A Pipeline for Screening Small Molecule-Enhanced Protein Stability in A Bacterial Orphan Receptor

James J. Siclari[a,b], Denize C. Favaro[a], Richard H. Huang[c], Kevin H. Gardner[a,d,e]

[a]: Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY 10031

[b]: Ph.D. Program in Biology, The Graduate Center – CUNY, New York, NY 10016

[c]: RedShift BioAnalytics, Inc., Boxborough, Massachusetts 01719

[d]: Department of Chemistry and Biochemistry, City College of New York, New York, NY 10031

[e]: Ph.D. Programs in Biochemistry, Biology, and Chemistry, The Graduate Center – CUNY, New York, NY 10016

ORCIDs: James J. Siclari: 0000-0001-5275-4515 / Denize C. Favaro: 0000-0001-5563-2548 / Richard H. Huang: 0000-0002-4725-4564 / Kevin H. Gardner: 0000-0002-8671-2556

Keywords: ligand binding – bacterial one-component signaling – thermal shift assay – microfluidic modulation spectroscopy – NMR spectroscopy

Running title: Efficient pipeline for ligand binding to bacterial one-component proteins

Abstract

Bacterial one-component signaling proteins integrate sensory and gene regulation functions within the same polypeptide, creating powerful natural sensors of environmental conditions which can also be adapted into powerful tools for synthetic biology and biotechnology. A key sensor motif within many of these proteins is the Per-ARNT-Sim (PAS) domain, known for its conserved fold yet highly divergent sequences, allowing for a broad range of ligands to control PAS protein function by changes in small molecule binding occupancy or configuration. This diversity introduces a challenging step – identification of ligands for "orphan" PAS proteins which show signatures of ligand binding but no copurifying high-affinity bound small molecules – into characterization and engineering of such proteins. In this study, we characterized CU228, a putative PAS-HTH transcription factor from *Candidatus Solibacter usitatus,* as a novel model system for searching for novel ligands by small molecule stabilization. Bioinformatics and structural analyses predicted a PAS domain with an Trp-rich internal cavity, suggesting potential small molecule interactions. Using a ~760 compound fragment library, differential scanning fluorimetry identified three ligands (KG-96, KG-408, and KG-484) that substantially increased CU228 thermal stability with ΔT_m values up to +10°C. Microfluidic modulation spectroscopy (MMS) revealed ligand-induced preservation of α-helical and β-sheet integrity under thermal stress. Saturation transfer difference NMR confirmed direct binding of all three ligands and enabled estimation of micromolar-range dissociation constants, consistent with expected fragment-level affinity. Our findings expand the analytical toolbox for probing protein-ligand interactions in flexible, signal-responsive systems, laying the groundwork for designing synthetic chemogenetic variants of one-component transcription factors.

Introduction

Responding to an ever-changing external environment is crucial for bacterial survival and proliferation. Many such adaptive responses are based on one component signaling proteins (Ulrich et al. 2005), which couple a sensor domain and a DNA-binding domain within the same polypeptide chain, switching gene expression on and off in response to changes in ligands from their surroundings. Exploiting the natural functions of bacterial proteins has driven significant advancements in biotechnology; indeed, natural one component signaling proteins have laid the foundations for many biosensors and, inducible expression systems, often through slight modifications to existing functionalities (Lazar and Tabor 2021).

One highly used sensor domain in many of these proteins is the Per-ARNT-Sim (PAS) domain, which is characterized by a highly conserved ~120 amino acid residue mixed □/β fold and often containing an internal hydrophobic pocket evolved for ligand binding. While the structural fold is highly conserved, sequence identity across PAS domains is less than 20%, enabling them to sense a wide range of stimuli (Taylor and Zhulin 1999; Henry and Crosson 2011; Stuffle et al. 2021).

An example of this from our lab is the blue-light-sensitive PAS transcription factor EL222, which contains an N-terminal light-oxygen-voltage LOV domain (a subfamily of PAS domains, evolved to bind flavin chromophores) and a C-terminal HTH domain. EL222 undergoes a conformational change upon blue light exposure, transitioning from a monomer to a homodimer and enabling DNA binding (Nash et al. 2011; Rivera-Cancel et al. 2012; Zoltowski et al. 2013; Chaudhari et al. 2025). Previous work has engineered this system for multiple uses in eukaryotic model systems (Motta-Mena et al. 2014;

Reade et al. 2016; Rullan et al. 2018; Zhao et al. 2018; Zhao et al. 2020; Cleere and Gardner 2024; Hoffman et al. 2025). LOV domains are easily identifiable by their primary sequences, including a conserved “NCRFL” motif that is used to bind flavin chromophores like FMN, FAD, and riboflavin (Henry and Crosson 2011). While many other PAS domain subfamilies exist in one-component systems and likely respond to a wide variety of ligands, the sequence conservation and ligand-binding specificities of these domains is less well understood.

Identifying and engineering LOV domains has the added advantage that LOV-binding cofactors are naturally available in organisms typically used for heterologous expression and bind with sufficient affinity for complexes to survive *in vitro* purification and characterization. Because this is not always the case for all ligands, we have built an in-house library of approximately 760 small-molecule fragments that has been useful for $^{15}N/^{1}H$ HSQC-based NMR-based screening of several PAS proteins (Amezcua et al. 2002; Scheuermann et al. 2009; Guo et al. 2013; Xu et al. 2021) and non-PAS proteins (Best et al. 2004). While this strategy has been successful for identifying useful compounds – some of which led to the successful development of a FDA-approved PAS domain inhibitor targeted cancer therapeutic (Cho et al. 2016; Jonasch et al. 2021), such protein-detected NMR screening approaches can be laborious due to high sample consumption.

In this study, we aimed to use this library to identify small molecule ligands which could bind and potentially activate a novel PAS-HTH protein as a new chemogenetically-controlled transcription factor. Unfortunately, the PAS-HTH protein we selected here, CU228, was not well behaved for protein-directed NMR screening,

requiring us to develop a novel multifaceted screening pipeline incorporating differential scanning fluorimetry (DSF), ligand-directed nuclear magnetic resonance (NMR), and microfluidic modulation spectroscopy (MMS). From this approach, we identified compounds that clearly bound CU228 with accompanying changes in protein stability suggestive of conformational changes within this novel PAS-HTH protein. Our results both seed the potential development of novel CU228-binding ligands and lay out a general strategy for rapidly conducting both binding and stabilization assays to identify new compounds capable of binding moderately behaved targets.

Results

Identification of a Protein Test Case

To identify a suitable test case for small molecule stabilization, we searched the SMART database (Letunic et al. 2021) for predicted one-component prokaryotic proteins containing both PAS and HTH domains. From this list of ~2600 sequences, we subsequently used a bioinformatics pipeline to select sequences with a single PAS domain (lacking the canonical NCRFL flavin-binding motif) and flanked to either side by a single HTH domain (**Fig. 1a**) (Edgar 2004; Cai et al. 2004; Robert and Gouet 2014; Zimmermann et al. 2018; Gabler et al. 2020; Gumerov and Zhulin 2020; Letunic et al. 2021). Out of the fifty sequences remaining at the end of this pipeline, we focused on a 228-residue protein from *Candidatus Solibacter usitatus* (Challacombe et al. 2011) (CU228). Conspicuously, AlphaFold3 modeling (Abramson et al. 2024) predicted that three tryptophan residues within the PAS domain would arrange to form a cage-like

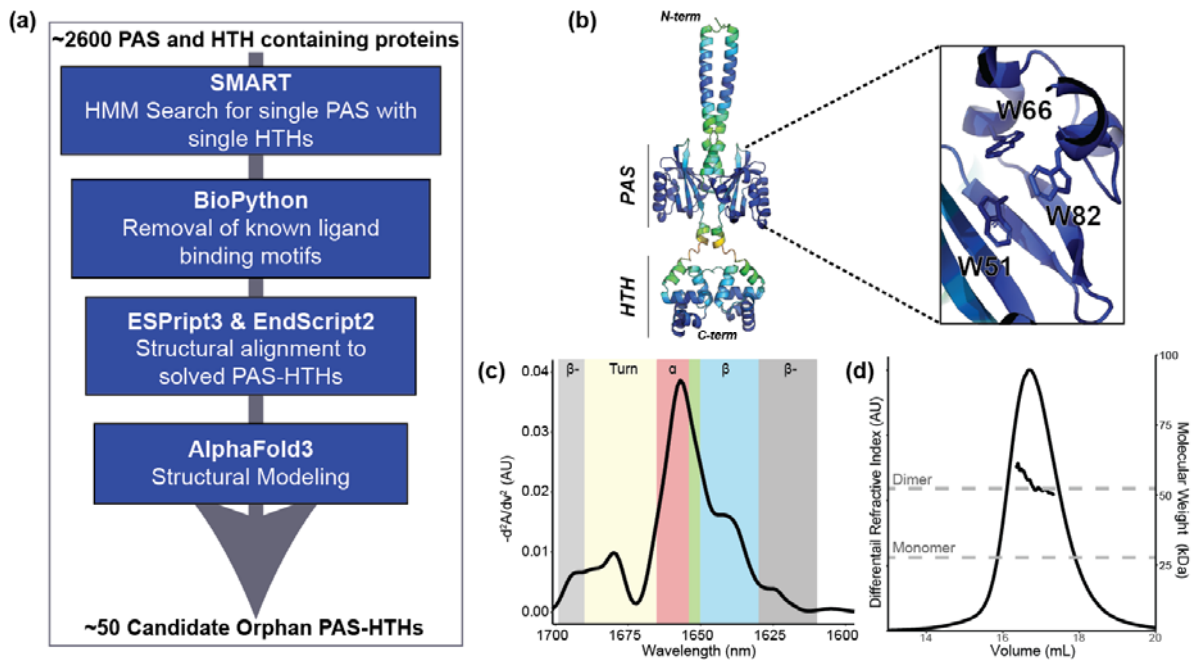


Figure 1. Initial characterization of test case protein for this study, CU228. (a) Schematic of bioinformatics pipeline for combinatorial domain hunting searching for a PAS-HTH with no FMN binding preferences. (b) AlpfaFold3 model colored by pLDDT score (blue = high confidence to red = low confidence) zoom window showing three internally facing tryptophan side chains in the PAS domain cavity. (c) MMS spectra of purified CU228 at 25°C, colors denote secondary structure elements, can be seen in Table S1. β-: Intermolecular β-sheet, α: α-helix, β: β-sheet, green box: unordered structure.

