Quantifying Secondary Structure Changes in Calmodulin using 2D-IR Spectroscopy

Lucy Minnes,1 Daniel J. Shaw,2 Benjamin P. Cossins,2 Paul M. Donaldson,3 Gregory M. Greetham,3 Michael Towrie,3 Anthony W. Parker,3 Matthew J. Baker,4 Alistair J. Henry,2 Richard J. Taylor,2 and Neil T. Hunt*1

1) Department of Physics, University of Strathclyde, SUPA, 107 Rottenrow East, Glasgow, G4 0NG, UK
2) UCB Pharma Ltd, Slough, UK, SL1 3WE
3) STFC Central Laser Facility, Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxon, OX11 0QX, UK
4) WestCHEM, Department of Pure and Applied Chemistry, Technology and Innovation Centre, 99 George Street, University of Strathclyde, Glasgow, G1 1RD, UK

* Corresponding author email address: neil.hunt@strath.ac.uk

ABSTRACT: Revealing the details of biomolecular processes in solution needs tools that can monitor structural dynamics over a range of time and length scales. We assess the ability of 2D-IR spectroscopy in combination with multivariate data analysis to quantify changes in secondary structure of the multifunctional calcium-binding messenger protein Calmodulin (CaM) as a function of temperature and Ca2+ concentration. Our approach produced quantitative agreement with circular dichroism (CD) spectroscopy in detecting the domain melting transitions of Ca2+-free (apo) CaM (reduction in α-helix structure by 13% (CD) and 15% (2D)). 2D-IR also allows accurate differentiation between melting transitions and generic heating effects observed in the more thermally-stable Ca2+-bound (holo) CaM. The functionally-relevant random-coil-α-helix transition associated with Ca2+ uptake that involves just 7-8 out of a total of 148 amino acid residues was clearly detected. Temperature-dependent Molecular Dynamics (MD) simulations show that apo-CaM exists in dynamic equilibrium with holo-like conformations while Ca2+ uptake reduces conformational flexibility. The ability to combine quantitative structural insight from 2D-IR with MD simulations thus offers a powerful approach for measuring subtle protein conformational changes in solution.

The link between structure and function in biological molecules is well-established but, to fully understand the molecular mechanisms occurring under physiological conditions, conformational changes spanning a range of time and length scales must be quantified. Ultrafast 2D-IR spectroscopy, in combination with isotopic substitution or non-natural labelling strategies has been used to extract site-specific information on structure, dynamics and function of biological molecules.1-3 However, in certain cases, larger-scale structural fluctuations of a protein may be central to function, while unravelling allosteric phenomena requires a more global perspective of bio-molecular behavior.

To explore such large scale structural changes in proteins there is thus a role for ‘whole molecule’ 2D-IR spectroscopy of unlabelled proteins to reveal a broader-scale picture of conformational change and create a platform for spatially-targeted studies. Vibrational coupling between peptide amide I modes in secondary structure elements are uniquely visible via off-diagonal peaks in 2D-IR spectra allowing quantification of contributions to equilibrium structure.10-14 Temperature jump or photochemical initiation methods15,16 extend the range of the accessible potential energy surface through ns to ms timescale dynamics.17 However, obstacles to whole molecule 2D-IR experiments lie in the difficulty of differentiating small changes to overlapping contributions to the amide I mode and correlating spectral changes with specific biological function.

Our goal is to develop tools using 2D-IR that can overcome these barriers using a stepwise approach whereby controlled variation of a protein structure generates a 2D-IR signature that can be correlated to the structural perturbation. Here, we report 2D-IR spectra probing the response of human Calmodulin (CaM) to temperature and Ca2+ ion binding and evaluate the use of multivariate data analysis (Principal Component Analysis, PCA, which has not previously been used in combination with protein 2D-IR spectroscopy) to quantify secondary structure change, benchmarking 2D-IR with CD spectroscopy.

CaM is a multifunctional calcium-binding messenger protein found in many eukaryotic cells. The predominantly α-helical structure of CaM is largely conserved across a range of organisms,18-22 comprising two (C- and N-terminal) globular domains, each consisting of helix-loop-helix E-F hand motifs (Fig.1, red) accommodating a pair of Ca2+ ion binding sites. The C- and N-terminal domains are connected by a central linker, which features two α-helices (Fig.1, orange). In apo-CaM (no Ca2+), these are joined by a short random coil (Fig.1(a), blue). Upon uptake of four Ca2+ ions in a cooperative two-step process
The apo-holo structural transition is well-established, but significant structural plasticity in both states is thought to underpin the adaptable nature of CaM in vivo. This complexity means that no clear picture of the relationship between structure, dynamics and function exists, making CaM a candidate for time-resolved 2D-IR studies, to complement the single previous time resolved study exploiting UV-visible absorption of the Tyr-138 residue that reported two-state unfolding of the C-terminal domain. CaM also provides an excellent test-bed for our tools designed to quantify secondary structure changes and establish the numbers of amino acid residues involved. With the capability established we will be able to follow protein structural dynamics in real time using time-resolved-2D-IR.

