PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2021, **23**, 26263

Protein confinement fine-tunes aggregation-induced emission in human serum albumin†

Ruibin Liang, * Debojyoti Das and Amirhossein Bakhtiiari

Luminogens exhibiting aggregation-induced-emission characteristics (AIEgens) have been designed as sensitive biosensors thanks to their "turn-on" fluorescence upon target binding. However, their AIE mechanism in biomolecules remains elusive except for the qualitative picture of restricted intramolecular motions. In this work, we employed *ab initio* simulations to investigate the AIE mechanism of two tetraphenylethylene derivatives recently developed for sensitive detection of human serum albumin (HSA) in biological fluids. For the first time, we quantified the *ab initio* free energy surfaces and kinetics of AIEgens to access the conical intersections on the excited state in the protein and aqueous solution, using a novel first-principles electronic structure method that incorporates both static and dynamic electron correlations. Our simulations accurately reproduce the experimental spectra and high-level correlated electronic structure calculations. We found that in HSA the internal conversion through the cyclization reaction is preferred over the isomerization around the central ethylenic double bond, whereas in the aqueous solution the reverse is true. Accordingly, the protein environment is able to moderately speed up certain non-radiative decay pathways, a new finding that is beyond the prediction of the existing model of restricted access to a conical intersection (RACI). As such, our findings highlight the complicated effects of the protein confinement on the competing non-radiative decay channels, which has been largely ignored so far, and extend the existing theories of AIE to biological systems. The new insights and the multiscale computational methods used in this work will aid the design of sensitive AIEgens for bioimaging and disease diagnosis.

Received 5th October 2021,
Accepted 2nd November 2021

DOI: 10.1039/d1cp04577f

rsc.li/pccp

Introduction

Molecules with aggregation-induced emission characteristics (or AIE luminogens, abbreviated as AIEgens below) are non-emissive in dilute solution but emissive upon the formation of molecular aggregates.[1] This "turn-on" fluorescence characteristic greatly enhances their detection sensitivity compared to the conventional "always-on" fluorescent probes.[1] Recently, various AIEgens were developed for biomolecular sensing.[2] For example, tetraphenylethylene (TPE)-derived AIEgens were designed for quantitative analysis of human serum albumin (HSA)[3–5] and protein aggregates[6–8] in complex biological fluids, which is critical for early diagnosis of many important diseases.[9–15]

The mechanism of AIE has been studied extensively since its first discovery.[1,16,17] Until now, the general consensus is that the restriction of intramolecular motions (RIM) in the aggregate slows down the non-radiative decay and thus enhances the radiative decay compared to the dilute solution. Early theoretical studies identified the important contribution of low-frequency vibrational modes to non-radiative decay.[18–21] A recent model attributed the AIE to the restricted access to the conical intersection (RACI) in the aggregate.[22–29] As a common type of AIEgen, TPE and its derivatives have been widely applied in fluorescence imaging, and their AIE mechanism has been extensively studied.[30–37] The isomerization around the central ethylenic double bond and the cyclization reaction (Fig. 1A) are two major non-radiative decay pathways in the gas phase and solution,[30–33,37–39] but their relative importance remains debated.[17]

Despite the qualitative pictures of RIM and RACI, the AIE mechanism in biomolecular settings has been much less studied than in the aggregate phase. For example, it remains elusive if the cyclization of the TPE derivatives can readily take place in the protein. Also, it is unclear how much the protein environment can slow down the isomerization compared with the solution. Quantitatively answering these mechanistic questions is critical for optimizing the AIEgens as effective biosensors, and *ab initio* simulation has a unique advantage for achieving this goal because it can fully track the photodynamics at atomistic-level detail.

Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX, 79409, USA. E-mail: rliang@ttu.edu

† Electronic supplementary information (ESI) available: Potential energy surface calculation of gas-phase TPE-2TA using the hh-TDA-BHLYP method, comparing the 6-31G* and the def2-TZVP basis sets. See DOI: 10.1039/d1cp04577f

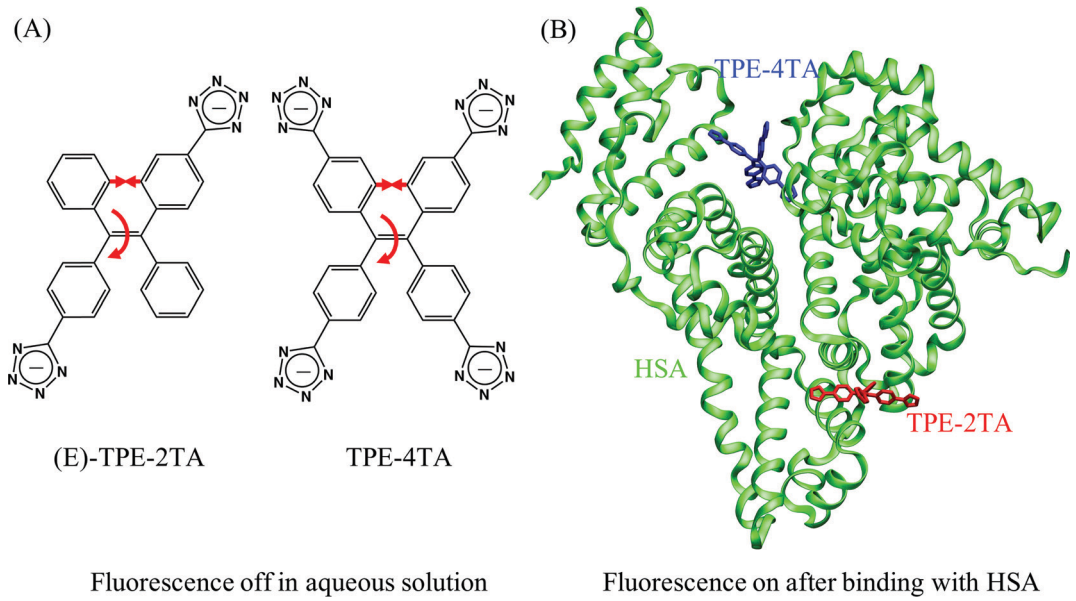

Fig. 1 In aqueous solutions, the rapid internal conversions of the AIEgens (TPE-2TA & TPE-4TA) turn off their fluorescence. After binding with HSA, the modified kinetics of the internal conversions turns on their fluorescence. (A) Chemical structures of TPE-2TA and TPE-4TA. The red arrows highlight the twisting of the central ethylenic C=C bond during the isomerization and the closure of the C–C distance during the cyclization reaction. (B) The binding poses of TPE-2TA (red) and TPE-4TA (blue) in the HSA (green) with the highest scores predicted by docking simulations. In each system setup, only one AIEgen (TPE-2TA or TPE-4TA) is bound to the protein.

