Supplementary Information

Impact of Frequent ARID1A Mutations on Protein Stability Provides Insights into Cancer Pathogenesis

Rajen K Goutam[1], Gangtong Huang[1], Exequiel Medina[2], Feng Ding[1], William J. Edenfield[3], Hugo Sanabria[1,*]

1. Department of Physics and Astronomy, Clemson University, Clemson, 29634, SC, USA
2. Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Chile.
3. Institute for Translational Oncology Research, Prisma Health, Greenville, SC, USA.

* Corresponding author: hsanabr@clemson.edu

Table of contents

Table 1: Proteins with ARID domains used in our study to sort the frequency of missense mutation.

Number	Name (protein id, range)
1	P41229-1 (76-165)
2	A6NKF2(110-201)
3	Q9UGL (94-183)
4	Q92833-2 (618-709)
5	KDM5C (21-~150)
6	Q4LE (303-392)
7	Q8NFD5-2(1050-1140)
8	P29374(306-397)
9	Q9BY66-1(80-170)
10	014497 (1018-1109)

Table 2: The primers for obtaining the mutants using plasmid of ARID (WT).

Mutants	sequence
R1020S_F	5’taagatgtgggttgatCGCTATCTGGCTTTCACTG3’
R1020S_R	5’ctttccggctcaccaccTAACTCGTACAGTTTGGTG3’
M1022K_F	5’gtgggttgatcgctatCTGGCTTTCACTGAGGAAAAAG3’
M1022K_R	5’ttcttacgttccggctcACCACCTAACTCGTACAG3’
K1047Q_F	5’caaccgctggacctgTACCGCCTGTACGTGAGC3’
K1047Q_R	5’acggccaactgccggCAGGTTCGTCATACCCATGG3’
G1063V_F	5’tttgacccaggttaatAAGAACAAGAAGTGGCGTGAATTG3’
G1063V_R	5’acgccaatttctttgacGCTCACGTACAGGCGGTA3’
A1089T_F	5’ctcgagcgctACCagctccctga3’
A1089T_R	5’tcagggagctGGTagcgctcgag3’

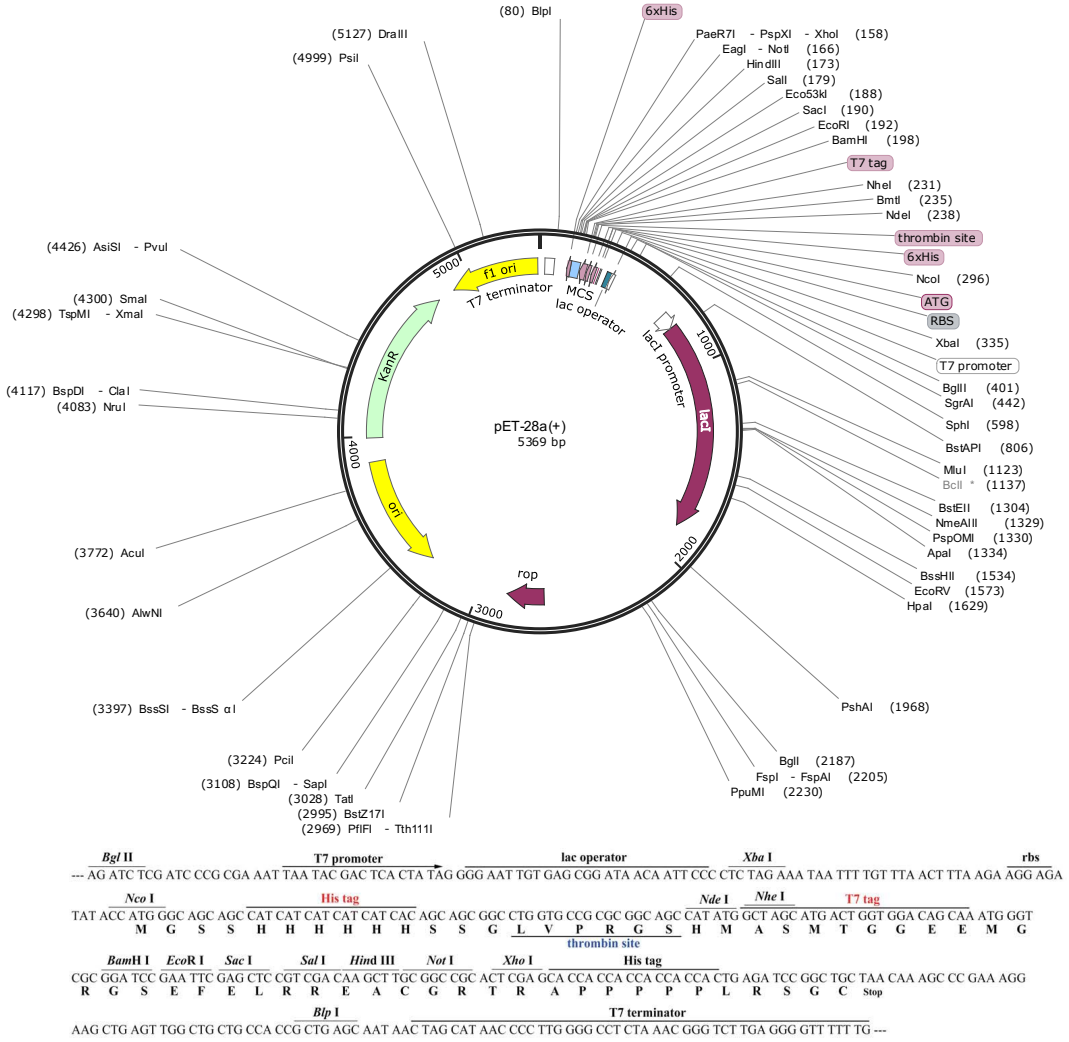

Fig1 S1: Plasmid vector pET-28a (+) and linear restriction sites. The top section shows the circular vector with unique restriction sites, an N-terminal His-Tag/thrombin/T7-Tag configuration. The T7 expression region is reversed due to pBR322 sequence numbering. The f1 origin orientation supports single-stranded DNA production for sequencing with the T7 terminator primer. The bottom section shows the cloning/expression region.

