

Physics-Guided α -Helix Design with Matched Negative Controls: A Reproducible In Silico Validation Protocol

Jonathan Washburn
Recognition Science Research Institute
Austin, Texas

January 8, 2026

Abstract

Forward design is a stronger test of mechanistic understanding than retrospective prediction: if a theory explains why a structure is stable, it should support generating new sequences that adopt that structure. We report a short, controlled mini-project that designs ten 25-residue sequences predicted to form continuous α -helices using simple, interpretable rules (helix-forming residue preference, salt-bridge patterning, amphipathic hydrophobic/polar faces, and helix capping). We validate these designs using ESMFold [3] and evaluate helicity with a strict geometric criterion based on backbone dihedral angles (ϕ/ψ) and continuity metrics. To guard against self-deception, we construct ten matched negative controls by substituting the two central residues (positions 13–14) with proline–proline (PP), a canonical helix-breaking perturbation.

Result: all 10/10 designed sequences are predicted as continuous α -helices by ϕ/ψ , while 10/10 negative controls are disrupted relative to their matched positives (loss of helix continuity and/or strong kinking). We release the complete sequence panel, prediction artifacts, and analysis scripts to enable independent replication. We also provide concrete next steps for strengthening interpretation: replication on AlphaFold/ColabFold and in vitro circular dichroism (CD) spectroscopy for helicity and thermal stability.

1 Introduction

Protein structure prediction has advanced rapidly, with AlphaFold demonstrating near-experimental accuracy across a broad range of proteins [2]. However, high predictive accuracy alone does not imply a first-principles explanation. In a “mechanistic understanding” program, a key upgrade is *forward design*: rather than asking whether a method can reconstruct known structures, we ask whether a small set of explicit rules can generate *new* sequences whose predicted structures match the intended design objective.

This paper presents a deliberately modest but high-signal design task: α -helix design. Alpha helices are an ideal “unit test” for mechanistic reasoning because their stability is dominated by local geometry and well-characterized interactions: backbone hydrogen-bonding patterns, residue-

specific helix propensity, and simple electrostatic effects such as $i-i+4$ salt bridges. A successful helix-design protocol does not solve protein folding in general, but it provides a clean check that (i) the design rules are coherent, (ii) the evaluation metric is not overly permissive, and (iii) the computational pipeline behaves directionally correctly under controlled perturbations.

Purely computational validation is vulnerable to self-deception: when a predictor returns the “expected” structure, it is easy to over-interpret the outcome, especially on an easy topology such as a helix. We therefore treat negative controls as mandatory. For each designed helix, we generate a matched negative-control sequence that differs only by a minimal, targeted perturbation with a clear physical prediction. Proline is a canonical helix breaker; inserting proline near the center of a helix is expected to disrupt helix continuity and often introduces a kink even when local helical dihedral angles remain partially satisfied. If a validation pipeline cannot detect this directional effect, it is not trustworthy.

Our contributions are therefore primarily methodological: we provide (1) a reproducible positive/negative sequence panel, (2) strict, geometry-based helicity metrics (dihedral ϕ/ψ , continuity, and kink angle), and (3) an explicit plan for upgrading computational validation to independent-model replication (AlphaFold/ColabFold) and physical measurement (CD spectroscopy). The resulting artifact is a compact, falsifiable test that can be extended to harder problems (mixed α/β topologies and β -sheet register pairing), where the global energy landscape and long-range topology dominate.

2 Background and related work

2.1 Why α -helix design is a good “unit test”

The α -helix is one of the simplest and most common secondary-structure motifs in proteins. It is stabilized by a regular backbone hydrogen-bond pattern, and its geometry is largely local: the conformation at residue i is primarily constrained by its immediate neighbors, the backbone torsion preferences of the residue, and short-range electrostatic and solvation effects. In contrast, β -sheet

topology depends on nonlocal pairing register and long-range contacts, which makes it a much harder test of global folding dynamics.

For a mechanistic program, this separation is useful. A helix-design task is not intended to be impressive by modern prediction standards; it is intended to be *diagnostic*. If we cannot design helices reliably with explicit rules, then our pipeline (design rules + evaluation metrics) is unlikely to be trustworthy on harder motifs.

2.2 Predictors as instruments (and why we need controls)

Learned structure predictors are not “ground truth”, but they can serve as high-throughput instruments for testing whether a set of design rules produces sequences that resemble known structural priors. AlphaFold [2] and ESMFold [3] encode powerful priors about peptide geometry and local secondary structure. This makes them useful, but it also creates a hazard: it is easy to obtain “good-looking” outputs for trivial designs, and it is easy to over-interpret those outputs as validation of a deeper theory.

