

# Thermodynamic Stabilization of Human Frataxin

Reyes Núñez-Franco<sup>†</sup>, Angel Torres-Mozas<sup>†</sup>, Claudio D. Navo<sup>†</sup>, Andreas Schedlbauer<sup>†</sup>, Mikel Azkargorta<sup>†</sup>, Ibon Iloro<sup>†</sup>, Félix Elortza<sup>†</sup>, Gabriel Ortega<sup>†,‡</sup>, Oscar Millet<sup>†,¶,\*</sup>, Francesca Peccati<sup>†,\*</sup>, Gonzalo Jiménez-Osés<sup>†,‡,\*</sup>

<sup>†</sup>Basque Research & Technology Alliance (BRTA), CIC bioGUNE, Bizkaia Technology Park, Building 800, 48160 Derio, Spain.

<sup>‡</sup>Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Spain.

<sup>¶</sup>ATLAS Molecular Pharma, Bizkaia Science and Technology Park, 48160 Derio, Spain.

<sup>‡</sup>Biomedical Research Network on Hepatic and Digestive Diseases (CIBEREHD), Instituto de Salud Carlos III, 28029 Madrid, Spain.

**ABSTRACT.** Recombinant proteins and antibodies are routinely used as drugs to treat prevalent diseases such as diabetes or cancer, while enzyme replacement and gene therapies are the main therapeutic intervention lines in rare diseases. In protein-based therapeutics, optimized *in vivo* stability is key as intrinsic denaturation and intracellular proteostatic degradation will limit potency, particularly in treatments requiring a sustained action, while clearance mechanisms may limit the amount of circulating protein. *In vivo* stability is ultimately correlated with the intrinsic thermodynamic stability of the biomolecule, but this is difficult to optimize because it often goes at the expense of reducing protein activity. Here, we have used *in silico* engineering approaches to thermodynamically stabilize human frataxin, a small mitochondrial protein that acts as an allosteric activator for the biosynthesis of Fe-S clusters, whose genetically-driven impairment results in a rare disease known as Friedreich ataxia. Specifically, we developed an efficient thermostability engineering computational approach that combines information on amino acid conservation, the Rosetta energy function, and two recent artificial intelligence tools – AlphaFold and ProteinMPNN – to produce thermodynamically stabilized variants of human frataxin. Such protein variants rescued the large destabilization exerted by well-known pathological mutations, with an increase over 20 °C in the melting temperature and a thermodynamic stabilization of more than 3 kcal·mol<sup>-1</sup> at the physiological temperature. This stability surplus is translated into an enhanced resistance to proteolysis, while maintaining the protein fully functional. This case-study highlights the power of our combined computational approach to generate optimized variants, adequate for protein-based therapeutics.

## Introduction

Protein-based therapeutics offer advantages over small molecule drugs in terms of potency and the versatility of their functions (e.g. catalysis, signaling, transport); also, having evolved to carry out very specific roles, they tend to induce less side-effects.<sup>1</sup> Since the first FDA approval of a recombinant protein (insulin, 1982), ~350 protein-based therapeutics have been developed, including antibodies (e.g. Blinatumomab<sup>2</sup> for the treatment of acute lymphoblastic leukemia), enzymes (e.g. Pegademase bovine<sup>3</sup> for the treatment of adenosine deaminase deficiency), coagulation factors (e.g. Eptacog Alfa for the treatment of haemophilia<sup>4</sup>), protein hormones (e.g. insulin detemir for the treatment of diabetes mellitus<sup>5</sup>) and cytokines (e.g. interferon beta-1b for the treatment of multiple sclerosis<sup>6</sup>). Protein therapeutics classified as Group I encompass exogenous enzymes and regulatory proteins used to mitigate the consequences of the deficiency of a natural protein, either caused by its absence or lack of function due to pathological mutations.<sup>7</sup> Classic examples include the use of insulin for the treatment of diabetes, or the use of pancreatic enzymes isolated from pigs for the treatment of cystic fibrosis.<sup>7</sup>

The success of protein-based therapeutics is often constrained by the stability of the biomolecular active principle. First, many examples show that an *in vitro* reduced thermodynamic stability is coupled with an impaired protein homeostasis *in vivo*<sup>8</sup> due to premature unfolding and subsequent degradation of the active species in the proteasome<sup>9</sup> or other protein clearance pathways.<sup>10</sup> In principle, this can be overcome with protein variants that stabilize the folded conformation *at the physiological temperature*. For this protein *thermodynamic stabilization* strategy to work, the optimized variants should resemble its natural analogue as closely as possible, to retain the native function and to avoid triggering an immune response.<sup>11</sup>

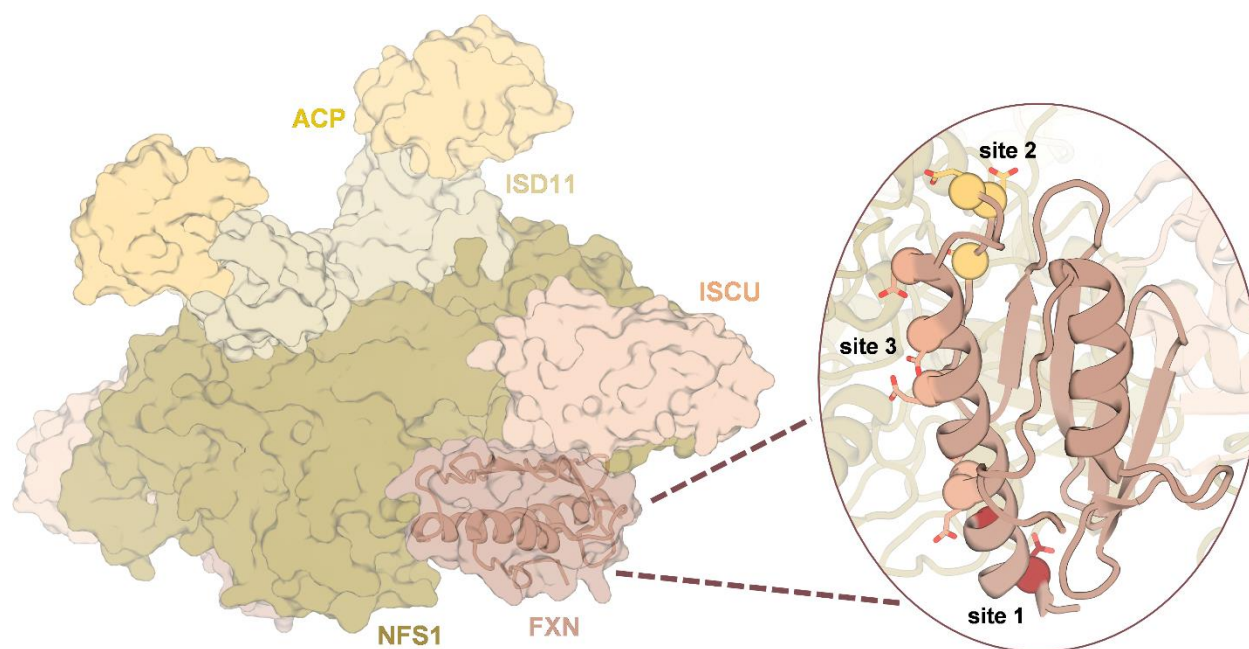
Several strategies have been used for optimizing protein-based therapeutics by engineering their physicochemical properties.<sup>1,11</sup> These approaches seek to control protein homeostasis, to improve their bioavailability and to increase the time they remain functional in the body. Some examples include strategies can improve bioavailability by increasing the hydrodynamic diameter of the protein to reduce kidney filtration through the chemical modification of the protein surface with polymer grafts such as polyethylene glycol, or their fusion to larger, more soluble proteins.

