

为面向生物场景，逆向折叠更名为**序列设计**。
同时对可选参数进行了更新，对其选择加以解释，便于用户理解使用。

输入

解码参数

解码策略 (Decoding Strategy)

☒ 贪婪搜索 (Greedy)

☐ 波束搜索 (Beam)

☐ 温度采样 (Sample)

每一步贪婪选择当前概率最高的氨基酸，是模型压缩知识的局部最优。**只能返回一条结果。**

策略对比

质量: 波束搜索 > 贪婪搜索 > 温度采样
速度: 贪婪搜索 > 温度采样 ≈ 波束搜索
多样性: 温度采样 >> 波束搜索 ≈ 贪婪搜索 = 0

特性	贪婪搜索 (Greedy)	波束搜索 (Beam)	温度采样 (Temperature)
输出确定性	✔ 确定	✔ 确定	✘ 不确定，随机性可调
多样性	✘ 无	✘ 较低	✔ 可高，取决于温度
句子质量	✘ 容易局部最优	✔ 接近全局最优	取决于随机性，可能好也可能差
算法复杂度	低	中等 (k 倍)	低 (采样本身简单)

运行序列设计

清空

输入

解码参数

解码策略 (Decoding Strategy)

☐ 贪婪搜索 (Greedy)

☒ 波束搜索 (Beam)

☐ 温度采样 (Sample)

同时保留 N 条最优候选路径（即束宽），探索更广的搜索空间，结果更接近全局最优。

束宽 (Beam Width)

10

返回条数

5

返回条数不能超过束宽。

输入

解码参数

解码策略 (Decoding Strategy)

☐ 贪婪搜索 (Greedy)

☐ 波束搜索 (Beam)

☒ 温度采样 (Sample)

按概率分布进行随机采样，并通过“温度因子T”控制随机程度，能够生成更多样化的序列。

温度 (Temperature)

0.0

T=0 等效于贪婪搜索。T越大，随机性越强，多样性越高。

返回条数

5

Protein structure





...ANRKVSLQRY
REKRKDRKFSK
AKKEKRKEKS...

蛋白质序列设计

V1.0

模型结合实验柔性信息和先进生成技术，高效设计能自发折叠实现功能的可合成氨基酸序列，将结构设想快速转化为实际分子，全面提升效率与精度。

蛋白质

序列设计

逆折叠

密码子优化的可选参数也进行了更新，对其选择加以解释，便于用户理解使用。

HiDimension

NevoBiotech

密码子优化 - 功能详情

← 回到功能目录

密码子优化

通过智能化的CDS序列重设计，提高在不同底盘细胞中的表达效率。一条序列的优化仅需分钟级别，可助您快速得到结果。

输入

可选参数

1. 选择底盘细胞

HEK293T

HEK293F

E.coli

2. 选择生成策略

贪婪搜索 (Greedy)

波束搜索 (Beam)

温度采样 (Sample)

每一步贪婪选择当前概率最高的氨基酸，是模型压缩知识的局部最优。只能返回一条结果。

3. 勾选要屏蔽的酶切位点

☐ BamH-I

☐ EcoR-I

☐ Hind-III

☐ Kpn-I

☐ Nco-I

☐ Nde-I

☐ Nhe-I

☐ Not-I

☐ Dpn-I

☐ Sac-I

☐ Sal-I

☐ Sma-I

☐ Xba-I

☐ Xho-I

策略对比

质量: 波束搜索 > 贪婪搜索 > 温度采样

速度: 贪婪搜索 > 温度采样 ≈ 波束搜索

多样性: 温度采样 >> 波束搜索 ≈ 贪婪搜索 = 0

特性	贪婪搜索 (Greedy)	波束搜索 (Beam)	温度采样 (Temperature)
输出确定性	✔ 确定	✔ 确定	✘ 不确定，随机性可调
多样性	✘ 无	✘ 较低	✔ 可高，取决于温度
句子质量	✘ 容易局部最优	✔ 接近全局最优	取决于随机性，可能好也可能差
算法复杂度	低	中等 (k 倍)	低 (采样本身简单)

运行密码子优化

清空

能力介绍

使用示例

能力介绍

密码子优化旨在在不改变蛋白质氨基酸序列的前提下，调整DNA序列中密码子的使用方式，以更契合宿主的偏好。这一过程是合成生物学与蛋白工程中的关键步骤，为功能验证、产业化生产和新型蛋白设计奠定基础。

