US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes

First Author: Chengxin Zhang
Published: August 29, 2022, in *Nature Methods*

Paper Presenter: Xinyu Wang

nature methods

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature methods › articles › article

Article | Published: 29 August 2022

US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes

Chengxin Zhang, Morgan Shine, Anna Marie Pyle & Yang Zhang ✉

Nature Methods **19**, 1109–1115 (2022) | Cite this article

10k Accesses | **101** Citations | **31** Altmetric | Metrics

OXFORD ACADEMIC Journals Books

Nucleic Acids Research

JOURNAL ARTICLE

TM-align: a protein structure alignment algorithm based on the TM-score

Yang Zhang ✉, Jeffrey Skolnick

Nucleic Acids Research, Volume 33, Issue 7, 1 April 2005, Pages 2302–2309,
https://doi.org/10.1093/nar/gki524
Published: 01 January 2005 **Article history** ▾

iScience

CellPress

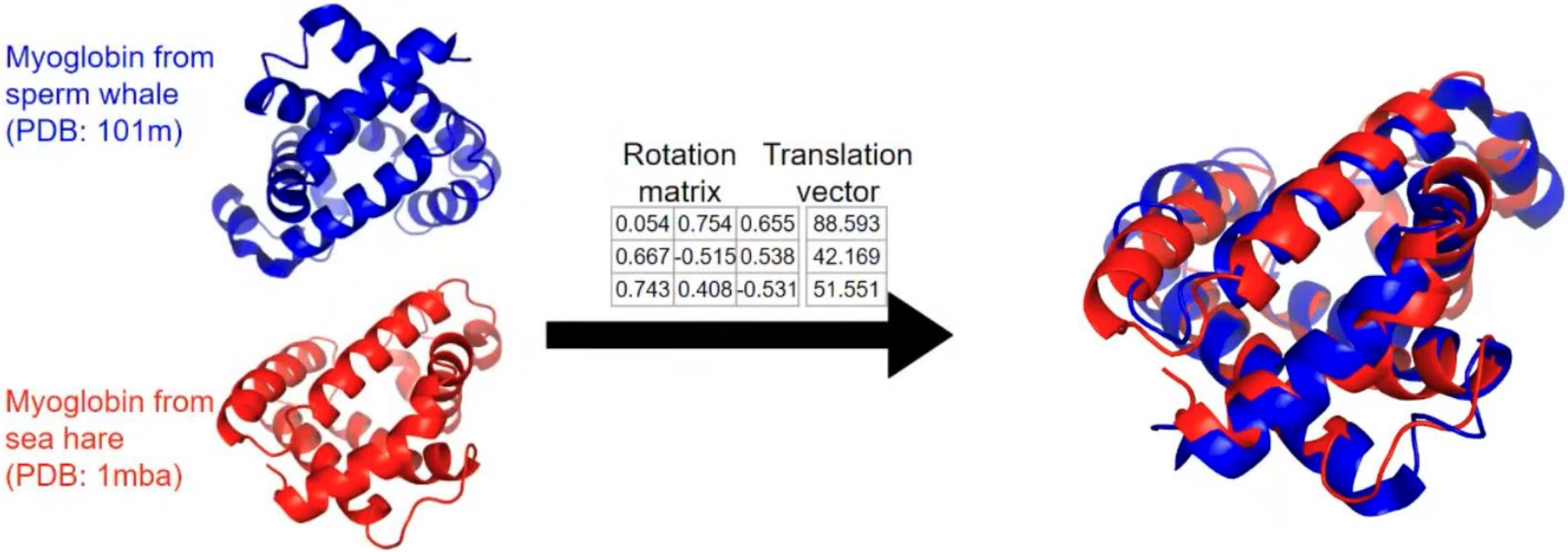
Volume 25, Issue 10, 21 October 2022, 105218

Article

A unified approach to sequential and non-sequential structure alignment of proteins, RNAs, and DNAs

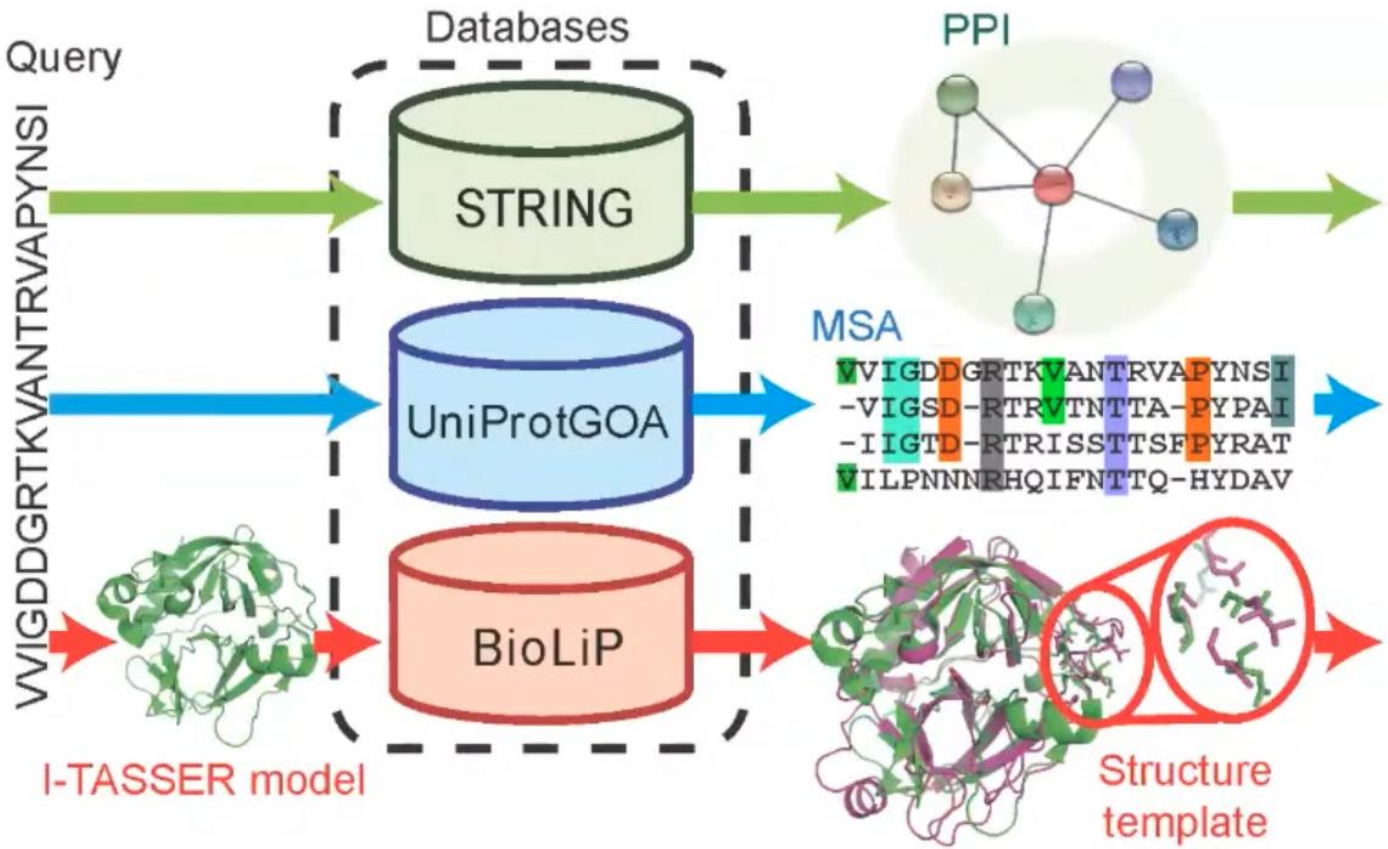
Chengxin Zhang [1 2 3], Anna Marie Pyle [1 2 4 5] ✉