In turn, thermodynamic stabilization and solubility can be achieved by engineering the amino acid sequence.<sup>1</sup> This approach is particularly attractive because it does not require chemical modifications, thus improving reproducibility and yield. It has been successfully used to modulate, for example, the pharmacokinetics of insulin. Engineering insulin's amino acid sequence to modulate its isoelectric point has provided the long-acting variant (24 h) glargine,<sup>12</sup> where the slower absorption is the consequence of an increased precipitation resulting from an isoelectric point closer to physiological pH, as well as the faster acting variant glulisine, where shifting the isoelectric point in the opposite direction reduces the formation of oligomers and facilitates

absorption.<sup>13</sup> The mutagenesis approach has also been used on interferon  $\beta$ -1B, interleukin-2, and the human growth hormone to cite a few.<sup>14–16</sup>

Friedreich's ataxia is a disease amenable to treatment with Group I protein-therapeutics. This autosomal recessive disorder with early onset (10–15 years) manifests with progressive ataxia, loss of tendon reflexes and dysarthria, with a prevalence of 1–2 per 50,000.<sup>17</sup> All Friedreich's ataxia patients present low levels of functional frataxin, a small iron storage protein that chaperones ferrochelatase and also acts as an allosteric modulator for iron-sulfur clusters biosynthesis by binding to the assembly complex in the mitochondria comprising proteins NFS1, ISD11, ACP, and ISCU (Figure 1). Ultimately, the disease leads to decreased maturation of iron-sulfur cluster proteins, including mitochondrial respiratory complexes, the proteins of the Krebs cycle, and proteins involved in DNA repair and replication.<sup>17</sup> Human frataxin undergoes maturation from a 210 amino acid precursor protein, and after proteolytic processing into mitochondria is converted to the mature, functional form which comprises amino acids 81-210.<sup>18</sup> From a structural point of view, it is composed of a flexible N-terminus, a conserved antiparallel  $\beta$ -sheet, involved in protein-protein recognition, two  $\alpha$ -helices, and an unstructured C-terminus.<sup>19</sup> Each frataxin unit can bind up to 7-10  $\text{Fe}^{2+}$  ions through a conserved anionic surface (i.e., the acidic ridge, Figure 1).<sup>20,21</sup>  $\text{Fe}^{2+}$ -bound frataxin specifically binds protoporphyrin IX — which features the same binding epitope as ferrochelatase — with micromolar affinity.<sup>22</sup>

In over 96% of the patients, homozygous expansions of GAA trinucleotide repeat in intron 1 of the frataxin (FXN) gene are observed, triggering local chromatin changes leading to transcriptional repression of the gene (FXN mRNA levels reduced by 70–95%).<sup>17</sup> In few other cases, Friedreich's ataxia patients present a GAA expansion on one allele and a *missense mutation* on the other allele of the FXN gene; over 10 missense mutations have been identified, among them I154F and I198R, which decrease frataxin levels compromising its stability, solubility, and function.<sup>23,24</sup> While no cure for this disease has been approved yet, there are indications that raising frataxin levels can inhibit and rescue the associated condition, opening opportunities for replacement approaches based on gene therapy.<sup>25–27</sup> Alternatively, protein replacement therapy is a promising option provided that frataxin can be delivered to the mitochondria.<sup>17</sup> In this regard, a Phase 2 study in humans is currently ongoing to explore the use as therapeutic of a fusion protein combining FXN into CTI-1601 (Larimar Therapeutics, Inc. ClinicalTrials.gov Identifier: NCT05579691), a protein that relies on the cell-penetrating ability of the trans-activator of transcription (TAT) peptide for targeted delivery.<sup>28,29</sup> In the context of improving protein replacement therapy alternatives, engineering the homeostasis of frataxin by increasing its thermodynamic stability, and solubility may improve its resistance to intracellular degradation and, eventually, its bioavailability, thus optimizing its therapeutic effect.



**Figure 1.** Cryo-EM structure of the frataxin-bound active human complex (PDB ID: 6NZU).<sup>30</sup> The complex contains two copies of the NFS1-ISD11-ACP-ISCU-FXN hetero-pentamer. NFS1: cysteine desulfurase, mitochondrial (colored in green); ISD11: LYR motif-containing protein 4 (colored in light brown); ACP: acyl carrier protein (colored in orange); ISCU: iron-sulfur cluster assembly enzyme (colored in orange); FXN: frataxin (colored in dark red). A close-up view of frataxin is shown in the circle. The spheres represent the residues involved in iron binding, with each color representing one potential iron binding site, capable of binding more than one ion<sup>31</sup> (site 1: E92, E96, site2: E121, D122, D124, site3: E100, E101, D104, E108, E111, D112, D115).

In this contribution, we focus on the structured section of mature human frataxin comprising residues 91–210, and use computational design to propose engineered thermostable variants of both the wild-type frataxin and the pathological single mutants I154F and L198R as candidates for protein replacement therapy. Advancements in Artificial Intelligence (AI) have recently provided programs capable of predicting highly accurate three-dimensional protein models from sequence (e.g. AlphaFold,<sup>32</sup> RoseTTAFold,<sup>33</sup> ESMFold<sup>34</sup>), which are progressively closing the sequence-structure gap<sup>35</sup> and provide exciting opportunities for the development of novel *in silico* protein design and engineering approaches.<sup>36–38</sup> Our group has recently combined AlphaFold structure ensembles and Rosetta energy predictions to develop a computational approach to predict protein thermostability changes upon mutation which has been validated on several families of enzymes engineered through directed evolution.<sup>39</sup> In this work, we combine i) mutational hotspot selection from evolutionary information, ii) sequence sampling on the selected hotspots with ProteinMPNN<sup>38</sup> — a deep learning approach for sequence optimization of a given fold — and iii) our AlphaFold/Rosetta-based approach for thermostability evaluation to design stabilized, biologically competent frataxin variants. These variants have been expressed in *Escherichia coli*

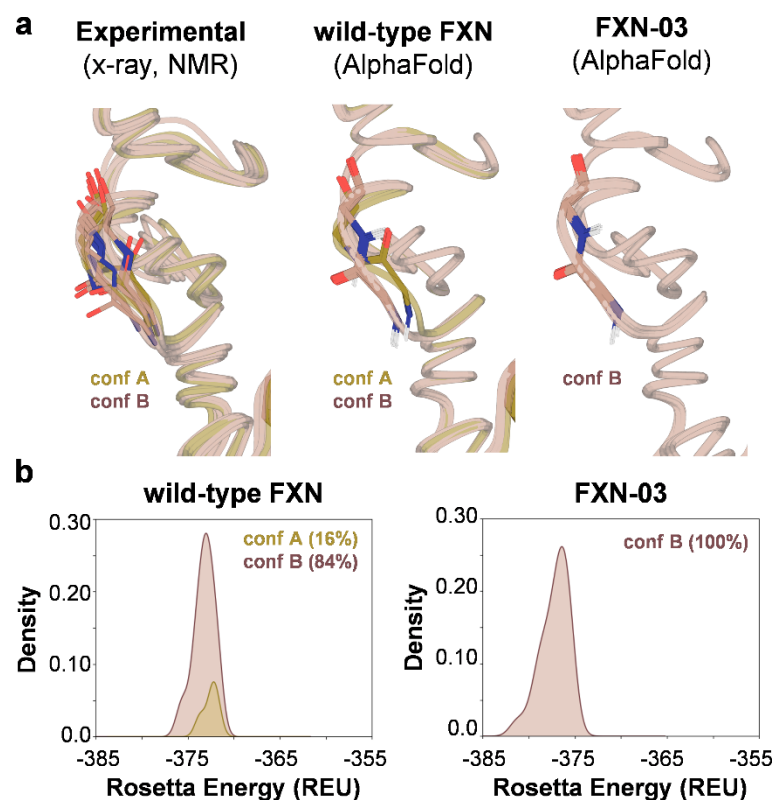
and their melting temperatures ( $T_m$ ) characterized. All 26 tested designs show equal or higher thermostability and unfolding reversibility than the corresponding natural sequence. Variant FXN-10 achieves a large thermostabilization of  $\Delta T_m = +23$  °C over the wild-type. Remarkably, the same variant also shows enhanced thermodynamic stabilization at physiological temperature, leading to increased resistance to proteolysis while maintaining the same ability to bind divalent cations and the FeS assembly complex as wild-type frataxin, representing a promising candidate for protein replacement therapy.

## RESULTS AND DISCUSSION

**Designs based on the consensus approach.** Thermostability is an important readout of protein thermodynamic stability, and engineering this property aims at broadening the temperature range of stability and function of proteins and is crucial for the development of industrial and biomedical applications.<sup>40</sup> The consensus approach has proven successful for the thermostabilization of enzyme families such as phytase-1 achieving stabilizations over 20 °C arising from the sum of small individual contributions.<sup>41,42</sup> Hence, and in order to set a baseline for the thermostabilization that can be achieved through evolutionary information while preserving biological function, we first set out to design a consensus variant (FXN-01, Table 1) using the 7 mutational hotspots derived from a conservation analysis of the frataxin-enriched multiple sequence alignment MSA80 (alignment of sequences with at least 80% identity to human frataxin; positions with conservation below 60% selected as mutational hotspots; see Supporting Information for details), and selecting the most abundant amino acid at each position. Consensus variant FXN-01 was predicted by our AlphaFold/Rosetta-based protocol to be more stable than the wild-type and indeed presents a  $\Delta T_m$  of +6.1 °C. The F120P single mutation – located in a loop region – carries the most thermostabilizing effect, as the remaining six mutations are nearly thermoneutral (variant FXN-02, Table 1;  $\Delta T_m = +0.4$  °C) and the single mutation alone (variant FXN-03, Table 1) induces a +5.1 °C change in  $T_m$ . Regarding the beneficial F120P substitution, loop mutations to proline locally increase the rigidity of the protein in both the native and denatured states,<sup>43</sup> and are a classic thermostabilizing strategy observed in thermophilic-mesophilic protein pairs.<sup>44</sup>

Inspection of the available of X-ray and NMR structures of wild-type frataxin and Friedreich's Ataxia single mutants<sup>19,30,45–47</sup> shows a flexible region encompassing residues 136–141 that is in direct contact with Phe120 (Figure 2a). Accordingly, AlphaFold ensembles of wild-type frataxin carry information on this flexibility showing significant population of two different backbone conformations for residues 138 and 139 (Figure 2a), confirming emerging reports on the ability of AlphaFold to predict not only static but also dynamic protein features.<sup>39,48</sup> In the ensembles obtained for FTX-03, one of those backbone conformations is almost exclusively selected, suggesting that the F120P mutation stabilizes not only of the loop region where it is inserted, but also of the neighboring 137–140 region. Of note, this selected conformation is more stable in terms of Rosetta energy, particularly in FXN-03 (Figure 2b).





