Conformational Shifts of Stacked Heteroaromatics: Vacuum vs. Water Studied by Machine Learning

Johannes R. Loeffler, Monica L. Fernández-Quintero, Franz Waibl, Patrick K. Quoika, Florian Hofer, Michael Schauperl and Klaus R. Liedl*

Center of Molecular Biosciences Innsbruck, Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria

Stacking interactions play a crucial role in drug design, as we can find aromatic cores or scaffolds in almost any available small molecule drug. To predict optimal binding geometries and enhance stacking interactions, usually high-level quantum mechanical calculations are performed. These calculations have two major drawbacks: they are very time consuming, and solvation can only be considered using implicit solvation. Therefore, most calculations are performed in vacuum. However, recent studies have revealed a direct correlation between the desolvation penalty, vacuum stacking interactions and binding affinity, making predictions even more difficult. To overcome the drawbacks of quantum mechanical calculations, in this study we use neural networks to perform fast geometry optimizations and molecular dynamics simulations of heteroaromatics stacked with toluene in vacuum and in explicit solvation. We show that the resulting energies in vacuum are in good agreement with high-level quantum mechanical calculations. Furthermore, we show that using explicit solvation substantially influences the favored orientations of heteroaromatic rings thereby emphasizing the necessity to include solvation properties starting from the earliest phases of drug design.

Keywords: machine learning, stacking, solvation, heteroaromatics, ANI

INTRODUCTION

Binding between targets and small molecule drugs depends on a small set of specific interactions (Bissantz et al., 2010). In structure-based drug design, the main goal is to optimize a small molecule to make use of all possible interaction sites provided by the protein's binding pocket (Bissantz et al., 2010; Kuhn et al., 2011). Computer simulations of protein ligand complexes and various approaches to predict the binding free energy are readily used in the drug design process (Chang et al., 2007; Chodera et al., 2011; Mobley and Klimovich, 2012; Limongelli et al., 2013; Hansen and Van, 2014). However, certain interactions, e.g., π-π stacking of heteroaromatics, are not properly parametrized in modern force fields to reliably make free energy estimations. Yet, these interactions play a major role in drug design (Burley and Petsko, 1985; Meyer et al., 2003; Williams et al., 2003; Adhikary et al., 2019). Heteroaromatic moieties or cores are found in the majority of drug molecules (Meyer et al., 2003; Wang et al., 2017) as they present ideal modification sites and allow for unique interactions, i.e., stacking (Meyer et al., 2003; Salonen et al., 2011). Stacking can occur as π-π (Huber et al., 2014), halogen-π (Wallnoefer et al., 2010), amide-π (Harder et al., 2013; Bootsma and Wheeler, 2018), cation-π (Gallivan and Dougherty, 1999), and even anion-π (Wheeler and Bloom, 2014) interactions.

The state-of-the-art approach to estimate stacking interactions and to identify favorable geometries is the application of high-level quantum mechanical calculations. This can either be done by using a grid-based approach (Huber et al., 2014; Bootsma et al., 2019) or by using descriptors derived from high-level quantum mechanical calculations (Bootsma and Wheeler, 2011). However, to obtain interaction energies via a grid-based approach, numerous calculations have to be performed and molecules are restrained in single-point calculations (Huber et al., 2014). Furthermore, these calculations are almost exclusively performed in vacuum or implicit solvent. Nevertheless, several studies have investigated the effect of solvation on stacking interactions and the resulting implications on thermodynamic properties (Kolář et al., 2011; Lee et al., 2019; Loeffler et al., 2020). In general, assessment of the desolvation penalty is crucial in drug design as it can reveal why certain molecules do not reflect the expected gain in binding affinity (Biela et al., 2012; Dobiaš et al., 2019; Loeffler et al., 2020). Therefore, a combination of approaches is inevitable to understand the energetics of molecules and to interpret and optimize SAR studies (Loeffler et al., 2020). Since quantum mechanical calculations come with an extreme computational cost, several ways to minimize calculation time have been developed, including fragmentation (Kitaura et al., 1999), semi-empirical methods (Dewar et al., 1985; Elstner, 2006; Stewart, 2009) and recently machine learning approaches (Smith et al., 2017, 2018). Machine learning is a powerful tool and has already been applied to address various challenges in chemistry, e.g., the prediction of binding affinity (Nguyen et al., 2019), atomic forces, nuclear magnetic resonance shifts (Ghosh and Hammes-Schiffer, 2015), and even the prediction of reaction pathways (Jiang et al., 2016). Additionally, it has been shown, that machine learning approaches allow substantially faster predictions of quantum mechanically calculated potential energy surfaces (Chmiela et al., 2017; Schütt et al., 2017; Smith et al., 2018; Yao et al., 2018), geometries and atomic charge models (Smith et al., 2017). In recent years, potentials based on deep neural networks have been developed and have already widely been applied to tackle several challenges, as they promise quantum accuracy at classical cost (Smith et al., 2017; Wang et al., 2018; Wang and Riniker, 2019; Xu et al., 2019; Ghanbarpour et al., 2020). These neural networks, in particular, the ANAKIN-ME (Accurate NeurAl networK engINe for Molecular Energies)—short ANI (Chmiela et al., 2017; Smith et al., 2017), have been trained to learn the potential energy surfaces. Similar to classical force fields, electrons are not treated explicitly in ANI. Additionally, the potential energy is calculated as a function of the geometric positions of atoms. In contrast, ANI does not use predefined properties such as atomic bonds, as in quantum mechanical calculations, and the energies in ANI are an artificial neural network. As the energy is not obtained by solving the Schroedinger equation, the computational effort of ANI is substantially reduced when compared to high-level QM calculations (Gao et al., 2020). From the potential energy surfaces

of organic molecules in a transferable way, including both the conformational and configurational space, ANI is able to predict the potential energy for molecules outside the training set.

To investigate protein-ligand interactions molecular dynamics simulations are a standard tool in computational drug design (Michel and Essex, 2010). Usually additive force fields are used to study the dynamic properties of proteins (Tian et al., 2020). These approaches are well-suited to describe protein properties and give valuable insights to all kinds of properties including flexibility (Fernández-Quintero et al., 2019a) and plasticity of binding sites (Fernández-Quintero et al., 2019b) and protein-protein interfaces (Fernández-Quintero et al., 2020). Using computer simulations requires a balance between cost and accuracy. Compared to classical force fields, quantum-mechanical methods are highly accurate but computationally expensive and not feasible for large systems. In classical force fields, stacking interactions of heterocycles with aromatic amino acid sidechains are still challenging to describe (Sherrill et al., 2009; Prampolini et al., 2015). Therefore, studies on stacking interactions almost exclusively rely on high-level quantum mechanical calculations (Bootsma and Wheeler, 2011, 2018; Huber et al., 2014; Bootsma et al., 2019). The use of Machine learning combines the best of both approaches.

