

This is the author's final, peer-reviewed manuscript as accepted for publication (AAM). The version presented here may differ from the published version, or version of record, available through the publisher's website. This version does not track changes, errata, or withdrawals on the publisher's site.

Reactive molecular dynamics at constant pressure via non-reactive force fields: extending the Empirical Valence Bond method to the isothermal-isobaric ensemble

Ivan Scivetti, Kakali Sen, Alin M. Elena and Ilian Todorov

Published version information

Citation: I Scivetti et al. 'Reactive molecular dynamics at constant pressure via nonreactive force fields: extending the Empirical Valence Bond method to the isothermal-isobaric ensemble.' *Journal of Physical Chemistry*, vol. A 124, no. 37 (2020): 7585-7597.

DOI: [10.1021/acs.jpca.0c05461](https://doi.org/10.1021/acs.jpca.0c05461)

This document is the unedited author's version of a Submitted Work that was subsequently accepted for publication in *Journal of Physical Chemistry A*, © American Chemical Society after peer review. To access the final edited and published work see DOI above.

Please cite only the published version using the reference above. This is the citation assigned by the publisher at the time of issuing the AAM. Please check the publisher's website for any updates.

Reactive molecular dynamics at constant pressure via non-reactive force fields: extending the Empirical Valence Bond method to the isothermal-isobaric ensemble

Ivan Scivetti,^{*,†,‡} Kakali Sen,[†] Alin M. Elena,[†] and Ilian Todorov[†]

[†]*Daresbury Laboratory, Sc. Tech. Keckwick Lane, Daresbury, WA4 4AD, Warrington, UK*

[‡]*Department of Chemistry, University of Liverpool, Liverpool L69 3BX, UK*

E-mail: ivan.scivetti@stfc.ac.uk

Abstract

The Empirical Valence Bond (EVB) method offers a suitable framework to obtain reactive potentials through the coupling of non-reactive force fields. In this formalism, most of the implemented coupling terms are built using functional forms that depend on spatial coordinates, while parameters are fitted against reference data to model the change of chemistry between the participating non-reactive states. In this work, we demonstrate that the use of such coupling terms precludes the computation of the stress tensor for condensed phase systems and prevents the possibility to carry out EVB molecular dynamics in the isothermal-isobaric (NPT) ensemble. Alternatively, we make use of coupling terms that depend on the energy gaps, defined as the energy differences between the participating non-reactive force fields, and derive an general expression for the EVB stress tensor suitable for computation. Implementation of this new methodology is tested for a model of a single reactive malonaldehyde solvated in

non-reactive water. Mass densities and probability distributions for the values of the energy gaps computed in the NPT ensemble reveal a negligible role of the reactive potential in the limit of low concentrated solutions, thus corroborating for the first time the validity of approximations based on the canonical NVT ensemble, customarily adopted for EVB simulations. The presented formalism also aims to contribute to future implementations and extensions of the EVB method to research the limit of highly-concentrated solutions.

1 Introduction

Molecular dynamics (MD) simulations offer a powerful computational tool to derive atomistic insight of complex phenomena from organic chemistry and biochemistry to heterogeneous catalysis.¹⁻³ The interatomic interactions in classical MD simulations are based on force field (FF) descriptors that allow for very fast computation of the interactions and access to simulate very large systems. Commonly, these FF descriptors have simple functional forms, with parameters either fitted to experimental data or derived from quantum mechanical calculations.^{4,5} In most of the available FF libraries, functional forms and fitted parameters remain unchanged during the course of the MD simulation. In reactive processes, however, the interactions inevitably change due to the breaking and/or formation of chemical species. Thus, standard FFs are not suitable to simulate chemical reactions and they are referred to as non-reactive.

An alternative to simulate chemical reactions with MD is given by Reactive FFs (RFFs),⁶⁻¹¹ that are designed to model interatomic interactions of multiple states representing different chemical species. The task of designing RFFs, however, is very challenging¹ and, despite the enormous progress over the last years,^{7,19} a general parameterization is not yet available.

The Empirical Valence Bond (EVB) method²⁰⁻²⁵ offers, instead, a simple general framework

¹Indeed, designing RFFs requires a high level of expertise to tackle a multi-dimensional problem,¹¹ where the modelled interactions are often expressed by complicated functional forms with many strongly coupled parameters that are optimized via the use of sophisticated tools.¹²⁻¹⁸

to model reactive processes through the coupling of multiple non-reactive FFs, where each FF corresponds to a different chemical state for the system. In this method, a suitable EVB matrix is built using the computed energies of the involved chemical states as well as appropriate coupling terms. Matrix diagonalization at each time step allows computation of reactive energy landscapes that account for the change in chemistry when sampling conformations between the participating, chemically different, states.

In contrast to RFFs, the advantage of the EVB method lies in the large availability of standard non-reactive FFs libraries. This feature has supported the implementation and development of the EVB method over the last four decades.²⁶⁻²⁹ In addition, despite the initial task to calibrate the coupling terms against reference data, research has demonstrated that these couplings are invariant to the surrounding electrostatics, making it possible to simulate the same reactive unit in different environments.³⁰ This convenient feature of the EVB method has widely increased its recognition as a powerful tool within the computational chemistry community.²⁴

The EVB method has successfully been applied to a wide range of complex phenomena, including reaction dynamics in gas phase and solution, infrared spectroscopy, enzyme reactions and catalysis, etc. Likewise, previous EVB research has allowed the study of intramolecular proton transfer, proton transport, solvation and transport of the hydrated protons, hydrogen atom transfer, and hydration structure in dilute acids (see for example Table 1 of ref. 7). We refer the interested reader to ref. 24 and references therein for a detailed list of applications investigated with EVB.

For condensed phase systems, reported MD simulations with the EVB method (herein, MD-EVB simulations) are often conducted either in the microcanonical (NVE) or the canonical (NVT) ensemble. In contrast, reported EVB simulations at constant pressure are less common in the literature. One example is the work of J. Åqvist et al.³¹ who used the Free Energy Perturbation method to investigate enzymatic reactions at ultra-high pressures via the computation of the free energy profile, but without including the actual reaction dynam-

ics in the simulation. Another example is the work P. Raiteri et al.³² who computed the proton diffusion in BaZrO₃, where the reactive force field was built using pair-wise interactions only. The purpose of this work is to present a general EVB formalism that allows MD-EVB simulations of systems with flexible molecular reactive sites at constant pressure and temperature, i.e. using the isothermal-isobaric (NPT) ensemble, while simultaneously accounting for the actual reaction dynamics of the whole system.

To date, there is no established protocol to perform MD-EVB simulations, as the most appropriate strategy depends on the system and the process to be simulated. A possible protocol for EVB simulations in condensed phase is to first consider only one of the chemical states of the system (preferably the state with lowest free energy) together with the surrounding, non-reactive, environment and carry out a standard NPT simulation (without EVB) at the target pressure and temperature.³³ The converged volume is then fixed and the MD-EVB simulation is performed using the NVT ensemble. This protocol is adopted here and named as NVT-based protocol. While this procedure appears to be a sensible strategy for aqueous systems with low concentrations of reactive sites, its validity to approximate real experimental conditions at constant pressure and temperature has never been corroborated to date.

In this work, we demonstrate that the use of the standard formulation to compute the stress tensor cannot be directly applied to derive the components of the EVB stress tensor. We argue that this limitation explains the absence of MD-EVB simulations in the NPT ensemble. In contrast to using complex spatial variables to fit the coupling terms of the EVB matrix, we instead propose to use of energy gaps,³⁴ defined as the energy difference between the non-reactive FFs. With this choice, we derive a general expression for the EVB stress tensor suitable for computational implementation, not only offering a solution to an overlooked limitation of EVB but also extending the applicability of MD-EVB simulations to NPT ensembles. The computational implementation of this new formalism is tested using a model of a solvated reactive malonaldehyde molecule in water. MD-EVB simulations at 300 K and 1 atm are used to quantify the role of the reactive potential in the computed density

and classical probability distributions of the energy gaps obtained from the sampling of the configurational space. Results also allow to evaluate the validity of the NVT-based protocol for MD-EVB simulations described above, while the derived method offers an opportunity to explore new strategies for future implementation and development of the EVB method. The structure of this paper is as follows. The fundamentals of the EVB method developed over the years are presented in a convenient notation in section 2. In section 3, we discuss the limitation of the standard formulation to calculate the EVB stress tensor, and propose a new alternative method. A brief overview of the computational implementation is given in section 4. Section 5 discusses general aspects of the coupling terms within the framework of the present paper. Details of the model and MD computational setting are provided in section 6, which is followed by section 7 with the results and discussion. Concluding remarks are finally addressed in section 8.

2 The EVB method

In this section we present the fundamentals of the EVB formalism in a convenient notation. Let us assume an atomic system composed of N_p particles with positions described by the set of vectors $\{\mathbf{R}\}$. The non-reactive force field (FF) for the chemical state m is described by the configurational energy $E_c^{(m)}(\{\mathbf{R}\})$ and the set of forces $\vec{F}_J^{(m)}(\{\mathbf{R}\})$, where the index J runs over the total number of particles. The configurational energy function $E_c^{(m)}(\{\mathbf{R}\})$ can be generally written as a sum of different terms as follows³⁵

$$\begin{aligned}
 E_c^{(m)}(\{\mathbf{R}\}) &= [E_{shell}^{(m)} + E_{teth}^{(m)} + E_{bond}^{(m)} + E_{ang}^{(m)} + E_{dih}^{(m)} + \\
 &+ E_{inv}^{(m)} + E_{3body}^{(m)} + E_{4body}^{(m)} + E_{ters}^{(m)} + \\
 &+ E_{metal}^{(m)} + E_{vdw}^{(m)} + E_{coul}^{(m)}](\{\mathbf{R}\})
 \end{aligned} \tag{1}$$

where $E_{shell}^{(m)}$, $E_{teth}^{(m)}$, $E_{bond}^{(m)}$, $E_{ang}^{(m)}$, $E_{dih}^{(m)}$, $E_{inv}^{(m)}$, $E_{3body}^{(m)}$, $E_{4body}^{(m)}$, $E_{ters}^{(m)}$, $E_{metal}^{(m)}$, $E_{vdw}^{(m)}$ and $E_{coul}^{(m)}$ are the interactions representing core-shell polarization, tethered particles, chemical bonds, valence angles, dihedrals, inversion angles, three-body, four-body, Tersoff, metallic, van der Waals and coulombic contributions, respectively. Following Eq. (1), the forces can be expressed using a similar decomposition. In the current notation, we shall use indexes m and k for the chemical states (and FFs), I and J for atoms and Greek letters for Cartesian coordinates. Indexes in parenthesis are used to emphasize the particular chemical state.

