Universal tight binding model for chemical reactions in solution and at surfaces. I. Organic molecules

Published in:
Journal of Chemical Physics

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

T. J. Sheppard, A. Y. Lozovoi, D. L. Pashov, J. J. Kohanoff, and A. T. Paxton

---

Articles you may be interested in

Universal tight binding model for chemical reactions in solution and at surfaces. III. Stoichiometric and reduced surfaces of titania and the adsorption of water
J. Chem. Phys. **141**, 044505 (2014); 10.1063/1.4890492

Universal tight binding model for chemical reactions in solution and at surfaces. II. Water
J. Chem. Phys. **141**, 044504 (2014); 10.1063/1.4890343

Ab initio calculation of proton-coupled electron transfer rates using the external-potential representation: A ubiquinol complex in solution
J. Chem. Phys. **126**, 224514 (2007); 10.1063/1.2737048

Modeling vibrational spectra using the self-consistent charge density-functional tight-binding method. I. Raman spectra
J. Chem. Phys. **121**, 5171 (2004); 10.1063/1.1775787

Surface-hopping modeling of photoinduced relaxation dynamics on coupled potential-energy surfaces
J. Chem. Phys. **107**, 6230 (1997); 10.1063/1.474288

Universal tight binding model for chemical reactions in solution and at surfaces. I. Organic molecules

T. J. Sheppard,[1] A. Y. Lozovoi,[1] D. L. Pashov,[2] J. J. Kohanoff,[1] and A. T. Paxton[2,a)]

[1]*Atomistic Simulation Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, United Kingdom*
[2]*Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom*

As is now well established, a *first order* expansion of the Hohenberg–Kohn total energy density functional about a trial input density, namely, the Harris–Foulkes functional, can be used to rationalize a non self consistent tight binding model. If the expansion is taken to *second order* then the energy and electron density matrix need to be calculated self consistently and from this functional one can derive a charge self consistent tight binding theory. In this paper we have used this to describe a polarizable ion tight binding model which has the benefit of treating charge transfer in point multipoles. This admits a ready description of ionic polarizability and crystal field splitting. It is necessary in constructing such a model to find a number of parameters that mimic their more exact counterparts in the density functional theory. We describe in detail how this is done using a combination of intuition, exact analytical fitting, and a genetic optimization algorithm. Having obtained model parameters we show that this constitutes a transferable scheme that can be applied rather universally to small and medium sized organic molecules. We have shown that the model gives a good account of static structural and dynamic vibrational properties of a library of molecules, and finally we demonstrate the model's capability by showing a real time simulation of an enolization reaction in aqueous solution. In two subsequent papers, we show that the model is a great deal more general in that it will describe solvents and solid substrates and that therefore we have created a self consistent quantum mechanical scheme that may be applied to simulations in heterogeneous catalysis. © *2014 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4887095]

I. INTRODUCTION

In this series of three papers we present a general theory and implementation of the tight binding (TB) approximation which can be successfully applied to a quantitative study of chemical reactions in gas phase, in solvent mixtures, and at surfaces in the context of heterogeneous catalysis. This first paper serves to introduce the method and describes the generation of TB models which are self consistent in the treatment of charge transfer and ionic polarizability. In Paper II[1] we describe models for the solvent and in Paper III we describe adsorption and dissociation of water at the surfaces of transition metal oxides.[2] Our approach is unique compared with existing classical and semi-empirical quantum mechanical models in that each particular element (C, O, H, and the transition metals) is described by a parameter set which is independent of the environment. So, for example, an oxygen atom may transfer seamlessly between an oxide phase and a water or alcohol molecule. Similarly single, double, or triple bonds are created as a consequence of the *bond order*, generated by the self consistent density matrix. In this sense the model is in the spirit of an *ab initio* quantum chemical scheme (and indeed the underlying theory derives from a second order expansion of the Hohenberg-Kohn functional) but is two to three orders of magnitude faster to compute due to the semi-empirical nature of the Hamiltonian. The price to pay, of course, is that

the parameterization of the Hamiltonian requires an extensive fitting strategy. We have managed to achieve this for the first time in a universal transferable parameter set.

To reach this goal has required some iteration between the three pillars of the model, namely, the gas phase molecules, the solvents, and the solid state. This means that we cannot present the process of parameterization in a logical linear sequence. Our universal model arises from earlier work in the construction of a model for water.[3] It turned out that this model contained a number of defects which have now been corrected and will be described in Paper II.[1] In particular the $s−p$-splitting in oxygen was quite unsuitable for a description of transition metal oxide bandstructures. This was subsequently corrected and using the new parameterization of oxygen as well as modified hydrogen parameters we created a model for water, employing polarization of the oxygen ion to the dipole level, which correctly renders the monomer polarizability, the relative densities of water and ice, the static dielectric constant, and self-diffusivity in the liquid and its radial distribution functions. So the task in this paper is to describe how we have augmented this model with a parameterization of the carbon atom and we show how with this freedom alone we are able to construct a transferable scheme that describes structural and dynamical properties of a broad set of molecules, including saturated, unsaturated and aromatic hydrocarbons, alcohols, ketones, aldehydes, and esters. Furthermore, our model is capable of atomistic simulations of chemical reactions using molecular dynamics.

---

a)Author to whom correspondence should be addressed. Electronic mail: Tony.Paxton@KCL.ac.uk.

The structure of this paper is as follows. A detailed description of the theory is given in Sec. II and in Sec. III we describe how parameters of our polarizable ion tight binding model are obtained, by intuition, analytical fitting, and genetic optimization. The full set of parameters applicable to organic molecules is given in Sec. IV, which means that a reader can in principle reproduce all of our results. In Sec. V we show structural, electronic, and dynamical properties of molecules included in the fitting, and in Sec. VI we demonstrate the transferability of the parameter set through test calculations on a library of small organic molecules featuring single, double, and triple bonds and a range of functional groups. In Sec. VII we demonstrate applicability into larger molecules, including a detailed analysis of the vibrational densities of states in hexane, hex-1-ene, 1-phenylbutane, and 1-phenyl-3-butanone (benzylacetone). Finally, in Sec. VIII we illustrate by showing snapshots from a real time molecular dynamics simulation, the enolization of benzylacetone initiated by the Grothuss proton transfer in aqueous solution. Discussion and conclusions are to be found in Sec. IX.

II. SELF CONSISTENT POLARIZABLE ION TIGHT BINDING (PITB)

TB theory bears a superficial resemblance to semi-empirical quantum chemistry schemes,[4–14] but whereas the latter are approximations to Hartree–Fock theory, tight binding is predicated upon the density functional theory (DFT).[15,16] Among other benefits this allows a more ready extension to highly correlated electron systems.[17,18] If the Hohenberg–Kohn total energy[19] is expanded to second order in the difference $\delta\rho$ between the self consistent density and some trial, input density $\rho^{\text{in}}(\mathbf{r})$, one obtains[20–22]

$$E^{(2)} = \sum_{\substack{n \\ \text{occ.}}} \langle \psi_n | H^0 | \psi_n \rangle$$
$$- \int d\mathbf{r} \rho^{\text{in}}(\mathbf{r}) V_{\text{xc}}^{\text{in}}(\mathbf{r}) + E_{\text{xc}}[\rho^{\text{in}}] - E_{\text{H}}[\rho^{\text{in}}] + E_{ZZ}$$
$$+ \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \left\{ e^2 \frac{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \right.$$
$$+ \left. \delta\rho(\mathbf{r}) \frac{\delta^2 E_{\text{xc}}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \delta\rho(\mathbf{r}') \right\}, \qquad (1)$$

where $E_{\text{xc}}[\rho^{\text{in}}]$ and $E_{\text{H}}[\rho^{\text{in}}]$ are the exchange-correlation and direct Coulomb (Hartree) energy functionals evaluated for the input density, $\rho^{\text{in}}(\mathbf{r})$. The sum is over occupied states and ψ_n are solutions of the Kohn–Sham equations[23] in the effective potential

$$V_{\text{eff}}^{\text{in}} = V_{\text{ext}} + V_{\text{H}}^{\text{in}} + V_{\text{xc}}^{\text{in}}$$

in which V_{ext} is the external potential due to the nuclei and the Hartree and exchange–correlation potentials, V_{H}^{in} and $V_{\text{xc}}^{\text{in}}$, are constructed from the input charge density. The non self consistent Hamiltonian H^0 is the sum of this effective potential and the non interacting kinetic energy of the electrons. E_{ZZ} is the pairwise Coulomb interaction between the nuclei.

In (1) the first two lines are a statement of the Harris–Foulkes approximation,[16,24] leading to a *first order* total energy $E^{(1)}$. Sutton *et al.*[15] and Foulkes and Haydock[16] showed that to within an approximation to the exchange and correlation, the second line is a pairwise energy as long as the input density is a sum of atom centered, spherical charge distributions. Therefore, since the first line is the one particle sum or *band energy* $E_{\text{band}}^{(1)}$, then the usual non self consistent TB approximation to the total energy, namely,

$$E_{\text{tot}}^{(1)} = E_{\text{band}}^{(1)} + E_{\text{pair}} \qquad (2)$$

may be rationalized as an approximation to the Harris–Foulkes density functional total energy. The two terms in (2) are the basic tenets of the standard TB approach pioneered by Slater and Koster[25] and built into a working scheme for much of the periodic table by Harrison.[26] The band energy may be written

$$E_{\text{band}}^{(1)} = \text{Tr}[\hat{\rho}^{\text{out}} H^0]$$

and $\hat{\rho}^{\text{out}}$ is the density operator obtained from the solution of the Kohn–Sham equation in the potential $V_{\text{eff}}^{\text{in}}$. In a TB calculation the trace is usually taken in the eigenvectors of the Hamiltonian obtained by diagonalization.[27] However, the theory is equally applicable if the density matrix is obtained in real space.[28,29] E_{pair} is a sum of usually repulsive pair potentials.

