Supporting Information

Kwon et al. 10.1073/pnas.1011569107

SI Text

SI Materials and Methods. ***Materials.*** All restriction enzymes were purchased from New England Biolabs (Beverly, MA). D,L-5,5,5-trifluoroleucine (Tfl) was purchased from Oakwood Products (West Columbia, SC). DNA oligomers were synthesized at Qiagen (Valencia, CA) or Integrated DNA Technologies (Coralville, IA). (4*S*)-2-amino-4-methylhexanoic acid (homoisoleucine, Hil) was prepared according to the methods of O'Donnell and Eckrich (1) and Dorizon and coworkers (2).

Plasmid construction. An *Eco*RI/*Hin*dIII fragment of pQEA1 (3) containing the A1 coding sequence was ligated into *Eco*RI/*Hin*dIII-digested pQE-80L (Qiagen) to yield pQE-80L/A1. The Asp residue at position 34 of A1 was mutated to Trp by site-directed mutagenesis. The resulting plasmid was designated pQE-80L/A1m. A Leu codon was introduced into either position 31 or position 37, yielding pQE-80L/S31L and pQE-80L/A37L, respectively. Plasmid pA1EL (4), which encodes both the protein A1 and a constitutively expressed leucyl-tRNA synthetase (LeuRS), was mutated using similar site-directed mutagenesis techniques. A Trp codon was introduced first at position 34 of A1, yielding pA1mEL. Introduction of leucine codons into either position 31 or position 37 resulted in the plasmids pS31LEL and pA37-LEL, respectively.

Expression of fluorinated proteins. M9 medium supplemented with 0.4% glucose, 3.5 mg/L thiamine, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 20 amino acids (40 mg/L), and 200 mg/L ampicillin was inoculated 1/50 with an overnight culture (M9) of *Escherichia coli* strain DH10B transformed with pQE-80L/A1m, pQE-80L/S31L, or pQE-80L/A37L and grown at 37 °C with shaking. After each culture reached $OD_{600} = 0.9–1.0$, the cells were harvested by centrifugation (6,000 × *g*, 4 °C, 6 min) and washed twice with cold 0.9% NaCl. The cell pellets were resuspended in M9 medium containing 19 amino acids (no Leu) and 1 mM Tfl. Protein expression was induced 10 min after the medium shift by addition of IPTG to a final concentration of 1 mM. After 3 h, the cells were harvested by centrifugation (6,000 × *g*, 4 °C, 10 min), and the cells were stored at −20 °C at least 12 h before purification. In the case of the protein S31L-T, one sample was made using the procedure for the production of proteins containing Hil described below.

Expression of proteins containing homoisoleucine. M9 medium supplemented with 0.4% glucose, 35 mg/L thiamine, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 20 amino acids (40 mg/L), 200 mg/L ampicillin, and 35 mg/L kanamycin was inoculated 1/50 with an overnight culture (M9) of *E. coli* strain LAM1000 transformed with pREP4 and pA1mEL, pS31LEL, or pA37LEL at 37 °C with shaking. After each culture reached $OD_{600} = 0.9–1.1$, the cells were harvested by centrifugation (5,000 × *g*, 4 °C, 15 min) and washed three times with cold 0.9% NaCl. The cell pellets were resuspended in M9 medium containing 19 amino acids (no Leu) and 0.5 mM Hil. Protein expression was induced 15 min after the medium shift by the addition of IPTG to a final concentration of 1 mM. Cells were harvested by centrifugation (5,000 × *g*, 4 °C, 15 min), resuspended in Qiagen buffer B (8M urea, 100 mM NaH_2PO_4, 10 mM TrisCl, pH 8.0) and sonicated for 10 min total process time with a pulse duration of 5 s and a wait duration of 5 s. The sonicated lysates were frozen at −80 °C for at least 12 h before proceeding with purification.

Expression of hydrogenated proteins. 2xYT medium was used instead of supplemented M9 medium. When the culture reached $OD_{600} = 0.9–1.0$, IPTG was added to a final concentration of 1 mM. After 3 h, the cells were harvested by centrifugation (6,000 × *g*, 4 °C, 10 min), and the cells were stored at 20 °C at least 12 h before purification. In the case of S31L-L, one sample was expressed using the cell strain LAM1000 containing pREP4 and pS31LEL and harvested using the procedure used for the production of proteins containing Hil.

