Two-dimensional codistribution spectroscopy applied to UVRR spectroscopy and ROA investigations of biomolecular transitions

Georg Ramer[1] and Lorna Ashton[2*]

[1] Vienna University of Technology, Institute for Chemical Technologies and Analytics, Getreidemarkt 9/164 UPA, Vienna, Austria
[2] Department of Chemistry, Lancaster University, Lancaster, UK, LA1 4YB
* Corresponding Author: l.ashton@lancaster.ac.uk

Abstract

The first Raman optical activity (ROA) two-dimensional correlation spectroscopy (2DCOS) study in 2006, monitoring the α-helix-to-β-sheet transition in poly(L-lysine) as a function of temperature, demonstrated the versatility of 2DCOS. The combination of ROA and 2DCOS provided new ROA band assignments, enabled a direct comparison between the simultaneously collected Raman and ROA data using heterocorrelations and probed sequential information. This study also confirmed that 2DCOS can be successfully used with bisignate data, although specific care is needed when interpreting the results. However, as time has passed, doubts have been raised about not only the sequential orders reported in the study but also the general reliability of sequential data. This issue has now been addressed with the introduction of 2D codistribution (2DCDS) which is specifically designed to provide the sequence of the distributed presence of species along the perturbation variable axis. We have revisited the original ROA data and in light of these and other new developments in 2D correlation techniques we present our updated results. Furthermore, we will demonstrate how 2DCDS can be successfully applied to bisignate data using new spectral data sets of perturbation-induced transitions in polynucleotides.

Introduction

The field of 2D correlation spectroscopy (2DCOS) has advanced rapidly since the publication of the first Raman optical activity (ROA) 2DCOS study in 2006.[1] There have been vast advancements in further 2DCOS techniques including, perturbation correlation moving windows[2], projection 2D correlation [3] and Pareto scaling[4] and the continued application of 2DCOS to a wide range of research areas demonstrates the versatility of the technique. The original ROA 2DCOS paper investigated the temperature-induced conformational transition from α-helix-to-β-sheet of poly(L-lysine) (PLYS) and successfully demonstrated the application of 2DCOS to bisignate ROA data. The synchronous plots enabled new ROA band assignments to be suggested while the asynchronous plots were used to probe sequential information and an attempt was made to relate these to specific steps in the structural

transition. The results of this study revealed unexpected sequential orders that suggested that β-sheet formed ahead of the unfolding of α-helical secondary structure. Close inspection of the ROA spectra indicated the presence of some β-sheet at the start of the investigation due to the observation of a negative peak observed at ~1220 cm[-1] and the unexpected sequential order was attributed to polypeptide aggregation occurring ahead of polypeptide unfolding as a result of this β-structure.[1] However, over the last decade many questions have been raised regarding the interpretation and reliability of asynchronous plots and this study was among others used as an example of incorrect sequential orders.[5]

Since the original study there has been much confusion in the literature as to whether the asynchronous plots do or do not provide reliable information on the local sequential order of events, the rate of different events or the average rate of concentration change of each species.[5-7] The majority of these misunderstandings are due to a lack of understanding of the theoretical foundations behind 2DCOS.[8, 9] Without an appreciation of the underlying fundamentals behind the rules of interpretation, commonly referred to as 'Noda's rules' it is easy to misinterpret or over interpret the available sequential information in 2DCOS plots.[8, 9] In 2014, Noda published an important paper highlighting these misunderstanding and introduced a new, complimentary method referred to as 2D Codistribution (2DCDS) to try to dispel the undue scepticism of 2DCOS.[9] The paper clearly states that 2DCOS can be used to probe the sequential order of spectral intensity variations but does not directly tell the orders of the distributed species (or structure). The new technique of 2DCDS has specifically been developed to address this problem and is reported to provide the sequence of distributed presence of species along the perturbation variable axis.[4, 9]

In light of these developments we have revisited the original ROA 2DCOS study and applied 2DCDS to this ROA data. Furthermore, in order to gain a greater understanding of the two techniques when applied to biological molecules we have used a model UV resonance Raman (UVRR) spectroscopy study of variations in polynucleotide concentrations to directly compare 2DCOS and 2DCDS.

