

Supplementary Materials:

A contraction index for protein stability predicts disease onset in stability-driven amyloidoses and identifies its own boundary of validity

Matthias Wurm

Contents

1	S1: Thermodynamic $\sigma(T)$ for four proteins	3
2	S2: Go model specification	3
2.1	Native contacts (helical cage)	4
2.2	Amyloid contacts (β -hairpin)	4
2.3	Native and amyloid reference states	4
2.4	Energy function	5
3	S3: Monte Carlo simulation parameters	5
4	S4: Temperature scan results	6
4.1	Native basin ($\alpha = 0$)	6
4.2	Amyloid basin ($\alpha = 1$)	6
5	S5: Full α scan data	6
6	S6: Alzheimer mutation data	7
7	S7: Intervention simulation details	8
7.1	Chaperone model	8
7.2	Stabilizer model	8
7.3	Complete intervention results at $\alpha = 0.45$	8

8 S8: Dose–response across disease stages	8
9 S9: Experimental validation data	9
9.1 S9.1: Experimental $\Delta\Delta G \rightarrow \sigma$ for 37 mutations	9
9.2 S9.2: Go model vs. experimental σ cross-validation	11
9.3 S9.3: σ -drift model parameters	11
10 S10: Large-scale validation across 20 proteins	12
11 S11: ClinVar VUS predictions for APP	12
11.1 S11.1: Calibration on known variants	12
11.2 S11.2: VUS predictions	12
12 S12: Code availability	14
13 S13: Transthyretin validation data	14
13.1 S13.2: Leave-one-out analysis	15
13.2 S13.3: Within-phenotype correlations	16
13.3 S13.4: Onset data caveats	16
14 S14: Benchmark against alternative early-warning indicators	16
15 S15: Hereditary lysozyme amyloidosis validation data	17
16 S16: Gelsolin amyloidosis validation data	17
17 S17: Negative control data	18
17.1 S17.1: SOD1/ALS	18
17.2 S17.2: Prion disease (PRNP)	18
17.3 S17.3: A β 42 cross-method disagreement	18

1 S1: Thermodynamic $\sigma(T)$ for four proteins

The stability index $\sigma(T) = \exp(-\Delta G(T)/NRT)$ was computed from published thermodynamic data using the Gibbs–Helmholtz equation. Table 1 shows σ at selected temperatures for each protein.

Table 1: $\sigma(T)$ for four proteins at selected temperatures.

T (K)	Trp-cage	Villin HP35	CI2	ACBP
280	0.634	0.890	0.866	0.857
290	0.708	0.882	0.870	0.868
300	0.799	0.885	0.882	0.887
310	0.910	0.898	0.904	0.915
320	1.042	0.920	0.933	0.949
330	1.202	0.952	0.970	0.991
340	1.393	0.991	1.014	1.040
350	1.621	1.039	1.066	1.097
360	1.892	1.096	1.127	1.161
370	2.216	1.162	1.195	1.234

Data sources:

- Trp-cage (TC5b): $T_m = 317$ K, $\Delta H_m = 230$ kJ/mol, $\Delta C_p = 2.5$ kJ/(mol·K). Neidigh et al. (2002) *Nat. Struct. Biol.* 9, 425–430; Streicher & Makhatadze (2007) *Biochemistry* 46, 2876–2880.
- Villin HP35: $T_m = 342$ K, $\Delta H_m = 155$ kJ/mol, $\Delta C_p = 3.1$ kJ/(mol·K). Kubelka et al. (2003) *J. Mol. Biol.* 329, 625–630.
- CI2: $T_m = 337$ K, $\Delta H_m = 278$ kJ/mol, $\Delta C_p = 5.0$ kJ/(mol·K). Jackson & Fersht (1991) *Biochemistry* 30, 10428–10435.
- ACBP: $T_m = 332$ K, $\Delta H_m = 370$ kJ/mol, $\Delta C_p = 6.3$ kJ/(mol·K). Kragelund et al. (1995) *J. Mol. Biol.* 256, 187–200.

2 S2: Go model specification

The dual-basin Go model uses $N = 20$ residues, each with $S = 8$ discrete conformational states (labeled 0–7). Two sets of 12 contacts each define the native and amyloid basins.

2.1 Native contacts (helical cage)

i	j	Type
1	5	helix $i, i + 4$
2	6	helix $i, i + 4$
3	7	helix $i, i + 4$
4	8	helix $i, i + 4$
5	11	cage
5	16	cage
5	17	cage
5	18	cage
6	16	packing
6	17	packing
8	15	packing
1	8	packing

2.2 Amyloid contacts (β -hairpin)

i	j	Type
0	19	hairpin
1	18	hairpin
2	17	hairpin
3	16	hairpin
4	15	hairpin
5	14	hairpin
6	13	hairpin
7	12	hairpin
8	11	hairpin
9	10	hairpin (turn)
0	9	turn
1	10	turn

2.3 Native and amyloid reference states

Native: [3, 2, 1, 1, 2, 0, 0, 2, 3, 5, 7, 4, 4, 4, 6, 3, 4, 4, 4, 4]