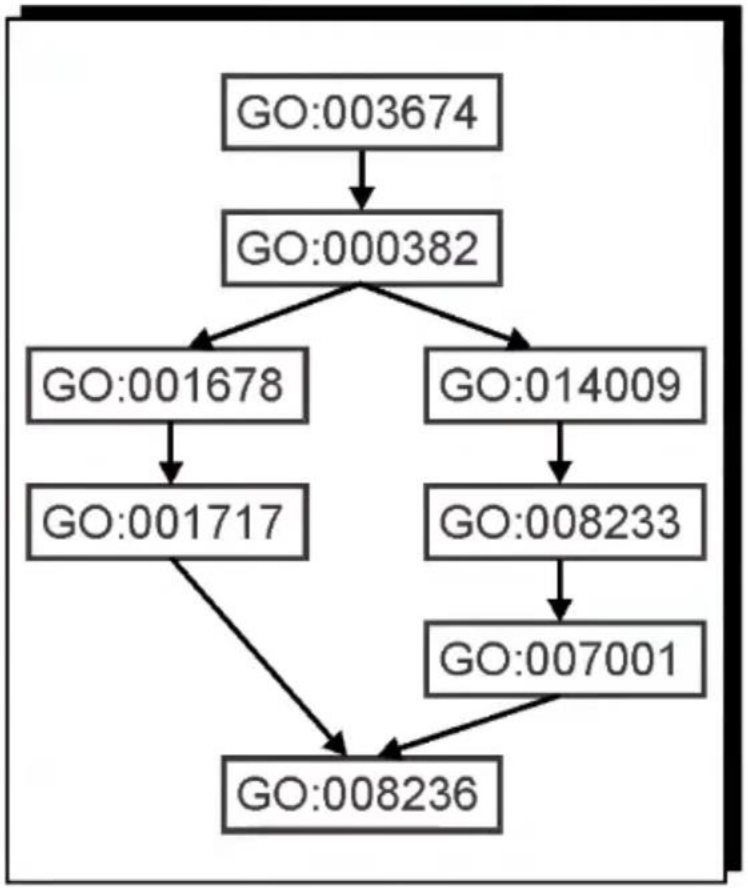
Introduction：structure superimposition and alignment



>101m
MVLSEGEWQLVLHVWAKVEADVAGHGQDIILRLFKSHPETLEKFDRVKHLKTEAEMKASEDLKKHGVTVLTALGAILKKK--G-
HHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGNFGADAQGAMKALELFRKDIAAKYKELGYQG
>1mba
-SLSAAEADLAGKSWAPVFANKANGLDFLVALFEKPDSANFFADFKGKS-VADIKASPQLRDVSSRIFTRLNEVNNAANASSKM
SAMFAKEHV--GFGVGSAQFENVRSMFPGFVASVAA--PPAGADAAWTKLFGLIIDALKA-A------GA

