



Integrating protein language models and automatic biofoundry for enhanced protein evolution

Received: 14 September 2024

Accepted: 24 January 2025

Published online: 11 February 2025

Check for updates

Qiang Zhang , Wanyi Chen^{1,3,4,11}, Ming Qin^{1,5,11}, Yuhao Wang , Zhongji Pu⁷, Keyan Ding , Yuyue Liu⁴, Qunfeng Zhang^{1,3,4}, Dongfang Li⁴, Xinjia Li⁷, Yu Zhao , Jianhua Yao , Lei Huang^{1,3,4}, Jianping Wu , Lirong Yang , Huajun Chen , & Haoran Yu

Traditional protein engineering methods, such as directed evolution, while effective, are often slow and labor-intensive. Advances in machine learning and automated biofoundry present new opportunities for optimizing these processes. This study devises a protein language model-enabled automatic evolution platform, a closed-loop system for automated protein engineering within the Design-Build-Test-Learn cycle. The protein language model ESM-2 makes zero-shot prediction of 96 variants to initiate the cycle. The biofoundry constructs and evaluates these variants, and feeds the results back to a multi-layer perceptron to train a fitness predictor, which then makes prediction of second round of 96 variants with improved fitness. With the tRNA synthetase as a model enzyme, four-rounds of evolution carried out within 10 days lead to mutants with enzyme activity improved by up to 2.4-fold. Our system significantly enhances the speed and accuracy of protein evolution, driving faster advancements in protein engineering for industrial applications.

Proteins play a vital role in various fields, including medicine, chemical manufacturing, energy, agriculture, and consumer products. However, for industrial applications, proteins often require engineering to enhance properties such as stability, activity, selectivity, and binding affinity^{1,2}. Numerous strategies have been developed for protein engineering, with directed evolution being a well-established and powerful approach^{3–5}. Traditional directed evolution relies on iterative cycles of random mutagenesis and high-throughput screening to identify variants with desired traits. While effective, this process is time-consuming and labor-intensive⁶. Additionally, since directed

evolution typically introduces one mutation at a time, it can be trapped in local fitness optima, limiting further improvement^{7,8}. Recently, targeted mutagenesis guided by structure or sequence information has become a popular way to produce so-called small-but-smart libraries⁹. These libraries, which contain a higher ratio of beneficial mutations and fewer deleterious ones, improve the efficiency of directed evolution. For example, iterative saturation mutagenesis (ISM) and CASTing methods have been effectively used for enzyme evolution, requiring screening of only moderate-sized libraries including 5000 to 20,000 variants^{10,11}. However, the success of targeted mutagenesis is highly

¹Zhejiang University, Hangzhou, Zhejiang 310058, China. ²ZJU-UIUC Institute, International Campus, Zhejiang University, Haining, Zhejiang 314400, China.

³Institute of Bioengineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, Zhejiang 310058, China. ⁴ZJU-Hangzhou Global Scientific and Technological Innovation Centre, Hangzhou, Zhejiang 311200, China. ⁵School of Software Technology, Zhejiang University, Hangzhou 315103, China. ⁶Polytechnic Institute, Zhejiang University, Hangzhou 310015, China. ⁷Xianghu Laboratory, Hangzhou 311231, China. ⁸AI Lab, Tencent, Shenzhen, Guangdong 518000, China. ⁹Zhejiang Key Laboratory of Intelligent Manufacturing for Functional Chemicals, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 311215, China. ¹⁰College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310027, China. ¹¹These authors contributed equally: Qiang Zhang, Wanyi Chen, Ming Qin. e-mail: huajunsir@zju.edu.cn; yuhoruan@zju.edu.cn

dependent on the choice of target site, which requires a deep understanding of the protein's structure-function relationship¹². Moreover, software tools such as Rosetta and HotSpot Wizard have shown promise in enzyme redesign and functional enhancement, but de novo enzyme design remains in its infancy, and these methods are generally effective only for relatively simple reactions^{13–15}.

Machine learning (ML) has recently emerged as a promising tool for exploring protein fitness landscapes¹⁶. One specific application, ML-assisted directed evolution (MLDE), employs supervised ML models to predict the fitness of protein variants that carry multiple mutations^{17,18}. This technique facilitates larger jumps in protein sequence space, helping to bypass local optima that often occur in landscapes with strong epistasis. Bayesian optimization (BO), a method within active learning, is particularly well-suited for identifying protein variants with substantially improved fitness. Several studies have applied Gaussian process models with BO to optimize proteins. For instance, an iterative BO has demonstrated its effectiveness in enhancing thermostability of a P450 enzyme¹⁹. Similar methods have been applied to engineer activity a halogenase²⁰, enantioselectivity of a putative nitric oxide dioxygenase²¹, fluorescence wavelength of green fluorescent protein (GFP)²² and so forth. Despite these advances, the widespread application of ML in protein engineering remains challenging, in both acquiring and modeling protein function data. Experimentally collecting functional data is time-consuming and labor-intensive, particularly for enzymes with polyspecificity due to the substrate diversity and complexity¹⁸. Additionally, it was difficult to know how to efficiently sample and utilize informative protein mutants to train the ML models, especially for the protein without any prior information about the relationship between structure and function.

Protein language models (PLMs) were trained on extensive datasets of protein sequences spanning the evolutionary tree of life, and hence learned the fundamental principles of protein structure and function. PLMs have proven powerful in modeling functional proteins²³, predicting the direction of natural evolution²⁴ and designing novel proteins²⁵. The captured knowledge in PLMs could be applied to "zero-shot" optimization of specific proteins. Affinity maturation of antibodies has been guided by PLMs, through which screening 20 or fewer variants improved the binding affinities up to 160-fold²⁶. PLMs have also been used to assist optimization of an uracil-N-glycosylase variant activity that enables programmable T-to-G and T-to-C base editing²⁷. However, a major open question is whether general evolutionary information learned from sequence variation across past evolution is sufficient to enable efficient evolution of a specific protein under specific selection pressure.

Additionally, laboratory automation aided by biofoundries would be invaluable for generating the large volumes of data needed to develop ML models for protein engineering. A fully integrated biofoundry combines high-throughput core instruments, including liquid handlers, thermocyclers, fragment analyzers, and high-content screening systems, with peripheral devices such as plate sealers, shakers, and incubators. These components are seamlessly coordinated by robotic arms and scheduling software²⁸. One example is PlasmidMaker which was developed for automated high-throughput plasmid design and construction²⁹. Furthermore, the combination of biofoundries and ML models has led to the application of BioAutoma, an automated closed-loop system designed for engineering the lycopene production pathway³⁰. An automated high-throughput genome editing platform was recently devised through which thousands of samples were automatically edited within a week³¹. However, protein engineering has yet to fully leverage recent developments in biofoundries. A so-called self-driving autonomous machine for protein landscape exploration (SAMPLE) platform was developed for fully autonomous protein engineering³². However, the platform only assembled pre-synthesized DNA fragments from

different homologues for exploring a small protein landscape containing 1352 protein sequences. A Bayesian optimization-guided evolutionary algorithm along with robotic experiments were also developed for protein engineering, and a four-site combinatorial library was explored by sampling 384 mutants each round for 4 rounds, 1536 mutants in total³³. However, the BO algorithm could not guide the mutant residue selection, and the generality of the algorithm remains unknown.

In this study, we propose a protein engineering strategy that integrates the predictive power of PLMs with the operational efficiencies of an automated biofoundry. Within the Design-Build-Test-Learn cycle, the Learning and Design phases utilize insights from PLMs to elucidate protein sequence-fitness relationships and to sample novel mutants, while the Build and Test phases are efficiently conducted using automated biofoundry. Specifically, when applying PLMs for protein variants design, two modules were developed to predict high-fitness mutants for proteins without knowing the mutation sites and the proteins with mutation sites known, respectively. Our robotic systems are adept at building protein variants and collecting protein variants functional data continuously, ensuring high reproducibility with comprehensive metadata tracking and real-time data sharing. By combining the advanced predictive capabilities of PLMs with the high-throughput functionalities of robotic systems, this method is designed to surpass traditional constraints and expedite the discovery and enhancement of proteins crucial for industrial applications. We used the *Methanocaldococcus jannaschii* p-cyanophenylalanine tRNA synthetase (pCNF-RS) as a model enzyme for validating the process. In each round, 96 variants were designed by PLMs or a supervised ML model, and then constructed and tested by an automated biofoundry. Four rounds were performed within half a month with the enzyme activity progressively improving, reaching its peak in the fourth round. The PLMeAE system exhibited superior performance compared to random selection and the traditional directed evolution strategy, and it has the potential to accelerate the engineering of other proteins.

Results

An overview for protein language model-enabled automatic evolution (PLMeAE)

Here, we devised a protein language model-enabled automatic evolution (PLMeAE) platform (Fig. 1), a closed-loop system for automated protein engineering within the Design-Build-Test-Learn (DBTL) cycle. The platform employs PLMs to facilitate the learning and design phases, while the build and test phases are executed by a biofoundry. The process begins with the creation of a variant library, informed by a zero-shot learning approach enabled by PLMs in the Design phase. In specific, two zero-shot tasks were solved by the PLMs, depending on the availability of mutation target sites. Firstly, with no prior information about the target protein, PLMs were used to predict high-fitness single mutants in a zero-shot setting. Secondly, when the mutation sites were already identified based on previous experiments or through physical modelling techniques such as docking, molecular dynamics simulations, the PLMs were used to predict zero-shot high-fitness multi-mutant variants at the given target sites. Following this, the proposed library was synthesized, expressed, and tested by automated facilities of biofoundry in the Build and Test steps. After collecting the experimental data, in the phase of Learn, the PLMs encode the protein sequences and a supervised machine learning model was trained to correlate these variants with their fitness levels. Subsequently, optimization algorithms were applied to explore the variant landscape, facilitating rational design and identifying promising variants for subsequent testing rounds. This iterative process, akin to an active learning strategy, continues until optimal variants are developed.

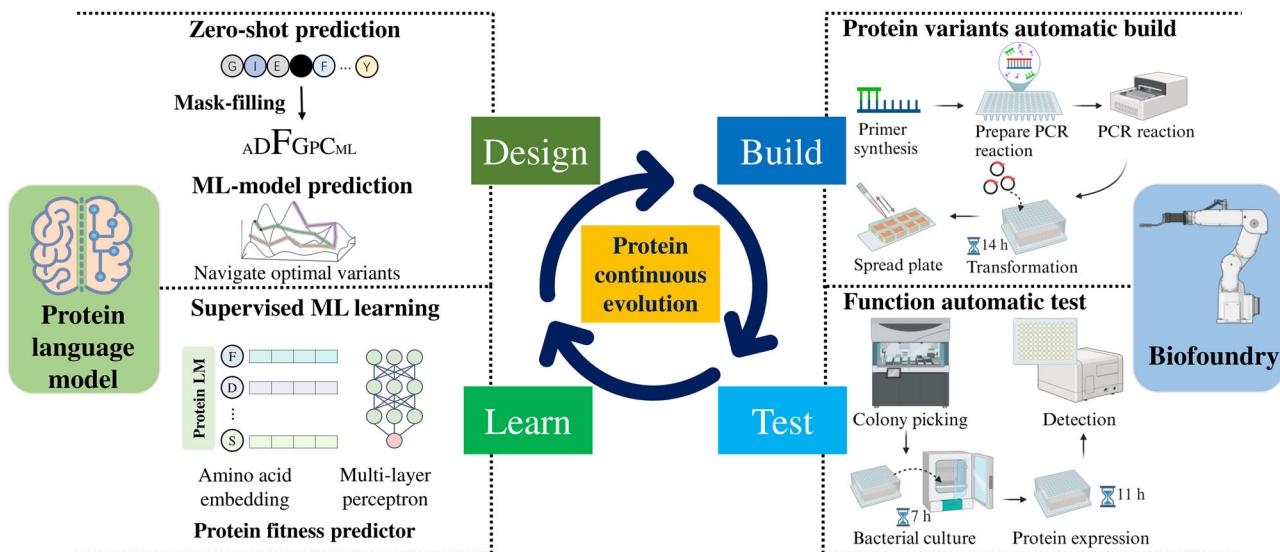


Fig. 1 | Overview of protein language model-enabled automatic protein evolution. In the Desing-Build-Test-Learn loop of protein engineering, PLMs are applied to facilitate the learning and design phases, while the build and test phases

are executed by a biofoundry. Created in BioRender. Yu, H. (2024) <https://Biorender.com/f92a776>.

