



RESEARCH ARTICLE

Engineering selective amyloid precursor protein inhibitors by machine learning and deep mutational scanning

Reut Meiri¹ | Oz Reuveni² | Michal Levi² | Evette S. Radisky³ |
Niv Papo^{2,4} | Yaron Orenstein^{1,5}

¹Department of Computer Science and Artificial Intelligence, Bar-Ilan University, Ramat Gan, Israel

²Avram and Stella Goldstein-Goren Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel

³Department of Cancer Biology, Mayo Clinic Comprehensive Cancer Center, Jacksonville, Florida, USA

⁴National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel

⁵The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel

Correspondence

Yaron Orenstein, Department of Computer Science and Artificial Intelligence, Bar-Ilan University, Ramat Gan, Israel.

Email: aron.orenstein@biu.ac.il

Niv Papo, Avram and Stella Goldstein-Goren Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

Email: papo@bgu.ac.il

Funding information

National Institutes of Health, Grant/Award Number: R01GM144393; Israel Cancer Association, Grant/Award Number: 20240036; Rosetrees Trust, Grant/Award Number: OoR2022/100004; Israel Cancer Research Fund, Grant/Award Number: 846497; United States-Israel Binalational Science Foundation, Grant/Award Number: 2019303

Review Editor: Nir Ben-Tal

Abstract

Deep mutational scanning (DMS) has proven effective for mapping protein–protein interactions (PPIs), but it cannot provide complete coverage of the mutation landscape, particularly for multi-mutant variants. To address this limitation, we trained machine-learning (ML) models on previously generated DMS data for a stabilized amyloid precursor protein inhibitor (APPI) binding to either of two serine proteases, mesotrypsin and kallikrein-6 (KLK6), which are implicated in various human disorders. We combined the models to accurately predict the binding selectivity of APPI variants, including double-mutant variants, for the two serine proteases. We achieved a Pearson correlation of 0.937 between predicted log₂ selectivity enrichment ratios and DMS-derived values. We further validated the predictions of our combined model by yeast-surface-display measurements and inhibition assays of purified APPI variants and revealed epistatic interactions that shape protease selectivity. Guided by binding selectivity predictions, we identified highly selective APPI variants, including the most selective mesotrypsin inhibitor reported to date. Together, these findings support the use of DMS and ML as a framework for predicting PPI selectivity and prioritizing selective therapeutic protein variants.

KEYWORDS

deep mutational scanning, neural networks, next-generation sequencing, protein engineering, protein–protein interactions

1 | INTRODUCTION

Targeting the protein–protein interactions (PPIs) that facilitate and regulate various cellular processes poses

significant challenges to the drug-development community. Some PPIs are highly specific, while others lack specificity, involving proteins that engage multiple, often structurally similar, targets on overlapping interfaces (Nooren & Thornton, 2003). Whereas highly specific PPIs have been leveraged to develop protein-based

Reut Meiri, Oz Reuveni, and Michal Levi have contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

therapeutics with reduced off-target toxicity, offering promising treatments for a variety of clinical conditions (Ebrahimi & Samanta, 2023; Slathia et al., 2023) engineering highly selective therapeutic proteins for closely related protein targets remains a far more difficult task, given their similar structural and chemical properties.

The binding specificity of proteins involved in non-specific PPIs with a particular target may nonetheless be manipulated by mutating those proteins (Sharabi et al., 2014). Such mutations may alter the free energy of the PPI, thereby either destabilizing or strengthening the interaction (Jubb et al., 2017). PPI selectivity is shaped primarily by interface residues, where “hot-spot” residues are interface positions that contribute significantly to binding free energy, such that mutations at hot-spot positions typically reduce affinity, whereas “cold-spot” residues are interface residues occupied by suboptimal amino acids, such that mutations at cold-spot positions typically increase affinity (Fry & Vassilev, 2005; Shirian et al., 2016). In addition, correlated mutations can also affect binding specificity to particular targets (Pazos et al., 1997), as they represent co-evolving residues that may be either in close proximity or separated from one another within a sequence. However, the combined effects of correlated mutations may not be simply additive, but rather epistatic (Starr & Thornton, 2016). The interactions between correlated mutations thus present challenges in protein engineering, as epistasis complicates predictions of binding affinity and selectivity outcomes (Miton & Tokuriki, 2016).

Two common high-throughput techniques for assessing PPIs comprise yeast two-hybrid screening (Coates & Hall, 2003) and yeast surface display (YSD) (Teymennet-Ramírez et al., 2022). However, these techniques are time-consuming, and their scalability is limited. Therefore, computational methods trained on high-throughput datasets, such as those generated through deep mutational scanning (DMS), have gained significant attention. DMS generates rich datasets that capture the functional impact of mutations on proteins and can be used to probe PPIs (Fowler & Fields, 2014). Furthermore, integrating DMS with statistical approaches enables the identification of complex underlying patterns (Leander et al., 2022) and reveals the effects of mutations on PPIs. Nonetheless, despite the promise of high-throughput approaches, DMS (which is based on data from experimental library screens) only partially covers the mutation space of a given protein. Since the search-space size increases exponentially with the number of mutated residue positions, there is a combinatorial explosion of the number of possible protein variants (Currin et al., 2019). Thus, this issue of partial coverage constitutes a significant obstacle to progress in obtaining a comprehensive picture of the mutational landscape of PPIs (Livesey & Marsh, 2020).

To impute the mutational landscape of a specific PPI, a variety of computational methods have been

developed to predict binding affinity based on the proteins' amino acid sequences. Among these methods, supervised methods have emerged as the most successful in predicting binding affinity (Livesey & Marsh, 2020). Supervised methods have advanced significantly with the development of deep neural networks, particularly convolutional neural networks (CNNs) (Gelman et al., 2021) and graph neural networks (Horne & Shukla, 2022). Still, many supervised methods are significantly less accurate at predicting beneficial mutations compared to deleterious mutations (Reeb et al., 2020). Their key limitation is their failure to capture epistatic effects, which are common across positions in a protein's amino acid sequence (Hsu et al., 2022). In addition, only very few studies have focused on selectivity prediction in the face of multiple targets, and therefore the challenge of engineering selective inhibitors remains to be addressed.

To take up this challenge, we chose two homologous complexes as model systems; each comprises a complex of amyloid precursor protein inhibitor (APPI)—a human serine protease inhibitor belonging to the Kunitz family—with a serine protease, either mesotrypsin or kallikrein-6 (KLK6), as the target protein (Scheidig et al., 1997) (Figure 1). The two proteases share high sequence homology and structural similarity, including the catalytic triad, His57, Asp102, and Ser195, that is highly conserved in serine proteases (Jiang et al., 2021). Both proteins constitute important targets for drug development: Mesotrypsin, normally expressed in the pancreas, brain, and colon (Salameh et al., 2010), is strongly implicated in tumor growth and the metastasis of epithelial cancers (Diederichs et al., 2004; Han et al., 2013; Hockla et al., 2010; Hockla et al., 2012; Jiang et al., 2010; Ma et al., 2015; Ma et al., 2019; Yang et al., 2008), and KLK6, normally expressed in the central nervous system (CNS), skin, and glands (Bayani & Diamandis, 2012), is associated with various cancers and neurodegenerative conditions (Bayani & Diamandis, 2012; Bayés et al., 2004; Blaber et al., 2007). Their similar structures constitute an obstacle to the development of high-affinity selective inhibitors, as described below.

Wild-type APPI (APPI_{WT}) binds nonspecifically to the active site of serine proteases, including mesotrypsin and KLK6, via its canonical binding loop (spanning positions T11 to I18) (PDB:1ZJD). The positions in this loop are tolerant to amino acid substitutions (with the exception of C14, which maintains the integrity, stability, and structure of the loop), thereby enabling the production of APPI variants with a broad range of affinities toward the two targets (Cohen et al., 2016; Cohen et al., 2018; Naftaly et al., 2018; Sananes et al., 2018). Since APPI_{WT} is susceptible to cleavage by mesotrypsin (Cohen et al., 2016; Salameh et al., 2010), we worked, in this study, with a more stable variant of APPI, namely, APPI_{M17G,I18F,F34V}, designated

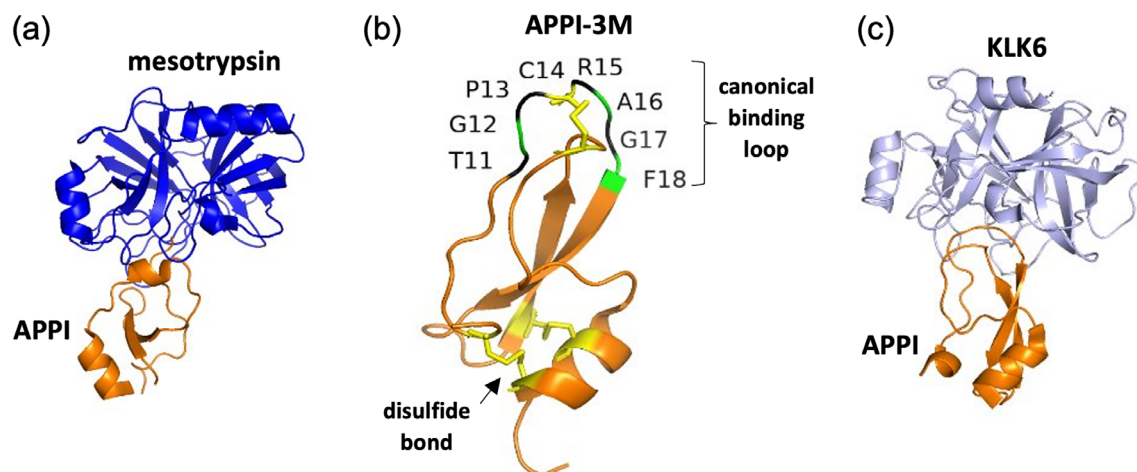


FIGURE 1 (a) Crystal structure of mesotrypsin complexed with APPI (PDB: 3L33) showing a molecule of mesotrypsin bound to a molecule of APPI. (b) Crystal structure of APPI-3M (PDB: 5C67), which, like APPI, is stabilized by three disulfide bonds. (c) Crystal structure of KLK6 complexed with APPI (PDB: 5NX1) showing a molecule of KLK6 bound to a molecule of APPI.

