Is Inhibition Process Better Described with MD(QM/MM) Simulations? The Case of Urokinase Type Plasminogen Activator Inhibitors

Florent Barbault*[a] and François Maurel[a]

Urokinase plasminogen activator (uPA) is an enzyme involved in cancer growth and metastasis. Therefore, the design of inhibitors of uPA is of high therapeutic value, and several chemical families have been explored, even if none has still emerged, emphasizing the need of a rationalized approach. This work represents a complete computational study of uPA complexed with five inhibitors, which present weak similarities. Molecular dynamics simulations in explicit solvent were conducted, and structural analyses, along with molecular mechanics (MM)/Poisson–Boltzmann surface area free energies estimations, yield precious structure–activity relationships of these inhibitors. Besides, we realized supplemental QM/MM computations that improved drastically the quality of our models providing original information on the hydrogen bonds and charge transfer effects, which are, most often, neglected in other studies. We suggest that these simulations and analyses could be reproduced for other systems involving protein/ligand molecular recognitions. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.21983

Introduction

The urokinase system consists of the urokinase type plasminogen activator (uPA), its receptor (uPAR), and its natural inhibitors (PAI-1 and PAI-2). All these proteins act on plasminogen transformation toward plasmin and induce several cellular processes, such as adhesion, migration, proliferation, and mobility.[1–5] These behaviors suggest that urokinase system is of great therapeutic interest. For example, from the inhibiting side of the urokinase system, large amounts of PAI-1 and/or PAI-2 have been observed in various thrombotic disorders,[6,7] diabetes,[8,9] obesity,[10,11] and syndrome X.[10,12] From the other side of the urokinase system, uPA enzyme has been shown to be able to transform plasminogen to plasmin. This last protease is then able to trigger the activation of many matrix metalloproteases.[1] All these proteins degrade collagen, extracellular matrix, and basement membrane proteins, which can lead to tissue destruction.[13] According to its ability of tissue destruction, uPA can be connected to several diseases and disorders, such as multiple sclerosis, rheumatoid arthritis, and AIDS.[14]

In addition to the diseases listed above, Reuning et al. have demonstrated that a disregulation of uPA can be related to cancer growth and metastasis.[15] Indeed, tumor cell invasion strongly depends on the dissociation of cell–cell and/or cell–matrix contacts. uPA, via plasmin formation, permits the digestion of extracellular matrix components and plays a central role in pericellular proteolysis. This way, over-expression of uPA or its receptor increases tumor growth and metastasis rate,[16] whereas over-expression of PAI-1 decreases it.[17] As uPA represents an attractive target against cancer, several approaches have been attempted to interfere with its expression. Antisense oligonucleotide and RNA, antibodies, recombinant or synthetic uPA as well as binding-site uPAR fragment strategies were used, providing various success but also deceptions.[15,18,19] To date, uPA synthetic inhibitors that directly influence the catalytic activity of this enzyme seem to be the most promising approach.[20,21]

One attractive characteristic of known uPA inhibitors is their high chemical diversity. For example, phenylguanidines,[20] pyridinylguanidines,[22] isocoumarins,[23] nonamidine multirings,[24] and diphenyl phosphonates[25] are all inhibitors of uPA but have various chemical structures and properties. All these compounds are reversible ligands with the enzyme. However, it is known that diphenyl phosphonates react with a serine in the active site leading to a stable phosphonylated enzyme.[26–28] Therefore, diphenyl phosphonates can be considered as irreversible ligands. In addition, epidemiological studies have shown that the clinical incidence of prostate cancer varies by geographical area.[29] From this observation eight nutraceutical compounds were spotted[30] by their abilities to inhibit uPA, indicating that a proper diet might support the prevention of prostate cancer. The addition of these eight nutraceutical compounds to the previously presented chemical families increases the chemical diversity of the known inhibitors, whereas, unfortunately, their inhibition constants remain always in the micromolar range. This emphasizes the need to rationalize the structure–activity relationships of known uPA inhibitors.

[a] F. Barbault, F. Maurel
Univ Paris Diderot, Sorbonne Paris Cité, ITODYS, UMR CNRS 7086, 15 rue J-A de Baif, 75205 Paris Cedex 13, France
E-mail: florent.barbault@univ-paris-diderot.fr

Contract/grant sponsor: CINES (Centre Informatique National de l'Enseignement Supérieur, France); Contract/grant number: 2009072208 (Grand Equipement National de Calcul Intensif).

Previously, X-ray structural study provided the structures of five ligands inside the uPA binding cavity with high resolution[21] along with their inhibition constants. Except UKI-1C and UKI-1D, the chemical structures of the inhibitors (shown with bold lines on Fig. 2) are very different. Despite this diversity, the elucidated dissociation constants show a rather small range of values, making these crystal structures a good start for a computational study.

In this work, we aimed at understanding the selectivity of several uPA inhibitors by using molecular dynamics (MD) simulations, molecular mechanics (MM)/Poisson–Boltzmann surface area (PBSA), and QM/MM methods. For this last method, several levels of theory and calculations schemes were used to produce the best correlation between free energy of complexation and biological activities. This correlation could be of great help for the chemical design of new active compounds. Finally, with these computational results in hands, we analyzed the variations of hydrogen bonds and the charge transfer effects, which are, most often, neglected in other studies. All the analyses performed yield meaningful structure–activity relationships, which enable, at last, the elaboration of a chemical pharmacophore.

Computational Methods

Structure preparation

The experimental structures of uPA complexed with the ligands benzamidine (benzenecarboxidiamide), amiloride (amino [(3,5-diamino-6-chloropyrazin-2-yl)carbonyl]amino methaniminium), UKI-1D (2,4,6-triisopropyl-phenylsulfonyl-L-(3-amidino)phenylalanine-piperazine-*N*-β-alanine, and WX293T (*N*-adamantyl-*N*′-(4-guanidinobenzyl)-urea) are available in the protein data base by their respective accession numbers 1F5K, 1F5L, 1F92, and 1EJN. No experimental structure exists for the ligand UKI-1C (2,4,6-triisopropyl-phenylsulfonyl-L-(3-amidino)phenylalanine-piperazine-*N*-ethyloxycarbonyl) in complex with uPA. This ligand differs only from UKI-1D by an ester function instead of an amido tail (see Fig. 2), this was manually modified with the help of the Sybyl 7.2 software,[31] and we checked that this modification did not generate a steric hindrance with the protein. The missing hydrogen positions were created from the X-ray structures according to the internal coordinates defined for each residue by the Amber set of parameters. These atomic positions are optimized within the MD protocol described in the following paragraph.

MD simulations

MD simulations were done with the AMBER 10 software.[32] All atomic partial charges of ligands were obtained from the AM1-BCC computational method.[33,34] The ff03 set of parameters for proteins[35] were used. To realize explicit solvent MD, a water box 10 Å away from any solute atom was created for the five protein/ligand complexes, and the particle mesh Ewald summation method[36] was used to compute the electrostatic forces. First, a minimization of 10,000 steps (2000 steps using the steepest descent method and the other ones with the conjugate gradient algorithm) for water molecules and counter ions was performed with harmonic restraints on solute. Then a similar global optimization of 10,000 steps was carried out without constraint to relax the whole system. The SHAKE algorithm[37] was used for all covalent bonds containing a hydrogen atom so that the time increment was set to 2 fs. The heating phase consisted in six 10-ps steps in NVT. The temperature was gradually increased by 50 K, from 0 to 300 K. The solute was weakly restrained in this stage. Finally, for the five complexes, trajectories of 10 ns were produced in the NTP ensemble. To observe the equilibration, root mean square deviations (RMSD) were recorded for all systems. These equilibrations were always reached before 6 ns, hence the structural analyses were done on the last 4 ns.

