



Study of SQ109 analogs binding to mycobacterium MmpL3 transporter using MD simulations and alchemical relative binding free energy calculations

Marianna Stampolaki^{1,3} · Ioannis Stylianakis¹ · Helen I. Zgurskaya² · Antonios Kolocouris¹

Received: 24 December 2022 / Accepted: 3 April 2023 / Published online: 2 May 2023
© The Author(s) 2023

Abstract

N-geranyl-*N'*-(2-adamantyl)ethane-1,2-diamine (SQ109) is a tuberculosis drug that has high potency against *Mycobacterium tuberculosis* (*Mtb*) and may function by blocking cell wall biosynthesis. After the crystal structure of MmpL3 from *Mycobacterium smegmatis* in complex with SQ109 became available, it was suggested that SQ109 inhibits MmpL3 mycolic acid transporter. Here, we showed using molecular dynamics (MD) simulations that the binding profile of nine SQ109 analogs with inhibitory potency against *Mtb* and alkyl or aryl adducts at C-2 or C-1 adamantyl carbon to MmpL3 was consistent with the X-ray structure of MmpL3 – SQ109 complex. We showed that rotation of SQ109 around carbon–carbon bond in the monoprotonated ethylenediamine unit favors two *gauche* conformations as minima in water and lipophilic solvent using DFT calculations as well as inside the transporter's binding area using MD simulations. The binding assays in micelles suggested that the binding affinity of the SQ109 analogs was increased for the larger, more hydrophobic adducts, which was consistent with our results from MD simulations of the SQ109 analogues suggesting that sizeable C-2 adamantyl adducts of SQ109 can fill a lipophilic region between Y257, Y646, F260 and F649 in MmpL3. This was confirmed quantitatively by our calculations of the relative binding free energies using the thermodynamic integration coupled with MD simulations method with a mean assigned error of 0.74 kcal mol⁻¹ compared to the experimental values.

Keywords SQ109 · MD simulations · Alchemical relative binding free energy · MM-GBSA · MmpL3 · Binding affinities

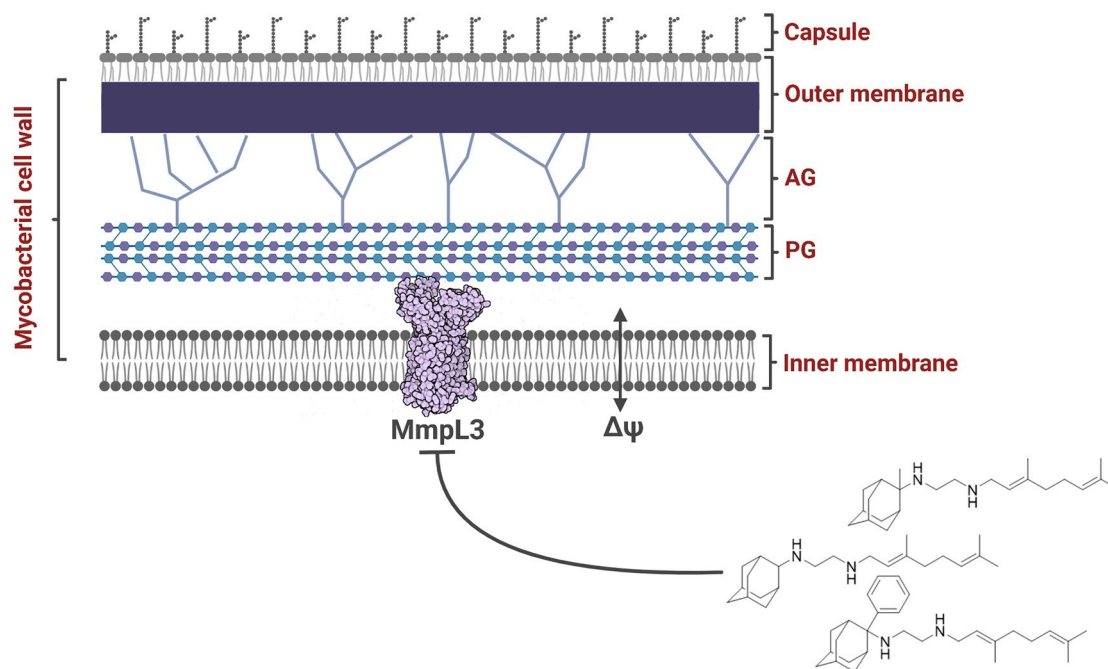
✉ Antonios Kolocouris
ankol@pharm.uoa.gr

¹ Laboratory of Medicinal Chemistry, Section of Pharmaceutical Chemistry, Department of Pharmacy, National and Kapodistrian University of Athens, Panepistimiopolis-Zografou, 15771 Athens, Greece

² Department of Chemistry and Biochemistry, University of Oklahoma, Stephenson Life Sciences Research Center, 101 Stephenson Parkway, Norman, OK 73019-5251, USA

³ Department of NMR-Based Structural Biology, Max Planck Institute for Multidisciplinary Sciences, Am Faßberg 11, 37077 Göttingen, Germany

Graphical abstract



Abbreviations

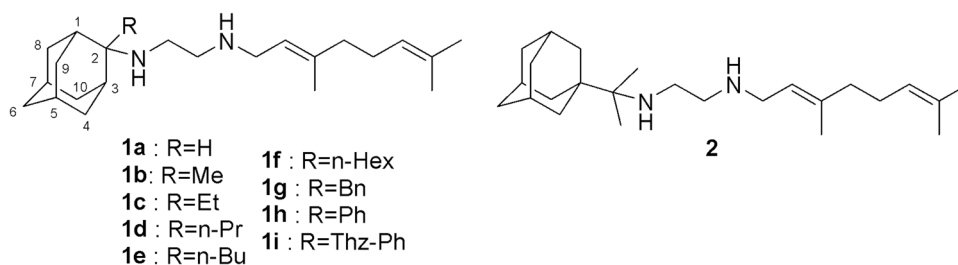
cryo-EM	Cryo-electron microscopy
DFT	Density Functional Theory
GAFF	Generalized Amber Force Field
MD	Molecular dynamics
TB	Tuberculosis
MM-GBSA	Molecular mechanics generalized born surface area
Mtb	<i>Mycobacterium tuberculosis</i>
Ms	<i>Mycobacterium smegmatis</i>
MtHN878	<i>M. tuberculosis</i> HN878
OPM	Orientations of Proteins in Membranes
POPC	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
RMSD	Root mean square deviation
TI/MD	Thermodynamic Integration coupled with MD simulations
TMM	Trehalose monomycolate
PMF	Proton motive force
RMSD	Root-mean-square deviation
SID	Simulation interaction diagram

Introduction

In 2020, an estimated 1.9 million people died from *tuberculosis* (TB), the leading cause of death among carriers of HIV [1]. The *N*-geranyl-*N'*-(2-adamantyl)ethane-1,2-diamine SQ109 (**1a**), [2] shown in Fig. 1, is a second generation ethylenediamine drug, after ethambutol against *Mycobacterium tuberculosis* (*Mtb*). Indeed, SQ109 (**1a**) has been in phase II clinical trials [3, 4] and shows high potency against drug resistant *Mtb* [5–7]. The importance of SQ109 (**1a**) as a highly potent therapeutic agent, triggered the synthesis of analogs [8–13] aiming at the improvement of drug potency, the spectrum of biological activity and the pharmacokinetic properties.

Ethambutol has been suggested to inhibit *Mtb* by binding to the membrane-embedded Emb proteins, EmbB and EmbC, involved in arabinan biosynthesis, [14] while SQ109 (**1a**) has been suggested that targets the trehalose monomycolate transporter, Mycobacterial membrane protein Large 3 (MmpL3), [4, 5] and like ethambutol, inhibits cell wall biosynthesis. MmpL3 is a membrane protein, essential for the translocation of mycolic acids in the form of trehalose monomycolates from their production site in the cytoplasm to the periplasmic space, where mycolic acids can be used in assembly of the *Mtb* outer

Fig. 1 Chemical structures of SQ109 (**1a**) and other ethylenediamine analogs **1b–i** or **2** with adducts at C-1 or C-2 adamantyl carbon, respectively, studied in the present work (Thz-Ph: 5-phenylthiazol-2-yl)



membrane [15, 16]. Its transporter activity is driven by a proton motive force (PMF) to describe that coupled with the movement of substrates toward the periplasm, protons flow into the cytoplasm to energize this translocation process. Two pairs of D-Y (D256-Y646 and D645-Y257) allow such proton translocation and these D-Y pairs are a conserved feature of the MmpL family of transporters [10, 16–18]. In 2019, the X-ray structure (PDB ID 6AJG [17]) of MmpL3 from *Mycobacterium smegmatis* (*Ms*) in complex with SQ109 (**1a**) or other *Mtb* inhibitors found inside the transporter's pore as well as the X-ray structure of the MmpL3 – 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (POPE) complex (PDB ID 6OR2) [19] became available. Afterwards, structures of the apo-protein [20, 21] or in complex with additional ligands [21, 22] were also reported, using cryo-electron microscopy (cryo-EM) or X-ray crystallography.

The ability of SQ109 (**1a**) to inhibit the function of MmpL3 can be explained on the basis of its direct binding to MmpL3 and disrupting transporter's proton translocation [17, 23] or by uncoupling activity on the PMF through another mechanism [5, 7, 24, 25] which is not specific to *Mycobacteria* [26]. The latter mechanism of action may be consistent with the broad-spectrum of activity [4] of SQ109 (**1a**) against pathogens that lack MmpL3 [27].

Using multiscale thermophoresis (MST) in native cell membrane nanoparticle environment, a dissociation constant (K_d) of $\sim 1.6 \mu\text{M}$ for SQ109 (**1a**) has been reported which is consistent with low μM potency of SQ109 (**1a**) against *Mtb*, while is much smaller than the ~ 0.4 – 1.2 mM measured using surface plasmon resonance (SPR) in micelles [17, 28]. In a previous paper [13] we synthesized the SQ109 (**1a**) analogs **1b–i**, **2** some of which exhibited low μM biological potency against *Mtb* and had mid-micromolar binding affinity to MmpL3 in micelles measured using SPR. While measurement of binding affinity in membranes with MST are more consistent with biological activities of SQ109 (**1a**) and its analogs against *Mtb* relative binding free energies even measured with SPR might provide useful values to compare them with calculated relative binding free energies as model that can support binding of **1b–i**, **2** in the same binding site as SQ109 (**1a**) in MmpL3.

Thus, here, we explored the binding profile of our synthesized SQ109 (**1a**) analogs **1b–g**, **2** to MmpL3 by applying: (a) molecular dynamics (MD) simulations ($\sim 7.2 \mu\text{s}$ total simulation time) using the X-ray structure of SQ109 (**1a**) with MmpL3 (PDB ID 6AJG [17]) as reference structure; (b) binding free energy calculations with the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) method; [29–31] (c) alchemical relative binding free energies calculations using the thermodynamic integration combined with MD simulations (TI/MD) method. [32–35] We next compared if the calculated relative binding free energies agreed with experimental relative binding free energies measured previously with SPR against *Mtb* MmpL3 (MtMmpL3) [13].

We explored the conformational properties of SQ109 (**1a**) in solution using Density Functional Theory (DFT) calculations to compare the conformational properties of its analogs **1b–i**, **2** in solution and in bound state with MmpL3 described by docking calculations and MD simulations. Thus, we calculated the free energies of the conformational minima of SQ109 (**1a**) by rotation around bonds that involve the ethylenediamine unit (three dihedral angles) in hydrophilic environment or hydrophobic environment using Density Functional Theory (DFT). The DFT calculations showed that in both hydrophilic or hydrophobic environments and inside the transporter's pore as showed from MD simulations the central ethylenediamine carbon–carbon bond in favored a *gauche* conformation.

Our MD simulations suggested that alkyl or aryl substituents at the adamantyl C-2 of SQ109 (**1a**) can fill the region between Y257, Y646, F260 and F649 in MmpL3 of *Ms* or *Mtb*. The TI/MD calculations of relative binding free energies were consistent with the experimental relative binding free energies measured using SPR and showing that binding strength is increased by increasing the size of the C-2 adamantyl adduct. Overall, the MD simulations combined with TI/MD calculation results suggested that MmpL3 can likely form stable complexes with ligands **1b–i**, **2** which bind MmpL3 at the same site with SQ109 (**1a**).