APPI-3M, as the scaffold protein for engineering selective binding. This variant is more resistant to cleavage by mesotrypsin (Cohen et al., 2016; Sananes et al., 2018) and has a greater affinity (by two orders of magnitude) for mesotrypsin than for KLK6. It is therefore a particularly suitable scaffold for engineering specificity switches and designing selective binding for the two model protein targets.

As the first steps toward better understanding the mechanisms and evolution of selective serine protease-inhibitor interactions and designing selective APPI inhibitors, we previously mapped the binding selectivity landscapes of APPI-3M with mesotrypsin and KLK6 by using DMS and pairwise selectivity screening (Naftaly et al., 2018). However, that study focused on single mutants and estimated binding selectivity through affinity screens and retrospective analysis without computational modeling, thereby limiting our ability to rank double mutants or to predict the impact of single and combined mutations on selectivity. In our follow-up study (Sarfati et al., 2022), we used machine-learning (ML) with DMS data to predict APPI variant selectivity for mesotrypsin and KLK6, but the drawbacks of that study were that the predictions were most accurate at the extremes and less reliable for moderate selectivity, and validation with soluble proteins was lacking.

The current study was designed to address the above drawbacks: We trained ML models on our previously generated DMS data of APPI-3M variants (Naftaly et al., 2018) with the aims to predict their binding selectivity for mesotrypsin and KLK6, to reveal epistatic interactions between mutated positions, and to highlight specific residues contributing to target selectivity. Thereafter, we used YSD to experimentally validate the predictions of the variants' selectivity. We then selected variants with pronounced selectivity toward either mesotrypsin or KLK6 for further purification and

evaluation via target inhibition assays. Ultimately, we identified two highly selective inhibitors, one for mesotrypsin and the other for KLK6, that were not observed in the DMS data. The inhibitor of mesotrypsin is, to the best of our knowledge, the most selective mesotrypsin inhibitor vs. KLK6 reported to date.

2 | METHODS

2.1 | Computational analysis of the high-throughput sequencing data

We previously generated the DMS data (Naftaly et al., 2018) used in this study, as follows. In brief, to select for binding specificity of APPI-3M variants for mesotrypsin vs. KLK6, we subjected a YSD mutagenesis library of variants of APPI-3M to sorting by FACS. We collected the presort library and two distinct fractions of the sorted population, namely, the fractions with preferential binding for mesotrypsin and KLK6. Three repetitions of each library fraction were analyzed by high-throughput sequencing (HTS). In the current study, we downloaded the data from GEO accession number GSE289251 and processed it as follows. We located the start of each coding segment by searching for the first two codons of APPI-3M (GAAGTT). We translated the coding segment of each sequence of the three above-mentioned fractions into their respective amino acid sequences. We discarded sequences with zero copies in the presort library and retained sequences with the same length as APPI_{WT} that did not contain internal stop codons or non-desired mutations (i.e., outside the binding interface). We excluded sequences with mutations outside the binding loop to minimize unintended epistatic effects that are not at the focus of this study.

We then counted the number of occurrences of each variant in each library and calculated the frequency of each variant var_i in the library lib , $f_{var_i,lib}$, as:

$$f_{var_i,lib} = \frac{\#reads_{var_i,lib}}{\sum_{j=1}^N \#reads_{var_j,lib}} \quad (1)$$

where $\#reads_{var_i,lib}$ is the number of reads of variant var_i in library lib out of N variants. Thereafter, based on the frequencies of each variant, we calculated the enrichment ratios (ERs) of each variant separately for the mesotrypsin and KLK6 libraries as follows:

$$ER_{var_i,mesotrypsin} = \frac{f_{var_i,mesotrypsin}}{f_{var_i,presort}} \quad (2)$$

$$ER_{var_i,KLK6} = \frac{f_{var_i,KLK6}}{f_{var_i,presort}} \quad (3)$$

where the ER is the ratio of the frequency of a given variant var_i in the mesotrypsin or KLK6 library to the frequency of the same var_i in the presort library.

We then calculated the selectivity ER as follows:

$$ER_{var_i,selectivity} = \frac{ER_{var_i,mesotrypsin}}{ER_{var_i,KLK6}} = \frac{f_{var_i,mesotrypsin}}{f_{var_i,KLK6}} \quad (4)$$

which is the ratio of the frequency of a given variant var_i in the mesotrypsin library to the frequency of the same variant in the KLK6 library.

The measure of the strength of the interaction between mutations at residue positions i and j in a double mutant variant is termed the coupling energy, $\Delta\Delta G_{int}$, which was previously described (Naftaly et al., 2018) as:

$$\Delta\Delta G_{int} = -RT \ln \left(\frac{ER_{ij,selectivity}}{ER_{i0,selectivity} \times ER_{0j,selectivity}} \right) \quad (5)$$

where R is the gas constant, T is the absolute temperature, $ER_{ij,selectivity}$ denotes the selectivity ER of the double mutant, $ER_{i0,selectivity}$ and $ER_{0j,selectivity}$ denote the corresponding single mutants, and 0 represents the APPI-3M background at the respective position.

2.2 | Machine-learning training dataset

By processing the raw DMS data, we generated a dataset containing the amino acid sequences, their three $\log_2$ ER labels (mesotrypsin, KLK6, and selectivity labels), and the $\log_2$ read count of each variant in each library. The dataset included 3177 variants. Of these,

2475 variants appeared in the mesotrypsin library and had a mesotrypsin $\log_2$ ER label, 1512 variants appeared in the KLK6 library and had a KLK6 $\log_2$ ER label, and 810 variants appeared in both libraries and had a selectivity label (Figure 2a). For each sequence, we retained the seven positions within the binding interface (positions 11, 12, 13, 15, 16, 17, and 18), as well as position 14 containing a cysteine residue which was not mutated. In addition, we included the four neighboring positions on each side of that region, which serve as padding for the CNN input, resulting in a total sequence length of 16. The dataset covered all 133 single-mutation variants and included 3043 multiple-mutation variants (Figure 2b).

2.3 | Machine-learning model architecture, input representation, and hyper-parameters search

We explored multiple input representations and model architectures for predicting $\log_2$ ER values. Specifically, we compared five models: (1) a fully connected (FC) network with flattened one-hot encoding, (2) an FC network using pretrained ESM2 embeddings extracted from the full-length protein sequence, (3) an FC network combining one-hot and ESM-derived features, (4) a CNN with one-hot encoding, and (5) a CNN with one-hot encoding concatenated with ESM-derived representations following the pooling layer. We did not compare more complex architectures since we were limited by the relatively small size of our dataset (3177 samples, where 2475 have a mesotrypsin $\log_2$ ER label, and 1512 have a KLK6 $\log_2$ ER label).

For ESM-based representations, we evaluated residue-level embeddings of 640 dimensions. This choice maintains a relatively small number of input features and model parameters, suitable for our dataset size. In addition, as reported before (Vieira et al., 2025), medium-sized ESM-2 models achieve strong performance, often comparable to larger models and in some DMS tasks even slightly outperforming them. We used the average difference in wild-type and variant embeddings over the L positions:

$$\Delta h = \frac{1}{L} \sum_{j=1}^L (h_j^{var} - h_j^{3M}) \quad (6)$$

where $h_j^{var} \in \mathbb{R}^{640}$ and $h_j^{3M} \in \mathbb{R}^{640}$ are the embedding vectors for position j in the variant and the APPI-3M, respectively, produced by the ESM-2 model.