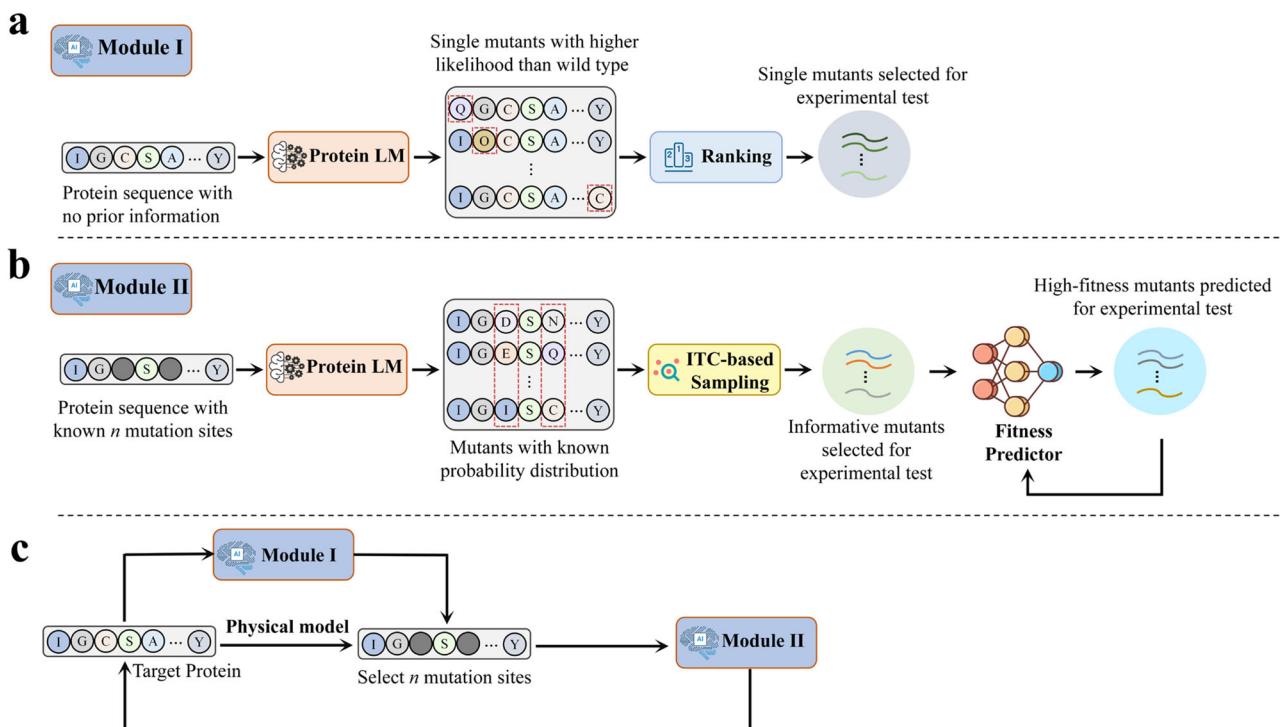


Fig. 2 | Protein language model used for protein automatic evolution. **a** Module I for engineering proteins without identified mutation sites. **b** Module II for engineering proteins with previously identified mutation sites. **c** Module I and Module II

used in combination or independently. Created in BioRender. Yu, H. (2025) <https://Biorender.com/g25x718>.

Protein language model used for protein variant design

In this study, we developed two modules based on PLMs to predict high-fitness mutants for two respective zero-shot tasks. The Module I is used for proteins without previously identified mutation sites (Fig. 2a). In this module, the PLM predicts single mutants with a high likelihood of improved fitness, using this likelihood as a proxy for fitness levels. These high-likelihood mutants are then employed to identify critical mutation sites. On the other hand, the Module II targets proteins with known mutation sites, and the PLM is used to sample informative

mutants for experimental characterization (Fig. 2b). Additionally, PLM is used to encode protein sequence for training a fitness predictor. Module I and Module II could be used in combination or independently (Fig. 2c).

PLMeAE Module I: Engineering proteins without previously identified mutation sites

In the Module I of the PLMeAE system, we focus on the engineering of proteins that lack predefined mutation sites, utilizing a systematic

approach to uncover and exploit novel sites for enhanced protein function (Fig. 2a). This module leverages the PLM to identify potential mutation sites in a zero-shot prediction setting, where no prior mutation data is available. The process begins with the wild-type sequence of the target protein. Each amino acid in the sequence is individually masked and analyzed by the PLM to predict the impact of potential mutations at that site. The model assesses all possible single-residue substitutions at each masked site, calculating the likelihood of each variant exceeding the fitness of the wild-type protein. Variants that demonstrate a high likelihood of improved function are then ranked based on their predicted fitness gains. The most promising candidates—typically the top 96 as determined by their likelihood—are selected for experimental characterization. Then, the automatic biofoundry will synthesize and test each variant to validate the model's predictions and to measure the actual fitness improvement over the wild type (Fig. 2a). The improved single variants identified through this process can be further selected as targets for additional fitness enhancements using Module II.

PLMeAE Module II: Engineering proteins with previously identified mutation sites

The Module II of the PLMeAE system targets proteins with identified mutation sites (Fig. 2b). The initial round evolution of Module II involves selection of informative variants for annotation, which subsequently serve as the dataset for constructing a supervised machine learning model. To achieve this, we adopted an advanced sampling approach that integrates a protein language model (PLM) with a novel metric derived from Information Transport Complexity (ITC). Leveraging the PLM, we can compute the probability distribution of each amino acid at specified masked positions within a protein sequence. To ensure the selected variants are sufficiently informative, mutants exhibiting both high probabilities and significant diversity, as evaluated by the ITC score, were chosen for inclusion in the subset. For example, suppose there are four types of amino acids sampled at each masked position, there are a total of 4845 ($C(20,4)$) subsets (Fig. 3a). Based on the probability distribution of amino acids calculated by PLMs and the similarity among amino acids calculated by PLM amino acid embeddings, our ITC-based method would identify the one subset with highest probability and the largest diversity among amino acids inside the subset (Fig. 3a).

The selected informative samples undergo fitness annotation via an automatic biofoundry, which subsequently feeds into a fitness predictor (FP) comprising a PLM and a multi-layer perceptron (MLP). To prevent overfitting, especially given the limited scale of annotated data, we maintain fixed parameters for the PLM while optimizing the MLP parameters. This optimization focuses on minimizing the mean absolute error in fitness predictions, thereby enhancing the model's accuracy and reliability in predicting functional outcomes. The fitness prediction model is then used to predict the fitness values of all the protein variants and the top ones are sent for biofoundry for build and test. The labeled data of the protein variants are then used to update the fitness predictor, and multiple such rounds are carried out until the satisfactory protein variants are obtained (Fig. 3b).

To validate the efficacy of Module II, we conducted an *in silico* experiment using the GB1 dataset, which includes four known mutation sites in B1 domain of protein G (GB1) and experimentally determined fitness values for nearly all 20^4 mutants³⁴. The experiment involved four rounds of sampling, each selecting 96 mutants. We utilized the experimentally measured fitness values as a proxy for the biofoundry annotation process. ESM are state-of-the-art protein language models that have been used for protein design, structure prediction and antibody or enzyme engineering^{26,27,35,36}. Several ESM models are available, each trained on different protein sequence dataset and with varying number of parameters. To understand which ESM model is more accurate for zero-shot prediction of

mutation, we explored the correlation between evolutionary score calculated by these ESM models and experimentally characterized fitness value in GB1 dataset. We used Spearman's correlation (ρ) to quantify the rank correlation. It was found that ESM2_t33_650-M_UR50D (ESM-2) performed best among all the models tested, with ρ values of 0.415, 0.331, 0.173 for single, double and triple mutants, respectively (Fig. 3c).

We then tested if ESM-2 was useful to identify potential positions for mutation when no information about mutation targets was available. We used two datasets, the SUMO-conjugating enzyme UBC9 dataset³⁷ and the Ubiquitin dataset³⁸, which contain activity data of almost all possible single variants of the two proteins. ESM-2 was used to predict top 96 single variants based on the two protein sequences, while another 96 variants were also randomly selected for comparison (Supplementary Data 1). As for UBC9 variants, the maximum fitness data obtained by EMS-2 prediction was 2.35, much higher than 1.72 achieved by random selection. Moreover, the mean fitness value of all variants predicted by ESM-2 was 0.53, higher than 0.43 of those randomly selected (Supplementary Fig. 1). Similarly, for ubiquitin, the mean fitness data of all variants predicted by ESM-2 was 0.75, significantly higher than the 0.44 of those randomly selected, although the random selection achieved a maximum fitness value similar to that of ESM-2 (Supplementary Fig. 1). We also found that high-fitness variants predicted by ESM-2 provided hot spots for further engineering. For example, E18C and D32A of ubiquitin predicted by ESM-2 ranked 44th and 87th in the dataset, respectively. If these two amino acids were selected for further engineering, D32K and E18M could be obtained, which ranked 13th and 19th in the dataset, respectively, indicating the potential of EMS-2 in identifying positions essential for protein engineering (Supplementary Data 1).

In the initial round of Module II, with the application of ITC-based sampling approach, four amino acids were selected at each mutation set, 256 variants in total, and 96 variants of them were selected according to their likelihoods given by the ESM-2. The mutants achieved a mean fitness of 0.744 and a maximum fitness of 5.45, which ranked 54th in the entire GB1 dataset, suggesting that the sampling approach was useful in identifying high-fitness variant (Supplementary Data 2). The second round achieved a mean fitness of 1.97, significantly higher than the first round, suggesting the great effect of supervised MLP model. Subsequent rounds also showed incremental improvements: the maximum fitness reached 5.50 in the second round, 5.73 in the third, and peaked at 6.20 in the fourth round, thereby enhancing the original wild-type protein's fitness by 520% (Fig. 3d and Supplementary Data 2). Despite this, the improvement over the last three rounds was not significant, possibly because the initial supervised ML model predictions had already achieved substantial gains, leaving fewer beneficial mutations available in the remaining sequence space. There are 149 361 data in the GB1 dataset, and the fitness value of 6.20 obtained in fourth round ranked 21st in the dataset. Although the largest fitness value of 8.76 in the dataset was not obtained yet, this method showed the potential to be used in protein engineering task. A comprehensive evaluation of both Module I and Module II is further conducted through wet-lab experiments, with detailed outcomes presented in the following sections.

