THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

Cite This: J. Phys. Chem. A 2018, 122, 2947−2955

pubs.acs.org/JPCA

Conformational Ensembles of Calmodulin Revealed by Nonperturbing Site-Specific Vibrational Probe Groups

Published as part of The Journal of Physical Chemistry virtual special issue "Time-Resolved Vibrational Spectroscopy".

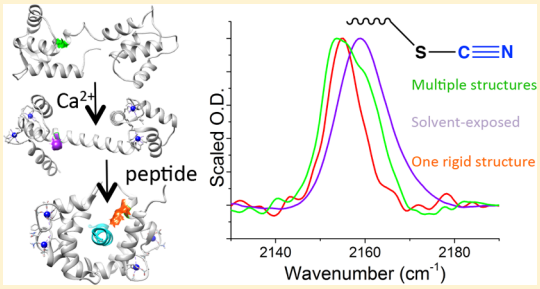
Kristen L. Kelly,[†] Shannon R. Dalton,[†,+] Rebecca B. Wai,[†] Kanika Ramchandani,[†] Rosalind J. Xu,[†] Sara Linse,[‡] and Casey H. Londergan*[,†]

[†]Department of Chemistry, Haverford College, Haverford, Pennsylvania 19041, United States

[‡]Department of Biochemistry and Structural Biology, Lund University, 221 00 Lund, Sweden

Supporting Information

ABSTRACT: Seven native residues on the regulatory protein calmodulin, including three key methionine residues, were replaced (one by one) by the vibrational probe amino acid cyanylated cysteine, which has a unique CN stretching vibration that reports on its local environment. Almost no perturbation was caused by this probe at any of the seven sites, as reported by CD spectra of calcium-bound and *apo* calmodulin and binding thermodynamics for the formation of a complex between calmodulin and a canonical target peptide from skeletal muscle myosin light chain kinase measured by isothermal titration. The surprising lack of perturbation suggests that this probe group could be applied directly in many protein−protein binding interfaces. The infrared absorption bands for the probe groups reported many dramatic changes in the probes' local environments as CaM went from *apo*- to calcium-saturated to target peptide-bound conditions, including large frequency shifts and a variety of line shapes from narrow (interpreted as a rigid and invariant local environment) to symmetric to broad and asymmetric (likely from multiple coexisting and dynamically exchanging structures). The fast intrinsic time scale of infrared spectroscopy means that the line shapes report directly on site-specific details of calmodulin's variable structural distribution. Though quantitative interpretation of the probe line shapes depends on a direct connection between simulated ensembles and experimental data that does not yet exist, formation of such a connection to data such as that reported here would provide a new way to evaluate conformational ensembles from data that directly contains the structural distribution. The calmodulin probe sites developed here will also be useful in evaluating the binding mode of calmodulin with many uncharacterized regulatory targets.



INTRODUCTION

The fast intrinsic time scale of vibrational spectroscopy (from tens of femtoseconds to tens of picoseconds, depending on frequency separation and wavelength) means that nearly all functional biomolecular structural fluctuations are expected to be in slow exchange on the IR time scale. Thus, the line shapes of vibrational resonances from proteins can directly report on the conformational distribution in the local environment of a chosen normal mode. Though functional groups with spectrally isolated vibrational frequencies have been used increasingly to provide site-specific information in proteins,[1−3] the one-dimensional line shapes for these vibrational probe groups have been largely ignored as sources of information about conformational ensembles. This study uses vibrational probe groups (in this case, thiocyanate groups) attached to a highly promiscuous and structurally modular regulatory protein (calmodulin) to document new features of the protein's conformational distribution as its structure and environment change. Although we show that the probes are themselves functionally nonperturbative, our data also reveal new and dramatic site-specific changes in calmodulin's dynamic conformational distribution as it binds calcium ions and then associates with a target peptide.

Calmodulin (CaM) is a 15 kDa calcium binding protein that is highly conserved among eukaryotes and is the primary calcium signaling protein for mammals. CaM serves as a secondary effector of Ca^{2+} regulatory signals by binding to proteins whose functions are regulated by changes in intracellular or extracellular Ca^{2+} levels but do not bind Ca^{2+} themselves. CaM binds >500 natural sequence targets,[4−6] and its promiscuity in binding is unmatched among proteins in the human genome. Because CaM plays such a central role in calcium-activated signaling, it is widely predicted to bind to many other regulatory proteins, often by rigidifying otherwise disordered domains of those proteins.[4,7] Only a fraction of CaM target motifs in target proteins have been characterized structurally, and in most cases, the structure of the calmodulin−

Received: January 15, 2018
Published: February 5, 2018

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

target complexes was determined using a fragment peptide as the target[8] rather than the full protein, thus enabling either crystallization of the complex or structural determination via multidimensional nuclear magnetic resonance (NMR). Many bound structures are uncharacterized due to CaM's inherent structural modularity, both in its own structure[9,10] and in its binding mode.[5,11] With a few well-characterized structures, its relative ease of recombinant expression, and its remarkable thermodynamic stability,[12] CaM is also a convenient model system to develop novel bioanalytical techniques related to dynamic conformational distributions and protein−protein binding.

CaM's sequence includes four modular EF-hand helix-loop helix subdomains, arranged in two pairs with a short flexible linker between the N- and C-terminal pairs (see Figure 1 for sequence and canonical published structures). The CaM sequence includes many acidic residues, resulting in high water solubility, and CaM is also highly enriched in hydrophobic residues, leading to highly collapsed structures. In all conditions, CaM's secondary structure is very helical and its main functional changes are in tertiary and quaternary contacts. *apo*-CaM's structure[13] (NMR structure shown in Figure 1B) is very helical, strongly collapsed, and likely modular in its tertiary arrangement of helices.[9,10] After binding Ca^{2+} with pM affinity, each EF-hand experiences an increase (on average) and rigidification of the interhelix angle, thus exposing hydrophobic side chains involved in target binding.[14,15] Though Ca^{2+}-saturated CaM is thus more rigid and less dynamic than *apo*-CaM, NMR data and simulations each suggest that it apparently still adopts a greater range of structures[14,16] than suggested by its rigid and "open" crystallized form[15] (shown in Figure 1C).