性能亮点:

成功在数十个蛋白，3种底盘细胞中提升蛋白表达量，至高提升300+%。一条序列的优化仅需分钟级别，快速得到结果。

模型优势:

传统密码子优化依赖CAI等统计指标，但忽视翻译速度、mRNA结构等因素。蜜维自研的HyperExpression以生成式AI革新范式，直接面向底盘细胞环境学习序列—表达数据，自动生成最优密码子序列，实现更高效全面的蛋白质表达。

操作简便:

用户只需选择目标底盘细胞，并输入蛋白质氨基酸序列，即可调用HyperExpression设计其最优密码子序列。输出为CDS的fasta文件，并提供下载。

引用文献

Zhao, Heng-Rui, et al. "Integration of Protein and Coding Sequences Enables Mutual Augmentation of the Language Model." bioRxiv, 2024, <https://doi.org/10.1101/2024.10.24.620004>

输入

可选参数

1. 选择底盘细胞

HEK293T

HEK293F

E.coli

2. 选择生成策略

贪婪搜索 (Greedy)

波束搜索 (Beam)

温度采样 (Sample)

同时保留 N 条最优候选路径（即束宽），探索更广的搜索空间，结果更接近全局最优。

束宽 (Beam Width)

10

返回条数

5

返回条数不能超过束宽。

3. 勾选要屏蔽的酶切位点

☐ BamH-I

☐ EcoR-I

☐ Hind-III

☐ Kpn-I

☐ Nco-I

☐ Nde-I

☐ Nhe-I

☐ Not-I

☐ Dpn-I

☐ Sac-I

☐ Sal-I

☐ Sma-I

☐ Xba-I

☐ Xho-I

策略对比

质量: 波束搜索 > 贪婪搜索 > 温度采样

速度: 贪婪搜索 > 温度采样 ≈ 波束搜索

多样性: 温度采样 >> 波束搜索 ≈ 贪婪搜索 = 0

特性	贪婪搜索 (Greedy)	波束搜索 (Beam)	温度采样 (Temperature)
输出确定性	✔ 确定	✔ 确定	✘ 不确定，随机性可调
多样性	✘ 无	✘ 较低	✔ 可高，取决于温度
句子质量	✘ 容易局部最优	✔ 接近全局最优	取决于随机性，可能好也可能差
算法复杂度	低	中等 (k 倍)	低 (采样本身简单)

输入

可选参数

1. 选择底盘细胞

HEK293T

HEK293F

E.coli

2. 选择生成策略

贪婪搜索 (Greedy)

波束搜索 (Beam)

温度采样 (Sample)

按概率分布进行随机采样，并通过“温度因子T”控制随机程度，能够生成更多样化的序列。

温度 (Temperature)

0.0

T=0 等效于贪婪搜索。T越大，随机性越强，多样性越高。

返回条数

5

3. 勾选要屏蔽的酶切位点

☐ BamH-I

☐ EcoR-I

☐ Hind-III

☐ Kpn-I

☐ Nco-I

☐ Nde-I

☐ Nhe-I

☐ Not-I

☐ Dpn-I

☐ Sac-I

☐ Sal-I

☐ Sma-I

☐ Xba-I

☐ Xho-I

策略对比

质量: 波束搜索 > 贪婪搜索 > 温度采样

速度: 贪婪搜索 > 温度采样 ≈ 波束搜索

多样性: 温度采样 >> 波束搜索 ≈ 贪婪搜索 = 0

特性	贪婪搜索 (Greedy)	波束搜索 (Beam)	温度采样 (Temperature)
输出确定性	✔ 确定	✔ 确定	✘ 不确定，随机性可调
多样性	✘ 无	✘ 较低	✔ 可高，取决于温度
句子质量	✘ 容易局部最优	✔ 接近全局最优	取决于随机性，可能好也可能差
算法复杂度	低	中等 (k 倍)	低 (采样本身简单)

RefSeq的S应该大写，更正了低同源检索中的相应内容

HiDimension

NovoBiotech

蛋白质低同源检索 - 功能详情

登录

← 回到功能目录

蛋白质低同源检索

基于 BiooSeek0.5 模型，高效完成低同源场景下的快速检索与比对任务。

输入

可选参数

输入蛋白质FASTA序列

MQYKLILNGKTLKGTTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

或上传FASTA文件

选择文件

未选择任何文件

运行检索

清空

能力介绍

使用示例

能力介绍

本功能旨在突破传统同源比对的局限性，尤其是在蛋白质序列相似度较低时，能够精准捕获其潜在的功能或结构关联。我们采用先进的蛋白质向量化表征与深度学习模型，以实现高效、精准的检索。

