
Article:
Fritzsch, Robby, Donaldson, Paul M, Greetham, Gregory M et al. (4 more authors) (2018) Rapid Screening of DNA-Ligand Complexes via 2D-IR Spectroscopy and ANOVA-PCA. Analytical Chemistry. pp. 2732-2740. ISSN 0003-2700

https://doi.org/10.1021/acs.analchem.7b04727

---

analytical chemistry

Article

Cite This: Anal. Chem. 2018, 90, 2732−2740

pubs.acs.org/ac

Rapid Screening of DNA−Ligand Complexes via 2D-IR Spectroscopy and ANOVA−PCA

Robby Fritzsch,[†] Paul M. Donaldson,[‡] Gregory M. Greetham,[‡] Michael Towrie,[‡] Anthony W. Parker,[‡] Matthew J. Baker,[§] and Neil T. Hunt*[,†]

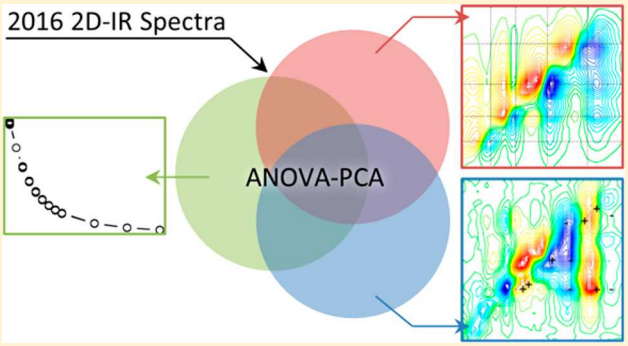
[†]Department of Physics, University of Strathclyde, SUPA, 107 Rottenrow East, Glasgow G4 0NG, U.K.

[‡]STFC Central Laser Facility, Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxon OX11 0QX, U.K.

[§]WestCHEM, Department of Pure and Applied Chemistry, Technology and Innovation Centre, University of Strathclyde, 99 George Street, Glasgow G1 1RD, U.K.

Supporting Information

ABSTRACT: Two-dimensional infrared spectroscopy (2D-IR) is well established as a specialized, high-end technique for measuring structural and solvation dynamics of biological molecules. Recent technological developments now make it possible to acquire time-resolved 2D-IR spectra within seconds, and this opens up the possibility of screening-type applications comparing spectra spanning multiple samples. However, such applications bring new challenges associated with finding accurate, efficient methodologies to analyze large data sets in a timely, informative manner. Here, we demonstrate such an application by screening 2016 2D-IR spectra of 12 double-stranded DNA oligonucleotides obtained in the presence and absence of binding therapeutic molecule Hoechst 33258. By applying analysis of variance combined with principal component analysis (ANOVA−PCA) to 2D-IR data for the first time, we demonstrate the ability to efficiently retrieve the base composition of a DNA sequence and discriminate ligand−DNA complexes from unbound sequences. We further show accurate differentiation of the induced-fit and rigid-body binding modes that is key to identifying optimal binding interactions of Hoechst 33258, while ANOVA−PCA results across the full sequence range correlate directly with thermodynamic indicators of ligand-binding strength that require significantly longer data acquisition times to obtain.



Molecular interactions of proteins and DNA are central to biological function, but our ability to understand these interactions in ever greater detail is closely linked to the development of new analytical technologies. Through direct measurement of the coupling of vibrational modes and the ability to reveal molecular dynamics with ∼100 fs time resolution, two-dimensional infrared (2D-IR) spectroscopy has provided new insights into the structure, dynamics, and function of a range of proteins[1−5] and enzymes.[6,7] Here, we focus upon applications of 2D-IR to study DNA, building on studies revealing the impact of Watson−Crick base pairing on the vibrational spectroscopy and solvation dynamics of DNA[8−11] and recent insights into the subtle structural changes induced in the double helix by ligand binding to the minor groove of B-type DNA.[12]

Alongside development of experimental methods and spectral interpretation, recent technological advances have significantly decreased the time needed to acquire a 2D-IR spectrum. The advent of mid-IR pulse shaping enabled spectral acquisition within 70 s,[13] while laser systems with pulse repetition rates of 100 kHz have further reduced spectral acquisition times to a few seconds.[14,15] This enhanced data collection efficiency opens up the possibility of applying 2D-IR in a more analytical context, for example, as a tool for screening novel bioactive compounds. However, such an application will produce large sets of data that need to be cross-compared to separate out and identify specific samples that fit the required category, meaning that currently employed data analysis approaches will be prohibitively slow and laborious. To realize the analytical potential of 2D-IR spectroscopy, it is therefore essential that data analysis tools advance in parallel with technology. In this study, the 2D-IR spectra of 12 double-stranded (ds)DNA oligomers (see Figure 1 for sequences) were measured in the presence and absence of binding molecule Hoechst 33258 (H33258) at a series of waiting times to retrieve sequence-specific information about ligand binding. Spectral replicates were treated as individual data

Received: November 15, 2017
Accepted: January 23, 2018
Published: January 23, 2018

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732−2740

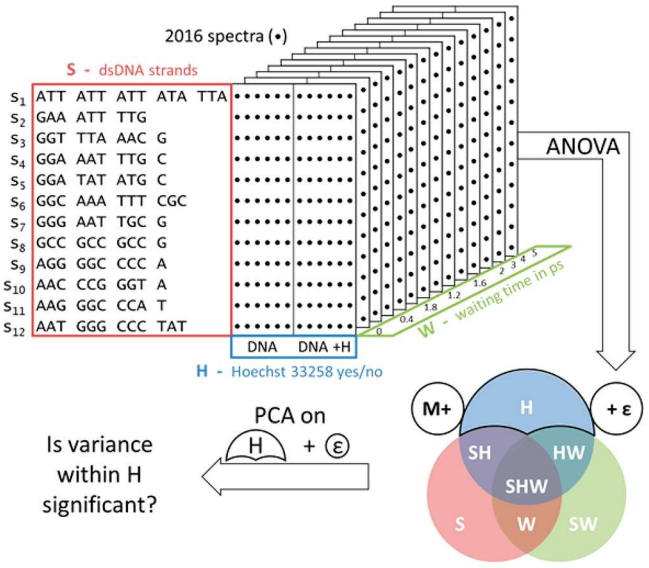


Figure 1. Experimental design and schematic representation of the ANOVA–PCA method. The set of 2D-IR spectra studied contains three main sources of variance (factors): the sequence of dsDNA, ***S***, the presence of H33258, ***H***, and the waiting time, ***W***. The ANOVA method generates subsets containing the variances attributable to each factor as well as the variance due to their interactions (e.g., (***SH***); the sequence-dependent impact of ligand binding) and the residual variance, ***ε*** (e.g., noise). Individual subsets are compared to residuals, ***ε***, and analyzed using PCA to test for significance.

points rather than being averaged, resulting in a stack of 2016 2D-IR spectra (Figure 1).

To analyze this data set we have employed analysis of variance combined with principal component analysis (ANOVA–PCA). This technique was developed to study large mass-spectrometry data sets analyzing amniotic fluid and has subsequently been applied in a variety of scenarios[16−18] including mid-infrared absorption experiments.[19] The essence of ANOVA–PCA is the separation of sources of variance within the data set according to experimental factors, which are controlled and systematically changed during the experiment. By using PCA to compare these subsets to a residual variance (e.g., spectral noise), ANOVA–PCA tests each subset for significance and analyzes them independently to reveal their individual influence on the data set (Figure 1).

