ORIGINAL PAPER

Correlation between biological activity and binding energy in systems of integrin with cyclic RGD-containing binders: a QM/MM molecular dynamics study

Mingli Xiang · Yuchun Lin · Gu He · Lijuan Chen · Mingli Yang · Shengyong Yang · Yirong Mo

Abstract We here report a combined quantum mechanical/ molecular mechanical (QM/MM) molecular dynamics (MD) study on the binding interactions between the $\alpha_V\beta_3$ integrin and eight cyclic arginine-glycine-aspartate (RGD) containing peptides. The initial conformation of each peptide within the binding site of the integrin was determined by docking the ligand to the reactive site of the integrin crystal structure with the aid of docking software FRED. The subsequent QM/MM MD simulations of the complex structures show that these eight cyclic RGD-peptides have a generally similar interaction mode with the binding site of the integrin to the cyclo(RGDf-N[M]V) analog found in the crystal structure. Still, there are subtle differences in the interactions of peptide ligands with the integrin, which contribute to the different inhibition activities. The averaged QM/MM protein-ligand interaction energy (IE) is remarkably correlated to the biological activity of the ligand. The IE, as well as a three-variable model which is somewhat interpretable, thus can be used to predict the bioactivity of a new ligand quantitatively, at least within a family of analogs. The present study establishes a helpful protocol for advancing lead compounds to potent inhibitors.

M. Xiang (✉) · G. He · L. Chen (✉) · S. Yang
State Key Laboratory of Biotherapy, West China Hospital,
West China Medical School, Sichuan University,
Chengdu 610041, China
e-mail: xiang_mingli@scu.edu.cn
e-mail: lijuan17@hotmail.com

Y. Lin
Department of Chemical and Biomolecular Engineering,
University of California,
Berkeley, CA 94720, USA

M. Yang
Institute of Atomic and Molecular Physics, Sichuan University,
Chengdu 610065, China

Y. Mo
Department of Chemistry, Western Michigan University,
Kalamazoo, MI 49008, USA

Introduction

Integrins consist of a large family of cell surface heterodimeric adhesion proteins responsible for a wide spectrum of cell-cell and cell-extracellular matrixes and cell-pathogen interactions [1, 2]. These proteins are non-covalently linked, heterodimeric systems composed of an α subunit and a β subunit [3–6], and present in many animal species, ranging from sponges to mammals [7]. Twenty four distinct heterodimeric integrins in mammals, which are assembled by totally 18 α subunits and eight β subunits [1, 4, 6, 8], have been identified and sequenced. Integrins act as receptors for extracellular matrix (ECM) and other cell-surface adhesins [9, 10]. Their ligands include cell surface counter receptors, ligands of the vasculature and ECM macromolecules [11]. Integrin–ligand interactions play essential roles in development, immune responses, leukocyte traffic, hemostasis, tumor genesis, tumor metastasis, and tumor angiogenesis by controlling diverse cell functions such as adhesion, shape, growth, differentiation and mobility [8, 9, 12, 13]. As integrins participate in so many important physiological activities, it is not surprising that they are at the heart of many human diseases of genetics, autoimmune, and others [1]. In fact, some of them have been considered as attractive targets for drug development [9].

Based on evolutionary relationship and ligand selectivity, integrin $\alpha\beta$ heterodimers can be grouped into several sub-families [1, 8]. Integrin $\alpha_V\beta_3$ is a particularly special and interesting member of one subfamily which has been shown to play an important role in regulating tumor angiogenesis [8, 9, 13, 14] and members of which share the ability to recognize the arginine-glycine-aspartate (RGD) sequence [1, 6, 14–18]. The specificity of $\alpha_V\beta_3$ lies in the fact that tumor cell growth and related malignant behavior depend mainly on the angiogenesis that is controlled by the endo-thelial cell $\alpha_V\beta_3$ integrin [14]. Experimental results have shown that the increased level of expression of the integrins $\alpha_V\beta_3$ is closely associated with the increased cell invasion and metastasis [8]. This integrin is directly involved in the evolution and diffusion of metastatic tumor cells and angio-genesis [2, 13] and acts as a regulator of disease pathology associated with cancer, osteoporosis and rheumatoid arthri-tis [11]. It has been found that sequences containing RGD as small as tetrapeptides can block the interaction between the $\alpha_V\beta_3$ integrin and its ligands [17]. Moreover, cyclic RGD-peptides show higher affinity for the $\alpha_V\beta_3$ receptor than linear ones [16]. Inspired by these hopeful pioneering works, a significant number of cyclic RGD peptides and non-peptidic heterocyclic mimetics of $\alpha_V\beta_3$ integrin antag-onists have been identified [2, 16, 17], although so far very few of them have progressed into clinical development [11]. Therefore, at the outset of our work, we focus on the design of novel antagonists of $\alpha_V\beta_3$ integrin by employing computer-aided drug design (CADD).

One of the upmost tasks of lead optimization is the measure of the interaction between a macromolecule and its ligands, as the affinity of a compound to its macromo-lecular receptor is closely related to the biological activity of this compound [19]. Thus, the first obstacle that should be overcome by CADD on the way to improving the biological activity of leads is to evaluate the receptor-ligand interac-tions as accurately as possible [20]. In general, CADD employs binding affinity to evaluate the interaction by means of two approaches, namely efficiency-oriented and accuracy-oriented [21]. Efficient approaches usually use simple scoring functions to gauge the macromolecule-ligand binding affinity, and have been broadly applied in the early stage of drug development for lead identification [22–26]. However, the correlation between the parameters derived from the efficient ways and experimental binding affinities is often unsatisfactory [27]. In the accuracy-oriented methods, the drawbacks inherent in scoring func-tions are tackled by employing time-consuming molecular dynamics (MD) techniques [28, 29], and consequently reli-able binding affinity can be obtained. Free energy perturba-tion (FEP), thermodynamics integration (TI) [30], molecular me-chanics/Poisson-Boltzmann surface area (MM/PBSA) [31] and linear interaction energy (LIE) [32] are the representatives

of MD-based accurate approaches. FEP and TI are probably the most rigorous and accurate approaches to calculate the receptor-ligand binding affinity [27, 28, 33, 34]. However, they are rarely used in the pharmaceutical industry because FEP and TI calculations are CPU demanding [20, 35, 36] and require the availability of validated molecular mechanics force field parameters to achieve high accuracy [35]. MM/PBSA and LIE are the simplified MD-based ways. By invoking a number of simple approximations, MM/PBSA and LIE could provide relatively good binding affinity values at a moderate computational cost [28, 33], but they have not been proven to be accurate enough to guide lead optimization [20].

