

An overview of the web infrastructure for the 2019 Model Metrics Challenge

[this document is available by clicking on the HELP link in any of the Results pages]

Contents

General information	4
Entry webpage for Results of all cryo-EM model Challenges.....	4
Overall scheme of the 2019 Metrics Challenge	5
List of software tools used in different evaluation tracks	6
Targets /reference structures / EM maps.....	7
Models.....	8
Predictors.....	9
Hierarchical organization of the results	9
Switching between targets.....	10
Sorting tables.....	10
Filtering models.....	11
Brief descriptions of scores.....	11
Downloading results as text files	11
Local accuracy plot: zoom in to region of interest.....	12
Evaluation based on model coordinates only	13
Overview	13
Measures.....	13
PHENIX (phenix.model_statistics) ¹	13
Molprobit scores ²	13
CaBLAM scores ³	13
ProQ3 (a-priori model accuracy assessment)	14
<i>Model coordinates only</i> web infrastructure.....	14
model coordinates only → Geometry Scores	14
model coordinates only → Atomic displacement parameters (ADP)	15

model coordinates only → Accuracy Estimate	16
Evaluation versus EM maps	17
Overview	17
Measures.....	17
PHENIX suite of scores for analysis of cryo-EM atomic models ²²	17
TEMPy scores ⁵⁻⁸	17
EMRinger score ⁹	18
Q- score ¹⁰	18
Atom Inclusion scores ¹¹	18
<i>Fit to EM map</i> web infrastructure	19
fit to EM map → Global Accuracy → Scores.....	19
fit to EM map → Global Accuracy → Plots → FSC versus Resolution.....	20
fit to EM map → Global Accuracy → Plots → Atom Inclusion.....	21
fit to EM map → Local Accuracy → Per-chain Summary.....	22
fit to EM map → Local Accuracy → Tempy /Phenix /EMRinger /Q-score /Atom Inclusion	23
fit to EM map → Local Accuracy → All Scores.....	26
Evaluation versus reference structure (monomeric mode).....	27
Overview	27
Measures.....	27
Superposition-based LGA family of scores (GDT_TS ^{13,23} , GDT_HA, GDC_SC ²⁴).....	27
Superposition-independent structure-based family of scores (LDDT, CADaa).....	27
Guide to web infrastructure.....	29
vs reference structure → Monomers → Global Accuracy → Scores.....	29
vs reference structure → Monomers → Global Accuracy → Plots	30
vs reference structure → Monomers → Local Accuracy → LGA /LDDT	31
Evaluation versus reference structure (multimeric mode)	32
Overview	32
Measures.....	32
QS scores	32
Superposition-based LGA family of scores (GDT_TS ^{13,23} , GDT_HA, GDC_SC ²⁴ , GDC_ALL).....	33

phenix.chain_comparison module	33
Hydrogen bonds scores.....	33
Guide to web infrastructure.....	34
vs reference structure → Multimers → QS score /LGA /chain_comparison /Hbonds	34
Comparison versus other models	35
Overview	35
Measures.....	35
Guide to web infrastructure.....	35
Comparative analyses.....	36
Overview	36
Guide to web infrastructure.....	36
Comparative analyses → Models Pair-wise Comparison	36
Comparative analyses → Scores Pair-wise Comparison → Scatter Plots.....	38
Comparative analyses → Scores Pair-wise Comparison → Correlation Summary	39
Comparative analyses → Model ranks (per target).....	41
Comparative analyses → Group ranks (across targets).....	42

General information

Entry webpage for Results of all cryo-EM model Challenges

Evaluation results for all cryo-EM model challenges are available from the EMDataResource website <https://model-compare.emdataresource.org>:

The screenshot shows the website <https://model-compare.emdataresource.org>. The header features the "EM Model Challenge" logo and social media icons for Facebook and RSS. Below the header, there are two main sections of challenge entries. The top section is for the year 2019, with the year "2019" centered in large blue text. It contains four challenge entries, each with a target image, a challenge ID, and a description: T0020 (Human Apoptosis), T0022 (Human Apoptosis), T0023 (Human Apoptosis), and T0024 (Human Liver Acetate). The bottom section is for the year 2016, with the year "2016" centered in large blue text. It contains a grid of challenge entries, including T0001 (Tobacco Mosaic Virus), T0002 (T205 Proteasome), T0003 (T205 Proteasome), T0004 (TSPV1 Channel), T0005 (Bovine Myxoma Virus), T0006 (β-Galactosidase), T0007 (α-Secretase), T0008 (T08 Ribosome), and T0009 (T08 Ribosome).

Results of the latest challenge are accessible by clicking on the '2019' block of target images.

