Quantifying Biomolecular Recognition with Site-Specific 2D Infrared Probes

Philip J. M. Johnson,* Klemens L. Koziol,** and Peter Hamm*†

Department of Chemistry, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland

Supporting Information

ABSTRACT: Azidohomoalanine (Aha) is an unnatural amino acid containing an infrared active azido side chain group that can, through frequency shifts of the azido stretch vibration, act as a probe of local structure. To realize the potential of such structural probes for protein science, we have developed a two-dimensional infrared spectrometer employing fast mechanical scanning and intrinsic phasing of the resulting spectra, leading to a lower sensitivity limit of ∼100 μOD level samples. Using this approach, we quantify the biomolecular recognition between a PDZ2 domain and two Aha-mutated peptides. It is shown that this method can distinguish different binding modes and that the energetics of binding can be determined.

Biomolecular recognition underlies enzymatic activity and signal transduction pathways in effectively all cellular processes, where the energetics of molecular recognition in combination with local concentrations determine the likelihood of protein–protein or protein–ligand binding and therefore the probability of occurrence of different cellular processes. Knowing these energetics is necessary to both understand the native biochemistry of the cell and identify, design, and characterize new drugs with which to control cellular function.1 Knowing these energetics is necessary to both understand the native biochemistry of the cell and identify, design, and characterize new drugs with which to control cellular function.1 Quantifying biomolecular recognition has commonly been the domain of isothermal titration calorimetry (ITC), where the heat of binding between a protein and ligand is monitored as a function of their relative molar ratio.2 While not a direct measure of structural interactions, through prior knowledge of the binding properties of the biomolecules of interest, it is possible to recover the free energy, enthalpy, and entropy of a binding event.2–4 Alternative approaches to measuring protein–ligand interactions, often with greater structural specificity, consist of nuclear magnetic resonance (NMR) spectroscopy (from the point of view of either the protein5–8 or the ligand9,10), X-ray crystallography,10 or scattering,11,12 and mass spectrometry techniques,13 among many others, each with their own strengths, weaknesses, and applicable affinity regimes.1

Here, we propose the use of unnatural amino acids with infrared activity in combination with two-dimensional infrared spectroscopy (2D IR) as a means of quantifying equilibrium protein–protein or protein–ligand interactions. Various unnatural amino acids with infrared-active side chains have been developed,17–21 which have proven to be sensitive reporters of local structural changes (e.g., protein folding22 or dynamics23–25). In contrast with thermodynamic approaches, single-point mutations to infrared-active unnatural amino acids allow for site-specific, and thus structural, information regarding biomolecular recognition.26 However, experimental limitations have to now hampered the general applicability of this method. The molecular vibrations of nitrile- or azido-modified amino acids have central frequencies in the range of 2100–2300 cm⁻¹, a spectral region free of native protein vibrations but with significant H₂O or D₂O solvent response.26 Among the various IR-active amino acids, azidohomoalanine (Aha) may be the most versatile because it can be incorporated into a protein at essentially any position using a Met auxotrophic expression strategy without the need for any additional chemical modification after protein expression.27 As a relatively strong absorber with an extinction coefficient of ∼300 M⁻¹ cm⁻¹,28 the Aha signal is nevertheless still miniscule compared to the solvent response due to the inability to concentrate most proteins beyond ∼1 mM. In contrast, the water extinction coefficient in this spectral region is some 2 orders of magnitude smaller but at a concentration of ∼56 M. The reliable extraction of information pertaining strictly to the protein is thus a significant challenge, particularly with linear infrared spectroscopy.28 2D IR spectroscopy,28 a vibrational analogue of 2D NMR spectroscopy, offers a compelling solution to the signal amplitude mismatch described above. As a nonlinear spectroscopy, 2D IR signals scale quadratically with the extinction coefficient, in contrast with the linear nature of FTIR. This acts to significantly enhance the infrared response of the unnatural amino acid vibration with respect to the broad H₂O or D₂O solvent response in a 2D spectrum. Nevertheless, background subtraction of the buffer response is still critical to reliably isolate the Aha signal, and a host of experimental and data processing techniques have been employed in the present work in concert to make that possible (see the Supporting Information for additional details). First, a syringe pump sample delivery system together with a solvent response in a 2D spectrum. Nevertheless, background subtraction of the buffer response is still critical to reliably isolate the Aha signal, and a host of experimental and data processing techniques have been employed in the present work in concert to make that possible (see the Supporting Information for additional details). First, a syringe pump sample delivery system together with a flow cell was used to exchange sample and buffer solutions multiple times with minimal perturbation to the interferometer. Second, we used
the boxcars geometry, which as a background-free method offers the highest possible sensitivity,\textsuperscript{29} in comparison to the experimentally easier pump–probe geometry.\textsuperscript{25,30} Third, fast mechanical scanning of the coherence time delays, taking only \(\sim 0.1\) s per single scan, effectively reduced the laser noise, which was not \(\delta\)-correlated but had a typical correlation time on the order of 1 s.\textsuperscript{29} Fast scanning was implemented through voice coil scanning stages, allowing for the acquisition of data at the Nyquist frequency of the mid-IR field between each successive laser shot (see Figure S1). Fourth, the deleterious effects of laser scattering were minimized through the use of quasi-phase-matched laser shot (see Figure S1). Fourth, the deleterious effects of laser scattering were minimized through the use of quasi-phase-matched laser shot (see Figure S1). Fourth, the deleterious effects of laser scattering were minimized through the use of quasi-phase-matched laser shot (see Figure S1). Fourth, the deleterious effects of laser scattering were minimized through the use of quasi-phase-matched laser shot (see Figure S1). Fourth, the deleterious effects of laser scattering were minimized through the use of quasi-phase-matched laser shot (see Figure S1).

