Chapter 5

Protein Thermal Stability Engineering Using HoTMuSiC

Fabrizio Pucci, Jean Marc Kwasigroch, and Marianne Rooman

Abstract

The rational design of enzymes is a challenging research field, which plays an important role in the optimization of a wide series of biotechnological processes. Computational approaches allow screening all possible amino acid substitutions in a target protein and to identify a subset likely to have the desired properties. They can thus be used to guide and restrict the huge, time-consuming search in sequence space to reach protein optimality. Here we present HoTMuSiC, a tool that predicts the impact of point mutations on the protein melting temperature, which uses the experimental or modeled protein structure as sole input and is available at the dezyme.com website. Its main advantages include accuracy and speed, which makes it a perfect instrument for thermal stability engineering projects aiming at designing new proteins that feature increased heat resistance or remain active and stable in nonphysiological conditions. We set up a HoTMuSiC-based pipeline, which uses additional information to avoid mutations of functionally important residues, identified as being too well conserved among homologous proteins or too close to annotated functional sites. The efficiency of this pipeline is successfully demonstrated on *Rhizomucor miehei* lipase.

Key words Protein melting temperature, Protein design, Thermal stability, Statistical potentials, Artificial neural network, Lipase

1 Introduction

In the last decades, lots of efforts have been devoted to analyze the molecular mechanisms associated to protein thermal stability, since their understanding is fundamental not only for advancing the theoretical comprehension of the protein folding process but also for potential applications to a wide series of biological processes that range from drug design to the synthesis of new protein nanomaterials. The design of new enzymes that remain active and stable at temperatures well above or below their physiological temperature can also lead to the improvement of the efficiency of catalytic processes while reducing their economic costs and their environmental impact [1–3].

Different experimental and computational approaches have been developed and largely used to enhance protein thermal resistance [4–6] such as:

Zoltán Gáspári (ed.), *Structural Bioinformatics: Methods and Protocols*, Methods in Molecular Biology, vol. 2112,
https://doi.org/10.1007/978-1-0716-0270-6_5,

- **Directed evolution methods** in which randomly distributed mutations are introduced in the target protein and are followed by screening and selection steps.
- **Rational protein design** in which the understanding of the protein structure/function relationships is the key ingredient.
- **Semi-rational approaches** that combine the benefits of the directed evolution and the rational design methods.

Despite impressive current achievements, the development of a systematic and cheap way to optimize a target protein remains a challenging goal. This is primarily due to the huge size of the sequence space to be explored and to the incomplete knowledge of the molecular mechanisms involved.

Here we present some advances in the computational protein design field by describing the HoTMuSiC software [7] that we recently developed. HoTMuSiC is an efficient tool that, given the experimental or modeled three-dimensional (3D) structure of the target protein as input, screens the protein sequence and predicts the impact of all possible single amino acid substitutions on the protein melting temperature in just a few minutes. Its performance makes it a perfect instrument to design proteins with improved thermal stability and to guide mutagenesis experiments that are usually expensive and time-consuming. Moreover, due to its speed, it can also be employed in large-scale investigations aimed at gaining insights into the relationship between protein thermal stability and natural evolution and into the adaptation mechanisms to extreme environmental conditions [8].

In the next sections, we briefly introduce the key ingredients and model structure of HoTMuSiC, show how it can be fruitfully applied for protein design applications, and discuss its performances in detail.

2 HoTMuSiC Key Instrument: The Statistical Potentials

The key instrument used in the construction of HoTMuSiC is the statistical potential formalism, known to be quite efficient when applied to a wide range of problems including protein structure prediction, protein design, and protein-ligand scoring.

These knowledge-based, effective potentials are derived from frequencies of sequence and structure elements in a nonredundant dataset of well-resolved protein 3D structures [9]. More precisely, let c be a structure element, consisting of the distance between two residues, the solvent accessibility of a residue, its backbone torsion angle, or combinations thereof, and let s be a sequence element consisting of the amino acid type of one or two residues. The free energy contribution of the association (c,s) is ruled by the Boltzmann law:

$$\Delta W(c,s) = -k_B\ T\ \mathrm{Log}\left[\frac{P(c,s)}{P(c)P(s)}\right] \quad (1)$$

where $P(c)$, $P(s)$, and $P(c,s)$ are the relative frequencies of c, s, and (c, s) in the dataset, k_B is the Boltzmann constant, and T is the absolute temperature.

