Label-Free Optical Detection of Protein Acetylation using *UV-Vis* Charge Transfer Spectra

Himanshi Maniram Devi [a#], Apoorva Badaya[b#], Arijit Maity [b], Simangka Bor Saikia [a], Ravindra Venkatramani[b] and Rajaram Swaminathan*[a]*

[a]Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

[b]Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India

[#]Equal Contributions

**Corresponding Authors ravi.venkatramani@tifr.res.in, rsw@iitg.ac.in*

Table of Contents

M1 Details of Experimental and Theoretical Methods

M1.1 Experimental methods:

M1.1.1: *Reaction condition for chemical acetylation of α_3C and α_3W*

Dialysis: After completion of the reaction, all the samples were dialyzed against deionised water (resistivity 18.2 MΩ.cm) to remove the remaining acetic anhydride and related impurities. The dialysis was performed for 4, 8, 12, and 24 hours at 4 °C. Subsequently, the dialyzed samples were checked by UV-Visible spectrophotometer and spectrofluorometer for presence of acetic anhydride and associated products in terms of its absorbance which extends slightly over 270 nm and emission which has a maximum around 340 nm. After 24 hours, both UV-Visible and fluorescence spectra showed minimal absorbance and emission, respectively from acetic anhydride and associated products (**Figure F3)**. Thus, all the samples after the reaction were dialyzed for 24 hours with frequent exchange of deionised water during the process to ensure complete removal of excess acetic anhydride or byproducts.

Centrifugation: After 24 hours of dialysis, all the samples were centrifuged at 12,000 rpm for 10 minutes at 4 °C, to remove any particles that may arise from precipitation and contribute to scattering. All measurements on the proteins post-acetylation were performed after the above steps of dialysis and centrifugation.

M1.1.2 *Steady State Luminescence*

Quantum Yield calculation: The quantum yields of native and acetylated proteins were measured at 280 and 355 nm in deionised water. Additionally, we calculated the relative quantum yield of the samples with respect to well-characterized references: N-acetyl-L-tryptophan amide (NATA) for 280 nm excitation and 9,10-diphenyl anthracene (DPA) for 355 nm excitation using:

$$\Phi_f^i = \frac{F^i f_s n_i^2}{F^s f_i n_s^2} \Phi_f^s \qquad 1$$

Here, Φ, *F, f, n* denote the quantum yield, integrated luminescence, Absorptance = $(1\text{-}10^{-Ax}$, A=absorbance) and refractive index, respectively. The superscript and subscript *s* and *i* for various symbols denote the parameters for the reference and sample, respectively[1]. Absorbance for all the protein samples and NATA was recorded in deionised water, while the DPA was recorded in cyclohexane. Excitation at 280 nm was excluded for α_3W as the protein contains Trp which can mask the luminescence due to strong emission between 300-500 nm. All the samples were excited with a slit width of 2 nm and emission spectra were recorded from 300-500 nm (for 280 excitation) and 370-650 nm (for 355 nm excitation) with a slit width of 15 nm. The blank was subtracted manually to avoid contamination by the solvent. QY calculations were adjusted to account for differences in the refractive indices of cyclohexane and deionised water. All the samples were recorded at room temperature (25 °C).

Tryptophan Fluorescence anisotropy: Anisotropy of native and acetylated samples were recorded with slit widths of 2 nm for excitation and 15 nm for emission, respectively. All the

samples were excited at 280 nm and emission was collected at their respective maxima and run in triplicate at room temperature. All the recorded data were the average of 10 individual readings; the samples were G-factor corrected. Steady State Fluorescence Anisotropy (r) is defined as follows:

$$r = \frac{(I_{VV} - GI_{VH})}{(I_{VV} + 2GI_{VH})} \quad 2$$

Here, I_{VH} and I_{VV} are emission intensities when the excitation polarizer is oriented vertically, while the emission polarizer is oriented horizontally and vertically, respectively.

Time-resolved fluorescence: The tryptophan fluorescence decay of native and acetylated α_3W were recorded by using Time-correlated single-photon counting instrument equipped with MCP detection (Horiba Jobin, Model: Ultrafast-01-DD). The decay analysis was performed to track the residual amount of acetic anhydride left after the acetylation reaction, dialysis was performed, and intensity decay was recorded for three different time durations (4, 12, and 24 hours). All the readings were performed in duplicates at room temperature and the absorbance for each sample was restricted to ≤ 0.08 to avoid the inner filter effects. The fluorescence of NATA was also recorded under the same condition as a control. The IRF was acquired using a chalk powder colloidal suspension[2]. The decay count recorded for all the samples was 1000, and the data was analyzed using DAS (Decay Analysis Software) software. The observed decay profile was analyzed by comparing it with the convolution of the chosen decay model (Equation 3) and the measured IRF.

$$I(t) = \sum_{i=1}^{2} \alpha_i exp\left(\frac{-t}{\tau_i}\right) \quad 3$$

M1.2 Computational methods:

We developed a general computational framework to simulate and predict the ProCharTS spectra of any protein with known 3D structure. The computational workflow is provided in **Figure F4.** Briefly, classical atomistic MD equilibration and production runs are carried out on the solvated protein to generate an equilibrium statistical ensemble with Boltzmann statistics. A cluster analysis based on a distance cut-off of amino acids is carried out on each structure derived from the MD statistical ensemble. Although we focus here only on charged amino acids, other amino acids of interest such as aromatic residues can also be included. The analysis yields a set of ProCharTS chromophores in the form of clusters (C_i) of i interacting amino acids for every structure in the ensemble. TDDFT calculations are carried out on the ProCharTS chromophores and their contributions added up to generate the protein spectra. Finally, the spectra is averaged over the MD statistical ensemble to compute the molar absorptivity for the proteins. Each step is explained in detail below where we apply the procedure to generate the ProCharTS profile of the two model proteins in native (α_3C, α_3W) and acetylated (Acn-α_3C, Acn-α_3W, where n denotes the number of lysines acetylated) forms. Here, initial protein structures for MD simulations were obtained from the RCSB Protein Data Bank (PDB)[3].

Clusters of charged amino acids were extracted from MD trajectories using VMD (Visual molecular dynamics)[4], and absorption spectra over a 380-800 nm wavelength range were generated using TDDFT calculations in Gaussian09[5].

M1.2.1 Generation of Protein Models

The starting structures for MD simulations for α_3C and α_3W proteins (PDB id: 2LXY and 1LQ7 respectively) were derived from NMR studies [6]. The proteins comprise of three helices-*H1*, *H2*, and *H3* (**Figure 1A**) and the 67 amino acid sequence is rich in charged amino acids (17 Lys, 17 GLU, 2 arginines (Arg). The proteins contain either a single Cys (α_3C) or Trp (α_3W), at position 34. To generate acetylated variants of α_3C and α_3W proteins, a custom VMD Tcl script was employed to mutate lysine (Lys) residues to acetyl-Lys (Aly) in the PDB structures obtained from the Protein Data Bank (PDB) (**Figure 1B**). The psfgen package in VMD was used to perform the mutations, incorporating the CHARMM36 force field. A total of 6 models for α_3C were generated: native, Ac1, Ac3, Ac5, Ac12 and Ac17 (Acn: n represents the number of acetylated Lys residues in the system). These systems were chosen to match the protein states obtained for different concentrations of acetic anhydride as verified by mass spectrometry (**Figure 1C-N**). To that end, relative solvent accessible surface area (rSASA) of each Lys residue in native α_3C protein were calculated for each structure and averaged over the trajectory as:

$$\mathrm{r}SASA(\%) = \frac{SASA_{Protein}}{SASA_{isolated}} \times 100 \qquad 4$$

Here $SASA_{protein}$ and $SASA_{isolated}$ is the solvent accessibility of the residue in the presence and absence of the rest of the protein. Lys residues were progressively acetylated in decreasing order of their rSASA value (**Figure F5A**). The Lys residues that we chose to acetylate for each model system are shown in **Figure F5(B-E)**. For α_3W we generated only the native and fully acetylated models (α_3W and Ac17-α_3W). In all cases including the fully acetylated case, the N-terminus was not acetylated in simulations. To summarize a total of eight native and acetylated protein systems were generated for MD simulations.

M1.2.2 Molecular Dynamics Simulations

Fully atomistic MD simulations on the protein models were carried under periodic boundary conditions out using Gromacs-5.1.7 [7–9] and the CHARMM36 force field [10]. Each of the eight protein models were solvated with explicit water molecules (TIP3P water model [11] and neutralized by adding suitable number of Na^+ or Cl^- ions. During simulations, electrostatic interactions were calculated using the Particle Mesh Ewald (PME) algorithm [12] with a Coulomb cutoff of 1.2 nm. Non-bonded van der Waals interactions were described by the Lennard-Jones potential with a switching function and a cutoff of 1 nm. Each protein system was first subjected to energy minimization to resolve steric clashes and unrealistic geometries using the steepest descent optimization algorithm, with a force tolerance of 500 kJ mol^{-1} nm^{-1}. All protein and

solvent atoms were allowed to relax during this process. The systems were then equilibrated for 1 ns under NVT (constant volume and temperature) conditions followed by 1 ns under NPT (constant pressure and temperature) conditions. During these steps, all protein heavy atoms were restrained to their positions determined by NMR using a harmonic spring constant of 1000 kJ mol^{-1} nm^{-2}. Temperature was controlled using the velocity-rescale thermostat [13] with a bath coupling constant of 0.1 ps. Pressure control was achieved using the Parrinello-Rahman barostat [14] with a coupling constant of 2 ps and compressibility of 4.5 x10^{-5} bar^{-1}. The LINCS algorithm [15] was used to constrain the H-bonds. Following this, unrestrained 100 ns MD trajectories were generated with a time step of 2 fs with the last 50 ns being considered as the production phase. As noted below in the next subsection, this protocol produces stable, locally equilibrated trajectories which sample protein conformations with Boltzmann population statistics. The positions and the velocities of the atoms were saved every 2 ps during the production stage yielding a total of 25,000 snapshots for each system which were taken forward for further analysis. System details including box dimensions, number of charged residues modified and counterions are summarized in **Table T3**.

M1.2.3 Protein Stability and Equilibration in MD Trajectories

To assess the overall conformational stability and equilibration, the CVCF (Cumulative Variance of Atomic Coordinate Fluctuations) analysis was performed on the protein backbone atoms of the structured (helical) residue segments in MD trajectories. The CVCF for a set of N atoms is defined as [16]

$$\sigma^2_{CVCF}(T) = \frac{1}{N_t} \sum_{t=1}^{N_t} \sum_{i=1}^{3N} (r_i(t) - \langle r_i \rangle)^2 \qquad 5$$

where, N_t is the total number of MD frames/snapshots extracted up to time T along the MD trajectory and counted by the index *t*. The coordinates $r_i(t)$ and $\langle r_i \rangle$ are the atomic coordinates at snapshot t and its average over N_t frames. Here the index *i* runs over the *x, y, z*- coordinates of all N atoms. Prior to evaluating the variance in Equation 4, the snapshots should be aligned relative to the positions of N atoms of a reference structure to eliminate variance contributions due to rigid body translation and rotation. The features of the CVCF trace over time provide insights into the sampling of the underlying energy landscape by the MD simulations: flat (plateau) regions indicate locally equilibrated sections, rising traces suggest the discovery of new energy minima, and decreasing traces indicate convergence towards local equilibration. Here, the term local equilibration indicates that Boltzmann sampling has been achieved for the region of the energy landscape sampled by the trajectory so that meaningful averaging of observables can be carried out over the entire trajectory segment [7].

We carried out CVCF-trace analysis on the 50 ns production MD trajectory segments of α_3C, Ac*n*-α_3C (with *n* = 1, 3,12, and 17 Lys acetylated), α_3W and Ac17-α_3W. The positions of the backbone atoms during the simulation were aligned with respect to the initial structure (at *t = 1*) of 50 ns production run. The data in **Figure F6** shows that all trajectories with the exception

of Ac3-α_3C are locally equilibrated at 50 ns. The Ac3-α_3C trajectory also achieves local equilibration within 25 ns as noted by the plateau in the CVCF trace for the 50 ns production segment in **Figure F6**. However, in the last 25 ns the local equilibrium is disturbed as the system transitions through a metastable state as indicated by the rising CVCF trace. Thus, for this system we consider only the first 25 ns locally equilibrated production segment for further analysis. In order to quantify the stability of each system, linear regression was carried out on the CVCF traces using all data points from the last 5 ns of each equilibrated trajectory segment to determine the best-fit slope. This analysis shows that in the last 5 ns, CVCF values converge to a plateau with slopes ranging between -0.0006 and 0.0013 $Å^2$/ns (**Figure F6**).

M1.2.4 Cut-off distance used in clustering of charged residues:

The cutoff distance (R_c) for electronic coupling used to define clusters is a critical parameter as it determines the size distribution of independent clusters used in our spectra calculations. Ideally, R_c should be taken as large as possible to obtain distinct electronically uncoupled clusters for which the additivity relation in Equation 6 holds. However, for large R_c the cluster sizes approach the size of the protein itself making the spectra calculations prohibitively expensive. Thus, we carried out control calculations to determine an optimal R_c which balances computational cost and accuracy. Specifically, we simulated the spectra for a sample decamer cluster of ten charged residues (The largest cluster which we could compute using TDDFT at a reasonable cost) extracted from the MD trajectory of α_3C protein as a function of R_c (**Figure F10(a)**). The decamer breaks down into multiple smaller clusters as we decrease the cut-off distances from 8-5Å (**Figure F10**). A comparison of the acetylation induced spectral changes at different R_c values shows that R_c = 6Å is sufficient to capture the spectral changes induced by acetylation of the model cluster as determined by a calculation on the full decamer.

M1.2.5 Site-Specific Sensitivity of ProCharTS to Lysine Acetylation in α_3C

To evaluate the sensitivity of ProCharTS to the choice of acetylation sites, three Ac12 variants of the α_3C system were constructed. Each variant contained acetylation 12 of the 17 Lys residues. While the standard strategy employed in the manuscript (e.g Figure 4 data) chooses these 12 residues (Ac12-v1) in terms of their solvent exposure (residues with the top 12 SASA), we also employed two additional variants (Ac12-v2 and Ac12-v3) swapping one of the acetylation sites. The sites which were swap were chosen by examining the contribution of each Lys residue to the overall ProCharTS spectrum of native α_3C protein. As shown in Fig. M1A, residue-wise spectral decomposition over 50 frames reveals substantial heterogeneity in participation of different Lys residues. Notably, Lys32 and Lys39 exhibit strong contributions to the spectra, whereas Lys10 contributes minimally. Further, an analysis of Lys–Glu distances in the MD trajectory of native α_3C (**Figure M1B**) demonstrates that Lys32 participates in a significantly larger number of interactions (4–6 Å) range compared to Lys10. Please note, these Lys–Glu contacts are part of different clusters ranging from dimers to decamers. Similarly, trimer interaction analysis (**Figure M1C**) indicates prominent involvement of Lys39 interacting with Glu in trimers in the MD trajectories. To summarize acetylation at Lys 32 or

Lys 39 should have more of an impact on the ProCharTS profile relative to acetylation at Lys 10.