Protein purification. N-terminally histidine-tagged A1 variants were purified under denaturing conditions by affinity chromatography using Ni-NTA resin (Qiagen, Chatsworth, CA) according to the manufacturer's instructions. For proteins containing Hil and for one batch each of S31L-L and S31L-T, the lysates in Qiagen Buffer B were thawed, sonicated, and then clarified using centrifugation (~75,000 × *g*, 25 °C, 10 min). Imidazole was added to Qiagen Buffer C (50 mM) and Qiagen Buffer E (250 mM) in order to improve purification efficiency. The purified protein solutions were dialyzed against 10 mM sodium acetate (pH 4)/100 mM NaCl, and were concentrated by ultrafiltration (Amicon Ultra-15 devices, mwco: 10,000 or 3,000, Millipore, Billerica, MA). The protein concentration was determined as measured by the absorbance at 280 nanometers of solutions, assuming extinction coefficients of 5,500 $M^{-1}\,cm^{-1}$ (5).

Amino acid analysis and sedimentation velocity analysis. Amino acid analysis of fluorinated proteins was performed at the W. M. Keck Facility at Yale University (New Haven, CT) on a Hitachi L-8900 amino acid analyzer (San Jose, CA) after hydrolysis at 115 °C in 70% formic acid. Sedimentation velocity analysis was performed at the National Analytical Ultracentrifugation Facility at the University of Connecticut (Storrs, CT) by using a Beckman XL-I Analytical Ultracentrifuge at 20 °C. The rotor was accelerated to 55,000 rpm, and interference scans were acquired at 1 min intervals for 7 h. The data were analyzed by using the program Sedfit (6) to obtain normalized $c(s)$ vs. sedimentation coefficient plots (Fig. S1).

Mass spectrometry. Liquid chromatography tandem mass spectrometry (LC/MS/MS) of proteins containing Hil or Tfl was performed at the Caltech Protein and Peptide Mass Analytical Laboratory. Trypsinized samples were subjected to liquid chromatography on an Eksigent (Dublin, CA) NanoLC-2D using a 6 cm long, 100 μm diameter C18 column, followed by MS/MS on an Applied Biosystems (Foster City, CA) QStar XL instrument. Data was analyzed using Analyst QS software provided by Applied Biosystems. Hil or Tfl incorporation levels were estimated using information contained within extracting ion currents (XIC) of trypsin-digested protein samples. For a given sample, a peak corresponding to a peptide globally substituted with the noncanonical amino acid and coding for multiple leucines was identified, and the related peak corresponding to replacement at a fraction of the leucine positions was also identified. Determination of the ratio of the partially substituted to globally substituted peak areas allowed for the estimation of amino acid incorporation levels assuming that leucines in the fragment were replaced statistically. An example calculation is shown in the section "LC/MS/MS Amino Acid Incorporation Estimates," which can be found below.

Steady-state measurements. Fluorescence absorption and emission spectra can be found in Fig. S2.

Time-correlated single-photon counting (TCSPC). The protein samples were prepared at 55 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution. The TCSPC measurements were performed by using femtosecond pulses (<100 fs) from a Ti-sapphire oscillator (Spectra-Physics, Mai Tai HP). Laser output, of which the repetition rate was attenuated from 80 MHz to 8 MHz utilizing a pulse picker (Spectra-Physics, Model 3980-5), was tuned to 885 nanometers and frequency-tripled to 295 nanometers using a time-plate tripler (Minioptic Technology, TP-2000B) for selective excitation of Trp. The UV beam, vertically polarized using a half waveplate, was introduced to a sample chamber and focused onto the sample cell. The residual frequency-doubled beam from the tripler was directed to a photodiode to trigger a TCSPC system (PicoQuant GmbH, FluoTime 200). Typically, the energy of the excitation pulse (attenuated) at the sample was ~10 picoJoules. In a right-angle geometry, the emitted fluorescence was collected at a magic angle (54.7°) with respect to the vertically polarized excitation beam and focused into a MCP-PMT (Hamamatsu, R3809U), which is attached to a double monochromator. The photomultiplier tube signal was routed to a time-to-amplitude converter as a start signal followed by a constant fractional discriminator (PicoQuant GmbH, SPC 630). To avoid possible photobleaching and photodegradation, samples were kept stirring using a micro magnetic stirrer. In this configuration, the instrument response has a full width at half maximum of ~30 ps. Multiexponential decays convoluted with instrumental response functions were analyzed using the FluoFit software package (PicoQuant).