Free energies calculations

The MM/PBSA and MM/generalized-born surface area methods [38–41] were used to evaluate protein ligand free energies (ΔG_{bind}) of interaction. Snapshot coordinates were extracted every 20 ps of the 4 ns trajectories. The ΔG_{bind} is calculated using the following equations:

$$\text{Protein} + \text{ligand} \rightarrow \text{Complex}$$

$$\begin{aligned}
\langle \Delta G_{\text{binding}} \rangle &= \langle G_{\text{complexe}} \rangle - (\langle G_{\text{protein}} \rangle + \langle G_{\text{ligand}} \rangle) \\
\langle G_i \rangle &= \langle G_i^{\text{gas}} \rangle + \langle G_i^{\text{solvent}} \rangle \\
\langle G_i^{\text{solvent}} \rangle &= \langle G_i^{\text{polar}} \rangle + \langle G_i^{\text{nonpolar}} \rangle \\
\langle G_i^{\text{gas}} \rangle &= \langle E_i^{\text{gas}} \rangle - T \langle S_i^{\text{gas}} \rangle \\
\langle \Delta G_{\text{binding}} \rangle &= \langle \Delta E_{\text{binding}}^{\text{gas}} \rangle - T \langle \Delta S_{\text{binding}}^{\text{gas}} \rangle + \langle \Delta G_{\text{binding}}^{\text{solvent}} \rangle \qquad (1)
\end{aligned}$$

Values in brackets denote the average obtained from calculations of snapshots coordinates, extracted every 20 ps, of the 4 ns analysis trajectories. Previous works showed us that this proportion of structure was able to provide comparable values to their experimental data for protein/ligand,[42] DNA/ligand,[43] RNA/ligand,[44] and even carbohydrates systems.[45,46] Subscript *i* represent the species used for the calculation (complex, protein, or ligand). E_i, S_i, and G_i are, respectively, the energy, entropy, and free energy of *i*. In MM/PBSA method, the solvent free energy is separated into nonpolar and polar contributions. The nonpolar term was linearly related to solvation accessible surface area[47] obtained from Molsurf calculations.[48] The polar solvation contribution was performed by solving the Poisson–Boltzmann (PB) equation using a finite-difference algorithm defined by Luo et al.[49]

For gas phase, the entropic contributions were divided into translational, rotational, and vibrational motions. The first two terms were obtained from classical statistical mechanics, whereas the contribution from vibrational motion was estimated from a normal mode analysis, for a subset of 10% of the total snapshots, according to the method developed by Kottalam and Case.[50] In this work, the gas-phase energies were calculated either with MM or with QM/MM calculations (see below).

We did not expect that MM/PBSA technique provides us the absolute free energies of binding as this should require other types of simulations, much more time consuming for protein/ligand systems.[51] However, a correlation with the logarithm of experimental inhibition constant was made and the Pearson's R^2 recorded as an evidence of this correlation.

QM/MM calculations

Hybrid QM/MM computations were performed for either selected representative conformations or the subsets of structures extracted for the entropy calculations. To do this, we used the QM/MM module[52] implemented in Amber software with, alternatively, AM1[53] and PM3[54] semiempirical hamiltonians. Such simulations require an accurate definition of the considered QM region. As the objectives of this work were to elucidate the interaction between several ligands and the uPA enzyme, the QM region was considered as the ligand and its surrounding residues defining the binding pocket. These residues were selected if the distance of at least one atom was lower than 6 Å of the inhibitor center of mass. Moreover, we check that no protein ionic residue (Lys, Arg, Asp, or Glu) was within 8 Å of the ligand center of mass to avoid long distance electrostatic effect. This way, the defined QM region contains ligand and the following uPA residues: Gly46, Asp90, Thr91, Leu92, Ala93, His94, Asp192, Ser193, Cys194, Gln195, Gly196, Asp197, Ser198, Val216, Ser217, Trp218, Gly219, Arg220, Gly221, Cys222, Ala223, Pro228, and Gly229. To get a view of the system size, the QM part contains from 316 to 390 atoms, in function of the ligand size. Boundary atoms were treated with the link atom method,[52,55,56] and hence hydrogen atoms were added to satisfy valence requirements.

QM/MM energies

The QM/MM energies were obtained from either single-point energies or by minimizing the systems with several steps of steepest descent and conjugated gradient. The influence of this optimization, along with the number of steps considered, is discussed in the result section of this manuscript. The binding energies were extracted according to the following equation where values in brackets represent an average:

$$\left\langle \Delta E^{\text{QM MM}}_{\text{binding}} \right\rangle = \left\langle E^{\text{QM MM}}_{\text{complex}} \right\rangle - \left\langle E^{\text{QM MM}}_{\text{protein}} \right\rangle - \left\langle E^{\text{QM MM}}_{\text{ligand}} \right\rangle$$

This energy is incorporated in the MM/PBSA equation by replacing the gas phase binding energy of eq. (1).

However, large variations of QM/MM interaction energies were sometimes encountered, indicating probable inconsistencies with the MM based structures. To overcome this artifact, one, two or, at worst, three points were sometimes removed from the whole dataset used to calculate the average. To select these points, we defined a fidelity coefficient as follows:

$$\%f = \sqrt{\frac{\sum_i^N (E_i - \bar{E})^2}{\bar{E}^2 (N-1)}} \times 100$$

E_i represents the QM/MM interaction energy of conformation i and $\bar{E}$ represents its average value. This fidelity coefficient tool is a good probe to locate which point should be discarded from the correlation. In this work, no variation up to 8% was observed.

Atomic charge variations

Atomic charge variations effect has been studied from these QM/MM simulations. Mulliken's atomic charges were recorded for each calculation. Differences between charges obtained in the complex state and the free protein or ligand states were then calculated and averaged for all conformations. To get values that can be visualized on the molecular structure, all have been normalized, defining a charge impact (CI). This way, the CI of an atom j, for N conformations, is described by the following equation:

$$\text{CI}_j = \frac{\sum |q_j^{\text{complex}} - q_j^{\text{free}}|}{N \times \max\{|q^{\text{complex}} - q^{\text{free}}|\}}$$

This definition implies that the CI is always between 0 and 1. A value of 1 signifies that this atom has the largest charge variation between the free and complex state, on the contrary of a value of 0, which indicates that there is no charge variation between these two states. These CI values could then be visualized for all atoms of the QM area as an occupancy factor with the help of visual molecular dynamics (VMD) software.[57] Blue, green, and red colors, respectively, indicate low, moderate, and high polarization effects (see Figs. 4 and 5).

Hydrogen bonds

In all MD trajectories, a hydrogen bond was considered when the donor–acceptor distance was smaller than 3.5 Å and the donor–acceptor–H angle smaller than 60°. These weak criteria were chosen deliberately to count as much as possible hydrogen bonds even those which represent weak interactions. Similar analyses were produced for the structures submitted to QM/MM computations. In all cases, distances and angles values, along with their respective standard deviations, were compiled and compared together.