Amyloid: [5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 2, 2, 2, 2, 2, 2, 2, 2, 2]

2.4 Energy function

The dual-basin energy is:

$$E(\alpha) = (1 - \alpha) \cdot E_{\text{nat}} + \alpha \cdot E_{\text{amy}}$$

where

$$E_{\text{nat}} = -\varepsilon \sum_{(i,j) \in C_{\text{nat}}} \delta(s_i, s_i^*) \delta(s_j, s_j^*) \quad (1)$$

$$E_{\text{amy}} = -\varepsilon \sum_{(i,j) \in C_{\text{amy}}} \delta(s_i, s_i^\dagger) \delta(s_j, s_j^\dagger) \quad (2)$$

with $\varepsilon = 1.0$, s_i^* the native reference state, and $s_i^\dagger$ the amyloid reference state.

3 S3: Monte Carlo simulation parameters

Table 2: Simulation parameters for all experiments.

Parameter	Temperature scan	α scan
MC steps per trial	5000–6000	6000
Number of trials	20–30	30
Measurement interval	10 steps	10 steps
σ window	400–500 steps	500 steps
Transient cutoff	first half	first half
Random seed	42	42
Temperature	0.15–0.75	$0.75 \times \min(T_c^{\text{nat}}, T_c^{\text{amy}})$

The stability index σ is measured during the **folding transient** (first half of each trajectory), not at equilibrium. At equilibrium, $\langle d(t + \Delta)/d(t) \rangle \approx 1$ by stationarity, which obscures the folding signal. The transient measurement captures the net contraction as the protein approaches its preferred basin from a random initial state.

The productive fraction D is computed from accepted MC moves: $D = n_{\text{productive}}/n_{\text{accepted}}$, where a move is productive if it increases Q (reduces distance) to the target basin.

T	Q_{nat}	Q_{amy}	σ_{nat}	σ_{amy}	D_{nat}	γ_{nat}
0.15	0.988	0.000	0.342	1.001	0.006	0.974
0.20	0.971	0.000	0.529	1.001	0.009	0.978
0.25	0.917	0.004	0.761	1.002	0.015	0.991
0.30	0.863	0.000	0.987	1.002	0.020	1.004
0.35	0.783	0.000	0.978	1.001	0.024	1.007
0.40	0.725	0.013	1.067	1.002	0.028	1.009
0.45	0.650	0.013	1.036	1.002	0.031	1.009
0.50	0.512	0.017	1.091	1.002	0.032	1.007

4 S4: Temperature scan results

4.1 Native basin ($\alpha = 0$)

$T_c(Q = 0.5) = 0.508$. σ_{nat} crosses 1.0 near $T = 0.30$, broadly consistent with the folding transition.

4.2 Amyloid basin ($\alpha = 1$)

T	Q_{nat}	Q_{amy}	σ_{nat}	σ_{amy}	D_{amy}	γ_{amy}
0.15	0.000	0.963	1.002	0.609	0.023	0.988
0.20	0.000	0.829	1.002	0.850	0.023	0.995
0.25	0.008	0.675	1.001	0.998	0.021	1.000
0.30	0.008	0.525	1.002	0.958	0.023	1.001
0.35	0.021	0.408	1.002	0.984	0.023	1.001
0.40	0.000	0.308	1.002	1.009	0.024	1.002

$T_c(Q = 0.5) = 0.311$. The amyloid basin has a lower critical temperature than the native basin, consistent with its different contact topology.

5 S5: Full α scan data

Table 3: Complete α scan results at $T = 0.233$.

α	Q_{nat}	Q_{amy}	σ_{nat}	σ_{amy}	D_{nat}	D_{amy}	γ_{nat}	γ_{amy}
0.00	0.892	0.008	0.616	1.002	0.013	0.012	0.982	1.000
0.05	0.925	0.006	0.673	1.002	0.014	0.013	0.985	1.000
0.10	0.919	0.011	0.736	1.003	0.016	0.013	0.990	1.000

α	Q_{nat}	Q_{amy}	σ_{nat}	σ_{amy}	D_{nat}	D_{amy}	γ_{nat}	γ_{amy}
0.15	0.875	0.006	0.844	1.003	0.017	0.013	0.996	1.000
0.20	0.886	0.008	0.940	1.005	0.019	0.014	1.002	1.001
0.25	0.867	0.017	1.012	1.007	0.021	0.014	1.006	1.001
0.30	0.833	0.008	1.042	1.004	0.023	0.014	1.009	1.001
0.35	0.764	0.039	1.029	1.008	0.026	0.015	1.010	1.001
0.40	0.733	0.028	1.037	1.013	0.029	0.016	1.010	1.001
0.45	0.697	0.061	1.063	1.017	0.032	0.017	1.010	1.001
0.50	0.511	0.103	1.127	1.016	0.035	0.020	1.009	1.001
0.55	0.408	0.147	1.088	1.021	0.035	0.023	1.007	1.001
0.60	0.206	0.289	1.113	1.001	0.031	0.025	1.004	1.001
0.65	0.142	0.358	1.026	0.998	0.026	0.024	1.002	1.001
0.70	0.056	0.444	1.018	0.999	0.020	0.024	1.001	1.001
0.75	0.025	0.492	1.013	0.993	0.017	0.023	1.001	1.001
0.80	0.014	0.586	1.007	1.008	0.014	0.022	1.001	1.002
0.85	0.014	0.558	1.004	0.994	0.012	0.022	1.001	1.001
0.90	0.006	0.631	1.003	0.982	0.012	0.021	1.001	1.001
0.95	0.003	0.681	1.002	1.017	0.010	0.022	1.000	1.002
1.00	0.006	0.672	1.002	0.977	0.010	0.021	1.000	1.000