structure that could bind an internal ligand (**Fig. 1b**). We assessed the structural integrity, secondary structure composition, and oligomeric state of CU228 using microfluidic modulation spectroscopy (MMS, an IR-based spectroscopy focused on the amide-I region) (Liu et al. 2020) and size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). Taken together, these studies revealed that the protein was a stable dimer and contained a mixed α/β fold with a particularly strong helical content, consistent with PAS-HTH architecture (**Fig. 1c,d**). Complementary $^{15}N/^{1}H$ TROSY NMR experiments (**Fig. S2**) displayed poor amide ^{1}H chemical shift dispersion and heterogeneous peak intensities, suggesting heterogenous tertiary

structure. Taken together, our data indicated CU228 is a structured yet dynamic and flexible protein, making it an attractive candidate for stabilization through small-molecule binding.

Screening for Small Molecule Stabilizers

We conducted a thermal shift assay using DSF to identify ligands that enhanced CU228 stability (**Fig. 2**). To capture medium-affinity binders, we screened our 760 compound fragment library (Amezcua et al. 2002) with 2.5 µM CU228 and 250 µM of each small molecule, each with 1.25% DMSO and performed in duplicate. CU228 alone exhibited a melting temperature (T_m) of 53.4±0.07°C, which was minimally impacted by the addition of most compounds in the library. However, approximately 50 compounds stabilized CU228 one standard deviation above the control (T_m > 55.7°C) (**Fig. 2a**). After eliminating samples with poor reproducibility, biphasic melting curves, and solubility issues, we focused on three ligands for further characterization: KG-96 (T_m = 57.2±2.8°C), KG-408 (T_m = 57.3±0.3°C), and KG-484 (T_m = 63.4±2.7°C). Notably, all three hit compounds included nitrogen-containing heterocycles and small aliphatic or aromatic side chains. (**Fig. 2b**).

Validation of Ligand Binding and Affinity Determination via NMR

Given the low quality of the CU228 $^{15}N/^{1}H$-TROSY spectrum, we confirmed direct binding of the three ligands using saturation transfer difference (STD) NMR. This ligand-observed NMR method requires a low concentration of protein (10-50 µM), with spectra

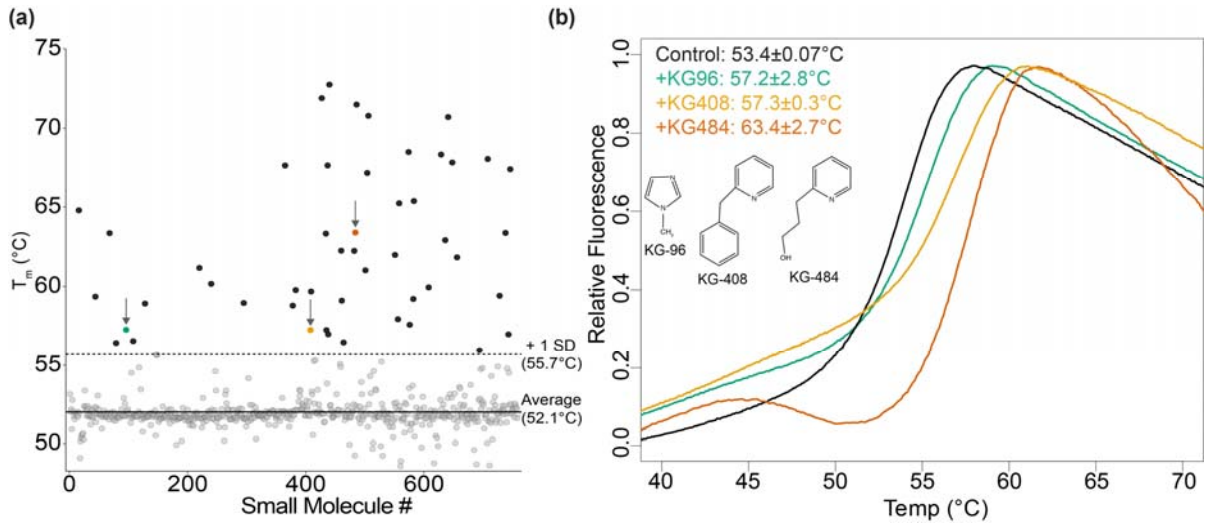


Figure 2. Initial CU228 small molecule fragment screen by thermal shift assay. (a) Average T_m (n=2) for all compounds in the screen calculated by thermal shift. (b) Raw fluorescence traces of Protein Thermal Shift™ Dye associating and dissociating to CU228 as a function of temperature in the control sample (T_m = 53.4±0.07°C) and three ligands: green: KG-96 (T_m = 57.2±2.8°C), yellow: KG-408 (T_m = 57.3±0.3°C), and red: KG-484 (T_m = 63.4±2.7°C). Insert shows chemical structure for respective compounds.

recorded in the presence of increasing ligand (0.02-5 mM). We acquired STD NMR data with an interleaved acquisition of two spectra, one with protein signals saturated (ON) and one without (OFF). Differences in intensity between these ON and OFF spectra reveal a transfer of saturation via intermolecular Nuclear Overhauser Effect (NOE) (Angulo et al. 2010) from a protein/ligand complex, indicating direct binding. We observed clear ligand peaks in the STD difference spectra, indicating saturation transfers from CU228 to KG-96, KG-408, and KG-484 consistent with direct interactions in the fast exchange regime (**Fig. 3a**).

To determine binding affinities, STD experiments were performed at various ligand concentrations and saturation times. We calculated STD amplification factors (AFs) for three spectral regions for each ligand (**Fig. 3b**; **Eq. 1-2**). We then generated

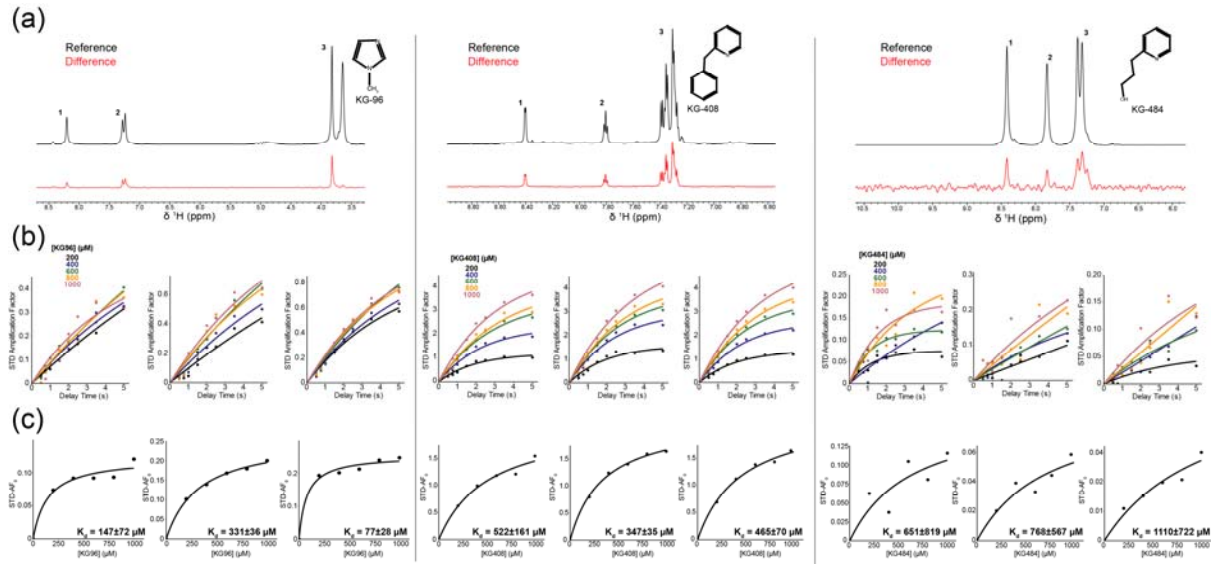


Figure 3. STD NMR of CU228 and KG-96, KG-408, or KG-484. (a) ^{1}H spectra of reference (black) and difference (red) for STD experiments (t_{sat} = 2 s) between CU228 and the respective compound at 1 mM, and 30 µM CU228. The insert shows the chemical structure of each. (b) STD buildup curves for each spectral region labeled in panel a utilizing varying saturation times. (c) STD binding curves for each spectral region labeled in panel a. Weighted average K_d from three epitopes for KG-96, KG-408, or KG-484 were 171±21 µM, 376±31 µM, and 841±392 µM respectively.

binding curves by plotting amplification factor growth rates (STF-AF_0) versus ligand concentration and fitting each to a Langmuir isotherm (**Eq. 3**), yielding average dissociation constants (K_d) for KG-96, KG-408, and KG-484 of 171±21 µM, 376±31 µM, and 841±392 µM respectively (**Fig. 3c**) (Mayer and Meyer 2001; Angulo et al. 2010).