Automation for protein variants build and test

For this work, we focused on engineering *Methanocaldococcus janaschii* p-cyanophenylalanine tRNA synthetase (*MjTyrRS*) for improved incorporation of non-canonical amino acids (ncAAs)³⁹. The enzyme activity was measured by detecting suppression of the sfGFP-UAG2 gene by ncAA, which was investigated by measuring the fluorescence intensities of the cells co-expressing sfGFP2TAG and pCNF-RS, a variant of *MjTyrRS*. The ML model designs proteins and sends them to the biofoundry for mutants build and test. Since site-directed

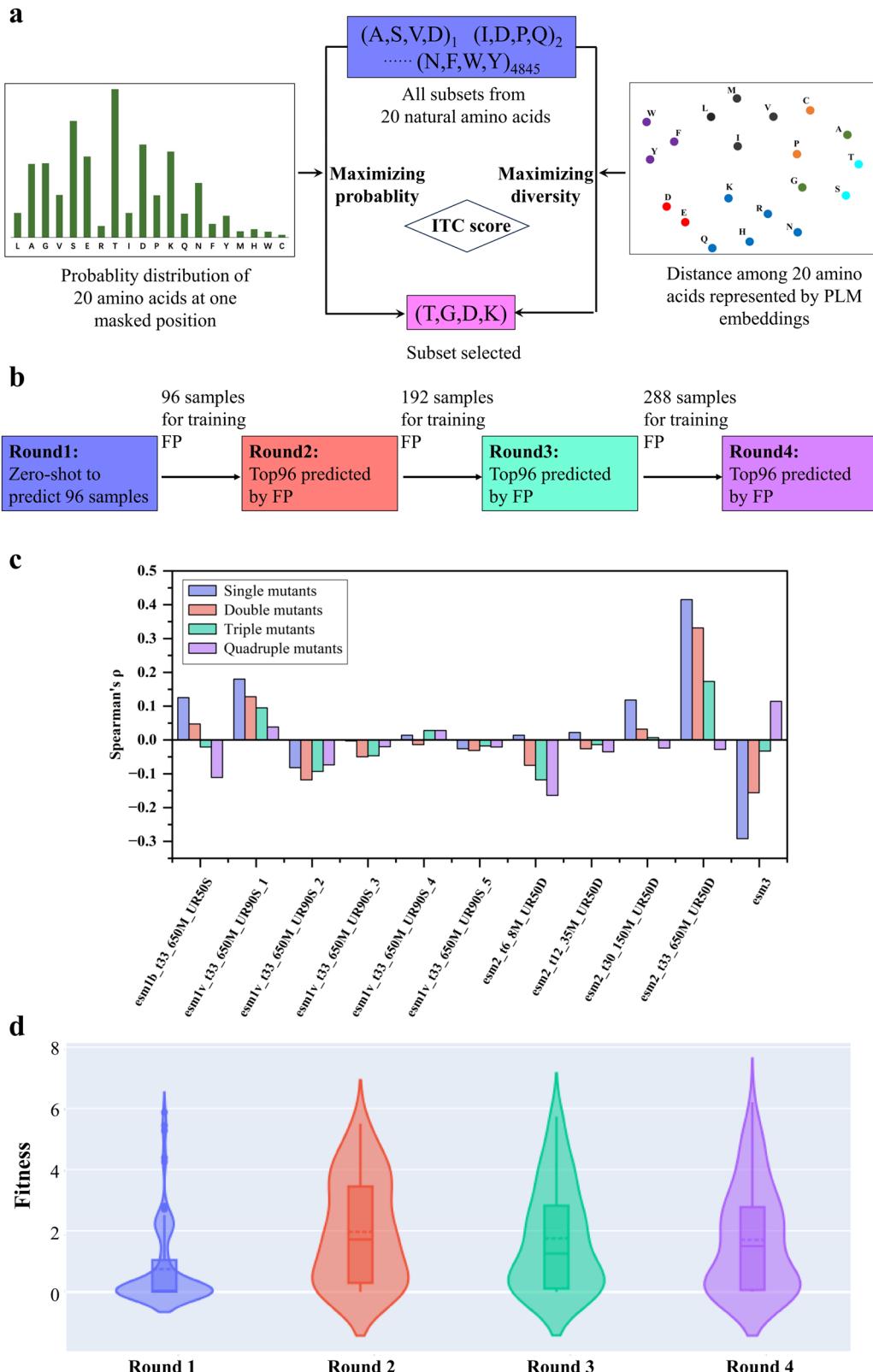


Fig. 3 | Protein language model used for engineering proteins with identified mutations. **a** A scheme illustrating application of PLM for sampling informative mutants at one mutation site, assuming that four amino acids are selected. **b** A flow chart illustrating the process of PLMeAE Module II. FP, fitness predictor.

c Evaluation of various ESM models in GB1 dataset. **d** Performance of PLMeAE Module II tested in the GB 1 dataset. The violin plots show the distribution of

protein variants fitness values. Inside each plot, a box-and-whisker diagram is included, where the whiskers represent the minimum and maximum values, the box spans from the first to the third quartile, the central solid line marks the median value and the dash line represents the mean value. Source data are provided as a Source Data file.

mutagenesis is a commonly used approach in protein engineering⁴⁰, we developed a highly streamlined, robust and general pipeline for automated making site-directed mutagenesis using QuikChange method, along with gene transformation, protein expression and enzyme biochemical characterization.

Protein variants build

The QuikChange method uses 30–35 base megaprimer with the intended mutation-bearing bases in the middle to amplify the entire recombinant plasmid. The methylated parent plasmids are removed using *DpnI* treatment before transformation. In the initial step, a Python script designed all primers after receiving 96 single mutant sequences from the ML model, which were then forwarded to provider for synthesis. The PCR preparation was carried out using the automation workstation (Evo), and the PCR plates were subsequently sealed by a plate sealer (ALPS) for the PCR reaction carried out by an automatic thermocycler (Fig. 4a). Following this, the plates were unsealed by automated plate seal removal (Xpeel) and the acoustic liquid handler (Echo) was used for *DpnI* digestion. BL21(DE3) competent cells containing sfGFP2TAG plasmid used for enzyme activity measurement were mixed with the PCR products by automation workstation (Evo) and cultured for 1 h. Transformant plating was conducted by spraying the cells onto eight-well agar plates with 8 individual pipetting channels in the automatic workstation (Fluent). The plates were tagged by a microplate labeler (Agilent Labeler), then incubated in the automated incubator (Cytomat 10 C) for 14 h (Fig. 4a, b). We optimized the high-throughput transformation condition and ensured the cell growth after transformed by digested PCR products (Supplementary Fig. 2).

Enzyme activity test

On the second day, the clones were picked into 800 µL of LB culture medium using automation workstation colony-picker (Fluent), and cultured for 7 h in the automated incubator with orbital shaking (Cytomat 2 C Tos). Then, a robotic centrifuge (Rotanta) was used to collect 200 µL of bacterial culture, which was then resuspended in GMML medium using a reagent dispenser (Multidrop Combi) and cultured for another 11 h in the automated incubator with orbital shaking (Cytomat 2 C Tos). The remaining bacterial culture was used for Sanger sequencing to verify the correct mutants and for glycerol stock preservation (Fig. 4a, b). On the third day, the activities of these mutants were measured by an automatic microplate reader (CLARIOstar), and the data were analyzed by Momentum DataMiner. With this process, we tested the reproducibility of the platform by detecting the incorporation of four different ncAAs into sfGFP (Supplementary Fig. 3), and found that the system reliably measured activity of pCNF-RS against four ncAAs with a standard error less than 5%.

We added several layers of exception handling and data quality control to further increase the reliability of the automatic platform (Fig. 4c). The system checks whether (1) the PCR has worked by assaying double-stranded DNA with SybrGreen (Supplementary Fig. 4), (2) colony grow on agar plate after transformation, and (3) the cells grow well when cultivated in presence of ncAA. Once the failure in the PCR was detected, the system will automatically repeat the PCR reaction, while when no colony was detected on the agar plate, the system will automatically repeat the process of cell transformation. To ensure the accuracy of pCNF-RS activity measurement, the OD₆₀₀ of cells expressing pCNF-RS and sfGFP was detected. Once the OD₆₀₀ was lower than 0.4 after 11-h cultivation, the system will automatically repick colonies for cell growth. The procedure for building and testing 96 single variants of pCNF-RS takes around 1 d for primers synthesis, 1 h for PCR, 0.5 h for *DpnI* digestion, 15 h for transformation and colonies picking, 7 h for protein expression, 11.5 h for measuring enzyme activity, and around 59 h overall to go from a requested

protein single mutant to a physical enzyme sample to a corresponding enzyme activity data (Fig. 4a, b).

PLMeAE validation: automated directed evolution of pCNF-RS

Methanocaldococcus jannaschii tyrosyl-tRNA synthetase (*MjTyrRS*), one of most widely used tRNA synthetases for incorporating ncAAs into proteins based on gene codon expansion technology, has been widely engineered for accepting a variety of ncAAs⁴¹. pCNF-RS, a variant of *MjTyrRS*, was engineered to accept *p*-cyano-L-phenylalanine (*p*CNF), but was also found to recognize at least 18 tyrosine derivatives, including *p*-acetylphenylalanine (*pAcF*), *p*-azido-L-phenylalanine (*pAzF*), and others⁴². This polyspecificity has enabled the rapid and broad application of the genetic ncAA incorporation technique without the need to evolve a new aminoacyl-tRNA synthetase (aaRS) variant⁴³. However, pCNF-RS exhibited a low incorporation efficiency towards these ncAAs, for example, only 60% expression level for proteins with *p*-acetylphenylalanine (*pAcF*) incorporated relative to wild-type GFP⁴². In order to engineer pCNF-RS for improved incorporation efficiency of ncAAs, we attempted to apply PLMeAE Module II firstly, followed by PLMeAE Module I, as a vast of previous studies regarding engineering of *MjTyrRS* could provide the insights into the selection of the mutation target sites. A previous study has modified four positions H283, P284, M285, and D286 of the *MjTyrRS* mutant pAzFRS, aiming to improve the efficiency of introducing *p*-azidophenylalanine (*pAzF*). A positive-negative selection procedure was carried out to obtain improved variants in the mutation library containing a theory of 160 000 mutations. In the positive selection, aaRS variants are tested in vivo for their ability to suppress in-frame amber codons in a chloramphenicol acetyl transferase (*cat*) reporter gene, which confers resistance to chloramphenicol. By contrast, negative selection involves abandoning aaRS variants that incorporate natural amino acids using a toxin protein barnase. The plasmid expressing aaRS mutation library was first transformed into *E. coli* cells containing plasmid expressing *cat* gene for positive screening. After that, the plasmid expressing aaRS mutants was isolated from surviving colonies and separated from positive screening plasmid, and then transformed into *E. coli* cells containing plasmid expressing barnase for negative screening. This process is complex, time-consuming and laborious (Supplementary Fig. 5). Through two rounds of positive-negative screening, a variant H283T/P284S/M285D/D286V was obtained, which showed an 8-fold increase in *pAzF* incorporation efficiency⁴⁴. These four residues H283, P284, M285, and D286 are in the C-terminal tRNA binding domain of *MjTyrRS* and play crucial roles in enhancing *pAzF* incorporation efficiency, yet their roles in pCNF-RS remain unclear (Fig. 5a). We hence selected these four target residues for combinatorial mutagenesis for improving the incorporation efficiency of the *pAcF*.

