QM/MM simulations for methyl transfer in solution and catalysed by COMT: ensemble-averaging of kinetic isotope effects†

Natalia Kanaan, J. Javier Ruiz Pernía and Ian H. Williams*

Sampling of structures from QM/MM molecular dynamics reveals distinct families of reactant-state conformers and yields kinetic isotope effects for reactions in enzyme active sites and in solution, averaged over thermal fluctuations of the environment, that allows meaningful comparison of computed with experimental values.

The experimental observation of a much more inverse secondary α-deuterium kinetic isotope effect (2° α-D KIE) $k(CH_3)/k(CD_3)$ for methyl transfer (1) (Scheme 1) to catecholate anion from S-adenosyl-L-methionine (SAM), catalyzed by catechol-O-methyltransferase (COMT) at 37 °C in water, than for uncatalyzed methyl transfer (2) (Scheme 1) to methoxide anion from S-methyldibenzothiophenium (SMBT) cation at 25 °C in methanol[1] was originally interpreted as evidence of compression by the enzyme causing a tighter S_N2 transition state than for the non-enzymic reaction.[2] Hybrid quantum-mechanics/molecular mechanics (QM/MM) calculations reproduced the trend in the 2° α-D_3 KIEs without evidence of compression,[3] but were based upon single structures for the reactant complex (RC) and transition state (TS) of the enzymic and non-enzymic reactions. We now report ensemble-averaged α-D_3 and α-^{13}C KIEs for reactions (1) and (2) (Scheme 1), as computed by means of AM1/TIP3P/OPLS potentials using DYNAMO.[4]

KIEs are frequently employed to deduce mechanistic details and to infer TS structure that may then inform design of TS analogues as potential inhibitors.[5] However, until now computational modelling of KIEs and TS structure has neglected the effects of thermal fluctuations of the protein and/or solvent environment that preclude a simple treatment involving a single RC configuration and a single TS configuration.
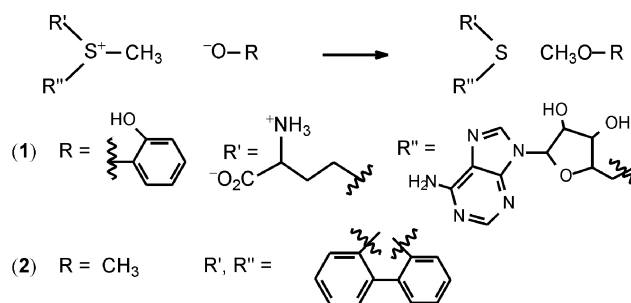
The key features of our procedure for ensemble averaging of KIEs are as follow. Initially we perform molecular dynamics (MD) simulations with umbrella sampling in overlapping windows along a distinguished coordinate for the reaction at the desired temperature and compute a potential of mean force (PMF). At the free-energy minimum (RC) and maximum (TS) of the PMF, we then compute extended QM/MM MD trajectories; the TS is subject to a constraint to prevent it collapsing to the RC or product. From the RC and TS trajectories, we then extract numerous independent configurations, and for each "snapshot", we perform QM/MM relaxation of the

Department of Chemistry, University of Bath, Bath, UK BA2 7AY. E-mail: i.h.williams@bath.ac.uk; Fax: 44 1225 386231; Tel: 44 1225 386625
† Electronic supplementary information (ESI) available: Details of the averaged KIE method. See DOI: 10.1039/b814212b

geometry to either a local minimum or saddle-point and at each stationary structure we compute the QM/MM Hessian for a subset of selected atoms. The sets of RC and TS snapshots are completely independent of each other; we do not consider pathways (e.g. intrinsic reaction coordinates) that connect particular TS structures with particular RC structures. Finally, we calculate semi-classical KIEs as averages over all possible independent combinations of RCs and TSs (see ESI†).

Our treatment of the enzyme-catalysed methyl transfer, reaction (1) (Scheme 1), is very similar to that described in our previous work.[3] The QM region comprised 63 atoms (catecholate + SAM) and the full QM/MM system for the MD simulations contained 17 162 atoms including 4614 water molecules in a cubic box of side 55.8 Å; the flexible region for local relaxation contained 2610 atoms; and the Hessian included the QM atoms + the active-site Mg^{2+} cation (64 atoms).