In this study we make use of the ANI potentials to calculate stacking interactions of heteroaromatics frequently occurring in drug design projects. We compare the calculated minimal energies with high-level quantum mechanical calculations in vacuum and in implicit solvation. Furthermore, we perform molecular dynamics simulations to generate an ensemble of energetically favorable and unfavorable conformations of heteroaromatics interacting with a truncated phenylalanine side chain, i.e., toluene, in vacuum and explicit solvation.
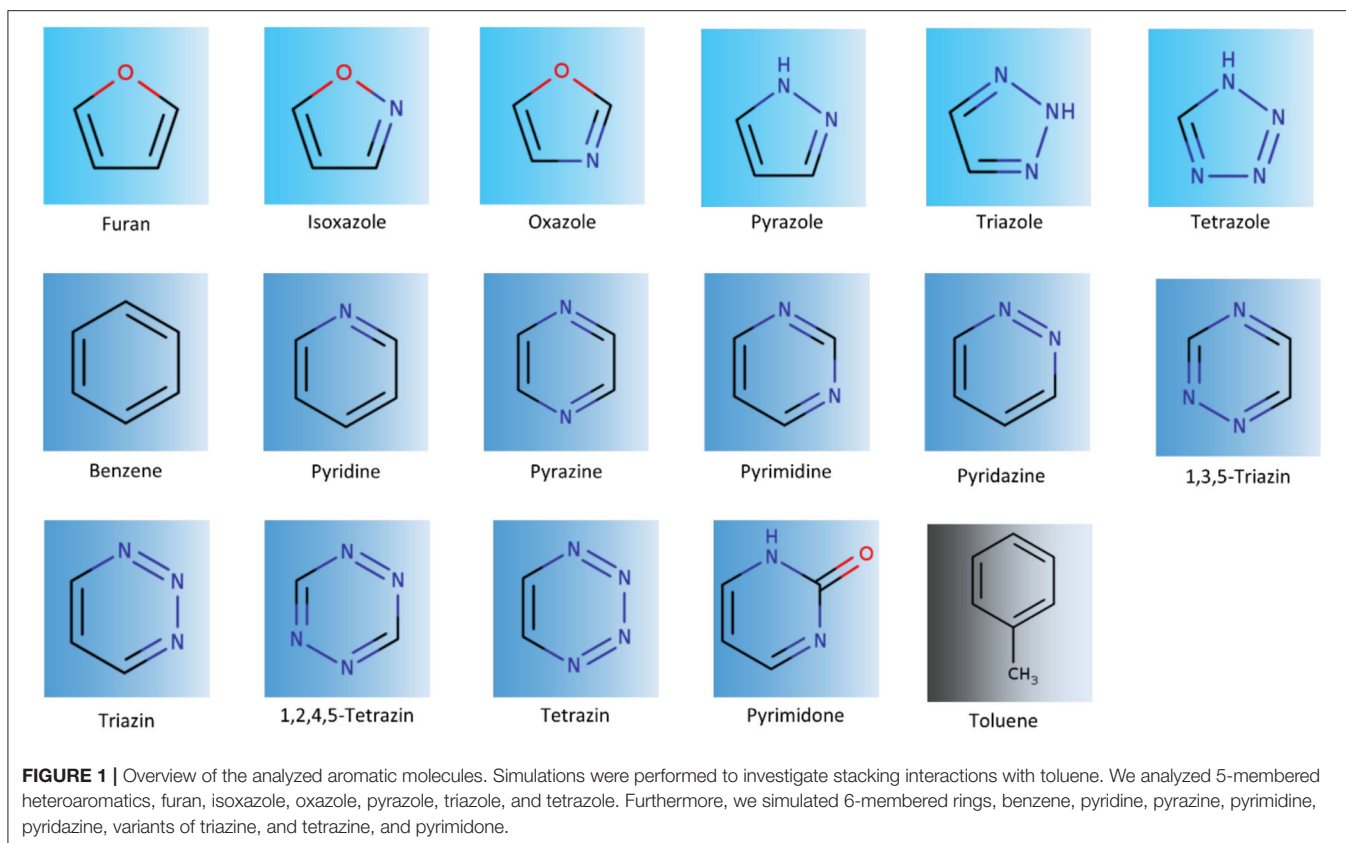
METHODS

Data Set
The set of molecules investigated in this study frequently occurs in drug molecules (Salonen et al., 2011) and has already been investigated in previous publications to characterize their stacking properties using quantum mechanical calculations and molecular mechanics based calculations to estimate their respective solvation properties as monomers as well as complexes (Huber et al., 2014; Bootsma et al., 2019; Loeffler et al., 2019) (**Figure 1**).

Quantum Mechanical Calculations
We followed the protocol recently introduced to perform energy optimization of heteroaromatics with toluene using Gaussian09 (Frisch et al., 2009) at the ωB97XD (Chai and Head-Gordon, 2008)/cc-pVTZ (Dunning, 1989) level. This combination has been benchmarked by Huber et al. (2014) and has been used in recent publications addressing similar questions (Loeffler et al., 2019, 2020). To better compare the geometries resulting from the simulations in water, we performed the geometry optimizations using an implicit water model. We used the polarizable continuum model, a reaction field calculation using the integral

---

FIGURE 1 | Overview of the analyzed aromatic molecules. Simulations were performed to investigate stacking interactions with toluene. We analyzed 5-membered heteroaromatics, furan, isoxazole, oxazole, pyrazole, triazole, and tetrazole. Furthermore, we simulated 6-membered rings, benzene, pyridine, pyrazine, pyrimidine, pyridazine, variants of triazine, and tetrazine, and pyrimidone.

equation formalism (Tomasi et al., 2005) implemented in Gaussian09 (Frisch et al., 2009).

ANI

This approach makes use of the Behler Parrinello symmetry functions to compute an atomic environment vector (AEV), G_i^X, which is composed of all elements, G_M probing regions of an atoms chemical surroundings. Each E_i^x is then used as input to a single neural network potential. The energy of a molecule is caluclated as the sum of all individual neural network potentials (**Supplementary Figure 1**).

The summation formalism to calculate E_τ shows two major advantages. Firstly, it allows fortransferability, and secondly, an even greater advantage is that due to the simple formalisma near linear scaling in computational complexity with added cores and/or GPUs is possible (**Supplementary Figure 1**).
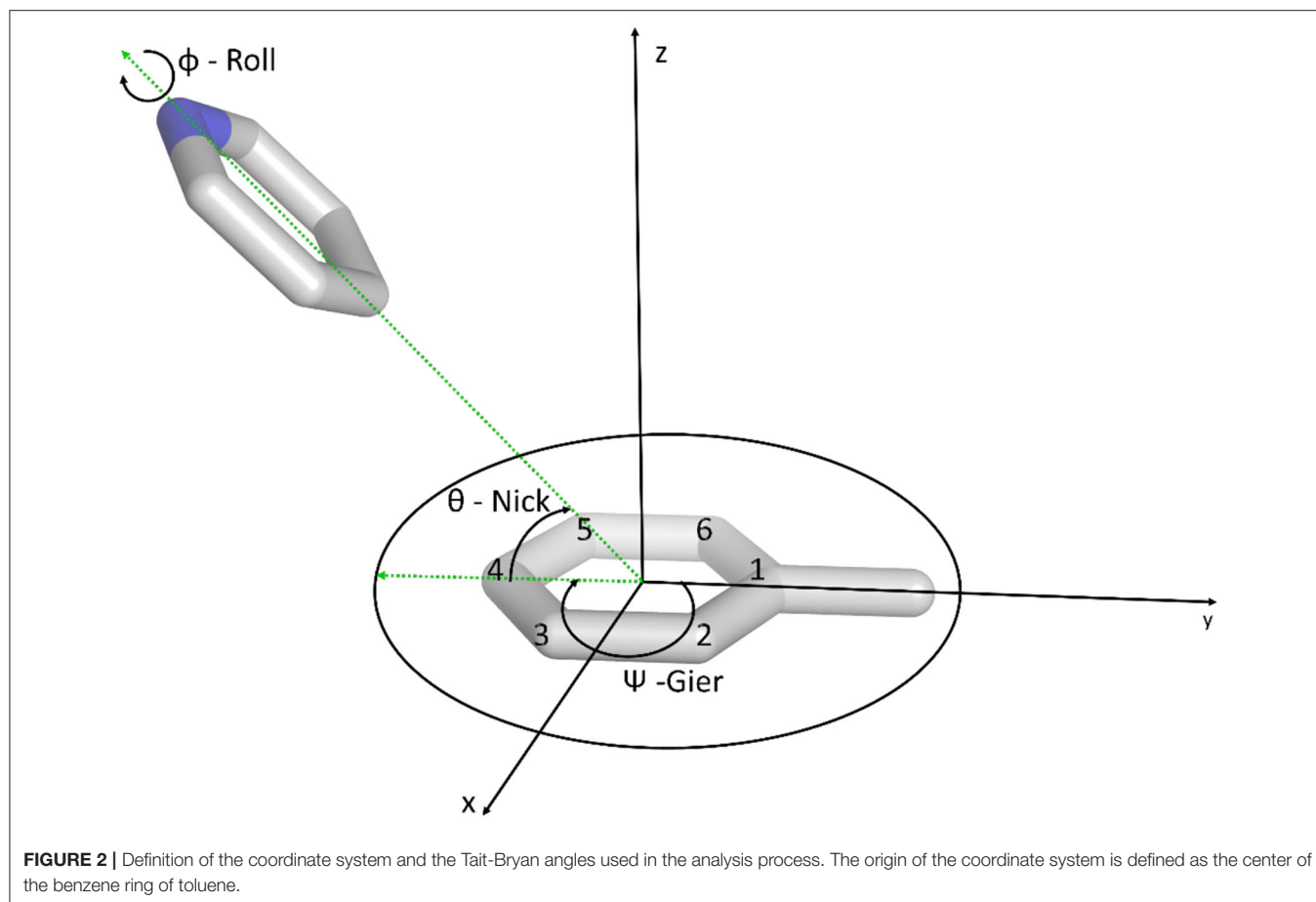
Simulation Setup

As starting structures for the simulations we used the minimum energy conformations provided in xyz-format in the **Supplementary Material** in the paper published by Bootsma et al. (2019). We solvated these conformations in a water box with a minimum wall distance of 10 Å using tleap resulting in approximately 1500 explicit water molecules (Case et al., 2018). To equilibrate the water box we performed a restrained equilibration allowing only the water molecules within the box to move as suggested in previous publications. During the

equilibration performed with the AMBER simulation package we restrained the aromatic molecules to keep the geometry obtained from high-level QM calculations. The final frame of the equilibration was then used as starting structure of the production run. For each step of the simulations we calculated the forces and energies using ANI (Smith et al., 2017). To perform the simulations we used the atomic simulation environment (ASE) engine, protocol included in the **Supplementary Material** (Larsen et al., 2017). We used a timestep of 0.25 fs. To keep the temperature constant at 300 K we used the Langevin algorithm with a friction coefficient of 0.02 atomic units. We employed periodic boundary conditions in x, y, and z directions. We performed a short LBFGS (Head and Zerner, 1985) optimization before initiating the production runs of 100 ps. We performed this setup 10 times with different starting velocities for each heteroaromatic molecule.