In order to construct energy functions that describe the impact of the temperature on the amino acid interactions, we introduced temperature-dependent statistical potentials [10, 11]. They are extracted from datasets of proteins with known structures and specified thermal stability properties. The potentials derived from mesostable proteins, denoted as $\Delta W(c,s)^{meso}$, describe the interactions at low T, while those derived from thermostable or hyperthermostable proteins, $\Delta W(c,s)^{thermo}$, represent the interactions at high temperatures.

Using these new potentials, the T-dependence of several types of amino acid interactions were unraveled [10]. Moreover, these potentials were successfully applied to the prediction of the protein stability curve as a function of the temperature [11, 12].

The potentials that are used in HoTMuSiC's model are the standard potentials $\Delta W(c,s)$ and the T-dependent potentials $\Delta W(c, s)^{meso}$ and $\Delta W(c,s)^{thermo}$, for various structure and sequence elements (c,s).

3 HoTMuSiC Harmony: The Artificial Neural Networks

The statistical potentials were combined to predict how the melting temperature T_m changes upon amino acid substitution:

$$\Delta T_m = T_m(\text{mutant}) - T_m(\text{wild type}) \quad (2)$$

We set up two different models to compute ΔT_m. The HoTMuSiC model requires as sole input the 3D structure of the wild-type protein and is based on a linear combination of the standard statistical potentials $\Delta W(c,s)$ for different sequence and structure elements (c,s). The T_m-HoTMuSiC model requires in addition the melting temperature of the wild-type protein and employs a combination of standard and T-dependent potentials $\Delta W(c,s)$, $\Delta W(c, s)^{thermo}$, and $\Delta W(c,s)^{meso}$.

Both models include three additional terms—a constant term **1** and two terms $\Delta V_{\pm}$ that represent the difference in volume between the wild-type and mutant residues—and describe the creation of stress or holes in the protein structure.

All these terms are weighted by sigmoid functions of the solvent accessibility A, which smoothly connect the surface to the protein interior while allowing different behaviors in these regions.

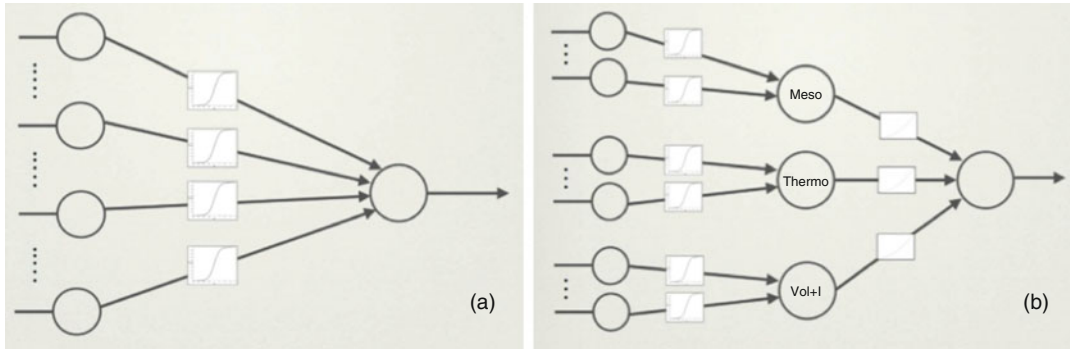

Fig. 1 Schematic representation of the ANNs used for the parameter identifications. (**a**) HoTMuSiC: two-layer ANN, consisting of a perceptron with sigmoid activation functions and 12 input neurons encoding 9 potentials, 2 volume terms and a constant term; (b) T_m-HoTMuSiC: three-layer ANN, consisting of 3 perceptrons with sigmoid weights; the first two perceptrons have 5 input neurons each, encoding 5 mesostable and 5 thermostable potentials, respectively; the third perceptron has 3 neurons for the volume and constant terms. The outputs of these three perceptrons (Meso, Thermo, and Vol + I) are the inputs of another perceptron with polynomial weight functions of the wild-type T_m value

Indeed, according to the type of potential, the mutational impact can be stronger at the surface or in the protein core.

Each sigmoid is a function of four parameters, which were identified using a cost function of the difference between experimental and computed ΔT_m values on a learning dataset. We used for that purpose the T1626 set [13] that contains 1626 mutations inserted in about 90 proteins of known X-ray structure of resolution below 2.5 Å, collected by literature screening.

