. 3, pp. 199–221, Jul. 1994.

[97] H. J. Böhm, “LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads,” *J. Comput. Aided. Mol. Des.*, vol. 6, no. 6, pp. 593–606, Dec. 1992.

[98] B. Gorelik and A. Goldblum, “High quality binding modes in docking ligands to proteins,” *Proteins Struct. Funct. Genet.*, vol. 71, no. 3, pp. 1373–1386, Dec. 2008.

[99] D. S. Goodsell and A. J. Olson, “Automated docking of substrates to proteins by simulated annealing,” *Proteins Struct. Funct. Bioinforma.*, vol. 8, no. 3, pp. 195–202, 1990.

[100] T. N. Hart and R. J. Read, “A multiple-start Monte Carlo docking method,” *Proteins Struct. Funct. Bioinforma.*, vol. 13, no. 3, pp. 206–222, 1992.

[101] G. M. Morris *et al.*, “Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function,” *J. Comput. Chem.*, vol. 19, no. 14, pp. 1639–1662, Nov. 1998.

[102] C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, "Flexible ligand docking using a genetic algorithm," *J. Comput. Aided. Mol. Des.*, vol. 9, no. 2, pp. 113–130, Apr. 1995.

[103] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Comput. Oper. Res.*, vol. 13, no. 5, pp. 533–549, 1986.

[104] F. Glover, "Tabu Search—Part I," *ORSA J. Comput.*, vol. 1, no. 3, pp. 190–206, Aug. 1989.

[105] F. Glover, "Tabu Search—Part II," *ORSA J. Comput.*, vol. 2, no. 1, pp. 4–32, Feb. 1990.

[106] S. J. Weiner *et al.*, "A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins," *J. Am. Chem. Soc.*, vol. 106, no. 3, pp. 765–784, Feb. 1984.

[107] C. I. Bayly *et al.*, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, May 1995.

[108] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.

[109] P. Kollman, "Free Energy Calculations: Applications to Chemical and Biochemical Phenomena," *Chem. Rev.*, vol. 93, no. 7, pp. 2395–2417, Nov. 1993.

[110] J. Aqvist, V. B. Luzhkov, and B. O. Brandsdal, "Ligand binding affinities from MD simulations," *Acc. Chem. Res.*, vol. 35, no. 6, pp. 358–365, 2002.

[111] H. A. Carlson and W. L. Jorgensen, "An extended linear response method for determining free energies of hydration," *J. Phys. Chem.*, vol. 99, no. 26, pp. 10667–10673, Jun. 1995.

[112] J. E. Jones, "On the Determination of Molecular Fields. I. From the Variation of the Viscosity of a Gas with Temperature," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 106, no. 738, pp. 441–462, Oct. 1924.

[113] C. Bissantz, G. Folkers, and D. Rognan, "Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations," *J. Med. Chem.*, vol.

43, no. 25, pp. 4759–4767, Dec. 2000.

[114] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, “Improved protein-ligand docking using GOLD,” *Proteins Struct. Funct. Genet.*, vol. 52, no. 4, pp. 609–623, Sep. 2003.

[115] J. M. Briggs, T. J. Marrone, and J. A. McCammon, “Computational science: New horizons and relevance to pharmaceutical design,” *Trends in Cardiovascular Medicine*, vol. 6, no. 6. Elsevier Inc., pp. 198–203, 1996.

[116] P. K. Weiner and P. A. Kollman, “AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions,” *J. Comput. Chem.*, vol. 2, no. 3, pp. 287–303, 1981.

[117] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, “An all atom force field for simulations of proteins and nucleic acids,” *J. Comput. Chem.*, vol. 7, no. 2, pp. 230–252, Apr. 1986.

[118] W. F. van Gunsteren and H. J. C. Berendsen, “Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry,” *Angewandte Chemie International Edition in English*, vol. 29, no. 9. John Wiley & Sons, Ltd, pp. 992–1023, 01-Sep-1990.

[119] W. L. Jorgensen and J. Tirado-Rives, “The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin,” *J. Am. Chem. Soc.*, vol. 110, no. 6, pp. 1657–1666, Mar. 1988.

[120] H. J. Böhm, “Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs,” *J. Comput. Aided. Mol. Des.*, vol. 12, no. 4, p. 309, 1998.

[121] D. K. Gehlhaar *et al.*, “Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming,” *Chem. Biol.*, vol. 2, no. 5, pp. 317–324, 1995.

[122] G. M. Verkhivker *et al.*, “Deciphering common failures in molecular docking of ligand-

protein complexes," *J. Comput. Aided. Mol. Des.*, vol. 14, no. 8, pp. 731–751, Nov. 2000.

[123] A. N. Jain, "Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities," *J. Comput. Aided. Mol. Des.*, vol. 10, no. 5, pp. 427–440, 1996.

[124] R. D. Head, M. L. Smythe, T. I. Oprea, C. L. Waller, S. M. Green, and G. R. Marshall, "VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands," *J. Am. Chem. Soc.*, vol. 118, no. 16, pp. 3959–3969, Jan. 1996.

[125] D. K. Gehlhaar, K. E. Moerder, D. Zichi, C. J. Sherman, R. C. Ogden, and S. T. Freer, "De Novo Design of Enzyme Inhibitors by Monte Carlo Ligand Generation," *J. Med. Chem.*, vol. 38, no. 3, pp. 466–472, Feb. 1995.

[126] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, "Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes," *J. Comput. Aided. Mol. Des.*, vol. 11, no. 5, pp. 425–445, 1997.

[127] I. Muegge and Y. C. Martin, "A general and fast scoring function for protein-ligand interactions: A simplified potential approach," *J. Med. Chem.*, vol. 42, no. 5, pp. 791–804, Mar. 1999.

[128] J. B. O. Mitchell, R. A. Laskowski, A. Alex, and J. M. Thornton, "BLEEP - Potential of mean force describing protein-ligand interactions: I. Generating potential," *J. Comput. Chem.*, vol. 20, no. 11, pp. 1165–1176, Aug. 1999.

[129] A. V. Ishchenko and E. I. Shakhnovich, "SMall Molecule Growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions," *J. Med. Chem.*, vol. 45, no. 13, pp. 2770–2780, Jun. 2002.

[130] M. Feher, E. Deretey, and S. Roy, "BHB: A simple knowledge-based scoring function to improve the efficiency of database screening," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 4, pp. 1316–1327, Jul. 2003.

[131] G. Verkhivker, K. Appelt, S. T. Freer, and J. E. Villafranca, "Empirical free energy

calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity," *Protein Eng. Des. Sel.*, vol. 8, no. 7, pp. 677–691, Jul. 1995.

[132] A. Wallqvist, R. L. Jernigan, and D. G. Covell, "A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design," *Protein Sci.*, vol. 4, no. 9, pp. 1881–1903, 1995.

[133] H. Gohlke, M. Hendlich, and G. Klebe, "Knowledge-based scoring function to predict protein-ligand interactions," *J. Mol. Biol.*, vol. 295, no. 2, pp. 337–356, Jan. 2000.

[134] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov, "Principles of docking: An overview of search algorithms and a guide to scoring functions," *Proteins: Structure, Function and Genetics*, vol. 47, no. 4. pp. 409–443, 01-Jun-2002.

[135] P. S. Charifson, J. J. Corkery, M. A. Murcko, and W. P. Walters, "Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins," *J. Med. Chem.*, vol. 42, no. 25, pp. 5100–5109, Dec. 1999.

[136] M. Feher, "Consensus scoring for protein-ligand interactions," *Drug Discovery Today*, vol. 11, no. 9–10. pp. 421–428, May-2006.

[137] H. Gohlke and G. Klebe, "Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors," *Angewandte Chemie - International Edition*, vol. 41, no. 15. Wiley-VCH Verlag, pp. 2644–2676, 02-Aug-2002.

[138] D. Mucs and R. A. Bryce, "The application of quantum mechanics in structure-based drug design," *Expert Opinion on Drug Discovery*, vol. 8, no. 3. pp. 263–276, Mar-2013.

[139] M. Xu and M. A. Lill, "Induced fit docking, and the use of QM/MM methods in docking," *Drug Discovery Today: Technologies*, vol. 10, no. 3. Sep-2013.

[140] K. Raha and K. M. Merz, "A Quantum Mechanics-Based Scoring Function: Study of Zinc Ion-Mediated Ligand Binding," *J. Am. Chem. Soc.*, vol. 126, no. 4, pp. 1020–1021, Feb. 2004.

[141] K. Raha and K. M. Merz, "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.*, vol. 48, no. 14, pp. 4558–4575, Jul. 2005.

[142] P. Mikulskis, S. Genheden, K. Wichmann, and U. Ryde, "A semiempirical approach to ligand-binding affinities: Dependence on the hamiltonian and corrections," *J. Comput. Chem.*, vol. 33, no. 12, pp. 1179–1189, May 2012.

[143] H. S. Muddana and M. K. Gilson, "Calculation of host-guest binding affinities using a quantum-mechanical energy model," *J. Chem. Theory Comput.*, vol. 8, no. 6, pp. 2023–2033, Jun. 2012.

[144] J. Řezáč, J. Fanfrlík, D. Salahub, and P. Hobza, "Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes," *J. Chem. Theory Comput.*, vol. 5, no. 7, pp. 1749–1760, Jul. 2009.

[145] J. Řezáč and P. Hobza, "A halogen-bonding correction for the semiempirical PM6 method," *Chem. Phys. Lett.*, vol. 506, no. 4–6, pp. 286–289, Apr. 2011.

[146] J. Řezáč and P. Hobza, "Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods," *J. Chem. Theory Comput.*, vol. 8, no. 1, pp. 141–151, Jan. 2012.

[147] J. Fanfrlík *et al.*, "Modulation of aldose reductase inhibition by halogen bond tuning," *ACS Chem. Biol.*, vol. 8, no. 11, pp. 2484–2492, Nov. 2013.

[148] A. Pecina *et al.*, "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*, vol. 117, no. 50, pp. 16096–16104, Dec. 2013.

[149] J. Fanfrlík *et al.*, "Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic Schistosoma mansoni cysteine peptidase by vinyl sulfone inhibitors," *J. Phys. Chem. B*, vol. 117, no. 48, pp. 14973–14982, Dec. 2013.

[150] E. Kellenberger, J. Rodrigo, P. Muller, and D. Rognan, "Comparative evaluation of eight docking tools for docking and virtual screening accuracy," *Proteins Struct. Funct. Genet.*,

vol. 57, no. 2, pp. 225–242, Nov. 2004.

[151] D. R. Westhead, D. E. Clark, and C. W. Murray, "A comparison of heuristic search algorithms for molecular docking," *Journal of Computer-Aided Molecular Design*, vol. 11, no. 3. pp. 209–228, May-1997.

[152] B. D. Bursulaya, M. Totrov, R. Abagyan, and C. L. Brooks, "Comparative study of several algorithms for flexible ligand docking," *J. Comput. Aided. Mol. Des.*, vol. 17, no. 11, pp. 755–763, Nov. 2003.

[153] D. H. Williams *et al.*, "Toward the Semiquantitative Estimation of Binding Constants. Guides for Peptide-Peptide Binding in Aqueous Solution," *J. Am. Chem. Soc.*, vol. 113, no. 18, pp. 7020–7030, Aug. 1991.

[154] L. Santana Azevedo *et al.*, "Recent Progress of Molecular Docking Simulations Applied to Development of Drugs," *Curr. Bioinform.*, vol. 7, no. 4, pp. 352–365, Nov. 2013.

[155] E. Yuriev and P. A. Ramsland, "Latest developments in molecular docking: 2010-2011 in review," *J. Mol. Recognit.*, vol. 26, no. 5, pp. 215–239, 2013.

[156] E. Yuriev, J. Holien, and P. A. Ramsland, "Improvements, trends, and new ideas in molecular docking: 2012-2013 in review," *J. Mol. Recognit.*, vol. 28, no. 10, pp. 581–604, 2015.

[157] A. R. Leach, B. K. Shoichet, and C. E. Peishoff, "Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps," *Journal of Medicinal Chemistry*, vol. 49, no. 20. pp. 5851–5855, 2006.

[158] B. K. Shoichet, "Virtual screening of chemical libraries," *Nature*, vol. 432, no. 7019. pp. 862–865, 2004.

[159] I. D. Kuntz, "Structure-based strategies for drug design and discovery," *Science (80-.).*, vol. 257, no. 5073, pp. 1078–1082, 1992.

[160] A. N. Jain and A. Nicholls, "Recommendations for evaluation of computational methods," *J. Comput. Aided. Mol. Des.*, vol. 22, no. 3–4, pp. 133–139, 2008.

[161] M. Kontoyianni, L. M. McClellan, and G. S. Sokol, “Evaluation of Docking Performance: Comparative Data on Docking Algorithms,” *J. Med. Chem.*, vol. 47, no. 3, pp. 558–565, 2004.

[162] T. Cheng, X. Li, Y. Li, Z. Liu, and R. Wang, “Comparative assessment of scoring functions on a diverse test set,” *J. Chem. Inf. Model.*, vol. 49, no. 4, pp. 1079–1093, 2009.

[163] G. L. Warren *et al.*, “A critical assessment of docking programs and scoring functions,” *J. Med. Chem.*, vol. 49, no. 20, pp. 5912–5931, 2006.

[164] S. Z. Grinter and X. Zou, “Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design,” *Molecules*, vol. 19, no. 7, pp. 10150–10176, 2014.

[165] J. Řezáč and P. Hobza, “Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications,” *Chemical Reviews*, vol. 116, no. 9. pp. 5038–5071, 2016.

[166] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, “AM1: A New General Purpose Quantum Mechanical Molecular Model1,” *J. Am. Chem. Soc.*, vol. 107, no. 13, pp. 3902–3909, 1985.

[167] J. J. P. Stewart, “Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi,” *J. Comput. Chem.*, vol. 12, no. 3, pp. 320–341, Apr. 1991.

[168] J. J. P. Stewart, “Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements,” *J. Mol. Model.*, vol. 13, no. 12, pp. 1173–1213, 2007.

[169] J. J. P. Stewart, “Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters,” *J. Mol. Model.*, vol. 19, no. 1, pp. 1–32, 2013.

[170] M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, and M. Elstner, “Parameterization of the DFTB3 method for Br, Ca, Cl, F, I, K, and Na in organic and biological systems,” *J.*

Chem. Theory Comput., vol. 11, no. 1, pp. 332–342, 2015.

[171] J. Řezáč, K. E. Riley, and P. Hobza, "Benchmark calculations of noncovalent interactions of halogenated molecules," *J. Chem. Theory Comput.*, vol. 8, no. 11, pp. 4285–4292, 2012.

[172] James Stewart, "Stewart Computational Chemistry - MOPAC Home Page," 2007. [Online]. Available: http://openmopac.net/. [Accessed: 28-Feb-2020].

[173] K. Bettayeb *et al.*, "CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases," *Oncogene*, vol. 27, no. 44, pp. 5797–5807, 2008.

[174] "WaterMap | Schrödinger." [Online]. Available: https://www.schrodinger.com/watermap. [Accessed: 28-Feb-2020].

[175] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking," *J. Med. Chem.*, vol. 55, no. 14, pp. 6582–6594, 2012.

[176] A. Sawai *et al.*, "Inhibition of Hsp90 down-regulates mutant epidermal growth factor receptor (EGFR) expression and sensitizes EGFR mutant tumors to paclitaxel," *Cancer Res.*, vol. 68, no. 2, pp. 589–596, 2008.

[177] C. E. Stebbins, A. A. Russo, C. Schneider, N. Rosen, F. U. Hartl, and N. P. Pavletich, "Crystal structure of an Hsp90-geldanamycin complex: Targeting of a protein chaperone by an antitumor agent," *Cell*, vol. 89, no. 2, pp. 239–250, 1997.

[178] J. W. M. Nissink, C. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole, and R. Taylor, "A new test set for validating predictions of protein-ligand interaction," *Proteins Struct. Funct. Genet.*, vol. 49, no. 4, pp. 457–471, Dec. 2002.

[179] G. Jones, P. Willett, and R. C. Glen, "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation," *J. Mol. Biol.*, vol. 245, no. 1, pp. 43–53, 1995.

[180] W. T. M. Mooij and M. L. Verdonk, "General and targeted statistical potentials for protein-ligand interactions," *Proteins Struct. Funct. Genet.*, vol. 61, no. 2, pp. 272–287,

Aug. 2005.

[181] O. Korb, T. Stützle, and T. E. Exner, “Empirical scoring functions for advanced Protein-Ligand docking with PLANTS,” *J. Chem. Inf. Model.*, vol. 49, no. 1, pp. 84–96, 2009.

[182] E. A. Wiley and G. Deslongchamps, “PostDock: A novel visualization tool for the analysis of molecular docking,” *Comput. Vis. Sci.*, vol. 12, no. 1, pp. 1–7, 2009.

[183] C. C. Group, “Chemical Computing Group (CCG) | Computer-Aided Molecular Design,” *Chemical Computing Group*, 2020. [Online]. Available: https://www.chemcomp.com/. [Accessed: 28-Feb-2020].

[184] L. Wright *et al.*, “Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms,” *Chem. Biol.*, vol. 11, no. 6, pp. 775–785, Jun. 2004.

[185] A. Nicholls, “What do we know and when do we know it?,” *J. Comput. Aided. Mol. Des.*, vol. 22, no. 3–4, pp. 239–255, Mar. 2008.

[186] H. Fan, J. J. Irwin, B. M. Webb, G. Klebe, B. K. Shoichet, and A. Sali, “Molecular docking screens using comparative models of proteins,” *J. Chem. Inf. Model.*, vol. 49, no. 11, pp. 2512–2527, Nov. 2009.

[187] N. Huang, B. K. Shoichet, and J. J. Irwin, “Benchmarking sets for molecular docking,” *J. Med. Chem.*, vol. 49, no. 23, pp. 6789–6801, Nov. 2006.

[188] R. D. Clark and D. J. Webster-Clark, “Managing bias in ROC curves,” *J. Comput. Aided. Mol. Des.*, vol. 22, no. 3–4, pp. 141–146, Mar. 2008.

[189] S. M. Vogel, M. R. Bauer, and F. M. Boeckler, “DEKOIS: Demanding evaluation kits for objective in silico screening - A versatile tool for benchmarking docking programs and scoring functions,” *J. Chem. Inf. Model.*, vol. 51, no. 10, pp. 2650–2665, Oct. 2011.

6. PRESENTATION OF THE RESULTS

- “In Silico Drug Design Based on the Semi-empirical Quantum Mechanical Scoring Function” 2nd International BAU Drug Design Congress “**Poster Presentation with the 2nd best Poster Awards**” Istanbul, Turkey – **2014**
- “Schrödinger-14th Annual European User Meeting.” Frankfurt, Germany – **2014**
- “Quantum Mechanical Scoring Function For Virtual Screening” 3rd International BAU Drug Design Congress “**Oral Presentation**”, Istanbul, Turkey – **2015**
- “In silico Drug Design Based on Quantum Mechanical Scoring” 21st EuroQSAR – Where Molecular simulations Meet Drug Discovery “**Poster Presentation**”, Verona, Italy – **2016**
- “Quantum Mechanical Scoring of Protein-Ligand Binding Enhanced with Explicit-Solvent Effects” 11th Triennial Congress of the World Association of Theoretical and Computational Chemists- WATOC “**Poster Presentation**”, Munich, Germany – **2017**
- “Quantum Mechanical Investigation of Non-Covalent Interactions in Protein-Ligand Complexes” Fourth International Conference on Drug Discovery, Development and Lead Optimization “**Poster Presentation**”, San Francisco, USA – **2018**
- “Impressive enrichment of Quantum Mechanics-Based Scoring Function for Virtual Screening” Gordon Research Conference on Computer Aided Drug Design “**Poster Presentation**”, Vermont, USA – **2019**

ÚOCHB AV ČR
IOCB PRAGUE
Ústav organické chemie a biochemie
Akademie věd České republiky, v. v. i.
Institute of Organic Chemistry and Biochemistry
of the Czech Academy of Sciences

7. DECLARATION OF THE AUTHORSHIP

Prohlášení

Prohlášení spoluautorů upřesňující podíl Mgr. Cemal Köprülüoğlu na publikacích přiložených k disertaci:

A. Haresh Ajani, Josef Jansa, **Cemal Köprülüoğlu**, Pavel Hobza, Vladimír Kryštof, Antonín Lyčka, Martin Lepsik `Imidazo[1,2-c]pyrimidin-5(6H)-one as a novel core of cyclin-dependent kinase 2 inhibitors: Synthesis, activity measurement, docking, and quantum mechanical scoring` *Journal of Molecular Recognition*, **2018** (https://doi.org/10.1002/jmr.2720).

B. **Cemal Köprülüoğlu**, Milan Dejmek, Michal Šála, Haresh Ajani, Hubert Hřebabecký, Jindřich Fanfrlík, Radek Jorda, Martin Dračínský, Eliška Procházková, Pavel Šácha, Vladimír Kryštof, Pavel Hobza, Martin Lepšík and Radim Nencka `Optimization of Norbornyl-based Carbocyclic Nucleoside Analogs as Cyclin-Dependent Kinase 2 Inhibitors` *Journal of Molecular Recognition*, **2020** (https://doi.org/10.1002/jmr.2842), (Cemal Köprülüoğlu and Milan Dejmek contributed equally).

C. Michaela Hylsova, Benoit Carbain, Jindrich Fanfrlík, Lenka Musilova, Susanta Haldar, **Cemal Köprülüoğlu**, Haresh Ajani, Pathik S. Brahmkshatriya, Radek Jorda, Vladimír Krystof, Pavel Hobza, Aude Echalier, Kamil Paruch, Martin Lepsík `Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by newpyrazolo[1,5-a]pyrimidines` *European Journal of Medicinal Chemistry*, **2017**. (https://doi.org/10.1016/j.ejmech.2016.12.023).

D. Saltuk M. Eyrilmez, **Cemal Köprülüoğlu**, Jan Řezáč, and Pavel Hobza `Impressive Enrichment of Semiempirical QuantumMechanics-Based Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys` *ChemPhysChem*, **2019** (https://doi.org/10.1002/cphc.201900628), (Saltuk M. Eyrilmez and Cemal Köprülüoğlu contributed equally).

E. Adam Pecina, Saltuk M. Eyrilmez, **Cemal Köprülüoğlu**, Vijay Madhav Miriyala, Martin Lepšík, Jindřich Fanfrlík, Jan Řezáč, and Pavel Hobza `SQM/COSMO Scoring Function: Reliable Quantum-Mechanical Tool for Structure-Based Drug Design` submitted to *ChemPlusChem*, **2020.**

Mgr. Cemal Köprülüoğlu je prvním autorem na většině publikací přiložených k disertaci, což jednoznačně vymezuje jeho podíl. Ve všech případech je tento podíl dominantní a to ve všech fázích přípravy publikace, od zadání tématu až k jejímu sepsání.

V Praze, 23. března 2020   Prof. Ing. Pavel Hobza, Dr.Sc., dr. h. c., FRSC

PUBLICATIONS

PUBLICATIONS

Publication A

Haresh Ajani, Josef Jansa, Cemal Köprülüoğlu, Pavel Hobza, Vladimír Kryštof, Antonín Lyčka,
Martin Lepsik

Imidazo[1,2-c]pyrimidin-5(6H)-one as a novel core of cyclin-dependent kinase 2 inhibitors: Synthesis, activity measurement, docking, and quantum mechanical scoring

J. Mol. Recognit., **2018**, 31, e32720

Received: 9 January 2018 | Revised: 26 February 2018 | Accepted: 21 March 2018
DOI: 10.1002/jmr.2720

RESEARCH ARTICLE

WILEY Journal of Molecular Recognition

Imidazo[1,2-*c*]pyrimidin-5(6*H*)-one as a novel core of cyclin-dependent kinase 2 inhibitors: Synthesis, activity measurement, docking, and quantum mechanical scoring

Haresh Ajani[1,2] | Josef Jansa[3,4] | Cemal Köprülüoğlu[1,2] | Pavel Hobza[1,2] | Vladimír Kryštof[5] | Antonín Lyčka[3,6] | Martin Lepsik

[1] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague 6, Czech Republic

[2] Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, Czech Republic

[3] Research Institute for Organic Syntheses (VUOS), Pardubice-Rybitví, Czech Republic

[4] Department of Organic Chemistry, Faculty of Science, Palacký University, Olomouc, Czech Republic

[5] Laboratory of Growth Regulators, Centre of the Region Haná for Biotechnological and Agricultural Research, Faculty of Science, Palacký University and Institute of Experimental Botany, Olomouc, Czech Republic

[6] Faculty of Science, University of Hradec Králové, Hradec Králové, Czech Republic

Correspondence
Martin Lepsik, Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo nam. 2, 166 10 Prague 6, Czech Republic.
Email: lepsik@uochb.cas.cz

Funding information
Ministry of Industry and Trade of the Czech Republic, Grant/Award Number: CZ.01.1.02/0.0/0.0/15_019/0004431; European Regional Development Fund, Grant/Award Number: CZ.02.1.01/0.0/0.0/16_019/0000729; Palacky University in Olomouc, Grant/Award Number: IGA_PrF_2018_006

Abstract

We report on the synthesis, activity testing, docking, and quantum mechanical scoring of novel imidazo[1,2-*c*]pyrimidin-5(6*H*)-one scaffold for cyclin-dependent kinase 2 (CDK2) inhibition. A series of 26 compounds substituted with aromatic moieties at position 8 has been tested in in vitro enzyme assays and shown to inhibit CDK2. 2D structure-activity relationships have ascertained that small substituents at position 8 (up to the size of naphtyl or methoxyphenyl) generally lead to single-digit micromolar IC_{50} values, whereas bigger substituents (substituted biphenyls) decreased the compounds' activities. The binding modes of the compounds obtained using Glide docking have exhibited up to 2 hinge-region hydrogen bonds to CDK2 and differed in the orientation of the inhibitor core and the placement of the 8-substituents. Semi-empirical quantum mechanics-based scoring identified probable favourable binding modes, which will serve for future structure-based design and synthetic optimization of substituents of the heterocyclic core. In summary, we have identified a novel core for CDK2 inhibition and will explore it further to increase the potencies of the compounds and also monitor selectivities against other protein kinases.

KEYWORDS

binding mode, physics-based scoring, protein-ligand binding

1 | INTRODUCTION

The nearly 30 years of development of protein kinase inhibitors as anti-cancer agents has resulted in more than 30 compounds approved as drugs and several others under advanced clinical evaluation.[1-4] Despite this remarkable progress, the treatment is hindered by many problems, such as lack of selectivity (ie, many off-target kinases may be inhibited) or drug resistance (decrease of the compounds' affinity due to kinase mutations). Thus, new strategies and novel compounds are urgently needed.

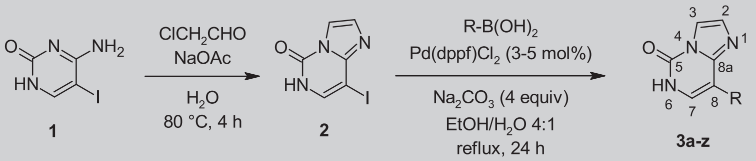
New kinase inhibitors are often developed by a multifaceted approach, including high-throughput screening of compound libraries using biochemical/cellular assays or analogue synthesis. During our decade-long research on kinase inhibitors,[5-11] we have ascertained

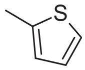
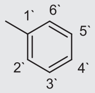
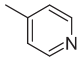
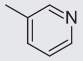
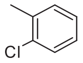
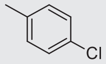
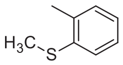
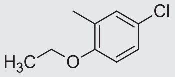
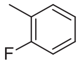
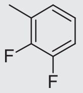
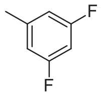
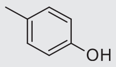
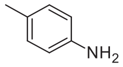
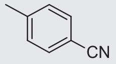
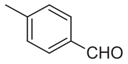
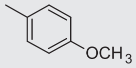
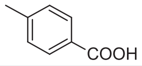
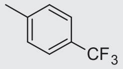
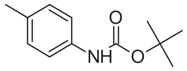
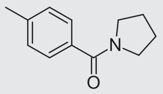
---

Haresh Ajani and Josef Jansa contributed equally to this work

Copyright © 2018 John Wiley & Sons, Ltd.

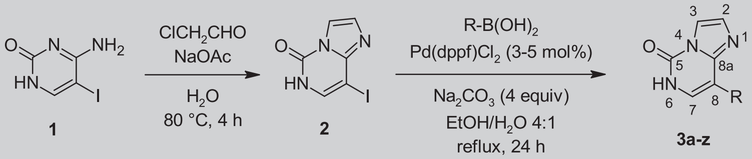
TABLE 1 Synthesis and 2D structures of the compounds, atom numbering, and activities against CDK2/cyclin E

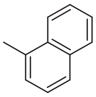
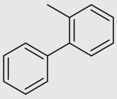
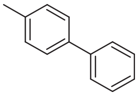
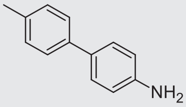
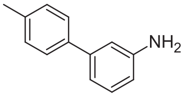
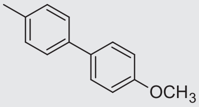

1
ClCH$_2$CHO
NaOAc
H$_2$O
80 °C, 4 h
2
R-B(OH)$_2$
Pd(dppf)Cl$_2$ (3-5 mol%)
Na$_2$CO$_3$ (4 equiv)
EtOH/H$_2$O 4:1
reflux, 24 h
3a-z
1
2
3
4
5
6
7
8
8a
R

Compound	R	IC$_{50}$ (µM)[a]	Yield (%)
3a		8.7 ± 5.3	61
3b		6.6 ± 0.6	88
3c		33.3 ± 17.7	62
3d		46.0 ± 13.0	83
3e		1.5 ± 0.3	76
3f		13.9 ± 1.5	75
3g		13.0 ± 1.3	87
3h		46.6 ± 5.8	62
3i		3.2 ± 0.9	75
3j		1.3 ± 0.4	21
3k		>25[b]	78
3l		4.8 ± 0.6	83
3m		8.3 ± 2.3	91
3n		14.0 ± 4.0	84
3o		19.7 ± 3.5	83
3p		29.3 ± 10.5	90
3q		>100	74
3r		41.5 ± 14.5	71
3s		42.1 ± 4.1	81
3t		>100	87

(Continues)

TABLE 1 (Continued)



Compound	R	IC$_{50}$ (μM)[a]	Yield (%)
3u		17.1 ± 2.9	70
3v		>100	34
3w		>12.5[b]	81
3x		>100	71
3y		23.2 ± 5.4	69
3z		>12.5[b]	78

[a] Measured at least as duplicates.

[b] Inhibition at higher concentrations could not be measured due to the limited solubility of the compounds.

the importance of the heterocyclic core[10,12,13] One less explored scaffold, imidazo[1,2-*c*]pyrimidine, was previously investigated randomly for biological activities, such as antimicrobial, antimycobacterial, antitubercular, inotropic, antiinflammatory, analgesic, antipyretic, and ulcerogenic.[14-18] Importantly, certain 5,7,8- and 2,5,8-trisubstituted imidazo[1,2-*c*]pyrimidine derivatives were described as potent Syk and Chk1 protein kinase inhibitors, respectively.[19-21] The described structure-activity relationships for Syk and Chk1 inhibitors as well the binding mode in the active sites of Chk1 and Lck kinases[21] inspired us to explore the imidazo[1,2-*c*]pyrimidin-5(6*H*)-one core as a starting point for development of cyclin-dependent kinase inhibitors.

Cyclin-dependent kinases (CDKs) are a family of Ser/Thr protein kinases, which are essential components of the cell-cycle regulation.[22] Their increased activity or mutations in their regulators often lead to an uncontrolled proliferation of cancer cells.[23-26] For these reasons, CDKs have become important targets for anticancer drug development.[4] CDK structure features the N-terminal β-sheet and C-terminal α-helical domains joined by the hinge region with an active-site cleft in between, which binds the ATP molecule. The ATP binding utilises 2 hinge-region hydrogen bonds (Glu81:O and Leu83:NH in cyclin-dependent kinase 2 [CDK2]), which are often mimicked by the small-molecule inhibitors. The majority of known CDK inhibitors are type I, ie, directly competing with ATP for its binding site,[2,4,27] while type II, III, and IV kinase inhibitors bind to allosteric sites. Detailed understanding of the binding of small-molecule inhibitors to CDK2 has been obtained through X-ray structures of co-crystal complexes,[4,27,28] recently also with CDK2 adopting the active conformation (ie, Thr160-phosphorylated and in a ternary complex with the regulatory cyclin subunit).[11,29,30] This wealth of structural information enables us to obtain insight into the binding modes of the imidazo[1,2-*c*]pyrimidin-5(6*H*)-one compounds in CDK2 by modelling and calculations.

Structure-based computer-aided design of new protein inhibitors has increasingly adopted quantum mechanical (QM) calculations[31-34] because they provide a quantitative treatment of protein-ligand (P-L) interactions,[35,36] metals (either in proteins[34,37] or ligands),[38] exotic ligands,[39] and noncovalent interactions of quantum origin, eg, σ-hole bonding,[31,40-42] or covalent binding of ligands.[43] The treatment of solvation in QM scoring needs to be fast and is thus usually done using continuum approaches (eg, COSMO).[34,44-46]

This paper reports on the discovery, synthesis, in vitro activity, docking, and semiempirical QM (SQM)-based scoring of a series of novel imidazo[1,2-*c*]pyrimidin-5(6*H*)-one CDK2 inhibitors. The 26 compounds (10 synthesised previously, 16 newly described here) have attained up to single-digit micromolar IC$_{50}$ values. 2D structure-activity relationships, binding mode analyses, and SQM-based scoring have helped to understand the measured activities and will serve for further design and synthetic optimization of the compounds to achieve better activities and selectivities.

2 | RESULTS AND DISCUSSION

2.1 | Compound design

Inspired by previous findings that substituted imidazo[1,2-c]pyrimidine derivatives can act as potent Syk and Chk1 protein kinase inhibitors and bind into their active sites,[19-21] we assayed 8 of the compounds based on the imidazo[1,2-c]pyrimidin-5(6H)-one core that we had synthesised previously (compounds **3a**, **3b**, **3c**, **3f**, **3n**, **3o**, **3p**, and **3w**[47] described below) and tested their inhibitory activity and selectivity against 3 pharmaceutically relevant kinase targets, ABL, CDK2, and PAK4. Several compounds acted as selective single-digit micromolar hits of CDK2 (Table S1 in the Supporting Information), which prompted us to explore this class of compounds more extensively.

2.2 | Compound synthesis

The synthesis of target imidazopyrimidines **3a-z** was adopted from our previous paper.[48] Firstly, iododerivative **2** was prepared by the reaction of 5-iodocytosine **1** with chloroacetaldehyde. The Suzuki-Miyaura cross-coupling of **2** with the corresponding boronic acids resulted in the series of 8-arylimidazo[1,2-*c*]pyrimidin-5(6*H*)-ones **3a-z** (Table 1). Catalysis by 3 mol% of Pd(dppf)Cl_2 in the mixture EtOH/H_2O 4:1 was applied in all cases to afford nearly complete conversions detected by TLC. Only the reactions of **2** with sterically more demanding biphenyl-2-boronic acid resulted in the lower conversion and lower yield of **3v**. Likewise, the introduction of another fluorine atom caused a lower conversion/yield of the reaction between **2** and 2,3-difluorophenylboronic acid (product **3j**) when compared with 2-fluoroderivative **3i**. Aminobiphenylboronic acids were prepared by the Suzuki-Miyaura reaction of bromoanilines with benzene-1,4-diboronic acid. Catalysis by Pd/C[48] allowed a reasonable proportion of the monocoupling reaction. After purification, 4′-amino- and 3′-aminobiphenylboronic acids were obtained in 1 reaction step in 29% and 20% yields, respectively.

The products were identified and characterised by IR, MS, and NMR spectroscopy. For all the new compounds **3d**, **3e**, **3g**, **3h**, **3i**, **3j**, **3k**, **3l**, **3m**, **3s**, **3t**, **3u**, **3v**, **3x**, **3y**, and **3z** (note that **3a**, **3b**, **3c**, **3f**, **3n**, **3o**, **3p**, **3q**, **3r**, and **3w** were described in our previous work[47]), the one- (1D) and 2-dimensional (2D) ^1H and ^{13}C NMR spectra were measured and analysed. The ^{19}F NMR spectra were measured for the fluorine-containing compounds **3i**, **3j**, and **3k**. 1D NMR spectra are collected in Table S2, and the numbering of the atoms is depicted in Scheme S1 (see the Supporting Information).

2.3 | Structure-activity relationship

The prepared imidazo[1,2-*c*]pyrimidin-5(6*H*)-ones were subjected to biochemical assays to determine their activity against recombinant CDK2/cyclin E. The results are summarised in Table 1. First, several

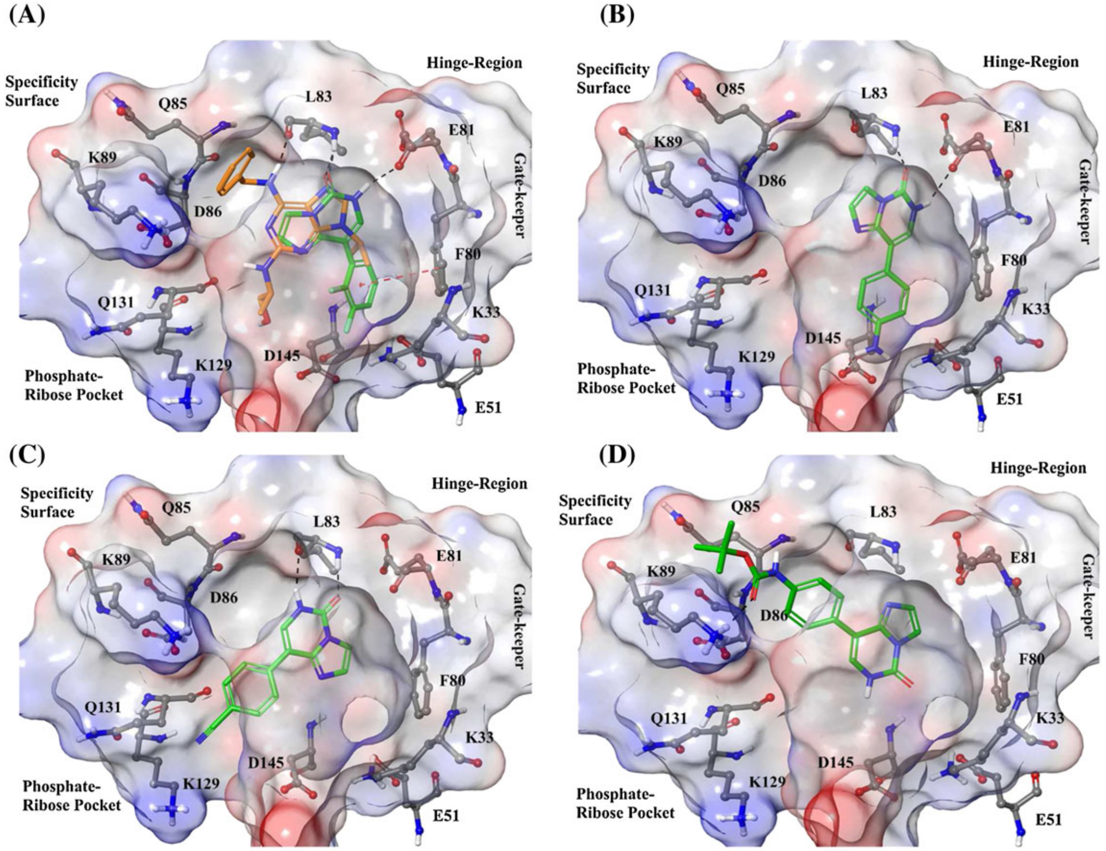

(A)
Specificity Surface
Hinge-Region
Q85
L83
E81
K89
D86
Gate-keeper
F80
Q131
K33
K129
D145
Phosphate-Ribose Pocket
E51
(B)
Specificity Surface
Hinge-Region
Q85
L83
E81
K89
D86
Gate-keeper
F80
Q131
K33
D145
K129
Phosphate-Ribose Pocket
E51
(C)
Specificity Surface
Hinge-Region
Q85
L83
E81
K89
D86
Gate-keeper
F80
Q131
K33
D145
K129
Phosphate-Ribose Pocket
E51
(D)
Specificity Surface
Hinge-Region
Q85
L83
E81
K89
D86
Gate-keeper
F80
Q131
K33
D145
K129
Phosphate-Ribose Pocket
E51

FIGURE 1 The binding modes of the imidazo[1,2-*c*]pyrimidin-5(6*H*)-one compounds (**3**) in CDK2. A, The active compound **3j** (green, bm1a) compared with the crystal structure of roscovitine (PDB code: 3DDQ).[49] B, The active compound **3m** (bm1b). C, The least active compound **3n** (bm2). D, The inactive compound **3s** (bm3). The ligand is shown in sticks, CDK2 as surface and important residues as ball and sticks. Colouring (C: green—ligand/grey—CDK2, N: blue, O: red, F: cyan, H: white). The figure was prepared by Maestro (Schrodinger)

aromatic R substituents were explored: 4′- and 3′-pyridyl (**3c**, **3d**) showed a weak inhibition, IC$_{50}$ = 33.3 μM and IC$_{50}$ = 46 μM, respectively. However, 2′-thienyl and phenyl (**3a**, **3b**) exhibited single-digit micromolar potency (IC$_{50}$ = 8.7 μM and IC$_{50}$ = 6.6 μM, respectively). Therefore, we further derived **3b** by small (**3e-3r**) and bigger (**3s-3z**) functional groups (Table 1). In the former class, slightly better activities were obtained for 2′-halogenation (**3e**, **3i**; IC$_{50}$ = 1.5 μM, 3.2 μM), 2′,3′-dihalogenation (**3j**; IC$_{50}$ = 1.3 μM) and 4-hydroxylation (**3 l**, IC$_{50}$ = 4.8 μM). Worse activities (up to 19.7 μM) were found for the derivatives with other substituents in this class except for compounds **3h**, **3k**, **3p**, **3q**, and **3r** with IC$_{50}$ > 25 μM. The molecular reasons for this behaviour will be described in the next section. All the bigger functional groups substantially reduced the activity (IC$_{50}$ > 12.5 μM).