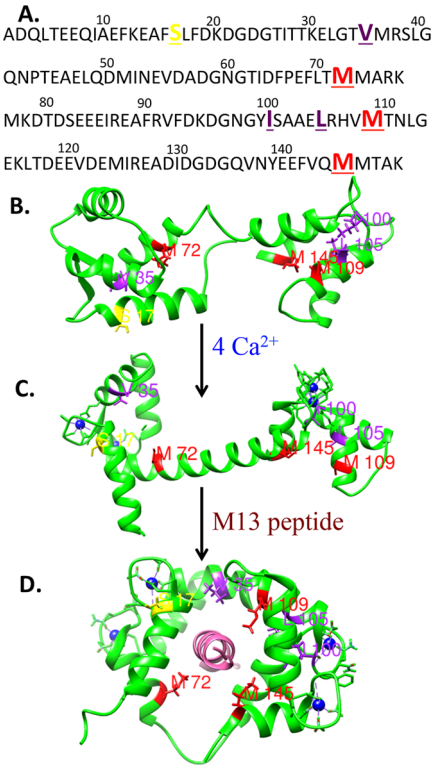


Figure 1. A. Mammalian calmodulin sequence with sites chosen for substitution by C* highlighted and underlined. B. NMR structure of *apo*-CaM (PDB: 1CFD).[13] C. X-ray crystal structure of Ca^{2+}-saturated CaM (PDB: 1CLL).[15] D. NMR structure of the complex between Ca^{2+}-saturated CaM and the skMLCK "M13" target peptide, colored in pink (PDB:2BBM).[17] All structures show selected sites for substitution of C* colored in red (key methionine resides), purple (other residues not directly involved in wraparound binding), or yellow ("control" S17 residue).

Although CaM binds its targets in several different possible geometries, the most common characterized "wraparound" binding geometry has CaM completely encircling a helical target sequence in what is usually an antiparallel manner (with the C-terminus of CaM around the N-terminal part of the target). The canonical example of the "wraparound" structure is CaM's complex with the CaM binding peptide (here called "M13" following prior convention) of skeletal muscle myosin light chain kinase (skMLCK),[17] shown in Figure 1D. "Wraparound" target-bound complexes of CaM tend to be structurally rigid and amenable to crystallization and/or NMR studies. From a conformational ensemble point of view, though CaM is always very collapsed and helical, its distribution of structures narrows greatly upon first binding calcium ions and then again upon binding to a target protein. CaM is a protein whose multiple environmentally induced functional changes provide a rich model for studying protein dynamics and structural ensembles.

Many of the known target sequences that bind CaM in "wraparound" mode are further classified by the pattern of conserved hydrophobic anchor residues on the target sequence: counting from the first hydrophobic binding "anchor" residue on the target CaM binding domain, the anchor patterns are usually at positions 1−5−10 or 1−8−14,[4−6] although there are many known exceptions to these anchor patterns, such as the anthrax exotoxin,[5,18] and there are likely to be further exceptions and novel binding modes discovered as more CaM−target complexes are structurally characterized. Although many novel CaM targets have been predicted using statistical methods[4,7] and/or experimentally identified via high-throughput screens,[19,20] identification of the exact CaM binding sequence and structural characterization of new CaM-bound complexes remains a slow process using conventional structural biology techniques.

Vibrational probe methodology has expanded in recent years to encompass many different functional groups with unique vibrational frequencies that can report on (unfortunately often indeterminate) features of their local structural environments. Probe functional groups tend to be very small compared to other labeling groups (often just two atoms)[1−3] and can be placed directly into binding interfaces between biomolecular species.[21−25] The covalently bound thiocyanate group can be introduced via otherwise bio-orthogonal chemical modification of either cysteine residues,[26−28] which can be placed at specific sequence sites by site-directed mutagenesis and recombinant protein expression, or other solution-exposed thiols.[29] The SCN group has a unique nitrile CN stretching vibration at about 2160 cm^{-1} (an otherwise clear spectral region) whose frequency and line shape are strongly sensitive to the local structural environment.[30−32] Replacement of native residues along a binding interface by β-thiocyano-alanine (or cyanylated cysteine, C* for short) is surprisingly nonperturbative both to the protein's local secondary structure[33] and to protein−protein binding thermodynamics.[25]

The line shapes of resonances from vibrational probe groups are uniquely able to report directly (rather than in a dynamically averaged and under-determined way) on the

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

changing conformational distributions of proteins. IR spectra can be collected for diverse samples, including large proteins and protein assemblies not accessible to NMR studies. Vibrational probe groups placed on CaM can report on new and otherwise unreportable features of CaM's conformational distribution, as it changes from *apo*- to Ca^{2+}-saturated to target-bound. Such probes might also be used to observe interactions of CaM with whole target proteins, most of which are substantially larger than CaM itself.

We recently reported[25] dramatic changes in the CN stretching IR absorption bands of SCN labels at six different sites on the M13 peptide as CaM wrapped around it. Furthermore, we showed that the artificial C* residues on the target peptide did not substantially perturb the binding between the species. The probe IR frequencies were consistent with expectations for solvent exposure based on previous measurements, but the relatively broad line shapes indicated a range of probe environments that appeared to report on local conformational dynamics, even in this relatively rigid (yet still microscopically dynamic[16,34]) bound complex.

In the current work, the SCN labels are covalently attached to CaM rather than the target peptide. This allows us to interrogate CaM under three different sets of conditions (*apo*, Ca^{2+}-saturated, and bound to the M13 peptide) and to view functionally important, site-specific changes in the protein's dynamic structure between the three conditions. We chose seven label sites (Figure 1) where C* was substituted for the native residue on CaM, including three methionine residues that are "key residues" for wraparound binding (M72, M109, and M145), three residues not directly involved in that binding mode but whose environments are expected to change in different forms of CaM (L105, I110, and V35), and one "control" residue (S17) far away from functionally interacting surfaces.

Our goals in the current work include further quantitative assessment of any probe-induced structural perturbation (by isothermal titration calorimetry of CaM's binding to M13); site-specific measurement of changes in CaM's conformational distribution as CaM rigidifies and its ensemble of structures narrows from *apo* to Ca^{2+} to target-bound; collection of a model data set for the development of vibrational probe methodology to determine conformational ensembles; and formation of a possible spectroscopic assay to determine the binding mode of CaM with otherwise uncharacterized and/or whole-protein targets.

EXPERIMENTAL METHODS

Materials. All chemicals were obtained from Sigma/Aldrich and used as received unless otherwise mentioned below. Cells and media were obtained from GE Life Sciences and MP Biomedical, respectively. All buffers were prepared using doubly deionized water (>17.8 mohm resistivity) and filtered and degassed before all uses, especially including chromatography under any applied pressure.

Site-Directed Mutagenesis. The plasmid for mammalian sequence calmodulin was mutated at the seven selected codon sites using a single megaprimer approach with primers designed using online tools and ordered from Eurofins Operon. The Stratagene QuikChange Ultra kit was used and successful PCR reactions were transformed and amplified using XL1-blue supercompetent cells.

