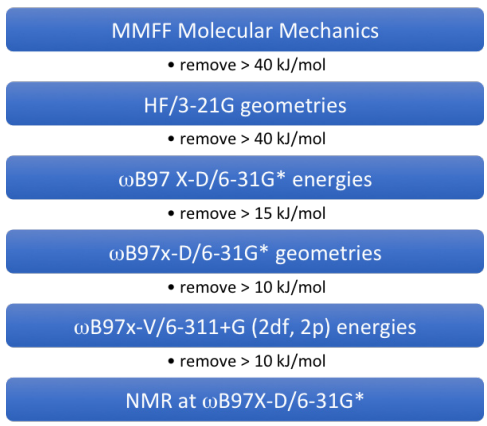
CALCULATING CHEMICAL SHIFTS IN CONFORMATIONALLY FLEXIBLE MOLECULES

Spartan'18 introduced a protocol for calculating ^{13}C chemical shifts in conformationally-flexible organic molecules, targeted at the natural products community with the goal of being able to identify problematic structure assignments [J. Nat. Prod. reference]. The protocol comprises a multi-step procedure to determine accurate Boltzmann weights starting from a systematic conformational search using MMFF molecular mechanics and terminating with a series of large basis set density functional energy calculations, and small basis set density functional chemical shift calculations that have been empirically corrected each of which is multiplied by its respective weight to produce a proper Boltzmann-averaged proton or ^{13}C spectrum. The procedure is depicted below.



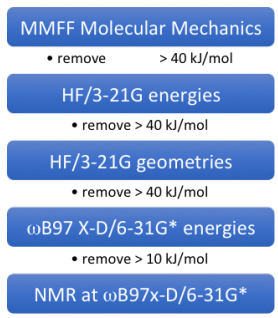
The protocol is fully automated requiring only input of a single conformer and designation of rotatable single bonds and flexible rings. Steps in the protocol may be eliminated, different quantum chemical methods employed and energy filters between steps adjusted. While the protocol has only been thoroughly assessed for ^{13}C chemical shifts, proton chemical shifts can be added to supplement (or replace) the ^{13}C data.

The performance of the protocol was first evaluated using ^{13}C chemical shifts for ~900 natural products the structures of which have been confirmed either by an X-ray crystal structure or by independent

synthesis. Only a very few outliers (rms deviations > 4 ppm) were noted, almost all of which can be attributed to line misassignments rather that incorrect structures. The protocol was then applied to a second set of ~2500 natural products, the structures of which rely solely on NMR data. Here, fully10-15% fall outside the anticipated error range and furthermore cannot readily be explained by line misassignments. These problematic structure assignments perhaps warrant further examination.

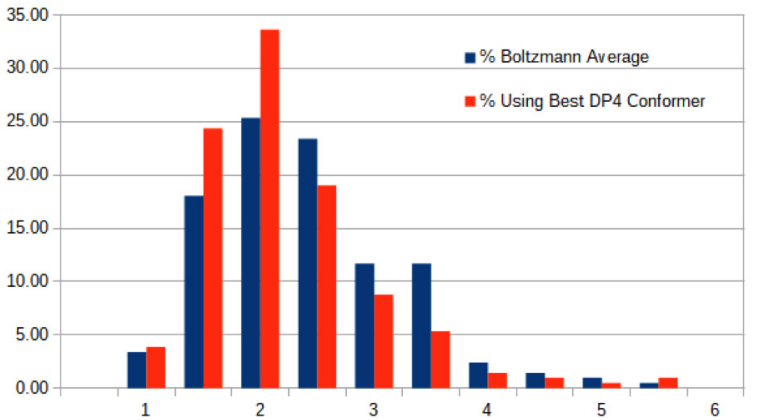
A major downside of the protocol available in ***Spartan'18*** is its high computation cost, due almost entirely due to the two final two steps required for evaluation of the Boltzmann weights. While present-generation personal computers allow its application to molecules with weights <500 amu and with limited conformational flexibility, computation times for larger natural products or those with more than a few hundred accessible conformers may in some cases exceed several days.

Spartan'20 introduces a lower-cost protocol that, instead of constructing a proper Boltzmann-weighted average, selects from the set of "reasonable" low-energy conformers the one that best matches the experimental ^{13}C and/or proton chemical shifts based on its DP4 score [DP4 reference]. Because this does not require calculation of accurate Boltzmann weights, it relies on simpler quantum chemical models and is roughly an order of magnitude less costly than the previous protocol, greatly extending its practical range. Of course it does require an experimental spectrum! The protocol is depicted below.



As with the previous protocol, alternative quantum chemical models and energy cutoff values may be substituted for those depicted above. In particular, the B3LYP/6-31G* density functional model may be substituted for ωB97X-D/6-31G* for both energy and NMR calculations, leading to an additional reduction in computation cost.

The two histograms below compare the performance of the original protocol based on Boltzmann averages (blue bars) and the alternative based on the best individual DP4 conformer score (red bars), for a xxx molecule subset of the natural products for which NMR structure assignments have been confirmed either by X-ray or independent synthesis. The metric is the smallest RMS deviation between calculated and experimental ^{13}C chemical shifts. The important conclusion is that the results are nearly identical. The fact that errors resulting from the new protocol are actually slightly smaller than those from the original should not be surprising as selection is based on the conformer that provides the best fit to the experimental data, irrespective of its energy.



The second pair of histograms compares the performance of the two protocols for a xxx molecule subset of the natural products for which structure assignments are based solely on NMR, and with overall RMS errors from the original protocol of 5 ppm or greater. Errors near the lower end of this range have previously been interpreted to indicate the possibility of structure or peak missassignments while those in the middle or higher range are symptomatic of structure or peak assignments that are most likely incorrect. Aside from the fact

that roughly 10% of the RMS values from the new protocol fall below 5 ppm, the two histograms are similar.