Why structure alignment





Metrics for Structure Similarities

$$RMSD = \sqrt{\frac{1}{L_{ali}} \sum_{i=1}^{L_{ali}} d_i^2}$$

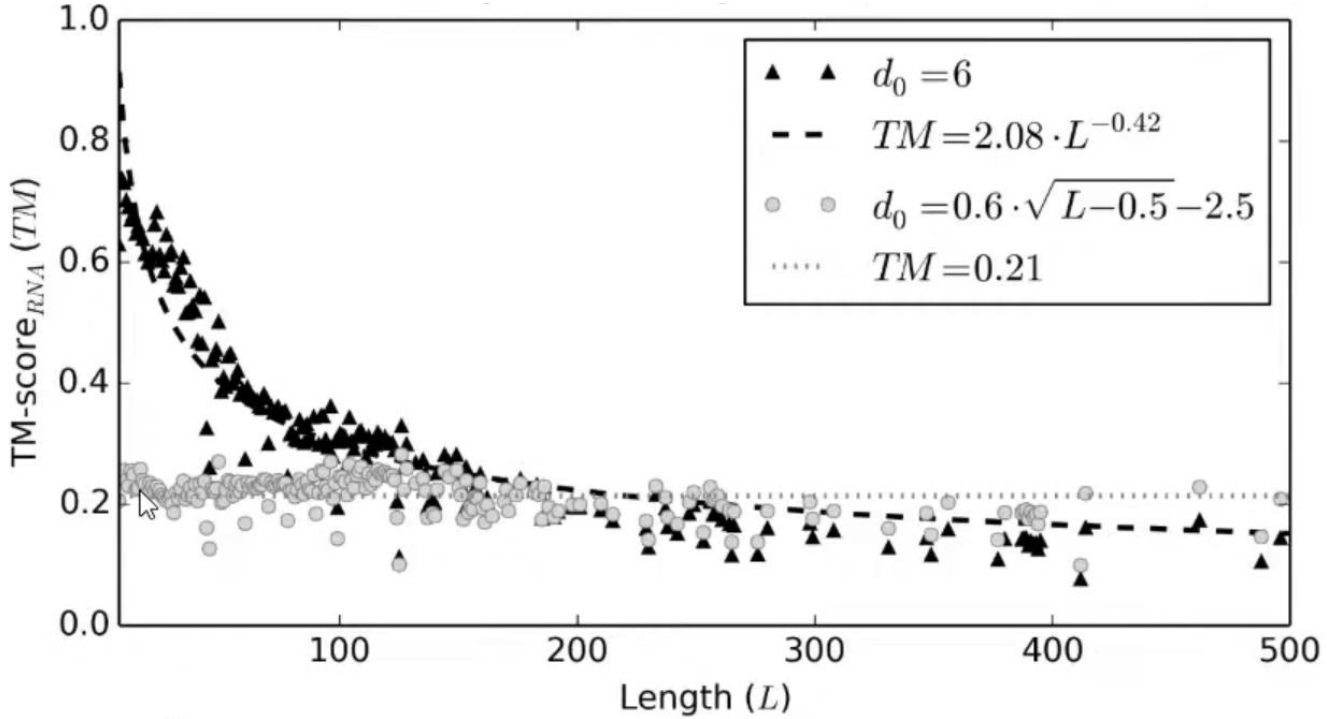
L - Total number of residues
L_{ali} - Number of aligned residues
d_i - Distance between the i-th aligned residue pair

$$Coverage = \frac{L_{ali}}{L}$$

$$TM_{score} = \frac{1}{L} \sum_{i=1}^{L_{ali}} \frac{1}{1 + \left(\frac{d_i}{d_0}\right)^2}$$

$$d_0 = \begin{cases} 1.24 \cdot \sqrt[3]{L - 15} - 1.8, & \text{for proteins} \\ 0.6 \cdot \sqrt{L - 0.5} - 2.5, & \text{for RNA} \end{cases}$$

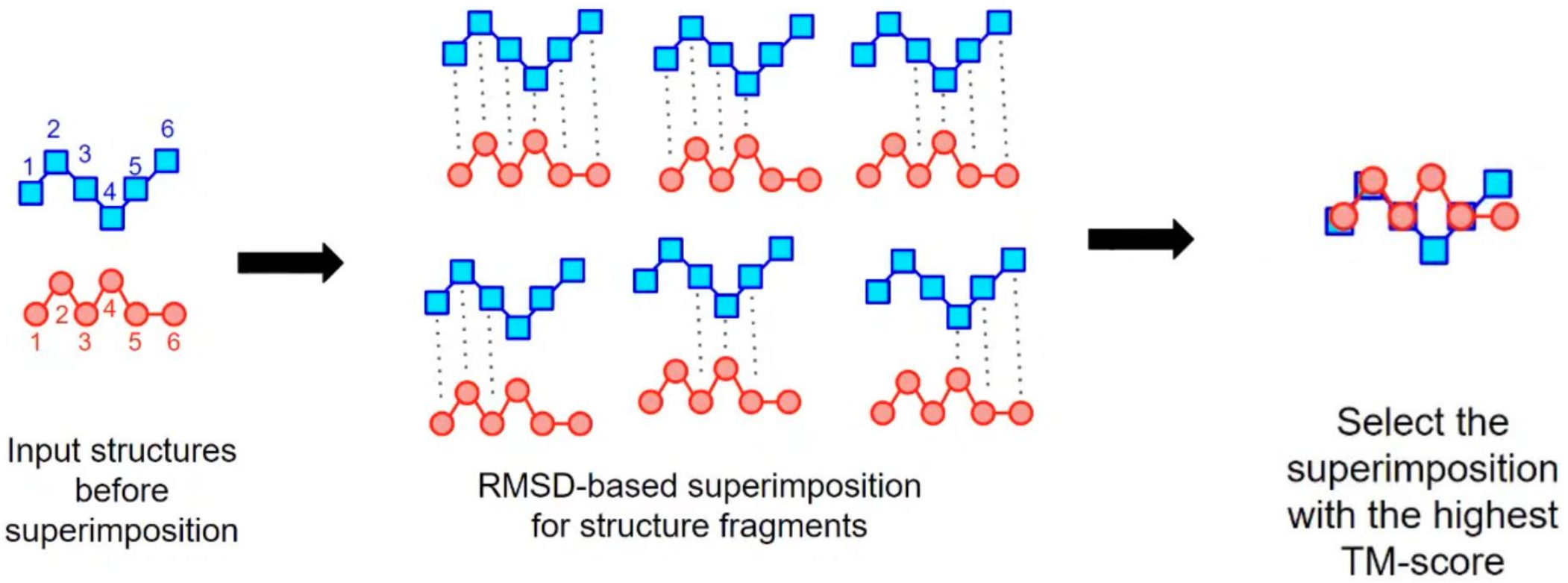
Size independency of TM-score Similarity



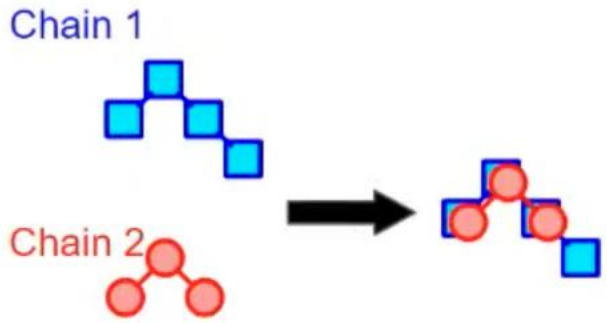
$$TM_{score} = \frac{1}{L} \sum_{i=1}^{L_{ali}} \frac{1}{1 + \left(\frac{d_i}{d_0}\right)^2}$$

$$d_0 = \begin{cases} 1.24 \cdot \sqrt[3]{L - 15} - 1.8, & \text{for proteins} \\ 0.6 \cdot \sqrt{L - 0.5} - 2.5, & \text{for RNA} \end{cases}$$

