H. Ihee et al.,
"Ultrafast X-ray diffraction of transient molecular structures in solution."

Supporting Online Materials

Materials and methods

Experimental setup. The experimental scheme (S1,S2,S3) is as follows. For the sample cell system, two quite different setups were used; a capillary-based system and an open-jet system. In the first system, a solution of $C_2H_4I_2$ dissolved in methanol (CH_3OH) was circulated through a quartz capillary (300 μm diameter) to provide a stable flow. In the open-jet system, the capillary was removed and a stable jet was provided by a high-pressure nozzle. The samples were purchased from Sigma Aldrich and used without further purification.
The concentration of $C_2H_4I_2$ was 60 mM, which corresponds to a solute-solvent ratio 1:412. This concentration was chosen to maximize the photo-product yield while keeping the inter-solute contribution negligible.
The fs laser pulses from amplified Ti:Sapphire laser system running at 1 kHz were first frequency tripled to a wavelength of 267 nm, and then stretched by passing them through two fused silica rods (length: 2 x 15 cm) to lower the peak power of the pulse, thereby avoiding multi-photon excitations. The resulting picosecond laser pulses ($\sim$2 ps, 267 nm, 35 μJ/pulse) were directed into the sample to initiate the reaction.
Polychromatic X-ray pulses of $\sim$100 ps duration and $\sim$5x10^8 photons per pulse were delivered from an in-vacuum undulator on beamline ID09B at the European Synchrotron Radiation Facility. The X-ray spectrum is peaked at 18.2 keV and has a bandwidth of 0.45 keV. The scattered X-ray signal was recorded on a 133 mm diameter fiber-coupled MarCCD at well-defined delay times (t) relative to the excitation laser pulse.
Ab Inition and DFT calculations. The possible structures of the parent molecule, the transient intermediates and the products in the gas phase and in solution, were calculated by the *ab initio* and Density Functional Theory (DFT) methods using the code GAUSSIAN 94 (S4). These species include $C_2H_4I_2$, bridged C_2H_4I, anti C_2H_4I, C_2H_4I-I (isomer), C_2H_4, I and I_2. For the geometry of the anti C_2H_4I radical, the Hartree-Fock (HF) and the MP2 methods were used because this anti form converges to the bridged structure in the DFT calculation. For all other structures, the DFT results were used. The structural parameters optimized with the MP2 and CCSD(T) methods were used for comparison. The Becke three parameter hybrid functional with the Lee-Yang-Parr correlation corrections (B3LYP) was used to perform the DFT calculation (S5,S6). The all electron basis

sets 6-311++G** were used for the C and H atoms, and the all electron basis sets 6-311G* with added d and f polarization functions and s and p diffuse functions for I atom, (S7,S8) which are referred to 6-311+G*, were used. Geometries were fully optimized, vibrational frequencies were calculated with analytical second derivatives, and zero point vibrational energies were derived. The solvent effects were calculated via self-consistent reaction field (SCRF) theory. The SCIPCM (S9) model, which allows geometry optimization at the HF and DFT level, were used to perform the solution calculation with the iso–density value of 0.0004 e/au^3.
To obtain the charge on each atom of all related species, the natural bond orbital analysis was performed. These structures and the charges of all species were used as starting points for the high angle analysis and for the MD simulations.
MD simulations. The MD simulations were performed using the program MOLDY (S10). The molecules were treated as rigid. For the description of methanol we used the so-called H1 model (S11).
For the atoms in the solute molecules we used an all-atom approach with the following parameters:

Atom	σ (Å)	4ε (kJ/mol)
C	3.905	0.472
H	2.600	0.444
I	4.100	10.444

No attempt to refine these parameters was made and the Lorentz-Berthelot's rules were used. The charges on different atoms were obtained from the DFT calculations and were kept fixed during the simulation.
For the simulation of the solvent response, we used as many as 2048 rigid methanol molecules contained in a cubic box with a side length of ~ 50 Å in order to avoid parasitic oscillations in the differentials at low q. The radial distribution functions (RDF) were calculated up to 50 Å in steps of 0.02 Å and used for the calculation of the scattered intensity up to 33 Å. Three thermodynamic conditions were simulated:

C1 : T=300 K, ρ=785.9 Kg/m^3;
C2 : T=328 K, ρ=785.9 Kg/m^3;
C3 : T=300 K, ρ=759.8 Kg/m^3;

Taking the difference between the scattered intensity of the system in condition C2 (I[C2]) from that in C1 and further dividing the difference by the temperature change, gives the "constant volume" component (the first curve of Fig. 2 C). In the same way, taking I[C3]-I[C1] and dividing it by the density change, we obtained the "density derivative" of the solvent (the second curve of Fig. 2 C). The basic assumption is that the change in solvent scattering, induced by the input of energy

from decaying solutes, can be written as (S12):

$$\Delta S = \left(\frac{\partial S}{\partial T}\right)_{\rho} \Delta T + \left(\frac{\partial S}{\partial \rho}\right)_{T} \Delta \rho$$

The first term is the response to a temperature rise at constant volume and the second the response to a change in density. Note that the constant volume derivative is the same as the constant density one because the number of particles is the same for all simulations and the experimental data have been normalized to a constant number of particles. The temperature term is activated at the early times in the reaction, typically between 0-10 ns, when energy flows into the solvent at constant volume. The second term is activated during the subsequent thermal expansion. The derivatives are determined by MD simulations and the temperature rise and the change in density can be extracted from global fitting.

For the MD calculation of the solute diffraction, we used 1 solute molecule ($C_2H_4I_2$, bridged C_2H_4I, anti C_2H_4I, C_2H_4I-I (isomer), C_2H_4, I and I_2) surrounded by 256 methanol molecules in a cubic box of ~ 25 Å (side length). These simulations were performed at ambient temperature (300 K) and at a density of 791 kg/m^3. The scattered intensity from the aforementioned 6 molecular systems was calculated considering the intra-solute RDFs and the solute-solvent cross term (cage). The effects of change in the solvent cages is summarized in Figure S1 in which we show (in both q and r space) the difference maps for different transitions with and without taking into account the solute–solvent cross terms. The systems were equilibrated at 300 K over 200 ps at constant temperature, through coupling to a Nose-Hoover thermostat (S13) with a time constant of 0.3-0.7 ps, depending on the system. The step size was 0.5 fs; the simulations were performed in the NVT ensemble and the trajectories were followed up to 1 ns. The equations of motion were integrated with a modification of the Beeman algorithm (S10). Periodic boundary conditions were used and the Van der Waals interaction was cutoff beyond 11 Å. Electrostatic interactions were calculated by the Ewald sum (S14); the alpha parameter was set to 0.19 or 0.27 Å^{-1} depending on the number of the molecules in the box. The scattered intensities were obtained from the RDFs (calculated every 50 fs during the simulation) using our own programs and tabulated atomic form factors.

