Optical fingerprinting of peptides using two-dimensional infrared spectroscopy: Proof of principle

Frédéric Fournier a,*, Elizabeth M. Gardner a, Rui Guo a, Paul M. Donaldson a, Laura M.C. Barter a, D. Jason Palmer b, Chris J. Barnett a, Keith R. Willison c, Ian R. Gould a, David R. Klug a

a Molecular Dynamics Research Group, Department of Chemistry, Imperial College London, London SW7 2AZ, UK
b European Laboratory for Nonlinear Spectroscopy, University of Florence, I-50019 Sesto Fiorentino (Firenze), Italy
c Institute of Cancer Research, Chester Beatty Laboratories, Cancer Research UK, Center of Cellular and Molecular Biology, London SW3 6JB, UK

Received 27 September 2007
Available online 13 November 2007

Abstract

We employ a particular form of two-dimensional infrared four-wave mixing (2DIR FWM) as a vibrational spectroscopic tool to quantify the amino acid content of a number of peptides. Vibrational features corresponding to ring modes of the aromatic groups of phenylalanine (Phe) and tyrosine (Tyr), as well as a methylene mode that is used as an internal reference, are identified. We show that the ratios of the integrated intensities, and the amplitudes, of the aromatic peaks of Phe and Tyr relative to the methylene integrated intensity, and amplitude, are proportional to the actual ratio of Phe and Tyr to CH2 in the samples within a precision of ±12.5%. This precision is shown to be sufficient to use this form of 2DIR spectroscopy as a possible proteins fingerprinting tool.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Two-dimensional infrared spectroscopy; 2DIR; Proteomics; Peptides; Phenylalanine; Tyrosine; Protein fingerprinting; Amino acid

We propose an alternative strategy for protein fingerprinting based on the identification of proteins’ amino acid content using a two-dimensional infrared (2DIR) spectroscopy. The number of amino acids that one needs to quantify depends on the proteins to be identified and on the precision of the measurements. Converting such an analytical strategy into a practicable proteomic method requires the attainment of particular levels of sensitivity, resolving power, and high throughput.

Optical techniques have been used for protein and peptide analysis, but overall they have been used with less widespread application than have the classical nuclear magnetic resonance (NMR) and mass spectrometry (MS) techniques. Raman vibrational spectroscopy and IR absorption spectroscopy can give secondary structure information and, when used in difference spectroscopy mode, can identify individual bonds within a protein [1–3]. Vibrational spectra are sensitive to chemical environment, secondary structure, and conformation [4,5]. The drop coating deposition Raman (DCDR) technique has also been used recently in the study of proteins and peptides [6,7] and has been applied to tyrosine (Tyr) phosphorylation [8] and protein mixtures [9].
Multidimensional nonlinear optical methods in recent years have become a practical and realizable approach in optical spectroscopy [10]. Coherent multidimensional vibrational spectroscopy has, in principle, the power to decongest vibrational spectra and to allow the fingerprinting of proteins without the need for significant sample manipulation. The most widely used 2D vibrational spectroscopy is based on a three-pulse photon echo approach [11,12]. Examples of its application to proteins and peptides include the structure determination of peptides by studying the amide I band that is associated with the amino acid backbone [13–15], the study of vibrational dynamics [16,17], and the study of hydrogen bond dynamics [18]. This technique has also been used to probe the coupling of DNA vibrational modes and secondary structure [19], exciton dynamics [20], and hydrogen bonds dynamics in water [21].

