CASP11 Target Classification

Special Thanks to:

Lisa Kinch, Wenlin Li, Dustin Schaeffer, and Hua Cheng from the Grishin Lab (FM Assessors)

Roland Dunbrack (TBM Assessor)

Andriy Kryshtafovych, Bohdan Monastyrskyy and Krzysztof Fidelis from the Prediction Center

Organizers
CASP11 Assessment would be impossible without

The Prediction Center:
Andriy Kryshtafovych, Krzysztof Fidelis, and Bohdan Monastyryskyy

The ECOD Classification:
Hua Cheng, Dustin Schaeffer, Yuxing Liao, Jimin Pei, Shuoyong Shi, Bong-Hyun Kim, and Nick Grishin
CASP11 Overview in Numbers

• 100 Targets released (45 server only)
  3 cancelled for all (3 cancelled for manual)
  4 have no structure
  7 for interaction prediction only
  =86 targets to split into domains

• 37 refinement targets
• 71 assisted prediction targets
• 38 CASP ROLL Targets (9 overlap)
• 207 registered groups (84 servers)
Groups with TS predictions (144 groups/44 servers)
  =36776 total models to evaluate
Contact assisted predictions (29 groups/6 servers)
  =5314 total models to evaluate

www.predictioncenter.org
CASP11 Overview in Numbers

Targets Contributed by:

- Joint Center for Structural Genomics (JCSG): 32
- Genomics Consortium (SGC): 4
- Midwest Center for Structural Genomics (MCSG): 8
- Northeast Structural Genomics Consortium (NESG): 5
- New York Structural Genomics Research Center (NYSGRC): 6
- Non-SGI research Centers and others (Others): 40
- Seattle Structural Genomics Center for Infectious Disease (SSGCID): 4
- NatPro PSI:Biology (NatPro): 1

www.predictioncenter.org
CASP11 Targets for Evaluation of Interactions (7 Targets)
CASP11 Interaction Predictions

T0787 & T0788: HIV-1 Env Complex

GP41 + GP120 → Heterodimer → Hexameric complex
CASP11 Interaction Predictions

T0797 & T0798: PKGII leucine zipper Rab11b complex

Zipper assembly + Rab11b + PKGII = Crystal unit
CASP11 Interaction Predictions

T0825: synthetic propeller construct

Chain A + Chain B → Homodimer
CASP11 Interaction Predictions

T0825: synthetic propeller construct

Chain A
4-blades

Chain B
5-blades

Homodimer
9-blade propeller
CASP11 Interaction Predictions

T0840 & T0841: RON Receptor Tyrosine kinase/ Macrophage Stimulating Protein (MSP)
**CASP11 Targets in Numbers**

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Number of Targets</strong></td>
<td>100</td>
</tr>
<tr>
<td>Cancelled for all (manual)</td>
<td>3 (3*)</td>
</tr>
<tr>
<td>No structure</td>
<td>4</td>
</tr>
<tr>
<td>Structures for Interaction only</td>
<td>7</td>
</tr>
<tr>
<td><strong>Remaining Targets for all (server only)</strong></td>
<td>46 (40)</td>
</tr>
<tr>
<td>Viral (phage)</td>
<td>7 (2)</td>
</tr>
<tr>
<td>Singletons (&lt;5 hits)</td>
<td>5 (4)</td>
</tr>
<tr>
<td>Environmental metagenome</td>
<td>1</td>
</tr>
<tr>
<td>Engineered</td>
<td>2</td>
</tr>
<tr>
<td><strong>Special sequence features</strong></td>
<td></td>
</tr>
<tr>
<td>Targets with Signal peptide</td>
<td></td>
</tr>
<tr>
<td>Transmembrane Targets</td>
<td></td>
</tr>
<tr>
<td>Targets with N-HisTag (-C)</td>
<td></td>
</tr>
<tr>
<td><strong>Evaluation Units for all (server only)</strong></td>
<td>78 (48)</td>
</tr>
<tr>
<td>TBM (server only)</td>
<td></td>
</tr>
<tr>
<td>FM (server only)</td>
<td></td>
</tr>
</tbody>
</table>
## CASP11 Targets in Numbers