In detail, the ingredients of a non self consistent TB model are the pair potentials, $\phi(r)$, between atoms labeled by their positions, $\mathbf{R}$ and $\mathbf{R}'$, with respect to some origin, which are functions of the distance between the atoms, so that

$$E_{\text{pair}} = \frac{1}{2} \sum_{\mathbf{R} \neq \mathbf{R}'} \phi_{\mathbf{R}\mathbf{R}'}(|\mathbf{R} - \mathbf{R}'|)$$

and the matrix elements of the Hamiltonian H^0 which are[25]

$$H^0_{\mathbf{R}L,\mathbf{R}'L'} = \begin{cases} \varepsilon_{\mathbf{R}\ell} \delta_{\ell\ell'}, & \mathbf{R} = \mathbf{R}' \\ E_{LL'}(\mathbf{R} - \mathbf{R}'), & \mathbf{R} \neq \mathbf{R}' \end{cases}. \qquad (3)$$

Here $L = \{\ell m\}$ is a composite angular momentum index. In an *spd* basis, L takes the values 0–8 for the orbitals

$$s, p_y, p_z, p_x, d_{xy}, d_{yz}, d_{3z^2-r^2}, d_{zx}, d_{x^2-y^2}.$$

We call $\varepsilon_{\mathbf{R}\ell}$ the on-site energies. In Harrison's "solid state table" these are Hartree–Fock term values. Note that there are no on-site, off diagonal matrix elements of H^0; these are *crystal field* terms and they are furnished by the PITB model, (10) below. Furthermore, the on-site energies are degenerate within each ℓ-manifold; again this degeneracy is lifted by crystal field effects in the PITB theory. The intersite Hamiltonian matrix elements in (3) are constructed from fundamental bond integrals, $V_{\ell\ell'm}$, which depend on the bond length, multiplied by angular dependent factors taken from the Slater–Koster table.[25,26] In an orthogonal *spd* basis, in the two center approximation, there are three on-site energies per atom species (ε_s, ε_p, and ε_d) and ten fundamental bond integrals for each pair of atom species, which depend on bond length, $|\mathbf{R} - \mathbf{R}'|$. They are[26]

$$V_{\ell\ell'm} = \{V_{ss\sigma}, V_{sp\sigma}, V_{pp\sigma}, V_{pp\pi},$$
$$V_{sd\sigma}, V_{pd\sigma}, V_{pd\pi}, V_{dd\sigma}, V_{dd\pi}, V_{dd\delta}\}$$

in which the first two indices indicate the orbitals contributing to the bond and the last the type of bond, σ or π, that results.

The tasks in constructing a TB model are then, (i) to find values of the on-site energies and fundamental bond integrals and parameters for the pair potentials and (ii) to find appropriate simple mathematical rules for how these scale with bond length or with the volume of a crystal. Both of these can be informed by canonical band theory as described by Harrison[26] in his construction of the solid state table from which an initial set of parameters is likely to get the user about three quarters of the way to a working model. In this work, we by and large adopt a scaling law due to Goodwin, Skinner, and Pettifor[30] (GSP) in which a matrix element or pair potential depends on distance like

$$\left.\begin{aligned} V_{\ell\ell'm}(r) &= V^0_{\ell\ell'm} \\ \phi(r) &= \phi^0 \end{aligned}\right\}$$
$$\times \left(\frac{r_0}{r}\right)^n \exp\left[n\left\{ -\left(\frac{r}{r_c}\right)^{n_c} + \left(\frac{r_0}{r_c}\right)^{n_c} \right\} \right]. \quad (4)$$

This is essentially a power law decaying as r^{-n} as in canonical band theory, whose value is $V^0_{\ell\ell'm}$ or ϕ^0 at the equilibrium bond length r_0, cut off with a multiplicative exponential function starting at the origin. The parameters r_c and n_c specify the cut off radius and decay of the cut off function. Exactly how these parameters are obtained is described in Sec. III.

In addition to the GSP ansatz (4), we shall be using the sum of exponentials times power law functions (EPL) as an alternative way to describe the distance dependence of pair potentials:

$$\phi(r) = \sum_i \phi^0_i \, (r_0/r)^{m_i} \, \exp[-p_i(r-r_0)]. \quad (5)$$

An example is the O–O pair potential which is repulsive at short separation but attractive at large distances due to the r^{-6} dispersion interactions (this is discussed in detail in Paper II[1]).

The diagonal matrix elements of $\hat{\rho}^{\text{out}}$, in the basis of a linear combination of atomic orbitals such as is used in TB, allow one to examine the Mulliken charge at each atomic site.[27] This leads to the construction of a self consistent TB scheme based in point charge electrostatics.[31–33] One can write

$$H = H^0 + H' \quad (6)$$

and H' now contains two terms: a Madelung potential energy and a Hubbard potential energy, which is a Hubbard U parameter multiplying the charge transfer at each site. These correspond, respectively, to inter- and intra-atomic electrostatic terms. This is the basis of the self consistent charge density functional tight binding (SCC-DFTB) method.[21] The two principal differences between our PITB and SCC-DFTB are (i) we do not seek the parameters of the model by explicit calculation, rather we adopt a fitting strategy, and (ii) we go beyond the point charge electrostatics.[20, 34] The second aspect is particularly important as it is this feature that allows us to study polarizability of the ionic charge, which has led to an understanding, among other things, of the stabilization of tetragonal and monoclinic phases in zirconia[35] and the dielectric constant of water.[1, 3]

In PITB theory we seek a self consistent field approximation to $E^{(2)}$. To this end we write an energy functional to replace (2)

$$E^{(2)}_{\text{tot}} = E^{(2)}_{\text{band}} + E_{\text{pair}} + E_2 \quad (7)$$

and

$$E^{(2)}_{\text{band}} = \text{Tr}[\hat{\rho} H^0]$$

using a density operator $\hat{\rho}$ constructed from the self consistent eigenvectors of $H = H^0 + H'$. In (7)

$$E_2 = \frac{1}{2} \sum_{\mathbf{R}} \left(e^2 \sum_L Q_{\mathbf{R}L} V_{\mathbf{R}L} + U_{\mathbf{R}} \delta q^2_{\mathbf{R}} \right). \quad (8)$$

The first and second terms are intended to model the third and fourth lines in (1), respectively.[18] Thus the intra-atomic Coulomb energy is simply a Hubbard U times the square of the charge transfer

$$\delta q_{\mathbf{R}} = q_{\mathbf{R}} - q^0_{\mathbf{R}} = \frac{1}{\sqrt{4\pi}} Q_{\mathbf{R}0},$$

where $q^0_{\mathbf{R}}$ is the nominal charge on the free atom or ion, now occupying the site at position $\mathbf{R}$—normally the model assumes neutral free atoms. $Q_{\mathbf{R}0}$ is the monopole moment of the charge transfer, but in the PITB we go beyond the point charge model by writing down point multipole moments of the charge which may be approximated from the eigenvectors $c^n_{\mathbf{R}L}$ of H if we define new crystal field "strength" parameters, $\Delta_{\ell\ell'\ell''}$. In this way we write[27, 36] for $\ell > 0$,

$$Q_{\mathbf{R}L} = \sum_{L'L''} \rho_{\mathbf{R}L'\mathbf{R}L''} \, \Delta_{\ell'\ell''\ell} \, C_{L'L''L},$$

where

$$\rho_{\mathbf{R}L'\mathbf{R}L''} = \sum_{\substack{n \\ \text{occ.}}} \bar{c}^n_{\mathbf{R}L'} c^n_{\mathbf{R}L''}$$

is an on-site density matrix block (in an orthogonal representation), and

$$C_{L''L'L} = \iint d\Omega \, Y_{L''} \, Y_{L'} \, Y_L$$

are Gaunt coefficients over real spherical harmonics. It is these that impose selection rules that limit the number of "strength," or crystal field parameters required by the model. In a basis of s, p, and d orbitals ($0 \le \ell', \ell'' \le 2$) the parameters are

$$\Delta_{011} = \Delta_{101} = \Delta_{spp},$$
$$\Delta_{112} = \Delta_{ppd},$$
$$\Delta_{022} = \Delta_{202} = \Delta_{sdd},$$
$$\Delta_{121} = \Delta_{211} = \Delta_{pdp},$$
$$\Delta_{222} = \Delta_{ddd},$$
$$\Delta_{123} = \Delta_{213} = \Delta_{pdf},$$
$$\Delta_{224} = \Delta_{ddg}.$$

For hydrocarbons and water we use just s and p orbitals in the TB basis and only the first two parameters are needed. In fact,

we only use the first, Δ_{spp}, which provides the strength of the $\ell = 1$ atomic dipole moment arising from a mixing of the s and p orbitals on a single site. The second, Δ_{ppd}, describes the formation of an $\ell = 2$ quadrupole moment from the redistribution of populations among the three p orbitals on a site. (In Paper III in this series[2] in which we deal with transition metal oxides we include d orbitals and require further crystal field parameters, Δ_{ddd} and Δ_{ddg}.) The multipole moments appear in the inter-atomic Coulomb, or Madelung, energy in (8) multiplied by quantities which are expansion coefficients of the electrostatic potential energy. If all atomic sites $\{\mathbf{R}'\}$ are decorated with multipole moments of the charge, $Q_{\mathbf{R}'L'}$, and if we ask what is the electrostatic potential in the neighborhood of an atomic site $\mathbf{R}$, then first we expand that potential into spherical harmonic polynomials,

$$V_{\mathbf{R}}(\mathbf{r}) = \sum_L V_{\mathbf{R}L}\, r^{\ell}\, Y_L(\hat{\mathbf{r}})$$

and then Poisson's equation can be used to relate the electrostatic potential components to the multipole moments on all atoms in the molecule, or in a periodic system. The solution to Poisson's equation in this case is[27,37]

$$V_{\mathbf{R}L} = \sum_{\mathbf{R}'L'} \tilde{B}_{\mathbf{R}L\,\mathbf{R}'L'}\, Q_{\mathbf{R}'L'}.$$

The quantities $\tilde{B}_{\mathbf{R}L\,\mathbf{R}'L'}$ are elements of a generalized Madelung matrix,

$$\tilde{B}_{\mathbf{R}L\,\mathbf{R}'L'}$$
$$= 16\pi^2 \sum_{L''} (-1)^{\ell}\, \frac{(2\ell''-1)!!}{(2\ell+1)!!(2\ell'+1)!!}\, C_{L'LL''}\, K_{L''}(\mathbf{R}-\mathbf{R}') \tag{9}$$

in which

$$K_L(\mathbf{r}) = r^{-\ell-1}\, Y_L(\mathbf{r})$$

is a solid Hankel function. In the point charge case, $\ell = \ell' = 0$, $\tilde{B}_{\mathbf{R}0\,\mathbf{R}'0} = |\mathbf{R}-\mathbf{R}'|^{-1}$ as expected and one is left with a Madelung potential. In a periodic system the structure constants are evaluated using generalized Ewald summations.[38] Now, by variation of the energy E_2 with respect to the density we can obtain the self consistent matrix elements of the Hamiltonian (6)[22,27]

$$H'_{\mathbf{R}L'\,\mathbf{R}L''} = U_{\mathbf{R}}\,\delta q_{\mathbf{R}}\,\delta_{L'L''} + e^2 \sum_L V_{\mathbf{R}L}\, \Delta_{\ell'\ell''\ell}\, C_{L'L''L}. \tag{10}$$

It is important to note that self consistency in an orthogonal TB basis modifies only the on-site Hamiltonian matrix elements. It is evident that this amounts to a shift in the on-site energies of the non self consistent Hamiltonian, H^0 (the first term in (10)) and new terms which are off-diagonal in the angular momentum and are not contained in H^0. These are self consistent *crystal field* terms and their interpretation is rather clear. The electrostatic potential at a site $\mathbf{R}$ causes a coupling between orbitals on that site, in this case the L-component of the crystal field causes a coupling between the ℓ' and ℓ'' orbitals. For example, the $\ell = 4$ component of the ligand field in certain transition metal oxides will cause a splitting of the

degeneracy of the metal d-bands into the well known t_{2g} and e_g manifolds. The PITB method achieves this splitting and its dependence on the metal's ligand environment (tetrahedral or octahedral) automatically once a value for Δ_{ddg} is chosen.[27,36] Of course the crystal field parameters are the same as the multipole strength parameters since the coupling between orbitals on-site (10) feeds back to generate multipole moments as explained above. The ability of the PITB theory to describe the development of multipoles is essential to a proper description of the polarizability of small organic molecules and of water.[1,39]

