Successful Acceleration of HMMER for the “PEZY-SC3” and General-Purpose CPUs Using Generative AI

— Up to approximately 70x speedup of hmmscan, a cornerstone of protein function analysis —

PEZY Computing, Inc. (hereinafter “PEZY Computing”) and QuettaAI, Inc. (hereinafter “QuettaAI”) are pleased to announce that they have successfully implemented an accelerated version of HMMER [1,2], a sequence search tool widely used in the field of bioinformatics, for the “PEZY-SC3” — an MIMD processor for HPC developed and manufactured by PEZY Computing — and for general-purpose CPUs.

The project was led by software engineer Toshiyuki Oda (QuettaAI). By making full use of a state-of-the-art generative AI coding agent (Claude [3] Code / Opus 5), code modification for a complex architecture was accomplished rapidly.

Project Background and the Use of AI

With the dramatic improvements in generative AI in recent years, AI is now able to provide powerful support for hardware-specific code modification and optimization, work that has traditionally required a high degree of specialized expertise. In this project, the implementation process was entrusted to Claude Code, and the complex HMMER codebase was successfully optimized for the PEZY-SC3 in a short period of time.

To ensure reliability in scientific computing, development was carried out under the requirement that computational results match the original version 100%. Parts of the computation in which the ordering of floating-point operations would be problematic were avoided, and numerical values such as E-values and bit scores were confirmed to match exactly in the final output.

Dramatic Performance Gains

Benchmark tests using a range of datasets confirmed dramatic improvements in processing speed.

- **Up to approximately 70x faster:** For hmmscan, a speedup of up to approximately 70x over the original version was achieved (processing 5,000 fragment sequences predicted by prodigal [5] from high-throughput sequencing (HTS) reads ERR056999 [4]; Table 1-1).
- **Substantially shorter analysis times in practice:** Using an environment equipped with four PEZY-SC3 units, Pfam [6] annotation of one HTS run (all 262,574 reads of ERR056999 translated by prodigal) can be completed in **approximately 11 minutes**, and one bacterial proteome (the *E. coli* proteome, 4,403 sequences) in **approximately 27 seconds** (roughly a 40x speedup). Even compared with the inversion technique [7], which uses hmmsearch in place of hmmscan, this represents a speedup of approximately 10x (Table 2-1).
- **Substantial speedup on CPUs as well:** Thanks to our own optimization of the host-side code, speedups of several-fold to several tens of fold for hmmscan are achieved even in CPU-only environments that do not use the PEZY-SC3.

Future Developments

The optimized source code developed in this project is scheduled for public release as open source. Using accelerators such as the PEZY-SC3 efficiently has until now demanded deep specialist knowledge, but recent advances in coding AI agents have made it possible to implement accelerator code with far greater ease. This is expected to remove bottlenecks in the processing of the vast volumes of life-science data generated by metagenomic analysis and other applications, and to significantly accelerate research and development in the life sciences.

PEZY Computing and QuettaAI will continue to combine cutting-edge hardware technology with artificial intelligence technology in order to open up new possibilities in high-performance computing (HPC).

Execution Environment

A node with an AMD EPYC 7702 64-Core Processor (64 cores, 128 threads) and 1 TB of RAM. HMMER 3.4 was built in the standard manner following README.md, using `./configure --prefix <path>` / `make`.

Parameter Search

For the `--cpu` option, using the *Escherichia coli* proteome, the values 0 (no parallelization), 1 (single-threaded, but the parallelization blocks are used), 2, 4, 8, 16, 32, 64 and 128 were tested for the original version. For the partially CPU-modified version and the SC3 version, however, no improvement was expected at 0, 1 or 2, so 4, 8, 16, 32, 64 and 128 were tested.

The shortest processing times were obtained with `--cpu 0` for the original hmmscan, `--cpu 64` for the partially CPU-modified and SC3 versions of hmmscan, and `--cpu 4` for the original hmmsearch.

Data	Seqs	Amino acids	Tool	Original (--cpu 0)	cpu_ex[a] (--cpu 64)	SC3 (--cpu 64)	CPU speedup	SC3 speedup
ERR056999[b]	262574	32571815	hmmscan(Pfam-A 38.2)	NA[c]	1598.306	673.983	NA	NA
ERR056999_5000[d]	5000	610556	hmmscan(Pfam-A 38.2)	979.648	33.710	14.555	29.061	67.307
hsa_2500[d]	2500	1375109	hmmscan(Pfam-A 38.2)	831.528	38.823	32.285	21.418	25.756
eco_proteome	4403	1354442	hmmscan(Pfam-A 38.2)	1057.031	41.100	27.318	25.719	38.694
eco100[d]	100	31686	hmmscan(Pfam-A 38.2)	29.018	3.215	2.933	9.026	9.894
eco100	100	31686	jackhmmer(sprot)[e,f]	723.888	173.329	42.403	4.176	17.072
eco100[g]	100	31686	hmmsearch(sprot)[e,f]	1677.944	491.041	113.733	3.417	14.753

