A comparative investigation of the consistent valence and extensible systematic force fields. A case study on the conformation of erythromycin A in benzene

José C. Martins,*[a] Rudolph Willem[ab] and Monique Biesemans[ab]

[a] *High Resolution NMR centre (HNMR) and* [b] *General and Organic Chemistry Department (AOSC), Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium*

Received (in Cambridge) 24th December 1998, Accepted 26th April 1999

The performances of the extensible systematic force field (esff), a general purpose force field recently developed for the molecular modeling of inorganic and organometallic compounds, and the well established consistent valence force field (cvff), known for its efficient performances, have been compared in their ability to produce accurately the conformation of organic moieties as well as for their conformational sampling properties. Erythromycin A (Ery-A), a compound with a well established conformation, was selected. After complete assignment of all ^{1}H and ^{13}C resonances of Ery-A in benzene solution, restrained MD simulations based on 44 nOe distance restraints were performed for both esff and cvff force fields, either by immediate application of the restraints on the X-ray structure, or on a set of 50 free MD conformers generated independently. The conformations generated by the two force fields were found to be highly similar, both mutually and to those found previously for Ery-A in the crystalline state and in chloroform. No distance violations larger than 0.22 Å were observed. The structures were also validated against 17 non-trivial $^{3}J(^{1}H–^{1}H)$ coupling constants and 26 $^{3}J(^{13}C–^{1}H)$ coupling constants, the latter determined from gradient-enhanced ^{1}H-{^{13}C} J-HMBC spectra. The conclusion from the present results is that both force fields generate directly comparable conformations for Ery-A.

Introduction

Erythromycin A (Ery-A) belongs to the macrolide group of antibiotics,[1] and has been used in medical treatment for nearly 5 decades.[1b] The structure consists of a polyfunctionalised 14 membered lactone ring, bearing L-cladinose and D-desosamine sugar residues (**1** in Fig. 1).[2] As for the majority of the clinically useful antibiotics,[3] the antimicrobial activity of Ery-A† originates from interference with the process of peptide biosynthesis on the bacterial ribosomes.[3,4] The mode of action has been extensively studied[4] and is known to proceed *via* interference of nascent polypeptide growth during the elongation step at the 23S RNA peptidyl transfer loop,[4a–e] and also prevention of correct 50S ribosomal subunit assembly.[4e,f] Extensive study of Ery-A and its derivatives has shown that various functional groups at specific positions in the macrolide cycle are involved at the molecular level. This has led to the development of new compounds with enhanced or complementary activity.[4e] Recently,[5] the introduction of a $(tBu_2SnO)_2O$ moiety into Ery-A gave rise to a regio- and stereospecifically pure compound which corresponds to [9*R*,12*S*]-$(tBu_2Sn)_2O$-Ery-A,[6] and displays significant anti-tumor activity under conditions where Ery-A does not.[6]

Its three-dimensional solution structure was determined using MD calculations restrained by involving nOe based distance restraints and torsional angles derived from coupling constants.[6] rMD is straightforward for many organic structures, including Ery-A, but not for organometallic compounds such as [9*R*,12*S*]-$(tBu_2Sn)_2O$-Ery-A. The problem is that classical force fields,[7] designed to model peptides,[7a] proteins,[7b,c] nucleic acids[7b,c] or organic compounds in general,[7d] such as cvff, lack the necessary parametrisation and mathematical expressions to straightforwardly process metallic centers in general, and tin atoms in particular. In fact, to the best of our knowledge, no parametrisation specific to tin has been presented so far in any of these force fields.

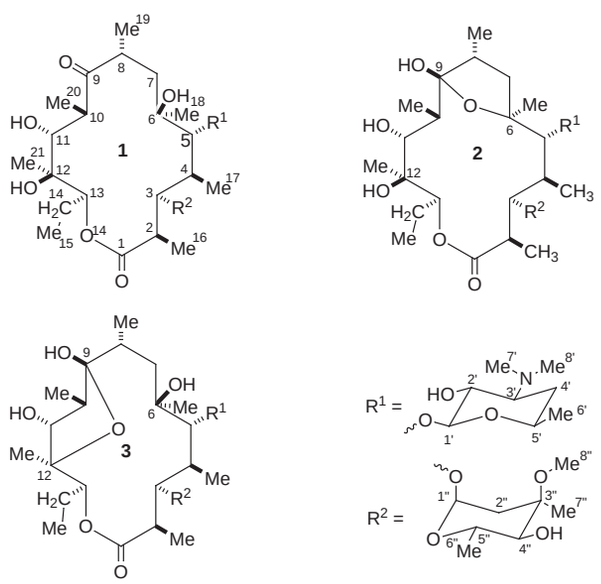

Fig. 1 Chemical structures of erythromycin A (Ery-A, **1**), the [9,6]-hemiketal of Ery-A or 9-deoxo-6-deoxy-9*S*-hydroxy-[6,9]-epoxy-erythromycin A (**2**), and the [9,12]-hemiketal of Ery-A or 9-deoxo-12-deoxy-9*S*-hydroxy-[9,12]-epoxy-erythromycin A (**3**). The desosamine and cladinose sugar residues are indicated as R^1 and R^2 respectively.

† List of abbreviations used in the text: 2D, two dimensional; DQF-COSY, double-quantum filtered correlation spectroscopy; 3D, three dimensional; cvff, consistent valence force field; Ery-A, erythromycin A; ErySn-A, [9*R*,12*S*]-$(tBu_2Sn)_2O$-erythromycin A; esff, extensible systematic force field; FID, free induction decay; ge, gradient enhanced; HMBC, heteronuclear multiple bond correlation; HMQC, heteronuclear multiple quantum correlation; HOHAHA, homonuclear Hartmann Hahn; NMR, nuclear magnetic resonance; nOe, nuclear Overhauser enhancement; NOESY, nuclear Overhauser effect spectroscopy, rMD, restrained molecular dynamics; rmsd, root mean square deviation.

Table 1 ^{1}H and ^{13}C resonance assignments of erythromycin A in benzene and comparison with those observed in chloroform[a]