2.4 | Inhibitor binding modes

Using structure-based approach, we have rationalised the observed structure-activity relationships. Docking was carried out in the active CDK2 structure using Glide SP and selection of the binding modes relied on SQM-based rescoring (see Methods and below).

Several types of binding modes (bm1-bm3) of the inhibitors were found. Not all of them were observed for all the compounds. The SQM-based rescoring sorted the 20 docking poses of each compound into groups with similar scores (within a few kcal/mol) which shared similar binding modes (within 0.5-Å root-mean-square deviation). The best PM6-D3H4X/COSMO-based score defined the preferred bm for each compound which was further analysed (Figure 1, Table 2).

The first one (bm1) was of the standard type I inhibitors featuring 2 hinge-region hydrogen bonds (E81:O…**3**:N(6)H and L83:NH…**3**:O5), similar to staurosporine[27] (Figure 1A). By tilting the core slightly, 2 sub-variants occurred: (1) bm1a, which placed the 8-aryl substituent towards the Lys33…Glu51 salt bridge (Figure 1A), and (2) bm1b with the 8-substituent, pointing towards the side chains of Lys33 and Asp145 (Figure 1B). The next binding mode (bm2) had the core reversed, with the 2 hinge-region hydrogen bonds being formed with Leu83 (L83:NH…**3**:O5 and L83:O…N(6)H), similar to purvalanol B[27] and the 8-aryl substituent pointing to the phosphate-ribose pocket of CDK2 (Figure 1C). The last binding mode (bm3) featured only 1 hinge-region hydrogen bond and the 8-aryl substituent pointing towards Q85 and K89 (Figure 1D). In relation to the CDK2 inhibitor roscovitine (Figure 1A), bm1 placed the 8-aryl moieties in the location of roscovitine N9-substituents, bm2 placed them in the location of roscovitine C2-substituents and bm3 placed them in the location of roscovitine C6-substituents.

The small substituents at position 8 (up to the size of 1-naphtyl or 4′-methoxyphenyl; compounds **3a-3r**) generally resulted in both bm1 and bm2 binding modes (with the exception of **3g**, **3h**, **3n** which, due to their bulkiness, exhibited only bm2 and **3q** with only bm3). The bm1 was characterised by the 2 hinge-region hydrogen bonds (Figure 1A, B) and the R-substituent interacting with the gatekeeper Phe80 via $\pi \cdots \pi$ stacking or CH$\cdots\pi$ interaction (bm1a, Figure 1A) and/or additional charge-assisted H-bonding interactions with Lys33, Glu51, and Asp145 (bm1a, bm1b; Figure 1A, B). The reversed-core binding mode (bm2) exhibited 2 different hinge-region hydrogen bonds and the placement of the 8-aryl substituents in the solvent-exposed pocket with no obvious direct P-L interactions (Figure 1C). The big R-substituents (**3s-3z**) resulted only in bm2/bm3 binding modes with the exception of **3u**, which bound via bm1, occupying simultaneously the regions around Glu51 and Asp145, as found in bm1a and bm1b binding modes, respectively.

TABLE 2 SQM-based scores and experimental ΔG_b^o (determined as ΔG_b^o = RT ln IC$_{50}$; all in kcal/mol) and IC$_{50}$ (μM)

Compound	PM6-D3H4X	DFTB3-D3H4X	ΔG_b^o	IC$_{50}$
bm1a				
3a	−26.5	−27.6	−7.1	8.7
3b	−24.7	−25.5	−7.3	6.6
3c	−24.3	−30.5	−6.4	33.3
3d	−24.7	−28.4	−6.2	46.0
3e	−23.0	−22.8	−8.2	1.5
3i	−22.5	−24.6	−7.7	3.2
3j	−22.3	−30.2	−8.2	1.3
3k	−19.4	−25.7	> −6.5	>25
bm1b				
3f	−17.8	−26.5	−6.9	13.9
3l	−22.7	−17.6	−7.5	4.8
3m	−9.8	−9.2	−7.2	8.3
3o	−16.5	−18.2	−6.7	19.7
3p	−18.9	−17.2	−6.4	29.3
3u	−16.4	−20.0	−6.8	17.0
bm2				
3g	−5.4	−13.7	−6.9	13.0
3h	−9.3	−11.7	−6.2	46.6
3n	−12.5	−14.3	−6.7	17.7
3r	−9.9	−13.8	−6.2	41.5
3y	−12.1	−14.4	−6.6	20.9
bm3				
3s	−13.6	−10.3	−6.2	42.1
3q	−6.9	−11.8	> − 5.7	>100
3t	−9.9	−12.1	> − 5.7	>100
3v	−5.9	−12.5	> − 5.7	>100
3w	−7.7	−8.6	> − 6.9	>12.5
3x	−9.9	−10.3	> − 5.7	>100
3z	−11.31	−14.61	> − 6.9	>12.5

2.5 | SQM-based scoring suggests molecular reasons for compound activities

Semiempirical quantum mechanics (SQM)-based scoring,[35,36] using PM6-D3H4X/COSMO and DFTB3-D3H4X/COSMO levels,[50,51] suggested molecular reasons for compound activities. The best-scoring pose at the PM6-D3H4X/COSMO level was selected for each compound and its PM6-D3H4X/COSMO and DFTB3-D3H4X/COSMO scores listed together with the experimental Gibbs free energies (ΔG_b^o) derived from the measured IC$_{50}$ values (Table 2). The compounds for which the predominant binding mode was bm1a attained the best scores (−19.4 to −26.5 kcal/mol for the former scoring function and − 22.8 to −30.5 kcal/mol for the latter). Only in the cases of

3c, **3d**, and **3k**, this prediction overestimated the binding affinity (for **3c** and **3d**, this might be caused by the overestimation of the strength of the charge-assisted hydrogen bond of their 4′- and 3′-pyridyl nitrogen and the Lys33 side chain). The bm1b was slightly worse, spanning from −9.8 to −22.7 and from −9.2 to −26.5 kcal/mol for the PM6-D3H4X/COSMO and DFTB3-D3H4X/COSMO levels, respectively. The **3m** compound seems to be underestimated, maybe due to the lack of receptor flexibility.

A reduction in the predicted activities was observed for the compounds with bm2 as the preferred binding mode (ranging from −5.4 to −12.5 and from −11.7 to −14.4 kcal/mol, respectively), for the PM6-D3H4X/COSMO and DFTB3-D3H4X/COSMO levels. The bm3 binding mode had similarly low scores, ranging from −5.9 to −13.6 and − 8.6 to −12.5, respectively.

It is important to mention that the docking followed by the SQM re-scoring gave only some insights into which binding modes would be favourable for binding. The fact that correlations with the experimental binding free energies were not obtained may be caused by narrow potency range, the lack of the crystal structure, lack of dynamical treatment, discrepancies between experimental affinity, and activity and other factors. Quantitative affinity calculations could be obtained by free-energy simulations. In any case, the current findings will be useful in further structure-based design and synthetic optimization.

3 | CONCLUSIONS AND OUTLOOK

This paper reports on the synthesis, activity measurements, and computational analyses of a series of 8-substituted imidazo[1,2-*c*]pyrimidine-5(6*H*)-ones as new CDK2 inhibitors. The measured IC_{50} values were in the micromolar range, up to single-digit values. The structure-activity relationships revealed that only smaller (up to the size of 1-naphtyl or 4′-methoxyphenyl) 8-substituents were compatible with the micromolar potency. The docking of the compounds into CDK2 provided several types of binding modes. This orientational freedom is allowed by the presence of only 1 substituent on the core. Semiempirical quantum-mechanical rescoring identified the probable favourable binding mode bm1a with 2 hinge-region hydrogen bonds. This finding will guide future structure-based design and synthetic optimization at positions 2, 3, and 8 (arylation, alkenylation, amination, halogenation, nitrosation, or nitration). Overall, this study is the first step in design and synthesis of tri-substituted imidazo[1,2-*c*]pyrimidine-5(6*H*)-one compounds to inhibit CDK2 and other kinases with higher potency and selectivity.

4 | METHODS

4.1 | Organic synthesis

4.1.1 | General methods

The NMR spectra were measured on Bruker Avance III HD 400 at 400.13 MHz (^{1}H), 100.62 MHz (^{13}C), and 376.50 MHz (^{19}F) in DMSO-d_6 at ambient temperature. The ^{1}H and ^{13}C chemical shifts were referenced to the residual signal of the solvent (δ = 2.50 (^{1}H) and 39.6 (^{13}C)). The ^{19}F chemical shifts were referred to internal $CFCl_3$ (δ = 0.0). Selected 2-dimensional gradient H,H-COSY, H,C-HMQC, H,C-HMQC-TOCSY, and H,C-HMBC spectra were measured using microprograms provided by the manufacturer (TOPSPIN 3.5) (data not shown). For 1D NMR spectra, all chemical shifts and their assignment to the particular atoms, see the Supporting Information (Table S2). Two types of ionisation for high-resolution mass spectrometry were used: electron impact ionisation and electrospray ionisation. The EI experiments were done on a GCT Premiere (Waters) machine. The measurement conditions were as follows: ionisation voltage of 70 eV, the positive mode, and the source temperature of 150°C. ESI experiments were done on a LTQ Orbitrap XL (ThermoFisher Sci.) under the following conditions: 80% methanol was used as a mobile phase, the source voltage of 4.3 kV, the capillary voltage of 9 V, the capillary temperature of 275°C, and the tube lens voltage of 155 V. The IR spectra were recorded on a Nicolet 6700 FT-IR with the MIRacle ATR equipment by Pike in the range of 640 to 4000 cm^{-1}. The MIRacle has a single-reflection ATR accessory with a diamond crystal. The melting points were determined on a Stuart SMP3 apparatus. The reaction progress was monitored by thin-layer chromatography (chloroform/methanol 5:1 as a mobile phase), which was performed on SiO_2 60 F_{254} plates with UV detection at 254 nm. All the starting materials were commercially available unless otherwise stated. The Pd/C catalyst, manufactured by Johnson Matthey, was obtained from D-Orland and has the following properties: type 394, Pd content: 5.13%, water content: 56.7%. The compounds **3a**, **3b**, **3c**, **3f**, **3n**, **3o**, **3p**, **3q**, **3r**, and **3w** were prepared according to our previous paper.[47]

4.1.2 | The general procedure for Suzuki-Miyaura cross-coupling

The mixture of **2** (2 g, 7.66 mmol), the corresponding boronic acid (11.50 mmol), sodium carbonate (3.25 g, 30.66 mmol), and Pd(dppf)Cl_2 (0.168 g, 0.23 mmol) were placed in a flask, into which argon atmosphere was introduced by 3 vacuum/argon cycles. Ethanol (48 mL) and then water (12 ml) were added, and 3 more vacuum/argon cycles were carried out. The reaction mixture was heated to reflux for 24 hours, after which the product was isolated from it (the isolation procedure for **3h**, **3k**, **3s**, **3u**, **3x**, **3y**, and **3z** is depicted below). Thereafter, the mixture was cooled to r.t., water (50 mL) was added, and then 35% hydrochloric acid (6 mL) was slowly added. Ethanol was evaporated at 50°C under reduced pressure. A little of active charcoal was added, and the mixture was stirred at 80°C for 10 minutes. The mixture was filtered over celite and washed with 4% hydrochloric acid (2 × 20 mL) at 80°C. After cooling to r.t., the pH of the filtrate was adjusted to 8 to 9 by solid sodium carbonate. The precipitated product was filtered off, washed with 100 mL of water and dried. The analytical samples of all the products were recrystallised from the solvent stated in the round brackets beyond mp and dried at 50°C/5 mbar. For NMR chemical shifts of new compounds, see the Supporting Information (Table S2).

4.1.3 | The isolation of the products 3h, 3k, 3s, 3u, 3x, 3y, and 3z

After the reaction mixture was refluxed for 24 hours (see above) and cooled to r.t., water (50 mL) was added, and the mixture was acidified with acetic acid (3.9 mL, 68.2 mmol). Ethanol was evaporated at 50°C under reduced pressure. The resulting suspension was filtered off, washed with 100 mL of water and dried. Crude products were recrystallised from the solvent stated in the round brackets beyond mp and dried at 50°C/5 mbar. The products recrystallised from DMF were washed with methanol (**3x**, **3y**) or acetone (**3z**).

4.1.4 | 8-(Pyridin-3-yl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3d)

White crystals, yield 1.35 g (83%), mp 284°C to 286°C (methanol), IR (ATR): 3159, 3063, 2587, 1704, 1625, 1549, 1483, 1423, 1402, 1280, 1255, 1194, 1132, 1105, 1047, 1027, 954, 895, 805, 777, 760, 732, 716, 706 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{11}H_8N_4O$: 212.0698, found 212.0699.

4.1.5 | 8-(2-Chlorophenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3e)

White crystals, yield 1.42 g (76%), mp 232°C to 234°C (acetonitrile), IR (ATR): 3167, 3062, 2728, 2623, 1716, 1619, 1549, 1503, 1489, 1470, 1426, 1407, 1335, 1284, 1259, 1241, 1133, 1108, 1064, 1035, 921, 880, 867, 756, 732, 717, 704, 652 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{12}H_8N_3OCl$: 245.0356, found 245.0354.

4.1.6 | 8-[2-(Methylsulfanyl)phenyl]imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3g)

White crystals, yield 1.72 g (87%), mp 193°C to 194°C (acetonitrile), IR (ATR): 3243, 3137, 2946, 1716, 1618, 1584, 1542, 1466, 1431, 1378, 1311, 1283, 1256, 1236, 1136, 1109, 1073, 1057, 1046, 966, 912, 875, 775, 754, 717, 694, 653 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{13}H_{11}N_3OS$: 257.0623, found 257.0621.

4.1.7 | 8-(5-Chloro-2-ethoxyphenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3h)

White crystals, yield 1.38 g (62%), mp 220°C to 223°C (acetonitrile), IR (ATR): 3174, 3089, 2931, 2613, 1751, 1733, 1697, 1596, 1544, 1491, 1465, 1389, 1325, 1286, 1266, 1237, 1128, 1111, 1032, 970, 924, 885, 854, 823, 804, 765, 739, 718, 703, 692, 652 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{14}H_{12}N_3O_2Cl$: 289.0618, found 289.0620.

4.1.8 | 8-(2-Fluorophenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3i)

White crystals, yield 1.32 g (75%), mp 254°C to 256°C (acetone), IR (ATR): 3064, 2925, 2837, 1721, 1623, 1546, 1480, 1443, 1405, 1333, 1280, 1256, 1240, 1215, 1153, 1141, 1113, 1098, 1053, 1031, 916, 874, 854, 816, 772, 765, 729, 717, 658 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{12}H_8N_3OF$: 229.0651, found 229.0649.

4.1.9 | 8-(2,3-Difluorophenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3j)

Unreacted starting iododerivative **2** was separated by filtration during recrystallisation from acetone. Then, the first crystals were discarded before crystallisation of pure **3j**. White crystals, yield 0.39 g (21%), mp 248°C to 250°C (acetone), IR (ATR): 3145, 3130, 3065, 2827, 2739, 1738, 1614, 1593, 1548, 1506, 1470, 1408, 1311, 1288, 1269, 1249, 1219, 1180, 1142, 1112, 1079, 1027, 930, 858, 813, 773, 745, 737, 713 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{12}H_7N_3OF_2$: 247.0557, found 247.0554.

4.1.10 | 8-(3,5-Difluorophenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3k)

White crystals, yield 1.48 g (78%), mp 325°C to 330°C (acetone), IR (ATR): 3145, 3128, 2829, 1738, 1618, 1595, 1547, 1511, 1474, 1431, 1341, 1326, 1286, 1252, 1147, 1114, 1018, 986, 929, 879, 845, 834, 774, 745, 733, 677 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{12}H_7N_3OF_2$: 247.0557, found 247.0559.

4.1.11 | 8-(4-Hydroxyphenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3 l)

White crystals, yield 1.45 g (83%), mp 342°C to 350°C (methanol), IR (ATR): 3380, 3157, 3118, 3043, 2728, 2607, 1705, 1610, 1547, 1515, 1508, 1433, 1411, 1349, 1271, 1246, 1216, 1171, 1143, 1112, 1105, 919, 874, 839, 822, 766, 725, 716, 652 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. For $C_{12}H_9N_3O_2$: 227.0695, found 227.0693.

4.1.12 | 8-(4-Aminophenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3 m)

4-Aminophenylboronic acid hydrochloride was used as starting material. Beige crystals, yield 1.57 g (91%), mp 297°C to 302°C decomposition (methanol), IR (ATR): 3424, 3352, 3218, 3078, 2919, 1706, 1608, 1542, 1514, 1478, 1429, 1406, 1373, 1330, 1275, 1238, 1178, 1139, 1107, 914, 882, 819, 761, 733, 653 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{12}H_{10}N_4O$: 226.0855, found 226.0857.

4.1.13 | 8-(4-Boc-aminophenyl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3 s)

White crystals, yield 2.03 g (81%), mp 208°C to 209°C decomposition (methanol), IR (ATR): 3449, 3060, 2929, 2836, 1726, 1709, 1616, 1548, 1519, 1495, 1408, 1367, 1318, 1275, 1240, 1218, 1153, 1111, 1053, 1029, 1015, 839, 826 766, 731, 718 cm^{-1}, HRMS (ESI) *m/z* [M + H]$^+$ calcd. for $C_{17}H_{19}N_4O_3$: 327.14517, found 327.14525.

4.1.14 | 8-(4-(pyrrolidine-1-carbonyl)phenyl)imidazo[1,2-c]pyrimidin-5(6H)-one (3 t)

White crystals, yield 2.05 g (87%), mp 248°C to 250°C (40% ethanol), IR (ATR): 3562, 3074, 2875, 1717, 1621, 1586, 1541, 1444, 1401, 1339, 1289, 1272, 1242, 1141, 1127, 1111, 1052, 917, 904, 845, 825, 768, 737, 711, 679 cm^{-1}, HRMS (ESI) *m/z* [M + H]$^+$ calcd. for $C_{17}H_{17}N_4O_2$: 309.13460, found 309.13465.

4.1.15 | 8-(Naphthalen-1-yl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3u)

Beige crystals, yield 1.40 g (70%), mp 275°C to 281°C (acetone), IR (ATR): 3149, 3122, 3058, 2823, 1722, 1625, 1616, 1544, 1501, 1487, 1411, 1341, 1282, 1243, 1141, 1110, 1020, 939, 915, 880, 866, 803, 781, 736, 724, 655 cm^{-1}, HRMS (EI) *m/z* [M]$^+$ calcd. for $C_{16}H_{11}N_3O$: 261.0902, found 261.0900.

4.1.16 | 8-(Biphenyl-2-yl)imidazo[1,2-*c*]pyrimidin-5(6*H*)-one (3v)

After recrystallisation from 40% ethanol, the product contained c. 8% of starting iododerivative **2**. This crude product was stirred with 50 mL of ethanol, and the remaining solid was filtered and discarded. Pure **3v** was subsequently crystallised from mother liquor. White crystals, yield 0.75 g (34%), mp 222°C to 229°C (ethanol), IR (ATR): 3156, 3064, 2830, 1706, 1626, 1545, 1477, 1439, 1402, 1332, 1281, 1236, 1141, 1116, 1072, 1049, 1010, 955, 920, 869, 767, 742, 700, 665 cm^{-1}, HRMS (EI) *m/z* $[M]^+$ calcd. for $C_{18}H_{13}N_3O$: 287.1059, found 287.1062.

4.1.17 | 8-(4′-amino-[1,1′-biphenyl]-4-yl) imidazo[1,2-c]pyrimidin-5(6H)-one (3x)

Starting boronic acid was prepared by the procedure stated below. Beige crystals, yield 1.65 g (71%), mp 330°C decomposition (DMF/MeOH), IR (ATR): 3407, 3347, 3240, 3081, 2869, 1705, 1616, 1543, 1494, 1406, 1287, 1260, 1243, 1175, 1140, 1130, 1107, 917, 843, 806, 778, 762, 751, 729 721, 706, 654 cm^{-1}, HRMS (ESI) *m/z* $[M + H]^+$ calcd. for $C_{18}H_{15}N_4O$: 303.12404, found 303.12411.

4.1.18 | 8-(3′-amino-[1,1′-biphenyl]-4-yl) imidazo[1,2-c]pyrimidin-5(6H)-one (3y)

Starting boronic acid was prepared by the procedure stated below. Grey solid, yield 1.61 g (69%), mp 285°C to 293°C (DMF/H_2O), IR (ATR): 3449, 3347, 3232, 3075, 2865, 2164, 1721, 1602, 1545, 1486, 1402, 1281, 1266, 1240, 1144, 1105, 933, 916, 866, 831, 779, 735, 685, 651 cm^{-1}, HRMS (ESI) *m/z* $[M + H]^+$ calcd. for $C_{18}H_{15}N_4O$: 303.12404, found 303.12409.

4.1.19 | 8-(4′-methoxy-[1,1′-biphenyl]-4-yl) imidazo[1,2-c]pyrimidin-5(6H)-one (3z)

Starting boronic acid was prepared according to Zang et al.52 Beige crystals, yield 1.87 g (78%), mp 309°C to 316°C (DMF), IR (ATR): 3136, 2838, 1748, 1606, 1546, 1526 1487, 1440, 1422, 1403, 1318, 1291, 1264, 1243, 1205, 1182, 1139, 1126, 1105, 1035, 1022, 1011, 918, 898, 823, 810, 743, 710, 654 cm^{-1}, HRMS (ESI) *m/z* $[M + H]^+$ calcd. for $C_{19}H_{16}N_3O_2$: 318.12370, found 318.12377.

4.1.20 | (4′-amino-[1,1′-biphenyl]-4-yl)boronic acid

The mixture of 4-bromoaniline (20 g, 116 mmol), benzene-1,4-diboronic acid (19.27 g, 116 mmol), sodium carbonate (24.65 g, 233 mmol), and wet 5% Pd/C catalyst (2.8 g, 0.58 mmol Pd) in methanol (200 mL) was stirred under reflux for 5 hours. Methanol was evaporated under reduced pressure, water (200 mL) was added to the residue, and the mixture was stirred overnight. The solid was filtered off, washed with water (200 mL), and dried to obtain 18.6 g of crude product which was suspended in 40% ethanol (200 mL). This mixture was heated to reflux to obtain a suspension. Hot suspension was filtered to remove the main impurity *p*-terphenyl-4,4″-diamine. Mother liquor was cooled overnight (≈5°C), and the precipitate was filtered off and dried to obtain 4′-aminobiphenylboronic acid. Beige crystals, yield 7.19 g (29%), mp 198°C to 200°C, ^1H NMR (400 MHz, DMSO-d_6): $\delta(^1H)$: 8.04 (2H, br s, $B(OH)_2$), 7.87 (2H, m), 7.57 (2H, m), 7.45 (2H, m), 6.72 (2H, m), 8 protons of two 1,4-disubstituted benzene rings, 5.27 (2H, br s, NH_2), ^{13}C NMR (100.6 MHz, DMSO-d_6): $\delta(^{13}C)$: 148.5, 142.2, 130.9, and 127.4 (all C), 134.4, 127.0, 124.3, and 114.3 (all 2 —CH═), IR (ATR): 3357, 3303, 3037, 1603, 1530, 1396, 1342, 1328, 1284, 1259, 1208, 1184, 1156, 1117, 1091, 1022, 995, 886, 813, 766, 741, 713 cm^{-1}, HRMS (ESI) *m/z* $[M + H]^+$ calcd. for $C_{12}H_{13}NO_2B$: 214.10339, found 214.10340.

4.1.21 | (3′-amino-[1,1′-biphenyl]-4-yl)boronic acid

The same procedure as for 4′-aminobiphenylboronic acid was applied with 3-bromoaniline. Yellowish crystals, yield 4.98 g (20%), mp 171°C to 174°C, ^1H NMR (400 MHz, DMSO-d_6): $\delta(^1H)$: 8.21 (2H, br s, $B(OH)_2$), 7.99, (2H, m), 7.67, (2H, m), 4 protons of 1,4-disubstituted benzene ring, 7.19, (1H, t), 7.01, (1H, d), 6.91, (1H, m), 6.69, (1H, m), 4 protons of 1,3-disubstituted benzene ring, 5.14 (2H, br s, NH_2), ^{13}C NMR (100.6 MHz, DMSO-d_6): $\delta(^{13}C)$: 149.1, 142.8, 141.6 and 132.9 (all C), 135.0, 125.8, (all 2 —CH═). 129.7, 115.2, 113.8, and 112.6 (all —CH═), IR (ATR): 3599, 3388, 3307, 3201, 3030, 1603, 1591, 1555, 1484, 1418, 1399, 1361, 1343, 1300, 1224, 1177, 1122, 1090, 1014, 993, 894, 872, 835, 790, 741, 687 cm^{-1}, HRMS (ESI) *m/z* $[M + H]^+$ calcd. for $C_{12}H_{13}NO_2B$: 214.10339, found 214.10340.

4.2 | Biochemical measurements

The tested compounds were dissolved in DMSO and diluted with water (the concentration of DMSO in the reaction was below 0.2%). The CDK2/cyclin E complex was produced in Sf9 insect cells via baculoviral infection and purified on a Ni^{2+}NTA column (Qiagen). Kinase (approx. 10 ng) was assayed using a mixture of the following: 1 mg/mL of histone H1, 15 μM of ATP, 0.05 of μCi [γ-^{33}P] ATP, the tested compound, and the reaction buffer, in the final volume of 10 μL. The reaction buffer consisted of: 60 mM of HEPES-NaOH, pH 7.5, 3 mM of $MgCl_2$, 3 mM of $MnCl_2$, 3 μM of Na-orthovanadate, 1.2 mM of DTT, and 2.5 μg/50 μL of $PEG_{20.000}$. The reactions were stopped by adding 5 μL of 3% aqueous H_3PO_4. Aliquots were spotted onto P-81 phosphocellulose (Whatman), washed 3 times with 0.5% aqueous H_3PO_4, and finally air-dried. Kinase inhibition was quantified using a FLA-7000 digital image analyser (Fujifilm). The concentration of the tested compounds required to reduce the CDK activity by 50% was determined from the concentration-response curves and reported as IC_{50} values.

4.3 | Computational methodology

The structure-based drug design strategy employed in this work entailed Glide docking and scoring followed by SQM-based rescoring.

4.3.1 | Molecular docking

Docking was carried out with Glide (version 75103, Schrodinger Suite) in the standard precision (SP) mode using the default Glide SP protocol.[53,54] The crystal structure of CDK2 in its active state (ie, in a ternary complex with cyclin A) bound to roscovitine inhibitor was used (the PDB code is 3DDQ).[49] The active-state CDK structures (even without the cyclin molecule) are preferable for structure-based design.[30] Prior to docking, the cyclin molecule, monothioglycerol, and water molecules were discarded. Only the chain A of CDK2 was used. By means of the Protein Preparation Wizard, hydrogen atoms were added to CDK2 using the default settings for titratable residues to match the experimental pH of 7.0. The hydrogen-bonding patterns were optimised using the OPLS3 force field.[55] The receptor grid for docking was created using the default settings. All the compounds **3a to 3z** were converted from 2D to 3D using the LigPrep module (Schrodinger Suite) and the default setting. Glide SP docking generated 20 poses which were subjected to SQM-based scoring.

4.3.2 | SQM-based scoring

SQM optimisation

All the 260 poses (ie, 10 poses for each of the 26 compounds) were scored in complex with CDK2 using our SQM-based methodology.[35,36] The best PM6-D3H4X/COSMO scores (see below) defined the poses for binding mode analyses (Figure 1, Table 2). Briefly, all the complexes were optimised at the PM6-D3H4X/COSMO level using the linear-scaling MOZYME algorithm[56] in MOPAC2016[57] via the Cuby4 interface[58] implementing the empirical corrections for noncovalent interactions.[59] We considered only the residues within 10 Å of the inhibitor, of which only the inner 8-Å surrounding of the inhibitor was allowed to move during the optimisation.

PM6-D3H4X/COSMO score

The SQM score was calculated according to Equation 1:[35,36]

$$\text{Score} = \Delta E_{\text{int}} + \Delta\Delta G_{\text{solv}} + \Delta G'_{\text{conf}}w(L) - T\Delta S. \qquad (1)$$

The individual terms describe 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 ($\Delta G'_{\text{conf}}{}^{w}(L)$), and optionally the entropy change upon the binding ($-T\Delta S$; not included in this work). The gas-phase interaction energies were calculated for the whole P-L systems and obtained as the difference between the energy of the complex and the energies of the protein and the ligand at the PM6-D3H4X.[59] The solvation free energy was determined using 2 implicit solvent models: COSMO[60] at the PM6 level for protein desolvation and SMD[61] at the HF/6-31G** level for the ligand for its greater accuracy.[44] The $\Delta G'_{\text{conf}}{}^{w}(L)$ term is the "free" energy change between the ligand in the protein-bound conformation and its optimal solution structure using PM6-D3H4X/SMD.

DFTB3-D3H4X/COSMO score

The gas-phase ΔE_{int} term was calculated at the DFTB3-D3H4X level, a third-order DFTB[62,63] with the 3OB[64] parameter set and a recent version of the D3H4X correction for noncovalent interactions.[65,66] This term replaced the PM6-D3H4X gas-phase term to yield the DFTB3-D3H4X/COSMO scores.

ASSOCIATED CONTENT

The inhibitory activities of previously prepared compounds against human purified kinases and NMR spectra of all the new compounds are presented in the Supporting Information.

ACKNOWLEDGEMENT

We thank Dr Radek Jorda, PhD for his critical reading and useful comments on the manuscript. This work has been supported by the Large Infrastructures for Research, Experimental Development and Innovations project "IT4Innovations National Supercomputing Center—LM2015070". VK thanks Palacky University in Olomouc (IGA_PrF_2018_006). H.A., C.K., P.H., and M.L. thank the research project RVO 61388963, awarded by the Czech Academy of Sciences and the European Regional Development Fund; OP RDE; Project: "Chemical biology for drugging undruggable targets (ChemBioDrug)" (No. CZ.02.1.01/0.0/0.0/16_019/0000729). J.J. and A.L. thank the Ministry of Industry and Trade of the Czech Republic for support (grant number CZ.01.1.02/0.0/0.0/15_019/0004431). We thank Martin Svoboda for mass spectrometry measurements, Jan Vaněček for recording IR spectra, and Dagmar Zvolánková for technical help.

ORCID

Martin Lepsik http://orcid.org/0000-0003-2607-8132

REFERENCES

1. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*. 2009;9(1):28-39.
2. Fischer PM. Approved and experimental small-molecule oncology kinase inhibitor drugs: a mid-2016 overview. *Med Res Rev*. 2017;37:314-367.
3. Fabbro D. 25 years of small molecular weight kinase inhibitors: potentials and limitations. *Mol Pharmacol*. 2015;87(5):766-775.
4. Asghar U, Witkiewicz AK, Turner NC, Knudsen ES. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat Rev Drug Discov*. 2015;14(2):130-146.
5. Krystof V, Cankar P, Frysova I, et al. 4-arylazo-3,5-diamino-1H-pyrazole CDK inhibitors: SAR study, crystal structure in complex with CDK2, selectivity, and cellular effects. *J Med Chem*. 2006;49(22):6500-6509.
6. Jorda R, Havlicek L, McNae IW, et al. Pyrazolo[4,3-d]pyrimidine bioisostere of roscovitine: evaluation of a novel selective inhibitor of cyclin-dependent kinases with antiproliferative activity. *J Med Chem*. 2011;54(8):2980-2993.
7. Haider C, Grubinger M, Reznickova E, et al. Novel inhibitors of cyclin-dependent kinases combat hepatocellular carcinoma without inducing chemoresistance. *Mol Cancer Ther*. 2013;12(10):1947-1957.
8. Gucky T, Jorda R, Zatloukal M, et al. A novel series of highly potent 2,6,9-trisubstituted purine cyclin-dependent kinase inhibitors. *J Med Chem*. 2013;56(15):6234-6247.
9. Mojzych M, Subertova V, Bielawska A, et al. Synthesis and kinase inhibitory activity of new sulfonamide derivatives of pyrazolo[4,3-e][1,2,4]triazines. *Eur J Med Chem*. 2014;78:217-224.
10. Vymetalova L, Havlicek L, Sturc A, et al. 5-Substituted 3-isopropyl-7-[4-(2-pyridyl)benzyl]amino-1(2)H-pyrazolo[4,3-d]pyrimidines with anti-proliferative activity as potent and selective inhibitors of cyclin-dependent kinases. *Eur J Med Chem*. 2016;110:291-301.
11. Hylsova M, Carbain B, Fanfrlik J, et al. Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: inhibition of CDK2 by new pyrazolo[1,5-a]pyrimidines. *Eur J Med Chem*. 2017;126:1118-1128.
12. Jorda R, Paruch K, Krystof V. Cyclin-dependent kinase inhibitors inspired by roscovitine: purine bioisosteres. *Curr Pharm Des*. 2012;18(20):2974-2980.
13. Nekardova M, Vymetalova L, Khirsariya P, et al. Structural basis of the interaction of cyclin-dependent kinase 2 with roscovitine and its analogues having bioisosteric central heterocycles. *Chemphyschem*. 2017;18(7):785-795.
14. Abignente E, Arena F, De Caprariis P, et al. Research on heterocyclic compounds. XXVII. Imidazo[1,2-c]pyrimidines. *Farmacoterapia*. 1991;46:1099-1110.

15. Bhuiyan MMH, Rahman KMM, Hossain MK, Rahim MA, Hossain MI. Fused pyrimidines. Part II: synthesis and antimicrobial activity of some furo[3,2-e]imidazo[1,2-c]pyrimidines and furo[2,3-d]pyrimidines. *Croat Chem Acta*. 2005;78:633-636.
16. Chhabria MT, Jani MH. Design, synthesis and antimycobacterial activity of some novel imidazo[1,2-c]pyrimidines. *Eur J Med Chem*. 2009;44(10):3837-3844.
17. Spitzer WA, Victor F, Pollock GD, Hayes JS. Imidazo[1,2-a]pyrimidines and imidazo[1,2-a]pyrazines: the role of nitrogen position in inotropic activity. *J Med Chem*. 1988;31(8):1590-1595.
18. Moraski GC, Markley LD, Chang M, et al. Generation and exploration of new classes of antitubercular agents: the optimization of oxazolines, oxazoles, thiazolines, thiazoles to imidazo[1,2-a]pyridines and isomeric 5,6-fused scaffolds. *Bioorg Med Chem*. 2012;20(7):2214-2220.
19. Hirabayashi A, Mukaiyama H, Kobayashi H, et al. Structure-activity relationship studies of imidazo[1,2-c] pyrimidine derivatives as potent and orally effective Syk family kinases inhibitors. *Bioorg Med Chem*. 2008;16(20):9247-9260.
20. Hirabayashi A, Mukaiyama H, Kobayashi H, et al. Structure-activity relationship studies of 5-benzylaminoimidazo[1,2-c]pyrimidine-8-carboxamide derivatives as potent, highly selective ZAP-70 kinase inhibitors. *Bioorg Med Chem*. 2009;17(1):284-294.
21. Meng ZY, Ciavarri JP, McRiner A, et al. Potency switch between CHK1 and MK2: discovery of imidazo[1,2-a]pyrazine- and imidazo[1,2-c] pyrimidine-based kinase inhibitors. *Bioorg Med Chem Lett*. 2013;23(10):2863-2867.
22. Malumbres M. Cyclin-dependent kinases. *Genome Biol*. 2014;15(6):122.
23. Hall M, Peters G. Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer. *Adv Cancer Res*. 1996;68:67-108.
24. Keyomarsi K, Conte D Jr, Toyofuku W, Fox MP. Deregulation of cyclin E in breast cancer. *Oncogene*. 1995;11(5):941-950.
25. Dobashi Y, Shoji M, Jiang SX, Kobayashi M, Kawakubo Y, Kameya T. Active cyclin A-CDK2 complex, a possible critical factor for cell proliferation in human primary lung carcinomas. *Am J Pathol*. 1998;153(3):963-972.
26. Porter DC, Zhang N, Danes C, et al. Tumor-specific proteolytic processing of cyclin E generates hyperactive lower-molecular-weight forms. *Mol Cell Biol*. 2001;21(18):6254-6269.
27. Huwe A, Mazitschek R, Giannis A. Small molecules as inhibitors of cyclin-dependent kinases. *Angew Chem Int Ed*. 2003;42(19):2122-2138.
28. Levin NMB, Pintro VO, de Avila MB, de Mattos BB, De Azevedo WF Jr. Understanding the structural basis for inhibition of cyclin-dependent kinases. New pieces in the molecular puzzle. *Curr Drug Targets*. 2017;18(9):1104-1111.
29. Coxon CR, Anscombe E, Harnor SJ, et al. Cyclin-dependent kinase (cdk) inhibitors: structure activity relationships and insights into the Cdk-2 selectivity of 6-substituted 2-arylaminopurines. *J Med Chem*. 2017;60(5):1746-1767.
30. Kontopidis G, McInnes C, Pandalaneni SR, et al. Differential binding of inhibitors to active and inactive CDK2 provides insights for drug design. *Chem Biol*. 2006;13(2):201-211.
31. Lange A, Gunther M, Buttner FM, et al. Targeting the gatekeeper MET146 of C-Jun N-terminal kinase 3 induces a bivalent halogen/chalcogen bond. *J Am Chem Soc*. 2015;137(46):14640-14652.
32. Berg L, Mishra BK, Andersson CD, Ekstrom F, Linusson A. The nature of activated non-classical hydrogen bonds: a case study on acetylcholinesterase-ligand complexes. *Chem Euro Journal*. 2016;22(8): 2672-2681.
33. Ryde U, Soderhjelm P. Ligand-binding affinity estimates supported by quantum-mechanical methods. *Chem Rev*. 2016;116(9):5520-5566.
34. Raha K, Merz KM. A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. *J Am Chem Soc*. 2004;126(4):1020-1021.
35. Fanfrlik J, Bronowska AK, Rezac J, Prenosil O, Konvalinka J, Hobza P. 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*. 2010;114(39):12666-12678.
36. Lepsik M, Rezac J, Kolar M, Pecina A, Hobza P, Fanfrlik J. The semiempirical quantum mechanical scoring function for in silico drug design. *ChemPlusChem*. 2013;78(9):921-931.
37. Pecina A, Lepsik M, Rezac J, et al. 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*. 2013;117(50):16096-16104.
38. Ciancetta A, Genheden S, Ryde U. A QM/MM study of the binding of RAPTA ligands to cathepsin B. *J Comput Aided Mol Des*. 2011;25(8):729-742.
39. Fanfrlik J, Brynda J, Rezac J, Hobza P, Lepsik M. Interpretation of protein/ligand crystal structure using QM/MM calculations: case of HIV-1 protease/metallacarborane complex. *J Phys Chem B*. 2008;112(47):15094-15102.
40. Fanfrlik J, Kolar M, Kamlar M, et al. Modulation of aldose reductase inhibition by halogen bond tuning. *ACS Chem Biol*. 2013;8(11): 2484-2492.
41. Fanfrlik J, Ruiz FX, Kadlcikova A, et al. The effect of halogen-to-hydrogen bond substitution on human aldose reductase inhibition. *ACS Chem Biol*. 2015;10(7):1637-1642.
42. Cousido-Siah A, Ruiz FX, Fanfrlik J, et al. IDD388 polyhalogenated derivatives as probes for an improved structure-based selectivity of AKR1B10 inhibitors. *ACS Chem Biol*. 2016;11(10):2693-2705.
43. Fanfrlik J, Brahmkshatriya PS, Rezac J, et al. Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic schistosoma mansoni cysteine peptidase by vinyl sulfone inhibitors. *J Phys Chem B*. 2013;117(48):14973-14982.
44. Kolar M, Fanfrlik J, Lepsik M, Forti F, Luque FJ, Hobza P. Assessing the accuracy and performance of implicit solvent models for drug molecules: conformational ensemble approaches. *J Phys Chem B*. 2013;117(19):5950-5962.
45. Soderhjelm P, Kongsted J, Ryde U. Ligand affinities estimated by quantum chemical calculations. *J Chem Theory Comput*. 2010;6(5):1726-1737.
46. Decherchi S, Masetti M, Vyalov I, Rocchia W. Implicit solvent methods for free energy estimation. *Eur J Med Chem*. 2015;91:27-42.
47. Jansa J, Lycka A, Padelkova Z, et al. New imidazo[1,2-c]pyrimidin-5(6H)-ones derived from cytosine: synthesis, structure, and cytotoxic activity. *J Heterocyclic Chem*. 2015;52(5):1382-1389.
48. Jansa J, Reznicek T, Jambor R, Bures F, Lycka A. Synthesis of hydroxy-substituted p-terphenyls and some larger oligophenylenes via palladium on charcoal catalyzed Suzuki-Miyaura reaction. *Adv Synth Catal*. 2017;359(2):339-350.
49. Bettayeb K, Oumata N, Echalier A, et al. CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases. *Oncogene*. 2008;27(44):5797-5807.
50. Pecina A, Haldar S, Fanfrlik J, et al. SQM/COSMO scoring function at the DFTB3-D3H4 level: unique identification of native protein-ligand poses. *J Chem Inf Model*. 2017;57(2):127-132.
51. Ajani H, Pecina A, Eyrilmez SM, et al. Superior performance of the SQM/COSMO scoring functions in native pose recognition of diverse protein-ligand complexes in cognate docking. *Acs Omega*. 2017;2(7): 4022-4029.
52. Zang ZQ, Zhang D, Wan XH, Zhou QF. The synthesis and property of liquid crystalline 4-alkoxyl-4 -cyano-p-terphenyls. *Mol Cryst Liq Cryst*. 2000;339(1):145-158.
53. Friesner RA, Banks JL, Murphy RB, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004;47(7):1739-1749.
54. Small-Molecule Drug Discovery Suite 2017–3,Glide, Schrödinger, LLC, New York, NY,. 2017.