<u>Assessing Secondary Structure Changes via MMS</u>

While the thermal shift assay works well to probe overall tertiary structure, it does so by testing the accessibility of a dye to the protein hydrophobic core, raising potential concerns in the case of suspected small molecule binders such as CU228 (Wu et al. 2023). To further assess the structural impacts of ligand binding on CU228 without any

bias or competition from a dye molecule, we employed MMS to monitor secondary structure changes of CU228 in the presence of KG-96, KG-408, and KG-484.

At ambient temperature, each ligand had minimal effects on the CU228 secondary structure, aside from KG-408's ability to increase both native and intermolecular β-sheet content (β-) (**Fig. S3**). Upon exposure to thermal stress, however, all three ligands stabilized β-sheet and α-helix at elevated temperatures, relative to the control as indicated by the increased T_m values (**Fig. 4, Table S2**). Additionally, ligand-treated samples exhibited delayed formation of β-aggregates during heating (**Fig. 4**), suggesting ligand-protein interactions might stabilize dimer formation involving predicted PAS domain β-sheets in the CU228 AlphaFold3 model. The degree

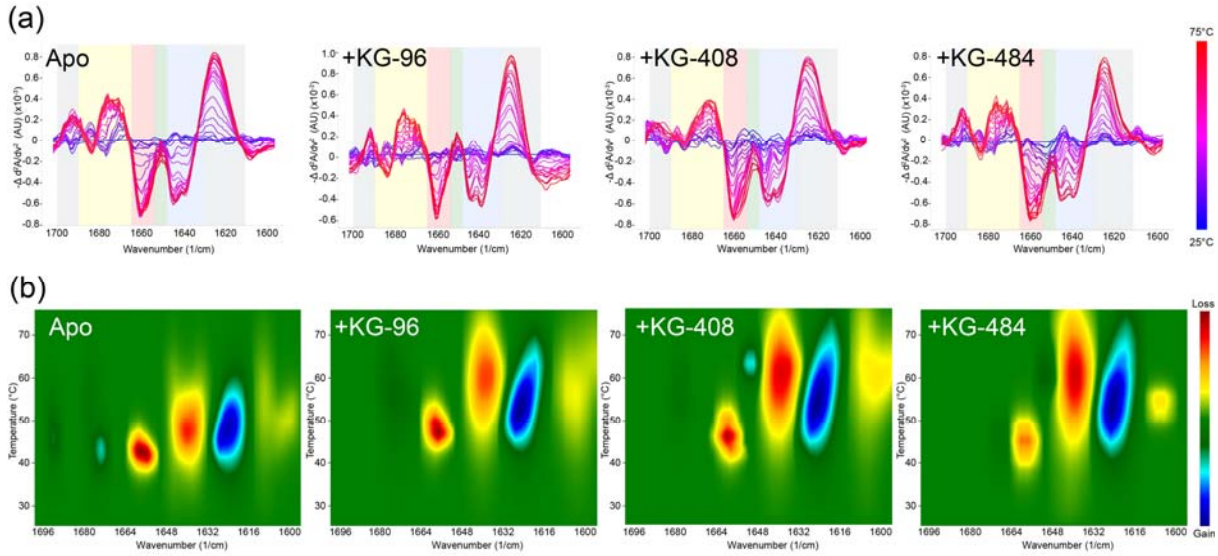


Figure 4. Thermal ramp MMS of CU228 and KG-96, KG-408, and KG-484 at 1 mM. (a) Difference spectra of respective samples along the temperature gradient at 2.5°C intervals with 1 mM ligand and 20 µM CU228 displaying changes in aggregated β-sheet (grey), turn (yellow), α-helices (red), unorder (green), and native β-sheet (blue). (b) Heat maps comparing temperature to rate of change along wavenumbers. Blue denotes increase in spectral intensity or gain in respective structure, red denotes decrease in spectral intensity or loss in respective structure.

of thermal stability was concentration dependent, with ranges comparable to K_d measurements acquired via NMR (**Fig. 5**). Maximal fractional contribution for α-helical and β-sheet content arises at temperatures around 35-40°C, which was not seen in the apo protein (**Fig. 5**), suggesting that full folding of CU228 requires the presence of a small molecule ligand.

As we increased ligand concentration in these experiments, the melting temperatures of all secondary structures concomitantly increased. However, while α-helical T_m values were moderately affected (4-5°C) by ligand binding, β-sheet resistance to thermal denaturation was increased more substantially, up to 14°C (**Table S2**). We interpret this observation as supporting evidence of the binding of these small compounds within the mixed α/β fold of the CU228 PAS domains, as the only predicted

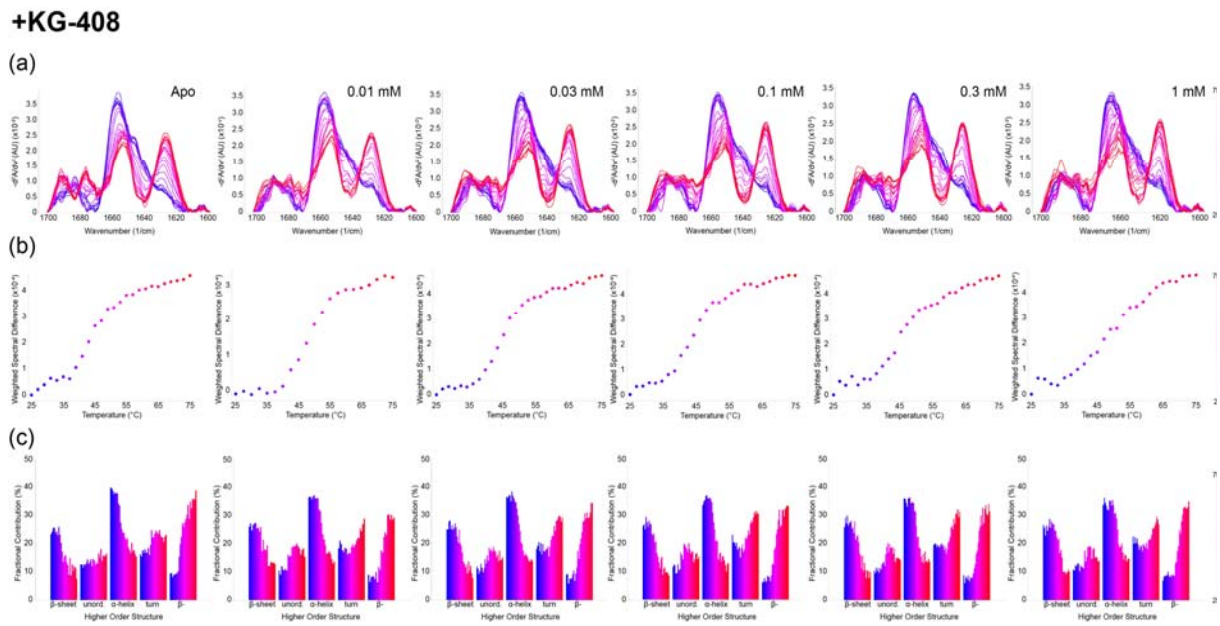


Figure 5. Concentration dependent thermal ramp MMS of CU228 with KG-408. (a) MMS similarity spectra across from 25-75°C at five different concentrations of KG-408. (b) Weighted spectral difference (compared to 25°C) across the thermal ramp range at each concentration (Kendrick et al. 2020). (c) Higher order structure fractional contribution as a function of temperature for each titration point.

beta-sheet containing element in the protein. In contrast, the interdomain linker and DNA binding domain of the protein are all predicted to be helical and less likely to be directly stabilized by ligand binding.

We also noticed changes to the 1600 nm region of the IR spectra during thermal stress, which is typically attributed to changes in amino acid side chains, particularly Tyr (Kong and Yu 2007; De Meutter and Goormaghtigh 2021). This was mainly observed with CU228 binding to KG-96 (**Fig. 4**), which coincidentally is the smallest of the three ligands tested, perhaps further implicating interior binding near the two tyrosine residues located within the PAS core.

Discussion

This study presents a multifaceted workflow for identifying and characterizing ligand interactions with a previously unstudied orphan PAS-HTH transcription factor. A structure-guided bioinformatic approach identified CU228 as a promising candidate, with a predicted Trp-rich pocket likely suitable for small-molecule targeting. Through thermal shift assays, multiple fragments were found to increase protein thermal stability. MMS provided insights into the preservation of secondary structure under thermal stress, and NMR confirmed direct ligand binding while quantitating high-micromolar binding affinities.

