Supporting information

Prediction of new stabilizing mutations based on mechanistic insights from Markov state models

Maxwell I. Zimmerman[*,1], Kathryn M. Hart[*,1,2], Carrie A. Sibbald[1], Thomas E. Frederick[1], John R. Jimah[3], Catherine R. Knoverek[1], Niraj H. Tolia[1,3], Gregory R. Bowman[1,4]

[1]*Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110*

[2]*Current address: Department of Chemistry, Williams College, 880 Main Street, Williamstown, MA 01267*

[3]*Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110*

[4]*Department of Biomedical Engineering, and Center for Biological Systems Engineering, Washington University in St. Louis, One Brookings Drive, St. Louis, MO, 63130*

*These authors contributed equally to this work.
Correspondence: G.R.B. (email: g.bowman@wustl.edu)

Contents

Figure S1. Analysis of N-terminal capping probabilities for each TEM variant.

Figure S2. Effect of M182T on the stability of helix 9, as judged by the distributions of distances between hydrogen-bonding partners.

Figure S3. Chemical melts of TEM variants.

Figure S4. Crystallographic model of TEM M182N.

Figure S5. Investigation of helix 9 stability in isolation between TEM variants.

Figure S6. Residue 182 χ_1 probabilities.

Figure S7. M182N distance distributions, conditional on Asn182's rotamer conformation, for three helix 9 backbone hydrogen bonding partners.

Figure S8. Solvent accessibility distributions at the domain interface, conditional on Asn182's rotamer conformation.

Figure S9. Chemical shift perturbations for each TEM variant.

Table S1. Crystallographic information for TEM M182N.

