

CAMEO - Beyond Single Protein Chains

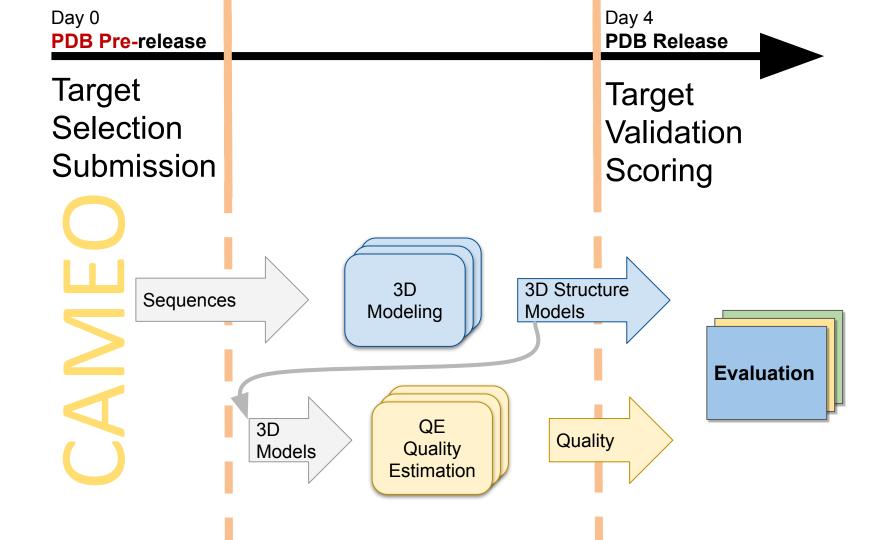
Xavier Robin

CASP14
December 2020









CASP CAMEO

Blind predictions

Blind predictions

CASP CAMEO

- Blind predictions
- Every 2 years

- Blind predictions
- Weekly

CASP CAMEO

- Blind predictions
- Every 2 years
- Human curation

- Blind predictions
- Weekly
- Fully automated

CASP CAMEO

- Blind predictions
- Every 2 years
- Human curation
- Human & Server predictors

- Blind predictions
- Weekly
- Fully automated
- Only servers

CASP

- Blind predictions
- Every 2 years
- Human curation
- Human & Server predictors
- Hand-picked targets
 - Obtained from crystallographers
 - Hard folding targets

CAMEO

- Blind predictions
- Weekly
- Fully automated
- Only servers
- Targets from PDB pre-release
 - Selection based on BLAST
 - Typically easier targets

CASP

- Blind predictions
- Every 2 years
- Human curation
- Human & Server predictors
- Hand-picked targets
 - Obtained from crystallographers
 - Hard folding targets
- Experts assessment

CAMEO

- Blind predictions
- Weekly
- Fully automated
- Only servers
- Targets from PDB pre-release
 - Remove sequence redundancy
 - Typically easier targets
- Automated evaluation

Major news of the past 2 years

- Refactoring & workflow updates
- Discontinued the Contact Prediction category (CP)
- Jürgen Haas moved to AstraZeneca
- New CAMEO 3D beta version

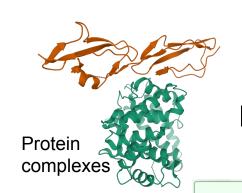


Jürgen Haas

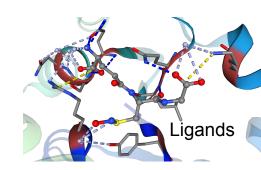


Rafal Gumienny

CAMEO - Beyond Single Protein Chains

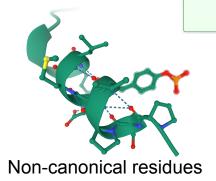


Introducing the new 3D category:



3D - Structures & Complexes

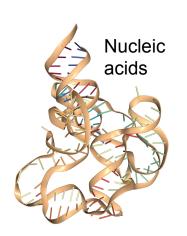
Registrations now open!