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Number of Targets</strong></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Cancelled for all (manual)</td>
<td>3</td>
<td>(3*)</td>
</tr>
<tr>
<td>No structure</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Structures for Interaction only</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td><strong>Remaining Targets for all (server only)</strong></td>
<td>46</td>
<td>(40)</td>
</tr>
<tr>
<td>Viral (phage)</td>
<td>7</td>
<td>(2)</td>
</tr>
<tr>
<td>Singletons (&lt;5 hits)</td>
<td>5</td>
<td>(4)</td>
</tr>
<tr>
<td>Environmental metagenome</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Engineered</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>Special sequence features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targets with Signal peptide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmembrane Targets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targets with N-HisTag (-C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Evaluation Units for all (server only)</strong></td>
<td>78</td>
<td>(48)</td>
</tr>
<tr>
<td>TBM (server only)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FM (server only)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### CASP11 Targets in Numbers

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Number of Targets</strong></td>
<td>100</td>
</tr>
<tr>
<td>Cancelled for all (manual)</td>
<td>3 (3*)</td>
</tr>
<tr>
<td>No structure</td>
<td>4</td>
</tr>
<tr>
<td>Structures for Interaction only</td>
<td>7</td>
</tr>
<tr>
<td><strong>Remaining Targets for all (server only)</strong></td>
<td>46 (40)</td>
</tr>
<tr>
<td>Viral (phage)</td>
<td>7 (2)</td>
</tr>
<tr>
<td>Singletons (&lt;5 hits)</td>
<td>5 (4)</td>
</tr>
<tr>
<td>Environmental metagenome</td>
<td>1</td>
</tr>
<tr>
<td>Engineered</td>
<td>2</td>
</tr>
<tr>
<td><strong>Special sequence features</strong></td>
<td></td>
</tr>
<tr>
<td>Targets with Signal peptide</td>
<td></td>
</tr>
<tr>
<td>Transmembrane Targets</td>
<td></td>
</tr>
<tr>
<td>Targets with N-HisTag (-C)</td>
<td></td>
</tr>
<tr>
<td><strong>Evaluation Units for all (server only)</strong></td>
<td>78 (48)</td>
</tr>
<tr>
<td>TBM (server only)</td>
<td></td>
</tr>
<tr>
<td>FM (server only)</td>
<td></td>
</tr>
</tbody>
</table>

![Image of a protein structure]
## CASP11 Targets in Numbers

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Number of Targets</strong></td>
<td>100</td>
</tr>
<tr>
<td>Cancelled for all (manual)</td>
<td>3 (3*)</td>
</tr>
<tr>
<td>No structure</td>
<td>4</td>
</tr>
<tr>
<td>Structures for Interaction only</td>
<td>7</td>
</tr>
<tr>
<td><strong>Remaining Targets for all (server only)</strong></td>
<td>46 (40)</td>
</tr>
<tr>
<td>Viral (phage)</td>
<td>7 (2)</td>
</tr>
<tr>
<td>Singletons (&lt;5 hits)</td>
<td>5 (4)</td>
</tr>
<tr>
<td>Environmental metagenome</td>
<td>1</td>
</tr>
<tr>
<td>Engineered</td>
<td>2</td>
</tr>
<tr>
<td><strong>Special sequence features</strong></td>
<td></td>
</tr>
<tr>
<td>Targets with Signal peptide</td>
<td></td>
</tr>
<tr>
<td>Transmembrane Targets</td>
<td></td>
</tr>
<tr>
<td>Targets with N-HisTag (-C)</td>
<td></td>
</tr>
<tr>
<td><strong>Evaluation Units for all (server only)</strong></td>
<td>78 (48)</td>
</tr>
<tr>
<td>TBM (server only)</td>
<td></td>
</tr>
<tr>
<td>FM (server only)</td>
<td></td>
</tr>
</tbody>
</table>
CASP11 Targets in Numbers

