PCCP

ARTICLE

Study of light-induced formation of photodimers in the *i*-motif nucleic acid structure by rapid-scan FTIR difference spectroscopy and hybrid hard- and soft-modelling

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Sanae Benabou *[a], Cyril Ruckebusch [b], Michel Sliwa[b], Anna Aviñó[c], Ramon Eritja[c], Raimundo Gargallo[a], Anna de Juan[a]

The *i*-motif is a DNA structure formed by cytosine-rich sequences, very relevant from a biochemical point of view and potentially useful in nanotechnology as pH-sensitive nanodevices or nanomotors. To provide a different view on the structural changes and dynamics of direct excitation processes involving i-motif structures, the use of rapid-scan FTIR spectroscopy is proposed. Hybrid hard- and soft-modelling based on the Multivariate Curve Resolution by Alternating least squares (MCR-ALS) algorithm has been used for the resolution of rapid-scan FTIR spectra and the interpretation of the photochemically induced time-dependent conformational changes of i-motif structures. The hybrid hard- and soft-modelling version of MCR-ALS (HS-MCR), which allows the introduction of kinetic models to describe process behavior, provides also rate constants associated with the transitions modeled. The results show that UV irradiation does not produce degradation of the studied sequences but induces the formation of photoproducts. The presence of these affect much more the stability of i-motif structures formed by short sequences than that of those formed by longer sequences containing additional structural stabilizing elements, such as hairpins.

Introduction

The photochemistry of nucleic acids is an important branch of photobiology and has drawn much attention[1–3]. Recently, understanding the interaction between light and DNA has been facilitated by the development of fast and ultrafast spectroscopies and powerful computational techniques[4–7]. When exposed to ultraviolet (UV) light, nucleic acids are promoted to the excited electronic states. Fortunately, DNA is intrinsically photostable and can dissipate in most cases the excess of electronic energy before photoreaction happens[2]. This is a mechanism of self-protection that is important for the maintenance of life. This photostability arises from the intrinsic geometry and electronic structure of this molecule[2].

To identify new therapeutic strategies for the treatment of cancer and to help the design of anticancer drugs, the structures and properties of harmful photolesions caused by DNA exposure to UV radiation should be identified[8–10]. More importantly, the knowledge of the mechanism involved in the generation of DNA photolesions and their repair can be used for the prevention of cancer and subsequently, the enhancement of the quality of life. However, the underlying mechanisms of these behaviors are not always fully understood[11] in spite of the study of DNA-repair mechanisms performed in *E. coli* and humans[12] using a combination of genetic and biochemical approaches.

The *in vitro* formation of *i*-motif structures in DNA sequences corresponding to centromeres, telomeres and to the promoter regions of several oncogenes has been demonstrated[13–15]. Recently, a new study provided the first direct evidence for the presence *in vivo* of *i*-motif structures in human cells and control regulatory functions[16]. In this context, the study of DNA structures different from the Watson-Crick double helix is of great interest because of their potential role in some diseases and ageing phenomena. These structures, known as non-canonical structures, include the triplex, the G-quadruplex, and the *i*-motif[17,18], among others. The *i*-motif structure is formed by cytosine-rich sequences and consists of parallel-stranded duplexes held together by intercalated base pairs[19]. The formation of the C·C$^+$ base pair needs the protonation of one of the cytosines at N3 (known as Hoogsteen binding) (Figure 1a), the pK_a value of which is around 4.5, depending on the temperature and ionic strength. For this reason, stable *i*-motif structures can only be formed at weakly acidic pH and can be lost or unfold by effect of changes in temperature, pH or by the action of light[14] (Figure 1b).

In this work, the effect of light irradiation on the *i*-motif structures formed by four different DNA sequences is studied.

[a] Department of chemical Engineering and Analytical Chemistry, University of Barcelona, Martí I Franquès 1-11, E-08028 Barcelona, Spain
[b] Université de Lille Sciences et Technologies, LASIR CNRS, Cité Scientifique, 59655 Villeneuve d'Ascq, France.
[c] Institute for Advanced Chemistry of Catalonia (IQAC), CSIC, Networking Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Jordi Girona 18-26, E-08034 Barcelona, Spain

†Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

The objective is to gain knowledge about the potential influence of the sequence and the presence of additional hairpins in the loops in the stability against UV irradiation. First, two short (25 nucleotides) DNA sequences, identified as TT and AA in Figure 1c and Figure 1d will be studied. These sequences only differ in two bases (adenine or thymine) in their first and third loops (Figure 1d); thus, whereas AA contains two opposite adenine bases, TT contains two opposite thymine bases.

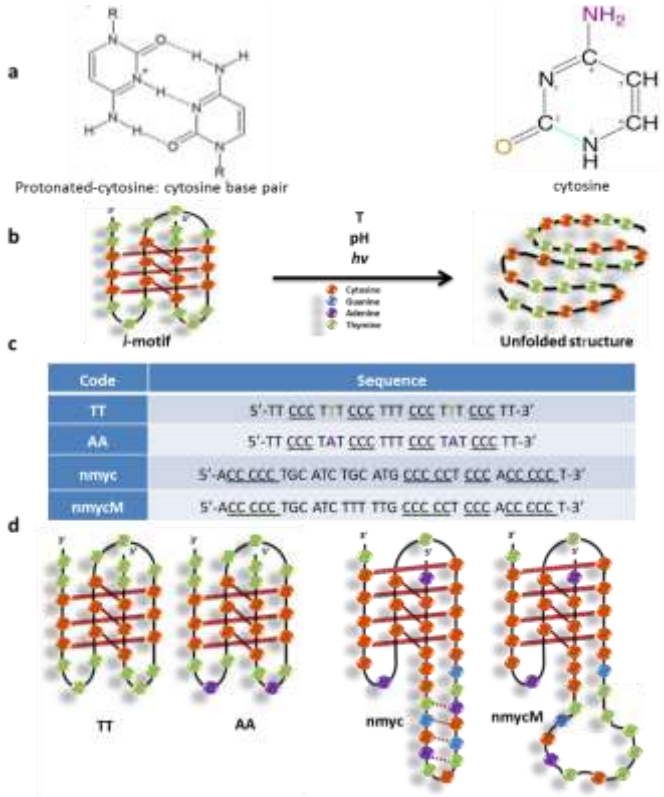


Figure 1. (a) Cytosine-protonated-cytosine base pair and free cytosine; (b) Schematic process of the unfolding of the intramolecular i-motif structures; (c) Sequences studied in this work. Underlined bases are those that could be involved in the formation of C+ · C base pairs; (d) Hypothetical scheme of the intramolecular i-motif structures adopted by the sequences proposed to be studied.

These sequences were already reported in a previous work performed on steady-state on the influence of the nature of internal bases located at the lateral loops on the thermal and acid–base stabilities of *i*-motif structures[20]. The results showed that the nature of the bases in the loops does not only affect the stability of *i*-motif structures with respect to temperature, but also produced structural modifications that could be detected by studying the acid–base behaviour. In terms of melting temperature (T_m) values, TT sequence formed the most stable i-motif structures at pH 5.6; whereas AA sequence formed the least stable one. Concomitantly, the results obtained from the acid–base titrations showed that the AA sequence exhibited the lowest stability with respect to pH changes. Here, TT and AA are used to study the effect of the nature of bases in the loop regions upon light irradiation of the *i*-motif structure.

In addition, two longer (34 nucleotides) sequences will be studied (nmyc and nmycM in Figure 1c and Figure 1d). The nmyc sequence corresponds to a cytosine-rich fragment found near the promoter region of the *n-myc* oncogene[21]. This gene is a member of the *myc* family of transcription factors and encodes a protein with a basic helix–loop–helix domain. Amplification of this gene is associated with a variety of tumors, most notably neuroblastoma[22]. The nmyc sequence shows an unusual 12-base long loop containing two complementary TGAC sequences that could promote the formation of a stable hairpin[21]. A mutated version of nmycM will be also studied. This one does not contain G, C or A bases at the loop (because they have been changed to T) and, consequently, this sequence cannot form the proposed hairpin shown in the hypothetical scheme of nmyc. The i-motif structures formed by both sequences contain additional C·C$^+$ base pairs that provide them greater stability than those formed by the TT or AA sequences.

To provide a detailed view on the fast-structural changes and dynamic processes occurring after direct irradiation of cytosine-rich DNA sequences, rapid-scan FTIR spectroscopy was used to monitor and interpret the spectral changes associated with the kinetics and mechanism of formation of *i*-motif structures. Rapid-scan FTIR spectroscopy is one of the preferred techniques to obtain rich structural information on photochemical processes of complex systems, such as proteins, in the millisecond scale[23]. It has been widely applied to the study of photoreactions such as those taking place at photosynthetic membranes[24–26] and to observe structural and hydration effects in B-DNA transitions and peptide recognition[27]. Besides, rapid-scan FTIR can be applied to any type of evolving systems, including nonreversible systems, and provides directly difference spectra that help to monitor small spectral changes. To our knowledge, rapid-scan FTIR spectroscopy has not been applied before to DNA systems or to *i*-motif structures to monitor light-induced changes as a function of time.

