RSC Advances

RSC Publishing

PAPER

Accurate prediction of ^{1}H chemical shifts in interstrand cross-linked DNA†

Cite this: *RSC Advances*, 2013, **3**, 3925

Ewald Pauwels,*[a] Diederica Claeys,‡[a] José C. Martins,[b] Michel Waroquier,[a] Giuseppe Bifulco,[c] Veronique Van Speybroeck[a] and Annemieke Madder*[d]

Received 5th October 2012,
Accepted 10th January 2013

DOI: 10.1039/c3ra22408b

www.rsc.org/advances

Structural analysis of modified DNA with NMR is becoming ever more difficult with increasingly complex compounds under scrutiny for use in medical diagnosis, therapeutics, material science and chemical synthesis. To facilitate this process, we developed a molecular modeling approach to predict proton chemical shifts in sufficient agreement with experimental NMR measurements to guide structure elucidation. It relies on a QM/MM partitioning scheme and first principle calculations to predict the spatial structure and calculate corresponding proton chemical shifts. It is shown that molecular dynamics simulations that take into account solvent and temperature effects properly are of utmost importance to sample the conformational space sufficiently. The proposed computational procedure is applicable to modified oligonucleotides and DNA, attaining a mean error for the proton chemical shifts of less than 0.2 ppm. Here, it is applied on the Drew–Dickerson d(CGCGAATTCGCG)$_2$ dodecamer as a benchmark system and the mispair-aligned N^{3}T-ethyl-N^{3}T cross-linked d(CGAAAT*TTTCG)$_2$ undecamer, illustrating its use as computational tool to assist in structure elucidation. For the proton chemical shifts in the cross-linked system our methodology yields a strikingly superior description, surpassing the predictive power of (semi-) empirical methods. In addition, our methodology is the only one available to make an accurate prediction for the protons in the actual cross-link. To the best of our knowledge, this is the first computational study that attempts to determine the chemical shifts of oligonucleotides of this size and at this level of complexity.

Introduction

Modified DNA is increasingly used in medical diagnosis[1,2] and therapeutics,[3,4] in materials science[5] and chemical synthesis.[6] New developments in these areas can be stimulated by understanding phenomena at the molecular level which requires structural insight into the DNA oligomer and its modifications. However, NMR structure analysis is becoming ever more stringent, especially with the growing size of the oligomer being considered. Molecular modeling can greatly facilitate this process: in a carefully designed computational approach, the spatial structure of a modified DNA oligomer can be obtained from simulation, along with its NMR properties. This allows comparison with experimental data and completion of gaps in the sequential resonance assignment. However, few modeling methodologies are accurately and universally applicable to (novel) chemical modifications. Here, we address this shortcoming and develop a density functional theory (DFT) approach to accurately predict proton chemical shifts in (modified) DNA oligomers. We focus on model interstrand cross-linked (ICL) oligonucleotides and concentrate on the cross-link region, where traditional semi-empirical methods fail to predict chemical shifts that are sufficiently close to the experimental NMR data to guide further structure elucidation.

ICLs are perhaps the most well known modifications to DNA. Induced by various chemical agents (*e.g.* nitrogen mustards),[7] the (mammalian) cell struggles to repair this damage,[8] requiring a complex but poorly understood process involving several enzymes and different pathways.[7,9] In cancer chemotherapy this very property is exploited to prevent proliferation of malignant cells, leading to a host of anti-tumor therapeutics such as cisplatin or nitrogen mustards[10] even though many tumors display resistance to such drugs.[11]

[a]*Center for Molecular Modeling, Ghent University, Technologiepark 903, B-9052 Zwijnaarde, Belgium, QCMM - Alliance Ghent-Brussels, Belgium. E-mail: ewald.pauwels@ugent.be*
[b]*NMR and Structure Analysis Unit, Department of Organic Chemistry, Ghent University, Krijgslaan 281 S4, B-9000 Ghent, Belgium*
[c]*Department of Pharmaceutical Science, Università degli studi di Salerno, via Ponte don Melillo, SA-84084 Fisciano, Italy*
[d]*Organic and Biomimetic Chemistry Research Group, Department of Organic Chemistry, Ghent University, Krijgslaan 281 S4, B-9000 Ghent, Belgium. E-mail: annemieke.madder@ugent.be*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3ra22408b

‡ Current address: Federal Agency for Medicines and Health Products, Eurostation building, Block 2, Place Victor Horta 40/40, B-1060 Brussels, Belgium

Short DNA oligomers function as model compounds to improve our understanding of cross-links, and furthermore show potential as biomimetic scaffolds.[12] Various strategies have been developed for synthesizing them, including—most recently—a new furan-oxidation based methodology to site-specifically cross-link nucleic acids in a very fast and efficient way.[13–16] The observed specificity is, however, still not entirely understood, requiring full structural characterization of the complete duplex structure before and after cross-linking.

Direct structural characterization of these oligonucleotides can be achieved *via* NMR analysis and site-specific labeling,[17–19] or, more cost-effectively, by sequential resonance assignment *via* proton hopping[20] based on distance restraints from the Nuclear Overhauser Effect (NOE). Unfortunately, this approach can easily fail if the dipolar couplings between protons of sequential nucleotides are missing or unresolved. The ICL region is particularly sensitive in this respect, since its canonical structure can be distorted or protons that are crucial for the assignment can be absent in the modification, leading to considerable ambiguity.

Molecular modeling is increasingly used as a complementary tool in this process. Using a molecular mechanics (MM) approach or by combining quantum mechanics with molecular mechanics (QM/MM), the spatial structure of the ICL is obtained through computational means,[18,21–27] and chemical shifts are calculated, allowing cross-reference with measurements. Several computational schemes exist to calculate chemical shifts based on a given structure. Empirical[28] or semi-empirical[29] schemes are very fast and accurate for unmodified DNA but rely on a reference set of chemical shift values and correction factors based on statistical analysis of experimental NMR data. Hence, these methods are either inherently unable to predict the chemical shifts of the ICL region, or require extensive additional parameterization. The DFT-based computation of NMR parameters—possibly in a QM/MM framework to include an extended molecular environment[30]—is not restricted by these limitations and has truly become an established tool,[31] progressively complementing the NMR experiment for structure elucidation.[32–34]

Yet, a sufficiently detailed molecular system has to be included in these simulations. The chemical nature of the molecular environment, including nucleotide sequence and solvent,[35] strongly influences the chemical shift of a nucleus. At the same time, chemically identical protons in the nucleotides show only a very small chemical shift dispersion[36] and computational methods need to be able to discriminate these as well. Finally, the flexibility of the ICL oligonucleotide and the effect of the resulting conformational dynamics on the chemical shifts need to be included in the computational protocol. Several studies on nucleotide constituents[37–39] have indicated that this leads to nontrivial changes in the chemical shifts. Inclusion of this temperature effect can be achieved by Boltzmann-weighing the NMR properties of a limited set of static conformations[38] or by averaging them for regularly sampled conformations taken from a molecular dynamics (MD) simulation at a finite temperature.[37,39]

In this work, we develop a QM/MM DFT approach to predict proton chemical shifts in (ICL) oligonucleotides that are sufficiently close to the experimental NMR data to assist structure elucidation. First, a successful computational protocol is established using an unmodified DNA oligomer as a benchmark system: the Drew–Dickerson dodecamer[40,41] d(CGCGAATTCGCG)$_2$, focusing on a central nucleotide pair. Compared with experimental NMR data,[42–46] all ingredients that are expected to substantially affect the computation of chemical shifts have been submitted to a thorough assessment: (i) the partitioning of the QM/MM scheme, (ii) the role of the solvent, and (iii) the type of calculation (static *versus* dynamic protocols). A satisfactory protocol is then selected and applied to an ICL oligonucleotide for which structural and chemical shift data are available in the literature: a mispair-aligned N^{3}T-ethyl-N^{3}T cross-linked undecamer[47] d(CGAAAT*TTTCG)$_2$. To the best of our knowledge, this is the first computational study that attempts to determine with DFT the ^{1}H chemical shifts of interstrand cross-linked oligonucleotides at this level of complexity. The derived procedure is applicable to modified oligonucleotides and DNA, and has the potential to be applicable to other biological systems.

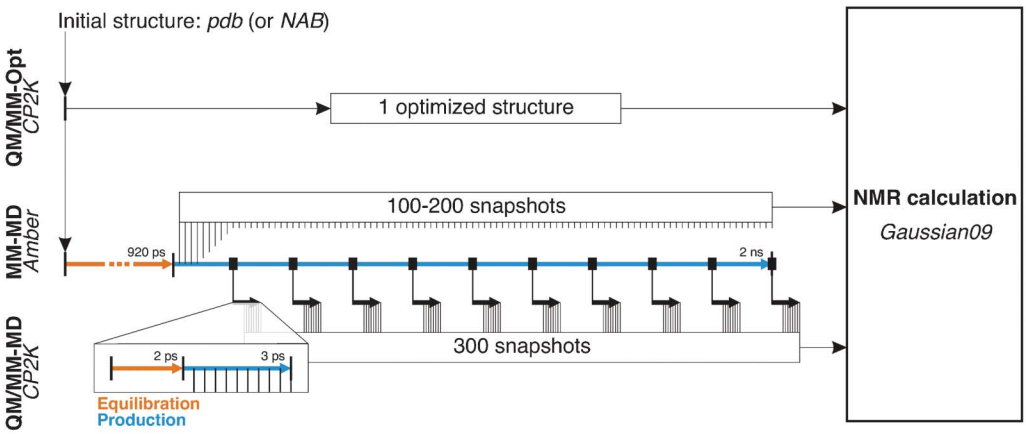
Methods and computational details

Structural models

Initial structures for both systems were taken from the pdb database: 7BNA[48] for the Drew–Dickerson dodecamer and 1S37[47] for the cross-linked d(CGAAAT*TTTCG)$_2$ undecamer. Topologies were constructed for these structures using the xLEAP interface and AmberTools 1.0.[49] The parmbsc00 refinement[50] of the Amber[51] force field was applied for all nucleic acids, whereas the cross-link modifications were specifically described by parameters taken from the general Amber force field (GAFF).[52] Hydrogens were added and each oligomer duplex was solvated in a cubic TIP3P[53] water box, with cell size 73 Å for the Dickerson dodecamer and 71 Å for the thymine mispair-aligned undecamer. Sodium counterions (22 respectively 20) were inserted in the water box to ensure charge neutrality. Scheme 1 illustrates the different computational procedures that were subsequently applied to these structures.