Femtosecond fluorescence upconversion. The protein samples were prepared at 550 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution. We used an amplified Ti-sapphire laser system (Spectra-Physics, Hurricane X), which produces ~110-fs pulses centered at 805 nanometers (fundamental), with a 1-kHz repetition rate and a 0.8-millijoule energy. The output beam was split into equal parts to generate the pump and the gate pulse trains. For the pump, the fundamental light was used to pump an optical parametric amplifier (Spectra-Physics, OPA-800C), the infrared idler output of which was sum-frequency mixed with the residual fundamental in a 0.5-mm thick β-barium borate (BBO) crystal (type I), recompressed with a prism pair, and frequency-doubled to provide the 295-nanometers pulses in a 1.0-mm thick BBO crystal. The pump pulses were focused, with a 24 cm focal length lens, on the rotating circular cell (1-mm thickness) containing the sample. Typically, the energy of the pump pulse (attenuated) at the sample was ~200 nJ. At these energies, the fluorescence signals from samples were linearly dependent on the pump energy. To check for sample degradation during experiments, fluorescence spectra were periodically measured right after the rotating cell by using a fiber-optic-coupled spectrometer (Acton Research, SpectraPro-300i) coupled to a charge-coupled device (Princeton Instruments, SpectruMM-256HB) before and after the collection of averaged transients for each sample. No difference between the spectra was observed.

The forward-scattered fluorescence from excited samples was collected and focused by two off-axis parabolic mirrors into a 0.5-mm thick BBO crystal. Cutoff filters were placed between the mirrors to reject scattered laser light and pass the desired fluorescence wavelengths. The gate pulses, attenuated to 23 μJ/pulse, passed through a computer-controlled optical delay line and were noncollinearly overlapped with the fluorescence in the BBO crystal. After the crystal, the upconverted signal was separated from the gate beam and the fluorescence by using an iris, and was focused on the entrance slit of a 0.25-m double-grating monochromator (Jobin Yvon, DH10) equipped with a photomultiplier tube at the exit slit. Upconversion efficiency was maximized by angle-tuning of the BBO crystal. The upconverted fluorescence transients were taken at the magic angle (54.7°) of the pump polarization relative to the gate polarization, parallel to the acceptance axis of the upconversion crystal, in order to eliminate the influence of induced sample anisotropy on the signal. The photomultiplier output was amplified (Stanford Research Systems (SRS), SR445) and processed by a gated integrator (SRS, SR250). The temporal response of the instrument was typically 350–450 fs. The observed fluorescence transients were fit to theoretical functions, using a Scientist nonlinear least-squares fitting program (Micromath), for the convolution of the Gaussian instrument response function with a sum of exponentials. All experiments were carried out at an ambient temperature of ~24 °C, and all fluorescence transients were obtained by the excitation of samples at 295 nanometers.

For fluorescence anisotropy measurements, the pump-beam polarization was rotated either parallel or perpendicular to the acceptance axis of the upconversion crystal to collect the parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) signals, respectively. These transients were used to construct time-resolved anisotropy: $r(t) = (I_{\parallel} - I_{\perp})/(I_{\parallel} + 2I_{\perp})$. The results of the time-resolved anisotropy are shown in Fig. S3. The ultrafast depolarization time constant, ϕ_I, attributed to fast internal conversion between the first two excited singlet states (^1L_a and ^1L_b) of Trp, varies dramatically with the time resolution. This process has a time scale of ~100 fs (7, 8). The limited resolution of our current apparatus (350–450 fs) does not allow us to fully resolve these dynamics and gives rise to a large uncertainty in the value of ϕ_I. The variability of ϕ_I will impact the fit of ϕ_{Trp}. The uncertainty in the amplitude of the anisotropy, r_{Trp} and r_∞, is not however affected by the limited time resolution of our data. Therefore we use the wobbling cone angle to reveal details about the local crowding near Trp. The wobbling cone angle is given by $1 - r_{Trp}/(r_{Trp} + r_\infty) = [(3\cos^2\theta - 1)/2]^2$ (9), and only depends upon the amplitude of the tryptophan wobbling motion and the anisotropy due to the rotation of the molecule.

