Fachhochschule Salzburg University of Applied Sciences
Information Technology & Systems Management

UNIVERSITÄT SALZBURG
Department of Molecular Biology

HMModeler, a new approach for designing profile HMMs for protein families

R. Graf [a], M. Aigner [a], M. Lechner [a], D. Schroffner [a], Peter Lackner [b] und Stefan Wegenkittl [a]
(a) Information Technology and Systems Management, Salzburg University of Applied Sciences GmbH, (b) Department of Molecular Biology, University of Salzburg

pHMM Technology in Chimera

We present *HMModeler V2.0* [1], a toolset for UCSF *Chimera* [2], that guides through the process of designing and applying profile-HMMs (pHMMs) for analyzing protein relationships.

We see this process as a sequence of steps in which the system supports the user by generating default configurations. The intuitive GUI offers a simultaneous display of multiple sequence alignment (MSA) *and* 3D structure and allows for manual refinement of each automatically generated solution wherever applicable. In particular, HMModeler offers a set of biologically meaningful meta-parameters to define areas in the MSA with varying properties such as degree of conservation, desired sequence patterns and structural stability.

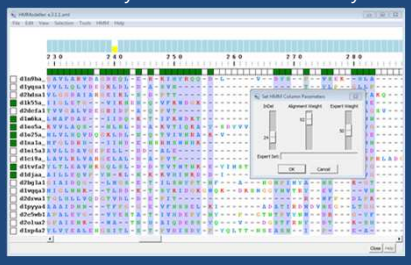
Start — Superimposed 3D Structures

High quality sequence alignments are derived from superimposed structures [3].
The authors' PSC++ [4] system, also integrated in Chimera, computes a set of optimal solutions.

Loop

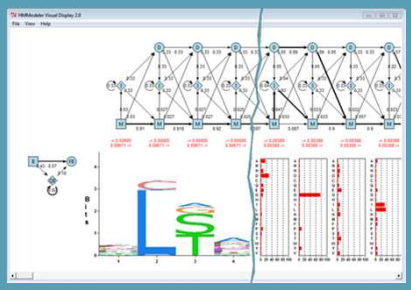
01 Alignment Editor and Expert Tuning of pHMM

Benefitting from the simultaneous 3D display, the MSA can be edited and refined manually. The model columns and all pHMM parameters are automatically set to defaults but may be changed by the user. A set of biologically meaningful meta-parameters is available to designate areas in the MSA with high/low degree of conservation and to define sequence patterns and degree of structural stability in order to more accurately trim the resulting pHMM for various applications.



02 pHMM Generation & Display

HMModeler then generates a 9-state pHMM for the family under consideration. The resulting model may be displayed offering both, histogram and sequence-logo display of the emission probabilities.



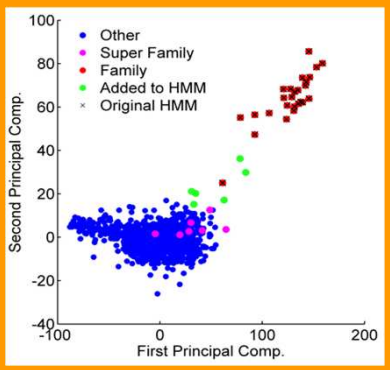
Advanced users may iteratively refine settings and meta-parameters on the basis of this representation.

03 Scoring

The user may trigger the computation of scores for single sequences or for whole databases.

HMModeler offers Forward and Viterby scores and various null-models, including backward sequence scoring.

04 Significant HITS Detection

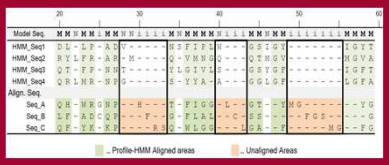


The most significant hits in a database query are displayed for further analysis.

Different scores may be used simultaneously in order to optimize the discriminating power of the pHMM.

Current work aims at improving the accuracy of e- and p-values by calibrating the models on the basis of order-2 Markov sampling.

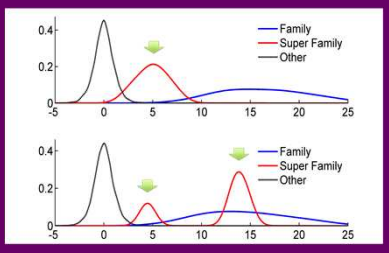
05 Align Hits



The resulting hits are automatically aligned to the original MSA using the pHMM for model columns and further sequence based methods elsewhere.

06 Results — Repeat >

The approach was tested with SCOP families. Improvements compared to conventional unsupervised purely sequence based HMM methods are shown.



After one iteration of steps 01-06, a good separation on family level is achieved. Iterating again with the compiled MSA shows an improved performance on the superfamily level.

References

[1] Optimierte Modelle zur Beschreibung von Proteinfamilien. Wegenkittl S, Lackner P, Auer F, Radlingmaier M. Proceedings der FFH 2009, 3. Forschungsforum der Österreichischen Fachhochschulen. FH Kärnten. 2009.

[2] UCSF Chimera - a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct; 25(13):1605-12.

[3] Improving model construction of profile HMMs for remote homology detection through structural alignment. Bernardes JS, Davila AM, Costa VS, Zaverucha G. BMC Bioinformatics. 2007 Nov 9;8:435.

[4] PSC++/ANTELOPE – An accurate approach to Multiple Protein Structure Comparison. Schwaiger CS, Zwifl T, Lirk G, Lackner P. Poster presentation at the 3DSIG 2007.