SoftwareX ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Software update

Q6: A comprehensive toolkit for empirical valence bond and related free energy calculations

Paul Bauer [a], Alexandre Barrozo [b], Miha Purg [a], Beat Anton Amrein [a], Mauricio Esguerra [a], Philippe Barrie Wilson [c], Dan Thomas Major [d], Johan Åqvist [a,*], Shina Caroline Lynn Kamerlin [a,*]

[a] Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden
[b] Department of Chemistry, University of Southern California, SGM 418, 3620 McClintock Ave., Los Angeles, CA 90089-1062, United States
[c] Leicester School of Pharmacy, De Montfort University, The Gateway, Leicester LE1 9BH, UK
[d] Department of Chemistry, Bar-Ilan University, Ramat-Gan 52900, Israel

ARTICLE INFO

Article history:
Received 17 August 2017
Received in revised form 16 November 2017
Accepted 3 December 2017

Keywords:
Free energy perturbation
Empirical valence bond
Molecular dynamics simulations
Path integral calculations

ABSTRACT

Atomistic simulations have become one of the main approaches to study the chemistry and dynamics of biomolecular systems in solution. Chemical modelling is a powerful way to understand biochemistry, with a number of different programs available to perform specialized calculations. We present here Q6, a new version of the Q software package, which is a generalized package for empirical valence bond, linear interaction energy, and other free energy calculations. In addition to general technical improvements, Q6 extends the reach of the EVB implementation to fast approximations of quantum effects, extended solvent descriptions and quick estimation of the contributions of individual residues to changes in the activation free energy of reactions.

Code metadata

Current code version	V. 6.0
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-17-00065
Legal code license	GNU General Public License version 2 (GPLv2)
Code versioning system used	Git
Software code languages, tools, and services used	Fortran, MPI 1.6.5
Compilation requirements, operating environments dependencies	GCC4.8 or newer, IFORT 13.0 or newer, PGI 12.1 or newer, OpenMPI 1.6.5 or newer or Intel MPI 3.0 or newer
Link to developer documentation	http://xray.bmc.uu.se/~aqwww/q/
Support email for questions	paul.bauer.q@gmail.com, qmoldyn@googlegroups.com

1. Motivation and significance

The empirical valence bond approach (EVB) [1,2] is an empirically-based hybrid valence bond/molecular mechanics (EVB/MM) approach, that uses classical force fields within a quantum mechanical framework to describe (bio)chemical reactivity. The theoretical details of this approach have been described in great detail elsewhere [1–4], and it is particularly powerful as a tool to model enzyme reaction mechanisms, as the use of classical force fields facilitates very fast calculations compared to density functional theory or *ab initio*-based QM/MM approaches, thus allowing for extensive configurational sampling. At the same time, well-parameterized force fields carry a tremendous amount of physical information, allowing for biochemical reactivity to be modelled in a meaningful way. This approach has been applied to a wide range of biological systems, from enzymes such as triosephosphate isomerase [5–7], to GTP hydrolysis on the ribosome [8] (for further examples of the application of EVB to both chemical and biochemical problems, see *e.g.* Refs. [9–22]).

The EVB approach is implemented into a number of software packages, including its original implementation into Warshel's MOLARIS simulation package [23], as well as AMBER [24], CHARMM [25] and Tinker [25,26]. The software presented here, Q6,

DOI of original article: http://dx.doi.org/10.1016/S1093-3263(99)00012-1.
* Corresponding authors.
E-mail addresses: aqvist@xray.bmc.uu.se (J. Åqvist), kamerlin@icm.uu.se (S.C.L. Kamerlin).

https://doi.org/10.1016/j.softx.2017.12.001
2352-7110/

is a general purpose simulation package for standard molecular dynamics simulations, empirical valence bond and other free energy calculations, including the linear interaction energy (LIE) approach for calculating ligand binding free energies [27]. We recently implemented an interface to Q, CADEE, which allows for the *in silico* directed evolution of enzymes using the EVB approach [28]. In addition, QGui provides a graphical interface for high-throughput setup and analysis of free energy calculations using Q [29]. Q6 is an overhaul of the original Q simulation package [30], with the addition of new solvents, force fields and thermostats, as well as, critically, the implementation of a path integral method for calculating kinetic isotope effects based on EVB calculations, as well as a residue deletion approach to rapidly screen the contribution of individual amino acid side chains to calculated activation free energies, which we believe is a useful screening tool for quantitative enzyme design studies.

2. Software description

The Q simulation package is primarily an engine for molecular dynamics simulations and free energy calculations, which can be performed using either the OPLS-AA (2001 and 2005) [31,32], AMBER (95 and 14SB) [33–35], CHARMM (v.22) [36,37], or GROMOS (87 and 96) [38–40] force fields. These simulation can be performed in both NVT and NPT ensembles, using either spherical [41] or periodic boundary conditions, and either the leapfrog [42] or velocity-verlet integrators [43]. Its distinguishing features are the implementation of the EVB approach, and the possibility to perform both standard free energy perturbation (FEP) and linear interaction energy (LIE) calculations, as originally developed by Åqvist [27]. The latter approaches provide a powerful tool for evaluating ligand binding free energies, making them very valuable in drug design [44–46], and their implementation in Q has been applied to a number of challenging systems, including most recently ligand binding free energies to GPCRs [47–49], potassium channels [50], and ribosomes [51–54]. In addition, novel computational schemes have been developed for computational mutagenesis [48,55,56], that provide more physical descriptions of ligand binding and successfully target the convergence problems that plague calculations of binding free energies upon residue substitution. Finally, the EVB approach is a very powerful tool for rapid screening for artificial enzyme design [28].

