

基于 5'mRNA 二级结构优化的卤醇脱卤酶高效表达及应用

朱钧豪, 汤晓玲, 郑仁朝*

浙江工业大学 生物工程学院, 浙江 杭州

朱钧豪, 汤晓玲, 郑仁朝. 基于 5'mRNA 二级结构优化的卤醇脱卤酶高效表达及应用[J]. 微生物学报, 2025, 65(11): 5105-5118.
ZHU Junhao, TANG Xiaoling, ZHENG Renchao. Efficient expression and application of a halohydrin dehalogenase based on optimization of 5'mRNA secondary structure[J]. *Acta Microbiologica Sinica*, 2025, 65(11): 5105-5118.

摘要: 【目的】卤醇脱卤酶作为一种生物催化剂, 兼具闭环与开环反应活性, 在手性环氧化合物等合成领域应用广泛, 其绿色高效制备具有重要意义。本研究通过对卤醇脱卤酶 HheC 编码序列 N 端进行基于 mRNA 二级结构的序列优化, 实现 HheC 的高效表达, 并将其应用于手性环氧氯丙烷的合成。【方法】采用 mRNA 预测工具预测 5'mRNA 二级结构与热力学性质, 以降低 mRNA 二级结构稳定性、提高折叠自由能(ΔG)为目标, 在不改变氨基酸序列的前提下设计高表达 5'mRNA 序列, 并表征酶的表达效率和催化性能。【结果】基于 5'mRNA 序列设计获得了 HheC 突变体, 其蛋白表达水平从 16.71% 提升至 33.39%, 对 1,3-二氯-2-丙醇的相对酶活提高了 3 倍。【结论】基于 5'mRNA 二级结构优化可显著提高 HheC 表达水平, 提升目标产物合成效率, 为构建酶法催化合成手性环氧氯丙烷路线奠定了基础。

关键词: 卤醇脱卤酶; mRNA 二级结构; 翻译表达调控

Efficient expression and application of a halohydrin dehalogenase based on optimization of 5'mRNA secondary structure

ZHU Junhao, TANG Xiaoling, ZHENG Renchao*

College of Biotechnology and Bioengineering, Zhejiang University of Technology, Hangzhou, Zhejiang, China

Abstract: [Objective] As biocatalysts, halohydrin dehalogenases can catalyze both cyclization and

资助项目: 国家重点研发计划(2021YFA1501400)

This work was supported by the National Key Research and Development Program of China (2021YFA1501400).

*Corresponding author. E-mail: zhengrc@zjut.edu.cn

Received: 2025-04-15; Accepted: 2025-06-20; Published online: 2025-07-17

ring-opening reactions and are widely used in the synthesis of chiral epoxides and other compounds. The eco-friendly and efficient preparation of halohydrin dehalogenases is thus of great significance. In this study, we optimized the N-terminus of the coding sequence of the halohydrin dehalogenase HheC based on the mRNA secondary structure to achieve the efficient expression of this enzyme and then applied this enzyme in the synthesis of chiral epichlorohydrin. **[Methods]** The mRNA prediction tools was used to predict the secondary structure and thermodynamic properties of 5'mRNA. To reduce the stability of the mRNA secondary structure and increase folding free energy (ΔG), we designed the 5'mRNA sequence without changing the amino acid sequence. Furthermore, we characterized the expression efficiency and catalytic performance of this enzyme. **[Results]** The HheC mutant was obtained *via* the design of the 5'mRNA sequence, with the protein level increasing from 16.71% to 33.39% and the relative activity towards 1,3-dichloro-2-propanol increasing by three folds. **[Conclusion]** The optimization based on the secondary structure of 5'mRNA improves the expression level of HheC and enhances the synthesis efficiency of the target product, laying a foundation for constructing the route of enzymatic catalytic synthesis of chiral epichlorohydrin.

Keywords: halohydrin dehalogenase; mRNA secondary structure; regulation of translation

卤醇脱卤酶 (halohydrin dehalogenases, HDDHs, EC 4.5.1.X) 是一类重要的生物催化剂, 它能够催化邻卤醇羟基对卤素进行分子内亲核取代反应, 生成对应的环氧化物与卤化氢; 同时, 也能催化相应的逆反应, 即受 N_3^- 、 CN^- 、 NO_2^- 等亲核试剂进攻, 催化环氧化物开环, 具有与转氨酶、大部分脱氢酶类似的双向催化功能。根据序列同源性与结构特征的差异, 卤醇脱卤酶可分为 A-G 等多个亚型^[1-2]。其中, HheC 对短链邻卤醇具有高活性和高选择性, 在手性环氧氯丙烷、他汀侧链、恶唑烷酮等合成中具有重要应用, 其高效制备也备受关注^[3-4]。

在现代生物技术领域, 外源基因在宿主生物中的高效表达已成为众多应用的基石。影响外源基因在宿主中高效表达的关键因素是多方面的, 包括基因序列的优化、合适启动子及调控元件的选择, 以及对宿主细胞翻译机制的调控。其中, mRNA 结构, 尤其是 5'mRNA 二级结构被视为关键因素之一。mRNA 二级结构能够显著影响翻译起始效率, 进而影响所编码蛋白质的整体表达水平^[5-8]。适当的 mRNA 二级结

构有助于蛋白质的正确折叠^[5], 改善核糖体的分配, 防止核糖体碰撞与堵塞, 提高翻译效率^[6,9]。反之, 高稳定性的二级结构可能阻碍核糖体的移动, 使核糖体需要更长时间解开配对的碱基, 导致“翻译停滞”^[8]; 核糖体下游高稳定性的二级结构甚至可能引发移码突变^[10]。此外, ΔG 是衡量 RNA 二级结构稳定性的关键热力学参数, 反映了 RNA 分子折叠成特定结构时的能量变化。 ΔG 越低(即负值越大)表明形成该结构需要释放更多能量, 二级结构越稳定。

对于翻译起始而言, 翻译起始区(translation initiation region, TIR)是启动翻译的重要功能区域, 包括核糖体结合位点(ribosome binding site, RBS)、SD 序列(Shine-Dalgarno sequence)、起始密码子及其下游的部分序列^[11]。TIR 序列对翻译起始的影响可归因于区域内 mRNA 二级结构引起的核糖体可及性差异。TIR 内一定程度的 mRNA 二级结构能够延长其半衰期, 防止其被过早降解^[12-13], 且由于 30S 亚基与 RBS 之间具有较强的亲和力, 并不影响二者的结合^[14-15]。当 mRNA 稳定性达到一定阈值时, 其稳定性与

翻译效率呈负相关,稳定的二级结构会阻碍二者结合,从而抑制翻译起始,显著影响目的基因表达^[16-17]。此外,mRNA二级结构在调控目的蛋白表达水平中发挥着多维度作用,涉及核糖体结合效率、翻译起始复合体组装、mRNA降解动力学及其与转录后调控网络的耦合机制。结构紧密的5'mRNA二级结构可遮蔽核糖体结合位点(如起始密码子AUG附近或SD序列),阻碍小亚基装载,显著抑制翻译起始^[18];还能通过限制eIF4E、eIF4G等起始因子对mRNA帽结构的结合,抑制翻译起始复合体的形成^[19]。由此可见,mRNA二级结构的折叠方式与稳定性能够通过多种机制影响蛋白质的合成速率,这为聚焦于5'mRNA二级结构设计,提升卤醇脱卤酶在大肠杆菌(*Escherichia coli*)中的异源表达水平提供了重要指导。

本研究以HHDH亚型之一的卤醇脱卤酶HheC为研究对象。通过降低编码序列5'mRNA二级结构稳定性在不改变氨基酸序列的前提下提高HheC在大肠杆菌中的异源表达水平,并表征酶的表达效率及其对1,3-二氯-2-丙醇(1,3-dichloro-2-propanol, 1,3-DCP)的催化性能,以期为手性环氧氯丙烷的酶法合成奠定基础。

1 材料与方法

1.1 材料

1.1.1 菌株

本研究使用的表达宿主为*E. coli* BL21(DE3),表达载体为pET-28b(+),二者均由本实验室保藏;研究对象卤醇脱卤酶HheC_{L143N/A159T/P175S/W249P}为实验室前期构建,标记为HheC_{M4}。