Multivariate data analysis based on hybrid hard- and soft-modeling multivariate curve resolution methods[28] has been used to interpret the light-induced conformational changes of the DNA sequences studied. The application of the hybrid hard- and soft-modelling multivariate curve resolution-alternating least squares (HS-MCR) provides the concentration profiles and related spectral signatures of all absorbing contributions in the monitored system (belonging to the kinetic process or not) and the rate constants associated with the kinetic events taking place[29]. As a process modelling algorithm, it has many advantages over pure hard-modelling and pure soft-modelling approaches.

As a summary of the main results of the paper, the photochemical process induced by UV irradiation is mostly explained in terms of the formation of dimeric photoproducts involving two adjacent pyrimidine bases, including cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoadducts (64PPs) [30]. The presence of a hairpin inside an i-motif structure modifies the development of this photochemical process, whereas the nature of bases at the central loop of the i-motif structure seems do not affect dramatically to it.

Materials and methods

Reagents and samples

The DNA sequences were synthesized on an Applied Biosystems 3400 DNA synthesizer using the 200 nmol scale synthesis cycle. Standard phosphoramidites were used and ammonia deprotection was performed overnight at 55 °C. The resulting products were purified using Glen-Pak Purification Cartridge (Glen Research, VA, USA). The length and homogeneity of the oligonucleotides was checked by reversed-phase HPLC using X-Terra® columns. DNA strand concentration was determined by absorbance measurements (at 260 nm) at 90 °C using the extinction coefficients calculated with the nearest neighbor method as implemented on the OligoCalc webpage[31]. NaCl, KH_2PO_4, K_2HPO_4, citric acid and NaOH were purchased from Panreac (Spain), Milli-Q® water was used in all experiments.

Instrumentation and experimental procedure

A Bruker's Tensor 27 FTIR spectrometer equipped with a photoconductive MCT-A detector and a KBr beam splitter was used in rapid-scan mode. Deuterated solutions of DNA sequences with an approximate concentration of 0.5 mM were introduced in a demountable liquid cell (Harrick Corp., USA) between two 25 mm diameter CaF_2 plates of 2 mm of thickness (Crystran Ltd, France), separated by a 56-micron Teflon spacer. Samples were prepared in deuterated phosphate buffer 50 mM with 100 mM NaCl and citric acid. Before measurements, sample solutions were first dried and then reconstituted in deuterated water.

To monitor the photoinduced reactions of interest, FTIR spectra were recorded under continuous illumination, in other words UV irradiation and data collection were synchronized, of the sample during 66 seconds with broad UV irradiation. LC8 (L9588-02A / 240-400 nm UV enhanced) model lamp from Hamamatsu equipped with a 280 - 400 nm band-pass filter (A9616-03) and a fiber to bring the light at the sample were used. The lamp is equipped with a mercury xenon lamp that gives a total power at the sample of 280 $mW \cdot cm^{-2}$. The spectrum of the light measured with an HR4000CG-UV-NIR at the sample is given in Supplementary Information. Background measurements were performed in the dark. Additional experiments (Supplementary Information) have been done to check that irradiation does not cause degradation of DNA strands.

FTIR difference spectra (ΔA) were calculated from the single beam spectrum obtained in the dark S(0) and the single-beam spectrum obtained at time S(t) according to the formula ΔA=−log[S(t)/S(0)]. Every experiment includes 300 FTIR difference spectra recorded between 1600 cm^{-1} and 1750 cm^{-1} at a resolution of 8 cm^{-1} in a global time window of 66 s (one spectrum every 0.22 seconds (time resolution), resulting from the average of 5 scans). Consequently, the dimensions of the related experimental data table (a data matrix **D**) are 300 × 66. The photochemical process of all the samples was performed at two different pH values: pH 4.0, where the *i*-motif folded structure is the major species, and pH 7.0, where the unfolded strand predominates. The data were acquired, exported and converted to MATLAB® (Mathworks, Natick, USA) readable files before being analyzed by appropriate chemometric methods (see below).

Data analysis

Data structure

Each process monitored by rapid-scan FTIR spectroscopy produces a data matrix **D** enclosing all spectral information, which can be described assuming a bilinear model based on the Beer-Lambert law:

$$\mathbf{D}_{(m,n)} = \mathbf{C}_{(m,p)} \cdot \mathbf{S}^{\mathbf{T}}{}_{(p,n)} + \mathbf{E}_{(m,n)} \quad (1)$$

When focusing on time-resolved spectroscopic data, **D**(*m*×*n*) contains the *m* measured difference spectra (at *n* wavenumbers) ordered as a function of time. **C**(*m*×*p*) is formed by columns describing the kinetic profiles of the *p* pure absorbing contributions in the sample and $\mathbf{S}^{\mathbf{T}}$(*p*×*n*) contains the related pure difference spectra. **E**(*m*×*n*) is the matrix containing the experimental error or, generally, the variance unexplained by the model.

Data pretreatment

FTIR difference spectra (**D**) were obtained directly from rapid-scan measurements. In difference spectra, there is no signal at time=0. Therefore, the interpretation of the spectra recorded in the following times should take into account that negative bands indicate the decay of spectroscopic signals related to the disappearance of initial compounds in a process, whereas positive bands indicate the emergence of spectroscopic signals linked to the formation of new products. The difference spectra were preprocessed by first derivative and smoothing by Savitzky-Golay (window 7 and polynomial order 2) before being further analyzed to enhance the small spectral variation recorded during the photochemical process[32].

Multivariate curve resolution alternating least squares (MCR-ALS)

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS)[33–35] is an iterative self-modelling approach that optimizes **C** and $\mathbf{S}^{\mathbf{T}}$ under constraints and has been successfully applied in numerous fields[36–38]. MCR-ALS has been widely used in process analysis because it provides pure component spectra and process profiles by using the information coming from the mixture spectra recorded during the evolution of the chemical system[39,40]. A good advantage is that no prior or little information is needed about the nature and composition of the compounds involved in the process of interest. MCR-ALS aims at resolving the bilinear model $\mathbf{D} = \mathbf{CS}^{\mathbf{T}}$ shown in Equation (1) by using the sole information contained in the raw data set **D**. The MCR-ALS *modus operandi* can be summarized in the following steps:

1. Determination of the number of components in the raw data set **D**;
2. Generation of initial estimates of either the evolving concentration profiles, **C**, or spectra, $\mathbf{S}^{\mathbf{T}}$;

3. Iterative least-squares calculation of the spectra matrix, $\mathbf{S}^T$, and the matrix of concentration profiles, **C,** under constraints until convergence is achieved (i.e. the LOF difference between two consecutive iterations is below a threshold or a predefined number of iterations is reached).

The percentage of lack of fit (LOF) and explained variance (R2) are used to determine the quality of the resolution model and are calculated by using Eqs. (2) and (3), respectively:

$$LOF = \sqrt{\frac{\sum_i \sum_j (d_{i,j} - \hat{a}i,j)^2}{\sum_i \sum_j d_{i,j}^2}} \quad (2)$$

$$R^2 = \frac{\sum_i \sum_j \hat{d}_{i,j}^2}{\sum_i \sum_j d_{i,j}^2} \quad (3)$$

$d_{i,j}$ and $\hat{d}_{i,j}$ are the ij^{th} element of **D** and the ij^{th} element of the reconstructed matrix by the MCR-ALS model, respectively.

More detailed description of the algorithm can be found elsewhere[26,34,39,40]. Briefly, the number of components can be estimated from singular value decomposition (SVD). The procedure needs an initial estimate of either **C** or $\mathbf{S}^T$. Here, the initial **C** matrix was generated using Evolving Factor Analysis (EFA)[41], which is a local rank analysis method[42], very appropriate to obtain initial estimates of concentration profiles from sequential evolutionary processes. The resolved profiles in **C** and $\mathbf{S}^T$ are not unique, being subject to rotational ambiguities. Constraints are therefore used to limit the number of possible solutions and to provide chemically meaningful concentration profiles and resolved spectra[41,43,44]. In this study non-negativity (concentration profiles of the resolved components must be positive) was applied to the concentrations. Non-negativity of the spectra cannot be applied to difference spectra signatures.

Hybrid hard- and soft-modelling multivariate curve resolution-alternating least squares

Hybrid hard- and soft-modelling multivariate curve resolution-alternating least squares is a variant of MCR-ALS that uses hard-modeling information. HS-MCR includes physicochemical models as an additional constraint to model the shape of the concentration profiles. Thus, in each iteration, the selected soft-modelled concentration profiles in **C** are passed to a non-linear fitting routine that constrains these concentration profiles to obey a preselected physicochemical model[39,40,42]. The outcome of the application of this constraint is the fitted profiles and the rate constants that define the model.