Full interpretation of 2D-IR spectra is facilitated by high level molecular dynamics (MD) simulations. The similarity between the time resolution of 2D-IR spectroscopy and the time-step of MD simulations means that the two approaches are complementary, allowing experimental validation of simulations to drive improved predictive abilities. Here we compare 2D-IR data to MD simulations, demonstrating how both provide insights into solution-phase molecular mechanisms.

**METHODS AND MATERIALS**

**Calmodulin expression:** *Escherichia coli* DH5α (Invitrogen) competent cells were used in the transformation of human CaM plasmid DNA. *E. coli* BL21 (DE3) Star (Novagen) competent cells were used in the expression of the protein. Sample integrity was verified by SDS Page analysis. Full experimental details are given in the Supporting Information. CaM cell pellets were re-dissolved and subject to AKTA purification. Purity was checked via size-exclusion chromatography and molecular weight verified with mass spectrometry.

**IR spectroscopy:** All samples for IR absorption and 2D-IR measurements were prepared at a CaM concentration of 0.8 mM in a Tris-DCI buffer (pD 7.4). Holo-CaM samples contained CaCl2 in a 10:1 stoichiometric excess. Samples were housed between two CaF2 windows separated by a 50 μm thick poly-tetrafluorethylene spacer in a temperature-controlled sample cell accurate to ±1 °C. Fourier Transform (FT)-IR absorption spectra were obtained using a Bruker Vertex 70v spectrometer with a resolution of 1 cm⁻¹ and were the average of 20 co-added scans. 2D-IR spectra were acquired using the ULTRA laser system via the FT-2D-IR method, reported previously, in which the mid IR pulses were arranged in a pseudo pump-probe geometry. The waiting time at which all spectra were acquired was 250 fs.

**IR Data Analysis:** IR absorption and 2D-IR spectra were analysed using difference spectroscopy and PCA. Difference spectra were obtained by subtraction of pairs of spectra following normalization to account for small fluctuations in signal intensity between measurements. Analysis of normalized and un-normalized 2D-IR data was carried out for comparison and no significant variation in the overall outcome was observed. Further examples of the use of non-normalized data are shown in the consideration of changes in the transition dipole moment of the sample in the discussion. PCA was applied using the Multivariate Analysis suite available with Origin Pro 2015, with spectral contributions to PCs reconstructed from the score values produced.

**Circular Dichroism (CD) spectroscopy:** CD spectra were recorded on a Chirascan qCD spectrometer with Peltier temperature control. CaM was diluted to ~12 μM (0.2 mg/ml) using Tris-HCl buffer, pH 7.4. The CD spectra were processed using algorithms from Dichroweb, and CDNN. The Dichroweb algorithms were a ridge regression method, CONTINLL, and a self-consistent method, SELCON3.

**Differential Scanning Calorimetry (DSC):** DSC experiments were performed using a VP-Capillary microcalorimeter with samples of CaM (300 μl) prepared at a concentration of 180 μM (3 mg/ml) in 20 mM HEPES buffer, 100 mM KCl, pH 7.5, with the addition of 2.5 mM EDTA and 2.5 mM EGTA for the apo-CaM samples, and the addition of 2 mM CaCl2 for the holo-CaM samples.

**MD simulations:** Structures of apo- (1CFD) and holo-CaM (1CLL) obtained from the Protein Data Bank (PDB) were the starting point for MD simulations. Missing residues were added using Maestro based on complete sequences from UniProt. The protein was solvated in an octahedral box with neutralising Na⁺ and Cl⁻ ions (~0.15 M). The AMBER ff14SB forcefield was used for the protein along with TIP3P water and recent Merz parameters for Ca²⁺ ions.

All MD simulations used GROMACS 5.0.4 with 2 fs time-steps and particle mesh Ewald for long range electrostatics with appropriate potential truncation and smoothing. Snapshots were taken every 10 ps. An energy minimization step relaxed the structure avoiding steric clashes or incorrect geometries. A position restraining force was applied to heavy atoms to stabilise the protein whilst equilibrating. The system was stabilized both isothermally and isobarically using an NVT temperature step and an NPT ensemble for 1 ns to test for stability. The production MD was run over 1 μs. The processed trajectories were analysed using the ‘Define Secondary Structure of Proteins (DSSP) hydrogen bonding algorithm’ to identify backbone hydrogen bonds and classify secondary structure. The root-mean squared (RMS) structural deviation was calculated with reference to the PDB structure.