Vacuum Interaction Energies

To calculate the interaction energies in vacuum we performed the geometry optimization of the complexes and the respective monomers individually. These calculations were performed for force fields using MOE, for QM using Gaussian09 and for the ANI potentials using the ASE environment. The vacuum stacking interaction energies were then calculated according to the supermolecular approach as previously published. It has been shown that Counterpoise-corrections can result in distortions of the hypersurface (Liedl, 1998). Thus, and to allow for better

FIGURE 2 | Definition of the coordinate system and the Tait-Bryan angles used in the analysis process. The origin of the coordinate system is defined as the center of the benzene ring of toluene.

comparability with the previous results no BSSE-correction was performed.

$$E_{interaction} = E_{complex} - E_{monomer\ A} - E_{monomer\ B} \quad (1)$$

Trajectory Analysis

The orientation of the stacked molecule during the simulation relative to the reference was described in terms of the Tait-Bryan angles (Markley and Crassidis, 2014). We especially focused on the *nick* and *gier* angles, as shown in **Figure 2**. Therefore, a reference coordinate system was defined using the toluene orientation. The y-axis is positioned in the direction from the ring C4 atom (para position) to the methyl carbon atom (cf. **Figure 2**). The x-axis was initially positioned in the direction from the center of mass of the C2 and C3 to the center of mass of the C4 and C5 atoms. From these two vectors we calculated the z-axis as the resulting cross product. The direction was chosen to obtain a right-handed coordinate system. To ensure an orthogonal coordinate system we recalculated the x-axis as the cross product of the y- and z-axis. The origin of the coordinate system was defined as the center of mass (COM) of the aromatic ring of the toluene molecule.

We aligned the obtained trajectories on the toluene molecule and then transformed the coordinates of the stacking

heteroaromatic molecule into the previously introduced coordinate system. Furthermore, we assigned a "nose" vector **r**. The atoms chosen for each molecule can be found in **Supplementary Figure 1**. The vector **r** was normalized to length 1, and the *nick angle* θ and *gier* angle Ψ were calculated as follows.

$$nick\ (\theta) = arcos\ (r_z) \cdot \frac{180}{\pi} - 90 \quad (2)$$

$$gier\ (\Psi) = \arctan\left(\frac{r_x}{r_y}\right) \cdot \frac{180}{\pi} \quad (3)$$

These angles were used to describe the molecular orientation in reference to the toluene molecule. In all frames where the center of mass was in the negative z-direction, the z-component of **r** was reversed, corresponding to mirroring the molecule by the xy-plane, i.e., the plane of the aromatic toluene (cf. **Figure 2**). Free energy profiles of the nick and gier angles obtained from kernel density estimation (KDE) with a kernel width of 0.1 radians.
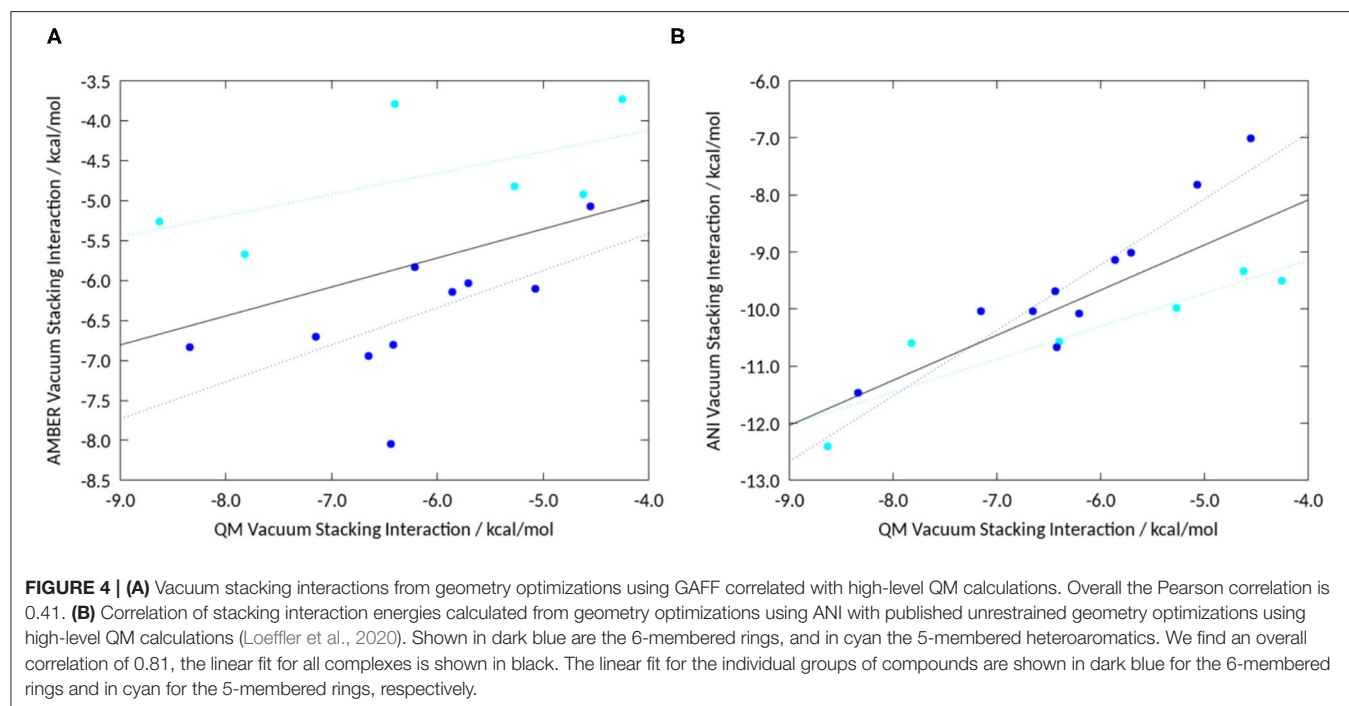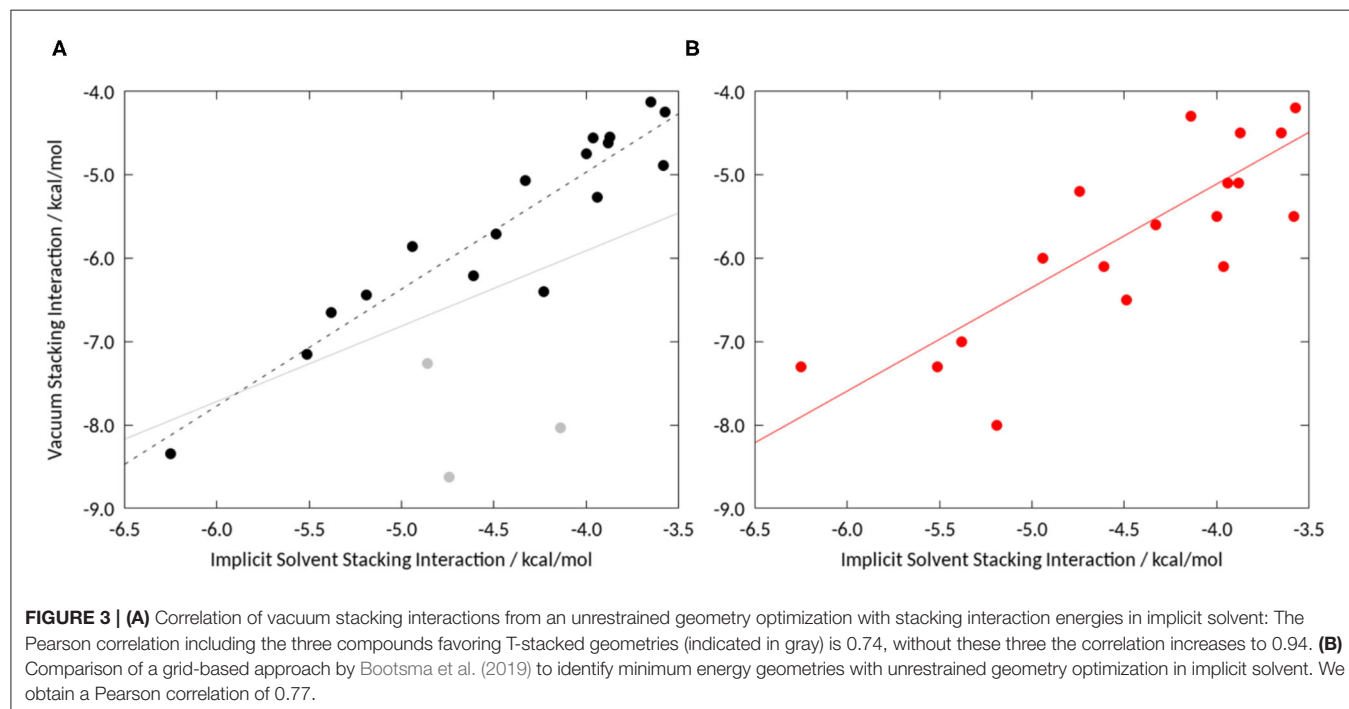
RESULTS

