

DOI: 10.1002/cmdc.201000085

High-Throughput Virtual Screening Using Quantum Mechanical Probes: Discovery of Selective Kinase Inhibitors

Ting Zhou and Amedeo Caflisch*^[a]

A procedure based on semi-empirical quantum mechanical (QM) calculations of interaction energy is proposed for the rapid screening of compound poses generated by high-throughput docking. Small molecules (consisting of 2–10 atoms and termed “probes”) are overlapped with polar groups in the binding site of the protein target. The interaction energy values between each compound pose and the probes, calculated by a semi-empirical Hamiltonian, are used as filters. The QM probe method does not require fixed partial charges and takes into account polarization and charge-transfer effects which are not captured by conventional force fields. The procedure is applied to screen ~100 million poses (of 2.7 million commercially available compounds) obtained by high-through-

put docking in the ATP binding site of the tyrosine kinase erythropoietin-producing human hepatocellular carcinoma receptor B4 (EphB4). Three QM probes on the hinge region and one at the entrance pocket are employed to select for binding affinity, while a QM probe on the side chain of the so-called gatekeeper residue (a hypervariable residue in the kinase) is used to enforce selectivity. The poses with favorable interactions with the five QM probes are filtered further for hydrophobic matching and low ligand strain. In this way, a single-digit micromolar inhibitor of EphB4 with a relatively good selectivity profile is identified in a multimillion-compound library upon experimental tests of only 23 molecules.

Introduction

Fast and accurate methods for computing the binding free energy between small molecules and proteins are required for computer-aided drug design (CADD).^[1–6] The increasing popularity of quantum mechanical (QM) methods in CADD is not just a consequence of ever-growing computing power, but is also due to the first-principle nature of QM, which should provide the highest accuracy.^[7–10] However, the computation time required for QM ranges from N^3 (semi-empirical) to N^5 (second-order Møller–Plesset perturbation theory and other post-Hartree–Fock methods), for which N is the number of basis functions.^[11] Therefore, QM is used for molecular systems of limited size, such as in the hybrid quantum mechanics/molecular mechanics (QM/MM) approach,^[12,13] or for a small subset of atoms, while a polarizable continuum model is employed to describe the protein and the solvent.^[14] In linear scaling QM methods, the computing time scales with N^2 or even N if the local character of chemical interactions is fully exploited.^[7,15–19] Using linear scaling theory, Stewart and Anikin et al. applied the software package MOZYME^[20] and LocalSCF,^[21] respectively, to calculate QM energies with localized molecular orbital (LMO) theory. Recently, we developed the quantum mechanical linear interaction energy model with continuum electrostatic solvation (QMLIECE), in which the electrostatic contribution to the binding energy is evaluated by a semi-empirical QM divide-and-conquer strategy. QMLIECE is useful for highly variable charge–charge interactions, as in the case of 44 peptidic inhibitors of a flaviviral nonstructural 3 serine protease.^[22] Nevertheless, neither LMO theory nor the QMLIECE approach are fast enough for evaluating multiple poses of small molecules gen-

erated by high-throughput virtual screening (HTVS). Therefore, Vasilyev and Bliznyuk first used fast-scoring functions for ranking, and then applied the LMO theory, as implemented in MOZYME, to only the 10–100 top-ranking poses.^[23] We have applied QMLIECE to < 10 000 poses at most.^[22] No application of QM methods to millions of poses in HTVS has been reported yet.

Herein, a procedure based on semi-empirical QM is developed for the *in silico* screening of millions of poses generated by high-throughput docking of large libraries of compounds. The interaction energies (IEs) between small molecules and individual polar groups (probes) in the binding pocket of the protein target are used to filter out poses that are not likely to bind. The method focuses on exploiting advantages of QM in HTVS, such as independence of force field parameters and the ability to capture charge transfer, polarization, and direction-specific effects, which are very important for describing hydrogen bonds (HBs).^[24] Flexibility of the functional groups of some

[a] T. Zhou, Prof. A. Caflisch
Department of Biochemistry, University of Zurich
Winterthurerstrasse 190, 8057 Zurich (Switzerland)
Fax: (+41) 44-635-6862
E-mail: caflisch@bioc.uzh.ch

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cmdc.201000085>: assessment of the QM probe method on CDK2, probe energies of minor conformers of four known EphB4 inhibitors, energy distributions of Probes 1–4 and 10, IE against distances between a water molecule and N-methylacetamide calculated by PM6 Hamiltonian, distribution of shape Tanimoto of 15 979 poses, and a list of kinases for screening.

of the side chains is taken into account by partial optimization of the structure of the complex. As a proof of principle, the QM probe approach is applied to the receptor tyrosine kinase erythropoietin-producing human hepatocellular carcinoma receptor B4 (EphB4), which is involved in cancer-related angiogenesis.^[25–27] Docking is performed at the ATP binding site, and five QM probes are used for filtering: three probes represent the backbone polar groups of the hinge region,^[28] one probe is located at the entrance pocket, and a fifth probe is selected at the gatekeeper side chain^[28–32] (Thr693 in EphB4) to bias the in silico screening toward selective inhibitors of EphB4.