Results

Conformational analysis of monoprotonated ethylenediamine unit in SQ109 (**1a**)

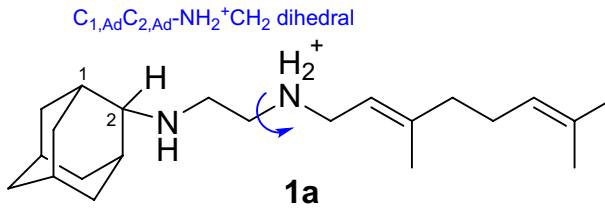
The ethylenediamine unit in SQ109 (**1a**) involves three dihedrals affecting the drug's orientation inside the MmpL3 pore through rotation around (2-Ad)NHCH₂–CH₂NH₂⁺Ger, (2-Ad)NHCH₂CH₂–NH₂⁺Ger and C_{1,Ad}C_{2,Ad}–NHCH₂CH₂NH₂⁺Ger bonds (where C_{1,Ad} or C_{2,Ad} are adamantyl carbons at position-1, or -2 in 2-adamantyl group) described in Tables 1–3. Using the B3LYP functional with dispersion interactions correction (B3LYP-D3) [36–39] and the 6-31G(d,p) basis set calculations we performed full geometry optimization and calculated the free energies of the conformational minima of SQ109 (**1a**), by manual rotation around these three dihedrals in a hydrophilic and a hydrophobic environment. Dispersion correction improves the calculation of the forces acting on the atoms in distances > 3 Å and the accuracy of relative conformational energies calculation which are shown in Table 1. The hydrophilic environment was simulated with an implicit water environment and a dielectric constant (ϵ) = 80 and the hydrophobic environment was simulated with an implicit chloroform environment and ϵ = 4.8 using the polarizable continuum model (PCM) [41] and taking advantage of a smooth switching function [42].

The rotation around the carbon–carbon bond, (2-Ad)NHCH₂–CH₂NH₂⁺Ger, in SQ109 (**1a**), (Fig. 2A) generated the *gauche*(–), *gauche*(+), *anti* and *eclipsed* conformations and changed the relative orientation of the (2-Ad)NH and NH₂⁺Ger groups and the hydrogen bonding profile

of the ethylenediamine unit inside the MmpL3 receptor. The *gauche* conformations had the lowest free energy following the *anti* conformation, which was seriously destabilized, while the *eclipsed* conformer corresponds to the rotation barrier and was not stable. Both *gauche* conformations were stabilized since a hydrogen bond was formed between protonated and unprotonated nitrogen atoms in the ethylenediamine unit.

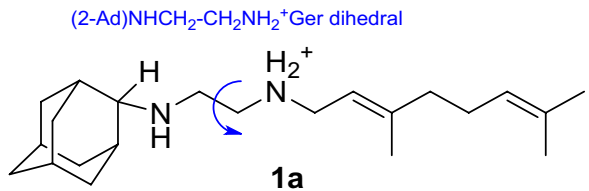
The rotation around (2-Ad)CH₂CH₂–NH₂⁺CH₂Ger dihedral favored the *anti* orientation (Table 2) since in the *gauche* conformation the steric energy increased due to repulsion between the geranyl and 2AdNH₂⁺ groups. In these *gauche* conformations the geranyl group of the

Table 2 Calculations of conformational energies by rotation of (2-Ad)CH₂CH₂–NH₂⁺Ger in SQ109 (**1a**) using the B3LYP-D3/6-31G(d,p) and PCM for water (ϵ = 80) or chloroform (ϵ = 4.8)



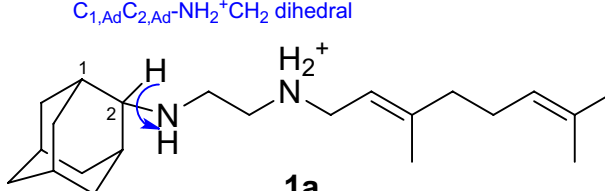
Dielectric	ΔG (kcal mol ⁻¹) for SQ109(1a)		
	Conformer		
	<i>gauche</i> (–)	<i>gauche</i> (+)	<i>anti</i>
ϵ = 80	2.77	2.78	0
ϵ = 4.8	2.85	2.94	0

Table 1 Calculations of conformational free energies by rotation of (2-Ad)CH₂–CH₂NH₂⁺Ger in SQ109 (**1a**) using the B3LYP-D3/6-31G(d,p) and PCM for water (ϵ = 80) or chloroform (ϵ = 4.8)



Dielectric	ΔG (kcal mol ⁻¹) for SQ109(1a)		
	conformer		
	<i>gauche</i> (–)	<i>gauche</i> (+)	<i>anti</i>
ϵ = 80	2.95	0	10.23
ϵ = 4.8	2.92	0	8.80

Table 3 Calculations of conformational energies by rotation around C_{1,Ad}C_{2,Ad}–NHCH₂ in SQ109 (**1a**) using the B3LYP-D3/6-31G(d,p) and PCM for water (ϵ = 80) and chloroform (ϵ = 4.8)



dielectric	ΔG (kcal mol ⁻¹) for SQ109(1a)		
	Conformer		
	<i>gauche</i> (–)	<i>gauche</i> (+)	<i>anti</i>
ϵ = 80	0	3.69	8.24
ϵ = 4.8	0	3.61	9.25

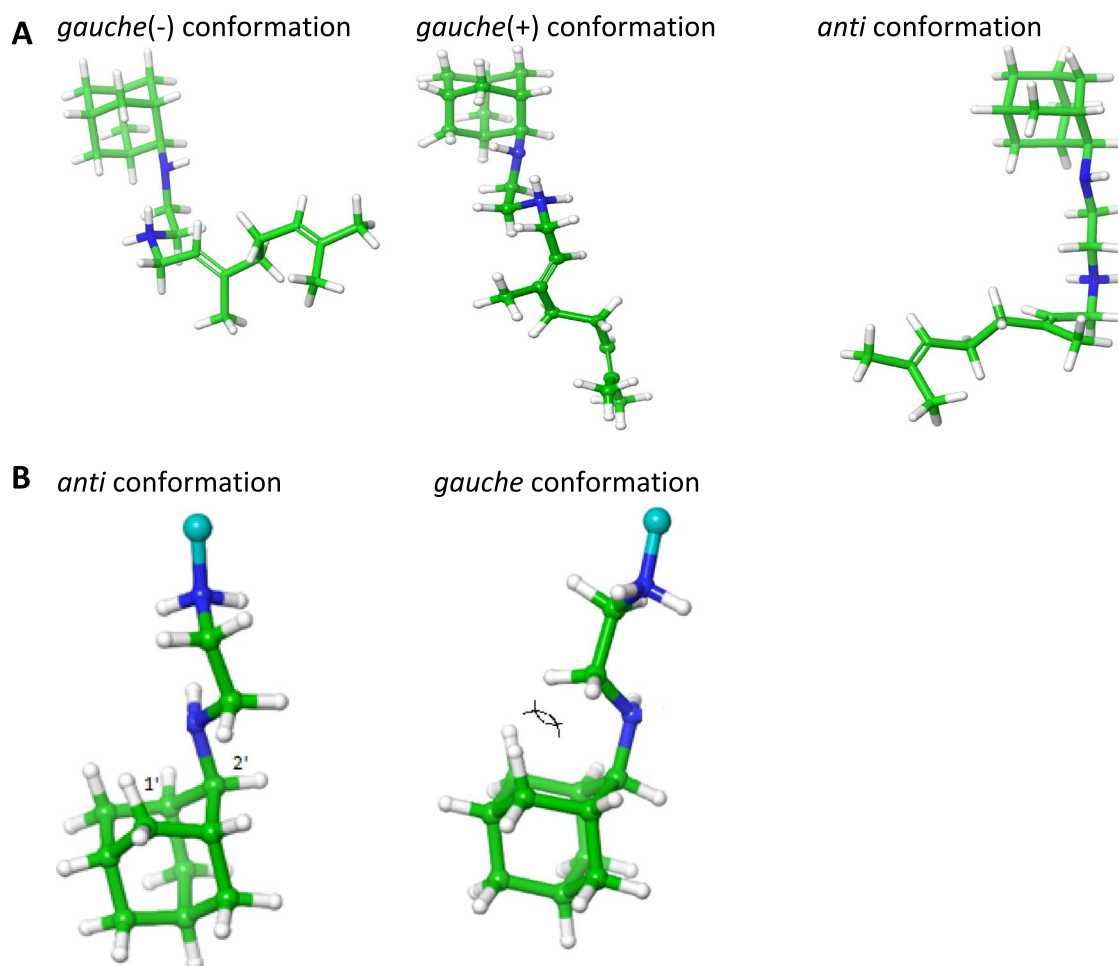


Fig. 2 Important conformational features of SQ109 (**1a**). In (A) are shown the conformations *gauche(-)* or *gauche(+)* or *anti* by rotation around the ethylenediamine's unit carbon-carbon bond, (2-Ad)CH₂-CH₂NH₂⁺CH₂Ger (see Table 1)), which defined the relative orientation of 2-adamantyl and geranyl groups. The conformation around the (2-Ad)CH₂CH₂-NH₂⁺CH₂Ger dihedral was *anti*. In (B) are shown the conformations generated by rotation around the C_{1,Ad}C_{2,Ad}-NHCH₂

dihedral (see Table 3) which defined the position of axial NHCH₂ as regards the cyclohexane subunit of 2-adamantyl group (geranyl group is shown with a light blue sphere); in the structure shown, in the right-hand part, the axial NHCH₂ brings CH₂ above the cyclohexane subunit of 2-adamantyl increasing steric repulsion (shown with two cross circular arc lines)

SQ109 (**1a**) could not adopt an extended structure that fits inside the narrow pore of the transporter.

The rotation around the C_{1,Ad}C_{2,Ad}-NHCH₂ dihedral defined the position of the alkyl and the 2-adamantyl ammonium groups and the DFT calculations results are shown in Table 3. The C_{1,Ad}C_{2,Ad}-NHCH₂ moiety can adopt only the equivalent two *gauche* conformations (which place the nitrogen of the ammonium group at C2-adamantyl position and C_{n-1} at symmetrical positions) due to the severe steric repulsions of the axial NHCH₂ in the *anti* conformation with the cyclohexane subunit group of the 2-adamantyl group (Fig. 2B).

In summary, the DFT study showed that rotation of SQ109 around (2-Ad)NHCH₂-CH₂NH₂⁺Ger dihedral favored two *gauche* conformations as minima while other

conformers were unpopulated in both dielectric media as well as inside the transporter's pore demonstrated by our MD simulations (see the next paragraph). Rotation around (2-Ad)NHCH₂CH₂-NH₂⁺Ger bond favored an *anti* orientation and rotation around (2-Ad)-NHCH₂CH₂NH₂⁺Ger bond showed that the position of NHCH₂ group above the cyclohexane subunit was prohibited.

MD simulations of the SQ109-MmpL3 complex

MmpL3 protein consists by 12 transmembrane (TM) α -helices and the TM region also contains two extra α -helices attached to the cytoplasmic membrane surface. We used the structure with PDB ID 6AJG [17] of the protein after excluding the C-terminus that has residues M1-E749.

The TM domain consists by the following α -helices and their residues: TM1 (14–33), TM2 (174–199), TM3 (208–224), TM4 (238–264), TM5 (271–301), TM6 (306–338), TM7 (396–415), TM8 (552–576), TM9 (583–601), TM10 (625–648), TM11 (660–690), TM12 (697–728).

In the PDB ID 6AJG, [17] SQ109 (**1a**) binds to the transmembrane domain of the MmpL3 transporter, from G641 to F649. It disrupts the hydrogen bonding interactions between the two D-Y pairs, where proton translocation takes place, blocking activation of the transporter. In particular, the ethylenediamine group of SQ109 (**1a**) intervenes between the D256-Y646 and D645-Y257 pairs [14] and forms hydrogen bonds with Asp645. The X-ray structure (PDB ID 6AJG) [17] showed that the complex is stabilized through numerous van der Waals interactions between the geranyl-ethylenediamine moiety of SQ109 (**1a**) and surrounding amino acids of MmpL3, including I249, I253, I297 in the upper part and L642, Y257, Y646, L686 in the bottom part of the pocket while the 2-adamantyl group lies close to F260 and F649. Ethylenediamine molecule has $pK_{a,1} = 10.71$ and $pK_{a,2} = 7.56$ [43] and at pH 7.4 the mono and diprotonated species will be equally populated. However, since the basic amino group close to adamantyl group is more hindered to protonation, we assumed that the monoprotated species must be predominated inside the hydrated MmpL3 pore, although actual protonation states are not known and would probably require neutron diffraction structures. Strikingly, in the X-ray structure (PDB ID 6AJG) [17] the ethylenediamine unit adopts a high energy conformation by rotation around its central carbon–carbon bond, described in (2-Ad) $\text{NHCH}_2\text{--CH}_2\text{NH}_2^+\text{Ger}$, having *eclipsed* the C–N bonds. Protein–ligand coordinates obtained from crystallography or cryo-em experiments provide a static model corresponding to a snapshot. While is not uncommon to find ligand conformers in complex with a protein that differ significantly from the lowest energy conformation in solution, due to stabilizing interactions inside the receptor, it is appropriate to judge the stability of the ligand's conformation in the experimental structure using MD simulations, as previously suggested [44–48].