Results and Discussion

Overall analysis

Monitoring of kinetic, potential, and overall energies along the trajectory, as well as the density, pressure, and temperature, demonstrates the stability of the five uPA complexes. For all systems, RMSD curves of selected residues were produced, along the 10 ns trajectories, to observe molecular movements and fluctuations. Only the curves of one complex are displayed on Figure 1. For the five uPA complexes, all curves reached a plateau after ~2 ns. Comparison of the average protein structures in all complexes shows that the secondary structures and globular form of uPA is kept during the simulations emphasizing, this way, the enzymatic role of uPA.

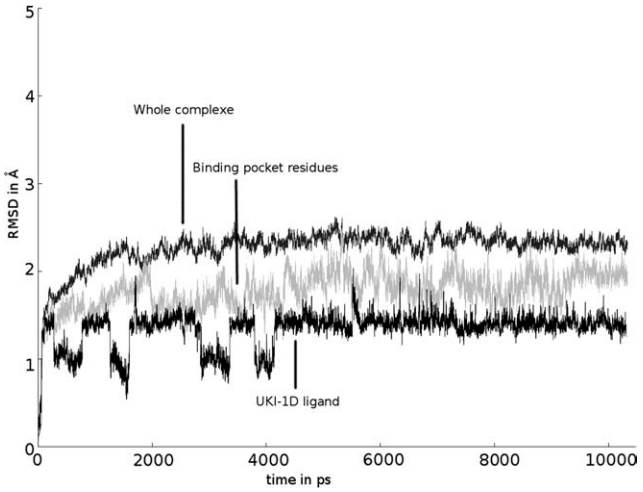

Figure 1. RMSD for the 10 ns MD of uPA/UKI-1D system. Red: RMSD curve for the whole complex; green: RMSD curve for the binding pocket residues; blue: RMSD curve for the UKI-1D inhibitor.

In addition, RMSD curves of the ligand/binding pocket provide meaningful information on the ligand binding mode. Surprisingly, these curves are similar for all complexes, despite the differences in sizes and shapes of the considered ligands. In all cases, RMSD graphs reach quickly a plateau, as shown on Figure 1 for ligand UKI-1D. For all systems, the RMSD curves profiles of the whole complex are similar to the RMSD of the ligand/binding pocket. This indicates that molecular movements are mainly described by the residues of the ligand/binding pocket.

Hydrogen bond

Table 1 and Figure 2 summarize the hydrogen bond networks recorded for all inhibitors at their uPA interface. For all trajectories, the percentage time of presence is reported and only the most representatives, up to 70% of time, are described here. These hydrogen bonds are considered essential for the binding on uPA, and their distances are reported in Table 1.

Three hydrogen bonds are observed for benzamidine ligand. One hydrogen bond is always reported (99%) along the trajectory and concerns the amino group of benzamidine with the carbonyl of Ser193. The second reported hydrogen bond implies the same amino group of benzamidine with the Asp192 carboxylate moiety. Another hydrogen bond is formed with a water molecule, which is present for 71% of time.

Because of the high number of amino groups on amiloride, this compound acts as hydrogen bond donor when it is bonded on uPA. Three hydrogen bonds, present more than 90% of time, were reported. The first one implies the amido part of ligand with the carbonyl oxygen of Gly219. The second is observed between one guanidinium hydrogen and one oxygen of the Asp192 carboxylate. The third one is recorded between the aromatic amino moiety, in para position of the chlorine atom, and the hydroxyl group of Ser198. These hydrogen bonds are, spatially, located at different positions presenting, this way, a better anchoring feature inside uPA than benzamidine. This observation is also supported by a significant decrease of free energy (see later) for these two compounds so that it may explain the decrease of inhibition constant of amiloride showing a K_i of 5.3 μM instead of 180 μM for benzamidine.

Table 1. Hydrogen bond networks for the five uPA inhibitors.

	X-ray	MD		QM/MM	
N°	*d*	%	*d*	*d*	Delta
Benzamidine					
1	3.48	99	2.89	2.96	−0.07
2	2.79	71	3.16	2.89	0.27
3	5.68	71	3.06	ND	ND
Amiloride					
1	3.31	99	2.94	3.01	−0.07
2	2.56	99	2.93	2.96	−0.03
3	2.80	90	3.02	3.35	−0.33
4	3.63	84	3.18	3.05	0.13
5	2.73	84	2.90	2.97	−0.07
UKI-1C					
1	ND	99	2.91	2.90	0.01
2	ND	96	3.01	2.89	0.12
3	ND	88	3.15	3.27	−0.12
4	ND	87	3.01	2.91	0.10
UKI-1D					
1	3.48	99	3.05	3.06	−0.01
2	3.08	98	2.93	3.17	−0.24
3	2.82	97	2.94	3.19	−0.25
4	2.83	93	3.04	2.89	0.15
5	2.93	85	2.94	2.90	0.04
WX293T					
1	3.06	99	2.97	3.18	−0.21
2	3.12	98	3.05	2.95	0.10
3	2.82	98	2.89	2.92	−0.03
4	3.68	97	3.07	3.16	−0.09
5	2.90	95	2.91	2.91	0.00
6	2.93	95	2.98	2.90	0.08

N° represents the number reported in Figure 2. % is the percentage time of presence along the MD trajectory, for the sake of clarity only the most representative (up to 70%) are described here. *d* is the distance between the donor–acceptors atoms. The column "delta" represents the difference between distance recorded in MD simulation and in the QM/MM computations. All distances are in Å.

Even if UKI-1C and UKI-1D possess high chemical similarities, their hydrogen bond networks are different (cf. Fig. 2), explaining this way the variations of interactions energies (see Table 2). These variations may be related to the distinct experimental inhibition constants of UKI-1C and UKI-1D with, respectively, 0.41 and 0.64 μM. However, from X-ray and MD analyses, no specific hydrogen bond was found to explain the difference of activity for the differing chemical part of these two ligands.

The guanidinium ion of WX293T is localized such as its similar group in amiloride ligand. Indeed, even if there is not a complete reproduction of all hydrogen bonds, a common feature of interaction with Asp192 is observed (see Fig. 2). For the other parts of the WX293T ligand, two supplementary strong hydrogen bonds are made with the carbonyl group of Ser217 and the two NH groups of the diamido part of WX293T. This way, these interactions locate the neighboring adamantane moiety through the uPA cavity generated by the histidines 94 and 46. This stabilizing contribution is described later in this article.

