Supporting Information for

Direct wavepacket dynamics with spin-orbit coupling: simulation of thioformaldehyde

Swagato Saha, Léon L. E. Cigrang, and Graham A. Worth

Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom.

Contents

S1 Active space orbitals

Orbitals used in the active space based methods included in this work are shown in Figure S1. The character of each orbital is given, along with a label indicating in which of the three possible active spaces it is included; (12,10), (10,10), and/or (10,6).

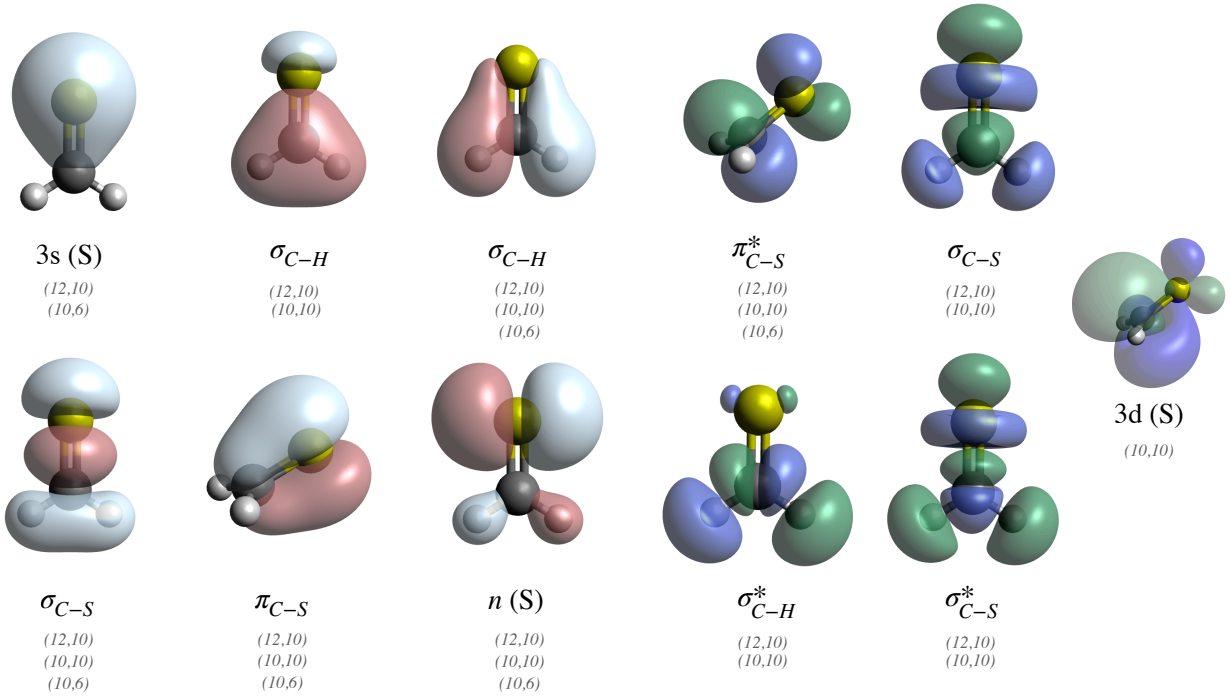

Figure S1 Orbitals included in various active spaces (grey labels) of thioformaldehyde. Red and light blue orbitals are initially doubly occupied, whereas green and dark blue orbitals are unoccupied.

S2 Vertical excitation energies

Table S1 Excitation energies of TF (in eV) computed using various methods, used in DD-vMCG dynamics in this work.

Method	Software	Basis set	S_1	T_1	T_2
MS-CASPT2(12,10)[a]	OpenMolcas	aug-cc-pVDZ	2.34	2.08	3.51
SA(2S+2T)-CASSCF(10,6)[b]	OpenMolcas	aug-cc-pVDZ	2.33	2.05	3.11
SA(2S+2T)-CASSCF(10,6)[c]	OpenMolcas	cc-pVDZ	2.31	2.03	3.08
MRCI(12,10)	Molpro	ano-rcc-pVDZ	2.34	2.13	3.28
SA(2S+2T)-CASSCF(12,10)	Molpro	aug-cc-pVDZ	1.96	1.80	2.63
MRCI(10,10)	Molpro	ano-rcc-pVDZ	2.12	1.86	3.14
SA(2S+2T)-CASSCF(10,10)	Molpro	aug-cc-pVDZ	1.77	1.58	2.50

[a] An IPEA shift of 0.25 was applied.
[b] Optimised geometry obtained from CCSD//aug-cc-pVDZ.
[c] Optimised geometry obtained from CASSCF(10,6)//cc-pVDZ, as per Ref. 1.