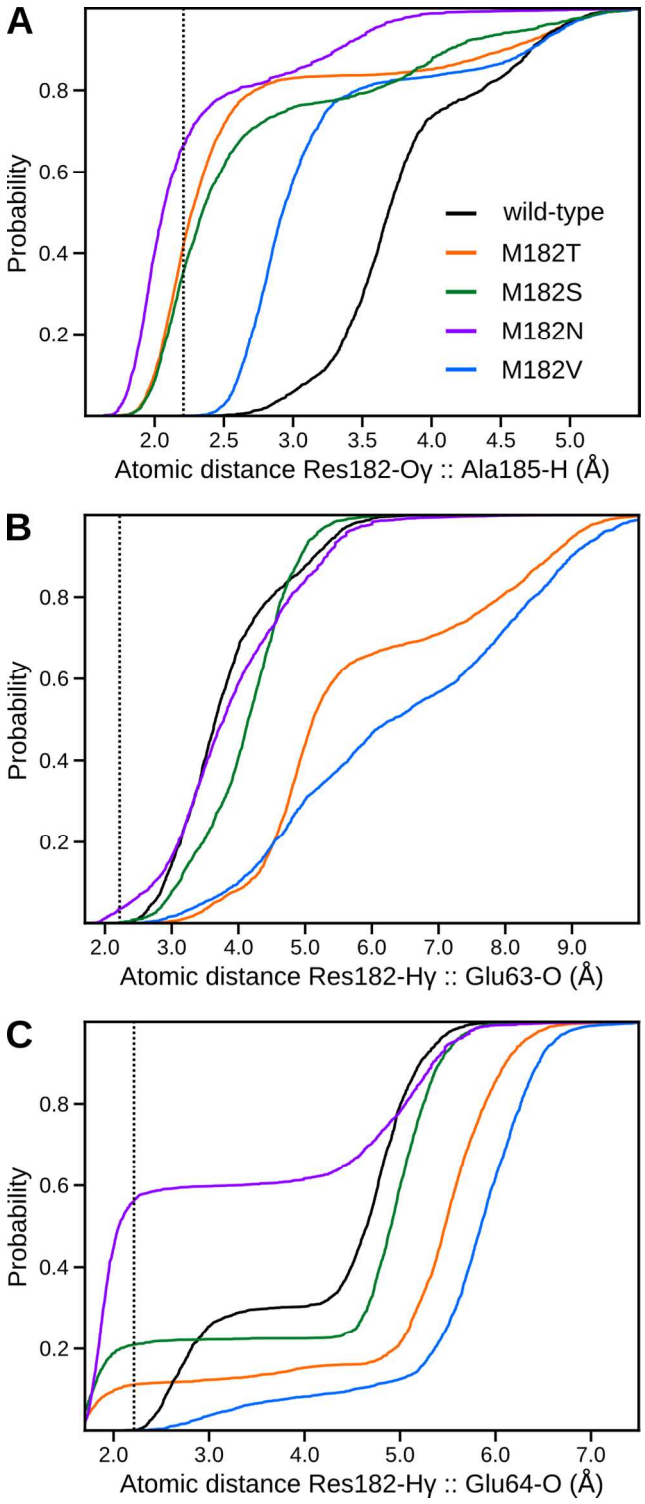

Figure S1. Analysis of N-terminal capping probabilities for each TEM variant. (A-C) Cumulative distributions functions, generated using population-weights from MSMs of our FAST simulations, of three distances: Res182-Oγ (or equivalent) to Ala185-H, Res182-Hγ (or equivalent) to Glu63-O, and Res182-Hγ (or equivalent) to Glu64-O, for five TEM variants: wild-type (black), M182T (red), M182S (green), M182N (purple), and M182V (blue). This indicates the probability of observing an atomic distance less than the specified value. The dotted line indicates the distance of transition from moderate to weak hydrogen bond strength (2.2 Å).