We performed 100 ns-MD simulations (two repeats, see Fig. S3) using the SQ109 (**1a**)-MmpL3 complex (PDB ID 6AJG) [17] embedded in a hydrated 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) bilayer using the amber99sb force field (ff99sb) [49] and a 500 ns-MD simulation (Fig. S4) for testing the stability of the protein complex in longer time scale. For each simulation, initial atom velocities were assigned randomly and independently. Structure of SQ109 (**1a**)-MmpL3 complex with PDB ID 6AJG [17] has ~ 2.6 Å resolution which is higher than other

structures resolved in refs [21, 22] while no other structure of MmpL3 in complex with an SQ109 (**1a**) analog has been resolved.

We observed a rotation of the carbon–carbon bond in the ethylenediamine unit that converts the *eclipsed* conformation of SQ109 (**1a**) in the MmpL3 X-ray structure (PDB ID 6AJG), [17] into the two possible *gauche* conformations, with no population of the *eclipsed* conformer over the whole trajectory. In each of the two *gauche* conformations the monoprotated ethylenediamine unit formed stabilizing hydrogen bond interactions with both D256 and D645, as compared to the *eclipsed* conformation which can form hydrogen bonds only with D645. Thus, in the *gauche* conformation the NH_2^+ group of the monoprotated ethylenediamine unit formed direct ionic hydrogen bonds with both D256 and D645 while the unprotonated NH acted as a donor in one direct and one water-mediated hydrogen bond with D256 (Fig. 3). We observed also that water molecules entered the pore forming hydrogen bonds between the monoprotated ethylenediamine unit and side chains of Y257/Y646, D256/D645, S293 or carbonyl groups backbone peptide bonds. The geranyl-ethylenediamine moiety was tightly enclosed in a narrow area of the transporter's pore with the geranyl chain surrounded by amino acids Y257, I297, L642, Y646 while at the bottom wider part of the binding area the 2-adamantyl group had van der Waals contacts with F260 and F649.

Simulations of the complexes of MmpL3 with SQ109 analogs **1b-i**, **2**

Docking calculations results

The X-ray structure of the MmpL3 – SQ109 (**1a**) complex (PDB ID 6AJG [17]) was used as the template structure for the docking calculations of ligands SQ109(**1a**), **1b-i**, **2** with MmpL3 after excluding C-terminus, consisting of M1-E749 residues as previously described. The highest score docking pose of SQ109 (**1a**) inside the MmpL3 pore (ChemScore [50] scoring function) had a root-mean-square-deviations of heavy atoms ($\text{RMSD}_{\text{ligand}}$) 1.7 Å compared to the structure with PDB ID 6AJG [17]. This suggested that the calculated docking poses could describe accurately the binding orientation of SQ109 analogues. The docking poses of the monoprotated ethylenediamine SQ109 analogues **1b-i**, **2** showed that the new adamantyl moieties can fit in the empty region at the bottom of the binding area where for example the bigger alkyl adducts linked at C-2 adamantyl position can be accommodated. The docking calculations showed that addition of alkyl, aryl or heteroaryl adducts, e.g. Me,

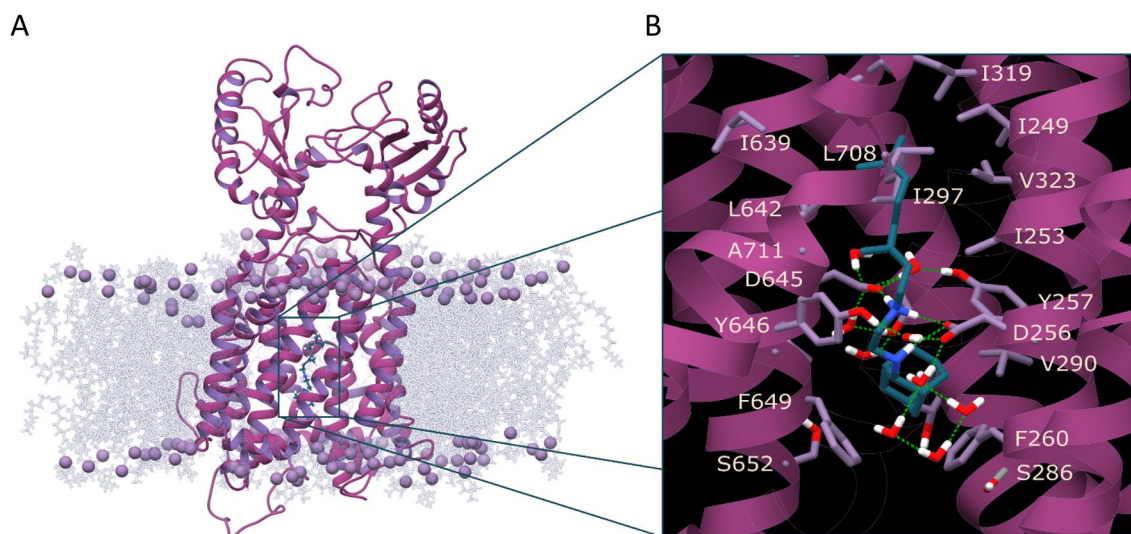


Fig. 3 **A** Structure of MmpL3 in complex with SQ109 (**1a**) after 100 ns MD simulations with ff99sb [49]. **B** Close-up view of the binding site. SQ109 (**1a**) formed hydrogen bond interactions with the two D-Y pairs, D256-Y646 and D645-Y257. Color scheme for frames: Ligand=petrol sticks, receptor=purple ribbons, residues as light purple sticks. Hydrogen bonding interactions=green dashes. For the protein, the experimental structure of SQ109 (**1a**) in com-

plex with MmpL3 (PDB ID 6AJG [17]) from *Ms*, was used as the starting structure for the MD simulations after excluding C-terminus, consisting by M1-E749 residues. The TM region included the following helices and their residues: TM1 (14–33), TM2 (174–199), TM3 (208–224), TM4 (238–264), TM5 (271–301), TM6 (306–338), TM7 (396–415), TM8 (552–576), TM9 (583–601), TM10 (625–648), TM11 (660–690), TM12 (697–728)

Et, nPr, nBu, Ph, Bn, Hex, 5-phenyl thiazol-2-yl (Ph-Thz) at the adamantyl C-2 of SQ109 (**1a**), in compounds **1b-i**, respectively, or replacement of the 2-adamantyl group by a 1-adamantyl-dimethylmethylene (C-1 dimethylmethylene) group in compound **2**, resulted in highest score docking poses having a *gauche*(–) or *gauche*(+) by rotation of (2-Ad)CH₂–CH₂NH₂⁺CH₂Ger that bind MmpL3 pore.

The docking algorithm produced 30 docking solutions. We visually inspected and realized that the first 4–5 highest docking solutions corresponded to a similar conformation of the ligand inside the receptor.

For the diprotonated SQ109 (**1a**) analogs, mainly the *gauche*(–) or *gauche*(+) by rotation of (2-Ad)CH₂–CH₂NH₂⁺CH₂Ger were obtained as highest score docking poses, but also in two cases the *anti* and *eclipsed* conformations were also observed (Table S2), since these later conformers stabilized inside the receptor with strong hydrogen bonding interactions despite their much lower stability in solution compared to the *gauche* conformation (Table 1). It was not unusual that the highest docking pose for each ligand was different since this reflected the high flexibility of the ligands and the random nature of the docking algorithm (genetic algorithm runs). For the MD simulations we used the highest scoring docking pose as starting conformation. We also tested as starting structure a conformation for the ligands **1b-i** resulted from superposition with

the last snapshot from 100 ns-MD simulation of MmpL3-SQ109 (**1a**) complex.

MD Simulations of SQ109 analogs in complex with MmpL3

To explore the dynamic behavior between the ligands **1a-i**, **2** and MmpL3 we performed 100 ns-MD simulations (two repeats, see Fig. S3) with starting structure the docking pose with the highest score. Thus, these docking poses were embedded in hydrated POPC bilayers of ~140 lipids and subjected to MD simulations using ff99sb. For each simulation, initial atom velocities were assigned randomly and independently. We performed the 100 ns-MD simulations using as starting structure a conformation for the ligands **1b-i** resulted from superposition with SQ109 (**1a**) but the results were similar.

For the sizeable alkyl substituents at C-2 adamantyl position, e.g. in analogs **1f**, **g**, **i**, 500 ns-MD simulations were also performed for testing the stability of the protein complexes in longer time scale (Fig. S4). For each simulation, initial atom velocities were assigned randomly and independently. We observed from the 100 ns-MD simulations or the 500 ns-MD simulations that the RMSD values of the protein TM C α carbons converged after 10–40 ns with RMSD_{protein} (Ca TM) ≤ 2.1 Å, suggesting small changes compared to

the X-ray structure (Fig. 4, S1, Table 4). A different stabilization between the full protein and the TM region was observed reflected by the different $\text{RMSD}_{\text{protein}}$ (Ca full protein) and $\text{RMSD}_{\text{protein}}$ (Ca TM) with values for the latter much smaller than the former. It is not unusual that the loops connecting the TM-helices are very flexible and increase the RMSD (Ca full protein). The ligands, which contained very flexible moieties such as the geranyl and ethylenediamine groups, shifted considerably from the starting docking pose, as revealed from the high values of RMSD ligand and in most cases the ligand binding conformation equilibrates in a stable position inside the transporter's pore after 70 ns (Fig. 4, S1, Table 4). The last frame described well the ligand–protein interaction frequency plots. Figure 4 shows last frames and ligand–protein interaction frequency plots for SQ109 (**1a**) and selected alkyl and aryl groups attached at the adamantyl C-2 of **1a** (R = H) in Fig. 1 including analogs **1b** (R = nBu), **1h** (R = Ph), **1i** (R = Ph-Thz) while the last frames, ligand–protein interaction frequencies for the other analogs, as well as the RMSD plots from MD simulations can be found in Fig. S1.

The MD simulations showed that the monoprotonated ethylenediamine unit of the ligands **1b-i, 2** adopted a *gauche* conformation, which favored hydrogen bonding interactions with 256 and/or D645 of MmpL3 (Fig. 4, S1). According to the MD simulations the complexes between **1b-i, 2** and MmpL3 formed common van der Waals interactions with the protein's residues along the narrow area of the transporter pore. Compared to SQ109 (**1a**) the geranyl-ethylenediamine moiety was surrounded similarly by the amino acid residues L642, Y646, Y257 while the 2-adamantyl group lied close to F260 and F649 at the bottom part of the binding area. However, the alkyl adducts increased the hydrophobic interactions at the bottom of the binding area with F260 and F649 and increased also the hydrophobic interactions with residues in the walls of narrow area of the pore, e.g. I253, see for example the analogs **1e, h, i** in Fig. 4. Water molecules formed hydrogen bonds between the monoprotonated ethylenediamine unit and receptor's residues, as described previously for SQ109 (**1a**).

We explored the disposition of the important residues for binding of the ligands (D256, D645, Y257, Y646, F260, F649) along the MD simulation and checked for other frames that could accommodate better the ligands. The RMSD (Ca) of these important residues for binding were plotted (Fig. S5). The RMSD (Ca) converged to values that range between 0.5 and 1.5 confirming not important disposition of these residues along the MD simulation.

While the monoprotonated form of SQ109 (**1a**) and analogs most likely predominated inside the hydrated MmpL3 pore, we also performed MD simulations for the diprotonated species (Fig. S2). The MD simulations showed that

the ligands in the diprotonated ethylenediamine form had similar coordinates inside the MmpL3 (Fig. S2) compared to the monoprotonated form.

