

## 研究论文

DOI: 10.12211/2096-8280.2026-059

## 祖先序列重建结合计算筛选的黄酮合酶I挖掘与优化用于芹菜素合成

谢东奇<sup>1</sup>, 张顺志<sup>2</sup>, 邵茜<sup>2</sup>, 朱国梁<sup>1</sup>, 丁霏希<sup>1</sup>, 温霖<sup>2</sup>, 郑良振<sup>1</sup>, 王晟<sup>1</sup>( <sup>1</sup> 上海智峪生物科技有限公司, 上海 201210; <sup>2</sup> 中国新时代控股集团有限公司, 北京 100034 )

**摘要:** 黄酮合酶I (FNS-I) 催化柚皮素 C2-C3 位脱氢生成芹菜素, 是酶法合成黄酮的关键步骤。为拓展有限的 FNS-I 序列空间并提高催化活性, 本研究以欧芹 PcFNS-I 为起点, 在伞形科内收集 22 条 FNS-I 候选现生序列, 经祖先序列重建 (ASR) 获得 AncFNS-I-48 和 AncFNS-I-50, 并以 CataPro 排序结合物种代表性选取 3 条天然同源酶, 连同 PcFNS-I 构成 6 个候选酶, 以等量纯化蛋白测定相对活性; 随后以 AncFNS-I-48 为亲本, 利用 PSSM、Pythia 和 CataPro 筛选 29 个单点突变体, 配套建立  $A_{380}$  终点法高通量筛选流程。在 3 次平行测定中, 以 AncFNS-I-48 活性归一化为 1.00, PcFNS-I、AncFNS-I-50、CcFNS-I、DcFNS-I 和 CmFNS-I 的相对活性分别为 0.63、0.63、0.26、0.06 和 0.05。50 mmol/L 磷酸盐缓冲液中的 pH 响应表明, AncFNS-I-48 在 pH 6.0-8.5 范围内保持较高产物生成能力。结合柚皮素在偏碱性条件下较好的溶解性及后续较高底物负荷与工艺评价的需要, 后续条件优化统一在 pH 8.0 下进行; 25 - 30 °C 更适合长时间反应, 250  $\mu$ mol/L FeSO<sub>4</sub> 时芹菜素生成量最高, 为  $2.674 \pm 0.101$  mmol/L。采用  $A_{380}$  终点法, 以 SDS-PAGE 目标条带净积分灰度校正的相对初始反应速率评价 29 个突变体; V159E 和 H261F 分别为 WT 的  $3.06 \pm 0.26$  和  $1.50 \pm 0.04$  倍, 均显著高于 WT ( $p < 0.01$ )。在采用 HPLC 检测的 20 mL 反应体系验证中, 以 WT 归一化为 1.00, I115V、V159E、M230D 和 N297D 分别为  $2.15 \pm 0.15$ 、 $2.14 \pm 0.11$ 、 $2.02 \pm 0.10$  和  $1.88 \pm 0.13$ , 均显著高于 WT; M85L 为  $0.35 \pm 0.03$ , 显著低于 WT ( $p < 0.01$ )。综合两类指标, 本文优先推荐低酶投量下相对初始反应速率最高的 V159E; I115V 在 20 mL 制备级反应中的产物生成指标均值最高, 但与 V159E 接近, 且现有分析未检验二者间差异, 因此将其作为侧重制备级产物生成的备选。本研究建立了“谱系序列收集-祖先重建-AI 预测-计算突变-高通量筛选”的酶工程路线, 为 FNS-I 及 2-ODD 家族酶的挖掘与改造提供方法参考。

**关键词:** 黄酮合酶I; 伞形科; 祖先序列重建; CataPro; 定点突变; 高通量筛选**中图分类号:** Q814 **文献标志码:** A

收稿日期: 2026-07-01 修回日期: 2026-10-10

引用本文: 谢东奇, 张顺志, 邵茜, 朱国梁, 丁霏希, 温霖, 郑良振, 王晟. 祖先序列重建结合计算筛选的黄酮合酶I挖掘与优化用于芹菜素合成[J]. 合成生物学, 2026, 7. DOI: 10.12211/2096-8280.2026-059

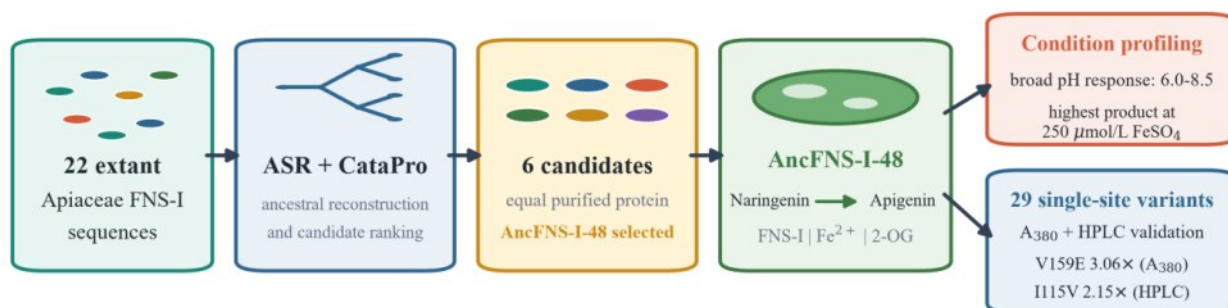
Citation: XIE Dongqi, ZHANG Shunzhi, SHAO Qian, ZHU Guoliang, DING Peixi, WEN Lin, ZHENG Liangzhen, WANG Sheng. Discovery and Optimization of Flavone Synthase I for Apigenin Synthesis by Combining Ancestral Sequence Reconstruction with Computational Screening [J]. Synthetic Biology Journal, 2026, 7. DOI: 10.12211/2096-8280.2026-059

# Discovery and Optimization of Flavone Synthase I for Apigenin Synthesis by Combining Ancestral Sequence Reconstruction with Computational Screening

XIE Dongqi<sup>1</sup>, ZHANG Shunzhi<sup>2</sup>, SHAO Qian<sup>2</sup>, ZHU Guoliang<sup>1</sup>, DING Peixi<sup>1</sup>, WEN Lin<sup>2</sup>,  
ZHENG Liangzhen<sup>1</sup>, WANG Sheng<sup>1</sup>

(<sup>1</sup>Shanghai Zelixir Biotech Company Limited, Shanghai 201210, China; <sup>2</sup>China New Era Group Corporation, Beijing 100034, China)

**Abstract:** Flavone synthase I (FNS-I) catalyzes C2-C3 dehydrogenation of naringenin to form apigenin and is a key enzyme for enzymatic flavone synthesis. To expand the limited FNS-I sequence space and improve catalytic activity, 22 extant Apiaceae FNS-I candidate sequences were collected using parsley PcFNS-I as the reference. Ancestral sequence reconstruction (ASR) yielded AncFNS-I-48 and AncFNS-I-50, while CataPro ranking combined with lineage coverage identified three natural homologs; together with PcFNS-I, these formed six candidates that were evaluated with equal amounts of purified protein. AncFNS-I-48 was then used as the parent for PSSM-, Pythia-, and CataPro-guided prioritization of 29 single-site variants and  $A_{380}$  endpoint high-throughput screening. Across three independent biological replicates, with AncFNS-I-48 normalized to 1.00, the relative activities of PcFNS-I, AncFNS-I-50, CcFNS-I, DcFNS-I, and CmFNS-I were 0.63, 0.63, 0.26, 0.06, and 0.05, respectively. Profiling in 50 mmol/L phosphate buffer showed that AncFNS-I-48 maintained comparatively high apigenin production from pH 6.0 to 8.5. Considering the improved apparent solubility of naringenin under mildly alkaline conditions and the needs of subsequent high-substrate-load and process evaluation, condition optimization was conducted at pH 8.0. Temperatures of 25 – 30 °C favored long-term product accumulation, and 250  $\mu\text{mol/L}$   $\text{FeSO}_4$  gave the highest apigenin concentration ( $2.674 \pm 0.101$  mmol/L). At low enzyme input,  $A_{380}$  endpoint measurements combined with SDS-PAGE target-band integrated-density correction were used to evaluate the relative initial reaction rates of 29 single-site variants. V159E and H261F reached  $3.06 \pm 0.26$  and  $1.50 \pm 0.04$  times the WT level, respectively, and were significantly higher than WT ( $p < 0.01$ ). In the HPLC-quantified 20 mL reaction validation system with higher enzyme input, I115V, V159E, M230D, and N297D reached  $2.15 \pm 0.15$ ,  $2.14 \pm 0.11$ ,  $2.02 \pm 0.10$ , and  $1.88 \pm 0.13$  relative to WT, respectively, whereas M85L was significantly lower than WT ( $0.35 \pm 0.03$ ,  $p < 0.01$ ). Considering both metrics, V159E is the preferred variant because it showed the highest relative initial reaction rate at low enzyme input, whereas I115V remains an alternative for preparative product formation, with a marginally higher mean in the 20 mL assay and no direct statistical comparison between the two variants. This study establishes an integrated enzyme-engineering route combining lineage-focused sequence collection, ancestral reconstruction, AI prediction, computational mutagenesis, and high-throughput screening, providing a methodological reference for the discovery and engineering of FNS-I and other 2-ODD enzymes.



Lineage-focused discovery, ancestral reconstruction, computational screening, and experimental validation

**Keywords:** flavone synthase I; Apiaceae; ancestral sequence reconstruction; CataPro; site-directed mutagenesis; high-throughput screening

## 1 引言

芹菜素 (Apigenin, 4',5,7-三羟基黄酮) 是黄酮亚类的代表性化合物, 广泛存在于欧芹、芹菜、柑橘及洋甘菊中, 其生物利用度、代谢及抗炎、抗肿瘤等活性受到持续关注<sup>[1-2]</sup>。黄酮类化合物是植物苯丙烷代谢途径的重要分支产物, 以 C6-C3-C6 二苯基色原酮为基本骨架, 迄今已鉴定超过 9000 种天然黄酮<sup>[3]</sup>; 它们还参与植物 UV 防护、抗氧化和微生物互作等过程<sup>[4-5]</sup>。芹菜素相关药理作用已有多项研究报告<sup>[6-11]</sup>。芹菜素的制备主要包括化学合成、植物提取和酶法合成。化学路线往往选择性有限且存在溶剂与污染负担<sup>[12-15]</sup>; 植物提取受原料含量和有机溶剂消耗限制; 酶法因反应条件温和、选择性高而具有替代潜力<sup>[16-17]</sup>。自然界中, FNS-I 直接催化柚皮素 C2-C3 位脱氢生成芹菜素 (图 1 (a)), 属于  $\text{Fe}^{2+}/2$ -酮戊二酸依赖的可溶性 2-ODD 家族<sup>[18]</sup>; FNS-II 则属于膜结合细胞色素 P450 (CYP93B) 家族, 依赖 NADPH 和  $\text{O}_2$ <sup>[19]</sup>。相较于 FNS-II, FNS-I 更便于在原核体系中重组表达和开展酶工程, 因此本研究选择 FNS-I 作为研究对象。

黄酮合酶 I (Flavone Synthase I, FNS-I; EC 1.14.20.6) 属 2-酮戊二酸依赖性双加氧酶 (2-ODD) 超家族, 催化柚皮素 (Naringenin) C2-C3 位脱氢生成芹菜素, 是该途径的关键步骤<sup>[18, 20]</sup>。首个 FNS-I 由 Martens 等人于欧芹 (*Petroselinum crispum*) 中克隆鉴定, 首次确认了该酶催化芹菜素合成的生化功能<sup>[21]</sup>。其后, 同一团队对欧芹中 FNS-I 和 FHT 的底物特异性和进化关系进行了系统比较, 发现 FNS-I 与黄烷酮 3 $\beta$ -羟化酶 (FHT) 在氨基酸序列层面的差异仅约 6%, 提示 FNS-I 可能由 FHT 经旁系同源进化而来, 属 2-ODD 家族内底物偏好性分化的典型案例<sup>[22]</sup>。国内已有研究者对黄酮合酶的异源表达和应用进行了专利布局<sup>[23]</sup>。迄今除伞形科少数物种外, FNS-I 在其他植物类群中报道有限, 已表征的天然 FNS-I 在异源表达系统中的可溶性表达水平和催化效率均不高, 大量同源蛋白的功能仍属未知。

蛋白质结构建模和功能预测工具已被用于扩展酶发现与设计的候选空间。AlphaFold3<sup>[24]</sup> 可对

蛋白质及其相互作用体系进行高精度结构预测, 为催化残基定位和结构比较提供依据。祖先序列重建 (Ancestral Sequence Reconstruction, ASR) 通过最大似然法推断进化树内部节点的氨基酸状态, 曾获得稳定性或可溶性优于部分现代酶的祖先蛋白<sup>[25-26]</sup>。CataPro 是面向酶催化参数预测的深度学习模型<sup>[27]</sup>, 可用于对候选序列进行优先级排序, 但不能替代实验测定。近年来, 基于同源建模、定点突变和机器学习的酶改造策略已得到广泛应用<sup>[28-29]</sup>。

本研究以 PcFNS-I 为出发点, 开展三个层次的工作。(i) 酶挖掘层: 在伞形科内收集 22 条 FNS-I 候选现生序列, 经 ASR 获得 2 条祖先序列, 并以 CataPro 排序结合物种代表性选取 3 条天然同源酶, 连同 PcFNS-I 共 6 个候选酶进入统一活性测定。(ii) 突变设计层: 以最优祖先酶 AncFNS-I-48 为亲本, 全位点虚拟突变扫描, 经 PSSM 保守位点分析与 Pythia 热稳定性过滤后, 由深度学习酶动力学参数预测模型 CataPro 评分选取候选突变体。(iii) 方法与评价层: 建立基于  $A_{380}$  终点法的 96 孔板高通量筛选平台, 比较 6 种纯化 FNS-I 在 50 mmol/L 磷酸盐缓冲液中的 pH 响应, 并在 pH 8.0 条件下优化反应温度和  $\text{FeSO}_4$  添加量。低酶水平测定经表达量校正的相对初始反应速率, 以评价低酶投量条件下的相对反应表现; 在 20 mL 制备级反应中采用 2 mg/mL 粗酶总蛋白浓度 (Bradford 法定量) 进一步验证其实际转化能力。

## 2 材料与方法

### 2.1 实验材料

#### 2.1.1 菌株与质粒

大肠杆菌 (*Escherichia coli*) BL21 (DE3) 感受态细胞用于蛋白表达, DH5 $\alpha$  用于质粒克隆扩增。表达载体 pET-28a (N 端 6 $\times$ His 标签, 卡那霉素抗性, Novagen) 用于 FNS-I 基因的克隆与表达。

#### 2.1.2 试剂与耗材

柚皮素标准品 (纯度  $\geq 98\%$ )、芹菜素标准品 (纯度  $\geq 98\%$ ) 购自 Sigma-Aldrich。底物柚皮素市售产品均为混旋光体, 本文所述底物均指混旋光

