CASP16
Macromolecular Ensembles

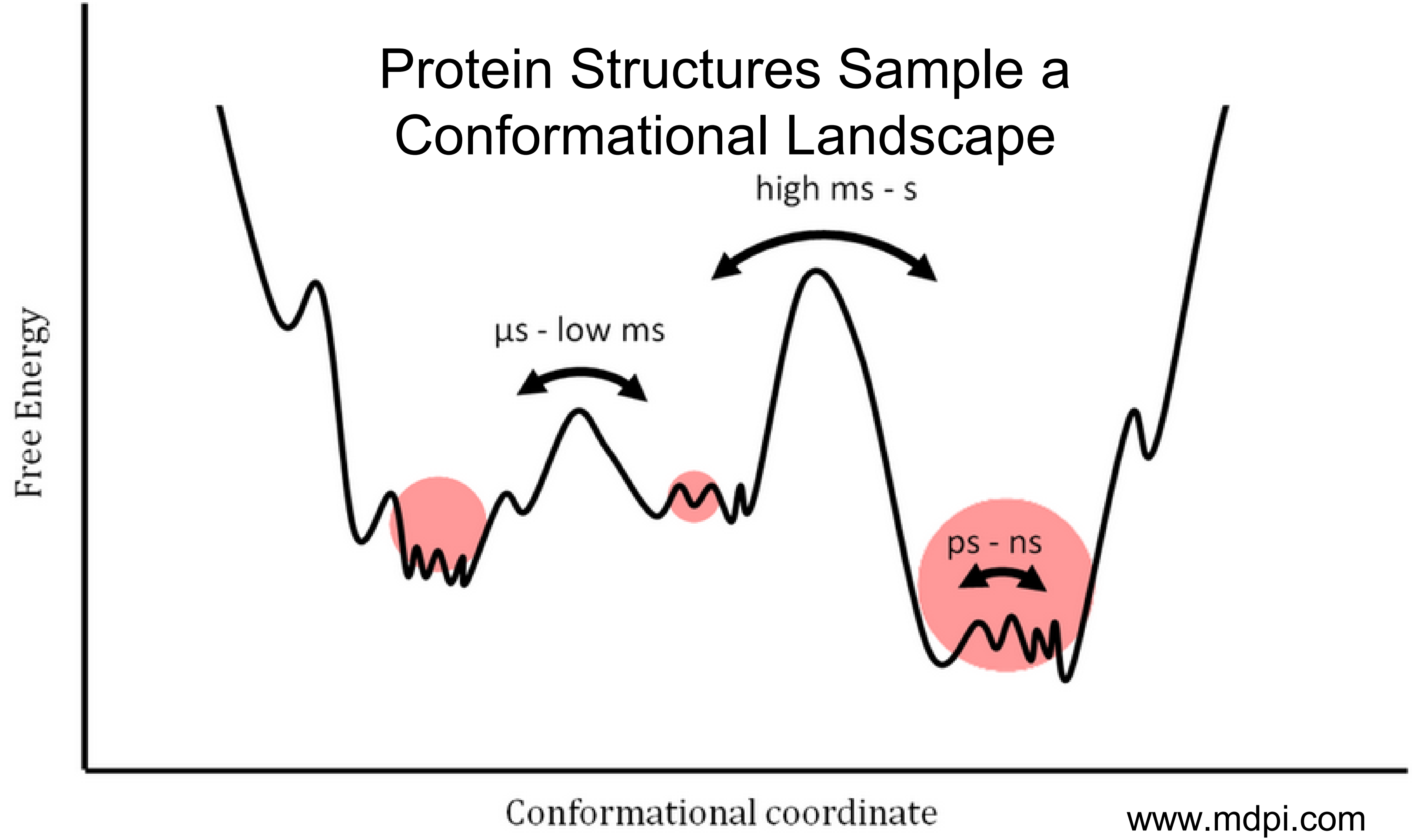
Namita Dube
Yuanpeng Janet Huang
Theresa Ramelot
GVT Swapna
T Benavides

Bruce Donald
Terry Oas
Susan Tsutakawa

Andriy Kryshtafovych
John Moult

CASP Special Interest Group on Ensembles of Conformations

Protein Dynamics



www.mdpi.com

Alternative Conformational States

- Hinges (HG)
- Lids / Cryptic Sites (LC)
- Rearrangements (RA)
- Fold Switching (FS)
- Intrinsically disordered proteins (ID)
  - linkers between domains
- Oligomer State

Modulators

- Boltzmann Distributions (BD)
  - Temperature
  - Pressure
  - “Invisible states”
- Ligand Binding (LB)
- Biomolecule Binding (BB)
- pH change (PH)
- Mutation / (MP) Postranslational modification

Alternative Conformational States

- Hinges (HG)
- Lids / Cryptic Sites (LC)
- Rearrangements (RA)
- Fold Switching (FS)
- Intrinsically disordered proteins (ID)
    - linkers between domains

- Oligomer State

Modulators

- Boltzmann Distributions (BD)
    - Temperature
    - Pressure
    - “Invisible states”

- Ligand Binding (LB)
- Biomolecule Binding (BB)
- pH change (PH)
- Mutation / (MP) Postranslational modification

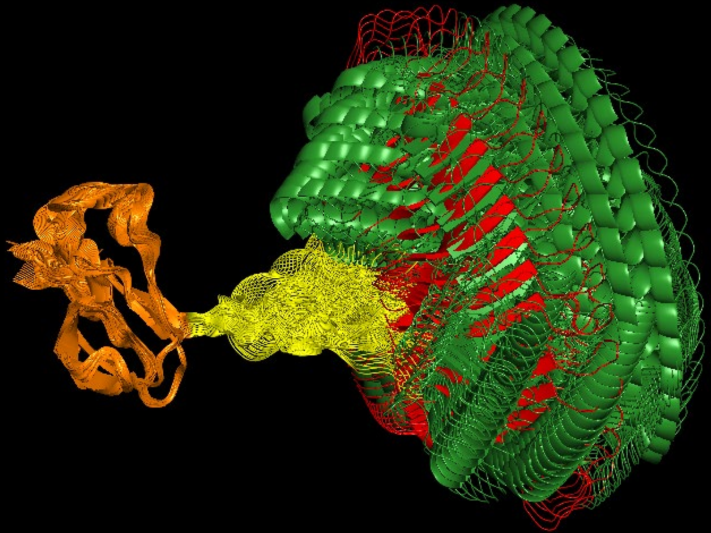
Features of CASP16 Ensemble Targets

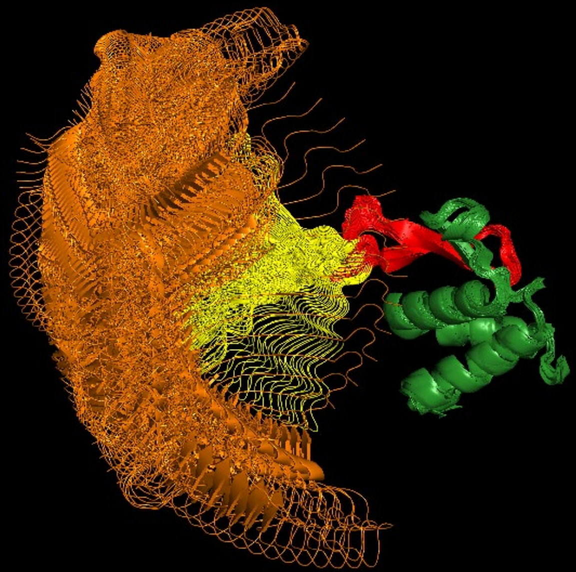
target	class	comments	nres	method	models	groups	models	class	
T1214holo	A1	Porin – ligand complex (apo in PDB)	677	EM	5	77	385	LC / LB	
M1228v1/v2	A4B2C2D2E3	2-state Protein-DNA complex	663	EM	5 + 5			HG / BB	
T1228v1/ v2	A1	2-state Protein-DNA complex	545	EM	5 + 5	69 / 68	345 / 340	HG / BB	
M1239 v1/v2	A4B2C2D2E2	2-state Protein-DNA complex	738	EM	5 + 5			HG/ BB	
T1239v1 / v2	A1	2-state Protein-DNA complex	620	EM	5 + 5	69 / 70	345 / 350	HG/ BB	
T1249v1 / v2	A3	2-state Protein trimer	488	EM	5 + 5	144 / 144	698 / 698	RA / BD	
T1294v1 / v2	A2	Protein dimer	214	X-ray	5 + 5	112 / 112	558 / 558	RA / BD	2-state
R1203v1 / v2	R1	RNA monomer	134	X-ray (2.85 Å)	5	43 / 49	183 / 183	HG / BD	2-state
R1253v1 / v2	R8	RNA octamer	574	EM	5+5	27 /27	121 / 118	RA / BD	2-state
R1283	R1, R4, R8	Monomer / tetramer / octamer	580	EM	5 + 5 + 5		32	156	OS

Features of CASP16 Ensemble Targets

target	class	comments	nres	method	models	groups	models	class	
T1214holo	A1	Porin – ligand complex (apo in PDB)	677	EM	5	77	385	LC / LB	
M1228v1/v2	A4B2C2D2E3	2-state Protein-DNA complex	663	EM	5 + 5			HG / BB	
T1228v1/ v2	A1	2-state Protein-DNA complex	545	EM	5 + 5	69 / 68	345 / 340	HG / BB	
M1239 v1/v2	A4B2C2D2E2	2-state Protein-DNA complex	738	EM	5 + 5			HG/ BB	
T1239v1 / v2	A1	2-state Protein-DNA complex	620	EM	5 + 5	69 / 70	345 / 350	HG/ BB	
T1249v1 / v2	A3	2-state Protein trimer	488	EM	5 + 5	144 /144	698 / 698	RA / BD	
T1294v1 / v2	A2	almost identical	214	X-ray	5 + 5	112 / 112	558 / 558	RA / BD	2-states
R1203v1 / v2	R1	RNA monomer	134	X-ray (2.85 Å)	5	43 / 49	183 / 183	HG / BD	2-state
R1253v1 / v2	R8	models	574	EM	5+5	27 /27	121 / 118	RA / BD	no good
R1283	R1, R4, R8	conformation in all states identical	580	EM	5 + 5 + 5		32	156	OS

MLV IN CTD Domain bound to BRD3 ET Domain





GVT
Swapna
S Aiyer
M Roth
J Hao
L Ma

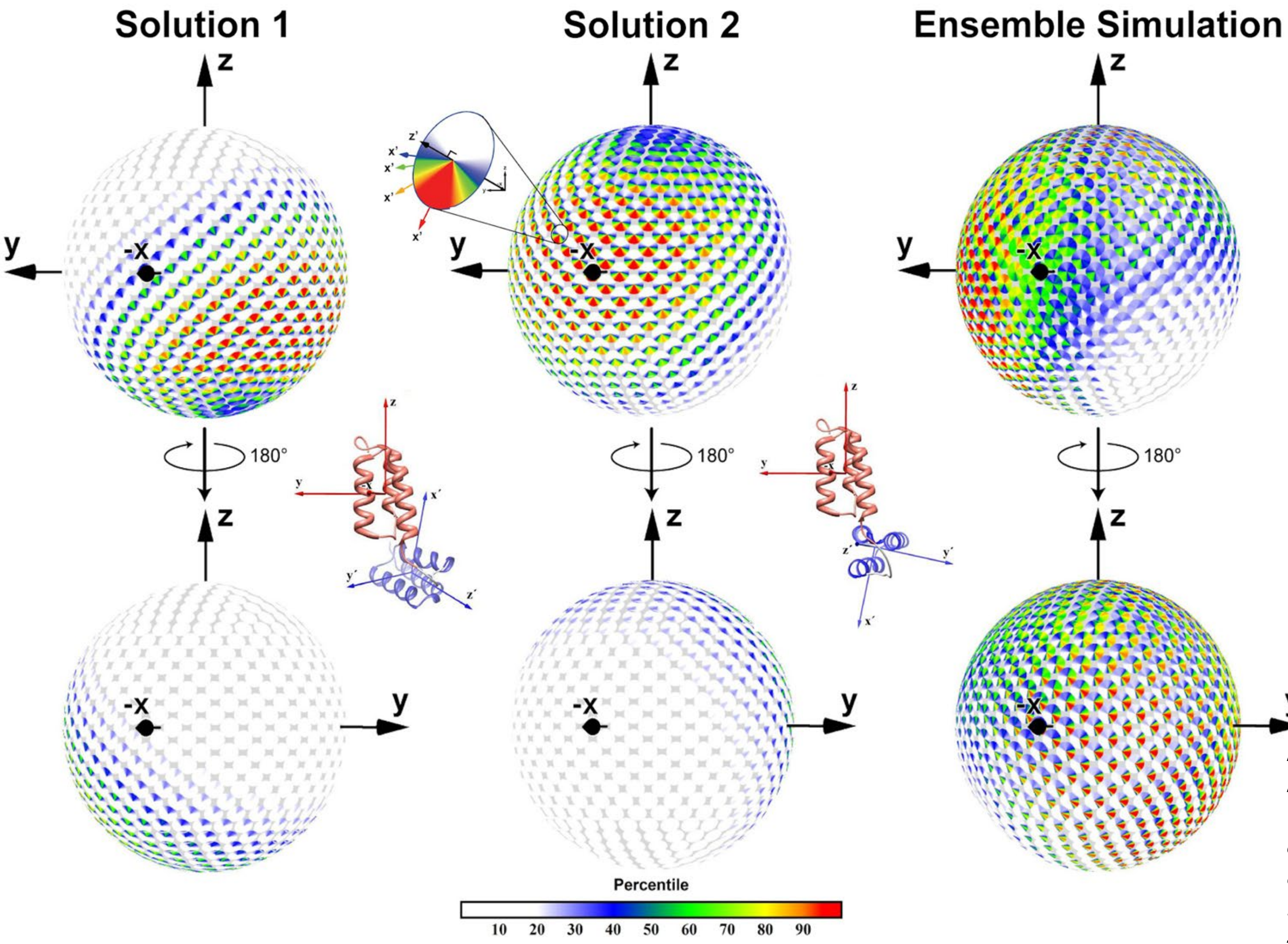
N Dube

Modeling Continuous Distributions

Protein A Challenge

Terry Oas
Bruce Donald
Susan Tsutakawa

T1200 / T1300



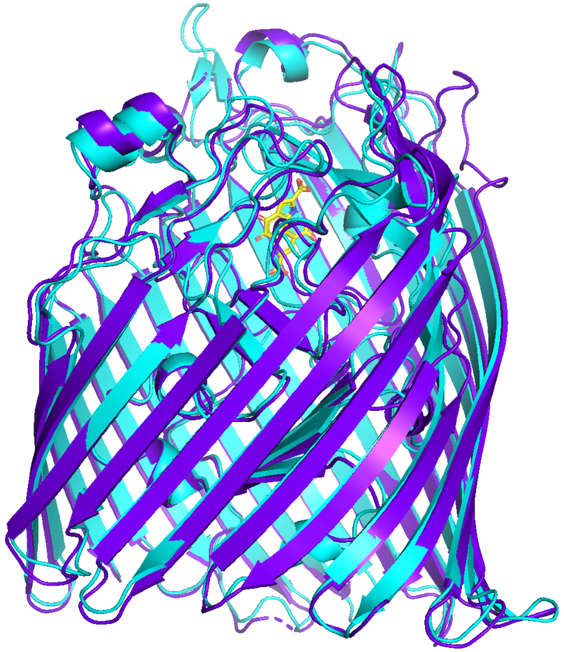
Y. Qi, J. W. Martin, A. W. Barb, F. ThÅLelot, A. K. Yan, B. R. Donald, and T. G. Oas, "Continuous interdomain orientation distributions reveal components of binding thermodynamics," J Mol Biol, 430, 3412, 2018.

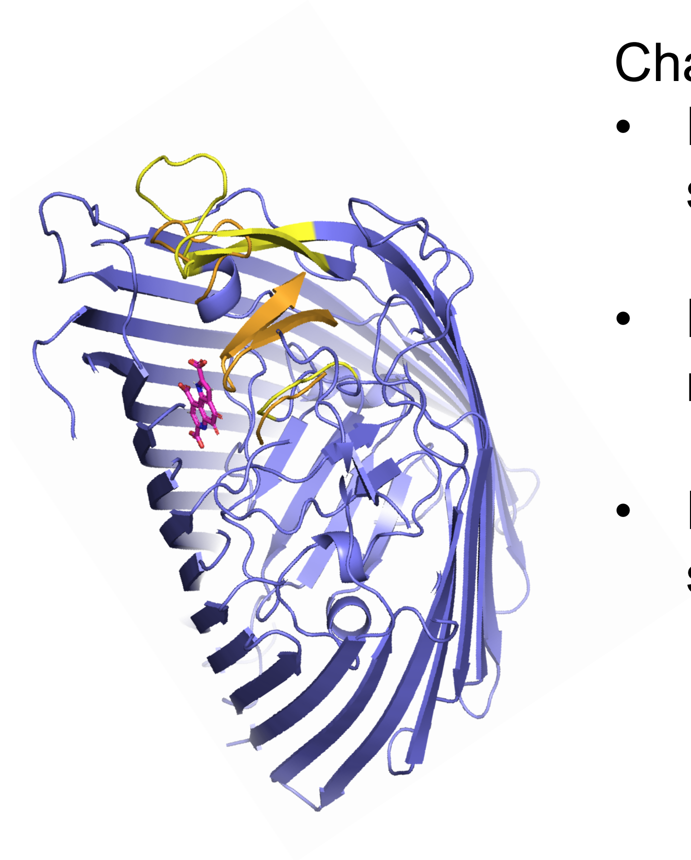
Challenges

- Differences may be localized – while overall structures of two states are nearly identical

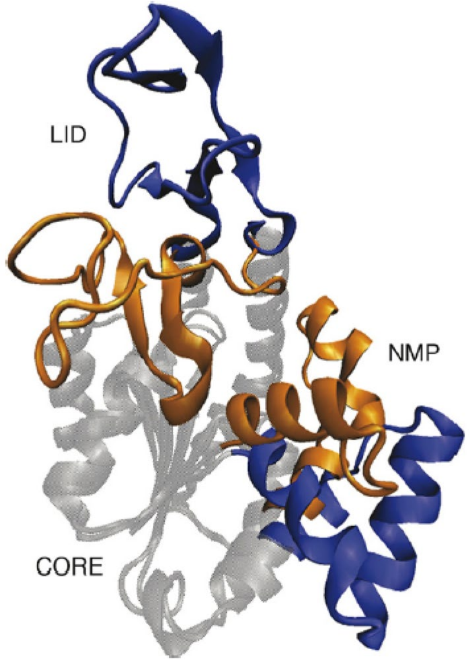
- May be differences just due to inaccuracy of model