The uncatalysed methyl transfer in solution, reaction (2) (Scheme 1), has not been studied previously by computational simulation. The QM region comprised 30 atoms (methoxide and SMBT) and the full QM/MM system for the MD simulations contained 5184 atoms, including 859 methanol molecules in a cubic box of side 38.6 Å. The flexible region for local relaxation was variable in size, being defined to include for each structure all solvent molecules (between 7 and 12 in number) having any atom within 2.8 Å of any solute atom. The Hessians included the same number (72–102) of atoms, corresponding to at least the first solvation shell, in order to satisfy the cut-off rule necessary to obtain converged values of isotope effects for solvated molecules.[6]

In each case the difference $d(S\cdots C) - d(C\cdots O)$ between the lengths of the breaking and making bonds in the methyl transfer was employed as the distinguished coordinate in the PMF determination; umbrella sampling was performed with 20 ps (and 10 ps of prior relaxation) of NVT MD with periodic boundary conditions in each of 60 overlapping windows. The extended MD trajectories for RC and TS were 1 ns in length,

Scheme 1

Table 1 Averaged kinetic isotope effects for COMT-catalysed reaction (1) and reference reaction (2) in methanol (number of RC,TS pairs shown in parentheses)

	α-D$_3$ KIE	^{13}C KIE
Reaction (1) (Scheme 1)		
Expt	0.83 ± 0.05	1.09 ± 0.05
Calc (10000)	0.82 ± 0.05	1.056 ± 0.002
Reaction (2) (Scheme 1)		
Expt	0.974 ± 0.016	1.083 ± 0.015
Calc (1666)	0.99 ± 0.16	1.044 ± 0.018

and 100 snapshots were taken at 10 ps intervals in order to ensure that all structures were independent. All simulations were performed at the temperature of the original experiment work: 310 K for reaction (1) and 298 K for reaction (2). Not all snapshots converged to stationary structures for which a Hessian could be evaluated. Each Hessian for a subset of N_s atoms was subjected to a projection to ensure that 6 zero frequencies are obtained for the translational and rotational modes and that the $3N_s - 6$ vibrational frequencies satisfy the Teller–Redlich product rule for isotopomeric species.[6] In practice we found that treating all $3N_s$ degrees of freedom of the subset as harmonic vibrations (whose frequencies did not satisfy the Teller–Redlich product rule) yielded KIEs with essentially the same values as those presented in Table 1.[7]

The calculated KIEs (Table 1) for the enzymic reaction (1) (Scheme 1) are mean values averaged over 10 000 combinations of RC and TS. The calculated 2° α-D$_3$ KIE (0.82 ± 0.05) appears to agree well with the experimental value (0.83 ± 0.05),[1a] while the calculated α-^{13}C KIE (1.056 ± 0.002) is somewhat lower than the experimental value (1.09 ± 0.05).[1a] Note that the standard deviations for the calculated KIEs reflect the uncertainty of averaging over precisely computed individual TS-theoretical rate-constant ratios for a limited sample of 100 RC and 100 TS configurations spanning a distribution of structures determined by the model potential, whereas the experimental errors arise from uncertainties in rate-constant determinations for (effectively) complete sampling of all thermally accessible RC and TS configurations which may also span a distribution of structures.

The key bond lengths and angles for enzymic TS configurations have small standard deviations from their mean values (Table 2) but there is considerable variation in the O···C and Mg···OH distances in the RCs. For a given RC, the KIEs may be averaged over all the TSs with which it may be paired. A

Table 2 Averaged geometries for RC and TS of enzymic reaction (1) and non-enzymic reaction (2)

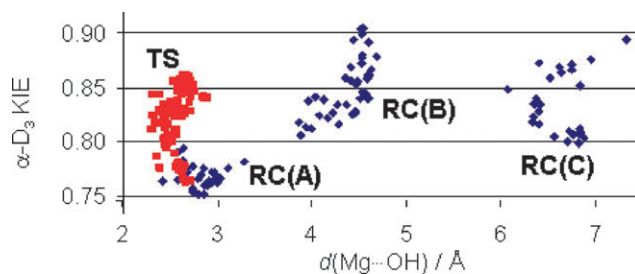|	RC	TS
Reaction (1) (Scheme 1)		
No. of structures	100	100
d(O···C)/Å	4.8 ± 1.3	2.060 ± 0.019
d(C···S)/Å	1.816 ± 0.003	2.107 ± 0.012
∠(O···C···S)/°	137 ± 8	171 ± 3
d(Mg···OH)/Å	4.6 ± 1.5	2.56 ± 0.13
Reaction (2) (Scheme 1)		
No. of structures	39	53
d(O···C)/Å	6.6 ± 2.5	2.18 ± 0.04
d(C···S)/Å	1.805 ± 0.004	2.00 ± 0.04
∠(O···C···S)/°	128 ± 30	175.4 ± 1.4

Fig. 1 Average α-D$_3$ KIE variation arising from groups of RC (blue) and TS (red) conformers for enzymic reaction (1) (Scheme 1).