Linear optical spectroscopies have, in principle, both the sensitivity and potential throughput to be used as a proteomic tool, but their resolving power generally is insufficient to unambiguously quantify enough amino acids in a protein. The choice of 2DIR spectroscopy is motivated by the power that this technique has to spread the spectral information over two or more dimensions in a manner highly analogous to that of 2D NMR. Successful fingerprinting of peptides and proteins by 2DIR methods requires accessing a wide range of disparate coupled vibrational states. To access these, we have adopted and adapted a particular version of 2DIR developed by Wright and coworkers [22–27]. In this approach, the frequencies of the excitation pulses are scanned independently across the vibrational spectrum. The coupling information is “read out” by a visible laser beam that scatters from the resulting polarization. The advantage of this particular variant of 2DIR is that it allows detection of photons in the visible optical spectrum. The ability to restrict the measurements to a narrow band, or even a single point, of the 2DIR spectrum greatly improves the overall throughput. The method also allows the spectrum to be decongested even further than is possible with other 2DIR methods [28].

In this article, we apply our 2DIR approach to peptides and demonstrate its utility in the detection of aromatic amino acid side chains. We demonstrate that the first stages of this fingerprinting strategy (i.e., amino acid identification and quantification) are workable and show the practicality of using an internal reference to achieve relative quantification of the amino acid level. Most important, we demonstrate that this 2DIR technique potentially has the required sensitivity and high throughput to be a viable proteomic tool; neither sensitivity to primary structure nor sample preparation and presentation is an insurmountable barrier.

We present results from eight peptide sequences having variable levels of phenylalanine (Phe) and tyrosine (Tyr), and give an indication of the accuracy and precision with which this information can be recovered using the current approach.

**Materials and methods**

**Principle of 2DIR four-wave mixing spectroscopy**

A complete description of this spectroscopic method can be found in the various publications from Wright’s group [22–27] and in some specific comments on its utility for protein fingerprinting in a previous article of ours [28]. Briefly, 2DIR four-wave mixing (FWM) is a nonlinear coherent optical technique giving 2D vibrational spectra. FWM is a third-order process where three laser beams interact with the medium through the third-order susceptibility tensor. By using two IR beams (ωg and ωp) and a third visible beam (ωv), this technique becomes a powerful vibrational spectroscopic tool. When the IR beams’ wavelengths are scanned over vibrational resonances, the FWM signal shows enhancement for the IR frequencies that are in resonance with vibrational transitions.

The benefit of this technique is to show multiplicatively enhanced cross-peaks when the IR beams are in resonance with coupled vibrational modes. This multiplicative enhancement comes from the fact that the hyperpolarizability describing the molecular response at the third order [25] contains mechanical and electric anharmonicity contributions [27]. These terms are nonzero only if the vibrational modes excited by ωg and ωp are coupled by intramolecular interactions or if they are composed of a fundamental mode and its overtone or a fundamental mode and a combination band. Fig. 1 shows the energy diagrams and coherence pathways involved for both processes giving multiplicative enhancements: DOVE–IR and DOVE–Raman.

**Sample preparation**

Automated synthesis of the peptides was performed using an Advanced ChemTech Apex 396 multiple peptide synthesizer. The syntheses followed a standard Fmoc/tBu (N-(9-fluorenylmethoxycarbonyl/tert-butyl) peptide
Peptides were purified to more than 98% homogeneity using reverse-phase high-performance liquid chromatography (HPLC) and then characterized by matrix-assisted laser desorption/ionization time-of-flight (MALDI–TOF) MS.

To avoid solvent/water contributions, the peptides were prepared in the form of dried films cast on glass slides. To prepare each film, a peptide solution of 50 mM concentration was made using nonbuffered ultrapure water. A 2-μl drop of the solution was deposited onto a standard, clear glass microscope slide 1.0 to 1.2 mm thick. The drop was left to dry in air at room temperature for approximately 30 min until it had formed a film. A second drop of 2 μl of the solution was then deposited on top of the film and also left to dry. The peptide films were further dried when put into the sample compartment of the experiment that was purged with nitrogen (<0.1% humidity). In several cases, the drops show a ring similar to the rings present in the DCDR technique [30]. We estimated an average of 10^{13} molecules present in the laser spots (typically 100 μm diameter).