According to the RACI model, to understand the AIE mechanism in the condensed phase, a fundamental property to quantify is the free energy surface (FES) of the AIEgens to reach the CIs between the ground and excited state, because it incorporates the thermal fluctuations of the environment at finite temperatures. However, this is a challenging task for *ab initio* simulations. Firstly, the electronic structure method needs to incorporate both the static and dynamic electron correlations. This is essential for accurately predicting the AIEgen's absorption/fluorescence spectra and the potential energy surface (PES), especially near the S_0/S_1 CIs.[40] Secondly, sufficient sampling of the conformational space is required, which entails efficient electronic structure calculations. To the best of our knowledge, only one study[27] has characterized the *ab initio* excited-state FES of diphenyldibenzofulvene (an AIEgen) in the aggregate phase, and no similar work has been carried out in any biomolecular system so far.

To fill this void, we employed a recently developed first-principles method, *i.e.*, the hole–hole Tamm–Dancoff approximated density functional theory (hh-TDA-DFT),[41,42] in a quantum mechanics/molecular mechanics (QM/MM) setting, and investigated the AIE mechanism of two tetrazolate-tagged TPE derivatives (TPE-TAs) designed for HSA analysis[4] (Fig. 1). This system was chosen as a model system to understand AIE in biological systems. The hh-TDA-DFT method efficiently incorporates both static and dynamic electron correlations,[41,42] and has been shown to accurately reproduce the experimental spectra and high-level *ab initio* benchmark calculations for azobenzene photodynamics.[43,44] The excited-state FES and non-equilibrium dynamics both indicate that the protein shifts the relative importance of the two non-radiative decay pathways, which points to the complicated nature of the AIE in biomolecular environments. Our findings highlight the necessity to explicitly consider multiple competing internal conversion pathways to optimize the performance of the AIEgens in complicated molecular environments.

Methods

Gas-phase benchmark calculations

The hh-TDA-DFT method was used to calculate the energies of the critical points of TPE-2TA and TPE-4TA and the S_1 state PES along their isomerization and cyclization pathways in the vacuum. The snapshots along the pathways were obtained by constrained optimization on the S_1 state. The constraints were on the distance between the two terminal carbon atoms during the cyclization or the central torsion angle during the isomerization (Fig. 1A). The critical points included the S_0 and S_1 state minima and the two S_0/S_1 state minimum energy conical intersections (MECIs) encountered along the two decay pathways. The critical points and snapshots along the pathways were optimized using the hh-TDA-DFT/6-31G* method with the BHLYP exchange–correlation functional[45–47] (abbreviated as hh-TDA-BHLYP below).

The energies for the same snapshots were benchmarked using the extended multistate complete active space second-order perturbation theory (XMS-CASPT2).[48] The reference wave function of the XMS-CASPT2 calculations was obtained from the state-average complete active space self-consistent field calculations, with an active space of 6 electrons and 6 orbitals and state-averaging over the five lowest singlet states (SA5-CASSCF(6,6)). The reference space of the XMS-CASPT2

calculation was spanned by the 5 CASSCF states. The XMS-CASPT2 calculations used a zero IPEA shift, a 0.25 a.u. imaginary shift and a 6-31G* basis set. All gas-phase calculations were performed using the TeraChem software package.[49–52]

System setup for TPE-2TA and TPE-4TA in HSA and aqueous solutions

Starting from a crystal structure of HSA (PDB code: 2vue), missing residues were first added to the crystal structure by homology modeling using the MODELLER software.[53] The Autodock Vina software[54] was then used to identify the binding poses of the TPE-2TA and TPE-4TA chromophores in the HSA. The binding pose with the most favorable binding energy was chosen for each AIEgen (Fig. 1B). Each of the two HAS + AIEgen complexes was then solvated by water molecules and sodium ions to neutralize the negative charges, resulting in a periodic boundary condition (PBC) simulation box with a size of $\sim 120 \times 147 \times 133$ Å^3. The setup of the simulation box was performed using AmberTools20.[55] To model the AIEgens in the aqueous solution, each of the two AIEgens was solvated by water molecules and NaCl ions, resulting in a physiological NaCl concentration of 0.16 M. The PBC simulation boxes have sizes of $\sim 65 \times 60 \times 60$ Å^3. The solvation of the AIEgens was performed using the Packmol software package.[56]

Classical MD simulations

Classical molecular dynamics (MD) simulation was used to equilibrate the systems. The force field parameters of TPE-2TA and TPE-4TA were obtained following the general Amber force field (GAFF) procedure.[57,58] The point charges of the AIEgens were derived from the restrained electrostatic potential[59] using the Hartree Fock/6-31G* wave functions. The protein and water molecules were treated with Amber14 and SPC/Fw force fields, respectively.[57,58,60–62]

The two HAS + AIEgen systems were first optimized with 100 kcal mol^{-1} Å^{-2} positional restraints on all heavy atoms of the protein and the AIEgens. With the same restraints, the systems were then equilibrated in the constant *NVT* ensemble for 10 ps. Following this, the systems were further equilibrated in the constant *NPT* ensemble, with the restraints only put on the protein backbone atoms and gradually reduced to 0.5 kcal mol^{-1} Å^{-2} over 4 ns. Then all restraints on the protein were fully released, and the *NPT* equilibration was performed for 50 ns. Due to the inaccuracy of the AIEgens' force field, the dihedral angles around the rotatable C–C and C=C bonds in the AIEgens were restrained to the values of the gas-phase S_0 minimum geometry using a harmonic potential with 250 kcal mol^{-1} radian^{-2} force constant. A similar equilibration procedure was followed for the aqueous solution systems, except that the steps regarding the protein restraints were not applicable and were skipped. The MD simulations were performed using the OpenMM software package.[63]

