PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 7523

Improved NOE fitting for flexible molecules based on molecular mechanics data – a case study with *S*-adenosylmethionine†

Jessica Bame,[a] Casper Hoeck,[a] Matthew J. Carrington,[b] Craig P. Butts,*[a] Christof M. Jäger*[b] and Anna K. Croft*[b]

Received 26th October 2017,
Accepted 20th February 2018

DOI: 10.1039/c7cp07265a

rsc.li/pccp

The use of molecular dynamics (MD) calculations to derive relative populations of conformers is highly sensitive to both timescale and parameterisation of the MD. Where these calculations are coupled with NOE data to determine the dynamics of a molecular system, this can present issues if these populations are thus relied upon. We present an approach that refines the highly accurate PANIC NMR methodology combined with clustering approaches to generate conformers, but without restraining the simulations or considering the relative population distributions generated by MD. Combining this structural sampling with NOE fitting, we demonstrate, for *S*-adenosylmethionine (aqueous solution at pH 7.0), significant improvements are made to the fit of populations to the experimental data, revealing a strong overall preference for the *syn* conformation of the adenosyl group relative to the ribose ring, but with less discrimination for the conformation of the ribose ring itself.

Introduction

Understanding the dynamic processes of molecules in solution is essential for many fields, ranging from material science to biology. These processes can be highly complex, involving a mixture of many different interchanging conformations, each of which can have specific outcomes on physical and chemical properties, including directing of reaction outcomes. As such, improved methods for identifying conformations accurately, including exchange of conformations, are highly valuable.

S-Adenosylmethionine (AdoMet or SAM) is one such molecule where accurate conformational data are critical to understanding its biological and chemical roles. SAM is utilised extensively as an important cofactor and co-substrate for both methylation and radical-based enzyme reactions. Here the role of different conformations is crucial when probing the enzymatic reactions in which SAM is involved, as they have significant impacts in tuning the reactivity of SAM. As an alkylating agent, SAM participates in biosynthetic production of numerous compounds as one of nature's key methyl donors.[1,2] In this role, SAM is also involved in metabolic process regulation through contributing to nucleic acid[3] and protein methylation reactions.[4,5]

In a different role, SAM is found as a central cofactor or co-substrate in the enzyme family of radical SAM enzymes.[6] These enzymes catalyse a broad set of radical reactions, from C–C bond formations,[7–9] to complex skeleton rearrangements[10–14] (see also ref. 15 for an extensive review and references therein). In contrast to heterolytic bond cleavage between the sulfur and the methyl carbon atom when acting as methyl donor, SAM is bound to a central iron–sulfur cluster in radical SAM enzymes and is cleaved homolytically – following a one electron transfer from the iron sulfur cluster – into methionine and the 5′-deoxyadenosyl radical, which acts as the active species for further radical reaction.[15] The major reason for this distinct reaction lies simultaneously in the stability and thus reactivity of the resulting radical and on the details of the interactions between SAM and the cluster[16,17] or, in other words, on SAM's conformation and how it interacts with the biological macromolecule. Here, the conformation has been shown to have an important impact on controlling and directing the reductive homolytic C–S bond cleavage.[18,19]

The physical properties and conformational behaviour of SAM in solution have been previously investigated by means of circular dichroism,[20] IR, UV,[21] and NMR spectroscopy.[22,23] Throughout these studies a relatively consistent picture arises, that SAM should adopt predominantly an *anti*-configuration around the glycosylic bond and that the ring puckering of the ribose ring should primarily adopt the 3′-*endo* conformation (Fig. 1). A recent NMR study by Markham *et al.*[23] added further NOESY and ROESY information regarding interproton distances,

[a] *University of Bristol, School of Chemistry, Clifton, Bristol BS8 1TS, UK. E-mail: craig.butts@bristol.ac.uk*
[b] *University of Nottingham, Department of Chemical and Environmental Engineering, University Park, Nottingham, NG7 2RD, UK. E-mail: christof.jaeger@nottingham.ac.uk, anna.croft@nottingham.ac.uk*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp07265a

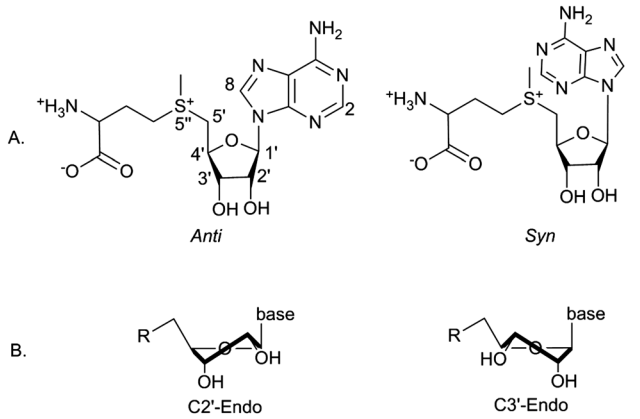

Fig. 1 Nomenclature for different *S*-adenosylmethionine conformations, including key numbering (A., *anti*). Different purine ring orientations, relative to the sugar (A) and different ring puckering (B).

which similarly found the *anti*-3′-*endo* conformation dominated. Such interproton distances derived from NOE (Nuclear Overhauser Enhancement) experiments can be used as part of evaluations of the quality of molecular force fields used in molecular dynamics simulations.[24,25] The data from Markham *et al.* have been used for SAM force field parametrisation by Saez and Vöhringer-Martinez,[26] where they introduced and improved the SAM force field for molecular simulations with a better description of involved conformers.