Based on these observations, we constructed three variants of α_3C, Ac12-α_3C-v1 (Lys10, Aly32, Aly39), Ac12-α_3C-v2 (Aly10, Lys32, Aly39), and Ac12-α_3C-v3 (Aly10, Aly32, Lys39). Thus, each variant preserves one spectrally important (or unimportant) Lys residue while acetylating the other two, enabling a direct test of site-specific spectral sensitivity. Note that while we swap the acetylation states of the three residues mentioned above, the remaining 9 acetylated Lys sites remain unchanged according to their SASA values. ProCharTS spectra were computed for 50 representative frames for each variant using the same protocol applied to native α_3C and its mutants. The resulting spectra (**Figure F18B**) reveals a clear distinction between the variants with Ac12-v1 exhibiting the lowest spectral intensity and Ac12-v3 showing the strongest signal. Notably, the spectra of Ac12-v2 and Ac12-v3 exceed that of Ac17, consistent with experimental observations (**Figure F18C-E**).These results demonstrate that the spectral response is highly sensitive to the identity of the acetylated lysine residues. This analysis establishes that ProCharTS is not only sensitive to the extent of acetylation but also to the precise site of modification. Even substitution of a single lysine residue in the acetylated set measurably alters the spectral profile, highlighting the residue-level resolution of the method.

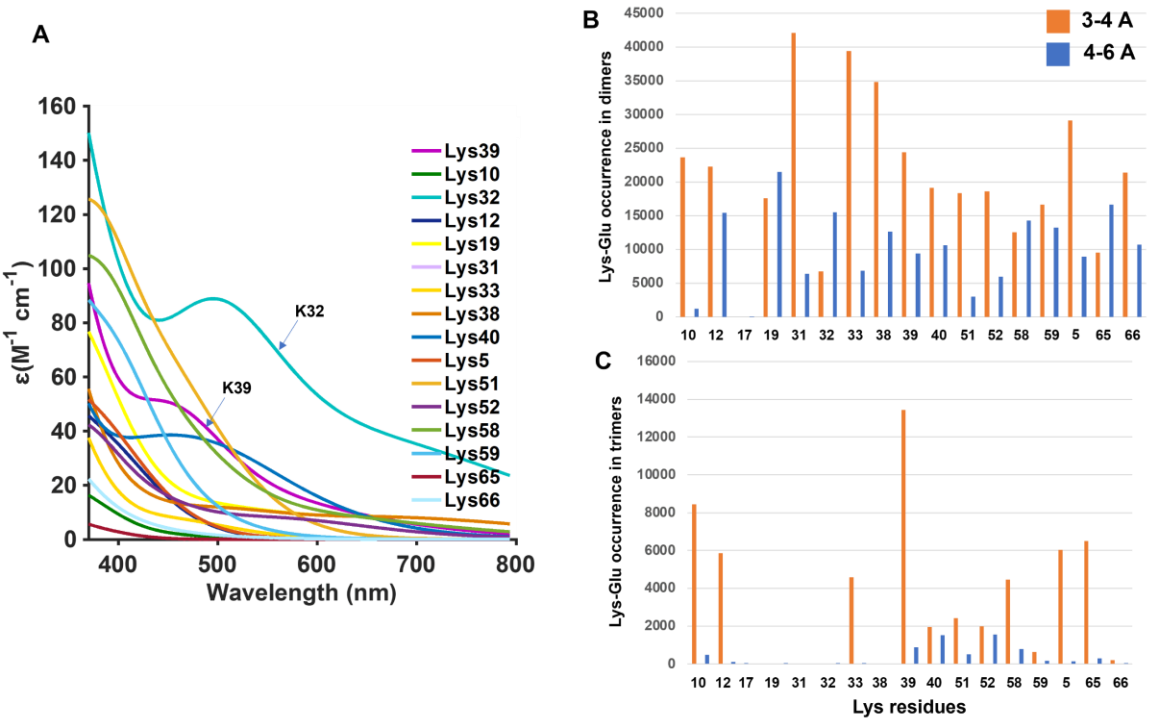


Figure M1: (A) Residue-wise contribution of individual lysine residues to the ProCharTS spectrum of native α_3C. Lys17 remained predominantly monomeric; therefore, its participation was not observed above 370 nm. (B) Lys–Glu dimer interaction analysis showing differential residue participation within two distance ranges (3–4 Å and 4–6 Å). These Lys–Glu contacts may contribute to chromophores ranging from dimers to decamers. (C) Trimer interaction analysis for lysine residues with Glu, across 50 frames of native α_3C, evaluated at 3–4 Å and 4–6 Å distance cutoffs.

Protein samples	ε of α_3C at 250 nm ($M^{-1}cm^{-1}$) before ProCharTS correction ±S.D (Conc *c*)	ε of α_3C at 250 nm ($M^{-1}cm^{-1}$) after ProCharTS correction ±S.D (Conc *c*)	ε of α_3W at 280 nm ($M^{-1}cm^{-1}$) before ProCharTS correction ±S.D (Conc *c*)	ε of α_3W at 280 nm ($M^{-1}cm^{-1}$) after ProCharTS correction ±S.D (Conc *c*)
Native	7351 ± 24 (10.5 µM)	8772 ± 5 (8.8 µM)	8451 ± 136 (10.4 µM)	9763 ± 319 (8.9 µM)
0.05 mM $(CH_3CO)_2O$	7375 ± 28 (10.1 µM)	8436 ± 4 (8.8 µM)	8264 ± 218 (10.3 µM)	9587 ± 173 (8.8 µM)
0.1 mM $(CH_3CO)_2O$	7548 ± 177 (10.2 µM)	8728 ± 20 (8.8 µM)	8819 ± 143 (10.2 µM)	10144 ± 146 (8.9 µM)
0.2 mM $(CH_3CO)_2O$	7709 ± 133 (10.0 µM)	8412 ± 3 (9.2 µM)	8220 ± 400 (10.1 µM)	8763 ± 205 (9.2 µM)
2 mM $(CH_3CO)_2O$	7515 ± 21 (10.0 µM)	7875 ± 52 (9.4 µM)	8357 ± 146 (10.2 µM)	8944 ± 170 (9.5 µM)
20 mM $(CH_3CO)_2O$	7344 ± 10 (10.3 µM)	7721 ± 3 (9.7 µM)	8058 ± 17 (10.2 µM)	8399 ± 41 (9.8 µM)

Table T1: Extinction coefficient values of α_3C and α_3W calculated by the Lowry method. The concentrations (*c*) of α_3C and α_3W were calculated using the Lowry assay, with a standard curve of BSA at known concentrations. The calculated concentrations are reported in parentheses both before and after subtraction of ProCharTS contribution at 550 nm. The Beer-Lambert law was used to calculate the extinction coefficient (ε = A/c, where A is absorbance) at 250 nm and 280 nm for α_3C and α_3W respectively. For instance, in native α_3C, A= 0.07719; therefore, before ProCharTS subtraction ε = 0.07719/10.5 µM = 7351 $M^{-1}cm^{-1}$ and after ProCharTS subtraction ε = 0.07719/8.8 µM = 8772 $M^{-1}cm^{-1}$.

Raman Peak (nm)	Excitation wavelength (nm)	Emission range (nm)
305	280	300-326
345	310	330-367
383	340	365-411
405	355	375-425
423	370	390-443
479	410	430-500

Table T2: Raman peaks observed at different excitation wavelengths. The Raman peak maxima of solvent (deionised water), along with the corresponding emission window summarised. The Raman peaks were observed using a 2 nm excitation and a 15 nm emission slit width.

Protein systems	Length of the simulation (ns)	No. of amino acids	No. of water molecules	Box type	Box dimensions (nm)	No. of charged residues added (Na^+ or Cl^-)
α_3C	100	67	8211	Cubic	6.39	2 Cl-
α_3W	100	67	8374	Cubic	6.43	2 Cl-
Ac1-α_3C	100	67	8212	Cubic	6.39	1 Cl-
Ac3-α_3C	100	67	8193	Cubic	6.39	1 Na+
Ac5-α_3C	100	67	8183	Cubic	6.39	3 Na+
Ac12-α_3C	100	67	8162	Cubic	6.39	10 Na+
Ac17-α_3C	100	67	8161	Cubic	6.39	15 Na+
Ac17-α_3W	100	67	8312	Cubic	6.43	15 Na+

Table T3: MD simulation conditions used for each of the 8 protein systems studied.

Sample	α-helix	β-sheet	Random coil
Native α_3C	94.06	0.06	5.88
α_3C+ 0.05 mM $(CH_3CO)_2O$	93.06	0.08	6.86
α_3C+ 0.1 mM $(CH_3CO)_2O$	93.06	0.08	6.86
α_3C+ 0.2 mM $(CH_3CO)_2O$	95.08	0.05	4.87
α_3C+ 2 mM $(CH_3CO)_2O$	95.08	0.05	4.87
α_3C+ 20 mM $(CH_3CO)_2O$	95.08	0.05	4.87
Native α_3W	95.56	0.06	4.38
α_3W+ 0.05 mM $(CH_3CO)_2O$	95.56	0.06	4.38
α_3W+ 0.1 mM $(CH_3CO)_2O$	95.56	0.06	4.38
α_3W+ 0.2 mM $(CH_3CO)_2O$	96.03	0.05	3.92
α_3W+ 2 mM $(CH_3CO)_2O$	95.88	0.08	4.04
α_3W+ 20 mM $(CH_3CO)_2O$	96.88	0.06	3.06

Table T4: Secondary structure content in native and acetylated proteins Analysing the effect of acetylation on the secondary structure of α_3C and α_3W. Alpha helix, beta sheet and random coil parameters were determined using the data processed by K2D3 software.

Sample	**r_{ss} (± S.D.)**
Native	0.083 (± 0.0006)
α_3W+0.05 mM $(CH_3CO)_2O$	0.075 (± 0.0023)
α_3W+0.1 mM $(CH_3CO)_2O$	0.077 (± 0.0044)
α_3W+0.2 mM $(CH_3CO)_2O$	0.086 (± 0.0043)
α_3W+2 mM $(CH_3CO)_2O$	0.077 (± 0.0034)
α_3W+20 mM $(CH_3CO)_2O$	0.052 (± 0.0004)

Table T5: Fluorescence anisotropy calculated for native and acetylated α_3W. Fluorescence anisotropy for all the protein samples was recorded at peak emission wavelength in deionised water upon excitation at 280 nm. The excitation slit was 2 nm, and the emission slit was 15 nm. The values in the brackets indicate standard deviations (SD) calculated from three independent measurements.

Sample	**α_1**	**α_2**	**τ_1 (ns)**	**τ_2 (ns)**	**τ_{mean} (ns)**	$\chi^2_{reduced}$
NATA	-	-	2.8	-	2.8	1.079
Native α_3W	0.25	0.75	1.0	2.8	2.4	1.051
α_3W+ 20 mM $(CH_3CO)_2O$ (4 hours Dialysis)	0.30	0.70	2.6	5.8	4.8	1.022
α_3W+ 20 mM $(CH_3CO)_2O$ (12 hours Dialysis)	0.25	0.75	1.6	4.9	4.1	1.014
α_3W+ 20 mM $(CH_3CO)_2O$ (24 hours Dialysis)	0.59	0.41	1.5	4.1	2.5	1.047

Table T6: Fluorescence lifetime parameters extracted from fitting tryptophan fluorescence intensity decay in NATA, native and acetylated α_3W. Protein α_3W or NATA was excited at 280 nm, and emission was collected at 347 nm for NATA; 327 nm for native α_3W and 345 nm for all the acetylated samples. The reduced chi-square value observed for all samples is nearly equal to unity, indicating a good fit for the lifetime model assumed for the decay (see Equation 3).

Sample	α_3C excited at 280 nm (± S.D.)
Native	0.016 (± 0.0012)
0.05 mM $(CH_3CO)_2O$	0.0184 (± 0.0034)
0.1 mM $(CH_3CO)_2O$	0.0177 (± 0.0014)
0.2 mM $(CH_3CO)_2O$	0.0139 (± 0.0021)
2 mM $(CH_3CO)_2O$	0.0118 (± 0.0017)
20 mM $(CH_3CO)_2O$	0.0101 (± 0.00172)

Sample	Native excited at 355 nm (± S.D.)	Acetylated (20 mM $(CH_3CO)_2O$) excited at 355 nm (± S.D.)
α_3C	0.0050 (± 0.00006)	0.0051 (± 0.00058)
α_3W	0.0032 (± 0.00022)	0.0033 (± 0.00046)

Table T7: Luminescence quantum yield (QY) of native and acetylated α_3C and α_3W for all the protein samples was measured in deionised water upon excitation at 280 and 355 nm. NATA (excitation at 280 nm) and 9,10-diphenylanthracene (DPA) (excitation at 355 nm) were used as standards and dissolved in deionised water and cyclohexane, respectively. For all the samples (including the standard), the excitation slit was 2 nm, and the emission slit was 15 nm. The values in the brackets show standard deviations calculated for three independent measurements.

Protein	R_g in Å (25000 frames)	R_g in Å (50 frames)
α_3C	11.74 ± 0.07	11.75 ± 0.06
Ac1-α_3C	11.83 ± 0.07	11.83 ± 0.08
Ac5-α_3C	11.95 ± 0.11	11.93 ± 0.10
Ac12-α_3C	12.09 ± 0.09	12.09 ± 0.10
Ac17-α_3C	12.22 ± 0.11	12.21 ± 0.13
α_3W	11.97 ± 0.08	11.97 ± 0.05
Ac17-α_3W	12.33 ± 0.08	12.30 ± 0.09

Table T8: Radius of gyration (R_g) of native and acetylated proteins from MD trajectories.

Protein	Lys-Glu Dimers (25000 frames)	Lys-Glu Dimers (50 frames)
α_3C	20322	38
Ac1-α_3C	19066	41
Ac5-α_3C	9068	13
Ac12-α_3C	1973	0
Ac17-α_3C	0	0
α_3W	28276	60
Ac17-α_3W	0	0

Table T9: The number of Lys-Glu dimers (with separation distance higher than 4 Å) calculated from MD snapshots of all the model protein systems.

