

SUPPORTING INFORMATION

Table S1: Nucleotide sequence of the cel5E.

Protein	Sequence
cel5E	ATGAAAGCGGTTGATCCGTTTGAAATGGTGCGCAAAATGGGCATGGGCACCAATCTGGGCAATACCCTGGAAGCGC CGTATGAAGGCAGCTGGAGCAAAAGCGCGATGGAATATTATTTTGACGACTTCAAGGCGGCGGGCTATAAGAATGT GCGCATTCCGGTGCGCTGGGATAATCATACCATGCGCACCTATCCGTATACCATTGATAAAGCGTTTCTGGACCGC GTGGAACAGGTGGTGGATTGGAGCTTAAGCCGCGGCTTTGTTACCATTATTAATAGCCATCATGACGACTGGATCA AGGAGGATTATAACGGCAATATCGAACGCTTCGAGAAGATCTGGGAGCAGATTGCGGAACGCTTTAAAAATAAAAG CGAGAACCTGCTGTTTCGAGATCATGAACGAACCGTTTGGCAATATTACCGATGAACAGATTGATGACATGAACAGC CGCATTCTGAAAATTATCCGCAAAACCAATCCGACCCGATTGTGATTATTGGCGGCGGCTATTGGAATAGCTATA ATACCCTGGTGAATATCAAGATCCCGGATGACCCGTATCTGATCGGCACCTTTCATTATTATGATCCGTATGAGTT CACCCACAAGTGGCGCGGCACCTGGGGTACTCAAGAAGATATGGATACCGTGGTGCGCGTGTGTTGATTTTGTGAAA AGCTGGAGCGATCGCAATAATATTCCGGTGATTTTGGCGAATTTGCGGTGATGGCGTATGCGGATCGCACCAGCC GTGTTAAATGGTATGATTTTATTAGCGACGCGGCGCTGGAACGCGGCTTTGCGTGTAGTGTGTTGGGATAATGGCGT GTTTGGCAGCCTGGATAATGATATGGCGATTATAATCGCGACACCCGCACCTTTGATACCGAAATTCTGAATGCG CTGTTTAAC

Table S2: Oligonucleotide primer sequences used for site-directed mutagenesis.

Name	Primer sequence
H205F_F	GTTACCTTTAAGTGGCGCGGCACC
H205F_R	CCACTTAAAGGTGAACCTCATACGGATC
S84A_F	TGGATTGGCGTTGAACGCCGGCGTTT
S84A_R	GGCTTAACGCCCAATCCACCACCTGT
A351_F	GCAAAAAGCATTATGGAATATTATTTTGACGACTTCA
A351_R	ATTTTAATCCTTTGTCCTCACGGTG
D259E_F	TTAAATGGTATGAATTTATTAGCGACGCGGC
D259E_R	CGCTAAATTAATCATACATTTAAACGGCTGG
N129D_F	AAGCGAGTCTGCTGTTGAGATCATG
N129D_R	CAGCAGATCCTCGCTTTATTTTAAAGCGTTC
N108Y_F	TATAACGGCTATATCGAACGCTTCGAGAAG
N108Y_R	CGTTCGATAGCGTTATAGCCGTTATATCTT
G88D_F	AGCGCGATTTTGTTACCATTATTAATAGCCATC
G88D_R	GTACAAATCGCGGCTTAGCT
T299E_F	CACCCCTTGATGAAGAAATTACCTGTG
T299E_R	GAATTTCTTCATCAAAGGTGCG
S152E_F	GCAATGAACAGCAATCTGAAAATTATCCG
S152E_R	GAATGCGTTCGTTTCATGTCATCAATCTGTTC
N106D_F	GGAGGATTATGATGGCAATATCGAACGCTC
N106D_R	GATATTGCCATCAATCCTCATCCGTGACGT
V227L_F	GTTTGATTTTATTAAGCTGGAGCGATCG
V227L_R	CCAGCTTTTAATAAAATCAAAACACGCGCAC
V220L_F	GATCCATTGTCGCGCTATTGATAGTT
V220L_R	CGCACAATGGTATCCATATCTTCTTGAG
N291D_F	GCGATTTATGATCGCGACACCCGCACC
N291D_R	GTGCTGGCGATATAAACGCCATATTCCAG
D218E_F	AAGATATGGAAACCGTGG TGC GCGTG
D218E_R	CCACGGTTTCATATCTTCTTGAGTACCC
S262I_F	ATTTTAGGATCGCGCGCTGGA
S262I_R	CCGCGTCAATAATAAACATCATACCATTAAACACGG