1.1.2 培养基

LB培养基(g/L):蛋白胨 10.0,酵母粉 5.0,NaCl 10.0,pH 7.4。

LB固体培养基:在LB液体培养基的基础上加入20 g/L的琼脂粉配制而成。

1.2 mRNA二级结构与热力学性质预测

1.2.1 基于ViennaRNA的mRNA二级结构与热力学性质分析

为提高mRNA二级结构预测结果的可信度,准确反映热力学性质变化趋势,本研究采用3种不同算法的预测工具对mRNA二级结构与热力学性质进行分析。利用ViennaRNA在线工具(<http://rna.tbi.univie.ac.at/>)^[20-21]对5'mRNA(基于起始密码子的80 nt序列)的二级结构、折叠自由能变化(ΔG)及核苷酸热力学性质进行预测。其mRNA折叠算法包括最小自由能(minimum free energy, MFE)模型与配分函数(partition function)模型。MFE模型通过搜索自由能最低的构象来预测序列最稳定的折叠结构;配分函数模型基于统计力学框架计算所有可能构象的热力学加权,从而获得碱基配对概率与结构多样性信息,并设置避免孤对碱基对(avoid isolated base pairs)以提高预测结果的合理性。选择Turner模型(2004)^[22]的热力学参数,设定折叠温度为37 °C,在1 mol/L盐浓度下进行能量计算。

1.2.2 基于Mfold的mRNA二级结构与热力学性质分析

源于UNAFold在线工具(<https://www.unafold.org/mfold/applications/rna-folding-form.php>)的Mfold模块^[23]同样是基于最小自由能原理开发的核酸二级结构预测算法。其核心采用Zuker等提出的动态规划(dynamic programming)算法^[24]与Turner模型(1999)^[25]的热力学参数系统地搜索所有可能的碱基配对组合,从而精确识别最稳定构象及多个亚稳态结构。该模块提供灵活的结构约束功能,支持强制配对、禁配设定和多链杂交模拟,在寡核苷酸设计、引物筛选、RNA干扰等领域均有广泛应用。设定折叠温度为37 °C,在1 mol/L盐浓度下进行能量计算。

1.2.3 基于UFold的mRNA二级结构分析

UFold^[26]基于深度学习框架,通过端到端的

图像化神经网络模型实现高精度与高效率的 mRNA 二级结构预测。其主体采用 U-Net 模型^[27]全卷积神经网络架构,能够在保持位置信息的同时逐层提取局部与全局特征。通过对 $L \times L$ 的配对矩阵进行对称性处理,筛除不符合 RNA 二级结构物理约束的配对,并结合线性规划进行结构优化,最终获得符合热力学与几何限制的稳定二级结构模型。该方法相比传统自由能最小化方法具有更高的预测准确性与处理效率,在处理假结预测与长程配对识别任务中表现更优秀。

1.3 重组菌株构建

以 HheC_{M4} 为模板,通过全质粒 PCR 对目标序列进行突变。质粒 DNA 由质粒小提试剂盒(购自北京擎科生物科技股份有限公司)提取获得。引物序列见表 1。PCR 反应体系(25 μ L): 2 \times Phanta Max Buffer (p505) 12.5 μ L, dNTP Mix (10 mmol/L) 0.5 μ L, Phanta Max Super-Fidelity DNA Polymerase 0.25 μ L, 均购自南京诺唯赞生物科技股份有限公司,上、下游引物(10 μ mol/L)各

1 μ L, DNA 模板 0.1 μ L, ddH₂O 补足至 25 μ L。PCR 反应条件: 95 $^{\circ}$ C 预变性 5 min; 95 $^{\circ}$ C 变性 30 s, 60 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 延伸 3.5 min, 30 个循环; 72 $^{\circ}$ C 终延伸 10 min。经 1.5% 琼脂糖凝胶电泳验证 PCR 产物后,用 Dpn I 于 37 $^{\circ}$ C 消化模板,通过热激法将产物转化至 *E. coli* BL21(DE3),涂布于含 50 μ g/mL 卡那霉素的 LB 固体培养基,37 $^{\circ}$ C 培养过夜,挑取单菌落送北京擎科生物科技股份有限公司测序验证。

1.4 目的蛋白诱导表达与分析

1.4.1 诱导表达

挑取重组菌落至含 50 μ g/mL 卡那霉素的 LB 液体培养基试管中,于 37 $^{\circ}$ C、200 r/min 培养过夜。按 2% 体积分数转接至含 50 μ g/mL 卡那霉素的 LB 液体培养基中,于 37 $^{\circ}$ C、180 r/min 培养至 OD_{600} 约为 0.6–0.8,加入 IPTG (终浓度 0.1 mmol/L)进行目标蛋白的诱导表达,28 $^{\circ}$ C、180 r/min 诱导 12 h 后,于 8 000 r/min 离心 10 min,去除上清液后得到湿菌体,用于后续的分析与催化反应。

表1 5'mRNA突变引物序列

Table 1 Mutant primers of 5'mRNA sequences

| Primers name | Primer sequences (5'→3') |
|-------------------------|--|
| ins2A ^{GCA} -F | ATATACCATGGCATCTACCGCTATTGTGA |
| ins2A ^{GCU} -F | ATATACCATGGCTTCTACCGCTATTGTGA |
| ins2A ^{GCC} -F | ATATACCATGGCCTCTACCGCTATTGTGA |
| ins2A ^{GCG} -F | ATATACCATGGCGTCTACCGCTATTGTGA |
| ^a ins2A-R | CATGGTATATCTCCTTCTTAAAGTTAAACAAAAT |
| L1-F | TGGCATGGGCTCTGCTCTGCGTCTGTCTG |
| L1-R | GCAGAGCCCATGCCACCGAAATGCTTTACGT |
| L2-F | TGTGACAAACGTTAAGCATTTTCGGTGGCATGGGCTCTGCTCTCCGTCTGTCTG |
| L2-R | GCCACCGAAATGCTTAAACGTTTGTGACAATAGCGGTAGACATGGTATATCTCC |
| H1-F | ATGTCTACAGCTATTGTGACTAACGTAAAG |
| H1-R | AATAGCTGTAGACATGGTATATCTCCTTCT |
| H2-F | TAAAGCATTTTCGGGGGAATGGGCTCTGCGCTGCGT |
| H2-R | CCGAAATGCTTTACGTTAGTCACAATAGCTGTAGA |

a: 下游通用引物。

a: Downstream universal primer.

1.4.2 蛋白表达分析

将 50 g/L 重组菌重悬于 PB 缓冲液 (20 mmol/L, pH 7.4), 在冰水浴下以 180 W 超声破碎 10 min (超声时间 2 s, 间隔时间 4 s) 至澄清, 4 °C、12 000 r/min 离心 20 min 后, 取上清液用 PB 缓冲液稀释 3 倍, 并与 5×loading buffer 按 4:1 体积比混合, 在沸水浴中放置 5 min, 取 4 μL 进行 SDS-PAGE 分析。经染色脱色仪 10 min 工作程序处理后, 在凝胶成像仪下观察电泳结果。采用图像处理软件 ImageJ 1.54f 对 SDS-PAGE 图像进行 Rolling ball 算法^[28]背景校正后取反色, 进行目的蛋白表达量的半定量灰度分析。

表达水平与酶活数据包含 3 个生物学平行重复组, 以平均值±标准差(mean±SD)表示。组间显著差异分析采用单因素方差分析(one-way ANOVA), 使用 Waller-Duncan 多重比较法进行差异检测。显著性水平设定为 $P < 0.05$, 所有统计分析均使用 SPSS 27.0 软件完成。

1.5 卤醇脱卤酶催化 1,3-DCP 合成手性环氧氯丙烷

1.5.1 反应体系与反应条件

在 20 mL Gly/NaOH 缓冲液 (600 mmol/L, pH 9.8) 与乙酸乙酯体积比为 1:1 的反应体系中加入含卤醇脱卤酶突变体的重组菌 0.2 g (终浓度为 10 g/L), 400 mmol/L 1,3-DCP, 于 30 °C 孵育 10 min 后, 180 r/min 水浴振荡反应 5 min 后取样, 12 000 r/min 离心 1 min, 取上层有机相经无水硫酸钠处理后进行气相色谱检测。