H33258 has been used both as a DNA binding dye and therapeutic molecule[20,21] and binds preferentially to A/T-rich regions of dsDNA,[22−27] showing particular affinity for A-tract sequences (A_nT_n motifs). It has been suggested that A-tract binding is a rigid body interaction promoted by a narrow minor groove,[26] but it has recently been shown that H33258 has very subtle effects on the propeller twists of A-tract dsDNA[28] that are reminiscent of an induced-fit interaction for optimal combinations of ligand and base sequences.[12] The ability of 2D-IR to resolve structural alterations of DNA sequences when bound to H33258 makes this couple an ideal system for testing the discriminating capability of ANOVA–PCA. We show that ANOVA–PCA, acting on the main variance factors (Figure 1), can differentiate between base sequence compositions of DNA oligomers and separate the spectra of those complexed from ligand-free sequences. Analysis of factor interactions reveal sequence-specific differences in the 2D-IR spectroscopy of DNA-H33258 complexes that correlate with traditional thermodynamic measurements of binding affinity, showing that 2D-IR and ANOVA–PCA has the potential for use as a rapid and effective probe of DNA binding interactions.

EXPERIMENTAL METHODS

Materials and Sample Preparation. Twelve complementary pairs of DNA oligonucleotides (salt-free, lyophilized) were purchased from Eurofins. TRIS base, NaCl, D_2O, DCl, and H33258 were obtained from Sigma-Aldrich. All chemicals were used without further purification.

Stock solutions of each oligomer (40 mM) were prepared using a deuterated TRIS buffer (100 mM TRIS, 100 mM NaCl, pD 7.0). Solutions of complementary single strands were mixed in equimolar ratios, diluted, and annealed at 95 °C for 5 min to form dsDNA samples. Samples containing H33258 were prepared by annealing the dsDNA in an equimolar ratio with a stock solution of H33258 in TRIS buffer. The final dsDNA concentration for all samples measured was 10 mM.

To characterize the binding affinity of H33258 with dsDNA, IR absorption experiments as a function of temperature were obtained. The IR absorbance of each dsDNA sequence was measured from 20 to 95 °C in 5 °C steps with and without H33258 at the same concentrations as used for the 2D-IR experiments. The melting temperature of dsDNA was obtained by applying PCA and fitting the temperature-induced variance to a sigmoidal function as previously reported.[11] The melting temperature stabilization (increase) caused by the ligand was used as an indicator of binding affinity. IR absorption measurements were also compared with UV–vis absorption and fluorescence measurements to ensure accurate sample characterization.

Infrared Spectroscopy. Samples for IR absorption and 2D-IR spectroscopy were held in a transmission cell consisting of two CaF_2 windows separated by a polytetrafluoroethylene (PTFE) spacer of 25 μm thickness. Fourier transform infrared (FT-IR) absorption spectra were acquired using a Thermo Scientific Nicolet iS10 Spectrometer at a resolution of 4 cm^{-1} between 1550 and 1800 cm^{-1}.

Time-resolved 2D-IR experiments were obtained using the STFC-Central Laser Facility LIFEtime instrument and the FT-2D-IR technique using a pseudo pump–probe beam geometry described in detail elsewhere.[14,29] Briefly, LIFEtime consists of two 100 kHz pulse repetition rate Yb:KGW amplified laser systems pumping optical parametric amplifiers equipped with difference-frequency generation. Mid-IR pulses centered at 1650 cm^{-1}, resonant with carbonyl and ring stretching modes of the nucleobases, were used for all experiments. The output of one OPA was directed into a mid-IR pulse shaper to generate the first two (pump) pulses of the 2D-IR experiment and scan the time delay (τ) between them. The output of the second OPA was used to provide the third (probe) pulse, which was delayed relative to the second pump pulse to control the waiting time, *W*, using an optical delay line.

A set of 14 waiting times from 0 to 5 ps were collected for each sample. ZZZZ polarization was used throughout, and a phase cycling methodology was employed to reduce any scattering of light from the samples.[13] Pulse durations of 300 fs for the pump and 200 fs for the probe limited the experimental time resolution. Using this approach, one 2D-IR spectrum at a fixed waiting time was acquired within 40 s (2 million laser shots), though little substantive improvement in the S/N ratio was observed after ∼10 s of acquisition. Six replicates were obtained for each sample and treated as individual spectra for

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732−2740

the purposes of data analysis as a measure of instrumental repeatability.

Data Processing and Analysis—ANOVA−PCA. The statistical analysis software R[30] has been used to process and analyze the data using a custom-made script. Spectra obtained with a waiting time of 600 fs were vector normalized (rescales each spectrum to have equal sum of squares of one) to minimize errors arising from sample thickness, concentration, and sequence length and to ensure that all DNA sequences have equal weighting when performing the PCA. The signal amplitude scaling between waiting times for a given DNA sequence was kept identical to that of the raw data to ensure that time-dependent information such as vibrational relaxation dynamics was preserved. Data with waiting times shorter than 400 fs were excluded from the analysis to remove effects due to pulse overlap.

Each 2D-IR spectrum was concatenated into vector-form so that each pixel is treated as an independent variable (Figure 2).

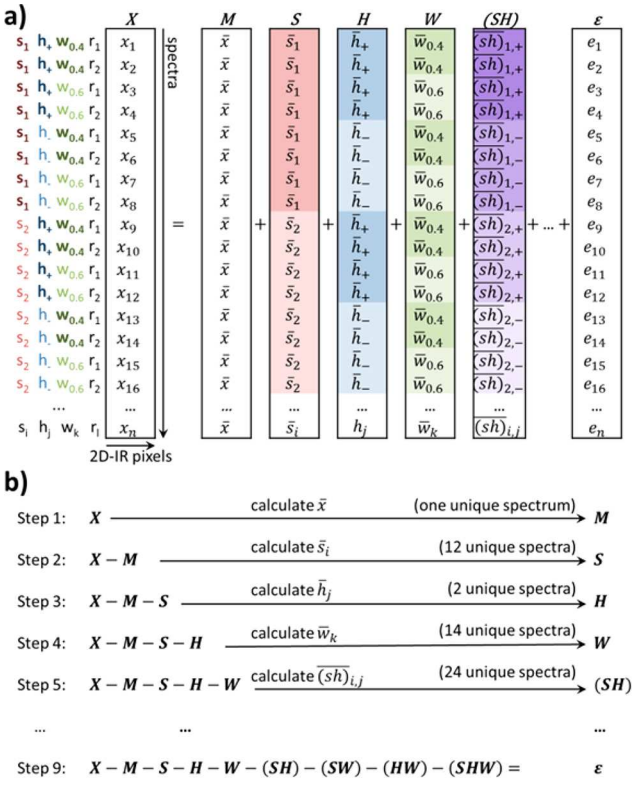


Figure 2. (a) Schematic representation of the data matrix $\boldsymbol{X}$ and its decomposition into factor matrices; blocks of color indicate averaged spectra in each factor matrix. Each row represents one 2D-IR spectrum concatenated to a vector. (b) Schematic structure of the stepwise subtraction to calculate factor matrices. The algorithm allocates variance from the raw data based on their origin into several new matrices with the same dimensions by calculating averages. The residual matrix ideally just contains the variance between repeats.

The complete data set, including spectra of 12 dsDNA sequences, with and without H33258, encompassing 12 waiting times thus forms a matrix, $\boldsymbol{X}$, where individual 2D-IR spectra form rows, and each column represents the pixel index.