QM/MM ground-state MD simulations

For each classical MD trajectory, twenty snapshots were taken from the last 20 ns dynamics with a time interval of 1 ns. For each snapshot, a subsystem with the open boundary condition was generated from the PBC system due to the current limitations in the TeraChem software to treat the PBC. Specifically, all protein atoms, the AIEgen and every solvent molecule with at least one atom within 15 Å of any protein atoms were retained. The remaining molecules were discarded. For each new system prepared in this way, the QM region included only the AIEgen (54 and 62 atoms for the TPE-2TA and TPE-4TA, respectively) and was treated with the BHLYP exchange–correlation functional. The MM region included the rest of the system and was treated with the same force field as in the classical MD simulations. The QM and MM subsystems were coupled through electrostatic embedding. In total, 20 snapshots were prepared for each system. Starting from each snapshot, a single ground-state QM/MM MD simulation was performed in the constant *NVT* ensemble with $T = 300$ K for 5 ps. For each trajectory, the first 2 ps simulation was discarded as equilibration and the last 3 ps was treated as the production run. Thus, in total, there were $\sim$60 ps of QM/MM sampling on the ground state PES, which originated from 20 ns of MM sampling of the conformational space. All torsions of the AIEgen were free to move during the QM/MM equilibration in the HSA and solution. The ground-state QM/MM MD simulations were performed using the TeraChem[49–52] interfaced with the OpenMM packages.[63]

Absorption spectra calculation

The absorption spectra of the TPE-2TA and TPE-4TA in both the HSA and the solution were calculated as below. Six thousand snapshots were taken from the production runs of the 20 ground state QM/MM trajectories, with an interval of 10 fs. The hh-TDA-BHLYP/MM method was then used to calculate the $S_0 \rightarrow S_1$ excitation energy and transition dipole moment for each snapshot. The absorption spectra were then calculated following the procedure of ref. 64, and the result was convolved in energy using a Lorentzian function with a full width at half maximum (FWHM) of 0.15 eV. It is worth noting that the $S_0 \rightarrow S_1$ excitation of the TPE derivatives at the Franck–Condon point is dominated by the $\pi \rightarrow \pi^*$ HOMO $\rightarrow$ LUMO transition.[30,31] Therefore, the $S_0 \rightarrow S_1$ excitation energy in principle can be well described by the $(N, N/2 + 1)$ active space of the hh-TDA-DFT method, where N is the total number of electrons in the AIEgen.[41]

Umbrella sampling simulations

Umbrella sampling simulations were performed on the S_1 state to calculate the free energy surfaces (FES) of the TPE-2TA and TPE-4TA in the HSA and solution. Two one-dimensional potentials of mean forces (PMFs) were calculated for each of the four simulation systems (four combinations between solution/HSA with TPE-2TA/TPE-4TA), resulting in eight PMFs in total. The PMFs were projected to one of the two reaction coordinates, the first being the C–C distance during the cyclization and the second being the torsion angle around the central C=C double bond (Fig. 1A). For each cyclization PMF, seventeen umbrella windows were placed along the reaction path, with window

centers ranging from 3.4 Å to 1.8 Å with a 0.1 Å interval, and a harmonic potential with 40 kcal mol^{-1} $Å^{-2}$ force constant was imposed on the reaction coordinate for each window. Ten umbrella windows were placed along the reaction path for each isomerization PMF, with window centers ranging from 180° to 90° with a 10° interval, and a harmonic potential with 100 kcal mol^{-1} $radian^{-2}$ force constant was imposed on the reaction coordinate for each window. The first umbrella window started from the last snapshot of a ground-state QM/MM MD trajectory. The initial structure of each window was obtained from the structure of its adjacent window after equilibration for ~1 ps. The hh-TDA-BHLYP/MM method was used to calculate the S_1 state forces on the atoms on-the-fly during the umbrella sampling simulations. For each simulation, the sampling was performed for at least 5 ps. The first 1 ps trajectory was discarded as equilibration, and the remaining 4 ps trajectory was treated as the production run. The PMFs were calculated by unbiasing the distributions of the reaction coordinates using the WHAM algorithm.[65,66] The umbrella sampling simulations were performed using the TeraChem[49–52] interfaced with the OpenMM packages.[63]

Non-equilibrium excited-state dynamics and fluorescence spectra calculations

In both the solution and the HSA, 20 non-equilibrium trajectories were initiated on the S_1 state, starting with the coordinates and momenta of the last snapshots of the 20 ground-state QM/MM trajectories. Each trajectory was propagated until 10 ps on the S_1 state or until the S_1–S_0 energy gap first dropped below 0.1 eV, where the system was assumed to reach the CI seam. The same QM/MM method for umbrella sampling was used for propagating the excited-state trajectories. The fluorescence spectra were calculated from these trajectories following the procedure of a previous study.[43] The excited-state trajectories were propagated using the TeraChem[49–52] interfaced with the OpenMM packages.[63]

Results and discussion

Gas phase PES characterization and benchmark calculation

Before characterizing the excited-state free energy profiles and dynamics of the AIEgens in the condensed phase environments, it is necessary to benchmark the accuracy of the hh-TDA-DFT method against high-level multireference *ab initio* calculations and experimental spectra. The BHLYP exchange–correlation functional[45–47] was used in all hh-TDA-DFT calculations (denoted as hh-TDA-BHLYP below). Fig. 2A illustrates the S_0 and S_1 energies of the TPE-2TA in the vacuum at several critical points of the PES. The $S_0 \rightarrow S_1$ excitation energies at the S_0 and S_1 minima predicted by the hh-TDA-BHLYP method are within 0.3 eV of the results of the XMS-CASPT2 calculations. The hh-TDA-BHLYP energies at the two S_0/S_1 MECIs encountered during the cyclization and isomerization are also within 0.3 eV of the XMS-CASPT2 results. The S_1 state PES along the isomerization and cyclization pathways of the TPE-2TA in the vacuum are also benchmarked (Fig. 2B and C). The hh-TDA-BHLYP (XMS-CASPT2) method predicts a 0.64 eV (0.43 eV) energy barrier along the cyclization pathway and a 0.10 eV (0.18 eV) energy barrier along the isomerization pathway. Therefore, both methods predict that the isomerization around the central ethylenic C=C double bond is energetically more favorable than the cyclization reaction in the vacuum.

Absorption and fluorescence spectra

To further benchmark the hh-TDA-DFT method's accuracy, the absorption spectra of the TPE-2TA and TPE-4TA were calculated in both the aqueous solution and the HSA (Fig. 3A and B). The calculated wavelength of maximum absorption is blue-shifted by ~0.3 eV compared to the experiment.[4] However, for each type of AIEgen, the wavelengths of maximum absorption are similar in both environments. In the experiment, the addition of the HSA to the solution incurs minimal changes in the absorption spectra.[4] In this regard, the relative shift of the absorption wavelength upon changing the environment is well reproduced by the simulation. Another key property of an AIEgen is the fluorescence spectrum. The time-integrated fluorescence spectra were calculated from the S_1 state trajectories of TPE-2TA and TPE-4TA in the HSA (see the "Methods" section). As shown in Fig. 3C and D, the calculated wavelengths of maximum emission for both AIEgens are within 0.21 eV blue-shift from the experimental values.