https://beta.cameo3d.org/



help-cameo3d@unibas.ch





Summary of the past 2 years

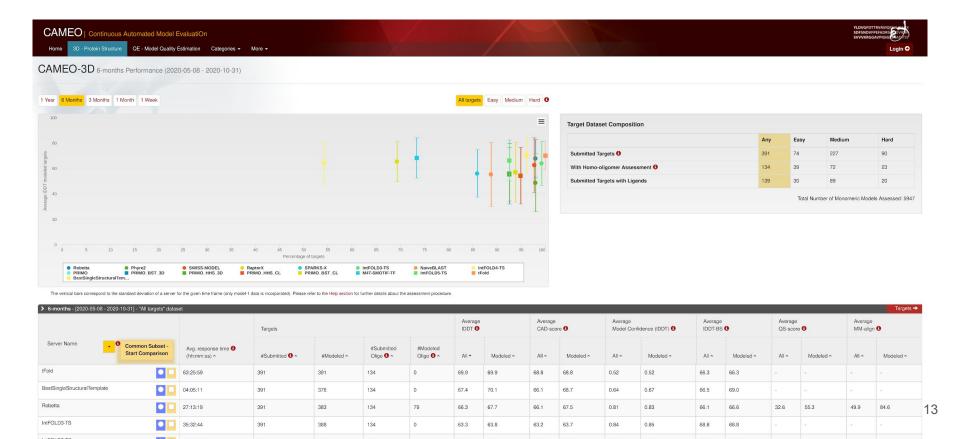
3D

- 1795 targets
 - o 117 513 predictions
- 15 active public servers
 - 25 development servers
- 9 (18) groups

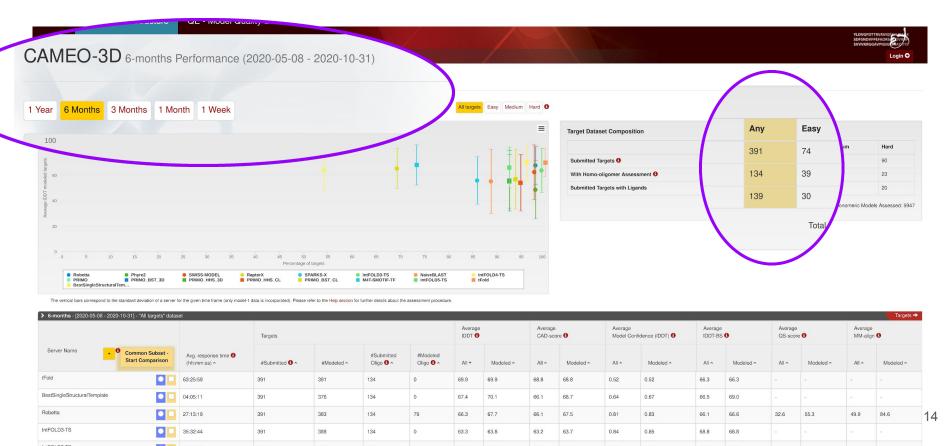
QE

- 19 159 targets
 - o 244 992 predictions
- 12 active public servers
 - 5 development servers
- 5 (6) groups