Geometry Optimizations

To assess the influence of solvation we initially performed unrestrained geometry optimizations, starting from the

geometries provided by Bootsma et al. (2019), in implicit solvent using the quantum mechanical setup as described in the Methods section. We investigated the stacking interactions of a set of compounds that was recently studied in two publications on a truncated phenylalanine sidechain, i.e., toluene (Bootsma et al., 2019; Loeffler et al., 2020). Comparing the resulting stacking interaction energies, we find a Pearson correlation of 0.74 for

the grid based approach (Bootsma et al., 2019) and 0.68 for the unrestrained energy optimizations (Loeffler et al., 2020). Comparing the obtained geometries, it is particularly striking that the compounds that prefer a T-stacked geometry in vacuum show a parallel displaced conformation in implicit solvent. If these compounds, (L09, L10, and L13), are excluded the correlation increases to 0.94 (see **Figure 3A**). This shows that

FIGURE 3 | (A) Correlation of vacuum stacking interactions from an unrestrained geometry optimization with stacking interaction energies in implicit solvent: The Pearson correlation including the three compounds favoring T-stacked geometries (indicated in gray) is 0.74, without these three the correlation increases to 0.94. **(B)** Comparison of a grid-based approach by Bootsma et al. (2019) to identify minimum energy geometries with unrestrained geometry optimization in implicit solvent. We obtain a Pearson correlation of 0.77.

FIGURE 4 | (A) Vacuum stacking interactions from geometry optimizations using GAFF correlated with high-level QM calculations. Overall the Pearson correlation is 0.41. **(B)** Correlation of stacking interaction energies calculated from geometry optimizations using ANI with published unrestrained geometry optimizations using high-level QM calculations (Loeffler et al., 2020). Shown in dark blue are the 6-membered rings, and in cyan the 5-membered heteroaromatics. We find an overall correlation of 0.81, the linear fit for all complexes is shown in black. The linear fit for the individual groups of compounds are shown in dark blue for the 6-membered rings and in cyan for the 5-membered rings, respectively.

even continuum models allow for different optimal stacking geometries compared, especially if T-stacked geometries are favored in vacuum.

Besides the high-level quantum mechanical calculations we performed simulations in water and vacuum using ANI (Smith et al., 2017) and the General Amber Force Field (GAFF) (Wang et al., 2004). To compare the resulting stacking interaction energies from ANI and GAFF with published QM data, we performed geometry optimizations of the complexes and of respective monomers (**Supplementary Table 1**). The stacking interaction was then calculated via the supermolecular approach, Equation (1) (Beljonne et al., 2000).

For the GAFF stacking interactions, we obtained an overall Pearson correlation of 0.41, as shown in **Figure 4A**. The lack of correlation between GAFF and QM data emphasizes that stacking interaction of different heteroaromatics with benzene is not well-parametrized in classical force field-based approaches. Individually, for the 5-membered rings the correlation increases to 0.61 and for the 6-membered rings to 0.60, indicated by the cyan and dark blue line in **Figure 4A**. The overall Pearson correlation for our set of compounds of QM vacuum stacking interactions with ANI stacking interactions results in 0.81. By only taking the 6 membered rings into account, the correlation increases to 0.93, depicted by the dark blue line in **Figure 4B**. For the 5-membered rings alone the correlation results in 0.91 (**Figure 4**-cyan line). The comparison between the results obtained with different methods is summarized in **Supplementary Table 1**.

Molecular Dynamics Simulations

As starting geometries for the molecular dynamics simulations we used the optimized structures published by Bootsma et al.
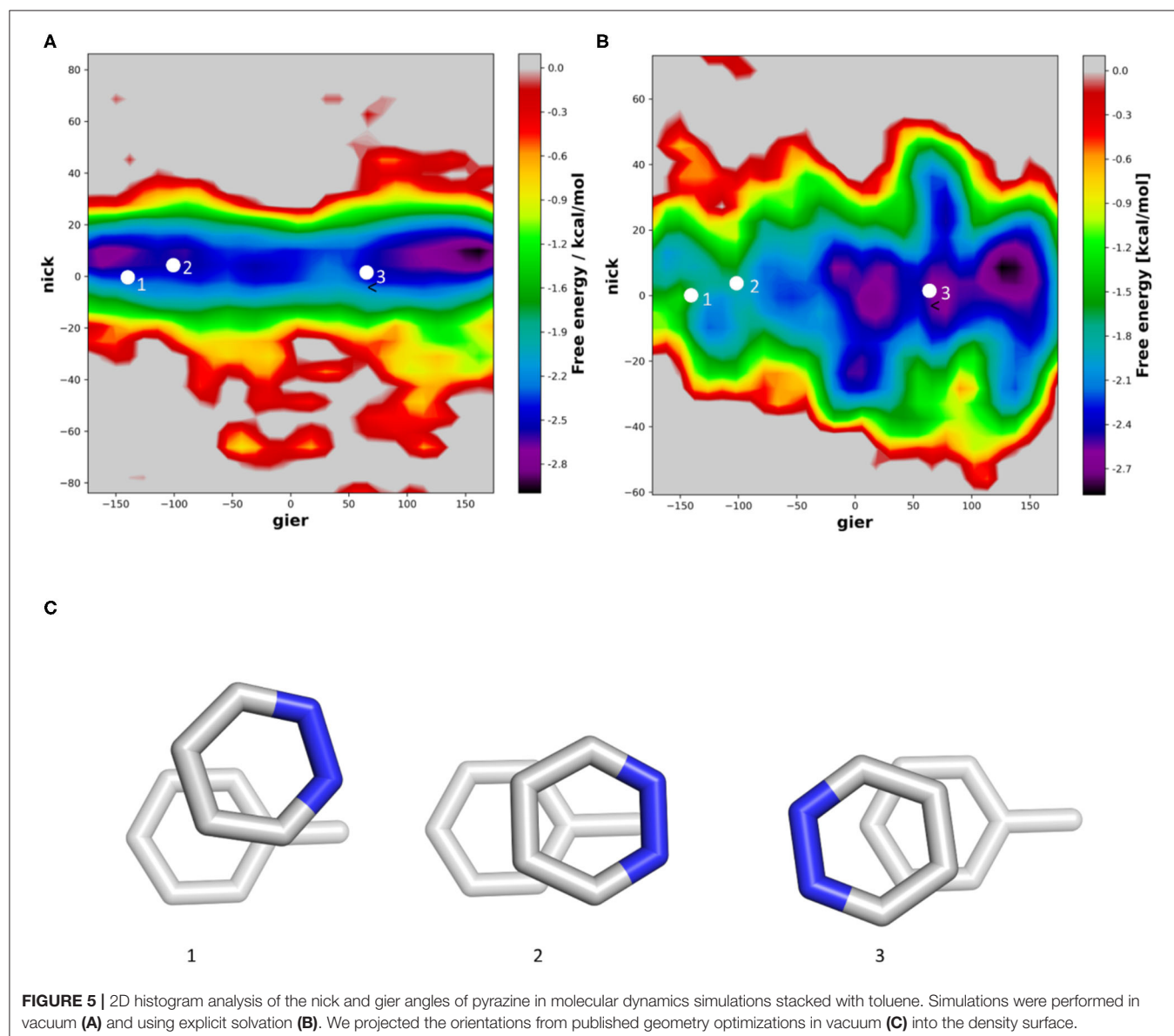