However, when dealing with simulation averages of molecules representing a mixture of two or more conformers in equilibrium, the comparison with NMR data comes with significant challenges. In particular, the r^{-6} averaging of the experimental NOE signal is very sensitive to short distance (r) distributions, but is insensitive to longer distance distributions. Hence, a limited NOE dataset may not be able to describe the entire conformational space of molecule. Moreover, when dealing with a conformer mixture, simplified averaging of simulated conformer distributions to compare single experimental NOE distance values is inappropriate, simply because many distance distribution combinations can lead to the exact same result upon averaging. Instead, a range of distance distributions over an ensemble of different molecular conformations should be employed, as pointed out by van Gunsteren and others.[25,27,28] Generating this ensemble of conformations is in itself a challenging task – in particular establishing the populations of the contributing conformers. These populations are generally derived from their computed energies, however computation of relative conformer energies, whether force field, DFT or *ab initio*, cannot deliver sufficient accuracy to account for a precise distribution of conformational equilibria even when these achieve so-called 'chemical accuracy' (typically considered to be <1 kcal mol^{-1}), which is considered to serve well for computational estimation of reaction kinetics and thermodynamics. However, changing the relative free energy difference between two conformers from +1 kcal mol^{-1} to -1 kcal mol^{-1} changes their relative populations from $\sim$70 : 30 to $\sim$30 : 70 at room temperature, *i.e.* their relative populations can invert within the error of even 'chemically accurate' energy calculations.

While individual conformers can often be represented structurally accurately in MD simulations – due to extensive parameterisation of the bonding parameters based on experimental and high-level computational data – relative conformer energies can be influenced significantly by intra- and intermolecular nonbonding interactions often dominated by electrostatics. In most force fields the electrostatics are represented by non-polarisable fixed point charges derived from quantum chemical calculations for one or more conformers of the target molecule. The choice of the conformers and the QM method to derive the charges significantly influences the derived point charges and thus the electrostatic interactions. Further, MD simulations have the need for extensive sampling in order to reach equilibrium across all of the conformational space, which can be tackled by approaches such as enhanced sampling methods. However, it still remains uncertain when conformational equilibrium is reached and other approaches developed therefore try to prevent this need.[29]

On the other hand, experimental methods able to accurately describe conformational equilibria – ideally without relying on the quality of populations derived from computations – promise a better solution to these structural challenges. Experimentally, we have demonstrated the potential for very high accuracy in NOE-distance analyses for conformationally rigid small molecules[30,31] operating in the fast-tumbling regime by using the PANIC[32] correction to NOE intensities prior to conversion into distances. This in turn can be applied to the simultaneous accurate determination of stereochemistry and/or conformation in semi-flexible[33,34] and flexible small molecules[35–37] through comparison to computation. While not applicable to large macromolecular systems with correlation times much longer than their conformational lifetimes, it has been shown to accurately quantify low population (1–2%) conformers,[33] demonstrating how powerful NOE-distance analysis can be in the study of conformationally flexible small molecule systems. This analysis has recently been shown to be sufficiently accurate to quantify very small changes in conformer populations induced by temperature ($<0.5\%/10\ ^{\circ}C$).[38] Throughout this study we have found that relying simply on conformer populations derived from computed energies (either force-field or density functional theory) does not provide sufficient accuracy to describe the experimental results.[37,39]

Herein we investigate the effect of removing reliance on MD-derived populations and introducing more accurate experimental NOE-distance data on the conformational analysis of *S*-adenosylmethionine in aqueous solution. Rather than relying on averaging over MD simulation trajectories with all of the inherent deficiencies highlighted above, we instead take conformer ensembles derived from classical molecular dynamics simulations and filter these such that averaged clusters are represented as individual conformers, without relying on their (calculated) population contribution. We combine this with accurate NOE-distance determinations to more precisely evaluate the molecule's dynamic conformational equilibrium in solution, and thus better fit the populations of the conformer ensembles to experiment.

Methods

NMR method

S-Adenosylmethionine (Cayman Chemicals) was prepared in solution containing 4–10 mM (1.1–2.7 mg) *S*-adenosyl-methionine, in 0.1 M (11 mg) d_{11}-Tris–HCl and 0.7 ml D_2O at pH 7.0 to provide physiological ionisation states. NMR experiments were performed on a 500 MHz Varian VNMRS Direct Drive NMR spectrometer equipped with an Agilent OneNMR probe. All NMR experiments were run at 25 °C. NMR experiments used for assignment of SAM were ^{1}H NMR (8 scans, spectral width 20 ppm (10 000 Hz)), ^{13}C NMR (125 MHz, 2000 scans, spectral width 275 ppm (34 345 Hz)), HSQC (4 scans, 200 t1 increments, spectral widths F2 20 ppm (10 000 Hz), F1 200 ppm (25 000 Hz)), H2BC (4 scans, 200 t1 increments, spectral widths F2 20 ppm (10 000 Hz), F1 200 ppm (25 000 Hz)), HMBC (8 scans, 200 t1 increments, spectral widths F2 20 ppm (10 000 Hz), F1 200 ppm (25 000 Hz)). NMR spectra were analysed and processed using MestreNova version 8.1.2-11880. To determine interatomic distances between protons in the molecule, DPFGSE 1D NOE spectra were collected (128 scans, mixing time 500 ms, spectral width 20 ppm (10 000 Hz)). The general method for generating interproton distances from the NOE intensities has been reported elsewhere,[36] and is described for this case in Section S3 of the ESI.†