We developed two models (one for each target) to predict $\log_2$ ER. We trained one model over the mesotrypsin labels, which we denoted $mesotrypsin_{model}$, and the other model over the KLK6 labels, which we denoted $KLK6_{model}$ (Figure 2c). We denoted the

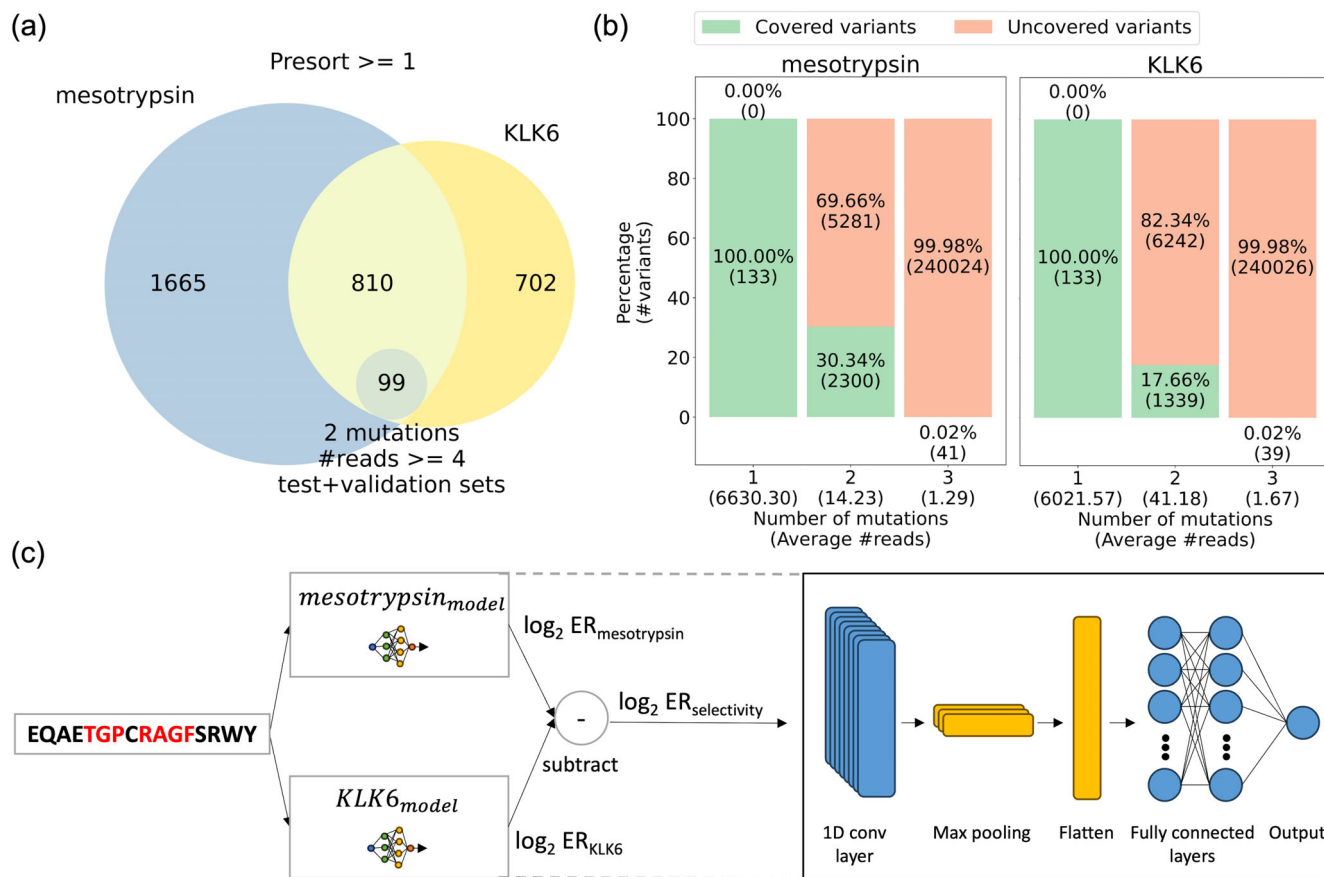


FIGURE 2 Data statistics and model architecture. (a) The overlap between mesotrypsin and KLK6 library fractions, where 99 variants in that overlap (≥ 4 repetitions) were used for test and validation. (b) Percentages of covered and uncovered variants with 1–3 mutations in the dataset. Exact counts and percentages are given, and average reads per variant are indicated below the plots. (c) Schematic representation of the CNN architecture used to predict the $\log_2$ ER values and the subtraction to predict the $\log_2$ selectivity ER.

subtraction of the output of the two models (the two predicted $\log_2$ ERs from *mesotrypsin_{model}* and *KLK6_{model}*) as the *selectivity_{model}* prediction.

To train the models and to select the architecture and hyper-parameters, we held-out a high-quality set of 99 variants that had two mutations, at least four reads in each of mesotrypsin and KLK6, and both mesotrypsin and KLK6 labels. We randomly partitioned this set into 50 variants as the validation set and 49 variants as the test set. For selecting the architecture and the hyper-parameter values, we sampled 100 random hyper-parameter combinations with a pre-defined range for each parameter (Table S1). We selected the hyper-parameter values that led to the highest Pearson correlation on the validation set of each model separately, and the architecture based on the *selectivity_{model}* that achieved the best performance.

The final two models (one for each target) to predict $\log_2$ ER were ensembles of CNNs (Figure 2c). In each model, the input sequence is one-hot encoded, resulting in a matrix of dimensions 20×16 , where the order of amino acids is arbitrary and kept consistent across all training and evaluation data. This matrix serves as

the input to a 1-dimensional convolutional layer. The output from the convolutional layer passes through a max-pooling layer with small pool sizes (2 for *mesotrypsin_{model}* and 3 for *KLK6_{model}* as selected in the hyper-parameter search), which reduces feature dimensionality and limits overfitting while preserving the strongest local activation signals. The resulting max-pooled matrices are then flattened into a numerical vector. That vector is fed into two consecutive fully connected layers with a rectified linear unit (ReLU) activation function. The output from the second fully connected layer serves as the input to a single neuron with a linear activation function, producing the final network prediction.

We additionally implemented a multi-task learning framework. In this model, convolutional layers are shared between tasks; this shared representation is fed to a two-dimensional output layer, corresponding to the predicted two $\log_2$ ERs for mesotrypsin and KLK6 respectively. We trained the multi-task model using the same training procedure and data splits as the single-task models. To ensure a fair comparison, we performed a hyper-parameter search over 100 values and

selected the model achieving the best performance on the validation set.

We implemented the models using Keras Python library with TensorFlow backend (version 2.18.0).

2.4 | Training and evaluation of the models

To increase model robustness, we used the random-ensemble-initialization approach, where we trained 100 models over the same data but with different randomly initialized weights (using standard Glorot initialization, the default for Conv1D and Dense layers in Keras/TensorFlow) and different random training data batches. In addition, so as to rely more on variants with a large statistical sample (a large number of reads), we used the $\log_2$ read counts of each variant in the mesotrypsin or KLK6 libraries as the sample weight in the training process. Then, for prediction, we obtained the final output by averaging the predictions of 100 trained models for each model type (i.e., $mesotrypsin_{model}$ or $KLK6_{model}$). To derive the prediction for the $selectivity_{model}$, we subtracted the average prediction of the $KLK6_{model}$ from that of the $mesotrypsin_{model}$. To assess confidence in the model's prediction, we calculated the coefficient of variation (CV) of a final prediction, which is an average of 100 predictions of the models in the ensemble, as follows:

$$CV = \frac{\sqrt{\frac{1}{100} \sum_{i=1}^{100} (M_i - \bar{M})^2 + \frac{1}{100} \sum_{i=1}^{100} (K_i - \bar{K})^2}}{\left| \frac{1}{100} \sum_{i=1}^{100} M_i - K_i \right|} \quad (7)$$

$$= \frac{\sqrt{\hat{\sigma}_{mesotrypsin}^2 + \hat{\sigma}_{KLK6}^2}}{|\hat{\mu}_{selectivity}|} = \frac{\hat{\sigma}_{selectivity}}{|\hat{\mu}_{selectivity}|}$$

where M and K are the 100-long prediction vectors of mesotrypsin and KLK6, respectively, $\hat{\sigma}$ is the estimated standard deviation, and $\hat{\mu}$ is the estimated mean. To evaluate the prediction performance of each model, we calculated the Pearson and Spearman correlation between the predictions and the experimental $\log_2$ ER values of a held-out test set of 49 double-mutant variants.

2.5 | Leave-one-position-out evaluation

To assess model robustness to predicting the effect of mutations at unobserved residue position and mitigate potential test–train leakage, we performed a leave-one-position-out evaluation. In this evaluation, we held out

all variants containing mutations at a given residue position (including both single and combinatorial mutations) as the test set. We then trained a model on all variants with mutations at the remaining six positions. We repeated this process for each position, ensuring that every residue position served as an independent test set. We evaluated model performance for mesotrypsin, KLK6, and selectivity predictions by Pearson and Spearman correlation coefficients.

2.6 | Evaluation of prediction performance as a function of sequencing depth

To evaluate the $mesotrypsin_{model}$ and $KLK6_{model}$ performance as a function of the sequencing depth, we used the same test set of 49 variants that had two mutations. We randomly selected 10 times 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.