Results - Overview

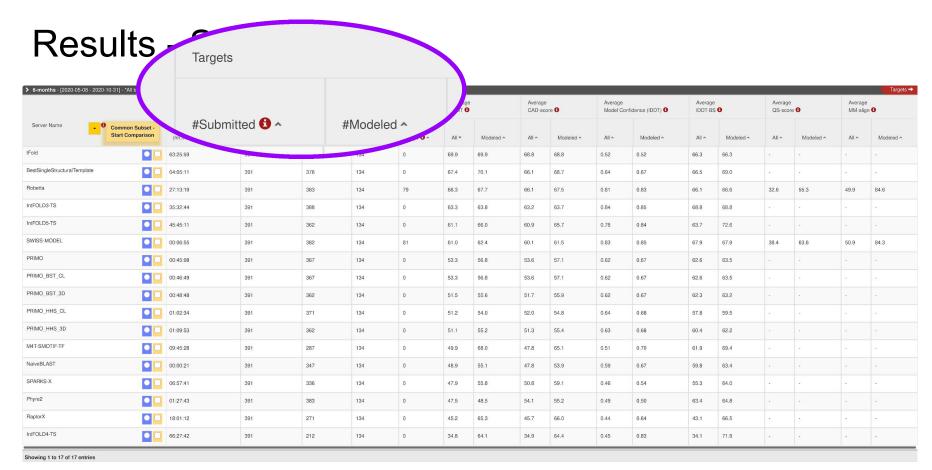


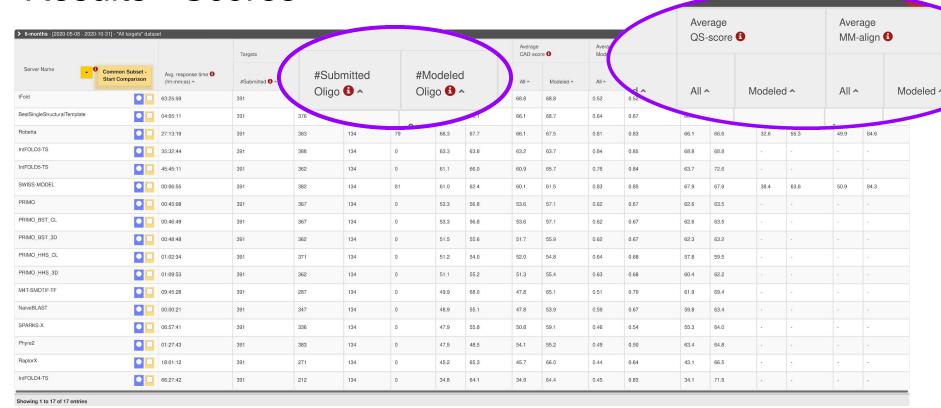
Results - Overview

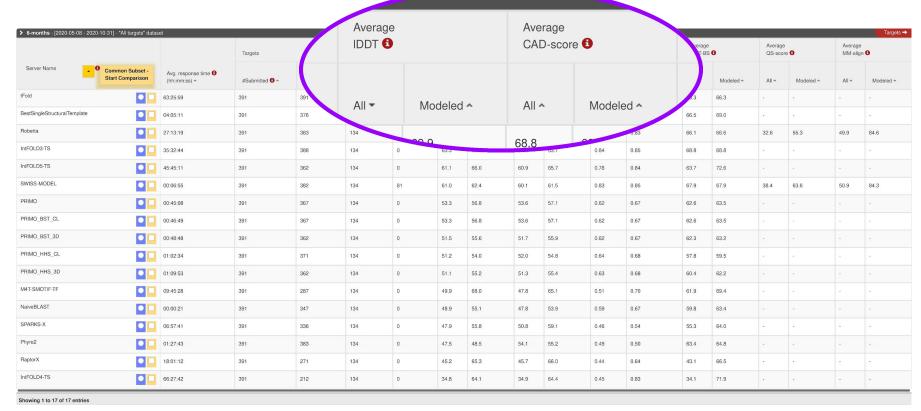


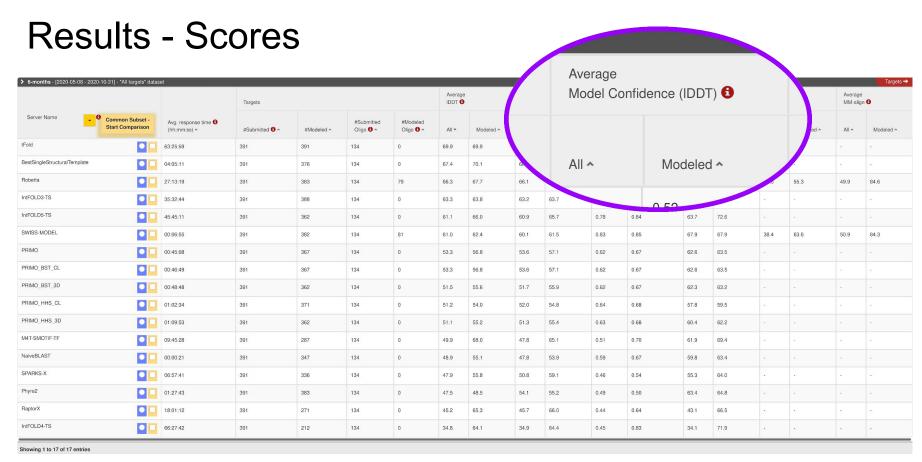
			Targets			Average IDDT 6			Average CAD-score 3		Average Model Confidence (IDDT) 6		Average IDDT-BS 6		Average QS-score 6		Average MM-align 6	
Server Name	Common Subset - Start Comparison	Avg. response time 6 (hh:mm:ss) ^	#Submitted 3 ^	#Modeled ^	#Submitted Oligo 6 ^	#Modeled Oligo 6 ^	All ▼	Modeled ^	All ^	Modeled ^	All ^	Modeled ^	All ^	Modeled ^	All ^	Modeled ^	All ^	Modeled ^
Fold		63:25:59	391	391	134	0	69.9	69.9	68.8	68.8	0.52	0.52	66.3	66.3				
estSingleStructuralTemplate		04:05:11	391	376	134	0	67.4	70.1	66.1	68.7	0.64	0.67	66.5	69.0				
lobetta		27:13:19	391	383	134	79	66.3	67.7	66.1	67.5	0.81	0.83	66.1	66.6	32.6	55.3	49.9	84.6
ntFOLD3-TS		35:32:44	391	388	134	0	63.3	63.8	63.2	63.7	0.84	0.85	68.8	68.8			-	
tFOLD5-TS		45:45:11	391	362	134	0	61.1	66.0	60.9	65.7	0.78	0.84	63.7	72.6	-	2	-	-