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Targets</td>
<td>100</td>
</tr>
<tr>
<td>Cancelled for all (manual)</td>
<td>3 (3*)</td>
</tr>
<tr>
<td>No structure</td>
<td>4</td>
</tr>
<tr>
<td>Structures for Interaction only</td>
<td>7</td>
</tr>
<tr>
<td>Remaining Targets for all (server only)</td>
<td>46 (40)</td>
</tr>
<tr>
<td>Viral (phage)</td>
<td>7 (2)</td>
</tr>
<tr>
<td>Singletons (&lt;5 hits)</td>
<td>5 (4)</td>
</tr>
<tr>
<td>Environmental metagenome</td>
<td>1</td>
</tr>
<tr>
<td>Engineered</td>
<td>2</td>
</tr>
<tr>
<td>Special sequence features</td>
<td></td>
</tr>
<tr>
<td>Targets with Signal peptide</td>
<td>34</td>
</tr>
<tr>
<td>Transmembrane Targets</td>
<td>3</td>
</tr>
<tr>
<td>Targets with N-HisTag (-C)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Evaluation Units for all (server only)</td>
<td>78 (48)</td>
</tr>
<tr>
<td>TBM (server only)</td>
<td></td>
</tr>
<tr>
<td>FM (server only)</td>
<td></td>
</tr>
</tbody>
</table>

- = 33 Signal peptides (38%)
- = 5 Targets with TMH
  - 1 N-Terminal TMH = SP (T0763)
  - 1 Incorrect TMH (T0793)
CASP11 Domain Definition and Classification

(86 Targets)
CASP11 Target Classification

Domain (Evaluation Unit) Definition:

1) Split targets based on visual domains
2) Splits also consider sequence/template information
3) Generate scores for models based on split/no split & compare performance (Grishin Plots)
4) Generate test splits for difficult cases & re-evaluate
CASP11 Domain Definition

T0759: Split targets based on visual domains
CASP11 Domain Definition

T0759: Consider sequence/template information

Conserved Domains (NCBI CDD)
CASP11 Domain Definition

T0759: Consider sequence/template information

1lm5
ECDO H-group:
β-hairpin
-α-hairpin repeat

Template homolog

Closest Templates
(by GDT_TS)

3cwx
ECDO H-group:
protein CagD

Template analog
T0759: Score models & Evaluate based on performance

Treat as two evaluation units
= T0759D1 & T0759D2
T0786: Split based on internal duplication of fold
CASP11Domain Definition

T0786: Consider Template

Homologous template 2q4h
ECOD H-groups:
N-terminal HIT-related
C-terminal HIT-related
CASP11 Domain Definition

T0786: Score models & Evaluate based on performance

T0786 “Grishin Plot”

No shift, No split
CASP11 Domain Definition

T0786: Score models & Evaluate based on performance

Treat as single evaluation unit = T0786D1

T0786 “Grishin Plot”

No shift, No split
CASP11 Domain Definition

T0775: Difficult case

Repeating units

C-terminus
CASP11 Domain Definition

T0775: Consider Sequence/Templates

Partial HH Pred
Template: 4a0t
CASP11 Domain Definition

T0775: Consider Sequence/Templates

Other Phage fiber
Templates: 3qr8
Start with meander
CASP11 Domain Definition

T0775: Generate Test Domains

Test Domains:
Split out all?
CASP11 Domain Definition

T0775: Generate Test Domains

Test Domains:
Split out all?