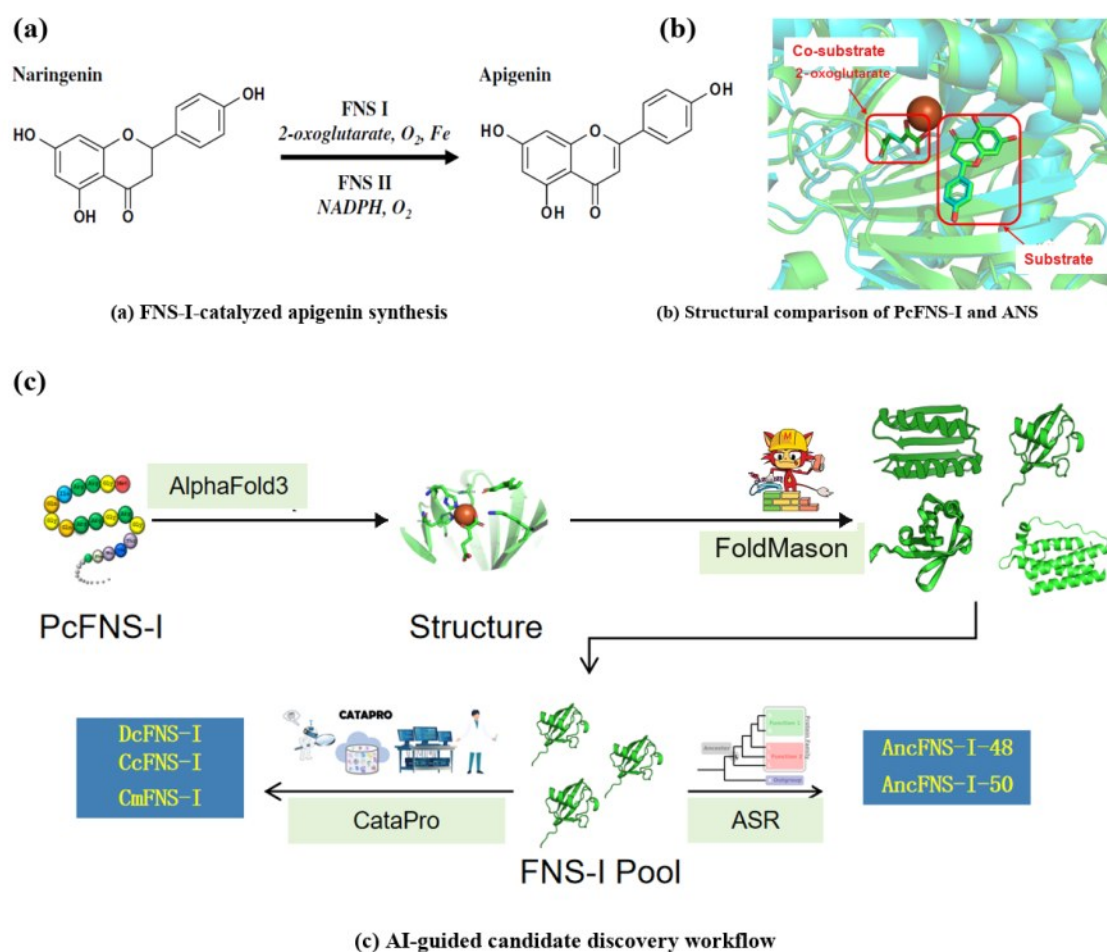


图1 FNS-I酶的挖掘策略与结构基础。

(a) FNS-I催化柚皮素生成芹菜素的反应路线，2-酮戊二酸为共底物， $\text{Fe}^{2+}$ 为辅因子。(b) PcFNS-I预测结构（青色）与拟南芥ANS晶体结构（PDB: 2BRT, 灰色）的比对。PcFNS-I配位残基编号为His218、Asp220和His276；2BRT中对应残基编号为His232、Asp234和His288。柚皮素构象仅作空间位置示意，不代表定量对接结果。(c) 以22条伞形科FNS-I候选现生序列记录为起点，经ASR获得2条祖先序列，并以CataPro排序结合谱系覆盖选取3条天然同源酶；连同PcFNS-I共6个候选酶进入实验验证。

Fig. 1 Discovery strategy and structural context of FNS-I

(a) FNS-I-catalyzed conversion of naringenin to apigenin, with 2-oxoglutarate as cosubstrate and  $\text{Fe}^{2+}$  as cofactor. (b) Structural comparison of predicted PcFNS-I (cyan) and Arabidopsis ANS (PDB 2BRT, gray). PcFNS-I residues are His218, Asp220, and His276; the corresponding 2BRT residues are His232, Asp234, and His288. The displayed naringenin conformation is a spatial illustration, not a quantitative docking result. (c) Starting from 22 extant Apiaceae FNS-I candidate sequence records, ASR yielded two ancestral sequences, while CataPro ranking combined with lineage coverage identified three natural homologs; together with PcFNS-I, six candidates entered experimental validation.

柚皮素。2-酮戊二酸（2-OG）、抗坏血酸（VC）、硫酸亚铁、磷酸盐及其他缓冲盐均为国产分析纯，购自Aladdin。限制性内切酶NdeI/XhoI、T4 DNA连接酶购自Thermo Fisher Scientific。甲醇（色谱纯）、甲酸（色谱纯）购自Merck。溶菌酶购自Aladdin。

### 2.1.3 仪器与设备

高效液相色谱系统（岛津LC-2010A，配紫外检测器），色谱柱为Hypersil BDS C18反相柱

（250 mm × 4.6 mm，5  $\mu\text{m}$ ，Thermo Fisher）。酶标仪（SpectraMax iD3），用于96孔板吸光度检测。恒温振荡培养箱、超声破碎仪、高速冷冻离心机等为实验室常规设备。

## 2.2 FNS-I候选酶的计算挖掘

### 2.2.1 AlphaFold3结构建模

以欧芹PcFNS-I（UniProt: Q7XZQ8）氨基酸

序列为查询序列, 利用 AlphaFold3<sup>[24]</sup> 进行三维结构预测, 并以拟南芥花青素合酶 ANS 晶体结构 (PDB ID: 2BRT) 作为结构参照。活性中心残基统一按 PcFNS-I 序列编号为 His218、Asp220、His276 和 Lys197; 2BRT 中对应  $\text{Fe}^{2+}$  配位残基为 His232、Asp234 和 His288。AncFNS-I-48 按其自身序列编号时, 对应残基为 His217、Asp219、His275 和 Lys196, 后文涉及 AncFNS-I-48 结构时均采用该编号。图 1(b) 中的柚皮素构象仅用于显示底物与催化口袋的相对空间位置, 不进行对接打分、结合能比较或定量机理推断。

### 2.2.2 祖先序列重建 (ASR)

祖先序列重建假设现生蛋白序列由一棵系统发育树描述, 各分支上的氨基酸替换服从给定替换模型。在已知叶节点 (现生序列) 的比对序列和树拓扑结构的条件下, 通过最大似然法逐位点估计内部节点 (祖先) 的氨基酸状态。本研究采用 PAML 软件包中的 `codeml` 程序<sup>[30]</sup> 完成重建。ASR 流程分为四步: (1) 以附录 A 表 1 所列 22 条伞形科 (*Apiaceae*) FNS-I 现生序列作为叶节点集合; (2) 采用 FoldMason<sup>[31]</sup> 进行结构感知的多序列比对, 同步生成氨基酸序列比对文件 (`foldmason_aa.fa`) 和 Newick 格式进化树 (`foldmason.nw`); (3) 将比对和树文件转换为 PAML `phylip` 格式输入; (4) 运行 `codeml` 进行祖先状态推断。

`codeml` 关键运行参数为 `seqtype=2` (氨基酸序列)、`model=2` (Jones 经验替换模型)、`aaRatefile=jones.dat`、`RateAncestor=1` (重建祖先状态)、`runmode=0` (使用用户提供的进化树)、`clock=0` (不强制分子钟假设) 和 `cleandata=1` (去除含模糊字符的位点)。输出文件 `rst` 包含各内部节点的边际重建和联合重建序列, 并标注各位点最可能氨基酸及后验概率。提取祖先序列后, 去除 `gap` 和模糊字符得到完整氨基酸序列。`codeml` 对进化树上全部内部节点给出祖先状态估计; 考虑到湿实验难以覆盖全部节点, 本研究按拓扑位置选取紧邻现生叶节点的浅层内部节点及其直接父节点。`rst` 输出中的 Node 48 与 Node 50 构成该叶侧节点-父节点对, 分别命名为 AncFNS-I-48 和 AncFNS-I-50 并进入异源表达与酶活测定。

为展示附录 A 表 1 所列 22 条现生序列记录之间的关系, 按完整氨基酸序列合并完全相同的记录, 得到 17 条唯一序列; 采用 MUSCLE 进行多序列比对, 以氨基酸 `p-distance` (成对比较时忽略 `gap` 位点) 构建 Neighbor-Joining (NJ) 无根拓扑, 并通过 200 次比对列重采样计算 `bootstrap` 支持度 (图 5)。图 5 中相同序列的 `accession` 以斜线合并标在同一叶节点。该 NJ 树仅用于描述现生序列之间的关系; ASR 采用上述 22 条叶节点、FoldMason 结构感知比对及其 Newick 树, 图 5 不参与 Node 48 和 Node 50 的选择。与候选选择直接相关的局部拓扑见附录 E 图 S3。

### 2.2.3 CataPro 活性预测与天然酶选取

CataPro<sup>[27]</sup> 是基于深度学习的酶催化动力学参数预测模型, 接受氨基酸序列和底物 SMILES 输入, 预测 `kcat`、`Km` 及催化效率 `kcat/Km`。以柚皮素为底物, 对上述 22 条序列记录进行预测, 结果按 `kcat/Km` 降序汇总于附录 A 表 1。评分最高的毒参 B3RFV8 与 ABG78795 的预测 `kcat/Km` 均为  $9.400 \text{ L} \cdot \text{mmol}^{-1} \cdot \text{s}^{-1}$ , 约为 PcFNS-I 的 1.05 倍; 同种序列 Q4QUR7 的预测值为  $9.342 \text{ L} \cdot \text{mmol}^{-1} \cdot \text{s}^{-1}$ 。为减少同种重复, 保留 B3RFV8 并命名为 CmFNS-I。欧芹 Q7XZQ8 作为文献对照 PcFNS-I, 不再重复计入天然酶优选; 其余物种结合谱系覆盖选取孜然 A6XFT1 (CcFNS-I) 和胡萝卜 Q4QUR9 (DcFNS-I)。据此, 3 条天然同源酶、文献对照 PcFNS-I 及 2 条祖先序列共构成 6 个候选酶, 进入统一实验验证。CataPro 输出仅用于候选优先级排序, 不作为催化性能的实验证据; 该模型随后还用于 AncFNS-I-48 突变体的评分 (见 2.5)。

### 2.2.4 Chai-1 复合物结构预测与比较

采用 Chai-1<sup>[32]</sup> 对 AncFNS-I-48 及 I115V、V159E、M230D、N297D、L186F 和 H261F 六个单点突变体与柚皮素的复合物进行结构预测。输入由对应蛋白质氨基酸序列和柚皮素分子结构组成; 将各突变体预测复合物与 WT 模型按蛋白质主链叠合, 比较突变位点、底物及邻近残基的空间关系。模型中未加入  $\text{Fe}^{2+}$  和 2-OG, 不进行结合能、催化能垒或动力学参数计算; 所得结构仅用于可视化和提出可检验的机制假说, 不作为催化机制的直接证据。

## 2.3 候选酶的异源表达与酶活测定

### 2.3.1 基因合成与克隆

将经计算筛选确定的6个候选酶编码基因(PcFNS-I、DcFNS-I、CcFNS-I、CmFNS-I、AncFNS-I-48和AncFNS-I-50)进行全基因合成,按大肠杆菌密码子偏好性优化,通过NdeI/XhoI位点克隆至pET-28a载体(N端6×His标签)。转化大肠杆菌BL21(DE3)感受态细胞,涂布于含50 μg/mL卡那霉素的LB平板上,37℃培养过夜。

### 2.3.2 蛋白表达

挑取单克隆接种至LB液体培养基(含50 μg/mL卡那霉素),37℃、220 r/min培养至 $OD_{600} \approx 0.6 - 0.8$ ,加入异丙基-β-D-硫代半乳糖苷(IPTG)至终浓度0.5 mmol/L,25℃诱导16 h。4℃、12000 × g离心10 min收集菌体。

### 2.3.3 粗酶液与纯化酶制备

菌体以PBS缓冲液(pH 7.4)洗涤后重悬,冰浴超声破碎(功率300 W,超声3 s/间歇5 s,总时长10 min),4℃、12000 × g离心15 min,取上清作为粗酶液。粗酶液和纯化酶均采用Bradford法定量。6个候选酶利用N端6×His标签进行亲和纯化,收集纯化组分并以SDS-PAGE核对目标条带;每种酶的3次平行作为独立培养所得的生物学重复。

### 2.3.4 酶活测定反应体系与HPLC检测

酶活测定反应体系(200 μL):底物柚皮素3 mmol/L,50 mmol/L磷酸盐缓冲液(pH 8.0),2-酮戊二酸4 mmol/L,抗坏血酸4 mmol/L,FeSO<sub>4</sub> 250 μmol/L及终浓度200 μg/mL的纯酶,30℃、200 r/min反应1 h。每种酶设置3次独立培养的生物学平行,反应结束后取适量反应液加入甲醇沉淀蛋白,涡旋混匀,12000 r/min离心1 min,取上清经0.22 μm微孔滤膜过滤后进样。

HPLC检测条件:C18反相色谱柱(250 mm × 4.6 mm, 5 μm),检测波长290 nm;柚皮素和芹菜素均在该波长采集色谱信号,并依据保留时间区分,芹菜素采用其标准品外标曲线定量。流动相为甲醇-0.5%甲酸水溶液梯度洗脱(0 min: 20% 甲醇 → 20 min: 60% 甲醇 → 25 min: 20% 甲醇),流速1.0 mL/min,柱温30℃,进样量10 μL。芹菜素标准曲线为 $y = 25.46x + 16.01$  ( $r = 0.9989$ )。

### 2.3.5 SDS-PAGE蛋白表达与纯化分析

取各候选酶超声破碎上清和纯化组分进行12.5% SDS-PAGE,考马斯亮蓝染色后观察目标条带。图2(a)泳道依次为:M,蛋白分子量标准;1/2, PcFNS-I上清/纯化;3/4, DcFNS-I上清/纯化;5/6, CmFNS-I上清/纯化;7/8, CcFNS-I上清/纯化;9/10, AncFNS-I-48上清/纯化;11/12, AncFNS-I-50上清/纯化。

## 2.4 高通量筛选方法的建立

### 2.4.1 检测原理

对底物柚皮素(Nar)和产物芹菜素(Api)进行紫外全波长扫描。以380 nm为检测波长时,Api在0.2 - 1.5 mmol/L范围内的吸光度随浓度增加而升高,而Nar在相同浓度下保持较低响应并随浓度逐步增加;Api在3.0 - 3.5 mmol/L出现平台。该结果表明380 nm能够区分Nar与Api,并提示Api定量应采用其线性工作区间。

### 2.4.2 终点法 $A_{380}$ 检测方案

本方法以终点法为主。反应进行1 h后,按2.3.4所述方法加入甲醇终止反应并沉淀蛋白,涡旋混匀、离心后取上清,必要时稀释后测定 $A_{380}$ 吸光值。以 $\Delta A_{380}$ (=样品 $A_{380}$  - 空白对照 $A_{380}$ )表征相对酶活性。基于附录E图S1,Api在0.2 - 1.5 mmol/L范围内与380 nm吸光度呈线性关系,校准方程为 $A_{380} = 1.8273C - 0.1597$ (C为Api浓度,mmol/L), $R^2 = 0.997$ ;因此该范围可作为Api的定量工作区间,超出范围的样品应稀释至线性区间后测定。附录E图S2进一步显示, $A_{380}$ 响应与粗酶投量(V/V)在2% - 20%范围内近似线性,拟合方程为 $A_{380} = 0.0867x + 0.3856$ , $R^2 = 0.9999$ ;30%点偏离该线性范围,x为酶投量百分数(V/V)。检测体系优化确定底物浓度3 mmol/L(约1 g/L)、粗酶量10% (V/V)、反应液稀释4倍为最佳检测条件,此时各酶量梯度间区分度显著,具备良好的定量检测能力。