For a complete theoretical background for the EVB approach, we refer the reader elsewhere [3,10,57–61]. Briefly, this approach uses classical force fields and free energy perturbation/umbrella sampling [2] (FEP/US) to interpolate between different reacting (valence bond) states by calculating a set of potential energies for each state for the given system coordinate. The potential energies for each system coordinate are saved during the molecular dynamics simulation, and are analysed in a post-processing step to obtain the overall free energy profile. This approach is extremely flexible in that once a given reference state (usually either the background reaction in vacuum or aqueous solution, or, for example, the wild-type enzyme compared to a series of mutant enzymes) has been parameterized, it is possible to move the same parameter set to a host of different environments, unchanged, without the need for further parametrization (see Ref. [62] for discussion of the phase-independence of the EVB off-diagonal term). In addition, while EVB simulations can be performed employing commonly used periodic boundary conditions [16], both MOLARIS [23,63] and Q [30] also allow for the use of spherical boundary conditions to describe the reacting system. In these software packages, the spherical boundary conditions are described in a stochastic multi-layer model in which the inner region of the sphere (*i.e.* the reacting atoms and all atoms within the typically inner 85% of the sphere) is completely flexible, all atoms outside the explicit sphere are restrained to their initial coordinates using a very strong (200 kcal mol^{-1} Å^{-2} or higher) restraint, and all atoms in the outer 15% of the sphere are restrained using a weaker (10 kcal mol^{-1} Å^{-2}) restraint to ease the transition between the fully flexible and fully restrained regions (Figure S1). The sphere is then described using the surface constrained all-atom solvent (SCAAS) model [41], with long-range interactions described using the local reaction field approach [64]. Such a truncation of the full reacting system is valuable as it can lead to substantial gains in computational speed compared to treating the full system explicitly, without loss of accuracy, provided that a large enough sphere (typically $>$ 20 Å radius) is used.

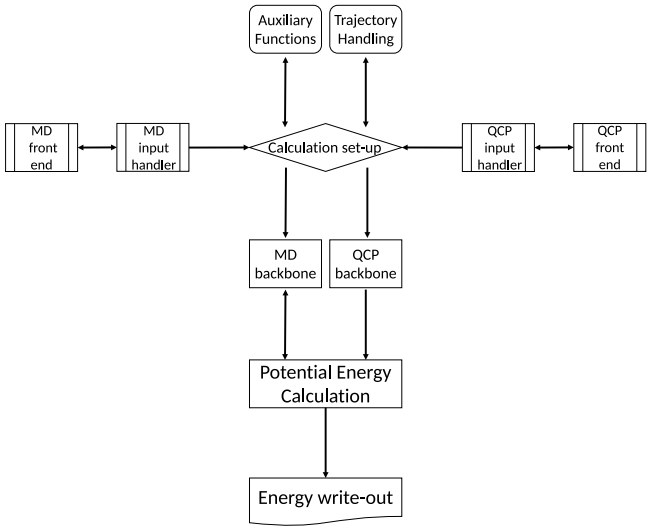

Fig. 1. Simplified overview of the Q6 code organization. Note that the post-processing for the free energy calculations (empirical valence bond and linear interaction energy calculations) is handled in a separate external module, Qcalc6. In addition, the FEP/US is handled in the same module as the potential energy calculation.

3. Software architecture

Q6 is based on several modules, combined into a number of individual programs that the user interacts with (see also the Q6 manual, available on the Q6 website: http://xray.bmc.uu.se/~aqwww/q/). These include the main dynamics routine, **Qdyn6**, the structure preparation routine, **Qprep6**, the free energy analysis module, **Qfep6**, the new quantum classical path tool, **Qpi6**, and the trajectory analysis tool, **Qcalc6**. The energy and force calculation routines are split up from other parts of the code, to allow their use in all the relevant parts of the program, with the same being true of the routines involved in the calculation of geometric properties. The individual sub-modules are implemented as FORTRAN90 modules, that are linked into the main programs as needed, as shown in Fig. 1.

4. Software functionalities

As mentioned in the Introduction, the main utilization of Q is for molecular dynamics simulations and free energy calculations, with particular focus on FEP/US, EVB and LIE calculations. In Q6: (1) path integral calculations can now be performed in order to calculate quantum corrections to classical free energies; (2) the residue or atom based contributions of different functional groups to a full free energy profile can be calculated using an electrostatic deletion approach; (3) simulations can be performed in a number of organic solvents, in addition to the already implemented TIP3P [65] and SPC [66] water models.

Please cite this article in press as: Bauer P., et al. Q6: A comprehensive toolkit for empirical valence bond and related free energy calculations. SoftwareX (2018), https://doi.org/10.1016/j.softx.2017.12.001.

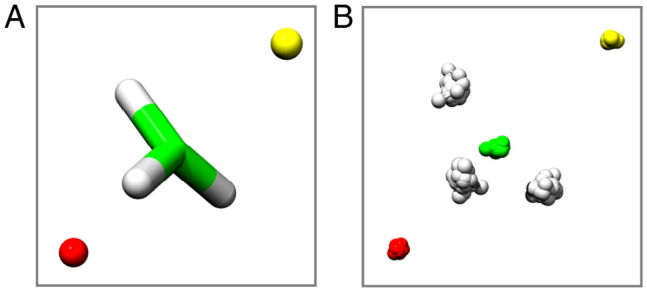

Fig. 2. (A) Schematic representation of the conversion between (A) the classical particle and (B) the ring polymer. Carbon atoms are shown in green, oxygen atoms in red, sulphur atoms in yellow and hydrogen atoms and ring polymers in white. Figure extracted from the methyl transfer reference reaction catalyzed by COMT studied as a test case (Figure S2), at the transition state. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Source: Figure prepared in Chimera [98].

4.1. Quantum classical path calculations

In Q6, we have now implemented the Bisection Quantum Classical Path (BQCP) approach, developed by Gao and Major [67–69], based on the original QCP methodology developed by Warshel [70,71] and the bisection sampling algorithm [72,73]. In these approaches, the nuclear quantum corrections to the classical energies of a system are calculated using a path integral formulation, with the classical particles replaced by a set of pseudoparticles (*i.e.* beads) that generate a ring polymer, centred on the classical coordinate of the particle (*i.e.* centroid, c) [69,73–75], as visualized in Fig. 2. The QM correction to a classical mechanical (CM) partition function is given by:

$$\frac{Q_{QM}}{Q_{CM}} = \left\langle \left\langle e^{-\frac{1}{k_B TP}\left(\sum_{j=1}^{P}(U_j - U_c)\right)} \right\rangle_{FP} \right\rangle_{U_c} \tag{1}$$

where P is the number of beads, U_j is the potential at the position of bead j, and U_c is the potential at the centroid (*i.e.* classical) position. The inner ensemble average is performed over the free-particle (FP) bead distribution, while the external ensemble average is performed over the classical potential. The BQCP description, which is described in great technical detail in Ref. [69] brings the additional advantage of sampling the FP distribution by an exact Monte Carlo sampling scheme. If the path-integral calculation is performed for different isotopes, one can then readily compute isotope effects for chemical reactions [67].

4.2. Group exclusion calculations

The estimation of the individual contributions of different amino acid side chains (or functional groups in a supramolecular complex) to the calculated activation free energy barriers is important in the search for possible mutation and engineering sites in macromolecules, as well as aiding protein engineering and understanding enzyme evolution [13,14,76–79].

