Fast infrared spectroscopy of protein dynamics: advancing sensitivity and selectivity

Koziol, Klemens L; Johnson, Philip J M; Stucki-Buchli, Brigitte; Waldauer, Steven A; Hamm, Peter

Abstract: 2D-IR spectroscopy has matured to a powerful technique to study the structure and dynamics of peptides, but its extension to larger proteins is still in its infancy, the major limitations being sensitivity and selectivity. Site-selective information requires measuring single vibrational probes at sub-millimolar concentrations where most proteins are still stable, which is a severe challenge for conventional (FT)IR spectroscopy. Besides its ultrafast time-resolution, a so far largely underappreciated potential of 2D-IR spectroscopy lies in its sensitivity gain. The present paper sets the goals and outlines strategies how to use that sensitivity gain together with properly designed vibrational labels to make IR spectroscopy a versatile tool to study a wide class of proteins.

DOI: https://doi.org/10.1016/j.sbi.2015.03.012

Originally published at:
DOI: https://doi.org/10.1016/j.sbi.2015.03.012
Fast Infrared Spectroscopy of Protein Dynamics: Advancing Sensitivity and Selectivity

Klemens L. Koziol, Philip J. M. Johnson, Brigitte Stucki-Buchli, Steven A. Waldauer, Peter Hamm*

Department of Chemistry,
University of Zurich,
Winterthurerstr. 190,
CH-8057 Zürich, Switzerland
*peter.hamm@chem.uzh.ch

(Dated: March 27, 2015)

Abstract

Abstract: 2D-IR spectroscopy has matured to a powerful technique to study the structure and dynamics of peptides, but its extension to larger proteins is still in its infancy, the major limitations being sensitivity and selectivity. Site-selective information requires measuring single vibrational probes at sub-millimolar concentrations where most proteins are still stable, which is a severe challenge for conventional (FT)IR spectroscopy. Besides its ultrafast time-resolution, a so far largely underappreciated potential of 2D-IR spectroscopy lies in its sensitivity gain. The present paper sets the goals and outlines strategies how to use that sensitivity gain together with properly designed vibrational labels to make IR spectroscopy a versatile tool to study a wide class of proteins.
I. INFRARED SPECTROSCOPY OF PROTEINS

The principle information content of vibrational (IR) spectroscopy is very similar to that of NMR spectroscopy. Both are sensitive to chemical structure, that is, in both cases certain molecular groups result in absorption bands at specific spectral positions, both respond to the coupling to the environment, and both allow one to determine connectivity of molecular groups and thereby ultimately the 3D structure of a molecule by employing multidimensional techniques [1–3]. Fig. 1 underlines the similar character of both techniques by showing the IR and NMR spectra of a small peptide, Ala-Gly-Ala-Aib. However, NMR spectroscopy has long eclipsed IR spectroscopy as a prominent tool for biophysical studies, and Fig. 1 also illustrates why. That is, IR spectroscopy does not provide high spectral resolution, but an \( N \)-atomic molecule has \( 3N - 6 \) vibrational modes. Hence, while an assignment of vibrational spectra on the level of individual molecular groups might still be possible for very small peptides, the same fails for proteins due to spectral congestion. In essence, IR spectroscopy loses its chemical selectivity for even the smallest proteins (unless one works with specific vibrational labels, as discussed below). Another issue is the sensitivity of IR spectroscopy due to the relatively small IR cross sections. In contrast, NMR spectroscopy, in particular multidimensional techniques, has been successfully extended to ever larger proteins, which has made NMR commonplace in structural biology, a development that has

![FTIR and 1H-NMR spectra](image)

**FIG. 1:** The IR (left) and \(^1\text{H}-\text{NMR}\) (right) spectra of a small peptide, Boc-Ala-Gly-Ala-Aib-OMe. In both cases, bands corresponding to certain molecular groups appear at characteristic absorption frequencies. The FTIR spectrum shows all \( 3N - 6 \) normal modes, while a \(^1\text{H}-\text{NMR}\) spectrum reports on only the H-atoms in a molecule.
been celebrated by various Nobel prizes.

