Viral Dark Matter: Protein Structure Prediction

Cathrine Aivati

May 3, 2012

Introduction

- Amino acid sequences – poorly match/do not match anything based upon sequence alone
- What about structural similarities?
- Structure determines function
- Secondary and Tertiary prediction software
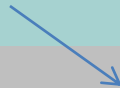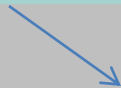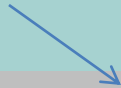- Supplementary to crystallography
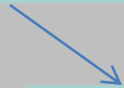
Pipeline

AA sequence

2° structure prediction (SSPro 3 and 8 class)

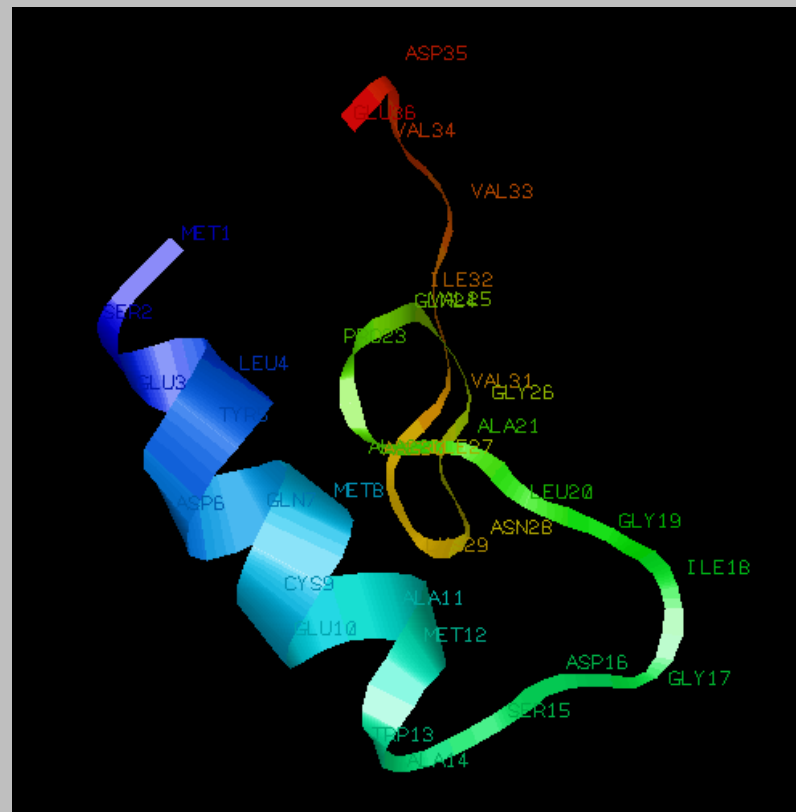**Choose 36 AA window**

PEPFOLD 3° prediction

PDBe FSSP (3D alignment)

Evaluation

Preliminary Results

- PDF (control) - 3° structure prediction and comparison from first 36 AA's in PDBe Fold = metaloprotease

Preliminary Results

Sugar Nucleotide Epimerase

AAs   1-36   → Tautomerase

AAs 37-73   → Ribosomal protein

AAs 74-110   → Ubiquitase domain?

AAs 111-147 → RAP74 (subunit of TF; recruits RNA Pol II)

AAs 148-184 → Ribosomal protein

AAs 185-211 → KIA7 ("prebiotic")/metaloprotease

AAs 212-248 → Ribosome protein

AAs 249-285 → LXR-alpha??

AAs 286-302 → EIF4E (TF?)

Aas 303-317 → Exportin

Additional Results

Also have tertiary prediction alignments for:

- Thioredoxin

- gp204

- Phosphate Starvation Induced Protein

- Car9 sequence 6683_4_1

- AA's 74-110 → DOC (death on curing) protein found in enterobacteria phage

- Protein involved in preventing host death…interesting

Future Direction

- Optimize selection of "windows"

- Ideal to have full length predictions (3DPro)

- Input VCIDs associated with physiological data