Following the PLMeAE Module II, ESM-2 model firstly carried out first-round zero-shot prediction of 96 informative variants with four amino acids substitutions at each site of four mutation target (Fig. 5b). Biofoundry automatically constructed and measured the activities of the 96 mutants. Suppression of the sfGFP-UAG2 gene by *pAcF* was investigated by measuring their fluorescence intensities, and the fitness was presented by fluorescence intensity of the variants relative to the wild type (Supplementary Fig. 6). In the first round, 91 variants were obtained with the best mutant H283S/P284A/M285A/D286Q (M-R1) showing a 1.3-fold improvement in fitness, while 84 variants showed lower activity compared to the wild type (Fig. 5c and Supplementary Data 3). The other six mutants include H283T/P284A/M285Q/D286E, which shows a 1.2-fold improvement, and H283K/P284S/M285L/D286A, which shows a 1.1-fold improvement in fitness, along with four variants exhibiting activity similar to that of the wild type (Supplementary Data 3). The first-round mutants and their enzyme activities data were then applied to train a two-layer MLP fitness predictor with the protein embedding by ESM-2 model. The fitness

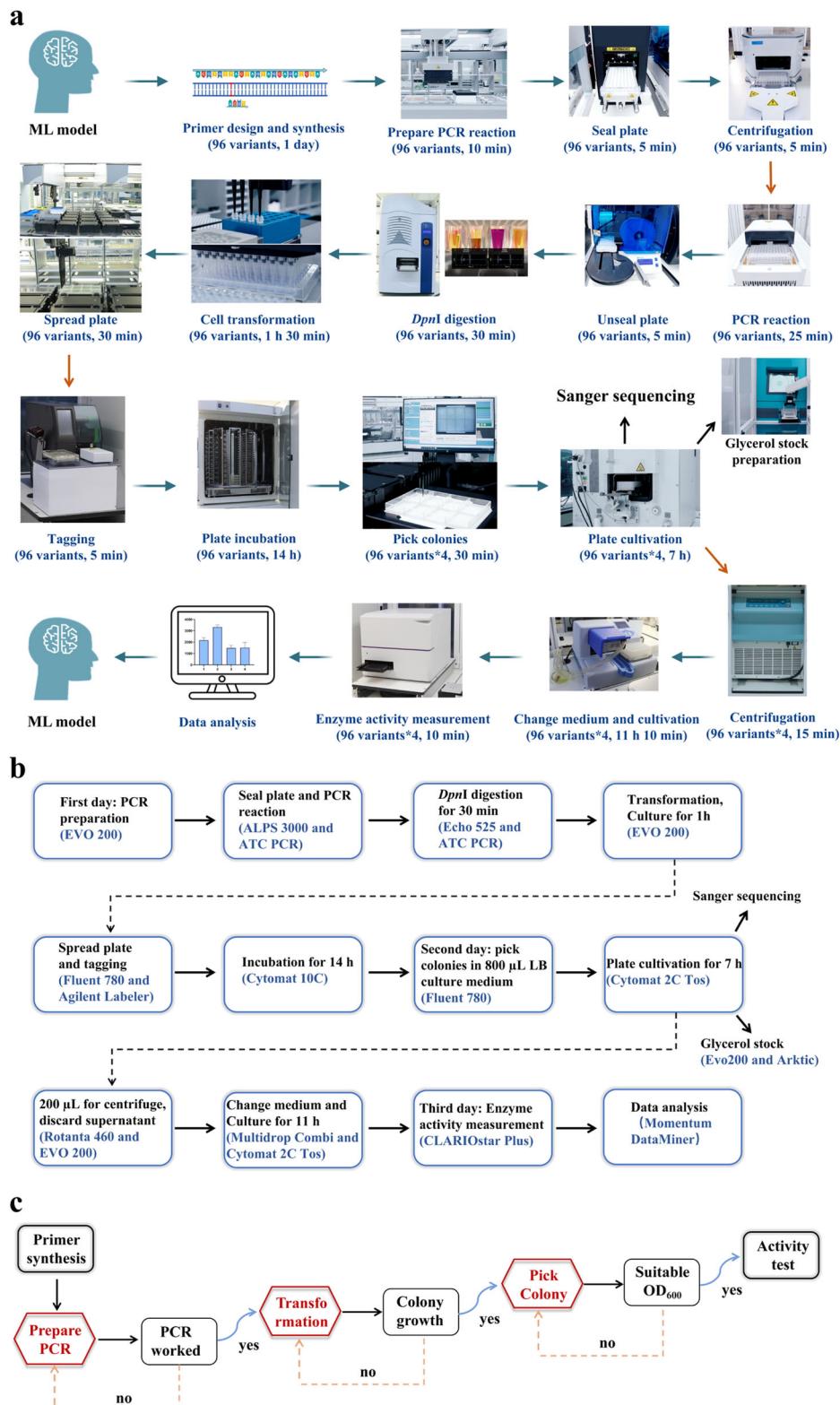


Fig. 4 | Overview of automatic protein variants build and test. a Workflow for protein variants build and test using biofoundry. **b** Flowchart showing different instruments and time used in the whole process which starts from preparing PCR

reaction, and ends at data analysis. **c** Multiple layers of exception handling and data quality control for failed experimental steps.

predictor was then applied to predict the fitness of all 160 000 variants and the top 96 were sent to biofoundry for build and test. Sequences of second-round variants varied a lot compared to the first-round variants indicating that the informative sampling strategy was effective to guide ML model to explore the sequence space (Fig. 5b). Forty-eight

second-round variants showed higher fitness than the wild type, and the best variant H283L/P284T/D286E (M-R2) improved the fitness of pCNF-RS by 2.0-fold (Fig. 5c). The number of variants with higher fitness than wild type was significantly higher for the second-round variants than the first-round variants, indicating a positive effect of the

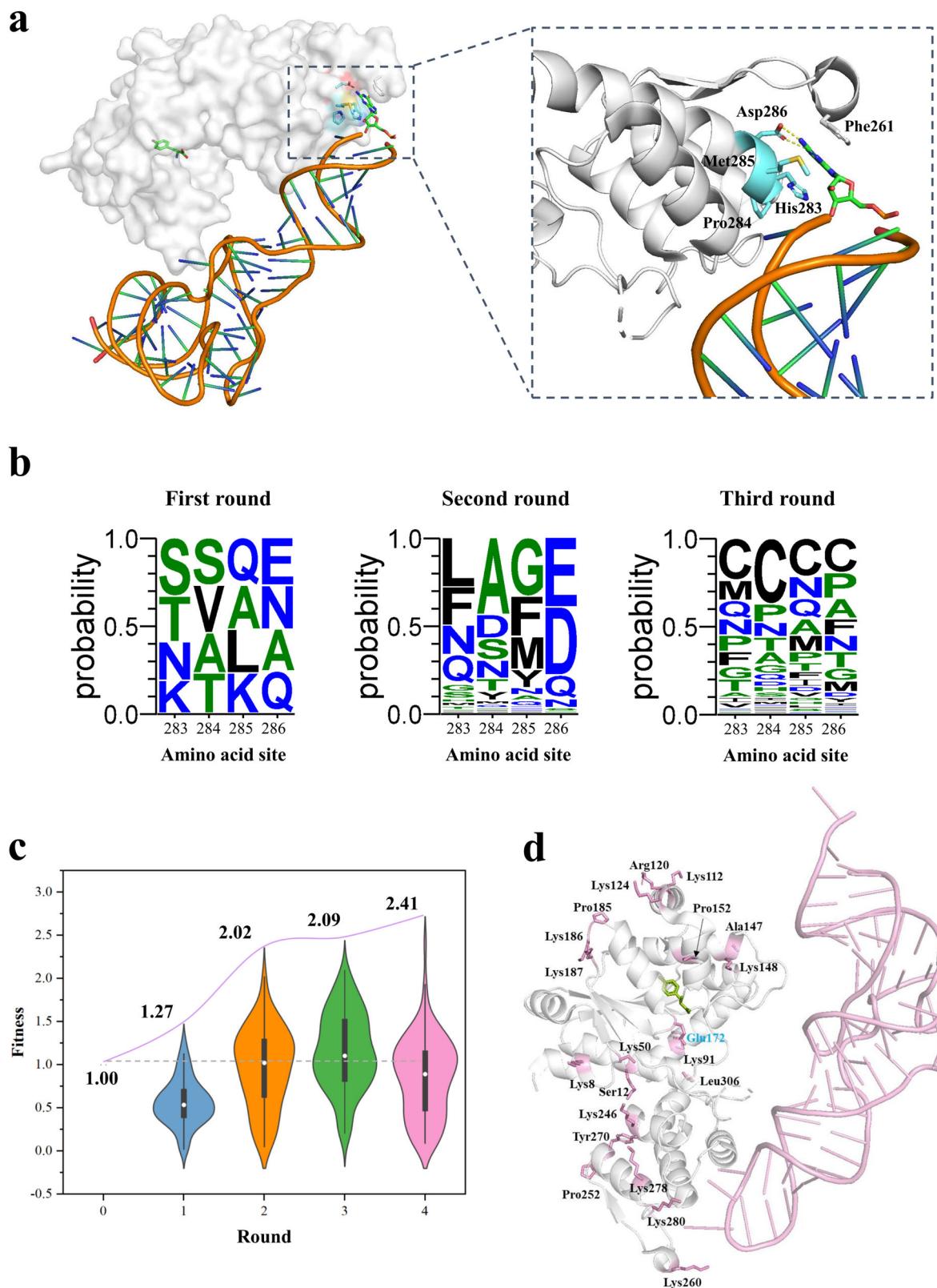


Fig. 5 | Automated evolution of pCNF-RS. **a** Mutation sites selected for constructing combinatorial mutagenesis with PLMeAE Module II, shown in the crystal structure (PDB:1J1U) of *Mj*TyrRS. **b** Sequence alignment of the variants obtained in four rounds of directed evolution. **c** Fitness values of variants obtained in four rounds of directed evolution. The activities of variants were tested in three independent repetitions. 91, 90, 94 and 96 variants are tested in first, second, third and

fourth round, respectively. The violin plots show the distribution of fitness values. A mini box-and-whisker diagram is included, where the whiskers represent the minimum and maximum values, the box spans from the first to the third quartile, and the central spot marks the median value. Source data are provided as a Source Data file. **d** Location of variants tested in fourth round evolution using PLMeAE Module I.

fitness predictor. The enzyme activity data of first two-round variants were then used to train the fitness predictor again to predict third-round 96 variants. These variants showed a higher sequence diversity compared to the first two-round variants and 60 of them showed higher fitness than wild type, reflecting a higher positive rate compared to first two rounds (Fig. 5b, c and Supplementary Table 1). However, the maximum fitness value was only 2.1-fold higher than the wild type, not significantly different from the maximum value of the second round.

Considering there might be no further improved variants obtainable using the PLMeAE Module II, we turned to PLMeAE Module I for exploring new mutation sites. With the best variant in the third round, M-R3 (H283G/P284C/D286K and a nonprogrammed mutant at lac operator of the plasmid) as the input (Supplementary Table 2), the ESM-2 was applied to predict 96 high-fitness single variants. Interestingly, the variants were located across the structure of the *MjTyrRS*, but none were found at the mutation sites selected in first three rounds (Fig. 5d). We also found that several substitutions could be suggested at one mutation site, such as K280G, K280D, K280C, K280S, K280A at position of Lys280, and hence a total of 32 mutation sites were tested (Supplementary Data 3 and Fig. 5d). Among the 96 variants characterized in round 4, 41 of them showed increased fitness compared to the wild type, and the best variant M-R3/E172D (M-R4) showed a 2.4-fold improved activity compared to the wild type, also 1.2-fold higher than the best variant obtained in the first-three rounds evolution (Supplementary Data 3). The mutation sites including Glu172, Lys186, Lys280 and Ser12 exhibited variants with fitness values improved by more than 1.5-fold compared to the wild type, which could be hence selected for constructing combinatorial mutants using PLMeAE Module II (Supplementary Fig. 7).

To further illustrate the effect of PLMeAE in engineering pCNF-RS, we constructed a random mutation library targeting residues H283, P284, M285, and D286 using primers containing four consecutive NNK codons. We then randomly picked 192 colonies into two 96-well plates for sequencing and testing for enzyme activity. Ninety variants with different sequences were finally obtained, among which 82 showed lower activity, 6 showed similar activity and 2 showed higher activity compared to the wild type (Supplementary Fig. 8). The two improved variants H283P/P284A/M285C/D286L and H283A/M285L/D286G improved the enzyme activity by 1.26-fold and 1.20-fold, respectively, which was comparable with M-R1, while significantly lower than that of M-R2, M-R3 and M-R4. The positive rate of random selection was around 2.2%, similar with that that of zero-shot predictions using PLM, but was significantly lower than 50% of second-round and 62.5% of third-round PLMeAE. This demonstrates the superior effectiveness of PLMeAE over random selection. Although the positive rate of random selection is similar to zero-shot predictions of PLM, it has been shown that the training data based on random sampling does not perform well in building supervised ML models^{45,46}.

