bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Main Manuscript for

The stability landscape of *de novo* TIM barrels explored by a modular design approach

Sergio Romero-Romero,[1,2] Miguel Costas,[3] Daniel-Adriano Silva Manzano,[4,5]† Sina Kordes,[2] Erendira Rojas-Ortega,[1] Cinthya Tapia,[1] Yasel Guerra,[1,]‡ Sooruban Shanmugaratnam,[2] Adela Rodríguez-Romero,[6] David Baker,[4,5,]* Birte Höcker,[2,]* D. Alejandro Fernández-Velasco.[1,]*

[1] Laboratorio de Fisicoquímica e Ingeniería de Proteínas, Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico.

[2] Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany.

[3] Laboratorio de Biofisicoquímica, Departamento de Fisicoquímica, Facultad de Química, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico.

[4] Department of Biochemistry, University of Washington, 98195 Seattle, USA.

[5] Institute for Protein Design, University of Washington, 98195 Seattle, USA.

[6] Instituto de Química, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico.

† Current address: Centro de Investigación para la Salud en América Latina, Facultad de Ciencias Exactas y Naturales, Pontificia Universidad Católica del Ecuador, 17012184 Quito, Ecuador.

‡ Current address: Neoleukin Therapeutics, 98102 Seattle, USA.

* Corresponding authors.

Correspondence

* D. Alejandro Fernández-Velasco. Laboratorio de Fisicoquímica e Ingeniería de Proteínas, Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico. Phone number: +525556232259. **Email:** fdaniel@unam.mx

* Birte Höcker. Department of Biochemistry, University of Bayreuth, Bayreuth 95447, Germany. Phone number: +490921557845. **Email:** birte.hoecker@uni-bayreuth.de

* David Baker. Institute for Protein Design, University of Washington, Seattle, WA 98195. Phone number: +12065431295. **Email:** dabaker@uw.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ORCID identifiers
Sergio Romero-Romero: 0000-0003-2144-7912
Miguel Costas: 0000-0001-8604-6381
Daniel-Adriano Silva Manzano: 0000-0002-3195-9009
Sina Kordes: 0000-0001-7489-3380
Erendira Rojas-Ortega: 0000-0003-4061-9580
Cinthya Tapia: 0000-0002-7256-6885
Yasel Guerra: 0000-0002-1419-9589
Sooruban Shanmugaratnam: 0000-0002-2614-6046
Adela Rodríguez-Romero: 0000-0001-7641-6545
David Baker: 0000-0001-7896-6217
Birte Höcker: 0000-0002-8250-9462
D. Alejandro Fernández-Velasco: 0000-0003-2368-6378

This PDF file includes:

- Main Text.
- Figures 1 to 5.
- Table 1.

One Sentence Summary

A family of designed TIM barrels with diverse thermodynamic properties shows epistatic effects on its stability landscape.

Abstract

The ability to design stable proteins with custom-made functions is a major goal in biochemistry with practical relevance for our environment and society. High conformational stability lowers protein sensitivity to mutations and changes in the environment; thus, understanding and manipulating protein stability will expand the applications of *de novo* proteins. Since the $(\beta/\alpha)_8$-barrel or TIM-barrel fold is one of the most common functional scaffolds, in this work we designed a collection of stable *de novo* TIM barrels (DeNovoTIMs), using a computational fixed-backbone and modular approach based on improved hydrophobic packing of sTIM11, the first validated *de novo* TIM barrel. DeNovoTIMs navigate a region of the stability landscape previously uncharted by natural TIM barrels, with variations spanning 60 degrees in melting temperature and 22 kcal per mol in conformational stability throughout the designs. Significant non-additive or epistatic effects were observed when stabilizing mutations from different regions of the barrel were combined. The molecular basis of epistasis in DeNovoTIMs appears to be related to the extension of the hydrophobic cores. This study is an important step towards the fine-tuned modulation of protein stability by design.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Significance Statement

De novo protein design expands our knowledge about protein structure and stability. The TIM barrel is a highly relevant fold used in nature to host a rich variety of catalytic functions. Here, we follow a modular approach to design and characterize a collection of *de novo* TIM barrels and subjected them to a thorough folding analysis. Non-additive effects modulate the increase in stability when different regions of the barrel are mutated, showing a wide variety of thermodynamic properties that allow them to navigate an unexplored region of the stability landscape found in natural TIM barrels. The design of stable proteins increases the applicability of *de novo* proteins and provides crucial information on the molecular determinants that modulate structure and stability.

Keywords

de novo protein design, protein folding and stability, stability landscape, non-additive effects, TIM barrel, $(\beta/\alpha)_8$ barrel.

Main Text

Introduction

Proteins are essential macromolecules capable of performing diverse and exquisite biological functions such as nutrient uptake, environmental stimuli sensing, immune protection, energy storage, cellular communication, molecule transportation, or enzymatic reactions. To guarantee such activities, the functional states must act under specific environmental conditions in a relevant time scale, that is, proteins must be "*stable*". Protein stability is required to maintain functional structures and it enhances the ability of proteins to evolve new properties (*1, 2*). The central role of proteins in the chemistry of life, as well as their increasing application in basic and applied research, implies that the understanding and manipulation of protein stability are both practically and academically relevant.

There are two main indicators of protein conformational stability at equilibrium. One is the difference of free energy between the native and unfolded states at a given temperature (ΔG), which is often obtained by chemical unfolding experiments carried out at 25 °C. In addition, stability is also assessed in the context of thermal unfolding, where the unfolding temperature (T_m), the temperature at the midpoint of the transition from native to the unfolded state, is the most common parameter employed to quantify stability. Both the ΔG and T_m parameters, usually determined as criteria for a "*stable*" protein, are related with the enthalpy (ΔH) and heat capacity (ΔC_P) changes through the Gibbs-Helmholtz equation, which describes the variation of ΔG with temperature, the so-called "*stability curve*" of proteins (*3*). Different mechanisms have been proposed to modify the stability curve of proteins (*4*), and numerous studies on natural proteins and their site-directed mutants have been used to rationalize the stability of thermophilic proteins and moreover to engineer thermostability (*5*).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Historically, the design of stable proteins has been one of the main objectives of computational protein design (*6*). Several strategies, such as increasing the hydrophobic area in internal cores, improvement of water-protein interactions, the introduction of disulfide bridges as well as the addition of salt bridges, have been proposed (*7-18*). The design of *de novo* proteins can further enhance our understanding of the physicochemical properties that modulate stability. For example, although folding behavior has been only addressed for very few cases, the kinetic analysis of the folding mechanism of two *de novo* βα proteins has revealed complex free energy surfaces (*19, 20*). The fine-tuning of conformational stability, that is, the manipulation of the protein stability curve, is an open challenge for protein design and engineering. Such a goal requires a comprehensive characterization of *de novo* proteins, describing the combination of thermodynamic parameters that can be reached in a particular fold.

Within the different topologies that a protein can adopt, the TIM-barrel or $(\beta/\alpha)_8$-barrel fold is one of the most abundant superfolds in nature (*21*). Furthermore, proteomic analysis shows that the TIM-barrel domain is within the mean size of the proteins present in *Escherichia coli* (*22*). Besides, the TIM-barrel fold is one of the most successful topologies used in nature to host catalytic activities. Due to its large variety of functions and its ubiquity in different types of enzymes, the TIM barrel represents a very suitable scaffold for protein function design and engineering (*23*). For these reasons, its construction has been an important objective for protein design over the years (*24-28*). Recently, the successful design of a *de novo* four-fold symmetric TIM barrel was described: the sTIM11 protein (*29*). Considering that sTIM11 presents a sequence distant from the naturally occurring TIM-barrel superfamilies, the potential of the TIM-barrel fold to bear functions is more significant than we know so far. sTIM11 shows a high melting temperature (T_m= 80 °C) but low conformational stability ($\Delta G_{25°C}$= ~4 kcal mol^{-1}) when compared to natural TIM barrels (*29-32*). Since low conformational stability often results in high sensitivity to mutations and changes in the environment, this can limit the design of novel proteins with new functions (*8*). Thus, fine-tuning the stability of the sTIM11 scaffold is a prerequisite to functionalize and generate tailor-made barrels for applications in biochemistry, biotechnology, and medicine. In this work, a fixed-backbone design with a modular approach was used to generate a collection of *de novo* TIM barrels. Their stability landscape and structural properties were characterized in detail increasing our knowledge on how stability can be fine-tuned by design.

Results and Discussion

Modular repacking of the TIM-barrel hydrophobic cores

The *de novo* protein sTIM11 is an idealized four-fold symmetric TIM barrel of 184 residues, which was designed to include two cysteines that, however, did not form the intended disulfide bond (Fig. 1). To avoid reactive free thiols, both residues were reverted to the residues in the original four-fold design (C8Q and C181V), resulting in sTIM11noCys. The base design DeNovoTIM0, which is the starting point for all further constructs in this work, additionally contains the changes W34V and A38G in all symmetry-related quarters. These residues are situated in every second α/β-loop, and in sTIM11, these tryptophan residues are the most highly solvent exposed. While different strategies have been explored to increase protein

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

stability (*8, 18*), here we focused on hydrophobic repacking. The structural analysis suggested two regions to be amenable to improvements in sTIM11, namely the central and the peripheral hydrophobic cores. The interior of the circular sheet forms the central core, whereas the outer face of the strands and the internal face of the helices constitute the peripheral core. In this latter, we identified two regions with internal cavities that are located in the lower and upper parts of the barrel, respectively (Fig. 1). The residues lining the three aforementioned regions were subjected to fixed-backbone Rosetta design according to the flow diagram depicted in Fig. S1.

Ten designs were selected for characterization in the first round: four with modifications in the internal core (DeNovoTIM1-4) as well as three designs each for the bottom peripheral or outer core (DeNovoTIM5-7) and the top peripheral core (DeNovoTIM8-10) (Fig. S2). For the inner core, no improved designs could be identified when four-fold symmetry was preserved. Therefore, in DeNovoTIM1-4, as well as all of its descendants, only a two-fold symmetry was enforced. An exploratory characterization by circular dichroism (CD) and differential scanning calorimetry (DSC) of these proteins (DeNovoTIMs 1-10) as well as DeNovoTIM0 showed that DeNovoTIM1, DeNovoTIM6, and DeNovoTIM8 were the best designs for each region (Fig. S3 and supporting text).