1.5.2 气相检测方法

底物 1,3-DCP 与产物环氧氯丙烷(epichlorohydrin, ECH)使用 GC (Agilent 公司) 进行检测, 具体检测条件如下: 色谱柱型号为 BGB-175 手性毛细管柱(0.25 mm×30 m×0.25 μm); 柱箱条件为初始柱温 100 °C, 保持 4 min 后以 2 °C/min 升温至 110 °C, 而后以 30 °C/min 升高

至 180 °C, 保持 3 min。进样口温度设定为 250 °C, FID 检测口温度设定为 250 °C。载气为高纯氮, 流量设定为 1.5 mL/min, 分流比为 80:1。

2 结果与分析

2.1 5'mRNA 序列设计

基于翻译起始区 TIR 中 5'mRNA 二级结构稳定性与目标蛋白表达水平呈负相关的研究结论^[5-8], 本研究通过降低 mRNA 二级结构稳定性以提高翻译起始效率, 对 HheC 编码序列的 5'mRNA 序列进行优化, 从而增加 HheC 表达。在不改变氨基酸序列的前提下, 分别以提高和降低 ΔG 为目标对序列进行设计, 并设计相应引物对 5'mRNA 核苷酸序列进行突变。其中, 2 号位丙氨酸插入突变 ins2A 为实验中意外所得, 上标为相应密码子编码的丙氨酸, 其余序列均编码与 HheC_{M4} 相同的氨基酸。

以起始密码子上游不同位点作为最小折叠自由能预测的起始位点均能得到相近的自由能预测结果($R^2=0.9385$)^[29]。因此, 本研究以起始密码子(AUG) 80 nt 的序列(-20-60)作为研究对象, 预测 mRNA 二级结构并计算自由能。如表 2 所示, 不同密码子对应的 2 号位丙氨酸插入突变体与出发酶 HheC_{M4} 相比, mRNA 二级结构确实存在折叠自由能差异。由此, 设计预期 mRNA 二级结构稳定性降低的高表达突变体 HheC_{M4/H1} ($\Delta G=-13.20$ kcal/mol) 与 HheC_{M4/H2} ($\Delta G=-12.00$ kcal/mol), 同时设计预期高稳定性低表达的突变体 HheC_{M4/L1} ($\Delta G=-21.70$ kcal/mol) 与 HheC_{M4/L2} ($\Delta G=-24.10$ kcal/mol) 用于反向验证。在此基础上进一步表征了对应突变体在宿主中的表达水平。

2.2 目标蛋白表达水平与催化活性研究

SDS-PAGE 图像与灰度扫描半定量分析结果如图 1 所示, 出发株 HheC_{M4} 中 HheC (28.7 kDa) 占宿主细胞中可溶性蛋白的比例为 16.71%。

表2 5'mRNA序列设计

Table 2 Design of synonymous mutation sequences of 5'mRNA

| Mutant | Sequence (-20-60) ^a | ΔG_1 (kcal/mol) ^b | ΔG_2 (kcal/mol) ^c |
|-----------------------|--|---|---|
| HheC _{M4} | CUUUAAGAAGGAGAUAUACCAUGUCUACCGCUAUUGUGACU AACGUAAAGC AUUUCGGUGGCAUGGGCUCUGCGCUGCU | -18.50 | -21.40 |
| ins2A ^{GCA} | CUUUAAGAAGGAGAUAUACCAUGGCAUCUACCGCUAUUGUGACU AACGUAA AGCAUUUCGGUGGCAUGGGCUCUGCGCUG | -16.80 | -19.10 |
| ins2A ^{GCU} | CUUUAAGAAGGAGAUAUACCAUGGCUUCUACCGCUAUUGUGACU AACGUAA AGCAUUUCGGUGGCAUGGGCUCUGCGCUG | -14.80 | -18.20 |
| ins2A ^{GCC} | CUUUAAGAAGGAGAUAUACCAUGGCCUCUACCGCUAUUGUGACU AACGUAA AGCAUUUCGGUGGCAUGGGCUCUGCGCUG | -15.30 | -18.20 |
| ins2A ^{GCG} | CUUUAAGAAGGAGAUAUACCAUGGCGUCUACCGCUAUUGUGACU AACGUAA AGCAUUUCGGUGGCAUGGGCUCUGCGCUG | -14.80 | -18.20 |
| HheC _{M4/L1} | CUUUAAGAAGGAGAUAUACCAUGUCUACCGCUAUUGUGACU AACGUAAAGC AUUUCGGUGGCAUGGGCUCUGCUCUGCGCU | -21.70 | -23.20 |
| HheC _{M4/L2} | CUUUAAGAAGGAGAUAUACCAUGUCUACCGCUAUUGUGACAAAACGUAAAGC AUUUCGGUGGCAUGGGCUCUGCUCUCCGU | -24.10 | -25.40 |
| HheC _{M4/H1} | CUUUAAGAAGGAGAUAUACCAUGUCUACAGCUAUUGUGACU AACGUAAAGC AUUUCGGUGGCAUGGGCUCUGCGCUGCGCU | -13.20 | -16.40 |
| HheC _{M4/H2} | CUUUAAGAAGGAGAUAUACCAUGUCUACAGCUAUUGUGACU AACGUAAAGC AUUUCGGGGGAAUGGGCUCUGCGCUGCGCU | -12.00 | -12.30 |

^a: 下划线代表起始密码; ^b: 源于ViennaRNA; ^c: 源于Mfold。

^a: The underline represents the start codon; ^b: Calculated by ViennaRNA; ^c: Calculated by Mfold.

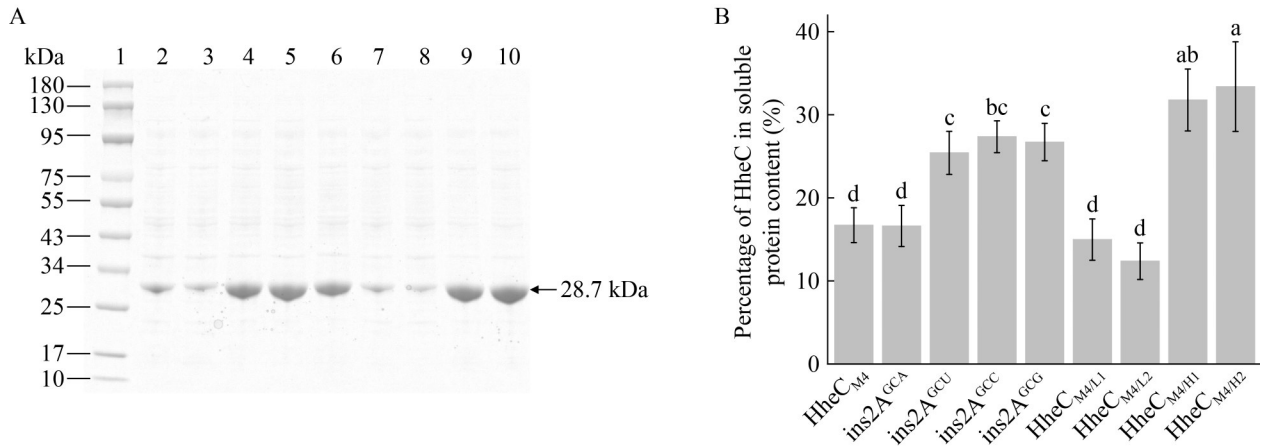
图1 HheC_{M4}及其突变体的表达水平

Figure 1 Expression levels of HheC_{M4} and its mutants. A: SDS-PAGE analysis chart of HheC_{M4} and its mutants (Lane 1: Protein marker; Lane 2: HheC_{M4}; Lane 3: ins2A^{GCA}; Lane 4: ins2A^{GCU}; Lane 5: ins2A^{GCC}; Lane 6: ins2A^{GCG}; Lane 7: HheC_{M4/L1}; Lane 8: HheC_{M4/L2}; Lane 9: HheC_{M4/H1}; Lane 10: HheC_{M4/H2}); B: Comparison of the soluble expression of HheC_{M4} and its mutants [Different lowercase letters indicate that there are significant differences in the expression level among different strains ($P < 0.05$)].

