The origin of efficient triplet state population in sulfur-substituted nucleobases

Sebastian Mai1, Marvin Pollum2, Lara Martínez-Fernández3, Nicholas Dunn2, Philipp Marquetand1, Inés Corral3, Carlos E. Crespo-Hernández2 & Leticia González1

Elucidating the photophysical mechanisms in sulfur-substituted nucleobases (thiobases) is essential for designing prospective drugs for photo- and chemotherapeutic applications. Although it has long been established that the phototherapeutic activity of thiobases is intimately linked to efficient intersystem crossing into reactive triplet states, the molecular factors underlying this efficiency are poorly understood. Herein we combine femtosecond transient absorption experiments with quantum chemistry and nonadiabatic dynamics simulations to investigate 2-thiocytosine as a necessary step to unravel the electronic and structural elements that lead to ultrafast and near-unity triplet-state population in thiobases in general. We show that different parts of the potential energy surfaces are stabilized to different extents via thionation, quenching the intrinsic photostability of canonical DNA and RNA nucleobases. These findings satisfactorily explain why thiobases exhibit the fastest intersystem crossing lifetimes measured to date among bio-organic molecules and have near-unity triplet yields, whereas the triplet yields of canonical nucleobases are nearly zero.
Chemically modified nucleobases provide DNA and, in particular, RNA with additional functions that are only beginning to be discovered. Thiobases, where the oxygen atom of a carbonyl group is substituted by a sulfur atom, have been detected in transfer-RNA and the lack of some of these substituted nucleobases can lead to metabolic diseases in the mitochondria. Thiobases are also employed in DNA-based nanotechnology, in photocrosslinking studies, in pharmacology, and—due to their interesting photophysical properties—in photo-chemotherapy. In fact, thiobases have recently been demonstrated as effective photo-chemotherapeutic agents for the treatment of skin cancer in tissue biopsies as well as of bladder cancer in animal models, and are being developed for clinical applications.

The canonical nucleobases absorb primarily in the ultraviolet-C range (100–280 nm). In contrast, substitution of the oxygen atom by sulfur induces a significant red-shift in the absorption spectrum, making thiobases effective absorbers of ultraviolet-B (280–315 nm) and ultraviolet-A (315–400 nm) radiation. An intriguing observation is that, while canonical nucleobases efficiently return to the electronic ground state following electronic excitation, thiobases populate long-lived, reactive triplet states in near-unity yields—a property that is exploited in their photonic applications. The photostability of canonical nucleobases is ascribed to ultrafast relaxation mechanisms that convert the photon energy into vibrational motion. Although intense research activity has been aimed at determining time scales and relaxation pathways in thiobases by using time-resolved spectroscopic techniques and theoretical methods, there exists no physical explanation of the electronic and structural factors that favour the formation of reactive triplet states instead of ultrafast relaxation to the ground state.

In this work, we provide a rationale for the origin of efficient population of triplet states in thiobases versus the ultrafast recovery of the ground state in canonical nucleobases. We use 2-thiocytosine (2tC) as a testbed to understand how thionation leads to a strong stabilization of different parts of the excited-state potential energy surfaces (PESs), which favours intersystem crossing (ISC) to the triplet manifold over internal conversion to the electronic ground state following electronic excitation. Thiobases populate long-lived, reactive triplet states in near-unity yields—a property that is exploited in their photonic applications. The photostability of canonical nucleobases is ascribed to ultrafast relaxation mechanisms that convert the photon energy into vibrational motion. Although intense research activity has been aimed at determining time scales and relaxation pathways in thiobases by using time-resolved spectroscopic techniques and theoretical methods, there exists no physical explanation of the electronic and structural factors that favour the formation of reactive triplet states instead of ultrafast relaxation to the ground state.

In this work, we provide a rationale for the origin of efficient population of triplet states in thiobases versus the ultrafast recovery of the ground state in canonical nucleobases. We use 2-thiocytosine (2tC) as a testbed to understand how thionation leads to a strong stabilization of different parts of the excited-state potential energy surfaces (PESs), which favours intersystem crossing (ISC) to the triplet manifold over internal conversion to the electronic ground state. Not only is 2tC specifically relevant for (ISC) to the triplet manifold over internal conversion to the electronic ground state in canonical nucleobases. We use 2-thiocytosine (2tC) as a testbed to understand how thionation leads to a strong stabilization of different parts of the excited-state potential energy surfaces (PESs), which favours intersystem crossing (ISC) to the triplet manifold over internal conversion to the electronic ground state. Not only is 2tC specifically relevant for (ISC) to the triplet manifold over internal conversion to the electronic ground state in canonical nucleobases. We use 2-thiocytosine (2tC) as a testbed to understand how thionation leads to a strong stabilization of different parts of the excited-state potential energy surfaces (PESs), which favours intersystem crossing (ISC) to the triplet manifold over internal conversion to the electronic ground state. Not only is 2tC specifically relevant for (ISC) to the triplet manifold over internal conversion to the electronic ground state in canonical nucleobases. We use 2-thiocytosine (2tC) as a testbed to understand how thionation leads to a strong stabilization of different parts of the excited-state potential energy surfaces (PESs), which favours intersystem crossing (ISC) to the triplet manifold over internal conversion to the electronic ground state. Not only is 2tC specifically relevant for (ISC) to the triplet manifold over internal conversion to the electronic ground state in canonical nucleobases.