- How well did modelers find two (or multiple) states





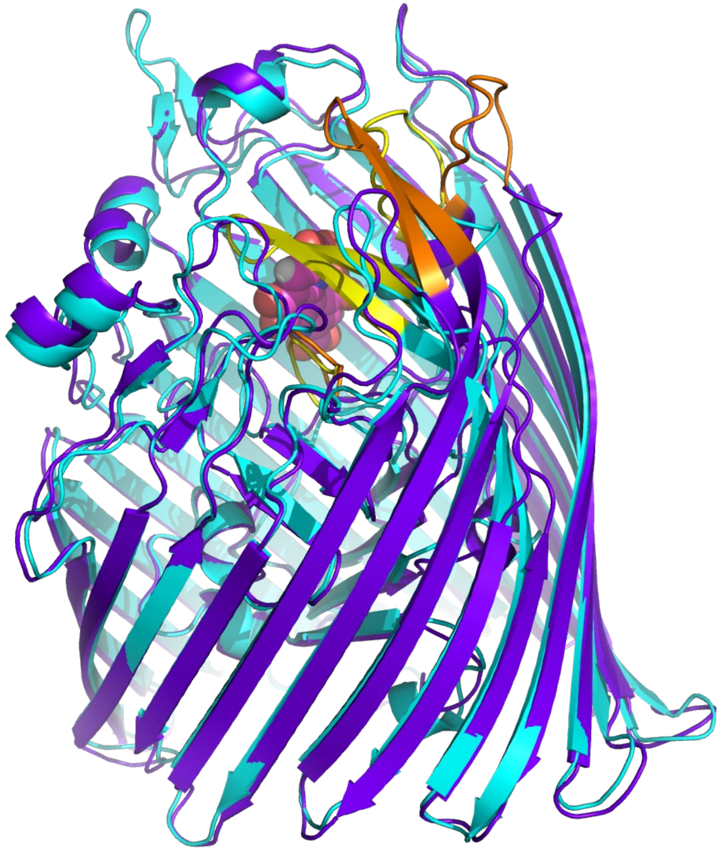
T1214

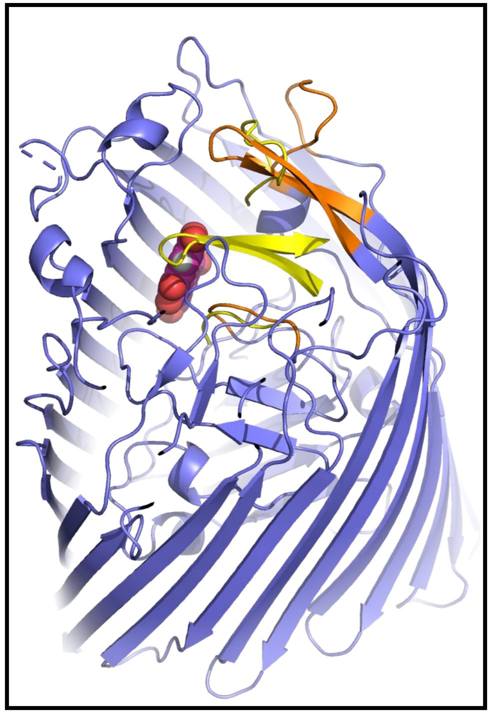


E coli adenylate kinase

adapted from Onuchic and co-workers JMB 2007

T1214





purple + orange: original apo structure in PDB
cyan + yellow: holo complex with PQQ bound

T1214, 677 residues,

Stoichiometry: A1 - Monomer,
Exp Method: EM

Ligand induced conformational changes
Protein has 2 conformations open/closed

apo – in PDB
challenge is to model to holo structure

Models Submitted 5

Rhys Grinter, Department of Microbiology, Monash University, Clayton, Australia

Scoring

Relative Overall Multi-Domain Orientations (0 to 100)

global GDT-TS

LDDT

For single domain or individual domain AU: Composite Score sensitive to both Global and Local Structural Variations (each one normalized to 0 to 100)

global GDT-TS

LDDT

local LDDT -- from selected regions

"local GDT-TS" -- LGA Dev value provides CA-CA distance -- from selected regions

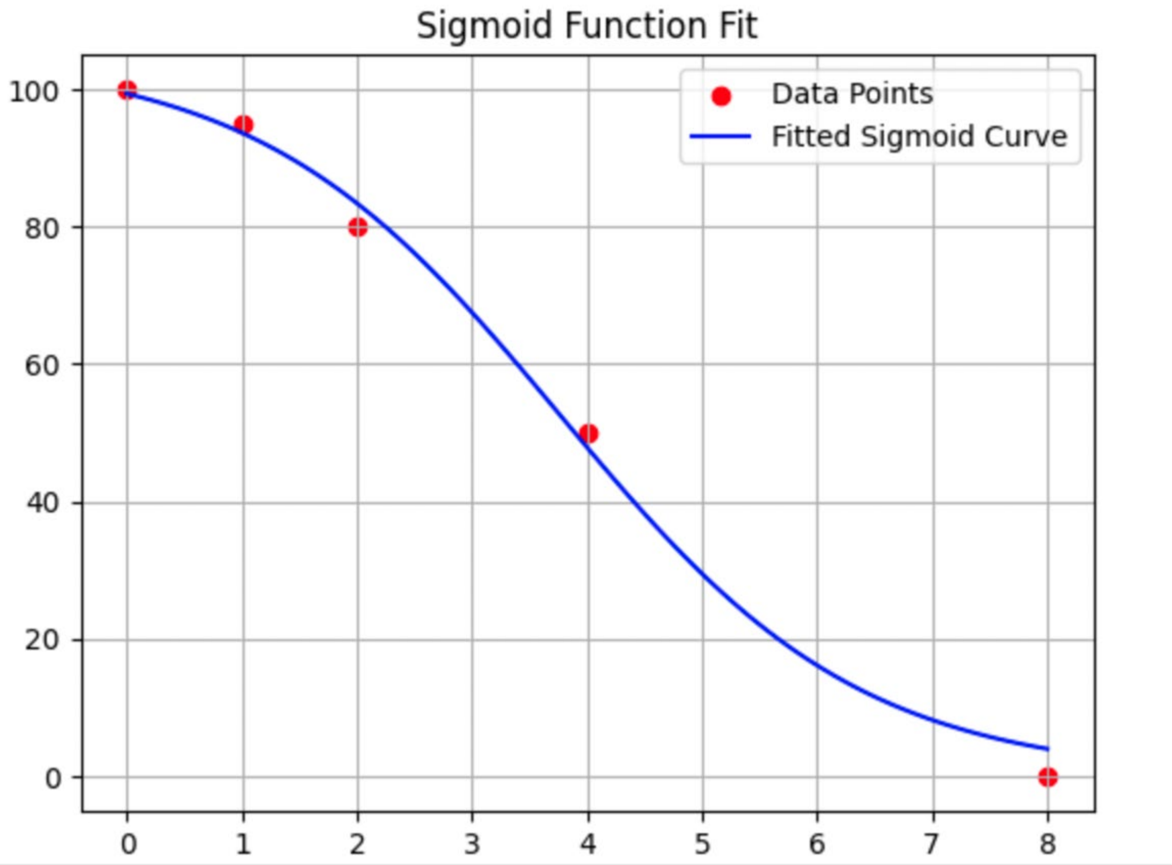
global RMSD - sensitive to overall global superimposition and local structural variations

Fitting of Sigmoid Function to RMSD values

RMSD = 0 Score = 100
RMSD = infinity Score =0

$L/(1+\exp(-k*(x-x0)))$, L=105, k=-0.77, x0=3.77.

σ_RMSD



RMSD (Å)

Use **RMSD** when you need a precise measure of average atomic displacement, especially for small changes.

Use **GDT-TS** for a more comprehensive assessment of structural similarity that is robust to variations and outliers.

RMSD = 0 Å: σ_RMSD ≈ 100 (perfect alignment since RMSD is zero).
RMSD = 1 Å: σ_RMSD ≈ 95-99 (most residues would fall within 1 Å, close to perfect).
RMSD = 2 Å: σ_RMSD ≈ 80-90 (many residues would fall within the 2 Å cutoff, good alignment).
RMSD = 4 Å: σ_RMSD ≈ 50-70 (residues spread more widely, some within the 4 Å cutoff).
RMSD = 8 Å: σ_RMSD ≈ 0-20 (most residues will be further away from the reference, poor alignment).
Used these points for fitting: (0, 100), (1, 95), (2, 80), (4, 50) and (8, 0)

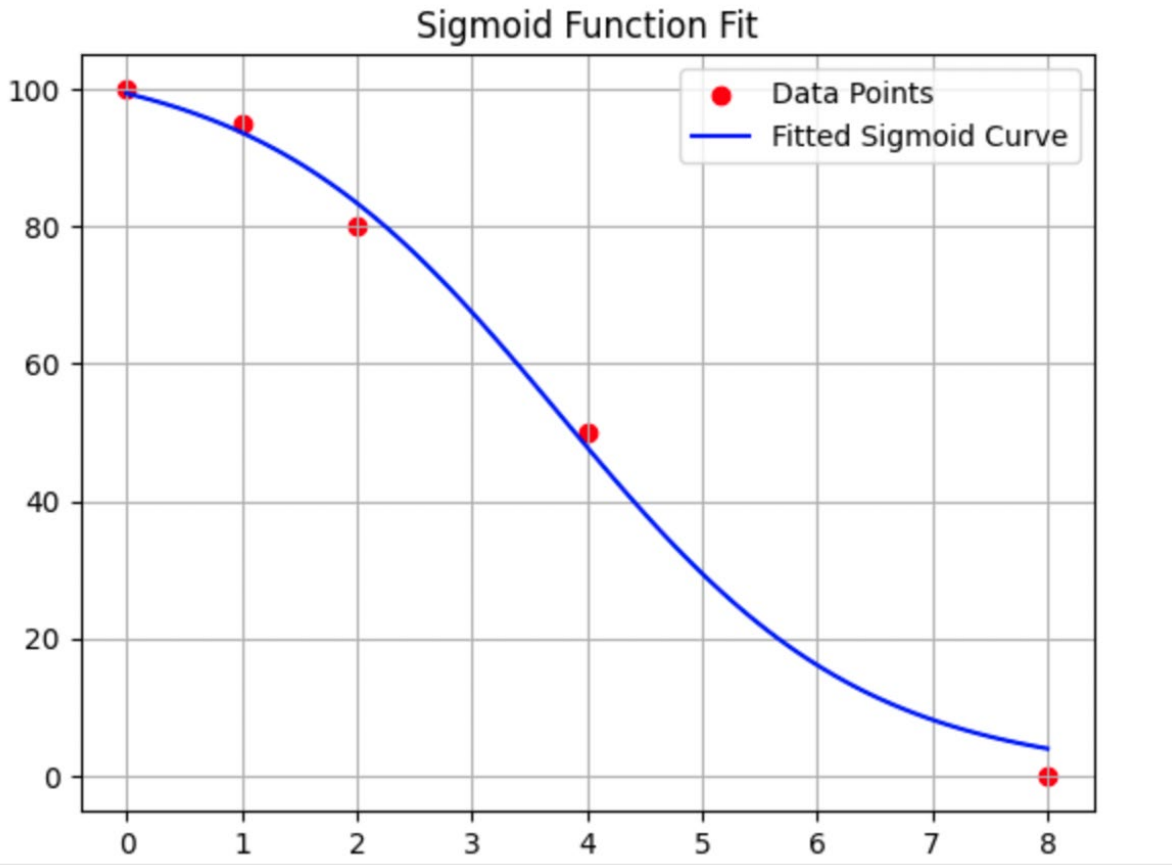
Fitting of Sigmoid Function to Local Ca Dev

Dev = 0 Score = 100

Dev = infinity Score =0

L/(1+exp(-k*(x-x0)), L=105, k=-0.77, x0=3.77.

σ_Dev



CA Dev (Å)

Dev = 0 Å: σ_Dev ≈ 100

Dev = 1 Å: σ_Dev ≈ 95-99

Dev = 2 Å: σ_Dev ≈ 80-90

Dev = 4 Å: σ_Dev ≈ 50-70

Dev = 8 Å: σ_Dev ≈ 0 - 20

Used these points for fitting: (0, 100), (1, 95), (2, 80), (4, 50) and (8, 0)

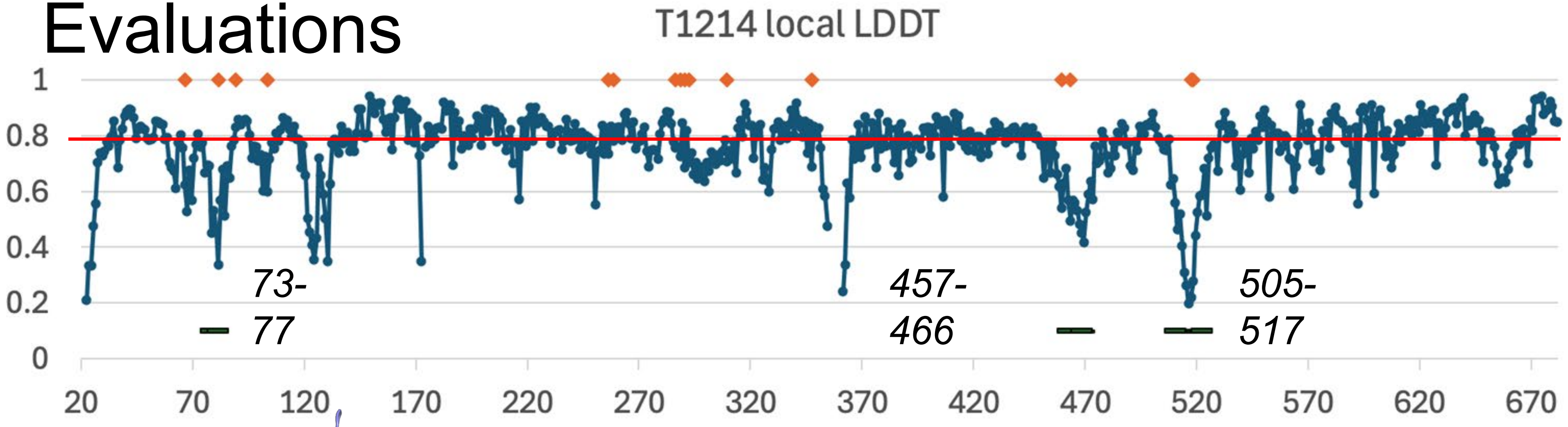
Composite Scores for Individual Domains

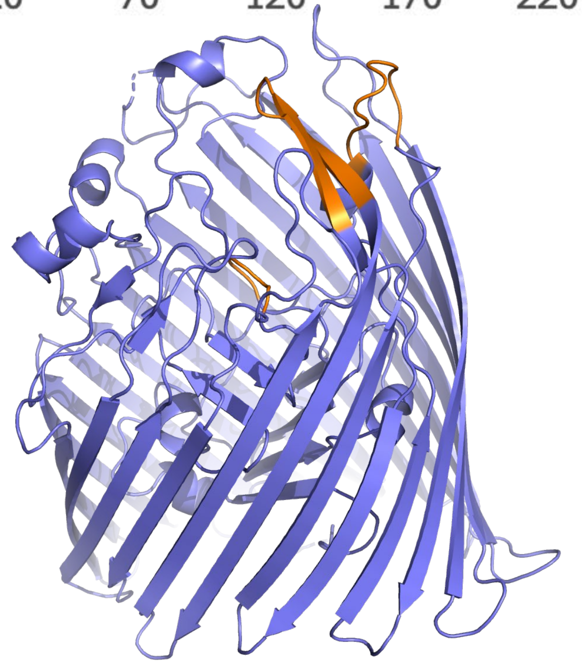
	LDDT	GDT_TS	σ_RMSD	local_LDDT	σ_Dev
Σ1	1.0*val	1.0*val	1.0*val	1.0*val	1.0*val
Σ2	1.0*val	1.0*val	--	1.0*val	1.0*val
Σ3	δ	δ	δ	1.0*val	1.0*val
Σ4	δ	δ	1.0*val	1.0*val	1.0*val

δ = 100.0 if value is > 80
else δ = 0

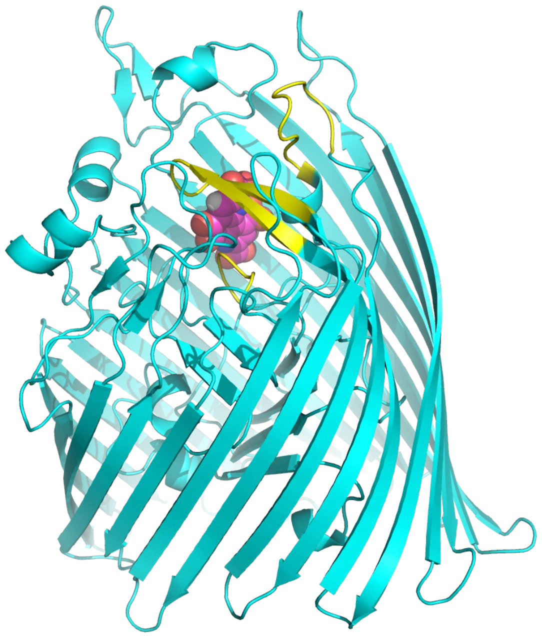
local_LDDT and σ_Dev are computed as average over residues in “regions of interest”

