Predictive Landscape of CASP

Nick V. Grishin

Howard Hughes Medical Institute and Departments of Biophysics and Biochemistry
University of Texas Southwestern Medical Center
Dallas, Texas, USA
A Random Walk on Predictive Landscape of CASP

Nick V. Grishin

Howard Hughes Medical Institute and Departments of Biophysics and Biochemistry
University of Texas Southwestern Medical Center
Dallas, Texas, USA
Why is CASP needed?

1. Prevent someone from premature declaring “folding problem” solved

2. Reveal the best achievements in structure modeling and stimulate further developments

3. Popularize structure modeling
Why is CASP needed?

1. Prevent someone from premature declaring “folding problem” solved

2. Reveal the best achievements in structure modeling and stimulate further developments

3. Popularize structure modeling
CASP5
“Fold Recognition” Assessors

Lisa Kinch and Nick Grishin

The Assessment Team

60% of the Grishin Lab

Yuan Qi
S. Sri Krishna
Indraneel Majumdar
Ruslan Sadreyev
Hua Cheng
Jamie Wrabl
Jimin Pei
It’s good to be a CASP assessor, because you get to see things
... and state-of-the-art models:
but ... it's **scary**, because things can happen to you ...
Generating data…
or not
Snapshot of Krishna’s computer screen during CASP data analysis

... and the result can be ...
Our main contribution to CASP
Shape of a target table:
FR category is disappearing

<table>
<thead>
<tr>
<th>TARGET</th>
<th>DIFFICULTY</th>
<th>CM</th>
<th>CM/FR</th>
<th>FR</th>
<th>FR/H</th>
<th>FR/A</th>
<th>NF/FR</th>
<th>NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td></td>
<td>T0099</td>
<td>(A1-A127)</td>
<td>T0111_2</td>
<td>T0089_1</td>
<td>T0096_1</td>
<td>T0086</td>
<td>T0086</td>
</tr>
<tr>
<td>T0111_1</td>
<td>(A1-A127)</td>
<td>T0113</td>
<td>(200-390)</td>
<td>T0089_3</td>
<td>T0096_2</td>
<td>T0102</td>
<td>T0087_1</td>
<td></td>
</tr>
<tr>
<td>T0122</td>
<td></td>
<td>T0090_2</td>
<td>(A58-A209)</td>
<td>T0096_1</td>
<td>T0102</td>
<td>T0087_2</td>
<td>T0090_1</td>
<td></td>
</tr>
<tr>
<td>T0123</td>
<td></td>
<td>T0092</td>
<td>(A6-A79)</td>
<td>T0102</td>
<td>T0102</td>
<td>T0087_2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0128_1</td>
<td>(A1-A240)</td>
<td>T0103</td>
<td>(19-260)</td>
<td>T0102</td>
<td>T0102</td>
<td>T0091</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0128_2</td>
<td>(A100-A222)</td>
<td>T0112_1</td>
<td>(A4-A149)</td>
<td>T0102</td>
<td>T0102</td>
<td>T0106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0125</td>
<td></td>
<td>T0110</td>
<td>(A296-A351)</td>
<td>T0102</td>
<td>T0102</td>
<td>T0106</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CASP4**

**FR**
Categories in CASP

Do we need any?

meta-groups that try to predict everything

Categories by methods

do we need to encourage new method development?

Categories by targets

the world is changing, categories do not
Categories in CASP

**CM**
- **TR** – template refinement
  - template is obvious, alignment is not a problem. Goals:
  - backbone refinement;
  - side-chain modeling

**FR**
- **TM** – template modeling
  - template can be found by sequence or FR techniques.
  - alignment;
  - modeling of differences

**NF**
- **FM** – free modeling
  - template-independent
  - assemble protein-like models
Grand Ideas from CASPs:

1. Computers help structure prediction: no more paper models
2. Knowledge-based potentials work better
3. Local “threading” and fragment assembly (Baker)
4. Averaging and consensus methods work: meta-servers (Ginalski-Rychlewski)
5. Sequence profile methods are as (or more powerful) than threading (Söding)
6. Jamming poorly similar templates together helps (Skolnick-Zhang)
Progress(ion) in structure prediction

Art
Intuition

Science
Hard numbers
Structure prediction
CASP1

Art

CASP1 - when people cared about protein structures
Structure prediction
CASP10

Is this a triumph?

CASP10 - when people care about algorithms
Structure prediction
CASP10

Science
Structure prediction
CASP10

Engineering
Main CASPx results

1. no new grand ideas
2. not much progress
3. some progress in moving away from templates

Tweakers vs. Transformers

Is it just engineering, or there is more to it?

Structure prediction today:

*it is not science, it is engineering!*
What is CASP about?

1. Physics;
2. Biology;

What are new basic principles we learned?

Protein name, not target number.
What is CASP about?

1. Physic(s);

2. Biology;


What are new basic principles we learned?