In this series of papers we construct *orthogonal* TB models. The extension to non orthogonal bases and to magnetic systems is described in Refs. 18 and 27. In orthogonal PITB the interatomic force is calculated simply, using the Hellmann–Feynman theorem.[27,36] The force on the atom labeled $\mathbf{R}$ is the negative of the gradient of the total energy with respect to the position $\mathbf{R}$,

$$\nabla_{\mathbf{R}} E_{\text{tot}}^{(2)} = \text{Tr}[\hat{\rho}\,\nabla_{\mathbf{R}} H^0] + \nabla_{\mathbf{R}} E_{\text{pair}} + e^2 \sum_L Q_{\mathbf{R}L}\,\nabla_{\mathbf{R}} V_{\mathbf{R}L}$$

and as long as the density matrix, ρ, is self consistent the only contribution to the force in addition to the first two terms, which are the same as for non self consistent tight binding, is the classical electrostatic force. It is easy in PITB to achieve the necessary self consistency so that energy is conserved to excellent precision in any ensemble in molecular dynamics. For example, in an NPT ensemble of a 128 molecule periodic system of water we find an energy drift of no more than 0.04 eV over a simulation time of 100 ps.

III. FITTING STRATEGY

As explained above, the task is to fit a number of parameters of the TB model. For a model for organic molecules we must fit parameters to the individual species oxygen, carbon, and hydrogen as well as bond integrals and pair potentials between each pair of species. Paper II in this series describes a TB model for water[1] developed in conjuction with the model for organic molecules described here. In order to develop both models concurrently we use the hydrogen atom as a reference fixing its on-site energy and Hubbard U. From the water model we inherit the onsite parameters for oxygen and oxygen–hydrogen bond integrals and pair potentials to be used in the present model for organic molecules. While this reduces the dimensions of the parameter space we are still left with the task of fitting of upwards of 20 parameters. Our strategy is to use a combination of fitting procedures with the genetic algorithm of Schwefel[40,41] forming the backbone of the process. This is a tool to be used in conjuction with intuitive fitting as well as analytical fitting where possible for optimal results. Any possible analytical fitting reduces the dimensions of the parameter space to be searched thereby allowing the genetic algorithm to search more efficiently. In intuitive fitting we aim to obtain useful *values* and *bounds* for parameters within the model. We try to identify relationships between parameters and properties which we can calculate within tight binding and try to determine if and how parameters may relate to one another. Intuitive fitting may give a

reasonable model but generally it is followed by genetic fitting to refine the parameters.[3] But from intuitive fitting we obtain sensible starting points and bounds for the parameters and identify any parameters which can be excluded from the genetic fitting procedure and given a fixed value. This reduces the size of the space which the genetic algorithm is required to search and eliminates unphysical parameter sets from consideration.

We reduce the complexity of the fitting problem by splitting it into two sections. First we consider a hydrocarbon model containing only carbon and hydrogen atoms. We fit carbon on-site parameters along with carbon–hydrogen and carbon–carbon bond integrals and pair potentials in this stage. We then introduce the oxygen atom and have carbon–oxygen bond integrals and pair potentials remaining to fit. The fitting is again further reduced by fitting pair potentials completely analytically in a manner described below leaving on-site parameters and bond integrals to be fitted by the genetic algorithm.

Which properties to include into the fitting is an important and perhaps difficult choice. In our approach we try to include as much of the basic physics of the system as possible into the fitting. By including only small molecules and small systems we reduce the computation time for a single parameter set allowing a larger region of the parameter space to be searched. When we then begin to study larger systems we can be confident that having included the basic physics the larger scale properties are accurately predicted. There are several features which we want our model to reproduce and so we include these into the objective function used by the genetic algorithm. The ability to describe all different bonding states of carbon and oxygen is essential. Carbon–carbon bonds can exist as single, double, or triple bonds as well as the resonance bonding found in aromatic rings while carbon–oxygen bonds can be either single or double. With this in mind we include ethane (C_2H_6), ethene (C_2H_4), ethyne (C_2H_2), and benzene (C_6H_6) into the fitting. We want the model to describe the structural and dynamical properties of molecules so we require bond lengths, stretching force constants, and bond angles of these molecules to be correctly rendered. We also want the model to be able to describe the charge transfers across different bonds so we include some simple three carbon molecules, namely, propane (C_3H_8), which contains two single C−C bonds, propene (C_3H_6), which contains a single and a double C=C bond and propyne (C_3H_4), which contains a single and a triple C≡C bond, and fit their dipole moments. Similarly for carbon–oxygen bonds we include methanol (single C−O bond) and methanal (formaldehyde, double C=O bond) into the fitting, looking at their structural, dynamical, and charge transfer properties.

A. Fitting hydrocarbon model

Carbon–hydrogen (C–H) and carbon–carbon (C–C) bond integrals along with carbon on-site parameters are extracted from the genetic algorithm. For our starting point we take parameters from two different sources. Our C–H and C–C bond integrals are taken from a model by Horsfield *et al.*[42] and car-

bon on-site energies from Harrison's solid state table.[26] The hydrogen on-site parameters are shared with the water model[1] being fixed to $\varepsilon_s = -1$ Ry and $U = 1$ Ry. Note that our model does not include any hydrogen–hydrogen bond integrals since we do not expect to include H_2 molecules into our simulations. We do however include a short range repulsive pair potential to prevent hydrogen atoms from approaching each other too closely. The Horsfield *et al.* model was constructed for hydrocarbons using a form of self consistency, namely, local charge neutrality (LCN) in which no charge transfer between atoms is allowed. This is the limit of infinite Hubbard U. Whereas we can take over the bond integrals from Horsfield *et al.* directly, their on-site energies were shifted to account for the LCN self consistency. In our case where we employ a finite Hubbard U, the term values from Harrison's solid state table[26] serve as better starting points for the genetic optimization. For the Hubbard U we take sensible values to begin with and as we do not expect the carbon atom to be significantly polarizable we simply set the crystal field parameter Δ_{spp} to zero.

As mentioned above the pair potentials are fitted analytically. To do this we use the bond length dependence of the binding energy of target molecules. For the carbon–hydrogen pair potential we use methane as a target while for the carbon–carbon pair potential we use ethane, ethene, and ethyne. We can accurately calculate the binding energy of a molecule using a coupled cluster method to which we can compare the TB binding energy,

$$E_{bind}^{TB} = E_{band}^{(2)} + E_{pair} + E_2 - E_{ref}, \qquad (11)$$

where E_{ref} is the nominal energy of free atoms; whereas in the coupled cluster method,

$$E_{bind}^{CC} = E_{total}^{CC}(\text{molecule}) - E_{total}^{CC}(\text{free atoms}).$$

The second term in (11) is the pairwise repulsive potential energy which we aim to fit for a particular set of bond integrals and on-site parameters. Consider a tight binding calculation of methane with the CH pair potential set equal to zero. We call this energy $E_{attractive}^{TB}$. For a given set of bond integrals and on-site parameters we can calculate $E_{attractive}^{TB}$ as a function of CH distance (since all CH distances are equal). We can also obtain E_{bind}^{CC} as a function of CH distance. From these we also obtain the respective forces $F_{attractive}^{TB}$ and F_{bind}^{CC}, the first derivatives of the two energies, as functions of CH distance. We now assert that the pair potential which we require is one which when added to $E_{attractive}^{TB}$ will produce forces equal to F_{bind}^{CC}. We use the first derivative rather than the binding energy as we are more concerned with reproducing the bond lengths of our molecules than the binding energies. Our only concerns with regards to binding energies are first that the molecules have a sufficiently large binding energy to remain bound and second that the relative binding energies of molecules are reasonable. Although we fit the pair potential analytically here we have not used the force constants (second derivatives of the energy) and so they, along with the other properties mentioned above, are included into the objective function which the genetic algorithm is required to minimize.

B. Fitting oxygen

To integrate the oxygen atom from the water model into our organic molecules model we begin by setting the carbon–oxygen bond integrals equal to the carbon–carbon bond integrals and use this as a starting point for the genetic algorithm. In this stage we also fit the carbon–oxygen pair potential to bond lengths in methanol and methanal in the same manner as described before using coupled cluster calculations. The remaining dipole moments and force constants of methanol and methanal are used in fitting the carbon–oxygen bond integrals. At this point it may occur that there is no set of carbon–oxygen bond integrals and pair potentials that reproduce the fitting properties. This means that the carbon and hydrogen model we have fitted is incompatible with the oxygen atom from the water model. In fact there are, in general, many carbon and hydrogen parameter sets which result in good hydrocarbon models but not all of these parameter sets are compatible with the oxygen atom we use. For this reason some iteration in the fitting procedure is required. It is however still advantageous to separate the fitting problem into sections as we have done as it allows more efficient searching of the entire parameter space.

C. Post genetic fitting

The genetic algorithm tests every parameter set generated and outputs the properties calculated. A genetic run can easily generate and test thousands of parameter sets so an enormous amount of data is generated. While we could simply take the parameter set corresponding to the smallest objective function it is more beneficial to carry out tests on a number of parameter sets with different properties to find a set which best fits our purposes. Such tests would take too long to be included within the genetic algorithm and they include checking that ethyne and benzene return to linear and planar form from non-planar starting configurations during relaxations. We also require that electrons are not transferred from a carbon atom *to* a hydrogen atom within our three carbon test molecules propane, propene, propyne as well as isopropyl alcohol (IPA), and propanone (during genetic fitting we only check this condition for methane and methanol). We also perform relaxations and molecular dynamics simulations on the molecules butanone, propanone, and IPA to ensure that the molecules are stable, selecting those sets which give reasonable geometries. We reject any parameter sets which do not pass these tests and in that way we can pick the best of the genetic algorithm output. Finally, some adjustments are made to the parameters to try and negotiate any final compromises between the predicted properties of our test molecules.

IV. PARAMETERS

The parameters of the organic molecule model generated through the fitting process are shown along with the parameters inherited from the water model in Table I. For C–H, C–C, and C–O interactions we use the GSP function to describe the bond length dependence (see Sec. II). These functions allow us an extra degree of control over how the bond integrals and

TABLE I. Tight binding parameters. All quantities are in Rydberg atomic units. Parameters describing oxygen, hydrogen, and their interactions are inherited from our TB model for water.[1] Only *s*-orbitals are attached to hydrogen atoms, others have *s*- and *p*-orbitals included in the LCAO basis. Functions "GSP" and "EPL" denote a particular form of a scaling law and are given in Eqs. (4) and (5), respectively. The distance dependence of bond integrals and pair potentials listed in the Table are shown in Fig. 1.