Methods

Design of QM probes

The QM probes are small molecular fragments (2–10 atoms) used for the efficient evaluation of the IE between polar groups of the protein and the ligand. The QM probes are “designed” 1) to reflect the local electronic structures in the binding pocket, 2) to be as simple as possible for computational efficiency, and 3) to distinguish sensitively between favorable and adverse contacts. Methanol, acetate anion, methylammonium cation, and guanidinium cation probes are essentially identical to the corresponding functional groups in the side chains of Ser/Thr/Tyr, Asp/Glu, Lys, and Arg, respectively (Table 1).

Table 1. QM probes for polar groups in proteins. ^[a]			
Functional Group	Location in Proteins	Probe Molecule	Overlapping Position
	backbone, Asn, Gln	H ₂ O	
	backbone, Asn, Gln	HF	
	Ser, Thr, Tyr	CH ₃ OH	–
	Asp, Glu	CH ₃ CO ²⁻	–
	Lys	CH ₃ NH ₃ ⁺	–
	Arg	C(NH ₂) ₃ ⁺	–

[a] Atoms in boldface are flexible during QM minimization of complex formation enthalpy, which is carried out with rigid ligand.

A water molecule is used as probe for the carbonyl group in the backbone as well as in the Asn and Gln side chains. Furthermore, the plane of the water molecule is perpendicular to that of the carbonyl group to have the same orientation of the lone pairs (Table 1).^[33,34] Water is more appropriate than acetone or acetamide because the additional methyl or amino group, respectively, may be involved in van der Waals (vdW) interactions with the ligand, and thus attenuate the energetic difference due to the formation of HB. Formaldehyde is not as sensitive as the water molecule in detecting the formation of HB because it has one carbon atom more than water, and the carbon atom may also form vdW interactions with the ligand.

The hydrogen fluoride (HF) probe is used to emulate the amide group in the backbone of the protein as well as in the Asn and Gln side chains for two reasons: First, HF is the strongest neutral HB donor, and thus is the most sensitive as a probe molecule to detect HB acceptors. Second, HF is a two-atom molecule, and has little additional vdW interactions with ligands.

The QM probe method was first assessed on cyclin-dependent kinase 2 (CDK2; see the Supporting Information) and then applied to EphB4. The application to CDK2 shows that the method is able to identify classical HBs as well as favorable polar interactions such as that between aromatic CH and carbonyl oxygen.^[35]

Calculation of interaction energy

Multiple poses of the ligands are determined by automatic docking and force field minimization (see below). Ligands are always fixed during the evaluation of the interaction energy $IE = H_{\text{ligand-probe}} - H_{\text{ligand}} - H_{\text{probe}}$ for which H is the formation enthalpy calculated with MOPAC^[36] and the semi-empirical Hamiltonian PM6.^[37] It has been reported that density functionals can quantitatively reproduce the HB energy of CCSD(T) (Coupled-Cluster with Single and Double and perturbative Triple excitations),^[38–44] but considering the efficiency required for filtering several million poses, rapid PM6 was selected eventually.^[45] The IEs between rigid probes and the ligand are calculated directly, while optimization of $H_{\text{ligand-probe}}$ is carried out for flexible probes. In particular, the coordinates of the atoms in boldface in Table 1 are optimized to find an energy minimum of the ligand–probe complex. The IE evaluation of rigid probes is fast (<0.5 s), while ~30 s are required for the flexible probes. To improve efficiency, the interactions of the ligand with the rigid probes are calculated first and can be used as filters. This strategy was used in the application to EphB4 (see the *Polar interactions: QM probe energies* section below).

Preparation of EphB4

Because the structure of the kinase domain of EphB4 was not available when we started this work, a homology model was built using the structure of EphB2 (mouse, PDB ID: 1JPA) as template. The sequence identity between human EphB4 and mouse EphB2 is 88%. A detailed description of the homology modeling procedure has been published.^[46] Notably, the RMSD

between the homology model and the crystal structure of EphB4 (PDB ID: 2VWX) is only 0.28 Å for 179 of 242 C α atoms. Moreover, the orientation of the side chains in the ATP binding site is essentially identical in the model and the X-ray structure, the largest discrepancy (1.1 Å) being at the tip of the Met668 side chain in the hydrophobic pocket.

Preparation of the library and initial filtering

Of the 9.8 million compounds in the 2007 version of the ZINC library,^[47] ~2.7 million had at least one HB donor and one HB acceptor, and a molecular weight (M_r) < 500 g mol⁻¹. The molecules were assigned protonation states at pH 7 and were prepared in multiple protonation states and multiple tautomeric forms. The presence of HB donor(s) and acceptor(s) is essential, as the QM probe method focuses on the evaluation of HB patterns,^[35] while the filter on M_r was employed because small molecules are more appropriate as lead compounds. The molecular properties used for filtering were calculated by DAIM.^[48] The program WITNOTP (Armin Widmer, Novartis Pharma, Basel, Switzerland) was used for automatic assignment of CHARMM atom types and parameters,^[49] including partial charges, which were determined by an iterative approach based on the partial equalization of orbital electronegativity.^[50,51]