55. Harder E, Damm W, Maple J, et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J Chem Theory Comput*. 2016;12(1):281-296.
56. Stewart JJ. Application of the PM6 method to modeling proteins. *J Mol Model*. 2009;15(7):765-805.
57. MOPAC2016, J. J. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://http://openmopac.net *(2016)*.
58. Rezac J. Cuby: an integrative framework for computational chemistry. *J Comput Chem*. 2016;37(13):1230-1237.
59. Rezac J, Hobza P. Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J Chem Theory Comput*. 2012;8(1):141-151.
60. Klamt A, Schuurmann G. COSMO—a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc-Perkin Trans*. 1993;2:799-805.
61. Marenich AV, Cramer CJ, Truhlar DG. 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*. 2009;113(18):6378-6396.
62. Gaus M, Cui QA, Elstner M. DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J Chem Theory Comput*. 2011;7(4):931-948.
63. Gaus M, Goez A, Elstner M. Parametrization and benchmark of DFTB3 for organic molecules. *J Chem Theory Comput*. 2013;9(1):338-354.
64. Gaus M, Lu XY, Elstner M, Cui Q. Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications. *J Chem Theory Comput*. 2014;10(4):1518-1537.
65. Kubillus M, Kubar T, Gaus M, Rezac J, Elstner M. Parameterization of the DFTB3 method for Br, Ca, Cl, F, I, K, and Na in organic and biological systems. *J Chem Theory Comput*. 2015;11(1):332-342.
66. Hostas J, Rezac J. Accurate DFT-D3 calculations in a small basis set. *J Chem Theory Comput*. 2017;13(8):3575-3585.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Ajani H, Jansa J, Köprülüoğlu C, et al. Imidazo[1,2-*c*]pyrimidin-5(6*H*)-one as a novel core of cyclin-dependent kinase 2 inhibitors: Synthesis, activity measurement, docking, and quantum mechanical scoring. *J Mol Recognit*. 2018;31:e2720. https://doi.org/10.1002/jmr.2720

PUBLICATIONS

Publication B

Cemal Köprülüoğlu, Milan Dejmek, Michal Šála, Haresh Ajani, Hubert Hřebabecký, Jindřich Fanfrlík, Radek Jorda, Martin Dračínský, Eliška Procházková, Pavel Šácha, Vladimír Kryštof, Pavel Hobza, Martin Lepšík and Radim Nencka

Optimization of Norbornyl-based Carbocyclic Nucleoside Analogs as Cyclin-Dependent Kinase 2 Inhibitors

J. Mol. Recognit., **2020**, doi.org/10.1002/jmr.2842

Journal Code	Article ID	Dispatch: 25-FEB-20	CE: Rajesh
JMR	2842	No. of Pages: 16	ME:

Received: 13 November 2019 | Revised: 26 January 2020 | Accepted: 17 February 2020

DOI: 10.1002/jmr.2842

RESEARCH ARTICLE

Journal of Molecular Recognition WILEY

Optimization of norbornyl-based carbocyclic nucleoside analogs as cyclin-dependent kinase 2 inhibitors

Cemal Köprülüoğlu[1,2] | Milan Dejmek[1] | Michal Šála[1] | Haresh Ajani[1,2] | Hubert Hřebabecký[1] | Jindřich Fanfrlík[1] | Radek Jorda[3] | Martin Dračínský[1] | Eliška Procházková[1] | Pavel Šácha[1] | Vladimír Kryštof[3] | Pavel Hobza[1,2] | Martin Lepšík[1] | Radim Nencka[1]

[1]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague 6, Czech Republic

[2]Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, Czech Republic

[3]Laboratory of Growth Regulators, Faculty of Science, Institute of Experimental Botany of the Czech Academy of Sciences, ~~Palacký University~~, Olomouc, Czech Republic

Correspondence
Martin Lepšík and Radim Nencka, Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo nam. 2, 166 10 Prague 6, Czech Republic.
Email: lepsik@uochb.cas.cz (M. L.) and nencka@uochb.cas.cz (R. N.)

Funding information
Academy of Sciences of the Czech Republic , Grant/Award Number: RVO 61388963; Czech Science Foundation , Grant/Award Number: 19-09086S; European Regional Development Fund, Grant/Award Number: CZ.02.1.01/0.0/0.0/16_019/0000729; Gilead Sciences, Inc.; Ministry of Education, Youth and Sports of the Czech Republic, Grant/Award Numbers: LM2015070, No. L01305

Abstract

We report on the discovery of norbornyl moiety as a novel structural motif for CDK2 inhibitors which was identified by screening a carbocyclic nucleoside analogue library. Three micromolar hits were expanded by the use of medicinal chemistry methods into a series of novel compounds with prevailingly micromolar activities against CDK2. The best compound of the series attained IC_{50} of 190 nM. The binding modes and affinities were explored at the molecular details by modeling, docking, and quantum mechanics-based scoring. In conclusion, the discovered 9-hydroxymethylnorbornyl moiety was shown by joint experimental-theoretical efforts to be able to serve as a novel substituent for CDK2 inhibitors making interactions with the gatekeeper pocket and opens a new space for the exploration of chemical space towards more effective derivatives targeting this important class of protein kinases.

KEYWORDS

ATP-competitive type I inhibitors, cyclin-dependent kinase 2, protein-ligand binding, quantum mechanical scoring

Uncorrected proofs

1 | INTRODUCTION

Cyclin-dependent kinases (CDKs) are ubiquitous enzymes in animals with several isoforms which are essential for numerous cell functions, including cell cycle regulation. Thus, their malfunctioning may lead to various types of cancer. Compounds targeting CDKs can thus be human anticancer therapeutics, as exemplified by three CDK4/CDK6 inhibitors palbociclib, ribociclib, and abemaciclib approved by FDA for the treatment of ER-positive and HER2-negative breast cancer.[1]

Like all other kinases, 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, that is, binding to active conformation of the kinase and directly competing with ATP for the binding site. Cyclin-dependent kinase 2 (CDK2) is the best known and studied member of CDK family.[2] Detailed understanding of the binding of small-molecule inhibitors to CDK2 comes from X-ray structures of co-crystal complexes.[3]

Cemal Köprülüoğlu and Milan Dejmek contributed equally.

https://doi.org/10.1002/jmr.2842

© 2020 John Wiley & Sons Ltd

Reliable prediction of protein-ligand binding affinity (scoring) is a major but still unsolved task of structure-based drug design. In our laboratory, we have been developing semiempirical quantum mechanics (SQM)-based scoring functions for a general and reliable description of diverse protein–ligand complexes.[4,5] These have successfully been applied to dozens of protein targets, including kinases,[6,7] binding up to thousands of ligands.[8–12]

Small molecule ATP-competitive kinase inhibitors must be adequately sized and shaped in order to fit into the active site. Screening nucleoside libraries might thus prove to be a rich source of new types of scaffolds for design of kinase inhibitors. During our past projects, we have been working on synthesis of numerous new compounds derived mostly from nucleosides. Namely, we have been preparing new carbocyclic and locked nucleoside analogues.[13–18]

These types of compounds have biological effects, most importantly antiviral and cytostatic activities. In addition, various nucleosides, including carboycylic and locked analogues, exert inhibitory potency against diverse medicinally relevant enzymes,[19] such as protein[20] and lipid kinases.[21,22]

Here, we report on a novel structural pattern for CDK2 inhibitors identified and optimized from the three screening hits discovered from our proprietary compound library of mostly nucleoside derivatives. The chemical space around the hits was expanded using our experience in kinase inhibitor design and 16 new compounds were synthesized. The molecular details of their binding were studied by modeling, docking, and quantum mechanics-based scoring.

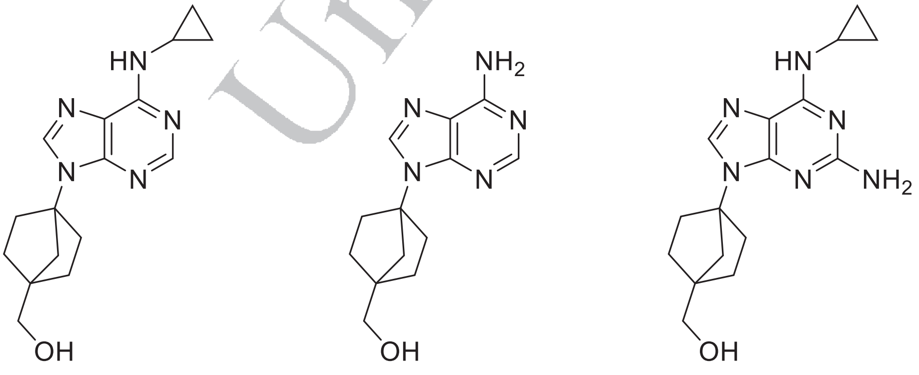
2 | RESULTS AND DISCUSSION

2.1 | Hit discovery and expansion

Previous campaigns launched at the Institute of Organic Chemistry and Biochemistry (IOCB) in Prague focused mostly on diverse nucleoside and nucleotide analogues and resulted in numerous clinically successful antiviral compounds.[23,24] These compounds, after the respective project end, were collected in the IOCB proprietary library, ready for further exploration. This unique collection of compounds allowed us to perform several screening campaigns, one of which focused on identification of novel protein kinase inhibitors. Over 1000 compounds from this database were taken forward for virtual screening against CDK2 kinase using quantum mechanics-based SQM/COSMO methodology.[5] In total, 200 best-scored structures were selected for activity testing against CDK2. We identified three norbornane-based compounds, **1**, **2**, and **3**, prepared in our lab, as hits for CDK2 inhibition (Figure 1). These compounds are nucleoside analogues and thus can occupy the ATP-binding site of kinases and thus potentially serve as type I inhibitors.

Molecular docking of these compounds in the CDK2 active site suggested two binding modes. The standard one, very similar to that of the purine-based inhibitor roscovitine,[25–28] featured two hinge-region hydrogen bonds with Leu83 and placed the 2,6,9-substituents into their respective canonical pockets (Figure S1-left). For **1**, we found in addition a reverse binding mode with the two hinge-region hydrogen bonds present but provided by different inhibitor atoms (Figure S1-right). The 2,9-substituents pointed to their respective pockets, yet under different angle. However, the potential 6-substituent would be swapped with the 9-substituent and would thus point toward the gatekeeper, which would hinder a productive hit optimization. Thus, only the canonical-binding mode was used for hit expansion.

Following the known structure-activity relationships (SARs) in purine-isostere-based kinase inhibitors,[29,30] we have selected 36 modifications of position 2 and 37 in position 6 and combined them with the hydroxynorbornyl moiety in position 9 on a purine scaffold. To prioritize among these 73 compounds for synthesis, we built the modifications on the purine core in the CDK2 active site and scored using SQM/COSMO approach.[5] Eight compounds with high scores and synthetic feasibility were selected for synthesis. Additionally, to explore the importance of the hydroxynorbornyl moiety in position 9, another eight compounds with the cyclohexyl substituent in position 9 were suggested for synthesis.



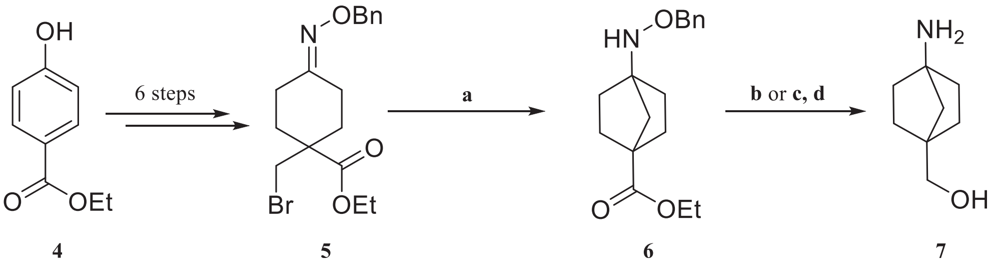
1 $IC_{50} = 5.7\ \mu M$

2 $IC_{50} = 4.7\ \mu M$

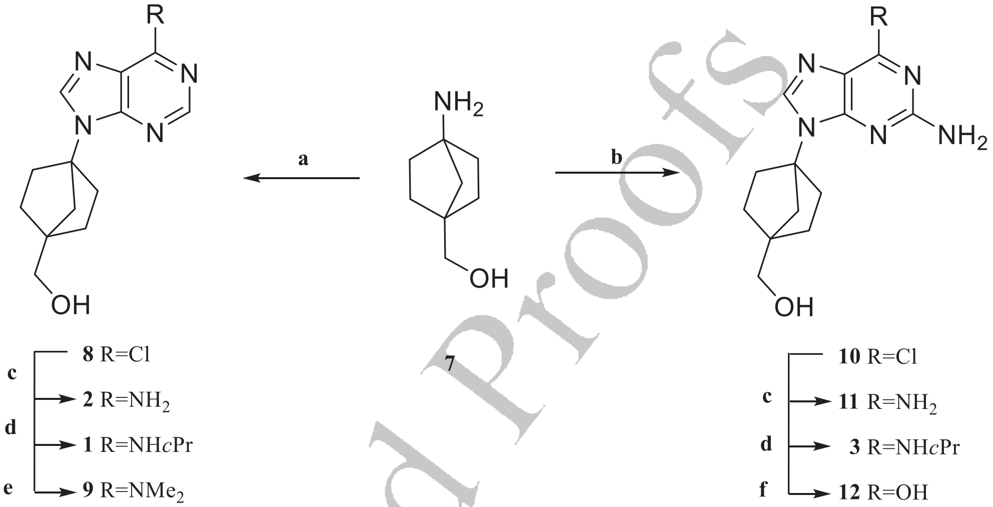
3 $IC_{50} = 12.6\ \mu M$

FIGURE 1 Three hits for CDK2 inhibition from the IOCB proprietary database

Uncorrected Proofs



SCHEME 1 (a) Bu3SnH, AIBN, toluene, reflux, 5 hours, 49%; (b) BH3-THF, diglyme, 110°C, 24 hours, 95%, (c) Pd(OH) 2/C, H2, MeOH, 24 hours, 85%, (d) LiAlH4, THF, reflux 5 hours, 67%



SCHEME 2 (a) 4,6-Dichloro-5-formamidopyrimidine, DIPEA, *n*-BuOH, MW, 160°C, 2 hours, 60%; (b) 2-amino-4,6-dichloro-5-formamidopyrimidine, DIPEA, n-BuOH, MW, 160°C, 2 hours, 80%; (c) NH3-EtOH, MW. 120°C, 30 minutes, 91% for **2** or 65% for **11**; (d) *c*-PrNH2, EtOH, MW 140°C, 30 minutes, 91% for **1** or 80% for **3**; (e) DMF, MW, 200°C, 2 minutes, 87%; (f) TFA, H2O, 12 hours, 69%

2.2 | Synthesis of nucleoside derivatives as novel CDK2 inhibitors

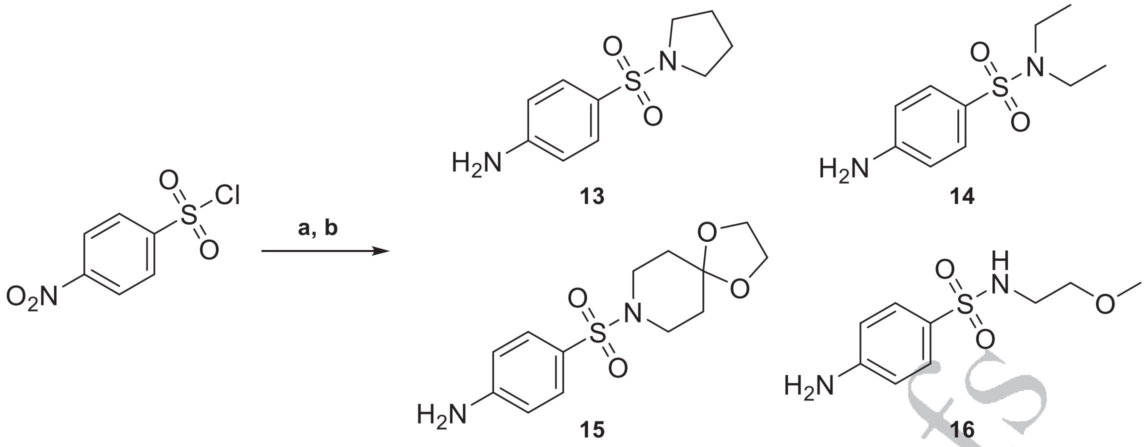
Synthesis of the original nucleoside analogues started with the preparation of double bridgehead substituted norbornane bicycle (**7**). First, oxime **5** was synthesized in a straightforward manner in 6 steps from ethylparaben **4**. Using Bu_3SnH-mediated radical cyclization of this intermediate, hydroxylamine **6** was obtained (Scheme 1). It is noteworthy that success of this reaction was strongly dependent on the thoroughness of degassing the reaction medium. In standard, undegassed solvent no cyclization occurred, when using simpler degassing procedures such as bubbling inert gas (N_2 or Ar) through the reaction medium yields rose to mediocre 20-50% and after thorough freeze-pump-thaw procedure acceptable 65% yield was achieved. Debrominated **5** was always present in the reaction mixture as a major impurity. The benzyloxy group, together with the ester function, was reduced to obtain key amine **7**, which was used in the following nucleobase construction according to our one-step protocol [31] to prepare the 6-chloropurine derivative **8** and 2-amino-6-chloropurine derivative **10**, respectively (Scheme 2). Simple modifications in the purine 6-position of **8** and **10** lead to a series of simple nucleoside derivatives, around which the chemical space was explored with the aim of establishing structure–activity relationships (SAR) for CDK2 inhibition.

The substituents in positions 2 and 6 suggested by computations (see above) were synthesized as two subseries of compounds. In the first subseries (**18a-21b**), we employed the known trans-cyclohexyldiamine as a substituent in position 2 [29] with a variety of different *p*-benzenesulfonamides in position 6. The sulfonamide-containing sidechains **13-16** were prepared in two simple steps from *p*-nitrobenzenesulfonyl chloride and corresponding amines (Scheme 3). In the second subseries (**24a-27b**), we used the phenylsulfonamidepyrrolidin sidechain in position 6, a moiety that already proved its applicability in other kinase inhibitors of heterocyclic origin,[32] and a variety of aromatic substituents in position 2.

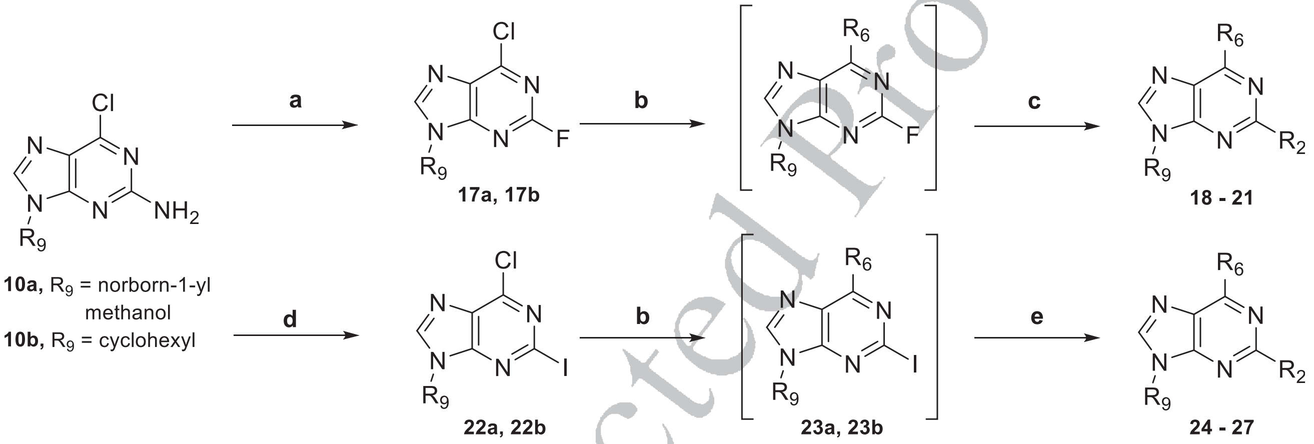
Two general routes were utilized for the synthesis of target compounds (Scheme 4). In the case of compounds bearing an aliphatic substituent in position 2, we exploited different reactivity of halogens in positions 2 and 6 as well as different nucleophilicity of anilines and aliphatic amines in S_NAr reactions. Using the Schiemann reaction, we prepared 2-fluoro-6-chloropurine derivatives bearing in position 9 cyclohexyl and 1-hydroxymethyl-norborn-4-yl, respectively. Both halogens were consecutively exchanged in one pot by sequential addition of appropriate nucleophiles - first the less reactive aniline reacted in position 6 of the purine and then the more reactive aliphatic amine, trans-1,4-cyclohexanediamine, was added to exchange the fluorine in position 2. In the case of sulfonamide **18** final deprotection of a ketal group was necessary.

In case of aromatic substituent in position 2, we used similar approach, where Sandmayer reaction was used ~~in order~~ to synthesize corresponding 2-iodo-6-chloropurine intermediates. These were first subjected to S_NAr reaction exchanging the chlorine in position 6, followed by Buchwald reaction coupling aromatic amine to position 2.

Uncorrected Proofs



SCHEME 3 (a) Amine, TEA, DCM, 0°C-rt, 12 hours; (b) Pd(OH)2, H2, MeOH—AcOEt, rt, 12 hours



SCHEME 4 (a) *i*-pentONO, HF-py, −30°C, 5 minutes 81% for **17a**, 85% for **17b;** (b) Aniline **13-16**, DIPEA, n-BuOH, 150°C, 12 hours; (c) *trans*-1,4,-diaminocyclohexane, n-BuOH, 175°C, 12 hours, 33-49% over two steps; (d) *i*-pentONO, CuI, CH2I2, THF, reflux, 4 hours, 74% for **22a**, 81% for **22b**; (e) aromatic amine, Cs2CO3, Pd2(dba)3, XantPhos, toluene-dioxane 1:1, 12 hours, 25–40% over two steps

For all the 8 synthesized compounds, a series of 9-cyclohexyl analogues was synthesized in the same manner to verify the utility of the norbornane bicycle. All the synthesized compounds were subjected to IC_{50} measurements against CDK2/cyclin E complex (Table 1).

We also preformed cytotoxicity assays of four different cell lines (HepG2, HL60, HeLa S3, and CCRF-CEM).

2.3 | The structure–activity relationship

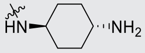
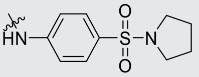
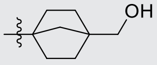
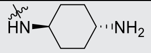
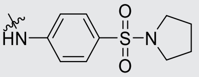
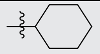
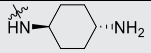
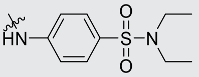
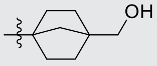
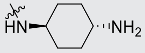
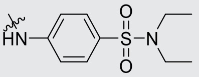
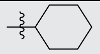
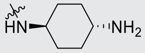
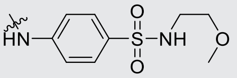
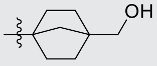
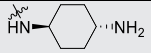
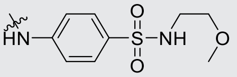
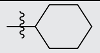
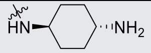
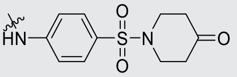
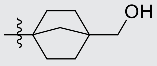
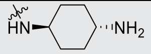
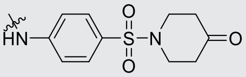
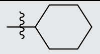
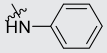
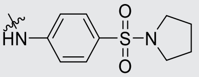
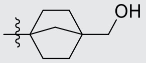
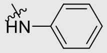
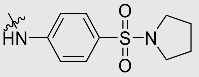
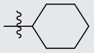
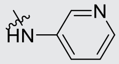
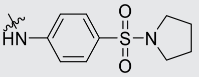
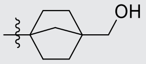
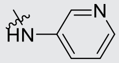
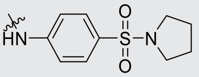
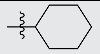
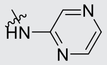
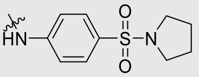
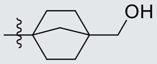
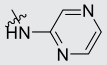
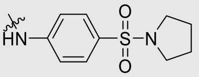
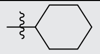
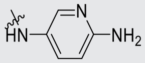
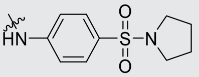
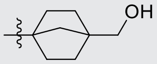
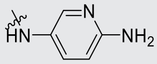
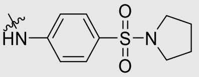
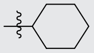
All the compounds with the 9-norbornyl substituent had values in the micro-submicromolar range, while the compounds with the 9-cyclohexyl substituent fell into two classes: (a) in case of trans-1,4-cyclohexanediamine in position 2, the compounds were submicromolar, sometimes even slightly more potent than their 9-norbornyl-substituted counterparts and (b) in case of phenylsulfonamidepyrrolidin in position 6 and various other substituents in position 2, the compounds were inactive (IC_{50} > 15 μM) (Table 1). The strongest-affinity compound **18a** (0.19 μM) combines trans-1,4-cyclohexanediamine at position 2, phenylsulfonamidepyrrolidin at position 6, and norbornyl at position 9. To obtain insight into the binding modes and affinities, we performed docking and scoring.

2.4 | Inhibitor binding modes

The standard binding mode (Figure S1-left) was observed for all the compounds. For **26a**, the reverse binding mode was also found but discarded because it cannot be used further for hit optimization. In the standard binding mode, the purine core featured two hinge-region hydrogen bonds (L83: NH…N(7) and L83:O…N(6)H) similar to roscovitine (Figure 2A) for all the compounds. Docking suggested four types of orientations of the phenylsulfonamidepyrrolidine substituent in position 6 across the compound series. We built all of them into the strongest-affinity compound **18a** (Figure 2A-E). They differed in interactions with Lys89: (a) via sulfonamide oxygen (Figure 2A,B), (b) via pyrrolidine nitrogen and Lys89 (Figure 2C), (c) shifted without any interaction (Figure 2D) and (d) shifted without any interaction, with 5-member ring slightly rotated (Figure 2E).

Uncorrected Proofs

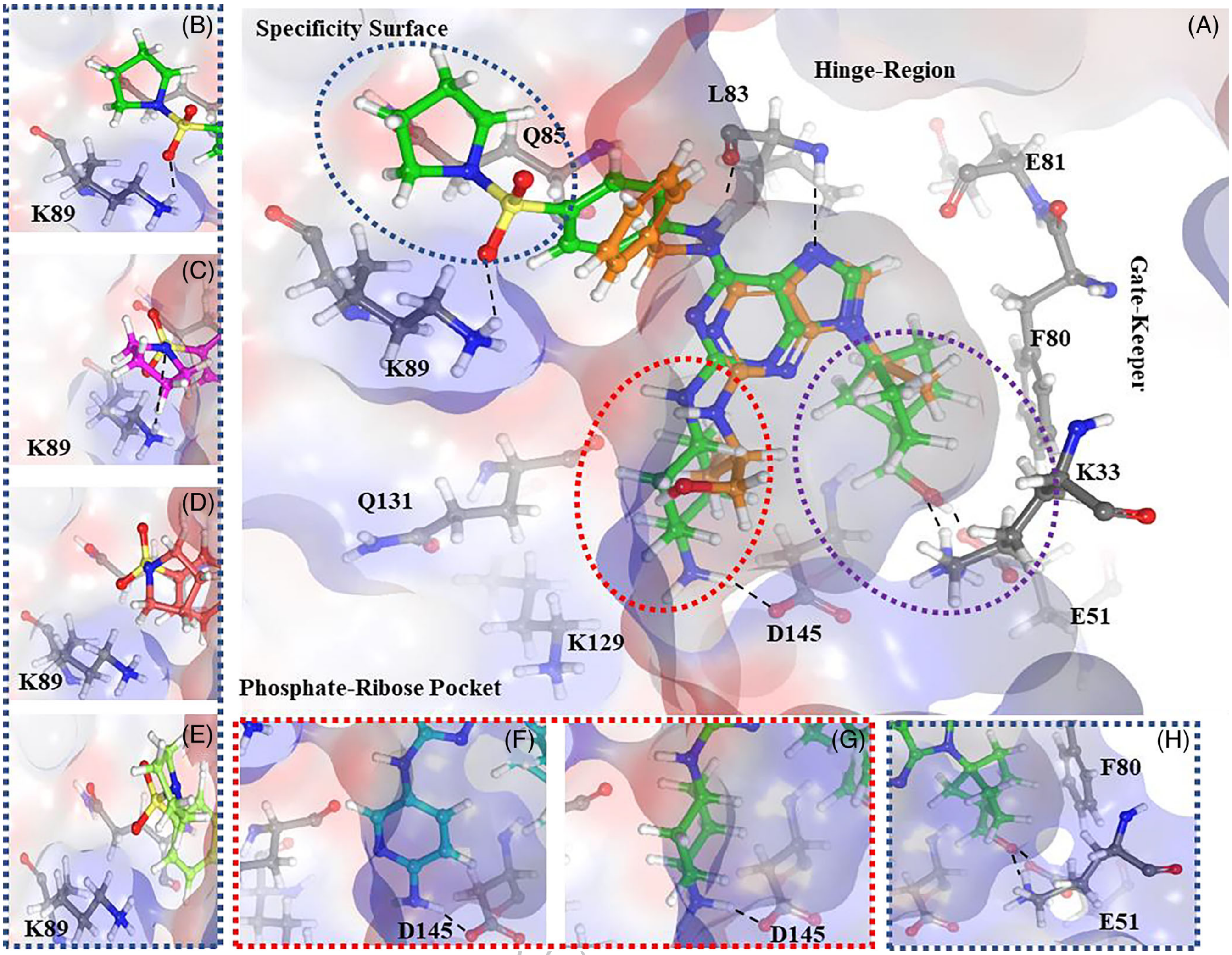
TABLE 1 2D structures of the 16 new compounds synthesized in this work and their measured activities against CDK2/cyclin E

Compound	R2	R6	R9	IC50 (μM)	Standard Deviation
18a				0.19	0.04
18b				0.43	0.12
19a				0.39	0.06
19b				0.68	0.14
20a				0.76	0.19
20b				0.58	0.04
21a				1.02	0.14
21b				0.58	0.14
24a				1.35	0.54
24b				>20	n.d.
25a				1.04	0.56
25b				>20	n.d.
26a				0.79	0.13
26b				>20	n.d.
27a				2.79	0.78
27b				14.84	3.40
Roscovitine *				0.17	0.06

[a] Reference compound.

The first type of interaction was identified as the most favorable based on the SQM/COSMO score values for compound **18a** (a) −63 kcal/mol, (b) −59 kcal/mol, (c) −52 kcal/mol, and (d) −54 kcal/mol score values were obtained. The most favorable orientation was (a) via sulfonamide oxygen and was thus modeled for all the compounds.

In position 2, the phenyl, 3-pyridyl, 2,5-pyrimidyl and cyclohexyl substituents fitted into hydrophobic cavity on one side and open to the solvent on the other side. The distal amino group (neutral for the aromatic substituents and charged for the cyclic aliphatic ones) always formed a hydrogen bond with Asp145 (Figure 2A and zoomed in Figure 2F,G). The experimental (Table 1) as well as computational



Color Figure - Print and Online

FIGURE 2 (a) The binding mode of **18a** (green sticks for carbon atoms) in CDK2 compared to the crystal structure of roscovitine with hydrogens added (orange sticks for carbon atoms; PDB code: 3DDQ). The ligand is shown in sticks and important CDK2 residues are shown as ball and sticks. Colors of atoms (C: green/orange—ligand/gray—CDK2, N:blue, O:red, S:yellow and H:white). Zoom into position 6 with different orientations distinguished by different carbon colors for the ligands: (b) the interaction between Lys 89 and sulfonamide oxygen (C: green), (c) the interaction between Lys 89 and pyrrolidine nitrogen (C: magenta) (d) shifted without any interaction (C: salmon) and (e) 5-member ring slightly rotated without any interaction (C: light green, f) Zoom into position 2 for compound **27a**, (g) Zoom into position 2 for **18a** h) Zoom into position 9 of **18a**. The figure was prepared with Maestro (Schrodinger)

Uncorrected

(Figure 3) data show that the charged salt bridge ($-NH_3^+ \ldots {}^-OOC-$) resulted in a stronger affinity than a charge-assisted hydrogen bond ($-NH_2 \ldots {}^-OOC-$). The 9-norbornyl part featured a nonpolar interaction with Phe80 gatekeeper and its terminal hydroxyl made two hydrogen bonds to the Lys33…Glu51 salt bridge (Figure 2A and zoomed in 2 ~~hours~~). The 9-cyclohexyl substituent only featured the nonpolar interaction with Phe80 gatekeeper ~~(Table 2)~~.

2.5 | QM-based binding affinities

As an estimate of binding free energy, we calculated the PM6-D3H4X/COSMO score (see Methods) of all the designed compounds (Table 1) complexed with CDK2. It must be stressed that binding entropies are not assessed fully (only solvation entropy is included via the implicit model) and thus the scores are offset by tens of kilocalories per mol to more negative values but should correlate with the experimental binding free energies under the assumption of fortuitous cancellation of errors for similar compounds.[5]

The experimental binding free energies are estimated from the measured IC_{50} values (Table 1) using the Cheng-Prusoff equation for competitive not tight-binding inhibitors.[5] Generally, IC_{50} values are measured in conditions where substrate concentration equals to the Michaelis constant of the enzyme, which transforms the equation into $\Delta G' = RT^* \ln (IC_{50}/2)$.

Figure 3 shows the plot of the PM6-D3H4X/COSMO scores vs experimental binding free energies. Two categories can be distinguished. The active compounds (IC_{50} from 0.19 to 2.79 μM) showed

FIGURE 3 PM6-D3H4X/COSMO scores plotted against experimental binding free energies expressed as RT*ln (IC_{50}/2), all in kcal/mol. Triangles are active compounds, filled circles with arrows representing the measured "worse than" for inactives

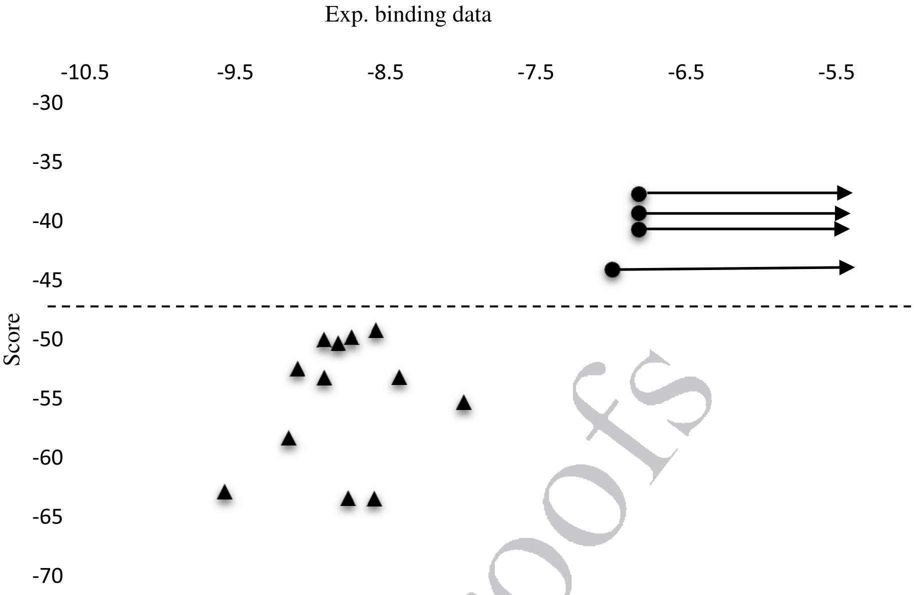


TABLE 2 Cytotoxicity of final compounds on HepG2, HL60, HeLa S3, and CCRF-CEM cells

Compound	IC50,μM			
	HepG2	HL60	HeLa S3	CCRF-CEM
18a	>10	1754 ± 0.228	>10	1572 ± 0.103
18b	>10	4866 ± 0.273	4275 ± 0.172	3752 ± 0.218
19a	>10	4347 ± 0.282	6757 ± 0.042	3131 ± 0.02
19b	6289 ± 0.226	3.53 ± 0.251	3646 ± 0.046	1653 ± 0.177
20a	>10	>10	>10	>10
20b	>10	>10	>10	>10
21a	>10	>10	>10	>10
21b	>10	4044 ± 0.321	>10	4299 ± 0.324
24a	2918 ± 0.217	2847 ± 0.194	0.983 ± 0.012	0.823 ± 0.073
24b	>10	>10	>10	>10
25a	1598 ± 0.087	1395 ± 0.045	0.839 ± 0.07	1039 ± 0.086
25b	>10	>10	>10	5265 ± 0.626
26b	>10	>10	2868 ± 0.084	4785 ± 0.089
27a	>10	>10	2442 ± 0.323	4846 ± 0.457
27b	>10	5012 ± 0.545	2.64 ± 0.279	1576 ± 0.107

high scores ranging from −64 to 49 kcal/mol (Figure 3, triangles). The inactive compounds (Figure 3, filled circles with arrows representing experimental "worse than") had the scores ranging from −44 to −37 kcal/mol. We can thus see that our score was able to separate the actives from the inactives.

The scoring results also corresponded to the SAR observed above for the 9-norbornyl vs 9-cyclohexyl group. In the first subseries with a potent trans-cyclohexyldiamine as a substituent in position 2 and a variety of different *p*-benzenesulfonamides in position 6, the IC_{50} showed a relative insensitivity to the identity of substituent 9 (2-fold change). The scores for these compounds (**18a-21b**) had high scores ranging from −63 to −50 kcal/mol. The probable structural reason is that in this subseries, the compounds are already bound to CDK2 via two strong interaction motifs: first, in position 2 where trans-cyclohexyldiamine part has electrostatic interaction with Glu145 (Figure 2G) and second, in position 6 with an electrostatic interaction between phenylsulfonpyrrolidin side chain and Lys89, which override the effect of position 9.

In the second subseries with phenylsulfonamidepyrrolidine in position 6 (compounds **24a**, **25a**, **26a**, **27a**), the 9-hydroxymethylnorbornyl group was crucial for activity. The score values ranged from −55 to −49 kcal/mol. On the contrary, the compounds with 9-cyclohexyl in this subseries (**24b**, **25b**, **26b**, **27b**) were inactive (IC_{50} > 14 μM) and their score values ranged from −44 to −37 kcal/mol (Figure 3).

Uncorrected proofs

Uncorrected Proofs

3 | CONCLUSIONS

Herein, we report the hit discovery and expansion via medicinal chemistry synthesis, activity measurements, and computational analyses of a series of carbocyclic nucleoside compounds featuring 9-hydroxymethyl norbornyl substituent as CDK2 inhibitors. The measured activities of the 16 newly synthesized compounds against CDK2/cyclin E were in the micromolar/submicromolar range and included as controls inactive compounds. The structure-activity relationships showed that structures with 9-hydroxymethylnorbornyl substituent had all submicromolar potencies.

The compounds with 9-cyclohexyl substituent fell into two classes (active/inactive) depending on the substituent in position 2. For a molecular understanding of the potency determinants, we have carried out computational modeling. Using docking and quantum-based SQM/COSMO scoring, we could separate the compounds into the actives and inactives and link the binding to individual interactions present in the binding cavity. Overall, the discovered 9-hydroxymethyl norbornyl substituent opens way to further explore it in the area of structure-based kinase design.

4 | METHODS

4.1 | Organic synthesis

Q5

~~4.1.1 | Experimental section~~

NMR spectra were recorded on Bruker Avance II 500 (^{1}H at 500 MHz) or Bruker Avance III HD (^{1}H at 400 MHz) spectrometer using DMSO-d_6 or $CDCl_3$ as a solvent and the solvent signal as a reference. Chemical shifts (δ) and coupling constants (J) were expressed in ppm and Hz, respectively. All structures were confirmed and ^{1}H and ^{13}C signals were assigned by a combination of 1D and 2D NMR (H,H-COSY, H,C-HSQC, H,C-HMBC, ROESY) techniques. Standard pulse programs from the library of the spectrometer were used; gradient selection was used in the 2D experiments. Mass spectra were measured on an LTQ Orbitrap XL using electrospray ionization (ESI). Elemental analyses were measured on Perkin Elmer CHN Analyzer 2400, Series II Sys or on SPECTRO iQ II. Microwave syntheses were carried out in a CEM Discover instrument with a single-mode cavity and focused microwave heating (microwave power supply 0-300 W, IR temperature sensor, sealed vessel mode). Column chromatography (both normal and reverse phase) was performed on a 40-60 μm silica gel using ISCO flash chromatography system or standard glass columns. Purity of all prepared compounds was higher than 98% unless stated otherwise.

Ethyl 4-[(benzyloxy)amino]bicyclo[2.2.1]heptane-1-carboxylate (**6**): A solution of **5** [15] (17 g, 46.2 mmol) in dry toluene (500 mL, distilled from sodium) was deoxygenated three times using the freeze-pump-thaw degassing method, and then heated to reflux. A solution of Bu_3SnH (37.3 mL, 138.5 mmol) and AIBN (1 g) in dry and deoxygenated toluene (150 mL) was added dropwise to this solution (3 hours), and the resulting reaction mixture was refluxed for further 1 hour. After cooling down, the reaction was quenched with careful addition of methyliodide (30 mL), and volatiles were evaporated and adsorbed on silica. Flash column chromatography (5-30% ethyl acetate in petroleum ether) afforded **3** (8.0 g, 60%) as a clear oil. Spectral characteristics match those described in literature [15].

(4-Aminobicyclo[2.2.1]hept-1-yl)methanol (**7**): To a solution of **6** (7.4 g, 25.6 mmol) in dry diglyme (200 mL) was added BH_3-THF complex (1 M solution in THF, 121 mL) and the mixture was heated in a sealed pressure vessel on 110°C for 48 hours. After cooling to RT, reaction was quenched by careful addition of water (10 mL) and volatiles were evaporated. The residue was adsorbed on silica from ethanol (50 mL) and purified by column chromatography (DCM to DCM: EtOH:(3 M NH_3-EtOH) = 7:2:1 to afford **7** (3.43 g, 95%) as clear oil which solidifies on standing. ^{1}H NMR: 1.12 (s, 2H, H-7), 1.16-1.26 (m, 2H, H-2endo, H-6endo), 1.34-1.50 (m, 4H, H-3, H-5), 1.54-1.59 (m, 2H, H-2exo, H-6exo), 3.34 (s, 2H, CH_2O). ^{13}C NMR: 32.28 (C-2, C-6), 37.40 (C-3, C-5), 48.20 (C-7), 48.98 (C-1), 62.40 (C-4), 65.17 (CH_2O). ESI MS *m/z* (%): 142.1 (100) [M + H]; HRMS ESI ($C_8H_{16}ON$) calculated: 142.12264; found: 142.12256.