In the context of the present example, there are some advantages in the HS-MCR algorithm over pure hard- and soft-modelling approaches. First, the introduction of the hard-modeling constraint significantly decreases the rotational ambiguity linked to the classical soft-modeling MCR-ALS approach and the rate constants are provided as additional information[39,40,42]. On the other hand, the physicochemical model can be applied to some or to all the absorbing species in **C**, which is not possible in classical hard-modelling approaches[45]. Moreover, as it happens with classical pure hard-modeling approaches, non-absorbing contributions can be included in the non-linear model fitting. This advantage allows dealing with the specificities of difference spectra, where the initial species should be taken into account and be incorporated as non-absorbing species (no measurable signal can be associated with it) to be able to postulate a realistic physicochemical model of the monitored process where the decay of initial species is also taken into account[29]. Finally, another main advantage of HS-MCR is the possibility to perform simultaneous analysis of several experiments and fit different kinetic models to them. HS-MCR can work with experiments behaving according to different unconnected kinetic models[46] and model-based and model-free experiments can be analyzed together[47].

Matlab 8.2 and the MCR-ALS graphic interface[48] were used for all calculations. The interface is freely available on the web at http://www.mcrals.info/.

Results and discussion

FTIR difference absorbance spectra of *i*-motif structure

The FTIR spectra of the oligonucleotides studied in this work recorded in D_2O at pH values 7.0 and 4.0 are shown in Figure 2. Most spectral variation is associated with bands assigned to in-plane double bond C=C ring stretching vibrations and carbonyl mode of the bases in the 1600-1750 cm^{-1} region[49]. These bands, which are sensitive to base pairing and base stacking, allowed to follow C·C$^+$ base pair formation of *i*-motif structures induced upon decreasing pH.

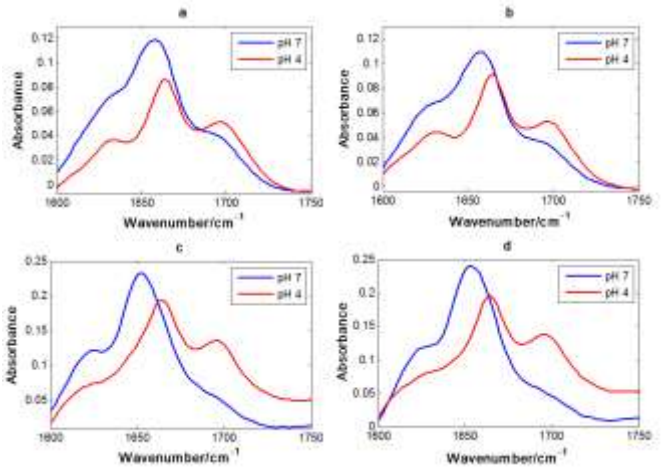


Figure 2. FTIR spectra corresponding to the cytosine-rich sequences considered in this work: (a) TT, (b) AA, (c) nmyc and (d) nmycM. DNA concentration: 1 mM. Spectra were acquired at pH 4.0 (50 mM citrate buffer, 100 mM NaCl, D_2O) and pH 7.0 (50 mM phosphate buffer, 100 mM NaCl, D_2O).

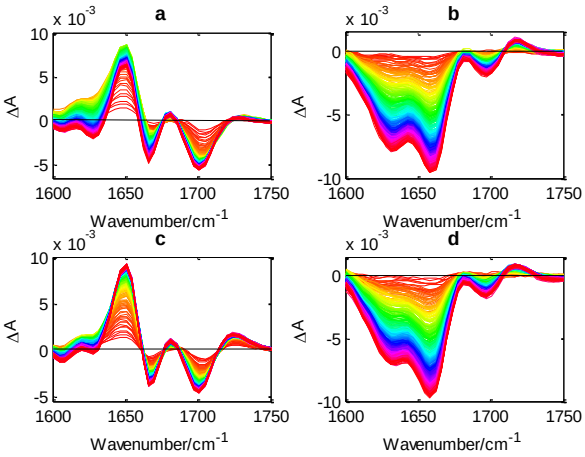
The FTIR spectra recorded at pH 7.0 showed an intense signal at 1655 cm^{-1} assigned to C(4)=O(4) thymine and C(2)=O(2) cytosine stretching modes, as well as two shoulders around 1620 (ND_2 scissoring of the neutral cytosines) and 1690 cm^{-1} (classically assigned to the C(2)=O(2) stretching mode of thymine)[49]. These features are common to all four sequences studied, suggesting that all of them show a similar secondary structure at pH 7.0, despite the clear differences present in their sequences (Figure 1c and 1d). At pH 7.0, in addition to the neutral cytosine bands, the thymine, adenine and guanine

residues produce different signatures[49,50] between 1600 cm^{-1} and 1700 cm^{-1}. It should be noted that the thymine signal around 1630 cm^{-1} (C=N ring in plane motion)[50] appears to be more pronounced in the case of nmyc and nmycM than in the case of TT and AA sequences.

The formation of the *i*-motif structure at pH 4.0 stabilized by $C{\cdot}C^+$ base pairs in TT and AA sequences was evidenced by the shift of the stretching bands above assigned at 1650-1655 cm^{-1} (C(2)=O(2) cytosine stretching modes) and 1690 cm^{-1} to high-frequency-shifted carbonyl stretching bands at 1665 and 1695 cm^{-1} (C=O stretching of protonated cytosine involved in Hoogsteen binding), respectively. Similarly, these bands are shifted to 1666 and 1700 cm^{-1} in the case of nmyc and nmycM, respectively. In all cases, the intensity of the band around 1695 cm^{-1} is higher than at pH 7.0.

Assessment of DNA integrity after irradiation

Before studying the nature and kinetics of the light-induced structural changes in cytosine-rich sequences, the potential degradation due to irradiation was studied on TT and nmyc sequences. Results obtained by using MALDI-TOF and MS-ESI techniques pointed to the conservation of molecular weights after irradiation, i.e., both techniques confirmed that DNA strands were not degraded after irradiation (Figures S2 and S3). Circular dichroism (CD) and molecular absorption spectra measured at pH 7 (where the unfolded strand is the major species) were similar before and after irradiation (Figures S4 and S5). On the other hand, CD spectra measured at pH 4 (where *i*-motif is the major species) showed a reduction of intensity after irradiation. This reduction of intensity was not accompanied by changes in the position of CD bands (Figure S4c and S4d). Additional CD-monitored melting experiments pointed to changes in the initial i-motif structure affecting the cooperativity of the unfolding process, especially in the case of



1. Figure 3. Rapid-scan FTIR raw spectra of TT and AA sequences: (a) TT at pH 4.0, (b) TT at pH 7.0, (c) AA at pH 4.0, and (d) AA at pH 7.0. DNA concentration: 0.5 mM, spectra were acquired at pH 4.0 (50 mM citrate buffer, 100 mM NaCl, D2O) and pH 7.0 (50 mM phosphate buffer, 100 mM NaCl, D2O). Process evolution follows the colormap sequence: red-yellow-green-blue-purple-dark red.

TT sequence (Figure S6). Taken together, all these data suggested that the i-motif structure formed by TT sequence was more affected by irradiation than that formed by nmyc. In both cases, no degradation was observed.

TT and AA sequences. Study of the effect of base sequences present at the lateral loops in *i*-motif formation.

Figure 3 shows the evolution of the difference raw spectra (ΔA) recorded during 66 seconds of UV irradiation for TT and AA DNA sequences at pH 4.0 and pH 7.0. These sequences differ only in the nature of two bases located at the first and third lateral loops (see Figure 1d). Whereas TT shows two opposite thymine bases, AA shows two opposite adenine bases. This difference could influence *i*-motif folding since adenine bases are bulkier than thymine ones. Besides, the TT sequence could produce additional non-canonical T·T base pairs.

As it was mentioned previously, negative bands in difference spectra are related to the decay of spectroscopic signals related to initial compounds in a process, whereas positive bands indicate the emergence of spectroscopic signals linked to new products[29]. At pH 4.0, the series of FTIR difference spectra for the TT sequence shows the rising of a band at 1650 cm^{-1} and the reduction of the bands at 1665 and 1700 cm^{-1}. Also, the intensity of the band around 1620 cm^{-1} increases slightly. All these spectral variations are in good agreement with those shown in Figure 2a, which suggests a certain degree of unfolding of the *i*-motif upon irradiation. At pH 7.0, where the TT sequence is expected to exist as an unfolded strand at the beginning of the experiment (Figure 3b), a decrease of intensity is observed in the range from 1600 to 1710 cm^{-1} (only a small increase is observed around 1720 cm^{-1}).