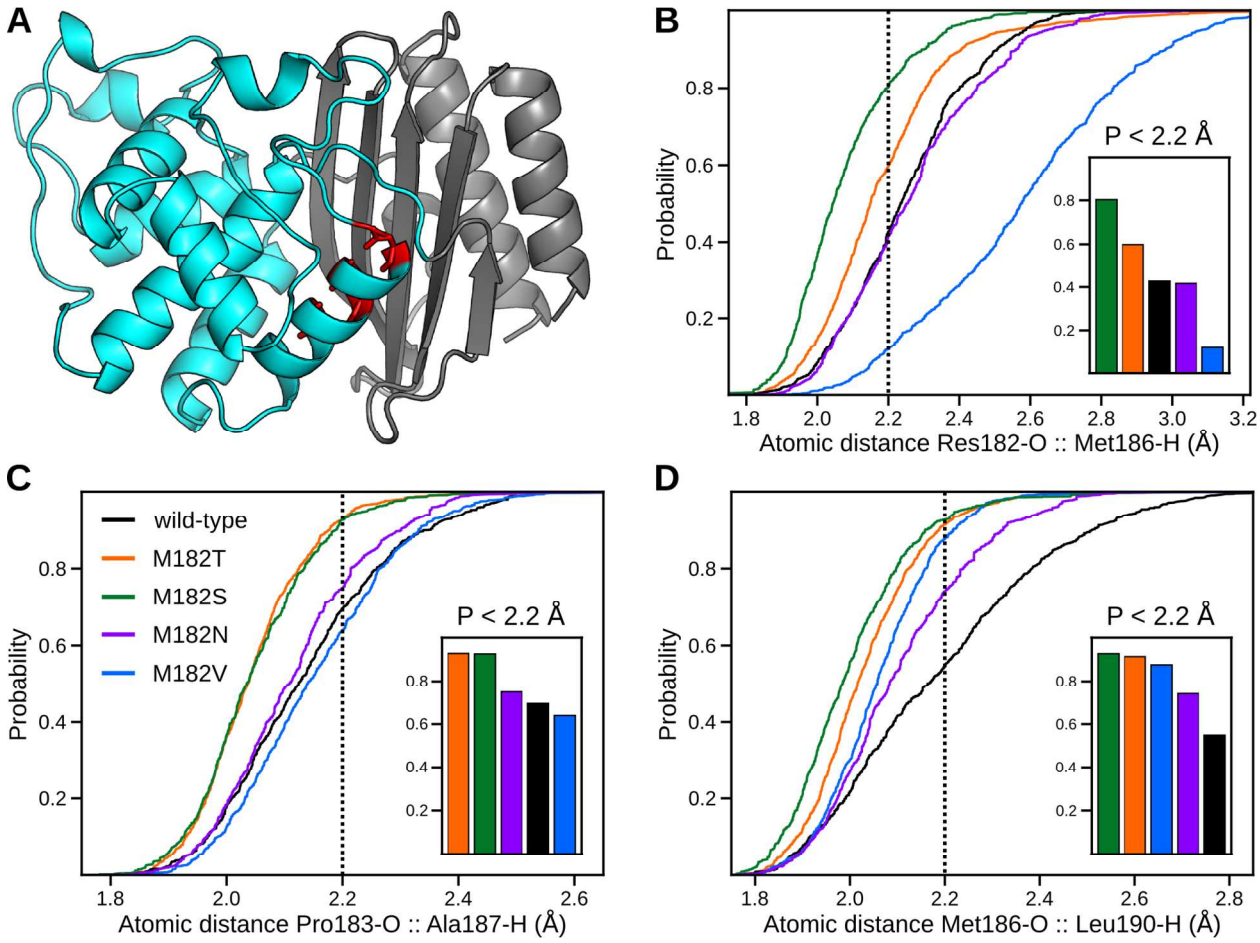

Figure S2. Effect of M182T on the stability of helix 9, as judged by the distributions of distances between hydrogen-bonding partners. (A) Structure highlighting hydrogen-bonding partners Residue 182 and Met186, Pro183 and Ala187, and Met186 and Leu190, which are colored red. (B-D) Cumulative distribution functions, generated using population-weights from MSMs of our FAST simulations, of the hydrogen bonding partners listed above for wild-type (black) and M182T (orange), M182S (green), M182N (purple), and M182V (blue). These plots indicate the probability of observing an atomic distance less than the specified value. Our cutoff distance for moderate hydrogen bonds, 2.2 Å, is shown as a dotted line. Probabilities of moderate hydrogen bonds for each pair are shown in the inset.

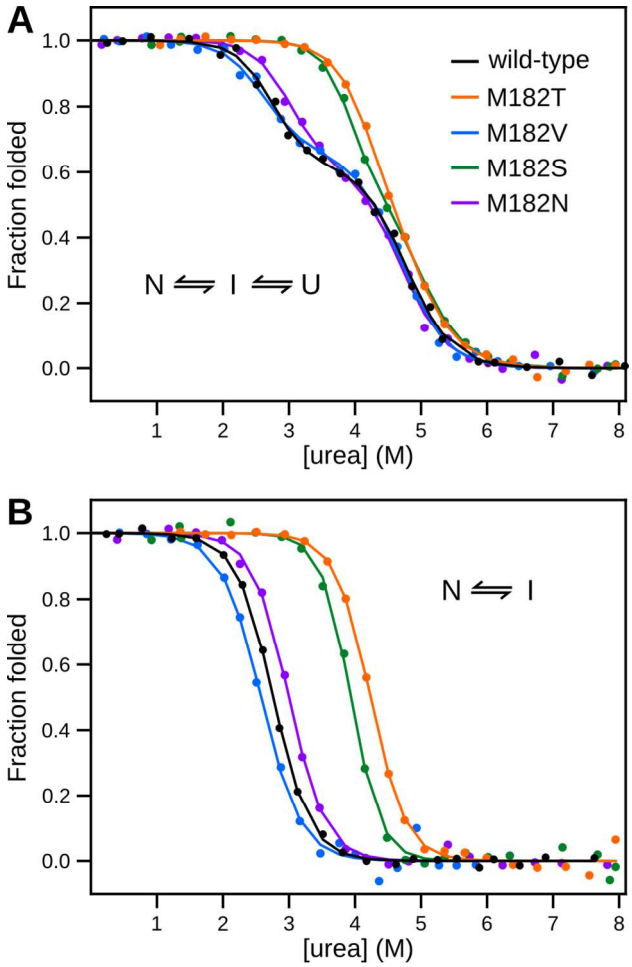

Figure S3. Chemical melts of TEM variants. Shown are the fractions of folded protein for wild-type TEM (black) and TEM M182T (red), M182V (blue), M182S (green), and M182N (purple) as a function of urea. (A) Monitoring signal from circular dichroism. (B) Monitoring signal from fluorescence.

