Microscopic description of acid–base equilibrium

Emanuele Grifoni^{a,b}, Giovanni Maria Piccini^{a,b}, and Michele Parrinello^{a,b,c,1}

^aDepartment of Chemistry and Applied Biosciences, Eidgenössische Technische Hochschule (ETH) Zürich, CH-6900 Lugano, Ticino, Switzerland; ^bInstitute of Computational Science, Università della Svizzera Italiana, CH-6900 Lugano, Ticino, Switzerland; and ^cItalian Institute of Technology, 16163 Genova, Italy

Contributed by Michele Parrinello, January 10, 2019 (sent for review November 29, 2018; reviewed by Christoph Dellago and David E. Manolopoulos)

Acid–base reactions are ubiquitous in nature. Understanding their mechanisms is crucial in many fields, from biochemistry to industrial catalysis. Unfortunately, experiments give only limited information without much insight into the molecular behavior. Atomistic simulations could complement experiments and shed precious light on microscopic mechanisms. The large free-energy barriers connected to proton dissociation, however, make the use of enhanced sampling methods mandatory. Here we perform an ab initio molecular dynamics (MD) simulation and enhance sampling with the help of metadynamics. This has been made possible by the introduction of descriptors or collective variables (CVs) that are based on a conceptually different outlook on acid–base equilibria. We test successfully our approach on three different aqueous solutions of acetic acid, ammonia, and bicarbonate. These are representative of acid, basic, and amphoteric behavior.

acid–base | metadynamics | collective variables | enhanced sampling

Acid–base reactions play a key role in many branches of chemistry. Inorganic complexation reactions, protein folding, enzymatic processes, polymerization, catalytic reactions, and many other transformations in different areas are sensitive to changes in pH. Understanding the pH role in these reactions implies having control over their reactivity and kinetics.

The crucial importance of pH has stimulated the collection of a large amount of data on acid–base equilibria. These are typically measured in gas and condensed phases, using spectroscopic and potentiometric techniques. However, there are practical limitations to the accuracy of these methods especially in condensed phases (1). Furthermore it is very difficult to extract from experimental data a microscopic picture of the processes involved. It is thus not surprising that acid–base equilibrium has been the subject of intense theoretical activity (1–12).

The acidity of a chemical species in water can be expressed in terms of pK_a , the negative logarithm of the acid dissociation constant. There are two ways of calculating these values, one static and the other dynamic.

The most standard approach is the static one in which solution-phase free energies, and consequently pK_a s, are obtained by closing a Born–Haber cycle composed of gas phase and solvation free energies (1, 3–7). While extremely successful in many cases, the static approach has some limitations. A solvation model needs to be chosen and continuum solvent models have a limited accuracy. This is particularly true in systems like zeolites or proteins characterized by irregular cavities in which an implicit description of the solvent is challenging. Obviously from such an approach dynamic information cannot be gained. Furthermore, there can be competitive reactions that cannot be taken into account unless explicitly included in the model.

In principle these limitations could be lifted in a more dynamical approach based on molecular dynamics (MD) simulations in which the solvent molecules are treated explicitly. If one had unlimited computer time, such simulations would explore all possible pathways and assign the relative statistical weight to the different states. Unfortunately the presence of kinetic bottlenecks frustrates this possibility by trapping the system in metastable states, since different protonation states are

separated by large barriers. Furthermore in acid–base reactions chemical bonds are broken and formed. This requires the use of ab initio MD in which the interatomic forces are computed on the fly from electronic structure theories. This makes the calculation more expensive and reduces further the time scale that can be explored.

To overcome this difficulty, the use of enhanced sampling methods (13) that accelerate configurational space exploration becomes mandatory. A very popular class of enhanced sampling methods is based on the identification of the degrees of freedom that are involved in the slow reaction of interest. These degrees of freedom are usually referred to as collective variables (CVs) and are expressed as explicit functions of the atomic coordinates **R**. Sampling is then enhanced by adding a bias that is a function of the chosen CVs (14–16). Furthermore, designing a proper set of good CVs has also a deeper meaning. Successful CVs capture in a condensed way the physics of the problem, identify its slow degrees of freedom, and lead to a useful modelistic description of the process.