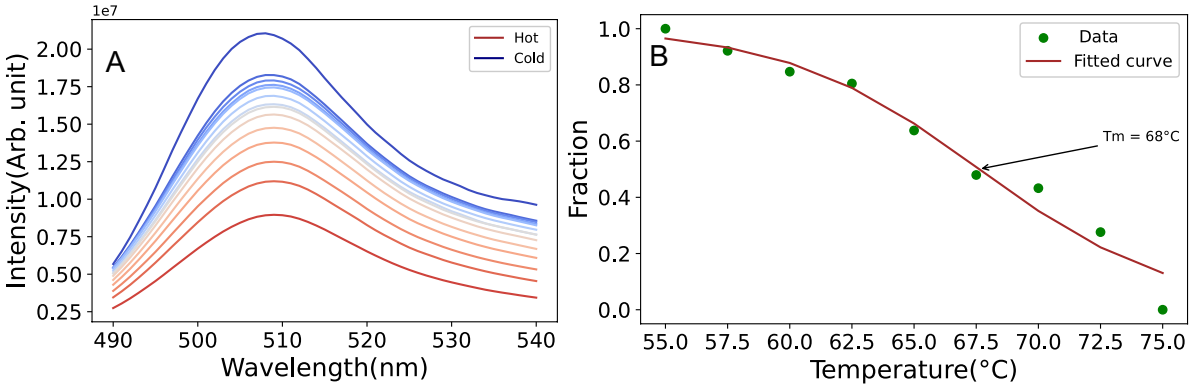

Fig. S2: Determination of the Melting Temperature (Tm) of GFP for System Calibration. A) Fluorescence spectra of GFP demonstrating a progressive quenching of fluorescence intensity as the temperature increases. B) Fitted melting curve for GFP, from area under curve yielding a Tm of 68°C, in close agreement with previously reported experimental values, validating system calibration.

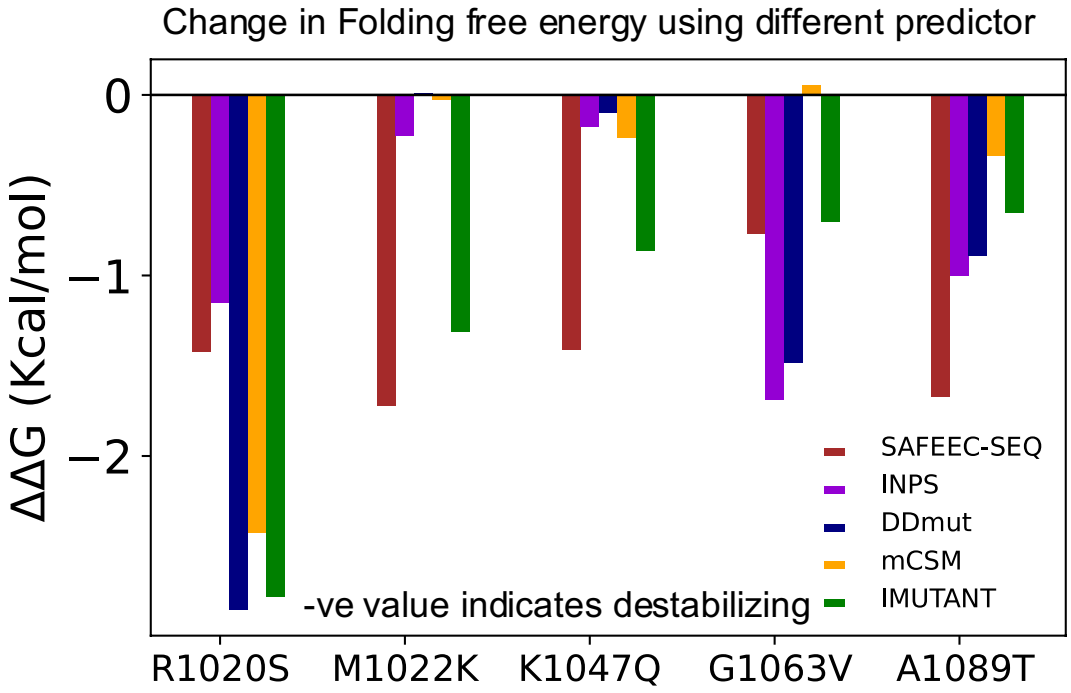

Fig. S3Stability changes in ARID1A predicted by both sequence-based and structure-based models with respect to the selected mutations.: The position and amino acid mutations are indicated by x axis along with the different web server tools as shown by the legend. Most of the mutations of interest show -ve value which means they are destabilizing.

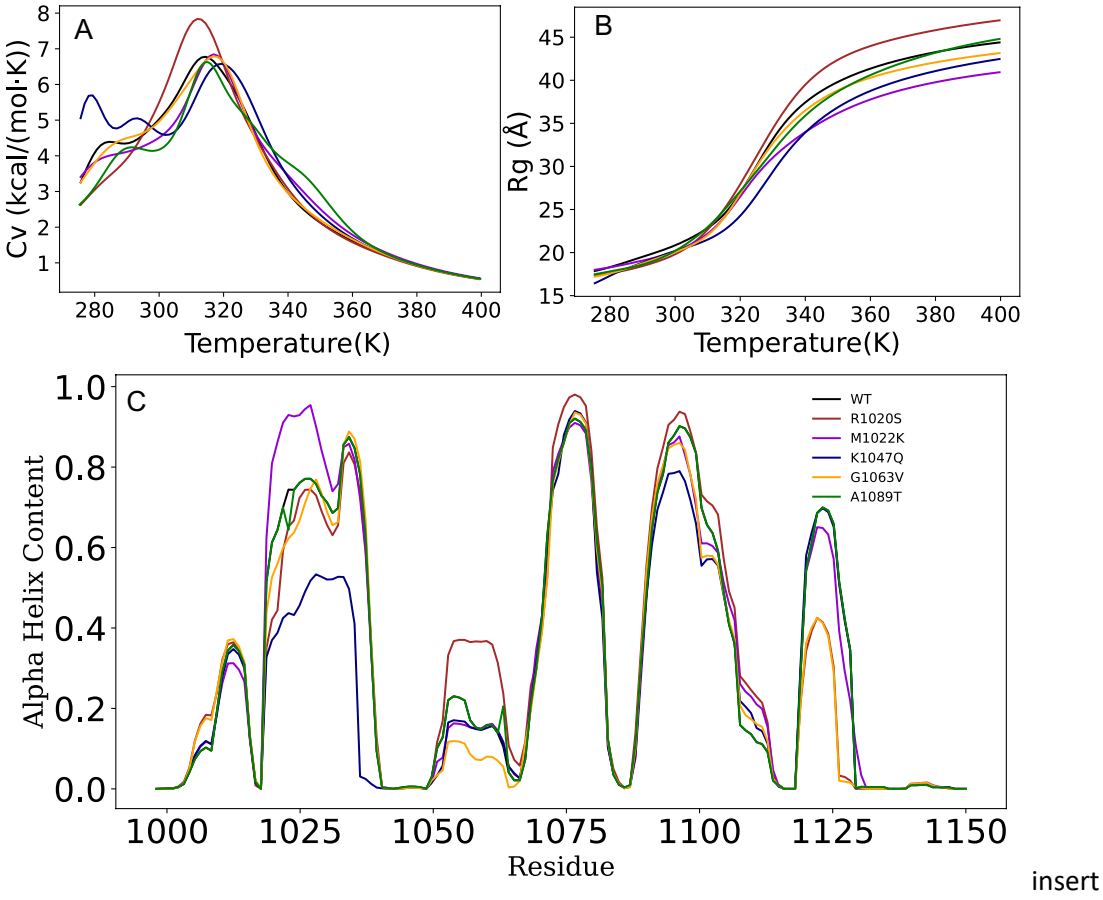

Fig: **S4**. **rexDMD showed variation on the molecular parameters.** A) The Cv as a function of temperature which peak value shows the Tm for the mutants, B) The average radius of gyrations (Rg) obtained for all the mutants as a function of temperature, C) The alpha-helix content for the mutants which shows M1022K has highest content compared to other mutants.