To test for additivity effects on stability and structure, mutations contained in the best design of each group were combined to generate the following double-region designs: DeNovoTIM11 (DeNovoTIM1 + DeNovoTIM6), DeNovoTIM12 (DeNovoTIM1 + DeNovoTIM8), and DeNovoTIM13 (DeNovoTIM6 + DeNovoTIM8). Finally, in the third design round the mutations of all three regions were combined resulting in DeNovoTIM14 (DeNovoTIM1 + DeNovoTIM6 + DeNovoTIM8). All these proteins as well as sTIM11, sTIM11noCys, and DeNovoTIM0 were characterized in detail (Fig. 1). Information on sequences, and mutations in each design are reported in the supporting information (Fig. S2 and tables S1-S4).

Physicochemical characterization of DeNovoTIMs

sTIM11, sTIM11noCys, and all DeNovoTIM variants presented the characteristic far-UV CD spectra observed for α/β proteins (Fig. 2A and Fig. S4). The near-UV CD and intrinsic fluorescence (IF) spectra showed that the aromatic residues are buried from the solvent and structured in the folded state (Fig. 2B-2C and Fig. S5-S6; see supporting text for details). All DeNovoTIMs adopt a monomeric and compact shape as revealed by the invariant value of the Stokes radius determined by analytical size exclusion chromatography over a twenty-fold protein concentration range (table S5). DeNovoTIM0 is also monomeric but shows a Stokes radius (26.1 ± 0.3 Å) slightly higher than that expected for a compact protein of this size, but still far away from the expected value for an unfolded conformation (22.5 ± 1.0 Å and 42.0 ± 1.0 Å, respectively; *33*). This is in agreement with the red shift in the IF spectra and suggests a slightly expanded conformation for DeNovoTIM0.

Thermal unfolding was then studied by CD and DSC (Fig. 2D-2E). All DeNovoTIMs showed cooperative transitions with a remarkably broad range of T_m values, from 47 °C (DeNovoTIM0) to 109 °C (DeNovoTIM12) (table 1); indeed at 90 °C many of the proteins still showed secondary and tertiary structure (Fig. S4B and Fig. S5B). All DeNovoTIMs, except

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DeNovoTIM13 and DeNovoTIM14, showed thermal unfolding reversibility (Fig. S7) and were well fitted to the two-state model (N⇋U) (Fig. S8 and table 1). This is remarkable because the temperature-induced unfolding of natural proteins of this size, particularly TIM barrels, is usually not reversible (*22, 30*). DeNovoTIM14 showed two endotherms, suggesting the presence of an unfolding intermediate (Fig. S8I), For DeNovoTIM13 and DeNovoTIM14, endotherms were well-fitted to an irreversible two-state mechanism (N→F) giving activation energies (E_{act}) of 120.3 and 37.4 kcal mol^{-1} (table 1), respectively, resulting in very different kinetic stabilities (Fig. S9-S10 and supporting text).

For DeNovoTIMs with a reversible thermal unfolding, the observed unfolding ΔH and ΔC_P also vary greatly (table 1); for some DeNovoTIMs these values are similar to the ones expected for a protein of 184 residues, whereas for others they are smaller (ΔH= 128.4 ± 3.5 kcal mol^{-1} and ΔC_P= 2.6 ± 0.04 kcal mol^{-1} K^{-1}, according to parametric equations reported in *34*). The ΔH values observed for the first and second design rounds (0.24 to 0.64 kcal mol^{-1} residue^{-1}) are similar to those reported for natural monomeric TIM barrels (0.25 to 0.67 kcal mol^{-1} residue^{-1}).

In order to obtain reliable ΔC_P values, multiple DSC experiments were performed at ten different protein concentrations (Fig. S8 and table 1), obtaining ΔC_P values which were independent of protein concentration and varied in a small standard deviation range (Fig. S11). A decrease in ΔC_P has been shown to result from residual structure in the unfolded state (*35*). This is observed in the far-UV CD spectra of those DeNovoTIMs that are unfolded at 90 °C. In addition, the low ΔH of DeNovoTIM14 increases in the presence of urea (Fig. 2F and Fig. S10). These results suggest that for some DeNovoTIMs, the reason for the low ΔH and ΔC_P is likely the high content of residual structure in the unfolded state (supporting text).

Stability at 25 °C was studied by chemical unfolding with urea or GdnHCl. Except for DeNovoTIM14, all designs were completely unfolded in 9.0 M urea (Fig. S4-S6). Unfolding and refolding transitions are coincident and the signal does not change after incubation for 12 hours, i.e. chemical unfolding is reversible and in equilibrium under the experimental conditions. For all DeNovoTIMs, except for DeNovoTIM14, CD and IF curves were monophasic, cooperative, coincident, and well globally-fitted to a two-state N⇋U model, indicating the absence of populated intermediates (Fig. 2G and Fig. S12-S13). DeNovoTIM14 presented a different behavior; no changes in CD or IF signal were observed in the presence of urea (Fig. 2G), even after incubation for 5 days. CD and IF spectra indicate that at 9.0 M urea DeNovoTIM14 presents native-like properties (Fig. S4-S6).

When chemical unfolding was carried out with GdnHCl, unfolding transitions were reversible and at equilibrium with this denaturant. IF data showed a monophasic transition in the 3-5 M GdnHCl range, while CD detected the presence of an unfolding intermediate between 3-4 M GdnHCl (Fig. 2H). Both traces were globally fitted to a three-state model with an intermediate: N⇋I⇋U (Fig. S13). All the selected first- and second-round designs presented a ΔG at 25 °C higher than DeNovoTIM0, whereas the triple-design, DeNovoTIM14, showed a pronounced increase in stability (ΔG_{Tot}= 23.6 kcal mol^{-1}; table 1). For DeNovoTIM14, the stability change related to the loss of the native state ($\Delta G_{N\text{-}I}$= 12.7 kcal mol^{-1}) is higher than the ΔG of second-round designs, whereas the stability of the intermediate is similar to them ($\Delta G_{I\text{-}U}$= 10.9 kcal mol^{-1}). For three-state folders, the change in free energy from the native to the intermediate state ($\Delta G_{N\text{-}I}$) has been termed “the relevant stability” because the intermediate is

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expected to be non-functional, whereas the stability change from the intermediate to the unfolded state ($\Delta G_{I\text{-}U}$) is designated as "residual stability" (*36*).

m values ($m = \partial\Delta G / \partial$[denaturant]) are proportional to the surface area exposed to the solvent upon unfolding (ΔASA); likewise, the buried area correlates very strongly with the number of residues (*37*). The *m* value calculated from the sTIM11noCys structure is in excellent agreement with the experimentally determined one (2.15 vs. 2.03 kcal mol^{-1} M^{-1}). For all the other DeNovoTIMs, the *m* value is similar to those observed for natural proteins with the same size, except for DeNovoTIM0 and DeNovoTIM8 where *m* decreases significantly, indicating that the native structure may not be completely well-packed or that the unfolded state has residual secondary structure (table 1). Although residual structure in the unfolded state is not clearly observed in CD spectra in 9.0 M urea (Fig. S4-S5), the persistence of native-like structure could be present at high urea concentration and not be identified by the techniques used here, as it has been reported for other proteins (*38, 39*).

The modular design approach used in this work improved both ΔG and T_m substantially and hence produced significantly more stable proteins, particularly in the second- and third-round designs. In this context, it is worth mentioning that over the years the combination of stabilizing mutations has been considered an effective strategy to enhance the stability of small proteins (*36, 40-43*). Previous work on small globular proteins with optimized hydrophobic cores and interactions on the surface exhibited increased thermal stability by up to 30 degrees (*9, 11, 15*). Extending these strategies from point mutants to regions appears to be useful for bigger folds such as the TIM barrel. In what follows, using the thermal and chemical unfolding data described above, the thermodynamic properties underlying the stability of DeNovoTIMs are analyzed.

Global thermodynamic stability and non-additive effects of DeNovoTIMs

As observed in natural proteins, the *m* values obtained from the chemical unfolding of sTIM11, sTIM11noCys, DeNovoTIM0, DeNovoTIM6, and DeNovoTIM8 correlate with their ΔC_P values determined by thermal unfolding (Fig. 3A), likely because both depend on the ΔASA upon unfolding. In contrast, ΔC_P values obtained for DeNovoTIM1, DeNovoTIM11, and DeNovoTIM12 are much lower than those expected from the reported correlation between *m* values and ΔC_P(Fig. 3A). According to the Rosetta models and the native state structures (see below), these differences are not exclusively due to properties of the native state since the calculated ΔASA is close to the expected value for the size of DeNovoTIMs (17 135 A^2; *37*). This suggests that the unfolded state reached at high temperatures is more structured than the one obtained by chemical unfolding.

The fact that many DeNovoTIMs show reversible temperature-induced unfolding allowed the assessment of their stability curves using the thermodynamic parameters obtained by DSC data (Fig. 3B). The ΔG values at 25 °C are in excellent agreement with those obtained from chemical unfolding experiments. According to the Gibbs-Helmholtz equation, conformational stability is modulated by changes in T_m, ΔH, and ΔC_P. For natural TIM barrels, it has been observed that changes in the stability curve are influenced mainly by modifying one or two of those parameters (*30, 31*). In contrast, the DeNovoTIMs differ in all three parameters.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Increasing ΔH is the most commonly found mechanism for stabilization of thermophilic proteins (*5*) and is also the most often exploited mechanism for engineering protein stability (*7, 40*). In DeNovoTIMs, this mechanism is used in all proteins but is especially important in DeNovoTIM6, that has the highest ΔH and, therefore, a ΔG higher than DeNovoTIM0, indicating an enthalpy-driven stabilization (Fig. 3B). Nevertheless, in the absence of a high-resolution structure (see below), it is difficult to rationalize how enthalpic stabilization was achieved in DeNovoTIM6 because considerable structural rearrangements take place when new interactions are introduced or molecular strain is removed. ΔC_P determines the magnitude of the curvature of the stability curve so that changes in this parameter triggers a more or less flattened curve. A decrease in ΔC_P has been postulated as a mechanism for thermostabilization (*35, 44*). For DeNovoTIMs, the reduction in ΔC_P combined with an increase in ΔH is the reason for the increase in both T_m and stability at 25 °C. The results presented here indicate that, as observed for natural proteins, in addition to the native state, the unfolded ensemble plays an important role in shaping the stability curve and should be considered in protein design.