	C_6D_6		$CDCl_3$	
Position	$\delta(^{1}H)$/ppm	$\delta(^{13}C)$/ppm	$\delta(^{1}H)$/ppm	$\delta(^{13}C)$/ppm
1	—	175.8	—	176.3
2	3.00	45.4	2.87	45.0
3	4.25	80.7	3.99	80.3
4	2.22	40.2	1.97	39.5
5	3.78	84.1	3.56	84.0
6	—	74.9	—	74.8
6-OH[b]	1.92	—	1.51[c]	—
7eq (pro-*S*)	2.03	39.2	1.93	38.5
7ax (pro-*R*)	1.83	39.2	1.74	38.5
8	2.70	4.52	2.68	44.9
9	—	220.8	—	221.9
10	3.08	38.4	3.08	38.1
11	4.15	69.5	3.82	68.8
11-OH[b]	4.45	—	3.95	—
12	—	74.9	—	74.5
12-OH[b]	3.38	—	3.13	—
13	5.52	77.4	5.03	77.1
14a (pro-*S*)	2.13	21.9	1.91	21.2
14b (pro-*R*)	1.59	21.9	1.48	21.2
Me15	0.98	11.1	0.84	10.7
Me16	1.25	16.2	1.18	15.9
Me17	1.50	9.6	1.10	9.2
Me18	1.55	26.9	1.46	26.4
Me19	0.83	18.1	1.16	18.4
Me20	1.28	12.3	1.14	12.0
Me21	1.24	16.7	1.12	16.2
1′	4.65	103.8	4.40	103.3
2′	3.37	71.2	3.21	71.1
2′-OH[b]	3.35	—	3.45[c]	—
3′	2.32	66.3	2.43	65.3
4′eq	0.91	29.0	1.22	29.2
4′ax	1.22	29.0	1.67	29.2
5′	3.48	69.0	3.48	68.6
Me6′	1.16	21.6	1.22	21.4
Me7′ + Me8′	1.88	40.1	2.29	40.3
1″	4.98	96.8	4.88	96.5
2″eq	2.06	35.0	2.35	35.0
2″ax	1.16	35.0	1.56	35.0
3″	—	72.9	—	72.7
4″	3.00	78.3	3.00	77.9
4″-OH	2.22	—	2.23	—
5″	4.22	66.1	3.99	65.7
Me6″	1.59	19.3	1.27	18.5
Me7″	1.00	21.5	1.23	21.4
Me8″	3.14	49.4	3.31	49.5

[a] Taken from reference 9. [b] Tentative assignments only, see text. [c] Taken from reference 10.

Rather than trying to parametrise tin into existing force fields, we used the extensible systematic force field, esff,[8] to derive the conformation of [9*R*,12*S*]-$(tBu_2Sn)_2O$-Ery-A. This recently released force field[8] is specifically conceived to model inorganic and organometallic compounds. Both its mathematical formulation, and the use of a modest set of atomic parameters from which the complete molecular force field is generated through a series of empirical rules,[8] provide the basis to adequately deal with the broad compositional and geometrical diversity of organometallic compounds.[8] Because of its recent development however, a comparison of the performances of this force field with those of other ones is still missing, only a few applications of the esff force field having appeared in the literature.[8a] The presence of a rather large organic moiety in [9*R*,12*S*]-$(tBu_2Sn)_2O$-Ery-A raised the question of whether the esff force field can appropriately handle this part of the structure. In this paper, the results of rMD modeling using both the well-established cvff and the esff force fields are presented and critically compared using Ery-A as a model compound. The choice of Ery-A in benzene was dictated by both the needs of performing the comparison on a compound with a well defined solution conformation, and of having the necessary reference conformation in the same solvent to allow investigation of the consequences of the organotin derivatisation on the conformation of Ery-A.[6]

Results and discussion

Resonance assignment and collection of structural data

The ^{1}H and ^{13}C NMR spectra display one set of major resonances, together with minor resonances. Using a combination of homo- and heteronuclear 2D techniques, a complete assignment of the ^{1}H and ^{13}C resonances, apart from those from exchangeable protons, associated with the major form was achieved (Table 1).[9,10] Protons of the same spin system were identified from HOHAHA[11] spectra and the coupling topology of their ^{1}H resonances, as deduced from the direct coupling connectivities in the DQF-COSY[12] spectrum. The remaining ^{1}H assignment ambiguities, including the four methyl singlets, were subsequently relieved using chemical shift and $^{3}J(^{1}H–^{13}C)$ scalar coupling multiplicity information from the 1D and DEPT[13] ^{13}C spectra. Combined with gradient enhanced 2D ^{1}H-{^{13}C} HMQC[14,15] and HMBC[14,16] spectroscopy, this completed simultaneously the ^{13}C NMR resonance assignment.

Labile hydroxy proton resonances were readily identified in the 2D NOESY spectrum from characteristic exchange cross-peaks.[17] However, no clear cut assignment was possible for all of them as none exhibited $^{2}J(^{1}H–^{13}C)$ correlation cross-peaks in the 2D ^{1}H-{^{13}C} HMBC spectrum, as a result of too fast proton exchange. The 4″-OH occurs at 2.22 ppm as unambiguously identified by a 2D DQF cross-peak with the H4″ resonance. Tentative assignment of the resonances at 4.45 ppm and 3.38 ppm to the 11-OH and 12-OH protons respectively, is based on a (only very weak) nOe cross-peak with the H11 and Me21 resonances respectively. The remaining hydroxy resonances at 3.35 and 1.92 ppm, are tentatively assigned to the 6-OH and 2′-OH protons respectively, only by comparison with the chemical shift as reported[9] for data from chloroform solutions (Table 1).

The presence of two carbonyl ^{13}C resonances assigned to C1 (175.8 ppm) and C9 (220.8 ppm) respectively (Table 1), shows that Ery-A occurs mainly (~95%) as the 9-keto isomer in benzene solution. As could be reasonably expected from previously reported NMR studies in various solvents,[18–20] the ^{1}H and ^{13}C spectra of Ery-A in benzene solution also reveal the presence of minor resonances. While the overall assignment of these minor forms was not achieved, the lack of minor ketone ^{13}C carbonyl resonances, combined with the observation of two minor ^{13}C resonances at 111.9 and 109.9 ppm indicate the presence of hemiketalic carbons in these minor isomers of Ery-A,[19,20] with hemiketal formation between the C9 carbonyl and either the C6 (**2**) or C12 (**3**) hydroxy groups.[18–20] The fact that, just like in chloroform, they total only 5% of the observed signal in benzene, matches perfectly well the observed dependence of the hemiketal population on solvent polarity, increasing from 5% in chloroform and benzene to 30% in water.[19]

A total of 44 non-trivial nOes were unambiguously identified from the 400 ms NOESY spectrum. These can be classified in 22 intra-lactone nOes, of which 10 are non-vicinal, 4 nOes within the cladinose residue and 8 nOes within the desosamine one, and 3 nOes indicative of close contacts between both sugar residues. The remaining nOes occur between lactone ring protons and either the cladinose (4 nOes) or the desosamine (3 nOes) sugar residue. A total of 17 $^{3}J(^{1}H–^{1}H)$ coupling constants relevant to conformation were likewise determined (Table 2).[22–24] The pairwise diastereotopic protons of the CH_2 groups at C7 and C14 were stereospecifically assigned using the $^{3}J(^{1}H–^{1}H)$ coupling constants and the nOes to the vicinal H8

Table 2 Comparison of relevant experimental $^3J(^1H–^1H)$ in C_6D_6, $CDCl_3$ and calculated from the rMD derived structure of Ery-A in C_6D_6

	$CDCl_3$ [a] ^3J/Hz	C_6D_6 ^3J/Hz	rMD [b] ^3J/Hz
J_{23}	9.5	9.2	8.7
J_{34}	1.5	1.5	1.6
J_{45}	7.5	7.7	7.2
J_{7ax8}	11.7	10.7	10.1
J_{7eq8}	2.4	1.8	1.6
J_{1011}	1.3	1.4	2.0
J_{1314a}	11.0	10.9	10.2
J_{1314b}	2.4	2.3	2.1
$J_{1''2''ax}$	4.8 [c]	4.8	4.3
$J_{1''2''eq}$	1.1 [c]	0.9	2.1
$J_{4''5''}$	9.7 [c]	9.2	10.1
$J_{1'2'}$	7.5 [c]	7.2	10.1
$J_{2'3'}$	10.3 [c]	10.2	10.2
$J_{3'4'ax}$	11.5 [c]	12.2	10.2
$J_{3'4'eq}$	4.5 [c]	3.6	2.7
$J_{4'ax5'}$	11.5 [c]	10.7	10.2
$J_{4'eq5'}$	2.2 [c]	1.9	2.9

[a] Taken from reference 22. [b] Calculated from the lowest energy conformer of the EF family using the Karplus equation with coefficients $A = 7.76$, $B = -1.10$ and $C = 1.40$.[24] [c] Taken from reference 23.