Cluster Size	α_3C	Ac1-α_3C	Ac5-α_3C	Ac12-α_3C	Ac17-α_3C	α_3W	Ac17-α_3W
1	121293	157589	158917	233015	358579	182031	366925
2	72173	107667	82591	66564	40672	108653	48213
3	53802	34964	23448	19762	11507	45509	11431
4	54875	26598	40561	42907	139	42166	589
5	16679	32777	6713	585	0	19410	0
6	20214	9618	24755	3	0	12552	0
7	3767	7741	4021	0	0	4221	0
8	1343	1015	539	0	2	1000	0
9	638	490	227	0	0	586	0
10	300	115	118	0	0	430	0

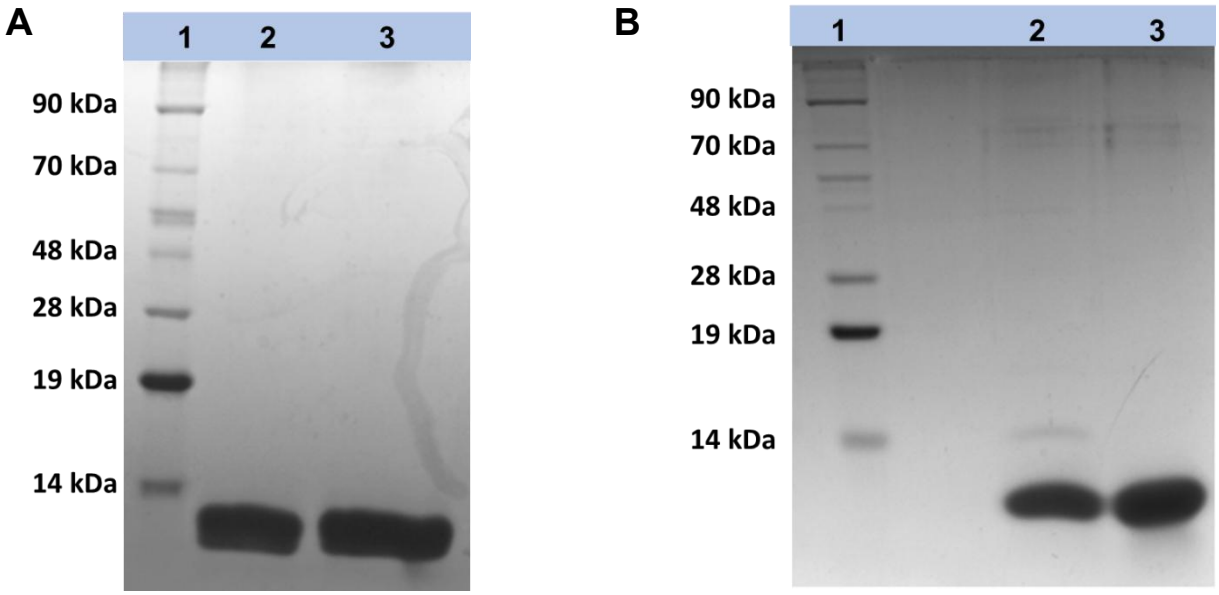
Table T10: Number of clusters of charged residues calculated from 25000 MD snapshots of all the model protein systems.

Cluster Size	α_3C	Ac1-α_3C	Ac5-α_3C	Ac12-α_3C	Ac17-α_3C	α_3W	Ac17-α_3W
1	246	318	314	464	709	354	747
2	122	218	164	128	83	221	89
3	110	67	44	38	25	87	25
4	120	55	92	83	0	90	0
5	33	69	19	2	0	44	0
6	39	17	45	0	0	23	0
7	11	12	5	0	0	7	0
8	3	2	1	0	0	2	0
9	0	2	0	0	0	0	0
10	0	1	0	0	0	1	0

Table T11: Number of clusters of charged residues calculated from 50 MD snapshots of all the model protein systems.

Proteins	Positively charge amino acids	Negatively charged amino acids	Total Charged percentage	Molecular Weight (in Da)
α_3C	19	17	53.7	7,460
α_3W	19	17	53.7	7,543
K-RAS	35	29	40.5	21,081
Histone H2A	30	9	30.0	14,007

Table T12: Comparison of charged amino acid residues and molecular weight of multiple proteins for which ProCharTS was recorded.



Lane: 1) Protein Ladder

Lane: 2) 100 µM α_3C

Lane: 3) 120 µM α_3C

Lane: 1) Protein Ladder

Lane: 2) 70 µM α_3W

Lane: 3) 100 µM α_3W

Figure F1: Purity analysis by gel electrophoresis: 15% SDS-PAGE was run to check the purity of the α_3C (A) and α_3W (B).

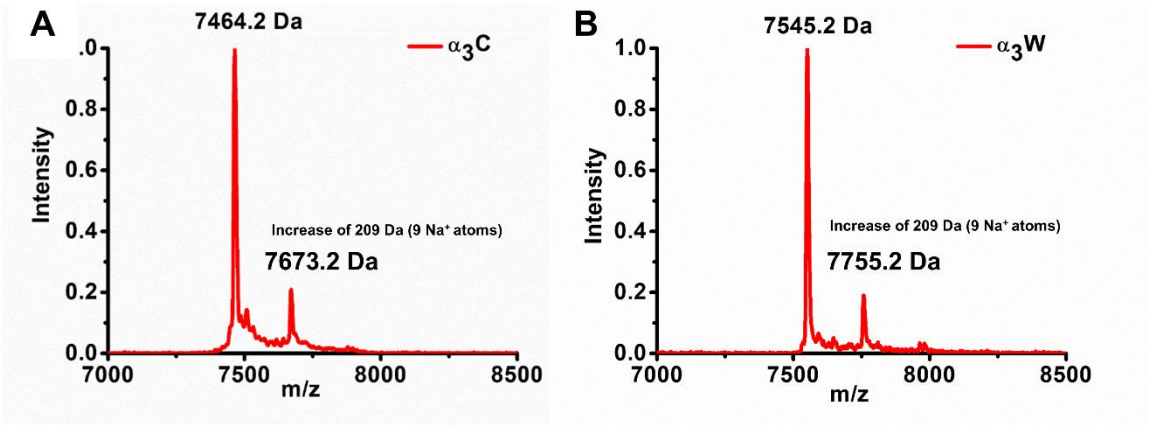


Figure F2: Mass spectra of purified α_3C and α_3W: The mass observed in MALDI-ToF for α_3C is 7464.2 Da (A), 7545.22 Da for α_3W (B), respectively, which matches the calculated mass. The additional low intensity peak shows the sodium adduct formation. The additional mass observed in sodium adducts in α_3C and α_3W corresponded to 9 sodium atoms.

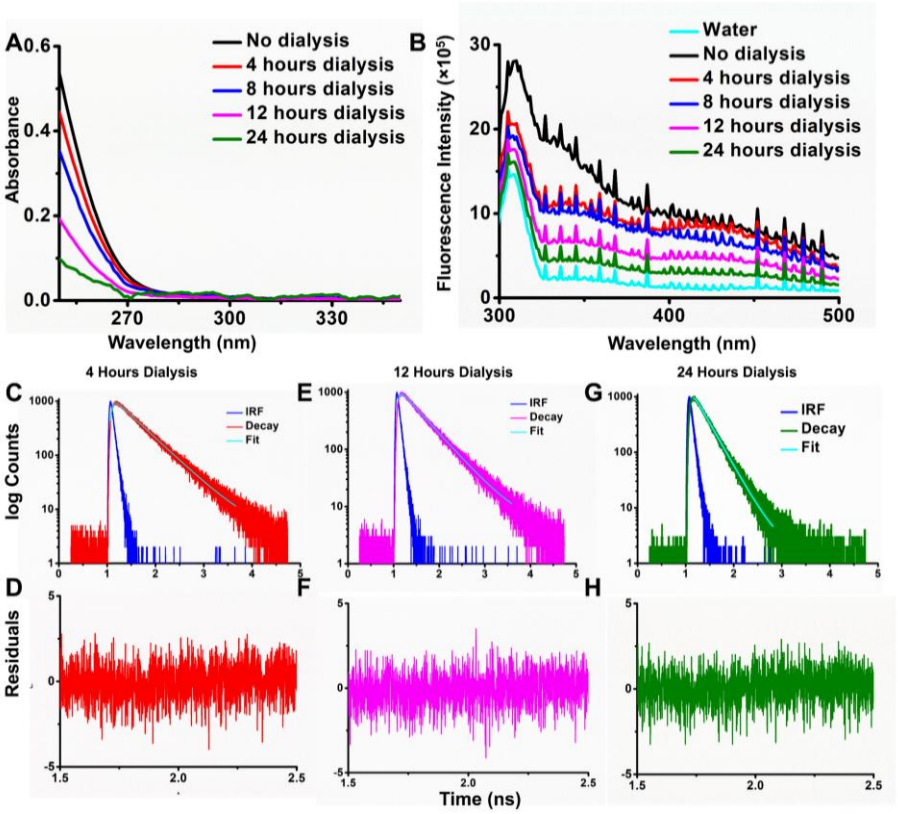


Figure F3: Spectroscopic analysis of acetic anhydride in deionised water: Absorption (A); steady state fluorescence spectra (B); and fluorescence intensity decays (C, E, G) of protein with acetic anhydride were recorded in deionised water at room temperature after dialysis for 4, 12, and 24 hours, respectively. The residuals of the fit for all the samples are also shown (D, F, H) below each decay. Note that the emission peak between 310-320 nm in panel B corresponds to the Raman peak for water. The parameters extracted from fitting are shown in Table T6.

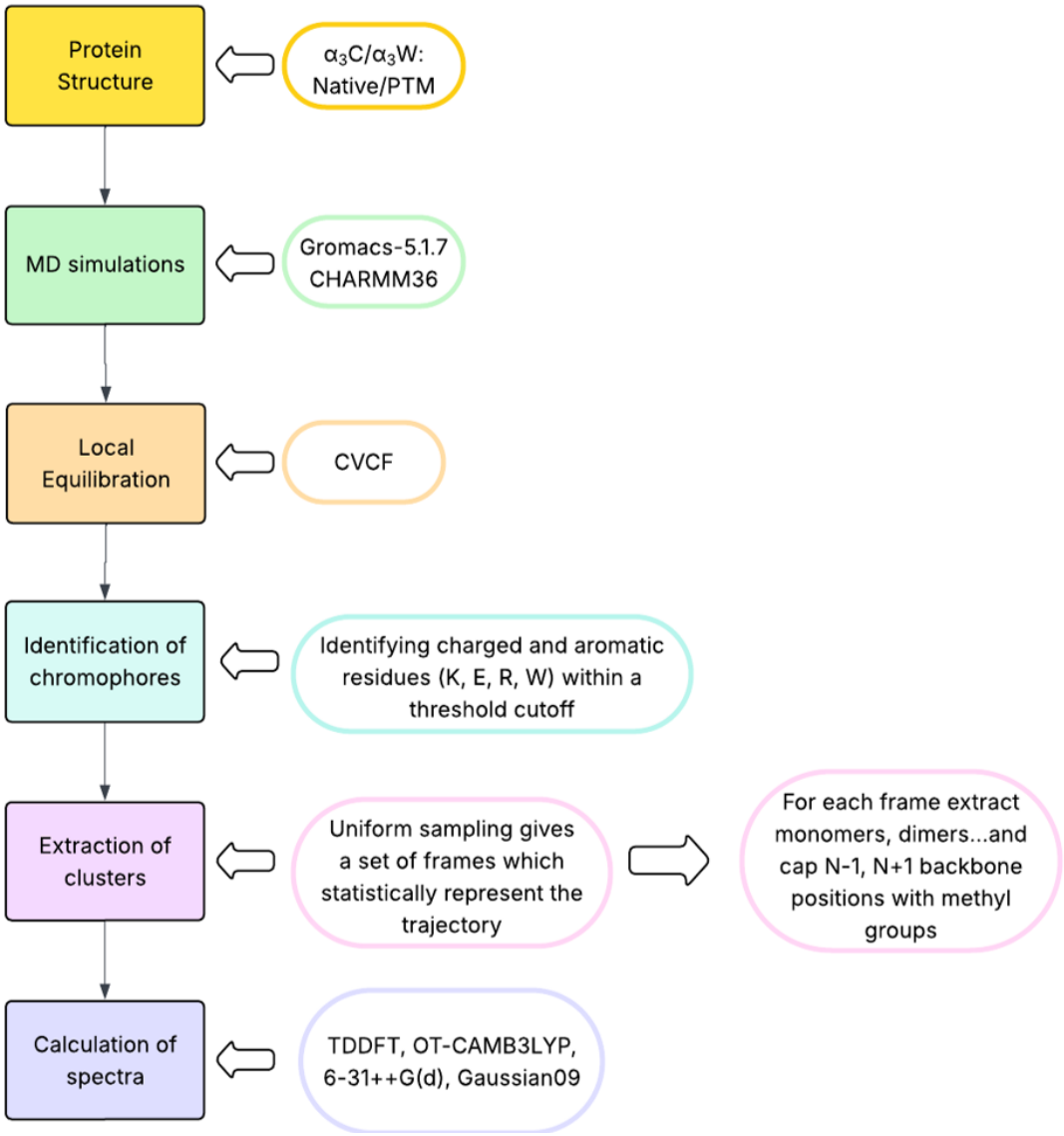


Figure F4: Workflow for generating spectral profiles of proteins (α_3C, Acn-α_3C, α_3W, and Ac17-α_3W in this manuscript).

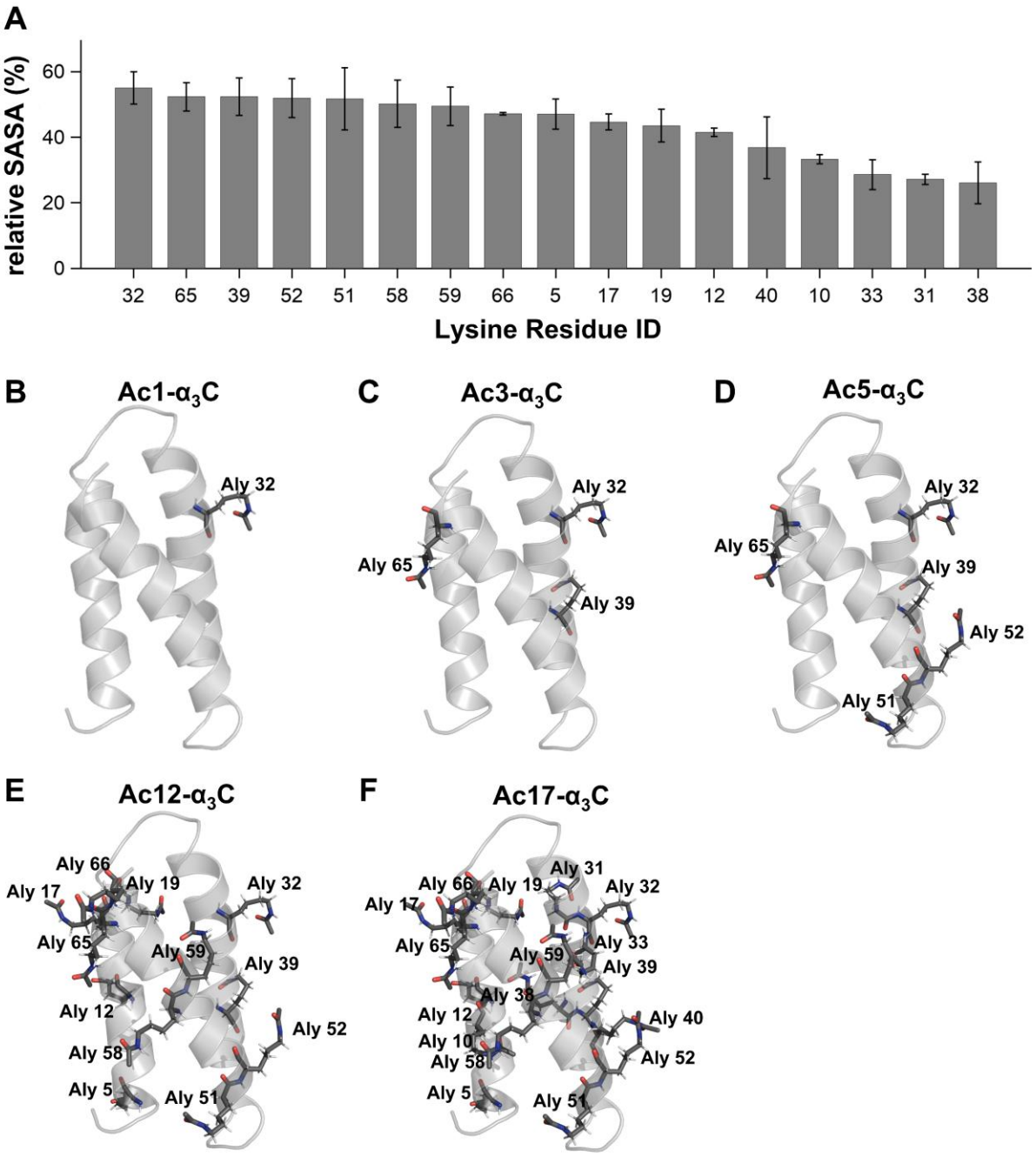


Figure F5: (A) Bar plot showing the relative solvent-accessible surface area (SASA) of Lys residues in the α3C protein, averaged over 50 ns MD simulation. Error bars represent standard deviations across the trajectory. (B–E) Cartoon representations of acetylated Lys (Aly) residues in progressively modified Acn-α3C variants, illustrating the extent of Lys acetylation: (B) Ac1-α3C, (C) Ac3-α3C, (D) Ac5-α3C, (E) Ac12-α3C and (F) Ac17-α3C.