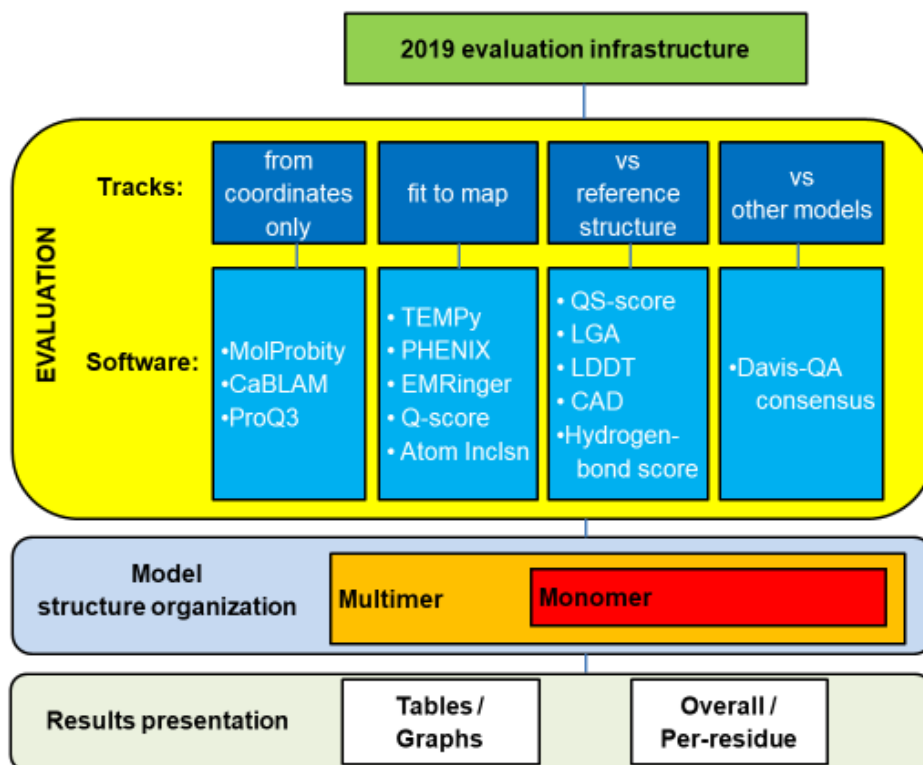
Overall scheme of the 2019 Metrics Challenge

The evaluation system for 2019 Model Metrics Challenge is based on the 2016/2017 Model Challenge system (Kryshtafovych A, Adams PD, Lawson CL, Chiu W. (2018). *Evaluation system and web infrastructure for the second cryo-EM model challenge. J Struct Biol.* 204, 96-108. doi: 10.1016/j.jsb.2018.07.006. PMID:30017700). The 2019 system was significantly redesigned with inclusion of new evaluation measures and analysis tools. The overall organization of the redesigned system is described in this Help document.

Submitted models are evaluated in 4 different tracks:

- exclusively from coordinates (stereochemistry and energy-based),
- evaluating model-to-map fit,
- comparing to reference structure(s),
- checking agreement with other submitted models.

Each track uses its own software tools, and evaluation results are grouped accordingly in 4 classes. A general evaluation scheme is presented below:



List of software tools used in different evaluation tracks

Exclusively from coordinates

- PHENIX ¹ (phenix.model_statistics),
- Molprobity ² (phenix.molprobity, phenix.cbetadev),
- CaBLAM ³ (phenix.cablam),
- ProQ3 ⁴ (machine learning energy and geometry-based single-model accuracy assessment method).

Model-to-map fit

- PHENIX ¹ (phenix.map_model_cc),
- TEMPy ⁵⁻⁸ (global and local model-map fit),
- EMRinger ⁹ (global and local model-map fit based on side-chain fit),
- Q-score ¹⁰,
- Atom Inclusion score (reimplemented from EMDB Visual Analysis pages ¹¹, e.g., <https://www.ebi.ac.uk/pdbe/entry/emdb/EMD-10101/analysis> – thanks to Zhe Wang and Ardan Patwardhan from EBI).

Comparing to reference structure(s)

- QS-score ¹² (multimers),
- LGA ¹³ (used for generating GDT-family based scores),
- LDDT ¹⁴ (superposition-free measure; compares difference in distance patterns),
- CAD score ¹⁵ (superposition-free measure; compares difference in contact areas),
- Hydrogen-bond score (hydrogen bonds identified with HBPLUS ¹⁶).

Agreement between the models

- Davis_QAconsensus ¹⁷.

Targets /reference structures / EM maps

Results for each target can be visited by clicking on the **target pictograph /name**

Targets in the 2019 edition of the challenge are numbered consecutively, from T0101 to T0104 (targets in the previous challenge started from T00...). In the results tables, biological target name is provided together with the target ID.

The extended target ID includes PDB ID of the reference structure and the name of the representative chain used in the evaluations. Several reference structures can be used for the same target. Results for the reference structures are highlighted in grey in all results tables.