**Optical setup**

A full description of the optical setup can be found elsewhere [28]. Briefly, to acquire 2DIR spectra, two IR beams (ω₁ and ω₀) and a visible beam (ω₂) are noncollinearly overlapped in time and space on the sample (Fig. 2B). The FWM signal at ω₃ = ω₁ + ω₀ − ω₂ is detected in the visible part of the electromagnetic spectrum with a photomultiplier, while the IR beam wavelengths are scanned over the spectral region of interest. For the experiment presented in this article, ω₁ is scanned from 1350 to 1570 cm⁻¹ and ω₀ is scanned from 2700 to 3200 cm⁻¹, both with 5 cm⁻¹ increments. The data obtained are 2D maps of the FWM signal level as a function of ω₁ and ω₀. Based on atmospheric water absorption bands are present in the first spectral region, the optical path of the IR beam at ω₀ is purged with nitrogen to avoid any spectral absorption that would cause pulse distortions and then artifacts in the 2D spectra. The visible beam wavelength is 790 nm (12,658 cm⁻¹). The FWM signal is then detected in the spectral range from 13,788 cm⁻¹ (725 nm) to 14,508 cm⁻¹ (689 nm).

The time delays of the ω₀ IR pulse and visible pulse relative to the ω₁ IR pulse are controlled by two delay stages (the pulse ordering is shown in Fig. 2A). The following delays were found to give a satisfying level of FWM signal for the systems studied in this article without nonresonant background or other processes, τ₁₂ = 1.5 ps and τ₂₃ = 1 ps. We found that in this configuration the spectral features are not significantly distorted by interference effects and, thus, are easier to exploit.

**Quantification procedure**

Once specific features corresponding to certain amino acids have been identified, the aim is to be able to establish the quantity of a particular amino acid in a sequence using the variations of its corresponding peak intensity from one 2D spectrum to another. Two different ways of measuring are used. First, to demonstrate the high-throughput potential, we use a simple peak amplitude measurement corrected for any experimental offset. Second, as a reference method (interference free and with potential broadening effects taken into account), we calculate the integrated intensity for each feature of interest. To perform the latter type of measurement, the spectra are fitted with 2D Gaussian functions (Eqs. (1a) and (1b)):

\[
S_{\text{FWM-CH}_{3}} = \text{Background} + \left| A_1 e^{-\left(\frac{1}{2\ln(2/\pi^2)\rho_1^2}\right)} \right|^2 ,
\]

(1a)

\[
S_{\text{FWM-CH}_{n}} = \text{Background} + \sum_{n \neq 1} A_n e^{\gamma n} e^{-\left(\frac{1}{2\ln(2/\pi^2)\rho_2^2}\right)} \left| e^{-\left(\frac{1}{2\ln(2/\pi^2)\rho_2^2}\right)} \right|^2 .
\]

(1b)

Experimental line shape studies and calculations using time-dependent perturbation theory have shown that when the delays τ₁₂ and τ₂₃ increase, the peak line shape changes.

![Fig. 2](image-url)
from Lorentzian (for zero delays) to roughly Gaussian. The spectra used in this work are taken at $\tau_{21} = 1.5$ ps and $\tau_{23} = 1$ ps, delays for which the Gaussian line shape dominates [28].

$\omega_n$ and $\omega_p$ are the laser frequencies and are the variables of the fit. $\omega_{2n}$ and $\omega_{2p}$ are the frequencies of the vibrational modes seen on the spectrum and are used as parameters of the fit. The subscript $n$ represents the peak, and the Greek letter represents either the $\omega_n$ direction or the $\omega_p$ direction. The parameter $A$ is the amplitude of the Gaussian functions, $\varphi_n$ is the phase of the peak $n$, and $\Gamma$ is the full-width at half-maximum (FWHM) for each peak in both frequency coordinates. These are also parameters of the fit.