DeNovoTIMs show a non linear correlation between ΔG at 25 °C and T_m, the most commonly used parameters that describe protein stability (inset in Fig. 3C). A similar trend between ΔG at the temperature where it is a maximum (ΔG_{Tmax}) and T_m has also been reported for natural and engineered proteins with different sizes and topologies (*34, 45, 46*). Additionally, the global thermodynamic stability can be conveniently described by the area (from 0 °C to T_m) under the stability curve (A). The advantage of A over ΔG at a given temperature is that A integrates the conformational stability in a temperature range (*47*). The relative global stability of DeNovoTIMs ($A/A_{\mathrm{DeNovoTIM0}}$) is also correlated with T_m (Fig. 3C). Notably, for DeNovoTIM6, DeNovoTIM11, and DeNovoTIM12, $A/A_{\mathrm{DeNovoTIM0}}$ is nearly ten-fold higher than for DeNovoTIM0 (Fig. 3C and table 1).

The modular strategy used to generate the DeNovoTIMs and the determination of their stabilities allowed us to calculate the contribution of each region to global stability, and also to evaluate the presence of non-additive effects between different regions of the barrel. Non-additive effects were evaluated as $\Delta\Delta G_{int}$ through an approach based on thermodynamic double mutant cycles (see Experimental Section). $\Delta\Delta G_{int}$ is also referred to as coupling energy, non-additive effect, interaction energy, and more recently epistatic effect (*48*). Thermodynamic cycles were constructed using the experimental $\Delta G_{25°C}$ values obtained from chemical unfolding experiments and linking single-region/double-region designs, and then double-region/triple-region designs as indicated in Fig. S14.

It was found that stabilization is non-additive, consequently, the different barrel regions are coupled, indicating that their contribution to protein stability depends on the structural context. A positive $\Delta\Delta G_{int}$ indicates that the introduction of favorable interactions has a higher stabilizing effect when a nearby region is already mutated. All the $\Delta\Delta G_{int}$ values calculated in Fig. S14 are summarized in the single cube shown in Fig. 3D. $\Delta\Delta G_{int}$ for single and double designs (upper face of the cube) are much smaller than those observed between double- and triple-region designs. The regions that are most energetically coupled in double-region designs are the inner core (DeNovoTIM1) and the upper peripheral core (DeNovoTIM8) ($\Delta\Delta G_{int}$= 6.1 kcal mol^{-1}, see the upper panel in Fig. S14 and arrows from top to bottom face of the cube in Fig. 3D). Coupling increases considerably when a third region is incorporated on the background of two already mutated regions ($\Delta\Delta G_{int}$> 6 kcal mol^{-1}, see lower panel in Fig. S14 and arrows on the bottom face of Fig. 3D). The largest $\Delta\Delta G_{int}$ was observed when the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DeNovoTIM8 mutations were added to DeNovoTIM11 ($\Delta\Delta G_{int}$= 14.2 kcal mol^{-1}, see purple arrow in the bottom face of Fig. 3D). Clearly, mutations in one region of the barrel can cause in a non-additive manner the loss or gain of one or more interactions in another distant region of the barrel. The latter indicates that the TIM-barrel fold is suitable for studying modularity and, in general, cooperative effects of proteins. Also, the results presented here suggest that the modular design strategy could be used in the future for the rational stability improvement in other protein topologies.

Structural analysis of DeNovoTIMs

The structural properties of DeNovoTIMs were examined by X-ray crystallography (table S6). High-resolution data were collected for sTIM11noCys and DeNovoTIM13 (1.88 and 1.64Å, respectively), whereas a low-resolution structure was obtained for DeNovoTIM6 (2.90 Å). All of them showed the designed globular compact TIM-barrel topology (Fig. 4). Structural comparison of the X-ray structures and Rosetta models for sTIM11, sTIM11noCys, DeNovoTIM6, and DeNovoTIM13 showed the lowest RMSD located in the second quarter of the barrel (ranging from 0.27 to 0.68 Å). As previously observed in sTIM11 (*29*), the main structural differences are found in the α-helices located at the amino- and carboxyl-terminal ends. In agreement, for all the barrel structures, the RMSD among quarters of the barrel is higher in the first and fourth ones (plot in Fig. 4A). Since the TIM barrel is a closed-repeat protein, contacts between the first and last helices depend on the precise curvature generated by each α/β unit, therefore geometrical strain may interfere with the proper closure of the barrel.

A comparison of the sTIM11noCys and sTIM11 structures showed that removal of the two cysteines causes some structural changes mainly localized in the first and last quarters; the most significant deviations are observed at the amino-terminal region where the first two helices are not well-formed. So even without forming the disulfide bridge, both cysteines in sTIM11 increase the stability and promote a proper closure of the barrel (Fig. 4B and table 1). The other parts of sTIM11noCys adopt almost the same structural arrangement as in sTIM11, except for the β_6/α_7 loop which was not modeled due to an absence of electron density in that region. Thus, although removing the cysteines has effects on stability and structure, sTIM11noCys maintains the general architecture corresponding to the expected TIM barrel.

The thermodynamic properties of DeNovoTIM6 are very similar to those expected for a natural protein (table 1). Unfortunately, due to the low quality of the crystals and therefore the low resolution obtained (2.90 Å), details such as side-chain conformations are not well resolved in the DeNovoTIM6 structure. Nevertheless, it could be verified that the protein is well folded into a compact TIM-barrel (Fig. 4C). As aforementioned for sTIM11 and sTIM11noCys, when the similarity between the structure and the Rosetta model is analyzed, the four quarters in DeNovoTIM6 show different RMSD values (Fig. 4C). The most similar quarter is located in the second region of DeNovoTIM6, whereas the main deviations are located in the first and last quarter of the barrel. Almost all α/β loops of the barrel are well defined and correspond to the model. However, for some residues within 5 of the 7 β/α loops no electron density was observed. The main differences observed in the structural analysis between the Rosetta model and the DeNovoTIM6 structure (table S7) are likely due to the low resolution of the data where

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

some residues and side chains are missing in the electron density map. In general, the DeNovoTIM6 structure has high B factors which may reflect higher disorder in the protein crystal or increased flexibility, similar to observations in some regions of sTIM11, namely the amino- and carboxyl-terminal α-helices. This could also explain difficulties in obtaining crystals that diffract at higher resolution despite many efforts (see Experimental Section).

As observed in all DeNovoTIMs, the similarities between the DeNovoTIM13 structure and the Rosetta model vary among the four quarters of the barrel (Fig. 4A). The second, third, and fourth quarters display minor differences between the structure and the Rosetta model, with the secondary structure elements and side chains superposing very well. The highest deviations are located at the amino- terminal region that closes the barrel (Fig. 4D). For DeNovoTIM13, the resolution of the crystal structure (1.64 Å) allowed a more in-depth analysis. Most of the hydrogen bonds and salt bridges designed are observed in the DeNovoTIM13 structure. As a consequence of the design strategy, this number is lower than the ones for sTIM11 and sTIM11noCys, stabilizing polar interactions being replaced by an increase in hydrophobic interactions in the DeNovoTIM series. For example, in going from sTIM11 to DeNovoTIM13, a 60 % increase in the total area in hydrophobic clusters was found (3765 vs. 6148 $Å^2$); most of this change comes from a three-fold increase in the area of the major hydrophobic cluster (1116 vs. 4351 $Å^2$). For DeNovoTIM13, both the area in the major hydrophobic cluster and the total hydrophobic area found in the structure are very similar to those designed (96 and 98 %, respectively; table S7).

One of the main proposed mechanisms for the stabilization of thermophilic proteins is an increase in the number of stabilizing interactions such as salt bridges and hydrogen-bond networks (*5*). In fact, in going from sTIM11 to DeNovoTIM0, a decrease in the number of electrostatic interactions is accompanied by a decrease in stability. In contrast, the structural analysis of DeNovoTIMs showed that these interactions are not clearly related to the observed changes in stability. For example, some of the designs that contained the highest number of polar stabilizing interactions (such as DeNovoTIM1 and DeNovoTIM8) were not the most stable ones, whereas some of the most stable designs (such as DeNovoTIM6 and DeNovoTIMs 12-14) showed a reduction in this type of interaction (table S7). On the contrary, the stability of DeNovoTIMs increases with the number of hydrophobic interactions. The total area, as well as the number of residues and contacts in hydrophobic clusters, are substantially increased in the best first-round designs along with the more stable second- and third-round designs (Fig. S15 and table S7). As discussed in more detail below, this suggests that repacking of the hydrophobic cores is one of the main mechanisms to increase the thermodynamic stability of DeNovoTIMs.

Epistasis on the stability landscape of *de novo* TIM barrels

To correlate the most common and informative parameters obtained from both temperature and chemical unfolding, T_m, ΔH, and $\Delta G_{25°C}$ were mapped onto a "stability landscape", a spatial representation of the observed combinations of these thermodynamic data (Fig. 5 and Experimental Section). Since DeNovoTIMs have different T_m values, experimental ΔH from DSC experiments can not be directly compared. To put the thermodynamic parameters on a similar ground for comparison, ΔH at 85 °C ($\Delta H_{85°C}$), the average T_m of the DeNovoTIM collection, was

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

calculated using ΔH and ΔC_P from DSC experiments (table 1).

The T_m range found in DeNovoTIMs is widely distributed, covering from 47 °C to 109 °C, a range of more than 60 °C in thermostability. Besides, it was possible to design TIM barrels with stabilizing mutations that led to huge differences in stability, even higher than other systems previously reported. The stability landscape of DeNovoTIMs can be compared to that constructed for natural proteins (Fig. S16). The latter is rough, with some regions more populated than others, and explores an ample space due to the diversity in size, topology, oligomeric state, function, and evolutionary history of the variety of natural proteins so far characterized. Interestingly, the comparison shows that several DeNovoTIMs are located in a region of the stability landscape corresponding to low ΔH and high $\Delta G_{25°C}$ values, which is not populated, as far as we know, by natural proteins. The modular strategy followed in the DeNovoTIM design rounds can be mapped in this stability landscape. sTIM11, sTIM11noCys, DeNovoTIM0, and most of the first-round designs cover a vast region of the landscape valley, whereas second-round designs are located in a higher stability region. Finally, the third-round design climbs to the highest region of the landscape (Fig. 5).