Docking into EphB4

Version 4 of AutoDock^[52] was used for flexible ligand docking of the ZINC subset of 2.7 million compounds using a rigid protein. First, the atom-specific affinity map files were generated by AutoGrid.^[53] The numbers of points in the x , y , and z directions were 62, 52, and 42, respectively, and the spacing between two adjacent grid points was 0.25 Å. AutoDock was then employed to generate multiple poses for further minimization by CHARMM.^[54] Because the AutoDock scoring function was not used for ranking, AutoDock energy evaluations were limited to 25 000 for each hybrid-genetic-local-search in order to speed up the docking procedure. The hybrid-genetic-local-search was run 400 times with different initial seeds to obtain multiple poses for each compound. (In preliminary docking runs of known EphB4 inhibitors, it was more likely to obtain the correct binding mode by using a large number of hybrid-genetic-local-searches than a large number of energy evaluations and only a few searches.) The docking was followed by energy minimization in the rigid protein using the CHARMM force field.^[49] Redundant poses were eliminated by clustering using an all-atom RMSD cutoff of 0.01 Å. For each pose, the IE values with the QM probes as well as electrostatics and vdW efficiencies were stored in a table of the data management system for distributed virtual screening (DVSDMS).^[55]

van der Waals filters

Upon energy minimization, loose vdW filters^[56] were applied to all poses to eliminate those with clashes and/or poor steric complementarity.^[7] The following cutoffs of the CHARMM intermolecular vdW energy were employed: $E_{\text{vdW}} < -20$ kcal mol⁻¹

and $E_{\text{vdW}}/M_r < -0.05$ kcal g⁻¹. Considering the efficiency and accuracy, the vdW filters are more appropriately calculated with force field methods than with QM.^[57]

Results and Discussion

Docking and van der Waals filters

A total of ~100 million poses of 2.7 million compounds were generated by flexible ligand docking into the rigid ATP binding site of EphB4. The vdW filters decreased the number of poses to ~90 million.

Polar interactions: QM probe energies

Thirteen polar groups were replaced by QM probes in the ATP binding site of EphB4 (Figure 1). According to the definitions of Traxler and Furet,^[28,58] Probes 1–3 lie in the adenine binding

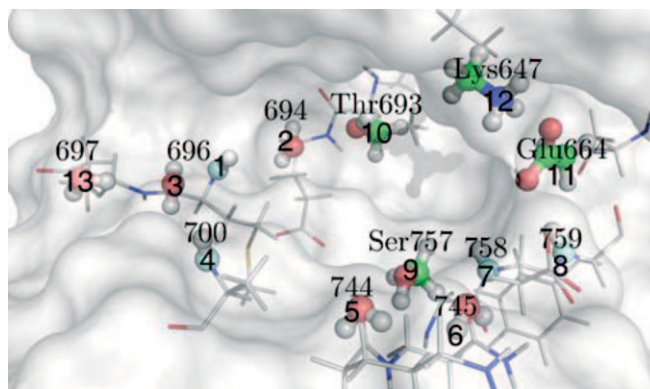


Figure 1. The 13 QM probes in the ATP binding site of the receptor tyrosine kinase EphB4. The probes are shown as spheres colored by atom type, with hydrogen in gray, carbon in green, nitrogen in blue, oxygen in red, and fluorine in cyan. The QM probes on side chains are labeled with the residue type and number, while those on the backbone are labeled only by the residue number.

region, Probe 4 is located at the entrance pocket, Probes 5 and 9 lie in the ribose binding pocket, Probe 6 is in the phosphate binding pocket, and Probes 10–12 lie in the hydrophobic pocket. The virtual screening focuses on binding at the hinge region; therefore Probes 1–3 were selected for filtering. Probe 4 was also taken into account for filtering, as it is very close to the hinge region. Moreover, the hydroxy group of the gatekeeper residue (Probe 10 at Thr693) was included to bias the search toward selective kinase inhibitors,^[28–32] as only 95 of the 507 human protein kinases have Thr as a gatekeeper.^[59] Whether to consider a specific probe depends on the requirement; for example, Probes 5 and 9 must be taken into account if the polar interactions in the ribose binding pocket need to be inspected. Therefore, the unused probes may be employed in future studies, such as combinatorial lead optimization or de novo design.^[60]

Table 2. QM probe energies and experimentally measured IC₅₀ values of four known EphB4 inhibitors.

Compd	Structure	M _r [g mol ⁻¹]	IC ₅₀ [μM]	Probe 1	E _{probe} [kcal mol ⁻¹]			
					Probe 2	Probe 3	Probe 4	Probe 10
ALTA2 ^[46]		353	1.4	-4.31	0.37	0.14	0.34	0.41
PP2 ^[69]		302	0.34	-3.71	-3.18	-1.24	0.12	-2.98
ONC102 ^[70]		365	0.10	-2.46	-2.43	-1.38	-0.22	-2.32
MIYA9f ^[71]		434	0.021	-3.04	-2.25	-1.16	-5.53	-3.47

Four known inhibitors of EphB4 were used to determine the cutoffs for filtering according to the QM probe energy. The four inhibitors were docked into the ATP binding site of EphB4 and minimized with the same protocols as used for the high-throughput docking. Table 2 shows structures and probe energies of these inhibitors. The probe energy values are robust upon minor shifts in the binding mode (see Supporting Information table S-I).

The QM probe energies of the four known EphB4 inhibitors and ~100 compounds (selected randomly as representatives of inactive compounds) were taken into account to determine the cutoffs for filtering. These are: $E_{\text{probe1}} < -2$ kcal mol⁻¹ (typical hydrogen bond energy of N-H...O)^[61] and E_{probe2} , E_{probe3} , E_{probe4} , and $E_{\text{probe10}} < 0.5$ kcal mol⁻¹. Furthermore, the filter based on the sum $E_{\text{probe1}} + E_{\text{probe2}} < -3.8$ kcal mol⁻¹ was employed to give more weight to the two most buried polar groups of the hinge region. Because the methanol Probe 10 at the gatekeeper residue is flexible, its interaction was calculated only if $E_{\text{probe1}} < 0$ kcal mol⁻¹. The filters based on QM probes decreased the number of poses from 90 million to 955 094 poses of 509 101 molecules. In other words, the QM probe filtering diminished the average number of poses per molecule from 37 to 2.