On the other hand, the spectral variations observed for AA sequence at pH 4.0 are slightly different from the one observed for TT at the same pH value. The increase of the band at 1720 cm^{-1} is more intense for AA whereas the decrease of the shoulder at 1620 cm^{-1} is less intense. Finally, at pH 7.0, both sequences show a very similar behavior.

Prior to multivariate analysis, the data shown in Figure 3 were preprocessed by calculating the first derivative and smoothing by using a Savitzky-Golay filter. This preprocessing enhances the small spectral differences occurring during process monitoring and decreases experimental noise. The preprocessed spectra are shown in Figure 4 (left plots). Soft-modelling MCR-ALS was then applied to investigate the potential presence of intermediates along the photochemical process and to calculate the corresponding concentration profiles (**C**) and pure spectra ($\mathbf{S}^T$). The results obtained for the TT and AA sequences at pH 7 and 4 are presented in Figure 4 (center and right plots, respectively). In this case, starting values for the evolving profiles of each species were calculated by means of Evolving Factor Analysis. MCR-ALS was performed applying non-negativity constraint to the concentration profiles.

Table 1. Lack of fit (LOF, in %) and explained variance (r2) obtained in the MCR-ALS analysis of FTIR data recorded along the photochemical process of TT and AA sequences.

Sequence	pH	% Lack of fit	% r^2
TT	4.0	3.4	99.89
	7.0	6.9	99.52
AA	4.0	3.2	99.89
	7.0	7.0	99.50

presence of this intermediate from visual inspection of the data in Figures 3a and 3c. The calculated pure spectra (**S**^{T}) for the intermediate and final species at pH 4.0 are similar for both TT and AA sequences.

On the other hand, the series of FTIR spectra of TT and AA at pH 7.0 were successfully modelled with just one component, whose pure spectrum was very similar for both sequences. This implies a unique transition from the initial species (with no signal) into a final species (in red in Figure 4).

In all TT and AA experiments, the calculated concentration profiles (**C**) resembled those expected for a kinetic process

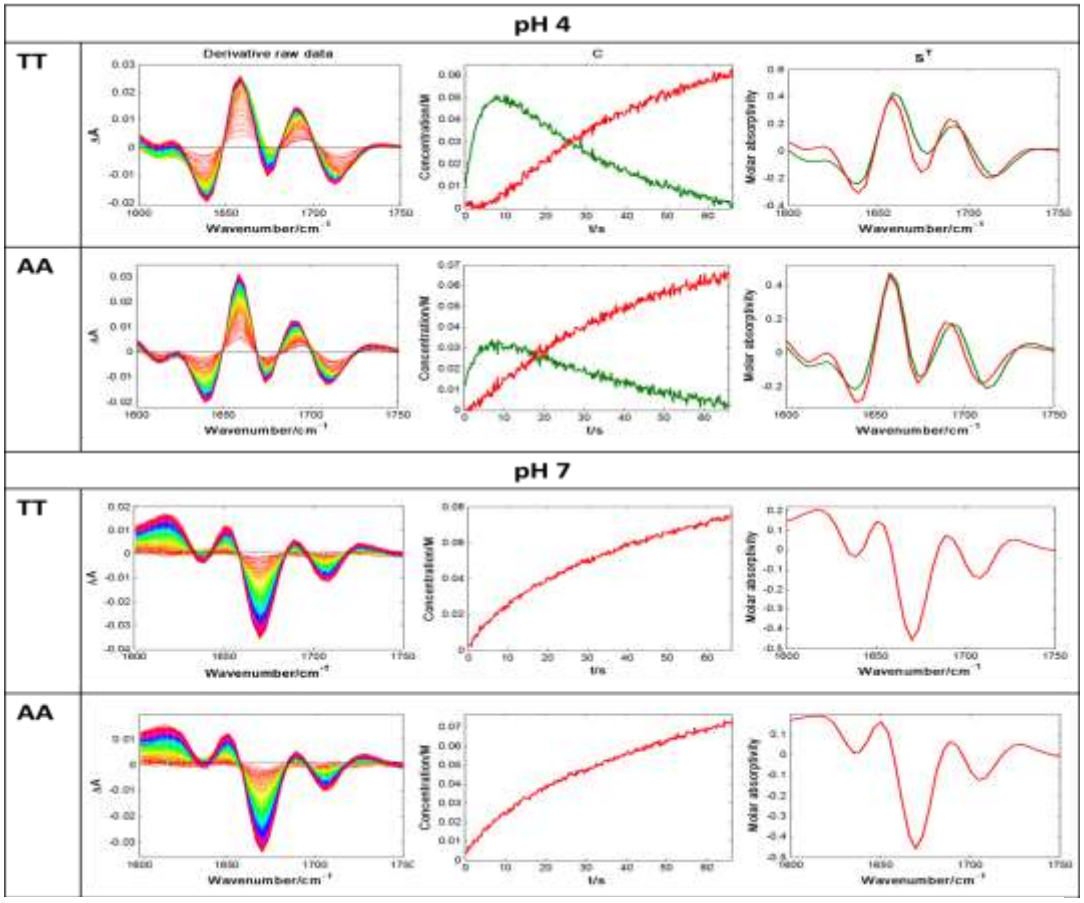


Figure 4. MCR-ALS analysis of first derivative rapid-scan FTIR spectra of 0.5 mM TT and AA at pH 4.0 and pH 7.0. Left plot: raw data (process evolution follows the colormap sequence: red-yellow-green-blue-purple-dark red). Middle plot: concentration profiles (**C**). Right plot: pure difference spectra (**S**^{T}).

Table 1 summarizes the figures of merit for the model fit obtained in the analysis of the different systems. The FTIR spectra of the light-induced processes undertaken by TT and AA at pH 4.0 were successfully modelled with two components. Since the process is monitored in difference spectroscopy, resolving two components indicates the presence of three species because processes monitored by difference spectroscopy always have the spectrum of the initial stage subtracted and, consequently, a rank-deficient data set is obtained. The three species would correspond to an initial species without spectral signal associated, an intermediate species (in green in Figure 4), and a final species (in red in Figure 4). This result is an example of the usefulness of multivariate analysis because it is very difficult to detect the passing through first-order consecutive reactions. At pH 7.0, both sequences showed visually a very similar kinetic profile whereas small differences were observed at pH 4.0.

To determine rate constants and to circumvent the rank-deficiency phenomenon linked to difference spectra, hybrid HS-MCR was applied to the TT and AA data sets previously analyzed by soft modelling MCR-ALS. Concentration profiles obtained from soft MCR-ALS analysis were used as initial estimates for HS-MCR. The results are shown in Table 2 and Figure 5.

At pH 4.0, a simple kinetic model of first-order consecutive reactions $((i-motif)_A \xrightarrow{k_1} (i-motif)_B \xrightarrow{k_2} (i-motif)_C)$ was proposed to describe the photochemical process involving the intramolecular *i*-motif structure formed by either TT or AA

sequences. The calculated concentration profiles and pure spectra by using HS-MCR resembled those calculated previously by using soft-modelling (Figure 4). The explained variance (r^2 = 99.8% and 99.8% for TT and AA sequences, respectively) is like that obtained by using the soft-modelling analysis, which indicates that the selected kinetic model was well-supported by the experimental data.

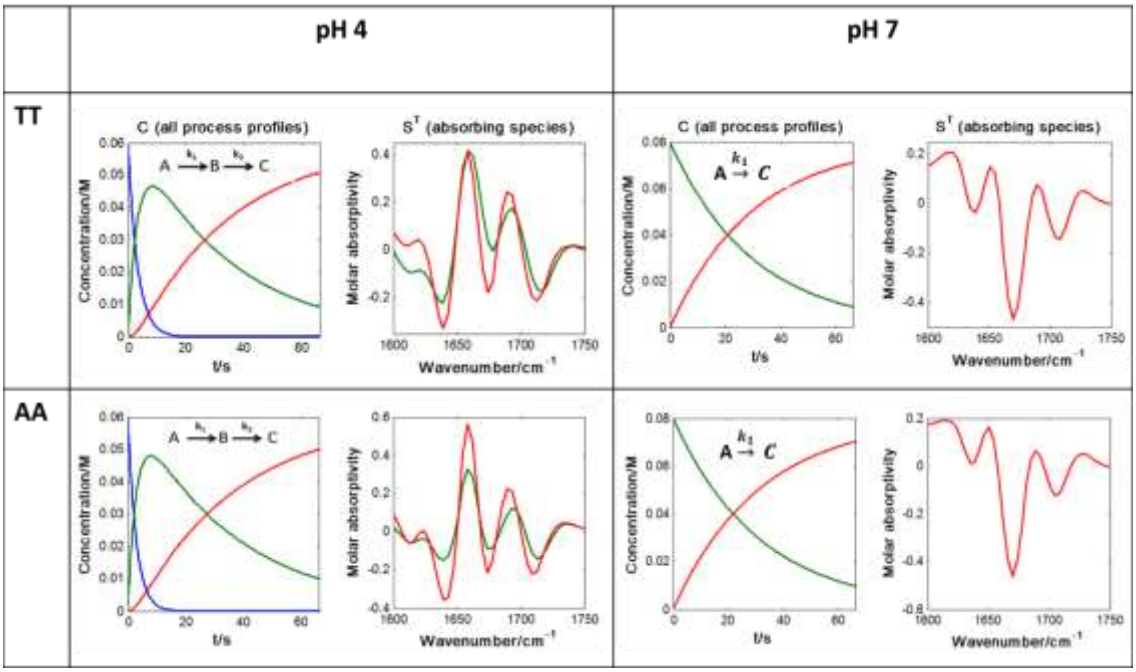


Figure 5. HS-MCR-ALS resolution results of first derivative rapid-scan FTIR spectra of 0.5 mM TT and AA at pH 4.0 and pH 7.0. Left plots: concentration profiles (initial decaying species is non-absorbing). Right plots: pure difference spectra.