S3 Equilibrium geometry & Normal modes

The optimised ground state geometry, and its normal mode frequencies, for all simulations performed in Molcas was obtained using CCSD//aug-cc-pVDZ (Tables S2 and S4). One exception to this statement is the SA-CASSCF(10,6) calculation. Here, the minimum energy geometry and frequencies were calculated using CASSCF. This was done to match the procedure used in Ref. 1.

On the other hand, the ground state geometries and frequencies (Tables S3 and S4) for dynamics simulations based on Molpro calculations were obtained using the same method as during the *on-the-fly* dynamics (e.g. MRCI(10,10) used for both optimisation and DD-vMCG).

For completeness, normal frequencies obtained from the other calculations are presented, comparing them to experimental data (Table S4).[2]

Table S2 Cartesian coordinates (in Å) for thioformaldehyde, optimised using CCSD//aug-cc-pVDZ.

Atom	X	Y	Z
C	0.000000	0.000000	-1.031512
H	0.000000	0.924833	-1.611352
H	0.000000	-0.924833	-1.611352
S	0.000000	0.000000	0.588236

Table S3 Cartesian coordinates (in Å) for thioformaldehyde, optimised using MRCI.

Atom	X	Y	Z
C	0.000000	0.000000	-0.085642
H	0.935518	0.000000	-0.670619
H	-0.935518	0.000000	-0.670619
S	0.000000	0.000000	1.557879

Table S4 Normal Mode Frequencies (in cm^{-1}) of thioformaldehyde, as calculated using different levels of electronic structure theory, and as experimentally observed.

Label	Frequency (cm^{-1})						Description
	CCSD	MRCI(12,10)	MRCI(10,10)	CASSCF(12,10)	CASSCF(10,10)	Expt.	
q_1	1028.09	971.48	969.52	410.09	437.11	990.2	CH_2 oop bend
q_2	1044.82	1000.77	1002.10	880.28	881.99	991.0	CH_2 rock
q_3	1100.30	1027.86	1021.26	822.14	825.42	1059.2	C-S stretch
q_4	1557.24	1483.98	1484.21	1399.79	1400.19	1457.3	CH_2 scissor
q_5	3134.95	2983.08	3002.59	3027.99	3034.29	2971.0	CH_2 symmetric stretch
q_6	3222.79	3074.81	3094.51	3145.80	3154.66	3024.6	CH_2 asymmetric stretch

S4 Convergence

For DD-vMCG calculations, the level of convergence cannot be assessed based on rigorous criteria. Hence a proxy for convergence must be used. A variable parameter is chosen and the dependence of certain properties with respect to different values of that parameter is evaluated. The only practical parameter one can use in this case is the the number of GWPs included in the nuclear basis. As for the properties, we choose to look at population transfer from S_1, and also compare the level of energy conservation during the dynamics simulations using different basis sizes. We note that all simulations were performed on the same quantum chemistry database (the CASPT2(12,10) one). In principle, the number of points in a database could also be used to assess convergence but given that our TF databases contain enough *ab initio* points (i.e. a minimal number of points is added with each additional simulation), we consider the databases "converged".

S4.1 Population transfer

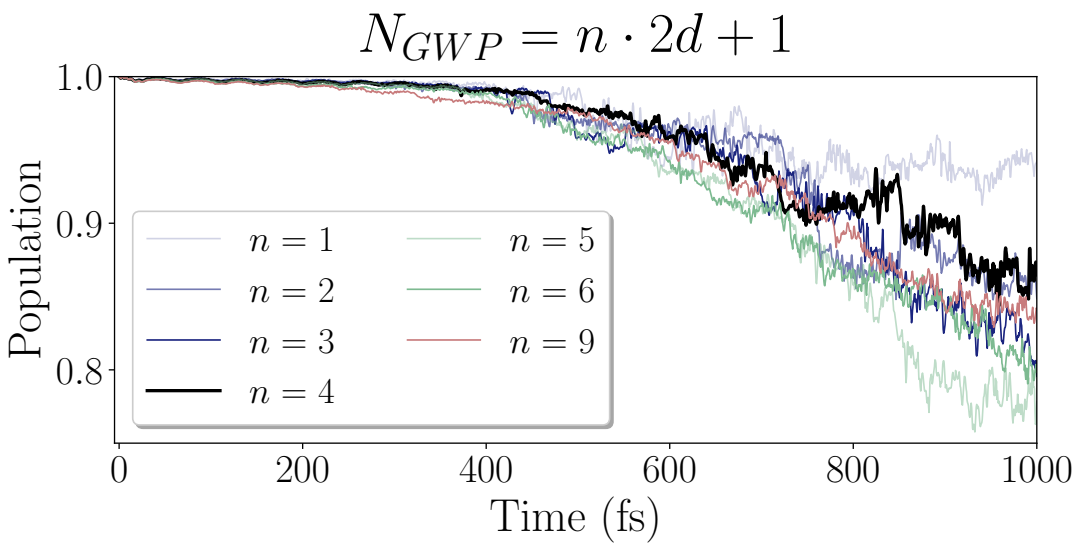

Figure S2 Population of the diabatic S_1 state during DD-vMCG simulations using 4 different basis set sizes. Using the $n \cdot 2d + 1$ heuristic, the increasing values of n correspond to 13, 25, 37, 49, 61, 73, and 109 GWPs, respectively.