Between the efficiency-oriented and accuracy-oriented approaches, however, it would be valuable to find alterna-tive routes that can be used to derive the binding affinity in a macromolecule-ligand complex with a reasonable accuracy at an acceptable cost. The combined quantum mechanical/molecular mechanical (QM/MM) hybrid method, which has primarily been proposed and developed for the study of enzyme-catalyzed reactions, [22, 26, 37] is the newly emerging way to approximate binding affinities [23, 25, 36]. It is rooted in the fundamental idea of combining the strengths of both QM (accuracy) and MM (efficiency) meth-ods to generate a powerful tool for the study of biological systems [38–42]. The combined QM/MM method is partic-ularly useful and reliable when the QM and MM regions are not covalently bound. Protein and solvent environment are in general represented by MM force fields, which are com-putationally efficient, while the ligand and several surround-ing important residues in the binding site of protein are treated at the QM level. The additional advantage of treating ligands quantum mechanically is the inclusion of ligand polarization upon binding [43, 44]. Furthermore, QM cal-culations can elegantly avoid the parameterization of force fields for ligands containing non-standard groups by explic-itly taking into account the electronic structures of whole molecules [38]. We note that often it is very difficult to parameterize certain functional groups [43]. Therefore the combined QM/MM method is supposed to be the best choice for modeling reactions [39, 45–52] in biological systems. Earlier, Alex and Finn estimated the binding affin-ity of thermolysin inhibitors based on combined QM/MM computations of static structures [53]. Later, molecular dy-namics was incorporated to get an ensemble of structures for the evaluation of averaged binding affinity. For instance, Alzate-Morales et al. performed combined QM/MM molec-ular dynamics simulations to determine the protein-ligand interaction energy between CDK2 and five inhibitors, and demonstrated that the QM/MM interaction energy is strongly correlated to the biological activity and can be used as a predictor [43]. Recently, Ciancetta et al. used the QM/MM-PBSA approach to estimate binding affinities of a series of six antitumor compounds to cathepsin B, reproducing

the antitumor activities of the complexes with a correlation coefficient of 0.35-0.86 after a conformational search [54]. Notably, Beierlein et al. developed a molecular simulation protocol by combing FEP or TI with QM/MM computations [55]. The advantage of this protocol is that electronic polarization effects are included in thermodynamically rigorous free energy calculations.

In this paper, we studied the interaction of the $\alpha_V\beta_3$ integrin with cyclic RGD-containing peptides, which are even more complex than the small molecular compounds studied ever before. The effectiveness of the proposed strategy for the set of RGD-containing ligands presented in Fig. 1 will pave a way for our further design and optimization of the $\alpha_V\beta_3$ integrin antagonists.

Computational details

Eight $\alpha_V\beta_3$ integrin-cyclopeptide complexes in total were simulated with the hybrid QM/MM method in this work. All calculations were conducted using the CHARMM suite of programs [44, 56]. Following are the details regarding the model building and computational protocols.

System preparation

The initial structure of the $\alpha_V\beta_3$ integrin was taken from the protein data bank (PDB ID:1L5G) [5], where the sugars, which are usually not involved in biological processes, were removed from the structure. Based on the heavy atom positions and the standard bond lengths and angles, hydrogen atoms were added to the protein structure using the HBUILD facility in CHARMM27 [44, 56]. In order to confirm that the 1L5G PDB file could be recognized properly by CHARMM, basic modifications had been done prior to computations and some important ones are mentioned in

the following. All the δ carbon atoms of isoleucines (77 isoleucines in total) labeled as CD1 in the original 1L5G PDB file were changed to CD. Similarly, the two oxygen atoms of the C-terminal COO group were labeled as OT1 and OT2 instead of O and OXT in the original file. The protonated states (HSD or HSE) of 19 histidine residues in the PDB file were determined manually (alternatively one can use a software such as Reduce [57]) by checking their individual micro-environments. For example, His521 and His813 in chain A and His274 in chain B were defined as HSE as a proton is best located on NE2 in the imidazole ring, whereas His91 in chain A was marked as HSD as there is a proton attached to ND1. Finally, disulfide bridges were built.

Eight cyclic RGD-containing peptides, as shown in Fig. 1, were selected from literature [2]. These RGD-containing cyclopeptides exhibit high affinity toward integrin. In the receptor high-affinity state [2], the IC_{50h} values of all these ligands are in the order of nanomolar concentrations.

Protein lL5G provides us the essential information of the integrin $\alpha_V\beta_3$ in complex with a ligand cyclo(RGDf-N [M] V) [5]. The binding site of the $\alpha_V\beta_3$ integrin is located at the interface of β-propeller domain from α_V and a βA domain from β_3 [5]. All of the eight cyclic RGD-containing peptide mimics were constructed [2] on the basis of imitating cyclo (RGDf-N[M]V). It was assumed that all of these eight mimetics have a similar interaction mode with the $\alpha_V\beta_3$ integrin to that of cyclo(RGDf-N[M]V). As such, the initial conformations of the eight complexes composed of the $\alpha_V\beta_3$ integrin and the ligands were constructed by the use of a protein-ligand docking program called fast rigid exhaustive docking, or FRED [58, 59]. To examine and verify the protocol, we docked cyclo(RGDf-N[M]V) to the integrin and compared the subsequent computational model with the known crystal structure. At first, conformations of the

Fig. 1 Chemical structures of the eight cyclic RGD-containing peptides ($\alpha_V\beta_3$ integrin binders) and their corresponding IC_{50h} values with standard errors of measurement expressed in nM (in parentheses) in the receptor high-affinity state [2]. For simplicity and consistency, the compounds are named the same as in literature [2]
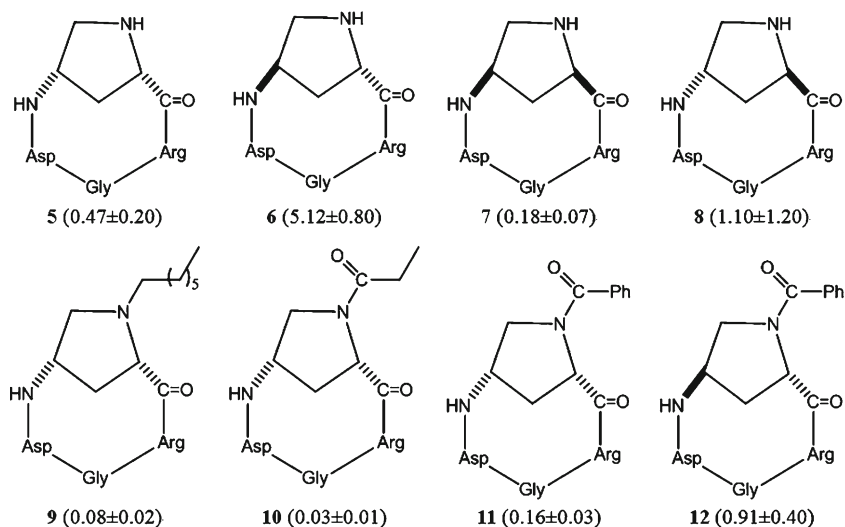
peptide cyclo(RGDf-N[M]V) were generated with OME-GA2, a tool in the FRED suite of software. The maximum number of constructed conformations is 2000. While chemical Gaussian overlay scoring function [60] was selected for conformation searching, rigid optimizing and evaluating interaction, the interaction mode of cyclo(RGDf-N[M]V) with the integrin within the binding site was reproduced, as shown in Fig. 2a. The reproduced conformation has the highest score among all conformations and it is almost identical to the crystal structure. The RMSD value which was calculated by using DS visualizer2.5, is only 0.84 Å.