These contributions can be rapidly screened through a "group exclusion" approach, by on-the-fly removal of the electrostatic and van der Waals contributions of the functional groups of interest to the total activation barrier in real time during the calculation, while still advancing the whole system on the same trajectory as without having performed this deletion. This protocol models a virtual deletion of the residue in question, similarly to performing conservative mutations of the residue in question (see also *e.g.* Refs. [10,80–83] among others for related approaches). In this approach, the total potential energy is first calculated as usual, shown in Eq. (2):

$$\begin{aligned} E_{total} = {} & E_{qq,el} + E_{qq,vdw} + E_{qp,el} + E_{qp,vdw} \\ & + E_{pp,el} + E_{pp,vdw} + E_{bond} + E_{angle} + E_{torsion} \end{aligned} \tag{2}$$

The interaction energies of the groups under investigation are then subtracted from the total energy E_{total} as shown in Eq. (3):

$$E_{exclude} = E_{Total} - E_{qq,el}^{q-exc} - E_{qq,vdw}^{q-exc} - E_{qp,el}^{q-exc} - E_{qp,vdw}^{q-exc} \tag{3}$$

Here, "q" indicates the "Q_{atoms}" (*i.e.* the atoms designated as reacting atoms), and "p" designates the surrounding protein and/or solvent atoms (or in the case of a supramolecular complex surrounding functional groups). The subscripts qq, qp and pp denote the Q_{atoms}–Q_{atoms}, Q_{atoms}–protein and protein–protein interactions respectively, and vdw, el, bond, angle and torsion stand for the van der Waals, electrostatic and geometric contributions to the total energy (E_{total}), respectively. Once E_{total} has been calculated, the energy contributions from the residue/functional group of interest are deleted from the total energy (designated as "exclude/exc" in Eq. (3)). The interaction energies are then removed at each of the reacting states. The main advantage of this approach as a screening tool is that only one trajectory needs to be followed for the calculation, as all atoms still move on the normal potential energy surface created by the chosen force field. Additionally, sampling is not limited to a small number of structures, extracted from a completed trajectory, but is rather performed on all configurations generated during the simulation. Its main disadvantage, however, is that large changes in the electrostatics that would result in major changes to the observed trajectories are not included, that these contributions are not necessarily additive [84], and that the system always follows the same path irrespectively of the nature of the interactions that have been excluded. This makes it less useful for predicting mutations that simply alter the shape of the binding pocket, but nevertheless excellent as a tool for predicting distant mutations that can alter the electrostatics at the active site.

4.3. Other technical developments

In addition to the major changes outlined above, the following technical improvements have also been made in Q6 compared to previous Q versions [30]. We have implemented: (1) the Langevin and Nosé-Hoover thermostats, based on the details provided in Tuckerman [85], (2) the LINCS constraint algorithm [86], and (3) the ability to describe more polyatomic solvent models, such as methanol, ethanol, chloroform and dichloromethane, based on the parameters provided by the Virtual Chemistry project [87,88].

One of the key goals during the development of the new software version has been to maintain full backward compatibility to calculations prepared for previous versions of the program, which has led us to be more conservative in setting new features as defaults, and the changes to default methods have therefore been minimal. However, the following minor changes have been made:

- Topology generation is now more lenient in placing solvent atoms next to other generated (non-crystallographic) solvent molecules, to allow tighter packing of complex solvents.
- Long range electrostatics are now always explicitly treated using the local reaction field (LRF) approach, unless explicitly disabled.
- When using periodic boundary conditions, interactions with all atoms not found in the local list are now described using LRF, not just those with a larger nonbonded cut-off.
- Constraints (SHAKE by default) are now always applied to solvent hydrogen atoms, unless explicitly disabled, and can be applied to both or either solvent and solute atoms in a more refined manner.

Please cite this article in press as: Bauer P., et al. Q6: A comprehensive toolkit for empirical valence bond and related free energy calculations. SoftwareX (2018), https://doi.org/10.1016/j.softx.2017.12.001.

- Temperature control is more refined to both allow fine-grained coupling groups, and to be more helpful in terms of the printed information provided during a calculation.
- Energy analysis can be performed on any size of data, and is not restricted by the program itself but rather only by the amount of memory available on the machine.

A number of intrinsic changes have also been made, which include a rework of the handling of mathematical functions and vector operations through the use of math libraries, the reworking of non-bonded interactions to allow all arbitrary molecule sizes, removal of most constraints on the maximum system size that can be used (except for the calculation speed), and the enforcement of sanity checking of the input for free energy calculations in order to enforce proper handling of bonded interactions. We note that, in contrast to for example MOLARIS [23,63,89], Q6 is currently unable to perform implicit pK_a calculations, and thus we couple to external servers. However, in light of our recent development of CADEE [28], this is a functionality we expect to be included in subsequent Q releases, as it will greatly assist in the automated enzyme design process.

Finally, performance benchmarks for Q6 are provided in Fig. 3. We note that we have not presented benchmarks for the residue deletion approach, as this calculation only contains the recalculation of a small number of interactions during the usual EVB/FEP/MD calculation. It can also be done in concert with the calculation of QCP energies as post processing, and has only limited performance impact there as well. The type of calculations typically performed with Q6, such as EVB simulations of a large number of enzyme variants, depend not on extreme parallelization of single runs but the possibility to run a large number of simulations in parallel. As can be seen from this figure, unsurprisingly, the program only scales reasonably over a larger number of cores when the system size is increased by using a very large sphere size (40 Å), but the absolute performance takes a large hit at the same time. Typical simulations with Q use 20–25 Å and sometimes 30 Å sphere sizes, and for this Q6 is fully adequate for running a large number of simulations using a few cores per trajectory (as is for example done in CADEE [28]).

5. Illustrative examples

5.1. Quantum Classical Path Calculations

In order to validate the correct implementation of the BQCP approach [69], we have calculated the kinetic isotope effects (KIE) for the E2 elimination reaction of 2-phenylethylbromide in ethanol [90] (allowing us also to test our new solvent models), as well as for the transmethylation reaction catalyzed by catechol *O*-methyltransferase [91,92] (see Figures S2 and S3 for the associated reaction mechanisms and Figure S4 for an overview of the COMT active site and the reacting region). In both cases, we employed the EVB approach to obtain the classical free energy profiles for the reaction, and the calculated the BQCP correction during post-processing. The EVB parameters and simulation details can be found in the Supporting Information.