MM calculations

Unrestrained MD simulations were performed after an equilibration protocol similar to that of ref. 54. Energy minimizations were performed on the initial structures, first with harmonic restraints (100 kcal mol^{-1} Å^{-2}) on all nucleic acid atoms, then without restraints. Each system was subsequently heated to 300 K during a 20 ps NVT MD run, again with restraints on the oligomer. These were gradually lifted during a subsequent 150 ps NPT MD simulation at 1 bar. Box dimensions were scaled corresponding with the average volume from the last 40 ps, followed by another MD run of 150 ps in the NVT ensemble. The velocities from the last



Scheme 1 Flowchart schematically illustrating the setup of the computational methodologies to calculate proton chemical shifts in interstrand cross-linked DNA. Energy minimization and molecular dynamics simulations were conducted with Amber[55,56] and CP2K (http://www.cp2k.org). NMR properties were calculated with Gaussian09.[71] Initial structures for each system were taken from the pdb databank, but in the general case could alternatively be generated using Nucleic Acid Builder (NAB), part of AmberTools.[49]

snapshot were rescaled to correspond to 300 K and used as a starting point for an NVE equilibration run of 600 ps. Finally, a 2 ns NVE production MD simulation was performed and snapshots – for use in later NMR calculations – were collected every 20 ps (10 ps in the dodecamer case). The average temperature during these runs was 299.2 K for the Drew–Dickerson dodecamer and 291.4 K for the d(CGAAAT*TTTCG)$_2$ undecamer; the average root-mean-square-deviation (RMSD) with respect to the crystal structures (7BNA and 1S37) was 4.06 Å, respectively 4.02 Å.

All force-field simulations were conducted under periodic boundary conditions with the AMBER 9 molecular simulation package.[55,56] A 15 Å cutoff was employed for the particle mesh Ewald summations[57] of nonbonding interactions. During MD simulations, the SHAKE algorithm[58] was applied to bonds involving hydrogen atoms, allowing a 2 fs time step. Temperature was regulated using Langevin dynamics[59] with a collision frequency of 1.0 ps^{-1}. Pressure regulation was achieved with isotropic position scaling, and the pressure relaxation time was set to 2.0 ps.

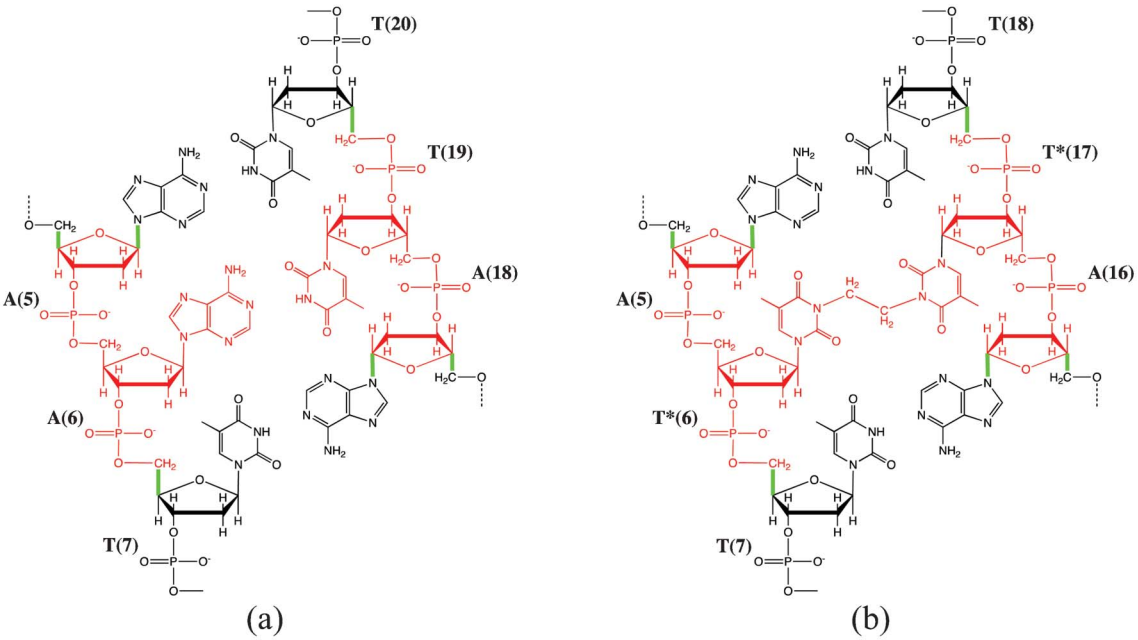
QM/MM calculations

The initial structures for all molecular systems were also subject to full energy minimization in a QM/MM scheme, employing the quasi-Newton LBFGS algorithm.[60] The resulting structures deviated from the corresponding crystal structures (7BNA and 1S37) with an RMSD of 3.21 Å for the Drew–Dickerson dodecamer and 3.45 Å for the d(CGAAAT*TTTCG)$_2$ cross-linked undecamer. In addition, MD simulations were performed for the latter system. However, unlike force-field MD simulations, the ns timescale is beyond reach of QM/MM methods at an affordable computational cost. To still include slower molecular motions to some extent, 10 QM/MM MD simulations were initiated in parallel, starting from 10 snapshots regularly sampled along the 2 ns MM production run (*i.e.* every 200 ps). In each of these MD runs, temperature was first equilibrated to 300 K during 2 ps, followed by a 3 ps production run, totaling 50 ps of QM/MM MD simulations. The production runs were sampled every 100 fs, generating in all 300 snapshots for later use in the NMR calculations. This methodology is illustrated in Scheme 1.

All QM/MM simulations were performed with the CP2K software (http://www.cp2k.org). For each molecular system, the (30.0 Å)^3 QM region contained the central or cross-linked nucleotide pair, at the 3′ end the next phosphate and –C(5′)H$_2$– groups, and at the 5′ end the next deoxyribose (except the –C(5′)H$_2$– group). This partitioning is illustrated in Scheme 2. A link atom approach was used to saturate with hydrogen the six broken bonds spanning the QM/MM boundary. The interaction between the MM and QM layers was calculated at the QM level, using the Gaussian expansion of the electrostatic potential method (GEEP) as developed by Laino *et al.*[61] For the QM layer, the PBE density functional[62] was used in combination with the Gaussian and plane waves (GPW) dual basis set method.[63] Norm-conserving GTH pseudopotentials[63–65] were expanded up to a density cutoff of 280 Ry and Gaussian triple-ζ basis sets[66] augmented with polarization functions were employed for H, C and N (TZVP), and for O and P (TZV2PX). For the MM layer, the parmbsc00 refinement[50] of the Amber[51] force field was maintained. A timestep of 1 fs was used throughout all QM/MM MD simulations, which were performed in the NVT ensemble. To regulate temperature, separate thermostats were applied to the QM and MM regions, employing a fast constant velocity rescaling algorithm[67] during the first 1 ps and the Nosé–Hoover algorithm[68,69] for the remainder of the MD simulation. A time constant of 10 fs was used in the former method, 100 fs in the latter. A Nosé–Hoover thermostat chain of length 5 was applied.

NMR calculations

Chemical shifts were calculated for QM/MM optimized structures and for snapshots collected during MM and QM/MM MD simulations. Initial attempts to also perform these calculations within the periodic scheme of CP2K[70] proved computationally challenging and unsatisfactory, after which a



Scheme 2 Adopted QM/MM partitioning scheme for energy minimizations and MD simulations on Drew-Dickerson dodecamer (a) and mispair-aligned N^3T-Ethyl-N^3T cross-linked undecamer d(CGAAAT*TTTCG)$_2$ (b). All atoms in red are included in the QM layer. Bonds indicated in green span the QM/MM partitioning and are saturated with hydrogen atoms in a link atom approach. A bond-scaling factor of 1.4 was used for all broken C–C bonds, a factor of 1.35 for all broken C–N bonds.

cluster approach was followed. To this end, each of the MM or QM/MM structures was again partitioned in a QM subsystem and MM subsystem. Except when otherwise mentioned, the former included the crosslink modification (if present), the central nucleotide pair, the three neighbouring nucleotide pairs on either side and all sodium ions or water molecules (typically about 30) closer than 2.3 Å to any atom of the central nucleotide pair. Terminal phosphate groups were replaced by hydroxyl groups and hydrogen link atoms were added to saturate the dangling bonds of the QM subsystem. Included in the MM subsystem were all sodium ions or water molecules (of the order of 2200) within a 25 Å radius of the center of mass of the central nucleotide pair, except those already taken up in the QM subsystem.