## 2.5 AI计算驱动的定点突变设计

### 2.5.1 全位点虚拟突变扫描

以AncFNS-I-48为亲本,基于AlphaFold3预测

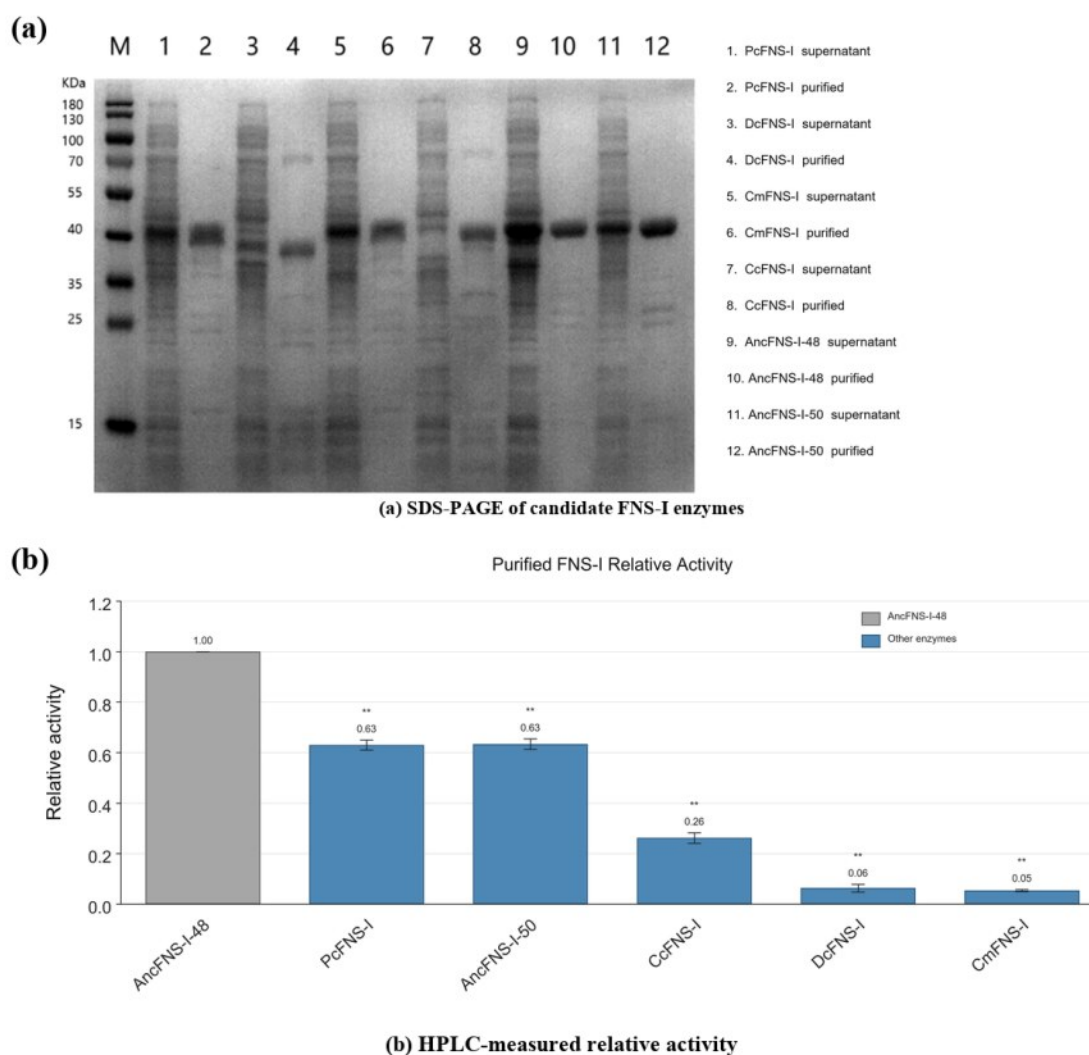


图2 候选FNS-I酶的表达、纯化与相对活性（HPLC检测）

(a) 6个候选酶的SDS-PAGE。M: 蛋白分子量标准; 1/2: PcFNS-I上清/纯化; 3/4: DcFNS-I上清/纯化; 5/6: CmFNS-I上清/纯化; 7/8: CcFNS-I上清/纯化; 9/10: AncFNS-I-48上清/纯化; 11/12: AncFNS-I-50上清/纯化。(b) 等量纯化蛋白在2.3.4所述pH 8.0基准反应体系(含250  $\mu\text{mol/L}$   $\text{FeSO}_4$ )中的相对活性。每个独立培养重复分别以同批AncFNS-I-48测定值归一化为1.00, 再由3个归一化比值计算均值 $\pm$ 标准差( $n=3$ ); 双侧单样本t检验与参考水平1.00比较, \*\*  $p < 0.01$ 。图2(b)用于固定pH 8.0条件下的候选酶横向比较, 图3(a)为另行设置的pH梯度实验。

Fig. 2 Expression, purification, and relative activity of candidate FNS-I enzymes

(a) SDS-PAGE of supernatant and purified fractions: M, protein marker; lanes 1/2, PcFNS-I; 3/4, DcFNS-I; 5/6, CmFNS-I; 7/8, CcFNS-I; 9/10, AncFNS-I-48; and 11/12, AncFNS-I-50. (b) HPLC-measured relative activities with equal amounts of purified protein in the pH 8.0 reference reaction described in Section 2.3.4, containing 250  $\mu\text{mol/L}$   $\text{FeSO}_4$ . Each independent-culture replicate was normalized to the AncFNS-I-48 value measured in the same batch, and means  $\pm$  SD were calculated from the three normalized ratios ( $n=3$ ). Two-sided one-sample t-test versus 1.00, \*\*  $p < 0.01$ . Panel (b) compares candidates under fixed pH 8.0 conditions, whereas Figure 3(a) reports the separately established pH-gradient experiment.

结构, 利用 PSSM (基于 PSI-BLAST 算法<sup>[33]</sup>) 进行保守性筛选, 并以 Pythia<sup>[34]</sup> 预测突变对稳定性的影响, 对各位点的19种非野生型替换进行虚拟扫描。Pythia与PSSM输出仅作为候选排序变量; Pythia  $\Delta\Delta G$  与 CataPro 原始预测分数列于附录C

表2, 其中负  $\Delta\Delta G$  表示预测稳定化。上述计算结果用于候选排序, 不替代实测活性或稳定性证据, 也不据此作确定性稳定机制判断。

## 2.5.2 两级排序与过滤

排除  $\text{Fe}^{2+}$  配位核心区域残基后, 将通过 PSSM/

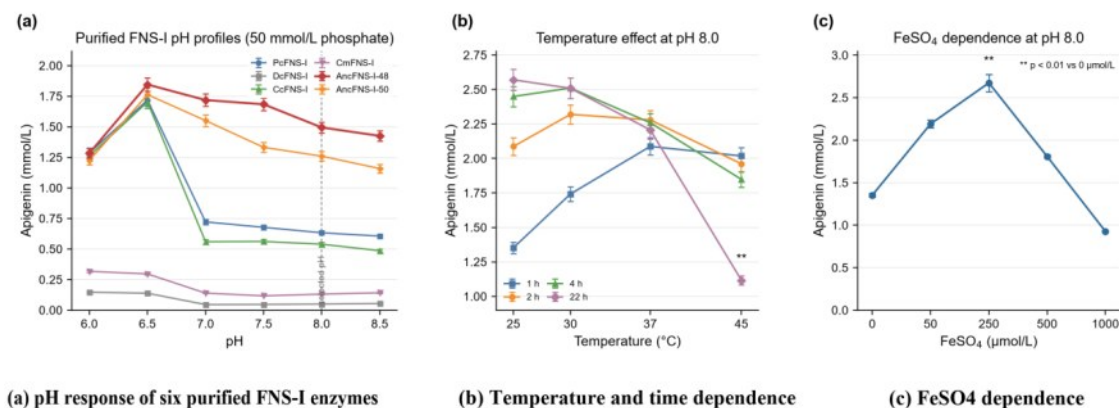


图3 候选FNS-I的pH响应及pH 8.0条件下AncFNS-I-48的反应条件分析

(a) 6种等量纯化FNS-I在50 mmol/L磷酸盐缓冲液、pH 6.0~8.5条件下反应1 h后的芹菜素生成量 (HPLC定量); 虚线表示后续实验采用的pH 8.0。(b) pH 8.0、含250 μmol/L FeSO<sub>4</sub>时, 25、30、37和45 °C条件下1、2、4和22 h的芹菜素浓度。(c) pH 8.0条件下, 0、50、250、500和1000 μmol/L FeSO<sub>4</sub>反应1 h后的芹菜素浓度。数据以均值±标准差表示 (n=3)。(b) 22 h时45 °C与30 °C采用双侧独立样本t检验; (c) FeSO<sub>4</sub>浓度效应应采用单因素方差分析, 250 μmol/L与0 μmol/L采用预设双侧独立样本t检验; \*\* p < 0.01。

Fig. 3 pH responses of candidate FNS-I enzymes and reaction-condition profiling of AncFNS-I-48 at pH 8.0

(a) HPLC-quantified apigenin production by equal amounts of six purified FNS-I enzymes after 1 h from pH 6.0 to 8.5 in 50 mmol/L phosphate buffer; the dashed line marks pH 8.0 used in subsequent experiments. (b) Apigenin concentrations after 1, 2, 4, and 22 h at 25, 30, 37, or 45 °C at pH 8.0 with 250 μmol/L FeSO<sub>4</sub>. (c) Apigenin concentrations after 1 h with 0, 50, 250, 500, or 1000 μmol/L FeSO<sub>4</sub> at pH 8.0. Data are means ± SD (n = 3). For panel (b), 45 versus 30 °C at 22 h was assessed by a two-sided independent-samples t-test. For panel (c), the FeSO<sub>4</sub> concentration effect was assessed by one-way ANOVA, with a prespecified two-sided independent-samples t-test comparing 250 versus 0 μmol/L FeSO<sub>4</sub>; \*\* p < 0.01.

Pythia初筛的候选输入CataPro进行活性排序, 选取综合得分靠前者29个单点突变体进入实验。计算评分用于优先级排序, 最终判断以重复实验结果为准。

### 2.5.3 高通量筛选流程

(1) 菌种规模化培养与诱导表达: 突变载体构建完成后挑取单克隆, 经96孔浅孔板 (每孔200 μL LB, 50 μg/mL卡那霉素, 37 °C、800 r/min培养至OD<sub>600</sub> ≈ 0.6 - 0.8)→96孔深孔板 (IPTG 0.5 mmol/L, 25 °C诱导16 h) 逐级放大。菌液均一化处理确保各反应单元起始生物量一致。

(2) 粗酶液制备: 菌体经溶菌酶 (1 mg/mL) 液消化裂解, 离心取上清为粗酶液。

(3) 反应与检测: 粗酶液10% (V/V), 底物柚皮素3 mmol/L, 50 mmol/L反应缓冲液 (pH 8.0), 2-酮戊二酸4 mmol/L, 抗坏血酸4 mmol/L, FeSO<sub>4</sub> 250 μmol/L, 30 °C、200 r/min反应1 h。随后按2.3.4所述方法加入甲醇终止反应并沉淀蛋白, 涡旋混匀、离心后取上清, 稀释后测定A<sub>380</sub>。每板设野生型AncFNS-I-48、空载体和无酶空白对照。为测定初始反应速率, 在反应早期线性区间取样,

采用A<sub>380</sub>终点法检测, 并以单位时间内芹菜素生成量计算初始反应速率; 每个独立培养重复分别以同批WT测定值归一化, 每个突变体设置3次独立培养的生物学平行。同时对WT及突变体粗酶液进行12.5% SDS-PAGE, 对目标条带的净积分灰度 (IOD) 进行定量, 以WT IOD/样品IOD作为校正系数。三次平行相对初始反应速率均为完成该灰度校正和同批WT归一化后的最终值, 并据此计算均值和标准差。

### 2.5.4 摇瓶验证与HPLC定量

挑取初筛的9个阳性H261F、L186F、A169K、H261A、N297D、I115V、M230D、V159E、T204P, 2个阴性Q71Y、M85L和野生型AncFNS-I-48进行摇瓶验证。甘油菌扩培后接种至50 mL LB培养基, 培养至OD<sub>600</sub> ≈ 0.6 - 0.8, IPTG 0.2 mmol/L、25 °C、180 r/min诱导16 h。PBS洗涤并超声破碎制备粗酶液。20 mL反应体系: 粗酶终浓度2 mg/mL (Bradford法定量), 柚皮素15 mmol/L、50 mmol/L磷酸盐缓冲液 (pH 8.0)、2-酮戊二酸20 mmol/L、抗坏血酸20 mmol/L和FeSO<sub>4</sub> 250 μmol/L, 30 °C、200 r/min反应2 h。该20 mL

体系采用HPLC定量芹菜素，用于评价较高粗酶总蛋白投量（2 mg/mL，Bradford法定量）下的试剂反应相对活性，而不用计算初始反应速率。每种酶设置3次独立培养的生物学平行，每个独立培养重复分别以同批WT测定值归一化为1.0，再由3个归一化比值计算均值和标准差。显著性采用双侧单样本t检验与归一化WT水平1.0比较，阈值设为 $p < 0.01$ 。

## 2.6 催化反应条件优化

以等量纯酶比较PcFNS-I、DcFNS-I、CcFNS-I、CmFNS-I、AncFNS-I-48和AncFNS-I-50在不同pH条件下的产物生成能力；随后以AncFNS-I-48为对象，在pH 8.0条件下考察反应温度、反应时间和FeSO<sub>4</sub>添加量。除FeSO<sub>4</sub>梯度实验外，其余反应体系统一补加250  $\mu\text{mol/L}$  FeSO<sub>4</sub>。每个条件设置3次独立培养的生物学平行，结果以均值  $\pm$  标准差表示。温度分析中比较22 h时45  $^{\circ}\text{C}$ 与30  $^{\circ}\text{C}$ ，FeSO<sub>4</sub>梯度采用单因素方差分析并预设250  $\mu\text{mol/L}$ 与0  $\mu\text{mol/L}$ 作双侧独立样本t检验，显著性阈值均为 $p < 0.01$ 。

纯酶pH响应：在50 mmol/L磷酸盐缓冲液中设置pH 6.0、6.5、7.0、7.5、8.0和8.5，分别加入

等量PcFNS-I、DcFNS-I、CcFNS-I、CmFNS-I、AncFNS-I-48或AncFNS-I-50纯化蛋白；柚皮素、2-酮戊二酸、抗坏血酸和FeSO<sub>4</sub>终浓度保持一致，30  $^{\circ}\text{C}$ 、200 r/min反应1 h后采用HPLC定量芹菜素，用于比较各候选酶的pH适应范围。AncFNS-I-48的pH适应性相关电荷位点空间分布见图6。

温度优化：统一控制pH为8.0，设置25、30、37和45  $^{\circ}\text{C}$ ，FeSO<sub>4</sub>维持250  $\mu\text{mol/L}$ ，其余反应组分保持一致，并于1、2、4和22 h取样。FeSO<sub>4</sub>依赖性分析：在pH 8.0及其他统一反应条件下设置0、50、250、500和1000  $\mu\text{mol/L}$  FeSO<sub>4</sub>，反应1 h后采用HPLC定量芹菜素，比较不同Fe<sup>2+</sup>水平对产物生成的影响。

