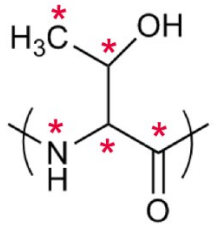
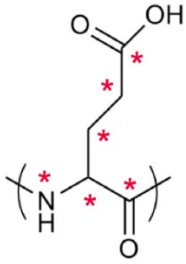
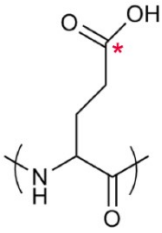
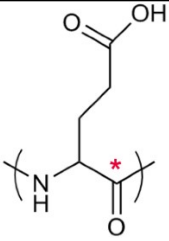
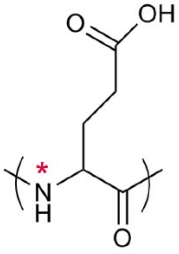
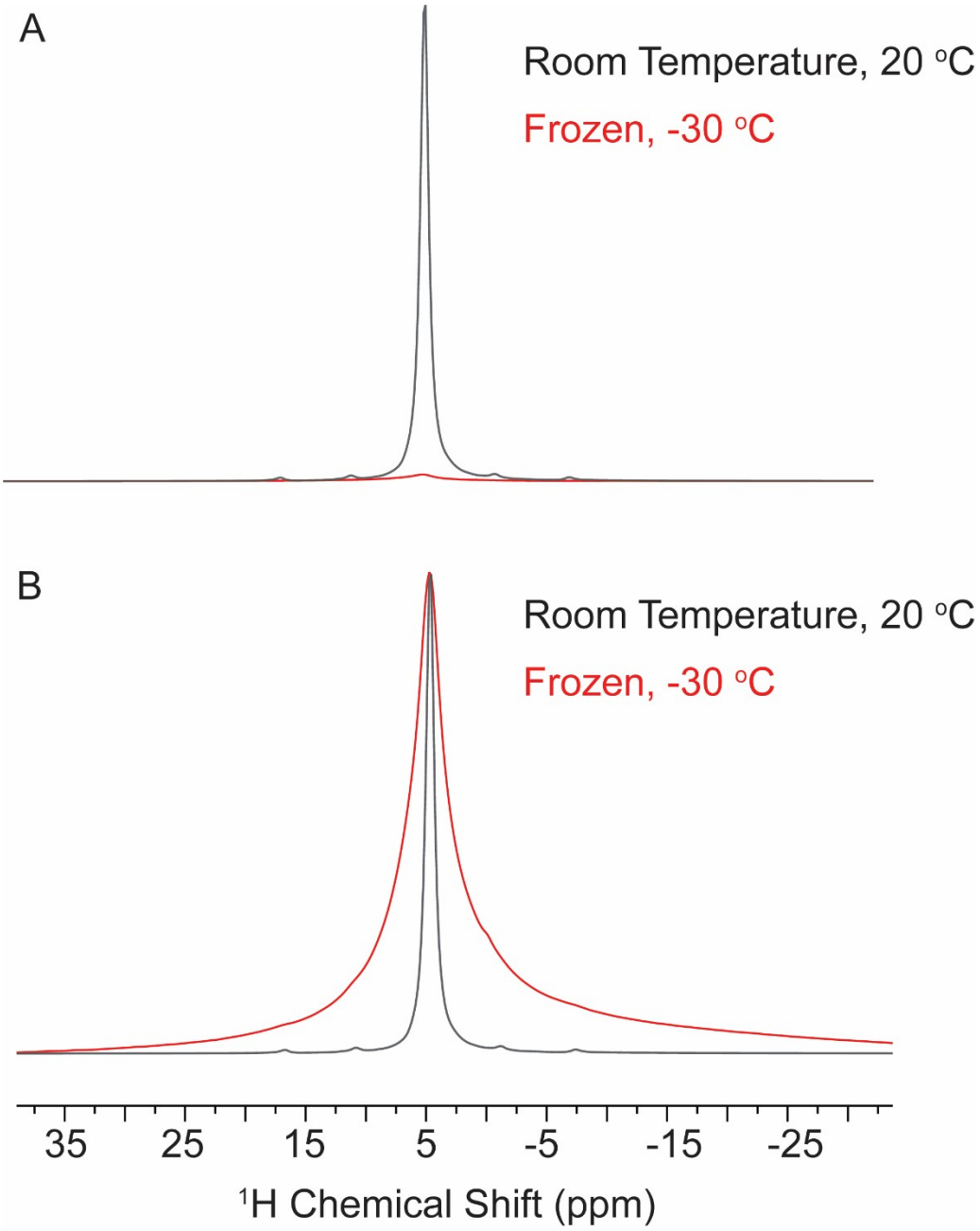
Elucidating molecular level interfacial interactions between a *de novo* protein and nucleated calcite with solid state NMR

Emily G. S. Close[1], Andrew S. Lipton[1], Marlo Zorman[2], Harley Pyles[3], Ying Chen[1], Thi Kim Hoang Trinh[1], Garry W. Buchko[1,4], Chun-Long Chen[1], Christopher J. Mundy[1,2], David Baker[3], Wendy J. Shaw[1]

1) Pacific Northwest National Laboratory, Richland, WA
2) Department of Chemical Engineering, University of Washington, Seattle, WA
3) Institute for Protein Design, University of Washington, Seattle, WA
4) School of Molecular Biosciences, Washington State University, Pullman, WA

Supplemental Table 1. Description of FD31 and FD31* names and corresponding labeling schemes.

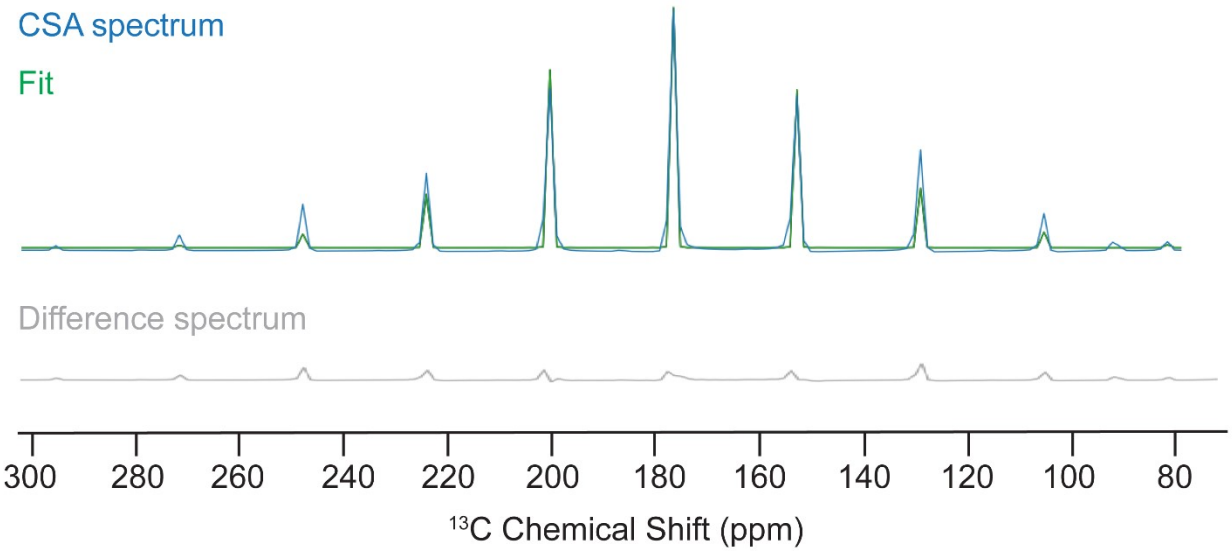
Name	Description	Location of labels
FD31-U-T	All threonine carbons and nitrogens are isotopically labeled in FD31	
FD31-U-E	All glutamic acid carbons and nitrogens are isotopically labeled in FD31	
FD31*-ECδ	The side chain carbonyl is ^{13}C labeled in glutamic acid in FD31*	
FD31*-EC′	The backbone carbonyl is ^{13}C labeled in glutamic acid in FD31*	
FD31*-EN	The backbone nitrogen is ^{15}N labeled in glutamic acid in FD31*	



