

# DeCypherHMM™

## Algorithm Module

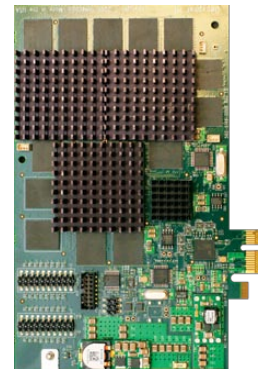
### The high-performance solution for protein characterization

DeCypherHMM provides the tools for connecting distantly related sequences and annotating the domain structure of protein families. The DeCypherHMM suite is available as a DeCypher module and as part of the CodeQuest workstation.

#### DeCypherHMM advantages:

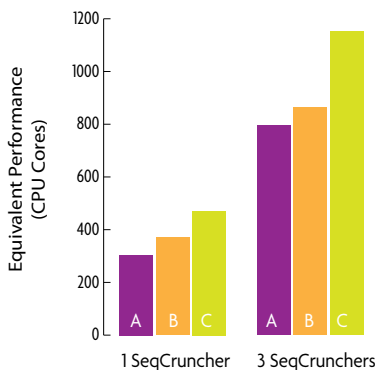
- Model-based analysis helps characterize multi-domain proteins
- Frameshift-tolerance helps you classify evolutionarily divergent proteins
- Ideal for classifying novel proteins for metagenomics studies
- Easily build and expand model databases from your sequence data
- Can return multiple hits per protein domain using J-state processing

SeqCruncher™  
FPGA Accelerator



TimeLogic's  
FPGA accelerator  
can process HMM  
searches with the  
performance  
of 304-471  
CPU cores

### DeCypherHMM delivers 304-1151 CPU cores of performance within a single server



Due to the highly parallel SeqCruncher circuitry, DeCypherHMM is able to substantially outperform traditional software such as HMMER. Running 3 typical HMM comparisons (A, B, C), DeCypherHMM—with a single SeqCruncher—completed the searches 304X, 369X and 471X faster than when they were run with HMMER software on 1 CPU core. With 3 SeqCruncher cards installed, DeCypherHMM demonstrated the performance equivalent of 794, 863, and 1151 CPU cores. Our HMM Whitepaper includes all search details.

Tests were completed on a Dell PowerEdge 2950 III server with two quad-core Xeon (3.0 GHz) processors and 8 GB of RAM (running 64-bit Fedora Core 6)



A 2U DeCypher host server can drive up to 3 SeqCruncher accelerators

### DeCypherHMM model-based search methods

Algorithm	Description	Applications
Hidden Markov Model Analysis	Comparison of nucleic or amino acid sequences to Hidden Markov Model databases	<ul style="list-style-type: none"> <li>• Assign protein families, domains and functions to cDNAs or genomic sequences (NT-HMM)</li> <li>• Find repetitive elements &amp; regulatory regions (e.g. transcription factor binding sites) (NT-NT HMM)</li> <li>• Assign domains and functions to protein sequences, predict 3D structures (AA-HMM)</li> <li>• Find new family members, train HMM models (HMM-AA)</li> </ul>
HMM-Framesearch	Frameshift-tolerant comparison of HMMs to nucleic data	<ul style="list-style-type: none"> <li>• Assign protein families, domains and functions to cDNAs and genomic sequences with tolerance for low-quality or cross-species sequence data (NT-HMM, HMM-NT)</li> </ul>
Profile Analysis	Comparison of nucleic or amino acid sequences to profile databases	<ul style="list-style-type: none"> <li>• Assign protein families, domains and functions to cDNAs or genomic sequences (NT-Profile)</li> <li>• Find repetitive elements &amp; regulatory regions (e.g. transcription factor binding sites) (NT-NT Profile)</li> <li>• Assign domains and functions to protein sequences, predict 3D structures (AA-Profile)</li> <li>• Find new family members, train profile models (Profile-AA)</li> </ul>
Profile-Framesearch	Frameshift-tolerant comparison of Profile models to nucleic data	<ul style="list-style-type: none"> <li>• Assign protein families, domains and functions to cDNAs or genomic sequences with tolerance for low-quality or cross-species sequence data (NT-Profile, Profile-NT)</li> </ul>

# DeCypherHMM™

## Frameshift-tolerant, model-based analysis

The DeCypherHMM suite includes both HMM profiles and standard Profiles. HMMs are often preferred because they provide more probabilistically correct modeling, which can translate to more sensitive results.