Map EMDB ID is shown in a separate column of the target-specific evaluation results. Unlike the previous challenge, results for different maps of the same protein are organized as different targets, e.g. Apoferritin comprises three targets (T0101-T0103) corresponding to maps at different resolution.



https://model-compare.emdataresource.org/2019/cgi-bin/em_results_summary.cgi

EM Model Challenge - 2019

Results
data repositiorium

Target ID	Target Name	Map ID
T0101	Human Apoferritin	emd_20026_1
T0102	Human Apoferritin	emd_20027_2
T0103	Human Apoferritin	emd_20028_2
T0104	Horse Liver Alcohol Dehydrogenase	emd_0406

Models

All models in the 2019 Challenge were submitted through the Challenges EMDataResource gateway at Rutgers. Information on the submissions with full metadata is available from:

https://challenges.emdataresource.org/sites/default/files/model_metrics_challenge_metadata.xlsx

Model summary statistics can be found here:

<https://challenges.emdataresource.org/sites/default/files/Submissions-Analysis.html>

Files used in the evaluation (models, targets, maps) and results (plain text files as generated by evaluation programs) are available by following the **'data repositiorium' link**

<https://model-compare.emdataresource.org/data/2019>)

from the main Results page:

Results
data repositiorium

Model ID	Target	Map ID
T0101	Human Apoferritin	emd_20026_1
T0102	Human Apoferritin	emd_20027_2
T0103	Human Apoferritin	emd_20028_2
T0104	Horse Liver Alcohol Dehydrogenase	emd_0406

The Model ids used in the Results tables (e.g. T0104EM010_1) are formed according to the following scheme:

- T0104 [target name]
- EM [electron microscopy]
- 010 [predictor group number (see below)]
- _1 [model number 1 from this predictor for this target]

Predictors

Each group participating in the EM Model Challenge is assigned a unique number. Predictor IDs corresponding to each model are encoded in the model name (see above). Before the June 2019 Face-to-Face Model Metrics Workshop, the group_ID – group_name correspondence was concealed from everyone but the organizers. The group names were revealed at the meeting and are now shown in the Group Name column on all Results pages.

Hierarchical organization of the results

As described above, models are evaluated in 4 different tracks: exclusively from coordinates, model-to-map fit, versus reference structure(s), and agreement between models. Evaluation results in each of the tracks are reported separately under **four different tabs** provided at the highest level of the Results page hierarchy:

EMDataResource
Sponsored by the US National Institute of General Medical Sciences (NIH/NIGMS)
Please address any questions or queries to: help@emddataresource.org
© 2015-2019

Evaluation data are available as sortable score tables or interactive (clickable) graphs.

Additionally, a **Comparative Analyses** Tab is available to the right of the four main tabs. Under this tab, a user can view summaries of relative accuracies of models, compare scores of models in different evaluation tracks or check correlations between pairs of scores.

Switching between targets

Switching to results for a different target can be done by returning to the main Challenge page and clicking on the desired target pictograph (see the *Models* section above) or by accessing the “Target” drop-down menu on any of the target-specific pages.

EM Model Challenge - 2019

Results

model coordinates only | fit to EM map | vs reference structure | vs other models | Comparative Analyses

Global Accuracy | Local Accuracy

Scores | Plots

Target: Horse Liver Alcohol Dehydrogenase: T0104 (emd_0406)

Filter by method: ab-initio optimized cryoEM model fully automated partially automated, with some manual steps

#	Model	Group Name	Method	EM Map	TempPy										Phenix				Q-score		Atom Inclusion										
					ab-initio/optimized	auto/man.	MapID	Res. ol.	CCC	CCC (ov)	LAP	MI (ov)	MI (not)	SMOC	box CC	CC(mask)	CC(vol)	CC(peaks)	Resol(FSC=0.5)	EMRinger score	Q-score	All	BB								
1	T0104EM035_1	phenix	ab-initio	man.	emd_0406	2.9	0.321	0.491	0.197	0.985	0.985	0.038	0.240	0.741	0.741	0.640	0.778	0.847	0.848	0.846	0.840	0.558	0.763	2.985	3.012	5.151	0.652	0.652	0.650	0.650	0.765

Sorting tables

Results tables can be sorted on any column by clicking on the column name (the column header will be highlighted). A repeat click will reorder the data in the reverse order.

EM Model Challenge - 2019

Results

model coordinates only | fit to EM map | vs reference structure | vs other models | Comparative Analyses

Geometry Scores | ADPs Histogram | Accuracy Estimate

Target: Human Apoferritin: T0101 (emd_20026_1)

Filter by method: ab-initio optimized cryoEM model fully automated partially automated, with some manual steps

#	Model	Gr.Name	Method	EM Map	Deviation from ideal model										Planarity				Dihedral				Molprobity				Cblam											
					ab-initio/optimized	auto/man.	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count		
1	T0101EM054_2	kihara	ab-initio	man.	0.020	0.142	1442	1.509	15.032	1943	0.096	0.318	204	0.008	0.027	257	9.347	77.749	871	0.000	0.000	0.590	96.470	1.20	4.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Filtering models

Results for all models, or only the models built using specific modeling techniques

ab initio / optimization

or

automatic / manual

can be displayed by marking the appropriate checkboxes and clicking on the 'filter' button.