	On-site parameters				
	ε_s	ε_p	Hubbard U	Δ_{spp}	Δ_{ppd}
Carbon	−1.8	−0.95	1.1	0	0
Oxygen	−2.1164	−1.1492	1.0775	−0.9430	0
Hydrogen	−1		1		

	C–H	C–C	C–O	O–H	O–O
	Bond integrals, $V_{\ell\ell'm}$, and scaling				
Function	GSP	GSP	GSP	GSP	GSP
$V^0_{ss\sigma}$	−0.6748	−0.3724	−0.1543	−0.5018	−0.0150
$V^0_{sp\sigma}$		0.4811	0.2125		0.0020
$V^0_{ps\sigma}$	−0.6356	−0.4811	−0.2932	−0.4362	−0.0020
$V^0_{pp\sigma}$		0.4500	0.1917		0.0500
$V^0_{pp\pi}$		−0.0600	−0.0548		−0.0200
$n_{ss\sigma}$	0.8007	2.95	2.4476	2.0963	2
$n_{sp\sigma}$		2.7	2.5873		2
$n_{ps\sigma}$	0.8007	2.7	2.4839	1.5019	2
$n_{pp\sigma}$		2.9	2.5088		3
$n_{pp\pi}$		2.9	3.3699		3
n_c	3.1955	6.5	6.5	4.0561	6
r_0	2.0485	2.9032	2.7	1.8094	5.6
r_c	2.2679	4.1196	4.4	3.7985	9
	Pair potentials, ϕ, and scaling				
Function	GSP	GSP	GSP	GSP	EPL
ϕ^0	1.2314	1.2980	0.3730	0.7367	
n	1.0717	3.5094	5.1	3.3502	
n_c	3.6543	6.5796	9.0	6.3096	
r_0	2.0485	2.9032	2.7	1.8094	5.6
r_c	2.2679	4.1196	4.4	3.3550	
ϕ^0_1					4.0306×10^{-3}
ϕ^0_2					-2.0265×10^{-3}
m_1					10
m_2					6
p_1					0
p_2					0
	Cut-off distances $[r^{(1)}_{\text{cut}}; r^{(2)}_{\text{cut}}]$				
$r^{(1)}_{\text{cut}}$	2.1	3.0	4.0	2.1	8
$r^{(2)}_{\text{cut}}$	3.9	4.55	5.0	5.5	11

pair potentials are cut off and decay through the r_c and n_c parameters (Fig. 1). To ensure all functions decay exactly to zero smoothly we multiply the functions (4) and (5) by 5th order polynomials defined between certain bond distances $r^{(1)}_{\text{cut}}$ and $r^{(2)}_{\text{cut}}$. The polynomials have value, first derivative and second derivative matching those of the function to be cut off at $r^{(1)}_{\text{cut}}$ and value, first derivative and second derivative equal to zero at $r^{(2)}_{\text{cut}}$.

Other parameters demand different considerations. For example, the Hubbard U is a parameter which has a role in controlling the amount of charge transfer allowed within a molecule. A larger value generally inhibits charge transfer between atoms. Consider a simple example using methane in which we begin with a large value for the Hubbard U of
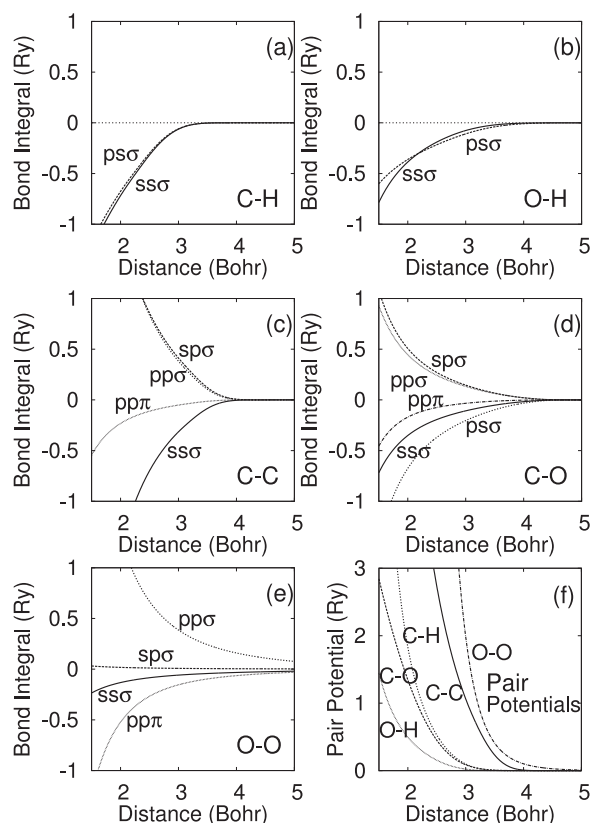
FIG. 1. Bond integrals and pair potentials using the parameters from Table I. (a) Carbon–hydrogen, (b) oxygen–hydrogen, (c) carbon–carbon, (d) carbon–oxygen, (e) oxygen–oxygen, and (f) pair potentials for all pairs.

angles within methanol and methanal leading to an indirect effect on the dipole moment, through the atomic positions rather than the charge transfer. The same is also true of $V_{pp\sigma}$ but to a lesser extent. $V_{pp\pi}$ however has a large effect on the dipole moment of methanal by controlling the charge transfer across the C–O bond while having a smaller effect in methanol. At first glance it would seem that $V_{pp\pi}$ could simply be adjusted to give the correct dipole moment for methanal but as $V_{pp\pi}$ is increased to this end it becomes impossible to fit a pair potential which gives a reasonable force constant. This is a case where a compromise needs to be made. Another point of note is that $V_{pp\sigma}$ controls the distance dependence of the attractive band energy, $E_{band}^{(2)}$. A larger value of $V_{pp\sigma}$ results in a stronger band energy at distances sampled during normal molecular vibration. While testing parameter sets we discovered that if $V_{pp\sigma}$ is too small then during molecular dynamics simulations of methanol normal molecular vibrations are enough to extend the C–O bond to a point where the quickly decaying band energy can be overcome by the repulsive pair potential resulting in bond breakage. The C–C $V_{pp\sigma}$ plays a similar role in carbon–carbon bonds. For example, in butanone the carbon–carbon bonds adjacent to the oxygen atom may contract slightly while the remaining carbon–carbon bond extends and may break. Increasing $V_{pp\sigma}$ is sufficient to counter this problem.

V. FITTED RESULTS

The results of the fitting process on the structural and dynamical properties of the training set of molecules are shown in Table II. All C–C bond lengths are practically equal to the target values thanks to the manner in which the pair potentials were fitted. The exception here is benzene, which is not included in the analytical fitting, with the bond length ∼0.05 bohr shorter than the target but still within 2%. Final adjustments to the C–O pair potential were made in the post genetic fitting stage to provide a closer fit to the C–O bond lengths in propanone and IPA. For this reason the C–O bond lengths of methanol and methanal have been compromised and are slightly longer than the target values but still well within acceptable bounds. The dynamical properties have also been fitted well for the linear molecules. Stretching force constants are within 5%–7% of the target with the exception of methanal. This is due to the compromise between its force constant and dipole moment as explained above. As it is still within 20% of the target we deem it acceptable. The force constant for the benzene ring stretch is stiffer than the target, but is still within 18%. Importantly we see that the relative ordering of the C–C and C–O force constants is correct, with C–C < C–O < C≃C < C=C < C=O < C≡C (as in the caption to Table II, C≃C designates an aromatic resonant bond as found in benzene).

We now turn to the charge transfer properties of the fitted molecules. In what follows, when we allude to charge transfer we imply always the transfer of electrons, that is, *negative* charge. In the fitting process we imposed a condition that no parameter sets in which charge was transferred from the carbon atom to the hydrogen atoms within the methane

carbon (∼1.5 Ry) and slowly lower it. Initially we are in a region of small charge transfer between carbon and hydrogen, typically <0.1e (the direction of the transfer may be dependent on other parameters). If we continue lowering the Hubbard U we enter a region of very large charge transfer in which self consistency can be unstable. If the Hubbard U is further reduced we enter a region where the charge transfer is again small and self consistency is again stable but the direction of charge transfer reversed. While this transition point occurs at different values for different parameter sets, in general a value of less than 1 Ry for the carbon Hubbard U runs the danger of entering an unstable region. Another factor which plays a role in the charge transfer between atoms within our organic molecules is the difference between the carbon-p and hydrogen-s on-site energies. As we have previously fixed the onsite energy and Hubbard U of the hydrogen atom we have only the carbon values to adjust. We find that for $\varepsilon_p \gtrsim -0.90$ Ry, within methane charge is transferred *from* carbon atoms *to* hydrogen atoms, which is unphysical. The C–H bond integrals also play a role in the amount of charge transfer within methane but they have a much larger role to play in the dipole moments of the simple three carbon molecules we have tested, namely, propane, propene, and propyne. The C–H bond integrals along with the carbon onsite energies and C–C $V_{sp\sigma}$ bond integral have greatest influence over the CCH bond angles in ethane and ethene.

Similar relationships are seen in the carbon–oxygen (C–O) bond integrals. $V_{sp\sigma}$ has the largest effect on the OCH

TABLE II. Fitted structural and dynamical properties: bond lengths, force constants, and bond angles. These quantities were used within the genetic algorithm for fitting. All targets are taken from CCSD(T)/aug-cc-pVTZ calculations performed using NWCHEM.[43] All bond lengths given in bohr, bond angles in degrees, and force constants, k_{ss}, in Ry/bohr^2. Single, double, and triple bond lengths are indicated by $-$, $=$, and $\equiv$, respectively. The resonating bond length in benzene is denoted C $\simeq$ C.