Despite the 0.2–0.35 eV blue-shift from the experimental spectra, it is worth noting that the hh-TDA-BHLYP method outperforms electronic structure methods without dynamic electron correlation. For example, the CASSCF(10,6)/6-31G* method predicts a $S_0 \rightarrow S_1$ excitation energy of 4.64 eV for the TPE-2TA at the S_0 minimum in the vacuum, further blue-shifting the hh-TDA-BHLYP's excitation energies by ~1 eV. The inclusion of dynamic correlation mitigates the blue-shifting in the calculated spectra. For example, the time-dependent DFT (TD-DFT) calculation using the BHLYP functional predicts an $S_0 \rightarrow S_1$ excitation energy of 3.73 eV at the same geometry, similar to the hh-TDA-BHLYP result (3.70 eV). However, because the hh-TDA-DFT also incorporates static correlation, it correctly describes the topology of the S_0/S_1 CIs,[40,41] which is a crucial advantage over the TD-DFT method when accurate trajectory propagation near the CI is necessary (see below).

Furthermore, after changing the AIEgen's environment from the vacuum to the solution or HSA, the excitation and emission energies of TPE-2TA are blue-shifted by ~0.3 eV and ~0.6 eV, respectively (Fig. 2A *vs.* Fig. 3). In other words, the condensed-phase environment introduces non-negligible shifts in the AIEgen's absorption and fluorescence spectra. Thus, there is a caveat with directly comparing the calculated gas-phase energies with the experimental spectra measured in the condensed phase. In other words, a meaningful comparison between simulation and experiment necessitates including the electrostatic and steric interactions between the environment and AIEgens. Therefore, the appropriate benchmark for the hh-TDA-DFT method is the XMS-CASPT2 results in the vacuum and the experimental spectra in the condensed phase.

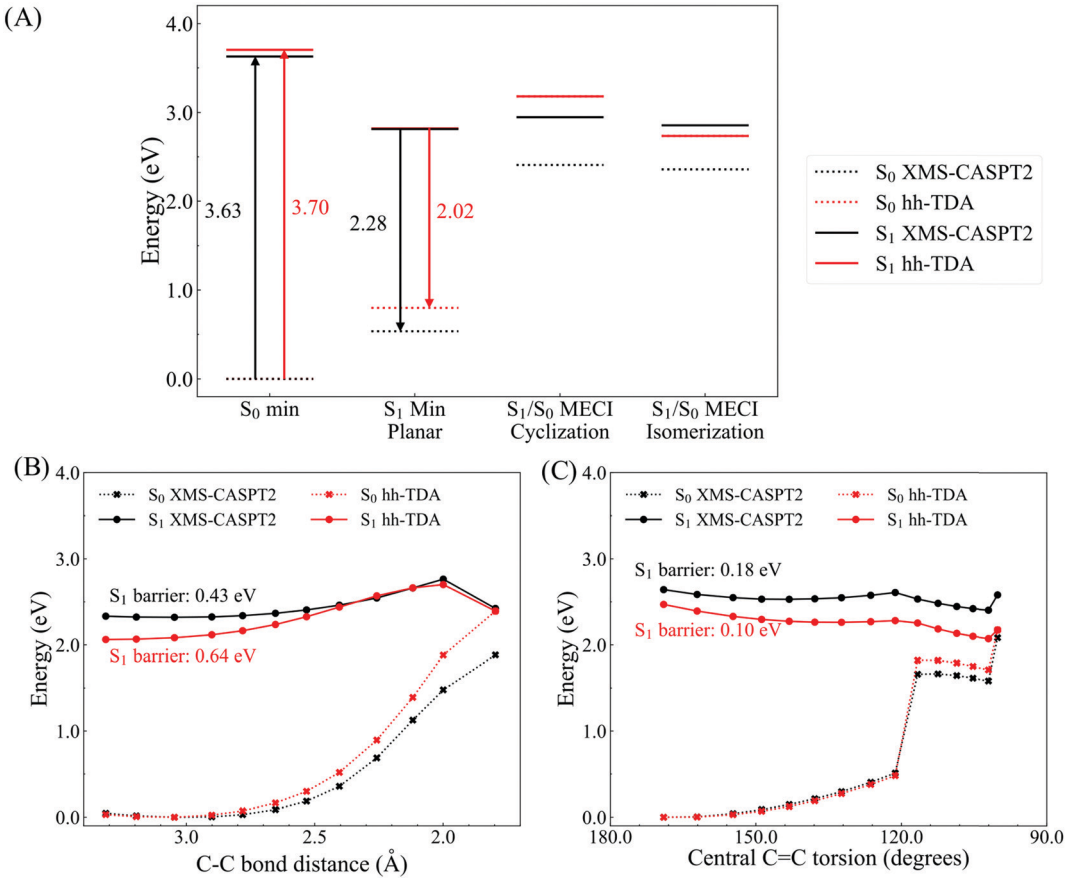

Fig. 2 The PES of the TPE-2TA in the vacuum. The isomerization is energetically preferred over the cyclization, and the benchmark XMS-CASPT2 calculations (black) confirm the hh-TDA-BHLYP results (red). (A) The S_0 and S_1 state energies corresponding to the critical points, including the S_0 and S_1 state minima (S_0 min and S_1 min) and the two S_0/S_1 MECIs encountered in the isomerization and cyclization pathways. The excitation and emission energies at S_0 and S_1 min are labeled with arrows. (B) PES along the C–C bond distance in the cyclization pathway. (C) PES along the torsion around the central ethylenic C=C bond for the isomerization pathway. The largest S_1 state energy barrier along each pathway is indicated. The solid and dotted lines indicate the S_1 and S_0 states, respectively. The zero-reference energy is taken as the lowest S_0 state energy.

Overall, the calculated spectra reasonably reproduce the experiments, which further validates the hh-TDA-DFT method.