MM-GBSA binding free energy calculations

We applied the MM-GBSA method [29–31] using the OPLS2005 force field [51] for the calculation of binding free energies (ΔG_{eff}) of ligands **1a-i, 2** inside the MmpL3 pore, using ensembles from 20 ns-MD simulations and calculated binding free energies without or with considering the membrane environment of the protein complex. For each simulation, initial atom velocities were assigned randomly and independently. Thus, we tested the membrane protein – ligand systems using an implicit membrane, i.e. a hydrophobic slab, [52–55] and considering explicitly water molecules inside the binding area [56]. The range of molecular weight of the ligands is 400–500 Da which corresponds to tripeptides [30] and their carbon skeleton was long enough to interact with many residues inside the receptor area. If a correct ranking of the binding affinities of **1b-i, 2** for the MmpL3 pore could be accomplished with the MM-GBSA method this would be a significant reduction in computational resources, compared to the accurate binding free energy perturbation methods for **1b-i, 2**–MmpL3 complexes in the POPC lipid bilayers containing $\sim 10^5$ atoms.

However, the calculated mean values of three repeats for the monoprotonated forms (Table 4) and diprotonated forms (Table S2) showed that the MM-GBSA method (applied with or without modifications to consider the membrane environment of the protein complex using a hydrophobic slab [52–55] and the explicit waters inside the binding area) [56] did not afford valuable results. The binding free energy values were dispersed and did not follow any trend, see Table 4, S2. Indeed, as a calculation method, MM-GBSA can show large deviations (e.g. 4 kcal mol⁻¹) in standard binding free energies compared to the experimental binding free energies. The method normally can provide useful results [57] for ranking of substituents [29] in the same series of ligands when the experimental binding affinities range is ~ 1000 or higher [30] which is not the case for **1b-i, 2** since their K_D values differ only by 4-fold.

Compared to the monoprotonated species, we observed that the diprotonated ethylenediamine moieties in compounds **1a-i, 2** form stronger hydrogen bonding interactions with the receptor area. This agreed with the MM-GBSA binding free energy values for the monoprotonated forms compared to the diprotonated forms since in the latter case the values were significantly lower consistent with enhanced hydrogen bonding interactions between the ligands and the receptor's binding site (Table 4, S2). Because of the stronger hydrogen bonding interactions, the ligands had smaller

Fig. 4 Last frames (left part) of monoprotonated ethylenediamine form for SQ109 (**1a**), and analogs **1e**, **h**, **i** inside the MmpL3 transporter in a POPC lipid bilayer from 100 ns-MD simulations with ff99sb. The receptor-ligand interaction frequency histograms plotted for the whole trajectory are shown on the right (the alkyl group connected at the adamantyl C-2 is shown inside the parenthesis following the compound number. Color scheme for frames: Ligand = petrol or purple or brown or orange sticks, receptor = white ribbons, residues in light purple sticks, hydrogen bonding interactions = dark grey dashes. For the protein, the experimental structure of SQ109 (**1a**) in complex with MmpL3 (PDB ID 6AJG 17) was used as the reference structure for the MD simulations after excluding C-terminus which included M1-E749 residues

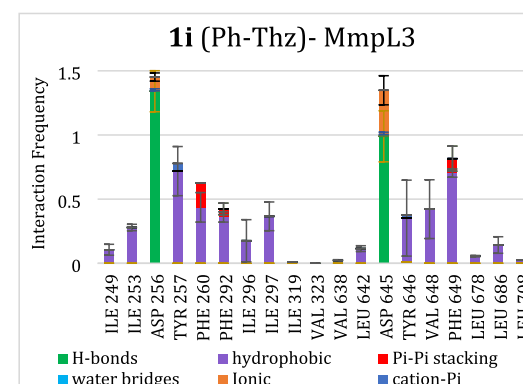
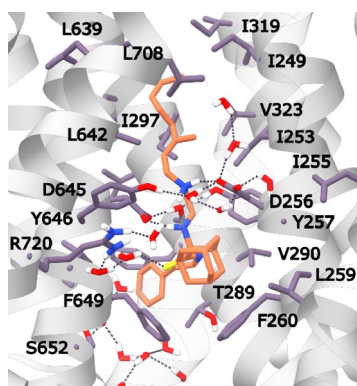
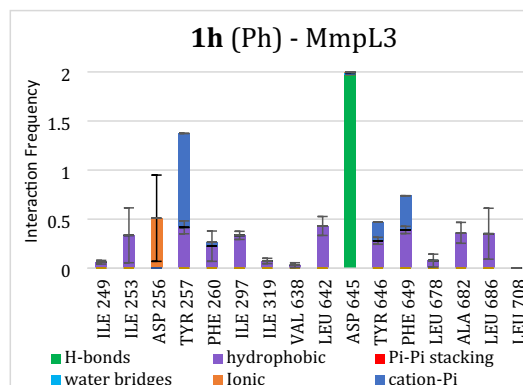
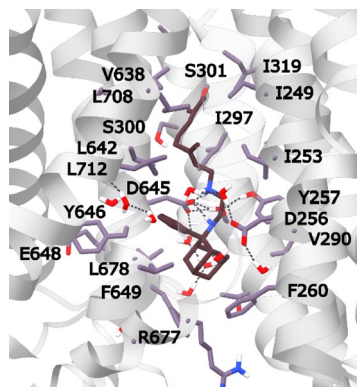
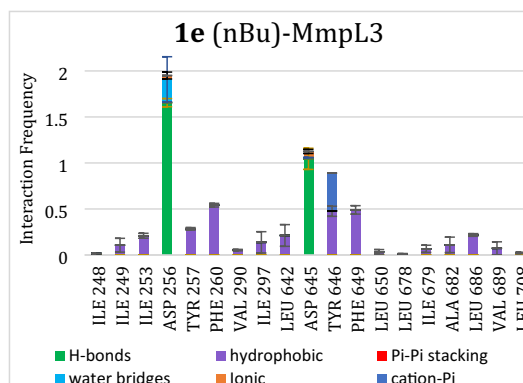
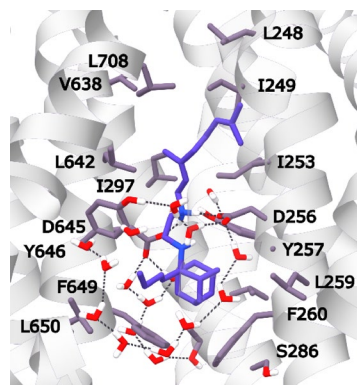
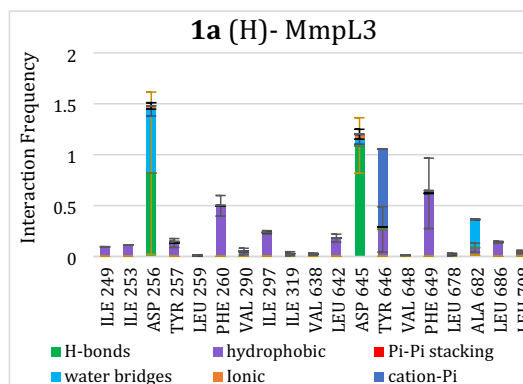
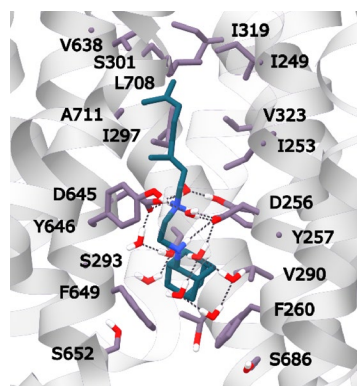
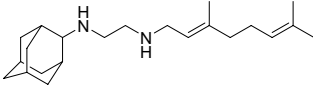
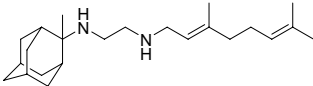
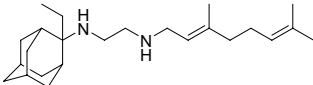
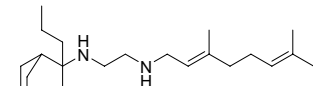
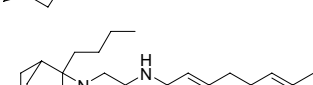
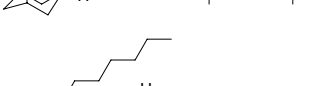
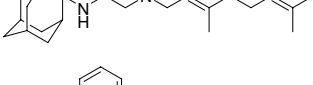
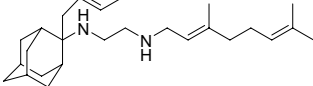
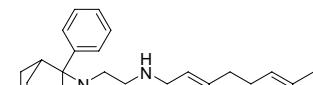



Table 4 Ligand-MmpL3 binding free energies (ΔG_{eff}) calculated using the MM-GBSA [29–31] method with the OPLS2005 [51] force field for the calculations of the intermolecular interactions without orwith using a hydrophobic slab [51–55] to model the membrane environment of the protein, RMSD_{ligand} and RMSD_{protein} (Ca TMD) mean values from 100 ns-MD simulations for **1a–i, 2**

Cmp No	Structure	RMSD _{ligand} ^a (Å)	RMSD _{protein} (Ca) ^b (Å)	RMSD _{protein} (Ca TMD) ^c (Å)	ΔG_{eff} ^d (kcal mol ⁻¹)
1a		3.29 ± 0.21	3.50 ± 0.12	1.17 ± 0.09	- 169.91 ± 7.37
1b		4.05 ± 0.28	4.28 ± 0.20	2.06 ± 0.07	- 131.80 ± 7.81
1c		3.02 ± 0.32	3.92 ± 0.19	1.83 ± 0.09	- 141.76 ± 9.72
1d		3.11 ± 0.21	3.57 ± 0.26	1.41 ± 0.08	- 159.92 ± 7.44
1e		4.00 ± 0.20	3.73 ± 0.17	1.29 ± 0.07	- 162.90 ± 8.18
1f		2.21 ± 0.21	3.61 ± 0.11	1.30 ± 0.05	- 160.48 ± 8.81
1g		5.43 ± 0.28	3.33 ± 0.15	1.37 ± 0.06	- 143.97 ± 7.68
1h		3.29 ± 0.18	2.71 ± 0.11	1.50 ± 0.07	- 153.60 ± 7.61
1i		3.12 ± 0.17	3.08 ± 0.17	1.38 ± 0.07	- 179.63 ± 8.75
2		3.21 ± 0.21	4.15 ± 0.30	1.26 ± 0.06	- 151.80 ± 8.74

The ligands have a monoprotonated ethylenediamine unit

^a Mean ± SD (Å); Ligand RMSD was calculated after superposition of each protein–ligand complex to that of the starting structure (snapshot 0) based on the C_α atoms of the protein, for the last 20 ns of the trajectories

^b Mean ± SD (Å); Protein RMSD was calculated for the C_α atoms of the whole protein, for the last 20 ns of the trajectories, using as starting structure snapshot 0 of the production MD simulation

^c Mean ± SD (Å); Protein RMSD was calculated for the C_α atoms of only the α-helices of the TM region, for the last 20 ns of the trajectories, using as starting structure snapshot 0 of the production MD simulation

^d Mean calculated effective binding free energy (kcal mol⁻¹) between ligand and MmpL3 receptor from three repeats. ΔG_{eff} was calculated from the last 20 ns of the trajectories using 40 ps intervals (i.e. 500 frames per trajectory) using a MM-GBSA model that considered the membrane as hydrophobic slab. [51–55]

flexibility as was shown by the lower $\text{RMSD}_{\text{protein}}$ (Ca TM) and lower $\text{RMSD}_{\text{ligand}}$ (Fig. S1, S2, Table 4, S2).

Structure–activity relationships of SQ109 analogs against Mmpl3 using alchemical binding free energy calculations with TI/MD

We previously measured [13] using SPR the binding affinities of SQ109 (**1a**) and the 9 active analogs **1b–i**, **2** against MtMmpl3 [23] with K_d values R=H (SQ109 (**1a**), 2060 μM); R=Me (**1b**, 248 μM); R=Et (**1b**, 190 μM); R=nPr (**1d**, 106 μM); R=nBu (**1e**, 108 μM); R=nHex (**1d**, 81 μM). It was observed that the K_d decreased showing tighter binding to Mmpl3 as the R substituent at C-2 adamantyl (which is an H in SQ109) became larger and more hydrophobic. A similar effect was observed with phenyl (**1h**, 136 μM); benzyl (**1g**, 74 μM) (Table S1). The C-1 dimethylmethylene analog **2** had a $K_d = 106 \mu\text{M}$ which was close to the isomeric **1d** (n-Pr) which had a $K_d = 120 \mu\text{M}$ and **1i** (Ph-Thz) with a $K_d = 91 \mu\text{M}$.