Fourier Transform. Since it is difficult to give an intuitive meaning to the result in q space, we present the result also in the real space. The relation between these two quantities is given by a sine Fourier transform (FT) as follows:

$$r\,\Delta S(r) = \frac{1}{2\pi^2} \int_0^\infty q\,\Delta S(q) \sin(q\,r)\,dq \qquad (1)$$

In our experiment the accessible q range is 0.5 – 8.5 Å^{-1}. For this reason we used the following equation:

$$r\,\Delta S(r) = \frac{1}{2\pi^2}\int_0^\infty q\,\Delta S(q)\sin(q\,r)\exp(-q^2\alpha)\,dq \qquad (2)$$

In this equation the "real" $\Delta S(r)$ is convoluted with a gaussian function ($\exp(-(\pi\,r)^2\alpha)$). We used a value of 0.03 Å^2 for α. The FTs have been done with simple numerical integration (no maximum entropy strategy has been used). Because of the absence of experimental points in the region 0–0.5 Å^{-1} we filled the gap with the theoretical curves to extend the data and in this way minimize the truncation in the FT. This procedure does not bias the $\Delta g(r)$.

Data treatment.

In this section we discuss the procedure to calculate the difference curves from the raw experimental data. It will be shown that the shape of the curves does not depend on the data treatment approach.

Conversion of the 2-D images into 1-D curves. The 2-dimensional diffraction patterns were converted to 1-dimensional intensity curves by calculating the average intensity as a function of the scattering angle (2θ) from the beam center. The data were corrected for the scattering from air and the capillary, space-angle conversion, polarisation and absorption.

How to obtain the difference maps Because the change in the scattered intensity between the ground and the excited states is very small ($\sim 0.1\%$), one cannot simply take the difference of the 2 images (laser on - laser off) but needs to scale the raw data with appropriately.

If an ideal incoming beam monitor of infinite precision was available, one could simply scale every image by the value measured by the monitor and then take the difference. In this way a "constant volume" difference curve could be generated. Unfortunately we need a very accurate scaling factor (0.005%) and such a precise counter is not available yet. For these reasons we have used another approach of scaling the images to the same number of molecules. Note that this approach does not require any incoming-beam counter because the image itself is used to obtain the scaling factor.

During the course of structural changes in the sample the scattering pattern changes across entirely q range but for isosbestic points, i.e. points at which the intensity does not change. If for all the possible changes (solvent heating, different reaction channels etc.) at least one isosbestic point exists, one could use this point for scaling and our approach would be exact to within statistical error on determination of the scattered intensity at that q. General situations are more complex than this ideal case for the following reasons. First, difference curves (at

constant number of molecules) for all possible changes should be known to identify isosbestic points. Second, in our experiment a polychromatic X-ray beam rather than the monochromatic beam was used to gain high flux of X-ray photons. As a consequence the effective scattering intensity at a given solid angle is a sum of scattering intensities by a range of scattering vectors with different wavelengths. This in general may shift or even "destroy" an isosbestic point. A typical effect of this polychromatic correction is given in Figure S2. Third, as the number of the reaction channels increases it becomes more difficult to identify isosbestic points, which makes our approach an approximation. To obtain optimal approximation we need to find the best normalisation range, which is the range where the average scattered intensity is calculated and used as a scaling factor. Of course this range depends on the sample/reaction investigated. For the photodissociation of $C_2H_4I_2$ possible reaction pathways and solvent response curves are shown in Figure S3. As one can see there is no common isosbestic point, but the range of 45.5-46.5 degrees, which corresponds to 6.84 - 6.99 Å^{-1}, seems to be the best one. To investigate the influence of the normalization range on the final result, we compared the same images processed using a different normalization range of 45-55 degrees as shown in Figure S4. Even if the curves do not coincide, the shape is almost the same within the experimental error. To confirm that our choice of the normalization range does not bias the results, we fit the difference maps obtained with 3 different normalization ranges (45-55, 45.9-46.6 and 52-54) and compared the results. All these analysis gave the same results that we present in the following sections.
Average over different repetitions and noise in the FT. The measurements have been performed using the following strategy:

$$\underbrace{delay(-3ns) \rightarrow delay(\#1) \rightarrow delay(-3ns) \cdots \rightarrow delay(\#N) \rightarrow delay(-3ns)}_{repetition\ \#1}$$

$$\underbrace{delay(-3ns) \rightarrow delay(\#1) \rightarrow delay(-3ns) \cdots \rightarrow delay(\#N) \rightarrow delay(-3ns)}_{repetition\ \#2}$$

$$\ldots$$

By grouping $2-3$ repetitions we obtain what we will call in the following of this report, a *sub–run*. We checked that the repetitions grouped in the same *sub–run* were equal, i.e. indistinguishable within the noise level. After this grouping we have still 6 or 7 *sub–run*s that can be used to find the error bar to assign to the measurement and to their FT.
The Figures S5 and S6 show the result of the procedure described above; not all

of the time-delays have been recorded for all the *sub–run*s.
In Figure S7, we show in an expanded scale the FT of the data for the 100 ps time delay. Note that the shape of the curves is almost the same within the experimental error and the peak positions are stable within 0.04 Å.
In the range 1 – 6 Å, the noise in the data does not introduce any artificial peaks or shoulders. We used the average over the different *sub–run* as data and the standard deviation as error bars. This procedure has been used in both real (r) and reciprocal (q) spaces.