The screenshot shows the Phenix interface with the following elements:

- Buttons: model coordinates only, fit to EM map, vs reference structure, vs other models, Comparative Analyses, Global Accuracy, Local Accuracy, Scores, Plots, help.
- Target: Horse Liver Alcohol Dehydrogenase (PDB: 4OJ6)
- Filter by method: ab-initio, optimized cryoEM model, fully automated, partially automated, with some manual steps. A 'filter' button is located to the right of these checkboxes.
- Table of results with columns: #, Model, Group Name, Method, EM Map, Res. ol., CCC, CCC (ov), LAP, TempPy (ENV, MI, MI (ov)), SMOCC (Orig., noh), box CC (Orig., noh), CC(mask) (Orig., BF=0), Phenix (CC(vol), CC(peaks), Resol.(FSC=0.5), EMRinger score), Q-score (Orig., noh), Atom Inclusion (A, noh), BB.
- A 'download csv' link is located above the table.
- A tooltip for 'CC(vol)' is visible, explaining it as the Cross Correlation coefficient (0-1) between a model and target density map regions with the highest density values.

Brief descriptions of scores

Hovering mouse over the column name pops up a window with a brief description of the selected measure. A more detailed description can be found by consulting this document or original papers. PMIDs are provided in the pop-up window.

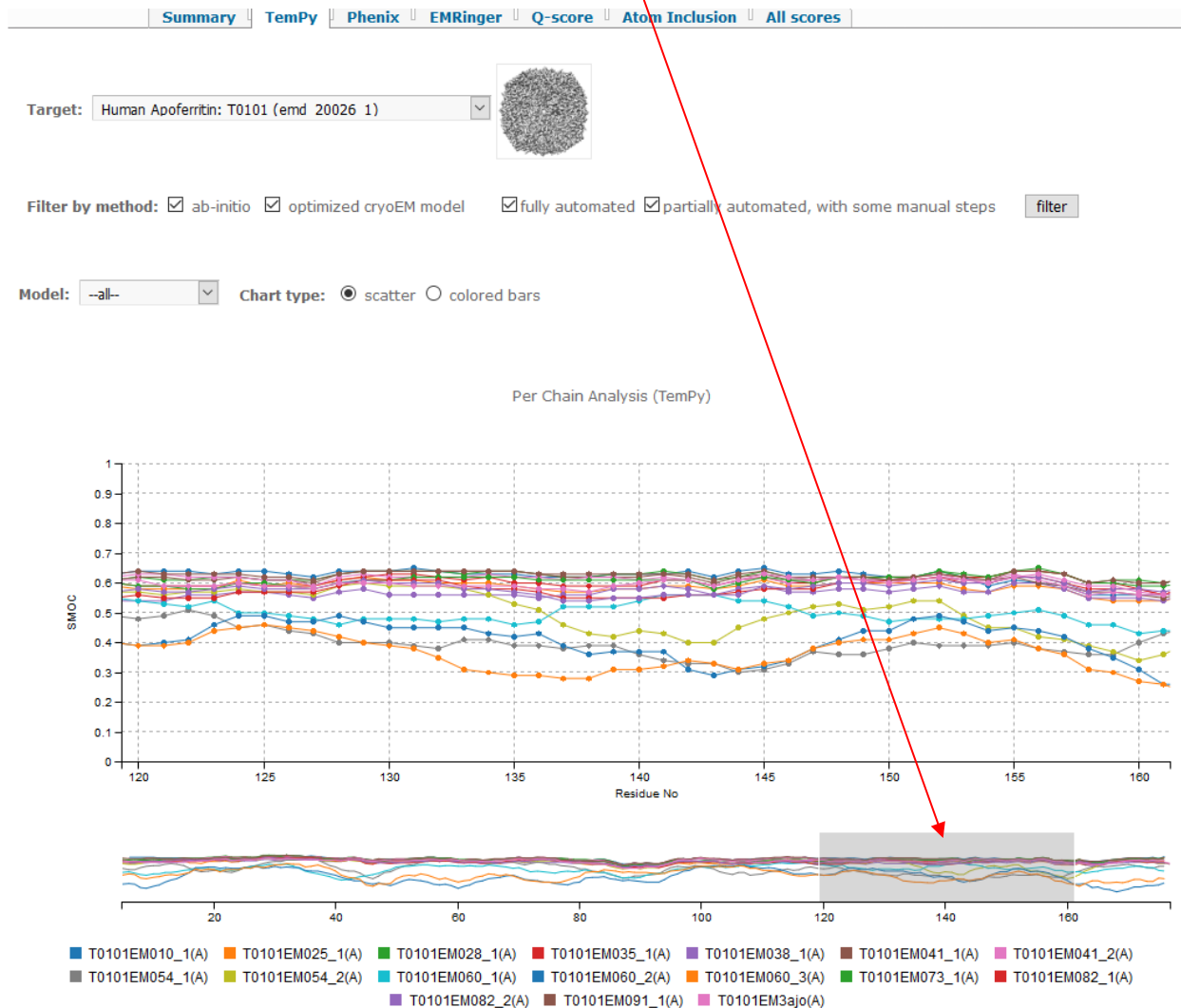
Downloading results as text files

Results presented in the interactive tables can be downloaded as text files in comma separated format by clicking on the 'download csv' link above the scores table.

Local accuracy plot: zoom in to region of interest

Per-residue scatter plots can be explored in more detail by narrowing the view to a region of interest.