Protein name, not target number
Evolution of the CASP Predictor...
Maybe structure prediction is a software engineering problem?

while I don’t know the answer, I tend to think that it is not
Main messages from recent CASPs

- Not that much progress in modeling since CASP5, i.e. one needs to look hard for progress = inching towards truth?

- Sometimes (accidentally?) predictors get good models for some targets, but not consistently, and they cannot really tell when these good models are obtained
General assessment of the (sad) situation:

(quotes from the best predictors in CASP9)

"A part of this problem is that it has not been a lot of progress during the last few years and that the progress that appeared is to a large degree due to tuning. This is not very exiting."

"I believe we are stuck in a very deep local minimum."

"the ideas that resulted in rapid progress some ten years ago have exhausted their potential."

"The present dead-lock situation in CASP comes from the fact that almost all participants apply the same methods, there are no innovators."

"In most cases we can either do modeling with psiblast or not at all."
General assessment of the (sad) situation:

(quotes from the best predictors in CASP9)

"A part of this problem is that it has not been a lot of progress during the last few years and that the progress that appeared is to a large degree due to tuning. This is not very exiting."

"I believe we are stuck in a very deep local minimum."

"the ideas that resulted in rapid progress ten years ago have exhausted their potential."

"The present dead-lock situation in CASP comes from the fact that almost all participants apply the same methods, there are no innovators."

"In most cases we can either do modeling with psiblast or not at all."
The bottom line: what’s best in structure prediction?

Conceptually, CASP community feeds on:

- HHpred
- ROSETTA
- Zhang-server
- PSI-pred

!!! Sequence methods work best for structure modeling
A typical highlight of CASPs:
CASP9: the largest “improvement” over the closest template structure

Amazing models: 12/13/2010

“improve” GDT by 44% compared to the best template
We did **not** see amazing models at the CASP10 meeting.
Sad, but true ...

Why is that?

Maybe FM category is not that important?

Most structures are known: ~70% proteins in a genome can be homology-modeled
We got 2 clean FM results:

1. Servers (380, 428, Zhang) and (321 Rosetta) are tied and statistically distinguished from the rest of the pack.

2. "Manual" groups who performed better were re-scoring server models and refining them - recycling of server models is a "winning" strategy for this FM casp.
How to "win" casp?

"the best servers just performed a grain better on each target".

Inching towards success by avoiding failure. How wise is it?

Is your rank higher because you make good predictions, or because you are better at avoiding failures?
CASP: competition or experiment?

CASP survives because it is a competition

CASP is interesting
because it is a (psychological?) experiment
CASP: what kind of experiment?

Definitely *in vivo*

Why? Because anything goes.

*Mouse lives, mouse eats, mouse dies*

Does not matter how you got the model, only the final result matters
**CASP: in vivo vs. in vitro**

*In vivo* has obvious advantages: it is clean and unrestricted.

*In vivo* has obvious drawbacks: it is hard to understand what actually works and why.

*In vitro* components of the experiment could help in pinpointing specific problems, and thus help the progress.
CASP: *in vitro* component

Let’s sacrifice some mice and do dissections!

- templates
- alignment
- 3D stuff
Interesting FM Model Problems

Target 578: 37_5

Poor quality secondary structures
“Strand” Problems:

Incorrect backbone torsion angles:

• No hydrogen-bonds with neighboring strand
• Compressed side chain distances limit contacts
• Shorter loops limit secondary structure angles

Poor quality secondary structures
Interesting FM Model Problems

Target 578: 37_5

Problem Source?

From the abstract book methods description:
• “picked server models”
• “refined and rebuilt” models
• “model quality evaluation”

From submitted pdb:
“PARENT N/A”

Poor quality secondary structures
Interesting FM Model Problems

Target 578: 37_5

The Answer?

From the abstract book methods description:
- “picked server models”
- “refined and rebuilt” models
- “model quality evaluation”

From submitted pdb:
“PARENT N/A”
Should state “ServerX_1”!

Poor quality secondary structures
Interesting FM Model Problems

Target 578: 37_5

What happened to?

From the abstract book methods description:
• “picked server models”
• “refined and rebuilt” models
• “model quality evaluation”

From submitted pdb:
“PARENT N/A”

Poor quality secondary structures
It is hard for assessors to help predictors in pure *in vivo* setup.

Organizers could tweak the procedures and formats for submission to help with that.

CASP is good “as is” as a competition, but it is not optimally designed as an experiment.
A couple of best things about CASP

Independent assessors, almost always different each time.

Assessment is a moving target, predictors don’t exactly know what to expect.

This is absolutely essential to keep predictors straight.
Main Obstacle in FM assessment

- **Model quality:**
  models are just too different from real structures

- **Model quality evaluation:**
  most models are so dissimilar to real structure that it is very hard to judge which one is better!
Free up your Minds
Never trust a single score!
Avoid overtraining on a single score

What is a better model?

Nobody knows,
but definitely not always the one
with a better GDT or TM score.