Discussion of the theoretical model.
In the following pages we describe the most important features of the data analysis. A detailed description of the hydrodynamical calculations applied to time resolved X-ray scattering can be found in (S15,S16).

Reaction. The reaction considered is represented in Figure S8. The syntax used is as follows :

label	molecule
g	ground state, $C_2H_4I_2$
t	anti OR bridge transient structure, C_2H_4I
iso	isomer geometry, C_2H_4I-I
$c2h4$	C_2H_4
$I2$	I_2
I	I

The black lines in Figure S8 represent possible dissociation pathways of the ground state. The colored ones represent ensuing reactions. The equations that describe the system are as follows (k and D stand for the rate coefficients):

$$\begin{aligned}
\partial c_g/\partial t &= -k_{g\to t+I}\cdot c_g - k_{g\to c2h4+I2}\cdot c_g - k_{g\to c2h4+I+I}\cdot c_g \\
\partial c_t/\partial t &= -D_{t+I\to iso}\cdot c_t\cdot c_I - k_{t\to c2h4+I}\cdot c_t + k_{g\to t+I}\cdot c_g \\
\partial c_{iso}/\partial t &= +D_{t+I\to iso}\cdot c_t\cdot c_I - k_{iso\to c2h4+I2}\cdot c_{iso} \\
\partial c_{c2h4}/\partial t &= +k_{t\to c2h4+I}\cdot c_t + k_{iso\to c2h4+I2}\cdot c_{iso} + k_{g\to c2h4+I2}\cdot c_g + \\
&\quad +k_{g\to c2h4+I+I}\cdot c_g \\
\partial c_{I2}/\partial t &= D_{I+I\to I2}\cdot n_I^2 + k_{iso\to c2h4+I2}\cdot c_{iso} + k_{g\to c2h4+I2}\cdot c_g \\
\partial c_I/\partial t &= k_{t\to c2h4+I}\cdot c_t - 2D_{I+I\to I2}\cdot n_I^2 - D_{t+I\to iso}\cdot c_t\cdot c_I + \\
&\quad +k_{g\to t+I}\cdot c_g + 2\,k_{g\to c2h4+I+I}\cdot c_g
\end{aligned}$$

By integrating the equations above (with a set of reaction constants), the population of the different species are obtained as a function of time. In our fitting

program the integration has been done using a Fortran library written at CERN (DDEQBS) (S17). The relative accuracy has been found better than 10^{-6} for all the time delays.

In the next 3 sections we explain how to link the response of the solvent to the populations of the solute species as a function of time. For this, the solvent temperature and the solvent density are obtained as a function of time by calculating the heat released to the solvent. More details can be found in (S15,S16).

Heat released as a function of time. The calculation of the heat released to the solvent is straightforward once the population of the different states are known, and can be done as (we set $E_g = 0$):

$$\begin{aligned} Q(t) &= \frac{N_A}{R}\left[h\,\nu\sum_k f_k(0) - \sum_k E_k f_k(t)\right] \\ &+ \frac{N_A}{R}\,h\,\nu\,f_{fast}\,[1-\exp(-t/\tau_{fast})] \\ k &= \{t, iso, c2h4, I2, I\} \end{aligned}$$

In this expression Q is the heat released per mole of solvent molecules; E_k is the energy of the state k (in J/molecule) and $h\nu$ is the energy of one photon (in J), R is the ratio between solvent molecules and solute molecules, N_A the Avogadro Number. The last term (f_{fast}) has been inserted to take into account the very fast recovery of the ground state via vibrational cooling; the relative time-scale τ_{fast} has been kept constant as 10 ps. The energies of the different states have been calculated using DFT/HF method (S18).

Hydrodynamics. Defining $\sigma = (C_v - C_p)\frac{\chi_T}{\alpha_P}$, where χ_T is the isothermal compression coefficient and α_P is the isobaric thermal expansion coefficient, we have

$$\begin{aligned} \Delta P(t) &= \frac{\alpha_P}{\chi_T\,Cv}\int_{-\infty}^{t}\frac{\partial Q}{\partial t}(t')\exp(-(v_s/(a\sqrt{2})\,(t-t'))^2)dt' \\ \Delta T(t) &= \frac{Q(t)-\sigma\,\Delta P(t)}{Cp} \\ \Delta\rho(t) &= \rho_0\,[\chi_T\,\Delta P(t) - \alpha_P\,\Delta T(t)) \end{aligned}$$

with ΔP in Pa (Pascal), ΔT in Kelvin and $\Delta\rho$ in kg/m^3, v_s is the speed of sound in the solution and a is the laser beamsize (RMS). The derivatives and integrations are performed numerically using spline interpolation (S19).

Scattered intensity calculation. The scattering intensity has been calculated by means of the following expression:

$$\begin{aligned} \Delta S(q,t) &= \frac{1}{R}\left[\sum_k f_k(t)\,S_k(q) - S_g(q)\sum_k f_k(0)\right] + \left(\frac{\partial S}{\partial T}\right)_\rho \Delta T(t) + \left(\frac{\partial S}{\partial \rho}\right)_T \Delta\rho(t) \\ k &= \{g, t, iso, c2h4, I2, I\} \end{aligned}$$

The $\Delta S(q,t)$ has been convoluted with the incoming pulse temporal profile (measured to ~ 1 ps resolution with a streak camera) :

$$\begin{aligned} \langle \Delta S(q,\tau) \rangle &= \int_{-\infty}^{+\infty} I(t')\, \Delta S(q, \tau - t') dt' \\ \Delta S(q, t - t' < 0) &= 0 \end{aligned}$$

χ^2 definition. The definition of the χ^2 of the model is following:

$$\begin{aligned} \chi^2 &= \sum_{j=\text{time delays}} \chi_j^2 \\ \chi_j^2 &= \sum_i \left(\frac{\langle \Delta S(q_i, \tau_j) \rangle - \Delta S_{exp}(q_i, \tau_j)}{\sigma_{i,j}} \right)^2 \end{aligned}$$

$\sigma_{i,j}$ is the error on the experimental curves calculated as the standard deviation of the different repetitions of the same time delay.