We compare the change in population of the S_1 state, plotted in Figure S2 for different simulations using increasing basis sizes according to the $n \cdot 2d + 1$ heuristic described in the main text. A clear trend cannot easily be identified, which may indicate that, in this case, it is not a suitable metric for convergence. On the other hand, it may provide some insight into the physics of the simulation. All calculations seem to predict the same behaviour, i.e. gradual

population transfer after about 400-500 fs, only to different degrees. The amount of transfer ranges between 75-95% at 1000 fs - the differences are amplified at longer times, in part due to numerical errors (as reflected in the energy fluctuations, presented hereafter) - and the rate at which population is transferred is very similar across all simulations. This likely indicates purely quantitative differences, and strengthens our claim that the $n = 4$ simulation adequately describes the photophysics in TF. Furthermore, the next section puts forward the claim that the $n = 4$ simulation is also the most numerically accurate, based on energy conservation arguments.

S4.2 Energy conservation

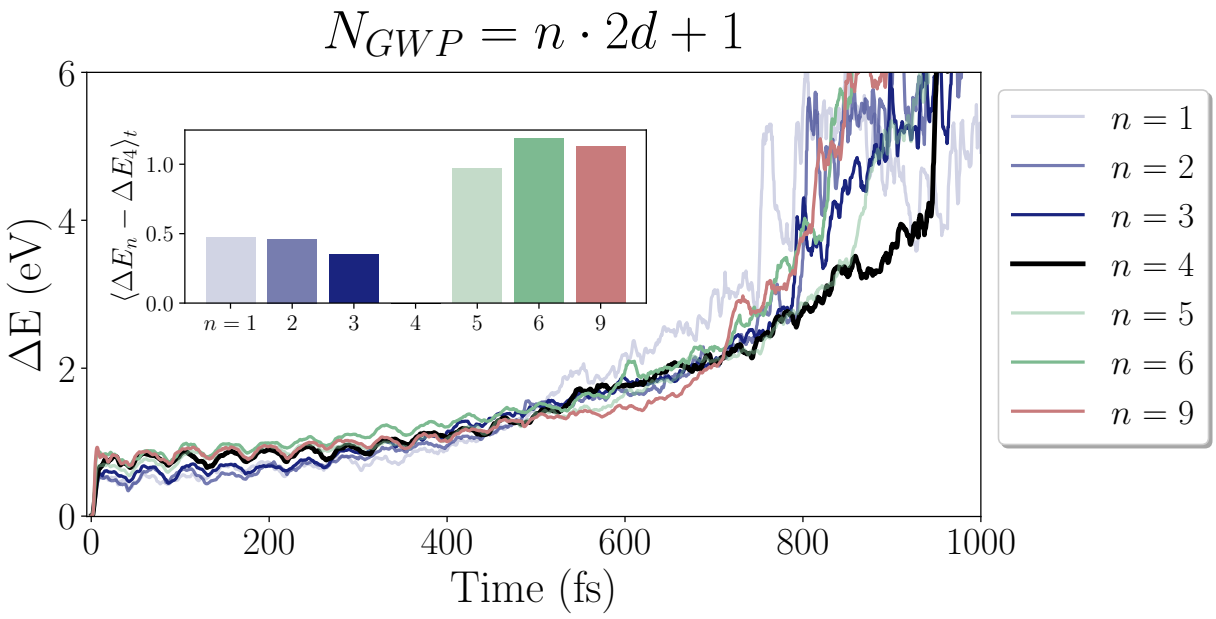

Figure S3 Error accumulated on the total energy during 1000 fs DD-vMCG simulations. Using the $n \cdot 2d + 1$ heuristic, the increasing values of n correspond to 13, 25, 37, 49, 61, 73, and 109 GWPs, respectively. Inset shows the total difference between the all results with the $n = 4$ result.