Excited-state free energy profiles

To characterize the AIE mechanism of TPE-2TA and TPE-4TA, umbrella sampling simulations were performed to calculate the S_1 state potential of mean forces (PMFs) for the cyclization and isomerization pathways of the AIEgens, in both the aqueous solution and HSA (Fig. 4). The protein environment of the HSA significantly reshaped the free energy surface of the AIEgens compared with the aqueous solution. In the solution, the rotation around the central C=C bond towards the CI at $\sim 90^\circ$ is almost barrierless. In contrast, in the HSA the rotation needs to overcome a barrier of more than 20 kcal mol^{-1} to reach the CI (Fig. 4A and C). The impact of the HSA on the free energy barrier is similar for both binding sites in the HSA (Fig. 1B). The origin of the large increase in the barrier is probably the steric repulsion between the AIEgen and the protein. The isomerization of the central ethylenic bond requires large displacements of the bulky phenyl rings, which are restricted in the binding pocket. (Fig. 1A) Furthermore, several hydrogen bonds are formed between the negatively-charged tetrazolate rings on the AIEgens and the positively-charged Lys and Arg residues in the HSA, which also hinder the rotation of the central ethylenic double bond.[4] The large free energy barrier to reach the CI significantly slows down the non-radiative decay through the isomerization channel. This observation corroborates the RACI model originally proposed for the molecular aggregate.[22] To the best of our knowledge, it is the first time that the RACI model is confirmed and quantified for AIE in biomolecular systems.

Interestingly, in contrast to the isomerization, the cyclization reaction is accelerated by the protein environment (Fig. 4B and D). In the solution, both AIEgens need to overcome a 13–14 kcal mol^{-1} barrier to reach the CI at $\sim$1.8 Å C–C distance. In the HSA, the barriers are reduced to 8–9 kcal mol^{-1}, regardless of the difference in the binding sites (Fig. 1B). The reduction of the cyclization barrier in the protein can be partly attributed to the shortened equilibrium distance between the two terminal carbon atoms during the cyclization ($\sim$3.2 Å *vs.* $\sim$2.9 Å at the free energy minima in Fig. 4B and D). In addition, the cyclization reaction does not involve large

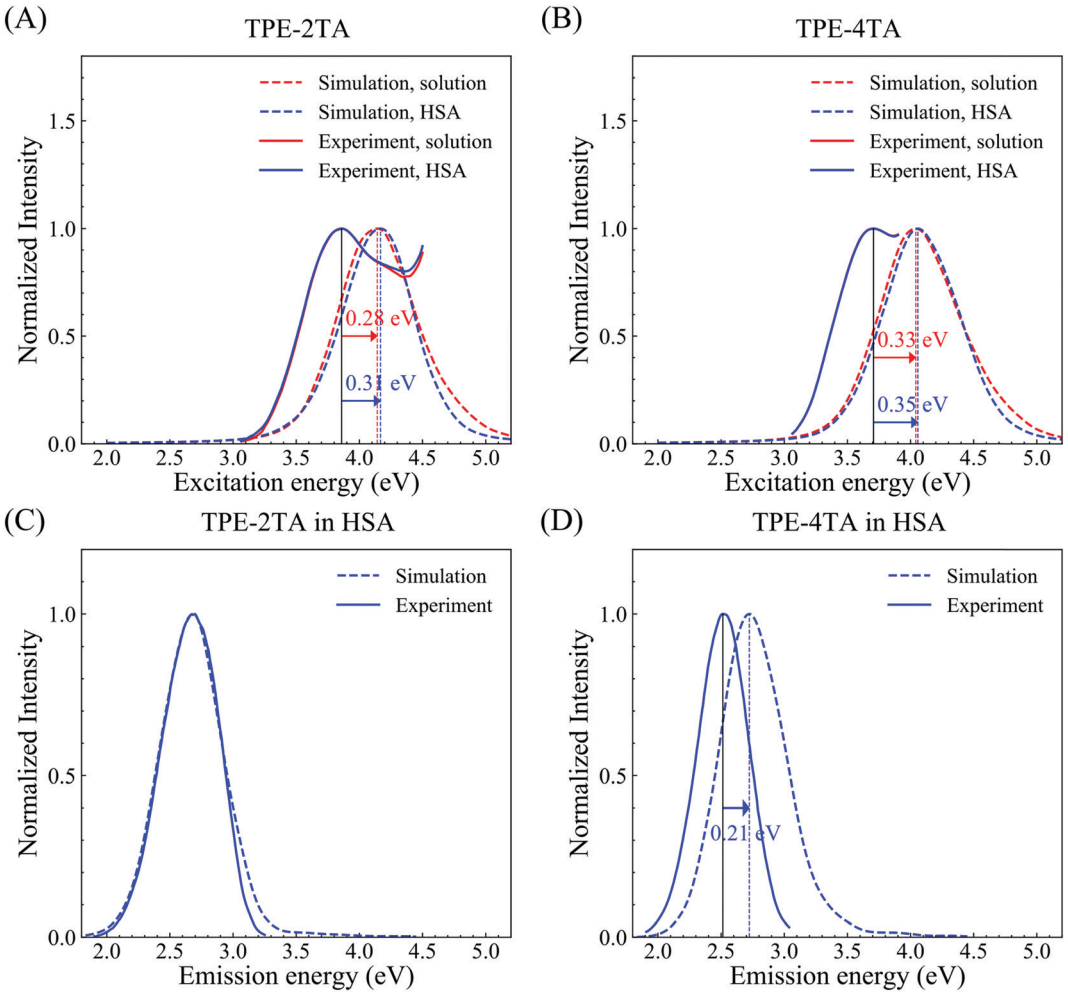

Fig. 3 The calculated absorption and steady-state fluorescence spectra are in good agreement with the experiment[4] for TPE-2TA and TPE-4TA. (A and B) Absorption spectra for TPE-2TA and TPE-4TA in the aqueous solution and HSA. (B and D) Fluorescence spectra for the TPE-2TA and TPE-4TA in the HSA. The blue and red curves indicate the AIEgens in the HSA and solution, respectively. The solid and dashed curves indicate the experimental and simulated spectra, respectively. The arrows and vertical lines depict the shift from the experimental to the simulated excitation/emission energies at maximum intensity. Only the $S_0 \rightarrow S_1$ ($S_1 \rightarrow S_0$) peaks are shown in the spectra because they were the dominant electronic transitions in the spectroscopic measurements.[4]

displacement of the phenyl rings, as opposed to the isomerization. Thus, contrary to the conventional RACI model, the protein confinement facilitates the AIEgen's access to the cyclization CI. Thus, the design of AIEgens needs to consider competing non-radiative decay pathways in the protein to maximize their fluorescence quantum yield. For example, methyl substitutions on the *ortho* positions of the phenyl rings may be necessary to increase the barrier along the cyclization pathway in the protein.[31] To the best of our knowledge, this is the first time that such a phenomenon is observed for the AIE in biomolecular systems.