Least-Squares Fit. The least-squares fit of the model against the experimental data was done using the minimization package MINUIT written at CERN (S20). The quantity minimized is χ^2 as defined above. The errors of the fitted parameters have been calculated by MINUIT and they represent 1 standard deviation; refer to the MINUIT manual for further informations. After the global χ^2 was minimized, we extracted the partial χ^2 (the χ_j^2 in the previous section) as a function of time to have a fine probe of the different models. The populations of the different states vs time calculated with a starting set of fitting parameters are shown in Figure S9. This essentially includes a very fast formation of I_2+C_2H_4 and of bridge + I (ps time scale), a two fold relaxation of the bridge structure via formation of the isomer (C_2H_4I-I) and decay into C_2H_4+I (on the ns time scale); finally a slower (hundreds of nanosecond) formation of I_2 via recombination of free atomic iodine and relaxation of the isomer.

In this approach we have direct access to the reaction pathways by controlling the reaction constants. The same plot *after* the fitting procedure is shown in Figure S10 and highlights the reaction true pathways. The fit results for a few selected time delays (100 ps, 10 ns, 1 us) are shown in Figure S11, where all components contributing to each curve are shown for both q-and r-space.

This result suggests (as discussed in the main text) the following. Firstly, the formation of the bridge structure occurs with very high yield (no C_2H_4 + I_2 or C_2H_4+I+I formation). Secondly, the main decay pathway of the bridge structure is the formation of the *isomer*, C_2H_4I-I . Thirdly, on the slower time scale (μs), the isomer decay into C_2H_4 + I_2. According to our calculation (S18), the isomer has a lower energy than the dissociated products; however this difference is only 4.9 kJ/mol ($\sim 2\, k_B T$) and there are no energy barriers for this dissociation, which indicates that this final stage is entropy driven.

Supporting figures

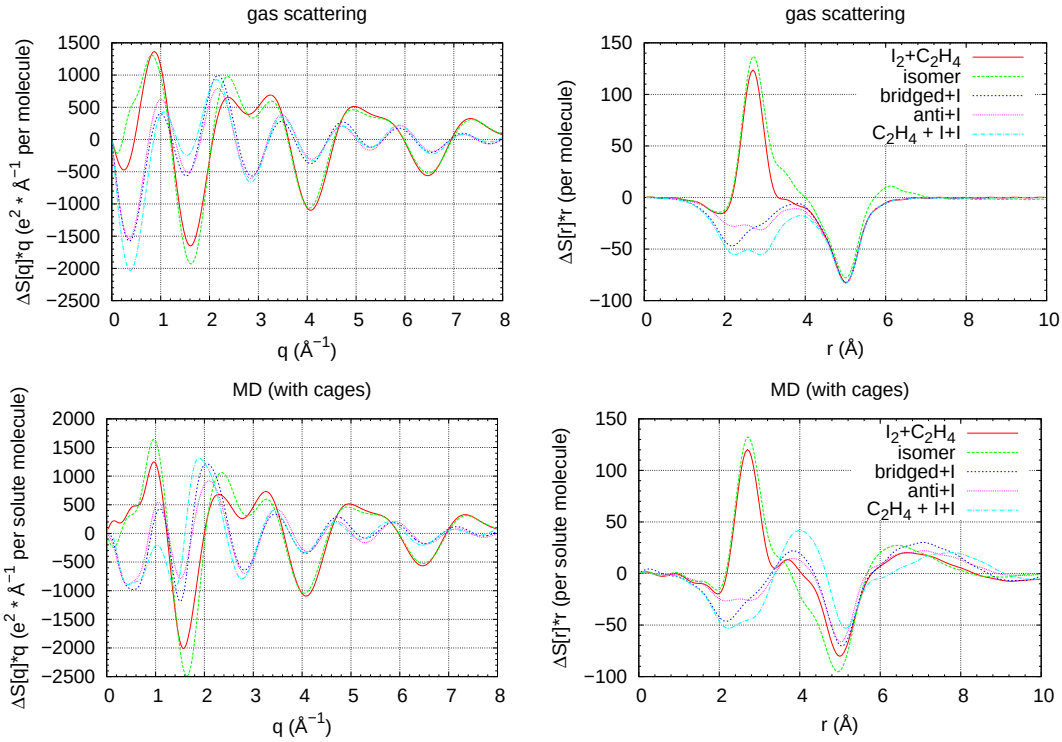

Figure S1: Influence of the solute–solvent cross terms in the difference maps.

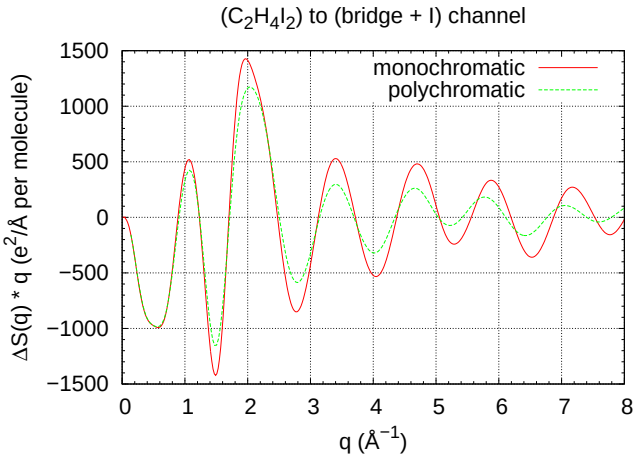

Figure S2: Typical effect of the polychromatic beam on the difference maps; note the shift of the isosbestic points, especially at high angles.