We move on to considering energy conservation during DD-vMCG dynamics. The use of a variational basis ensures, by construction, energy conservation provided the Gaussian integrals are done exactly - independent of basis-set completeness.

$$\frac{dE}{dt} = 2\Re\Big(\sum_u \dot{\lambda}_u K_u^{\text{exact}}\Big), \qquad K_u^{\text{exact}} = c^\dagger H^{(u)} c - c^\dagger H S^{-1} S^{(u)} c. \tag{1}$$

λ_u denotes the parameters of the Gaussian basis functions, c describes the variational coupling between the GWPs,

while the real part of the summation relates to the drift in total energy. Using the projector onto the basis span, the coefficient may be written as:

$$K_u^{\text{exact}} = \langle \psi | \hat{H}(I-P)\partial_{\lambda_u}\psi \rangle. \tag{2}$$

Under the stationarity condition of the time-dependent variational principle in vMCG, energy is conserved.

$$\langle \psi | \hat{H}(I-P)\partial_{\lambda_u}\psi \rangle = 0 \quad \forall u. \tag{3}$$

Note that this fact is independent of basis set size, and hence this parameter cannot generally be used to asses the level of accuracy of a vMCG simulation. However, as a result of using the LHA (to second order) for the solution of the Gaussian integrals, some error is introduced in the form of an energy drift. In other words, energy is conserved by construction with respect to the approximate Hamiltonian only ($\hat{H}^{LHA}$), and the resulting energy drift during propagation, under the exact Hamiltonian, can be compared across simulations with increasing basis size (Figure S3). Furthermore, it may be shown that the resulting loss of energy conservation is conditioned to the completeness of the GWP basis, i.e., one expects the total energy to be better conserved for a more complete basis.

$$\Delta\hat{H} = \hat{H}^{exact} - \hat{H}^{LHA}, \qquad K_u^{\text{LHA}} = 0. \tag{4}$$

$$\frac{dE}{dt} = 2\Re\Big(\sum_u \dot{\lambda}_u \langle \psi | \Delta\hat{H}(I-P)\partial_{\lambda_u}\psi \rangle \Big). \tag{5}$$

This allows us - comparing energy conservation as a function of time for different basis sizes - to make a quantitative assessment of how converged the simulations are.

In Figure S3 we perform such a comparison. The drift/error on the total system energy ($\Delta E = E^{tot} - E^{error}$) is plotted for different values of n. Importantly, note that the energy reported in the figure is the expectation value of the whole wavefunction, i.e. it combines the energy drift on all individual GWPs. To get a better sense of the absolute error, i.e., how well is energy conserved for an 'effective DD-vMCG trajectory', see the normalised energy drift as reported in Figure S8. The inset in Figure S3 shows the time-averaged difference in energy drift *compared to n = 4* ($\langle \Delta E_n - \Delta E_4 \rangle_t$). This difference is positive in all cases, meaning the time-averaged energy drift is minimised for $n = 4$, i.e., it conserves energy best for the longest. For $n < 4$, the simulations show an increased drift in energy due to basis

set incompleteness, as expected. For $n > 4$, energy conservation is compromised when the linear dependencies due to overlapping GWPs - more pronounced as the basis set approaches completeness - become too acute to be resolved by regularisation. Note that the initial jump in energy, seen in all simulations, is also due to linear dependencies present during the first few time-steps since all GWPs start at the same point in position space. This is naturally resolved as the propagation causes the basis functions to spread out.

Given that the total energy is acceptably conserved for the $n = 4$ basis - barring numerical errors that inflate energy at long propagation times (regardless of the GWP basis) - the results could be said to be quantitatively accurate.

S5 Distribution of SOC elements

Figure S4 shows the distribution of effective SOC elements (see Equation 13 in main text) for each record stored in the database, i.e. every geometry for which properties have been calculated, for both reference simulations. The main observation to note is that the El-Sayed allowed coupling between states (i.e. S_0-T_1 and S_1-T_2) remains relatively constant across the entire configuration space explored during the dynamics simulation. Between states of the same orbital symmetry, the S_1-T_1 coupling is close to zero in most regions while the elements between S_0-T_2 show regions with larger magnitudes. Note however that the S_0-T_2 energy gap is the largest among all pairs of states and direct ISC between these states is not thought to be viable.