Separation of $\boldsymbol{X}$ into Subsets. The format of $\boldsymbol{X}$ is shown schematically in Figure 2a. Each individual spectrum in $\boldsymbol{X}$ is represented by x_n (n = 1−2016) spectra. The experiment is set up to investigate the effect of three main factors on the 2D-IR spectrum. These are the sequence of the DNA strand, S, the presence of the ligand, H, and the waiting time, W. Thus, each row of $\boldsymbol{X}$, x_n, is described in terms of the sequence of the dsDNA, s_i (i = 1−12 sequences); the presence of the binder H33258, h_j (j = ± for ligand presence/absence), and the waiting time, which is indicated by w_k (k = 0.4−5, where the value indicates the waiting time in ps). Each sample was repeated six times, indicated by r_l (l = 1−6) Thus, the second replicate of sequence 5, with H33258 present at a waiting time of 0.8 ps is indicated in this notation by $s_5h_+w_{0.8}r_2$.

Analysis of variance was carried out on matrix $\boldsymbol{X}$ according to established methods.[16] The structure of the experimental design is used to sequentially calculate a series of matrices from $\boldsymbol{X}$ based on the three factors defined in the experiment (see Figure 2a). These matrices are then subtracted from $\boldsymbol{X}$ to obtain the residual matrix, $\boldsymbol{\varepsilon}$. The outcome of the whole procedure is thus to separate the variance of $\boldsymbol{X}$ into matrices of three main experimental factors $\boldsymbol{S}$, $\boldsymbol{H}$, and $\boldsymbol{W}$; four interactions $(\boldsymbol{SH})$, $(\boldsymbol{SW})$, $(\boldsymbol{HW})$, and $(\boldsymbol{SHW})$, and residual matrix $\boldsymbol{\varepsilon}$ based on known information about the data set

$$\boldsymbol{X} = \boldsymbol{M} + \boldsymbol{S} + \boldsymbol{H} + \boldsymbol{W} + (\boldsymbol{SH}) + (\boldsymbol{SW}) + (\boldsymbol{HW}) + (\boldsymbol{SHW}) + \boldsymbol{\varepsilon} \quad (1)$$

The matrices for eq 1 are calculated stepwise. In the first step (Figure 2b), an average spectrum $\bar{x}$ for all spectra in $\boldsymbol{X}$ is calculated, and a new matrix $\boldsymbol{M}$ is created. $\boldsymbol{M}$ has the same dimensions as $\boldsymbol{X}$ but contains $\bar{x}$ for every row as shown in Figure 2a. Subtraction $\boldsymbol{X} - \boldsymbol{M}$ generates a residual matrix, which is used in subsequent steps. In step 2, the matrix representing the sequence factor, $\boldsymbol{S}$, is generated. All rows of $\boldsymbol{X} - \boldsymbol{M}$ belonging to the same sequence s_i are averaged to give spectra $\bar{s}_i$. The 12 average spectra $\bar{s}_i$ are repeated according to their dedicated rows to form matrix $\boldsymbol{S}$. The subtraction $\boldsymbol{X} - \boldsymbol{M} - \boldsymbol{S}$ then results in a new residual matrix containing reduced variance. In step 3, $\boldsymbol{X} - \boldsymbol{M} - \boldsymbol{S}$ is used to calculate two average spectra $\bar{h}_j$, thus forming matrix $\boldsymbol{H}$ accounting for the ligand factor. $\boldsymbol{X} - \boldsymbol{M} - \boldsymbol{S} - \boldsymbol{H}$ is then used to create matrix $\boldsymbol{W}$, in which all spectra with a common waiting time (w_k) are averaged. This is subtracted in turn accounting for the factor waiting time.

The spectral impact of different combinations of the three main factors is then considered by interactions between factors, where average spectra are calculated for each possible combination (e.g., $(\boldsymbol{SH})$ consists of 24 unique spectra $\overline{(sh)}_{i,j}$ for 12 sequences (i) each measured with and without H33258 (j)). The method is repeated until matrices for all factors and interactions are calculated and subtracted from the raw data to give residual matrix $\boldsymbol{\varepsilon}$, which theoretically contains only instrumental noise.

It is important to note that this method requires a balanced data set, meaning that the number of spectra analyzed is exactly the same for each level of each factor.[16] If this is not the case, the order of the sequential subtraction will have an impact on the results, and the variance of one factor may end up in the matrix of another. This has implications for outliers. Three out of six repeats of one DNA spectrum were obscured by scatter-artifacts (apparent from visual inspection) and had to be excluded. To maintain a balanced data set, three repeats from all other measurements had to be removed, reducing the total number of spectra analyzed to 864.

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732−2740

Principal Component Analysis. Once the factor and interaction matrices are assembled, they are tested for significance using PCA. One or more factors or interactions are added to the residual matrix, $\boldsymbol{\varepsilon}$, and a PCA is performed to test whether the principal variance of this subset is dominated by noise or a systematic change. If the variance of any systematic effects from the factor outweighs the residuals, then the scores of the first principal component (PC1) will be dominated by the factor. Conversely, if the variance of the residuals outweighs any systematic effects from the factor, the scores of the first principal component (PC1) will be dominated by noise from the residuals, and the factor has no significant impact on the spectrum. By adding more than one factor or interaction to the residuals, one can gradually increase the complexity of the data analyzed.

For comparison with the 2D-IR results, the ANOVA–PCA analysis was also performed using IR absorption data. The results are consistent and shown in Figures S-1 to S-4.

RESULTS AND DISCUSSION

Base Sequence Dependence. Before the ANOVA–PCA method can be used to determine previously unobtainable information from the data set, it is important to validate it against known results. The first step is to determine whether the ANOVA–PCA accurately recovers information from the 2D-IR spectra relating to the DNA base composition of the sample or, more formally, whether the main factor $\boldsymbol{S}$ is significant. The score plot of the first two principal components (PCs) of $\boldsymbol{S} + \boldsymbol{\varepsilon}$ (Figure 3a) shows that every spectrum falls into

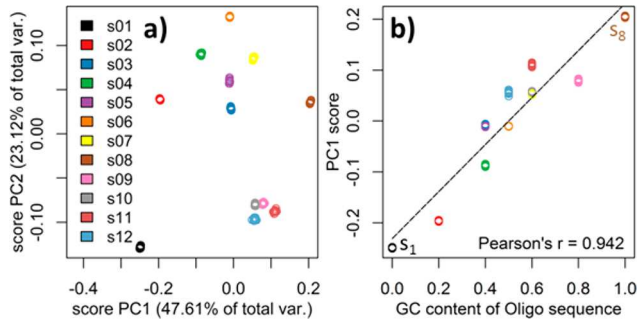


Figure 3. (a) ANOVA–PCA results for dsDNA sequence dependence $\boldsymbol{S} + \boldsymbol{\varepsilon}$. Each point in the scores plot represents one 2D-IR spectrum and similar scores between spectra indicate common spectral features. (b) The scores in PC1 show a good correlation (Pearson's correlation coefficient of 0.94) to the G/C content of the sequences measured.

one of 12 clearly separated groups representing the 12 average spectra of matrix $\boldsymbol{S}$. The residuals in $\boldsymbol{\varepsilon}$ create the variance within each group and show that the dsDNA strands can be differentiated from one another by using 2D-IR data. The spread of each point cloud indicates the variance of the spectral response relative to the differentiation of the sequences.

