Supporting Information

Fully Automated Quantum-Chemistry-Based Computation of Spin–Spin-Coupled Nuclear Magnetic Resonance Spectra

Stefan Grimme,* Christoph Bannwarth, Sebastian Dohm, Andreas Hansen, Jana Pisarek, Philipp Pracht, Jakob Seibert, and Frank Neese

anie_201708266_sm_miscellaneous_information.pdf

Contents

1 General outline

The procedure starts with a chemically reasonable three-dimensional structure (Cartesian coordinates) as input, i.e., this geometry must be in line with the constraints imposed by the bonding pattern (cf. constitutional isomerism) and the relative configuration of the compound. The following steps are conducted:

1. Generation of the conformer/rotamer ensemble (CRE) using the GFN-xTB semiempirical tight-binding method[1] and a newly developed CRE search algorithm

2. DFT based free energy ranking of conformers to obtain equilibrium (Boltzmann) populations

3. Computation of chemical NMR shielding constants (chemical shifts, δ) and spin-spin coupling constants (J, SSCC) for all conformers found with more than 4% population

4. Fully automatic analysis of the CRE to determine magnetically as well as chemically equivalent atoms, averaging of the J/δ values over the CRE, possibly fragmentation of the spin-system for large cases, and diagonalization of the spin-Hamiltonian

For steps 2 and 3, well established quantum chemistry methods are employed, which are mainly based on (double-)hybrid DFT calculations with sufficiently large (converged) atomic orbital (AO) basis sets. The conformational search algorithm was developed in the course of this project and is implemented in the `xtb` code[2] and attached scripts. For step 4, new programs were written by S. Grimme.

In principle the approach is applicable to compounds composed of the first 86 elements of the periodic table for which GFN-xTB has been parametrized. However, some of the special (property optimized) Gaussian AO basis sets used for the NMR parameter calculations are only available for elements up to Ar, so that in such cases standard (energy optimized) basis sets must be used (e.g., def2-TZVP[3]). In addition, for heavier elements
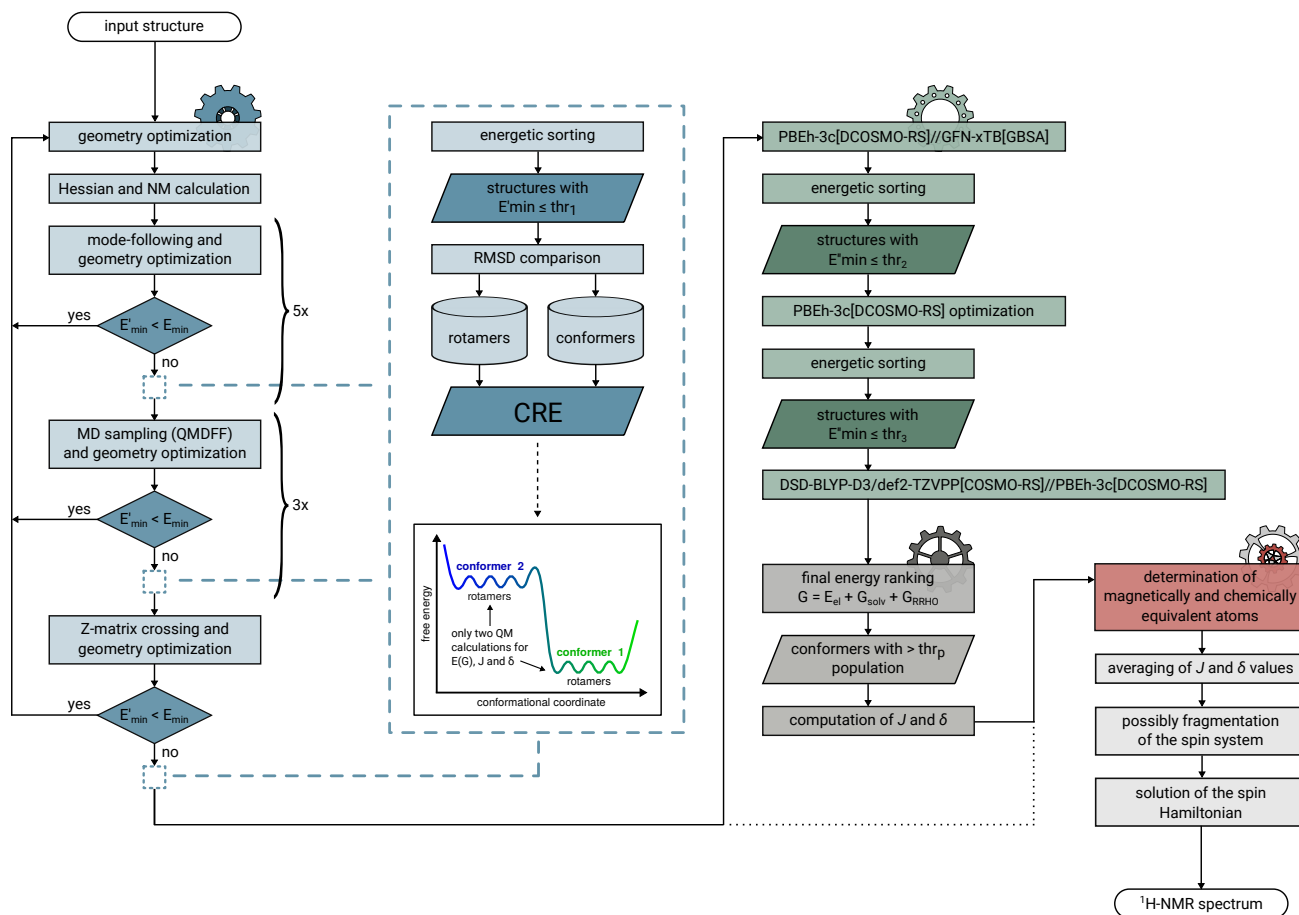
Figure 1: Detailed flow-chart of the composite procedure for the automatic computation of NMR spectra.

inclusion of scalar relativistic effects in the NMR parameter calculation is necessary which is not considered here.

The following list summarizes the standard theoretical levels (which are outlined in more detail below) for the various steps (acronyms in brackets denote a solvation model):

- conformer/rotamer ensemble: GFN-xTB[GBSA]

- optimized structures for all further single-point calculations: PBEh-3c[4][DCOSMO-RS] (TPSS-D3/def2-TZVP[DCOSMO-RS] as an alternative for transition metal compounds)

- electronic energies for Boltzmann populations: DSD-BLYP/def2-TZVPP[5] (D3(BJ) version with with damping parameters s_6=0.57, a_1=0, s_8=0, a_2=5.4[6])

- solvation free energies for Boltzmann populations: COSMO-RS

- vibrational frequencies for thermostatistical corrections to free energy: GFN-xTB[GBSA] (alternatively: PBEh-3c or TPSS-D3/def2-TZVP)

- NMR shieldings: PBE0[7]/pcSseg-2[8][CPCM]

- NMR spin-spin coupling constants: PBE0/pcJ-0+pol[9][CPCM]

The chosen methods represent the best compromise between computational effort and target accuracy. Solvent effects are implicitly accounted for in all parts by appropriate continuum solvation models (GBSA,[10] COSMO-RS,[11,12] DCOSMO-RS,[13] CPCM[14]). The SMD model[15] was considered for testing in initial stages of the project and may replace (D)COSMO-RS but so far, no detailed comparisons between both methods for conformational/NMR problems are available.

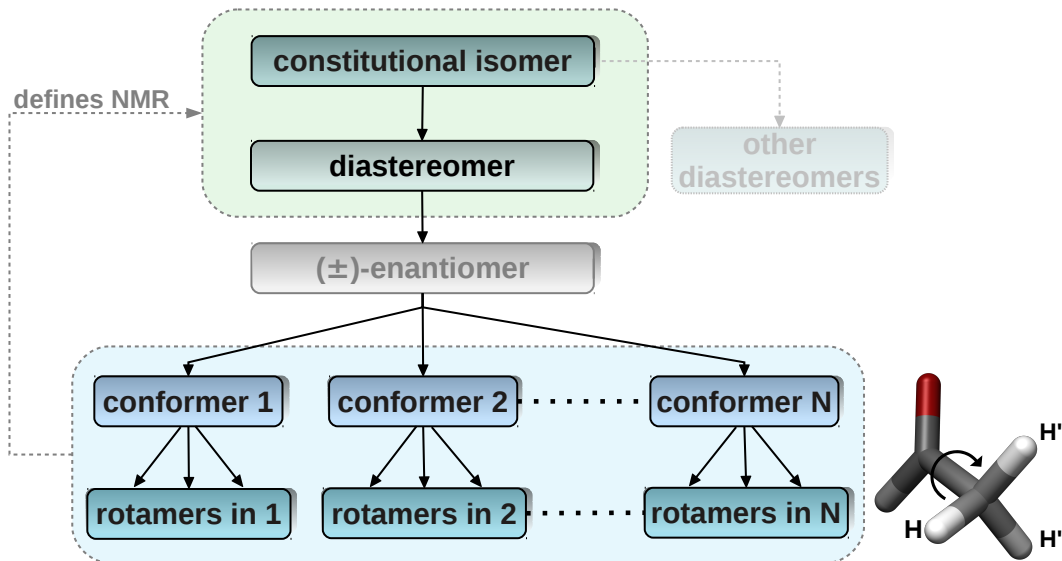# 2 Generation of the conformation/rotamer ensemble

Figure 2: Hierarchy of isomerism that is relevant for the presented NMR procedure. A compound with a given constitution and relative configuration (in case stereocenters are present) will show a distinctive NMR spectrum, making it distinguishable from other diastereomers, as well as other constitutional isomers. The absolute configuration (i.e., the choice of the enantiomer) does not play a role for NMR, hence, for chiral compounds, the choice of the enantiomer is arbitrary in the computation. The slow time scale of the NMR measurement allows the ensemble of solvated molecules to pass through all thermally accessible conformations including the respective rotamer sub-minima. This CRE determines the NMR spectrum of the molecule. As an example, the only conformer of acetaldehyde is shown, which consists of three rotamers that are generated by rotation around the C–C bond.

A conformer belongs to a set of stereoisomers, each of which is characterized by a conformation corresponding to a distinct potential energy minimum.[16] Rotamers arise from restricted rotation around chemical bonds[17] leading to interchange of nuclei that belong to the same group of nuclides (e.g., the interchange of ^{1}H nuclei at a methyl group). This leads to minima, which are not distinguishable by any spin-independent quantum mechanical observable computed at the respective minima. Since nuclei are treated classically in Born-Oppenheimer approximated calculations, we can identify rotamers as having different Cartesian coordinates (per atom), while being identical in the aforementioned observables

(we use rotational constants and energies). Rotamers contribute to the molecular entropy. Furthermore, they are of utmost importance in the context of NMR, because they embody the rapid nuclei interchange (sub ps time scale). On the slow time scale of the experiment, this rapid interchange of nuclei results in an averaging of NMR parameters (shifts and SSCC). Hence, for each molecule (of a given relative configuration), only a single set of CRE-averaged NMR parameters is employed in the spin-Hamiltonian. In passing, we note that the time-dependency of the spin-Hamiltonian is not considered here. Life-time (Heisenberg) line-broadening is included by the finite line-width of the Lorentzian type signals. Note that the distinction between conformers and rotatmers is mainly made to reduce computational effort, i.e., the NMR parameter calculation is conducted only once for one rotamer of each conformer.

If Y_A denotes an NMR parameter for nucleus A (or the coupling constant for an atom pair), the average is computed as

$$Y_A^{av} = \sum_j^{CRE} p_j Y_A^j \tag{1}$$

over the complete CRE with normalized populations p_j for species j. They are given by

$$p_j = \frac{\exp(-\Delta G_j/RT)}{\sum_i^{CRE} \exp(-\Delta G_i/RT)} \tag{2}$$

at absolute temperature T (R is the molar gas constant) for a given relative free energy ΔG_j of species j. In the above formula the average is taken over the complete GFN-xTB CRE, i.e., it is assumed that each conformer has the same (dynamic) spin-symmetry which should hold under normal conditions and elevated temperatures. In the rare case that artificial conformers from GFN-xTB are included in the CRE, such structures have to be removed manually before the spin-symmetry analysis is conducted.

The absolute free energy G of a minimum structure (conformer) is obtained as the

sum of the electronic energy E_{el}, solvation free energy G_{solv}, and a modified rigid-rotor-harmonic-oscillator (RRHO) term[18]

$$G = E_{el} + G_{solv} + G_{RRHO} \qquad (3)$$

and the difference is conveniently taken with respect to the structure that has the lowest free energy. For benchmarks of gas phase conformational energies and a performance evaluation of standard density functionals see Ref.[19] Due to the very good performance observed in those studies, a double-hybrid functional[20] (i.e., DSD-BLYP) method was chosen as default (single-point) QC method. As a fall-back method, e.g., for complicated transition metal cases we propose the standard PW6B95 functional,[21] which is employed for many years in our group as a default hybrid.

2.1   The composite search procedure for the CRE

The proposed algorithm consists of three steps: normal mode following (MF), molecular dynamics (MD), and 'genetic' structure crossing (GC). The procedure is dubbed MF-MD-GC (or more precisely MF-MD-GC//GFN-xTB because in all steps GFN-xTB is used for the underlying quantum chemistry).

The basic idea of MF has been put forward already some time ago[22] and suggested for finding transition states at the semiempirical PM6-DH+ level (see Ref.[23]). The search can be conducted in the gas phase or in solvent simulated by the Generalized Born Solvent Area (GBSA) implicit solvation model.[10] The structure of interest is first fully optimized and subsequently the harmonic vibrational frequencies are computed. The main idea is that energetically low-lying, thermally accessible conformers and rotamers can be generated in a physically plausible way by displacement along the low-frequency normal modes (NMs). New minima are automatically detected on the one-dimensional potential energy curves generated this way and fully optimized. The approach can work in the standard NM

basis or with Pipek-Mezey chemical group localized modes.[24,25] All modes refer to a mass-unweighted coordinate system. One of the basic ideas is to avoid costly and possibly non-robust non-linear coordinate systems as far as possible. However, it is well known[26] that large nuclear displacements are not possible in the NM system. Therefore, the simple displacement of the Cartesian coordinates r along the target (or search) NM ϕ_i (both are vectors of length $3N$ where N is the number of atoms) with step length s

$$r_{new} = r_{old} + s\phi_i \tag{4}$$

has to be modified. Our algorithm is based on the fact that the structure optimizer in the `xtb` code also works in (approximate) NM coordinates. Thus, after displacement (which refers to one point on the potential energy curve), an incomplete geometry optimization for n_o steps ($n_o = 5 - 15$) is conducted where the target mode ϕ_i is projected out from the internal NM optimization coordinate system. This ensures that all degrees of freedom other than the search (or reaction) coordinate are relaxed. This way, non-linear effects are also effectively included. In complicated cases or for large displacements, the search direction can optionally be updated after each displacement according

$$\phi_i^{k+1} = \phi_i^k + t_u(r^{k-1} - r^k) \tag{5}$$

where k is a counter for the number of displacements and t_u is an update scaling factor with a typical value of 0.1-0.3.