In standard chemical reactions, this is relatively simple since well-defined structures can be assigned to reactants and products (17–19). This is not the case for acid–base reactions in which a proton is added to or subtracted from the solute. Once this process has taken place, water ions (H^+ or/and OH^-) are solvated and their structure becomes elusive. In fact, water ions can rapidly diffuse in the medium via a Grotthuss mechanism (20). They become highly fluxional and the identity of the atoms taking part in their structure changes continuously. The nature of these species is thus difficult to capture in an explicit analytic function of **R**. However, given the relevance of acid–base reactions, many attempts have been made at defining these entities (8–12). Unfortunately these CVs have an ad hoc nature and, while successful in this or that case, cannot be generally applied.

Significance

Acid–base reactions are among the most important chemical processes. Yet we lack a simple way of describing this class of reactions as a function of the atomic coordinates. In fact, once dissolved in water, H^+ and its conjugate anion OH^- have a highly fluxional structure difficult to pin down. Here we solve this issue by taking the point of view of describing acid–base reactions as an equilibrium between the solute and the whole solvent. This allows identifying generally applicable descriptors. As a consequence it is now possible to perform quantitative enhanced sampling simulation of acid–base reaction in water and in other environments such as the zeolite cavities or at surfaces.

Author contributions: E.G., G.P., and M.P. designed research; E.G. and G.P. performed research; E.G., G.P., and M.P. analyzed data; and E.G., G.P., and M.P. wrote the paper.

Reviewers: C.D., University of Vienna; and D.E.M., University of Oxford.

The authors declare no conflict of interest.

Published under the PNAS license.

¹To whom correspondence should be addressed. Email: parrinello@phys.chem.ethz.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1819771116/-DCSupplemental.

Published online February 14, 2019.

We then group the acid–base sites in species. For instance, in the case of the simplest amino acid glycine in aqueous solution the number of species N^s will be equal to 3. All water oxygen atoms belong to one species; then one counts in another species the two carboxylic oxygen atoms, and finally one considers as the third species the nitrogen atom of the amino group.

In the spirit of this work we count the total excess or defects of protons associated to each species:

$$q_k = \sum_{i \in k} \delta_i. \quad [5]$$

This implies that we are not interested in the specific identity of the reacted site, but whether or not the k th species in its entirety has increased ($q_k = +1$), has decreased ($q_k = -1$), or has not changed its number of protons. If we consider a solute with only one reactive moiety, then each possible state of the system can be described by one of the three 2D vectors (0,0), (−1,1), or (1,−1).

In the general case each protonation state can be described by a vector $\vec{q} = (q_0, q_1, \dots, q_{N^s-1})$ with dimension equal to the number of inequivalent reactive sites, N^s . A more exhaustive explanation is provided in [SI Appendix](#).

For use in enhanced sampling these vectors need to be expressed as a scalar function $f = f(\vec{q})$ such that, for each physically relevant $\vec{q}$, f attains values able to distinguish the different overall protonation states. There are infinitely many ways of constructing a scalar from a vector. Possibly the simplest choice is to write $f(\vec{q}) = \vec{X} \cdot \vec{q}$ and, to distinguish between different protonation states, to choose $\vec{X} = (2^0, 2^1, 2^2, \dots, 2^{N^s-1})$.

This leads to the following definition for the CV that is used to describe the protonation state of the system,

$$s_p = \sum_{k=0}^{N^s-1} 2^k \cdot q_k, \quad [6]$$

where k are the indexes used to label the respective reactive site groups. In [SI Appendix](#) an example is worked out in detail. Of course the CV is made continuous by the use of W_i in the calculation of the δ_i needed to evaluate q_k in Eq. 5.