[4-(6-Chloro-9H-purin-9-yl)bicyclo[2.2.1]hept-1-yl]methanol (**8**): To a solution of amine **7** (282 mg, 2 mmol) in *n*-BuOH (10 mL), was added 4,6-dichloro-5-formamidopyrimidine (460 mg, 2.4 mmol) and DIPEA (1.05 mL, 6 mmol) and the reaction mixture was microwave irradiated in a sealed vessel on 160°C for 2 hours. Flash chromatography (1-2% methanol in ethyl acetate) followed by crystallization from toluene −cyclohexane mixture afforded **8** (335 mg, 60%) as white crystals (m.p. = 176.5-178°C). ^{1}H NMR: 1.41-1.49 (m, 2H, H-2endo, H-6endo), 1.77-1.85 (m, 2H, H-2exo, H-6exo), 2.05-2.14 (m, 2H, H-3exo, H-5exo), 2.13 (bs, 2H, H-7), 2.24-2.32 (m, 2H, H-3endo, H-5endo), 3.51 (d, 2H, $J_{CH2\text{-}OH}$ = 5.3, CH_2O), 4.68 (t, 1H, $J_{OH\text{-}CH2}$ = 5.3, OH), 8.68 (s, 1H, H-8′), 8.77 (s, 1H, H-2′). ^{13}C NMR: 31.29 (C-2, C-6), 34.55 (C-3, C-5), 43.57 (C-7), 48.51 (C-1), 64.20 (CH_2O), 66.22 (C-4), 131.72 (C-5′), 146.54 (C-8′), 149.37 (C-6′), 151.24 (C-2′), 152.35 (C-4′). ESI MS *m/z* (%): 279.1 (100) [M + H], 301.1 (86) [M + Na]. Anal. calc. For $C_{13}H_{15}N_4OCl$ x 1/5 C_7H_8 (297.17): C 58.20, H 5.63, N 18.85, Cl 11.93; found: C 57.96, H 5.63, N 18.73, Cl 11.73.

[4-(6-Amino-9H-purin-9-yl)bicyclo[2.2.1]hept-1-yl]methanol (**2**): A solution of **8** (150 mg, 0.53 mmol) in ethanolic ammonia (3.5 M, 3 mL) was heated in a microwave reactor at 120°C for 30 minutes. Product was isolated by flash chromatography (5-20% methanol in ethyl acetate) and subsequent crystallization from aqueous methanol. Yield 128 mg, 91%, colorless needles (m.p. = 225-226°C). ^{1}H NMR: 1.38-1.46 (m, 2H, H-2endo, H-6endo), 1.74-1.82 (m, 2H, H-2exo, H-6exo), 2.06 (bs, 2H, H-7), 2.03-2.11 (m, 2H, H-3exo, H-5exo), 2.16-2.24 (m, 2H, H-3endo, H-5endo), 3.50 (d, 2H, $J_{CH2\text{-}OH}$ = 5.3, CH_2O), 4.62 (t, 1H, $J_{OH\text{-}CH2}$ = 5.3, OH), 7.15 (bs, 2H, NH_2), 8.07 (s, 1H, H-8′), 8.11 (s, 1H, H-2′). ^{13}C NMR: 31.38 (C-2, C-6), 34.60 (C-3, C-5), 43.61 (C-7), 48.48 (C-1), 64.39 (CH_2O), 65.37 (C-4), 119.81 (C-5′), 139.59 (C-8′), 150.11 (C-4′), 152.17 (C-2′), 156.29 (C-6′). ESI MS *m/z* (%): 260.2 (100) [M + H], 282.2 (32) [M + Na]. Anal. calc. For $C_{13}H_{17}N_5O$ (259.31): C 60.21, H 6.61, N 27.01; found: C 59.92, H 6.60, N 26.80.

{4-[6-(Cyclopropylamino)-9H-purin-9-yl]bicyclo[2.2.1]hept-1-yl}methanol (**1**): A solution of **8** (150 mg, 0.53 mmol) and cyclopropylamine (367 μL, 5.3 mmol) in ethanol (5 mL) was heated in a microwave reactor at 140°C for 20 minutes. Volatiles were evaporated, crude product was adsorbed on silica and purified by flash chromatography (1-5% methanol in ethyl acetate), and subsequent crystallization from toluene-cyclohexane mixture to afford **1** (146 mg, 91%) as white crystals (m.p. = 157-158°C). ^{1}H NMR: 0.58-0.62 and 0.68-0.73 (m, 2H, CH_2-cyclop), 1.40-1.46 (m, 2H, H-2endo, H-6endo), 1.73-1.82 (m, 2H, H-2exo, H-6exo), 2.06 (bs, 2H, H-7), 2.03-2.12 (m, 2H, H-3exo, H-5exo), 2.15-2.23 (m, 2H, H-3endo, H-5endo), 3.02 (bs, 1H, CH-cyclop), 3.50 (d, 2H, $J_{CH2\text{-}OH}$ = 5.3, CH_2O), 4.62 (t, 1H, $J_{OH\text{-}CH2}$ = 5.3, OH), 7.81 (bs, 1H, NH), 8.07 (s, 1H, H-8′), 8.22 (bs, 1H, H-2′). ^{13}C NMR: 6.54 (CH_2-cyclop), 31.38 (C-2, C-6), 34.61 (C-3, C-5), 43.63 (C-7), 48.47 (C-1), 64.38 (CH_2O), 65.39 (C-4), 120.18 (C-5′), 139.42 (C-8′), 149.7 (C-4′), 152.07 (C-2′), 155.83 (C-6′). ESI MS *m/z* (%): 300.2 (100) [M + H], 322.2 (3) [M + Na]. Anal. calc. For $C_{16}H_{21}N_5O$ (299.37): C 64.19, H 7.07, N 23.39; found: C 64.11, H 7.02, N 23.40.

{4-[6-(Dimethylamino)-9H-purin-9-yl]bicyclo[2.2.1]hept-1-yl}methanol (**9**): A solution of **8** (150 mg, 0.53 mmol) in DMF (3 mL) was subjected to microwave irradiation (sealed vessel, 200°C, 2 minutes). Volatiles were evaporated, crude product was adsorbed on silica and purified by flash chromatography (1-5% methanol in ethyl acetate), and subsequent crystallization from toluene - ethyl acetate mixture. Yield 135 mg, 87%, colorless crystals (m.p. = 155-156°C). ^{1}H NMR: 1.39-1.46 (m, 2H, H-2endo, H-6endo), 1.73-1.82 (m, 2H, H-2exo, H-6exo), 2.00-2.09 (m, 2H, H-3exo, H-5exo), 2.06 (bs, 2H, H-7), 2.19-2.26 (m, 2H, H-3endo, H-5endo), 3.44 (bs, 6H, CH_3), 3.50 (d, 2H, $'J_{CH2\text{-}OH}$ = 5.3, CH_2O), 4.63 (t, 1H, $J_{OH\text{-}CH2}$ = 5.3, OH), 8.08 (s, 1H, H-8′), 8.19 (s, 1H, H-2′). ^{13}C NMR: 31.37 (C-2, C-6), 34.50 (C-3, C-5), 37.97 (CH_3), 43.60 (C-7), 48.42 (C-1), 64.38 (CH_2O), 65.42 (C-4), 120.36 (C-5′), 138.47 (C-8′), 150.95 (C-4′), 151.48 (C-2′), 154.54 (C-6′). ESI MS *m/z* (%): 288.2 (100) [M + H]. Anal. calc. For $C_{15}H_{21}N_5O$ (287.36): C 62.70, H 7.37, N 24.37; found: C 62.43, H 7.44, N 24.00.

[4-(2-Amino-6-chloro-9H-purin-9-yl)bicyclo[2.2.1]hept-1-yl]methanol (**10**): To a solution of amine **7** (500 mg, 3.5 mmol) in *n*-BuOH (20 mL), was added 2-amino-4,6-dichloro-5-formamidopyrimidine (870 mg, 4.2 mmol) and DIPEA (1.83 mL, 10.5 mmol) and the reaction mixture was microwave irradiated in a sealed vessel on 160°C for 2 hours. Flash chromatography (1%-5% methanol in ethyl acetate) followed by crystallization from ethyl acetate–acetone mixture afforded **10** (846 mg, 81%) as pink crystals (m.p. = 150°C-151°C). ^{1}H NMR: 1.38-1.45 (m, 2H, H-2, H-6endo), 1.72-1.79 (m, 2H, H-2, H-6exo), 2.00-2.06 (m, 2H, H-3exo, H-5exo), 2.04 (bs, 2H, H-7), 2.14-2.23 (m, 2H, H-3endo, H-5endo), 3.46-3.51 (m, 2H, CH_2O), 4.63-4.68 (m, 1H, OH), 6.82 (bs, 2H, NH_2), 8.08 (s, 1H, H-8′). ^{13}C NMR: 31.30 (C-2, C-6), 34.28 (C-3, C-5), 43.34 (C-7), 48.44 (C-1), 64.31 (CH_2O), 65.45 (C-4), 124.27 (C-5′), 142.15 (C-8′), 149.64 (C-6′), 154.55 (C-4′), 159.49 (C-2′). ESI MS *m/z* (%): 294.2 (37) [M + H], 316.2 (19) [M + Na], 608.8 (100) [2 M + Na]; HRMS ESI ($C_{13}H_{17}ON_5Cl$) calculated: 294.11161; found: 294.11167. Anal. calc. For $C_{13}H_{16}ClN_5O$ (293.75): C 53.15, H 5.49, N 23.84, Cl 12.07; found: C 53.33, H 5.56, N 23.61, Cl 12.30.

(4-(2-amino-6-chloro-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**10b**): Compound was prepared according to a published procedure reported by author [31].

[4-(2,6-Diamino-9H-purin-9-yl)bicyclo[2.2.1]hept-1-yl]methanol (**11**): A solution of **10** (120 mg, 0.41 mmol) in ethanolic ammonia (3.5 M, 3 mL) was heated in a microwave reactor at 120°C for 30 minutes. Product was isolated by flash chromatography (15%-30% methanol in ethyl acetate) and subsequent crystallization from aqueous methanol. Yield 73 mg, 65%, pale orange crystals (m.p. = 268°C-269°C). ^{1}H NMR: 1.33-1.44 (m, 2H, H-2, H-6endo), 1.69-1.79 (m, 2H, H-2, H-6exo), 2.00 (bs, 2H, H-7), 1.98-2.07 (m, 2H, H-3exo, H-5exo), 2.09-2.17 (m, 2H, H-3endo, H-5endo), 3.48 (d, 2H, $J_{CH2,OH}$ = 5.3, CH_2O), 4.61 (t, 1H, $J_{OH,CH2}$ = 5.3, OH), 5.64 (bs, 2H, 2′-NH_2), 6.58 (s, 2H, 6′-NH_2), 7.64 (s, 1H, H-8′). ^{13}C NMR: 31.39 (C-2, C-6), 34.40 (C-3, C-5), 43.44 (C-7), 48.39 (C-1), 64.47 (CH_2O), 64.83 (C-4), 114.32 (C-5′), 136.17 (C-8′), 152.41 (C-4′), 156.32 (C-6′), 159.90 (C-2′). ESI MS *m/z* (%): 275.3 (100) [M + H]; HRMS ESI ($C_{13}H_{19}ON_6$) calculated: 275.16149; found: 275.16145. Anal. calc. For $C_{13}H_{18}N_6O$ (274.32): C 56.92, H 6.61, N 30.64; found: C 57.00, H 6.49, N 30.64.

{4-[2-Amino-6-(cyclopropylamino)-9H-purin-9-yl]bicyclo[2.2.1]hept-1-yl}methanol (**3**): A solution of **10** (588 mg, 2 mmol) and cyclopropylamine (1.39 mL, 20 mmol) in ethanol (20 mL) was heated in a microwave reactor at 140°C for 30 min. Volatiles were evaporated, crude product was adsorbed on silica and purified by flash chromatography (5%-10% methanol in ethyl acetate), and subsequent crystallization from acetone to afford **3** (502 mg, 80%) as off-white crystals (m.p. = 247-248°C). ^{1}H NMR: 0.53 to 0.61 and 0.61 to 0.68 (m, 2H, CH_2-cyclop), 1.34-43 (m, 2H, H-2, H-6endo), 1.68-1.79 (m, 2H, H-2, H-6exo), 2.00 (bs, 2H, H-7), 1.96-2.07 (m, 2H, H-3exo, H-5exo), 2.09-2.19 (m, 2H, H-3endo, H-5endo), 3.01 (bs, 1H, CH-cyclop), 3.48 (d, 2H, $J_{CH2,OH}$ = 5.0, CH_2O), 4.61 (t, 1H, $J_{OH,CH2}$ = 5.2, OH), 5.70 (bs, 2H, NH_2), 7.18 (bs, 1H, NH), 7.63 (s, 1H, H-8′). ^{13}C NMR: 6.61 (CH_2-cyclop), 31.40 (C-2, C-6), 34.41 (C-3, C-5), 43.46 (C-7), 48.39 (C-1), 64.48 (CH_2O), 64.84 (C-4), 114.58 (C-5′), 135.91 (C-8′), 151.9 (C-4′), 156.10 (C-6′), 159.82 (C-2′).ESI MS *m/z* (%): 315.3 (100) [M + H]; HRMS ESI ($C_{16}H_{23}ON_6$) calculated: 315.19279; found: 315.19268. Anal. calc. For $C_{16}H_{22}N_6O$ (314.39): C 61.13, H 7.05, N 26.73; found: C 60.95, H 7.02, N 26.81.

2-Amino-9-[4-(hydroxymethyl)bicyclo[2.2.1]hept-1-yl]-1,9-dihydro-6H-purin-6-one (**12**): A solution of **10** (120 mg, 0.41 mmol) in TFA–water mixture (2:1, 6 mL) was stirred at RT overnight. Volatiles were evaporated and crude product was codistilled with ethanol (3 × 10 mL), NH_4OH (10 mL), and ethanol (2 × 10 mL), adsorbed on silica gel and purified by flash chromatography (mobile phase: 15-30% methanol in ethyl acetate). Subsequent crystallization from water-methanol mixture afforded **12** (78 mg, 69%) as light brown powder (m.p. > 360°C [decomp.]). ^{1}H NMR: 1.33-1.43 (m, 2H, H-2, H-6endo), 1.67-1.79 (m, 2H, H-2, H-6exo), 1.98 (bs, 2H, H-7), 1.93-2.03 (m, 2H, H-3exo, H-5exo), 2.10-2.19 (m, 2H, H-3endo, H-5endo), 3.47 (d, 2H, $J_{CH2,OH}$ = 5.3, CH_2O), 4.62 (t, 1H, $J_{OH,CH2}$ = 5.3, OH), 6.36 (bs, 2H, NH_2), 7.62 (s, 1H, H-8′), 10.57 (bs, 1H, H-1′). ^{13}C NMR: 31.38 (C-2, C-6), 34.53 (C-3, C-5), 43.50 (C-7), 48.37 (C-4), 64.40 (CH_2O), 65.13 (C-1), 117.86 (C-5′), 136.17 (C-8′), 151.75 (C-4′), 152.89 (C-2′), 157.04

Uncorrected Proofs

(C-6′). ESI MS *m/z* (%): 276.2 (14) [M + H], 298.2 (100) [M + Na]; HRMS ESI ($C_{13}H_{18}O_2N_5$) calculated: 276.14550; found: 276.14549. Anal. calc. For $C_{13}H_{17}N_5O_2$ (275.31): C 56.31, H 6.22, N 25.44; found: C 56.24, H 6.30, N 25.31.

4.2 | General procedure for the preparation of sulfonamides 13-16

A solution of 4-nitrobenzenesulfonyl chloride (665 mg, 3 mmol) in DCM (30 mL) was added dropwise to a stirred solution of amine (3.05 mmol) and triethylamine (418 μL, 3 mmol) in DCM (30 mL) at 0°C and stirred at room temperature for 12 hours. Reaction mixture was diluted with DCM (50 mL) and washed with water, organic phase was dried with sodium sulfate and exaporated in vacuo. This intermediate was, without further purification, dissolved in a MeOH-AcOEt mixture (1:1, 30 mL), Pd/C (10%, 100 mg) was added, and reaction mixture was hydrogenated under balloon for 24 hours. Catalyst was filtered off on a celite pad, and product was purified by flash column chromatography.

4-(Pyrrolidin-1-ylsulfonyl)aniline (**13**): Chromatography: ethyl acetate in petrol ether (40-80%), yield 648 mg (95% over 2 steps). Analytical data are consistent with literature.[33]

4-Amino-N,N-diethylbenzenesulfonamide (**14**): Chromatography: ethyl acetate in petrol ether (30-100%), yield 560 mg (82% over 2 steps). Analytical data are consistent with literature.[34]

4-((1,4-Dioxa-8-azaspiro[4.5]decan-8-yl)sulfonyl)aniline (**15**): Chromatography: ethyl acetate in petrol ether (40-100%), yield 720 mg (80% over 2 steps). ^1H NMR: δ 1.61-1.67 (m, 4H, H-6′, H-10′), 2.86-2.92 (m, 4H, H-7′, H-9′), 3.81 (s, 4H, H-2′, H-3′), 6.08 (s, 2H, NH_2), 6.61-6.61 (m, 2H, H-2), 7.32-7.38 (m, 2H, H-3). ^{13}C NMR: δ 33.68 (C-6′, C-10′), 44.39 (C-7′, C-9′), 63.75 (C-2′, C-3′), 105.33 (C-5′), 112.69 (C-2), 119.71 (C-4), 129.48 (C-3), 153.19 (C-1). ESI MS *m/z* (%): 299.1 (4) [M + H], 321.1 (51) [M + Na], 619.2 (100) [2 M + Na]; HRMS ESI ($C_{13}H_{18}O_4N_2NaS$) calculated: 321.08795; found: 321.08806.

4-Amino-N-(2-methoxyethyl)benzenesulfonamide (**16**): Chromatography: ethyl acetate in petrol ether (40-100%), yield 570 mg (83% over 2 steps). Analytical data are consistent with literature.[35]

(4-(6-Chloro-2-fluoro-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**17a**): **10a** (100 mg, 0.34 mmol), placed in a plastic falcon tube, was dissolved in 60% HF-pyridine (2 mL) at −50°C. Isoamylnitrite (104 μL, 0.51 mmol) was added, and reaction mixture was stirred at −30°C for 5 minutes, after which it was poured on ice and product was extracted with chloroform. Organic phase was washed with sat. $NaHCO_3$ and water and flash chromatography (5-10% methanol in DCM) afforded **17a** (82 mg, 81%) as a white amorphous solid. ^1H NMR: δ 1.41-1.49 (m, 2H, H-2b, H-6endo), 1.75-1.84 (m, 2H, H-2a, H-6exo), 2.00-2.12 (m, 4H, H-3exo, H-5exo, H-7), 2.20-2.29 (m, 2H, H-3endo, H-5endo), 3.50 (d, 2H, $J_{CH2,OH}$ = 5.3, CH_2O), 4.68 (t, 1H, $J_{OH,CH2}$ = 5.3, OH), 8.69 (s, 1H, H-8′). ^{13}C NMR: δ 31.00 (C-2, C-6), 34.27 (C-3, C-5), 43.23 (C-7), 48.29 (C-1), 63.92 (CH_2O), 66.07 (C-4), 130.61 (d, $J_{5',F}$ = 4.8, C-5′), 147.31 (d, $J_{8',F}$ = 3.0, C-8′), 150.37 (d, $J_{6',F}$ = 18.3, C-6′), 153.82 (d, $J_{4',F}$ = 17.6, C-4′), 155.59 (d, $J_{2',F}$ = 212.5, C-2′). ESI MS *m/z* (%): 297.3 (100) [M + H]; HRMS ESI ($C_{13}H_{15}ON_4ClF$) calculated: 297.09129; found: 297.09136.

6-Chloro-9-cyclohexyl-2-fluoro-9H-purine (**17b**): **10b** (100 mg, 0.4 mmol), placed in a plastic falcon tube, was dissolved in 60% HF-pyridine (2 mL) at −50°C. Isoamylnitrite (121 μL, 0.6 mmol) was added, and reaction mixture was stirred at −30°C for 5 minutes, after which it was poured on ice and product was extracted with chloroform. Organic phase was washed with sat. $NaHCO_3$ and water and flash chromatography (30-100% ethyl acetate in petrol ether) afforded **17b** (86 mg, 85%) as a white amorphous solid. ^1H NMR: δ 1.27 (tt, 1H, J_{GEM} = 13.0, $J_{4'ax,3'eq}$ = 3.6, H-4′ax), 1.52-1.38 (m, 2H, H-3′ax), 1.64-1.76 (m, 1H, H-4′eq), 1.98-1.80 (m, 4H, H-2′eq, H-3′eq), 1.99-2.07 (m, 2H, H-2′ax), 4.41 (tt, 1H, $J_{1',2'a}$ = 11.9, $J_{1',2'b}$ = 3.8, H-1′), 8.81 (s, 1H, H-8). ^{13}C NMR: δ 24.82 (C-4′), 25.07 (C-3′), 31.97 (C-2′), 55.07 (C-1′), 130.31 (d, $J_{5,F}$ = 4.8, C-5), 146.97 (d, $J_{8,F}$ = 3.1, C-8), 150.39 (d, $J_{6,F}$ = 18.3, C-6), 153.47 (d, $J_{4,F}$ = 17.5, C-4), 156.07 (d, $J_{2,F}$ = 213.5, C-2). ESI MS *m/z* (%): 255.2 (100) [M + H]; HRMS ESI ($C_{11}H_{13}ON_4ClF$) calculated: 255.08073; found: 255.08080.

4.3 | General procedure for the preparation of compounds 18-21

A mixture of **17a** or **17b**, corresponding aniline (**13-16**, 1.2 eq), DIPEA (2 eq), and *n*-BuOH (2-5 mL) was heated in a sealed pressure vessel to 150°C for 12 hours. When TLC showed complete disappearance of the starting material, *trans*-1,4-cyclohexandiamine (3 eq) was added and the reaction mixture was heated to 175°C for 12 hours. Product was purified by reverse phase column chromatography in H_2O-MeCN gradient (10-100% MeCN, 0,1% TFA) and freeze-dried.

(4-(2-(([1r,4r]-4-Aminocyclohexyl)amino)-6-((4-[pyrrolidin-1-ylsulfonyl]phenyl)amino)-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**18a**): Starting from **17a** (80 mg, 0.31 mmol), **13** (85 mg, 0.38 mmol), DIPEA (109 mL, 0.63 mmol), *n*-BuOH (2 mL) in stage 1 and *trans*-1,4-cyclohexandiamine (108 mg, 0.94 mmol) in stage 2, yield 69 mg (32%) as a TFA salt. ^1H NMR: δ 1.31-1.51 (m, 6H, H-3″´b, H-3b, H-2b), 1.65 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.77 (m, 2H, H-3a), 2.05 (s, H-5), 2.07-2.22 (m, 6H, H-2a, H-2b, H-2″´a), 3.03 (bm, H-4″´), 3.13 (m, 4H, $N\mathbf{CH_2}CH_2$), 3.51 (s, 2H, $\mathbf{CH_2}OH$), 3.62 (bm, 1H, H-1″´), 6.74 (bs, 2´-NH), 7.69 (m, 2H, H-2″ or H-3″), 7.69 (bd, 3H, 4″´-NH_3^+), 7.98 (s, 1H, H-8′), 8.30 (m, 2H, H-2″ or H-3″), 9.98 (s, 1H, 6´-NH). ^{13}C NMR: δ 24.86 ($NCH_2\mathbf{CH_2}$), 29.53 (C-3″´), 30.09 (C-2″´), 31.44 (C-3), 34.31 (C-2), 43.47 (C-5), 47.98 ($N\mathbf{CH_2}CH_2$), 48.37 (C-4), 49.12 (C-4″´), 49.82 (C-1″´), 64.49 ($\mathbf{CH_2}OH$), 65.28 (C-1), 114.19 (C-5′), 119.39 (C-2″ or C-3″), 128.20 (C-2″ or C-3″), 127.8* (C-1″ or C-4″), 137.6* (C-8′), 144.83 (C-1″ or C-4″), 152.7 (C-4′), 151.56 (C-6′), 157.80 (C-2′). ESI MS *m/z* (%): 581.3 (100) [M + H], 603.3 (15) [M + Na]; HRMS ESI ($C_{29}H_{41}O_3N_8S$) calculated: 581.30168; found: 581.30174.

N^2-([1r,4r]-4-Aminocyclohexyl)-9-cyclohexyl-N^6-(4-[pyrrolidin-1-ylsulfonyl]phenyl)-9H-purine-2,6-diamine (**18b**): Starting from **17b** (80 mg, 0.31 mmol), **13** (85 mg, 0.38 mmol), DIPEA (109 mL, 0.63 mmol), *n*-BuOH (2 mL) in stage 1 and *trans*-1,4-cyclohexandiamine

Uncorrected Proofs

(108 mg, 0.94 mmol) in stage 2, yield 85 mg (42%) as a TFA salt. ^{1}H NMR: δ 1.16-1.56 (m, 7H, H-3”b, H-4”b, H-2‴b, H-3‴b), 1.61-1.67 (m, 4H, NCH$_2$**CH$_2$**), 1.67-1.74 (m, 1H, H-4”a), 1.81-1.93 (m, 4H, H-2”b, H-3”a), 1.94-2.05 (m, 4H, H-2”a, H-3‴a), 2.05-2.17 (m, 2H, H-2‴a), 2.97-3.10 (m, 1H, H-4‴), 3.08-3.19 (m, 4H, N**CH$_2$**CH$_2$), 3.60-3.73 (m, 1H, H-1‴), 4.15-4.27 (m, 1H, H-1″), 6.89 (bs, 1H, 2-NH), 7.67-7.73 (m, 2H, H-3′), 7.83 (bs, 3H, NH$_2$), 8.10 (s, 1H, H-8), 8.24-8.34 (m, 2H, H-2′), 10.05 (s, 1H, 6-NH). ^{13}C NMR: δ 24.68 (NCH$_2$**CH$_2$**), 24.84 (C-4″), 25.23 (C-3″), 29.32 (C-3‴), 30.01 (C-2‴), 32.06 (C-2″), 47.81 (N**CH$_2$**CH$_2$), 48.86 (C-4‴), 49.6 (C-1‴), 53.2 (C-1″), 113.7 (C-5), 119.15 (C-2′), 127.9 (C-3′), 128.01 (C-4′), 137.00 (C-8), 144.63 (C-1′), 151.32 (C-6), 151.60 (C-4), 157.8 (C-2). ESI MS *m/z* (%): 539.5 (100) [M + H], 561.3 (12) [M + Na]; HRMS ESI ($C_{27}H_{39}O_2N_8S$) calculated: 539.29112; found: 539.29121.

4-((2-(([1r,4r]-4-Aminocyclohexyl)amino)-9-(4-(hydroxymethyl)bicyclo[2.2.1]heptan-1-yl)-9H-purin-6-yl)amino)-N,N-diethylbenzenesulfonamide (**19a**): Starting from **17a** (95 mg, 0.32 mmol), **14** (108 mg, 0.47 mmol), DIPEA (137 mL, 0.79 mmol), *n*-BuOH (4 mL) in stage 1 and *trans*-1,-4-cyclohexandiamine (109 mg, 0.96 mmol) in stage 2. Yield 96 mg (43%) as a TFA salt. ^{1}H NMR: δ 1.04 (t, 6H, $J_{CH3,CH2}$ = 7.1, NCH$_2$**CH$_3$**), 1.28-1.53 (m, 6H, H-3”b, H-5”endo, H-2‴b, H-3‴b), 1.71-1.84 (m, 2H, H-3”a, H-5”exo), 1.97-2.25 (m, 10H, H-2″, H-6″, H-7″, H-2‴a, H-3‴a), 2.98-3.10 (m, 1H, H-4‴), 3.15 (q, 4H, $J_{CH2,CH3}$ = 5.8, N**CH$_2$**CH$_3$), 3.51 (s, 2H, CH$_2$O), 3.56-3.67 (m, 1H, H-1‴), 6.92 (bs, 1H, 2′-NH), 7.65-7.72 (m, 2H, H-3), 7.92 (bs, 3H, NH$_2$), 8.19 (s, 1H, H-8′), 8.22-8.27 (m, 2H, H-2), 10.05 (s, 1H, 6′-NH). ^{13}C NMR: δ 14.14 (NCH$_2$**CH$_3$**), 29.34 and 29.88 (C-2‴, C-3‴), 31.24 (C-3″, C-5″), 34.05 (C-2″, C-6″), 41.81 (N**CH$_2$**CH$_3$), 43.25 (C-7″), 48.21 (C-4″), 48.90 (C-1‴), 49.71 (C-4‴), 64.26 (CH$_2$O), 65.44 (C-1″), 112.45 (C-5′), 119.46 (C-2), 127.44 (C-3), 131.97 (C-4), 137.33 (C-8′), 144.02 (C-1), 150.95 (C-6′), 152.10 (C-4′), 157.73 (C-2′). ESI MS *m/z* (%): 583.4 (100) [M + H]; HRMS ESI ($C_{29}H_{43}O_3N_8S$) calculated: 583.31733; found: 583.31732.

4-((2-(([1r,4r]-4-Aminocyclohexyl)amino)-9-cyclohexyl-9H-purin-6-yl)amino)-N,N-diethylbenzenesulfonamide (**19b**): Starting from **17b** (100 mg, 0.39 mmol), **14** (108 mg, 0.47 mmol), DIPEA (137 mL, 0.79 mmol), *n*-BuOH (mL) in stage 1 and *trans*-1,4-cyclohexandiamine (135 mg, 1.18 mmol) in stage 2. Yield 123 mg (49%) as a TFA salt. ^{1}H NMR: δ 1.03 (t, 6H, $J_{CH3,CH2}$ = 7.1, NCH$_2$**CH$_3$**), 1.16-1.25 (m, 1H, H-4”b), 1.27-1.44 (m, 4H, H-3”b, H-2‴b), 1.45-1.58 (m, 2H, H-3‴b), 1.61-1.70 (m, 1H, H-4”a), 1.75-2.20 (m, 10H, H-2″, H-3”a, H-2‴a, H-3‴a), 2.95-3.06 (m, 1H, H-4‴), 3.13 (q, 4H, $J_{CH2,CH3}$ = 7.0, N**CH$_2$**CH$_3$), 3.60-3.75 (m, 1H, H-1‴), 4.14-4.28 (m, 1H, H-1″), 6.93 (bs, 1H, 2′-NH), 7.59-7.81 (m, 2H, H-3), 8.13 (s, 1H, H-8), 8.18 (bs, 3H, NH$_2$), 8.24-8.34 (m, 2H, H-2), 10.05 (bs, 1H, 6′-NH). ^{13}C NMR: δ 14.19 (NCH$_2$**CH$_3$**), 24.99 (C-4″), 25.35 (C-3″), 29.44 (C-3‴), 30.23 (C-2‴), 32.18 (C-2″), 41.95 (N**CH$_2$**CH$_3$), 49.08 (C-4‴), 49.83 (C-1‴), 53.5 (C-1″), 113.29 (C-5′), 119.52 (C-2), 127.53 (C-3), 132.01 (C-4), 137.06 (C-8′), 144.36 (C-1), 151.45 (C-4′), 157.96 (C-2′). ESI MS *m/z* (%): 541.4 (100) [M + H], 563.4 (21) [M + Na]; HRMS ESI ($C_{27}H_{41}O_2N_8S$) calculated: 541.30677; found: 541.30679.

4-((2-(([1r,4r]-4-aminocyclohexyl)amino)-9-(4-(hydroxymethyl)bicyclo[2.2.1]heptan-1-yl)-9H-purin-6-yl)amino)-N-(2-methoxyethyl)benzenesulfonamide (**20a**): Starting from **17a** (100 mg, 0.34 mmol), **16** (93 mg, 0.4 mmol), DIPEA (117 mL, 0.67 mmol), *n*-BuOH (3 mL) in stage 1 and *trans*-1,4-cyclohexandiamine (115 mg, 1 mmol) in stage 2. Yield 77 mg (33%) as a TFA salt. ^{1}H NMR: δ 1.28-1.53 (m, 6H, H-3”b, H-5”endo, H-2‴b, H-3‴b), 1.72-1.84 (m, 2H, H-3”a, H-5”exo), 1.94-2.26 (m, 10H, H-2″, H-6″, H-7″, H-2‴a, H-3‴a), 2.89 (q, 2H, $J_{CH2,CH2}$ = $J_{CH2,NH}$ = 5.8, N**CH$_2$**CH$_2$), 2.96-3.11 (m, 1H, H-4‴), 3.17 (s, 3H, OCH$_3$), 3.31 (t, 2H, $J_{CH2,CH2}$ = 5.8, NCH$_2$**CH$_2$**), 3.51 (s, 2H, CH$_2$O), 3.56-3.71 (m, 1H, H-1‴), 7.56 (t, 1H, $J_{NH,CH2}$ = 6.0, SNH), 7.63-7.77 (m, 2H, H-2), 7.90 (bs, 3H, NH$_2$, 2′-NH), 8.14 (s, 1H, H-8′), 8.16-8.28 (m, 2H, H-3), 9.98 (s, 1H, 6′-NH). ^{13}C NMR: δ 29.36 and 29.90 (C-2‴, C-3‴), 31.25 (C-3″, C-5″), 34.06 (C-2″, C-6″), 42.15 (N**CH$_2$**CH$_2$), 43.26 (C-7″), 48.20 (C-4″), 48.92 (C-4‴), 49.64 (C-1‴), 57.89 (OCH$_3$), 64.28 (CH$_2$O), 65.37 (C-1″), 70.57 (NCH$_2$**CH$_2$**), 112.8 (C-5′), 119.29 (C-3), 127.21 (C-2), 132.85 (C-1), 137.2 (C-8′), 143.79 (C-4), 151.09 (C-4′), 157.72 (C-2′). ESI MS *m/z* (%): 585.3 (100) [M + H], 607.3 (43) [M + H]; HRMS ESI ($C_{28}H_{41}O_4N_8S$) calculated: 585.29660; found: 585.29656.

4-((2-(([1r,4r]-4-Aminocyclohexyl)amino)-9-cyclohexyl-9H-purin-6-yl)amino)-N-(2-methoxyethyl)benzenesulfonamide (**20b**): Starting from **17b** (100 mg, 0.39 mmol), 16 (108 mg, 0.47 mmol), DIPEA (137 mL, 0.79 mmol), *n*-BuOH (mL) in stage 1 and *trans*-1,4-cyclohexandiamine (135 mg, 1.18 mmol) in stage 2. Yield 119 mg (46%) as a TFA salt. ^{1}H NMR: δ 1.51-1.20 (m, 7H, H-3”b, H-4”b, H-2‴b, H-3‴b), 1.67-1.74 (m, 1H, H-4”a), 2.14-1.80 (m, 10H, H-2″, H-3”a, H-4”a, H-2‴a, H-3‴a), 2.89 (q, 2H, $J_{CH2,CH2}$ = $J_{CH2,NH}$ = 5.8, N**CH$_2$**CH$_2$), 2.99-3.04 (m, 1H, H-4‴), 3.17 (s, 3H, OCH$_3$), 3.31 (t, 2H, $J_{CH2,CH2}$ = 5.8, NCH$_2$**CH$_2$**), 3.58-3.72 (m, 1H, H-1‴), 4.16-4.31 (m, 1H, H-1″), 7.56 (t, 1H, $J_{NH,CH2}$ = 6.0, SNH), 7.68-7.75 (m, 2H, H-2), 7.90 (bs, 4H, NH$_2$, 2′-NH), 8.18-8.24 (m, 2H, H-3), 8.29 (s, 1H, H-8′), 10.11 (s, 1H, 6′-NH). ^{13}C NMR: δ 24.83 (C-4″), 25.18 (C-3″), 29.33 and 30.00 (C-2‴, C-3‴), 31.95 (C-2″), 42.16 (N**CH$_2$**CH$_2$), 48.86 (C-4‴), 49.57 (C-1‴), 53.5 (C-1″), 57.89 (OCH$_3$), 70.57 (NCH$_2$**CH$_2$**), 112.17 (C-5′), 119.41 (C-3), 127.22 (C-2), 133.12 (C-1), 136.95 (C-8′), 143.64 (C-4), 151.13 (C-4′), 157.60 (C-2′). ESI MS *m/z* (%): 543.3 (100) [M + H], 565.2 (36) [M + H]; HRMS ESI ($C_{28}H_{34}O_3N_9S$) calculated: 543.28603; found: 543.28607.

1-((4-((2-(([1r,4r]-4-Aminocyclohexyl)amino)-9-(4-(hydroxymethyl)bicyclo[2.2.1]heptan-1-yl)-9H-purin-6-yl)amino)phenyl)sulfonyl)piperidin-4-one (**21a**): Starting from **17a** (75 mg, 0.25 mmol), **15** (90 mg, 0.30 mmol), DIPEA (88 mL, 0.5 mmol), *n*-BuOH (3 mL) in stage 1 and *trans*-1,4-cyclohexandiamine (87 mg, 0.75 mmol) in stage 2. Final deprotection of the ketal was accomplished by diluting the crude reaction mixture with 5% TFA in H$_2$O-MeCN (1:1) and stirring this mixture at room temperature for 12 hours. Yield 57 mg (35%) as a TFA salt. ^{1}H NMR: δ 1.27-1.52 (m, 6H, H-2‴b, H-3‴b, H-3”a, H-5”endo), 1.70-1.82 (m, 2H, H-3”b, H-5”exo), 1.95-2.23 (m, 10H, H-2″, H-6″, H-7″, H-3‴a, H-2‴a), 2.38-2.46 (m, 4H, OC**CH$_2$**), 3.00-3.10 (m, 1H, H-4‴), 1.23-3.35 (m, 4H, N**CH$_2$**CH$_2$), 3.51 (s, 2H, CH$_2$O), 3.58-3.68 (m, 1H, H-1‴), 6.78 (bs, 1H, 2′-NH), 7.64-7.74 (m, 2H, H-3), 7.86 (bs, 3H, NH$_2$), 8.00 (s, 1H, H-8′), 8.27-8.36 (m, 2H, H-2), 10.05 (s, 1H, 6′-NH). ^{13}C NMR: δ 29.35 and 29.88 (C-2‴, C-3‴), 31.25 (C-3″, C-5″), 34.09 (C-2″, C-6″), 39.5 (NCH$_2$**CH$_2$**), 43.26 (C-7″), 45.16

Uncorrected Proofs

($N\mathbf{CH_2}CH_2$), 48.20 (C-4″), 48.90 (C-4‴), 49.70 (C-1‴), 64.28 (CH_2O), 65.30 (C-1″), 113.15 (C-5′), 119.44 (C-2), 127.40 (C-4), 128.13 (C-3), 137.4 (C-8′), 144.83 (C-1), 151.09 (C-4′), 152.36 (C-6′), 157.67 (C-2′), 205.58 (CO). ESI MS *m/z* (%): 609.4 (100) [M + H]; HRMS ESI ($C_{30}H_{41}O_4N_8S$) calculated: 609.29660; found: 609.29667.

1-((4-((2-((([1r,4r]-4-Aminocyclohexyl)amino)-9-cyclohexyl-9H-purin-6-yl)amino)phenyl)sulfonyl)piperidin-4-one (**21b**): Starting from **17b** (80 mg, 0.31 mmol), **15** (112 mg, 0.38 mmol), DIPEA (109 mL, 0.63 mmol), *n*-BuOH (2 mL) in stage 1 and *trans*-1,4-cyclohexandiamine (108 mg, 0.94 mmol) in stage 2. Final deprotection of the ketal was accomplished by diluting the crude reaction mixture with 5% TFA in H_2O-MeCN (1:1) and stirring this mixture at room temperature for 12 hours. Yield 81 mg (38%) as a TFA salt. ^{1}H NMR: δ 1.15-1.55 (m, 7H, H-3”b, H-4”b, H-2‴b, H-3‴b), 1.67-1.74 (m, 1H, H-4”a), 1.80-1.95 (m, 4H, H-3”a, H-2”b), 1.95-2.02 (m, 4H, H-3‴a, H-2”a), 2.04-2.15 (m, 2H, H-2‴a), 2.36-2.48 (m, 4H, $OCCH_2$), 2.99-3.09 (m, 1H, H-4‴), 3.21-3.39 (m, 4H, NCH_2), 3.60-3.69 (m, 1H, H-1‴), 4.16-4.28 (m, 1H, H-1″), 6.95 (bs, 1H, 2′-NH), 7.64-7.77 (m, 2H, H-2), 7.84 (bs, 3H, NH_2), 8.12 (s, 1H, H-8′), 8.28-8.37 (m, 2H, H-3), 10.11 (bs, 1H, 6′-NH). ^{13}C NMR: δ 24.83 (C-4″), 25.23 (C-3″), 29.32 (C-3‴), 30.00 (C-2‴), 31.95 (C-2″), 40.3 ($OC\mathbf{CH_2}$), 45.12 (NCH_2), 48.86 (C-4‴), 49.75 (C-1‴), 53.88 (C-1″), 112.1 (C-5′), 119.65 (C-2), 128.15 (C-3), 137.07 (C-8′), 144.90 (C-1), 151.00 (C-4′), 150.74 (C-4′), 157.61 (C-2′), 205.57 (CO). NegESI MS *m/z* (%): 565.3 (100) [M-H]; HRMS negESI ($C_{28}H_{37}O_3N_8S$) calculated: 565.27148; found: 556.27057.

(4-(6-chloro-2-iodo-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**22a**): To a mixture of **9** (881 mg, 3 mmol), CuI (571 mg, 3 mmol) and CH_2I_2 (967 μL, 12 mmol) in THF (20 mL) was added isoamyl nitrite (1.22 mL, 6 mmol) dropwise and the reaction mixture was heated to reflux for 4 hours. Even though the TLC retention was almost identical to the starting material (ethyl acetate - methanol 9:1), reaction on *p*-nitrobenzyl pyridine stain was different. Volatiles were evaporated, crude mixture was adsorbed on silica and flash chromatography (0-30%methanol in ethyl acetate) afforded **19** (900 mg, 74%) as pale brown foam. ^{1}H NMR: δ 1.40-1.50 (m, 2H, H-2b, H-6endo), 1.74-1.84 (m, 2H, H-2a, H-6exo), 1.99-2.07 (m, 2H, H-3exo, H-5 endo), 2.08 (s, 2H, H-7), 2.21-2.31 (m, 2H, H-3endo, H-5endo), 3.51 (d, 2H, $J_{CH2,OH}$ = 5.3, CH_2O), 4.69 (t, 1H, $J_{OH,CH2}$ = 5.3, OH), 8.62 (s, 1H, H-8′). ^{13}C NMR: δ 31.36 (C-2, C-6), 34.73 (C-3, C-5), 43.62 (C-7), 48.62 (C-1), 64.30 (CH_2O), 66.49 (C-4), 117.42 (C-2′), 131.89 (C-5′), 146.84 (C-8′), 148.87 (C-6′), 153.43 (C-4′). ESI MS *m/z* (%): 405.1 (100) [M + H]; HRMS ESI ($C_{13}H_{15}N_4ClI$) calculated: 404.99736; found: 404.99743.