With this preamble, one might ask the question why one would ever want to measure an IR spectrum of a protein? There are niche (but still very important) problems, such as amyloid fibril formation or membrane bound systems, where NMR spectroscopy becomes more difficult, as it relies on rotationally diffusing molecules. But the untapped potential of IR spectroscopy lies in its high “direct time resolution”. NMR spectroscopy can elucidate dynamics on many timescales, even down into the picosecond regime, but it does so via relaxation experiments, which works only for equilibrium fluctuations. With “direct time resolution”, we have in mind to study non-equilibrium processes in a pump-probe fashion, such as the triggered (un)folding of a protein [4–11] or the propagation of a perturbation through an allosteric system [12–14]. The direct time resolution of a spectroscopic method is dictated by the dephasing time of its transitions, which is a few picoseconds for typical vibrational bands in the solution phase, in contrast to millisecond dephasing of NMR. IR spectroscopy is fast enough to time-resolve essentially any process of biological relevance, however, it is its fast dephasing that broadens the spectra and thus limits the spectral resolution. High spectral and temporal resolutions are mutually exclusive, hence, what appears to be the limitation of IR spectroscopy should rather be seen as its great opportunity.

II. ADVANCING THE SENSITIVITY

The vast majority of proteins aggregate at the concentrations typically used in FTIR spectroscopy (i.e., \( \geq 10 \text{ mM} \)). As a guideline of what is required, we note that NMR spectroscopy typically works at concentrations around 0.5 mM, but even in that regime only about 20% of nonmembrane proteins remain soluble [15]. The likelihood of a given protein to aggregate increases very steeply with concentration. Myoglobin, often called the hydrogen atom of biology, is still stable at 20 mM, and as such is a distinct exception in this regard. Site selective information will require detection of the vibration of a single molecular group. For what would be considered already a medium-strong IR absorber with an extinction coefficient of 400 M\(^{-1}\)cm\(^{-1}\) (e.g., azidohomoalanine discussed below) measured in a cuvette of 25 \(\mu\text{m}\) thickness, a band with 0.5 mOD absorption is expected at 0.5 mM concentration, which will sit atop a water (D\(_2\)O) background of 250 mOD at 2120 cm\(^{-1}\). These numbers set the goal necessary to make IR spectroscopy a versatile tool to study a wide class of proteins.
FIG. 2: 2D IR (left) and 2D NMR (right) exchange spectra, measuring the hydrogen bonding of phenol to benzene, or the complexation dynamics of Li$^+$ (green) to a crown ether (monobenzo-15-crown-5), respectively. Adapted from Refs. [16, 17] with permission.

2D IR spectroscopy is a relatively new technique [18], which transfers concepts from 2D NMR spectroscopy to vibrational transitions [2, 3]. Spreading spectra into two dimensions increases the spectral resolution, and at the same time allows one to correlate various transitions with each other. The most direct comparison of the two techniques can be made for exchange spectroscopy (EXSY) [16, 19, 20], for which virtually identical plots have been produced (Fig. 2). The only difference is the vastly disparate timescales probed by the two methods – picoseconds versus milliseconds –, which refers to the “direct time resolution” inherent to both techniques. In the context of protein dynamics, 2D IR spectroscopy has been applied to investigate the structure of membrane peptides [21, 22], fibril formation [23, 24], the structural flexibility of enzyme active sites [25], ligand migration in heme proteins [26], ligand binding [27, 28], and the dynamics of the protein hydration shell [29, 30], to list just a few examples. Recent reviews can be found in Refs. [31–34]
The time window that can be addressed by simple 2D IR spectroscopy is limited by the lifetime of the vibrational transition, which is short with a few 10s of picoseconds in the most favorable case. While proteins are also dynamical on these fast timescales, biological relevant processes are typically orders of magnitudes slower. Pump-probe type of experiments, also termed transient 2D IR spectroscopy, do not have that limitation and can span timescales from picoseconds to practically infinite, but their application is still relatively rare. In such an experiment, an actinic pump pulse triggers a non-equilibrium process and a transient 2D IR pulse sequence is used at a later delay time to explore the response of the sample (see Refs. [35–38] for reviews). Examples include studies of the folding/unfolding of small peptides after the photoreaction of an artificially incorporated molecular group [6, 39, 40] or after a temperature jump [9, 10].

A so far largely underappreciated potential of 2D IR spectroscopy, which might turn out to be the most important one in the context of the goal set above, lies in its sensitivity gain. First, this is the result of using coherent, highly brilliant laser-based IR light sources, in contrast to the black-body radiator in an FTIR spectrometer, together with the fact that detection can be background free when using a so-called box-CARS geometry [3]. More importantly, however, the 2D IR signal strength scales quadratically with the extinction coefficient, in contrast to linear scaling for IR absorption spectroscopy. For the example given above, the signal ratio of IR label to water background is expected to improve from 1:500 to 1:2.3. That is, the water background is huge just because its concentration is huge (56 M), but its extinction coefficient is only 1.8 M\(^{-1}\)cm\(^{-1}\).

