Episode Book

EPISODE #6316

Why Protein Design Ignores Astronomical Counts

PUBLISHED	27 October 2026 at 19:50 UTC
DURATION	22:02
LISTEN	myweirdprompts.com/episode/protein-design-sequence-space-49838/
HOSTS	Corn, Herman, Siobhan, Hilbert
SCRIPT MODEL	openai/gpt-6.1-sol
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.22
SOURCES	19 cited by the research (listed at the end)
BOOK GENERATED	28 October 2026 at 09:53 UTC

SYNOPSIS

Generative AI can churn out hundreds of thousands of candidate molecules, but evaluating these numbers requires facing the staggering scale of biological sequence space. Exploring the mathematics behind amino acid arrangements and chemical universes reveals why exhaustive searches are fundamentally impossible. The discussion breaks down how evolutionary reuse, energy landscapes, and automated design campaigns narrow down infinite theoretical combinations into physically testable binders.

TRANSCRIPT



Corn

Daniel has been reading about a model generating something like twenty thousand candidate proteins, and he wants to know what that means against the number of proteins that could exist. Twenty thousand sounds enormous until you discover what sits underneath the fraction. How many arrangements can biology make from its building blocks, and how many possible molecules might qualify as starting points for medicines?



Herman

A hundred amino acids already gives you roughly ten to the power of one hundred and thirty possible sequences. Twenty thousand candidates barely disturbs the exponent.



Corn

The progress bar will be disappointing.



Herman

Yet those candidates can still be useful. You don't need to inspect every possible protein to find one that binds a particular target. You need a way to propose sequences that have a substantially better chance than arbitrary strings.



Corn

That's the part I want to separate, Herman. A computer producing twenty thousand objects sounds like production. Producing twenty thousand objects worth making sounds like discovery. The headline quietly lets those become the same thing.



Herman

The laboratory separates them rather firmly. Sometimes expensively. The sequence count tells us why exhaustive search cannot work, while the experimental results tell us whether the alternative does.



Corn

I'll predict your next move. You'll find an elegant shortcut through this space and briefly forget that anything has to dissolve.



Herman

Dissolving is on my list.



Corn

Beneath the elegant shortcut.



Herman

Within striking distance of it. There are two different spaces here. Protein sequence space counts arrangements along a chain. Small molecule chemical space counts possible molecular structures, with different atoms and connections. Those aren't interchangeable denominators.



Corn

A protein can itself become a medicine, though. We aren't sorting molecules into drugs and non-drugs by size.



Herman

Correct. Insulin is a protein medicine. The familiar estimate of ten to the sixtieth drug-like molecules usually concerns small molecules under particular chemical restrictions. It doesn't count every conceivable therapeutic, and it certainly doesn't mean that many effective drugs exist.



Corn

Drug-like has rather generous admissions.



Herman

It means meeting broad property criteria. We'll unpack those. For protein sequences, the arithmetic is cleaner. Twenty standard amino acids, any one allowed at each position, gives twenty raised to the chain length. Actual function imposes much tighter restrictions afterward.



Corn

Twenty thousand also isn't a campaign we need to identify from Daniel's rounded recollection. There are concrete examples nearby.



Herman

TpsGPT generated twenty-eight thousand sequences. Bits to Binders received twelve thousand designs. An autonomous Claude campaign generated three hundred and eighteen thousand, four hundred and eighty-four protein designs.



Corn

The friendly AI down the road has acquired inventory.



Herman

Those are proposals, importantly. The last campaign sent only a small selection for physical testing. Before we get there, take the sequence arithmetic literally. Every extra position multiplies the possible arrangements by twenty, regardless of whether those arrangements are useful.



Corn

Repetition allowed, order matters. Twenty choices for the beginning doesn't become nineteen choices for the next position.



Herman

Yes. An amino acid can appear repeatedly. Two positions give four hundred sequences. Three give eight thousand. Nothing has folded yet, and already the multiplication is doing most of the work.



Corn

That word permutations would have sent me toward arranging twenty different pieces once each. Biology lets you order another bag.



Herman

For a hundred positions, twenty to the hundredth is approximately ten to the one hundred and thirtieth. A one followed by about a hundred and thirty zeroes. The observable universe contains on the order of ten to the eightieth atoms.