6-Chloro-9-cyclohexyl-2-iodo-9H-purine (**22b**): Compound was prepared according to a published procedure.[36]

4.4 | General procedure for the preparation of compounds 22

A solution of **22a** or **22b** (100 mg, 0.25 or 0.28 mmol), DIPEA (2 equiv, 86 μL, 0.5 mmol or 96 μL, 0.55 mmol), and **13** (1.2 equiv, 67 mg, 0.3 mmol or 75 mg, 0.33 mmol) in *n*-BuOH (3 mL) was heated in a pressure vessel to 150°C overnight. Volatiles were thoroughly removed in vacuo, residue was suspended in ethyl acetate, filtered through a plug of celite and evaporated to afford crude intermediate **23a** or **23b**. A sample of the intermediate was isolated and subjected to NMR and HRMS analysis.

In a separate flask, $Pd_2(dba)_3$ (0.05 equiv, 11 mg, 0.012 mmol or 13 mg, 0.014 mmol) and XantPhos (0.1 equiv, 14 mg, 0.025 mmol or 16 mg, 0.028 mmol) were mixed in dry toluene (2 mL) under argon atmosphere and heated to 50°C for 10 minutes (color turns from dark-purple to yellow). To this formed Pd-complex was added Cs_2CO_3 (1.1 equiv, 89 mg, 0.27 mmol or 99 mg, 0.30 mmol) followed by a slow addition of a solution of aniline “(0.32 mmol, 1.3 equiv) and the crude intermediate in dry dioxane (2 mL). Reaction mixture was stirred at room temperature overnight, all volatiles were evaporated, and the product was isolated by a combination of normal phase (MeOH in $CHCl_3$) and reverse phase (10-100% MeCN in H_2O, 0.2% TFA) flash chromatography.

(4-(2-iodo-6-((4-[pyrrolidin-1-ylsulfonyl]phenyl)amino)-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**23a**): ^{1}H NMR: δ 1.40-1.50 (m, 2H, H-2b, H-6endo), 1.62-1.69 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.73-1.86 (m, 2H, H-2a, H-6exo), 2.12-1.97 (m, 4H, H-3exo, H-5exo, H-7), 2.19-2.29 (m, 2H, H-3endo, H-5endo), 3.10-3.18 (m, 4H, $N\mathbf{CH_2}CH_2$), 3.51 (d, 2H, $J_{CH2,OH}$ = 5.4, CH_2O), 4.68 (d, 1H, $J_{OH,CH2}$ = 5.5, OH), 7.75-7.82 (m, 2H, H-3″), 8.11-8.19 (m, 2H, H-2″), 8.30 (s, 1H, H-8′), 10.57 (s, 1H, 6′-NH). ^{13}C NMR: δ 24.88 ($NCH_2\mathbf{CH_2}$), 31.30 (C-2, C-6), 34.74 (C-3, C-5), 43.59 (C-7), 47.97 ($N\mathbf{CH_2}CH_2$), 48.50 (C-1), 64.23 (CH_2O), 65.82 (C-4), 118.62 (C-2′), 119.94 (C-2″), 120.01 (C-5′), 128.40 (C-3″), 129.65 (C-4″), 141.45 (C-8′), 143.52 (C-1″), 150.94 (C-4′), 151.26 (C-6′).ESI MS *m/z* (%): 594.2 (100) [M + H]; HRMS ESI ($C_{23}H_{27}IN_6O_3S$) calculated: 594.09100; found: 594.09109.

9-Cyclohexyl-2-iodo-N-(4-[pyrrolidin-1-ylsulfonyl]phenyl)-9H-purin-6-amine (**23b**): ^{1}H NMR: δ 1.23-1.31 (m, 1H, H-4”b), 1.41-1.53 (m, 2H, H-3”a), 1.62-1.68 (m, 4H, H-4”a), 1.69-1.76 (m, 1H, H-4”a), 1.90-1.80 (m, 4H, H-2”a, H-3”a), 1.99-2.07 (m, 2H, H-2”a), 3.11-3.18 (m, 4H, $N\mathbf{CH_2}CH_2$), 4.32-4.42 (m, 1H, H-1″), 7.73-7.81 (m, 2H, H-3′), 8.12-8.18 (m, 2H, H-2′), 8.44 (s, 1H, H-8), 10.60 (bs, 1H, 6-NH). ^{13}C NMR: δ 24.87 ($NCH_2\mathbf{CH_2}$), 24.92 (C-4″), 25.19 (C-3″), 32.52 (C-2″), 47.97 ($N\mathbf{CH_2}CH_2$), 53.95 (C-1″), 118.84 (C-2), 119.95 (C-2′), 120.33 (C-5), 128.40 (C-3′), 129.68 (C-4′), 140.78 (C-8), 143.51 (C-1′), 150.46 (C-4), 150.85 (C-6). ESI MS *m/z* (%): 552.2 (100) [M + H]; HRMS ESI ($C_{21}H_{25}IN_6O_2S$) calculated: 552.08044; found: 552.08031.

(4-(2-(phenylamino)-6-((4-[pyrrolidin-1-ylsulfonyl]phenyl)amino)-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**24a**): Aniline (29 μL) was used in the second stage pf reaction, FCC in 1-10% MeOH in $CHCl_3$, yield 54 mg (39%) ^{1}H NMR (401 MHz, DMSO-d_6) δ 1.55-1.41 (m, 2H, H-2a, H-6endo), 1.70-1.61 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.90-1.76 (m, 2H, H-2b, H-6exo), 2.13 (s, 2H, H-7), 2.29-2.16 (m, 4H, H-3, H-5), 3.12-3.17 (m, 4H, NCH_2), 3.53 (s, 2H, CH_2O), 6.94 (t, 1H, *J* = 7.3 Hz,, H-4‴), 7.24-7.32 (m, 2H, H-3‴), 7.66-7.73 (m, 2H, H-3″), 7.80-7.86 (m, 2H, H-2‴), 8.10 (s, 1H, H-8′), 8.32-8.38 (m, 2H, H-2″), 9.25 (s, 1H, 2′-NH), 10.15 (s, 1H, 6′-NH). ^{13}C NMR (101 MHz, DMSO) δ 24.88 ($NCH_2\mathbf{CH_2}$), 31.45 (C-2, C-6), 34.39 (C-3, C-5), 43.52 (C-7), 48.00 (NCH_2), 48.46 (C-1), 64.44 (CH_2O), 65.35 (C-4), 115.76 (C-5″), 118.81

Uncorrected Proofs

Uncorrected Proofs

(C-2‴), 119.92 (C-2″), 121.03 (C-4‴), 128.25 (C-3″), 128.49 (C-4″), 128.57 (C-3‴), 138.63 (C-8′), 141.30 (C-1‴), 144.53 (C-1″), 151.60 (C-2′, C-4′), 155.16 (C-6′). NegESI MS *m/z* (%): 558.2 (100) [M-H]; HRMS negESI ($C_{28}H_{37}O_3N_8S$) calculated: 565.27148; found: 565.27112.

9-cyclohexyl-N^2-phenyl-N^6-(4-[pyrrolidin-1-ylsulfonyl]phenyl)-9H-purine-2,6-diamine ((**24b**): Aniline (33 μL) was used in the second stage pf reaction, FCC in 1-3% MeOH in $CHCl_3$, yield 48 mg (34%). ^{1}H NMR: δ 1.26-1.34 (m, 1H, H-4‴a), 1.38-1.50 (m, 2H, H-3‴a), 1.62-1.68 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.71-1.79 (m, 1H, H-4‴b), 1.86-1.93 (m, 2H, H-2‴a), 1.97-2.08 (m, 4H, H-2‴b, H-3‴b), 3.10-3.17 (m, 4H, $N\mathbf{CH_2}CH_2$), 4.27-4.37 (m, 1H, H-1‴), 6.91-6.97 (m, 1H, H-4″), 7.24-7.31 (m, 2H, H-3″), 7.66-7.72 (m, 2H, H-2′), 7.77-7.84 (m, 2H, H-2″), 8.15 (s, 1H, H-8), 8.27-8.34 (m, 2H, H-3′), 9.24 (s, 1H, 2-NH). 10.16 (s, 1H, 6-NH). ^{13}C NMR: δ 24.67 ($NCH_2\mathbf{CH_2}$), 24.93 (C-4‴), 25.24 (C-3‴), 31.99 (C-2‴), 47.79 ($N\mathbf{CH_2}CH_2$), 53.85 (C-1‴), 115.52 (C-5), 118.83 (C-2″), 119.87 (C-2′), 120.85 (C-4″), 127.99 (C-3′), 128.29 (C-4′), 128.32 (C-3″), 138.21 (C-8), 141.14 (C-1″), 144.32 (C-1′), 150.99 (C-4), 151.41 (C-6), 155.21 (C-2).ESI MS *m/z* (%): 518.2 (100) [M + H], 540.2 (12) [M + Na]; HRMS ESI ($C_{27}H_{32}O_2N_7S$) calculated: 518.23327; found: 518.23326.

(4-(2-(Pyridin-3-ylamino)-6-((4-[pyrrolidin-1-ylsulfonyl]phenyl)amino)-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**25a**): 3-aminopyridine (30 mg) was used in the second stage pf reaction, FCC in 5-20% MeOH in $CHCl_3$, yield 35 mg (25%). ^{1}H NMR: δ 1.52 (m, 2H, H-2b, H-6endo), 1.66 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.85 (m, 2H, H-2a, H-6exo), 2.20-2.05 (m, 4H, H-3exo, H-5exo, H-7), 2.31 (m, 2H, H-3endo, H-5endo), 3.15 (m, 4H, $N\mathbf{CH_2}CH_2$), 3.55 (s, 2H, CH_2O), 7.74 (m, 2H, H-3″), 7.91 (dd, 1H, $J_{5''',4'''}$= 8.8, $J_{5''',6'''}$= 5.3, H-5‴), 8.18 (s, 1H, H-8′), 8.29 (m, 2H, H-2″), 8.47 (dd, 1H, $J_{6''',5'''}$= 5.4, $J_{6''',4'''}$= 1.1, H-6‴), 8.65 (ddd, 1H, $J_{4''',5'''}$= 8.8, $J_{4''',2'''}$= 2.5, $J_{4''',6'''}$= 1.2, H-4‴), 9.34 (d, 1H, $J_{2''',4'''}$= 2.4, H-2‴), 10.16 (s, 1H, 2′-NH), 10.38 (s, 1H, 6′-NH). ^{13}C NMR: δ 24.70 ($NCH_2\mathbf{CH_2}$), 31.31 (C-2, C-6), 34.45 (C-3, C-5), 43.60 (C-7), 47.81 ($N\mathbf{CH_2}CH_2$), 48.21 (C-1), 64.15 (CH_2O), 65.27 (C-4), 117.03 (C-5′), 120.02 (C-2″), 126.49 (C-5‴), 128.14 (C-3″), 128.90 (C-4″), 131.61 (C-4‴), 131.86 (C-2‴), 134.43 (C-6‴), 139.59 (C-8′), 140.32 (C-3‴), 143.99 (C-1″), 150.87 (C-4′), 151.76 (C-2′), 153.48 (C-6′).ESI MS *m/z* (%): 561.2 (100) [M + H]; HRMS ESI ($C_{28}H_{39}O_3N_8S$) calculated: 561.23908; found: 561.23901.

9-Cyclohexyl-N^2-(pyridin-3-yl)-N^6-(4-[pyrrolidin-1-ylsulfonyl]phenyl)-9H-purine-2,6-diamine (**24b**): 3-aminopyridine (34 mg) was used in the second stage pf reaction, FCC in 1-10% MeOH in $CHCl_3$, yield 55 mg (38%). ^{1}H NMR: δ 1.28-1.37 (m, 1H, H-4‴b), 1.44-1.57 (m, 2H, H-3‴b), 1.61-1.69 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.70-1.77 (m, 1H, H-4‴a), 1.85-2.01 (m, 4H, H-2‴b, H-3‴a), 2.02-2.11 (m, 2H, H-2‴a), 3.18-3.11 (m, 4H, $N\mathbf{CH_2}CH_2$), 4.37-4.46 (m, 1H, H-1‴), 7.72-7.77 (m, 2H, H-3′), 7.91 (dd, 1H, $J_{5'',4''}$ = 8.7, $J_{5'',6''}$ = 5.3, H-5″), 8.24-8.30 (m, 2H, H-2′), 8.31 (s, 1H, H-8), 8.47 (dd, 1H, $J_{6'',5''}$ = 5.4, $J_{6'',4''}$ = 1.1, H-6″), 8.62 (ddd, 1H, $J_{4'',5''}$ = 8.7, $J_{4'',2''}$ = 2.6, $J_{4'',6''}$ = 1.2, H-4″), 9.49 (d, 1H, $J_{2'',4''}$ = 2.5, H-2″), 10.22 (bs, 1H, 2-NH), 10.40 (bs, 1H, 6-NH). ^{13}C NMR: δ 24.70 ($NCH_2\mathbf{CH_2}$), 24.88 (C-4‴), 25.10 (C-3‴), 32.52 (C-2‴), 47.81 ($N\mathbf{CH_2}CH_2$), 53.90 (C-1‴), 116.39 (C-5), 120.09 (C-2′), 126.59 (C-5″), 128.15 (C-3′), 128.96 (C-4′), 131.37 (C-2″), 132.11 (C-4″), 134.22 (C-6″), 139.13 (C-8), 140.42 (C-3″), 143.95 (C-1′), 150.29 (C-4), 151.64 (C-2), 153.76 (C-6). ESI MS *m/z* (%): 519.3 (100) [M + H], 563.4 (21) [M + H]; HRMS ESI ($C_{26}H_{31}O_2N_8S$) calculated: 519.22852; found: 519.22853.

(4-(2-(pyrazin-2-ylamino)-6-((4-[pyrrolidin-1-ylsulfonyl]phenyl)amino)-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**26a**): Aminopyrazine (31 mg) was used in the second stage pf reaction, FCC in 2-15% MeOH in $CHCl_3$, yield 48 mg (44%) ^{1}H NMR: δ 1.48-1.53 (m, 2H, H-2a, H-6endo), 1.64-1.67 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.81-1.86 (m, 2H, H-2b, H-6exo), 2.11 (bs, 2H, H-7), 2.13-2.19 (m, 2H, H-3a, H-5a), 2.26-2.31 (m, 2H, H-3b, H-5b), 3.12-3.15 (m, 4H, $N\mathbf{CH_2}CH_2$), 3.53 (s, 2H, CH_2O), 7.67-7.70 (m, 2H, H-3″), 8.15 (s, 1H, H-8′), 8.20 (d, 1H, $J_{5''',6'''}$= 2.5, H-5‴), 8.35 (dd, 1H, $J_{6''',5'''}$= 2.5, $J_{6''',3'''}$= 1.5, H-6‴), 8.44-8.47 (m, 2H, H-2″), 9.58 (d, 1H, $J_{3''',6'''}$= 1.5, H-3‴), 10.09 (s, 1H, 2′-NH), 10.30 (s, 1H, 6′-NH). ^{13}C NMR: δ 24.86 ($NCH_2\mathbf{CH_2}$), 31.51 (C-2, C-6), 34.56 (C-3, C-5), 43.73 (C-7), 47.97 ($N\mathbf{CH_2}CH_2$), 48.37 (C-1), 64.43 (CH_2O), 65.42 (C-4), 117.02 (C-5′), 120.22 (C-2″), 128.21 (C-3″), 128.64 (C-4″), 135.66 (C-3‴), 137.02 (C-5‴), 139.50 (C-8′), 142.46 (C-6‴), 144.35 (C-1″), 150.49 and 151.25 (C-2‴, C-4′), 151.87 and 153.38 (C-2′, C-6′). ESI MS *m/z* (%): 562.2 (13) [M + H], 584.2 (100) [M + H]; HRMS ESI ($C_{27}H_{31}O_3N_9S$) calculated: 584.21628; found: 584.21624.

9-Cyclohexyl-N^2-(pyrazin-2-yl)-N^6-(4-[pyrrolidin-1-ylsulfonyl]phenyl)-9H-purine-2,6-diamine (**26b**): Aminopyrazine (34 mg) was used in the second stage pf reaction, FCC in 1-5% MeOH in $CHCl_3$, yield 41 mg (29%), ^{1}H NMR: δ 1.23-1.36 (m, 1H, H-4‴a), 1.39-1.52 (m, 2H, H-3‴a), 1.63-1.68 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.71-1.78 (m, 1H, H-4‴b), 1.84-1.93 (m, 2H, H-3‴b), 1.95-2.12 (m, 4H, H-2‴), 3.10-3.17 (m, 4H, $N\mathbf{CH_2}CH_2$), 4.33-4.43 (m, 1H, H-1‴), 7.66-7.72 (m, 2H, H-3′), 8.20 (d, 1H, $J_{5'',6''}$ = 2.6, H-5″), 8.27 (s, 1H, H-8), 8.35 (dd, 1H, $J_{6'',5''}$ = 2.6, $J_{6'',3''}$ = 1.5, H-6″), 8.40-8.45 (m, 2H, H-2′), 9.59 (d, 1H, $J_{3'',6''}$ = 1.5, H-3″), 10.13 (s, 1H, 2-NH), 10.32 (s, 1H, 6-NH). ^{13}C NMR: δ 24.67 ($NCH_2\mathbf{CH_2}$), 24.90 (C-4‴), 25.20 (C-3‴), 31.99 (C-2‴), 47.79 ($N\mathbf{CH_2}CH_2$), 54.12 (C-1‴), 116.16 (C-5), 120.14 (C-2′), 128.00 (C-3′), 128.48 (C-4′), 135.61 (C-3″), 136.81 (C-5″), 139.00 (C-8), 142.19 (C-6″), 144.08 (C-1′), 150.32 and 150.43 (C-2″, C-4), 151.49 and 153.42 (C-2, C-6). ESI MS *m/z* (%): 520.5 (100) [M + H], 542.5 (27) [M + Na]; HRMS ESI ($C_{25}H_{30}O_2N_9S$) calculated: 520.22377; found: 520.22387.

(4-(2-([6-aminopyridin-3-yl]amino)-6-((4-[pyrrolidin-1-ylsulfonyl]phenyl)amino)-9H-purin-9-yl)bicyclo[2.2.1]heptan-1-yl)methanol (**27a**): 2,5-diaminopyridine dihydrochloride (59 mg) was used in the second stage pf reaction, ratio of regioisomers 8:1 (HPLC), FCC in 5-20% MeOH in $CHCl_3$, only one isomer isolated with yield 40 mg (23%). ^{1}H NMR: δ 1.44-1.54 (m, 2H, H-2b, H-6endo), 1.61-1.70 (m, 4H, $NCH_2\mathbf{CH_2}$), 1.80-1.90 (m, 2H, H-2a, H-6exo), 2.04-2.13 (m, 4H, H-3exo, H-5exo, H-7), 2.23-2.34 (m, 2H, H-3endo, H-5endo), 3.10-3.16 (m, 4H, $N\mathbf{CH_2}CH_2$), 3.54 (s, 2H, CH_2O), 7.04 (d, 1H, $J_{5''',4'''}$= 9.4, H-5‴), 7.67-7.75 (m, 2H, H-3″), 7.89 (bs, 2H, NH_2), 8.07 (s, 1H, H-8′), 8.13 (dd, 1H, $J_{4''',5'''}$= 9.5, $J_{4''',2'''}$= 2.5, H-2‴), 8.25-8.35 (m, 2H, H-2″), 8.57 (s, 1H, H-2‴), 9.47 (s, 1H, 2′-NH), 10.57 (s, 1H, 6′-NH). ^{13}C NMR: δ 24.70 ($NCH_2\mathbf{CH_2}$), 31.22 (C-2, C-6), 34.40 (C-3, C-5), 43.58 (C-7), 47.82 ($N\mathbf{CH_2}CH_2$), 48.22 (C-1), 64.10 (CH_2O), 65.16 (C-4), 113.53 (C-5‴), 116.36 (C-5′), 119.77 (C-2″), 128.09 (C-3″), 128.38 (C-4″), 128.51 (C-3‴), 138.01 (C-4‴), 138.81 (C-8′), 144.24 (C-1″), 150.28

(C-6‴), 151.21 (C-4′), 151.70 and 154.44 (C-2′, C-6′).ESI MS *m/z* (%): 576.2 (100) [M + H], 598.2 (62) [M + Na]; HRMS ESI ($C_{28}H_{34}O_3N_9S$) calculated: 576.24998; found: 576.24991.

N^2-(6-aminopyridin-3-yl)-9-cyclohexyl-N^6-(4-[pyrrolidin-1-ylsulfonyl] phenyl)-9H-purine-2,6-diamine (**27b**)): 2,5-diaminopyridine dihydrochloride (65 mg) was used in the second stage pf reaction, ratio 3:1 (HPLC), FCC in 2-10% MeOH in $CHCl_3$, separation of isomers on RP FCC, yield 57 mg (32%) of **27b** and 14 mg (8%) of the opposite regioisomer. ^{1}H NMR: δ 1.26-1.37 (m, 1H, H-4‴a), 1.41-1.55 (m, 2H, H-3‴a), 1.61-1.68 (m, 4H, NCH_2**CH_2**), 1.68-1.75 (m, 1H, H-4‴b), 1.84-1.97 (m, 4H, H-3‴b, H-2‴a), 2.01-2.09 (m, 2H, H-2‴b), 3.09-3.18 (m, 4H, N**CH_2**CH_2), 4.30-4.39 (m, 1H, H-1‴), 7.04 (d, 1H, $J_{3',4'}$ = 9.5, H-3′), 7.67-7.75 (m, 2H, H-3″), 7.88 (bs, 2H, NH_2), 8.12 (dd, 1H, $J_{4',3'}$ = 9.5, $J_{4',6'}$ = 2.5, H-4′), 8.21-8.29 (m, 2H, H-2″), 8.33 (s, 1H, H-8), 8.63 (bs, 1H, H-6′), 9.56 (bs, 1H, 2-NH), 10.33 (bs, 1H, 6-NH). ^{13}C NMR: δ 24.95 (NCH_2**CH_2**), 25.10 (C-4‴), 25.38 (C-3‴), 32.38 (C-2‴), 48.09 (N**CH_2**CH_2), 54.35 (C-1‴), 113.96 (C-3′), 115.10 (C-5), 120.12 (C-2″), 123.20 (C-6′), 128.41 (C-3″), 128.50 (C-5′), 128.91 (C-4″), 138.64 (C-8, C-4′), 144.33 (C-1″), 150.39 (C-2′), 150.74 (C-4), 151.58 and 155.16 (C-2, C-6). ESI MS *m/z* (%): 534.2 (100) [M + H], 556.2 (47) [M + Na]; HRMS ESI ($C_{26}H_{32}O_2N_9S$) calculated: 534.23942; found: 534.23938.

4.5 | 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 Ni^{2+}NTA column (Qiagen). Kinase (approx. 10 ng) was assayed using a mixture of the following: 1 mg/mL of histone H1, 15 μM of ATP, 0.05 of μCi [γ-^{33}P]ATP, the tested compound, and reaction buffer, in a final volume of 10 μL. The reaction buffer consisted of: 60 mM of HEPES-NaOH, pH 7.5, 3 mM of $MgCl_2$, 3 mM of $MnCl_2$, 3 μM of Na-orthovanadate, 1.2 mM of DTT, and 2.5 μg/50 μL of $PEG_{20.000}$. The reactions were stopped by adding 5 μL of 3% aqueous H_3PO_4. Aliquots were spotted onto P-81 phosphocellulose (Whatman), washed three times with 0.5% aqueous H_3PO_4, 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 IC_{50} values were determined from the dose-response curve.

4.6 | Computational methodology

4.6.1 | Virtual screening and compound design

The final compounds presented in this work issued from computer-aided iterative design. First, conformers of over 1000 compounds from the IOCB proprietary database had been docked into CDK2 structure and scored using SQM/COSMO methodology.[5] We used the CDK2 structure from complex with a large inhibitor staurosporine (PDB: 1AQ1)[37] to allow larger compounds to bind. For docking, the Glide programme of Schrodinger[38] in the standard precision (SP) mode was used with the default settings. Waters beyond 5 Å from the ligand were removed and the bond orders assigned. The receptor grid was created using the default settings. A grid box of 20 × 20 × 20 Å^3 was generated, 10 × 10 × 10 Å^3 inner box was centered on the corresponding ligand. Hydrogen atoms were added to the protein by means of the Protein Preparation Wizard using the default settings. To determine preferable protonation of protein titratable residues, pH was set to 7.0. OPLS3[45] force field was used for hydrogen optimization. Ligand structures were converted from 2D to 3D using the LigPrep module (Schrodinger Suite) with default settings. Glide was set to yield 10 best-scored poses per ligand.

For scoring, we employed the SQM/COSMO method at the PM6-D3H4X level[5] using the linear-scaling MOZYME algorithm[42] in MOPAC 2016 [43] via the Cuby3 interface.[44] To speed up the calculations, only residues within 10 Å of the inhibitor (the union across the whole series, that is, the same for all the complexes) were taken into account. During geometry optimizations, only residues within 8 Å from the inhibitors were allowed to move. The interaction energies were obtained by subtracting the SQM/COSMO energies of the protein and the ligand from those of the complex. Adding the interaction solvation free energies and ligand deformation free energies gave SQM/COSMO scores. Based on the SQM/COSMO scores, selected compounds were put forward for activity testing against CDK2/Cyclin E.

The three hits discovered (compounds **1**, **2**, **3**) defined the 2,6-diamino purine core and 9-norbornyl substituent as a base for further exploration of chemical space to define the structure-activity relationships. We combined this core with selected 36 modifications of position 2 and 37 modifications in position 6 according to literature.[29,30] For these 73 compounds, we first placed the purine core in the active site to maintain the main hinge region H-bonds using docking with settings as above. The modifications were built manually using PyMol, ver. 0.99[46] so that no steric clashes with the protein resulted. This approach is based on our extensive experience with SQM/COSMO scoring,[8,40,41] which states that docking is useful for exploring various binding modes, while building diminishes the risk of energy variations due to small changes in the structures.[41] SQM/COSMO scoring[5] with settings described above was used to prioritize compounds for synthesis.

Eight compounds with high scores and synthetic feasibility (**18a**, **19a**, **20a**, **21a**, **24a**, **25a**, **26a**, **27a**; Table 1) were suggested for synthesis. To explore the importance of the hydroxynorbornyl moiety in position 9, another eight compounds with the cyclohexyl substituent in position 9 (**18b**, **19b**, **20b**, **21b**, **24b**, **25b**, **26b**, **27b**; Table 1) were additionally suggested for synthesis. Thus, a total of 16 compounds were synthesized based on computational design.

4.6.2 | Refined scoring of CDK2/ligand series

In order to shed light on the molecular reasons of the 16 new compounds' activities or the lack thereof, we carried out a refined scoring.

Uncorrected Proofs

Uncorrected Proofs

Most importantly, the protein conformation now was that of CDK2/roscovitin complex (PDB: 3DDQ)[25] because we had learnt that the three hits had purine core. Docking followed the previously published protocol using Glide (see also above).[5,40] Only protein chain A of the CDK2 structure from its complex with roscovitine was used after the ligands and solvent molecules had been discarded. A careful analysis of all the obtained binding modes and their SQM/COSMO scores was performed. The identified best-scoring orientations of modifications in position 6 were built for all the compounds with the modification in question of the series followed by a short molecular dynamics-based quenching for small structural rearrangements. Atomic velocities were assigned following Maxwell–Boltzmann distribution at temperature of 1000 K. The temperature profile was: 1500 K for 1 ps and then cooling down to 0 K over 2 ps.[6,41] The resulting structures were used for SQM/COSMO scoring (see above). The best-scoring orientation of position 6 in compound **18a** was then built into other compounds with the same modification.

ACKNOWLEDGMENTS

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." V.K. and R.J. thank the Czech Science Foundation for support (19-08410S). J.F., C.K., H.A., P.H., and M.L. thank the research project RVO 61388963, awarded by the Academy of Sciences of the Czech Republic and European Regional Development Fund; OP RDE; Project: "Chemical biology for drugging undruggable targets (ChemBioDrug)" (No. CZ.02.1.01/0.0/0.0/16_019/0000729). Part of the work was supported by Gilead Sciences, Inc.

CONFLICT OF INTEREST

The study was partially supported by Gilead Sciences, Inc.

AUTHOR CONTRIBUTIONS

C.K. and M.D. contributed equally to the study. C.K., H.A., J.F., and M.L. performed docking and molecular modeling. M.D., M.Š., and H.H. performed synthesis of compounds. R.J. and V.K. were responsible for biological testing of compounds. M.D and E.P. measured and interpreted analytical data of the prepared compounds. P.H., M.L., and R.N. designed the study, wrote the manuscript, and led the team.

ORCID

Martin Lepšík https://orcid.org/0000-0003-2607-8132
Radim Nencka https://orcid.org/0000-0001-6167-0380

REFERENCES

1. Klein ME, Kovatcheva M, Davis LE, Tap WD, Koff A. *CDK4/6 inhibitors: the mechanism of action may not be as simple as once thought*. Cancer Cell. 2018;34:9-20.
2. Malumbres M. *Cyclin-dependent kinases*. Genome Biol. 2014;15:122.
3. Huwe A, Mazitschek R, Giannis A. *Cyclin-dependent kinases small molecules as inhibitors of cyclin-dependent kinases*. Angew Chem Int Ed. 2003;42:2122-2138.
4. Fanfrlík J, Bronowska AK, Řezáč J, Přenosil O, Konvalinka J, Hobza P. *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. 2010;114:12666-12678.
5. Lepšík M, Řezáč J, Kolář M, Pecina A, Hobza P, Fanfrlík J. *The semiempirical quantum mechanical scoring function for in silico drug design*. ChemPlusChem. 2013;78:921-931.
6. Hylsová M, Carbain B, Fanfrlík J, et al. *Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: inhibition of CDK2 by new pyrazolo[1,5-a]pyrimidines*. Eur J Med Chem. 2017;126:1118-1128.
7. Nekardová M, Vymětalová L, Khirsariya P, et al. *Structural basis of the interaction of cyclin-dependent kinase 2 with roscovitine and its analogues having bioisosteric central heterocycles*. ChemPhysChem. 2017;18:785-795.
8. Pecina A, Meier R, Fanfrlík J, et al. *The SQM/COSMO filter: reliable native pose identification based on the quantum-mechanical description of protein-ligand interactions and implicit COSMO solvation*. Chem Commun (Camb). 2016;52:3312-3315.
9. Pecina A, Brynda J, Vrzal L, et al. *Ranking power of the SQM/COSMO scoring function on carbonic anhydrase II–inhibitor complexes*. ChemPhysChem. 2018;19:873-879.
10. Pecina A, Haldar S, Fanfrlík J, et al. *SQM/COSMO scoring function at the DFTB3-D3H4 level: unique identification of native protein–ligand poses*. J Chem Inf Model. 2017;57:127-132.
11. Ajani, H.; Pecina, A.; Eyrilmez, S. M.; ~~Ich~~ Fanfrlík, J.; Haldar, S.; ~~Ezáč~~ezáč, J. R.; Hobza, P.; Lepš, M. Superior performance of the SQM/COSMO scoring functions in native pose recognition of diverse protein–ligand complexes in cognate docking. 2017
12. Eyrilmez SM, Köprülüoğlu C, Řezáč J, Hobza P. *Impressive enrichment of semiempirical quantum mechanics-based scoring function: HSP90 protein with 4541 inhibitors and decoys*. ChemPhysChem. 2019;20:2759-2766.
13. Nencka R, Sala M, Dejmek M, Dracinsky M, Holy A, Hrebabecky H. *Two convergent approaches toward novel carbocyclic C-nucleosides*. Synthesis. 2010;4119-4130.
14. Šála M, De Palma AM, Hřebabecký H, et al. *Design, synthesis, and biological evaluation of novel coxsackievirus B3 inhibitors*. Bioorg Med Chem. 2010;18:4374-4384.
15. Hřebabecký H, Dejmek M, Dračínský M, et al. *Synthesis of novel azanorbornylpurine derivatives*. Tetrahedron. 2012;68:1286-1298.
16. Dejmek M, Hrebabecky H, Sala M, et al. *From norbornane-based nucleotide analogs locked in south conformation to novel inhibitors of feline herpes virus*. Bioorg Med Chem. 2014;22:2974-2983.
17. Šála M, Dejmek M, Procházková E, et al. *Synthesis of locked cyclohexene and cyclohexane nucleic acids (LCeNA and LCNA) with modified adenosine units*. Org Biomol Chem. 2015;13:2703-2715.
18. Hrebabecky H, Dracinsky M, Prochazkova E, Sala M, Mackman R, Nencka R. *Control of alpha/beta Anomer formation by a 2′,5′ bridge: toward nucleoside derivatives locked in the south conformation*. J Org Chem. 2017;82:11337-11347.
19. Hercik K, Kozak J, Sala M, et al. *Adenosine triphosphate analogs can efficiently inhibit the Zika virus RNA-dependent RNA polymerase*. Antiviral Res. 2017;137:131-133.
20. Enkvist E, Raidaru G, Vaasa A, Pehk T, Lavogina D, Uri A. *Carbocyclic 3′-deoxyadenosine-based highly potent bisubstrate-analog inhibitor of basophilic protein kinases*. Bioorg Med Chem Lett. 2007;17:5336-5339.
21. Dejmek M, Sala M, Hrebabecky H, et al. *Norbornane-based nucleoside and nucleotide analogues locked in north conformation*. Bioorg Med Chem. 2015;23:184-191.
22. Klima, M.; Baumlova, A.; Chalupska, D.; Hrebabecky, H.; Dejmek, M.; Nencka, R.; Boura, E. Crystal Structure of Phosphatidyl Inositol 4-Kinase II Alpha in Complex with Nucleotide Analog. *Worldwide Protein Data Bank*, 2015.

23. Holý A. *Antiviral acyclic nucleoside phosphonates structure activity studies*. Antiviral Res. 2006;71:248-253.
24. Eyer L, Nencka R, de Clercq E, Seley-Radtke K, Růžek D. *Nucleoside analogs as a rich source of antiviral agents active against arthropod-borne flaviviruses*. Antiviral Chem Chemother. 2018;26:2040206618761299.
25. Bettayeb K, Oumata N, Echalier A, et al. *CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases*. Oncogene. 2008;27:5797-5807.
26. ~~Jorda, R.; Zatloukal, M.; Bazgier, C.; Berka, K.; Be, T.; Strnad, M.; Krys, V. A novel series of highly potent 2,6,9-trisubstituted purine cyclin-dependent kinase inhibitors. 2013.~~ Q7
27. Martin MP, Olesen SH, Georg GI, Schönbrunn E. *Cyclin-dependent kinase inhibitor dinaciclib interacts with the acetyl-lysine recognition site of bromodomains*. ACS Chem Biol. 2013;8:2360-2365.
28. Jorda R, Havlíček L, Šturc A, et al. *3,5,7-substituted pyrazolo[4,3-d] pyrimidine inhibitors of cyclin-dependent kinases and their evaluation in lymphoma models*. J Med Chem. 2019;62:4606-4623.
29. Gucký T, Jorda R, Zatloukal M, et al. *A novel series of highly potent 2,6,9-trisubstituted purine cyclin-dependent kinase inhibitors*. J Med Chem. 2013;56:6234-6247.
30. Jorda R, Paruch K, Krystof V. *Cyclin-dependent kinase inhibitors inspired by roscovitine: purine bioisosteres*. Curr Pharm des. 2012;18:2974-2980.
31. Dejmek M, Kovackova S, Zbornikova E, et al. *One-pot build-up procedure for the synthesis of variously substituted purine derivatives*. RSC Adv. 2012;2:6970-6980.
32. Mojzych M, Šubertová V, Bielawska A, et al. *Synthesis and kinase inhibitory activity of new sulfonamide derivatives of pyrazolo[4,3-e][1,2,4]triazines*. Eur J Med Chem. 2014;78:217-224.
33. Myers SM, Bawn RH, Bisset LC, et al. *High-throughput screening and hit validation of extracellular-related kinase 5 (ERK5) inhibitors*. ACS Comb Sci. 2016;18:444-455.
34. Di Gioia ML, Leggio A, Guarino IF, Leotta V, Romio E, Liguori A. *A simple synthesis of anilines by LiAlH4/TiCl4 reduction of aromatic nitro compounds*. Tetrahedron Lett. 2015;56:5341-5344.
35. Thomas AP, Breault GA, Beattie JF, Jewsbury PJ. Imidazo[1,2-a]pyridine and Pyrazolo[2,3-a]pyridine Derivatives, WO2001014375A1. 2001.
36. Legraverend M, Ludwig O, Bisagni E, et al. *Synthesis and in vitro evaluation of novel 2,6,9-trisubstituted purines acting as cyclin-dependent kinase inhibitors*. Bioorg Med Chem. 1999;7:1281-1293.
37. Lawrie AM, Noble MEM, Tunnah P, Brown NR, Johnson LN, Endicott JA. *Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2*. Nat Struct Biol. 1997;4:796-801.
38. Friesner RA, Banks JL, Murphy RB, et al. *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. J Med Chem. 2004;47:1739-1749.
39. Fanfrlík J, Ruiz FX, Kadlčíková A, et al. *The effect of halogen-to-hydrogen bond substitution on human aldose reductase inhibition*. ACS Chem Biol. 2015;10:1637-1642. Q8
40. Ajani H, Jansa J, Köprülüoğlu C, et al. *Imidazo[1,2-c]pyrimidin-5(6H)-one as a novel core of cyclin-dependent kinase 2 inhibitors: synthesis, activity measurement, docking, and quantum mechanical scoring*. J Mol Recognit. 2018;31:e2720.
41. Brahmkshatriya S, Dobes P, Fanfrlik J, et al. *Quantum mechanical scoring: structural and energetic insights into cyclin-dependent kinase 2 inhibition by Pyrazolo[1,5-a]pyrimidines*. Curr Comp Aided-Drug Design. 2013;9:118-129.
42. Stewart JJP. *Application of the PM6 method to modeling proteins*. J Mol Model. 2009;15:765-805.
43. ~~JJP~~ Stewart, MOPAC2012. Stewart Computational Chemistry, Version 7.263W web: HTTP://OpenMOPAC.net
44. Řezáč J. *Cuby: an integrative framework for computational chemistry*. J Comput Chem. 2016;37:1230-1237.
45. Harder E, Damm W, Maple J, et al. *OPLS3: a force field providing broad coverage of drug-like small molecules and proteins*. J Chem Theory Comput. 2016;12:281-296.
46. De Lano WL. *The PyMOL Molecular Graphics System*. San Carlos, CA: DeLano Scientific LLC; 2002.

SUPPORTING INFORMATION

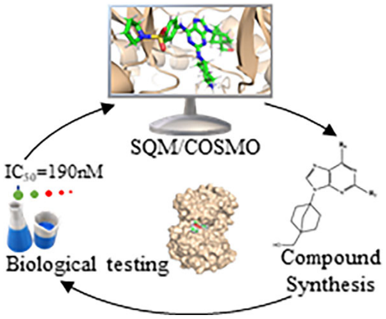
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Köprülüoğlu C, Dejmek M, Šála M, et al. Optimization of norbornyl-based carbocyclic nucleoside analogs as cyclin-dependent kinase 2 inhibitors. *J Mol Recognit*. 2020;e2842. https://doi.org/10.1002/jmr.2842

Uncorrected Proofs

Graphical Abstract

The contents of this page will be used as part of the graphical abstract of html only.
It will not be published as part of main.


SQM/COSMO
IC_{50}=190nM
Biological testing
Compound
Synthesis

Uncorrected Proofs

PUBLICATIONS

Publication C

Michaela Hylsova, Benoit Carbain, Jindrich Fanfrlík, Lenka Musilova, Susanta Haldar, Cemal Köprülüoğlu, Haresh Ajani, Pathik S. Brahmkshatriya, Radek Jorda, Vladimír Krystof, Pavel Hobza, Aude Echalier, Kamil Paruch, Martin Lepsík

Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: Inhibition of CDK2 by newpyrazolo[1,5-a]pyrimidines

Eur J Med Chem, **2017**, 126, 1118-1128

European Journal of Medicinal Chemistry 126 (2017) 1118–1128

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Research paper

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, 1], Benoit Carbain [a, f, 1], Jindřich Fanfrlík [b, 1], Lenka Musilová [a], Susanta Haldar [b, c], Cemal Köprülüoğlu [b, c], Haresh Ajani [b, c], Pathik S. Brahmkshatriya [b], Radek Jorda [d], Vladimír Kryštof [d], Pavel Hobza [b, c], Aude Echalier [e, 2], 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 nam. 2, 166 10 Prague 6, Czech Republic*
[c] *Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, 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*
[f] *International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic*

ARTICLE INFO

Article history:
Received 15 September 2016
Received in revised form
7 December 2016
Accepted 9 December 2016
Available online 11 December 2016

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.

© 2016 Elsevier Masson SAS. All rights reserved.

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 scoring [1] has expanded the methodological repertoire and enabled a quantitative treatment of P–L interactions [2,3], metals (either in proteins [1,4] or ligands) [5], exotic ligands [6], and

---

* Corresponding author. Department of Chemistry, CZ Openscreen, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic.

** Corresponding author.

E-mail addresses: paruch@chemi.muni.cz (K. Paruch), lepsik@uochb.cas.cz (M. Lepšík).

[1] These authors contributed equally to this work.

[2] 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.

http://dx.doi.org/10.1016/j.ejmech.2016.12.023
0223-5234/© 2016 Elsevier Masson SAS. All rights reserved.