Historically, protein receptors have been deorphanized *in vitro* through mass spectrometry or atomic structure determination to identify the natural ligands (Laschet et al. 2018). This approach requires the ligand to be present in the expression system and

be of high enough affinity to remain bound through all sample preparation steps. Screening of artificial ligands has been beneficial in guiding natural ligand discovery as well as developing triggers for protein engineering but can require a large amount of sample when run in certain formats. For example, our prior $^{15}N/^{1}H$ HSQC NMR-based screening of our 760 compound library of PAS domains (Amezcua et al. 2002; Scheuermann et al. 2009; Guo et al. 2013) typically entailed the use of 150-200 mg of uniformly ^{15}N-labeled protein. While this protein-detected NMR approach is extremely data-rich, providing very early binding site information and affinity, the sample requirements limit its application to many systems. In contrast, the pipeline described here requires a low quantity of protein – approximately 4 mg for all of the experimental work describe herein – and can be scaled up to perform high-throughput screens.

The relatively low affinities we observed for each ligand is consistent with typical expectations for fragment-based screening, where focused collections of small compounds (MW ~ 200 Da) are used to quickly screen a fraction of chemical space to lay the foundation for further optimizations (Barelier and Krimm 2011; Kirkman et al. 2024). More specifically, the mid- to high micromolar affinities seen for KG-96, KG-408, and KG-484 are also consistent with our experience with this small fragment library against other targets with affinities between 100-500 µM (Amezcua et al. 2002; Best et al. 2004; Guo et al. 2013), some of which were rapidly improved to low micromolar or high nanomolar potency (Scheuermann et al. 2009; Key et al. 2009). In addition to getting these initial leads, we also obtained some early structure-activity relationship (SAR) information that could assist future development of higher affinity derivatives from

these fragment scaffolds. The potential for this can already be seen as the difference of a phenyl ring between KG-408 and KG-484 results in the doubling of K_d value.

In addition to developing a deorphanization strategy, our findings offer new insights into the confirmational behavior of CU228 that can be extended to other one-component signaling proteins. The observed increase in β-sheet melting temperature upon ligand binding and prolonged intermolecular β-sheet interactions are consistent with the predicted regions for stimulus recognition and homodimerization interfaces within the PAS domain suggest stabilization of the dimer complex though the PAS domain β-sheets. Seeing the greatest structural resilience in the β-sheet content of CU228 further supports that each ligand directly interacts with the PAS domain, while the remaining helical content is less perturbed (**Table S2**).

Accordingly, literature suggests the 1650-1652 cm^{-1} region of the IR spectra is indicative of coiled-coil formation (Heimburg et al. 1999; Dong et al. 2008). Absorption in this region became more prominent in our results when the CU228 was heated above its melting temperature (**Fig. 5**). In the apo state, this prominent peak offers more support for the AlphaFold3 model of CU228 (**Fig. 1**) where the N-terminal α-helices are one of the predicted dimerization interfaces**.** Upon CU228 binding to KG-96, this coiled-coil region (1650-1652 cm^{-1}) retains structure despite thermal stress, while the non-coiled coil α-helical structures (1656 cm^{-1}) found in the PAS, linker, and HTH segments melt (**Fig. S4**).

Overall, our integrative approach demonstrates the utility of combining computation modeling, biophysical assays, and NMR to study flexible, dynamic proteins and their interactions with small molecules, particularly when no prior knowledge of

endogenous ligand exists. Strengths of this approach are its speed, low sample consumption, and mix of label-free and label-containing approaches to quickly get a slate of fragment and protein information for subsequent higher-content assays and optimization, enabling larger classes of protein targets to be surveyed.

Materials and Methods

Cloning, Protein Expression and Purification

DNA encoding the CU228 sequence (GenBank: HEY1493056.1) was ordered from Twist Biosciences and cloned into a pHisGβ1-parallel expression vector (Harper et al. 2003). CU228 was recombinantly expressed with a His_6-Gβ1 tag in BL21(DE3) *E. coli* (Stratagene) in LB media at 16°C overnight. Cells were pelleted and resuspended in Buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl). The cells were lysed by sonification, and the lysate was cleared by centrifuging at 10,000 x g for 30 minutes at 4°C. The supernatant was collected and filtered through a 0.2 µm syringe filter.

For affinity purification, the protein was loaded onto a 5 mL HisTrap (Cytiva) column preequilibrated with Buffer A. The column was washed with 10 column volumes (CVs) of Buffer A before eluting the protein with buffer B (50mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM imidazole). The eluent was diluted 1:10 with Buffer C (50 mM Tris-HCl, pH 8.0) and the His_6-Gβ1 tag was cleaved by adding 1 mg His_6-TEV protease (Kapust et al. 2001; Phan et al. 2002) per 30 mg of fusion protein. After overnight cleavage, the sample was again applied to a preequilibrated 5 mL HisTrap column to remove tag and TEV protease. The flow through was then concentrated using an 30

kDa MWCO Amicon Ultra concentrator to a volume of 3 mL. The concentrated flowthrough was applied to a Superdex 200 10/300 (Cytiva) column preequilibrated with assay buffer (10 mM sodium phosphate pH 7.0, 50 mM NaCl, 0.5 mM DTT).

Thermal Shift Assay

Sourcing for compounds in the small molecule library is outlined in Guo et al., 2013 (Guo et al. 2013). Thermal shift assays were performed using Applied Biosystems Protein Thermal Shift™ Dye kit as per the manufacturer's instructions. CU228 was used at a concentration of 2.5 µM. All small molecule fragments were diluted to 250 µM in 1.25% DMSO. Fluorescence was recorded using a QuantStudio 7 Flex Real-time qPCR machine. All data were processed and analyzed using TSA-CRAFT (Lee et al. 2019).

Microfluidic Modulation Spectroscopy

For initial characterization of CU228 and folding confirmation, MMS measurements were made on an Aurora instrument (RedShift BioAnalytics, Inc.) at 1 mg/mL protein concentration. The samples were analyzed against the matching buffer (50 mM sodium phosphate pH 7.0, 50 mM NaCl, 0.5 mM DTT) at a backing pressure between 7-12 psi with an average flow rate of 1.4 µL/s. The differential absorbance spectra were collected across the Amide-I band from 1588-1712 cm^{-1} with a resolution of 1 cm^{-1}.

For thermal ramping measurements, an Aurora TX (RedShift BioAnalytics, Inc.) was used to capture the differential spectral data of sample-buffer over temperature with 0.5 mg/mL CU228 concentration and 1 mM ligand concentration. Equivalent concentrations of ligands were also added to reference buffer for baseline subtraction. Spectra were captured from 1588-1712 cm^{-1} at 1 cm^{-1} resolution. The flow rate was calibrated to 1 µL/min at 25°C and each fluid exchange lasted 0.4 s to remove carryover effects in the MMS cell. Thermal ramping was carried out continuously at 1°C/min from 25 to 75°C. Raw spectra were captured approximately every 17 s along with cell temperature measured by a thermistor attached to the MMS flow cell. Dark scan, sample channel, and reference channel measurements were captured 3 s apart. Due to the continuous temperature ramp, dark scan and reference detector signals were phase-corrected to align them in time and temperature with sample measurements. The sample and reference channel signals were dark offset corrected, and water vapor was subtracted to yield differential absorbance.

The *delta* analytics software (Lucas et al. 2025) was used for sample analysis and spectral processing. The differential absorbance spectra were normalized to absolute spectra using a 0.63 nominal displacement factor using apo-CU228 as the model protein for the fitting. Absolute spectra were then converted to similarity spectra by first taking the second derivative, then inversed and baselined. Savitsky-Golay smoothing was applied to the second derivative plot using a window of 19 wavenumbers. HOS structural elements were calculated using the similarity spectra by Gaussian curve fitting. A list of Gaussians with designated wavenumbers and

corresponding structural types is shown in **Supplemental Table 1** (Kendrick et al. 2020; Liu et al. 2020; Ma et al. 2023; Wei et al. 2025).

Nuclear Magnetic Resonance

Solution nuclear magnetic resonance spectra were collected with a Bruker Avance III HD 700 MHz spectrometer with a 5 mm QCI-F CryoProbe at 298K using standard Bruker pulse sequences. Samples contained 20 µM CU228 and KG-408 concentrations of 200, 400, 600, 800, and 1000 µM. Saturation transfer difference spectra were collected using the pulse program “stddiffesgp.3” with varying saturation times of 0.5, 0.75,1, 1.5, 2, 2.5, 3.5, and 5 s. The interscan relaxation delay (d1) was kept constant at 5 s for all the experiments. Binding affinities were calculated as previously reported (Mayer and Meyer 2001; Angulo et al. 2010). To calculate saturation transfer difference (STD) values, absolute integrals of the difference spectra for three separate spectral regions (8.42-8.39 ppm, 7.83-7.78 ppm, and 7.42-7.26 ppm) were calculated and normalized over the reference integral (I_0). The STD amplification factor (STD-AF) was calculated using:

$$STD - AF = \varepsilon \cdot STD \quad (1)$$

where $\varepsilon = ([L]_{total}/[P]_{total})$, making STD-AF proportional to the fraction of bound protein. Once plotted against delay time, the data were fit to the STD buildup curve equation:

$$STD - AF(t_{sat}) = STD - AF_{max} \cdot (1 - \exp(-k_{sat} \cdot t_{sat})) \quad (2)$$

Individual variables can be seen in **Table S3**. STD-AF_0 was then calculated by multiplying STD-AF_{max} x $k_{sat.}$ Plotting STD-AF_0 versus ligand concentration allowed least-squares fitting to a Langmuir curve:

$$STD - AF_0([L]) = \frac{\alpha_{STD}[L]}{[L] + K_d} \quad (3)$$

Data processing and absolute integrals were calculated using NMRfx Analyst (Norris et al. 2016). Curve fitting and analysis were done using R Statistical Software v4.4.2 (Elzhov TV 2023; Wickham H 2023; RCoreTeam 2024).