## 2.7 统计学分析

除另有说明外，所有实验均采用3次独立培养所得的生物学重复，结果以均值  $\pm$  标准差表示。图2(b)中，每个独立培养重复分别以同批AncFNS-I-48测定值归一化；图4(c)、(d)中，每个独立培养重复分别以同批WT测定值归一化。归一化后获得3个独立比值，并采用双侧单样本t检验与参考水平1.0比较。图3(b)在22 h时采用双侧独立样本t检验比较45  $^{\circ}\text{C}$ 与30  $^{\circ}\text{C}$ ，图3(c)采用单因

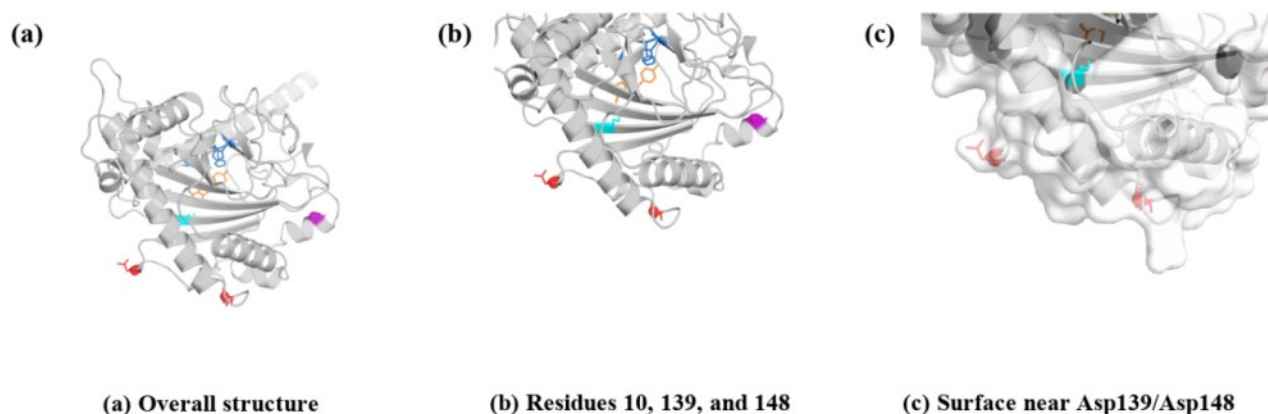


图6 AncFNS-I-48/50差异位点及其预测电荷环境

(a) 整体结构。红色：Asp139/Asp148（局部净电荷-2）；紫色：Ala10（电荷不变）；橙色：柚皮素；天蓝色：H217/D219/H275；青色：K196。(b) 第10/139/148位放大。(c) Asp139/Asp148附近分子表面。上述空间解释为结构假说，不代表已经验证的pH适应机制。

Fig. 6 Differing sites between AncFNS-I-48 and AncFNS-I-50 and their predicted charge environment

(a) Overall structure. Asp139/Asp148 are shown in red and form a locally acidic surface with a net charge of -2; Ala10 is shown in purple and is charge-neutral; naringenin is orange; H217/D219/H275 are sky blue; and K196 is cyan. (b) Enlarged view of residues 10, 139, and 148. (c) Molecular surface near Asp139/Asp148. The spatial interpretation is a structural hypothesis and does not establish a validated pH-adaptation mechanism.

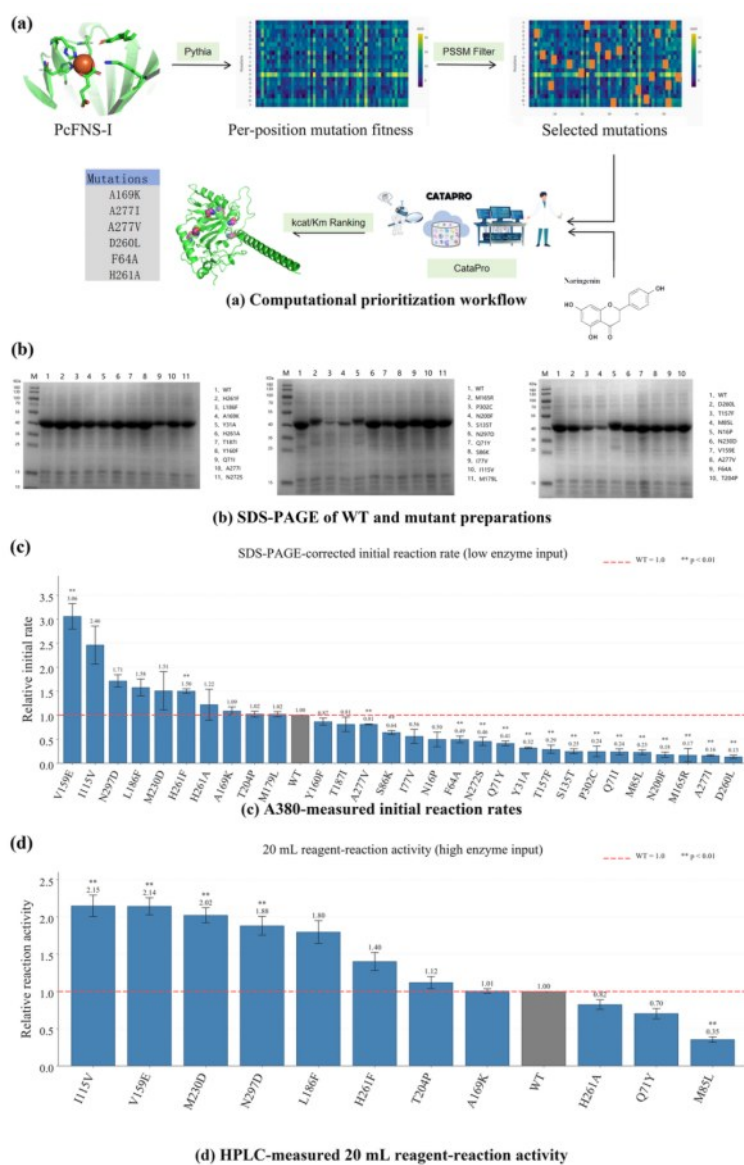


图4 AncFNS-I-48突变体的计算筛选、SDS-PAGE与相对活性

(a) 全位点虚拟扫描、PSSM/Pythia初筛、CataPro排序及实验验证流程。(b) WT及29个突变体粗酶液的SDS-PAGE，三块胶平行排列；每块胶泳道1均为WT，其余泳道的突变体名称见各胶右侧标识。(c) 低酶投量下测得并经过目标条带IOD校正的采用 $A_{380}$ 检测的相对初始反应速率。(d) 所选突变体在20 mL试剂反应体系中以相同且较高的粗酶总蛋白投量（终浓度2 mg/mL，Bradford法定量）测得的采用HPLC检测的相对反应活性。(c)和(d)中，每个独立培养重复分别以同批WT测定值归一化为1.0，再由3个归一化比值计算均值±标准差（ $n=3$ ）；两组的粗酶总蛋白投量及评价指标不同，不作数值上的直接横向比较。红色虚线表示WT水平且绘制于柱形和误差棒上层；双侧单样本t检验：\*\* $p < 0.01$ ；未标记者 $p \geq 0.01$ 。

Fig. 4 Computational prioritization, SDS-PAGE, and relative activity of AncFNS-I-48 mutants

(a) Computational workflow. (b) Three horizontally aligned SDS-PAGE gels of WT and 29 crude-enzyme mutant preparations; lane identities are shown beside each gel. (c)  $A_{380}$ -measured relative initial reaction rates at low enzyme input, corrected using the corresponding target-band IOD. (d) HPLC-measured relative reagent-reaction activity of selected mutants in a 20 mL system using equal crude-extract total-protein input (2 mg/mL, quantified by the Bradford assay). For panels (c) and (d), each independent-culture replicate was normalized to the WT measured in the same batch, and means  $\pm$  SD were calculated from the three normalized ratios ( $n=3$ ). The panels use different crude-extract total-protein inputs and response metrics and are not intended for direct quantitative comparison. Red dashed line, WT; two-sided one-sample t-test, \*\* $p < 0.01$ ; unmarked,  $p \geq 0.01$ .

素方差分析评价 $\text{FeSO}_4$ 浓度效应，并按预设比较采用双侧独立样本t检验比较250  $\mu\text{mol/L}$ 与0  $\mu\text{mol/L}$ 。

所有检验均为双侧检验，显著性阈值设为 $p < 0.01$ ；未标注者表示 $p \geq 0.01$ 。

### 3 结果与讨论

#### 3.1 FNS-I同源酶的计算挖掘与实验验证

以欧芹 PcFNS-I (UniProt: Q7XZQ8) 为结构参照, 并按“伞形科 FNS-I 序列收集-ASR/CataPro 双路径筛选-实验验证”开展候选酶挖掘 (图 1(c))。AlphaFold3 预测结构呈典型 2-ODD 超家族双链  $\beta$ -螺旋折叠。与拟南芥 ANS (PDB: 2BRT) 的结构比对 (图 1(b)) 显示, PcFNS-I 的 His218、Asp220 和 His276 构成  $\text{Fe}^{2+}$  配位环境, Lys197 位于活性中心附近; 2BRT 中对应配位残基编号为 His232、Asp234 和 His288。图中的柚皮素仅作空间位置示意, 未据此开展对接打分或提出定量结合机制。

在伞形科内收集 22 条 FNS-I 候选现生序列, 并将其作为 FoldMason/PAML 祖先序列重建的叶节点集合; 按拓扑位置选取 AncFNS-I-48 和 AncFNS-I-50 两个相邻内部节点。CataPro 对全部 22 条序列记录的预测结果见附录 A 表 1; 结合预测 kcat/Km 与谱系覆盖, 选取 DcFNS-I、CcFNS-I 和 CmFNS-I, 连同 PcFNS-I 及两条祖先序列共构成 6 个候选酶。SDS-PAGE 结果 (图 2(a)) 显示, 6 种酶的纯化泳道在预期分子量附近均可见目标条带, 表明相对活性测定所用样品均经过纯化。该胶图用于定性确认, 不据其条带灰度换算酶量。

基于附录 A 表 1 所列 22 条伞形科 FNS-I 候选现生序列记录, 合并完全相同的氨基酸序列后得到 17 条唯一序列; 采用 MUSCLE 多序列比对和氨基酸 p-distance 构建的 NJ 无根拓扑见图 5。重复组的 accession 以斜线合并标在同一叶节点。图中采用矩形水平布局, 内部节点旁标注 200 次比对列重采样获得的 bootstrap 支持度; 枝长与距离标尺不显示, 矩形布局不表示进化根或祖先方向。

ASR 使用附录 A 表 1 所列 22 条现生序列作为叶节点集合, 经 FoldMason 结构感知比对及其 Newick 树输入 PAML codeml; 在该树的拓扑中, AncFNS-I-48 与 AncFNS-I-50 分别代表叶侧内部节点及其直接父节点。两者的比较用于说明祖先节点采样策略, 不将相邻节点间的序列差异单独解释为活性因果位点。图 5 为合并完全相同序列后的

17 叶 NJ 展示树, 不参与 Node 48 和 Node 50 的选择。该局局部拓扑见附录 E 图 S3; 其中 Node 48 位于 Node 50 下游, 二者编号均来自 codeml rst 输出。

6 个候选纯酶的 HPLC 相对活性见图 2(b)。每个独立培养重复分别以同批 AncFNS-I-48 测定值归一化。以 AncFNS-I-48 归一化为  $1.000 \pm 0.000$ , PcFNS-I、AncFNS-I-50、CcFNS-I、DcFNS-I 和 CmFNS-I 分别为  $0.630 \pm 0.020$ 、 $0.634 \pm 0.021$ 、 $0.262 \pm 0.021$ 、 $0.063 \pm 0.015$  和  $0.054 \pm 0.004$  ( $n = 3$ )。在双侧单样本 t 检验中, 上述 5 种酶均与归一化 AncFNS-I-48 水平存在显著差异 ( $p < 0.01$ )。AncFNS-I-48 在该反应体系中具有最高相对活性。

#### 3.2 催化反应条件优化

首先比较 6 种纯化 FNS-I 在 50 mmol/L 磷酸盐缓冲液中的 pH 响应, 再以 AncFNS-I-48 为对象, 在 pH 8.0 下评价温度、反应时间和  $\text{FeSO}_4$  添加量对芹菜素积累的影响 (图 3); 除  $\text{FeSO}_4$  梯度外, 各反应体系统一补加 250  $\mu\text{mol/L}$   $\text{FeSO}_4$ 。

纯酶 pH 响应 (图 3(a), 反应 1 h) 显示, AncFNS-I-48 在 pH 6.0、6.5、7.0、7.5、8.0 和 8.5 时的芹菜素浓度 (HPLC 定量) 分别为 1.284、1.845、1.718、1.684、1.494 和 1.425 mmol/L, 在整个测试范围内均保持较高产物生成能力。AncFNS-I-50 在 pH 8.0 和 8.5 时分别为 1.260 和 1.158 mmol/L; PcFNS-I 和 CcFNS-I 在 pH 6.5 达到较高水平, 但在 pH 7.0 以上明显下降。相比之下, AncFNS-I-48 兼具较高活性和更宽的 pH 适应范围。柚皮素为弱酸性多酚, 在偏碱性条件下表观溶解度通常提高; 因此选择 pH 8.0 可在保持 AncFNS-I-48 较高产物生成能力的同时改善底物分散和投料条件, 更符合高底物负荷及工业放大的工艺需求。

AncFNS-I-48 与 AncFNS-I-50 仅在 A10S、D139S 和 D148K 三个位点不同。AncFNS-I-48 在 pH 6.5 时产物生成量最高, 但在 pH 6.0 - 8.5 范围内均保持较高水平。综合考虑柚皮素在偏碱性条件下的溶解性和后续工艺评价需要, 后续反应条件选择 pH 8.0, 而非将其表述为酶活峰值。将这一 pH 响应与 AncFNS-I-48/50 的三处序列差异对照, 可提出局部静电环境假说 (图 6)。AncFNS-I-48 在

