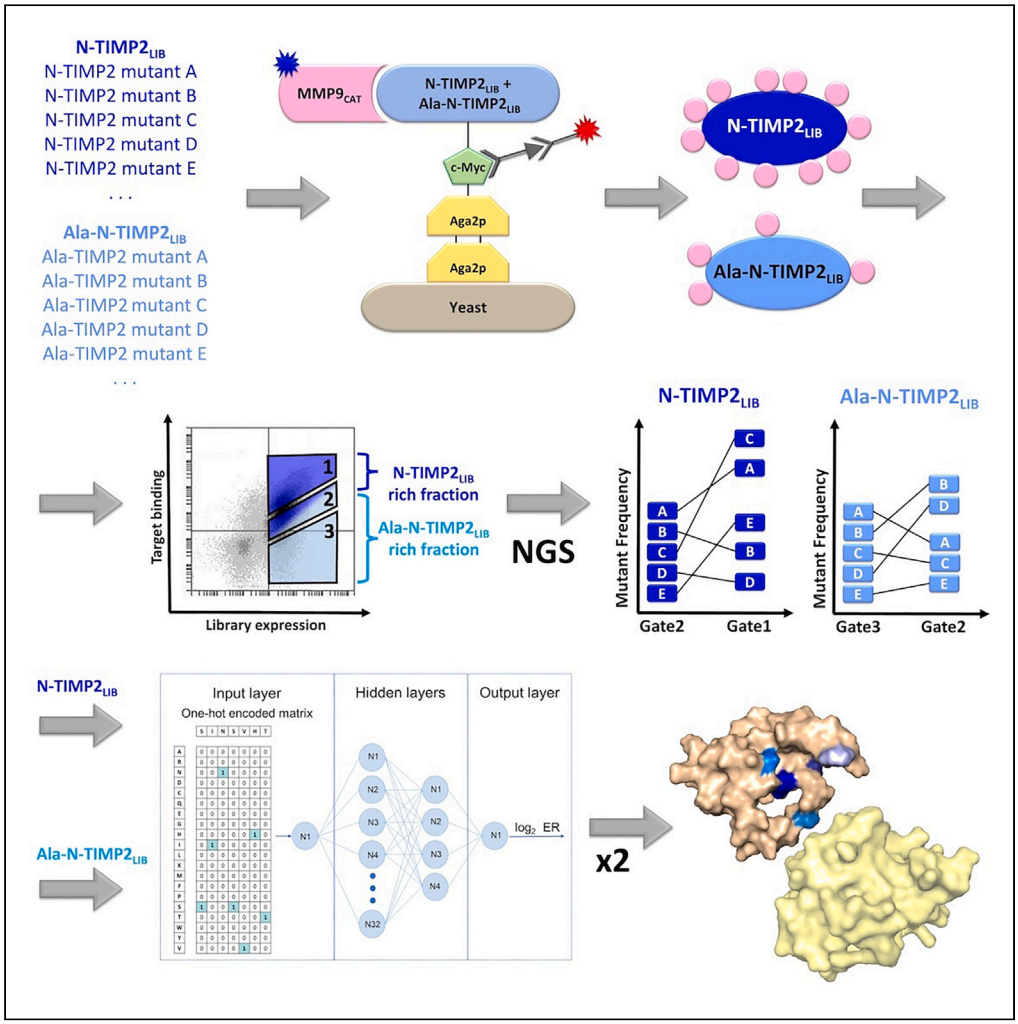
iScience

CellPress
OPEN ACCESS

Article

Deep neural networks for predicting the affinity landscape of protein-protein interactions



Reut Meiri, Shay-Lee Aharoni Lotati, Yaron Orenstein, Niv Papo

yaron.orenstein@biu.ac.il (Y.O.)
papo@bgu.ac.il (N.P.)

Highlights

An approach to map the binding specificity landscape over a broad affinity range

Models trained on a protein complex predict absolute affinities with high accuracy

Interpretability reveals cold and hot spots and suggests experimental improvements

Increased variability vs. read depth within a library improves affinity predictions

Meiri et al., iScience 27, 110772
September 20, 2024 © 2024 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.110772

iScience

CellPress
OPEN ACCESS

Article

Deep neural networks for predicting the affinity landscape of protein-protein interactions

Reut Meiri,[1,5] Shay-Lee Aharoni Lotati,[2,5] Yaron Orenstein,[3,4,*] and Niv Papo[2,6,*]

SUMMARY

Studies determining protein-protein interactions (PPIs) by deep mutational scanning have focused mainly on a narrow range of affinities within complexes and thus include only partial coverage of the mutation space of given proteins. By inserting an affinity-reducing N-terminal alanine in the N-terminal domain of the tissue inhibitor of metalloproteinases-2 (N-TIMP2), we overcame the limitation of its narrow affinity range for matrix metalloproteinase 9 ($MMP9_{CAT}$). We trained deep neural networks (DNNs) to quantitatively predict the binding affinity of unobserved wild-type variants and variants carrying an N-terminal alanine. Good correlation was obtained between predicted and observed log_2 enrichment ratio (ER) values, which also correlated with the affinity of N-TIMP2 variants to $MMP9_{CAT}$. Our ability to predict affinities of unobserved N-TIMP2 variants was confirmed on an independent dataset of experimentally validated N-TIMP2 proteins. This ability is of significant importance in the field of PPI prediction and for developing therapies targeting these interactions.

INTRODUCTION

Protein-protein interactions (PPIs) play essential roles in diverse biological processes. Understanding how mutations in a protein's sequence modulate its interactions is important for decoding the evolution of protein-binding interfaces and for developing therapies targeting these interfaces.[1,2] On the basis of their contributions to PPIs, interfacial residues can be categorized as either hot spots or cold spots, where hot spots are positions where the residues make a major contribution to the protein-protein binding free energy and cold spots are positions occupied by amino acids that make suboptimal contributions. Thus, mutations in hot-spot positions may lead to a dramatic decrease in the binding affinity of a specific PPI (e.g., Cukuroglu et al.[3]) and mutations in cold spots may enhance binding affinity (e.g., Shirian et al.[4]).

Numerous experimental mutagenesis approaches have been used to map protein-binding interfaces, particularly those measuring affinity changes in purified variants with site-specific single mutations to alanine[5] or those identifying binding epitopes by using affinity screens of yeast-surface-displayed (YSD) protein libraries.[6] However, both these approaches enable the analysis of only a limited number of positions and mutations.[7,8] Thus, to predict the affinity and selectivity of protein variants more accurately and on a much larger scale[9] and to generate more comprehensive affinity and specificity PPI landscapes,[10,11] high-throughput affinity screens of protein libraries combined with deep mutational scanning (DMS) have been developed and implemented.

Nonetheless, despite the promise of high-throughput approaches, they, too, have some limitations. First, DMS studies have focused mainly on variants spanning a narrow range of affinities, which curtails the characterization of interactions within protein complexes having wide binding landscapes (i.e., a broad range of affinities). This constraint on the range of affinities results from limitations of scale (library size) and also from limitations in the signal detection abilities of the experimental techniques designed to probe the PPIs. Recently, however, the use of multiple affinity sorting gates for screening protein libraries has enabled the discrimination between variants with similar affinities by broadening the range of target affinities that can be detected and improving the accuracy of affinity measurements for individual variants.[12–14] Second, DMS approaches are limited in their ability to experimentally map the binding affinity landscape of high-affinity complexes. Thus, to accurately evaluate variants with high affinity for their target protein, a low target concentration—which may be difficult to follow—is used. In contrast, high target concentrations may be appropriate for low-affinity but not for high-affinity variants (since they may result in saturated binding signals). Therefore, a complete picture of the binding affinity landscape, i.e., identification of both affinity-reducing and affinity-enhancing mutations, cannot be achieved in a single experimental affinity screen. Third, DMS enables only partial coverage of the mutation space of a given protein. Since the search-space size increases exponentially with the number of tested residue positions in the amino acid sequence of the protein, there is a combinatorial explosion in the number of possible protein variants that would have to be measured

[1]School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[2]Avram and Stella Goldstein-Goren Department of Biotechnology Engineering and the National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[3]Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel
[4]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel
[5]These authors contributed equally
[6]Lead contact
*Correspondence: yaron.orenstein@biu.ac.il (Y.O.), papo@bgu.ac.il (N.P.)
https://doi.org/10.1016/j.isci.2024.110772

experimentally.[15] This issue of partial coverage therefore constitutes a significant obstacle to progress in obtaining a comprehensive picture of the mutational landscape. Many investigations have thus focused solely on mutations in a single nucleotide, resulting in a significant reduction (approximately 70%) in the potential coverage of all feasible single amino acid substitutions. For example, the study of Livesey and Marsh collected 31 datasets from DMS experiments, but the majority of those datasets demonstrated only partial coverage of the mutation space.[16]

As indicated earlier, full coverage of the mutation space is required to produce an accurate prediction of the affinities of variants spanning a wide range of binding affinities. The need for full coverage is especially critical when only low-throughput affinity screens are feasible. Thus, to impute this mutation space, a variety of computational methods have been developed to predict binding affinity based on the protein's amino acid sequence. These include four major approaches—each with its advantages and limitations—to produce predictive models for protein properties (e.g., binding affinity): calculation-based empirical models, meta-predictors that integrate multiple model predictions as input features, unsupervised models based on evolutionary information (that does not include any experimental measurements), and supervised models based on experimental (assay-derived) datasets of protein variants.[16,17] While unsupervised models can accurately predict the effect of mutations on the inherent protein function, they fail to predict protein properties that have not been subjected to long-term evolutionary selection.[18] In a recent benchmark study, supervised methods emerged as the most successful in predicting protein functions, which was demonstrated by their consistent ranking in the average-to-high range and their attainment of top correlation scores in 20 out of 31 datasets.[16]

Supervised methods have advanced significantly thanks to the rise of deep neural networks (DNNs), with convolutional neural networks and graph neural networks being the most successful architectures.[19] Envision, a stochastic gradient-boosting learning algorithm of single-mutation effects, was trained on nine DMS mutagenesis datasets. Envision used 27 biological, structural, and physicochemical features for each protein variant to predict the variant's molecular effect. However, Envision's predictions for variants with intermediate effects (e.g., moderate binding affinities) are poor, and its performance deteriorates when structural and evolutionary features are missing.[20] Following Envision, Song et al.[21] pointed out that the lack of negative examples, i.e., variants with decreased functionality, in most DMS datasets makes it difficult to directly estimate the impact of sequence on function. To tackle this problem, they developed a statistical sequence-function model that showed good predictive performance on 10 DMS datasets of protein binding, folding, and enzymatic activity. However, when trying to model epistatic effects, the model suffered from long runtimes when its complexity was increased. Gelman et al.[18] benchmarked various neural-network architectures (fully connected, sequence convolutional, and graph convolutional) trained on various DMS datasets. They tested the limitations and dependence of the model's performance on the number of training examples and the quality of the experimental measurements. They suggested that it might be preferable to limit the overall quantity of unique variants examined as a means of guaranteeing that each variant receives a sufficient number of sequencing reads for accurate functional score calculations.

Informed by the background described earlier, we sought to develop DNN to comprehensively map PPIs and to accurately predict affinity-enhancing or affinity-reducing mutations, not necessarily observed in library affinity sorts. To develop the ability to predict both observed and unobserved variants over a broad range of affinities in tightly bound protein complexes, we took as a case study the complex of matrix metalloproteinase (MMP)9 with the N-terminal domain of the tissue inhibitor of metalloproteinases-2 (N-TIMP2). N-TIMP2 is the inhibitory domain of TIMP2 that exerts its action by binding to the active site of MMPs. The wild-type (WT) form, N-TIMP2$_{WT}$, exhibits sub-nanomolar affinity for the MMP9 catalytic domain (MMP9$_{CAT}$) (inhibition constant [K_i] value of 108.9 pM). The active site of MMP9, like that of all MMPs, comprises a catalytic zinc ion chelated to three histidine residues (His-401, His-405, and His-411 in MMP9) and a catalytic glutamate residue (Glu-402 in MMP9).[22] The N-TIMP2–MMP9 interface was previously shown to be highly tolerant to N-TIMP2 residue substitution or incorporation of additional amino acids without impairing stability,[23] making N-TIMP2 an ideal candidate for mapping binding-affinity landscapes of PPIs. Importantly, it has been shown that TIMP2 can inhibit MMP activity only when the N terminus of the Cys1 residue of the N-TIMP2 domain is free.[24] For example, an alanine residue insertion followed by Cys1 disrupts secondary interactions that putatively stabilize the interaction of N-TIMP2 with the MMP active site.[24,25] In this study, we leveraged this alanine insertion to generate an N-TIMP2 library comprising pairs of variants, one with low affinity to MMP and the other with high affinity, thereby allowing us to expand the range of affinities identified in our affinity library screens and to overcome experimental limitations in detecting a broad affinity range in the N-TIMP2–MMP9 complex.

The research strategy for this study thus comprised the following steps (Figure 1). (1) Two N-TIMP2 mutagenesis libraries were generated, one having variants with high affinities toward MMP9 (designated N-TIMP2$_{LIB}$) and the other having variants with low affinities toward MMP9 by virtue of an insertion of alanine at the N terminus of N-TIMP2 (designated Ala-N-TIMP2$_{LIB}$). (2) The two libraries were mixed in a 1:1 ratio (designated N-TIMP2$_{MIX}$), YSD, and subjected to fractional sorting by fluorescence-activated cell sorting (FACS) for binding to the MMP9 catalytic domain (MMP9$_{CAT}$). (3) The sorted library fractions were subjected to high-throughput sequencing (HTS) to determine the enrichment ratio (ER) of each individual variant. (4) Models were trained to predict the ER and hence to enable accurate and quantitative prediction of the impact on the binding affinity of N-TIMP2 variants that were not included or not observed in the libraries, and also to enable mapping of the binding-affinity landscape of the N-TIMP2/MMP9$_{CAT}$ complex. (5) Finally, the predictions were validated experimentally by determining the MMP9$_{CAT}$ inhibition potency of N-TIMP2 variants having a large range of affinities for MMP9$_{CAT}$.