Apolar interactions: Hydrophobic matching

The QM probes do not account for hydrophobic surface matching upon binding, which is important to evaluate the interaction at the hydrophobic pocket for kinases with a small gatekeeper residue, such as EphB4.^[62] The atoms of the ligand were classified as polar or nonpolar according to their QM-calculated partial charges (semi-empirical PM6 Hamiltonian and Mulliken population analysis). By comparing the QM charges of typical polar and nonpolar atoms in small molecules, 0.22 electronic units was selected as the threshold, that is, those atoms with partial charge in the range from -0.22 to 0.22 electronic units were considered apolar, while the remaining atoms were considered polar. Although this assignment requires the choice of an arbitrary threshold value, it was adopted because of its efficiency and

simplicity. The hydrophobic matching was approximated by the vdW interactions between the residues within the hydrophobic pocket of EphB4 (Val629, Ala645, the nonpolar part of Lys647, Met668, Ile691, and Thr693) and the nonpolar atoms of the ligand. A hydrophobic matching of at least -5 kcal mol⁻¹ was used to significantly decrease the number of poses for further analysis (15 979 poses belonging to 13 823 molecules). Visual inspection of some of the discarded poses confirmed that they do not fill the hydrophobic pocket with significant apolar surface matching.

Ligand strain filter

To evaluate the strain of the ligand, a minimization was performed in the absence of the protein starting from the bound conformation. The program MOPAC with a semi-empirical Hamiltonian RM1^[63] was used for the minimization. The program ROCS^[64] was employed to overlap the conformation minimized in the absence of the protein with the pose used as the starting point of the minimization, and to calculate the shape Tanimoto. The latter is defined as $O_{AB}/(V_A + V_B - O_{AB})$, in which O_{AB} is the volume overlap between conformer A and conformer B, and V_A and V_B are the volumes of conformers A and B, respectively.^[65] A shape Tanimoto close to 1 implies that the two conformations are essentially identical. The distribution of the shape Tanimoto of the 15 979 poses is shown in Supporting Information figure S-VII. A threshold for shape Tani-

moto >0.9 was chosen to further decrease the number of poses for visual inspection. In total, 8461 poses belonging to 7536 molecules passed this filter. Finally, 23 compounds were selected for experimental validation upon visual inspection of the first 1000 poses sorted according to QM probe energies and CHARMM intermolecular energy (see Supporting Information table S-IV). Sortings were simply carried out according to the sum of CHARMM intermolecular nonbonding energy terms, and the five probe energies (Probes 1–4 and 10) separately. The top ~ 166 poses of each ranking were selected. The main criteria used to filter out poses during visual inspection were the involvement of highly flexible functional groups such as $-(\text{CH}_2)_n-\text{OH}$ ($n \geq 1$) in HBs with the hinge region, and the desolvation of polar groups in the ATP binding site that were not involved in intermolecular HBs. The former was required, as the QM probe filters do not take into account the loss in conformational entropy upon binding. Visual inspection, although subjective, is an inevitable procedure for discarding poses with unfavorable interactions and/or unlikely conformations.^[66] Filtered by the QM probe and vdW energies, the remaining poses are inspected very efficiently, as key polar and apolar interactions have already been verified.

Computational requirements

The docking approach requires ~ 10 min per compound on a single Opteron CPU 244 (2.4 GHz). The energy minimization requires ~ 10 min for an average of 50 poses for each compound. The CPU time for calculating an IE with a rigid QM probe is <0.5 s mainly for the program input/output process, while ~ 30 seconds are needed for calculating the IE with a flexible probe. Despite the large quantity of molecules and poses, distributing docking and minimization jobs to hundreds of CPUs in two Beowulf clusters, selecting specific molecules and poses, and applying filters was efficiently managed by the database management system DVSDMS.^[55]

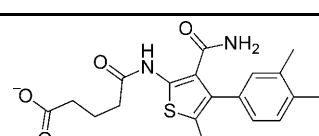
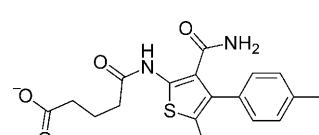
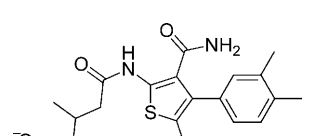
Experimental validation

The 23 compounds selected for validation were tested in two different enzymatic assays with the kinase catalytic domain of EphB4 in solution. One assay is based on fluorescence-resonance energy transfer (FRET) between coumarin and fluorescein (Omnia[®] Tyr recombinant kit KNZ4051, BioSourceTM^[67]), while the other measures the amount of radiolabeled phosphate transferred from $[\gamma\text{-}^{33}\text{P}]\text{ATP}$ to the substrate upon phosphorylation by EphB4.^[68] Three of the 23 compounds tested share a 2-formamido-4-phenylthiophene-

3-carboxamide scaffold. Two of these three thiophene derivatives have an inhibitor concentration for half-maximal activity (IC_{50}) $<10 \mu\text{M}$ in both assays, while the third has respective IC_{50} values of 9 and $17 \mu\text{M}$ (Table 3).