Expression and Purification of CaM Single-Cysteine Variants. Recombinant expression and purification of all CaM variants followed precedent from the Linse lab.[9,35] 1L cultures of transformed BL21 Star (DE3)pLysS One Shot *E. coli* cells were induced to overexpression and grown overnight at 18 °C, and sonicated protein lysate was boiled before column purification on (1) a DEAE fast flow column (GE Healthcare), (2) a home-packed Phenyl Sepharose FF column (media from GE), and (3) a Superdex 75 10/300 column (GE) used following reduction with 100× DTT (see next paragraph).

Cyanylation of Cysteine. Approximately 3×10^{-8} mol of cysteine-variant proteins purified by Phenyl Sepharose was transferred via PD-10 size exclusion column to a denaturing reaction buffer of 9 M Urea, 20 mM Hepes, 10 mM EDTA, pH 7.0 that enabled chemical access to each of the selected cysteines and treated with 100-fold molar excess of dithiothreitol (DTT) to reduce disulfide dimers. DTT was separated via a Superdex 75 10/300 column equilibrated in the same urea-based reaction buffer, and proteins were isolated and reconcentrated to approximately 500 μL volume using Sartorius Vivaspin centrifugal concentrators with 3000 molecular weight cutoff. 8× molar quantity of a stock solution of 25 mM 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) in 100 mM sodium phosphate buffer, pH 7.0 was directly added to the protein solution and the reaction's progress was monitored using the UV absorbance at 412 nm.[36] After approximately 1 h and observation of near-quantitative reaction with cysteine thiols, 55-fold molar excess of NaCN predissolved in the reaction buffer was added. After 30 min, the protein (now cyanylated at cysteine) was reisolated using a PD-10 size exclusion column pre-equilibrated in the *apo* buffer (see below). Centrifugal reconcentration and an additional PD-10 separation were used to change buffers to the Ca^{2+}-containing conditions (see below).

Peptide Synthesis and Purification. The M13 peptide (sequence Ac-KRRWKKNFIAVSAANRFKKISSSGAL-NH_2) was synthesized using solid-phase synthesis on the PAL resin using fmoc chemistry and HOBt/HATU as activator cocktail, cleaved under standard conditions in neat trifluoroacetic acid solvent, and purified using HPLC, following previous successful procedures reported for this peptide.[25]

Isothermal Titration Calorimetry. A TA Instruments NanoITC low-volume instrument with a working volume of 194 μL held at a temperature of 25.0 °C was used for all ITC experiments. Approximately 350 μL of an 80 μM solution (measured to ±20 uM by UV spectroscopy; the extinction coefficient of CaM is low and does not permit a highly accurate concentration determination) of a selected SCN-labeled CaM variant was prepared via overnight dialysis at 4 °C via a 3.5 kDa MWCO D-Tube Dialyzer Maxi (Novagen) against filtered and degassed 20 mM Hepes, 10 mM $CaCl_2$, pH 7.0 buffer, and that solution was placed in the ITC sample cell. An 80 μL aliquot of a 400 μM solution of the M13 peptide (measured to ±20 μM by UV spectroscopy) was prepared via overnight dialysis at 4 °C against exactly the same batch of buffer via a 1 kDa MWCO Tube-o-dialyzer (G-Biosciences), and the peptide solution was loaded into the injection syringe. The instrument's buret delivered 25 injections of 2.0 μL into the sample cell at a rate of one injection every 5 min. Titration data were minimally baseline-corrected and analyzed using a simple independent, two species binding model in Nanoanalyze software (from TA Instruments, see Supporting Information file). Reported error estimates for thermodynamic binding parameters are a convolution of concentration accuracies and variances for the fitted parameters from the nonlinear titration curve fits.

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

After each ITC titration was complete, the mixed sample of CaM and M13 peptide was retrieved from the sample cell via syringe and immediately concentrated to approximately 25−30 μL for IR spectroscopy (see below).

Infrared Spectroscopy. Approximately 6 μL of each sample was placed between two CaF_2 windows (13 mm × 2 mm) (Harrick) with a path length of 56 μm set by a Teflon spacer. All spectra were collected using a Vertex 70 FTIR spectrometer (Bruker Optics) with a photovoltaic MCT detector (Kolmar Technologies), with 512 scans at 2 cm^{-1} resolution. The appropriate buffer in the same cell with the identical spacer was used as the background.

All CaM samples were concentrated to approximately 1 mM CaM concentration (and a volume near 25 μL) using Sartorius Vivaspin centrifugal concentrators with a 3000 molecular weight cutoff. Concentrations were verified with UV spectroscopy using a 1 μL sample placed in the window of a 0.025 cm path length DMV Bio-Cell (Starna Cells). *apo*-CaM spectra were collected in 20 mM Hepes, 10 mM EDTA, pH 7.0, and both Ca^{2+}-saturated and M13-bound spectra were collected in 20 mM Hepes, 10 mM $CaCl_2$, pH 7.0. Post-ITC samples that showed clear saturation by M13 titration were concentrated and used for IR spectroscopy to guarantee that only one CaM species would contribute to each of the M13-bound spectra.

The baseline in the CN stretching region was corrected to highlight the CN peak because the combination band of water (which constitutes an absorptive background in this spectral region) shifts depending on solute and because the protein excludes solvent volume between the cell windows. All data between 2145 and 2185 cm^{-1} were ignored and the spectrum from 2100 to 2220 cm^{-1} was fit to an eighth-order polynomial, which was then subtracted from the complete spectrum between 2100 and 2220 cm^{-1}. The resulting line shapes were analyzed by determining the mode frequency by inspection, the mean frequency by averaging data between 2148 and 2175 cm^{-1}, the full width at half-maximum by inspection, and quantities associated with higher-order central moments (variance, skewness) by calculating the appropriate central moments over the same spectral region.

Far-UV Circular Dichroism. An AVIV model 410 spectropolarimeter was used to collect far-UV CD spectra for all the Ca^{2+}-saturated versions of each of the C*-containing CaM variants and wildtype CaM. Samples previously prepared in 20 mM Hepes, 10 mM $CaCl_2$, pH 7.0 for ITC, or IR experiments were diluted at least 10× with deionized water H_2O to 300 μL to ensure that the Hepes concentration was below 2 mM. All CD spectra were acquired at 25 °C from 180 to 260 nm in a 0.1 cm quartz cell, scanning at 20 nm/min with 2 scans at a bandwidth of 1 nm.