Experimental

Polynucleotide UVRR Spectra
Polyadenylic acid - polyuridylic acid sodium salt (polyAU) and polycytidylic acid - polyguanylic acid sodium salt (polyGC) were purchased from Sigma and each polynucleotide prepared by dissolving the dry material in distilled H_2O without any further purification at a concentration of ~5 mg/mL. Samples were then prepared by mixing polyAU and polyGC at concentrations from 0-100% in 10% intervals. For example sample 1 contained 100% polyAU and 0% polyGC while sample 2 contained 90% polyAU and 10% polyGC until 6 samples had been prepared. Each the 6 samples had a total polynucleotide concentration of 5 mg/mL.

UVRR spectra were acquired using a Raman microscope (Renishaw, Wotton-under-edge, Gloucestershire, UK). Each sample (30 μL) was pipetted into a small well on the lid of a polystyrene microplate as previously described.[10] Approximately 0.2 mW of power were delivered to the sampling point using a Lexel Laser emitting at 244 nm (Cambridge Lasers, USA). The UV laser was focused into the solution with care being taken not to focus on the lid itself and each spectrum was acquired for 60s as the well was continuously rotated under the laser to avoid photodegradation. Cosmic spikes in the data set were electronically removed from spectra using Grams. Before 2DCOS or 2DCDS was applied the polynucleotide data were smoothed (using a triangular average) and baseline corrected (using attenuated least squares) using MATLAB software version 2013b (The Math Works, MA, USA).

ROA Poly(L-lysine) Spectra

A sample data set taken from the original data previously reported for the ROA poly(L-lysine) (PLYS) temperature-induced α-helix-β-sheet transition was used for this study and experimental details are available in the original paper.[1] Briefly, spectra were acquired for 2-7 hours at 4, 12, 20, 28, 36, 44 and 52 °C using the ICP ROA instrument at Glasgow University.[11] For this study the wavenumber region 1000-1400 cm^{-1} was investigated and no additional preprocessing was carried out further to the original data pretreatments of baseline subtraction, normalisation to data acquisition times and smoothing.[1]

2D correlation analysis

2DCOS and 2DCDS were carried out using MATLAB. For the polynucleotide data all synchronous plots have been drawn using a maximum number of contours of 15 contours and asynchronous plots at a maximum of 12 contours in order to only display the most significantly changing peaks as identified from the spectra (Figure 1A). For the PLYS data set a maximum of 12 contours was used for all 2DCOS and 2DCDS synchronous and asynchronous plots.

Results and Discussion

PolyAU-PolyGC concentration profiles: 2DCOS

UVRR spectra were acquired of differing concentrations of the polynucleotides polyAU and polyGC in water in 10% increments with a total concentration of 5 mg/mL (Figure 1A). As polyAU concentration decreases from 100% to 0% a clear loss in band intensity can be observed for the adenine (A) and uracil (U) assigned peaks at ~1340 and 1580 cm^{-1} [12-14] Only one peak positioned at 1485 cm^{-1} assigned to guanine (G)[12-14] increases in intensity with increasing concentration of polyGC from 0 to 100%.
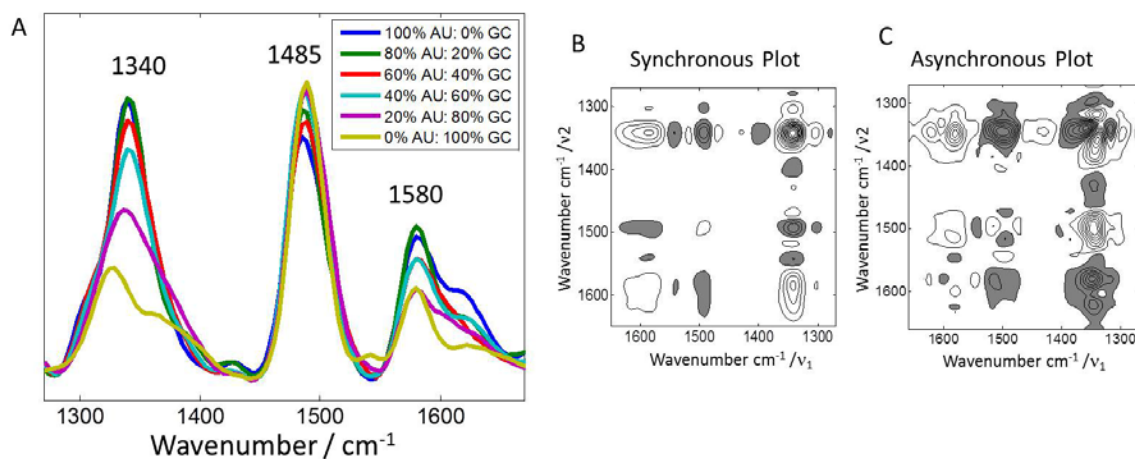
Figure 1. (A) UVRR spectra of differing concentrations of the polynucleotides polyAU and polyGC in water in 10% increments with a total concentration of 5 mg/mL alongside (B) the synchronous plot and (C) asynchronous plot generated from the polynucleotide spectral data set. No shading indicates positive cross peaks and grey shading indicates negative cross peaks.