Further characterization of the variants obtained through PLMeAE

We further characterized the best variants M-R3 and M-R4 obtained in the third and fourth-round evolution, respectively. The incorporation efficiency of the *pAcF* was first measured by detection of sfGFP expression whose second position was substituted by UAG. Wild-type pCNF-RS and two variants showed higher amount of sfGFP expression in presence of *pAcF* compared to in absence of *pAcF*, indicating the successful UAG translation. The two variants exhibited higher activity than wild type, as indicated by both fluorescence intensity and yield of purified sfGFP incorporated with *pAcF* (Fig. 6a). The best variant obtained in the fourth-round evolution M-R4 produced around 1.7 mg/mL purified sfGFP, 2.8-fold higher than the M-R3 obtained in the third-round evolution, and 12.1-fold higher than the wild type (Fig. 6a). Since the pCNF-RS exhibited the misincorporation of canonical amino acids,

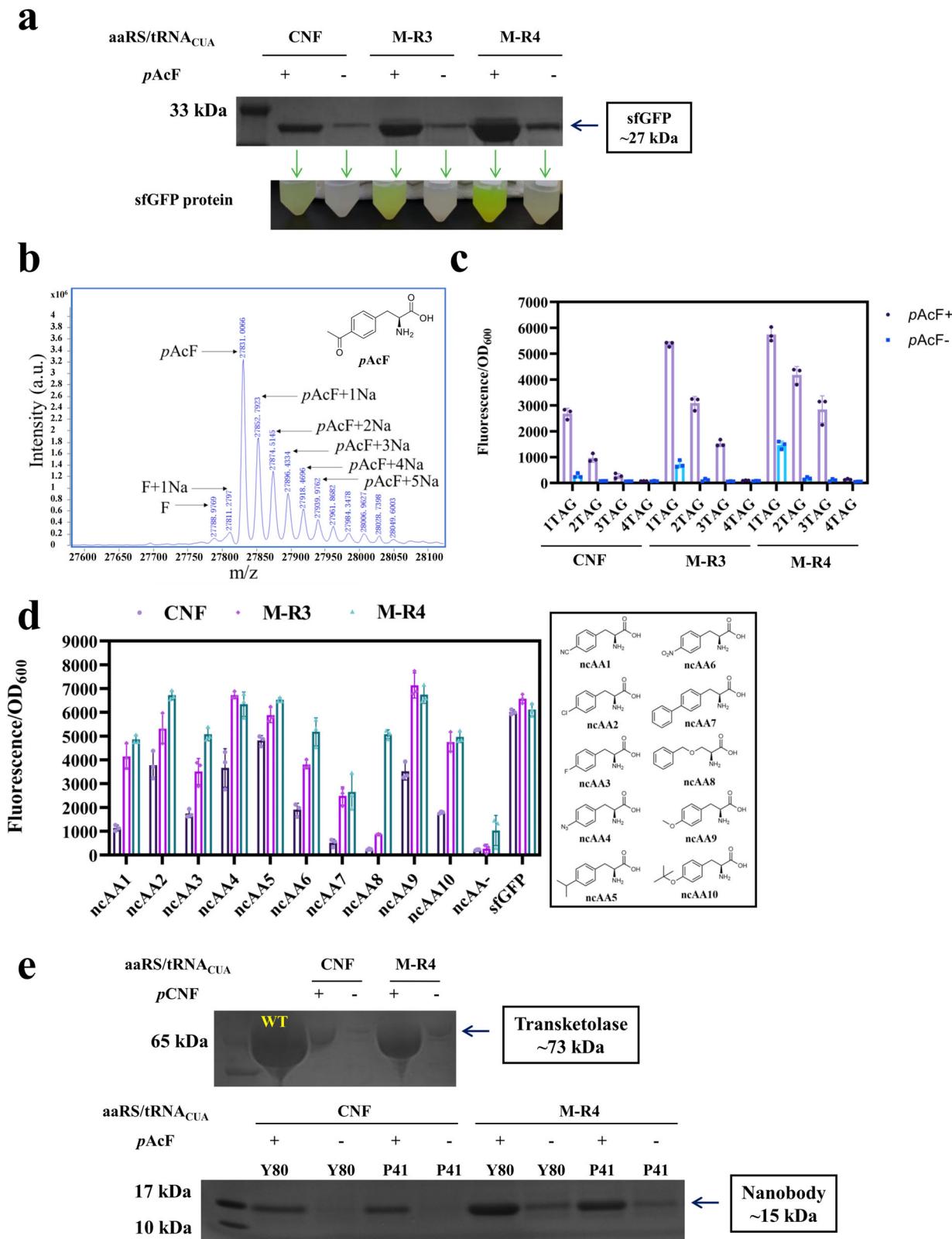
M-R4 also increased expression level of sfGFP with natural amino acids incorporated, which was confirmed by the mass-spectra analysis (Fig. 6b). We also tested the ability of *MjTyrRS* variants to suppress multiple amber codons. When suppressing two and three amber codons in presence of *pAcF*, M-R3 and M-R4 showed significantly higher efficiency over the wild type pCNF-RS, and M-R4 exhibited better performance than the M-R3. Specifically, we observed a 1.6-, 4.4-, and 11.6-fold increase in incorporation of *pAcF* into sfGFP(1TAG), sfGFP(2TAG) and sfGFP(3TAG), respectively, for the M-R4 compared to the wild type (Fig. 6c).

As the pCNF-RS showed polyspecificity toward various substrates, we tested if the variants M-R3 and M-R4 also improved the incorporation efficiency of other ncAAs. The variants did show improved activities against 10 ncAAs compared to the wild type, while the two variants showed different improvement effect (Fig. 6d). SDS-PAGE revealed that sfGFP with these 10 ncAAs incorporated could be purified in high yield (Supplementary Fig. 9) and the amount of protein purified was higher for M-R4 than pCNF-RS. Mass-spectra analysis confirmed the correct incorporation of these ncAAs, although sodium adducts were observed in the mass spectrum (Supplementary Table 3 and Supplementary Fig. 10–19), which was commonly in electrospray ionization ESI-MS measurements⁴⁷. Interestingly, although M-R4 showed higher incorporation efficiency toward *pAcF* compared to M-R3, M-R3 displayed similar activity toward ncAA4, ncAA5, ncAA7, ncAA9 and ncAA10 relative with M-R4, indicating that both mutations might have influenced the enzyme substrate specificity through modifying the interactions with tRNA.

We also tested the effect of M-R4 variant on improving expression of other proteins with ncAAs incorporated. A previous study showed that incorporation of *pCNF* at position 385 of *E. coli* transketolase 3M variant (TK, S385Y/D469T/R520Q) increased k_{cat} of the enzyme by 100%⁴⁸. We hence compared the amount of TK Y385*pCNF* expressed in presence of pCNF-RS or M-R4. The same weight of wet cells was used to purify the target protein, and as expect, M-R4 generated more Y385*pCNF* than the pCNF-RS, with the concentration of purified protein improved by 6.7-fold, indicating that M-R4 significantly improved the incorporation efficiency of *pCNF* at TK (Fig. 6e). One of main applications of incorporating *pAcF* in proteins is to couple with other molecules for preparing antibody-drug conjugates (ADCs) or bispecific antibody. We hence tested if M-R4 could improve the expression of *pAcF* containing nanobody. A VHH nanobody targeting chemokine (C-C motif) ligand 5 (CCL5) was incorporated with *pAcF* at positions Y80 and P41. Since VHH contains disulfide bonds and is challenging to be expressed in high yield in *E. coli*, the amount of nanobody obtained was significantly lower than TK. However, it clearly showed a higher amount of Y80*pAcF* and P41*pAcF* were obtained enabled by M-R4 compared to pCNF-RS (Fig. 6e). This suggested that the variant M-R4 is promising to be used to construct ADCs or bispecific antibody for drug development.

Discussion

In this study, we developed an automated protein evolution platform, PLMeAE. PLMeAE tightly integrates protein language models for mutant design, and biofoundry facilities for automatic protein variants build and test, which effectively explores protein fitness landscapes and discovers optimized proteins. We successfully applied PLMeAE to engineer pCNF-RS for increasing enzyme activity for enhanced incorporation efficiency of ncAAs into proteins. After four rounds of evolution, the activity of the enzyme was increased by 2.4-fold and the yield of protein with *pAcF* incorporated was increased by 12.2-fold enabled by the best pCNF-RS variant obtained. During the automatic evolution, 96 mutants were automatically constructed and tested for activity in a single round, which is corresponding to the automated 96-channel electronic pipette in biofoundry. A total of 384 mutants were constructed and tested during the four rounds evolution. In the case of



engineering pCNF-RS, a single round of experimental testing takes around 59 h on our biofoundry including around 24 h of primers shipping delay, while ML model training and new variants prediction take less than 1 h. The four design-build-test-learn cycles only took 240 h, around 10 days. Previously, a combinatorial mutation library was constructed at H283, P284, M285, D286 of *pAzFRS* and screened using a strategy of positive and negative selections⁴⁴. After two rounds

of positive-negative selections using *cat/barnase* system, they obtained improved variants against *pAzF*. The whole process took around 5 days for preparations, 7 days for one round positive selection, 7 days for one round negative selection and 5 days for final screening based on fluorescence, 38 days in total for two rounds of selections (Supplementary Fig. 5). However, experimental failure could happen at steps such as mutation library construction, cells

Fig. 6 | Detailed characterization of pCNF-RS variants. **a** Yield of sfGFP incorporated with pAcF aided by pCNF-RS and its variants. +, cell culture in presence of pAcF; -, cell culture in absence of pAcF. The concentration of purified sfGFP in presence of pAcF was 0.1 mg/mL, 0.6 mg/mL and 1.7 mg/mL for pCNF-RS, M-R3 and M-R4, respectively. **b** Molecular weight determination of the protein sfGFP-pAcF. Expected molecular weight is 27,832.0 Da and the peak of mass 27,831.0 Da was observed. **c** Efficiency of pCNF-RS and its variants suppressing multiple amber codons. Data are presented as mean values ± SD over three independent

repetitions. **d** Incorporation efficiency of various ncAAs by pCNF-RS, M-R3 and M-R4. Data are presented as mean values ± SD over three independent repetitions. Structures of ncAAs tested are shown. **e** Expression of transketolase and nanobody with ncAAs incorporated by pCNF-RS and M-R4. The protein concentration was 1.8 mg/mL and 12 mg/mL for TK Y385pCNF expressed with pCNF-RS and M-R4 in presence of pCNF, respectively. Three independent experiments were conducted to confirm the results of Fig. 6a, e. Source data are provided as a Source Data file.

transformation, plasmid Midiprep and isolation. Additionally, more rounds of positive-negative screening might be needed to obtain improved variants. This would result in the whole process taking much longer in real situation.

Other studies have developed automated and semi-autonomous system for protein engineering. A self-driving autonomous machines for protein landscape exploration (SAMPLE) platform was developed for fully autonomous protein engineering³². The SAMPLE agent uses bayesian optimization method with Gaussian process models trained on a cytochrome P450 dataset for navigating the protein landscape. With the purpose of engineering glycoside hydrolase for improved thermostability, the algorithm was applied to predict effect of DNA constructs assembled with DNA fragments from natural GH1 family members, 1352 unique GH1 sequences in total. The platform assembled pre-synthesized DNA fragments from different homologues and three sequences per round, 20 rounds in total. Additionally, a bayesian optimization-guided evolutionary algorithm (BO-EVO) along with robotic experiments were also developed for protein engineering, and a four-site combinatorial library of RhlA, an enzyme responsible for synthesizing the lipid moieties of rhamnolipid, was explored by sampling 384 mutants each round for 4 rounds, 1536 mutants in total³³.