The two figures below show exemplary potentials with very low (n-butane) and high (o-dichlorbiphenyl) conformational barriers. It is seen that smooth potentials without any spikes or discontinuities are obtained. From those, minima can be identified and fully (unconstrained) optimized. At first glance, it seems to be in principle possible to obtain also the barriers, in addition to the minima, which would give rise to the kinetics and possibly line-broadening effects in the NMR experiment. However, inspection of many
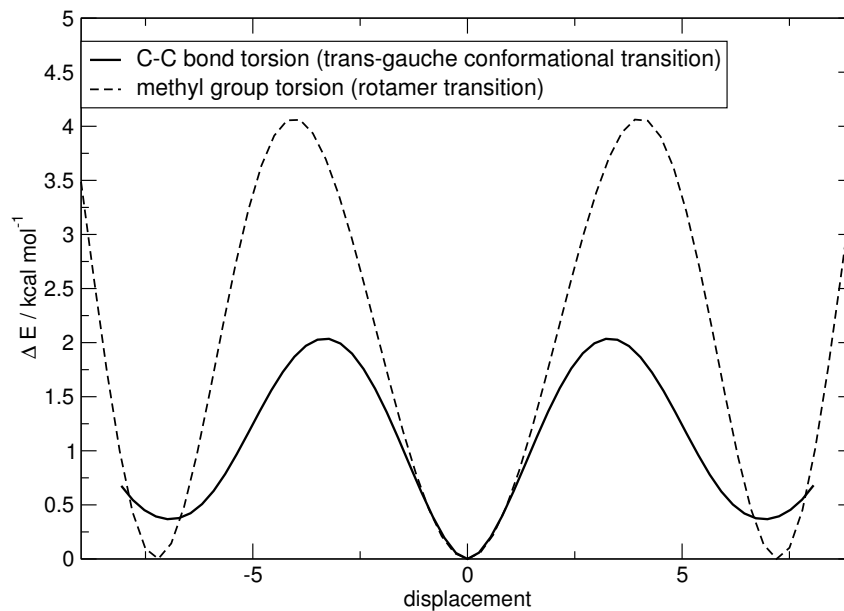
9

Figure 3: Potential energy curves obtained from two lowest followed NM (C-C bond and methyl group torsion, respectively) in n-butane using $t_u = 0.2$, $n_o = 10$, and $s = 1$.
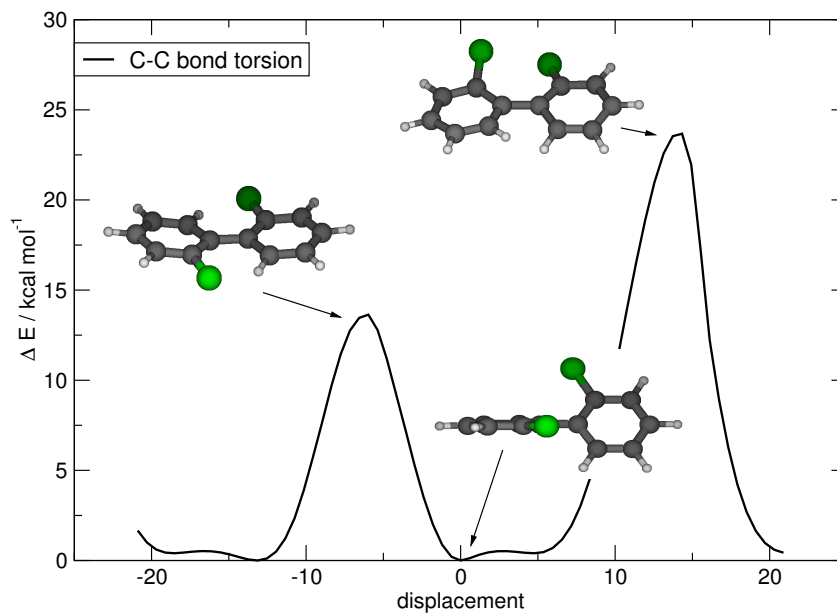

Figure 4: Potential energy curve obtained from lowest followed NM in o-dichlorbiphenyl using $t_u = 0.2$, $n_o = 5$, and $s = 2$.

potential curves in more complicated molecules show that this is not possible in an automatic way. Empirical evidence indicates that following modes with a harmonic frequency of 300–400 cm^{-1} leads to a CRE containing structures which are thermally accessible at around room temperature. A cross check on a few high-barrier cases (e.g., 1,3-di-tbutyl-octatetraene and 2-chloro-trans-4,6-dimethyl-1,3,2-dioxarsenane) show that barriers of 20–25 kcal mol^{-1} cannot be overcome correctly with the default settings of our algorithm.

More critical for the completeness of the CRE (which is crucial for the entire proposal) are the technical settings for the algorithm. In addition to the parameters described above, the number of displacements n_p is also varied. After some computer experiments for a wide variety of molecules, we propose a composite procedure in which the complete procedure for all modes below the frequency threshold (i.e., typically 5-100 modes) is run several times with different settings as given in the table below.

Table 1: Technical settings for the MF-MD-GC algorithm including additional steps, see text for details.

run	modes	s	n_o	t_u	n_p
1	normal	2	5	0.5	20
2	normal	0.5	5	0.0	30
3	normal	3	15	0.1	40
4	local^a	1	5	0.5	20
5	local^a	2	5	0.1	40
6-7	MD			—	
8	genetic crossing			—	

^{a} Reduced number of modes ($\frac{2}{3}$ of normal value).

The optimized structures from all eight runs are gathered and analyzed as described in the following. In steps 6-7 in Table 1, molecular dynamics runs of 500 ps length at three temperatures (T=300,400,500 K) in the NVT ensemble are conducted from which equidistantly taken snapshots are optimized (100 per temperature frame) which are added to the CRE. This additional step ensures that strongly delocalized, but mostly uncoupled conformational movements (e.g., many methyl groups) are treated adequately. Because the relatively long trajectory calculation would become a bottleneck of the entire search

11

procedure, the QMDFF[27] molecule specific force-field is used for the MD instead of GFN-xTB itself (the FF is, however, derived from GFN-xTB). Alternatively, however, GFN-xTB can be used in critical cases also for the MD.

If in any step of the above procedure an energetically lower-lying conformer is found, the whole treatment is restarted with a Hessian and NM calculation.

This algorithm works very well for small and medium sized systems but often fails to provide a sufficiently large number of rotamers for larger cases with many weakly coupled rotatable functional groups. To circumvent this problem an extension of the treatment termed 'genetic crossing' (GC) was implemented which employs ideas from genetic optimization algorithms.

This approach employs internal (Z-matrix, R) coordinates and takes the energetically lowest structure found so far (R_{ref}) as reference, i.e., to define the connectivities. From the existing CRE, all pairs of structures are considered and a new structure is generated by taking the differences to the reference over all internal coordinates (i.e., bond length, bond angles, an dihedral angles) according to

$$R_{new} = R_{ref} + R_i - R_j \qquad (6)$$

where ij label the pairs, and R_{new} is the generated new structure, which is subjected to a full geometry optimization. In this way, structural differences (e.g., a methyl group rotation) relative to R_{ref} present only in i and j are combined in the resulting new conformer/rotamer. The repeated application of the crossing algorithm is in principle able to recover the full permutational complexity of the rotamer generation problem. However, in practice only one run is conducted and the number of optimized structures is limited to a user defined value (2000 by default). In order to enhance diversity in the so extended CRE, the generated (crossed) structures are sorted according to Cartesian root-mean-square deviation (RMSD) with respect to R_{ref}, i.e., more different ones are taken preferentially. Improved efficiency is also achieved by considering only R_{new} for optimization which avoid any close atomic

contacts (i.e., applying 'clash' checks based on changes of the D3 coordination number[28]). The number of surviving R_{new} is further increased by working in a specially designed Z-matrix coordinate system which minimizes the number of inter-dependent dihedral angle definitions. This only requires a re-ordering of the atoms in the Cartesian coordinates in the very beginning of the entire procedure (automatically done by a sorting tool).

An important task in the procedure after each step in the above table is to eliminate identical molecules and to discriminate between conformers and rotamers. Two molecules are taken as identical if their Cartesian RMSD as obtained from an all atom best-fit using a quaternion algorithm[29] is less than a default threshold $t_{RMSD} = 0.2$ Å. Non-identical molecules are rotamers if their relative deviation in the average rotational constants is less than $t_{rot} = 0.02$ and their relative deviation of the nuclear repulsion energy is less than $t_{rep} = 0.0001$. These values are based on very tight `xtb` code optimization thresholds and need re-adjustment if more sloppy geometry optimization criteria, e.g., for very large molecule are applied.

The entire CRE search procedure is solid and robust but requires thousands of energy/gradient evaluations already for a medium sized molecule and hence is only possible at a semiempirical QM level. Without the reasonably accurate GFN-xTB method (including the GBSA solvation model), the entire approach would not be possible. However, major parts of the computations can be run in parallel so that reasonable turnaround times are observed on typical 16-core workstations. The computer timings are roughly minutes for about 20-30 atoms, about an hour for 50-100 atoms, and overnight for 200 atoms. These values are very crude estimates because the computational effort depends on many factors (e.g., quality of the starting structure, conformational flexibility, electronic structure of the molecule).

2.2 Completeness of the ensemble

The above described so-called MF-MD-GC//GFN-xTB CRE generation method has been tested on a very large number of molecules including transition metal complexes. In great detail we have considered organic molecules and in particular a set of 100 drug molecules for which comparisons to an established conformer search algorithm, which is applied routinely in industry, can be made (see section 6). Here, the approach is first tested for the observable molecular entropy which strongly depends on the completeness of the conformer *and* rotamer ensemble. For medium sized test molecules, harmonic vibrational values augmented by the conformer/rotamer (CR) entropy given by

$$S_{tot} = S_{RRHO} + S_{CR} \tag{7}$$

where S_{RRHO} is the modified rigid-rotor-harmonic-oscillator entropy[18] and

$$S_{CR} = R \sum_i^{CRE} p_i \log p_i \tag{8}$$

is computed as a sum over all species found with population p_i and relative solvation-exclusive free energy ΔG_i. In Eq. 7, S_{RRHO} refers to the value computed for the lowest energy conformer. These computations are entirely done at the GFN-xTB level with E_i referring to the corresponding conformational electronic energy. Reference entropies are taken from the NIST database[30] and a comparison of MF-MD-GC//GFN-xTB values obtained with the default settings is given in Tab. 2. Note that our approach includes a special rotor-treatment in the RRHO part for the low-energy modes in order to avoid artifacts by the harmonic approximation and numerical errors.[18]

As can be seen from the data, the CR contribution is substantial with around 10% for alkane chains. Addition of the CR entropy to the RRHO result improves the agreement with experiment although we note a systematic overestimation for the smaller alkanes compared to the larger ones which is likely caused by the RRHO value (neglected anhar-

Table 2: Comparison of experimental and MF-MD-GC//GFN-xTB computed absolute gas phase entropies S (in J/mol K at $T =$ 298 K) for various alkanes. The conformational and theoretical RRHO values are also given.

molecule	sum formula	$S_{exptl.}$	$S_{theor.}$	S_{RRHO}	S_{CR}
n-pentane	C_5H_{12}	347.8	369.5	334.6	34.9
cyclo-hexane	C_6H_{12}	298.2	280.9	274.8	6.1
n-hexane	C_6H_{14}	388.8	405.9	364.5	41.4
3-methyl-pentane	"	382.9	409.1	367.0	42.1
3-ethyl-pentane	C_7H_{16}	411.5	430.9	386.1	44.8
2,4-dimethyl-pentane	"	396.7	432.3	388.7	43.6
cyclo-octane	C_8H_{16}	366.8	391.9	360.9	31.0
n-octane	C_8H_{18}	467.1	469.3	422.4	46.9
3-methyl-heptane	"	465.9	474.2	426.3	47.9
n-decane	$C_{10}H_{22}$	545.8	530.4	479.3	51.1
2-methyl-decane	$C_{11}H_{24}$	578.3	563.3	511.3	52.0
n-dodecane	$C_{12}H_{26}$	622.5	589.1	535.5	53.6

monicity for the low-frequency vibrational modes, see e.g., Ref.[31] for a recent proposal to include them). In any case, the data show that for the very flexible alkanes our approach yields a relatively complete CRE which is a prerequisite for the successful computation of NMR spectra of many molecules.

2.3   Free energy ranking and filtering

All GFN-xTB[GBSA] conformers with relative free energies (excluding the RRHO free energy contribution at this point but including the solvation part) less than an energy window threshold of $thr_1 = 6$ kcal mol^{-1} are always kept within the MF-MD-GC treatment in the CRE (left part of Figure 1). They are subjected to a few well defined filtering steps to maintain high accuracy at low computational cost. These are

1. PBEh-3c[DCOSMO-RS]//GFN-xTB[GBSA] single-point energy (for all GFN-xTB[GBSA] conformers)

2. PBEh-3c[DCOSMO-RS] optimization (if single-point PBEh-3c[DCOSMO-RS]//GFN-xTB[GBSA] energy is less than $thr_2 = 3$ kcal mol^{-1})

3. DSD-BLYP[COSMO-RS]//PBEh-3c[DCOSMO-RS] single-point energy (if PBEh-3c[DCOSMO-RS] energy is less than $thr_3 = 3$ kcal mol^{-1})

The final ranking includes the GFN-xTB RRHO free energy contribution and the solvation free energy term as given in Eq.3. All conformers with populations larger than $thr_p = 4$ % at this final free energy level are subjected to the NMR parameter calculations.

The PBEh-3c composite hybrid functional method[4] is our default QM optimization level. Alternatively, the somewhat slower TPSS[32]-D3/def2-TZVP(-f)[3] level can be used which performs robustly also for electronically difficult (e.g., transition metal) complexes where hybrids are often worse. For transition metal complexes, double-hybrid density functionals are less accurate and we recommend to switch to the PW6B95-D3 functional in such cases.[21] If the GFN-xTB[GBSA] conformational space is small, it is recommended to skip step one in the above filtering procedure and to directly start with the PBEh-3c optimizations. For long lists (hundreds of conformers), however, it is more efficient to restrict the number of calculations by the first PBEh-3c filtering step. All DFT energies are D3 dispersion-corrected,[28,33] while GFN-xTB and PBEh-3c have D3 'built-in'.