**Figure 2.** a) Flexible region between residues 137 to 140 in a) an overlay of NMR and X-ray structures (PDB ID: 1EKG, 3S4M, 3S5D, 3S5E, 3S5F, 3T3J, 3T3L, 3T3T, 3T3X, 1LY7, 6NZU) of different frataxin variants, and AlphaFold ensembles (25 structures) of wild-type frataxin and F120P single mutant FXN-03. While two backbone conformations are observed for residues 138 and 139 in both the experimental and AlphaFold predicted structures for wild-type frataxin, only one conformation is observed for FXN-03. b) Normalized kernel density estimates of the Rosetta energy distributions calculated for each variant in the two conformations. REU: Rosetta energy units.

Intrigued by the evolutionary significance of the highly conserved F120P mutation, we performed a phylogenetic analysis of the frataxin-enriched MSA80 alignment which clearly identified this substitution, together with two additional mutations present in consensus variant FXN-01 (S160T and T191S), as branch points between different orders of mammals and amphibians/reptilians/birds (see Supporting Information). Of note, by combining these mutations into the triple mutant FXN-04, predicted by our protocol to be at least as stabilizing as FXN-01, a further increase in  $T_m$  of +8.6 °C over the wild-type was achieved.

**ProteinMPNN sequence design.** Recently, a very fast and robust deep learning-based protein sequence approach (ProteinMPNN) has been developed by the Baker lab, which achieves very high protein sequence recovery on native backbones without carrying out large scale sidechain packing calculations.<sup>38</sup> By sampling with ProteinMPNN the same 7 mutational hotspots found in the MSA80 dataset (see Supporting Information), three more variants were produced: FXN-05 and

FXN-06 (5 mutations each) predicted to be the two most stable sequences, and FXN-07 as the most frequent solution (5 mutations) (Table 2). All three designs showed increased  $T_m$  over wild-type FXN between +2.5 and +7.3 °C, matching or even improving the thermostabilization achieved by the consensus approach.

To ascertain whether a more extensive sequence sampling can lead to better thermostabilization, we sampled with ProteinMPNN the larger pools of hotspots generated from the MSA70 and MSA60 sequence alignments (sequences with a minimum 70% and 60% sequence identity to human frataxin, where 11 and 26 positions show conservation below 60%, respectively, and are set as mutable; see Supporting Information), obtaining FXN-08 (9 mutations) and FXN-09 (16 mutations) as the variants predicted by our AlphaFold/Rosetta protocol to be the most stable from each pool, and more stable than all previous designs. With these variants, we achieved an even larger increase in  $T_m$  over the wild-type (+11.9 °C and +12.7 °C, respectively) (Table 2). These results strongly suggest that improved thermostabilization can indeed be achieved by more extensive sequence design. Of note, ProteinMPNN only mutates a fraction of the available hotspots to achieve a significant improvement of the target properties.<sup>38</sup>

An alternative way of exploring evolutionary information without imposing an arbitrary per-residue conservation threshold, is to compute the per-position Shannon entropy<sup>49</sup> (SE) on a given MSA. SE not only accounts for the degree of conservation but also for the global distribution of amino acid identities across different sequences. Hence, those positions which vary more frequently and to more diverse amino acids across the MSA are characterized by the highest SE values, and can be selected as mutational hotspots (see Supporting Information). Sampling with ProteinMPNN the 20 mutational hotspots identified through Shannon entropy on MSA80 yielded FXN-10 (13 mutations) – predicted to be the most stable of this pool – which shows a sizeable +23.3 °C increase in  $T_m$  over the wild-type. (Table 2). Of note, our computational protocol produced a superstable frataxin variant by efficiently selecting only 13 mutations out of a relatively small mutational space (20 positions); importantly, this significant thermostability increase was achieved in a single design step, unlike the multiple rounds of extensive mutation required by iterative saturation mutagenesis or directed evolution,<sup>50,51</sup> and providing thermostabilization comparable to the most successful computational approaches<sup>52–54</sup> at a fraction of the computational cost (less than 1 hour per variant on a mid-range workstation, see Supporting Information).

**Table 2.** Summary of the thermostabilized designs for wild-type frataxin, with indication of the method used for hotspot selection, the number of mutational hotspots (N. hotspots), the approach followed to generate the mutations, the name of the variant, the number of mutations (N. mutations), the difference in melting temperature ( $\Delta T_m$ ), and the predicted unfolding free energy change ( $\Delta\Delta G^{calc}$ ) (REU: Rosetta energy units). All quantities are referred to wild-type frataxin. The identity of the mutations for each variant are presented in Table S10. See Supporting Information for the definition of multiple sequence alignments MSA80, MSA70 and MSA60. All measured  $T_m$  values are summarized in Table S11.

Hotspot detection	N. hotspots	Mutation approach	Variant	N. mutations	$\Delta T_m$ (°C)	$\Delta\Delta G^{calc}$ (REU)
Per-residue conservation $\leq 60\%$ on MSA80	7	Consensus	FXN-01	7	6.1 <sup>a</sup>	8.2
			FXN-02	6	0.4 <sup>a</sup>	3.1
			FXN-03	1	5.1	4.1
			FXN-04	3	8.6	7.7
		ProteinMPNN	FXN-05	5	7.3	11.9
			FXN-06	5	4.8	10.6
			FXN-07	5	2.5	9.0
Per-residue conservation $\leq 60\%$ on MSA70	11	ProteinMPNN	FXN-08	9	11.9	12.2
Per-residue conservation $\leq 60\%$ on MSA60	26	ProteinMPNN	FXN-09	16	12.7	14.3
Shannon entropy (SE > 0.65) on MSA80	20	ProteinMPNN	FXN-10	13	23.3	12.7

<sup>f</sup>The difference in melting temperature ( $\Delta T_m$ ) for these variants is calculated with respect to non-6xHis-tagged wild-type frataxin.

### Rescue of pathological mutants I154F and L198R

Having succeeded in thermostabilizing wild-type FXN, we turned to thermostabilize two pathological single mutants, FXN-I154F and FXN-L198R, which induce large decreases in  $T_m$  over 10 °C. The role of pathological mutations in Friedreich's ataxia has been widely discussed, as different mutations lead to specific phenotypes of the disease tuning to different extents the protein levels, localization and function.<sup>55</sup> Full-length frataxin has an unstructured region at the N-terminus encompassing the first 41 amino-acids, and its processing to the mature, functional form (residues 81-210) takes place through at least one intermediate (residues 42-210).<sup>55</sup> Pathological mutation I154F ( $\Delta T_m = -12.9$  °C, Figure S4) has been shown to decrease both mature frataxin levels and accumulation of the intermediate without impairing its association with mitochondria,<sup>23</sup> as well as promoting precipitation upon iron binding due to an increased flexibility.<sup>56</sup> On the other hand, pathological mutation L198R ( $\Delta T_m = -20.5$  °C, Figure S4), located in the C-terminal region,



places a positive charge in an apolar environment and destabilizes the native state inducing a more dynamic behavior in the microsecond/millisecond timescale around the mutation.<sup>21,55,57</sup>

For each of these two pathological variants, eight designs were generated by i) transplanting either one or the three stabilizing mutations identified in the phylogenetic analysis (F120P, S160T, T191S), ii) resampling with ProteinMPNN the same 26 mutational hotspots sampled in the wild-type, and iii) transplanting the mutations of the most stable engineered variants of the wild-type (FXN-09 and FXN-10; the L198I mutation of FXN-09 was not transplanted to FXN-L198R in order to preserve the original pathogenicity). Designs FXN-11 to FXN-18 (derived from FXN-I154F) and FXN-19 to FXN-26 (derived from FXN-L198R), were therefore produced (Tables 3 and S8). Regarding the I154F mutation, five variants achieve at least +20 °C increase in  $T_m$ , which not only fully compensates the destabilizing effect of the pathological mutation, but also rises the stability over that of the wild-type. Similar results were obtained for the pathological mutation L198R, where four variants show  $\Delta T_m \geq +24$  °C. In both cases, the largest stabilizations were achieved by transferring the mutations from FXN-10 (variants FXN-18 and FXN-26), although resampling with ProteinMPNN (variants FXN-15/16 and FXN-23/24) achieved very similar results in terms of both thermostability ( $\Delta T_m$  +20 to +24 °C) and introduced mutations (96–98% sequence identity), as expected. Importantly, almost all our variants increase stability without decreasing folding reversibility; this is particularly remarkable for variants designed with ProteinMPNN, which in many cases even enhanced reversibility (Table S11, Figures S7 and S9). This feature is important for improving protein homeostasis *in vivo*, as irreversibly unfolded proteins are more prone to aggregation and degradation.