PLMeAE differed from these earlier studies in both the ML models used and the function of automatic facilities. PLMeAE used the protein language model for zero-shot prediction of informed variants and encoding the protein for training fitness predictor based on a multi-layer perceptron, while previous studies mainly applied machine learning models such as bayesian optimization trained on public dataset, which were not capable of carrying out zero-shot prediction. Over the nature evolution, proteins adhere to a set of inherent principles for achieving optimal stability, function, and efficiency. PLMs trained on vast datasets of natural proteins learned and exploited these principles, which allowed zero-shot optimization of specific proteins. However, enzymes possess remarkable diversity in terms of their classes and catalytic mechanisms, remaining a challenge for selecting promising variants using PLM. Additionally, when engineering an enzyme for accepting unnatural substrates to possess new function, it has a high possibility that no natural enzyme possesses this function. In this situation, the PLMs might not learn the principle to acquire this new function, and hence might not be able to provide improved variants. For example, in our study, with 4 mutation target sites, PLM designed 96 first-round variants and only 6 of them showed higher fitness value compared to the wild type. Hence, PLMeAE applied PLM primarily to propose informed mutants and new mutation sites, and combined the PLM with a supervised machine learning model for navigating the fitness landscape of enzyme. This took advantages of principles learned in PLMs to guide the evolution of enzymes, and ensured the protein fitness improvement as the evolution went. As for the more challenging task that the target fitness landscape is distant from that of the native protein, PLM would predict the variants with a very low accuracy, leading to no positive variants in the training dataset. However, since we use an ITC-based sampling strategy, the variants predicted by the PLM have a high diversity in sequence, which is helpful for supervised ML model to learn the fitness landscape. Additionally, since only several amino acids are selected as mutation targets, the sequence space is limited. It is still highly possible to obtain

improved variants using PLMeAE, although more rounds of evolution may be needed for such a challenging task.

Automatic protein variants build and test was composed of two critical tasks including the construction of plasmid expressing protein mutants and automatic measurement of enzyme activity. A variety of strategies have been developed for automating the construction of plasmid DNA using the DNA fragments assembly methods such as Golden gate assembly³², Gibson assembly⁴⁹ and those using *Pyrococcus furiosus* Argonaute based artificial restriction enzymes²⁹. In our study, as the four mutation target sites are consecutive, we automated the QuikChange method to construct mutants, which is one of the most widely used site-directed mutagenesis strategies.

Another critical task in automatic protein variants build and test is the automatic measurement of protein functions, which is challenging for some enzymes due to low-throughput activity measurement methods such as gas chromatography and liquid chromatography. Hence, it was important to develop high-throughput enzyme activity measurement methods such as those based on fluorescence and colorimetric assay. Here, we applied PLMeAE to engineer the *MjTyrRS* for improved enzyme activity which could be measured based on fluorescence of sfGFP with amber codon inserted. This study showcased the successful application of PLMeAE in engineering tRNA synthetase for improved incorporation efficiency of ncAAs into a protein. A potent technique called genetic code expansion (GCE) makes it possible to add ncAAs to a protein at specific positions⁵⁰. However, the commonly used tRNA synthetase variants in GCE showed low activity towards ncAAs and hence led to low yield of protein incorporated with ncAAs. A method has been developed for the directed evolution of aaRS involving both positive and negative selections based on the ability to suppress a nonsense mutation in the presence of the desired ncAA⁵¹. However, this process requires multiple rounds of selection for each new substrate, which is time-consuming and laborious. The success of PLMeAE in engineering pCNF-RS for improved activity could be potentially used for engineering other tRNA synthetases such as pyrolysyl tRNA synthetase⁵², and hence expand the applications of GCE in protein engineering or exploring relationship between protein structure and function.

The PLMeAE platform has the potential to expand applications to engineer enzymes with activity measured by liquid chromatography (LC), gas chromatography (GC) and mass spectrometry (MS). Various automation set up has been developed to accelerate chromatographic and mass spectrometric analysis⁵³. For example, a number of liquid chromatography systems are equipped with 96-well plate auto-samplers, which are easily operated by robotic arms and scheduling software. Moreover, employing a low-pressure GC setup combined with low thermal mass (LTM) ovens enables rapid cycle times of under 1 min for detecting samples⁵⁴. Various technologies including liquid handler robots, microfluidics, and automated solvent extraction systems, have been leveraged to automate sample preparation for chromatographic analysis⁵³. Recent advancements have also facilitated high-throughput mass spectrometry (MS) by enabling direct introduction of analytes into mass spectrometers through methods such as laser, microfluidics, and acoustics⁵⁵. An automated matrix-assisted laser desorption/ionization time of flight (MALDI-ToF) MS has been developed for the direct analysis of bacterial and yeast colonies, achieving a throughput of approximately 2 s per sample^{56,57}.

Additionally, a robotic platform has been designed to integrate a liquid handler, plate dryer, sample stacker, and mass spectrometer, enabling fully automated MALDI MS analysis and the processing of over a million samples in a week³⁸. Integrating these automated, high-throughput assays into the BioFoundry would significantly expand the application of PLMeAE in engineering other enzymes.

Despite the potential of PLMeAE in protein engineering, developing a new PLMeAE system from scratch is challenging due to the specialized expertise required at the intersection of synthetic biology, computer science, and laboratory automation and robotics. Additionally, establishing new biofoundries is costly. To overcome these challenges, it is crucial to foster collaboration among researchers from diverse disciplines and to train the next generation of synthetic biology researchers with development in artificial intelligence and laboratory automation²⁸. In the future, researchers will be able to efficiently perform protein engineering with minimal human intervention.

Methods

PLM-enabled protein-fitness learning and variant design

In mathematical terms, the process of directed evolution process can be modeled as a black-box optimization problem, wherein the goal is to identify the amino acid sequence S^* that achieves the highest fitness within an extensive combinatorial search space. This can be formulated as

$$S^* = \arg \max_{S \in S^*} F(S) \quad (1)$$

where the function $F : S \rightarrow R$ evaluates the fitness of a sequence S .

Module I: Single-point mutations design. For the protein without knowing the mutation target sites, we use prior knowledge of pre-trained models for computational analysis to identify at which sites of the protein mutations should occur. The study began with a wild type protein sequence, systematically mutated at all sites to generate a comprehensive library of single amino acid variants. This process generated a comprehensive library encompassing all possible single amino acid variants (excluding the original residue) across the protein's length, resulting in a dataset of N^*20 sequences, where η is the number of mutable positions. We employed the ESM-2 model, for computational evaluation of the mutation library. Sequences were batch-processed to compute their likelihood scores, reflecting the model's assessment of each variant's compatibility with the evolutionary patterns learned during training. According to the computed likelihoods, we selected the top 96 sequences with the highest scores for further wet-lab experimental analysis. This threshold of 96 was chosen to balance the need for a sufficiently diverse set of mutants for experimental validation against the constraints of experimental throughput and resources. The selected variants underwent experimental fitness assays to measure their functional performance relative to the wild-type protein. These assays provided empirical data on the mutations' effects, contributing to a labeled dataset for subsequent rounds of fitness prediction, variant sampling, and experiment validation. The mutation sites in these experimentally measured high-fitness variants are considered highly potential for enhancing the fitness level of the wild-type protein. These sites are thus used in subsequent rounds of variant design. The improved single variants identified through this process can be further selected as targets for additional fitness enhancements using Module II.

Module II: Combinatorial variants design

Design of informative variants. To tackle the challenge of choosing the most informative mutants for annotation and training in machine learning-assisted directed evolution, we propose an innovative variant

sampling strategy that leverages a PLM with a novel measure of sample informativeness based on the concept of information transport complexity (ITC). This strategy employs a mask-prediction technique in PLMs and assesses the mutations informativeness by calculating how well the selected subset represents the whole set of 20 amino acids. In our methodology, we initially identify critical mutation sites within a protein sequence and then mask these sites. Then this masked protein sequence, represented as $S' = [s_1, \dots, [MASK], \dots, s_N]$, is processed by a PLM, which predicts the possible amino acids for each masked site based on the contextual relationship within the sequence.

For each mask, a selection of amino acids is made based on variant richness. It assesses how well a subset of selected amino acids, V_t , captures the essential information of the entire set of 20 natural amino acids, V . This is approached by the ITC score between the full set and the subset, aiming to minimize the distance indicating a high level of information capture. This involves solving an optimization problem to find an efficient information transport plan, utilizing advanced techniques to measure the discrepancy in information content between V and V_t .

$$r_a = \text{ITC}(V, V_t), t \in [1, C_{20}^c] \quad (2)$$

The optimization problem involving two amino acid sets is defined using a probability distribution for the original set $g(V)$ and its subset $g(V_t)$. To address this, the ITC score is implemented as an entropic regularization-based unbalanced optimal transport problem, which is formulated as:

$$\begin{aligned} \text{ITC}(V, V_t) = \min_{\eta} & (\eta, M)_F + \varepsilon \cdot \Psi(\eta) + \alpha \cdot (\text{KL}(\eta, g(V)) \\ & + \text{KL}(\eta^T, g(V_t))) \text{ s.t. } \eta \geq 0 \end{aligned} \quad (3)$$

Here, η represents the optimal transport plan, while M denotes the cost matrix containing pairwise Euclidean distances of token embeddings. The notation $(\cdot, \cdot)_F$ refers to the Frobenius dot product. The term $\Psi(\eta) = \sum_{i,j} \eta_{i,j} \log(\eta_{i,j})$ is the entropic regularization term, and KL represents the Kullback-Leibler divergence. The Coefficients ε and α correspond to the entropy regularization and marginal relaxation, respectively.

Training the fitness predictor. The annotated data samples are subsequently employed to train a fitness predictor F through a regression loss approach. The predictor integrates a protein language model (PLM), denoted as G , and a multi-layer perceptron (MLP), denoted as H . Given the limited size of the annotated data, the parameters of the PLM are frozen to mitigate overfitting, focusing only on optimizing the MLP parameters. The mean absolute error is minimized during training as follows:

$$F = \frac{1}{M} \sum_{S \in \tilde{S}} |H(G(S)) - y(S)| \quad (4)$$

where y denotes the annotated fitness value of the mutant S , and M represents the total number of mutants.

Evaluation of ESM models on GB1 dataset

The GB1 dataset comprises protein variants derived from the well-characterized B1 domain of protein G (GB1), a 56-amino-acid immunoglobulin-binding protein. Mutations were introduced at four specific positions (V39, D40, G41, and V54) within this domain³⁴. The dataset was used to systematically evaluate the fitness of these variants. We employed a variety of ESM models including both the ESM1 and ESM2 models^{36,59}, as well as the ESM3 model⁶⁰, to compute fitness scores for the GB1 dataset. The specific models used in this study include ESM1b (T33, 650 M, UR50S), ESM1v models (T33, 650 M, UR90S), ESM2 models (from T12 35 M to T33 650 M), ESM3 model

(esm3-sm-open-v1). Each model was pre-trained on large-scale protein sequence datasets and is capable of accurately capturing evolutionary information embedded in protein sequences.

Fitness scores were calculated using the Masked Marginal Score method, which evaluates the likelihood of mutated amino acids relative to the wild-type sequence. For each sequence, specific mutations were introduced by masking positions corresponding to the mutations in the GB1 dataset. The fitness score for each mutation was computed as the difference between the log probabilities of the wild-type and mutant amino acids at each mutated position, as follows:

$$\text{Fitness} = \sum_{i=1}^n (\log P(\text{mutant}_i) - \log P(\text{wild-type}_i)) \quad (5)$$

where $P(\text{mutant}_i)$ and $P(\text{wild-type}_i)$ are the predicted probabilities at position i. Higher scores indicate more favorable mutations. The Masked Marginal Score method was applied across all models to ensure consistency and comparability of the results. To assess the accuracy of the predictions, the Spearman rank correlation coefficient was calculated between the predicted fitness scores and experimentally measured fitness values.