**Figure 1:** Molecular structures of a) apo-CaM and b) holo-CaM showing α-helices (red); β-strands (yellow); flexible central linker (orange and blue); Ca²⁺ ions (grey).
Figure 2: a) Temperature dependence of the amplitude of the 220 nm peak in CD spectra of apo-CaM (black) and holo-CaM (red). Solid black line shows the result of fitting to a sigmoidal function (see text). The blue dashed line is to guide the eye. b) DSC data for apo-CaM (red) and holo-CaM (blue). Vertical dashed lines indicate the melting temperatures (T_m) determined via Gaussian fitting (see text).

RESULTS

CD spectroscopy and DSC: CD spectroscopy (Fig.2(a)) and DSC measurements (Fig.2(b)) on apo- and holo-CaM provide a benchmark for 2D-IR results. In particular, CD spectroscopy provides an accepted standard for quantifying secondary structure content. The CD and DSC results agreed well with previous reports and are not discussed in detail, but are presented fully in the supporting information.

The CD spectra of both apo- and holo-CaM featured two negative peaks at 208 and 222 nm, consistent with a mainly α-helical protein (Fig.S-1). The peaks diminished in size with increased temperature following a sigmoidal trend for apo-CaM, (mid-point: ~55 °C, Fig.2 (a), black) but an approximately linear trend with no sigmoidal component for holo-CaM (Fig 2(a), red). Secondary structure content was estimated using the SELCON3, 57,58 CONTINLL, 55,56 CDNN 54 algorithms and the results averaged. The detailed results are shown in Table S-1.

The DSC measurements on apo-CaM showed a peak near 60 °C which was well-represented by two Gaussian functions with melting temperatures of (T_m) ~46 and ~58 °C (Fig.2(b), red). For holo-CaM (Fig.2(b), blue) the two peaks were shifted to much higher temperatures (96 and 111 °C respectively). Comparison of the DSC and CD data with previous results allow assignment of the observed features to thermal unfolding of apo-CaM in a two-step process, with the C-terminal domain unfolding at ~46 °C followed by the more stable N-terminal domain at ~ 58 °C. Addition of Ca^{2+} forms the more thermally stable holo-protein, for which no domain melting occurs below 96 °C.

Figure 3: a) IR absorption spectra of apo-CaM. b) Difference IR absorption spectra for apo-CaM. Blue arrows show trends with increased temperature. The spectral distribution of PC2 is shown by dashed line (red). c) Temperature-dependent variation in the peak to peak magnitude of difference spectral signal (black) and PC2 coefficients (red). Relative Trend Value compares the peak-to-peak amplitude of the difference spectral data (black) and the PC2 coefficients (red) as a function of temperature. For ease of comparison these have been normalized using standard normal variate method. d) Comparison of the PC2 results for apo- (black) and holo-CaM (red). Absolute PC2 coefficients for each conformation have been scaled by percentage of total variance accounted for by PC2 to enable a valid comparison. Dashed lines in c) and d) are to guide the eye.

IR Absorption Spectroscopy: IR absorption spectra of apo-CaM at temperatures between 20 and 80 °C show a broad peak near 1645 cm^{-1}, assigned to the amide I mode of the protein (Fig.3(a)). The band became broader and the peak shifted to higher wavenumber as the temperature increased (see arrows). A clear isosbestic point is visible at 1650 cm^{-1} (Figs.3 (a&b)).

Difference spectral analysis, performed by subtracting the 20 °C spectrum from spectra obtained at elevated temperatures (Fig.3(b)) showed that the results could be well-represented by the sum of two Gaussian lineshape functions: a negative peak, corresponding to a loss of intensity, at ~1636 cm^{-1} and a positive peak at ~1671 cm^{-1} (See column FTIR_D in Table S2).

Application of PCA to the IR absorption data for apo-CaM (a summary of the PCA method and results are shown Figure S-2) showed that the variance in the IR absorption spectra is accounted for by two principal components (PC, 99.97% of the total variance). The first PC (PC1, 99.29%) was temperature independent (the relative loading of PC1 did not vary with temperature) while the second (PC2, 0.68%) showed a strong temperature-dependence. The spectral density associated with PC1 closely resembled the average IR absorption spectrum over the temperature range studied (Figure S-2). The PC2 spectrum (Fig.3(b)) was similar to the average of the temperature-dependent difference spectra (Fig.3(b)) and was well-represented by two Gaussian lineshape functions located at 1634 (negative) and 1667 cm^{-1} (positive) respectively (Table S-2, column FTIR_{PCa}).