Plotting the PC1 score of each spectrum against the G/C content of each sequence (Figure 3b) shows a linear correlation, indicating that the method accurately determines the base composition of a given sequence. Some of the DNA sequences studied contain the same base-pair composition but different sequences. These appear in Figure 3b at G/C content values near 0.5 and can be seen to possess slightly different PC1 scores. This indicates that PC1 is not completely independent from the order of the bases, suggesting that further spectral variation arising from nearest-neighbor effects in the sequences will be visible.

Having ascertained that PC1 describes the overall base-pair composition, spectra for the A–T and G–C base pairs were obtained by multiplying the PC1 score of DNA sequences s_1 (all A/T) and s_8 (all G/C) with the loading vector of PC1 (p_1) and adding the global mean spectrum ($PC1$ score $\cdot\ p_1 + \bar{x}$). These calculated spectra (Figure 4a,b) are compared to the raw

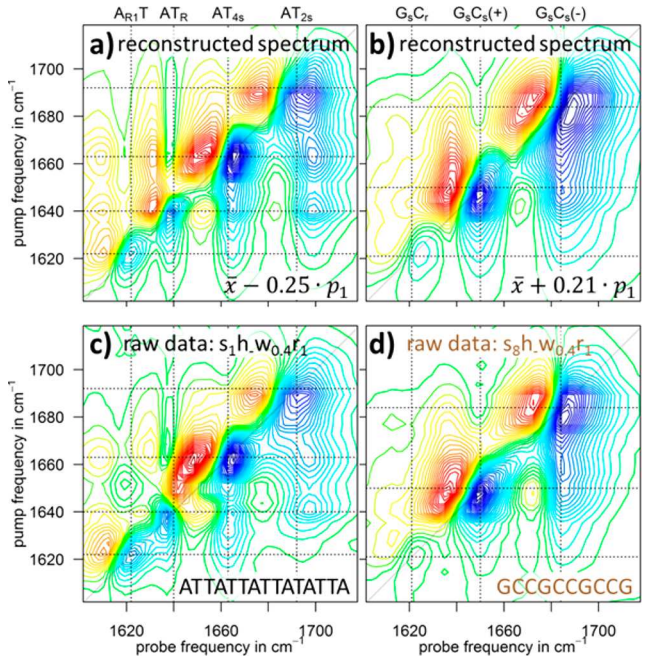


Figure 4. Top: Reconstructed spectra using the average PC1 scores of s_1 (all A/T) (a) and s_8 (all G/C) (b). Peak assignments according to literature are shown.[8,29] $\bar{x}$ and p_1 are vectors defined as the global average spectrum and the loading vector of PC1, respectively. Bottom: Raw 2D-IR data at 0.4 ps of (c) DNA sequence s_1 consisting purely of A–T base pairs and (d) sequence s_8 with only G–C base pairs.

spectra of s_1 and s_8 at 400 fs (Figure 4c,d). The agreement is excellent, showing that the ANOVA–PCA method can extract the sequence-related aspects of the DNA spectra. However, it is noted that as the calculated A–T and G–C base-pair spectra (Figure 4a,b) show a general spectral response of the Watson–Crick base pairs independent from the DNA sequence then they will provide a useful comparison to allow differentiation of spectral features arising from interactions along the strand. The black gridlines in Figure 4a,d indicate literature values for CO stretching and ring modes of the nucleobases and agree very well within ±2 cm^{-1} of the peak positions in the reconstructed spectra.

Spectral Change upon Addition of H33258. Having established that ANOVA–PCA can differentiate DNA sequences by base composition, we move on to further test the method by using it to differentiate the spectra of ligand-free from complexed sequences. To do this, the main factor matrix $\boldsymbol{H}$ was added to the residual matrix $\boldsymbol{\varepsilon}$ and interrogated with PCA. Previous experiments have shown that the concentration of H33258 used in the samples in this study is too small to be detected by infrared spectroscopy. Any changes in the spectrum can therefore be attributed to changes in DNA vibrations.[12]

The manner in which the main factor matrix $\boldsymbol{H}$ is generated means that it contains only two unique spectra. These

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732–2740

correspond to the average spectrum over all samples with and without H33258, respectively. A PCA on the sum $\boldsymbol{H} + \boldsymbol{\varepsilon}$ shows two clusters of spectra separated by a significant distance in PC1 (Figure 5a). In contrast, the PC2 scores are spread out

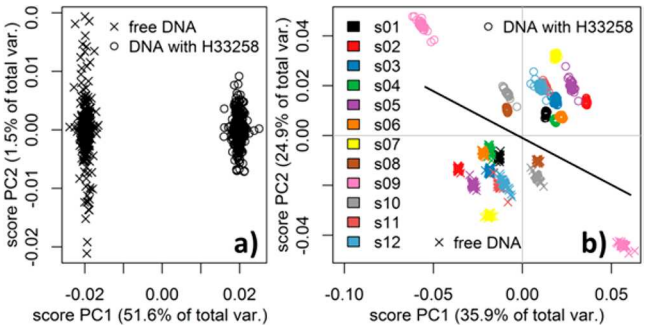


Figure 5. (a) PCA scores plot of subset $\boldsymbol{H} + \boldsymbol{\varepsilon}$. Separation of ligand-bound (circles) and -free (crosses) DNA sequences along PC1 shows a significant change in the 2D-IR spectrum occurs when adding H33258. (b) PCA scores plot for subset $\boldsymbol{H} + (\boldsymbol{SH}) + \boldsymbol{\varepsilon}$. Circles show bound sequences, and crosses show unbound sequences. Individual sequences are separated by color. The bound and unbound sequences are clearly separated into two groups by the black diagonal line.

arbitrarily, and the PC2 loading plot only shows noise. From this result, it can be confirmed that adding a binder has a measurable effect on the vibrational modes of the DNA bases. However, because $\boldsymbol{H}$ only contains responses averaged over all DNA strands, the subset is neither able to explain whether all DNA sequences interact with the binder in the same way or whether some of the sequences interact at all.

Sequence Dependence of H33258 Binding. Although it is clear that the ANOVA–PCA method can differentiate spectra containing H33258 from those without, this does not provide sequence-dependent information. To investigate base sequence-dependent impacts of H33258 binding, it is necessary to include the interaction matrix $(\boldsymbol{SH})$. A PCA was performed on the sum $\boldsymbol{H} + (\boldsymbol{SH}) + \boldsymbol{\varepsilon}$. The result (Figure 5b) is that the scores plot now shows individual clusters corresponding to each base sequence (separated by color in Figure 5b) and that spectra where H33258 was present (circles) are differentiated from those without the ligand (crosses). This clearly shows that the spectral impact of H33258 binding varies according to base sequence. Spectra of free DNA tend to have negative PC1 and PC2 scores and can be clearly separated from spectra of DNA with ligand (black line), because they essentially invert their sign, and higher principal components indicate that there is even more spectral information available for this differentiation.

The sign inversion can be explained from the experimental setup: factor $\boldsymbol{H}$ only consists of two unique spectra, which were calculated from data where the average spectrum of each sequence was subtracted (for calculation of $\boldsymbol{S}$). The analyzed subset $\boldsymbol{H} + (\boldsymbol{SH}) + \boldsymbol{\varepsilon}$ is therefore centered with respect to the sequence, and every variance in $(\boldsymbol{SH})$ must have a counterpart of equal magnitude.