Regions of Interest (RoI) for Local Score Evaluations

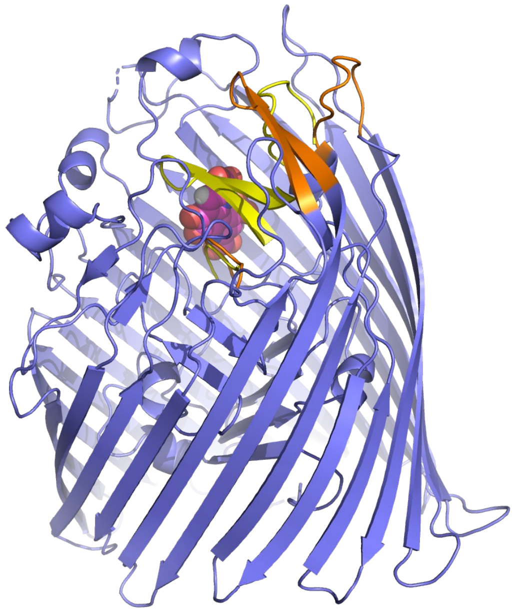




apo



holo

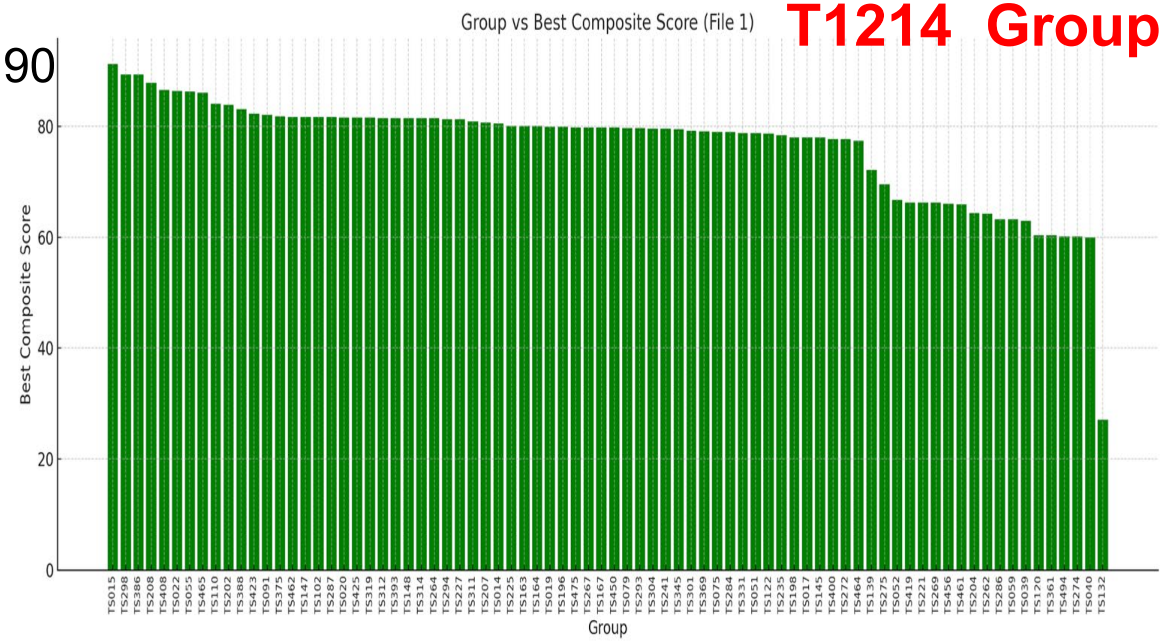


orange - apo
yellow – bound to PQQ

T1214 Group Rank by Best Model

90

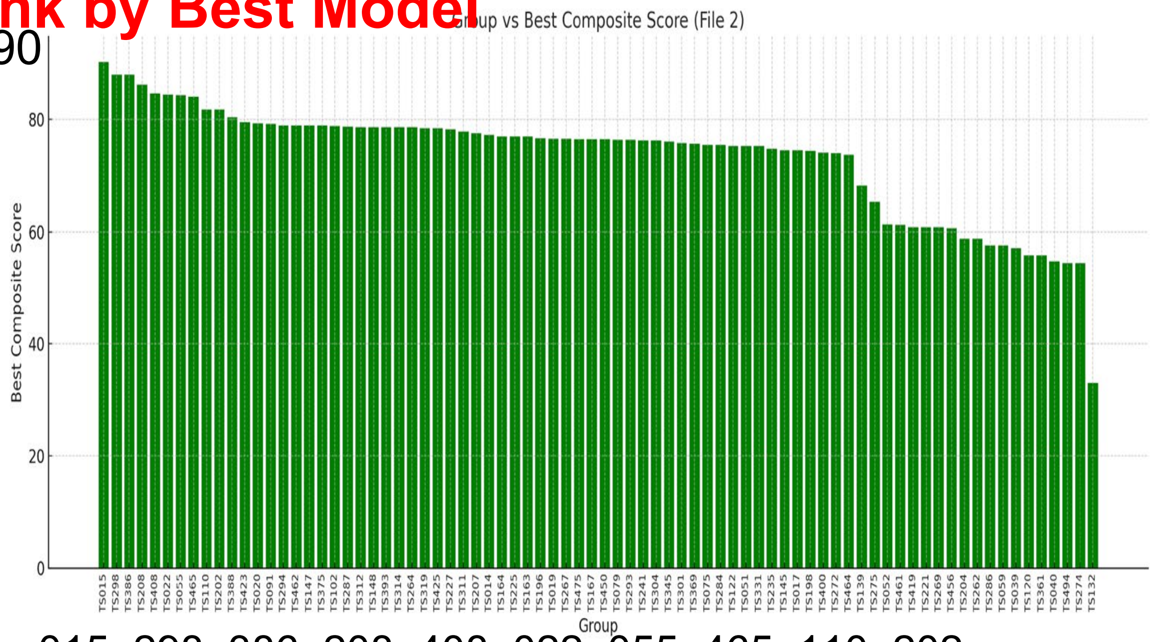
Σ1



015, 298, 386, 208, 408, 022, 055, 465, 110, 202

90

Σ2



015, 298, 386, 208, 408, 022, 055, 465, 110, 202

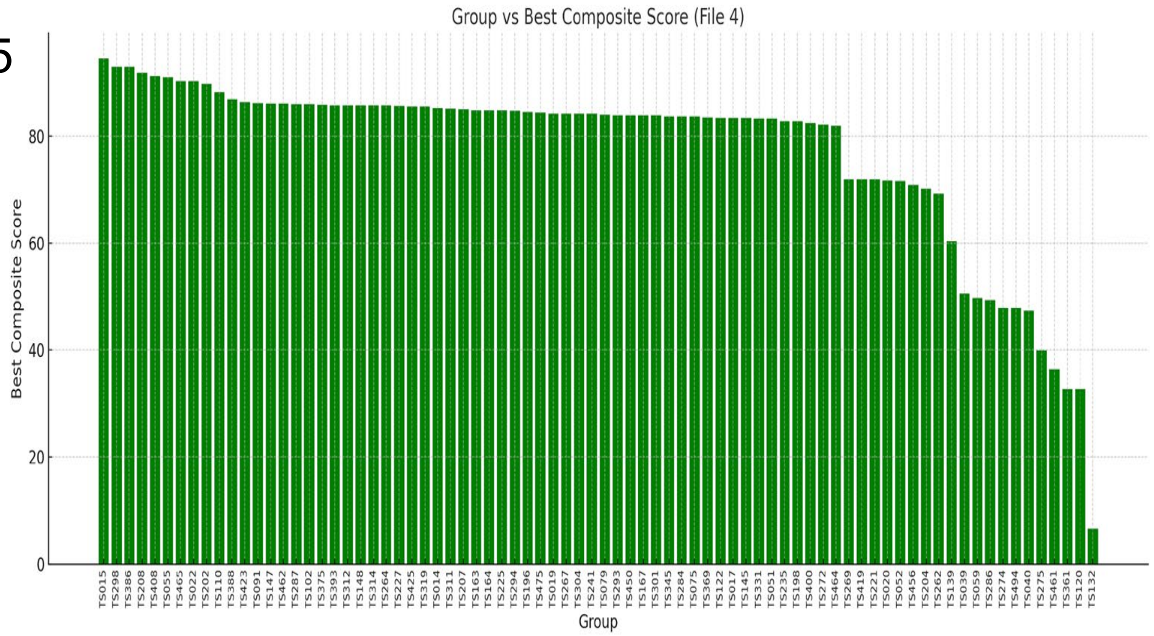
95

Σ3

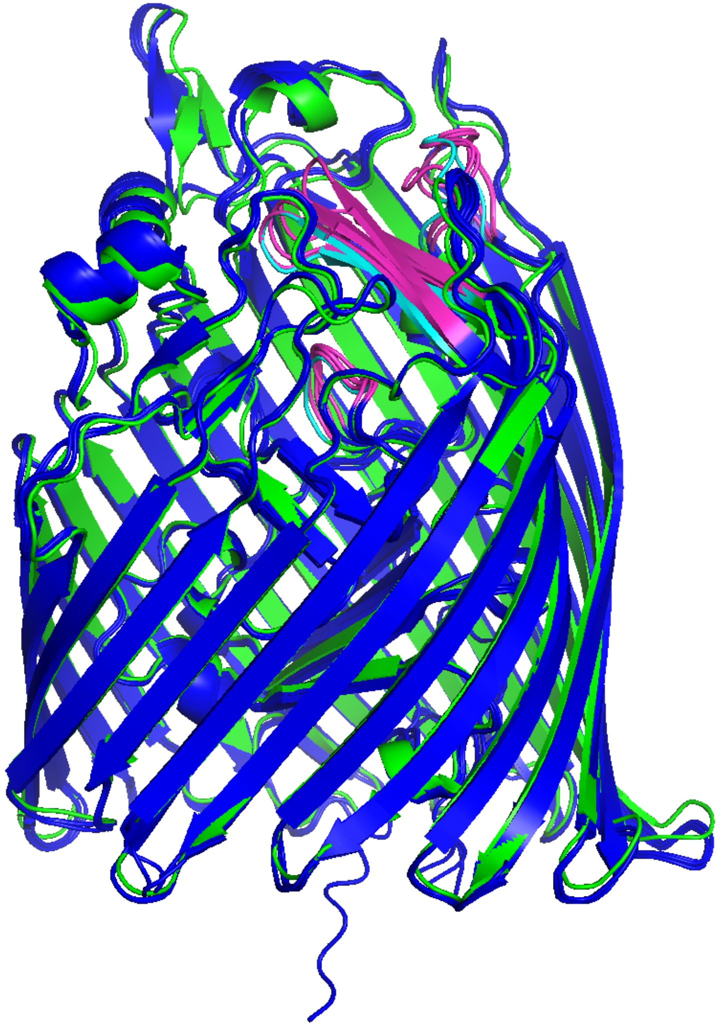
015, 386, 298, 208, 408, 055, 465, 022, 202, 110

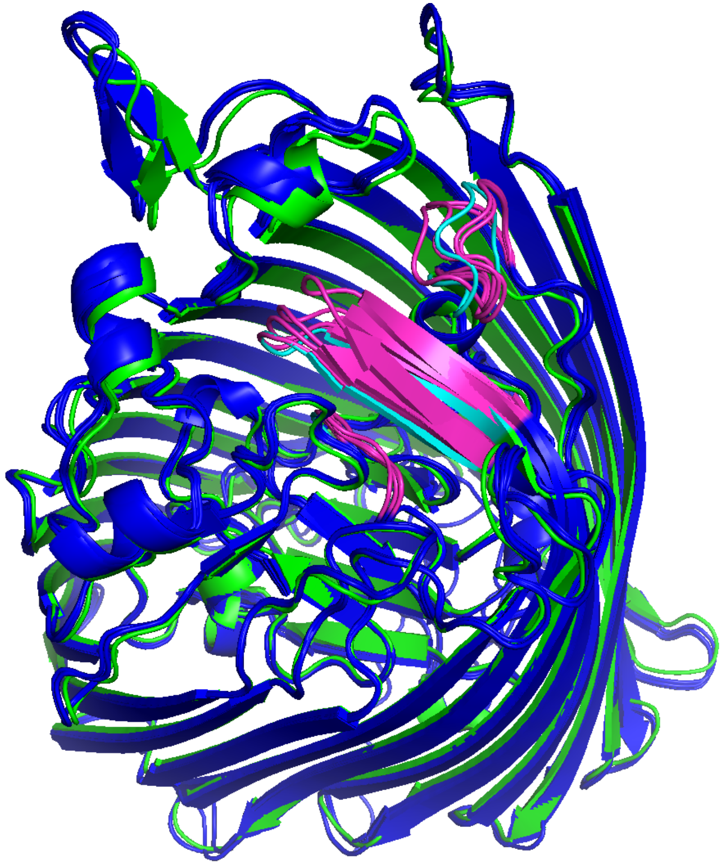
95

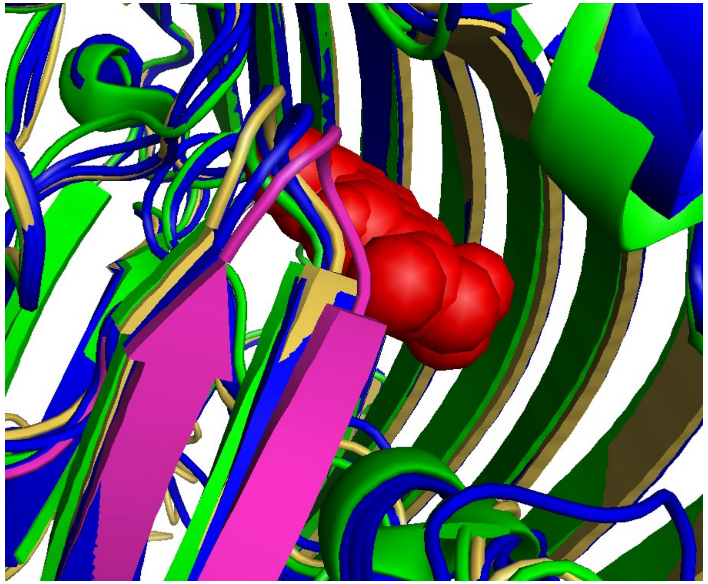
Σ4



015, 298, 386, 208, 408, 055, 465, 022, 202,110







015 – Yellow
208 – Magenta
298, 386, 408 – Blue
Xray - green

015 PEZYFoldings
298 ShanghaiTech-human
386 ShanghaiTech-Ligand
208 falcon2
408 SNU-CHEM-lig
55 LCDD-team
(462 Zheng)

Reference T1214v1
Top Groups (015, 298, 386, 208, 408, 055) Groups that don't have ligands
Regions of interest

M1228

T1228_v1, 545 residues
T1228_v2, 545 residues

Exp Method: EM

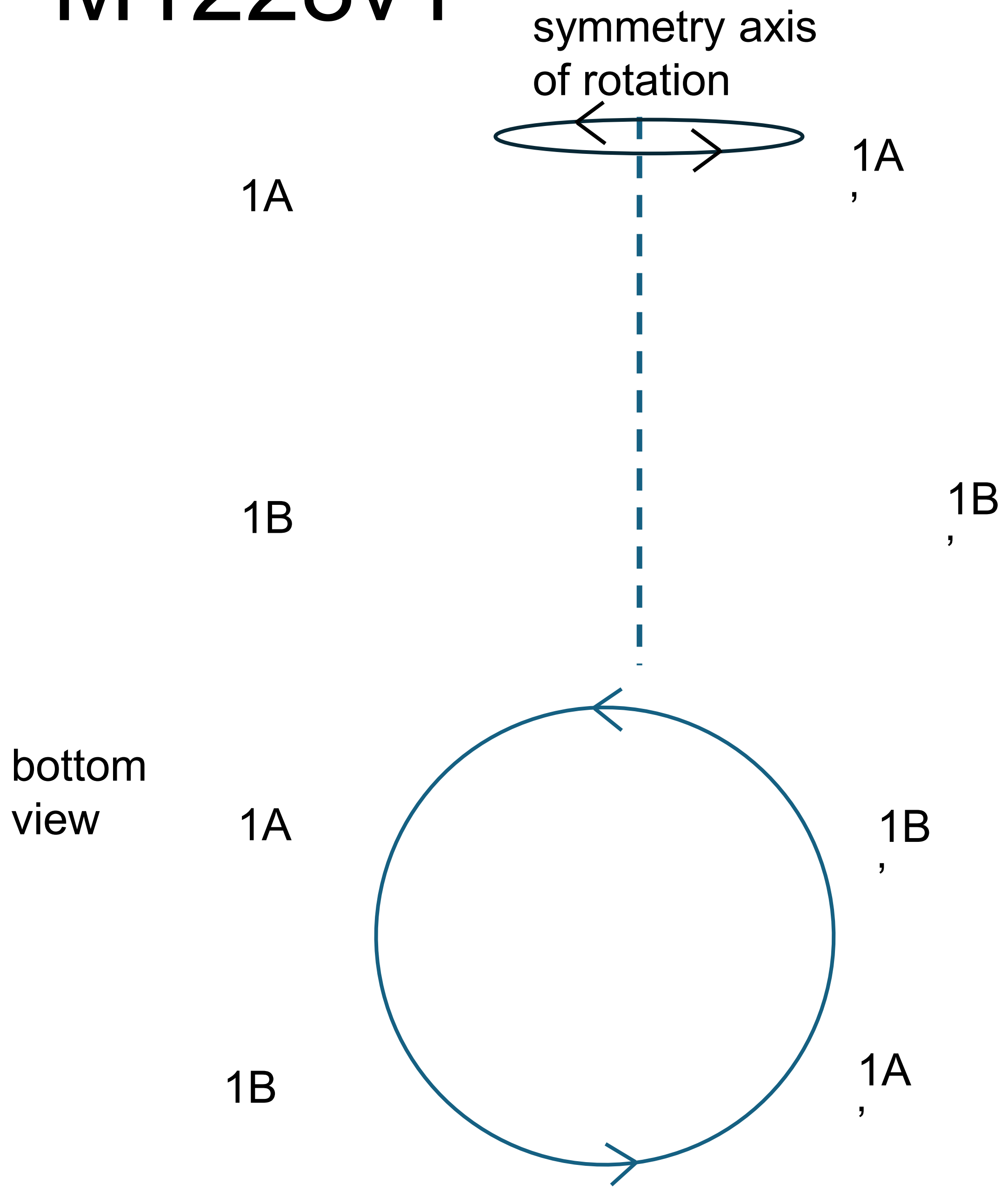
Protein subunits of M1228 complex
Two distinct conformations in the EM data

Challenge is to model both states

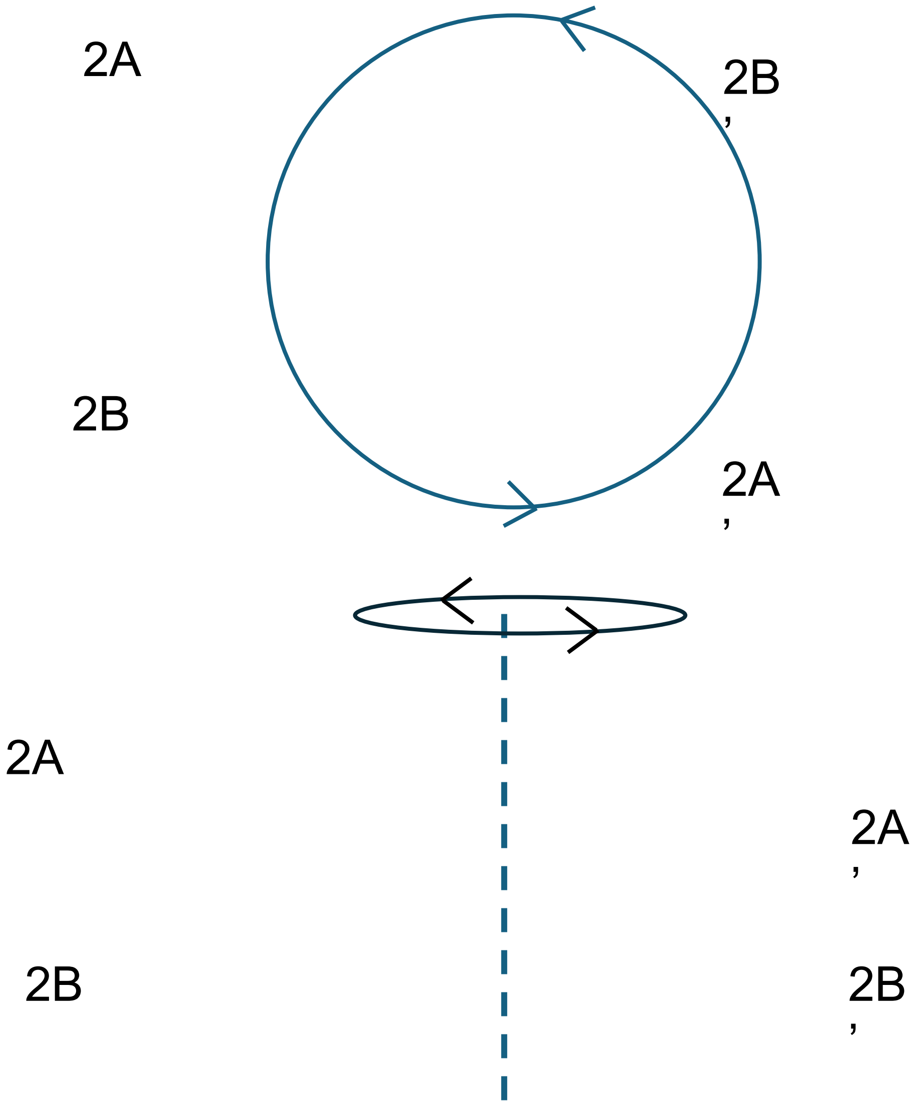
Models Submitted 5 V1 + 5 V2

Phoebe Rice, University of Chicago, Biochemistry & Molecular Biophysics

M1228v1



M1228v2



TM-score = 0.696

TM-score = 0.716

TM-score = 0.469

TM-score = 0.468

M1228v1
M1228v1TS294_1

M1228v1
M1228v1TS462_3

M1228v2
M1228v2TS052_5

M1228v2
M1228v2TS022_2

v1 state

v2 state

USalign

5 models | Predictor v1 | State 1A | V-shape

5 models | Predictor v2 | State 1B | Compact

Two State Score: Did the predictors do well in predicting two distinct states?