The purpose of the EVB method is to couple N_F non-reactive force fields to obtain a reactive potential. These FFs are coupled through the Hamiltonian \hat{H}_{EVB} with a matrix representation $H_{EVB} \in \mathcal{R}^{N_F \times N_F}$ that has the following components

$$H_{EVB}^{mk}(\{\mathbf{R}\}) = \begin{cases} E_c^{(m)}(\{\mathbf{R}\}) & m = k \\ C_{mk}(\epsilon_{mk}) & m \neq k \end{cases} \quad (2)$$

where each diagonal element corresponds to the configurational energy $E_c^{(m)}(\{\mathbf{R}\})$ of the non-reactive FF that models the interactions as if the system was in the chemical state (m), whereas the off-diagonal terms C_{mk} are the couplings between states m and k . For convenience in the notation, we shall omit hereinafter the dependence on the set of coordinates $\{\mathbf{R}\}$ for the particles. Even though there are different possible choices for the coupling terms, in the above definition we have set C_{mk} to depend on $\epsilon_{mk} = E_c^{(m)} - E_c^{(k)} = -[E_c^{(k)} - E_c^{(m)}] = -\epsilon_{km}$, where ϵ_{mk} is commonly referred to as energy gap and defines a possible reaction coordinate for the reactive process.^{22,24,34,36} Since the H_{EVB} matrix is Hermitian by construction and the C_{mk} terms are real, the condition of $C_{mk} = C_{km}$ must be imposed to the off-diagonal elements. Diagonalization of H_{EVB} leads to N_F possible eigenvalues $\{\lambda_1, \dots, \lambda_{N_F}\}$ with

$$H_{EVB}\Psi_{\lambda_m} = \lambda_m\Psi_{\lambda_m}, \quad m = 1, \dots, N_F. \quad (3)$$

The EVB energy, E_{EVB} , is defined as the lowest eigenvalue

$$E_{\text{EVB}} = \min(\lambda_1, \dots, \lambda_{N_F}) \quad (4)$$

with the corresponding normalized EVB eigenvector

$$\Psi_{\text{EVB}} = \Psi_{\min(\lambda_1, \dots, \lambda_{N_F})}. \quad (5)$$

and

$$E_{\text{EVB}} = \langle \Psi_{\text{EVB}} | \hat{H}_{\text{EVB}} | \Psi_{\text{EVB}} \rangle. \quad (6)$$

Since the eigenvector Ψ_{EVB} is real and normalized we have

$$\sum_{k=1}^{N_F} |\Psi_{\text{EVB}}^{(k)}|^2 = 1 \quad (7)$$

from which we can interpret $|\Psi_{\text{EVB}}^{(k)}|^2$ as the fraction of the chemical state (k) being part of the EVB state. The eigenvector Ψ_{EVB} can also be represented as a column vector $\in \mathcal{R}^{N_F \times 1}$ where $\Psi_{\text{EVB}}^{(k)}$ is the element of the k -row. Thus, Eq. (6) is expressed as a matrix multiplication

$$E_{\text{EVB}} = \sum_{m,k=1}^{N_F} \tilde{\Psi}_{\text{EVB}}^{(m)} H_{\text{EVB}}^{mk} \Psi_{\text{EVB}}^{(k)} \quad (8)$$

where $\tilde{\Psi}_{\text{EVB}}$ is the transpose of Ψ_{EVB} . In Appendix A we demonstrate that the decomposition of E_{EVB} into different types of interactions (bonds, angles, etc) as for $E_c^{(m)}$ in Eq. (1) is not well defined.

The resulting EVB force over the particle J , \vec{F}_J^{EVB} , follows from the Hellman-Feynman

theorem³⁷

$$\begin{aligned}\vec{F}_J^{\text{EVB}} &= -\nabla_{\vec{R}_J} E_{\text{EVB}} = -\langle \Psi_{\text{EVB}} | \nabla_{\vec{R}_J} \hat{H}_{\text{EVB}} | \Psi_{\text{EVB}} \rangle \\ &= \sum_{\alpha=x,y,z} F_{J\alpha}^{\text{EVB}} \check{\alpha}\end{aligned}\quad (9)$$

where $\check{\alpha}$ corresponds to each of the orthonormal Cartesian vectors and

$$F_{J\alpha}^{\text{EVB}} = -\langle \Psi_{\text{EVB}} | \frac{\partial \hat{H}_{\text{EVB}}}{\partial R_{J\alpha}} | \Psi_{\text{EVB}} \rangle. \quad (10)$$

From Eq. (2) the matrix components of the operator $\frac{\partial \hat{H}_{\text{EVB}}}{\partial R_{J\alpha}}$ are given as follows

$$\frac{\partial H_{\text{EVB}}^{mk}}{\partial R_{J\alpha}} = \begin{cases} \frac{\partial E_c^{(m)}}{\partial R_{J\alpha}} = -F_{J\alpha}^{(m)} & m = k \\ \frac{dC_{mk}}{dR_{J\alpha}} = \frac{dC_{mk}(\epsilon_{mk})}{d\epsilon_{mk}} \frac{\partial \epsilon_{mk}}{\partial R_{J\alpha}} & m \neq k \\ = \frac{dC_{mk}(\epsilon_{mk})}{d\epsilon_{mk}} \left[\frac{\partial E_c^{(m)}}{\partial J\alpha} - \frac{\partial E_c^{(k)}}{\partial J\alpha} \right] \\ = C'_{mk} [F_{J\alpha}^{(k)} - F_{J\alpha}^{(m)}] \end{cases} \quad (11)$$

where $C'_{mk} = \frac{dC_{mk}(\epsilon_{mk})}{d\epsilon_{mk}}$ and $F_{J\alpha}^{(k,m)}$ is the α component of the total configurational force over particle J in the chemical state (k, m) . Similarly to Eq. (8), Eq. (10) can be expressed as a matrix multiplication

$$F_{J\alpha}^{\text{EVB}} = - \sum_{m,k=1}^{N_F} \tilde{\Psi}_{\text{EVB}}^{(m)} \left(\frac{\partial H_{\text{EVB}}^{mk}}{\partial R_{J\alpha}} \right) \Psi_{\text{EVB}}^{(k)}. \quad (12)$$

The above equations define the standard EVB force field (EVB-FF). Even though the EVB formalism was first developed to compute molecular systems, EVB is also applicable to extended systems, customarily modelled using the supercell approximation and periodic boundary conditions (PBCs). Nevertheless, MD-EVB simulations have only been conducted for the NVE and NVT ensembles, to the best of our knowledge, as there is no evidence of

a previously reported method to compute the EVB stress tensor. In the next section, we discuss the intricacies related to computing the stress tensor using the standard formulation and propose a new method that allows extending the applicability of MD-EVB to NPT ensembles the first time.

3 The EVB stress tensor

The key requirement for a NPT simulation with the EVB method is to being able to compute the EVB stress tensor σ^{EVB} . Similarly to the energy and forces, the configurational stress tensor for the force field m , $\sigma^{c(m)}$, can be decomposed in a general expression equivalent to Eq. (1), where each contribution is computed separately using well-known functional forms.^{35,38} For *bonded* interactions, for example, the $\alpha\beta$ contribution to the stress tensor from particle J due to the bonded interactions with the surrounding particles, $\sigma_{J,\alpha\beta}^{\text{bond}(m)}$, is given by

$$\sigma_{J,\alpha\beta}^{\text{bond}(m)} = \sum_I R_{JI,\alpha} f_{IJ,\beta}^{\text{bond}(m)} \quad (13)$$

where $R_{JI,\alpha}$ is the α component of the vector separation $\vec{R}_{JI} = \vec{R}_J - \vec{R}_I$ between particles I and J , and $\vec{f}_{IJ}^{\text{bond}(m)}$ the bond force over particle J from its bonded interaction with particle I . In Eq. (13) the sum runs over all particles I interacting with particle J via bonds. Analogously, we could in principle propose an expression for the $\alpha\beta$ component of the EVB stress tensor resulting from the EVB bonded forces, $f_{IJ,\beta}^{\text{EVB}}$, as follows,

$$\sigma_{J,\alpha\beta}^{\text{EVB}} = \sum_I R_{JI,\alpha} f_{IJ,\beta}^{\text{EVB}}. \quad (14)$$

In the present case of bonded interactions, the evaluation of Eq. (14) requires of each individual EVB-bonded force over particle J from interaction with particles I , given by $\vec{f}_{IJ}^{\text{EVB}}$. Nevertheless, the EVB force given in Eq. (10) represents the total force, \vec{F}_J^{EVB} , resulting from the interaction of particle J with all the neighboring particles, which generally include

other type of interactions apart from bonding interactions. As far as we can discern, each individual contribution to the force $\vec{f}_{IJ}^{\text{EVB}}$ cannot be computed from the EVB formalism presented in last section and, consequently, the evaluation of the stress tensor via Eq. (14) is not possible. The same reasoning applies to other type of interactions. This limitation precludes the computation of the stress tensor within the EVB formalism via standard formulae and, consequently, MD simulations using the NPT ensemble. Surprisingly, this inherent limitation of the EVB method has not been previously discussed in the literature, to the best of our knowledge.

To circumvent this problem, we propose to use the well-known relation between the configurational energy and the configurational stress tensor³⁹

$$\frac{\partial E_c^{(k)}}{\partial h_{\alpha\beta}} = -V \sum_{\gamma=x,y,z} \sigma_{\alpha\gamma}^{c(k)} h_{\beta\gamma}^{-1} \quad (15)$$

where h is the set of lattice vectors of the supercell with volume $V=\det(h)$. Multiplying to the left by $h_{\nu\beta}$ and summing over β we obtain the inverse relation to Eq. (15)

$$\sigma_{\alpha\beta}^{c(k)} = -\frac{1}{V} \sum_{\gamma=x,y,z} h_{\beta\gamma} \frac{\partial E_c^{(k)}}{\partial h_{\alpha\gamma}} \quad (16)$$

which can be used to define the EVB stress tensor

$$\sigma_{\alpha\beta}^{\text{EVB}} = -\frac{1}{V} \sum_{\gamma=x,y,z} h_{\beta\gamma} \frac{\partial E_{\text{EVB}}}{\partial h_{\alpha\gamma}}. \quad (17)$$

Similar to the definition of the EVB force, we evaluate $\partial E_{\text{EVB}}/\partial h_{\alpha\gamma}$ using the Eq. (6) and the Hellman-Feynman theorem³⁷

$$\frac{\partial E_{\text{EVB}}}{\partial h_{\alpha\beta}} = \langle \Psi_{\text{EVB}} | \frac{\partial \hat{H}_{\text{EVB}}}{\partial h_{\alpha\beta}} | \Psi_{\text{EVB}} \rangle. \quad (18)$$

The matrix components of the operator $\frac{\partial \hat{H}_{\text{EVB}}}{\partial h_{\alpha\beta}}$ follow from the definition of the EVB matrix

(2) and the use of relation (15)

$$\frac{\partial H_{\text{EVB}}^{mk}}{\partial h_{\alpha\beta}} = \begin{cases} \frac{\partial E_c^{(m)}}{\partial h_{\alpha\beta}} = -V \sum_{\gamma} \sigma_{\alpha\gamma}^{c(m)} h_{\beta\gamma}^{-1} & m = k \\ \frac{dC_{mk}}{dh_{\alpha\beta}} = \frac{dC_{mk}(\epsilon_{mk})}{d\epsilon_{mk}} \frac{\partial \epsilon_{mk}}{\partial h_{\alpha\beta}} & m \neq k \\ = \frac{dC_{mk}(\epsilon_{mk})}{d\epsilon_{mk}} \left[\frac{\partial E_c^{(m)}}{\partial h_{\alpha\beta}} - \frac{\partial E_c^{(k)}}{\partial h_{\alpha\beta}} \right] & \\ = -VC'_{mk} \sum_{\gamma} [\sigma_{\alpha\gamma}^{c(m)} - \sigma_{\alpha\gamma}^{c(k)}] h_{\beta\gamma}^{-1}. & \end{cases}$$

Finally, the EVB stress tensor of Eq. (17) can be expressed as a matrix multiplication

$$\sigma_{\alpha\beta}^{\text{EVB}} = -\frac{1}{V} \sum_{\gamma=x,y,z} h_{\beta\gamma} \sum_{m,k=1}^{N_F} \tilde{\Psi}_{\text{EVB}}^{(m)} \left(\frac{\partial H_{\text{EVB}}^{mk}}{\partial h_{\alpha\beta}} \right) \Psi_{\text{EVB}}^{(k)}. \quad (19)$$

These expressions provide an alternative to compute the stress tensor σ^{EVB} from the configurational stress tensors of each non-reactive FF, $\sigma_{\alpha\gamma}^{c(k)}$. It is important to note that this new scheme to compute σ^{EVB} can only be derived if one uses functional forms for C_{mk} that depend on the energy differences ϵ_{mk} , for which one can evaluate $\frac{\partial E_c^{(m)}}{\partial h_{\alpha\beta}} - \frac{\partial E_c^{(k)}}{\partial h_{\alpha\beta}}$ and use relation (15) with the computed configurational stress tensor for each chemical state. In contrast, if the choice was to use coupling terms that do not depend on ϵ_{mk} but other degrees of freedom such as spatial coordinates (see refs. 40–46), we cannot discern a clear logic to derive an expression for σ^{EVB} , which might explain the fact there is no evidence of any previous reported method to compute the stress tensor using EVB.