FIGURE 5 | 2D histogram analysis of the nick and gier angles of pyrazine in molecular dynamics simulations stacked with toluene. Simulations were performed in vacuum **(A)** and using explicit solvation **(B)**. We projected the orientations from published geometry optimizations in vacuum **(C)** into the density surface.
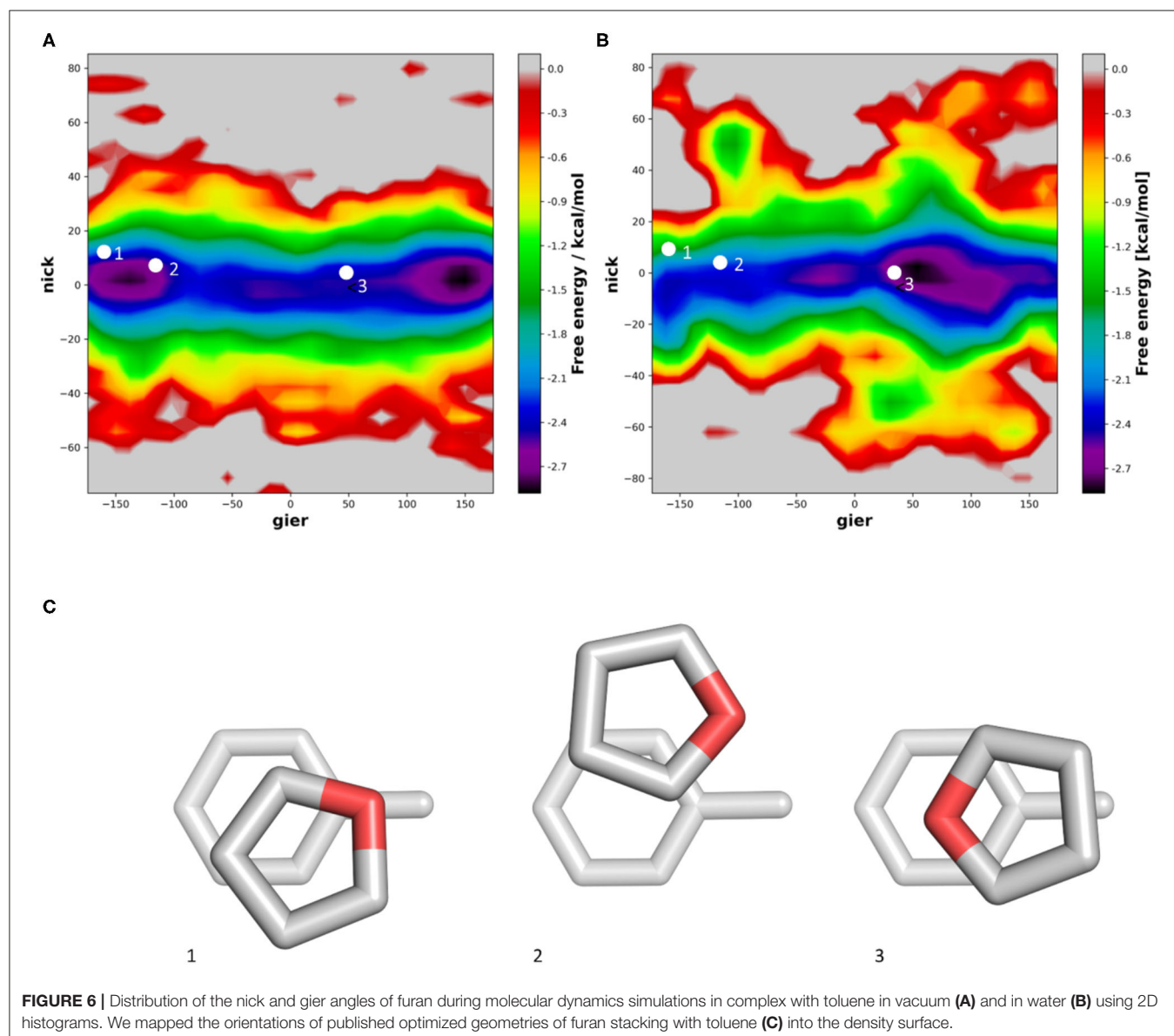
(2019), and solvated these structures as described in the Methods section. The geometries by Bootsma et al. (2019), were obtained by performing elaborate high-level quantum mechanical calculations. It has been shown that the potential energy surface of stacked heteroaromatics is rather shallow, therefore, we focused our analysis on the relative orientation of the respective heteroaromatic rings rather than x,y, and z coordinates. Thus, we analyzed the trajectories using the relative orientations of the stacked heteroaromatics to toluene, i.e., the nick and the gier angle, as described in the Methods section. We highlight four systems in these sections, additional plots can be found in the **Supplementary Material**.
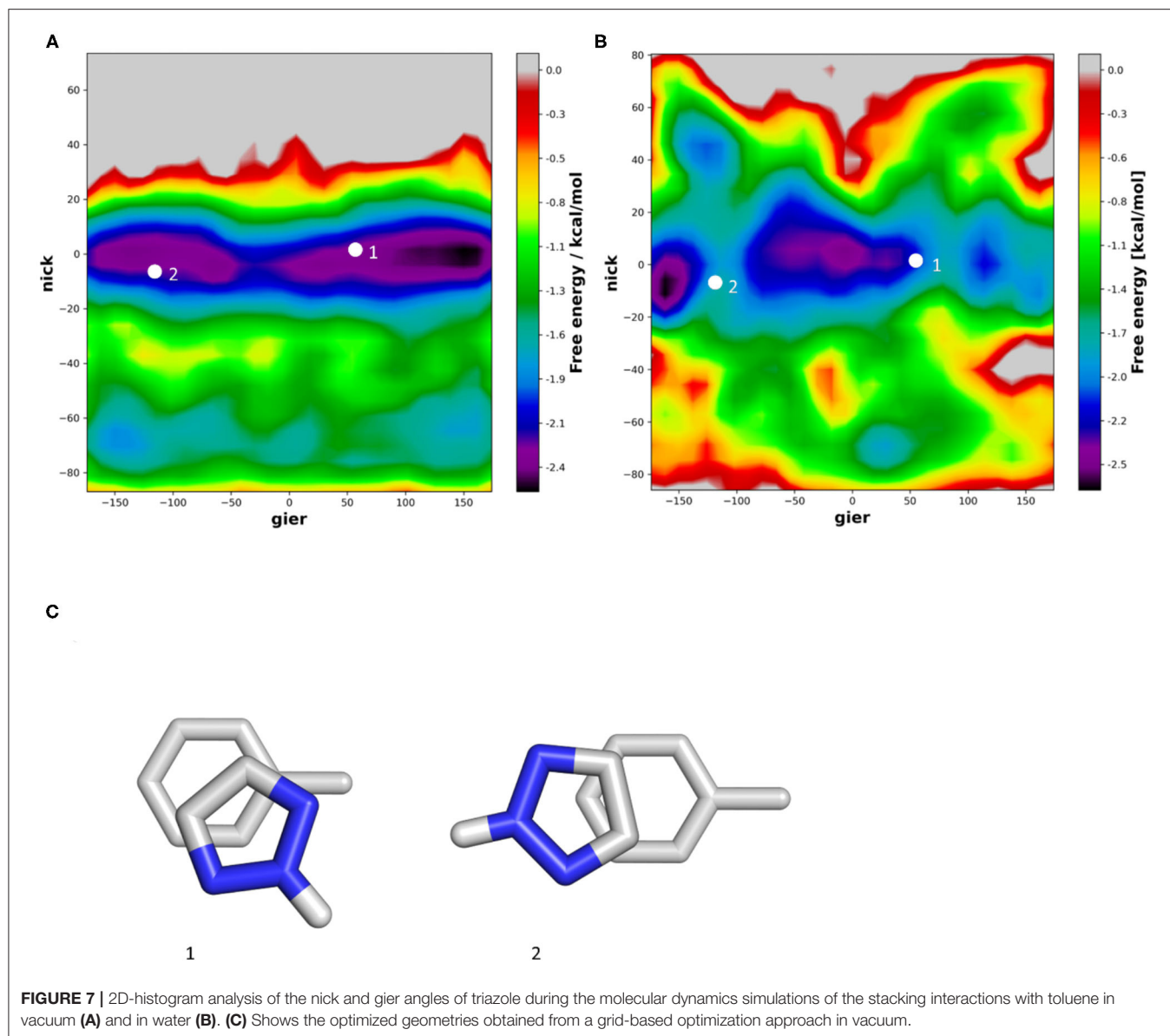
In general, we can see that the nick angle shows less variation than the gier angle regardless if the simulation is performed in vacuum or water (cf. **Supplementary Figure 4**). However, comparing the individual systems, either simulated in vacuum or water, different population distributions can be observed.