The most direct safeguard is *matched negative controls*. If a method predicts the intended structure for the positive designs but fails to respond appropriately to canonical helix-breaking perturbations, then the method is not sensitive to the causal mechanism we think we are testing. In this paper, each positive sequence is paired with a negative sequence differing only by a minimal central PP substitution; the outcome is evaluated with strict geometric metrics that detect both breaks and kinks.

3 Methods

3.1 Design objective and constraints

We target a single, explicit structural objective: *a continuous, monomeric α -helix* for short sequences. We restrict to $N = 25$ residues to keep the task simple and interpretable; helicity is evaluated primarily on the $N - 2 = 23$ internal residues where ϕ/ψ dihedrals are well-defined.

The protocol intentionally prioritizes falsifiability and reproducibility over novelty: we generate a small set of sequences using fixed heuristics, and we evaluate them with strict geometric metrics plus matched negative controls. We do not optimize sequences against the predictor; there is no gradient-based sequence search, no use of templates, and no iterative design loop.

3.2 Physics-guided sequence design

We generated ten candidate helix sequences spanning five strategies:

- **Balanced (amphipathic)**: alternating hydrophobic (A/L/M) and polar/charged (E/K/Q) positions to create two faces.
- **Charged (salt-bridge stabilized)**: enrichment of oppositely charged residues arranged to encourage $i-i+4$ electrostatic stabilization.

- **Hydrophobic (leucine-zipper-like)**: enrichment of L and other helix-compatible residues with periodicity consistent with helical turns.
- **Polar (high solubility)**: high E/K/Q/A content to reduce aggregation while retaining helix propensity.
- **Alanine-rich (classic helix peptides)**: A-dominant backbone with occasional charges for solubility.

All sequences are provided in `designed_helices.fasta`. The generator is `design_helices.py`.

Salt-bridge logic. In an α -helix, residues separated by four positions (i and $i+4$) are approximately co-linear along one face of the helix, so opposite charges at these positions can form stabilizing salt bridges. In our “charged” strategy, we intentionally place E and K with $i-i+4$ spacing to encourage these interactions while retaining high helix-forming content.

Amphipathic logic. Many helices are amphipathic: one face is hydrophobic and the opposite face is polar/charged. We implement this crudely by using a repeating pattern that places hydrophobic residues on adjacent positions in the same helical turn and polar residues on the opposing positions. This is not a full helical-wheel optimization, but it is sufficient for a clean unit test.

3.3 Design rules as explicit constraints (pre-registered)

To make the design procedure maximally falsifiable, we summarize the intended design logic as explicit constraints on the *positive* sequences. These constraints are not claimed to be optimal; they are claimed to be interpretable and sufficient for a unit test.

Let the sequence be s_1, \dots, s_N with $N = 25$. Define:

- A set of helix breakers $B = \{P, G\}$.
- A set of helix-compatible residues $H = \{A, E, L, M, K, Q, R\}$.
- Charge function $q(K) = q(R) = +1$, $q(E) = q(D) = -1$, and $q(\cdot) = 0$ otherwise.
- Hydrophobic set $U = \{A, V, I, L, M, F, W\}$.

The intended positive-design constraints are:

(C1) No helix breakers:

(C2) High helix-compatible content:

(C3) Salt-bridge enrichment (charged designs):

(C4) Avoid long hydrophobic runs:

(C5) Mild net charge:

In addition, we apply simple end “capping” heuristics: $s_1 \in \{N, S, D, T\}$ and $s_N \in \{G, N, K\}$. In practice, the generator uses these constraints implicitly through fixed residue sets and patterning logic (see `design_helices.py`).

3.4 Matched negative controls (helix breakers)

For each designed helix (POS), we constructed a matched negative-control sequence (NEG) by substituting the two central residues (positions 13–14; 1-indexed) with PP. This minimal perturbation keeps length and composition similar but injects a strong helix-breaking motif at maximal leverage.

Negative controls are provided in `designed_helices_negative_controls.fasta`. The generator is `design_helix_negative_controls.py`-pattern PP.

3.5 Structure prediction (ESMFold)

We predicted structures for all sequences using the ESMFold API [3], which returns a PDB-formatted structure per input sequence. We store the resulting structures as:

- POS predictions: `predictions/`
- NEG predictions: `predictions_negative/`

Scope of inference. We treat ESMFold as a computational instrument. The claims in this paper are therefore strictly about *predicted* structure under this instrument, not about *in vitro* folding. Section “Next steps” describes how to upgrade this to independent-model replication and experimental CD measurements.