For minimizing the cost function, artificial neural networks (ANN) were considered. For HoTMuSiC, a standard single-layer ANN was used (Fig. 1a), in which each neuron is associated with one input term ($\Delta W(c,s)$, ΔV, **1**) and the activation functions are sigmoid functions of A.

The ANN of T_m-HoTMuSiC is composed of three layers, where the additional, hidden, layer gets activated by functions of the T_m of the wild-type protein and confers more weight to mesostable or thermostable perceptrons according to the thermal properties of the wild-type protein (*see* Fig. 1b and [7] for details).

A standard back-propagation algorithm was employed in the training of the neural network. To test the predictor, we performed fivefold cross-validation using an early stopping technique to avoid overfitting.

4 HoTMuSiC Sound: The Results

4.1 Performances

To validate the method, we first applied the two HoTMuSiC models to the T1626 learning dataset using a fivefold cross-validation procedure and compared their performances with another ΔT_m predictor developed in the literature, i.e., AutoMute [14]. The results are reported in [7]. Here, in addition, we compared the HoTMuSiC performances with popular $\Delta\Delta G$ prediction methods, namely, PoPMuSiC [15], FoldX [16], and Rosetta [17]. Indeed, there is the tendency in the literature to use thermodynamic stability prediction methods to compute thermal stability changes even though the two quantities $\Delta\Delta G$ and ΔT_m are only partially correlated [13]. These comparisons give us an idea of the additional error that one makes by using $\Delta\Delta G$ predictors to predict ΔT_m. The results are given in Table 1.

To extend the performance analysis and get more robust results, we also did a manual literature search and used dedicated annotation servers such as ProTherm [18] and Brenda [19] to collect mutations that are not in T1626, i.e., newly characterized substitutions inserted in experimentally solved protein structures, and mutations inserted in structures that are not yet solved but were modeled by homology modeling. In this way, we obtained a second dataset, called T526, containing 526 mutations in 42 experimental and 58 modeled protein structures.

As seen in Table 1, T_m-HoTMuSiC is slightly more accurate than HoTMuSiC, which is normal as it relies on additional information, i.e., the experimental T_m of the wild type. Both HoTMuSiC models outperform the other predictors, especially on the T526 test set, which is not part of any of the predictors' learning sets. This could indicate that some of these predictors suffer from overfitting problems. Finally, the performance of all the tested $\Delta\Delta G$ predictors is significantly lower. We may thus conclude that using $\Delta\Delta G$ predictors to predict ΔT_m leads to a drop of the performance, of at least 0.1 in correlation coefficient.

4.2 Webserver

HoTMuSiC and T_m-HoTMuSiC are freely available for academic use on the webserver dezyme.com. The site contains an introductory and an explanatory page (**Software** and **Help**) and three working pages:

- **My Data** privately stores the user's personal files, containing protein structures in PDB [20] format or lists of mutations.
- **Query**: On the "HoT" query page, the user must choose either a PDB code that is automatically retrieved from the PDB data bank [20] or a private structure file stored in his My Data page. He must also choose among three options:

Table 1
Performances of the predictors on the T1626 [13] and T526 datasets as evaluated by the Pearson correlation coefficients *R* between the predictor output and the experimental ΔT_m values

Predictors	T1626	T526
	R	*R*
***T_m*-HoTMuSiC**	**0.61**	**0.59**
HoTMuSiC	0.59	0.56
AutoMute 2.0	0.54	0.22
PoPMuSiC v3.0	-0.50	-0.35
FoldX	-0.42	-0.27
Rosetta	-0.47	-0.26

The three predictors on a blue background are ΔT_m predictors and the three on a green background are $\Delta\Delta G$ predictors; the latter show negative correlations as, by convention, negative $\Delta\Delta G$ and positive ΔT_m values are stabilizing

1. *Systematic* mutation option in which all the possible amino acid substitutions for each residue in the protein sequence are computed.
2. Uploading a mutation *File* containing a list of single-site mutations.
3. Giving *Manual*ly a list of single-site mutations.

The last option gives the choice between HoTMuSiC and T_m-HoTMuSiC. When selecting the latter, the T_m of the wild type must be specified.