The predicted binding mode of compound 1 (Figure 2) indicates that its amide group N1-H , carbonyl group $\text{C}_2=\text{O}$, and

Table 3. Experimental validation of EphB4 inhibitors identified by high-throughput docking and the QM probe approach.

Compd	Structure	M_r [g mol^{-1}]	IC_{50} [μM] FRET ^[a] $[\gamma\text{-}^{33}\text{P}]\text{ATP}$ ^[b]
1		373	8.4, 7.6 2
2		360	7.1, 4.6 2
3		388	17.5, 16.7 9

[a] FRET-based enzymatic assay with the recombinant catalytic domain of human EphB4 in solution and an ATP concentration of $20 \mu\text{M}$. Each inhibitor was tested twice. To provide evidence against nonspecific effects such as aggregation, compound 2 was also tested upon addition of the detergent Triton X-100 (0.1% v/v). Similar values of percent inhibition (compound 2 at 30 and $10 \mu\text{M}$) were measured with and without detergent. [b] Enzymatic assay with the recombinant catalytic domain of human EphB4 and $[\gamma\text{-}^{33}\text{P}]\text{ATP}$ at $1 \mu\text{M}$ (performed at Reaction Biology Corp).

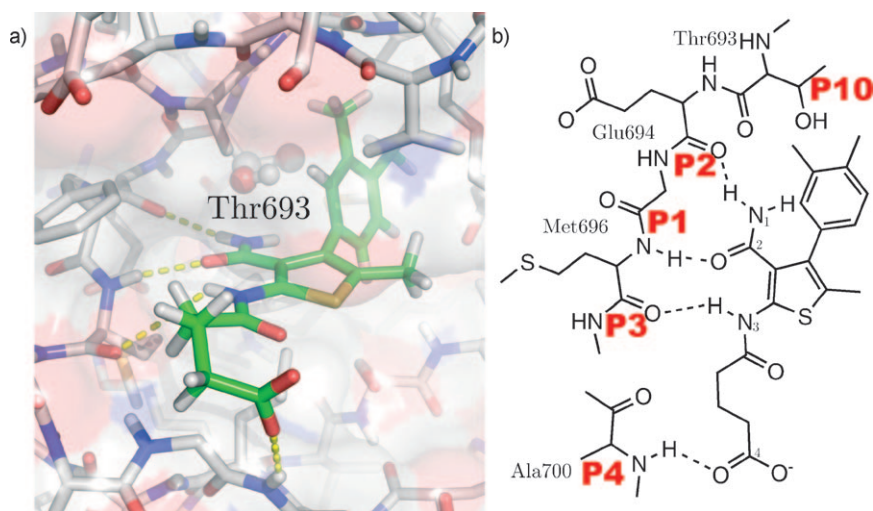
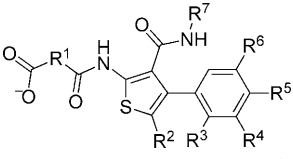



















Figure 2. Binding mode of compound 1 predicted by docking. a) The intermolecular HBs to the hinge loop and entrance loop are shown with yellow dashed lines. The pose was minimized in the rigid EphB4 structure (PDB ID: 2VWX). The atoms of the side chain of the gatekeeper residue Thr693 are represented as spheres. Nonpolar hydrogen atoms are omitted for clarity. b) The five QM probes used as filters are highlighted with red labels. Integer labels on compound 1 indicate functional groups mentioned in the text. The side chain of Phe695 is omitted for clarity.

Table 4. Structure and inhibitory activity of 23 commercially available derivatives of compound 1.

Compd	R ¹	R ²	R ³	R ⁴				IC ₅₀ [μM] ^[a]	Inhibition [%] at 20 μM compd ^[a]
					R ⁵	R ⁶	R ⁷		
4		CH ₃	H	H	F	H	H	–	49
5		CH ₃	H	H	H	H	H	–	50
6	CH ₂ CH ₂ CH ₂	CH ₃	H	H	C(CH ₃) ₃	H	H	–	38
7	CH ₂ CH ₂	H	H	CH ₃	CH ₃	H	H	–	46
8		H	CH ₃	H	H	CH ₃	H	–	50
9		H	Cl	H	Cl	H	H	9.7	–
10		CH ₃	H	H	CH ₃	H	H	13.7	–
11		CH ₃	Cl	H	Cl	H	H	7.8	–
12	CH ₂ CH ₂ CH ₂	H	H	H	CH ₃	H		–	7
13		CH ₃	H	H	CH ₃	H	H	–	40
14		H	H	CH ₃	CH ₃	H	H	–	44
15		CH ₃	H	CH ₃	CH ₃	H	H	–	31
16		H	H	CH ₃	CH ₃	H	H	–	21
17		CH ₃	H	CH ₃	CH ₃	H	H	–	16
18	CH ₂ CH ₂ CH ₂	H	H	CH ₃	CH ₃	H	H	–	55
19	CH ₂ CH ₂ CH ₂	CH ₃	H	H	CH ₂ CH ₃	H	H	16.3	–
20		CH ₃	H	CH ₃	CH ₃	H	H	–	27
21		H	H	H	CH ₃	H	H	–	15
22	CH ₂ CH ₂ CH ₂	CH ₃	Cl	H	Cl	H	H	–	47
23	CH ₂ CH ₂ CH ₂	H	Cl	H	Cl	H	H	–	54
24		CH ₃	H	H	CH ₃	H	H	5.2	–
25		CH ₃	H	CH ₃	CH ₃	H	H	–	2
26		H	H	H	Cl	H	H	–	20