MD simulation and data clustering

Molecular-dynamics simulations were performed using the GPU implementation[40–42] of the Amber16[43] molecular dynamics package. The force field parameters for SAM are based on updated force field parameters from Saez and Vöhringer-Martinez[26] suitable for use with the Amber force fields. Electrostatic point charges were reparametrised following the restrained electrostatic potential (RESP) fitting procedure by Kollman *et al.*[44] and are based on multiconfigurational fitting of three different conformers. The structures for RESP fitting were taken from the crystal structures of butirosin biosynthetic enzyme,[45] BtrN (pdb entry 4M7T), tRNA-wybutosine synthesising enzyme, TYW2 (pdb entry 3A25),[46] and 7-carboxy-7-deazaguanine synthase, QueE (pdb entry 4NJI),[11] representing bent and stretched SAM conformations. The structures were geometry optimised at the B3LYP[47–49]/6-31+G(d)[50,51] level of theory including diffuse functions[52] applying the polarisable continuum model (PCM)[53] as the implicit solvation model with Gaussian09.[54] Two sets of charges were subsequently derived. The first set was prepared following the standard RESP procedure at the HF/6-31G(d) level, and a second set was generated based on PCM-B3LYP/cc-PVTZ[51] calculations in implicit solvent with a dielectric constant of 4.335, which is suitable for representing the electrostatic environment in a protein more closely. The derived point charges can be found in the ESI.†

The simulations were carried out for SAM in the +1 charged state in explicit solvent, using the SPC/E[55] water model. The simulations were conducted at a temperature of 300 K using periodic boundary conditions. In total 2515 water molecules have been added to the system to form a truncated octahedral unit cell big enough to prevent significant interactions of the solvated SAM molecule with its own mirror images (48.5 Å cell length). Following a combined steepest descent and conjugate gradient minimisation for 2500 steps, the simulations were carried out for at least 500 ns, applying constant pressure (*NPT*) molecular dynamics at one atmosphere using a Langevin directed dynamics for pressure control. Electrostatic long range interactions were treated with the Particle Mesh Ewald (PME)[56] method and a 12 Å cut-off for nonbonding interactions.

The MD simulation data was not taken to fit NOE data directly, but was subdivided into clusters, representing specific structural features, with the rotation around the glycosylic bond and the conformation of the sugar determined as the most important conformational features. A standard clustering method based on root mean square differences was not efficient in discriminating clusters that differed in these features, thus the dihedral angles representing those structural changes were taken for clustering directly using the programme DASH[57] together with Amber's analysis programme cpptraj.[58] The individual clusters were subsequently taken for further analysis.

S-Adenosylmethionine tends to invert its stereochemistry at the sulfur in solution rather quickly. Thus, all NMR experiments described were conducted with a mixture of diastereomers (differing in stereochemistry at the sulfur). A comparison of simulation of both diastereomers of SAM (see ESI,† for details) showed only slight differences in the interatomic spacing related to the adenine ring orientation and the sugar puckering (the largest deviation was <0.05 Å), hence the analysis is robust within the experimental accuracy described.

Results and discussion

NOE-determined interproton distances

The interproton distances extracted from the NOE spectra for a solution of *S*-adenosylmethionine in D_2O (pH 7.0) are outlined in Table 1. The NOE intensities were extracted using the PANIC methodology,[32] which corrects for differential relaxation between NOE spin pairs and improves the accuracy of interproton distances measured by NOE spectroscopy. A larger number of NOEs were measured than were reported by Markham *et al.*[23] The common distances measured matched broadly with those reported by Markham *et al.*,[23] but with differences in H2′–H8 and H3′–H8 of 0.32 Å and 0.31 Å, respectively, which are higher than expected deviations for the experimental techniques reported herein. These distances directly report on the rotational position of the adenine ring relative to the sugar (*i.e. anti vs. syn* conformations) and hence this difference can be considered potentially significant.

MD clustering

The time series arising from MD simulations of SAM were clustered along the dihedral angles around the glycosylic bond (O–C1′–N9–C8) and the sugar central bond (C4′–C3′–C2′–C1′) into six distinctive clusters (Fig. 2 and 3, see ESI,† for details) that represent the different orientations of the adenine ring and conformations of the sugar ring. In order to ensure that the resulting clusters represent individual NOE-relevant conformers, all distances corresponding to experimental NOE values were

Table 1 Interproton distances from NOE measurements in this study, literature, and MD simulation (500 ns, PCM parametrisation)

		NOE-derived distances [Å]		NOE r^{-6} averaged simulation derived interatomic distances [Å]							
H1	H2	This study	Markham *et al.*[23]	Cluster 1 *syn*-2′-*endo*	Cluster 2 *syn*-3′-*endo*	Cluster 3 *anti*$_1$-2′-*endo*	Cluster 4 *anti*$_1$-3′-*endo*	Cluster 5 *anti*$_2$-2′-*endo*	Cluster 6 *anti*$_2$-3′-*endo*	Cluster-averaged (MD population)	Cluster averaged (NOE population)
1′	8	2.68	2.6	2.50	2.50	3.68	3.70	3.66	3.72	2.78	2.68
1′	2′	2.92	2.9	3.01	2.83	3.01	2.81	3.01	2.79	2.85	2.89
1′	4′	2.99	NR	3.00	2.86	2.72	2.88	2.75	2.89	2.87	2.89
2′	8	2.78	3.1	4.33	3.78	2.75	3.22	2.24	2.18	2.89	2.82
2′	1′	2.89	2.9	3.01	2.83	3.01	2.81	3.01	2.79	2.85	2.89
2′	3′	2.40	NR	2.31	2.30	2.29	2.33	2.30	2.34	2.31	2.31
3′	8	3.21	2.9	5.94	5.34	4.14	2.86	3.75	2.46	3.15	3.26
4′	1′	2.89	NR	3.00	2.86	2.72	2.88	2.75	2.89	2.87	2.89
3′	2′	2.36	NR	2.31	2.30	2.29	2.33	2.30	2.34	2.31	2.31
5″	8	3.79	NR	5.87	6.07	3.15	2.74	3.61	3.71	3.21	3.55
5″	2′	2.74	NR	2.53	4.17	2.76	4.15	2.63	3.97	3.22	2.93
8	1′	2.68	NR	2.50	2.50	3.68	3.70	3.66	3.72	2.78	2.68
8	2′	2.83	NR	4.33	3.78	2.75	3.22	2.24	2.18	2.89	2.82
8	3′	3.22	NR	5.94	5.34	4.14	2.86	3.75	2.46	3.09	3.17
MAD				26.5%	24.6%	13.1%	15.8%	13.0%	16.5%	4.7%	2.0%
STD				35.4%	30.3%	18.1%	22.2%	17.7%	23.5%	7.0%	3.0%