For the benzene-toluene complex, we sample both the π-π stacked and the T-stacked conformations (cf. **Supplementary Figure 5**). However, we can see a clear preference for the π-π stacked geometry in vacuum and explicit solvation. The T-stacked geometry can only be found stabilized in simulations using explicit solvent. However, even in the simulations performed in vacuum, we can show that the two molecules are hardly ever completely parallel, but almost always slightly tilted (**Supplementary Figure 1**), a fact that is very difficult to include in grid-based approaches using single point calculations.

In contrast to benzene, pyridazine has a substantial dipole, due to the two neighboring heteroatoms. In vacuum, we can clearly observe that the orientations proposed from QM

FIGURE 6 | Distribution of the nick and gier angles of furan during molecular dynamics simulations in complex with toluene in vacuum **(A)** and in water **(B)** using 2D histograms. We mapped the orientations of published optimized geometries of furan stacking with toluene **(C)** into the density surface.

FIGURE 7 | 2D-histogram analysis of the nick and gier angles of triazole during the molecular dynamics simulations of the stacking interactions with toluene in vacuum **(A)** and in water **(B)**. **(C)** Shows the optimized geometries obtained from a grid-based optimization approach in vacuum.
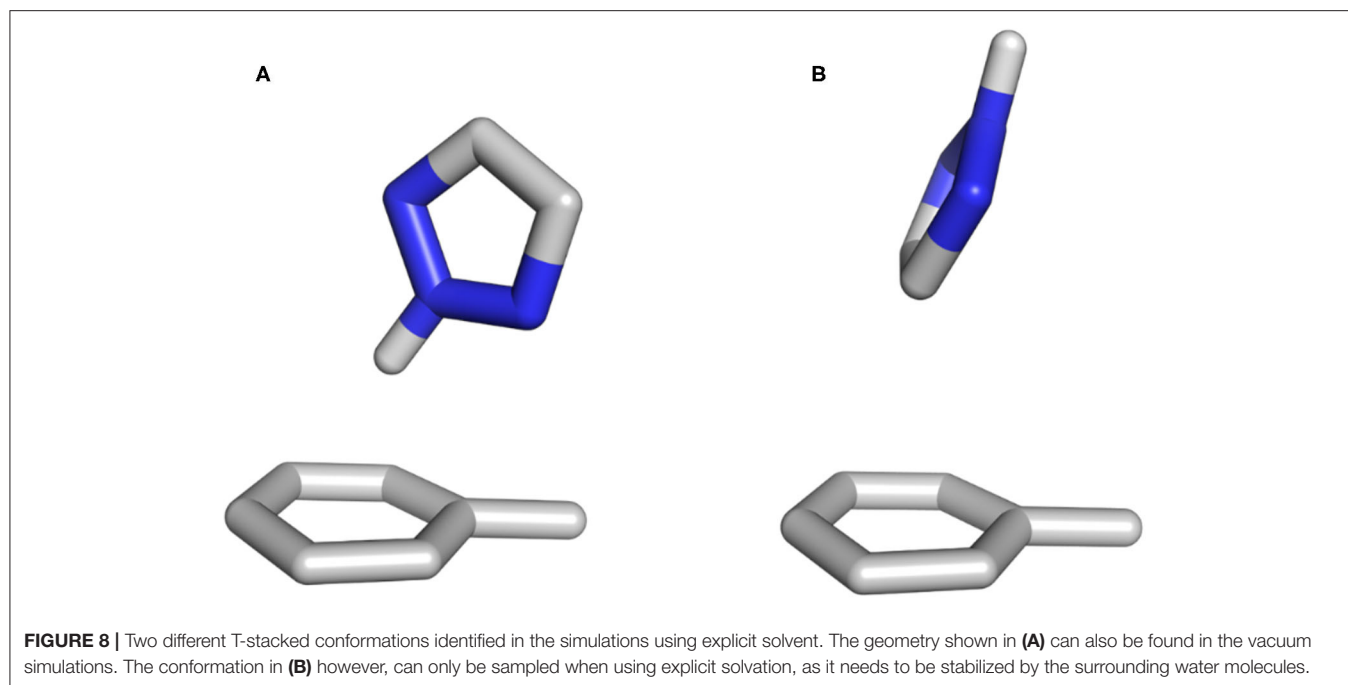
simulations represent the two main minima (**Figure 5A**). In our trajectories, the main orientation is found when the two dipoles are aligned but pointing into opposite directions (**Figure 5C**). In the presence of a solvent, no deep minimum can be identified, but we can clearly see, that an orientation in which the two Nitrogen atoms are orientated directly toward the methyl group of toluene is substantially less likely (**Figure 5B**). This is well in line with previously published results, where a second minimum was identified in implicit solvent geometry optimization (Loeffler et al., 2020). In the violin plots (**Supplementary Figure 4**), we can see that in the gier angle the distribution of the minima is ∼30°, which corresponds to a rotation by one aromatic bond of the aromatic ring.

For five-membered rings, the inserted heteroatoms play a crucial role for the stacking interaction strength and

conformations. In the example of furane we can find one orientation sampled very commonly. As mentioned previously, vacuum quantum mechanical calculations show low energy conformations when the dipole of furan and toluene are aligned. In our simulations we find that this orientation is indeed favorable, when performing the simulations in vacuum (**Figure 6A**). However, when performing the simulations in water, we can clearly observe a shift in the population (**Figure 6B**). In the violin plot (**Supplementary Figure 4**), this population shift is especially visible in the nick angle, clearly showing a more favorable tendency for T-stacked geometries in water compared to the vacuum distributions. Similar to the simulations of pyrazine, we can now identify the most favored orientation where the Oxygen atom is orientated toward the solvent rather than the methyl group of toluene (**Figure 6C**).

FIGURE 8 | Two different T-stacked conformations identified in the simulations using explicit solvent. The geometry shown in **(A)** can also be found in the vacuum simulations. The conformation in **(B)** however, can only be sampled when using explicit solvation, as it needs to be stabilized by the surrounding water molecules.

This conformation is stabilized by the surrounding solvent. Furthermore, we can observe a slightly higher occurrence of T-stacked geometries in water, which are also stabilized by interactions of the heteroatom and the aromatic π-cloud with surrounding water molecules.

Introducing a protonated Nitrogen atom to a five membered heteroaromatic system substantially influences its electrostatic properties and thereby stacking interaction (Bootsma et al., 2019). In our simulations we do not only see π-stacking but also various conformations of T-stacking. In vacuum, the T-stacking is observed exclusively as an interaction of the protonated Nitrogen atom with the toluene π-cloud (**Figure 7A**). During the simulations performed in water we additionally capture a conformation where the protonated Nitrogen atom interacts with the surrounding water molecules while the stacking interaction occurs between one of the carbon-bound hydrogen atoms (**Figure 7B**). Despite the different stacking geometries, we are able to identify a preference of orientation. In vacuum the strong dipole of triazole is aligned with the toluene dipole, while in water it is clearly favorable for the protonated Nitrogen atom to be orientated away from the methyl group of toluene, thereby allowing an improved interaction with the surrounding water molecules. These observations can also be confirmed in the violin plots (**Supplementary Figure 4**), where the distribution of the nick angles is substantially broader, indicating the occurrence of different T-stacked geometries.

DISCUSSION

In this study we performed molecular dynamics simulations of heteroaromatics, stacking with toluene in vacuum and in explicit

solvent. It has been shown previously, that even implicit solvation can influence stacking interaction energies and geometries. In our results we observe this most prominently for heterocycles where a protonated Nitrogen atom is present. In vacuum, T-stacking is almost always favored in unrestrained geometry optimizations, while the parallel displayed geometry is more favorable when using an implicit solvent. Furthermore, we also calculated the vacuum stacking interactions by using ANI. Overall, we find a good correlation of the resulting energies with DFT calculations, despite an offset in the absolute energy values (see **Figure 3**). However, for the 5-membered rings, three complexes reveal a substantially stronger stacking interaction with ANI, namely furan, isoxazole, and oxazole. If these three complexes are neglected, the correlation increases to 0.93. This might indicate that the Oxygen atom in aromatic rings is not yet perfectly trained within the ANI network to characterize such subtle intermolecular interactions.