Figures 1B and 1C display the 2DCOS synchronous and asynchronous plots, respectively, for the polynucleotide UVRR data. Interpretation of the plots requires the application of the established rules, referred to a 'Noda's rules' [8] which state that a positive synchronous cross peak (unshaded throughout this paper) indicates that both wavenumbers ν_1 and ν_2 are changing in intensity in the same direction, while a negative cross peak (shaded throughout) indicates that ν_1 is changing in intensity in the opposite direction to ν_2. From Figure 1B a positive cross peak can be observed at (1340,1580) and negative cross peaks can be observed at (1340,1485) and (1485,1580) indicating that the peaks at ~1340 and 1580 cm^{-1} are decreasing in intensity while the peak at 1485 cm^{-1} is increasing in intensity. In this model polynucleotide example these changes can just as easily be observed in the original spectra (Figure 1A) and there is no advantage to applying 2D correlation analysis to this data. This is a frequent criticism of 2D correlation spectroscopy and it can often be applied unnecessarily. However, the key advantage of 2D correlation techniques comes to the forefront when they are applied to far more complex data sets that are difficult to analysis visually and can aid greatly in the extraction of information by providing a detailed, alternative method of visualisation.[15-17] Further cross peaks can be observed for spectral features such as those occurring at ~1300, 1400, and 1542 cm^{-1} , however, it can be observed from the spectra (Figure 1A) that these wavenumbers fall between the three most significantly changing peaks. As only six percentage mixtures have been measured it is difficult to determine whether these features are a real change as a result of changing percentages or due to baseline fluctuations not removed by the baseline correction. However, as the aim of the paper is to compare 2DCOS with 2DCDS we need only focus on the most significantly changing peaks.

Figure 1C displays the asynchronous plot of the polynucleotide spectral data set. By applying Noda's rules we can probe the sequential spectral information from the sign of corresponding cross peaks.[8] If the synchronous and corresponding asynchronous cross peak are of the same sign (either both positive or both negative) then wavenumber ν_1 changes predominately before wavenumber ν_2 conversely if the two cross peaks are of the opposite sign then ν_2 changes predominately before wavenumber ν_1. If there is no corresponding asynchronous cross peak then both peaks are changing simultaneously. In Figure 1C the positive asynchronous cross peak at (1340,1485) corresponds to a negative synchronous cross peak and therefore indicates that the band at 1485 cm^{-1} changes predominately before the band at 1340 cm^{-1}. Furthermore, the negative asynchronous cross peak at (1340,1580) corresponds to a positive synchronous cross peak and the negative asynchronous cross peak at (1485,1580) corresponds to a negative synchronous cross peak which together indicate that 1485 cm^{-1} changes in intensity predominantly before 1580 cm^{-1} which in turn changes in intensity predominately before 1340 cm^{-1}.