**Results**

**Photochemical mechanism of 2-thiocytosine.** A combination of steady-state and femtosecond transient absorption spectroscopy with ab initio computations and nonadiabatic dynamics is used to elucidate the excited-state dynamics of 2tC. Femtosecond transient absorption spectra (fs-TAS) were recorded exciting with UV-B (308 nm) and UV-A (321 nm) wavelengths, which correspond to the low-energy tail of the steady-state absorption spectrum (see Supplementary Fig. 1). According to calculations (see Supplementary Fig. 2, Supplementary Table 1 and Supplementary Note 1), the UV-A or UV-B wavelengths excite the lowest-energy \( \pi, \pi^* \) state (the S2 at the equilibrium geometry) or the tail of the second \( \pi, \pi^* \) state (S5), respectively. The TAS obtained after 308 nm excitation (Fig. 1, the TAS exciting at 321 nm is shown in Supplementary Fig. 3) shows an initial featureless rise in the absorbance covering the full spectral probe window from −360 fs until a time delay of −120 fs (that is, within the cross correlation of the pump and probe beams), when two distinct absorption maxima begin to emerge: one in the UV range (350–400 nm) and another in the visible range (around 580 nm). These absorption bands continue to grow until 320 fs, at which point the UV band reaches its maximum. As the UV band decays, the visible band continues to increase in amplitude and blueshifts to \( \lambda_{max} = 525 \, \text{nm} \) at 3.76 ps, with an apparent isosbestic point at 420 nm. After 3.76 ps, no significant changes in the absorption maxima are observed for a time delay of up to 20 ps (not shown), the maximum time delay investigated in this work.

The TAS data were subjected to a target and global fitting analysis. Two lifetimes (\( \tau_1 = 210 \pm 50 \, \text{fs} \) and \( \tau_2 = 480 \pm 60 \, \text{fs} \)) were needed to model the dynamics during the first 4 ps. Figure 2 shows the decay-associated spectra (DAS) and the fitted representative kinetic traces taken at the transient absorption maximum, that is, 355, 523 and 581 nm (see Supplementary Fig. 4 for the traces exciting at 321 nm).

In order to interpret the TAS, and to deduce the relaxation pathways associated with the obtained lifetimes, the deactivation pathways out of the initially excited \( S_2 (\pi, \pi^*) \) state were calculated using the MS-CASPT2 method (multi-state complete active space second-order perturbation theory). TAS were simulated (Fig. 3a, the corresponding data is given in Supplementary Table 2) from the excited-state minima (\( n, \pi^* \), \( n, \pi^* \), \( n, \pi^* \) and \( n, \pi^* \)) using a sum of Gaussians centred at the calculated transition energies, with the amplitudes proportional to the oscillator strengths. The \( \pi, \pi^* \) state absorbs strongly around 363 nm and the \( \pi, \pi^* \) state dominates the visible part of the spectrum with two transitions around 441 and 546 nm.
giving rise to a broad absorption band. In contrast, the two respective minimum geometries (MS-CASPT2, convoluted using Gaussians) are very similar, with transitions at 333 and 576 nm versus 333 and 574 nm, respectively.

**Figure 2 | Target and global analysis of the TAS.** (a) Decay associated spectra obtained by globally fitting the transient absorption data to a two-component sequential kinetic model (A and B) with a constant offset (C). (b) Kinetic traces and fitted functions from the global fit analysis, taken at representative probe wavelengths, as indicated with coloured dashed lines in (a).

**Figure 3 | Simulated absorption spectra of the excited states.** (a) Calculated absorption spectra from the given excited states at their respective minimum geometries (MS-CASPT2, convoluted using Gaussians with 0.7 eV full-width at half-maximum). (b) Linear combinations of the calculated spectra with the given proportions. (c) The experimental transient absorption spectra at the indicated time delays.

According to the calculations, the initial broad and featureless increase in absorption amplitude across the full spectral probe window until ~120 fs (see also DAS A in Fig. 2a) can be assigned to a linear combination of the singlet states with \( \pi, \pi^* \) characters. This assignment is further supported by dynamics simulations, which predict that a \( 1_{\pi, \pi^*}/1_{\pi, \pi^*} \) conical intersection (CoIn) can be reached within 160 fs upon excitation, from which the molecule populates simultaneously two singlet minima with \( 1_{\pi, \pi^*} \) and \( 1_{\pi, \pi^*} \) characters.

At a time delay of about ~80 fs, two distinct absorption bands begin to evolve, with maxima around 360 and 550 nm, respectively (Fig. 1 and DAS B in Fig. 2a). These absorption bands are assigned to a linear combination of the absorption from the \( 1_{\pi, \pi^*}, 1_{\pi, \pi^*}, 3_{\pi, \pi^*} \) and \( 3_{\pi, \pi^*} \) states (Fig. 3b). The UV absorption band matches most closely the spectrum computed for the \( 1_{\pi, \pi^*} \) state, thus the subsequent decay in this region is likely associated with a decay of \( 1_{\pi, \pi^*} \) population.

Finally, the TAS at a time delay of ca. 4 ps, which has an absorption band peaking at 525 nm (Fig. 1 and DAS C in Fig. 2a), compares satisfactorily with the spectrum calculated for the \( 3_{\pi, \pi^*} \) state, possibly with some contribution from the \( 3_{\pi, \pi^*} \) state (Fig. 3b).