The Gaussian09 software package[71] was used to calculate the chemical shieldings, employing gauge including atomic orbitals (GIAO)[72] and the mPW1PW91[73] density functional in combination with the 6–31 g(d,p) basis set[74–76] for all atoms of the QM subsystem. This particular level of theory was shown to yield satisfactory results at a reasonable computational expense.[77] All atoms in the MM subsystem were represented by point charges, taken from the Amber parmbsc0 force field refinement.[50]

Proton chemical shifts were calculated as $\delta = \sigma(\text{ref}) - \sigma$, in which σ is the isotropic component of the chemical shielding tensor of a particular proton and σ(ref) is a reference ^{1}H shielding constant. Frequently, this latter value is obtained from a separate DFT calculation on tetramethylsilane (TMS), which is usually used as a reference compound. However, this approach depends critically on an accurate calculation for TMS and can easily suffer from systematic errors because of the computational method.[78] Hence, a reference proton shielding constant was determined in each case *via* linear regression of the experimental shifts and calculated shieldings for all molecular systems, adopting an approach similar to that of ref. 79 and 80. These values range from 31.3 to 31.8 ppm (see further on) and are quite comparable to the shielding values obtained from separate DFT calculations on a tetramethylsilane molecule.

For comparison, proton chemical shifts were also calculated using the (semi)-empirical methods of Altona[28] and Wijmenga.[29] To this end, the sequences of the Drew–Dickerson dodecamer and the cross-linked undecamer were submitted to the DSHIFT server.[81]

Results and discussion

Benchmarking the NMR computational scheme

As mentioned earlier, calculated proton chemical shifts are known to be very sensitive to the extent of the molecular environment that is included in the calculation.[35] Ideally, one should account for as much as possible of the true molecular system. Of course, a larger molecular environment in the calculation increases the computational cost substantially. A careful benchmark is therefore needed, in which the extent of the environment is gradually increased until near-convergence of the calculated properties is reached. Here, such benchmarking was performed for the QM/MM optimized structure of the Drew–Dickerson dodecamer, focusing on the central A(6)–T(19) basepair, as schematically shown in Scheme 2. Various calculations of the proton NMR chemical shieldings were performed using the Gaussian09 software[71] (adopting a

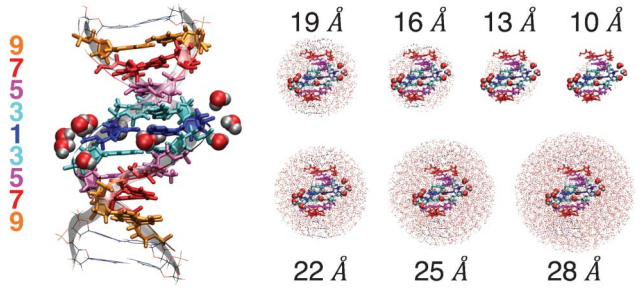


Fig. 1 QM/MM optimized structure of the Dickerson–Drew dodecamer, illustrating the variations in the QM/MM partitioning for the subsequent calculation of proton chemical shieldings.

procedure similar to that outlined in the Methods and Computational Details section) in which the composition of the QM and MM subsystems was varied, as illustrated in Fig. 1. Three main alterations were examined:

(a) the number of nucleotide pairs, neighboring the central nucleotide pair, taken up in the QM subsystem (ranging from 1-blue to 9-orange),

(b) whether or not water molecules and sodium ions in close proximity (within 2.3 Å) of the central nucleotide pair were taken up in the QM subsystem (space-filling models in the left part of Fig. 1),

(c) the amount of point charges in the MM subsystem. As a measure, the radius of the sphere centered at the center of mass of the central nucleotide pair is given (ranging from 10 to 28 Å).

The resulting variation of the proton chemical shieldings for the central adenosine (6) as a function of the QM/MM composition in the NMR calculation is displayed in Fig. 2 (a similar dependence is observed for the central thymidine—not shown).

Disregarding at first the medium and point charge description of the MM subsystem, gradual convergence of the chemical shielding values is observed when increasingly more neighboring nucleotide pairs are taken up in the QM subsystem (illustrated in the left part of Fig. 2). Proton shifts in DNA are very sensitive to the aromatic shielding of the neighboring bases, and here especially the H2′ and H2′′ deoxyribose shieldings, and also H62 of the nucleobase seem to be strongly affected (consider, *e.g.*, the variation in size for 1 to 5 nucleotide pairs). Although convergence quickly sets in, the mean absolute error (MAE, relative to the results of the most elaborate model) only drops below 0.1 ppm once 7 nucleotide pairs are included. This corresponds to the methodology of the very first empirical ^{1}H chemical shift prediction methods[82,83] where also the three nearest nucleotide pairs on 3′ and 5′ side are taken into account.

Long and short-range interactions with the medium also affect the chemical shieldings. Considering only the electrostatic interactions, by incorporating point charges within 19 Å to the center of mass of the central nucleotide pair, clear effects on the polarizable protons are apparent. As illustrated in the right part of Fig. 2, the chemical shieldings are

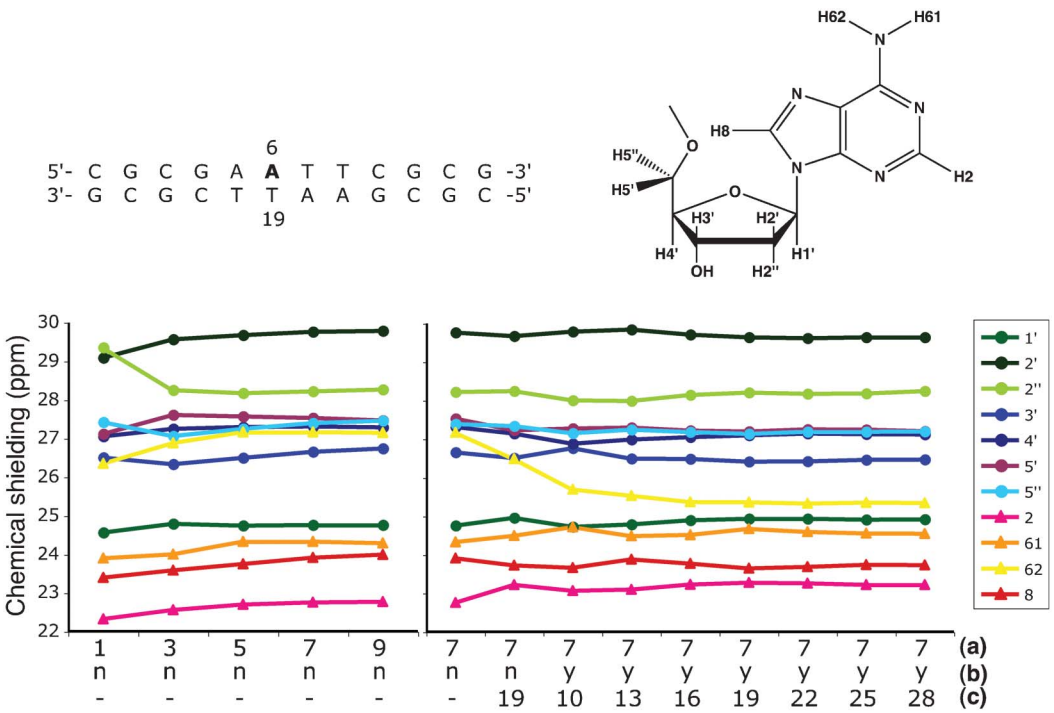


Fig. 2 Fluctuation of central adenosine (6) proton chemical shieldings as a function of QM/MM partitioning in the NMR calculation. (a), (b) and (c) refer to the number of nucleotide pairs, the inclusion (yes/no) of medium in the QM subsystem, and the radius (in Å) of the inclusion sphere for MM point charges, respectively (see text). The sequence of the Dickerson–Drew dodecamer, position of the central nucleotide pair, chemical structure and proton numbering scheme for adenosine are given at the top.

considerably shifted with respect to the previous 7-nucleotide calculation, with H2 and H62 deviating the most. Yet, the prime effect of the medium—hydrogen bonding—is not merely electrostatic in nature.[84] This is clearly visible when incorporating those water molecules in the QM subsystem that are involved in hydrogen bonds with the central A(6)–T(19) nucleotide pair. The H62 shielding undergoes the most spectacular shift (−1 ppm), as it is involved in a tight hydrogen bond (2.11 Å) with a water molecule (see Fig. 1). Extending (or reducing) the inclusion of longer-range electrostatic interactions on top of the explicit water description, by altering the radius of the sphere of point charges in the MM subsystem, has a much less pronounced effect. Convergence is also observed for radii larger than 16 Å, which describes the first sphere fully surrounding all 7 nucleotide pairs of the QM subsystem (see Fig. 1). Given the low computational cost of adding point charges to an NMR calculation, selecting a sphere radius of 25 Å for the MM subsystem is entirely feasible, while minimizing the MAE (0.02 ppm) and maximum error (0.06 ppm) with respect to the most elaborate model system (at the far right of Fig. 2). A reduction of the number of nucleotide pairs taken up in the QM subsystem can not be compensated by a more extended point charge description in the MM subsystem (*e.g.* 5 nucleotide pairs and a 19 Å spherical MM subsystem). These schemes systematically lead to much larger MAEs.