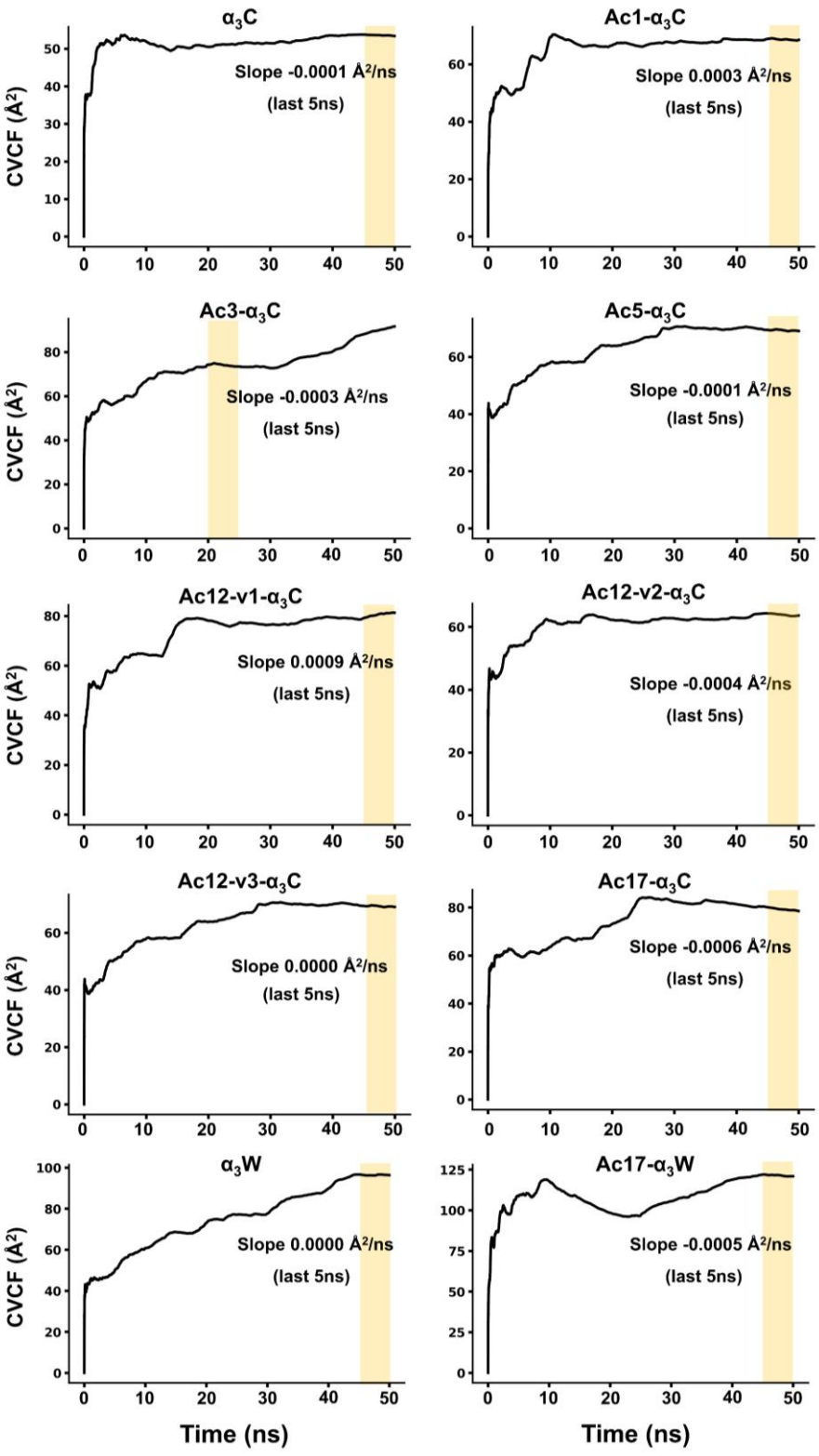


Figure F6: Cumulative Variance of Atomic Coordinate Fluctuation (CVCF) traces of structured backbone atoms of α_3C, its acetylated variants (Acn- α_3C), α_3W and Ac17-α_3W over 50 ns production MD trajectories.

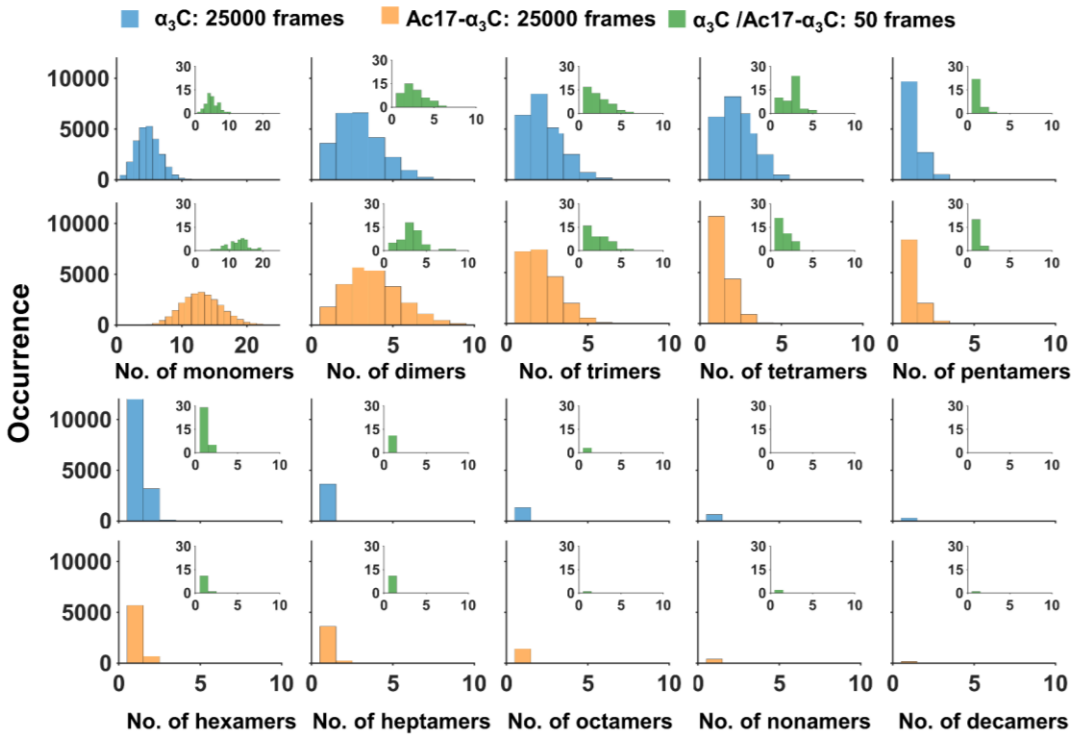


Figure F7: Cluster distribution (size ranging from monomers to decamers) across 50 (green) and 25,000 (blue and orange) frames in α_3C and Ac17-α_3C protein trajectories.

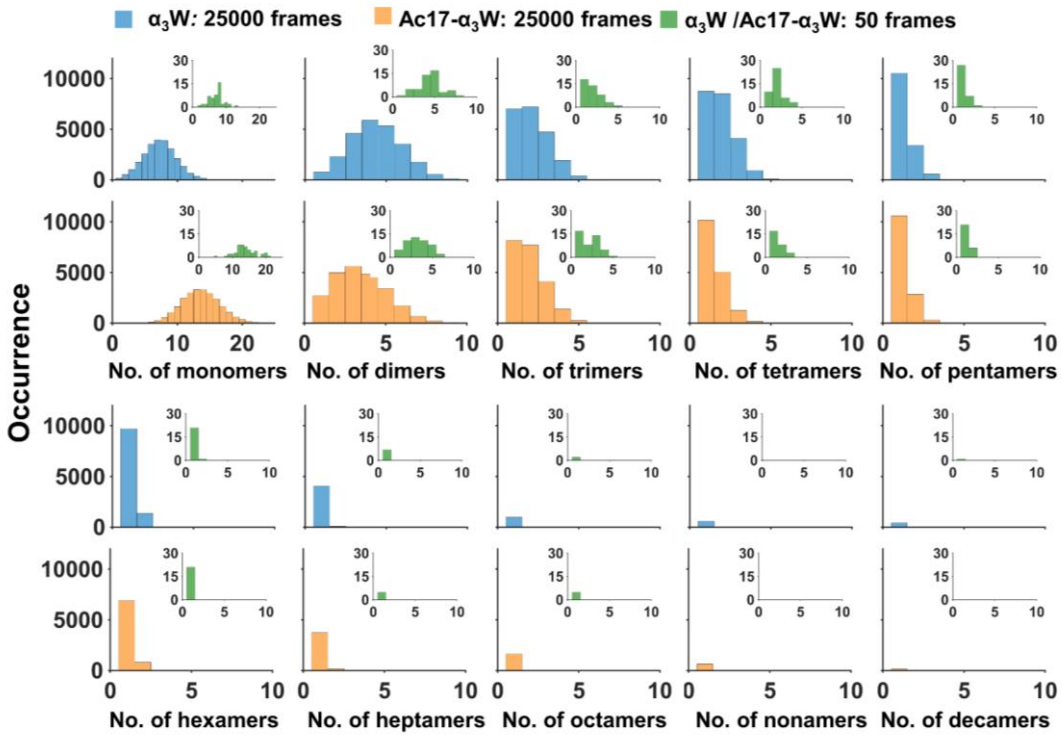


Figure F8: Cluster distribution (size ranging from monomers to decamers) across 50 (green) and 25,000 (blue and orange) frames in α_3W and Ac17-α_3W protein trajectories.

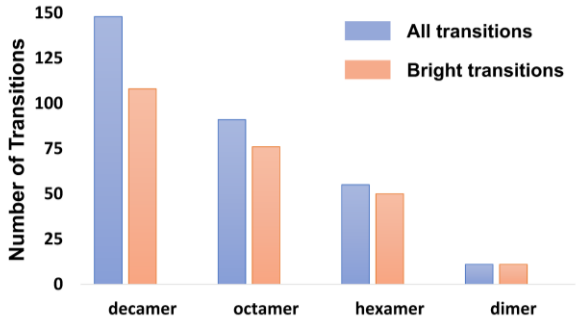


Figure F9: Number of transitions calculated using TDDFT as a function of cluster size. Bright transitions are those with oscillator strength > 0.0000