Adopting the same set of docking strategy, we subsequently built the complex models of the integrin with the eight integrin ligands which were docked to the binding site. For each ligand, the conformation having the highest score was retained. In this way, the initial conformation of each cyclic RGD-containing peptide at the binding site of the integrin was obtained. As demonstrated in Fig. 2b, all eight peptide ligands interact with the integrin in a similar mode to the cyclo(RGDf-N[M]V). We note that in the subsequent QM/MM MD simulations, Mg^{2+} ions in $\alpha_V\beta_3$ integrin were modeled instead of Mn^{2+} ions due to the lack of force field parameters for the Mn^{2+} ion. This kind of replacement has been demonstrated to be practical in literature [10, 61, 62].

QM/MM setup and simulations

Each biological system was partitioned into inner regions and outer regions [41, 48]. Atoms within the inner region were treated quantum-mechanically, and referred to the QM region. In contrast, the outer region was described by a force field, and referred to the MM region. The cyclic RGD-containing peptide in the complex, which was the sole component of the QM region here, was treated quantum mechanically as a single residue. The QM atoms were

described with the semi-empirical AM1 method [63] whereas the integrin and solvent environment were treated by a classical CHARMM27 force field [44, 56].

All the complexes went through a multi-step treatment including energy minimization, solvation, heating, followed by QM/MM MD simulations. At the first step, QM/MM energy minimization (150 steps of ABNR) was performed for each system. The second step is solvation. The complex was immersed into a TIP3P [64] water sphere with a radius of 30 Å, centered at the QM residue. Water molecules beyond the range of 30 Å from the centroid and the overlapping water molecules whose oxygen atoms closer than 2.5 Å to any other heavy atoms were removed. The complex was further solvated in the same way in another water sphere whose moment of inertia was rotated in a certain degree from the previous one. These solvation procedures (sphere rotation, sphere addition and water deletion) were repeated several times to avoid solvent cavities. Afterward, water molecules were relaxed using optimization algorithms (200 steps of ABNR) followed by a QM/MM minimization of the overall solvated system (100 steps of ABNR).

At the third step, each system was slowly heated from 10 to 310 K in 50 ps interval by hybrid QM/MM velocity verlet dynamics. Finally, the complex was equilibrated by means of QM/MM leapfrog Langevin dynamics at 310 K with stochastic boundary conditions [65] for 100 ps, followed by the production run for 50 ps with data collected. In the QM/MM MD simulations, the 30 Å radius sphere was partitioned into the reaction zone (0–25 Å), buffer zone (25–28 Å) and reservoir zone (28–30 Å). Atoms in the reaction zone were propagated using Newton's equations of motions. Atoms within the buffer region were treated using the Langevin equations of motion, where friction coefficients of 200 ps^{-1} and of 62 ps^{-1} were used for non-hydrogen protein atom and water oxygen atoms, respectively.
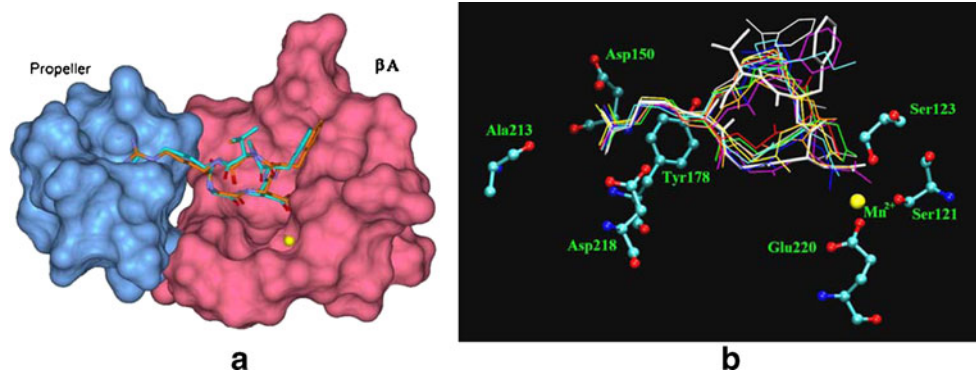

Fig. 2 Initial conformations of cyclic RGD-containing peptides within the binding site of integrin $\alpha_V\beta_3$. **a** Comparison between the crystal structure of cyclo(RGDf-N[M]V) (represented with sky blue stick) within the binding site of $\alpha_V\beta_3$ and the computational structure (depicted with orange stick) generated by FRED. A Mn^{2+} within the site is shown as a yellow ball. Ligand binding site was represented with solid surfaces. **b** Crystal structure of cyclo(RGDf-N[M]V) and the constructed initial conformations of eight cyclic RGD-containing ligands within the binding site of $\alpha_V\beta_3$ integrin(PDB ID:1L5G) [5]. Peptide coloring: white stick, cyclo(RGDf-N[M]V); red, 5; orange, 6; yellow, 7; green, 8; cyan, 9; blue, 10; purple, 11; silver, 12; yellow ball, Mn^{2+}. This figure was generated with the program VMD [76]

Amino acid residues that were more than 28 Å away from the center of the sphere were fixed during the MD simulations. To maintain the correct average distribution of water molecules and prevent the escape of any water molecule, a spherical boundary potential at 30 Å was imposed [66]. The SHAKE [67] algorithm was employed to constrain all bonds containing hydrogen atoms. The time step was set to 1 fs in molecular dynamics simulations. A non-bonded cutoff with an atom switching function in the region from 12 to 13 Å was used to smoothly scale down the interaction to zero at 14 Å.

Results and discussion

In this section the structural features of each of the integrin-inhibitor complexes were discussed, and the relationship between the interaction energies and the biological activity (measured as IC_{50h}) was established. All geometrical data and energy terms had been averaged during the production run, and the corresponding standard deviations were presented to justify the reliability of the average values. As in simulations conformation was saved every 100 steps, a 50 ps simulation in the production run resulted in 500 conformations which were used for analyses.

Interaction mode analysis

X-ray crystal diffraction experiment unveiled [5] that the RGD sequence makes the main contact area with the binding site of integrin in the complex of the $\alpha_V\beta_3$ integrin with cyclo(RGDf-N[M]V). The arginine guanidinium group of the cyclo-peptide, which inserts into a groove at the top of the propeller domain (Fig. 2a), is held in place by a bidentate salt bridge to Asp218 at the bottom of the groove and by an additional salt bridge to Asp150 at the rear (Fig. 2b). As we had assumed in the docking process that all eight mimetics have a similar interaction mode with the $\alpha_V\beta_3$ integrin to that of cyclo(RGDf-N[M]V), the arginine guanidinium group of the eight cyclo-peptides interacts with Asp218 of the propeller domain in the simulated average structures and most of the upper portion of the arginine side chain exposed to solvent (Fig. 3). Of course the strengths of interactions with Asp218 are different and vary with eight cyclo-peptides.