Find best match of all 10 models to each of two states

if best Predictor model v_m matches Model State 1A
  assign v_m = 1A
  chose best v_n to 1B
  compute Σ* = Σ_m_1A + Σn_1B

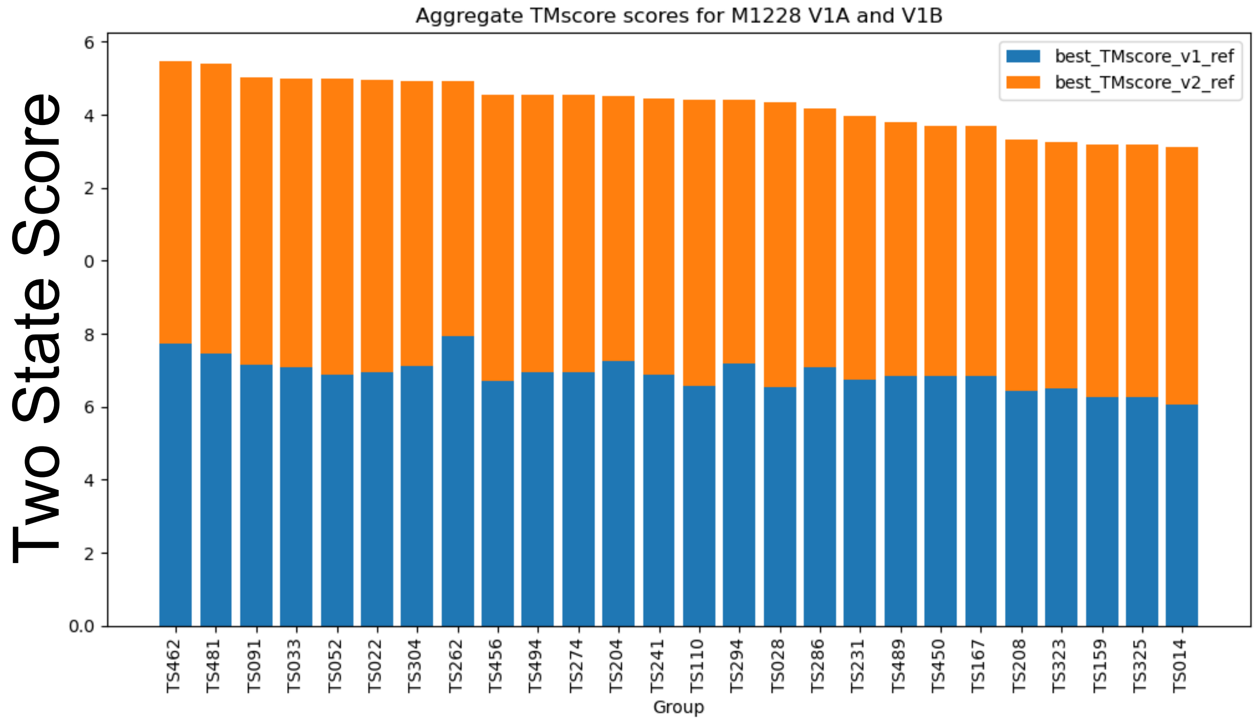
else

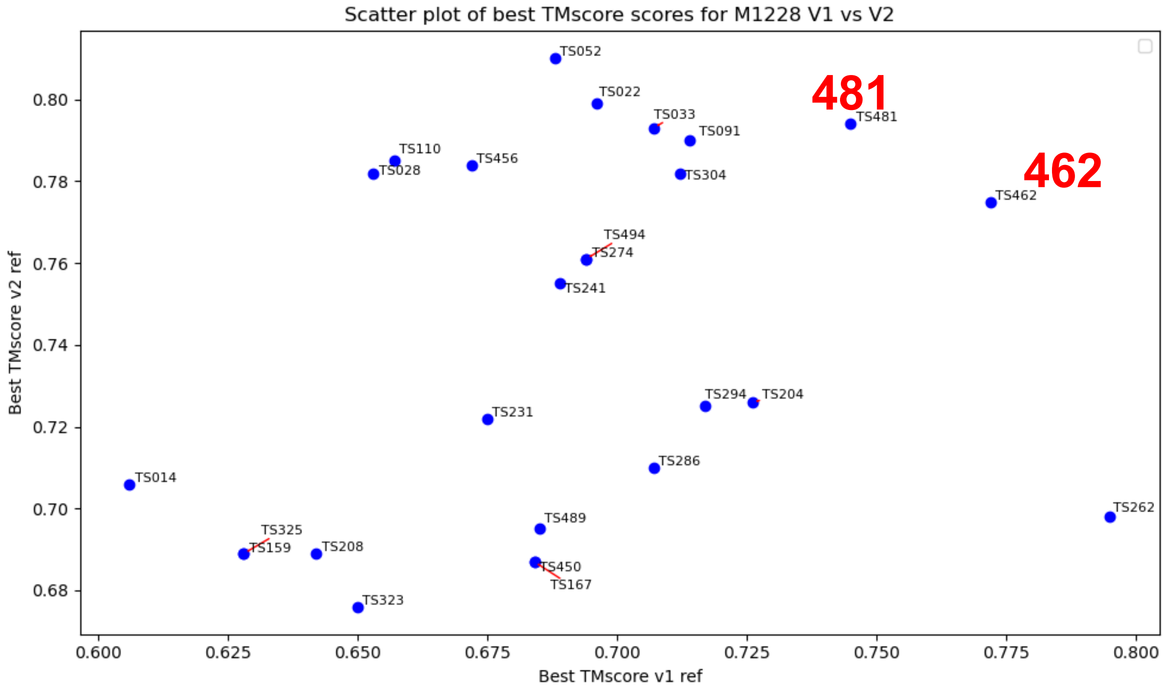
if best Predictor model v_m matches Model State 1B
  assign v_m = 1B
  chose best v_n to 1A
  compute Σ* = Σ_m_1B + Σn_1A

Rank groups by Σ*

Used Σ = GDT_TS score

TM scores





462 **Zheng**
481 **Vfold**
091 **Huang-HUST**
033 **Diff**
052 **Yang-Server**
304 **AF3-server**
262 **CoDock**

Ranking for M1228
No group did very well with both states - best cases were 462 Zheng / 481 Vfold

Group 462
Zheng

TM-score = 0.716

TM-score = 0.438

v1

v2

M1228v1
M1228v1TS462_3

M1228v2
M1228v1TS462_1

top dimers have
opposite conformation

T1228v1/v2

1A

1B

V-shaped

compact

2A

2B

bottom dimers have
(almost the) same
conformation

chains 1A and 2A
(0.93 global
LDDT)

chains 1B and 2B
(0.87 global
LDDT)

T1228v1

1A

2A

T1228v2

1B

2B

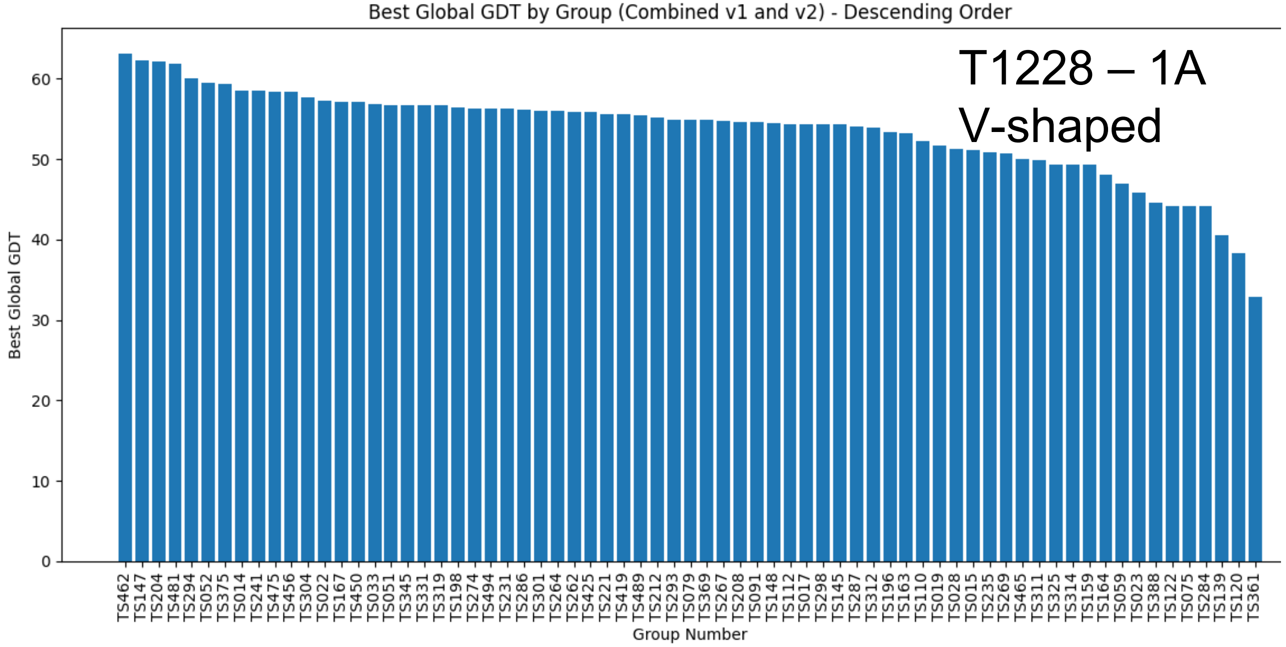
T1228v1 vs v2

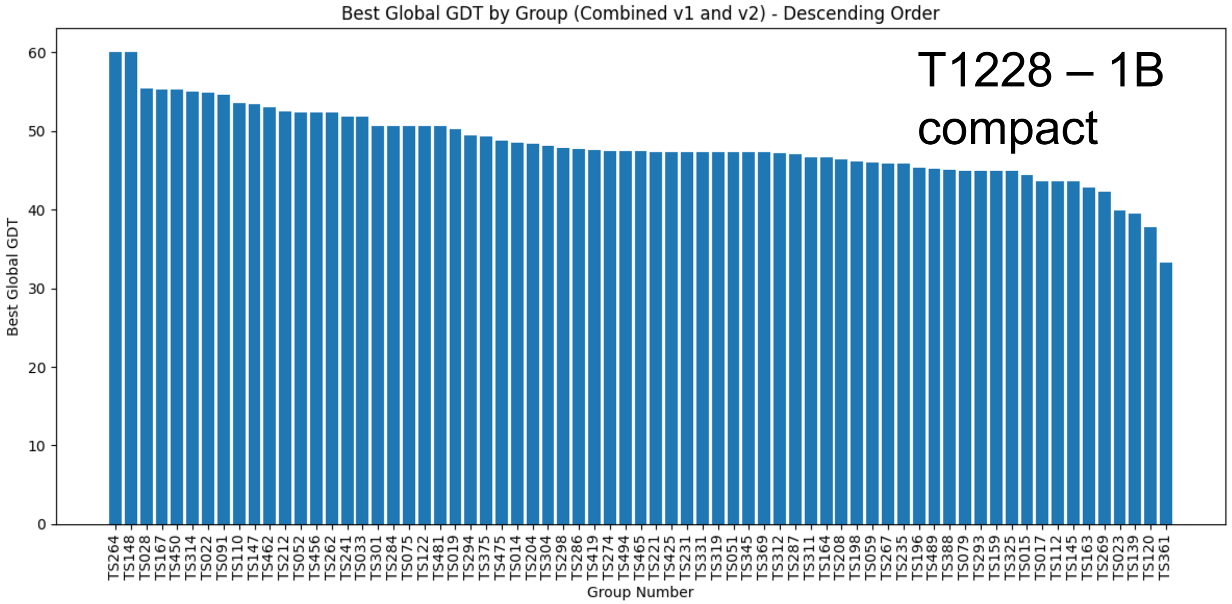
“V-shaped” state
1A, 2A

“compact” state
1B, 2B

T1228 Ranking of Best Model by Group against 2 Reference States

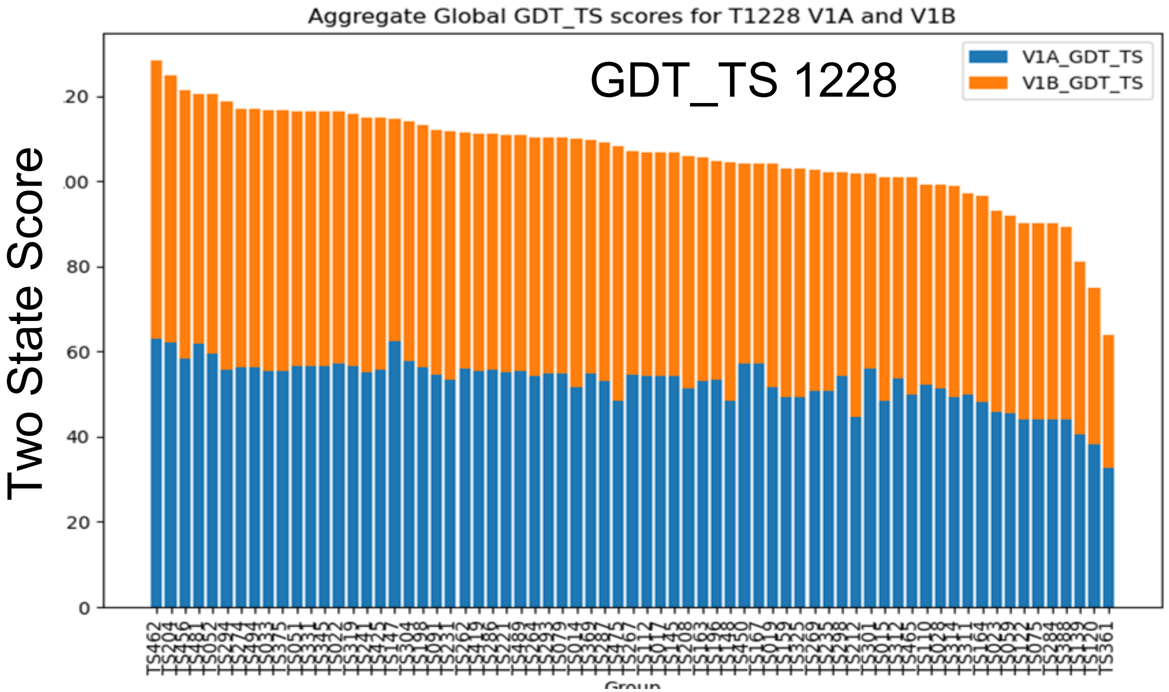
GDT_max ~ 62

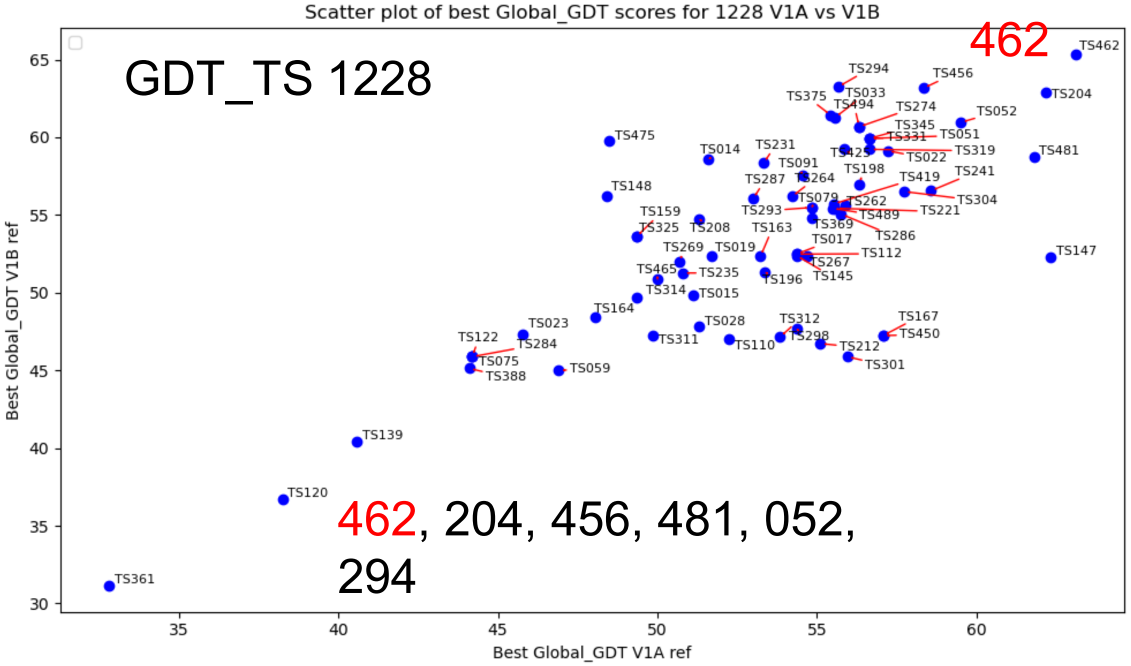


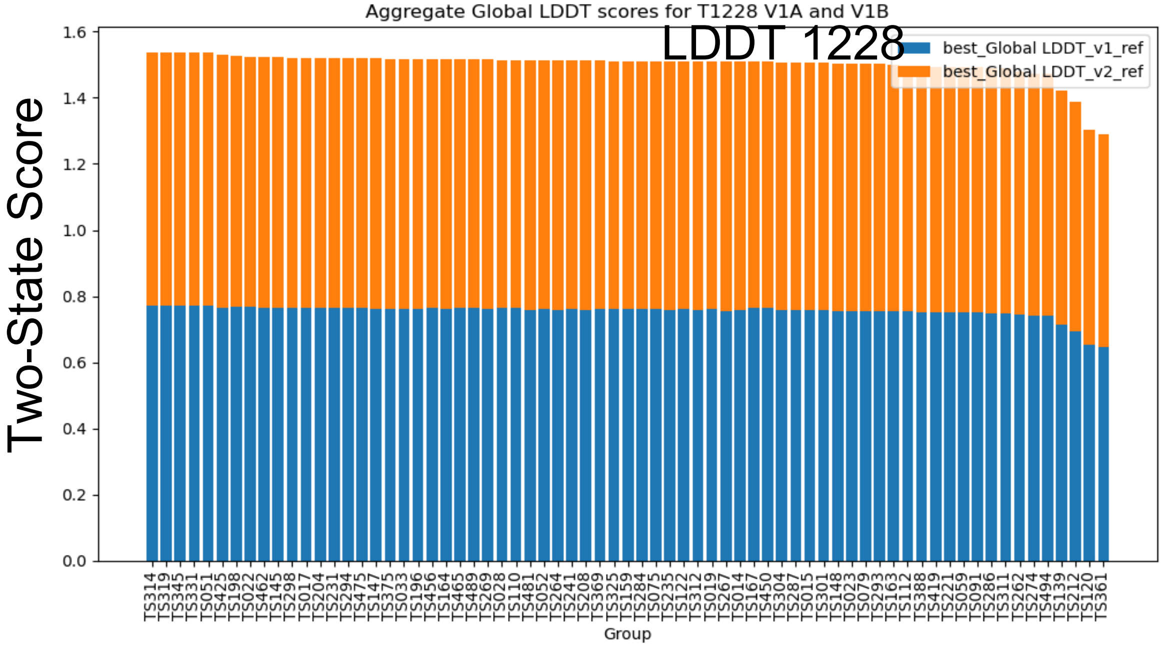


462, 147, 204, 481, 294,
052, 375, 014, 241, 475

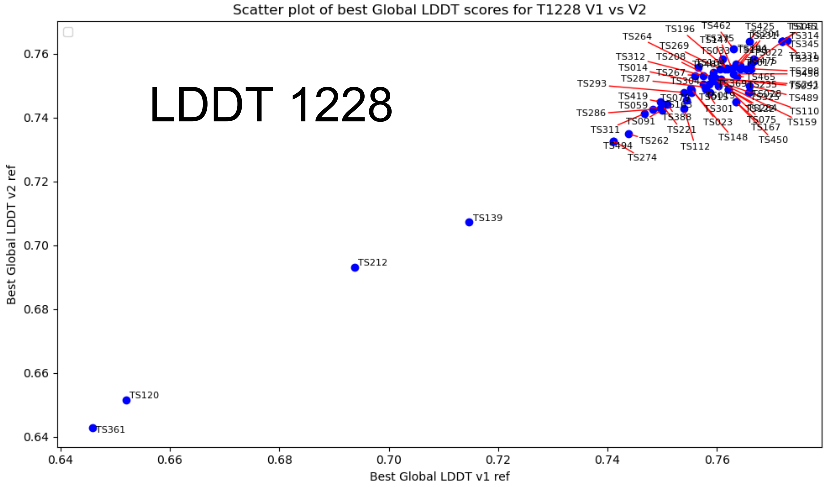
264, 148, 028, 167, 450,
314, 022, 091, 110, 147