Excited-state dynamics

The non-equilibrium excited-state dynamics following the photoexcitation confirm the conclusions from the equilibrium FES. For the TPE-2TA in the HSA, none of the 20 S_1 state trajectories reached any CI (defined as a less than 0.1 eV S_0–S_1 energy gap) within ~10 ps following the Frank–Condon excitation. In contrast, for the TPE-2TA in the aqueous solution, 40% of trajectories reached the isomerization CI within 10 ps, and none reached the cyclization CI. Thus, the HSA blocks the isomerization decay channel that can be readily accessed in the solution. Notably, although the protein lowers the cyclization energy barrier, none of the trajectories reached the cyclization CI within 10 ps even with the extra kinetic energy from the photoexcitation.

Assuming the intramolecular motions of the AIEgen quickly reaches equilibrium with the environment on the picosecond timescale following the photoexcitation,[34] we can estimate the reaction rate constants for accessing the CI on the S_1 state using the transition state theory:

$$k = k_B T/h \cdot \exp(-\Delta G^{\ddagger}/k_B T),$$

where k_B is the Boltzmann factor, T is the temperature, h is the Planck constant and $\Delta G^{\ddagger}$ is the free energy barrier. In the HSA,

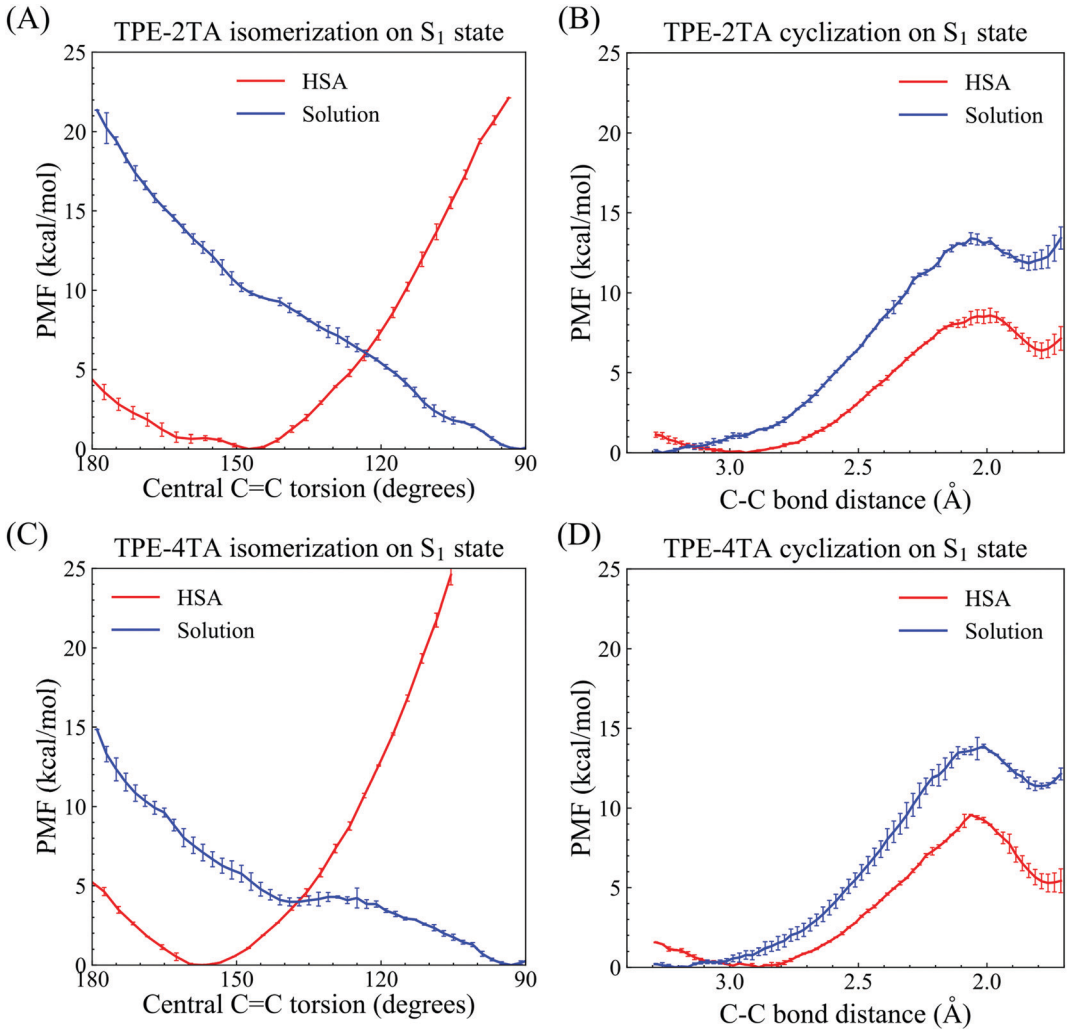

Fig. 4 Compared to the aqueous solution, the HSA impedes the isomerization but accelerates the cyclization of the TPE-2TA and TPE-4TA on the S_1 state. The S_1 state PMFs of the TPE-2TA (A and B) and TPE-4TA (C and D) in the aqueous solution (blue) and HSA (red) were calculated by the hh-TDA-BHLYP/MM umbrella sampling simulations. For both AIEgens, the cyclization is more favorable than the isomerization in the HSA, whereas the reverse is true in the aqueous solution.

the cyclization rate constant is $\sim$9.3 μs^{-1}, assuming $\Delta G^{\ddagger}$ = 8 kcal mol^{-1} and T = 300 K. The isomerization rate constant is $\sim$0.02 s^{-1} assuming $\Delta G^{\ddagger}$ = 20 kcal mol^{-1} at the same temperature, which is orders of magnitude slower than the cyclization. The typical fluorescence lifetime of TPE-derived AIEgens is on the sub-nanosecond to nanosecond timescale in confined environments such as protein aggregates.[8,67] Thus, in the HSA, the internal conversions are slow enough to allow auto-fluorescence of the AIEgens. In the solution, however, the internal conversion through the isomerization occurs on the picosecond timescale as observed from the non-equilibrium excited-state dynamics, which quenches the radiative decay through fluorescence.

Effect of basis set and docking structure of the ligand

Throughout our QM and QM/MM simulations, the 6-31G* basis set was used due to its computational efficiency for the excited-state umbrella sampling and non-equilibrium excited-state dynamics simulations. The accuracy of this basis set in combination with the hh-TDA-BHLYP has been benchmarked against the gas-phase XMS-CASPT calculations and condensed-phase experimental spectra (Fig. 2 and 3). Although the 6-31G* is not a large basis set, it is noteworthy that most conclusions in this work are based on the comparison of free energy barriers between different non-radiative decay pathways (cyclization *vs.* isomerization) and different molecular environments (solution *vs.* protein). In this regard, it is the difference in these free energy barriers that matter the most in the explanation of AIE phenomena, instead of the absolute energy barriers. Thus, although there could be errors in the absolute free energy barriers due to the use of a moderate basis set (6-31G*), such errors are not likely to alter the qualitative comparison between the PMFs. To further benchmark our choice of basis set, we carried out gas phase hh-TDA-BHLYP calculations of TPE-2TA using a larger basis set def2-TZVP, and the new gas-phase potential energy surfaces confirm our conclusion that the

isomerization is energetically preferred over the cyclization (Fig. S1, ESI†).