Table 1-1. Processing time for each dataset, in seconds.

a. Version in which only the CPU-side modifications are applied, without using the SC3.

b. Predicted from the reads by prodigal.

c. Omitted because a processing time of 10 hours or more is anticipated.

d. 5,000, 2,500 and 100 sequences respectively drawn at random from ERR056999, the *Homo sapiens* proteome (UP000005640) and the *Escherichia coli* proteome (UP000000625) downloaded from UniProt [8].

e. Search against swissprot [8]. When a large database such as trembl [8] is the target, the computation falls back to the CPU even when the SC3 is in use, and the processing time was almost the same as that of the partially CPU-modified version (results omitted).

f. For jackhmmer and hmmsearch(sprot), thread parallelism is effective; in the comparison at --cpu 32 (Table 1-2), the original version and the partially CPU-modified version are roughly equivalent (speedup of about 1.0).

g. The HMM produced by jackhmmer in the row above was used.

Data	Tool	Original (--cpu 32)	cpu_ex[a] (--cpu 32)	SC3 (--cpu 32)	CPU speedup	SC3 speedup
ERR056999[b]	hmmscan(Pfam-A 38.2)	NA[c]	1595.11	665.149	NA	NA
ERR056999_5000[d]	hmmscan(Pfam-A 38.2)	2246.959	34.014	14.719	66.060	152.657
hsa_2500[d]	hmmscan(Pfam-A 38.2)	1176.821	39.174	32.423	30.041	36.296
eco_proteome	hmmscan(Pfam-A 38.2)	1971.131	41.407	27.607	47.604	71.400
eco100[d]	hmmscan(Pfam-A 38.2)	45.372	3.266	2.95	13.892	15.380
eco100	jackhmmer(sprot)[e]	475.765	480.393	120.524	0.990	3.947
eco100[g]	hmmsearch(sprot)[e]	168.961	170.961	45.997	0.988	3.673

Table 1-2. Results for the comparison in Table 1-1 with the --cpu option set to 32. Sequence counts, amino acid counts and footnotes are the same as in Table 1-1.

Data	hmmsearch (--cpu 4)	CPU(--cpu 64) speedup	SC3(--cpu 64) speedup
ERR056999	7816.99	4.891	11.598
ERR056999_5000	174.179	5.167	11.967
hsa_2500	366.68	9.445	11.358
eco_proteome	303.274	7.379	11.102

Table 2-1. Processing time (in seconds) of the original hmmsearch when the inversion technique [7], which uses hmmsearch for Pfam annotation, is applied, together with the speedup of the customized hmmscan relative to it. For the hmmsearch of the inversion technique, the results at --cpu 32 are also given in Table 2-2.

Data	hmmsearch (--cpu 32)	CPU(--cpu 32) speedup	SC3(--cpu 32) speedup
ERR056999	9488.088	5.948	14.265
ERR056999_5000	248.975	7.320	16.915
hsa_2500	435.164	11.108	13.421
eco_proteome	394.97	9.539	14.307

Table 2-2. Results for the comparison in Table 2-1 with the --cpu option set to 32.

References

[1] http://hmmer.org/ (accessed 2026/09/24)

[2] Eddy, Sean R. "Accelerated profile HMM searches." *PLoS computational biology* 7.10 (2011): e1002195.

[3] https://claude.ai/ (accessed 2026/09/24)

[4] https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=ERR056999, retrieved on 2026-08-31 using sratoolkit [9]

[5] Hyatt, Doug, et al. "Prodigal: prokaryotic gene recognition and translation initiation site identification." *BMC bioinformatics* 11.1 (2010): 119.

[6] Paysan-Lafosse, Typhaine, et al. "The Pfam protein families database: embracing AI/ML." *Nucleic acids research* 53.D1 (2025): D523-D534.

[7] http://cryptogenomicon.org/hmmscan-vs-hmmsearch-speed-the-numerology.html (accessed 2026/09/24)

[8] Bateman, Alex, et al. "UniProt: the universal protein knowledgebase in 2025." *Nucleic acids research* 53.D1 (2025).
[9] https://github.com/ncbi/sra-tools version 3.4.1 (accessed 2026/09/24)

Released: September 25, 2026

Media Contact
QuettaAI, Inc.
6F Chiyoda Ogawamachi Crosta, 1-11 Kanda-Ogawamachi,
Chiyoda-ku, Tokyo 101-0052, Japan
https://www.quetta-ai.co.jp/?page_id=178&lang=en