2.7 | YSD assays for experimental validation

Given the resource-intensive nature of producing soluble proteins and our need to test a large number of variants, we employed YSD as a preliminary selectivity screening technique to experimentally validate the $selectivity_{model}$ and to identify selective APPI variants (Figure S1). This technique allowed us to determine the binding levels of APPI variants to the two target proteases (in their soluble form) by displaying each variant on the surface of yeast cells in the presence of fluorescently labeled target proteases. The fluorescence signal thus served as a measure of the binding level of each displayed variant to the respective protease. We selected variants according to the following three categories, where different combinations of the single-mutant variants formed the double-mutant variants: (i) “single mutation-observed,” which included variants with a single mutation (all of which were observed in the DMS data); (ii) “double mutations-observed,” which included variants with two mutations that were observed in the DMS data; and (iii) “double mutations-unobserved,” which included variants with two mutations that were not observed in the DMS data. The YSD technique and the construction and cloning of APPI variants in YSD format are described in the Data S1 (Methods).

To calculate the binding selectivity of each APPI protein variant, we divided the fluorescence intensity of that variant bound to a mesotrypsin-FITC conjugate by the fluorescence intensity of the variant bound to KLK6-APC conjugate on a logarithmic scale. To account for the variability in protein expression levels in the different APPI variants, we normalized the fluorescence intensities to the expression levels of each variant [determined using a phycoerythrin (PE)-labeled anti c-Myc antibody]. Additionally, to establish a reference protein for comparison of the selectivity of the different YSD variants, we used the APPI_{P13Y,G17Y} variant for further normalization, as it exhibited a moderate log₂ selectivity ER and a moderate binding level for both targets and thus served as a good reference for comparison. We further normalized the fluorescence binding signals of all variants to the binding level of this variant. We then used the normalized data (Figure S2a–d) to calculate the selectivity index_{YSD} based on the YSD assay; this index reflects the preference of each APPI variant for binding to mesotrypsin vs. KLK6, as follows:

$$\text{Selectivity index}_{\text{YSD}} =$$

$$\log_2 \left(\frac{\text{fluorescence}_{\text{mesotrypsin}} / \text{fluorescence}_{\text{mesotrypsin, ref}}}{\text{fluorescence}_{\text{KLK6}} / \text{fluorescence}_{\text{KLK6, ref}}} \right) \quad (8)$$

where $\text{fluorescence}_{\text{mesotrypsin}}$ and $\text{fluorescence}_{\text{KLK6}}$ are the fluorescence intensity of the variant obtained from binding to mesotrypsin-FITC or KLK6-APC conjugates, respectively, and $\text{fluorescence}_{\text{mesotrypsin, ref}}$ and $\text{fluorescence}_{\text{KLK6, ref}}$ are the fluorescence intensity of the reference variant APPI_{P13Y,G17Y} obtained from binding to the same conjugates.

2.8 | Inhibition assays for experimental validation

We used the competitive inhibition equation (Equation 8) and the Morrison tight-binding equation (Equation 9) to calculate the inhibition constant (K_i) for each purified APPI variant:

$$V = \frac{V_{\max} * [S]}{K_m * \left(1 + \frac{[I]}{K_i}\right) + [S]} \quad (9)$$

where V is the formation rate of the cleavage product [p-nitroaniline (pNA) for mesotrypsin and 7-amino-4-methyl coumarin (AMC) for KLK6], $v_{\max}$ is the maximal rate of substrate (Z-GPR-pNA for mesotrypsin and BOC-FSR-AMC for KLK6) cleavage, K_m is the Michaelis constant, K_i is the inhibitor (APPI variant) constant,

$[S]$ is the substrate concentration, and $[I]$ is the inhibitor concentration (Kakkar et al., 1999):

$$\frac{V_0 - V_i}{V_i} = \frac{[I]_0}{K_i * \left(1 + \frac{[S]_0}{K_m}\right)} \quad (10)$$

where V_0 is the steady-state rate in the absence of the inhibitor, V_i is the steady-state rate in the presence of the inhibitor, K_m is the Michaelis constant for substrate cleavage and $[S]_0$ and $[I]_0$ are the initial concentrations of the substrate and inhibitor, respectively. We performed calculations by using K_m values of $24.66 \pm 1.30 \mu\text{M}$ for mesotrypsin and $329.30 \pm 2.50 \mu\text{M}$ for KLK6 (Cohen et al., 2018).

We then calculated the selectivity index_{inhibition} of each purified variant as the logarithmic ratio of its K_i value against KLK6 to the K_i value for the same variant against mesotrypsin normalized to the corresponding K_i of APPI-3M, as follows:

$$\text{selectivity index}_{\text{inhibition}} = \log_2 \left(\frac{\frac{K_i, \text{KLK6}}{K_{i3M}, \text{KLK6}}}{\frac{K_i, \text{mesotrypsin}}{K_{i3M}, \text{mesotrypsin}}} \right) \quad (11)$$

where $K_i, \text{KLK6}$ and $K_i, \text{mesotrypsin}$ are the inhibition constants of the variant for KLK6 and mesotrypsin, respectively, and $K_{i3M}, \text{KLK6}$ and $K_{i3M}, \text{mesotrypsin}$ are the inhibition constants of APPI-3M for these proteases. Protein production and purification and the determination of the inhibition constant (K_i) of APPI variants to mesotrypsin and KLK6 are described in Data S1 (Methods).

3 | RESULTS

3.1 | Training and evaluating ML models to predict the selectivity of APPI-3M variants

We developed ML models to predict, for any given variant, the log₂ selectivity ER, which is an experimental measurement correlated to the selectivity of the variant. We trained two models, designated *mesotrypsin_{model}* and *KLK6_{model}*, to predict the log₂ ER values based on a specific library. The training dataset comprised single- and multi-mutation variants (Figure 2b), along with their respective log₂ ER labels (i.e., mesotrypsin and KLK6 labels). We held out 99 high-quality variants, of which half were used for tuning hyper-parameters and the other half were used for testing the trained model. We used the remaining dataset, consisting of single-mutant and multiple-mutant variants with lower read counts (Figure 2b) for training. During the training process, we employed the logarithmic value of the read

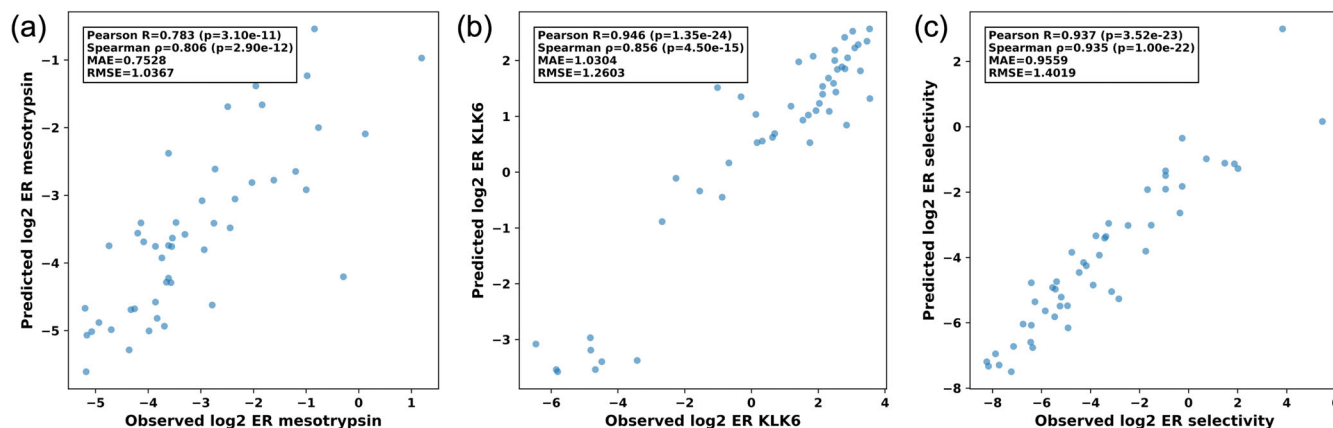


FIGURE 3 Predictive performance (correlation between predicted and observed $\log_2$ ERs) of the ML models for a held-out high-quality test set of 49 double-mutant variants for (a) *mesotrypsin*_{model}, (b) *KLK6*_{model}, and (c) *selectivity*_{model}.

count for each variant for sample weighting. We evaluated the performance of the *mesotrypsin*_{model} and the *KLK6*_{model} in predicting $\log_2$ ERs in terms of the Pearson correlation of predicted vs. observed $\log_2$ ERs.

We evaluated five architectures differing in input representation and model type. For each architecture, we selected the best-performing hyper-parameter combination separately for *mesotrypsin*_{model} and *KLK6*_{model} prediction, and the resulting models were used to compute *selectivity*_{model} performance. The FC models with one-hot, one-hot combined with ESM2 representation, and ESM2-only inputs achieved Pearson correlations of 0.920, 0.916, and 0.914 for the *selectivity*_{model}, respectively. The CNN models yielded Pearson correlations 0.929, and 0.927 for *selectivity*_{model}, for the one-hot and one-hot combined with ESM2 representation models, respectively (Figure S3). Overall, the CNN with one-hot encoding achieved the highest performance across tasks. Accordingly, we selected the CNN with one-hot input as our final model and used it in subsequent analyses. Moreover, most hyper-parameter combinations for the CNN model led to high prediction performance (>0.88 Pearson correlation, Figure S4), demonstrating relative robustness to hyper-parameter choices.