4种插入突变体中, $ins2A^{GCA}$ 表达水平变化不显著(16.61%), 而 $ins2A^{GCU}$ 、 $ins2A^{GCC}$ 、 $ins2A^{GCG}$ 表达水平提升幅度相近, HheC 占比分别为 25.42%、27.36% 与 26.71%。序列设计结果中具有更高 mRNA 稳定性的 HheC_{M4/L2} (12.38%) 表达水平低于 HheC_{M4/L1} (14.98%), 且较出发株均有所下降; 而具有更低稳定性的 HheC_{M4/H2} (33.39%) 表达水平高于 HheC_{M4/H1} (31.78%), 较出发株提升显著, 与设计预期一致。

HheC_{M4} 及其突变体的全细胞催化活性测定结果如图 2 所示, 与表达水平测定结果对应, 4种插入突变体中 $ins2A^{GCA}$ 具有最低的催化活性, 较出发株 HheC_{M4} 不存在显著变化, 而其余 3 种插入突变体催化活性均有显著提升。序列设计结果中具有高 mRNA 稳定性的 HheC_{M4/L1} 与 HheC_{M4/L2} 相对酶活均不及出发株, 且具有更低 ΔG 的 HheC_{M4/L2} (57.95%) 催化活性低于 HheC_{M4/L1} (72.18%); 而 mRNA 稳定性更低的

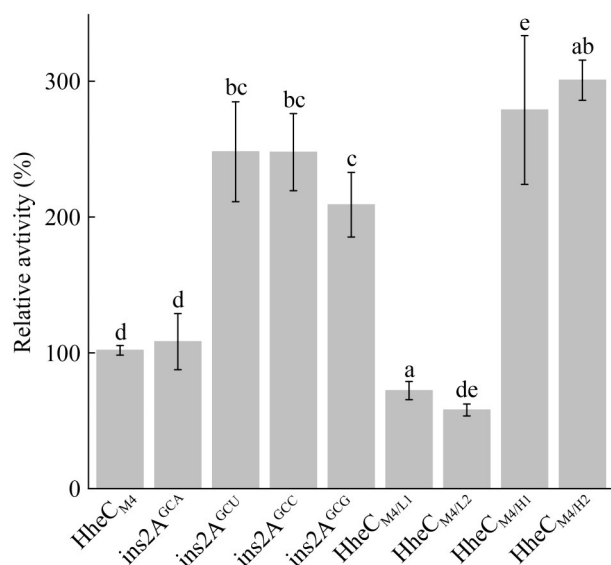


图2 HheC_{M4}及其突变体的相对酶活比较

Figure 2 Comparison of the relative enzyme activities of HheC_{M4} and its mutants. Different lowercase letters indicate that there are significant differences in the relative activity among different strains ($P < 0.05$).

HheC_{M4/H1} (278.85%)与 HheC_{M4/H2} (300.78%)催化活性均存在显著提升, 且具有最高 ΔG 的 HheC_{M4/H2} 表现出最高的催化活性, 与表达水平以及 mRNA 结构稳定性预测结果一致。

基于 HheC 表达水平与 ΔG 相关性绘制的散点图如图 3 所示。采用 2 种不同算法工具得到的 ΔG 结果均反映 mRNA 热力学稳定性与 HheC 表达量呈负相关性, 即随着自由能 ΔG 的升高 (mRNA 稳定性降低) HheC 表达量显著增加, 且 2 种算法的预测结果在分布趋势与数据密度上基本一致。

2.3 mRNA 二级结构分析

MFE 结构是依据自由能最小化原则得到的、mRNA 二级结构最稳定的构象, 代表了热力学上最可能形成的结构。质心结构 (centroid structure) 是通过概率模型计算所得的 mRNA 二级结构, 表示热力学构象集合中最接近平均构象的一种代表性结构^[30]。

源于 ViennaRNA 的 5'mRNA MFE 二级结构与质心二级结构预测结果如图 4 所示。基于熵值对结构进行着色, 反映了 mRNA 在该位置存在不同构象的可能性。熵值趋于 0 (暖色调) 表明

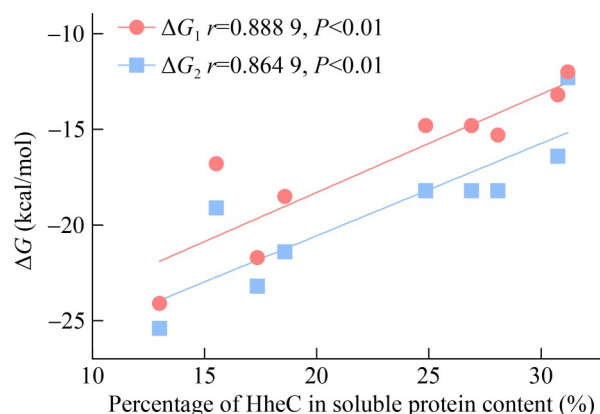


图3 mRNA二级结构折叠自由能(ΔG)与HheC表达水平相关性分析

Figure 3 Correlation analysis of the folding free energy (ΔG) of mRNA secondary structure and the expression level of HheC.