The question if the systematically longer bond lengths with TPSS than with hybrids[4,34] are beneficial for NMR parameter calculations (because they are on average closer to R_0 than to R_e values, i.e., the calculations are done on an effective equilibrium structure) is presently not clearly answered and has to await further studies. For details on vibrationally averaged shielding computations where this is considered explicitly see Ref.[35]

The final conformational energy method DSD-BLYP was chosen based on its very good performance for general chemistry (conformational energies and non-covalent interactions in particular) in the GMTKN55 database.[19] According to the results from this database, ωB97x-D3[36] is competitive and represents an alternative. All calculations reported here, refer to a standard energy protocol which is DSD-BLYP[COSMO-RS]/def2-TZVPP//PBEh-3c[DCOSMO-RS], although we note that similar results for conformational free energies may be obtained at, e.g., the ωB97x-D3[SMD] level. Further studies of the best practical

theory level for conformational energies in solution are ongoing in our laboratory. One issue here is the numerical accuracy of the applied solvation models for large systems which presently is not satisfactory and not allowing tightly converged geometry optimizations.

As a typical example we show in Figure 5 the conformational free energies at the three relevant theoretical levels (GFN-xTB[GBSA], PBEh-3c[DCOSMO-RS], DSD-BLYP[COSMO-RS]//PBEh-3c[DCOSMO-RS]) for the case of guaiol consisting of 20 conformers in the 3 kcal mol^{-1} default energy window ($thr_2 = thr_3$, see Figure 1).
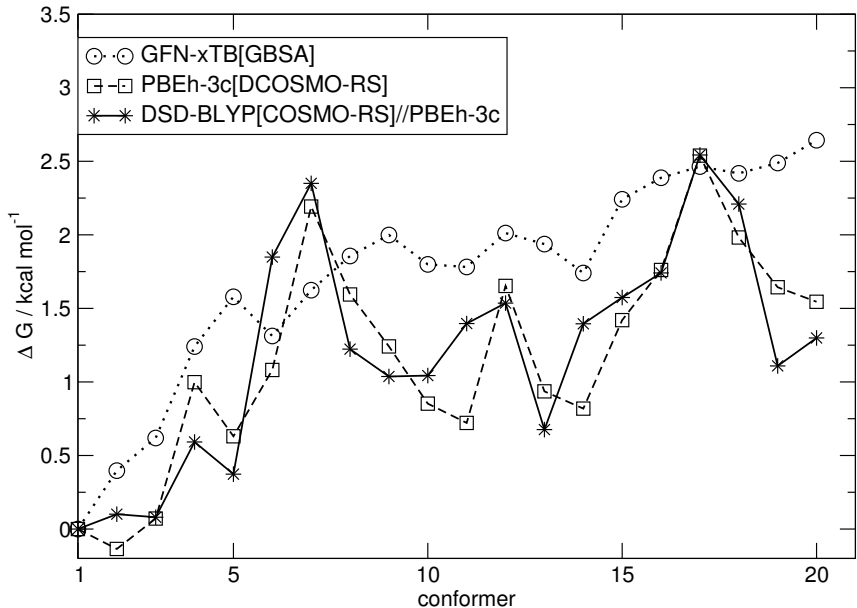

Figure 5: Conformational free energies (using the GFN-xTB RRHO contribution) of guaiol in CHCl$_3$. The lines in between the data points are just drawn to guide the eye and the structures are sorted according to the GFN-xTB result without the RRHO term.

It can be seen that for this relatively simple case all three methods agree well indicating that both, the electronic energy as well as solvation free energy surfaces of the methods are rather parallel. Although we note (not unexpectedly) larger deviations of GFN-xTB from both DFT energies than between the low- and high-level DFT results, the semiempirical method correctly reproduces the trend and provides correctly the low-lying structures.

A more difficult case is shown in Figure 6 with the example of the natural product salvinorin A (see inset for the Lewis formula, same technical settings as in the previous
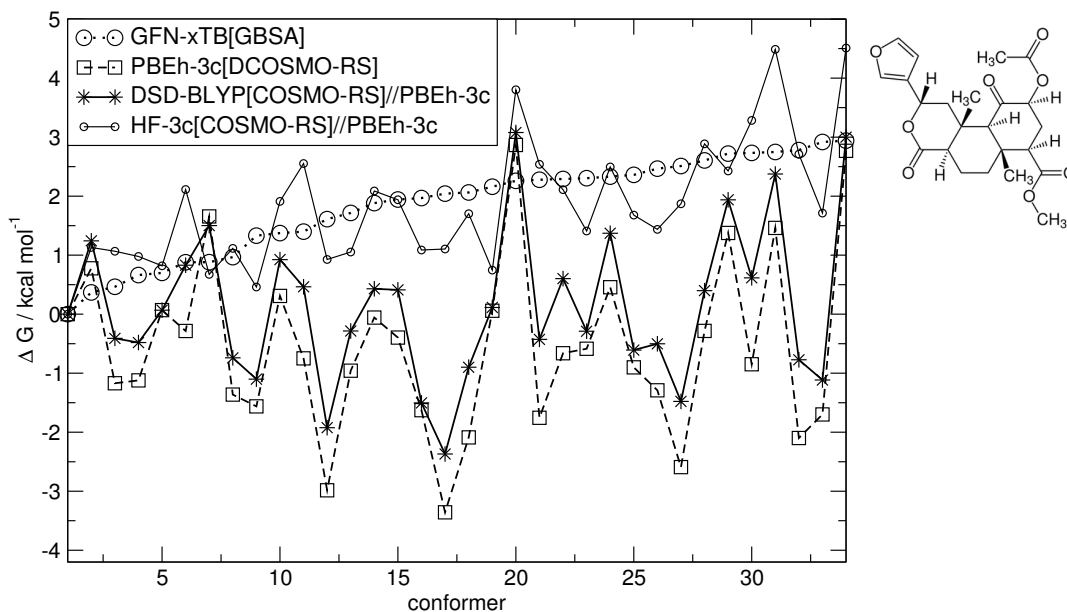
example).

Figure 6: Conformational free energies (without the GFN-xTB RRHO contribution) of salvinorin A in CHCl$_3$. The lines in between the data points are just drawn to guide the eye and the structures are sorted according to the GFN-xTB result without the RRHO term.

Here, small deviations between both DFT methods but a substantially different conformer ordering at the GFN-xTB level are observed. That this example is really difficult becomes clear by comparison of the DFT with the HF-3c[37] result which also employs COSMO-RS but a minimal AO basis set similar to GFN-xTB. Thus, the gas phase electronic problem (and not the geometry or the solvation model) is crucial. We show this example of a seemingly simple organic compound here to demonstrate the complexity of the conformation problem with its various aspects and stress that further quantum chemical method development is definitely needed. Nevertheless, the ensemble generated from the 3 kcal mol^{-1} default GFN-xTB energy window very likely contains the most stable species at the highest DFT level which is the key prerequisite for the subsequent calculation steps.

3 Computation of chemical NMR shielding and spin-spin coupling constants (SSCC)

3.1 General

We employ hybrid DFT combined with Jensen's property optimized Gaussian AO basis sets for nuclear magnetic shielding (pcS-X[8]) and spin-spin coupling (pcJ-X[9]) calculations where X denotes a cardinal number indicating increasing basis set completeness. These basis sets provide already for small to medium sized sets (X=0-2) results closer to the complete basis set (CBS) limit than standard, energy optimized sets.[38] For shieldings we use the segmented variant[39] pcSseg-2 which is essentially a basis set of about triple-ζ to quadruple-ζ quality. For elements for which the pcS/J basis sets are not available (i.e. $Z > 18$), the Ahlrichs' def2-TZVP set[3] is taken as a substitute.

The computed hydrogen chemical shifts δ with medium sized AO basis sets show small very systematic deviations from experimental or high level CCSD(T) values.[38,40,41] This can be corrected by linear regression of computed chemical shieldings σ to experimental values or by simple scaling of the δ values. However, we just convert shieldings to the shift scale as usual using TMS as standard without any empirical scaling, i.e.,

$$\delta_A^{theor.} = \sigma_{TMS} - \sigma_A \tag{9}$$

where σ_A is the computed shielding in ppm for the nucleus of interest and σ_{TMS} is the value of the reference compound. We employ σ_{TMS} values of 31.538 ppm (^{1}H) and 187.34 ppm (^{13}C) which refer to our standard theoretical level.

The computed hydrogen-hydrogen SSCC with medium sized AO basis sets show small and very systematic deviations from the CBS which can be corrected by simple scaling, i.e.,

$$J_{HH}^{CBS} = 1.07 J_{HH}^{pcJ-0+pol} \tag{10}$$

where J_{CBS} is the estimated basis set limit (as obtained from pcJ-2 calculations) and pcJ-0+pol is the standard pcJ-0 basis augmented with p (on hydrogen atoms) and d polarization functions (on C,N,O,F). The pcJ-0 basis is already relatively large and of about (unpolarized) triple-ζ character. A similar relation was found for the unpolarized basis, i.e.,

$$J_{HH}^{CBS} = 1.12 J_{HH}^{pcJ-0} \ . \tag{11}$$

The scaled pcJ-0+pol basis provides results which deviate on average by less than 0.1 Hz from the pcJ-2 result (maximum deviation around 0.5 Hz) for typical organic molecules. The deviations for the scaled pcJ-0 variant are slightly larger (maximum deviation around 0.7 Hz, average deviation of 0.2 Hz). The J^{HH} SSCC are dominated by the Fermi-contact (FC) term[42] and hence, we restrict the calculation to this term which results in a speed-up of a factor > 5 in the SSCC treatment.

3.2 Efficiency considerations

The double-hybrid DFT energy and hybrid DFT SSCC calculations employ the RI-JK approximation for the Coulomb and exchange integrals with matching auxiliary AO basis sets.[43,44] In the shielding constant calculations, the COSX approximation[45–47] for the exchange part is applied in addition to the RI treatment of the Coulomb term.[48] The double-hybrid DFT calculations employ RI-JK in the SCF as well as RI in the MP2 part.[49,50] In the PBEh-3c optimizations, RI was used for the Coulomb term.

For not too many conformations in the GFN-xTB start ensemble, the calculation time is determined by the SSCC part even if only the FC term is considered. Inclusion of spin-dipolar terms increases the computational cost for the SSCC by a factor of 2-3 while the computation of the diamagnetic (DSO) and paramagnetic (PSO) spin-orbit terms yields an additional overhead of a factor of 2-3.

4 Fragmentation and solution of the spin-Hamiltonian

The procedure outlined here is formulated with an emphasis on coupled ^{1}H-NMR spectra, but is general and can treat any set of magnetically active nuclei. The theory of the quantum mechanical computation of an NMR spectrum for given chemical shifts, SSCC, and spin I of the nuclei involved is well known.[51] For n spins with $I = \frac{1}{2}$ there are $N_S = 2^n$ product basis functions. If the input parameters J and δ are exact, the full treatment will yield the exact energy levels and dipole-transition intensities from which (by applying Lorentzian line-broadening to each transition) a practically exact spectrum can be computed.

The Hamilton matrix elements are relatively easy to compute, but already for small n the solution of the exponentially growing dimension of the special eigenvalue problem becomes intractable on normal workstation computers (e.g., for $n = 16$ there are $N_S = 65536$ basis functions). If magnetically equivalent nuclei are grouped together (e.g., the three hydrogen atoms of a methyl group form a so-called composite particle (CP) with effective $I = \frac{3}{2}$), the effort can be reduced significantly.[52] With full exploitation of spin-symmetry and employing the composite particle technique, the in-house written code for solution of the nuclear spin-problem (`anmr`[53]) can treat typical organic molecules with around 14-20 hydrogen atoms exactly (depending on the number of CPs). For larger spin-systems, the Hamiltonian matrix is separated into overlapping fragments, which are diagonalized individually (so-called divide-and-conquer approach). The largest system treated in this way by the present code had $n = 70$.

The fragmentation is based on the fact that the SSCC decay quickly with chemical distance between the coupling nuclei and, e.g., ^5J or higher couplings are rarely observed experimentally.[51] This is exploited in the calculation of SSCC in ORCA by a restriction to atom pairs with less than 8 Å distance, which decreases the computational effort significantly for large systems.

In the fragmentation algorithm, spins with maximum couplings are grouped together

and extended by neighboring spins with large SSCC. The fragmented (sub) spin-systems contain m 'real' spins and m' 'buffer' spins. The spins in the set m' can appear in several fragments and hence, transitions involving these must be removed in the intensity calculations such that no double counting occurs.

The fragments and buffer spins are generated as follows. The fragments should be created in a way, which keeps their mutual influence as small as possible. To determine the nuclei with the smallest interaction, the complete spin-system is treated as a graph, with the spins as nodes connected via their couplings and the inverse SSCC as the distances between the nodes. Dijkstra's algorithm is then used to compute the shortest path through this graph for each pair of spins. The pair with the longest of those shortest paths is chosen as the pair with the smallest mutual influence. Now, in the shortest path between those spins, the smallest coupling constant is set to zero. This is equivalent with disconnecting the graph at that particular position. Dijkstra's algorithm is used again to find the next shortest path in-between the spins. This procedure is iterated until there is no path left connecting the spins, resulting in two decoupled systems. The overall algorithm is repeated, until the size of all fragments is below the threshold. This entire procedure is repeated for different numbers of real and buffer spins (for the same given number of total spins, code flag $-$mss) in the fragments until a minimum value for the largest neglected coupling constant (typically <0.2-0.5 Hz) is found. The included buffer spins are always selected according to the magnitude of the SSCC. The default maximum number of spins including buffer spins in one fragment is 12-14 depending on the available memory of the computer.

As examples for the fragmentation of large systems we show below vitamin D2 with 73 atoms containing 44 hydrogens as well as adrenosterone (46 atoms, 24 protons) in Figure 7 and 8

The splitting into fragments is chemically reasonable and reflects the spin-spin coupling between protons.