Assuming additivity, the expected change in stability calculated for DeNovoTIM14 would be the sum of the individual stabilizations provided by all the single-region designs (DeNovoTIM1 + DeNovoTIM6 + DeNovoTIM8) giving a value of 11.2 kcal mol^{-1}. However, the stability of DeNovoTIM14 is 23.6 kcal mol^{-1}, indicating that more than half of the stabilization comes from positive non-additive effects. The thermodynamic cube presented in Fig. 3D shows that the $\Delta\Delta G_{int}$ mentioned above increases in going from the first- to the second- and third-round designs. Non-additive effects or interaction energies may be referred to as epistasis, a concept traditionally used in genetics to describe the phenotype dependency of a mutation on the genetic state at other sites (*48-50*). Previous studies have explored and analyzed the mechanisms of epistasis within proteins, especially regarding their implications for protein function, evolution, and stability (*51-55*).

Rearrangements in the TIM barrel can influence local changes in other parts of the protein, and these epistatic effects are quantified in the $\Delta\Delta G_{int}$ values whose magnitude for DeNovoTIMs is considerable. The structural analyses suggest that one of the molecular basis of the epistatic effect observed in DeNovoTIMs is likely related to the extension of the hydrophobic cores, particularly to the increase of the major hydrophobic cluster located in the interface between the inner β-barrel and the outer α-helices (Fig. S15 and table S7). From the first- to the second-round designs, the highest area in hydrophobic clusters was found for DeNovoTIM12, and this corresponds to the highest positive epistatic effect in this round ($\Delta\Delta G_{int}$= 6.1 kcal mol^{-1}), whereas the decrease of the hydrophobic cluster area in DeNovoTIM11 (compared to DeNovoTIM1 and DeNovoTIM6) correlates with a negative $\Delta\Delta G_{int}$= -1.8 kcal mol^{-1}. From the second- to the third-round designs, the most notable change in hydrophobic area is observed in going from DeNovoTIM11 to DeNovoTIM14, resulting in the highest positive epistatic effect ($\Delta\Delta G_{int}$= 14.2 kcal mol^{-1}). The relevance and magnitude of the epistatic or non-additive effects found in DeNovoTIMs, as well as those observed in other reports, suggest that modeling such interactions can improve the success in protein design and engineering.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Conclusions

Design requires a deep understanding of the relationship between sequence, structure, and stability, and therefore, the combination of thermodynamic and structural data is fundamental to achieve this goal. Here, we designed a family of stable TIM barrels and explored their stability landscape. The TIM-barrel collection reported in this work exhibits a considerable range in thermostability (more than 60 degrees in T_m) and conformational stability at 25 °C (more than 22 kcal mol^{-1} in ΔG). These data can now be used to accelerate the development of future custom design protein stability curves which, in turn, will expand the biomedical and biotechnological applications of *de novo* proteins. For example, by fusion to another *de novo* protein, one of the stabilized scaffolds reported here (DeNovoTIM13) has been successfully used to create a reaction chamber on the top of the barrel (56), confirming the convenience of working with robust and stable TIM barrels in the path towards functional *de novo* proteins.

In the same way that one explores the sequence space by studying homologous proteins from different organisms, *de novo* design with a fixed backbone follows a similar strategy generating new sequences within the same topology. It is well known that highly stable proteins can be generated by computational design. However, one of the unexpected findings resulting from the thermodynamic characterization of this family of DeNovoTIMs is that very stable proteins can be obtained in unexplored regions of the stability landscape. The paths followed in the stability landscape of DeNovoTIMs are severely influenced by epistatic effects that appear to arise from an increase in hydrophobic clusters. The design and characterization of stable *de novo* proteins is an essential step on the route to the next generation of new protein functions charting novel sequence space.

Acknowledgments

We acknowledge financial support and allocation of beamtime by PSI and HZB. We thank the beamline staff at the SLS and at BESSY for assistance, and LANEM-IQ-UNAM for the support in crystal characterization. We thank María Isabel Velázquez López, Laura Iliana Alvarez Añorve, Alma Jessica Díaz Salazar, and Georgina Espinosa Pérez for their competent technical support, Gregor Wiese for generating and crystallizing sTIM11noCys, Noelia Ferruz-Capapey for her help in the structural analyses, as well as Po-Ssu Huang for his comments on the manuscript. We kindly thank all the members of the Fernandez-Velasco, Höcker, and Baker Labs for their constructive suggestions to improve the research.

Funding

This work was supported by scholarships from CONACYT (749489 to C.T., 387653, 291062, 14401, and 27897 to S.R.R), UNAM-DGAPA-PAPIIT (IN220516 to S.R.R.), and UNAM-DGAPA (postdoctoral fellowship to Y.G.). This research was also financed by grants from CONACYT (221169 to A.R.R., 254514 to D.A.F.V.), UNAM-DGAPA-PAPIIT (IN220519 to M.C., IN208418 to A.R.R., IN219519 and IN220516 to D.A.F.V.), and Programa de Apoyo a la Investigación y el Posgrado FQ-UNAM (5000-9018 to M.C.). B.H. gratefully acknowledges financial support by

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the European Research Council (ERC Consolidator Grant 647548 'Protein Lego') and by HZB to visit the beamlines at BESSY.

Author Contributions

S.R.R., M.C., D.B., B.H., and D.A.F.V. designed the research. D-A.S.M. wrote program code. S.R.R., E.R., C.T., and M.C. collected thermodynamic data. S.R.R., S.K., S.S., Y.G., and A.R.R. solved the crystal structures. S.R.R., M.C., and D.A.F.V. analyzed all the experimental data. S.R.R., B.H., and D.A.F.V. wrote the manuscript. All authors discussed and commented on the manuscript.

Competing interests

Authors declare no competing interests.

Data and materials availability

All data to support the conclusions of this manuscript are included in the main text and supporting information. Coordinates and structure files have been deposited to the Protein Data Bank (PDB) with accession codes: 6YQY (sTIM11noCys), 6Z2I (DeNovoTIM6), and 6YQX (DeNovoTIM13).

Supporting Information

Supporting Text, Experimental Section, Supplementary References, Supplementary Figures S1 to S16, and Supplementary Tables S1 to S7 are available in the supporting information of the paper.

References and Footnotes

1. Tokuriki, N.; Tawfik, D. S. Stability Effects of Mutations and Protein Evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19* (5), 596–604. https://doi.org/10.1016/j.sbi.2009.08.003.
2. Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (15), 5869–5874. https://doi.org/10.1073/pnas.0510098103.
3. Becktel, W. J.; Schellman, J. A. Protein Stability Curves. *Biopolymers* **1987**, *26* (11), 1859–1877. https://doi.org/10.1002/bip.360261104.
4. Nojima, H.; Ikai, A.; Oshima, T.; Noda, H. Reversible Thermal Unfolding of Thermostable Phosphoglycerate Kinase. Thermostability Associated with Mean Zero Enthalpy

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Change. *J. Mol. Biol.* **1977**, *116* (3), 429–442. https://doi.org/10.1016/0022-2836(77)90078-X.

5. Razvi, A.; Scholtz, J. M. Lessons in Stability from Thermophilic Proteins. *Protein Sci.* **2006**, *15* (7), 1569–1578. https://doi.org/10.1110/ps.062130306.
6. Huang, P. S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327. https://doi.org/10.1038/nature19946.
7. Islam, M. M.; Kobayashi, K.; Kidokoro, S. I.; Kuroda, Y. Hydrophobic Surface Residues Can Stabilize a Protein through Improved Water–Protein Interactions. *FEBS J.* **2019**, *286* (20), 4122–4134. https://doi.org/10.1111/febs.14941.
8. Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87* (1), 105–129. https://doi.org/10.1146/annurev-biochem-062917-012102.
9. Kim, D. N.; Jacobs, T. M.; Kuhlman, B. Boosting Protein Stability with the Computational Design of β-Sheet Surfaces. *Protein Sci.* **2016**, *25* (3), 702–710. https://doi.org/10.1002/pro.2869.
10. Huang, P. S.; Oberdorfer, G.; Xu, C.; Pei, X. Y.; Nannenga, B. L.; Rogers, J. M.; DiMaio, F.; Gonen, T.; Luisi, B.; Baker, D. High Thermodynamic Stability of Parametrically Designed Helical Bundles. *Science (80-.).* **2014**, *346* (6208), 481–485. https://doi.org/10.1126/science.1257481.
11. Borgo, B.; Havranek, J. J. Automated Selection of Stabilizing Mutations in Designed and Natural Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (5), 1494–1499. https://doi.org/10.1073/pnas.1115172109.
12. Kuhlman, B.; Baker, D. Exploring Folding Free Energy Landscapes Using Computational Protein Design. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 89–95. https://doi.org/10.1016/j.sbi.2004.01.002.
13. Kwok, S. C.; Hodges, R. S. Clustering of Large Hydrophobes in the Hydrophobic Core of Two-Stranded α-Helical Coiled-Coils Controls Protein Folding and Stability. *J. Biol. Chem.* **2003**, *278* (37), 35248–35254. https://doi.org/10.1074/jbc.M305306200.
14. Selvaraj, S.; Gromiha, M. M. Role of Hydrophobic Clusters and Long-Range Contact Networks in the Folding of (α/β)8 Barrel Proteins. *Biophys. J.* **2003**, *84* (3), 1919–1925. https://doi.org/10.1016/S0006-3495(03)75000-0.
15. Malakauskas, S. M.; Mayo, S. L. Design, Structure and Stability of a Hyperthermophilic Protein Variant. *Nat. Struct. Biol.* **1998**, *5* (6), 470–475. https://doi.org/10.1038/nsb0698-470.
16. Dahiyat, B. I.; Mayo, S. L. Probing the Role of Packing Specificity in Protein Design. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (19), 10172–10177. https://doi.org/10.1073/pnas.94.19.10172.
17. Colón, W.; Elöve, G. A.; Wakem, L. P.; Sherman, F.; Roder, H. Side Chain Packing of the N- and C-Terminal Helices Plays a Critical Role in the Kinetics of Cytochrome c Folding. *Biochemistry* **1996**, *35* (17), 5538–5549. https://doi.org/10.1021/bi960052u.
18. Fersht, A. R.; Serrano, L. Principles of Protein Stability Derived from Protein Engineering Experiments. *Curr. Opin. Struct. Biol.* **1993**, *3* (1), 75–83. https://doi.org/10.1016/0959-440X(93)90205-Y.
19. Basak, S.; Nobrega, P. R.; Tavella, D.; Deveau, L. M.; Koga, N.; Tatsumi-Koga, R.; Baker, D.; Massi, F.; Matthews, C. R. Networks of Electrostatic and Hydrophobic Interactions Modulate the Complex Folding Free Energy Surface of a Designed Bα Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (14), 6806–6811.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

https://doi.org/10.1073/pnas.1818744116.