Acknowledgements

This work was supported by a grant from the NIH (R01 GM106239 to K.H.G.). The authors would like to thank members of the Gardner Laboratory for useful discussions, Dr. Jia Liu (Director, CUNY ASRC Epigenetics Core) for assistance with the QuantStudio 7 Flex Real-time qPCR machine protocol to undertake the thermal shift assay, and Dr. Bruce Johnson (Senior Research Director, Computational Sciences, CUNY ASRC Structural Biology Initiative) for advice on the use of NMRfx Analyst.

References

Abramson J et al. 2024. Accurate structure prediction of biomolecular interactions with alphafold 3. Nature. 630(8016):493-500. https://dx.doi.org/10.1038/s41586-024-07487-w. doi:10.1038/s41586-024-07487-w.

Amezcua CA, Harper SM, Rutter J, Gardner KH. 2002. Structure and interactions of pas kinase n-terminal pas domain. Structure. 10(10):1349-1361. https://dx.doi.org/10.1016/S0969-2126(02)00857-2. doi:10.1016/s0969-2126(02)00857-2.

Angulo J, Enríquez-Navas PM, Nieto PM. 2010. Ligand–receptor binding affinities from saturation transfer difference (std) nmr spectroscopy: The binding isotherm of std initial growth rates. Chemistry – A European Journal. 16(26):7803-7812. https://dx.doi.org/10.1002/chem.200903528. doi:10.1002/chem.200903528.

Barelier S, Krimm I. 2011. Ligand specificity, privileged substructures and protein druggability from fragment-based screening. Current Opinion in Chemical Biology. 15(4):469-474. https://dx.doi.org/10.1016/j.cbpa.2011.02.020. doi:10.1016/j.cbpa.2011.02.020.

Best JL et al. 2004. Identification of small-molecule antagonists that inhibit an activator:Coactivator interaction. Proceedings of the National Academy of Sciences. 101(51):17622-17627. https://dx.doi.org/10.1073/pnas.0406374101. doi:10.1073/pnas.0406374101.

Cai W, Pei J, Grishin NV. 2004. Reconstruction of ancestral protein sequences and its applications. BMC Evolutionary Biology. 4(1):33. https://dx.doi.org/10.1186/1471-2148-4-33. doi:10.1186/1471-2148-4-33.

Challacombe JF et al. 2011. Biological consequences of ancient gene acquisition and duplication in the large genome of candidatus solibacter usitatus ellin6076. PLoS ONE. 6(9):e24882. https://dx.doi.org/10.1371/journal.pone.0024882. doi:10.1371/journal.pone.0024882.

Chaudhari AS et al. 2025. Light-dependent flavin redox and adduct states control the conformation and DNA-binding activity of the transcription factor el222. Nucleic Acids Research. 53(6). https://dx.doi.org/10.1093/nar/gkaf215. doi:10.1093/nar/gkaf215.

Cho H et al. 2016. On-target efficacy of a hif-2α antagonist in preclinical kidney cancer models. Nature. 539(7627):107-111. https://dx.doi.org/10.1038/nature19795. doi:10.1038/nature19795.

Cleere MM, Gardner KH. 2024. Optogenetic control of phosphate-responsive genes using single-component fusion proteins in *saccharomyces cerevisiae*. ACS Synthetic Biology. 13(12):4085-4098. https://dx.doi.org/10.1021/acssynbio.4c00529. doi:10.1021/acssynbio.4c00529.

De Meutter J, Goormaghtigh E. 2021. Amino acid side chain contribution to protein ftir spectra: Impact on secondary structure evaluation. European Biophysics Journal. 50(3-4):641-651. https://dx.doi.org/10.1007/s00249-021-01507-7. doi:10.1007/s00249-021-01507-7.

Dong H, Paramonov SE, Hartgerink JD. 2008. Self-assembly of α-helical coiled coil nanofibers. Journal of the American Chemical Society. 130(41):13691-13695. https://dx.doi.org/10.1021/ja8037323. doi:10.1021/ja8037323.

Edgar RC. 2004. Muscle: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32(5):1792-1797. https://dx.doi.org/10.1093/nar/gkh340. doi:10.1093/nar/gkh340.

Elzhov TV MK, Spiess A, Bolker B. Minpack.Lm: R interface to the levenberg-marquardt nonlinear least-squares algorithm found in minpack, plus support for bounds. In:

Gabler F et al. 2020. Protein sequence analysis using the mpi bioinformatics toolkit. Current Protocols in Bioinformatics. 72(1). https://dx.doi.org/10.1002/cpbi.108. doi:10.1002/cpbi.108.

Gumerov VM, Zhulin IB. 2020. Trend: A platform for exploring protein function in prokaryotes based on phylogenetic, domain architecture and gene neighborhood analyses. Nucleic Acids Research. 48(W1):W72-W76. https://dx.doi.org/10.1093/nar/gkaa243. doi:10.1093/nar/gkaa243.

Guo Y et al. 2013. Regulating the arnt/tacc3 axis: Multiple approaches to manipulating protein/protein interactions with small molecules. ACS Chemical Biology. 8(3):626-635. https://dx.doi.org/10.1021/cb300604u. doi:10.1021/cb300604u.

Harper SM, Neil LC, Gardner KH. 2003. Structural basis of a phototropin light switch. Science. 301(5639):1541-1544. https://dx.doi.org/10.1126/science.1086810. doi:10.1126/science.1086810.

Heimburg T, Schünemann J, Weber K, Geisler N. 1999. Ftir-spectroscopy of multistranded coiled coil proteins. Biochemistry. 38(39):12727-12734. https://dx.doi.org/10.1021/bi983079h. doi:10.1021/bi983079h.

Henry JT, Crosson S. 2011. Ligand-binding pas domains in a genomic, cellular, and structural context. Annual Review of Microbiology. 65(1):261-286. https://dx.doi.org/10.1146/annurev-micro-121809-151631. doi:10.1146/annurev-micro-121809-151631.

Hoffman SM, Espinel-Ríos S, Kwartler SK, Lalwani MA, Avalos JL. 2025. Balancing doses of el222 and light improves optogenetic induction of protein production in *komagataella phaffii*. Biotechnology and Bioengineering. https://dx.doi.org/10.1002/bit.29027. doi:10.1002/bit.29027.

Jonasch E et al. 2021. Belzutifan for renal cell carcinoma in von hippel–lindau disease. New England Journal of Medicine. 385(22):2036-2046. https://dx.doi.org/10.1056/NEJMoa2103425. doi:10.1056/nejmoa2103425.

Kapust RB et al. 2001. Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Engineering, Design and Selection. 14(12):993-1000. https://dx.doi.org/10.1093/protein/14.12.993. doi:10.1093/protein/14.12.993.

Kendrick BS, Gabrielson JP, Solsberg CW, Ma E, Wang L. 2020. Determining spectroscopic quantitation limits for misfolded structures. Journal of Pharmaceutical Sciences. 109(1):933-936. https://dx.doi.org/10.1016/j.xphs.2019.09.004. doi:10.1016/j.xphs.2019.09.004.

Key J, Scheuermann TH, Anderson PC, Daggett V, Gardner KH. 2009. Principles of ligand binding within a completely buried cavity in hif2α pas-b. Journal of the American Chemical Society. 131(48):17647-17654. https://dx.doi.org/10.1021/ja9073062. doi:10.1021/ja9073062.

Kirkman T, Dos Santos Silva C, Tosin M, Bertacine Dias MV. 2024. How to find a fragment: Methods for screening and validation in fragment-based drug

discovery. ChemMedChem. 19(24). https://dx.doi.org/10.1002/cmdc.202400342. doi:10.1002/cmdc.202400342.

Kong J, Yu S. 2007. Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochimica et Biophysica Sinica. 39(8):549-559. https://dx.doi.org/10.1111/j.1745-7270.2007.00320.x. doi:10.1111/j.1745-7270.2007.00320.x.

Laschet C, Dupuis N, Hanson J. 2018. The g protein-coupled receptors deorphanization landscape. Biochemical Pharmacology. 153(62-74. https://dx.doi.org/10.1016/j.bcp.2018.02.016. doi:10.1016/j.bcp.2018.02.016.

Lazar JT, Tabor JJ. 2021. Bacterial two-component systems as sensors for synthetic biology applications. Current Opinion in Systems Biology. 28(100398. https://dx.doi.org/10.1016/j.coisb.2021.100398. doi:10.1016/j.coisb.2021.100398.

Lee P-H, Huang XX, Teh BT, Ng L-M. 2019. Tsa-craft: A free software for automatic and robust thermal shift assay data analysis. SLAS Discovery. 24(5):606-612. https://dx.doi.org/10.1177/2472555218823547. doi:10.1177/2472555218823547.