To assess whether shared representation learning across proteases could further improve performance, we additionally trained a multi-task CNN in which shared convolutional layers feed a two-dimensional output layer simultaneously predicting the two $\log_2$ ERs for mesotrypsin and KLK6. The multi-task model achieved a Pearson correlation of 0.9287 for selectivity prediction, which was slightly lower than the corresponding single-task CNN model (Pearson correlations 0.9291). Thus, while multi-task learning is conceptually motivated by shared biochemical features of APPI interactions, it did not improve predictive performance in our dataset, and we chose not to use it in downstream analyses.

Both the *mesotrypsin*_{model} and the *KLK6*_{model} exhibited high performance on the held-out test set (Figure 3a,b): Pearson correlations of 0.783 and 0.946,

respectively (p values = $3.10\text{e-}11$ and $1.35\text{e-}24$, respectively) were obtained. In addition, strong rank-based agreement was observed, with Spearman correlations of 0.806 and 0.856, respectively (p values = $2.90\text{e-}12$ and $4.50\text{e-}15$, respectively). Error-based metrics further supported model performance, with mean absolute errors (MAE) of 0.753 and 1.030 and root mean square errors (RMSE) of 1.037 and 1.260 for the *mesotrypsin*_{model} and the *KLK6*_{model}, respectively. The *selectivity*_{model}, defined as *mesotrypsin*_{model} - *KLK6*_{model}, also exhibited high performance on the held-out test set (Figure 3c). A Pearson correlation of 0.937 (P value = $3.52\text{e-}23$) and a Spearman correlation of 0.935 (P value = $1.00\text{e-}22$) were observed between predicted and observed log ratio of selectivity ERs. The corresponding MAE and RMSE values were 0.956 and 1.402, respectively.

To further assess model robustness to predicting the effect of unobserved mutations and mitigating potential test–train leakage, we performed a leave-one-position-out evaluation. Performance varied across positions, with high prediction performance observed for positions 11 and 13, where Pearson correlations for selectivity were 0.908 and 0.959, respectively. High performance was also observed for KLK6 predictions across most positions. However, performance was reduced for certain positions, such as 12, 15, and 17, where lower or negative correlations were observed for mesotrypsin and KLK6 predictions. Overall, while performance varied across positions, the model still demonstrated predictive capability under this challenging evaluation, suggesting some ability to generalize to unobserved mutations (Figure S5).

3.2 | Effect of sequencing depth and sample weighing on the performance of the ML models

Next, we aimed to test how the sequencing depth influences the performance of our ML models. The

sequencing depth affects the statistical sample size of each variant and consequently the robustness of the log₂ ER values measured in the experiment. We assessed how the number of reads influences prediction performance of the models *mesotrypsin*_{model} and *KLK6*_{model} by resampling in silico the whole dataset to generate simulated datasets with different numbers of DNA library sequencing reads. We trained the *mesotrypsin*_{model} and *KLK6*_{model} on each simulated dataset and tested each model on the held-out test set.

As expected, predictive performance generally improved with increasing dataset size for both proteases (Figure S6). This effect was most apparent for mesotrypsin, where the Pearson correlation increased from 0.504 using 10% of the data to 0.770 using the full dataset (slightly below the previously reported full dataset result due to a different random seed and number of shuffles). In contrast, KLK6 predictions remained consistently high even at lower sampling levels, with Pearson correlations above 0.887 across all sequencing depths. These results suggest that, despite incomplete sampling of the sequence space, the dataset contains sufficient information to support robust model training, while indicating that increased sequencing depth could further improve performance, particularly for mesotrypsin prediction.

To evaluate the impact of sample weighting based on sequencing depth, we compared three training schemes: (i) no sample weighting, (ii) raw read-count-based weighting, and (iii) log₂-transformed read-count weighting (Figure S7). Overall, both the unweighted and log₂-weighted models achieved similarly high predictive performance, with average Pearson correlations of 0.9146 and 0.9159, respectively, on held-out validation set. In contrast, using raw read counts as weights substantially degraded performance, particularly for mesotrypsin prediction (0.665 compared to 0.897 for the log₂-weighted model), indicating that untransformed counts overemphasize high-coverage variants and impair generalization. These results suggest that while incorporating read depth information does not strongly improve predictive performance, log-transformation is important for preventing extreme weighting effects.

3.3 | Identifying positions that affect the selectivity of APPI for mesotrypsin vs. KLK6

To investigate the binding landscape of the interaction between APPI-3M and each of the two target serine proteases, mesotrypsin and KLK6, we predicted, using the *selectivity*_{model}, all possible single-mutant and double-mutant log₂ selectivity ERs. Given the predictions, we calculated the selective epistasis effects, that is, the deviation of a double-mutant variant's predictions from the sum of its corresponding single-mutant

predictions. Positive selective epistasis is indicated when a combination of two mutations results in a change in the log₂ selectivity ER ratio (mesotrypsin vs. KLK6) that is greater than the sum of the individual changes in selectivity, whereas negative selective epistasis is indicated when the combined effect is less than expected based on the sum of the individual effects.

We revealed that mutations at positions 12 and 17 produced the highest levels of positive selective epistasis (Figure 4a). Similarly, a mutation at position 12 combined with a mutation at position 15 or 16 yielded positive selective epistasis. In contrast, mutations at position 13 combined with mutations at another position resulted in negative selective epistasis, as did most mutations at position 11, where mutations at both positions 11 and 13 (cold spots for mesotrypsin) exhibited the highest levels of negative selective epistasis. A possible explanation for the pronounced negative epistasis, particularly in combinations involving two cold spots for mesotrypsin, is that individual mutations at these positions already yield high predicted selectivity for the target protease, and therefore combining them does not enhance selectivity beyond the sum of their individual effects.

Additionally, we calculated the contribution of each position and amino acid residue to the selectivity of double-mutant variants for one of the two protease targets (Figure 4b). The results confirmed previous findings and demonstrated that adding a second mutation at position 13 consistently increased selectivity for mesotrypsin. Likewise, most substitutions at position 11 led to enhanced selectivity for mesotrypsin. Conversely, mutations at position 17 consistently increased selectivity for KLK6. These findings align with the known roles of position 13 as a cold spot for mesotrypsin and position 17 as a cold spot for KLK6 (Naftaly et al., 2018). Furthermore, these selectivity-switch residues (13 and 17) enable a shift in selectivity from one protease to the other (mesotrypsin to KLK6 or vice versa) (Naftaly et al., 2018).

3.4 | Selecting candidate variants for experimental validation

The next step was to experimentally validate, by using YSD, our *selectivity*_{model} in predicting double-mutant variants, followed by analysis of the predicted selective epistasis effects and residue contributions to selectivity. For this experimental validation, we selected 42 variants, including 20 double-mutant variants, 21 single-mutant variants corresponding to the double-mutant variants, and the APPI-3M scaffold variant (Table S2). The double-mutant variants were selected on the basis of two criteria: (i) known cold-spot selectivity-switch positions (11 and 13 for mesotrypsin, and 17 and 18 for KLK6) (Naftaly et al., 2018) and their combinations in two mutations, and (ii) variants with double mutations

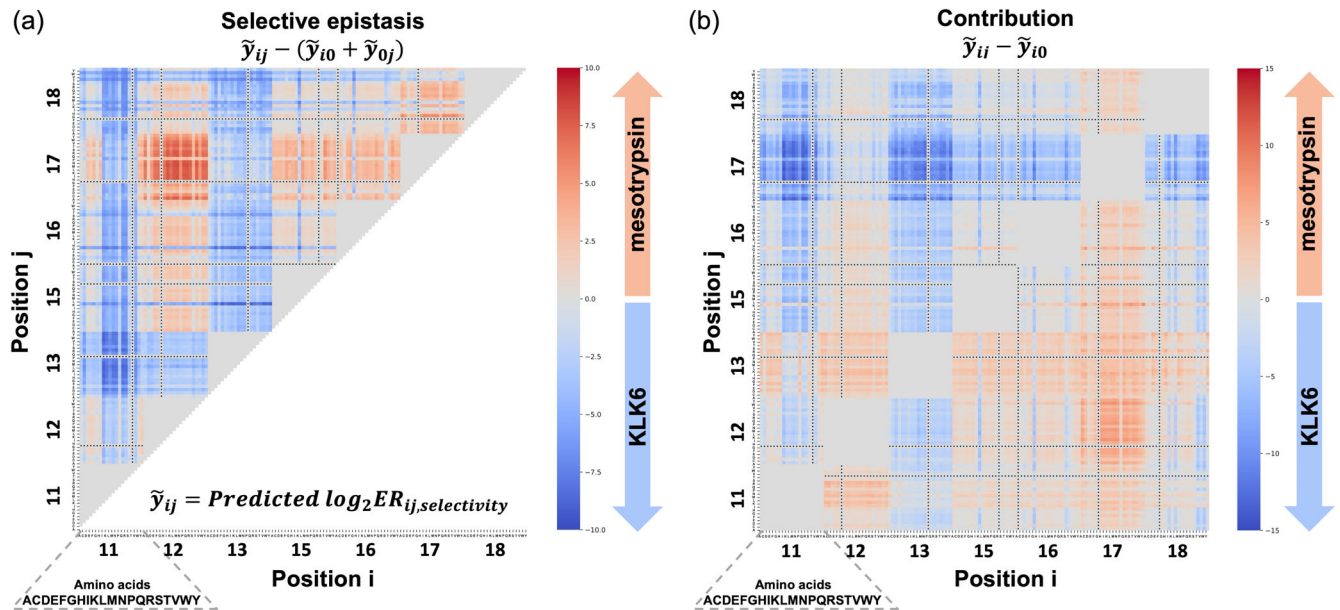


FIGURE 4 Analysis of double-mutant vs. single-mutant selectivity predictions. (a) Selective epistasis heatmap. (b) Double-mutation selectivity heatmap. Each square is a 20 × 20 matrix containing the difference in predicted log₂ selectivity ER values of double- and single-mutant couplings between a mutation at position i and a mutation at position j.