On the basis of static MS-CASPT2 calculations, the relaxation pathways sketched in Fig. 4 can be postulated (see Supplementary Table 3 for the corresponding data). After one-photon absorption (308 or 321 nm) the \( 1_{\pi, \pi^*} \) state at 3.65 eV is populated (S2 at the Franck–Condon geometry). A barrierless pathway leads to the \( 1_{\pi, \pi^*}/1_{\pi, \pi^*} \) CoIn at 3.02 eV, which connects to two different singlet state minima with \( 1_{\pi, \pi^*} \) (3.02 eV) or \( 3_{\pi, \pi^*} \) (2.95 eV) character. From these two minima, several relaxation pathways are plausible. Relaxation to the ground state involves overcoming energy barriers of at least 0.8 eV, making this process unlikely. Instead, two singlet-triplet \( (1_{\pi, \pi^*}/1_{\pi, \pi^*}) \) and \( 3_{\pi, \pi^*}/3_{\pi, \pi^*} \) minimum-energy crossing points (MECPs) located at 3.05 and 3.08 eV, respectively, are accessible due to their small energy barriers of at least 0.8 eV, making this process unlikely.
barriers and associated spin–orbit couplings of up to 170 cm\(^{-1}\). Energetically and geometrically close to these MECPs lies a CoIn between the \(^3\nu_5\nu^*\) and \(^3\pi\pi^*\) states, at an energy of 3.03 eV. Hence, there exists a three-state near-degeneracy region involving the lowest singlet and the two lowest triplet states. From the \(^3\nu_5\nu^*/^3\pi\pi^*\) CoIn, the system can relax either to the \(^3\pi\pi^*\) minimum at 3.02 eV or to the \(^3\nu_5\nu^*\) minimum at 2.85 eV.

A time-resolved interpretation of the relaxation mechanism of 2tC is obtained by performing surface-hopping nonadiabatic simulations using the SHARC method (surface hopping including arbitrary couplings)\(^{38}\). Associated energies, gradients and relevant couplings are obtained at the multi-reference configuration interaction including single excitations (MRCIS) level of theory (see Supplementary Fig. 5 and Supplementary Note 2 for validation of the method). Figure 5 shows the excited-state populations based on 137 trajectories and the result of monoeponential fitting. The relaxation pathways in the dynamics simulations involve mainly four electronic states—labelled according to their energetic ordering at each geometry as \(S_1\), \(S_2\), \(T_1\) and \(T_2\)—as foreseen by the quantum chemical calculations. These states encompass the \(^3\nu_5\nu^*\), \(^1\nu_5\nu^*\), \(^3\pi\pi^*\) and \(^3\pi\pi^*\) characters, but note that due to state crossing and mixing, no one-to-one correspondence between the \(S/T\) labels and the characters exists beyond the Franck–Condon geometry.

After excitation, the population in the \(S_2\) state decays to the \(S_1\) state with a lifetime of 160 fs (Supplementary Table 4)—a value that might be slightly overestimated due to a small barrier to reach the \(S_2/S_1\) crossing (0.13 eV), which is present with MRCIS but not with the more accurate MS-CASPT2 method (Fig. 4). The population of the \(S_1\) state rises to a maximum of about 30% during the first 200 fs due to population transfer from the \(S_2\) state. Afterwards, 2tC undergoes ISC to the triplet manifold with an average time constant of \(~250\) fs. The population of the \(S_1\) state amounts to only 13% after 1 ps.

The most important ISC pathways predicted by the dynamics simulations are \(S_1\rightarrow T_2\) and \(S_1\rightarrow T_1\), with a minor contribution from \(S_2\rightarrow T_2\). The characters of the triplet states \(T_1\) and \(T_2\) are strongly mixed and cannot be assigned as strictly \(^3\pi\pi^*\) or \(^3\nu_5\nu^*\) during the dynamics simulations. It is precisely this mixed wavefunction character that leads to large spin–orbit couplings between \(S_1\) and both triplet states, explaining why both \(S_1\rightarrow T_1\) and \(S_1\rightarrow T_2\) transitions are possible, and why ISC is very efficient. According to the simulations, \(T_2\) is populated faster (in \(~100\) fs) than \(T_1\) (in \(330\) fs) because the average \(S_1\rightarrow T_2\) energy gap is smaller than the \(S_1\rightarrow T_1\) gap. Note that the spin–orbit couplings vary in these regions but are on average \(~160\) cm\(^{-1}\) between \(S_1\) and \(T_2\), as well as \(50\) cm\(^{-1}\) between \(S_1\) and \(T_1\). However, the \(T_1\) state receives most of the population within 1 ps due to subsequent \(T_2\rightarrow T_1\) internal conversion, leading to an average ISC lifetime of 250 fs. This value is in full agreement with the lifetime of 210 ± 50 fs obtained experimentally. After 1 ps, 74% of the total population resides in either of the triplet states, and the remaining 16% of the population in the excited singlet manifold likely undergoes ISC at later times. A small fraction of the trajectories (10%) returned to the ground state because MRCIS underestimates the energy of the relevant \(S_1/S_0\) CoIns, as compared with MS-CASPT2 (Supplementary Table 3). This ground-state repopulation should be regarded as an artefact of employing MRCIS in the dynamics simulations and we therefore estimate a >90% total ISC yield from the simulations.

From a structural point of view, there are no large out-of-plane deformations during the dynamics as anticipated from the similarity and quasi-planar nature of all excited-state minima (Supplementary Note 3). During the \(S_2\rightarrow S_1\) internal conversion, there is a bond inversion within the conjugated double bond system of the pyrimidine ring. As the trajectories relax to \(S_1\) and the triplet states, ring puckering increases, probably due to the increased available energy or due to a more flexible pyrimidine ring in these states. The triplet minima are structurally similar to the singlet minima and thus motion on the \(T_1\) and \(T_2\) PESs cannot be distinguished from motion on the \(S_1\) PES. In fact, the structural similarity is one of the factors leading to the efficient ISC observed in this work.