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 trials [26]) 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 R^3) 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 IC_{50} values below 20 nM [28]. The significant increase in activity upon replacing $R^3 = H$ 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 $R^3 = H$ but not with $R^3 = Br/Et$ [28]. However, individual contributions of these factors to the overall binding affinity have not been reported to date.

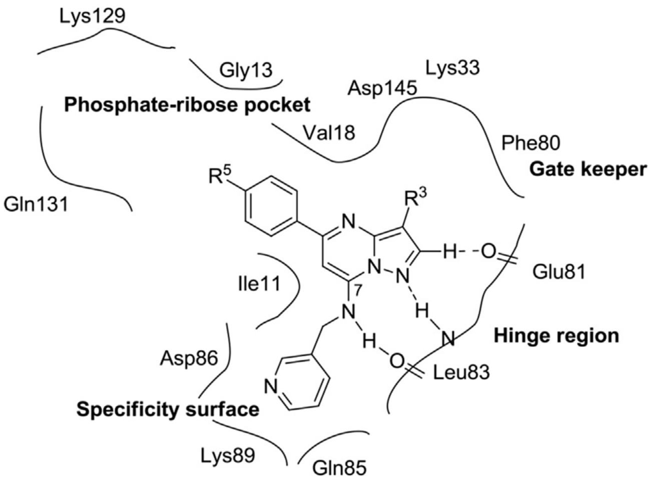


Fig. 1. Schematical representation of interaction of pyrazolo[1,5-*a*]pyrimidines with the CDK2 active site with indicated key amino acids. Three hinge region hydrogen bonds are dashed.

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

Table 1
In vitro activity of CDK2 inhibitors **1a**-**1u** measured as triplicates.

Compound	R^3	R^5	CDK2 IC_{50} [μM][a]
1a	H	Ph	12.58 ± 5.27
1b	Br	Ph	0.49 ± 0.07
1c	H	*p*-OH-Ph	3.86 ± 0.08
1d	Br	*p*-OH-Ph	0.14 ± 0.06
1e	H	*p*-NH_2-Ph	4.91 ± 2.36
1f	Br	*p*-NH_2-Ph	0.14 ± 0.04
1g	H	*p*-CH_2OH-Ph	6.11 ± 2.11
1h	Et	*p*-CH_2OH-Ph	0.13 ± 0.04
1i	Br	*p*-CH_2OH-Ph	0.045 ± 0.015
1j	H	*p*-$CONH_2$-Ph	3.94 ± 1.06
1k	Et	*p*-$CONH_2$-Ph	0.018 ± 0.001
1l	Br	*p*-$CONH_2$-Ph	0.037 ± 0.014
1m	H	*p*-$CONMe_2$-Ph	1.97 ± 0.77
1n	Br	*p*-$CONMe_2$-Ph	0.064 ± 0.013
1o	H	*p*-$CONH_2$-*m*-OMe-Ph	5.53 ± 1.27
1p	H	*p*-$CONH_2$-*m*-OH-Ph	2.67 ± 0.48
1q	Et	*p*-$CONH_2$-*m*-OH-Ph	0.031 ± 0.007
1r	Br	*p*-$CONH_2$-*m*-OH-Ph	0.052 ± 0.015
1s	Br	*p*-$CONH_2$-*o*-OH-Ph	0.015 ± 0.004
1t	H	*p*-$CONH_2$-*m*-$(CH_2)_2CN$-Ph	5.45 ± 1.58
1u	Br	*p*-$CONH_2$-*m*-$(CH_2)_2CN$-Ph	0.053 ± 0.008

[a] Performed at least for three independent biological replicates.

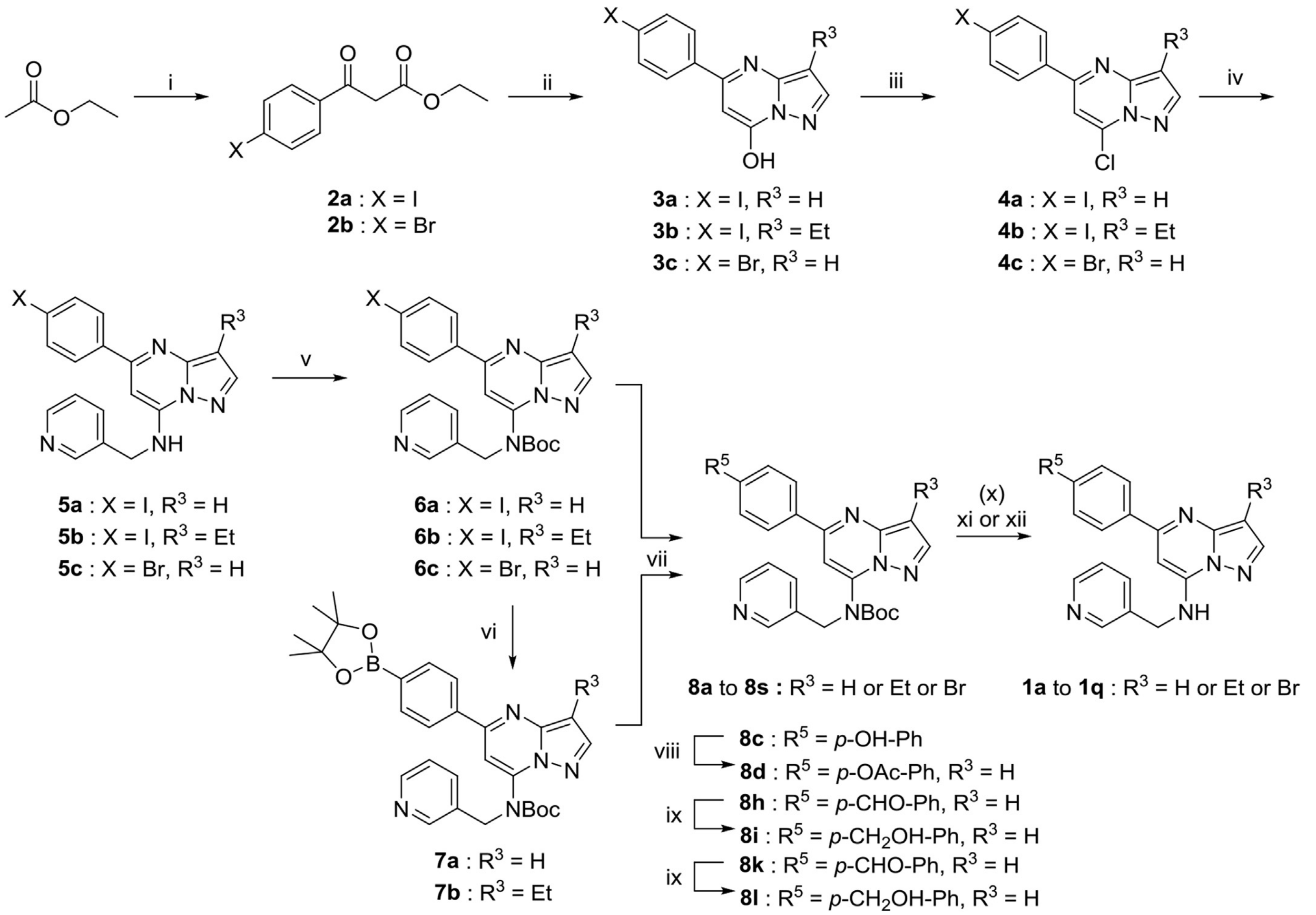
were used for the subsequent installation of the substituted biphenyl moiety via Suzuki coupling (Scheme 1). The syntheses of the biphenyls containing single substituents were quite straightforward and provided the target compounds **1a**–**1q** indicated in Table 1.

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).

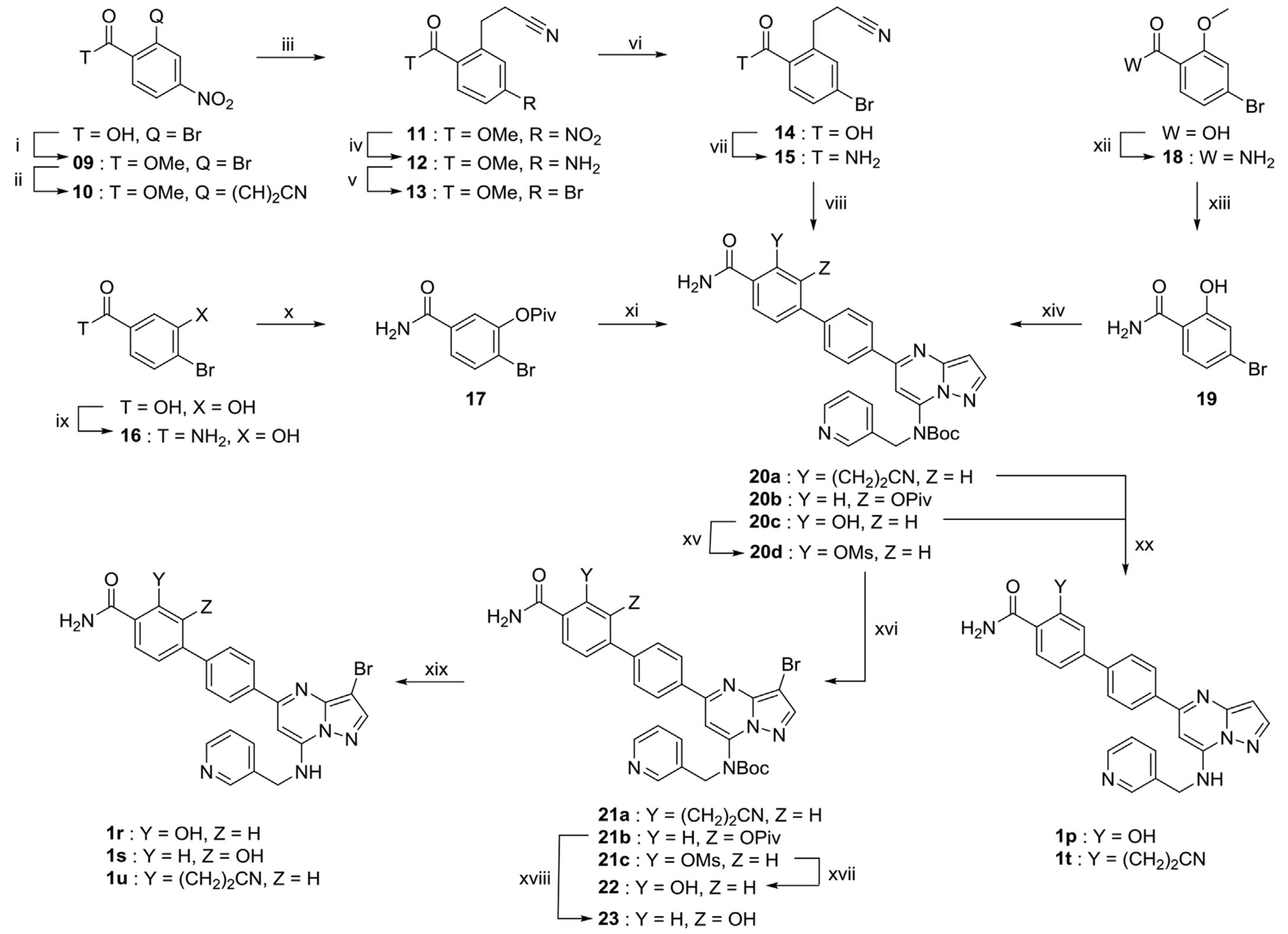
The synthesis of the analogs with nitrile-containing side chains was also relatively challenging. Direct Heck couplings of acrylonitrile with commercially available 2-bromo-4-nitrobenzoic acid or its amide were unsuccessful. However, the reaction did proceed on ester **9**, providing the acrylic nitrile 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 IC_{50} 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 ($R^5 = p$-CONH$_2$-o-OH-Ph) and $R^3 = $ Br (0.015 μM) (Table 1). In general, the compounds could be clearly divided into two groups based on the identity of the R^3 substituent: compounds with $R^3 = H$ were weaker binders with $IC_{50} \geq 2$ μM, whereas compounds with $R^3 = $ Br or Et were significantly more potent ($IC_{50} \leq 0.5$ μM, Table 1). This structure–activity relationship (SAR), compatible with the previous observations for PP-based CDK2 inhibitors with smaller (phenylic) R^5 substituents [28], was previously explained by the filling of the hydrophobic cavity near



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.** $POCl_3$, pyr., r.t., (**4a**: 85%), (**4b**: 42%), (**4c**: 95%); **iv.** 3-picolylamine, MeCN, reflux, (**5a**: 93%), (**5b**: 79%), (**5c**: 97%); **v.** Boc_2O, DMAP, CH_2Cl_2, r.t., (**6a**: 92%), (**6b**: 90%), (**6c**: 94%); **vi.** B_2Pin_2, K_3PO_4, $PdCl_2dppf$, DME, H_2O, reflux, (**7a**: 72% from **6a**, 85% from **6c**), (**7b**: crude used as such in next step); **vii.** K_3PO_4, $PdCl_2dppf$, DME, H_2O, reflux; **viii.** AcCl, CH_2Cl_2, NEt_3, 0 °C to r.t., (**8d**: 82%); **ix.** $NaBH_4$, MeOH, 0 °C to r.t., (**8i:** 88%), (**8l:** 88%); **x.** NBS, MeCN, r.t.; **xi.** 3 M aq. HCl, EtOH, 60 °C; **xii.** TFA, CH_2Cl_2, r.t.



Scheme 2. Synthesis of compounds **1p**, **1r**, **1s**, **1t** and **1u**.
Reagents and conditions: **i.** $SOCl_2$, MeOH, reflux, 95%; **ii.** acrylonitrile, $Pd(P(t\text{-}Bu)_3)_2$, DIPEA, dioxane, 100 °C, 49%; **iii.** H_2, $Pd(OAc)_2$, PPh_3, toluene, 80 °C, 70%; **iv.** $SnCl_2$, EtOAc, H_2O, reflux, 55%; **v.** isoamylnitrite, $CuBr_2$, CH_3CN, 60 °C, 77%; **vi.** NaOH, THF, r.t., quant.; **vii.** $SOCl_2$, DMF, THF, r.t., then NH_3 in dioxane, THF, 0 °C, 72%; **viii.** Compound **7a**, $PdCl_2dppf$, K_3PO_4, DME, H_2O, reflux, (**20a:** 77%); **ix.** $SOCl_2$, THF, reflux then NH_3 in MeOH, THF, 0 °C, 66%; **x.** PivCl, NEt_3, CH_2Cl_2, THF, 75%; **xi.** Compound **7a**, $PdCl_2dppf$, K_3PO_4, DME, H_2O, reflux, (**20b**: 40%); **xii.** Na_2CO_3, $SOCl_2$, NH_3 in MeOH, 80 °C, 68%; **xiii.** BCl_3, CH_2Cl_2, r.t., 90%; **xiv.** Compound **7a**, K_3PO_4, $PdCl_2dppf$, DME, H_2O, reflux, (**20c**: 80%); **xv.** MsCl, NEt_3, DMF, 0 °C, (**20d**: 87%); **xvi.** NBS, CH_2Cl_2, 0 °C, (**21a**: 91%), (**21b**: 86%), (**21c**: 83%); **xvii.** NaOH, THF, reflux, (**22:** 90%); **xviii.** K_2CO_3, 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%).

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 R^3 substituents of the PP core face the gatekeeper Phe80 and the pyridyl moiety occupies the specificity surface (Fig. 1) of the ATP-binding domain of CDK2. The proximal phenyl ring of the R^5 biphenyl moiety (C19-C24, Fig. 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, Fig. 2) extends in the phosphate-binding pocket toward Lys129 on the surface of the enzyme (Figs. 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, Fig. 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 (Figs. 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 (Fig. 2). The corresponding water molecule bound to the CDK2 chain C (W171) has been modeled with full occupancy.

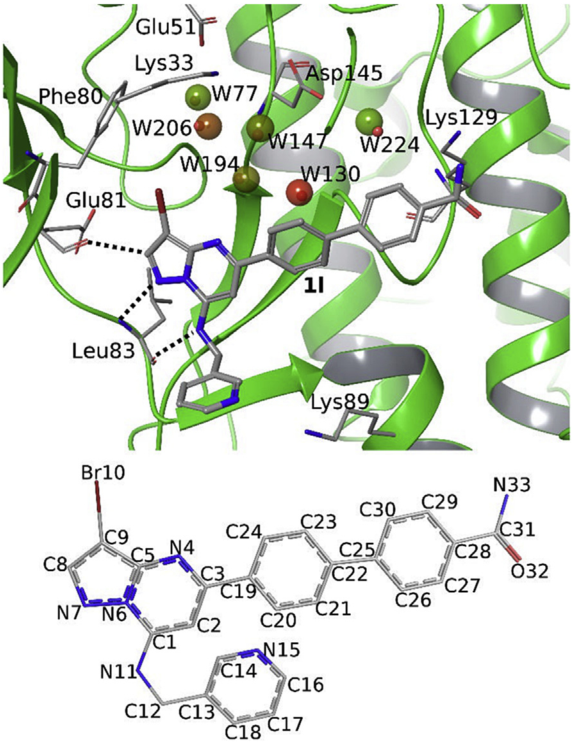


Fig. 2. Inhibitor binding mode, atom numbering and structural water molecules in the crystallographic complex of CDK2/**1l**. Three hinge-region hydrogen bonds between the protein and the inhibitor are shown as dotted lines. Color coding of sticks: gray–carbon, blue–nitrogen, red–oxygen, brown–bromine. Spheres represent water sites as obtained from WaterMap. Their coloring ranges from green (the smallest value of ΔΔG, Table 2) via brown to red (the largest value of ΔΔG, Table 2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.4. *The thermodynamics of explicit active-site waters in the CDK2/**1l** complex*

The CDK2/**1l** 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 (Fig. 2, Table S2). 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 **1l**, respectively. As discussed below, the thermodynamic parameters of water are crucial in QM scoring.

2.5. *Molecular dynamics of the CDK2/**1l** complex*

To understand the full dynamics of the CDK2/**1l** 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 **1l**: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 (Fig. S2, blue and gray bars, respectively).

Table 2
Thermodynamics (free energy, ΔG; enthalpy, ΔH, and entropy, -TΔS; all in kcal/mol) of water molecules in the CDK2/**1l** crystal structure as calculated by WaterMap.

Water	ΔG	ΔH	-TΔS
W77	1.2	−3.2	4.4
W206	4.4	0.5	3.9
W194	2.4	−1.6	4.0
W147	2.4	−2.0	4.5
W224	0.8	−4.4	5.1
W130	6.2	2.1	4.1

2.5.1. 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 more stable 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) (Fig. 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 (Fig. S3B). This was concomitant with the formation of the H-bond between the inhibitor's carboxamide in the conformation **B** and Lys129 (**1l**:O32•••NZ:Lys129) after 40 ns and its maintenance after 60 ns (Fig. S3A). The plotting of the populations of the O32-C31-C3-N4 dihedral throughout the whole 80 ns trajectory (Fig. S2, violet curve) has revealed that both conformations **A** and **B** are present and correspond to the crystallographic values (Fig. S2, blue and gray bars). Figs. S2 and 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 strucsture of the CDK2/**1l** complex.

2.5.2. 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 (Fig. S5). These water molecules are present throughout most of the simulation time, forming a hydrogen-bonded chain and linking the inhibitor N4 nitrogen (Fig. 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 **1l** (Fig. 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 (Fig. 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. Fig. 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 R^3 substituents

As described in Section 2.2, the studied inhibitors can be clearly divided into two groups based on the identity of the R^3 substituent: compounds with $R^3 = H$ are weaker binders with $IC_{50} \geq 2$ μM while compounds with bromine or ethyl moiety as R^3 have $IC_{50} \leq 0.5$ μM. The previous qualitative structure-based explanation was based on filling the hydrophobic cavity near the gatekeeper residue Phe80 [27,28].

2.6.1. Direct R^3-Gatekeeper interactions

Here, this question is approached quantitatively based on QM/COSMO interaction 'free' energies between three inhibitors possessing all three substituent types at R^3 (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 $R^3 = H$. Based on this indication of the affinity difference between the R^3 subseries, we proceeded to the modeling and scoring of the whole 21-compound series (Table 1), using the advanced QM/COSMO scoring function [2,3].

2.7. The QM/COSMO scoring of the entire inhibitor series

2.7.1. 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 more stable 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. Fig. S4A). We therefore used the conformation **B** for the QM scoring of all inhibitors possessing the terminal carboxamide (and N,N-dimethyl carboxamide) motif.

2.7.2. 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 the QM-based scoring [3,8], i.e. no explicit water molecules were present and the solvent was modeled by the implicit COSMO model [36]. As shown in Fig. 3A, there was a poor correlation of the QM score with the experimentally derived ΔG_{bind} (expressed as $RT^*\ln(IC_{50}/2)$). 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

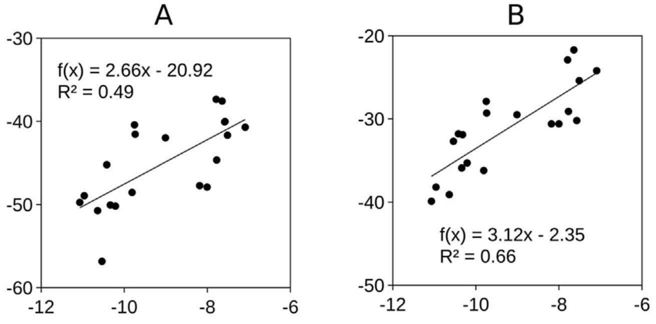


Fig. 3. Calculated QM/SQM/COSMO scores plotted against experimental $RT^*\ln(IC_{50}/2)$, all in kcal/mol, A) pure implicit solvent model, B) mixed explicit/implicit solvent model, i.e. six explicit water molecules added on top of the implicit solvent treatment.

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 $R^3 = H$ and $R^3 = Br/Et$ subseries cannot be explained solely by the small interaction with Phe80 (as calculated in Section 2.6), which corresponds to ca. 1 kcal/mol.

2.7.3. *Mixed explicit/implicit solvent QM scoring*

We have recently realized that structural waters have important consequences for docking and QM scoring [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 $R^3 = H$ 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, Fig. 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 0.66) and the predictive index (PI of 0.80) – see Fig. 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 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 (Fig. 4). Such interactions were responsible for the highest QM/SQM score of **1s**, due to the most negative ΔE_{int}.

3. Conclusions

Herein we report the design, synthesis, activity measurements, and crystallographic and computational analyses of a series of 21 new pyrazolo[1,5*a*]pyrimidine-based CDK2 inhibitors. The compounds bear 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/**1l**.

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 R^3 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 Ni^{2+}NTA column (Qiagen). Kinase (approx. 10 ng) was assayed using a mixture of the following: 1 mg/mL of histone H1, 15 μM of ATP, 0.05 of μCi [γ-33P]ATP, the tested compound, and reaction buffer, in a final volume of 10 μL. The reaction buffer consisted of: 60 mM of HEPES-NaOH, pH 7.5, 3 mM of $MgCl_2$, 3 mM of $MnCl_2$, 3 μM of Na-orthovanadate, 1.2 mM of DTT, and 2.5 μg/50 μl of $PEG_{20.000}$. The reactions were stopped by adding 5 μL of 3% aqueous H_3PO_4. Aliquots were spotted onto P-81 phosphocellulose (Whatman), washed 3 times with 0.5% aqueous H_3PO_4, 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 IC_{50} 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 $(NH_4)_2SO_4$ and KCl as the precipitant agents [39], after the incubation of the protein complex with 2 mM of inhibitor **1l** 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 **1l** was solved by molecular replacement using the program Phaser

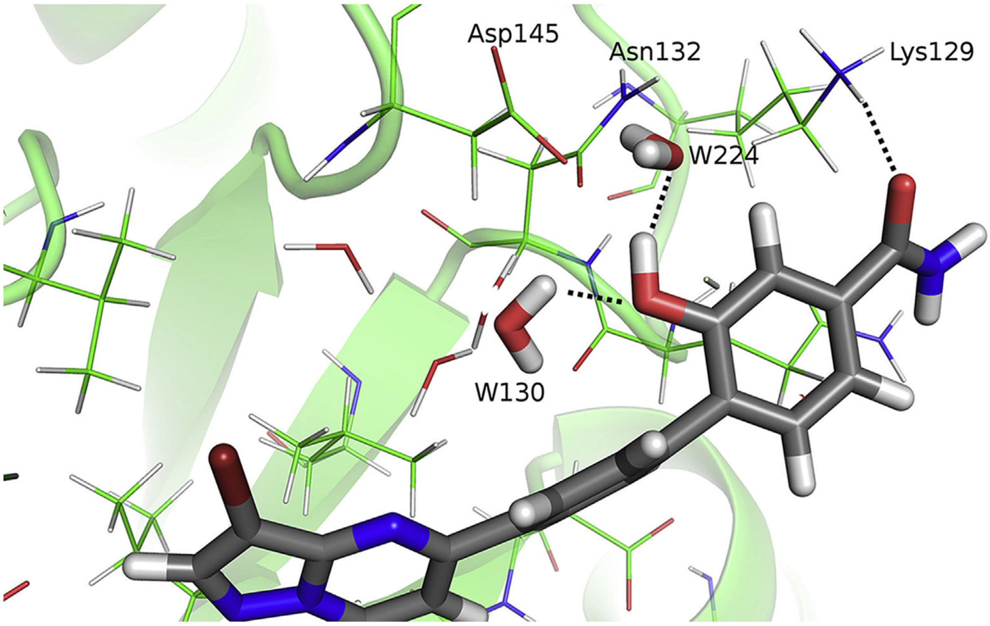


Fig. 4. Interactions of the R^5 substituents of the strongest inhibitor **1s** with CDK2 and two active-site waters. The geometry was obtained by modeling and QM/SQM optimization. Color coding as in Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[41] 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 Coot [43] and refinement in Refmac [44] 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/**1l** 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

4.4.1.1. System setup. The missing loop between the residues 38–41 was modeled using the MODELLER program [45] 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 package [47] using 500 cycles of the steepest descent, followed by 300 cycles of the conjugate gradient method. The CDK2/**1l** 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 field [48] was used. For the ligands, the general AMBER force field (GAFF) [49] with HF/6-31G* RESP charges [50] and fragment-based approach [51] were used. Explicit sigma-hole (ESH) using the “all-fit” (AF) approach [52] was utilized for the ligands with $R^3 = Br$ to describe potential $X \cdots \pi$ interactions.

4.4.1.2. MD simulations. All the MD simulations and most of the analyses were run in GROMACS 4.5.4 [53] 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 tool [55] in VMD 1.9.1.

4.4.2. WaterMap calculations

The CDK2/**1l** and CDK2/**1j** complexes were prepared with the Protein Preparation Wizard in Maestro [56] using the default options. The crystallographic water molecules were retained in CDK2/**1l** 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

4.4.3.1. System setup. Residues 38 and 41 were capped by N-methyl amide and acetyl, 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 (Fig. 2). For the ligands with $R^3 = H$, 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 AMBER14 [47] was used to determine potential flips of Asn, Gln and His residues. Histidines were modeled as neutral, monoprotonated, based on the visual inspection of the surroundings: His121 was protonated on Nδ and other His residues on Nε. 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 AMBER14 [47] 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).

4.4.3.2. Modeling. All 21 inhibitors in complex with CDK2 were modeled from the crystallographic CDK2/**1l** complex in PyMol [59] 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 (see Section 2.5). 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].

4.4.3.3. QM/SQM optimization. We used an improved multi-layer setup [8] 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 [10]. 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.

4.4.3.4. 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 [61]. The rest of the system (SQM region) was calculated by PM6-D3H4X [3] using MOPAC2009 [62].

The binding free energy was approximated by the total score expressed by Eq. (1) [2] with the new consistent notation described in Ref. [3]:

$$\mathrm{Score} = \Delta E_{\mathrm{int}} + \Delta\Delta G_{\mathrm{solv}} + \Delta G'^{\,\mathrm{W}}_{\mathrm{conf}}(\mathrm{L}) - T\Delta S_{\mathrm{solv}} \qquad (1)$$

The individual terms describe the gas-phase interaction energy (ΔE_{int}), the interaction solvation/desolvation free energy ($\Delta\Delta G_{\mathrm{solv}}$), the change of the conformational “free” energy of the ligand ($\Delta G'^{\mathrm{W}}_{\mathrm{conf}}(\mathrm{L})$), and optionally the entropy of the explicit water molecules ($T\Delta S_{\mathrm{solv}}$). It should be mentioned, however, that another portion of the entropy contribution is already included in the $\Delta\Delta G_{\mathrm{solv}}$ 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: COSMO [36] at the PM6 level and SMD [63] at the HF/6-31G** level (the latter only for the ligands to increase the accuracy) [10]. For the evaluation of the change of the conformational “free” energy of the ligand ($\Delta G'^{\mathrm{W}}_{\mathrm{conf}}(\mathrm{L})$), 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.

4.4.3.5. QM fragmentation. The interaction energies between the inhibitor and protein fragments (side chains and peptide bonds) were determined at the DFT-D3 (TPSS/TZVPP) [61] level combined with the COSMO implicit solvent model [36].

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”. VK thanks the Czech Science Foundation (15-15264S). RJ thanks Czech Science Foundation (15-17282Y). 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-OPENSCREEN: National Infrastructure for Chemical Biology (Identification code: LM2015063).

Supplementary data

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 are presented in the Supporting Information. Furthermore, NMR, IR and MS spectra of all the compounds are presented as Appendix to Supplementary data.

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejmech.2016.12.023.

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. Fanfrlik, 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. Rezác, 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. Rezáč, E. Munusamy, M. Lepšík, P. Matějíček, 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. Brahmkshatriya, 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. Soderhjeim, 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. Kekenes-Huskey, A. McCammon, J. Dzubiella, Solvent fluctuations in hydrophobic cavity–ligand binding kinetics, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 1197–1202.

[16] B. Vorlová, D. Nachtigallová, J. Jirásková-Vaníčková, H. Ajani, P. Jansa, J. Rezáč, 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) 130–146.

[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. Brahmkshatriya, P. Dobeš, J. Fanfrlik, J. Rezác, 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. Havlícek, 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. Rezác, 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. Penaflor, 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. Echalier, Y. Ferandin, J.A. Endicott, H. Galons, L. Meijer, CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases, Oncogene 27 (2008) 5797–5807.

[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 Sci. 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, University of California, San Francisco, 2014.

[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 J. 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] 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.

[62] J.J.P. Stewart, MOPAC2012, Stewart Computational Chemistry, Colorado Springs, CO, 2012. http://OpenMOPAC.net.

[63] 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.

PUBLICATIONS

Publication D

Saltuk M. Eyrilmez, Cemal Köprülüoğlu, Jan Řezáč, and Pavel Hobza

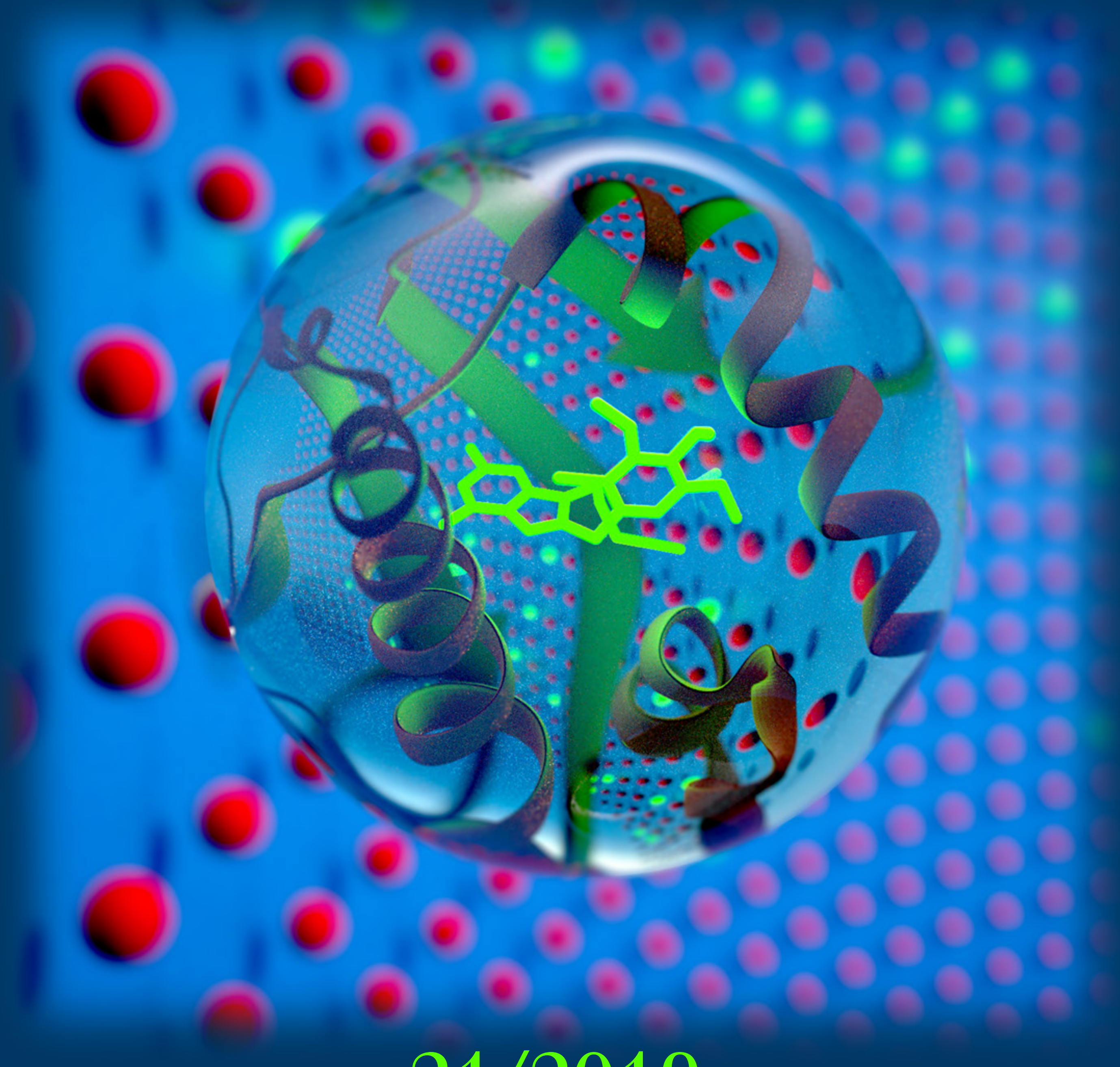
Impressive Enrichment of Semiempirical QuantumMechanics-Based Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys

ChemPhysChem, **2019**, 20, 2759-2766

A EUROPEAN JOURNAL

CHEMPHYSCHEM

OF CHEMICAL PHYSICS AND PHYSICAL CHEMISTRY



21/2019

Front Cover:

S. M. Eyrilmez, C. Köprülüoğlu et al.

Impressive Enrichment of Semiempirical Quantum Mechanics-Based Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys

WILEY-VCH

www.chemphyschem.org

A Journal of ChemPubSoc Europe

ChemPubSoc
Europe

DOI: 10.1002/cphc.201900975

CHEMPHYSCHEM
Cover Profile

Impressive Enrichment of Semiempirical Quantum Mechanics-Based Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys

Saltuk M. Eyrilmez

Cemal Köprülüoğlu

Jan Řezáč

Pavel Hobza

ÚOCHB AV ČR
IOCB PRAGUE

www.uochb.cz

The image shows the ability of SQM-based frame to separate the actives (background: green spheres) from inactives (red spheres) while maintaining a powerful sampling (front: HSP90 crystal complex). Read the full text of the Article at 10.1002/cphc.201900628.

What is the most significant result of this study?

The enrichments obtained with nine standardly used docking/scoring functions were low, sometime even below the random-value limit. An impressive enrichment increase was achieved when protein-ligand structure was optimised and score determined at the quantum mechanical level.

What prompted you to investigate this topic/problem?

Ligands are bound in the active site of the protein by non-covalent interactions which have been studied in our laboratory for decades. We have repeatedly shown the importance of quantum effects being the case of protein-ligand complexes as well. This has prompted us to introduce a quantum mechanics-based scoring function.

What other topics are you working on at the moment?

Our main project is characterisation of non-covalent interactions; currently we solve two sub-projects – Non-Covalent interactions in biomolecular systems including protein-ligand and protein-protein complexes and Non-Covalent interactions in advanced functional materials. Needless to say, that these interactions are the same in both mediums.

What was the biggest surprise (on the way to the results presented in this paper)?

Undoubtably, a dramatic improvement of Enrichment Factor passing from the single scoring function to SQM-based scoring function was the most surprising finding.

Acknowledgements

We are grateful to Dr. Federico Urban for helpful discussion. This work was part of the Research Project RVO: 61388963 of the Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences. We acknowledge the support from the European Regional Development Fund; OP RDE; Project: 'Chemical Biology for Drugging Undruggable Targets (ChemBioDrug)' (No. CZ.02.1.01/ 0.0/0.0/16_019/0000729). This work was also supported by the Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project 'IT4Innovations National Supercomputing Center – LM2015070'.

A EUROPEAN JOURNAL

CHEMPHYSCHEM

OF CHEMICAL PHYSICS AND PHYSICAL CHEMISTRY

21/2019

Front Cover:
S. M. Eyrilmez, C. Köprülüoğlu et al.
Impressive Enrichment of Semiempirical Quantum Mechanics-Based
Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys

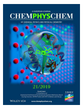
WILEY-VCH

www.chemphyschem.org

A Journal of
ChemPubSoc
Europe

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

DOI: 10.1002/cphc.201900628



Impressive Enrichment of Semiempirical Quantum Mechanics-Based Scoring Function: HSP90 Protein with 4541 Inhibitors and Decoys

Saltuk M. Eyrilmez+,[a, b] Cemal Köprülüoğlu+,[a, b] Jan Řezáč,[a] and Pavel Hobza*[a, b]

This paper describes the excellent performance of a newly developed scoring function (SF), based on the semiempirical QM (SQM) PM6-D3H4X method combined with the conductor-like screening implicit solvent model (COSMO). The SQM/COSMO, Amber/GB and nine widely used SFs have been evaluated in terms of ranking power on the HSP90 protein with 72 biologically active compounds and 4469 structurally similar decoys. Among conventional SFs, the highest early and overall enrichment measured by EF_1 and AUC% obtained using single-scoring-function ranking has been found for Glide SP and Gold-ASP SFs, respectively (7, 75% and 3, 76%). The performance of other standard SFs has not been satisfactory, mostly even decreasing below random values. The SQM/COSMO SF, where P–L structures were optimised at the advanced Amber level, has resulted in a dramatic enrichment increase (47, 98%), almost reaching the best possible receiver operator characteristic (ROC) curve. The best SQM frame thus inserts about seven times more active compounds into the selected dataset than the best standard SF.

1. Introduction

The determination of the structure and properties of protein-ligand (P–L) complexes is a key task in structure-based drug design.[1] To this end, numerous docking and scoring functions (DF, SF) have been developed and tested. Besides classical approach to docking/scoring based on empirical, knowledge- or physics-based methods,[2] a quantum mechanics (QM) approach was pioneered by Merz et al.[3] The former approaches, dependent on the existence of a sufficiently broad training set, are based on molecular mechanics (MM) methods (or their simplification), whereas the latter approach utilises the 'objective' QM method. The main advantage of QM methods is the fact that they cover quantum effects (e.g. charge transfer or σ-hole binding), which might play an important role in P–L interactions. Due to the size of P–L complexes, mostly semiempirical QM (SQM) methods as AM1,[4] PM3,[5] PM6,[6] PM7[7] and DFTB3[8] are applied. None of these methods is, however, directly suitable for the investigation of noncovalent complexes.[9] For this purpose, correction terms ensuring the proper description of dispersion, electrostatic and σ-hole interactions should be included. Throughout the present paper, the advanced D3H4X[10] correction term has been used in combination with the PM6 method (PM6-D3H4X). The method can be applied to extended P–L complexes with several thousand atoms because the PM6 method can be combined with the MOZYME linear scaling algorithm implemented in MOPAC.[11] P–L complexes exist in a solvent environment, which affects their structure and properties. To model the solvent, we used the conductor-like screening implicit solvent model (COSMO).[12] The binding free energy between the protein and ligand in a solvent is approximated by the score [Eq. (1)]:

$$SCORE = \Delta E_{int} + \Delta\Delta G_{solv} + \Delta G'^{w}_{conf}(P) + \Delta G'^{w}_{conf}(L) + T\Delta S_{int} \qquad (1)$$

expressed as the sum of the gas-phase P–L interaction energy (the first term), the change of solvation/desolvation free energy upon complex formation (the second term), the change of the conformation 'free' energies of the protein and ligand (the third and fourth terms), and the entropy change upon binding (the fifth term).[13,14] The first two terms, having the opposite sign (the first term is always stabilising while the second one is destabilising), are clearly dominant and SQM/COSMO-based SFs are mostly based only on them.[15]

The evaluation of the performance of DFs and SFs is mainly based on the estimation of their sampling and ranking power, where the first refers to the ability of SF to predict the position of a native ligand correctly. We have recently investigated the sampling power of several widely used classical SFs as well as quantum mechanics-based SFs developed in our laboratory. PM6 and DFTB3 SQM methods were combined with PM6/COSMO, systematically based on the PM6 characteristics. Four[16] and seventeen[17] different P–L complexes were studied, and the SQM/COSMO SFs clearly outperformed all classical SFs. Slight improvement was achieved when the less empirical but considerably more expensive DFTB3 method was applied.[18–20] Ranking power, describing the ability of SF to rank different ligands of the same target protein (based on binding affinities), is a more difficult task. In the first study,[21] we investigated a

---

[a] *S. M. Eyrilmez,+ C. Köprülüoğlu,+ Dr. J. Řezáč, Prof. P. Hobza*
Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences
Flemingovo nam. 2, 16610 Prague 6 (Czech Republic)
E-mail: pavel.hobza@uochb.cas.cz
[b] *S. M. Eyrilmez,+ C. Köprülüoğlu,+ Prof. P. Hobza*
Regional Centre of Advanced Technologies and Materials
Palacký University, 77146 Olomouc (CzechRepublic)

[+] *These authors contributed equally to this work*

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

relatively easy case, where the structures of the complexes of carbonic anhydrase II with ten different ligands were known, what reduces the problem to the determination of binding free energy. In addition, in this case the SQM/COSMO SF provided much better results than ten different classical SFs.