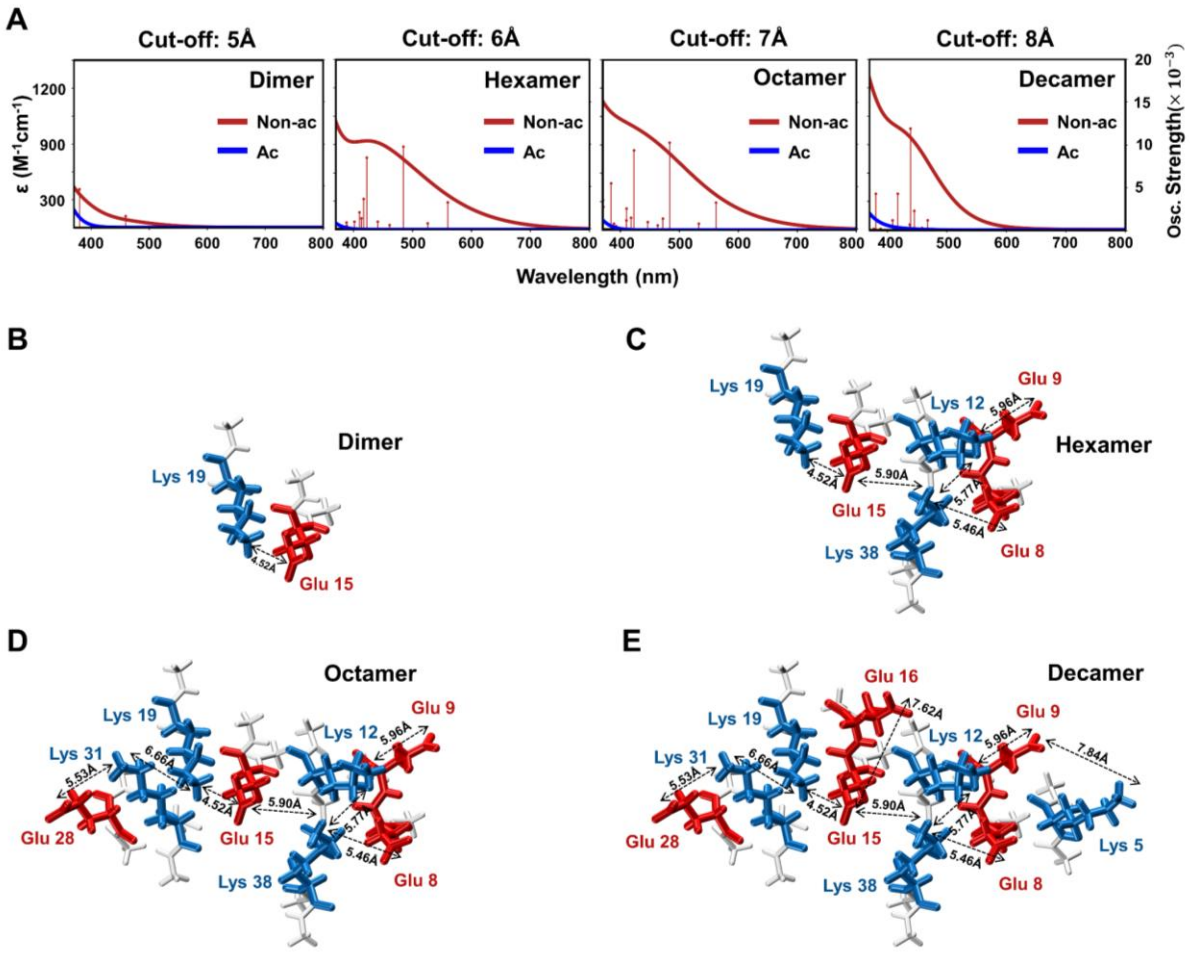
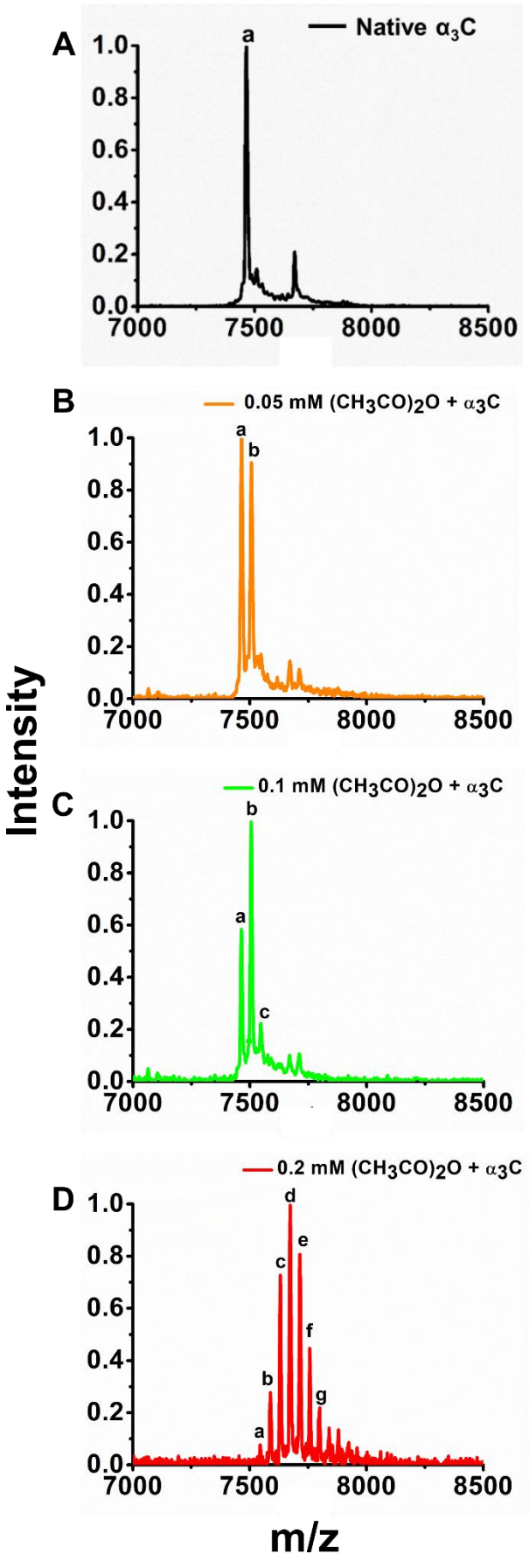


Figure F10: (A) Simulated spectra of native and acetylated (Ac) decamer using different cutoff distances. Stem lines represent the oscillator strengths of underlying transitions. Biggest cluster size in each case is mentioned at the top right side of the plot. (B) to (E) are the conformations of the biggest clusters mentioned in each cutoff distance selected. Higher order interactions present in multimers but not accounted when we apply 6Å as distance criteria, do not cause any significant change in spectra of the entire cluster both before and after acetylation.

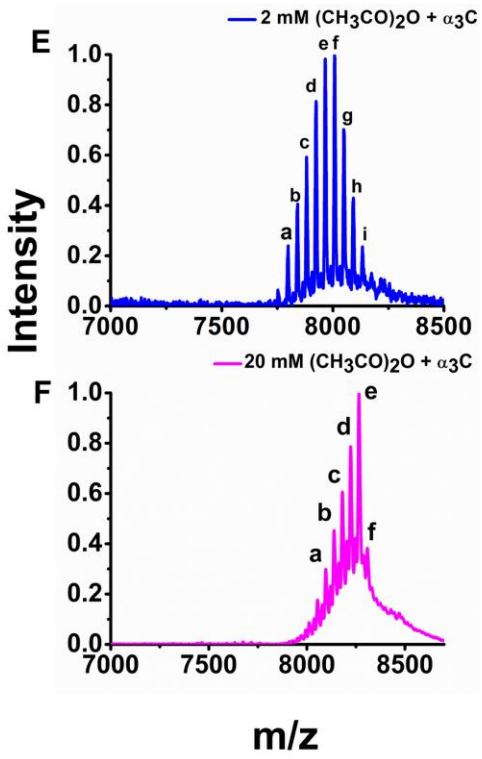


Sample	Molecular Mass (in Da)	No. of residues modified
a	7464.3	0

S.No	Molecular Mass (in Da)	No. of residues modified
a	7464.8	0
b	7506.8	1

S.No	Molecular Mass (in Da)	No. of residues modified
a	7464.8	0
b	7506.8	1
c	7547.5	2

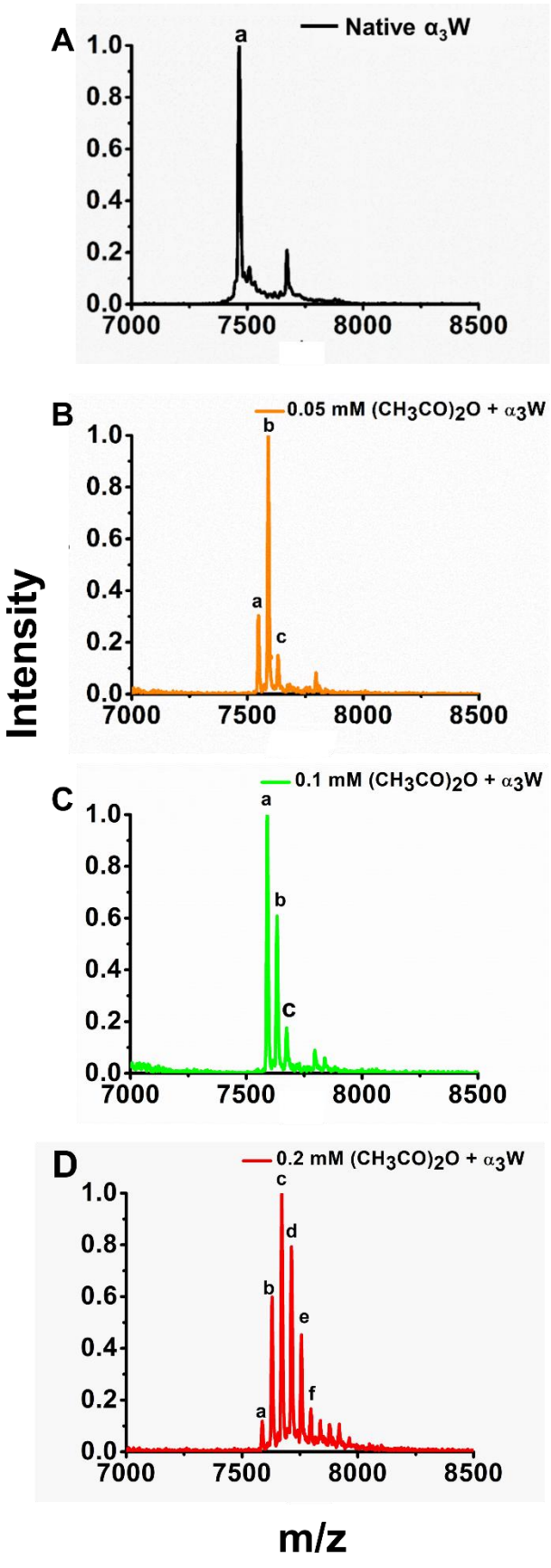
S.No.	Molecular Mass (in Da)	No. of residues modified
a	7548.0	2
b	7589.5	3
c	7631.5	4
d	7673.0	5
e	7716.5	6
f	7758.3	7
g	7800.2	8



S.No.	Molecular Mass (in Da)	No. of residues modified
a	7797.0	8
b	7840.6	9
c	7881.7	10
d	7925.1	11
e	7966.0	12
f	8008.0	13
g	8050.2	14
h	8090.8	15
i	8132.6	16

S.No.	Molecular Mass (in Da)	No. of residues modified
a	8094.3	15
b	8136.2	16
c	8178.0	17
d	8220.5	18
e	8262.3	19
f	8304.5	20

Figure F11: Mass spectra of native and different degrees of acetylated α_3C**:** The molecular mass was recorded for (A) native and acetylated with (B) 0.05 mM, (C) 0.1 mM, (D) 0.2 mM, (E) 2 mM, (F) 20 mM $(CH_3CO)_2O$ (acetic anhydride) α_3C. In the plot, the different peaks represent the different acetylated species. The molecular weights of all the peaks are listed in the figure.

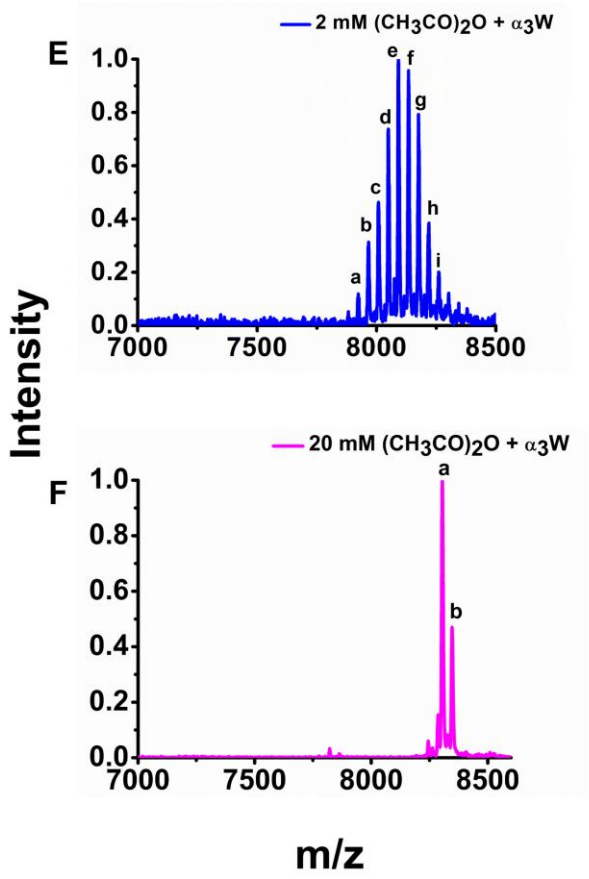


S.No.	Molecular Mass (in Da)	No. of residues modified
a	7545.3	0

S.No	Molecular Mass (in Da)	No. of residues modified
a	7546.2	0
b	7590.2	1
c	7634.2	2

S.No	Molecular Mass (in Da)	No. of residues modified
a	7590.2	1
b	7632.1	2
c	7676.7	3

S.No.	Molecular Mass (in Da)	No. of residues modified
a	7588.3	1
b	7630.0	2
c	7670.9	3
d	7714.0	4
e	7755.7	5
f	7797.5	6



S.No.	Molecular Mass (in Da)	No. of residues modified
a	7926.5	9
b	7967.5	10
c	8007.5	11
d	8049.2	12
e	8092.0	13
f	8134.0	14
g	8176.0	15
h	8219.8	16
i	8262.0	17

S.No.	Molecular Mass (in Da)	No. of residues modified
a	8305.0	18
b	8347.1	19

Figure F12: Mass spectra of native and acetylated α_3W: The molecular mass was recorded for (A) native and acetylated with (B) 0.05 mM, (C) 0.1 mM, (D) 0.2 mM, (E) 2 mM, (F) 20 mM $(CH_3CO)_2O$ (acetic anhydride) α_3W. In the plot, the different peaks represent the different acetylated species. The molecular weight values of all the peaks are mentioned in the figure.

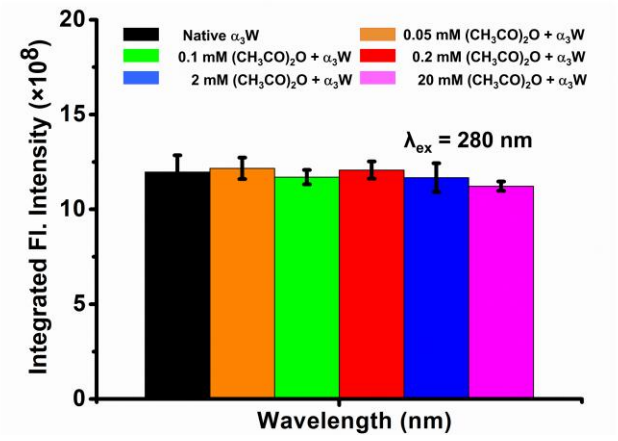


Figure F13: Area under the curve of native and acetylated proteins: The area under the emission curve for steady-state fluorescence was plotted upon excitation at 280 nm for native and different extents of acetylated samples. All the samples were recorded in deionised water with 2 nm and 15 nm excitation and emission slit widths, respectively.

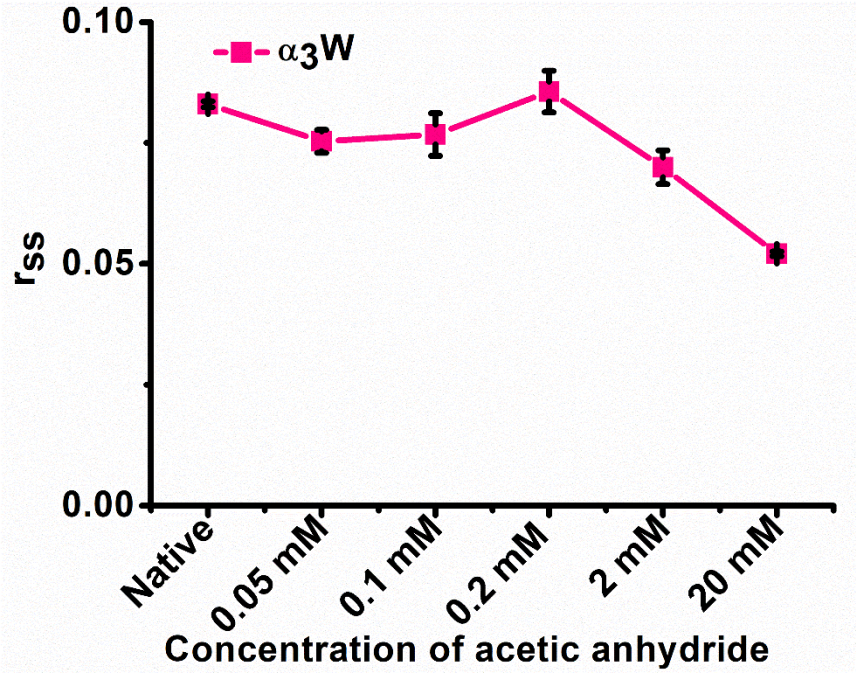


Figure F14: Tryptophan fluorescence anisotropy of native and acetylated α_3W. The fluorescence anisotropy of tryptophan was recorded for native and acetylated α_3W as a function of different concentrations of acetic anhydride used. The samples were excited at 280 nm with a slit width of 2 nm and emission was collected at peak wavelength with a slit width of 15 nm. The error bar in the figure shows the standard deviation from three independent experiments.