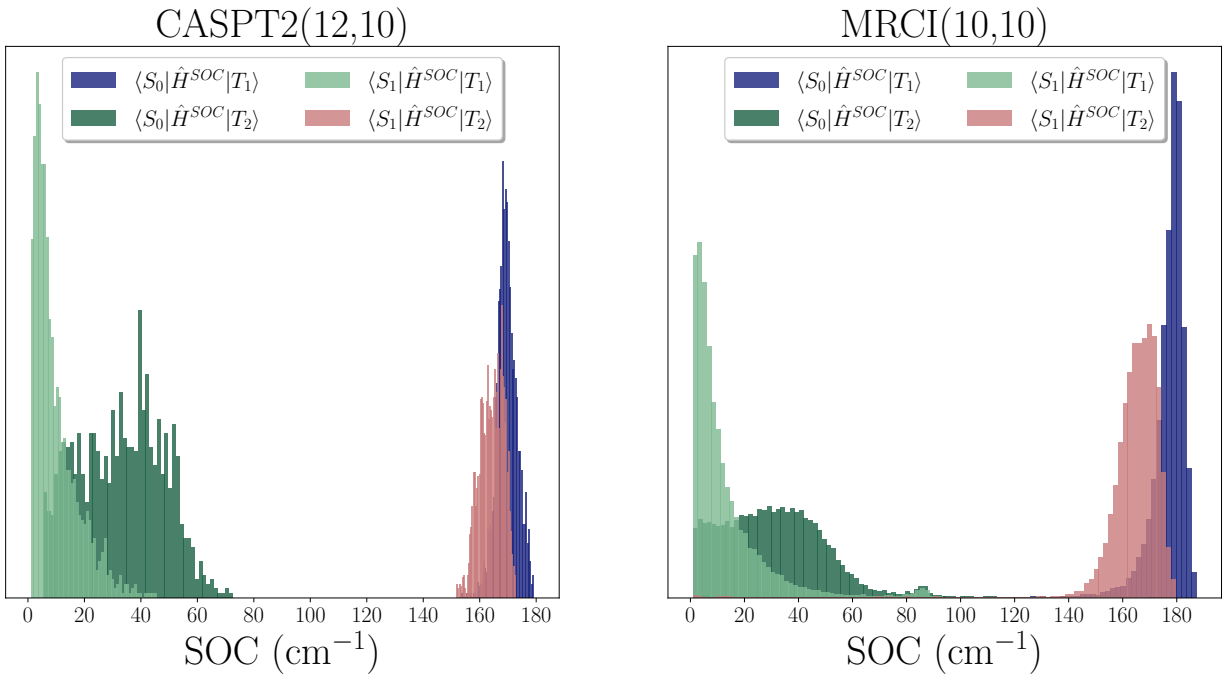

Figure S4 Distribution of SOC elements in the database of both reference calculations.

This analysis can be used to asses the suitability of employing a spin-diabatic formalism with effective SOC elements, independent of the dynamics method. In systems exhibiting a strong geometry dependence of the SOCs (i.e. showing a much broader distribution), dropping the SOC phase can create problems. This is not the case however, if the SOC oscillations are mere QC artifacts (components change signs randomly, magnitudes are constant). The fact that the SOC elements in TF only show a slight geometry dependence, and interstate coupling is prominent only

for 1-singlet-1-triplet blocks, i.e. (S_1, T_2) & (S_0, T_1), means one can safely adopt the effective SOC approach. The advantage of doing so is that it can trivially rectify the random phase oscillations found in QC calculations of SOCMEs, by extracting their raw magnitudes and neglecting the phase.

In order to confirm that the slight geometry dependence observed in the distributions is negligible, we repeated our reference simulations (CASPT2(12,10) and MRCI(10,10)) with constant SOC elements by taking the magnitude at the FC point - under this approximation, the SOC phase factor may be neglected. Furthermore, the El Sayed forbidden pairs were not coupled (SOC = 0). A significant departure from the reference calculations would suggest these conditions do not apply to TF. The results are shown in Figure S5, and do not exhibit any major deviations. This also confirms the S_1-S_0-T_1 population transfer channel - IC followed by ISC. S_0-T_1 populations are strictly correlated. The validity of this SOC protocol, outside such simplifications, depends on the nature of the dynamics method.