Comparing the normalized temperature-variation of the contribution of PC2 (Fig.3(c), red points) to that of the peak-to-peak magnitude of the difference spectral feature (black points)
Although the spectra obtained for *holo*-CaM and *apo*-CaM, were broadly similar, the magnitude of the variation in signal was smaller for *holo*-CaM (Fig. 3(d)) and the trend was linear, rather than sigmoidal. A departure from linearity is noted above 70 °C for *holo*-CaM. This consistent with CD and DSC spectroscopy results which indicate a more thermally-stable protein after Ca\(^{2+}\) uptake.

**2D-IR Spectroscopy:** At 20 °C, the 2D-IR spectrum of *apo*-CaM contains a negative diagonal peak (red) at a pump and probe frequency of ~1645 cm\(^{-1}\) (Fig. 4(a)). This is assigned to the v=0-1 transition of the *apo*-CaM amide I mode.\(^{10-14}\) A positive off-diagonal feature (blue), assigned to the v=1-2 amide I transition, is shifted to lower probe frequency by the anharmonicity of the mode. The anharmonicity was not constant across the amide I band: fitting slices through the 2D-IR data at pump frequencies of 1640 and 1660 cm\(^{-1}\) to two Gaussian functions returned anharmonic shifts of 17 cm\(^{-1}\) and 23 cm\(^{-1}\) respectively (Table S2, columns 2DIR\(_{1640cm^{-1}}\) and 2DIR\(_{1660cm^{-1}}\)).

The 2D-IR spectrum of *apo*-CaM at 80 °C (Fig. 4(b)) qualitatively resembles that at 20 °C, but difference 2D-IR spectra reveal temperature-dependent changes. The full set of difference 2D-IR spectra are shown in Fig. S-4, the average is displayed in Fig. 4(c) as an example. A positive diagonal feature near 1635 cm\(^{-1}\) alongside a negative feature located just below the diagonal near 1660 cm\(^{-1}\) occur in the region of the amide I (0-1) band. These indicate a loss of the v=0-1 intensity at 1635 cm\(^{-1}\) (Fig. 4(c), black arrow) and a gain at 1660 cm\(^{-1}\) (Fig. 4(c), red arrow), corresponding to the changes observed in the IR absorption spectra. Accompanying peaks with the opposite sign in the position of the v=1-2 transitions are observed (blue and purple arrows respectively) as would be expected. The values of the anharmonic shift between the v=0-1 and v=1-2 components of the difference spectral features at 1635 and 1660 cm\(^{-1}\) were 16 and 28 cm\(^{-1}\) respectively. The former is similar to that obtained from the 2D-IR spectrum at 20 °C, the latter is somewhat larger (\(\pm\) 23 cm\(^{-1}\)), reflecting the shift of the 1660 cm\(^{-1}\) feature off the spectrum diagonal.

Applying PCA revealed two PCs (99.39%, Figure S-5). PC1 (98.62%) was temperature independent while PC2 (0.77%) showed a strong temperature dependence. The spectral distribution of PC2 (Fig. 4(d)) compared well with the average difference spectrum (Fig. 4(c)). The temperature-dependent contribution of PC2 (Fig. 4(e) red) closely matched the behavior of the features obtained from difference spectroscopy (black). The close agreement between PCA and difference spectroscopy results confirms the validity of PCA to extract the salient features of the 2D-IR spectra and we will consider only PCA-derived results henceforth. In addition to the more efficient analysis, an advantage of PCA is the improved signal to noise ratio of the PC2 spectral density in comparison to the difference 2D-IR spectrum. This is because variations from random noise do not correlate with temperature in the same way as protein spectral changes and so appear in a different PC. It is noticeable that the 2D-IR data do not return the clear sigmoidal profile for *apo*-CaM as observed by IR absorption spectroscopy. This arises from the nature of the 2D-IR signals and is discussed in detail below.