Table 3 Contributions of the various force field terms of the total energy (in kJ mol^{-1}) of the low energy conformers obtained from both cvff and esff using either rMD protocol I or II [a]

Low energy conf.	Total energy	Non-bond energy [b]	Internal energy [c]	Forcing energy
CX (50)	396.9 ± 11.0	196.5 ± 5.2	244.1 ± 3.0	20.1 ± 0.5
CF (49)	389.5 ± 17.9	184.9 ± 8.2	248.4 ± 8.0	21.0 ± 0.7
EX (50)	280.8 ± 23.7	119.1 ± 11.6	217.8 ± 22.3	21.8 ± 0.8
EF (36)	296.4 ± 25.6	128.4 ± 11.6	210.6 ± 24.6	24.6 ± 0.6

[a] The conformers originating from the 4 different rMD runs are labelled by a two-letter code, in which the first stands for the force field used (C for cvff, E for esff), and the second refers to the rMD protocol used (X for protocol I which applies the restraints to the X-ray structure and F for protocol II which contains a free MD run to generate starting conformers). [b] Comprising van der Waals and Coulomb terms. [c] Energy associated with all terms describing the covalent connectivities.

and H13 protons respectively.[25] Analysis of the pattern of large and small $^3J(^1H–^1H)$ coupling constants and close interproton contacts within the pyranose rings deduced from the corresponding nOe cross-peaks, show that both sugar residues adopt the same chair conformation as observed in the X-ray structure of Ery-A,[2,22] including the α- and β-glycosidic linkage for the cladinose and desosamine residues repectively. In the process, stereospecific assignments were obtained for the CH_2 protons in both sugar residues.

Conformational calculations and solution structure

Distance restraints derived from the nOes were used in restrained MD simulations within the cvff and esff force fields, in order to generate conformers representative of the solution structure of Ery-A in benzene. To this end, the X-ray structure of 6-*O*-methyl-erythromycin A [26] in which the 6-methyl group was simply substituted for a hydrogen atom, was chosen as a suitable starting structure. In a first restrained MD simulation, the distance restraints were gradually imposed on the starting structure by increasing the weight of the distance restraint potential term from 0 to 1 in 4 steps of 1 ps rMD at 1000 K. During the following 250 ps rMD trajectory at this temperature, 50 restrained conformers were sampled and subjected to stepwise cooling followed by restrained energy optimization. This protocol I yielded a family of low energy conformers in both the cvff and esff force fields, subsequently referred to as CX and EX respectively (see Table 3). In protocol II, a possible bias of the starting structure on the end result (constraints in protocol I reaching full effect after only 4 ps) was investigated using the starting structure as the origin of a 250 ps free MD trajectory at 1000 K, from which a set of 50 conformers were sampled. These were then used as the starting point for restrained MD simulations using a protocol otherwise similar to protocol I. When applied to the cvff and esff force fields, these yielded the CF and EF families of conformers respectively (Fig. 2, a and b). Simulation statistics and restraint violation analysis for each group of conformers are collected in Tables 3 and 4. All simulations had a high yield (>98%) in low energy conformers, except for the EF run, where 13 of the resulting conformations had energies exceeding those of the lowest one by a factor of at least 2, and were discarded. The low energy conformers collected within each group have converged to very similar if not identical conformations, as judged from the mutual rmsds of the lowest energy conformer from each family (Table 5). Families generated from the same force field are generally coincident (rmsd virtually zero and identical dihedral angles). When comparing families from different force fields, the rmsds are non-zero but small, nevertheless resulting in quasi-identical conformations (Fig. 2c). Dihedral angles differ slightly between both sets but are all within 5°; differences in bond lengths are all smaller than 1 to 3 pm. These small differences most probably result from the different formulation

Table 4 Statistics of distance restraint violations for the low energy conformers obtained from both cvff and esff using either rMD protocol I or II

	$N > 0.1$ Å [a]	Sum [b]/Å	Max [c]/Å
CX	2.62 ± 0.53	0.73 ± 0.07	0.21 ± 0.00
CF	3.32 ± 0.62	0.51 ± 0.16	0.21 ± 0.00
EX	2.00 ± 0.00	0.56 ± 0.06	0.20 ± 0.01
EF	2.31 ± 0.44	0.56 ± 0.06	0.20 ± 0.01

[a] Average number of violations larger than 0.1 Å. [b] Average sum of all violations. [c] Average maximum violations occurring within the low energy conformers.

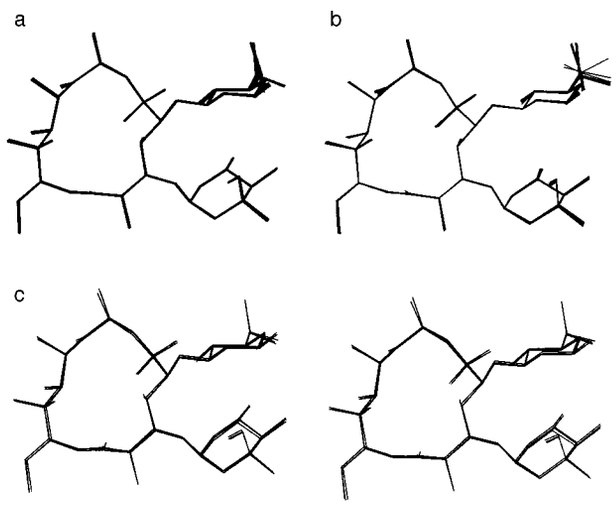

Fig. 2 Comparison of all low energy conformations of Ery-A obtained from rMD simulations. In (a), the 49 low energy conformers representing the CF family obtained with cvff, and in (b) the 36 low energy conformers representing the EF family, obtained with esff, are shown superimposed. For clarity, the hydrogen atoms are not shown. All non-hydrogen atoms were considered for superposition. The stereoview in (c) shows the superposition of the four lowest energy conformers resulting from each of the four families (CX, CF, EX and EF) and allows us to appreciate the close correspondence in the conformations obtained from both force fields using either protocol I or II.