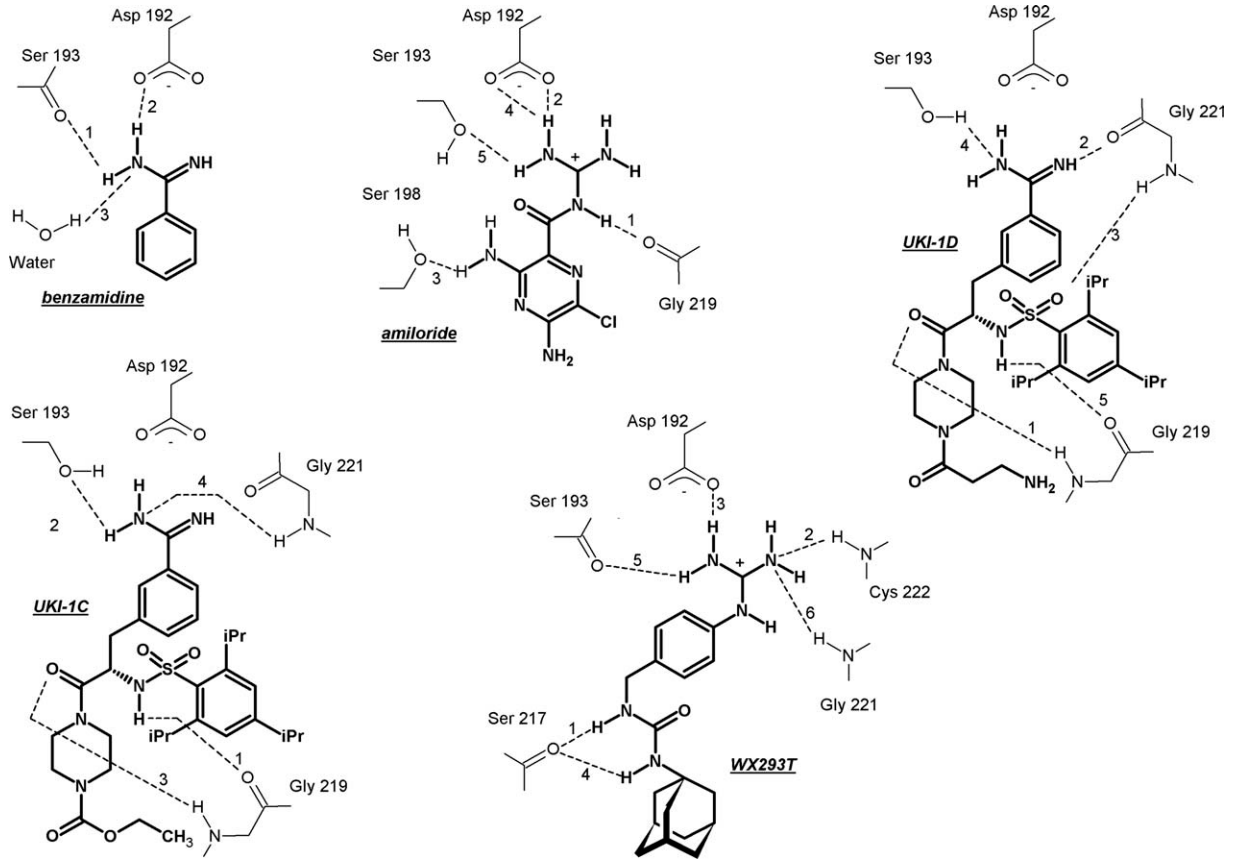

Figure 2. Schemes of the main hydrogen bond recorded during the MD simulations. The numbers refer to Table 1.

For all ligands, the recorded hydrogen bonds above were also described on the experimental X-ray structures.[21] These distances are compiled on Table 1, and we observe only small variations of their values. This emphasizes the quality of the MD simulations and indicates that the movements explored represent only minor structural adaptations of the ligands inside the uPA active site.

structural examination of interacting mode emphasized the essential role of uPA Asp192, already highlighted on several uPA/inhibitor structural studies.[21,59] Ligands WX293T and amiloride, which are ions, take advantage of this accessible negative charge, and therefore have gas electrostatic energies values around twice more important than the neutral compounds (see Table 2).

Free energy

All calculated MM/PBSA free energies have negative values, hence demonstrating a favorable interaction for the ligands. These values, along with their decompositions, are compiled on Table 2 and are linearly compared with the logarithm of the experimental inhibition constants (log K_i). This correlation, shown on the top of Figure 3, provides a Pearson coefficient (R^2) of 0.728, indicating this way an acceptable reproduction of the ligands affinities.[58]

The solvent contributions are positive for all systems and balance the favorable gas contributions. uPA interaction with benzamidine is mainly described, with around 60% of its gas energy (−38.4 kcal/mol), by its Van der Waals contribution (−22.6 kcal/mol). Similar ratios are also obtained for ligands UKI-1C and UKI-1D. In all cases, the

Table 2. MM/PBSA and (QM/MM)PBSA energies for the five uPA inhibitors.

Ligand name	Benzamidine	Amiloride	UKI-1D	UKI-1C	WX293T
Gas contribution					
ΔE_{elec}	−15.9	−73.3	−37.6	−38.5	−74.0
ΔE_{vdw}	−22.6	−30.7	−60.0	−59.6	−31.4
ΔE^{gas} (MM)	−38.4	−104.0	−97.6	−98.1	−105.4
ΔE^{gas} (QM/MM)	−35.0	−97.5	−102.9	−100.6	−81.6
$-T\Delta S^{gas}$	+19.3	+22.2	+26.0	+29.6	+24.7
ΔG^{gas} (MM)	−57.7	−126.2	−123.6	−127.6	−130.2
ΔG^{gas} (QM/MM)	−54.3	−119.8	−128.9	−130.1	−106.3
Solvent contribution					
$\Delta G_{nonpolar}$	−2.0	−3.6	−7.0	−6.7	−5.3
ΔG_{polar}	+29.3	+72.5	+69.8	+59.6	+57.1
$\Delta G_{solvent}$	+27.3	+68.8	+62.8	+52.9	+51.9
$\Delta G_{binding}$ (MMPBSA)	−30.5	−57.4	−60.8	−74.8	−78.3
$\Delta G_{binding}$ (QM/MMPBSA)	−27.1	−50.9	−66.1	−77.3	−54.5
Experimental K_i (μM)	180	5.3	0.64	0.41	2.4
log(K_i)	−3.745	−5.276	−6.387	−6.194	−5.620

All values are in kcal/mol.

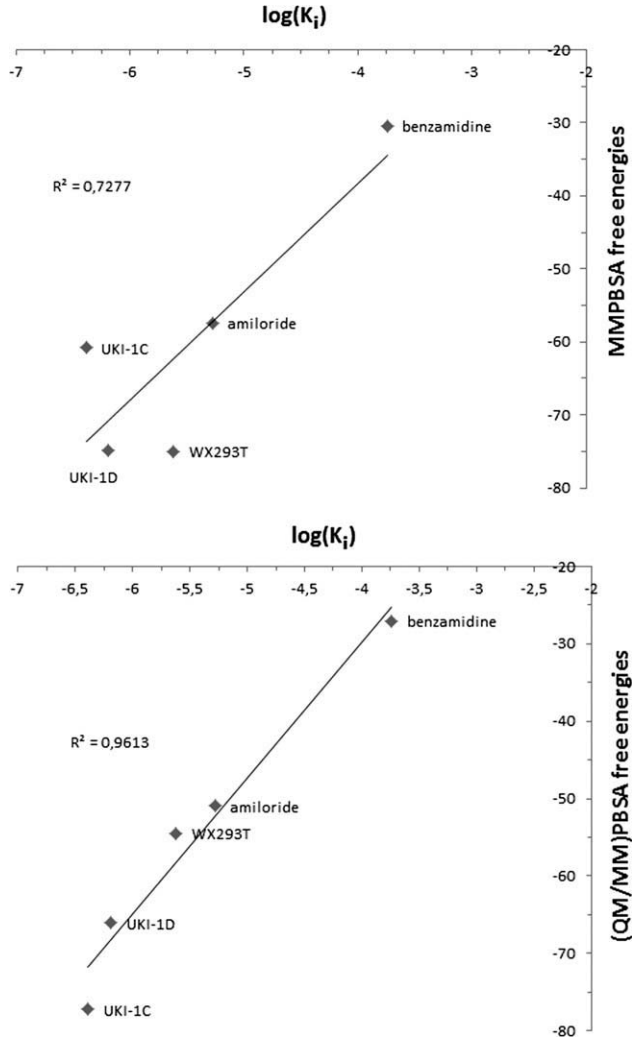

Figure 3. Correlations curves for MM/PBSA (top) and (QM/MM)PBSA (bottom) with the experimental free energies. For both correlations the calculated *p*-values are under the significance level of 0.01 ($p < 0.01$).