The convergence of the fragmentation procedure with the maximum size of the fragment
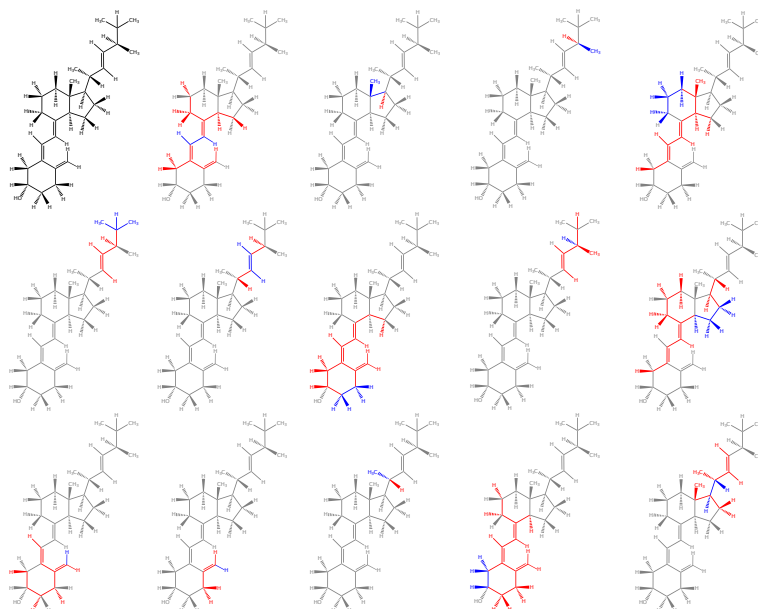
Figure 7: Fragmentation of vitamin D2 into overlapping spin-systems. The full structures are shown in black or gray while colored atoms indicate fragments (blue: 'real' spins, red: 'buffer' spins)
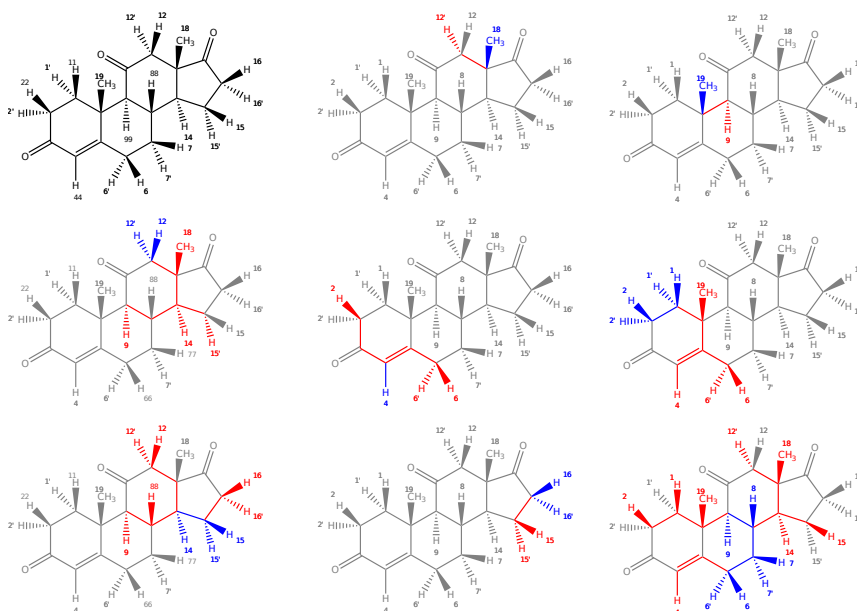

Figure 8: Fragmentation of adrenosterone into overlapping spin-systems. The full structures are shown in black or gray while colored atoms indicate fragments (blue: 'real' spins, red: 'buffer' spins)

spin-system is shown for the example of adrenosterone (500 MHz, CDCl$_3$) in Figure 9 and for the all-trans-polyene C$_{12}$H$_{14}$ in Figure 10 (400 MHz, CDCl$_3$).
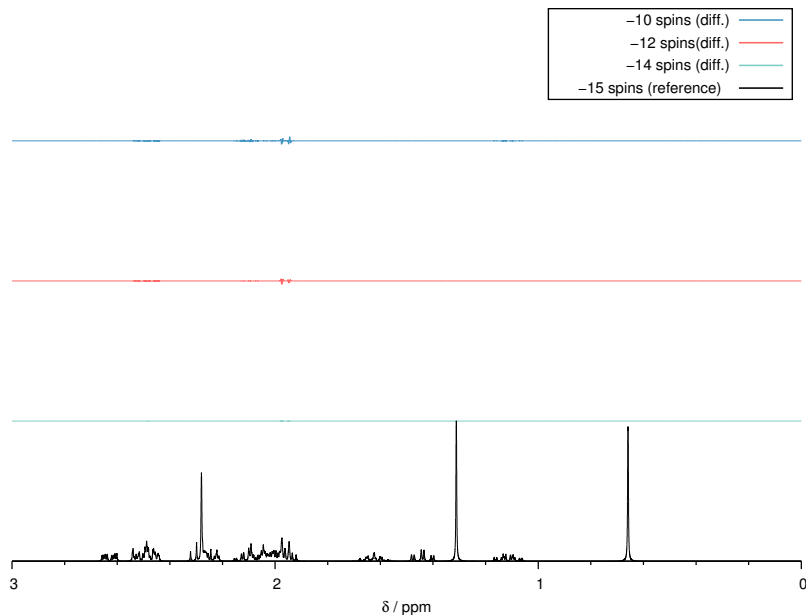

Figure 9: Simulated ^{1}H-NMR spectra for adrenosterone with different maximum fragment system sizes. The largest performed calculation with 15 spins is taken as reference and the differences for smaller sizes are shown.

As can be seen for this non-trivial example, the fragmentation approach works well yielding only very small residual differences for a limited number of 12–14 spins in the fragments. Similar observations have been made for other systems as well. The differences are in practice even smaller than suggested by this figure because tiny frequency or intensity errors are amplified by the small line width of the signals when taking difference spectra. As a worst-case scenario for a very strongly coupled unsaturated system with many significant ^{4}J we show the errors of the fragmentation procedure for the polyene example where larger but still tolerable deviations to the exact solution are observed.
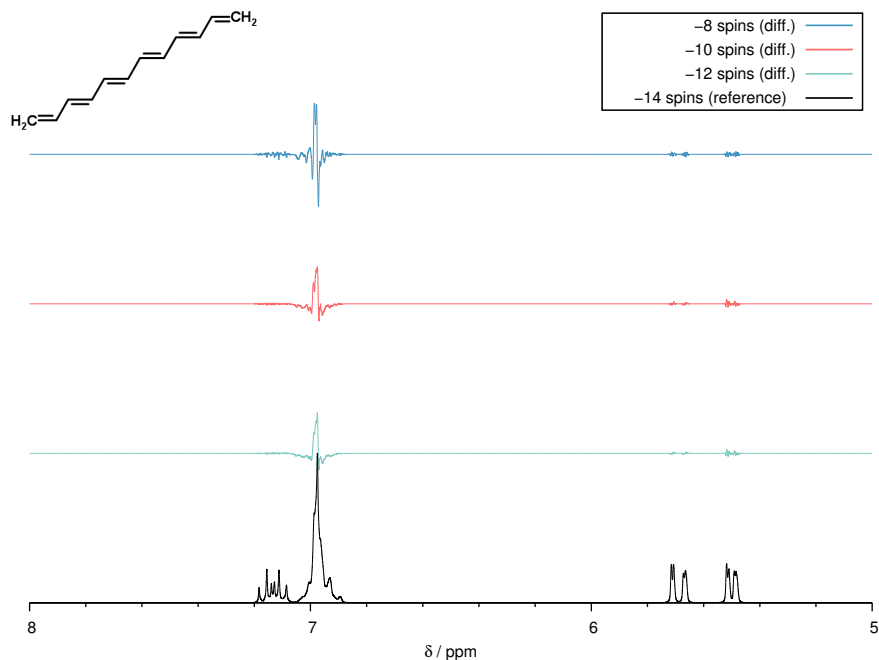
Figure 10: Simulated ^{1}H-NMR spectra for (3E,5E,7E,9E)-1,3,5,7,9,11-dodecahexaene with different maximum fragment system sizes. The largest performed calculation with 14 spins (full system) is taken as reference and the differences for smaller sizes are shown.

5 Additional example ^{1}H-NMR spectra

5.1 Discrimination of diastereomers in adrenosterone

As an example for a diastereomer assignment problem we present in Figure 11 computed spectra of four adrenosterone diastereomers, differ in the configuration of the two methyl groups attached to C10 and C13. It is demonstrated that only the computed spectrum of the correct adrenosterone diastereomer with (10R,13S) configuration (bottom, in black) agrees well with the experimental spectrum (inverted, in gray) in terms of chemical shifts and details of the multiplets.

5.2 1-Chlor-butane

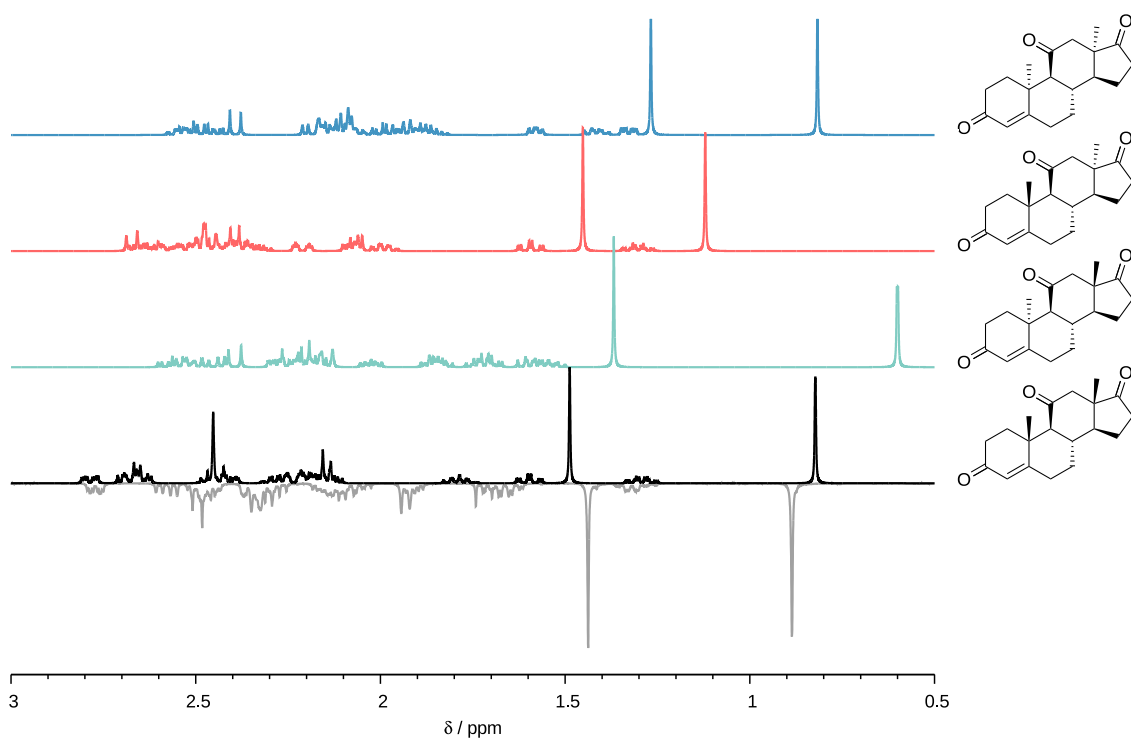See Figure 12 for a comparison of experimental and computed spectra.

Figure 11: Comparison of experimental and computed ^{1}H-NMR spectra for adrenosterone (500 MHz, CDCl$_3$). The ^{1}H-NMR spectra are also computed for three diastereomers with different configuration at the C10 and C13 centers.
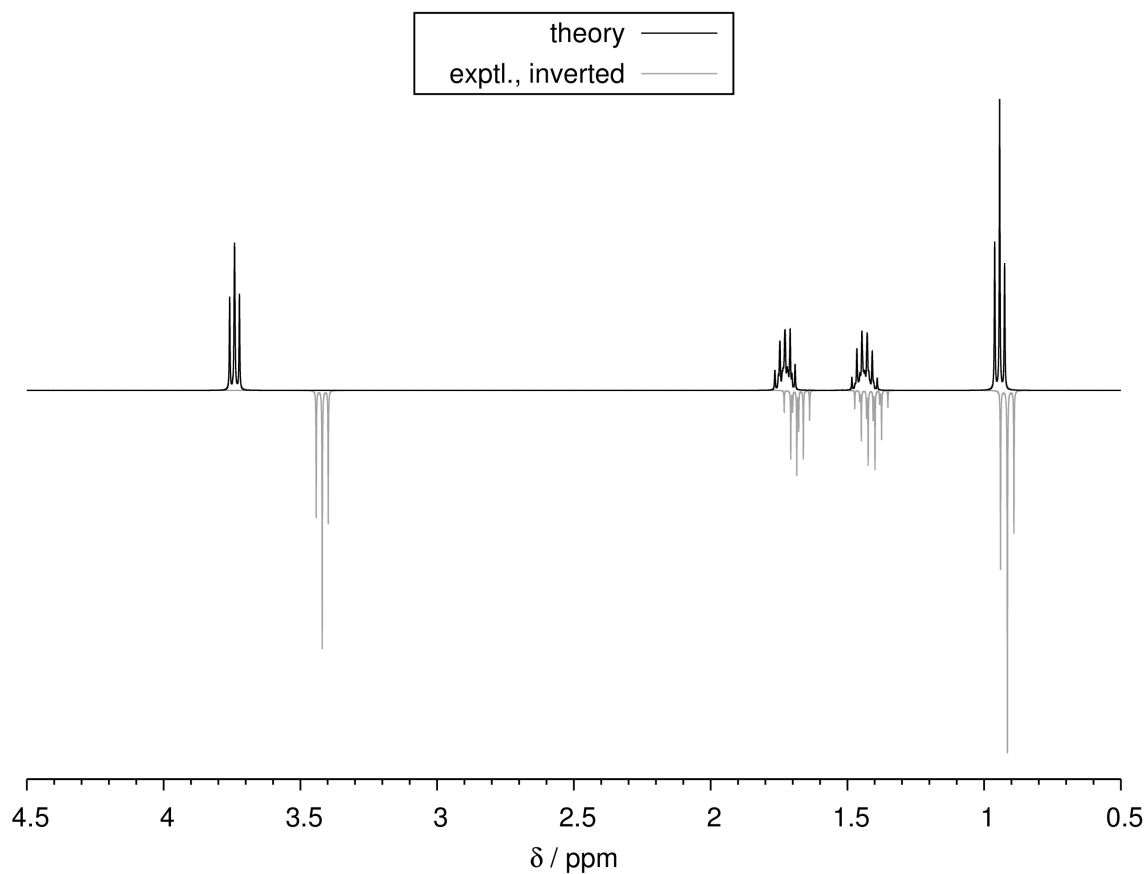
Figure 12: Comparison of experimental (derived from peak list) and computed ^{1}H-NMR spectra for 1-chlor-butane (400 MHz, CDCl$_3$). The theoretical spectrum refers to four conformers with population between 5 and 32%.

5.3 4-Phenylmorpholine

See Figure 13 for a comparison of experimental and computed spectra.