20. Watters, A. L.; Deka, P.; Corrent, C.; Callender, D.; Varani, G.; Sosnick, T.; Baker, D. The Highly Cooperative Folding of Small Naturally Occurring Proteins Is Likely the Result of Natural Selection. *Cell* **2007**, *128* (3), 613–624. https://doi.org/10.1016/j.cell.2006.12.042.
21. Sterner, R.; Höcker, B. Catalytic Versatility, Stability, and Evolution of the (Betaalpha)8-Barrel Enzyme Fold. *Chem. Rev.* **2005**, *105* (11), 4038–4055. https://doi.org/10.1021/cr030191z.
22. Braselmann, E.; Chaney, J. L.; Clark, P. L. Folding the Proteome. *Trends Biochem. Sci.* **2013**, *38* (7), 337–344. https://doi.org/10.1016/j.tibs.2013.05.001.
23. Romero-Romero, S.; Kordes, S.; Michel, F.; Höcker, B. Evolution, Folding, and Design of TIM Barrels and Related Proteins. *Curr. Opin. Struct. Biol.* **2021**, *68*, 94–104. https://doi.org/10.1016/j.sbi.2020.12.007.
24. Nagarajan, D.; Deka, G.; Rao, M. Design of Symmetric TIM Barrel Proteins from First Principles. *BMC Biochem.* **2015**, *16* (1), 1–22. https://doi.org/10.1186/s12858-015-0047-4.
25. Figueroa, M.; Oliveira, N.; Lejeune, A.; Kaufmann, K. W.; Dorr, B. M.; Matagne, A.; Martial, J. A.; Meiler, J.; Van de Weerdt, C. Octarellin VI: Using Rosetta to Design a Putative Artificial (β/α)8 Protein. *PLoS One* **2013**, *8* (8), e71858. https://doi.org/10.1371/journal.pone.0071858.
26. Höcker, B.; Lochner, A.; Seitz, T.; Claren, J.; Sterner, R. High-Resolution Crystal Structure of an Artificial (Bα) 8-Barrel Protein Designed from Identical Half-Barrels. *Biochemistry* **2009**, *48* (6), 1145–1147. https://doi.org/10.1021/bi802125b.
27. Höcker, B.; Claren, J.; Sterner, R. Mimicking Enzyme Evolution by Generating New (Betaalpha)8-Barrels from (Betaalpha)4-Half-Barrels. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (47), 16448–16453. https://doi.org/10.1073/pnas.0405832101.
28. Offredi, F.; Dubail, F.; Kischel, P.; Sarinski, K.; Stern, a S.; van de Weerdt, C.; Hoch, J. C.; Prosperi, C.; Francois, J. M.; Mayo, S. L.; Martial, J. a. De Novo Backbone and Sequence Design of an Idealized Alpha/ Protein: Evidence of Stable Tertiary Structure. *J. Mol. Biol.* **2004**, *325* (1), 163–174. https://doi.org/10.1061/S022-2836(02)01206-8.
29. Huang, P. S.; Feldmeier, K.; Parmeggiani, F.; Fernandez-Velasco, D. A.; Hocker, B.; Baker, D. De Novo Design of a Four-Fold Symmetric TIM-Barrel Protein with Atomic-Level Accuracy. *Nat. Chem. Biol.* **2016**, *12* (1), 29–34. https://doi.org/10.1038/nchembio.1966.
30. Romero-Romero, S.; Costas, M.; Rodríguez-Romero, A.; Fernández-Velasco, D. A. Reversibility and Two State Behaviour in the Thermal Unfolding of Oligomeric TIM Barrel Proteins. *Phys. Chem. Chem. Phys.* **2015**, *17* (32), 20699–20714. https://doi.org/10.1039/c5cp01599e.
31. Cipolla, A.; Delbrassine, F.; Da Lage, J. L.; Feller, G. Temperature Adaptations in Psychrophilic, Mesophilic and Thermophilic Chloride-Dependent Alpha-Amylases. *Biochimie* **2012**, *94* (9), 1943–1950. https://doi.org/10.1016/j.biochi.2012.05.013.
32. Rudolph, R.; Siebendritt, R.; Kiefhaber, T. Reversible Unfolding and Refolding Behavior of a Monomeric Aldolase from Staphylococcus Aureus. *Protein Sci.* **1992**, *1* (5), 654–666. https://doi.org/10.1002/pro.5560010511.
33. Uversky, V. N. Use of Fast Protein Size-Exclusion Liquid Chromatography to Study the Unfolding of Proteins Which Denature through the Molten Globule. *Biochemistry* **1993**, *32* (48), 13288–13298. https://doi.org/10.1021/bi00211a042.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

34. Robertson, A. D.; Murphy, K. P. Protein Structure and the Energetics of Protein Stability. *Chem. Rev.* **1997**, *97* (5), 1251–1267. https://doi.org/10.1021/cr960383c.
35. Robic, S.; Guzman-Casado, M.; Sanchez-Ruiz, J. M.; Marqusee, S. Role of Residual Structure in the Unfolded State of a Thermophilic Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (20), 11345–11349. https://doi.org/10.1073/pnas.1635051100.
36. Lamazares, E.; Clemente, I.; Bueno, M.; Velázquez-Campoy, A.; Sancho, J. Rational Stabilization of Complex Proteins: A Divide and Combine Approach. *Sci. Rep.* **2015**, *5* (9129), 1–11. https://doi.org/10.1038/srep09129.
37. 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* (10), 2138–2148. https://doi.org/10.1002/pro.5560041020.
38. Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. Long-Range Interactions within a Nonnative Protein. *Science (80-.).* **2002**, *295* (5560), 1719–1722. https://doi.org/10.1126/science.1067680.
39. Shortle, D.; Ackerman, M. S. Persistence of Native-like Topology in a Denatured Protein in 8 M Urea. *Science (80-.).* **2001**, *293* (5529), 487–489. https://doi.org/10.1126/science.1060438.
40. Haruki, M.; Tanaka, M.; Motegi, T.; Tadokoro, T.; Koga, Y.; Takano, K.; Kanaya, S. Structural and Thermodynamic Analyses of Escherichia Coli RNase HI Variant with Quintuple Thermostabilizing Mutations. *FEBS J.* **2007**, *274* (22), 5815–5825. https://doi.org/10.1111/j.1742-4658.2007.06104.x.
41. Serrano, L.; Day, A. G.; Fersht, A. R. Step-Wise Mutation of Barnase to Binase. A Procedure for Engineering Increased Stability of Proteins and an Experimental Analysis of the Evolution of Protein Stability. *J. Mol. Biol.* **1993**, *233*, 305–312. https://doi.org/https://doi.org/10.1006/jmbi.1993.1508.
42. Stearman, R. S.; Frankel, A. D.; Freire, E.; Liu, B.; Pabo, C. O. Combining Thermostable Mutations Increases the Stability of λ Repressor. *Biochemistry* **1988**, *27* (19), 7571–7574. https://doi.org/10.1021/bi00419a059.
43. Matsumura, M.; Signor, G.; Matthews, B. W. Substantial Increase of Protein Stability by Multiple Disulphide Bonds. *Nature* **1989**, *342* (6247), 291–293. https://doi.org/10.1038/342291a0.
44. Hart, K. M.; Harms, M. J.; Schmidt, B. H.; Elya, C.; Thornton, J. W.; Marqusee, S. Thermodynamic System Drift in Protein Evolution. *PLoS Biol.* **2014**, *12* (11), e1001994. https://doi.org/10.1371/journal.pbio.1001994.
45. Howell, S. C.; Inampudi, K. K.; Bean, D. P.; Wilson, C. J. Understanding Thermal Adaptation of Enzymes through the Multistate Rational Design and Stability Prediction of 100 Adenylate Kinases. *Structure* **2014**, *22* (2), 218–229. https://doi.org/10.1016/j.str.2013.10.019.
46. Rees, D. C.; Robertson, A. D. Some Thermodynamic Implications for the Thermostability of Proteins. *Protein Sci.* **2001**, *10* (6), 1187–1194. https://doi.org/10.1110/ps.180101.
47. Alfano, C.; Sanfelice, D.; Martin, S. R.; Pastore, A.; Temussi, P. A. An Optimized Strategy to Measure Protein Stability Highlights Differences between Cold and Hot Unfolded States. *Nat. Commun.* **2017**, *8* (May), 1–9. https://doi.org/10.1038/ncomms15428.
48. Horovitz, A.; Fleisher, R. C.; Mondal, T. Double-Mutant Cycles: New Directions and Applications. *Curr. Opin. Struct. Biol.* **2019**, *58* (Oct), 10–17.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

https://doi.org/10.1016/j.sbi.2019.03.025.