The correlation plotted on top of Figure 3 shows a dispersion of points around the least-square curve. To improve this correlation, we decided to perform QM/MM calculations to obtain better gas phase energies [see eq. (1)].

QM/MM studies

To date, several QM/MM studies applied to biomolecules have been successfully undertaken to study the ligand/macromolecule recognition. However, most of these studies are limited to only one representative structure[43,60,61] extracted from MD trajectories. Here, we aimed to perform a QM/MM study on a set of structures, and we investigated whether the minimization and the level of theory could improve the gas phase energies used in the MM/PBSA calculations [see eq. (1)], thus providing (QM/MM)PBSA energies.

Our first attempt was to perform the QM/MM calculations with only one representative structure of the five uPA/ligand systems. To obtain these, we extracted, from their MD trajectories, the structures which displayed the lowest RMSD of their average coordinates. (QM/MM)PBSA values, realized with AM1 hamiltonian, are compiled in Table 3, and a correlation with experimental data yielded a R^2 of 0.161. This low correlation value specifies that the inhibition variations of ligands are far from being reproduced, emphasizing the difficulty of taking only one structure into account to describe, efficiently, uPA/ligands interactions processes.

To overcome the fact that one representative structure failed to provide meaningful results, we realized these QM/MM calculations on a subset of structures. Single-point interactions energies were then calculated for all complexes using the AM1 semiempirical hamiltonian for the QM part. These values are reported on Table 3. For some structures, energies values were very far from the others. These weird points damped the average so that we decided to remove them. Therefore, we defined a statistical tool, named fidelity coefficient (see computational details), and removed the points that provide an error up to 8%. A leave-one-out (LOO) strategy[58] was then undertaken—although a simple view of the energies values could also localize these outliers. The number of conformations, given in the Table 4, is then the minimal number of conformations used to calculate the interaction energy of any uPA/ligand complex. To definitely validate this approach, we compared the results with the correlations produced with the whole dataset of conformers, that is, without removing outliers highlighted by the fidelity coefficient. These correlations produced R^2 coefficients, compiled in Table 3, which have their

Table 3. MM/PBSA and (QM/MM)PBSA values for the five uPA inhibitors (in kcal/mol).

			(QM/MM)PBSA						
			AM1					PM3	
Ligands	Log(K_i)	MM/PBSA	Rep	SP	SP*	Min	Min*	Min	Min*
Benzamidine	−3.745	−30.5	−40.4	−22.1	−26.2	−27.1	−33.7	−33.0	−35.9
Amiloride	−5.276	−57.4	−145.7	−29.9	−24.7	−50.9	−50.9	−43.0	−39.1
UKI-1D	−6.194	−60.8	−135.5	−29.8	−30.4	−66.1	−62.9	−61.9	−60.5
UKI-1C	−6.387	−74.8	−65.6	−34.6	−33.1	−77.3	−77.3	−59.6	−59.6
WX293T	−5.620	−78.3	−192.1	−35.6	−29.9	−54.5	−46.5	−48.6	−48.6
Num conformers		200	1	15	Full	18	Full	17	Full
R^2		0.728	0.161	0.683	0.553	0.961	0.804	0.910	0.798

AM1 Rep stands for representative structures and AM1 SP for single point. Num of conformers represents the minimal number of conformers used for the correlation. R^2 is the Pearson coefficient.

Table 4. Mulliken's atomic partial charges for a part of WX293T interacting with uPA Asp 192.

Atom	Free state	Complex state	Delta	%CI
1′	−0.1748 ± 0.0107	−0.1977 ± 0.0479	0.0229	19
1	−0.1772 ± 0.0217	−0.1773 ± 0.0426	0.0001	0
2	0.1558 ± 0.0150	0.1639 ± 0.0852	−0.0081	7
3	−0.1479 ± 0.0247	−0.1682 ± 0.0859	0.0203	17
4	0.3178 ± 0.0214	0.3123 ± 0.0578	0.0055	5
5	0.2640 ± 0.0135	0.2954 ± 0.0551	−0.0314	26
6	−0.2708 ± 0.0063	−0.2919 ± 0.0455	0.0211	17
7	0.2542 ± 0.0027	0.3035 ± 0.0440	−0.0493	41
8	0.2541 ± 0.0031	0.3052 ± 0.0425	−0.0511	42
9	−0.2262 ± 0.0129	−0.3468 ± 0.0656	0.1206	100
10	0.2550 ± 0.0225	0.1917 ± 0.1143	0.0633	52
11	0.2578 ± 0.0282	0.1969 ± 0.1153	0.0609	50
12	−0.6172 ± 0.0380	−0.5523 ± 0.0611	−0.0649	54
13	−0.6120 ± 0.0653	−0.5186 ± 0.0983	−0.0934	77
14	0.3375 ± 0.0344	0.3056 ± 0.0277	0.0319	26
15	0.0736 ± 0.0203	0.0493 ± 0.0222	0.0243	20
16	0.1157 ± 0.0378	0.0966 ± 0.0303	0.0191	16
17	−0.2500 ± 0.0305	−0.2649 ± 0.0399	0.0149	12
18	0.0847 ± 0.0723	0.0890 ± 0.0662	−0.0043	4

Free and complex states values are reported with their standard deviations. Delta values represent the free state minus the complex state. %CI is the percentage of CI (see computational details).

values decreased by only around 0.1 than those obtained with the improved dataset. This result demonstrates the reliability of this method, and hence the fidelity coefficient appears a good decision tool, giving a neutral interpretation of how outliers were identified. Therefore, this tool has been used for the rest of the QM/MM computations.

The use of several conformations yields a real improvement when compared with the previous result. However, the obtained R^2 of 0.683 is still lower than the value obtained with the classical use of MM/PBSA ($R^2 = 0.728$). To obtain better results from QM/MM simulations, energetic minimizations were performed for all conformers (see Supporting Information). The final energy values for the AM1 hamiltonian are represented in Table 3, whereas the correlation graph with $\log(K_i)$ is shown on Figure 3 and exhibits a R^2 of 0.961. This represents a deep improvement when compared with classical MM/PBSA analysis, indicating this way that these QM/MM models are now able to reproduce, with high fidelity, the variations of ligands activities toward the uPA enzyme.

Prior to perform structural analysis on these successful results, similar calculations were reproduced with PM3 semiempirical hamiltonians. The results, presented in Table 3, exhibit a similar correlation with experimental data. This demonstrates that, from practical point of view, the most important parameters to control for (QM/MM)PBSA computations are the number of structures considered and their energy minimization. All the remaining QM/MM analyses shown below are those performed with the AM1 hamiltonian.

Hydrogen bonds variations

The hydrogen bonds identified in the MD trajectories (see above) were analyzed with the resulting QM/MM structures. For all hydrogen bonds, distances were computed and averaged. All these data are reported in Table 1 along with the differences of distances obtained on the two types of simulations.