Too short for evaluation (<30 residues)
CASP11 Domain Definition

T0775: Consider Sequence/Templates

Shift to reflect Template bounds
T0775: Consider Sequence/Templates

Shift to reflect Template bounds
## CASP11 Targets in Numbers

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Number of Targets</strong></td>
<td>100</td>
</tr>
<tr>
<td>Cancelled for all (manual)</td>
<td>3 (3*)</td>
</tr>
<tr>
<td>No structure</td>
<td>4</td>
</tr>
<tr>
<td>Structures for Interaction only</td>
<td>7</td>
</tr>
<tr>
<td><strong>Remaining Targets for all (server only)</strong></td>
<td>46 (40)</td>
</tr>
<tr>
<td>Viral</td>
<td>7</td>
</tr>
<tr>
<td>Singletons (&lt;5 hits)</td>
<td>5 (4)</td>
</tr>
<tr>
<td>Environmental metagenome</td>
<td>1</td>
</tr>
<tr>
<td>Engineered</td>
<td>2</td>
</tr>
<tr>
<td><strong>Special sequence features</strong></td>
<td></td>
</tr>
<tr>
<td>Targets with Signal peptide</td>
<td>34</td>
</tr>
<tr>
<td>Transmembrane Targets</td>
<td>3</td>
</tr>
<tr>
<td>Targets with N-HisTag (-C)</td>
<td>2 (4)</td>
</tr>
<tr>
<td><strong>Evaluation Units for all (server only)</strong></td>
<td>78 (48)</td>
</tr>
<tr>
<td>TBM (server only)</td>
<td></td>
</tr>
<tr>
<td>FM (server only)</td>
<td></td>
</tr>
</tbody>
</table>

126 Domains for classification
CASP11 Target Classification by ECOD
Evolutionary Classification of Protein Domains: ECOD

The ECOD group

Dustin Schaeffer
Hua Cheng
Jimin Pei
Shuoyong Shi
Yuxing Lao
Bong-Hyun Kim

Evolutionary Classification of Protein Domains: ECOD

The ECOD classification scheme

Topology

- Architectures
  - X-groups

Evolution

- H-groups
  - T-groups
  - F-groups

Evolutionary Classification of Protein Domains: ECOD

The ECOD classification scheme

- Topology
  - X-groups
  - H-groups
  - T-groups
  - F-groups

- Evolution

Evolutionary Classification of Protein Domains: ECOD

The ECOD classification scheme

- **Architectures**
  - X-groups
  - H-groups
  - T-groups
  - F-groups

Evolutionary Classification of Protein Domains: ECOD

The ECOD classification scheme

- Topology
  - Architectures
    - X-groups
  - H-groups
    - T-groups
    - F-groups

Evolutionary Classification of Protein Domains: ECOD

The ECOD classification scheme

Evolutionary Classification of Protein Domains: ECOD

The ECOD classification scheme

Topology
- Architectures
  - X-groups
  - H-groups
  - T-groups
  - F-groups (PFAM)

Evolution

RimM

CDC48_N

ECOD Architectures of CASP11 Targets

- α/β barrels
- α+β obligate
- α+β complex
- α+β 4-layers
- α+β 3-layers
- Mixed α+β and α/β
- Few SSEs
- β barrels
- β meanders
- β sandwiches
- β obligate
- β complex
- α arrays
- α bundles
- α superhelices
- α obligate
ECOD Architectures of CASP11 Targets

- mixed α + β and α/β
- few SSEs
- β barrels
- β meanders
- α/β sandwiches
- β sandwiches
- β obligate
- α + β obligate
- α + β complex
- α + β 4-layers
- α + β 3-layers
- α + β 2-layers

Legend:
- All α
- All β
- α/β
- α + β
ECOD Architectures of CASP11 Targets

**α arrays**

- Helix-turn-helix
  - F:T0792, H:T0793

**α bundles**

- Transmembrane heme-binding four-helical bundle
  - H:T0836

- AhpD-like
  - X:T0837

- 6-phosphogluconate dehydrogenase C-terminal domain-like
  - F:T0813, T0851

- GAT-like domain
  - F:T0839

- Spectrin repeat-like
  - X:T0831
ECOD Architectures of CASP11 Targets

Top LGA_S template:
alpha complex topology
alpha-helical domain in nickel-iron hydrogenase
3rgwL

Spectrin repeat-like
X:T0831
ECOD Architectures of CASP11 Targets

Top LGA_S template:
alpha complex topology
alpha-helical domain in nickel-iron hydrogenase
3rgwL