- The **Results** page contains all the predictions performed by the user, which can easily be downloaded. If the systematic mutation mode is chosen, the following files are available:
  - [pdbname].hot contains the predicted ΔT_m value for all possible single-site substitutions inserted in the protein of structure [pdbname] as well as other biophysical characteristics of the mutated residue, i.e., its solvent accessibility and its secondary structure.
  - [pdbname].hots contains information at the residue level: the mean ΔT_m of all substitutions at each position and the sum of all positive and of all negative ΔT_m values.
  - [pdbname].html shows a histogram picture in which the ΔT_m sum of all stabilizing mutations at each position is indicated. Figure 2 contains an example of this output.
  - [pdbname.zip] is an archive containing the three abovementioned files, ready to download.

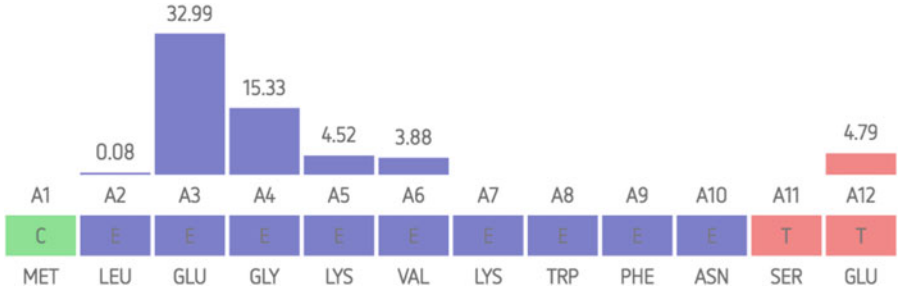

Fig. 2 Example of the webserver's histogram results for the systematic scanning of a target protein. The bars represent the sum of the positively predicted ΔT_m values at each sequence position; these sums are indicated above the bars (in °C). The largest bars represent residues whose mutations are likely to thermostabilize the protein and on which protein engineering experiments should focus first

4.3 Application to Modeled Structures

The quality of the input structures is of fundamental importance for stability predictions. Indeed, the more precise the structure in terms of resolution, the more accurate the predictions. This point is often overlooked when the performances of predictors are discussed. However, it will probably become even more important in the near future since, for example, a growing number of structures will be resolved with cryo-electron microscopy techniques, which usually have lower resolutions (at least for now) than those obtained with X-ray crystallography.

In the newly developed dataset T526, we have also included structures obtained via homology modeling using the Swiss-Model server [21], in view of analyzing the robustness of HoTMuSiC with respect to structural inaccuracies. We compared the performance on proteins for which we have a good-resolution X-ray structure or a modeled structure obtained with a template of which the sequence identity (SI) with respect to the target is either greater than 98% or between 22% and 98%.

As seen from Table 2, the proteins whose X-ray structure is available or can reliably be modeled from templates with SI $\geq$ 98% show good performances with a linear correlation coefficient of about 0.60, whereas the proteins that are modeled from templates SI < 98% (with a mean <SI> of 69%) have, as expected, lower but still reasonably good performances measured by a correlation coefficient $R \approx 0.45$.

Another important aspect that impacts the prediction accuracy is the experimental technique used to determine protein structures and the associated resolution. Clearly, the best results are obtained with X-ray structures of resolution of 2.0–2.5 Å at most, as expected from the above analysis on modeled structures.

These results show that HoTMuSiC can be reliably applied not only to good-resolution structures but also to low-resolution and modeled structures. This further broadens the field of applicability of the method.

Table 2
Performances of the predictors on the T526 datasets as evaluated by the Pearson correlation coefficients *R* between the predictor output and the experimental ΔT_m values. "≥98%" and "<98%" indicate the sequence identity with respect to the template used for homology modeling

Predictors	*R*		
	X-ray	≥98%	<98%
T_m-HoTMuSiC	0.59	0.62	0.45
HoTMuSiC	0.56	0.58	0.46

4.4 Protein Design with HoTMuSiC Pipeline

In order to optimize the thermal stability of a target protein to allow it, for example, to remain stable and active at temperatures higher than the physiological temperature, one has to select mutations that increase the protein melting temperature without affecting the protein function. We constructed a systematic pipeline for that purpose, which includes the HoTMuSiC tool but also the ConSurf software [22] that computes the evolutionary conservation score of each residue, as well as annotations from UniProt [23]. This pipeline aims at selecting combinations of point mutations likely to achieve a substantial increase of the protein thermoresistance without affecting the function of the target protein. In what follows, we describe step by step how this pipeline works, as illustrated in Fig. 3:

1. In a first stage, the impact of all possible point substitutions on the melting temperature of a target protein is evaluated with HoTMuSiC and/or T_m-HoTMuSiC in the systematic mode (*see* Subheading 4.2), using as input the (experimental or modeled) protein structure in PDB format. If multiple PDB structures of the target protein are available, HoTMuSiC is applied to all of them (with usually a cutoff of about 2.5 Å on the X-ray resolution), and the mean ΔT_m of each substitution is computed. The mutations are then classified according to their mean ΔT_m values.
2. The following step consists of running ConSurf [22] to estimate the evolutionary conservation score (from 0 to 9) of each amino acid of the target sequence aligned to its homologous sequences. At this level, we performed the first selection by imposing that the mutated positions have a conservation score of at most 7. This filters out mutant residues that are too conserved and thus probably too important for functional or structural reasons.
3. The next step is to retrieve from UniProt [23] the available annotations about catalytic residues and binding sites, and to

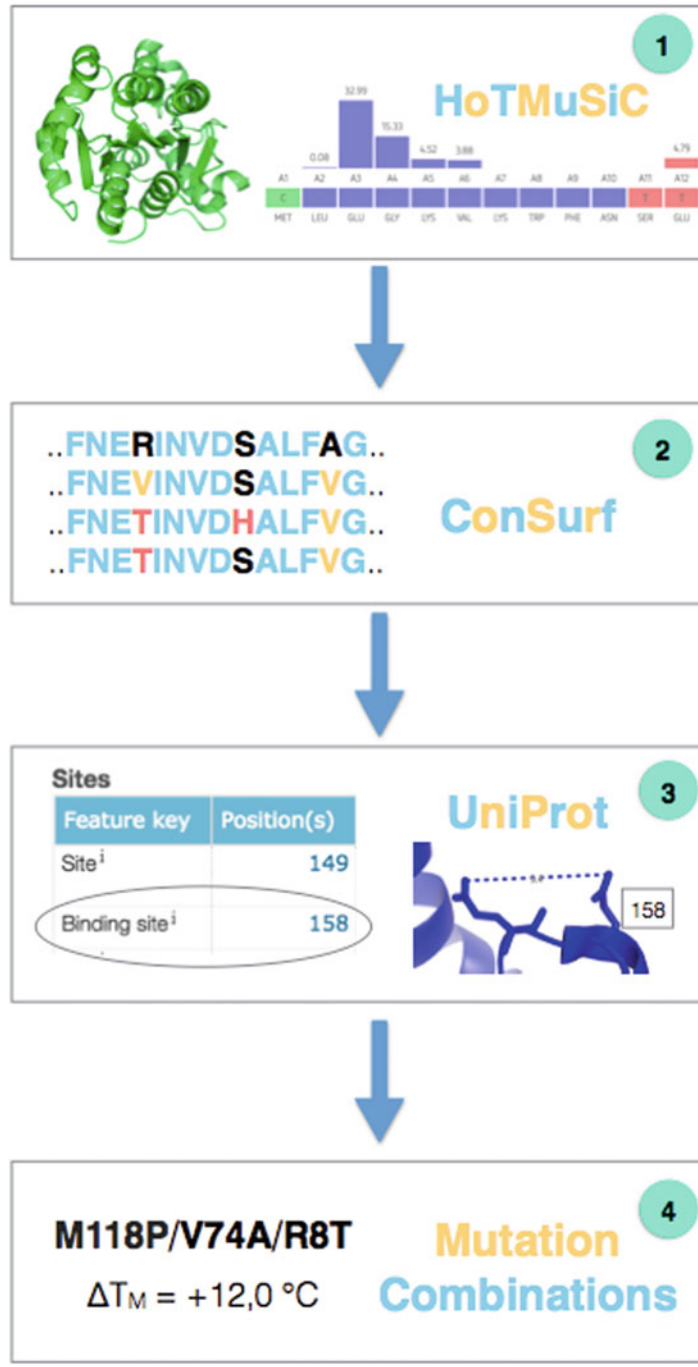

Fig. 3 Schematic picture of the HoTMuSiC-based pipeline for protein design

compute the distance between these sites and all the protein residues. All positions that are closer than 7 Å from one of the functional sites were overlooked. This ensures that the selected mutations do not touch or influence the catalytic activity or the binding to other biomolecules.