49. Miton, C. M.; Tokuriki, N. How Mutational Epistasis Impairs Predictability in Protein Evolution and Design. *Protein Sci.* **2016**, *25* (7), 1260–1272. https://doi.org/10.1002/pro.2876.
50. Starr, T. N.; Thornton, J. W. Epistasis in Protein Evolution. *Protein Sci.* **2016**, *25* (7), 1204–1218. https://doi.org/10.1002/pro.2897.
51. Risso, V. A.; Manssour-Triedo, F.; Delgado-Delgado, A.; Arco, R.; Barroso-delJesus, A.; Ingles-Prieto, A.; Godoy-Ruiz, R.; Gavira, J. A.; Gaucher, E. A.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M. Mutational Studies on Resurrected Ancestral Proteins Reveal Conservation of Site-Specific Amino Acid Preferences throughout Evolutionary History. *Mol. Biol. Evol.* **2015**, *32* (2), 440–455. https://doi.org/10.1093/molbev/msu312.
52. Ashenberg, O.; Gong, L. I.; Bloom, J. D. Mutational Effects on Stability Are Largely Conserved during Protein Evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (52), 21071–21076. https://doi.org/10.1073/pnas.1314781111.
53. Natarajan, C.; Inoguchi, N.; Weber, R. E.; Fago, A.; Moriyama, H.; Storz, J. F. Epistasis Among Adaptive Mutations in Deer Mouse Hemoglobin. *Science (80-.).* **2013**, *340* (6138), 1324–1327. https://doi.org/10.1126/science.1236862.
54. Lynch, V. J.; May, G.; Wagner, G. P. Regulatory Evolution through Divergence of a Phosphoswitch in the Transcription Factor CEBPB. *Nature* **2011**, *480* (7377), 383–386. https://doi.org/10.1038/nature10595.
55. Ortlund, E. A.; Bridgham, J. T.; Redinbo, M. R.; Thornton, J. W. Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science (80-.).* **2007**, *317* (5844), 1544–1548. https://doi.org/10.1126/science.1142819.
56. Caldwell, S. J.; Haydon, I. C.; Piperidou, N.; Huang, P. S.; Bick, M. J.; Sebastian Sjöström, H.; Hilvert, D.; Baker, D.; Zeymer, C. Tight and Specific Lanthanide Binding in a de Novo TIM Barrel with a Large Internal Cavity Designed by Symmetric Domain Fusion. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (48), 30362–30369. https://doi.org/10.1073/pnas.2008535117.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and Tables

List of Figures

- Fig. 1: Modular design approach to obtain the DeNovoTIM collection.
- Fig. 2: Conformational properties and equilibrium unfolding of DeNovoTIMs.
- Fig. 3: Stability and energetic coupling in DeNovoTIMs.
- Fig. 4: Three-dimensional structures of DeNovoTIMs.
- Fig. 5: Stability landscape of *de novo* TIM barrels.

List of Tables

- Table 1: Thermodynamic properties of DeNovoTIMs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

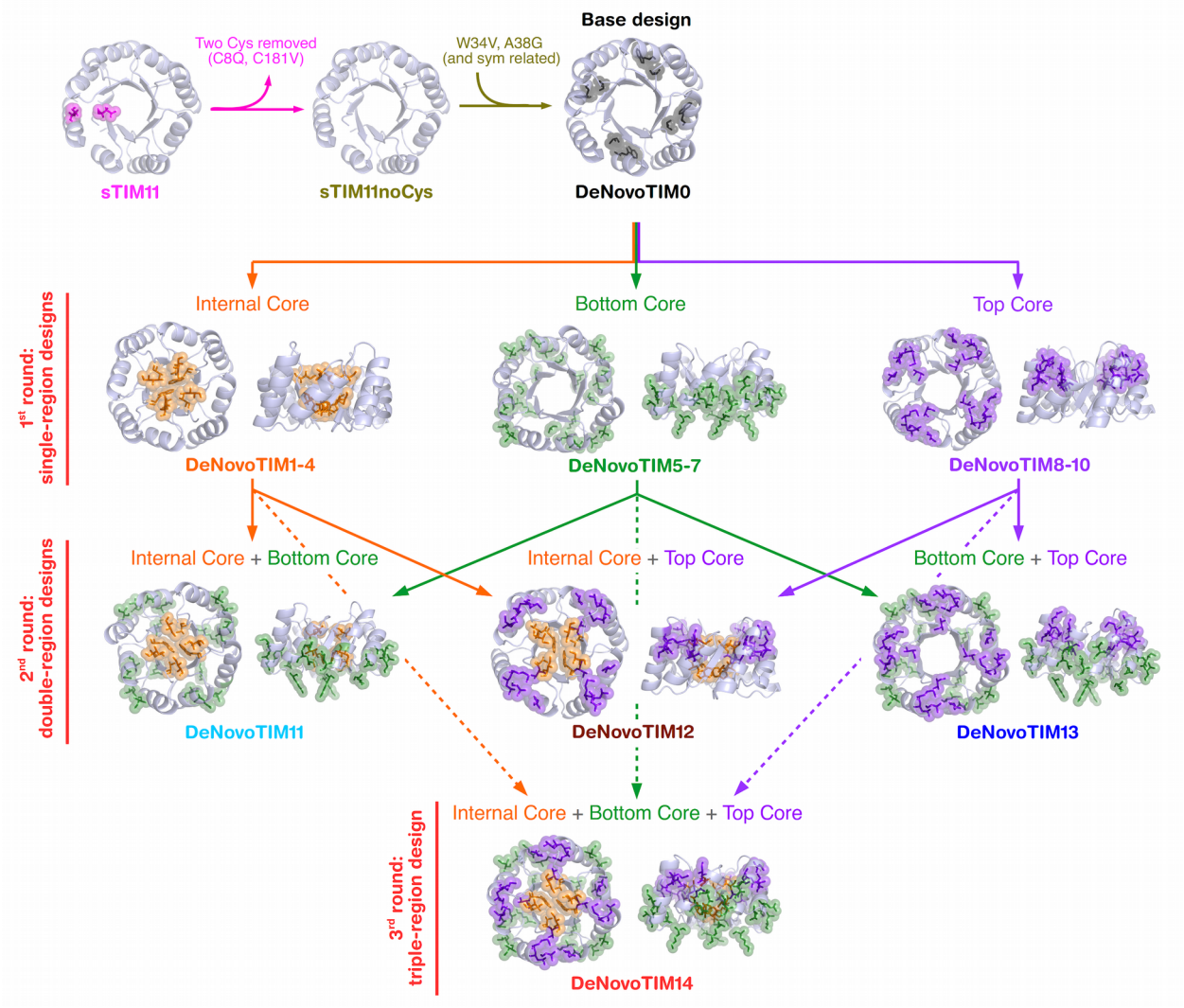

Fig. 1. Modular design approach to obtain the DeNovoTIM collection. Cartoon representation of the regions and the corresponding residues modified in each design round. The two cysteine residues present in sTIM11 that were reverted to the corresponding symmetry-related residues in sTIM11noCys are shown in magenta (C8Q and C181V). Mutations W34V and A38G (as well as their 4-fold-symmetry related residues) introduced in DeNovoTIM0 are shown in black. The inner core, formed by the β-barrel residues A21, R23, I40, I42, A67, R69, I86, and I88 (as well as their 2 fold-symmetry related residues) is shown in orange. The peripheral bottom core, formed by the N-terminal region of even β-strands and the C-region of the flanking α-helices, that is, residues Q11, E15, T18, K31, and V34 (as well as their 4-fold-symmetry related residues) is colored green. Peripheral top core situated at the C-terminal region of the odd β-strands and the N-terminal region of the flanking α-helices formed by residues K2, A5, W6, Y22, S24, and D29 (as well as their 4-fold-symmetry related residues) is shown in purple. All the sequences analyzed in this work are reported in Fig. S2 and tables S1-S2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