In summary, the experimental and theoretical results suggest that excitation of 2tC to the lowest-energy \(^1\pi\pi^*\) state (\(S_2\)) leads to very efficient ISC to the \(^3\pi\pi^*\) \((T_1)\) state in three general steps, which are schematically depicted in Fig. 6. The first step is the relaxation from the vertically excited \(S_2\) \((^1\pi\pi^*)\) state to the \(S_1\) state through an \(S_2/S_1\) CoIn. After transition from the \(S_2\) to the \(S_1\) state, the system populates the \(S_1\) PES, which in some regions of the configurational space has dominant \(^1\pi\pi^*\) character and in other regions \(^3\nu_5\nu^*\) character, according to MS-CASPT2 calculations. This is also consistent with the observation of a band assigned to the \(^1\pi\pi^*\) state in the TAS at short time delays. The second step is ultrafast ISC from the \(S_1\) state to the triplet manifold \((T_2\) and \(T_1\) states). In the time-resolved experiments, this process leads to the appearance of an absorption band assigned to the \(^3\pi\pi^*\) state. This band begins to appear at approximately the same time as the \(^1\pi\pi^*\) absorption band, indicating that internal conversion in the singlet manifold and ISC occur on very similar time scales. The experimental time constant \(\tau_1\) for this second step is 210 ± 50 fs, in close agreement to the ISC lifetime of 250 fs observed theoretically. As the calculations show, the three-state near-degeneracy between \(S_1\), \(T_1\) and \(T_2\), together with the large spin–orbit matrix elements, are responsible for the ultrafast ISC lifetime of 2tC. Finally, the third step is internal conversion from the \(T_2\) to the \(T_1\) state and subsequent vibrational energy redistribution to populate the lowest-energy minimum of the \(T_1\) \((^3\pi\pi^*)\) state. The second experimental time constant of 480 ± 60 fs is attributed to the latter step, together with residual ISC, as suggested by the decay of the UV transient absorption band. The eventual relaxation from the \(T_1\) state to the ground state appears to involve significant energetic barriers and does not play a role on the time scales.
considered in this work, in agreement with the long-lived nature of the $T_1$ state of thiobases in solution\textsuperscript{15}.

**Thiobases versus canonical bases.** Having revealed the electronic relaxation mechanisms of $2tC$, we are now in a favourable position to compare the structural and electronic features of the PESs in the thiobases and the canonical nucleobases, in order to address the intriguing question as to why they are so photo-physically different. To this end, Table 1 collects the energetic PESs in the thiobases and the canonical nucleobases, in order to position to compare the structural and electronic features of the thiobases versus canonical bases.

The first conspicuous fact is that sulfur-substitution significantly lowers the energies of the singlet state minima. Both $1\pi\pi^*$ and $1\pi\pi^*$ state minima are stabilized by up to 1 eV for all three thiobases ($2tC$, $2tU$, $6tG$) in comparison with the corresponding canonical nucleobase. This stabilization is well-known in thiocarbonyl compounds\textsuperscript{40} and thiobases\textsuperscript{15,30}. However, it is surprising that the energy of the most accessible (that is, lowest-energy) CoIn of each molecule stays approximately constant upon thionation ($3.8/3.8$ eV for $C/2tC$, $4.0/4.0$ eV for $U/2tU$ and $4.3/3.8$ eV for $G/6tG$). As a consequence, canonical nucleobases exhibit efficient ground-state recovery after excitation, whereas the population of the thiobases rapidly decays to deep singlet state minima. The presence of one or more triplet states with energies similar to the singlet state minima and with large spin–orbit couplings facilitates subsequent ISC in the thiobases. While the energetic stabilizations can be correlated directly with the excited-state behaviour of the thiobases, it is not straightforward to rationalize why thionation lowers the energies of the states’ minima so strongly, but has little impact on the energies of relevant CoIns.

We argue that a plausible explanation relates to the moiety of the molecule where the electronic excitation is localized. We focus on whether the excitation (that is, the excited electron or, more importantly, the hole left behind) is localized significantly on the (thio)carbonyl group or on the aromatic ring. States characterized by an excitation localized on the chalcogene atom are expected to stabilize considerably upon thionation because it takes less energy to excite an electron from an $n$ or a $\pi$ orbital of a sulfur atom into the $\pi^*$ system than from an $n$ or $\pi$ orbital of an oxygen atom (compare the lower ionization potential of S versus O)\textsuperscript{41}. Conversely, if the electronic excitation is localized primarily on the pyrimidine ring (that is, electronic excitation from pyrimidine $\pi$ to pyrimidine $\pi^*$) no stabilization is expected upon thionation. Using the TheoDOR program\textsuperscript{42}, the electronic wave functions were analysed to assign the localization of the relevant $S_n\rightarrow S_1$ excitations. The results are shown in Table 1 (the corresponding difference densities of the excitations can be found in Supplementary Fig. 6, the numeric data in Supplementary Table 5). Remarkably, the stabilization of the excited-state minima of the thiobases can now be easily explained. Since the $\pi\pi^*$ states are always localized on the carbonyl/thiocarbonyl group, the $n\pi^*$ state minima are stabilized strongly upon thionation by $\sim 1$ eV. Similar arguments explain why the $\pi\pi^*$ states stabilize, but to lesser extent because they are at least partially delocalized over the aromatic rings.