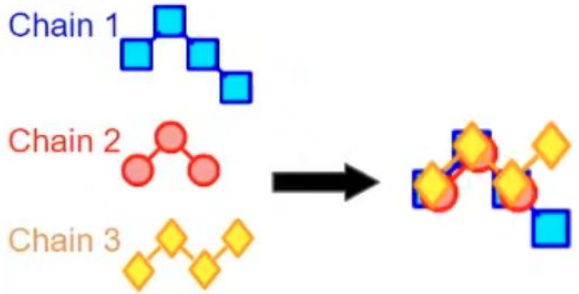
Heuristic approach to optimize TM-score



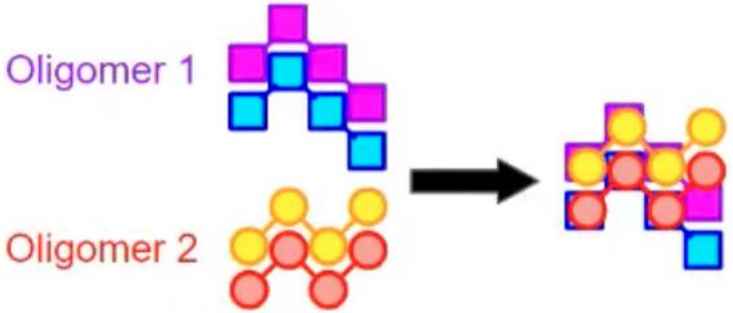
Different alignment modes



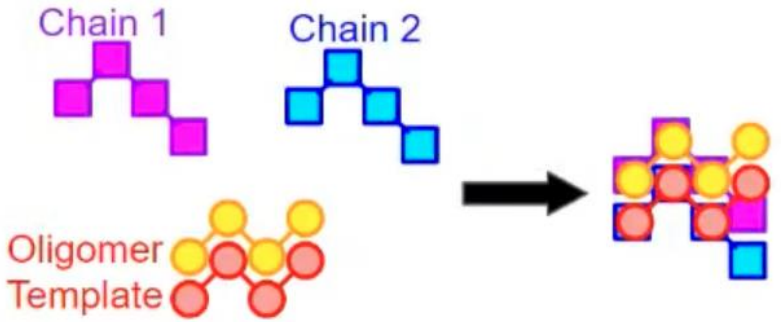
Pairwise monomer structure alignment



Multiple structure alignment
(≥3 monomers)

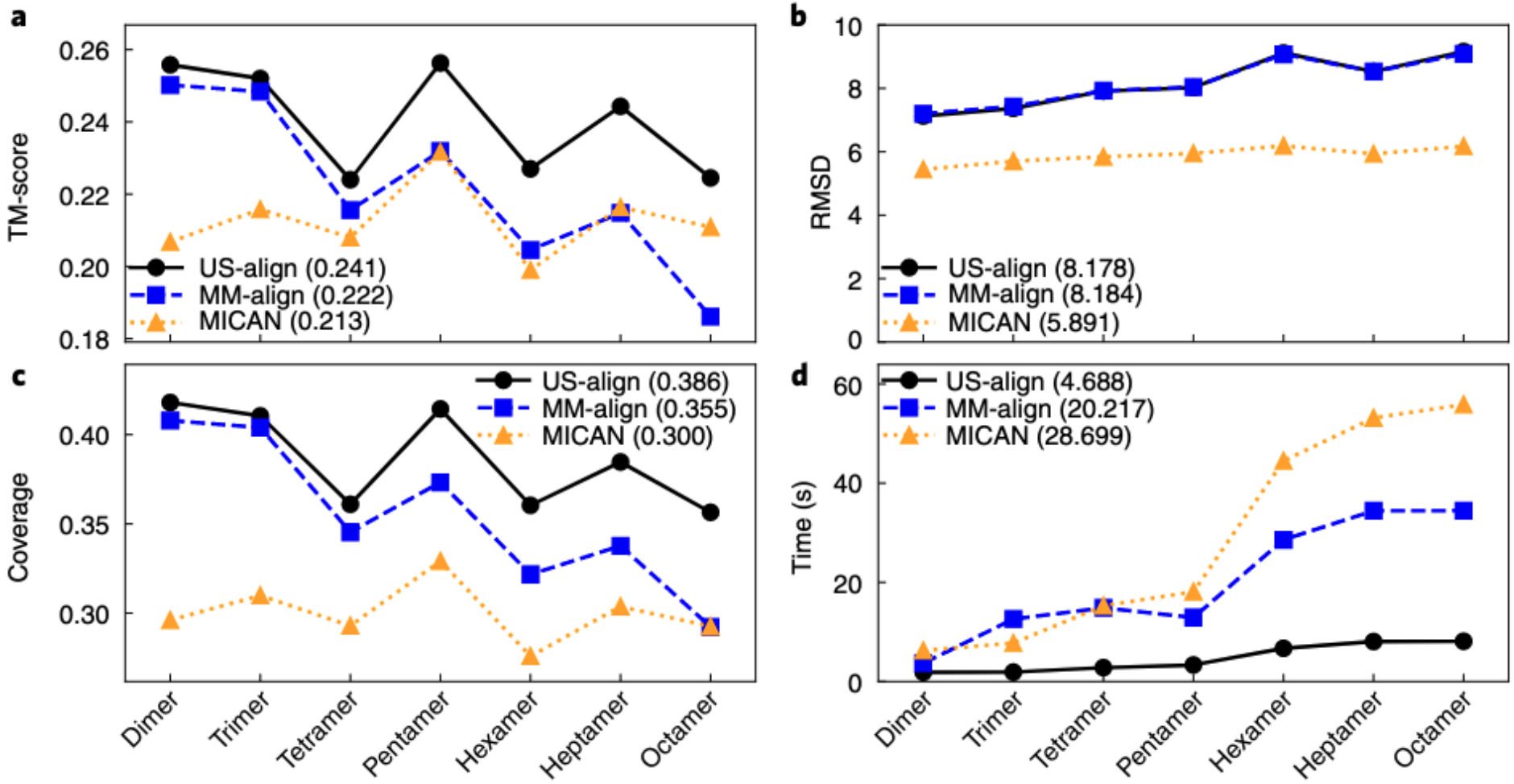


Oligomer structure alignment
(2 oligomers with unknown
residue/chain correspondence)