So far we have presented an alternative to compute the stress tensor $\sigma_{\alpha\beta}^{\text{EVB}}$ but have not discussed the total virial \mathcal{V}_{EVB} . Similarly to the stress tensor, the inability to compute individual contributions of the EVB force prevents the evaluation of the virial using the standard formulation,³⁵ and the usual decomposition of the virial depending of the type of interaction under consideration. Within the presented formalism, we compute the virial \mathcal{V}_{EVB} from $\sigma_{\alpha\beta}^{\text{EVB}}$

as follows

$$\mathcal{V}_{\text{EVB}} = - \sum_{\alpha=x,y,z} \sigma_{\alpha\alpha}^{\text{EVB}}. \quad (20)$$

In contrast to the EVB energy, it is possible to decomposed the virial into different type of interactions, as we demonstrate in Appendix B. The total stress tensor, σ^T , is given by the following general expression

$$\sigma^T = \sigma^{\text{kin}} + \sigma^{\text{EVB}} + \sigma^{\text{RB}} + \sigma^{\text{bc}} \quad (21)$$

where σ^{kin} , σ^{RB} and σ^{bc} are the contributions to the stress tensor from the kinetic energy, rigid bodies (RB) and bond constraints (bc), respectively. The EVB method only accounts for the configurational interactions, as described. The kinetic stress tensor is computed as usual from the instantaneous velocities of the particles.³⁵ For a particle that is part of a rigid body, the only possible interactions are intermolecular non-bonded interactions (such as coulombic and van der Waals interactions) with other neighboring particles that are not part of the same rigid body. Following the computation of the EVB forces via Eq. (10), the contribution to the stress from the rigid bodies follows from refs. 38 and 39

$$\sigma_{\alpha\beta}^{\text{RB}} = \sum_{\mathcal{B}=1}^{N_{\text{RB}}} \sum_{I=1}^{\eta_{\mathcal{B}}} F_{I_{\mathcal{B}},\alpha}^{\text{EVB}} d_{I_{\mathcal{B}},\beta} \quad (22)$$

where $\vec{F}_{I_{\mathcal{B}}}$ is the total force over particle I of rigid body \mathcal{B} and $\vec{d}_{I_{\mathcal{B}}}$ the vector distance from atom $I_{\mathcal{B}}$ to the center of mass of the rigid body \mathcal{B} . In the above expression, index \mathcal{B} runs over all the rigid bodies. Each rigid body is composed of $\eta_{\mathcal{B}}$ particles. Since, by definition, the topology of rigid bodies remain unaltered during the simulation, the use of RBs within in the present framework is meaningful only to model the environment interacting reactive EVB site. A common example is the use of rigidly constrained water molecules to model a solution. Contributions to the stress tensor from bond constraints, $\sigma_{\alpha\beta}^{\text{bc}}$, are obtained using the SHAKE/RATTLE algorithm^{47,48} during the course of the simulation. This algorithm is independent of the EVB

formalism, and corrects for the dynamics of the constrained particles. Finally, frozen particles do not contribute to the stress tensor and are not considered in the formalism. It is important to note that the topology defined via the setting of RBs, frozen atoms and bond constraints must be consistent for all the coupled FFs, as they impose well defined conditions for the dynamics. For example, if a group of atoms form a rigid body, they must remain a rigid body independently of chemical state under consideration.

4 Overview of the computational implementation

The EVB method described in section 2 and its extension for the computation of the stress tensor (section 3) were implemented within the DL_POLY_4 code.^{49,50} In the standard format, DL_POLY_4 reads the initial coordinates, velocities and forces from the CONFIG file. Each particle is labelled according to its specification in the FIELD file, which contains the information of the FF type and parameters for the interactions between the particles. Settings for the MD simulation are specified in the CONTROL file. Initially, the code was modified to allow i) reading multiple (N_F) CONFIG and FIELD files, ii) allocating arrays of dimension N_F for the relevant quantities, iii) checking consistency of specification between all force fields and initial coordinates (including any possible constraint such as rigid bodies), iv) reading EVB settings such as coupling terms and v) preventing the execution if there are MD or FF options that are not consistent with a EVB simulation. With regards to this last point, not all type of interactions in the energy decomposition of Eq. (1) are suitable to describe reactive interactions. For example, three-body, four-body, Tersoff and metallic interactions are, by construction, not designed to account for different chemical states. Thus, such interactions should only be used to model the surrounding atomic environment interacting with the EVB site.

Regarding the EVB method in itself, modifications to the code required to allow for the computations of energies, forces, stress tensor and virials for each of the N_F force-fields

separately. From the computed configurational energy of each FF and the choice of the functional forms for the coupling terms, the EVB matrix (2) is built and diagonalized, and the lowest eigenvalue and the corresponding vector are assigned to E_{EVB} and Ψ_{EVB} , respectively. Matrix (11) is computed for each particle’s Cartesian components and the resulting EVB force is obtained via the matrix multiplication of Eq. (12). From the stress tensors computed for each FF, matrix (19) is built for all the $\alpha\beta$ terms and the $\alpha\beta$ component of the EVB stress tensor obtained via Eq. (19), and the total virial from Eq. (20). Such EVB calculations are conducted for each time step taking advantage of the domain decomposition as implemented in DL_POLY_4.^{49,50}

In this implementation, all the N_F force fields are computed in a loop architecture, i.e. one after the other, before being coupled via the EVB method. This means that all the available processors are used to compute each force-field, in contrast to the alternative strategy of dividing processors for each force field. For extended systems, this choice is convenient given the relative high computational cost of the long range Coulombic part in comparison with all the other contributions to the configurational energy. This loop structure increases the computational time by a multiplicative factor of approximately N_F with respect to the required time to compute only a single force field.

5 Coupling terms

The quality of EVB method depends on the choice for the coupling terms C_{mk} , particularly to reproduce accurate interactions at the intermediate region between chemical states m and k where the change of chemistry occurs. Several sophisticated EVB coupling recipes have been proposed over the years.^{23,40–46,51,52} Despite their proven success, these recipes use complex internal (spatial) coordinates to couple the force fields. Here, however, we aim to use functional forms C_{mk} that depend on the energy gaps $\epsilon_{mk} = E_c^{(m)} - E_c^{(k)}$, because these variables not only constitute a possible generalized reaction coordinate^{22,24,34,36} but also allow

to compute the EVB stress tensor as described in Sec. 3, which is the main purpose of the present work.

The dependence of coupling terms C_{mk} on energy gaps has been previously investigated by B. Hartke *et al.*²² In their work, Density Functional Theory (DFT)^{53,54} was first used to compute the minimum energy path (MEP) between reactant and product states for the bond breaking-formation of several molecules in gas phase. Via highly accurate DFT-derived FFs^{55,56} for the involved chemical states m and k , the authors computed the coupling terms C_{mk} for selected configurations along the corresponding MEP from the individual energies $E_c^{(m)}$ and $E_c^{(k)}$ and the reference DFT energy. By plotting C_{mk} as a function of ϵ_{mk} , the data was fitted to constants and Gaussian-type of functions.

The implementation of such procedure necessarily requires the use of force-fields i) consistent with the level of theory that is used to compute the explicit electronic problem for the reaction and ii) accurate enough far from the reference geometry for which they were fitted. Ultimately, meeting these requirements is a non-trivial challenge, generally impossible in many cases, particularly for large systems. In addition, previous research claimed that for several reactions the resulting EVB energy leads to large errors along the MEP, especially in the transition state region where, artificial minima are created in the worst cases. To overcome these limitations, a combination of Gaussian functions were proposed to model the coupling terms.⁴⁵ This methodology offers a promising route for calibration and development of EVB potentials, and provides a suitable framework for future applications of the EVB extension proposed in this work.

6 Model study and MD settings

To test the implementation of EVB method and its extension to the NPT ensemble, we considered a single malonaldehyde molecule in water solution as a model system. Malonaldehyde (MA) is an archetypal example of intramolecular proton transfer between two oxygen atoms.

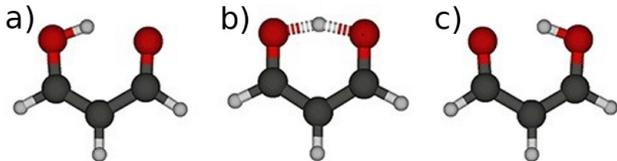


Figure 1: Proton transfer process for malonaldehyde between the two conformations (a and c) via the transition state (b). Oxygen (red), Carbon (dark grey) and hydrogen (light grey).

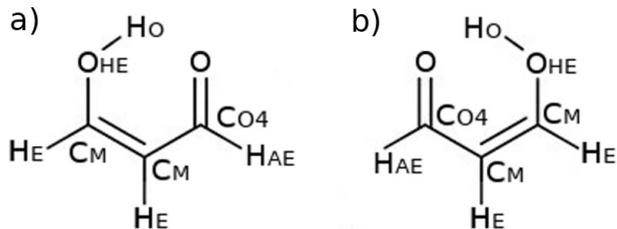


Figure 2: Schematic representation for the two conformations of MA, whose force fields are fitted to model integrations when the proton H_O is bonded to a) the oxygen at the left b) the oxygen at the right. Note the change of labelling for O, C and H atoms as well as the change of the double bonds. Atom labelling is consistent to the OPLS-2005 FF library.

Each conformation corresponds to a different chemical configurations for the same molecule, as shown in Fig. 1 a and c. In a classical description, the system swaps between both conformations only when vibrations promote the proton to overcome the energy barrier via the transition state (TS), as depicted in Fig. 1b.