Bacterial strains and reagents

p-cyanophenylalanine (*p*CNF, ncAA1), *p*-chlorophenylalanine (*p*ClF, ncAA2), *p*-fluorophenylalanine (*p*FF, ncAA3), *p*-zidophenylalanine (*p*AzF, ncAA4), *p*-isopropylphenylalanine (*p*-isopropylF, ncAAS), *p*-nitrophenylalanine (*p*NO₂F, ncAA6), *p*-phenylphenylalanine (ncAA7), benzylserine (ncAA8), *O*-methyl-Tyr (OmeY, ncAA9), *O*-tert-butyltyrosine (ncAA10) were purchased from Aladdin (Shanghai, China), Bidepharm (Shanghai, China) and Macklin (Shanghai, China). The 100 mM ncAA solutions were prepared by dissolving the amino acids in water, with addition of NaOH to ensure complete solubilization. *DpnI* enzyme was obtained from TaKara (Kusatsu, Japan), while Super-Fidelity DNA Polymerase (phanta) was ordered from Vazyme Biotech Co., Ltd (Nanjing, China). The expression vector of sfGFP used was pET-21a, and the host strain used was *E.coli* C321.ΔA. expT7 (a gift from Xiaowei Gao⁶¹). The pULTRA-CNF plasmid expressing pCNF-RS was kindly provided by Prof. Peter Schultz (Addgene plasmid #48215) (Supplementary Table 4). The plasmid pQR711 for expressing transketolase was a gift from Prof. John Ward⁶² and Prof. Paul Dalby. The gene of sfGFP2TAG and oligonucleotide primers for constructing sfGFP mutations with amber codon insertion were synthesized by Tsingke Biotechnology Co., Ltd (Beijing, China) (Supplementary Tables 5 and 6).

Automation of *MjTyrRS* variants construction and test

In the first step, a Python script was used to generate all primers sequences for site-directed mutagenesis and sent to Tsingke (Beijing, China) for synthesis (Supplementary Data 4). The 25 μL PCR reaction mixture, prepared by the automated workstation (Evo, Tecan Trading AG, Switzerland), consisted of 12.5 μL of 2× Phanta Flash Master Mix (Vazyme, Nanjing, China), 1 μL of pCNF-RS plasmid (10 ng/μL), 0.5 μL of each primer, and 10.5 μL of deionized water. Then the PCR plates were sealed by a plate sealer (ALPS, ThermoFisher, Waltham, MA, USA) and centrifuged (Vispin, Agilent, Santa Clara, CA, USA) for the PCR reaction carried out by an automatic thermocycler (ThermoFisher, Waltham, MA, USA). The PCR was carried out under the following steps: 98 °C for 30 s, 30 circles of 95 °C for 10 s, 60 °C for 5 s and 72 °C for 30 s, and a final extension of 72 °C for 1 min. Subsequently, the plates were unsealed by automated plate seal removal (Xpeel, Sopachem, Zwevegem, Belgium) and transferred to the acoustic liquid handler (Echo, Tecan Trading AG, Switzerland) for *DpnI* digestion for 30 min at 37 °C. The reaction solution contains 10 μL of PCR product, 1 μL *DpnI* and 1.2 μL buffer. *E.coli* C321.ΔA. expT7 competent cells containing sfGFP2TAG plasmid were used for variants construction and further enzyme activity measurement. 2 μL of digested solution was added to

20 μL competent cells by automation workstation (Evo, Tecan Trading AG, Switzerland). After incubating at 4 °C for 30 min, heat shock at 42 °C for 90 s was performed. The samples were then instantly placed on ice for 2 min. Following this, 200 μL of Luria-Bertani (LB) medium was added to each sample, and they were incubated in a shaker for 1 hour. Transformant plating was performed by spraying the cells onto eight-well agar plates using the 8-channel pipetting system in the automated workstation (Fluent, Tecan Trading AG, Switzerland). The plates were labeled with a microplate labeler (Agilent Labeler, Santa Clara, CA, USA) and then incubated in the automated incubator (Cytomat 10 C, ThermoFisher, Waltham, MA, USA) for 14 h.

On the second day, the clones were picked into 800 μL of LB culture medium containing 100 μM ampicillin sodium (Amp) and 50 μM spectinomycin dihydrochloride pentahydrate (Spe) using automation workstation colony-picker (Fluent, Tecan Trading AG, Switzerland), and cultured at 37 °C 600 rpm for 7 h in the automated incubator with orbital shaking (Cytomat 2 C Tos, ThermoFisher, Waltham, MA, USA). Then, a robotic centrifuge (Rotanta, Hettich, Tuttlingen, Germany) was used to collect 200 μL of bacterial culture, which was then resuspended in GMML medium (M9 salts, 1% glycerol, 1 mM MgSO₄ and 0.1 mM CaCl₂ and 0.3 mM Leucine) containing 50 μM Amp, 100 μM Spe, 0.2 mM IPTG in presence or absence of 1 mM *p*AcF using a reagent dispenser (Multidrop Combi, ThermoFisher, Waltham, MA, USA) and cultured for another 11 h in the automated incubator with orbital shaking (Cytomat 2 C Tos, ThermoFisher, Waltham, MA, USA). The remaining bacterial culture was used for sanger sequencing to verify sequence accuracy and for glycerol stock preservation in a -80 °C freezer (Arktic, Sptlabtech, Shanghai, China).

On the third day, 200 μL of overnight cultured bacterial suspension was transferred to a black microplate, the activities of these mutants were measured by an automatic microplate reader (CLARIOstar, Cary, NC, USA), and the data were analyzed by Momentum DataMiner.

Measurement of aminoacyl tRNA-Synthetase/tRNA pair activity suppressing multiple codons

After four rounds of mutant construction and activity measurement, we measured the activity of aminoacyl-tRNA synthetase-tRNA pairs with multiple codon suppression for the mutants, M-R3 from the third round and M-R4 from the fourth round. Firstly, pET21a-sfGFP with one amber sites (sfGFP-1amb, corresponding to position D36), two amber sites (sfGFP-2amb, corresponding to positions D36, K101), three amber sites (sfGFP-3amb, corresponding to positions D36, K101, D190) and four amber sites (sfGFP-4amb, corresponding to positions D36, K101, D190, E213) were constructed (Supplementary Table 5). PCR was performed using Phanta Flash Master Mix (Nanjing, China) at 95 °C for 2 min, with 30 circles of 95 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min and a final extension of 72 °C for 2 min followed by *DpnI* digestion. The constructed plasmids were transformed to *E. coli* BL21(DE3) competent cells and confirmed by DNA sequencing using T7 and T7ter universal primer. Next, proceed with plasmid extraction and separately co-transform the plasmids with pCNF-RS, M-R3, and M-R4 into *E. coli* C321.ΔA. expT7 competent cells. Five clones of each transformant were picked into 800 μL LB liquid medium containing 100 μM Amp and 50 μM Spe, and cultured at 37 °C 600 rpm for 7 h. Then, the medium was changed to GMML containing Amp, Spe, IPTG (0.2 mM) and with or without *p*AcF (1 mM) as described before. Finally, the activities of pCNF-RS, M-R3, and M-R4 suppressing multiple amber codons with *p*AcF were measured by the microplate reader (CLARIOstar).

Measurement of activity of aminoacyl tRNA-Synthetase/tRNA pair towards various substrates

Ten different ncAAs were purchased and dissolved in water to a final concentration of 100 mM, with a few drops of NaOH added to assist

dissolution. The plasmids expressing pCNF-RS, M-R3, and M-R4 were separately co-transformed into *E. coli* C321.ΔA. expT7 competent cells containing pET21a-sfGFP2TAG plasmid. The clones of each transformant were picked and cultured in 96-deep-well plates (2.2 mL) also containing 100 μM Amp and 50 μM Spe, and induced with 0.2 mM IPTG in presence or absence of ten different ncAAs (each 1 mM). The activities of these aminoacyl tRNA-Synthetase/tRNA pair towards various substrates were determined by the microplate reader (CLARIOstar).

Purification of sfGFP protein with ncAAs incorporated

The expression and purification of sfGFP with ncAAs incorporated was carried out as before⁶³. Five milliliters of LB medium containing 50 μM spectinomycin and 100 μM ampicillin was inoculated with *E. coli* C321.ΔA. expT7 cells expressing sfGFP2TAG and *Mj*TyrRS-tRNA pair for incorporating ncAAs at the 2nd position of sfGFP. The cells were cultured at 37 °C and 220 rpm for 12–14 h. After that, 1 mL of the cultures were transferred to 100 mL of GMML medium and grown under the same conditions. The cells were then induced with 0.2 mM IPTG, with or without 1 mM ncAAs, when the OD₆₀₀ reached 0.6–0.8. After 18 h of cultivation at 18 °C and 220 rpm, the cells were harvested by centrifugation at 2272 g for 20 minutes. The cell pellets were re-suspended in 50 mM Tris-HCl buffer (pH 8.0) and sonicated on ice. The soluble supernatant fraction was then collected by centrifugation at 7012 g for 20 min, and filtered using a PES syringe filter. Ni-NTA column was then used for sfGFP purification. Lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0) was first used to equilibrate the column, which was then loaded with the filtered supernatant. Impurities were washed away with wash buffer containing 50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH 8.0. The proteins were finally eluted with elution buffer containing 50 mM NaH₂PO₄, 300 mM NaCl, 250 mM imidazole, pH 8.0. The proteins were concentrated using Amicon Ultra centrifugal filter (Millipore, Billerica, MA). Protein purity was confirmed by 4–12% FuturePAGE SDS-PAGE (ACE Biotechnology, Changzhou, China). Concentrations were determined by NanoDrop One A280 (Thermo Scientific, Waltham, MA).

Purification of transketolase with ncAAs incorporated

The amberless *E. coli* strain C321.ΔA. expT7 was utilized for transketolase expression (Supplementary Table 5). The mutant S385TAG (with a UAG stop codon at position Ser385), was co-expressed overnight in C321.ΔA. expT7 cells with the ncAA-incorporation machinery including pCNF-RS and M-R4. Following this, clones were transferred to 5 mL of LB medium containing 100 μM ampicillin and 50 μM spectinomycin. A 100 mL culture was then grown to an OD₆₀₀ of 0.6–0.8 in a 250 mL shake flask. Expression was induced with 0.2 mM IPTG and 1 mM pCNF at 37 °C and 220 rpm for 7 h. Cells were harvested, resuspended in lysis buffer, and sonicated to release the target protein. The protein was purified using Ni-NTA spin columns. The lysis buffer contained 500 mM NaCl, 5 mM imidazole, 2.4 mM thiamine diphosphate (ThDP), and 9 mM MgCl₂ in 20 mM Tris-HCl. For elution, 500 mM NaCl and 500 mM imidazole were dissolved in 20 mM Tris-HCl. NaCl and imidazole were removed by filtration, using a buffer containing 2.4 mM ThDP and 9 mM MgCl₂ in 50 mM Tris-HCl. The pH of all solutions was adjusted to 7.0. Protein concentration was measured using the Bradford Reagent (Hercules, California, USA).

Purification of nanobody with ncAAs incorporated

The wild-type nanobody sequence was cloned into the pET-32a(+) plasmid between the XbaI and NotI restriction sites, and a UAG stop codon was introduced at positions Tyr80 and Pro41 in nanobody through site-directed mutagenesis (Supplementary Table 5). To incorporate the ncAA into the nanobody, the bacteria were transformed with a plasmid encoding the evolved orthogonal aminoacyl-tRNA synthetase (pCNF-RS, M-R4) and its corresponding tRNA. These

plasmids were then introduced into *E. coli* TransB (DE3) cells, which were cultured in LB medium containing ampicillin and spectinomycin. Protein expression was induced under the same conditions as previously described, with the addition of 1 mM pAcF for ncAA incorporation. The protein was purified, concentrated and analyzed following the same protocol used for sfGFP purification.

Mass spectrometry

LC-ESI-MS was performed as before, using an Agilent 1290 Infinity II LC system connected to a 6545 Q-TOF mass spectrometer (Agilent, UK)⁵². Samples of 5 μL of 0.4 mg/mL purified sfGFP were injected into an Agilent PLRP-S (50 mm × 2.1 mm, 1000 Å, 5 μm) column maintained at 30 °C. The QTOF mass spectrometer scanned m/z from 100 to 3100 Da. Positive electrospray ionization (ESI) was applied with a capillary voltage of 4000 V, a nozzle voltage of 500 V, a fragmentor voltage of 175 V, a skimmer voltage of 65 V, and an octopole RF peak of 750 V. Nitrogen was used as the nebulizer gas with a pressure of 45 psi and a desolvation gas flow rate of 5 L/min. Data were analyzed in MassHunter BioConfirm software (version B.10.00) and deconvolved with the maximum entropy algorithm. Protparam (<http://us.expasy.org/tools/protparam.html>) was used to calculate theoretical masses of wild-type protein, and the theoretical masses for ncAA-incorporated proteins were manually adjusted. Totally, 11 protein samples were analyzed with one biological replicate.