We were able to successfully reproduce the kinetic isotope effects for both reactions with our implementation of the BQCP approach, with the corresponding results given in Table 1. The values show that the approach can describe both the normal and inverse isotope effects in the respective reactions. We note that our calculated values for the COMT reaction underestimate the experimental activation free energy. This is most likely caused by the fact that we calibrated our EVB reference potential to the values previously established by Warshel for a similar substrate, and not dopamine itself [93]. However, as our aim in this validation was to reproduce the kinetic isotope effects rather than the absolute activation free energies, this was not deemed critical to the evaluation of the BQCP implementation. We would like to point out in addition that the BQCP approach we have implemented into Q6 is a specialized variation of Warshel's original QCP approach [71]; the original approach is based on *a posteriori* QCP sampling that circumvents convergence problems with the stiff necklace and works well also for systems with relatively high values of H/D kinetic isotope effects [94], and is accessible through the MOLARIS software package [23,63].

Table 1
Results of the QCP calculations performed on the COMT methyl transfer reaction and E2 elimination in ethanol[a].

System	$\Delta G^{\ddagger}_{exp}$	$\Delta G^{\ddagger}_{light}$	$\Delta G^{\ddagger}_{heavy}$	KIE_{Calc}	KIE_{Exp}
COMT	18.5	16.0 ± 0.2	15.9 ± 0.2	0.80	0.79
E2 elimination	20.0	20.0 ± 0.1	21.0 ± 0.1	5.5	7.1

[a] Calculated activation free energies for both the light ($\Delta G^{\ddagger}_{light}$) and heavy ($\Delta G^{\ddagger}_{heavy}$) isotopes are shown in kcal mol^{-1}. The experimental activation free energies and kinetic isotope effects were obtained from Refs. [92] and [90] respectively. All simulation details can be found in the Supporting Information. The calculations show that the BQCP approach is able to describe both the normal and inverse kinetic isotope effects observed experimentally in these two systems, and works both for simulations in water and in organic solvent.

5.2. Group exclusion calculations

The second main addition to Q6, *i.e.* calculation of selected residue contributions to the potential energy of the reactive region over a complete reaction, has been tested on the epoxide ring opening reaction catalyzed by *Solanum tuberosum* epoxide hydrolase 1 [13,95]. In this system, two catalytically important active site tyrosine residues, Y154 and Y235, are an ideal target to estimate their contribution to the free energy barrier of the reaction, as their truncation to phenylalanine has been studied both experimentally [95] and computationally [13]. These active site tyrosines form, together, an oxyanion hole that is critical for stabilizing the substrate (see the Supporting Information); their conservative truncation to phenylalanine removes their ability to form this oxyanion hole. As this reaction is not just a multi-step process, but also both enantio- and regioselective [13,95], which is computationally expensive to model, we have used as our model system only the ring opening of the preferred (*R,R*)-enantiomer of trans-stilbene oxide [13], with nucleophilic attack of the active site nucleophile on the preferred carbon atom. This allows us also to compare directly to our previous work [13], where we used the linear response approximation [96,97], to estimate the contribution of individual residues to catalysis. We have considered here the effect of excluding either tyrosine side chain individually, as well as the effect of excluding both at once. We note that simply excluding these tyrosines from the free energy calculation is not biochemically equivalent to mutating them to phenylalanine; however, both provide an approximation of the contribution of these residues to the calculated activation free energies. A comparison of experimental and calculated results is shown in Table 2. It can be seen that the methods give a good estimate for the energy difference in all cases, showing that this kind of approach can be useful to identify possible residues of interest in a studied enzyme.

5.3. Testing new solvents

The improvement in the handling of solvent molecules made it possible to describe more complicated solvents than three-point water models such as SPC [66] or TIP3P [65]. The solvents dichloromethane, chloroform, methanol and ethanol have been