LC/MS/MS amino acid incorporation estimates. LC/MS/MS was used to estimate the replacement levels of leucine in some protein samples. Fig. S4 depicts the total ion currents (TIC, Fig. S4*A*) and three extracted ion currents (XICs, Fig. S4 *B–D*) from a digested A1m-H sample. The large peak in Fig. S4*B* corresponds to a peptide in which all of the Leu residues are replaced by Hil, the smaller peak in Fig. S4*C* corresponds to a mixture of two peptides containing one Leu and one Hil residue, and the very small peak in Fig. S4*D* corresponds to a peptide containing only Leu residues. The areas in the three XICs allow determination of the extent of incorporation of noncanonical amino acids in place of leucine. Assuming that there is a probability p of homoisoleucine substitution in place of leucine, the distribution of peak areas should correspond to the binomial distribution

$$A[(1-p)^2 + 2(1-p)p + p^2], \qquad \textbf{[S1]}$$

where A is a multiplication factor equal to the total area of the three peaks and the three terms of the polynomial correspond to nonsubstituted, singly substituted, and doubly substituted peptides, respectively (The term for singly substituted peaks takes into account both positional isomers of singly substituted peptides). Because these three terms represent the only combinations of substitutions possible in the peptide, the relationship

$$(1-p)^2 + 2(1-p)p + p^2 = 1 \qquad \textbf{[S2]}$$

also holds. The ratio between two peaks in a peptide series depends only on the probability of incorporation and not on the value of A. Therefore, the ratio of two peaks from experimental

data can be used in order to get an estimate of p. The ratio of the peak areas of singly substituted to doubly substituted peptides is

$$\frac{2(1-p)p}{p^2}. \quad \textbf{[S3]}$$

Rearranging the above expression,

$$\frac{2p-2p^2}{p^2} = \frac{2(1-p)p}{p^2} = X, \quad \textbf{[S4]}$$

where X is the experimentally observable ratio of singly substituted to doubly substituted peptides. Solving for p gives

$$p = 0 \quad \textbf{[S5]}$$

or

$$p = \frac{2}{X+2} \quad \textbf{[S6]}$$

with the root of interest being the nonzero root. Substituting for the ratio of peak areas gives an estimate of the incorporation level p.

In some cases, peaks corresponding to peptides containing three leucine or leucine analogs were observed and used to quantify incorporation levels. In these cases, the peak area distribution is represented by

$$A[(1-p)^3 + 3(1-p)^2p + 3(1-p)p^2 + p^3] \quad \textbf{[S7]}$$

with

$$(1-p)^3 + 3(1-p)^2p + 3(1-p)p^2 + p^3 = 1. \quad \textbf{[S8]}$$

Defining the ratio of doubly substituted to triply substituted peak area as X, substituting X into Eq. **S8**, and solving for p yields the nonzero root

$$p = \frac{3}{X+3}, \quad \textbf{[S9]}$$

again enabling an estimation of the incorporation level of noncanonical amino acids in place of leucine.

In some cases, the tandem mass spectrometry did not enable positive identification of all possible positional isomers of a peptide. For example, in some cases, only two out of three of the possible doubly substituted positional isomers containing three possible substitution locations were identified in the tandem mass spectrometry data. In these cases, X was multiplied by an appropriate factor to account for peptides that were not observed (again invoking the assumption of completely random incorporation). Using the above example, when only two out of three doubly substituted peptides could be identified, X was multiplied by a factor of 1.5 in order to estimate what the peak area ratio would have been with all three peaks present in equal weights.

Using this methodology, the homoisoleucine-containing proteins used were found to have 90% or greater Hil in place of Leu, and the sample of S31L-T that was analyzed in this fashion contained approximately 99% Tfl in place of Leu. These results were obtained by examining three separate series of peptides from each protein sample. These peptides had the following sequences: AEIGDLNNTSGIR, GSHHHHHHGSMASGDLENEVAQLER, and SLEWEAAEL EQK (A1m), LLEWEAAELEQK (S31L), or SLEWEALELEQK (A37L).