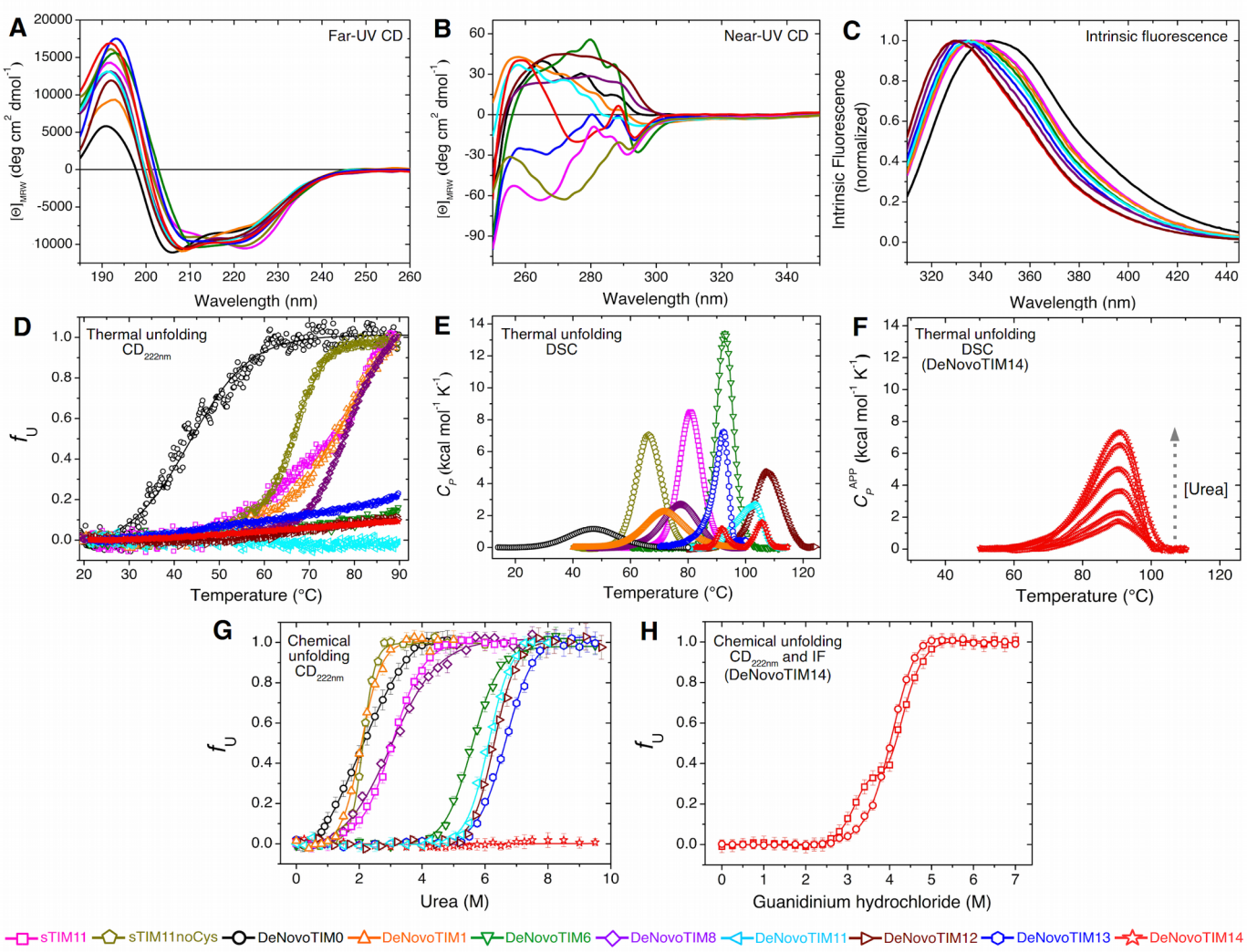

Fig. 2. Conformational properties and equilibrium unfolding of DeNovoTIMs. A) Far-UV CD spectra. **B)** Near-UV CD spectra. **C)** Intrinsic Fluorescence (IF) spectra (λ_{exc}= 295 nm). **D)** Thermal unfolding followed by CD_{222nm} (scan rate: 1.5 K hr^{-1}). **E)** Differential Scanning Calorimetry (DSC) endotherms (scan rate: 1.5 K hr^{-1}; for easy comparison, the physical and chemical baselines have been subtracted). **F)** DSC endotherms of DeNovoTIM14 in the presence of increasing concentrations of urea (2.0 to 6.0 M) from bottom to top (scan rate: 1.5 K hr^{-1}). For clarity, in panels E and F only a small part of the pre- and post-transition baselines are shown. **G)** Chemical unfolding using urea and followed by CD (notice that DeNovoTIM14 does not unfold with urea). **H)** Chemical unfolding induced by guanidinium hydrochloride for DeNovoTIM14 (squares: CD, circles: IF).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

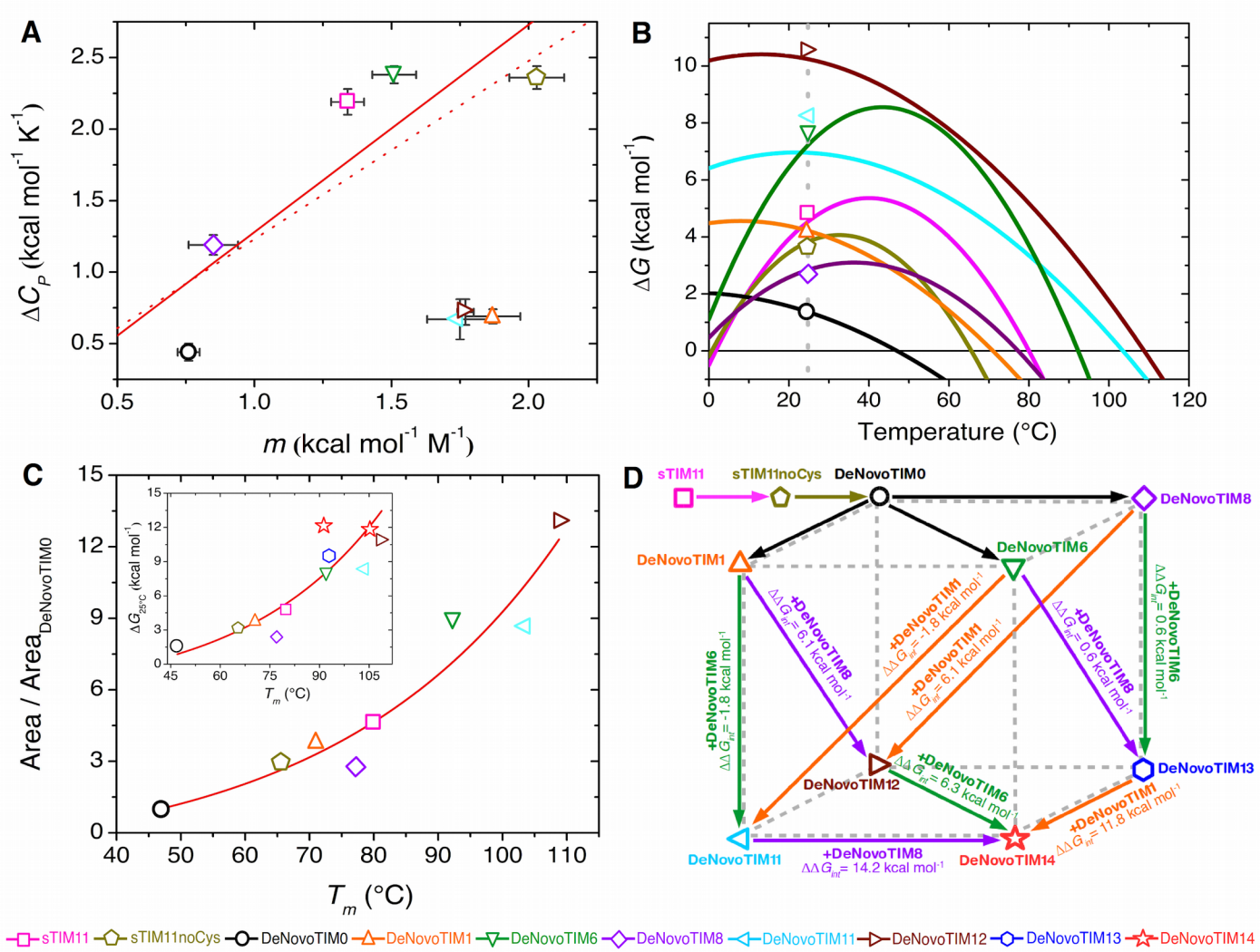

Fig. 3. Stability and energetic coupling in DeNovoTIMs. A) Correlation between two parameters which are proportional to the exposed surface area: *m* value from chemical unfolding and ΔC_P from temperature-induced unfolding (solid line: linear regression excluding DeNovoTIM1, DeNovoTIM11, and DeNovoTIM12 data; R^2: 0.76. Dotted line: correlation reported by *37*). **B)** Stability curves calculated from DSC data (lines) using the Gibbs-Helmholtz equation (open symbols show ΔG values determined by chemical unfolding at 25 °C. Grey dashed line indicates 25 °C). **C)** Correlation between the relative global thermodynamic stability ($Area/Area_{DeNovoTIM0}$) and thermostability (T_m) (R^2: 0.93). Inset: correlation between ΔG at 25 °C determined by chemical unfolding and T_m (R^2: 0.87). For DeNovoTIM14, where two transitions were found, it was assumed that the one observed at lower [GdnHCl] corresponds to the lower T_m. **D)** Thermodynamic cube showing the coupling energy ($\Delta\Delta G_{int}$) between different regions of DeNovoTIMs. $\Delta\Delta G_{int}$ values were calculated from the double-mutant cycles shown in Fig. S14. $\Delta\Delta G_{int}$ values between single-region mutants are depicted as colored arrows from the top face to the bottom face. $\Delta\Delta G_{int}$ values calculated for the addition of a single-region design to a double-region design are shown as colored arrows in the bottom face.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