To that end, we study the equilibrium binding of two Aha-mutated peptides (i.e., Val\((0)\)Aha and Val\((−3)\)Aha), derived from the Ras-associating guanine nucleotide exchange factor 2 (RA-GEF), to a PDZ2 domain from the protein tyrosine phosphatase 1E.\textsuperscript{34,35} PDZ2 domains consist of six \(\beta\)-strands and two \(\alpha\)-helices, structured such that a binding groove spans strand \(\beta 2\) and helix \(\alpha 2\) (see Figure 1), and biomolecular recognition occurs primarily through interactions with the C-terminus of target proteins\textsuperscript{36–38} in a 1:1 stoichiometry.\textsuperscript{39} Assuming that the structure of the bound ligand remains the same upon mutation, the Aha side chain is expected to be buried in the hydrophobic recognition pocket of the binding groove in the case of Val\((0)\)Aha but points outward in Val\((−3)\)Aha (see Figure 1). Previous 2D IR experiments support that structural picture.\textsuperscript{25}

To determine the binding affinity directly from the infrared response of the Aha-mutated ligands, a series of 2D IR experiments were performed, fixing the initial free ligand concentration (and thus also the Aha concentration) at 500 \(\mu\)M and varying the PDZ2 concentration (see Figure 2a,c). 2D IR spectra of uncoupled vibrational modes consist of two peaks of opposite sign, related to the transition between the ground and first excited states of the vibrator (0–1 transition, shown in blue in Figure 2a,c) and the transition between the first and second excited states (1–2 transition, in red).\textsuperscript{28} Here, we focus on the 0–1 transition, which is found along the diagonal of the 2D spectrum, while the 1–2 transition is off-diagonal and for the most part outside of the spectral window shown in Figure 2a,c. The 0–1 signal along the diagonal, which has been extracted in Figure 2b,d, is free from overlapping contributions of the 1–2 band, allowing for an unambiguous measure of the population of the corresponding ensemble.

The results for Val\((−3)\)Aha are shown in Figure 2a,b. With no PDZ2 domain present (leftmost panel), we observe a broad peak at \(\sim 2112\) cm\(^{-1}\) due to the Aha in the unbound peptide. With increasing molar ratio of the PDZ2 domain, the band initially broadens and gradual shifts toward \(\sim 2107\) cm\(^{-1}\) due to binding of the ligand to the protein. Binding is essentially saturated already at a molar ratio of 1.4, with minimal deviations up to a molar ratio of 5. For a global fitting of the data, we first determined the central frequencies and widths of two Gaussian distributions of unbound and fully bound ligands, respectively, from the data at molar ratios of 0 (Figure 2b, leftmost panel) and 5 (Figure 2b, rightmost panel) and subsequently extracted the amplitudes of these two ensembles as a function of protein concentration. The results of that analysis are shown in Figure 3a, where the populations are fit to a two-state binding model

\begin{equation}
P + L \leftrightarrow PL
\end{equation}

revealing a dissociation constant of \(K_D = 22 \pm 10\) \(\mu\)M, which is in excellent agreement with that obtained from ITC (\(K_D = 33 \pm 5\) \(\mu\)M; see Figure S3 in Supporting Information).