In the lower (line-only) graph, place the cursor over the plot area (marker should turn into cross) and then click above the first residue of the desired interval, then drag cursor to the end residue (the area of interest will be highlighted grey) and release it – the top plot will change accordingly. The width of the selected window can be increased /decreased by placing the cursor at the edge of the grey box (marker will become a two-sided arrow) and then clicking and dragging the edge. Once the width of the grey window is selected (40 residues as shown), you can move this window along the whole graph by clicking on the main (upper) plot and moving the cursor to the left and right. Scrolling while the cursor is in the upper plot expands /narrows the selected window of residues.



Evaluation based on model coordinates only

Overview

This group evaluates models based exclusively on their atomic coordinates. Phenix and MolProbity report agreement of stereo-chemical features of the models with those of high-resolution experimental structures, while ProQ3 estimates global and local accuracy of the models based on knowledge-based potentials and features derived from the models and predicted from target's sequence.

Measures

*PHENIX (phenix.model_statistics)*¹

PHENIX model_statistics quantifies deviations of **bond distances**, **angles**, **chirality**, **planarity** and **dihedral angles** from ideal values¹⁸. For each parameter, three values are provided: RMSD, maximum deviation (in Ångstroms for distances or degrees for angles), and number of measured bonds, angles, etc.

*Molprobity scores*²

MolProbity validates agreement of model geometric parameters with high-resolution experimental structures (2 Å or better). Four Molprobity scores are reported:

Clash score reports the number of serious steric clashes per 1000 atoms. A clash is considered “serious” if steric overlap between any two atoms is > 0.4 Å. A good quality structure typically has clash-score < 20 .

Rot-out reports the percentage of sidechains classified as poor rotamers, from those sidechains that can be evaluated. A sidechain conformation is poor if its set of torsion angles falls outside the bounds of the rotamer definition.

Ram-out quantifies the percentage of residues with backbone conformations classified as outliers (i.e., those for which the combination of ϕ and ψ torsion angles is unusual), while

Ram-fav quantifies percentage of residues with conformations in favored Ramachandran plot regions, from those residues that can be evaluated.

*CaBLAM scores*³

CaBLAM (or $C\alpha$ Based Low-resolution Annotation Method) is a tool for validating low-resolution structures. In the 2.5–4.0Å resolution range, CaBLAM can be more robust in validating protein backbone than the Ramachandran analysis; for high-quality models, it typically provides little information beyond the Ramachandran validation.

Ca-out reports the percentage of $C\alpha$ geometry outliers.

Conf-out quantifies the percentage of backbone conformations classified as outliers.

Conf-disfav quantifies percentage of disfavored conformations (including outliers), from those residues that can be evaluated.

ProQ3 (a-priori model accuracy assessment)

ProQ3¹⁹ is based on a machine learning algorithm that combines knowledge-based Rosetta energy terms²⁰ with comparison of predicted and observed structural features, including contacts between different atom types, secondary structure and surface accessibility, and features predicted from sequence profiles. Local, per-residue accuracy is described in terms of S-score (0-1)²¹, and global accuracy is a normalized sum of the local values (in 0-1 range). Higher values correspond to more reliable estimates.

Model coordinates only web infrastructure

The **'model coordinates only'** tab of the website provides results in three subsections: Geometry Scores, ADP Histogram and Accuracy Estimate.

model coordinates only → Geometry Scores

The **Geometry Scores** tab reports scores calculated using the measures discussed in this section.

EM Model Challenge - 2019

Results

model coordinates only | fit to EM map | vs reference structure | vs other models | Comparative Analyses