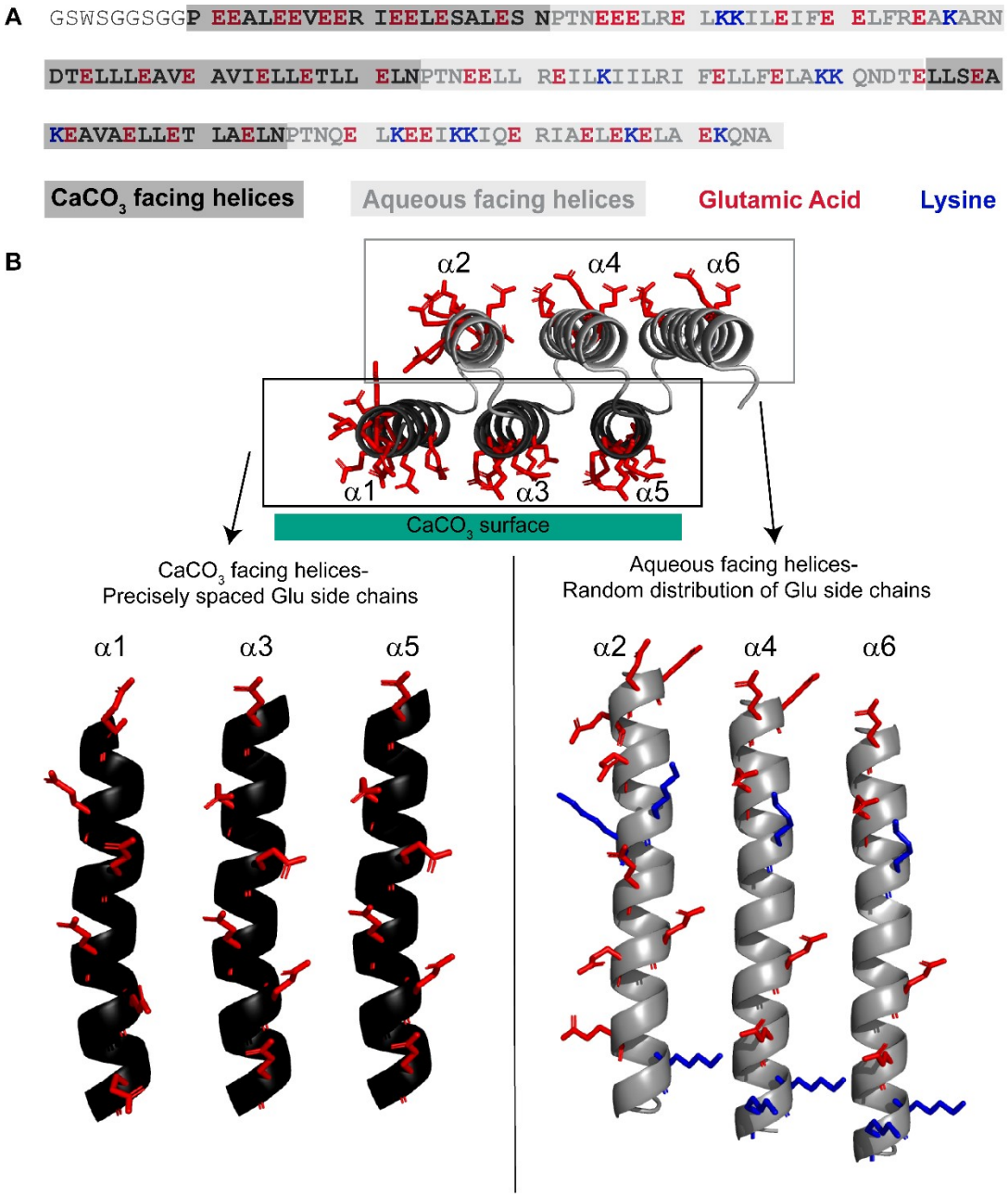
Supplemental Figure 1. Examples of ^{1}H NMR spectra of water present in FD31-nucleated calcite samples that confirm they were frozen. (A) Liquid water at room temperature (black) has a much sharper resonance compared to frozen water (red), resulting in a much higher signal intensity. (B) Spectra from (A) with the frozen water resonance scaled to the same height as the liquid water resonance highlight the differences in line broadness. Spectra were collected frozen (-30 °C; red) and at room temperature (20 °C; black) spinning at 5 kHz on an Agilent VNMRS spectrometer operating at 300 MHz ^{1}H resonance frequency.

Supplemental Table 2. Herzfeld Berger Analysis (HBA) for ^{13}C-labeled Glycine. σiso = (1/3)(σ11 + σ22 + σ33), η = (σ22 – σ11)/(σ33 – σiso), where |σ11 – σiso| < |σ33 - σiso|, and Ω = σ11 - σ33. All HBA parameters are defined in Figure 3. HBA fits are shown in Figure S5.

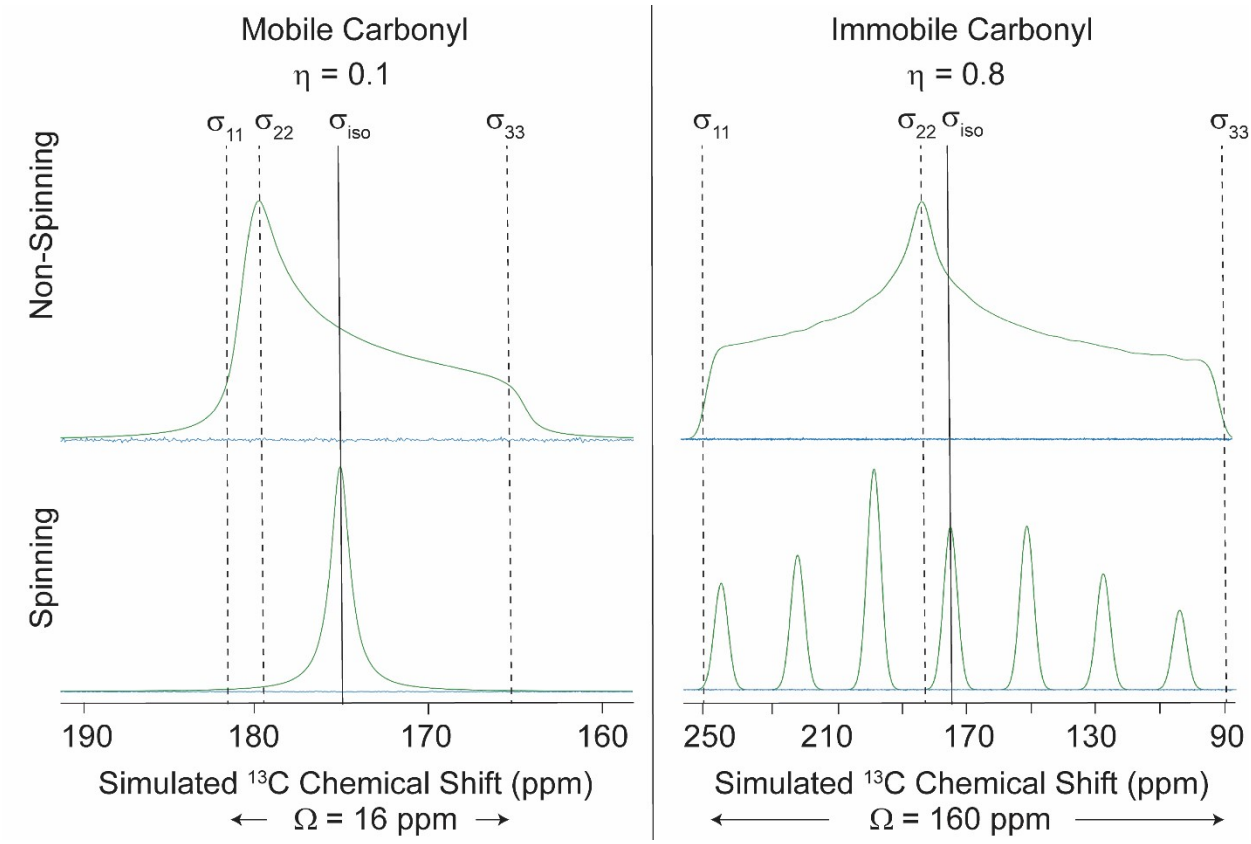
	Ω (ppm)	η	σ_{iso}	Linewidth (ppm)	Result	nRMSD
Glycine, 20°C	112.2	0.9	175.5	3.5	Immobile	0.03