To speed up the fitting process, the spectra are cut into three regions of interest (Fig. 3): the aliphatic peak region, the low-frequency aromatic region, and the high-frequency aromatic region. Eq. (1a) is used to fit the aliphatic peak only; this mathematical form was chosen empirically and represents the diagonal behavior of this particular peak very well. For the two other spectral regions, we used Gaussian functions with relative phases (Eq. (1b)). This allows the possible interferences between peaks to be taken into account and allows interference-free amplitudes and widths of each peak to be extracted. We also included a background in the equation used for the fit to correct from any experimental offset. Fig. 3 shows the fit performed for the YGGFF sample superimposed on the experimental spectrum.

Peak amplitudes and FWHM are extracted from the fits, allowing calculation of the integral of each peak independently from the others. Taking the modulus square of a 2D Gaussian peak having amplitude $A$ and FWHMs $\Gamma_n$ and $\Gamma_p$, the integral from $-\infty$ to $+\infty$ is easily calculated as

$$I_{\text{peak}} = \frac{A^2\Gamma_{|n\rangle\Gamma_{|m\rangle}}}{8\ln 2}.$$  \hspace{1cm} (2a)

For the aliphatic peak, taking into account the mathematical form used for the fit, the integral is calculated as

$$I_{\text{CH}_2} = \frac{A^2_{\text{CH}_2}\Gamma_{|\text{CH}_2\rangle\Gamma_{|\text{CH}_2\rangle}}}{16\ln 2}.$$  \hspace{1cm} (2b)

With the FWM signal being proportional to the module square of the third order susceptibility, and then to the square of the number of molecules [31], the square root of the integral is calculated to give a quantity having a linear dependence with the number of molecules. The aliphatic peak at 1450/2850 (all of the peaks are noted by their coordinates in the frequency plane, $\omega_n/\omega_p$, in cm$^{-1}$) is taken as an internal reference, and the integral of the other peaks is taken relative to this one. For any peak $n$, the ratio of Eqs. (2a) and (2b) is calculated. With the square of the amplitude $A$ being proportional to the square of the number of molecules $N (A^2 \propto N^2)$, we obtain the following expression for the ratio (Eq. (3)):

$$\text{Ratio}_n = \sqrt{\frac{I_{\text{peak}}}{I_{\text{CH}_2}}} = \sqrt{\frac{2A^2_{\text{CH}_2}\Gamma_{|n\rangle\Gamma_{|m\rangle}}}{A^2_{\text{CH}_2}\Gamma_{|\text{CH}_2\rangle\Gamma_{|\text{CH}_2\rangle}}}}.$$  \hspace{1cm} (3)

The number of CH$_2$ groups ($N_{\text{CH}_2}$) in each sequence is easily calculated, as are the numbers of phenylalanine and tyrosine groups ($N_{\text{Phe}}$ and $N_{\text{Tyr}}$). The experimental quantity, Ratio$_n$, is then plotted as a function of the ratio $\frac{N_{\text{Phe}}}{N_{\text{CH}_2}}$ for each Phe and Tyr peak and should show a linear variation.

Results

2D spectra of peptides

2D vibrational spectra of eight peptides are shown in Fig. 4. The spectra have in common a strong peak at around 1450/2850. This has been identified as corresponding to vibrational modes of CH$_2$ [32]; this feature is used as an internal reference for the calculation of the ratio of the integrated intensities. Other complex features appear around 1470/3050, 1520/3100, and 1480/2920.

For sequences in which Phe is the only aromatic acid present, two groups of peaks are measured in the upper part of the spectra (LRRFSLG and LRRFFLG,

---

**Fig. 3.** Example of fits using Eqs. (1a) and (1b). The spectrum shown is the one obtained for the YGGFF sample. On the 2D graph, the experimental data are shown in color levels and the fits are represented by black contours. The regions of interest used for the separated fitting procedure are shown in red, green, and orange. These correspond to the aliphatic region, the low-frequency aromatic region, and the high-frequency aromatic region, respectively. Three profiles for the three spectral regions are shown at 1450, 1475, and 1515 cm$^{-1}$. The black dots are the data points, and the color lines are the fits restricted to their respective regions. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
Three peaks are present along the vertical at $\omega_\alpha \approx 1515$ cm$^{-1}$ and $\omega_\beta = 3110, 3080,$ and $3040$ cm$^{-1}$; this part of the spectra is called the high-frequency aromatic region. Two peaks are also identified at 1470/3050 and 1480/3070; we call this spectral region the low-frequency aromatic region.