RESULTS AND DISCUSSION

Cyanylation of Cysteine in CaM Variants. Calmodulin is a compact, stably folded protein in all conditions: it is nearly impossible to denature to any significant extent except at very high denaturant concentrations (from which it refolds reversibly),[12] and it runs substantially smaller than its molecular weight on SDS-PAGE denaturing gels. To provide complete chemical access to the sites of interest, cyanylation of cysteine reactions was performed in 9 M urea solutions, which are greater than 50% urea by weight and thus favors the unfolded state of *apo*-CaM. The cyanylation chemistry originally developed by Degani and Patchornik[27] is compatible with denaturant solutions and mixed solvents, and despite the increased viscosity of the reaction solutions due to the high urea content, cyanylation was quantitative at all sites of interest according to both UV/visible spectroscopy during the first reaction step and infrared spectroscopy of concentrated samples.

Far-UV Circular Dichroism. CD spectra of both *apo* and Ca^{2+}-loaded CaM variants indicated that there was no significant change in the highly helical secondary structure of the protein due to any of the probe groups, as compared to the CD spectrum of the wild type protein (see Figures S1 and S2 in the Supporting Information for spectra). This indicates that at least the secondary structure of CaM was not significantly perturbed by any of the label placements. CD is only a useful measure of secondary structure and not tertiary or quaternary interactions: ITC was used as a more functional assay of the possible perturbations by the probe groups.

Isothermal Titration Calorimetry. Previous results for wild-type CaM binding to SCN-labeled versions of the M13 peptide[25] indicated only very small perturbation due to the probe groups on the target peptide. ITC titration curves for wild type and all seven probe-labeled CaM's are presented in Figure S2 (Supporting Information), and thermodynamic binding parameters from fitting these curves are shown in Table 1. The wild-type binding parameters are very slightly different from those reported earlier due to different buffer conditions.

Table 1. Fitted Parameters from the ITC Titration Curves for M13 Peptide Titrated into All CaM Variants[a]

variant	K_D (nM) (±50%)	n (±0.2)	ΔH (kJ/mol) (±20)	$\Delta S°$ (J/mol K)[b] (±30)
WT	100	1.0	−50	−40
M145C*	100	1.0	−70	−90
M109C*	100	0.9	−70	−90
M72C*	500	0.8	−50	−60
L105C*	400	0.7	−50	−40
I100C*	200	0.7	−40	5
V35C*	100	0.7	−50	−30
S17C*	100	1.1	−60	−70

[a]Estimated errors for each quantity are indicated in the column headers. [b]Calculated at 25 °C and based on standard concentration of 1 M.

Only small differences from the wild-type binding behavior were observed for all SCN probe sites: dissociation constants (K_D) for all of the M13-bound complexes were within a factor of 5 compared to those of unlabeled CaM, with four of the sites exhibiting no measurable perturbation of the dissociation equilibrium constant for the bound complex. Binding stoichiometries (n) were all reasonably close to 1:1 within our ability to measure peptide and CaM concentrations, and all signals were approximately the same strength leading to similar ΔH values. These fitted binding parameters, and the uniformly tight binding of CaM to this target peptide, indicate that the probe IR spectra collected for the M13-bound complex (see below) all display signals from only the bound complex and not the unbound, Ca^{2+}-saturated form of CaM.

The lack of measurable functional perturbation here is quite surprising, given that the C* side chain was substituted for a variety of different amino acids of varying shapes and sizes. It is especially provocative that probe groups placed at the three "key" methionine residues (M72C*, M109C*, and M145C*)

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

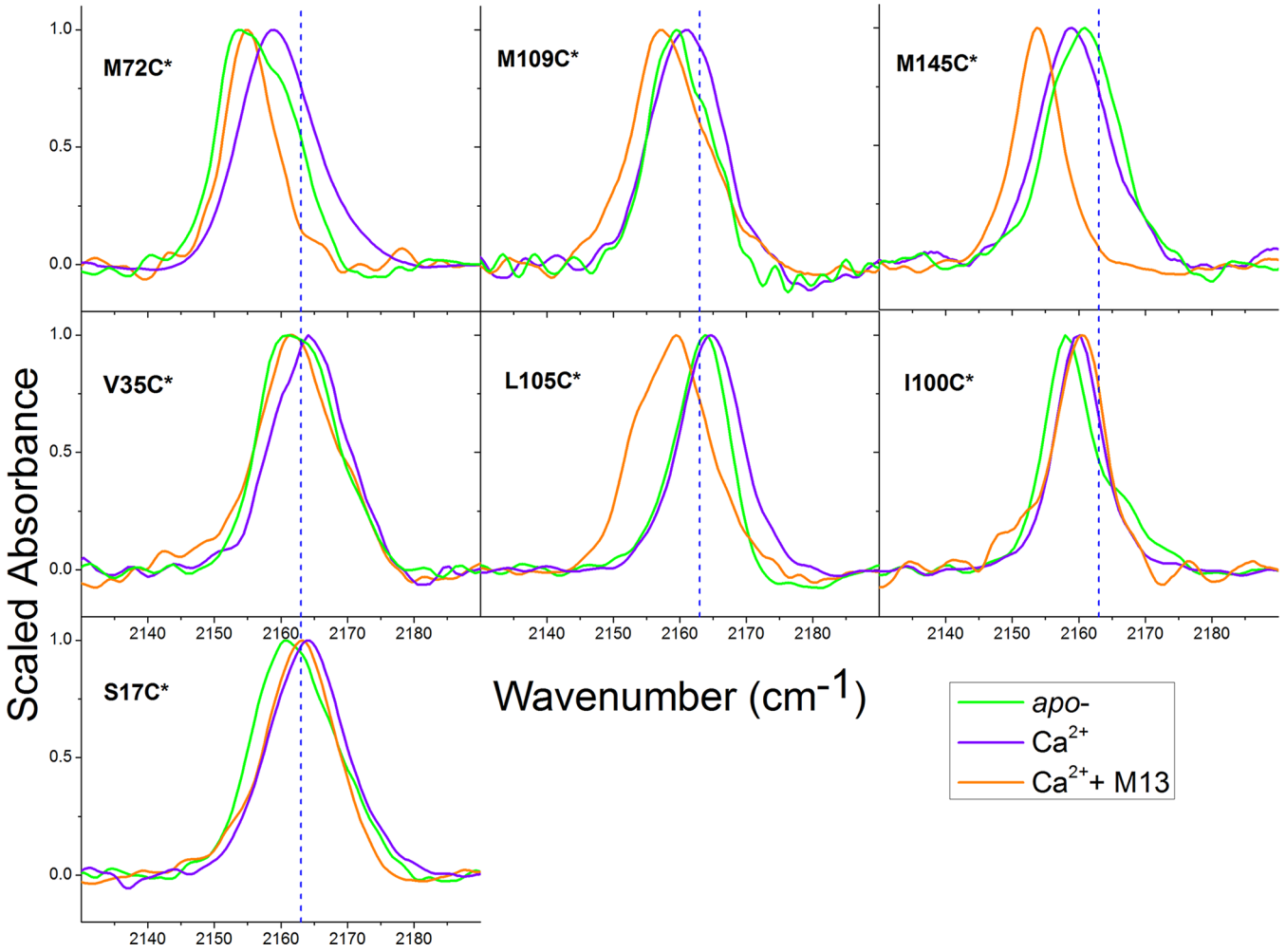


Figure 2. Infrared absorption spectra in the CN stretching region for C* probes placed at seven sites (indicated in frames) in CaM, in *apo*-conditions (green), in Ca^{2+}-saturating conditions (purple), and bound to the M13 peptide (orange). Dashed vertical lines appear at 2163 cm^{-1}, the mean frequency for a uniformly solvent-exposed SCN group.