III. ADVANCING THE SELECTIVITY

Site selective information from transient IR spectroscopy requires localized vibrational labels. In this regard, isotope labeling is most attractive, as it influences the chemical properties of a protein only marginally (if at all). For most cases, the –CO groups of the peptide backbone have been considered, either as single (\(^{13}\)C\(^{16}\)O) or as double (\(^{13}\)C\(^{18}\)O) substitution, where the latter is preferred since it completely separates the corresponding vibrational mode from the main amide I band. This approach has been used extensively to study structure and dynamics of small peptides [5, 7–9, 21, 22, 24, 34, 40–43], but unfortunately, it is essentially no longer feasible even for small proteins (elegant exceptions to
that statement exist [23]). First, the synthesis becomes extremely tedious once the size of
the protein exceeds that which can be handled on a peptide synthesizer (>50 amino acids),
so the protein has to be expressed in *E. coli*. Second and more importantly, the amide I
frequency is downshifted into a spectral range were many amino acid side chains absorb
as well, in particular those containing carboxylate (Glu, Asp) or amine (Asn, Gln, Arg)
groups [3]. While small model peptides can be designed to circumvent such interferences,
these amino acids are often crucial for the function and stability of wild-type proteins.

Vibrational labeling of larger proteins will therefore have to utilize special molecular
groups that appear in the spectral window between $\approx 1800 \text{ cm}^{-1}$–$2800 \text{ cm}^{-1}$ (see Fig. 1,
left). The only band from a naturally occurring amino acid in that frequency range originates
from the $-\text{SH}$ group of cysteine [44]. Alternatively, isotope labeling of $-\text{CD}$ vibrations of
deuterated amino acids has been proposed [45–47], but both are weak IR absorbers and as
such have been little explored as of yet.

The remaining option is using non-natural amino acids containing $-\text{C}=\text{O}$ [29, 30, 48],
$-\text{N}_3$ [27, 49–51], or $-\text{C}=\text{N}$ [28, 52] groups, many of which have been summarized in recent
reviews [53, 54]. Of course, with such labels one gives up one of the beauties of IR spec-
troscopy, i.e., one has to chemically modify the protein, but many of these IR labels are
small perturbations in comparison to for example fluorescence labels. Exploring the applic-
cability of these amino acids for time resolved studies is still in its infancy. In our opinion,
the following possibilities are most promising in terms of a high extinction coefficient (which
tentatively excludes $-\text{C}=\text{N}$ groups) and the straightforwardness to incorporate them into a
protein. The first is the asymmetric $\text{N}_3$ stretch vibration of azidohomoalanine (Aha) with a
reasonably high extinction coefficient of 300-400 M$^{-1}\text{cm}^{-1}$ [27] (note that the corresponding
number is wrong in Ref. [53]). It is a methionine analog that can be incorporated into a
protein using a Met auxotrophic mutant strategy [55]. Genetically encoded unnatural amino
acids is another possibility to incorporate an azide group into a protein [51]. In fact the orig-
inal purpose for which amino acids containing an azide group has been designed was “click”
chemistry, and consequently one may also add metal-carbonyl complexes with even stronger
C$=\text{O}$ bands to it in a post-translational coupling step as a second alternative [48]. Finally,
Re-carbonyl or Ru-carbonyl complexes have been bound selectively to histidines [30, 56–58].
In all three cases, the required amino acids (Met or His) can be placed essentially anywhere
in a protein by site-directed mutagenesis, provided the protein remains stable. The protein
should have only one of these amino acids at the desired position, and in this regard, Met is advantageous as it is relatively rare in natural proteins, whereas His is often crucial for their function. Note that we have disregarded cysteine as another possible anchor point for a label [29], since we want to keep that amino acid free for adding other functionalities, such as a photoswitch [6, 8, 12, 13, 59] used to initiate a non-equilibrium process (see the example highlighted below). We reiterate that being able to initiate a non-equilibrium process is a prerequisite to extend the accessible time window beyond the vibrational lifetime of the label.