Compared to the crystal structure, Asp150 of the propeller domain shifts away from the position of interacting with cyclo-peptides while Tyr178 has closer contact with the guandinium group of the peptide inhibitors (Table 1). Notably in the crystal structure, the phenyl plane of Tyr178 is approximately perpendicular to the side chain of ligand Arg that is a little curved. However, in the simulated structures, the plane is almost parallel to the guardinium group and the side chain of the Arg is more stretched (Table 1 and Fig. 3).

This change, which is out of our expectation, may enhance the interaction between the ligand and the binding site by forming a cation-π interaction or probably hydrophobic interaction. Cation-π interaction makes an important contribution in protein-ligand binding process and has ever been observed in a wide range of biological systems [68–73]. Interestingly, an additional interaction was observed between the guanidinium group of potent inhibitors (except compound **5**) and carbonyl oxygen of Ala213 instead of one of the Asp150 carboxylate oxygen atoms.

It seems that the cation-π interaction of the guanidinium part of ligand Arg with Tyr178 and the additional interaction with Ala213 have a prominent contribution to the whole $\alpha_V\beta_3$ integrin-peptide binding. As shown in Fig. 3, the additional interaction and/or cation-π interaction were observed among the $\alpha_V\beta_3$ integrin-potent inhibitor complexes. In contrast, those interactions were not noticed in the complex composed of $\alpha_V\beta_3$ integrin and compound 6, the weakest one among these inhibitors.

The presence of divalent cations in protein is one of the major determinants which affect integrin –ligand interactions [4]. Based on the findings in the crystal structure [5], the ligand Asp side chain is completely buried in the βA domain and the Asp carboxylate oxygen contact a Mn^{2+} ion in the simulated structures. In the simulated ones, however, the magnesium ion Mg^{2+} is six-coordinated by oxygen atoms (Fig. 4). The hydroxyl oxygen of Ser121 and Ser123 in the βA domain and the carboxylate oxygen atom(s) of ligand Asp are always involved in the coordination with the metal ion. In complexes with ligands **5**, **7** and **9**, Ser121, Ser123 and ligand Asp provide four coordination oxygen atoms for the Mg^{2+} ion and the remaining two are supplemented by Asp119 alone. In complexes with compounds **8** and **11**, Ser121, Ser123 and ligand Asp supply three coordination oxygen atoms for the Mg^{2+} ion. For the remaining three, two of them are from Asp119 and one from Glu220. While in the averaged complex structures with ligands **6**, **10** and **12**, ligand Asp and residues within the binding site do not provide enough oxygen atoms to coordinate the magnesium ion. Averaged structures or snapshots (Fig. 4b, f and h) of the integrin-inhibitor complexes as obtained in our QM/MM MD simulations provide us additional information. In the complexes liganded by inhibitors **6** and **12**, Ser121, Ser123,Glu220 and ligand Asp contribute four atoms to coordinate the magnesium ion, apart from two bound solvent water molecules. In the simulated complex structure with the cyclic RGD-containing peptide **10**, three of six coordination oxygen atoms come from the residues including Ser121, Ser123 and Glu220. Among the remaining three, two of them are from carboxylate oxygen of the ligand Arg and one is supplied by a solvent water molecule.

Fig. 3 Interactions between the integrin $\alpha_V\beta_3$ and the guanidinium part of ligand Arg. The structures were gained by averaging 500 conformations collected in the analysis period of QM/MM MD simulations. Cation-π interaction between the guanidinium part and Tyr178 are highlighted. Ligands in (**a–h**) correspond to compounds **5**–**12**, respectively. This figure was generated with DS visualizer2.5
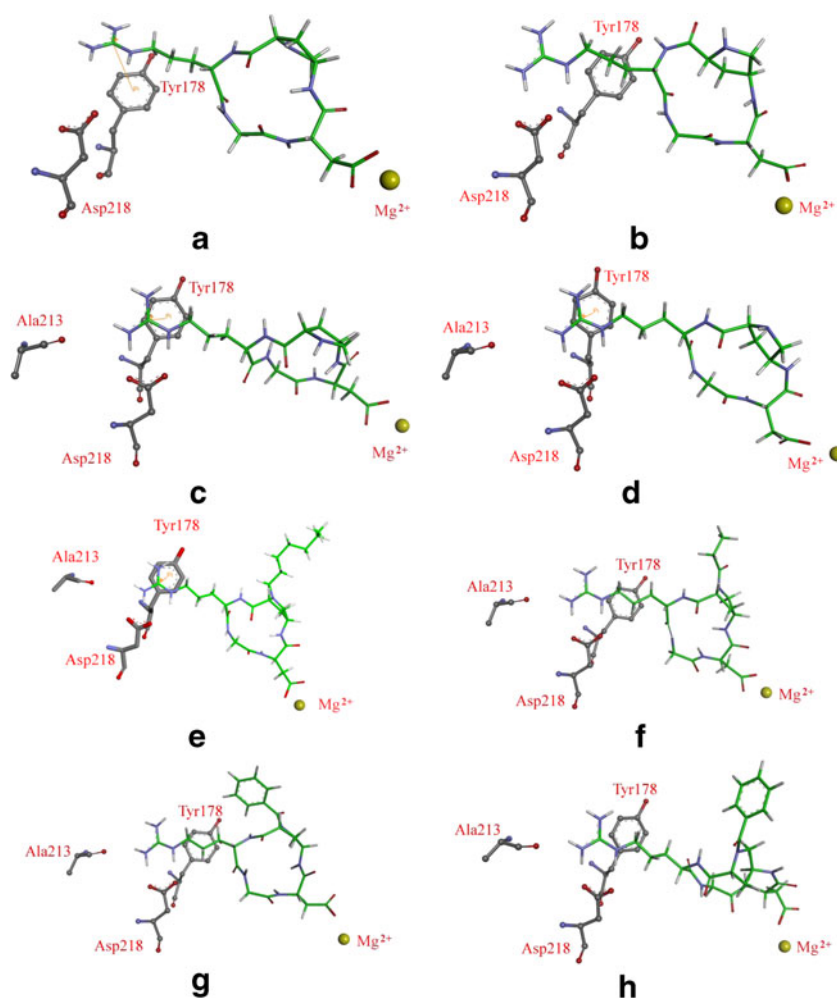
Correlation between IC$_{50h}$ and interaction energy

Averaged QM/MM protein-ligand interaction energies ($\Delta E_{QM/MM}$) for each of the inhibitors with integrin $\alpha_V\beta_3$, together with their electrostatic (ES) and van der Waals (vdW) components, are compiled in Table 2. The biological activity and solvent accessible surface area (SASA) of ligands in complexes are also provided. The IC$_{50h}$ values of potent inhibitors are approximately 1 nM or less.