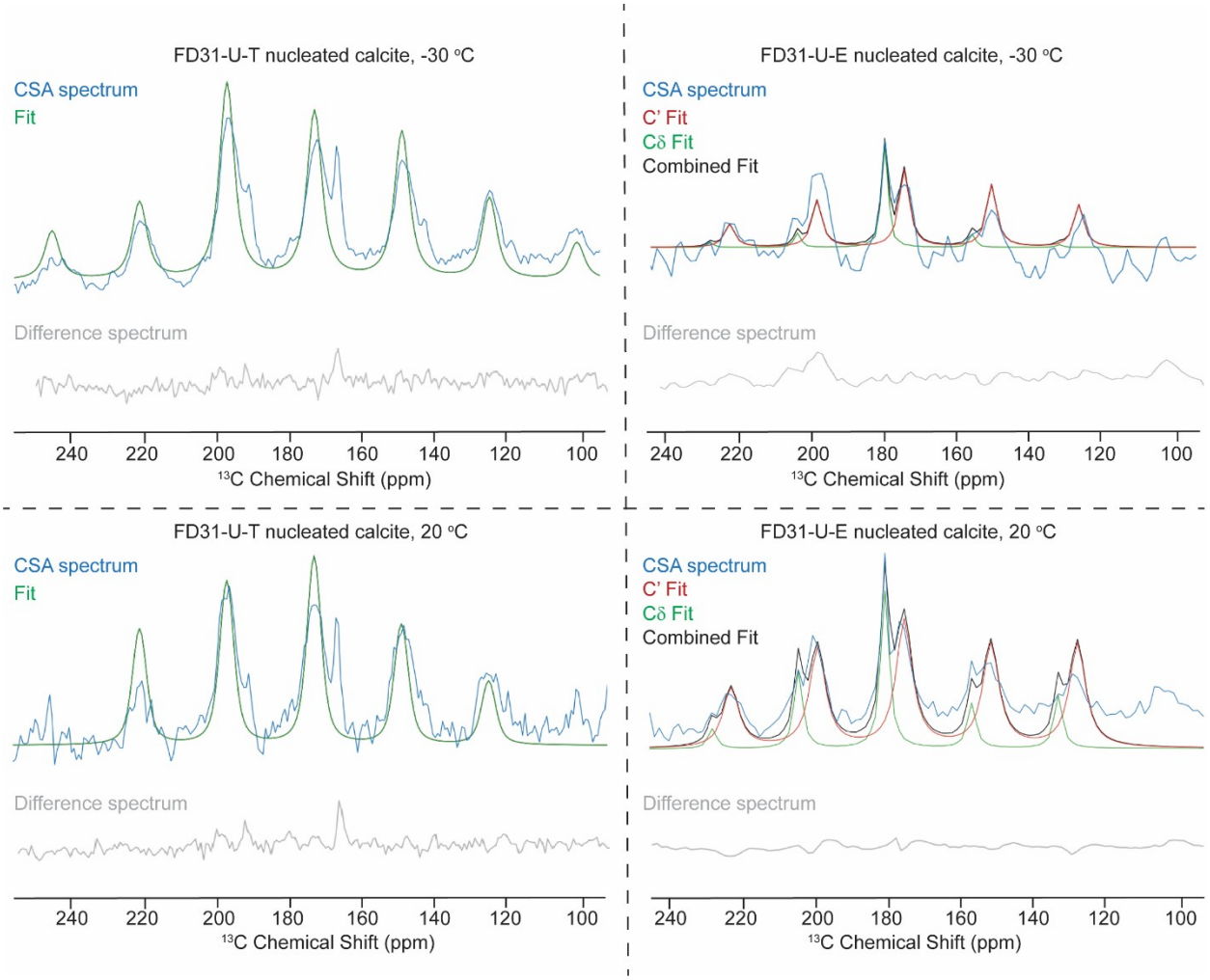
Supplemental Figure 2. Fits showing agreement between CSA spectra (blue) with the Herzfeld-Berger Analyses (green) for the uniformly ^{13}C-labeled glycine control sample. The ^{13}C {^{1}H} CP-MAS spectra was collected with 3.2k scans spinning at 1.8 kHz on an Agilent VNMRS spectrometer operating at a ^{1}H resonance frequency of 300 MHz at room temperature (20 °C). Difference spectra (gray) show overall good agreement between the raw data and the HBA fit. The nRMSD value for this fit (0.03) serves as a reference for an expected nRMSD value corresponding to a high-quality fit.



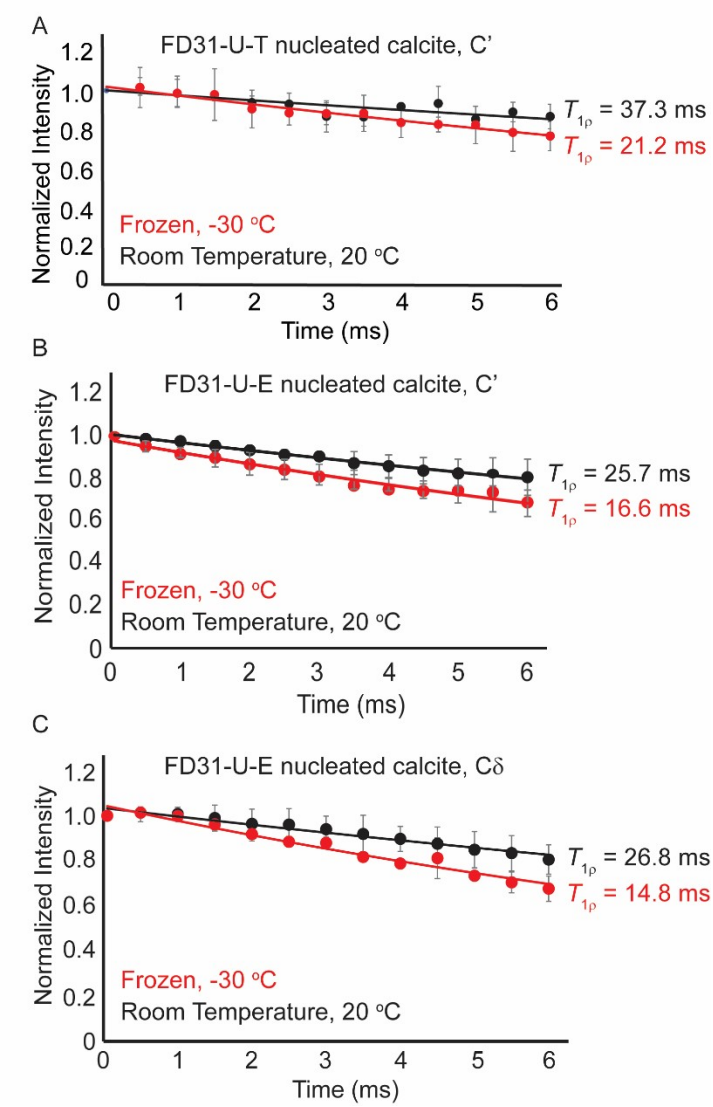
Supplemental Figure 3. (A) Primary amino acid sequence for FD31 with $CaCO_3$ facing helices highlighted in shaded dark grey, aqueous facing helices shaded light grey, and glutamic acid residues highlighted in red. (B) Surface of the $CaCO_3$ binding face (left; α1, α3, and α5; black cartoon) and aqueous exposeed face (right; α2, α4, and α6; grey cartoon), with the glutamic acid side chain shown in red sticks and lysine residues shown as blue sticks. The $CaCO_3$ binding face contains precisely spaced glutamic acid residues capable of lattice matching to calcium ions while the aqueous face contains randomly spaced glutamic acid residues that are not lattice matched to $CaCO_3$. The aqueous facing region also contains positively charged residues that help to proclude potential electrostatic interactions between negatively charged glutamic acid residues and positively charged calcium ions in the surface.