The benchmarking for the Drew–Dickerson dodecamer indicates that a good trade-off between extent of the model system and computational expense is reached for a computational scheme in which 7 neighboring nucleotide pairs and the closest surrounding water of the central pair are taken up in the QM subsystem, combined with a MM subsystem comprising point charges for all atoms within a 25 Å radius of the center of mass of the central nucleotide pair. All further NMR calculations were conducted with this scheme (see Methods and Computational Details).

Drew–dickerson dodecamer: comparing theory and experiment

In the previous section, the chemical shieldings were compared between various computational models, *i.e.* with respect to a benchmark in which the molecular model was systematically extended to approximate the true molecular environment. An even more stringent test for the computational protocol is its ability to reproduce the experimental chemical shifts. This comparison is taken up in Table 1 and in Fig. 3 for all proton chemical shifts of the central A–T nucleotide pair. Comparison is made with the experimental data of ref. 43 and 45. For the QM/MM optimized structure (blue dots in Fig. 3), the agreement between theory and experiment is quite satisfactory, but deviations of 0.5 ppm and more occur in the region between 2–3 ppm and 6–9 ppm. The prediction correlates well with the experimental data and a linear fit is characterized by an R^2 coefficient of determination of 0.974, quite close to the diagonal. The MAE with respect to the experimental data is 0.32 ppm, largely due to the poor prediction of the chemical shifts for the H1′, H2′ and H2′′ carbohydrate protons and H2 of the adenine.

In part, these differences are the result of the static nature of the calculations, which only consider one point on the potential energy surface. The oligonucleotide conformational dynamics are not accounted for, whereas this does have an impact on the NMR chemical shifts. The sugar ring conformation considerably influences the deoxyribose shifts[85] and nucleobase protons are heavily affected by these dynamics and the interaction with the solvent.[39] To account for these

Table 1 Overview of calculated and measured chemical shift data for the central A(6)–T(19) nucleotide pair of the Drew–Dickerson dodecamer. All values are in ppm, except for R^2 (dimensionless). For the dynamic data, the standard deviation is given between parentheses. All experimental chemical shifts were taken from ref. 45 except for H5′ and H5′′, where the values reported in ref. 43 were used

		Experiment[43,45]	QM/MM-opt	MM-MD	Wijmenga[29]	Altona[28]
	H5′	*4.19*	4.42	4.30 (.35)	4.17	4.28
	H5′′	*4.24*	4.47	4.15 (.41)	4.13	4.22
	H4′	*4.44*	4.53	4.39 (.41)	4.27	4.44
	H1′	*6.13*	6.75	6.51 (.49)	6.14	6.12
A(6)	H8	*8.07*	7.92	7.99 (.65)	8.15	8.12
	H2	*7.61*	8.44	8.09 (.46)	7.80	7.58
	H3′	*4.98*	5.20	5.32 (.41)	4.92	4.99
	H2′	*2.54*	2.02	2.59 (.47)	2.58	2.54
	H2′′	*2.88*	3.47	2.97 (.44)	2.77	2.88
	H5′	*4.14*	4.39	4.33 (.34)	4.02	4.32
	H5′′	*4.14*	4.37	4.31 (.40)	3.98	4.20
	H4′	*4.11*	4.45	4.25 (.44)	4.17	4.22
	H1′	*5.87*	6.33	6.12 (.47)	5.95	5.95
T(19)	H6	*7.07*	7.02	7.50 (.58)	7.30	7.19
	H3′	*4.78*	4.82	4.99 (.37)	4.72	4.84
	H2′	*1.95*	1.64	1.95 (.46)	2.08	2.00
	H2′′	*2.53*	3.02	2.74 (.48)	2.47	2.53
	H5(CH3)	*1.26*	1.35	1.39 (.32)	1.50	1.35
	R^2		0.974	0.994	0.996	0.999
	MAE		0.32	0.19	0.11	0.05
	σ(ref)		31.67	31.78		

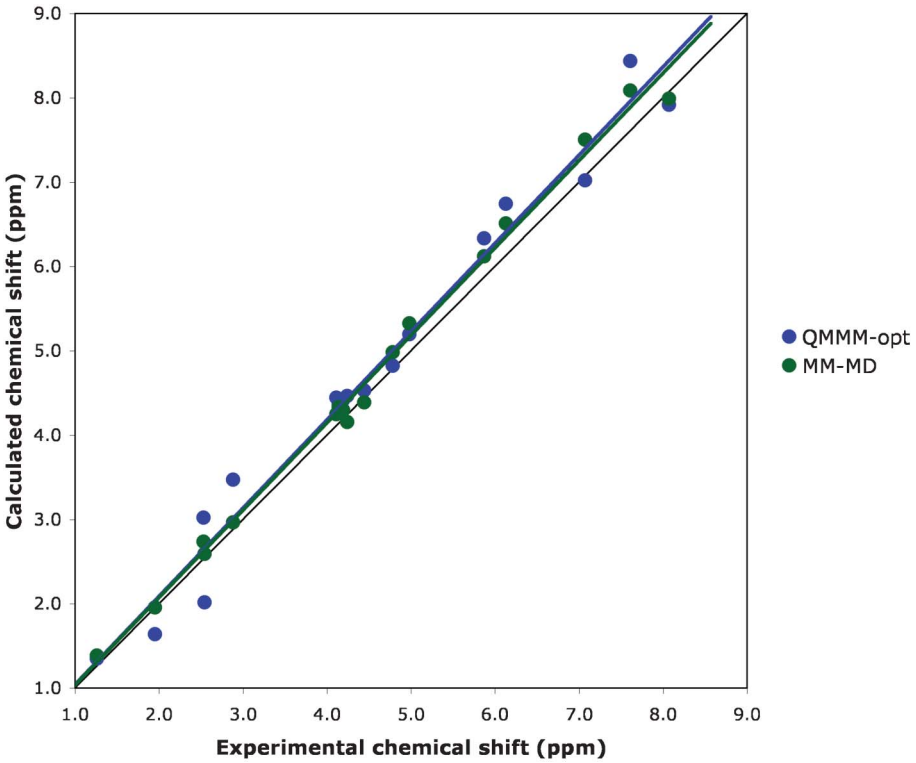


Fig. 3 Correlation between theoretical and experimental proton chemical shifts for the central A–T nucleotide pair of the Dickerson–Drew dodecamer. *Via* linear regression, a reference proton shielding constant of 31.67 ppm was determined for the QM/MM optimized structure (blue), 31.78 ppm for the results of the MM dynamics (green).

typical motions (on the nanosecond timescale) and their impact on the chemical shifts, NMR calculations were performed on 200 snapshots from a 2 ns MM MD simulation (as outlined in Scheme 1). The resulting average shifts are also indicated in Fig. 3 (green dots) and Table 1. These calculated values present an overall better agreement with experiment: the MAE with experiment is reduced to 0.19 ppm and the R^2 coefficient is 0.994. Only one proton still deviates considerably from the linear fit—adenine H8—but all others correspond much better to the fit than in the NMR calculation for the QM/MM optimized geometry. The slope of the linear fit is 1.0367, which indicates an increasing systematic error downfield. In principle, this could be easily improved by scaling—commonly used[86]—but this was not done here, allowing an unadulterated assessment of the adopted computational scheme.

The variation throughout the dynamics of the calculated chemical shift values for the central adenosine is illustrated in Fig. 4. The distribution of shift values is broad for all protons, with standard deviations ranging from 0.35 ppm for the H5′ proton to 1.14 ppm for the H62 amine proton (see Table 1). In fact, both amine protons display an extremely broad chemical shift distribution. This is caused by rapid fluctuations in the solvent environment of these protons, and the resulting variations in hydrogen bonding. But also non-exchangeable protons can suffer from this effect: the same applies for the aromatic H8 and—to a lesser extent—H2 protons of the nucleobase, as already noted in another study.[39]

The dynamics considerably improve the agreement between theory and experiment for the deoxyribose protons, especially in the region 1–3 ppm. Clearly, the snapshots taken from the MD represent a sufficient conformational ensemble to describe the rapid ring puckering and fluctuations in the exocyclic torsion angles, which are known to influence the sugar proton chemical shifts.[85] In Fig. 5, the dependence of the H1′ chemical shift of adenosine (6) on the deoxyribose ring conformation is illustrated. The sugar ring does not assume just one conformation (as obtained from the static QM/MM optimization), but rather switches between the south (C2′-*endo*) and north (C3′-*endo*) conformations, with the former clearly predominating. In this C2′-*endo* conformer, the H1′ chemical shift attains rather high values (7–8 ppm) but the less populated conformations attenuate the shift. Accounting for the full conformational ensemble obtained from the MM-MD simulations causes the H1′ shift to drop (from 6.75 to 6.51 ppm), yielding a better agreement with experiment (6.13 ppm). The aromatic protons of the nucleobases (adenine H8, H2 and thymine H6) remain poorly predicted by the dynamics with deviations of up to 0.4 ppm from the experimental values. This method even seems to perform worse for H6 of T(19). However, the MM-MD value for this proton is consistent with the linear fit (Fig. 3), contrary to the QM/MM-opt result. Adenine H8, on the other hand, falls outside the linear fit, even though the predicted value is almost identical to the experiment shift.