The FEP/MD [58–60] or TI/MD [61, 33, 34] methods which are based on statistical mechanics can provide accurate results for relative binding free energies with an error 1 kcal mol⁻¹ and have been applied in membrane protein–ligand complexes, e.g. GPCRs. [32, 35, 62–65] We applied the TI/MD method combined with a thermodynamic cycle method to examine if the binding profile of the ethylenediamine analogs **1b–i**, **2** was the same with SQ109 (**1a**) in its complex with Mmpl3 (PDB ID 6AJG [17]). This might be likely if there is an agreement between calculated and experimental relative binding free energies for alchemical transformations between pairs of compounds **1a–i**, **2**. We performed TI/MD calculations for alchemical transformations in selected pairs of diamine SQ109 analogs that were not accompanied with large changes in ligand's alkyl. Thus, we calculated perturbations by one or two methylene groups in the C-2 alkyl adduct (Table 5).

The end states in the alchemical calculations tested were similar to the structure in the corresponding complexes resulted from the converged 100 ns-MD simulations. This was checked to certify that the 2 ns-MD simulation in each

λ -state was enough for the complexes to converge in an equilibrium structure. Two repeats of TI/MD calculations were performed for each alchemical transformation.

The effect in binding free energy by increasing the length of the alkyl chain by one methylene, which was examined with the alchemical perturbations **1a** (H) \rightarrow **1b** (Me) or **1b** (Me) \rightarrow **1c** (Et) or **1c** (Et) \rightarrow **1d** (n-Pr) or **1d** (n-Pr) \rightarrow **1e** (n-Bu), was to increase binding affinity (Table 5). As noted previously, in the 100 ns-MD simulations of Mmpl3—**1a–e** complexes the $\text{RMSD}_{\text{protein}}$ (Ca TMD) was $\leq 2.1 \text{ \AA}$ (Table 4). Thus, the last snapshots of the MD simulations were suitable starting structures for the TI/MD simulations of the studied perturbations (Fig. 4, S1).

The biggest change in experimental binding free energy was noted when H (SQ109) changed to Me (**1b**), $\Delta\Delta G_{b,\text{exp}} = -1.30 \text{ kcal mol}^{-1} \pm 0.79 \text{ kcal mol}^{-1}$, and when Ph (**1h**) changed to Bn (**1g**), $\Delta\Delta G_{b,\text{exp}} = -0.38 \pm 0.02 \text{ kcal mol}^{-1}$ or Et (**1c**) changed to Pr (**1d**), $\Delta\Delta G_{b,\text{exp}} = -0.36 \text{ kcal mol}^{-1} \pm 0.29 \text{ kcal mol}^{-1}$.

The binding free energy changes for **1a** (H) \rightarrow **1b** (Me) were $\Delta\Delta G_{b,\text{exp}} = -1.30 \text{ kcal mol}^{-1} \pm 0.79 \text{ kcal mol}^{-1}$, $\Delta\Delta G_{b,\text{TI/MD}} = -0.49 \pm 0.06 \text{ kcal mol}^{-1}$, for **1b** (Me) \rightarrow **1c** (Et) were $\Delta\Delta G_{b,\text{exp}} = -0.16 \pm 0.14 \text{ kcal mol}^{-1}$, $\Delta\Delta G_{b,\text{TI/MD}} = 0.06 \pm 0.08 \text{ kcal mol}^{-1}$, for **1c** (Et) \rightarrow **1d** (n-Pr) were $\Delta\Delta G_{b,\text{exp}} = -0.36 \pm 0.29 \text{ kcal mol}^{-1}$, $\Delta\Delta G_{b,\text{TI/MD}} = -0.87 \pm 0.09 \text{ kcal mol}^{-1}$ and for **1d** (n-Pr) \rightarrow **1e** (n-Bu) were $\Delta\Delta G_{b,\text{exp}} = 0.01 \pm 0.30 \text{ kcal mol}^{-1}$, $\Delta\Delta G_{b,\text{TI/MD}} = 0.20 \pm 0.11 \text{ kcal mol}^{-1}$.

We considered next the perturbation in the C-2 adamantyl alkyl by two methylenes in **1e** (n-Bu) \rightarrow **1f** (n-Hex) and we studied the change by one methylene from phenyl to benzyl in C-2 substituent in **1h** (Ph) \rightarrow **1g** (Bn) where the perturbation in conformational space should be relatively important. The binding free energy changes were $\Delta\Delta G_{b,\text{exp}} = -0.18 \text{ kcal mol}^{-1} \pm 0.01 \text{ kcal mol}^{-1}$, $\Delta\Delta G_{b,\text{TI/MD}} = -1.42 \pm 0.15 \text{ kcal mol}^{-1}$ and $\Delta\Delta G_{b,\text{exp}} = -0.38 \pm 0.02 \text{ kcal mol}^{-1}$, $\Delta\Delta G_{b,\text{TI/MD}} = -1.83 \pm 0.15 \text{ kcal mol}^{-1}$, respectively. We did not test larger perturbations that were not consistent with the method's principles. [61]

In general, the deviation from experimental values was smaller than 1 kcal mol⁻¹ when the perturbation was one

Table 5 Free energy perturbation values computed with the MBAR method [66] for alchemical simulations performed with TI/MD for pairs of ligands bound to Mmpl3

Alchemical transformation	$\Delta\Delta G_{b,\text{exp}}$ (kcal mol ⁻¹)	$\Delta\Delta G_{b,\text{TI/MD}}$ (kcal mol ⁻¹)	$ \Delta\Delta G_{b,\text{TI/MD}} - \Delta\Delta G_{b,\text{exp}} $ (kcal mol ⁻¹)
1a (H) \rightarrow 1b (Me)	-1.30 ± 0.79	-0.49 ± 0.06	0.81
1b (Me) \rightarrow 1c (Et)	-0.16 ± 0.14	0.06 ± 0.08	0.23
1c (Et) \rightarrow 1d (n-Pr)	-0.36 ± 0.29	-0.87 ± 0.09	0.51
1d (n-Pr) \rightarrow 1e (n-Bu)	0.01 ± 0.30	0.20 ± 0.11	0.18
1e (n-Bu) \rightarrow 1f (n-Hex)	-0.18 ± 1.01	-1.42 ± 0.12	1.245
1h (Ph) \rightarrow 1g (Bn)	-0.38 ± 0.02	-1.83 ± 0.15	1.45

methylene, e.g. for **1a** (H) → **1b** (Me), **1b** (Me) → **1c** (Et), **1c** (Et) → **1d** (n-Pr), **1d** (n-Pr) → **1e** (n-Bu), see Table 5. When the perturbation in the conformational space was bigger, e.g. was two methylene groups in **1e** (n-Bu) and **1f** (n-Hex) or from phenyl to benzyl in **1h** (Ph) and **1g** (Bn) the deviation was larger, i.e. 1.25 kcal mol⁻¹ or 1.45 kcal mol⁻¹, but in both these two cases was smaller than 1.5 kcal mol⁻¹. Overall, the mean assigned error (mue) was 0.739 kcal mol⁻¹ which was consistent with the fact that **1a-h** bind similarly with SQ109 (**1a**) to MmpL3 in its experimental structure (PDB ID 6AJG [17]), and that alkyl or aryl substituents at the adamantyl C-2 of SQ109 can fill the lipophilic region between Y257, Y646, F260 and F649 in MmpL3 pore and increasing the binding affinity.

Discussion

SQ109 (**1a**) [5] is an ethylenediamine-based inhibitor of MmpL3 undergoing clinical trials [3, 4] that also has activity against a broad range of bacteria, protozoa and even some yeasts/fungi [26]. Previous research suggested that SQ109 (**1a**) can block the MmpL3-mediated transport of trehalose monomycolates [5, 6] through preventing the proton transportation by (a) binding directly to the transporter's pore [17, 23] in *Mtb*, as supported by the X-ray structure of the MmpL3 from *M. smegmatis* in complex with SQ109 (**1a**) (PDB ID 6AJG [17]) or (b) indirectly [7, 26, 13, 67, 68] which in principle can be accomplished by membrane structure perturbation [7, 26, 32, 44, 56, 57] leading to increased membrane lipid disorder/fluidity and, arguably, to uncoupler activity on the PMF [5, 7, 24, 25].

The sequence of MmpL3 is highly conserved across *Mycobacteria* [68]. Of the MmpL proteins encoded by mycobacterial genomes, MmpL3 and MmpL11 are the only MmpL genes conserved across *Mycobacteria* [16]. The importance of MmpL3 is illustrated by the fact that it is the only MmpL gene that cannot be successfully knocked out in *Mtb* [15]. That MtMmpL3 could rescue the viability of the *Ms* MmpL3 knockout mutant further indicates that these MmpL3 orthologs have highly conserved functions [6]. Therefore, *Ms* MmpL3 is a reasonable model for the *Mtb* counterpart since the two MmpL3 orthologs can substitute each other to function [19].

Here, based on SPR data we previously obtained [13] (Table S1) showing the binding of **1a-i, 2** to MtMmpL3 and the tighter binding of the bigger adducts at C-2 adamantyl group, we investigated the binding profile of compounds **1b-i, 2** using MD simulations and alchemical relative binding free energy calculations based on the X-ray structure of the MmpL3-SQ109 (**1a**) complex from *Ms* (PDB ID 6AJG [17]).

We performed MD simulations of the X-ray structure with PDB ID 6AJG [17] in POPC bilayers containing ~140 lipids and showed that the *eclipsed* conformer observed in the X-ray structure represents the transition state for rotation around NHCH₂-CH₂NH₂⁺ dihedral compared to the preferred *gauche* conformations inside the MmpL3 which we observed in docking calculations and MD simulations.

To fully understand the conformational properties of SQ109 (**1a**) we performed DFT calculations of the *gauche* and *anti* conformations generated by rotation around (2-Ad)NHCH₂-CH₂NH₂⁺Ger, (2-Ad)NHCH₂CH₂-NH₂⁺CH₂Ger and C_{1,Ad}C_{2,Ad}-NHCH₂CH₂NH₂⁺Ger bonds. Thus, we calculated the free energies of SQ109 (**1a**) conformation in an implicit water environment ($\epsilon = 80$) or in a lipophilic environment (chloroform, $\epsilon = 4.8$) using the B3LYP-D3/6-31G(d,p) [69] theory (Tables 1–3, Fig. 2) and identified two *gauche* conformations as minima by rotation around AdNHCH₂-CH₂NH₂⁺Ger dihedral. These *gauche* conformations were stabilized with a hydrogen bond between the protonated ammonium group and the unprotonated nitrogen; the next more stable was the *anti* conformation, being more than ~9 kcal mol⁻¹ higher in energy and thus unpopulated (Table 1, Fig. 2), while the eclipsed conformer observed in the X-ray structure represents the transition state by rotation around AdNHCH₂-CH₂NH₂⁺Ger bond. In both dielectric media and inside the transporter's pore, rotation around AdNHCH₂CH₂-NH₂⁺CH₂Ger bond favored the *anti* orientation (Table 2), in agreement with the extended geranyl chain structure that fits inside the narrow pore of MmpL3 transporter, since in the *gauche* conformation the steric energy increased due to repulsion between the geranyl and NH₂⁺(2-Ad) groups. The DFT calculations for the C_{1,Ad}C_{2,Ad}-NHCH₂CH₂NH₂⁺Ger bond rotation, which defined the position of axial NHCH₂ as regards the cyclohexane subunit of adamantyl group, showed that the position of axial NHCH₂ brings CH₂ above the cyclohexane subunit is prohibited as increasing considerably the steric repulsion.

In addition, we performed MD simulations of the complexes between MmpL3 and **1b-i, 2** which showed that, comparatively to SQ109 (**1a**) the ligands formed also hydrogen bonding interactions with D256 and/or D645 of MmpL3 (Fig. 4, S1) and had common van der Waals interactions with the protein's residues along the transporter's pore axis. In these complexes the geranyl-ethylenediamine moiety was surrounded by the amino acid residues L642, Y646, Y257 while the 2-adamantyl group lied close to F260 and F649 at the bottom part of the binding area. We observed that increasing the length of the alkyl chain the hydrophobic interactions with F260 and F649 were increased as well as with residues in the pore, e.g. I253 (Fig. 4, S1) which was consistent with the SPR binding affinities.

To further confirm that the new SQ109 analogs **1b-i, 2** bind to the MmpL3 pore according to the experimental structure

of SQ109(**1a**) bound to MmpL3 (PDB ID 6AJG [17]), we compared the calculated binding free energy values with the MM-GBSA method [29–31] (without or with using an implicit membrane model [52–55] and considering also explicitly water molecules inside the binding area [56]) with the binding strength ranking from the experimental SPR results. The results showed that no valuable correlation was observed.