The potency of a compound is largely related to its binding free energy [19]. On the basis of assumption [43] that solvation/desolvation energies, enzyme deformation energy, and entropic changes are proportional to the magnitude of the interactions between an inhibitor and a protein, the

Table 1 Distances(Å), which are gained from relevant average structures, between ligand Arg of each peptide and concerned residues in binding site with standard deviations

| Compound	cyclo(RGDf-N[M]V)	5	6	7	8	9	10	11	12
Asp218[a]	4.02	3.94 (±0.22)	3.79 (±0.17)	3.66 (±0.09)	3.82 (±0.16)	3.80 (±0.18)	3.70 (±0.09)	3.66 (±0.10)	3.66 (±0.12)
Tyr178[b]	4.17	3.79 (±0.18)	4.40 (±0.24)	3.71 (±0.14)	3.68 (±0.14)	3.65 (±0.16)	4.00 (±0.25)	4.02 (±0.21)	4.22 (±0.37)
Asp150[a]	5.53	7.53 (±0.34)	6.21 (±0.43)	6.96 (±0.24)	9.16 (±0.27)	7.82 (±0.31)	6.56 (±0.36)	6.36 (±0.40)	6.93 (±0.44)
Ala213[c]	5.21	6.26 (±0.36)	6.61 (±0.59)	5.19 (±0.27)	5.29 (±0.43)	5.12 (±0.49)	4.35 (±0.38)	4.55 (±0.27)	4.14 (±0.27)
Ligand Arg[d]	5.71	6.22 (±0.10)	6.13 (±0.13)	6.11 (±0.14)	6.10 (±0.19)	6.17 (±0.17)	6.24 (±0.07)	6.23 (±0.07)	6.10 (±0.14)

[a] distance between the centroid of charged section in the ligand guanidinium group and that of the carboxylate oxygens of binding site residue

[b] distance between the centroid of whole side chain of ligand Arg and that of the phenyl group of Tyr178

[c] distance between the carbonyl oxygen of Ala213 and the centroid of charged section in the ligand guanidinium group

[d] distance between the centroid of charged section in the ligand guanidinium group and the Cα of the ligand Arg

Fig. 4 Averaged structures or snapshots(b, f, h) which demonstrate the interactions between the integrin $\alpha_V\beta_3$ and the carboxylate part of ligand Asp, obtained from 500 conformations collected in the analysis period of the QM/MM MD simulations. The Mg^{2+} ion is illustrated with yellow ball and solvent waters are shown as red-white sticks. Ligands in (**a–h**) are the same as in Fig. 3. This figure was generated with DS visualizer2.5
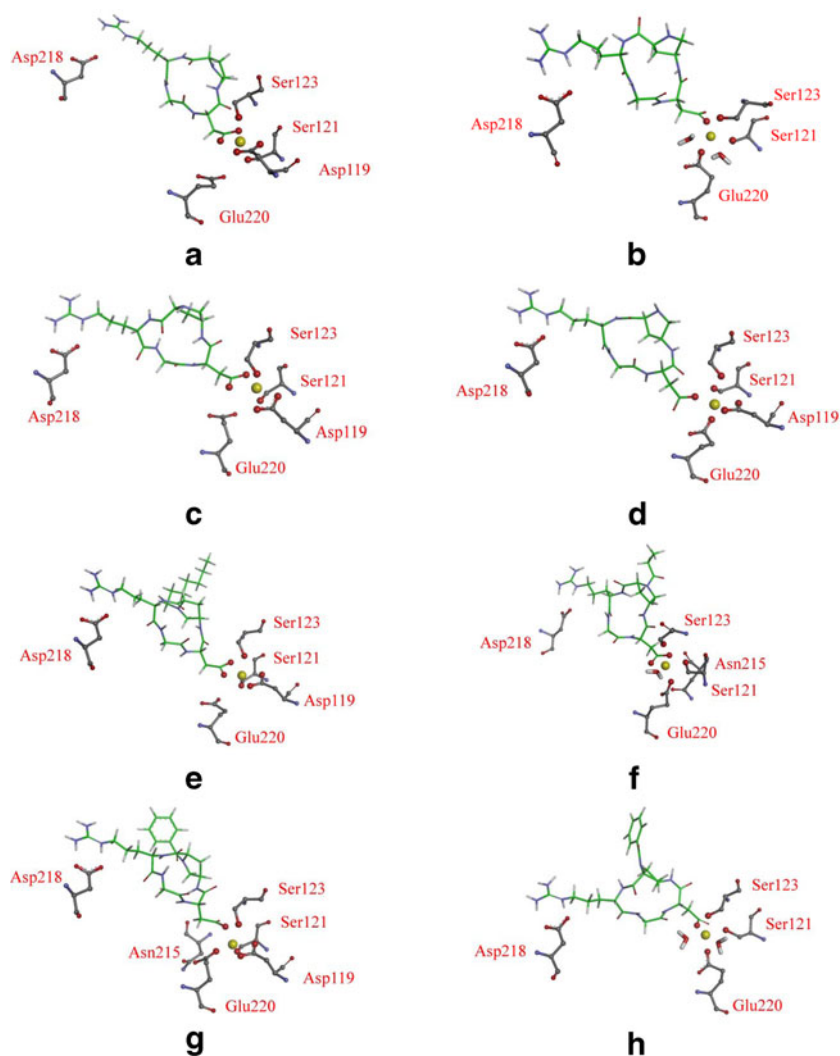
biological activity of a compound could be approximated as a linear function of the interaction energy as

$$log_{10}IC_{50h} = \alpha\Delta E_{QM/MM} + \beta, \qquad (1)$$

where α and β are undetermined constants.

The result of a least-square fit of Eq. 1 using data of Table 2 is shown in Fig. 5. The high correlation coefficient

value (R=0.894) is encouraging and endorses the direct correlation between IC_{50h} values and interaction energies. It further suggests that the biological activity of a given inhibitor can be approximately expressed with QM/MM interaction energies, at least when comparing them with a series of compounds belonging to a similar group.

Table 2 Averaged $\Delta E_{QM/MM}$ and its components (kcal/mol) for the integrin $\alpha_V\beta_3$ inhibitors studied with standard deviations

Property	5	6	7	8	9	10	11	12
$\Delta E_{QM/MM}$	−531.28 (±13.42)[a]	−490.93 (±18.68)	−528.69 (±14.91)	−523.38 (±14.66)	−556.75 (±14.61)	−616.33 (±17.77)	−535.66 (±16.43)	−517.67 (±15.44)
ΔE_{ES}	−490.89 (±13.72)	−449.00 (±19.04)	−491.65 (±15.76)	−489.81 (±15.50)	−504.89 (±14.77)	−573.06 (±18.52)	−484.09 (±16.45)	−468.49 (±15.88)
ΔE_{vdW}	−40.39 (±4.44)	−41.93 (±4.33)	−37.04 (±4.84)	−33.57 (±5.13)	−51.86 (±4.89)	−43.27 (±4.59)	−51.57 (±4.76)	−49.18 (±4.33)
$log_{10}IC_{50h}$	−9.33	−8.29	−9.74	−8.96	−10.10	−10.52	−9.80	−9.04
SASA[b]	211.02 (±7.82)	266.22 (±9.92)	230.81 (±8.99)	240.53 (±8.48)	315.48 (±13.45)	338.42 (±12.29)	274.17 (±13.51)	401.03 (±12.11)