Geometry Scores | ADPs Histogram | Accuracy Estimate

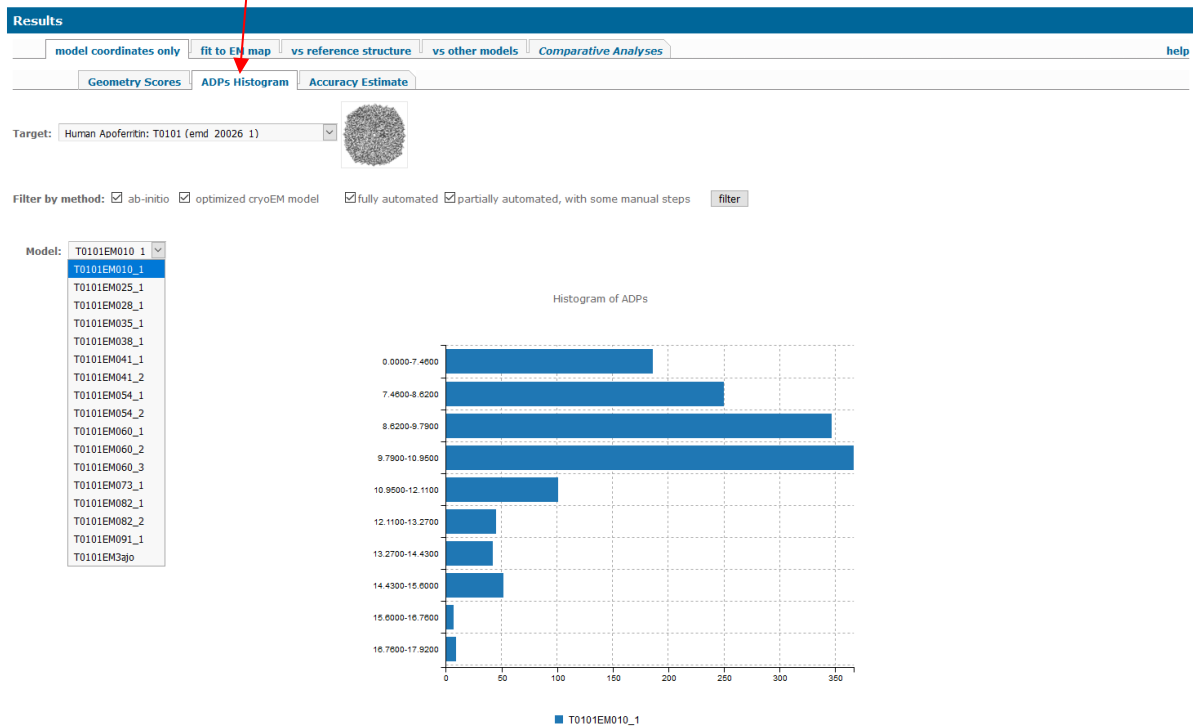
Target: Human Apoferritin: T0101 (emd 20026 1)

Filter by method: ab-initio optimized cryoEM model fully automated partially automated, with some manual steps

#	Model	Gr.Name	Method	Deviation from ideal model												Molprobity				Cblamb						
				ab-initio	auto	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	rmsd	max	count	Clash	Rot-out	Rama-out	Rama-fv	Conf-out	Conf-disfv
1	T0101EM054_2	khara	ab-initio	man.	0.020	0.142	1442	1.599	15.032	1943	0.096	0.318	204	0.005	0.027	257	9.347	77.749	871	0.000	0.000	0.590	98.470	1.20	4.00	0.00
2	T0101EM073_1	singharoy	optimized	auto	0.022	0.071	1443	2.310	9.235	1945	0.138	0.476	204	0.010	0.126	258	22.110	176.516	872	0.000	1.310	0.000	99.410	1.20	2.40	0.00
3	T0101EM091_1	chiu	optimized	man.	0.005	0.028	1513	0.817	8.655	2051	0.044	0.139	215	0.005	0.027	272	18.238	177.040	940	0.340	0.000	0.000	98.240	0.60	1.80	0.00
4	T0101EM025_1	cdmd	optimized	auto	0.016	0.075	1442	2.185	10.746	1942	0.144	0.502	204	0.012	0.058	256	17.637	177.003	872	0.360	1.960	0.000	97.060	1.20	2.40	0.00
5	T0101EM038_1	fzjewlich	optimized	auto	0.032	0.165	1443	1.979	18.628	1945	0.093	0.363	204	0.020	0.188	258	20.545	174.284	872	0.360	1.310	0.590	98.830	1.20	2.40	0.00
6	T0101EM082_2	rosetta	ab-initio	auto	0.017	0.219	1435	1.989	53.667	1932	0.069	0.291	202	0.009	0.121	255	10.727	137.135	868	0.360	0.000	0.590	98.830	1.20	3.60	0.00
7	T0101EM082_1	rosetta	ab-initio	auto	0.022	0.219	1435	1.994	53.105	1932	0.079	0.305	202	0.009	0.120	255	11.041	137.410	868	0.730	0.000	0.590	98.820	1.80	3.60	0.00
8	T0101EM010_1	yu	ab-initio	man.	0.004	0.033	1435	0.665	7.188	1932	0.039	0.156	202	0.003	0.016	255	13.527	140.426	868	1.090	0.000	0.000	98.220	0.60	2.40	0.00
9	T0101EM049		optimized	man.	0.006	0.038	1513	0.937	8.652	2051	0.072	0.312	215	0.004	0.020	272	20.746	179.679	940	1.370	0.810	0.000	98.240	0.60	1.80	0.00
10	T0101EM029_1	ccpem	ab-initio	man.	0.020	0.091	1442	1.984	8.865	1942	0.105	0.299	204	0.015	0.056	256	13.760	158.268	872	1.440	0.000	0.000	99.410	0.60	3.00	0.00
11	T0101EM041_2	arpwarp	ab-initio	auto	0.012	0.050	1450	1.842	9.831	1953	0.089	0.395	205	0.011	0.101	258	13.068	84.049	877	3.240	3.250	1.170	97.080	1.20	3.60	0.00
12	T0101EM035_1	phenix	ab-initio	man.	0.005	0.033	1442	0.542	4.975	1942	0.038	0.126	204	0.003	0.019	256	13.019	179.820	872	4.330	0.000	0.000	98.820	0.60	1.20	0.00
13	T0101EM054_1	khara	ab-initio	man.	0.027	0.135	1442	2.468	22.996	1943	0.142	0.529	204	0.010	0.035	257	9.898	65.823	871	4.340	0.000	4.120	86.470	12.50	22.60	2.38
14	T0101EM041_1	arpwarp	ab-initio	man.	0.012	0.044	1450	1.930	11.573	1953	0.084	0.324	205	0.007	0.020	258	13.978	149.523	877	4.670	2.600	0.560	98.490	2.40	3.60	0.00
15	T0101EM060_2	deeptracer	ab-initio	man.	0.031	0.103	1442	2.398	11.128	1942	0.135	0.461	204	0.003	0.039	256	18.480	179.994	872	36.100	0.000	0.000	94.710	5.40	10.10	1.79
16	T0101EM060_3	deeptracer	ab-initio	man.	0.031	0.103	1442	2.430	11.790	1942	0.115	0.348	204	0.002	0.017	256	18.410	179.991	872	41.160	0.000	0.000	95.290	6.00	8.90	2.38
17	T0101EM060_1	deeptracer	ab-initio	auto	0.047	0.400	1443	2.572	58.192	1944	0.246	1.574	204	0.019	0.092	257	9.760	59.867	873	42.280	0.650	2.940	92.940	7.10	10.70	0.60