5.4 Methylcyclohexane

See Figure 14 for a comparison of experimental and computed spectra including plots of the contributing axial and equatorial conformers.

5.5 Strychnine

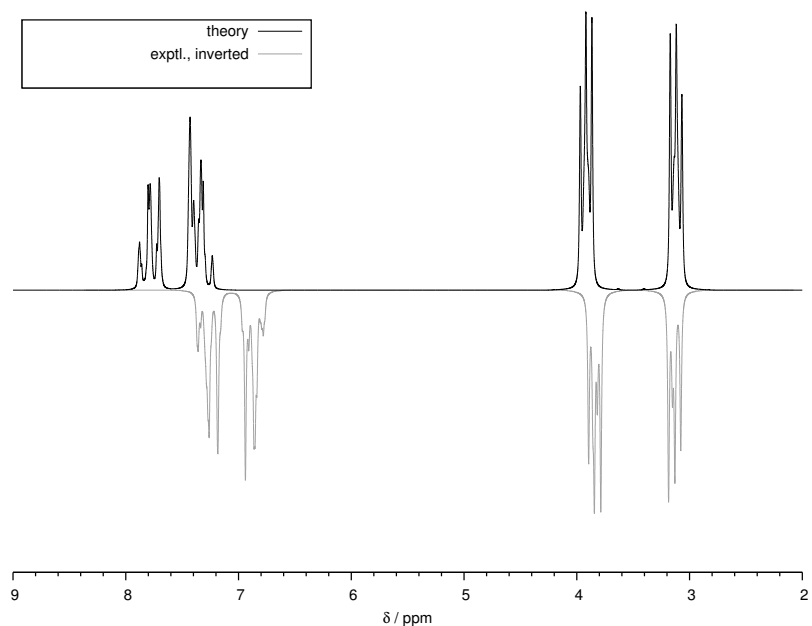See Figure 15 for a comparison of experimental and computed spectra.

Figure 13: Comparison of experimental (derived from peak list) and computed ^{1}H-NMR spectra for 4-phenylmorpholine (90 MHz, CDCl$_3$). Only a single conformer is considered.

5.6 α-Ionone

See Figure 16 for a comparison of experimental and computed spectra.

5.7   Guaiol

See Figure 17 for a comparison of experimental and computed spectra.

5.8   D-sucrose

As a very challenging case we briefly discuss the spectrum of D-sucrose shown in Figure 18. Here, the huge conformational space of sugars in general[54] as well as the competition between intramolecular and intermolecular hydrogen bonding makes it a worst case scenario.

For D-sucrose the GFN-xTB and DFT conformational energy surfaces are rather 'non-parallel' so that more conformations have to be re-evaluated at the PBEh-3c level. The conformational search was started from the X-ray derived conformation. The energy win-
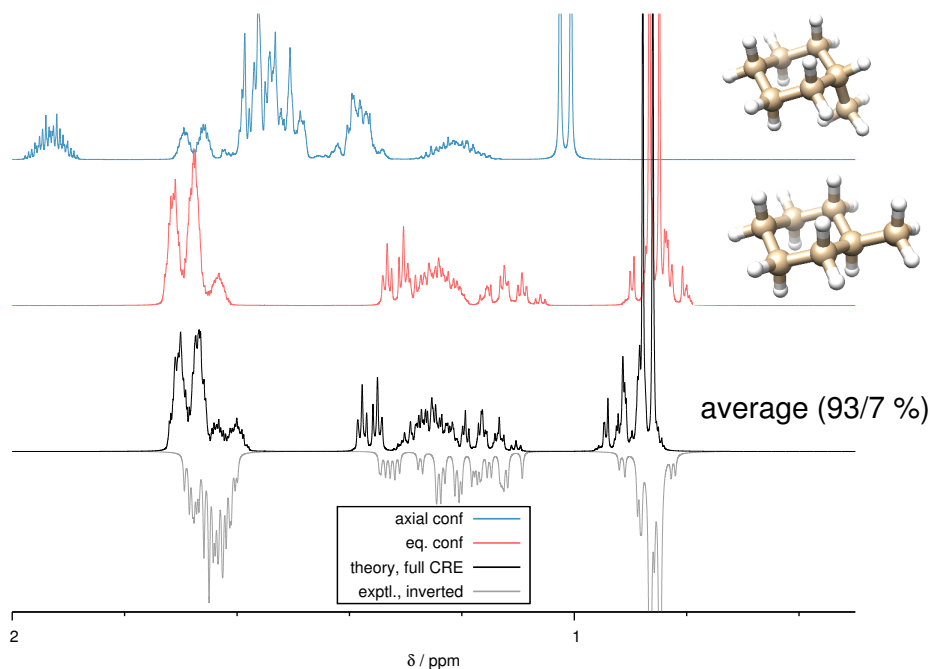
Figure 14: Comparison of experimental (derived from peak list) and computed ^{1}H-NMR spectra for methylcyclohexane (400 MHz, CDCl$_3$). The theoretical spectrum in black refers to the parameter average of axial and equatorial conformers with population of 93 and 7%, respectively.

dow was increased in this case from 6 to 10 kcal mol^{-1} and 458 conformations were considered further at the DFT level. From these the 31 lowest in a 3 kcal mol^{-1} window at the PBEh-3c[DCOSMO-RS] level were selected and treated by DSD-BLYP[COSMO-RS]. The final ensemble consisting of nine structures with populations >4% is shown below.

As can be seen, two structures are dominating which is in agreement with the results of a force-field MD study in explicit water.[55] However, in details the preferred conformation in our and in the MD study differ. Our lowest free energy conformer in water shows strong resemblence to the conformation found in the cystal.[56] In general, the dominantly contributing conformers have at least one (the preferred one even two) inter-ring hydrogen bonds.

The simulated NMR spectrum shown in Figure 18 differs in one main aspect from the experimental one. The two CH$_2$ protons on C7 of the pyranose ring are chemically in-
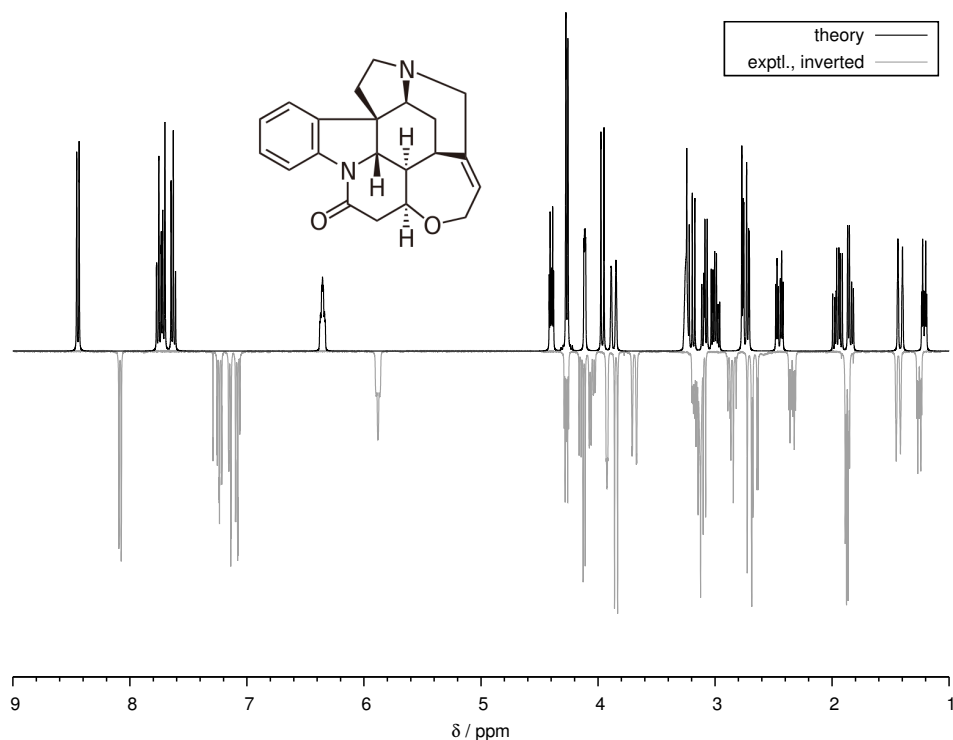
Figure 15: Comparison of experimental and computed ^{1}H-NMR spectra for strychnine (400 MHz, CDCl$_3$). The theoretical spectrum refers to two conformers with population of 96 and 4 %, respectively.

equivalent in all considered conformations and also on average their chemical shifts differ by about 0.1 ppm. As a result two doublets with a geminal splitting of about 13 Hz are computed. This is expected because the neighboring carbon atom is chiral and hence the environment for the two protons must be different in all conformers. Intriguingly, however, various experimental spectra show clearly a singlet without any splitting or broadening for these protons. This finding can be explained by accidentally the same chemical shifts (to less than about 0.002 ppm) for the two protons in the CH$_2$ group[57] or by an unknwon mechanism which interchanges them. As expected, the corresponding shifts in all conformations entering our simulation are rather different and hence, we compute more signals than observed. In passing we note that replacement of the OH group in the respective CH$_2$OH unit by Cl leads to the appearance of inequivalent (doublet) signals.[58] Further studies including other polysacharides may reveal more understanding of these interesting
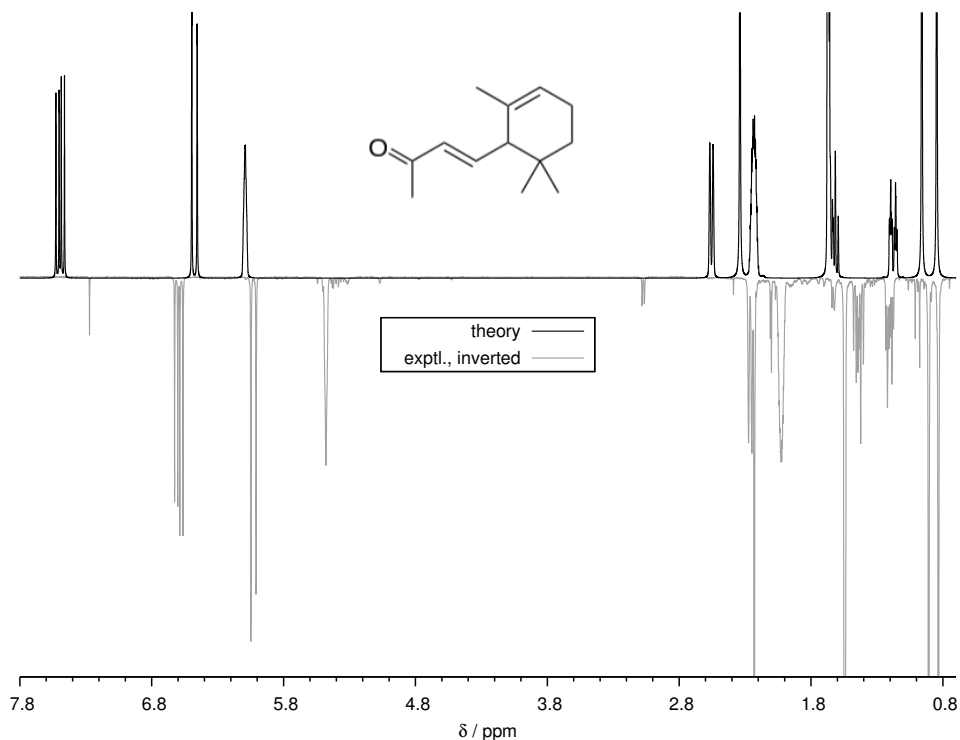
Figure 16: Comparison of experimental and computed ^{1}H-NMR spectra for α-ionone (400 MHz, CDCl$_3$). The theoretical spectrum refers to four conformers with population between 10 and 43%.

observations which remain unresolved at this point.

6   Drug molecule conformer benchmark

The performance of the CRE generation algorithm was tested on a conformer benchmark set consisting of 100 medium sized, mainly organic molecules (22 to 73 atoms), which is shown in Figures 20a to 20f. For comparison with other conformer generating algorithms the energy was chosen to be the most reliable criterion. It is questionable, whether the widely used RMSD (root-mean-square deviation of atomic positions) comparison[59] of generated conformers to experimental crystal structure geometries is sufficient, in particular if conformers of molecules in solution or gas phase are of interest. The CPU time was not a benchmarking criterion either, because our algorithm primarily aims for high accuracy
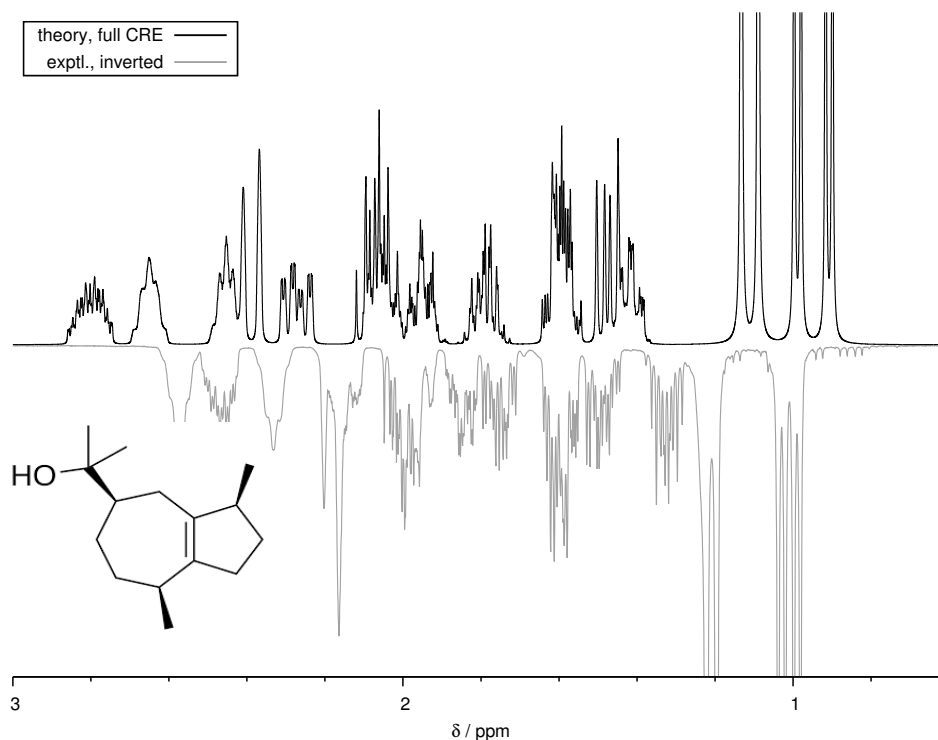
Figure 17: Comparison of experimental and computed ^{1}H-NMR spectra for guaiol (400 MHz, CDCl$_3$). The theoretical spectrum refers to six conformers with population between 7 and 21%.