'NR' = not reported. 'MAD' = mean average deviation between cluster-averaged distances and NOE-derived distances from this study. 'STD' = standard deviation between cluster-averaged distances and NOE-derived distances from this study.

analysed and found to show single narrow ranges of distances for each interatomic contact, as shown in Fig. 4 for H8–H1′ and H8–H2′. All clusters show flexibility in the methionine part of SAM, which has not been considered further for this clustering approach since, as described later, this flexibility is not discriminatory in the NOE data.

It should be noted that while using one dihedral angle for clustering the ribose ring puckering allows comparison with the experimental NOE data, it does not represent an entirely precise description of all the ring puckering conformers. The dominating 3′-*endo* and 2′-*endo* conformations represented by the clusters always show additional contributions of the closely related 4′-*exo* and 1′-*exo* conformations respectively according to analysis based on the nomenclature convention by Altona and Sundaralingam.[59] A detailed ring puckering analysis for the simulations is presented in the ESI.†

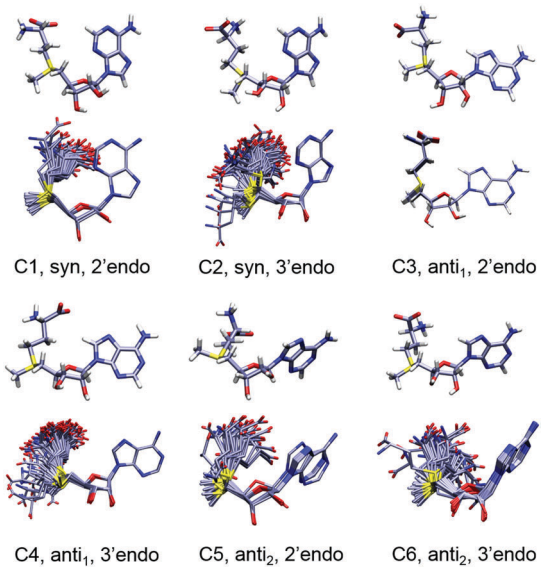

Fig. 2 Representation of the six conformer clusters found in 500 ns MD simulation using PCM-derived point charges, shown as superimposed ensembles with one example structure depicted for each.

Motion in the adenine and sugar part of SAM is restricted, as represented in the well-defined peaks in the dihedral angle histograms. Fig. 3 depicts the total distribution and the individual dihedral angle distribution for each cluster after analysis of 500 ns simulation (using PCM derived point charges).

From this data, the MD simulation suggests a roughly equal population of the *anti* and *syn* conformations of the adenine moiety although the *anti*-descriptor comprises two different angular orientations (*anti*$_1$ and *anti*$_2$) of which *anti*$_1$ is slightly preferred. Both the 2′-*endo* and *the* 3′-*endo* puckering of the sugar ring are observed with a computed preference for the 3′-*endo* puckering. This is all broadly in line with previous studies, although Markham *et al.*[23] do not discriminate between the *anti*$_1$ and *anti*$_2$ forms.

In order to compare the MD results to experimental NOE-distance data, the distance distributions for each H–H pair were considered individually. Fig. 4 illustrates two example distance distributions for H8–H2′ and H8–H1′, which represent rotation around the glycosidic bond (O–C1′–N9–C8). The H8–H1′ distribution shows roughly equal populations of short and long distances, representing the roughly equal populations of both *syn* and *anti* conformations, highlighted in the dihedral plots. Similarly, the H8–H2′ distribution is dominated by longer

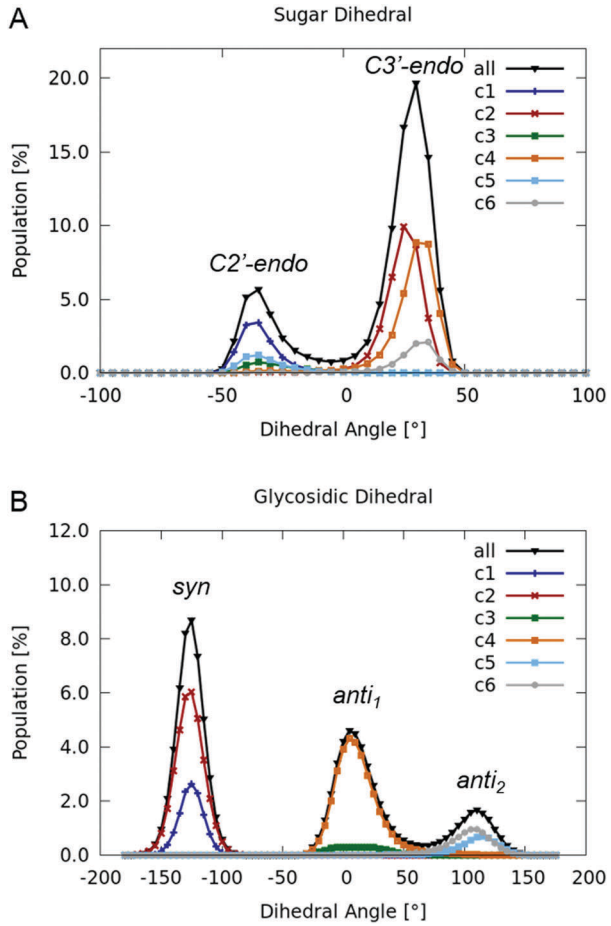

Fig. 3 Total and individual cluster dihedral angle distributions for SAM from 500 ns MD simulation using PCM-derived point charges.