Table 2. Kinetic parameters obtained by HS-MCR on FTIR data recorded along the photochemical process of TT and AA sequences.

Sequence	pH	k_1 (s^{-1})	k_2 (s^{-1})	% Lack of fit	% r^2
TT	4.0	0.305 ± 0.004	0.0300 ± 0.0001	4.6	99.78
	7.0	0.0332 ± 0.0001	-	7.8	99.39
AA	4.0	0.346 ± 0.008	0.0284 ± 0.0002	4.2	99.82
	7.0	0.0319 ± 0.0002	-	8.1	99.34

The explanation of the model is based on complementary experimental data. As already commented, it has been observed that the final product has the same molecular weight than the initial i-motif structure (Figures S2 and S3). On the other hand, its CD spectrum has a similar CD signature than the initial i-motif structure, while its intensity is lower (Figure S4a). Finally, the unfolding process of the final product is less cooperative than that observed for i-motif before irradiation (Figure S61.b).

On the other hand, the proposed model was also based on the results obtained by enzymatic digestion followed by HPLC separation (see Figure S7). At pH 7, the chromatogram for the non-irradiated TT sequence showed two peaks corresponding to C and T nucleotides. Upon irradiation at pH 7, the chromatogram shows the appearance of a major peak at higher retention time, which was related to a major type of photoproduct. At pH 4, on the other hand, more peaks are observed that may be related to a mix of photoproducts. Baggesen *et al.*, using calf thymus DNA, demonstrated the possibility of formation of both photoproducts[54] (CDP and 64PP), being the possibility of 64PP formation higher at pH 4 than pH 7 due to the short distance between the thymine bases. A very close look at absorption data of TT sequence at pH 4 (Figure S5) also reveals a small increase of absorbance at 325 nm, indicating the minor presence of 64PPs.

Altogether, the species "$(i\text{-motif})_A$" corresponds to the initial *i*-motif structure. In this case, this initial species was flagged as non-absorbing in the kinetic model fitting because the measurement is difference spectra [29,51,52]. However, this species was needed to provide a complete and correct description of the whole process evolution. The last species "$(i\text{-motif})_C$" would correspond to an i-motif containing cyclobutane pyrimidine dimers (CPDs) between adjacent thymine or cytosine bases. This i-motif structure has the same molecular weight and overall shape than the initial i-motif. However, as some bases are now forming dimers resulting from irradiation, its unfolding is clearly affected, as observed from CD-monitored melting experiments (Figure S6). On the other hand, the intermediate species "$(i\text{-motif})_B$" would involve the formation of pyrimidine (6-4) pyrimidone photoadducts (64PPs) between the C5–C6 double bond of the 5'-end thymines and the C4 carbonyl group of the 3'-end

thymines. Due to the short distance between these bases when the *i*-motif structure is formed, the formation of this photoproduct would be more relevant at acidic pH[53]. For both sequences, the transition from species "$(i\text{-motif})_A$" to species "$(i\text{-motif})_B$" is fast, as denoted by the relatively high rate constant, k_1. On the contrary, the transition from species "$(i\text{-motif})_B$" to species "$(i\text{-motif})_C$" is ruled by a slower process, which may indicate subtle changes in the nucleic acid strand. Finally, the rate constants at pH 7.0 are similar for both sequences, which indicate a very similar kinetic evolution. It should be stressed that the value of the rate constant at pH 7.0 is similar to the second rate constant at pH 4.0, which could be associated with the formation of CPDs photoproducts. At this point, it should be noted that the unfolded strand "$(open\ strand)_A$" makes difficult the formation of 64PPs at pH 7 due the long distance between the bases involving in this type of photoproducts. So, the CPDs would be the only photoproduct that could form at pH 7 "$(open\ strand)_C$", which could be also present in species "$(i\text{-motif})_C$" at pH 4.0. Hence, the only transition observed at pH 7.0 could be equivalent to the second transition modelled at pH 4.

In previous studies[20], the sequence 5'-TT CCC TXT CCC TTT CCC TXT CCC TT-3', where X are thymine (T), adenine (A), cytosine (C) or guanine (G) was studied as a model to understand the influence of the nature of bases located at the lateral loops on the stability of the *i*-motif structure (Figure 1c). It was shown that the TT sequence forms one of the most stable *i*-motif structures to temperature and pH changes, whereas AA is the least stable sequence. Therefore, it was concluded that the stability of the *i*-motif structure decreases when two adenine bases are placed opposite to each other. On the other hand, Geinguenaud[49] *et al.* studied the effect of incorporating purine residues in DNA sequences by FTIR spectroscopy and other techniques. The results showed that purine inserts into an oligomeric cytosine sequence makes the formation of the tetraplex *i*-motif less favorable. The results obtained in the present work show that the TT and AA sequences with formed *i*-motif structures (at pH 4.0) show different behavior upon irradiation than the corresponding unfolded strands (at pH 7.0). Even though the rate constants ruling the formation of photodimers in *i*-motif structure were on the same order of magnitude for AA and TT, a slightly faster decay of the *i*-motif structure was observed for the AA sequence (seen in the slightly higher k_1 value at pH 4.0), in agreement with the lower stability of the *i*-motif linked to the presence of adenine in the DNA sequence, as reported above[15]. The resolved spectral signatures of the species involved in the process studied are also slightly different among sequences, which points to a certain influence of the nature of the base (T or A) on the overall *i*-motif folding.

nmyc *vs.* nmycM sequences. Influence of a lateral hairpin in the *i*-motif formation.

Figure 6 shows the evolution of the difference raw spectra under light during 66 seconds of UV irradiation for nmyc and nmycM sequences, at pH 4.0 and pH 7.0, respectively. As shown in Figure 1c and 1d, nmyc includes a short stretch of bases able to form a lateral hairpin stabilized by Watson-Crick A·T and G·C base pairs. This hairpin can be formed either at pH 7.0 or at pH 4.0. On the contrary, nmycM, which is a version of nmyc where bases at the lateral hairpin have been changed to T, cannot form this lateral hairpin.

Overall, the dependence of the spectral changes shown in Figure 6 with pH has some similarities with those provided in Figure 3 for TT and AA sequences. As an example, large differences were observed when changing from pH 4.0 to pH 7.0 for both sequences, nymc and nmycM. However, a closer look reveals that some spectral changes for nmyc and nmycM are different from those for TT and AA sequences.

At pH 4.0, raw rapid-scan FTIR difference spectra of nmyc (Figure 6a) showed an enhancement of the bands around 1625 cm^{-1} (ND2 scissoring of the neutral cytosines) and 1650 cm^{-1} (C(2)=O(2) cytosine stretching modes) and a decrease of the band around 1695 cm^{-1} (C(2)=O(2) stretching mode of thymine and protonated cytosine involved in Hoogsteen binding). These changes are in good agreement with those seen in the direct FTIR spectra shown in Figure 2c. At pH 7.0, an overall decrease was observed from 1600 to 1700 cm^{-1}, which was related to the complete unfolding of the initial structure. A closer inspection of the raw data for nmyc and nmycM at pH 4.0 reveals clear differences between both sequences at 1625 and 1665 cm^{-1}; thus, whereas the decrease of the band at 1665 cm^{-1} is more pronounced in the case of nmycM, the increase of the band at 1625 cm^{-1} is more intense in the case of nmyc.