LDDT far less discriminatory



Ground Truth – T1228

Group 462 - Zheng

State 1A
V-shaped State

462 **Zheng**
204 **Zou**
456 **Yang-Multimer**
481 **Vfold**

State 1B
Compact State

Ranking for
T1228

D4

M1239

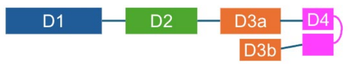
T1239_v1, 620 residues
T1239_v2, 620 residues

Exp Method: EM

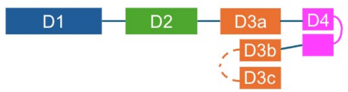
Protein subunits of M1239 complex
Two distinct conformations in the EM data

challenge is to model both states
Models Submitted 5 V1 + 5 V2

T1228



T1239



Phoebe Rice, University of Chicago, Biochemistry & Molecular Biophysics

TM-score
= 0.457

TM-score
= 0.460

TM-score
= 0.435

TM-score
= 0.489

M1239v1
M1239v1TS262_3

M1239v1
M1239v1TS294_1

M1239v2
M1239v2TS235_4

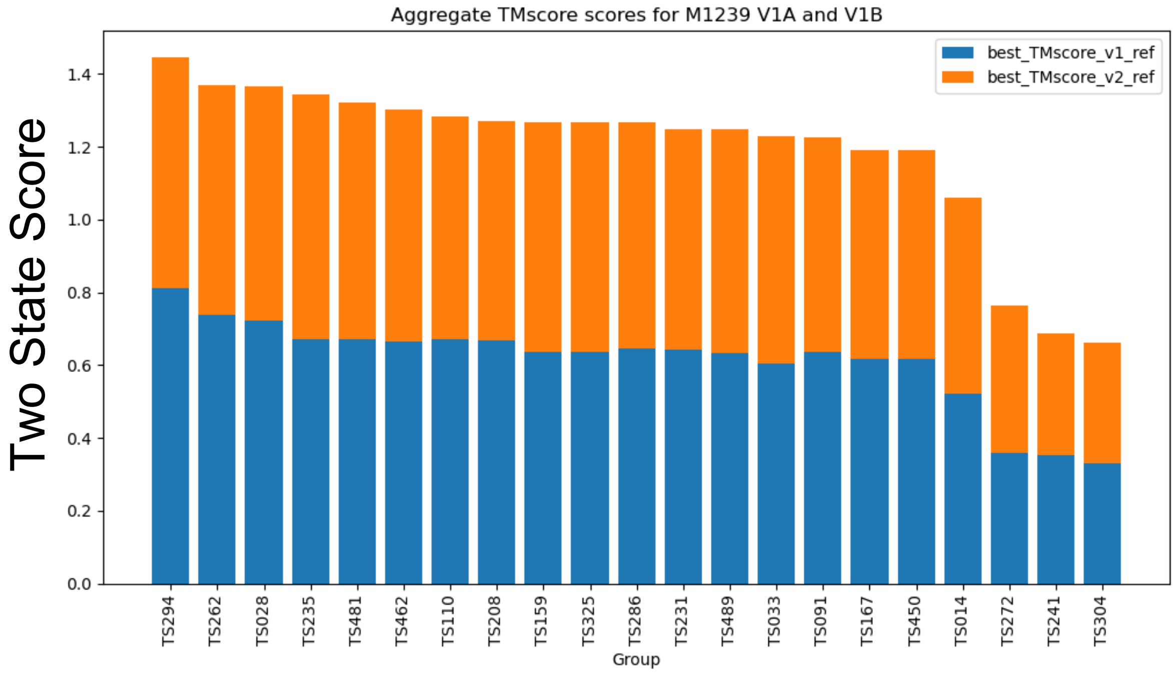
M1239v2
M1239v2TS028_2

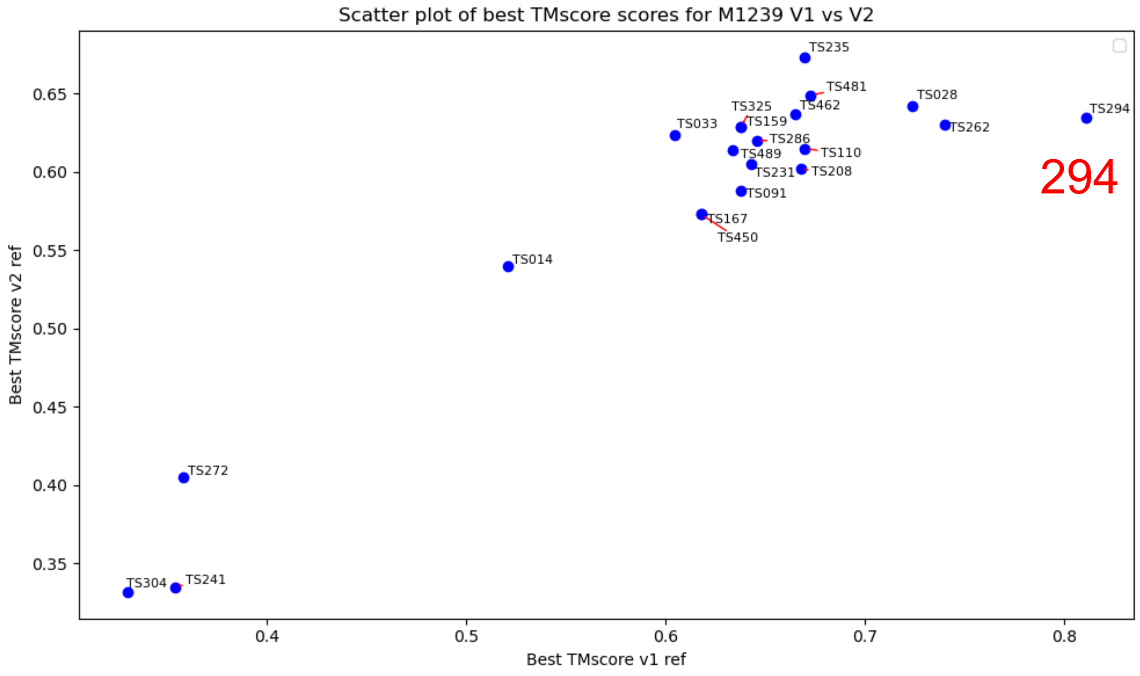
v1 state

v2 state

USalign

M1239





294 **KiharaLab**
262 **CoDock**
208 **DeepFold-server**
235 **isyslab-hust**
481 **Vfold**
462 **Zheng**
...

ranking for M1239

Group 294
KiharaLab

TM-score
= 0.460

TM-score=
0.664

M1239v1
M1239v1TS294_1

M1239v2
M1239v1TS294_2

v1 state

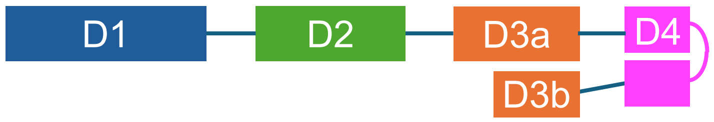
v2 state

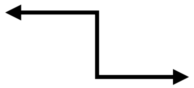
USAlign

T1228

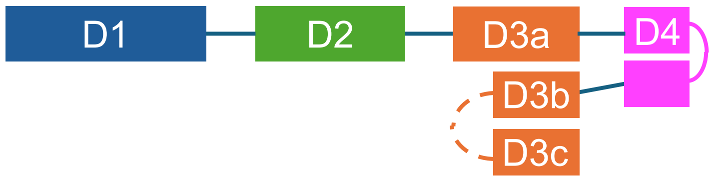
1A
V-shaped

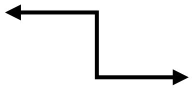
1B
Compact State

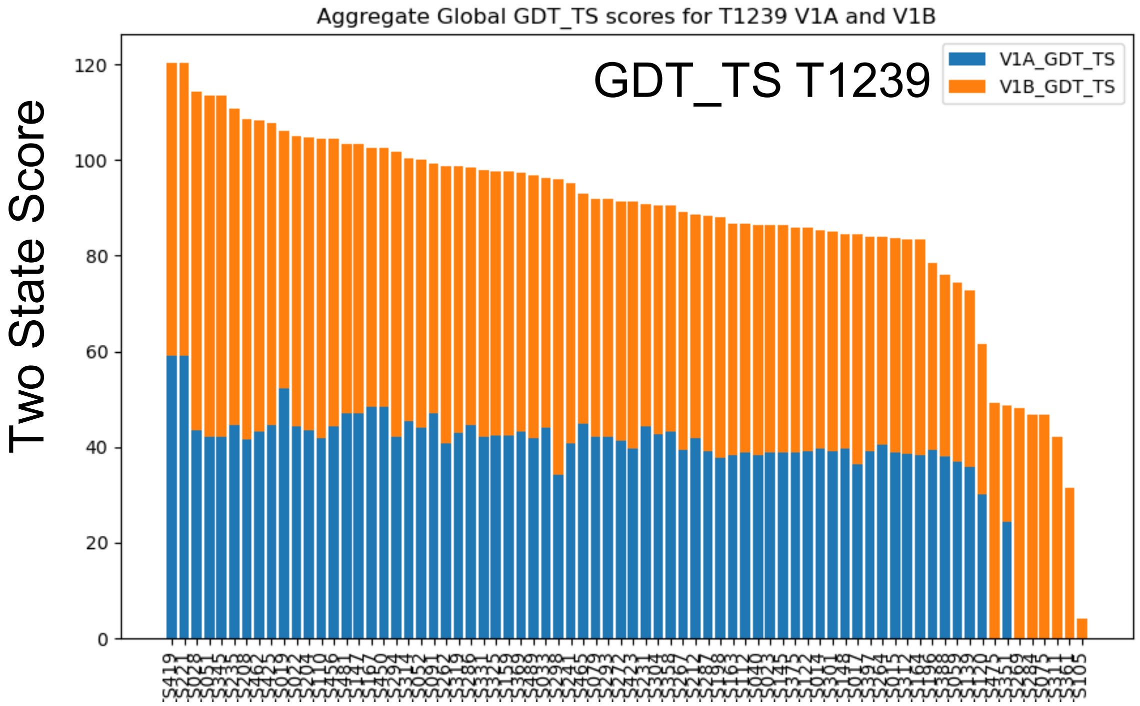


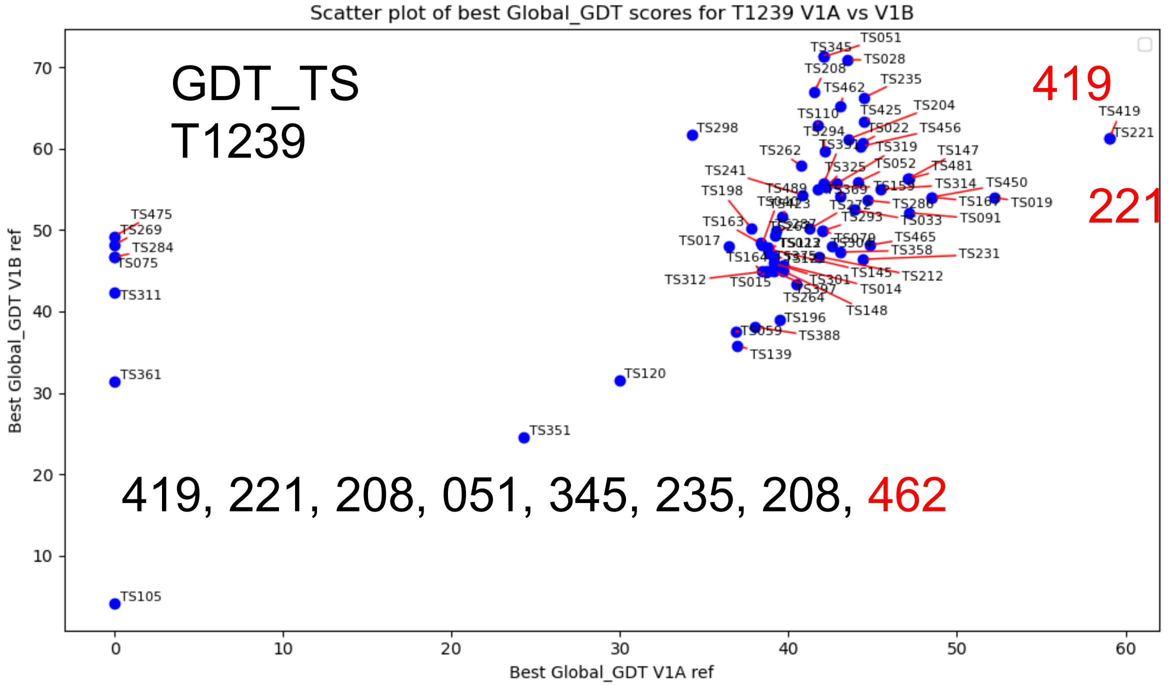


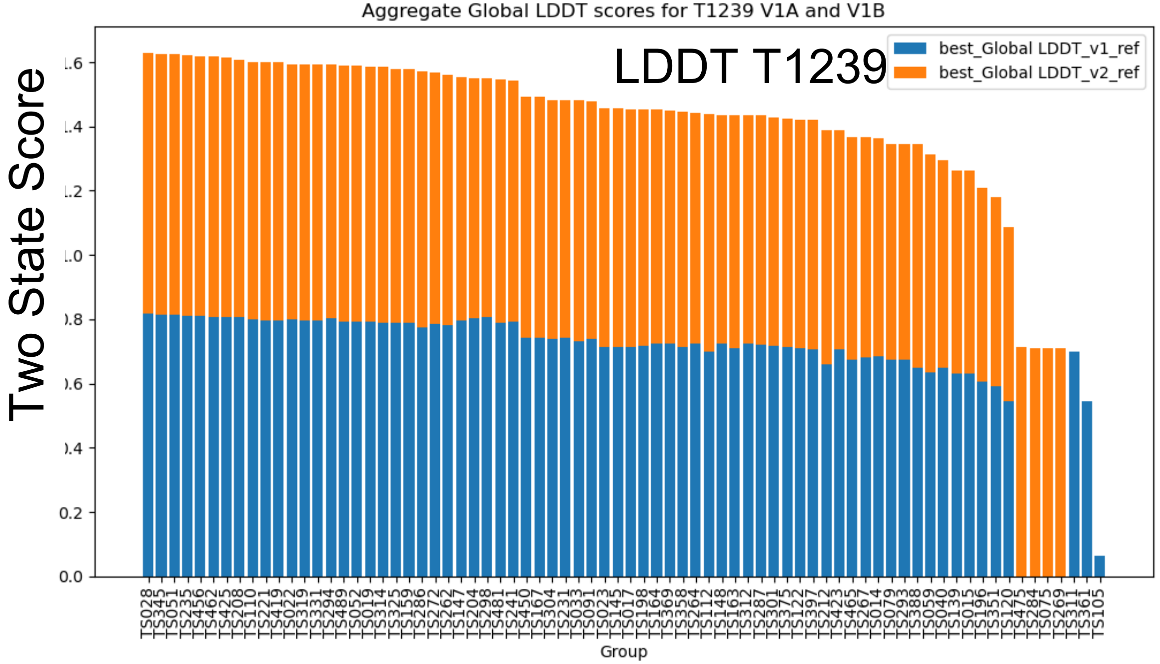
T1239



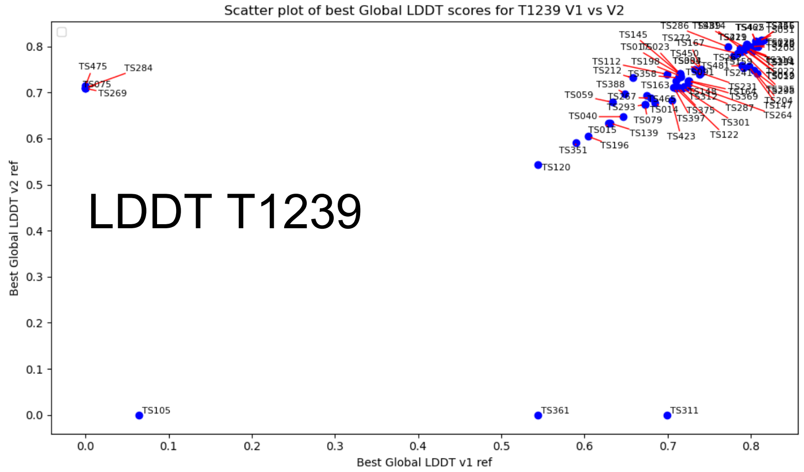








LDDT far less discriminatory



Ground Truth – T1239
Group 462

State 1A
V-shaped State
Group 462

419 **CSSB-Human**
221 **CSSB_FAKER**
028 **NKRNA-s**
051 **MULTICOM**
345 **MULTICOM_human**
235 **isyslab-hust**
208 falcon2
462 **Zheng**

State 1B
Compact State
Group 462

D4

ranking for T1239

State 1B
Compact State
Group 419

D4

Assessment Units

T1224 (1 domain – 1 target)
Global + Local composite

T1228 (4 domains – 20 targets)
1A (v1) Global only GDT_TS
1B (v1_1) Global only GDT_TS
2A (v2) Global only GDT_TS
2B (v2_1) Global only GDT_TS

1A.D1 1A.D2 1A.D3 1A.D4 Global + Local composite
1B.D1 1B.D2 1B.D3 1B.D4 Global + Local composite
2A.D1 2A.D2 2A.D3 2A.D4 Global + Local composite
2B.D1 2B.D2 2B.D3 2B.D4 Global + Local
composite

T1239 (4 domains – 10 targets)
1A (v1) GDT_TS
1B (v1_1) GDT_TS

1A.D1 1A.D2 1A.D3 1A.D4 Global + Local composite 1B.D1 1B.D2 1B.D3 1B.D4 Global + Local composite

31 Assessment Units

Assessment Units

T1224 (1 domain – 1 target)
Global + Local composite

T1228 (4 domains – 20 targets)
1A (v1) Global only GDT_TS
1B (v1_1) Global only GDT_TS
2A (v2) Global only GDT_TS
2B (v2_1) Global only GDT_TS

1A.D1 1A.D2 1A.D3 1A.D4 Global + Local composite
1B.D1 1B.D2 1B.D3 1B.D4 Global + Local composite
2A.D1 2A.D2 2A.D3 2A.D4 Global + Local composite
2B.D1 2B.D2 2B.D3 2B.D4 Global + Local composite

T1239 (4 domains – 10 targets)
1A (v1) GDT_TS
1B (v1_1) GDT_TS

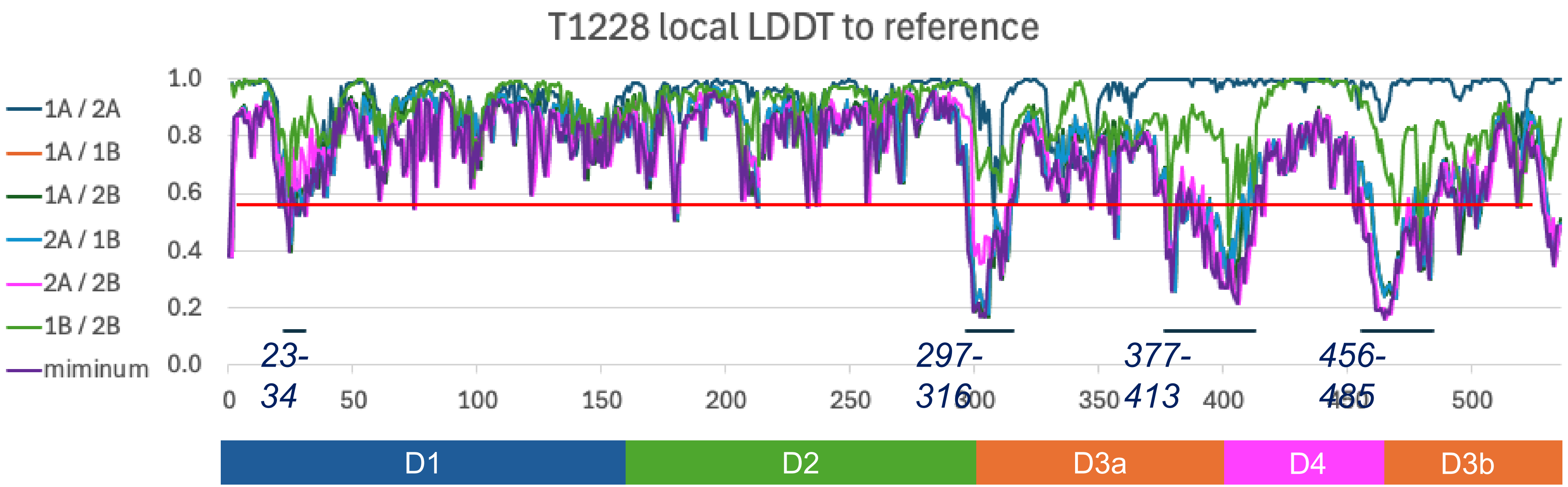
1A.D1 1A.D2 1A.D3 1A.D4 Global + Local composite 1B.D1 1B.D2 1B.D3 1B.D4 Global + Local composite