[a] FRET-based enzymatic assay with the recombinant catalytic domain of human EphB4 in solution and ATP at 20 μM; each inhibitor was tested twice, and average values are reported.

amide group N3–H are involved in HBs with the backbone polar groups in the hinge region. Moreover, its carboxyl group is involved in a HB with the entrance pocket, and the dimethylphenyl ring is buried in the hydrophobic pocket. To validate the predicted binding mode of compound **1**, a set of 23 commercially available derivatives with the same 2-formamido-4-phenylthiophene-3-carboxamide scaffold were purchased and tested (Table 4). The structure–activity relationship (SAR) data are consistent with the binding mode. In particular, the cyclopropyl group at R⁷ causes a major loss in activity (compound **12**), in agreement with the hinge region HB of the amide group N1–H. Moreover, the similar inhibitory activity of compounds **22** and **23** (~50% at 20 μM), which differ only by a methyl group at R², is consistent with the orientation of the R² substituent toward the solvent (Figure 2).

The selectivity profile of compound **1** was tested against a panel of 85 protein kinases (National Centre for Protein Kinase Profiling at the University of Dundee; see Supporting Information table S-II). At 10 μM, the activity of Aurora B remained 37% relative to a DMSO control, while six other kinases retained 40–60% activity (Table 5). Note that three of these seven kinases have Thr as gatekeeper residue. Importantly, the inhibitory activity of compound **1** on the remaining 78 kinases is either zero or extremely modest. The binding mode obtained by docking into EphB4 suggests that the phenyl ring of compound **1** interacts favorably with the hydroxy group of the gatekeeper Thr693 ($E_{\text{probe10}} = -1.6 \text{ kcal mol}^{-1}$). These results indicate that compound **1** is rather selective, which is partially due to the use of the QM Probe 10 at the gatekeeper side chain (only ~20% of human kinases have Thr as gatekeeper).^[59] Interestingly, the known inhibitors PP2,^[69] ONC102,^[70] and MIYA9f^[71] have very favorable interaction energy with QM Probe 10 (Table 2), which is consistent with their good selectivity for protein kinases with Thr as the gatekeeper.

Table 5. Selectivity profile of compound **1** tested on a panel of 85 protein kinases.

Kinase	Activity [%] ^[a]	Gatekeeper ^[b]
Aurora B	37	Leu
EphA2	41	Thr
VEGFR	42	Thr
p38B MAPK	51	Thr
FGFR1	60	Val
LKB1	60	Met
CK1δ	60	Met
5 kinases	60 < % activity ≤ 70	–
13 kinases	70 < % activity ≤ 80	–
10 kinases	80 < % activity ≤ 90	–
50 other kinases	> 90	–

[a] Percent activity remaining with compound **1** present at 10 μM; values are relative to a control with 100% DMSO. Measurements were carried out at the National Centre for Protein Kinase Profiling at the University of Dundee. [b] The 78 un-nominated kinases have the following residue as gatekeeper with occurrence in parentheses: Met (35), Leu (16), Phe (12), Thr (10), Gln (2), Glu (1), Ile (1), and Tyr (1). The sequence information and gatekeeper residues of the whole panel of 85 kinases are given in the Supporting Information (table S-III); the ATP concentration in the assay is also shown in the Supporting Information (table S-II).

Conclusions

A new procedure for filtering millions of poses of small molecules based on QM calculations is described by an application to the tyrosine kinase EphB4. Polar groups in the protein binding site are substituted by small probes consisting of 2–10 atoms such as water and methanol for the carbonyl and hydroxy groups, respectively, and the interaction energy values between each pose and individual QM probes are calculated using a semi-empirical Hamiltonian. The use of a first-principle method is an advantage over classical force fields with fixed partial charges. As an example, the QM probes are able to detect favorable polar interactions such as the aromatic –CH...O=C interaction, which can be rather strong depending on the electronegativity of substituents on the aromatic ring.

The QM probe filtering is applied to ~100 million poses of small molecules generated by automatic docking into the ATP binding site of the kinase catalytic domain of EphB4. Only 1% of the poses pass the filters of favorable interaction energy with five QM probes: three in the hinge region, one at the entrance pocket, and one at the gatekeeper side chain (Thr693 in EphB4). The latter is a non-conserved residue in the kinome, and is therefore used to bias the virtual screening toward selective inhibitors. Upon further filtering based on nonpolar interactions, ligand strain, and visual inspection, 23 compounds are selected and tested in an enzymatic assay. Importantly, the QM probe filters as well as the additional filters used for post-processing require rather arbitrary threshold values, which might seem to contradict the use of first-principle methods. The QM probe method is an approximation of the real binding free energy because it takes into account only part of the protein target. Moreover, entropic terms are neglected.

Of the 23 compounds tested, three molecules with a 2-formamido-4-phenylthiophene-3-carboxamide scaffold are active in the low micromolar range in two different enzymatic assays. Additional evidence for the binding mode of compound **1**, and in particular its favorable interactions with the protein functional groups approximated by the QM probes, is provided by the SAR of 25 commercially available derivatives. Enzymatic assays on a panel of 85 protein kinases indicate that compound **1** is not promiscuous, as no inhibitory activity is observed for most of these kinases and modest inhibitory activity for only seven of them (three of which have Thr as a gatekeeper residue). Thus, compound **1** has potential for further development into a lead candidate, because of its low micromolar inhibitory activity for EphB4, low *M_r* (373 g mol⁻¹), and good selectivity profile.