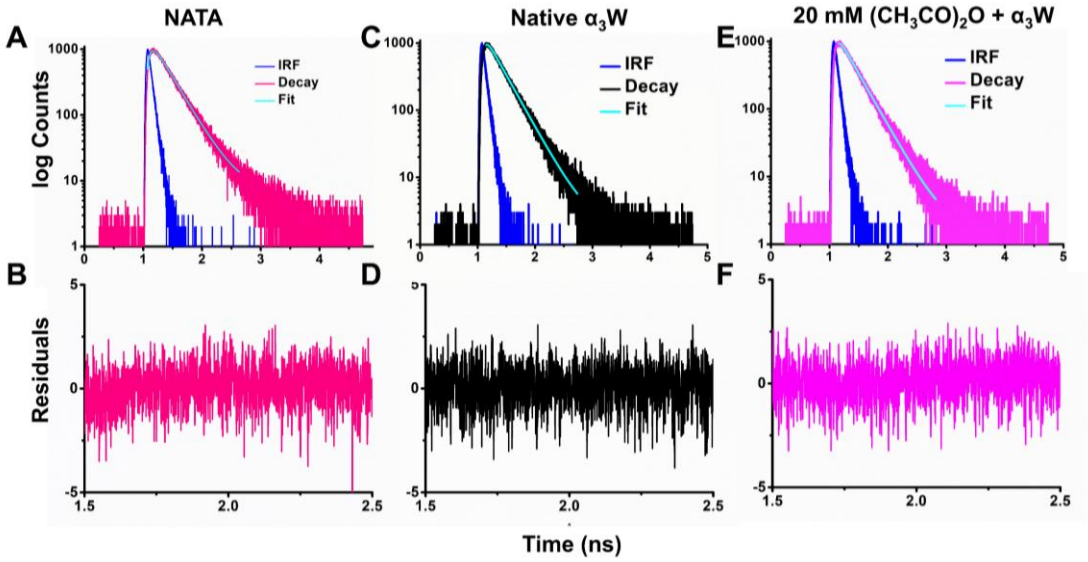


Figure F15: Tryptophan fluorescence intensity decay observed upon excitation at 280 nm: NATA (A); native α_3W (C); and α_3W acetylated with 20 mM acetic anhydride (E). Decays were obtained by exciting the sample at 280 nm and collecting the emission at 347 nm for NATA; 327 nm for native α_3W and 345 nm for acetylated sample. Residuals for the fitted decay are shown below the decay for NATA (B); native α_3W (D); and acetylated α_3W (F). Fitted parameters are listed in Table T6.

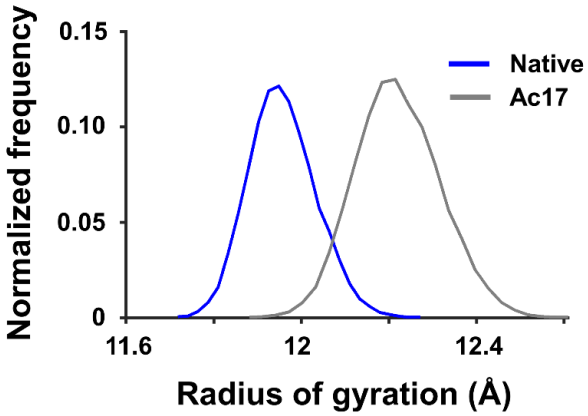


Figure F16: The normalized frequency distribution of the radius of gyration (R_g) for native-α_3W (blue) and Ac17-α_3W (gray) over 50 ns MD trajectory.

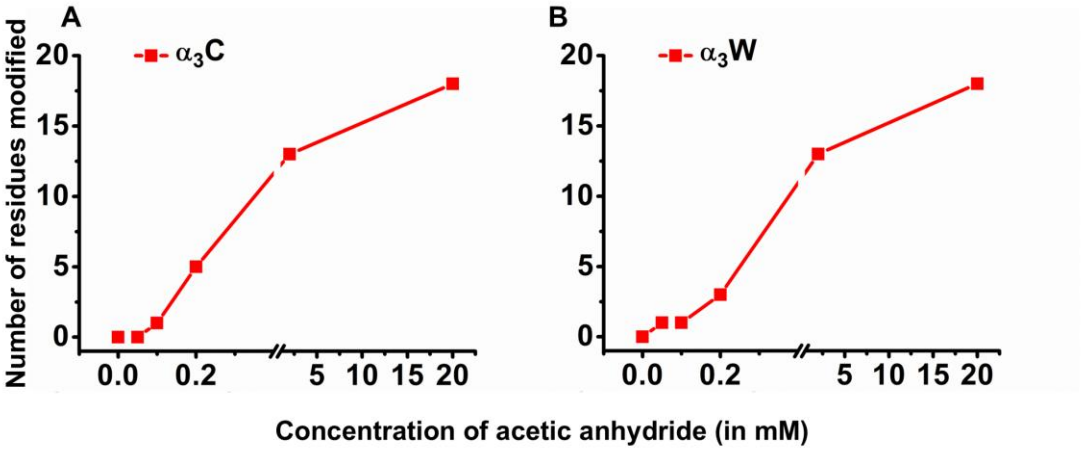


Figure F17: Number of lysine residues modified plotted against the different concentrations of acetic anhydride. The break in the X-axis of α_3C and α_3W is between 0.4 mM and 0.5 mM.

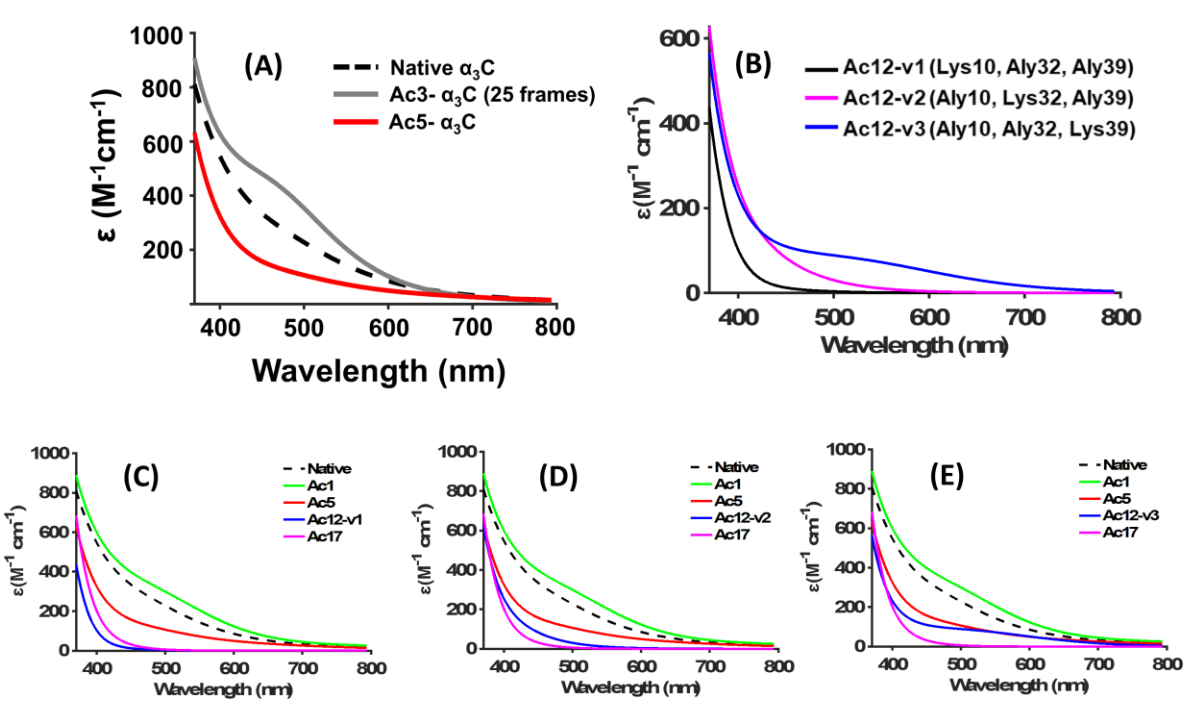


Figure F18: (A) Computed spectra for native α_3C (dotted line), Ac3-α_3C (grey solid line), and Ac5-α_3C (red solid line). (B) Comparison of the three Ac12 variants differing in the acetylation state of Lys10, Lys32, and Lys39. (C–E) ProCharTS profile of Ac12-v1, Ac12-v2, and Ac12-v3, respectively, shown alongside Native, Ac1, Ac5, and Ac17 systems.

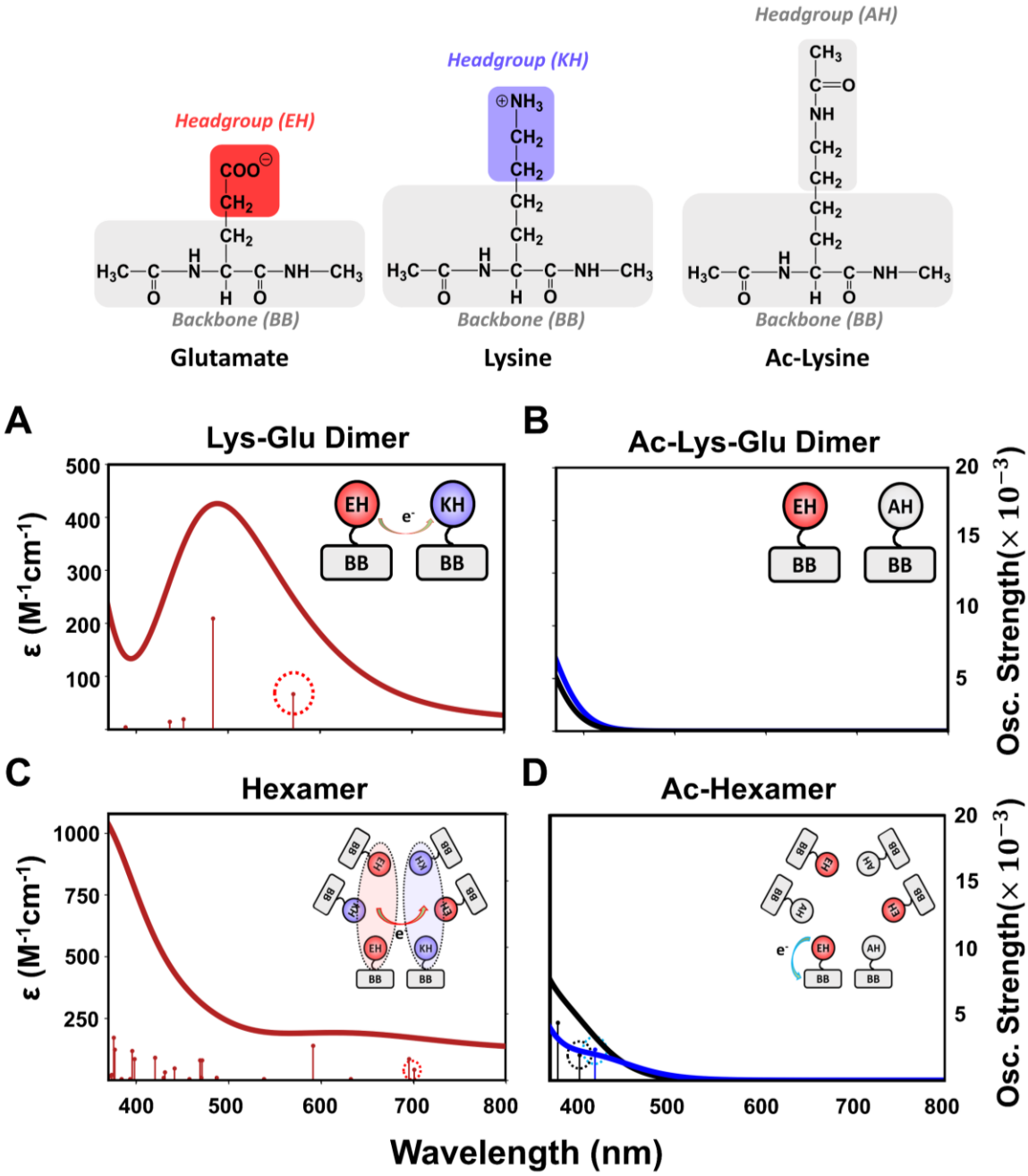


Figure F19: Absorption spectra of small clusters before and after acetylation. (A) and (C) are absorption spectra of lysine-glutamate dimer and hexamer respectively. (B) and (D) are absorption spectra of dimer and hexamer after acetylation of lysine residues. Blue line represents the spectra of cluster where acetylated lysines are present and black line represents the cluster with only glutamate residues (without acetylated lysines). Stem lines are oscillator strengths of the underlying transitions. Charge density difference (CDD) for the redmost transitions are indicated by schematic representations at top right corner in each panel. Actual CDD are plotted for those transitions in **Figure F20**.