In order to rationalize the energy shifts of the CoIns upon thionation, we consider C and $2tC$ as the initial example. Several different $S_n/S_1$ CoIns have been reported for C (Table 1)\textsuperscript{43–49} These CoIns have different electronic character and hence we expect different localization of the excitations. The excitation within the semiplanar CoIn is localized on the chalcogen atom and thus stabilized in $2tC$. However, the ethylenic CoIn (which is the lowest-energy CoIn of C, and thus primarily responsible for its ultrafast relaxation) is not stabilized because its excitation is centred on the pyrimidine ring. The amino-out-of-plane (oop-NH$_2$) CoIn is stabilized slightly (by 0.3 eV). Although the semiplanar CoIn is also stabilized by 1.4 eV in $2tC$ and becomes the lowest-energy CoIn, it is still too high in energy compared with the $1\pi\pi^*$ minimum, precluding ultrafast internal conversion.

Table 1 | Relative energies of stationary points and crossings for $2tC$, $2tU$, $6tG$, and their parent nucleobases.

| Critical Point | Localization on O | $\Delta E$ (C ref. 43) | $\Delta E$ (2tC this work) | Localization on S | $\Delta E$ (G ref. 66) | $\Delta E$ (6tG ref. 25) |
|---------------|-------------------|-------------------------|---------------------------|-------------------|-------------------------|---------------------------|
| $1\pi\pi^*$ min | Yes | 4.2 | Yes | 3.0 | 4.3 | Yes | 3.1 |
| $1\pi\pi^*$ min | Partially | 4.0 | Partially | 3.0 | 4.1 | Yes | 3.8 |
| $S_n/S_1$ semiplanar | Yes | 5.2 | Yes | 3.8 | No | 4.0 | Yes | 4.0 |
| $S_n/S_1$ oop-NH$_2$ | No | 4.1 | Partially | 3.8 | No | 4.0 | Partially | 3.8 |
| $S_n/S_1$ ethylenic | No | 3.8 | No | 3.8 | No | 3.8 | No | 3.8 |
| Behaviour | Relaxation (2–3 ps (refs 67,68)) | ISC (210 ± 50 fs) | | | Relaxation (2.5 ps (refs 68,69)) | ISC (360 ± 30 fs (ref. 12)) | |
| $1\pi\pi^*$ min | Yes | 4.1 | Yes | 3.3 | 4.3 | Yes | 3.1 |
| $1\pi\pi^*$ min | No | 4.5 | Yes | 3.8 | No | 4.0 | Yes | 3.8 |
| $S_n/S_1$ oop-Y (Y = O/S) | Yes | 5.8 | Yes | 4.0 | No | 4.0 | Partially | 3.8 |
| $S_n/S_1$ ethylenic | No | 4.0 | No | 3.8 | No | 4.3 | No | 3.9 |
| Behaviour | Relaxation (2.5 ps (refs 68,69)) | ISC (360 ± 30 fs (ref. 12)) | | | Relaxation (360 fs (ref. 70)) | ISC (310 ± 50 fs (ref. 20)) | |

All energies ($\Delta E$) are given in eV above the ground state minimum of the respective molecule. The classification whether the $S_n\rightarrow S_1$ excitation is localized on the chalcogen atom (Y = O/S) is based on the density plots presented in Supplementary Fig. 6 and Supplementary Table 5. These plots were calculated at the CASSCF(14,10) level of theory (CASSCF(18,13) for G and 6tG).

*Not reported in the literature, but presumed to have no localization on S in analogy to the equivalent CoIn in the parent nucleobase.
to the ground state and favouring instead rapid decay to the $^1\text{nt}^*$ minimum and subsequent ISC to the triplet manifold.

Satisfactorily, a similar rationale applies to the $\text{U/2tU}$ pair. According to calculations, $\text{U}$ relaxes primarily through the ethylenic CoIn$^{29-33}$, with a twist at the carbon-carbon double bond, whereas the oxygen-out-of-plane (oop-O) CoIn$^{26,52}$ is much higher in energy. In 2tU, on the other hand, the oop-S CoIn is stabilized, whereas the ethylenic CoIn is much higher in energy$^{26,24}$. The stabilization of the oop-S CoIn is, however, not sufficient enough to provide a fast relaxation channel to the ground state$^{24}$. To our knowledge, there are no calculations on 2-thiothymine, but we expect that similar arguments hold for the thymine/2-thiothymine pair, because the methyl group is not expected to affect the electronic structure significantly.

The CoIns in the G/6tG pair also follow the discussed trend: the oop-S CoIn is stabilized more than the oop-NH$_2$ CoIn—although the effect is less pronounced than in the pyrimidine bases. Ground-state relaxation from the $^1\pi^*$ minimum in 6tG would be energetically possible, but the shape of the PES favours the population of the $^1\text{nt}^*$ minimum instead, from where all possible deactivation pathways to $S_0$ involve very large energy barriers.

From the above analysis, it can be proposed that thionation of the nucleobases stabilizes all states in which the electronic excitation is primarily localized on the sulfur atom (for example, $n\pi^*, \pi\pi^*$), whereas the other states ($n\pi^*, \pi\pi^*$), and their corresponding CoIns, are not stabilized to the same extent. This leads to the conclusion that the intrinsic relaxation pathways in the canonical nucleobases—which involve extensive ring puckering but no extensive motion of the carbonyl group—are blocked in the thiobases because they are too high in energy relative to the considerably stabilized excited-state minima. Although there exist S/S$^+_s$ CoIns involving sulfur-localization that are also stabilized$^{25,29,54}$, they are too high in energy with respect to the excited-state minima and do not play a role in the deactivation.