RESULTS

Sorting the mixture of N-TIMP2$_{LIB}$ and Ala-TIMP2$_{LIB}$ libraries that bind to MMP9$_{CAT}$

To comprehensively and accurately determine, using a single sort, the K_i values of N-TIMP2 variants spanning a large range of affinities for the catalytic domain of MMP9 (i.e., MMP9$_{CAT}$), we generated—and then mixed together—two coding-sequence libraries. We first generated the high-affinity N-TIMP2$_{LIB}$ library with random single and multiple mutations at seven residue positions (i.e., 4, 35, 38, 68, 71, 97, and 99,

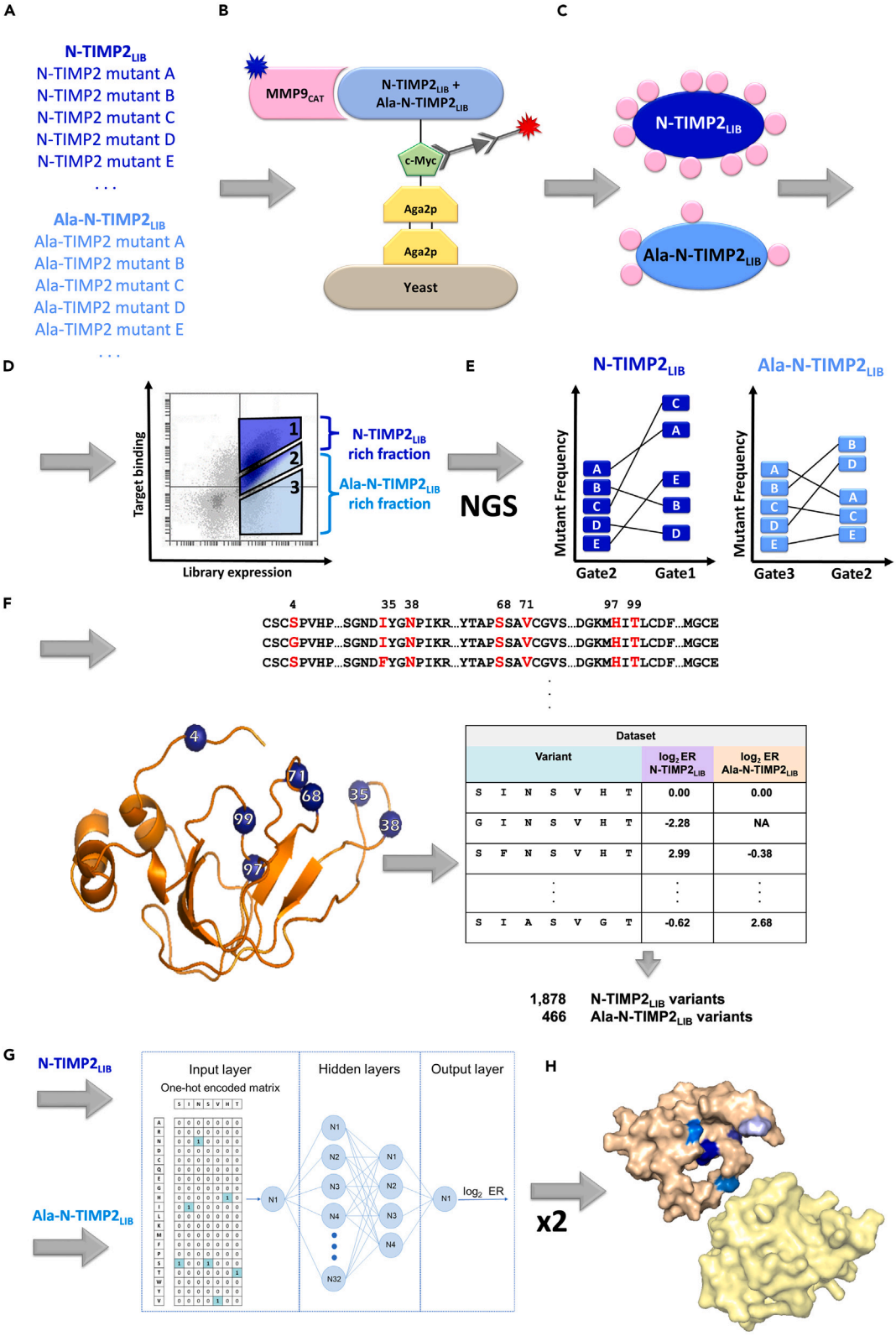

A
N-TIMP2LIB
N-TIMP2 mutant A
N-TIMP2 mutant B
N-TIMP2 mutant C
N-TIMP2 mutant D
N-TIMP2 mutant E
. . .
Ala-N-TIMP2LIB
Ala-TIMP2 mutant A
Ala-TIMP2 mutant B
Ala-TIMP2 mutant C
Ala-TIMP2 mutant D
Ala-TIMP2 mutant E
. . .
B
MMP9CAT
N-TIMP2LIB + Ala-N-TIMP2LIB
c-Myc
Aga2p
Aga2p
Yeast
C
N-TIMP2LIB
Ala-N-TIMP2LIB
D
Target binding
Library expression
1
2
3
N-TIMP2LIB rich fraction
Ala-N-TIMP2LIB rich fraction
NGS
E
N-TIMP2LIB
Ala-N-TIMP2LIB
Mutant Frequency
Gate2
Gate1
Gate3
Gate2
F
4 35 38 68 71 97 99
CSCSPVHP...SGNDIYGNPIKR...YTAPSSAVCGVS...DGKMHITLCDF...MGCE
CSCGPVHP...SGNDIYGNPIKR...YTAPSSAVCGVS...DGKMHITLCDF...MGCE
CSCSPVHP...SGNDFYGNPIKR...YTAPSSAVCGVS...DGKMHITLCDF...MGCE
Dataset
Variant
log2 ER N-TIMP2LIB
log2 ER Ala-N-TIMP2LIB
S I N S V H T 0.00 0.00
G I N S V H T -2.28 NA
S F N S V H T 2.99 -0.38
S I A S V G T -0.62 2.68
1,878 N-TIMP2LIB variants
466 Ala-N-TIMP2LIB variants
G
N-TIMP2LIB
Ala-N-TIMP2LIB
Input layer
One-hot encoded matrix
Hidden layers
Output layer
log2 ER
H
x2

Figure 1. Research strategy for identifying affinity-enhancing and affinity-reducing mutations

(A) Two N-TIMP2-derived libraries, designated N-TIMP2$_{LIB}$ (high affinity for MMP9) and Ala-N-TIMP2$_{LIB}$ (containing an Alanine insertion at the N terminus of the protein; low affinity for MMP9), were generated.

(B and C) The two libraries were mixed, and the mixed library was yeast surface displayed and subjected to flow cytometry screening for affinity to MMP9$_{CAT}$, the catalytic domain of MMP9, with (C) variants of N-TIMP2$_{LIB}$ having higher affinities for MMP9$_{CAT}$ than those derived from Ala-N-TIMP2$_{LIB}$.

(D) The library was sorted using three gates, each having different affinities to MMP9$_{CAT}$. The fractions were then analyzed by next generation sequencing (NGS, also termed HTS). Based on the scattering of the N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$ variants and that of unmodified N-TIMP2 and Ala-N-TIMP2 proteins, respectively, we determined which gate to use for the analysis of each library. To find affinity-enhancing mutations in N-TIMP2$_{LIB}$, we used Gate 2 and Gate 1, which had N-TIMP2-like affinity and higher affinity than N-TIMP2, respectively. Similarly, we used Gate 3, which had Ala-N-TIMP2-like affinity, and Gate 2, which conferred higher affinity compared to Ala-N-TIMP2, to identify affinity-enhancing mutations originating from Ala-N-TIMP2$_{LIB}$.

(E) Computational analysis of the NGS results was followed by calculation of the frequency of each variant in each of the three gates. The enrichment ratio (ER) was calculated separately for N-TIMP2$_{LIB}$ between Gate 1 and Gate 2, and for Ala-N-TIMP2$_{LIB}$ between Gate 2 and Gate 3.

(F) We generated a sequence–$\log_2$ ER dataset to train a DNN to predict the $\log_2$ ER of protein variants.

(G) Based on the $\log_2$ ER values for each of the two libraries (i.e., N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$), a neural network was trained to accurately and quantitatively predict the impact of unobserved potentially interacting mutations on the binding affinity. A matrix of one-hot encoded seven binding residues is given as input to a cascade of two fully connected layers with a rectified linear unit (ReLU) activation function, and finally through a single neuron with linear activation. The output is the predicted $\log_2$ ER of each variant.

(H) By combining the aforementioned two trained models, we revealed a variety of mutations at the interface of N-TIMP2 that modulate the binding affinity in the high-affinity complex, N-TIMP2/MMP9; these mutations would not have been discovered unless the two models were used together.

based on the N-TIMP2$_{WT}$ PDB: 1BUV) located in the N-TIMP2/MMP9$_{CAT}$ binding interface; all positions lie within 4 Å of the MMP structure,[1] and six of them (namely 35 & 38, 68 & 71, and 97 & 99) are coupled in pairs as a result of their close proximity (<5.7 Å) to one another.[26] All seven positions were previously shown to be highly tolerant to mutagenesis.[1,23] We then generated the low-affinity Ala-N-TIMP2$_{LIB}$ library in which an alanine residue was added at the N terminus of each N-TIMP2 variant (Figure 2A). The need to generate two separate libraries derived from the intrinsic high affinity of N-TIMP2$_{LIB}$ to its target, MMP9$_{CAT}$, which makes it difficult to identify in the same experiment mutations that increase or decrease the affinity of N-TIMP2$_{LIB}$ to the target. The low-affinity library, Ala-N-TIMP2$_{LIB}$, allowed us to screen against a concentration of the target that is of the same order of magnitude as the K_i of the N-TIMP2/MMP9$_{CAT}$ complex. To obtain a broad range of affinities, we combined the N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$ to yield the N-TIMP2$_{MIX}$ library, with a total size of 4.3×10^5 variants, which is three orders of magnitude larger than the theoretical diversity that could be obtained from random mutagenesis at single positions.

We then used a YSD platform to sort N-TIMP2$_{MIX}$ library variants according to their affinity to MMP9$_{CAT}$. Using the YSD pCHA plasmid, we displayed N-TIMP2$_{MIX}$ library variants on the *Saccharomyces cerevisiae* yeast surface as fusion proteins with the Aga2p/Aga1p system. This display allows the N terminus of the variants, which lies in the binding interface and is thus crucial for binding to MMP9, to be freely exposed to the solvent and available for binding to MMP9$_{CAT}$ (Figure 2B). Thereafter, by using flow cytometry of N-TIMP2$_{MIX}$, we conducted a preliminary sort (designated Pre-sort) to exclude variants with low expression and very low affinity to MMP9$_{CAT}$ (which probably indicates protein misfolding); the Pre-sort provided an N-TIMP2$_{MIX}$ library with variants having a wide range of affinities to MMP9$_{CAT}$ (Figures S1A and S1B).

We then subjected the N-TIMP2$_{MIX}$ library to fractional sorting to collect three sub-populations with different affinities to MMP9$_{CAT}$; to this end, three diagonal sorting gates were applied according to the binding and expression signal distributions of the different variants in the N-TIMP2$_{MIX}$ (Figure 2C). We chose the MMP9$_{CAT}$ concentration (100 nM) for the fractional sorting to be in the range between the *equilibrium* binding constants (K_i values) for binding of N-TIMP2 and Ala-N-TIMP2 to MMP9$_{CAT}$ (Table 1), such that maximal scattering of the affinity signal of the library members was obtained. We discriminated between three sub-populations with different binding affinities to MM9$_{CAT}$, i.e., higher affinity than N-TIMP2$_{WT}$ (Gate 1), WT-like affinity (Gate 2), and lower affinity than N-TIMP2$_{WT}$, which was also comparable to the Ala-N-TIMP2 affinity (Gate 3) (Figure 2C). We determined the WT-like affinity gate according to the flow cytometry signal that was produced by N-TIMP2$_{WT}$ binding to MMP9$_{CAT}$ under the same conditions, i.e., MMP9$_{CAT}$ at a concentration of 100 nM (Figure S1A). An additional flow cytometry analysis of the fractionated sub-populations was used to verify the differences in the binding signal between the three gates (Figure S2). It was subsequently found that each variant in the sorted N-TIMP2$_{MIX}$ library was distributed differently in the population in terms of affinity to MMP9$_{CAT}$, which means that each variant was enriched in a specific gate and was depleted in the others.