A reactive FF for MA would aim to model the interatomic interactions for the whole domain with the forming and breaking of the O-H bonds. An example of such a FF was proposed by Y. Yang *et al.*⁵⁷ based on an extension of the molecular mechanics with proton transfer method⁵⁸ to non-linear hydrogen bonds. More recently, reactive force fields for MA have been derived using machine learning⁵⁹ and neural networks.⁶⁰ Here, nevertheless, we use different non-reactive FFs to describe interactions in the vicinity of each conformation, as schematically shown in Fig 2. The two FFs were generated with the DL_FIELD program⁶¹ in a format suitable to DL_POLY_4 using the OPLS-2005 FF library,^{62,63} which is not only specially designed for liquid simulations but also constitutes an example of non-reactive FF available in the literature. Atoms are labelled differently depending on the FF. For example, for the conformation of Fig 2a (FF₁ from now on), the proton H_O is chemically bonded to

the O_{HE} site and only interacts with oxygen O via van der Waals and coulombic interactions. For this FF_1 topology, the conformational energy would be rather large for geometries where the proton H_{O} is at the vicinity of the O site and the realistic chemistry would be better represented by interactions according to the topology of the FF of Fig 2b (FF_2 from now on). For this reason, if one only used FF_1 to described the interactions, atom H_{O} would unlikely explore the vicinity of the O site during the course of a MD simulation. The same reasoning applies to the complementary non-reactive FF_2 .

Each water molecule of the solvent was simulated with the TIP4P scheme,⁶⁴ which uses a four-site water model with an off-center point charge for oxygen. To maximize the effect of the EVB reactive potential on the solution, the number of the non-reactive water molecules has to be minimized. In MD simulations with FFs, this choice is restricted by the van der Waals cutoff radius, which is routinely set to 12 Å. Thus, for a cubic supercell with periodic boundary conditions, the minimum size of the box length should be 24 Å. To this purpose, models were built to contains 599 rigid water molecules arranged around the MA molecule within a cubic box of 27 Å, while using an initial separation criteria of 1.9 Å between the molecules. This amount of water prevented box length values below the limit of 24 Å in all the simulations. Such a model already represents an aqueous solution with a rather low concentration of 9.19×10^{-2} molality [mol(MA)/kg(H_2O)].

Using the initial arrangement of atoms, the system was initially computed in the NVT ensemble at 300K using only FF_1 for MA and a Nose-Hoover^{65,66} thermostat with a relaxation time of 0.5 ps. Equilibration was conducted for 5 ps, scaling the system temperature every 5 fs and resampling instantaneous system momenta distribution every 9 fs. Production MD followed for 30 ps. The last snapshot with positions, velocities and forces served as the starting point of a NPT simulation at 1 atm, this time using a Nose-Hoover thermostat and barostat with relaxation times of 3.0 and 1.0 ps for the thermostat and barostat, respectively. Equilibration was conducted for 2 ps while allowing for a variation of 10% in the system density. This was followed by a 3.0 ns of MD production run. The average supercell dimension

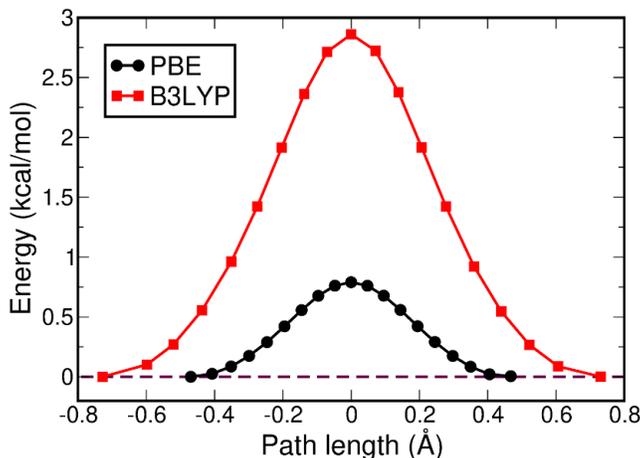


Figure 3: DFT energy profile along the computed MEPs for a single MA molecule in vacuum using the B3LYP (red) and the PBE (black) approximations for the electronic exchange and correlation.

was used for all the EVB-NVT simulations and as starting point for the EVB-NPT runs. All the MD simulations used a timestep of 1 fs, while the electrostatic interactions are computed through the smooth particle mesh Ewald method.^{39,50} Details of the EVB simulations are given in the next section.

7 Results and discussion

Proton transfer is a quantum mechanical process.⁶⁷⁻⁶⁹ Accounting for the full quantum problem of the nuclei, however, is computationally prohibited for sufficiently large systems, and several approximations have been developed over the years (See ref. 70 and references therein). In particular, the EVB method has been extended to its Multistate version (MS-EVB) to successfully capture the essential physics and chemistry in different protonated systems, both in the classical and quantum regime.^{52,71-78} Nevertheless, MS-EVB inherits the limitation of EVB with respect to the stress tensor and its application has only been restricted to NVE and NVT ensembles.

With regards to the model of a single MA in water, A. Yamada *et al.* has previously used the quantum-classical molecular dynamics method⁷⁹ to compute the quantum reaction dy-

namics.⁷⁰ On the other hand, Y. Yang *et al.*⁵⁷ assumed the whole solution as classical, and used MD to compute proton transfer rates in an effective potential for MA that included zero point energy effects.

The purpose of this work, however, is to compute the solvated MA using the EVB method in the NPT ensemble, and compare the results with the NVT-based protocol. Thus, nuclear quantum effects are neglected in the following simulations as well as zero point energy corrections, in line with a previous density-functional tight-binding QM/MM study.⁸⁰ Even though this represents an over simplification of the problem, the assumption of classical mechanics for the whole system has demonstrated to provide a reasonable framework to compute the lower limits for proton transfer.⁵⁷

We started our study by considering the explicit quantum electronic problem of a single MA in vacuum at zero temperature, and used the computed quantities as a reference to calibrate the coupling terms of the EVB matrix. By means of the Nudge Elastic Band (NEB) method^{81,82} combined with DFT calculations, we computed the minimum energy path (MEP) to transfer the proton between the two conformations. Details for these calculations are provided in section Settings for DFT simulations. Following the geometry relaxation of each MA conformation, the converged structures were used as fixed end-points of the MEP, which was built by using 17 intermediate images. Fig. 3 shows the computed energy profiles along the converged MEP for the B3LYP⁸³ and the PBE⁸⁴ electronic exchange and correlation (XC) functionals. Dispersive van der Waals interaction are included via the Grimme’s DFT-D3 formalism.⁸⁵ We compute energy barriers of 2.86 and 0.79 kcal/mol for B3LYP and PBE, respectively, which is in agreement with previous work⁸⁶ and corroborates the crucial dependence of the XC functional on the energy barrier and length of the MEP. These values underestimate previous coupled-cluster calculations, which predicted energy barriers between 4.0-4.3 kcal/mol.^{57,87-89} Despite this underestimation, we shall proceed with the computed DFT values, as achieving chemical accuracy is not the main purpose of the present work. Even though this choice represents a departure from a more realistic

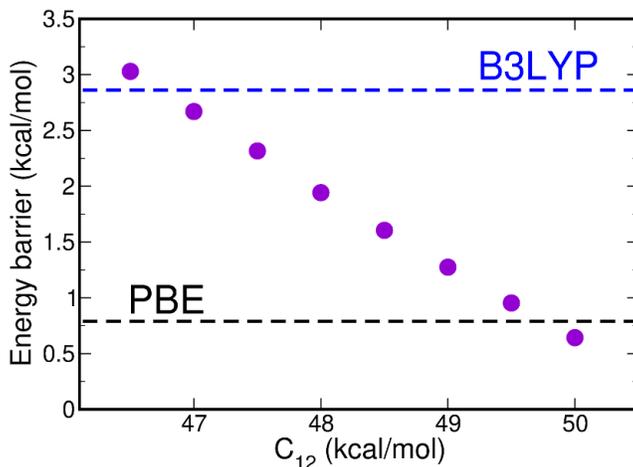


Figure 4: EVB barrier for proton transfer for a MA molecule in vacuum as a function of the coupling term C_{12} . Computed energy barriers obtained from the DFT calculations of Fig. 3 are shown as a reference.

chemistry, lower energy barriers are more convenient to MD, as less computational time is needed to switch between conformations, thus allowing a better sampling of the configurational space.

On the other hand, energy barriers computed with the EVB method depend on the term C_{12} . Here, we have assumed C_{12} to be a constant. Note that the choice of a constant for the coupling term, even trivial, complies with the functional form requirement for the coupling terms to compute the EVB stress tensor. Figure 4 shows that the computed EVB barrier for MA decreases as the value of C_{12} increases. For the adopted OPLS-2005-FFs and to the purpose of comparison with the DFT energy barriers of Fig. 3, we have only considered values of C_{12} in the range between 46.5 and 50.0 kcal/mol. These results constitute an example of how coupling terms can be used to calibrate EVB potentials against a reference value, in this case obtained from DFT. The advantage of the EVB formalism lies in the assumption that the coupling terms calibrated for reactive molecules in gas phase do not change significantly when transferring the reactive system from one phase to the another.²⁴ This approximation has been rigorously validated via Constrained-DFT calculation.³⁰ Moreover, the use of constants for the coupling terms is the most common choice in the execution of EVB simulations for solvated reactive sites, as a constant can be finely adjusted until a

calculated property (usually free energy) agrees with the experimental value.^{24,33} Here, we are not interested in comparing with experiments but evaluating how the results are affected by the use of different ensembles.

Clearly, the reactive EVB potential is different from any of the individual non-reactive FFs, particularly in the TS region. Thus, it is natural to argue as to which extent this reactive EVB potential affects the stress tensor and the converged volume (and density) of the solution, and how results compare with the density resulting from a NPT simulation using one of the involved FFs, as in the NVT-based protocol. If the energy barrier is sufficiently large, the system will only sample the vicinity of one of the possible configurations. Even though the system might occasionally swap conformation, the TS region will be hardly sampled. Thus, for cases where the conformations are chemically equivalent and the barrier is large enough, the NVT-based protocol appears to be a sensible approximation. In contrast, if the barrier is sufficiently low, the TS region will be better sampled during the course of a MD simulation and the average potential will depart from any of the individual FFs.

Following a NPT simulation of the solvated MA at 1 atm and 300 K using only one of the two FFs, the computed density of the system is predicted to be in the region between the horizontal red dashed lines, as shown in Fig. 5. From the set of C_{12} values considered in Fig. 4 the classical barrier for proton transfer can be artificially changed. Thus, the EVB potentials generated using these different C_{12} values can be considered as different reactive models for MA. For each of these reactive models, we run a full EVB-NPT simulation and compute the density and its uncertainty, indicated by the green filled circles with error bars. Results demonstrate that the converged density for the solvated MA is statistically independent of the energy barrier for intramolecular proton transfer. In addition, the fact that computed EVB-NPT values statistically fall within the boundary of the red dashed lines supports the validity of the NVT-based protocol in determining the size of the system, as least for the present test case of low concentrated aqueous MA.