6 S6: Alzheimer mutation data

Table 4: Alzheimer-associated APP mutations: known clinical data and model predictions.

Mutation	α	σ_{nat}	Q_{nat}	D_{nat}	Onset	Known effect
A2V (protective)	0.08	0.695	0.923	0.016	—	Reduced aggregation
Icelandic (A673T)	0.10	0.707	0.910	0.016	—	40% less A β
Wild type	0.15	0.822	0.877	0.017	age	Normal
Swedish (K670N/M671L)	0.30	1.091	0.823	0.023	50–60 yr	6–8 $\times$ A β
Arctic (E693G)	0.40	1.032	0.710	0.030	50–60 yr	Protofibrils
Dutch (E693Q)	0.42	0.986	0.713	0.031	40–50 yr	CAA
Iowa (D694N)	0.45	1.009	0.637	0.033	50–60 yr	Severe CAA
Flemish (A692G)	0.48	1.077	0.600	0.035	40–50 yr	Mixed
London (V717I)	0.50	1.113	0.573	0.036	40–55 yr	A β 42/40 shift
Osaka (E693 Δ)	0.55	1.098	0.417	0.035	40–50 yr	Oligomers

α **assignment rationale:** The mutation load α was assigned based on published data on

aggregation propensity, production rate, and clinical severity:

- Protective variants (A2V, Icelandic): $\alpha < 0.15$ based on reduced aggregation/production.
- Swedish: $\alpha = 0.30$ based on $6\text{--}8\times$ increased $A\beta$ production (BACE1 cleavage enhancement), but no intrinsic aggregation change.
- Arctic, Dutch, Iowa, Flemish: $\alpha = 0.40\text{--}0.48$ based on enhanced aggregation propensity from in vitro studies.
- London, Osaka: $\alpha = 0.50\text{--}0.55$ based on combined production shift and aggregation enhancement.

7 S7: Intervention simulation details

7.1 Chaperone model

The chaperone intervention restricts the conformational space available to selected residues. For each target residue r , a set of “forbidden” states is defined: all states except the native state and its two nearest neighbors (modular arithmetic on $S = 8$). This reduces the accessible states from 8 to 3 per residue.

Residues are prioritized by their number of native contacts (most connected first). The “dose” is the number of residues restricted.

7.2 Stabilizer model

The stabilizer adds an energy bonus $\varepsilon_{\text{boost}}$ to each native contact:

$$E_{\text{nat}} = -(\varepsilon + \varepsilon_{\text{boost}}) \sum_{(i,j) \in C_{\text{nat}}} \delta(s_i, s_i^*) \delta(s_j, s_j^*)$$

This deepens the native energy minimum without changing the amyloid contacts.

7.3 Complete intervention results at $\alpha = 0.45$

8 S8: Dose–response across disease stages

All stages were treatable. The dose–response relationship is non-linear: early intervention ($\alpha = 0.25$, $\varepsilon = 0.1$) requires $15\times$ less stabilization energy than late intervention ($\alpha = 0.60$, $\varepsilon = 1.5$).

Strategy	Dose	σ_{nat}	Q_{nat}	Q_{amy}	Rescued?
None	—	1.070	0.678	0.064	—
Chaperone	2 residues (10 blocked)	1.093	0.683	0.053	no
Chaperone	4 residues (20 blocked)	1.083	0.678	0.033	no
Chaperone	6 residues (30 blocked)	1.102	0.722	0.069	no
Chaperone	8 residues (40 blocked)	1.068	0.794	0.078	no
Chaperone	10 residues (50 blocked)	1.009	0.856	0.061	near
Chaperone	12 residues (60 blocked)	0.926	0.892	0.039	✓
Stabilizer	$\varepsilon_{\text{boost}} = 0.1$	1.041	0.753	0.072	no
Stabilizer	$\varepsilon_{\text{boost}} = 0.2$	1.112	0.806	0.033	no
Stabilizer	$\varepsilon_{\text{boost}} = 0.3$	1.029	0.842	0.064	no
Stabilizer	$\varepsilon_{\text{boost}} = 0.5$	0.854	0.847	0.069	✓
Stabilizer	$\varepsilon_{\text{boost}} = 0.7$	0.769	0.928	0.039	✓
Stabilizer	$\varepsilon_{\text{boost}} = 1.0$	0.513	0.931	0.053	✓
Combined	4 res. + $\varepsilon = 0.5$	0.896	0.878	0.078	✓
Combined	8 res. + $\varepsilon = 0.5$	0.749	0.914	0.061	✓

Table 5: Minimum stabilizer dose to restore $\sigma_{\text{nat}} < 1$ at each α .