核心模型:

依托自主研发的 BiooSeek0.5 模型与 RefSeq、UniProt 向量数据库。

高效检索:

在海量蛋白质库中实现近乎实时的搜索与比对。

操作简便:

只需输入氨基酸序列，即可运行任务并获得可供下载分析的表格化结果 (.csv)。

引用文献

Zhao, Heng-Rui, et al. "Integration of Protein and Coding Sequences Enables Mutual Augmentation of the Language Model." *bioRxiv*, 2024, <https://doi.org/10.1101/2024.10.24.620004>

(4)

.

Protein sequence

GIVEQCCTSIC
SLYQLENYCN...

RefSeq

打破同源束缚，大幅扩展可检索空间与效率，可发现序列相似性低但结构同源性高的蛋白质。此模型曾成功挖掘出世界目前已知最小长度且具有活性的TnpB。

蛋白质

检索

向量数据库

Embedding

V0.5

- 蛋白质热稳定性优化中，对结果展示部分新增序列对比，帮助用户快速定位优化前后的序列差异
- 并且提示，热稳定性优化可能对蛋白质活性造成影响。

结果预览

优化已完成。下方展示优化后的氨基酸序列，可在比对视图中快速定位突变/固定位点。

提示：提高热稳定性可能伴随蛋白质活性下降。请在比对视图中重点核查催化残基、配位或结合位点等关键氨基酸，剔除涉及关键位点的候选序列。

突变比对

FASTA原文

下载结果 (.fasta)

```
>optimized_example_Tm
MFKKYSSLENHYNYSKFIEKLYSLGLTKGEWVAREKIHGTNFSLIIERDKVTC AKRTGPIL
PAEDFFGYEIIILKNYADSIKAVQDIMETSAVVS YQVFGEFAGPGIQKNVDYCDKDFYVFD
IIVTTESGDVTYVDDHMMESFCNTFKFKMAPLLGRGKFEELIKLPNGLDSVVQDYNFTVD
HAGQVDANKCVWNAEAKGEVFTAKGYVLKPCYPSWLRNGNRVAIKCKNSKFSEKKKSDKP
IKAKVYLSEADNKLVGILACYVT LNRVNNVISKIGEIGPKDFGKVMGLTVQDILEETSRE
GITLTQADNP SLIKKELVKMVQDVLRPAWIELVS
```

结果预览

优化已完成。下方展示优化后的氨基酸序列，可在比对视图中快速定位突变/固定位点。

提示：提高热稳定性可能伴随蛋白质活性下降。请在比对视图中重点核查催化残基、配位或结合位点等关键氨基酸，剔除涉及关键位点的候选序列。

突变比对

FASTA原文

optimized_example_Tm						
1	10	20	30	40	50	60
Orig	MFKKYSSLENHYNYSKFIEKLYSLGLT	GGEWVAREKI	IGTNFSLIIERDKVTC	AKRTGPIL		
Opti	MFKKYSSLENHYNYSKFIEKLYSLGLT	KGEWVAREKI	HGTNFSLIIERDKVTC	AKRTGPIL		
61	70	80	90	100	110	120
Orig	PAEDFFGYEIIILKNYADSIKAVQDIMETS	SAVVS	YQVFGEFAGPGIQKNVDYCDKDFYVFD			
Opti	PAEDFFGYEIIILKNYADSIKAVQDIMETS	SAVVS	YQVFGEFAGPGIQKNVDYCDKDFYVFD			
121	130	140	150	160	170	180
Orig	IIVTTESGDVTYVDD	YMMESFCNTFKFKMAPLLGRGKFEELIKLPND	LDSVVQDYNFTVD			
Opti	IIVTTESGDVTYVDD	HMMESFCNTFKFKMAPLLGRGKFEELIKLPNG	LDSVVQDYNFTVD			
181	190	200	210	220	230	240
Orig	HAGLV	DANKCVWNAEAKGEVFTA	EGYVLKPCYPSWLRNGNRVAIKCKNSKFSEKKKSDKP			
Opti	HAGQV	DANKCVWNAEAKGEVFTA	KGYVLKPCYPSWLRNGNRVAIKCKNSKFSEKKKSDKP			
241	250	260	270	280	290	300
Orig	IKAKV	ELSEADNKLVLGILACYVT	LNRVNNVISKIGEIGPKDFGKVMGLTVQDILEETSRE			
Opti	IKAKV	YELSEADNKLVLGILACYVT	LNRVNNVISKIGEIGPKDFGKVMGLTVQDILEETSRE			
301	310	320	330			
Orig	GITLTQADNP SLIKKELVKMVQDVL RP AWIELVS					
Opti	GITLTQADNP SLIKKELVKMVQDVL RP AWIELVS					