**2D-IR spectroscopy on holo-CaM** showed a similar pattern to that observed for the *apo*-protein. The change in the v=0-1 feature at 1635 cm\(^{-1}\) in the PC2-derived spectrum dominated any positive contribution near 1660 cm\(^{-1}\), consistent with the IR absorption spectroscopy results (Fig. S-6). The signals showed
MD Simulations: The time-averaged changes in secondary structure between apo- and holo-CaM determined using GROMACS and DSSP (Fig.5(a&b)) agree well with experiments. The predicted α-helical content was found to be consistent to within 4% of the PDB values in both cases. A plot of RMS-distance as a function of time provides an indication of the relative stability of each CaM conformation. The results for apo-CaM (Fig.5(c, black)) show that the RMS-distance parameter adopts an overall average of 7 Å, indicating a degree of variance from the starting structure. Sudden changes in the RMS-distance value provide evidence of conformational sampling (flexibility) by apo-CaM. In the initial 0.1 μs of the simulation, the apo-form adopted a conformation similar to that of the published structure (PDB 1CFD) giving an RMS-distance value of 4 Å. This structure was dominant during the simulations but others included a structure in which the central α-helix was completely disrupted, appearing to bend back upon itself (RMS-distance = 7 Å, Fig.5-S-7). The structure that minimized the RMS-distance parameter (3.4 Å) showed a straighter (but still disrupted) central helical section, sharing characteristics of the apo- and holo-conformations. Finally, an RMS-distance value of 9 Å was produced by the central helix folding over upon itself. Simulations of the holo-CaM conformation (Fig.5(c, red)) showed that the holo form is the more thermally stable of the two studied, undergoing fewer departures from the dominant structure.

Figure 5: MD-derived structures of a) apo- and b) holo-CaM. (c) RMS distance time course plot for apo- (black) and holo-CaM (red). (d) RMS fluctuations per residue. Positions for the amino acids displaying greatest RMS fluctuation are marked in the structures using numbers in (a) and (b).

Additional insight into the differences between the proteins is available from the RMS-fluctuation plot (Fig.5(d)), which identifies the protein residues displaying the greatest fluctuation. Residues 90-148, comprising the apo-C-terminal domain (black) exhibit the largest deviation from their starting position, reflecting the flexibility of this domain. Conversely, the holo-CaM state (red) shows a marked decrease in fluctuations although the largest degree of movement occurs in the C-terminal domain.

Repeating the MD simulation at higher temperatures with the default GROMACS method predicted negligible structural changes occurring up to 80 °C. The lack of any detectable change chiefly arises because the default simulation is unable to sufficiently sample the timescales required at higher temperatures over the relatively short 1 μs simulation time. To circumvent the timescale limitation at higher temperatures, an enhanced form of sampling, which increases the sampling rate using a number of replicates at different temperatures known as REST2 sampling was applied to 20 and 80 °C simulations of apo-CaM. Comparing the two approaches at 20 °C showed good agreement, with 54% and 52% α-helix content obtained via REST2 and GROMACS models respectively. The REST2 simulation predicted a 9% decrease in α-helix content at 80 °C with a full convergence of the simulation attained 0.6 μs into the 1 μs simulation time (Fig.5-S-8). Applying this approach to holo-CaM predicted a 4% decrease in α-helix content at 80 °C, reflecting the increased thermal stability of holo-CaM. The detailed results of the MD structure calculations are summarized in Table S3.

Figure 6: a) Buffer-subtracted unnormalized IR absorption spectra and b) 2D-IR diagonal (absolute value, raw data) of apo-CaM. c) Integrated area for each temperature plotted for FTIR (black) and 2D-IR diagonal data (red). d) Calculated d(o)/dH spectra for apo-CaM.

DISCUSSION

Spectral assignment: The IR absorption and 2D-IR data for apo-CaM indicate that the amide I band of the protein shifts toward higher frequency as the temperature increases. In both experiments spectral density was lost at 1635 cm⁻¹ and gained near 1660 cm⁻¹. The results for the holo-CaM sample were similar, though subtle differences in the thermal responses of the two states were observed.

DSC results show that heating apo-CaM causes C- and N-terminal domain melting at 46 and 57 °C respectively, but that holo-CaM is thermally stable up to 96 °C. The fact that the IR spectra for apo- and holo-CaM show subtle differences while the CD and DSC data indicate substantially different thermal stabilities for the two forms of CaM makes accurate assignment of IR spectra imperative. Thermally-induced IR spectral changes have two possible origins: 1) general heating, which shifts the amide I band to higher frequency via weakened H-bonds within the protein and between the protein and solvent. 2) A change in the secondary structure content of the protein may cause a spectral shift.
Several factors indicate that the changes in the apo-CaM spectra originate from a secondary structure transition: An isosteric point in the IR absorption data is consistent with a two-state transition, but is inconclusive in isolation. The spectral feature near 1635 cm$^{-1}$ that decreases in amplitude at elevated temperatures is very close to the expected amide I frequency of an $\alpha$-helix.\textsuperscript{72} The feature gaining in amplitude at 1665 cm$^{-1}$ (Fig.4(c)) is likely to contain contributions from weakened H-bonding and from a gain in random coil content accompanying the reduction in $\alpha$-helix. The amide I mode of a random coil typically appears at $\sim$1650 cm$^{-1}$ but shifts to higher wavenumber at higher temperatures. The overlap with the negative peak caused by loss of the $\alpha$-helix will bias the random coil peak to higher wavenumber. The appearance of the 1665 cm$^{-1}$ peak just below the 2D-IR spectrum diagonal suggests assignment to the gain of a more dynamic secondary structure component. This is because structural fluctuations lead to anti-diagonal elongation of inhomogeneously-broadened 2D-IR lineshapes and so much of the change in the lineshape caused by heating a random coil will occur in the off-diagonal region of the spectrum.\textsuperscript{73-77}