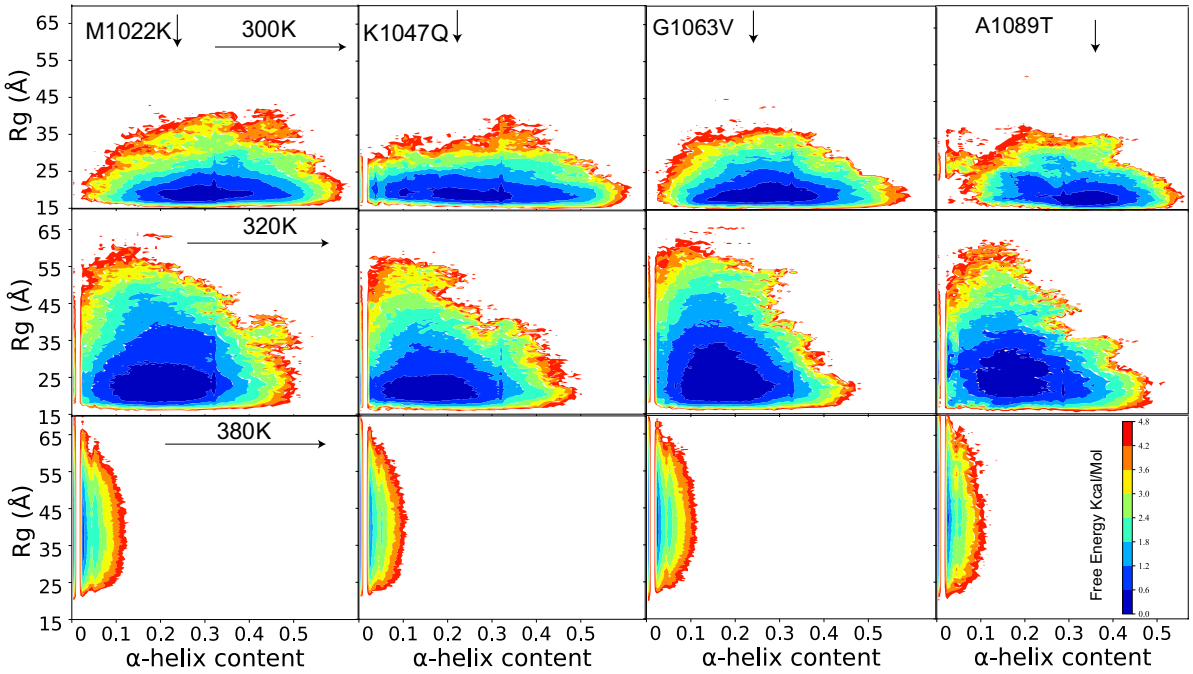

Fig. S5. **2D PMF plot from rexDMD.** The 2D PMF plot of the mutants at different temperatures. Top row at 300K, middle row at 320K and bottom row at 380K. The first column M1022K, second column for K1047Q, third column for G1063V followed by A1089T. Within the same temperature there are different dynamics of structure for different mutants as shown by the alpha-helices contents. As the temperatures increases there is reduction in alpha-helices content and the increment of R_g for all samples.

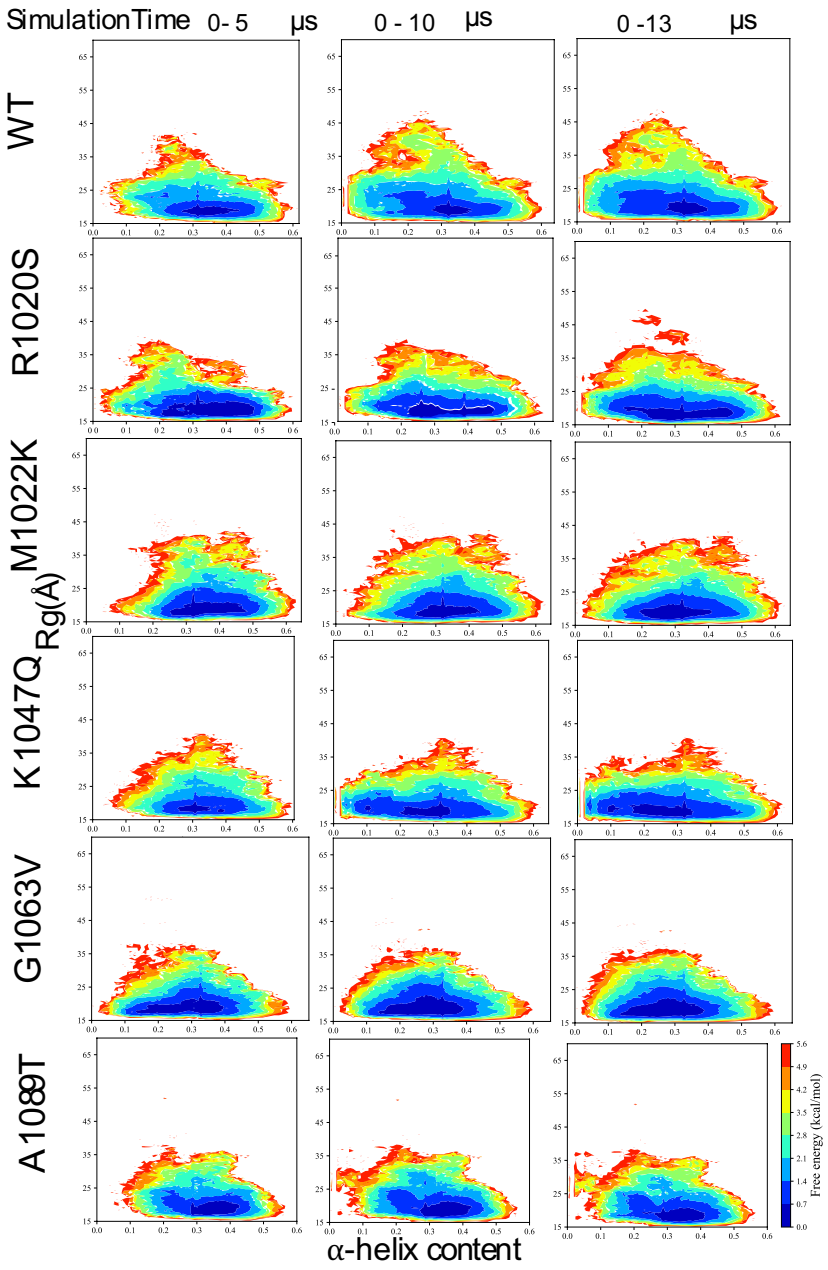

Figure S6. 2D PMF plots from rexDMD simulations show the evolution of unfolded conformations for all mutants. The radius of gyration (Rg) and α-helix content were used to assess unfolding. The higher Rg at 10 μs and 13 μs compared to 5 μs indicates protein unfolding and approaching equilibrium. The PMF at 0.65 μs reveals energy minima with low α-helix content.