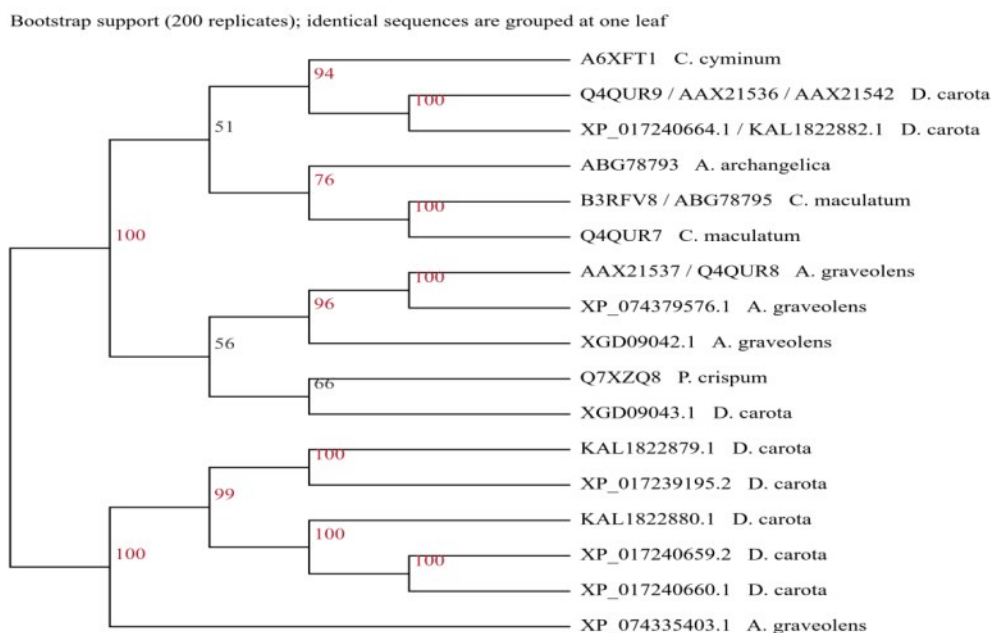


**图5** 代表附录A表1所列22条伞形科FNS-I候选现生序列记录的17条唯一序列无根NJ拓扑。完全相同的氨基酸序列合并为一个叶节点，标签以斜线并列其accession。采用MUSCLE进行多序列比对，以氨基酸p-distance（成对比较时忽略gap位点）构建Neighbor-Joining拓扑；内部节点旁仅标注200次比对列重采样获得的bootstrap支持度数值。枝长和距离标尺不显示，矩形水平布局不表示进化根或祖先方向。图5仅用于描述现生序列之间的关系；ASR以22条序列作为叶节点，使用FoldMason结构感知比对及其Newick树，并由PAML codeml完成祖先状态推断；本图不参与Node 48和Node 50的选择。其局部父子关系见附录E图S3。

**Fig. 5** Unrooted Neighbor-Joining topology of 17 unique amino-acid sequences representing the 22 extant Apiaceae FNS-I candidate sequence records listed in Appendix A Table 1. Records encoding identical sequences were collapsed to one leaf, with their accessions separated by slashes in the leaf label. MUSCLE was used for multiple-sequence alignment, and pairwise amino-acid p-distances (gapped positions excluded pairwise) were used to infer the NJ topology. Internal labels report bootstrap support values from 200 alignment-column resampling replicates. Branch lengths and a distance scale are not shown, and the rectangular horizontal layout does not imply an evolutionary root or ancestral direction. This tree only describes relationships among extant sequences. ASR used the 22 sequence records as leaves, the FoldMason structure-aware alignment and its Newick tree, followed by ancestral-state inference with PAML codeml; this figure was not used to select Node 48 or Node 50. Their local parent-descendant relationship is shown in Appendix E Figure S3.

第139、148位保留两个Asp，局部净电荷为-2；AncFNS-I-50对应为Ser139/Lys148，局部净电荷为+1（相对AncFNS-I-48净增+3）。其中D148K发生电荷翻转（Asp -1 → Lys +1， $\Delta q = +2$ ），D139S失去一个负电荷（ $\Delta q = +1$ ），A10S位于N端且电荷不变。这些差异可能改变局部静电势、水化状态或构象稳定性，进而影响不同pH条件下的反应表现；现有模型不足以确定具体作用途径，仍需K148D等回复突变及电势分析验证。

在pH 8.0条件下，温度结果（图3(b)，HPLC定量）呈时间依赖性。2 h时25、30、37和45 °C分别产生2.087、2.320、2.280和1.961 mmol/L芹

菜素；22 h时分别为2.571、2.509、2.208和1.116 mmol/L。在250  $\mu\text{mol/L}$   $\text{FeSO}_4$ 条件下，22 h时45 °C显著低于30 °C（ $p < 0.01$ ）。25-30 °C更有利于长时间维持产物积累。

在pH 8.0、反应1 h条件下， $\text{FeSO}_4$ 添加量结果（图3(c)）呈明显浓度依赖性。0、50、250、500和1000  $\mu\text{mol/L}$ 时的芹菜素浓度分别为 $1.353 \pm 0.024$ 、 $2.192 \pm 0.047$ 、 $2.674 \pm 0.101$ 、 $1.809 \pm 0.016$ 和 $0.926 \pm 0.007$  mmol/L（ $n = 3$ ）。单因素方差分析显示 $\text{FeSO}_4$ 浓度效应显著（ $p < 0.01$ ）；其中250  $\mu\text{mol/L}$ 组最高，且显著高于0  $\mu\text{mol/L}$ 组（ $p < 0.01$ ）。继续升高至500和1000  $\mu\text{mol/L}$ 后产物生成量下降，因

此 250  $\mu\text{mol/L}$  用于除  $\text{FeSO}_4$  梯度实验外的其余反应体系。

### 3.3 高通量筛选方法的建立

基于底物与产物紫外吸收光谱差异,建立了  $A_{380}$  终点法高通量筛选平台。附录E图S1显示,在 380 nm 处 Api 的吸收明显高于相同浓度的 Nar; Api 在 0.2 - 1.5 mmol/L 范围内呈线性响应,校准方程为  $A_{380} = 1.8273C - 0.1597$  ( $C$  为 Api 浓度, mmol/L),  $R^2 = 0.997$ 。Api 在 3.0 - 3.5 mmol/L 出现平台,因此该高浓度区间不纳入线性回归。附录E图S2进一步验证了  $A_{380}$  对酶投量 (V/V) 的响应: 2% - 20% 范围内近似线性 ( $A_{380} = 0.0867x + 0.3856$ ,  $R^2 = 0.9999$ ), 30% 点偏离线性范围。上述结果证明 380 nm 可用于区分底物 Nar 和产物 Api, 并可在 Api 浓度 0.2 - 1.5 mmol/L 范围内作为 Api 定量检测波长, 同时为低酶投量筛选设定了酶投量 (V/V) 线性工作区间。

检测条件优化确定 3 mmol/L 底物浓度、10% 酶量、反应液稀释 4 倍为最佳参数组合。样品通过稀释保持在 Api 的 0.2 - 1.5 mmol/L 线性工作区间内后测定  $A_{380}$ ; 该条件下, 不同酶量梯度的  $A_{380}$  信号区分度显著。每板设立野生型 AncFNS-I-48 对照、空载体对照及无酶空白对照, 以  $\Delta A_{380}$  为相对活性指标, 完成 96 孔板规模突变体库的快速初筛, 阳性克隆再经摇瓶培养和 HPLC 定量验证。

### 3.4 AI 计算指导的 AncFNS-I-48 突变体设计

以 AncFNS-I-48 为亲本, 按“全位点虚拟扫描-PSSM/Pythia 初筛-CataPro 排序-实验验证”的路线推进 (图4(a))。各计算模块用于压缩实验候选空间, 不将预测分数直接等同于催化或稳定性实验证据。Pythia 与 CataPro 原始预测分数列于附录C表2。Pythia  $\Delta\Delta G$  为预测热稳定性变化, 负值表示预测稳定化; N297D、H261F、V159E、M230D 和 Q71Y 的  $\Delta\Delta G$  分别为 -0.660、-13.021、-0.505、-2.388 和 -1.225, CataPro 预测的  $\log_{10}$  (kcat/Km) 分别为 0.8996、0.8404、0.9079、0.9178 和 0.8847。这些结果仅用于候选排序, 不替代实测活性或稳定性证据。

PSSM 对每个位点的 19 种非野生型替换进行保守性/共进化评分, Pythia 评估替换对热稳定性的预测影响; 排除  $\text{Fe}^{2+}$  配位核心区域后, 再以 CataPro 对候选进行催化参数排序。三类评分用于逐级压缩实验空间, 最终以实验测得的相对初始反应速率和 20 mL 试剂反应活性作判断。

排除  $\text{Fe}^{2+}$  配位核心残基后, 选取综合排序靠前的 29 个单点突变体进行实验。WT 及突变体粗酶液的 SDS-PAGE 见图 4(b)。图 4(c) 为低酶投量初始线性区间内测得、并按对应目标条带 IOD 完成表达量校正的采用  $A_{380}$  检测的相对初始反应速率; 图 4(d) 为所选突变体在粗酶总蛋白 2 mg/mL (Bradford 法定量) 的 20 mL 试剂反应体系中采用 HPLC 检测的 WT 归一化相对活性。两组数据均由 3 次独立培养重复分别以同批 WT 测定值归一化后计算均值  $\pm$  标准差 ( $n = 3$ ), 但因蛋白投量和评价指标不同, 不作数值上的直接横向比较。显著性采用双侧单样本 t 检验与归一化 WT 水平 1.0 比较, 仅以 \*\* 标示  $p < 0.01$ 。

低酶投量下经 SDS-PAGE 灰度校正的相对初始反应速率见图 4 (c)。V159E 相对初始反应速率最高, 为  $3.06 \pm 0.26$  倍 ( $p < 0.01$ ); H261F 为  $1.50 \pm 0.04$  倍 ( $p < 0.01$ )。两者是在既定显著性阈值下显著高于 WT 的单点突变体。

计算排序靠前的位点并不全部对应显著的实测增强, 说明 PSSM、Pythia 和 CataPro 更适合作为候选优先级工具; 表达量、折叠状态、底物进入和反应体系因素仍会共同影响最终表型。

其他均值高于 WT 的突变体包括 I115V ( $2.46 \pm 0.40$ )、N297D ( $1.71 \pm 0.13$ )、L186F ( $1.58 \pm 0.18$ )、M230D ( $1.51 \pm 0.40$ )、H261A ( $1.22 \pm 0.33$ )、A169K ( $1.09 \pm 0.08$ )、T204P ( $1.02 \pm 0.06$ ) 和 M179L ( $1.02 \pm 0.05$ ); 这些数值均为相对初始反应速率, 在  $p < 0.01$  阈值下未达到显著提高, 不能仅凭均值判定为稳定的正向突变。

粗酶总蛋白投量为 2 mg/mL (Bradford 法定量) 的 20 mL 反应体系的结果见图 4(d)。I115V、V159E、M230D 和 N297D 的相对活性分别为  $2.15 \pm 0.15$ 、 $2.14 \pm 0.11$ 、 $2.02 \pm 0.10$  和  $1.88 \pm 0.13$ , 均显著高于归一化 WT 水平; M85L 为  $0.35 \pm 0.03$ , 显著低于 WT ( $p < 0.01$ )。L186F、H261F、

T204P、A169K、H261A 和 Q71Y 的均值分别为 1.80、1.40、1.12、1.01、0.82 和 0.70，在  $p < 0.01$  阈值下未达到显著差异。该组结果表征较高粗酶总蛋白投量试剂反应条件下的产物生成能力，不等同于图 4(c) 的低酶投量初始反应速率。

为观察六个代表性高活力位点在预测复合物中的空间环境，将野生型与突变体的 Chai-1 柚皮素复合物模型进行叠合（图 7）。AncFNS-I-48 采用自身序列编号时，催化残基 His217、Asp219、His275 和 Lys196 分别对应 PcFNS-I 的 His218、Asp220、His276 和 Lys197。图 7 仅展示突变位点与底物及周围结构的邻近关系，不构成已证实的催化机制。六个突变均不与柚皮素直接接触，预测模型中的配位三联体几何未见明显改变，提示这些突变的作用更可能发生在直接配位中心之外，但该模型不足以排除间接影响。I115V、L186F 和 H261F 位于入口环或 DHGH 盖子的第二壳层，V159E、M230D 和 N297D 位于相对远端区域。

预测模型提示若干可检验的结构假说：I115V 位于 F113-T117 入口环附近，Ile 替换为 Val 可能减小局部空间占据；L186F 可能改变 DHGH 模体附近

的疏水填充；H261F 位于 DHGH 模体的第一个 His，可能改变该区域的极性相互作用；V159E 可能影响其与 K163 附近的局部静电环境；M230D 可能引入新的局部极性接触；N297D 位于口袋盖表面，可能改变附近的氢键或水化网络。上述解释均基于未包含  $\text{Fe}^{2+}$  和 2-OG 的预测复合物，不能据此确定催化机制，仍需含辅因子的复合物建模、分子动力学、热稳定性测定或回复突变验证。

### 3.5 讨论

以伞形科为边界收集现生 FNS-I 候选序列，并结合 ASR 与 CataPro 排序，在保持谱系相关性的同时扩展了实验候选空间；但本研究的初始集合仅含 22 条访问记录，分类范围与祖先节点采样仍然有限，且仅实验测试了两个相邻祖先节点。纯酶活性及 pH 响应共同表明，AncFNS-I-48 在基准反应体系中活性最高，并在 pH 6.0 – 8.5 范围内保持较高产物生成能力。该宽 pH 响应使其能够在 pH 8.0 下兼顾反应表现与柚皮素溶解性，可作为后续较高底物负荷、传质和放大评价的候选亲本，但现有结果尚不足以判定其工业适用性。这一潜力

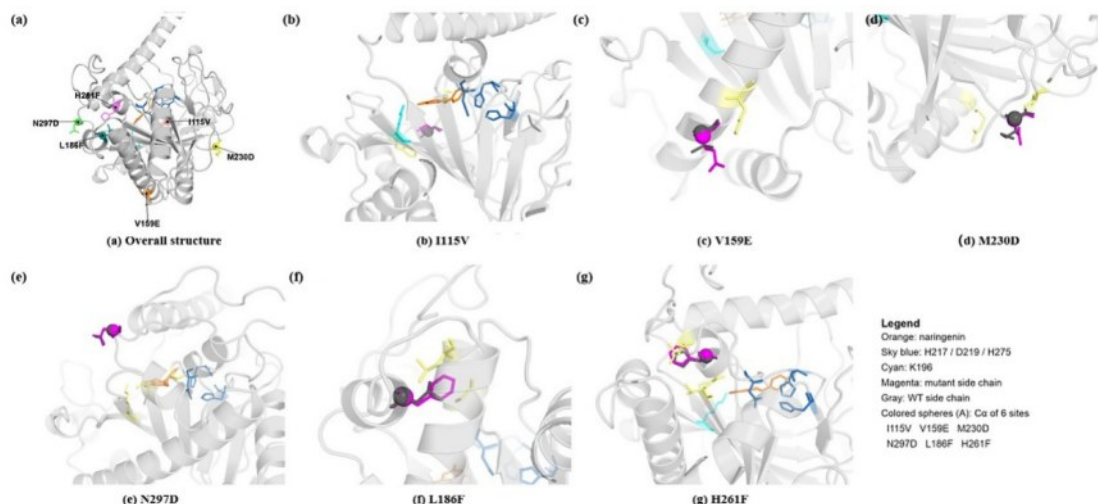


图 7 AncFNS-I-48 - 柚皮素预测复合物中六个代表性候选突变的结构环境

(a) 六个突变位点在整体结构上的分布；橙色棍状为柚皮素。(b - g) 各突变与野生型叠合的局部放大：(b) I115V，(c) V159E，(d) M230D，(e) N297D，(f) L186F，(g) H261F。品红色为突变体侧链，灰色为野生型；模型未包含  $\text{Fe}^{2+}$  和 2-OG。所示空间关系为预测性假说，不构成已验证的催化机制。

Fig. 7 Predicted structural environment of six representative candidate mutations in the AncFNS-I-48-naringenin complex

(a) Overall distribution of the six mutation sites; naringenin is shown as orange sticks. (b-g) Local overlays of each mutant with the wild type: (b) I115V, (c) V159E, (d) M230D, (e) N297D, (f) L186F, and (g) H261F. Mutant side chains are magenta and wild-type side chains are gray. The model does not include  $\text{Fe}^{2+}$  or 2-OG. The displayed spatial relationships are predictive hypotheses and do not establish a validated catalytic mechanism.