Comprehensive HTS analysis of N-TIMP2 interface mutations

To determine the affinity for MMP9$_{CAT}$ of each variant in N-TIMP2$_{MIX}$ and hence to map the N-TIMP2/MMP9$_{CAT}$ binding interface, we extracted the DNA plasmid of the pre-sorted N-TIMP2$_{MIX}$ library and the three sorted N-TIMP2$_{MIX}$ fractions and performed NGS analysis of the libraries using Illumina MiSeq. Most (~98%) of the total sequenced read pairs (381,481, 394,278, 662,060, and 459,124 reads for Gate 1, Gate 2, Gate 3, and Pre-sort, respectively) passed our quality filtering and merging processes. Thereafter, we translated the sequences into their respective amino acid sequences by using the N-TIMP2$_{WT}$ sequence (PDB: 1BUV) as a reference. We filtered out short reads and reads with stop codons: thus, for N-TIMP2$_{LIB}$, 335,834, 329,673, 46,948, and 149,225 of merged single sequences translated into valid amino acid sequences in Gate 1, Gate 2, Gate 3, and Pre-sort, respectively; and for Ala-N-TIMP2$_{LIB}$, 1,309, 6,564, 527,052, and 252,542 of merged single sequences translated into valid amino acid sequences in Gate 1, Gate 2, Gate 3, and Pre-sort, respectively. There were very few variants of Ala-N-TIMP2$_{LIB}$ in Gate 1 (higher affinity than N-TIMP2$_{WT}$) and Gate 2 (WT-like affinity) due to the significant decrease (by three orders of magnitude) in affinity caused by the alanine insertion.

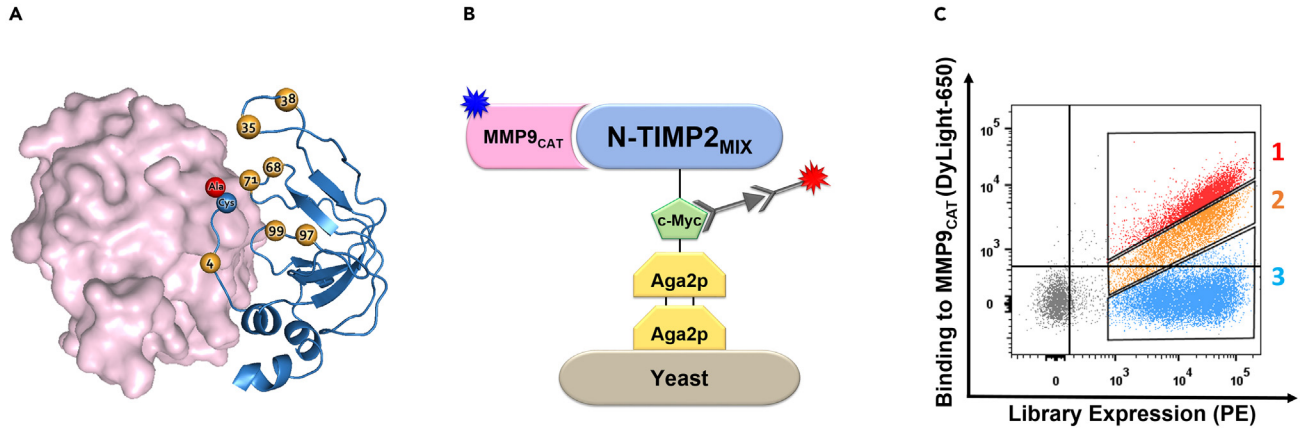


Figure 2. Affinity screening of the N-TIMP2$_{MIX}$ library by yeast surface display

(A) Illustration of the Ala-N-TIMP2$_{LIB}$ in complex with MMP14 (PDB: 1BUV). The Ala-N-TIMP2 library contains one mutation per clone in seven positions (orange) at the interface with the N-TIMP2 protein, and an alanine residue insertion (red) at the N terminus of the protein, following the Cys1 position (blue). This alanine insertion disrupts the interaction of Ala-N-TIMP2 with the Zn^{2+} ion buried in the MMP catalytic pocket and reduces the binding affinity of Ala-N-TIMP2 for MMP$_{CAT}$ in the complex.

(B) Schematic representation of yeast-surface-displayed N-TIMP2$_{MIX}$ interacting with the soluble target, MMP9$_{CAT}$, conjugated to DyLight-650 (blue star). An anti c-Myc tag antibody (labeled with phycoerythrin [PE; red star]), used to follow the expression of N-TIMP2 variants, is also shown.

(C) Flow cytometry was used to screen and fractionate N-TIMP2$_{MIX}$ variants according to their binding to MMP9$_{CAT}$ (100 nM). The x axis signal corresponds to yeast surface expression levels (as determined from the PE signal), while the y axis signal corresponds to MMP binding levels (as determined by the DyLight-650 signal). The polygonal shapes represent the three affinity-based sorting gates used to select the YSD library fractions.

We analyzed the sequence of each variant separately based on its frequency in the three sorted library fractions. Since we were interested in identifying mutations causing changes in the affinity to MMP9$_{CAT}$ in comparison to the WT in each library, we used the scattering of N-TIMP2 and Ala-N-TIMP2 as the criterion for choosing which of the sorted gates would be used for analysis of each library (Figure S1). For example, we used Gate 2, which had N-TIMP2$_{WT}$-like affinity, to analyze N-TIMP2$_{LIB}$, together with Gate 1, which had a higher affinity than Gate 2. Similarly, we based the Ala-N-TIMP2$_{LIB}$ analysis on library fractions that had comparable and higher affinities to that of Ala-N-TIMP2, i.e., Gate 3 and Gate 2, respectively.

After we had filtered out sequences that did not contain reads in at least one of the gates and sequences with mutations outside the binding interface, we calculated the frequency of each variant in each gate according to the ratio between the number of reads of each variant and the total number of reads in its parent library (i.e., either N-TIMP2$_{LIB}$ or Ala-N-TIMP2$_{LIB}$) within a certain gate (Equation 1; see STAR methods). This approach allowed us to overcome differences in read coverage, i.e., the number of total reads, in each library. To compare the affinity of MMP9$_{CAT}$ for each variant to that for N-TIMP2 or Ala-N-TIMP2 in each library, we calculated the normalized frequency (NF) of each variant as the frequency of that variant normalized to the frequency of the WT N-TIMP2 or Ala-N-TIMP2 within the same sorting gate (Equation 2; see STAR methods). To scale the binding affinity of each variant to MMP9$_{CAT}$ according to the frequency of that variant in a specific affinity gate, we calculated the ER of each variant (Equation 3; see STAR methods). Based on the flow cytometry scattering profile of the sorted N-TIMP2$_{MIX}$ library, we determined the ER values of N-TIMP2$_{LIB}$ variants and those of Ala-N-TIMP2$_{LIB}$ variants separately. We calculated the ER value for N-TIMP2$_{LIB}$ as the ratio of the NF of an individual variant having higher (Gate 1) and comparable (Gate 2) affinities to that of the WT, or variants having higher (Gate 2) and comparable (Gate 3) affinities to that of Ala-N-TIMP2, based on the premise that, if a specific mutation enhances the binding affinity to MMP9$_{CAT}$, that variant will be more abundant in an upper gate than in a lower gate, resulting in an ER value that is greater than 1. In opposition, if a mutation has deleterious effects on the binding affinity, it will be less abundant in an upper gate compared to a lower gate, leading to an ER value that is smaller than 1. Combining all the data obtained from the enrichment or depletion of each variant (from both N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$ libraries) in the three affinity gates (Gates 1–3) allowed us to cover a broad range of affinities for N-TIMP2 variants in complex with MMP9$_{CAT}$.

Training and performance evaluation of our models to predict ER of new N-TIMP2 variants

Since we only observed a subset of all possible single-mutant variants, 98 and 84 of N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$ out of 133 possible variants, respectively, we developed a machine-learning (ML) model to predict the $\log_2$ ER of any given variant. To accurately predict variants that were not observed in either N-TIMP2$_{LIB}$ or Ala-N-TIMP2$_{LIB}$, we trained two models, designated N-TIMP2$_{MODEL}$ and Ala-N-TIMP2$_{MODEL}$, to predict the $\log_2$ ER values based on a specific library. To train the two models, we used the dataset of variants that we generated, which includes single- and multi-mutation variants and their two $\log_2$ ER labels (i.e., N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$ labels). For each model, we held out 20% of high-quality data, of which 10% of the variants were used for tuning hyper-parameters and the other 10% were used for testing the trained model. We trained the model on the remaining 80% of variants. During the training process, we employed the logarithmic value of the

Table 1. K_i values of the N-TIMP2 variants against $MMP9_{CAT}$

Clone	K_i (nM)
N-TIMP2 (WT)	0.1089 ± 0.0181
Ala-N-TIMP2	441.20 ± 81.23
N-TIMP2 variant	
S4A	0.85 ± 0.2[a]
S4E	15.5 ± 1.2[a]
S4P	1.147 ± 0.175
S4Q	1.39 ± 0.14[a]
S4R	0.55 ± 0.05[a]
I35E	2.50 ± 0.13[a]
I35K	0.58 ± 0.08[a]
N38D	0.08965 ± 0.00604
N38Q	2.40 ± 0.55[a]
S68M	0.04702 ± 0.00714
S68N	0.05521 ± 0.00348
S68V	0.02808 ± 0.00260
S68W	0.37 ± 0.04[a]
S68Y	0.8 ± 0.2[a]
V71L	0.2033 ± 0.0244
V71N	4.68 ± 0.08[a]
V71R	0.6338 ± 0.0729
V71W	0.08306 ± 0.00801
H97M	0.04943 ± 0.00577
H97R	0.02214 ± 0.00188
H97C	0.2484 ± 0.0312
T99G	0.04742 ± 0.00268
T99M	0.0333 ± 0.0023
T99Q	0.09534 ± 0.00860
T99Y	1.1 ± 0.3[a]

[a] K_i measurements reported in the study of Sharabi et al.[1]

read count for each variant for sample weighting. We evaluated the performance of the $N\text{-}TIMP2_{MODEL}$ and the $Ala\text{-}N\text{-}TIMP2_{MODEL}$ in predicting log_2 ERs by the Pearson correlation of predicted and observed log_2 ERs.

When comparing multiple ML models on the validation set, in the $N\text{-}TIMP2_{MODEL}$ we observed Pearson correlations of 0.9495, 0.8816, 0.8968, and 0.9401 for the neural network, random forest, support vector regression with a linear kernel, and support vector regression with a polynomial kernel, respectively. In the $Ala\text{-}N\text{-}TIMP2_{MODEL}$ model, the Pearson correlations were 0.7742, 0.7737, 0.7018, and 0.7411 for the same models, respectively (Figure S3). Thus, due to the improved prediction performance achieved by the neural-network models compared to the non-neural-network models, we chose the neural-network models to be our $N\text{-}TIMP2_{MODEL}$ and the $Ala\text{-}N\text{-}TIMP2_{MODEL}$. Since all models achieved high correlations, we conclude that the high correlations stem mostly from the nature of the dataset. Both the $N\text{-}TIMP2_{MODEL}$ and the $Ala\text{-}N\text{-}TIMP2_{MODEL}$ achieved high performance on the held-out test set (Figures 3A and 3B), with Pearson correlations of 0.963 and 0.819, respectively (p values = 2.20×10^{-107} and 1.97×10^{-12}, respectively), meaning that both models can accurately predict the observed log_2 ERs.

Experimentally validating log_2 ER predictions by using purified variants

For a final prediction of log_2 ER for a specific variant, we generated predictions for pairs of N-TIMP2 and Ala-N-TIMP2 variants. To this end, we used both the $N\text{-}TIMP2_{MODEL}$ (library with a high initial binding affinity) and the $Ala\text{-}N\text{-}TIMP2_{MODEL}$ (library with a low initial binding affinity) to assess the affinity of each of the two versions of each variant, enabling the detection of numerous N-TIMP2 variants that span a very large range of affinities to $MMP9_{CAT}$. The sum of the two predicted log_2 ERs (from the $N\text{-}TIMP2_{MODEL}$ and the $Ala\text{-}N\text{-}TIMP2_{MODEL}$ that contain unique complementary information)—which we denote the $N\text{-}TIMP2_{MIX\_MODEL}$ prediction—represents mutations with a broad spectrum

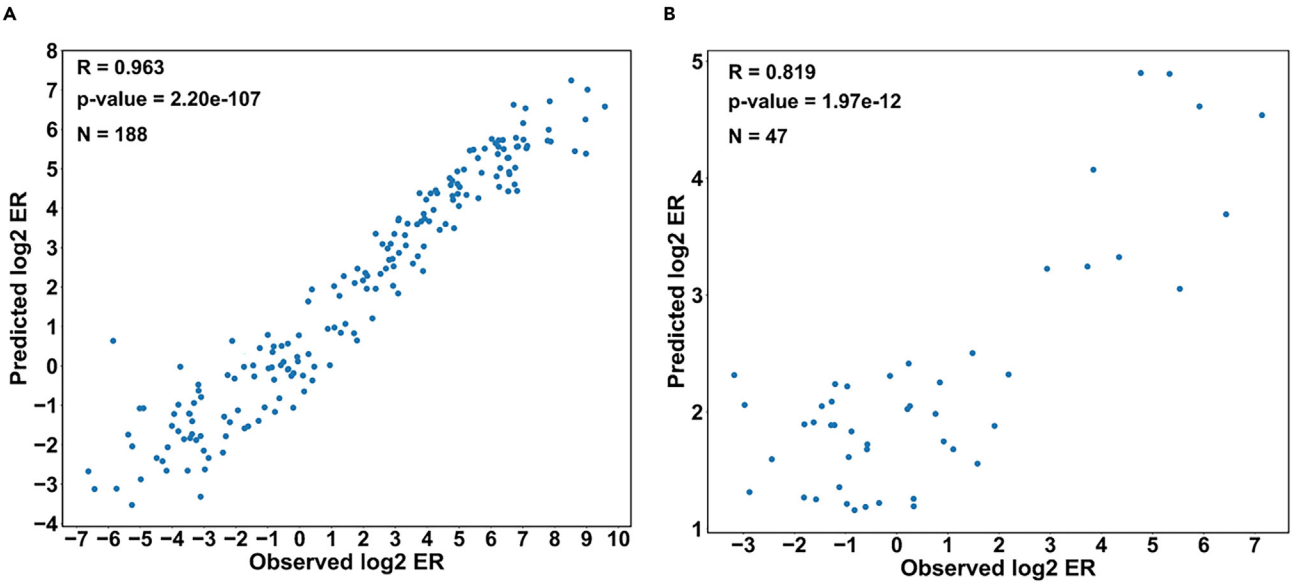


Figure 3. Performance evaluation of the neural-network models

(A and B) Results on a held-out high-quality test set of 10% of the variants for (A) N-TIMP2$_{MODEL}$ and (B) Ala-N-TIMP2$_{MODEL}$. The Pearson correlation was calculated between predicted and observed $\log_2$ ERs.

of affinity to MMP9$_{CAT}$. All N-TIMP2 variants with interface mutations that were not included in the experimental data could be identified by combining the predictions obtained from these two models (i.e., N-TIMP2$_{MIX_MODEL}$).