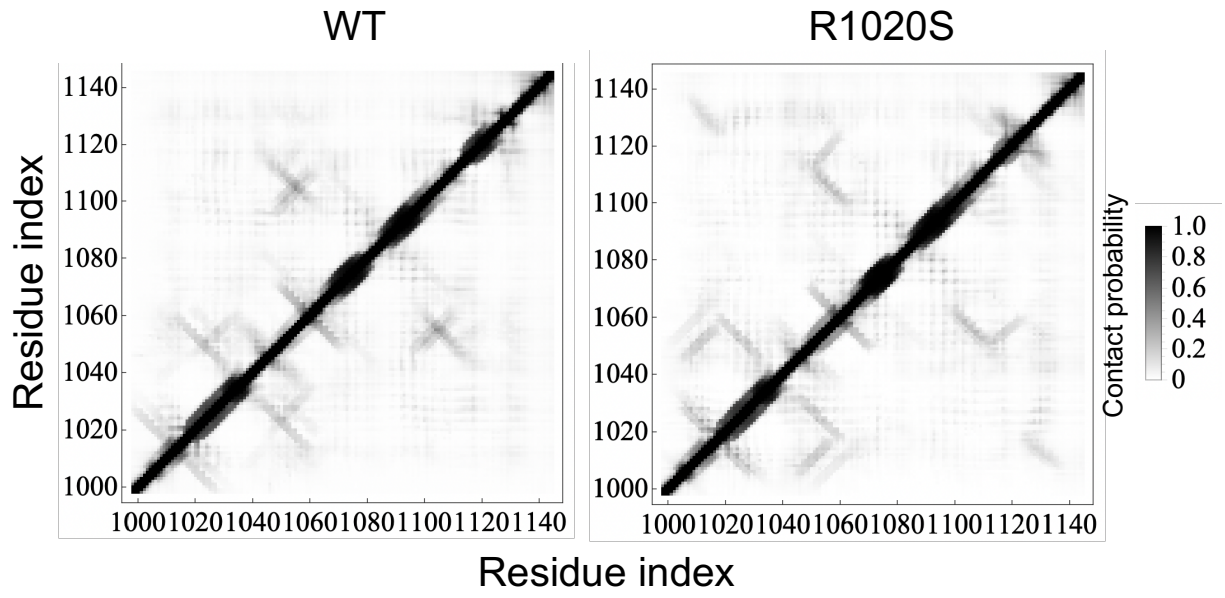

Figure S7).Calculated intra-chain contact maps of the WT and R1020S Arid1A.

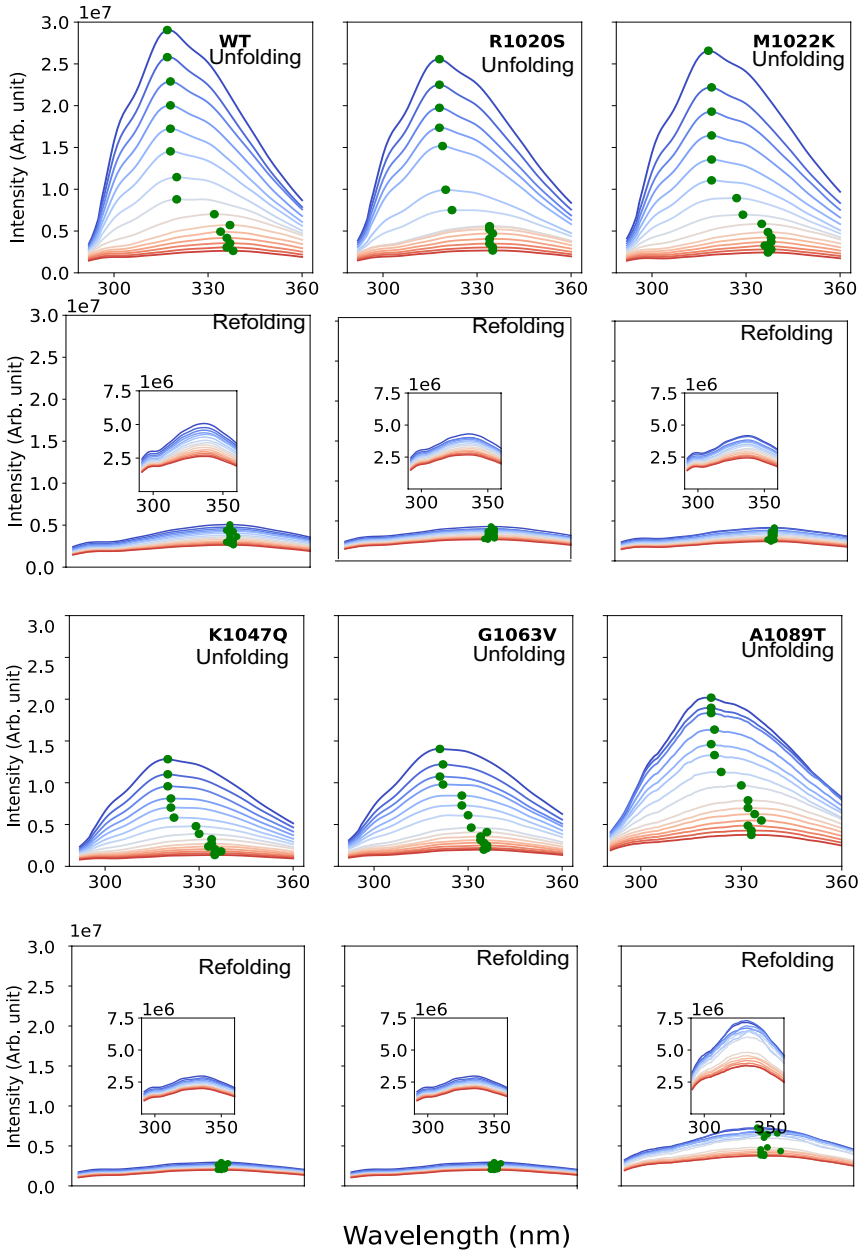

Fig. S8. **Thermal denaturation spectra for all the samples**. The first and third rows represent the unfolding whereas the second and fourth rows represent the refolding for the mutants as shown by legend in the unfolding curve. The inset in refolding curve is the zoom in view for clarity with different scale. This spectrum shows the wavelength shift at, maximum intensity (green dots) shift towards right as well as the fluorescence intensity quenched for unfolding on refolding the reverse is not true only partial recovery of fluorescence but not blue shifting of the green dots i.e. maximum wavelength for confirming irreversibility of the process.