This sequential order on first analysis could be seen as counterintuitive as it suggests PolyGC changes before PolyAU yet the concentration changes are consistent and therefore changes in the population of the two species are simultaneous. However, when it is remembered that 2DCOS only provides the sequential order of spectral intensity variations and not species populations then the sequential orders are to be expected. From the UVRR spectra (Figure 1A) it can clearly be observed that each peak responds differently to the changing concentrations, the extent of intensity change and rate varies between each peak. The differences in spectral variations are particularly significant in this data set as it is UVRR spectra and it is well established that depending on specific wavelength Raman scattering from different nucleic acid residue can be enhanced preferentially relative to each other.[14, 18] At ~244 nm we can expect G residues to be enhanced preferentially to A residues[14, 18] and this can be observed in Figure 1A by the much greater variation in intensity from 0-100% concentrations of the polyAU assigned peaks at 1340 and 1580 cm^{-1} compared to the polyGC assigned peak at 1485 cm^{-1}. As there is a greater enhancement of the Raman scattering arising from the polyGC residues the peak changes only slightly with decreasing concentration, whereas the relatively weaker enhancement of polyAU residues results in much lower peak intensities at lower concentrations and therefore greater spectral variation. Consequently, the sequential orders, as determined by the 2DCOS asynchronous plot, are due to the differences in response of each peak to the changing concentrations. The enhanced sensitivity of the peak observed at 1485 cm^{-1} to polyGC may result in this peak changing ahead of the less sensitive peaks polyAU peaks occurring at ~1340 and 1580 when a excitation wavelength of 244 nm is applied. These results clearly demonstrate that 2DCOS can be used to probe the sequential order of spectral intensity variations but cannot be used to determine the orders of the distributed species (or structure).

PolyAU-PolyGC concentration profiles: 2DCDS

Due to the continued request for 2D correlation techniques to be able to determine differences in species population rather than spectral intensity variations Noda developed 2DCDS[9] and by applying 2DCDS to the model polynucleotide data we should be able to determine variations in population species. Figure 2 displays the 2DCDS (A) synchronous and (B) asynchronous plots of the model polynucleotide data. As previously reported in the synchronous plot the cross peaks are now all positive so limiting the usefulness of this plot.[9] There are also less cross peaks due to……
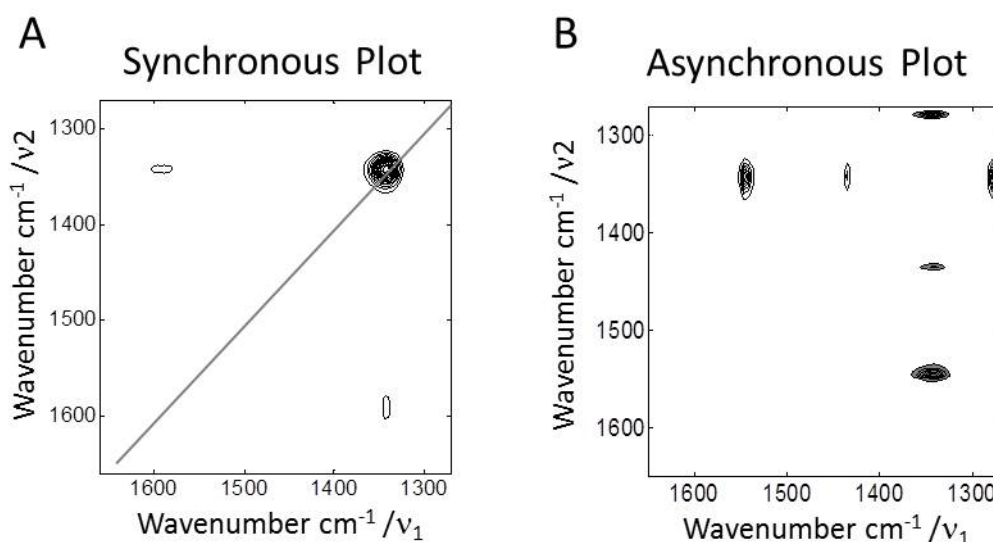
Figure 2. 2DCDS (A) synchronous and (B) asynchronous plots generated from the polynucleotide spectral data set. No shading indicates positive cross peaks and grey shading indicates negative cross peaks.

The asynchronous plot also has less cross peaks.

