BASECAMP RESEARCH — EDEN MCP CONNECTOR

DOCUMENTATION

AUTHENTICATION
OAuth · no API keys

CAPABILITIES
AMP design · Immunogenicity

SUPPORT
support@basecamp-research.com

EDEN is Basecamp Research's biological foundation model. This connector lets you use two of EDEN's capabilities directly in a conversation with Claude:

1 **Antibiotic design** — generate novel antimicrobial peptides against drug-resistant bacteria.

2 **Vaccine target prioritisation** — predict whether a protein antigen is likely to trigger an immune response.

You interact with both in plain language. Claude handles the tool calls, file uploads, and result formatting for you.

RESEARCH USE ONLY

EDEN's outputs are model predictions, not experimental results. They are intended to guide and prioritise laboratory work, not replace it. Nothing produced through this connector should be used to inform medical decisions, clinical diagnoses, or patient care. Always validate candidates in the lab.

CONTENTS

01 GETTING STARTED

Connecting

Add EDEN from the connectors menu in Claude and authenticate when prompted. EDEN uses OAuth — there are no API keys to manage. Once connected, you can start asking Claude to design peptides or score antigens straight away.

What you can ask

You don't need to know the tool names or syntax. Natural requests work, for example:

- *"Design some antimicrobial peptides against MRSA."*
- *"Generate AMP candidates for Pseudomonas aeruginosa PA14 and show me the most potent ones."*
- *"Predict the immunogenicity of this coding sequence: ATG..."*
- *"I've uploaded a FASTA of candidate antigens — score them all and rank by immunogenicity."*

Claude will pick the right capability, run it, and return the results.

02

CAPABILITY 1 — ANTIMICROBIAL PEPTIDE (AMP) GENERATION

What it does

Given a single target pathogen, EDEN generates novel antimicrobial peptide candidates and predicts how potent each one is against that strain. Each call returns up to 30 candidates. For every candidate you get:

- a sequence identifier
- the amino-acid sequence of the peptide
- a predicted minimum inhibitory concentration (MIC) for the target strain

Lower predicted MIC indicates higher predicted potency.

How it was validated

EDEN's AMP generation was validated in collaboration with the de la Fuente Lab at the University of Pennsylvania. In that work, **97% of EDEN-designed peptides were active against WHO-priority pathogens** when tested in vitro. One candidate, EDEN-7, was tested in mice infected with multidrug-resistant *Acinetobacter baumannii* and showed efficacy in the same range as a last-line antibiotic — despite being generated zero-shot, with no optimisation.

Supported target strains

Provide exactly one strain per request, chosen from the list below:

PATHOGEN	STRAIN
S. aureus (MRSA)	`ATCC BAA–1556`
S. aureus	`ATCC 12600`
Vancomycin-resistant E. faecalis	`ATCC 700802`
Vancomycin-resistant E. faecium	`ATCC 700221`
A. baumannii	`ATCC 19606`
E. coli	`AIC221`
E. coli	`AIC222`
E. coli	`ATCC 11775`
K. pneumoniae	`ATCC 13883`
P. aeruginosa	`PA01`
P. aeruginosa	`PA14`

If you ask for a strain that isn't on this list, Claude will tell you and suggest the closest supported option.

Example

YOU

Generate antimicrobial peptides against vancomycin-resistant E. faecium and show me the five with the lowest predicted MIC.

CLAUDE

Calls EDEN, then returns a ranked table of candidate IDs, sequences, and predicted MICs against ATCC 700221.

Interpreting the results

- Predicted MIC is a model estimate of potency against the requested strain only. It does not tell you about activity against other strains, toxicity, stability, or manufacturability.
- Candidates are starting points for laboratory testing. The 97% in-vitro hit rate above reflects EDEN's overall performance in a validated setting; it is not a guarantee for any individual peptide.
- Predicted potency against one strain does not imply broad-spectrum activity.

03

CAPABILITY 2 —
IMMUNOGENICITY
PREDICTION

What it does

EDEN predicts the probability that a protein-coding antigen will induce an immune response, working directly from the antigen's native nucleotide coding sequence (CDS). This is useful for prioritising vaccine targets: instead of testing every candidate antigen empirically, you can rank them by predicted immunogenicity first and focus laboratory effort on the most promising.

Each sequence returns a probability between 0 and 1.

How it works

The pipeline embeds your sequence using EDEN — a 28-billion-parameter genomic language model trained on 9.7 trillion tokens of metagenomic sequence data — and applies a logistic regression probe to the mean-pooled final-layer embedding. The probe was trained on 3,911 natural nucleotide coding sequences from NCBI and tested on a further 1,111. On an independent external dataset (ImmunoDB) it achieves an AUROC of 0.85.

Input requirements

This is the part to get right. Each sequence you submit must be:

- a natural / native nucleotide coding sequence (DNA: A, C, G, T)
- forward strand
- in-frame — length a multiple of 3
- between 150 and 8,192 nucleotides
- a complete protein-coding sequence

The model will not give meaningful results for:

- ☐ amino-acid sequences (it needs nucleotides, not protein)
- ☐ back-translated or codon-optimised sequences — these shift the embedding distribution and degrade accuracy
- ☐ partial proteins, epitope-only fragments, or subsequences
- ☐ sequences trimmed or padded just to satisfy the length limits

If a sequence is invalid, it's returned as rejected with a reason; valid sequences in the same batch are still scored. If you don't have a valid native CDS, Claude will tell you rather than working around it — it won't back-translate or truncate to force a result.