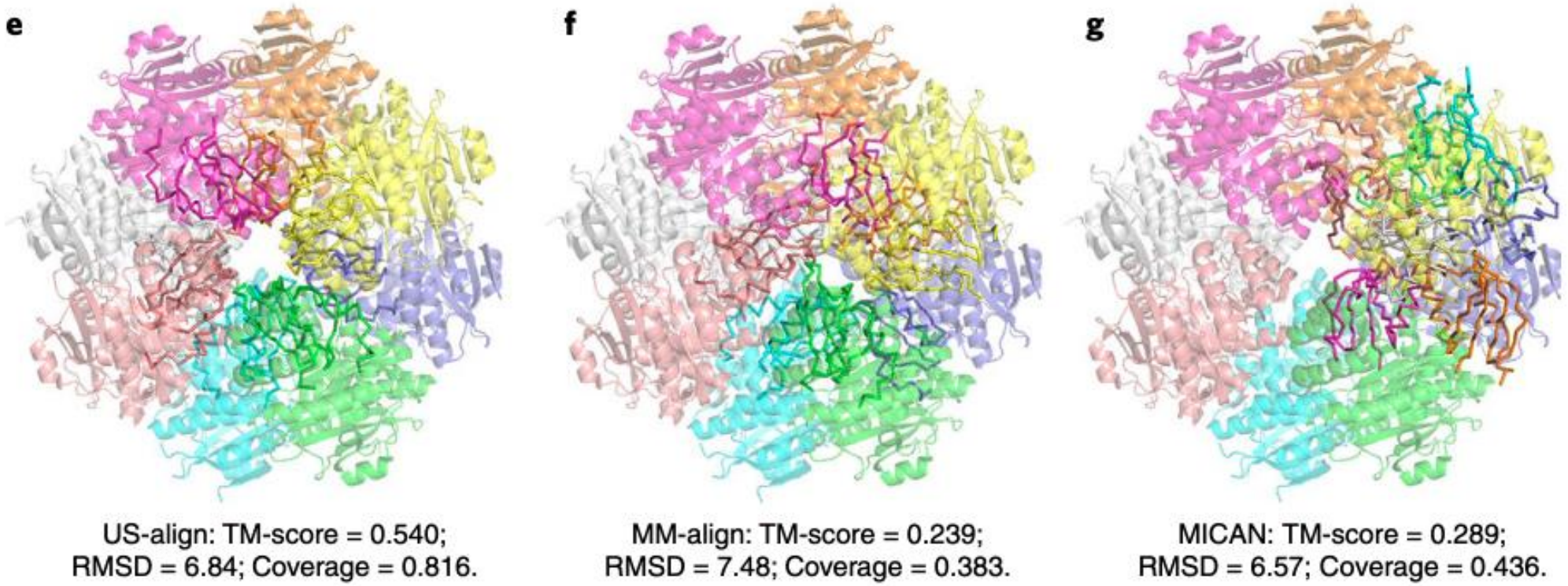


Template-based docking
(a set of chains and an oligomer structure template)

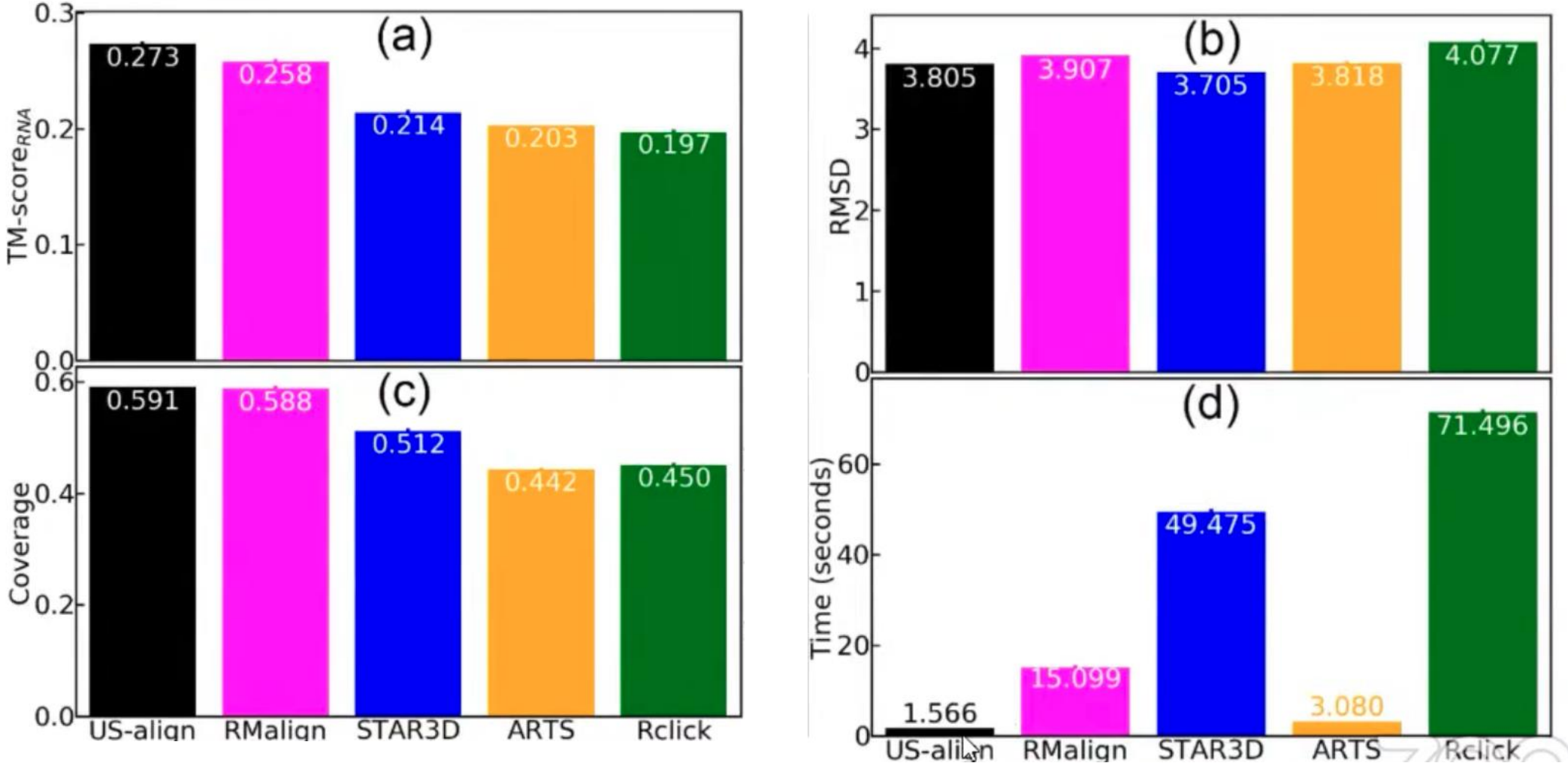
Oligomer structure alignment Performance