1. O'Donnell MJ, Eckrich TM (1978) Synthesis of amino acid derivatives by catalytic phase-transfer alkylations. *Tetrahedron Lett.* 1978:4625–4628.
2. Dorizon P, et al. (1999) Stereoselective synthesis of highly functionalized cyclopropanes. Application to the asymmetric synthesis of (1*S*,2*S*)-2,3-methanoamino acids. *J. Org. Chem.* 64:4712–4724.
3. Tang Y, et al. (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. *Angewandte Chemie International Edition* 40:1494–1496.
4. Tang Y, Tirrell DA (2001) Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *J. Am. Chem. Soc.* 123:11089–11090.
5. Swiss Institute of Bioinformatics (2003) ProtParam tool. http://ca.expasy.org/tools/protparam.html.
6. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* 78:1606–1619.
7. Shen XH, Knutson JR (2001) Subpicosecond fluorescence spectra of tryptophan in water. *J. Phys. Chem. B* 105:6260–6265.
8. Zhong DP, Pal SK, Zewail AH (2001) Femtosecond studies of protein—DNA binding and dynamics: Histone I. *Chemphyschem* 2:219–227.
9. Steiner RF (1991) *Top. Fluoresc. Spectrosc.*, ed Lakowicz JR (Plenum Press, New York), Vol 2, pp 1–52.

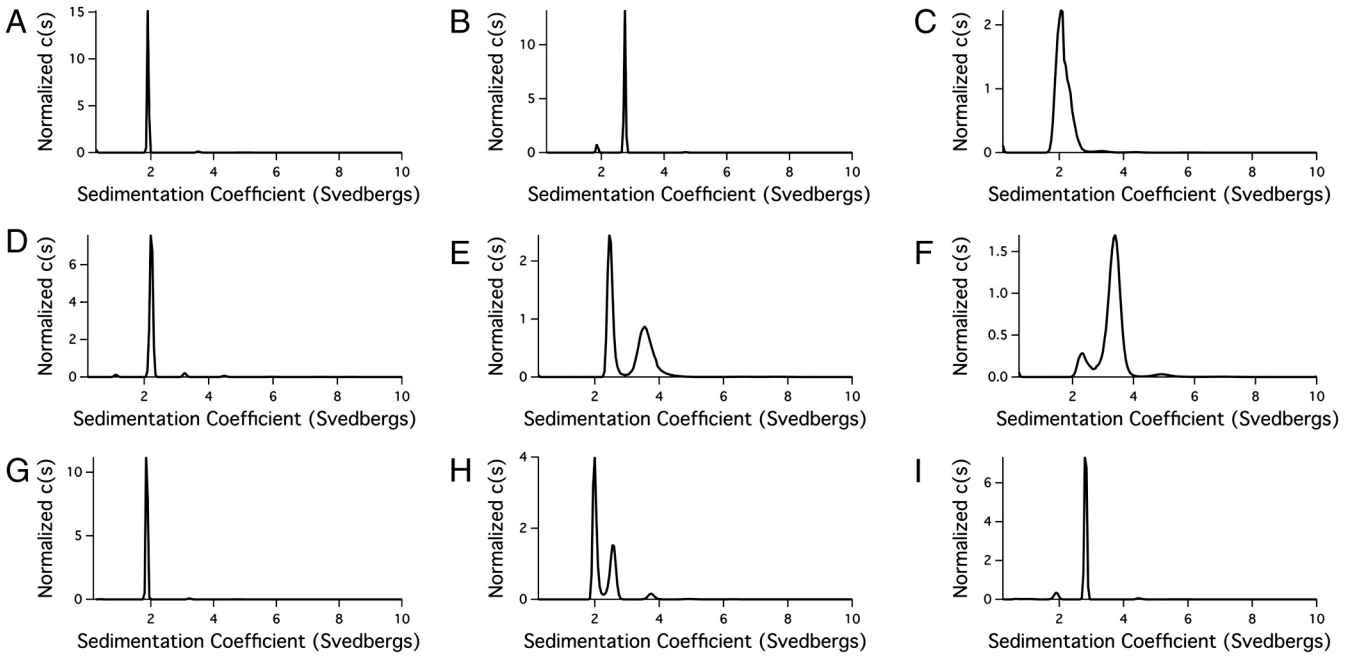

Fig. S1. Normalized plots from the Sedfit $c(s)$ analysis for A1m-L (*A*), S31L-L (*B*), A37L-L (*C*), A1m-T (*D*), S31L-T (*E*), A37L-T (*F*), A1m-H (*G*), S31L-H (*H*), and A37L-H (*I*). The protein samples were prepared at 550 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution. The broad, single peak of the A37L-L trace may be the result of an equilibrium mixture between trimeric and higher-order species. It seems likely that the breadth of the traces derived from sedimentation velocity analysis of A1m-T, S31L-T, and A37L-T is a consequence of incomplete replacement of Leu with Tfl.