仍需通过不同底物负荷下的实测溶解度、连续反应稳定性、比活和稳态动力学进一步验证。CataPro将部分有活性序列纳入候选,但6种纯化酶的实测排序与模型排序并非完全一致,说明模型适合用于优先级筛选,不能替代实验。

计算流程将谱系内现生序列收集、FoldMason结构感知比对、祖先节点重建和催化参数预测串联起来,但当前序列集合仅包含伞形科22条叶节点记录,且仅实验验证了两个相邻祖先节点和单点突变,仍可能遗漏其他谱系分支、深层祖先节点及组合突变的协同效应。

29个单点突变体在低酶水平下测定初始反应速率,并经SDS-PAGE目标条带灰度校正,以降低可溶性表达量差异对候选排序的影响;其中V159E和H261F在 $p < 0.01$ 阈值下显著高于WT。在20 mL制备级反应中采用2 mg/mL粗酶总蛋白浓度(Bradford法定量)进一步验证其实际转化能力,I115V、V159E、M230D和N297D显著高于WT,而M85L显著低于WT。前者反映初始线性阶段的相对速率,后者反映较高粗酶总蛋白投量和既定反应时间下的产物生成能力,两类数值不作直接定量比较,也不能替代纯化酶比活或稳态动力学参数。这些结果显示单点空间仍有优化潜力;多点组合可能存在协同或拮抗效应,需在单点证据基础上设计组合并进行评价。

本研究仍有以下局限:(1)祖先序列仅选取叶侧节点及其父节点这一对相邻节点,未探索更深层或不同分支的祖先序列。(2)虽有Chai-1复合物模型,关键位点的结构-功能关系尚缺晶体或高分辨实验结构及分子动力学模拟的直接证据。

后续可在现有PSSM/Pythia/CataPro框架上引入蛋白语言模型嵌入和组合突变设计,并结合含 $\text{Fe}^{2+}/2\text{-OG}$ 的复合物建模、分子动力学及回复突变实验验证结构假设。

## 4 结 论

本研究在伞形科内收集22条FNS-I候选现生序列记录,并结合ASR和CataPro排序获得6个FNS-I候选酶。纯酶3次平行测定中,以AncFNS-I-48活性归一化为1.00, PcFNS-I、AncFNS-I-50、

CcFNS-I、DcFNS-I和CmFNS-I分别为0.63、0.63、0.26、0.06和0.05。50 mmol/L磷酸盐缓冲液中的pH响应显示, AncFNS-I-48在pH 6.0 - 8.5范围内均保持较高产物生成能力;结合柚皮素在偏碱性条件下较好的溶解性,pH 8.0被确定为后续条件测试和放大评价的起始条件。25 - 30 °C更适合长时间反应;  $\text{FeSO}_4$ 梯度中250  $\mu\text{mol/L}$ 的芹菜素生成量最高,为 $2.674 \pm 0.101$  mmol/L。29个单点突变体在低酶水平下以SDS-PAGE灰度校正的相对初始反应速率评价,V159E和H261F分别达到WT的3.06和1.50倍,并在 $p < 0.01$ 阈值下显著提高。粗酶总蛋白投量为2 mg/mL (Bradford法定量)的20 mL试剂反应体系中,I115V、V159E、M230D和N297D的WT归一化相对活性分别为2.15、2.14、2.02和1.88,均显著高于WT ( $p < 0.01$ )。综合两类指标,本文优先推荐低酶投量下相对初始反应速率最高的V159E;I115V在20 mL制备级反应中的产物生成指标均值最高,但与V159E接近,且现有分析未检验二者间差异,因此将其作为侧重制备级产物生成的备选。结果支持将谱系序列收集、祖先重建和计算排序用于压缩实验候选空间,并表明AncFNS-I-48具有进一步开展高底物负荷与放大工艺评价的潜力。该“谱系序列收集-祖先重建-AI预测-计算突变-高通量筛选”联合策略为FNS-I及其他2-ODD家族酶的挖掘与改造提供方法参考。

## 参 考 文 献

- [1] Salehi B, Venditti A, Sharifi-Rad M, et al. The therapeutic potential of apigenin[J]. International Journal of Molecular Sciences,2019,20(6):1305.
- [2] Hostetler GL,Ralston RA,Schwartz SJ. Flavones: food sources, bioavailability, metabolism, and bioactivity[J]. Advances in Nutrition,2017,8(3):423-435.
- [3] Williams CA,Grayer RJ. Anthocyanins and other flavonoids[J]. Natural Product Reports,2004,21(4):539-573.
- [4] Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology[J]. Plant Physiology,2001,126(2):485-493.
- [5] He M, Min JW, Kong WL, et al. A review on the pharmacological effects of vitexin and isovitexin[J]. Fitoterapia,2016,115:74-85.

- [6] Shukla S, Gupta S. Apigenin: a promising molecule for cancer prevention[J]. *Pharmaceutical Research*, 2010, 27(6): 962-978.
- [7] Madunić J, Madunić IV, Gajski G, et al. Apigenin: a dietary flavonoid with diverse anticancer properties[J]. *Cancer Letters*, 2018, 413: 11-22.
- [8] 苑林宏, 夏薇, 赵秀娟, 等. 芹菜素通过抑制 PKB/Akt 激酶活性诱导人胃癌细胞凋亡[J]. *科学通报*, 2007, 52(13): 1525-1529.  
YUAN L H, XIA W, ZHAO X J, et al. Apigenin induces apoptosis of human gastric cancer cells by inhibiting PKB/Akt kinase activity[J]. *Science Bulletin*, 2007, 52(13): 1525-1529.
- [9] 蔡婧, 刘文文, 赵向丽, 等. 芹菜素对肝癌细胞生长及基因表达的影响[J]. *世界华人消化杂志*, 2010, 18(6): 542-549.  
CAI J, LIU A W, ZHAO X L, et al. Effects of apigenin on hepatocellular carcinoma cell growth and gene expression[J]. *World Chinese Journal of Digestology*, 2010, 18(6): 542-549.
- [10] 宋晓童. 8-(6"-伞形酮)-芹菜素衍生物的抗糖尿病活性研究[D]. 天津: 天津科技大学, 2019.  
SONG X T. Antidiabetic activity of 8-(6"-umbelliferone) - apigenin derivatives[D]. Tianjin: Tianjin University of Science and Technology, 2019.
- [11] 周云峰, 罗飞, 付欢欢, 等. 芹菜素对肺癌细胞氨肽酶N抑制作用机制的研究[J]. *解放军医药杂志*, 2015, 27(8): 21-24.  
ZHOU Y F, LUO F, FU H H, et al. Mechanism of apigenin-mediated inhibition of aminopeptidase N in lung cancer cells[J]. *Medical & Pharmaceutical Journal of Chinese People's Liberation Army*, 2015, 27(8): 21-24.
- [12] 闻永举, 申秀丽. 柚皮苷半合成芹菜素的研究[J]. *江苏农业科学*, 2012, 40(5): 189-190.  
WEN Y J, SHEN X L. Semisynthesis of apigenin from naringin[J]. *Jiangsu Agricultural Sciences*, 2012, 40(5): 189-190.
- [13] 陈平, 韩俊杰, 刘涛, 等. 水溶性芹菜素的合成[J]. *食品科学*, 2009, 30(18): 69-72.  
CHEN P, HAN J J, LIU T, et al. Synthesis of water-soluble apigenin[J]. *Food Science*, 2009, 30(18): 69-72.
- [14] 陈新, 李平. 新的半合成芹菜素制备工艺[P]. 中国发明专利, CN 200410000704, 2004.  
CHEN X, LI P. A new semisynthetic process for preparing apigenin[P]. Chinese invention patent, CN 200410000704, 2004.
- [15] 肖金霞, 靳莎, 杨雪峰. 天然产物芹菜素的化学全合成[J]. *生物技术世界*, 2014(9): 10-11.  
XIAO J X, JIN S, YANG X F. Total chemical synthesis of the natural product apigenin[J]. *Biotechnology World*, 2014(9): 10-11.
- [16] 李凤霞. 查尔酮异构酶基因过表达对新疆雪莲类黄酮生物合成的调控[D]. 北京: 中国科学院植物研究所, 2006.  
LI F X. Regulation of flavonoid biosynthesis in *Saussurea involucrata* by overexpression of chalcone isomerase[D]. Beijing: Institute of Botany, Chinese Academy of Sciences, 2006.
- [17] 李春霞. 黄酮类化合物芹菜素酶法合成体系的建立及条件优化[D]. 扬州: 扬州大学, 2017.  
LI C X. Establishment and optimization of an enzymatic system for apigenin synthesis[D]. Yangzhou: Yangzhou University, 2017.
- [18] Martens S, Mithofer A. Flavones and flavone synthases[J]. *Phytochemistry*, 2005, 66(20): 2399-2407.
- [19] Ayabe S, Akashi T. Cytochrome P450s in flavonoid metabolism[J]. *Phytochemistry Reviews*, 2006, 5(2-3): 271-282.
- [20] Gebhardt Y, Witte S, Forkmann G, et al. Molecular evolution of flavonoid dioxygenases in the family Apiaceae[J]. *Phytochemistry*, 2005, 66(11): 1273-1284.
- [21] Martens S, Forkmann G, Matern U, et al. Cloning of parsley flavone synthase I[J]. *Phytochemistry*, 2001, 58(1): 43-46.
- [22] Martens S, Forkmann G, Britsch L, et al. Divergent evolution of flavonoid 2-oxoglutarate-dependent dioxygenases in parsley[J]. *FEBS Letters*, 2003, 544(1-3): 93-98.
- [23] 周志华, 周金林, 叶德晓, 等. 黄酮合酶及其应用[P]. 中国发明专利, CN 110616204 B, 2019.  
ZHOU Z H, ZHOU J L, YE D X, et al. Flavone synthase and its application[P]. Chinese invention patent, CN 110616204 B, 2019.
- [24] Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. *Nature*, 2024, 630(8016): 493-500.
- [25] Gumulya Y, Back JM, Wun SJ, et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction[J]. *Nature Catalysis*, 2018, 1(11): 878-888.
- [26] Trudeau DL, Kaltenbach M, Tawfik DS. On the potential origins of the high stability of reconstructed ancestral proteins[J]. *Molecular Biology and Evolution*, 2016, 33(10): 2633-2641.
- [27] Wang Z, Xie D, Wu D, et al. Robust enzyme discovery and engineering with deep learning using CataPro[J]. *Nature Communications*, 2025, 16(1): 2736.
- [28] 李贵祥, 徐铮, 李莎, 等. 基于同源建模和定点突变技术研究嗜热型L-阿拉伯糖异构酶与D-半乳糖的亲和作用[J]. *催化学报*, 2012, 33(10): 1701-1707.  
LI G X, XU Z, LI S, et al. Homology modeling and site-directed mutagenesis of thermophilic L-arabinose isomerase for interaction with D-galactose[J]. *Chinese Journal of Catalysis*, 2012, 33(10): 1701-1707.
- [29] 郭发旭, 冯全, 张建华, 等. 人工智能驱动的酶改造与设计研究进展[J]. *生物技术通报*, 2025, 41(12): 50-65.

GUO F X, FENG Q, ZHANG J H, et al. Progress in artificial-intelligence-driven enzyme engineering and design[J]. Biotechnology Bulletin, 2025, 41(12): 50-65.

- [30] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood[J]. Molecular Biology and Evolution, 2007, 24(8): 1586-1591.
- [31] Gilchrist C L M, Mirdita M, Steinegger M. Multiple protein structure alignment at scale with FoldMason[J]. Science, 2026, 391(6784): 485-488.
- [32] Chai Discovery, Boitreau J, Dent J, et al. Chai-1: decoding the molecular interactions of life[EB/OL]. bioRxiv, 2024: 2024.10.10.615955. DOI: 10.1101/2024.10.10.615955.
- [33] Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs[J]. Nucleic Acids Research, 1997, 25(17): 3389-3402.
- [34] Sun J, Zhu T, Cui Y, et al. Structure-based self-supervised learning enables ultrafast protein stability prediction upon mutation[J]. The Innovation, 2025, 6(1): 100750.



**通讯作者:** 王晟(1983—), 男, 博士, 上海智峪生物创始人兼CEO, AI蛋白质结构预测先驱, 早期工作为 AlphaFold 奠定基础。曾任腾讯 AI Lab 高级专家, CASP 竞赛冠军, 谷歌引用近 8000 次, 入选全球前 2% 顶尖科学家。主导开发 fastAF2 算法, 推动 AI 生物技术从实验室到工业闭环, 树立生物制造新标杆。研究方向为人工智能驱动蛋白质结构预测、酶设计与合成生物学。

E-mail: wangsheng@zelixir.com。



**第一作者:** 谢东奇(1994—), 男, 硕士, 高级工程师, 现任上海智峪生物科技有限公司研发经理。研究方向为合成生物学驱动的工业酶创制, 涉及定向进化、蛋白制备及生物过程工程, 聚焦高效生物催化剂开发与产业化工艺研究。

E-mail: xiedongqi@zelixir.com。

## 作者贡献声明

王晟、温霖、谢东奇: 提出概念、基本框架、提出研究方向、设计论文框架; 谢东奇、张顺治、邵茜: 设计研究思路、设计研究方案、研究方案可行性调查分析、实施研究过程; 谢东奇、朱国梁、丁霏希: 进行实验、进行对比实验、实验分析; 谢东奇、张顺治、邵茜: 文献调研与整理; 谢东奇、张顺治、邵茜: 确定研究对象范围、收集数据、采集、清洗与分析数据; 谢东奇、朱国梁、郑良振: 承担技术开发、绘制图谱; 丁霏希、郑良振、温霖: 参与提出构思, 参与撰写论文; 谢东奇、郑良振、王晟: 起草论文、修订论文、审核论文; 张顺治、邵茜、温霖: 参与论文修订、论文最终版本修订。

所有作者均已阅读并同意文章的已发布版本。

## 机构审查委员会声明

不适用。

## 知情同意声明

不适用。

## 数据可用性声明

支持本研究结论的数据均已包含在本文及其补充材料中; 其他原始实验数据可向通讯作者合理请求获取。

## 利益冲突

王晟是本刊编委, 在同行评议过程中未进行影响公平性的行为。其他作者没有利益冲突。

附录A FNS-I 候选酶计算挖掘所用序列及CataPro 预测结果

本研究在伞形科内收集 22 条 FNS-I 候选现生序列记录。表 1 汇总其登录号、物种、以柚皮素为

底物的 CataPro 预测 kcat/Km 及天然酶选取依据。ASR 按 v2.3 所述流程以 22 条序列作为叶节点，使用 FoldMason 结构感知比对及其 Newick 树。CataPro 排序保留全部 22 条访问记录；图 5 仅为现生序列关系展示，将完全相同的氨基酸序列合并为 17 个叶节点，不参与 Node 48 和 Node 50 的选择。