4. The remaining list thus contains point mutations that have a high chance to improve the thermoresistance of the target protein without affecting its function. Often, however, one cannot reach the required increase in melting temperature by just a single mutation. The strategy is then to combine several point mutations. For this purpose, we made the approximation that if two mutated sites are separated by a spatial distance of more than 10 Å, the stabilization effect is additive. This final criterion led us to select groups of two, three, or more mutations that optimize the target protein.

4.5 HoTMuSiC-Based Pipeline Applied to Rhizomucor Miehei Lipase

To illustrate how the HoTMuSiC pipeline can be used to optimize a protein, we applied it to thermally stabilize the lipase from *Rhizomucor miehei* (RML). This enzyme catalyzes a wide range of reactions such as the hydrolysis of oil, the esterification of fatty acids, and the transesterification and alcoholysis of glycerides. Therefore, it received a lot of attention, and different studies tried to improve its thermal stability to increase the efficiency of these biocatalytic reactions [24].

There are several available X-ray structures of the enzyme, and we considered here the two structures with PDB code 3TGL and 4TGL, which have a resolution of 1.9 and 2.6 Å, respectively. RML folds into a ß-sheet surrounded by helices (Rossmann fold) and has a melting temperature of 58.7 °C [25]. It contains a Ser-His-Asp trypsin-like catalytic triad, in which the active serine is buried under a short helical lid that undergoes conformational changes (*see* Fig. 4a). By exposing or protecting the catalytic pocket, the lid movement controls the enzymatic activity [25].

The first step to predict thermostabilizing mutations consists in running HoTMuSiC in the systematic mode on the two PDB structures of RML to compute the ΔT_m of all possible point mutations. These values, stored in the 3TGL.hot and 4TGL.hot files on the Results page, were first averaged for each individual mutation and then ranked decreasingly according to their average ΔT_m value.

The top 15 mutations in the list are given in Table 3. Note that glycine and proline substitutions were excluded from this list since their mutations are likely to induce changes in conformation which are not taken into account in the prediction model.

In the next steps, the residue conservation across homologous sequences was evaluated using an in-house implementation of the ConSurf algorithm, and the spatial distance between the mutated residues and the catalytic triad Ser-His-Asp was computed. The mutations inserted at highly conserved positions (ConSurf

Table 3
The 15 point mutations in RML that are predicted as the most thermostabilizing by HoTMuSiC

Mutations	ΔT_m^{pred} (°C)	ΔT_m^{exp} (°C)	ConSurf	Distance (Å)
E221V	3.2	-	9	6.3
E221I	2.8	-	9	6.3
Q174F	2.5	-	6	5.3
Q174Y	2.3	-	6	5.3
E230F	2.0	1.6	3	8.5
E230I	2.0	5.7	3	8.5
E230Y	1.8	2.1	3	8.5
E230V	1.8	5.4	3	8.5
Q176F	1.8	-	8	5.2
E221L	1.8	-	9	6.3
Q174W	1.8	-	4	5.3
Q176Y	1.8	-	8	5.2
E230L	1.7	5.4	3	8.5
A64I	1.7	-	8	8.0
E230W	1.7	2.7	3	8.5

The predicted and experimental [25] ΔT_m values are reported in columns 2 and 3. The ConSurf conservation scores and the spatial distance from the catalytic residues, computed between average side chain centroids, are shown in columns 4 and 5. The substitutions that are dropped on the basis of their conservation scores and of their distance to the catalytic triad are on a yellow and green background, respectively

score ≥ 8) or that are too close to the catalytic site (distance$\leq$7 Å) were then removed, as they could impact the protein function or other biophysical characteristics that we do not want to modify.

The substitutions filtered out due to their high conservation scores or their proximity to the catalytic triad are indicated in Table 3. Among the 15 top mutations predicted in **step 1**, we kept only the 6 mutations of Glu at position 230.

To get a larger number of candidate mutations, we have to relax the first, ΔT_m-based, criterion. In this way we obtained the 15 most stabilizing mutations that satisfy all our selection filters, shown in Table 4.