The second CV is a summation of distances between all acid–base sites multiplied by their partial charge δ_i ,

$$s_d = \sum_{i,m>i} -r_{im} \cdot \delta_i \cdot \delta_m, \quad [7]$$

where the indexes i and m run over all the acid–base sites belonging to different k groups, and r_{im} is the distance between the two atoms. In this way, just the acid–base pair that has exchanged a proton gives a contribution different from zero. Eq. 7 is valid only when one single-conjugate acid–base pair is present. However, due to the action of bias, it may occur occasionally that several acid–base pairs are formed. To avoid sampling these very unlikely events we apply a restraint on the number of pairs. Further details are provided in [SI Appendix](#).

Results

We have applied our method to three aqueous solutions of acetic acid, ammonia, and bicarbonate as representations of a weak acid, a weak base, and an amphoteric compound, respectively. The setups of all three simulations are identical except for the identity of the solvated molecules. This ensures that the outcome reflects the different chemistry of these three systems and that there is no bias due to the initial condition.

Each simulation of the systems was performed with Born–Oppenheimer MD simulations combined with well-tempered metadynamics (14, 26) using the CP2K package (27) patched with PLUMED 2 (28) and strongly constrained and appropriately normed functional (29) for the xc energy, E_{xc} . See [SI Appendix](#) for details.

In Fig. 3 we plot the free-energy surfaces (FESs) as a function of s_p and s_d . These FESs vividly reproduce the expected behavior. They all have a minimum at $s_p = 0$ that corresponds to the state in which no charges are present in the solvent. In the acetic acid FES (Fig. 3A) a second minimum close to $s_p = -1$ reflects its acid behavior. By contrast, the ammonia FES (Fig. 3B) shows a second minimum close to $s_p = 1$. The shapes of ammonia and acetic acid FESs are approximately related by a mirror symmetry

Fig. 3. (A–C) Free-energy surfaces along s_p and s_d of acetic acid (A), ammonia (B), and bicarbonate (C) in aqueous solution. Color bars indicate the free energy expressed in $\text{kJ} \cdot \text{mol}^{-1}$ units. The CV s_d is expressed in angstroms.

reflecting their contrasting behavior. Similarly the bicarbonate symmetric FES (Fig. 3C) mirrors its amphoteric character.

As the conjugate pair is formed s_d starts to assume positive values corresponding to the separation and diffusion of the conjugate pair. Compared with the undissociated state in which only $s_d = 0$ is allowed, states where a conjugate pair is present show an elongated shape of the basins along this variable. This is caused by the diffusive behavior of the hydronium and hydroxide ions in solution that makes accessible a continuum range of distances. Moreover, along this CV we can observe a barrier around 1.5 corresponding to the breaking of the covalent bond between the hydrogen atom and the acid–base site.

Conclusions

The general applicability of this method to systems with different natures is an important step made in their understanding and description. The scheme can be extended to include quantum nuclear effects with the use of path integral molecular dynamics

(30). This would be of quantitative significance since for instance pK_a values are affected by deuteration. Moreover, the absence of assumptions or impositions about reactive candidates or reaction paths allows extending this method to systems of increasing complexity which cannot be addressed with traditional methods. Examples of questions that can now be answered are tautomeric

equilibria in biochemical processes and acid behavior in zeolites and on the surface of oxides exposed to water.

ACKNOWLEDGMENTS. Calculations were carried out on the ETH Euler cluster and on the Mönch cluster at the Swiss National Supercomputing Center. This research was supported by the European Union Grant ERC-2014-AdG-670227/VARMET.