Name	Formula		TB	Target
Methane	CH$_4$	C$-$H	2.0531	2.0596
		k_{ss}	0.7305	0.6994
Ethane	C$_2$H$_6$	C$-$C	2.8942	2.8915
		k_{ss}	0.6363	0.5839
		$\angle$CCH	107.99	111.20
Ethene	C$_2$H$_4$	C$=$C	2.5305	2.5291
		k_{ss}	1.2651	1.1966
		$\angle$CCH	119.55	121.32
Ethyne	C$_2$H$_2$	C$\equiv$C	2.2873	2.2870
		k_{ss}	2.1971	2.0694
Benzene	C$_6$H$_6$	C $\simeq$ C	2.6108	2.6694
		k_{ss}	1.1531	0.9788
Methanol	CH$_3$OH	C$=$O	2.7015	2.6939
		k_{ss}	0.6591	0.6967
		$\angle$COH	101.78	107.90
		$\angle$OCH	102.77	106.31
Methanal	CH$_2$O	C$=$O	2.3349	2.2883
		k_{ss}	1.3428	1.6536
		$\angle$OCH	120.07	121.64

molecule were accepted. This has ensured that charge is transferred from the hydrogen atoms to the carbon atom as expected from experiment[44,45] and from arguments based on the relative electronegativities of carbon and hydrogen (2.5 and 2.2, respectively). The charge transferred between atoms as deduced from atom centered Mulliken charges is shown in Table III for our simple carbon and hydrogen molecules. The charge transfer is greatest in methane as the carbon is only bonded to hydrogen atoms from which it can easily draw charge. As the hybridization state of the carbon atom changes from sp^3 to sp^2 to sp more charge is accumulated on the carbon atoms despite them being bonded to fewer hydrogen atoms. This is in line with experimental infrared measurements.[44,45] The fitted dipole moments for our three-carbon molecules are shown in Table IV and by considering the Mulliken charges, $\delta = \delta q_{\mathbf{R}}$, in units of the proton charge e, of the carbon atoms in each of these molecules we can understand the origin of these dipoles. Starting with propane, the end chain carbon atoms have charges $\delta = -0.18e$

TABLE III. Mulliken charges on carbon atoms in the tight binding model. All charges given in units of the proton charge, e.

Molecule	Carbon charge (e)
CH$_4$	-0.2855
C$_2$H$_6$	-0.1573
C$_2$H$_4$	-0.1620
C$_2$H$_2$	-0.1662

TABLE IV. Fitted dipole moments, **p**, of some simple hydrocarbons. Target taken from MP2 calculations using NWCHEM with aug-cc-pVDZ basis set. Dipole moments given in Debye. Orientations of propene and methanol shown in Fig. 2.

		TB	Target
Propane	$|\mathbf{p}|$	0.061	0.087
Propene	$|\mathbf{p}|$	0.249	0.351
	p_x	0.248	0.343
	p_y	0.025	-0.03
Propyne	$|\mathbf{p}|$	0.822	0.755
Methanol	$|\mathbf{p}|$	1.403	1.711
	p_x	0.946	1.360
	p_y	-1.042	-1.040
Methanal	$|\mathbf{p}|$	1.880	2.411

with the central carbon having $\delta = -0.05e$. The symmetry of this charge distribution results in a practically zero dipole moment. In propene and propyne the three carbon atoms have charges $\delta = -0.24e$, $-0.01e$, $-0.16e$ and $\delta = -0.33e$, $+0.06e$, $-0.12e$, respectively (the double or triple bond is between the first and second carbon atoms). This indicates a non-zero bond moment between the first and second carbon atoms with a build up of charge on one side of the bond. This breaks the symmetry seen in the case of propane and results in a dipole moment. This non-zero bond moment is to be expected considering that the different hybridized states of the carbon atom have different electronegativities[46] with $sp^3 < sp^2 < sp$ and shows our model is able to describe the interaction of these different hybridized states of carbon correctly.

In Table IV we also show the fitted dipole moments of methanol and methanal. Fig. 2 shows the orientations of the x and y axes referred to in Table IV. For these molecules we have been able to fit the dipole moments reasonably well. As mentioned before, in the case of methanal a compromise needed to be made between the dipole moment and the C–O stretching force constant. In methanol the x component of the dipole moment (due mostly to the O–H bond) is smaller than the target value.

VI. APPLICATION OF MODEL TO SMALL ORGANIC MOLECULES

We have made simulations on a range of small organic molecules using our TB model and compared the results with calculations using second order Møller–Plesset[47] (MP2)
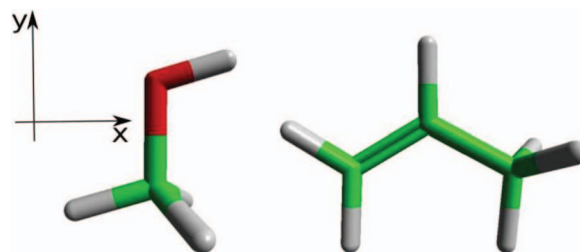

FIG. 2. Orientations of methanol and propene used to define x and y components of the dipole moment in Table IV.

corrections to the Hartree–Fock total energy, and experimental results in order to check the predictive power of our model. We have considered a library of molecules covering a range of different functional groups with the aim of showing that our TB model gives useful descriptions across many different environments and can be applied with confidence to a large range of organic molecules. The functional groups and molecules we have considered are alkanes (propane), alkenes (propene), alkynes (propyne), primary and secondary alcohols (ethanol and IPA), aldehydes (ethanal), ketones (propanone), carboxylic acids (ethanoic and propenoic acid), ethers (dimethyl ether), and esters (ethyl acetate). We begin by looking at the structures of the molecules before moving on to look at dynamical properties in the form of vibrational frequencies. Finally, we consider charge transfer properties, looking at Mulliken charges and dipole moments of the molecules in question.

A. Structures

To evaluate the accuracy of the structures predicted by our TB model we have made geometry optimizations on each of the molecules in our library and compared the relaxed structures to geometries obtained from MP2 geometry

TABLE V. Predicted structural data for a library of organic molecules. Bond lengths are in bohr and bond angles in degrees. Comparisons are made with MP2 calculations using NWCHEM[43] with the aug-cc-pVDZ basis set. Single, double, and triple bonds lengths are indicated as in Table II.

Functional group	Molecule	Property	Tight binding	MP2
Alkane		C−C	2.8812	2.8903
		∠CCC	107.18	112.14
Alkene		C=C	2.5257	2.5474
		C−C	2.8114	2.8395
		∠CCC	118.60	124.61
Alkyne		C≡C	2.2881	2.3263
		C−C	2.7667	2.7732
1° alcohol		C−C	2.8485	2.8644
		C−O	2.6799	2.7169
		∠CCO	103.49	107.09
2° alcohol		C−C	2.8522	2.8803
		C−O	2.6609	2.7210
		∠CCC	113.03	112.30
Aldehyde		C−C	2.8229	2.8438
		C=O	2.3136	2.3109
		C−H	2.0394	2.1067
		∠OCH	119.65	119.96
Ketone		C−C	2.8233	2.8584
		C=O	2.3027	2.3191
		∠CCC	119.49	117.26
Carboxylic acid		C−C	2.8455	2.8418
		C−O	2.5677	2.5834
		C=O	2.2733	2.3000
		∠OCO	133.62	122.56
α,β-unsaturated carboxylic acid		C=C	2.5360	2.5489
		C−C	2.7314	2.8004
		C=O	2.2849	2.3068
		C−O	2.5712	2.5489
		∠OCO	132.90	122.69
		∠CCC	115.98	123.71
Ether		C−O	2.6927	2.6840
		∠COC	103.72	110.76
Ester		C−C	2.8493	2.8454
		C=O	2.3010	2.3040
		C−O	2.5260	2.5697
		O−C	2.7937	2.7303
		∠CCO	119.10	125.99
		∠CCO	107.77	110.72
		∠COC	101.29	113.86

TABLE VI. Frequencies of normal modes of vibration for methane, comparing the tight binding model results with experimental frequencies.[48] Frequencies given in cm^{-1}.

	TB	Expt.
Symmetric stretch	3094	2917
Asymmetric stretch	3064	3019
CH scissors	1851	1534
CH bend	1719	1306

relaxations. All MP2 relaxation simulations were done using NWCHEM[43] with an aug-cc-pVDZ basis set. We present the results of this comparison in Table V. It is clear that in general the TB structures accurately predict the MP2 relaxed structures. The TB model predicts several trends observed in the MP2 relaxed structures. For the alkane, alkene, and alkyne molecules the model predicts the effect of carbon–carbon double and triple bonds on neighboring single bonds causing them to shorten. In the primary and secondary alcohols, aldehydes, and ketones the MP2 values predict a shortening of carbon–carbon bonds adjacent to carbon–oxygen bonds and we see this also in the TB model.

B. Vibrational frequencies

To determine how well our TB model describes the dynamical properties of organic molecules we have calculated the vibrational frequencies of the normal modes of vibration using the dynamical matrix approach[49] for our library of molecules. We have compared the results against experimental infrared results from Refs. 48 and 50. We begin by looking at some of the predicted vibrational frequencies of the molecules which were used in the fitting of our model as these may be indicative of some of the properties of all other organic molecules. Table VI contains the predicted frequencies of the normal modes of vibration of methane. Although we do not correctly predict the relative ordering of the symmetric and asymmetric stretches both are quite close to the experimental value. The C–H bending modes are shifted by $\sim$300–400 cm^{-1} but are still within 20%–30% of the exper-

FIG. 4. Predicted TB vibrational frequencies *vs.* experimental[48] vibrational frequencies for benzene ring stretches and deformations.

imental value and we have retained the relative ordering of these bending modes.

We show the predicted C–H bending and stretching frequencies for ethane, ethene, ethyne, and benzene against the experimental values in Fig. 3. The central line corresponds to the point on the graph where the predicted and experimental values are equal while the outer lines mark a 20% deviation which is the maximum we would like to accept. The predicted frequencies mostly lie within this range with a few exceptions. In the approximate range of 1500–2000 cm^{-1} in the predicted frequencies there are deviations. These are identified as the in-plane C–H bends of benzene, the CH_3 bends of ethane, and the CH_2 bends of ethene. In our model they are all shifted to higher frequencies compared to experiment, as was the case for the C–H bending frequencies within methane. To the upper right of the plot are the C–H stretching frequencies. The model predicts that the value of the asymmetric stretching frequencies follow the ordering ethane < ethene $\simeq$ benzene < ethyne in accordance with experimental data. Predicted frequencies of less than $\sim$1350 cm^{-1} correspond to benzene and ethyne out of plane C–H bends as well as the movements of entire CH_3 and CH_2 groups such as rocking and wagging motions. Fig. 4 shows a similar plot for the ring stretches and deformations of benzene. Those to the upper right correspond
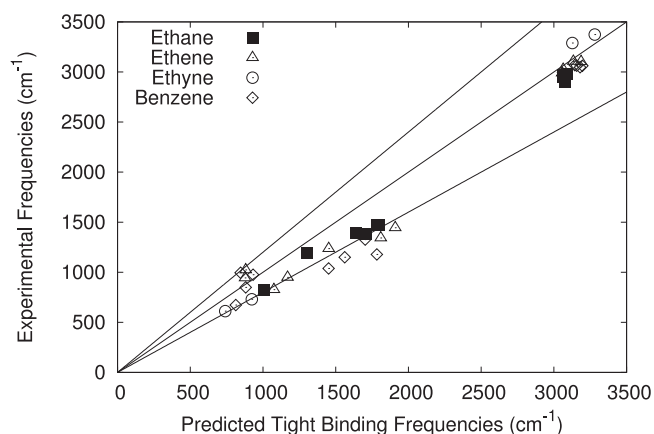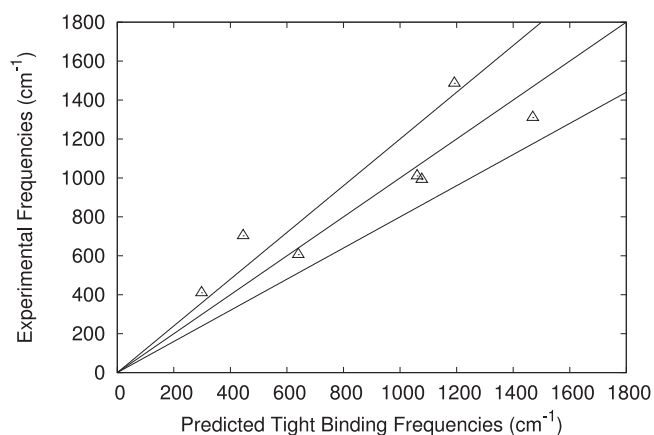

FIG. 3. Predicted TB vibrational frequencies vs. experiment[48] for ethane, ethene, ethyne, and benzene C–H bending and stretching modes. The dotted lines indicate an error envelope of $\pm$20%.