Corn

We have run out of atoms before labels.



Herman

The sequence count exceeds that atom count by roughly ten to the fiftieth. Giving every atom its own sequence wouldn't come remotely close to recording them all. This is direct arithmetic from the alphabet, not an experimentally measured inventory.



Corn

Nor does every sequence correspond to a stable protein. We could count all hundred-character messages and get an impressive number without improving your shopping lists.



Herman

My shopping lists are quite precise.



Corn

They're arranged by biochemical function.



Herman

That is why we remember the lentils. But your distinction matters. Some sequences won't adopt a useful structure. Some will aggregate. Some might fold but do nothing relevant. The unconstrained count includes all of them.



Herman

Corn, would you count a sequence that cannot remain soluble as part of the search space?



Corn

In the mathematical count, yes. In a manufacturer's list of candidates, I would want it removed. Otherwise you can claim a vast discovery programme by generating things you already know you cannot use.



Herman

That difference between possible and eligible will follow us into chemistry. Ten to the sixtieth is a familiar estimate for drug-like chemical space, but the answer depends heavily on which atoms, structures and properties you're allowing.



Corn

There isn't a census office for imaginary tablets.



Herman

An influential estimate by Bohacek, McMartin and Guida in nineteen ninety-six put small drug-like possibilities at at least ten to the sixty-third. Other treatments give vastly smaller counts, around ten to the twentieth through ten to the twenty-fourth, for restricted enumerations up to thirty atoms.



Corn

Those cannot all mean the identical set.



Herman

They don't. The chemical restrictions and counting assumptions matter enormously. I wouldn't combine them into a single precise population. The useful conclusion is that plausible chemical universes exceed the libraries we can physically test by extraordinary margins.



Corn

Explain the drug-like filter without making it sound like a medical entrance exam.



Herman

Lipinski's rule of five is a familiar guide to properties associated with oral absorption. Molecular weight no more than about five hundred daltons, a measure of molecular mass. Then logP no greater than five, describing preference for an oily environment over water.



Corn

Swallowing it only solves the front door.



Herman

It still has to get into the body. The other familiar limits concern hydrogen bonding, no more than ten acceptors and five donors. Those are chemical features that affect interactions with water and biological surfaces.



Corn

None of that says it stops a tumour or kills a bacterium.



Herman

Nor guarantees good absorption. They're guidelines with exceptions, not a test of therapeutic activity. A molecule can fit the rules and be useless, toxic, or impossible to deliver at a helpful concentration.



Corn

Twenty thousand divided by ten to the sixtieth is two times ten to the minus fifty-sixth. We can write the fraction. We just cannot pretend it's meaningful coverage.



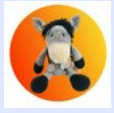
Herman

Even a billion candidates barely changes that conclusion. Yet practical libraries can yield discoveries. In twenty nineteen, roughly three hundred and fifty million virtual molecules were docked against targets including a bacterial enzyme and a dopamine receptor.



Corn

Docked means predicted to fit, rather than individually poured into little wells.



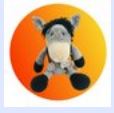
Herman

Predicted interactions first, selected compounds tested afterward. That campaign yielded a particularly potent inhibitor of the bacterial enzyme and a new activator of the receptor. Its success didn't require covering a substantial fraction of all chemistry.



Corn

It needed a useful selection process.



Herman

We should take a call before that word selection acquires another layer. Proteins use a twenty-letter alphabet, and that multiplication is what makes these counts so dizzying.

S

Siobhan

Sorry, I've been nodding along for weeks now, but what actually is a protein, is it a bit like a recipe, or a piece of string you'd get in a shop?

S

Siobhan

And when you say the space is huge, do you mean physical space, like the size of a room?



Herman

A protein is a little like a string of different beads that curls into a useful shape, and the huge space means all the different bead arrangements we could make, not a very large room.



Corn

Your shoulders dropped on the word beads.



Herman

Selection happens in nature too. Many proteins contain domains, often roughly fifty to two hundred and fifty amino acids long, that can fold as units. Evolution reuses and recombines those units rather than inventing every sequence without a starting point.