We performed alchemical relative binding free energy calculations using the accurate TI/MD method [61] which, along with FEP/MD method, [58] have been shown to perform with an error of 1 kcal mol⁻¹. For full accuracy, we included in the TI/MD calculations the whole protein-membrane system consisting of 10⁵ atoms. We applied the TI/MD method in alchemical transformations including changes in alkyl adduct at C-2 adamantyl by one or two methylene groups (Table 5) and examined how the calculated relative binding free energies were compared with experimental values that we measured using SPR (Table S1) [13]. For one methylene perturbations the deviation from experimental values was smaller than 1 kcal mol⁻¹ and for two methylenes or for Ph to Bn perturbations the deviation was bigger, but smaller than 1.5 kcal mol⁻¹. We observed a $\mu_{\text{e}}=0.74$ kcal mol⁻¹ that is consistent with the fact that alkyl or aryl substituents at the adamantyl C-2 of SQ109 (**1a**) can fill the empty lipophilic region close to F260 and F649.

Altogether, the MD simulations data that we produced based on the X-ray structure of the SQ109 (**1a**) and MmpL3 complex (PDB ID 6AJG [17]) agreed that our previously synthesized SQ109 (**1a**) analogs **1b-i**, **2** bind to the same binding area with SQ109 (**1a**). Compared to SQ109 (**1a**), in the analogs with larger alkyl or aryl adducts in the adamantane ring, the geranyl-ethylenediamine moiety was similarly surrounding by the amino acid residues L642, Y646, Y257. However, the larger adducts at 2-adamantyl carbon can fit close to F260 and F649 increasing the hydrophobic interactions at the bottom of the binding area.

Methods

DFT calculations

For the DFT calculations was used the B3LYP functional [36–38] in combination with D3 Grimme's correction for dispersion. [39, 69] All structures were fully optimized at B3LYP-D3/6-31G(d,p) level using the GAUSSIAN 03 [70] package; frequency calculations were also performed to locate minima.

Ligands preparation

The 2D structures of the compounds SQ109(**1a**), **1b-i**, **2** were sketched with Marvin Program (Marvin version 21.17.0, ChemAxon, <https://www.chemaxon.com>), model-built with

Schrödinger 2017–1 platform (Schrödinger Release 2021–1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2021; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2021) and the compounds' 3D structures in their monoprotonated form were energy minimized using the conjugate gradient method, the MMFF94 [71] force field and a distance-dependent dielectric constant of 4.0 until a convergence threshold of $2.4 \cdot 10^{-5}$ kcal mol⁻¹ Å⁻¹ was reached. The ionization state of the compounds at pH 7.5 were checked using the Epik program [72] implemented in Schrödinger suite (Prime Version 3.2, Schrödinger, LLC, New York, NY, 2015). The most likely state for the ethylenediamine unit is the mono- protonated but we also performed all the calculations for the diprotonated state as well.

Docking calculations

The X-ray structure of the MmpL3 – SQ109 (**1a**) complex (PDB ID 6AJG [17]) was used as the template structure for the docking calculations of ligands SQ109(**1a**), **1b-i**, **2** with MmpL3. The part of the protein sequence that extended to the periplasmic area and included amino acids F750-H929, was deleted as this part was very distant from the binding site. Additionally, the 34 amino acid sequence K355-G388 that was missing from the X-ray structure (PDB ID 6AJG [17]) was added using the Prime module of Maestro (Schrödinger Release 2021–1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2021; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2021). In the next step, the MmpL3 – SQ109 (**1a**) complex was optimized using the Protein Preparation Wizard implementation in Schrödinger suite (Schrödinger Release 2021–1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2021; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2021). [73] In this process, the bond orders and disulfide bonds were assigned, and missing hydrogen atoms were added. The N- and C-termini of the protein model were capped by acetyl and N-methyl-amino groups, respectively. All hydrogens of each protein complex were minimized with the OPLS2005 force field [74, 75] by means of Maestro/Macromodel 9.6 [76] using a distance-dependent dielectric constant of 4.0. The molecular mechanics minimizations were performed with the conjugate gradient method and a threshold value of $2.4 \cdot 10^{-5}$ kcal mol⁻¹ Å⁻¹ as the convergence criterion. Each protein was subjected in an all atom minimization using the OPLS2005 [74, 75] force field with heavy atom root mean square deviation (RMSD) value constrained to 0.30 Å until the RMS of conjugate-gradient reached values <0.05 kcal·mol⁻¹·Å⁻¹. Then SQ109 (**1a**), utilized as a reference ligand, and the apo protein MmpL3, utilized as

template protein, were saved separately or the docking calculations of the tested compounds SQ109(**1a**), **1b-i**, **2** to MmpL3 using GOLD software [77] (GOLD Suite, Version 5.2; Cambridge Crystallographic Data Centre: Cambridge, U.K., 2015. GOLD Suite, version 5.2; Cambridge Crystallogr. Data Cent. Cambridge, U.K., 2015) and ChemScore [50] as the scoring function. Each compound was docked in the binding site of SQ109(**1a**) in area of 10 Å around the experimental coordinates of SQ109 (**1a**) and 30 genetic algorithm runs were applied for each docking calculation. The “allow early termination” option, which terminated ligand conformational sampling if the top three solutions had an RMSD difference less than 1.5 Å was inactivated, and the “Generate Diverse Solutions” option, which sets the smallest inter-cluster RMSD to 1.5 Å, was activated. All other parameters were used with their default values. We performed the docking calculations also for the SQ109 analogs **1a-i**, **2** in the diprotonated form of ethylenediamine unit. The visualization of produced docking poses was performed using the program Chimera, [78] and the top-scoring docking poses were used as starting structures for the complexes for MD simulations to investigate the binding profile of the SQ109 (**1a**) and analogs **1b-i**, **2** inside the MmpL3 pore.

MD simulations

Each protein–ligand complex from docking calculations was inserted in a pre-equilibrated hydrated POPC membrane bilayer according to Orientations of Proteins in Membranes (OPM) database [79]. The protein was added in the hydrated lipid bilayer extended by 10 Å, 10 Å, 18 Å in *x*, *y*, *z* axes from the protein, consisting by ca. 140 lipids and 22,000 TIP3P water molecules, [80] using the System Builder utility of Desmond v4.9 (Schrödinger Release 2021–1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021). Sodium and chloride ions were added randomly in the water phase to neutralize the systems and reach the experimental salt concentration of 0.150 M NaCl. The total number of atoms of the complex was approximately 100,000 and the orthorhombic simulation box dimensions was (86 × 83 × 141 Å³) and applied periodic boundary conditions. We used the Desmond Viparr tool to assign amber99sb [49] force field parameters for the calculations of the protein and lipids and intermolecular interactions, and Generalized Amber Force Field (GAFF) [81] for assigning parameters to ligands. Ligand atomic charges were computed according to the RESP procedure [82, 83] using the Gaussian03 program [70] and the antechamber module of Amber18 [84].

The MD simulation of each protein–ligand complex inside the lipid bilayer was performed using the default protocol provided with Desmond v4.9 program. Thus, the MD

simulations protocol consisted of a series of MD simulations designed to relax the system, while not deviating substantially from the initial coordinates. During the first stage, a simulation was run for 200 ps at a temperature of 10 K in the NVT ensemble (constant number of atoms, volume and temperature), with solute-heavy atoms restrained by a force constant of 50 kcal mol⁻¹ Å⁻². The temperature was raised to 310 K during a 200 ps MD simulation in the NPT ensemble (constant number of atoms, pressure and temperature), with the same force constant applied to the solute atoms. The temperature of 310 K was used in MD simulations to ensure that the membrane state was above the main phase transition temperature of 296 K for POPC bilayers. [85] The heating was then followed by equilibration simulations. First, two 1 ns stages of NPT equilibration were performed. In the first 1 ns stage, the heavy atoms of the system were restrained by applying a force constant of 10 kcal mol⁻¹ Å⁻², and in the second 1 ns stage, the heavy atoms of the protein–ligand complex were restrained by applying a force constant of 2 kcal mol⁻¹ Å⁻² to equilibrate water and lipid molecules. In the production phase, the relaxed systems were simulated without restraints under NPT ensemble conditions for 100 ns or 500 ns.

Particle Mesh Ewald (PME) [86, 87] was employed to calculate long-range electrostatic interactions with a grid spacing of 0.8 Å. The SHAKE method was used to constrain heavy atom-hydrogen bonds at ideal lengths and angles [88]. Van der Waals and short-range electrostatic interactions were smoothly truncated at 10 Å. The Nosé-Hoover thermostat [89] was utilized to maintain a constant temperature in all simulations, and the Martyna-Tobias-Klein method [90] was used to control the pressure. The equations of motion were integrated using the multistep reversible reference system propagator algorithms (RESPA) integrator [91] with an inner time step of 2 fs for bonded interactions and non-bonded interactions within a cutoff of 10 Å. An outer time step of 6.0 fs was used for non-bonded interactions beyond the cutoff. Replicas of the system were saved every 10 ps. Within the 100 ns-MD simulation time, the total energy (not shown) and RMSD_{protein} (C_α TM) atoms reached a plateau, and the systems were considered equilibrated and suitable for statistical analysis (see Table 4, S2). The calculated RMSD_{protein} (C_α TM) for the last 50 ns was < 2.0 Å (see blue curves in Fig. S1). Two MD simulation repeats (Fig. S3) were performed for each system using the same starting structure and by applying in the MD simulations randomized velocities. We also used the same protocol and performed the MD simulations for the SQ109 analogs **1a-i**, **2** in their doubly protonated form of ethylenediamine unit in complex with MmpL3 (Fig. S2). All the MD simulations with Desmond or Amber software were run on GTX 1060 GPUs in lab workstations or the ARIS Supercomputer.

The visualization of the trajectories was performed using the graphical user interface (GUI) of Maestro and the protein–ligand interaction analysis was done with the Simulation Interaction Diagram (SID) tool, available with Desmond v4.9 program. For hydrogen bonding interactions, a 2.5 Å distance between donor and acceptor heavy atoms, and an angle $\geq 120^\circ$ between donor-hydrogen-acceptor atoms and $\geq 90^\circ$ between hydrogen-acceptor-bonded atom were applied. Non-specific hydrophobic contacts were identified when the side chain of a hydrophobic residue fell within 3.6 Å from a ligand's aromatic or aliphatic carbon, while π - π interactions were characterized by stacking of two aromatic groups face-to-face or face-to-edge. Water-mediated hydrogen bonding interactions were characterized by a 2.7 Å distance between donor and acceptor atoms, as well as an angle $\geq 110^\circ$ between them.

MM-GBSA calculations

For these calculations, structural ensembles were extracted in intervals of 40 ps from three 20 ns MD simulation repeats for each complex running with randomized velocities. Prior to the calculations all water molecules, ions, and lipids were removed, except 20 waters in the vicinity of the ligand, [92] and the structures were positioned such that the geometric center of each complex is located at the coordinate origin. The MD trajectories were processed with the Python library MDAAnalysis [93] in order to extract the 20 water molecules closest to any atom in the ligand for each of the 501 frames. During the MM-PBSA calculations, the explicit water molecules were considered as being part of the protein. Binding free energies of compounds in complex with MmpL3 were estimated using the 1-trajectory MM-GBSA approach. [29–31] The binding free energy for each complex was calculated using Eqs. (1)–(6)

$$\Delta G_{bind} = \langle G_{complex} - G_{protein} - G_{ligand} \rangle_{complex} \quad (1)$$

$$G_i = V_{MM} - T \langle S_{MM} \rangle + \Delta G_{solv} \quad (2)$$

$$V_{MM} = V_{bonded} + V_{coul} + V_{LJ} \quad (3)$$

$$\Delta G_{solv} = \Delta G_P + \Delta G_{NP} \quad (4)$$

The binding free energy for each complex can be calculated according to Eq. (5)

$$\Delta G_{bind} = \langle \Delta E_{coul} + \Delta E_{LJ} \rangle - T \langle \Delta S_{MM} \rangle + \Delta \Delta G_{solv} \quad (5)$$

and after neglecting entropy Eq. (5) is converted to Eq. (6)

$$\Delta G_{eff} = \langle \Delta E_{coul} + \Delta E_{LJ} \rangle + \Delta \Delta G_{solv} \quad (6)$$

In Eqs. (1)–(4) G_i is the free energy of system i , that being the ligand, the protein, or the complex; V_{MM} is the potential energy in vacuum as defined by the molecular mechanics (MM) model, which is composed of the bonded potential energy terms (V_{bonded}) and nonbonded Coulombic (V_{coul}) and Lennard–Jones (V_{LJ}) terms; S_{MM} is the entropy; ΔG_{solv} is the free energy of solvation for transferring the ligand from water in the binding area calculated using the PBSA model, composed by a polar (ΔG_P) and nonpolar (ΔG_{NP}) term; T is the temperature and angle brackets represent an ensemble average. Molecular mechanics energies for Lennard–Jones (V_{LJ}) and Coulombic electrostatic (Coul) V_{coul} were calculated with OPLS2005 [94] force field; in these calculations $\Delta V_{bonded} = 0$ as the single trajectory method was adopted and $\Delta V_{MM} = \Delta V_{LJ}$ and ΔV_{coul} . The polar part of the solvation free energy was determined by calculations using the Generalized-Born model. [95] The nonpolar term was considered proportional to the solvent accessible surface area (SASA), $\Delta G_{NP} = \gamma \cdot SASA$, where $\gamma = 0.0227 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$. Because the SQ109 analogues tested are very similar entropy term was neglected and ΔG_{bind} is termed as effective binding energy, ΔG_{eff} which is calculated according to Eq. (6). [96] We applied a dielectric constant $\epsilon_{solute} = 1$ to the binding area and to account for the lipophilic environment of the protein an heterogeneous dielectric implicit membrane model was used along the bilayer z-axis. [52–55]. The post-processing thermal_mmgbsa.py script of the Schrodinger Suite was used which takes snapshots from the MD simulations trajectory and calculated ΔG_{eff} according to Eq. (6).