At pH 4.0, raw rapid-scan FTIR difference spectra of nmyc (Figure 6a) showed an enhancement of the bands around 1625 cm^{-1} (ND2 scissoring of the neutral cytosines) and 1650 cm^{-1} (C(2)=O(2) cytosine stretching modes) and a decrease of the band around 1695 cm^{-1} (C(2)=O(2) stretching mode of thymine and protonated cytosine involved in Hoogsteen binding). These changes are in good agreement with those seen in the direct FTIR spectra shown in Figure 2c. At pH 7.0, an overall decrease was observed from 1600 to 1700 cm^{-1}, which was related to the complete unfolding of the initial structure. A closer inspection of the raw data for nmyc and nmycM at pH 4.0 reveals clear differences between both sequences at 1625 and 1665 cm^{-1}; thus, whereas the decrease of the band at 1665 cm^{-1} is more pronounced in the case of nmycM, the increase of the band at 1625 cm^{-1} is more intense in the case of nmyc.

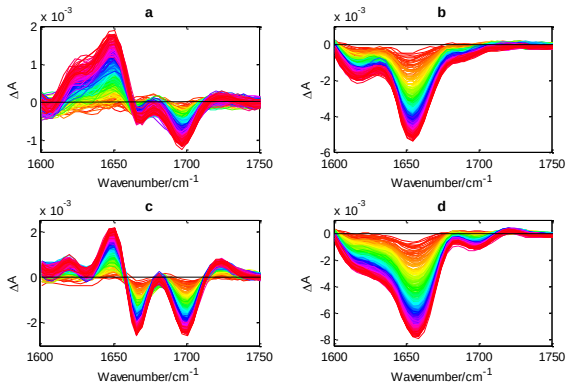


Figure 6. Raw rapid-scan FTIR spectra of 0.5 mM nmyc and nmycM at pH 4 and pH 7; (a) nmyc at pH4, (b) nmyc at pH7, (c) nmycM at pH4 and (d) nmycM at pH7. Process evolution follows the colormap sequence: red-yellow-green-blue-purple-dark red.

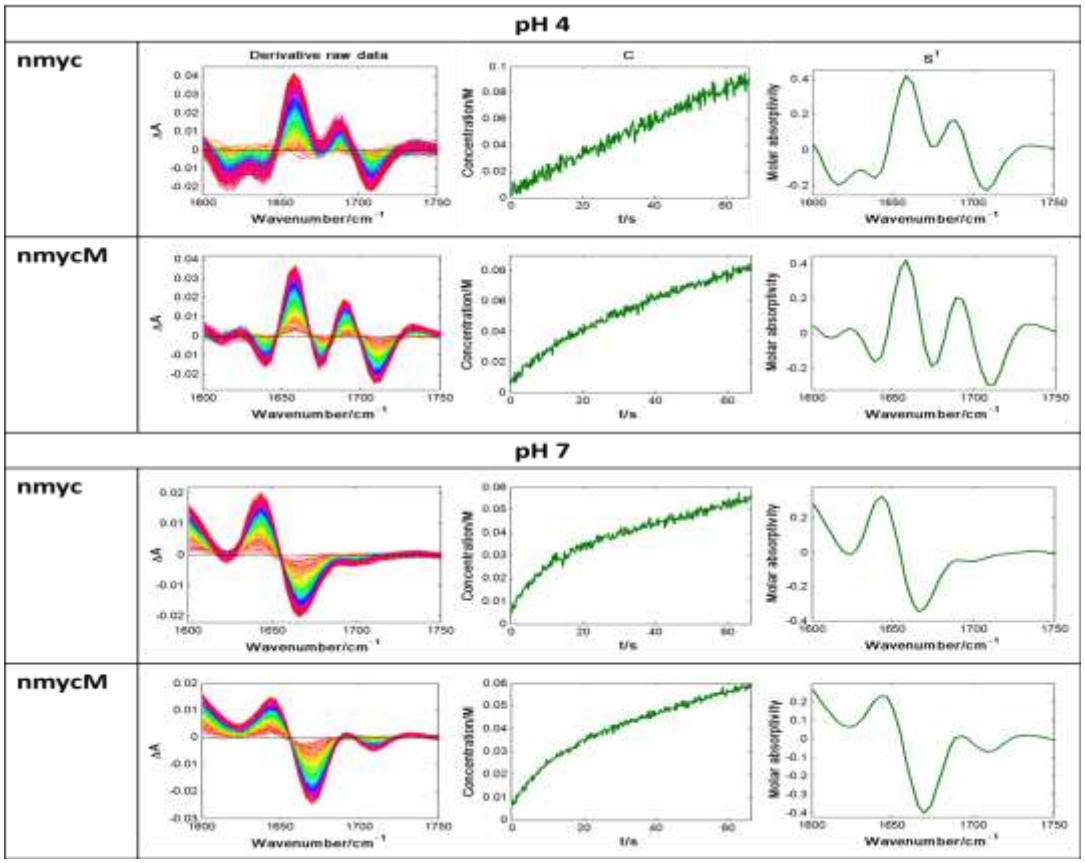


Figure 7. MCR-ALS analysis of first derivative rapid-scan FTIR spectra of 0.5 mM nmyc and nmycM at pH 4 and pH 7. Left plot: raw data (process evolution follows the colormap sequence: red-yellow-green-blue-purple-dark red). Middle plot: concentration profiles (**C**). Right plot: pure difference spectra ($\mathbf{S}^T$).

Also, a small increase is observed around 1720 cm^{-1} in the case of nmycM, which is not observed in the case of the wild sequence. At pH 7.0, there are also some differences between nmyc and nmycM, the clearest being the larger decay at 1665 cm^{-1} and the slight increase of the band at 1690 cm^{-1}, in the case of nmycM. The presence of additional base pairs in the lateral hairpin in nmyc produces more changes in the corresponding FTIR spectrum than in the case of nmycM. This was not observed for TT and AA sequences, which showed a more similar spectral and photochemical evolution.

As in the case of the TT and AA sequences, after preprocessing the raw difference spectra, soft-modelling MCR-ALS was applied following the procedure described in section 3.2. The results of soft-modelling MCR-ALS analysis of nmyc and nmycM at pH 4.0 and 7.0 are shown in Figure 7.

The rapid-scan FTIR spectra of both sequences at different pH values were successfully modelled with one component, which in difference spectroscopy means that there are two contributions involved in the decay of the *i*-motif. This result is different from that obtained for both TT and AA sequences at pH 4.0. The extracted pure difference spectra of both sequences at pH 4.0 show clearly noticeable differences in intensity at 1630 and 1665 cm^{-1}. Also, the shape of the calculated concentration profiles is slightly different. Thus, the resolved concentration profile for nmyc shows a slower evolution, almost a linear shape, which is different from the faster and exponential evolution found for nmycM in the same time scale. At pH 7.0, the kinetic profiles and spectral signatures are practically identical in both sequences.

Table 3. Lack-of-fit (LOF, in %) and explained variance (r^2) obtained in the MCR-ALS analysis of FTIR data recorded along the photochemical process of nmyc and nmycM sequences.

Sequence	pH	% Lack of fit	% r^2
Nmyc	4.0	20.7	95.73
	7.0	9.3	99.13
nmycM	4.0	11.5	98.68
	7.0	6.8	99.54

Table 3 shows the quality of MCR-ALS model fit. The percentage of lack of fit is higher than in the case of TT and AA sequences due to the lower signal-to-noise ratio, associated with less intense spectral changes linked the *i*-motif transitions than in the previously studied shorter DNA sequences.

HS-MCR was also applied to the nmyc and nmycM data analyzed previously by soft modelling MCR-ALS. Figure 8 and Table 4 summarize the results obtained. As shown previously for TT and AA sequences, complementary experimental data were used in this phase. A first-order reaction kinetic model in one step was proposed to model the photochemical process of nmyc and nmycM sequences at both pH values. HS-MCR concentration profiles (**C**) and pure spectra ($\mathbf{S}^T$) resemble in

shape to the soft-modelling results. Explained variances, similar to the soft-modelling analysis, were obtained, indicating that the selected kinetic model was appropriate. At pH 4.0, the model $(\text{i-motif})_A \rightarrow (\text{i-motif})_C$ was set, where the species "$(\text{i-motif})_A$", which corresponds to the sequence with the *i*-motif formed (i.e., the initial state of the photochemical experiment), was set to be a non-absorbing species [29,51,52]. The species "$(\text{i-motif})_C$" would correspond to the i-motif structure containing CPD photodimers. Only two species were required to fully describe the kinetic evolution of these two sequences at pH 4.0 in contrast to the behavior with AA and TT sequences. This different behavior could be related to the lower amount of pyrimidine bases at the 5' and 3' ends, which would prevent the formation of 64DD photoproducts in nmyc and nmycM i-motifs.

and, hence, the present results agree with literature[56]. Furthermore, the nmyc sequence at pH 4.0 shows a lower rate constant than nmycM and this can be related to the additional stabilization provided by the lateral hairpin stabilized by Watson-Crick base pairs that further hinders the formation of photodimers in i-motif. It should be reminded that this hairpin cannot be formed in the nmycM sequence. At pH 7.0, on the contrary, the calculated rate constants for both sequences are quite similar, indicating that the hairpin does not provide additional stability to the initial conformation of the open strand.