In an effort to explain the distribution by sequence shown in Figure 5b, we compare the results from ANOVA–PCA with FT-IR measurements of the melting temperature stabilization (ΔT_m) of each of the sequences studied upon binding of H33258. The results are shown in Figure 6a and can be used as a proxy for binding affinity of the ligand to a given sequence. The length of the bars shows the observed change in dsDNA melting temperature, ΔT_m. Strong interactions are indicated by

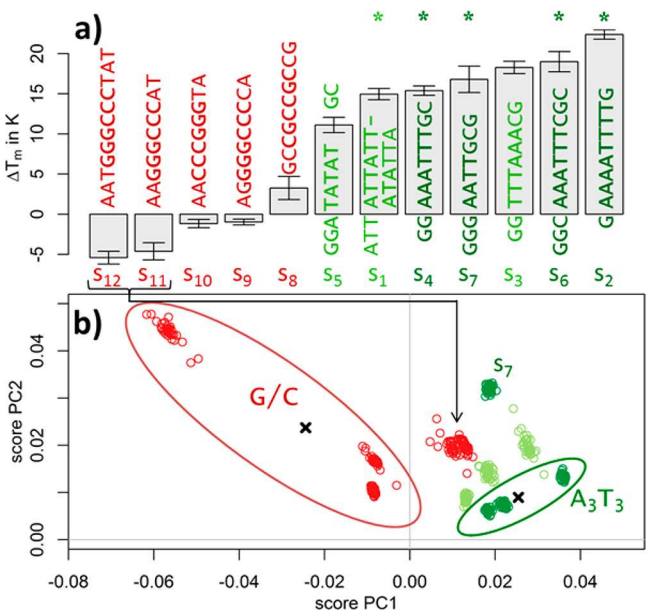


Figure 6. (a) The change in dsDNA melting temperature upon addition of H33258 as observed by FT-IR spectroscopy for all DNA sequences studied. Asterisks indicate sequences exhibiting an increase in H33258 fluorescence greater than 10-fold (see Figure S-5). (b) PCA scores plot for $\boldsymbol{H} + (\boldsymbol{SH}) + \boldsymbol{\varepsilon}$. Scores for free DNA spectra omitted. Ellipses show two groups of distinctive DNA–ligand interactions. Colors qualitatively represent binding affinity for H33258: red corresponds to insignificant minor-groove binding; dark green indicates classic minor-groove binding (see text).

a positive ΔT_m, as seen in dsDNA strands s_1 to s_7, each of which features either an A-tract (A_nT_n where $n \geq 2$) or an A/T-rich section of the sequence. The results are consistent with what is known from previous studies.[26,31,32] Those sequences exhibiting a 10-fold or greater increase in H33258 fluorescence upon binding, which indicates minor groove binding, are indicated with an asterisk (see also Figure S-5). Adding the ligand to G/C-rich DNA sequences (s_8 to s_{12}) either lowered the melting point of the dsDNA or had no effect. To correlate the results of the thermodynamic measurements with the scores plot derived from ANOVA–PCA, Figure 6b shows a simplified depiction of Figure 5b that only includes spectra of DNA with H33258 bound. The color scheme is a qualitative representation of the ΔT_m value obtained for each sequence. Red indicates zero, small, or negative ΔT_m, while green indicates a large, positive ΔT_m. A darker green furthermore indicates sequences with the classic A-tract binding site. Thus, sequences with darker green colors exhibit closer to optimal binding conditions for H33258.

Figure 6b shows that, according to ANOVA–PCA, sequences that experience a large ΔT_m have positive PC1 scores, whereas those exhibiting small/negative ΔT_m values produce a negative score in PC1. Further, sequences with well-defined A_nT_n sites of length 3 and higher (green ellipse) have comparably low scores in PC2. This suggests the important result of a correlation of the PC1 and PC2 scores with the thermodynamic measurement of binding affinity as would be required for a successful validation of ANOVA–PCA and 2D-IR as an analytical technique.

While there seems to be a general relation between the first two principal components and the binding affinity, this alone is perhaps too simplistic to completely explain the observed interactions. For example dsDNA strands with oligomers

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732–2740

AAGGGCCCAT (s_{11}) and AATGGGCCCTAT (s_{12}) show positive PC1 scores (black arrow) even though the negative ΔT_m and the lack of increased fluorescence indicates no minor-groove binding. It is perhaps relevant that both s_{11} and s_{12} sequences contain A/T substructures at the ends of the double strands. As the minor groove only forms for sequence lengths of >3 bases, this motif does not represent a classic binding site for H33258. It could however be possible that the ligand interacts with these A/T-rich ends rather than the inaccessible G/C cores leading to a spectroscopic effect recognized by ANOVA−PCA as binding to an A/T-rich sequence. The melting point stabilization seems to support this picture: While successful minor-groove binding stabilizes the double helix and increases the melting point (s_1 to s_7), interactions at the end of the strands may facilitate end fraying of the double helix and decrease the melting temperature in s_{11} and s_{12}. This is consistent with our observed deviation of their PC1 score with respect to the scores of s_8, s_9, and s_{10}.

The dsDNA oligomer s_7 (GGGAATTGCG) has a minor groove suitable for H33258 but reaches higher PC2 scores than other A_nT_n sequences. Previous studies suggest that H33258 spans approximately five base pairs along the minor groove.[24] The target sequence in this strand is only four base pairs long, and the binder is likely to be in close proximity to a G−C base pair. This would be expected to change the spectral response upon binding in comparison to strands with a longer target sequence such as s_2, s_4, and s_6.

It is important to note that subset $\boldsymbol{H} + (\boldsymbol{SH}) + \boldsymbol{\varepsilon}$ does not simply represent a dependence on the G−C or A−T base-pair amount. This dependency has been subtracted from the data when calculating $\boldsymbol{S}$, and the PC scores from $\boldsymbol{H} + (\boldsymbol{SH}) + \boldsymbol{\varepsilon}$ do not show a clear correlation to the base-pair composition as the PC1 scores of subset $\boldsymbol{S} + \boldsymbol{\varepsilon}$.

The scores plot in Figure 6b suggests that there are at least two distinctive groups of binding interactions; one being the response of DNA strands containing A_nT_n ($n \geq 3$, s_2, s_4, s_6) and the other are strands with a G/C core and a negative PC1 score (s_8, s_9, s_{10}). These are shown as green and red ovals in Figure 6b, and the average PC1 and PC2 scores for each group are marked with a black cross. The average score was used to reconstruct the spectral response for the two binding cases. The aforementioned sign inversion of the scores implies that it is possible to generate difference spectra directly from the scores of the bound spectra. We note that this would not be the case if more than one type of ligand was measured, and distances in the scores plot would need to be used to reproduce a difference spectrum. Average scores of the first ten principal components have been used for the reconstruction to retain as much relevant data as possible. The reconstructed spectra for the two groups are shown in Figure 7.

The spectra can be read like difference 2D-IR spectra, where negative (blue) signals indicate a change to lower amplitudes, and positive (red) signals increases in amplitude due to ligand interactions. Both of the ANOVA−PCA-derived 2D-IR difference spectra show a complex response. The sequences with a G/C-core (red group) show changes in the vibrational modes along the diagonal of the spectrum without apparently affecting the off-diagonal peaks to any great degree (Figure 7b). By contrast, the positive green grouping, which accommodates binding of H33258, leads to significant changes in the off-diagonal region of the 2D-IR spectrum.