表1 22条伞形科FNS-I候选现生序列及CataPro预测催化效率（底物：柚皮素）

Table 1 Candidate extant Apiaceae FNS-I sequences and CataPro-predicted catalytic efficiencies (substrate: naringenin)

| 登录号            | 物种                    | kcat/Km | 相对 FNS1 | 筛选说明            |
|----------------|-----------------------|---------|---------|-----------------|
| B3RFV8         | Conium maculatum      | 9.400   | 1.055   | 入选 CmFNS-I      |
| ABG78795       | Conium maculatum      | 9.400   | 1.055   | 同种,未入选          |
| Q4QUR7         | Conium maculatum      | 9.342   | 1.048   | 同种,未入选          |
| Q7XZQ8         | Petroselinum crispum  | 8.912   | 1.000   | 文献对照 PcFNS-I    |
| XP_017240664.1 | Daucus carota         | 8.287   | 0.930   |                 |
| KAL1822882.1   | Daucus carota         | 8.287   | 0.930   |                 |
| XGD09043.1     | Daucus carota         | 8.285   | 0.930   |                 |
| XGD09042.1     | Apium graveolens      | 7.802   | 0.875   |                 |
| A6XFT1         | Cuminum cyminum       | 7.700   | 0.864   | 入选 CcFNS-I      |
| ABG78793       | Angelica archangelica | 7.682   | 0.862   |                 |
| XP_074379576.1 | Apium graveolens      | 7.559   | 0.848   |                 |
| AAX21537       | Apium graveolens      | 7.491   | 0.840   |                 |
| Q4QUR8         | Apium graveolens      | 7.491   | 0.840   |                 |
| Q4QUR9         | Daucus carota         | 7.291   | 0.818   | 入选 DcFNS-I      |
| AAX21536       | Daucus carota         | 7.291   | 0.818   | 与 AAX21542 序列相同 |
| AAX21542       | Daucus carota         | 7.291   | 0.818   | 与 AAX21536 序列相同 |
| XP_074335403.1 | Apium graveolens      | 6.474   | 0.726   | FNS-I-like      |
| KAL1822880.1   | Daucus carota         | 6.034   | 0.677   |                 |
| XP_017240659.2 | Daucus carota         | 5.452   | 0.612   | isoform X1      |
| XP_017239195.2 | Daucus carota         | 5.201   | 0.584   |                 |
| XP_017240660.1 | Daucus carota         | 5.077   | 0.570   |                 |
| KAL1822879.1   | Daucus carota         | 4.983   | 0.559   |                 |

注：按预测 kcat/Km 降序排列；kcat/Km 单位为 L·mmol<sup>-1</sup>·s<sup>-1</sup>，‘相对 FNS1’ 为相对欧芹 Q7XZQ8 的倍数。22 条访问记录对应 17 条唯一全长序列；重复组为 AAX21537/Q4QUR8、B3RFV8/ABG78795、Q4QUR9/AAX21536/AAX21542 和 XP\_017240664.1/KAL1822882.1。ASR 以 22 条记录作为 FoldMason/PAML 流程的叶节点；图 5 为避免视觉冗余，将完全相同序列的 accession 合并标在同一叶节点。CataPro 预测仅用于候选优先级排序，不替代实测催化性能。

附录B 候选酶氨基酸序列

B. 1 PcFNS-I (Petroselinum crispum, UniProt: Q7XZQ8)

MAPTTITALAKEKTLNLDVFVRDEDERPKVA  
YNQFSNEIPIISLAGLDDSDGRRPEICRKIVKAC

EDWGIFQVVDHGDIDSLISEMTRLISREFFALPAE  
EKLEYDTTGKKRGGFTISTVLQGDDAMDWREF  
VTYFSYPINARDYSRWPKKPEGWRSTTEVYSEK  
LMVLGAKLLEVLSEAMGLEKGDLTACVDME  
QKVLINYYPTCPQPDLTGVRRTDPTITILLQ  
DMVGGLQATRDGGKWTWTVQPVGEAFVVLGD  
HGHYLSNGRFRNADHQAVVNSTSSRLSIATFQN

PAQNAIVYPLKIREGEKAILDEAITYAEMYKKCM  
TKHIEVATRKKLAKEKRLQDEKAKLEMKSAD  
ENLA

## B.2 DcFNS-I (Daucus carota, UniProt: Q4QUR9)

MAPTTITALAKEKTLNSDFVRDEDERPKVA  
YNQFSTEIPIISLAGIDDDSNRRPEVCRKIVEAF  
EDWGIFQVVDHGDIDSLAEMSRLSREFFALPAE  
EKLRDYYTGGKRGFTISTHLQGDDVKDWREF  
VVYFSYPVDARDYSRCPDKPEGWRSVTEVYSE  
KLMALGAKLLEVLSEAMGLEKEALTEACVNME  
QKVLINYYPTCPQPDITLGVRRTDPTGITILLQ  
DMVGGLQATRDGGKWTWITVQPVGEAFVNLGD  
HGHYLSNGRFGKNADHQAVVNSTSSRLSIATFQN  
PAQNAIVYPLKIREGEKPILEEAMTYAEMYKKN  
MTKHIEVATQKKLAKEKRLQNEKAKLETKF

## B.3 CcFNS-I (Cuminum cyminum, UniProt: A6XFT1)

MAPTTITALAQEKTLNSDFVRDEDERPKVA  
YNQFSTEIPIISLAGIDDDSKGRRPEVCRKIVEAF  
EDWGIFQVVDHGVDSALISEMSRLSREFFALPAE  
EKLRDYYTGGKRGFTISTHQQGDDVRDWREF  
VTYFSYPVDARDYSRWPEKPEGWRSVTEVYSE  
KLMVLGAKLLEVLSEAMGLDKGALTACVNM  
EQKVLINYYPTCPEPDITLGVRRTDPTGITILLQ  
DMVGGLQATRDGGKWTWITVQPVGEVFNLDG  
HGHYLSNGRFGKNADHQAVVNSTSSRLSIATFQN  
PAQNAIVYPLKIREGEKPILEEAITYAEMYKKNM  
TKHIEVATQKKLAKEKRLQEEKAKLETKTKSAD  
GILA

## B.4 AncFNS-I-48 (Ancestral sequence)

MAPTTITALAKEKTLNLDVFRDEDERPKVA  
YNQFSNEIPIISLAGLDDNGRRPEICRKIVEAFE  
DWGIFQVVDHGDIDSLISEMSRLSREFFALPAEE  
KLRYDYYTGGKRGFTISTHLQGDDVRDWREFV  
TYFSYPIDARDYSRWPDKPEGWRSTTEVYSEKL

MVLGAKLLEVLSEAMGLEKEALTKACVNMEQ  
KVLINYYPTCPQPDITLGVRRTDPTGITILLQD  
MVGGLQATRDGGKWTWITVQPVGEAFVNLGDH  
GHYLSNGRFGKNADHQAVVNSTSSRLSIATFQNP  
AQNAIVYPLKIREGEKAILDEAITYAEMYKKNM  
TKHIEVATQKKLAKEKRLQDEKAKLEM

## B.5 AncFNS-I-50 (Ancestral sequence)

MAPTTITALSKEKTLNLDVFRDEDERPKVA  
YNQFSNEIPIISLAGLDDNGRRPEICRKIVEAFE  
DWGIFQVVDHGDIDSLISEMSRLSREFFALPAEE  
KLRYDYYTGGKRGFTISTHLQGDDVRDWREFV  
TYFSYPISARDYSRWPKKPEGWRSTTEVYSEKL  
MVLGAKLLEVLSEAMGLEKEALTKACVNMEQ  
KVLINYYPTCPQPDITLGVRRTDPTGITILLQD  
MVGGLQATRDGGKWTWITVQPVGEAFVNLGDH  
GHYLSNGRFGKNADHQAVVNSTSSRLSIATFQNP  
AQNAIVYPLKIREGEKAILDEAITYAEMYKKNM  
TKHIEVATQKKLAKEKRLQDEKAKLEM

## B.6 CmFNS-I (Conium maculatum, UniProt: B3RFV8)

MAPTTITALAQEKTLNLAIVFRDEDERPKVA  
YNEFSNEIPIISLAGLENDSDGRRPEICRKIVEAFE  
NWGIFQVVDHGDIDSLISEMSRLSREFFALPAEE  
KLRYDYYTGGKRGFTISTHLQGDDVRDWREFV  
TYFSYPIDARDYSRWPDKPEGWRSITEVYSERL  
MVLGAKLLEVLSEAMGLEKEALTKACVNMEQ  
KVLINYYPTCPEPDITLGVRRTDPTGITVLLQD  
MVGGLQATRDGGKWTWITVQPVGEAFVNLGDH  
GHYLSNGRFGKNADHQAVVNSSSSRLSIATFQNP  
QNAIVYPLKIREGEKAILDEAITYAEMYKKNMT  
KHIEVATLKKLAKEKRLQDEKANMEKKSASAH  
GISA

## 附录 C 计算筛选评分表

进入实验验证的 29 个 AncFNS-I-48 突变体的  
Pythia  $\Delta\Delta G$  与 CataPro 原始预测分数如下表所示，

$\Delta\Delta G$  为 Pythia 预测的热稳定性变化，负值表示稳定化；log10 kcat、log10 Km 和 log10 (kcat/Km) 为 CataPro 以柚皮素为底物输出的原始预测分数。该表用于记录计算排序变量，不能替代实测活性。

表2 29个 AncFNS-I-48 突变体的 Pythia  $\Delta\Delta G$  与 CataPro 原始预测分数  
Table 2 Pythia  $\Delta\Delta G$  and raw CataPro prediction scores for 29 AncFNS-I-48 variants

| 突变体   | $\Delta\Delta G$ | log10 kcat | log10 Km | log10(kcat/Km) |
|-------|------------------|------------|----------|----------------|
| WT    | 0                | -0.9012    | -1.6531  | 0.7627         |
| N297D | -0.660           | -0.6889    | -1.7147  | 0.8996         |
| H261F | -13.021          | -0.7562    | -1.6892  | 0.8404         |
| V159E | -0.505           | -0.6827    | -1.7178  | 0.9079         |
| M230D | -2.388           | -0.6705    | -1.7160  | 0.9178         |
| Q71Y  | -1.225           | -0.7097    | -1.7267  | 0.8847         |
| L186F | -2.810           | -0.6771    | -1.7105  | 0.9106         |
| A169K | -7.122           | -0.6765    | -1.7328  | 0.9156         |
| A277V | -5.261           | -0.6978    | -1.7201  | 0.8907         |
| S86K  | -4.435           | -0.6695    | -1.7250  | 0.9139         |
| F64A  | -2.735           | -0.6392    | -1.7110  | 0.9375         |
| T204P | -2.171           | -0.6720    | -1.7207  | 0.9169         |
| Y31A  | -3.942           | -0.6499    | -1.7287  | 0.9309         |
| I77V  | -1.415           | -0.6765    | -1.7199  | 0.9162         |
| T187I | -2.529           | -0.6814    | -1.7297  | 0.9105         |
| H261A | -5.188           | -0.6662    | -1.7309  | 0.9282         |
| Q71I  | -1.805           | -0.7405    | -1.7316  | 0.8733         |
| Y160F | -6.374           | -0.6501    | -1.7151  | 0.9296         |
| A277I | -4.016           | -0.7126    | -1.7194  | 0.8755         |
| M179L | -4.563           | -0.6512    | -1.7209  | 0.9257         |
| N272S | -1.730           | -0.6750    | -1.7319  | 0.9219         |
| N200F | -5.802           | -0.8435    | -1.6849  | 0.7826         |
| M165R | -6.027           | -0.6476    | -1.7284  | 0.9321         |
| S135T | -3.758           | -0.6788    | -1.7222  | 0.9120         |
| P302C | -1.314           | -0.7805    | -1.7272  | 0.8503         |
| M85L  | -5.528           | -0.6630    | -1.7217  | 0.9187         |
| I115V | -3.575           | -0.6689    | -1.7196  | 0.9195         |
| N16P  | -8.916           | -0.6783    | -1.7234  | 0.9151         |
| T157F | -1.742           | -0.6462    | -1.7342  | 0.9354         |
| D260L | -4.925           | -0.8403    | -1.6748  | 0.7825         |

附录D 突变体构建引物序列

位点。

D.1 模板序列

以祖先重构酶 AncFNS-I-48 编码基因为定点突变模板。基因全长 1062 bp，经密码子优化为大肠杆菌偏好，合成后克隆至 pET-28a 载体 NdeI/XhoI

D.2 定点突变引物

AncFNS-I-48 编码序列（5'→3'）：  
ATGGCCCCGACCACCATTACGGCGCTGGC  
GAAGGAGAAAACCTTAAATCTGGATTTTGTGC

GCGACGAAGATGAACGTCCGAAAGTGGCATAT TGAATTTGTGACGTA~~CTTT~~AGCTATCCGATTGA  
AATCAGTTTCAGCAACGAAATTCGGATTATTAGC TGCACGTGACTATAGCCGTTGGCCGGATAAAC  
CTGGCGGGTCTGGATGATGACAACGGTCGTCG CGGAAGGCTGGCGTAGCACCACCGAAGTGTAT  
TCCGGAGATTATTCGTAAAATTGTGGAAGCGG AGCGAGAAACTGATGGTTCTGGGCGCGAAGC  
CAGAAGATTGGGGCATT~~TTTT~~TATGTGGTGGAC TGCTGGAAGTTTAAAGTGAAGCGATGGGCTTA  
CACGGTATTGATAGCGGCTTAATCAGCGAGATG GAGAAAGAAGCGCTGACGAAAGCCTGCGTGA  
AGCCGCCTGAGTCGTGAGTTTTTTGCGTTACC ACATGGAACAGAAAGTGCTGATCAATTACTAC  
AGCGGAAGAAAAGCTGCGTTATGATACCACGG CCGACCTGCCCCGAGCCGGATCTGACCCTGGG  
GCGGTAAACGTGGCGGCTATACCATTAGCACC CGTGCGTCGTCATACCGATCCGGGACCATCA  
CATCTGCAAGGCGATGATGTGCGTGATTGGCG CGATTCTGCTGCAAGATATGGTGGGTGGCCTG