for binding[37−40] did not lead to substantial perturbation of the CaM/M13 binding thermodynamics. These residues directly participate in the binding interaction, and the fact that the C* group at these sites does not perturb the binding thermodynamics (contrary to expectations based on other substitutions at these sites) is a strong indicator that C* is particularly well-suited to placement at binding interfaces between proteins. Despite its sterically distinct footprint and the artificial SCN functional group, C* appears to provide a reasonable steric mimic for native methionine residues. Substitutions at the other four sites were expected to be less perturbing, but given that CaM is essentially "perfectly evolved" for its function and exhibits very few evolutionary sequence variations across animal species,[41,42] we expected that any placement of artificial amino acids into CaM could perturb its functionality to some extent.

The ITC results summarized in Table 1 strongly indicate that, at least for this peptide target whose CaM-bound structure is a reasonable representative of a major binding mode of CaM, the C* group is an innocent reporter on its local environment even when placed directly in the functional protein−protein interface. ITC measurements are the most functional and directly relevant way to evaluate perturbation by the artificial probe groups in this system. Measurements in other systems meant to evaluate perturbation by nitrile probe groups have included CD melting curves of helical peptides,[33] thermal unfolding studies of a nitrile-labeled SH3 domain,[43] and an enzymatic activity assay that indirectly reported the functional formation of a Ras/GTPase adduct.[21,23] But most studies using nitrile vibrational labels have not directly measured the label-induced perturbations, and our ITC results suggest strongly that within certain limits, C* could be broadly used along otherwise poorly characterized binding interfaces between proteins without introducing substantial or functionally relevant structural perturbations.

This conclusion, combined with previous and similar data from labels on the peptide in the same complex,[25] is highly significant for future possible application of C* to examine protein−protein binding. In this study we followed our previous recommendations for where C* might reasonably be placed, refraining from editing any of CaM's many charged residues, and instead we focused directly on neutral (and mostly hydrophobic) amino acids of different shapes, sizes, and known structural environments. The success of our approach means that C* could be a widely useful probe of protein binding interfaces.

Infrared Spectroscopy. IR spectra for seven C* probe groups in three different forms of CaM (apo, Ca^{2+}-saturated, and bound to both Ca^{2+} and M13) are shown in Figure 2. At all

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

sites, there are obvious and sometimes dramatic changes in the CN frequency and line shape as the environment of CaM changes and its conformational ensemble readjusts. Empirically determined line shape parameters calculated without any assumptions about the underlying distributions for each probe CN band appear in Table S1 (Supporting Information).

The mean and mode CN frequencies are mostly lower than the 2163 cm^{-1} frequency that would be indicative of uniform solution exposure (dashed lines in Figure 2). The generally low observed frequency shifts are in accordance with the collapsed nature of CaM in all of its forms and the fact that the SCN probe groups are mainly located in environments that look more like the "interior" rather than the "exterior" of a protein. With that in mind, there is a very large range of mean and mode frequencies found in this data set compared to all previous data reported for the SCN group, and it is likely that the main physical drivers of these large frequency variations are local dispersion and exchange−repulsion interactions with closely neighboring pieces of the protein[32,44−46] rather than varying levels of solvent exposure (which reasonably explains the data from SCN labels on the M13 peptide[25]). Solvent is also expected to be a strong contributor to the CN frequency at some of these sites according to previously reported structures.

The IR CN stretching line shapes are especially heterogeneous and rich and provide a direct readout at these specific probe sites of the environments around the probe groups. This includes information about the dynamic conformational ensemble of CaM that cannot be directly reported by other methods. Previous reports of the long lifetime and slow dephasing dynamics of the SCN group[47,48] indicate that the CN line shapes observed are largely from an inhomogeneous frequency distribution that reports directly on the structural distribution. Several of the line shapes in Figure 1 (i.e., the M72C* and M109C* *apo* forms and the L105C* M13-bound complex) display obvious evidence of multiple spectral subpopulations, indicating that the probe groups in these forms of CaM appear in multiple different environments due to local steric variations, heterogeneity in the overall structure, or some combination of the two factors. In CaM where the secondary structure is largely invariant, the CN probe groups are expected to report on changes and heterogeneity/modularity of tertiary and quaternary contacts both across solution conditions and within a single protein ensemble.

The CN IR signals indicate that each site in CaM behaves very differently through the three conditions investigated here. The S17C* site shows the smallest changes across conditions, with frequencies consistent with water exposure and relatively symmetric line shapes, consistent with S17's relative distance from functional surfaces of the protein. The S17C* *apo* spectrum appears at slightly lower frequency, indicating perhaps that this probe site occupies an environment that is slightly more solvent-excluded in the collapsed *apo* globule than in the Ca^{2+}-bound forms of CaM. The I100C* site exhibits narrower CN bands that appear at substantially lower frequencies than for the S17C* probe groups: this suggests that most of the time the I100C* side chain is located in a structurally invariant environment which is red-shifted and thus could be more solvent-excluded in the *apo* form with the possibility of a lightly populated solvent-exposed or otherwise blue-shifted subpopulation. Both V35C* and L105C* display broad and shifting CN bands with interesting line shapes, and the changes at these two sites are very different as CaM goes from *apo* to binding Ca^{2+} ions to binding its target. Although V35C* displays more subtle changes from lower, to higher, back to lower frequencies in its relatively broad line shape, L105C* displays high frequencies until the target peptide is present, at which point the CN band shifts to much lower frequency and displays what appears to be two major spectral subpopulations in the M13-bound complex. The NMR structure of the CaM−M13 complex indicates that the V35C* probe group could be pointing directly at the bound peptide while the L105C* group is probably further away from the peptide, but the structural changes at each of these probe sites appear to affect the L105C* group much more.