HMM-Framesearch and Profile-Framesearch extend these methods for use with nucleic sequences. By adding the capability to shift alignments across reading frames (also known as frame-interleaving), these become valuable tools for comprehensive nucleic sequence annotation. This is especially important when using low-coverage genomes or performing cross-species comparisons, where frameshifts and splice variants are more common.

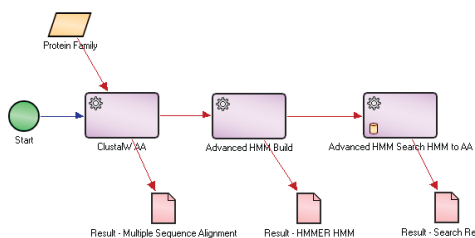
## Interfaces for all levels of bioinformatics expertise

DeCypherHMM's graphical user interface features extensive parameter settings and permits you to submit thousands of sequences per search. Color-coded search results, returned as HTML or text, reflect biologically relevant alignment details, including % identity, gapping & frameshifts (on each strand), mismatches and multiple domains. Target databases may also be built via the web (up to 200MB), simplifying administration.



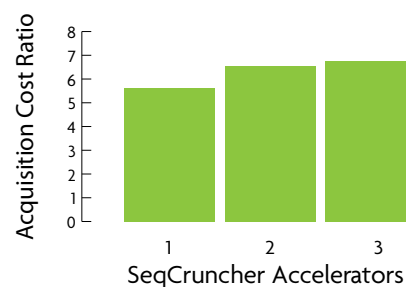
Larger, more complex searches may be submitted through the Command Line client, which can be installed on Windows, Linux, and Solaris operating systems. This efficient connection to the DeCypher system relies on a set of concise command arguments for total control over your search jobs. These commands are easily scripted for automated pipelines.

With our optional PipeWorks™ client/server software for Windows, DeCypherHMM searches may be interconnected with other search methods and filters to create comprehensive analysis workflows.



## DeCypher vs. the commodity cluster

Computational performance is only meaningful when solution cost is considered as well. In the graph below, we compare the cost of a DeCypher system (including the DeCypherHMM module, 1, 2, or 3 SeqCruncher cards and a Dell 2950 III server) to the cost of a cluster with equivalent performance (Dell 1950 IU servers with 2 quad-core CPUs). The cluster would require 5-6.6X more budget to acquire.



Once you factor additional cluster operational costs (power, HVAC cooling, IT staff, server room rackspace), the DeCypher system becomes even more cost-effective.

## Integration with other TimeLogic solutions

The DeCypher® Enterprise system is available with DeCypherHMM™, TeraBLAST®, DeCypherSW™ and the GeneDetective™ algorithm modules. With built-in 2-dimensional scalability that enables searches to leverage multiple SeqCruncher accelerator cards and multiple accelerated nodes, DeCypher enables you to configure an optimal biocomputing solution for your research environment.

DeCypherHMM is also included with CodeQuest™, a plug-and-play workstation for personal or lab-wide genomics projects. CodeQuest includes all four algorithm modules, a powerful desktop workstation and 1 or 2 SeqCruncher accelerators.

## DeCypherHMM expands your research capacity

With DeCypherHMM, your entire research team can run *ad hoc* queries or schedule comprehensive, repetitive analyses for confident, up-to-date functional assignments from your expanding genomic resources. In either case, DeCypherHMM processes your analyses at the speed of hundreds to thousands of CPUs cores, but at a fraction of the cost and physical requirements.

Contact TimeLogic today to learn more about DeCypherHMM or to schedule a free performance benchmark on your data!