The binding of H33258 to sequences s_4 and s_5, which fall into the green group, has been studied recently by 2D-IR

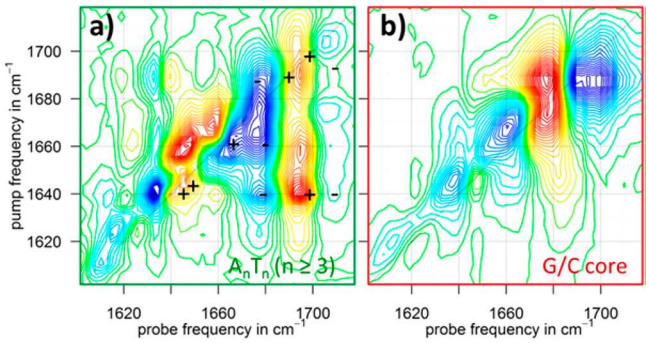


Figure 7. Reconstructed difference spectra for two distinctive DNA−ligand interactions calculated from the first ten principal components of subset $\boldsymbol{H} + (\boldsymbol{SH}) + \boldsymbol{\varepsilon}$. (a) Shows the results for the spectra included in the group identified by the green oval in Figure 6b). (b) Shows the results for the spectra included in the red oval in Figure 6b. Black ± symbols in (a) indicate position and sign of prominent bands in difference 2D-IR spectra relating to H33258 binding to sequence s_4 as reported previously.[12]

spectroscopy using a different instrument with broader pump bandwidth and phosphate buffer rather than a TRIS buffer.[12] The difference spectra of s_4 and s_5 obtained from ANOVA−PCA are in good agreement with these experimental results, again showing that ANOVA−PCA is recovering the salient spectral features arising from this complex intermolecular interaction. Plus and minus symbols marked on Figure 7a show the positions of prominent peaks reported in the difference 2D-IR spectra of H33258 binding to sequence s_4.[12] The reported blue shift of the AT_{2s} mode around 1700 cm^{-1} can be clearly seen in the ANOVA−PCA reconstruction, albeit with reduced amplitude due to smaller bandwidth of the LIFEtime laser (80 cm^{-1} full width half-maximum, fwhm, compared to ca. 300 cm^{-1} fwhm[12]). Understanding the complex, spectral response of the DNA to a ligand requires further analysis, but the results show how ANOVA−PCA can highlight subtle changes due to different DNA−ligand interactions and enables analysis of large data sets of 2D-IR spectra.

Spectral Change due to Waiting Time and DNA Sequence. The last main factor influencing the variance of the data set is the waiting time ($\boldsymbol{W}$) that can be used to obtain information relating to vibrational relaxation mechanisms. In contrast to $\boldsymbol{S}$ and $\boldsymbol{H}$, this factor represents a continuous variable rather than categorical variables and therefore requires a slightly different approach for analysis. The matrix $\boldsymbol{W}$ contains the average variance between waiting times and due to the prior subtraction of $\boldsymbol{M}$, just includes the change from the global average spectrum. By adding $\boldsymbol{M}$ back to $\boldsymbol{W}$, the correct time-dependent behavior is obtained. Subset $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{\varepsilon}$ contains the average evolution of all measured 2D-IR spectra with waiting time, including processes such as vibrational relaxation, spectral diffusion, or energy transfer.[29] As every excited molecular vibration will decay back to the ground state, the first principal component should derive from the principal vibrational lifetime of all modes in the 2D-IR spectrum. Spectral diffusion or energy transfer will have minor impacts on peak shapes and amplitudes and so are likely to be present in higher principal components. This will make such dynamics challenging to extract with an unsupervised method like PCA and will be complicated for ANOVA-type methods because the behavior can be highly mode-specific; a point which we return to later.

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732−2740

For the purposes of this article, we focus on the vibrational relaxation of the DNA modes, as derived from PC1.

The PC1 score of subset $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{\varepsilon}$ is plotted against the corresponding waiting time in Figure 8a, inset. The results

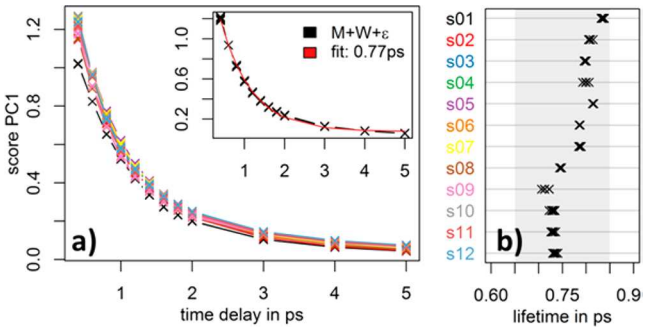


Figure 8. (a, inset) PC1 score of subset $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{\varepsilon}$ plotted against waiting time. (a) PC1 score of subset $\boldsymbol{M} + \boldsymbol{S} + \boldsymbol{W} + (\boldsymbol{SW}) + \boldsymbol{\varepsilon}$ plotted against waiting time. (b) Vibrational relaxation of different DNA sequences obtained from ANOVA–PCA. Monoexponential functions were used for fitting throughout. Gray area illustrates fwhm of the probe pulse.

show a decay that can be well represented by monoexponential function with a lifetime of 770 fs. The spectral features in the loading plot of PC1 closely resemble the global average spectrum in $\boldsymbol{M}$, which is dominated by the $\nu = 0{-}1$ and $\nu = 1{-}2$ transitions on or near the diagonal of the spectrum (see Figure S-6). It is instructive to compare this to previous results. 2D-IR experiments on sequence s_8 (all A/T) reported lifetimes of around 750 ± 200 fs for this region.[29] In that paper, the diagonal peaks of only one sequence were analyzed conventionally in a peak-by-peak manner, and an average lifetime was calculated. This strongly indicates that the PC1 score following ANOVA analysis gives an accurate picture of the principal vibrational relaxation of the data set. It is however noted that the IR spectrum of DNA in the base region may be particularly amenable to this type of treatment because all modes in the spectral region studied exhibit very similar vibrational relaxation dynamics. While this is not unusual for biological molecules, for which the most studied IR modes (e.g., amide I of proteins and base modes of DNA) all show very similar ps time scale vibrational relaxation, it is not clear how ANOVA–PCA will differentiate significant mode-to-mode variation within a data set. This will be a topic for further study.

We now extend the analysis of the $\boldsymbol{M} + \boldsymbol{W}$ subset to investigate sequence-dependent vibrational relaxation. For this, the subset $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{S} + (\boldsymbol{SW}) + \boldsymbol{\varepsilon}$, is considered, where $\boldsymbol{W}$ contains the average relaxation and $(\boldsymbol{SW})$ includes sequence-specific deviations from the average relaxation in $\boldsymbol{W}$. The PC1 score of this subset plotted against waiting time (Figure 8a) shows individual relaxation dynamics for every sequence. A monoexponential fit of these decays obtains very similar lifetimes for the measured dsDNA strands of between 700 and 850 fs and the corresponding loading plot of PC1 again resembles the global average spectrum. Considering the derived lifetimes more carefully (Figure 8b) indicates that G/C-rich sequences (s_8, all G/C, 750 fs) relax slightly more quickly than A/T-rich oligomers (s_1, all A/T, 830 fs). It is noted however that this difference of about 100 fs is just half the size of the fwhm of the probe pulse duration (200 fs, illustrated as gray area in Figure 8b).