We then validated the affinity constants predicted by the N-TIMP2$_{MIX_MODEL}$. To this end, the WT protein and 25 variants were used, namely, 11 variants for which K_i values were available in the literature[1] and 14 purified variants for which we determined K_i values, as follows. The protein variants were expressed in the yeast *Pichia pastoris*. Use of the pPICZαA plasmid as the expression vector enabled production of the N-TIMP2 variants with a free N terminus (important for binding to MMP9$_{CAT}$) and C-terminal His and c-Myc epitope tags (used for purification and labeling, respectively). We purified the protein variants using affinity chromatography, followed by size-exclusion chromatography (SEC) (Figure S4A). Mass spectrometry analysis confirmed the correct mass of the purified mutants (Figure S4B), and SDS-PAGE analysis confirmed their high expression level and purity (Figure S4C). MMP9$_{CAT}$ with an N-terminal His$_6$ tag was expressed in *Escherichia coli* Bl21(DE3)pLysS cells using a pET28 vector. The protein was purified by affinity chromatography (Figure S5A), followed by SEC (Figure S5B), and SDS-PAGE analysis was used to confirm its expression and purity (Figure S5C). To assess *in vitro* the binding affinity of the N-TIMP2 variants to MMP9$_{CAT}$, we performed catalytic inhibition assays. In these assays, MMP9$_{CAT}$ was incubated with increasing concentrations of N-TIMP2$_{WT}$ or one of the purified N-TIMP2 variants, and the cleavage of a fluorescent MMP substrate was determined as a function of time. The slope of each reaction was calculated, and the acquired data were fitted by multiple regression to Morrison's tight-binding equation (Equation 5) to determine the K_i value for the binding of each complex.

The K_i value for each variant was normalized to that of N-TIMP2$_{WT}$ from the same experimental set to enable comparison of K_i values obtained in the current study with those from the literature. We note that all the variants were purified without the alanine insertion because our goal was to identify mutations in N-TIMP2 that change the binding affinity to MMP9$_{CAT}$ (Table 1; Figure 4A). N-TIMP2$_{WT}$ bound to/inhibited MMP9$_{CAT}$ with a K_i of 108.90 $\pm$ 18.09 pM, a finding consistent with a previous study.[26] As expected, the N-TIMP2 variants spanned a large range of affinities to MMP9$_{CAT}$: half of the purified N-TIMP2 variants (i.e., N-TIMP2$_{S68M}$, N-TIMP2$_{S68N}$, N-TIMP2$_{S68V}$, N-TIMP2$_{H97M}$, N-TIMP2$_{H97R}$, N-TIMP2$_{T99G}$, and N-TIMP2$_{T99M}$) showed higher affinities compared to the WT, with K_i values 2- to 5-fold lower than that of the WT. Some variants (i.e., N-TIMP2$_{N38D}$, N-TIMP2$_{V71W}$, and N-TIMP2$_{T99Q}$) showed a small improvement compared to the WT affinity binding, and some variants (i.e., N-TIMP2$_{S4P}$, N-TIMP2$_{V71R}$, N-TIMP2$_{V71L}$, and N-TIMP2$_{H97C}$) showed a decrease in binding affinity compared to the WT affinity.

To further validate the performance of the N-TIMP2$_{MIX_MODEL}$ (i.e., the ability to predict not only the observed $\log_2$ ERs of YSD variants but also the K_i values of purified variants), we held out an independent test set comprising 26 K_i values (not used in our algorithm) (Figure 4B). This test set included variants over a wide distribution of affinities allowing us to test absolute-affinity prediction performance of variable affinities. As we did not have sufficient data to train a complex function for a weighted combination of the two models in this K_i-prediction task, we decided to output the sum of predictions from both models (N-TIMP2$_{MODEL}$ and Ala-N-TIMP2$_{MODEL}$). Individually, each model yielded a lower correlation of the K_i values to the predicted $\log_2$ ER (Pearson's correlations of −0.494 for the N-TIMP2$_{MODEL}$ and −0.327 for the Ala-N-TIMP2$_{MODEL}$) than the sum of the predictions (Pearson's correlation of −0.545 for the N-TIMP2$_{MIX_MODEL}$) (p value = 0.004) (Figure S6).

This result demonstrates the ability of the N-TIMP2$_{MIX_MODEL}$ to successfully learn the affinity of unobserved variants (those not used in the development of the algorithm) and hence to identify variants of N-TIMP2 that were not observed in the original high-affinity complex.

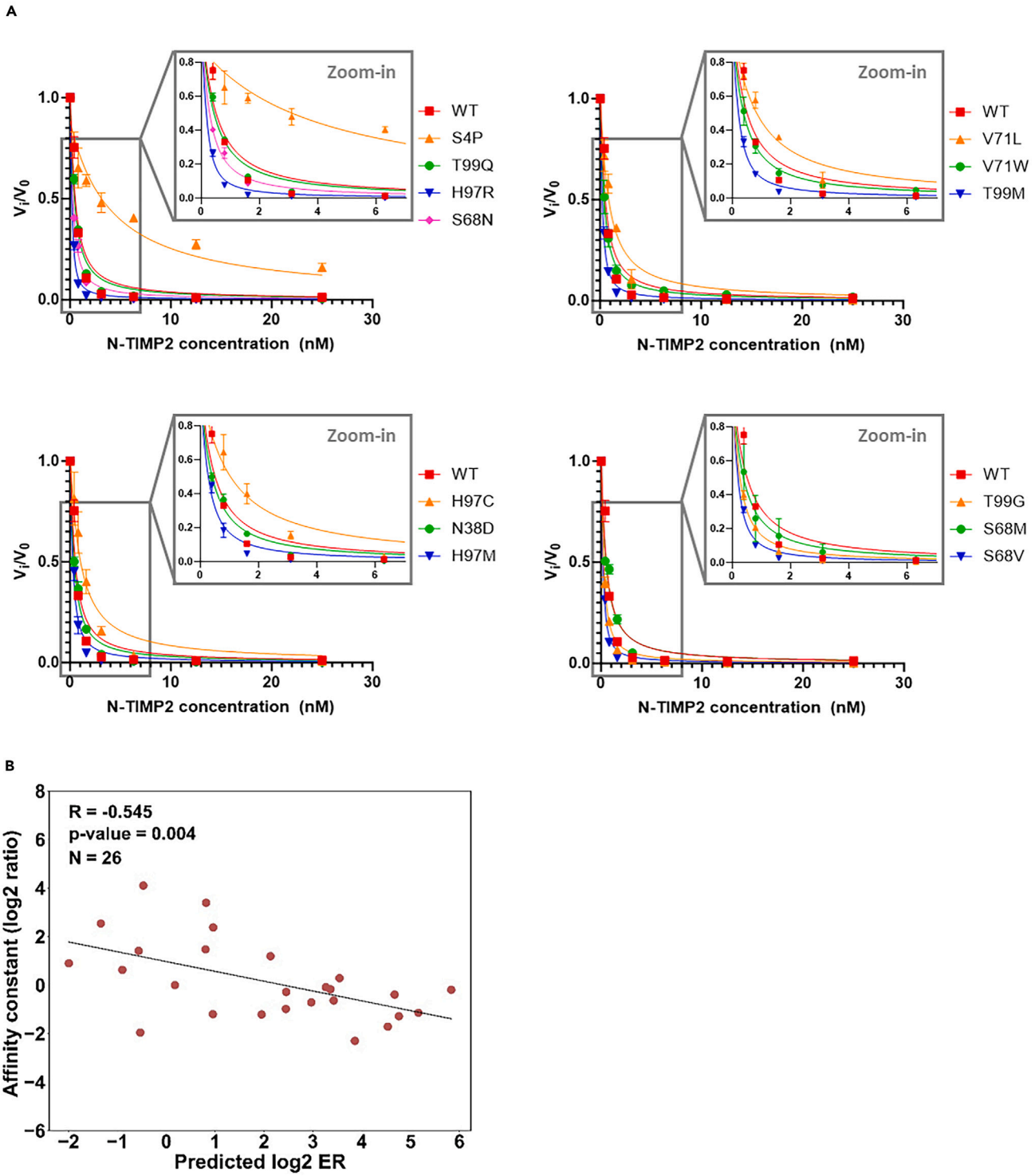


Figure 4. Inhibition of $MMP9_{CAT}$ by purified $N\text{-}TIMP2_{WT}$ and selected N-TIMP2 variants

(A) The cleavage of the fluorescent MMP substrate Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH2·TFA by $MMP9_{CAT}$ was measured over time. The initial reaction velocity was determined at each concentration of inhibitor (0.4–25 nM). The curves were fitted to Morrison's tight-binding inhibition equation (Equation 5) to obtain K_i values. Data shown are the averages of independent triplicate experiments, and error bars represent the standard deviation.

(B) Experimental validation of the $N\text{-}TIMP2_{MIX\_MODEL}$. We summed the output of the two neural-network models' predictions to score the WT and the 25 selected variants.

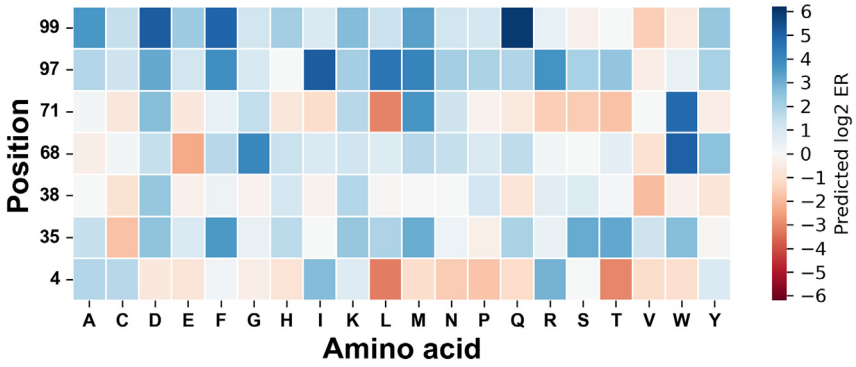


Figure 5. Effect of single mutations on the affinity of N-TIMP2 to MMP9$_{CAT}$

The figure shows a heatmap demonstrating the prediction of the affinity constant (predicted ER) of N-TIMP2 to MMP9$_{CAT}$. The substituted positions are shown on the y axis, and the substituting amino acids are shown on the x axis.

However, the correlation between the DMS $\log_2$ ERs (i.e., observed $\log_2$ ERs) and K_i values is limited, probably because of the innate differences between FACS-based YSD protein library affinity screening (DMS and ERs) and affinity measurements of purified proteins in solution; the former examines the interaction of yeast-anchored N-TIMP2 variants with a soluble target protein (MMP9$_{CAT}$), while the latter assesses the binding between two soluble proteins (N-TIMP2 variant and MMP9$_{CAT}$).

Identifying positions that affect the affinity of N-TIMP2 to MMP9$_{CAT}$

To comprehensively investigate the single-mutant binding landscape of the interaction between N-TIMP2 and MMP9$_{CAT}$, we predicted all possible single-mutant $\log_2$ ERs by using the N-TIMP2$_{MIX_MODEL}$ (Figure 5). Analysis of the binding landscape enabled identification of single mutations that change the affinity between N-TIMP2 and MMP9$_{CAT}$ across the binding interface of N-TIMP2/MMP9$_{CAT}$. We identified the majority of affinity-reducing mutations in position 4 (Figure 5) in agreement with it being a "hot spot" on the N terminus binding region of TIMP2. We also identified two clear "cold spots" in which there was affinity enhancement for the majority of N-TIMP2 variants with mutations at positions 97 or 99; these included mutations to methionine, phenylalanine, and aspartic acid.

Effect of data quality on the performance of the ML models

The size and quality of a DMS dataset affect the performance of a ML model trained on it. We aimed to test two factors that could influence the quality of our ML: the number of mutations in observed variants and the number of sequenced reads. The number of mutations in each variant controls the variability observed over the dataset and thus affects the ability of the ML model to learn and predict mutations that were not observed in the data. The number of reads controls the statistical sample size of each variant and consequently the robustness of the $\log_2$ ER values measured in the experiment. We assessed how these experimental factors influence prediction performance of the neural-network model by resampling *in silico* the whole N-TIMP2$_{LIB}$ dataset to generate simulated datasets with different numbers of unique variants based on their number of mutations and different numbers of DNA library sequencing reads. We trained the models on each simulated dataset and tested each model on randomly selected held-out test sets.