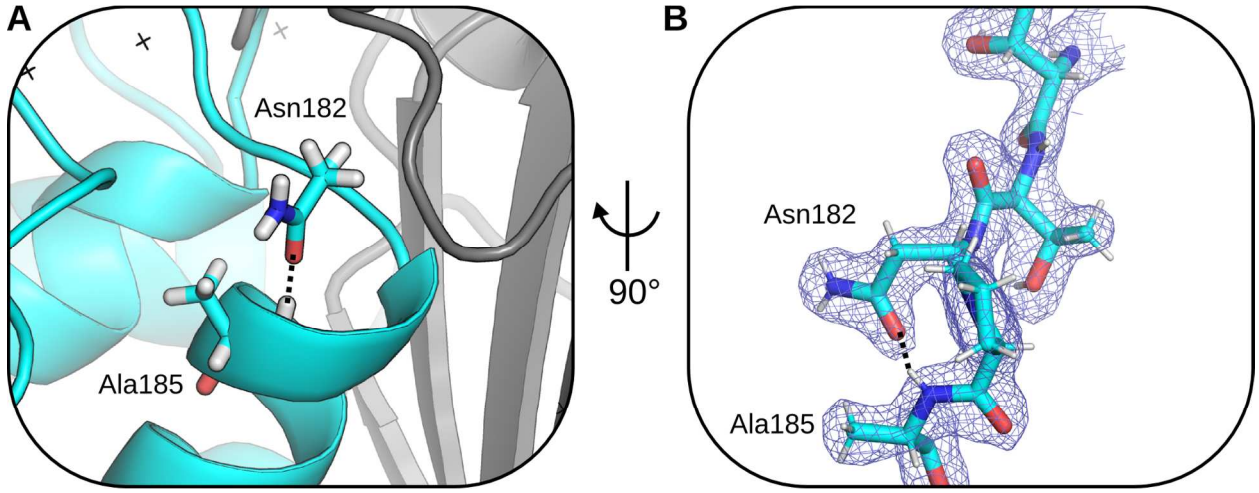

Figure S4. The best fit rotamer of Asn182 from the crystal structure of M182N. Shown is a representative TEM structure from the crystal lattice, solved to 2.0 Å. Asn182 is observed to form a hydrogen bond with Ala185. Additionally, the sidechain amine has no hydrogen bonding partner and points outward to solvent. (A) Asn182 and Ala185 are represented as sticks, with the backbone of the α-helix domain (cyan) and β-sheet domain (gray) represented as a cartoon. (B) Electron density around Asn182.

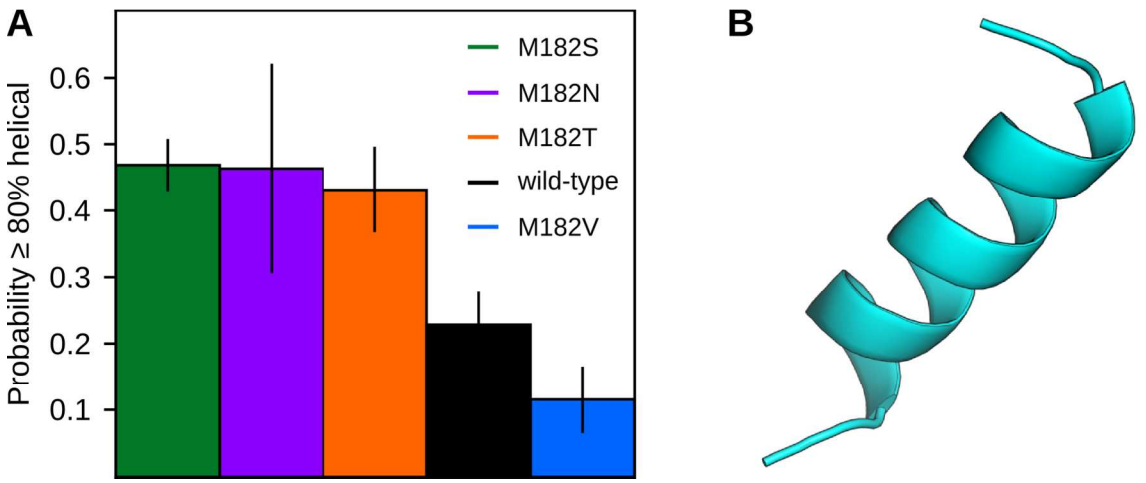

Figure S5. Investigation of helix 9 stability in isolation between TEM variants. (A) Probability of each variant's helix 9 having greater than or equal to 80% of its native helicity. Probabilities come from the MSMs of the isolated helix 9 for wild-type (black), M182T (orange), M182S (green), M182V (blue), and M182N (purple). (B) Helix 9 in isolation (residues 181-197), and the starting structure for simulations.

