

## TEAM

Y. Igarashi, P. Cieplak, K. Gramatikoff,  
Y. Zhang, S. Boyd, A. Eroshkin,  
A. Osterman, A. Godzik, and J. Smith

## Introduction

PMAP is an integrated web-based bioinformatics resource developed at CPP. The main goals of this development are:

- provide a bioinformatics support for the on-going Technology Research Projects (TRP) in CPP;
- develop a platform supporting exploration of and reasoning about proteolytic pathways for CPP and the entire community;
- pursue research applications of this platform in computational analysis of proteolysis.

Understanding the underlying mechanisms of high selectivity of **protease-substrate recognition** is one of the central problems in the field of regulatory proteolysis that plays a critical role in mammalian cells.

In our studies we emphasize the importance of topological (structural) aspects of protease substrate specificity. Here we illustrate the current developments that address this problem by exploring structural properties of **protease active sites (Part I)** and **substrate cut-sites (Part II)**.

A combination of structural and cellular context of proteolytic events (captured within the framework of CutDB database) sets up the framework for the reconstruction and exploration of **proteolytic pathways and networks**.

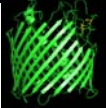
## Discovering New Proteases

Mining PDB for 'hidden' proteases using Active Site Recognition tool

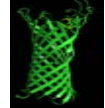
### candidate novel proteases

Serine protease	Aspartic protease	Metallo protease	Cysteine protease
6	334	83	476

Candidate novel aspartic protease 1NQG - outer membrane cobalamin transporter (BTUB) from *E. coli*. Active site found: Asp193-Asp195

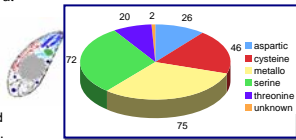


Precedent - a membrane protease 1I78 - membrane aspartic protease OMP1 from *E. coli*. Active site found: ASP83-ASP85



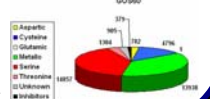
## Proteases Mining in *Toxoplasma gondii* Genome

- More than 60 mln in USA at risk of toxoplasmosis
- *T. gondii* proteases are important targets in host cell invasion
- On-going CPP project (M. Bogoy)
- Genome survey revealed 241 candidate proteases.



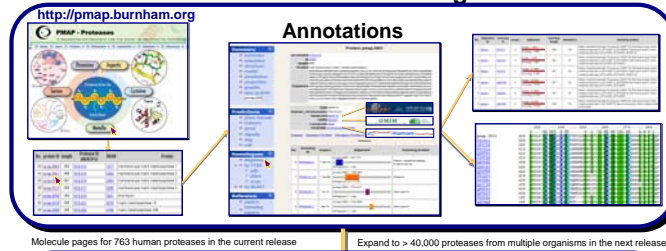
## Protease Mining in Global Ocean Sampling (GOS)

- Collaboration with C.Venter Institute: Environmental metagenomics using sequencing of ocean micro organisms.
- There are 222,000 potential proteases in GOS data. Data on the figure clustered at 60% sequence identity level.



## Protease Molecule Pages

Annotations

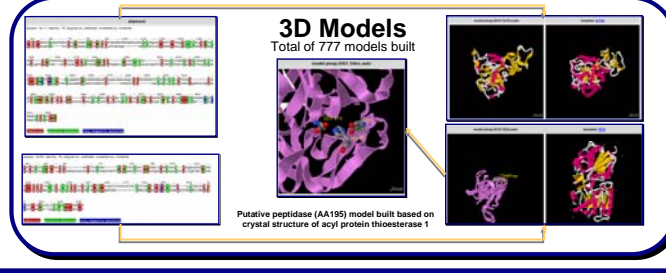


Annotations

Molecule pages for 763 human proteases in the current release | Expand to >40,000 proteases from multiple organisms in the next release

## 3D Models

Total of 777 models built



Putative peptidase (AA195) model built based on crystal structure of acyl protein thioesterase 1

## Annotation Pipeline

### Basic annotations

- Homologs (NR, Swiss-Prot)
- Domains (PFAM)
- Features (SignalP, TMHMM)
- Cellular location (by MultiLoc\* algorithm)
- Links (genes, expression, disease)



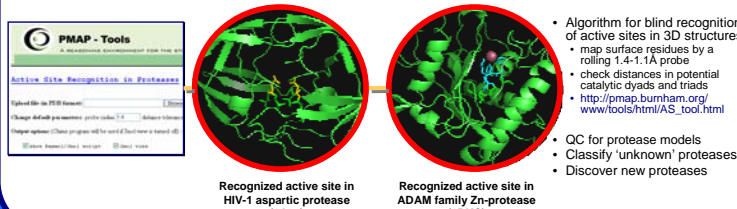
### Structural Modeling

- Protease 3D models were built using Modeller software (A. Sali)
- High quality within conserved cores
- QC program developed and applied