For sequences in which Tyr is the only aromatic amino acid present, only a very weak peak is measured in the low-frequency region at 1470/3040 and a stronger one is measured in the high-frequency region at 1530/3130 (LRRYSLGY and LRRYYLG in Fig. 4).

The assignments of the modes of the aromatic peaks are confirmed by density functional theory (DFT) calculations of 2DIR FWM spectra of benzene derivatives and aromatic amino acids. The details of the DFT calculations, the detailed spectra calculations, and the assignments concerning the aromatic amino acids will be the subject of a future article, with some information already made available in a recent publication [28]. The identified peaks are depicted in Fig. 5. The spectral assignments are listed in Table 1. There are two vibrational ring stretching modes contributing to the aromatic peaks of Phe and Tyr: CC + HCC in-plane ($\nu_{13}$) and CC + HCC in-plane + CCC in-plane ($\nu_{16}$) [33].

We focus on the Phe’s doublet, the most intense peak of the Phe’s triplet, and the most intense Tyr peak. We extract the FWHM and amplitudes, as well as the frequencies, of the aliphatic Tyr and Phe peaks, and the relative amplitudes and integrated intensities are calculated using the procedure described in the previous section.

Quantification of amino acid levels

Figs. 6 and 7 show plots of the normalized integrated intensities (Ratio in Eq. (3)) and the normalized peak amplitudes of the Phe and Tyr peaks against the known number of Phe and Tyr present in each peptide relative.
The normalization is achieved by ratioing the sizes of the side chain features against the size of the \( \text{CH}_2 \) feature. Because the experiments are all homodyne, all signals are proportional to the square of the concentration of molecules and, therefore, are square-rooted before ratioing.

The IR beam intensities are wavelength dependent, but because these variations are the same for each scan, the ratios of the peaks’ integrated intensities are not affected and are directly comparable from one scan to another.

The correlation curves obtained are fitted with linear functions constrained to go through the origin. Indeed, a spectrum of a sample without Phe (LRRYYLG and

### Table 1

Spectral assignments of the main peaks depicted in Fig. 5

<table>
<thead>
<tr>
<th>Peak position (cm(^{-1}))</th>
<th>Group and amino acid</th>
<th>Mode assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1450/2850</td>
<td>CH(_2), all</td>
<td>( \delta \text{CH}_2/2\delta \text{CH}_2 ) (first overtone)</td>
</tr>
<tr>
<td>1470/3050</td>
<td>Aromatic, Phe</td>
<td>( (6\delta \text{CH}<em>2, v</em>{13})/(6\delta \text{CH}<em>2, v</em>{13}) ) + (6( \delta \text{CH}<em>2, v</em>{16}) )</td>
</tr>
<tr>
<td>1480/3070</td>
<td>Aromatic, Phe</td>
<td>( v_{13}/v_{13} + (6\delta \text{CH}<em>2, v</em>{16}) )</td>
</tr>
<tr>
<td>1515/3040</td>
<td>Aromatic, Phe</td>
<td>Not assigned</td>
</tr>
<tr>
<td>1515/3080</td>
<td>Aromatic, Phe</td>
<td>( v_{13}/v_{13} + (6\delta \text{CH}<em>2, v</em>{16}) )</td>
</tr>
<tr>
<td>1515/3110</td>
<td>Aromatic, Phe</td>
<td>( v_{13}/v_{13} + v_{16} )</td>
</tr>
<tr>
<td>1530/3130</td>
<td>Aromatic, Tyr</td>
<td>( v_{13}/v_{13} + v_{16} )</td>
</tr>
</tbody>
</table>

**Note.** \( \delta \), deformation; \( v \), stretch; \( + \), combination band. Modes in parentheses are complex modes with two contributions.