To further investigate on the role of the ensemble, we use the converged volume from the

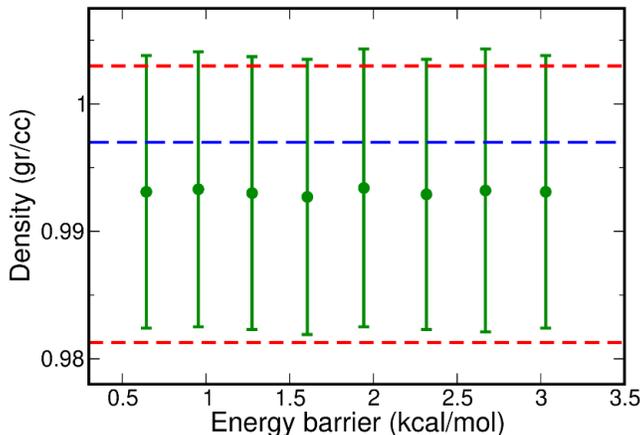


Figure 5: Computed density (green filled circles) of the model solution composed of one MA and 599 rigid water molecules. Different choices of the coupling C_{12} lead to different energy barriers for proton transfer of a single MA in vacuum (see Fig. 4). Each energy barrier (values in the x-axis) can be considered as a different reactive model for MA. The region between horizontal red dashed lines corresponds to the range of possible densities from a standard NPT simulation using only one of the non-reactive FFs. The horizontal blue line refers to the experimental density of pure water. Reference pressure and temperature are 1 atm and 300K, respectively.

first NPT simulation (with only one FF) and run EVB-NVT simulations for each reactive field. To compare EVB simulation for both ensembles, here we propose to use the energy gap ϵ_{12} , obtained from the energy difference between the FF_1 and FF_2 at each time step. The range of computed values for ϵ_{12} are grouped using a total of 150 bins, each bin with an energy window of 20 kcal/mol. To remove the dependence on the simulation time (i.e. number of configurations sampled), histograms are normalized such as the total area is equal to one. Such normalized distributions can be interpreted as the classical probability of finding the system at a given value of ϵ_{12} . Figure 6a shows the computed distribution following EVB-NVT and EVB-NPT simulations for the solvated model with an energy barrier for MA of 1.94 kcal/mol. Distributions exhibit two broad peaks centered at approximately -350 and 350 kcal/mol, which indicates the system mainly samples configurations in the vicinity of the conformations of Figure 2 and resembles the well-known probability distribution of a proton in a double well. The observed small asymmetry in the distributions is attributed to the finite time of MDs run and the lack of control for both configurations to be equally sampled.

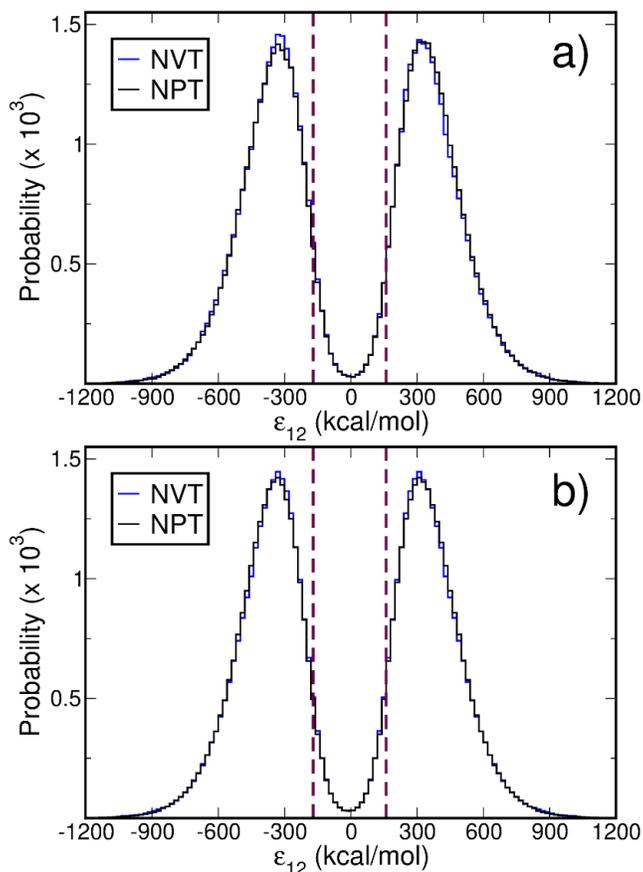


Figure 6: Normalized distribution for the ϵ_{12} energy gap following NVT and NPT simulations for the model of MA with an energy barrier of 1.94 kcal/mol a) after MD runs b) upon symmetrization. Note values in the y-axis are scaled for the sake of visualization. The energy domain within the brown vertical dashed lines is arbitrarily assigned to the transition state region (see text).

In fact, the reactive potential for MA in gas phase at zero temperature is symmetric along the MEP (Fig. 3), and the same is expected for MA in solution at 300 K despite the electrostatic field created by the surrounding water. Symmetric distributions, however, can only be achieved by increasing the sampling of the conformational phase space and, hence, the computational time for the MD runs, which is beyond the purpose of this work. Alternatively, we use the computed data and make these distributions to be symmetric around $\epsilon_{12} = 0$, as shown in Fig 6b. To the best of our knowledge, this is the first time such probability distributions are used within the framework of EVB as a tool to evaluate the sampling of the configurational space. Moreover, these computed distributions support previous research

that suggest the convenience of ϵ as an alternative reaction coordinate.^{22,24,34,36} In fact, in contrast to the MEP, ϵ represents a coordinate that implicitly accounts not only for the reactive site but also includes the effect of the surrounding solvent.

The symmetric distributions obtained for the NVT and NPT ensembles indicate a good level of agreement. Nevertheless, it is convenient to quantify this agreement for a better comparison. To this purpose, we first estimate the width for both peaks, which is of the order of 320 kcal/mol. This range for ϵ_{12} was used to define the TS domain, indicated by the region between the dashed lines of Fig. 6, located at $\epsilon_{12}^{TS} = \pm 160$ kcal/mol. This TS region is also adopted to be independent of the reactive model for MA. We define the probability for the the system to be in the TS region as follows

$$\text{TS Probability}(\epsilon_{12}^{TS}) = \int_{-\epsilon_{12}^{TS}}^{\epsilon_{12}^{TS}} \mathcal{P}(\epsilon_{12}) d\epsilon_{12} \quad (23)$$

where $\mathcal{P}(\epsilon_{12})$ is the normalized distribution obtained from the histograms, as shown in Fig. 6. Clearly, the computed TS probability depends on the choice for the extension of the TS region, which is completely arbitrary. However, Eq. (23) provides a method to quantify the relevance of the TS region and allow the comparison between different ensembles. Figure 7 shows the computed values for the TS Probabilities for the different reactive models of MA. As expected, the classical probability for the system to sample the TS region increases as the barrier reduces. Additionally, the computed probability distributions do not depend on the assumed ensemble for the present model.

Finally, it is important to evaluate how the presented results depend on the duration of the MD-EVB runs. To this purpose, we have computed mass densities and TS-Probabilities for different durations of the MD runs (0.3, 1.0, 2.0 and 3.0 ns) and report the results in section S1 and S2 of the supporting information (SI). Analysis of the computed data shows that mass densities are already converged after 0.3 ns of EVB-NPT simulation, within statistical uncertainty. In contrast, TS-Probabilities are not always converged after 0.3 ns and

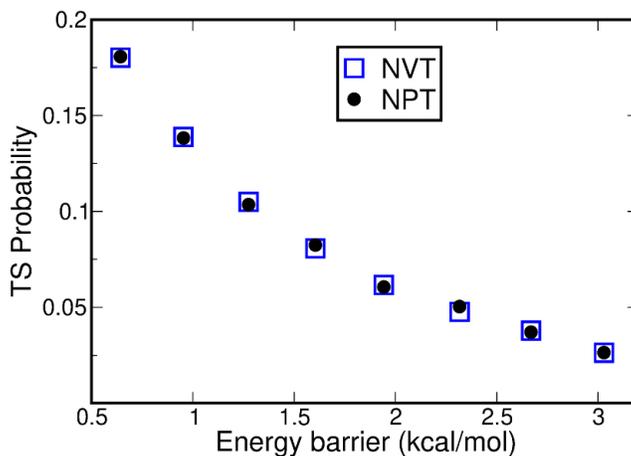


Figure 7: Probability to sample the TS for solvated MA in the NVT and NPT ensembles. Results are plotted as a function of the classical energy barrier for proton transfer in MA.

convergence depends on the reactive model for MD and the assumed ensemble. However, we observe that the computed difference for the NVT and NPT ensembles generally become smaller as the simulation time is increased. To further corroborate these findings, we have computed the reactive model of Fig. 6 up to 10.0 ns, and observe that TS-Probabilities are already converged after 3.0 ns (Fig. S3 of the SI).

Based on these findings, we conclude that the NVT-based protocol to compute single reactive sites in water solution with EVB using the NVT ensemble is already a remarkably good approximation to the self-consistent EVB-NPT simulation, at least for concentrations of reactive sites lower than 9.19×10^{-2} molality. In the present case study of solvated MA, the flexible-reactive MA is composed of nine atoms and contributes with 27 degrees of freedom (d.o.f), whereas the 599 rigid water molecules can only undergo translation and rotation, contributing another 3594 d.o.f. Thus, the rather small ratio 27:3594 and the rather low compressibility of water might explain why the reactive degrees of freedom play a negligible role in the dynamics of the whole system.

In future studies, it would be interesting to investigate systems with higher concentrations for the reactive MA. As discussed in section 6, due to restriction on the size of the simulation cell, higher concentration of MA could only be achieved here by replacing water molecules with MA molecules. However, the standard EVB method is designed for only one reactive

component, either in gas phase or interacting with a surrounding non-reactive environment. Thus, increasing the concentration of reactive components would necessarily require an extension of the EVB method to account for multiple reactive sites in a single simulation. This new capability would be ideal not only to the prospect of computing higher concentrations in solutions, but also to perform EVB simulations of molecular crystals composed of reactive units, as for the family of Ketohydrazone-Azoenol systems.^{90,91} Such a development is part of current research in our group.

8 Concluding remarks

In this work we propose a new formalism to derive the stress tensor within the EVB method, thus allowing EVB simulations of flexible molecular reactive sites in condensed phase at constant pressure and temperature for the first time. This formalism is based on the use of energy gaps as reactive coordinates to parameterize the coupling terms of the EVB matrix. As a test case, we considered the intramolecular proton transfer in MA molecule solvated in water by means of molecular dynamics, while neglecting the role of quantum nuclei and zero point energy corrections.

In comparison to the NVT-based protocol of converging the system volume using only one of the non-reactive FFs, results from EVB-NPT simulations for different reactive models of MA (i.e. different energy barriers) demonstrate a negligible effect of the EVB potential in the computed density of the solution.

In addition, we performed EVB-NVT simulations using the converged volume from the NVT-based protocol. To compare the sampling of the configurational space with respect to EVB-NPT simulations, we use the energy gap as a variable to compute probability distributions of the reactive system. Detailed analysis of these distributions also demonstrate a negligible difference between both ensembles. We attribute these findings to the relative low concentration for the model of MA in water, where the non-reactive dynamics of the rigid

water molecules dominates over the reactive dynamics of the single MA. Therefore, future EVB simulations of solvated reactive molecules with higher concentrations would be beneficial to quantify the role of the reactive dynamics on the whole system. To this purpose, developments to extend the EVB method are required to accommodate multiple reactive sites at the same time.