We also note that the scoring function from the docking simulation has limited accuracy. For each TPE derivative, the 2–3 docking structures having the highest scores belong to the same binding site and are geometrically similar, so we only take a single structure with the highest score as the starting structure. To mitigate the bias in the selection of the initial binding poses, we equilibrated the protein–ligand complexes in MD simulations for ~50 ns, where the ligands were allowed to adjust their position in the protein. Furthermore, we performed non-equilibrium QM/MM excited-state dynamics simulations starting from different snapshots during the MD trajectory covering a ~20 ns sampling of the ligand binding poses, which further mitigates the initial structural bias. Moreover, we simulated two TPE derivatives at two distinct positions inside the HSA (Fig. 1). Similar trends in the QM/MM PMF barriers were observed for these two binding sites (Fig. 4), which corroborates our conclusions. Thus, we do not expect running more QM/MM simulations will qualitatively change our conclusions. In practice, it is too computationally expensive to use *ab initio* excited-state QM/MM umbrella sampling simulations (as presented here) to evaluate the PMFs for all possible binding poses. Future work is necessary to explore the possibility of parameterizing reactive force fields to describe the isomerization and cyclization on the excited state, which could facilitate the free energy calculations.

Conclusions

By integrating the hh-TDA-DFT method in a multiscale simulation framework, we found that (1) the isomerization around the central ethylenic bond is the dominant non-radiative decay pathway for the TPE-2TA and TPE-4TA in the solution; (2) upon binding with the HSA, the protein environment significantly increases the free energy barrier for the non-radiative decay through photoisomerization, thus enhancing its fluorescence quantum yield in the protein; (3) interestingly, the protein environment accelerates the non-radiative decay through the cyclization reaction, but not to the extent that quenches the fluorescence. Our new findings reveal the complicated nature of the AIE phenomena in biological systems. To maximize the fluorescence quantum yield of the AIEgens, one needs to understand how the environment reweights the importance of competing non-radiative decay pathways. To this end, excited-state free energy barriers for accessing the CIs need to be quantified using an accurate first-principles electronic structure method combined with enhanced sampling techniques. Our work demonstrates that such a multiscale simulation framework can provide new insights into the AIE mechanisms that are not accessible by standard gas-phase simulations, which will benefit the design of effective AIEgens for disease diagnosis in the future.

Author contributions

The manuscript was written through contributions of all authors. Ruibin Liang designed the research, performed the simulations, analyzed the data and wrote the manuscript. Debojyoti Das and Amirhossein Bakhtiiari revised the manuscript. All authors have given approval to the final version of the manuscript.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

This work was supported by the startup funds from Texas Tech University. The researcher used GPU computing facilities provided by the High-Performance Computing Center at Texas Tech University.

References

1 J. Mei, N. L. C. Leung, R. T. K. Kwok, J. W. Y. Lam and B. Z. Tang, *Chem. Rev.*, 2015, **115**, 11718–11940.
2 R. T. K. Kwok, C. W. T. Leung, J. W. Y. Lam and B. Z. Tang, *Chem. Soc. Rev.*, 2015, **44**, 4228–4238.
3 Y. Hong, C. Feng, Y. Yu, J. Liu, J. W. Y. Lam, K. Q. Luo and B. Z. Tang, *Anal. Chem.*, 2010, **82**, 7035–7043.
4 Y. Tu, Y. Yu, Z. Zhou, S. Xie, B. Yao, S. Guan, B. Situ, Y. Liu, R. T. K. Kwok, J. W. Y. Lam, S. Chen, X. Huang, Z. Zeng and B. Z. Tang, *ACS Appl. Mater. Interfaces*, 2019, **11**, 29619–29629.
5 Z. Wang, P. Zhang, H. Liu, Z. Zhao, L. Xiong, W. He, R. T. K. Kwok, J. W. Y. Lam, R. Ye and B. Z. Tang, *ACS Appl. Mater. Interfaces*, 2019, **11**, 17306–17312.
6 C. W. T. Leung, F. Guo, Y. Hong, E. Zhao, R. T. K. Kwok, N. L. C. Leung, S. Chen, N. N. Vaikath, O. M. El-Agnaf, Y. Tang, W.-P. Gai and B. Z. Tang, *Chem. Commun.*, 2015, **51**, 1866–1869.
7 M. Kumar, Y. Hong, D. C. Thorn, H. Ecroyd and J. A. Carver, *Anal. Chem.*, 2017, **89**, 9322–9329.
8 A. Das, A. Gupta, Y. Hong, J. A. Carver and S. Maiti, *Biochemistry*, 2020, **59**, 1813–1822.
9 S. P. Clavant, T. M. Osicka and W. D. Comper, *Lab. Med.*, 2007, **38**, 35–38.
10 A. Farrugia, *Transfus. Med. Rev.*, 2010, **24**, 53–63.
11 A. Gatta, A. Verardo and M. Bolognesi, *Intern. Emerg. Med.*, 2012, **7**, 193–199.
12 A. S. Levey, C. Becker and L. A. Inker, *JAMA*, 2015, **313**, 837–846.
13 A. Akirov, H. Masri-Iraqi, A. Atamna and I. Shimon, *Am. J. Med.*, 2017, **130**, 1465.e1411–e1419.
14 D. Kumar and D. Banerjee, *Clin. Chim. Acta*, 2017, **469**, 150–160.
15 V. L. Villemagne, V. Doré, S. C. Burnham, C. L. Masters and C. C. Rowe, *Nat. Rev. Neurol.*, 2018, **14**, 225–236.
16 J. Luo, Z. Xie, J. W. Y. Lam, L. Cheng, H. Chen, C. Qiu, H. S. Kwok, X. Zhan, Y. Liu, D. Zhu and B. Z. Tang, *Chem. Commun.*, 2001, 1740–1741.