Structure-based virtual screening can only be successful if the method can reliably predict the geometry of the P–L complex (the binding mode) and the SF used provides reliable ranking. It is now widely accepted that it is beyond the ability of the currently used SFs to meet both requirements. As shown above, the SQM/COSMO-based SF was more successful in both these scenarios. The present study has applied this methodology (with further extensions) to a larger and more diverse set of P–L complexes, the HSP90 protein and its about 5000 compounds from the DUD–E database. The HSP90 (heat shock protein 90) is a protein that stabilises various growth receptors[22] and some signalling molecules[23] required for the survival of cancer cells.

The structures of P–L complexes were not known and were thus determined by docking. The PM6/COSMO and Amber/GB SFs and nine widely used classical SFs were evaluated in terms of docking and scoring. The SQM/COSMO and Amber/GB SFs were only applied for ranking, and the respective poses were generated by different classical SFs. Since the biological activities of all ligands and decoys are not known in detail and are characterised only in terms 'active' and 'inactive', direct correlation between activities and theoretical scores is impractical. It should be added here that even if ligand affinities are known, correlation between them and calculated scores is difficult, sometimes[24] even denoted as being 'beyond the current methods'. To describe the ability of the method to distinguish between active and inactive ligands, we use the enrichment factor, a quantity that distinguishes known ligands from decoys.

1.1. Strategy

The knowledge of the native structure of the P–L complex is crucial for the estimation of the biological activity of a ligand. If the experimental structure of the P–L complex is missing, it is possible to use the theoretical structure determined by the gradient optimisation of binding free energy (Eq. 1). Such an approach is not only CPU-time demanding but, and this is more serious, it mostly leads to a local minimum at the free energy landscape. There is an enormous number of the local minima for P–L complexes, and it is clearly impractical to search the whole landscape at the SQM level. We have chosen an alternative route; DFs are used to generate a large number of poses, which will be scored in the next step by means of SQM/COSMO SF. We are aware that a reliable identification of a native P–L pose with a single DF is a difficult task. Therefore, we have intentionally used nine different DFs to increase the possibility of finding a native binding pose. The disadvantage of this procedure is clear – the SQM/COSMO calculations should be performed for thousands of P–L structures. Therefore, the SQM/COSMO SF should be computationally as efficient as possible. For that reason, we have introduced a semiempirical quantum mechanics-based virtual screening frame, which eliminates redundant poses and produces high-quality structures to increase the efficiency and applicability of demanding PM6/COSMO calculations. The structures of P–L complexes determined by DFs should be optimised at the molecular mechanics level. This work has used the AMBER[25] biomolecular simulation package for geometry optimisations (MM_N). To increase the reliability of subsequent SQM scoring, restrained AMBER optimisations have been applied as well. The bond length and bond angle values have been taken from PM6 optimised compound structures (MM_A). In the present study, we have used test compounds taken from 'A Database of Useful Decoys: Enhanced' (DUD–E)[24] for the HSP90 target. This set contains 4850 decoys, 25 experimental inactive compounds with similar physical properties (e.g. molecular weight, calculated logP) but dissimilar 2-D topology, and 88 actives with known experimental binding affinities. The ligands included 19 macrocycle-containing molecules. Since considerable effort might be needed to rationalise the protocol, these compounds were excluded from the actives in the first place.[26] 406 of the decoys were also not considered due to the computational reasons.

1.2. Scoring

Within the present scoring framework, the score was approximated without the entropy change [Eq. (2)]:

$$SCORE = \Delta E_{int} + \Delta\Delta G_{solv} + \varDelta \mathrm{G}^{'\mathrm{w}}_{\mathrm{conf}}(P^{H\ atoms}) + \varDelta \mathrm{G}^{'\mathrm{w}}_{\mathrm{conf}}(L) \qquad (2)$$

where the first, second and fourth terms were identical to these in Equation (1), while in the third term only hydrogens were considered in optimisation.

Three types of scoring were applied: MM scoring using Amber/GB SF and two types of SQM scoring based on SQM/COSMO SFs, denoted as SQM_1 and SQM_2, where MM_N and MM_A optimised structures were utilised.

In the case of multiple protonation states, each state was scored individually and the one with the minimum score was used for enrichment analysis.

MM scoring: The scoring scheme shown in Equation (2) was applied for MM scoring using the MM_N optimised structures of the complex and ligands. Since the application of MM_A optimisations deteriorated the Amber energies, the MM scoring over MM_A-optimised structures was not performed.

SQM scoring: The key point for any SQM scoring in this virtual screening study was to decrease the redundant poses before processing them at the PM6/COSMO level. To achieve this, we first applied RSMD clustering with a 1 Å cut-off to eliminate similar poses produced by MM_N optimisations. Representative poses were selected as the MM_N minimum complex structures. The complexes for the subsequent SQM scoring were selected on the basis of MM_N optimisations (the complexes within the 10 kcal/mol energy interval were taken into consideration).

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Two types of SQM scoring (SQM_1 and SQM_2) using the same SQM/COSMO SF but different optimisation schemes, denoted as MM_N and MM_A, were considered. We have used Cuby4[27] software to automate our fragmentation, optimisation and energy calculation protocols.

1.3. Analysis

For each of these SFs, the score of all ligand poses binding to the respective target protein was calculated, ranked and plotted. The performance of the scoring functions was evaluated based on the analysis of the enrichment factor (EF) and receiver operator characteristic (ROC) plots.[28] The accuracy of virtual screening was evaluated using EF. Calculated score values were ranked, and EF was defined as [Eq. (3)]:

$$EF_{subset} = (ligand_{selected}/N_{subset})/(ligand_{total}/N_{total}) \tag{3}$$

where $ligand_{total}$ is the number of known ligands with activity against the target, N_{total} is the number of all compounds in the dataset, $ligand_{selected}$ is the number of found ligands in a given subset, and N_{subset} is the total number of the compounds in the subset. EF_{subset} provides information about the number of the true positives among the decoys in the given subset in comparison with a random selection.[29] Generally, the top part of the library of the ranked compounds was used for further evaluation and was strongly dependent on the initial library size. The size might range from 0.1 % to 10 % and it was considered as 1 % in the present study.[30]

ROC curves were obtained by plotting sensitivity (Se) and specificity (Sp), where:

$$Se_{subset} = (ligand_{selected}/Ligands_{total}) \times 100$$

$$Sp_{subset} = [(Decoys_{total} - Decoys_{selected})/Decoys_{total}] \times 100$$

The ROC curves were plotted as (100%–Sp%) (i.e. % of selected decoys) versus Se% (i.e. % of selected active compounds).[31]

The AUC was defined as the area under a ROC curve. It is simply the probability that a randomly chosen active has a higher score than a randomly chosen inactive. In other words, the AUC is the average of this property over all inactive fractions.[28]

The results were also supported by the pROC AUC values, which focus on early enrichment.[32][33] pROC AUC values for random enrichment were determined as follows [Eq. (4)]:

$$\lim_{a\to 0}\int_a^1 (-\log_{10}X)dX = \frac{-1}{\log 10}\lim_{a\to 0}\int_a^1 (\log X)dX$$
$$= 0.434\lim_{a\to 0}\{X - X\log X\}|0pt1a = 0.434 \tag{4}$$

Table 1. The ROC enrichment factors (EF1), AUC (in %) and pROC AUC obtained for single-docking-function ranking (DF).

DF	EF1	AUC [%]	pROC AUC
AD4	1	49	0.383
VINA	0	30	0.192
SMINA	0	34	0.224
GlideSP	7	75	0.880
GlideXP	4	71	0.730
ASP	3	76	0.787
Gscr	0	60	0.488
Cscr	0	34	0.270
PLP	1	51	0.383

2. Results and Discussion

Table 1 shows the performance of nine different SFs where a single SF was used for both scoring and ranking. The analysis is based on EF_1, AUC and pROC AUC characteristics determined for the average property over all inactive fractions.

Evidently, the results are not satisfactory, especially concerning the early-stage enrichment (EF_1 values). The EF_1 values of six out of nine SFs were equal to or below random values ($EF_1 = 1$) and only SP, XP and ASP SFs provided EF_1 above this limit. The overall performance measured by AUC values was slightly better – the AUC values of five SFs were below random values (50 %), whereas SP, XP, ASP as well as Gscr were above them, while PLP equalled the random performance. The best among the single-scoring-function ranking SFs were GlideSP, GlideXP and ASP, having the highest EF_1, AUC% and pROC AUC characteristics (7, 75 %, 0.880; 4, 71 %, 0.730 and 3, 76 %, 0.787). Note that GlideSP exhibits higher early enrichment, which is more important for drug-design purposes, while ASP has better overall performance. The combined performance of all SFs was, however, poor. In all the cases, the single-scoring-function ranking SFs had been applied. It was thus necessary to decide whether the problem originated in incorrect structures determined by docking or in the incorrect score determined by ranking. The question was to be answered in the next step by a combined study using different SFs for scoring and ranking. It is known that rescoring with the different docking functions can improve the enrichment significantly.[34] Each scoring function has been therefore sampled extensively to fill the active pocket as complete as possible. We collected 100 poses from each docking software. Comparing results presented in the Tables 1 and 2 we found that none of empirical functions (cf. Table 1) could reach the respective results presented in the Table 2.

Table 2 summarises the enrichment where docking was made by a single standard DF while ranking was performed by MM, SQM_1 and SQM_2 SFs. A comparison with the corresponding values from Table 1 clearly shows the dramatic improvement when binding free energies (SCOREs) have been evaluated at the MM and both SQM levels. The SQM_1 results will be discussed first. All combinations of SQM_1 ranking with poses generated by different DFs have provided the enrichment values considerably above the random values. The highest early and overall enrichment was obtained for Gscr, SMINA, AD4, GlideSP and PLP structures. Considering the pROC AUC values

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. The ROC enrichment factors (EF$_1$), AUC (in %) and pROC AUC obtained for SQM$_2$//DF (a combination of scoring and docking; the P–L structures were optimised with the MM$_A$ method), SQM$_1$//DF (the P–L structures were optimised with the MM$_N$ method) and MM//SF.

SF//DF	EF1	AUC [%]	pROC AUC
SQM$_2$//AD4	40	91	2.104
SQM$_1$//AD4	25	70	1.262
MM//AD4	15	86	1.304
SQM$_2$//VINA	42	93	2.052
SQM$_1$//VINA	27	67	1.277
MM//VINA	13	83	1.208
SQM$_2$//SMINA	37	93	1.997
SQM$_1$//SMINA	31	69	1.371
MM//SMINA	17	84	1.343
SQM$_2$//SP	34	81	1.670
SQM$_1$//SP	15	82	1.368
MM//SP	15	76	1.277
SQM$_2$//XP	32	85	1.710
SQM$_1$//XP	14	65	0.872
MM//XP	23	83	1.493
SQM$_2$//ASP	31	93	1.877
SQM$_1$//ASP	24	66	1.093
MM//ASP	18	91	1.418
SQM$_2$//Gscr	44	97	2.329
SQM$_1$//Gscr	31	74	1.350
MM//Gscr	14	90	1.352
SQM$_2$//Cscr	29	89	1.757
SQM$_1$//Cscr	19	62	0.880
MM//Cscr	14	82	1.229
SQM$_2$//PLP	31	95	2.096
SQM$_1$//PLP	27	68	1.193
MM//PLP	3	88	1.333

the VINA DF shows a better performance than PLP even though both DFs have similar EF$_1$ and AUC results. Glide SP exhibits the best overall performance but low EF$_1$ values. On the other hand, the pROC AUC result was the second best. The SP DF thus provides early stage success as demonstrated by pROC AUC value and the best overall performance (see AUC% value in the Table 2). For drug discovery, as mentioned above, early evaluation is more important; therefore, preference should be given to SMINA, SP, Gscr and VINA DFs combined with SQM/COSMO SF. Much better enrichment performed by combined SQM$_1$ ranking and DF docking provides evidence that all standard SFs have problems with the determination of binding free energies while their geometries are reliable. Surprisingly high enrichment, especially an overall one, was obtained when MM SF was applied. The best results in the overall performance were obtained with ASP and Gscr DFs. The very good performance of MM is promising for the future investigation of extended P–L complexes, because MM is much less CPU-time demanding. It should be noted that the SQM$_1$ results discussed above were obtained with the P–L structures optimised with the standard MM$_N$ method. On the other hand, the SQM$_2$ results in Table 2 were obtained with the P–L structures optimised with the MM$_A$ method. Evidently, this systematically resulted in significantly higher enrichment. Considering the SQM$_1$ values, only the Gscr structures provided AUC values higher than 70%. When the SQM$_2$ values were considered, five of the SF structures exceeded 90% limit and the PLP and Gscr values even reached 95 and 97%, respectively.

A similar dramatic increase was found for the early enrichment, where five out of nine EF$_1$ values were higher than 31 (this value was not exceeded by any EF$_1$ for SQM$_1$ and MM) and the highest EF$_1$ was detected for SQM$_2$//Gscr (44), beside this, pROC AUC value is five times better than the random. A comparison of the entries in the Table 2 clearly shows that high enrichment is only obtained if reliable binding modes are used. Evidently, the poses generated by docking are not sufficiently accurate and significant enrichment increase is only obtained after their re-optimisation at the MM$_A$ level. The question arises whether comparable results can be expected for other proteins as well. The necessary condition for it is the generation of reliable structures. There is no reason to expect that a DF that has generated reliable structures for some protein will also succeed for another one. To make the method more robust, it is thus beneficial to use more DFs for the generation of ligand poses. To test this approach, we have collected ligand poses from all the DFs considered in the present paper; the subsequent ranking was performed with SQM$_1$, SQM$_2$ and MM SFs. The consideration of the poses from all DFs provided an enrichment increase when SQM$_1$ and SQM$_2$ SFs were used (cf Table 3). When SQM$_1$, SQM$_2$ and MM methods were applied, the AUC values reached 75%, 98% and 92%, respectively, and highest enrichment was achieved when the SQM$_2$ method was used. The highest AUC values obtained with the same methods where only the structures generated by a single DF were used equalled 74%, 97% and 91%, respectively. The EF$_1$ values for SQM$_1$, SQM$_2$ and MM methods (where the structures of all DFs were used) amounted to 32, 47, 10, and, again, the highest EF$_1$ was obtained for SQM$_2$. The pROC AUC value is 2.477 which means the performance is six times better than the random case. When only the structures generated by a single DF were used EF$_1$ equalled 31, 44 and 23, respectively. The consideration of the structures from all DFs improved early and overall enrichment for SQM$_1$ and SQM$_2$, the effect was not dramatic. We have seen a decline of EF$_1$ for MM but improved overall performance. The reason for the decline might be due to the energy ranking of all the structures generated by all DFs. The above-mentioned results are valid for the present protein. For different targets situation might be different and the use of structures from more DFs is thus recommendable.

Table 3. The ROC enrichment factors (EF$_1$), AUC (in %) and pROC AUC obtained for SQM$_2$//ALL (a combination of scoring and docking; the P–L geometries from all SFs were optimised with the MM$_A$ method), SQM$_1$//ALL (a combination of scoring and docking; the P–L geometries from all SFs were optimised with the MM$_N$ method) and MM//ALL.

SF//DF	EF1	AUC [%]	pROC AUC
SQM$_2$//ALL	47	98	2.477
SQM$_1$//ALL	32	75	1.426
MM//ALL	10	92	1.395

3. Conclusions

The enrichment obtained with single-scoring-function ranking was low for all nine conventional SFs. Several SFs provided enrichment even below the random-value limit. Only four SFs (SP, XP, ASP and Gscr) provided enrichment above random values. Evidently, no single SF succeeds in both docking/scoring and ranking.

The enrichment increased when SQM_1, SQM_2, and MM ranking was determined for poses generated by standard SFs. This gives evidence that standardly used SFs provide reliable poses but fail for ranking. On the other hand, SQM_1, SQM_2 as well as MM SFs yield reliable ranking.

A significant enrichment increase was achieved when P–L structure optimisation was performed within the SQM_2 frame. The enrichment (AUC) obtained by five out of nine SFs exceeded 90 %, and the PLP and Gscr AUC values even reached 95 % and 97 %. Impressive enrichment in terms of both EF_1 and AUC resulted when the Gscr, AD4 and PLP structures were re-optimised at the MM_A level (44, 97 %, 2.329; 40, 91 %, 2.104 and 31, 95 %, 2.096 respectively). Using the PM6 parameters in the MM treatment improves the geometry of the ligand, what leads to better geometries of the P–L complex and, consequently, to higher enrichment. The consideration of all poses provided an enrichment increase for SQM_1 and SQM_2 methods, EF_1, AUC and pROC AUC values rose to 32, 75 %, 1.426 and 47, 98 %, 2.477 respectively.

The overall enrichment after MM_A application to P–L structure optimisation was very close to the best ROC AUC limits.

The standard approach to virtual screening is based on the use of single-scoring-function ranking. The highest enrichment in EF_1, AUC and pROC AUC (7, 75 %, 0.880 and 3, 76 %,0.787 respectively) was obtained using the GlideSP and ASP SFs. Passing from the best single-scoring-function ranking to the advanced SQM treatment led to a dramatic increase. A combination of the SQM SF with P–L optimisation using MM_A provided impressively high EF_1, AUC and pROC AUC values (47, 98 %, 2.477). The enrichment factor obtained included 34 (out of 72) experimentally active structures in the subset. This means that nearly 50 % of actives are found in 1 % of the whole dataset. In other words, the present SQM_2 SF frame inserts about seven times more active compounds into the selected dataset and three times better pROC performance than the best SF. This clearly demonstrates the impressive performance of the SQM_2 frame in both early and overall enrichment. The values of the overall enrichment are close to the best ROC curve. We are certainly aware that all these findings are based on the investigation of a single protein. Intensive work in our laboratory is currently being performed for targets from other protein families.

The above-mentioned findings clearly demonstrate the advantage of using SQM SFs over the standard ones. We believe that despite higher CPU demands, the wider application of SQM SFs could be beneficial not only for structure-based drug design but also for related applications.

Figure 1 shows the visualisation of docking results using a novel Post Dock tool[35] implemented in MOE Software.[36] Six (A, C, D, E, F, G) out of nine DFs provided binding modes close to the crystal pose. Their transparencies were high which means the respective DFs provided correct poses with the worse score. In another word, they are good in sampling but failed in ranking. On the contrary, figures B, H and I demonstrated that the individual DFs failed to generate the correct binding modes. As it seen in the Figure 1J SQM provided less transparent yellow colour which means that the binding mode totally matches with the crystal pose. It implies that SQM SF was able to select the crystal pose with the highest score. These results show that increasing the number of poses for the individual DFs was important for finding the crystal pose. For instance, in the case of SQM SF (Figure 1J) the binding mode with the best score fully agreed with the crystal pose while AD4 DF (Figure 1A) found the best agreement with the crystal pose for pose number 90 having the worse score. Evidently that the use of SQM SF is required for obtaining both successful sampling and ranking.

Computational Section

Compound Preparation

The compounds were downloaded in the SMILES format and prepared using the LigPrep module with an Optimised Potentials for Liquid Simulations (OPLS3e)[37] force field. Their ionisation states were generated at pH 7.0 ± 2.0 using Epik[38] in LigPrep.[39] Specific chiralities were retrained during the ligand preparations. The structures generated within the state penalty value of 0–1 were saved for docking calculations.

Protein Preparation

The process of protein preparation requires special care in physics-based SFs.[16] Protein has been downloaded from the Protein Databank[40] with the 1UYG PDB code.[41] We have decided to keep three conserved water molecules (W2121, W2123 and W2236). To implement the selection, we first aligned the PDB structures with 100 % sequence similarity to 1UYG. According to the Ref.[42] we selected the intersection set of the most favourable water molecules (W1, W3,W4). Hydrogens of the protein were added by using reduce program, which is part of the AMBER18 suite. The protonation states of each histidine residue were assigned manually based on hydrogen-bonding patterns. Hydrogen positions were relaxed by the simulated annealing protocol using short molecular dynamics (MD). The protocol includes the optimisation of hydrogens, annealing and optimisation in the solvent igb7[43] model. The MD protocol was the following: the initial temperatures were assigned following Maxwell Boltzmann distribution to the target temperature of 1000 K. They were kept at 1500 K for 1 ps and then cooled down to 0 K over 2 ps. Optimisation was carried out employing the Broyden-Fletcher-Goldfarb-Shanno algorithm using a limited amount of computer memory with the igb7 solvent model.

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

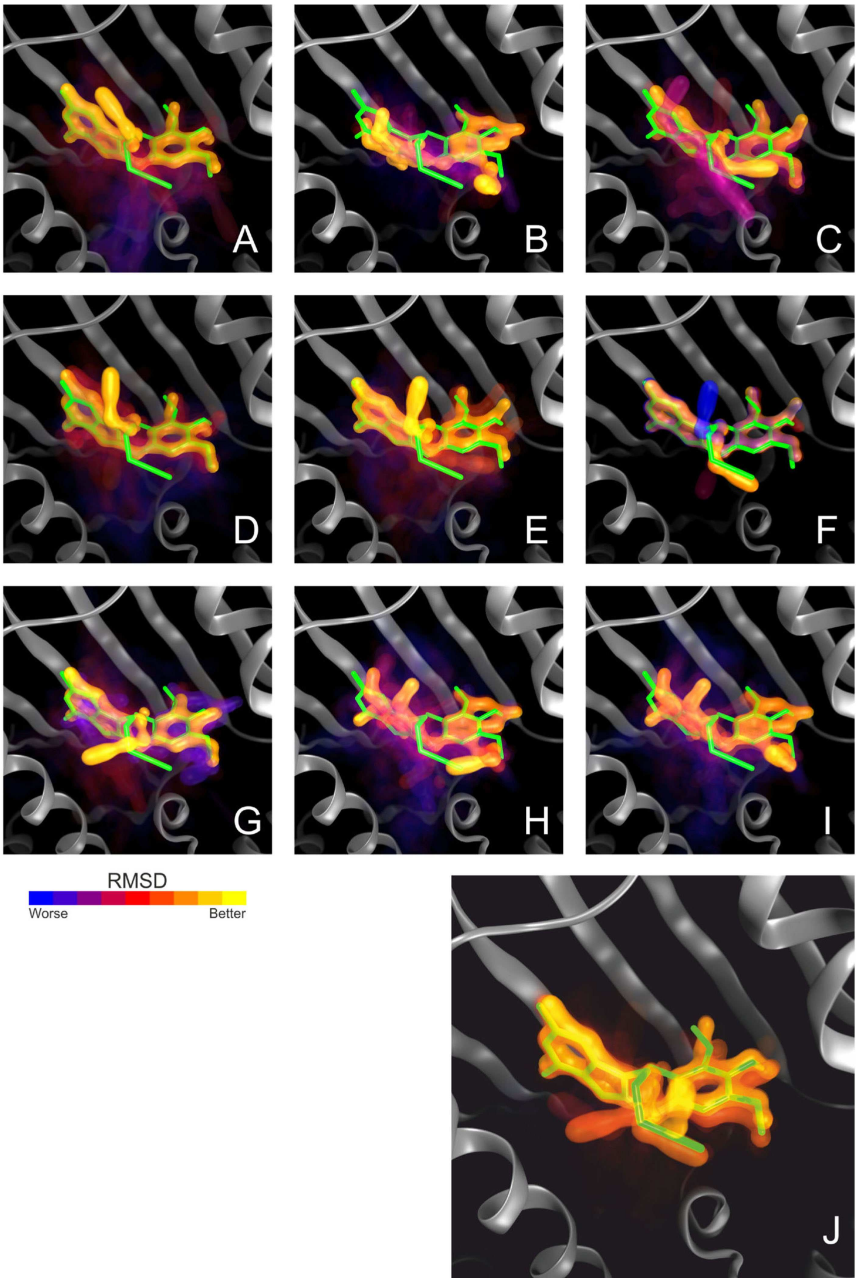


Figure 1. Individual representation of RMSD vs Energy for docking and scoring results of 9 different DFs and the scoring result of SQM-based SF: A) AD4, B) ASP, C) CSCR, D) GSCR, E) GlideSP, F) GlideXP, G) PLP, H) SMINA, I) VINA and J) represent the results of SQM. For these results, 1UYF crystal ligand has been used for the comparison in Post Dock.[35] It displays an interactive pseudo-3D snapshot of multiple docked ligand poses such that both the docking poses and docking scores are encoded visually for rapid assessment. The docking energies are represented by a transparency scale whereas the docking poses are visually encoded by a colour scale. Reference ligand localization in the binding site is shown in green colour with the stick model. The poses from the docking functions are shown in the surface model. The colours from the tinted yellow to the faint blue represent the RMSD values. Yellow colour corresponds to the lowest RMSD and a blue colour corresponds to the highest RMSD. Regarding the opaqueness, the opaquest surface represents the lowest energy pose with a better score and the score is getting worse when the transparency increase.

Dockings

We have examined the poses generated by nine docking functions: Glide(SP,XP),[44] AutoDock4,[45] Vina,[46] Smina,[47] and GOLD software,[48–50] using ASP,[51] GoldScore,[48][49] ChemPLP[52] and ChemScore.[51] All hydrogens of the compounds and the receptor were explicitly preserved during all docking calculations to make it possible to see every possible interaction for different protonation states of the same molecule. We have changed the upper limit of the pose production to 100 for all DFs while keeping other settings as default. The centre coordinates of the grid were assigned as the geometrical centre of crystal inhibitor and used for all docking functions.

The grid centre was adjusted in MGLtools. In our grid box all the possible interactions have been checked and presented. Based on the x,y,z centres 20 Å^3 grid box covered all the poses. 20 Å^3 grid

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

box was prepared for AutoDock4, AutoDock Vina as well as SMINA. Grid earch space size was specified in “grid points” (0.375 Angstrom), as in AutoDock 4. Docking receptor grid in Glide was generated using the Glide Receptor Grid Generation module. A cubic box of 20 $Å^3$ was placed at the grid box centre. For Gold dockings, Grid centre coordinates were used for binding site origin. The radius value was defined as 12.4 Å in order to produce the same volume as the grid boxes for previous DFs.

Fragmentation

The fragmentation of the protein step was applied to reduce the computational cost for demanding PM6/COSMO calculations for the complex. For this reason, all docked pose coordinates were gathered to generate a reference volume for the fragmentation. The fragmented protein part (receptor) was defined as a selection of protein residues within a 4 Å distance from the reference volume, truncated, and capped by using Cuby4. Hydrogen sampling and optimisation processes were applied as explained in the protein preparation section. MM_N and PM6/COSMO energies of the receptor were noted for scoring calculations.

Scoring Preparation

We used ffPM3[53] for protein, tip3p[54] for water molecules, gaff2 for compounds and the igb7 model for the solvation of AMBER calculations. We assigned partial atomic charges by means of the AM1-BCC[55] charge model implemented in antechamber.[56] Individual input complex structures were generated from docked poses and the receptor. The MM preparation of the complexes were initiated as a 2 ps MD step and the optimisation of the ligand hydrogen atoms and the surrounding H atoms of the receptor within 4 Å with respect to ligand heavy atoms. The hydrogen sampling of the complex step was followed by another optimisation of all ligand atoms along with 4 Å surrounding hydrogen atoms of the receptor.

All compound conformations were also optimised by MM_N and MM_A protocols and single-point PM6/COSMO energy calculations were applied for further deformation penalty inclusion.

Acknowledgements

We are grateful to Dr. Federico Urban for helpful discussion. This work was part of the Research Project RVO: 61388963 of the Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences. We acknowledge the support from the European Regional Development Fund; OP RDE; Project: ‘Chemical Biology for Drugging Undruggable Targets (ChemBioDrug)’ (No. CZ.02.1.01/0.0/0.0/16_019/0000729). This work was also supported by the Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project ‘IT4Innovations National Supercomputing Center – LM2015070’.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: docking · enrichment · non-covalent interactions · semiempirical quantum mechanics-based scoring function · virtual screening

[1] Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, T. Hou, *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964–12975.
[2] H. Gohlke, M. Hendlich, G. Klebe, *J. Mol. Biol.* **2000**, *295*, 337–356.
[3] K. Raha, K. M. Merz, *J. Am. Chem. Soc.* **2004**, *4*, 1020–1021.
[4] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *13*, 3902–3909.
[5] J. J. P. Stewart, *J. Comput. Chem.* **1991**, *3*, 320–341.
[6] J. J. P. Stewart, *J. Mol. Model.* **2007**, *12*, 1173–1213.
[7] J. J. P. Stewart, *J. Mol. Model.* **2013**, *1*, 1–32.
[8] M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, M. Elstner, *J. Chem. Theory Comput.* **2015**, *11*, 332–342.
[9] J. Řezáč, P. Hobza, *Chem. Rev.* **2016**, *116*, 5038–5071.
[10] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
[11] J. J. P. Stewart, MOPAC 2016, Stewart Computational Chemistry, Colorado Springs, CO: **2016**.
[12] A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2* **1993**, *0*, 799–805.
[13] J. Fanfrlík, A. K. Bronowska, J. Řezáč, O. Přenosil, J. Konvalinka, P. Ho, *J. Phys. Chem. Btry B* **2010**, *114*, 12666–12678.
[14] M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, J. Fanfrlík, *ChemPlusChem* **2013**, *78*, 921–931.
[15] M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, J. Fanfrlík, *ChemPlusChem* **2013**, *78*, 921–931.
[16] A. Pecina, R. Meier, J. Ich Fanfrlí, M. Lepš, Í K. J. Řezá, P. Hobza, C. Baldauf, *Chem. Commun. Chem. Commun* **2016**, *3312*, 3312–3315.
[17] H. Ajani, A. Pecina, S. M. Eyrilmez, J. Fanfrlík, S. Haldar, J. Řezáč, P. Hobza, M. Lepšík, *ACS Omega* **2017**, *2*, 4022–4029.
[18] J. Řezáč, *J. Chem. Theory Comput.* **2017**, *13*, 4804–4817.
[19] J. Řezáč, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 141–151.
[20] J. Řezáč, K. E. Riley, P. Hobz, *J. Chem. Theory Comput.Ion* **2011**, *7*, 2427–2438.
[21] A. Pecina, J. Brynda, L. Vrzal, R. Gnanasekaran, M. Hořejší, S. M. Eyrilmez, J. Řezáč, M. Lepšík, P. Řezáčová, P. Hobza, P. Majer, V. Veverka, J. Fanfrlík, *ChemPhysChem* **2018**, *7*, 873–879.
[22] A. Sawai, S. Chandarlapaty, H. Greulich, M. Gonen, Q. Ye, C. L. Arteaga, W. Sellers, N. Rosen, D. B. Solit, *Cancer Res.* **2008**, *68*, 589–596.
[23] C. E. Stebbins, A. A. Russo, C. Schneider, N. Rosen, F. U. Hartl, N. P. Pavletich, *Cell* **1997**, 89, 239–250.
[24] M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, *J. Med. Chem.* **2012**, *55*, 6582–6594.
[25] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668–1688.
[26] O. Gutten, D. Bím, J. Řezáč, L. Rulíšek, *J. Chem. Inf. Model.* **2018**, *58*, 48–60.
[27] J. Řezáč, *J. Comput. Chem.* **2016**, *37*, 1230–1237.
[28] A. Nicholls, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239–255.
[29] H. Fan, J. J. Irwin, B. M. Webb, G. Klebe, B. K. Shoichet, A. Sali, *J. Chem. Inf. Model.* **2009**, *11*, 2512–2527.
[30] H. Chen, P. D. Lyne, F. Giordanetto, T. Lovell, J. Li, *J. Chem. Inf. Model.* **2005**, *1*,401-415.
[31] N. Huang, B. K. Shoichet, J. J. Irwin, *J. Med. Chem.* **2006**, *23*, 6789–6801.
[32] R. D. Clark, D. J. Webster-Clark, *J. Comput.-Aided Mol. Des.* **2008**, *3–4*, 141–146.
[33] S. M. Vogel, M. R. Bauer, F. M. Boeckler, *J. Chem. Inf. Model.* **2011**, *10*, 2650–2665.
[34] O. Korb, T. Ten Brink, F. R. D. V. P. Raj, M. Keil, T. E. Exner, *J. Comput.-Aided Mol. Des.* **2012**, *2*,185-197.
[35] E. A. Wiley, G. Deslongchamps, *Computing and Visualization in Science* **2009**, *1*, 1–7.
[36] Chemical Computing Group ULC, *Molecular Operating Environment (MOE)*, Montreal, Quebec, Canada **2018**.
[37] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
[38] J. C. Shelley, A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin, M. Uchimaya, *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.
[39] Schrödinger, *Schrödinger Release 2018–2* **2018**.

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[40] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–42.
[41] L. Wright, X. Barril, B. Dymock, L. Sheridan, A. Surgenor, M. Beswick, M. Drysdale, A. Collier, A. Massey, N. Davies, *Chem. Biol.* **2004**, *11*, 775.
[42] K. Haider, D. J. Huggins, *J. Chem. Inf. Model.* **2013**, *53*, 2571–2586.
[43] J. Mongan, C. Simmerling, J. A. McCammon, D. A. Case, A. Onufriev, *J. Chem. Theory Comput.* **2006**, *1*, 156–169.
[44] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, P. S. Shenkin, *J. Med. Chem.* **2004**, *7*, 1739–1749
[45] G. M. Morris, D. S. Goodsell, M. E. Pique, R. Huey, S. Forli, W. E. Hart, S. Halliday, R. Belew, A. J. Olson, *J. Comput. Chem.* **2009**, *16*, 2785–2791.
[46] O. Trott, A. J. Olson, *J. Comput. Chem.* **2009**, *31*, 455–461.
[47] D. R. Koes, M. P. Baumgartner, C. J. Camacho, *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904.
[48] J. W. M. Nissink, C. Murray, M. Hartshorn, M. L. Verdonk, J. C. Cole, R. Taylor, *Proteins Struct. Funct. Genet.* **2002**, *49*, 457–471.
[49] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727–748.
[50] G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *1*, 43–53.
[51] W. T. M. Mooij, M. L. Verdonk, *Proteins Struct. Funct. Bioinf.* **2005**, *61*, 272–287.
[52] O. Korb, T. Stützle, T. E. Exner, *J. Chem. Inf. Model.* **2009**, *1*, 84–96.
[53] A. M. Wollacott, K. M. Merz, *J. Chem. Theory Comput.* **2006**, *4*,1070-1077.
[54] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[55] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.
[56] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

Manuscript received: June 26, 2019
Revised manuscript received: August 21, 2019
Accepted manuscript online: August 28, 2019
Version of record online: September 11, 2019

© 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

PUBLICATIONS

Publication E

Adam Pecina, Saltuk M. Eyrilmez, Cemal Köprülüoğlu, Vijay Madhav Miriyala, Martin Lepšík, Jindřich Fanfrlík, Jan Řezáč, and Pavel Hobza

SQM/COSMO Scoring Function: Reliable Quantum-Mechanical Tool for Structure-Based Drug Design

ChemPlusChem, **2020**, submitted

Manuscript

Click here to download Manuscript Pecina_CPCminireview2020.docx

SQM/COSMO Scoring Function:

Reliable Quantum-Mechanical Tool for Structure-Based Drug Design

Adam Pecina*[a], Saltuk M. Eyrilmez[a,b], Cemal Köprülüoğlu[a,b], Vijay Madhav Miriyala[a], Martin Lepšík[a], Jindřich Fanfrlík[a], Jan Řezáč[a], and Pavel Hobza*[a,b]

[a] Dr. A. Pecina, S. M. Eyrilmez, C. Köprülüoğlu, Dr. V. M. Miriyala, Dr. M. Lepšík, Dr. J. Fanfrlík, Dr. J. Řezáč and Prof. P. Hobza
Institute of Organic Chemistry and Biochemistry of Czech Academy of Sciences
Flemingovo namesti 2, CZ-166 10, Prague, Czech Republic
E-mails: pecina.adam@gmail.com, pavel.hobza@uochb.cas.cz

[b] S. M. Eyrilmez, C. Köprülüoğlu, Dr. V. M. Miriyala and Prof. P. Hobza
Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacky University
CZ-771 46, Olomouc, Czech Republic
E-mail: pavel.hobza@uochb.cas.cz

Abstract: Our decade-long experience in reliable description of noncovalent interactions in biomolecules using quantum mechanical methods at different levels of theory enabled us to develop a semiempirical quantum mechanical (SQM)-based scoring approach for structure-based drug design. The experience gained in applying it to tens of protein targets and thousands of ligands resulted in setting up a faster SQM/COSMO approach, which outperforms standard academic and commercial scoring procedures in native pose identification, activity ranking of different compounds as well as early enrichment in virtual screening. Due to its superior performance, feasibility and chemical generality, we propose the SQM/COSMO approach as an efficient tool in structure-based computer-aided drug design.

1. Introduction

The ultimate goal of *in silico* structure-based drug design is to find and optimize ligands (L) of a pharmaceutically relevant target (usually protein, P) in terms of affinity and specificity. To this end, the known three-dimensional structure of the protein is probed whether it could bind compounds from a diverse set of small organic molecules in a process called *virtual screening*.[1] Its integral parts are *docking* which seeks to recognize a geometrical fit and *scoring* which estimates the binding affinity in the P-L complex.[2]

Given the number of compounds which need to be tested *in silico* (amounting to 10^7), multiplied by their numerous protonation states and tautomers as well as orientations and conformations in the protein binding site, the total number of P-L complexes to be evaluated can be as high as 10^{12}-10^{13}.[3] That is why there has been an enormous pressure on the efficiency of docking/scoring calculations. They have thus historically relied on drastic approximations which often lead to compromising their accuracy.[3] The main avenues along which improvement has been achieved over the last decades are: i) flexibility, entropy and kinetics of P-L complexes, ii) the role of water in P-L recognition, and iii) scoring function developments. The former problematic is broad and important and we refer the reader to the excellent reviews.[4] The second topic ranges from developments of implicit solvent models[5] to inclusions of explicit bridging water molecules to P-L docking/scoring.[6] The third area has had many developments in knowledge-based or empirical scoring functions or using machine learning[4a] but our focus is on chemically general and reliable description of P-L noncovalent interactions, i.e. using quantum mechanical (QM) methods. Moreover, QM methods treat inorganic and covalently-bound ligands without need of any special-purpose parametrization.[7] Other reviews describe the application of QM methods to evaluating P-L binding.[8]

In our laboratory, we have paid special attention to QM description of non-classical noncovalent interactions, such as halogen or chalcogen bonds (σ-hole interactions)[9], which occur in P-L complexes.[10] However, accurate QM calculations are very demanding - their cost scales steeply with the system size. To avoid this bottleneck, we and others have been using two approaches: i) hybrid QM/MM approach in which density functional theory (DFT) is used to treat the ligand and its closest environment and MM accounts for the rest of the P-L complex[8a, 8c, 11] and ii) semiempirical quantum-mechanical (SQM) methods which are applicable to thousands of atoms but needed further parametrization to overcome their limitations in the description of noncovalent interactions.[12]

In this minireview, we cover the use of SQM methods in structure-based drug design since our last review[13] and highlight the latest advances, especially in the area of virtual screening.[14]

WILEY-VCH

Adam Pecina received his PhD from the Charles University, Prague (under the supervision of Pavel Hobza). He did postdoctoral studies at the Institute of Organic Chemistry and Biochemistry (IOCB) of the Czech Academy of Sciences with Pavel Hobza and at the Italian Institute of Technology (IIT) with Marco De Vivo. His research aims at computationally driven drug design and molecular mechanisms of pharmaceutically relevant enzymes and nanozymes. He is currently a Marie Skłodowska-Curie Fellow hosted at IIT in Italy.

Pavel Hobza is a Distinguished Chair and Group leader at IOCB and Professor of physical chemistry at Charles University, Prague and Palacky University, Olomouc. He specializes in theoretical description of non-covalent interactions and related applications in chemistry and life-sciences. He is the author of more than 500 papers, 4 books and 35 review papers (with H-index of 104). He is in the top one percentage of cited chemists in the world (awarded the "Highly Cited Researcher" by Thomson Reuters list, later Clarivate Analytics) and also a Schrödinger Medal laureate.

2. SQM methods applicable to protein – ligand complexes

The SQM methods used nowadays fall into two broader classes. The "classical" SQM methods based on the neglect of differential diatomic overlap approximations include the MNDO, AM1, PM3, PM6, PM7 and OMx methods.[15] Similar in complexity are the methods based on self-consistent-charge tight-binding, either DFT-based[16] (SCC-DFTB) or empirical (e.g. GFN2-xTB[17]). The works reviewed here use mainly PM6, PM7 and third-order SCC-DFTB (abbreviated as DFTB3).[15d, 15e, 18] These methods cover the chemical space needed for applications to biomolecules and, unlike MM, they do not need any system-specific parametrization. They are very fast, and can be combined with linear-scaling algorithms (MOZYME)[19] for PM6 and PM7 or divide-and-conquer for SCC-DFTB[20] that make them applicable to systems with thousands of atoms.

However, the approximations making the SQM methods so efficient lead to rather poor description of noncovalent interactions. As a result of the one-electron approximation, London dispersion is missing. The electrostatic interactions are also substantially simplified. This, combined with the use of sub-minimal basis set that limits atomic polarization leads to the most important deficiency of SQM methods - a strong underestimation of the strength of hydrogen bonds.

Recently, many of these deficiencies were addressed by specific empirical corrections, which in turn enabled the applications discussed here. London dispersion corrections (D, D3) had been added to semiempirical methods[21] analogously to the extremely successful DFT-D methodology,[22] and additional corrections for hydrogen bonding (H, H+, H2, H4, H5)[23] had been introduced. Besides London dispersion and hydrogen-bonding, non-classical noncovalent interactions such as halogen bonding from the family of σ-hole interactions[9b] deserved their own parametrization (X).[24] All these corrections have been parametrized to reproduce accurate benchmark CCSD(T)/CBS interaction energies[25] in databases of small model complexes[9a] where they can achieve accuracy better than 1 kcal/mol.