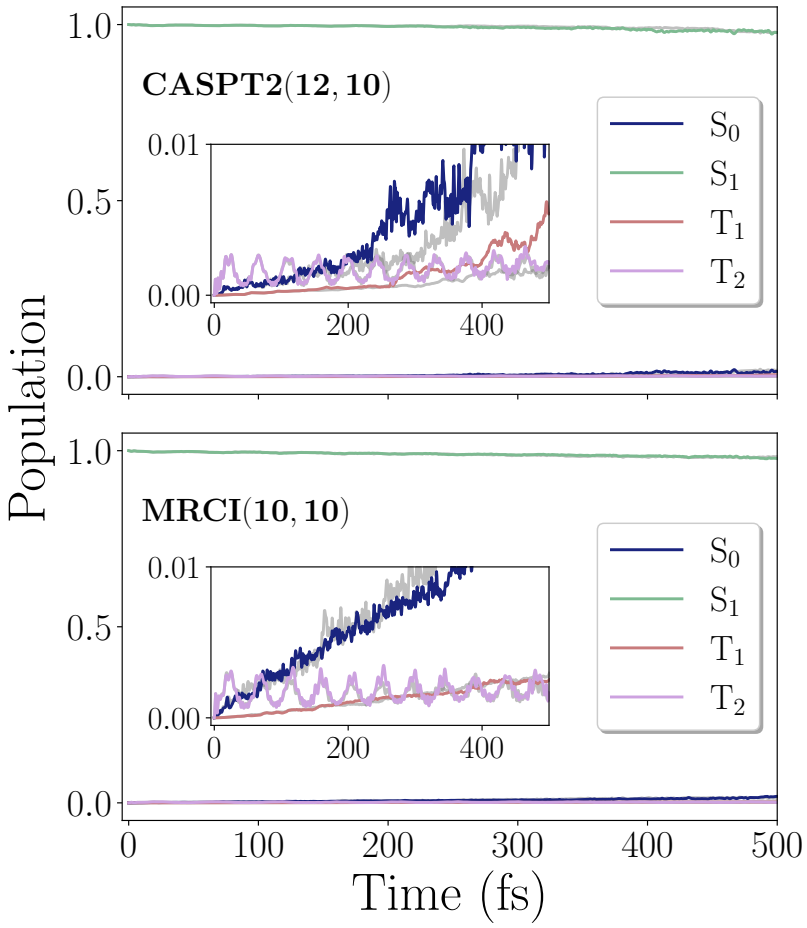

Figure S5 Diabatic state populations of reference simulations performed up to 500 fs with constant SOC elements, defined at the FC point. The grey lines indicate the reference result, as presented in Figure 5 in the main text.

S6 Distribution of NAC elements

The distribution of all non-adiabatic coupling (NAC) elements between states S_0 and S_1 are shown in Figure S6. The magnitude of the coupling is of a similar order as the El Sayed allowed SOC elements of Figure S4, and on average slightly higher. This is consistent with the observation that the amount of IC exceeds that of ISC. Note also that a few NAC elements have a very large magnitude (not shown), corresponding to geometries where (near) degeneracies occur.

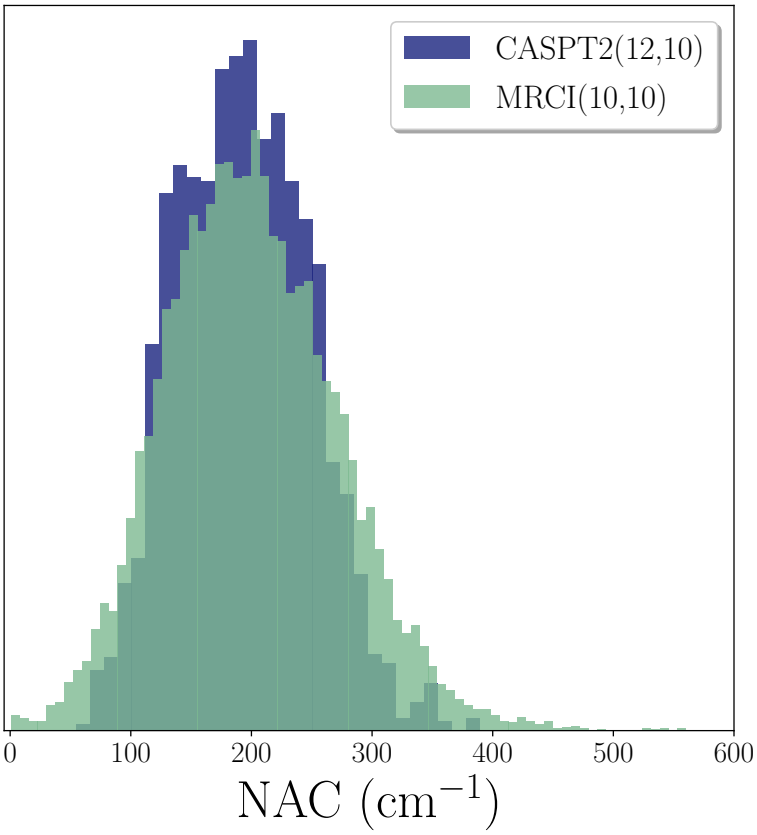

Figure S6 Distribution of NAC elements in the database of both reference calculations.

S7 Results of additional simulations

In Figure S7, the change in population of the S_0, T_1, and T_2 states for all additional DD-vMCG simulations (summarised in Table 3 of the main text) is shown.