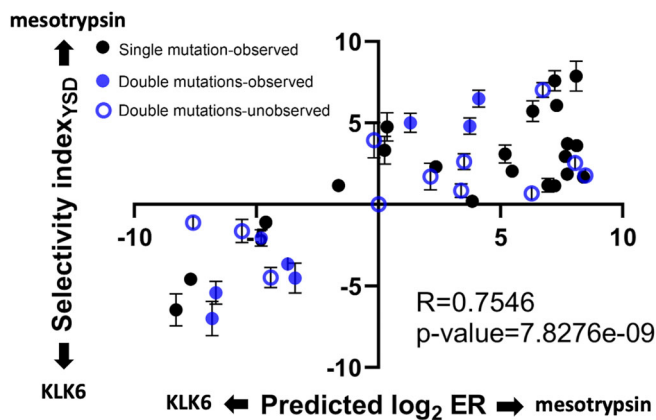


FIGURE 5 Experimental validation of selectivity predictions based on YSD measurements. Correlation of log₂ selectivity ER, as estimated by the *selectivity_{model}*, with the selectivity index_{YSD} determined using the YSD assay.

that exhibited relatively high selective epistasis. Seventeen out of the 20 double-mutant variants awarded high-confidence predictions (i.e., a CV smaller than 1.5), except for three variants where the absolute predicted value was smaller than 1, leading to a higher CV (Figure S8).

3.5 | Experimental validation by a YSD screen

We used YSD as a preliminary selectivity screening technique to test the utility of the *selectivity_{model}* for the

identification selective single or double APPI variants, both for variants observed in the DMS data and for unobserved variants. The screening step revealed that predicted log₂ selectivity ER values obtained from the *selectivity_{model}* correlated well with the selectivity index_{YSD} (Figure 5) (Pearson correlation 0.755; P value = 7.828e-09). Specifically, most variants predicted to exhibit selectivity toward KLK6 or mesotrypsin did indeed consistently demonstrate such selectivity. Notably, the *selectivity_{model}* accurately predicted the selectivity preferences of double-mutant variants that were not observed in the DMS data, emphasizing the model's generalizability capabilities. However, two variants with intermediate log₂ selectivity ER values that were predicted to exhibit selectivity to KLK6 showed slightly stronger binding to mesotrypsin, thereby suggesting that the *selectivity_{model}* is less accurate for some variants predicted with moderate selectivity scores. In addition, the *selectivity_{model}* showed some inconsistency in predicting the exact level of selectivity, with a number of variants that had been predicted to be highly selective for either mesotrypsin or KLK6 showing instead only a moderate level of selectivity by YSD.

To evaluate how well the model predicts selective epistasis effects, we compared experimentally measured selective epistasis values (based on YSD measurements) with selective epistasis inferred from *selectivity_{model}* predictions (Figure S9). Overall, the *selectivity_{model}* captured the qualitative selective epistatic landscape, correctly reproducing the selective epistatic trend in 8 out of 10 double-mutant variants. Notably, APPI_{G17Y,F18K} and APPI_{G17Y,F18R} both

involving the same substitution at position 17, were predicted in the wrong direction. However, *selectivity_{model}* overestimated the magnitude of epistatic effects in several cases, as reflected by longer deviations between predicted and measured values. For example, variant such as APPI_{T11K,P13I} and APPI_{T11M,P13D} exhibit strong negative epistasis experimentally, which was correctly identified in direction but somewhat exaggerated in magnitude by the model. These results indicate that while the *selectivity_{model}* epistasis is not perfectly calibrated quantitatively, it provides a reliable qualitative approximation of experimentally observed mutational interactions.

3.6 | Experimental validation by inhibition assays

We followed our preliminary YSD screen by validation of the binding selectivity using purified soluble APPI variants. To confirm the selectivity of single- and double-mutant variants through inhibition assays, we chose highly selective APPI variants based on two criteria: (i) highest or lowest predicted $\log_2$ selectivity ER values, and (ii) highest or lowest selectivity index_{YSD} scores (Figure S2a,b). In addition, we also produced and purified single- and double-mutant variants that exhibited both intermediate $\log_2$ selectivity ER values and intermediate selectivity index_{YSD} scores (Figure S2c,d). The APPI-3M scaffold and 10 of the 18 selected variants had been produced in previous work, and their K_i inhibition constants were therefore available (Cohen et al., 2016; Naftaly et al., 2018). The production and purification procedure for APPI variants and K_i measurements are described under Supplementary Results and shown in Figures S10–S13.

The $\log_2$ selectivity ER predictions correlated well with the selectivity index_{inhibition} derived from the K_i values of the purified proteins, including both observed and unobserved variants (Figure 6a,b). However, the *selectivity_{model}* was less accurate for some variants predicted with intermediate $\log_2$ selectivity ER values, such as APPI_{T11I,G17S}, APPI_{P13L,G17S} and APPI_{T11R,G17V}, compared to extreme selectivity ER values. Those three double-mutant variants were predicted to exhibit moderate selectivity, with a preference for KLK6 but, in practice, showed a stronger affinity for mesotrypsin when purified (Figure 6a). Similarly to the results obtained from selectivity index_{YSD} derived from YSD measurements, these results may suggest that the predictions of the *selectivity_{model}* for variants with moderate $\log_2$ selectivity ER values are less accurate than those for variants with a strong preference for one target vs. the other. This reduced accuracy may result from the design of the sorting experiment, in which only variants showing clear binding preference for either mesotrypsin or KLK6 were collected. As a result, variants with no preference or only moderate selectivity were underrepresented in the training dataset.

The selectivity index_{YSD} derived from YSD measurements of APPI variants precisely reflected the selectivity predictions, yielding a Pearson correlation of 0.8728 (P value = 2.2971e-06) (Figure S14b). In contrast, the K_i -based measurements of purified APPI variants yielded a lower Pearson correlation of 0.5869 (P value = 0.0105) (Figure 6a). The superior precision of the YSD-based determination of selectivity is likely attributable to the experimental setup for the YSD-based measurements and the DMS (used as an input to the *selectivity_{model}* to obtain $\log_2$ selectivity ER values): in the YSD set-up, the two target proteases—

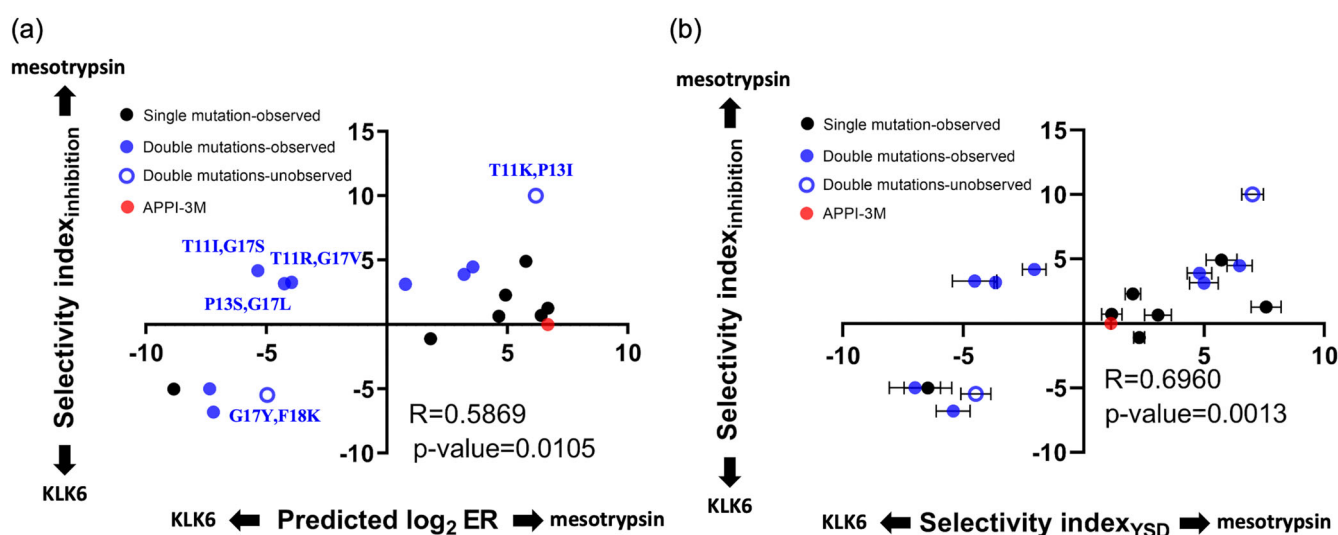


FIGURE 6 Validation of selectivity prediction determined via the inhibition constants obtained for purified vs. YSD APPI variants. (a) Correlation between the selectivity index_{inhibition} determined by affinity (K_i) constants of the purified proteins and selectivity $\log_2$ ER determined from DMS, followed by prediction by the *selectivity_{model}*. (b) Correlation between the selectivity index_{YSD}, determined by YSD and the selectivity index_{inhibition} based on the K_i values of purified variants.

mesotrypsin and KLK6—are both present in the reaction tube [thereby mimicking the competitive binding observed during the DMS (Naftaly et al., 2018)] and interact with the APPI variant immobilized to a surface. This experimental setup differs from that for the K_i measurements in which the soluble form of the APPI variant competes with a single protease substrate (either mesotrypsin or KLK6) for binding at its active site.