Overall, these results show that our approach is able to allosterically rescue the pathological single mutations by extensive substitutions in different regions of the protein.

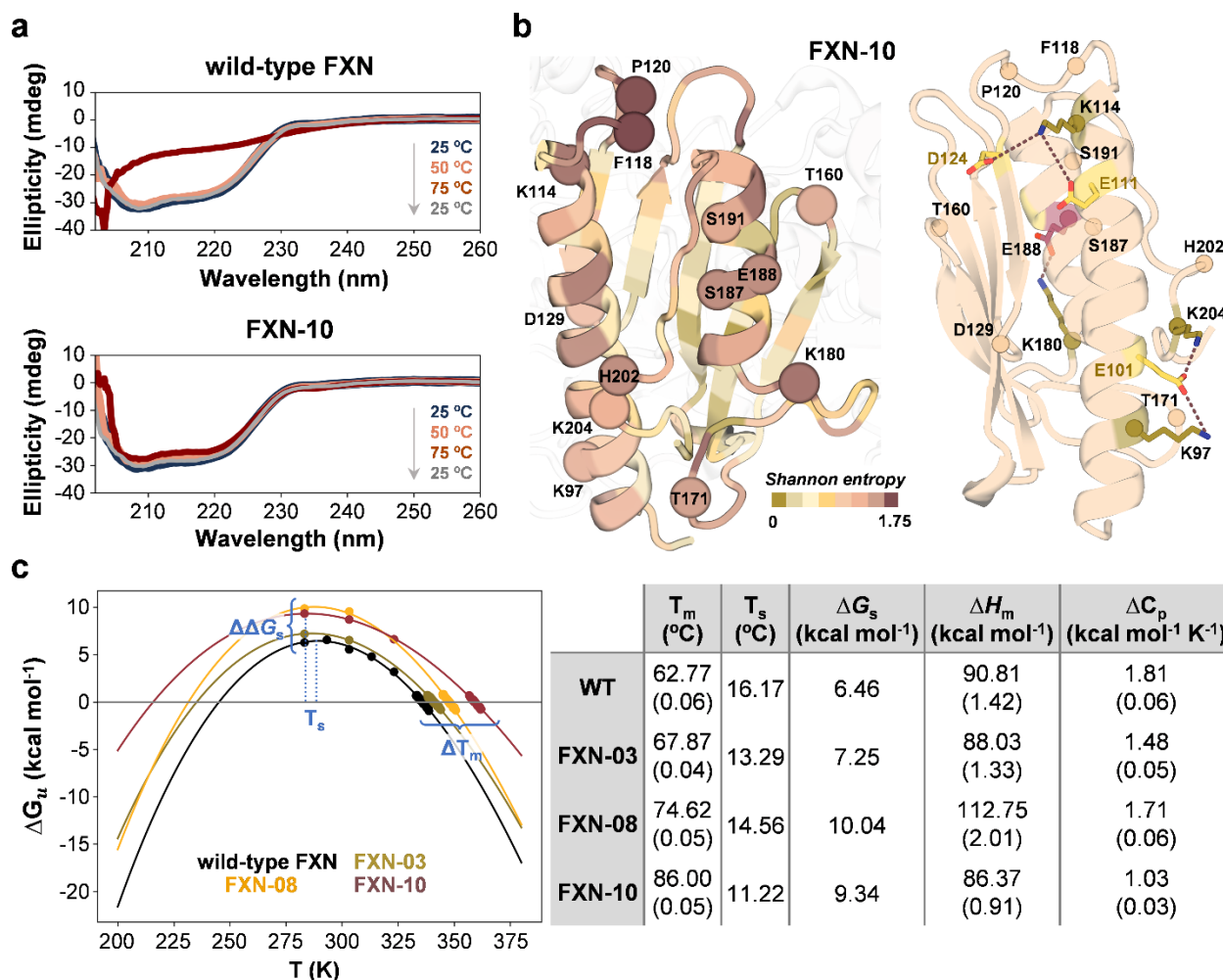
**Table 3.** Summary of the thermostabilized designs for FXN-I154F and FXN-L198R, with indication of the method used for hotspot selection, the number of mutational hotspots (N. hotspots), the method used to generate the mutations, the name of the variant, the number of mutations (N. mutations), the difference in melting temperature ( $\Delta T_m$ ), and the predicted unfolding free energy change ( $\Delta\Delta G^{calc}$ ) (REU: Rosetta energy units). All quantities are referred to the parent pathological mutant of each variant (FXN-I154F for designs FXN-11 to FXN-18 and FXN-L198R for designs FXN-19 to FXN-26). The identity of the mutations for each variant are presented in Table S10. See Supporting Information for the definition of multiple sequence alignments MSA80, MSA70 and MSA60. All measured  $T_m$  values are summarized in Table S11.

Hotspot detection	N. hotspots	Mutation approach	Variant	N. mutations	$\Delta T_m$ (°C)	$\Delta\Delta G^{calc}$ (REU)
I154F + mutations from FXN-04			FXN-11	3	9.9	9.5
I154F + mutations from FXN-03			FXN-12	1	6.7	5.7
Per-residue conservation $\leq 60\%$ on MSA60	26	ProteinMPNN	FXN-13	16	20.3	8.8
			FXN-14	15	21.1	9.0
Shannon entropy (SE > 0.65) on MSA80	20	ProteinMPNN	FXN-15	11	21.1	12.5
			FXN-16	11	20.5	13.3
I154F+ mutations from FXN-09			FXN-17	16	16.4	14.0
I154F+ mutations from FXN-10			FXN-18	13	23.4	15.2
L198R + mutations from FXN-04			FXN-19	3	9.3	2.6
L198R + mutations from FXN-03			FXN-20	1	4.1	1.7
Per-residue conservation $\leq 60\%$ on MSA60	26	ProteinMPNN	FXN-21	12	19.8	10.5
			FXN-22	14	22.9	8.1
Shannon entropy (SE > 0.65) on MSA80	20	ProteinMPNN	FXN-23	13	23.7	11.7
			FXN-24	12	23.9	9.6
L198R+ mutations from FXN-09			FXN-25	15	16.6	15.1
L198R + mutations from FXN-10			FXN-26	13	24.1	12.3

**Physical origin of improved thermostability.** Figure 3a shows the CD spectra of wild-type frataxin and superstable variant FXN-10 at different temperatures. While the former is unfolded at 75 °C, the latter maintains a secondary structure that closely matches the wild-type one even at such high temperature, showing that the mutations introduced in FXN-10 do not significantly alter the folded structure of the protein. The most stable variants share many common mutations (8 between FXN-08 and FXN-10; 6 between FXN-09 and FXN-10) including hydrophobic-to-charged (A114K, A188E/K, A204K), hydrophobic-to-polar (A187S), charged-to-polar (K171T/S), substitution to closely related amino acids (R97K, Y118F, S160T, T191S), interconversion of polar amino acids (S202H), and substitution to proline (F120P, discussed

above) (Table S10). These types of superficial mutations to either charged amino-acids or proline are commonly observed in the thermophilic counterparts of mesophilic proteins.<sup>44</sup> Accordingly, ProteinMPNN is known to favor the introduction of charged amino acids at low sampling temperatures, which may contribute to thermostabilization of the designed sequences.<sup>38</sup> In fact, the four charged residues introduced in the most stable variant FXN-10 (Lys97, Lys114, Lys180, Glu188 and Lys204) create a dense network of salt bridges not present in wild-type frataxin (Figure 3b), in line with previous observations that optimization of charge-charge interactions at the protein surface is a mechanism for stabilization.<sup>58–60</sup> Still, the theoretical isoelectric point (IP = 5.4) and total charge at pH=7.4 ( $Z = -8.1$ ) of FXN-10 are nearly identical to those of the wild-type (IP = 5.2;  $Z = -8.1$ ) (Table S10).