The anharmonicity, derived from 2D-IR spectroscopy, of the 1635 cm$^{-1}$ feature is small (\textsim\ 17 cm$^{-1}$, Table S2) compared to the feature at 1665 cm$^{-1}$ (23-28 cm$^{-1}$). The coupling of amide I oscillators in the $\alpha$-helix reduces the anharmonicity with increasing helix size,\textsuperscript{72} further suggesting that a decrease in the quantity of a secondary structure element with long-range structure is occurring, consistent with a helix to coil transition.

To confirm our assignment to an $\alpha$-helix to random coil process in apo-CaM, we apply the method of Grechko and Zanni,\textsuperscript{72} which combines linear absorption and 2D-IR to extract a spectrum of the transition dipole strength of the sample, $d(\omega)$. The method is based on the fact that the dependences of the signal intensity on the transition dipole moment of the linear absorption ($\propto|\mu|^2$) and 2D-IR techniques ($\propto|\mu|^4$) are different, leading to changes in the relative magnitudes of the FTIR and 2D-IR signals if the degree of vibrational coupling within the molecule changes.\textsuperscript{78} Specifically, the integrated area of an IR absorption spectrum is conserved when the coupling changes, but that of the 2D-IR spectrum diagonal is not. This can be visualized via the frequency-dependent ratio of the 2D-IR diagonal and IR absorption signals: the former increases relative to the latter in the event of increased vibrational coupling. Rather than applying the Grechko and Zanni methodology to extract the absolute transition dipole moment, which requires accurate calibration,\textsuperscript{72} we employ a modified approach to extract changes in the transition dipole strength relative to those at a reference temperature (20 °C), $d(\omega)^{rel}$:

\[
d(\omega)^{rel} = \frac{2DIR_{\beta}(\omega) \ OD_{20}}{FTIR_{\gamma}(\omega) \ AOD_{20}}
\]

Where $d(\omega)^{rel}$ is the ‘spectrum’ of the relative transition dipole strength. $2DIR_{\beta}(\omega)$ and $FTIR_{\gamma}(\omega)$ are the absolute values of the diagonal of the 2D-IR spectrum and the absorption spectrum at temperature $T$ respectively. $OD_{20}$ and $AOD_{20}$ are the maxima of the absorption spectrum and the 2D-IR diagonal signal at 20 °C respectively.\textsuperscript{72} It is important to note that this approach is valid only when the coupling does not affect the fundamental transition dipole strength (i.e. that of the random coil amide I vibration).

The results of applying this process (Fig.6) show that although the shape of the IR absorption spectrum changes with temperature (Fig.6(a)), the integral (Fig.6(c), black) is constant.\textsuperscript{72} Conversely, the integral of the 2D-IR spectrum diagonal (Fig.6(b) and 6(c), red) decreases as the temperature increases. This is due to changes on the low frequency edge of the amide I lineshape. Calculating $d(\omega)^{rel}$ (Fig.6(d)) reveals a peak near 1635 cm$^{-1}$, which decreases with increasing temperature. This is a firm indication of a loss of the vibrational coupling associated with an $\alpha$-helix. The region near 1660 cm$^{-1}$, assignable to a random coil, is unchanged with increasing temperature as the measured transition dipole reflects only changes in coupling, upon which the impact of random coil elements are negligible. The transition dipole method thus supports the conclusion that the temperature-induced spectral changes in apo-CaM are due to a helix-coil transition.