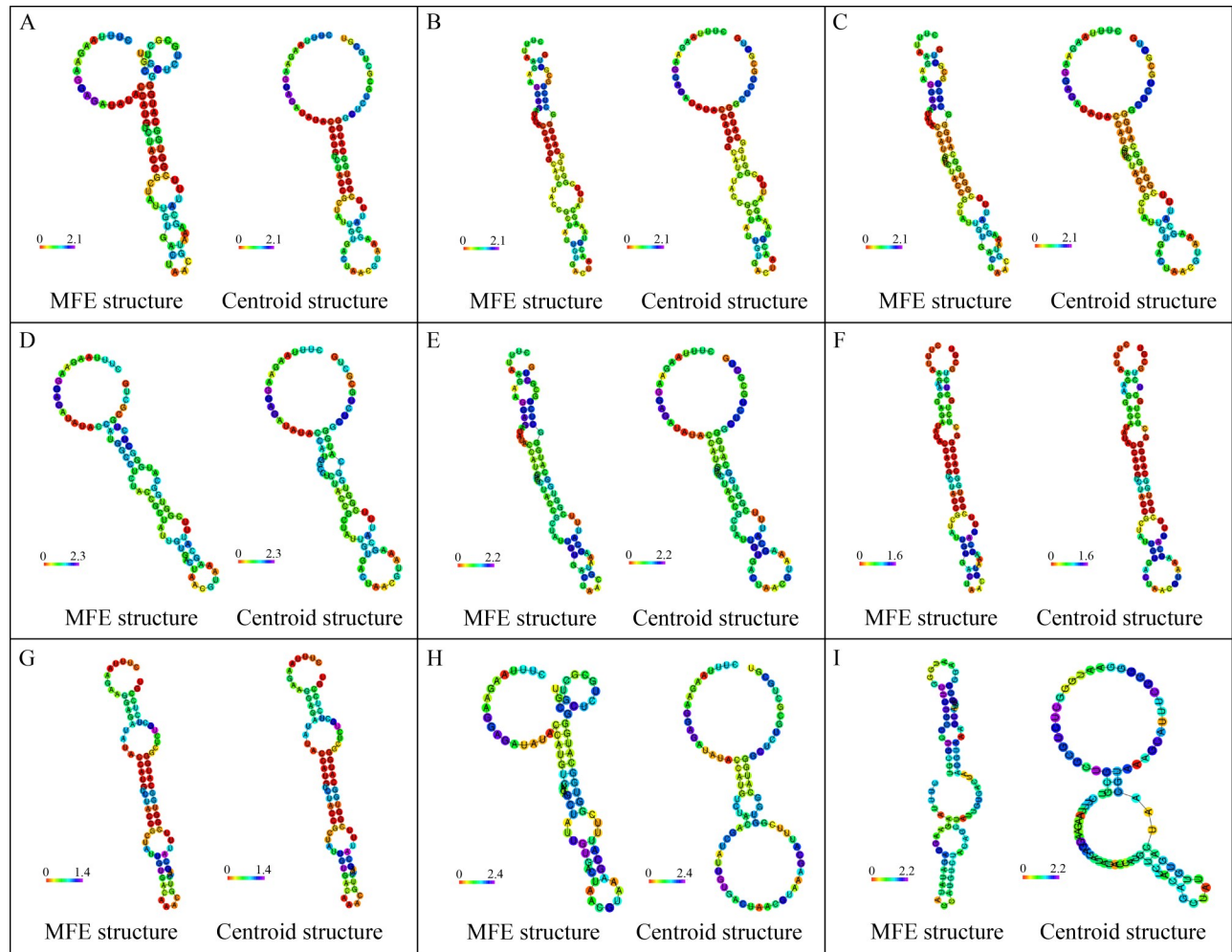


图4 ViennaRNA mRNA二级结构预测结果

Figure 4 Prediction of mRNA structure by ViennaRNA. A: HheC_{M4}; B: ins2A^{GCA}; C: ins2A^{GCU}; D: ins2A^{GCC}; E: ins2A^{GCG}; F: HheC_{M4/L1}; G: HheC_{M4/L2}; H: HheC_{M4/H1}; I: HheC_{M4/H2}.

该位点的结构较不稳定，可形成多种不同构象，如环或单链结构；熵值趋于冷色调则表明该位点的结构较稳定，通常位于茎(stem)结构中，更易于配对并形成稳定的双链结构。HheC_{M4}二级结构的大部分区域形成了一个紧凑的发夹结构，且MFE结构与质心结构具有较高的相似性(图4A)，表明该结构具有一定程度的稳定性。HheC_{M4/H1}(图4H)与HheC_{M4/H2}(图4I)的MFE结构与质心结构相似度较低，存在更多不稳定的非配对结构，因而易于解旋，有助于核糖体快速结合，通过提高翻译效率增加HheC表达量。相比之

下，HheC_{M4/L1}(图4F)与HheC_{M4/L2}(图4G)的MFE结构与质心结构均呈高度相似的发夹结构，表明两者具有高度稳定的mRNA二级结构，从而通过阻碍翻译起始降低HheC表达量。

源于Mfold的mRNA二级结构预测结果如图5所示。由于Mfold输出的mRNA结构仅包括MFE结构或热力学能量接近MFE的次优结构，因此大多呈现为相似度较高的连续茎环结构。HheC_{M4}(图5A)存在多个茎环结构，双链配对区域较为集中；ins2A^{GCA}(图5B)结构与其高度相似，配对密度略低；而ins2A^{GCU}(图5C)、

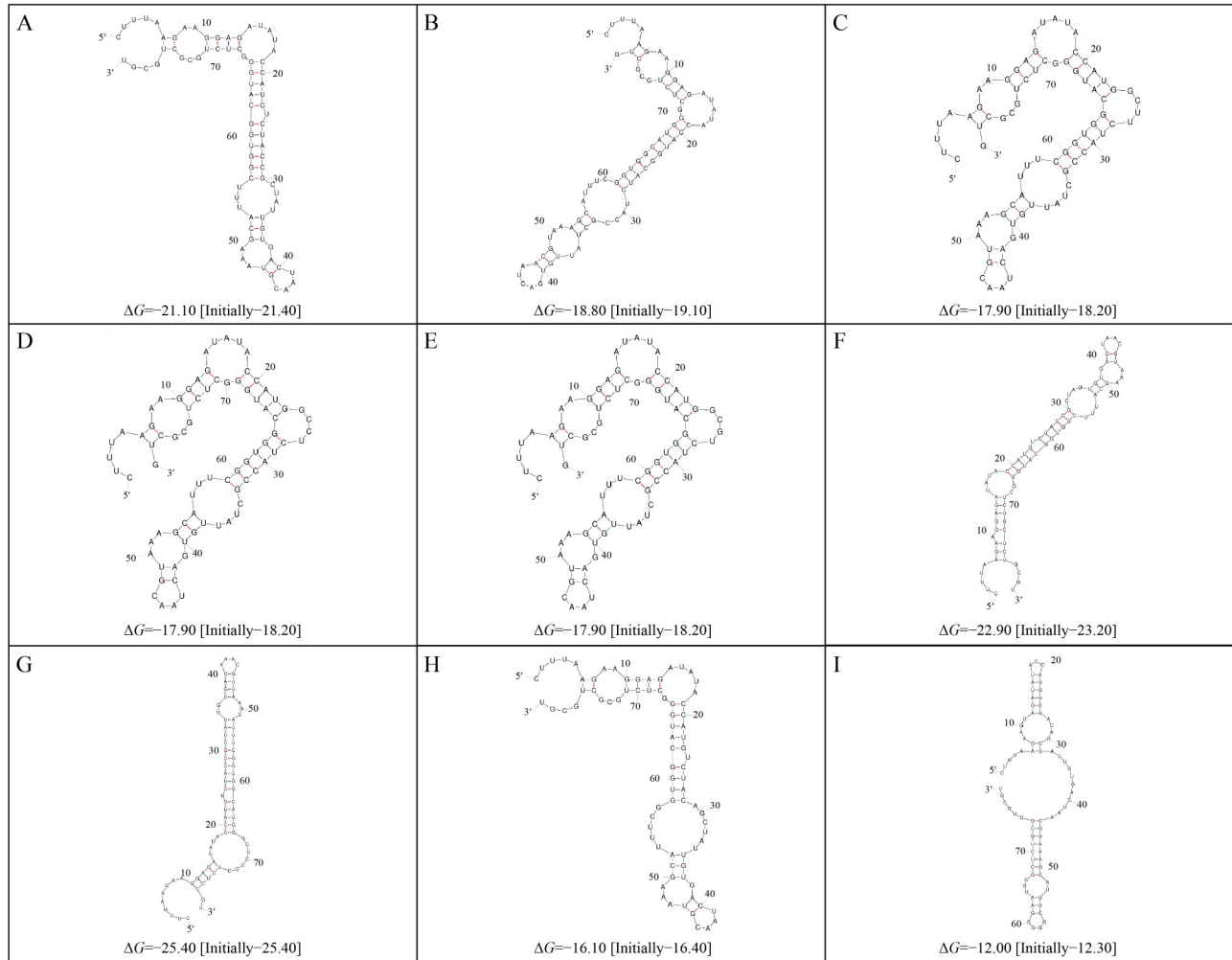


图5 Mfold mRNA二级结构预测结果

Figure 5 Prediction of mRNA structure by Mfold. A: HheC_{M4}; B: ins2A^{GCA}; C: ins2A^{GCU}; D: ins2A^{GCC}; E: ins2A^{GCG}; F: HheC_{M4/L1}; G: HheC_{M4/L2}; H: HheC_{M4/H1}; I: HheC_{M4/H2}.

ins2A^{GCC} (图 5D)、ins2A^{GCG} (图 5E)三者结构高度相似, 发夹短, 环结构分布更多, 配对密度更低。HheC_{M4/L1} (图 5F)与HheC_{M4/L2} (图 5G)存在连续且紧密的长茎结构, 双链配对密度显著提升。HheC_{M4/H1} (图 5H)虽与HheC_{M4}同样具有较高的相似度, 但配对密度进一步下降, HheC_{M4/H2} (图 5I)则仅在少量碱基间形成较短的配对。

源于UFold的mRNA二级结构预测结果如图6所示。HheC_{M4} (图6A)主链存在连续的茎区双链配对, 而插入突变体(图6B-6E)茎区长度降

低, 主链存在更多环结构。HheC_{M4/L1} (图6F)与HheC_{M4/L2} (图6G)存在更长的茎区和更致密的碱基配对, 且在HheC_{M4/L2}中尤为显著。相比之下, HheC_{M4/H2} (图6I)配对结构少, 茎区结构较短, 发夹未完全闭合, 同时更复杂的长程相互作用表明其存在更大的三维构象自由度。