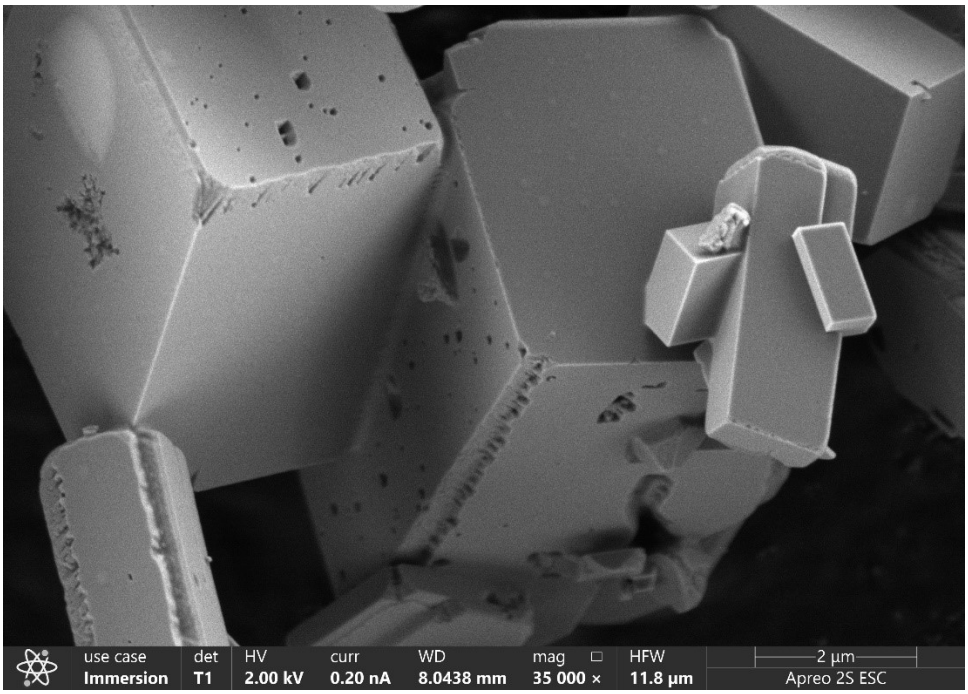
Supplemental Figure 4. Simulated data for a mobile versus immobile ^{13}C carbonyl CSA pattern with corresponding Herzfeld-Berger parameters labeled. For a mobile carbonyl, the symmetry value (η) will be closer to 0 and the span (Ω) will be relatively small compared to an immobile system, where the symmetry value η will be closer to 1 and the Ω will be larger. CSA parameters σ_{11}, σ_{22}, σ_{33}, σ_{iso} and Ω are labeled on the simulated data.



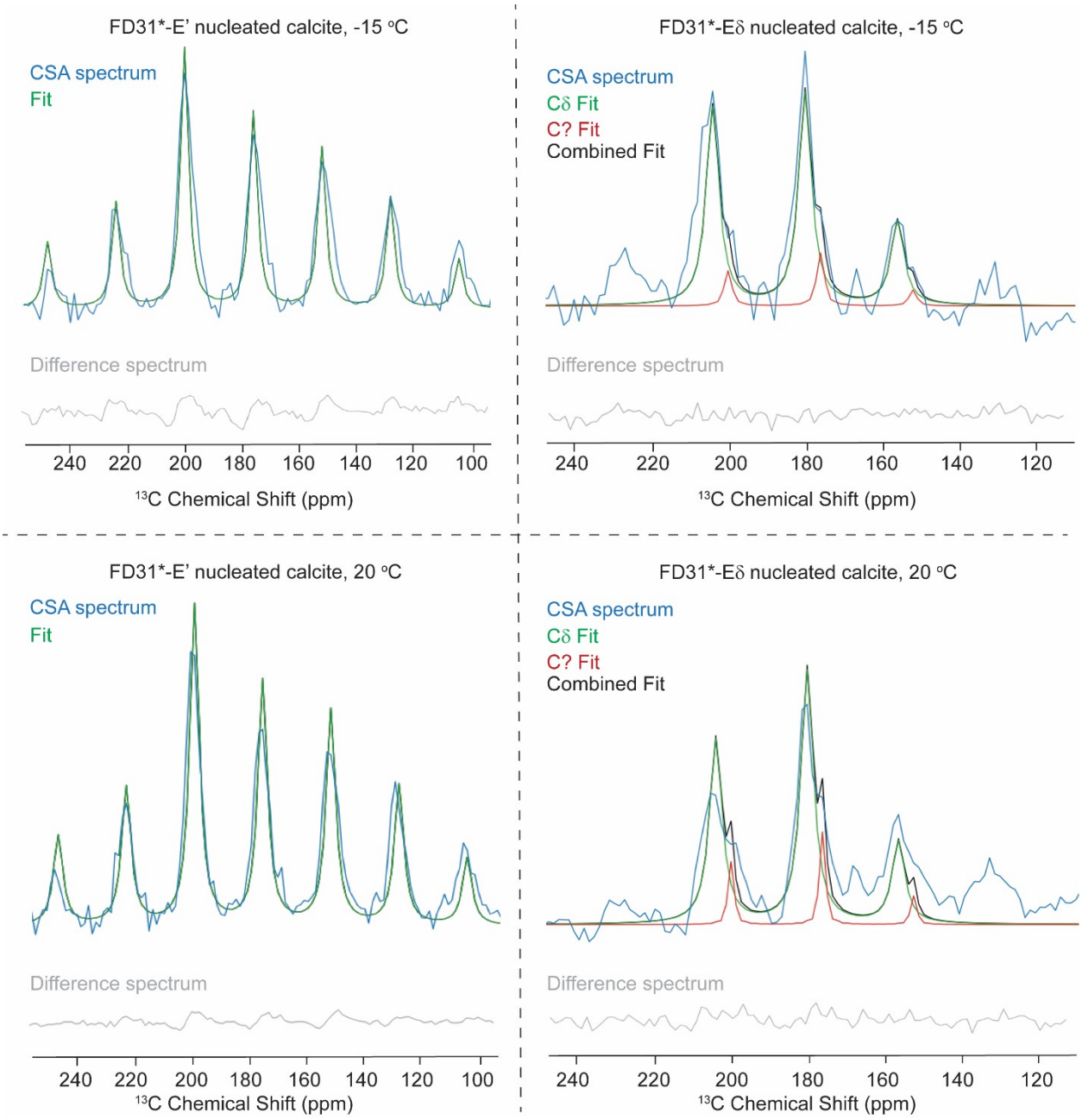
Supplemental Figure 5. Fits showing agreement between CSA spectra (blue) with the Herzfeld-Berger Analyses of FD31-U-T and FD31-U-E nucleated calcite samples from Figure 2 and Table 1. Difference spectra (gray) are plotted below each corresponding fit and show an overall good agreement between the raw data and the HBA fit for each construct. The strong residual peak in the FD31-U-T difference spectra at both temperatures corresponds to calcite. The ^{13}C $\{^{1}H\}$ CP-MAS spectra were all collected with 10 k scans spinning at 1.8 kHz on an Agilent VNMRS spectrometer operating at a ^{1}H resonance frequency of 300 MHz at either room temperature (20 °C) or frozen (-30 °C).