In our previous study[21], we confirmed that the intramolecular *i*-motif from nmyc sequence is stable throughout the pH range 2–6, with maximal stability at pH 4.5. The potential presence of the base pairing between the TGCA repeats in the long loop was also studied and confirmed by NMR spectroscopy and

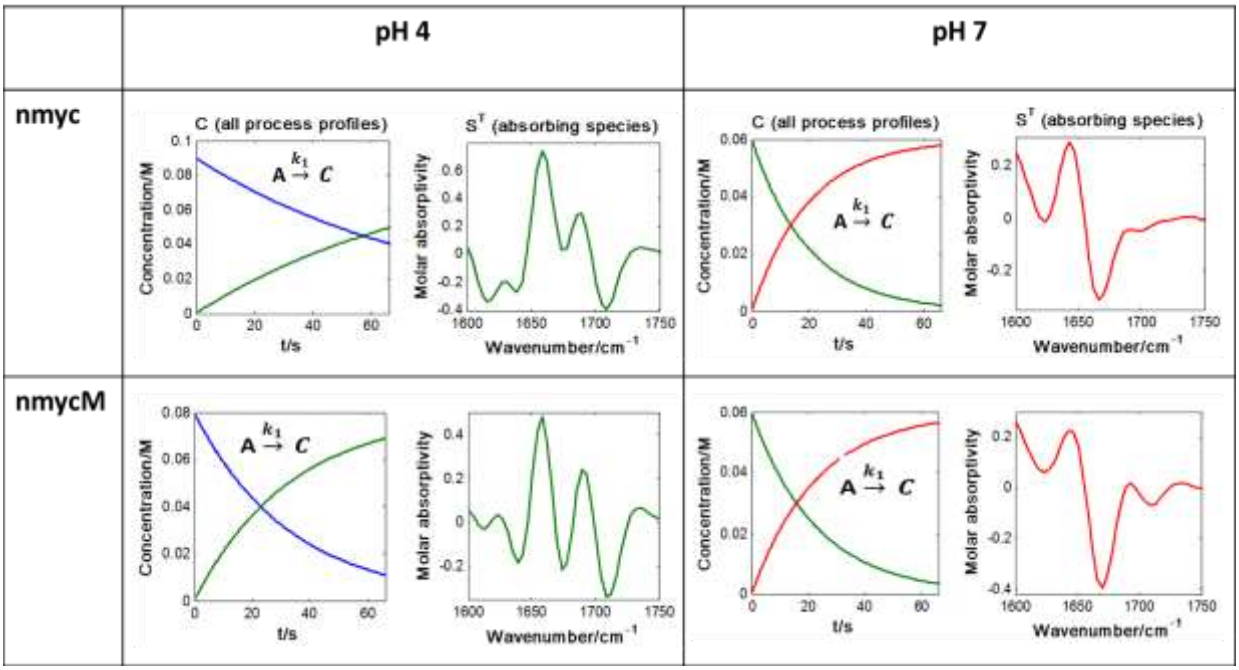


Figure 8. HS-MCR-ALS resolution of first derivative rapid-scan FTIR spectra of 0.5 mM nmyc and nmycM at pH 4 and pH 7. Left plots: concentration profiles (initial decaying species is non-absorbing). Right plots: pure difference spectra signatures.

For the experiments at pH 7.0, a one-step first-order reaction model, encoded $(\text{open strand})_A \rightarrow (\text{open strand})_C$, was postulated. Again, the species B is the non-absorbing initial unfolded strand at pH 7.0, and the species C is related to the formation of CPD in this open strand by the action of light for both nmyc and nmycM sequences. The nmyc and nmycM sequences show clearly lower rate constants associated with the decay of the *i*-motif than TT and AA sequences (values at pH 4.0 in Table 4 and k_1 in Table 2, respectively). This fact could be explained as a result of two factors. First, the greater stability of nmyc and nmycM structures[21], which is related to the presence of an additional $C{\cdot}C^+$ base pair in the *i*-motif structure. Second, the lower presence of pyrimidine bases in their sequences, which are the precursors of all these photoproducts. Besides, Douki et al. have shown that TpT and TpC DNA fragments are the most photoreactive sequences [55], whereas lower amounts of damage were produced at CpT and CpC sites. Nmyc sequence shows less amounts of TpT and TpC other techniques. The second TGCA sequence was replaced by four T bases in mnycM, which impedes the formation of the suspected stem-loop structure. At pH 7.0 and pH 5.0, the presence of such base pairs was seen dramatically reduced in comparison to the wild sequence nmyc, which confirmed the absence of a stem. To conclude, we suggested the formation of a stable Watson–Crick hairpin by the bases in the first loop, stabilizing the *i*-motif structure of nmyc[21]. Considering these results, it was clear that the nmyc sequence studied in the present work was expected to display a slightly stronger stabilization and, consequently, a relatively higher resistance to the direct photoexcitation by UV light, showing a low rate constant linked to a slow photochemical process.

Table 4. Kinetic parameters obtained from the analysis with HS-MCR of FTIR data recorded along the photochemical process of nmyc and nmycM sequences.

Sequence	pH	k_1 (s^{-1})	% Lack of fit	% r^2
nmyc	4.0	0.0121 ± 6.8e-05	21.8	95.25
	7.0	0.0507 ± 0.0006	11.8	98.60
nmycM	4.0	0.0300 ± 0.0002	12.7	98.38
	7.0	0.0437 ± 0.0004	9.2	99.15

Conclusions

In this work, light-induced formation of dimeric photoproducts in *i*-motif structures of four different DNA sequences were studied by combining rapid-scan FTIR difference spectroscopy and hybrid hard- and soft-modelling. Most of the differences observed for changes in *i*-motif structure were seen in experiments carried out at pH 4.0. At this pH and room temperature, the *i*-motif structure formed by the TT sequence, which can be considered as a model sequence, exists naturally. First, the effect of the length of DNA sequences was assessed. Longer sequences (nmyc and nmycM) proved to be clearly more stable under illumination[57]. These sequences showed rate constants linked to the decay of the *i*-motif that are one order of magnitude lower than those of short sequences (AA and TT). The fact that the *i*-motif formed by nmyc and nmycM has one $C{\cdot}C^+$ pair more than the ones formed by AA and TT additionally supports the higher stability of the *i*-motif in these longer DNA sequences and the more difficult formation of photodimers. Besides, the base composition of these long sequences plays a role in the low formation of photoproducts. Less pyrimidine bases are available to form adjacent or non-adjacent photodimers due to the formation of compact *i*-motif structure. A second issue consisted of checking the influence of base sequence at the lateral loops on the i-motif stability. These differences can produce modifications in stability, like that already discussed for TT and AA sequences. They may also induce the formation of a lateral hairpin structure as for nmyc, whereas this is not the case for nmycM. Although less significant than the effect of the length of DNA sequences and *i*-motif size, these minor differences play a role in the formation of photoproducts in *i*-motif structures. When comparing nmyc and nmycM, we observed that the wild structure nmyc including the hairpin is slightly more stable than the mutated nmycM, where this structural element does not exist. The rate constant of the nmycM sequence associated with the formation of photoproducts in *i*-motif decay is three times higher than that of the nmyc sequence because of the presence of available thymine bases in the loop. More subtle is the difference in behavior between TT and AA, where TT only presents a rate constant associated with the decay of the *i*-motif slightly lower than that of the AA sequence, indicating that the nature of the lateral loops in the *i*-motif is less crucial than other factors in the stabilization of this structural element.

In contrast to the experiments developed at pH 4.0, essays carried out at pH 7.0 showed rather similar behavior in most DNA sequences. This is understandable if it is considered that the DNA sequences studied are supposed to be in the form of open strand at this pH value and, hence, differences are less noticeable among transition of the formation of CPD photoproduct in open strand. Su *et al.* have discovered a new DNA photoproduct that can only be produced in acidic pH. This photoproduct is a non-adjacent CPD formed between bases at the loops[53]. Overall, our results are in very good agreement with the previous studies and this kind of photoproduct could form in *i*-motif structures at pH 4 in addition to the common CPD photoproducts.

From a data analysis point of view, MCR-ALS has shown to be a powerful tool to study the dynamics of the conformational changes of *i*-motif structures, monitored by rapid-scan FTIR difference spectroscopy. The combination of MCR-ALS and rapid–scan FTIR difference spectroscopy allows proposing a kinetic mechanism associated with the photochemical transitions related to the formation of photoproducts in the *i*-motif structures and helps in the detection of intermediate if existing. This approach can be easily extended to *in vitro* data coming from the analysis of living cells.