Table 5 Mutual rmsd (in Å) of the lowest energy conformers obtained from both cvff and esff using either rMD protocol I or II[a]

	CX	CF	EX	EF
CX	—	0.000	0.125	0.156
CF	0.000	—	0.125	0.156
EX	0.083	0.083	—	0.052
EF	0.086	0.086	0.005	—

[a] Rmsd values representing all heavy atoms (upper triangle) or the ring atoms comprising the lactone ring (lower triangle). 0.000 indicates values smaller than 0.5×10^{-3} Å.

Table 6 Comparison of relevant experimental $^3J(^{13}C,^1H)$ coupling constants in C_6D_6 with those in $CDCl_3$ and calculated from the rMD derived structure of Ery-A in C_6D_6

	$CDCl_3$[a]	C_6D_6	rMD[b]
lactone			
J_{C4H2}	—	3.8	4.3
J_{C1H3}	—	3.2	3.7
J_{C5H3}	3.3	3.5	3.3
J_{C3H5}	5.8	5.5	4.5
J_{C7H3}	—	4.5	6.5
J_{C18H3}	3.0	4.4	3.1
J_{C5H7}	—	4.3	2.6
J_{C18H7}	—	6.4	6.7
J_{C9H11}	—	3.7	2.8
J_{C13H11}	—	2.8	2.7
J_{C20H11}	—	5.8	6.7
J_{C1H13}	4.0	4.1	5.5
J_{C11H13}	—	3.3	2.8
J_{C15H13}	3.2	3.8	2.1
J_{C21H13}	—	4.3	6.7
glycosidic			
$J_{C3H1''}$	3.6	4.6	3.5
$J_{C1''H3}$	4.0		
$J_{C5H1'}$	2.6	3.9	3.3
$J_{C1'H5}$	5.6	6.5	5.4
cladinose			
$J_{C3''H1''}$	5.2	5.7	6.5
$J_{C3''H1''}$	6.5	6.7	6.6
$J_{C3''H5''}$	—	<2.5	1.0
$J_{C4'H2'ax}$	—	6.8	6.7
desosamine			
$J_{C3'H1'}$	[c]	<2.5	1.1
$J_{C5'H1'}$	[c]	<2.5	1.7
$J_{C3'H5'}$	—	~2.5	1.2

[a] Taken from reference 29. [b] Calculated from the lowest energy conformer using the Karplus equation with coefficients $A = 5.7$, $B = -0.6$ and $C = 0.5$.[27] [c] Close to zero according to reference 28.

of the mathematical model and the parameters used in both force fields, described below. As they are small, and in view of comparisons with the conformation of [9*R*,12*S*]-$(tBu_2Sn)_2O$-Ery-A, which were obtained with rMD according to protocol II in the esff force field,[6] further analysis of the solution conformation of Ery-A in benzene is based on the EF family of conformers only.

Since the experimental $^3J(^1H,^1H)$ and $^3J(^{13}C,^1H)$ coupling constant data were not imposed during the rMD simulations, they afforded an independent validation of the conformation of Ery-A obtained from rMD. $^3J(^1H,^1H)$ coupling constants calculated from the lowest energy conformer of the EF family compare well with the experimental ones (Table 2). Comparison of the calculated and experimental $^3J(^{13}C,^1H)$ couplings from the 2D ge-J-HMBC experiment, was more difficult, as only a few Karplus-type equations exist for such couplings in compounds other than peptides or proteins. The parameters used here are those for a C–O–C–H segment representing glycosidic linkages, and were obtained from a rather limited set of compounds.[24,27] This specific Karplus equation can therefore not be considered as the optimal one for macrolide antibiotics such as Ery-A. Nevertheless, it has provided reasonable angular estimations in conformational studies and the errors are believed to be in the range of ±20°.[28] Thus, while we felt it inappropriate to enforce these as dihedral angle restraints, they still have qualitative value in the validation of the conformations produced from the nOe data alone (Table 6).[27–29] In general there is a good qualitative correspondence between the experimentally observed $^3J(^{13}C,^1H)$ coupling constants and those calculated from the lowest energy conformer with the Karplus equation, as the large and small coupling constant values are very well reproduced. On a more quantitative basis, many calculated couplings are within 1 Hz of the measured ones. The exceptions are found in the Karplus function dihedral angle intervals where $dJ/d\theta$ is large or where the application of this equation can be questioned because the nature of the atoms involved differ significantly from those used in the glycosidic parametrisation set.

Table 7 Dihedral angle comparison between the X-ray structure of Ery-A and the NMR structure of Ery-A in benzene solution

No.	θ_{ij}	X-ray[a]/°	NMR/°
1	O14–C1–C2–C3	115.9	120.4
2	C1–C2–C3–C4	−61.2	−82.1
3	C2–C3–C4–C5	164.8	159.1
4	C3–C4–C5–C6	−116.1	−91.2
5	C4–C5–C6–C7	−68.5	−75.5
6	C5–C6–C7–C8	175.0	174.9
7	C6–C7–C8–C9	−77.0	−72.1
8	C7–C8–C9–C10	−60.8	−68.5
9	C8–C9–C10–C11	122.0	119.9
10	C9–C10–C11–C12	−173.3	−166.0
11	C10–C11–C12–C13	167.8	167.2
12	C11–C12–C13–C14	−74.9	−73.2
13	C12–C13–O14–C1	107.3	110.1
14	C13–O14–C1–C2	171.3	176.4
15	H3–C3–O3–C1″	28	23.2
16	C3–O3–C1″–H1″	43	38.3
17	H5–C5–O5–C1′	14	12.9
18	C5–O5–C1′–H1′	46	40.6

[a] Lactone angles taken from reference 22, glycosidic angles from reference 10.

Although a wide variety of solvents have been used for conformational studies of Ery-A by NMR,[9,18–21,29] benzene, in the present investigation, is only the second solvent for which a full study, including rMD, has been performed. The conformation of Ery-A in benzene is very similar to that described previously in chloroform[9,18,29] as are to a large extent the NMR data[9,18–20,29] for the main isomer (Tables 2 and 6). Differences are larger between the ^1H chemical shifts than the ^{13}C ones for Ery-A in benzene and chloroform (Table 1). This indicates that the changes are due to contributions to the ^1H shift from the solvent, rather than to important conformational changes. A comparison of the θ_{ij} dihedral angles obtained here in the EF conformer family (Table 7), with those from the literature confirms the high similarity in the lactone ring conformation in benzene, chloroform[22] and the crystal structure.[22] Only θ_{23} and θ_{45} differ noticeably, by 21° and 25° respectively, while all other differences are smaller than 10°, in most cases even smaller than 5°.