In an attempt to gain further insights into the thermostabilization mechanisms operating in our designs, we measured the protein stability curves as a function of temperature<sup>61</sup> of selected frataxin variants, characterizing the temperature dependence of the unfolding free energy  $\Delta G_u(T)$  of wild-type, FXN-03, FXN-08 and FXN-10 (Figure 3c and Supporting Information). The four variants show similar temperatures of maximal stability ( $T_s = 11–16$  °C) and a raised stability curve at all temperatures – as observed for the vast majority of mesophile-thermophile pairs<sup>44,62–65</sup> – with variants FXN-08 and FXN-10 achieving a thermodynamic stabilization of  $\Delta\Delta G_s \sim 3$  kcal mol<sup>-1</sup>. While the FXN-08 stability curve has essentially the same curvature as the wild-type one – implying similar heat capacities of unfolding ( $\Delta C_p$ ) – and is shifted upwards due to a larger enthalpy of unfolding at  $T_m$  ( $\Delta H_m$ ), FXN-03 and FXN-10 show comparable  $\Delta H_m$  and lower  $\Delta C_p$  values than wild-type frataxin. For the variant showing the largest increase in  $T_m$ , FXN-10,  $\Delta C_p$  drops by 0.8 kcal mol<sup>-1</sup> K<sup>-1</sup> resulting in a substantial flattening of the curve, a stabilization mechanism observed in heat adaptation.<sup>62,64,65</sup> Importantly, the fact that our variants feature higher stability at physiological temperatures supports the value of our approach to improve the homeostasis of the protein *in vivo*.



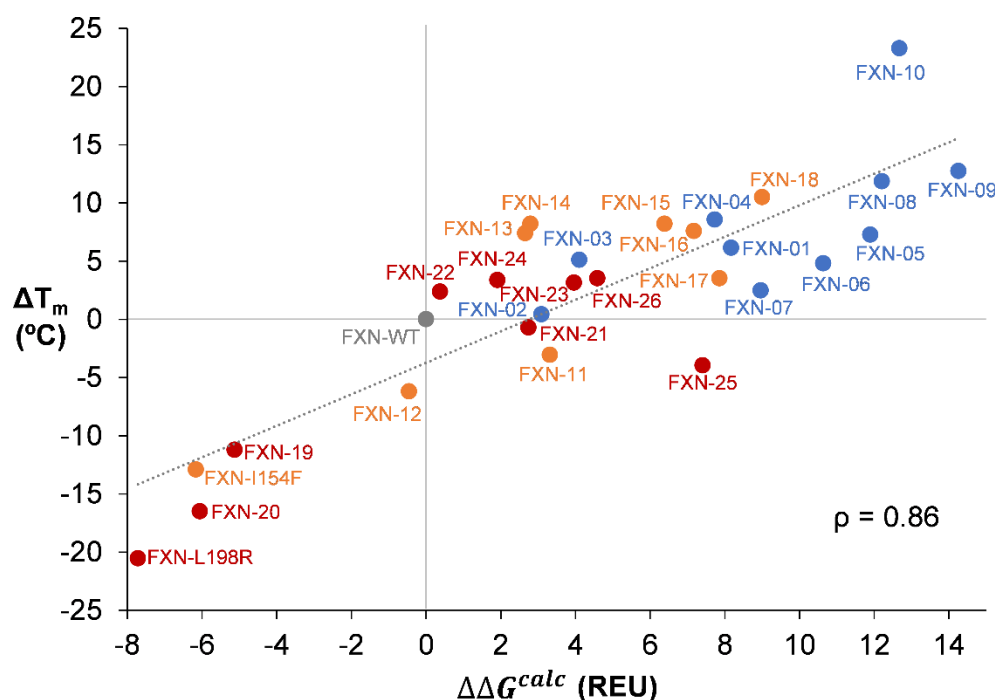
**Figure 3.** a) CD spectra at 25, 50, 75 °C (heating phase) and 25 °C (cooling phase) for wild type frataxin and design FXN-10. b) AlphaFold model for and FXN-10 colored according to Shannon entropy with indication of the mutated positions as spheres (left). Salt bridges introduced in thermostable design FXN-10 and not present in wild-type frataxin (right). The mutated charged residues engaged in new salt bridges are shown as green (K97, K,114, K180, K204) and magenta (E188) sticks. Non-mutated charged residues are shown as yellow sticks. c) Stability curves measured for wild-type frataxin, FXN-03, FXN-08 and FXN-10, together with the melting temperature ( $T_m$ ), temperature of maximum thermodynamic stability ( $T_s$ ), maximum thermodynamic stability ( $\Delta G_s$ ), enthalpy of unfolding at  $T_m$  ( $\Delta H_m$ ), and heat capacity of unfolding ( $\Delta C_p$ ) derived from these curves. Numbers in parenthesis indicate the standard errors arising from the fitting. Note that FXN-08 is the most *thermodynamically stable* variant (i.e., largest  $\Delta \Delta G_s$ ), while FXN-10 is the most *thermostable* one (i.e., largest  $\Delta T_m$ ).

**Evaluation of *in silico* prediction accuracy.** To evaluate the global accuracy and predictivity of our method, we compared the calculated ( $\Delta\Delta G^{calc}$ ) and experimentally measured changes in thermostability for the set of variants designed in this work. As described above, the curvature of the stability curve might not be the same for all variants, so the common approximation that changes in melting temperatures and unfolding free energies are linearly related<sup>61</sup> (see Supporting Information), cannot be adopted in our case; on the other hand, a rigorous determination of the thermodynamic stabilization ( $\Delta\Delta G_s$ ) of the nearly 30 assayed variants from their stability curves is beyond the scope of this work. Therefore, a direct comparison between Rosetta and experimental energies is not possible, and hence we directly correlated  $\Delta\Delta G^{calc}$  with  $\Delta T_m$  values with respect to wild-type frataxin instead. Overall, we obtained a strong correlation between computed and experimental stabilities within a  $T_m$  range spanning over 40 °C (Figure 4), achieving better predictivity than popular force-field and machine-learning-based approaches.<sup>66</sup>

Outliers from the generally good agreement are designs FXN-10 and FXN-25, whose  $\Delta T_m$  are under- and overestimated, respectively. For FXN-10, the thermodynamic parameters derived from the stability curves suggest that the underestimation may arise from the flattening of the curve due to lower  $\Delta C_p$ , which differentially increases  $\Delta T_m$  over  $\Delta\Delta G_s$ . In fact, for the thermodynamically characterized variants (Figure 3c), Rosetta energies correlate much better with the latter (Figure S11), suggesting that our approach might be best suited for predicting changes in *thermodynamic stability* (i.e.,  $\Delta\Delta G_s$ ) rather than *thermostability* ( $\Delta T_m$ ).

Interestingly, a decomposition of Rosetta energies shows that FXN-10 is also an outlier in terms of the contributions to the computed values. In particular, we detected notable differences in the relative importance of Rosetta's *ref* term in the computed energy of each variant with respect to the wild-type. This term accounts for the contribution of the unfolded state to the unfolding free energy;<sup>67</sup> since explicit calculations on the inherently complex and unknown unfolded state are prohibitively expensive and error prone, the *ref* term is structure-independent and assigns to each amino acid of the sequence a weight empirically determined to maximize native sequence recovery.<sup>67</sup> In FXN-10, this contribution is much larger than in the other variants and dominates  $\Delta\Delta G^{calc}$  (Figure S5), suggesting an important contribution of the unfolded state to the thermodynamic stabilization. Since  $\Delta C_p$  values have been shown to correlate with changes in the solvent accessible surface area between unfolded and folded states ( $\Delta SASA$ ),<sup>68</sup> we tentatively propose that the observed decrease in  $\Delta C_p$  for FXN-10 might be due to a reduced SASA in its unfolded state compared to the wild-type. This hypothesis is supported by the small SASA difference (~2%) computed for the *folded* states of the two proteins (Table S9). These results suggest that our AlphaFold/Rosetta-based protocol is not only capable of correctly predicting the sign and relative intensity of thermostability changes upon mutation, but also provides qualitative insights into the underlying biophysical thermostabilization mechanisms.