The results for Val\((0)\)Aha are richer in content (Figure 2c,d). With no PDZ2 domain present, the spectrum is essentially the same as that for Val\((−3)\)Aha with a center frequency of 2112.5 cm\(^{-1}\). With increasing molar ratio of the PDZ2 domain, an additional band is observed, initially appearing as elongation of the band to lower frequencies (molar ratio 0.4) and then evolving into a second distinct red-shifted band at \(\sim 2097\) cm\(^{-1}\) at molar ratios of 1.6 and beyond. However, the original band does not disappear even at the highest protein concentration, as one might expect; instead, a careful inspection of the 2D IR spectra reveals a small frequency red shift of that peak of about 2.5 cm\(^{-1}\) (see vertical lines in Figure 2c,d). It has been shown that the Aha frequency reflects the degree of solvation;\textsuperscript{39} hence, we can conclude that the three peaks represent the following conformations: at 2112.5 cm\(^{-1}\), the unbound and random coiled peptide, where the broad line shape is indicative of a heterogeneous local solvent environment; at 2110 cm\(^{-1}\), a loosely bound peptide with a solvated side chain such that the azido group experiences a similarly heterogeneous local environment as in the unbound case; and at 2097 cm\(^{-1}\), a more tightly bound peptide with the side chain buried in the hydrophobic recognition pocket, where the lack of solvent exposure leads to a more homogeneous local environment and thus a narrow line shape.

Recent molecular dynamics simulations of the binding of the wild-type peptide to the PDZ2 domain\textsuperscript{39} suggest that the loosely bound peptide is in fact an encounter complex, in which the ligand is not yet bound inside of the binding groove. Rather, stabilized by non-native salt bridges formed with residues from \(\alpha 2\) and \(\beta 2\), the ligand spans across the binding groove (see Figure 6 of ref 38). In such an encounter complex, the Aha label remains solvated to almost the same extent as that in the
unbound peptide, leading to the only minor frequency shift of its vibration. This assignment suggests a model

\[ P + L \rightleftharpoons PL_1 \rightleftharpoons PL_2 \]  \hspace{1cm} (2)

with a dissociation constant of the encounter complex \( K_{D1} \equiv [P][L]/[PL_1] \) and an equilibrium constant between the encounter complex and tightly bound peptide \( K_{12} \equiv [PL_2]/[PL_1] \). A fit as for Val(−3)Aha turned out to be under-determined in this case because of the three overlapping distributions with a very small frequency shift of unbound peptide vs. encounter complex and the fact that binding is not completely saturated even at a molar ratio of 5.9. Rather, we included the model in eq 2 in a single global fitting step of the complete data set, thereby fixing the relative amplitudes of the three contributions as a function of protein concentration to the model (see the Supporting Information for details). This procedure rendered the fit stable and revealed \( K_{D1} = 470 \pm 100 \mu M \) and \( K_{12} = 0.8 \pm 0.2 \) (Figures 2d and 3b).

To connect these results to ITC, which cannot distinguish the two binding modes and thus measures the combined effect of both, we note that the model in eq 2 is mathematically identical to the model in eq 1 when replacing \([P][L]/[PL_1]\) and \([PL_2]/[PL_1]\) with an overall dissociation constant \( K_{D1} \equiv [P][L]/[PL_1] + [PL_2]/[PL_1] \) and defining an overall dissociation constant

\[ K_{D1}^{\text{all}} \equiv \frac{[P][L]}{[PL_1] + [PL_2]} = \frac{K_{D1}}{1 + K_{12}} \]  \hspace{1cm} (3)

With that, we obtain \( K_{D1}^{\text{all}} = 250 \pm 50 \mu M \), again in excellent agreement with the ITC result of 250 ± 60 \( \mu M \) (see Figure S3b in the Supporting Information).