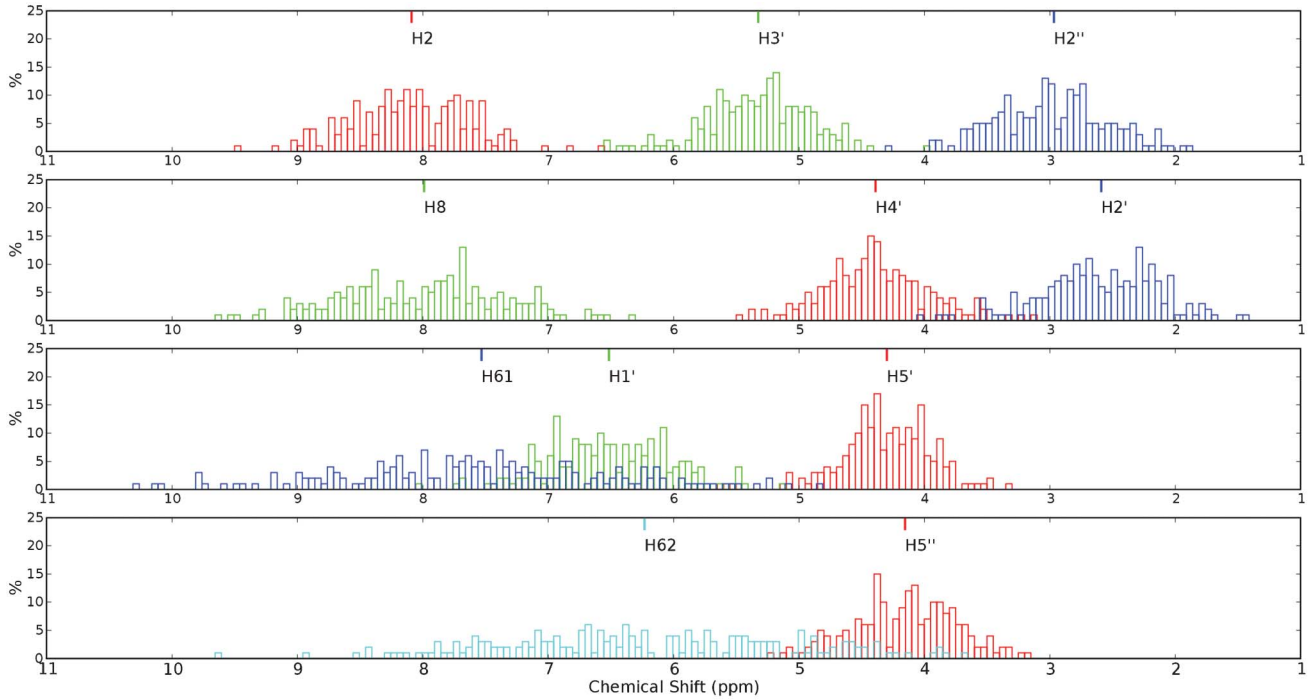


Fig. 4 Histograms of the proton chemical shifts for the central adenosine (6) residue in the Dickerson–Drew dodecamer, calculated for 200 snapshots from a 2 ns MM-MD simulation. Averaged values for each distribution are shown at the top of the plots.

Overall, the MAE with experiment for the MM-MD results (0.19 ppm) is almost on par with that for the chemical shifts predicted with the Wijmenga[29] approach (see Table 1). This method (with a standard deviation of about 0.17 ppm) relies on the semi-empirical relations of Giessner-Prettre and Pullman[87] but is only applicable for canonical B-DNA. Even

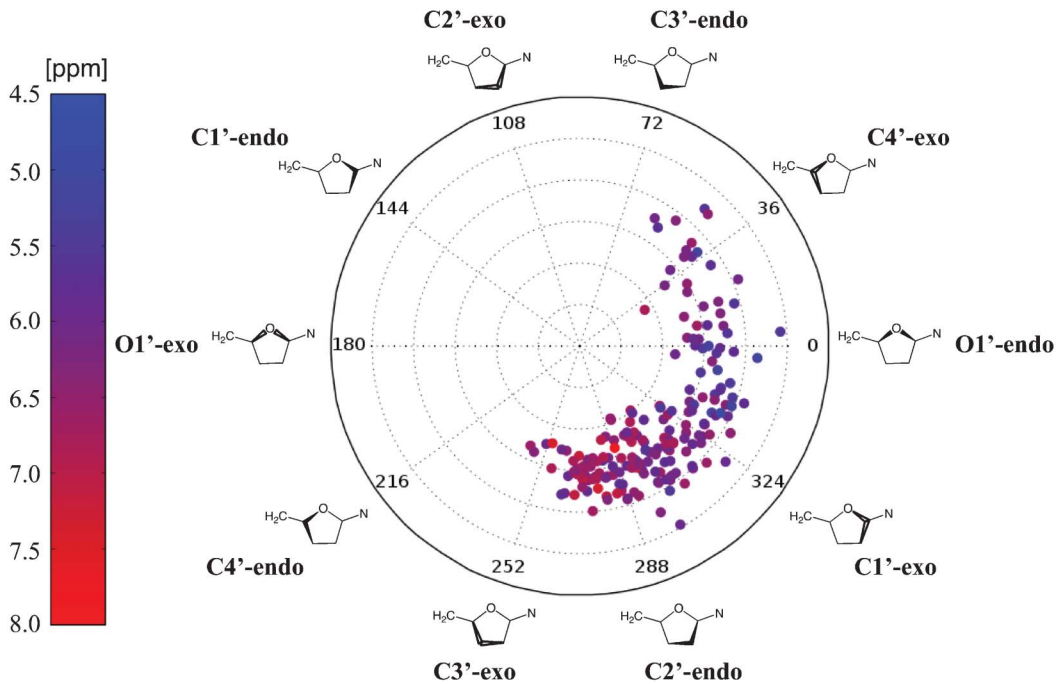


Fig. 5 Polar plot of the puckering coordinates for the deoxyribose in the adenosine(6) residue of the Dickerson–Drew dodecamer throughout 200 snapshots of a 2 ns MM-MD run. Representative stages along the pseudorotation pathway are indicated at the corresponding phase angles (in degrees). The radius is the puckering amplitude (from 0.0 to 1.0 Å). Data points are colored according to the chemical shift value of H1′. Polar coordinates were calculated with the MD-TRACKS toolkit.[88]

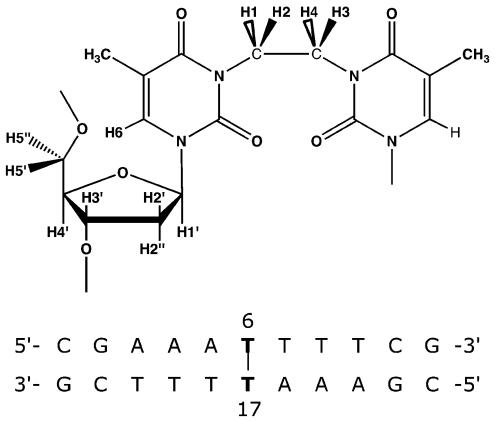


Fig. 6 Sequence and position of the N^3T-ethyl-N^3T cross-link in the undecamer. Chemical structure and atom numbering scheme of the cross-link region.

though the MAE is 0.11 ppm, larger deviations persist for the aromatic nucleobase protons, analogous with the MM-MD results. The empirical Altona approach[28] yields an even closer agreement with experiment, further reducing the MAE to only 0.05 ppm. This is quite logical since it is based on statistical analysis of experimental shifts for reference state B-DNA forms, of which the Drew–Dickerson dodecamer is a prototypical example.

N^3T-ethyl-N^3T cross-linked undecamer

The grand test for our computational methodology is its performance in predicting chemical shifts of novel chemical modifications to DNA, such as interstrand cross-linked oligonucleotides. Few, if any, modeling methodologies are able to do so and remain universally applicable. The benchmarked computational NMR methodology was therefore applied to the mispair-aligned N^3T-ethyl-N^3T cross-linked undecamer[47] d(CGAAAT*TTTCG)$_2$ for which the structure and sequence is shown in Fig. 6. NMR calculations were performed on three types of structures: on the QM/MM optimized geometry, on 100 snapshots taken from a 2 ns MM-MD trajectory, and on 300 snapshots taken from ten 3 ps QM/MM-MD trajectories (see Scheme 1). The latter set of extensive simulations was specifically performed to examine whether the cross-link region is properly described by the GAFF[52] force field parameters (see Supporting Information†). This cannot *a priori* be assumed since the parameters were not calibrated for these specific cross-linked sequences. A QM/MM description of this region is therefore much less biased by the choice of parameters than an MM description. In addition, as previous results indicate that proper account of molecular flexibility is of utmost importance, additional QM/MM molecular dynamics runs were performed. In Fig. 7 and Table 2, the calculated proton chemical shifts are compared to the measured NMR data.[47] The simulations confirm our earlier findings that the static description of the oligonucleo-