In the case of holot-CaM, the spectral features assigned to loss of $\alpha$-helix in the apo-CaM sample are still present, suggesting a reduction in $\alpha$-helix has occurred. Calculating $d(\omega)^{rel}$ for holot-CaM (Fig.S-9) also suggests a reduction in $\alpha$-helix content though on a smaller scale than for apo-CaM. Despite this apparent similarity between the two proteins, the gain in the 1665 cm$^{-1}$ random coil signal, suggested by PCA and difference spectroscopy above, is of much smaller magnitude in the holot-CaM data, suggesting that the two states behave differently.

Table 1: Summary of percentage changes of $\alpha$-helix and random coil induced by increasing the temperature of apo-CaM and holot-CaM from 20-80 °C.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha$-helix (%)</th>
<th>random coil (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(20-80°C)</td>
<td>(20-80°C)</td>
</tr>
<tr>
<td>CD</td>
<td>67-27%</td>
<td>17-38%</td>
</tr>
<tr>
<td>2D-IR</td>
<td>54-28%;</td>
<td>16-26%</td>
</tr>
<tr>
<td>$d(\omega)^{rel}$</td>
<td>54-22%</td>
<td>-</td>
</tr>
<tr>
<td>FTIR</td>
<td>54-50%</td>
<td>16-19%</td>
</tr>
<tr>
<td>Rest2 MD</td>
<td>54-45%</td>
<td>16-21%</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>71-39%</td>
<td>16-32%</td>
</tr>
<tr>
<td>2D-IR</td>
<td>58-36%</td>
<td>13-19%</td>
</tr>
<tr>
<td>$d(\omega)^{rel}$</td>
<td>58-31%</td>
<td>-</td>
</tr>
<tr>
<td>FTIR</td>
<td>58-54%</td>
<td>16-19%</td>
</tr>
<tr>
<td>Rest2 MD</td>
<td>58-53%</td>
<td>13-15%</td>
</tr>
</tbody>
</table>

Figure 7: Predicted $\alpha$-helix percentages as a function of temperature for apo-CaM (black) and holot-CaM (red). a) Results obtained from CD spectroscopy. b) Results obtained from transition dipole analysis ($d(\omega)^{rel}$) of 2D-IR data. Lines are to guide the eye.

Quantification of secondary structure change: Changes in the 2D-IR spectral distribution derived from PCA alongside the $d(\omega)^{rel}$ result indicate that a helix to coil transition occurs in apo-
CaM as the temperature increases. Holo-CaM exhibits a decrease in the signal arising from α-helices. This is qualitatively consistent with CD measurements, but we seek to determine the quantitative accuracy of 2D-IR.

To do this, we use the assignment of the 1635 cm⁻¹ and 1665 cm⁻¹ features to α-helix and random coil elements to allow direct quantitative comparison with CD results and with MD simulations. Although it is feasible to obtain absolute values of secondary structure content directly from IR data, it is non-trivial. Instead, we quantified changes in secondary structure from the spectral signals assuming a starting point of the α-helix and random coil content predicted by MD simulations at 20 °C (Table S-2), which are consistent with structural studies. The fractional change in the difference spectral amplitude, the PC2 coefficient multiplied by the PC2 spectral amplitude or the fractional change in d(ω) as a function of temperature was then used to estimate the fractional change in absolute secondary structure.

The results (Table 1; Fig.S-10) show that 2D-IR spectroscopy and the d(ω) approach both provide good agreement with CD spectroscopy, suggesting that the quantity of α-helix decreases steadily with increasing temperature for both states of CaM (Table 1). This apparent loss of α-helix yields increases in random coil content, though it is slightly more marked for the apo state. It is noted that the differing CD data analysis algorithms return a spread of results, leading to an uncertainty in the values recovered that is reflected with error bars on the CD analysis (Figure S-10).

It is important to consider the similar results obtained for apo- and holo-CaM from both 2D-IR and CD data in the context of DSC data, which implies that the two states of CaM respond differently to elevated temperature with holo-CaM resisting domain melting below 96 °C. The reasons for this discrepancy are that heating has two effects on the spectroscopy of the protein: 1) heating the α-helix without triggering a secondary structure transition; 2) the occurrence of a transition. DSC shows that the former is observed in both apo- and holo-CaM while the unfolding transition is exclusive to apo-CaM.

It is clear that heating the α-helix has a significant effect on both CD and 2D-IR (including the d(ω) approach) datasets, causing an apparent reduction in α-helix percentage for the thermally stable holo-CaM. The magnitude of both CD and 2D-IR measurements is sensitive to the interactions of residues within the α-helix, which explains the good quantitative agreement between them, but heating will increase average inter-residue separations, reducing interactions and so signal amplitudes without what would be considered to be a loss of a helical structure. This is especially true for the d(ω) approach, which specifically targets coupling-induced spectral changes. It is thus an oversimplification to interpret reductions in α-helix 2D-IR signal as a quantitative ‘loss’ of structure, as opposed to a thermal modification of the signal amplitude.