### Fig. 6

Correlation between the known Phe to \( \text{CH}_2 \) ratios and the ratios measured by 2DIR. The square root of the ratio of the integral (and amplitude) of the features relative to the aliphatic peak integral (and amplitude) is plotted as a function of the known number of Phe relative to the known number of \( \text{CH}_2 \) groups. Data points are shown by open circles (and open triangles) and are labeled by the amino acid sequence of the corresponding sample. The dashed red line is the linear fit constrained to go through the origin. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

### Fig. 7

Correlation between the known Tyr to \( \text{CH}_2 \) ratios and the ratios measured by 2DIR. The square root of the ratio of the integral (and amplitude) of the features relative to the aliphatic peak integral (and amplitude) is plotted as a function of the known number of Tyr relative to the known number of \( \text{CH}_2 \) groups. Data points are shown by open circles (and open triangles) and are labeled by the amino acid sequence of the corresponding sample. The dashed red line is the linear fit constrained to go through the origin. Note that the scales are smaller than those in Fig. 6, making the dispersion appear bigger than it really is (see Table 2).
LRRYSGLY) or Tyr (LRRFFLG and LRRFSLGF) does not show any features at the Phe or Tyr frequencies, and so these peak intensities are zero. The standard deviations for the dispersion of the data points for each feature for both the integral and amplitude data are shown in Table 2.

### Discussion

The linear correlation between the relative integrated intensities (or amplitudes) of specific vibrational features and the relative amount of the corresponding amino acid content shown in Figs. 6 and 7 confirms the validity of Eq. (3). This demonstrates that relative quantification of the amino acid content by 2DIR is possible. The deviations from the linear behavior shown in Table 2 can have two different origins: structural effects (primary and secondary structures) and sample-to-sample variations.

The sample-to-sample variations include fluctuations in time of the intensity profiles of the lasers and pointing variations inducing changes in the phase matching conditions. It is not clear at this point whether the precision is limited by structural effects or instrumental/optical variations from sample to sample. Nevertheless, we have achieved sufficient precision for our purpose (see Table 2), as discussed at the end of this section.

The 2D spectra were acquired at 0.5 s per data point, with each spectrum comprising 4545 points (Fig. 4). The precisions resulting from the amplitude measurements compared with the integral measurements (Table 2) show that there is neither a need for a full 2DIR spectrum nor for a 2D fit of the entire spectrum for relative quantification. Once features are located, assigned, and identified, the peak strengths can be retrieved from one well-chosen data point. So long as a point on the spectrum that has relatively little interference from surrounding features (i.e., the spectral feature is resolved) can be found, the measurements of peak amplitudes are accurate enough. The data, therefore, suggest that it would be possible to measure a number $N$ of amino acid quantity ratios (so $N + 1$ peak amplitudes) in a time of $(N + 1)/2$ s with sufficient precision. It is likely that with the current state of the art the initial throughput will be somewhat slower due to automation overheads, and so probably a time of 1 to 5 s per amino acid feature would be more realistic in the case of protein samples.

Given that quantification of amino acid levels using 2DIR seems feasible, the outstanding question is how many amino acid ratios one would need to measure to identify a protein unambiguously. To answer this question, a full bioinformatics study is needed and currently is under way (it will be presented in detail elsewhere). Nevertheless, we have constructed a database of the amino acid ratios of 48,000 human proteins (using the National Center for Biotechnology Information [NCBI] reference sequence collection) and have performed preliminary tests to investigate the number $N$ of amino acid quantity ratios required. Table 3 summarizes our results.

The number of ratios required depends on the protein to be identified and on the type of amino acids that are effectively discovered and quantified by 2DIR FWM. That is why the number of ratios (peaks) in Table 3 is given as intervals. From this preliminary investigation, we can say that, with the current state of the experimental procedures ($\pm 12.5$% precision), four to nine peaks would need to be measured to uniquely identify a protein. This means that protein identification should take between 4 and 45 s considering, as discussed previously, an acquisition time of 1 to 5 s per point. These acquisition times fulfill the requirement of high throughput necessary for the development of a protein fingerprinting technique.