Acknowledgements

We thank Dr. Danzhi Huang for performing some of the enzymatic assays and for interesting discussions and comments regarding the manuscript. We thank Dr. Philipp Schütz and Christian Bolliger for maintaining the Beowulf Linux clusters Etna and Schrödinger, respectively, which were used for most of the computations. We are grateful to Armin Widmer for providing the

modeling program WITNOTP, which was used for generating topology files and assigning CHARMM atom types. This work was supported by grants from the Swiss National Science Foundation to A.C.

Keywords: gatekeepers • high-throughput screening • inhibitors • kinases • quantum mechanics

- [1] J. Apostolakis, A. Caflisch, *Comb. Chem. High Throughput Screening* **1999**, *2*, 91–104.
- [2] R. C. Glen, S. C. Allen, *Curr. Med. Chem.* **2003**, *10*, 763–777.
- [3] W. P. Walters, M. Namchuk, *Nat. Rev. Drug Discovery* **2003**, *2*, 259–266.
- [4] D. Huang, A. Caflisch, *J. Med. Chem.* **2004**, *47*, 5791–5797.
- [5] W. Jorgensen, *Science* **2004**, *303*, 1813–1818.
- [6] M. Wang, C. F. Wong, *J. Chem. Phys.* **2007**, *126*, 026101.
- [7] K. Raha, K. M. Merz, *J. Med. Chem.* **2005**, *48*, 4558–4575.
- [8] A. Cavalli, P. Carloni, M. Recanatini, *Chem. Rev.* **2006**, *106*, 3497–3519.
- [9] M. B. Peters, K. Raha, K. M. Merz, *Curr. Opin. Drug Discov. Devel.* **2006**, *9*, 370–379.
- [10] T. Zhou, D. Huang, A. Caflisch, *Curr. Top. Med. Chem.* **2010**, *10*, 33–45.
- [11] A. Van der Vaart, V. Gogonea, S. Dixon, K. Merz, *J. Comput. Chem.* **2000**, *21*, 1494–1504.
- [12] H. M. Senn, W. Thiel, *Angew. Chem.* **2009**, *121*, 1220–1254; *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
- [13] P. Fong, J. P. McNamara, I. H. Hillier, R. A. Bryce, *J. Chem. Inf. Model.* **2009**, *49*, 913–924.
- [14] V. Buback, M. Mladenovic, B. Engels, T. Schirmeister, *J. Phys. Chem. B* **2009**, *113*, 5282–5289.
- [15] S. Gadre, R. Shirsat, A. Limaye, *J. Phys. Chem.* **1994**, *98*, 9165–9169.
- [16] S. Dixon, K. Merz, *J. Chem. Phys.* **1996**, *104*, 6643–6649.
- [17] T. Lee, J. Lewis, W. Yang, *Comput. Mater. Sci.* **1998**, *12*, 259–277.
- [18] Y. R. Mo, J. L. Gao, S. D. Peyerimhoff, *J. Chem. Phys.* **2000**, *112*, 5530–5538.
- [19] D. W. Zhang, Y. Xiang, A. M. Gao, J. Z. Zhang, *J. Chem. Phys.* **2004**, *120*, 1145–1148.
- [20] J. J. P. Stewart, *Int. J. Quantum Chem.* **1996**, *58*, 133–146.
- [21] N. A. Anikin, V. M. Anisimov, V. L. Bugaenko, V. V. Bobrikov, A. M. Andreyev, *J. Chem. Phys.* **2004**, *121*, 1266–1270.
- [22] T. Zhou, D. Huang, A. Caflisch, *J. Med. Chem.* **2008**, *51*, 4280–4288.
- [23] V. Vasilyev, A. Bliznyuk, *Theor. Chem. Acc.* **2004**, *112*, 23.
- [24] R. A. Friesner, *Adv. Protein Chem.* **2005**, *58*, 79–104.
- [25] R. H. Adams, *Semin. Cell Dev. Biol.* **2002**, *13*, 55–60.
- [26] G. Martiny-Baron, T. Korff, F. Schaffner, N. Esser, S. Eggstein, D. Marmé, H. G. Augustin, *Neoplasia* **2004**, *6*, 248–257.
- [27] N. Kertesz, V. Krasnoperov, R. Reddy, L. Leshanski, S. R. Kumar, S. Zozulya, P. S. Gill, *Blood* **2006**, *107*, 2330–2338.
- [28] P. Traxler, P. Furet, *Pharmacol. Ther.* **1999**, *82*, 195–206.
- [29] M. E. M. Noble, J. A. Endicott, L. N. Johnson, *Science* **2004**, *303*, 1800–1805.
- [30] P. J. Alaimo, Z. A. Knight, K. M. Shokat, *Bioorg. Med. Chem.* **2005**, *13*, 2825–2836.
- [31] J. J. Liao, *J. Med. Chem.* **2007**, *50*, 409–424.
- [32] M. Azam, M. A. Seeliger, N. S. Gray, J. Kuriyan, G. Q. Daley, *Nat. Struct. Mol. Biol.* **2008**, *15*, 1109–1118.
- [33] J. A. Odutola, T. R. Dyke, *J. Chem. Phys.* **1980**, *72*, 5062–5070.
- [34] P. Bobadova-Parvanova, B. Galabov, *J. Phys. Chem. A* **1998**, *102*, 1815–1819.