The challenge is to differentiate thermal effects from secondary structure transition in the data analysis. Comparison of the apo and holo-samples facilitates this. Comparing the α-helix content from CD (Fig.7(a)) and 2D-IR (d(ω) approach (Fig. 7(b)) methods for apo- (black) and holo-CaM (red) shows that the magnitude of the differences between the apo and holo datasets peaks at 60 °C, when domain melting occurs, following the DSC observations (Fig.2(b)). Taking the difference in predicted α-helix content at 60 °C (grey dashed lines and arrows, Fig.7) indicates a reduction in α-helix content of 13% (CD) and 15% (2D-IR) is associated with the unfolding transition. It is noteworthy that this brings the predicted structure change of CD and 2D-IR closer to the values from IR absorption (5%) and MD (9%).

**Detecting the Apo to Holo Transition:** The 2D-IR data reproduces the temperature-induced secondary structure change. We now consider the sensitivity of 2D-IR to small changes by probing the Ca²⁺-binding transition, which involves a 5% change in the α-helix content; this equates to just 7-8 residues.

The signal to noise ratio of the IR absorption difference spectrum for the apo-holo transition (Fig.8(a)) is poor, but a loss of spectral density at 1650 cm⁻¹ is accompanied by a gain at 1635 cm⁻¹. As expected if the linker undergoes a random coil to α-helix transition (Fig.1). This is much clearer in the 2D-IR spectrum (Fig.8(b)), where a gain of signal near 1640 cm⁻¹ (black arrow) accompanies a loss at higher wavenumber (red arrow). No thermal effects need to be accounted for in this transition. The random coil signal is dominated by the α-helix signal, as would be expected due to the larger transition dipole moment of the α-helix amide I mode.

Using the magnitude of the 2D-IR difference spectrum to evaluate the percentage change in α-helix content reveals a value of 12%. This overestimates the data from structural studies and MD predictions, as well as IR absorption (~6%)²⁵. Applying the d(ω) approach returns 6%, reinforcing the benefits of the transition dipole approach.

**Figure 8:** apo- to holo-CaM transition detected by a) IR absorption difference spectroscopy (holo- minus apo-) and b) 2D-IR difference spectroscopy. The color bar shows the magnitude of the difference spectrum following normalization of the raw 2D-IR spectra (see text).

**CONCLUSIONS**

Experiments to measure the thermal stability and Ca²⁺-binding transition of human CaM with IR absorption and 2D-IR spectroscopy, corroborated with DSC and CD measurements, show that 2D-IR can be used to reliably quantify relative changes in secondary structure elements in proteins using the
amide I band. We have shown that 2D-IR in combination with PCA quantitatively reproduces the temperature-dependent trends recovered from CD spectroscopy of the same protein samples and so can be considered, by this benchmark, to be an accurate reporter of protein secondary structure change that is robust, efficient to employ and non-supervised. The results reveal a nuanced situation in that changes in the 2D-IR signal caused by large changes in temperature include thermally-induced alterations in coupling as well as secondary structure transitions. The benefit of a reference sample (holo-CaM in this case) is vital to extract quantitative information relating to secondary structure change under these conditions. The 2D-IR results clearly go beyond conventional IR absorption whereby the absorption signal is not as sensitive to thermally-induced ‘relaxation’ of coupling within the α-helix as CD and 2D-IR. Combining IR absorption with 2D-IR spectroscopy of apo- and holo-CaM via the transition dipole approach and cross-referencing with DSC data suggests that the derived lower values for the secondary structure changes are likely to be closer to the actual values than numbers derived independently from CD or 2D-IR without comparing to a reference sample. When considering Ca²⁺-induced structural change in human CaM it has been demonstrated that 2D-IR spectroscopy is sensitive to changes in secondary structure at the few percent (7-8 residues in 148) level, providing superior contrast to IR absorption.

ASSOCIATED CONTENT
Supporting Information
Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION
Corresponding Author
*neil.hunt@strath.ac.uk

Author Contributions
The manuscript was written through contributions of all authors.

ACKNOWLEDGMENT
We thank Dr Oliver Durrant (CD and DSC measurements) and Dr Niall Simpson (protein production) for assistance. Funding from STFC (PhD studentship for LM and Programme Access to the ULTRA laser) is gratefully acknowledged. The authors declare no conflict of interest in the production of this work.

REFERENCES

(42) Andreasen, T. J.; Luetje, C. W.; Heideman, W.; Storm, D.