For the data presented here, the average number of molecules in the laser spot is approximately $10^{11}$ (17 pmol), and good fingerprints can be obtained at approximately $10^{11}$ to $10^{12}$ molecules. With small modifications to the existing apparatus, but using the same general approach, we would expect $10^{11}$ molecules in the beam to be achievable; however, this has yet to be demonstrated, and we do not yet know the absolute lower limit for sensitivity. There are other variants of nonlinear optical spectroscopy that we plan to assess for viability as fingerprinting methods, but neither their utility nor the possibilities of analyzing post-translational modifications have yet been established.

### Conclusion

We have demonstrated the basis of an alternative analytical approach to protein identification based on 2DIR spectroscopy by measuring 2D spectra of peptide side chains. Specific features corresponding to the aromatic vibrational modes of Tyr and Phe, as well as an aliphatic mode, were identified. Using the spectra obtained for various amino acid sequences containing different amounts of Phe and Tyr, it was shown that the relative integrated signal strengths and the relative amplitudes are propor-

---

### Table 2

Standard deviations of the data points (Figs. 6 and 7) out of the linearity, reflecting the dispersion from linear behavior

<table>
<thead>
<tr>
<th>Peak</th>
<th>Integral data (%)</th>
<th>Amplitude data (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe 1470/3050</td>
<td>±14.0</td>
<td>±13.5</td>
</tr>
<tr>
<td>Phe 1480/3070</td>
<td>±15.0</td>
<td>±10.5</td>
</tr>
<tr>
<td>Phe 1515/3110</td>
<td>±13.0</td>
<td>±10.5</td>
</tr>
<tr>
<td>Tyr 1530/3130</td>
<td>±11.5</td>
<td>±12.0</td>
</tr>
<tr>
<td>Average value</td>
<td>±13.4 (±1.5)</td>
<td>±11.5 (±1.6)</td>
</tr>
</tbody>
</table>

---

### Table 3

Numbers of amino acid quantity ratios and (number of peaks) necessary to identify a protein as a function of the available mass information and the precision of the 2DIR measurements

<table>
<thead>
<tr>
<th>Precision on</th>
</tr>
</thead>
<tbody>
<tr>
<td>ratios (%)</td>
</tr>
<tr>
<td>mass (±2%)</td>
</tr>
<tr>
<td>mass (±5%)</td>
</tr>
<tr>
<td>No mass information</td>
</tr>
<tr>
<td>±10</td>
</tr>
<tr>
<td>±20</td>
</tr>
<tr>
<td>±30</td>
</tr>
</tbody>
</table>

Note. Numbers of peaks are in parentheses.
tional to the number of Phe or Tyr in the peptide relative to
the number of methylene groups. This proportionality is
achieved with a precision of approximately ±12.5%, which
in principle would be enough for the purpose of finger-
printing. Our data did not show any evidence of structural
effects that would affect the precision of the data.

It should be noted that the number of amino acid peaks
to be measured depends a priori on three factors: the search
pattern of amino acids chosen, the protein to be identified,
and the precision of the 2DIR measurements. With the pre-
cision demonstrated here (±12.5%), the number of peaks
required would be, in principle, between four and nine.
Moreover, it was shown that simple measures of the
2DIR signal amplitude at the peak positions is sufficient
for this purpose and that there is no need to measure full
2D spectra. This suggests that protein identification could
take between 4 and 45 s.

Acknowledgments

We acknowledge Dr. Darek Kedra of the Bioinformatics
Support Service at the Imperial College London. This work
was financially supported by the Engineering and Physical
Sciences Research Council (EPSRC), the Royal Society,
and the Chemical Biology Centre Doctoral Training Centre.