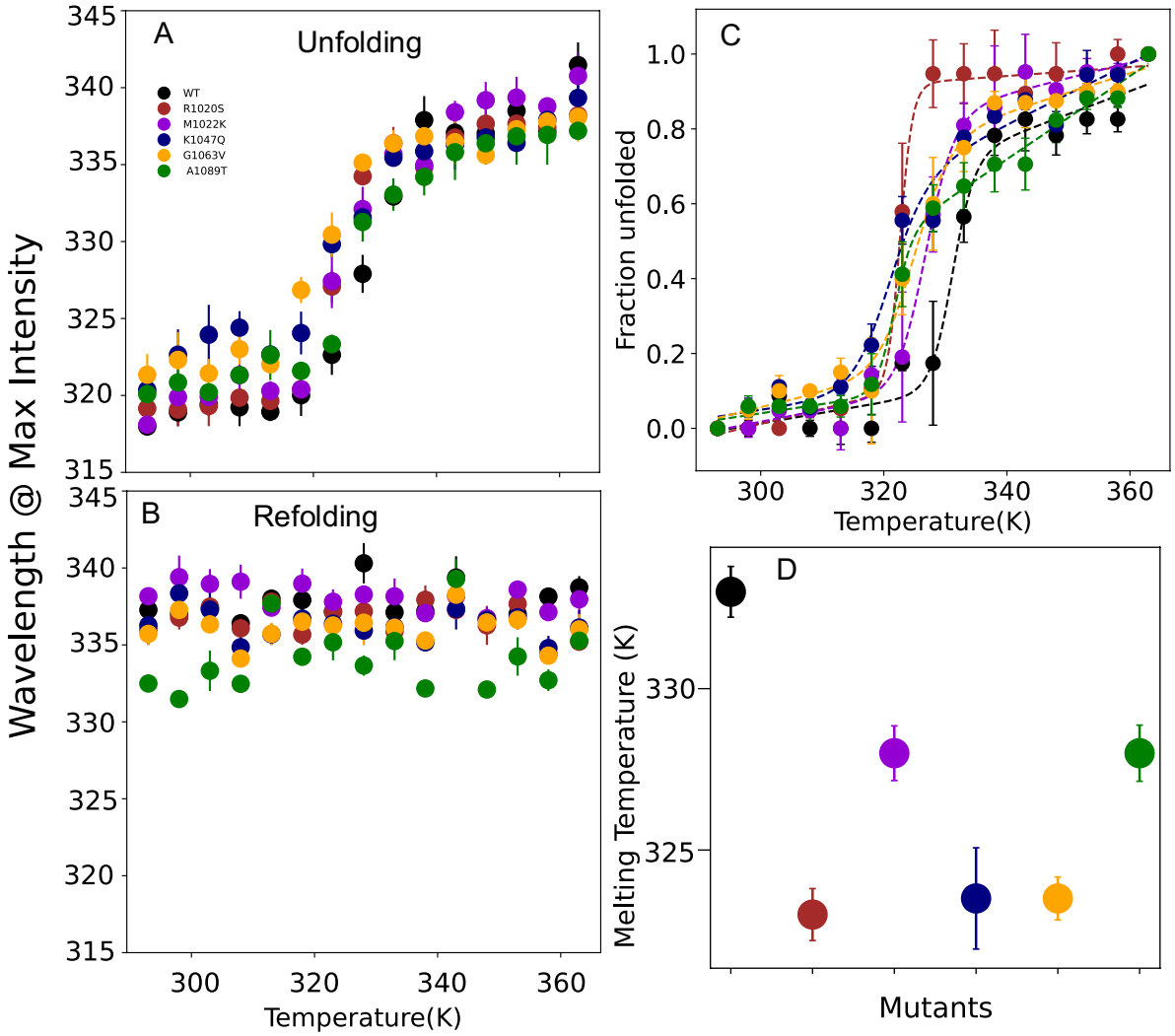

Fig. S9. **Thermal denaturation parameters for all the mutants**. A) Wavelength shift at maximum intensity for unfolding as a function of temperature, B) Wavelength shift at maximum intensity for refolding as a function of temperature, C) Normalized unfolded fraction with fit function for unfolding. D) The melting temperature for the mutants obtained from fit function at which unfolded fraction is 50%.

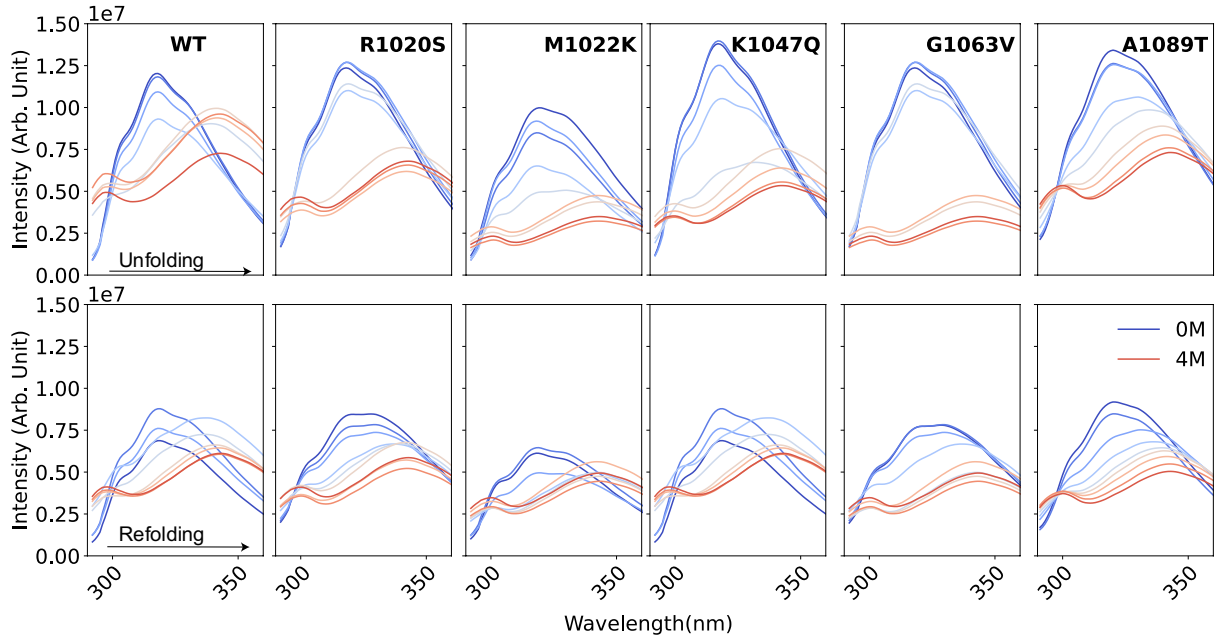

Fig. S10. **Chemical denaturation spectra for all the samples**. The top row represents the unfolding whereas the bottom row represents refolding. This clearly shows wavelength shift towards red as well as fluorescence quenching on unfolding. Similarly for the refolding process blue shifting and regain in the fluorescence confirms the reversibility of the process.