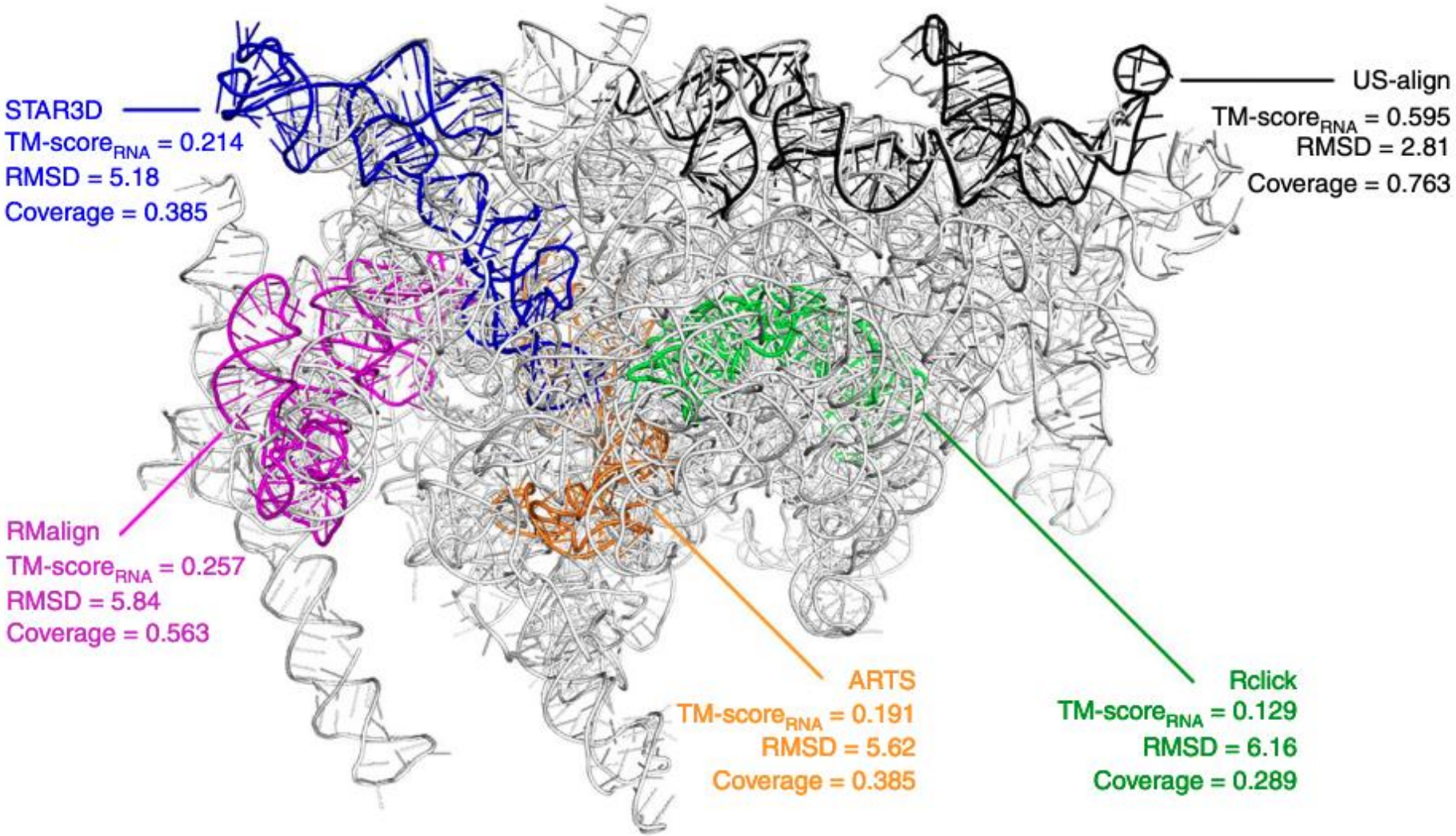
Case study： 4jhm



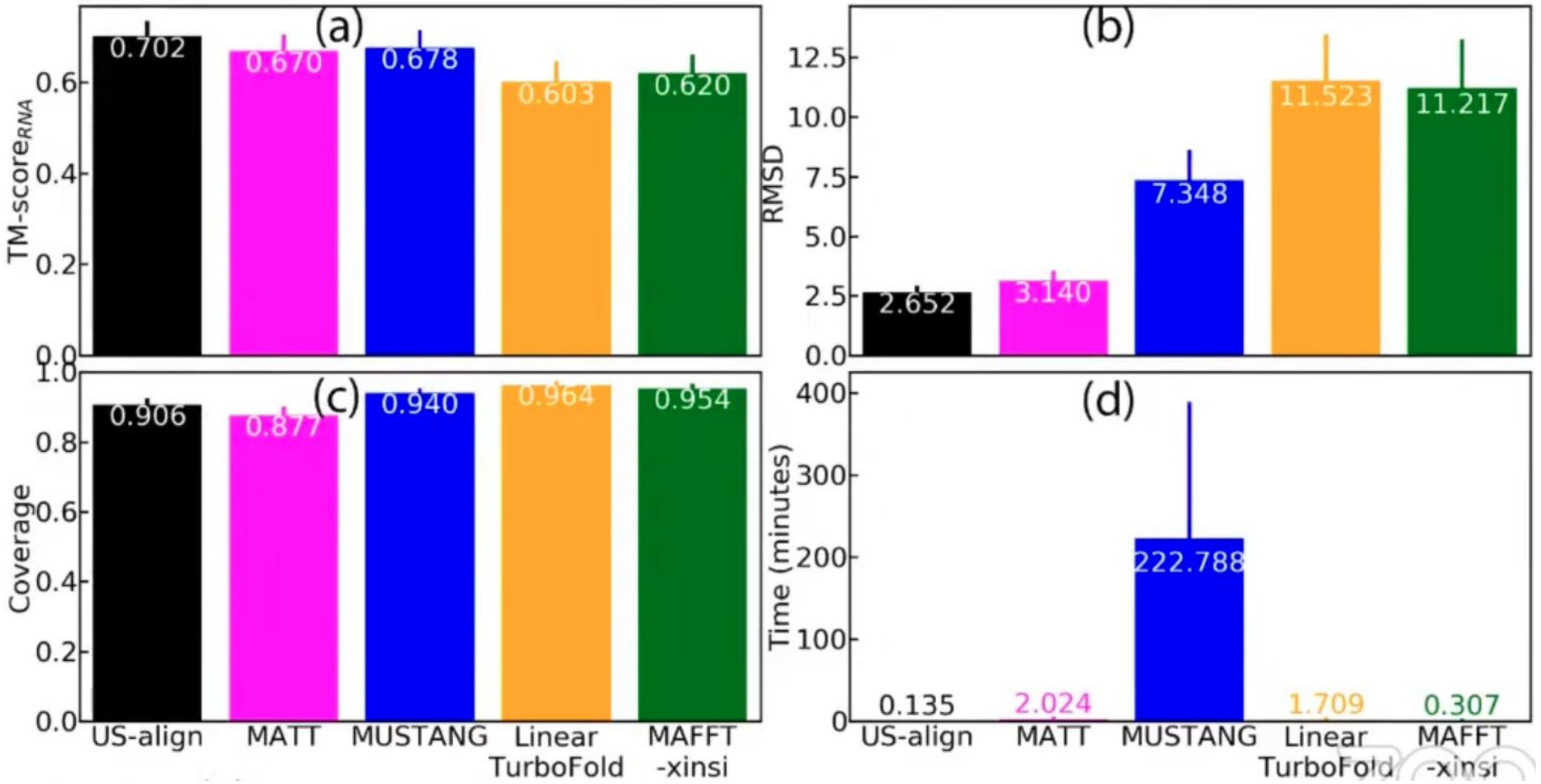
Monomeric RNA structure alignment



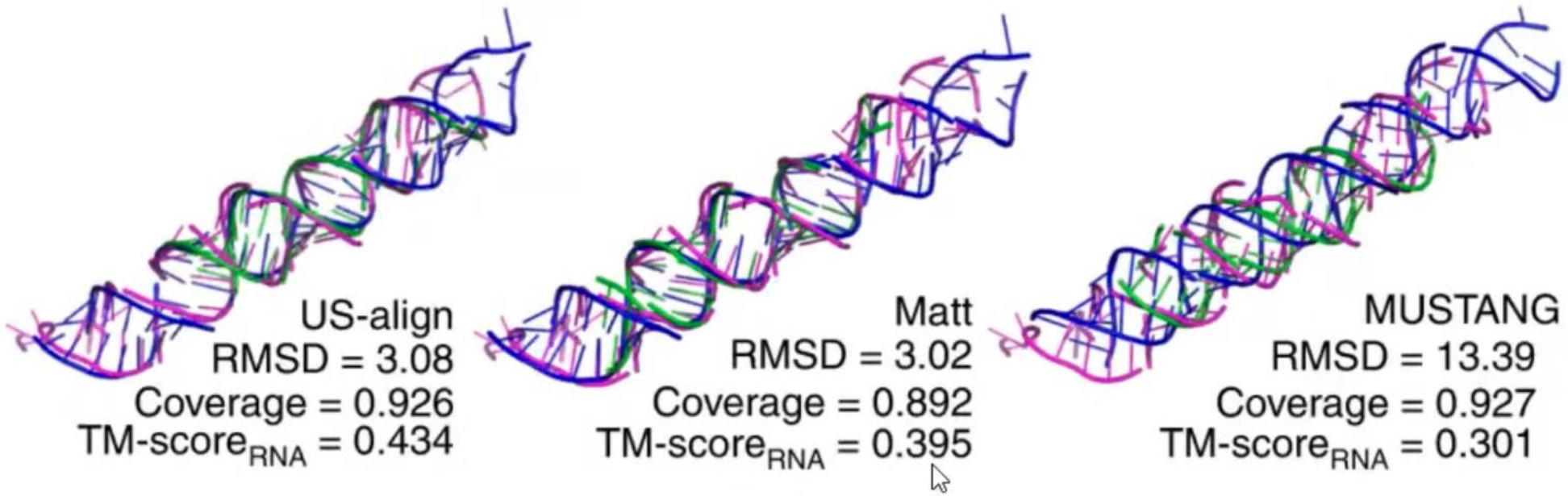
Case study： 6y2l