Corn

That sounds closer to renovation than searching.



Herman

Existing structure constrains what gets tried. Calling evolution a generative model with a prior is a useful shorthand, although evolution has no desired final design. The prior here means some arrangements are much more likely than others because of what already exists.



Corn

The same distinction applies to folding. A chain doesn't try every pose before choosing one.



Herman

That's Levinthal's paradox. Enumerating every possible conformation could take longer than the age of the universe. Yet some proteins fold on microsecond to millisecond timescales. Others take longer, but none needs to inspect every theoretical configuration.



Corn

Energy and chemistry do the steering.



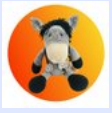
Herman

The landscape has preferred routes and intermediate states. Sequence space and folding space are different problems, but both punish the assumption that solving something requires visiting every possibility.



Corn

A model can therefore be useful while being extravagantly biased. We'd actually prefer it biased toward things that fold.



Herman

Condition it on a target surface, a desired function, or a structural requirement, and you change which proposals are likely. You're drawing from a tiny, selected region. It needn't even be one tidy region of similar sequences.



Corn

That last qualification hurt you.



Herman

Only because tidy regions are pleasant.



Corn

In the actual campaign, then, Claude wasn't simply typing amino acid letters until the weekend ended.



Herman

It orchestrated specialised protein design tools. In the August scorpion-toxin campaign, it generated three hundred and eighteen thousand, four hundred and eighty-four designs over sixty-four hours, targeting five venom proteins.



Corn

A weekend with unusually poor catering.



Herman

It called twelve design tools and ran more than five hundred and fifty thousand protein and toxin structure predictions. Those predictions estimate how a proposed binder might sit against its target. They aren't photographs of an interaction.



Corn

When you say autonomous, someone still chose the objective and paid for the work.



Herman

Humans set up the campaign and supplied prompts. No human made individual design decisions within that automated process. Autonomy has a boundary here, and it doesn't include the laboratory materialising proteins by itself.



Corn

This is perilously close to my ideal job. Propose things while nobody watches, let someone else wash the glassware.



Herman

You would still have to submit the selection.



Corn

I knew there was a catch.



Herman

The system selected a hundred and eighty designs. Additional selections brought the total sent for testing to two hundred and eleven. Adaptyv Bio and Twist Bioscience handled the physical work.



Corn

Already the generated count tells us almost nothing about the size of the experiment.



Herman

Seventeen met the binding criteria. Two binders against one toxin reached single-digit nanomolar dissociation constants, indicating tight binding under the measurement conditions.



Corn

Tight binding to venom sounds promising. It does not tell me the venom has stopped working.



Herman

They weren't tested for neutralisation in cells or animals, and their structures weren't experimentally solved. Anthropic's limitation statement was explicit. The evidence is binding, not structure or function. Every reported pose remains a prediction.



Corn

The docking illustration has excellent posture, though.



Herman

A beautiful predicted interface can be persuasive. Too persuasive. Corn, if you could afford to make only a small batch, would you choose the highest-scoring designs or deliberately spread your bets?



Corn

Spread them, unless the score has earned my trust on comparable experiments. Ten variants sharing one mistaken structural assumption could give me ten expensive copies of the same disappointment.



Herman

That is a real selection problem. Diversity can protect against correlated mistakes. The highest numerical score isn't automatically the most informative experiment, either. A slightly lower-ranked design might test a different binding arrangement.



Corn

Now the denominator I care about is successful experiments per available laboratory slot.



Herman

With the endpoint specified. Successful binding, successful neutralisation and successful treatment are different counts. A model optimised for one can disappoint on another without contradicting its original result.



Corn

Does the scoring system actually know which proposals are best, or is it mostly confident about things that resemble its training examples?



Herman

Sometimes it ranks usefully. Sometimes it does not. A blinded antibody benchmark found only about ten to fourteen percent of AI submissions beat the relevant control, compared with roughly thirty-nine percent for randomly selected clones.



Corn

Random has submitted a competitive application.



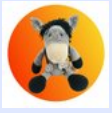
Herman

Only one participant exceeded that random baseline. That was a particular benchmark, not proof that every model loses to chance. Still, it exposes a serious gap between proposing plausible sequences and selecting the strongest performers.