ECOD assignment:
alpha bundle
Spectrin repeat-like
3zcjA1

41.08
LGA_S

42.8
LGA_S

Spectrin repeat-like
X:T0831
ECOD Architectures of CASP11 Targets

α superhelices

ARM repeat
H:T0810, T0827d1, T0828, T0852

ARM repeat
H:T0770

Repetitive alpha hairpin
X:T0831

GST-C
F:T0849

ARM repeat
F:T0821, T0839

alpha/alpha toroid
F:T0835
ECOD Architectures of CASP11 Targets

\(\alpha\) complex topology

No Template
T0827d2

\(\alpha\) obligate multimer

No Template
T0777

No Template
T0820d1
ECOD Architectures of CASP11 Targets

**β sandwiches**

- **Concanavalin_A-like:**
  - duf4425  **F:T0784**

- **Concanavalin_A-like:**
  - galactose binding domain-like  **H:T0808, T0812, T0822**

- **Immunoglobulin-related:**
  - Ig, fibronectinIII, E-set, PapD  **H:T0814d1-d3, T0828, T0845, T0857**

- **Domain in virus attachment proteins**
  -  **H:T0785, T0795, T0802, T0804**

- **Concanavalin_A-like:**
  - lectins/glucanases  **F:T0772, T0856; H:T0808**

- **Immunoglobulin-related:**
  - Prealbumin  **F:T0774d1&d2**
ECOD Architectures of CASP11 Targets

**β barrels**

- **OB-fold**
  - F:T0760d1
- **Lipocalins/Streptavidin**
  - X:T0782
- **SH3?**
  - X:T0793d1

**β complex topology**

- **Aerolisin/ETX pore-forming domain-related**
  - F:T0796

**β meander**

- **Uncharacterized protein BF3112**
  - F:T0833
ECOD Architectures of CASP11 Targets

Pathogenicity island AAA ATPase similar to MCM (Minichromosome maintenance protein) helicases

Potential function:

SH3:chromo DNA bound template 1bnz

LGA_S 45.4

SH3:chromo Top LGA_S template 1g6zA1
ECOD Architectures of CASP11 Targets

β Obligate multimers

- beta-propeller (7-blade)  
  T:T0845

- Phage fibre proteins  
  H:T0775, T0799

- Triple_beta-spiral  
  H:T0804

- Leucine-rich_repeats  
  F:T0768

- Pectin_lyase-like  
  H:T0800
ECOD Architectures of CASP11 Targets

β Obligate multimers

Phage fibre proteins
H:T0775, T0799

Triple_beta-spiral
H:T0804
ECOD Architectures of CASP11 Targets

β Obligate multimers
ECOD Architectures of CASP11 Targets

**α+β Obligate multimers**

- No template: T0820
- beta-hairpin-alpha-hairpin repeat: F:T0759

**α+β Complex topology**

- No template: T0834
- Insert subdomain of RNApol α: H:T0780
ECOD Architectures of CASP11 Targets

Sensor domain of Histidine Kinase (Duf3404): 4R7Q

Arm repeat?

- Profilin-like?
- FAD-binding/transporter-associated domain-like?

No template

T0834
ECOD Architectures of CASP11 Targets

Sensor domain of Histidine Kinase (Duf3404): 4R7Q

1) Profilin-like
   - LuxQ sensor domain of histidine kinase
   - Also top GDT template

2) FAD-binding/transporter-associated domain-like

D1

No template
T0834
ECOD Architectures of CASP11 Targets

Sensor domain of Histidine Kinase (Duf3404): 4R7Q

Arm repeat?

HorA hit

2awiA

Also top GDT template (4gyo)

No template

T0834
ECOD Architectures of CASP11 Targets

$\alpha+\beta$ 3 layers

HIT-like $\text{H:T0786}$

Mog1p/PsbP-like $\text{H:T0838}$

Profilin-like $\text{X:T0834}$

$\alpha+\beta$ 4 layers

Carbon-nitrogen hydrolase $\text{F:T0794}$
ECOD Architectures of CASP11 Targets

\[ \alpha + \beta \] 2 layers

- **Alpha-beta plaits**
  - \( X: T0765, T0769, T0773 \)

- **Alpha-beta plaits**
  - \( F: T0805, T0817 \)

- **Beta-Grasp**
  - \( X: T0853 \)