TABLE VII. Comparison of frequencies of normal modes of vibration for methanol between TB and experiment.[48] Only the CO stretch frequency was fitted. Frequencies are in cm^{-1}.

Description	TB	Expt.
CO stretch	1033	1033
OH stretch	3459	3681
OH bend	1258	1345
CH sym stretch	2999	2844
CH asym stretch	3140	3000
	3127	2960
CH bends	1745	1477
	1750	1477
CH_3 angle bend	1455	1455
CH_3 wag	972	1165
CH_3 twist	1031	1066

TABLE VIII. Comparison of frequencies of normal modes of vibration for methanal between TB and experiment.[48] Only the CO stretch frequency was fitted. Frequencies are in cm^{-1}.

Description	TB	Expt.
CO stretch	1625	1764
CH_2 sym stretch	2918	2783
CH_2 asym stretch	3132	2743
CH_2 scissors	1491	1500
CH_2 rock	1113	1249
CH_2 wag	1077	1167

to ring stretches while those to the lower left correspond to ring deformation normal modes with the out of plane deformations being underestimated by 100–250 cm^{-1}.

For methanol and methanal the calculated vibrational frequencies shown in Tables VII and VIII are in agreement with experimental data. For methanol, again, the C–H bending frequencies are shifted with respect to experimental values just as in the hydrocarbon molecules.

Moving on to the three-carbon hydrocarbon molecules in our library, the C–C stretching and bending frequencies presented in Table IX show that all the frequencies correspond well with the experimental values. For our remaining molecules the vibrational frequencies involving functional groups are collected together and shown in Table X. Overall our TB model compares very well with experimental results. With IPA and propanone we can investigate the effect of the carbon–oxygen bond on the carbon–carbon vibrational frequencies as shown in Table IX. Experimental results suggest that the carbon bending frequencies should not change significantly in propanone compared to propane and this is reproduced in the TB model. Experimental observations also suggest that the carbon–carbon symmetric stretching frequencies for IPA should be very similar to those of propane and that for propanone the symmetric and asymmetric stretches should be shifted to lower and higher frequen-

cies, respectively, compared with propane. These facts are predicted by our TB model.

C. Charge transfer properties

Finally, we examine the charge transfer properties of our library of molecules. Table XI shows the predicted dipole moments compared with the dipole moments calculated from MP2. For each case the molecules have first been relaxed with the respective methods. Our model in general underestimates target dipole moments. This is a consequence in part of the compromise in the fitting between the C–O force constant and dipole moment of methanal. Although our model underestimates the size of the dipole moment we still regard the predicted values as being more than acceptable. Comparing the Mulliken charges of our TB calculations for propane to those of propanone and IPA provides insight into the effect of the oxygen atom on neighboring carbon atoms. In propane the end chain carbon atoms both have a Mulliken charge of $\delta = -0.18e$ while the central carbon is effectively neutral with $\delta = -0.05e$. Comparing this with propanone which has $\delta = -0.40e$ for end chain carbons and $\delta = +0.72e$ for the central carbon atom, we see the electronegative oxygen atom pulling charge towards itself from the central carbon atom. The large molecular dipole moment is a consequence of this. A similar effect is seen in IPA with $\delta = -0.28e$ for end chain carbons and $\delta = +0.35e$ for the central carbon although the amount of charge drawn by the oxygen atom from the central carbon atom is smaller as it is also bonded to a hydrogen from which it easily draws charge. The charge on the oxygen atom is roughly the same in both cases with $\delta = -0.64e$ and $\delta = -0.67e$ for propanone and IPA, respectively, but the size of the induced dipole differs having values of 1.01 D and 0.809 D, respectively. Our model also reproduces the dipole moments of carboxylic acids adequately. The Mulliken charges and carbon–oxygen bond lengths in ethanoic acid are shown in Table XII and compared with its carboxylate anion in which the hydroxyl hydrogen atom has been removed. The two

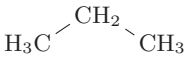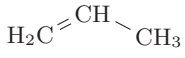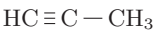TABLE IX. Three-carbon molecule CC normal mode vibrations compared with experiment.[48] Frequencies are in cm^{-1}.

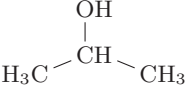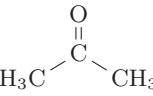| Molecule		Description	TB	Expt.
Propane	H_3C–CH_2–CH_3	CCC bend	370	369
		CC sym stretch	965	869
		CC asym stretch	1059	1054
Propene	H_2C=CH–CH_3	CCC bend	447	428
		C−C stretch	1052	920
		C=C stretch	1711	1652
Propyne	$HC \equiv C - CH_3$	CCC bend	383	328
		C−C stretch	991	931
		C≡C stretch	2241	2140
Iso-propanol	H_3C–$CH(OH)$–CH_3	CCC bend	344	
		CC sym stretch	925	
		CC asym stretch	1103	1072
Propanone	H_3C–$C(O)$–CH_3	CCC bend	350	385
		CC sym stretch	801	777
		CC asym stretch	1371	1216

TABLE X. Frequencies of normal modes of vibration in some organic molecules compared with experiment.[48] Frequencies are in cm^{-1}.

Functional group	Molecule	Property	Tight binding	Expt.
2° alcohol		OH stretch	3506	3657
		OH bend	1132	1251
		CO stretch	1034	1152
Dldehyde		CO stretch	1791	1743
		CCO bend	478	509
		CH stretch	3007	2822
		CH bend in plane	1354	1400
		CH bend out of plane	826	763
Ketone		CO stretch	1897	1731
		CO bend out of plane	454	484
		CO bend in plane	468	530
Carboxylic acid		CO double stretch	1990	1788
		CO single stretch	1144	1182
		OH stretch	3484	3583
		OH bend	1245	1250
Ether		CO sym stretch	981	928
		CO asym stretch	1108	1102
		COC bend	384	418

oxygen atoms in the carboxylate anion have the same Mulliken charge and both carbon–oxygen bond lengths are equal. This is a purely electronic effect due to stabilization of the anion through resonance bonding in which the extra electron is shared between the two oxygen atoms. This type of prediction is a major advantage of our model over classical force field methods.

D. Proton affinities

We have calculated proton affinities for several molecules and anions using the formula

$$E^{\mathrm{PA}} = E_{\mathrm{bind}}^{\mathrm{A}} + E_{\mathrm{bind}}^{\mathrm{H}^+} - E_{\mathrm{bind}}^{\mathrm{AH}}.$$

Note that the sign convention is opposite to the usual for chemical reactions. A positive proton affinity denotes an exothermic reaction.

Table XIII shows proton affinities for the hydroxyl ion, methoxide ion, the enolate anion of propenol, the acetate anion, water and ethene. For the negatively charged species our model gives very good agreement with the MP2 reference values, especially when taking into account that none of proton affinities were used for the fitting of the model. The agreement for neutral molecules water and ethene somewhat worsens. Importantly, the relative ordering of the proton affinities is re-

TABLE XI. Dipole moments (in Debye) of some organic molecules. Comparison is made with MP2 calculations using NWCHEM[43] with the aug-cc-pVDZ basis set.

Molecule	Molecule	Tight binding	MP2
Ethanol		1.275	1.620
Isopropyl alcohol		1.712	1.711
Ethanal (acetaldehyde)		2.480	2.796
Propanone		2.835	2.919
Ethanoic (acetic) acid		1.515	1.648
Prop-2-enoic (acrylic) acid		1.759	2.045
Ether		0.881	1.351

TABLE XII. Mulliken charges and bond lengths in ethanoic (acetic) acid and its carboxylate anion. Mulliken charges are in units of proton charge e and bond lengths are in bohr.

	Acid	Carboxylate anion
δ_O hydroxyl	-0.8209	-0.4984
δ_O carbonyl	-0.7439	-0.4984
C$-$O hydroxyl	2.5677	2.3994
C=O carbonyl	2.2733	2.3994

TABLE XIII. Proton affinities (kcal/mol) of several molecules obtained using our TB model. Comparisons are calculated with MP2 using NWCHEM with an aug-cc-pVDZ basis.

Molecule	TB	MP2
OH^-	409.8	391.8
CH_3O^-	400.2	386.8
$CH_3COCH_2^-$	369.6	360.7
CH_3COO^-	359.1	351.6
H_2O	118.4	169.1
C_2H_4	70.1	167.1

tained within our TB model. It is the relative proton affinities that define the direction of proton transfer reactions.

VII. LARGER MOLECULES

Our model is capable of describing larger organic molecules. As before we look at the relaxed structures, dipole moments, and vibrational frequencies of several organic molecules. Studying these larger molecules allows us to see long range effects of functional groups which would not be apparent with smaller molecules as well as allowing us to test phenyl groups within the model.