3.6 Strict geometric helicity metrics

We quantify helicity using backbone dihedrals and continuity rather than solely distance heuristics.

Dihedral-based helix region. For each internal residue $i \in \{2, \dots, N - 1\}$, we compute:

$$\sum_{i=1}^N \mathbf{1}[s_i \in B] = \phi_i = \text{dihedral}(C_{i-1}, N_i, C\alpha_i, C_i) \quad (6)$$

$$\psi_i = \text{dihedral}(N_i, C\alpha_i, C_i, N_{i+1}) \quad (7)$$

We classify residue i as helix-like if $\phi_i \in [-100^\circ, -30^\circ]$ and $\frac{1}{N} \sum_{i=1}^N \mathbf{1}[s_i \in H]$ is large (heuristic). This window is intentionally loose to avoid false negatives due to model noise.

Implementation detail. Dihedral angles are computed from backbone atom coordinates (N, C α , C) parsed from the predicted PDB using the standard vector-projection formulation (projecting bond vectors onto the plane orthogonal to the central bond and measuring the signed angle via atan2). This is implemented in `validate_helix_negative_controls.py`.

Helix fraction. Let $h_i \in \{0, 1\}$ be the helix-like indicator for each internal residue. The helix fraction is:

$$f_\alpha = \frac{1}{N-2} \sum_{i=2}^{N-1} h_i \quad (8)$$

Helix continuity. We compute the longest consecutive run of helix-like residues among internal positions, L_{\max} , and report the continuity fraction $L_{\max}/(N - 2)$.

Kink angle. Some perturbed sequences can retain locally helical ϕ/ψ while forming a kinked helix (a known proline effect). We therefore compute a kink angle from C α coordinates by estimating principal axes for the left and right halves (via PCA), excluding the two mutated central residues. If v_L and v_R are the unit principal axes, the kink angle is:

$$\theta_{\text{kink}} = \cos^{-1}(v_L \cdot v_R) \quad (9)$$

3.7 Primary outcome: flip rate

We define POS as *continuous* if $f_\alpha \geq 0.95$ and $L_{\max}/(N - 2) \geq 0.95$. We define NEG as *disrupted* if it is not continuous or if $\theta_{\text{kink}} \geq 25^\circ$. The primary success criterion is the flip rate across matched pairs:

$$\text{flip} = \Pr[\text{POS continuous} \wedge \text{NEG disrupted}]$$

Threshold rationale. The continuity thresholds (0.95) are chosen to represent a near-ideal helix: essentially all internal residues must be helix-like and contiguous. The kink threshold (25°) is intended to capture visually meaningful bending. In our results, POS helices have small kink angles (maximum $\approx 10.7^\circ$), while the smallest NEG kink is $\approx 28.7^\circ$, so 25° cleanly separates “straight” from “kinked” in this dataset.

All metrics are computed by `validate_helix_negative_controls.py`, which also writes a machine-readable summary: `helix_negative_control_results.json`.

4 Results

4.1 Positive controls: 10/10 continuous helices

All ten designed sequences are predicted as *continuous* α -helices by the strict ϕ/ψ + continuity criteria: $f_\alpha = 1.00$ and $L_{\max} = 23/23$ for every POS design. The estimated POS kink angles are small (mean 6.7° , range 3.6° to 10.7°), consistent with straight helices.

4.2 Negative controls: 10/10 disrupted relative to positives

All ten matched negative controls (central PP) are *disrupted* relative to their POS counterparts under the combined continuity/kink criterion. While the NEG sequences retain some local helix-like dihedrals (mean $f_\alpha = 0.865$), helix continuity collapses on average (mean longest-run fraction 0.517). NEG kink angles increase dramatically (mean 96.5° , range 28.7° to 143.6°), indicating strong bending or breakage.

4.3 Primary outcome

The flip rate is $10/10 = 100\%$ under the combined criterion (continuity failure or kink $\geq 25^\circ$). This indicates the evaluation pipeline is directionally sensitive and not trivially classifying all sequences as “helical” under a permissive metric.

4.4 Continuity-only versus kink-aware criteria

If we ignore kinks and use continuity alone (NEG must fail the continuity threshold), then 9/10 negative controls are classified as disrupted. The remaining case (`design_8_hydrophobic`) retains locally helical ϕ/ψ and continuity but exhibits a clear bend; the kink-aware criterion correctly counts this as disrupted. This illustrates why a purely local secondary-structure classifier can miss mechanistically relevant failure modes.