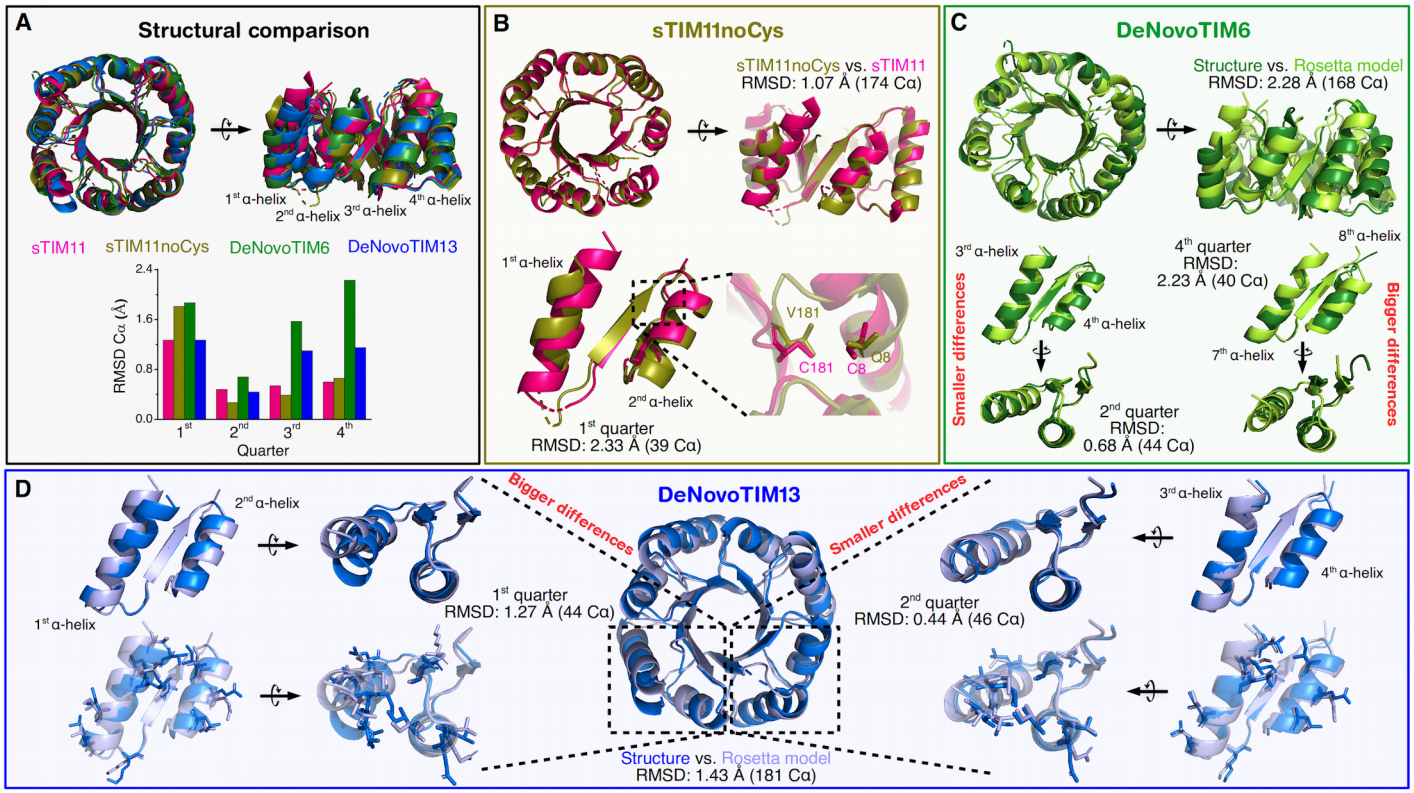

Fig. 4. Three-dimensional structures of DeNovoTIMs. A) Structural alignment of X-ray structures of sTIM11 (PDB ID: 5BVL), sTIM11noCys (PDB ID: 6YQY), DeNovoTIM6 (PDB ID: 6Z2I), and DeNovoTIM13 (PDB ID: 6YQX). The RMSD Cα between the structure and the Rosetta model among the quarters in each protein is shown in the lower part of the panel. B) Comparison of sTIM11noCys and sTIM11 structures (RMSD: 1.07 Å -174 Cα-). The mutated residues 8 and 181 in sTIM11noCys are zoomed in the bottom part. C) Comparison of the DeNovoTIM6 structure with the Rosetta model (RMSD: 2.28 Å -168 Cα-). The quarters with the highest and lowest structural similarity are highlighted (bottom left and bottom right, respectively). D) Comparison of the DeNovoTIM13 structure with the Rosetta model (RMSD: 1.43 Å -181 Cα-). The quarters with the highest and lowest structural similarity are highlighted (right and left, respectively). Sidechains of the mutated residues are shown in sticks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

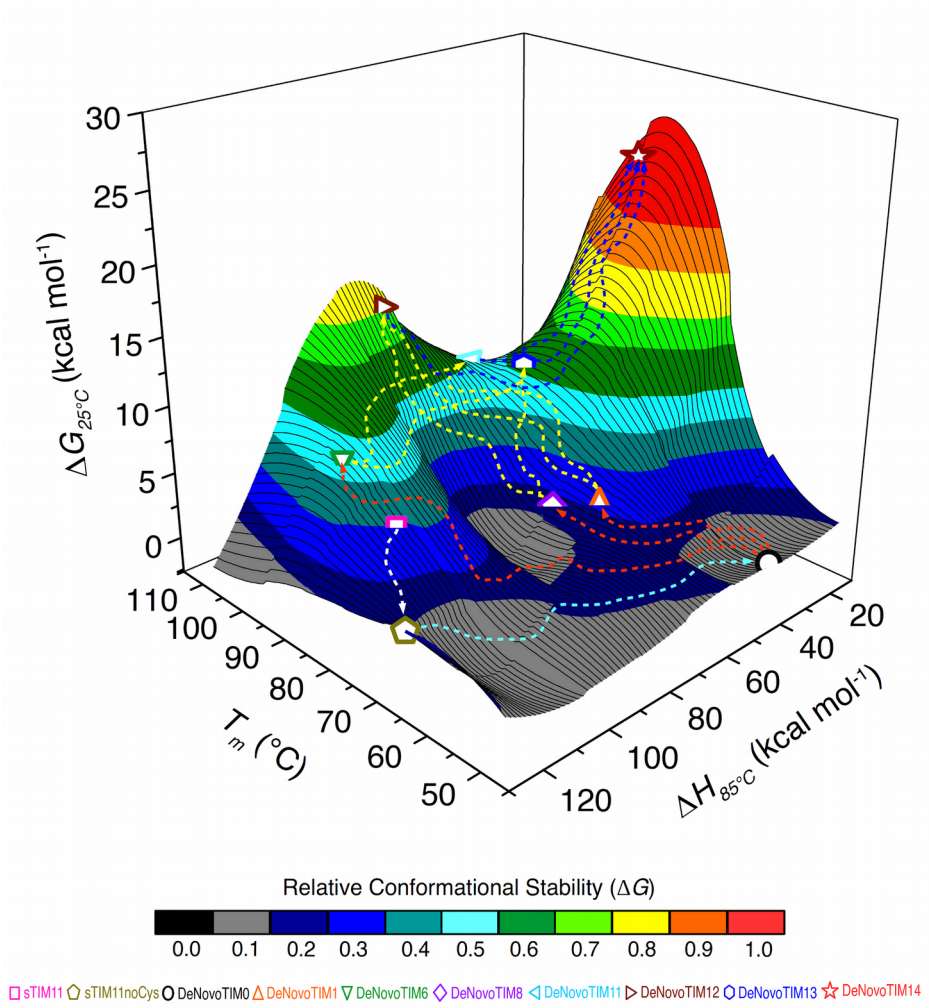

Fig. 5. Stability landscape of *de novo* TIM barrels. The stability surface is colored according to normalized $\Delta G_{25°C}$ values in 0.1 bins. Colored lines represent a possible pathway from one design to another and were drawn only as a guide to the eye: white from sTIM11 to sTIM11noCys, cyan from sTIM11noCys to DeNovoTIM0, orange from DeNovoTIM0 to the first-round DeNovoTIMs, yellow from single-region designs to the double-region ones, and blue from double-region designs to the triple-region design. T_m and ΔH data were obtained from thermal unfolding, whereas ΔG values derive from chemical unfolding. $\Delta H_{85°C}$ was calculated as indicated in the experimental section, except for DeNovoTIM13 and DeNovoTIM14, where ΔH at T_m was plotted.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.29.319103; this version posted March 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Thermodynamic properties of DeNovoTIMs.

de novo TIM barrel	Type of modular design	Thermal unfolding (by DSC)[a]						Chemical unfolding (by CD and IF)[e]		
		T_m (°C)	ΔH (kcal mol^{-1})	$\Delta H_{85°C}$ (kcal mol^{-1})[b]	ΔC_P (kcal mol^{-1} K^{-1})	ΔH_{vH} / ΔH	Global Thermodynamic stability (kcal K mol^{-1})[c]	$\Delta G_{25°C}$ (kcal mol^{-1})	m (kcal mol^{-1} M^{-1})	$D_{[1/2]}$ (M)
sTIM11	First reported *de novo* TIM barrel	80.0 ± 0.2	92.9 ± 1.3	104.0 ± 1.8	2.19 ± 0.09	0.99 ± 0.05	279.6	4.8 ± 0.3	1.34 ± 0.06	3.1
sTIM11noCys	sTIM11 without cysteines	65.6 ± 0.1	82.0 ± 0.6	127.8 ± 2.2	2.36 ± 0.08	0.99 ± 0.03	175.8	3.2 ± 0.2	2.03 ± 0.10	1.9
DeNovoTIM0	Base design	47.0 ± 0.2	24.7 ± 0.6	41.4 ± 2.9	0.44 ± 0.06	1.09 ± 0.08	60.5	1.5 ± 0.1	0.76 ± 0.02	2.1
DeNovoTIM1	Single-region design: Internal core	71.0 ± 0.4	48.1 ± 0.8	57.8 ± 1.5	0.69 ± 0.05	0.98 ± 0.05	225.9	3.8 ± 0.1	1.87 ± 0.10	2.0
DeNovoTIM6	Single-region design: Bottom cavity	92.3 ± 0.1	124.9 ± 1.5	107.5 ± 1.1	2.38 ± 0.06	1.03 ± 0.02	542.2	7.9 ± 0.2	1.51 ± 0.08	5.6
DeNovoTIM8	Single-region design: Top cavity	77.3 ± 0.3	51.9 ± 0.8	61.1.81.3	1.19 ± 0.07	0.95 ± 0.09	165.2	2.5 ± 0.2	0.85 ± 0.09	2.9
DeNovoTIM11	Double-region design: Internal core + bottom cavity (DeNovoTIM1 + DeNovoTIM6)	103.5 ± 0.2	61.7 ± 2.4	49.3 ± 0.9	0.67 ± 0.14	1.02 ± 0.09	523.2	8.4 ± 0.4	1.75 ± 0.08	6.0
DeNovoTIM12	Double-region design: internal core + top cavity (DeNovoTIM1 + DeNovoTIM8)	108.8 ± 0.3	79.3 ± 1.9	62.2 ± 1.0	0.72 ± 0.08	1.01 ± 0.08	791.5	10.9 ± 0.2	1.77 ± 0.03	6.2
DeNovoTIM13	Double-region design: bottom cavity + top cavity (DeNovoTIM6 + DeNovoTIM8)	92.8 ± 0.4	46.7 ± 4.5	ND E_{act}: 120.3 ± 2.8 kcal mol^{-1} [d]			ND	9.5 ± 0.2	1.54 ± 0.03	6.6
DeNovoTIM14	Triple-region design: internal core + bottom cavity + top cavity (DeNovoTIM1 + DeNovoTIM6 + DeNovoTIM8)	91.5 ± 0.1	5.4 ± 0.2	ND E_{act}: 37.4 ± 0.5 kcal mol^{-1} [d]			ND	Tot: 23.6 ± 2.0 N to I: 12.7 ± 1.9 I to U: 10.9 ± 0.6	N to I: 4.3 ± 0.9 I to U: 2.5 ± 0.1	N to I: 3.2 I to U: 4.4
		105.3 ± 0.1	8.6 ± 0.4							

[a] The thermodynamic parameters reported are the average of ten experiments carried out at different protein concentrations (0.25-2.5 mg mL^{-1}; Fig. S8 and Fig. S11), ± indicate the standard deviation calculated from these 10 experiments.
[b] ΔH at 85 °C ($\Delta H_{85°C}$) was calculated using the experimental ΔH and ΔC_P values as indicated in the experimental section.
[c] The global thermodynamic stability was calculated from the area of the stability curve evaluated between 0 °C and T_m (Fig. 3B).
ND: Not determined due to irreversibility in the thermal unfolding. Instead, activation energy (E_{act}) was calculated from an irreversible two-state mechanism.
[d] E_{act} value and ± are the average and standard deviation, respectively, from three different calculation methods (Fig. S9 and Fig. S10).
[e] ± indicate the standard error from global fitting (Fig. S13).