## Active Site Recognition and Modeling

PMAP - Tools

Active Site Recognition in Proteases



Recognized active site in HIV-1 aspartic protease (1A30)

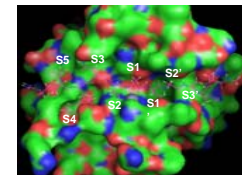
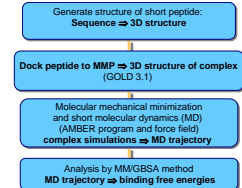
Recognized active site in ADAM family Zn-protease (1BKIC)

- Algorithm for blind recognition of active sites in 3D structures
  - map surface residues by a rolling 1.4-1.1Å probe
  - check distances in potential catalytic dyads and triads
- QC for protease models
- Classify 'unknown' proteases
- Discover new proteases

http://pmap.burnham.org/www/tool/html/AS\_tool.html

## Substrate Docking

Virtually docking approach to the analysis of protease substrate specificity



Binding pocket of the MMP2-10IB with peptide (RWNSLAAA) from docking experiment

**Experimentally observed cut-site has the best ΔG.**

Relative binding free energies computed for MMP2 and peptide PIPIAV.

Peptide	ΔΔG	ΔG
Ace-PIPIA   VAARA-Nme	0	-53.97
Ace-AP1PI   AVAAA-Nme	19.25	-34.72
Ace-AAPII   IAVAA-Nme	18.52	-35.45
Ace-AAAP1   IAVAA-Nme	23.44	-30.53
Ace-AAAP1   IPIAV-Nme	28.24	-25.73

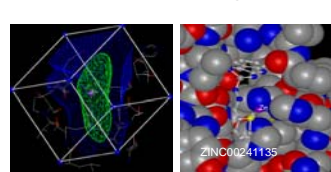
\* experimentally observed cleavage site

Sequence from phage display moved along initially docked 10AA scaffold and optimized using MM and MD methods. Assessment of binding free energy (entropy included) using MM/GBSA (Molecular mechanical/Generalized Born/Surface area) approach.

## Screening for New Inhibitors

Virtual screening for human MMP-2 small molecule inhibitors

First generation of possible inhibitors



ZINC00241135

Compound	Total Score	Steric	Desolvation	Acc	Donor
ZINC00241135	-80.564629	-74.0389	20.097867	0	-25.7236
ZINC0800205	-79.409714	-64.7961	14.826523	-6.89183	-32.36
ZINC0457717	-75.300468	-56.698	13.684132	-0.71545	-31.5711
ZINC0042119	-74.363083	-74.6483	5.036962	-0.00016	-2.29228
ZINC0021332	-73.854919	-69.2653	13.397898	-1.53697	-21.0264
ZINC0028207	-72.746376	-58.2493	10.421024	0	-24.9181
ZINC0043401	-72.369316	-65.1397	14.648669	-6.38275	-35.0314
ZINC0030989	-72.352539	-69.9812	7.20942	0	-8.88077
ZINC0102884	-72.108203	-73.8348	1.986262	0	0
ZINC0069370	-71.995628	-69.9464	10.132626	-0.65266	-20.1188
ZINC0812991	-71.531006	-69.2218	7.446223	0	-10.4058
ZINC0022071	-71.474686	-65.8605	6.795144	0	-13.7108
ZINC0813025	-71.341156	-59.9871	11.543693	-6.18113	-16.2966
ZINC0033788	-71.264657	-68.0255	16.375376	-1.99647	-30.548
ZINC0101789	-71.263115	-64.9981	15.045433	-1.19071	-19.1197
ZINC0869524	-71.112068	-60.3621	12.193067	-4.62886	-18.1742
ZINC0244616	-70.802666	-62.0762	8.697891	-0.60616	-8.69789
ZINC0028290	-70.765358	-62.3653	6.541446	-2.80153	-11.54
ZINC0116227	-70.670308	-65.9857	6.26743	0	-8.89032
ZINC0656718	-70.581543	-65.6826	10.884795	-2.35444	-13.4293

• Collaboration with Argonne National Lab (R. Stevens)

• A Docking Pipeline (FRED, DOCK5 and AutoDock)

• 1.3 million compounds from ZINC database

• IBM BlueGene supercomputer

Next steps - experimental testing of 500 best-scoring compounds