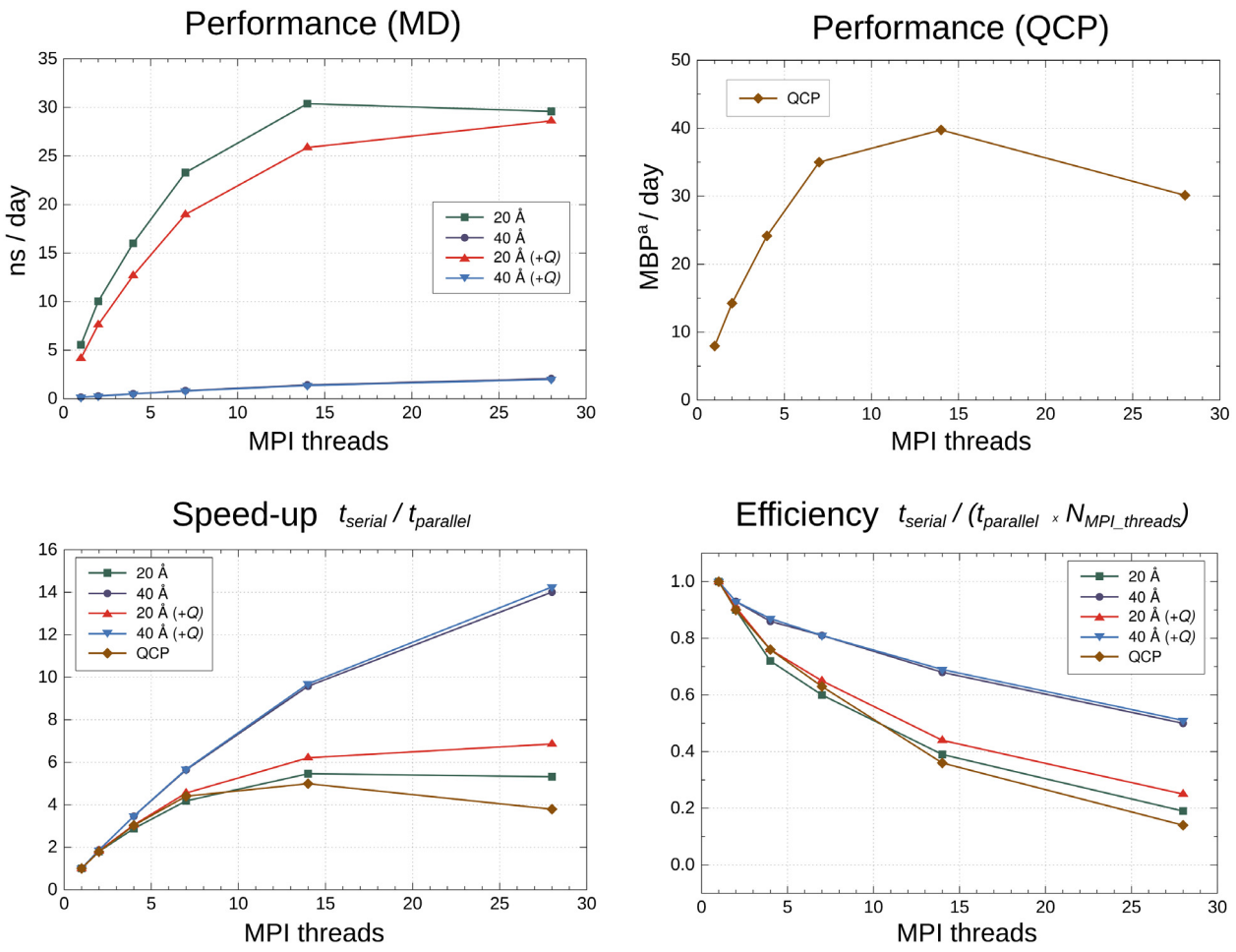

Fig. 3. Q6 performance benchmarks. Shown here are the (top left) performance of classical MD simulations, free energy (EVB simulations, indicated as +Q on the figure to indicate the inclusion of "Q-atoms", which have an infinite cut-off for nonbonded interactions), and (top right) QCP calculations. The corresponding speed-up (bottom left) and efficiency (bottom right) are also shown here. The model reaction is the COMT reaction described in the main text, and all simulations were performed with spherical boundary conditions, considering two different sphere radii (20 and 40 Å). The MD was performed at 300 K using the leapfrog algorithm [42], the Berendsen thermostat [99], a 1 fs step size, and a 100 fs bath coupling time. A 10 Å cutoff was used with a charge-group scheme for all non-bonded interactions with the exception of those involving "Q-atoms", where a 99 Å (in principle infinite) cutoff was used. SHAKE [100] was applied to all solvent atoms, and long-range electrostatics were described using the local reaction field (LRF) approach [64]. The system with the 20 Å sphere contains 2375 solute atoms and 424 water molecules in the mobile regions in the simulations (not including excluded atoms outside the sphere), and the system with the 40 Å sphere contains 3423 solute atoms and 7948 water molecules. Where a Q-region was used in free energy calculations, it contained 72 atoms. MBP denotes million bead-positions, and is calculated as: N(classical snapshot) × N(beads) × N(QCP sampling time). All simulations were performed on the supercomputer cluster Kebnekaise at HPC2N, using Intel Xeon E5-2690v4 processors. Q6 was compiled with Intel Parallel Studio XE 2017.1-132.

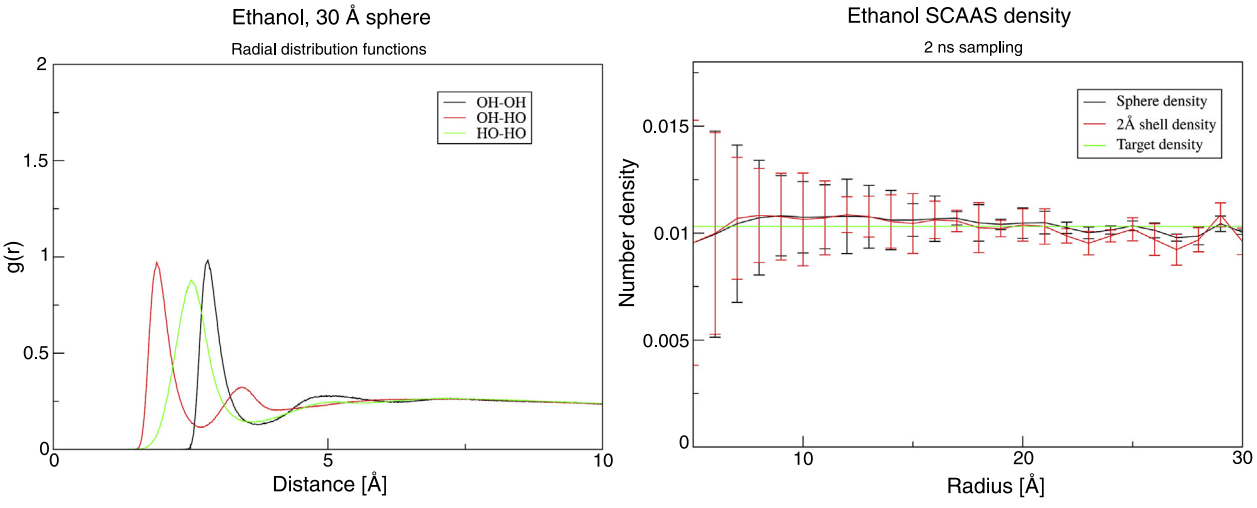

Fig. 4. Plotted radial distribution functions (A) and solvent densities at different radii in the solvent sphere (B) for the model system of ethanol as a solvent, in a 30 Å solvent sphere. Additional results for other solvents can be found in the Supporting information.

Please cite this article in press as: Bauer P., et al. Q6: A comprehensive toolkit for empirical valence bond and related free energy calculations. SoftwareX (2018), https://doi.org/10.1016/j.softx.2017.12.001.

Table 2
A comparison of calculated and experimental activation barriers for the ring opening reaction of the *(R, R)*-enantiomer of trans-stilbene oxide by wild-type (WT) StEH1 and the Y154F, Y235F and Y154F/Y235F StEH1 variants[a].

	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})	$\Delta\Delta G^{\ddagger}_{Exp}$	$\Delta\Delta G^{\ddagger}_{Calc}$	$\Delta\Delta G^{\ddagger}_{Est}$
WT	23 ± 2	2400 ± 210	–	–	–
Y154F	0.08 ± 0.008	4 ± 0.5	4.9	1.5	0.22
Y235F	0.07 ± 0.01	1.3 ± 0.2	4.9	3.7	1.86
Y154F/Y235F	n.d.	n.d.	n.d.	NA	2.09

[a] The experimental data was obtained from Ref. [95]. The explicitly calculated values were obtained from Ref. [13] and estimated by our group exclusion approach, as described in the Supporting Information. Shown here are both the absolute and relative activation free energies ($\Delta G^{\ddagger}$ and $\Delta\Delta G^{\ddagger}$), with the subscripts Exp, Calc and Est denoting experimental, explicitly calculated and estimated values, with the estimated values from group contributions [97] and the calculated values obtained using the residue deletion approach, respectively. All energies in this table are shown in kcal mol^{-1}. NA – not explicitly calculated in Ref. [13]. n.d. – no experimentally observable rate.

tested during simulations using spherical boundary conditions, to confirm the correct behaviour concerning radial distribution functions and solvent densities at different positions of the solvent sphere. The results of 1 ns simulations at 300 K are shown in Fig. 4.