2.4 热力学性质分析

山形图(mountain plot)能够基于mRNA二级结构的MFE结构自由能、质心结构自由能以及配对(partition function, PF)结构展现序列特定区域的自由能变化。如图7所示, HheC_{M4/L1} (图7F)

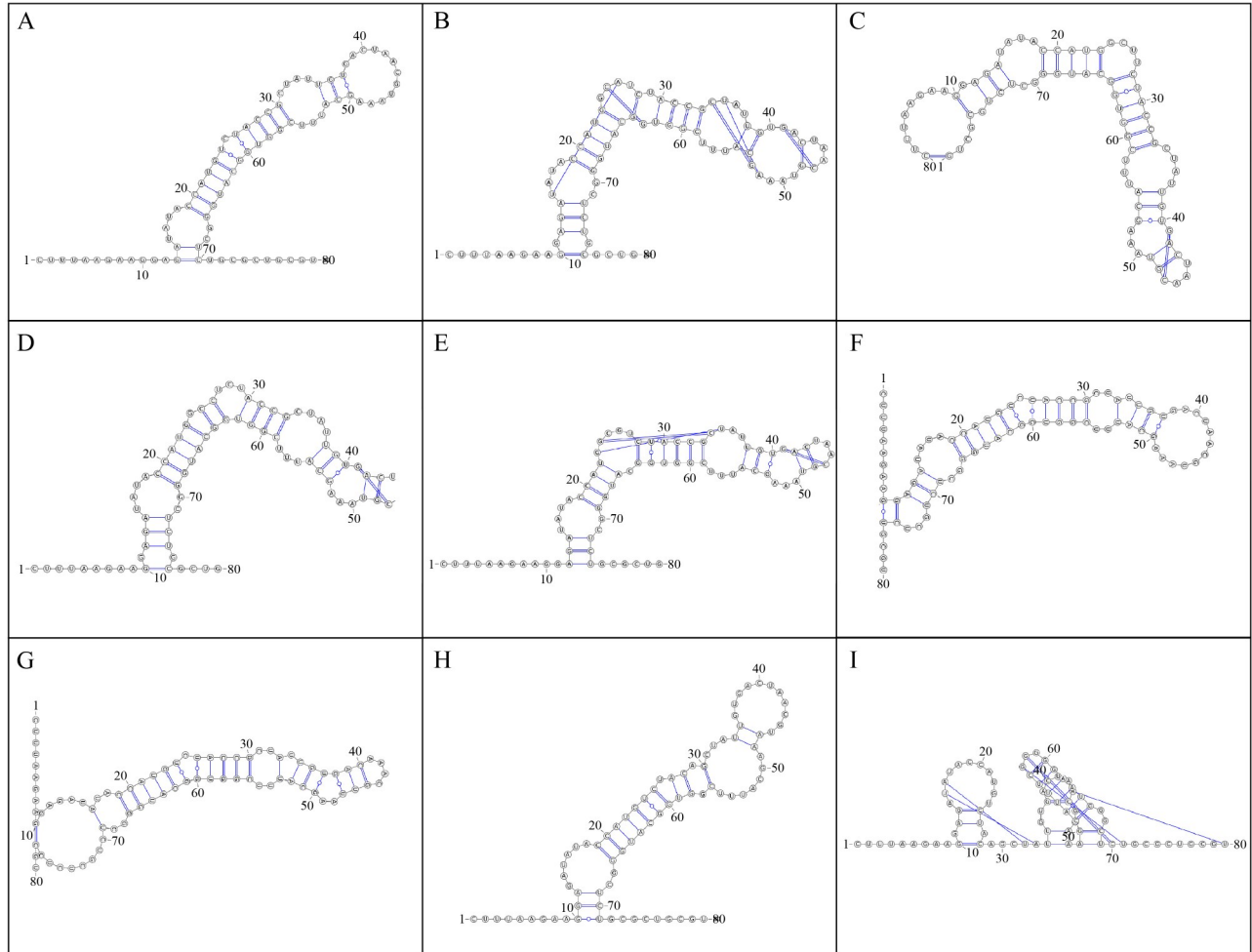


图6 Ufold mRNA二级结构预测结果

Figure 6 Prediction of mRNA structure by Ufold. A: HheC_{M4}; B: ins2A^{GCA}; C: ins2A^{GCU}; D: ins2A^{GCC}; E: ins2A^{GCG}; F: HheC_{M4/L1}; G: HheC_{M4/L2}; H: HheC_{M4/H1}; I: HheC_{M4/H2}.

与 HheC_{M4/L2} (图 7G) 的三条曲线高度相似, 且 HheC_{M4/L2} 的 MFE 与 Centroid 曲线完全重合, 这进一步证明了其 mRNA 二级结构具有高度稳定性。HheC_{M4/H1} (图 7H) 与 HheC_{M4/H2} (图 7I) 相较于 HheC_{M4} (图 7A) 自由能更低, 并且 HheC_{M4/H2} 中存在一个显著的低能量区域, 这或许更有利于 mRNA 二级结构的部分解旋。

熵(entropy)是量化二级结构稳定性变化的一个重要参数, 可用于衡量 mRNA 不同区域的结构可变程度。低熵表示该位置的结构稳定, 构象选择性低, 在大多数情况下均能折叠成特定

结构; 而高熵值则相反, 表明二级结构高度动态, 构象变化较多。通过对各突变体的 mRNA 序列进行熵值分析能够进一步揭示它们之间二级结构的可变性和动态特性。HheC_{M4}、HheC_{M4/L1}、HheC_{M4/L2} 中存在更多的低熵区域, 表明其大部分区域构象变化受限; 而 HheC_{M4/H1}、HheC_{M4/H2} 中熵普遍较高, 表明其二级结构在多个区域存在较大的可变空间。