31 Assessment Units

T1228 domains

T1228: 545 residues
Domains and :
“Regions of Interest”

D1: 1 - 164
D2: 165 – 300
D3: 301 – 401, 468-535
D4: 402 – 467



D1
overlay
1A vs 1B

Composite Scores for Individual Domains

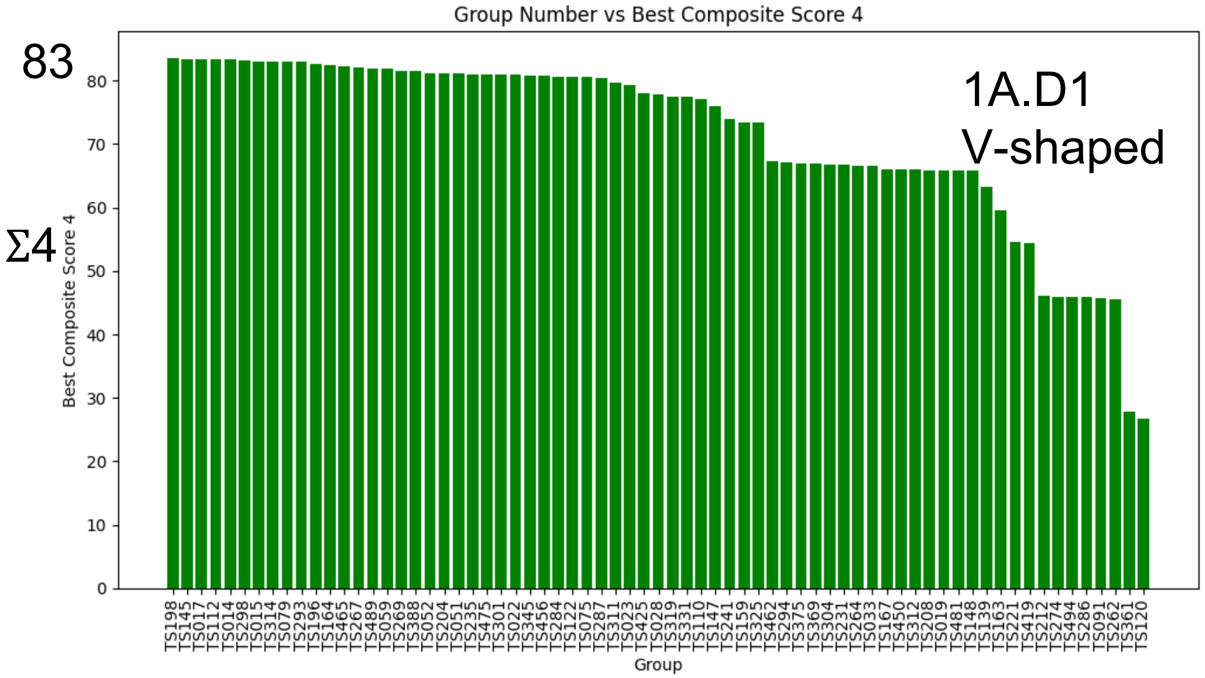
	LDDT	GDT_TS	σ_RMSD	local_LDDT	σ_Dev
Σ1	1.0*val	1.0*val	1.0*val	1.0*val	1.0*val
Σ2	1.0*val	1.0*val	--	1.0*val	1.0*val
Σ3	δ	δ	δ	1.0*val	1.0*val
Σ4	δ	δ	1.0*val	1.0*val	1.0*val

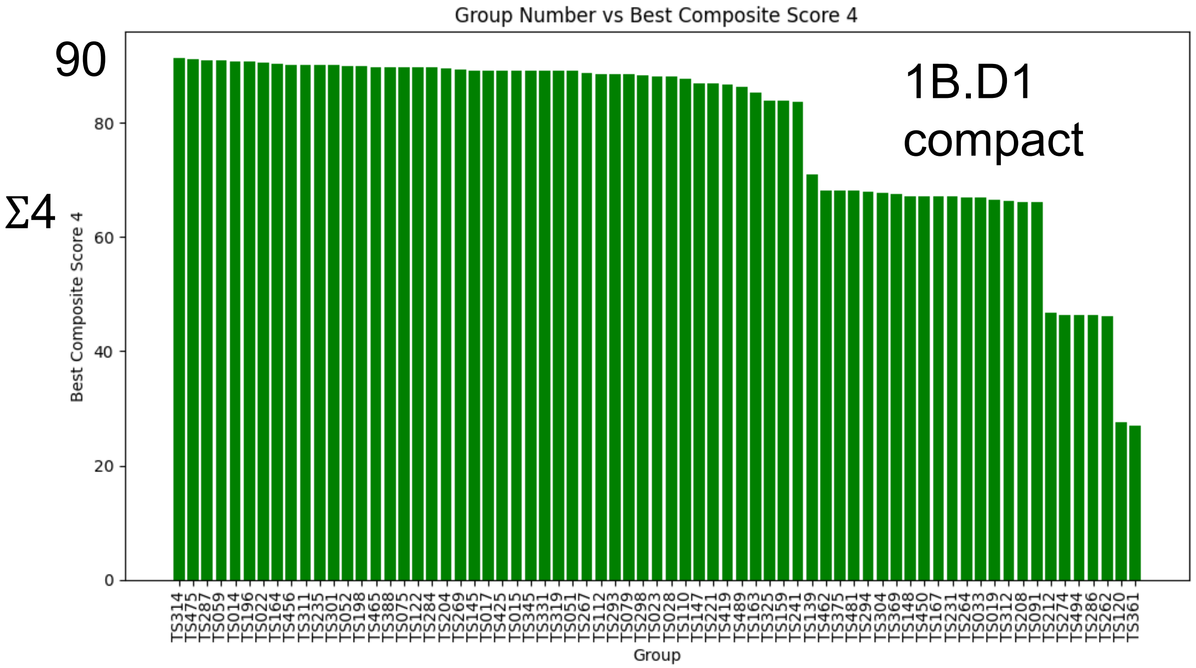
$\delta \triangleq$ 100.0 if value is > 80

else $\delta = 0$

local_LDDT and σ_Dev are computed as average over residues in "regions of interest"

T1228 Group Rank by Best Model - 2 examples of 16 reference domain states





T1224 - Composite Global and Local Score Σ4

	Rank 1	2	3	4	5	6	7	8	9	10
1A.D1	198	145	017	112	014	298	015	314	079	293
1A.D2	304	481	419	204	221	375	019	241	033	319
1A.D3	293	079	052	196	014	235	091	015	465	267
1A.D4	419	221	267	015	369	462	139	120	465	481
1B.D1	314	475	287	059	014	196	022	164	456	311
1B.D2	147	019	091	241	456	052	462	022	450	167
1B.D3	051	048	345	456	311	022	235	052	167	450
1B.D4	148	286	304	033	419	221	204	014	262	264
2A.D1	052	456	475	022	388	196	312	314	051	462
2A.D2	204	241	287	475	481	312	212	425	319	331
2A.D3	293	079	052	489	196	235	147	198	267	231
2A.D4	267	286	369	375	241	284	139	462	120	164
2B.D1	462	147	287	425	319	331	051	345	075	284
2B.D2	110	314	462	028	147	450	167	019	418	264
2B.D3	051	475	311	345	235	456	022	052	167	450
2B.D4	314	033	241	262	294	148	462	204	231	345

some top groups

462	**Zhang**	4
450	**OpenComplex_Server**	4
304	**AF3-server**	2
419	**CSSB-Human**	2
304	**AF3-server**	2

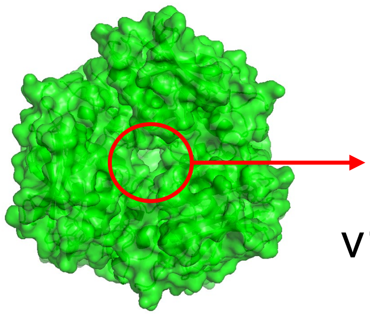
Ranking based on Composite Score does not discriminate performance between groups - many groups did equally well

Similar results with domains of T1239

T1249 v1/v2

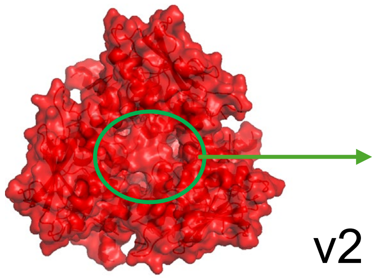
- 488 residue long
- A3, Trimer
- Cryo-EM structure of an arenaviral spike complex
- 5+5 conformations
- EM experimental method
- 2 distinct conformations in EM data (open/closed)
- Metal ions are present in both conformations, but their coordination sites changes

Ron Discin lab



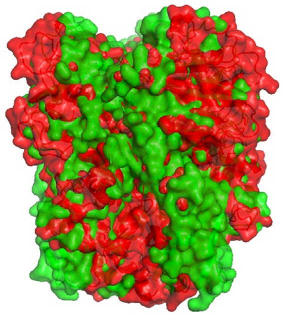
opening

v1-open

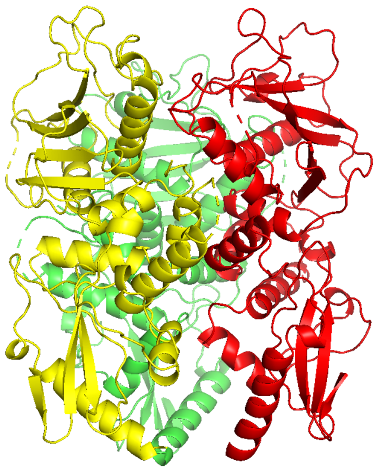


no openings

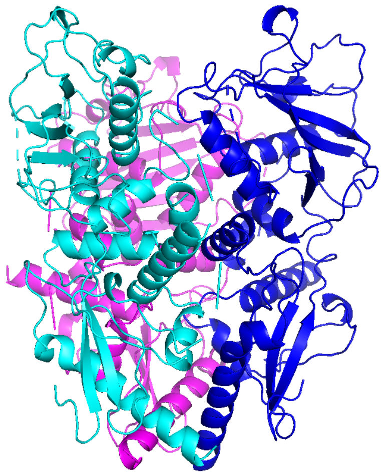
v2 - closed



v1



v2

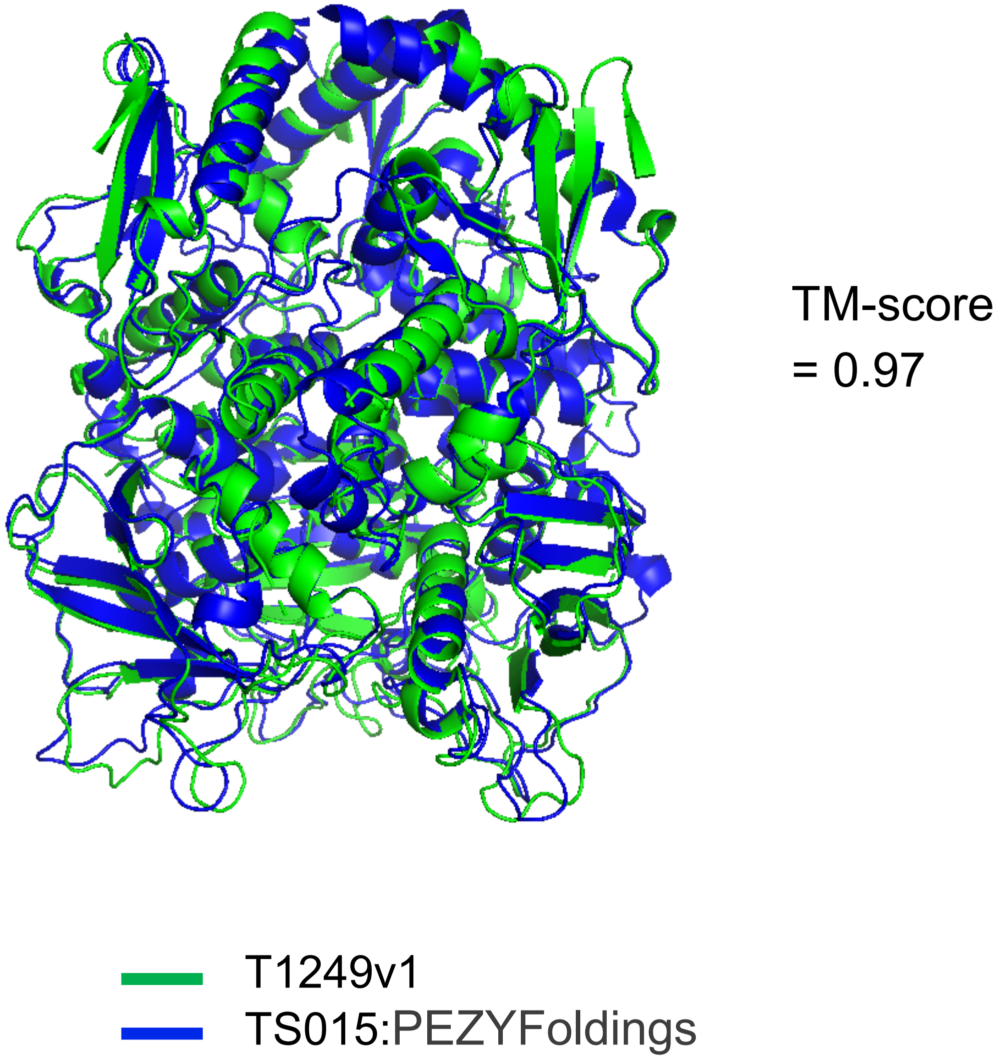


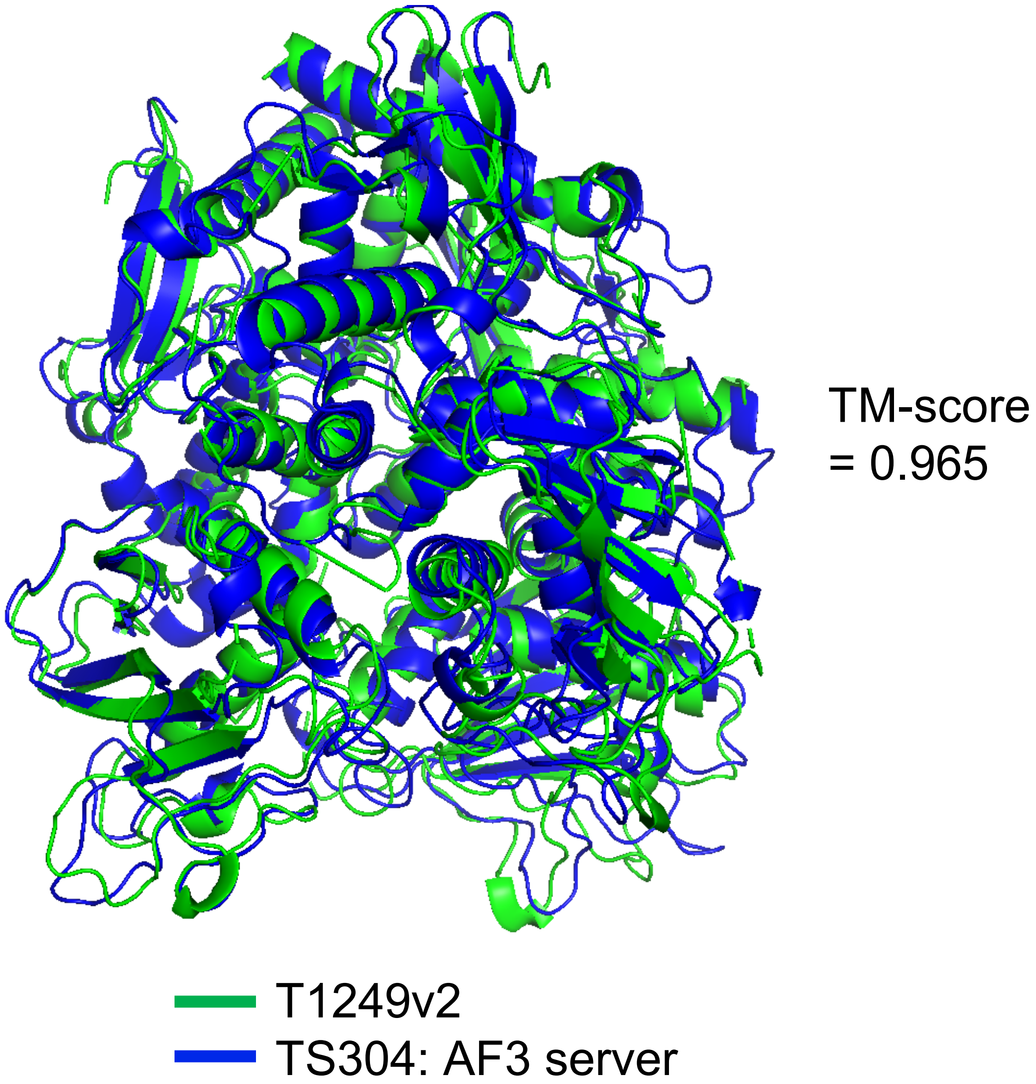
TM-score= 0.838

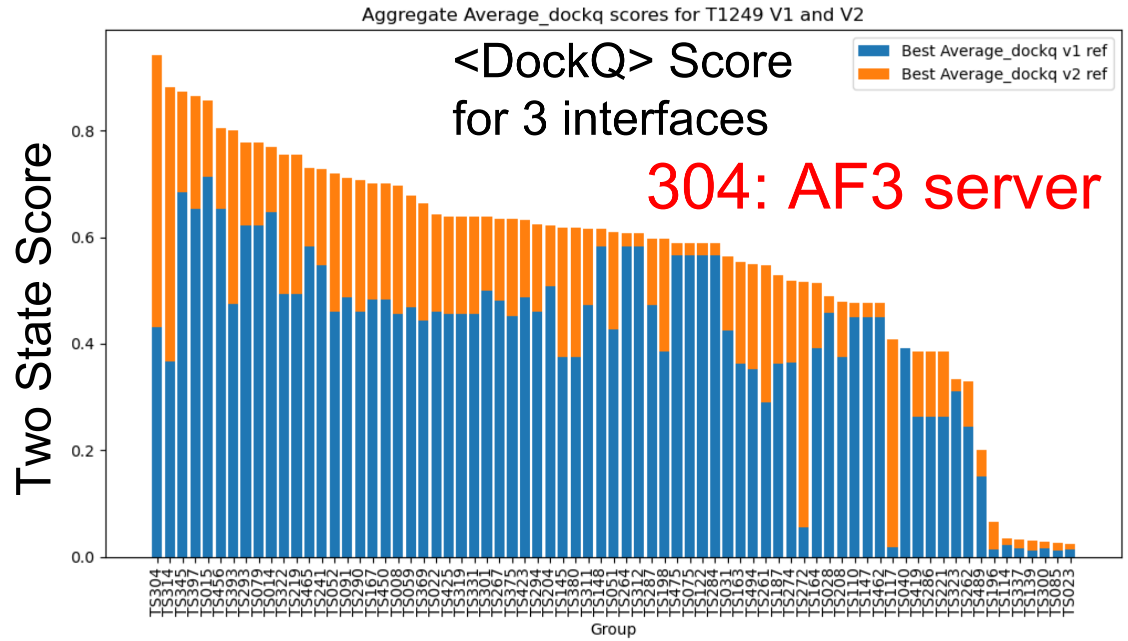
v1/v2

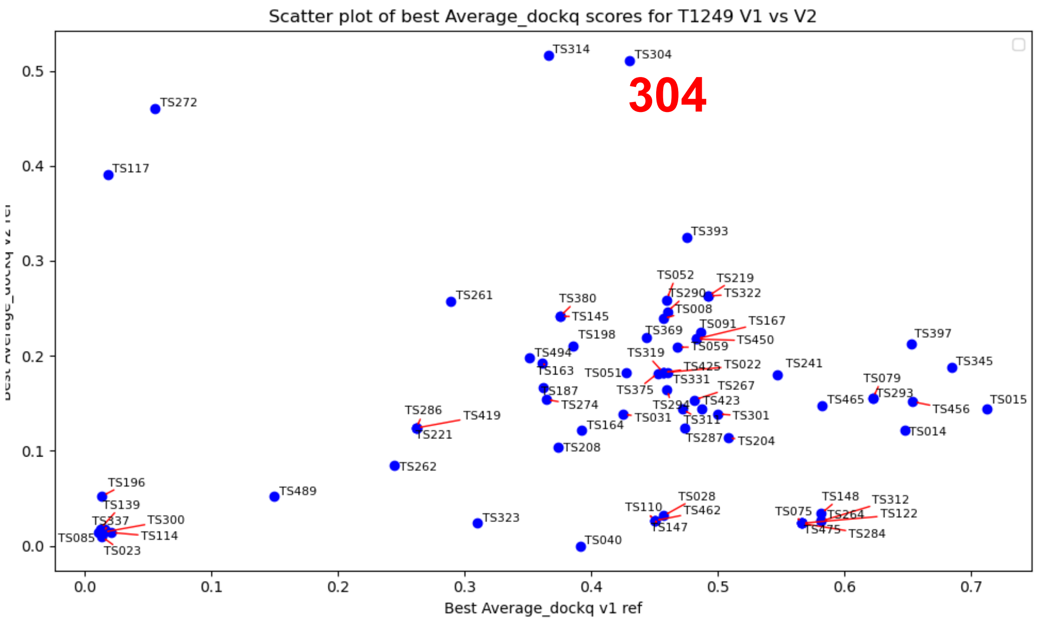
Scores used:

- **Local Quality**: lDDT provides insight into local structural accuracy.
- **Interface Quality**: DockQ_Avg assesses the quality of protein-protein interfaces.