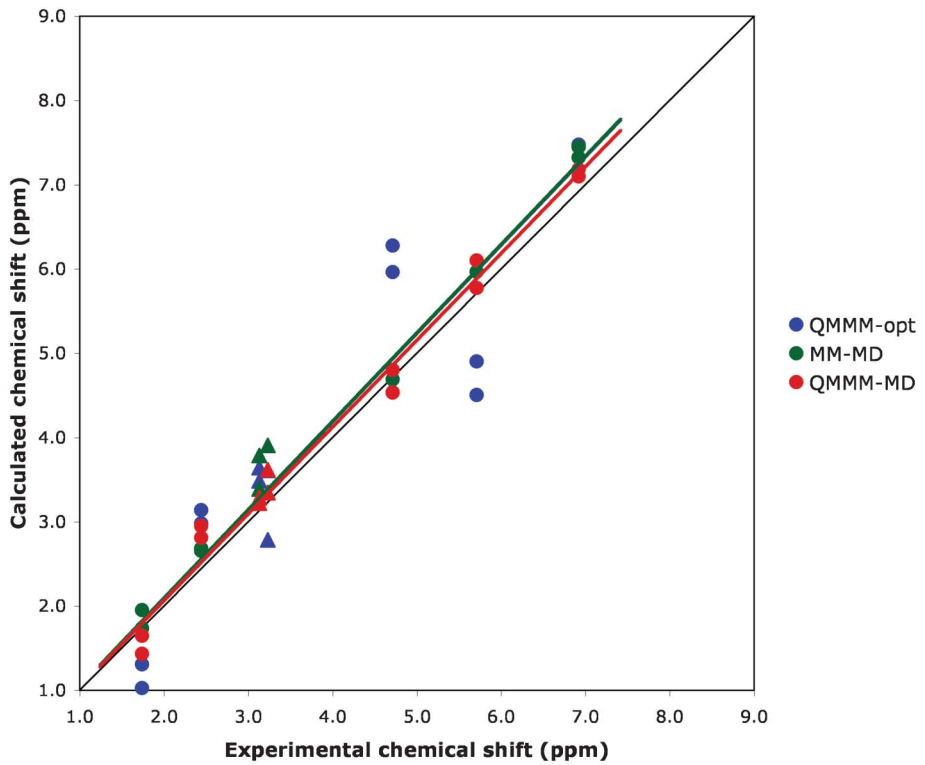


Fig. 7 Correlation between theoretical and experimental proton chemical shifts for the central cross-linked T–T pair in the d(CGAAAT*TTTCG)$_2$ undecamer. The ethyl proton chemical shifts of the cross-link are represented by triangles. *Via* linear regression, a reference proton shielding constant of 31.64 ppm was determined for the QM/MM optimized structure (blue), 31.77 ppm for the results of the MM dynamics (green), and 31.29 ppm for the QM/MM dynamics (red).

Table 2 Overview of calculated and measured chemical shift data for the central T–T nucleotide pair of the N^3T-ethyl-N^3T cross-linked undecamer. All values are in ppm, except for R^2 (dimensionless). For the dynamic data, the standard deviation is given between parentheses

		Experiment[47]	QM/MM-opt	MM-MD	QM/MM-MD	Wijmenga[29]	Altona[28]
T*(6)	H1′	*5.71*	4.51	5.97 (.47)	5.78 (.89)	5.95	5.95
	H6	*6.92*	7.48	7.32 (.48)	7.10 (.79)	7.30	7.19
	H3′	*4.71*	5.97	4.54 (.49)	4.53 (.72)	4.72	4.84
	H2′	*1.74*	1.03	1.73 (.65)	1.43 (1.06)	2.08	2.00
	H2′′	*2.44*	2.98	2.66 (.42)	2.95 (1.10)	2.47	2.53
T*(17)	H1′	*5.71*	4.90	6.10 (.46)	6.10 (.78)	5.95	5.95
	H6	*6.92*	7.46	7.45 (.52)	7.18 (.90)	7.30	7.19
	H3′	*4.71*	6.28	4.69 (.49)	4.80 (.76)	4.72	4.84
	H2′	*1.74*	1.31	1.95 (.52)	1.65 (.92)	2.08	2.00
	H2′′	*2.44*	3.14	2.69 (.40)	2.81 (1.02)	2.47	2.53
ICL	H1	*3.13*	3.48	3.39 (.48)	3.28 (.98)		
	H2	*3.23*	3.61	3.90 (.52)	3.62 (1.08)		
	H3	*3.13*	3.64	3.78 (.73)	3.22 (1.03)		
	H4	*3.23*	2.79	3.35 (.49)	3.34 (1.07)		
	R^2		0.845	0.983	0.985	0.994	0.999
	MAE		0.71	0.30	0.23	0.20	0.20
	σ(ref)		31.64	31.77	31.29		

tide results in large errors for the predicted chemical shifts, for deoxyribose protons (H1′ and H3′ in particular) but also for the ethyl protons of the cross-link (H1 to H4). For non-cross-link protons, MM as well as QM/MM dynamics considerably improve the agreement with experiment. The cross-link protons require a QM/MM-MD treatment (see further on). Overall, the MM-MD and QM/MM-MD results yield MAE values of 0.30 ppm and 0.23 ppm respectively, compared to 0.72 ppm for the QM/MM optimized structure. The linear fits are also more consistent in the dynamic cases, with R^2 values much closer to 1.00 as compared to the data of the QM/MM optimized structure.

A striking feature in Fig. 7, for all calculation methods, is that the chemical shifts of equivalent protons in both thymine nucleotides are different. In reality, these protons are chemically identical, yielding the same chemical shifts and giving rise to a deposited pdb-structure (1S37)[47] that is perfectly symmetrical. Such a situation never arises in our structural calculations, partly because the solvent and ionic environment of the oligomer is inherently asymmetric, but

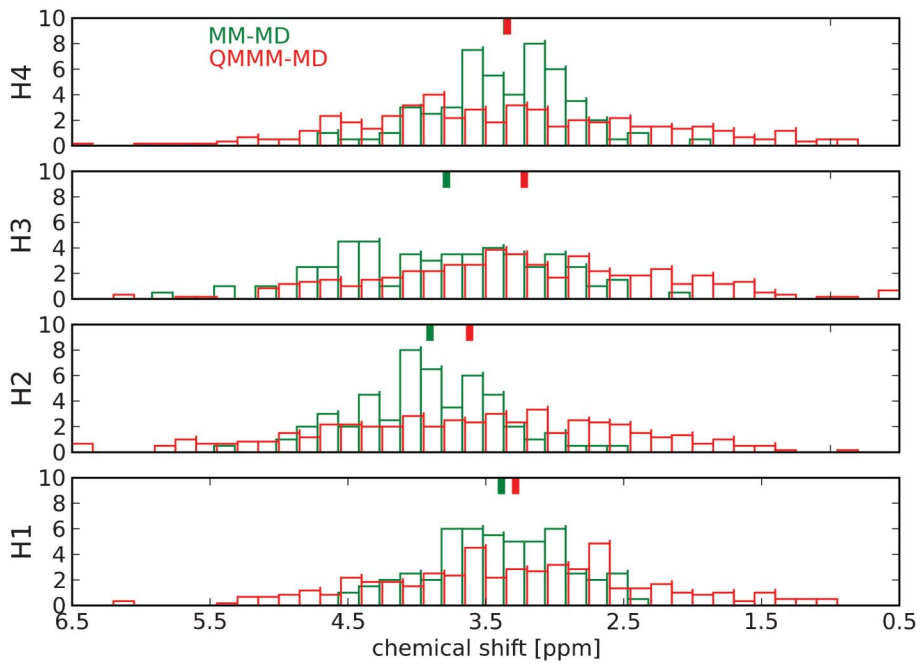


Fig. 8 Histograms of the cross-link ethyl proton chemical shifts for the ICL d(CGAAAT*TTTCG)$_2$ undecamer, calculated for 100 snapshots from a 2 ns MM-MD simulation (green) and 300 snapshots from ten 3 ps QM/MM-MD trajectories (red). Averaged values for each distribution are shown at the top of the plots.

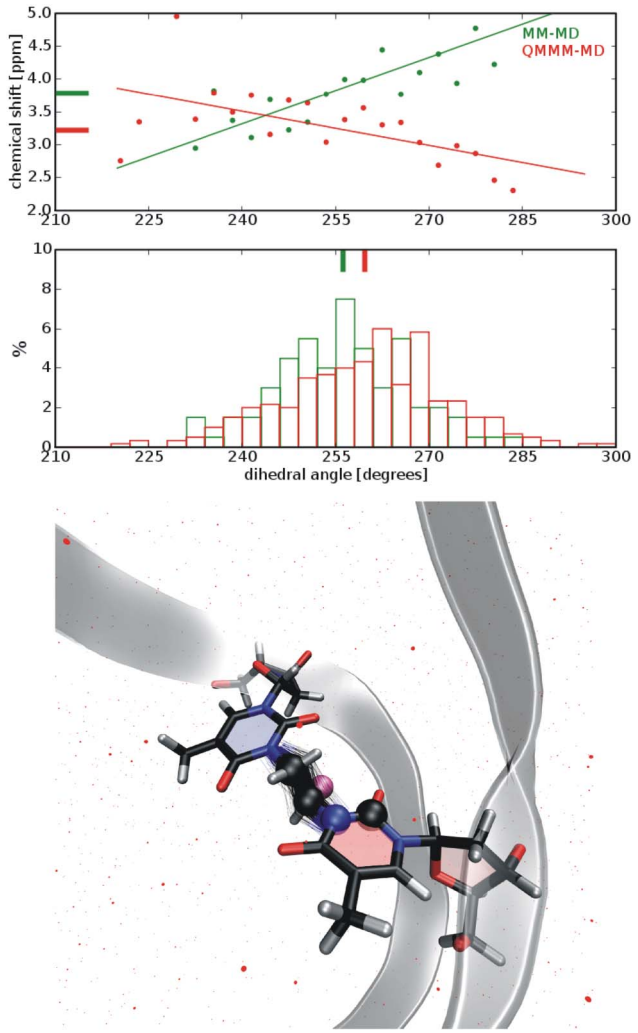


Fig. 9 (Center) Histogram of the C2(T17)-N3(T17)-C(ICL)-C(ICL) dihedral angle values throughout the MM-MD (green) and QM/MM-MD simulations (red). Averaged dihedral angle values for each distribution are shown as lines at the top of the plot. (Top) Influence of this dihedral angle on the H3 chemical shift. Dots indicate the shift values obtained by averaging over all structures in the corresponding dihedral angle bins. Linear fits merely illustrate the trend for the MM-MD and QM/MM-MD simulations. Grand average H3 chemical shifts (as reported in Table 2) are also shown by thick horizontal lines on the left of the plot. (Bottom) Structural representation of a snapshot from the QM/MM-MD simulations, zooming in on the T-ethyl-T cross-link. Spheres indicate the C2(T17)–N3(T17)–C(ICL)–C(ICL) dihedral; the H3 proton is marked by a smaller, purple sphere. Thymine-6 is colored blue, thymine-17 in red. Transparent lines illustrate the conformational flexibility of the ethyl crosslink throughout the dynamics. The oligonucleotide backbone is represented by a grey ribbon, water molecules by red spheres.