Letunic I, Khedkar S, Bork P. 2021. Smart: Recent updates, new developments and status in 2020. Nucleic Acids Research. 49(D1):D458-D460. https://dx.doi.org/10.1093/nar/gkaa937. doi:10.1093/nar/gkaa937.

Liu LL, Wang L, Zonderman J, Rouse JC, Kim H-Y. 2020. Automated, high-throughput infrared spectroscopy for secondary structure analysis of protein biopharmaceuticals. Journal of Pharmaceutical Sciences. 109(10):3223-3230. https://dx.doi.org/10.1016/j.xphs.2020.07.030. doi:10.1016/j.xphs.2020.07.030.

Lucas L, Tsoi PS, Nair A, Ferreon ACM, Ferreon JC. 2025. Sensing protein structural transitions with microfluidic modulation infrared spectroscopy. Biosensors. 15(6):382. https://dx.doi.org/10.3390/bios15060382. doi:10.3390/bios15060382.

Ma TM, Huang RH, Collins VI, Wikström M, Batabyal D. 2023. Advancement of microfluidic modulation spectroscopy as a highly sensitive protein characterization technique for the detection of small structural changes. Analytical Biochemistry. 683(115350. https://dx.doi.org/10.1016/j.ab.2023.115350. doi:10.1016/j.ab.2023.115350.

Mayer M, Meyer B. 2001. Group epitope mapping by saturation transfer difference nmr to identify segments of a ligand in direct contact with a protein receptor. Journal of the American Chemical Society. 123(25):6108-6117. https://dx.doi.org/10.1021/ja0100120. doi:10.1021/ja0100120.

Motta-Mena LB et al. 2014. An optogenetic gene expression system with rapid activation and deactivation kinetics. Nature Chemical Biology. 10(3):196-202. https://dx.doi.org/10.1038/nchembio.1430. doi:10.1038/nchembio.1430.

Nash AI et al. 2011. Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (lov-hth) DNA-binding protein. Proceedings of the National Academy of Sciences. 108(23):9449-9454. https://dx.doi.org/10.1073/pnas.1100262108. doi:10.1073/pnas.1100262108.

Norris M, Fetler B, Marchant J, Johnson BA. 2016. Nmrfx processor: A cross-platform nmr data processing program. Journal of Biomolecular NMR. 65(3-4):205-216. https://dx.doi.org/10.1007/s10858-016-0049-6. doi:10.1007/s10858-016-0049-6.

Phan J et al. 2002. Structural basis for the substrate specificity of tobacco etch virus protease. Journal of Biological Chemistry. 277(52):50564-50572. https://dx.doi.org/10.1074/jbc.M207224200. doi:10.1074/jbc.m207224200.

RCoreTeam (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Reade A et al. 2016. Tael: A zebrafish-optimized optogenetic gene expression system with fine spatial and temporal control. Development. 144(2):345-355. https://dx.doi.org/10.1242/dev.139238. doi:10.1242/dev.139238.

Rivera-Cancel G, Motta-Mena LB, Gardner KH. 2012. Identification of natural and artificial DNA substrates for light-activated lov–hth transcription factor el222. Biochemistry. 51(50):10024-10034. https://dx.doi.org/10.1021/bi301306t. doi:10.1021/bi301306t.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new endscript server. Nucleic Acids Research. 42(W1):W320-W324. https://dx.doi.org/10.1093/nar/gku316. doi:10.1093/nar/gku316.

Rullan M, Benzinger D, Schmidt GW, Milias-Argeitis A, Khammash M. 2018. An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation. Molecular Cell. 70(4):745-756.e746. https://dx.doi.org/10.1016/j.molcel.2018.04.012. doi:10.1016/j.molcel.2018.04.012.

Scheuermann TH et al. 2009. Artificial ligand binding within the hif2α pas-b domain of the hif2 transcription factor. Proceedings of the National Academy of Sciences. 106(2):450-455. https://dx.doi.org/10.1073/pnas.0808092106. doi:10.1073/pnas.0808092106.

Stuffle EC, Johnson MS, Watts KJ. 2021. Pas domains in bacterial signal transduction. Current Opinion in Microbiology. 61(8-15. https://dx.doi.org/10.1016/j.mib.2021.01.004. doi:10.1016/j.mib.2021.01.004.

Taylor BL, Zhulin IB. 1999. Pas domains: Internal sensors of oxygen, redox potential, and light. Microbiology and Molecular Biology Reviews. 63(2):479-506. https://dx.doi.org/10.1128/mmbr.63.2.479-506.1999. doi:10.1128/mmbr.63.2.479-506.1999.

Ulrich LE, Koonin EV, Zhulin IB. 2005. One-component systems dominate signal transduction in prokaryotes. Trends in Microbiology. 13(2):52-56. https://dx.doi.org/10.1016/j.tim.2004.12.006. doi:10.1016/j.tim.2004.12.006.

Wei B et al. 2025. Real-time monitoring of higher-order structure of rnas by temperature-course size exclusion chromatography and microfluidic modulation spectroscopy. Analytical Chemistry. 97(10):5632-5642. https://dx.doi.org/10.1021/acs.analchem.4c06343. doi:10.1021/acs.analchem.4c06343.

Wickham H FR, Henry L, Müller K, Vaughan D. (2025). Dplyr: A Grammar of Data Manipulation. R package version 1.1.4, https://dplyr.tidyverse.org

Wu T et al. 2023. Conformationally responsive dyes enable protein-adaptive differential scanning fluorimetry. https://dx.doi.org/10.1101/2023.01.23.525251. doi:10.1101/2023.01.23.525251.

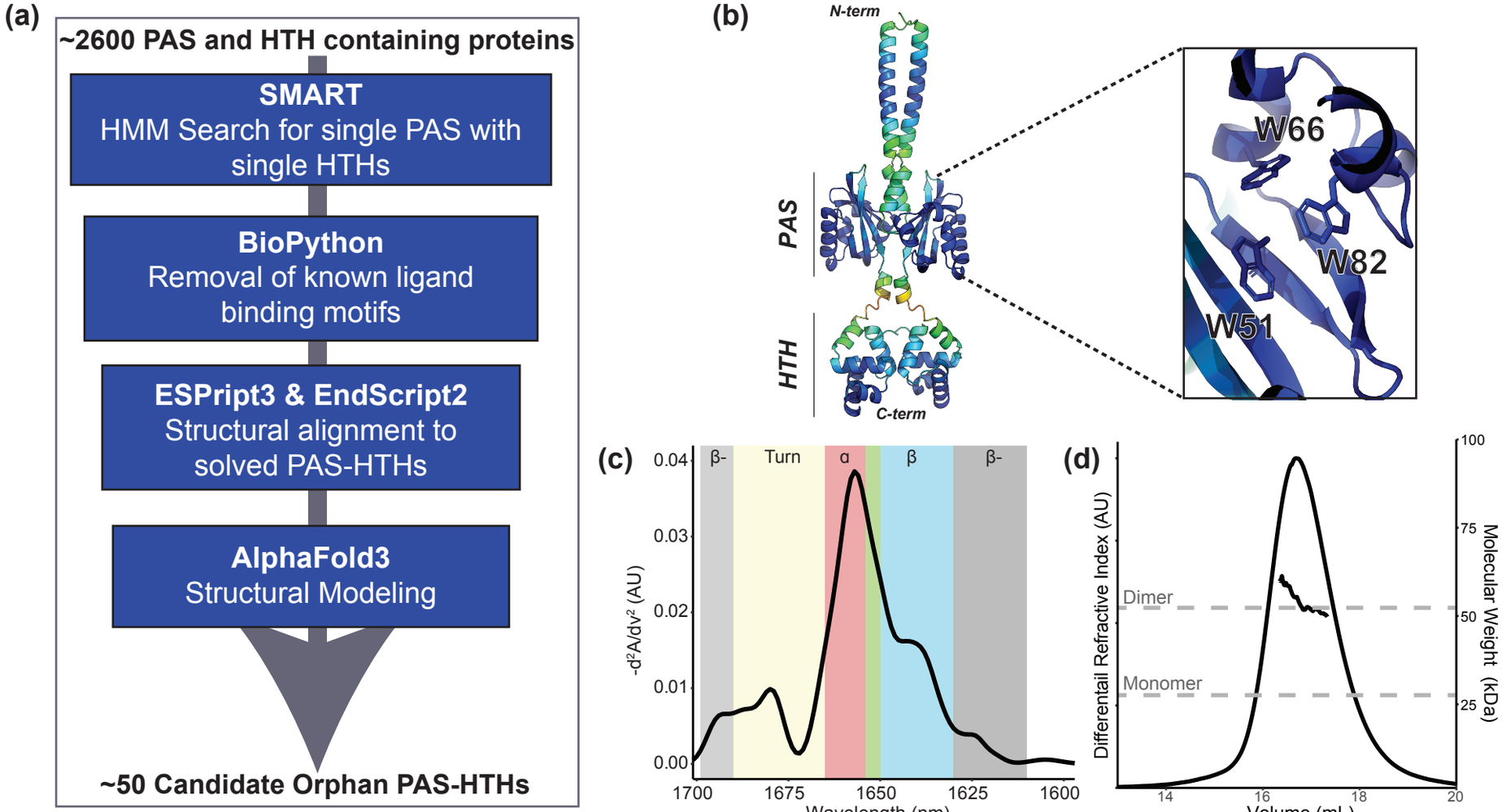
Xu X, Dikiy I, Evans MR, Marcelino LP, Gardner KH. 2021. Fragile protein folds: Sequence and environmental factors affecting the equilibrium of two interconverting, stably folded protein conformations. Magnetic Resonance. 2(1):63-76. https://dx.doi.org/10.5194/mr-2-63-2021. doi:10.5194/mr-2-63-2021.