It is the C-terminal carboxyl group together with the hydrophobic interactions of the Val(−0) side chain in the
Consequently, mutation of Val(−0) with the longer and weakly polar Aha side chain, which is still buried in the hydrophobic recognition pocket according to the 2D IR results, destabilizes the tightly bound state, as evidenced by the rather large $K_D$ ($250 \pm 50 \mu M$). This, in turn, increases the relative population of the encounter complex so that it becomes experimentally observable. Val(−3)Aha, on the other hand, has a much larger binding affinity ($K_D = 22 \pm 10 \mu M$), similar to that of the wild-type peptide. On the basis of the high binding affinity, we may assume that the structure in the binding groove is the same as the wild-type (Figure 1); hence, the Aha label points outward and thus is solvated to a significant extent even if the ligand is tightly bound. We would therefore not be able to spectroscopically distinguish the tightly bound peptide from an encounter complex. However, when assuming that the dissociation constant $K_{D1}$ of the encounter complex is about the same in both mutants (because it binds less specifically), we can still conclude from eq 3 that the encounter complex is hardly populated in Val(−3)Aha.

The spectroscopic approach proposed here is similar to NMR28–30 with a “chemical shift” that reports on the environment of a spectroscopic probe. In the affinity regime of the system studied here, NMR would be in the limit of coalescing lines,40 in which case only one band is observed, whose frequency position reflects the average over all states. While methods based on relaxation dispersion have been developed to recover exchange rates and chemical shift of those “invisible states”,40,41 2D IR is inherently an ultrafast spectroscopy, and coalescence will not occur for any process slower than ~1 ps. This is why, in the present study, the three states of Val(−0)Aha indeed reveal three bands. Both spectroscopic techniques, NMR and 2D IR, are thus complementary.

While the technical challenges of measuring weak infrared signals have to now limited the application of IR methods in biomolecular recognition, these measurements pave the way for greater adoption of unnatural amino acids as infrared probes for protein–protein and protein–ligand interactions. Different vibrational labels that respond more strongly to the environment, such as ester groups in non-natural amino acid side chains,32 might be able to resolve even more binding modes. Furthermore, with recent advances in high-repetition-rate amplified laser systems, multidimensional IR spectroscopy of equilibrium biomolecular recognition will be accessible within seconds,32,43 allowing for collection of a complete binding interaction curve in a fraction of the time needed for thermodynamic measures of binding while simultaneously yielding greater structural information reflecting possible conformational heterogeneity. These measurements also serve as a proof of principle for the application of unnatural amino acids to nonequilibrium experiments, in which an actinic pulse initiates a structural process such as ligand unbinding, which is then followed by transient 2D IR spectroscopy.31 Together with similar sensitivity requirements as in the present case, the observation of such nonequilibrium processes will be informative for the critical feature distinct from established methods used to determine biomolecular recognition.

**ASSOCIATED CONTENT**

Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcllett.7b00742.

Details on the protein expression and peptide synthesis, the 2D IR experimental setup including a demonstration of the advantage of fast scanning and the intrinsic phasing, the fitting procedure, as well as the ITC measurements (PDF)

**AUTHOR INFORMATION**

ORCID

Philip J. M. Johnson: 0000-0002-7251-4815
Peter Hamm: 0000-0003-1106-6032

Author Contributions


Notes

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We thank Rolf Pfister for help with peptide synthesis, Marjorie Sonnay for help with peptide purification, Ilian Jelesarov for very valuable discussions regarding the ITC measurements, Steven Waldauer for the ITC data of Val(−3)Aha, Robbert Bloem and Brittige Stucki-Buchli for contributions at an early stage of this project, and Oliver Zerbe, Amedeo Calfaisch, and Ben Schuler for helpful discussions on various other aspects of the work. The work has been supported in part by a European Research Council (ERC) Advanced Investigator Grant (DYNALLO) and by the Swiss National Science Foundation (SNF) through the NCCR MUST as well as Grant 200021_165789/1.

**REFERENCES**


Conformationally Sensitive IR Probe of Protein Folding, Protein Dyer, R. B.; Carrico, I.; Raleigh, D. P. Azidohomoalanine: A 7475. Specific Infrared Probes of Proteins. IUCrJ 2014, 1, 523−529.


(43) Greetham, G. M.; Donaldson, P. M.; Nation, C.; Szaraniovich, I. V.; Clark, I. P.; Shaw, D. J.; Parker, A. W.; Towrie, M. A 100 kHz Time-Resolved Multiple-Probe Femtosecond to Second Infrared Absorption Spectrometer. Appl. Spectrosc. 2016, 70, 645−653.