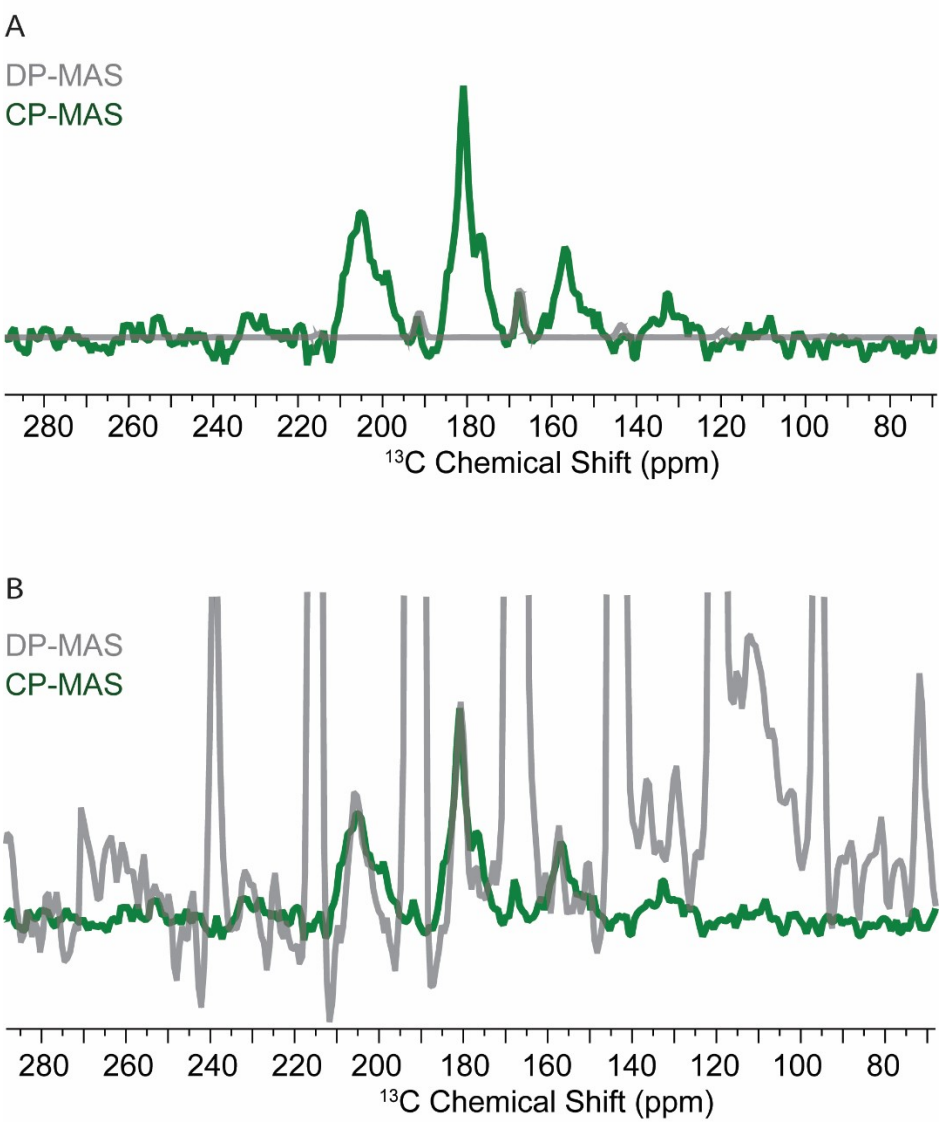
Supplemental Figure 6. The $^{13}C'$ $T_{1\rho}$ plots for (A) FD31-U-T, (B) FD31-U-E'C′ and (C) FD31-U-E-Cδ nucleated calcite. The $T_{1\rho}$ experiments were collected with three trials and averaged together to calculate the rate constant. Points represent average normalized intensity for each spin lock time and error bars represent the experimental standard deviation. Spectra were collected frozen (-30 °C; red) and at room temperature (20 °C; black) spinning at 5 kHz on an Agilent VNMRS spectrometer operating at a ^{1}H resonance frequency of 300 MHz.



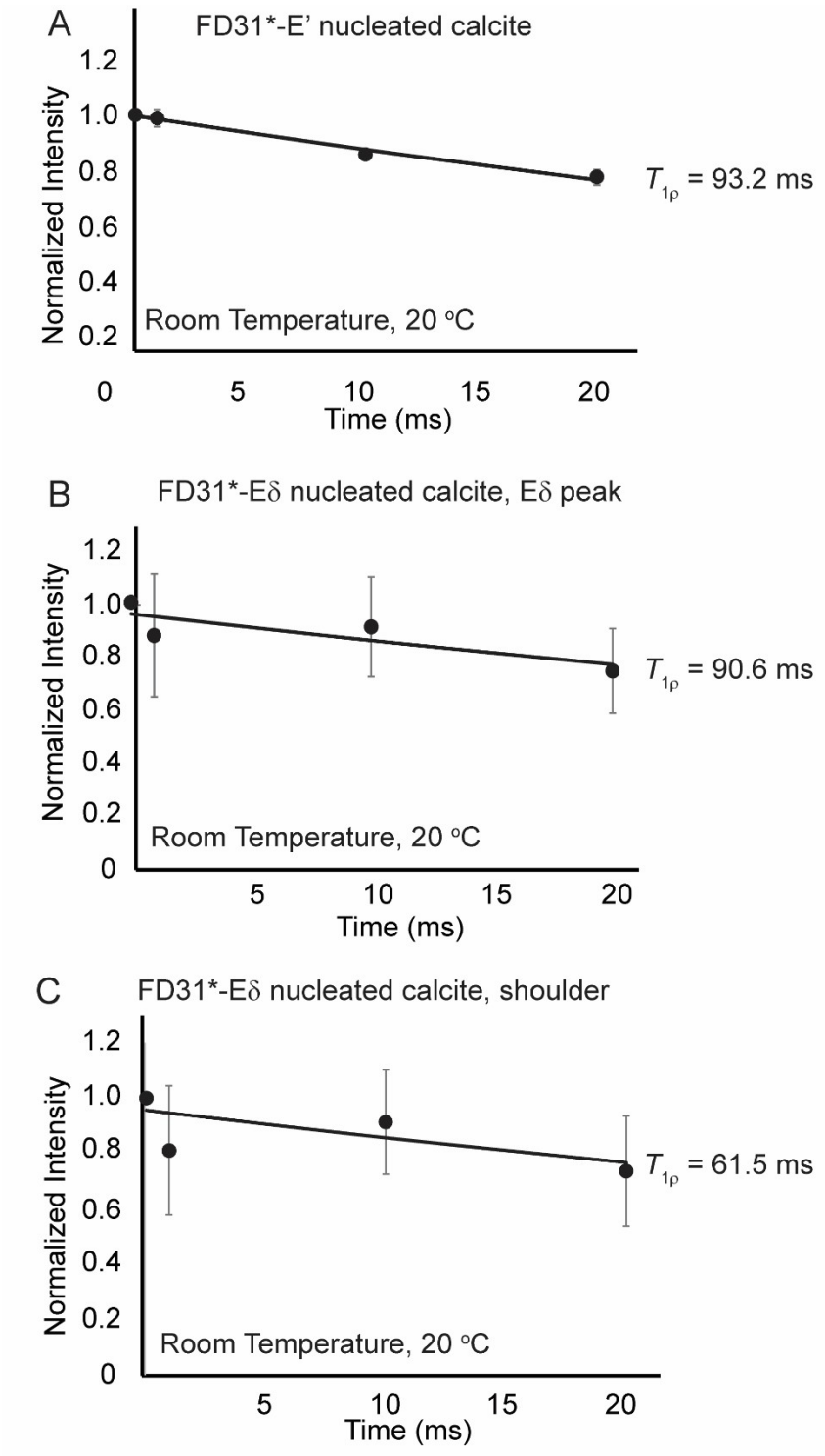
Supplemental Figure 7. SEM of calcite prepared in the absence of protein. Without the protein additive, the calcium carbonate formed appears to have the characteristic rhombohedron morphology of calcite {104}.