Globally, variations of hydrogen bond distances are weak, underlining, this way, the simulations coherence. However, some distance differences are more than 0.2 Å and may be considered as significant. For benzamidine, the hydrogen bond number 2, involving Asp192, became reinforced by the QM/MM simulations, whereas for UKI-1D, the hydrogen bonds 2 and 3, implicating for both Gly221, appear to be weakened. With the highest variation, 0.33 Å, the hydrogen bond 3 of amiloride decreases strongly and may not be considered as a strong hydrogen bond interaction. Similar analysis could be drawn for WX293T where the interaction number 1 became weakened.

All these results exhibit one supplemental advantage of QM/MM computations: the ability to discriminate the real hydrogen bonds from these elucidated geometrically in the MD trajectory.

Atomic charge variations

Comparison of correlations, shown on Figure 3, indicates which ligands have been the most influenced by the QM/MM energy refinement. With energies differences of 6.43 and 23.84 kcal/mol, amiloride and WX293T exhibit the strongest variations. These compounds are the only ones bearing a +1 charge, and we suggest that their electrostatic energies, through MM of MM/PBSA, are overestimated.

This problem has been already demonstrated for five homologous α-aminoalkylphosphonate diphenyl esters uPA inhibitors with QM studies on a highly simplified protein environment.[59] More generally, the overestimation of electrostatic potential is a recurrent problem in docking scoring functions for charged molecules like ligand/RNA systems.[62–65] In MM, the electrostatic overestimation comes from the use of a model where the atomic partial charges never evolve during a molecular interaction. In QM simulations, the variations of atomic charges are intrinsically taken into account through simulation of the wavefunction.

To analyze this effect, differences of atomic partial charges, found in the complex state and its two isolated partners (protein and ligand), have been computed. WX293T molecule, with the largest variation of (QM/MM)PBSA free energy, yields the best illustration of this effect. Mulliken atomic partial charges were measured and averaged for each conformation for the complex uPA/WX293T, uPA alone and WX293T alone. These values were then normalized, according to the equation defined in the computational details, to be visualized on a molecular viewer, such as VMD. This way, Figures 4 and 5 illustrate the charge transfer process of WX293T toward uPA enzyme.

In WX293T, three distinct parts appear to be involved in a charge transfer process: the adamantane, diamide, and guanidinium moities. The adamantane part, which is nested inside a V shaped cavity defined by His46 and His94, globally loses atomic charges, whereas both interacting histidines gain them

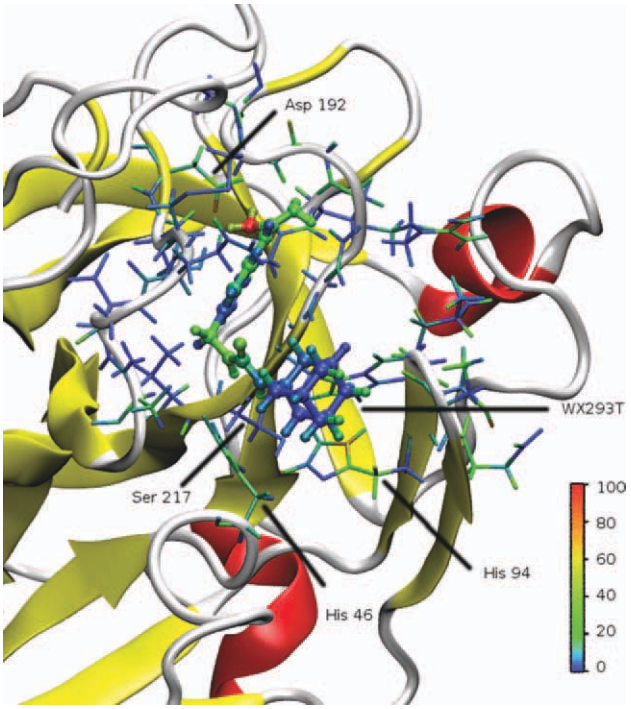

Figure 4. Inhibitor WX293T complexed with uPA protein in ribbon. Ligand is represented in ball and stick whereas binding pocket residues in stick. Coloration is made according to the CI percentage (see computational details).

(see Fig. 4). The second part of WX293T ligand concerns the diamide moiety which presents two strong hydrogen bonds with the backbone oxygen of Ser217. These interactions are reinforced during the QM/MM simulations (see hydrogen bonds analyses), and therefore have an impact on charge transfer. When interaction occurs, we observe that both hydrogens of the diamide, and their bonded nitrogens, increase their electropositive behavior and strengthen the electronegativity of the Ser217 oxygen. The largest charge transfer effect is given by the ionic interaction between the positive guanidinium part of WX293T and the facing negative charge of uPA Asp192. Table 4 and Figure 5 illustrate this process and clearly demonstrate that guanidinium becomes less positive, with a loss of 0.152 charge, and conversely, that carboxylate group becomes less negative, with a gain of 0.072 charge. No accordance is found between gain and loss of charge, showing that the charge transfer of guanidinium also occurs for the aromatic ring of WX293T, with a gain charge of 0.023, and also as a dispersion (0.057) onto the uPA protein environment. This way, the electrostatic interaction is less important than expected with MM calculations.

By correcting the electrostatic effect through QM/MM calculations, the (QM/MM)PBSA method presented here is able to predict, with a high reliability, the variations of inhibition constants for several uPA inhibitors. Therefore, the procedure issued in this work might be used as a supplementary tool for accurate evaluation of selectivity when computer-aided drug design (CADD) strategies are undertaken.

Implication for drug design

More than its methodological aspect, this study evidences several key interactions desired for uPA inhibition. These allow us the elaboration of a pharmacophoric scheme represented on Figure 6.

The MD trajectories emphasized the structural rigidity of uPA binding site. Therefore, future inhibitors should present a correct shape to bind the enzyme active site, as its flexibility is weak.

This study highlights the key role of Asp192 through electrostatic interaction. One might suppose that an ionic ligand should enhance the molecular recognition with this amino acid. However, the QM/MM calculations demonstrated that, due to the atomic charge transfer, such ionic compounds would be disappointing. Therefore, we propose to put an aromatic amidine moiety in front of the Asp192 residue.

On the meta position of this moiety, a hydrogen bond donor group (Hd on Fig. 6), such as OH or NH_2, may generate one hydrogen bond with Ser198. Chirality of the asymmetric carbon must be respected to locate ideally the molecular extensions on their respective protein cavities.

On the axis of the aromatic amidine group, the first extension part may contain an electronegative atom (A), such as oxygen or sulfur, and a NH group. These atoms may produce two hydrogen bonds with, respectively, Gly219 and Ser217. A hydrophobic group

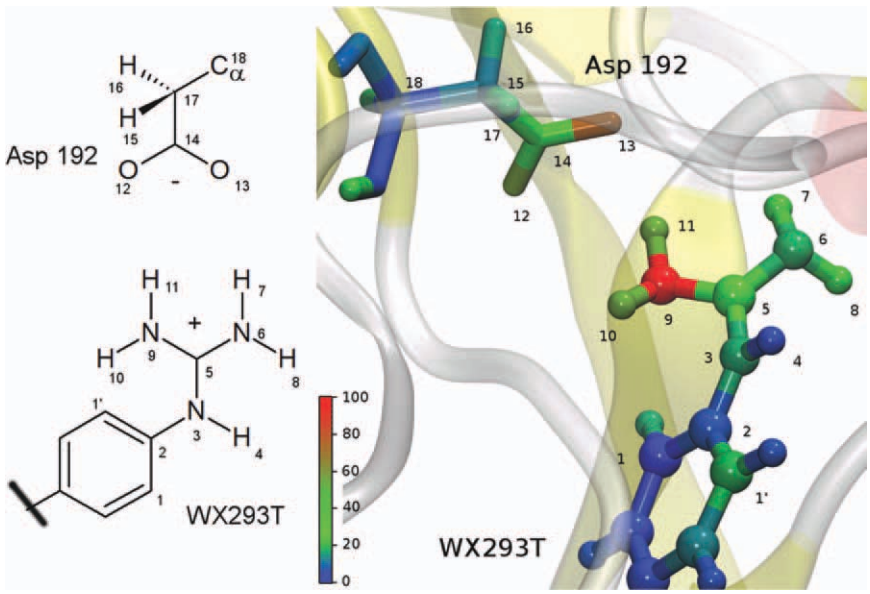

Figure 5. Left, atom numbering of a part of WX293T along with Asp 192 of uPA protein. Atomic charge values are reported in Table 4. Right, zoom of this interaction colored according to the CI percentage (see computational details).