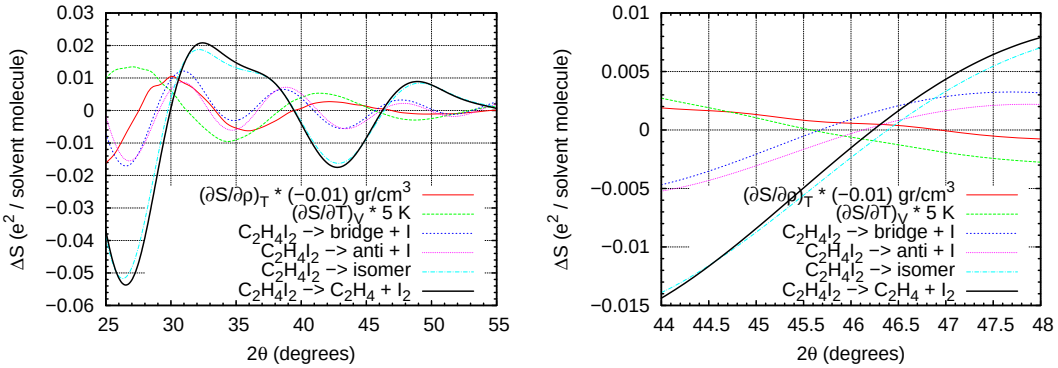

Figure S3: Difference curves for the contributing components. The plots have been scaled to 2θ using $\lambda = 0.71$Å. Numerical ratio Molecule/Solvent = 1/500, excitation 10%

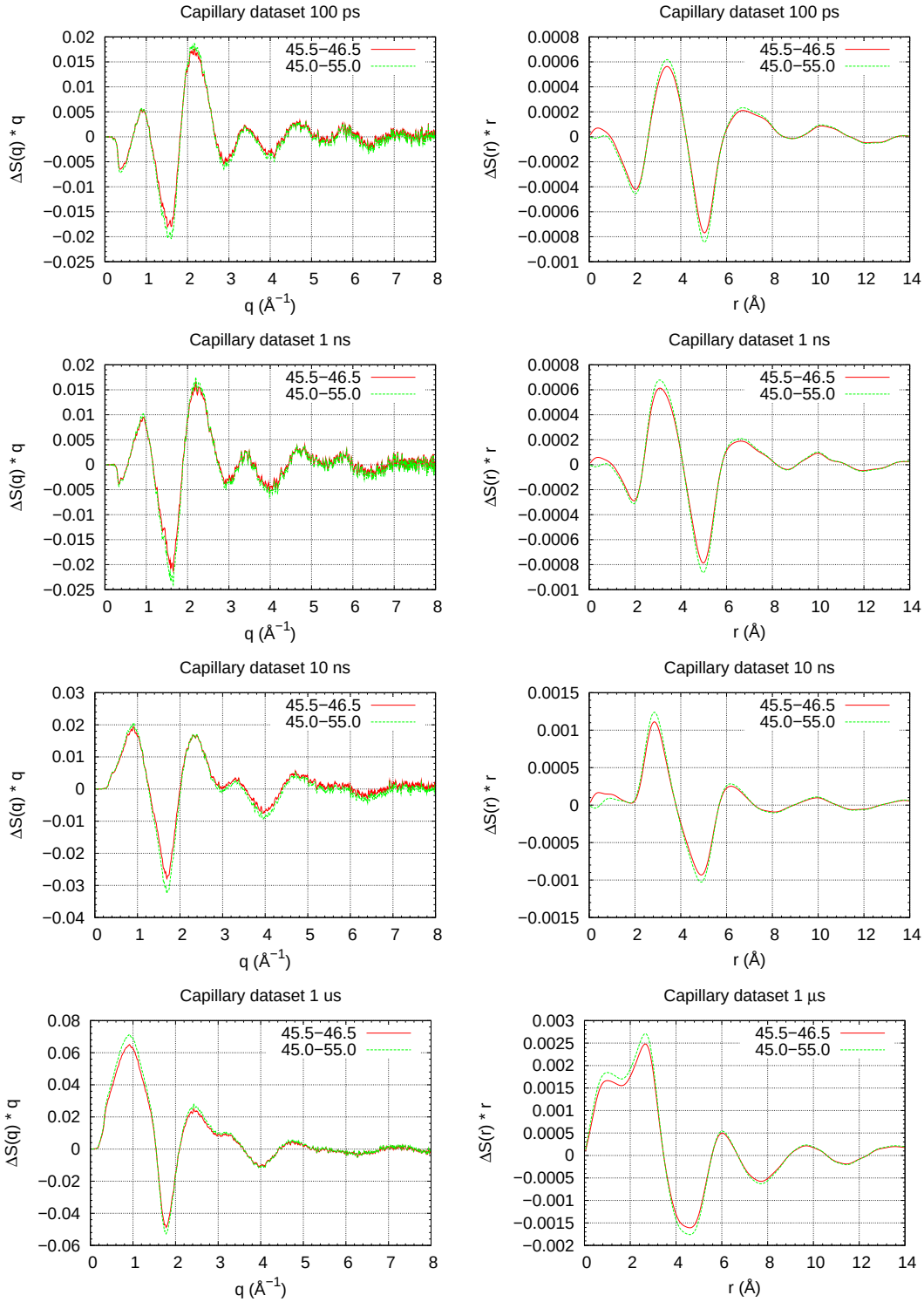

Figure S4: Influence of the normalization range on the calculation of the difference curves