A. Structures and charge transfer properties

We compare the TB structures with MP2 optimized geometries. For all MP2 optimizations in this section we have used NWCHEM with a cc-pVDZ basis set.[43] In Table XIV we present the results of geometry optimizations on the molecules 1-phenylbutane, 1-phenyl-2-butene, methyl benzene (toluene), and 1-phenyl-3-butanone (benzylacetone) along with MP2 target geometries given in parentheses. The planar structure of each of these molecules is shown in Fig. 5. The linear chain carbon–carbon bond lengths are almost identical to the respective three carbon molecules investigated in Sec. VI in Table V and the phenyl group similar to benzene shown in Table II. This is true of both the MP2 values and our TB model. The bond joining the chain and the phenyl group
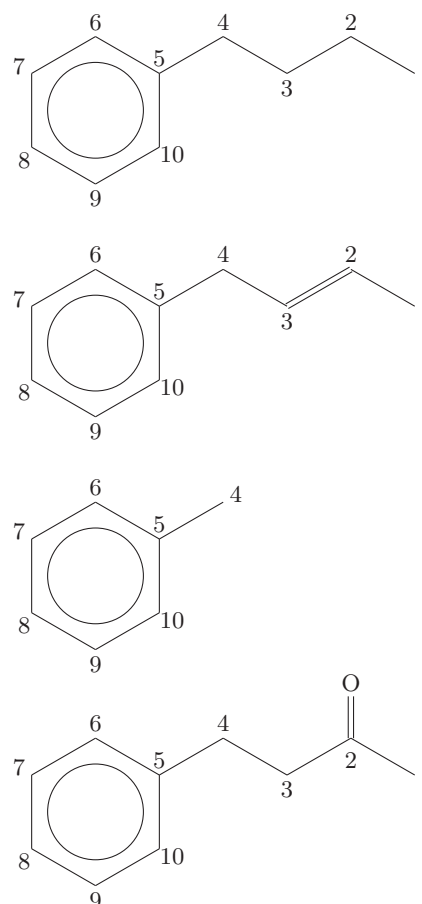

FIG. 5. The molecules of Table XIV. From the top, these are 1-phenylbutane, 1-phenyl-2-butene, toluene, and benzylacetone.

is consistently shorter than the MP2 values and may be an effect of the phenyl group whose bonds are also shorter than the target. The fact that the phenyl group geometries are virtually identical in each of these molecules shows that the effect of the functional groups is relatively short ranged vindicating our fitting approach in which we fitted only properties of small molecules. The dipole moments of these molecules are

TABLE XIV. Predicted structural data and dipole moments, **p**, for larger organic molecules, toluene (methyl benzene), 1-phenylbutane, phenylbutene (1-phenyl-2-butene), and benzylacetone (1-phenyl-3-butanone). Bond lengths are in bohr and dipoles in Debye. Values given in parentheses are taken from MP2 calculations using NWCHEM with cc-pVDZ basis set.[43] Labeling for the carbon atoms is as shown in Fig. 5.

| Bond	Toluene	Phenylbutane	Phenylbutene	Benzylacetone
C_1C_2		2.8832 (2.8901)	2.8135 (2.8403)	2.8234 (2.8691)
C_2C_3		2.8697 (2.8923)	2.5233 (2.5500)	2.8223 (2.8784)
C_3C_4		2.8847 (2.8885)	2.8096 (2.8469)	2.9077 (2.8945)
C_4C_5	2.7998 (2.8519)	2.7929 (2.8696)	2.7947 (2.8698)	2.7872 (2.8531)
C_5C_6	2.5994 (2.6603)	2.5999 (2.6656)	2.6000 (2.6581)	2.5997 (2.6525)
C_6C_7	2.6051 (2.6534)	2.6051 (2.6486)	2.6051 (2.6548)	2.6051 (2.6537)
C_7C_8	2.6058 (2.6536)	2.6058 (2.6555)	2.6059 (2.6519)	2.6058 (2.6525)
C_8C_9	2.6058 (2.6536)	2.6059 (2.6495)	2.6059 (2.6557)	2.6062 (2.6537)
C_9C_{10}	2.6052 (2.6534)	2.6052 (2.6571)	2.6053 (2.6503)	2.6050 (2.6525)
$C_{10}C_5$	2.5993 (2.6603)	2.6001 (2.6597)	2.5998 (2.6628)	2.6004 (2.6618)
CO				2.3103 (2.3108)
$\lvert p \rvert$	0.334 (0.325)	0.319 (0.462)	0.311 (0.154)	2.567 (2.990)
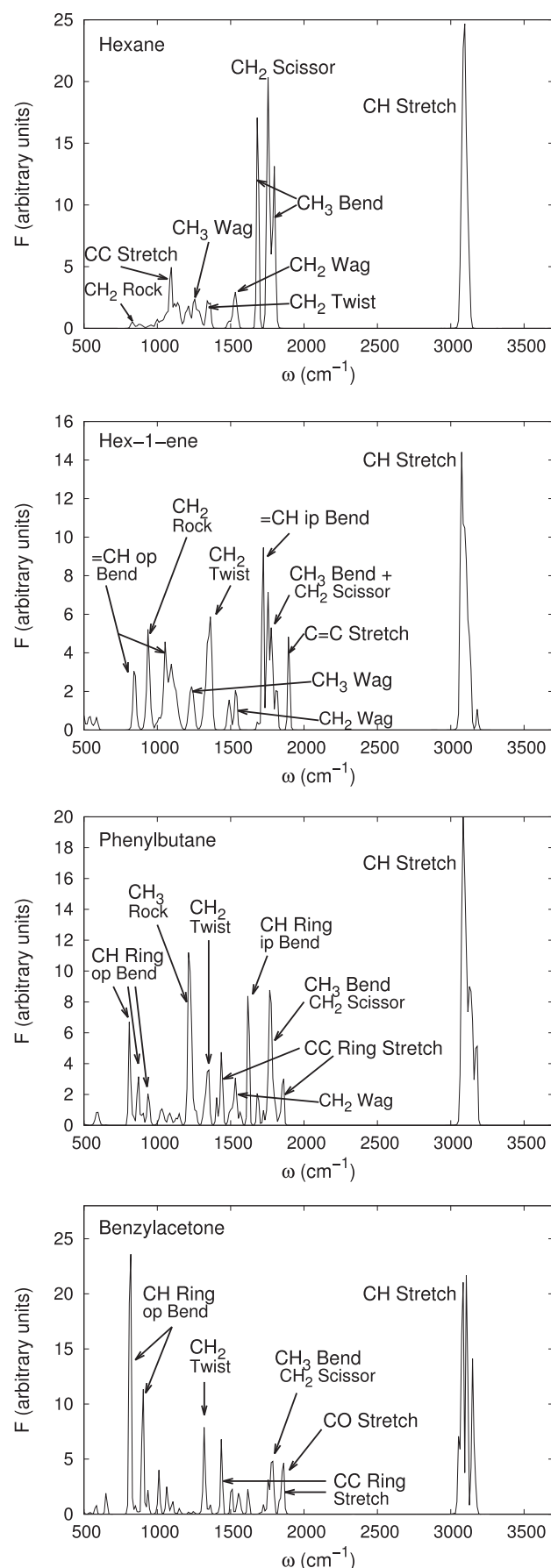
FIG. 6. Vibrational density of states for hexane, hex-1-ene, 1-phenylbutane, and benzylacetone. Peaks are identified by the dynamical matrix approach[49] and labeled accordingly.

shown in Table XIV which are again in good agreement with the targets.

B. Vibrational frequencies

In experimental infrared spectroscopy molecules can be identified by peaks arising from molecular vibrations which are characteristics of their functional groups. To test how our model compares with experiments we calculate the vibrational density of states (VDOS) for hexane, hex-1-ene, 1-phenylbutane, and benzylacetone. The VDOS shows the entire vibrational spectrum of the molecule with all normal modes of vibration producing peaks, unlike the infra-red spectrum in which only vibrational modes which are infra-red active are observed. The VDOS are calculated from molecular dynamics trajectories as the Fourier transform of the velocity autocorrelation function,

$$C_v(t) = \langle v_i(0)v_i(t)\rangle. \tag{12}$$

We have made molecular dynamics simulations of the molecules in the gas phase using an NVE ensemble at 20 K for 50 ps with a time step of 0.5 fs, recording atom positions and velocities every 2 fs. To identify the peaks of the vibrational density of states we have also calculated the vibrational frequencies using the dynamical matrix approach as was used in Sec. VI. Studying the eigenvectors of the dynamical matrix is sufficient for eigenmode and peak identification. In Fig. 6 we show the spectra and have labeled the relevant eigenmodes alongside the peaks. All of the molecules share the eigenmodes specific to alkanes while 1-phenylbutane and benzylacetone also share several modes specific to aromatic molecules. Note that the C–O stretching mode and C–C ring stretching modes of benzylacetone coincide into a single peak in our spectrum. We also show how these frequencies correspond with the characteristic experimental infra-red frequencies of different functional groups in Table XV. As was the case for the smaller molecules the CH_3 bending and CH_2 scissoring modes are shifted to higher frequencies compared to observation but all other frequencies correspond well to the experimental targets.

TABLE XV. Comparison of calculated vibrational frequencies of larger molecules using the TB model with experiment. Frequencies are calculated using the dynamical matrix approach. Frequencies are in cm^{-1}.

Functional group	Mode	TB	Expt.
Alkanes	CH stretch	3100–3050	3000–2850
(hexane)	CH_2 scissor	1800–1755	1470–1450
	CH_3 bend	1680	1370–1350
	CH_2 wag(long chain)	835	725–720
Alkenes	=CH stretch	3160–3050	3100–3000
(hex-1-ene)	C=C stretch	1890	1680–1640
	=CH bend	840–1050	650–1000
Aromatic	CH stretch	3160–3055	3100–3000
(phenylbutane)	CC ring stretch	1860	1600–1585
	CC ring stretch	1420	1500–1400
	CH oop bend	810–935	675–900
Ketone	CO stretch	1860	1705–1725
(benzylacetone)			

VIII. REACTIONS

In this section we illustrate the power of the TB model by demonstrating the dynamics of a chemical reaction. The reaction we model is enolization of benzylacetone to 1-phenyl-2-buten-3-ol. It is well known that enolization reactions take place with active solvent participation.[51] The rate limiting step is the removal of a proton from a methyl group. This proton is then transferred through the solvent by means of the Grothuss mechanism to the oxygen atom of the molecule. To model this reaction in quantitative detail would require the use of advanced sampling techniques but this is beyond our goal. To this end we bypass the rate limiting step by simply removing a proton from the benzylacetone molecule and placing it in the solvent.

We therefore take an equilibrated system of a cluster of 64 water molecules surrounding a molecule of benzylacetone in periodic boundary conditions. We then remove a proton from the carbon atom adjacent to the carbonyl group and place it in the surrounding water where it combines with a water

molecule to form H_3O^+. We then relax the system again in a process consisting of short molecular dynamics simulations and relaxations. We then perform molecular dynamics in an NVT ensemble at 300 K. To demonstrate that the TB theory reproduces the hydrogenation step we show in Fig. 7 a set of nine snapshots at selected timesteps. Between $t = 0$ and $t = 38$ fs a proton is seen in free flight between two water molecules, a snapshot from a proton transfer Grothuss mechanism. At $t = 41$ fs the water molecule at the center of the image is clearly a hydronium ion, H_3O^+, the proton at the top right of the ion itself being about to embark on a second Grothuss step heading towards the water molecule to its right. That second proton transfer occurs between $t = 44$ fs and $t = 50$ fs at which point the proton at the right of the resulting hydronium ion is in the process of transfer to the acceptor C=O group of the benzylacetone. At $t = 55$ fs the enolization is complete and all surrounding water molecules are neutral.