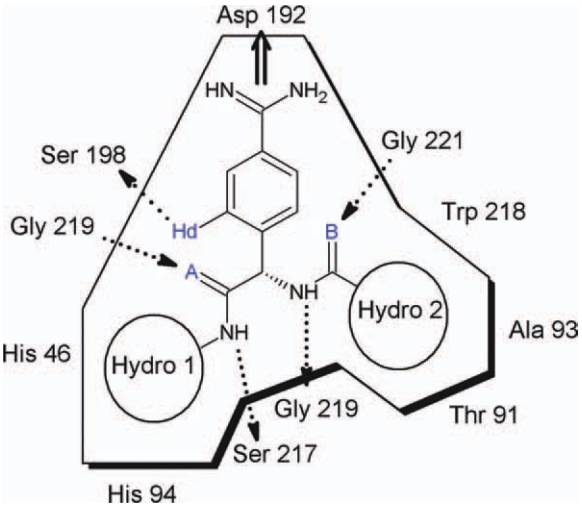

Figure 6. Pharmacophoric scheme proposal based on the interactions elucidated for the five uPA inhibitors.

(hydro 1) should complete this molecular extension and fill one uPA cavity created by histidines 46 and 94.

A supplemental extension may be connected on the asymmetric carbon. Similarly of the previous extension, this one can contain a NH group followed by an electronegative atom (B). These atoms may create two hydrogen bonds with the glycines 219 and 221. At last, a hydrophobic group (hydro2) could fill the uPA cavity created by residues Thr91, Ala93, and Trp218. This cavity is rather small and so a simple aromatic group, branched or not, may have enough room on this pocket.

Conclusions

This work represents a complete computational study of five uPA inhibitors. Starting from crystal structures of their complexes, MD simulations were undertaken with explicit solvent for 10 ns. The hydrogen bond analyses yielded crucial informations about the binding mode of the inhibitors. Free energies of binding, first obtained through MM/PBSA analyses, was able to reproduce a linear relationship with their experimental $\log(K_i)$. To further improve our models, we realized QM/MM computations and replaced the interaction energies in the MM/PBSA module, then implementing a (QM/MM)PBSA analysis.

Many protocols of QM/MM were tried, and we proved that the use of several optimized conformations, instead of one 'representative,' significantly improve the correlation with the experimental data. We obtained a remarkable correlation, with a R^2 of 0.961, with the AM1 hamiltonian. However, this result may not be directly transferable to any other systems because this approach needs an extensive validation to a more important dataset. Moreover, when applied to more complex biological systems, such as, proteins complexed with a metal, the AM1 or PM3 hamiltonians will probably fail to provide meaningful results.

Some hydrogen bonds were enhanced by their QM description, which then appears to be able to discriminate between a true or false interaction. We also demonstrated that the improvement of the correlation made by QM/MM calculations came mainly from the atomic partial charge variations. This phenomenon was quantified and normalized to be visualized on three-dimensional structures, as exemplified on Figures 4 and 5. We proved that the interaction of ionic ligands is always overestimated, which must be taken into account in the context of computer-aided drug design (CADD) strategies. Indeed, ionic molecules are well known to present artificial high score on molecular docking, so they are usually put into a specific subgroup for being evaluated separately.[66]

The computational procedure introduced in this article enabled an accurate description of the binding mode for uPA inhibitors with different chemical structures, shapes, functional groups, and/or global charges. From these invaluable information, a pharmacophoric scheme has been generated with the will of being useful to the future design of more powerful inhibitors.

Keywords: uPA · urokinase · molecular dynamics · hybrid methods · QM/MM · MM/PBSA · (QM/MM)PBSA · free energy of binding · Hydrogen bond · atomic charges variations

How to cite this article: F. Barbault, F. Maurel, *J. Comput. Chem.* **2012**, *33,* 607–616. DOI: 10.1002/jcc.21983

Additional Supporting Information may be found in the online version of this article.

[1] J. P. Irigoyen, P. Munoz-Canoves, L. Montero, M. Koziczak, Y. Nagamine, *Cell. Mol. Life. Sci.* **1999**, *56,* 104.
[2] A. E. May, S. M. Kanse, T. Chavakis, K. T. Preissner, *Fibrinol. Proteol.* **1998**, *12,* 205.
[3] J. C. Porter, N. Hogg, *Trends Cell. Biol.* **1998**, *8,* 390.
[4] S. A. Rabbani, P. Harakida, D. J. Davidson, J. Henkin, A. P. Mazar, *Int. J. Cancer* **1995**, *63,* 840.
[5] S. A. Rabbani, R. H. M. Xing, *Int. J. Oncol.* **1998**, *12,* 911.
[6] V. Grimaudo, F. Bachmann, J. Hauert, M. A. Christe, E. K. Kruithof, *Thromb. Haemostasis* **1992**, *67,* 397.
[7] B. Wiman, *Thromb. Haemostasis* **1995**, *74,* 71.
[8] I. Juhan-Vague, M. C. Alessi, P. Vague, *Ann. Med.* **1996**, *28,* 371.
[9] S. R. Torr-Brown, B. E. Sobel, *Thromb. Res.* **1994**, *75,* 473.
[10] I. Juhan-Vague, M. C. Alessi, *Thromb. Haemostasis* **1997**, *78,* 656.
[11] A. Quattrone, A. Gambardella, A. M. Carbone, R. L. Oliveri, A. Lavano, E. V. De Marco, D. Civitelli, F. Bono, M. Zappia, K. Pardatscher, G. DiMinno, *J. Neurol.* **1999**, *246,* 1086.
[12] I. Juhan-Vague, S. G. Thompson, J. Jespersen, *Arterioscler. Thromb.* **1993**, *13,* 1865.
[13] P. Mignatti, D. B. Rifkin, *Physiol. Rev.* **1993**, *73,* 161.
[14] T. W. Rockway, V. Nienaber, V. L. Giranda, *Curr. Pharm. Des.* **2002**, *8,* 2541.
[15] U. Reuning, V. Magdolen, O. Whilem, K. Fischer, V. Lutz, H. Graeff, M. Schmitt, *Int. J. Oncol.* **1998**, *13,* 893.
[16] R. L. Shapiro, J. G. Duquette, D. F. Roses, I. Nunes, M. N. Harris, H. Kamino, E. L. Wilson, D. B. Rifkin, *Cancer Res.* **1996**, *56,* 3597.
[17] R.-P. Czekay, D. J. Loskutoff, *J. Cell. Phys.* **2009**, *220,* 655.
[18] F. Fazioli, F. Blasi, *Trends Pharmacol. Sci.* **1994**, *15,* 25.
[19] M. Schmitt, N. Harbeck, C. Thomssen, O. G. Wilhelm, V. Magdolen, U. Reuning, K. Ulm, H. Hölfer, F. Jänicke, H. Graeff, *Thromb. Haemostasis* **1997**, *78,* 285.
[20] S. Sperl, U. Jacob, N. Arroyo de Prada, J. Stürzebecher, O. G. Wilhelm, W. Bode, V. Magdolen, R. Huber, L. Moroder, *Proc. Natl. Acad. Sci. USA* **2000**, *97,* 5113.
[21] E. Zeslawska, A. Schweinitz, A. Karcher, P. Sondermann, S. Sperl, J. Stürzebecher, U. Jacob, *J. Mol. Biol.* **2000**, *301,* 465.
[22] C. G. Barber, R. P. Dickinson, V. A. Horne, *Bioorg. Med. Chem. Lett.* **2002**, *12,* 181.