Alchemical TI/MD binding free energies calculated with MBAR method

Method's principles

The TI/MD method has been described [61]. Free energy is a state function, and thus the free energy difference between states is independent of the path that connects them. To compare two ligands 0 and 1 binding to a receptor the calculation of $\Delta A_1(b)$ and $\Delta A_0(b)$, respectively, is needed and then the difference $\Delta \Delta A_{0 \rightarrow 1}(b)$ or $\Delta \Delta A_{0,1}(b)$. The calculation of $\Delta A_1(b)$ and $\Delta A_0(b)$ is computationally demanded and subjected to big errors because its includes large changes between the two states. Thus, the calculation of the relative binding free energies for two ligands bound to MmpL3 (for the 6 pairs of ligands shown in Table 5, respectively) was performed instead using the MBAR method [66] and applying a thermodynamic cycle. [33, 34, 97] (Fig. 5), i.e. using the ΔG values obtained for the transformations of the ligands in the bound (b) and the solvent (s), i.e. water states, respectively, $\Delta G_{0,1}(b)$ and $\Delta G_{0,1}(s)$, according to Eq. (7)

$$\Delta\Delta A_{b,0\rightarrow 1} \text{ or } \Delta\Delta A_{b,0,1} = \Delta A_1(b) - \Delta A_0(b) = \Delta A_{0,1}(b) - \Delta A_{0,1}(s) \quad (7)$$

Using this method, we calculated the difference between $\Delta A_{0,1}(b)$ and $\Delta A_{0,1}(s)$ which corresponds to the unphysical alchemical transformation $0 \rightarrow 1$ in the bound state and in the water state known as alchemical transformation which may be chosen to include small change (perturbation) of ligand structure, eg. H to CH₃ at 2-position of adamantyl group, to lower the error for the free energy perturbation calculation $\Delta A_{0,1}(b)$ or $\Delta A_{0,1}(s)$.

Because the phase space overlap between two states 0, 1 of interest can be near zero, doing free energy calculations for the two states alone will often have very large errors. Free energy is a state function, we can construct a thermodynamic path that takes us through a set of states that improves phase space overlap between states that can be unphysical. By this, we mean that intermediate states do not have to be observable experimentally. To put this mathematically, we can improve our results by constructing high phase space overlap intermediates and calculating our free energy difference $\Delta\Delta A_{0\rightarrow 1}$ by the sum of the binding free energy differences between the intermediate states.

Briefly, a thermodynamic parameter λ was used that smoothly connects states 0 and 1 through a λ -dependent potential $U(r^N; \lambda)$, such that $U(r^N; 0) = U_0(r^N)$ and $U(r^N; 1) = U_1(r^N)$. The transformation was broken down into a series of M steps corresponding to a set of λ values $\lambda_1, \lambda_2, \dots, \lambda_M$ ranging from 0 to 1, such that there was sufficient phase space overlap between neighboring intermediate λ states. TI computes the free energy change of transformation $0 \rightarrow 1$ by integrating the Boltzmann averaged $dU(\lambda)/d\lambda$ as is described in Eq. (8)

$$\Delta A_{0\rightarrow 1} = \int_0^1 d\lambda \left\langle \frac{dU(r^N; \lambda)}{d\lambda} \right\rangle_\lambda \quad (8)$$

$$= \Delta A_{0\rightarrow 1} \approx \sum_{k=1}^M w_k \left\langle \frac{dU(r^N; \lambda)}{d\lambda} \right\rangle_{\lambda_k}$$

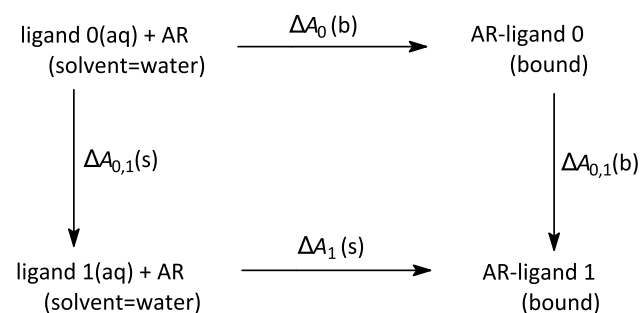


Fig. 5 Thermodynamic cycle used for the calculation of relative binding free energies

where the second sum indicates numerical integration over M quadrature points (λ_k , for $k = 1, \dots, M$) with associated weights w_k . A linear extrapolation between states can be applied for the construction of $U_1(r^N; \lambda)$ while with Amber18 softcore potentials [34, 59, 98] the LJ and Coulomb term potentials are described according to Eq. (9)

$$U(r^N; \lambda) = U_0^{SC}(r^N; \lambda) + \lambda \Delta U^{SC}(r^N; \lambda) \quad (9)$$

$$= U_0^{SC}(r^N; \lambda) + \lambda (U_1^{SC}(r^N; 1 - \lambda) - U_0^{SC}(r^N; \lambda))$$

MBAR [66] calculated the free energy difference between neighboring intermediate states using Eq. (10)

$$\Delta A_{\lambda \rightarrow \lambda+1} = -\ln \ln \frac{\langle w \exp(-\beta U_{\lambda+1}) \rangle_\lambda}{\langle w \exp(-\beta U_{\lambda+1}) \rangle_{\lambda+1}} \quad (10)$$

where w is a function of $A(\lambda)$ and $A(\lambda + 1)$. The equation was solved iteratively to give the free energy change of neighboring states $\Delta A(\lambda \rightarrow \lambda + 1)$, which via combination yielded the overall free energy change. MBAR method has been shown to minimize the variance in the calculated free energies, by making more efficient use of the simulation data [66, 99–101].

TI/MD simulations protocol

For the TI/MD calculations performed with ff14sb, [102] the relaxed complexes of compounds **1a-e**, **g**, **h** with MmpL3 from the 100 ns-MD simulations in a POPC lipid bilayer with the ff99sb [49] were used as starting structures for the alchemical calculations. The setup procedure was the same as previously reported for the MD simulations with Amber18 program. [84] TI/MD calculations were also performed for the ligands in solution. The bond constraint SHAKE [88] algorithm was disabled for TI mutations in AMBER GPU-TI module pmemdGTI, [103] and therefore a time step of 1 fs was used for all MD simulations.

Initial geometries were minimized using 20,000 steps of steepest descent minimization at $\lambda = 0.5$. These minimized geometries were then used for simulations at all λ values. Eleven λ values were used, equally spaced between 0.0 and 1.0. Each simulation was heated to 310 K over 500 ps using the Langevin thermostat, [104, 105] with a collision frequency set to 2 ps^{-1} . The Berendsen barostat [106] was used to adjust the density over 500 ps at constant pressure (NPT), with a target pressure of 1 bar and a 2 ps coupling time. Then, 500 ps of constant volume equilibration (NVT) was followed by 2 ns NVT production simulations. Energies were recorded every 1 ps, and coordinates were saved every 10 ps. Production simulations recalculated the

potential energy at each λ value every 1 ps for later analysis with MBAR [66].

For each calculation, the 1-step protocol was performed, ie. disappearing one ligand and appearing the other ligand simultaneously, and the electrostatic and Van der Waals interactions were scaled simultaneously using softcore potentials from real atoms that were transformed into dummy atoms. [34] We carried out the calculations using the 1-step protocol which changed charges and Van der Waals interactions in a single simulation by activating both Lennard–Jones and Coulomb softcore potentials simultaneously, reducing the computational cost. [107] The 1-step protocol offers a less computational expensive and more accurate approach to free energy estimates according to recent studies. [98]

The final states 0 and 1 of the alchemical calculations $0 \rightarrow 1$ or $1 \rightarrow 0$, ie. the structures of ligand 0-AR and 1-AR complexes as resulted from the alchemical transformations were compared with these complexes structure resulted from converged 100 ns-MD simulations. This was performed to certify that the 2 ns MD simulation for each λ -state during the alchemical calculations was enough for the complexes 0-AR and 1-AR to converge to same structure with 100 ns-MD simulations. Two repeats were performed for the TI/MD calculation for each alchemical transformation (Table 5).

In summary, we performed for the $\sim 100,000$ atoms protein complexes studied here for the single protonation state 100 ns-MD simulations in 2 repeats \times 9 ligands (18 MD simulations) and for the double protonation state 80 ns-MD simulations in 9 ligands (9 MD simulations). Additionally, we performed for 4 representative ligands 500 ns-MD simulations (4.52 μ s MD simulation time). We performed the simulations using Desmond program, which performed much faster than Amber or Gromacs programs using an amber force field (ff99sb) that can fairly describe the protein conformation.

We tested the MM-GBSA calculations using ensembles from 20 ns-MD simulations with ff99sb for 10 ligands in 3 repeats \times 2 protonation states (60 MD simulations). For the monoprotonated form of the ligands we tested an environment without or with an implicit model for membrane (2 environments). The simulation time for these simulations was 2.4 μ s. We calculated the MM-GBSA interaction energies with free available OPLS2005 force field with Desmond software.

We used the last snapshot of the converged MD simulations and performed TI/MD simulations to calculate the relative binding free energies using alchemical perturbations and the amber software and ff14sb. Thus, we performed 2 repeats \times 6 ligands \times 10- λ values (120 2 ns-MD simulations). The total simulation time for the simulations performed in this study was 7.16 μ s.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10822-023-00504-6>.

Acknowledgements We are grateful to Chiesi Hellas for supporting this research. This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility ARIS (pr010007) to AK.

Author contributions This research is part of the PhD thesis of MS. AK conceived and supervised the project. The binding affinities were measured in HIZ lab. MS performed the biomolecular simulations and IS performed the DFT calculations. MS and AK wrote the manuscript.

Funding Open access funding provided by HEAL-Link Greece.