Finally, localization of the excitation on the sulfur atomic orbitals also enhances spin–orbit coupling. This is because the spin–orbit operator is of short range nature$^{55}$ and depends on the atomic number, thus making this coupling largest if the excitations are localized on the heaviest atom. Further enhancement of population transfer from the singlet to the triplet manifold is due to the smaller singlet-triplet energy gaps in the thiobases compared with the canonical nucleobases because heavy atom substitution often leads to an increased density of states.

**Discussion**

Using transient absorption spectroscopy and theoretical calculations, we demonstrated that photoexcited 2-thiocytosine populates the triplet manifold on a femtosecond timescale and with close-to-unity triplet yield. Its relaxation after excitation proceeds in three general steps, which partially overlap: (i) barrierless deactivation from the $S_2$ to the $S_1$ state; (ii) ISC to the triplet manifold with a time constant of $210 \pm 50$ fs, enhanced by the presence of a three-state near-degeneracy region involving the $S_1$, $T_2$ and $T_1$ states; and (iii) internal conversion to the $T_1$ state. The latter step, together with some residual ISC is responsible for a time constant of $480 \pm 60$ fs. More importantly, we have rationalized and generalized the influence of thionation on the excited-state dynamics of the canonical nucleobases. While it is generally recognized that thionation leads to a strong decrease in excited-state energies, here we show that the extent of this stabilization differs significantly in different parts of the PESs. Furthermore, we have shown that the stabilization depends sensitively on the electronic wavefunction character and favours wave functions where the excitation is localized on the sulfur atom. Hence, excited-state minima (for example, $^1n\pi^*$) and conical intersections involving the thiocarbonyl group are stabilized upon thionation, whereas the conical intersections intrinsic to the pyrimidine skeleton are not. This observation explains satisfactorily the remarkably different behaviour of thiobases compared with their parent nucleobases and provides design criteria for the synthesis of prospective photochemotherapeutic drugs.

**Methods**

**Materials.** 2-Thiocytosine (97%, Sigma-Aldrich), sodium dihydrogenphosphate (Sigma-Aldrich), and sodium hydrogenphosphate (99.0%, Sigma-Aldrich) were used as received. PBS solutions were prepared using 0.15 g of sodium dihydrogenphosphate and 0.27 g of sodium hydrogenphosphate dissolved in 200 ml of ultrapure water to give pH 7.4 aqueous phosphate buffer with a total phosphate concentration of 16 mM.

**Femtosecond transient absorption spectroscopy.** Femtosecond broadband transient absorption spectroscopy was performed using a Libra-HF pulsed laser system (Coherent, Inc.), which produces 800 nm, 100 fs pulses with a total output power of $4 \text{ W}$. A small fraction of this fundamental beam ($< 2 \text{ mW}$) was split off and focused into a continuously moving 2 mm CaF$_2$ crystal to generate the white light continuum probe pulses (320–710 nm). The remainder of the fundamental beam was directed through an optical parametric amplifier (TOPAS-800, Light Conversion, Ltd.) to generate the excitation (that is, pump) pulses at either 308 or 321 nm. All other wavelengths were removed from the excitation beam using a reflective wavelength filter and a Glan–Taylor polarizer.

The spectrometer (Helios, Ultrafast Systems, LLC.) uses an optical delay line to delay the probe pulses in time with respect to the pump. The pump power was attenuated to 1 $\mu$W at the sample and its polarization was randomized using a depolarizing plate in order to prevent contribution of rotational effects to the dynamics. At each delay time the spectrum of the probe is collected with and without pumping the sample in order to produce the difference spectra (AA). The excitation power at 321 nm was adjusted to $0.5 \pm 0.1 \text{ W}$. Absorbance data were prepared from the concentrations of 5 µM of various PBS. All transient absorption experiments were performed in 2 mm optical path length quartz cuvettes (Starna Cells, Inc.). Sample solutions were stirred continuously throughout data collection and replaced with fresh sample, from the same stock, every 5–7 scans so that all scans had no more than 5% degradation as determined by the steady-state absorption. A home-made LabView program (National Instruments, Inc.) was used to correct all transient absorption data for group velocity dispersion of the white light probe$^{35}$. Global and target analysis of the data were performed using a sequential kinetic model$^{32}$, convoluted with the instrument response function fixed at 200 fs, as determined by the coherence signal of neat methanol$^{32}$.

The sequential kinetic model included two lifetimes: $A \rightarrow B \rightarrow C$. (1)

The spectrum of the third component C can be considered as an offset and represents the transient species that has not decayed during the first 20 ps investigated in this work.

The results of five separate experiments, three exciting at 308 nm and two exciting at 321 nm, were analysed. The target and global fitting analysis was performed independently for each of the five experiments in order to estimate the uncertainty of the lifetimes. Up to 91 kinetic traces from the broadband transient absorption data were used in this analysis. The uncertainties are reported as twice the standard deviation. The DAS are extracted from the global fit of the data and correspond to the wavelength-dependent amplitudes for each of the lifetime components in the kinetic traces.

