Does inclusion of residue-residue contact information boost protein threading?

Sutanu Bhattacharya
Auburn University

MCBIOS 2019
Residue-residue contact map

Contact map
Research question

Does contact map boost protein threading?
What is protein threading?

- align sequence to template

![Diagram showing sequence alignment between query protein and template.]

**Query protein**

ABCDEFGH

**Template**

ABCXYZGH

Sequence alignment

ABCDEFGH

ABC - - - XYZ GH

Down arrow pointing to a protein structure.
Challenges in protein threading

- Challenge 1:
  Finding **TOP 1 template** from the template library (fold recognition)

- Challenge 2:
  Getting optimal **query-template alignment**
Our hypothesis

Contact information on top of protein **threading** boosts the **performance** of purely threading-based methods.
Does Inclusion of Residue-Residue Contact Information Boost Protein Threading?

Sutanu Bhattacharya, Debswapna Bhattacharya

First published: 18 March 2019 | https://doi.org/10.1002/prot.25684
Flow chart of our work

Query Sequence
CQETRNCEV

Template Library

Predicted contact map

Al-eigen

Native contact map

F_{score} = Z-score + (CMO*10)

Final model

(Lena and coworkers, 2010
Bhattacharya and coworkers, 2019)
Results

benchmark on 500 protein targets
Results (1)

<table>
<thead>
<tr>
<th>dataset</th>
<th>comparisons with threading method</th>
<th>measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test500</td>
<td>MUSTER</td>
<td>TM-score of top-ranked model</td>
</tr>
</tbody>
</table>
Test500 (MUSTER vs. This work)

- **better** average TM-score than MUSTER
- statistically significant (p-value < 0.05)
Results (2)

benchmark on 150 protein targets
## Results (2)

<table>
<thead>
<tr>
<th>dataset</th>
<th>comparisons with ab-initio folding method</th>
<th>measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSICOV150</td>
<td>CONFOLD2</td>
<td>TM-score of top-ranked model</td>
</tr>
</tbody>
</table>

(Cheng and coworkers, 2018)
PSICOV150 (CONFOLD2 vs. This work)

- **better** average TM-score than CONFOLD2
- statistically significance (p-value < 0.05)
Results (3)

benchmark on 20 CASP13 protein targets
## Results (3)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Comparisons with state-of-the-art contact-assisted threading methods</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASP13 (20 full-length proteins in total of 32 domains)</td>
<td>- EigenTHREADER</td>
<td>TM-score of top-ranked model</td>
</tr>
<tr>
<td></td>
<td>- map-align</td>
<td></td>
</tr>
</tbody>
</table>

(Jones and coworkers, 2017
Baker and coworkers, 2017)
CASP13 performance

![Bar chart showing performance for different methods: map_align, EigenTHREADER, and Our work. The chart compares performance in Full-length and Domain level.]
Case study: 2f2ba (245 residues)

Native 3D structure

Predicted model
Contact + Threading

Predicted model
Only Threading

TM-score: 0.59

TM-score: 0.28
Case study: T0966 (494 residues)

Our predicted model (rainbow) superimposed to the native (gray)

EigenTHREADER predicted model (rainbow) superimposed to the native (gray)

TM-score: 0.8

TM-score: 0.19
Conclusions and Future work

- Test500: contact + threading better than purely threading-based methods

- PSICOV150: contact + threading better than contact-assisted ab initio folding methods

- Contact-assisted threading is a promising avenue for fold recognition.

- What about getting best fit query-template alignment (Challenge 2)?

- What about residue-residue distance instead of contacts?
Acknowledgements

Dr. Debswapna Bhattacharya

Auburn University
Query Sequence

CQETRKKCTEMKKFKNCEV

Template Library

Hydrophobicity match
SS & SA match
Profile match
Torsion Angles match

Score (S)

Template sequence

E

C

Query sequence

V

DP table

Z_{\text{score}}

Native contact map

CMO

1

20

F_{\text{score}} = Z_{\text{score}} + (\text{CMO} \times 10)

Final model