(rather than, e.g., for the speed of conformer generation), since the need for an almost complete CRE, including the correct ranking, is a fundamental prerequisite to simulate NMR in solution. The performance evaluation is based on the electronic energy of the most stable conformer that was generated compared to a reference. This reference was the most stable conformer generated by another algorithm, called '*best*' algorithm, which is used in industry.[60] Conformers were generated using the same input structure for both algorithms, and the most stable conformers were then determined at the PBE0-D3/def2-TZVP//GFN-xTB[1,7,28,33,61,62] level of theory. The technical details of our algorithm were tuned to generate the best possible count of most stable conformers that are either equally good or better compared to the most stable reference conformer found by the '*best*' algorithm. Here, 'equally good' means that the PBE0-D3(BJ)/def2-TZVP//GFN-xTB energy difference ΔE of the compared conformers is within ± 0.2 kcal mol^{-1}, while 'better' is de-
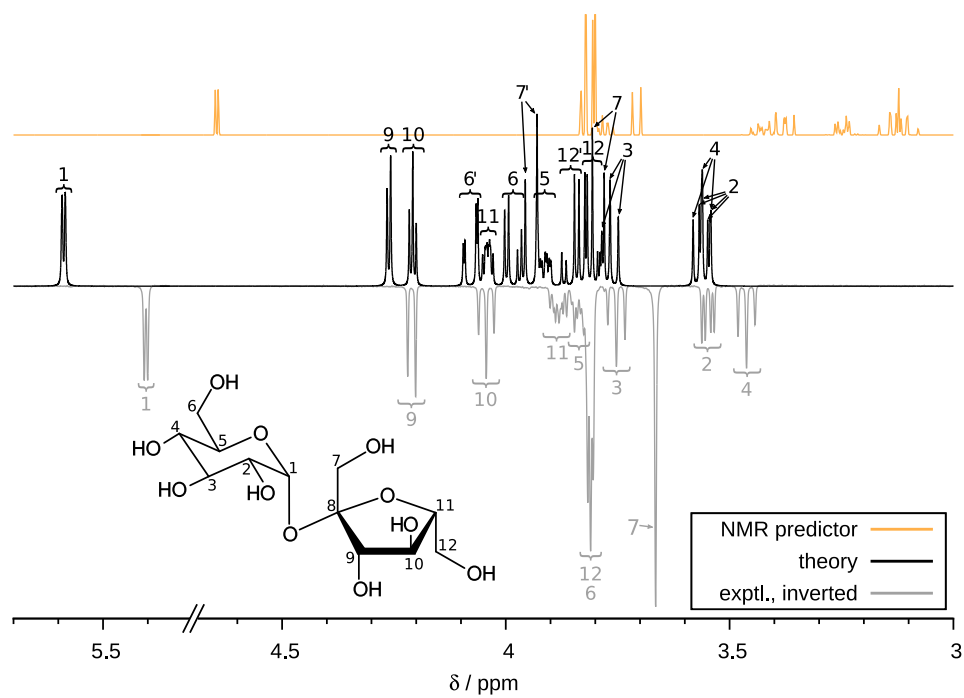
Figure 18: Experimental and calculated ^{1}H-NMR spectra of D-sucrose (500 MHz, in D_2O) with peak assingment. The conformer ensemble consists of nine species with populations in the range 5 to 33 %. The default GFN-xTB conformational energy selection window was increased in this case from 6 to 10 kcal mol^{-1} in order to include all relevant conformers.
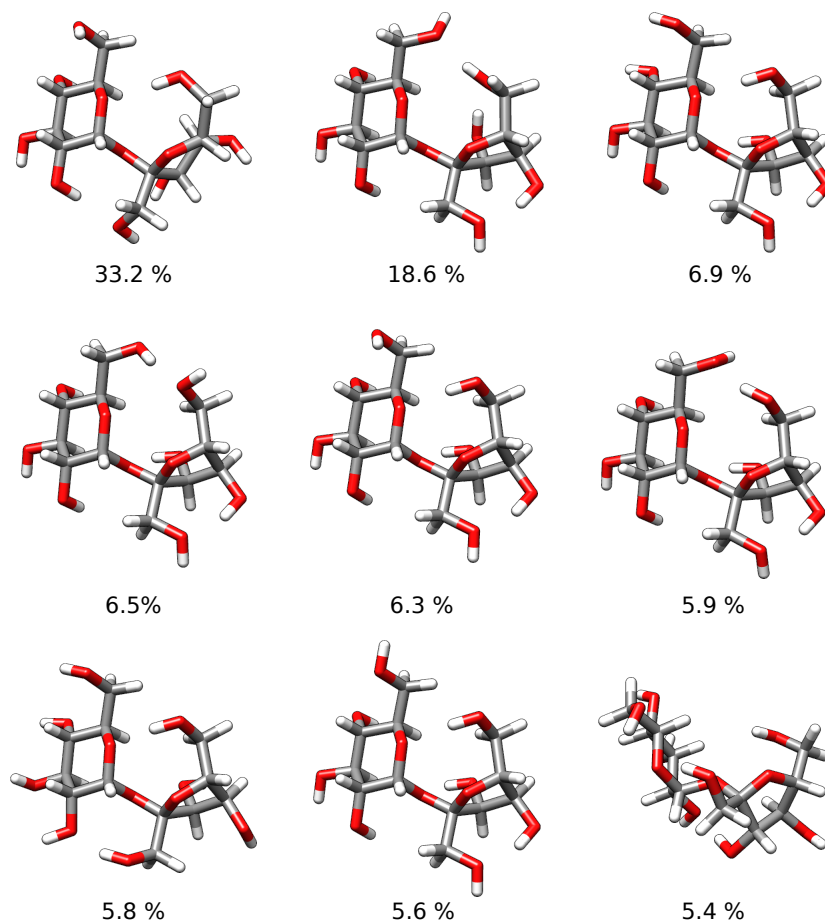
Figure 19: DFT optimized (PBEh-3c[DCOSMO-RS]) structures of the dominant conformers in the D-sucrose ensemble in water. The populations at T=298 K derived from the standard theory level (DSD-BLYP[COSMO-RS]//PBEh-3c[DCOSMO-RS]) are also given.

fined by ΔE being less than that ($\Delta E < -0.2$ kcal mol^{-1}). RMSD comparisons between the most stable conformers of the two algorithms were not conducted systematically, since already a comparison of two rotamers can lead to misleading high RMSD values. The count of molecules for the benchmark set, in which the most stable conformer is equally good or better compared to the reference is shown in Table 3. All calculations were conducted in the gas phase.

Table 3: Number of molecules in the benchmark set (100 systems in total) in which the most stable conformer found with the MF-MD-GC algorithm is equally good or better than the reference. Data are shown for different versions of the algorithm, the lowermost being the final version used. MF $\hat{=}$ normal mode following with canonical modes, LMF $\hat{=}$ local mode following, MD $\hat{=}$ molecular dynamics simulation.

MF-MD-GC algorithm details	equal	better
3x MF + 1x LMF + 40 ps MD(GFN-xTB)	39	21
3x MF + 2x LMF + 40 ps MD(GFN-xTB)	49	28
3x MF + 2x LMF + 80 ps MD(GFN-xTB)	50	32
3x MF + 2x LMF + 0.5 ns MD(QMDFF[27])	56	29
3x MF + 1 ns MD(QMDFF)	55	27
3x MF + 2x LMF + 3x 0.5 ns MD(QMDFF)	54	34

As can be seen from Table 3, the results can be improved by tuning different parts of the composite algorithm. For example, the usage of a larger number of local modes was proven to be important for the algorithm. Significant improvements are achieved by prolonging the molecular dynamics simulation, which is also an essential step for the generation of rotamers. A longer MD simulation can partially compensate for truncating the local mode following steps of the algorithm, however, doing so can negatively affect cases in which more challenging geometry perturbation of the molecule is required, e.g., for macrocyclic compounds.

Results for the final version of the algorithm (electronic energies and ΔE of the lowest conformer found by MF-MD-GC compared to the 'best' algorithm) are shown in Table 4.
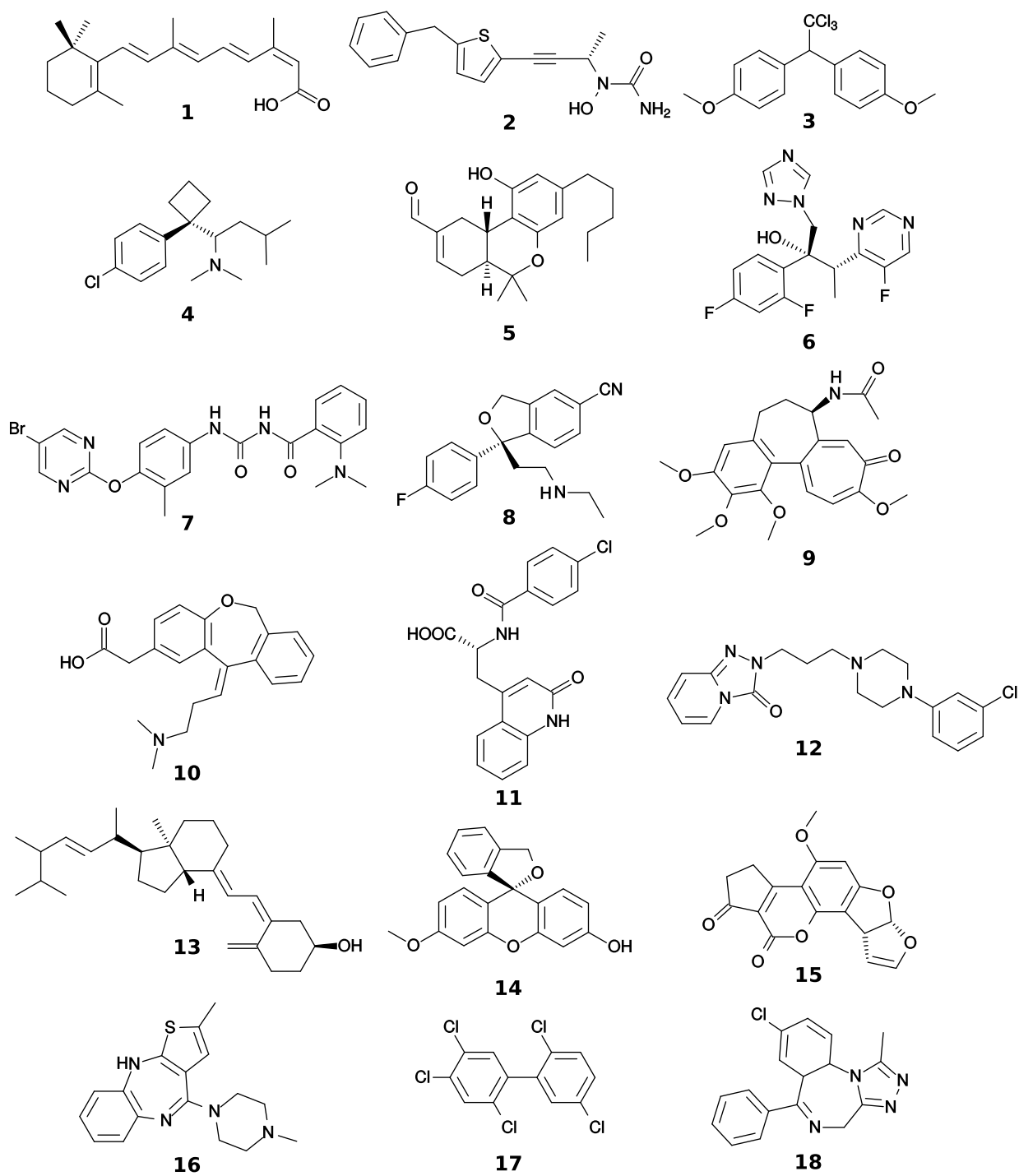
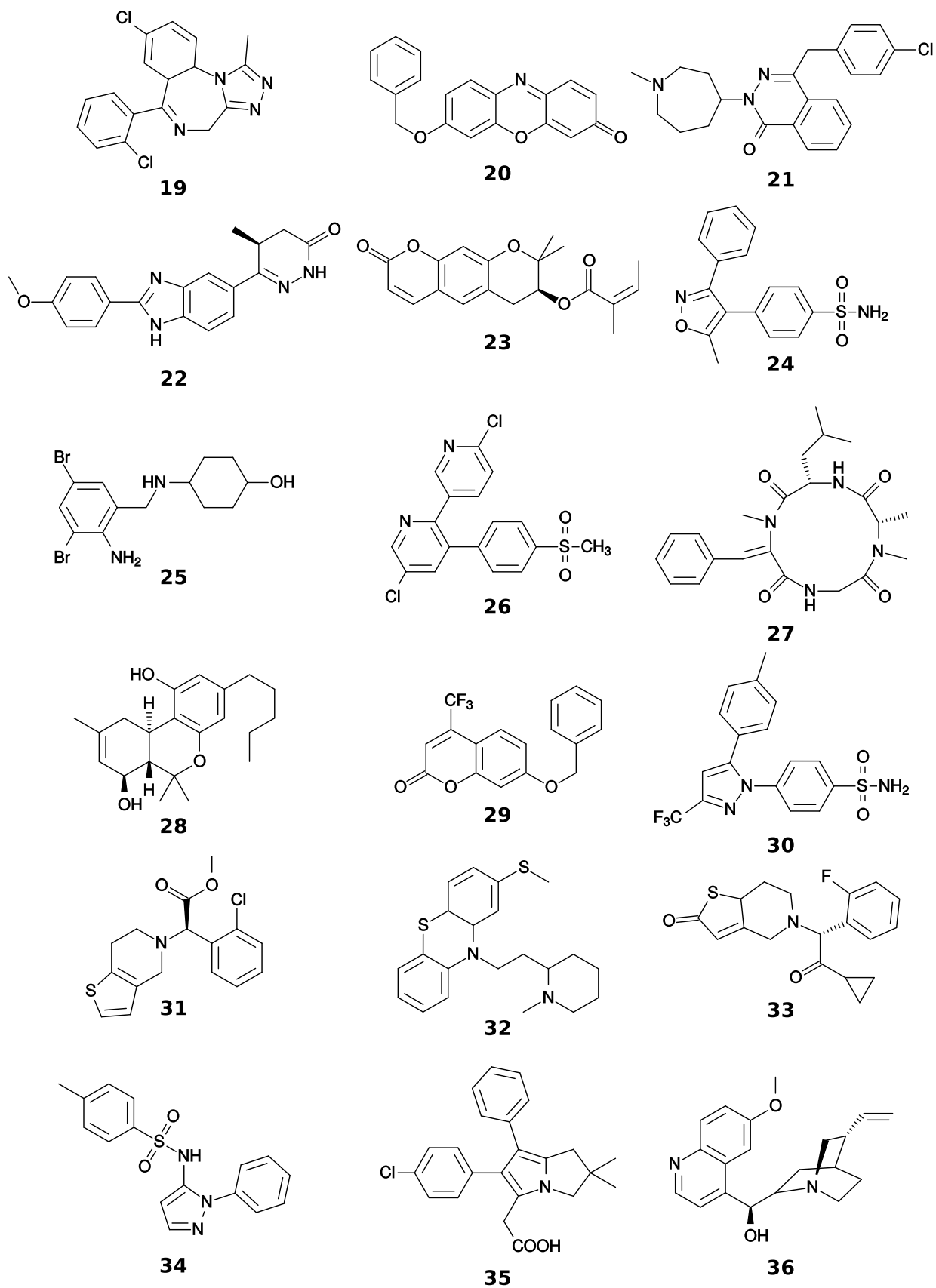Figure 20a: Molecules 1 to 18 of the benchmark set.