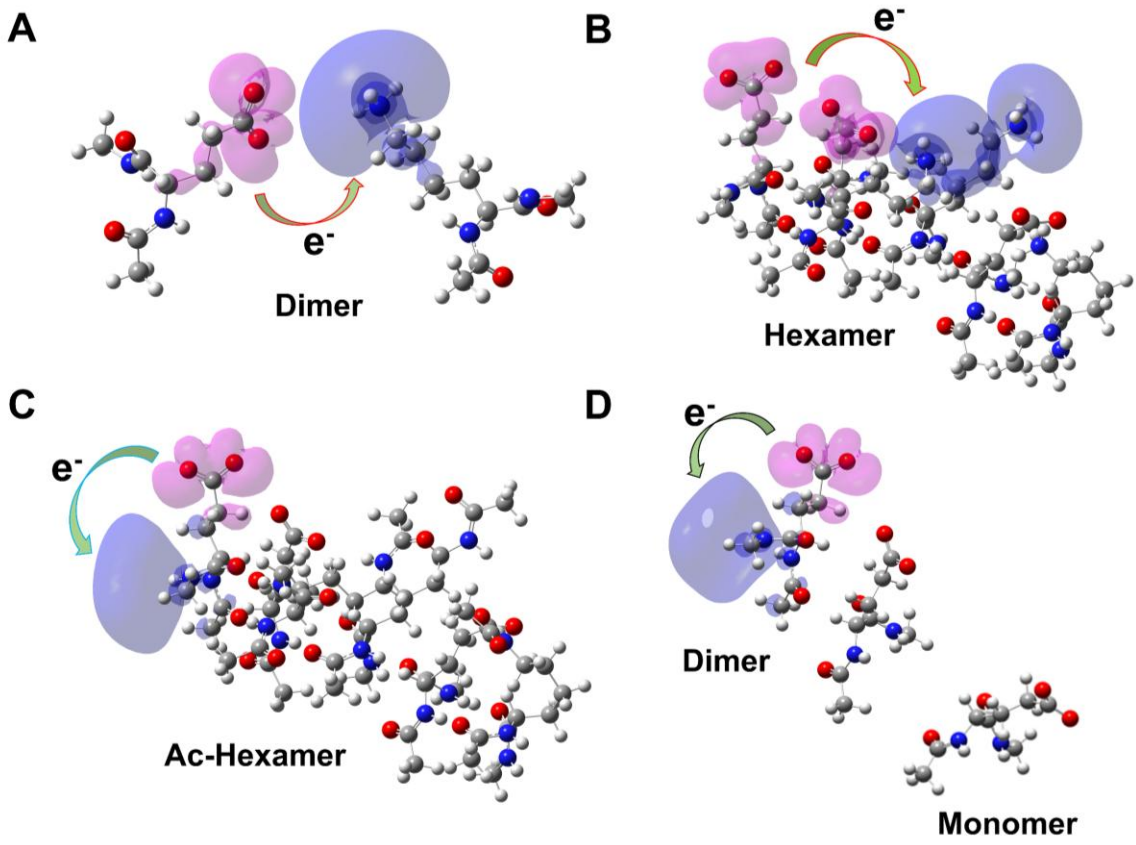


Figure F20: (A-D) Hole and electron density distribution for the transitions marked by dotted circle in panels (A-D) of **Figure F19** respectively. (D) is the same acetylated hexamer (C) but without the acetylated lysine residues. Hence, it is a combination of dimer and monomer glutamates.

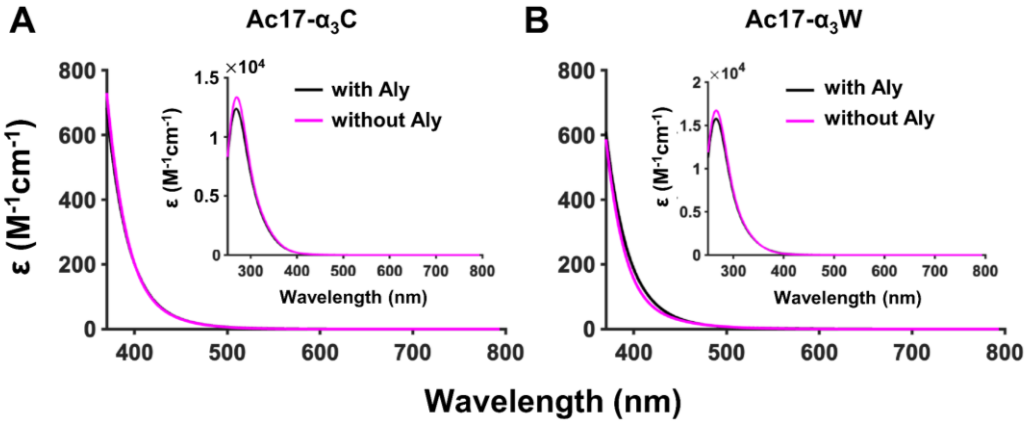


Figure F21: Comparison of computed spectra for acetylated (A) α_3C and (B) α_3W proteins using two approaches. (1) Including all charged residues and aromatic amino acids with acetylated lysine (Aly), and (2) including only charged residues and aromatic amino acids, excluding Aly.

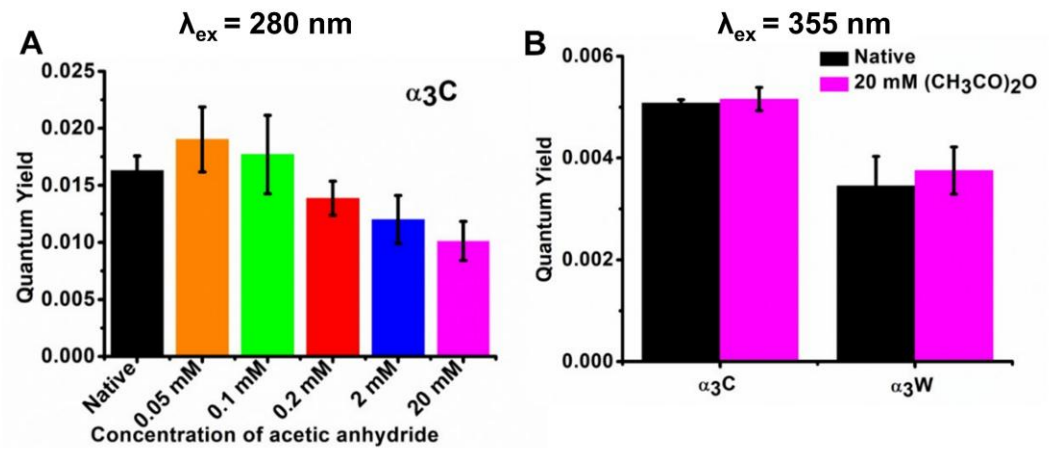


Figure F22: Luminescence quantum yield of native and different degrees of acetylated α_3C and α_3W: Samples were excited at a wavelength of 280 nm for α_3C (A). The sample was excited at 355 nm, and the quantum yield was calculated for native and acetylated α_3C and α_3W (B).

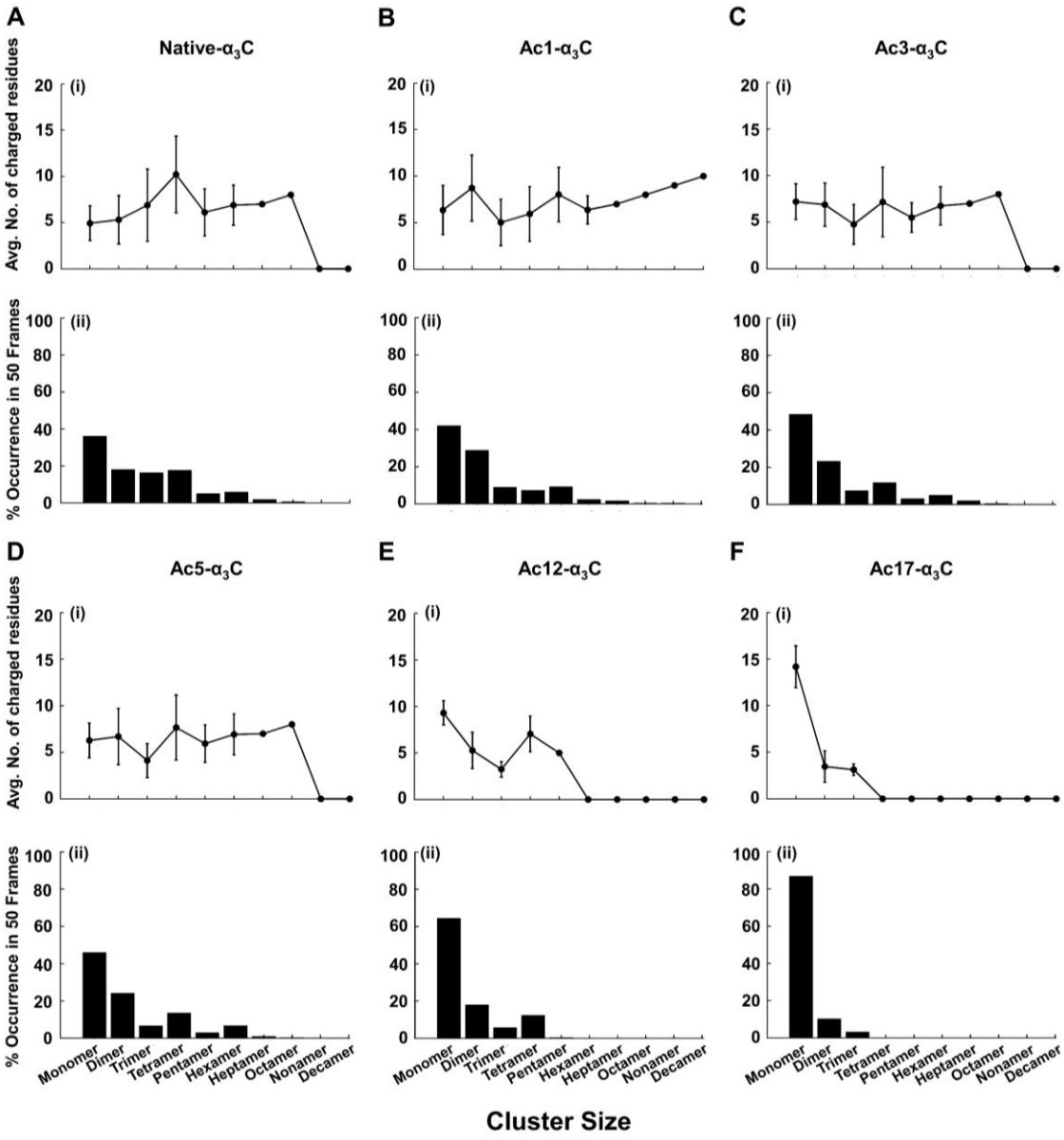


Figure F23: Comparison of charge-mediated clustering behavior across different protein variants: (A) Native-α3C, (B) Ac1-α3C, (C) Ac3-α3C, (D) Ac5-α3C, (E) Ac12-α3C and (F) Ac17-α3C. (i) Average number of charged residues involved per cluster across different cluster sizes. (ii) Percentage occurrence of various cluster sizes over 50 sampled frames.

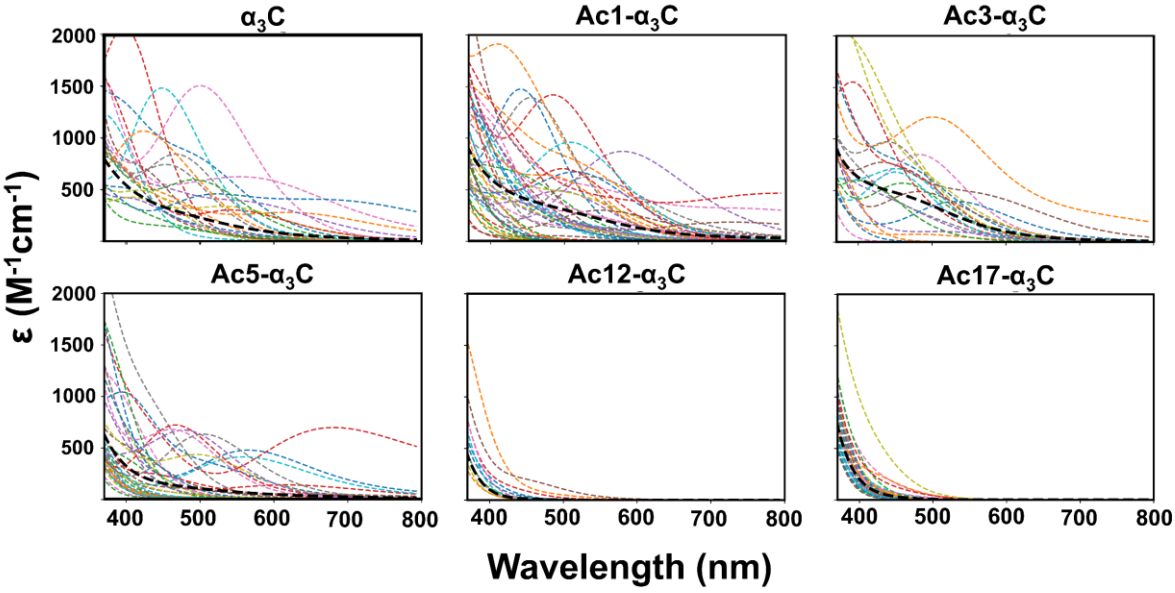


Figure F24: Averaged spectra (black dashed line) of α_3C proteins differing the number of acetylated lysines. Spectra at selected frames are represented by colored dashed line. 25 frames were selected in case of Ac3-α_3Cwhile 50 frames were selected for rest of the systems.

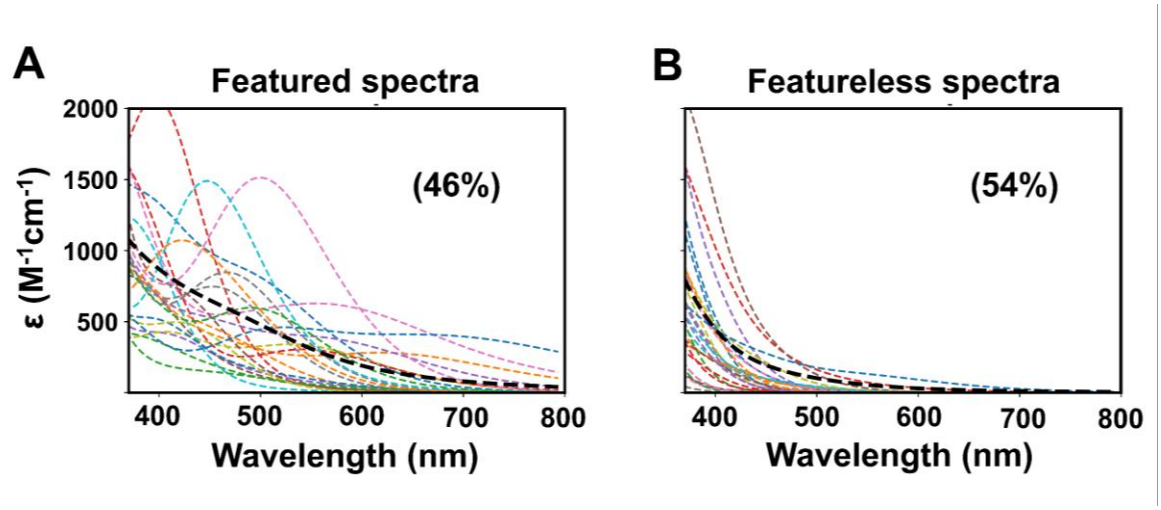


Figure F25: Spectra of all 50 frames for α_3C protein divided into featured (A) and featureless (B) spectra. Average spectra for both the cases is shown by black dotted line.