Comparative assessment of the esff and cvff force fields

The performances of the well-established cvff and the novel esff force fields, when used for molecular simulations of organic molecules using rMD, are now critically compared, using Ery-A as a model compound. As stated before, irrespective of the force field or protocol used, all simulations CX and EX (protocol I) and CF and EF (protocol II) produced low energy conformers (Tables 3), which satisfy all the experimental NMR data (Table 4) and are very similar to one another (Table 5,

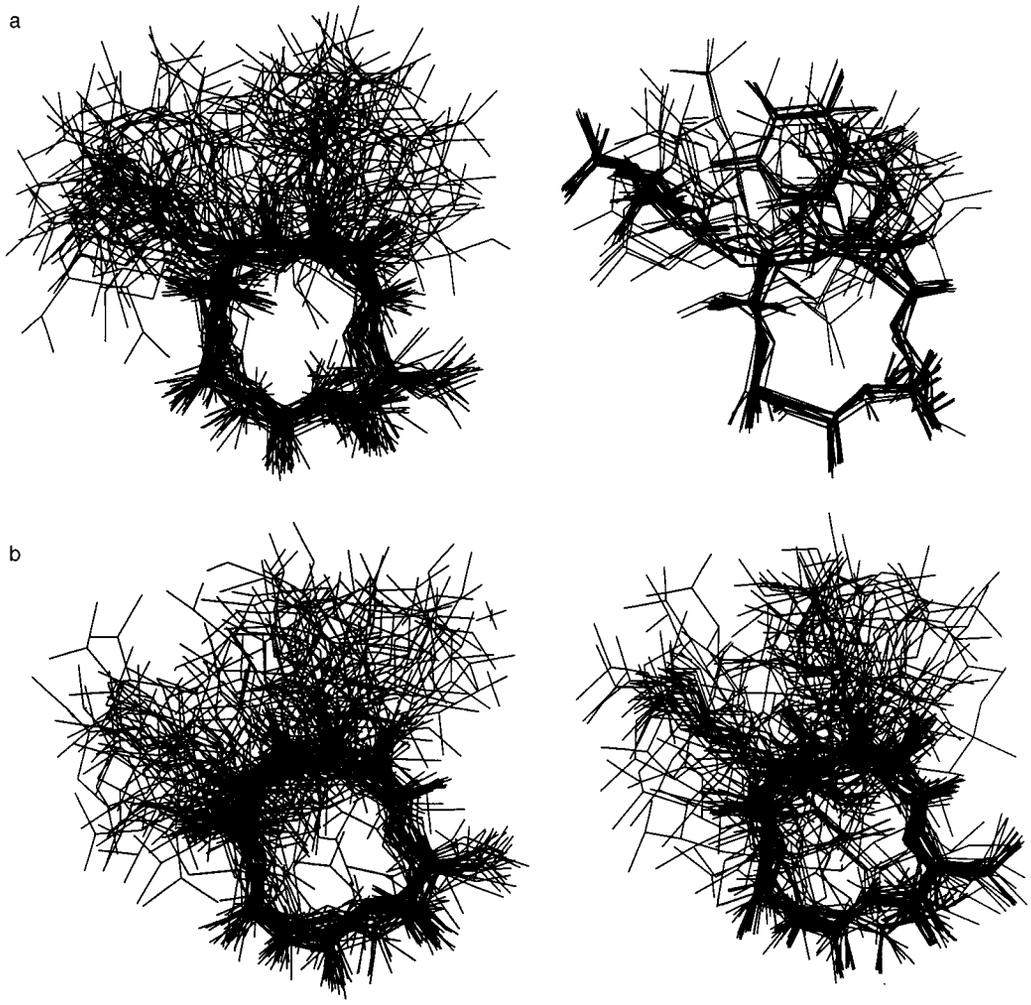

Fig. 3 Comparison of the conformational space sampled during a 250 ps free MD simulation as described by protocol II, using either the cvff (a) or the esff (b) force field. In each representation the atoms of the lactone cycle were used to superimpose conformers 2 to 49 on conformer 1. The superpositions on the left hand side represent those prior to the cooling of the temperature using an annealing protocol, while those on the right hand side were obtained after the final unrestrained energy optimization.

Fig. 2c). The rmsd difference over all non-hydrogen atoms averaged at only 0.156 Å when comparing conformers obtained with cvff and esff, and is virtually zero when comparing the conformers which result when both protocols I and II are applied with the same force field (Table 5). Also, all rMD runs yield similar, if not identical, restraint violation statistics (Table 4). The violated distance restraints referred to were the same in both force fields. The same holds for their average number, total sum and maximum, indicating the equivalent weight of these constraints when applied in the cvff and esff force fields. The total energy and the contribution of the various force field terms are summarised in Table 3. Conformations issued from cvff are about 100 to 125 kJ mol^{-1} higher in energy as compared to those from esff. Ruling out direct energy comparisons between different force fields because of their different mathematical formulation and parametrisation, the near coincidence in structure and violation statistics suggests, however, that the global energy minimum defined by the force field and the experimental restraints occurs for the same configuration of atoms, justifying qualitative comparisons. First, it can be noted that the total internal energies which result from the esff force field are smaller by about 10 to 15% (Table 3). The major contribution to the difference in total energy is provided by the non-bonding energy term, and more specifically the van der Waals term, which is substantially lower when using the esff force field. While at first glance this is rather surprising, it should be noted that cvff uses a r^{-12} dependence while esff uses a r^{-9} dependence in the formulation of the Lennard-Jones potential, and therefore these energy terms are not expected to be identical. Finally, while many distinct conformations, generated from the free MD run in protocol II were used as starting points for rMD calculations, these all converged to the same final conformation (Fig. 2a, b), in all aspects identical to that obtained from protocol I, as evidenced by the mutual rmsd comparison within each force field (Table 5, Fig. 2c), therefore ruling out bias.

Sampling properties of the cvff and esff force fields

The sets of 50 free MD conformers from cvff and esff (CF and EF respectively) were generated to assess the possible bias of the starting structure on the end result using protocol I. A detailed analysis of the various conformations of Ery-A generated during free MD runs with the cvff force field has been presented previously.[10,21] We focus here on a comparison of the unrestrained structures of Ery-A generated by both force fields.

A visual comparison of the 50 free MD conformers issued from cvff and esff, either before annealing (Fig. 3a) or after final energy minimisation (Fig. 3b), had already indicated that the esff force field samples a larger conformational space. More quantitative support was obtained by analyzing the differences in dihedral angle fluctuations in the lactone ring and glycosidic bonds using order parameters. For each dihedral angle, these were calculated[30] over all 50 free MD conformers sampled at 1000 K using either the cvff or the esff force field (Fig. 4). Of the 18 dihedrals monitored in this way, the esff conformers

Table 8 Statistics of the energies of the set of 50 free MD conformers issued from cvff and esff, prior to annealing and after energy minimisation

	Energy/kcal mol^{-1}	Variance[a] (%)	Range[b]/kcal mol^{-1}	No. > σ (2σ)[c]
cvff (upper left)[d]	418.1 ± 21.3	5.1	372.5–469.8	20 (3)
esff (lower left)[d]	425.7 ± 26.7	6.3	371.2–524.6	17 (3)
cvff (upper right)[d]	91.4 ± 3.6	4.0	86.8–101.8	11 (3)
esff (lower right)[d]	61.6 ± 4.9	8.0	52.6–70.5	17 (0)

[a] Variance coefficient. [b] Minimum and maximum energy observed within the group of conformers. [c] Number of structures with energy deviating from the mean value by one (two) standard deviation(s) σ (2σ). [d] The position of the set of conformers as shown in Fig. 3 is indicated in parentheses. The first two and last two rows are for the sets of conformers prior to annealing and after energy minimisation respectively.