- [35] S. Sarkhel, G. R. Desiraju, *Proteins Struct. Funct. Bioinf.* **2004**, *54*, 247–259.
- [36] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209–220.
- [37] J. J. P. Stewart, *J. Mol. Model.* **2007**, *13*, 1173–1213.
- [38] H. G. Korth, M. I. de Heer, P. Mulder, *J. Phys. Chem. A* **2002**, *106*, 8779–8789.
- [39] N. Turki, A. Milet, O. Ouamerali, R. Moszynski, E. Kochanski, *THEOCHEM* **2002**, *577*, 239–253.
- [40] R. A. Klein, *J. Comput. Chem.* **2003**, *24*, 1120–1131.
- [41] J. Ireta, J. Neugebauer, M. Scheffler, *J. Phys. Chem. A* **2004**, *108*, 5692–5698.
- [42] Y. Zhao, O. Tishchenko, D. G. Truhlar, *J. Phys. Chem. B* **2005**, *109*, 19046–19051.
- [43] Y. X. Wang, B. Paulus, *Chem. Phys. Lett.* **2007**, *441*, 187–193.
- [44] H. R. Leverentz, D. G. Truhlar, *J. Phys. Chem. A* **2008**, *112*, 6009–6016.
- [45] *Comparison of PM6 and X-ray Structures of Hydrogen-Bonded Systems*, <http://openmopac.net/Hydrogen%20bonds.html> (accessed May 12, 2010).
- [46] P. Kolb, C. B. Kipouros, D. Huang, A. Caflisch, *Proteins Struct. Funct. Bioinf.* **2008**, *73*, 11–18.
- [47] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
- [48] P. Kolb, A. Caflisch, *J. Med. Chem.* **2006**, *49*, 7384–7392.
- [49] F. Momany, R. Rone, *J. Comput. Chem.* **1992**, *13*, 888–900.
- [50] K. No, J. Grant, H. Scheraga, *J. Phys. Chem.* **1990**, *94*, 4732–4739.
- [51] K. No, J. Grant, M. Jhon, H. Scheraga, *J. Phys. Chem.* **1990**, *94*, 4740–4746.
- [52] D. S. Goodsell, A. J. Olson, *Proteins* **1990**, *8*, 195–202.
- [53] *AutoGrid*, <http://autodock.scripps.edu/wiki/AutoGrid> (accessed May 12, 2010).
- [54] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- [55] T. Zhou, A. Caflisch, *J. Chem. Inf. Model.* **2009**, *49*, 145–152.
- [56] P. Kolb, D. Huang, F. Dey, A. Caflisch, *J. Med. Chem.* **2008**, *51*, 1179–1188.
- [57] T. Giese, D. York, *Int. J. Quantum Chem.* **2004**, *98*, 388–408.
- [58] M. Cherry, D. H. Williams, *Curr. Med. Chem.* **2004**, *11*, 663–673.
- [59] D. Huang, T. Zhou, K. Lafleur, C. Nevado, A. Caflisch, *Bioinformatics* **2010**, *26*, 198–204.
- [60] F. Dey, A. Caflisch, *J. Chem. Inf. Model.* **2008**, *48*, 679–690.
- [61] H. Adalsteinsson, A. H. Maulitz, T. C. Bruice, *J. Am. Chem. Soc.* **1996**, *118*, 7689–7693.
- [62] N. G. Ahn, K. A. Resing, *Science* **2005**, *308*, 1266–1267.
- [63] G. B. Rocha, R. O. Freire, A. M. Simas, J. J. P. Stewart, *J. Comput. Chem.* **2006**, *27*, 1101–1111.
- [64] ROCS, OpenEye Scientific Software, Santa Fe, NM (USA), <http://www.eyesopen.com> (accessed May 12, 2010).
- [65] J. Grant, M. Gallardo, B. Pickup, *J. Comput. Chem.* **1996**, *17*, 1653–1666.
- [66] E. Vangrevelinghe, K. Zimmermann, J. Schoepfer, R. Portmann, D. Fabbro, P. Furet, *J. Med. Chem.* **2003**, *46*, 2656–2662.
- [67] S. M. Rodems, B. D. Hamman, C. Lin, J. Zhao, S. Shah, D. Heidary, L. Makings, J. H. Stack, B. A. Pollok, *Assay Drug Dev. Technol.* **2002**, *1*, 9–19.
- [68] *HotSpot™*, <http://reactionbiology.com/pages/kinase.htm> (accessed May 12, 2010).
- [69] A. Sturz, B. Bader, K. H. Thierach, J. Glienke, *Biochem. Biophys. Res. Commun.* **2004**, *313*, 80–88.
- [70] C. Berset, S. V. Audetat, A. Barberis, T. Gunde, J. Tietz, P. Traxler, A. Schumacher, *Protein Kinase Inhibitors*, US Patent 2007/0149535, <http://www.google.com/patents/about?id=wqigAAAAEBAJ> (accessed May 12, 2010).
- [71] Y. Miyazaki, M. Nakano, H. Sato, A. T. Truesdale, J. D. Stuart, E. N. Nartey, K. E. Hightower, L. Kane-Carson, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 250–254.

Received: February 26, 2010

Revised: April 26, 2010

Published online on June 10, 2010