Training on variants with a single mutation per variant yielded a low correlation of 0.567, on average, between predicted and observed $\log_2$ ERs of variants with multiple mutations. However, including variants with two and three mutations in the training set led to a significant increase in the correlation, i.e., 0.658 and 0.855, respectively. By training the models with variants containing three or more mutations, the Pearson correlation increased to reach a value greater than 0.8 (Figure 6A). With regard to the effect of the number of reads, the model that was trained on a mere 10% of the data demonstrated the worst performance of all models but still achieved excellent performance, with an average Pearson correlation of 0.9 (Figure 6B).

This result shows that a diversity of variants and mutations in the training data is crucial as incorporating variants with two or three mutations led to a significant improvement in model performance compared to increasing the number of reads. This outcome is consistent with the phenomenon of epistasis, wherein the collective impact of multiple mutations diverges from the impact that would be expected by simply adding up their individual effects.[27] Specifically, previous findings have demonstrated that beneficial single mutations within the N-TIMP2/MMP9$_{CAT}$ complex, when incorporated together, result in negative epistasis.[28] Therefore, we conclude that training a model on variants that possess multiple mutations enables the model to better learn the epistatic effect than training on single mutations and as a result its overall performance will be enhanced.

DISCUSSION

In this study, we introduced a novel approach to accurately and quantitatively predict the impact of numerous N-TIMP2 variants on the binding affinity in the N-TIMP2/MMP9$_{CAT}$ complex with the goal of comprehensively mapping the affinity landscape of this complex. We generated two N-TIMP2 mutagenesis libraries, one with high (N-TIMP2$_{LIB}$) and the other with low (Ala-N-TIMP2$_{LIB}$) affinity toward MMP9$_{CAT}$ and

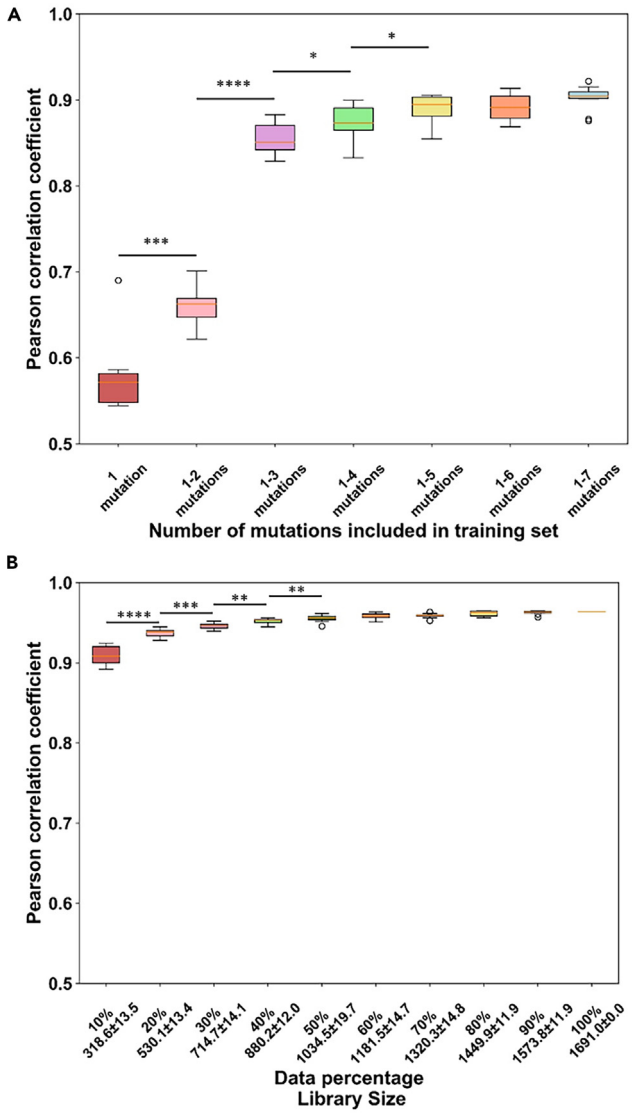


Figure 6. Prediction performance as a function of the number of mutations in each variant in the training set or the size of the training set

(A) Prediction performance as a function of the number of mutations in each variant in the training set. We evaluated the model on a test set of 10% of the data containing variants with a high read count (at least 40 reads) excluding variants having single mutations. For 1 to X ($2 \leq X \leq 7$) mutations in each variant, a training set of 350 variants was randomly selected 10 times (excluding 1 mutation with a randomly selected training set of 80 variants).

(B) Prediction performance as a function of the size of the training set. Results are reported for the same held-out test set of 10% of the variants. A training set of 10%–100% of the rest of the data was randomly selected 10 times (excluding 100% of the data with a fixed training set). The library size of the training sets is shown as the mean ± standard deviation of 10 repeats. We compared the results by Wilcoxon rank-sum test *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

performed affinity-dependent fractional sorting of a mixture of these libraries by flow cytometry. We determined the $\log_2$ ERs separately for each library based on the NGS of the sorted library fractions and trained DNNs to learn the pattern of each library, i.e., which positions and amino acids are most important for binding affinity. Implementation of these procedural measures enabled us to predict the affinity (K_i values) of any N-TIMP2 variant to $MMP9_{CAT}$, even variants that were absent from the mutagenesis libraries.

To the best of our knowledge, we demonstrate here for the first time that a single mutation in the N-TIMP2 ligand, specifically, an N-terminal alanine insertion, can be used as a means to overcome the limitation of the narrow affinity range of N-TIMP2 for $MMP9_{CAT}$. By leveraging this finding to artificially generate two versions of the same ligand—one having low affinity and the other having high affinity to $MMP9_{CAT}$—we were able to use a fixed concentration of the target protein in each affinity sorting screen. By assessing the affinity of each of the two versions of each variant (high and low affinity) in each affinity sorting fraction, we gathered more comprehensive data on the affinity to the target compared to assessing only one of the versions and on only a single sorting fraction. This comprehensive approach enabled us to enhance the accuracy and sensitivity of our affinity predictions and represents an applied solution when experiments (e.g., affinity screens) are possible

only with a fixed, single concentration of the target. In our case, the two versions of the same ligand enabled us to make predictions for variants across a broad range of affinities to $MMP9_{CAT}$ while using a single concentration of $MMP9_{CAT}$.

We plan to extend our study in multiple directions. First, for future DMS experiments, we plan to increase the variability of the variants rather than their read depth. We posit that, with greater variability, the model can better learn epistatic effects, which will have a greater impact on its prediction performance than increasing the read depth to reduce sampling noise (Figures 6A and 6B). Second, we plan to improve our model by adding additional features, such as structural-based features, solvent accessibility, protein stability, and evolutionarily conserved information.[16,18] Third, we plan to perform additional low-throughput experiments to measure K_i values and to test a different binning of the sorting gate. These steps may help us to improve our DNN models, currently trained on high-throughput NGS data, for predicting absolute-affinity values. Moreover, we plan to further improve prediction performance of K_i values by unsupervised learning, e.g., by using large protein language models.[29]

Given the difficulties encountered during earlier research involving DMS in predicting novel variants spanning a broad range of affinities, our ability to predict the affinity of unobserved variants without including them in an experiment is of significant importance to the field of PPI prediction and for developing therapies targeting these interactions. More generally, we plan to apply our innovative approach to many other protein-function datasets as an improved method to provide a rich characterization platform for any PPI affinity landscape. By mapping the affinity landscape across a wide range of affinities, researchers may gain a more complete understanding of these interactions. Our approach is also expected to shorten the screening process of protein-based therapeutics. In practical terms, more comprehensive knowledge of N-TIMP2 protein inhibitor is essential for drug discovery, where understanding the strength of N-TIMP2/$MMP9_{CAT}$ interactions can be critical for identifying and targeting this interface for optimal inhibition of MMP9, a protease involved in various pathologies.[30,31]

Limitations of the study

Our study tackled several challenges. First, generating pairs of N-TIMP2 and Ala-N-TIMP2 variants led to measurements of two distinct $\log_2$ ERs for each variant. To overcome the challenge of combining the two distinct $\log_2$ ERs, we trained two different models (i.e., N-$TIMP2_{MODEL}$ and Ala-N-$TIMP2_{MODEL}$) to predict the effect of each type of variant in each pair separately. The summation of the two predicted $\log_2$ ERs from the high-affinity N-$TIMP2_{MODEL}$ and the low-affinity Ala-N-$TIMP2_{MODEL}$, which we denoted the N-$TIMP2_{MIX\_MODEL}$ prediction, yielded the highest correlation with K_i values but is a simple non-parametrized function and thus is likely suboptimal. Second, one of the main limitations of DMS studies is the partial coverage of the mutation space of a given protein, i.e., not all possible variants are present in the library. We overcame this limitation by using a state-of-the-art DNN model to predict the activity of any given variant. Our experimental dataset contained 84 and 98 single-mutation variants in N-$TIMP2_{LIB}$ and Ala-N-$TIMP2_{LIB}$, respectively, covering only 63% and 74% of all possible single mutations. This partial coverage could be attributed to common challenges encountered during the synthesis of the libraries and subsequent experiments (e.g., limitation in library transformation yield).[16] Since the N-$TIMP2_{MIX\_MODEL}$ predicts the binding affinity of any variant of N-TIMP2, it solves the limitation of partial coverage of the mutation space of a given protein in DMS studies, but, still, it is only a prediction and not an experimental measurement. Third, the number of repetitions of each variant in a DMS is limited, resulting in specific variants having too few repetitions to enable accurate calculation of their $\log_2$ ERs. Previous studies have addressed this issue by excluding from the training data variants with reads below a specific quality threshold.[16,32] We opted to partition the variants into training, validation, and test sets, while ensuring that the validation and test sets comprised high-confidence variants, i.e., those with a high number of repetitions. Additionally, during the training process, we employed the logarithmic value of the read count for each variant for sample weighting, thereby putting more trust in high-confidence variants. Still, the sequencing depth we used limited the number of variants we observed and the accuracy of their observed ER.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Niv Papo (papo@bgu.ac.il).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- HTS data have been deposited at SRA and are publicly available as of the date of publication. Accession number is listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Dr. Uzi Hadad, Ilse Katz Institute for Nanoscale Science & Technology, for his consultation and technical assistance. We would also like to thank Dr. Gal Yosef for her suggestions and assistance. R.M. thanks the Israel Data Science and AI Initiative for providing cloud credit for this work. This work was supported by the National Institutes of Health (R01GM144393) to Y.O. and N.P.; the Israeli Council for Higher Education (CHE) via Data Science Research Center, Ben-Gurion University of the Negev, Israel, the Israel Cancer Association (20240036) to Y.O.; and the Rosetrees Trust (OoR2022/100004), the Worldwide Cancer Research (grant number 20-0238), the Israel Cancer Research Fund (ICRF) (grant number 846497), and the United States-Israel Binational Science Foundation (grant number 2019303) to N.P.

AUTHOR CONTRIBUTIONS

R.M. and S.-L.A.L. conducted the experiments. R.M., S.-L.A.L., Y.O., and N.P. designed the experiments, performed the analysis, and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Generating a mixture of N-TIMP2 and Ala-N-TIMP2 mutagenesis libraries
  - PCR reaction to generate the Ala-N-TIMP2$_{LIB}$
  - Yeast-displayed fractional MMP9$_{CAT}$ affinity screens
  - HTS of the library fractions
  - Computational analysis of the HTS data
  - Merging and filtering the high-throughput sequencing data
  - Dataset
  - Model architectures
  - Selection of model hyper-parameters
  - Training and evaluation of the models
  - Validation of the model predictions
  - Evaluation of prediction performance as a function of data characteristics
  - Generating N-TIMP2 variants by site-directed mutagenesis
  - Production and purification of N-TIMP2 variants and MMP9$_{CAT}$
  - Characterization of MMP9$_{CAT}$
  - Catalytic activity and inhibition assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110772.