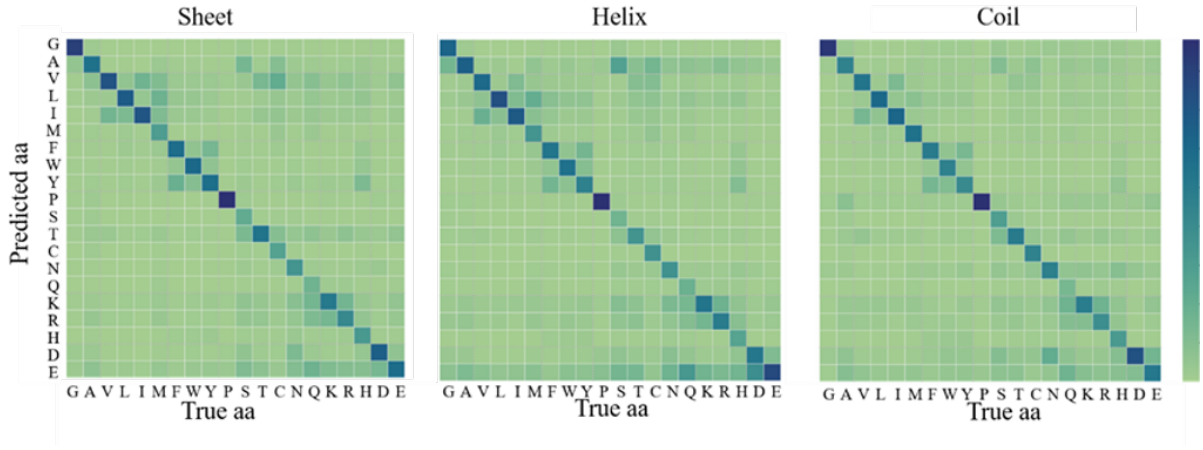


Figure S1: Heatmap illustrating amino acid replacement preferences between generated and true sequences across Sheet, Helix, and Coil secondary structures. Darker diagonal elements indicate high recovery accuracy for individual residues, and structured off-diagonal clusters highlight biochemically conservative mutations maintained by EZPro-Fit.

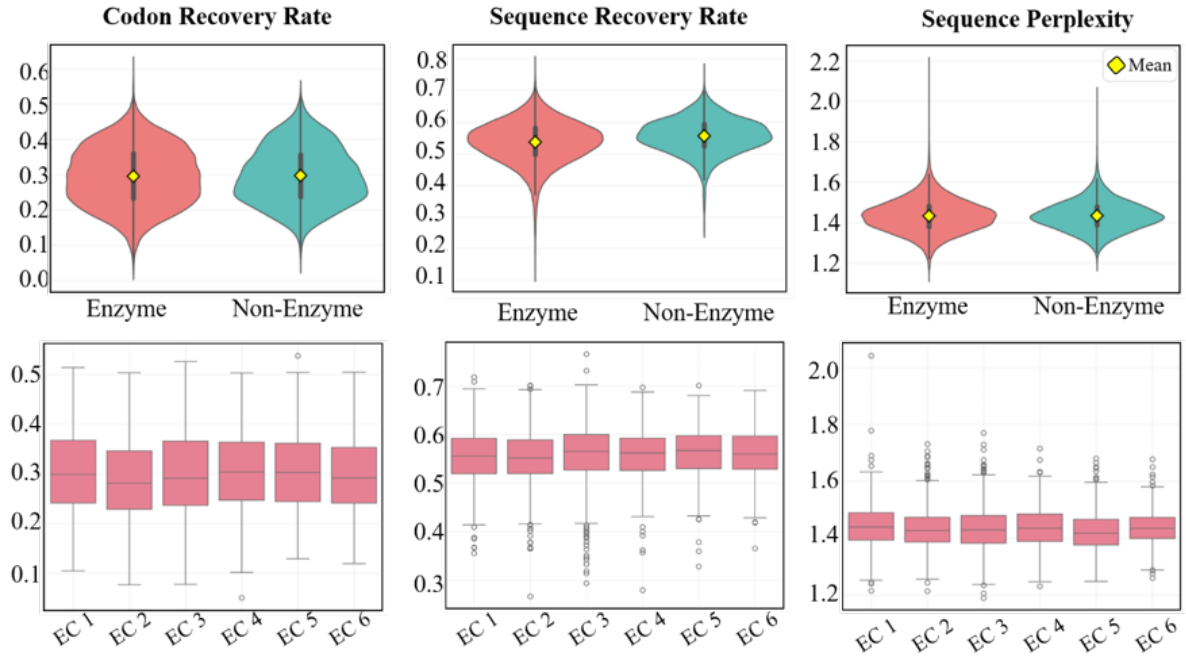


Figure S2: Sequence recovery metrics and perplexity distributions of EZPro-Fit. (Top) Violin plots comparing codon recovery, sequence recovery, and sequence perplexity between enzyme and non-enzyme datasets. (Bottom) Box plots showing consistent performance across the six Enzyme Commission (EC 1–6) classes, indicating uniform generative capacity across diverse catalytic functions.