Although the three methionine residues display qualitatively similar behaviors, there are substantial differences between the IR probe signals from these three binding-related sites. None of the probe groups appears to be completely solvent-exposed in any of the three conditions. In the *apo* structure, M72C* displays a low-frequency, bimodal line shape, whereas M109C* and M145C* probe groups have higher frequencies and relatively broad line shapes. The Ca^{2+}-saturated bands for these three sites are all very similar, with frequencies of about 2159 cm^{-1} and relatively symmetric line shapes. These probe spectra appear to display evidence of a rigidified structure whose binding interface is "ready" for target binding. When CaM binds M13, the probes at all three sites shift to much lower frequency, in a semiquantitatively similar way to what was observed for hydrophobic groups on the M13 peptide in previous work.[25] The M72C* and M145C* line shapes are very narrow in the M13-bound complex, indicating that these probes are in structurally very homogeneous environments. The M109C* band also shifts to much lower frequency upon binding to M13, but the line shape stays relatively broad and suggests that there is residual structural mobility at that site in the binding interface. Though it is known that the C-terminus of CaM binds more tightly to its wraparound targets than the N-terminus, the variations at these three methionine sites indicate that the bound structure appears most structurally heterogeneous at either end of the peptide and not necessarily in the middle.

The narrow M13-bound probe bands at M72C* and M145C* are much narrower than the signals observed previously for hydrophobic sites W4C* and F8C* on the M13 peptide, where the CN bands stayed quite broad perhaps due to the smallness of the C* residue compared to tryptophan or phenyalanine. C* is much more similar in size to methionine residues, so perhaps these narrow line shapes indicate that the C* residue "fits" better inside the available space when a similarly sized native residue inside is replaced by a relatively tight and rigid bound structure (that nonetheless still includes some substantial local flexibility and conformational entropy). These narrow bands are similar to the IR band for the very structurally constrained SCN-homoalanine residue in the Boxer group's labeled and reconstituted RNase complex,[49] and we interpret the narrowness of the M72C* and M145C* sites as reporting a rigid and invariant local environment around the probe group. The M109C* residue is apparently in a more heterogeneous range of local environments in the M13-bound complex according to the data in Figure 2.

General Discussion. All 21 of the spectra in Figure 2 are very rich and contain new information about the conformational ensembles of CaM in these three conditions. Due to the extremely fast time scale of IR spectroscopy compared to protein structural fluctuations, and in the case of this particular probe group the long lifetime and slow dephasing of the SCN signal,[47,48] the probe line shapes directly reflect the conforma-

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

tional ensembles in a way that NMR line shapes cannot. However, the best possible current interpretation of these signals is semiquantitative and empirical. Many of the mean and mode frequencies observed here reflect details that would be expected on the basis of the reported crystal or average NMR structures of CaM in these three conditions, but some of the probe frequencies, and all of the line shapes, contain new information that does not appear in PDB-deposited structures. A truly quantitative interpretation of the data in Figure 2 would involve comparing the data to well-sampled structural ensembles, likely generated by all-atom and explicit-solvent molecular dynamics calculations. Such ensembles have been reported for CaM,[16] and we and others are currently building the computational machinery to make direct comparisons between our vibrational probe data and those ensembles via explicit inclusion of the C* probe group and calculating its vibrational frequency in carefully sampled members of the CaM ensembles. One likely issue in comparing our spectral line shapes to simulated ensembles is that many all-atom force fields have known biases toward more collapsed and highly structured conformations; the recent release of CHARMM36 was designed in part to counter that bias.[50]

Because CaM always has a high degree of secondary structure and is thus less "disordered" than many other proteins whose conformational ensembles are of biological interest, there are still interactions between functional groups whose parametrizations might not provide an accurate sampling of the overall structural ensemble. When a more direct and systematic connection between experiment and simulation has been forged, perhaps using the data here as a test set, it could become possible to use the direct ensemble information in vibrational probe spectra to evaluate and readjust classical force fields to accurately predict other structural ensembles. The high quality and richness of the data in Figure 2 suggest that there is a great deal of biologically relevant information about CaM's context-dependent conformational distributions that can be extracted from vibrational probe data, which can be collected at a relatively low cost and without functional perturbation of the structure (as we showed above using ITC). The next challenge is to establish a systematic and direct connection between simulated ensembles and experiments like those performed here.

One thing that is possible through a semiquantitative interpretation of data such as that in Figure 2, especially through the differences between Ca^{2+}-saturated and target-bound CaM, is a first-order diagnosis of the binding mode of CaM. Targets that bind to CaM via the wraparound binding mode should induce large shifts and characteristic changes in the IR probe spectra of probes at the key methionine residues M72C*, M109C*, and M145C*. However, targets that bind in alternate, nonwraparound configurations should not introduce the same changes and might induce different changes at the other sites investigated here (or at other sites not probe-labeled in this work). The nonperturbativity of these probes on CaM suggests that, avoiding changes to the charged residues, IR spectra of a few probe sites on CaM could be used to provide a preliminary estimate of the binding mode of CaM to some of its hundreds of uncharacterized targets. These probe spectra, which can be acquired with very few sampling limitations due to size and/or complexity of protein targets, could also be used to determine whether the binding mode changes when CaM binds to whole proteins rather than just the binding peptide sequence for a known CaM regulatory target. Future work with the probe-labeled variant CaM's produced here will also focus on CaM targets that are difficult to characterize by more conventional structural biology techniques.

CONCLUSIONS

SCN vibrational probe groups substituted for seven different native residues on CaM do not perturb the protein's secondary structure (according to CD spectra) and only minimally change the binding thermodynamics for the formation of a complex between CaM and the M13 target peptide from skMLCK. This lack of perturbation, even when the probe side chain was substituted for key methionine residues along the hydrophobic binding interface, suggests that this probe strategy might be used quite broadly in many protein−protein binding interfaces.

Infrared spectra of the SCN probe groups in three different forms of CaM (without calcium, saturated with calcium ions, and bound to the M13 target peptide) display a wide range of frequencies and line shapes that report directly on the local structural distribution around the probe groups, and each site exhibits different behavior in sometimes dramatic ways. Although some of the CN frequencies are consistent with the degree of solvent exposure expected from reported structures, the frequencies and line shapes contain a great deal of new information about the dynamic conformational ensembles of calmodulin in different conditions. A more direct link between simulated ensembles and experiment is needed before a fully quantitative interpretation of this rich data is possible. Constitution of such a direct link would allow evaluation of conformational ensembles directly from vibrational probe data. The infrared probe sites on CaM reported here can also be used to determine the possible binding modes of calmodulin with novel or otherwise uncharacterized targets, from peptides to full proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.8b00475.