WISS-MODEL		00:06:55	391	382	134	81	61.0	62.4	60.1	61.5	0.83	0.85	67.9	67.9	38.4	63.6	50.9	84.3
RIMO		00:45:08	391	367	134	0	53.3	56.8	53.6	57.1	0.62	0.67	62.6	63.5				
RIMO_BST_CL		00:46:49	391	367	134	0	53.3	56.8	53.6	57.1	0.62	0.67	62.6	63.5	-			(2)
RIMO_BST_3D	•	00:48:48	391	362	134	0	51.5	55.6	51.7	55.9	0.62	0.67	62.3	63.2	-			
RIMO_HHS_CL		01:02:34	391	371	134	0	51.2	54.0	52.0	54.8	0.64	0.68	57.8	59.5				
RIMO_HHS_3D		01:09:53	391	362	134	0	51.1	55.2	51.3	55.4	0.63	0.68	60.4	62.2				
14T-SMOTIF-TF		09:45:28	391	287	134	0	49.9	68.0	47.8	65.1	0.51	0.70	61.9	69.4	-			
aiveBLAST		00:00:21	391	347	134	0	48.9	55.1	47.8	53.9	0.59	0.67	59.8	63.4			-	
PARKS-X	• •	06:57:41	391	336	134	0	47.9	55.8	50.8	59.1	0.46	0.54	55.3	64.0	-			
nyre2		01:27:43	391	383	134	0	47.5	48.5	54.1	55.2	0.49	0.50	63.4	64.8	-	9	-	-
aptorX	• •	18:01:12	391	271	134	0	45.2	65.3	45.7	66.0	0.44	0.64	43.1	66.5			-	
FOLD4-TS	00	66:27:42	391	212	134	0	34.8	64.1	34.9	64.4	0.45	0.83	34.1	71.9	-			-