As all the synchronous cross peaks are positive interpretation of the 2CDS asynchronous plot is far easier than in 2DCOS. A positive asynchronous cross peak indicates the presence of the species reflected by the spectral intensity at v_1 is distributed predominantly before v_2, while a negative asynchronous cross peak indicates that v_2 is distributed predominantly before v_1.[4, 9] However, in Figure 2B no asynchronous cross peaks can be observed for the coordinates 1340, 1485 and 1580 cm^{-1} and this indicates that the population species change in a similar way, as expected as polyAU decreases in concentration polyGC increases in both in 10% steps. The only observed cross peaks fall at coordinates between peaks, and as previously discussed these are most likely due to changes in spectral background not corrected for by preprocessing.
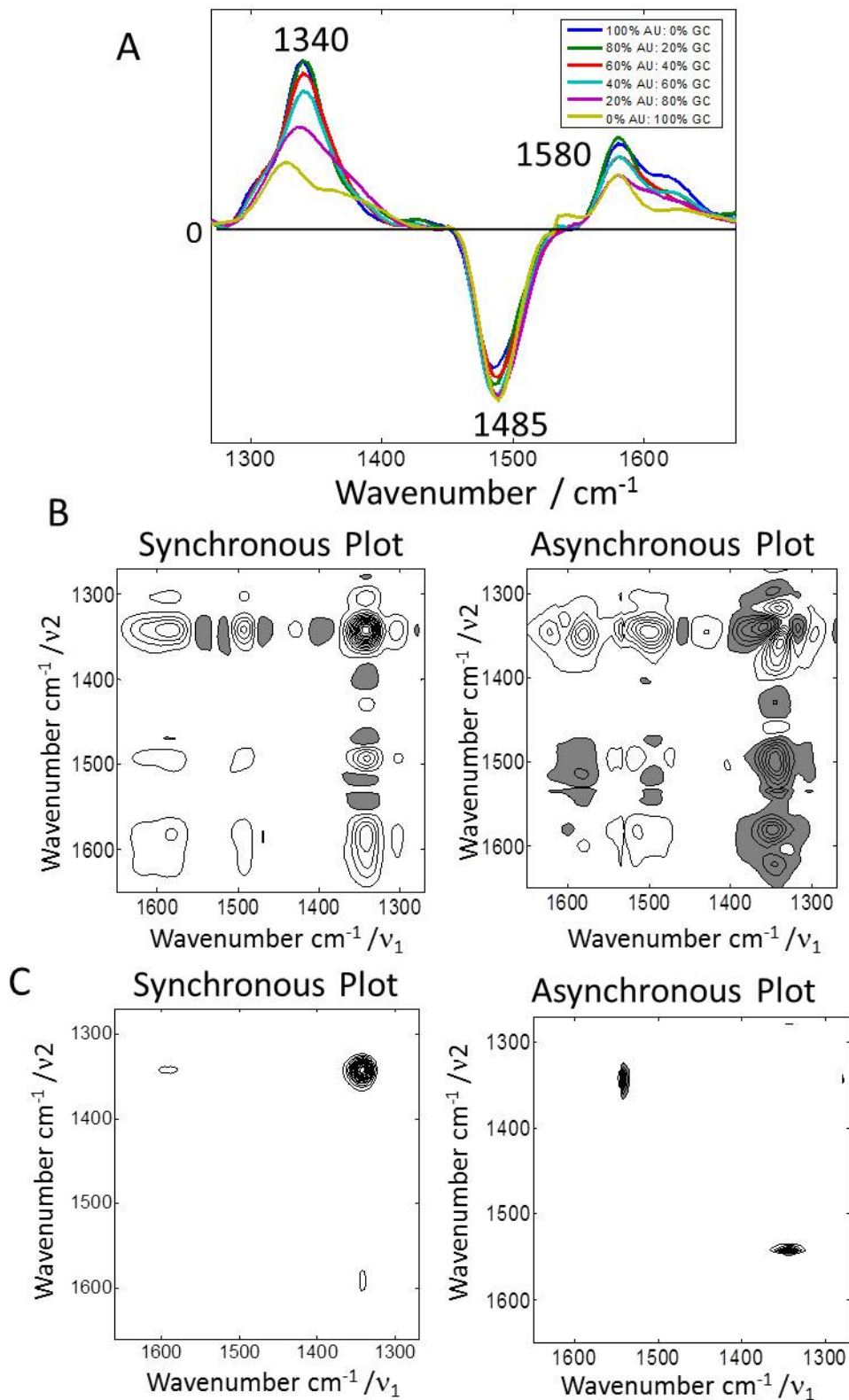
Bisignate Spectra

One of the initial concerns when first applying 2DCOS to ROA data was the bisignate nature of a ROA spectrum, in that ROA spectra have peaks that are both positive and negative in intenisty.[1, 19] Through the investigation of both ROA data and stimulated data we demonstrated that although the mathematical basis of 2DCOS allows for its successful application to bisignate data care does need to be taken when interpreting the synchronous plots. [19] In ROA spectra peaks can increase in intensity by either becoming more positive or more negative, and *vice versa* they decrease in intensity by becoming either less positive or less negative. The result of this for interpretation of 2DCOS synchronous plots is that a cross peak between a

For example a negative peak that is decreasing in intensity in response to a perturbation will become less negative which in turn means that it is also becoming more positive, however, if this is compared to a positive peak that

Take from original paper

Before applying 2DCDS to ROA data it was again important to establish the effect of using bisignate data. In order to test this approach the model polynucleotide data was adapted to be bisignate with the peak at 1485 cm^{-1} being converted to a negative peak (Figure 3A).