mostly because very long MD trajectories would be required to reach completely symmetric conformational ensembles for both sides of the oligomer. One way to circumvent this would be to average the chemical shifts of equivalent protons over both thymine residues (T(6) and T(17)), yielding smaller MAE values with respect to experiment. However, this was not done here, since such a procedure is not possible for asymmetrically cross-linked DNA oligomers.

The overall performance of the MM and QM/MM dynamic methods is similar in predicting the chemical shifts of non-cross-link protons. Neither outperforms the other, with larger deviations from the measurements for the aromatic nucleobase protons in the former and for some carbohydrate protons in the latter dynamic method. Comparable errors are now also apparent for the (semi)-empirical methods (see Table 2), which do not take into account the deviation from canonical B-DNA structure in the cross-linked undecamer. Furthermore, no chemical shift values are readily available for the interstrand ethyl cross-link using the Wijmenga or Altona approaches.

Our computational NMR methodology also has the ability to predict the proton chemical shifts of this interstrand ethyl cross-link. The corresponding values (H1 to H4) are taken up in Table 2. For the cross-link region specifically, the results of the QM/MM dynamics are much better than those of the MM dynamics, especially for the H2 and H3 protons pointing towards the major groove. Fig. 8 compares the variation of the chemical shifts for all ICL protons for both types of molecular dynamics simulations. The resulting distributions are very broad, easily spanning 4–6 ppm, comparable to the amine protons of the central adenosine in the Drew–Dickerson dodecamer (see Fig. 4). The chemical shift of the ethyl protons is very sensitive to the dynamic variations of the cross-linked structure, but also to the surrounding solvent, despite their apolar nature. Clearly, the varying hydrogen bonding pattern with the nearby aromatic nitrogens and carbonyl groups throughout the dynamics leads to microsolvation of the ICL region, with a concomitant effect on the ethyl proton chemical shifts. The width of the MM-MD distributions is more compact, but the predicted H2 and H3 chemical shifts are also apparently higher than in the QM/MM-MD case. This clear distinction indicates that the conformational space of the cross-link is fundamentally different in the MM and QM/MM descriptions. As such, different preferential conformations are attained during molecular dynamics, affecting the resulting proton chemical shifts. This is explicitly illustrated in Fig. 9 for the C2(T17)–N3(T17)–C(ICL)–C(ICL) dihedral and the H3 chemical shift (the other ICL protons show comparable, but less pronounced trends). Histograms of the dihedral angle values during both MD simulations show that the QM/MM-MD distribution is biased towards a slightly higher dihedral angle, as compared to the MM-MD distribution: the preferred conformation for this dihedral is different in both simulations. This conformational change in itself will affect the H3 chemical shift. More fundamentally though, the correlation between the H3 chemical shift and the C2(T17)–N3(T17)–C(ICL)–C(ICL) dihedral is significantly different in the QM/MM and MM methodologies. This is not easy to interpret: the reported correlations are rather low (0.48 for MM-MD and −0.33 for QM/MM-MD) and many more structural variables will affect the C2(T17)–N3(T17)–C(ICL)–C(ICL) dihedral and H3 shift. Nevertheless, this difference further points out the fundamentally different description of the ICL conformation in the MM and QM/MM schemes. The latter method, though computationally more expensive, seems best fit to predict and reproduce the experimental chemical shifts in the ICL region.

Conclusions

An efficient way to obtain structural insight is duly needed to promote development of novel applications of modified DNA. Solving complications encountered in NMR structure analysis when coping with oligonucleotides of increasing size can be facilitated by complementary use of molecular modeling, yielding structural as well as spectroscopic information through simulation. Yet, computational prediction of NMR chemical shifts remains challenging, especially for molecular systems as extensive as (modified) DNA. We have developed a computational methodology relying on first principles and a QM/MM partitioning scheme that is able to satisfactorily predict the proton chemical shifts of (modified) DNA systems, significantly facilitating structural elucidation using NMR spectroscopy. For instance, in interstrand cross-linked oligonucleotides, our method could provide a complete assignment of all chemical shifts related to the cross-link region, allowing gaps in a sequential resonance assignment to be filled in. For such molecular systems in particular, a QM/MM methodology is essential, since standard (semi-) empirical chemical shift prediction methods are less accurate and a better description of the electronic structure of the modified DNA is required.

Benchmarking on a central A–T nucleotide pair of the $d(CGCGAATTCGCG)_2$ Drew–Dickerson DNA dodecamer, our QM/MM DFT scheme was successfully tested, reaching convergence below 0.1 ppm once sufficient nucleotide pairs and water molecules surrounding the central nucleotide pair were taken up in the QM subsystem, and enough point charges were included in the MM subsystem. This computational methodology performs quite well in reproducing the experimental chemical shifts, with a mean error of less than 0.2 ppm, provided that the dynamics of the oligonucleotide are properly taken into account. This is achieved with (force field) molecular dynamics simulations to incorporate solvent and temperature effects in the calculation of NMR properties.

This protocol was subsequently applied to the mispair-aligned N^3T-ethyl-N^3T cross-linked $d(CGAAAT^*TTTCG)_2$ undecamer. Relying on the conformational sampling of MD simulations in a force field and mixed QM/MM description of the cross-linked oligonucleotide, dynamic proton chemical shifts were calculated and compared with their experimental counterparts. The resulting average deviations are of the order of 0.3 ppm and 0.2 ppm respectively, with the QM/MM scheme clearly providing a better prediction of the proton chemical shifts in the cross-link region. This most advanced method treats that complete region at the QM level and is best able to describe its conformational space, since it does not crucially depend on force field parameters for the cross-link.

Overall, the combined application to interstrand cross-linked DNA oligomers of our derived methodology for NMR calculations with molecular dynamic simulations, using either a force field or QM/MM description, matches the predictive power of (semi)-empirical methods. In addition, our methodology is able to yield reliable information about the NMR properties of the cross-link region. Furthermore, our methodology can be successfully applied to other (modified) DNA systems, and is likely applicable to other (bio)molecular systems.

Acknowledgements

This work is supported by the Fund for Scientific Research—Flanders (FWO) and the Research Board of the Ghent University. All computational resources (Stevin Supercomputer Infrastructure) and services were provided by Ghent University, the Hercules Foundation and the Flemish Government—department EWI. The authors kindly acknowledge Dr Jeff Klauda from the Laboratory of Molecular & Thermodynamic Modeling at the University of Maryland, for providing a translation of the Amber-parmbsc00 topologies and parameters to the Charmm format.

References

1 S. Karmakar, B. A. Anderson, R. L. Rathje, S. Andersen, T. B. Jensen, P. Nielsen and P. J. Hrdlicka, *J. Org. Chem.*, 2011, **76**, 7119–7131.
2 U. V. Schneider, I. Geci, N. Johnk, N. D. Mikkelsen, E. B. Pedersen and G. Lisby, *PLoS One*, 2011, **6**, 10.
3 C. F. Bennett and E. E. Swayze, in *Annual Review of Pharmacology and Toxicology*, Annual Reviews, Palo Alto,2010, vol. 50, pp. 259–293.
4 G. F. Deleavey and M. J. Damha, *Chem. Biol.*, 2012, **19**, 937–954.
5 S. R. Gerrard, C. Hardiman, M. Shelbourne, I. Nandhakumar, B. Norden and T. Brown, *ACS Nano*, 2012, **6**, 9221–9228.
6 G. Sun, A. Noronha and C. Wilds, *Tetrahedron*, 2012, **68**, 7787–7793.
7 A. J. Deans and S. C. West, *Nat. Rev. Cancer*, 2011, **11**, 467–480.
8 D. M. Noll, T. M. Mason and P. S. Miller, *Chem. Rev.*, 2006, **106**, 277–301.
9 P. A. Muniandy, J. Liu, A. Majumdar, S. T. Liu and M. M. Seidman, *Crit. Rev. Biochem. Mol. Biol.*, 2010, **45**, 23–49.
10 O. D. Scharer, *ChemBioChem*, 2005, **6**, 27–32.
11 H. I. Wu, J. A. Brown, M. J. Dorie, L. Lazzeroni and J. M. Brown, *Cancer Res.*, 2004, **64**, 3940–3948.
12 M. A. Catry and A. Madder, *Molecules*, 2007, **12**, 114–129.
13 S. Halila, T. Velasco, P. De Clercq and A. Madder, *Chem. Commun.*, 2005, 936–938.
14 K. Stevens and A. Madder, *Nucleic Acids Res.*, 2009, **37**, 1555–1565.
15 M. Op de Beeck and A. Madder, *J. Am. Chem. Soc.*, 2011, **133**, 796–807.
16 K. Stevens, D. D. Claeys, S. Catak, S. Figaroli, M. Hocek, J. M. Tromp, S. Schürch, V. Van Speybroeck and A. Madder, *Chem.–Eur. J.*, 2011, **17**, 6940–6953.
17 I. D. Kozekov, L. V. Nechev, M. S. Moseley, C. M. Harris, C. J. Rizzo, M. P. Stone and T. M. Harris, *J. Am. Chem. Soc.*, 2003, **125**, 50–61.
18 Y. J. Cho, H. Y. Kim, H. Huang, A. Slutsky, I. G. Minko, H. Wang, L. V. Nechev, I. D. Kozekov, A. Kozekova, P. Tamura, J. Jacob, M. Voehler, T. M. Harris, R. S. Lloyd,