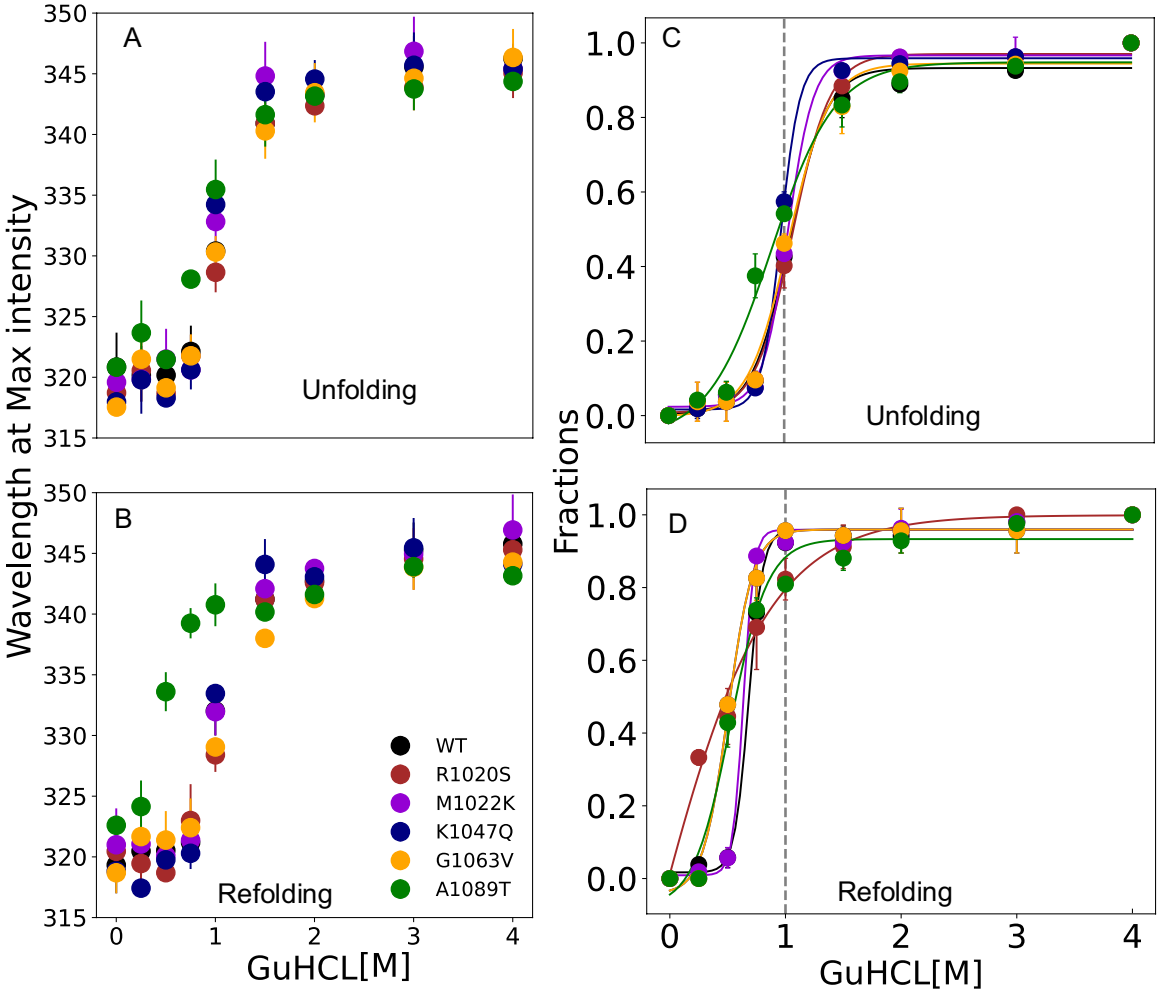

Fig. S11. **Chemical denaturation parameters for all samples**. A: Wavelength shift at maximum intensity unfolding, B: Wavelength shift at maximum intensity refolding, C: normalized unfolded fraction for unfolding and D: normalized refolded fraction for refolding.

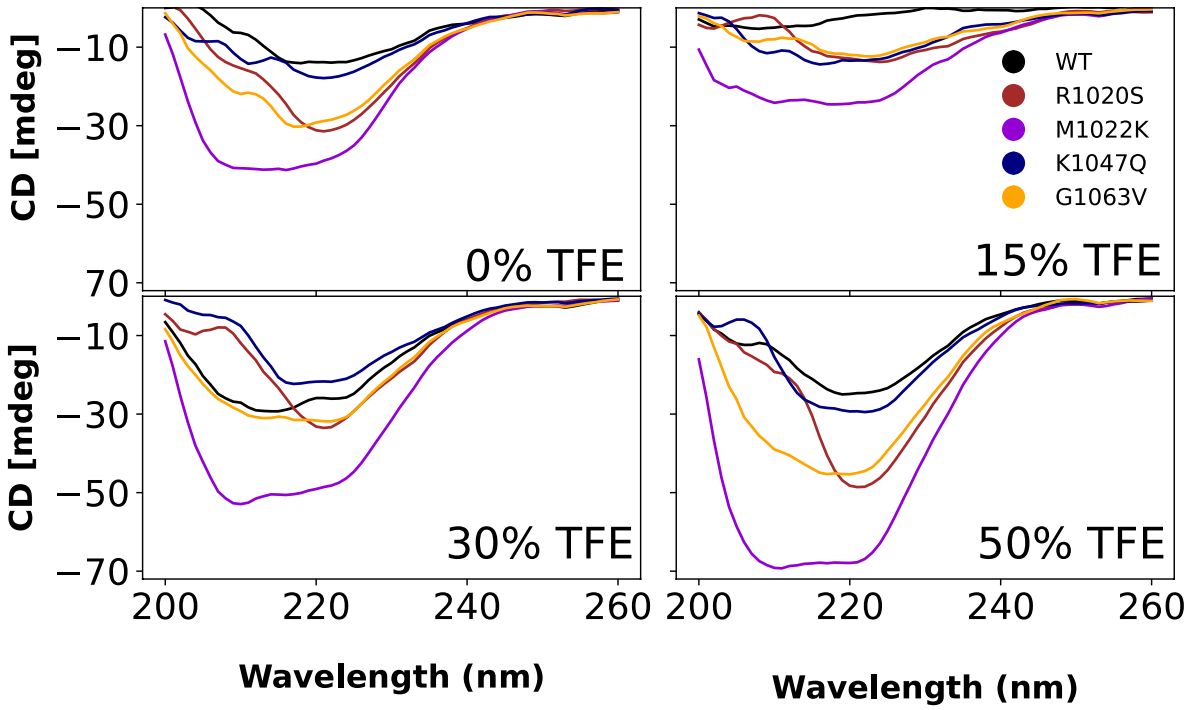

Fig. S12. **CD Spectra at different %TFE (trifluoroethanol) showed M1022K most stabilizing.** As the percentage of TFE (trifluoroethanol) increased, the stability of the helical content improved, with this effect being particularly pronounced for the M1022K mutation, where the depth of the signal increased with higher TFE concentrations. The A1089T mutation is absent from this figure because the sample was not purified during the measurements.