The 2DCOS synchronous and asynchronous are shown in Figure 3B and a change in the sign of the synchronous cross peaks can be observed at (1340,1485) and (1485,1580) in response to the peak at 1485 cm^{-1} now being negative and it is therefore essential that this is taken into account when interpreting the direction of intensity change from the 2DCOS synchronous plot. Corresponding asynchronous peaks have also changed in sign, however, as these cross peaks have to be related to the synchronous cross peaks it does not affect interpretation of the asynchronous plots.

Figure 3. (A) The adapted UVRR spectra of differing concentrations of the polynucleotides polyAU and polyGC with the peak at 1485 cm^{-1} changed to a negative peak to simulate bisignate data. (B) The 2DCOS synchronous plot asynchronous plot generated from the polynucleotide spectral data set. (C) The 2DCDS synchronous plot asynchronous plot
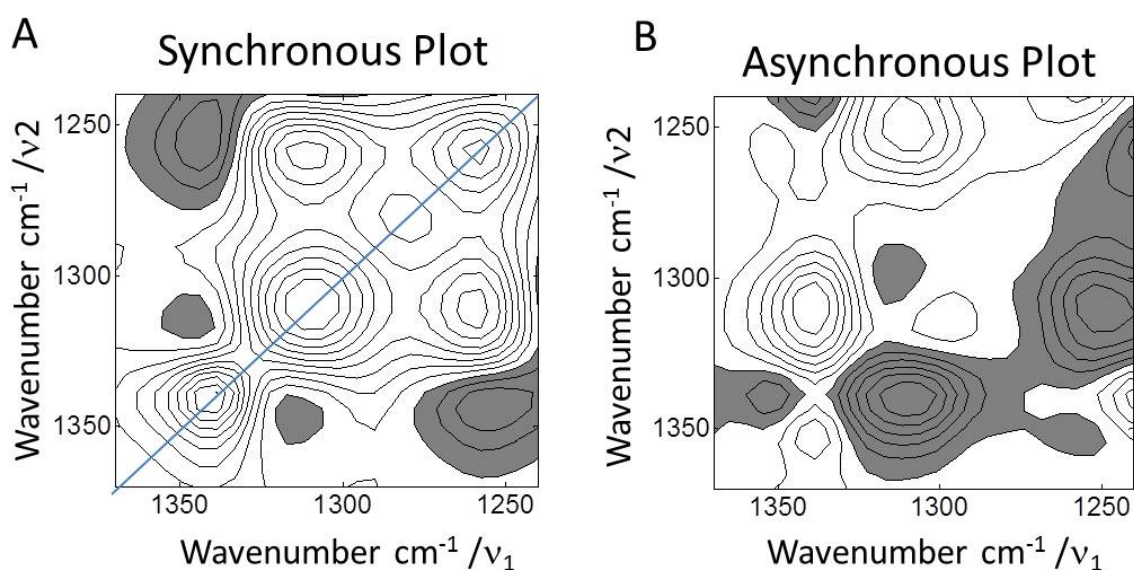
generated from the polynucleotide spectral data set. No shading indicates positive cross peaks and grey shading indicates negative cross peaks.

Figure 3C displays the 2DCDS plots for the adapted polynucleotide data and it can be observed that the plots are very similar to the original 2DCDS plots (Figure 2) with no additional cross peaks. These results confirm the suggestion made by Noda that the new technique of 2DCDS can be applied to bisignate data. Check Noda's wording

α-helix-to-β-sheet transition in PLYS

In 2006 we presented the full Raman and ROA data sets of temperature-induced α-helix-to-β-sheet transition in PLYS and analysed the 2DCOS results in detail.[1] To simplify the comparison of 2DCOS with 2DCDS a much smaller spectral region from 1240-1370 cm^{-1} has been used for this study, focusing on 3 specific ROA bands occurring at ~1260 and 1310 cm^{-1} assigned to β-structure and the band at ~1341 cm^{-1} assigned to α-helical structure. The redrawn 2DCOS synchronous and asynchronous plot can be observed in Figure 4A and 4B, respectively.