[a] Values in parentheses represent the standard deviations

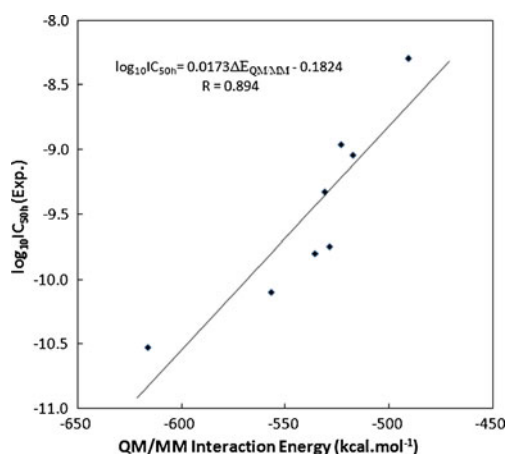[b] Solvent accessible surface area, unit in $Å^{-2}$

Fig. 5 Correlation between the experimentally determined $\log_{10}IC_{50h}$ and the calculated QM/MM interaction energy (kcal mol^{-1}) for the eight cyclic RGD-containing peptides

It can be seen from Table 2 that electrostatic energy (ΔE_{ES}) dominates the overall interaction energy, whereas the van der Waals interactions (ΔE_{vdW}) account for less than 10 % of the interaction energy. The binding site of the integrin $\alpha_V\beta_3$ is an open pocket. Part of a ligand directly contacts the site and part of it exposes to solvent. Obviously, the difference of bioactivities among compounds cannot be fully expressed by the total interaction energy alone. There must be some other independent factors which will affect the biological property. We speculated that the SASA of ligands in binding site, which is related to the non-polar contribution to solvation [74, 75], may be one of the key factors and it was not directly considered in the present QM/MM interaction energy calculation. The SASAs of the eight ligands in the integrin binding site, which are listed in Table 2, were calculated by executing a command file of CHARMM [44, 56]. It is a 'COOR SURFACE' command based file. In the process of calculating SASA, 500 collected conformations were used. Ultimately, a three-variable model was setup (Eq. 2) and the undetermined parameters were solved by employing toolkit solver in Microsoft Office Excel 2007 so as to express the bioactivity of the ligand in terms of both the interaction energy and the SASA:

$$\log_{10}IC_{50h} = \alpha\Delta E_{ES} + \beta\Delta E_{vdW} + \gamma\,SASA$$
$$= 1.659 \times 10^{-2}\Delta E_{ES} + 4.972 \times 10^{-3}\Delta E_{vdW} - 3.125 \times 10^{-3}SASA \quad (2)$$

Equation 2 can be used to predict the IC_{50h} values, which are quoted in Table 3. The experimentally determined IC_{50h}

values were compared with the predicted ones by plotting experimental values against the calculated. As shown in Fig. 6, the comparison yields a rather good correlation coefficient R=0.936.

We also tried to correlate the IC_{50h} values with descriptors like ΔE_{ES}, ΔE_{vdW} or SASA individually, but the correlation coefficients obtained were significantly worse. There seems no direct correlation between the bioactivities and vdW energies among these peptides. It further verifies that the inhibition behavior of these peptides within the binding site of the $\alpha_V\beta_3$ integrin must result from the synergic contribution of a few factors. Each element has its important role in whole interaction between the peptide ligand and the residues at the binding site. It is not the individual descriptors but a multiple regression model with respect to three terms (Eq. 2) or the whole QM/MM interaction energy (Eq. 1) which can be used to predict the IC_{50h} for a new inhibitor quantitatively.

Conclusions

We have conducted hybrid QM/MM MD simulations on integrin $\alpha_V\beta_3$ complexed with eight cyclic RGD-containing inhibitors. The common features and several obvious differences of interactions between the peptide inhibitors and the binding site of integrin $\alpha_V\beta_3$ have been revealed. In agreement with the analysis of the crystal structure of integrin $\alpha_V\beta_3$ complexed with cyclo(RGDf-N[M]V), simulations showed that the arginine guanidinium group of eight cyclic RGD-containing peptide mimics interacts with the carboxylate oxygen(s) of Asp218 in the propeller domain. Tyr178 in the propeller domain is close to the arginine guanidinium group of the peptide binders and its phenyl plane is almost parallel to the stretched guanidinium chain to enhance the protein-ligand binding by cation-π interaction or hydrophobic interaction, while Asp150 in the same domain moves away from the guanidium group. On the other hand, for most potent inhibitors except compound **5**, the carbonyl oxygen atom of Ala213 approaches the guanidium group of ligand Arg. Only in complex composed of $\alpha_V\beta_3$ integrin and compound **6**, the weakest one among these inhibitors, neither cation-π interaction of guanidium group with Tyr178 nor the interaction with Ala213 was observed.

Table 3 Results of multiple regression analysis (see Eq. 1) for the compounds studied

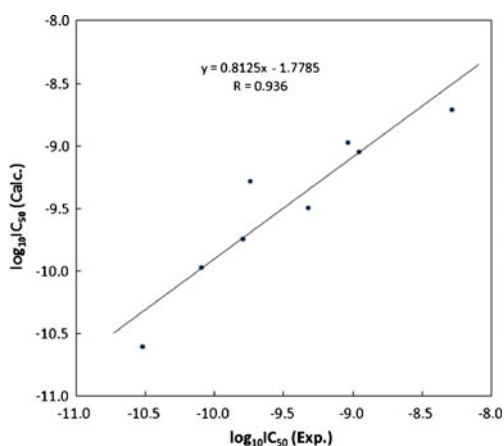| Compound	5	6	7	8	9	10	11	12
Experimental IC_{50h}(nM)	0.47	5.12	0.18	1.1	0.08	0.03	0.16	0.91
Experimental $\log_{10}IC_{50h}$ values	−9.33	−8.29	−9.74	−8.96	−10.10	−10.52	−9.80	−9.04
Calculated IC_{50h}(nM)	0.32	1.98	0.53	0.90	0.11	0.02	0.18	1.08
Calculated $\log_{10}IC_{50h}$ values	−9.49	−8.70	−9.28	−9.05	−9.97	−10.60	−9.74	−8.97

Fig. 6 Graph of calculated $\log_{10}IC_{50h}$ versus experimental $\log_{10}IC_{50h}$ for the eight RGD cyclo-peptides studied and equation obtained by a least-square fit

Similar to the observation in the X-ray structure, the side chain of ligand Asp, one of the most important residues in RGD sequences recognized by integrin $\alpha_V\beta_3$, is completely buried in the βA domain. As observed in the analysis of individual interactions, the influence of Ser121 and Ser123 in the βA domain seems to be crucial as well. Together with the carboxylate oxygen of ligand Asp, they participate in the coordination with the Mg^{2+} ion. In different interaction pairs, Asp119, Glu220 and solvent water molecule(s) are also involved in the coordination.

Apart from the observed features of interaction mode, averaged QM/MM interaction energies provide us useful information. There is a good correlation between the calculated IC_{50h} values and the experimental ones. For a new inhibitor, the IC_{50h} can thus be predicted from the whole interaction energy or three-variable model quantitatively. Furthermore, the three-variable model is somewhat interpretable.