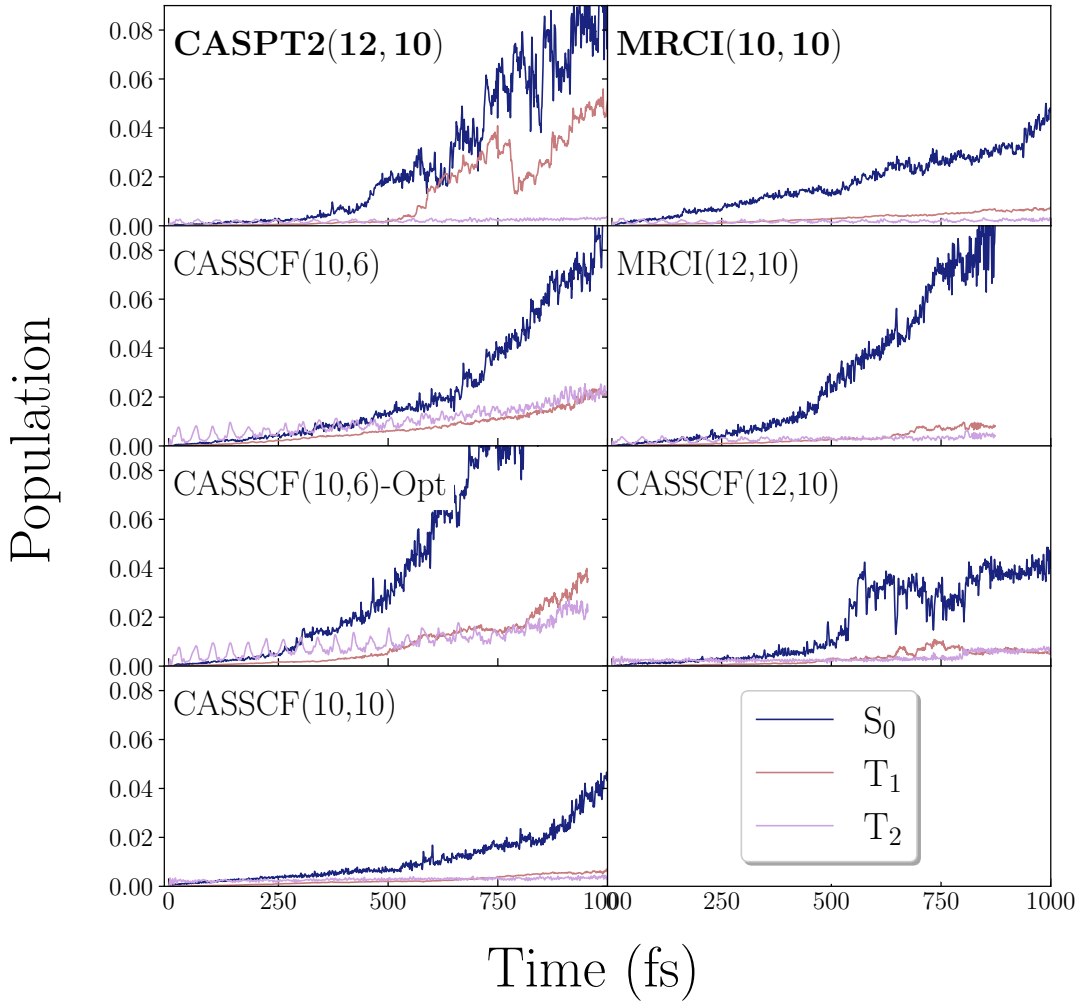

Figure S7 Diabatic state population of S_0, T_1, and T_2.