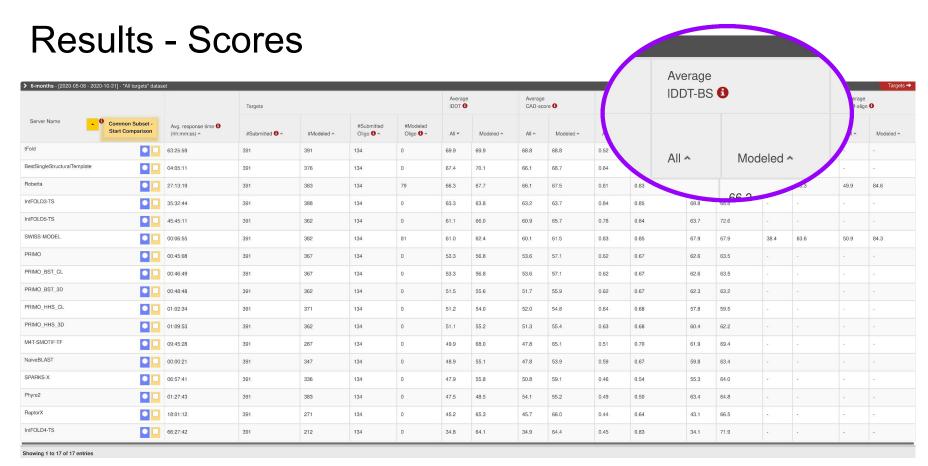
						Average IDDT 6		Averag CAD-so		Average Model Co	onfidence (IDDT) 6	Average IDDT-B		Average QS-scor		Average MM-alig	
Server Name	Avg.	#Modeled Oligo 6 ^	All ▼	Modeled ^	All ^	Modeled ^	All ^	Modeled ~	All ^	Modeled ^	All ^	Modeled ^	All ^	Modeled			
old				J	0	69.9	69.9	68.8	68.8	0.52	0.52	66.3	66.3				-
stSingleStructuralTemplate	(1111:11	(hh:mm:ss) ^						66.1	68.7	0.64	0.67	66.5	69.0				
petta				134	79	66.3	67.7	66.1	67.5	0.81	0.83	66.1	66.6	32.6	55.3	49.9	84.6
OLD3-TS	35:32:44	391	388	134	0	63.3	63.8	63.2	63.7	0.84	0.85	68.8	68.8			-	
OLD5-TS	45:45:11	391	362	134	0	61.1	66.0	60.9	65.7	0.78	0.84	63.7	72.6	-	-		-
SS-MODEL	00:06:55	391	382	134	81	61.0	62.4	60.1	61.5	0.83	0.85	67.9	67.9	38.4	63.6	50.9	84.3
MO	00:45:08	391	367	134	0	53.3	56.8	53.6	57.1	0.62	0.67	62.6	63.5				
MO_BST_CL	00:46:49	391	367	134	0	53.3	56.8	53.6	57.1	0.62	0.67	62.6	63.5				
MO_BST_3D	00:48:48	391	362	134	0	51.5	55.6	51.7	55.9	0.62	0.67	62.3	63.2		-	-	
MO_HHS_CL	01:02:34	391	371	134	0	51.2	54.0	52.0	54.8	0.64	0.68	57.8	59.5				
MO_HHS_3D	01:09:53	391	362	134	0	51.1	55.2	51.3	55.4	0.63	0.68	60.4	62.2				
r-SMOTIF-TF	09:45:28	391	287	134	0	49.9	68.0	47.8	65.1	0.51	0.70	61.9	69.4				-
veBLAST	00:00:21	391	347	134	0	48.9	55.1	47.8	53.9	0.59	0.67	59.8	63.4				
ARKS-X	06:57:41	391	336	134	0	47.9	55.8	50.8	59.1	0.46	0.54	55.3	64.0		2		121
e2	01:27:43	391	383	134	0	47.5	48.5	54.1	55.2	0.49	0.50	63.4	64.8	-	-		-
torX	18:01:12	391	271	134	0	45.2	65.3	45.7	66.0	0.44	0.64	43.1	66.5			-	
OLD4-TS	66:27:42	391	212	134	0	34.8	64.1	34.9	64.4	0.45	0.83	34.1	71.9		1.	1.	1.