[23] J. J. Heynekamp, L. A. Hunsaker, T. A. Vander Jagt, L. M. Deck, D. L. Vander Jagt, *BMC Chem. Biol.* **2006**, *6,* 1.
[24] C. W. West, M. Adler, D. Arnaiz, D. Chen, K. Chu, G. Gualtieri, E. Ho, C. Huwe, D. Light, G. Phillips, R. Pulk, D. Sukovich, M. Whitlow, S. Yuan, J. Bryant, *Bioorg. Med. Chem. Lett.* **2009**, *19,* 5712.
[25] J. Joossens, P. Van der Veken, A.-M. Lambeir, K. Augustyns, A. Haemers, *J. Med. Chem.* **2004**, *47,* 2411.
[26] A. Belyaev, X. Zhang, K. Augustyns, A. M. Lambeir, I. De Meester, I. Vedernikova, S. Scharpe, A. Haemers, *J. Med. Chem.* **1999**, *42,* 1041.
[27] J. A. Bertrand, J. Oleksyszyn, C. M. Kam, B. Boduszek, S. Presnell, R. R. Plaskon, F. L. Suddath, J. C. Powers, L. D. Williams, *Biochemistry* **1996**, *35,* 3147.
[28] J. Oleksyszyn, B. Boduszek, C. M. Kam, J. C. Powers, *J. Med. Chem.* **1994**, *37,* 226.
[29] M. M. Rafi, *Nutrition* **2004**, *20,* 78.
[30] J. Jankun, S. H. Selman, J. Aniola, E. Skrzypczak-Jankun, *Oncol. Rep.* **2006**, *16,* 341.
[31] SYBYL, Tripos Inc.: 1699 South Hanley Rd., St Louis, Missouri, 63144 USA. Available online at http://tripos.com.
[32] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, B. Wang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. F. Wong, F. Paesani, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, University of California: San Francisco, **2008**. Available online at http://ambermd.org/.
[33] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2000**, *21,* 132.
[34] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23,* 1623.
[35] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. A. Kollman, *J. Comput. Chem.* **2003**, *24,* 1999.
[36] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98,* 10089.
[37] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23,* 327.
[38] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, 3rd, *Acc. Chem. Res.* **2000**, *33,* 889.
[39] I. Massova, P. A. Kollman, *J. Am. Chem. Soc.* **1999**, *121,* 8133.
[40] J. Wang, P. Morin, W. Wang, P. A. Kollman, *J. Am. Chem. Soc.* **2001**, *123,* 5221.
[41] W. Wang, P. A. Kollman, *Proc. Natl. Acad. Sci. USA* **2001**, *98,* 14937.
[42] A. Li, F. Barbault, F. Maurel, M. Delamar, B. Wang, *J. Mol. Struct: THEOCHEM* **2008**, *849,* 62.
[43] A. Li, F. Maurel, F. Barbault, M. Delamar, B. Wang, X. Zhou, P. Wang, *Eur. J. Med. Chem.* **2010**, *45,* 983.
[44] F. Barbault, B. Ren, J. Rebehmed, C. Teixeira, Y. Luo, O. Smila-Castro, F. Maurel, L. Zhang, L. Zhang, *Eur. J. Med. Chem.* **2008**, *43,* 1648.
[45] C. Gourmala, Y. Luo, F. Barbault, Y. Zhang, S. Ghalem, F. Maurel, B. T. Fan, *J. Mol. Struct: THEOCHEM* **2007**, *821,* 22.
[46] Y. Luo, F. Barbault, C. Gourmala, Y. Zhang, F. Maurel, H. Yongzhou, B. T. Fan, *J. Mol. Model* **2008**, *14,* 901.
[47] D. Sitkoff, K. A. Sharp, B. Honig, *Biophys. Chem.* **1994**, *51,* 397; discussion 404–399.
[48] D. Sitkoff, D. J. Lockhart, K. A. Sharp, B. Honig, *Biophys. J.* **1994**, *67,* 2251.
[49] R. Luo, L. David, M. K. Gilson, *J. Comput. Chem.* **2002**, *23,* 1244.
[50] J. Kottalam, D. Case, *Biopolymers* **1990**, *29,* 1409.
[51] A. R. Leach, Molecular Modelling: Principles and Applications; Harlow: Pearson Education EMA, **2001**.
[52] R. Walker, M. Crowley, D. A. Case, *J. Comput. Chem.* **2007**, *29,* 1019.
[53] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107,* 3902.
[54] J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10,* 209.
[55] M. J. Field, M. Albe, C. Bret, F. Proust-De Martin, *J. Comput. Chem.* **2000**, *21,* 1088.
[56] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1986**, *7,* 718.
[57] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14,* 33.
[58] S. Wold, C. Albano, W. J. Dunn, U. Edlund, K. Esbenson, P. Geladi, S. Hellberg, W. Lindburg, M. Sjostrom, Chemometrics: Mathematics and Statistic in Chemistry; Dordrecht, The Netherlands: Kluwer, **1984**.
[59] R. Grzywa, E. Dyguda-Kazimierowicz, M. Sienczyk, M. Feliks, W. A. Sokalski, J. Oleksyszyn, *J. Mol. Model.* **2007**, *13,* 677.
[60] R. Kaushik, K. Merz, *J. Am. Chem. Soc.* **2004**, *126,* 1020.
[61] P. Fong, J. P. McNamara, I. H. Hillier, R. A. Bryce, *J. Chem. Inf. Model.* **2009**, *49,* 913.
[62] F. Barbault, L. R. Zhang, L. H. Zhang, B. T. Fan, *Chemom. Intell. Lab. Syst.* **2006**, *82,* 269.
[63] S. Fulle, H. Gohlke, *J. Mol. Recognit.* **2010**, *23,* 220.
[64] S. D. Morley, M. Afshar, *J. Comput-Aided. Mol. Des.* **2004**, *18,* 189.
[65] P. Pfeffer, H. Gohlke, *J. Chem. Inf. Model.* **2007**, *47,* 1868.
[66] H.-J. Böhm, M. Stahl, In Reviews in Computational Chemistry; K. B. Lipkowsky, D. B. Boyd, Eds.; New-York: Wiley, **2002**.

Received: 8 June 2011
Revised: 4 October 2011
Accepted: 5 October 2011
Published online on 13 January 2012