Acknowledgements

Sanae Benabou thanks the LAboratoire de Spectrochimie Infrarouge et Raman (LASIR) for all the support, especially, Dr. Isabelle Waele for her help in the acquisition of rapid-scan FTIR measurements. Funding from Spanish government (CTQ2014-61758-EXP, CTQ2014-52588-R and CTQ2015-66254-C2-2-P) and recognition from the Autonomous Catalan government (2014SGR1106) are acknowledged.

Notes and References

1. C. E. Crespo-Hernández, B. Cohen, P. M. Hare, and B. Kohler, *Chem. Rev.*, 2004, **104**, 1977–2019.
2. C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández, and B. Kohler, *Annu. Rev. Phys. Chem.*, 2009, **60**, 217–239.
3. A. Sancar, *Adv. Protein Chem.*, 2004, **69**, 73–100.
4. D. Markovitsi, *Photochem. Photobiol.*, 2016, **92**, 45–51.
5. L. Martinez-Fernandez, A. Banyasz, L. Esposito, D. Markovitsi, and R. Improta, *Signal Transduct. Target. Ther.*, 2017, **2**, 17021.
6. P. M. Keane, M. Wojdyla, G. W. Doorley, J. M. Kelly, I. P. Clark, A. W. Parker, G. M. Greetham, M. Towrie, L. M. Magno, and S. J. Quinn, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6307.
7. P. M. Keane, M. Wojdyla, G. W. Doorley, J. M. Kelly, A. W. Parker, I. P. Clark, G. M. Greetham, M. Towrie, L. M. Magno, and S. J. Quinn, *Chem. Commun.*, 2014, **50**, 2990–2992.

8. P. ALEXANDER and H. MOROSON, *Nature*, 1960, **185**, 678.
9. R. Das Gupta and S. Mitra, *Biochim. Biophys. Acta*, 1974, **374**, 145–158.
10. H. Chen, R. Li, S. Li, J. Andréasson, and J. H. Choi, *J. Am. Chem. Soc.*, 2017, **139**, 1380–1383.
11. T. Gustavsson, R. Improta, and D. Markovitsi, *J. Phys. Chem. Lett.*, 2010, **1**, 2025–2030.
12. J. Hu, C. P. Selby, S. Adar, O. Adebali, and A. Sancar, *J. Biol. Chem.*, 2017, **292**, 15588–15597.
13. S. Benabou, A. Aviñó, S. Lyonnais, C. González, R. Eritja, A. De Juan, and R. Gargallo, *Biochimie*, 2017, **140**, 20–33.
14. S. Benabou, A. Aviñó, R. Eritja, C. González, and R. Gargallo, *RSC Adv.*, 2014, **4**, 26956–26980.
15. M. Garavís, N. Escaja, V. Gabelica, A. Villasante, and C. González, *Chem. - A Eur. J.*, 2015, **21**, 9816–9824.
16. M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger, and D. Christ, *Nat. Chem.*, 2018.
17. M. Kaushik, S. Kaushik, K. Roy, A. Singh, S. Mahendru, M. Kumar, S. Chaudhary, S. Ahmed, and S. Kukreti, *Biochem. Biophys. Reports*, 2016, **5**, 388–395.
18. N. Saini, Y. Zhang, K. Usdin, and K. S. Lobachev, *Biochimie*, 2013, **95**, 117–123.
19. H. a. Day, P. Pavlou, and Z. a E. Waller, *Bioorganic Med. Chem.*, 2014, **22**, 4407–4418.
20. S. Benabou, M. Garavís, S. Lyonnais, R. Eritja, C. González, and R. Gargallo, *Phys. Chem. Chem. Phys.*, 2016, **18**, 7997–8004.
21. S. Benabou, R. Ferreira, A. Aviñó, C. González, S. Lyonnais, M. Solà, R. Eritja, J. Jaumot, and R. Gargallo, *Biochim. Biophys. Acta - Gen. Subj.*, 2014, **1840**, 41–52.
22. E. Bell, L. Chen, T. Liu, G. M. Marshall, J. Lunec, and D. A. Tweddle, *Cancer Lett.*, 2010, **293**, 144–157.
23. C. Zscherp and A. Barth, *Biochemistry*, 2001, **40**, 1875–1883.
24. A. Mezzetti and W. Leibl, *Eur. Biophys. J.*, 2005, **34**, 921–936.
25. A. Mezzetti, E. Nabedryk, J. Breton, M. Y. Okamura, M. L. Paddock, G. Giacometti, and W. Leibl, *Biochim. Biophys. Acta - Bioenerg.*, 2002, **1553**, 320–330.
26. E. Nabedryk, M. L. Paddock, M. Y. Okamura, and J. Breton, *Biochemistry*, 2005, **44**, 14519–14527.
27. H. Khesbak, O. Savchuk, S. Tsushima, and K. Fahmy, *J. Am. Chem. Soc.*, 2011, **133**, 5834–5842.
28. A. de Juan and R. Tauler, *Crit. Rev. Anal. Chem.*, 2006, **36**, 163–176.
29. L. Blanchet, C. Ruckebusch, J. P. Huvenne, and A. de Juan, *Chemom. Intell. Lab. Syst.*, 2007, **89**, 26–35.
30. J. Ravanat, T. Douki, and J. Cadet, *J. Photochem. Photobiol. B Biol.*, 2001, **63**, 88–102.
31. W. A. Kibbe, *Nucleic Acids Res.*, 2007, **35**, 43–46.
32. P. A. Gorry, *Anal. Chem.*, 1990, **62**, 570–573.
33. S. Navea, A. De Juan, and R. Tauler, *Anal. Chem.*, 2003, **75**, 5592–5601.
34. A. De Juan and R. Tauler, *Anal. Chim. Acta*, 2003, **500**, 195–210.
35. R. Tauler, *Chemom. Intell. Lab. Syst.*, 1995, **30**, 133–146.
36. T. Li, L. Zhang, J. Ai, S. Dong, and E. Wang, *ACS Nano*, 2011, **5**, 6334–6338.
37. J. Jaumot, N. Escaja, R. Gargallo, C. González, E. Pedroso, and R. Tauler, *Nucleic Acids Res.*, 2002, **30**, e92.
38. J. Jaumot, V. Marchan, R. Gargallo, A. Grandas, and R. Tauler, *Anal. Chem.*, 2005, **76**, 7094–7101.
39. A. de Juan, S. C. Rutan, and R. Tauler, in *Comprehensive Chemometrics*, Elsevier, Oxford, 2009, pp. 325–344.
40. R. Tauler, M. Maeder, and A. de Juan, in *Comprehensive Chemometrics*, Elsevier, Oxford, 2009, pp. 473–505.
41. M. Maeder, *Anal. Chem.*, 1987, **59**, 527–530.
42. A. De Juan, M. Maeder, M. Martínez, and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2000, **54**, 123–141.
43. S. Steenken, *Chem. Rev.*, 1989, **89**, 503–520.
44. R. Tauler, A. Smilde, and B. Kowalski, *J. Chemom.*, 1995, **9**, 31–58.
45. M. Maeder and A. D. Zuberbühler, *Anal. Chem.*, 1990, **62**, 2220–2224.
46. A. De Juan, M. Maeder, M. Martínez, and R. Tauler, *Anal. Chim. Acta*, 2001, **442**, 337–350.
47. S. Mas, R. Tauler, and A. de Juan, *J. Chromatogr. A*, 2011, **1218**, 9260–9268.
48. J. Jaumot, A. de Juan, and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2015, **140**, 1–12.
49. F. Geinguenaud, J. Liquier, M. G. Brevnov, V. Olga, Y. I. Alexeev, E. S. Gromova, E. Taillandier, and O. V Petrauskene, 2000, 12650–12658.
50. M. Banyay, M. Sarkar, and A. Gräslund, *Biophys. Chem.*, 2003, **104**, 477–488.
51. L. Blanchet, A. Mezzetti, C. Ruckebusch, J. P. Huvenne, and A. De Juan, *Anal. Bioanal. Chem.*, 2007, **387**, 1863–1873.
52. L. Blanchet, C. Ruckebusch, A. Mezzetti, J. P. Huvenne, and A. De Juan, *J. Phys. Chem. B*, 2009, **113**, 6031–6040.
53. D. G. T. Su, J. L. F. Kao, M. L. Gross, and J. S. A. Taylor, *J. Am. Chem. Soc.*, 2008, **130**, 11328–11337.
54. L. M. Baggesen, S. V. Hoffmann, and S. B. Nielsen, *Photochem. Photobiol.*, 2014, **90**, 99–106.
55. T. Douki, *Photochem. Photobiol. Sci.*, 2013, **12**, 1286–1302.
56. T. Douki and J. Cadet, *Biochemistry*, 2001, **40**, 2495–2501.
57. S. P. Gurung, C. Schwarz, J. P. Hall, C. J. Cardin, and J. A. Brazier, *Chem. Commun.*, 2015, **51**, 5630–5632.