17 K. Kokado and K. Sada, *Angew. Chem., Int. Ed.*, 2019, **58**, 8632–8639.
18 Q. Peng, Y. Yi, Z. Shuai and J. Shao, *J. Am. Chem. Soc.*, 2007, **129**, 9333–9339.
19 Q. Wu, Q. Peng, Y. Niu, X. Gao and Z. Shuai, *J. Phys. Chem. A*, 2012, **116**, 3881–3888.
20 T. Zhang, Y. Jiang, Y. Niu, D. Wang, Q. Peng and Z. Shuai, *J. Phys. Chem. A*, 2014, **118**, 9094–9104.
21 S. Yin, Q. Peng, Z. Shuai, W. Fang, Y.-H. Wang and Y. Luo, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**, 205409.
22 Q. Li and L. Blancafort, *Chem. Commun.*, 2013, **49**, 5966–5968.
23 Q.-S. Li, W.-L. Ding, X.-L. Peng, G. Cui, Z.-S. Li and L. Blancafort, *ChemPhotoChem*, 2019, **3**, 814–824.
24 S. Ruiz-Barragan, K. Morokuma and L. Blancafort, *J. Chem. Theory Comput.*, 2015, **11**, 1585–1594.
25 X.-L. Peng, S. Ruiz-Barragan, Z.-S. Li, Q.-S. Li and L. Blancafort, *J. Mater. Chem. C*, 2016, **4**, 2802–2810.
26 M. Dommett, M. Rivera and R. Crespo-Otero, *J. Phys. Chem. Lett.*, 2017, **8**, 6148–6153.
27 N. Yamamoto, *J. Phys. Chem. A*, 2020, **124**, 4939–4945.
28 M. Dommett and R. Crespo-Otero, *Phys. Chem. Chem. Phys.*, 2017, **19**, 2409–2416.
29 L. Stojanović and R. Crespo-Otero, *J. Phys. Chem. C*, 2020, **124**, 17752–17761.
30 A. Prlj, N. Došlić and C. Corminboeuf, *Phys. Chem. Chem. Phys.*, 2016, **18**, 11606–11609.
31 Y.-J. Gao, X.-P. Chang, X.-Y. Liu, Q.-S. Li, G. Cui and W. Thiel, *J. Phys. Chem. A*, 2017, **121**, 2572–2579.
32 T. Tran, A. Prlj, K.-H. Lin, D. Hollas and C. Corminboeuf, *Phys. Chem. Chem. Phys.*, 2019, **21**, 9026–9035.
33 G.-J. Zhao, K.-L. Han, Y.-B. Lei and Y.-S. Dou, *J. Chem. Phys.*, 2007, **127**, 094307.
34 S. Kayal, K. Roy and S. Umapathy, *J. Chem. Phys.*, 2018, **148**, 024301.
35 Y. Cai, L. Du, K. Samedov, X. Gu, F. Qi, H. H. Y. Sung, B. O. Patrick, Z. Yan, X. Jiang, H. Zhang, J. W. Y. Lam, I. D. Williams, D. Lee Phillips, A. Qin and B. Z. Tang, *Chem. Sci.*, 2018, **9**, 4662–4670.
36 K. Kokado, T. Machida, T. Iwasa, T. Taketsugu and K. Sada, *J. Phys. Chem. C*, 2018, **122**, 245–251.
37 J. Guan, R. Wei, A. Prlj, J. Peng, K.-H. Lin, J. Liu, H. Han, C. Corminboeuf, D. Zhao, Z. Yu and J. Zheng, *Angew. Chem., Int. Ed.*, 2020, **59**, 14903–14909.
38 E. Lenderink, K. Duppen and D. A. Wiersma, *J. Phys. Chem.*, 1995, **99**, 8972–8977.
39 R. W. J. Zijlstra, P. T. van Duijnen, B. L. Feringa, T. Steffen, K. Duppen and D. A. Wiersma, *J. Phys. Chem. A*, 1997, **101**, 9828–9836.
40 B. G. Levine, C. Ko, J. Quenneville and T. J. MartÍnez, *Mol. Phys.*, 2006, **104**, 1039–1051.
41 J. K. Yu, C. Bannwarth, E. G. Hohenstein and T. J. Martínez, *J. Chem. Theory Comput.*, 2020, **16**, 5499–5511.
42 C. Bannwarth, J. K. Yu, E. G. Hohenstein and T. J. Martínez, *J. Chem. Phys.*, 2020, **153**, 024110.
43 R. Liang, *J. Chem. Theory Comput.*, 2021, **17**, 3019–3030.
44 J. K. Yu, C. Bannwarth, R. Liang, E. G. Hohenstein and T. J. Martínez, *J. Am. Chem. Soc.*, 2020, **142**, 20680–20690.
45 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
46 C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
47 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372–1377.
48 T. Shiozaki, W. Győrffy, P. Celani and H.-J. Werner, *J. Chem. Phys.*, 2011, **135**, 081106.
49 I. S. Ufimtsev and T. J. Martinez, *J. Chem. Theory Comput.*, 2009, **5**, 2619–2628.
50 A. V. Titov, I. S. Ufimtsev, N. Luehr and T. J. Martinez, *J. Chem. Theory Comput.*, 2013, **9**, 213–221.
51 S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, S. I. L. Kokkila-Schumacher, N. Luehr, J. W. Snyder, C. Song, A. V. Titov, I. S. Ufimtsev and T. J. Martinez, *J. Chem. Phys.*, 2020, **152**, 224110.
52 S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder Jr, C. Song, A. V. Titov, I. S. Ufimtsev, L.-P. Wang and T. J. Martínez, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1494.
53 B. Webb and A. Sali, *Curr. Protoc. Bioinform.*, 2016, **54**, 5.6.1–5.6.37.
54 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
55 D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, C. Jin, K. Kasavajhala, M. C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, Y. Xue, D. M. York, S. Zhao and P. A. Kollman, University of California, San Francisco, 2021.
56 L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.
57 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
58 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.
59 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
60 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
61 C. J. Dickson, B. D. Madej, Å. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J. Chem. Theory Comput.*, 2014, **10**, 865–879.
62 Y. J. Wu, H. L. Tepper and G. A. Voth, *J. Chem. Phys.*, 2006, **124**, 024503.

63 P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts and V. S. Pande, *J. Chem. Theory Comput.*, 2013, **9**, 461–469.
64 B. F. E. Curchod, A. Sisto and T. J. Martínez, *J. Phys. Chem. A*, 2017, **121**, 265–276.
65 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
66 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1339–1350.
67 W. Wang, T. Lin, M. Wang, T.-X. Liu, L. Ren, D. Chen and S. Huang, *J. Phys. Chem. B*, 2010, **114**, 5983–5988.