α	Stage	$\sigma_{\text{untreated}}$	$\varepsilon_{\text{rescue}}$	σ_{rescued}
0.20	Healthy	0.965	0.0	0.965
0.25	Early	1.016	0.1	0.883
0.30	Early	1.096	0.1	0.991
0.35	Moderate	1.023	0.2	0.991
0.40	Moderate	1.061	0.3	0.953
0.45	Moderate	1.074	0.5	0.878
0.50	Severe	1.149	0.7	0.902
0.55	Severe	1.097	1.0	0.863
0.60	Critical	1.081	1.5	0.691

9 S9: Experimental validation data

9.1 S9.1: Experimental $\Delta\Delta G \rightarrow \sigma$ for 37 mutations

Table 6 summarizes the experimental $\Delta\Delta G$ values and computed $\sigma = \exp(\Delta\Delta G/NRT)$ at $T = 310$ K for single-point mutations across four proteins, using per-residue normalization consistent with Eq. 5 of the main text. All 32 destabilizing mutations give $\sigma > 1$ and all 5 stabilizing mutations give $\sigma < 1$ (100% classification accuracy). Data sources: Barnase Fersht et al. (1995), CI2 Jackson & Fersht (1991), T4 Lysozyme Eriksson et al. (1992), Human Lysozyme Canet et al. (2002).

Table 6: Experimental $\Delta\Delta G$ and computed $\sigma = \exp(\Delta\Delta G/NRT)$ for 37 mutations.

Protein (N)	Mutation	$\Delta\Delta G$ (kJ/mol)	σ	Effect
Barnase (110)	A32G	+4.6	1.016	destabilizing
Barnase (110)	I51A	+13.4	1.048	destabilizing
Barnase (110)	I76A	+12.1	1.044	destabilizing
Barnase (110)	I88A	+8.8	1.032	destabilizing
Barnase (110)	V10A	+7.1	1.025	destabilizing
Barnase (110)	L14A	+10.9	1.039	destabilizing
Barnase (110)	Y17A	+6.3	1.022	destabilizing
Barnase (110)	F56A	+9.2	1.033	destabilizing
Barnase (110)	T16S	+2.9	1.010	destabilizing
Barnase (110)	D8A	+1.3	1.005	destabilizing
Barnase (110)	K27A	+0.8	1.003	destabilizing
Barnase (110)	E73A	+0.4	1.001	destabilizing
Barnase (110)	N58A	-0.4	0.999	stabilizing
Barnase (110)	D93N	-1.7	0.994	stabilizing
CI2 (64)	I20A	+10.0	1.062	destabilizing
CI2 (64)	L32A	+12.6	1.079	destabilizing
CI2 (64)	I37A	+5.0	1.031	destabilizing
CI2 (64)	A16G	+6.7	1.041	destabilizing
CI2 (64)	V47A	+3.3	1.020	destabilizing
CI2 (64)	E26A	+1.3	1.008	destabilizing
CI2 (64)	K2A	+0.4	1.002	destabilizing
CI2 (64)	D52A	-0.8	0.995	stabilizing
T4 Lysozyme (164)	L99A	+22.2	1.054	destabilizing
T4 Lysozyme (164)	L99G	+27.2	1.066	destabilizing
T4 Lysozyme (164)	A98V	-2.5	0.994	stabilizing
T4 Lysozyme (164)	V149I	-1.3	0.997	stabilizing
T4 Lysozyme (164)	L121A	+13.4	1.032	destabilizing
T4 Lysozyme (164)	F153A	+15.5	1.037	destabilizing
T4 Lysozyme (164)	I3A	+7.5	1.018	destabilizing
T4 Lysozyme (164)	M102A	+3.8	1.009	destabilizing
T4 Lysozyme (164)	T152S	+4.2	1.010	destabilizing

Protein (N)	Mutation	$\Delta\Delta G$ (kJ/mol)	σ	Effect
T4 Lysozyme (164)	A42G	+2.1	1.005	destabilizing
Human Lys. (130)	I56T	+25.1	1.078	amyloidogenic
Human Lys. (130)	D67H	+18.8	1.058	amyloidogenic
Human Lys. (130)	W64R	+12.6	1.038	destabilizing
Human Lys. (130)	T70N	+5.0	1.015	destabilizing
Human Lys. (130)	F57I	+20.9	1.064	amyloidogenic

9.2 S9.2: Go model vs. experimental σ cross-validation

The Spearman rank correlation between Go-model σ_{nat} and experimental $\sigma_{\text{exp}} = \exp(\Delta\Delta G/NRT)$ (with $N = 42$ for A β 42) for the 10 A β mutations is $\rho = 0.988$, with Pearson $r = 0.978$ and 9/10 classification agreement. With per-residue normalization, both σ_{exp} and σ_{Go} fall in the range 0.95–1.12, enabling direct quantitative comparison. The sole discrepancy (Dutch E693Q, $\sigma_{\text{exp}} = 1.036$ vs. $\sigma_{\text{Go}} = 0.986$) reflects this mutation’s atypical mechanism through cerebrovascular amyloid angiopathy rather than parenchymal plaque formation.