Received: March 11, 2024
Revised: June 27, 2024
Accepted: August 15, 2024
Published: August 19, 2024

REFERENCES

1. Sharabi, O., Shirian, J., Grossman, M., Lebendiker, M., Sagi, I., and Shifman, J. (2014). Affinity- and specificity-enhancing mutations are frequent in multispecific interactions between TIMP2 and MMPs. PLoS One *9*, e93712. https://doi.org/10.1371/journal.pone.0093712.
2. Jubb, H.C., Pandurangan, A.P., Turner, M.A., Ochoa-Montaño, B., Blundell, T.L., and Ascher, D.B. (2017). Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. Prog. Biophys. Mol. Biol. *128*, 3–13. https://doi.org/10.1016/j.pbiomolbio.2016.10.002.
3. Cukuroglu, E., Engin, H.B., Gursoy, A., and Keskin, O. (2014). Hot spots in protein-protein interfaces: towards drug discovery. Prog. Biophys. Mol. Biol. *116*, 165–173. https://doi.org/10.1016/j.pbiomolbio.2014.06.003.
4. Shirian, J., Sharabi, O., and Shifman, J.M. (2016). Cold Spots in Protein Binding. Trends Biochem. Sci. *41*, 739–745. https://doi.org/10.1016/j.tibs.2016.07.002.
5. Morrison, K.L., and Weiss, G.A. (2001). Combinatorial alanine-scanning. Curr. Opin. Chem. Biol. *5*, 302–307. https://doi.org/10.1016/s1367-5931(00)00206-4.
6. Rosenfeld, L., Heyne, M., Shifman, J.M., and Papo, N. (2016). Protein Engineering by Combined Computational and In Vitro Evolution Approaches. Trends Biochem. Sci. *41*, 421–433. https://doi.org/10.1016/j.tibs.2016.03.002.
7. Kowalsky, C.A., Klesmith, J.R., Stapleton, J.A., Kelly, V., Reichkitzer, N., and Whitehead, T.A. (2015). High-resolution sequence-function mapping of full-length proteins. PLoS One *10*, e0118193. https://doi.org/10.1371/journal.pone.0118193.
8. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807. https://doi.org/10.1038/nmeth.3027.
9. Fowler, D.M., Stephany, J.J., and Fields, S. (2014). Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat. Protoc. *9*, 2267–2284. https://doi.org/10.1038/nprot.2014.153.
10. Naftaly, S., Cohen, I., Shahar, A., Hockla, A., Radisky, E.S., and Papo, N. (2018). Mapping protein selectivity landscapes using multi-target selective screening and next-generation sequencing of combinatorial libraries. Nat. Commun. *9*, 3935. https://doi.org/10.1038/s41467-018-06403-x.
11. Whitehead, T.A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S.J., De Mattos, C., Myers, C.A., Kamisetty, H., Blair, P., Wilson, I.A., and Baker, D. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat. Biotechnol. *30*, 543–548. https://doi.org/10.1038/nbt.2214.
12. Aharon, L., Aharoni, S.L., Radisky, E.S., and Papo, N. (2020). Quantitative mapping of binding specificity landscapes for homologous targets by using a high-throughput method. Biochem. J. *477*, 1701–1719. https://doi.org/10.1042/BCJ20200188.
13. Heyne, M., Papo, N., and Shifman, J.M. (2020). Generating quantitative binding landscapes through fractional binding selections combined with deep sequencing and data normalization. Nat. Commun. *11*, 297. https://doi.org/10.1038/s41467-019-13895-8.
14. Jenson, J.M., Xue, V., Stretz, L., Mandal, T., Reich, L.L., and Keating, A.E. (2018). Peptide design by optimization on a data-parameterized protein interaction landscape. Proc. Natl. Acad. Sci. USA *115*, E10342–E10351. https://doi.org/10.1073/pnas.1812939115.
15. Currin, A., Kwok, J., Sadler, J.C., Bell, E.L., Swainston, N., Ababi, M., Day, P., Turner, N.J., and Kell, D.B. (2019). GeneORator: An Effective Strategy for Navigating Protein Sequence Space More Efficiently through Boolean OR-Type DNA Libraries. ACS Synth.

Biol. *8*, 1371–1378. https://doi.org/10.1021/acssynbio.9b00063.

16. Livesey, B.J., and Marsh, J.A. (2020). Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. Mol. Syst. Biol. *16*, e9380. https://doi.org/10.15252/msb.20199380.
17. Hsu, C., Nisonoff, H., Fannjiang, C., and Listgarten, J. (2022). Learning protein fitness models from evolutionary and assay-labeled data. Nat. Biotechnol. *40*, 1114–1122. https://doi.org/10.1038/s41587-021-01146-5.
18. Gelman, S., Fahlberg, S.A., Heinzelman, P., Romero, P.A., and Gitter, A. (2021). Neural networks to learn protein sequence-function relationships from deep mutational scanning data. Proc. Natl. Acad. Sci. USA *118*, e2104878118. https://doi.org/10.1073/pnas.2104878118.
19. Horne, J., and Shukla, D. (2022). Recent Advances in Machine Learning Variant Effect Prediction Tools for Protein Engineering. Ind. Eng. Chem. Res. *61*, 6235–6245. https://doi.org/10.1021/acs.iecr.1c04943.
20. Gray, V.E., Hause, R.J., Luebeck, J., Shendure, J., and Fowler, D.M. (2018). Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. Cell Syst. *6*, 116–124.e3. https://doi.org/10.1016/j.cels.2017.11.003.
21. Song, H., Bremer, B.J., Hinds, E.C., Raskutti, G., and Romero, P.A. (2021). Inferring Protein Sequence-Function Relationships with Large-Scale Positive-Unlabeled Learning. Cell Syst. *12*, 92–101.e8. https://doi.org/10.1016/j.cels.2020.10.007.
22. Yano, H., Nishimiya, D., Kawaguchi, Y., Tamura, M., and Hashimoto, R. (2020). Discovery of potent and specific inhibitors targeting the active site of MMP-9 from the engineered SPINK2 library. PLoS One *15*, e0244656. https://doi.org/10.1371/journal.pone.0244656.
23. Arkadash, V., Radisky, E.S., and Papo, N. (2018). Combinatorial engineering of N-TIMP2 variants that selectively inhibit MMP9 and MMP14 function in the cell. Oncotarget *9*, 32036–32053. https://doi.org/10.18632/oncotarget.25885.
24. Wingfield, P.T., Sax, J.K., Stahl, S.J., Kaufman, J., Palmer, I., Chung, V., Corcoran, M.L., Kleiner, D.E., and Stetler-Stevenson, W.G. (1999). Biophysical and functional characterization of full-length, recombinant human tissue inhibitor of metalloproteinases-2 (TIMP-2) produced in Escherichia coli. Comparison of wild type and amino-terminal alanine appended variant with implications for the mechanism of TIMP functions. J. Biol. Chem. *274*, 21362–21368. https://doi.org/10.1074/jbc.274.30.21362.
25. Yosef, G., Arkadash, V., and Papo, N. (2018). Targeting the MMP-14/MMP-2/integrin alphavbeta3 axis with multispecific N-TIMP2-based antagonists for cancer therapy. J. Biol. Chem. *293*, 13310–13326. https://doi.org/10.1074/jbc.RA118.004406.
26. Arkadash, V., Yosef, G., Shirian, J., Cohen, I., Horev, Y., Grossman, M., Sagi, I., Radisky, E.S., Shifman, J.M., and Papo, N. (2017). Development of High Affinity and High Specificity Inhibitors of Matrix Metalloproteinase 14 through Computational Design and Directed Evolution. J. Biol. Chem. *292*, 3481–3495. https://doi.org/10.1074/jbc.M116.756718.
27. Starr, T.N., and Thornton, J.W. (2016). Epistasis in protein evolution. Protein Sci. *25*, 1204–1218. https://doi.org/10.1002/pro.2897.
28. Shirian, J., Arkadash, V., Cohen, I., Sapir, T., Radisky, E.S., Papo, N., and Shifman, J.M. (2018). Converting a broad matrix metalloproteinase family inhibitor into a specific inhibitor of MMP-9 and MMP-14. FEBS Lett. *592*, 1122–1134. https://doi.org/10.1002/1873-3468.13016.
29. Livesey, B.J., and Marsh, J.A. (2023). Advancing variant effect prediction using protein language models. Nat. Genet. *55*, 1426–1427. https://doi.org/10.1038/s41588-023-01470-3.
30. Li, Z., Wei, J., Chen, B., Wang, Y., Yang, S., Wu, K., and Meng, X. (2023). The Role of MMP-9 and MMP-9 Inhibition in Different Types of Thyroid Carcinoma. Molecules *28*, 3705. https://doi.org/10.3390/molecules28093705.
31. Hu, J., Van den Steen, P.E., Sang, Q.X.A., and Opdenakker, G. (2007). Matrix metalloproteinase inhibitors as therapy for inflammatory and vascular diseases. Nat. Rev. Drug Discov. *6*, 480–498. https://doi.org/10.1038/nrd2308.
32. Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics *200*, 413–422. https://doi.org/10.1534/genetics.115.175802.
33. Mata-Fink, J., Kriegsman, B., Yu, H.X., Zhu, H., Hanson, M.C., Irvine, D.J., and Wittrup, K.D. (2013). Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. J. Mol. Biol. *425*, 444–456. https://doi.org/10.1016/j.jmb.2012.11.010.
34. Angelini, A., Chen, T.F., de Picciotto, S., Yang, N.J., Tzeng, A., Santos, M.S., Van Deventer, J.A., Traxlmayr, M.W., and Wittrup, K.D. (2015). Protein Engineering and Selection Using Yeast Surface Display. Methods Mol. Biol. *1319*, 3–36. https://doi.org/10.1007/978-1-4939-2748-7_1.
35. Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D. (2006). Isolating and engineering human antibodies using yeast surface display. Nat. Protoc. *1*, 755–768. https://doi.org/10.1038/nprot.2006.94.
36. Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507.
37. Sela-Passwell, N., Kikkeri, R., Dym, O., Rozenberg, H., Margalit, R., Arad-Yellin, R., Eisenstein, M., Brenner, O., Shoham, T., Danon, T., et al. (2011). Antibodies targeting the catalytic zinc complex of activated matrix metalloproteinases show therapeutic potential. Nat. Med. *18*, 143–147. https://doi.org/10.1038/nm.2582.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti c-Myc antibody	Abcam	Ab32072
goat anti-mouse secondary antibody conjugated to phycoerythrin (PE)	Sigma-Aldrich	P9670
restriction enzyme SacI-HF	New England Biolabs	10166195
anti-mouse secondary-antibody conjugated to alkaline phosphatase	Jackson ImmunoResearch	AB_2338528
Bacterial and virus strains		
Escherichia coli DH10β bacteria	(OriGene Technologies, MD, USA)	CC100004
Chemicals, peptides, and recombinant proteins		
Phusion HF polymerase	New England Biolabs	M0530S
AmpureXP beads	Beckman Coulter	A63882
Zeocin	Invitrogen	R25005
Instant Blue	CBS Scientific	R-250
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich	367931
Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2·TFA	Merck Millipore	BML-P276-0001
Critical commercial assays		
E.Z.N.A. Yeast Plasmid Mini Kit	Omega Bio-tek	D6942-00S
HiYield Gel/PCR Fragments Extraction Kit	RBC Bioscience	YDF100
Access Array barcode library	Fluidigm	N/A
DeNovix dsDNA High Sensitivity Assay kit	DeNovix	TN144
MaxiPrep	Geneaid	N/A
Deposited data		
Raw data	This paper	SRA: PRJNA1071009
Recombinant DNA		
pPICZαA construct	RBC Bioscience	N/A
Software and algorithms		
Fast Length Adjustment of Short reads (FLASH) software	The Center for Computational Biology (CCB), Johns Hopkins University	N/A
Code for predicting the affinity landscape of N-TIMP2/MMP9CAT	This paper	https://doi.org/10.5281/zenodo.13294007
Other		
Superdex 75 column	GE Healthcare Life Sciences	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

S. cerevisiae EBY100 yeast strain.
yeast *Pichia pastoris X-33* strain.
Escherichia coli DH10β bacteria.

METHOD DETAILS

Generating a mixture of N-TIMP2 and Ala-N-TIMP2 mutagenesis libraries

We generated two N-TIMP2 libraries with different initial affinities to $MMP9_{CAT}$. We derived the high-affinity library, $N\text{-}TIMP2_{LIB}$, from $N\text{-}TIMP2_{WT}$ mutagenized individually at seven highly tolerant binding-interface positions: 4, 35, 38, 68, 71, 97, and 99 [based on the $N\text{-}TIMP2_{WT}$ gene (PDB ID 1BUV)[1,12]]. We generated the low-affinity library, $Ala\text{-}N\text{-}TIMP2_{LIB}$, based on $N\text{-}TIMP2_{LIB}$, but with an alanine residue added at the N-terminus of the $N\text{-}TIMP2_{LIB}$ gene. We numbered the mutated positions in $Ala\text{-}N\text{-}TIMP2_{LIB}$ according to the $N\text{-}TIMP2_{WT}$ gene, i.e., starting from the cysteine in the first position and ignoring the added alanine residue in the counting.

We mixed the $N\text{-}TIMP2_{LIB}$ and $Ala\text{-}N\text{-}TIMP2_{LIB}$ libraries and expressed the mixed DNA library, i.e., $N\text{-}TIMP2_{MIX}$, in a YSD system using *S. cerevisiae* EBY100 yeast strain according to an established protocol. For the construction of this yeast-displayed library, we used the pCHA-VRC01-scFv vector (obtained from Dane Wittrup, Massachusetts Institute of Technology), such that the N-terminus of the displayed proteins in the libraries would be distant from the yeast membrane and available for interaction with the target protein $MMP9_{CAT}$.[33] In the pCHA-VRC01-scFv construct, we fused the C-terminus of $N\text{-}TIMP2_{MIX}$ to the N-terminus of the yeast membrane Aga-2 protein, leaving the N-terminus of the displayed N-TIMP2 (or Ala-N-TIMP2) library protein variants exposed to the solvent.[34]

To generate the $Ala\text{-}N\text{-}TIMP2_{LIB}$, we designed forward and reverse primers with homology to the sequence of the pCHA-VRC01-scFv vector and with an overlapping sequence to the N-TIMP2 gene (Table S1). To generate the forward primer, we added a DNA sequence coding for alanine at the 5′ end (at the beginning of the sequence that overlaps with the N-TIMP2 gene). We performed a PCR reaction using these primers, Phusion HF polymerase (New England Biolabs, Ipswich, MA, USA), and the $N\text{-}TIMP2_{LIB}$ gene as a template. The PCR reaction is described below.