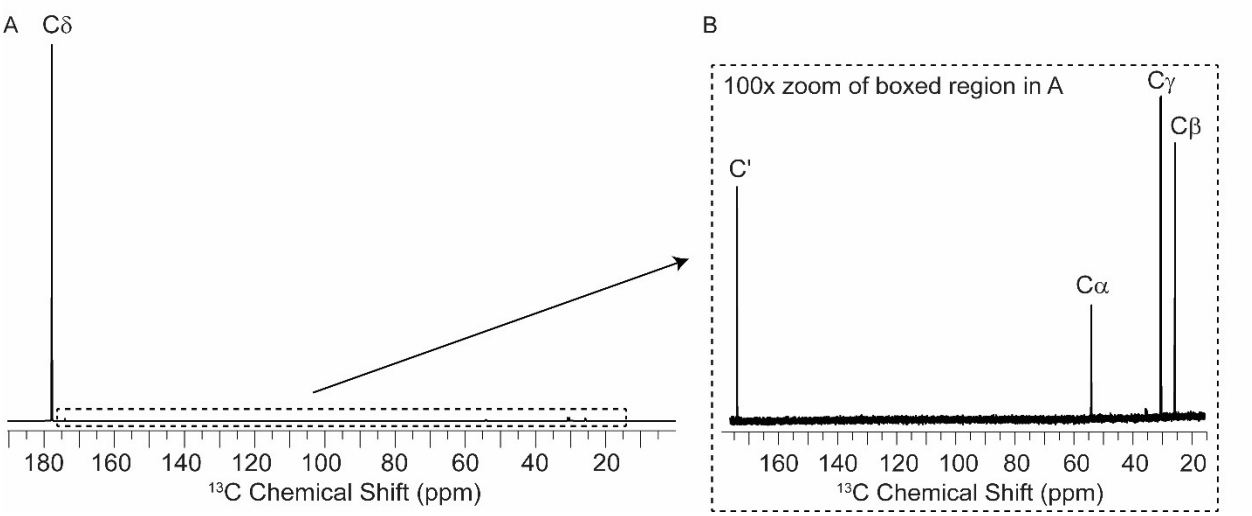
Supplemental Figure 8. Fits showing agreement between CSA spectra (blue) with the Herzfeld-Berger Analyses from Figure 4 and Table 2. Fits for the glutamic acid C′ (left) and Cδ (right) are shown in green and fits for the shoulder tetnatively assigned as the glutamine Cδ (right) are shown in red. Difference spectra (gray) are plotted below each corresponding fit and show an overall good agreement between the raw data and the HBA fit for each construct. The $^{13}C\{^{1}H\}$ CP-MAS spectra were all collected with 10k scans spinning at 1.8 kHz on an Agilent VNMRS spectrometer operating a ^{1}H resonance frequency of 300 MHz at either room temperature (20 °C) or frozen (-15 °C).



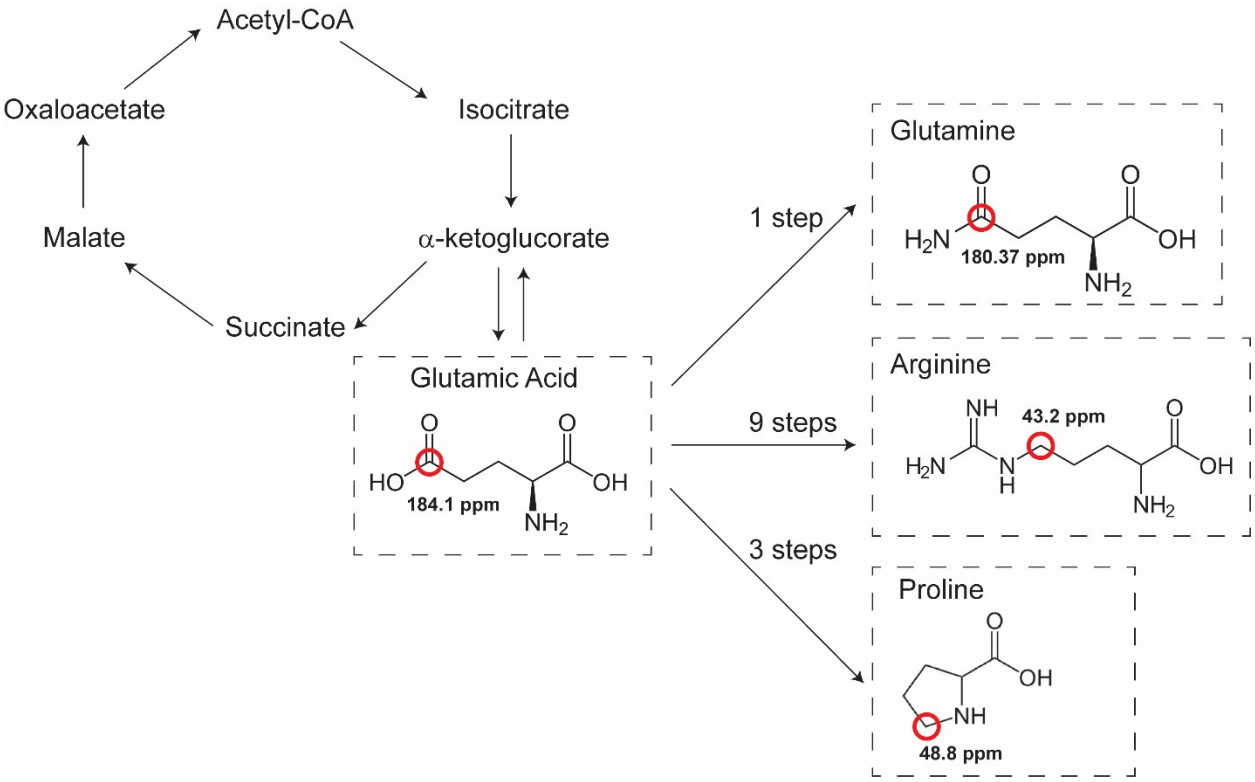
Supplemental Figure 9. The $^{13}C\{^{1}H\}$ cross-polarization magic angle spinning (CP-MAS) spectrum and direct-excitation magic angle spinning (DP-MAS) spectrum of the carbonyl region of FD31*-ECδ with nucleated calcite. Both samples were prepared under identical nucleation conditions, but the CP-MAS spectrum was acquired from a sample with natural abundance calcite while the DP-MAS spectrum was acquired from a sample with ^{13}C-labeled calcite. Spectra are scaled by (A) the calcite peak or (B) the side-chain carbonyl peak. Due to longer relaxation times and fewer scans, the DP-MAS spectrum was collected using 60K scans and required 5 days of signal averaging. Despite the lower spectral resolution in the DP-MAS spectra, the CSA pattern of the glutamic acid side-chain carbonyl appears to be identical in both the CP-MAS and DP-MAS spectra.



Supplemental Figure 10. The $T_{1\rho}$ plots for (A) FD31*-EC′ and FD31*-ECd nucleated calcite, (B) main peak, and (C) shoulder. The $T_{1\rho}$ experiments were collected with two trials and averaged together to calculate the rate constant. Points represent average normalized intensity for each spin lock time and error bars represent the experimental standard deviation. Spectra were collected at 20 °C spinning at 5 kHz on an Agilent VNMRS spectrometer operating at a ^{1}H resonance frequency of 300 MHz.