Variants with predominantly selective binding to mesotrypsin included those with mutations at position 13, while mutations at position 17 enhanced selectivity for KLK6 (Figure 6, Table S3). Examples include APPI_{T11K,P13I} and APPI_{G17Y,F18K} with binding preferences for mesotrypsin and KLK6, respectively (Figure 6a). These variants were not observed in the DMS data (i.e., were not used for *selectivity_{model}* training) and were thus predicted solely by the *selectivity_{model}*. APPI_{T11K,P13I} exhibited tight-binding inhibition behavior with mesotrypsin, with a K_i value of 1.8 ± 1.0 nM (Figure S12h), and exemplified a classic inhibition model against KLK6, with a K_i constant of 6.9 ± 1.2 μ M (Figure S13c). This K_i value translates into a 3832-fold increase in selectivity for mesotrypsin over KLK6 (Table S3) and also indicates a 2254-fold increase in selectivity compared to APPI_{WT} (Cohen et al., 2018). To the best of our knowledge, APPI_{T11K,P13I} is the most selective mesotrypsin inhibitor vs. KLK6 reported to date. The APPI_{G17Y,F18K} variant also exhibited classic inhibition behavior with mesotrypsin, with a K_i value of 458.6 ± 46.0 nM (Figure S12g) and a classic inhibition model with KLK6 with a K_i constant of 38.0 ± 7.9 nM (Figure S13f). This K_i value represents a 12.1-fold increase in selectivity for KLK6 over mesotrypsin (Table S3). The above findings, obtained for variants carrying two mutations that were not observed in the DMS data, are in agreement with the *selectivity_{model}* predictions.

4 | DISCUSSION

In this study, we developed an ML-based method to predict the selectivity of multi-mutation APPI variants for homologous serine protease targets across a broad range of binding affinities: By leveraging deep neural networks trained on our previously generated DMS data for the scaffold protein APPI-3M, we predicted the selectivity ER of APPI variants for mesotrypsin vs. KLK6. We followed the computational prediction by experimental validation using APPI protein variants produced on a YSD platform. This platform provided a rapid and resource-efficient solution for identifying highly selective protein-based inhibitors, allowing us to prioritize promising protein variants for further purification and evaluation via target inhibition assays. The combined computational-experimental methodology accurately predicted binding selectivity and selective epistatic interactions, enabling

the identification of novel APPI variants that were not observed in the original library. Among those variants were highly selective inhibitors for mesotrypsin and KLK6, including the most selective mesotrypsin inhibitor vs. KLK6 reported to date.

One of the key findings of our study was the identification of specific amino acid positions, particularly positions 11, 13, 17, and 18, that play a critical role in determining target selectivity. Mutations at positions 11 and 13, which are cold spots for mesotrypsin, enhanced selectivity for mesotrypsin, while mutations at positions 17 and 18 increased selectivity for KLK6. This position-specific influence on selectivity was supported by both our *selectivity_{model}* and experimental validation via YSD and inhibition assays.

Another major novelty of this work was the demonstration of non-additive, epistatic effects of mutations on selectivity. By comparing predicted selectivity values of double-mutant variants to the summed effects of their single mutations, we uncovered multiple cases of both positive and negative selective epistasis. For example, combinations involving positions 12 and 17 showed strong positive epistasis, while combinations involving positions 11 and 13, each already known as cold spots for mesotrypsin, exhibited negative epistasis. These non-additive interactions highlight the necessity of taking into consideration both the presence of multiple mutations and epistasis in selectivity prediction and protein engineering.

Furthermore, our results showed that epistatic interactions are not rare exceptions in the mutational landscape and that they play a crucial role in determining the functional outcome of mutations. Models that do not take these effects into consideration may miss promising mutation combinations or wrongly predict how engineered variants will behave. In our case, the ability to detect and quantify selective epistasis allowed us to rationally select and validate double-mutant variants with enhanced or shifted target specificity.

The promising results described above notwithstanding, the study has several limitations. First, while our *selectivity_{model}* performed well in predicting selectivity, its performance was less successful for variants with moderate selectivity scores compared to scores at the high or low extremes. This outcome suggests potential biases in the training dataset, with fewer variants representing intermediate selectivity regimes, as seen in the comparison between DMS measurements and K_i -based inhibition constants (Figure S14a). From a protein design perspective, this limitation implies that the model is most reliable for identifying variants with strongly enhanced or diminished selectivity, which are often of primary interest in early-stage screening. In addition, while multiple targets may exist in a cellular environment, the current study focuses on selectivity prediction between only two proteases. Still, it is a step toward ML-based prediction of protein selectivity in

more complex cellular environments as most studies focus only on a single protease. Second, we validated our predictions primarily by using YSD, which, while effective, may not fully reproduce the conditions of soluble protein interactions. The differences between YSD-based selectivity indexes and K_i -based inhibition constants (Figure 6b) highlight the need for further improvement in the experimental validation techniques. Third, our study focused on double-mutant variants, leaving the combinatorial effects of higher-order mutations unexplored. Future studies focusing on variants with more than two mutations will be necessary to fully capture the complexity of these APPI-serine protease PPIs.

In light of the above, we aim to expand our study in several directions. To improve predictive performance for K_i inhibition constants, we plan to develop models based on more advanced ML techniques—advanced architectures, such as graph neural networks, and unsupervised learning techniques, including large protein language models (Livesey & Marsh, 2023). Additionally, we will expand the training dataset to include a broader range of multiple-mutant variants to enhance model generalizability. Integrating data from YSD validation as part of a transfer-learning strategy could further improve predictions, as our analysis indicates that YSD-derived selectivity indexes are better correlated with K_i inhibition constants than current model predictions. Furthermore, cellular experimental validation followed by in-vivo assays will be crucial in assessing the therapeutic potential of the designed inhibitors. More generally, we plan to extend this approach to additional protein families and inhibitor scaffolds to assess how broadly it can support the prediction of PPI selectivity.

Given the inherent challenges of distinguishing selectivity for two protease targets with similar structural characteristics and predicting the effects of multiple mutations on protein selectivity, our findings underscore the combined effectiveness of YSD screening and ML-driven predictions in discovering potent double-mutant inhibitors that specifically target highly conserved regions of serine proteases. More generally, by reducing the need for extensive experimental screening, our ML-based approach can accelerate the discovery of novel protein variants with optimized selectivity and binding affinity. This methodology could be extended beyond protease inhibitors to other protein classes, including antibodies, unlocking significant therapeutic potential.

AUTHOR CONTRIBUTIONS

Reut Meiri: Data curation; investigation; formal analysis; visualization; writing – review and editing; writing – original draft. **Niv Papo:** Conceptualization; methodology; formal analysis; supervision; funding acquisition; visualization; project administration; writing – review and editing; writing – original draft. **Evette S. Radisky:** Conceptualization; funding acquisition. **Yaron Orenstein:** Conceptualization; methodology; formal analysis; supervision; project

administration; visualization; funding acquisition; writing – original draft; writing – review and editing. **Oz Reuveni:** Data curation; investigation; validation; formal analysis; writing – original draft; writing – review and editing; visualization. **Michal Levi:** Data curation; investigation; validation; formal analysis; visualization; writing – review and editing; writing – original draft.

ACKNOWLEDGMENTS

We thank Dr. Uzi Hadad, Ilse Katz Institute for Nano-scale Science & Technology, for consultation and technical assistance. Reut Meiri thanks the Israel Data Science and AI Initiative for providing cloud credit for this work. The research was funded in part through Teva's fellowship and mentoring program, The National BioInnovators Forum. This work was supported by the National Institute of Health (R01GM144393) to Yaron Orenstein, Niv Papo and Evette S. Radisky; 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 Yaron Orenstein; 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 Niv Papo.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Unprocessed FASTQ files were obtained from the GEO under accession number GSE289251. Code and processed data used in this study are available at <https://github.com/OrensteinLab/APPI-KLK6-Meso> and have been archived at [10.5281/zenodo.15823190](https://zenodo.org/record/15823190).