Note on non-circularity: The α values used to compute σ_{Go} were derived from aggregation propensity rankings (relative aggregation rates), while σ_{exp} was computed from independently measured $\Delta\Delta G$ values (thermodynamic stability changes). The two datasets are methodologically independent, so the correlation is non-trivial.

9.3 S9.3: σ -drift model parameters

The σ -drift model uses:

- Wild-type $\sigma_0 = 0.822$ at age 30
- Proteostasis decline rate: $\Delta\sigma = 0.03$ per decade (based on Brehme et al. 2014, Hsp70 decline $\sim 2\%$ /decade; Saez & Bhatt 2009, proteasome decline $\sim 1.5\%$ /decade)
- $\sigma(t) = \sigma_0 + (t - 30) \times 0.003$ where t is age in years
- Onset predicted when $\sigma(t) \geq 1.0$

Predicted onset ages: London ~ 37 yr (clinical 40–55), Swedish ~ 47 yr (clinical 50–60), wild type ~ 89 yr (sporadic 75–85), Icelandic > 130 yr (clinically protective).

10 S10: Large-scale validation across 20 proteins

Experimental $\Delta\Delta G$ values were compiled for 277 mutations across 20 proteins (Table ??). Sources include ProTherm Kumar et al. (2006), Guerois et al. (2002), and published mutagenesis studies. For each mutation, $\sigma = \exp(\Delta\Delta G/NRT)$ was computed at $T = 310$ K with protein-specific N .

Classification results: all 277 mutations are correctly classified at the $\sigma = 1$ threshold (destabilizing $\rightarrow \sigma > 1$, stabilizing $\rightarrow \sigma < 1$). This is mathematically guaranteed by the monotonicity of the exponential. The non-trivial observation is that the cross-protein Pearson correlation is $r = 0.774$ and Spearman $\rho = 0.796$, reflecting the N -dependence of the per-residue normalization.

The protein set spans $N = 53$ (Arc repressor) to $N = 316$ (thermolysin) and includes α -helical, β -sheet, α/β , and all- α topologies. Mean σ for destabilizing mutations: 1.025 (median 1.023); for stabilizing mutations: 0.996 (median 0.997).

11 S11: ClinVar VUS predictions for APP

11.1 S11.1: Calibration on known variants

Of 25 known pathogenic APP variants, 24 (96%) receive $\sigma > 1$ using the empirical substitution matrix. The single exception, T714I (Iranian), has negative $\Delta\Delta G$ in the generic matrix because I $\rightarrow$ T is generally stabilizing, despite this mutation’s known pathogenicity through an aggregation-specific mechanism.

Of 3 known benign variants, all receive $\sigma > 1$: A673T ($\sigma = 1.061$), E665D ($\sigma = 1.002$), K670R ($\sigma = 1.002$). Within the σ -drift framework, E665D and K670R are not false positives but late-onset candidates: their σ values are so close to 1.0 that clinical manifestation would not be expected within a normal lifespan (> 120 yr at $\Delta\sigma = 0.03/\text{decade}$). A673T is a genuine limitation—its protective effect operates through reduced β -secretase cleavage rather than thermodynamic stabilization, a mechanism outside the scope of equilibrium $\Delta\Delta G$.

11.2 S11.2: VUS predictions

Twenty-six APP variants classified as VUS in ClinVar were analyzed. Twenty-five receive $\sigma > 1$ (predicted destabilizing). The complete list:

Table 7: σ predictions for APP variants of uncertain significance. Risk category: “high” ($\sigma > 1.05$), “moderate” ($1.02 < \sigma \leq 1.05$), “low” ($\sigma \leq 1.02$). No onset ages are predicted because the σ -drift model has been validated only for TTR, not for APP.

Variant	Position	Region	$\Delta\Delta G$ (kcal/mol)	σ	Risk category
G696R	696	A β	+2.87	1.117	high
D678H	678	A β	+2.33	1.094	high
G700E	700	A β	+2.21	1.089	high
K687E	687	A β	+2.19	1.088	high
G700D	700	A β	+1.94	1.078	high
V695A	695	A β	+1.73	1.069	high
A692T	692	A β	+1.52	1.061	high
V689M	689	A β	+1.42	1.056	high
V710F	710	A β	+1.42	1.056	high
G696S	696	A β	+0.91	1.036	moderate
G709S	709	A β	+0.91	1.036	moderate
V711L	711	A β	+0.89	1.035	moderate
K687R	687	A β	+0.73	1.029	moderate
V695I	695	A β	+0.68	1.027	moderate
V711I	711	A β	+0.68	1.027	moderate
V710I	710	A β	+0.68	1.027	moderate
E682G	682	A β	+0.35	1.014	low
A692V	692	A β	+0.29	1.011	low
L720R	720	APP	+3.06	1.007	low
I716T	716	APP	+2.57	1.005	low
G681A	681	A β	+0.13	1.005	low
T719P	719	APP	+2.31	1.005	low
M722K	722	APP	+2.30	1.005	low
V717A	717	APP	+1.73	1.004	low
T719N	719	APP	+1.36	1.003	low
G696V	696	A β	−0.08	0.997	stabilizing