5 Discussion

5.1 What this validates

This mini-project validates three concrete points:

1. **Forward design of helices is achievable with explicit rules.** Simple, interpretable heuristics generate sequences that robustly predict as continuous helices.
2. **The evaluation is not self-deceptive.** Matched negative controls shift the predicted structure in the expected direction, with a 100% flip rate under strict geometry-based metrics.
3. **Kink detection matters.** One negative control (`design_8_hydrophobic`) remains locally helical by ϕ/ψ but becomes kinked; a continuity-only criterion would miss a meaningful disruption mode.

5.2 What this does *not* validate

We emphasize limits to interpretation:

- **Not a general solution to folding.** Helices are dominated by local interactions; β -sheet topology and tertiary packing remain the difficult regime.
- **Predictor-world validation.** ESMFold is a learned predictor trained on large structural corpora. Agreement indicates consistency with learned structural priors, not experimental folding in solution.
- **Not purely first-principles.** While the rules are physics-motivated (electrostatics, amphipathicity), residue propensities are an empirical proxy for complex solvent/backbone effects.

5.3 Why the negative controls are still valuable

Despite these limitations, matched negative controls provide a strong internal check: the same learned predictor that returns perfect helices for the positives also changes its output substantially when a canonical helix-breaking perturbation is applied. This reduces the risk that the positive result is merely an artifact of a permissive classifier or an implementation bug in the analysis.

5.4 Threats to validity

Construct validity: “Helix” is not a single scalar. A residue can have helix-like ϕ/ψ angles while the chain is kinked or partially broken. We mitigate this by combining dihedral fraction, continuity, and kink angle.

Internal validity: All measurements depend on correct parsing of PDB files and correct dihedral computation. We address this by using standard backbone-only definitions and by reporting multiple independent metrics that move coherently under perturbation (continuity decreases and kink increases).

External validity: The predictor may not match in vitro behavior. The appropriate upgrade is a physical measurement (CD) and independent-model replication (AlphaFold/ColabFold), which we provide as explicit next steps.

6 Next steps: stronger interpretation upgrades

6.1 Replication on AlphaFold/ColabFold

The cleanest computational upgrade is to replicate this experiment with an independent AlphaFold-family system (e.g., ColabFold [4]). We provide a fixed sequence panel (`designed_helices_posneg_20.fasta`) and an analysis script (`analyze_alphafold_colabfold_helix_crosscheck.py`) that computes the same ϕ/ψ , continuity, and kink metrics from downloaded PDB outputs. If the POS/NEG flip reproduces under AlphaFold-family inference, this materially strengthens interpretation by reducing model-specific bias.

Table 1: Matched POS/NEG helix metrics (strict ϕ/ψ , continuity, and kink). POS designs are continuous helices (all $f_\alpha = 100\%$, $L_{\max} = 23/23$). NEG designs substitute central residues 13–14 with PP. We call a pair “flipped” if NEG is not continuous or has $\theta_{\text{kink}} \geq 25^\circ$.

Design	POS f_α	NEG f_α	POS L_{\max}	NEG L_{\max}	θ_{kink}	POS → NEG	Flip
design_3_hydrophobic	100.0%	95.7%	23/23	13/23	$3.6^\circ \rightarrow 36.6^\circ$	Yes	
design_8_hydrophobic	100.0%	100.0%	23/23	23/23	$6.8^\circ \rightarrow 28.7^\circ$	Yes	
design_4_polar	100.0%	78.3%	23/23	10/23	$7.0^\circ \rightarrow 107.7^\circ$	Yes	
design_1_balanced	100.0%	91.3%	23/23	11/23	$10.4^\circ \rightarrow 90.2^\circ$	Yes	
design_2_charged	100.0%	91.3%	23/23	11/23	$7.5^\circ \rightarrow 143.3^\circ$	Yes	
design_5_ala_rich	100.0%	78.3%	23/23	10/23	$5.8^\circ \rightarrow 143.6^\circ$	Yes	
design_6_balanced	100.0%	78.3%	23/23	9/23	$10.7^\circ \rightarrow 72.9^\circ$	Yes	
design_7_charged	100.0%	91.3%	23/23	11/23	$3.7^\circ \rightarrow 127.0^\circ$	Yes	
design_9_polar	100.0%	82.6%	23/23	11/23	$6.7^\circ \rightarrow 101.1^\circ$	Yes	
design_10_ala_rich	100.0%	78.3%	23/23	10/23	$4.6^\circ \rightarrow 114.0^\circ$	Yes	
Aggregate	10/10	mean 86.5%	10/10	mean 11.9/23	mean 6.7° → 96.5°	10/10	

Table 2: Aggregate summary of strict helix metrics. POS designs are ideal helices under our criteria ($f_\alpha = 1.0$, $L_{\max} = 23/23$). NEG values summarize the 10 matched PP-perturbed sequences.