Previous publications have shown that vacuum stacking interactions are stronger when heteroatoms are positioned outside the toluene π-cloud (Huber et al., 2014; Bootsma et al., 2019). When checking the position of the heteroatoms during our simulations, we can confirm for pyrazine that in both vacuum and water the Nitrogen atoms are outside the underlying toluene for more than 70% of the frames. However, as the system reveals a high flexibility, the nitrogen atoms can also be found oriented toward the π-cloud. The vacuum simulations of furan show that the oxygen atom is favorable outside the π-cloud in ∼70% of the simulation. This even increases to more than 80% for the simulation in water, where the oxygen atom of furan can interact with the surrounding water molecules. In the case of triazole, this observation could not be confirmed in vacuum. On the one hand, the protonated Nitrogen atom of triazole is the main

interaction partner for the T-stacked geometries (**Figure 8A**), and on the other hand, in vacuum, the positive polarization of the protonated Nitrogen atom is the only possible interaction partner for the π-cloud of the underlying toluene. The influence of solvation was not only visible from our molecular dynamics simulations, but also from the geometry optimizations using implicit solvation. In contrast to the optimization performed in vacuum, the unrestrained optimization using implicit solvation resulted in a π-π stacked geometry rather than a T-stacked geometry. However, the protonated Nitrogen atom group is still positioned inside the π-cloud. Our simulations in water show that for more than 65% of all frames the protonated Nitrogen atom group is located outside of the π-cloud, interacting with the surrounding water molecules. Additionally, we can identify two different T-stacked conformations in our simulations in water as shown in **Figures 7B**, **8**. On the one hand, we observe a T-stacked geometry stabilized by the interaction of the protonated Nitrogen atom with the underlying π-cloud (**Figure 8A**). This geometry can be seen in vacuum as well as in explicit solvent simulations (**Figure 7**). On the other hand, we identify a T-stacked geometry where the protonated Nitrogen does not interact with the π-cloud but rather with the surrounding water molecules (**Figure 8B**).

ANI allows to explore the conformational space of organic molecules at lower computational cost and facilitates the characterization and understanding of non-covalent interactions i.e., stacking interactions and hydrogen bonds. Nevertheless, in its current form ANI cannot be used to analyze protein-ligand interactions, as the ANI potentials are not yet parametrized for proteins. Furthermore, the water molecules in ANI still need to be evaluated and compared to classical water models, e.g., OPC, SPC, and the TIP water models. Future work on ANI will aim to develop and include new methods to better describe long-range interactions by including coulomb interactions. The constant addition of more data to machine learning methods will make ANI even more generalizable and improve calculations in different chemical environments, the treatment of ions and the applicability to describe reactions (Smith et al., 2019).

In this study we can show that by using neural networks we can get information not only on geometries but also on intermolecular interactions correlating well with state-of-the-art QM calculations. Furthermore, using neural networks we now can generate ensembles of stacked heteroaromatic complexes including explicit solvation. Both of these points can give crucial information in the early stages of computational drug design.

CONCLUSION

In our study we investigated the influence of solvation on complexes of stacked heteroaromatics using implicit solvent geometry optimizations and molecular dynamics simulations including explicit solvation. We demonstrate that potentials derived from machine learning can be used to perform molecular dynamics simulations as the geometries obtained using high level quantum mechanical simulations are present within the ensemble in solution with shifted populations. Additionally, the calculated stacking interactions using neural networks energies calculated in vacuum correlate well with high level quantum mechanical calculations. However, heterocycles containing an oxygen, i.e., furan, oxazole and isoxazole are overpredicted in terms of stacking interaction energies. The ensembles from the molecular dynamics simulations are well in line with previously published results and show that heteroatoms are in general favorable outside of the π-cloud. This is true for heteroatoms except for secondary amines, which, especially in vacuum, show beneficial interactions with the underlying π-cloud. Furthermore, we highlight the necessity of including solvation properties of aromatic molecules as the optimal geometries can differ substantially depending on whether water molecules are present as possible interaction sites or not. The effect of population shifts naturally increases with the polarity of the aromatic ring and is especially notable if secondary amines are present.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem. 2021.641610/full#supplementary-material

REFERENCES

Adhikary, R., Zimmermann, J., Stanfield, R. L., Wilson, I. A., Yu, W., Oda, M., et al. (2019). Structure and dynamics of stacking interactions in an antibody binding site. *Biochemistry* 58, 2987–2995. doi: 10.1021/acs.biochem.9b00119

Beljonne, D., Cornil, J., Silbey, R., Millié, P., and Brédas, J. L. (2000). Interchain interactions in conjugated materials: the exciton model versus the supermolecular approach. *J. Chem. Phys.* 112, 4749–4758. doi: 10.1063/1.481031

Biela, A., Khayat, M., Tan, H., Kong, J., Heine, A., Hangauer, D., et al. (2012). Impact of ligand and protein desolvation on ligand binding to the S1 pocket of thrombin. *J. Mol. Biol.* 418, 350–366. doi: 10.1016/j.jmb.2012.01.054

Bissantz, C., Kuhn, B., and Stahl, M. (2010). A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* 53, 5061–5084. doi: 10.1021/jm100112j

Bootsma, A. N., Doney, A. C., and Wheeler, S. E. (2019). Predicting the strength of stacking interactions between heterocycles and aromatic amino acid side chains. *J. Am. Chem. Soc.* 141, 11027–11035. doi: 10.1021/jacs.9b00936

Bootsma, A. N., and Wheeler, S. E. (2011). Converting SMILES to stacking interaction energies. *J. Chem. Inf. Model.* 59, 3413–3421. doi: 10.1021/acs.jcim.9b00379

Bootsma, A. N., and Wheeler, S. E. (2018). Stacking interactions of heterocyclic drug fragments with protein amide backbones. *ChemMedChem* 13, 835–841. doi: 10.1002/cmdc.201700721

Burley, S. K., and Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* 229, 23–28. doi: 10.1126/science.3892686

Case,. D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Cruzeiro, V. W. D. III, et al. (2018). *Amber 18*. San Francisco, CA: University of California.

Chai, J.-D., and Head-Gordon, M. (2008). Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* 10, 6615–6620. doi: 10.1039/b810189b

Chang, C.-e. A., Chen, W., and Gilson, M. K. (2007). Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U. S.A.* 104, 1534–1539. doi: 10.1073/pnas.0610494104

Chmiela, S., Tkatchenko, A., Sauceda, H. E., Poltavsky, I., Schütt, K. T., and Müller, K.-R. (2017). Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* 3:e1603015. doi: 10.1126/sciadv.1603015

Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K., and Pande, V. S. (2011). Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* 21, 150–160. doi: 10.1016/j.sbi.2011.01.011

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985). Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 107, 3902–3909. doi: 10.1021/ja00299a024

Dobiaš, J., Ondruš, M., Hlaváč, M., Murár, M., Kóna, J., Addová, G., et al. (2019). Medicinal chemistry: an effect of a desolvation penalty of an amide group in the development of kinase inhibitors. *Chem. Pap.* 73, 71–84. doi: 10.1007/s11696-018-0576-6

Dunning, T. H. (1989). Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* 90, 1007–1023. doi: 10.1063/1.456153

Elstner, M. (2006). The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.* 116, 316–325. doi: 10.1007/s00214-005-0066-0

Fernández-Quintero, M. L., Hoerschinger, V. J., Lamp, L. M., Bujotzek, A., Georges, G., and Liedl, K. R. (2020). VH-VL interdomain dynamics observed by computer simulations and NMR. *Proteins* 88, 830–839. doi: 10.1002/prot.25872

Fernández-Quintero, M. L., Kraml, J., Georges, G., and Liedl, K. R. (2019a). CDR-H3 loop ensemble in solution – conformational selection upon antibody binding. *mAbs* 11, 1077–1088. doi: 10.1080/19420862.2019.1618676

Fernández-Quintero, M. L., Loeffler, J. R., Waibl, F., Kamenik, A. S., Hofer, F., and Liedl, K. R. (2019b). Conformational selection of allergen-antibody complexes - surface plasticity of paratopes and epitopes. *PEDS* 32, 513–523. doi: 10.1093/protein/gzaa014

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09*. Wallingford CT: Gaussian Inc.