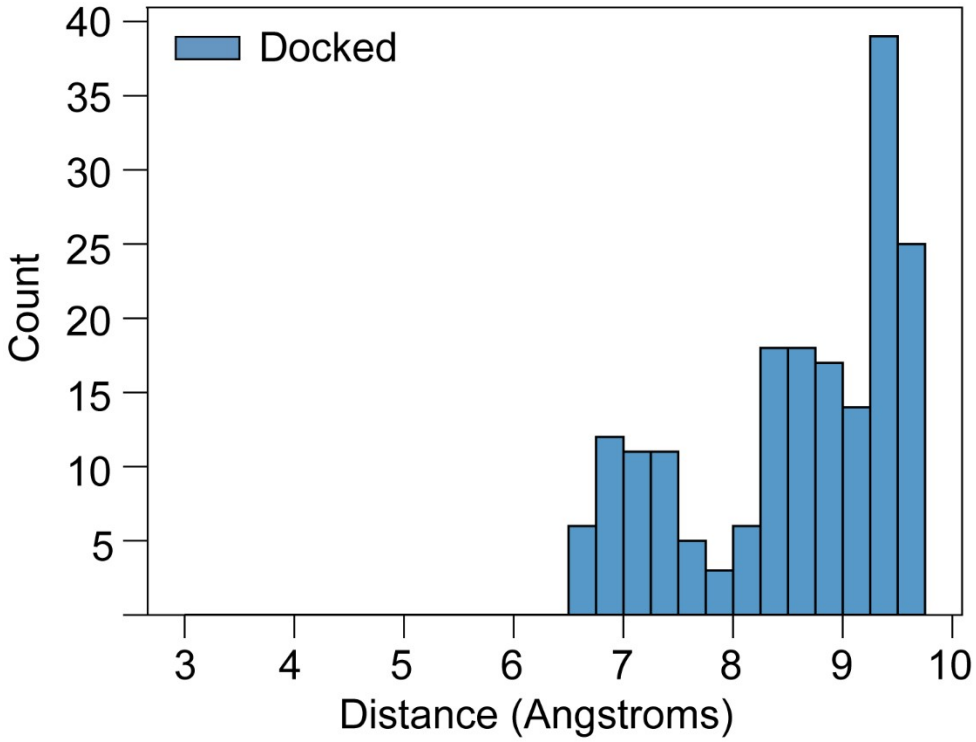
Supplemental Figure 11. One-dimension ^{13}C solution NMR spectrum of 5-^{13}C-Glu used as starting material for recombinant protein expression of FD31*-Eδ in *E. Coli*. (A) Full spectrum shows a dominant peak for the Cδ of glutamic acid. (B) Expansion of the dashed spectral region in (A) shows the other carbons in glutamic acid are visible via natural abundance but are significantly weaker in intensity than the labeled Cδ.



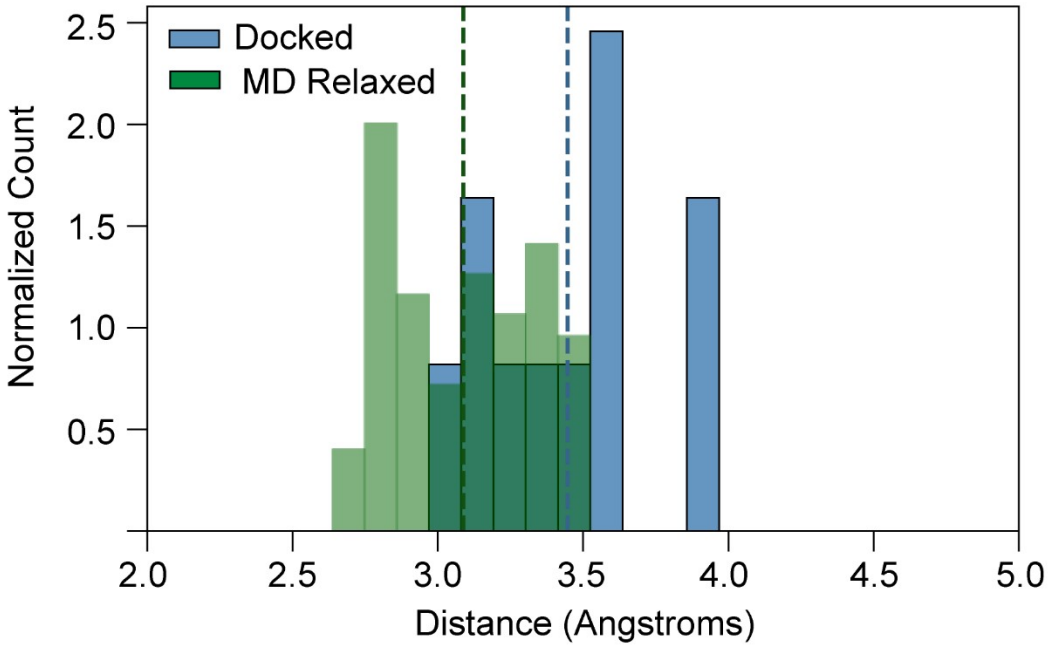
Supplemental Figure 12. Biosynthentic pathway showing that glutamic acid is a precursor for glutamine, arginine, and proline. The location of the isotopically labeled ^{13}C in the original 5-^{13}C-Glu starting material and in each of the scrambled amino acids is highlighted with a red circle, and the corresponding average chemcial shift for each ^{13}C is shown in bold. Only the scrambled ^{13}C in glutamine has a chemical shift in the frequency range that matches the shoulder in Figure 4C, therefore, if the shoulder was a result of isotopic scrambling it could only correspond to glutamine $^{13}C\delta$. Additionally, scrambling of glutamic acid in to arginine requires nine steps, to proline requires three steps and to glutamine requires just one step, making glutamine the most likely amino acid to result from scrambling of glutamic acid.

Supplemental Table 3. REDOR distance measurements and simulations.

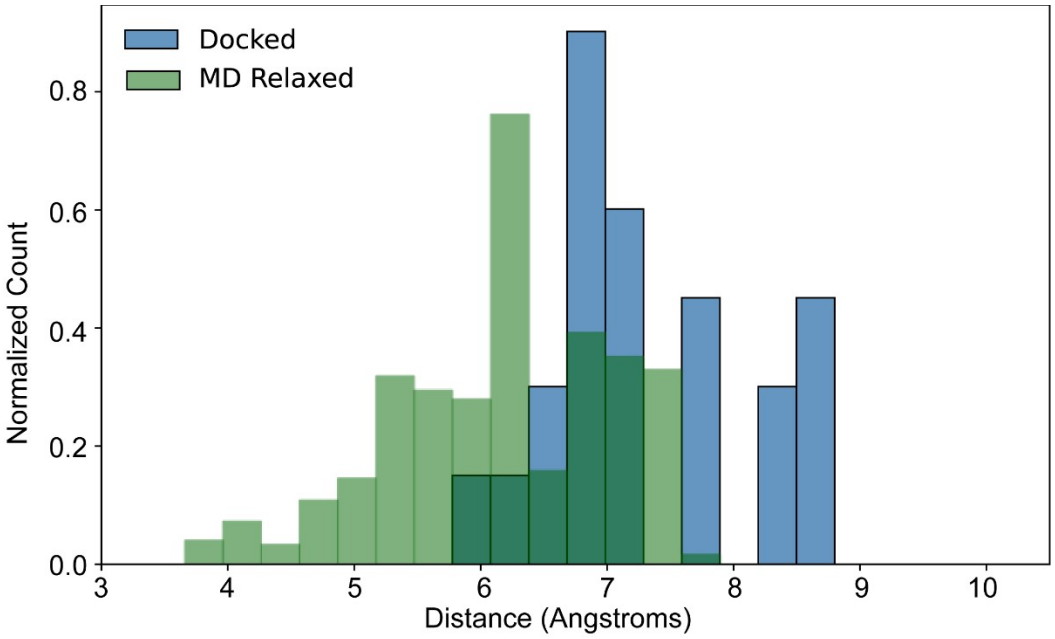
	Simulation					Experimental	
	5.7Å	5.4Å	5.1Å	4.8Å	4.5Å	^{15}N-Glu mutant nucleated Ca^{13}CO3	
Rotor Cycle	S/S0	S/S0	S/S0	S/S0	S/S0	S/S0	st dev
0.00	1.00	1.00	1.00	1.00	1.00	0.99	0.03
32.00	1.00	1.00	0.98	1.00	1.00	0.97	0.04
64.00	0.99	0.98	0.91	0.96	0.94	0.87	0.08
96.00	0.95	0.93	0.81	0.86	0.80	0.86	0.04
128.00	0.82	0.76	0.68	0.56	0.41	0.78	0.09
χ^2	0.0236	0.0174	**0.0139**	0.0541	0.1426		