- 蛋白质结构预测中新增结果可视化展示（目前仅展示固定的示例结果PDB文件）
- 可选择展示颜色和颜色。可拖动缩放查看

结果预览

使用3Dmol.js

结构预测完成！您可以在下方进行三维结构预览，或下载PDB文件用于本地分析。

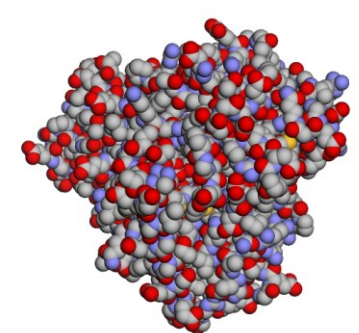
样式: 飘带 球 棍 线 颜色: 彩虹 链 下载结果 (.pdb)



结果预览

结构预测完成！您可以在下方进行三维结构预览，或下载PDB文件用于本地分析。

样式: 飘带 球 棍 线 颜色: 彩虹 链 下载结果 (.pdb)



结果预览

结构预测完成！您可以在下方进行三维结构预览，或下载PDB文件用于本地分析。

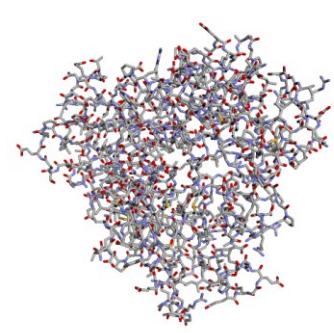
样式: 飘带 球 棍 线 颜色: 彩虹 链 下载结果 (.pdb)



结果预览

结构预测完成！您可以在下方进行三维结构预览，或下载PDB文件用于本地分析。

样式: 飘带 球 棍 线 颜色: 彩虹 链 下载结果 (.pdb)



结果预览

结构预测完成！您可以在下方进行三维结构预览，或下载PDB文件用于本地分析。

样式: 飘带 球 棍 线 颜色: 彩虹 链 下载结果 (.pdb)



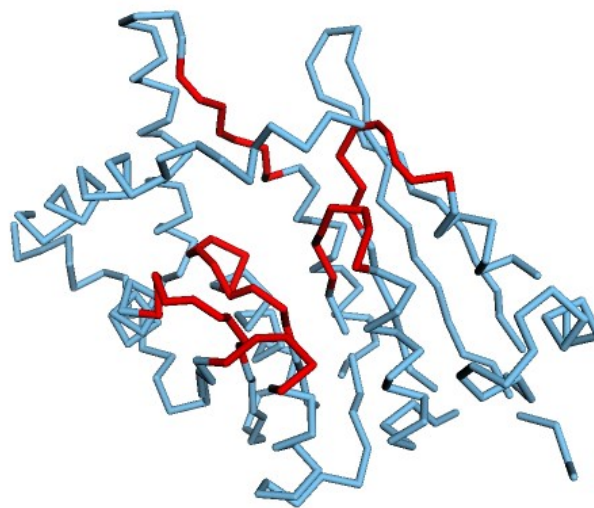
- 蛋白质结构补齐中新增结果可视化展示（目前仅展示固定的示例结果PDB文件）。
- 默认颜色和样式，原先部分淡蓝色，补齐部分则以红色高亮显示。
- 目前版本只生成碳骨架，所以使用球棍（stick）模型展示。后续加入氧原子后可增加cartoon等展示模式。
- 可拖动缩放查看

结果预览

结构补齐完成！下方是补齐后的三维结构（基于示例文件）。原有部分为浅蓝色，新增片段已用红色高亮标出。您可以旋转/缩放视图进行详细查看，也可以下载补齐后的PDB文件。

提示：目前版本只生成了碳骨架，不包含氧原子等，因此以球棍模型展示。未来版本中我们可对全原子进行生成，并加入更多样式选择。

补齐后结构



下载结果 (.pdb)

使用3Dmol.js