IV. OUTLOOK AND CONCLUSION

To demonstrate that the anticipated concentration regime can indeed be reached with good signal-to-noise ratio and in fact with quite some room, Fig. 3a shows the 2D IR spectrum of Aha at 0.1 mM. The 2D IR band of Aha is sitting atop a broad water background (Fig. 3b), but owing to the quadratic extinction coefficient dependence of the 2D IR response, the water background is reduced to the extent that it can be readily subtracted off (Fig. 3c). We have studied the structure and flexibility of an Aha-labeled ligand bound to an allosteric protein, the PDZ2 domain, in this way [27].

Allostery is the propagation of a signal between two sites of a protein. As an approach to study its dynamics, Fig. 4a shows a PDZ2 domain with an azobenzene-derivative covalently linked across its binding groove [59] in such a way that photo-isomerization of the
FIG. 4: (a) A photoswitchable PDZ2 domain constructed to mimic the conformational transition after binding of a ligand to the binding groove. (b) Response of the protein measured via a localized mode situated on the photoswitch (green) and the amide I mode (red). Three phases can be distinguished: vibrational cooling (phase I), opening of the binding groove (phase II), and the subsequent structural response of the remainder of the protein (phase III). Adapted from Ref. [12] with permission.

A linker induces a conformational transition of the protein, which mimics that after ligand binding [12]. By chance, one vibrational mode localized on the azobenzene linker could be spectroscopically isolated in this case (Fig. 4b, green), which enabled us to draw very detailed conclusions on the dynamics of the binding groove [12, 13]. However, the subsequent response of the amide I band (Fig. 4, red) averages over the whole protein without any site-specific information, hence vibrational labels of the type discussed above in connection with transient 2D IR spectroscopy will be needed to study the allosteric response of more remote parts of the protein.
To conclude, while 2D IR spectroscopy has matured to a powerful technique to study the structure and dynamics of small peptides, the same is not yet the case for proteins, the major limitations being sensitivity and selectivity. The present paper sets the goals and outlines strategies to overcome these limitations. They pave the way towards a post structure-function-relationship era, in which we will explore the functional dynamics of protein structures.

V. ACKNOWLEDGEMENT

The work discussed here has been supported in parts by an ERC advanced investigator grant (DYNALLO) and the Swiss National Science foundation through the NCCR MUST. Valuable contributions of our coworkers, in particular Rolf Pfister, Reto Walser, Robbert Bloem, Amedeo Caflisch and Gerhard Stock are highly acknowledged.

  ● A comprehensive textbook on 2D IR spectroscopy, approaching the technique from a very practical point of view.


- One of the few recent examples, using transient 2D IR spectroscopy to study the unfolding of a peptide after a temperature jump, in connection with isotope labeling for site-selective information. That allowed the authors to retrieve very detailed information on the folding pathway.


- Another example of transient 2D IR spectroscopy of the temperature-jump induced unfolding of a peptide, however, without utilizing isotope labeling.


- Transient IR spectroscopy to study the structural response of an allosteric protein. A perturbation is initiated by a photoswitch that is covalently linked to the protein in a way that mimics ligand binding.


**One of the few examples of the study of a larger protein by utilizing 2D-IR spectroscopy in connection with isotope labeling. While not achieving single-amino acid selectivity, the authors selectively isotope-labeled complete protein domains.**


**One of the most beautiful examples, illustrating the power of time-resolved 2D IR spectroscopy to study peptide aggregation. Isotope labeling reveals very detail insights into the aggregation mechanism.**


- This paper illustrates that the Aha label, even though being only a medium-strong IR absorber, can be used to retrieve site-selective information from larger proteins down into a concentration regime which makes 2D IR spectroscopy a versatile tool to study a wide class of proteins. It also introduces methods to reliably subtract the water background (as shown in Fig. 3).


- This paper utilizes a Ru-carbonyl complex bound to the surface of a protein via a His residue to study the solvation dynamics of the protein.


[35] Bredenbeck, J., Helbing, J., Kolano, C., Hamm, P.. Ultrafast 2D-IR spectroscopy of


• Transient 2D IR spectroscopy to study the refolding of a peptide after triggering a photo-switch in the middle of an α-helix.


- The paper utilizes a Re-carbonyl complex bound to a protein via an Aha residue and a post-translational coupling step using “click” chemistry.


Santos-Silva, T., Mukhopadhyay, A., Seixas, J.D., Bernardes, G.J.L., Romão, C.C., Romão,