- Ho J, Coote ML (2011) First-principles prediction of acidities in the gas and solution phase. *Wiley Interdiscip Rev Comput Mol Sci* 1:649–660.
- Elstner M, Hobza P, Frauenheim T, Suhai S, Kaxiras E (2001) Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J Chem Phys* 114:5149–5155.
- Saracino GAA, Improta R, Barone V (2003) Absolute pK_a determination for carboxylic acids using density functional theory and the polarizable continuum model. *Chem Phys Lett* 373:411–415.
- Schüürmann G, Cossi M, Barone V, Tomasi J (1998) Prediction of the pK_a of carboxylic acids using the ab initio continuum-solvation model PCM-UAHF. *J Phys Chem A* 102:6706–6712.
- Ho J, Coote ML (2009) A universal approach for continuum solvent pK_a calculations: Are we there yet? *Theor Chem Acc* 125:3–21.
- Silva CO, da Silva EC, Nascimento MAC (2000) Ab initio calculations of absolute pK_a values in aqueous solution II. Aliphatic alcohols, thiols, and halogenated carboxylic acids. *J Phys Chem A* 104:2402–2409.
- Rebollar-Zepeda AM, Galano A (2016) Quantum mechanical based approaches for predicting pK_a values of carboxylic acids: Evaluating the performance of different strategies. *RSC Adv* 6:112057–112064.
- Davies JE, Doltsinis NL, Kirby AJ, Roussev CD, Sprik M (2002) Estimating pK_a values for pentaoxyphosphoranes. *J Am Chem Soc* 124:6594–6599.
- Park JM, Laio A, Iannuzzi M, Parrinello M (2006) Dissociation mechanism of acetic acid in water. *J Am Chem Soc* 128:11318–11319.
- Tummanapelli AK, Vasudevan S (2014) Dissociation constants of weak acids from ab initio molecular dynamics using metadynamics: Influence of the inductive effect and hydrogen bonding on pK_a values. *J Phys Chem B* 118:13651–13657.
- Ortiz APdA, Tiwari A, Puthenkalathil RC, Ensing B (2018) Advances in enhanced sampling along adaptive paths of collective variables. *J Chem Phys* 149:072320.
- Lee J-G, et al. (2006) Deprotonation of solvated formic acid: Car-Parrinello and metadynamics simulations. *J Phys Chem B* 110:2325–2331.
- Bernardi RC, Melo MCR, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim Biophys Acta* 1850:872–877.
- Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99:12562–12566.
- Valsson O, Parrinello M (2014) Variational approach to enhanced sampling and free energy calculations. *Phys Rev Lett* 113:1–5.
- Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* 23:187–199.
- Piccini G, McCarty JJ, Valsson O, Parrinello M (2017) Variational flooding study of a SN_2 reaction. *J Phys Chem Lett* 8:580–583.
- Mendels D, Piccini G, Parrinello M (2018) Collective variables from local fluctuations. *J Phys Chem Lett* 9:2776–2781.
- Piccini G, Polino D, Parrinello M (2017) Identifying slow molecular motions in complex chemical reactions. *J Phys Chem Lett* 8:4197–4200.
- Agmon N (1995) The Grotthuss mechanism. *Chem Phys Lett* 244:456–462.
- Wicke E, Eigen M, Ackermann Th (1954) Über den zustand des protons (hydroniumions) in wäßriger lösung [About the state of the proton (hydronium ion) in aqueous solution]. *Z Phys Chem* 1:340–364.
- Zundel G, Metzger H (1968) Energiebander der tunnelnden überschuss-protonen in flüssigen sauren. Eine IR-spektroskopische untersuchung der natur der gruppierungen $H_5O_2^+$ [Energy bands of tunneling excess protons in liquid acids. An IR spectroscopic study of the nature of the $H_5O_2^+$ species]. *Z Phys Chem* 58:225–245.
- Marx D, Tuckerman ME, Hutter J, Parrinello M (1999) The nature of the hydrated excess proton in water. *Nature* 397:601–604.
- Hulthe G, Stenhagen G, Wennerström O, Ottosson CH (1997) Water clusters studied by electrospray mass spectrometry. *J Chromatogr A* 777:155–165.
- Iyengar SS, et al. (2005) The properties of ion-water clusters. I. The protonated 21-water cluster. *J Chem Phys* 123:1–9.
- Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* 100:1–4.
- Vandevondele J, et al. (2005) Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput Phys Commun* 167:103–128.
- Brandenburg JG, Bates JE, Sun J, Perdew JP (2016) Benchmark tests of a strongly constrained semilocal functional with a long-range dispersion correction. *Phys Rev B* 94:17–19.
- Peng H, Yang Z-H, Sun J, Perdew JP (2016) Versatile van der Waals density functional based on a meta-generalized gradient approximation. *Phys Rev X* 6:041005.
- Parrinello M, Rahman A (1984) Study of an F center in molten KCl. *J Chem Phys* 80:860–867.