Far-UV circular dichroism spectra, isothermal titration calorimetry data and fits, and line shape analysis of infrared spectra (PDF)

AUTHOR INFORMATION

Corresponding Author

*C. H. Londergan. E-mail: clonderg@haverford.edu.

ORCID

Sara Linse: 0000-0001-9629-7109

Casey H. Londergan: 0000-0002-5257-559X

Present Address

[+]CSL Behring, King of Prussia, PA.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Dr. Karin S. Akerfeldt (Haverford College) is acknowledged for intellectual contributions and for assistance in synthesis of the M13 peptide. Bashkim Kokona (Haverford College) provided help with peptide synthesis, HPLC purification, and mass spectroscopic analysis of the M13 peptide. This work was funded by grants R15-GM088749-01 and R15-GM088749-02 from the National Institute of General Medical Sciences and by

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

CAREER grant CHE-1150727 from the National Science Foundation.

REFERENCES

(1) Ma, J.; Pazos, I. M.; Zhang, W.; Culik, R. M.; Gai, F. Site-Specific Infrared Probes of Proteins. *Annu. Rev. Phys. Chem.* **2015**, *66* (1), 357−377.

(2) Adhikary, R.; Zimmermann, J.; Romesberg, F. E. Transparent Window Vibrational Probes for the Characterization of Proteins With High Structural and Temporal Resolution. *Chem. Rev.* **2017**, *117* (3), 1927−1969.

(3) Kim, H.; Cho, M. Infrared Probes for Studying the Structure and Dynamics of Biomolecules. *Chem. Rev.* **2013**, *113* (8), 5817−5847.

(4) Yap, K. L.; Ames, J. B.; Swindells, M. B.; Ikura, M. Diversity of Conformational States and Changes within the EF-Hand Protein Superfamily. Proteins-Struct. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 499−507.

(5) Vetter, S. W.; Leclerc, E. Novel Aspects of Calmodulin Target Recognition and Activation. *Eur. J. Biochem.* **2003**, *270*, 404−414.

(6) Hoeflich, K. P.; Ikura, M. Calmodulin in Action. *Cell* **2002**, *108* (6), 739−742.

(7) Radivojac, P.; Vucetic, S.; O'Connor, T. R.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. Calmodulin Signaling: Analysis and Prediction of a Disorder-Dependent Molecular Recognition. Proteins-Struct. *Proteins: Struct., Funct., Genet.* **2006**, *63*, 398−410.

(8) Ishida, H.; Vogel, H. J. Protein-Peptide Interaction Studies Demonstrate the Versatility of Calmodulin Target Protein Binding. *Protein Pept. Lett.* **2006**, *13*, 455−465.

(9) Shuman, C. F.; Jiji, R.; Akerfeldt, K. S.; Linse, S. Reconstitution of Calmodulin from Domains and Subdomains: Influence of Target Peptide. *J. Mol. Biol.* **2006**, *358*, 870−881.

(10) Finn, B. E.; Evenas, J.; Drakenberg, T.; Waltho, J. P.; Thulin, E.; Forsen, S. Calcium-Induced Structural Changes and Domain Autonomy in Calmodulin. *Nat. Struct. Mol. Biol.* **1995**, *2* (9), 777−783.

(11) Feng, Y. D.; Wei, Q. Diversity of the Modes for Calmodulin Binding to Its Targets. *Prog. Chem.* **2012**, *24* (10), 2028−2039.

(12) Guerini, D.; Krebs, J. Influence of Temperature and Denaturing Agents on the Structural Stability of Calmodulin: A 1H-Nuclear Magnetic Resonance Study. *FEBS Lett.* **1983**, *164* (1), 105−110.

(13) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. Solution Structure of Calcium-Free Calmodulin. *Nat. Struct. Mol. Biol.* **1995**, *2* (9), 768−776.

(14) Chou, J. J.; Li, S. P.; Klee, C. B.; Bax, A. Solution Structure of Ca2+-Calmodulin Reveals Flexible Hand-like Properties of Its Domains. *Nat. Struct. Biol.* **2001**, *8* (11), 990−997.

(15) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin Structure Refined at 1.7 Å Resolution. *J. Mol. Biol.* **1992**, *228* (4), 1177−1192.

(16) Gsponer, J.; Christodoulou, J.; Cavalli, A.; Bui, J. M.; Richter, B.; Dobson, C. M.; Vendruscolo, M. A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction. *Structure* **2008**, *16* (5), 736−746.

(17) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. Solutions Structure of a Calmodulin-Target Peptide Complex by Multidimensional NMR. *Science* **1992**, *256*, 632−638.

(18) Drum, C. L.; Yan, S. Z.; Bard, J.; Shen, Y. Q.; Lu, D.; Soelaiman, S.; Grabarek, Z.; Bohm, A.; Tang, W. J. Structural Basis for the Activation of Anthrax Adenylyl Cyclase Exotoxin by Calmodulin. *Nature* **2002**, *415* (6870), 396−402.

(19) O'Connell, D. J.; Bauer, M. C.; O'Brien, J.; Johnson, W. M.; Divizio, C. A.; O'Kane, S. L.; Berggard, T.; Merino, A.; Akerfeldt, K. S.; Linse, S.; et al. Integrated Protein Array Screening and High Throughput Validation of 70 Novel Neural Calmodulin-Binding Proteins. *Mol. Cell. Proteomics* **2010**, *9*, 1118−1132.

(20) Lipstein, N.; Göth, M.; Piotrowski, C.; Pagel, K.; Sinz, A.; Jahn, O. Presynaptic Calmodulin Targets: Lessons from Structural Proteomics. *Expert Rev. Proteomics* **2017**, *14* (3), 223−242.

(21) Stafford, A. J.; Ensign, D. L.; Webb, L. J. Vibrational Stark Effect Spectroscopy at the Interface of Ras and Rap1A Bound to the Ras Binding Domain of RalGDS Reveals an Electrostatic Mechanism for Protein-Protein Interaction. *J. Phys. Chem. B* **2010**, *114* (46), 15331−15344.

(22) Ritchie, A. W.; Webb, L. J. Understanding and Manipulating Electrostatic Fields at the Protein-Protein Interface Using Vibrational Spectroscopy and Continuum Electrostatics Calculations. *J. Phys. Chem. B* **2015**, *119* (44), 13945−13957.

(23) Stafford, A. J.; Walker, D. M.; Webb, L. J. Electrostatic Effects of Mutations of Ras Glutamine 61 Measured Using Vibrational Spectroscopy of a Thiocyanate Probe. *Biochemistry* **2012**, *51*, 2757−2767.