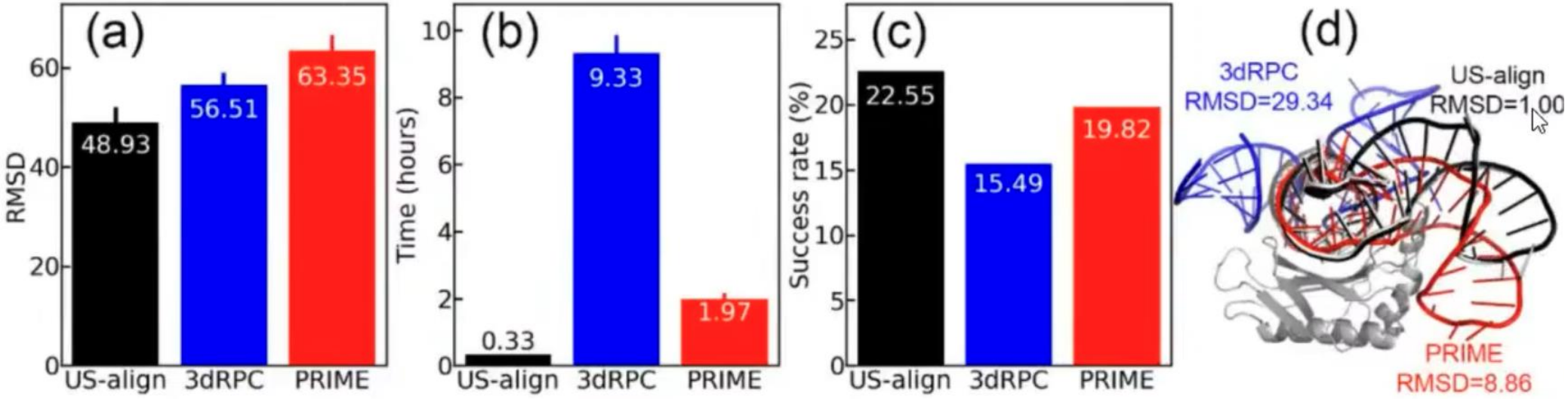
RNA multiple structure alignment



Case study:6v5b



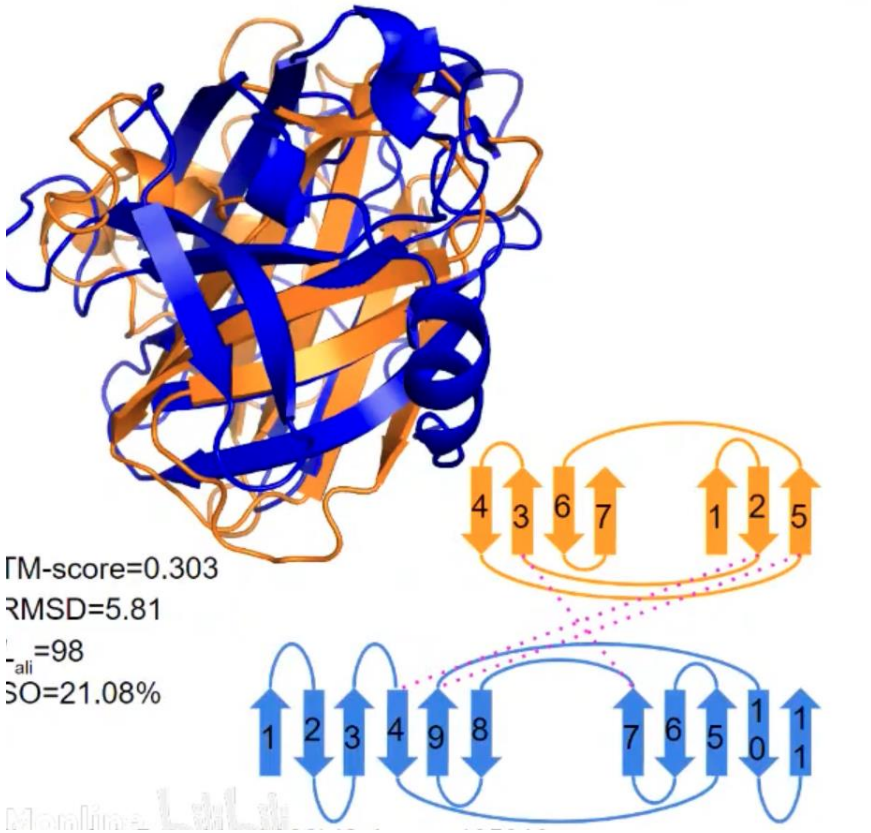
RNA-Protein docking

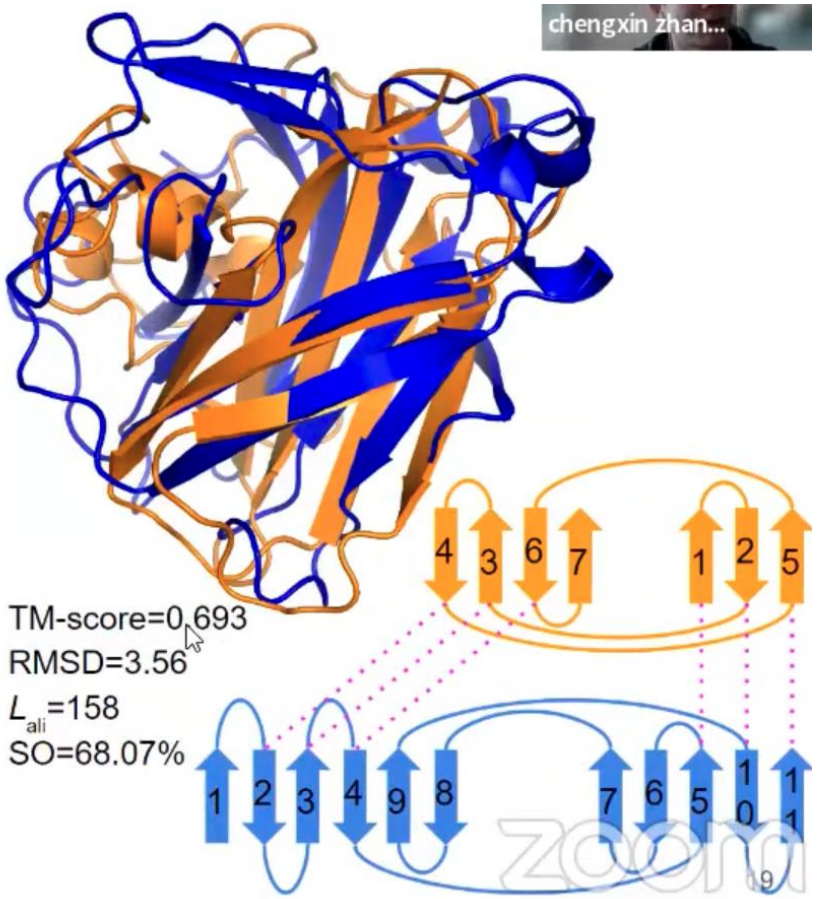


US-align2

- Structure alignment for proteins,nucleic acids, and macormolecular
- Similar structure=sequential?
- Usalign2-non-sequential

Sequential alignment vs non-sequential alignment





Questions