Corn

Especially when a generator gets judged by its best survivor, while the selector escapes responsibility for the discarded possibilities.



Herman

The broader autonomous design campaign had more encouraging results. Of a thousand three hundred and twenty synthesised designs, three hundred and fifty-four bound. That's about twenty-seven percent.



Corn

Better than a curiosity, worse than the pictures suggest.



Herman

Almost half of the top-ranked designs bound in that broader campaign. Ranking carried useful information there. We can't flatten all these experiments into a claim that ranking never works, any more than we can call it solved.



Corn

Different targets, different candidate pools, different tests. Your beautiful mechanism has acquired paperwork.



Herman

Paperwork I support. The scorpion campaign also had no binders against one of its targets. A headline average hides where the method still struggles.



Corn

Faster generation might make this worse for the person choosing. If every proposal looks plausible, the shortlist becomes a full-time occupation.



Herman

RFD4-Proteina, from the University of Washington's protein design institute and NVIDIA, reports fivefold to tenfold improvements in generation speed and design success against previous models. It's open source, so more groups can produce candidates.



Corn

More proposals arrive at the same pipette.



Herman

Unless experimental throughput improves alongside them. Competitions are increasingly including physical validation. Bits to Binders tested twelve thousand designed minibinders in living T cells, rather than stopping at a structural score.



Corn

Living cells are notoriously indifferent to presentation quality.



Herman

The Changjiang Cup is offering free laboratory testing for more than five thousand designed sequences. PROTEINA's public pitch also centres on identifying legitimate candidates among vast generated lists. Validation is becoming a product in its own right.



Corn

TpsGPT gives us the smaller version of the same funnel. Twenty-eight thousand generated sequences, seven passing the filters, activity confirmed in at least two.



Herman

Yes, designed enzymes rather than toxin binders. You have to preserve that distinction when comparing hit rates. Corn, would a cheaper experiment persuade you more than a better predictor?



Corn

If the predictor is wrong in ways we cannot recognise, cheaper experiments. If the experiment measures the wrong thing, neither. I'd spend some money checking what our test actually rewards before buying more of either.



Herman

An excellent binding assay can still answer a narrower question than the programme needs.



Hilbert

Herman, does your amino acid count assume equal portions? My protein spreadsheet has seven point four percent cysteine. I've been logging everything I eat by sequence where I can get one. That percentage is statistically impossible. I suspect the nutritional yeast.



Herman

Equal choices describe the unrestricted count. Real proteins don't use every amino acid equally often. Your spreadsheet percentage wouldn't establish that the yeast was mislabeled.



Hilbert

I'm taking it to Nissim at the ingredient shop on Thursday. He's supplied me for long enough to recognise a column heading. There are eighteen hundred rows. Eighteen hundred and forty-seven, counting sauces only when they have their own bowl. I haven't cooked the same recipe twice. The last seitan came out wrong, and I'm having him initial the cysteine column before I order again.



Corn

He cannot initial you a control group.



Hilbert

He can stop writing yeast on bags that might be something else. He's there at six, weighing powders, answering delivery calls and finding substitutions. Then I arrive with a printout. This week I'm booking his counter for a sequence audit. You took care to distinguish binding from neutralising, Herman. I can distinguish a receipt from an amino acid sequence.



Herman

He'll need the product records. The food label alone won't give you the sequences of all its proteins.



Hilbert

The spreadsheet is on the laptop in the freezer. Heat degrades storage media. The paper backup is beside it in a ziplock bag marked Protein Archive, Do Not Eat. I'll take that copy to the shop. The screen fogs when it comes out, and Nissim charges for the counter by the quarter hour.



Corn

Defrosting has become an administrative expense.



Hilbert

He wants the audit before his delivery. He has allowed me eleven minutes.



Herman

Eleven minutes wouldn't leave much room to establish which ingredient changed. In an actual design campaign, keeping the records consistent is part of making the experiment interpretable.



Corn

A candidate identifier sounds trivial until the generated sequence, the manufactured sample and the measured result stop agreeing. Then your learning loop is learning from the wrong protein.



Herman

Provenance is getting attention too. DeepMind's SynthIDBio embeds identifying signatures in designed sequences and predicted structures. The aim is to preserve function while making origin traceable.