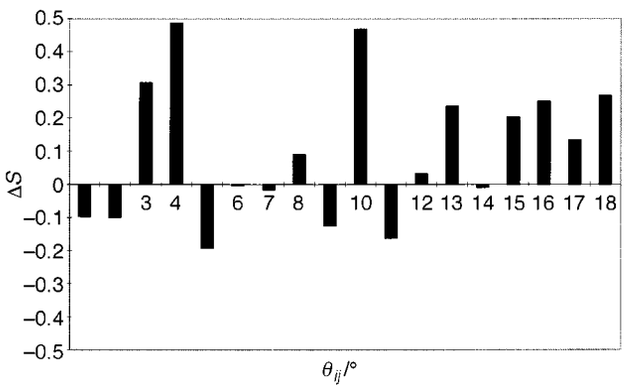

Fig. 4 The difference in sampling efficiency of the cvff and esff force fields as inferred from the difference in order parameter $\Delta S = S_{cvff} - S_{esff}$ observed for selected dihedral angles in Ery-A. The numerical labeling of the dihedral angles θ_{ij} along the x-axis reflects the ranking in Table 7. For each force field, the order parameter S for each dihedral angles was obtained[30] from the 50 free MD conformers sampled at 1000 K using protocol II. Positive and negative ΔS values correspond to the presence of more or less flexibility in the associated dihedral angle in the esff force field as compared to the cvff force field, and are indicative of better or worse conformational sampling, respectively.

displayed three with higher order, seven with similar order parameters ($\Delta S \leq 0.1$) and nine with substantially lower order parameters, with ΔS ranging from 0.20 up to 0.48. A more detailed analysis of the order parameters shows that the C3–C5 region of the lactone ring displays more flexibility when the esff force field is used (angles 3 and 4 in Fig. 4). This may be coupled to the lower order for the glycosidic bonds located in this part of the structure, which display order parameters, smaller by 0.2 to 0.3, as compared to cvff (angles 15–18 in Fig. 4). Linear regression analysis yields only a poor linear relationship between S_{esff} and S_{cvff} with a slope of 0.51 and a correlation coefficient r^2 of only 0.46. By contrast, equivalent calculations for the set of 50 restrained MD conformers collected before annealing in protocol I yield similar order parameters for both force fields with values generally over 0.95, consistent with the limitation in the sampling of conformational space by the immediate application of restraints. In this case, a linear relationship, with slope 1.04 and a correlation coefficient r^2 of 0.98, exists between the S_{esff} and S_{cvff} order parameters for the same dihedral angles.

Importantly, the increased coverage of conformational space observed when using esff rather than cvff during the 250 ps free MD simulation, is accompanied by a slight increase in the energy distribution within the group of sampled conformers (Table 8). This becomes more evident when the standard deviations from the mean energies are expressed as variance coefficients (Table 8). While slightly larger before annealing (6.3% *vs.* 5.1%), the variance coefficient of the energy for the esff conformers after energy optimisation, while remaining small, is twice that observed for cvff (8.0% *vs.* 4.0%). This is also evident from the number of outliers, *i.e.* the number of conformers with energy beyond one standard deviation (Table 8). Their number is significantly reduced after energy optimisation within cvff, but not so for esff, nicely parallelling the visual impression afforded by Fig. 3. This would indicate that, within the duration of the MD simulation, the area of the global energy minimum for Ery-A in the esff force field is characterised by a larger amount of local, higher energy minima than for cvff.

It is as yet unclear whether the increased conformational sampling of esff reported here is linked to the nature of the compound in this specific modeling study, or may turn out to be a persistent property of the esff force field. The latter might result from the quite distinct mathematical formulation of esff as compared to cvff. This should become evident as more such comparative studies are reported. If confirmed, the presence of more distinct conformations corresponding to local energy minima in the potential energy surface, especially for constrained systems such as macrocycles, may turn out to be a potential concern in the conformational exploration of the low energy minima of such molecules. One might indeed argue that in reality, Ery-A is more constrained than the esff results suggest. Nevertheless, these considerations should not be used to judge the relative merit of the esff force field with respect to cvff, as the cause of the broader sampling characteristics, and the wider energy distribution linked to it, is currently not understood, and needs more investigation. It should be pointed out however, that a conformationally and energetically more heterogeneous set of free MD conformers as obtained here from esff can be turned into an advantage, when these are subsequently to be used for restrained MD, as the extended sampling of conformational space allows possible bias to be reduced.

Conclusion

This paper has investigated the performances of the esff force field, designed to handle the broad compositional and geometrical diversity of organometallic and inorganic compounds, in modeling purely organic moieties, *i.e.* including only elements of the 2nd period. Comparison of the rMD results obtained from both the novel esff and the well-established cvff force fields, indicates that the resulting conformations are similar showing that use of the esff force field is appropriate in the restrained simulation of organic moieties as well. Therefore it can be expected that, when applied to modeling of compounds including both organic and metallorganic moieties, esff should be able to adequately handle the organic moiety. The results reported here for Ery-A also indicate that when used for free MD, at least for the modeling protocols used here, esff results in a larger coverage of the conformational space, linked to a somewhat higher energy distribution. The origin of this observation is yet to be determined and therefore, the present study does not provide a sufficient base to extrapolate these properties towards a broad panel of organic structures. The observation that esff may provide a broader screening of the conformational space is an important consideration needing further attention. At the same time, it may also

decrease bias risks linked to the choice of the starting model in restrained MD modeling using experimental NMR data.