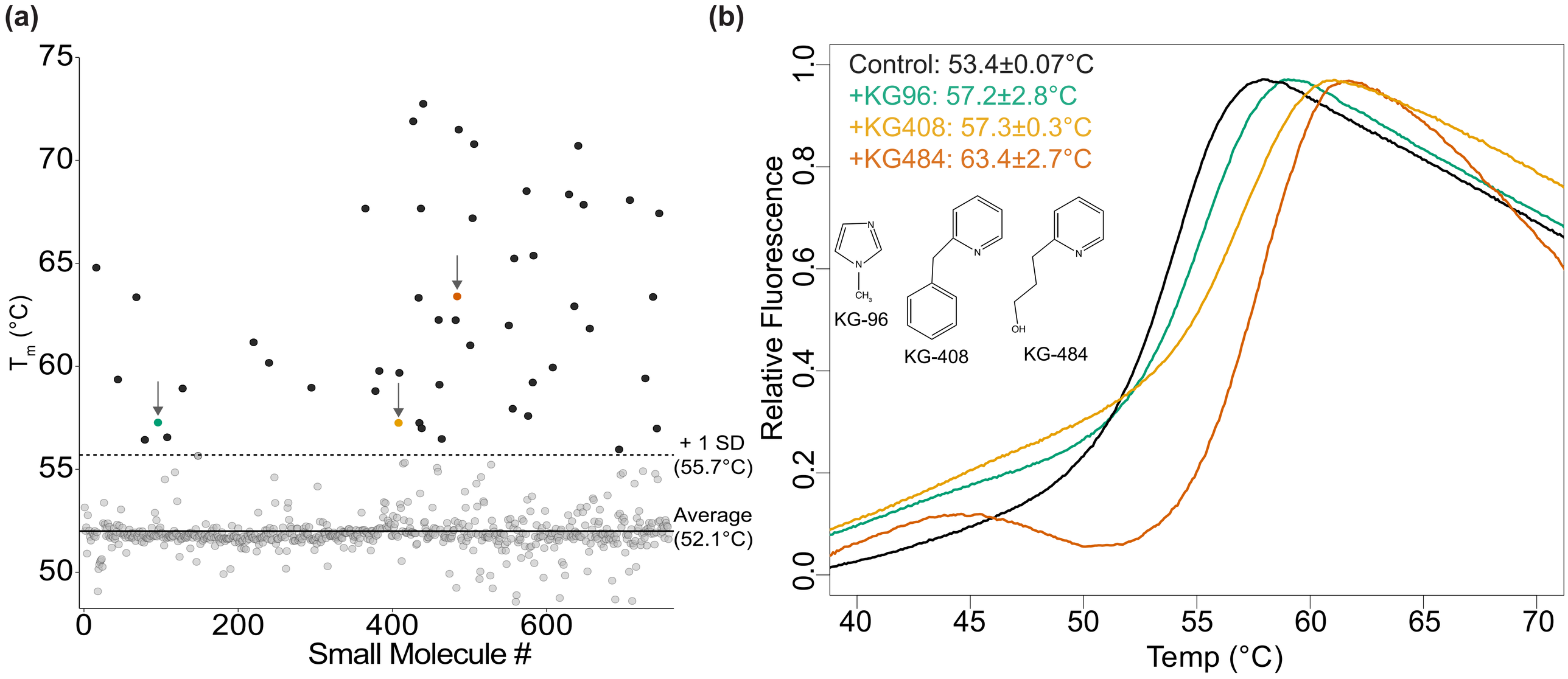
Zhao EM et al. 2020. Design and characterization of rapid optogenetic circuits for dynamic control in yeast metabolic engineering. ACS Synthetic Biology. 9(12):3254-3266. https://dx.doi.org/10.1021/acssynbio.0c00305. doi:10.1021/acssynbio.0c00305.

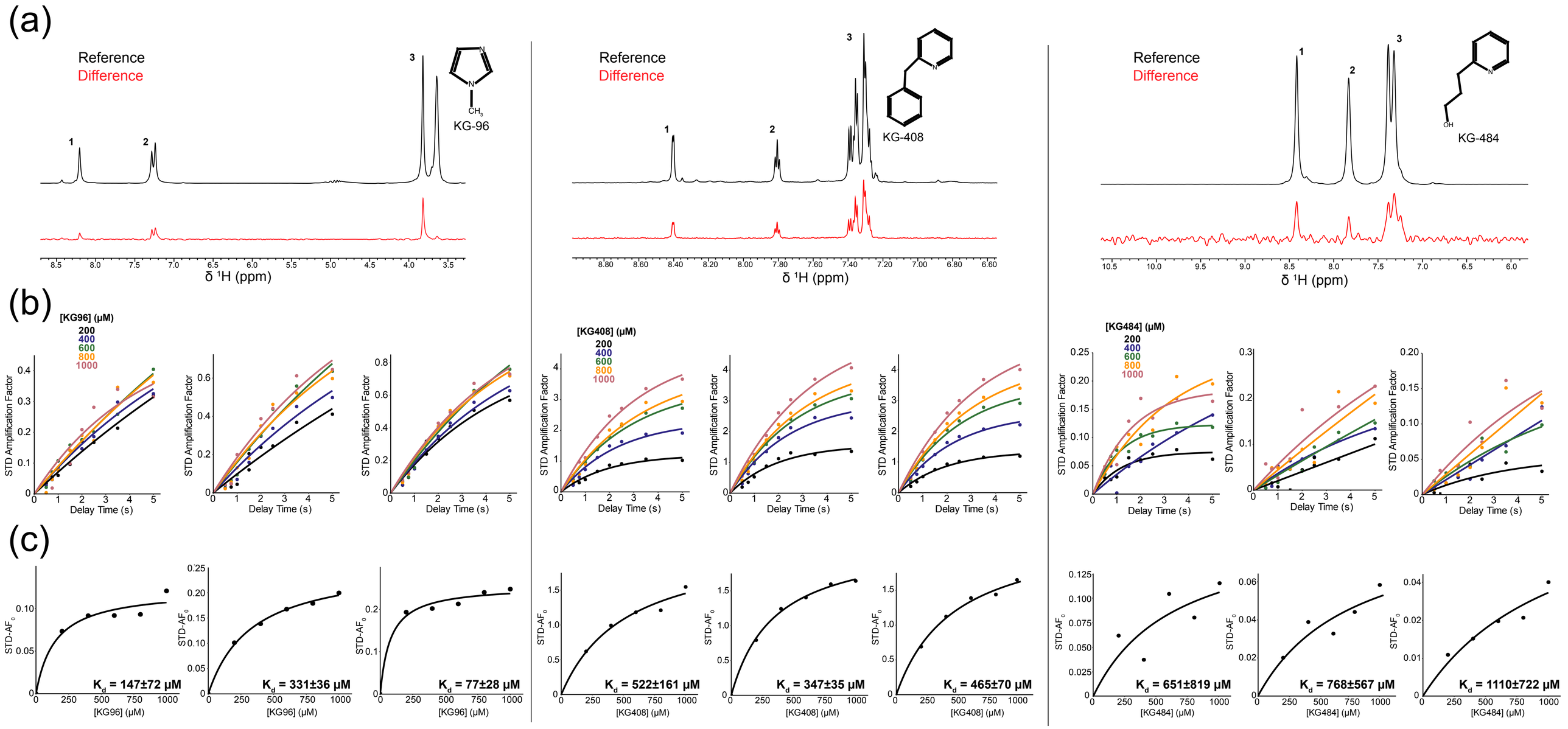
Zhao EM et al. 2018. Optogenetic regulation of engineered cellular metabolism for microbial chemical production. Nature. 555(7698):683-687. https://dx.doi.org/10.1038/nature26141. doi:10.1038/nature26141.