Declarations

Conflict of interest The authors declare no competing financial interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Walker TM et al (2022) The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: a genotypic analysis. *The Lancet Microbe* 3:e265–e273
- Lee RE et al (2003) Combinatorial lead optimization of [1,2]-diamines based on ethambutol as potential antituberculosis preclinical candidates. *J Comb Chem* 5:172–187
- Heinrich N et al (2015) Early phase evaluation of SQ109 alone and in combination with rifampicin in pulmonary TB patients. *J Antimicrob Chemother* 70:1558–1566
- Sacksteder KA, Protopopova M, Barry CE, Andries K, Nacy CA (2012) Discovery and development of SQ109: a new antitubercular drug with a novel mechanism of action. *Future Microbiol* 7:823–837
- Tahlan K et al (2012) SQ109 Targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of mycobacterium tuberculosis. *Antimicrob Agents Chemother* 56:1797–1809
- Grzegorzewicz AE et al (2012) Inhibition of mycolic acid transport across the Mycobacterium tuberculosis plasma membrane. *Nat Chem Biol* 8:334–341
- Li W et al (2014) Novel insights into the mechanism of inhibition of MmpL3, a target of multiple pharmacophores in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 58:6413–6423
- Malwal SR et al (2021) Structure, in vivo detection, and antibacterial activity of metabolites of SQ109, an anti-infective drug candidate. *ACS Infect Dis* 7:2492–2507

9. Meng Q, Luo H, Chen Y, Wang T, Yao Q (2009) Synthesis of novel [1,2]-diamines with antituberculosis activity. *Bioorg Med Chem Lett* 19:5372–5375
10. Onajole OK et al (2010) Synthesis and evaluation of {SQ}109 analogues as potential anti-tuberculosis candidates. *Eur J Med Chem* 45:2075–2079
11. Onajole OK et al (2011) SQ109 analogues as potential antimicrobial candidates. *Med Chem Res* 20:1394–1401
12. Li K et al (2015) Oxa, Thia, heterocycle, and carborane analogues of SQ109: bacterial and protozoal cell growth inhibitors. *ACS Infect Dis* 1:215–221
13. Stampolaki M et al (2023) Synthesis and Testing of Analogs of the Tuberculosis Drug Candidate SQ109 against Bacteria and Protozoa: Identification of Lead Compounds against Mycobacterium abscessus and Malaria Parasites. *ACS Infect Dis* 9:342–364
14. Zhang L et al (2020) Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol. *Science* 368:1211–1219
15. Domenech P, Reed MB, Barry CE (2005) Contribution of the Mycobacterium tuberculosis MmpL protein family to virulence and drug resistance. *Infect Immun* 73:3492–3501
16. Viljoen A et al (2017) The diverse family of MmpL transporters in mycobacteria: from regulation to antimicrobial developments. *Mol Microbiol* 104:889–904
17. Zhang B et al (2019) Crystal structures of membrane transporter MmpL3, an anti-TB drug target. *Cell* 176:636
18. Bernut A et al (2016) Insights into the smooth-to-rough transition in Mycobacterium boletii unravels a functional Tyr residue conserved in all mycobacterial MmpL family members. *Mol Microbiol* 99:866–883
19. Su C-C et al (2019) MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine. *Proc Natl Acad Sci USA* 116:11241–11246
20. Adams, O. et al. (2021) Cryo-EM structure and resistance landscape of M. tuberculosis MmpL3: An emergent therapeutic target. *Structure* 29, 1182–1191.e4.
21. Su CC et al (2021) Structures of the mycobacterial membrane protein MmpL3 reveal its mechanism of lipid transport. *PLOS Biol* 19:e3001370
22. Yang X et al (2020) Structural basis for the inhibition of mycobacterial MmpL3 by NITD-349 and SPIRO. *J Mol Biol* 432:4426–4434
23. Li W et al (2019) Direct inhibition of MmpL3 by novel antitubercular compounds. *ACS Infect Dis* 5:1001–1012
24. Shetty A et al (2018) Novel acetamide indirectly targets mycobacterial transporter MmpL3 by proton motive force disruption. *Front Microbiol* 9:2960
25. Xu Z, Meshcheryakov VA, Poce G, Chng S-S (2017) {MmpL}3 is the flippase for mycolic acids in mycobacteria. *Proc Natl Acad Sci* 114:7993–7998
26. Li K et al (2014) Multitarget drug discovery for tuberculosis and other infectious diseases. *J Med Chem* 57:3126–3129
27. Stevens CM et al (2022) Proton transfer activity of the reconstituted Mycobacterium tuberculosis MmpL3 is modulated by substrate mimics and inhibitors. *Proc Natl Acad Sci U S A* 119:e2113963119
28. Qiu W, Guo Y (2022) Analysis of the oligomeric state of mycobacterial membrane protein large 3 and its interaction with SQ109 with native cell membrane nanoparticles system. *Biochim Biophys Acta - Biomembr* 1864:183793
29. Massova I, Kollman PA (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect Drug Discovery Des* 18:113–135
30. Weng G et al (2019) Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein-peptide complexes. *Phys Chem Chem Phys* 21:10135–10145
31. Wang E et al (2019) End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev* 119:9478–9508
32. Lenselink EB et al (2016) Predicting binding affinities for GPCR ligands using free-energy perturbation. *ACS Omega* 1:293–304
33. He X et al (2020) Fast, accurate, and reliable protocols for routine calculations of protein-ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* 5:4611–4619
34. Song LF, Lee T-S, Zhu C, York DM, Merz KM (2019) Using AMBER18 for relative free energy calculations. *J Chem Inf Model* 59:3128–3135
35. Stampelou M et al (2022) Dual A1/A3 adenosine receptor antagonists: binding kinetics and structure-activity relationship studies using mutagenesis and alchemical binding free energy calculations. *J Med Chem* 65:13305–13327
36. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652
37. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37:785–789
38. Francl MM et al (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J Chem Phys* 77:3654
39. Grimme S, Ehrlich S, Goerigk L (2011) Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 32:1456–1465
40. Davidson ER, Feller D (1986) Basis set selection for molecular calculations. *Chem Rev* 86:681–696
41. Tomasi J, Mennucci B, Cammi R (2005) Quantum Mechanical Continuum Solvation Models. *Chem Rev* 105:2999–3093
42. York DM, Karplus M (1999) A smooth solvation potential based on the conductor-like screening model. *J Phys Chem A*. 103:11060–11079
43. Weast RC (1969) Handbook of chemistry and physics Am J Med Sci 49
44. Peach ML, Cachau RE, Nicklaus MC (2017) Conformational energy range of ligands in protein crystal structures: The difficult quest for accurate understanding. *J Mol Recognit* 30:e2618
45. Robertson MJ, Meyerowitz JG, Panova O, Borrelli K, Skiniotis G (2022) P. Nat Struct Mol Biol 29: 210–217
46. Fusani L, Palmer DS, Somers DO, Wall ID (2020) Exploring ligand stability in protein crystal structures using binding pose metadynamics. *J Chem Inf Model* 60:1528–1539
47. Malde AK, Mark AE (2011) Challenges in the determination of the binding modes of non-standard ligands in X-ray crystal complexes. *J Comput Aided Mol Des* 25:1–12
48. van Zundert GCP, Moriarty NW, Sobolev OV, Adams PD, Borrelli KW (2021) Macromolecular refinement of X-ray and cryoelectron microscopy structures with Phenix/OPLS3e for improved structure and ligand quality. *Structure* 29:913–921.e4
49. Hornak V et al (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Genet* 65:712–725
50. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring functions: I The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 11:425–445
51. Rizzo RC, Jorgensen WL (1999) OPLS all-atom model for amines: Resolution of the amine hydration problem. *J Am Chem Soc* 121:4827–4836
52. Greene D, Qi R, Nguyen R, Qiu T, Luo R (2019) Heterogeneous dielectric implicit membrane model for the

- calculation of MMPBSA binding free energies. *J Chem Inf Model* 59:3041–3056
53. Botello-Smith WM et al (2013) Numerical poisson-boltzmann model for continuum membrane systems. *Chem Phys Lett* 555:274–281
54. Botello-Smith WM, Luo R (2015) Applications of MMPBSA to Membrane Proteins I: efficient numerical solutions of periodic poisson-boltzmann equation. *J Chem Inf Model* 55:2187–2199
55. Xiao L, Diao J, Greene D, Wang J, Luo R (2017) A continuum poisson-boltzmann model for membrane channel proteins. *J Chem Theory Comput* 13:3398–3412
56. Aldeghi M, Bodkin MJ, Knapp S, Biggin PC (2017) Statistical analysis on the performance of molecular mechanics poisson-boltzmann surface area versus absolute binding free energy calculations: bromodomains as a case study. *J Chem Inf Model* 57:2203–2221
57. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10:449–461
58. Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. *J Phys Chem B* 114:10235–10253
59. Lee TS et al (2020) Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J Chem Inf Model* 60:5595–5623
60. Cournia Z, Allen B, Sherman W (2017) Relative binding free energy calculations in drug discovery: recent advances and practical considerations. *J Chem Inf Model* 57:2911–2937
61. Kirkwood JG (1935) Statistical mechanics of fluid mixtures. *J Chem Phys* 3:300–313
62. Deflorian F et al (2020) Accurate prediction of GPCR ligand binding affinity with free energy perturbation. *J Chem Inf Model* 60:5563–5579
63. Matricon P et al (2017) Fragment optimization for GPCRs by molecular dynamics free energy calculations: Probing druggable subpockets of the A_{2A} adenosine receptor binding site. *Sci Reports* 7:6398
64. Wan S et al (2020) Hit-to-lead and lead optimization binding free energy calculations for G protein-coupled receptors. *Interface Focus* 10:20190128
65. Keränen H, Åqvist J, Gutiérrez-De-Terán H (2015) Free energy calculations of A_{2A} adenosine receptor mutation effects on agonist binding. *Chem Commun* 51:3522–3525
66. Procacci P (2013) Multiple Bennett acceptance ratio made easy for replica exchange simulations. *J Chem Phys* 139:124105
67. Feng X et al (2015) Antiinfectives targeting enzymes and the proton motive force. *Proc Natl Acad Sci U S A* 112:E7073–E7082
68. Varela C et al (2012) MmpL genes are associated with mycolic acid metabolism in mycobacteria and corynebacteria. *Chem Biol* 19:498–506
69. Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 132:154104
70. Frisch, M., Trucks, G., Schlegel, H. & Scuseria, G. Gaussian 03, revision C. 02; Gaussian, Inc.: Wallingford, CT, 2004. (2013).
71. Halgren TTA (1996) Merck molecular force field I Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 17:490–519
72. Shelley JC et al (2007) Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* 21:681–691
73. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 27:221–234
74. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236
75. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487
76. Mohamadi F et al (1990) Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J Comput Chem* 11:440–467
77. Jones G, Willett P, Glen RC, Leach R, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748
78. Pettersen EF et al (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
79. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res* 40:D370–D376
80. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935
81. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general Amber force field. *J Comput Chem* 25:1157–1174
82. Wang J, Cieplak P, Kollman PA (2000) How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J Comput Chem* 21:1049–1074
83. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J Phys Chem* 97:10269–10280
84. Case DA et al (2018) AMBER 2018. University of California, University of California, San Francisco
85. Koynova R, Caffrey M (1998) Phases and phase transitions of the phosphatidylcholines. *Biochim Biophys Acta - Rev Biomembr* 1376:91–145
86. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
87. Essmann U et al (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593
88. 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* 23:327–341
89. Martyna GJ, Klein ML, Tuckerman M (1992) Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J Chem Phys* 97:2635–2643
90. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101:4177–4189
91. Humphreys DD, Friesner RA, Berne BJ (1994) A multiple-time-step molecular dynamics algorithm for macromolecules. *J Phys Chem* 98:6885–6892
92. Maffucci I, Contini A (2013) Explicit ligand hydration shells improve the correlation between MM-PB/GBSA binding energies and experimental activities. *J Chem Theory Comput* 9:2706–2717
93. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* 32:2319–2327
94. Banks JL et al (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* 26:1752–1780

95. Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA continuum model for solvation a fast analytical method for the calculation of approximate Born radii. *J Phys Chem A* 101:3005–3014
96. Gohlke, H., Case, D. A., Biology, M., Scripps, T. & Rd, N. T. P. Converging Free Energy Estimates : MM-PB (GB) SA Studies on the Protein – Protein Complex Ras – Raf. 238–250 (2003).
97. Genheden S, Nilsson I, Ryde U (2011) Binding affinities of factor Xa inhibitors estimated by thermodynamic integration and MM/GBSA. *J Chem Inf Model* 51:947–958
98. Steinbrecher T, Joung I, Case DA (2011) Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations. *J Comput Chem* 32:3253–3263
99. Shirts MR, Pande VS (2005) Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J Chem Phys* 122:144107
100. Paliwal H, Shirts MR (2011) A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods. *J Chem Theory Comput* 7:4115–4134
101. Tan Z, Gallicchio E, Lapelosa M, Levy RM (2012) Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J Chem Phys* 136:144102
102. Maier JA et al (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11:3696–3713
103. Lee T-S, Hu Y, Sherborne B, Guo Z, York DM (2017) Toward fast and accurate binding affinity prediction with pmemdGTI: an efficient implementation of GPU-accelerated thermodynamic integration. *J Chem Theory Comput* 13:3077–3084
104. Lzaguirre JA, Catarello DP, Wozniak JM, Skeel RD (2001) Langevin stabilization of molecular dynamics. *J Chem Phys* 114:2090–2098
105. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* 103:4613
106. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690
107. Su P-C, Johnson ME (2016) Evaluating thermodynamic integration performance of the new amber molecular dynamics package and assess potential halogen bonds of enoyl-ACP reductase (FabI) benzimidazole inhibitors. *J Comput Chem* 37:836–847

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.