As a final step of our pipeline, we determined subsets of mutations of residues whose relative distances are equal to 10 Å or more and are thus assumed to be independent (Fig. 4). For example, this procedure yields the sets of multiple mutations:

- E230I/S103W/K53I
- E230F/R68V/K53F
- E230M/K106I/N63V

Table 4
Final list of proposed mutations in RML, which satisfy all our selection criteria

Mutations	ΔT_m^{pred} (°C)	ΔT_m^{exp} (°C)	ConSurf	Dist (Å)
E230F	2.0	1.6	3	8.5
E230I	2.0	5.7	3	8.5
E230Y	1.8	2.1	3	8.5
E230V	1.8	5.4	3	8.5
E230L	1.7	5.4	3	8.5
E230W	1.7	2.7	3	8.5
S103W	1.7	–	3	8.5
R68V	1.6	–	6	17.0
E230M	1.6	1.3	3	8.5
K53I	1.5	–	2	13.3
K106I	1.5	–	4	13.2
K53F	1.4	–	2	13.3
K106F	1.3	–	4	13.2
N63V	1.3	–	6	7.4
N63I	1.3	–	6	7.4

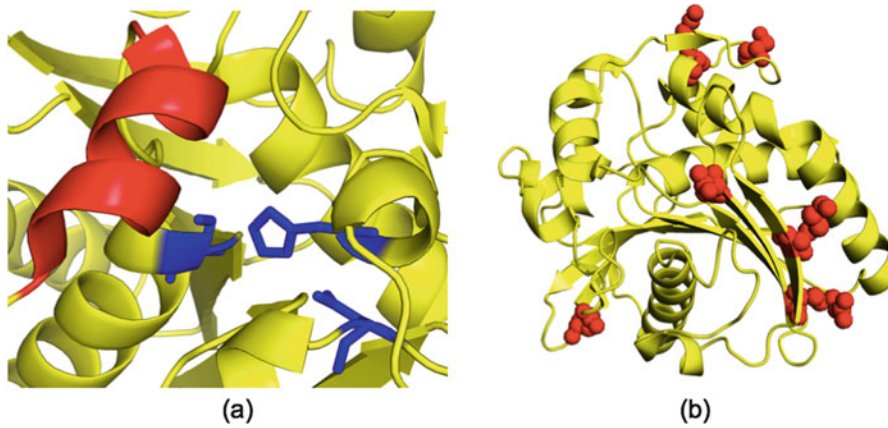

Fig. 4 Three-dimensional structure of *Rhizomucor miehei* lipase (PDB code 3TGL). (**a**) Zoom on the Ser-His-Asp catalytic triad (in blue sticks) with the helical lid (in red) that undergoes conformational changes and modulate the protein activity. (**b**) Residues to be mutated (in red spheres) for the thermostabilization of the lipase

Finally, we compared our computational predictions on lipase with the available experimental data reported in [25]. The prediction score of HoTMuSiC on this set of 36 point mutations was found to outperform the competitor methods reported in Table 1 and to reach quite a good accuracy evaluated by a root mean square deviation between predicted and experimental ΔT_m values of 1.7 °C and by a linear correlation coefficient of 0.6. This score is the same as the one obtained from the other tests shown in Tables 1 and 2.

Moreover, four sets of multiple mutations have been experimentally shown to lead to a substantial increase of the RML thermoresistance [25]. If we assume that the effect of mutations with a sufficient spatial separation is additive, these multiple mutations are also very well predicted by HoTMuSiC. The final score of our pipeline, when the 4 multiple mutations are added to the 36 point mutations, reaches a linear correlation coefficient higher than 0.8.

Finally note that the predicted ΔT_m values tend to be lower than the experimental values. Indeed, as we already noted [26], the training datasets are dominated by destabilizing mutations and this induces biases toward these types of mutations and leads to an underestimation of the predicted stabilization effects.

5 Conclusion

In this chapter, we presented recently developed protein thermal stability predictors, and their application to efficiently optimize targeted proteins. We focused more specifically on the HoTMuSiC predictor, which predicts the ΔT_m values of all the possible point mutations in a medium-size protein in a few minutes, on the basis of its experimental or modeled 3D structure. Our tool outperforms the other available ΔT_m predictors, as well as $\Delta\Delta G$ predictors when applied to thermal stability even though they are designed for thermodynamic stability.

The fastness of HoTMuSiC allows scanning and predicting all possible substitutions inserted in a target protein. It can thus guide mutagenesis experiments aimed to improve the thermoresistance of proteins and be employed in the optimization of a wide series of biotechnological processes.

To optimize the selection of mutations that need to be tested experimentally, we complemented the HoTMuSiC results with additional information on the residue conservation among homologous proteins and the distance from annotated functional sites. This novel pipeline is designed to filter out mutations that are likely to affect functionally or structurally important residues. The point mutations that satisfy the criteria are then combined into subsets of non-interacting mutations that are assumed to be independent.