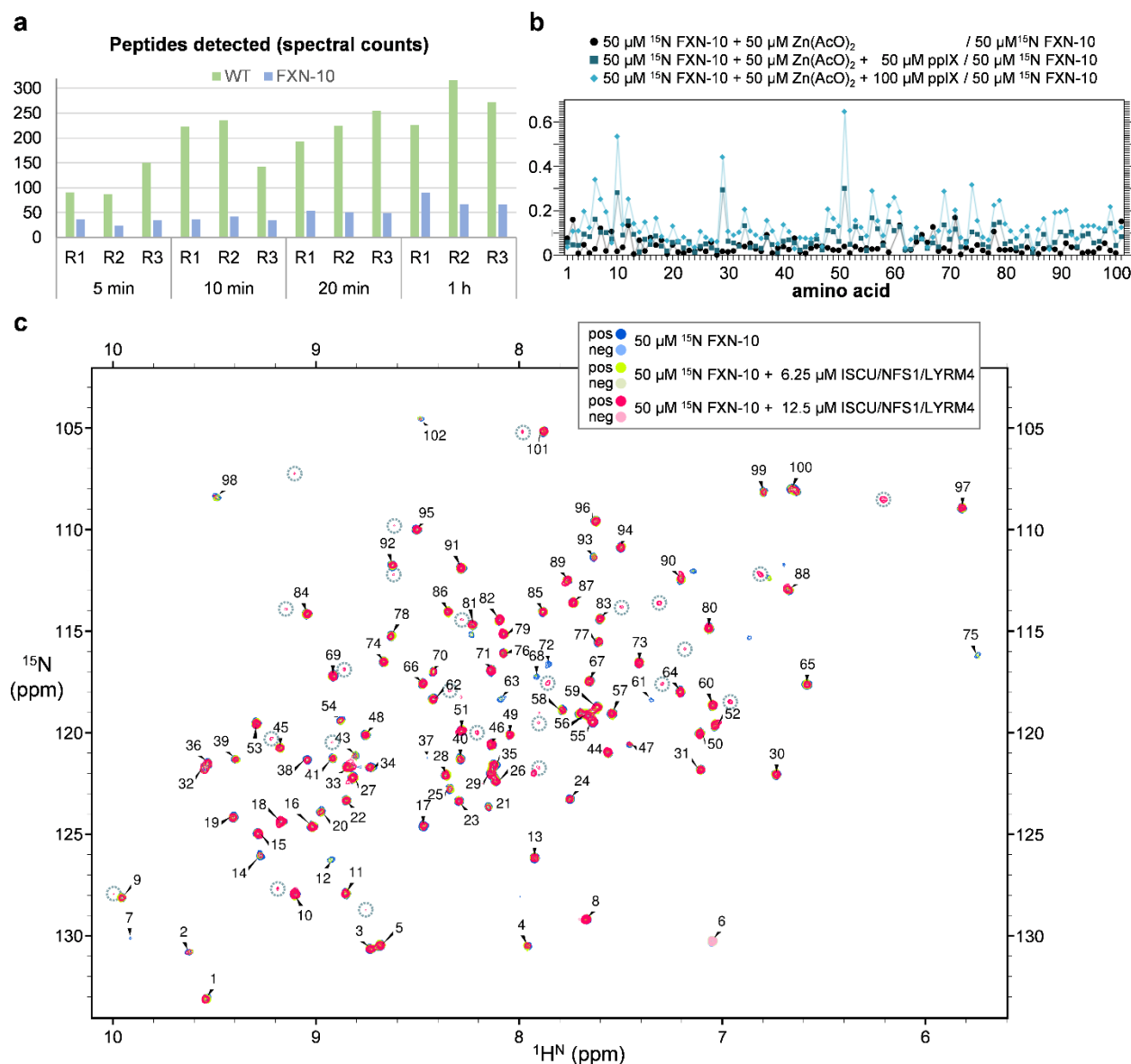




**Figure 4.** Correlation between theoretical  $\Delta\Delta G^{calc}$  and experimental  $\Delta T_m$  values for the 26 variants designed to stabilize wild-type frataxin (in blue) and the two pathological mutants I154F (in orange) and L198R (in red), using wild-type frataxin (in grey) as a reference. Positive and negative values indicate stabilization and destabilization, respectively. Each data point corresponds to a different protein variant.  $\rho$ : Pearson correlation coefficient. The dashed line indicates a linear regression with slope 1.35 and intercept -3.74.

**Stabilizing mutations confer improved resistance to proteolytic degradation.** To evaluate the extent to which our variants would improve protein homeostasis in a biological environment, we measured their propensity to enzymatic degradation. Early studies on extremely thermophilic bacteria demonstrated that enzymes derived from these organisms seemed to be more resistant to proteolysis than the corresponding enzymes from mesophilic organisms,<sup>69</sup> suggesting that this resistance may be a general feature of thermostable proteins.<sup>70</sup> In fact, resistance to proteolytic degradation is used nowadays to assess the thermodynamic folding stability of large libraries of proteins.<sup>71</sup> Accordingly, the most thermostable frataxin variant designed in this work, FXN-10 ( $\Delta T_m = 23.3$  °C), exhibited a notable resilience to proteolysis by trypsin when compared to the wild-type (Figure 5a and Supporting Information). This improved resistance to proteolytic degradation will increase the lifetime of the protein in biological environments, further supporting the use of our engineered frataxin variants as therapeutics.

**Stabilizing mutations preserve biological function.** Besides being thermally and enzymatically stable, an engineered protein must retain the biological activity of interest in order to be used as a potential therapeutic. By performing a series of NMR experiments developed in our laboratory,<sup>22</sup> we verified that superstable FXN-10 is able to bind  $\text{Zn}^{2+}$  ions (as a proxy for  $\text{Fe}^{2+}$ ) and protoporphyrin IX (ppIX) – which, when bound to metal-loaded frataxin shares the binding epitope with ferrochelatase – as well as to engage in protein-protein interactions with the iron-sulfur assembly cluster (Figure 5b and Supporting Information). Measurement of chemical shift perturbation (CSP) upon addition of  $\text{Zn}^{2+}$  and ppIX show changes upon binding consistent with those observed for wild type frataxin.<sup>22</sup> Also, interaction of frataxin with the iron-sulfur assembly complex shows the appearance of a new set of signals (Figure 5c and Supporting Information), in line with the slow exchange regime of the interaction. This is again in agreement with our prior observations for wild-type frataxin, proving the biological competency of the designed, stable variant.



**Figure 5.** a) Protein stability assay by trypsin degradation and mass spectrometry analysis shows decreased proteolysis for engineered variant FXN-10 compared to wild-type frataxin. Spectral counts refer to peptide-spectrum match (PSM), i.e., peptides identified after protein digestion. b) CSP observed for u-<sup>15</sup>N, <sup>13</sup>C-labelled FXN-10 in the presence of Zn<sup>2+</sup> and protoporphyrin IX (ppIX). c) 800 MHz 2D <sup>15</sup>N-sf-HMQC spectra of u-<sup>15</sup>N, <sup>13</sup>C-labelled FXN-10 in the presence of increasing amount of FeS assembly complex. Residue numbers are arbitrary due to lack of complete assignment of the protein.

## Conclusions

Highly stabilized human frataxin variants have been designed *in silico* by combining conservation-based mutation hotspot selection from evolutionary information with deep-learning-based sequence sampling using ProteinMPNN, and stability evaluation using AlphaFold ensembles and the Rosetta energy function. These predictions have been made in one single step, with minimal computational cost, and with a 100% success ratio (i.e., all designed variants exhibit improved thermostability compared to the wild-type, with varying levels of enhancement). The most stable design FXN-10, whose stabilization is largely associated to a reduced  $\Delta C_p$  of unfolding resembling observed heat adaptation mechanisms, is highly resistant to proteolytic degradation and retains the wild-type's ability to bind metal ions and interact with the FeS assembly complex. All these beneficial properties confer this variant a great potential for being used in protein replacement therapy for Friedreich's ataxia. We believe these studies might establish a generalizable, rational strategy for the design of proteins with improved properties for its use as therapeutics, as well as in other biotechnological applications.

## ASSOCIATED CONTENT

### Supporting Information

Detailed methods, characterization and computational data, and additional figures (PDF)

## ACKNOWLEDGMENTS

This research was supported by the Agencia Estatal Investigación of Spain (AEI; grants PID2021-125946OB-I00, PID2021-124171OB-I00, PDC2022-133725-C22, PDC2022-133677-I00, CEX2021-001136-S and a predoctoral fellowship to R.N.F.) and the Partnership for Advanced Computing in Europe (PRACE-ICEI, reference icei-prace-2021-0001). F.P. thanks the Ministerio de Economía y Competitividad for a Juan de la Cierva Incorporación (IJC2020-045506-I) research contract. G.O.Q. acknowledges support from Ikerbasque, the Basque Foundation for Science, and from the Ramón y Cajal Program (AEI; grant RYC2021-031074-I).