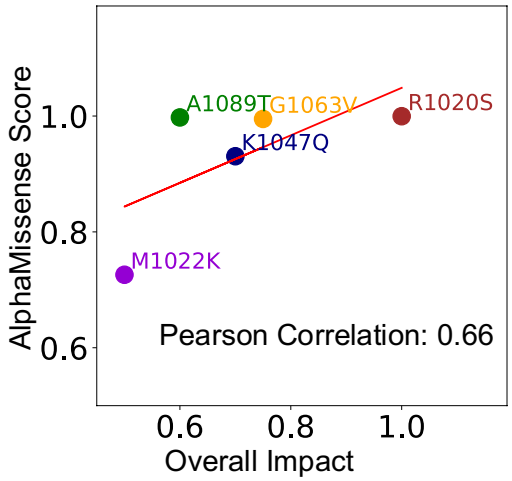

Fig. S13: The AlphaMissense Score showed all mutants are pathogenic correlate with overall impact. The Pearson correlation score is 0.66 agreements with our overall impacts for the most impacted mutants R1020S and least impacted mutants M1022K.

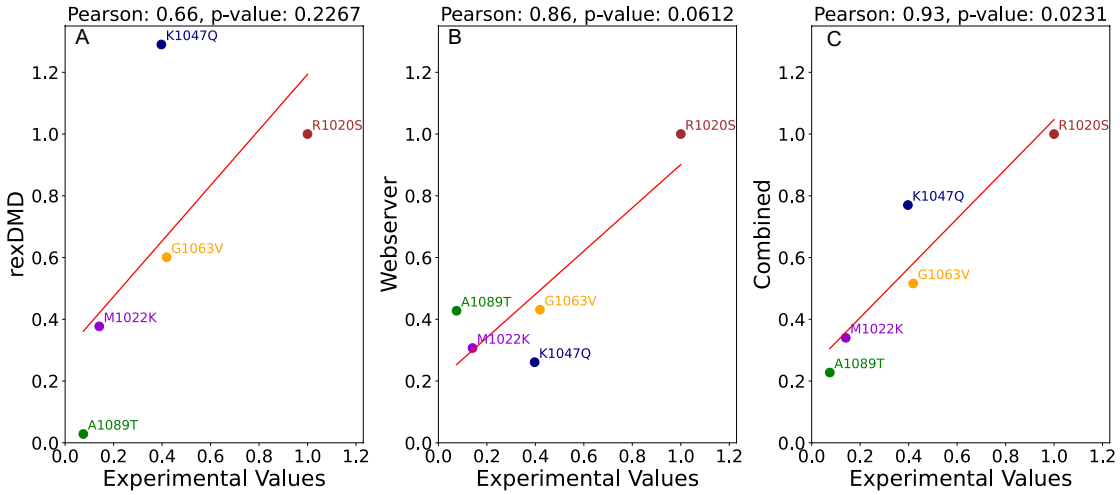

Figure S14: Improvement in Pearson correlation coefficient and validation of results with a p-value of 0.0231 using combined approaches. A) Correlation between experimental impact (averaged from thermal and chemical data as described in *Overall Impact*) and rexDMD impact, showing a Pearson coefficient of 0.66 and a p-value of 0.226. B) Correlation between experimental impact and Web server impact, with a Pearson coefficient of 0.86 and a p-value of 0.0612.C) Correlation between experimental impact and combined impact (average of rexDMD and Web server impacts), demonstrating a Pearson coefficient of 0.93 and a significant p-value of 0.0231.

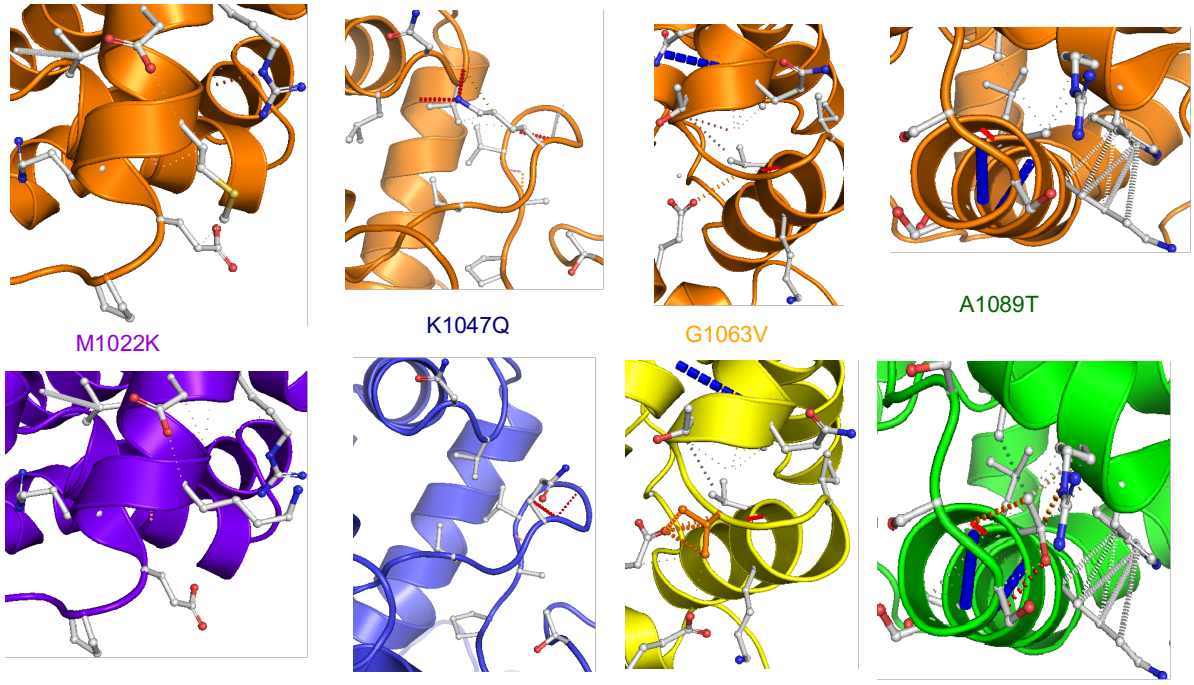

Fig. S15: Interacting sites for WT and mutants in specific position (top panel) represents interacting sites for WT (unmutated) for M1022K, K1047Q, G1063V and A1089T. (bottom panel) represents the interacting for the mutant accordingly.