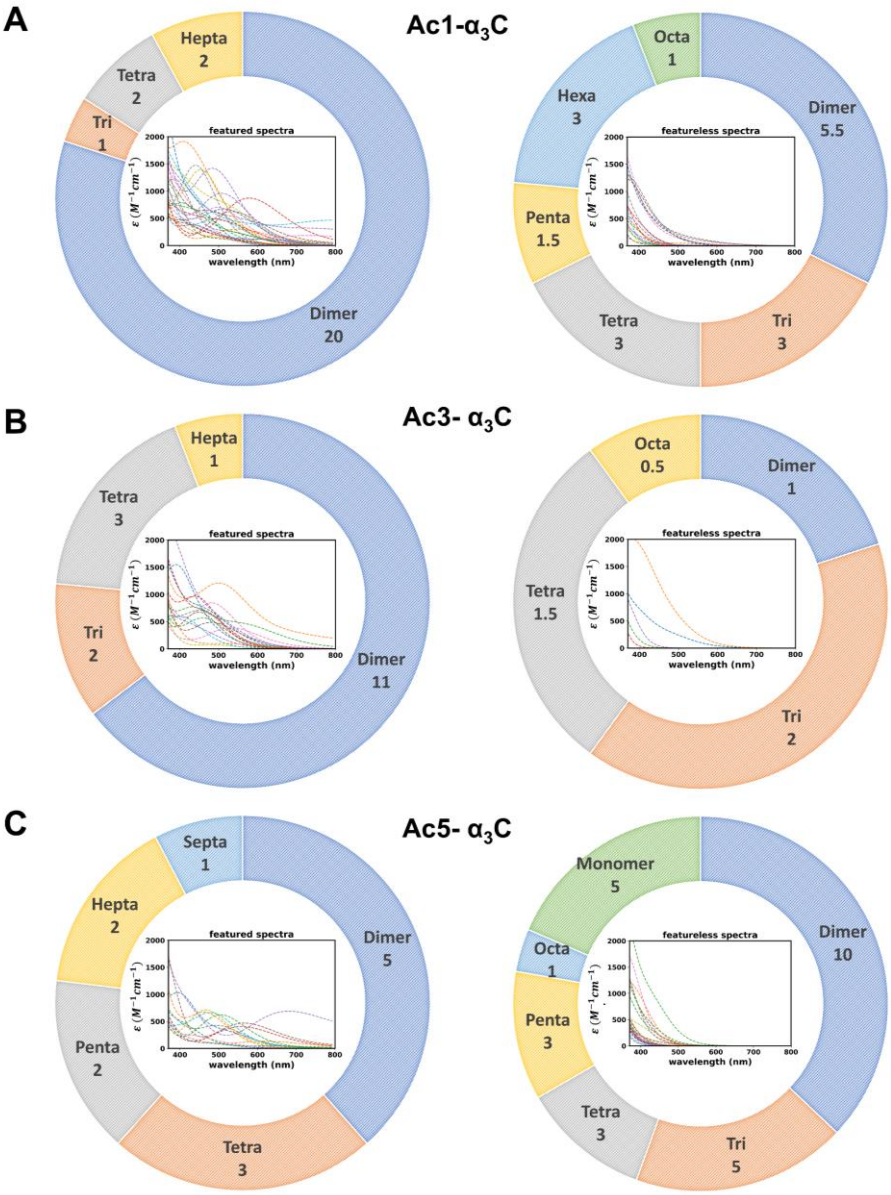


Figure F26: Contributions of clusters to the featured (left) and featureless (right) spectra of Ac1-α_3C (A), Ac3-α_3C (B) and Ac5-α_3C (C). The number of frames of which spectral intensity (above 450nm (left) and 350-450nm (right)) is predominantly contributed, is indicated for each cluster. Few out of 50 frames have zero transition above 350nm and they are discarded from analysis.

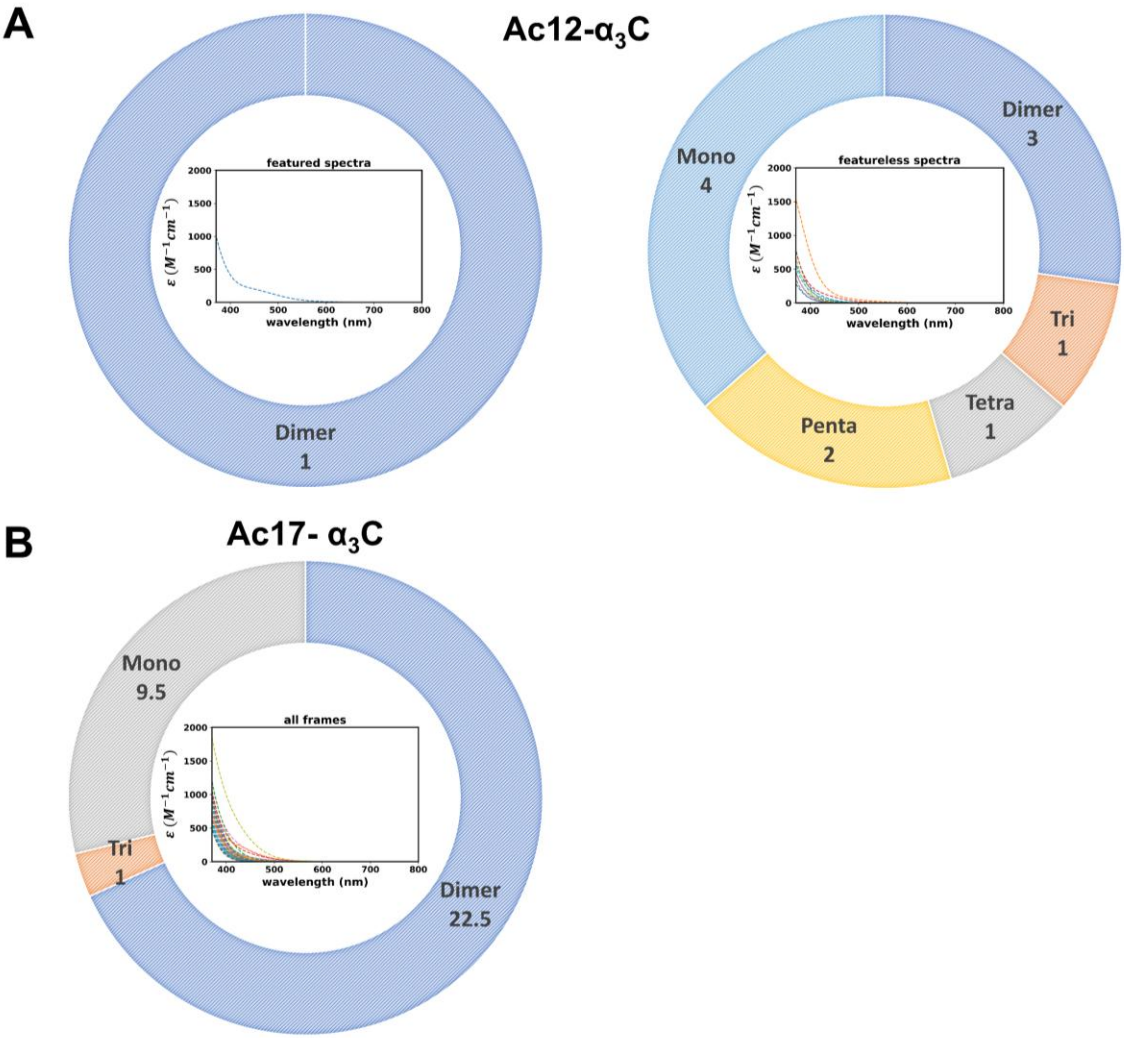


Figure F27: Contributions of clusters to the featured (left) and featureless (right) spectra of Ac12 (A) and Ac17 (B). The number of frames of which spectral intensity (above 450nm (left) and 350-450nm (right)) is predominantly contributed, is indicated for each cluster. Few out of 50 frames have zero transition above 350nm and they are discarded from analysis.

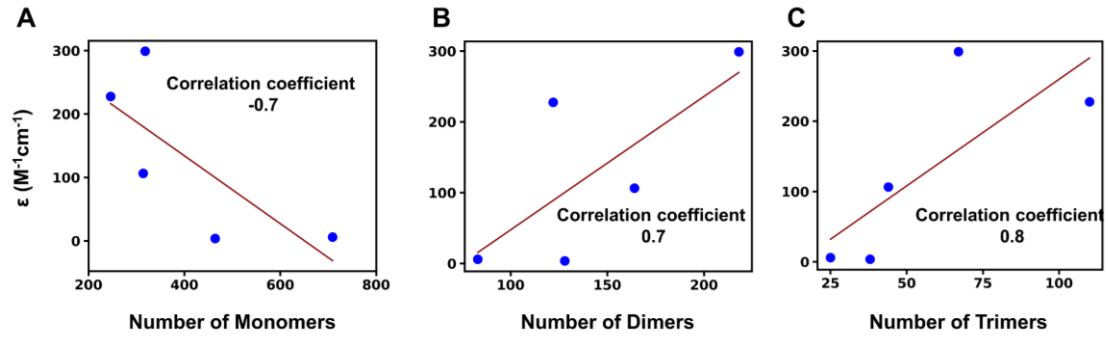


Figure F28: Correlation between spectral intesity (at 500nm) and total number of different sized clusters sampled from 50 frames: A) number of monomers, B) number of dimers and C) numbers of trimers.

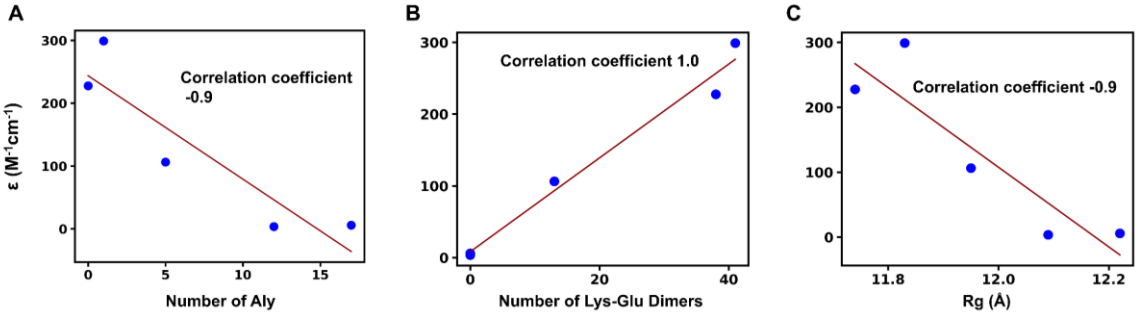


Figure F29: Correlation between spectral intesity (at 500nm) and A) number of Aly residues, B) number of Lys-Glu dimers with separation distance higher than 4Å, and C) R_g of proteins.

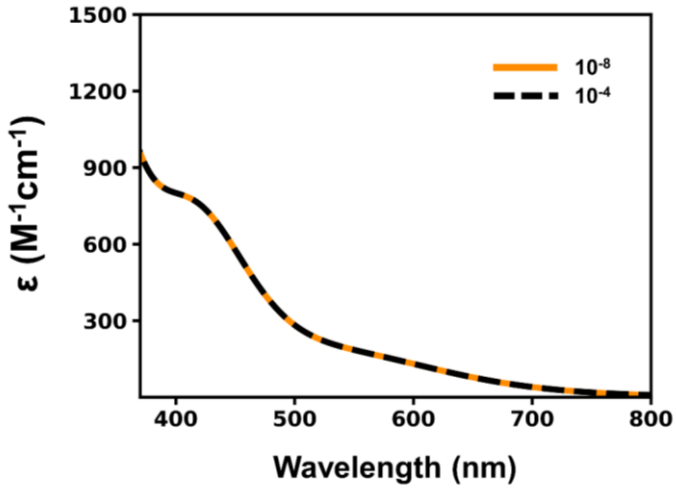


Figure F30: Comparison of spectra of a cluster (hexamer) simulated using two different SCF convergence criteria for root mean square (RMS) change density matrix. Two criteria are mentioned in the figure. 10^{-4} is used throughout all the calculations.

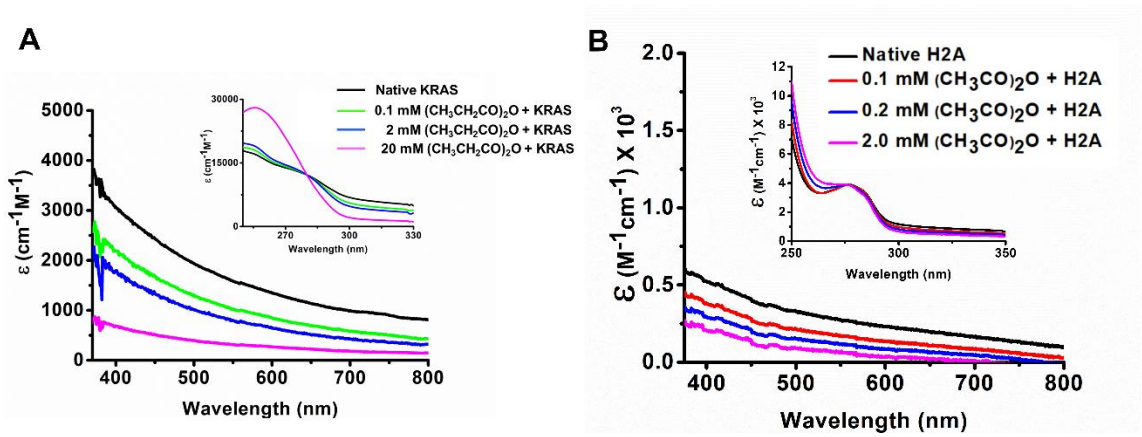


Figure F31: Absorption spectra of K-RAS (A) and Histone H2A (B) proteins. The acetylation reaction with KRAS (16 µM) and Histone H2A (20 µM) was performed for 30 minutes at 4˚C at pH ~8 in 100 mM HEPES buffer. Different concentrations of acetic anhydride were added to proteins, as shown in the figure, to generate different degrees of acetylation. Both spectra exhibit monotonic decrease in spectra intensity above 400 nm with increasing acetic anhydride concentrations.

References:

1 J. R. Lakowicz, Ed., *Principles of Fluorescence Spectroscopy*, Springer US, Boston, MA, 2006.
2 S. Sahu, T. Debnath and K. Sahu, *Journal of Physical Chemistry Letters*, 2024, **15**, 3677–3682.
3 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic acids research*, 2000, **28**, 235–242.
4 W. Humphrey, A. Dalke and K. Schulten, *Journal of molecular graphics*, 1996, **14**, 27-28,33-38.
5 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, D. J. Fox, et al., 2013.
6 Q. H. Dai, C. Tommos, E. J. Fuentes, M. R. A. Blomberg, P. L. Dutton and A. J. Wand, *Journal of the American Chemical Society*, 2002, **124**, 10952–10953.
7 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. Van Der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.
8 B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
9 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J Comput Chem*, 2005, **26**, 1701–1718.
10 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *Journal of Computational Chemistry*, 2010, **31**, 671–690.
11 P. Mark and L. Nilsson, *Journal of Physical Chemistry A*, 2001, **105**, 9954–9960.
12 T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089–10092.
13 G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**, 014101.
14 M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182–7190.
15 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
16 S. Paul, S. R. K. Ainavarapu and R. Venkatramani, *Journal of Physical Chemistry B*, 2020, **124**, 4247–4262.