Experimental

NMR experiments

Erythromycin displays limited solubility in benzene. For most NMR experiments, 8.5 mg Ery-A (Aldrich, erythromycin hydrate) was dissolved into 0.5 mL of C_6D_6 (Aldrich). For some ^{13}C experiments, a saturated solution was prepared by dissolving 31.5 mg in 0.5 mL of C_6D_6, from which the undissolved fraction was removed by centrifugation. All ^1H and ^{13}C NMR experiments, at respectively 500.13 and 125.77 MHz, were performed at 303 K on a Bruker AMX500 instrument equipped with a X32 computer, a BSMS digital lock and a BGU II gradient unit. Chemical shifts were referenced to the residual solvent peak and converted to the TMS scale by adding 7.15 ppm and 128.0 ppm for ^1H and ^{13}C respectively. The 1D ^1H spectra covered a spectral width of 10 ppm. 32 FIDs with 16K data points each were acquired, with a total recycling delay of 3.37 s. For the 1D ^{13}C as for the DEPT-90 and DEPT-135 spectra,[13] 15000 or 2048 FIDs with 32K data points each were acquired, with a total recycling delay of 2.98 s, under continuous ^1H-broadband decoupling. All homonuclear 2D spectra were recorded using a spectral width of 6.02 ppm and in the phase sensitive mode, using TPPI.[31] Typically 256 to 512 t_1 increments, 24 to 64 FIDs of 2K data points each, were recorded. The total recycling delay was 2 s. Processing consisted of multiplication with a $\pi/4$ shifted squared sinebell in F2 and a $\pi/3$ shifted sinebell in F1, followed by zero-filling to a 4K by 2K data matrix, prior to Fourier transformation. HOHAHA[11] spectra were obtained with a MLEV-17 spin lock using a 10 kHz B_1 field. Spin lock times were 32, 50 and 70 ms. The mixing times in the NOESY[17] spectra were 500 and 800 ms. The data set was completed by a DQF-COSY[12] spectrum. ge-^1H-{^{13}C} HMQC[13,14] and ge-^1H-{^{13}C} HMBC[14,16] 2D spectra were recorded on the saturated sample with 512 t_1 increments and 16 to 64 scans of 2K data points each. Delays for heteronuclear coupling evolution were 3.85 ms for the HMQC and 60 or 120 ms for the HMBC spectra. The ^{13}C spectral width was set to 120 ppm, with appropriate folding of the carbonyl resonances. No ^1H decoupling was applied during acquisition. The total recycling delay was 2 s. Gradients were applied with 50 : 30 : 40 amplitude ratio (100 = 0.5 T m^{-1}). Processing consisted of multiplication with $\pi/4$ shifted squared sinebell in F2 and $\pi/3$ shifted sinebell in F1, followed by zero-filling to a 4K by 2K data matrix, prior to Fourier transformation. As the gradient selection causes phase-modulation in t_1, spectra were represented in the magnitude mode. The 3D ge-^1H-{^{13}C}-J-HMBC spectrum[32] was recorded as a succession of 16 2D ge-^1H-{^{13}C} HMBC spectra, in which the effective evolution time τ for the $^3J(^{13}C–^1H)$ couplings was linearly varied from 0 s to 0.5 s ($=\tau_{max}$) through the use of a moving 180° $^nJ(^{13}C–^1H)$ refocusing pulse on the ^{13}C nuclei, within a fixed time period of 0.5 s. For each individual spectrum, a total of 100 FIDs consisting of 4K data points, 128 scans each were recorded. Processing consisted of zero-filling along the ^{13}C dimension, followed by squared cosine-bell multiplication and Fourier transformation along the ^1H and ^{13}C frequency dimensions, followed by magnitude calculation. All spectra were subjected to third and second order polynomial base line correction.

Structural data collection

Interproton distances were determined from the build-up of the corresponding nOe cross-peak intensities measured from a series of 8 NOESY spectra with mixing times ranging from 200 ms to 1200 ms, and recorded as described above. Integration areas were defined at 500 ms with the UXNMR software and subsequently applied to all NOESY spectra. To extend the linearity of the nOe build-up, the cross-peak intensities are normalised against a diagonal peak.[33] As considerable overlap is present however, only a limited number of diagonal peaks are well resolved. Therefore, individual resonances were grouped according to their estimated T_1 relaxation time constants and normalised against a resolved diagonal peak within each group. Linear regression of the build-up curves yielded correlation coefficients well above 0.95, with most superior to 0.98, indicating linearity of the build-up. Finally the interproton distances corresponding to each nOe were obtained through calibration of the associated slope against the slope of the 7-CH_2 geminal nOe, corresponding to a distance of 1.79 Å. These were translated into upper limit distance restraints through multiplication by 1.1. Lower limits were set to the sum of the hydrogen van der Waals radius. Distance restraints involving methyl groups were increased by 0.5 Å so as to refer to the corresponding pseudo-atom.[34] Stereospecific assignments for each of the four CH_2 moieties were derived from the combined analysis of $^3J(^1H–^1H)$ couplings and geminal nOes within the associated -CH–CH_2- fragments. $^3J(^1H–^1H)$ coupling constants were obtained from 1D ^1H spectra, processed with mild Gaussian apodisation. $^3J(^{13}C–^1H)$ couplings were extracted from the 3D ge-^1H-{^{13}C}-J-HMBC spectrum by fitting the intensity I of the associated $^1H–^{13}C$ correlation in each of the 16 individual 2D $^1H–^{13}C$ spectra to $I = I_0\sin(\pi J\tau)$ using a two-parameter (I_0, J) non-linear regression as implemented with SigmaPlot 4.1 (Jandel Scientific). With $\tau_{max} = 0.5$ s, $^3J(^{13}C–^1H) \geq 2.5$ Hz could be determined, while imposing an upper limit of 2.5 Hz for smaller couplings. Additional Fourier transformation of the $^3J(^{13}C–^1H)$ evolution time, which would lead to a 3D J-resolved $^1H–^{13}C$ spectrum, was not performed. Indeed, the limited resolution which can be achieved along the J-resolved frequency dimension with the experimental setup used, precludes more accurate $^3J(^{13}C–^1H)$ measurements than those obtained from the simulations above.

Molecular modeling

All molecular modeling was performed using the Biosym 95.0 software purchased from MSI (Scranton Road 9685, San Diego, CA) on a Silicon Graphics Indigo2 computer. The Discover 3.0 module was used throughout. The results were analysed using the Analysis and Decipher modules within Biosym 95.0. Both cvff and esff force fields were used. All simulations were performed *in vacuo*. No cut-offs were considered to calculate non-bonding interactions. A relative permittivity ε of 2.4 (benzene) was chosen and considered distance dependent in order to (crudely) mimic solvent screening effects. Distance and torsion angle restraints were imposed using a harmonic potential function with forcing constants of 167.36 kJ mol^{-1} Å^{-2} (40 kcal mol^{-1} Å^{-2}). After initial energy optimisation, the starting structure[26] was submitted to 5 ps of free MD using random starting velocities corresponding to a Maxwell–Boltzmann distribution at 1000 K. In protocol I, this was followed by 5 ps of MD during which the distance restraints were gradually imposed through linear scaling of the associated force constants (from 0 to 1 in 4 steps). This was followed by a 250 ps rMD trajectory which was sampled every 5 ps, resulting in a set of 50 conformers. Each conformer was annealed from 1000 to 300 K by lowering the temperature in steps of 100 K, followed by 1 ps rMD. At 300 K, another 5 ps of rMD was applied prior to energy optimisation, using a cascade of steepest descent, conjugate gradient and Newton–Raphson energy minimisation until the maximum derivative dropped below 4.184 J mol^{-1} Å^{-1}. Protocol II differed only from protocol I in that no distance restraints were imposed in the first stage of the computation, yielding a set of 50 conformers judged to be representative of the conformational space available to Ery-A. Each of these were used as starting conformations for 5 ps rMD at 1000 K during which the

distance restraints were gradually imposed to full effect. Following another 5 ps of 1000 K rMD, the conformers were annealed and their energy optimised as for Protocol I. Chiralities were continuously restrained to avoid chirality inversion with a force constant of 418.4 kJ mol^{-1} rad^{-2}.

Acknowledgements

The authors are indebted to one of the referees whose valuable comments helped to improve the discussion. J. C. M. is a post-doctoral research associate of the Fund for Scientific Research Flanders (FWO-Belgium). The authors are indebted to the Fund for Scientific Research Flanders (Belgium) for financial support: FKFO, Grant no. 2.0094.94; Nationale Loterij, Grant no. 9.0006.93 and 9.0192.98; FWO Grant no. G.0192.98.