These predictions are exploratory: they apply the σ framework to variants whose clinical

significance is unknown, using generic substitution propensities rather than experimentally measured $\Delta\Delta G$. $\sigma > 1$ indicates predicted thermodynamic destabilization, not a clinical diagnosis. The risk categories are qualitative and should be interpreted as prioritization for experimental follow-up.

12 S12: Code availability

All code is available at <https://github.com/forgottenforge/protein-sigma>.

All scripts require only Python 3 with NumPy, SciPy, and Matplotlib. No specialized molecular dynamics software is needed.

13 S13: Transthyretin validation data

Table 9 presents the complete TTR mutation dataset used in Section 3.8 of the main text. Spearman $\rho(\sigma, \text{onset}) = -0.977$ ($p < 10^{-15}$, $n = 24$). MAE of σ -drift predicted onset = 9.2 years (bootstrap 95% CI: 7.0–11.4 years, 10^4 resamples). T119M correctly classified as protective. $\Delta\Delta G$ data from Sekijima et al. (2005) and Hammarström et al. (2002); phenotype and onset data from Connors et al. (2003).

13.1 S13.2: Leave-one-out analysis

We performed a jackknife analysis, removing each of the 24 pathogenic mutations in turn and recomputing $\rho(\sigma, \text{onset})$. Across all 24 iterations, ρ ranges from -0.968 to -0.982 (mean -0.977), confirming that no single mutation drives the correlation. The three most influential mutations—D18G, L55P, and G6S—span the full σ range (0.930–0.975), indicating that the correlation is not anchored by outliers at either extreme.

13.2 S13.3: Within-phenotype correlations

Phenotype-stratified analysis: FAP mutations ($n = 16$) show $\rho = -0.965$; FAC mutations ($n = 4$) show $\rho = -0.80$ (n too small for significance); CNS mutations ($n = 2$) cannot be tested. The correlation is not driven by between-phenotype differences.

13.3 S13.4: Onset data caveats

Published “onset ages” for TTR mutations are typically age at diagnosis, not true symptom onset, introducing right-censoring bias of uncertain magnitude. For mutations with reported

onset ranges, we used the most commonly reported value:

- V30M: 33 yr (Portuguese early-onset; Swedish late-onset at 50–60 yr exists but represents a distinct population with modifier alleles).
- V122I: 60 yr (African-American population; onset varies 55–70 yr).

A sensitivity analysis using alternative V30M onset (50 yr, Swedish) changes overall ρ from -0.977 to -0.964 ; using 25 yr (earliest reported) gives $\rho = -0.981$. The correlation is robust to the specific V30M onset value chosen.

14 S14: Benchmark against alternative early-warning indicators

To test whether the early-warning property is specific to σ or a generic feature of the Go model’s critical transition, we benchmarked σ against four alternative indicators across all 72 parameter combinations from the robustness sweep (S3):

1. **ΔG threshold:** native basin free energy crosses zero.
2. **Q -variance susceptibility (χ):** fluctuation of native contact fraction exceeds 50% of its maximum value.
3. **Critical slowing down (CSD):** relaxation time τ exceeds $5\times$ the baseline τ_0 .
4. **D -only:** D_{nat} drops below half its initial value.

Results. All five indicators achieve 72/72 early warning. Mean lead times: Q -variance 55.5%, CSD 47.6%, σ 35.9%, ΔG threshold 36.0%, D -only 36.3%. The Go model’s transition is sufficiently sharp that multiple indicators detect it; σ ’s contribution is interpretability (product decomposition, protein-independent threshold, drift model), not superior detection.

15 S15: Hereditary lysozyme amyloidosis validation data

Human lysozyme (LYZ, $N = 130$) causes hereditary systemic amyloidosis through partial unfolding of the globular domain. $\Delta G_{\text{wt}} = 9.3$ kcal/mol at pH 5.0, 37°C Booth et al. (1997). σ computed via Eq. 6 of the main text.

Spearman $\rho(\sigma, \text{onset}) = -0.714$ ($n = 6$). At $n = 6$, the minimum detectable $|\rho|$ at 80% power ($\alpha = 0.05$, two-tailed) is approximately 0.81; this result does not reach conventional

significance on its own. The σ -drift model (same 0.03/decade rate as TTR, not refit) gives MAE = 7.2 years. One $\Delta\Delta G$ value (L84S) is estimated from clinical severity rather than direct measurement.