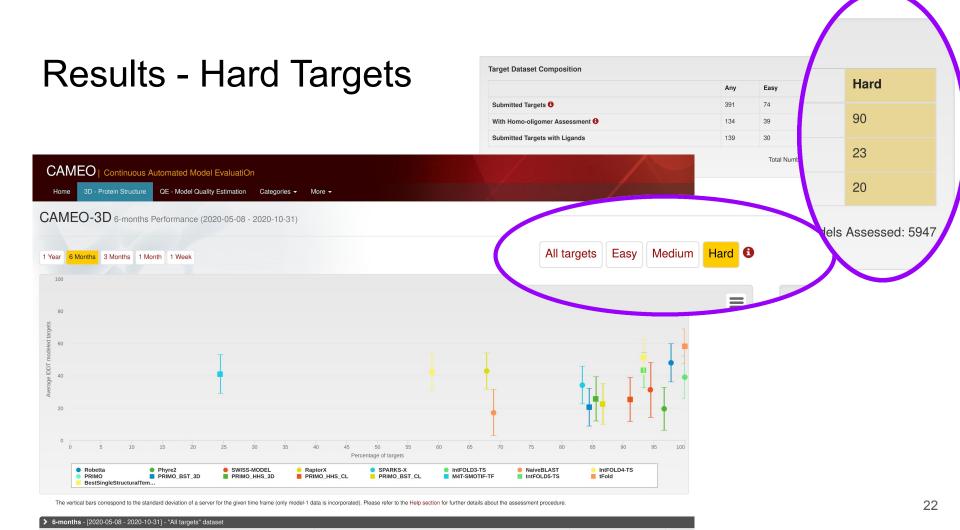


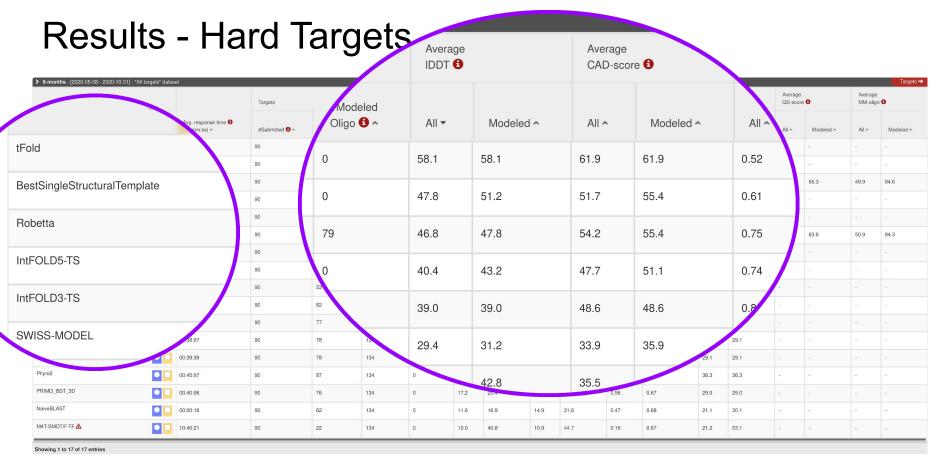










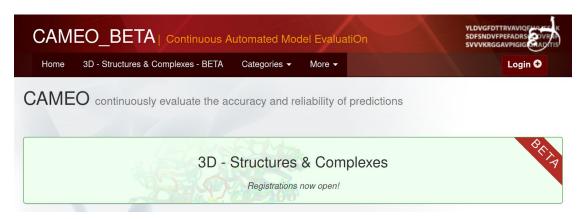


Modeling of single protein chains...

What's the next frontier?

3D Structures & Complexes

- Provides full complexes as prediction targets to servers that support it:
 - One or more protein sequences
 - Ligands as InChl or PDB component ID (3 letter codes)
 - Canonical or modified sequence
 - o DNA/RNA



Complexes

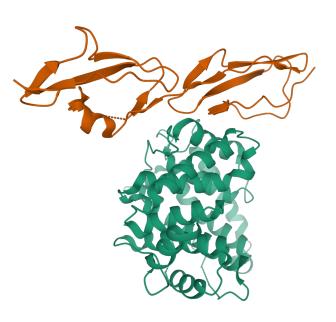
- Homo-oligomers already scored in CAMEO
 - o QS-score & oligomeric IDDT



Se	erver Evaluations o	All scores	es Superposition dependent			sition independ	dent Bind	ng site scores	Oligomeric scores				
Sh	now 10 ✔ entries										F	ilter:	
	Server Name	Predictions 6	Resp. time (hh:mm:ss)	From	To ^	Cov. (%)	QS-score	Oligo details	QS-score Best Match 🕏 ^	IDDT Oligo	MM-align TM- score 3 ^	MM-align RMSD 🕙 ^	QScore
	SWISS-MODEL	Oligo 1 🖺 🖾	00:21:28	-	÷	-	0.83	Details	0.83	61.02	0.92	2.60	-
	Robetta 🗹	Oligo 1 🖺 🛂	23:39:34	-	•	-	0.61	Details	0.64	57.36	0.93	2.64	-

Complexes

- Homo-oligomers already scored in CAMEO
- Hetero-oligomers
 - Submitted as multiple sequences
- Target selection:
 - Difficulty of the members + novelty of the complex
- Scoring
 - Oligo-IDDT
 - QS-score
 - o Model Confidence?
 - CAPRI
 - Collaboration with Elixir 3D-BioInfo