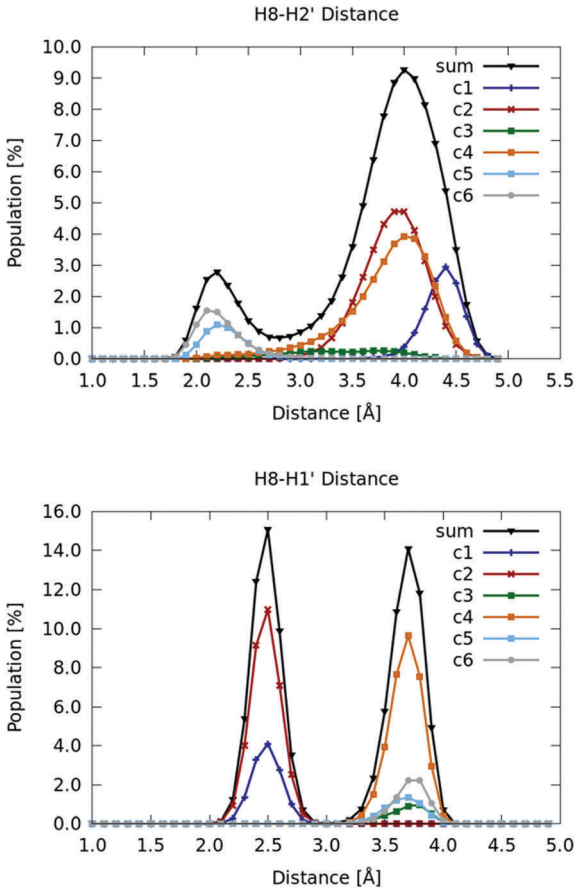

Fig. 4 Total and individual cluster distance distributions for H8–H2′ and H8–H1′ of SAM found in 500 ns MD simulation using PCM-derived point charges.

distances reflecting the relatively lower contribution of the $anti_2$ conformation to the overall calculated conformational ensemble. The computational results specifically highlighted the challenge in calculating accurate time-averaged internuclear distances directly from MD simulations alone. The double peak potentials observed in the H–H distance plots (Fig. 4) impose a need for high accuracy in the relative weighting/population of the individual peaks in order to be useful for NOE data fitting – especially given the sensitivity of NOE data to short distances over long distance.

However, as can be seen from Fig. 5 the relative population of individual clusters varies significantly (increasing/decreasing by nearly a factor of two in some cases) during the first 400 ns of simulation. This is true for all of the charge parameterisations used, and even after 500 ns one cannot be sure whether any of the simulations has yet reached equilibrium. This population variance demonstrates that cluster/conformer populations based solely on MD must be considered highly suspect without clear evidence that conformational equilibrium has been reached. Similarly, repeating the simulation with a different electrostatic parametrisation of point charges changes the relative population of the individual clusters (Table 2). This sensitivity of the population to the timescale and parameterisation of the MD creates an uncertainty as to whether the populations derived from any given MD simulation can be reliably used for comparison to experiment at all.

In contrast, geometries, and hence interatomic distances, within the individual clusters are not significantly sensitive to the MD method or parameterisation of electrostatics. There is negligible change in these throughout the simulation (between 300 and 500 ns, the mean average distance difference of all NOE relevant interatomic distances is 0.004 Å) and repeating the simulation with a different electrostatic parametrisation of point charges does also not influence the average interatomic distances of each individual cluster significantly (mean average distance difference 0.034 Å, see ESI† for details).

At this point it appears that the populations of conformers arising from MD cannot be considered reliable, however the interatomic distance data for each conformer/cluster from MD simulations can. Consequently, for each pair of protons an effective H–H distance was calculated for each cluster 1–6 by taking corresponding distance in each individual conformer within a cluster, and reducing these by r^{-6} to provide an effective NOE for each conformer as described elsewhere,[33] then weighting these NOEs by their calculated population within that cluster (relative populations within each cluster

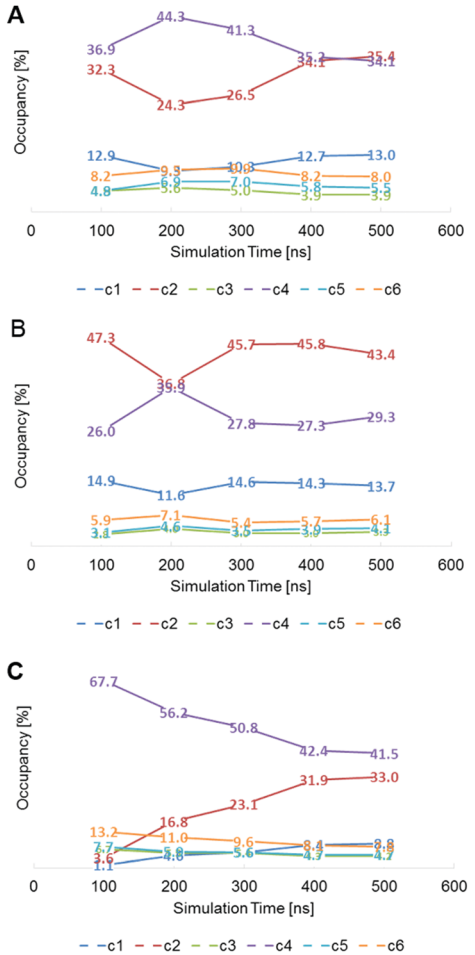

Fig. 5 Relative cumulative cluster contributions over simulation time for different simulations using (a) PCM derived point charges, (b) HF derived point charges, (c) *R*,*S*-isomer with PCM derived point charges.