PCR reaction to generate the $Ala\text{-}N\text{-}TIMP2_{LIB}$

The PCR reaction was performed with primers (Table S1), Phusion HF polymerase (New England Biolabs, Ipswich, MA, USA), and the gene library $N\text{-}TIMP2_{LIB}$ as a template. The PCR conditions were as follows: 98°C for 30 s, followed by 35 cycles of 10-s, 20-s, and 11-s incubations at 98°C, 69°C and 72°C, respectively, and then a 10-min incubation at 72°C. The resulting DNA constructs of the $Ala\text{-}N\text{-}TIMP2_{LIB}$ were mixed with those of the $N\text{-}TIMP2_{LIB}$ in a 1:1 ratio, giving a mixed DNA library ($N\text{-}TIMP2_{MIX}$). The pCHA plasmid was linearized with the restriction enzymes BamHI-HF and Nhe-HF (New England Biolabs). The open plasmid (~2 μg) together with ~6.8 μg of the mixed DNA library ($N\text{-}TIMP2_{MIX}$) were transformed by homologous recombination into a freshly prepared competent EBY100 *Saccharomyces cerevisiae* yeast strain using a MicroPulser electroporator (Bio-Rad, CA, USA), as previously described.[35] The transformed yeast cells were grown in SDCAA selective medium (2% dextrose, 0.67% yeast nitrogen base, 0.5% BactoÔ Casamino Acids, 1.47% sodium citrate, and 0.429% citric acid monohydrate, adjusted to pH 4.5) overnight at 30°C to an OD_{600} of 10.0 (10^8 cells/mL). The library size was estimated by plating serial dilutions on SDCAA plates (2% dextrose, 0.67% yeast nitrogen base, 0.5% Bacto Casamino Acids, 1.54% Na_2HPO_4, 1.856% $Na_2HPO_4 \cdot H_2O$, 18.2% sorbitol, and 1.5% agar).

Yeast-displayed fractional $MMP9_{CAT}$ affinity screens

Library screening was performed as described. The yeast-displayed mixed library $N\text{-}TIMP2_{MIX}$ that had been grown in SDCAA medium overnight was transferred into the selective SGCAA medium (2% galactose, 0.67% yeast nitrogen base, 0.5% Bacto Casamino Acids, 1.47% sodium citrate, 0.429% citric acid monohydrate, adjusted to pH 4.5) for overnight growth at 30°C to induce expression of the library on the surface of the yeast cells.[35] For screening the library against $MMP9_{CAT}$, ~10^6 of the yeast cells were collected and washed with binding buffer [50 mM Tris, pH 7.5, 100 mM NaCl, 5 mM $CaCl_2$, and 1% bovine serum albumin (BSA)]. To determine expression levels of the displayed proteins, the cells were incubated with a primary antibody, i.e., mouse anti c-Myc antibody (Abcam, Cambridge, UK), at a dilution of 1:50, for 1 h at room temperature. Thereafter, the cells were washed with the binding buffer and incubated with goat anti-mouse secondary antibody conjugated to phycoerythrin (PE) (Sigma-Aldrich, St. Louis, MO, USA) in a 1:50 dilution, together with soluble DyLight-650-labeled $MMP9_{CAT}$ (purified as described in the production and purification of N-TIMP2 variants and MMP9CAT section in STAR methods), in the dark for 30 min on ice. The cells were washed and sorted with a FACSAria (Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, Israel). We designed the first sort for enriching the $MMP9_{CAT}$-binding $N\text{-}TIMP2_{MIX}$ population by using 1 μM DyLight-650-labeled $MMP9_{CAT}$ such that all sorted variants could be detected when incubated with 100 nM $MMP9_{CAT}$.

We then performed a second sort to generate three library fractions, each having a different binding affinity toward $MMP9_{CAT}$, such that the lowermost gate (Gate 3) had an Ala-N-TIMP2-like affinity, the middle gate (Gate 2) had an $N\text{-}TIMP2_{WT}$-like affinity, and the uppermost gate (Gate 1) included variants with a higher affinity to $MMP9_{CAT}$ than that of the variants in the other two gates. We conducted this fractional sort using 100 nM fluorescently labeled $MMP9_{CAT}$. We chose this target concentration of $MMP9_{CAT}$ for the fractional sort based on two considerations: (i) by using the target at an intermediate concentration between the K_i values of the two complexes, $N\text{-}TIMP2/MMP9_{CAT}$ and $Ala\text{-}N\text{-}TIMP2/MMP9_{CAT}$, all the variants of $N\text{-}TIMP2_{MIX}$ would be included in a single sort; and (ii) the selected concentration would facilitate a broad scattering of the variants according to their binding signal and hence enable the separation of subpopulations (library fractions) with different affinities. We then used dual-color flow cytometry for analysis of the mixed library expression and binding to $MMP9_{CAT}$ in an Accuri C6 flow cytometer (BD Biosciences, San Jose, CA, USA) according to the same labeling protocol as that for the flow cytometry sorting.

HTS of the library fractions

The preparation for the HTS of the library fractions and the filtering process on the HTS data is described. Plasmid DNA was extracted individually for each library from the parental pre-sorted $N\text{-}TIMP2_{MIX}$ library and the three affinity-sorted library fractions (each yielding ~10^8 yeast

cells) using the E.Z.N.A. Yeast Plasmid Mini Kit (Omega Bio-tek, Norcross, GA, USA) according to the manufacturer's protocol. The isolated DNA plasmids of the libraries were run on 1% agarose gel and then purified with the HiYield Gel/PCR Fragments Extraction Kit (RBC Bioscience, Taiwan). The purified DNA plasmid product (~1 ng) of each of the above libraries was used for gene amplification in a PCR reaction, containing 2% DNA template, 5% forward primer (10 μM), 5% reverse primer (10 μM) (Table S2), 20% Phusion reaction buffer, 2% deoxynucleotide triphosphates (dNTPs), and 1% Phusion HF polymerase (New England Biolabs) in doubly distilled water. The PCR conditions were as follows: 98°C for 30 s, followed by 35 cycles of 10-s, 20-s, and 11-s incubations at 98°C, 72°C and 72°C, respectively, and then a 10-min incubation at 72°C. The PCR amplified products were run on a 3% agarose diagnostic DNA gel and then purified with HiYield Gel/PCR Fragments Extraction Kit (RBC Bioscience). These purified PCR products were sent for next generation sequencing (NGS; Hylabs, Rehovot, Israel), in which a second PCR was performed using the Access Array barcode library (Fluidigm, San Francisco, CA, USA) to add to each sample the adaptor and index sequences used by the Illumina sequencing technology. Thereafter, the samples were purified with AmpureXP beads (Beckman Coulter, CA, USA), and their DNA concentrations were determined in a DNA high sensitivity assay performed in a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and the DeNovix dsDNA High Sensitivity Assay kit (DeNovix, Wilmington, DE, USA). Finally, the samples were run on a TapeStation (Agilent, CA, USA) to verify the size of the PCR product. The pools were then loaded for sequencing on an Illumina Miseq (Illumina, San Diego, CA, USA), using the v2 (500 cycles) kit.

Computational analysis of the HTS data

We translated the DNA sequences of the pre-sorted and the three sorted library fractions into the amino acid sequences that they encode. We then aligned these sequences to the N-TIMP2 sequence. We filtered out sequences of lengths different from the WT length, or sequences with a stop codon before the end. In the sorted library fractions, we analyzed the sequences of N-TIMP2$_{LIB}$ separately from those of Ala-N-TIMP2$_{LIB}$. For each library, we performed the analysis based on two out of the three sorted library fractions to obtain the ratio of the fraction that increases affinity to the fraction with WT-like affinity, i.e., we analyzed the N-TIMP2$_{LIB}$ based on library fractions having higher (Gate 1) and comparable (Gate 2) affinities to that of the WT protein sequence in the respective gates. In the same manner, to analyze Ala-N-TIMP2$_{LIB}$, we used two library fractions, one with higher affinity than Ala-N-TIMP2 and the other with Ala-N-TIMP2-like affinity, i.e., Gates 2 and 3, respectively. For each sequence, we counted the number of occurrences of each variant in each library fraction. Then, we calculated the frequency of each variant, f_{mut_i}, as

$$f_{mut_i} = \frac{\#reads\ mut_i}{\sum_{n}^{1} \#reads\ mut_i} \quad \text{(Equation 1)}$$

where $\#reads\ mut_i$ is the number of reads of a variant in library fraction i, and $\sum_{n}^{1}\#reads\ mut_i$ is the sum of the number of all the reads for all variants in its library (N-TIMP2$_{LIB}$ or Ala-N-TIMP2$_{LIB}$) in library fraction (gate) i.

Next, to compare the frequencies of each variant to that of the WT in the same library, we calculated NF_{mut_i} by

$$NF_{mut_i} = \frac{f_{mut_i}}{f_{WT}} \quad \text{(Equation 2)}$$

which is the ratio of the frequency of a given variant in library fraction i to the frequency of N-TIMP2 or Ala-N-TIMP2 clones in library fraction (gate) i.

Based on the NFs, we calculated the ERs of each variant in the N-TIMP2$_{MIX}$ library (separately for N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$) as follows:

$$ER_{N-TIMP2_{LIB}} = \frac{NF_{mut_i\ Gate1}}{NF_{mut_i\ Gate2}} \quad and \quad ER_{Ala-N-TIMP2_{LIB}} = \frac{NF_{mut_i\ Gate2}}{NF_{mut_i\ Gate3}} \quad \text{(Equation 3)}$$

where NF_{mut_i} is the NF of a variant in a specific gate.

Merging and filtering the high-throughput sequencing data

An average Illumina quality score was calculated for each read in a given set of paired-end reads. Read pairs in which either of the reads had an average quality score <20 (i.e., the odds that the corresponding base call are incorrect are 1/100) were discarded. The remaining read pairs were merged into a single sequence by Fast Length Adjustment of Short reads (FLASH) software (The Center for Computational Biology (CCB), Johns Hopkins University, Baltimore, Maryland, USA).[36] This filtering process was applied for the sequences of all libraries.

Dataset

We generated a dataset (Figure 1F) containing the amino acid sequences, their two $\log_2$ ER labels (N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$ labels), and the $\log_2$ read count of each variant in each of the library fractions. We represent the sequences as the seven highly tolerant binding interface positions which were mutated in the experiment. In total, this dataset contains 1,878 N-TIMP2$_{LIB}$ variants and 466 Ala-N-TIMP2$_{LIB}$ variants (including multi-mutation variants).

Model architectures

We developed two ER predictors, each an ensemble of multi-layered neural networks. The first predictor was trained over the N-TIMP2$_{LIB}$ dataset, which we denote as N-TIMP2$_{MODEL}$, and the second over the Ala-N-TIMP2$_{LIB}$ dataset, which we denote as Ala-N-TIMP2$_{MODEL}$ (Figure 1G). Each input sequence is one-hot encoded resulting in a matrix of dimensions 20*7. The flattened input matrix is fed into a fully connected layer with 32 and 8 neurons in the N-TIMP2$_{MODEL}$ and Ala-N-TIMP2$_{MODEL}$, respectively. The output of this layer goes through another fully connected layer with 4 and 2 neurons (in the N-TIMP2$_{MODEL}$ and the Ala-N-TIMP2$_{MODEL}$, respectively). Finally, each output is governed by a single neuron with a linear activation function. We implemented the model using Keras python library with TensorFlow backend (version 2.2.0). The runtime, CPU usage, and maximum memory for training the model on an Intel(R) Xeon(R) Gold 6130 CPU were 61.8 s, 576%, and 0.26 GB, respectively, for the Ala-N-TIMP2$_{MODEL}$ and 478.7 s, 981%, and 0.26 GB, respectively, for the N-TIMP2$_{MODEL}$.

Selection of model hyper-parameters

To train the models and select the optimal hyper-parameters, we excluded 10% of the data from the training set and used it as a validation set. We held out another 10% of the data to use as a test set. The validation set and the test set both contained variants with a high read count (40 and 100 in N-TIMP2$_{LIB}$ and Ala-N-TIMP2$_{LIB}$, respectively) to obtain high-quality and reliable test sets. We chose the hyper-parameter values using a random search over 100 random hyper-parameter combinations with a pre-defined range for each parameter (Table S4). We chose the hyper-parameter values that lead to the highest Pearson correlation on the validation set of each model separately.

Training and evaluation of the models

Once the hyper-parameters had been set, to increase model robustness, we used the random-ensemble-initialization approach, where we trained 10 models over the same data but with different randomly initialized weights and different random orders in which the data was used for model training in batches. We used the $\log_2$ sum of read counts of each variant over all gates as the sample weight in the training process so as to place greater reliance on variants with a large statistical sample (a large number of reads). Then, for prediction, we used as the output the average of the predictions of the 10 trained models. To evaluate the prediction ability of two models, we calculated the Pearson correlation between the predictions and the experimental $\log_2$ ER values of a held-out test of 10% of the data. We denote the sum of the two predicted $\log_2$ ERs (from N-TIMP2$_{MODEL}$ and Ala-N-TIMP2$_{MODEL}$) as N-TIMP2$_{MIX_MODEL}$ prediction.

Validation of the model predictions

To experimentally validate the N-TIMP2$_{MIX_MODEL}$ predictions, we used experimentally obtained K_i values as an independent test set, on which the models were neither developed nor trained. This set comprised 26 K_i values, including those for the WT and for 7 purified variants determined in this study (see below), those for 7 purified variants reported in a previous study,[12] and those for 11 variants with K_i values taken from the literature.[1] K_i values for all variants, including those taken from the literature, were determined using the same method (described below under "catalytic activity and inhibition assays" subsection). Each test-set variant contains a single mutation in one of the seven residue positions mutated in our library (i.e., positions 4, 35, 38, 68, 71, 97, and 99). Out of these variants, 15 showed a decrease in affinity while 10 variants showed an increase in affinity compared to the WT, whose K_i value was determined in the same experiment. We calculated the correlation between the K_i values and the predictions. Since we used K_i values of proteins purified in different purification systems, in different sets of experiments and in different laboratories, we normalized each K_i measurement to the K_i value of the WT in the same experiment and used the $\log_2$ transformation of these values.