TM-score
= 0.97
T1249v1
TS015:PEZYFoldings


TM-score
= 0.965
T1249v2
TS304: AF3 server





ranking for T1249 304,314,345,397,015,456,393,293,079,014,322

TM-score = 0.945

TS304_2o : AF3 server for V1

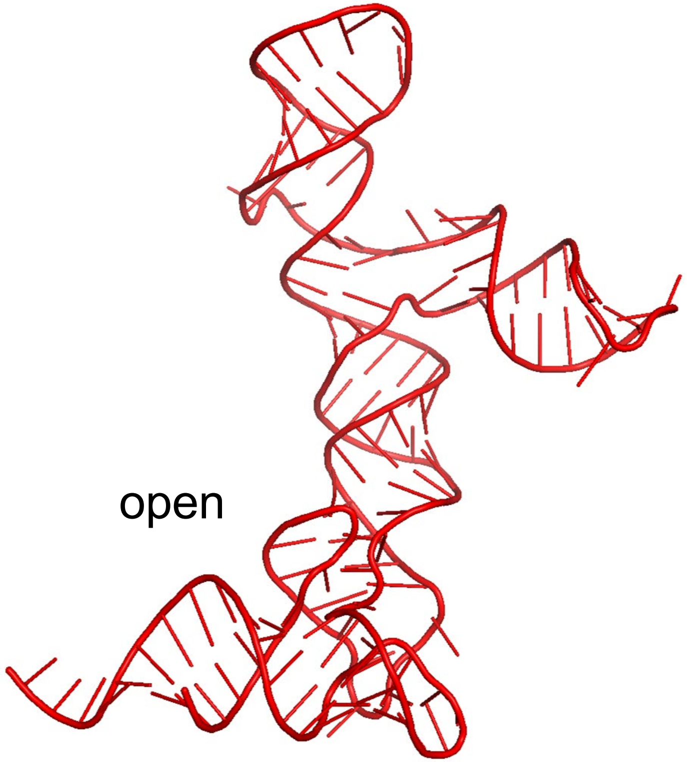
TM-score = 0.965

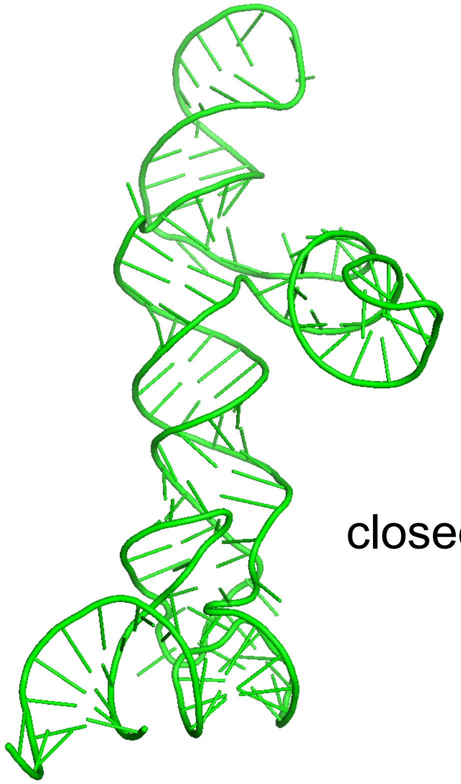
TS304_1o: AF3 server for V2

R1203

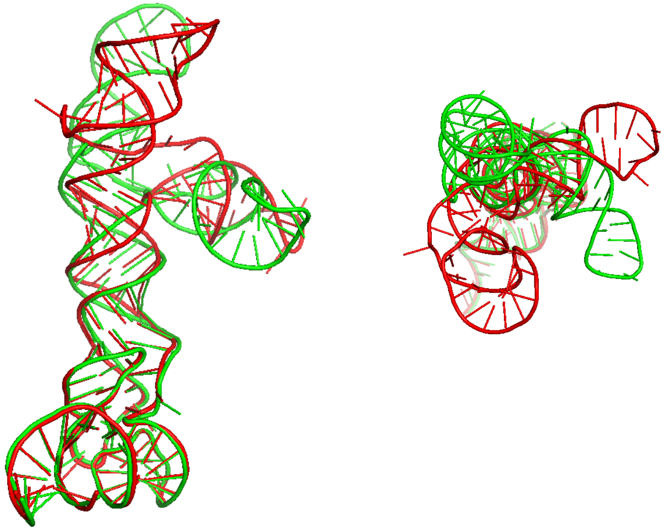
HIV-1 Rev Response Element (RRE) Stem -Loop II (SLII) open and closed conformations

- 134 residue long
- R1: RNA monomer
- 5 models submitted
- Structure determined in 2 conformations



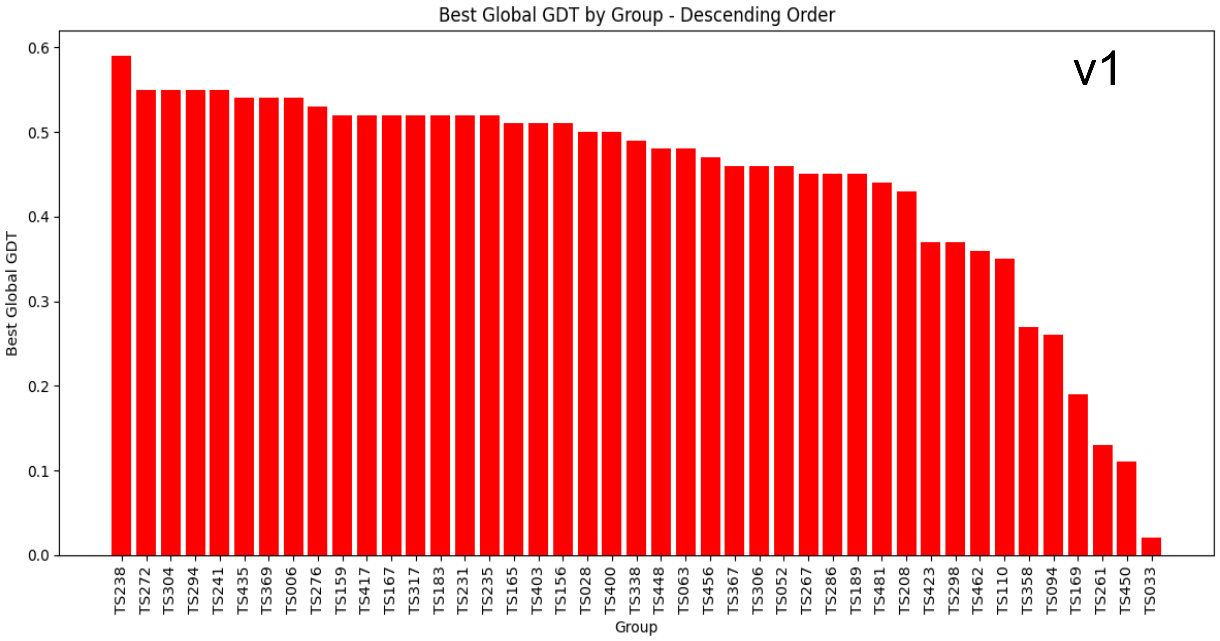


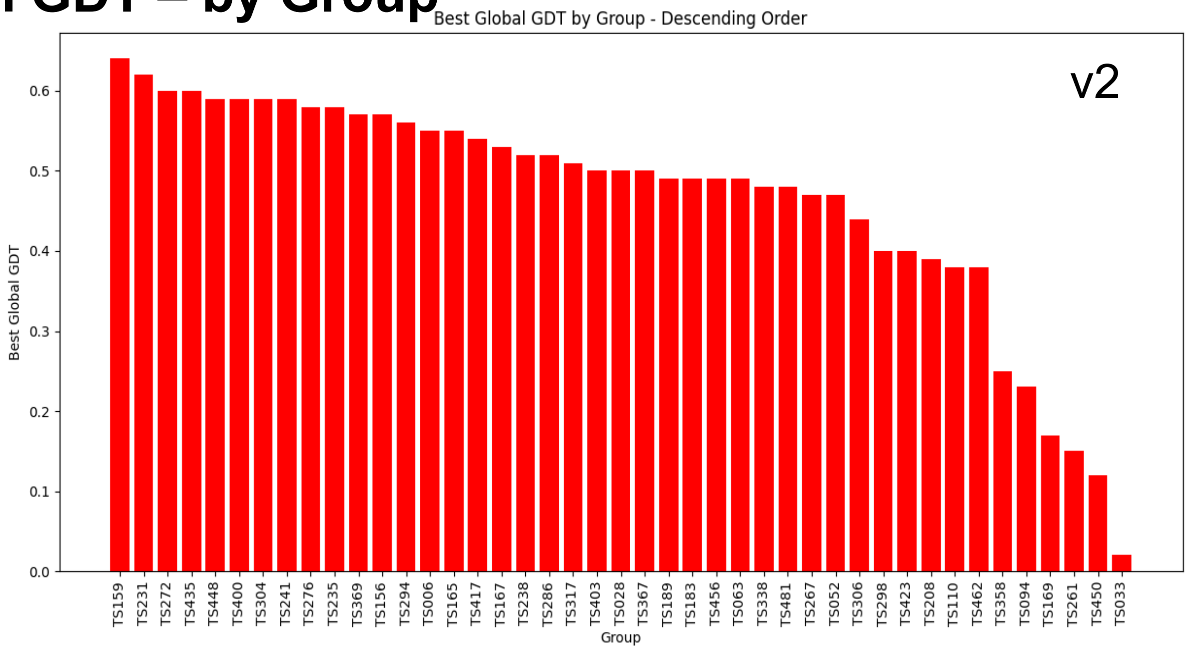
TMScore =0.575

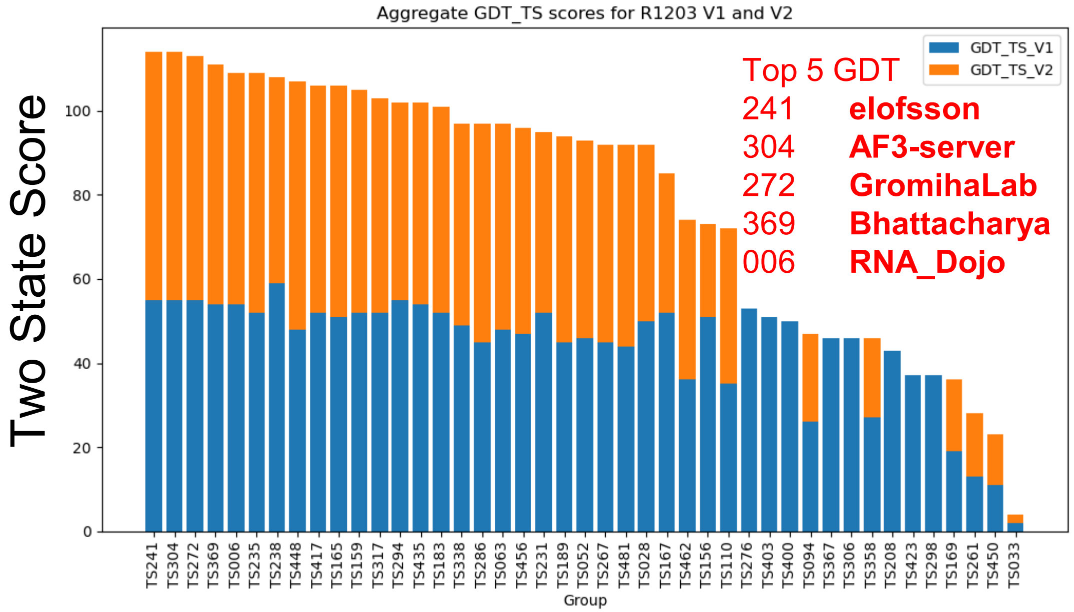


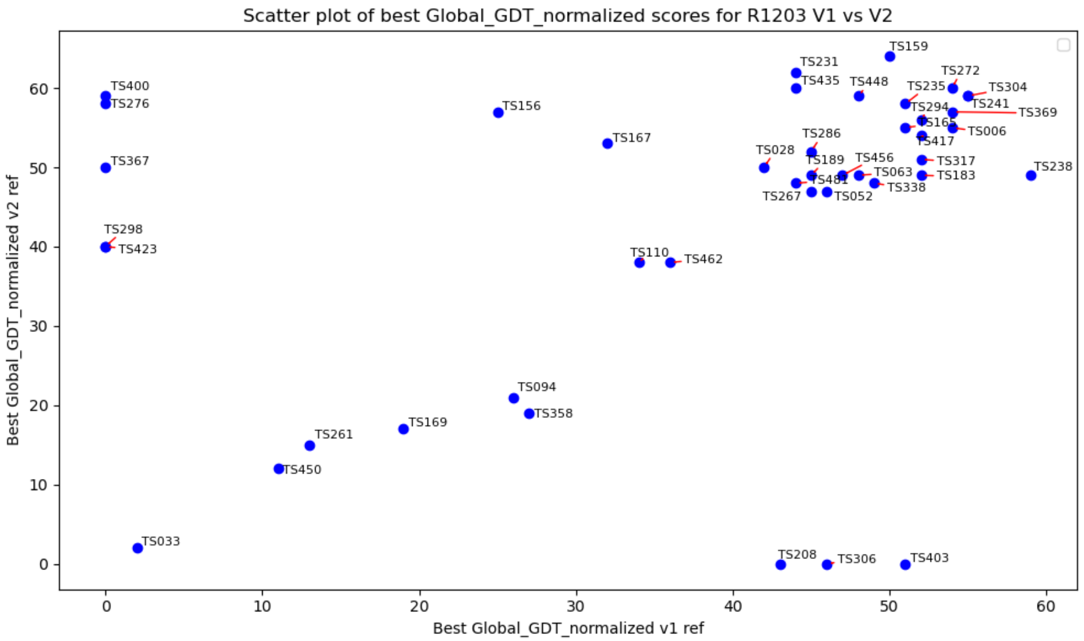
Tipo et al Nat Comm 15: 4198 (2024)

Best Global GDT – by Group



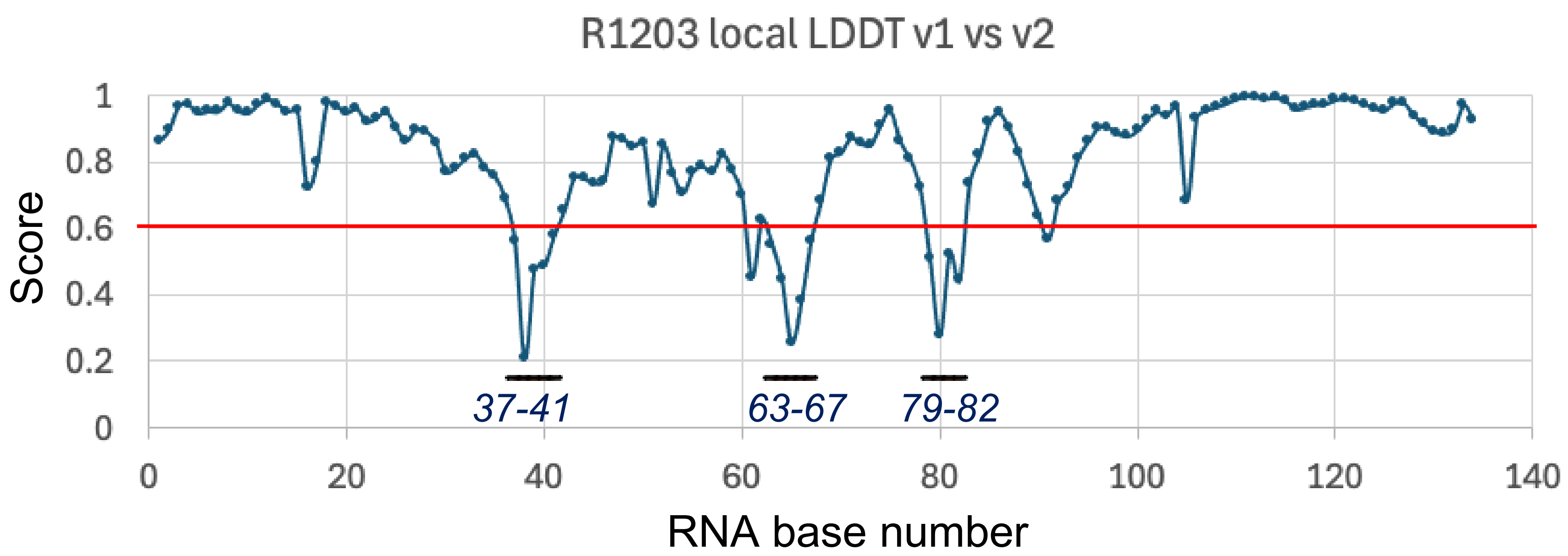




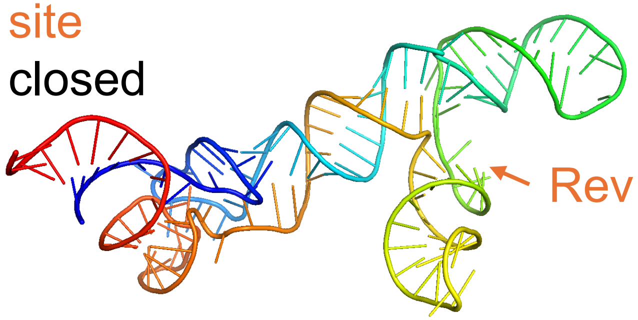


HIV-1 Rev Response Element (RRE) Stem-Loop II (SLII) open and closed conformations