16 S16: Gelsolin amyloidosis validation data

Gelsolin (GSN) causes Finnish hereditary amyloidosis (AGel) through destabilization of domain 2 ($N = 117$, $\Delta G_{\text{wt}} \approx 6.0$ kcal/mol).

For the three mutations with onset data, $\rho = -1.0$ and MAE = 3.8 years. With $n = 3$, no meaningful statistical inference is possible. The result is consistent with the framework but gains evidential weight only as part of the multi-protein pattern (Table 5 of the main text).

17 S17: Negative control data

Pre-specified negative control diseases were selected based on mechanistic criteria before σ was computed: (i) toxic gain-of-function (SOD1/ALS), (ii) templated conformational conversion (prion disease), (iii) intrinsically disordered substrate (A β 42). All data sources and analysis code are provided in the repository.

17.1 S17.1: SOD1/ALS

SOD1 ($N = 153$, $\Delta G_{\text{wt}} \approx 13.5$ kcal/mol). Ten mutations with published $\Delta\Delta G$ and ALS onset ages.

$\rho(\sigma, \text{onset}) = +0.28$ ($p = 0.43$)—wrong sign. G37R (modest $\Delta\Delta G$) has earliest onset; H46R (large $\Delta\Delta G$) has late onset. Stability loss is necessary but not sufficient; toxic gain-of-function determines severity.

17.2 S17.2: Prion disease (PRNP)

PRNP ($N_{\text{domain}} = 104$, $\Delta G_{\text{wt}} \approx 6.0$ kcal/mol). Seven mutations with published $\Delta\Delta G$.

$\rho(\sigma, \text{onset}) = -0.29$ ($p = 0.53$), not significant. All $\sigma < 1$ (protein remains thermodynamically stable). Drift-model onset predictions (160–270 yr) are clearly wrong—prion disease requires PrP^{Sc}-templated conversion, not spontaneous unfolding.

17.3 S17.3: A β 42 cross-method disagreement

Two stability-prediction tools (DynaMut2: structure-based $\Delta\Delta G$; ESM-1v: evolutionary log-likelihood ratio) were applied to 31 A β 42 variants (pathogenic + VUS). Inter-method Spearman correlation: $\rho = -0.097$ ($p = 0.65$, $n = 31$). This confirms that folding-stability tools do not produce consistent results for intrinsically disordered peptides, as expected: A β 42 has no stable reference fold for $\Delta\Delta G$ to be defined against.

References

- Canet, D., Last, A. M., Tito, P., et al. (2002). Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nature Struct. Biol.*, 9, 308–315.
- Eriksson, A. E., Baase, W. A., Zhang, X.-J., et al. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, 255, 178–183.
- Fersht, A. R., Matouschek, A., & Serrano, L. (1995). The folding of an enzyme. *J. Mol. Biol.*, 224, 771–782.
- Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. *Biochemistry*, 30, 10428–10435.
- Connors, L. H., Lim, A., Prokaeva, T., Roskens, V. A., & Costello, C. E. (2003). Tabulation of human transthyretin (TTR) variants. *Amyloid*, 10, 160–184.
- Hammarström, P., Jiang, X., Hurshman, A. R., Powers, E. T., & Kelly, J. W. (2002). Sequence-dependent denaturation energetics: a major determinant in amyloid disease diversity. *PNAS*, 99, 16427–16432.
- Sekijima, Y., Wiseman, R. L., Matteson, J., et al. (2005). The biological and chemical basis for tissue-selective amyloid disease. *Cell*, 121, 73–85.
- Kumar, M. D. S., Bava, K. A., Gromiha, M. M., et al. (2006). ProTherm and ProNIT: thermodynamic databases for proteins and protein–nucleic acid interactions. *Nucleic Acids Res.*, 34, D204–D206.
- Booth, D. R., Sunde, M., Bellotti, V., et al. (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature*, 385, 787–793.

Table 8: Simulation scripts.

Script	Description
protein_folding_cbfi.py	Thermodynamic $\sigma(T)$ for real proteins from published data
protein_telescope_v2.py	Single-basin Go model with σ_{macro} validation
protein_dual_basin.py	Dual-basin model: native vs. amyloid, α scan, transient σ
protein_dual_basin_intervention.py	Chaperone and stabilizer interventions, dose-response
protein_alzheimer_mutations.py	Alzheimer mutation analysis, therapeutic targets
protein_validation.py	Experimental $\Delta\Delta G$ validation, cross-protein analysis, σ -drift
protein_largescale_validation.py	Large-scale validation (20 proteins, ClinVar VUS predictions)
robustness_sweep.py	Parameter sweep: early-warning robustness across N , S , contact topologies
ttr_validation.py	Transthyretin mutation validation against clinical onset ages
hdx_proxy.py	HDX-MS/NMR order parameter proxy for σ
drift_sensitivity.py	Drift-rate sensitivity analysis (0.02–0.05/decade)
benchmark_early_warning.py	Benchmark: σ vs. 4 alternative early-warning indicators
lyz_validation.py	Hereditary lysozyme amyloidosis validation
gelsolin_validation.py	Gelsolin amyloidosis validation
sod1_validation.py	SOD1/ALS negative control analysis
prnp_validation.py	Prion disease (PRNP) negative control analysis
structure_based_ddg.py	DynaMut2 + ESM-1v cross-method analysis for A β 42
paper5_figures.py	Figure generation for main text and supplementary