Settings for DFT simulations

Geometry optimization and NEB calculations for the two conformations of MA were conducted using the library DL_FIND⁹² of the ChemShell program⁹³ via its interface⁹⁴ with the ORCA package⁹⁵ for the DFT computation of energies and forces. Both PBE and B3LYP functionals were used together with DFT-D3 dispersion correction.⁸⁵ The def2-TZVP basis set⁹⁶ was used for all the atoms.

Author contributions

I.S. and A. M. E. designed the implementation of the EVB in DL_POLY_4; I.T. has lead DL_POLY_4 and participated in initial stages of the EVB project; A. M. E. led the refactoring of DL_POLY_4; I.S. developed the mathematical formalism for the EVB stress tensor, implemented the required changes within DL_POLY_4, built the models and performed the MD simulations; A. M. E. provided technical support for the EVB implementation; K. S. prepared the input files for the ChemShell simulations; I.S. wrote the manuscript; A. M. E., K. S. and I. T. revised the manuscript.

Acknowledgement

This work made use of computational support by CoSeC, the Computational Science Centre for Research Communities, through CCP5: The Computer Simulation of Condensed Phases,

EPSRC grant no EP/M022617/1. I. S. acknowledges i) fruitful discussions with Vlad Sokhan, Silvia Chiacchiera, Fausto Martelli, Alfredo Caro and Gilberto Teobaldi, ii) Chin Yong for his help with the DL_FIELD code, iii) John Purton for his initial contributions to the EVB funding project iv) Jim Madge for his heroic contribution to the refactoring of DL_POLY_4. A. M. E. acknowledges support of EPSRC via grant EP/P022308/1.

Supporting Information Available

The Supporting Information is available free of charge.

- supporting-info.pdf: Computed densities vs. simulation time (section S1). TS probability vs. simulation time (section S2).

A On the decomposition of the EVB energy

Similarly to Eq. (1) of the main manuscript, it is interesting to investigate if E_{EVB} can be decomposed in different energy contribution. To this purpose we express the diagonal terms of H_{EVB}^{mk} as a sum of the individual contributions

$$H_{\text{EVB}}^{mk} = \begin{cases} \sum_{\text{type}} U_{\text{type}}^{(m)} & m = k \\ C_{mk}(\epsilon_{mk}) & m \neq k \end{cases} \quad (\text{A1})$$

where the index type runs over all type of possible interactions (bonds, angles, coulombic, etc). In contrast, $C_{mk}(\epsilon_{mk})$ cannot be decomposed in terms of the individual contributions $U_{\text{type}}^{(m)}$. Consequently, matrix H_{EVB} cannot be decomposed for each type of interaction. One might consider the particular case of constant coupling terms $C_{mk}(\epsilon_{mk}) = C_{mk}$, $\forall m, k = 1, \dots, N_F$, with $m \neq k$ to check if a separation into individual terms is possible. For the sake of simplicity, let us consider the case of two FFs with $C_{12} = C_{21}$. Without loss of

generality, we can write \mathcal{C}_{12} as a sum of a set of constants

$$\mathcal{C}_{12} = \sum_{\text{type}} \mathcal{C}_{12}^{\text{type}} \quad (\text{A2})$$

and

$$H_{\text{EVB}} = \sum_{\text{type}} H_{\text{EVB}}^{\text{type}}, \quad \text{with} \quad H_{\text{EVB}}^{\text{type}} = \begin{pmatrix} U_{\text{type}}^{(1)} & \mathcal{C}_{12}^{\text{type}} \\ \mathcal{C}_{12}^{\text{type}} & U_{\text{type}}^{(2)} \end{pmatrix} \quad (\text{A3})$$

Using the computed EVB eigenvector, Ψ_{EVB} , from diagonalization of the H_{EVB} matrix we have

$$E_{\text{EVB}} = \sum_{\text{type}} E_{\text{EVB}}^{\text{type}} \quad (\text{A4})$$

where

$$E_{\text{EVB}}^{\text{type}} = \langle \Psi_{\text{EVB}} | \hat{H}_{\text{EVB}}^{\text{type}} | \Psi_{\text{EVB}} \rangle \quad (\text{A5})$$

which, in principle, offers a possible way to decompose the EVB energy in terms of individual types of interactions. However, such a decomposition is not unequivocally defined, as there are infinite ways of writing the sum for \mathcal{C}_{12} in Eq. (A2). This demonstrates that an EVB energy decomposition in individual terms as in Eq. (1) is not well defined. In fact, only E_{EVB} is well defined.

B On the EVB stress tensor and virial decomposition

The purpose of this section is to demonstrate that the EVB stress tensor $\sigma_{\alpha\beta}^{\text{EVB}}$ and virial \mathcal{V}_{EVB} can be decomposed in different components according to the type of the interaction. We note that $\sigma_{\alpha\gamma}^{c(m)}$ in Eq. (19) can be decomposed in different contributions, namely $\sigma_{\alpha\gamma}^{c(m)} = \sum_{\text{type}} \sigma_{\alpha\gamma}^{\text{type}(m)}$. Thus,

$$\frac{\partial H_{\text{EVB}}^{mk}}{\partial h_{\alpha\beta}} = \sum_{\text{type}} \frac{\partial H_{\text{type}}^{mk}}{\partial h_{\alpha\beta}} \quad \text{where} \quad (\text{B1})$$

$$\frac{\partial H_{\text{type}}^{mk}}{\partial h_{\alpha\beta}} = \begin{cases} -V \sum_{\gamma} \sigma_{\alpha\gamma}^{\text{type}(m)} h_{\beta\gamma}^{-1} & m = k \\ -VC'_{mk} \sum_{\gamma} [\sigma_{\alpha\gamma}^{\text{type}(m)} - \sigma_{\alpha\gamma}^{\text{type}(k)}] h_{\beta\gamma}^{-1} & m \neq k \end{cases}$$

from which, similarly to Eq. (20), we have

$$\sigma_{\alpha\beta}^{\text{type}} = -\frac{1}{V} \sum_{\gamma=x,y,z} h_{\beta\gamma} \sum_{m,k=1}^{N_F} \tilde{\Psi}_{\text{EVB}}^{(m)} \left(\frac{\partial H_{\text{type}}^{mk}}{\partial h_{\alpha\beta}} \right) \Psi_{\text{EVB}}^{(k)} \quad (\text{B2})$$

and

$$\sigma_{\alpha\beta}^{\text{EVB}} = \sum_{\text{type}} \sigma_{\alpha\beta}^{\text{type}} \quad (\text{B3})$$

with the following decomposition for the virial

$$\mathcal{V}_{\text{EVB}} = \sum_{\text{type}} \mathcal{V}_{\text{EVB}}^{\text{type}} = - \sum_{\text{type}} \sum_{\alpha=x,y,z} \sigma_{\alpha\alpha}^{\text{type}}. \quad (\text{B4})$$

References

- (1) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **2002**, *9*, 646–652.
- (2) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.
- (3) Grajciar, L.; Heard, C. J.; Bondarenko, A. A.; Polynski, M. V.; Meeprasert, J.;

- Pidko, E. A.; Nachtigall, P. Towards operando computational modeling in heterogeneous catalysis. *Chem. Soc. Rev.* **2018**, *47*, 8307–8348.
- (4) González, M.A., Force fields and molecular dynamics simulations. *JDN* **2011**, *12*, 169–200.
- (5) Ballone, P. Modeling Potential Energy Surfaces: From First-Principle Approaches to Empirical Force Fields. *Entropy* **2014**, *16*, 322–349.
- (6) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A* **2001**, *105*, 9396–9409.
- (7) Farah, K.; Müller-Plathe, F.; Böhm, M. C. Classical Reactive Molecular Dynamics Implementations: State of the Art. *ChemPhysChem* **2012**, *13*, 1127–1151.
- (8) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. T. The ReaxFF reactive force-field: development, applications and future directions. *Npj Computational Materials* **2016**, *2*, 15011 EP –, Review Article.
- (9) Yun, K.-S.; Pai, S. J.; Yeo, B. C.; Lee, K.-R.; Kim, S.-J.; Han, S. S. Simulation Protocol for Prediction of a Solid-Electrolyte Interphase on the Silicon-based Anodes of a Lithium-Ion Battery: ReaxFF Reactive Force Field. *The Journal of Physical Chemistry Letters* **2017**, *8*, 2812–2818.
- (10) Islam, M. M.; Kolesov, G.; Verstraelen, T.; Kaxiras, E.; van Duin, A. C. T. eReaxFF: A Pseudoclassical Treatment of Explicit Electrons within Reactive Force Field Simulations. *Journal of Chemical Theory and Computation* **2016**, *12*, 3463–3472.
- (11) Liang, T.; Shin, Y. K.; Cheng, Y.-T.; Yilmaz, D. E.; Vishnu, K. G.; Verners, O.; Zou, C.;

- Phillpot, S. R.; Sinnott, S. B.; van Duin, A. C. Reactive Potentials for Advanced Atomistic Simulations. *Annual Review of Materials Research* **2013**, *43*, 109–129.
- (12) Pahari, P.; Chaturvedi, S. Determination of best-fit potential parameters for a reactive force field using a genetic algorithm. *Journal of Molecular Modeling* **2012**, *18*, 1049–1061.
- (13) Larsson, H. R.; van Duin, A. C. T.; Hartke, B. Global optimization of parameters in the reactive force field ReaxFF for SiOH. *Journal of Computational Chemistry* **2013**, *34*, 2178–2189.
- (14) Li, Y.; Hartke, B. Approximate photochemical dynamics of azobenzene with reactive force fields. *The Journal of Chemical Physics* **2013**, *139*, 224303.
- (15) Jaramillo-Botero, A.; Naserifar, S.; Goddard, W. A. General Multiobjective Force Field Optimization Framework, with Application to Reactive Force Fields for Silicon Carbide. *Journal of Chemical Theory and Computation* **2014**, *10*, 1426–1439.
- (16) Larentzos, J. P.; Rice, B. M.; Byrd, E. F. C.; Weingarten, N. S.; Lill, J. V. Parameterizing Complex Reactive Force Fields Using Multiple Objective Evolutionary Strategies (MOES). Part 1: ReaxFF Models for Cyclotrimethylene Trinitramine (RDX) and 1,1-Diamino-2,2-dinitroethene (FOX-7). *Journal of Chemical Theory and Computation* **2015**, *11*, 381–391.
- (17) Rice, B. M.; Larentzos, J. P.; Byrd, E. F. C.; Weingarten, N. S. Parameterizing Complex Reactive Force Fields Using Multiple Objective Evolutionary Strategies (MOES): Part 2: Transferability of ReaxFF Models to C-H-N-O Energetic Materials. *Journal of Chemical Theory and Computation* **2015**, *11*, 392–405.
- (18) Dittner, M.; Müller, J.; Aktulga, H. M.; Hartke, B. Efficient global optimization of reactive force-field parameters. *Journal of Computational Chemistry* **2015**, *36*, 1550–1561.