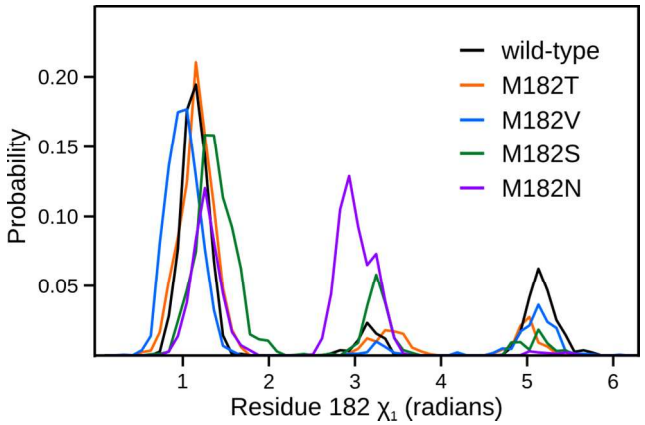

Figure S6. Residue 182 χ_1 probabilities. Shown are the χ_1 probabilities of each TEM sequence: wild-type (black), M182T (red), M182V (blue), M182S (green), and M182N (purple). These probabilities come from MSMs of the full protein.

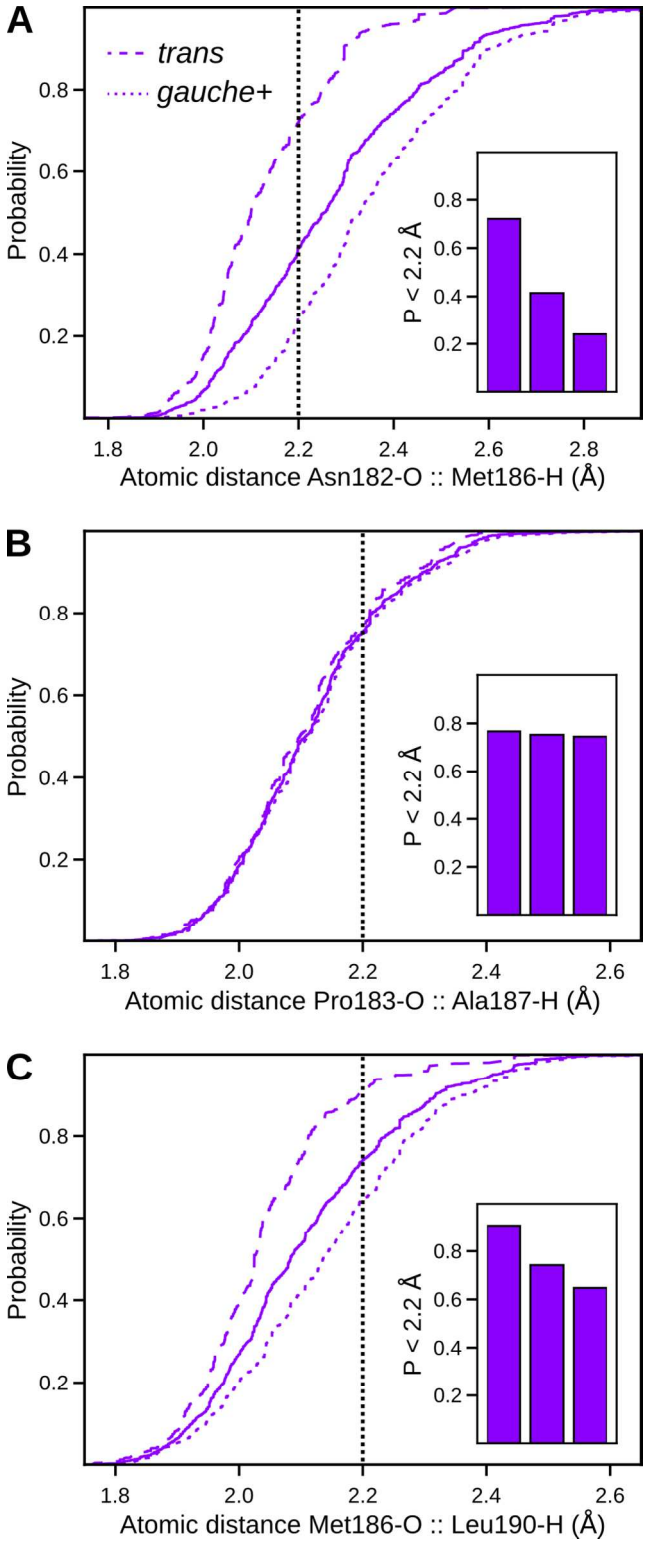

Figure S7. M182N distance distributions, conditional on Asn182's rotamer conformation, for three helix 9 backbone hydrogen bonding partners. (A-C) Cumulative distribution plots, conditional to M182N's rotamer, of the three distances represented in Fig S2. These distributions are generated using population-weights from MSMs of our FAST simulations. Shown are the distributions for Asn182 adopting the *trans* rotamer (dashed lines), the

gauche+ rotamer (dotted lines), and all rotamers (solid lines). These plots indicate the probability of observing an atomic distance less than the specified value. Our cutoff distance for moderate hydrogen bonds, 2.2 Å, is