References

1 (*a*) *Macrolide Antibiotics*, *Chemistry*, *Biology and Practice*, ed. S. Omura, Academic Press, New York, 1984; (*b*) J. A. Washington II, J. A. Wilson and W. R. Wilson, *Mayo Clin. Proc.*, 1985, **60**, 189; (*c*) A.-S. Malmborg, *J. Antimicrob. Chemother.*, 1986, **18**, 293.
2 D. R. Harris, S. G. McGeachin and H. H. Mills, *Tetrahedron Lett.*, 1965, **11**, 679.
3 E. F. Gale, E. Cundliffe, P. E. Reynolds, M. H. Richmond and M. J. Waring, *The Molecular Basis of Antibiotic Action*, John Wiley and Sons, London, 1981.
4 (*a*) J. C.-H. Mao and R. G. Wiegland, *Biochim. Biophys. Acta*, 1968, **157**, 404; (*b*) H. Teraoka and K. H. Nierhaus, *J. Mol. Biol.*, 1978, **126**, 185; (*c*) J. R. Menninger and D. P. Otto, *Antimicrob. Agents Chemother.*, 1982, **21**, 811; (*d*) S. Douthwaite, T. Powers, J. Y. Lee and H. F. Noller, 1989, *J. Mol. Biol.*, 1989, **209**, 655; (*e*) L. H. Hansen, P. Mauvais and S. Douthwaite, *Mol. Microbiol.*, 1999, **31**, 623; (*f*) W. S. Champney and R. Burdine, *Antimicrob. Agents Chemother.*, 1996, **40**, 1301.
5 J. C. Martins, F. Kayser, M. Gielen, R. Willem and M. Biesemans, *J. Magn. Reson.*, 1997, **124**, 218.
6 J. C. Martins, R. Willem, F. A. G. Mercier, M. Gielen and M. Biesemans, *J. Am. Chem. Soc.*, 1999, **121**, 3284.
7 (*a*) G. Némethy, M. S. Pottle and H. A. Sheraga, *J. Phys. Chem.*, 1983, **87**, 1883; (*b*) S. J. Weiner, P. A. Kollmann, D. A. Case, C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *J. Am. Chem. Soc.*, 1984, **106**, 765; (*c*) B. R. Brooks, R. E. Bruccoleri, B. O. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187; (*d*) P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, M. Genest and A. T. Hagler, *Proteins: Struct., Funct. Genet.*, 1988, **4**, 1243.
8 (*a*) S. Barlow, A. L. Rohl, S. Shengua, C. M. Freeman and D. O'Hare, *J. Am. Chem. Soc.*, 1996, **118**, 7578; (*b*) S. Shengua, L. Yan, Y. Yang and J. Shaulsky, *Biosym Solutions August Issue*, 1995, 8.
9 J. R. Everett and J. W. Tyler, *J. Chem. Soc.*, *Perkin Trans. 1*, 1985, 2599.
10 J. Gharbi-Benarous, P. Ladam, M. Delaforge and J. P. Girault, *J. Chem. Soc.*, *Perkin Trans. 2*, 1993, 2303.
11 A. Bax and D. G. Davis, *J. Magn. Reson.*, 1985, **65**, 355.
12 U. Piantini, O. W. Sørensen and R. R. Ernst, *J. Am. Chem. Soc.*, 1982, **104**, 6800.
13 D. M. Doddrell, D. T. Pegg and M. R. Bendall, *J. Magn. Reson.*, 1982, **48**, 323.
14 J. Keeler, R. Clowes, A. L. Davis and E. D. Laue, *Methods Enzymol.*, 1994, **239**, 145.
15 A. Bax, R. Griffey and B. L. Hawkins, *J. Magn. Reson.*, 1983, **55**, 301.
16 A. Bax and M. F. Summers, *J. Am. Chem. Soc.*, 1986, **108**, 2093.
17 J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546.
18 J. R. Everett and J. W. Tyler, *J. Chem. Soc.*, *Chem. Commun.*, 1987, 815.
19 J. R. Everett, E. Hunt and J. W. Tyler, *J. Chem. Soc.*, *Perkin Trans. 2*, 1991, 1481.
20 J. Barber, J. I. Gyi, G. A. Morris, D. A. Pye and J. K. Sutherland, *J. Chem. Soc.*, *Chem. Commun.*, 1990, 1040.
21 J. Gharbi-Benarous, P. Ladam, M. Delaforge and J. P. Girault, *J. Chem. Soc.*, *Perkin Trans. 2*, 1992, 1989.
22 J. R. Everett, I. K. Hatton, E. Hunt, J. W. Tyler and D. J. Williams, *J. Chem. Soc.*, *Perkin Trans. 2*, 1989, 1719.
23 J. M. McGill and R. Johnson, *Magn. Reson. Chem.*, 1993, **31**, 273.
24 C. A. G. Haasnoot, F. A. A. M. De Leeuw and C. Altona, *Tetrahedron*, 1980, **36**, 2783.
25 (*a*) G. Wagner, W. Braun, T. F. Havel, T. Schaumann, N. Go and K. Wüthrich, *J. Mol. Biol.*, 1987, **196**, 611; (*b*) K. Wüthrich, M. Billeter and W. Braun, *J. Mol. Biol.*, 1983, **169**, 949.
26 Since the X-ray structure of erythromycin A is not available from the Cambridge Crystallographic Database, the coordinates of the otherwise very similar clarithromycin (6-methoxy-erythromycin A) were used instead: H. Iwasaki, Y. Sugawara, T. Adachi, S. Morimoto and Y. Watanabe, *Acta Crystallogr.*, *Sect. C*, 1993, **49**, 1227.
27 I. Tvaroska, M. Hricovini and E. Petrakova, *Carbohydr. Res.*, 1989, **189**, 359.
28 P. Ladam, J. Gharbi-Benarous, M. Piotto, M. Delaforge and J. P. Girault, *Magn. Reson. Chem.*, 1994, **32**, 1.
29 J. R. Everett and J. W. Tyler, *J. Chem. Soc.*, *Perkin Trans. 2*, 1987, 1659.
30 S. G. Hyberts, M. S. Goldberg, T. F. Havel and G. Wagner, *Protein Sci.*, 1992, **1**, 736.
31 D. Marion and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 1983, **113**, 967.
32 R. Willker and D. Leibfritz, *Magn. Reson. Chem.*, 1995, **32**, 665.
33 (*a*) S. Macura, B. T. Farmer II and L. R. Brown, *J. Magn. Reson.*, 1986, **70**, 493; (*b*) R. Willem, M. Biesemans, K. Hallenga, G. Lippens, F. Malaisse-Lagae and W. J. Malaisse, *J. Biol. Chem.*, 1992, **267**, 210.
34 K. Wüthrich, M. Billeter and W. Braun, *J. Mol. Biol.*, 1983, **169**, 949.

Paper 8/10003I