## REFERENCES

- (1) Ebrahimi, S. B.; Samanta, D. Engineering Protein-Based Therapeutics through Structural and Chemical Design. *Nat. Commun.* **2023**, *14*, 2411.
- (2) Przepiorka, D.; Ko, C. W.; Deisseroth, A.; Yancey, C. L.; Candau-Chacon, R.; Chiu, H. J.; Gehrke, B. J.; Gomez-Broughton, C.; Kane, R. C.; Kirshner, S.; Mehrotra, N.; Ricks, T. K.; Schmiel, D.; Song, P.; Zhao, P.; Zhou, Q.; Farrell, A. T.; Pazdur, R. FDA Approval: Blinatumomab. *Clin. Cancer Res.* **2015**, *21*, 4035–4039.
- (3) Booth, C.; Gaspar, H. B. Pegademase Bovine (PEG-ADA) for the Treatment of Infants and Children with Severe Combined Immunodeficiency (SCID). *Biol. Targets Ther.* **2009**, *3*, 349–358.

- (4) Lyseng-Williamson, K. A.; Plosker, G. L. Recombinant Factor VIIa (Eptacog Alfa): A Pharmacoeconomic Review of Its Use in Haemophilia in Patients with Inhibitors to Clotting Factors VIII or IX. *Pharmacoeconomics* **2007**, *25*, 1007–1029.
- (5) Keating, G. M. Insulin Detemir: A Review of Its Use in the Management of Diabetes Mellitus. *Drugs* **2012**, *72*, 2255–2287.
- (6) Marziniak, M.; Meuth, S. Current Perspectives on Interferon Beta-1b for the Treatment of Multiple Sclerosis. *Adv. Ther.* **2014**, *31*, 915–931.
- (7) Leader, B.; Baca, Q. J.; Golan, D. E. Protein Therapeutics: A Summary and Pharmacological Classification. *Nat. Rev. Drug Discov.* **2008**, *7*, 21–39.
- (8) Winklhofer, K. F.; Tatzelt, J.; Haass, C. The Two Faces of Protein Misfolding: Gain- and Loss-of-Function in Neurodegenerative Diseases. *EMBO J.* **2008**, *27*, 336–349.
- (9) Blouin, J.-M.; Duchartre, Y.; Costet, P.; Lalanne, M.; Ged, C.; Lain, A.; Millet, O.; de Verneuil, H.; Richard, E. Therapeutic Potential of Proteasome Inhibitors in Congenital Erythropoietic Porphyria. *Proc. Natl. Acad. Sci.* **2013**, *110*, 18238–18243.
- (10) Balch, W. E.; Morimoto, R. I.; Dillin, A.; Kelly, J. W. Adapting Proteostasis for Disease Intervention. *Science* **2008**, *319*, 916–919.
- (11) Kontos, S.; Hubbell, J. A. Drug Development: Longer-Lived Proteins. *Chem. Soc. Rev.* **2012**, *41*, 2686–2695.
- (12) Wang, F.; Carabino, J. M.; Vergara, C. M. Insulin Glargine: A Systematic Review of a Long-Acting Insulin Analogue. *Clin. Ther.* **2003**, *25*, 1541–1577.
- (13) Becker, R. H. A.; Frick, A. D. Clinical Pharmacokinetics and Pharmacodynamics of Insulin Glulisine. *Clin. Pharmacokinet.* **2008**, *47*, 7–20.
- (14) Murphy, R. M.; Tsai, A. M. *Misbehaving Proteins: Protein (Mis)Folding, Aggregation, and Stability*; Springer New York: New York, NY, 2006.
- (15) ben Bdira, F.; González, E.; Pluta, P.; Laín, A.; Sanz-Parra, A.; Falcon-Perez, J. M.; Millet, O. Tuning Intracellular Homeostasis of Human Uroporphyrinogen III Synthase by Enzyme Engineering at a Single Hotspot of Congenital Erythropoietic Porphyria. *Hum. Mol. Genet.* **2014**, *23*, 5805–5813.
- (16) Chen, W.; Kong, L.; Connelly, S.; Dendle, J. M.; Liu, Y.; Wilson, I. A.; Powers, E. T.; Kelly, J. W. Stabilizing the CH2 Domain of an Antibody by Engineering in an Enhanced Aromatic Sequon. *ACS Chem. Biol.* **2016**, *11*, 1852–1861.
- (17) Zhang, S.; Napierala, M.; Napierala, J. S. Therapeutic Prospects for Friedreich’s Ataxia. *Trends Pharmacol. Sci.* **2019**, *40*, 229–233.
- (18) Condò, I.; Ventura, N.; Malisan, F.; Rufini, A.; Tomassini, B.; Testi, R. In Vivo Maturation of Human Frataxin. *Hum. Mol. Genet.* **2007**, *16*, 1534–1540.
- (19) Dhe-Paganon, S.; Shigeta, R.; Chi, Y.-I.; Ristow, M.; Shoelson, S. E. Crystal Structure of Human Frataxin \*. *J. Biol. Chem.* **2000**, *275*, 30753–30756.
- (20) Cavadini, P.; O’Neill, H. A.; Benada, O.; Isaya, G. Assembly and Iron-Binding Properties of Human Frataxin, the Protein Deficient in Friedreich Ataxia. *Hum. Mol. Genet.* **2002**, *11*, 217–227.



- (21) Faraj, S. E.; González-Lebrero, R. M.; Roman, E. A.; Santos, J. Human Frataxin Folds Via an Intermediate State. Role of the C-Terminal Region. *Sci. Rep.* **2016**, 6, 20782.
- (22) Bernardo-Seisdedos, G.; Schedlbauer, A.; Pereira-Ortuzar, T.; Mato, J. M.; Millet, O. Protoporphyrin IX Binds to Iron(II)-Loaded and to Zinc-Loaded Human Frataxin. *Life* **2023**, 13, 222.
- (23) Clark, E.; Butler, J. S.; Isaacs, C. J.; Napierala, M.; Lynch, D. R. Selected Missense Mutations Impair Frataxin Processing in Friedreich Ataxia. *Ann. Clin. Transl. Neurol.* **2017**, 4, 575–584.
- (24) Li, H.; Gakh, O.; Smith, D. Y.; Ranatunga, W. K.; Isaya, G. Missense Mutations Linked to Friedreich Ataxia Have Different but Synergistic Effects on Mitochondrial Frataxin Isoforms. *J. Biol. Chem.* **2013**, 288, 4116–4127.
- (25) Piguet, F.; de Montigny, C.; Vaucamps, N.; Reutenauer, L.; Eisenmann, A.; Puccio, H. Rapid and Complete Reversal of Sensory Ataxia by Gene Therapy in a Novel Model of Friedreich Ataxia. *Mol. Ther.* **2018**, 26, 1940–1952.
- (26) Perdomini, M.; Belbellaa, B.; Monassier, L.; Reutenauer, L.; Messaddeq, N.; Cartier, N.; Crystal, R. G.; Aubourg, P.; Puccio, H. Prevention and Reversal of Severe Mitochondrial Cardiomyopathy by Gene Therapy in a Mouse Model of Friedreich's Ataxia. *Nat. Med.* **2014**, 20, 542–547.
- (27) Lynch, D. R.; Kichula, E.; Lin, H. Frataxin Restoration in the Nervous System: Possibilities for Gene Therapy. *Mol. Ther.* **2018**, 26, 1880–1882.
- (28) Gonzalez, E.; Wagner, E.; Mess, N.; Wang, A.; Payne, M.; Bettoun, D.; Ottinger, E.; Rup, B.; Xu, X. P55 - Characterization of in Vivo Disposition of CTI-1601: A Mitochondria Targeted Therapy for Friedreich's Ataxia. *Drug Metab. Pharmacokinet.* **2020**, 35, S38.
- (29) Payne, R. M. Cardiovascular Research in Friedreich Ataxia: Unmet Needs and Opportunities. *JACC Basic to Transl. Sci.* **2022**, 7, 1267–1283.
- (30) Fox, N. G.; Yu, X.; Feng, X.; Bailey, H. J.; Martelli, A.; Nabhan, J. F.; Strain-Damerell, C.; Bulawa, C.; Yue, W. W.; Han, S. Structure of the Human Frataxin-Bound Iron-Sulfur Cluster Assembly Complex Provides Insight into Its Activation Mechanism. *Nat. Commun.* **2019**, 10, 2210.
- (31) Huang, J.; Dizin, E.; Cowan, J. A. Mapping Iron Binding Sites on Human Frataxin: Implications for Cluster Assembly on the ISU Fe-S Cluster Scaffold Protein. *J. Biol. Inorg. Chem.* **2008**, 13, 825–836.
- (32) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, 596, 583–589.
- (33) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C.