- (19) Shin, Y. K.; Kwak, H.; Zou, C.; Vasenkov, A. V.; van Duin, A. C. T. Development and Validation of a ReaxFF Reactive Force Field for Fe/Al/Ni Alloys: Molecular Dynamics Study of Elastic Constants, Diffusion, and Segregation. *The Journal of Physical Chemistry A* **2012**, *116*, 12163–12174.
- (20) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society* **1980**, *102*, 6218–6226.
- (21) Åqvist, J.; Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chemical Reviews* **1993**, *93*, 2523–2544.
- (22) Hartke, B.; Grimme, S. Reactive force fields made simple. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16715–16718.
- (23) Carpenter, B. K.; Harvey, J. N.; Glowacki, D. R. Prediction of enhanced solvent-induced enantioselectivity for a ring opening with a bifurcating reaction path. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8372–8381.
- (24) Duarte, F.; Kamerlin, S. C. L. *Theory and Applications of the Empirical Valence Bond Approach*; Wiley, 2017; Chapter 2, pp 27–61.
- (25) Zhang, X.; Harvey, J. N. EVB and polarizable MM study of energy relaxation in fluorine–acetonitrile reactions. *Phys. Chem. Chem. Phys.* **2019**, *21*, 14331–14340.
- (26) Warshel, A. MOLARIS-XG. Version 9.15. https://laetro.usc.edu/doc/theory_molaris_9.15.pdf **2015**,
- (27) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J Mol Graph Model.* **1998**, *16*, 213–261.

- (28) Yamashita, T.; Peng, Y.; Knight, C.; Voth, G. A. Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *Journal of Chemical Theory and Computation* **2012**, *8*, 4863–4875, PMID: 25100924.
- (29) Case, D.; Betz, R.; Cerutti, D.; et. al., AMBER 2016. University of California, San Francisco. **2016**,
- (30) Hong, G.; Rosta, E.; Warshel, A. Using the Constrained DFT Approach in Generating Diabatic Surfaces and Off Diagonal Empirical Valence Bond Terms for Modeling Reactions in Condensed Phases. *The Journal of Physical Chemistry B* **2006**, *110*, 19570–19574, PMID: 17004821.
- (31) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **2004**, *384*, 288 – 294.
- (32) Raiteri, P.; Gale, J. D.; Bussi, G. Reactive force field simulation of proton diffusion in BaZrO₃ using an empirical valence bond approach. *Journal of Physics: Condensed Matter* **2011**, *23*, 334213.
- (33) Duboue-Dijon, E.; Pluharova, E.; Domin, D.; Sen, K.; Fogarty, A. C.; Cheron, N.; Laage, D. Coupled Valence-Bond State Molecular Dynamics Description of an Enzyme-Catalyzed Reaction in a Non-Aqueous Organic Solvent. *The Journal of Physical Chemistry B* **2017**, *121*, 7027–7041, PMID: 28675789.
- (34) Mones, L.; Kulhánek, P.; Simon, I.; Laio, A.; Fuxreiter, M. The Energy Gap as a Universal Reaction Coordinate for the Simulation of Chemical Reactions. *The Journal of Physical Chemistry B* **2009**, *113*, 7867–7873, PMID: 19432459.
- (35) Todorov, I. T.; Smith, W. *The DL-POLY-4 User Manual, version 4.09* **2018**,
- (36) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; 1991.

- (37) Feynman, R. P. Forces in Molecules. *Phys. Rev.* **1939**, *56*, 340–343.
- (38) Smith, W. *CCP5 Info. Quart.* **1987**, *26*.
- (39) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.
- (40) Chang, Y. T.; Miller, W. H. An empirical valence bond model for constructing global potential energy surfaces for chemical reactions of polyatomic molecular systems. *The Journal of Physical Chemistry* **1990**, *94*, 5884–5888.
- (41) Kim, Y.; Corchado, J. C.; Villá, J.; Xing, J.; Truhlar, D. G. Multiconfiguration molecular mechanics algorithm for potential energy surfaces of chemical reactions. *The Journal of Chemical Physics* **2000**, *112*, 2718–2735.
- (42) Schlegel, H. B.; Sonnenberg, J. L. Empirical Valence-Bond Models for Reactive Potential Energy Surfaces Using Distributed Gaussians. *Journal of Chemical Theory and Computation* **2006**, *2*, 905–911.
- (43) Sonnenberg, J. L.; Schlegel, H. B. Empirical valence bond models for reactive potential energy surfaces. II. Intramolecular proton transfer in pyridone and the Claisen reaction of allyl vinyl ether. *Molecular Physics* **2007**, *105*, 2719–2729.
- (44) Sonnenberg, J. L.; Wong, K. F.; Voth, G. A.; Schlegel, H. B. Distributed Gaussian Valence Bond Surface Derived from Ab Initio Calculations. *Journal of Chemical Theory and Computation* **2009**, *5*, 949–961.
- (45) Steffen, J.; Hartke, B. Cheap but accurate calculation of chemical reaction rate constants from ab initio data, via system-specific, black-box force fields. *The Journal of Chemical Physics* **2017**, *147*, 161701.

- (46) Steffen, J. A new class of reaction path based potential energy surfaces enabling accurate black box chemical rate constant calculations. *The Journal of Chemical Physics* **2019**, *150*, 154105.
- (47) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*, 327 – 341.
- (48) Andersen, H. C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **1983**, *52*, 24 – 34.
- (49) Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T. DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. *J. Mater. Chem.* **2006**, *16*, 1911–1918.
- (50) Bush, I.; Todorov, I.; Smith, W. A DAFT DL_POLY distributed memory adaptation of the Smoothed Particle Mesh Ewald method. *Computer Physics Communications* **2006**, *175*, 323 – 329.
- (51) Glowacki, D. R.; Rose, R. A.; Greaves, S. J.; Orr-Ewing, A. J.; Harvey, J. N. Ultra-fast energy flow in the wake of solution-phase bimolecular reactions. *Nature Chemistry* **2011**, *3*, 850–855.
- (52) Glowacki, D. R.; Orr-Ewing, A. J.; Harvey, J. N. Non-equilibrium reaction and relaxation dynamics in a strongly interacting explicit solvent: F + CD₃CN treated with a parallel multi-state EVB model. *The Journal of Chemical Physics* **2015**, *143*, 044120.
- (53) Hohenberg, P. C.; Kohn, W. Inhomogeneous electron gas. *Physical Review* **1964**, *136*, 864–871.
- (54) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133–A1138.

- (55) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDF) for Molecules and Condensed Phase Simulations. *Journal of Chemical Theory and Computation* **2014**, *10*, 4497–4514.
- (56) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A general intermolecular force field based on tight-binding quantum chemical calculations. *The Journal of Chemical Physics* **2017**, *147*, 161708.
- (57) Yang, Y.; Meuwly, M. A generalized reactive force field for nonlinear hydrogen bonds: Hydrogen dynamics and transfer in malonaldehyde. *The Journal of Chemical Physics* **2010**, *133*, 064503.
- (58) Lammers, S.; Lutz, S.; Meuwly, M. Reactive force fields for proton transfer dynamics. *Journal of Computational Chemistry* **2008**, *29*, 1048–1063.
- (59) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nature Communications* **2017**, *8*, 872.
- (60) Unke, O. T.; Meuwly, M. A reactive, scalable, and transferable model for molecular energies from a neural network approach based on local information. *The Journal of Chemical Physics* **2018**, *148*, 241708.
- (61) Yong, C. W. Descriptions and Implementations of DL_F Notation: A Natural Chemical Expression System of Atom Types for Molecular Simulations. *Journal of Chemical Information and Modeling* **2016**, *56*, 1405–1409, PMID: 27455451.
- (62) Banks, J. L. et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of Computational Chemistry* **2005**, *26*, 1752–1780.
- (63) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the

- OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.
- (64) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.
- (65) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.
- (66) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (67) Kiefer, P. M.; Hynes, J. T. Adiabatic and nonadiabatic proton transfer rate constants in solution. *Solid State Ionics* **2004**, *168*, 219 – 224, Proceedings of the Workshop on Hydrogen: Ionic, Atomic and Molecular Motion.
- (68) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The nature of the hydrated excess proton in water. *Nature* **1999**, *397*, 601–604.
- (69) Tuckerman, M. E.; Marx, D.; Parrinello, M. The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature* **2002**, *417*, 925–929.
- (70) Yamada, A.; Kojima, H.; Okazaki, S. A molecular dynamics study of intramolecular proton transfer reaction of malonaldehyde in solutions based upon mixed quantum-classical approximation. I. Proton transfer reaction in water. *The Journal of Chemical Physics* **2014**, *141*, 084509.
- (71) Schmitt, U. W.; Voth, G. A. Multistate Empirical Valence Bond Model for Proton Transport in Water. *The Journal of Physical Chemistry B* **1998**, *102*, 5547–5551.
- (72) Schmitt, U. W.; Voth, G. A. The computer simulation of proton transport in water. *The Journal of Chemical Physics* **1999**, *111*, 9361–9381.

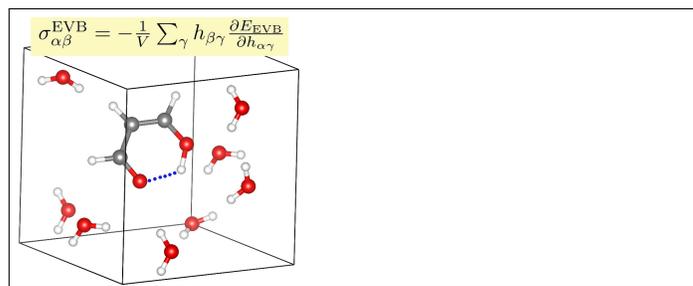
- (73) Day, T. J. F.; Soudackov, A. V.; Čuma, M.; Schmitt, U. W.; Voth, G. A. A second generation multistate empirical valence bond model for proton transport in aqueous systems. *The Journal of Chemical Physics* **2002**, *117*, 5839–5849.
- (74) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport. *The Journal of Physical Chemistry B* **2008**, *112*, 467–482, PMID: 17999484.
- (75) Swanson, J. M. J.; Maupin, C. M.; Chen, H.; Petersen, M. K.; Xu, J.; Wu, Y.; Voth, G. A. Proton Solvation and Transport in Aqueous and Biomolecular Systems: Insights from Computer Simulations. *The Journal of Physical Chemistry B* **2007**, *111*, 4300–4314, PMID: 17429993.
- (76) Wick, C. D.; Dang, L. X. Investigating Hydroxide Anion Interfacial Activity by Classical and Multistate Empirical Valence Bond Molecular Dynamics Simulations. *The Journal of Physical Chemistry A* **2009**, *113*, 6356–6364, PMID: 19391589.
- (77) Park, K.; Lin, W.; Paesani, F. A Refined MS-EVB Model for Proton Transport in Aqueous Environments. *The Journal of Physical Chemistry B* **2012**, *116*, 343–352, PMID: 22107267.
- (78) Biswas, R.; Tse, Y.-L. S.; Tokmakoff, A.; Voth, G. A. Role of Presolvation and Anharmonicity in Aqueous Phase Hydrated Proton Solvation and Transport. *The Journal of Physical Chemistry B* **2016**, *120*, 1793–1804, PMID: 26575795.
- (79) Yamada, A.; Okazaki, S. A quantum equation of motion for chemical reaction systems on an adiabatic double-well potential surface in solution based on the framework of mixed quantum-classical molecular dynamics. *The Journal of Chemical Physics* **2008**, *128*, 044507.
- (80) Walewski, L.; Bala, P.; Elstner, M.; Frauenheim, T.; Lesyng, B. Fast QM/MM method

- and its application to molecular systems. *Chemical Physics Letters* **2004**, *397*, 451 – 458.
- (81) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics* **2000**, *113*, 9901–9904.
- (82) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of Chemical Physics* **2000**, *113*, 9978–9985.
- (83) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- (84) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (85) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.
- (86) Sadhukhan, S.; Muñoz, D.; Adamo, C.; Scuseria, G. E. Predicting proton transfer barriers with density functional methods. *Chemical Physics Letters* **1999**, *306*, 83 – 87.
- (87) Schröder, M.; Gatti, F.; Meyer, H.-D. Theoretical studies of the tunneling splitting of malonaldehyde using the multiconfiguration time-dependent Hartree approach. *The Journal of Chemical Physics* **2011**, *134*, 234307.
- (88) Hammer, T.; Manthe, U. Intramolecular proton transfer in malonaldehyde: Accurate multilayer multi-configurational time-dependent Hartree calculations. *The Journal of Chemical Physics* **2011**, *134*, 224305.