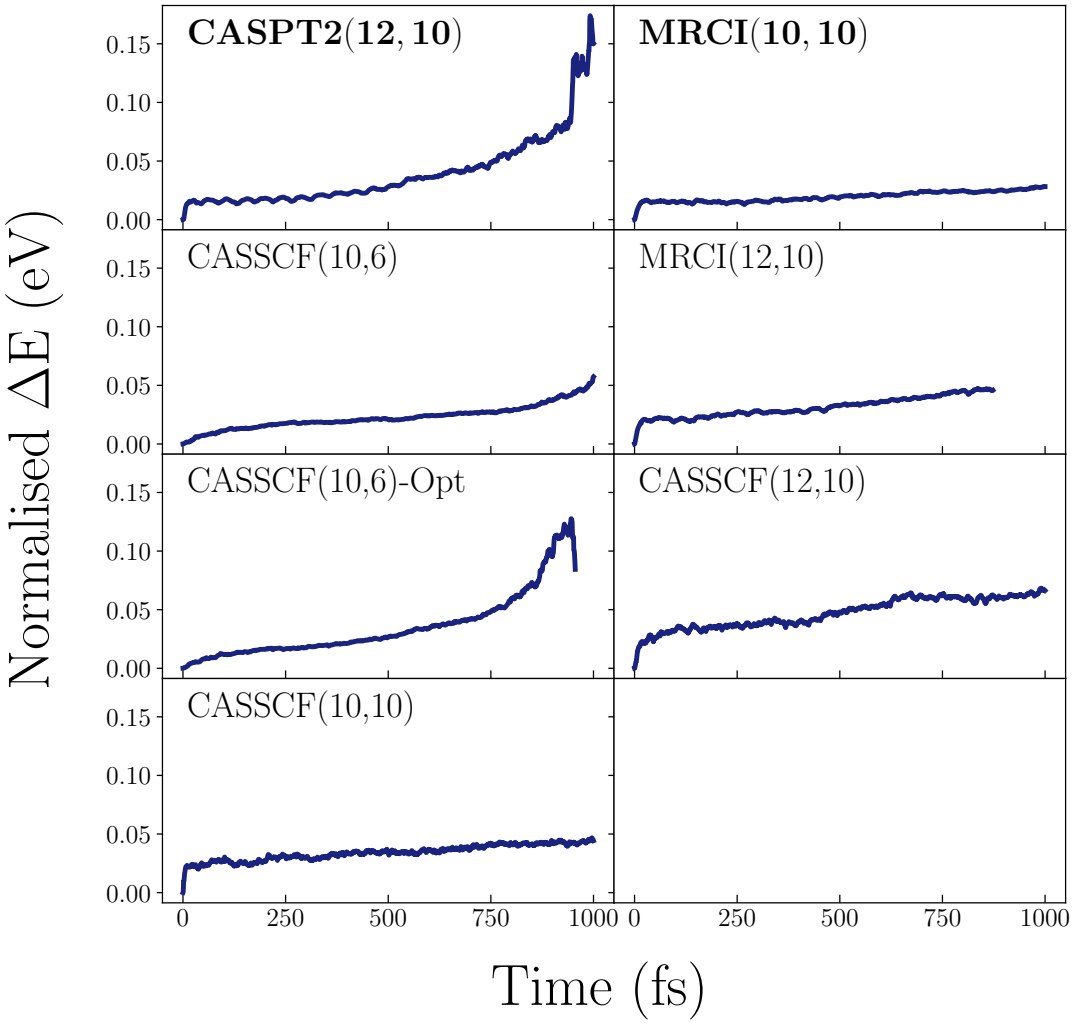

Figure S8 Normalised error on the total energy

The error accumulated in the total energy for all DD-vMCG simulations is plotted in Figure S8 as a normalised energy difference:

$$(Normalised)\Delta E(\mathrm{t}) = \frac{E(t) - E(t_0)}{N_{GWP}} \tag{6}$$

The normalisation is due to the fact that the GWPs in DD-vMCG simulations are coupled, and therefore the total error in the energy is spread across all individual basis functions. As such, this metric gives an indication of the error associated with each individual GWP (an 'effective DD-vMCG trajectory') and thus allows for a proper comparison between simulations with a different number of GWPs. Additionally, one could compare these values directly with, for example, trajectory-based methods where the error on the energy is also calculated per individual trajectory.

S8 Wigner-sampled initial conditions

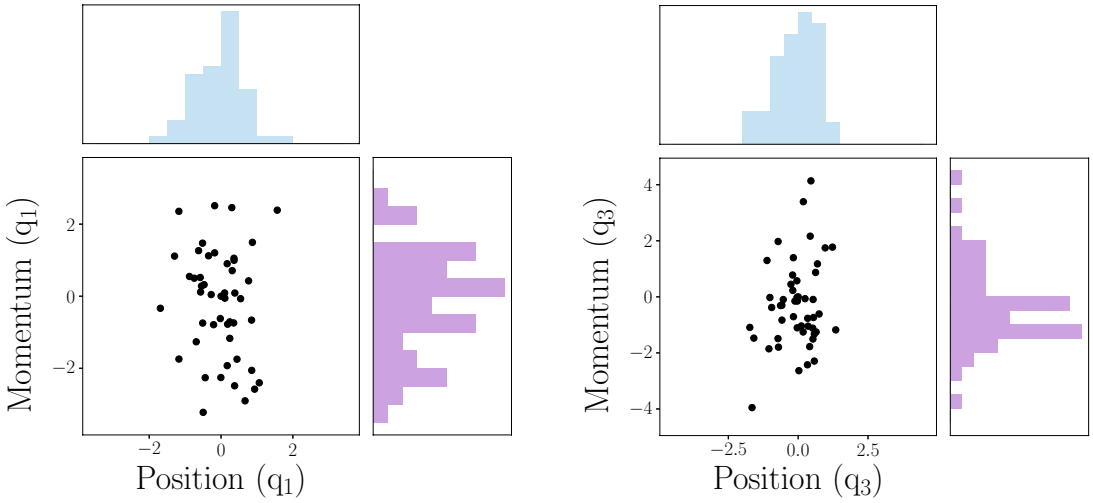

Figure S9 Distribution of initial conditions for 49 GWPs obtained from Wigner sampling. Positions and momenta are plotted along the q_1 mode (left) and the q_3 mode (right).

References

[1] S. Mai, A. J. Atkins, F. Plasser and L. González, *J. Chem. Theory Comput.*, 2019, **15**, 3470–3480.

[2] D. J. Clouthier and D. A. Ramsay, *Annu. Rev. Phys. Chem.*, 1983, **34**, 31–58.