Two ways to submit sequences

Inline — for one or a few sequences

Paste the nucleotide CDS directly into the chat. Best for a single sequence or a quick check. Limited to 5 sequences per call.

YOU

Predict the immunogenicity of this antigen: `ATGACC…`

CLAUDE

Runs the prediction and reports the probability.

Dataset upload — for many sequences

Upload a FASTA file and Claude will score the whole set. This is the preferred path for anything beyond a handful of sequences.

- Format: FASTA
- Maximum file size: 100 MB
- Maximum 100 valid sequences per scoring call — Claude splits larger sets into batches automatically
- The upload is verified against a SHA-256 digest of the file, so the bytes scored are exactly the bytes you provided
- Uploaded datasets are private to your account

YOU

Here's a FASTA of 240 candidate antigens from this pathogen. Score them all and give me the top 20 by predicted immunogenicity.

CLAUDE

Uploads the file, runs the predictions in batches, and returns a ranked list.

Interpreting the score

The result is a model-estimated probability, not a confirmed biological measurement. As a rough guide:

SCORE	INTERPRETATION
≥ 0.90	Upper range of the model's output; high predicted immunogenicity. Strong candidate for experimental follow-up.
0.50 – 0.89	Intermediate. Treat as a positive signal worth corroborating with other evidence.
< 0.50	No strong immunogenicity signal predicted. Validation still advised before ruling a candidate out.

A few things to keep in mind:

- The score reflects how similar your sequence looks — through EDEN's embeddings — to immunogenic sequences in the training data.
- Bacterial and viral antigens are best represented in the training data and the model is most reliable there. For antigens from other organisms, assume lower reliability.
- Real-world performance on novel inputs may differ from the 0.85 AUROC benchmark.
- The score says nothing about mechanism, protective efficacy, or clinical outcome — only the predicted probability of inducing an immune response.
- Always validate in the lab, regardless of the score, and never use these results to make clinical decisions.

04

BUILT ON BASEDATA

EDEN is trained on BaseData™, the largest, fastest-growing and most information-rich biological database in the world. To build it, Basecamp Research has run sampling expeditions to over 200 locations across more than 30 countries — thermal springs, deep-sea sediment, polar ice, high-altitude plateaus — documenting more than a million species new to science and over 10 billion new genes, roughly ten times the content of every public database combined.

Every sample is collected under informed-consent and benefit-sharing agreements, with each sequence traceable to one of hundreds of country-specific collection permits. This allows a portion of value to flow back to the countries and communities who steward the biodiversity it comes from.

05 DATA AND PRIVACY

- Sequences you paste and datasets you upload are scoped to your account and private by default.
- Dataset uploads use short-lived, presigned URLs and expire automatically.
- For full details, see the Basecamp Research Online Privacy Notice.

06 LIMITATIONS AT A GLANCE

- **Outputs are predictions, not experimental results.** Validate everything in the lab.
- **Not for clinical use.** No medical decisions, diagnoses, or patient care.
- **AMP generation:** one supported strain per request; predicted MIC applies to that strain only.
- **Immunogenicity:** natural nucleotide CDS only — no amino-acid, codon-optimised, back-translated, or partial sequences; 150–8,192 nt, in-frame; up to 5 inline or 100 valid sequences per dataset call.
- **Reproducibility:** scores are tied to a specific model checkpoint; results from a different checkpoint or software stack are not directly comparable.

07 SUPPORT

Questions, issues, or access requests

Email: support@basecamp-research.com
EDEN preprint: biorxiv.org/content/10.64898/2026.01.12.699009v2

08 ACKNOWLEDGMENTS & ATTRIBUTIONS

This connector, and Basecamp Research services that build on EDEN model output (including antimicrobial peptide generation), incorporate third-party open-source software. Basecamp Research gratefully provides the attribution below.

APEX-pathogen (APEX 1.1)

Antimicrobial peptide (AMP) generation incorporates APEX-pathogen (APEX 1.1), open-source software developed by Dr. César de la Fuente and the Machine Biology Group at the University of Pennsylvania, which is used to curate and prioritize the antimicrobial peptide candidates derived from EDEN model output. APEX 1.1 is made available by its authors under the MIT License at gitlab.com/machine-biology-group-public/apex-pathogen. The AMP capability was developed in collaboration with Dr. de la Fuente's group and is provided free of charge for research use only, in accordance with the applicable Terms of Service. Basecamp Research gratefully acknowledges Dr. de la Fuente and his laboratory for making this work openly available to the research community.

Citation

Torres, M. D. T., Wan, F. & de la Fuente-Nunez, C. "Deep learning reveals antibiotics in the archaeal proteome." *Nature Microbiology* 10, 2153–2167 (2025). doi:10.1038/s41564-025-02061-0.

This acknowledgment does not imply that Dr. de la Fuente, the Machine Biology Group, or the University of Pennsylvania endorses the Service, Basecamp Research, or any Output.