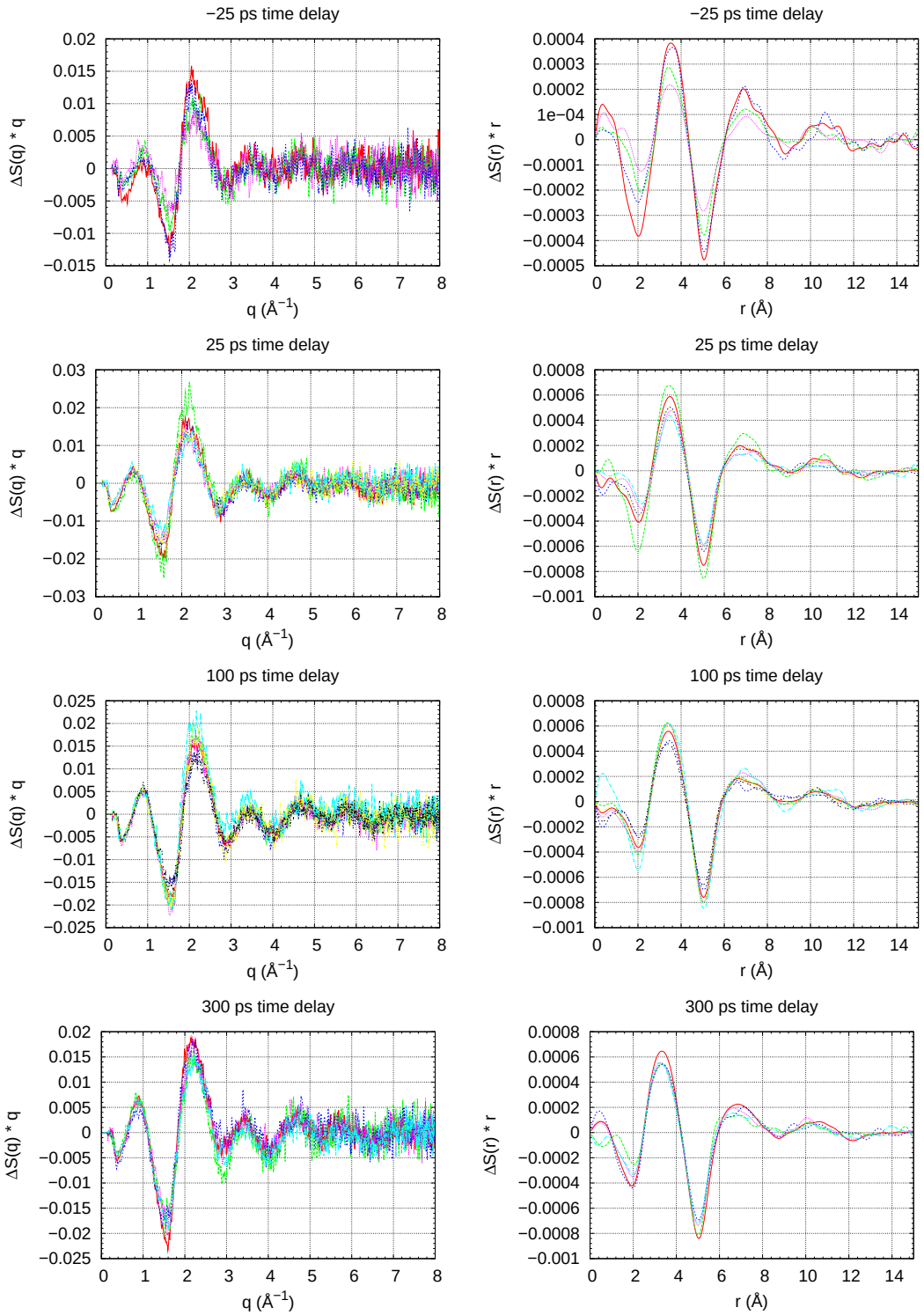

Figure S5: Comparison between different *sub–run*s

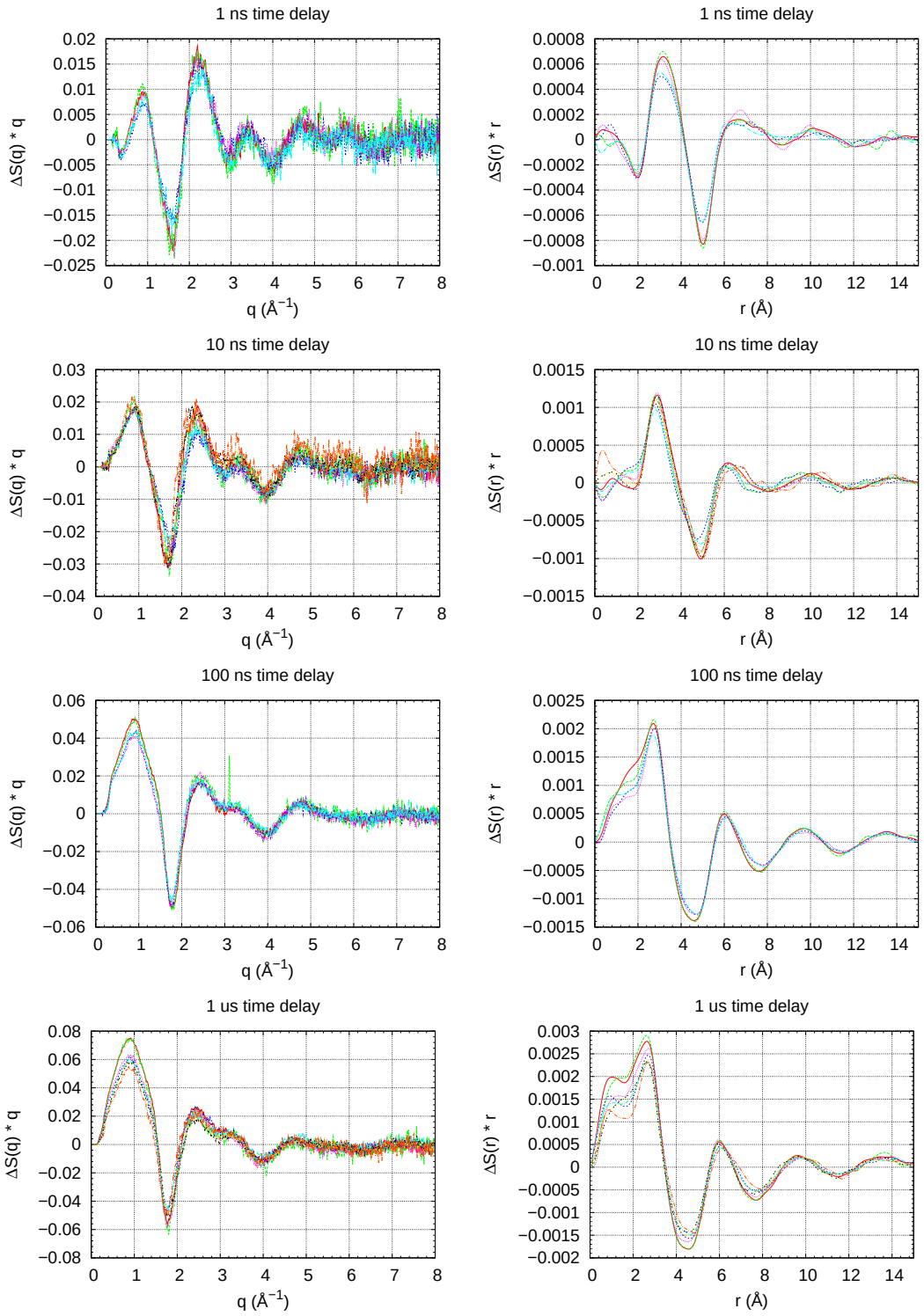

Figure S6: Comparison between different *sub–run*s

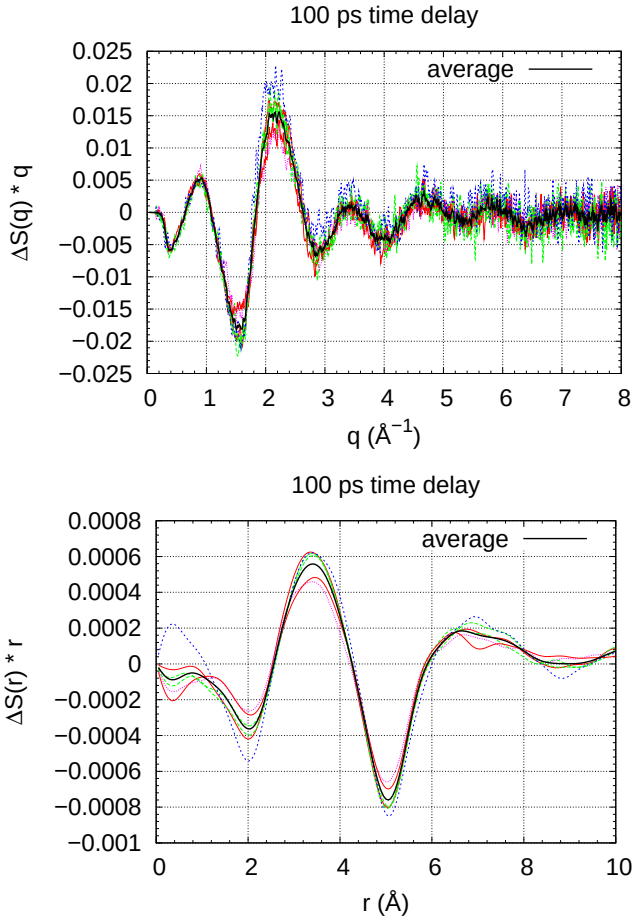

Figure S7: Data from all runs and their average at 100 ps delay.