Data availability

Authors declare that all data supporting the findings of this study are available within the paper and its supplementary information files. The GB1 dataset is available from BioProject database with accession code PRJNA278685. The UBC9 dataset is available from the Biostudies database with accession number S-BSST60 [www.ebi.ac.uk/biostudies]. And the ubiquitin dataset has been reported in the literature³⁸. The PDB structure 1J1U is used. LC-MS data underlying Fig. 6b, Supplementary fig. 10-19 have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD058768. Source data are provided with this paper.

Code availability

All code and the necessary data to run that code are accessible at GitHub (<https://github.com/HICAI-ZJU/PLMeAE>). The source code has also been deposited to Zenodo⁶⁴ at <https://doi.org/10.5281/zenodo.1461351>.

References

1. Chiu, M. L. & Gilliland, G. L. Engineering antibody therapeutics. *Curr. Opin. Struct. Biol.* **38**, 163–173 (2016).
2. Bottcher, D. & Bornscheuer, U. T. Protein engineering of microbial enzymes. *Curr. Opin. Microbiol.* **13**, 274–282 (2010).
3. Dalby, P. A. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **21**, 473–480 (2011).
4. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
5. Rubingh, D. N. Protein engineering from a bioindustrial point of view. *Curr. Opin. Biotechnol.* **8**, 417–422 (1997).
6. Wang, Y. et al. Directed Evolution: Methodologies and Applications. *Chem. Rev.* **121**, 12384–12444 (2021).
7. Bloom, J. D. & Arnold, F. H. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. USA* **106**, 9995–10000 (2009).
8. Yang, J., Li, F. Z. & Arnold, F. H. Opportunities and Challenges for Machine Learning-Assisted Enzyme Engineering. *ACS Cent. Sci.* **10**, 226–241 (2024).
9. Yu, H., Ma, S., Li, Y. & Dalby, P. A. Hot spots-making directed evolution easier. *Biotechnol. Adv.* **56**, 107926 (2022).

10. Reetz, M. T., Wang, L. W. & Bocola, M. Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem. Int Ed. Engl.* **45**, 1236–1241 (2006).
11. Reetz, M. T., Carballera, J. D. & Vogel, A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem. Int Ed. Engl.* **45**, 7745–7751 (2006).
12. Sebestova, E., Bendl, J., Brezovsky, J. & Damborsky, J. Computational tools for designing smart libraries. *Methods Mol. Biol.* **1179**, 291–314 (2014).
13. Kalvet, I. et al. Design of Heme Enzymes with a Tunable Substrate Binding Pocket Adjacent to an Open Metal Coordination Site. *J. Am. Chem. Soc.* **145**, 14307–14315 (2023).
14. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
15. Sumbalova, L., Stourac, J., Martinek, T., Bednar, D. & Damborsky, J. HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res* **46**, W356–W362 (2018).
16. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
17. Wittmann, B. J., Johnston, K. E., Wu, Z. & Arnold, F. H. Advances in machine learning for directed evolution. *Curr. Opin. Struct. Biol.* **69**, 11–18 (2021).
18. Mazurenko, S., Prokop, Z. & Damborsky, J. Machine Learning in Enzyme Engineering. *ACS Catal.* **10**, 1210–1223 (2019).
19. Romero, P. A., Krause, A. & Arnold, F. H. Navigating the protein fitness landscape with Gaussian processes. *Proc. Natl. Acad. Sci. USA* **110**, E193–E201 (2013).
20. Buchler, J. et al. Algorithm-aided engineering of aliphatic halogenase WelO5* for the asymmetric late-stage functionalization of soraphens. *Nat. Commun.* **13**, 371 (2022).
21. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. USA* **116**, 8852–8858 (2019).
22. Saito, Y. et al. Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins. *ACS Synth. Biol.* **7**, 2014–2022 (2018).
23. Ferruz, N. & Höcker, B. Controllable protein design with language models. *Nat. Mach. Intell.* **4**, 521–532 (2022).
24. Hie, B. L., Yang, K. K. & Kim, P. S. Evolutionary velocity with protein language models predicts evolutionary dynamics of diverse proteins. *Cell Syst.* **13**, 274–285.e276 (2022).
25. Madani, A. et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **41**, 1099–1106 (2023).
26. Hie, B. L. et al. Efficient evolution of human antibodies from general protein language models. *Nat. Biotechnol.* **42**, 275–283 (2024).
27. He, Y. et al. Protein language models-assisted optimization of a uracil-N-glycosylase variant enables programmable T-to-G and T-to-C base editing. *Mol. Cell* **84**, 1257–1270.e1256 (2024).
28. Yu, T., Boob, A. G., Singh, N., Su, Y. & Zhao, H. In vitro continuous protein evolution empowered by machine learning and automation. *Cell Syst.* **14**, 633–644 (2023).
29. Enghiad, B. et al. PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction. *Nat. Commun.* **13**, 2697 (2022).
30. Hamedirad, M. et al. Towards a fully automated algorithm driven platform for biosystems design. *Nat. Commun.* **10**, 5150 (2019).
31. Li, S. et al. Automated high-throughput genome editing platform with an AI learning in situ prediction model. *Nat. Commun.* **13**, 7386 (2022).
32. Rapp, J. T., Bremer, B. J. & Romero, P. A. Self-driving laboratories to autonomously navigate the protein fitness landscape. *Nat. Chem. Eng.* **1**, 97–107 (2024).
33. Hu, R. et al. Protein engineering via Bayesian optimization-guided evolutionary algorithm and robotic experiments. *Brief Bioinform.* **24**, 1–9 (2023).
34. Wu, N. C., Dai, L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *Elife* **5**, e16965 (2016).
35. Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
36. Rives, A. et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. USA* **118**, e2016239118 (2021).
37. Weile, J. et al. A framework for exhaustively mapping functional missense variants. *Mol. Syst. Biol.* **13**, 957 (2017).
38. Roscoe, B. P. & Bolon, D. N. Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast. *J. Mol. Biol.* **426**, 2854–2870 (2014).
39. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498–500 (2001).
40. Hsieh, P. C. & Vaisvila, R. Protein engineering: single or multiple site-directed mutagenesis. *Methods Mol. Biol.* **978**, 173–186 (2013).
41. Krahn, N., Tharp, J. M., Crnkovic, A. & Soll, D. Engineering aminoacyl-tRNA synthetases for use in synthetic biology. *Enzymes* **48**, 351–395 (2020).
42. Young, D. D. et al. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894–1900 (2011).
43. Li, J. C., Liu, T., Wang, Y., Mehta, A. P. & Schultz, P. G. Enhancing Protein Stability with Genetically Encoded Noncanonical Amino Acids. *J. Am. Chem. Soc.* **140**, 15997–16000 (2018).
44. Gan, R. et al. Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins. *Biotechnol. Bioeng.* **114**, 1074–1086 (2017).
45. Qiu, Y. & Wei, G. W. CLADE 2.0: Evolution-Driven Cluster Learning-Assisted Directed Evolution. *J. Chem. Inf. Model* **62**, 4629–4641 (2022).
46. Qiu, Y., Hu, J. & Wei, G. W. Cluster learning-assisted directed evolution. *Nat. Comput. Sci.* **1**, 809–818 (2021).
47. Karki, S., Shi, F., Archer, J. J., Sistani, H. & Levis, R. J. Direct Analysis of Proteins from Solutions with High Salt Concentration Using Laser Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **29**, 1002–1011 (2018).
48. Wilkinson, H. C. & Dalby, P. A. Fine-tuning the activity and stability of an evolved enzyme active-site through noncanonical amino-acids. *FEBS J.* **288**, 1935–1955 (2021).
49. Chao, R. et al. Fully Automated One-Step Synthesis of Single-Transcript TALEN Pairs Using a Biological Foundry. *Acs Synth. Biol.* **6**, 678–685 (2017).
50. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
51. Bryson, D. I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
52. Liu, K. et al. An evolved pyrrolysyl-tRNA synthetase with poly-substrate specificity expands the toolbox for engineering enzymes with incorporation of noncanonical amino acids. *Bioresources Bio-process.* **10**, 92 (2023).
53. Zhang, J. et al. Accelerating strain engineering in biofuel research via build and test automation of synthetic biology. *Curr. Opin. Biotechnol.* **67**, 88–98 (2021).
54. Fialkov, A. B., Lehotay, S. J. & Amirav, A. Less than one minute low-pressure gas chromatography - mass spectrometry. *J. Chromatogr. A* **1612**, 460691 (2020).

55. Zhang, S. et al. Directed evolution of a cyclodipeptide synthase with new activities via label-free mass spectrometric screening. *Chem. Sci.* **13**, 7581–7586 (2022).
56. Xue, P. et al. A mass spectrometry-based high-throughput screening method for engineering fatty acid synthases with improved production of medium-chain fatty acids. *Biotechnol. Bioeng.* **117**, 2131–2138 (2020).
57. Si, T. et al. Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **139**, 12466–12473 (2017).
58. Lin, S. et al. Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **361**, eaar6236 (2018).
59. Meier, J. et al. Language models enable zero-shot prediction of the effects of mutations on protein function. *Adv. Neur. In.* **34**, 29287–29303 (2021).
60. Hayes, T. et al. Simulating 500 million years of evolution with a language model. *Science* **0**, eads0018 (2025).
61. Yi, H. et al. Comparative analyses of the transcriptome and proteome of *Escherichia coli* C321. ΔA and further improving its non-canonical amino acids containing protein expression ability by integration of T7 RNA polymerase. *Front Microbiol.* **12**, 744284 (2021).
62. French, C. & Ward, J. M. Improved production and stability of *E. coli* recombinants expressing transketolase for large scale biotransformation. *Biotechnol. Lett.* **17**, 247–252 (1995).
63. Chen, W. et al. Non-canonical amino acids uncover the significant impact of Tyr671 on Taq DNA polymerase catalytic activity. *FEBS J.* **291**, 2876–2896 (2024).
64. Zhang, Q. et al. Integrating protein language models and automatic biofoundry for enhanced protein evolution. <https://doi.org/10.5281/zenodo.14613518> (2025).

Acknowledgements

This work was supported by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (No. 2025C01097) to H.Y., Q. Z., K. D., L. Y., the Key Research and Development Program of China (Grant No. 2022YFA0913000), the National Natural Science Foundation of China (Grant No. 22108245), and the Fundamental Research Funds for the Central Universities (Grant No. 226-2022-00214) to H.Y. We would like to thank iBioFoundry and Core Facility at the Institute for Intelligent Bio/Chem Manufacturing, ZJU-Hangzhou Global Scientific and Technological Innovation Center. The authors also would like to thank AI + High Performance Computing Center of ZJU-ICI. We thank Prof. He Huang for kindly providing the plasmid for expressing nanobody.

Author contributions

H.Y., Q.Z., H.C. and L.Y. conceived the project and supervised the work. The ML model was designed, built and tested by M.Q., Y.W., Z.P., K.D., Y.Z., and J.Y. Automatic facilities were performed by Y.L., W.C., L.H., and J.W. Experiments were performed by W.C., Q.F.Z., D.L., and X.L. Data was analyzed by H.Y., W.C., and Q.Z. H.Y., Q.Z., and W.C. wrote the manuscript, which was edited and approved by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56751-8>.

Correspondence and requests for materials should be addressed to Huajun Chen or Haoran Yu.

Peer review information *Nature Communications* thanks Jiri Damborsky, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025