**Stationary quantum chemistry.** According to Kasha’s rule, the relaxation from $S_0$ via $S_1$ to $S_2$ should be ultrafast. This is in agreement with the experimental fs-TAS results, which showed that excitation at 321 nm (mainly to $S_1$) and excitation at 308 nm (which also excites $S_2$) induce the same excited-state dynamics. Hence, the computational investigations focused on the relaxation dynamics starting from $S_2$. Minima of the PESs were first optimized at the CASSCF(14,10)/ANO-S-VDZP level of theory, whereas for the optimization of CoIns and MECPs CASSCF(14,10)/6-31G* was used. Using CASSCF(14,10)/ANO-S-VDZP, all minima and crossing points were connected with minimum-energy paths to check for the absence of energetic barriers. If a minimum-energy path successfully finishes from a surface crossing to a minimum, it implies that there are no energetic barriers in between. The minima, CoIns and MECPs were then reoptimized at the MS-CASPT2/ANO-L-VTZP level of theory using the methodologies described in refs. 57, 58 and the ORCA optimizer$^{35}$. For all MS-CASPT2 calculations, an imaginary level shift of 0.3 a.u. and the default IPEA shift of 0.25 a.u. has been used. All calculations were conducted in the gas phase and using MOLCAS 7.8 (ref. 60), except for the
CASSCF optimizations of the Cols and MECPs, where MOLPRO 2009 (ref. 61) was used.

The levels of theory used in the various calculations, including the numbers of averaged states, are also compiled in Supplementary Table 6. The active space or reference space orbital composition is presented in Supplementary Fig. 7. All calculations were performed with the aminothion tautomers of 2-thiocytosine, which is justified in Supplementary Fig. 2 and Supplementary Table 7.

All optimized geometries are reported in Supplementary Note 3. While MS-CASPT2 is, in many cases, an excellent method to describe excited states, it cannot be yet routinely employed in dynamics simulations, due to the lack of efficient implementations of analytical MS-CASPT2 nuclear gradients. Hence, all stationary point electronic structure calculations were performed with MRCI and within the SA-CASSCF calculations. The CAS(6,5) has been created removing the least occupied π*-orbital and the most occupied two π-orbitals from the CAS(10,8), and therefore the reduced reference space still contains all orbitals involved in the primary excitations of all the relevant states. Inner shells were kept frozen in the MRCI stage of the calculations. All MR-CIS calculations were performed with COLUMBUS 7 (ref. 62), with integrals from MOLCAS60, which also provided scalar-relativistic one-electron integrals according to the Douglas–Kroll–Hess method and the spin–orbit integrals using atomic mean-field integrals 39,40.
The electronic Schrödinger equation was integrated with a time step of 0.02 fs using the local diabatization methodology 63, which employs wavefunction overlaps 64 to provide the information about the nonadiabatic couplings. An energy-based decoherence correction 65 was applied to the electronic states in the diagonal basis. As stated above, the simulations were carried out on spin-adiabatic potentials, obtained by diagonalization of the spin–orbit Hamiltonian including states S1 to S5 and T1 to T2. However, for the analysis of the results the state populations were transformed back into the spin-free basis, since this allows for easier interpretation of the results. Note that the energies of S1, T1 and T2 (but no gradients, spin–orbit couplings or wavefunction overlaps) were also calculated at each time step in order to scrutinize whether neglecting these states is justified. Furthermore, since the S1 population decays with only 20% to the S2 state, for all analysis and discussion purposes the S2 population was merged into the S1 population.

Wave function localization analysis. The calculations were performed using optimized geometries reported in the literature for cytosine 48, uracil 52, guanine 46, 2-thiouracil 48 and 6-thioguanine 25,29, whereas for 2-thiocytosine the geometries reported here were used: for each molecule, the nπ* and ππ* excited states were minimized and relaxed. Cols and MECPs are displayed. At each geometry, a CASSCF/ANO-L calculation was performed, averaging over 4 singlet states. For cytosine, uracil, 2-thiocytosine, and 2-thiouracil, a CAS(14,10) was used, whereas for guanine and 6-thioguanine a CAS(18,13) was employed. These calculations were performed with MOLCAS60. The one-electron difference density matrices between the ground state and the first excited state were then analysed using the TheoDORe program 42 in terms of the Mulliken detachment populations in order to arrive at the classification in Table 1. The Mulliken populations are given in Supplementary Table 5.

Data availability. All data is available from the authors upon request.