Table 2 Relative conformer distribution and cluster variability from different simulations. MSD relates to the whole SAM molecule, 'msd part' represents the mean square deviation of the adenine and sugar part of SAM only (values in Å^2)

	PCMfit 500 ns			HFfit 500 ns			PCMfit-*R*-*S*-isomer		
	occ [%]	msd	msd part	occ [%]	msd all	msd part	occ [%]	msd all	msd part
1	13.0	0.793	0.195	13.7	0.796	0.193	8.8	0.929	0.204
2	35.4	1.084	0.221	43.4	1.083	0.219	33.0	1.293	0.225
3	3.9	1.198	0.336	3.3	1.316	0.342	4.2	1.304	0.337
4	34.1	1.164	0.300	29.3	1.205	0.297	41.5	1.321	0.304
5	5.5	1.178	0.303	4.1	1.164	0.308	4.7	1.247	0.310
6	8.0	1.200	0.326	6.1	1.171	0.320	7.8	1.383	0.342

were not found to vary significantly during or between MD simulations, in contrast to what was observed for the relative populations between each cluster). Each population-weighted NOE was then converted back into a single distance, again using the r^{-6} NOE-distance relationship. The resulting cluster-averaged NOE-distances for each H–H pair are shown in Table 1 (the full method describing their calculation can be found in Section S3.2 of the ESI†).

NOE-MD fitting

The unreliability of MD-derived cluster populations was confirmed by firstly using the MD-derived cluster populations to calculate ensemble-averaged NOE-interproton distances. The 500 ns PCM data were compared to the experimental NOE-interproton distances and found to provide an approximate, but less than ideal, fit (MAD 4.7%, StDev 7.0%) compared to the accuracy expected for NOE-distance measurements based on rigid molecules (typically <4% for both MAD and StDev).[30] Fitting using the HF parameterisation gave a grossly similar result (4.4% MAD, StDev 6.9%) This suggests that both MD results are equally poor descriptions of the NOE data, although it should be noted that the cluster-averaged distances offer a substantially better fit to the NOE data than any single cluster alone (see final two rows of Table 1). The likely sources of the poor fit are either missing clusters (reflecting a lack of convergence in the MD simulation timescale or insufficient quality of the force field parameterisation) or poor description of the cluster populations by the MD simulation. We and others have previously shown that the populations can be better described by consideration of the experimental interproton distances[36,39] and experimental residual dipolar coupling,[60] and methodologies are being developed aiming for describing conformational equilibria of more challenging molecules like flexible sugars[61] and biomolecules.[62] Consequently, the populations of the six MD-derived clusters were perturbed to offer a best-fit to the experimental NOE-distance data in Table 1 (full method for this fitting can be found in Section S3.3 of the ESI†). An excellent match to the NOE-distance data can be achieved, with MAD and StDev of 2.0% and 2.0%, respectively. These NOE-fitted populations (Table 3) suggest that the *syn* conformer (clusters 1 and 2) present in 61% of populated conformers, and 3′-*endo* present in 58% (Table 4). This conformer distribution contrasts with those reported by Markham *et al.* and earlier studies,[20–23] which suggest a preference for the *anti* and 3′-*endo* conformations. Refining the conformer populations using the values reported by Markham *et al.* (Table 1) for their 5 experimental distances gave less than satisfactory results, as no solution could be found which gave a standard deviation of <3%, presumably arising from the less accurate NOE-distance analysis used in that earlier report.

In order to test if the ensemble of measured NOE-distances is either over-fitted or not very sensitive to the relative populations of MD clusters, the sensitivity of the fitting procedure to changes in conformer populations was tested. A least-squares minimisation of

Table 3 Best fit of NOE distance data to MD cluster populations 1–6 (300 ns, PCM simulation) derived by non-linear least squares optimisation of the standard deviation between experimental and calculated NOE-distances data as described in detail in Section S3 of the ESI

	Population %
Cluster 1	31
Cluster 2	31
Cluster 3	4
Cluster 4	16
Cluster 5	7
Cluster 6	11

Table 4 Adenine *syn*/*anti* and ribose 2′/3′-*endo* best fit conformer population to NOE distance data

Ade	Pop %	Sugar $H_{3'}$	Pop %
syn (C1 + C2)	62	2′-*endo* (C1 + C3 + C5)	42
anti$_1$ (C3 + C4)	20	3′-*endo* (C2 + C4 + C6)	58
anti$_2$ (C5 + C6)	18		

the standard deviation between experimental and calculated NOE-distances was conducted for a series of fixed total populations of each grouped cluster *e.g.* 2′-*endo*, while allowing the relative populations within that grouped cluster, along with the populations of all the other clusters, to vary. The results (Fig. 6) confirm that the *syn* form (clusters 1 + 2) has to be present in a limited range of populations (around 50–75%), while *anti*$_1$ (clusters 3 + 4) and *anti*$_2$ (clusters 5 + 6) can only each be present in 30% or lower populations if an acceptable fit to the experimental result (<3% standard deviation) is to be achieved. The NOE data is much less sensitive to the 2′-*endo*/3′-*endo* conformation of the sugar, with the populations lying between 35–80% and 20–65% for 3′-*endo* and 2′-*endo*, respectively. In other words, this sensitivity test confirms the found preference for the *syn* conformation with very high confidence, while the distinction between 2′- and 3′-*endo* preference needs to be treated with care on the basis of NOE data alone.