6. Conclusions and impact

In order to calculate the free energies for chemical and biological processes with both high accuracy and high efficiency, it is necessary to have access to approaches that can perform extensive computational sampling to obtain chemical free energies, while still providing a physical description of the processes involved. The EVB approach is an extremely powerful computational tool that fulfils both these criteria, and has been extensively applied to a host of chemical and biochemical problems. Q was originally developed as a tool for performing empirical valence bond and other free energy calculations [30]. In Q6, we extend the capabilities of Q to allow for the study of nuclear quantum effects and perform rapid screening of the effect of different functional groups on the calculated activation free energies, while also improving the efficiency of the program and providing a range of technical upgrades, as outlined above. We believe Q6 will be a valuable tool for any researcher interested in studying (bio)chemical reactivity.

Acknowledgements

The European Research Council provided funding for this research under European Community's Seventh Framework Programme (FP7/2007–2013). Code development was supported by Swedish Research Council (VR) funding (2014-3688, 2014-2118 and 2015-04928) to JÅ and SCLK. SCLK is a Wallenberg Academy Fellow. The computational work was performed through the allocation of computing time by the Swedish National Infrastructure for Computing (SNIC 2016/34-27) on the Abisko and Kebnekaise clusters. We thank Åke Sandgren from HPC2N (Umeå) for help with improvements to the Q6 code.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.softx.2017.12.001.

References

[1] Warshel A, Weiss RM. An empirical valence bond approach for comparing reactions in solutions and in enzymes. J Am Chem Soc 1980;102:6218–26.
[2] Warshel A. Computer modelling of chemical reactions in enzymes and solutions. Wiley Interscience; 1991.
[3] Kamerlin SCL, Warshel A. The empirical valence bond model: Theory and applications. WIREs Comput Mol Sci 2011;1:30–45.
[4] Shurki A, Derat E, Barrozo A, Kamerlin SCL. How valence bond theory can help you understand your (bio)chemical reaction. Chem Soc Rev 2015;44:1037–52.
[5] Åqvist J, Fothergill M. Computer simulation of the triosephosphate isomerase catalyzed reaction. J Biol Chem 1996;26:10010–6.
[6] Åqvist J. Cold adaptation of triosephosphate isomerase. Biochemistry 2017;56:4169–76.
[7] Kulkarni YS, Liao Q, Petrović D, Krüger DM, Strodel B, Amyes TL, Richard JP, Kamerlin SCL. Enzyme architecture: Modeling the operation of a hydrophobic clamp in catalysis by triosephosphate isomerase. J Am Chem Soc 2017;139:10514–25.
[8] Åqvist J, Kamerlin SCL. Exceptionally large entropy contributions enable the high rates of GTP hydrolysis on the ribosome. Sci Rep 2015;5:15817.
[9] Vendrell O, Gelabert R, Moreno M, Lluch JM. A potential energy function for heterogenous proton-wires. ground and photoactive states of the proton-wire in the green fluorescent protein. J Chem Theory Comput 2008;4:1138–50.
[10] Frushicheva MP, Cao J, Chu ZT, Warshel A. Exploring challenges in rational enzyme design by simulating the catlaysis in artificial kemp eliminase. Proc Natl Acad Sci USA 2010;107:16869–74.
[11] Bellucci MA, Coker DF. Molecular dynamics of excited state intramolecular proton transfer: 3-hydroxyflavone in solution. J Chem Phys 2012;136:194505.
[12] Prasad Bora R, Plotnikov NV, Lameira J, Warshel A. Quantitative exploration of the molecular origin of the activation of GTPase. Proc Natl Acad Sci USA 2013;110:20509–14.
[13] Amrein BA, Bauer P, Duarte F, Janfalk Carlsson Å, Naworyta A, Mowbray SL, Widersten M, Kamerlin SCL. Expanding the catalytic triad in epoxide hydrolases and related enzymes. ACS Catal 2015;5:5702–13.
[14] Barrozo A, Duarte F, Bauer P, Carvalho ATP, Kamerlin SCL. Cooperative electrostatic interactions drive functional evolution in the alkaline phosphatase superfamily. J Am Chem Soc 2015;137:9061–76.
[15] Schopf P, Mills MJ, Warshel A. The entropic contributions in vitamin B12 enzymes still reflect the electrostatic paradigm. Proc Natl Acad Sci USA 2015;112:4328–33.
[16] Dunning GT, Glowacki DR, Preston TJ, Greaves SJ, Greetham GM, Clark IP, Towrie M, Harvey JN, Orr-Ewing AJ. Reaction dymamics. Vibrational relaxation and microsolvation of DF after F-atom reactions in polar solvents. Science 2015;347:530–3.
[17] Glowacki DR, Orr-Ewing AJ, Harvey JN. Non-equilibrium reaction and relaxation dynamics in a strongly interacting explicit solvent: F + CD_3CN treated with a parallel multi-state EVB model. J Chem Phys 2015;143:044120.
[18] Isaksen GV,Åqvist J, Brandsdal BO. Enzyme surface rigidty tunes the temperature dependence of catalytic rates. Proc Natl Acad Sci USA 2016;113:7822–7.
[19] Isaksen GV, Hopmann KH,Åqvist J, Brandsdal BO. Computer simulations reveal substrate specificity of glycosidic bond cleavage in native and mutant human purine nucleoside phosphorylase. Biochemistry 2016;55:2153–62.
[20] Kazemi M, Himo F,Åqvist J. Enzyme catalysis by entropy without Circe effect. Proc Natl Acad Sci USA 2016;113:2406–11.
[21] Zhan S, Mårtensson D, Purg M, Kamerlin SCL, Ahlquist MSG. Capturing the role of explicit solvent in the dimerization of RuV (bda) water oxidation catalysts. Angew Chem Int Ed 2017;56:6962–5.
[22] Duboué-Dijon E, Pluhařová E, Domin D, Sen K, Fogarty AC, Chéron N, Laage D. Coupled valence-bond state molecular dynamics description of an enzyme-catalyzed reaction in a non-aqueous organic solvent. J Phys Chem B 2017;121:7027–41.
[23] Warshel A, Chu Z, Villa J, Strajbl M, Schutz C, Shurki A, Vicatos S, Chakrabarty S, Plotnikov N, Schopf P. Molaris-XG, v 9.11. Los Angeles: University of Southern California; 2012.
[24] Case DA, Bertz RM, Cerutti DS, Cheatham III TE, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Homeyer N, Izadi S, Janowski P, Kaus J, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Luchko T, Luo R, Madej B, Mermelstein D, Merz KM, Monard G, Nguyen H, Nguyen HT, Omelyan I, Onufriev A, Roe DR, Roitberg A, Sagui C, Simmerling CL, Botello-Smith WM, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, Kollman PA. AMBER 2016. San Francisco: University of California; 2016.
[25] Glowacki DR, Orr-Ewing AJ, Harvey JN. A parallel multistate framework for atomistic non-equilibrium reaction dynamics of solutes in strongly interacting organic solvents, 2014, arXiv preprint arXiv:1412.4180.
[26] Carpenter BK, Harvey JN, Glowacki DR. Prediction of enhanced solvent-induced enantioselectivity for a ring opening with a bifurcating reaction path. Phys Chem Chem Phys 2015;17:8372–81.
[27] Åqvist J, Marelius J. The linear interaction energy method for predicting ligand binding free energies. Comb Chem High Throughput Screen 2001;4:613–26.
[28] Amrein BA, Steffen-Munsberg F, Szeler I, Purg M, Kulkarni Y, Kamerlin SCL. CADEE: Computer-Aided Directed Evolution of Enzymes. IUCrJ 2017;4:50–64.