plot (Fig. 1, blue points) of <α-D$_3$ KIE> averaged in this way against the RC(Mg···OH) distance (for coordination of the non-nucleophilic catechol OH group to the active-site Mg^{2+} of COMT) shows three distinct groups of RC configurations in the 1 ns MD trajectory, which have different average KIEs of 0.77 ± 0.01, 0.85 ± 0.03 and 0.84 ± 0.03, respectively, for Mg···OH = 2.82 ± 0.18, 4.36 ± 0.23 and 6.6 ± 0.3 Å, corresponding to different orientations of the catecholate substrate. A similar plot (Fig. 1, red points) of KIEs averaged over all RCs for each TS shows only one group, indicating that all TSs belong to a single family of conformers. The α-^{13}C KIEs for these three groups of reactant-state conformers are, respectively, 1.0536 ± 0.0008, 1.0561 ± 0.0011 and 1.0573 ± 0.0009, which do not differ very significantly.

A typical structure from each of the three groups of RC conformers is shown in Fig. 2(a, b and c). Structures **A**, **B** and **C** all show the Mg^{2+} octahedrally coordinated to Asp141, Asp169, Asn170, Glu199 and water molecule Wat51. The sixth coordination site is occupied by the non-nucleophilic OH of catechol in **A**, but, in **B** and **C**, is also occupied by the other oxygen of Asp141. Structures **B** and **C** have longer Mg···OH distances and **C** shows the catechol moiety rotated away from SAM. The three structures are representative of populations sampled during the 1ns QM/MM MD trajectory for RC at 310 K which shows the catechol ligand dissociating from Mg^{2+} within the active site of COMT. Curiously, analogous MD trajectories at both 298 and 310 K initiated from the single RC used in our earlier study appear to be stable with a structure similar to **A** in this work, in which the catechol OH is tightly coordinated to magnesium. Fig. 2(d) shows a typical TS structure in which the catechol OH makes an even shorter contact with Mg^{2+} than in **A**.

In Lau and Bruice's model[8] for the COMT ternary complex, Mg^{2+} interacts with both the catechol OH and O$^-$, whereas Kuhn and Kollman[9] found a structure with the cation coordinated by Asp169, Asn170, Wat51, the catechol O$^-$ and doubly coordinated by Asp141; their structure remained stable in molecular-mechanical MD at 300 K for at least 500 ps. Although both those studies started from the same X-ray structural coordinates (for an inhibitor complex, 1vid) as we have used, our initial RC was obtained by following a PMF down from a TS similar to that shown in Fig. 2(d): this structure allows the nucleophilic O$^-$ of catechol to attack the positive charged methyl group of SAM while (in principle) leaving the coordination shell of Mg^{2+} unaffected. There may be five distinct families of RC conformers for the ternary
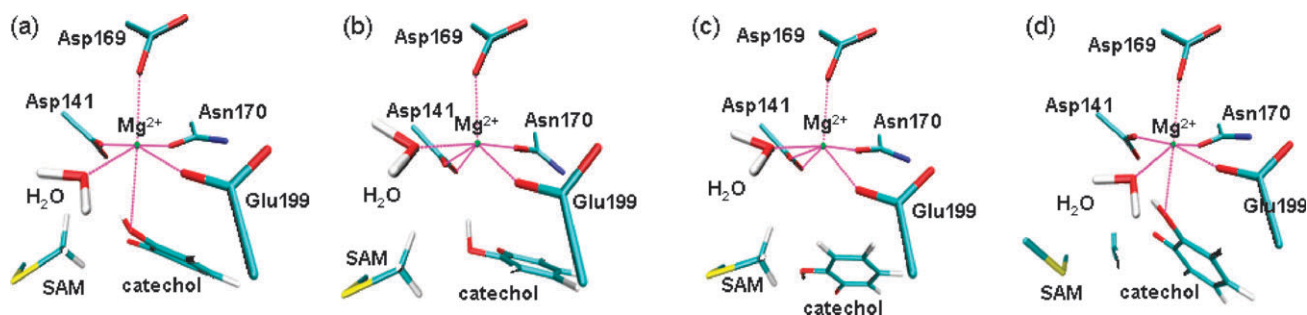
Fig. 2 Representative structures (a, b, c) of groups of RC conformers **A**, **B** and **C**, respectively, and (d) TS for reaction (1).

complex of COMT with SAM and a catechol substrate! There is no clear experimental evidence to indicate which of these corresponds to the reactant state for the $k(CH_3)/k(CD_3)$ determinations for the enzymic methyl transfer, or indeed whether the observed state includes fractional contributions from more than one of these families of conformers under the experimental conditions.