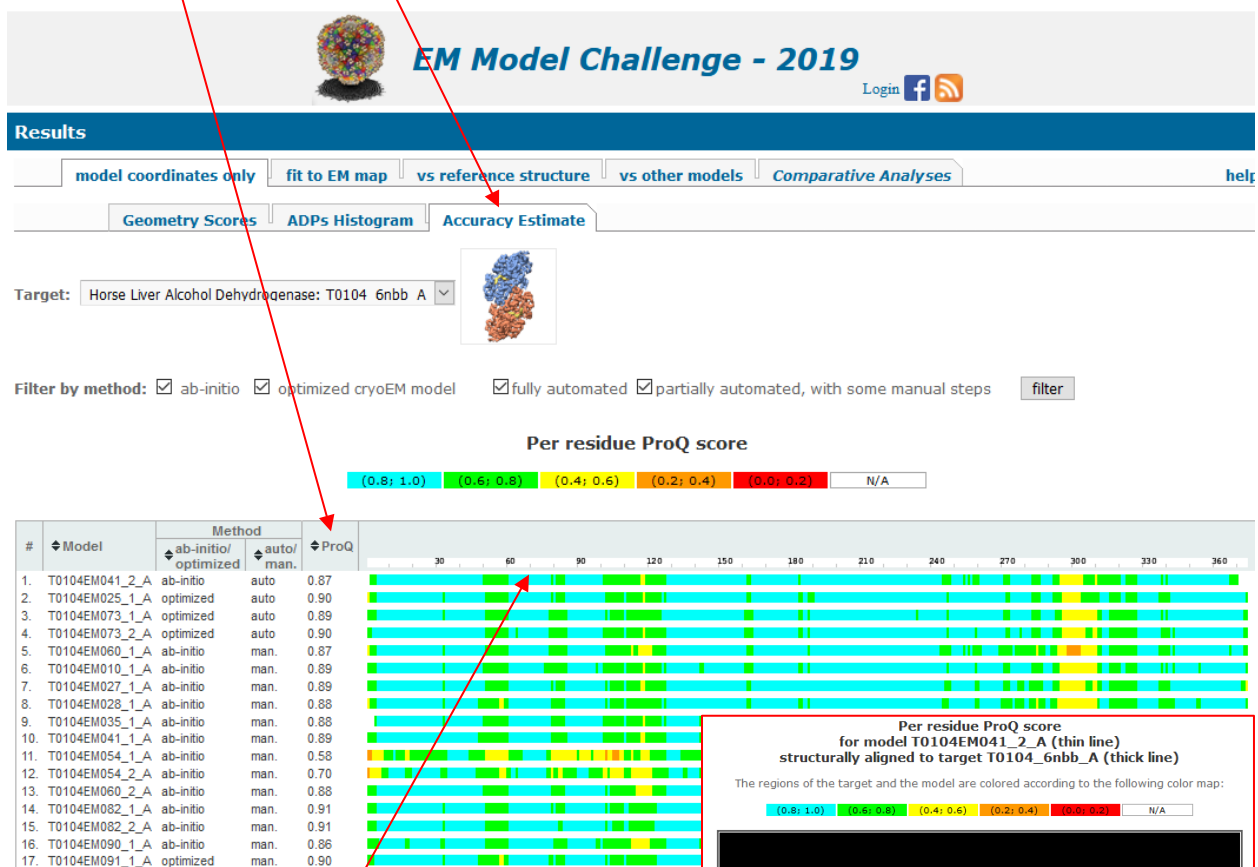
model coordinates only → Atomic displacement parameters (ADP)

The **ADP Histogram** tab presents a histogram of ADPs (B-factors) for the model selected in the 'Model' dropdown menu.



model coordinates only → Accuracy Estimate

The **Accuracy Estimate** tab shows color-coded barplots of the predicted local accuracy of each residue in the model in terms of the ProQ3 error S-function (see above). Residue numbers are provided in the header section of the plot. The cumulative accuracy estimate for the model (**ProQ score**) is provided to the left of the barplot.