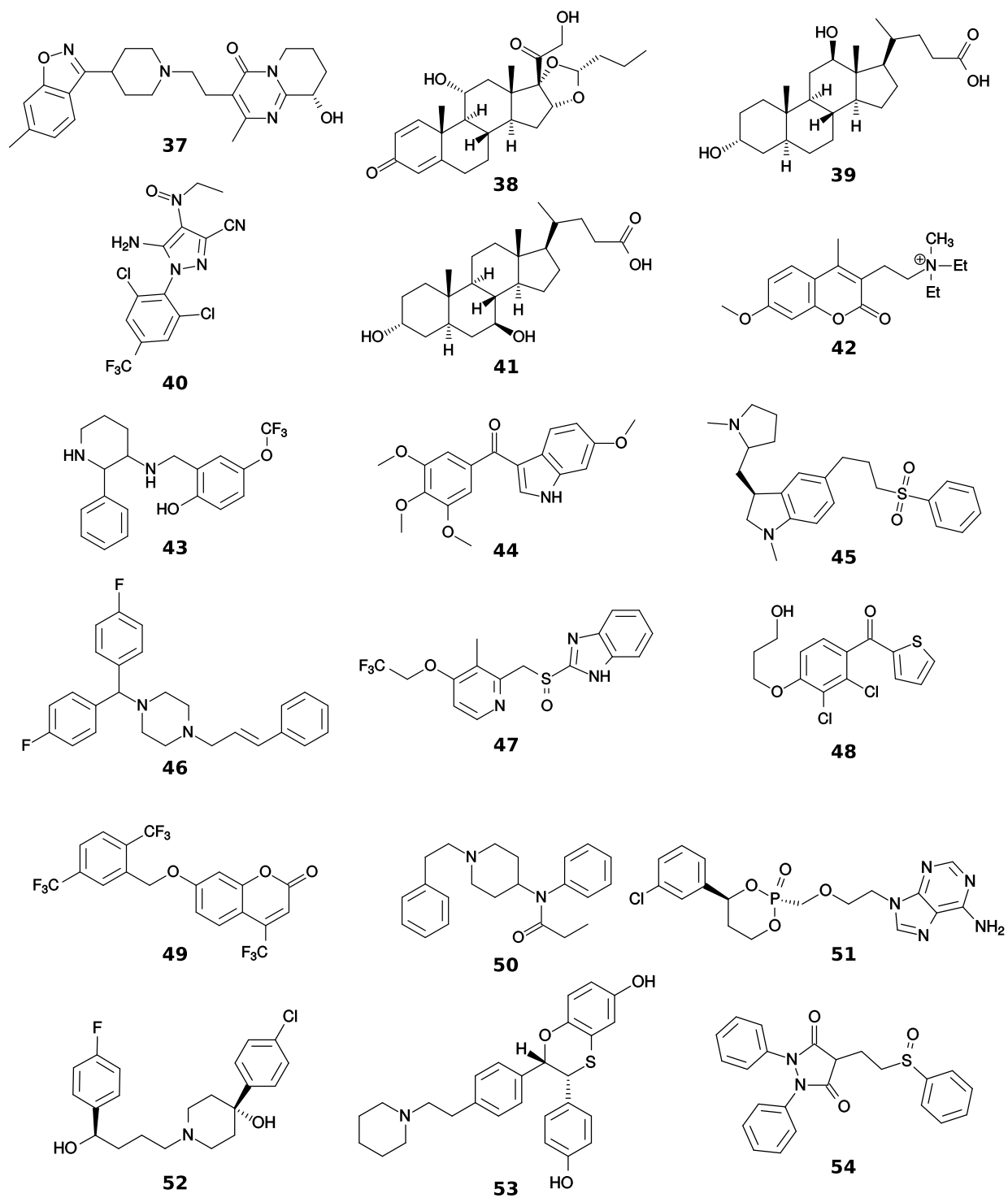Figure 20b: Molecules 19 to 36 of the benchmark set.
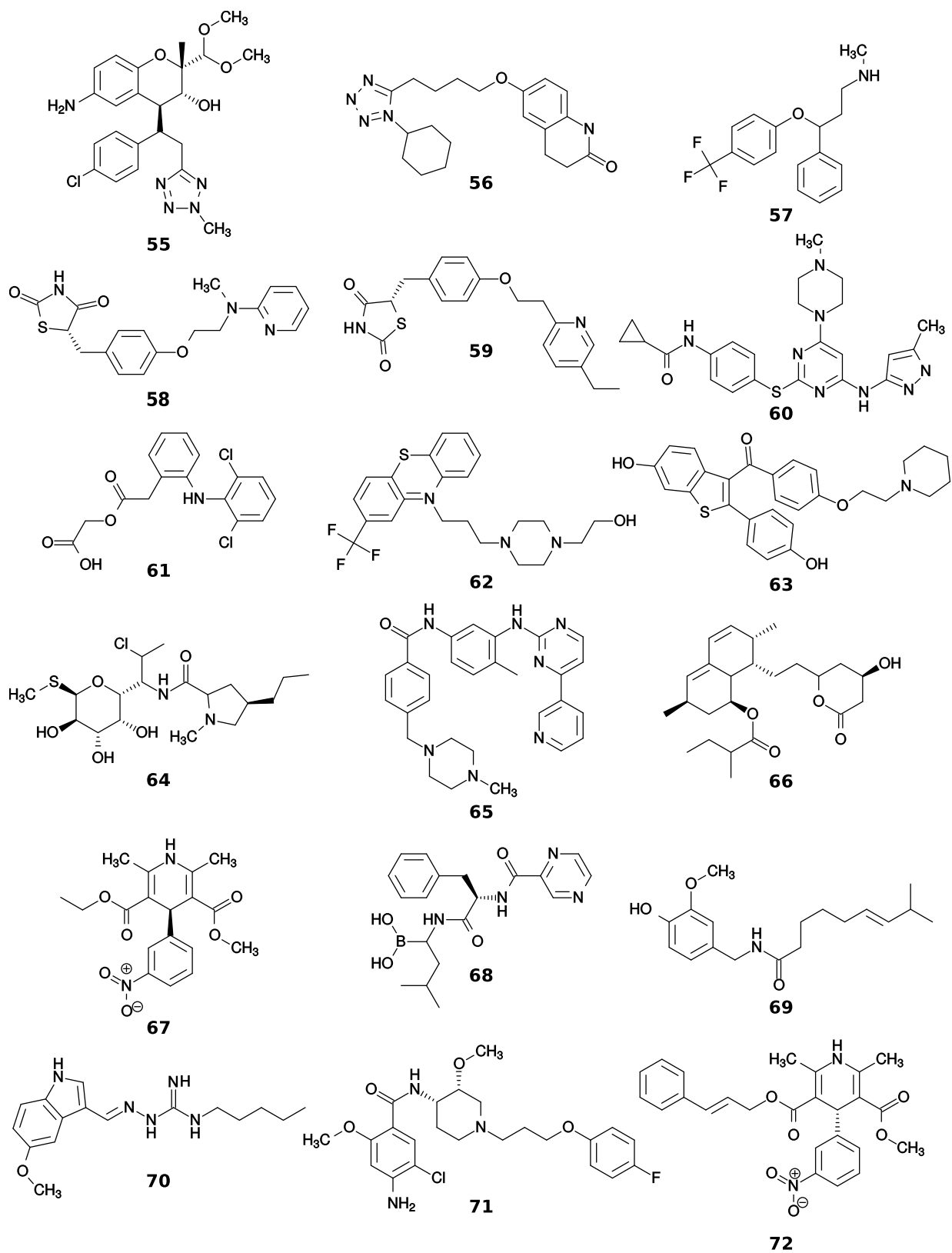
Figure 20c: Molecules 37 to 54 of the benchmark set.

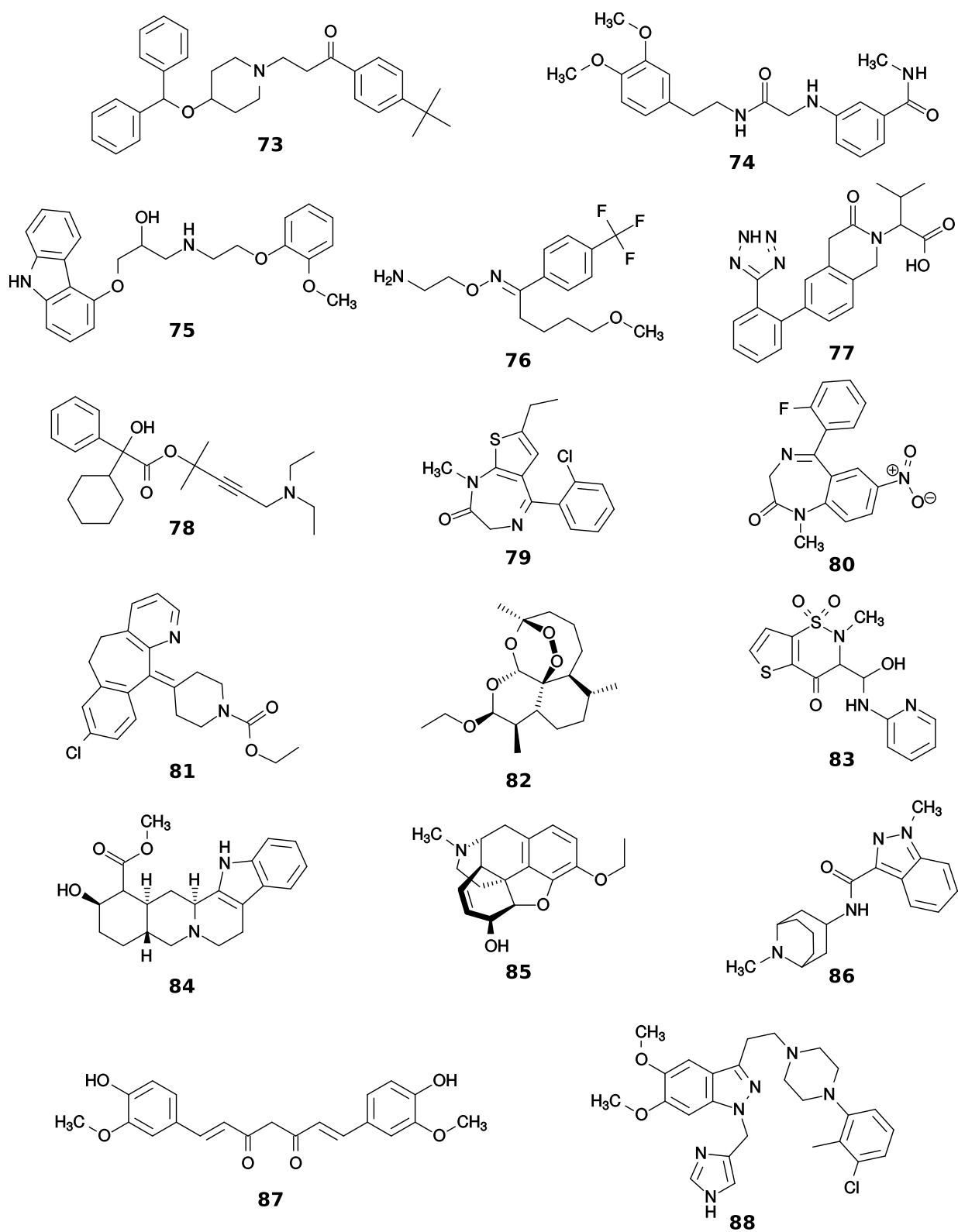Figure 20d: Molecules 55 to 72 of the benchmark set.

Figure 20e: Molecules 73 to 88 of the benchmark set.
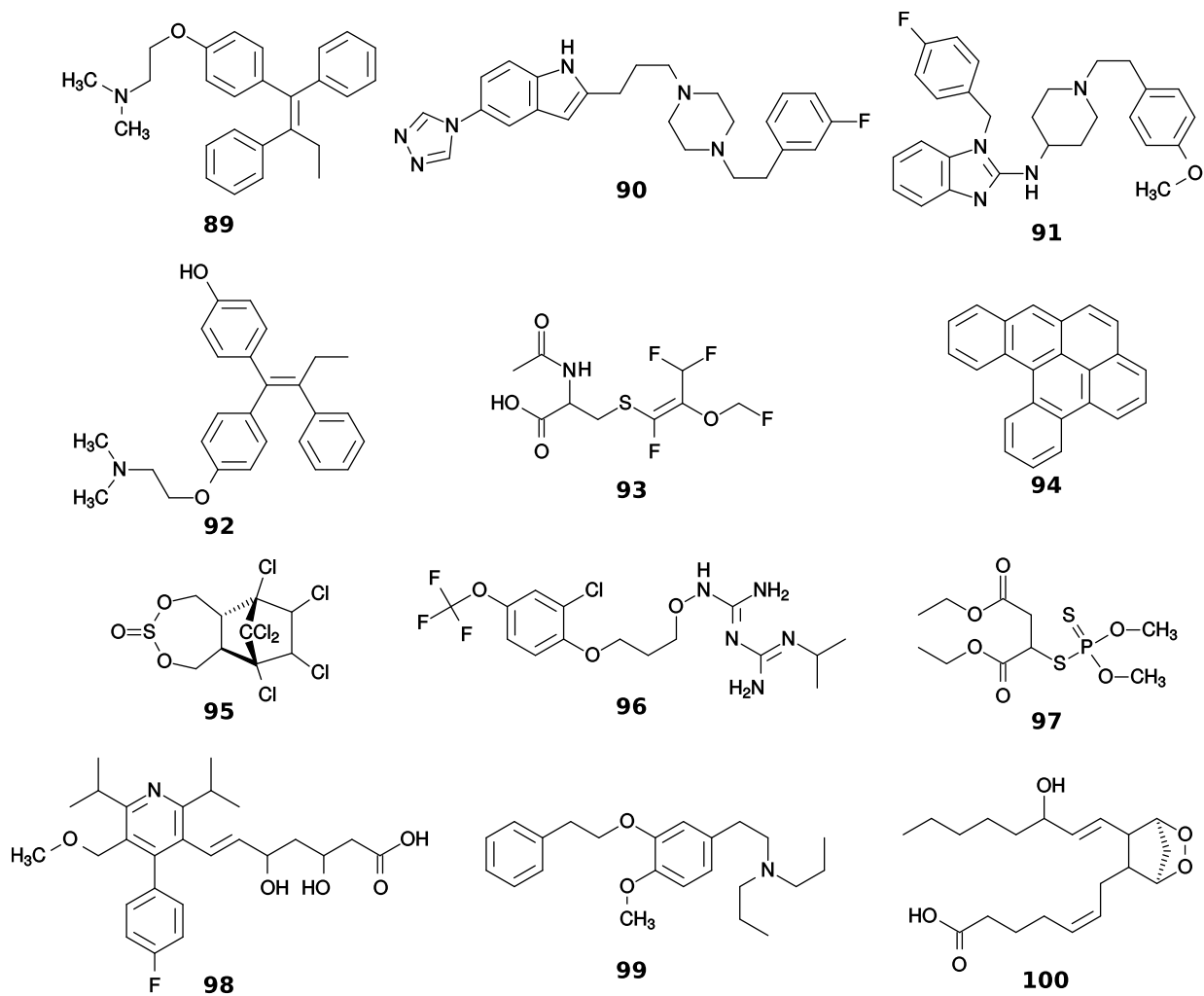
Figure 20f: Molecules 89 to 100 of the benchmark set.