2.5 高表达 HheC 催化 1,3-DCP 合成 ECH 反应进程

本研究考察了 HheC_{M4} 及其突变体(5 g/L 湿

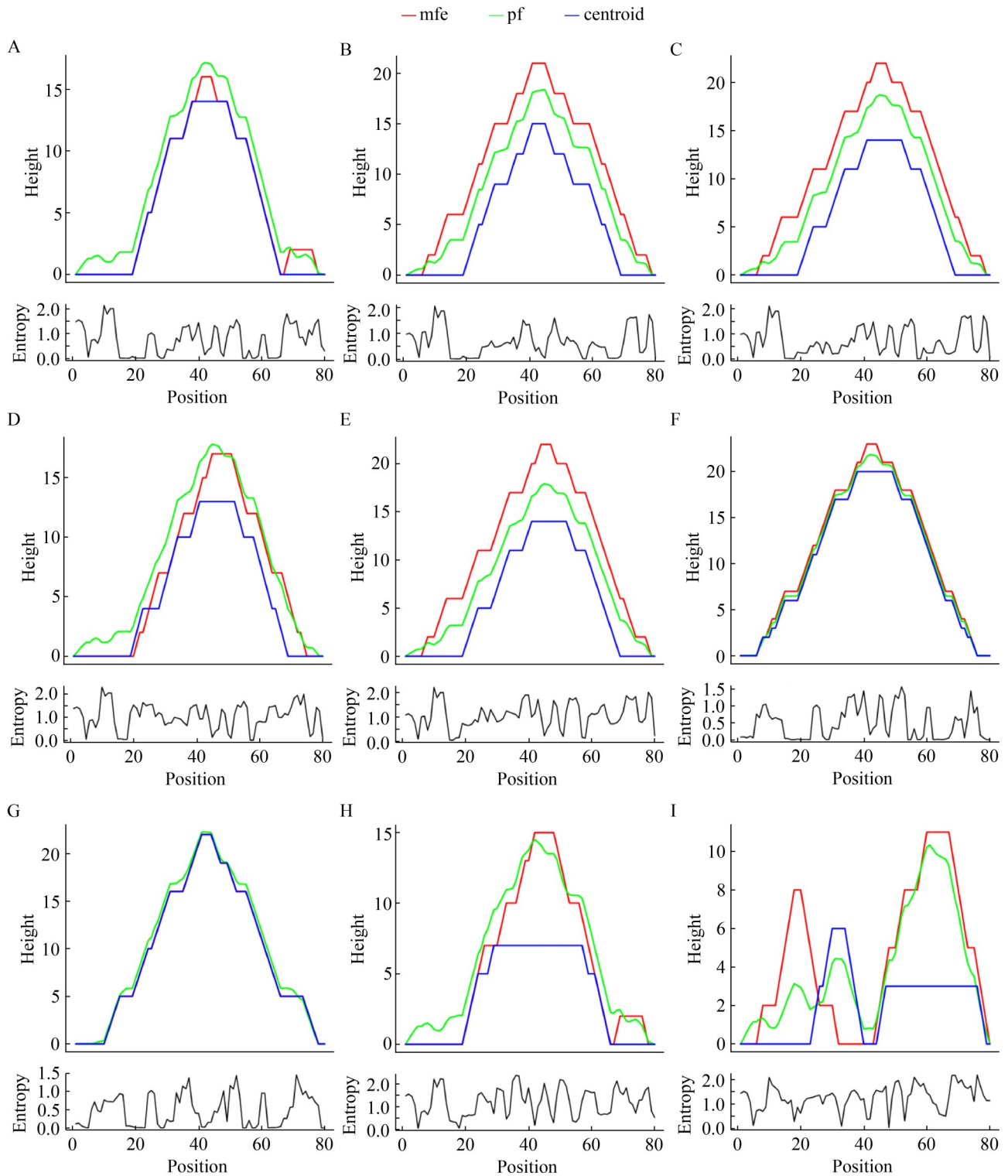


图7 mRNA二级结构山形图

Figure 7 Mountain plot of mRNA secondary structure. A: HheC_{M4}; B: ins2A^{GCA}; C: ins2A^{GCU}; D: ins2A^{GCC}; E: ins2A^{GCG}; F: HheC_{M4/L1}; G: HheC_{M4/L2}; H: HheC_{M4/H1}; I: HheC_{M4/H2}.

细胞)催化 100 mmol/L 底物合成 ECH 的反应进程, 结果如图 8 所示。受限于卤醇脱卤酶的可逆反应, 在该体系下的最大产率仅能达到约 88%。HheC_{M4/H1} 与 HheC_{M4/H2} 反应 5 min 时产率即可分别达到 14.31% 与 17.17%, 且仅需约 50 min 即可达到最大产率; 而出发株 HheC_{M4} 反应 5 min 时产率仅为 7.54%, 达到同一水平产率需将反应时间延长至约 120 min。因此, 在实际生产中本研究的高表达突变体能够有效加速反应进程, 在现有体系下完成底物的高效转化, 缩短生产周期, 为卤醇脱卤酶高效催化 1,3-DCP 合成 ECH 奠定重要基础。

3 讨论与结论

卤醇脱卤酶作为一种能够催化卤代醇与环氧化物相互转化的生物催化剂, 在环境修复与监测^[31-32]、合成药物手性中间体^[3-4]等领域均具有重要的应用价值。另一方面, 原核生物的基因表达受 mRNA 结构的影响。在蛋白质合成过程中, 翻译起始是其中限速且受调控程度最高的阶段^[11], 易受到 mRNA 二级结构稳定性的影响。稳定的 mRNA 二级结构会阻碍核糖体 30S 亚基与 RBS 的结合, 从而抑制翻译起始, 显著影响目的基因的表达^[9,16-17], 而较低的 mRNA 二级结构稳定性有利于两者结合, 进而高效启动翻译^[33-35]。因此, 通过序列优化降低翻译起始区 mRNA 二级结构的稳定性能够有效提高翻译起始效率, 实现外源基因在宿主细胞中的高效表达。

本研究基于卤醇脱卤酶 HheC 编码序列 5'端 mRNA 二级结构与折叠自由能, 借助 mRNA 二级结构预测工具进行计算, 在维持氨基酸序列不变的前提下实现了核苷酸序列的优化。最终, 使 HheC 在 *E. coli* BL21(DE3) 中的表达水平由占宿主细胞可溶性蛋白的 16.71% 提升至 33.39%, 相对酶活提升至原来的 3 倍。此外, 通过负向优化得到的高稳定性 5'mRNA 序列表现出低表达水平, 进一步验证了 HheC 5'mRNA 稳定性与

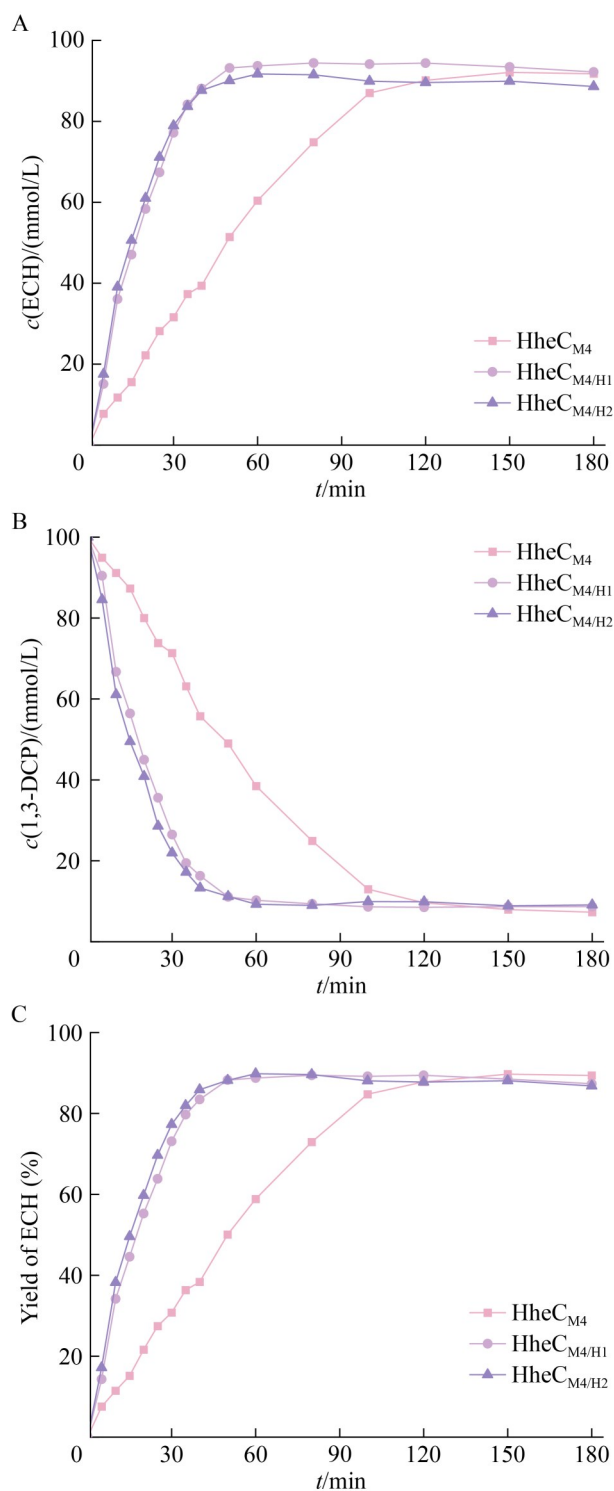


图8 HheC_{M4}及其突变体催化1,3-DCP合成ECH的反应进程

Figure 8 Synthesis of ECH from 1,3-DCP catalyzed by HheC_{M4} and its mutants.