[29] Isaksen GV, Andberg TA,Åqvist J, Brandsdal BO. Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems. J Mol Graph Mod 2015;60:15–23.
[30] Marelius J, Kolmodin K, Feierberg I,Åqvist J. Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. J Mol Graph Mod 1998;16:213–25.
[31] Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 2001;105:6474–87.
[32] Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved peptide and protein torsional energetics with the OPLS-AA force field. J Chem Theory Comput 2015;11:3499–509.
[33] Cornell WD, Cieplak P, Bayly C, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic-molecules. J Am Chem Soc 1995;117:5179–97.
[34] Hornak V, Abel R, Okur A, Strockbine BA, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Prot Struct Funct Bioinform 2006;65:712–5.
[35] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB : Improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 2015;11:3696–713.
[36] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1981;4:187–217.
[37] Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 2004;25:1400–15.
[38] van Gunsteren WF, Berendsen HJC. Groningen molecular simulation (GROMOS) library manual. Groningen: BIOMOS b.v.; 1987.
[39] van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: The GROMOS96 manual and user guide. Zürich, Groningen: Hochshulverlag AG an der ETH Zürich and BIOMOS b.v.; 1996.
[40] Scott WRP, Huenenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krueger P, van Gunsteren WF. The GROMOS biomolecular simulation program package. J Phys Chem A 1999;103:3596–607.
[41] King G, Warshel A. A surface constrained all-atom solvent model for effective simulations of polar solutions. J Chem Phys 1989;91:3647–61.
[42] Verlet L. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.. Phys Rev 1967;159:98–103.
[43] Swope WC, Andersen HC, Berens PH, Wilson KR. A computer-simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Applications to small water clusters. J Chem Phys 1982;76:637–49.
[44] Hansson T, Marelius J,Åqvist J. Ligand binding affinity prediction by linear interaction energy methods. J Comput Aided Mol Des 1998;12:27–35.
[45] Åqvist J, Medina C, Samuelsson J-E. A new method for predicting binding affinities in computer-aided drug design. Prot Eng Des Sel 1994;Yes:385–91.
[46] Gutiérrez-de Terán H,Åqvist J. Linear interaction energy: Method and applications in drug design. New York: Springer; 2012.
[47] Xu B, Fällmar H, Boukharta L, Pruner J, Lundell I, Mohell N, Gutiérrez-de Terán H, Larhammar D. Mutagenesis and computational modeling of human G protein-coupled receptor Y2 for neuropeptide Y and peptide YY. Biochemistry 2013;52:7987–98.
[48] Keränen H,Åqvist J, Gutiérrez-de Terán H. Free energy calculations of A_{2A} adenosine receptor mutation effects on agonist binding. Chem Commun 2014;51:3522–5.
[49] Gutiérrez-de Terán H, Keränen H, Azuaje J, Rodríguez D,Åqvist J, Sotelo E. Computer-aided design of GPCR ligands. New York: Springer; 2015.
[50] Boukharta L, Keränen H, Weinzinger AS, Wallin G, de Groot BL,Åqvist J. Computer simulations of structure–activity relationships for hERG channel blockers. Biochemistry 2011;50:6146–56.
[51] Sund J, Andér M,Åqvist J. Principles of stop-codon reading on the ribosome. Nature 2010;465:947–50.
[52] Lind C, Sund J,Åqvist J. Codon-reading specificities of mitochondrial release factors and translation termination at non-standard stop codons. Nat Commun 2013;4:2940.
[53] Satpati P, Sund J,Åqvist J. Structure-based energetics of mRNA decoding on the ribosome. Biochemistry 2014;53:1714–22.
[54] Lind C,Åqvist J. Principles of start codon recognition in eukaryotic translation initiation. Nucl Acids Res 2016;44:8425–32.
[55] Boukharta L, Gutiérrez-de Terán H,Åqvist J. Computational prediction of alanine scanning and ligand binding energetics in G-protein coupled receptors.. PLoS Comput Biol 2014;10:e1003585.
[56] Shamsudin Khan Y, Gutiérrez-de Terán H, Boukharta L,Åqvist J. Toward an optimal docking and free energy calculation scheme in ligand design with application to COX-1 inhibitors. J Chem Inf Mod 2014;54:1488–99.
[57] Warshel A, Weiss RM. An empirical valence bond approach for comparing reactions in solutions and in enzymes. J Am Chem Soc 1980;102:6218–26.
[58] Hwang J-K, Warshel A. Microscopic examination of free-energy relationships for electron transfer in polar solvents. J Am Chem Soc 1987;109:715–20.
[59] Luzhkov V,Åqvist J. Computer simulation of phenyl ester cleavage by β-cyclodextrin in solution. J Am Chem Soc 1998;120:6131–7.
[60] Roca M, Vardi-Kilshtain A, Warshel A. Towards accurate screening in computer-aided enzyme design. Biochemsitry 2009;48:3046–56.
[61] Fuxreiter M, Mones L. The role of reorganization energy in rational enzyme design. Curr Opin Chem Biol 2014;21:34–41.
[62] Hong GY, Rosta E, Warshel A. Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. J Phys Chem B 2006;110:19570–4.
[63] Warshel A, Creighton S. MOLARIS. In: van Gunsteren WF, Weiner PK, editors. Computer simulation of biomolecular systems. Leiden: ESCOM; 1989.
[64] Lee FS, Warshel A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. J Chem Phys 1992;97:3100–7.
[65] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–35.
[66] Berendsen HJC, Postma J, van Gunsteren WF, Hermans J. Intermolecular forces. In: Proceedings of the fourteenth Jerusalem symposium on quantum chemistry and biochemistry. 1981. p. 13-16.
[67] Major DT, Gao J. Implementation of the bisection sampling method in path-integral simulations. J Mol Graph Mod 2005;24:121–7.
[68] Major DT, Garcia-Viloca M, Gao J. Path-integral simulations of proton transfer reactions in aqueous solution using a combined QM/MM potential. J Chem Theory Comput 2006;2:236–45.
[69] Major DT, Gao JL. An integrated path integral and free-energy perturbation-umbrella sampling method for computing kinetic isotope effects of chemical reactions in solution and in enzymes. J Chem Theory Comput 2007;3:949–60.
[70] Hwang JK, Chu ZT, Yadav A, Warshel A. Simulations of quantum mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions. J Phys Chem 1991;95:8445–8.
[71] Hwang JK, Warshel A. A quantized classical path approach for calculations of quantum-mechanical rate constants. J Phys Chem 1993;97:10053–8.
[72] Pollock EL, Ceperley DM. Simulation of quantum many-body systems by path-integral methods. Phys Rev B 1984;30:2555–68.
[73] Ceperley DM. Path-integrals in the theory of condensed helium. Rev Modern Phys 1995;67:279–355.
[74] Feynman RP, Hibbs AR. Quantum mechanics and path integrals. McGraw-Hill; 1965.
[75] Gao JL, Wong KY, Major DT. Combined QM/MM and path integral simulations of kinetic isotope effects in the proton transfer reaction between nitroethane and acetate ion in water. J Comput Chem 2008;29:514–22.
[76] Nes WD, Jayasimha P, Song ZH. Yeast sterol C24-methyltransferase: Role of highly conserved tyrosine-81 in catalytic competence studied by site-directed mutagenesis and thermodynamic analysis. Arch Biochem Biophys 2008;477:313–23.
[77] Badieyan S, Bevan DR, Zhang CM. Probing the active site chemistry of beta-glucosidases along the hydrolysis reaction pathway. Biochemistry 2012;51:8907–18.
[78] Mujika JI, Lopez X, Mulholland AJ. Mechanism of C-terminal intein cleavage in protein splicing from QM/MM molecular dynamics simulations. Org Biomol Chem 2012;10:1207–18.
[79] Blaha-Nelson D, Krüger DM, Szeler K, Ben-David M, Kamerlin SCL. Active site hydrophobicity and the convergent evolution of paraoxonase activity in structurally divergent enzymes: The case of serum paraoxonase 1. J Am Chem Soc 2017;139:1155–67.
[80] Bash PA, Field MJ, Davenport RC, Petsko GA, Ringe D, Karplus M. Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. Biochemistry 1991;30:5826–32.
[81] Gerber PR, Mark AE, van Gunsteren WF. An approximate but efficient method to calculate free-energy trends by computer-simulation - Application to dihydrofolate-reductase inhibitor complexes. J Comput Aid Mol Des 1993;7:305–23.
[82] Liu HY, Mark AE, van Gunsteren WF. Estimating the relative free energy of different molecular states with respect to a single reference state. J Phys Chem 1996;100:9485–94.
[83] Ito M, Brinck T. Novel approach for identifying key residues in enzymatic reactions: Proton abstraction in ketosteroid isomerase. J Phys Chem B 2014;118:13050–8.
[84] Mark AE, van Gunsteren WF. Decomposition of the free-energy of a system in terms of specific interactions - Implications for theoretical and experimental studies. J Mol Biol 1994;240:167–76.