(24) Stafford, A. J.; Walker, D. M.; Webb, L. J. Electrostatic Effects of Mutations of Ras Glutamine 61 Measured Using Vibrational Spectroscopy of a Thiocyanate Probe. *Biochemistry* **2012**, *51* (13), 2757−2767.

(25) Dalton, S. R.; Vienneau, A. R.; Burstein, S. R.; Xu, R. J.; Linse, S.; Londergan, C. H. Cyanylated Cysteine Reports on Protein-Protein Binding Interfaces without Strong Perturbation, submitted for publication.

(26) Degani, Y.; Patchornik, A. Cyanylation of Sulfhydryl Groups by 2-nitro-5-thiobenzoic acid- High Yield Modification and Cleavage of Peptides at Cysteine Residues. *Biochemistry* **1974**, *13*, 1−11.

(27) Patchornik, A.; Degani, Y.; Neumann, H. Selective cyanylation of sulfhydryl groups. *J. Am. Chem. Soc.* **1970**, *92* (23), 6969−6970.

(28) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. Site-Specific Conversion of Cysteine Thiols into Thiocyanate Creates an IR Probe for Electric Fields in Proteins. *J. Am. Chem. Soc.* **2006**, *128*, 13356−13357.

(29) Johnson, M. N. R.; Londergan, C. H.; Charkoudian, L. K. Probing the Phosphopantetheine Arm Conformations of Acyl Carrier Proteins Using Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2014**, *136* (32), 11240−11243.

(30) Maienschein-Cline, M. G.; Londergan, C. H. The CN Stretching Band of Aliphatic Thiocyanate Is Sensitive to Solvent Dynamics and Specific Solvation. *J. Phys. Chem. A* **2007**, *111*, 10020−10025.

(31) Oh, K.; Choi, J.; Lee, J.; Han, J.; Lee, H.; Cho, M. Nitrile and Thiocyanate IR Probes: Molecular Dynamics Simulation Studies. *J. Chem. Phys.* **2008**, *128* (15), 154504.

(32) Blasiak, B.; Ritchie, A. W.; Webb, L. J.; Cho, M. Vibrational Solvatochromism of Nitrile Infrared Probes: Beyond the Vibrational Stark Dipole Approach. *Phys. Chem. Chem. Phys.* **2016**, *18* (27), 18094−18111.

(33) Edelstein, L.; Stetz, M. A.; McMahon, H. A.; Londergan, C. H. The Effects of Alpha-Helical Structure and Cyanylated Cysteine on Each Other. *J. Phys. Chem. B* **2010**, *114*, 4931−4936.

(34) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational Entropy in Molecular Recognition by Proteins. *Nature* **2007**, *448* (7151), 325−U3.

(35) Carey, J.; Lindman, S.; Bauer, M.; Linse, S. Protein Reconstitution and Three-Dimensional Domain Swapping: Benefits and Constraints of Covalency. *Protein Sci.* **2007**, *16* (11), 2317−2333.

(36) Riddles, P. W.; Blakeley, R. L.; Zerner, B. Reassessment of Ellman Reagent Methods Enzymol. *Methods Enzymol.* **1983**, *91*, 49−60.

(37) Vougier, S.; Mary, J.; Dautin, N.; Vinh, J.; Friguet, B.; Ladant, D. Essential Role of Methionine Residues in Calmodulin Binding to Bordetella Pertussis Adenylate Cyclase, as Probed by Selective Oxidation and Repair by the Peptide Methionine Sulfoxide Reductases. *J. Biol. Chem.* **2004**, *279* (29), 30210−30218.

(38) Yamniuk, A. P.; Ishida, H.; Lippert, D.; Vogel, H. J. Thermodynamic Effects of Noncoded and Coded Methionine Substitutions in Calmodulin. *Biophys. J.* **2009**, *96* (4), 1495−1507.

(39) Yuan, T.; Vogel, H. J. Substitution of the Methionine Residues of Calmodulin with the Unnatural Amino Acid Analogs Ethionine and Norleucine: Biochemical and Spectroscopic Studies. *Protein Sci.* **1999**, *8* (1), 113−121.

(40) Yuan, T.; Ouyang, H.; Vogel, H. J. Surface Exposure of the Methionine Side Chains of Calmodulin in Solution: a Nitroxide Spin

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955

Label and Two-Dimensional NMR Study. *J. Biol. Chem.* **1999**, *274* (13), 8411−8420.
(41) Friedberg, F.; Rhoads, A. R. Evolutionary Aspects of Calmodulin. *IUBMB Life* **2001**, *51* (4), 215−221.
(42) Nojima, H. Molecular Evolution of the Calmodulin Gene. *FEBS Lett.* **1987**, *217* (2), 187−190.
(43) Adhikary, R.; Zimmermann, J.; Dawson, P. E.; Romesberg, F. E. IR Probes of Protein Microenvironments: Utility and Potential for Perturbation. *ChemPhysChem* **2014**, *15* (5), 849−853.
(44) Blasiak, B.; Lee, H.; Cho, M. Vibrational Solvatochromism: Towards Systematic Approach to Modeling Solvation Phenomena. *J. Chem. Phys.* **2013**, *139* (4), 044111.
(45) Blasiak, B.; Cho, M. Vibrational Solvatochromism. II. A First-Principle Theory of Solvation-Induced Vibrational Frequency Shift Based on Effective Fragment Potential Method. *J. Chem. Phys.* **2014**, *140* (16), 164107.
(46) Blasiak, B.; Cho, M. Vibrational Solvatochromism. III. Rigorous Treatment of the Dispersion Interaction Contribution. *J. Chem. Phys.* **2015**, *143* (16), 164111.
(47) van Wilderen, L. J. G. W.; Kern-Michler, D.; Mueller-Werkmeister, H. M.; Bredenbeck, J. Vibrational Dynamics and Solvatochromism of the Label SCN in Various Solvents and Hemoglobin by Time Dependent IR and 2D-IR Spectroscopy. *Phys. Chem. Chem. Phys.* **2014**, *16* (36), 19643−19653.
(48) Levin, D. E.; Schmitz, A. J.; Hines, S. M.; Hines, K. J.; Tucker, M. J.; Brewer, S. H.; Fenlon, E. E. Synthesis and Evaluation of the Sensitivity and Vibrational Lifetimes of Thiocyanate and Selenocyanate Infrared Reporters. *RSC Adv.* **2016**, *6* (43), 36231−36237.
(49) Fafarman, A. T.; Boxer, S. G. Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease S. *J. Phys. Chem. B* **2010**, *114*, 13536−13544.
(50) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14* (1), 71−73.

DOI: 10.1021/acs.jpca.8b00475
J. Phys. Chem. A 2018, 122, 2947−2955