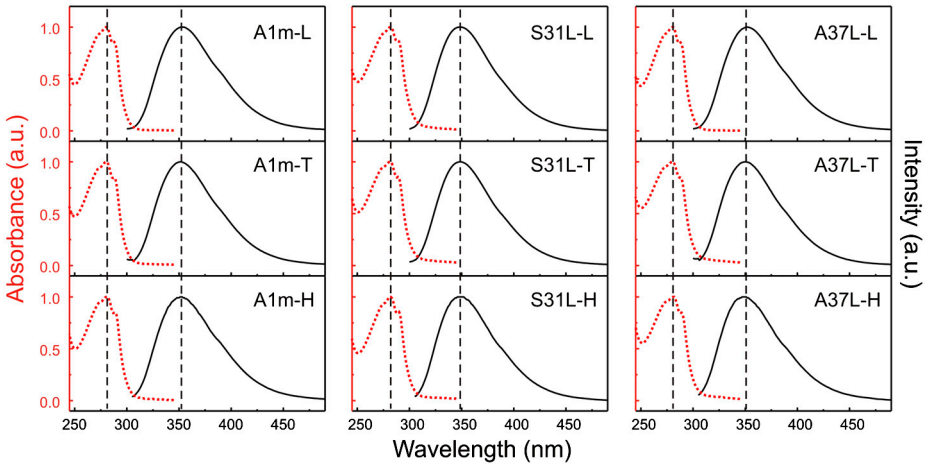

Fig. S2. Steady-state UV-visible absorption (red) and fluorescence emission spectra (black) of proteins excited at 295 nanometers. The protein samples were prepared at 550 μM concentration in 10 mM acetate (pH 4)/100 mM NaCl solution.

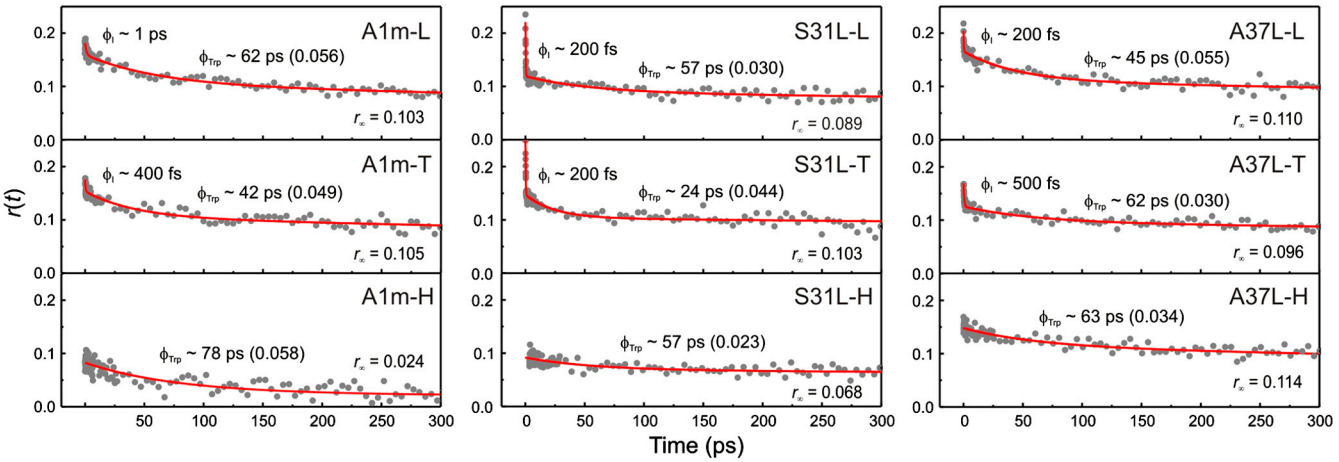

Fig. S3. Time-resolved anisotropy, $r(t)$, of the proteins. All anisotropy decays were fitted to $r(t) = r_I \exp(-t/\phi_I) + r_{Trp} \exp(-t/\phi_{Trp}) + r_\infty$, where -$r_I$ is the initial ultrafast anisotropy, r_{Trp} is the Trp motion-related anisotropy (value given in parentheses in figure), r_∞ is the offset anisotropy, ϕ_I is the initial ultrafast internal-conversion time constant of Trp (≤1 ps), and ϕ_{Trp} is the Trp-rotational correlation time constant.