This result from ANOVA–PCA delivers reasonable agreement with experimental vibrational lifetimes (830 fs compared to 750 ± 200 fs[29] for s_1), which seem to be largely unaffected by the composition of the DNA. To further validate these findings, a PCA was carried out on the raw data for each sequence individually. The principal variance for these data sets originates from vibrational relaxation, and the scores of PC1 could again be well represented by a monoexponential decay function. The lifetimes found by individual PCA agree very well with the values obtained from the ANOVA–PCA method (Figure S-7). We conclude therefore that our technique is able to distinguish the principal vibrational lifetime of different DNA sequences in a multivariate approach without the need of individual peak-by-peak analysis, with the caveat that mode-to-mode variation will not be clearly identified using PC1. The results indicate a slightly faster vibrational relaxation for GC-rich double strands.

Spectral Change due to Waiting Time and DNA–Ligand Interaction. The final question that we seek to address is to determine whether DNA–ligand interactions affect the vibrational relaxation dynamics of the sequences. Little information is available on this, and so, our ANOVA–PCA study can be used to go beyond validation to provide new insight. In this case, subset $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{H} + (\boldsymbol{HW}) + \boldsymbol{\varepsilon}$ was analyzed to detect any overall difference in the vibrational relaxation of free and ligand-bound DNA. It is noted that no sequence component is present in this subset. The PC1 score (see inset of Figure 9a) returns decay functions with virtually identical lifetimes for free- (770 fs) and ligand-bound DNA (780 fs). A comparison to the result from subset $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{\varepsilon}$ shows how ligand-specific matrix $(\boldsymbol{HW})$ contains negligible variance compared to the average relaxation matrix $\boldsymbol{W}$. This

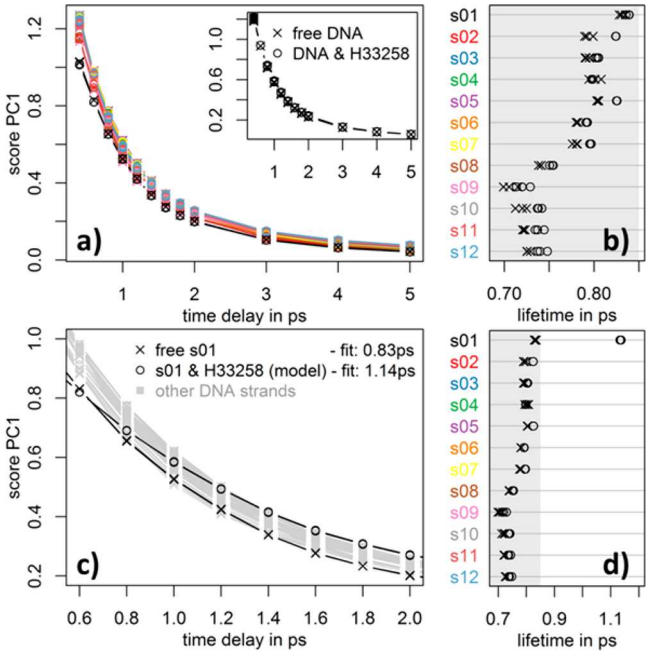


Figure 9. (a, inset) PC1 score of subset $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{H} + (\boldsymbol{HW}) + \boldsymbol{\varepsilon}$ plotted against waiting time. (a) PC1 score of a PCA using the complete data set $\boldsymbol{X}$ plotted against waiting time. (b) Vibrational lifetime with (circles) and without (crosses) H33258 obtained from PCA on $\boldsymbol{X}$ using a monoexponential function. (c) A PCA performed on $\boldsymbol{X}_{\text{model}}$. The PC1 score successfully retrieves an artificially induced slow vibrational relaxation for oligomer s_{01} with H33258. (d) Retrieved lifetime data from $\boldsymbol{X}_{\text{model}}$.

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732–2740

suggests that interactions with H33258 are more subtle and do not cause an overall change to the principal vibrational relaxation dynamics of DNA.

Extending this subset to sequence-specific interactions requires the addition of all remaining matrices, ($\boldsymbol{S}$), ($\boldsymbol{SH}$), ($\boldsymbol{SW}$), and ($\boldsymbol{SHW}$), resulting in a PCA of the initial data set $\boldsymbol{X}$. Figure 9b shows that the principal variance of the complete data set again characterizes the relaxation across all vibrational modes. The loading plot resembles again the global average spectrum $\bar{x}$ and fitting of the PC1 score to a monoexponential function reveals that none of the oligomers measured show a significant change in vibrational lifetime when adding H33258 (Figure 9b). This result was validated as before by performing individual PCAs for each sequence−ligand combination and fitting the obtained PC1 score to a monoexponential function (Figure S-7). Lifetimes recovered from this approach agree well with the dynamics acquired from ANOVA−PCA.

In light of the demonstrated relative insensitivity of DNA vibrational relaxation to sequence or ligand binding, a further test was carried out to determine the ability of ANOVA−PCA to retrieve different relaxation dynamics should they have existed in the data set. To achieve this, a model data set $\boldsymbol{X}_{\mathbf{model}}$ was generated, in which the amplitudes of all spectra for s_1 in the presence of H33258 were rescaled to show a vibrational decay time scale of 1100 fs, significantly slower than observed in $\boldsymbol{X}$ and different to all other samples, which were kept the same as in $\boldsymbol{X}$. The PC1 score from $\boldsymbol{X}_{\mathbf{model}}$ (all elements were analyzed) is shown in Figure 9c and successfully retrieves the 300 fs slower decay for the vibrational modes of s_1 when interacting with the binder (Figure 9d). This result confirms that the principal vibrational relaxation captured in the 2D-IR spectrum is indeed unaffected by the H33258 interaction.

CONCLUSION

It has been shown that the ANOVA−PCA method is able to separate a large, highly dimensional data set into tangible subsets that can be analyzed in a step by step manner. The separation of variance corresponding to well-defined factors allows selective analysis of the information of interest and exclusion of otherwise inseparable data. This method applied for the first time to a 2D-IR data set of 2016 DNA spectra clearly separates generic A−T base-pair vibrations from generic G−C vibrations as well as accurately revealing sequence composition and will allow the study of sequence-dependent shifts of these vibrations due to nearest-neighbor interactions along the strand. We have shown that the principal vibrational lifetime of these modes remains largely unaffected by the base composition of the double strands and ligand binding. This information was extracted without labor-intensive analysis of individual spectra, and the results obtained from ANOVA−PCA are comparable to published results that were collected using conventional methods.

Perhaps most usefully for novel applications, ANOVA−PCA highlights small changes in 2D-IR spectra due to sequence-dependent DNA−ligand interactions. Conventional methods like fluorescence measurements or even the DNA melting point stabilization obtained from FT-IR data can only give a simple indication as to whether the ligand is binding to the target sequence or not. These methods are unable to give information about the molecular details of the binding interaction. An important distinction between the ANOVA−PCA results extracted from 2D-IR presented here and similar experiments using IR absorption is the more information-rich view of how the ligand is interacting with the DNA that is obtained from the vibrational coupling patterns detected in the off-diagonal region of the 2D-IR spectrum. We show that, although they can be broadly categorized, each of the sequences studied gives a unique 2D-IR response to H33258 binding. The most distinctive interactions can be summarized in two groups. The first group, with DNA strands containing target sequence A_nT_n showed spectral changes of A/T modes in both on- and off-diagonal regions consistent with an induced-fit-type interaction[12] of the ligand with DNA. The second group of sequences with a G/C-rich minor groove showed predominantly changes along the diagonal region of the 2D-IR spectrum and an absence of dominant off-diagonal changes. Most importantly a correlation of ANOVA−PCA-derived parameters with melting temperature stabilization measurements was shown and indicates that further study will reveal new subtle layers of insight from 2D-IR screening-type studies. Although it has been shown to be possible to obtain comparable results regarding DNA sequence and ligand binding from FT-IR spectroscopy, the additional layer of spectral insight is uniquely available form 2D-IR spectroscopy at little or no additional time overhead for data collection and shows the potential for 2D-IR to become more widely applicable as the laser technology becomes more accessible.