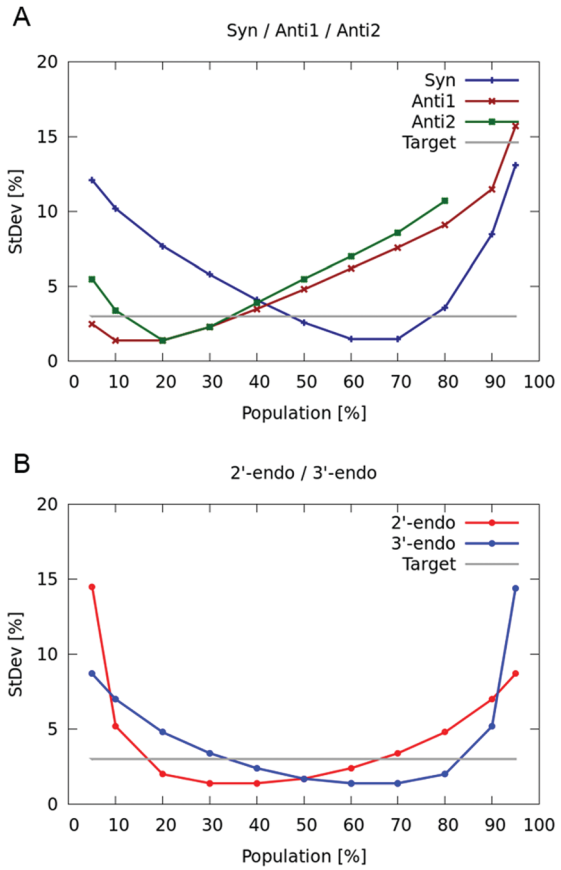

Fig. 6 Quality of NOE fit presented as standard deviation (StDev) of fit *vs.* populations of various clustered MD conformers.

Conclusions

The conformational flexibility of *S*-adenosylmethionine presents an excellent example for a detailed evaluation of conformational equilibria by NMR NOE experiments in combination with molecular modelling. Our evaluation particularly highlights the strengths and weaknesses of this type of approach. By combining MD-derived conformer geometries with NOE-fitted conformer populations derived from more accurate NOE-distance measurements, we avoid the poor computational descriptions of conformer populations, and show that the SAM conformational space in aqueous solution differs significantly from previous reports. In particular, we find that the conformation of the adenine moiety prefers to sit *syn* rather than *anti* to the sugar, but with some contribution from the *anti* conformations certainly required to properly model the experimentally-derived NOE distances. There is less discrimination in the ratio of 2′:3′-*endo* sugar conformations, where both must be populated by at least 20% in order to fulfil the observed NOE data, but with a bias towards the 3′-*endo* conformation appearing likely in order to obtain the best fit to the experimental NOE data.

This study highlights once again how difficult it is to deliver adequate structural preferences from molecular modelling for accurate NOE data fitting. Classical atomistic molecular force fields often struggle to predict conformational equilibrium distributions that are accurate enough for this procedure. This poor prediction, and the above-mentioned reasons, show why it is problematic to either take MD equilibria for NOE fitting or utilise NOE data for quality control of the force fields itself.

In contrast to energetic data, if the force fields are parametrised precisely enough (*e.g.* against high level QM and experimental data), they can accurately describe individual conformers. Our clustering approach from MD data showed that, for the given example, the available force field generates equilibrated individual conformers that can be used for accurate NOE fitting. The quality of the force field and, in particular, the question of whether the force field represents all relevant conformers in turn can also be assessed by the quality check of the NOE fitting itself.

To summarise, using geometries from unrestrained MD simulations avoids errors arising from computed populations and over-restraining with NOE experimental data. Clustering these MD geometries to represent single peak distributions for (NOE-relevant) interatomic distances, reduces problems arising from the r^{-6} dependence of the NOE while still reflecting the entire conformational space of the molecules, including secondary effects such as steric clashes, hydrogen bonding and solvation. While this does not mean that the conformational space of every molecule can be both precisely and accurately described by this approach, where appropriate experimental NOE data is available a more reliable description should be obtained by this approach.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

MJC was funded by BP through the Nottingham Engineering Research Placement Scheme. AKC and CMJ would like to acknowledge support from the European Cooperation in Science and Technology (COST) network CM1201. CMJ acknowledges funding through the Dean's Prize for Engineering (UoN), Nottingham Advanced Research Fellowship and EU FP7 Marie Curie Actions – People, Co-funding of Regional, National and International Programmes (COFUND) under Grant Agreement no PCOFUND-GA-2012-600181. AKC and CMJ also gratefully acknowledge support and access to the University of Nottingham High Performance Computing Facility.