shown as a dotted line. Probabilities of moderate hydrogen bonds for each pair are shown in the inset, which show a significant difference between the *trans* and *gauche+* rotamer for two out of three distances.

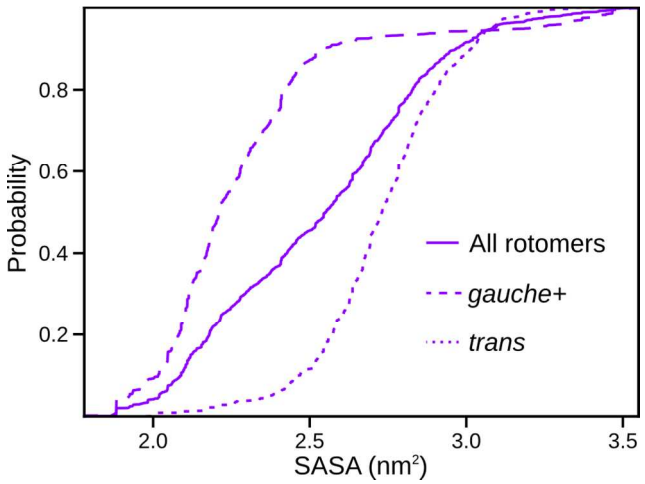

Figure S8. Solvent accessibility distributions at the domain interface, conditional on Asn182's rotamer conformation. Shown are the cumulative distribution functions, generated using population-weights from MSMs of our FAST simulations, for the solvent accessible surface area of six residues: Tyr46, Ile47, Pro62, Glu63, Pro183, and Ala184, illustrated in Fig. 6. These residues are located at the interface of the s2h2 loop, helix 9, and the β-sheet domain. Shown are the distributions for Asn182 adopting the *trans* rotamer (dashed line), the *gauche+* rotomer (dotted line), and all rotamers (solid line). These plots indicate the probability of observing solvent-accessible surface area less than the specified value.