Table 4: Electronic energies for the most stable conformers of the benchmark molecules, generated with the '*best*' and MF-MD-GC conformation search algortihms. The energies are calculated at the PBE0-D3(BJ)/def2-TZVP//GFN-xTB level of theory. Negative ΔE values indicate that the structure found by MF-MD-GC is lower than with the '*best*' algorithm.

| Molecule | PBE0-D3(BJ)/def2-TZVP//GFN-xTB | | ΔE [kcal mol^{-1}] |
	'*best*' algorithm	MF-MD-GC algorithm	
1	-928.6952962	-928.6952314	0.04
2	-1276.6177896	-1276.6164262	0.86
3	-2148.7682227	-2148.7681578	0.04
4	-1177.6771440	-1177.6769288	0.14
5	-1041.9960984	-1041.9965435	-0.28
6	-1265.7623173	-1265.7633892	-0.67
7	-3884.6387154	-3884.6568291	-11.37
8	-1019.2507769	-1019.2508813	-0.07
9	-1358.4416003	-1358.4405824	0.64
10	-1093.6913322	-1093.6846518	4.19
11	-1602.7823113	-1602.7822981	0.01
12	-1545.5033180	-1545.5037959	-0.30
13	-1167.7427526	-1167.7415529	0.75
14	-1183.9522546	-1183.9521845	0.04
15	-1105.5880452	-1105.5881488	-0.07
16	-1276.2416182	-1276.2422996	-0.43
17	-2760.2664319	-2760.2664856	-0.03
18	-1334.0279515	-1334.0279323	0.01
19	-1793.4947621	-1793.4947082	0.03
20	-1011.4957264	-1011.4956578	0.04
21	-1551.5124875	-1551.5124528	0.02
22	-1103.5128558	-1103.5127377	0.07

23	-1111.4054505	-1111.4053554	0.06
24	-1350.6462735	-1350.6461820	0.06
25	-5838.1716128	-5838.1717680	-0.10
26	-1812.3228112	-1812.3227543	0.04
27	-1375.3492212	-1375.3524960	-2.05
28	-1043.1995776	-1043.2015997	-1.27
29	-1178.8581898	-1178.8581672	0.01
30	-1667.6866435	-1667.6865294	0.07
31	-1682.0552677	-1682.0547838	0.30
32	-1720.8199303	-1720.8182105	1.08
33	-1399.1090195	-1399.1090085	0.01
34	-1330.7959023	-1330.7956886	0.13
35	-1554.1841147	-1554.1840435	0.04
36	-1035.7147198	-1035.7148711	-0.09
37	-1436.1990413	-1436.1964911	1.60
38	-1423.4332782	-1423.4354490	-1.36
39	-1238.6788981	-1238.6812282	-1.46
40	-2411.9935657	-2411.9931444	0.26
41	-1238.6772210	-1238.6772004	0.01
42	-981.0549289	-981.0550760	-0.09
43	-1295.2839616	-1295.2861223	-1.36
44	-1165.4450633	-1165.4448757	0.12
45	-1512.3755359	-1512.3766378	-0.69
46	-1314.6064572	-1314.6065273	-0.04
47	-1629.6048656	-1629.6046926	0.11
48	-2084.0071149	-2084.0070413	0.05
49	-1852.6193676	-1852.6192701	0.06
50	-1039.1437804	-1039.1435424	0.15
51	-2034.4504806	-2034.4509438	-0.29

52	-1579.5764826	-1579.5869451	-6.57
53	-1799.0093580	-1799.0102587	-0.57
54	-1620.8037681	-1620.8035636	0.13
55	-1942.9368769	-1942.9388001	-1.21
56	-1201.7155148	-1201.7164349	-0.58
57	-1087.4080668	-1087.4080321	0.02
58	-1484.6034125	-1484.6034508	-0.02
59	-1468.5737105	-1468.5747826	-0.67
60	-1803.5733106	-1803.5721867	0.71
61	-1892.6536004	-1892.6536074	0.00
62	-1790.1040146	-1790.1110067	-4.39
63	-1835.9150574	-1835.9152815	-0.14
64	-2048.7129056	-2048.7250076	-7.59
65	-1581.0804845	-1581.0810440	-0.35
66	-1311.4200104	-1311.4192252	0.49
67	-1257.8579464	-1257.8578967	0.03
68	-1283.0617319	-1283.0628408	-0.70
69	-981.8691870	-981.8691763	0.01
70	-971.7780533	-971.7789838	-0.58
71	-1917.9225103	-1917.9226336	-0.08
72	-1526.7830699	-1526.7828962	0.11
73	-1407.6980286	-1407.7053074	-4.57
74	-1241.4515722	-1241.4520382	-0.29
75	-1339.6123519	-1339.6175397	-3.26
76	-1142.8875850	-1142.8886322	-0.66
77	-1430.4171681	-1430.4178849	-0.45
78	-1215.0915664	-1215.0933817	-1.14
79	-1660.9955824	-1660.9957069	-0.08
80	-1105.8341925	-1105.8340155	0.11

81	-1571.4019709	-1571.4013990	0.36
82	-1079.2069959	-1079.2072813	-0.18
83	-1762.6261284	-1762.6261203	0.01
84	-1150.2594936	-1150.2594243	0.04
85	-1017.5123185	-1017.5123023	0.01
86	-993.8709773	-993.8710609	-0.05
87	-1262.6685850	-1262.6686386	-0.03
88	-1947.3154132	-1947.3191369	-2.34
89	-1137.2857454	-1137.2858346	-0.06
90	-1397.4182715	-1397.4202791	-1.26
91	-1478.6361814	-1478.6392649	-1.93
92	-1212.4686987	-1212.4688143	-0.07
93	-1501.9880478	-1501.9884064	-0.22
94	-922.3229777	-922.3230643	-0.05
95	-3730.2561335	-3730.2558257	0.19
96	-1842.3107313	-1842.3129615	-1.40
97	-1980.7543866	-1980.7542159	0.11
98	-1540.8509762	-1540.8546314	-2.29
99	-1100.6467244	-1100.6442567	1.55
100	-1156.5708559	-1156.5708856	-0.02

References

(1) S. Grimme, C. Bannwarth, P. Shushkov *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(2) S. Grimme. xtb, a tight-binding quantum chemistry code for large molecules, University of Bonn, 2017. The program and additional codes for the conformational search and the computation of the NMR spectrum (program anmr) can be obtained for

academic use upon request at xtb@thch.uni-bonn.de.

(3) F. Weigend, R. Ahlrichs *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(4) S. Grimme, J. G. Brandenburg, C. Bannwarth, A. Hansen *J. Chem. Phys.* **2015**, *143*, 054107.

(5) S. Kozuch, D. Gruzman, J. M. L. Martin *J. Phys. Chem. C* **2010**, *114*, 20801–20808.

(6) S. Kozuch, J. M. L. Martin *J. Comput. Chem.* **2013**, *34*, 2327–2344.

(7) C. Adamo, V. Barone *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(8) F. Jensen *J. Chem. Theory Comput.* **2008**, *4*, 719.

(9) F. Jensen *Theor. Chem. Acc.* **2010**, *126*, 371.

(10) W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(11) F. Eckert, A. Klamt *AiChE Journal* **2002**, *48*, 369.

(12) A. Klamt *WIREs Comput. Mol. Sci.* **2011**, *1*, 699–709.

(13) S. Sinnecker, A. Rajendran, A. Klamt, M. Diedenhofen, F. Neese *J. Phys. Chem. A* **2006**, *110*, 2235–2245.

(14) V. Barone, M. Cossi *J. Phys. Chem. A* **1998**, *102*, 1995.

(15) A. V. Marenich, C. J. Cramer, D. G. Truhlar *J. Phys. Chem. B* **2009**, *113*, 6378.

(16) IUPAC Gold Book, see http://goldbook.iupac.org/C/C01262.html.

(17) IUPAC Gold Book, see http://goldbook.iupac.org/R/R05407.html.

(18) S. Grimme *Chem. Eur. J.* **2012**, *18*, 9955–9964.

(19) L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme *Phys. Chem. Chem. Phys.* **2017**, submitted.

(20) L. Goerigk, S. Grimme *WIREs Comput. Mol. Sci.* **2014**, *4*, 576–600.

(21) Y. Zhao, D. G. Truhlar *J. Phys. Chem. A* **2005**, *109*, 5656–5667.

(22) I. Kolossváry, W. C. Guida *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.

(23) T. Kamachi, K. Yoshizawa *J. Chem. Inf. Model.* **2016**, *56*, 347–353.

(24) J. Pipek, P. G. Mezey *J. Chem. Phys.* **1989**, *90*, 4916–4926.

(25) C. R. Jacob, M. Reiher *J. Chem. Phys.* **2009**, *130*, 084106.

(26) B. Njegic, M. S. Gordon *J. Chem. Phys.* **2008**, *129*, 164107.

(27) S. Grimme *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.

(28) S. Grimme, J. Antony, S. Ehrlich, H. Krieg *J. Chem. Phys.* **2010**, *132*, 154104.

(29) E. A. Coutsias, C. Seok, K. A. Dill *J. Comput. Chem.* **2004**, *25*, 1849–1857.

(30) NIST Standard Reference Database, Release September 2012. Editor: Russell D. Johnson III. See http://webbook.nist.gov/chemistry/ (accessed November 2012).

(31) Y.-P. Li, A. T. Bell, M. Head-Gordon *J. Chem. Theory Comput.* **2016**, *12*, 2861–2870.

(32) J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria *Phys. Rev. Lett.* **2003**, *91*, 146401.

(33) S. Grimme, S. Ehrlich, L. Goerigk *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(34) S. Grimme, M. Steinmetz *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.

(35) K. Ruud, P.-O. Astrand, P. R. Taylor *J. Am . Chem. Soc.* **2001**, *123*, 4826–4833.

(36) Y. Lin, G. Li, S. Mao, J. Chai *J. Chem. Theory. Comput.* **2013**, *9*, 263–272.

(37) R. Sure, S. Grimme *J. Comput. Chem.* **2013**, *34*, 1672–1685.

(38) D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville, C. Ochsenfeld *J. Chem. Theory Comput.* **2014**, *10*, 572–578.

(39) F. Jensen *J. Chem. Theory Comput.* **2014**, *10*, 1074–1085.

(40) M. W. Lodewyk, M. R. Siebert, D. J. Tantillo *Chem. Rev.* **2012**, *112*, 1839–1862.

(41) E. Benassi *J. Comput. Chem.* **2017**, *38*, 87–92.

(42) A. Bagno *Chem. Eur. J* **2001**, *7*, 1652–1661.

(43) F. Weigend *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(44) F. Weigend *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.

(45) F. Neese, F. Wennmohs, A. Hansen, U. Becker *Chem. Phys.* **2009**, *356*, 98–109.

(46) S. Kossmann, F. Neese *Chem. Phys. Lett.* **2009**, *481*, 240–243.

(47) R. Izak, F. Neese *J. Chem. Phys.* **2011**, *135*, 144105.

(48) K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs *Chem. Phys. Lett.* **1995**, *240*, 283–289.

(49) F. Weigend, M. Häser *Theor. Chem. Acc.* **1997**, *97*, 331–340.

(50) F. Weigend, A. Köhn, C. Hättig *J. Chem. Phys.* **2002**, *116*, 3175–3183.

(51) H. Günther *NMR Spectroscopy - Basic Principles, Concepts and Applications in Chemistry, 3rd ed.*; Wiley VCH: Weinheim, 2013.

(52) D. R. Whitman, L. Onsager, D. H. T *J. Am. Chem. Soc.* **1960**, *32*, 67.

(53) S. Grimme. ANMR, a nuclear spin system solver for larges systems, University of Bonn, 2017. The program can be obtained for academic use upon request at xtb@thch.uni-bonn.de.

(54) H. B. Mayes, L. J. Broadbelt, G. T. Beckham *J. Am. Chem. Soc.* **2013**, *136*, 1008–1022.

(55) S. Immel, F. W. Lichtenthaler *Liebigs Ann.* **1995**, 1925–1937.

(56) J. C. Hanson, L. C. Sieker, L. H. Jensen *Acta Cryst. B* **1973**, *29*, 797–808.

(57) K. Bock, R. U. Lemieux *Carbohydrate Research* **1982**, *100*, 63–74.

(58) C. Christofides, B. D. Davies, J. A. Martin, E. B. Rathbone *J. Am. Chem. Soc.* **1986**, *108*, 5738–5743.

(59) P. C. D. Hawkins *J. Chem. Inf. Model.* **2017**, DOI: 10.1021/acs.jcim.7b00221.

(60) A. Goeller, Benchmark input structures and most stable conformers generated with the 'best'-algorithm provided by the Bayer AG (Wuppertal), personal communication.

(61) A. D. Becke, E. R. Johnson *J. of Chem. Phys.* **2005**, *123*, 154101.

(62) F. Weigend, R. Ahlrichs *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.