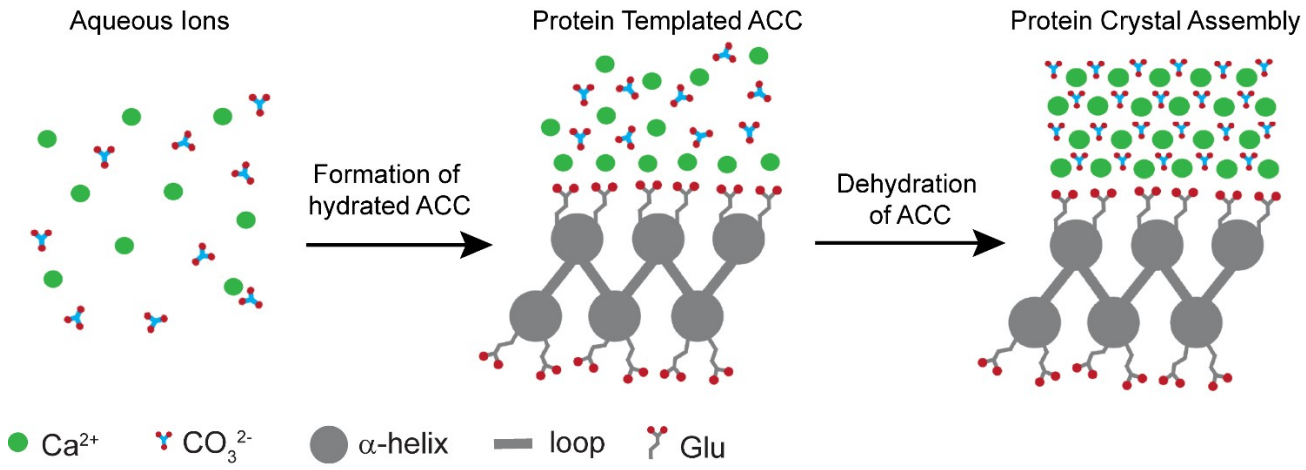
Supplemental Figure 13. Measurement wizard in PyMOL was used to measure distances between nitrogen atoms in the glutamic acid backbone to carbons in the calcite surface. The average distance range for this interaction spanned ~6.5 to 10 Å.



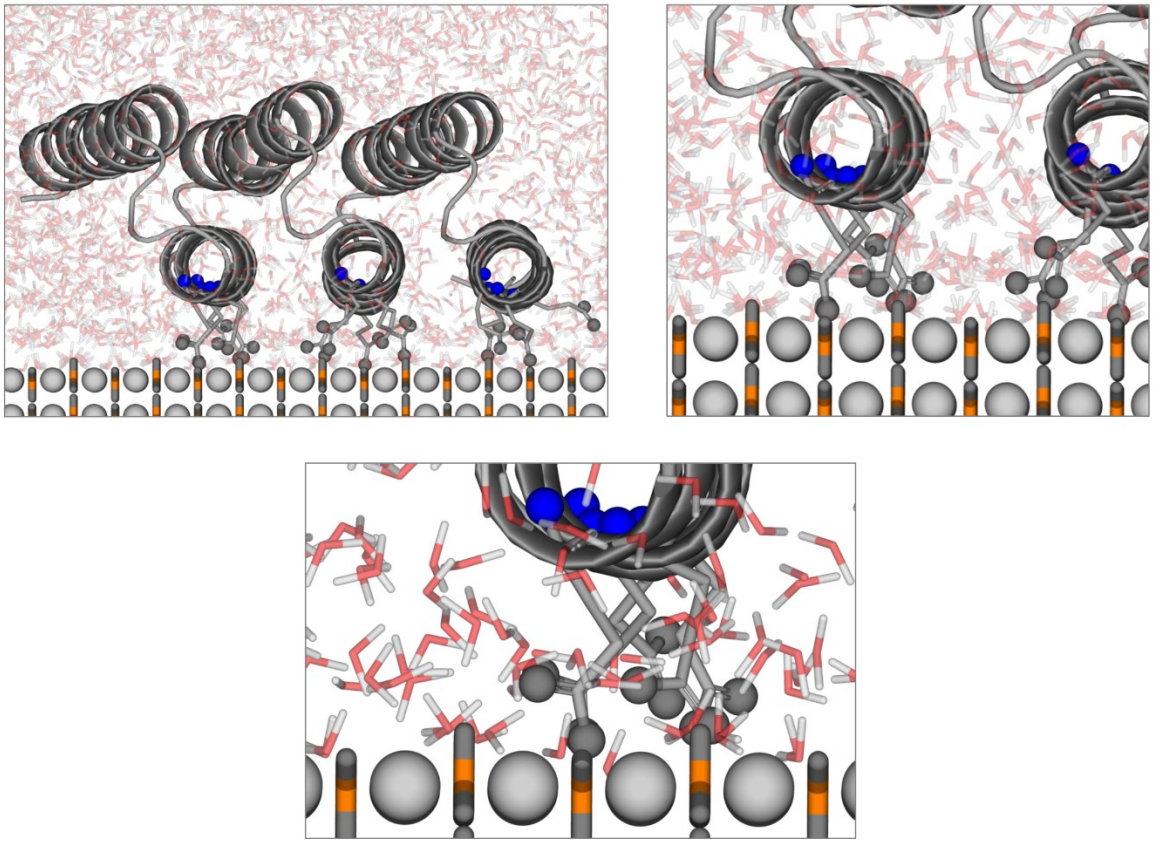
Supplemental Figure 14. Average distances of the FD31* δ-carbon from carbon in the calcite surface measured from a 100 ns MD simulation in which FD31 was docked onto calcite{110} in the presence of water.



Supplemental Figure 15. Average distances between the FD31* backbone nitrogen atoms and the calcite surface carbons from a 100 ns molecular dynamics (MD) simulation. In this simulation, FD31 was initially docked onto the calcite {110} surface in the presence of water, with the protein positioned 4 Å from the surface. This initial positioning created a sterically hindered conformation, where sidechains clashed with the surface. Subsequent energy minimization allowed the sidechains to relax into a more stable conformation, resulting in protein-surface distances that closely resembled those observed in NMR experiments. These findings confirm that an average backbone-to-surface distance of ~5.1–5.7 Å is achievable when the sidechains are fully elongated.



Supplemental Figure 16. Negatively charged glutamic acid sidechains in the protein template interact directly with positively charged calcium ions in solution, promoting the nucleation of dehydrated amorphous calcium carbonate (ACC). Subsequently, the ACC undergoes structural rearrangement and is directly transformed into calcite {110}.



Supplemental Figure 17. A 100 ns molecular dynamic simulation was run of FD31*-E^{15}N on Ca$^{13}CO_3$ in the presence of water. While glutamic acid sidechains appear to displace some surface bound waters to interact with calcite directly, residual water molecules still remain. Water molecules are shown as red and white sticks.