These subsets are identified to strongly enhance the effect on thermal stability.

The HoTMuSiC pipeline was applied to the thermal stabilization of lipase from *Rhizomucor miehei*. Several series of multiple mutations were predicted for this enzyme. For those mutations whose ΔT_m values were measured experimentally, the prediction score was shown to be high.

Finally note that the potentiality of HoTMuSiC is not restricted to the protein design field. Due to its speed, it can also be applied on a proteomic scale to gain important theoretical insights into the thermal and evolutionary adaptation of proteins to extreme environments.

Acknowledgments

We thank Raphael Bourgeas for the help in implementing HoTMuSiC on the webserver and César Ngabo for his contribution to the HoTMuSiC pipeline. We acknowledge support from the Fund for Scientific Research (FNRS) through a PDR research project. F.P. and M.R. are FNRS postdoctoral researcher and research director, respectively.

References

1. Pucci F, Rooman M (2017) Physical and molecular bases of protein thermal stability and cold adaptation. Curr Opin Struct Biol 42:117–128
2. Kumar S, Tsai CJ, Nussinov R (2000) Factors enhancing protein thermostability. Protein Eng 13:179–191
3. Razvi A, Scholtz JM (2006) Lessons in stability from thermophilic proteins. Protein Sci 15:1569–1578
4. Fowler DM, Fields S (2014) Deep mutational scanning: a new style of protein science. Nat Methods 11:801–807
5. Jäckel C, Kast P, Hilvert D (2008) Protein design by directed evolution. Annu Rev Biophys 37:153–173
6. Chica R, Doucet N, Pelletier JN (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol 16:378–384
7. Pucci F, Bourgeas R, Rooman M (2016) Predicting protein thermal stability changes upon point mutations using statistical potentials: introducing HoTMuSiC. Sci Rep 6:23257
8. Pucci F, Rooman M (2016) Improved insights into protein thermal stability: from the molecular to the structurome scale. Philos Trans A Math Phys Eng Sci 374:2080
9. Dehouck Y, Gilis D, Rooman M (2006) A new generation of statistical potentials for proteins. Biophys J 90:4010–4017
10. Folch B, Dehouck Y, Rooman M (2010) Thermo- and mesostabilizing protein interactions identified by temperature-dependent statistical potentials. Biophys J 98:667–677
11. Pucci F, Dhanani M, Dehouck Y, Rooman M (2014) Protein thermostability prediction within homologous families using temperature-dependent statistical potentials. PLoS One 9:e91659
12. Pucci F, Kwasigroch JM, Rooman M (2017) SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. Bioinformatics 33:3415–3422
13. Pucci F, Bourgeas R, Rooman M (2016) High-quality thermodynamic data on the stability changes of proteins upon single-site mutations. J Phys Chem Ref Data 45:023104
14. Masso M, Vaismann II (2010) AUTO-MUTE: web-based tools for predicting stability changes in proteins due to single amino acid replacements. Protein Eng Des Sel 23:683–687

15. Dehouck Y, Grosflis A, Folch B, Gilis D, Bogaerts P, Rooman M (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. Bioinformatics 25:2537–2543
16. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320:369–387
17. Kellogg EH, Leaver-Fay A, Baker D (2010) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins 79:830–838
18. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Res 32:D120–D121
19. Placzek S (2017) BRENDA in 2017: new perspectives and new tools in BRENDA. Nucleic Acids Res 45:D380–D388
20. Berman HM et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
21. Waterhouse A et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303
22. Ashkenazy H, Abadi S, Martz E, Chav O, Mavrose I, Pupko T, Ben-Tai N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 44: W344–W350
23. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169
24. Rodrigues RC, Fernandez-Lafuente R (2010) Lipase from Rhizomucor miehei as a biocatalyst in fats and oils modification. J Mol Catal B Enzym 66:15–32
25. Li G, Fang X, Su F, Chen Y, Xu L, Yan Y (2018) Enhancing the thermostability of Rhizomucor miehei lipase with a limited screening library by rational-design point mutations and disulfide bonds. Appl Environ Microbiol 84: e02129–e02117
26. Pucci F, Bernaerts KV, Kwasigroch JM, Rooman M (2018) Quantification of biases in predictions of protein stability changes upon mutations. Bioinformatics 34:3659–3665