Extraction of more subtle dynamics like energy transfer processes is less straightforward with PCA due to the separation of components simply according to covariance. The use of difference spectra on the other hand is only able to return relative dynamics rather than absolute time-evolution of the actual signals. A different approach might therefore be favorable to extract these dynamics.

Overall, our results exemplify how the use of ANOVA−PCA could facilitate large-scale screening tests of ligands or drugs using time-resolved 2D-IR spectroscopy: The variation in DNA sequence can be generalized to a set of possible target structures where its corresponding main factor, S, defines characteristic vibrational signatures of these targets. The variance in H33258 can be replaced by known or unknown ligands, where main factor H contains ligand-specific changes to the 2D-IR spectra of the targets. Any change to the principal vibrational relaxation due to binding is available from interactions with factor W. This multivariate approach can utilize the time resolution and abundance of structural information in 2D-IR spectroscopy to its full, analytical potential and could help assessing the selectivity of novel ligands by understanding the underlying binding mechanisms.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.7b04727.

ANOVA−PCA results for FT-IR data; fluorescence enhancement of H33258 with DNA; global average 2D-IR spectrum $\bar{x}$, PC1 loading of $\boldsymbol{M} + \boldsymbol{W} + \boldsymbol{\varepsilon}$; extracting vibrational relaxation using individual PCA for each DNA sequence (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: neil.hunt@strath.ac.uk.

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732−2740

ORCID

Paul M. Donaldson: 0000-0002-0305-9142

Neil T. Hunt: 0000-0001-7400-5152

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding is gratefully acknowledged from STFC for programme access to the Central Laser Facility systems (ST/M000125/1) and from BBSRC (BB/L014335/1) for the LIFEtime spectrometer.

■ REFERENCES

(1) Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A. *Acc. Chem. Res.* **2008**, *41*, 432−441.
(2) Adamczyk, K.; Candelaresi, M.; Robb, K.; Gumiero, A.; Walsh, M. a; Parker, A. W.; Hoskisson, P. a; Tucker, N. P.; Hunt, N. T. *Meas. Sci. Technol.* **2012**, *23*, 062001.
(3) Elsaesser, T. *Chem. Rev.* **2017**, *117*, 10621−10622.
(4) Hunt, N. T. *Chem. Soc. Rev.* **2009**, *38*, 1837−1848.
(5) Johnson, P. J. M.; Koziol, K. L.; Hamm, P. *J. Phys. Chem. Lett.* **2017**, *8*, 2280−2284.
(6) Thielges, M. C.; Chung, J. K.; Fayer, M. D. *J. Am. Chem. Soc.* **2011**, *133*, 3995−4004.
(7) Pagano, P.; Guo, Q.; Kohen, A.; Cheatum, C. M. *J. Phys. Chem. Lett.* **2016**, *7*, 2507−2511.
(8) Krummel, A. T.; Mukherjee, P.; Zanni, M. T. *J. Phys. Chem. B* **2003**, *107*, 9165−9169.
(9) Greve, C.; Elsaesser, T. *J. Phys. Chem. B* **2013**, *117*, 14009−14017.
(10) Sanstead, P. J.; Stevenson, P.; Tokmakoff, A. *J. Am. Chem. Soc.* **2016**, *138*, 11792−11801.
(11) Hithell, G.; González-Jiménez, M.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Wynne, K.; Hunt, N. T. *Phys. Chem. Chem. Phys.* **2017**, *19*, 10333−10342.
(12) Ramakers, L. A. I.; Hithell, G.; May, J. J.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Hunt, N. T. *J. Phys. Chem. B* **2017**, *121*, 1295−1303.
(13) Shim, S.; Strasfeld, D. B.; Ling, Y. L.; Zanni, M. T. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14197−14202.
(14) Greetham, G. M.; Donaldson, P. M.; Nation, C.; Sazanovich, I. V.; Clark, I. P.; Shaw, D. J.; Parker, A. W.; Towrie, M. *Appl. Spectrosc.* **2016**, *70*, 645−653.
(15) Luther, B. M.; Tracy, K. M.; Gerrity, M.; Brown, S.; Krummel, A. T. *Opt. Express* **2016**, *24*, 4117−4127.
(16) Harrington, P. D. B.; Vieira, N. E.; Espinoza, J.; Nien, J. K.; Romero, R.; Yergey, A. L. *Anal. Chim. Acta* **2005**, *544*, 118−127.
(17) de Haan, J. R.; Wehrens, R.; Bauerschmidt, S.; Piek, E.; van Schaik, R. C.; Buydens, L. M. C. *Bioinformatics* **2007**, *23*, 184−190.
(18) Sarembaud, J.; Pinto, R.; Rutledge, D. N.; Feinberg, M. *Anal. Chim. Acta* **2007**, *603*, 147−154.
(19) Climaco Pinto, R.; Bosc, V.; Nocairi, H.; Barros, A. S.; Rutledge, D. N. *Anal. Chim. Acta* **2008**, *629*, 47−55.
(20) White, C. M.; Heidenreich, O.; Nordheim, A.; Beerman, T. A. *Biochemistry* **2000**, *39*, 12262−12273.
(21) Disney, M. D.; Stephenson, R.; Wright, T. W.; Haidaris, C. G.; Turner, D. H.; Gigliotti, F. *Antimicrob. Agents Chemother.* **2005**, *49*, 1326−1330.
(22) Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291−309.
(23) Bailly, C.; Colson, P.; Henichart, J.; Houssier, C. *Nucleic Acids Res.* **1993**, *21*, 3705−3709.
(24) Spink, N.; Brown, D. G.; Skelly, J. V.; Neidle, S. *Nucleic Acids Res.* **1994**, *22*, 1607−1612.
(25) Gavathiotis, E.; Sharman, G. J.; Searle, M. S. *Nucleic Acids Res.* **2000**, *28*, 728−735.
(26) Breusegem, S. Y.; Clegg, R. M.; Loontiens, F. G. *J. Mol. Biol.* **2002**, *315*, 1049−1061.
(27) Fornander, L. H.; Wu, L.; Billeter, M.; Lincoln, P.; Nordén, B. *J. Phys. Chem. B* **2013**, *117*, 5820−5830.
(28) Bostock-Smith, C. E.; Harris, S. A.; Laughton, C. A.; Searle, M. S. *Nucleic Acids Res.* **2001**, *29*, 693−702.
(29) Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, A. W.; Hunt, N. T. *J. Phys. Chem. B* **2016**, *120*, 4009−4018.
(30) R Core Team. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014; https://www.r-project.org/.
(31) Furse, K. E.; Corcelli, S. A. *J. Am. Chem. Soc.* **2008**, *130*, 13103−13109.
(32) Bazhulina, N. P.; Nikitin, A. M.; Rodin, S. A.; Surovaya, A. N.; Kravatsky, Y. V.; Pismensky, V. F.; Archipova, V. S.; Martin, R.; Gursky, G. V. *J. Biomol. Struct. Dyn.* **2009**, *26*, 701−718.

DOI: 10.1021/acs.analchem.7b04727
Anal. Chem. 2018, 90, 2732−2740