The sums of the mean $O \cdots C$ and $C \cdots S$ partial bond lengths in the TSs for enzymic and non-enzymic reactions do not differ significantly. Thus the more inverse $2°$ α-D_3 KIE for the enzymic reaction (1) does not arise from compression of the S_N2 transition state relative to that for the non-enzymic reaction (2). The KIE averaged over all RCs correlates inversely with the magnitude of the imaginary reaction coordinate frequency (in the range $440i$ to $540i$ cm^{-1} for reaction (1) and $200i$ to $500i$ cm^{-1} for reaction (2)) for each TS for both the enzymic and non-enzymic reactions: the greater the curvature along the reaction coordinate mode in the TS, the more inverse is the $2°$ α-D_3 KIE. For the COMT-catalysed reaction, the more inverse KIEs are associated with tighter coordination of the non-nucleophilic catechol OH to Mg^{2+}, as in **A**, in contrast to **B** and **C**.

The calculated $2°$ α-D KIE (0.99 ± 0.16) for the non-enzymic reaction (2), averaged over 1666 combinations of RC and TS, agrees with the experimental value (0.974 ± 0.016),[1b] while the calculated α-^{13}C KIE (1.044 ± 0.018) is again lower than the experimental value (1.083 ± 0.015).[1b] The non-enzymic TS configurations also have small standard deviations from the mean values of key bond lengths and angles (Table 2) but again there is considerable variation in the $O \cdots C$ distance in the RCs, corresponding to a very broad free-energy minimum in the PMF for this coordinate; however, the mean KIEs do not differ significantly between these groups of RCs.

AM1 gives a $2°$ α-D KIE for S_N2 reaction of cyanide with chloroethane significantly more inverse than the average of results from 29 *ab initio* and DFT methods in the gas phase,[10] but the apparently good agreement of the latter with experiment in DMSO is not significant owing to neglect of solvation effects. Moreover, it is not valid to compare that benchmark reaction with the present reactions (1) and (2) (Scheme 1) since the nucleophiles and nucleofuges are different.

However, although these ensemble-averaged KIEs may be subject to error due to use of the AM1 hamiltonian for the QM region and possibly due to (as yet) inadequate sampling of configurations, nonetheless they may be compared meaningfully with experiment and offer valuable insight into how different KIEs may arise from different families of configurations. We note that apparent differences between KIEs calculated by different QM methods are often smaller than the fluctuations found between different configurations within the same method. Use of a "high-level" correction (*e.g.* density functional theory with a respectable basis set) for the QM region with a dual-level methodology[11] is expected to provide more quantitatively accurate determinations of the KIEs, but still it will be essential to consider the effects of thermal fluctuations as demonstrated here.

Notes and references

1 (*a*) M. F. Hegazi, R. T. Borchardt and R. L. Schowen, *J. Am. Chem. Soc.*, 1979, **101**, 4359–4365; (*b*) C. H. Gray, J. K. Coward, K. B. Schowen and R. L. Schowen, *J. Am. Chem. Soc.*, 1979, **101**, 4351–4358.
2 J. Rodgers, D. A. Femec and R. L. Schowen, *J. Am. Chem. Soc.*, 1982, **104**, 3263–3268.
3 G. D. Ruggiero, I. H. Williams, M. Roca, V. Moliner and I. Tuñón, *J. Am. Chem. Soc.*, 2004, **126**, 8634–8635.
4 M. J. Field, M. Albe, C. Bret, F. Proust-De Martin and A. Thomas, *J. Comput. Chem.*, 2000, **21**, 1088–1100.
5 V. L. Schramm, Enzyme Kinetics And Mechanism, Pt E, *Methods Enzymol.*, 1999, **308**, 301–355.
6 G. D. Ruggiero, S. J. Guy, S. Martí, V. Moliner and I. H. Williams, *J. Phys. Org. Chem.*, 2004, **17**, 592–601.
7 I. H. Williams, *Chem. Phys. Lett.*, 1982, **88**, 462–466; I. H. Williams, *J. Mol. Struct. (THEOCHEM)*, 1983, **11**, 275–284.
8 E. Y. Lau and T. C. Bruice, *J. Am. Chem. Soc.*, 2000, **122**, 7165–7171.
9 B. Kuhn and P. A. Kollman, *J. Am. Chem. Soc.*, 2000, **122**, 2586–2596.
10 Y. Fang, Y. Gao, P. Ryberg, J. Eriksson, M. Kołodziejska-Huben, A. Dybała-Defratyka, S. Madhavan, R. Danielsson, P. Paneth, O. Matsson and K. C. Westaway, *Chem. Eur. J.*, 2003, **9**, 2696–2708.
11 S. Martí, V. Moliner, I. Tuñón and I. H. Williams, *J. Phys. Chem. B*, 2005, **109**, 3707–3710.