表3 29个突变体的引物序列

Table 3 Primer sequences for 29 variants

| 突变体   | 氨基酸替换      | 正向引物 (F, 5'→3')                                                       | 反向引物 (R, 5'→3')         |
|-------|------------|-----------------------------------------------------------------------|-------------------------|
| N297D | Asn297→Asp | TTTCAGAATCCGGCGCAG <del>GAT</del> GCGATTGTGTATCCACTGAAAAT             | CGCCGGATTCTGAAAGGT      |
| H261F | His261→Phe | GTTGTGAACCTGGGTGAT <del>TTT</del> GGCCACTATTTAAGCAATGG                | ACCCAGGTTCAACAACAAAC    |
| V159E | Val159→Glu | TGGCGTAGCACCACCGAA <del>GAA</del> TATAGCGAGAAACTGATGGT                | GGTGGTGCTACGCCAGCCT     |
| M230D | Met230→Asp | ACGATTCTGCTGCAAGAT <del>GAT</del> GTGGGTGGCCTGCAAGCGAC                | TTGCAGCAGAATCGTGATG     |
| Q71Y  | Gln71→Tyr  | GAAGATTGGGGCATT <del>TTTT</del> <del>TAT</del> GTGGTGGACCACGGTATTGAT  | AATGCCCCAATCTTCAAACG    |
| L186F | Leu186→Phe | GGCTTAGAGAAAGAAGCG <del>TTT</del> TACGAAAGCCTGCGTGAACAT               | TTCTTTCTCTAAGCCCATCG    |
| A169K | Ala169→Lys | AAACTGATGGTTCTGGGC <del>AAA</del> AGCTGCTGGAAGTTTAAAG                 | CAGAACCATCAGTTTCTCG     |
| A277V | Ala277→Val | AAGAATGCGGATCATCA <del>AGT</del> GTGGTGAACAGTACCAGCAG                 | ATGATCCGCATTCTTGAACAC   |
| S86K  | Ser86→Lys  | GGCTTAATCAGCGAGATG <del>AAA</del> CGCCTGAGTCGTGAGTTTTT                | CTCGTGATTAAAGCCGCT      |
| F64A  | Phe64→Ala  | CGTAAAATTGTGGAAGCG <del>GCA</del> GAAAGATTGGGGCATT <del>TTT</del> CAG | TTCCACAATTTACGGCAAAT    |
| T204P | Thr204→Pro | CTGATCAATTACTACCCG <del>CCG</del> TCCCCGCAGCCGGATCTGAC                | GTAGTAATTGATCAGCACTTT   |
| Y31A  | Tyr31→Ala  | GAACGTCCGAAAGTGGCAG <del>GCA</del> AATCAGTTCAGCAACGAAATT              | CACTTTCGGACGTTTCATCT    |
| I77V  | Ile77→Val  | CAGGTGGTGGACCACGGT <del>GTT</del> GATAGCGGCTTAATCAGCGAG               | GTGGTCCACCACCTGAAAAAT   |
| T187I | Thr187→Ile | TTAGAGAAAGAAGCGCTG <del>ATT</del> AAAGCCTGCGTGAACATGGAAC              | CGCTTCTTTCTCTAAGCCCAT   |
| H261A | His261→Ala | GTTGTGAACCTGGGTGAT <del>GCA</del> GGCCACTATTTAAGCAATGG                | ACCCAGGTTCAACAACAAAC    |
| Q71I  | Gln71→Ile  | GAAGATTGGGGCATT <del>TTTT</del> <del>ATT</del> GTGGTGGACCACGGTATTGAT  | AATGCCCCAATCTTCAAACG    |
| Y160F | Tyr160→Phe | CGTAGCACCACCGAAGTG <del>TTT</del> AGCGAGAAACTGATGGTTCT                | TTCCGTGGTGCTACGCCAG     |
| A277I | Ala277→Ile | AAGAATGCGGATCATCA <del>ATT</del> GTGGTGAACAGTACCAGCAG                 | ATGATCCGCATTCTTGAACAC   |
| M179L | Met179→Leu | GAAGTTTAAAGTGAAGCG <del>CTG</del> GGCTTAGAGAAAGAAGCGCT                | TTCACTTAAAACTTCCAGCAG   |
| N272S | Asn272→Ser | AGCAATGGCCGTTTCAAG <del>AGC</del> GCGGATCATCAAGCCGTGGT                | GAAACGGCCATTGCTTAAAT    |
| N200F | Asn200→Phe | GAACAGAAAGTGCTGATC <del>TTT</del> TACTACCCGACCTGCCCCGAG               | CAGCACTTTCTGTTCCATG     |
| M165R | Met165→Arg | GTGTATAGCGAGAAACTG <del>CGT</del> GTTCTGGGCGCGAAGCTGCT                | TTTCTCGCTATACACTTCGGT   |
| S135T | Ser135→Thr | GAATTTGTGACGTA <del>CTTT</del> <del>ACCT</del> TATCCGATTGATGCACGTGAC  | GTACGTCACAAATTCACGCCAAT |
| P302C | Pro302→Cys | CAGAACGCGATTGTGTAT <del>TGT</del> CTGAAAATTCGTGAGGGCGAAAAAG           | CACAATCGCGTTCTGCGCCGGAT |
| M85L  | Met85→Leu  | AGCGGCTTAATCAGCGAG <del>CTG</del> AGCCGCCTGAGTCGTGAGTT                | GCTGATTAAGCCGCTATCAAT   |
| I115V | Ile115→Val | AAACGTGGCGGCTTACC <del>GTT</del> AGCACCCATCTGCAAGGCGAT                | AAAGCCGCCACGTTTACC      |
| N16P  | Asn16→Pro  | GCGAAGGAGAAAACCTTAC <del>CCG</del> CTGGATTTTGTGCGCGACGAAG             | GGTTTTCTCCTTCGCCAG      |
| T157F | Thr157→Phe | GAAGGCTGGCGTAGCACC <del>TTT</del> GAAAGTGATAGCGAGAAACT                | GCTACGCCAGCCTTCCGGT     |
| D260L | Asp260→Leu | TTTGTGTGAACCTGGGT <del>CTG</del> CATGGCCACTATTTAAGCAAT                | CAGGTTCAACAACACGCGCCT   |

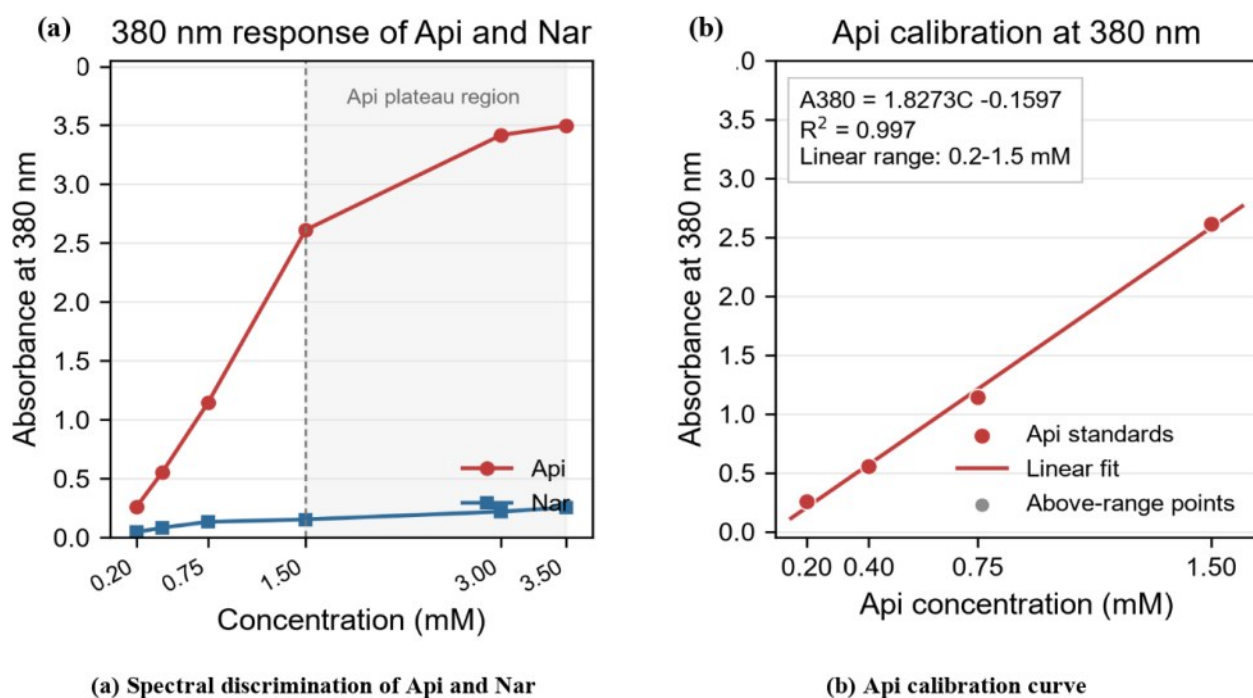
注：引物由北京擎科生物科技有限公司合成，纯化等级按供应商订单记录执行。突变位点以粗体下划线标注于序列中。全质粒 PCR 条件：98 °C 30 s；98 °C 10 s、55 °C 30 s、72 °C 3 min，25 个循环；72 °C 5 min。

CAAGCGACGCGTGACGGTGGCAAAACCTGGA  
TTACCGTGCAACCGGTGGAAGGCGCGTTTGT  
GTGAACCTGGGTGATCATGGCCACTATTTAAG  
CAATGGCCGTTTCAAGAATGCGGATCATCAAG  
CCGTGGTGAACAGTACCAGCAGTCGTCTGAGC  
ATTGCGACCTTTCAGAATCCGGCGCAGAACGC  
GATTGTGTATCCACTGAAAATTCGTGAGGGCG  
AAAAAGCGATTTTAGATGAAGCCATTACCTATG  
CGGAAATGTATAAGAAGAATATGACCAAACATA  
TTGAAGTGCGACCCAGAAGAAATTAGCCAAG  
GAAAAACGTCTGCAGGACGAAAAGGCAAAATT

AGAAATG

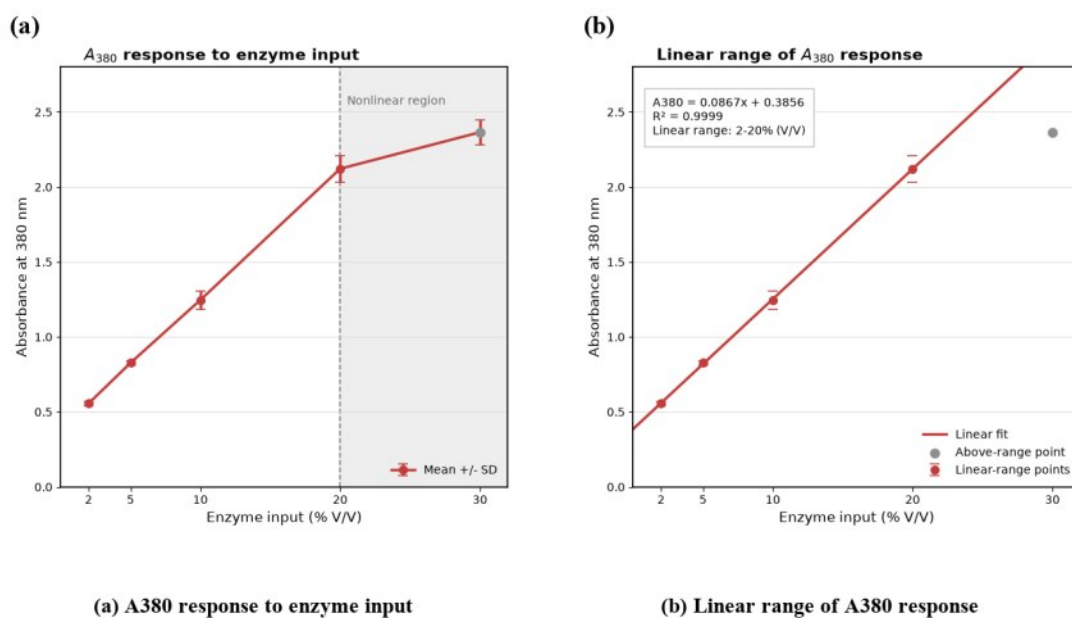
所有单点突变体以 AncFNS-I-48 为模板, 采用全质粒 PCR 扩增法构建。正向引物 (F) 在突变位点处引入目标碱基替换, 反向引物 (R) 与 F 的 5' 端互补。扩增产物经 DpnI 消化模板后转化大肠杆菌 DH5 $\alpha$ , 测序验证后提取质粒转化 BL21 (DE3) 进行表达。共计 29 个突变体, 引物序列见表3 (按氨基酸位点排序):

## 附录 E 支撑图片



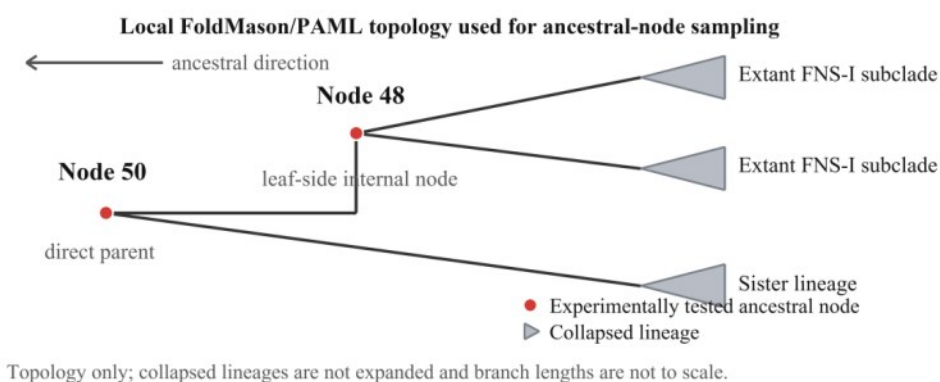
**图S1** 380 nm处柚皮素 (Nar) 与芹菜素 (Api) 的吸收响应及芹菜素定量工作区间。(a) 不同浓度标准品在380 nm处的吸光度。Api在0.2 - 1.5 mmol/L范围内随浓度增加而升高, 在3.0 - 3.5 mmol/L出现平台; Nar在0.2 - 3.5 mmol/L范围内保持较低且持续上升的响应, 与Api在380 nm处具有明显区分。灰色区域表示Api超出线性工作区间的浓度点。(b) Api在0.2 - 1.5 mmol/L范围内的校准曲线,  $A_{380} = 1.8273C - 0.1597$  (C为Api浓度, mmol/L),  $R^2 = 0.997$ ; 灰色点为超出线性工作区间的观测值。

**Figure S1** Absorbance response of naringenin (Nar) and apigenin (Api) at 380 nm and the validated Api quantification range. (a) Absorbance of standards at 380 nm across the tested concentrations. Api increased over 0.2 - 1.5 mmol/L and reached a plateau from 3.0 to 3.5 mmol/L, whereas Nar remained lower and increased progressively, providing clear discrimination at 380 nm. The gray region denotes Api concentrations outside the linear working range. (b) Api calibration curve over 0.2 - 1.5 mmol/L:  $A_{380} = 1.8273C - 0.1597$ , where C is Api concentration in mmol/L;  $R^2 = 0.997$ . Gray points are observations outside the validated linear range.



**图S2**  $A_{380}$  对酶投量 (V/V) 的响应及线性区间。(a) 2%、5%、10%、20%和30% (V/V) 酶投量下的 $A_{380}$  均值 ± 标准差 ( $n = 3$ ); 2% - 20% (V/V) 范围内响应近似线性, 30%点偏离线性范围。(b) 2% - 20% (V/V) 范围内的线性拟合,  $A_{380} = 0.0867x + 0.3856$ ,  $R^2 = 0.9999$ ;  $x$ 为酶投量百分数 (V/V)。反应时间均为1 h。

**Figure S2**  $A_{380}$  response to enzyme input (% V/V) and the validated linear range. (a) Mean  $\pm$  SD ( $n = 3$ )  $A_{380}$  values at enzyme inputs of 2%, 5%, 10%, 20%, and 30% (V/V); the response was approximately linear from 2% to 20% (V/V), whereas the 30% point deviated from the linear range. (b) Linear fit over 2% - 20% (V/V):  $A_{380} = 0.0867x + 0.3856$ ,  $R^2 = 0.9999$ , where  $x$  is the enzyme-input percentage (V/V). All reactions were conducted for 1 h.



**图S3** ASR 候选节点的局部拓扑与Node 48/Node 50 选择关系。ASR 输入由附录A 表1 所列22 条FNS-I 现生序列、FoldMason 结构感知比对及其Newick 树组成, 并由PAML codeml 完成祖先状态推断。Node 48 为紧邻现生叶节点的浅层内部节点, Node 50 为其直接父节点; 红色节点为进入实验验证的祖先序列, 灰色三角表示未展开的现生或姐妹谱系。Node 编号为codeml rst 输出中的内部节点标签, 不是图5 NJ 拓扑的bootstrap 支持度。本图仅展示与候选选择直接相关的局部父子拓扑, 枝长不按比例绘制。

**Figure S3** Local topology and selection relationship of Node 48 and Node 50 in ASR. The ASR input comprised the 22 extant FNS-I sequence records listed in Appendix A Table 1, the FoldMason structure-aware alignment and its Newick tree, followed by ancestral-state inference with PAML codeml. Node 48 is the shallow internal node adjacent to extant leaves, and Node 50 is its direct parent. Red nodes denote the ancestral sequences tested experimentally, and gray triangles denote collapsed extant or sister lineages. Node numbers are internal-node labels from the codeml rst output, not bootstrap support values in the Figure 5 NJ topology. Only the local parent-descendant topology relevant to candidate selection is shown; branch lengths are not drawn to scale.