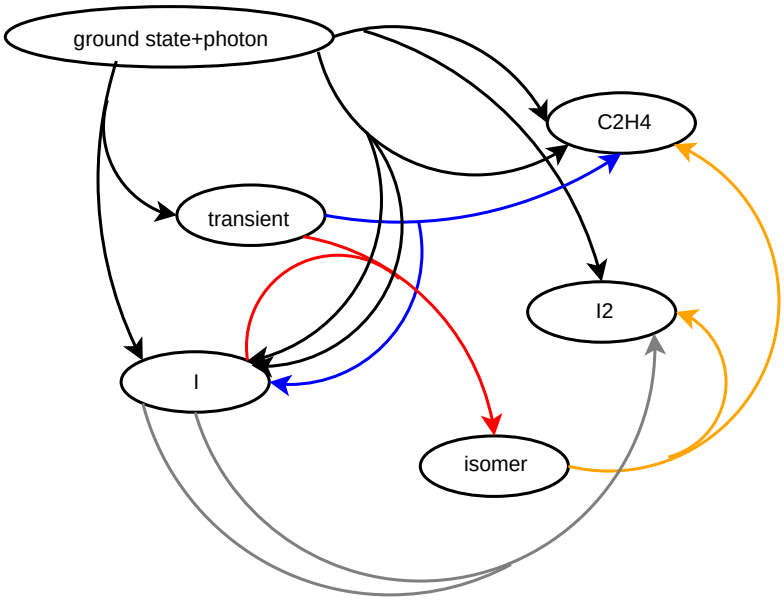

Figure S8: Reaction channels considered in the model

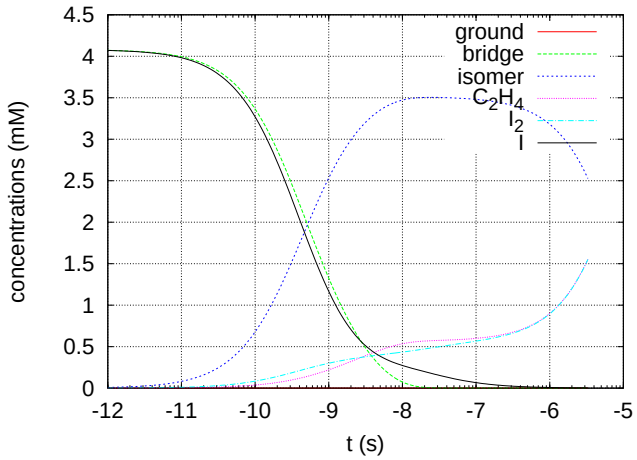

Figure S9: Concentration of the different molecules as a function of time with a possible set of starting parameters