It could therefore be concluded that QM/MM MD simulation is a promising approach which can be used to derive the binding affinity in a macromolecule-ligand complex with a reasonable accuracy at an acceptable cost and the present study establishes a helpful protocol for advancing lead compounds to potent inhibitors. The interaction energy, as well as a three-variable interpretable model, could be employed to predict the bioactivity of a new ligand quantitatively, at least within a family of analogs. Although it seems impractical to use the averaged QM/MM interaction energy as a scoring function in virtual screening, QM/MM MD simulation could be a powerful tool of improving the identified lead compounds. Indeed, we hope to further improve the predictive power of the model in our future work and envision that the accurate QM/MM MD based IE or multiple-variable interpretable model will routinely be utilized in the process of lead optimization, a critical and challenging part of drug discovery.

References

1. Hynes RO (2002) Integrins: bidirectional, allosteric signaling machines. Cell 110:673–687
2. Zanardi F, Burreddu P, Rassu G, Auzzas L, Battistini L, Curti C et al. (2008) Discovery of subnanomolar arginine-glycine-aspartate-based $\alpha_V\beta_3/\alpha_V\beta_5$ integrin binders embedding 4-aminoproline residues. J Med Chem 51(6):1771–1782
3. Xiong JP, Stehle T, Diefenbach B, Zhang R, Dunker R, Scott DL et al (2001) Crystal structure of the extracellular segment of integrin $\alpha_V\beta_3$. Science 294(5541):339–345
4. Arnaout MA, Goodman SL, Xiong JP (2002) Coming to grips with integrin binding to ligands. Current Opin Cell Biol 14(5):641–652
5. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, Goodman SL et al. (2002) Crystal structure of the extracellular segment of integrin $\alpha_V\beta_3$ in complex with an Arg-Gly-Asp ligand. Science 296:151–155
6. Humphries JD, Byron A, Humphries MJ (2006) Integrin ligands at a glance. J Cell Sci 119:3901–3903
7. Spitaleri A, Mari S, Curnis F, Traversari C, Longhi R, Bordignon C et al. (2008) Structural basis for the interaction of isoDGR with the RGD-binding site of $\alpha_v\beta_3$ integrin. J Biol Chem 283(28):19757–19768
8. Jin H, Varner J (2004) Integrins: roles in cancer development and as treatment targets. Brit J Cancer 90(3):561–565
9. Bella J, Humphries MJ (2005) $C\alpha$ -H$\cdots$O = C hydrogen bonds contribute to the specificity of RGD cell-adhesion interactions. BMC Struct Biol 5:4
10. Paradise RK, Lauffenburger DA, van Vliet KJ (2011) Acidic extracellular pH promotes activation of integrin $\alpha_v\beta_3$. PloS one 6(1):e15746
11. Elliot D, Henshaw E, MacFaul PA, Morley AD, Newham P, Oldham K et al. (2009) Novel inhibitors of the $\alpha_v\beta_3$ integrin-lead identification strategy. Bioorg Med Chem Lett 19:4832–4835
12. Sukopp M, Marinelli L, Heller M, Brandl T, Goodman SL, Hoffmann RW et al (2002) Designed beta turn mimic based on the allylic strain concept: evaluation of structural and biological features by incorporation into a cyclic RGD peptide (Cyclo(-L-arginylglycyl-L- α-aspartyl-)). Helv Chim Acta 85 (12):4442–4452
13. Haubner R, Weber WA, Beer AJ, Vabuliene E, Reim D, Sarbia M et al (2005) Noninvasive visualization of the activated $\alpha_v\beta_3$ integrin in cancer patients by positron emission tomography and [^{18}F] galacto-RGD. PLoS Med 2(3):e70
14. Varner JA, Cheresh DA (1996) Integrins and cancer. Curr Opin Cell Biol 8(5):724–730
15. Takagi J (2004) Structural basis for ligand recognition by RGD (Arg-Gly-Asp)-dependent integrins. Biochem Soc T 32(Pt3):403–406
16. Locardi E, Mullen DG, Mattern RH, Goodman M (1999) Conformations and pharmacophores of cyclic RGD containing peptides which selectively bind integrin $\alpha_V\beta_3$. J Pept Sci 5(11):491–506
17. Tucker GC (2006) Integrins: molecular targets in cancer therapy. Curr oncol rep 8(2):96–103
18. Belvisi L, Bernardi A, Colombo M, Manzoni L, Potenza D, Scolastico C et al. (2006) Targeting integrins: insights into structure and activity of cyclic RGD pentapeptide mimics containing azabicycloalkane amino acids. Bioorg Med Chem 14(1):169–180
19. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl 41 (15):2644–2676