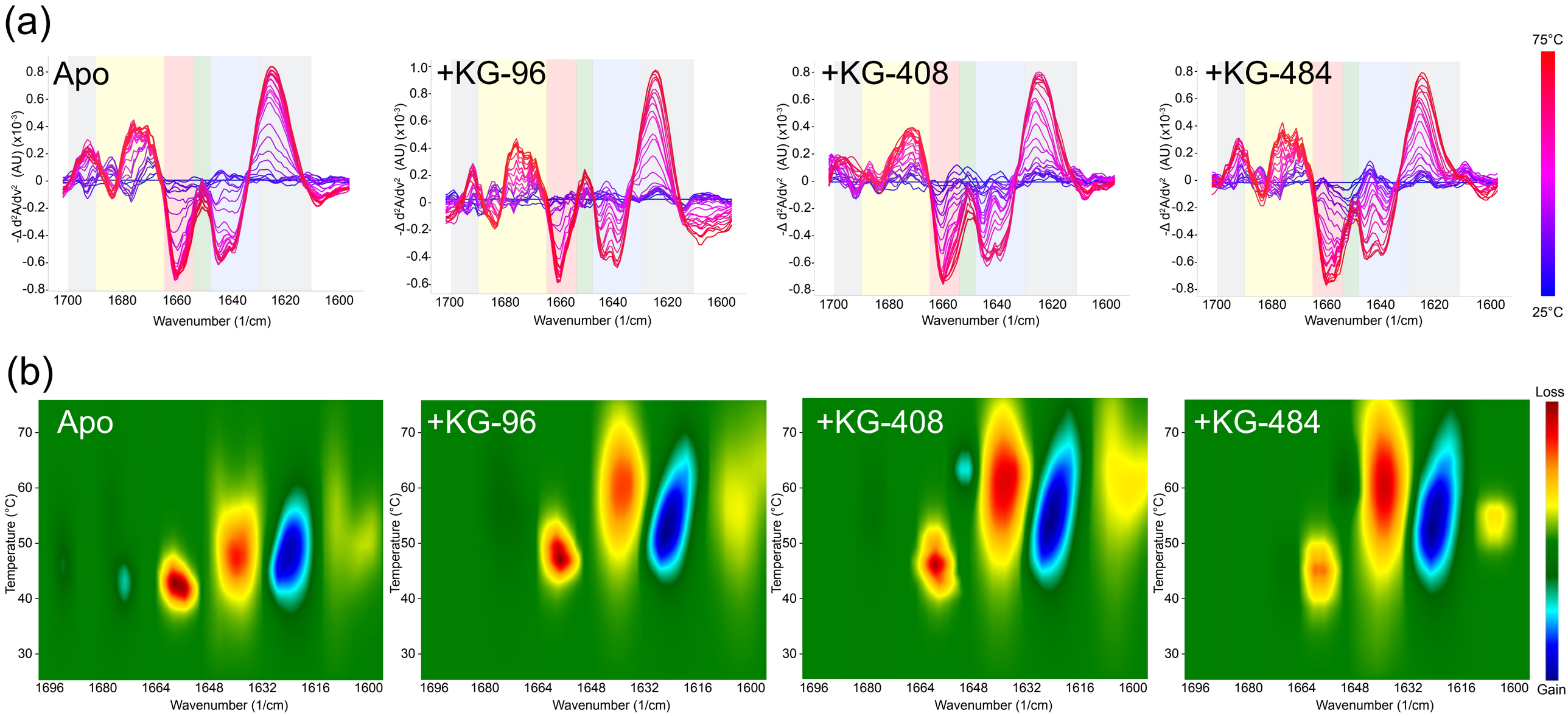
Zimmermann L et al. 2018. A completely reimplemented mpi bioinformatics toolkit with a new hhpred server at its core. Journal of Molecular Biology. 430(15):2237-2243. https://dx.doi.org/10.1016/j.jmb.2017.12.007. doi:10.1016/j.jmb.2017.12.007.

Zoltowski BD, Motta-Mena LB, Gardner KH. 2013. Blue light-induced dimerization of a bacterial lov–hth DNA-binding protein. Biochemistry. 52(38):6653-6661. https://dx.doi.org/10.1021/bi401040m. doi:10.1021/bi401040m.


(a)
~2600 PAS and HTH containing proteins
SMART
HMM Search for single PAS with single HTHs
BioPython
Removal of known ligand binding motifs
ESPript3 & EndScript2
Structural alignment to solved PAS-HTHs
AlphaFold3
Structural Modeling
~50 Candidate Orphan PAS-HTHs
(b)
N-term
PAS
HTH
C-term
W66
W82
W51
(c)
β-
Turn
α
β
β-
-d²A/dv² (AU)
0.04
0.03
0.02
0.01
0.00
1700
1675
1650
1625
1600
Wavelength (nm)
(d)
Differentail Refractive Index (AU)
Molecular Weight (kDa)
Dimer
Monomer
100
75
50
25
14
16
18
20
Volume (mL)

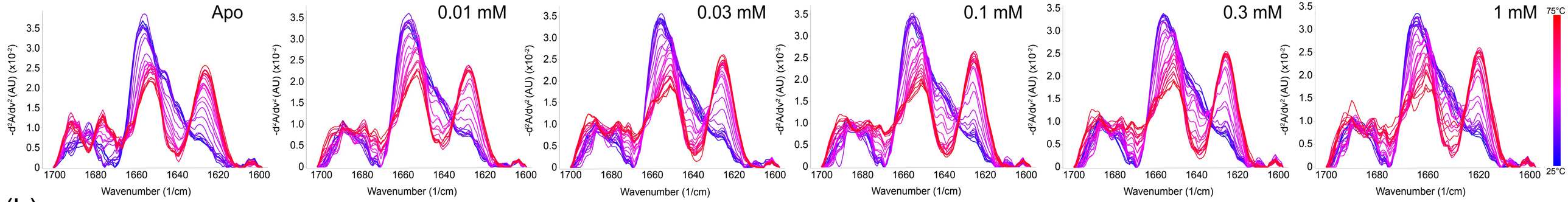

(a)
75
70
65
60
55
50
T_m (°C)
0
200
400
600
Small Molecule #
+ 1 SD
(55.7°C)
Average
(52.1°C)
(b)
Control: 53.4±0.07°C
+KG96: 57.2±2.8°C
+KG408: 57.3±0.3°C
+KG484: 63.4±2.7°C
KG-96
KG-408
KG-484
Relative Fluorescence
0.0
0.2
0.4
0.6
0.8
1.0
40
45
50
55
60
65
70
Temp (°C)


(a)
Reference
Difference
KG-96
KG-408
KG-484
δ ¹H (ppm)
(b)
[KG96] (μM)
[KG408] (μM)
[KG484] (μM)
200
400
600
800
1000
STD Amplification Factor
Delay Time (s)
(c)
STD-AF0
[KG96] (μM)
[KG408] (μM)
[KG484] (μM)
Kd = 147±72 μM
Kd = 331±36 μM
Kd = 77±28 μM
Kd = 522±161 μM
Kd = 347±35 μM
Kd = 465±70 μM
Kd = 651±819 μM
Kd = 768±567 μM
Kd = 1110±722 μM

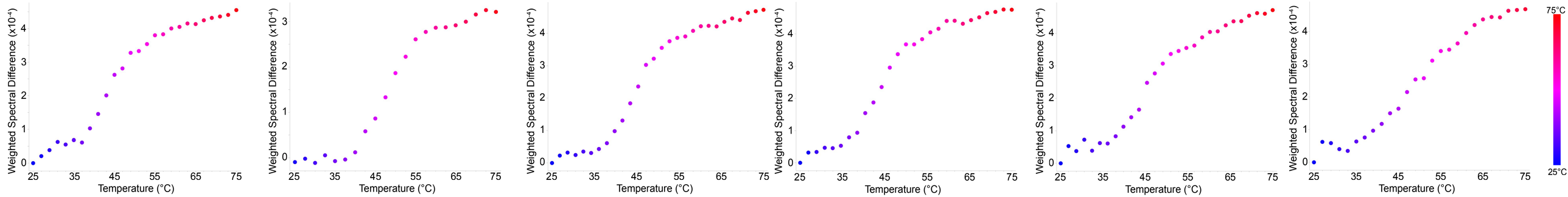

(a)
Apo
+KG-96
+KG-408
+KG-484
-Δ d²A/dν² (AU) (x10⁻³)
Wavenumber (1/cm)
75°C
25°C
(b)
Apo
+KG-96
+KG-408
+KG-484
Temperature (°C)
Wavenumber (1/cm)
Loss
Gain

+KG-408

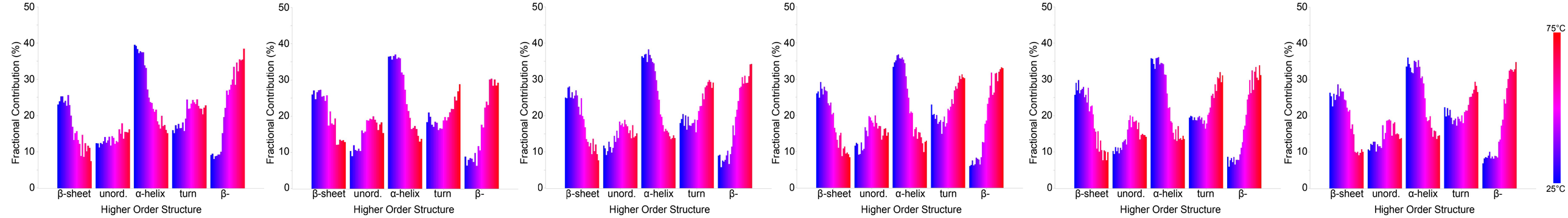
(a)


Apo
0.01 mM
0.03 mM
0.1 mM
0.3 mM
1 mM
-d²A/dν² (AU) (x10⁻²)
Wavenumber (1/cm)
75°C
25°C

(b)


Weighted Spectral Difference (x10⁻⁴)
Temperature (°C)
75°C
25°C

(c)


Fractional Contribution (%)
β-sheet
unord.
α-helix
turn
β-
Higher Order Structure
75°C
25°C