- (89) Wang, Y.; Braams, B. J.; Bowman, J. M.; Carter, S.; Tew, D. P. Full-dimensional quantum calculations of ground-state tunneling splitting of malonaldehyde using an accurate ab initio potential energy surface. *The Journal of Chemical Physics* **2008**, *128*, 224314.
- (90) Gilli, P.; Bertolasi, V.; Pretto, L.; Lyčka, A.; Gilli, G. The Nature of Solid-State N–H...O/OH...N Tautomeric Competition in Resonant Systems. Intramolecular Proton Transfer in Low-Barrier Hydrogen Bonds Formed by the ...OC–CN–NH ... \rightleftharpoons ...HO–CC–NN...Ketohydrazone–Azoenol System. A Variable-Temperature X-ray Crystallographic and DFT Computational Study. *Journal of the American Chemical Society* **2002**, *124*, 13554–13567.
- (91) Durlak, P.; Latajka, Z. Car–Parrinello and Path Integral Molecular Dynamics Study of the Proton Transfer in the Intramolecular Hydrogen Bonds in the Ketohydrazone–Azoenol System. *The Journal of Physical Chemistry B* **2018**, *122*, 7862–7873.
- (92) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *The Journal of Physical Chemistry A* **2009**, *113*, 11856–11865, PMID: 19639948.
- (93) et. al., P. S. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *Journal of Molecular Structure: THEOCHEM* **2003**, *632*, 1 – 28.
- (94) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—a modular software package for QM/MM simulations. *WIREs Computational Molecular Science* **2014**, *4*, 101–110.
- (95) Neese, F. The ORCA program system. *WIREs Computational Molecular Science* **2012**, *2*, 73–78.

- (96) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

Graphical TOC Entry



Supporting Information for manuscript
"Reactive molecular dynamics at constant
pressure via non-reactive force fields:
extending the Empirical Valence Bond
method to the isothermal-isobaric ensemble"

Ivan Scivetti,^{*,†,‡} Kakali Sen,[†] Alin M. Elena,[†] and Ilian Todorov[†]

[†]*Daresbury Laboratory, Sc. Tech. Keckwick Lane, Daresbury, WA4 4AD, Warrington, UK*

[‡]*Stephenson Institute for Renewable Energy, Department of Chemistry, University of
Liverpool, Liverpool L69 3BX, England*

E-mail: ivan.scivetti@stfc.ac.uk

S1-Computed densities vs. simulation time

Figure S1 shows how the computed densities for the solution depend on the total simulation time of the MD-EVB run. Each reactive model for the MA molecule is identified by the classical energy barrier for proton transfer in vacuum (x-axis). Computed error values (not shown in the figure) are 0.011 gr/cc on average, which represents four times the largest deviation between the computed values after 0.3 ns. Thus, computed densities after 0.3 ns can be considered as converged. Clearly, results are well converged after 1.0 ns.

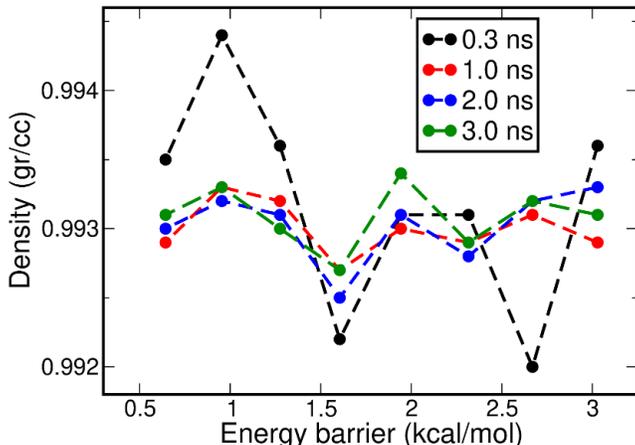


Figure S1: Computed densities of the model solution composed of one MA and 599 rigid water molecules. Each reactive model for MA is identified by the energy barrier for proton transfer in vacuum (values in the x-axis). Results are reported for simulation times of 0.3ns, 1.0ns, 2.0ns and 3.0ns. Error values (not shown) are 0.011 gr/cc on average. Reference pressure and temperature are 1 atm and 300K, respectively.

S2-TS probability vs. simulation time

Figure S2 shows how the computed transition state (TS) probabilities for the different reactive models of MA depend on the total time for the MD simulation. In contrast to the computed density of Fig. S1, TS-Probability values are not always converged after 0.3 ns. In addition, differences between NVT and NPT ensembles become closer as the simulation time is increased, with the largest computed difference of 2.3% after 3.0 ns for model of panel e).

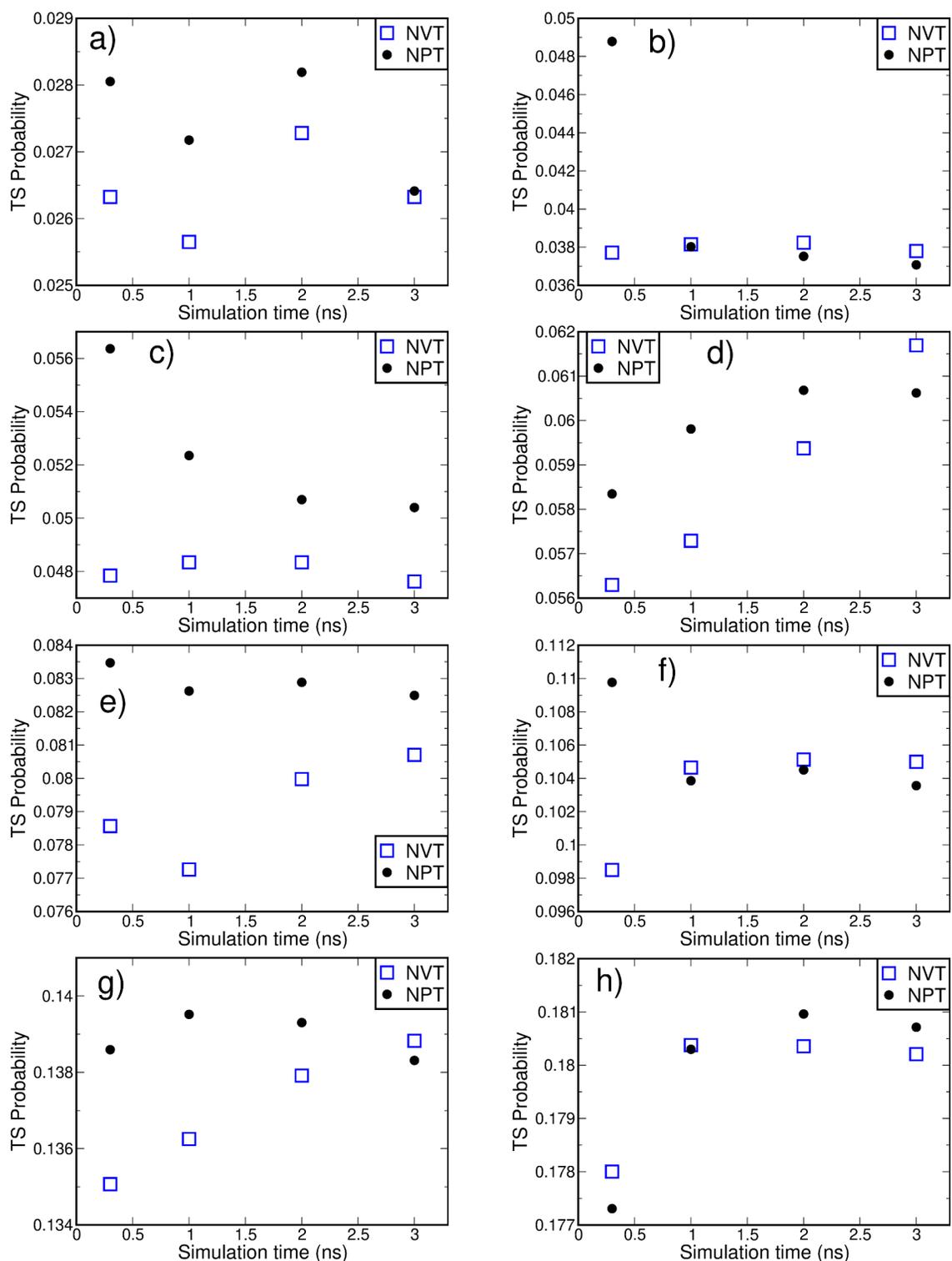


Figure S2: Probability to sample the TS for solvated MA in the NVT and NPT ensembles. Results are plotted as a function of the total time for the MD simulation. Each panel corresponds to different reactive models for MA in vacuum, each model characterised by the energy barrier for proton transfer a) 3.03, b) 2.67, c) 2.32, d) 1.94, e) 1.61, f) 1.28, g) 0.95 and h) 0.64 kcal/mol.

To further evaluate the role of the simulation time, we considered the reactive model of Fig. 6 of the main manuscript, which is the same as the model of panel d) of the Fig. S2 of this Supporting Information (1.94 kcal/mol of energy barrier of proton transfer), and carried out MD-EVB simulations up to 10.0 ns, both in the NVT and NPT ensembles. Results not only indicate that 3.0 ns is already sufficient to converge the TS-Probability but also demonstrate that computed values are independent of the assumed ensemble within numerical accuracy, at least in the limit of low-concentration solution as in the present case.

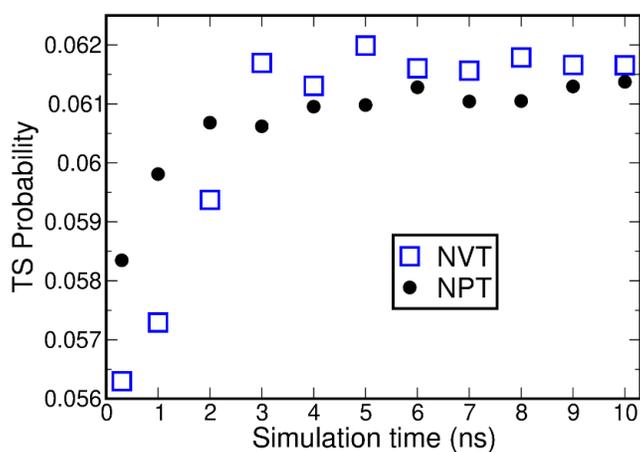


Figure S3: Probability to sample the TS for solvated MA in the NVT and NPT ensembles. Results are plotted as a function MD simulation time.