Ligands

- PDB "non-polymer" entities
- Submitted as InChl or 3 letter codes
- Target Selection
 - Easy protein modeling targets?
- Scoring
 - o RMSD
 - Collaboration with Elixir 3D-BioInfo

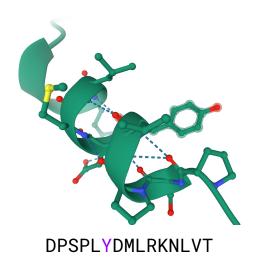


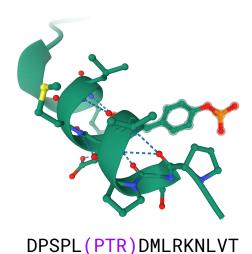
Haas et al. (2018). Proteins. 86, 387-398.



Side-chains / modified residues

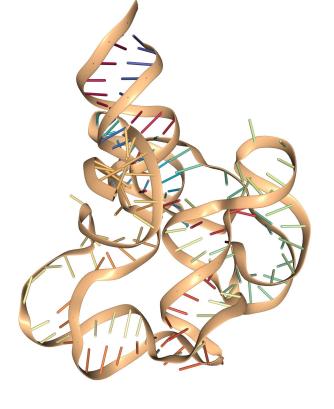
- Modified / non natural residues
- Submitting non-canonical sequences
- All atom scores (IDDT, CAD-score)





DNA/RNA

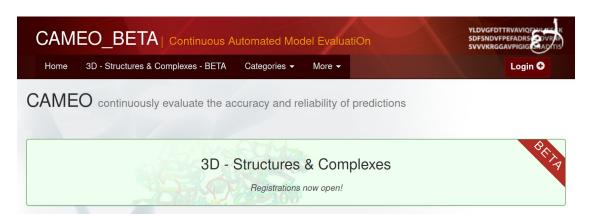
- Nucleic acids are part of the PDB pre-release
- Can already be submitted to servers
- Target selection?
- Scoring
 - o CAD score?
- Elixir 3D-BioInfo Community





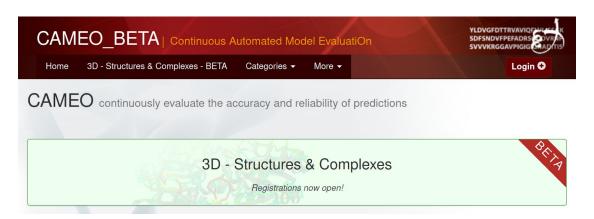
3D Structures & Complexes

- Will replace current 3D modeling category
- Opt-in
 - Servers can select what they can do



Aspects of modeling

- Predict correct oligomeric state from the sequence
- Model the full complex
- Return predictions in mmCIF or PDB format



Work in progress

- Selection of relevant targets through flexible target sets
 - Free modeling with few single chain very hard targets
 - Template-based, more, easier targets
 - DNA & RNA
 - Ligand modeling: high sequence identity with new ligands
 - Variants modeling: high sequence identity
 - Core subset of targets
- Opt-in
- Selection of relevant scores
- Handling mmCIF predictions

Community White Paper

- Reflect on achievements so far
 - Looking for stories and input from participants
- Outline future developments
- Input from the community on future directions

Recommendations

- Servers are useful for:
 - Access to everyone
 - Reproducibility
 - Open Science
- Benefits for server developers:
 - Continuous benchmarking of server performance
 - Many targets to collect statistics
 - Development servers anonymized to benchmark new features



Acknowledgements

CAMEO contributors: Juergen Haas, Rafal Gumienny & many others

Torsten Schwede group

PDB

CAMEO participants

A. Sali, L. McGuffin, T. Schwede, J. Soeding, D. Baker, A. Fiser, M. Sternberg, Y. Zhang, C. Floudas, S. Tosatto, J. Xu, Y. Zhou, O. Brock, B. Wallner, A. Eloffson, D. Labudde, C. Venclovas, J. Cheng, O. Taştan Bishop, Y. An-Suei, T. Sosnick, C. Kaesar, P. Winn, C. Seok., S. Wang

https://beta.cameo3d.org/

help-cameo3d@unibas.ch