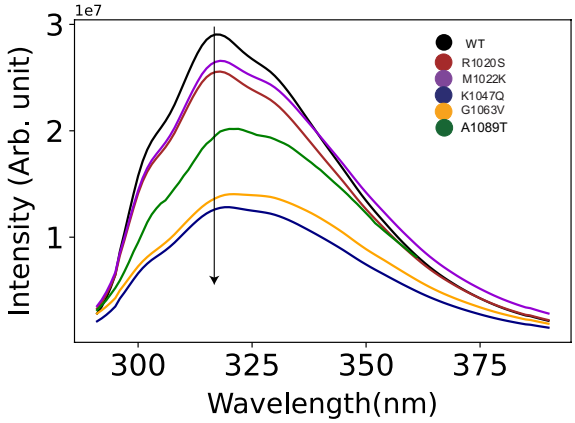

Fig. S16: The native state spectra for the mutants showed variation. For the same concentration and under the same condition the native states vary both in intensity and maximum wavelength.

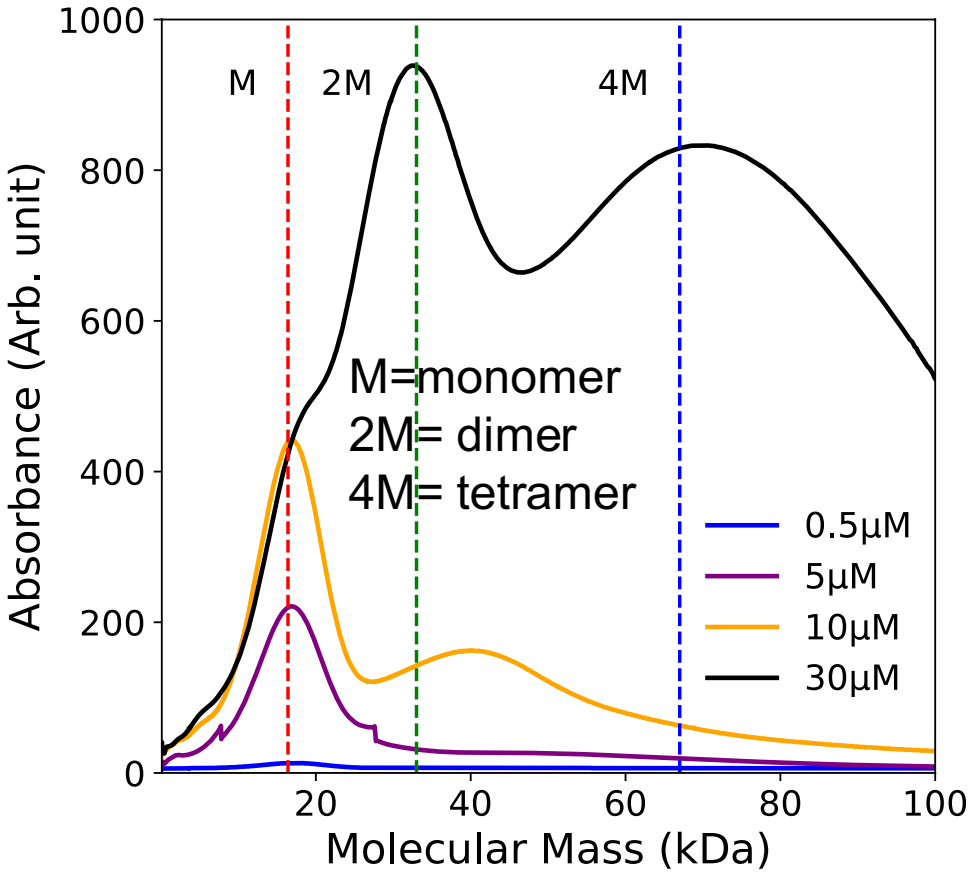

Figure S17: Size Exclusion Chromatography (SEC) for WT at varying concentrations. The SEC profiles, as indicated by the legend, demonstrate the separation of protein at different concentrations. At concentrations below 5µM, no dimerization is observed. However, as the concentration increases to 10µM, the peaks shift to the right, indicating larger molecular mass and the onset of dimerization. At 30µM, monomers are no longer detected, and significant amounts of tetramers and dimers are evident in the profile.

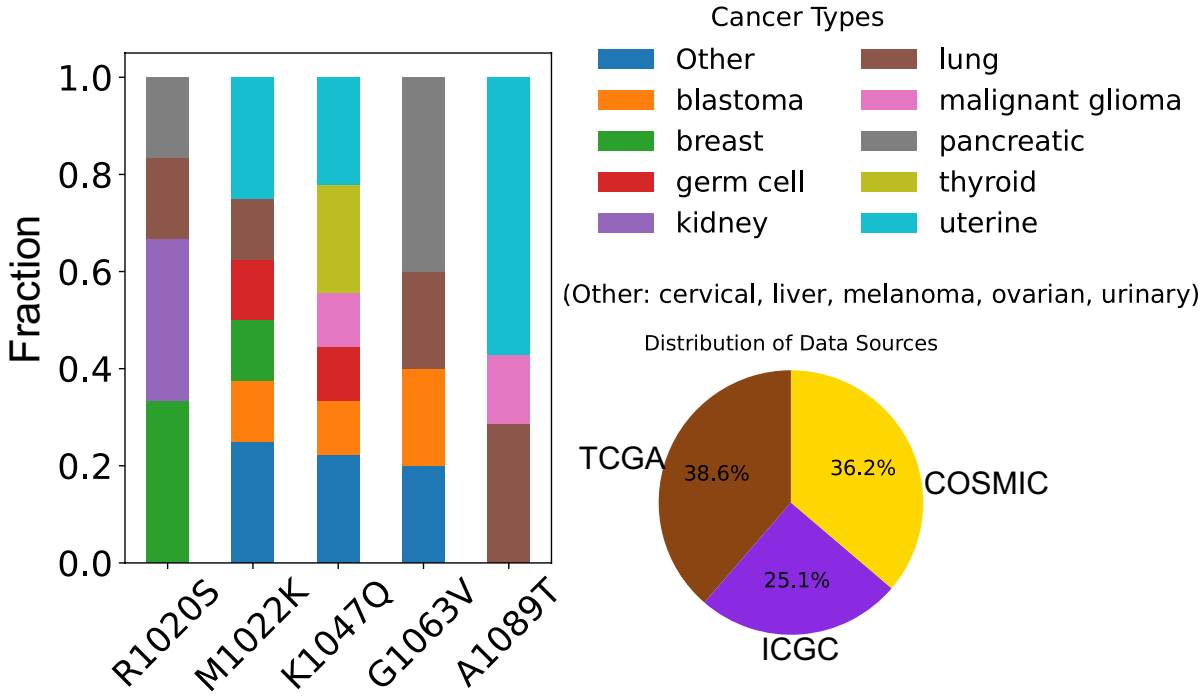

Figure S18: Distributions of types of cancers for the selected mutants and their Data sources. The highest instable has less cancer types than the least impacted mutants.