IX. DISCUSSION AND CONCLUSIONS

It is remarkable that a simple, orthogonal TB model has the capability to describe atomic and electronic structure, dynamics, solvation, and chemical reaction in a wide range of organic molecules. The rendering of single, double, and triple bonds is completely automatic as is the making and breaking of bonds, and indeed any crystal field effects. TB occupies a space between *ab initio* methods and classical force fields. In the present case, all parameterized interactions are short ranged, as they are in a force field, but the solution of the Schrödinger equation involving a Hamiltonian with only short ranged hopping leads to a density matrix that has an arbitrary range of interaction so that ultimately the model is sensitive to the environment of a molecule, or the long range forces in an extended system. We must add the caveat that we have focused upon *structure* and *dynamics*, and not *energetics*, in fitting to our training set of molecules. Therefore, further refinement and benchmarking will be needed in order to reproduce the enthalpy changes in chemical reactions.

One key to the success of the PITB is in the treatment of the electrostatics. The imposition of local charge neutrality is too severe an approximation for all except the simple hydrocarbons. As we have shown above, *point charge* electrostatics requiring a finite Hubbard U is quite adequate to describe the formation of the correct dipole moments and the prediction of structural and dynamic properties of all the hydrocarbons we have studied. This is clear since we use a zero crystal field parameter on the carbon atom (see Table I). In contrast, to describe oxygen it is essential to include anion polarizability at least to the dipole level.

Another key aspect of our modeling, as a keen reader will have noticed, is that although the theory behind the PITB is couched in terms of a density functional, we do not rely on local or gradient corrected density functional calculations to obtain our parameters. This is in contrast to the SCC-DFTB method.[21] In fact it is widely known that existing density functionals have great difficulty dealing with the decaying charge density in between molecules.[52] In our approach we side step this issue by fitting only to experimental data or
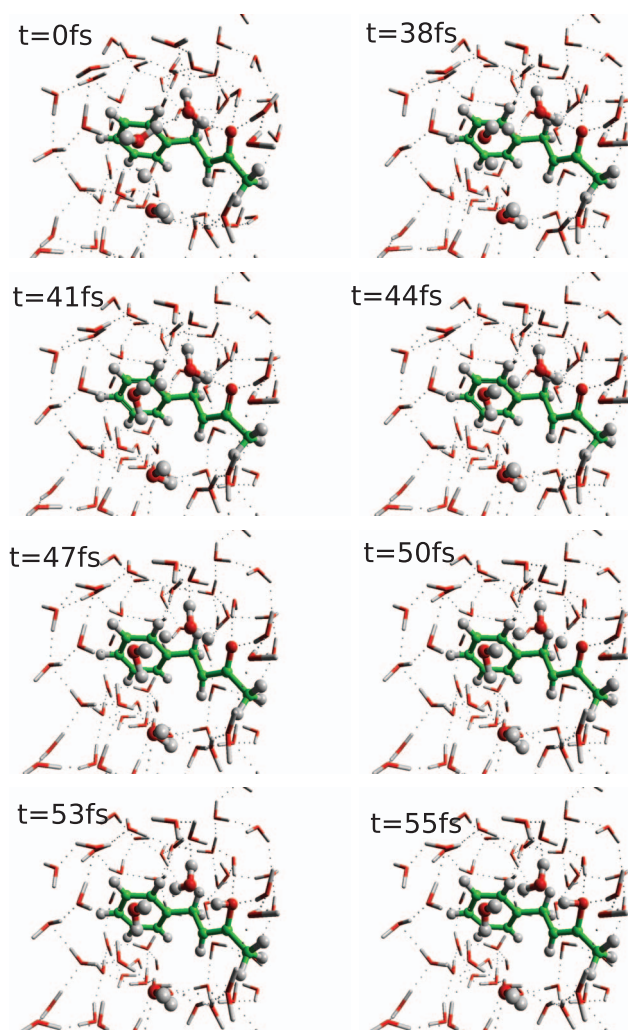
FIG. 7. A series of snapshots from a molecular dynamics simulation demonstrating the enolization reaction of benzylacetone to 1-phenyl-2-buten-3-ol in water. A proton is transferred between water molecules by the Grothuss mechanism, via hydronium ions, until it attaches to the acceptor C=O group of the benzylacetone.

to calculations using quantum chemistry methods that admit electron correlations. We should remark that it is not inconsistent to build a theory upon the DFT but then to claim we can side step its deficiencies. The point is that Eq. (1) is not restricted to the LDA; indeed one can imagine constructing an orbital dependent, self interaction corrected, "LDA+U"-type tight binding scheme which will hence go beyond LDA.[17,18] Our contention is that by fitting the model of Eq. (7) to experiment and CCSD(T) we are not in any way attempting to mimic the LDA and so are, in principle at any rate, not tied in to its shortcomings. The treatment of longer range tails is discussed in more detail in the following paper.[1]

In Papers II and III, we will show how the theory is extended to describe hydrogen bonding and processes taking place on crystal surfaces. In the future we expect there to be applications in many diverse fields; the explicit treatment of both charge and atomic migration will eventually allow applications in corrosion and electrochemistry.

ACKNOWLEDGMENTS

[1] A. Y. Lozovoi, T. J. Sheppard, D. L. Pashov, J. J. Kohanoff, and A. T. Paxton, J. Chem. Phys. **141**, 044504 (2014).
[2] A. Y. Lozovoi, D. L. Pashov, T. J. Sheppard, J. J. Kohanoff, and A. T. Paxton, J. Chem. Phys. **141**, 044505 (2014).
[3] A. T. Paxton and J. J. Kohanoff, J. Chem. Phys. **134**, 044130 (2011).
[4] J. J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods* (Cambridge University Press, Cambridge, 2006).
[5] J. A. Pople, D. P. Santry, and G. A. Segal, J. Chem. Phys. **43**, S129 (1965).
[6] J. A. Pople and G. A. Segal, J. Chem. Phys. **43**, S136 (1965).
[7] J. A. Pople, D. L. Beveridge, and P. A. Dobosh, J. Chem. Phys. **47**, 2026 (1967).
[8] R. C. Bingham, M. J. S. Dewar, and D. H. Lo, J. Am. Chem. Soc. **97**, 1294 (1975).
[9] M. J. S. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4899 (1977).
[10] W. Thiel, J. Am. Chem. Soc. **103**, 1413 (1981).
[11] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 (1985).
[12] J. J. P. Stewart, J. Comput. Chem. **10**, 209 (1989).
[13] J. J. P. Stewart, J. Mol. Model. **13**, 1173 (2007).
[14] J. J. P. Stewart, J. Mol. Model. **19**, 1 (1989).
[15] A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, J. Phys. C: Solid State Phys. **21**, 35 (1988).
[16] W. M. C. Foulkes and R. Haydock, Phys. Rev. B **39**, 12520 (1989).
[17] S. Sanna, B. Hourahine, T. Gallauner, and T. Frauenheim, J. Phys. Chem. **111**, 5665 (2007).
[18] A. T. Paxton and M. W. Finnis, Phys. Rev. B **77**, 024428 (2008).
[19] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[20] M. W. Finnis, A. T. Paxton, M. Methfessel, and M. van Schilfgaarde, Phys. Rev. Lett. **81**, 5149 (1998).
[21] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[22] M. W. Finnis, *Interatomic Forces in Condensed Matter* (Oxford University Press, Oxford, UK, 2003).
[23] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[24] J. Harris, Phys. Rev. B **31**, 1770 (1985).
[25] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
[26] W. A. Harrison, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond* (Freeman, San Francisco, 1980).
[27] A. T. Paxton, in *Multiscale Simulation Methods in Molecular Sciences*, NIC Series Vol. 42, edited by J. Grotendorst, N. Attig, S. Blügel, and D. Marx (Institute for Advanced Simulation, Forschungszentrum Julich, 2009), pp. 145–174, see http://webarchiv.fz-juelich.de/nic-series/volume42/volume42.html.
[28] A. P. Horsfield, A. M. Bratkovsky, M. Fearn, D. G. Pettifor, and M. Aoki, Phys. Rev. B **53**, 12694 (1996).
[29] A. P. Horsfield, A. M. Bratkovsky, D. G. Pettifor, and M. Aoki, Phys. Rev. B **53**, 1656 (1996).
[30] L. Goodwin, A. J. Skinner, and D. G. Pettifor, Europhys. Lett. **9**, 701 (1989).
[31] W. A. Harrison, Phys. Rev. B **31**, 2121 (1985).
[32] J. A. Majewski and P. Vogl, Phys. Rev. Lett. **57**, 1366 (1986).
[33] J. A. Majewski and P. Vogl, *The Structures of Binary Compounds* (North-Holland, Amsterdam, 1989), pp. 287–362.
[34] P. K. Schelling, N. Yu, and J. W. Halley, Phys. Rev. B **58**, 1279 (1998).
[35] S. Fabris, A. T. Paxton, and M. W. Finnis, Phys. Rev. B **61**, 6617 (2000).
[36] M. W. Finnis, A. T. Paxton, M. Methfessel, and M. van Schilfgaarde, in *Tight Binding Approach to Computational Materials Science*, MRS Symposium Proceedings Vol. 491, edited by P. E. A. Turchi, A. Gonis, and L. Colombo (Materials Research Society, Pittsburgh, PA, 1998), pp. 265–274.
[37] O. K. Andersen, in *The Electronic Structure of Complex Systems*, edited by P. Phariseau and W. M. Temmerman (Plenum, New York, 1984), p. 11.
[38] E. Bott, M. Methfessel, W. Krabs, and P. C. Schmidt, J. Math. Phys. **39**, 3393 (1998).
[39] A. T. Paxton, T. N. Todorov, and A. M. Elena, Chem. Phys. Lett. **483**, 154 (2009).
[40] H.-P. Schwefel, *Numerische Optimierung von Computer–Modellen mittels der Evolutionsstrategie*, Interdisciplinary Systems Research Vol. 26 (Birkhäuser, Basle, 1977).
[41] H.-P. Schwefel, *Evolution and Optimum Seeking (Sixth Generation Computer Technologies)* (Wiley, New York, 1995).
[42] A. P. Horsfield, P. D. Godwin, D. G. Pettifor, and A. P. Sutton, Phys. Rev. B **54**, 15773 (1996).
[43] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, Comput. Phys. Commun. **181**, 1477 (2010).
[44] A. J. Petro, J. Amer. Chem. Soc. **80**, 4230 (1958).
[45] P. Lazzeretti, R. Zanasi, and W. T. Raynes, J. Chem. Phys. **87**, 1681 (1987).
[46] R. McWeeney, *Coulson's Valence* (Oxford University Press, Oxford, 1979).
[47] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
[48] *Tables of Molecular Vibrational Frequencies Consolidated*, edited by T. Shimanouchi (National Bureau of Standards, Washington, D.C., 1972), Vol. I.
[49] G. P. Srivastava, *The Physics of Phonons* (Taylor and Francis, Abingdon, 1990).
[50] *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69, edited by P. J. Linstrom and W. G. Mallard (National Institute of Standards and Technology, 2011), see http://webbook.nist.gov/chemistry.
[51] C. S. Cucinotta, A. Ruini, A. Catellani, and A. Stirling, ChemPhysChem **7**, 1229 (2006).
[52] S. Grimme, WIREs Comput. Mol. Sci. **1**, 211 (2011).