Table 9: Complete TTR validation data. σ computed from $\sigma_{\text{mut}} = \sigma_{\text{wt}} \cdot \exp(\Delta\Delta G/NRT)$ with $N = 127$, $T = 310$ K, $\Delta G_{\text{unfold}}^{\text{wt}} = 25$ kJ/mol. Onset envelope spans drift rates 0.02–0.05/decade (central: 0.03/decade) from age 30.

Mutation	$\Delta\Delta G$ (kcal/mol)	σ	Phenotype	Onset (actual)	Onset (central)	Onset (envelope)
T119M	−0.8	0.917	protective	—	never	never
G6S	+0.3	0.930	FAC	72 yr	53	44–65
V14A	+0.4	0.931	mixed	70 yr	53	44–64
A97S	+0.6	0.934	FAP	68 yr	52	43–63
R104H	+0.8	0.936	FAC	65 yr	51	43–62
H88R	+0.9	0.937	FAP	62 yr	51	43–61
E89K	+1.0	0.938	FAP	57 yr	51	42–61
Y78F	+1.1	0.940	FAP	58 yr	50	42–60
V71A	+1.2	0.941	FAP	54 yr	50	42–60
S77Y	+1.3	0.942	FAP	55 yr	49	42–59
I107V	+1.4	0.943	FAC	55 yr	49	41–59
V122I	+1.5	0.944	FAC	60 yr	49	41–58
G47R	+1.5	0.944	FAP	52 yr	49	41–58
E54K	+1.6	0.946	FAP	50 yr	48	41–57
F64L	+1.7	0.947	FAP	48 yr	48	41–57
T60A	+1.8	0.948	FAP+FAC	45 yr	47	40–56
S50R	+1.9	0.949	FAP	43 yr	47	40–56
I84S	+2.0	0.950	FAP+FAC	42 yr	47	40–55
V30M	+2.1	0.952	FAP	33 yr	46	40–54
L58H	+2.3	0.954	FAP	40 yr	45	39–53
A36P	+2.5	0.957	FAP	38 yr	45	39–51
Y114C	+2.8	0.960	FAP	30 yr	43	38–50
A25T	+3.2	0.965	CNS	28 yr	42	37–47
L55P	+3.6	0.970	FAP	20 yr	40	36–45
D18G	+4.0	0.975	CNS	25 yr	38	35–43

Table 10: Complete lysozyme validation data.

Mutation	$\Delta\Delta G$ (kcal/mol)	σ	Onset (actual)	Onset (central)	Source
T70N	+0.8	0.899	60 yr	64	experimental
L84S	+1.2	0.904	55 yr	62	estimated
D67H	+1.5	0.907	55 yr	61	experimental
F57I	+1.8	0.911	50 yr	60	experimental
I56T	+2.1	0.914	55 yr	59	experimental
W64R	+3.5	0.930	40 yr	53	experimental

Table 11: Complete gelsolin validation data.

Mutation	$\Delta\Delta G$ (kcal/mol)	σ	Onset	MAE	Source
N184K	+2.0	0.946	55 yr	2.0	estimated
D187Y	+2.5	0.953	45 yr	3.0	experimental
D187N	+3.0	0.959	40 yr	6.4	experimental
A174P	+2.8	0.957	—	—	DynaMut2
G167R	+3.5	0.966	—	—	DynaMut2

Table 12: SOD1 negative control data.

Mutation	$\Delta\Delta G$ (kcal/mol)	σ	Onset (actual)	Mechanism
G37R	+1.8	0.918	35 yr	GOF
H46R	+3.0	0.925	55 yr	metal-binding loss
G85R	+2.5	0.922	45 yr	GOF
A4V	+4.0	0.931	40 yr	GOF (aggressive)
D90A	+1.2	0.915	50 yr	recessive
E100G	+2.0	0.919	48 yr	GOF
L38V	+1.5	0.917	42 yr	GOF
G93A	+2.2	0.920	45 yr	GOF
I113T	+1.0	0.914	52 yr	GOF
L144F	+2.8	0.924	50 yr	GOF

Table 13: PRNP negative control data.

Mutation	$\Delta\Delta G$ (kcal/mol)	σ	Onset (actual)	Onset (σ -drift)	Disease
D178N	+1.5	0.947	50 yr	207	FFI
E200K	+1.0	0.940	58 yr	230	CJD
V210I	+0.8	0.937	55 yr	241	CJD
T188K	+1.2	0.943	52 yr	219	CJD
F198S	+2.0	0.954	45 yr	183	GSS
P102L	+1.8	0.951	48 yr	193	GSS
A117V	+0.5	0.933	40 yr	262	GSS