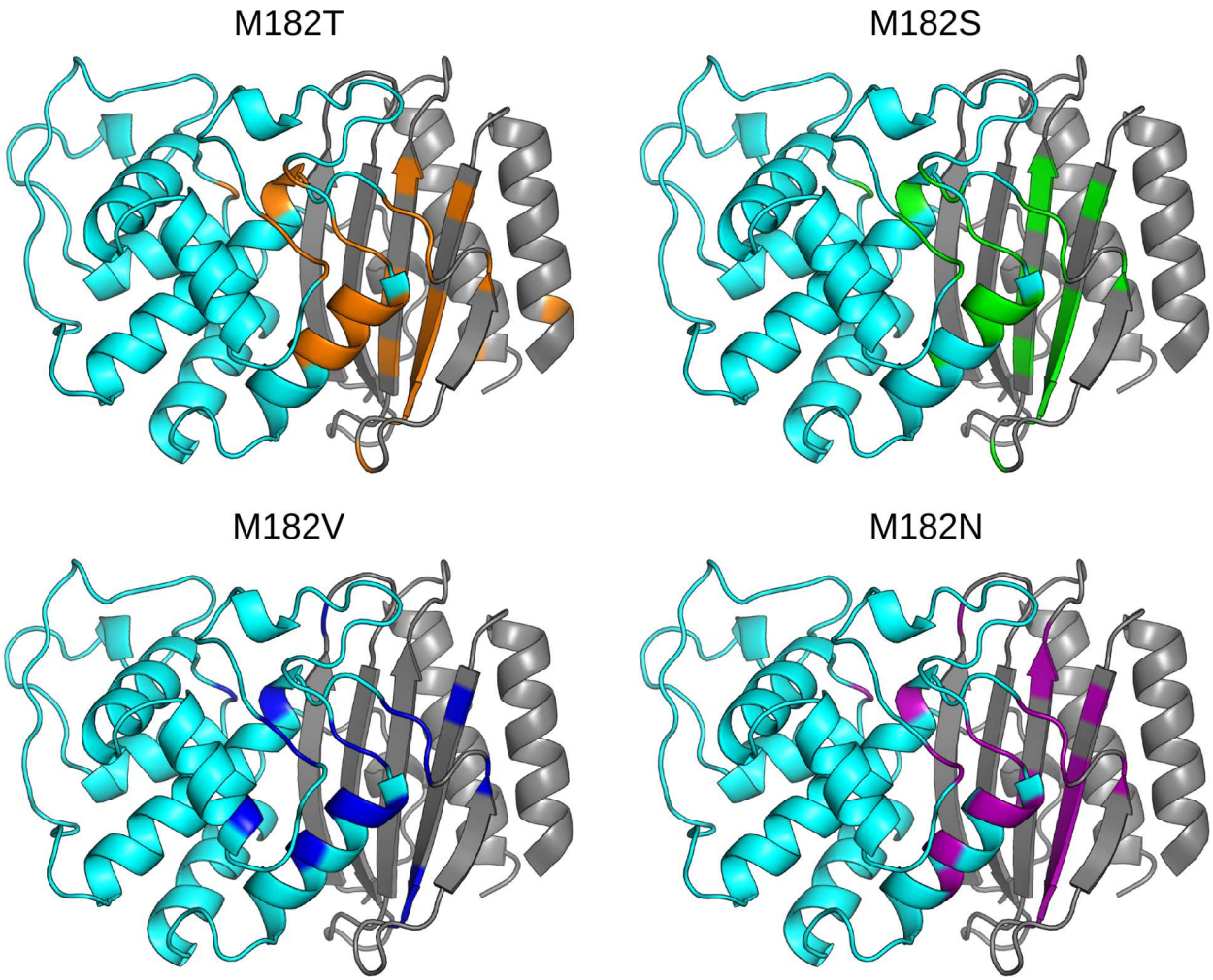

Figure S9. Chemical shift perturbations for each TEM variant. The backbone of the α-helix domain (cyan) and β-sheet domain (gray) are represented as a cartoon. Highlighted residues indicate the locations of backbone amide chemical shifts that are perturbed significantly relative to wild-type. Chemical shift perturbations are shown for each TEM variant: M182T (orange), M182S (green), M182V (blue) and M182N (purple).

Table S1. Crystallographic information for TEM M182N.

	M182N
Data collection	
Space group	P 2_1
Cell dimensions	
a, *b*, *c* (Å)	81.91, 49.66, 122.16
α, β, γ (°)	90, 90.102, 90
Resolution (Å)	20 – 2 (2.1 – 2.0)
R_{sym} (%)	9.9 (50.6)
$I / \sigma I$	8.95 (2.34)
Completeness (%)	98.6 (98.1)
Redundancy	3.53 (3.47)

Refinement	
Resolution (Å)	20 – 2
No. reflections	66,017
R_{work} / R_{free} (%)	22.14/28.08
No. atoms	
Protein	8,094
Ligands	20
Solvent	212
Average B-factors	
Protein	39.28
Ligands	72.49
Solvent	21.01
Number of TLS groups	24
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.61
Ramachandran favored (%)	96.44
Ramachandran allowed (%)	3.56
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	4.32

Each data set was collected from a single crystal
Highest resolution shell is shown in parenthesis