ORCID

Reut Meiri  <https://orcid.org/0009-0003-9640-448X>
Niv Papo  <https://orcid.org/0000-0002-3583-3112>
Yaron Orenstein  <https://orcid.org/0000-0002-7056-2418>

REFERENCES

- Bayani J, Diamandis EP. The physiology and pathobiology of human kallikrein-related peptidase 6 (KLK6). *Clin Chem Lab Med*. 2012; 50:211–33.
- Bayés A, Tsetsenis T, Ventura S, Vendrell J, Aviles FX, Sotiropoulou G. Human kallikrein 6 activity is regulated via an autoprolytic mechanism of activation/inactivation. 2004;385:517–24.
- Blaber SI, Yoon H, Scarisbrick IA, Juliano MA, Blaber M. The autolytic regulation of human kallikrein-related peptidase 6. *Biochemistry*. 2007;46(17):5209–17. <https://doi.org/10.1021/bi6025006>
- Coates P, Hall P. The yeast two-hybrid system for identifying protein–protein interactions. *J Pathol*. 2003;199:4–7.
- Cohen I, Kayode O, Hockla A, Sankaran B, Radisky DC, Radisky ES, et al. Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *Biochem J*. 2016;473:1329–41.

- Cohen I, Naftaly S, Ben-Zeev E, Hockla A, Radisky ES, Papo N. Pre-equilibrium competitive library screening for tuning inhibitor association rate and specificity toward serine proteases. *Biochem J*. 2018;475:1335–52.
- Curran A, Kwok J, Sadler JC, Bell EL, Swainston N, Ababi M, et al. GeneORator: an effective strategy for navigating protein sequence space more efficiently through Boolean OR-type DNA libraries. *ACS Synth Biol*. 2019;8:1371–8.
- Diederichs S, Bulk E, Steffen B, Ji P, Tickenbrock L, Lang K, et al. S100 family members and trypsinogens are predictors of distant metastasis and survival in early-stage non-small cell lung cancer. *Cancer Res*. 2004;64:5564–9.
- Ebrahimi SB, Samanta D. Engineering protein-based therapeutics through structural and chemical design. *Nat Commun*. 2023;14:2411.
- Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. *Nat Methods*. 2014;11:801–7.
- Fry DC, Vassilev LT. Targeting protein–protein interactions for cancer therapy. *J Mol Med*. 2005;83:955–63.
- Gelman S, Fahlberg SA, Heinzelman P, Romero PA, Gitter A. Neural networks to learn protein sequence–function relationships from deep mutational scanning data. *Proc Natl Acad Sci U S A*. 2021;118:e2104878118.
- Han S, Lee CW, Trevino JG, Hughes SJ, Gas J. Autocrine extrapancreatic trypsin 3 secretion promotes cell proliferation and survival in esophageal adenocarcinoma. *PLoS One*. 2013;8:e76667.
- Hockla A, Miller E, Salameh MA, Copland JA, Radisky DC, Radisky ES. PRSS3/Mesotrypsin is a therapeutic target for metastatic prostate cancer. *Mol Cancer Res*. 2012;10:1555–66.
- Hockla A, Radisky DC, Radisky ES. Mesotrypsin promotes malignant growth of breast cancer cells through shedding of CD109. *Breast Cancer Res Treat*. 2010;124:27–38.
- Horne J, Shukla D. Recent advances in machine learning variant effect prediction tools for protein engineering. *Ind Eng Chem Res*. 2022;61:6235–45.
- Hsu C, Nisonoff H, Fannjiang C, Listgarten J. Learning protein fitness models from evolutionary and assay-labeled data. *Nat Biotechnol*. 2022;40:1114–22.
- Jiang G, Cao F, Ren G, Gao D, Bhakta V, Zhang Y, et al. PRSS3 promotes tumour growth and metastasis of human pancreatic cancer. *Gut*. 2010;59:1535–44.
- Jiang L, Yuan C, Huang M. A general strategy to inhibit serine protease by targeting its autolysis loop. *FASEB J*. 2021;35:e21259.
- Jubb HC, Pandurangan AP, Turner MA, Ochoa-Montañó B, Blundell TL, Ascher DB. Mutations at protein–protein interfaces: small changes over big surfaces have large impacts on human health. *Prog Biophys Mol Biol*. 2017;128:3–13.
- Kakkar T, Boxenbaum H, Mayersohn M. Estimation of K_i in a competitive enzyme-inhibition model: comparisons among three methods of data analysis. *Drug Metab Dispos*. 1999;27:756–62.
- Leander M, Liu Z, Cui Q, Raman S. Deep mutational scanning and machine learning reveal structural and molecular rules governing allosteric hotspots in homologous proteins. *Elife*. 2022;11:e79932.
- Livesey BJ, Marsh JA. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Mol Syst Biol*. 2020;16:e9380.
- Livesey BJ, Marsh JA. Advancing variant effect prediction using protein language models. *Nat Genet*. 2023;55:1426–7.
- Ma H, Hockla A, Mehner C, Coban M, Papo N, Radisky DC, et al. PRSS3/Mesotrypsin and kallikrein-related peptidase 5 are associated with poor prognosis and contribute to tumor cell invasion and growth in lung adenocarcinoma. *Sci Rep*. 2019;9:1844.
- Ma R, Ye X, Cheng H, Ma Y, Cui H, Chang X. PRSS3 expression is associated with tumor progression and poor prognosis in epithelial ovarian cancer. *Gynecol Oncol*. 2015;137:546–52.
- Miton CM, Tokuriki N. How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci*. 2016;25:1260–72.
- Naftaly S, Cohen I, Shahar A, Hockla A, Radisky ES, Papo N. Mapping protein selectivity landscapes using multi-target selective screening and next-generation sequencing of combinatorial libraries. *Nat Commun*. 2018;9:3935.
- Nooren IMA, Thornton JM. Diversity of protein–protein interactions. *EMBO J*. 2003;22:3486–92.
- Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein–protein interaction 1. *J Mol Biol*. 1997;271:511–23.
- Reeb J, Wirth T, Rost B. Variant effect predictions capture some aspects of deep mutational scanning experiments. *BMC Bioinformatics*. 2020;21:107.
- Salameh MA, Robinson JL, Navaneetham D, Sinha D, Madden BJ, Walsh PN, et al. The amyloid precursor protein/protease nexin 2 Kunitz inhibitor domain is a highly specific substrate of Mesotrypsin *. *J Biol Chem*. 2010;285:1939–49.
- Sananes A, Cohen I, Shahar A, Hockla A, Vita ED, Miller AK, et al. A potent, proteolysis-resistant inhibitor of kallikrein-related peptidase 6 (KLK6) for cancer therapy, developed by combinatorial engineering. *J Biol Chem*. 2018;293:12663–80.
- Sarfati H, Naftaly S, Papo N, Keasar C. Predicting mutant outcome by combining deep mutational scanning and machine learning. *Proteins Struct Funct Bioinform*. 2022;90:45–57.
- Scheidig AJ, Hynes TR, Pelletier LA, Wells JA, Kossiakoff AA. Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of alzheimer's amyloid β -protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities. *Protein Sci*. 1997;6:1806–24.
- Sharabi O, Shirian J, Grossman M, Lebendiker M, Sagi I, Shifman J. Affinity- and specificity-enhancing mutations are frequent in multispecific interactions between TIMP2 and MMPs. *PLoS One*. 2014;9:e93712.
- Shirian J, Sharabi O, Shifman JM. Cold spots in protein binding. *Trends Biochem Sci*. 2016;41:739–45.
- Slathia PS, Sagrika, Sharma E, Khan IA, Thakur RS, Sharma P. Recombinant production of therapeutic proteins. In: Singh DB, Tripathi T, editors. *Protein-based therapeutics*. Singapore: Springer Nature; 2023. p. 101–29. https://doi.org/10.1007/978-981-19-8249-1_4
- Starr TN, Thornton JW. Epistasis in protein evolution. *Protein Sci*. 2016;25:1204–18.
- Teymennet-Ramírez KV, Martínez-Morales F, Trejo-Hernández MR. Yeast surface display system: strategies for improvement and biotechnological applications. *Front Bioeng Biotechnol*. 2022;9:794742. <https://doi.org/10.3389/fbioe.2021.794742/full>
- Vieira LC, Handojo ML, Wilke CO. Medium-sized protein language models perform well at transfer learning on realistic datasets. *Sci Rep*. 2025;15:21400.
- Yang L, Zhang L, Wu Q, Boyd DD. Unbiased screening for transcriptional targets of ZKSCAN3 identifies integrin $\beta 4$ and vascular endothelial growth factor as downstream targets *. *J Biol Chem*. 2008;283:35295–304.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Meiri R, Reuveni O, Levi M, Radisky ES, Papo N, Orenstein Y. Engineering selective amyloid precursor protein inhibitors by machine learning and deep mutational scanning. *Protein Science*. 2026; 35(8):e70712. <https://doi.org/10.1002/pro.70712>