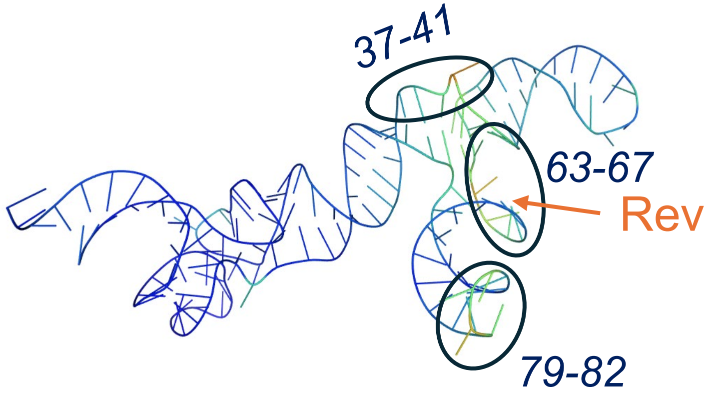
from: https://swissmodel.expasy.org/assess



Rev-binding site closed



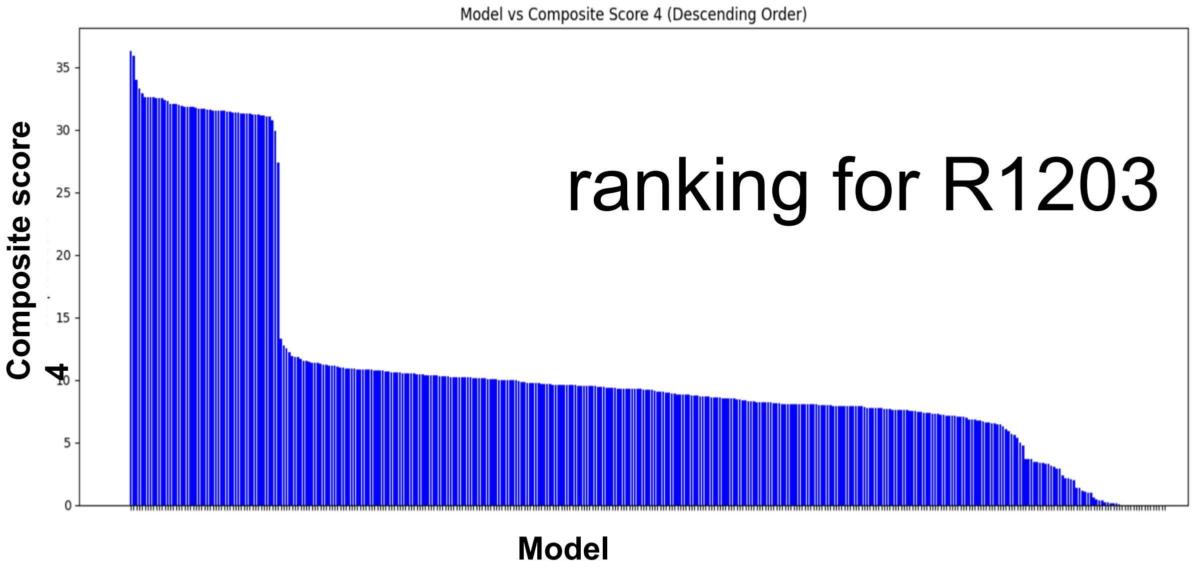
major groove open

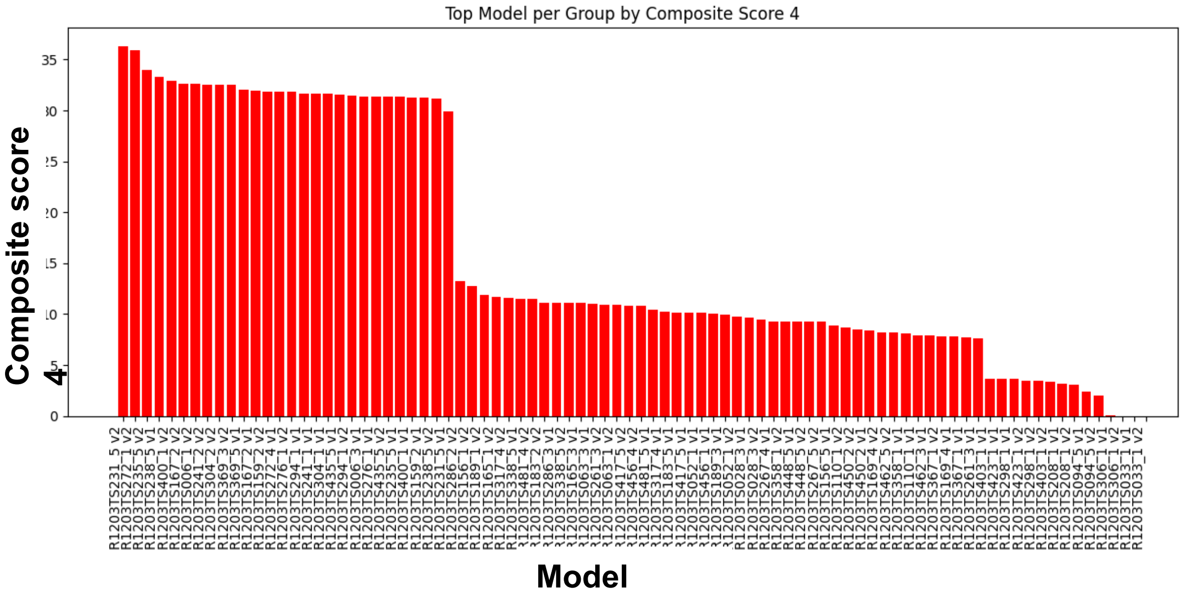


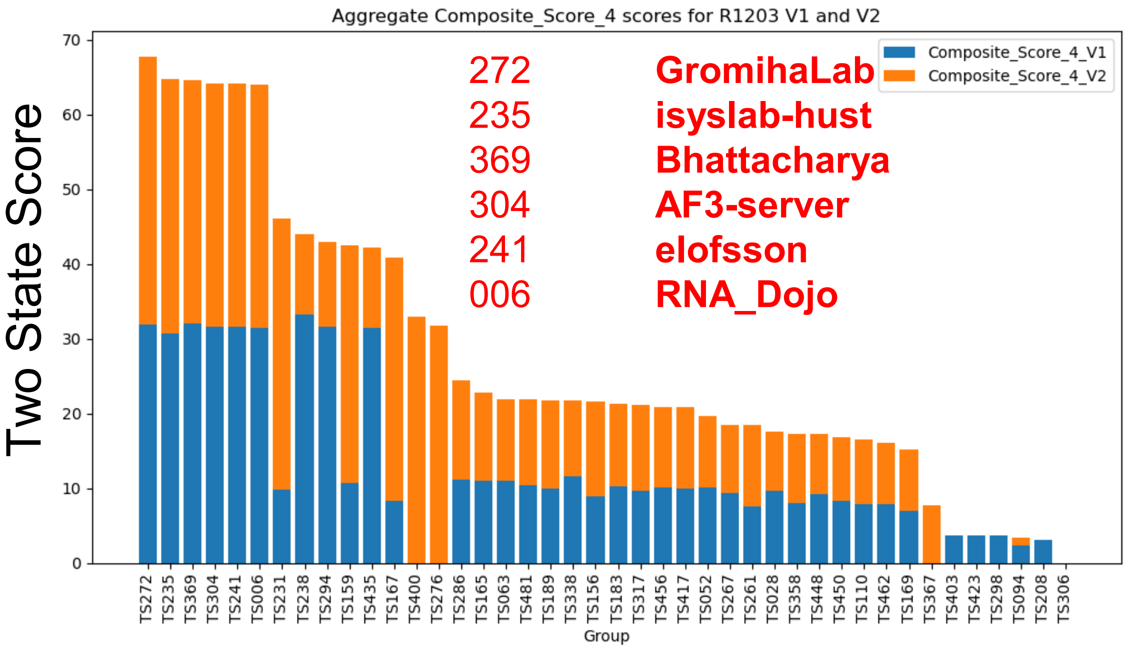
colored blue to red for bases 1 - 134

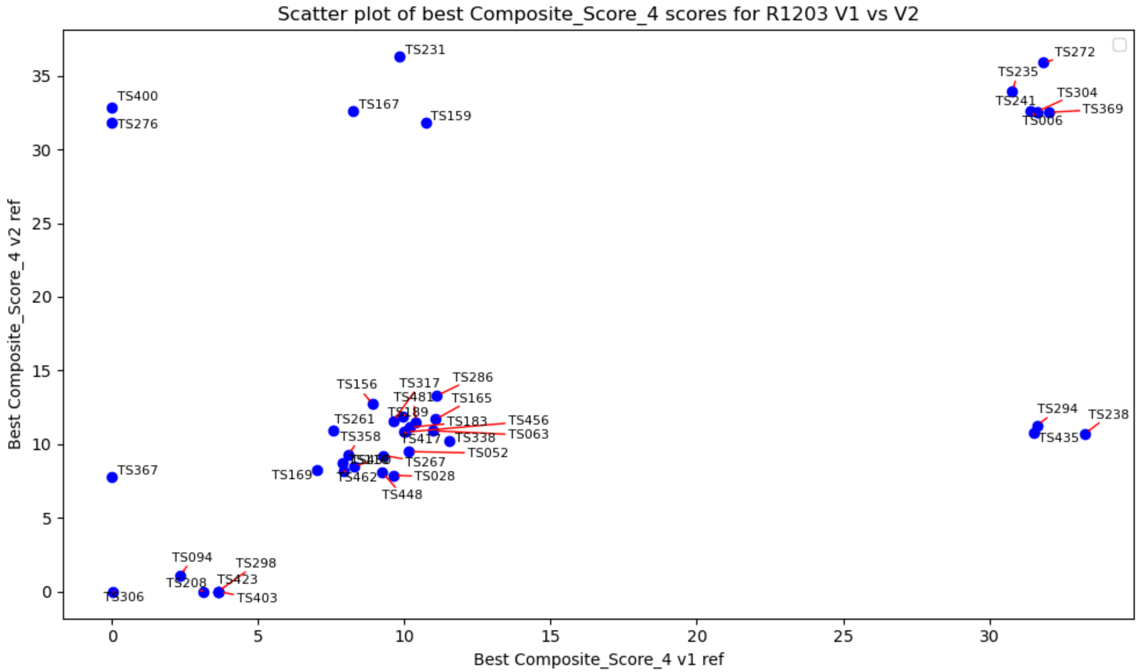
lLDDT colored onto closed form

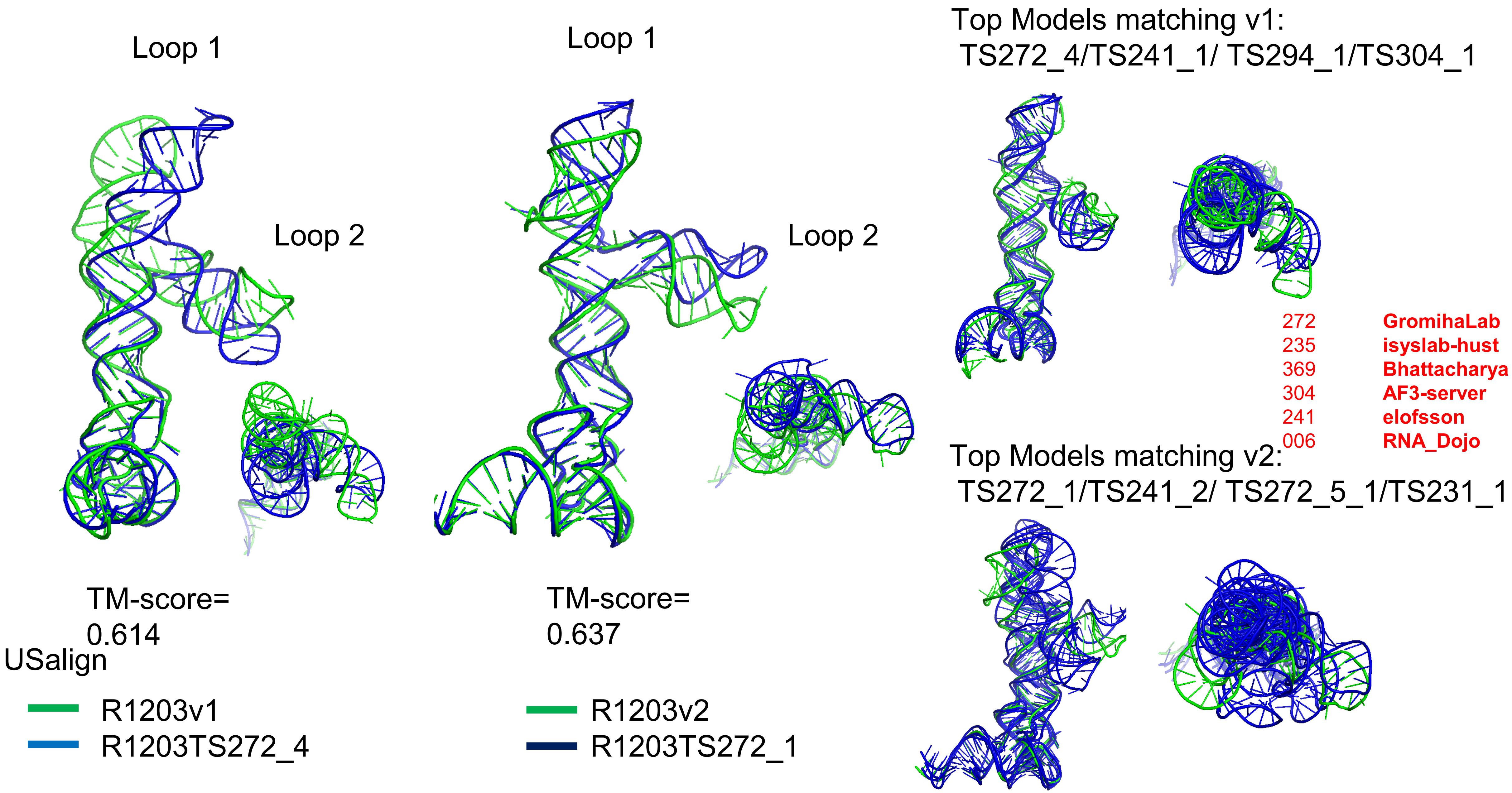
Best Composite Global-Local Score – by Group



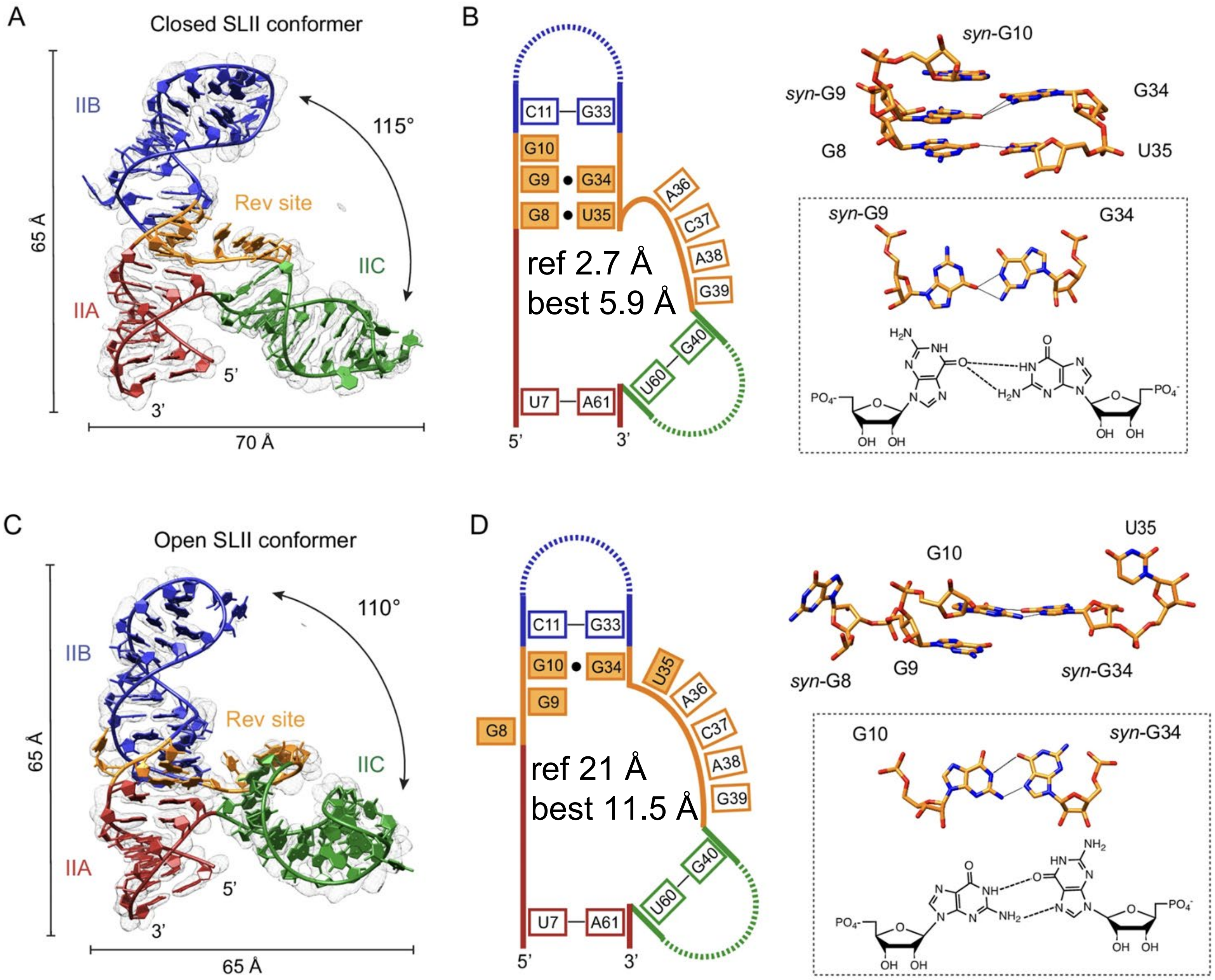








Loop 1
Loop 2
Loop 1
Loop 2
Top Models matching v1:
TS272_4/TS241_1/ TS294_1/TS304_1
272 GromihaLab
235 isyslab-hust
369 Bhattacharya
304 AF3-server
241 elofsson
006 RNA_Dojo
Top Models matching v2:
TS272_1/TS241_2/ TS272_5_1/TS231_1
TM-score=
0.614
TM-score=
0.637
USalign
R1203v1
R1203TS272_4
R1203v2
R1203TS272_1



Tipo et al Nat Comm 15: 4198 (2024)

Conclusions

Many predictors delivered excellent models of holo state for membrane porin T1214 – even when modeled without ligand bound

The alternative states of M1228 and M1239 protein-DNA complexes were identified by some groups – but with low accuracy

Some predictors can identify “V-shaped” vs “compact” states of T1228 and T1239. However, the GDT-TS scores for each of these states are only 50 – 60.

AF3 Server had best performance for modeling alternative conformational states of trimeric target T1249

Several groups delivered good models of open/closed states of HIV-1 Rev Response Element (RRE) Stem-Loop II (SLII) R1203. Mostly AF3-based methods

Future CASP experiments should better distinguish multiple conformations accessible for the apo protein from alternative conformations induced by partner binding.