

Greedy

Decoding Loop

Sequence

ESM-3 Model to Predict

FoldX DDG

acts as reward

Original Wild-Type Sequence: For e.g.: 1PGA.pdb
MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

Fold

RepairPDB to create FoldX ready, parse Wild-Type sequence for baseline

Starting Mutations = { 10: "F", 25: "Y", 30: "W", 41: "I", 55: "R" }

Mutated Sequence:
MTYKLILNGFTLKGETTTEAVDAAYAEKVWKQYANDNGVDIEWTYDDATKTFTVRE

Masking %: User defined (e.g., 0.4%), Maskable Positions: 56 - 5 (frozen) = 51 for design

Number of Masks: 51 * 0.4 ≈ 20 Randomly selects 20 positions for '_' mask.

Generation Step 1

Generation Step 2

Generation Step 1

Generation Step 1

Generation Step 2

Candidate n

ESMFold Meta 98B
650M-15B Progen2
Salesforce
6B
Meta 93M
Rostlab
TAPE 420M
Berkeley
92M
UniRep
Harvard
18M
2017 2018 2019 2020 2021 2022 2023 2024

Year

- The optimization algorithm uses proportional schedule. It is adaptive, efficient, and robust.
- Guided Generation Step Total run step
- Number of Candidate Denoised Prediction in each step.
- Number of Position to Unmask = max(1, masked_position // remaining_steps)
- The proportional schedule is designed to finish unmasking right at last step.

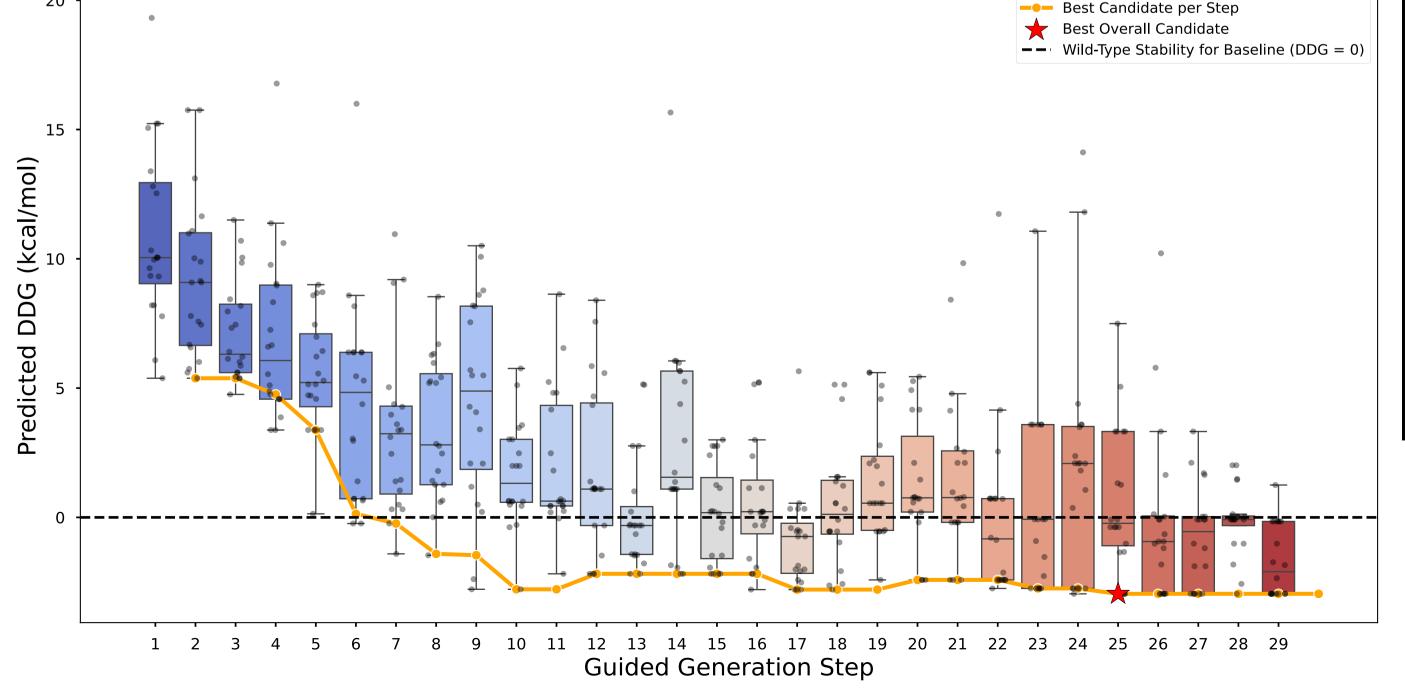
ESM3-FoldX Guided Generation using DDG scores

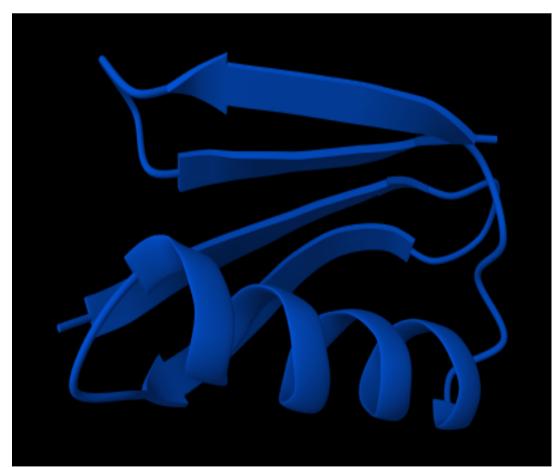
w.r.t WT in parallel

Candidate 2

FoldX calculates DDG scores for each candidate

Candidate



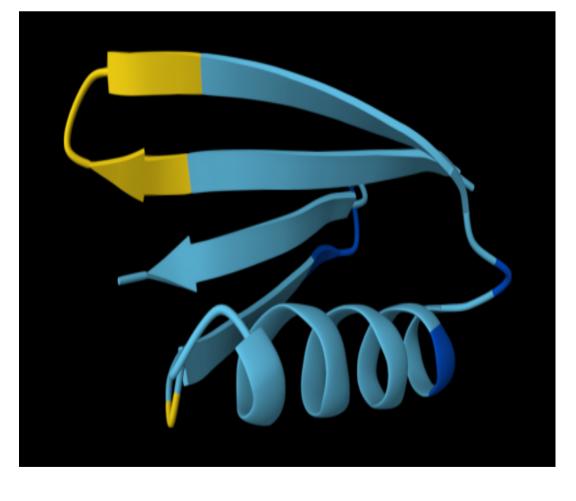


ESM generates n partial sequence

Candidate with best DDG selected

for next step

Original structure of 1PGA



Best Optimized structure using Guided-Generation Framework