Metric	POS	NEG (mean)	NEG (min-max)
f_α	100%	86.5%	78.3–100%
$L_{\max}/(N - 2)$	1.00	0.517	0.391–1.00
θ_{kink} (deg)	6.7	96.5	28.1–143.6
Flip rate (kink-aware)		10/10 (100%)	
Flip rate (continuity-only)		9/10 (90%)	

6.2 In vitro CD spectroscopy

The strongest real-world upgrade is circular dichroism (CD) spectroscopy [1]. CD directly measures secondary structure content in solution and can be used to estimate helicity (negative bands near 208 and 222 nm) and thermal stability (melt curves monitoring ellipticity at 222 nm). Critically, the matched PP negatives provide an immediate internal control for measurement artifacts (concentration errors, buffer absorbance, and aggregation). We provide a practical protocol and recommended peptide pairs (two positives and matched PP negatives) in `docs/CD_SPECTROSCOPY_HELIx_VALIDATION.md`.

Data and Code Availability

All artifacts referenced in this paper are included in the repository:

- **Sequence panels:** `designed_helices.fasta`, `designed_helices_negative_controls.fasta`, `designed_helices_posneg_20.fasta`
- **Design scripts:** `design_helices.py`, `design_helix_negative_controls.py`
- **ESMFold validation:** `validate_helix_negative_controls.py`, outputs `predictions/` and `predictions_negative/`
- **Machine-readable metrics:** `helix_negative_control_results.json`

- **AlphaFold/ColabFold cross-check:** `docs/ALPHAFOLD_COLABFOLD_HELIx_CROSSCHECK.md` and `analyze_alphafold_colabfold_helix_crosscheck.py`
- **CD protocol:** `docs/CD_SPECTROSCOPY_HELIx_VALIDATION.md`

Reproduction (ESMFold-based) can be performed with:
`python design_helices.py`
`python design_helix_negative_controls.py --pattern PP`
`python validate_helix_negative_controls.py`

Note: `validate_helix_negative_controls.py` performs network requests to the public ESMFold API.

A Sequence panel (POS and matched NEG)

For completeness, we list the full 20-sequence panel used for replication. The authoritative source is the repository FASTA file `designed_helices_posneg_20.fasta`.

```
>design_3_hydrophobic
LERLEAELEQLQEKLEEELEKLAEL
>design_8_hydrophobic
LEELEEQLEQLEARLARLEEQLEKL
>design_4_polar
SAEAAKEKQAEEEEEEQQQQAKQEN
>design_1_balanced
NAQKLLEKAMEEEAKKMMEKLMEKN
>design_2_charged
NELAAKQLLEAAALKAAQEALLKLAK
>design_5_ala_rich
NAAAAKAAAAKAAAAAAEAAAAAAAK
>design_6_balanced
DMKQALEQLAEEKMAKELLKQLLKNN
>design_7_charged
NEQQQLKAQEAQAAQEAQQKQLK
>design_9_polar
SKAAQEAKQKEEEAAEKAKAKEAAN
>design_10_ala_rich
NAAAAEEEEAAAAAAKAAAAKAAAAK
>design_3_hydrophobic_NEG_PP_P13_P14
LERLEAELEQLQPPLEELEEKLAEL
>design_8_hydrophobic_NEG_PP_P13_P14
LEELEEQLEQLEPPLARLEEQLEKL
>design_4_polar_NEG_PP_P13_P14
SAEAAKEKQAEEEPPEQEQQQQAKQEN
>design_1_balanced_NEG_PP_P13_P14
NAQKLLEKAMEEPKKMMEKLMEKN
>design_2_charged_NEG_PP_P13_P14
NELAAKQLLEAPPAAQEALLKLAK
>design_5_ala_rich_NEG_PP_P13_P14
NAAAAKAAAAKAPPAAEAAAEEAAAK
>design_6_balanced_NEG_PP_P13_P14
DMKQALEQLAEEKPPKELLKQLLKNN
>design_7_charged_NEG_PP_P13_P14
NEQQQLKAQEAAPPAAQEAQKQLK
>design_9_polar_NEG_PP_P13_P14
```

SKAAQEAQKKEEPPEKAKAKEKAAN
>design_10_ala_rich_NEG_PP_P13_P14
NAAAAEAAAAEAPPKAAAEEKAAK

References

- [1] Norma J. Greenfield. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 2006.
- [2] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.
- [3] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan Dos Santos Costa, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *bioRxiv*, 2022.
- [4] Martin Mirdita, Konstantin Schütze, Yosuke Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature Methods*, 2022.