Gallivan, J. P., and Dougherty, D. A. (1999). Cation-π interactions in structural biology. *PNAS* 96, 9459–9464. doi: 10.1073/pnas.96.17.9459

Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J. S., and Roitberg, A. E. (2020). TorchANI: A free and open source PyTorch based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. Model.* 60, 3408–3415. doi: 10.1021/acs.jcim.0c00451

Ghanbarpour, A., Mahmoud, A. H., and Lill, M. A. (2020). On-the-fly prediction of protein hydration densities and free energies using deep learning. *arXiv[Preprint].arXiv:2001.02201.*

Ghosh, S., and Hammes-Schiffer, S. (2015). Calculation of electrochemical reorganization energies for redox molecules at self-assembled monolayer modified electrodes. *J. Phys. Chem. Lett.* 6, 1–5. doi: 10.1021/jz5023784

Hansen, N., and Van, W. G. (2014). Practical aspects of free-energy calculations: a review. *J. Chem. Theory Comput.* 10, 2632–2647. doi: 10.1021/ct500161f

Harder, M., Kuhn, B., and Diederich, F. (2013). Efficient stacking on protein amide fragments. *ChemMedChem* 8, 397–404. doi: 10.1002/cmdc.201200512

Head, J. D., and Zerner, M. C. (1985). A Broyden—Fletcher—Goldfarb—Shanno optimization procedure for molecular geometries. *Chem. Phys. Lett.* 122, 264–270. doi: 10.1016/0009-2614(85)80574-1

Huber, R. G., Margreiter, M. A., Fuchs, J. E., von Grafenstein, S., Tautermann, C. S., Liedl, K. R., et al. (2014). Heteroaromatic π-stacking energy landscapes. *J. Chem. Inf. Model.* 54, 1371–1379. doi: 10.1021/ci500183u

Jiang, B., Li, J., and Guo, H. (2016). Potential energy surfaces from high fidelity fitting of ab initio points: the permutation invariant polynomial - neural network approach. *Int. Rev. Phys. Chem.* 35, 479–506. doi: 10.1080/0144235X.2016.1200347

Kitaura, K., Ikeo, E., Asada, T., Nakano, T., and Uebayasi, M. (1999). Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* 313, 701–706. doi: 10.1016/S0009-2614(99)00874-X

Kolár, M., Fanfrlík, J., and Hobza, P. (2011). Ligand conformational and solvation/desolvation free energy in protein–ligand complex formation. *J. Phys. Chem. B* 115, 4718–4724. doi: 10.1021/jp2010265

Kuhn, B., Fuchs, J. E., Reutlinger, M., Stahl, M., and Taylor, N. R. (2011). Rationalizing tight ligand binding through cooperative interaction networks. *J. Chem. Inf. Model.* 51, 3180–3198. doi: 10.1021/ci200319e

Larsen, A. H., Mortensen, J. J., Blomqvist, J., Castelli, I. E., Christensen, R., Dulak, M., et al. (2017). The atomic simulation environment—a python library for working with atoms. *J. Phys.* 29:273002. doi: 10.1088/1361-648X/aa680e

Lee, H., Dehez, F., Chipot, C., Lim, H. K., and Kim, H. (2019). Enthalpy-entropy interplay in π-stacking interaction of benzene dimer in water. *J. Chem. Theory Comput.* 15, 1538–1545. doi: 10.1021/acs.jctc.8b00880

Liedl, K. R. (1998). Dangers of counterpoise corrected hypersurfaces. Advantages of basis set superposition improvement. *J. Chem. Phys.* 108, 3199–3204. doi: 10.1063/1.475715

Limongelli, V., Bonomi, M., and Parrinello, M. (2013). Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U. S.A.* 110, 6358–6363. doi: 10.1073/pnas.1303186110

Loeffler, J. R., Fernández-Quintero, M. L., Schauperl, M., and Liedl, K. R. (2020). STACKED – Solvation theory of a romatic complexes as key for estimating drug binding. *J. Chem. Inf. Model.* 60, 2304–2313. doi: 10.1021/acs.jcim.9b01165

Loeffler, J. R., Schauperl, M., and Liedl, K. R. (2019). Hydration of aromatic heterocycles as adversary of π-stacking. *J. Chem. Inf. Model.* 59, 4209–4219. doi: 10.1021/acs.jcim.9b00395

Markley, F. L., and Crassidis, J. L. (2014). "Euler angles," in Fundamentals of Spacecraft Attitude Determination and Control, eds F. L. Markley and J. L. Crassidis (New York, NY: Space Technology Library; Springer), 361–364. doi: 10.1007/978-1-4939-0802-8_9

Meyer, E. A., Castellano, R. K., and Diederich, F. (2003). Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed.* 42, 1210–1250. doi: 10.1002/anie.200390319

Michel, J., and Essex, J. W. (2010). Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aided Mol. Des.* 24, 639–658. doi: 10.1007/s10822-010-9363-3

Mobley, D. L., and Klimovich, P. V. (2012). Perspective: alchemical free energy calculations for drug discovery. *J. Chem. Phys.* 137:230901. doi: 10.1063/1.4769292

Nguyen, D. D., Cang, Z., Wu, K., Wang, M., Cao, Y., and Wei, G.-W. (2019). Mathematical deep learning for pose and binding affinity prediction and ranking in d3r grand challenges. *J. Comput. Aided Mol. Des.* 33, 71–82. doi: 10.1007/s10822-018-0146-6

Prampolini, G., Livotto, P. R., and Cacelli, I. (2015). Accuracy of quantum mechanically derived force-fields parameterized from dispersion-corrected DFT data: the benzene dimer as a prototype for aromatic interactions. *J. Chem. Theory Comput.* 11, 5182–5196. doi: 10.1021/acs.jctc.5b00642

Salonen, L. M., Ellermann, M., and Diederich, F. (2011). Aromatic rings in chemical and biological recognition: energetics and structures. *Angew. Chem. Int. Ed.* 50, 4808–4842. doi: 10.1002/anie.201007560

Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R., and Tkatchenko, A. (2017). Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* 8:13890. doi: 10.1038/ncomms13890

Sherrill, C. D., Sumpter, B. G., Sinnokrot, M. O., Marshall, M. S., Hohenstein, E. G., Walker, R. C., et al. (2009). Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π-π, CH/π, and SH/π interactions. *J. Comput. Chem.* 30, 2187–2193. doi: 10.1002/jcc.21226

Smith, J. S., Isayev, O., and Roitberg, A. E. (2017). ANI-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.* 8, 3192–3203. doi: 10.1039/C6SC05720A

Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O., and Roitberg, A. E. (2018). Less is more: sampling chemical space with active learning. *J. Chem. Phys.* 148:241733. doi: 10.1063/1.5023802

Smith, J. S., Nebgen, B. T., Zubatyuk, R., Lubbers, N., Devereux, C., Barros, K., et al. (2019). Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* 10:2903. doi: 10.1038/s41467-019-10827-4

Stewart, J. J. P. (2009). Application of the PM6 method to modeling proteins. *J. Mol. Model.* 15, 765–805. doi: 10.1007/s00894-008-0420-y

Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., et al. (2020). Ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* 16, 528–552. doi: 10.1021/acs.jctc.9b00591

Tomasi, J., Mennucci, B., and Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chem. Rev.* 105, 2999–3094. doi: 10.1021/cr9904009

Wallnoefer, G. H., Fox, T., Liedl, K. R., and Tautermann, C. S. (2010). Dispersion dominated halogen–π interactions: energies and locations of minima. *Phys. Chem. Chem. Phys.* 12, 14941–14949. doi: 10.1039/c0cp00607f

Wang, H., Zhang, L., and Han, J. E. W. (2018). DeePMD-Kit: a deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* 228, 178–184. doi: 10.1016/j.cpc.2018.03.016

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035

Wang, L., Deng, Y., Wu, Y., Kim, B., LeBard, D. N., Wandschneider, D., et al. (2017). Accurate modeling of scaffold hopping transformations in drug discovery. *J. Chem. Theory Comput.* 13, 42–54. doi: 10.1021/acs.jctc.6b00991

Wang, S., and Riniker, S. (2019). Use of molecular dynamics fingerprints (MDFPs) in SAMPL6 octanol–water log P blind challenge. *J. Comput. Aided Mol. Des.* 34, 393–403. doi: 10.1007/s10822-019-00252-6

Wheeler, S. E., and Bloom, J. W. G. (2014). Anion–π interactions and positive electrostatic potentials of N-heterocycles arise from the positions of the nuclei, not changes in the π-electron distribution. *Chem. Commun.* 50, 11118–11121. doi: 10.1039/C4CC05304D

Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinković, D., and Jhoti, H. (2003). Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* 424, 464–468. doi: 10.1038/nature01862

Xu, M., Zhu, T., and Zhang, J. Z. H. (2019). Molecular dynamics simulation of zinc ion in water with an ab initio based neural network potential. *J. Phys. Chem. A* 123, 6587–6595. doi: 10.1021/acs.jpca.9b04087

Yao, K., Herr, J. E., Toth, D. W., Mckintyre, R., and Parkhill, J. (2018). The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.* 9, 2261–2269. doi: 10.1039/C7SC04934J