Compression tricks Cartesian scores.
There is a need for a distance score.
A favorite “cheat”
CAD-score (Česlovas Area Difference)

CAD-score: a new method for the evaluation of protein structural models

For pairs of CASP9 models for which CAD-score and GDT-TS rankings are in conflict, MolProbity score was used to judge which ranking is more consistent with the physical realism of models. The pie chart shows the results for 7578 pairs of models (having GDT-TS>60%) with MolProbity score difference bigger than 0.9:
Three kinds of scores

XYZ comparison: TM, GDT, RMSD, etc.

Distance comparison: CAD, Q-score, etc.

Model geometry: MolProb

All need to be used in method development and tests
Currently, we assess models against models.

Models - computational predictions.

Models built from experimental constraints, these models are called “structures”.
Ideal assessment

Seems best to assess models against experimental data,
e.g. structure factors, or even intensities

We are not there yet ...
Testing methods

CASP is not an ideal setup –
  too much at stake;
  too small sample.

CASP is a celebration and
  a “mouth piece” of structure prediction
Continuous evaluation

**CAMEO** - every structure that gets released in PDB is a subject of prediction efforts.

**CASP-ROLL** - harder structures are continuously released as targets.
Future of structure modeling in CASP

Ability to test FM will continuously reduce in time, as more structures are determined.

Modeling new segments is essential to test.

Loop modeling offers this opportunity.
Future of structure modeling in CASP

Refinement is an essential category

LM – loop modeling as a new category?

Why?

TBM assessors don’t have time to look at models

FM assessors don’t have time to look at loops in TBM targets
Another assessment suggestion

CASP10 assessors are fantastic

They figured out and statistically supported reasonable ranking

However, no comments were made on specific models and methods used to build them

Why? Assessors didn’t have time!!!
Another assessment suggestion

Assessor to look at the models and tell us what works and what does not

This assessor should not be bothered with ranking – who really cares

This assessor should dissect models element by element, residue by residue, method by method, and tell the predictors what’s up (or down)
What works in structure modeling today

What’s deduced from evolution

They are not predictions, they are deductions.

They are most useful for biologists
What is iffy in structure modeling today

Anything \textit{ab-initio}, be it loops or complete structures

Most terrible fact is that we don’t know when it works and when it does not:

\textbf{there is no reliable E-value}
Future of structure modeling

Exploration of evolution is unlikely to bring revolutionary changes.

The future is likely to be in “physical chemistry”, or statistical mechanics of proteins.
Statistical mechanics of proteins

Not exploring evolutionary information: no alignments, no templates

Current semi-stagnation is observed probably because bioinformatics approaches largely exhausted their potential – like sec.str.

Scoring functions,
Conformational sampling,
New alternatives to moving the polypeptide chain about?
How will this be achieved?

Either through diligent engineering – tweak software to improve scores

Or by looking at structures in the process and thinking about innovation.

Bring the art back?
Interesting observation: prediction of Local Cores in FM
Models Tend to Predict Common Local Cores

Position Specific Alignment (Prediction Center)

Target 531
Local Core
Models Tend to Predict Common Local Cores

Local Core

Position Specific Alignment (Prediction Center)

Target 550d2 Local Core
Models Tend to Predict Common Local Cores

Position Specific Alignment (Prediction Center)

Target 550d2 Local Core
Models Tend to Predict Common Local Cores

Local Core

Position Specific Alignment (Prediction Center)

Group Model 386_1

Target 550d2 Local Core
Models Tend to Predict Common Local Cores

Local Core

Position Specific Alignment (Prediction Center)

Target 578 Local Core
Models Tend to Predict Common Local Cores

Local Core

Position Specific Alignment (Prediction Center)

Target 578 Local Core
Models Tend to Predict Common Local Cores

Position Specific Alignment (Prediction Center)

Target 624 Local Core
Models Tend to Predict Common Local Cores

Local Core

Position Specific Alignment (Prediction Center)

Target 624
Local Core
FM Models Tend to Predict Common Local Cores

Potential reasons for common local cores:

Uninteresting byproduct of superimposing structures

Result of many manual groups picking and refining similar server models

Local cores are easy

Potential Uses for the community?:
Quality assessment: mimic predictions – use server models, enhance with refinements using different energy functions? Identify and fix cores and use as a building block for the rest of the secondary structure elements/fragments?
How good are we in structure prediction?

Where are we on the path to folding?

CASP - Emphasizes on relative

CASP - Neglects the absolute

Would be nice, in assessment papers, tell something about the absolute state-of-the-art, in terms of simple measures biologists understand, and that is sequence identity and RMSD.
Most serious prediction problems

Domains and Evaluation Units: their parsing, definition and related problems, like arrangements and swaps. Problem for both assessors and predictors.

Recognition of the best model in the pool of junk

Unclear reasons why some models turn out better than others.
Summary

CASP is the most amazing experiment in protein science

The rigor and desire for perfection is unprecedented at all levels throughout the process

CASP should not be shy to change with the changing word, while keeping the best traditions