Figure 4. 2DCOS (A) synchronous and (B) asynchronous plots generated from the temperature-induced PLYS spectral data set in the wavenumber region of 1240-1370 cm^{-1}. No shading indicates positive cross peaks and grey shading indicates negative cross peaks.

The synchronous peaks negative at (1260,1341) and (1310,1341) and positive at (1260,1310) indicate that the peak at 1341 cm^{-1} is changing in the opposite direction to the peaks at 1260 and 1310 cm^{-1}. The corresponding negative asynchronous cross peak negative at

(1310,1341) indicates that the peak at 1341 cm[-1] changes in intensity predominantly before the peak at 1310 cm[-1]. As previously discussed, in the original paper these results were suggested to imply that the appearance of β-structure was occurring ahead of the unfolding of α-helix,[1] what should have been made clearer was that it was the actual appearance of the β-structure assigned spectral peaks that was occurring ahead of the loss of α-helical assigned peaks. The small presence of β-structure at the start of the experiment combined with the sensitivity of ROA spectra to changes in β-structure resulting in the spectral variations being dominated by these features.[20, 21]
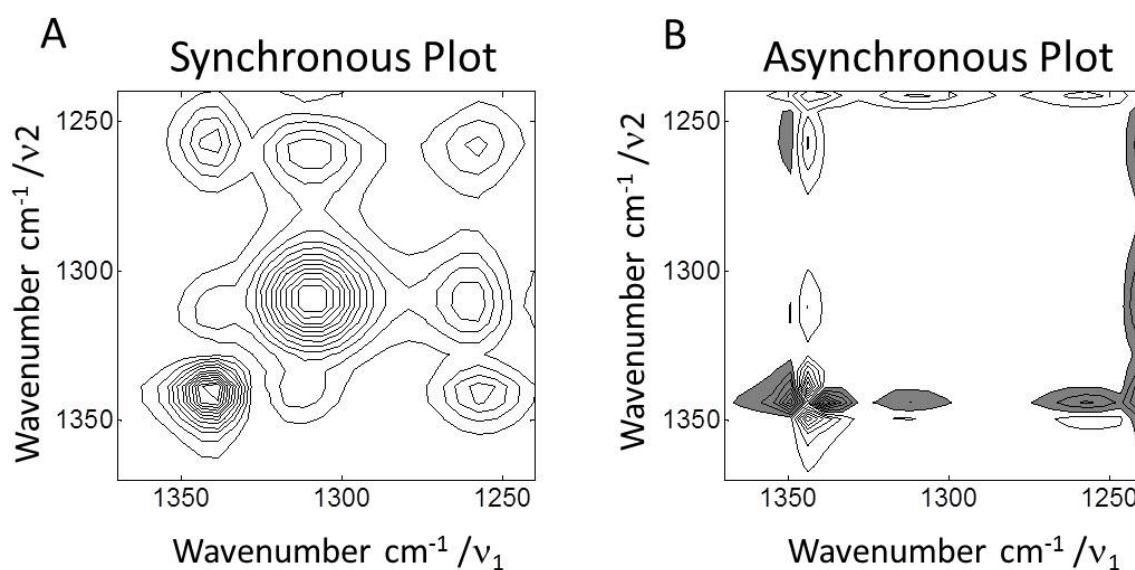
Figure 5. 2DCDS (A) synchronous and (B) asynchronous plots generated from the temperature-induced PLYS spectral data set in the wavenumber region of 1240-1370 cm[-1]. No shading indicates positive cross peaks and grey shading indicates negative cross peaks.

Figure 5 displays the 2DCDS synchronous and asynchronous plots. As before, the synchronous plot now only has positive cross peaks so limiting its usefulness for data analysis, and also has slightly less cross peaks. The 2DCDS asynchronous plot (Figure 5B) looks very different from the 2DCOS asynchronous plot (Figure 4B). In the 2DCDS plot two negative cross peaks can be observed at (1260,1341) and (1310,1341) which indicate that ~1341 is changing ahead of 1310 and 1260 cm[-1] check with Noda's paper how to word this. These results therefore suggest that the species population of α-helical structure is changing ahead of change in β-structure. Furthermore, in the 2DCOS asynchronous plot a negative cross peak corresponding to a positive synchronous cross peak at (1260,1310) indicated that despite both of these peaks being assigned to β-structure the peak at 1310 cm[-1] is changing predominantly before the peak at ~1260 cm[-1]. However in the 2DCDS asynchronous plot no

cross peak can be observed at (1260,1310) indicating that the two peaks are changing simultaneously as would be expected.