Corn

Useful as those designs circulate between organisations. It still doesn't certify that the design works.



Herman

Origin and performance need separate evidence. A watermark can answer where something came from without answering what it does.



Corn

The most common wrong belief here is that generating hundreds of thousands of candidates means we've searched a meaningful portion of protein space.



Herman

We've proposed a tiny, heavily conditioned selection. Its value comes from experimentally confirmed performance, not the fraction of all possible sequences covered. Even seventeen confirmed binders remain a long way from seventeen medicines.



Corn

I'll be watching who gets to choose the laboratory shortlist, and whether their records let everyone else learn from the failures as well as the successes. The next useful improvement may be in those decisions rather than another mountain of proposals.



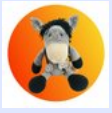
Herman

Thanks to Hilbert Flumingtop for producing the show.



Corn

For more along these lines, there's episode four, If Your Voice Ages, Does Your Fine-Tune Become Useless. This has been My Weird Prompts, the human-AI collaboration podcast.



Herman

If you enjoyed it, leave us a review wherever you listen.



Corn

See you tomorrow.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **UW Medicine Newsroom, RFD4-Proteina announcement [primary]**
   https://newsroom.uw.edu/news-releases/researchers-achieve-major-advance-in-protein-design

2. **36Kr, Claude designed 310,000 kinds of proteins within 64 hours (2026)**
   https://eu.36kr.com/en/p/4034833727100038

3. **UbiTiter Issue #12, AI Protein Design: Generation Works, Ranking Does Not (2026-08-20)**
   https://unibiointelligence.com/titer/2026-08-20/

4. **Autonomous De Novo Protein Binder Design with Claude, alphaXiv listing**
   https://www.alphaxiv.org/abs/2608.claude-de-novo

5. **Carterra press release on the Anthropic study (2026-08-18)**
   https://carterra-bio.com/news/carterra-enables-anthropics-autonomous-ai-protein-design-study-via-its-ht-spr-binding-platforms/

6. **ACS *Accounts of Chemical Research*, The Chemical Space Project**
   https://pubs.acs.org/achre4/article/48/3/722/1248286/The-Chemical-Space-Project

7. ***Nature* news, Ultra-large virtual molecular libraries throw open chemical space (2019)**
   https://www.nature.com/articles/d41586-019-00482-6

8. ***C&EN*, Hunting for drugs in chemical space (Bohacek 10^63 estimate)**
   https://cen.acs.org/pharmaceuticals/drug-discovery/Hunting-drugs-chemical-space/100/i23

9. **Wikipedia, last edited 2026-08-11**
   https://en.wikipedia.org/wiki/Lipinski%27s_rule_of_five

10. **Wikipedia, last edited 2026-10-11**
    https://en.wikipedia.org/wiki/Levinthal%27s_paradox

11. **Wikipedia, last edited 2026-10-26**
    https://en.wikipedia.org/wiki/Protein_domain

12. **Wikipedia, last edited 2026-10-06**
    https://en.wikipedia.org/wiki/Drug_discovery

13. **TpsGPT, Ramanathan et al., revised 2025-12-15**
    https://arxiv.org/abs/2512.08772v2

14. **Reinforcement-guided generative protein language models... AAV capsids (2026-03-19)**
    https://arxiv.org/abs/2603.19473v1

15. **Swarms of Large Language Model Agents for Protein Sequence Design (2025-11-27)**
    https://arxiv.org/abs/2511.22311v1

16. **Bits to Binders competition coverage**
    https://bioengineer.org/global-ai-protein-design-contest-puts-12000-computer-made-cancer-binders-to-the-test-in-living-t-cells/

17. **Changjiang Cup 2026 announcement (2026-09-24)**
https://www.biointron.com/company-news/2026-global-ai.html

18. **Edaily on PROTEINA Co. validation of AI-designed candidates (2026-09)**
https://en.edaily.co.kr/news/eda202609145031/

19. **SynthIDBio / DeepMind *Nature* paper (2026-10-12)**
https://carterra-bio.com/news/google-deepminds-nature-study-on-watermarking-ai-designed-proteins-validated-with-carterra-ht-spr/