其表达水平的相关性。将所得成果应用于催化 1,3-DCP 合成 ECH, 在本研究体积比为 1:1 的 Gly/NaOH (600 mmol/L, pH 9.8) 与乙酸异丁酯的双相体系下, 高表达 HheC 通过湿细胞催化 100 mmol/L 1,3-DCP 合成 ECH 达到产率上限的反应时间由 120 min 缩短至 50 min。HheC 的高效表达不仅能够提高目标酶的产量, 使微量活性差异更易被检测与量化, 从而提升蛋白质工程中突变筛选的灵敏度和效率; 而且对于生产应用而言, 更易于实现反应体系的高效催化, 显著提高经济效益, 降低生产成本, 增强酶催化工艺路线的竞争力。

作者贡献声明

朱钧豪: 参与设计研究, 开展实验与数据分析, 撰写论文; 汤晓玲: 整体实验设计及文章修改; 郑仁朝: 设计、指导本研究。

作者利益冲突公开声明

作者声明不存在任何可能会影响本文所报告工作的已知经济利益或个人关系。

参考文献

- [1] Van HYLCKAMA VLIEG JET, TANG LX, LUTJE SPELBERG JH, SMILDA T, POELARENDS GJ, BOSMA T, van MERODE AEJ, FRAAIJE MW, JANSSEN DB. Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases[J]. *Journal of Bacteriology*, 2001, 183(17): 5058-5066.
- [2] De JONG RM, TIESINGA JJW, ROZEBOOM HJ, KALK KH, TANG LX, JANSSEN DB, DIJKSTRA BW. Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site[J]. *The EMBO Journal*, 2003, 22(19): 4933-4944.
- [3] MA SK, GRUBER J, DAVIS C, NEWMAN L, GRAY D, WANG A, GRATE J, HUISMAN GW, SHELDON RA. A green-by-design biocatalytic process for atorvastatin intermediate[J]. *Green Chemistry*, 2010, 12(1): 81-86.
- [4] YOU ZY, LIU ZQ, ZHENG YG. Properties and biotechnological applications of halohydrin dehalogenases: current state and future perspectives[J]. *Applied Microbiology and Biotechnology*, 2013, 97(1): 9-21.
- [5] BUCHAN JR, STANSFIELD I. Halting a cellular production line: responses to ribosomal pausing during translation[J]. *Biology of the Cell*, 2007, 99(9): 475-487.
- [6] TULLER T, VEKSLER-LUBLINSKY I, GAZIT N, KUPIEC M, RUPPIN E, ZIV-UKELSON M. Composite effects of gene determinants on the translation speed and density of ribosomes[J]. *Genome Biology*, 2011, 12(11): R110.
- [7] TULLER T, CARMİ A, VESTSIGIAN K, NAVON S, DORFAN Y, ZABORSKE J, PAN T, DAHAN O, FURMAN I, PILPEL Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation[J]. *Cell*, 2010, 141(2): 344-354.
- [8] CHEN CL, ZHANG HB, BROITMAN SL, REICHE M, FARRELL I, COOPERMAN BS, GOLDMAN YE. Dynamics of translation by single ribosomes through mRNA secondary structures[J]. *Nature Structural & Molecular Biology*, 2013, 20(5): 582-588.
- [9] MAO YH, LIU HL, LIU YL, TAO SH. Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *Saccharomyces cerevisiae*[J]. *Nucleic Acids Research*, 2014, 42(8): 4813-4822.
- [10] SIPLEY J, GOLDMAN E. Increased ribosomal accuracy increases a programmed translational frameshift in *Escherichia coli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1993, 90(6): 2315-2319.
- [11] LAURSEN BS, SØRENSEN HP, MORTENSEN KK, SPERLING-PETERSEN HU. Initiation of protein synthesis in bacteria[J]. *Microbiology and Molecular Biology Reviews*, 2005, 69(1): 101-123.
- [12] KUDLA G, MURRAY AW, TOLLERVEY D, PLOTKIN JB. Coding-sequence determinants of gene expression in *Escherichia coli*[J]. *Science*, 2009, 324(5924): 255-258.
- [13] LENZ G, DORON-FAIGENBOIM A, RON EZ, TULLER T, GOPHNA U. Sequence features of *E. coli* mRNAs affect their degradation[J]. *PLoS One*, 2011, 6(12): e28544.
- [14] STUDER SM, JOSEPH S. Unfolding of mRNA secondary structure by the bacterial translation initiation complex[J]. *Molecular Cell*, 2006, 22(1): 105-115.
- [15] De SMIT MH, van DUIN J. Translational initiation on structured messengers another role for the Shine-Dalgarno interaction[J]. *Journal of Molecular Biology*, 1994, 235(1): 173-184.
- [16] CHIARUTTINI C, GUILLIER M. On the role of mRNA secondary structure in bacterial translation[J]. *Wiley Interdisciplinary Reviews RNA*, 2020, 11(3): e1579.
- [17] TULLER T, ZUR H. Multiple roles of the coding sequence 5' end in gene expression regulation[J]. *Nucleic Acids Research*, 2015, 43(1): 13-28.
- [18] BARRINGTON CL, GALINDO G, KOCH AL, HORTON ER, MORRISON EJ, TISA S, STASEVICH TJ, RISSLAND OS. Synonymous codon usage regulates translation initiation[J]. *Cell Reports*, 2023, 42(12): 113413.
- [19] CHAKARBORTY S, IRSHAD IU, MAHIMA, SHARMA AK. TIR predictor and optimizer: web-tools for accurate prediction of translation initiation rate and

- precision gene design in *Saccharomyces cerevisiae*[J]. *Biotechnology Journal*, 2024, 19(5): 2400081.
- [20] GRUBER AR, LORENZ R, BERNHART SH, NEUBOCK R, HOFACKER IL. The Vienna RNA websuite[J]. *Nucleic Acids Research*, 2008, 36: W70-W74.
- [21] LORENZ R, BERNHART SH, HÖNER ZU SIEDERDISSEN C, TAHER H, FLAMM C, STADLER PF, HOFACKER IL. ViennaRNA package 2.0[J]. *Algorithms for Molecular Biology*, 2011, 6: 26.
- [22] MATHEWS DH, DISNEY MD, CHILDS JL, SCHROEDER SJ, ZUKER M, TURNER DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(19): 7287-7292.
- [23] ZUKER M. Mfold web server for nucleic acid folding and hybridization prediction[J]. *Nucleic Acids Research*, 2003, 31(13): 3406-3415.
- [24] ZUKER M, STIEGLER P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information[J]. *Nucleic Acids Research*, 1981, 9(1): 133-148.
- [25] MATHEWS DH, SABINA J, ZUKER M, TURNER DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure[J]. *Journal of Molecular Biology*, 1999, 288(5): 911-940.
- [26] FU LY, CAO YX, WU J, PENG QK, NIE Q, XIE XH. Ufold: fast and accurate RNA secondary structure prediction with deep learning[J]. *Nucleic Acids Research*, 2022, 50(3): e14.
- [27] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015: 234-241.
- [28] STERNBERG SR. Biomedical image processing[J]. *Computer*, 1983, 16(1): 22-34.
- [29] GOODMAN DB, CHURCH GM, KOSURI S. Causes and effects of N-terminal codon bias in bacterial genes[J]. *Science*, 2013, 342(6157): 475-479.
- [30] DING Y, CHAN CY, LAWRENCE CE. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble[J]. *RNA*, 2005, 11(8): 1157-1166.
- [31] GUL I, WANG Q, JIANG QF, FANG RQ, TANG LX. Enzyme immobilization on glass fiber membrane for detection of halogenated compounds[J]. *Analytical Biochemistry*, 2020, 609: 113971.
- [32] GUL I, LE W, JIE Z, FANG RQ, BILAL M, TANG LX. Recent advances on engineered enzyme-conjugated biosensing modalities and devices for halogenated compounds[J]. *TrAC Trends in Analytical Chemistry*, 2021, 134: 116145.
- [33] YANG JR, CHEN XS, ZHANG JZ. Codon-by-codon modulation of translational speed and accuracy via mRNA folding[J]. *PLoS Biology*, 2014, 12(7): e1001910.
- [34] WU HW, PEI JJ, WU GG, SHAO WL. Overexpression of GH10 endoxylanase XynB from *Thermotoga maritima* in *Escherichia coli* by a novel vector with potential for industrial application[J]. *Enzyme and Microbial Technology*, 2008, 42(3): 230-234.
- [35] DUAN JB, WAINWRIGHT MS, COMERON JM, SAITOU N, SANDERS AR, GELERNTER J, GEJMAN PV. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor[J]. *Human Molecular Genetics*, 2003, 12(3): 205-216.