Notes and references

1 H. Lin, *Bioorg. Chem.*, 2011, **39**, 161–170.
2 M. Fontecave, M. Atta and E. Mulliez, *Trends Biochem. Sci.*, 2004, **29**, 243–249.
3 F. Yan and D. G. Fujimori, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3930–3934.
4 M. I. Jackson, J. Cao, H. Zeng, E. Uthus and G. F. Combs, *J. Biol. Chem.*, 2012, **287**, 36455–36464.
5 Y. Shi and J. R. Whetstine, *Mol. Cell*, 2007, **25**, 1–14.
6 H. J. Sofia, G. Chen, B. G. Hetzler, J. F. Reyes-Spindola and N. E. Miller, *Nucleic Acids Res.*, 2001, **29**, 1097–1106.
7 O. T. Magnusson, H. Toyama, M. Saeki, R. Schwarzenbacher and J. P. Klinman, *J. Am. Chem. Soc.*, 2004, **126**, 5342–5343.
8 L. E. Cooper, D. Fedoseyenko, S. H. Abdelwahed, S.-H. Kim, T. Dairi and T. P. Begley, *Biochemistry*, 2013, **52**, 4592–4594.
9 I. Barr, J. A. Latham, A. T. Iavarone, T. Chantarojsiri, J. D. Hwang and J. P. Klinman, *J. Biol. Chem.*, 2016, **291**, 8877–8884.
10 G. Sicoli, J.-M. Mouesca, L. Zeppieri, P. Amara, L. Martin, A. L. Barra, J. C. Fontecilla-Camps, S. Gambarelli and Y. Nicolet, *Science*, 2016, **351**, 1320–1323.
11 D. P. Dowling, N. A. Bruender, A. P. Young, R. M. McCarty, V. Bandarian and C. L. Drennan, *Nat. Chem. Biol.*, 2014, **10**, 106–112.
12 L. D. Palmer and D. M. Downs, *J. Biol. Chem.*, 2013, **288**, 30693–30699.
13 P. Hänzelmann and H. Schindelin, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12870–12875.
14 C. M. Jäger and A. K. Croft, *Chem. – Eur. J.*, 2017, **23**, 953–962.
15 J. B. Broderick, B. R. Duffus, K. S. Duschene and E. M. Shepard, *Chem. Rev.*, 2014, **114**, 4229.
16 J. A. Kampmeier, *Biochemistry*, 2010, **49**, 10770–10772.
17 J. B. Broderick, *Nature*, 2010, **465**, 877–878.
18 A. Dey, Y. Peng, W. E. Broderick, B. Hedman, K. O. Hodgson, J. B. Broderick and E. I. Solomon, *J. Am. Chem. Soc.*, 2011, **133**, 18656–18662.
19 K. A. Shisler and J. B. Broderick, *Arch. Biochem. Biophys.*, 2014, **546**, 64–71.
20 H. Follmann, I. Kuntz and W. Zacharias, *Eur. J. Biochem.*, 1975, **58**, 31–41.
21 W. A. Klee and S. H. Mudd, *Biochemistry*, 1967, **6**, 988–998.
22 M. L. Stolowitz and M. J. Minch, *J. Am. Chem. Soc.*, 1981, **103**, 6015–6019.
23 G. D. Markham, P.-O. Norrby and C. W. Bock, *Biochemistry*, 2002, **41**, 7636–7646.
24 T. A. Soares, X. Daura, C. Oostenbrink, L. J. Smith and W. F. van Gunsteren, *J. Biomol. NMR*, 2004, **30**, 407–422.
25 B. Zagrovic and W. F. van Gunsteren, *Proteins: Struct., Funct., Bioinf.*, 2006, **63**, 210–218.
26 D. A. Saez and E. Vöhringer-Martinez, *J. Comput.-Aided Mol. Des.*, 2015, **29**, 951–961.
27 R. Bürgi, J. Pitera and W. F. van Gunsteren, *J. Biomol. NMR*, 2001, **19**, 305–320.
28 X. Daura, I. Antes, W. F. van Gunsteren, W. Thiel and A. E. Mark, *Proteins*, 1999, **36**, 542–555.
29 E. Haensele, N. Saleh, C. M. Read, L. Banting, D. C. Whitley and T. Clark, *J. Chem. Inf. Model.*, 2016, **56**, 1798–1807.
30 C. P. Butts, C. R. Jones, E. C. Towers, J. L. Flynn, L. Appleby and N. J. Barron, *Org. Biomol. Chem.*, 2011, **9**, 177–184.
31 M. G. Chini, C. R. Jones, A. Zampella, M. V. D'Auria, B. Renga, S. Fiorucci, C. P. Butts and G. Bifulco, *J. Org. Chem.*, 2012, **77**, 1489–1496.
32 H. Hu and K. Krishnamurthy, *J. Magn. Reson.*, 2006, **182**, 173–177.
33 C. P. Butts, C. R. Jones and J. N. Harvey, *Chem. Commun.*, 2011, **47**, 1193–1195.
34 C. P. Butts, C. R. Jones, Z. Song and T. J. Simpson, *Chem. Commun.*, 2012, **48**, 9023–9025.
35 S. Di Micco, A. Zampella, M. V. D'Auria, C. Festa, S. De Marino, R. Riccio, C. P. Butts and G. Bifulco, *Beilstein J. Org. Chem.*, 2013, **9**, 2940–2949.
36 M. Burns, S. Essafi, J. R. Bame, S. P. Bull, M. P. Webster, S. Balieu, J. W. Dale, C. P. Butts, J. N. Harvey and V. K. Aggarwal, *Nature*, 2014, **513**, 183–188.
37 C. R. Jones, C. P. Butts and J. N. Harvey, *Beilstein J. Org. Chem.*, 2011, **7**, 145.
38 C. R. Jones, M. D. Greenhalgh, J. R. Bame, T. J. Simpson, R. J. Cox, J. W. Marshall and C. P. Butts, *Chem. Commun.*, 2016, **52**, 2920–2923.
39 J. Wu, P. Lorenzo, S. Zhong, M. Ali, C. P. Butts, E. L. Myers and V. K. Aggarwal, *Nature*, 2017, **547**, 436–440.
40 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.
41 S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.
42 A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.
43 D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, *AMBER 2016*, University of California, San Francisco, 2016.

44 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
45 P. J. Goldman, T. L. Grove, S. J. Booker and C. L. Drennan, *Proc. Natl. Acad. Sci.*, 2013, **110**, 15949–15954.
46 M. Umitsu, H. Nishimasu, A. Noma, T. Suzuki, R. Ishitani and O. Nureki, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 15616–15621.
47 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
48 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785.
49 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623.
50 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.
51 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
52 T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294–301.
53 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3093.
54 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, N. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, 2009.
55 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.
56 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
57 D. W. Salt, B. D. Hudson, L. Banting, M. J. Ellis and M. G. Ford, *J. Med. Chem.*, 2005, **48**, 3214–3220.
58 D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
59 C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, 1973, **95**, 2333–2344.
60 A. Kolmer, L. J. Edwards, I. Kuprov and C. M. Thiele, *J. Magn. Reson.*, 2015, **261**, 101–109.
61 J. Xia, C. J. Margulis and D. A. Case, *J. Am. Chem. Soc.*, 2011, **133**, 15252–15255.
62 N. Salvi, L. Salmon and M. Blackledge, *J. Am. Chem. Soc.*, 2017, **139**, 5011–5014.