C. J. Rizzo and M. P. Stone, *J. Am. Chem. Soc.*, 2005, **127**, 17686–17696.
19 H. Ding, A. Majumdar, J. R. Tolman and M. M. Greenberg, *J. Am. Chem. Soc.*, 2008, **130**, 17981–17987.
20 K. Wüthrich, Wiley, New York, 1986.
21 C. J. Wilds, E. Palus and A. M. Noronha, *Can. J. Chem.*, 2007, **85**, 249–256.
22 J. H. Zhu, Y. M. Zhao, Y. Y. Zhu, Z. Y. Wu, M. X. Lin, W. J. He, Y. Wang, G. J. Chen, L. Dong, J. F. Zhang, Y. Lu and Z. J. Guo, *Chem.–Eur. J.*, 2009, **15**, 5245–5253.
23 S. Teletchea, T. Skauge, E. Sletten and J. Kozelka, *Chem.–Eur. J.*, 2009, **15**, 12320–12337.
24 A. Kamal, N. Shankaraiah, C. R. Reddy, S. Prabhakar, N. Markandeya, H. K. Srivastava and G. N. Sastry, *Tetrahedron*, 2010, **66**, 5498–5506.
25 L. A. Truflandier, K. Sutter and J. Autschbach, *Inorg. Chem.*, 2011, **50**, 1723–1732.
26 K. Spiegel, U. Rothlisberger and P. Carloni, *J. Phys. Chem. B*, 2004, **108**, 2699–2707.
27 J. Garrec, C. Patel, U. Rothlisberger and E. Dumont, *J. Am. Chem. Soc.*, 2012, **134**, 2111–2119.
28 C. Altona, D. H. Faber and A. J. A. W. Hoekzema, *Magn. Reson. Chem.*, 2000, **38**, 95–107.
29 S. S. Wijmenga, M. Kruithof and C. W. Hilbers, *J. Biomol. NMR*, 1997, **10**, 337–350.
30 X. He, B. Wang and K. M. Merz, *J. Phys. Chem. B*, 2009, **113**, 10380–10388.
31 M. Kaupp, M. Bühl and V. G. Malkin, *Calculation of NMR and EPR parameters*, Wiley-VCH, Weinheim, 2004.
32 A. Bagno, *Chem.–Eur. J.*, 2001, **7**, 1652–1661.
33 G. Bifulco, P. Dambruoso, L. Gomez-Paloma and R. Riccio, *Chem. Rev.*, 2007, **107**, 3744–3779.
34 J. S. Mugridge, R. G. Bergman and K. N. Raymond, *J. Am. Chem. Soc.*, 2011, **133**, 11205–11212.
35 S. Taubert, H. Konschin and D. Sundholm, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2561–2569.
36 S. L. Lam and L. M. Chi, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2010, **56**, 289–310.
37 A. Bagno, F. Rastrelli and G. Saielli, *J. Org. Chem.*, 2007, **72**, 7373–7381.
38 A. Bagno, F. Rastrelli and G. Saielli, *Magn. Reson. Chem.*, 2008, **46**, 518–524.
39 S. Komin, C. Gossens, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *J. Phys. Chem. B*, 2007, **111**, 5225–5232.
40 H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 2179–2183.
41 R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature*, 1980, **287**, 755–758.
42 D. R. Hare, D. E. Wemmer, S. H. Chou, G. Drobny and B. R. Reid, *J. Mol. Biol.*, 1983, **171**, 319–336.
43 S. J. Glaser, M. L. Remerowski and G. P. Drobny, *Biochemistry*, 1989, **28**, 1483–1487.
44 S. I. Yamakage, T. V. Maltseva, F. P. Nilson, A. Foldesi and J. Chattopadhyaya, *Nucleic Acids Res.*, 1993, **21**, 5005–5011.
45 N. A. Froystein and E. Sletten, *J. Am. Chem. Soc.*, 1994, **116**, 3240–3250.
46 P. A. Fagan, H. P. Spielmann, S. T. Sigurdsson, S. M. Rink, P. B. Hopkins and D. E. Wemmer, *Nucleic Acids Res.*, 1996, **24**, 1566–1573.
47 M. W. da Silva, C. J. Wilds, A. M. Noronha, O. M. Colvin, P. S. Miller and M. P. Gamcsik, *Biochemistry*, 2004, **43**, 12549–12554.
48 S. R. Holbrook, R. E. Dickerson and S. H. Kim, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1985, **41**, 255–262.
49 J. M. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.
50 A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817–3829.
51 J. M. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.
52 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
53 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
54 K. E. Furse and S. A. Corcelli, *J. Phys. Chem. B*, 2010, **114**, 9934–9945.
55 D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross and P. A. Kollman, *AMBER 9*, University of California, San Francisco, 2006.
56 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
57 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
58 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
59 B. P. Uberuaga, M. Anghel and A. F. Voter, *J. Chem. Phys.*, 2004, **120**, 6363–6374.
60 R. H. Byrd, P. H. Lu, J. Nocedal and C. Y. Zhu, *SIAM J. Sci. Comput.*, 1995, **16**, 1190–1208.
61 T. Laino, F. Mohamed, A. Laio and M. Parrinello, *J. Chem. Theory Comput.*, 2005, **1**, 1176–1184.
62 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
63 G. Lippert, J. Hutter and M. Parrinello, *Mol. Phys.*, 1997, **92**, 477–487.
64 S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B: Condens. Matter*, 1996, **54**, 1703–1710.
65 M. Krack, *Theor. Chem. Acc.*, 2005, **114**, 145–152.
66 J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, -.
67 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 7.
68 S. Nose, *Mol. Phys.*, 1984, **52**, 255–268.
69 S. Nose, *J. Chem. Phys.*, 1984, **81**, 511–519.
70 V. Weber, M. Iannuzzi, S. Giani, J. Hutter, R. Declerck and M. Waroquier, *J. Chem. Phys.*, 2009, **131**, 11.
71 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, Jr., 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, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. 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. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *GAUSSIAN 09 (Revision A.1)*, Gaussian, Inc., Wallingford, CT, 2009.

72 R. Ditchfield, *Mol. Phys.*, 1974, **27**, 789–807.

73 C. Adamo and V. Barone, *J. Chem. Phys.*, 1998, **108**, 664–675.

74 M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.

75 P. C. Harihara and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.

76 V. A. Rassolov, J. A. Pople, M. A. Ratner and T. L. Windus, *J. Chem. Phys.*, 1998, **109**, 1223–1229.

77 C. Bassarello, P. Cimino, L. Gomez-Paloma, R. Riccio and G. Bifulco, *Tetrahedron*, 2003, **59**, 9555–9562.

78 K. K. Baldridge and J. S. Siegel, *J. Phys. Chem. A*, 1999, **103**, 4038–4042.

79 R. Jain, T. Bally and P. R. Rablen, *J. Org. Chem.*, 2009, **74**, 4017–4023.

80 P. R. Rablen, S. A. Pearlman and J. Finkbiner, *J. Phys. Chem. A*, 1999, **103**, 7357–7363.

81 S. L. Lam, *Nucleic Acids Res.*, 2007, **35**, W713–W717.

82 P. A. Hader, D. Alkema, R. A. Bell and T. Neilson, *J. Chem. Soc., Chem. Commun.*, 1982, 10–12.

83 R. A. Bell, D. Alkema, J. M. Coddington, P. A. Hader, D. W. Hughes and T. Neilson, *Nucleic Acids Res.*, 1983, **11**, 1143–1149.

84 Q. Cui and M. Karplus, *J. Phys. Chem. B*, 2000, **104**, 3721–3743.

85 A. P. Dejaegere and D. A. Case, *J. Phys. Chem. A*, 1998, **102**, 5280–5289.

86 I. A. Konstantinov and L. J. Broadbelt, *J. Phys. Chem. A*, 2011, **115**, 12364–12372.

87 C. Giessner-Prettre and B. Pullman, *Q. Rev. Biophys.*, 1987, **20**, 113–172.

88 T. Verstraelen, M. Van Houteghem, V. Van Speybroeck and M. Waroquier, *J. Chem. Inf. Model.*, 2008, **48**, 2414–2424.