Evaluation of prediction performance as a function of data characteristics

To evaluate the N-TIMP2$_{MODEL}$ performance as a function of the number of mutations in each variant in the training set, we used a random test set of 10% of the data (excluding variants with single mutations so that they appear only in the training set) that comprised of variants with a high read count (40 repetitions) to obtain high-quality and reliable test sets. For X ($1 \leq X \leq 7$) mutations within each variant, we randomly selected a training set of 350 variants, each having at most X mutations (excluding 1 mutation with a training set of 80 variants). To evaluate the performance of N-TIMP2$_{MODEL}$ as a function of the training data size, we used the same test set of 10% of the data. We randomly selected a training set of X% ($10 \leq X \leq 100$) of the rest of the data (excluding 100% of the data with a fixed training set). After training of each model, we evaluated it in terms of the Pearson correlation between the predictions and the experimental $\log_2$ ER values of the test set. We repeated the training and evaluation process 10 times for each X, and report the mean and standard deviation of the Pearson correlation.

Generating N-TIMP2 variants by site-directed mutagenesis

The evaluation of prediction performance as a function of data characteristics is described in this and the next two subsections. For purification of the variants, the pPICZαA construct, containing the N-TIMP2 gene[26] with the AOX1 promoter at its N-terminus and a hexahistidine tag at its C-terminus, and the Zeocin resistance gene, was used. The plasmid was propagated in *Escherichia coli* DH10β bacteria (OriGene Technologies, MD, USA) and then purified with a HiYield plasmid mini kit (RBC Bioscience). Thereafter, a site-directed mutagenesis procedure was carried out to generate the N-TIMP2 variants in a PCR reaction using specific primers (Table S3). These primers contained the codon encoding the desired amino acid mutation in the middle of their sequence, and 15 bp flanking the codon from each side of the primer, which were complementary to the template N-TIMP2 DNA sequence. The PCR mixture comprised 2% plasmid DNA template (~50 ng), 5% forward

primer (10 μM), 5% reverse primer (10 μM), 20% Phusion HF buffer, 3% DMSO, 2% dNTPs, and 1% Phusion HF polymerase (New England Biolabs) in doubly distilled water. The PCR conditions were as follows: 98°C for 3 min followed by 25 cycles of 10-s, 30-s and 10-min incubations at 98°C, 55°C–65°C (depending on the specific primer) and 72°C, respectively, followed by a 10-min incubation at 72°C. Thereafter, the PCR products were loaded on a diagnostic 1% agarose gel to verify the procedure's success by detecting bands of the correct size, and then transformed into competent *E. coli* DH10β. The transformed bacteria were plated on LB agar plates containing 50 μg/mL Zeocin (Invitrogen, Grand Island, NY, USA). Plasmids were extracted from several bacterial colonies, and the correct sequence with the inserted mutation was verified for each N-TIMP2 variant (Genetics Unit, National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Israel).

Production and purification of N-TIMP2 variants and $MMP9_{CAT}$

To purify the selected N-TIMP2 variants, we used the yeast *Pichia pastoris X-33* strain (which upon induction secretes proteins into the growth medium), according to the pPICZα protocol (Invitrogen, Carlsbad, CA, USA) with minor modifications. In brief, for preparing a sufficient quantity of plasmids from each N-TIMP2 variant, the transformed *E. coli* DH10β cells containing the plasmids were grown overnight at 37°C in 300 mL of LB medium containing 50 μg/mL Zeocin (Invitrogen), and the plasmids were extracted using MaxiPrep (Geneaid, New Taipei City, Taiwan). Thereafter, ~100 μg of plasmids from each variant were linearized with the restriction enzyme SacI-HF (New England Biolabs) and then transformed into freshly prepared electro-competent *P. pastoris X-33* according to the pPICZα protocol (Invitrogen). The transformed yeast cells, each containing a different N-TIMP2 variant, were grown on YPDS plates (18.2% sorbitol, 2% peptone, 2% D-glucose, 2% agar, 1% yeast extract and 50 μg/mL Zeocin) for 72 h at 30°C. N-TIMP2 variants were purified as previously described[26]: For each N-TIMP2 variant, 10 yeast colonies were selected and grown overnight in 5 mL of BMGY medium (2% peptone, 1% yeast extract, 0.23% $K_2H(PO_4)$, 1.181% $KH_2(PO_4)$, 1.34% yeast nitrogen base, 4 × 10^{-5}% biotin, 1% glycerol) at 30°C, and then transferred into 5 mL of BMMY medium (2% peptone, 1% yeast extract, 0.23% $K_2H(PO_4)$, 1.181% $KH_2(PO_4)$, 1.34% yeast nitrogen base, 4 × 10^{-5}% biotin, 0.5% methanol) for protein induction of 72 h, with the addition of 1% methanol once a day in the last 2 days. Overexpression of the secreted proteins was determined by western blot, using a 1:3000 dilution of mouse anti-His_6 primary antibody (Abcam), followed by a 1:5000 dilution of anti-mouse secondary-antibody conjugated to alkaline phosphatase (Jackson ImmunoResearch, West Grove, PA, USA), and detection by incubation in 2 mL of 5-bromo-4-chloro-3-indolyl phosphate reagent (Sigma-Aldrich). For large-scale production of the proteins, the yeasts expressing the N-TIMP2 variants and exhibiting the highest protein expression, were grown in 50 mL of BMGY medium overnight, followed by 72 h of growth in 500 mL of BMMY medium, with daily additions of 1% methanol. The proteins were purified by centrifugation of the yeast cell suspension at 3800 × g for 20 min and filtration of the supernatant, followed by the addition of 300 mM NaCl and 10 mM imidazole at pH 8.0. The supernatant was incubated for 1 h at 4°C, centrifuged at 3800 × g for 10 min, filtered using a 0.22-μm cutoff, and then loaded on nickel-nitrilotriacetic acid-Sepharose beads (Invitrogen), followed by washing with 50 mM Tris, pH 7.5, 300 mM NaCl and 10 mM imidazole, and elution with 20 mL of 50 mM Tris, pH 7.5, 300 mM NaCl and 250 mM imidazole. The elution fraction was concentrated in a Vivaspin centrifugal concentrator with a 5-kDa cutoff (GE Healthcare Life Sciences, Pittsburgh, PA, USA). The proteins were then further purified using a Superdex 75 column (GE Healthcare Life Sciences) with elution buffer (50 mM Tris, pH 7.5, 300 mM NaCl and 5 mM $CaCl_2$) in an ÄKTA Pure instrument (GE Healthcare Life Sciences). SDS-PAGE analysis on a 15% polyacrylamide gel under reducing conditions was then performed for the purified proteins. Bands were visualized by staining with Instant Blue (CBS Scientific, CA, USA). Purified protein samples were subjected to mass spectrometry analysis (Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, Israel), and protein concentrations were determined by UV–Vis absorbance at 280 nm, using a NanoDrop Spectrophotometer (Thermo Fisher Scientific), with an extinction coefficient (ε280) of 13,325 M^{-1} cm^{-1} for N-TIMP2 and all its variants.

The human MMP9 catalytic domain ($MMP9_{CAT}$), lacking the fibronectin-like domain (residues 107–215 and 391–443), was purified as previously described,[37] with the following modifications: The $MMP9_{CAT}$ gene was expressed in *E. coli* Bl21(DE3)pLysS cells in a pET28 vector (with an N-terminal His_6 tag) and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich) overnight at 30°C. The cells were harvested and subjected to three rounds of sonication and centrifugation at 12,000 × g to isolate the inclusion bodies. The inclusion bodies were solubilized in 8 M urea and 25 mM Tris, pH 7.5. Thereafter, the protein was loaded onto a nickel column and eluted with 8 M urea, 25 mM Tris, pH 7.5, 30 mM NaCl and 200 mM imidazole. The eluted protein was refolded by slow dialysis over 3 days at 4°C with a gradient of decreasing urea concentrations, from 8 M to 0 M. Finally, the enzyme was purified by size exclusion on a Superdex 75 (GE Healthcare Life Sciences) gel filtration column with buffer containing 20 mM Tris, pH 7.5, 50 mM NaCl and 5 mM $CaCl_2$. The $MMP9_{CAT}$ concentration was determined by UV-Vis absorbance at 280 nm, using a NanoDrop Spectrophotometer (Thermo Fisher Scientific), with an extinction coefficient (ε280) of 33,920 M^{-1} cm^{-1}. The process yielded ~0.1 mg protein per liter of culture cells. The purity of $MMP9_{CAT}$ was determined by SDS-PAGE analysis. For flow cytometry, the purified $MMP9_{CAT}$ was fluorescently labeled for 1 h at room temperature with DyLight-650 Amine Reactive Dye (Thermo Fisher Scientific), which contains an *N*-hydroxysuccinimide ester that forms a covalent bond with primary amines, as previously described.[26]

Characterization of $MMP9_{CAT}$

The catalytic activity (K_m value) of the purified $MMP9_{CAT}$ enzyme was determined by measuring 0.5 nM $MMP9_{CAT}$ activity against different concentrations (0–50 μM) of the fluorogenic substrate Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2·TFA [where Mca is (7-methoxycoumarin-4-yl)acetyl, Dpa is N-3-(2,4-dinitrophenyl)-L-2,3-diaminopropionyl and TFA is trifluoroacetic acid)] (Merck Millipore, CA, USA) in TCNB buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM $CaCl_2$, and 0.05% Brij). Fluorescence was monitored for 1 h using a Synergy 2 plate reader

with 340/30 excitation and 400/30 emission filters (BioTek, Winooski, VT, USA) at 37°C. The data was fitted to the Michaelis-Menten equation (Equation 4) using GraphPad Prism 7 (San Diego, CA, USA).

$$V = \frac{V_{max}[S]}{K_m+[S]} \quad \text{(Equation 4)}$$

where V - enzyme velocity; V_{max} - enzyme maximum velocity achieved at maximum substrate concentration; S - substrate concentration, and K_m – the Michaelis-Menten constant.

The calculated value of K_m (6.175 ± 0.528 μM) was obtained from three independent experiments.

The catalytic activity of the fluorescently labeled $MMP9_{CAT}$ enzyme was confirmed in an assay performed in TCNB buffer with a final concentration of 15 μM of the fluorogenic substrate Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2·TFA.

Catalytic activity and inhibition assays

N-TIMP2 and its variants were tested for their inhibitory activity against 0.2 nM $MMP9_{CAT}$. To this end, $MMP9_{CAT}$ was incubated with 0.4–25 nM WT N-TIMP2 or N-TIMP2 variants, or with 3.9–1000 nM Ala-N-TIMP2, in TCNB buffer for 1 h at 37°C. Thereafter, the fluorogenic substrate Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2·TFA was added to the reaction mixture at a final concentration of 15 μM. Fluorescence was monitored (with 340/30 excitation and 400/30 emission filters) using a Synergy 2 plate reader (BioTek) at 37°C. The reactions were followed spectroscopically for 60 min, and initial rates were determined from the linear portion of the increase in the fluorescence signal caused by the cleavage of the fluorescent substrate. The data was globally fitted by multiple regression to Morrison's tight-binding inhibition equation (Equation 5) using GraphPad Prism 7.

$$\frac{V_i}{V_0} = \frac{1 - \left([E]+[I]+K_i^{app}\right) - \sqrt{\left([E]+[I]+K_i^{app}\right)^2 - 4[E][I]}}{2[E]} \quad \text{(Equation 5)}$$

where V_i - enzyme velocity in the presence of an inhibitor; V_0 - enzyme velocity in the absence of an inhibitor; E - enzyme concentration; I - inhibitor concentration and K_i^{app} - the apparent inhibition constant, which is given by Equation 6:

$$K_i^{app} = K_i\left(1 + \frac{[S]}{K_m}\right) \quad \text{(Equation 6)}$$

where S - inhibitor concentration; K_m – the Michaelis-Menten constant, and K_i - the inhibition constant.

The inhibition constant, K_i, was calculated by plotting the initial velocities against different concentrations of the inhibitors. K_i values are given as means ± SE for three independent experiments.

QUANTIFICATION AND STATISTICAL ANALYSIS

For performance evaluation of the DNN models a Pearson correlation was calculated between predicted and observed $\log_2$ ERs. Details (including the exact value of n) are given in Figure 3 and its legend.

For the determination of the inhibition of $MMP9_{CAT}$ by purified $N\text{-}TIMP2_{WT}$ and selected N-TIMP2 variants, the inhibition curves were fitted to Morrison's tight-binding inhibition equation (Equation 5) obtain K_i values. Data shown in Figure 4 are the averages of independent triplicate experiments, and error bars represent the standard deviation. Details (including the exact value of n) are given in Figure 3 and its legend. For the experimental validation of the $N\text{-}TIMP2_{MIX\_MODEL}$ similar statistical parameters are provided.

For the prediction performance as a function of the number of mutations in each variant in the training set and the size of the training set, the library size of the training sets is shown as the mean ± standard deviation of 10 repeats. We compared the results by Wilcoxon rank-sum test *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Details are given in Figure 6 and its legend.