Please cite this article in press as: Bauer P., et al. Q6: A comprehensive toolkit for empirical valence bond and related free energy calculations. SoftwareX (2018), https://doi.org/10.1016/j.softx.2017.12.001.

[85] Tuckerman M. Statistical mechanics: Theory and molecular simulation. Oxford University Press; 2010.
[86] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–72.
[87] van der Spoel D, van Maaren PJ, Caleman C. GROMACS molecule & liquid database. Bioinformatics 2012;28:752–3.
[88] Caleman C, van Maaren PJ, Hong MY, Hub JS, Costa LT, van der Spoel D. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. J Chem Theory Comput 2012;8:61–74.
[89] Sham Y, Chu ZT, Warshel A. Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and microscopic approaches.. J Phys Chem B 1997;101:4458–72.
[90] Saunders WH, Edison DH. Mechanisms of elimination reactions. 4. Deuterium isotope effects in E2 reactions of some 2-phenylethyl derivatives. J Am Chem Soc 1960;82:138–42.
[91] Zhang JY, Kulik HJ, Martinez TJ, Klinman JP. Mediation of donor–acceptor distance in an enzymatic methyl transfer reaction. Proc Natl Acad Sci USA 2015;112:7954–9.
[92] Zhang JY, Klinman JP. Enzymatic methyl transfer: Role of an active site residue in generating active site compaction that correlates with catalytic efficiency. J Am Chem Soc 2011;133:17134–7.
[93] Jindal G, Warshel A. Exploring the dependence of QM/MM calculations of enzyme catalysis on the size of the QM region. J Phys Chem B 2016;120:9913–21.
[94] Mavri J, Matute RA, Chu ZT, Vianello R. Path integral simulation of the H/D kinetic isotope effect in monoamine oxidase B catalyzed decomposition of dopamine. J Phys Chem B 2016;120:3488–92.
[95] Elfstrom LT, Widersten M. Catalysis of potato epoxide hydrolase, StEH1. Biochem. J. 2005;390:633–40.
[96] Lee FS, Chu ZT, Bolger MB, Warshel A. Calculations of antibody antigen interactions—microscopic and semimicroscopic evaluation of the free-energies of binding of phosphorylcholine analogs to Mcpc603. Protein Eng 1992;5:215–28.
[97] Muegge I, Tao H, Warshel A. A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding. Protein Eng 1997;10:1363–72.
[98] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12.
[99] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular-dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–90.
[100] Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. J Comput Phys 1977;23:327–41.

Please cite this article in press as: Bauer P., et al. Q6: A comprehensive toolkit for empirical valence bond and related free energy calculations. SoftwareX (2018), https://doi.org/10.1016/j.softx.2017.12.001.