20. Jorgensen WL (2009) Efficient drug lead discovery and optimization. Acc Chem Res 42(6):724–733
21. Xiang M, Cao Y, Fan W, Chen L, Mo Y (2012) Computer-aided drug design: lead discovery and optimization. Comb Chem High Throughput Screen 15(4):328–337
22. Schneidman-Duhovny D, Nussinov R, Wolfson HJ (2004) Predicting molecular interactions in silico: II. Protein-protein and protein-drug docking. Curr Med Chem 11(1):91–107
23. Shaikh SA, Jain T, Sandhu G, Latha N, Jayaram B (2007) From drug target to leads–sketching a physicochemical pathway for lead molecule design in silico. Curr Pharm Des 13(34):3454–3470
24. Zhou Z, Felts A, Friesner R, Levy R (2007) Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets. J Chem Inf Model 47(4):1599–1608
25. Adane L, Bharatam PV (2008) Modelling and informatics in the analysis of P. falciparum DHFR enzyme inhibitors. Curr Med Chem 15(16):1552–1569
26. Song CM, Lim SJ, Tong JC (2009) Recent advances in computer-aided drug design. Brief Bioinform 10(5):579–591
27. Mohan V, Gibbs AC, Cummings MD, Jaeger EP, DesJarlais RL (2005) Docking: successes and challenges. Curr Pharm Des 11 (3):323–333
28. Alonso H, Bliznyuk AA, Gready JE (2006) Combining docking and molecular dynamic simulations in drug design. Med Res Rev 26(5):531–568
29. Takeuchi H, Okazaki K (1990) Molecular dynamics simulation of diffusion of simple gas molecules in a short chain polymer. J Chem Phys 92(9):5643–5652
30. Kollman PA (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 93:2395–2417
31. Swanson JM, Henchman RH, McCammon JA (2004) Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys J 86:67–74
32. Aquist J, Marelius J (2001) The linear interaction energy method for predicting ligand binding free energies. Comb Chem High Throughput Screen 4:613–626
33. Tan JJ, Cong XJ, Hu LM, Wang CX, Jia L, Liang XJ (2010) Therapeutic strategies underpinning the development of novel techniques for the treatment of HIV infection. Drug Discov Today 15(5–6):186–197
34. Reddy MR, Singh UC, Erion MD (2011) Use of a QM/MM-based FEP method to evaluate the anomalous hydration behavior of simple alkyl amines and amides: application to the design of FBPase inhibitors for the treatment of type-2 diabetes. J Am Chem Soc 133:8059–8061
35. Reddy MR, Erion MD (2007) Relative binding affinities of fructose-1, 6-bisphosphatase inhibitors calculated using a quantum mechanics-based free energy perturbation method. J Am Chem Soc 129(30):9296–9297
36. Menikarachchi LC, Gascón JA (2010) QM/MM approaches in medicinal chemistry research. Curr Top Med Chem 10:46–54
37. Warshel A, Levitt M (1976) Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103:227–249
38. Spiegela K, Magistrato A (2006) Modeling anticancer drug–DNA interactions via mixed QM/MM molecular dynamics simulations. Org Biomol Chem 4:2507–2517
39. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. Angew Chem Int Ed Engl 48:1198–1229
40. Zhou T, Huang D, Caflisch A (2010) Quantum mechanical methods for drug design. Curr Top Med Chem 10:33–45
41. Xenides D, Randolf BR, Rode BM (2005) Structure and ultrafast dynamics of liquid water: a quantum mechanics/molecular mechanics molecular dynamics simulations study. J Chem Phys 122(17):174506
42. Alves CN, Marti S, Castillo R, Andres J, Moliner V, Tunon I et al. (2007) Calculation of binding energy using BLYP/MM for the HIV-1 integrase complexed with the S-1360 and two analogues. Bioorg Med Chem 15(11):3818–3824
43. Alzate-Morales JH, Contreras R, Soriano A, Tunon I, Silla E (2007) A computational study of the protein-ligand interactions in CDK2 inhibitors: using quantum mechanics/molecular mechanics interaction energy as a predictor of the biological activity. Biophys J 92(2):430–439
44. Brooks BR, Brooks CL III, Mackerell A Jr, Nilsson L, Petrella R, Roux B et al. (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30(10):1545–1614
45. Martin FP-D, Dumas R, Field MJ (2000) A hybrid-potential free-energy study of the isomerization step of the acetohydroxy acid isomeroreductase reaction. J Am Chem Soc 122(32):7688–7697
46. Gleeson MP, Hillier IH, Burton NA (2004) Theoretical analysis of peptidyl α-ketoheterocyclic inhibitors of human neutrophil elastase: insight into the mechanism of inhibition and the application of QM/MM calculations in structure-based drug design. Org Biomol Chem 2(16):2275–2280
47. Cui Q, Li G, Ma J, Karplus M (2004) A normal mode analysis of structural plasticity in the biomolecular motor F1-ATPase. J Mol Biol 340(2):345–372
48. Lin Y, Cao Z, Mo Y (2006) Molecular dynamics simulations on the Escherichia coli ammonia channel protein AmtB: mechanism of ammonia/ammonium transport. J Am Chem Soc 128 (33):10876–10884
49. Rowley CN, Woo TK (2007) Generation of initial trajectories for transition path sampling of chemical reactions with ab initio molecular dynamics. J Chem Phys 126:024110
50. Cheng Y, Cheng X, Radic Z, McCammon JA (2007) Acetylcholinesterase: mechanisms of covalent inhibition of wild-type and H447I mutant determined by computational analyses. J Am Chem Soc 129(20):6562–6570
51. Hu H, Lu Z, Parks JM, Burger SK, Yang W (2008) Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: sequential sampling and optimization on the potential of mean force surface. J Chem Phys 128(3):034105
52. Rowley CN, Woo TK (2011) Counteranion effects on the zirconocene polymerization catalyst olefin complex from QM/MM molecular dynamics simulations. Organometallics 30:2071–2074
53. Alex A, Finn P (1997) Fast and accurate predictions of relative binding energies. J Mol Struct THEOCHEM 398:551–554
54. Ciancetta A, Genheden S, Ryde U (2011) A QM/MM study of the binding of RAPTA ligands to cathepsin B. J Comput Aided Mol Des 25(8):729–742
55. Beierlein FR, Michel J, Essex JW (2011) A simple QM/MM approach for capturing polarization effects in protein − ligand binding free energy calculations. J Phys Chem B 115:4911–4926
56. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4(2):187–217
57. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of sidechain amide orientation. J Mol Biol 285:1735–1747
58. McGann M, Almond H, Nicholls A, Grant J, Brown F (2003) Gaussian docking functions. Biopolymers 68(1):76–90
59. McGaughey G, Sheridan R, Bayly C, Culberson J, Kreatsoulas C, Lindsley S et al. (2007) Comparison of topological, shape, and docking methods in virtual screening. J Chem Inf Model 47 (4):1504–1519

60. McGann M (2011) FRED pose prediction and virtual screening accuracy. J Chem Inf Model 51:578–596

61. Bártová I, Koča J, Otyepka M (2008) Regulatory phosphorylation of cyclin-dependent kinase 2: insights from molecular dynamics simulations. J Mol Model 14(8):761–768

62. Wells GA, Müller IB, Wrenger C, Louw AI (2009) The activity of Plasmodium falciparum arginase is mediated by a novel inter-monomer salt-bridge between Glu295-Arg404. FEBS J 276 (13):3517–3530

63. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc 107(13):3902–3909

64. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935

65. Brunger A, Brooks CL III, Karplus M (1985) Active site dynamics of ribonuclease. Proc Natl Acad Sci USA 82(24):8458–8462

66. Brooks CL III, Karplus M (1983) Deformable stochastic boundaries in molecular dynamics. J Chem Phys 79:6312–6325

67. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341

68. Fernández-Recio J, Romero A, Sancho J (1999) Energetics of a hydrogen bond (charged and neutral) and of a cation-π interaction in apoflavodoxin1. J Mol Biol 290(1):319–330

69. Wintjens R, Liévin J, Rooman M, Buisine E (2000) Contribution of cation-π interactions to the stability of protein-DNA complexes. J Mol Biol 302(2):395–410

70. Zacharias N, Dougherty DA (2002) Cation-π interactions in ligand recognition and catalysis. Trends Pharmacol Sci 23 (6):281–287

71. Lummis SCR, Beene DL, Harrison NJ, Lester HA, Dougherty DA (2005) A cation-π binding interaction with a tyrosine in the binding site of the GABAC receptor. Chem Biol 12(9):993–997

72. Xiu X, Puskar NL, Shanata JAP, Lester HA, Dougherty DA (2009) Nicotine binding to brain receptors requires a strong cation–π interaction. Nature 458(7237):534–537

73. Tantry S, Ding FX, Dumont M, Becker JM, Naider F (2010) Binding of fluorinated phenylalanine α-factor analogues to ste2p: Evidence for a cation-π binding interaction between a peptide ligand and its cognate G protein-coupled receptor. Biochemistry 49(24):5007–5015

74. Zhou Z, Madura JD (2004) Relative free energy of binding and binding mode calculations of HIV–1 RT inhibitors based on dock–MM–PB/GS. Proteins Struct Funct Bioinf 57(3):493–503

75. Bonnet P, Bryce RA (2004) Molecular dynamics and free energy analysis of neuraminidase–ligand interactions. Protein Sci 13 (4):946–957

76. Humphrey W, Dalke A, Schulten K (1996) VMD - visual molecular dynamics. J Mol Graph 14(1):33–38