Conclusions

The primary goal of his paper was to clarify the distinction between two types of sequential order information i.e. variation vs distributed presence.

…designed to directly provide the sequence of distributed presence of species along the perturbation variable axis. This techniques can be used as a complementary tool to augment the deficiency of traditional two-dimensional correlation spectroscopy in directly identifying the presence of intermediate species.

Is most effectively used in dealing with the type of spectral intensity distirbutions which are monotonic or unimodal in nature

…segmentation of dataset into smaller blocks prior to the codistribution analysis is recommended

Thus the analysis may be in pricnicple applied to bisognate spectra. However it is important to make sure the sign of the actual spectral intensity does not change during the observation.

1.   L. Ashton, L. D. Barron, B. Czarnik-Matusewicz, L. Hecht, J. Hyde and E. W. Blanch, *Molecular Physics*, 2006, 104, 1429-1445.
2.   S. Morita, H. Shinzawa, I. Noda and Y. Ozaki, *Appl. Spectrosc.*, 2006, 60, 398-406.
3.   I. Noda, *J. Mol. Struct.*, 2010, 974, 116-126.
4.   I. Noda, *Biomed. Spectrosc. Imaging*, 2015, 4, 109-127.
5.   V. A. Shashilov and I. K. Lednev, *J. Am. Chem. Soc.*, 2008, 130, 309-317.
6.   Q. Jia, N. N. Wang and Z. W. Yu, *Appl. Spectrosc.*, 2009, 63, 344-353.
7.   H. Huang, *Analytical Chemistry*, 2007, 79, 8281-8292.
8.   I. Noda, *Vib. Spectrosc.*, 2012, 60, 146-153.
9.   I. Noda, *J. Mol. Struct.*, 2014, 1069, 50-59.
10.   L. Ashton, Y. Xu, V. L. Brewster, D. P. Cowcher, C. A. Sellick, A. J. Dickson, G. M. Stephens and R. Goodacre, *Analyst*, 2013, 138, 6977-6985.
11.   L. Hecht, L. D. Barron, E. W. Blanch, A. F. Bell and L. A. Day, *Journal of Raman Spectroscopy*, 1999, 30, 815-825.
12.   Q. Wu, T. Hamilton, W. H. Nelson, S. Elliott, J. F. Sperry and M. Wu, *Analytical Chemistry*, 2001, 73, 3432-3440.

13. Z. Q. Wen and G. J. Thomas, *Biopolymers*, 1998, 45, 247-256.
14. J. R. Perno, C. A. Grygon and T. G. Spiro, *Journal of Physical Chemistry*, 1989, 93, 5672-5678.
15. L. Ashton and E. W. Blanch, *J. Mol. Struct.*, 2010, 974, 132-138.
16. L. Ashton, A. Hobro, G. L. Conn, M. Rouhi and E. W. Blanch, *J. Mol. Struct.*, 2008, 883-884, 187-194.
17. Y. Wang, K. Murayama, Y. Myojo, R. Tsenkova, N. Hayashi and Y. Ozaki, *J. Phys. Chem. B*, 1998, 102, 6655-6662.
18. I. Mukerji, M. C. Shiber, J. R. Fresco and T. G. Spiro, *Nucleic Acids Res.*, 1996, 24, 5013-5020.
19. L. Ashton, B. Czarnik-Matusewicz and E. W. Blanch, *J. Mol. Struct.*, 2006, 799, 61-71.
20. I. H. McColl, E. W. Blanch, A. C. Gill, A. G. O. Rhie, M. A. Ritchie, L. Hecht, K. Nielsen and L. D. Barron, *J. Am. Chem. Soc.*, 2003, 125, 10019-10026.
21. L. D. Barron, L. Hecht, I. H. McColl and E. W. Blanch, *Molecular Physics*, 2004, 102, 731-744.