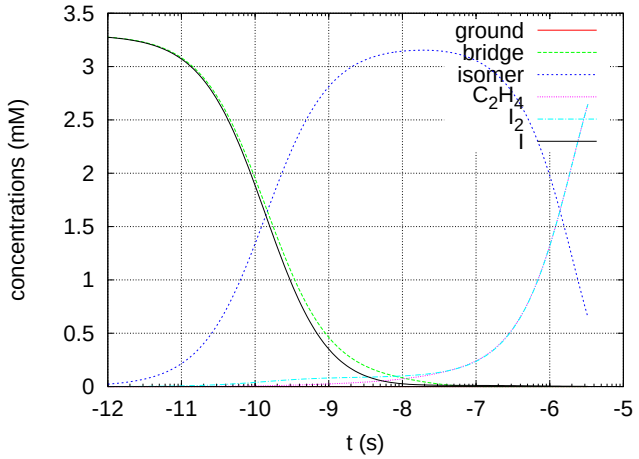

Figure S10: Population vs time obtained after the fitting procedure

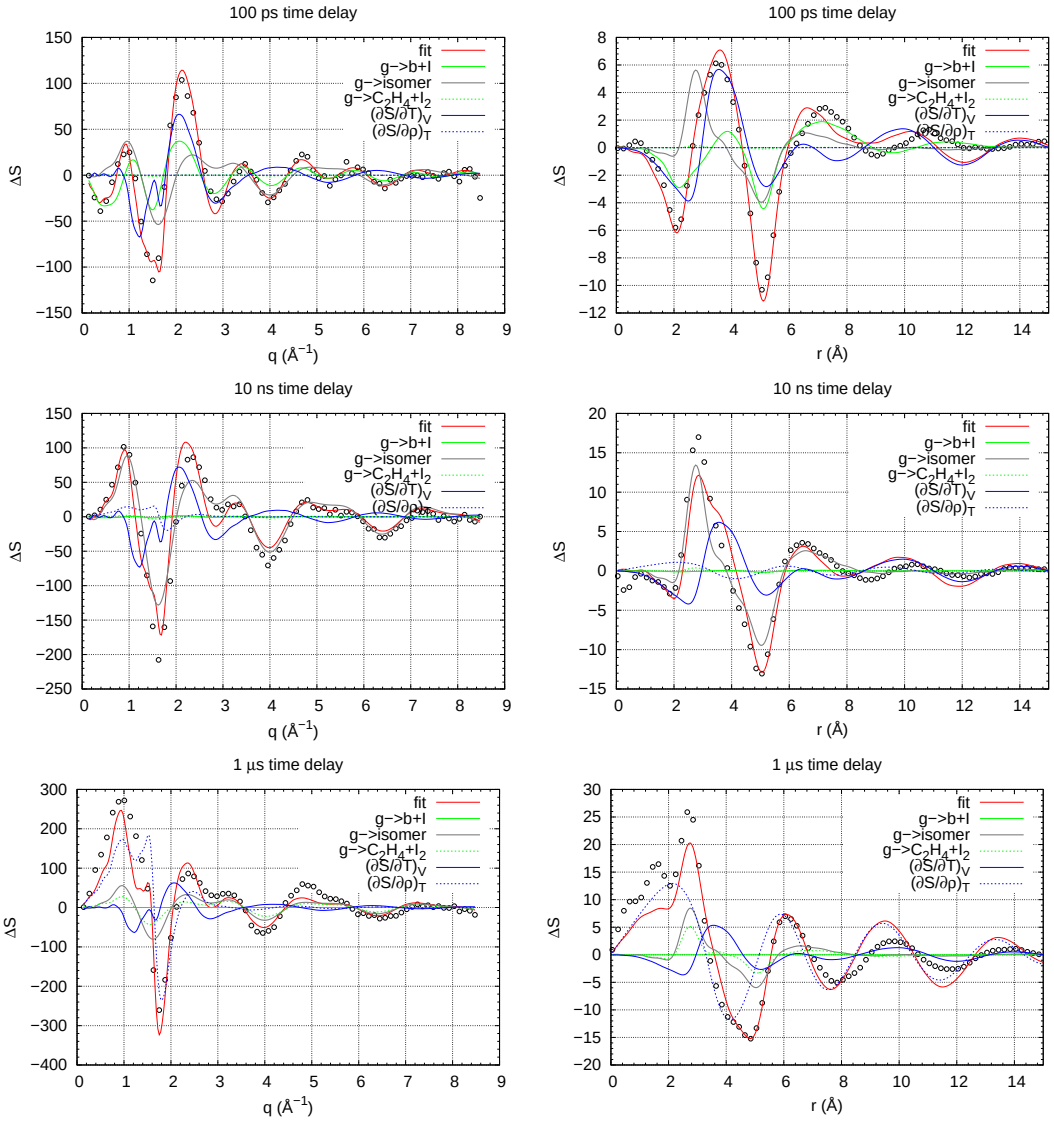

Figure S11: Different contributions to the total difference map.

Supporting References

(S1) R. Neutze, et al., Phys. Rev. Lett. **87**, 195508 (2001).

(S2) A. Plech, et al., Phys. Rev. Lett. **92**, 125505 (2004).

(S3) M. Wulff, Femtochemistry and Femtobiology: Ultrafast Events in Molecular Science, M. M. Martin, J. T. Hynes, eds. (Elsevier, 2004), 337–347.

(S4) M.J. Frisch, G.W. Trucks, Gaussian 94, P. PA, ed. (Elsevier, 1995).

(S5) A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

(S6) C. Lee, W. Yang, R. G. Parr, Phys. Rev. B **37**, 785 (1998).

(S7) A. Stromberg, O. Gropen, U. Wahlgren, J. Comput. Chem. **98**, 5648, (1983).

(S8) M. N. Glukhovtsev, A. Pross, J. Chem. Phys. **103**, 1878 (1995).

(S9) J. B. Foresman, T. A. Keith, K. B. Wiberg, J. Snoonian, M. J. Frisch, J. Phys. Chem. **100**, 16098 (1996).

(S10) K. Refson, Comp. Phys. Comm. **126**, 309 (2000).

(S11) M. Haughney, M. Ferrario, I. R. McDonald, J. Phys. Chem. **91**, 4934 (1987).

(S12) F. Mirloup, S. Bratos, R. Vuilleumier, M. Wulff, A. Plech, Femtochemistry and Femtobiology: Ultrafast Events in Molecular Science, M. M. Martin, J. T. Hynes, eds. (Elsevier, 2004), pp. 349-352.

(S13) W. Hoover, Phys. Rev. A **31**, 1695 (1985).

(S14) M. Berthaut, J. Phys. Rad. **13**, 499 (1952).

(S15) F. Mirlopup, Ph.D. thesis, University Paris VI (2004).

(S16) S. Bratos, F. Mirloup, R. Vuilleumier, M. Wulff, J. Chem. Phys. **116**, 10615 (1992).

(S17) Cernlib, First-order Differential Equations [Gragg-Bulirsch-Stoer], http://wwwasdoc.web.cern.ch/wwwasdoc/shortwrupsdir/d201/top.html.

(S18) Q. Kong, in preparation (2005).

(S19) J. Burkardt, Source Codes in Fortran 90, http://www.csit.fsu.edu/ burkardt/f_src/f_src.html.

(S20) CERN, Minuit; Function Minimization and Error Analysis, http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html.