- **C-terminal domain in some PLP-dependent transferases**
  - \( F: T0819, T0843 \)

- **Cystatin-like**
  - \( X: T0761, H: T0766, T0781, T0818, T0848, F: T0815 \)

- **Hydrogenase expression protein**
  - \( X: T0789, T0790, T0791 \)
ECOD Architectures of CASP11 Targets
\(\alpha+\beta\) 2 layers (cont’d)

Pili subunits
H:T0803

NE0471 N-terminal domain-like
X:T0829

No template
T0855

TolA/TonB C-terminal domain
X:T0771

Uncharacterized protein
ZP02042476.1 X:T0763

Hypothetical protein
RUMGNA 01148 H:T0767
ECOD Architectures of CASP11 Targets

\(\alpha/\beta\) 3-layered sandwiches

Other Rossmann-like X group:

- Nucleotide-diphospho-sugar transferases \textit{F:T0783}
- \(\alpha/\beta\)-Hydrolases \textit{F:T0764, T0852}
- PLP-dependent transferases \textit{F:T0801, T0843}
- Alkaline phosphatase-like \textit{F:T0826}
- Uracil-DNA glycosylase-like \textit{X:T0806}
- Periplasmic binding protein-like II \textit{F:T0762}
ECOD Architectures of CASP11 Targets
\(\alpha/\beta\) 3-layered sandwiches (cont’d)

- Rossmann-like: X:T0783, F:T0813, T0851
- HAD-like: F:T0854
- Phospholipase D /nuclease: H:T0847
- SGNH hydrolase: F:T0776
- STT3/PgIB/AgIB core domain: H:T0830
- P-loop nucleoside triphosphate hydrolases: H:T0793
ECOD Architectures of CASP11 Targets

**α/β barrels**

TIM barrels \( F: T0807, T0811, T0853 \)

**Mixed α+β and α/β**

Thioredoxin-like

\( F: T0849 \)

**Few SSEs**

Glycoside hydrolase/deacetylase

\( H: T0810 \)

No template

\( T0793d2 \)
ECOD Classification of CASP11 Targets

- H/T-group
- Family
- X-group
- New
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

![Graph showing the correlation between Difficulty (Template LGA_S) and Average Server GDT_TS for targets T0774-D1 and T0759-D2.](Image)
ECOD Classification of CASP11 Targets

ECOD F: DUF3988

LGA 39.37

T0774-D1

T0759-D2

Average Server GD

Difficulty (Template LGA_S)
ECOD Classification of CASP11 Targets

1lm5 Template for both domains

$\text{LGA}=41.8$

To D2

ECOD F: Plectin

Average Server GD

Difficulty (Template LGA_S)
**ECOD Classification of CASP11 Targets**

- **T0761 & T0781**: Cystatin-like
- **T0789, T0790 & T0791**: hydrogenase expression protein-like
- **T0808**: Concanavalin_A-like
- **T0814 (triplicate)**: Immunoglobulin-like
- **T0817**: Other Rossmann-like
- **T0853**: beta-Grasp
- **T0854**: HAD domain-like

Other duplication examples that could evolve quickly:

- **ECOD F: Plectin**
- **1lm5**: Template for both domains, \(LGA=41.8\)
- **To D2**
ECOD Classification of CASP11 Targets

**Special Cases: Duplications**

T0814: Triplication of IG-like domains (T group: Immunoglobulin/ Fibronectin type III/ E set domains/ PapD-like)
ECOD Classification of CASP11 Targets

**Special Cases: Duplications**

T0814: Triplication of IG-like domains (T group: Immunoglobulin/ Fibronectin type III/ E set domains/ PapD-like)

Permuted with strand invasion (Interrupted)
Correlation of Performance with Difficulty

ECOD Classification of CASP11 Targets

- T0775: Phage T4 tail fibre
- T0795: Raptor Adenovirus 1 fibre head
- T0799: Phage T5 tail fibre
- T0804: Mastadenovirus fibre head

= Fast Evolving
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

T0773 & T0769: engineered alpha-beta plaits
T0765: alpha-beta plaits

= Predicting Analogs
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

Analogous Templates Close to structures

New Folds...
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

Target 820-D1

Template (1yxiA)

LGA 52.4

Analogous Templates Close to structures

New Folds...
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

New Folds...
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

New Folds...
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

New Folds...
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

Difficult Classification Cases...
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

Difficult Classification Cases...