References

[1] M. Iwaki, A. Puustinen, M. Wikstrom, P.R. Rich, Structural and
chemical changes of the P–M intermediate of
Paracoccus denitrificans
cytochrome c oxidase revealed by IR spectroscopy with labeled

Shannon, J. Hirst, R. Lawrence, P.R. Rich, ATR–FTIR redox
difference spectroscopy of
Yarrowia lipolytica
and bovine complex I,

catalytic site by FTIR spectroscopy, Biochem. Soc. Trans. 33 (2005)
886–889.


as building blocks for proteins: Comparative theoretical and spec-

[6] D. Zhang, Y. Xie, M.F. Mrozek, C. Ortiz, V.J. Davisson, D. Ben-
Amotz, Raman detection of proteinase analytes, Anal. Chem. 75

drop coating deposition Raman method for protein analysis, Anal.

detection of peptide tyrosine phosphorylation, Anal. Biochem. 332

[9] J. Filik, N. Stone, Drop coating deposition Raman spectroscopy of

[10] J.C. Wright, Coherent multidimensional vibrational spectroscopy,


[12] R.M. Hochstrasser, M.C. Asplund, P. Hamm, N-H. Ge, Femtosec-
ond two-dimensional infrared spectroscopy, J. Chin. Chem. Soc. 47

[13] A.M. Moran, S-M. Park, J. Dreyer, S. Mukamel, Linear and
nonlinear infrared signatures of local (α- and 310-helical structures

[14] J. Bredenbeck, P. Hamm, Peptide structure determination by two-
dimensional infrared spectroscopy in the presence of homogeneous
1578.

dimensional infrared spectroscopy displays signatures of structural

[16] J-H. Ha, Y.S. Kim, R.M. Hochstrasser, Vibrational dynamics of N-
64508/1–64508/10.

[17] Y.S. Kim, R.M. Hochstrasser, Dynamics of amide-I modes of the

hydrogen-bond dynamics in a β-turn transient two-dimensional infrared

two-dimensional infrared spectroscopy: Insight into why vibrational
spectroscopy is sensitive to DNA structure, J. Phys. Chem. B 110

exciton dynamics revealed by two-dimensional optical spectroscopy,

spectroscopy of water: I. Vibrational dynamics in two-dimensional IR

[22] K.A. Meyer, J.C. Wright, Interference, dephasing, and coherent
control in time-resolved frequency domain two-dimensional vibra-

[23] W. Zhao, J.C. Wright, Spectral simplification in vibrational spec-
roscopy using doubly vibrationally enhanced infrared four wave

[24] D.E. Thompson, J.C. Wright, Model for spectral artifacts in two-
dimensional four-wave mixing spectra from absorption and refractive
11282–11289.

[25] W. Zhao, J.C. Wright, Doubly vibrationally enhanced four wave
1411–1414.

[26] J.C. Wright, P.C. Chen, J.P. Hamilton, A. Zilian, M.J. LaBuda,
Theoretical foundations for a new family of infrared four-wave

[27] K. Kwak, S. Cha, M. Cho, J.C. Wright, Vibrational interactions of
acetonitrile: Doubly vibrationally resonant IR–IR-visible four-wave

identification and decongestion of Fermi resonances by control of
pulse time ordering in two-dimensional IR spectroscopy, J. Chem.

[29] E. Atherton, R.C. Sheppard, in: D. Rickwood, B.D. Hames (Eds.),
Solid State Phase Peptide Synthesis, Oxford University Press, Oxford,

[30] V. Kopecky Jr., V. Baumruk, Structure of the ring in drop coating
deposited proteins and its implication for Raman spectroscopy of

[31] W. Zhao, J.C. Wright, Measurement of γ(3) for doubly vibrationally
enhanced four wave mixing spectroscopy, Phys. Rev. Lett. 83 (1999)
1950–1953.

[32] H.F. Shurvell, in: J.M. Chalmers, P.R. Griffiths (Eds.), Handbook of
1791–1794.

constants of liquid benzene-h, at 25 °C extended to 11.5 cm-1
and molar polarizabilities and integrated intensities of benzene-h, between