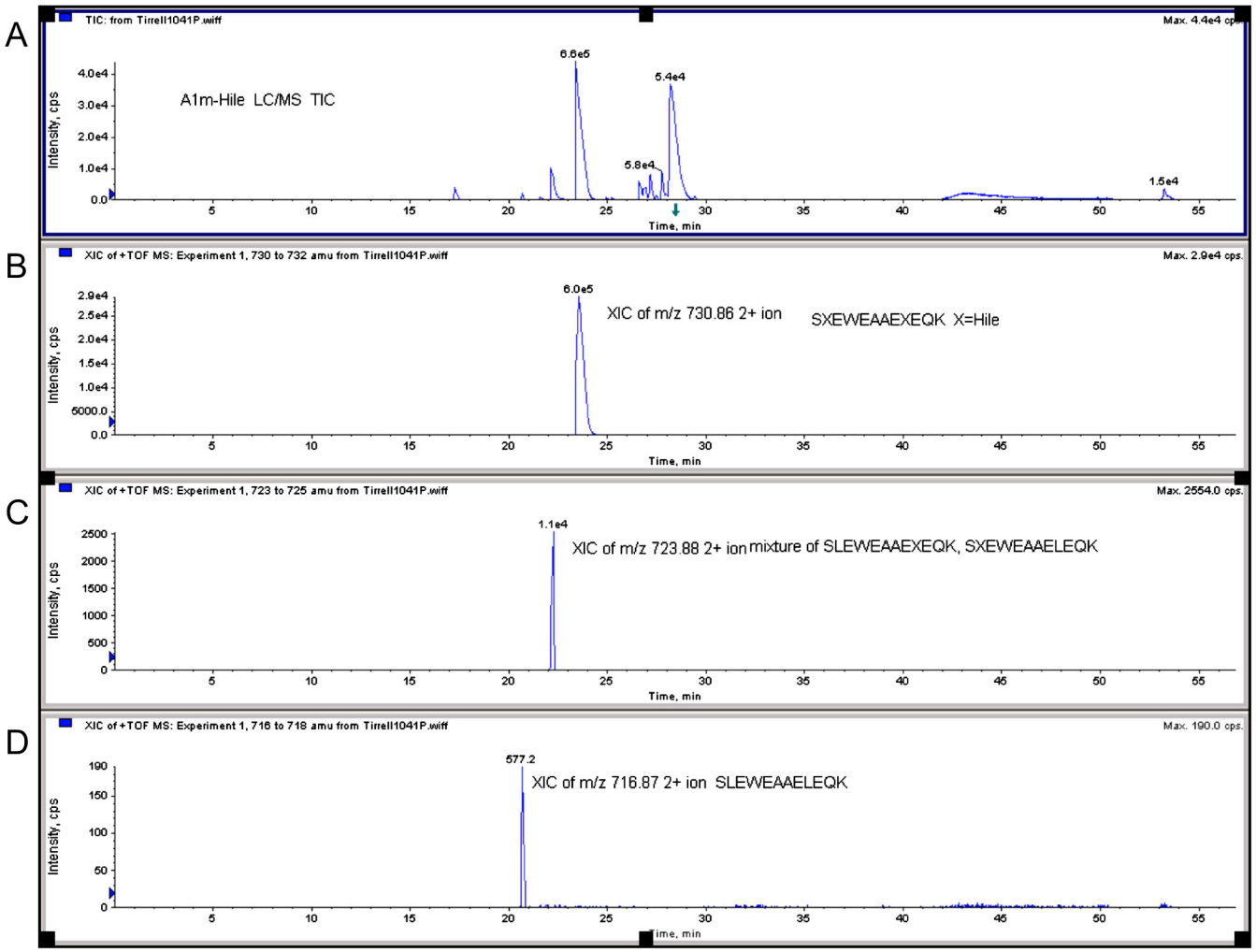

Fig. S4. LC/MS/MS of trypsinized A1m-H. (*A*) TIC of digested protein sample. (*B*)–(*D*) XICs of peptides containing two (*B*), one (*C*), or no (*D*) leucine to homo-isoleucine substitutions in the peptide SLEWEAAELEQK. The ratios of the peak areas obtained in the XICs can be used to estimate the extent of leucine replacement in the protein sample. Peptide masses: SXEWEAAEXEQK: 2+ ion: 730.86 Da observed, 730.87 Da expected. SLEWEAAEXEQK, SXEWEAAEXEQK: 2+ ion: 723.88 Da observed, 723.86 Da expected. SLEWEAAELEQK 2+ ion: 716.87 Da observed, 716.85 expected.