T0824-D1
NucB DNase

Template (1g8tA)
NUCLEASE SM2

Average Serv

Difficulty (Template LGA_S)
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

Cases...

T0824

T0820-D2

Template (1dmw)

T0832-D1

PHENYLALANINE HYDROXYLASE

Average Score

0 20 40

Difficulty (Template LGA_S)

0 20 40 60 80 100
ECOD Classification of CASP11 Targets

Correlation of Performance with Difficulty

T0820-D2 - Template (2cmjB)

Dimerization domains

Average Serve vs Difficulty (Template LGA_S)
CASP11 Target Classification into TBM/FM Evaluation Categories
CASP11 Domain Classification

Differences between Free Modeling (FM) and Template-Based Modeling (TBM) categories

• Are there sequence detectable templates?

• What is the distance to closest template?

• How do the server groups perform?
Differences between Free Modeling (FM) and Template-Based Modeling (TBM) categories

• Are there sequence detectable templates?
  - HH-PRED targets: probabilities/coverage
  - Server declared ECOD-related parents

• What is the distance to closest template?

• How do the server groups perform?
Differences between Free Modeling (FM) and Template-Based Modeling (TBM) categories

• Are there sequence detectable templates?
  HHPRED targets: probabilities/coverage
  Server declared ECOD-related parents

• What is the evolutionary distance to closest template?
  ECOD classification
  Target to template LGA_S

• How do the server groups perform?

CASP11Domain Classification
CASP11 Domain Classification

Differences between Free Modeling (FM) and Template-Based Modeling (TBM) categories

• Are there sequence detectable templates?
  HHpred targets: probabilities/coverage
  Server declared ECOD-related parents

• What is the evolutionary distance to closest template?
  ECOD classification
  Target to template LGA_S

• How do the server groups perform?
  Scores above random, top scores, first scores, etc.
CASP11 Targets: TBM or FM?

Correlation of Performance with Difficulty

Where do we draw the line?
CASP11 Targets: TBM or FM?

Correlation of Performance with Difficulty

"Expert" Opinion & ECOD Mapping
CASP11 Targets: TBM or FM?

Correlation of Performance with Difficulty

Play with Plots
Playing with Plots
Playing with Plots: The Results

• **Template Scores**
  1) Top HHpred probability to homologous template
  2) Target to template LGA_S

• **Performance Scores**
  3) GDT average of server models
  4) GDT average of first server models above random
  5) Number of first server models above random

Convert each to Z-scores & Sum
Incorporate 5 scores for a better separation

more objective way to draw the line?

- Z-score sum of 5 measures

- confident TBM
- notSure
- confident FM

LGA_S to template
Use machine learning techniques to draw a line

- SVM (linear kernel)
- Logistic regression

Z-score sum of 5 measures

LGA_S to template

- confident TBM
- notSure
- confident FM
Discuss Targets in the ‘Danger Zone’

- SVM (linear kernel)
- Logistic regression

LGA_S to template

Z-score sum of 5 measures

- Green: confident TBM
- Orange: notSure
- Brown: confident FM
Discuss Targets in the ‘Danger Zone’

SVM (linear kernel)

T0853-D2 should be TBM, It is a duplication!

LGA S to template

Z-score sum of 5 measures
Discuss Targets in the ‘Danger Zone’

T0793-D5 should be FM due to poor performance; top template 1v3f is not very similar.

Logistic regression

SVM (linear kernel)

T0793-D5

LGA_S to template

Z-score sum of 5 measures

Avg GDT ~20

LGAS ~50

Template 1v3f

confident TBM
notSure
confident FM

due to poor performance; top template 1v3f is not very similar
CASP11 Target Classification: The Results

Z-score sum of 5 measures

LGA_S to template