- R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876.
- (34) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379*, 1123–1130.
- (35) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Zidek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50*, D439–D444.
- (36) Wang, J.; Lisanza, S.; Juergens, D.; Tischer, D.; Watson, J. L.; Castro, K. M.; Ragotte, R.; Saragovi, A.; Milles, L. F.; Baek, M.; Anishchenko, I.; Yang, W.; Hicks, D. R.; Expòsit, M.; Schlichthaerle, T.; Chun, J. H.; Dauparas, J.; Bennett, N.; Wicky, B. I. M.; Muenks, A.; DiMaio, F.; Correia, B.; Ovchinnikov, S.; Baker, D. Scaffolding Protein Functional Sites Using Deep Learning. *Science* **2022**, *377*, 387–394.
- (37) Wicky, B. I. M.; Milles, L. F.; Courbet, A.; Ragotte, R. J.; Dauparas, J.; Kinfu, E.; Tipps, S.; Kibler, R. D.; Baek, M.; DiMaio, F.; Li, X.; Carter, L.; Kang, A.; Nguyen, H.; Bera, A. K.; Baker, D. Hallucinating Symmetric Protein Assemblies. *Science* **2022**, *378*, 56–61.
- (38) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN. *Science* **2022**, *378*, 49–56.
- (39) Peccati, F.; Alunno-Rufini, S.; Jiménez-Osés, G. Accurate Prediction of Enzyme Thermostabilization with Rosetta Using AlphaFold Ensembles. *J. Chem. Inf. Model.* **2023**, *63*, 898–909.
- (40) Modarres, H. P.; Mofrad, M. R.; Sanati-Nezhad, A. Protein Thermostability Engineering. *RSC Adv.* **2016**, *6*, 115252–115270.
- (41) Sternke, M.; Tripp, K. W.; Barrick, D. Consensus Sequence Design as a General Strategy to Create Hyperstable, Biologically Active Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *166*, 11275–11284.
- (42) Lehmann, M.; Loch, C.; Middendorf, A.; Studer, D.; Lassen, S. F.; Pasamontes, L.; Van Loon, A. P. G. M.; Wyss, M. The Consensus Concept for Thermostability Engineering of Proteins: Further Proof of Concept. *Protein Eng.* **2002**, *15*, 403–411.
- (43) Baxa, M. C.; Haddadian, E. J.; Jha, A. K.; Freed, K. F.; Sosnick, T. R. Context and Force Field Dependence of the Loss of Protein Backbone Entropy upon Folding Using Realistic Denatured and Native State Ensembles. *J. Am. Chem. Soc.* **2012**, *134*, 15929–15936.

- (44) Kumar, S.; Nussinov, R. How Do Thermophilic Proteins Deal with Heat? *Cell. Mol. Life Sci.* **2001**, *58*, 1216–1233.
- (45) Tsai, C. L.; Bridwell-Rabb, J.; Barondeau, D. P. Friedreich's Ataxia Variants I154F and W155R Diminish Frataxin-Based Activation of the Iron-Sulfur Cluster Assembly Complex. *Biochemistry* **2011**, *50*, 6478–6487.
- (46) Bridwell-Rabb, J.; Winn, A. M.; Barondeau, D. P. Structure - Function Analysis of Friedreich's Ataxia Mutants Reveals Determinants of Frataxin Binding and Activation of the Fe - S Assembly Complex. *Biochemistry* **2011**, *50*, 7265–7274.
- (47) Musco, G.; Stier, G.; Kolmerer, B.; Adinolfi, S.; Martin, S.; Frenkiel, T.; Gibson, T.; Pastore, A. Towards a Structural Understanding of Friedreich's Ataxia: Solution Structure of Frataxin. *Structure* **2000**, *8*, 695–707.
- (48) Stein, R. A.; McHaourab, H. S. SPEACH\_AF: Sampling Protein Ensembles and Conformational Heterogeneity with AlphaFold2. *PLoS Comput. Biol.* **2022**, *18*, e1010483.
- (49) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
- (50) Ahmad, S.; Kamal, M. Z.; Sankaranarayanan, R.; Rao, N. M. Thermostable Bacillus Subtilis Lipases: In Vitro Evolution and Structural Insight. *J. Mol. Biol.* **2008**, *381*, 324–340.
- (51) García-Marquina, G.; Núñez-Franco, R.; Peccati, F.; Tang, Y.; Jiménez-Osés, G.; López-Gallego, F. Deconvoluting the Directed Evolution Pathway of Engineered Acyltransferase LovD. *ChemCatChem* **2022**, *14*.
- (52) Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **2016**, *63*, 337–346.
- (53) Barber-Zucker, S.; Mindel, V.; Garcia-Ruiz, E.; Weinstein, J. J.; Alcalde, M.; Fleishman, S. J. Stable and Functionally Diverse Versatile Peroxidases Designed Directly from Sequences. *J. Am. Chem. Soc.* **2022**, *144*, 3564–3571.
- (54) Weinstein, J. J.; Goldenzweig, A.; Hoch, S. Y.; Fleishman, S. J. PROSS 2: A New Server for the Design of Stable and Highly Expressed Protein Variants. *Bioinformatics* **2021**, *37*, 123–125.
- (55) Castro, I. H.; Pignataro, M. F.; Sewell, K. E.; Espeche, L. D.; Herrera, M. G.; Noguera, M. E.; Dain, L.; Nadra, A. D.; Aran, M.; Smal, C.; Gallo, M.; Santos, J. Frataxin Structure and Function. In *Subcellular Biochemistry*; Harris, J. R., Marles-Wright, J., Eds.; Subcellular Biochemistry; Springer International Publishing: Cham, 2019; Vol. 93, pp 393–438.
- (56) Correia, A. R.; Pastore, C.; Adinolfi, S.; Pastore, A.; Gomes, C. M. Dynamics, Stability and Iron-Binding Activity of Frataxin Clinical Mutants. *FEBS J.* **2008**, *275*, 3680–3690.
- (57) Faraj, S. E.; Roman, E. A.; Aran, M.; Gallo, M.; Santos, J. The Alteration of the C-Terminal Region of Human Frataxin Distorts Its Structural Dynamics and Function. *FEBS J.* **2014**, *281*, 3397–3419.
- (58) Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.;

- Loladze, V. V; Makhatadze, G. I. Protein Stability and Surface Electrostatics: A Charged Relationship. *Biochemistry* **2006**, *45*, 2761–2766.
- (59) Sanchez-Ruiz, J. M.; Makhatadze, G. I. To Charge or Not to Charge? *Trends Biotechnol.* **2001**, *19*, 132–135.
- (60) Makhatadze, G. I.; Loladze, V. V; Ermolenko, D. N.; Chen, X.; Thomas, S. T. Contribution of Surface Salt Bridges to Protein Stability: Guidelines for Protein Engineering. *J. Mol. Biol.* **2003**, *327*, 1135–1148.
- (61) Bechtel, W. J.; Schellman, J. A. Protein Stability Curves. *Biopolymers* **1987**, *26*, 1859–1877.
- (62) Hollien, J.; Marqusee, S. A Thermodynamic Comparison of Mesophilic and Thermophilic Ribonucleases H. *Biochemistry* **1999**, *38*, 3831–3836.
- (63) Fu, H.; Grimsley, G.; Scholtz, J. M.; Pace, C. N. Increasing Protein Stability: Importance of DeltaC(p) and the Denatured State. *Protein Sci.* **2010**, *19*, 1044–1052.
- (64) Luke, K. A.; Higgins, C. L.; Wittung-Stafshede, P. Thermodynamic Stability and Folding of Proteins from Hyperthermophilic Organisms. *FEBS J.* **2007**, *274*, 4023–4033.
- (65) Razvi, A.; Scholtz, J. M. Lessons in Stability from Thermophilic Proteins. *Protein Sci.* **2006**, *15*, 1569–1578.
- (66) Huang, P.; Chu, S. K. S.; Frizzo, H. N.; Connolly, M. P.; Caster, R. W.; Siegel, J. B. Evaluating Protein Engineering Thermostability Prediction Tools Using an Independently Generated Dataset. *ACS Omega* **2020**, *5*, 6487–6493.
- (67) Alford, R. F.; Leaver-Fay, A.; Jeliaskov, J. R.; O’Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048.
- (68) Myers, J. K.; Nick Pace, C.; Martin Scholtz, J. Denaturant m Values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. *Protein Sci.* **1995**, *4*, 2138–2148.
- (69) Daniel, R. M.; Cowan, D. A.; Morgan, H. W.; Curran, M. P. A Correlation between Protein Thermostability and Resistance to Proteolysis. *Biochem. J.* **1982**, *207*, 641–644.
- (70) Parsell, D. A.; Sauer, R. T. The Structural Stability of a Protein Is an Important Determinant of Its Proteolytic Susceptibility in Escherichia Coli. *J. Biol. Chem.* **1989**, *264*, 7590–7595.
- (71) Tsuboyama, K.; Dauparas, J.; Chen, J.; Laine, E.; Mohseni Behbahani, Y.; Weinstein, J. J.; Mangan, N. M.; Ovchinnikov, S.; Rocklin, G. J. Mega-Scale Experimental Analysis of Protein Folding Stability in Biology and Design. *Nature* **2023**, *620*, 434–444.