The results discussed here were obtained mostly with the advanced version of the corrections we have developed, a method-specific re-parametrization of the D3 dispersion and H4 correction for hydrogen bonding and X for halogen bonding. Among other methods, these are applicable to PM6 (PM6-D3H4X) and DFTB3 (DFTB3-D3H4). The corrections are independent of the SQM calculation and a standalone implementation is available[26], but nowadays they are available directly in MOPAC[27] for PM6/PM7 and DFTB+[28] for DFTB3 and no post-processing of the results is needed. For DFTB3, a more advanced H5 correction recently became available[23a], and preliminary results show that it improves the description of P – L complexes further.[29] For systems containing halogens, a halogen-bonding correction should be applied, and it is available for both PM6 and DFTB3.[12d, 24] The PM7 method already contains analogous corrections and no further modifications are needed. It is more robust than PM6, which is useful in more exotic systems, but not as accurate as PM6-D3H4, especially in larger complexes.[30]

Until recently, the accuracy of these methods was measured only in small model systems[9a] and their performance in P – L complexes was evaluated only indirectly by comparing the final score with experimental data. Recently, we have built data sets of model systems directly related to this application, the PLA15 set of 15 active site models and PLF475 set of their smaller fragments, featuring DLPNO-CCSD(T) benchmarks.[29] With these, we can directly evaluate the accuracy of SQM interaction energies used in the applications reviewed here. Figure 1 summarizes the main results in the PLA15 data set, a set of 15 different proteins in which the interaction between the ligand and a large model of the active site is evaluated.

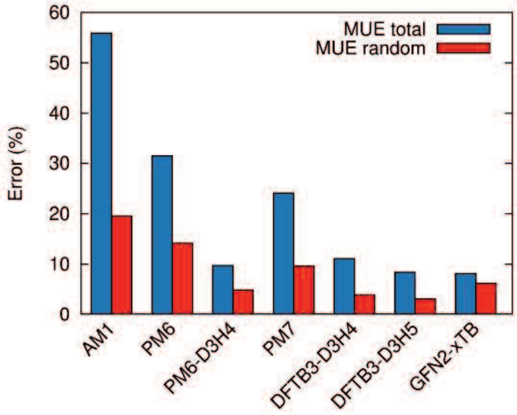


Figure 1. Error of selected SQM methods in the PLA15 data set of 15 ligand – active site models. Reported is the total mean unsigned error (MUE) and its random part

Here, we evaluate the total mean unsigned error (MUE, evaluated from relative errors to normalize the differences between the systems) as a measure of an absolute accuracy of the methods, and its random part (MUE calculated after the systematic error of each method is removed). The latter error measure is closer to the applications in scoring where only relative energies in a series of systems matter. The corrected SQM methods reach absolute accuracy of around 10%, and the random part of the error can be less than half of that. These results suggest that the methods discussed here most often, PM6-D3H4 and DFTB-D3H4, are best in their class.

3. Solvation

For biological applications, a major phenomenon to be described besides the P–L interactions *in vacuo* are the solvation phenomena of both, the P–L complexes as well as the binding partners prior to binding. This is certainly a challenging task which warrants the use of different approximations. For the approaches discussed here which use static geometries, we resort to implicit solvent models, reviewed in breadth in Decherchi et al.[5a] Out of a dozen of available implicit solvent models, we describe below only those pertinent to our work. For the description of protein solvation, especially in conjunction with force field calculations, two widely used approaches are the *Generalized Born* (GB)[31] and *Poisson-Boltzmann (PB)*[32] model. The pair-wise GB model is very fast but rather approximate. PB is more accurate but also more time demanding, as it relies on the solution of Poisson-Boltzmann (PB) equations. We have tested these molecular mechanics (MM)-based models against three QM-based solvation models; MST[33], SMD[34] and COSMO-RS[35]. We have shown that the QM-based models yielded more accurate solvation free energies than the MM-based ones but at a higher computational cost.[5b]

A seemingly ideal compromise is the use of the COSMO model employing the scaled-conductor approximation[36] which is implemented for SQM calculations in MOPAC.[27] However, the implementation covers only the electrostatic contribution to the solvation free energy and neglects the other terms important in non-polar systems. We have thus introduced a simple non-polar term and reparametrized the COSMO model for PM6 and PM7, yielding COSMO2.[37] It has improved the accuracy not only in the calculations of solvation free energies of small molecules, but also in the scoring of protein-ligand complexes. Most notably, we have observed significant improvement in correlation between the score computed using PM6-D3H4/COSMO2 and experimental binding energies in complexes where structural waters are present in the active site, but not included in the model.[37] For DFTB3, multiple solvent models are available, but their implementation is not optimized for large systems and makes the P – L calculations impractically slow. Therefore, we usually resort to combining DFTB3 interaction energies with separately calculated PM6/PM7 COSMO solvation energy.

4. SQM-based Scoring Functions

SQM-based Scoring Function (SF) was firstly introduced by Kenneth Merz group, by using AM1 method augmented with empirical dispersion and combined with Poisson-Boltzmann implicit solvent model and force-field dispersion term.[38] They showed a superior performance of the QMScore over other SFs in the case of metalloprotein-ligand binding,[39] however further corrections were needed, especially for hydrogen bonding and dispersion.[40]

Some recent studies have also employed QM/SQM scoring functions to calculate P-L binding affinities.[41] Another strategy of combining QM and SQM is to involve fragmentation-based methods[8a, 42], such as in the three layer MIM3 protocol designed for estimating the relative strength of P-L binding for a set of structurally similar ligands.[43] To speed up the binding affinity prediction, the combination of SQM/MM approach have been applied on different P-L systems. For example, SCC-DFTB energies and CHARMM force field was successfully used in the on-the-fly QM/MM study of high-level X-ray structures of zinc metalloproteins.[44] The same method, coupled with the Attracting Cavities docking algorithm[45] was applied on Astex Diverse data set with comparable performance with standard SFs, however the method outperformed standard ones in the zinc metalloprotein data set and a heme protein data with dominating iron or covalent ligand binding.

Number of new studies have successfully used PM6 method in combination with our corrections for noncovalent interactions to score P-L systems. The new systematic study of Kenneth Merz group introduced a local sampling algorithm for producing binding ensembles of small aromatic ligands with T4 Lysozyme L99A mutant and they showed the best correlation with experimental binding affinities using PM6-DH2/COSMO-based score.[46] SQM-based scoring was recently used also for binding energy calculations of peptide-HLA (human leukocyte antigen) system. Methodology, based on PM6-D3H4, PM7 and FMO-DFTB3 potentials with inclusion of structural and solvent effects, successfully differentiated between binders and non-binders and showed a strong correlation with experimental data.[47]

Other methods also aim to rescore MD trajectories by SQM-based end point methods as shown in several applications on human protein kinase Lck, C-terminal domain of breakpoint cluster region (BRC) protein or estrogen receptors alpha.[48]

In our laboratory, we have taken a systematic approach to design a modular general-purpose SQM SF.[13] We have adopted the MM-PB/GBSA-like master equation[49] (Eq. 1) in which the interaction, solvation and ligand deformation terms are evaluated at the SQM/COSMO levels.[50] The modularity resides in the fact that the most up-to-date version of the SQM/COSMO methods is always used.

The score which approximates the binding free energy is computed on the P-L complex optimized in water environment (Eq 1)[13]

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

The individual terms describe the gas-phase interaction energy (ΔE_{int}), the change of solvation free energy upon complex formation ($\Delta\Delta G_{solv}$), the change of conformational ‘free’ energy ($\Delta G'^{w}_{conf}$) of the protein and ligand in water environment and the

WILEY-VCH

change of entropy upon ligand binding ($-T\Delta S_{int}$). The first term is dominant since it describes gas-phase (undamped) interaction between protein and ligand. It is the only favorable term in the equation and its reliable estimation is crucial. The unfavorable solvation/desolvation term represents the second largest term. Polar or even charged ligands are strongly hydrated. Before entering the active site of protein, they must be dehydrated and the respective desolvation free energy is very large, in absolute value might even be comparable to the gas-phase interaction energy. Calculation of solvation free energy is difficult and is connected with larger uncertainties than gas phase energy, especially for charged systems.[37] It is important to mention that gas-phase interaction energies and solvation/desolvation free energies do not correlate simply because they are due to different physicochemical properties. Both terms partially compensate for each other and the final score is considerably smaller than their absolute values. The other terms of Eq.1 are usually smaller than the first two terms, however any variations in them can change the ranking of ligands. Up to date, the above mentioned scheme at different levels have been successfully applied on series of inhibitors of various kinases[6d, 13, 51], proteases[13, 52], aldo-keto reductases[10a, 10b, 53] and also of DNA polymerase[54] and serine racemase.[6b]

The scoring function (SF) has also been extended to treat covalent inhibitor binding.[7b] We have also used it for interaction 'free' energy evaluation in modified insulin analogues[55], SH3 domains in complex with their ligands[56] or in combination with virtual glycine scan procedure[57] for quantifying the contribution of individual protein residues to inorganic ligand-metalloprotein binding[57-58] or at protein-protein interfaces.[59]

Despite the above-mentioned list of successful applications of SQM-based scoring, the full version of the SQM-based SF is too time-consuming to be applicable to drug design process in industrial context. The optimization of P-L complex at the SQM level is the most demanding step. Moreover, the gradient optimization leads just to the closest local minimum on the potential energy surface (PES). To find the global minimum, one should apply molecular dynamics or other methods to sample PES. Consequently, we simplified and accelerated the SQM based SF by employing only the dominant ΔE_{int} and $\Delta\Delta G_{solv}$ terms and replacing the costly full SQM optimization by sampling of PES of P-L binding by extensive docking. We thus generate an ensemble of binding poses of the ligand in the active site of its target and just apply short hydrogen atom relaxation of such complexes by AMBER forcefield[60] to relieve the majority of close contacts that may be introduced by the soft repulsive potentials used in docking algorithms. This novel scheme is about two orders of magnitude faster than the full version of SQM based SF.

We have made the first step toward QM-based high-throughput virtual screening[14a] by validating the ability of the SQM/COSMO SF at PM6-D3H4X level to identify the native binding pose and compare it with several standard SFs. We selected four difficult-to-handle P-L systems: i) HIV-1 PR with large, flexible and charged peptidomimetic ligand; ii) aldose reductase representing the enzymes with cofactor in the active site; iii) acetylcholine esterase with two binding pockets and halogenated ligand and iv) TNF-α converting enzyme (TACE) representing Zn metalloproteins. To this aim, we generated a large amount of sensible and non-redundant alternative ligand binding poses and checked how well the different scoring approaches (physics-based AMBER/GB[60-61], empirical/regression-based Glide XP[6a], Plants PLP[62], AutoDock Vina[63], Gold's Chemscore[64], Goldscore[65], PLP[62] and knowledge-based Astex Statistical Potential ASP[66]) were able to differentiate between the alternative and native states (Figure 2). The best performing SQM/COSMO SF identified the X-ray pose as the minimum-free-energy structure in 3 out of 4 cases. The second-best SFs were ASP and ChemScore having over 50 ligand poses that were scored better than the native one (so called False-Positive solutions; FP). Other SF estimated from 80 to 350 FPs. None of the SFs were able to correctly estimate the native pose of the TACE metalloprotein. Here all SFs have FPs but to a different extent, where SQM/COSMO SF performed the best, having the smallest number of 39 FPs.[14a] The follow-up study showed that the challenges of Zn metalloprotein can be overcome by invoking a more robust SQM method (i.e. DFTB3-D3H4X) in combination with $\Delta\Delta G_{solv}$ term evaluated at the COSMO level from PM6-D3H4X calculation.[14b] The less empirical DFTB3 described the interaction between the ligand and Zn(II) ion with a smaller error than PM6-D3H4, which resulted in zero FPs for the TACE complex. Importantly, the description of the other three systems stayed at a high quality – zero FP were retained. This study clearly showed the importance of higher-level electronic-structure theory for reliable description of protein-ligand binding.[14b]

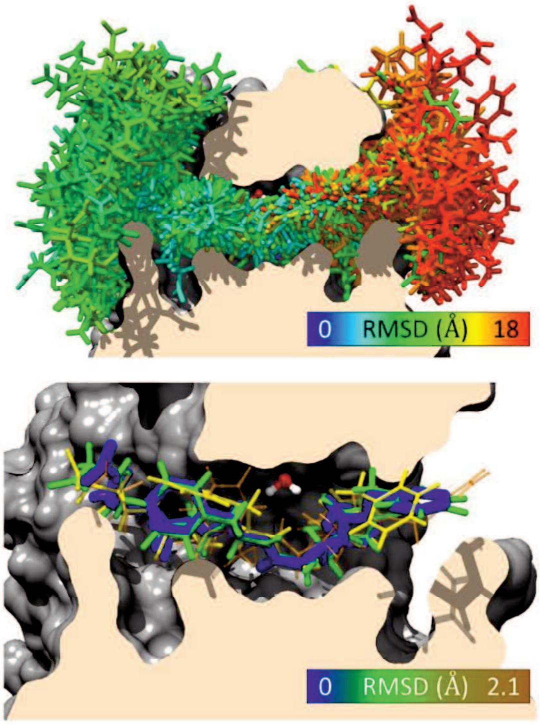


Figure 2. The sampling power of SQM/COSMO SF as shown in the case of HIV-1 PR[14a] All ligand poses generated by docking programs, that are color-coded by RMSD relative to the xray (native) pose (top panel). The native pose (in blue sticks) uniqly recognized by SQM/COSMO SF as the minimum-free-energy structure, in comparison with higher energy values of decoy poses with small changes in the binding geometry (bottom panel).

The sampling power of the SQM/COSMO SFs at PM6 and DFTB levels were further tested on an extended dataset of 17 P-L systems from diverse protein families (including 3 enzyme classes, 1 chaperone and 2 nuclear receptors classes)[67] We compared SQM/COSMO SFs with 4 standard SFs (Glide XP[6a], AutoDock4[68], AutoDock Vina[63], and UCSF Dock[69]). The SQM/COSMO SF performed best, having one order of magnitude smaller number of FPs (40 and 42 for DFTB3 and PM6, respectively) than standard ones (from 211 to 635 FPs). Moreover, the SQM/COSMO SF's number of FPs for neutral ligands were single digits only (1 and 2, respectively) in comparison with a range from 18 to 85 for standard SFs.[67]

Identifying the correct native pose of the P-L systems (sampling power) is a critical prerequisite for affinity estimation in physics-based scoring. The second step is a reliable prediction of P-L affinities (ranking power). For this task, we prepared a dataset of 10 carbonic anhydrase II (CAII)-inhibitor complexes, determined their high-resolution crystal structures (resolution 1.1.-1.4 Å) and consistently measured their inhibition constants (K_i).[14c] Using all the information from the X-ray structures, e.g. binding mode, position of active-site waters and protein conformation, we compared scoring results of SQM/COSMO SF at DFTB3-D3H4X level (because CAII is a zinc metalloenzyme) with 5 widely used classical SFs (AMBER/GB, GOLD, DOCK 6, AutoDock Vina, Autodock4). For SQM/COSMO, we obtained a reasonable correlation with experimental binding data (R^2 of 0.69, predictive index PI[70] of 0.81), which greatly outperformed all commonly used classical SFs (R^2 < 0.4 and PI < 0.7).[14c] Then, we scrutinized the effects of omitting crystal waters and using a single rigid conformation of the target protein, still retaining the unique correlation (R^2 of 0.56 and PI 0.64 for SQM/COSMO vs. R^2 < 0.3 and PI < 0.5 for classical SFs). In the last step, we allowed small conformational movements of ligands by restrained docking which again improved prediction of the affinities (R^2/PI of 0.77/0.92 for SQM/COSMO).[14c] The SQM/COSMO method thus showed a great potential to become a general tool during the hit-to-lead stage of structure-based drug design after further testing.

SQM/COSMO SF: Library Enrichment

To obtain "good results" in structure-based virtual screening for good reasons, we need a protocol which reliably predicts geometry of the complex and on top of that provides reliable ranking. Despites decades of development, the fulfillment of both requirements is beyond the ability of current docking and scoring approaches.[4a] Having witnessed the successes of the SQM/COSMO SF in both sampling and ranking power, we entered the area of virtual screening with the aim to prioritize active ligands over inactive ones toward a target protein. We selected HSP90 (heat shock protein 90) and a set of 4541 compounds which contained 72 inhibitors and 4469 decoys from the DUD-E database[71]. Nine widely used docking functions (Autodock4[68], AutoDock's VINA[63] and SMINA[72], Glide (SP,XP)[6a], Gold's ASP[66], GoldScore[65], ChemScore[64] and ChemPLP[62]) were applied for docking and PM6-D3H4X/COSMO SF was used for rescoring.[73] The performance was evaluated by analyzing the enrichment factor (EF) while testing 1% of the compounds (the ratio of the identified actives from the total number of compounds; EF_1) and receiver operator characteristics (ROC and pROC) plots[74] which provide a detailed view of overall and early enrichment. ROC curves show the capability of the method to distinguish between active and inactive compounds, while pROC focuses on the early stage enrichment (see Figures 3 and 4).

Table 1 shows that all the standard docking/scoring functions failed, especially for early enrichment (EF_1 values sometimes even below random values; EF_1 equal to 1).

Method	EF_1	AUC (%)	AUC pROC
Autodock4 (AD4)	1	49	0.38
SQM//AD4	40	91	2.10
Autodock VINA	0	30	0.19
SQM//VINA	42	93	2.05
Autodock SMINA	0	34	0.22
SQM//SMINA	37	93	2.00
GlideSP	7	75	0.88
SQM//SP	34	81	1.67
GlideXP	4	71	0.73
SQM//XP	32	85	1.71
Gold ASP (ASP)	3	76	0.79
SQM//ASP	31	93	1.88
GoldScore (GS)	0	60	0.49
SQM//GS	44	97	2.33
ChemScore (CS)	0	34	0.27
SQM//CS	29	89	1.76
ChemPLP (PLP)	1	51	0.38
SQM//PLP	31	95	2.10
SQM//all	47	98	2.48

Table 1. The ROC Enrichment Factors EF_1, AUC and pROC AUC obtained by using a single method or a combination of methods (scoring//docking). Data were taken from Ref.[73]

Among conventional SFs, the best performance was found for Glide SP and XP SFs (EF_1 of 7 and 4, respectively) and for Gold's ASP in the case of overall enrichment. The SQM/COSMO SF rescoring increased both early and overall enrichments tremendously. Area under the ROC curve (AUC) reached 90 % in 6 out of 9 cases. When all the poses were combined and optimized with a modified AMBER force field, very encouraging EF_1 and AUC values of 47 and 98%, respectively, were obtained. These values were even approaching best possible characteristics (i.e. 64 and 100%). The advanced SQM/COSMO SF thus covered almost seven times more active ligands than the most efficient standard SF. Results presented in Table 1 clearly demonstrated that almost all standard SFs provided reasonable geometries but scored poorly. However, there was not a single standard SF which would succeed in both docking and scoring.

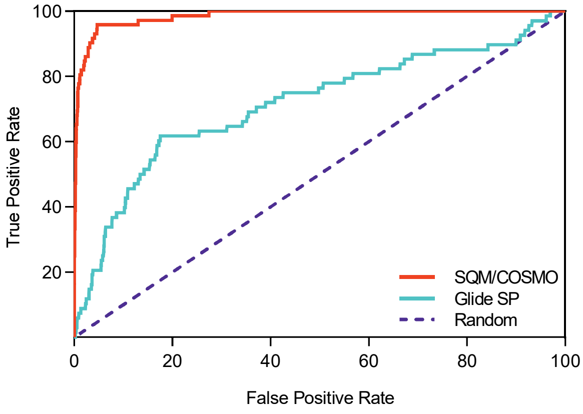


Figure 3. ROC curve for the comparison of the overall enrichment performance of the best standard SF (Glide SP) in blue, SQM/COSMO SF in red and random (dark blue).

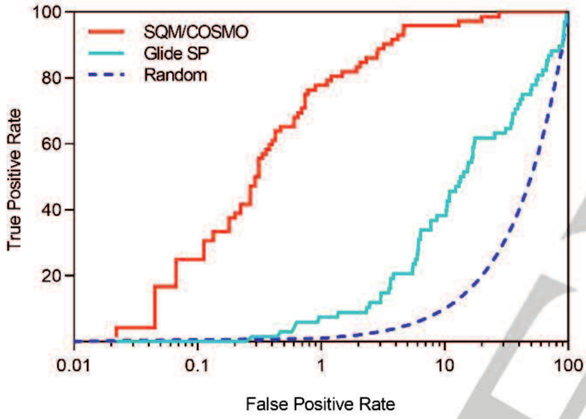


Figure 4. pROC curve for the comparison of the early enrichment performance of the best standard SF (Glide_SP) in blue, SQM/COSMO SF in red and random (dark blue).

The success in early enrichment which we have observed for HSP90 protein may in part be caused by the high rigidity of the protein. Currently, we are extending this computational strategy to other proteins which are known to be more flexible.

As mentioned in the Introduction, computer-aided structure-based drug design can profit from developments in three areas: i) by accounting for the P-L complex flexibility, ii) by better describing the P-L complex environment, such as improving the implicit solvation model or adding explicit water molecules and iii) by adopting a better model for P-L binding. During our decade-long research on the development of SQM-based and SQM/COSMO SFs, we have improved the accuracy of both the interaction and solvation free energy methods and have successfully applied it to dozens of protein targets and thousands of ligands. We have attained a remarkable sampling and ranking power, outperforming standard scoring functions, which allowed us to extend the development toward virtual screening. In this latter area, we have achieved a first successful application and currently are working on others.

On the other hand, we admit that in order to reach general applicability, we need to tackle flexibility and entropy on a much broader scale. The receptor flexibility is, however, one of the biggest challenges, because conformational changes of proteins range from small rearrangements of side chains to large movements of whole domains and their incorporation is of utmost importance for various target classes, such as kinases. This is particularly important in virtual screening, where we must not lose active compounds via docking procedure due to the steric clashes with rigid protein target. Some part of flexibility can be introduced already in docking by an induced fit procedure which allows a limited movement of some active site residues.[75]

Molecular Dynamics (MD) simulations can provide an adequate sampling of binding protein-ligand conformations, allowing us to compute free energy and kinetics of binding, e.g. via dynamic docking procedure.[76] Despite the simulations are still very demanding, the enormous increase of computer power and faster software implementations advances over the last decade have prompted their use in the virtual screening pipeline. It should be kept in mind that potential/free energy surfaces of P-L complexes are extremely complicated and contain huge number of local minima, which are separated from each other only by small barriers. Thus, using higher number of P-L conformations as starting geometries makes it possible to evaluate many microstates with shorter MD and less demanding optimization calculations. Our preliminary results show dramatic improvements of SQM scores after enabling the optimization of the P-L complex where the whole ligand and a part of the protein was considered flexible. It should be noted that the conformational response of proteins could range from motion of residues surrounding the binding pocket to movement of loops and domains, which can cause a significant change in the tertiary structure of protein. For these reasons, we are not only trying to investigate scale and definition of the flexibility to the maximum extent possible but also achieve the desired higher accuracy at minimal computational cost.

5. Conclusion

Due to the enormous advances in computational power, QM-based methods have been incorporated into the process of drug design in all its preclinical stages. In principle, they can provide high accuracy and better prediction of binding affinities of ligands and their biological targets and result in rational design of more potent drugs saving huge expenses and time.

Over the last decade, we have developed fast and reliable protocols for estimation of total binding free energies of P-L complexes. The first more accurate one dubbed **SQM-based SF** can be used for estimation of binding free energies of preselected smaller set of proteins and ligands where the P-L complex conformations are obtained from computationally costly full gradient optimization. This concept is ideal for the hit-to-lead optimization stage of drug design. The second, more efficient methodology called **SQM/COSMO SF** approximates the total score by the two leading terms only, the P-L interaction energy in vacuum and solvation/desolvation free energy. This approach is ideal for fast estimates of binding free energies of large sets of ligands and proteins for which the structures are generated by docking. We have shown the superior performance of **SQM/COSMO SF** in sampling, ranking and library enrichment setups in several proof-of-concept studies, and we are working towards a fully automatized SQM/COSMO protocol applicable within the pharmaceutical pipeline.

Acknowledgements

This work was part of the Research Project RVO: 61388963 of the Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences. We acknowledge the support from the European Regional Development Fund; OP RDE; Project: 'Chemical Biology for Drugging Undruggable Targets (ChemBioDrug)' (No. CZ.02.1.01/ 0.0/0.0/ 16_019/ 0000729). This work was also supported by the Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project 'IT4Innovations National Supercomputing Center – LM2015070'.

Keywords:

sampling power • native pose prediction • enrichment • virtual screening • ranking power • quantum chemical semiempirical methods • implicit solvation • protein-ligand binding • affinity • *in silico* drug design

References

[1] a) G. Klebe, *Drug Discov Today* **2006**, *11*, 580-594; b) G. Schneider, *Nature Reviews Drug Discovery* **2010**, *9*, 273-276.

[2] D. C. Young, *Computational Drug Design: A Guide for Computational and Medicinal Chemists*, John Wiley & Sons, Inc., **2009**.

[3] J. J. Irwin, B. K. Shoichet, *J Med Chem* **2016**, *59*, 4103-4120.

[4] a) E. Yuriev, J. Holien, P. A. Ramsland, *J Mol Recognit* **2015**, *28*, 581-604; b) M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, *J Med Chem* **2016**, *59*, 4035-4061.

[5] a) S. Decherchi, M. Masetti, I. Vyalov, W. Rocchia, *Eur J Med Chem* **2015**, *91*, 27-42; b) M. Kolar, J. Fanfrlik, M. Lepsik, F. Forti, F. J. Luque, P. Hobza, *J Phys Chem B* **2013**, *117*, 5950-5962.

[6] a) R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz, *J Med Chem* **2006**, *49*, 6177-6196; b) B. Vorlova, D. Nachtigallova, J. Jiraskova-Vanickova, H. Ajani, P. Jansa, J. Rezac, J. Fanfrlik, M. Otyepka, P. Hobza, J. Konvalinka, M. Lepsik, *Eur J Med Chem* **2015**, *89*, 189-197; c) J. M. Fox, K. Kang, M. Sastry, W. Sherman, B. Sankaran, P. H. Zwart, G. M. Whitesides, *Angew Chem Int Edit* **2017**, *56*, 3833-3837; d) M. Hylsova, B. Carbain, J. Fanfrlik, L. Musilova, S. Haldar, C. Kopruluoglu, H. Ajani, P. S. Brahmkshatriya, R. Jorda, V. Krystof, P. Hobza, A. Echalier, K. Paruch, M. Lepsik, *Eur J Med Chem* **2017**, *126*, 1118-1128; e) T. Beuming, Y. Che, R. Abel, B. Kim, V. Shanmugasundaram, W. Sherman, *Proteins* **2012**, *80*, 871-883.

[7] a) A. Ciancetta, S. Genheden, U. Ryde, *J Comput Aid Mol Des* **2011**, *25*, 729-742; b) J. Fanfrlik, P. S. Brahmkshatriya, J. Rezac, A. Jilkova, M. Horn, M. Mares, P. Hobza, M. Lepsik, *J Phys Chem B* **2013**, *117*, 14973-14982.

[8] a) U. Ryde, P. Soderhjelm, *Chem Rev* **2016**, *116*, 5520-5566; b) N. D. Yilmazer, M. Korth, *Int J Mol Sci* **2016**, *17*; c) D. Mucs, R. A. Bryce, *Expert Opinion on Drug Discovery* **2013**, *8*, 263-276; dA. Heifetz, *Quantum Mechanics in Drug Discovery*, Humana, New York, NY, New York, USA, **2020**.

[9] a) J. Rezac, P. Hobza, *Chem Rev* **2016**, *116*, 5038-5071; b) M. H. Kolar, P. Hobza, *Chem Rev* **2016**, *116*, 5155-5187; c) P. Politzer, J. S. Murray, T. Clark, G. Resnati, *Phys Chem Chem Phys* **2017**, *19*, 32166-32178.

[10] a) J. Fanfrlik, M. Kolar, M. Kamlar, D. Hurny, F. X. Ruiz, A. Cousido-Siah, A. Mitschler, J. Rezac, E. Munusamy, M. Lepsik, P. Matejicek, J. Vesely, A. Podjarny, P. Hobza, *Acs Chem Biol* **2013**, *8*, 2484-2492; b) J. Fanfrlik, F. X. Ruiz, A. Kadlcikova, J. Rezac, A. Cousido-Siah, A. Mitschler, S. Haldar, M. Lepsik, M. H. Kolar, P. Majer, A. D. Podjarny, P. Hobza, *Acs Chem Biol* **2015**, *10*, 1637-1642; c) K. Kriz, J. Fanfrlik, M. Lepsik, *Chemphyschem* **2018**, *19*, 2540-2548.

[11] a) M. J. Field, P. A. Bash, M. Karplus, *J Comput Chem* **1990**, *11*, 700-733; b) H. M. Senn, W. Thiel, *Angew Chem Int Edit* **2009**, *48*, 1198-1229; c) S. K. Burger, D. C. Thompson, P. W. Ayers, *J Chem Inf Model* **2011**, *51*, 93-101.

[12] a) J. Rezac, J. Fanfrlik, D. Salahub, P. Hobza, *J Chem Theory Comput* **2009**, *5*, 1749-1760; b) M. Korth, M. Pitonak, J. Rezac, P. Hobza, *J Chem Theory Comput* **2010**, *6*, 344-352; c) J. Rezac, P. Hobza, *J Chem Theory Comput* **2012**, *8*, 141-151; d) J. Rezac, P. Hobza, *Chem Phys Lett* **2011**, *506*, 286-289.

[13] M. Lepsik, J. Rezac, M. Kolar, A. Pecina, P. Hobza, J. Fanfrlik, *Chempluschem* **2013**, *78*, 921-931.

[14] a) A. Pecina, R. Meier, J. Fanfrlik, M. Lepsik, J. Rezac, P. Hobza, C. Baldauf, *Chem Commun* **2016**, *52*, 3312-3315; b) A. Pecina, S. Haldar, J. Fanfrlik, R. Meier, J. Rezac, M. Lepsik, P. Hobza, *J Chem Inf Model* **2017**, *57*, 127-132; c) A. Pecina, J. Brynda, L. Vrzal, R. Gnanasekaran, M. Horejsi, S. M. Eyrilmez, J. Rezac, M. Lepsik, P. Rezacova, P. Hobza, P. Majer, V. Veverka, J. Fanfrlik, *Chemphyschem* **2018**, *19*, 873-879; d) S. M. Eyrilmez, C. Kopruluoglu, J. Rezac, P. Hobza, *Chemphyschem* **2019**, *20*, 2759-2766.

[15] a) W. T. Michael J. S. Dewart, *J Am Chem Soc* **1977**, *99*, 4899-4907; b) M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J Am Chem Soc* **1985**, *107*, 3902-3909; c) J. J. P. Stewart, *J Comput Chem* **1991**, *12*, 320-341; d) J. J. P. Stewart, *J Mol Model* **2007**, *13*, 1173-1213; e) J. J. P. Stewart, *J Mol Model* **2013**, *19*, 1-32; f) W. Weber, W. Thiel, *Theor Chem Acc* **2000**, *103*, 495-506.

[16] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, *J Chem Phys* **2001**, *114*, 5149-5155.

[17] C. Bannwarth, S. Ehlert, S. Grimme, *J Chem Theory Comput* **2019**, *15*, 1652-1671.

[18] a) M. Gaus, Q. A. Cui, M. Elstner, *J Chem Theory Comput* **2011**, *7*, 931-948; b) M. Gaus, X. Y. Lu, M. Elstner, Q. Cui, *J Chem Theory Comput* **2014**, *10*, 1518-1537; c) M. Kubillus, T. Kubar, M. Gaus, J. Rezac, M. Elstner, *J Chem Theory Comput* **2015**, *11*, 332-342; d) X. Y. Lu, M. Gaus, M. Elstner, Q. Cui, *J Phys Chem B* **2015**, *119*, 1062-1082.

[19] J. J. P. Stewart, *Int J Quantum Chem* **1996**, *58*, 133-146.

[20] a) W. T. Yang, T. S. Lee, *J Chem Phys* **1995**, *103*, 5674-5678; b) Y. Nishimura, H. Nakai, *J Comput Chem* **2018**, *39*, 105-116.

[21] a) J. P. McNamara, I. H. Hillier, *Phys Chem Chem Phys* **2007**, *9*, 2362-2370; b) T. Tuttle, W. Thiel, *Phys Chem Chem Phys* **2008**, *10*, 2159-2166.

[22] a) P. Jurecka, J. Cerny, P. Hobza, D. R. Salahub, *J Comput Chem* **2007**, *28*, 555-569; b) J. Antony, S. Grimme, *Phys Chem Chem Phys* **2006**, *8*, 5287-5293.

[23] a) J. Rezac, *J Chem Theory Comput* **2017**, *13*, 4804-4817; b) V. M. Miriyala, J. Rezac, *J Comput Chem* **2017**, *38*, 688-697.

[24] J. Rezac, *J Comput Chem* **2019**, *40*, 1633-1642.

[25] J. Rezac, P. Hobza, *J Chem Theory Comput* **2013**, *9*, 2151-2155.

[26] J. Rezac, *J Comput Chem* **2016**, *37*, 1230-1237.

[27] J. J. P. Stewart, HTTP://OpenMOPAC.net., Stewart Computational Chemistry, Colorado Springs, CO, USA, **2016**.

[28] B. Aradi, B. Hourahine, T. Frauenheim, *J Phys Chem A* **2007**, *111*, 5678-5684.

[29] R. Kriz K., J., *J Chem Inf Model* **2020**, *accepted*, doi: 10.1021/acs.jcim.1029b01171.

[30] J. Hostas, J. Rezac, P. Hobza, *Chem Phys Lett* **2013**, *568*, 161-166.

[31] V. Tsui, D. A. Case, *Biopolymers* **2001**, *56*, 275-291.

[32] B. Honig, K. Sharp, A. S. Yang, *J Phys Chem-Us* **1993**, *97*, 1101-1109.

[33] a) C. Curutchet, M. Orozco, F. J. Luque, *J Comput Chem* **2001**, *22*, 1180-1193; b) I. Soteras, C. Curutchet, A. Bidon-Chanal, M. Orozco, F. J. Luque, *J Mol Struc-Theochem* **2005**, *727*, 29-40.

[34] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J Phys Chem B* **2009**, *113*, 6378-6396.

[35] A. Klamt, *Abstr Pap Am Chem S* **1995**, *210*, 122-COMP.

[36] A. Klamt, G. Schuurmann, *Journal of the Chemical Society-Perkin Transactions 2* **1993**, 799-805.

[37] K. Kriz, J. Rezac, *J Chem Inf Model* **2019**, *59*, 229-235.

[38] K. Raha, K. M. Merz, *J Am Chem Soc* **2004**, *126*, 1020-1021.

[39] K. Raha, K. M. Merz, *J Med Chem* **2005**, *48*, 4558-4575.

[40] a) H. S. Muddana, M. K. Gilson, *J Chem Theory Comput* **2012**, *8*, 2023-2033; b) P. Mikulskis, S. Genheden, K. Wichmann, U. Ryde, *J Comput Chem* **2012**, *33*, 1179-1189.

[41] L. Rao, I. Y. Zhang, W. P. Guo, L. Feng, E. Meggers, X. Xu, *J Comput Chem* **2013**, *34*, 1636-1646.

[42] a) M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko, *Chem Rev* **2012**, *112*, 632-672; b) J. F. Liu, X. W. Wang, J. Z. H. Zhang, X. He, *Rsc Adv* **2015**, *5*, 107020-107030.

[43] B. Thapa, D. Beckett, J. Erickson, K. Raghavachari, *J Chem Theory Comput* **2018**, *14*, 5143-5155.

[44] P. Chaskar, V. Zoete, U. F. Rohrig, *J Chem Inf Model* **2014**, *54*, 3137-3152.

[45] V. Zoete, T. Schuepbach, C. Bovigny, P. Chaskar, A. Daina, U. F. Rohrig, O. Michielin, *J Comput Chem* **2016**, *37*, 437-447.

[46] M. N. Ucisik, Z. Zheng, J. C. Faver, K. M. Merz, *J Chem Theory Comput* **2014**, *10*, 1314-1325.

[47] C. A. Ortiz-Mahecha, H. J. Bohorquez, W. A. Agudelo, M. A. Patarroyo, M. E. Patarroyo, C. F. Suarez, *J Chem Inf Model* **2019**, *59*, 5148-5160.

[48] a) L. D. Alvarez, A. S. Veleiro, G. Burton, *Proteins* **2015**, *83*, 1297-1306; b) C. N. Cavasotto, N. S. Adler, M. G. Aucar, *Front Chem* **2018**, *6*.

[49] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, *Accounts Chem Res* **2000**, *33*, 889-897.

[50] J. Fanfrlik, A. K. Bronowska, J. Rezac, O. Prenosil, J. Konvalinka, P. Hobza, *J Phys Chem B* **2010**, *114*, 12666-12678.

[51] a) M. Nekardova, L. Vymetalova, P. Khirsariya, S. Kovacova, M. Hylsova, R. Jorda, V. Krystof, J. Fanfrlik, P. Hobza, K. Paruch, *Chemphyschem* **2017**, *18*, 785-795; b) J. Snasel, P. Naus, J. Dostal, A. Hnizda, J. Fanfrlik, J. Brynda, A. Bourderioux, M. Dusek, H. Dvorakova, J. Stolarikova, H. Zabranska, R. Pohl, P. Konecny, P. Dzubak, I. Votruba, M. Hajduch, P. Rezacova, V. Veverka, M. Hocek, I. Pichova, *J Med Chem* **2014**, *57*, 8268-8279; c) C. Köprülüoğlu, M. Šála, H. Ajani, H. Hřebabecký, J. Fanfrlík, R. Jorda, M. Dračínský, E. Procházková, B. Klepetářová, P. Šácha, V. Kryštof, P. Hobza, M. Lepšík, R. Nencka, *J Mol Recognit* **2020**, accepted.

[52] a) H. M. Houstecka M., Fanfrlik J., Brynda J., Pallova L., Hanova I., Mertlikova-Kaiserova H., Lepsik M., Horn M, Smrčina M., Majer P., Mares M., *J Med Chem* **2020**, *just accepted*; b) J. Dostal, A. Pecina, O. Hruskova-Heidingsfeldova, L. Mareckova, I. Pichova, P. Rezacova, M. Lepsik, J. Brynda, *Acta Crystallogr D* **2015**, *71*, 2494-2504.

[53] A. Cousido-Siah, F. X. Ruiz, J. Fanfrlik, J. Gimenez-Dejoz, A. Mitschler, M. Kamlar, J. Vesely, H. Ajani, X. Pares, J. Farres, P. Hobza, A. D. Podjarny, *Acs Chem Biol* **2016**, *11*, 2693-2705.

[54] H. Cahova, A. Panattoni, P. Kielkowski, J. Fanfrlik, M. Hocek, *Acs Chem Biol* **2016**, *11*, 3165-3171.

[55] L. Zakova, E. Kletvikova, M. Lepsik, M. Collinsova, C. J. Watson, J. P. Turkenburg, J. Jiracek, A. M. Brzozowski, *Acta Crystallogr D* **2014**, *70*, 2765-2774.

[56] J. Gemperle, R. Hexnerova, M. Lepsik, P. Tesina, M. Dibus, M. Novotny, J. Brabek, V. Veverka, D. Rosel, *Sci Rep-Uk* **2017**, *7*.

[57] A. Pecina, M. Lepsik, J. Rezac, J. Brynda, P. Mader, P. Rezacova, P. Hobza, J. Fanfrlik, *J Phys Chem B* **2013**, *117*, 16096-16104.

[58] P. Mader, A. Pecina, P. Cigler, M. Lepsik, V. Sicha, P. Hobza, B. Gruner, J. Fanfrlik, J. Brynda, P. Rezacova, *Biomed Res Int* **2014**.

[59] D. E. Honda, J. B. L. Martins, M. M. Ventura, S. M. Eyrilmez, M. Lepsik, P. Hobza, A. Pecina, S. M. de Freitas, *Eur J Org Chem* **2018**, *2018*, 5203-5211.

[60] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang, P. Kollman, *J Comput Chem* **2003**, *24*, 1999-2012.

[61] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J Comput Chem* **2004**, *25*, 1157-1174.

[62] O. Korb, T. Stutzle, T. E. Exner, *J Chem Inf Model* **2009**, *49*, 84-96.

[63] O. Trott, A. J. Olson, *J Comput Chem* **2010**, *31*, 455-461.

[64] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, R. P. Mee, *J Comput Aid Mol Des* **1997**, *11*, 425-445.

[65] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J Mol Biol* **1997**, *267*, 727-748.

[66] W. T. M. Mooij, M. L. Verdonk, *Proteins* **2005**, *61*, 272-287.

[67] H. Ajani, A. Pecina, S. M. Eyrilmez, J. Fanfrlik, S. Haldar, J. Rezac, P. Hobza, M. Lepsik, *Acs Omega* **2017**, *2*, 4022-4029.

[68] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J Comput Chem* **2009**, *30*, 2785-2791.

[69] W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz, R. C. Rizzo, *J Comput Chem* **2015**, *36*, 1132-1156.

[70] D. A. Pearlman, P. S. Charifson, *J Med Chem* **2001**, *44*, 3417-3423.

[71] M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, *J Med Chem* **2012**, *55*, 6582-6594.

[72] D. R. Koes, M. P. Baumgartner, C. J. Camacho, *J Chem Inf Model* **2013**, *53*, 1893-1904.

[73] S. M. Eyrilmez, C. Köprülüoğlu, J. Rezac, P. Hobza, *Chemphyschem* **2019**, *20*, 2721-2721.

[74] A. Nicholls, *J Comput Aid Mol Des* **2008**, *22*, 239-255.

[75] a) W. Sherman, H. S. Beard, R. Farid, *Chem Biol Drug Des* **2006**, *67*, 83-84; b) W. Sherman, T. Day, M. P. Jacobson, R. A. Friesner, R. Farid, *J Med Chem* **2006**, *49*, 534-553.

[76] M. De Vivo, A. Cavalli, *Wires Comput Mol Sci* **2017**, *7*.

ACCEPTED MANUSCRIPT

Entry for the Table of Contents

SQM/COSMO
ENRICHMENT
SAMPLING
RANKING
SCORING

The minireview covers a challenging journey of our semiempirical QM-based scoring function (SQM/COSMO SF) toward its successful application in structure-based drug design. On the way, the SQM/COSMO SF has acquired crucial abilities, i.e. the sampling, ranking and enrichment power. It has proven its superiority over conventional SFs in several proof-of-concept studies and is thus ready for new challenges, e.g. within the pharmaceutical pipeline.

Institute and/or researcher Twitter usernames: @IOCBPrague and @madAniceP, @salezyum, @Cemal_Kp, @Lepsik_science

WILEY-VCH