References
1. Carell, T. et al. Structure and function of noncanonical nucleobases. Angew. Chem. Int. Ed. Engl. 51, 7110–7131 (2012).
2. Carbon, J., David, H. & Studier, M. H. Thiobases in Escherichia coli transfer RNA: 2-thiocytosine and 5-methylaminomethyl-2-thiouracil. Science 161, 1146–1147 (1968).
3. Zhang, S., Blain, J. C., Zielinska, D., Gryaznov, S. M. & Szostak, J. W. Fast and accurate nonenzymatic copying of an RNA-like synthetic genetic polymer. Proc. Natl Acad. Sci. USA 110, 17732–17737 (2013).
4. Wei, F.-Y. et al. Deficit of rRNA synthesis modification by cdk1 affects the development of type 2 diabetes in mice. J. Clin. Invest. 121, 3598–3608 (2011).
5. Wojciechowski, F. & Leumann, C. J. Alternative DNA base-pairs: from efforts to expand the genetic code to potential material applications. Chem. Soc. Rev. 40, 5669–5679 (2011).
6. Fricke, A., Saintome, C., Fourrey, J.-L., Clivo, P. & Laugà, P. Thionucleobases as intrinsically photofinestable of nucleic acid structure and nucleic acid-protein interactions. J. Photochem. Photobiol. B 42, 109–124 (1998).
7. Wilhelmsson, L. M. Fluorescent nucleic acid base analogues. Q. Rev. Biophys. 43, 159–183 (2010).
8. Perigaud, C., Gosselin, G. & Imbach, J. L. Nucleoside analogues as chemotherapeutic agents: a review. Nucleosides Nucleotides 11, 903–945 (1992).
9. Jordan, L. P., Duran, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. Nat. Rev. Drug Disc. 12, 447–464 (2013).
10. Attard, N. R. & Karran, P. UVA photosensitization of thiopurines and skin cancer in organ transplant recipients. Photochem. Photobiol. Sci. 11, 62–68 (2012).
11. Pridgeon, S. W. et al. Thiopyrimidine combined with UVA as a potential novel therapy for bladder cancer. Br J. Cancer 104, 1869 (2011).
12. Pollum, M., Jockusch, S. & Crespo-Hernández, C. E. Increase in the photoactivity of uracil derivatives by doubling thionation. Phys. Chem. Chem. Phys. 17, 27851–27861 (2015).
13. Gemenetzidis, E., Shavorskaya, O., Xu, Y.-Z. & Trigante, G. Topical 4-thiopyrimidine is a viable photosensitizer for the photodynamic therapy of skin malignancies. J. Dermatol. Treat. 24, 209–214 (2013).
14. Trigante, G. & Xu, Y.-Z. In Hugo, A. G. (ed.) Photodynamic Therapy: Fundamentals, Applications and Health Outcomes 193–206 (Nova Science Publishers, 2015).
15. Pridgen, M., Martinez-Fernández, L. & Crespo-Hernández, C. E. In Barbatti, M., Borin, A. C. & Ullrich, S. (eds) Photoinduced Phenomena in Nucleic Acids I, Vol. 356 of Topics in Current Chemistry 245–327 (Springer Berlin Heidelberg, 2015).
16. Crespo-Hernández, C. E., Cohen, B., Hare, P. M. & Kohler, B. Ultrafast excited-state dynamics in nucleic acids. Chem. Rev. 104, 1977–2020 (2004).
17. Giussani, A., Segarra-Martí, J., Roca-Sanjuán, D. & Merchán, E. Excitation of nucleobases from a computational perspective I: reaction paths. In Barbatti, M., Borin, A. C. & Ullrich, S. (eds) Photoinduced Phenomena in Nucleic Acids I, Vol. 355 of Topics in Current Chemistry, 57–97 (Springer Berlin Heidelberg, 2015).
18. Reichardt, C. & Crespo-Hernández, C. E. Ultrasinful spin crossover in 4-thiopyrimidine in an ionic liquid. Chem. Commun. 46, 5963–5965 (2010).
19. Reichardt, C. & Crespo-Hernández, C. E. Room-temperature phosphorescence of the DNA monomer analogue 4-thiopyrimidine in aqueous solutions after UVA excitation. J. Phys. Chem. Lett. 1, 2239–2243 (2010).
20. Reichardt, C., Guo, C. & Crespo-Hernández, C. E. Excited-state dynamics in 6-thioguanosine from the femtosecond to microsecond time scale. J. Phys. Chem. B 115, 3263–3270 (2011).
21. Taras-Goślin´ska, K., Burdzin´ski, G. & Wenska, G. Relaxation of the T1 excited state of 6-thio-2-thioguanine. J. Phys. Chem. A 115, 2239–2243 (2010).
22. Pollum, M. & Crespo-Hernández, C. E. Communication: The dark singlet state as a doorway state in the ultrafast and efficient intersystem crossing dynamics in 2-thiouracil and 2-thiouracil. J. Chem. Phys. 140, 071101 (2014).
23. Pollum, M., Jockusch, S. & Crespo-Hernández, C. E. 2,4-Dithiothymine as a potent UVA chemotherapeutic agent. J. Am. Chem. Soc. 136, 17930–17933 (2014).
24. Pollum, M., Ortiz-Rodríguez, L. A., Jockusch, S. & Crespo-Hernández, C. E. The triplet state of 6-thio-2'-deoxyguanosine: Intrinsic properties and reactivity toward molecular oxygen. Photochem. Photobiol. 92, 286–292 (2016).
25. Martinez-Fernández, L., González, I. & Corral, I. An ab initio mechanism for efficient population of triplet cytosine from cytosine photodegradation. Chem. Commun. 48, 2134–2136 (2012).
26. Gobbo, J. P. & Borin, A. C. 2-thiouracil deactivation pathways and triplet states population. Comput. Theor. Chem. 1040-1041, 195–201 (2014).
27. Cui, G. & Fang, W.-H. State-specific heavy-atom effect on intersystem crossing processes in 2-thiouracil: a potential photodynamic therapy photosensitizer. J. Phys. Chem. A 108, 14341–14346 (2004).
28. Cui, G. & Thiel, W. Intersystem crossing enables 4-thiopyrimidine to act as a photosensitizer in photodynamic therapy: an ab initio QM/MM study. J. Phys. Chem. Lett. 5, 2682–2687 (2014).
31. Yu, H.
39. Crespo-Hernández, C. E. et al.
43. Nakayama, A., Harabuchi, Y., Yamazaki, S. & Taketsugu, T. Photophysics of uracil and guanine. Eur. J. Inorg. Chem. 2011, 1487–1495 (2011).
49. Mai, S., Marquetand, P., González, L. A static picture of the relaxation and nonradiative transitions in the nitro-polycyclic aromatic hydrocarbons: a time resolved picture. J. Phys. Chem. B 117, 12796–12801 (2013).
54. Mai, S., Marquetand, P. & González, L. Antivir. Chem. Chemother. 24, 2920–2931 (2013).
50. Yamazaki, S. & Taketsugu, T. Nonradiative deactivation mechanisms of uracil, cytosine and uracil moieties. Eur. J. Med. Chem. 46, 4715–4724 (2011).