Clicking on the color-coded bar for any model (e.g. **T0104EM041_2_A**) shows the selected model structurally aligned to the reference structure and colored according to the per-residue *ProQ* score (picture to the right).

Evaluation versus EM maps

Overview

Evaluation of each model's fit to its cryo-EM density map includes calculation of global and per-residue goodness-of-fit scores generated with PHENIX¹, TEMPy⁵⁻⁸, EMRinger⁹, Q-score¹⁰ and Atom Inclusion¹¹.

Some scores are highly sensitive to the presence/absence of hydrogen atoms in the model. For such measures, scores are calculated both for models including all atoms (*Orig.Model*) and for models stripped of their hydrogen atoms (*noH*).

The calculation of cross-correlation scores implicitly takes into account ADP/B-factors as well as x,y,z coordinates. Since some models were submitted with unreasonable (or missing) ADP/B-factors, we calculated the cross-correlations with the B-factors as submitted by modelers (*Orig.Model*) and without them (*BF=0*).

Measures

*PHENIX suite of scores for analysis of cryo-EM atomic models*²²

Box CC: real space cross-correlation coefficient (0-1) between a model and entire target density map. Higher values usually signify a better fit to map. Low values do not necessarily mean that the model does not fit the map well, but may instead indicate that there are uninterpreted map regions or poorly connected densities.

CC(mask): cross-correlation coefficient (0-1) between a model and target density map values inside a mask calculated around the macromolecule.

CC(vol): cross-correlation coefficient (0-1) between a model and target density map regions with the highest density values. The regions are defined by the N highest value points in the model-calculated map, with N being the number of grid points inside the molecular mask.

CC(peaks): cross-correlation coefficient (0-1) between a model and target density map regions with the highest density values. The regions are defined by the N highest value points in the model-calculated map and the N highest value points in the experimental map.

Resol. (FSC=0.5): model-map Fourier Shell Correlation resolution at FSC = 0.5.

The cross-correlation scores are calculated with originally submitted B-factors and with B-factors set to 0.

*TEMPy scores*⁵⁻⁸

CCC (cross-correlation coefficient) scores goodness of fit between the original map and the map calculated from the model coordinates at the author-specified resolution of the

experimental map (or an updated user-provided map). *CCC* is calculated by the array multiplication of density values at the same points in the model and target maps.

LAP (*Laplacian-filtered CCC*) is computed similarly to *CCC*, using density maps pre-processed with a Laplacian filter.

ENV (*Envelope*) estimates how much of the density map is filled with atoms, and penalizes protrusions from the map envelope. Larger *ENV* values denote better fits.

MI: (*Mutual Information*) is a statistical measure that quantifies the extent of register between two binned densities relative to their background distributions.

SMOC (*Segment Mander's Overlap Coefficient*) is a per-residue model-to-map fit measure, which calculates the Mander's overlap coefficient for overlapping residue fragments and assigns the score to the central residue in the fragment. The score is in [0-1] range with higher values indicating a better fit (*Note: The score can also take negative values when the density values in one of the maps are negative*). The *SMOC* score is also calculated for the whole structure by averaging the per-residue scores.

The *ENV* and *SMOC* scores are calculated for all atoms and non-hydrogen atoms. The other *TEMPy* scores are insensitive to the presence of hydrogen atoms in models.

*EMRinger score*⁹

EMRinger evaluates accuracy of side-chain placement within map density. There is a strong (negative) correlation between resolution and the overall *EMRinger* score. Side chain density is generally only resolvable for resolutions better than 4.5 Å. In general, for maps better than 3.5 Å resolution, the minimum expected score is 1. Most structures which have been carefully refined score above 1.5, with some getting scores above 3.

*Q-score*¹⁰

Q-score measures the resolvability of atomic-model features in a density map at each model atom position (non-hydrogen atoms only). Per-residue *Q-scores* are generalized into the global *Q-score* for the whole model. The global *Q-score* generally correlates well with the estimated resolution of a density map.

*Atom Inclusion scores*¹¹

Atom Inclusion reports the fraction of atoms within the target map contour. An atom is within the contour if its position in the map has a density value above the current threshold. Several variants of the score are calculated: *All:Orig.Model* is calculated on all atoms present in the submitted model; *All:noH* is calculated on all non-hydrogen atoms; *BB* is calculated on backbone atoms only.

Fit to EM map web infrastructure

The *fit to EM map* tab provides evaluation results for overall model-to-map fit (under *Global Accuracy* tab) and fit on a per-residue basis (*Local Accuracy* tab).

The *Global Accuracy* results calculated with the measures discussed above are presented under the *Scores* and *Plots* tabs.

fit to EM map → Global Accuracy → Scores

The *Scores* table provide all scores calculated in the 'fit to EM map' evaluation track. Values in five columns of the *Scores* table (**Resol. (2)**, **EMRinger** and **Atom Inclusion (2)**) are clickable.

Results

model coordinates only fit to EM map vs reference structure vs other models Comparative Analyses help

Global Accuracy Local Accuracy

Scores Plots

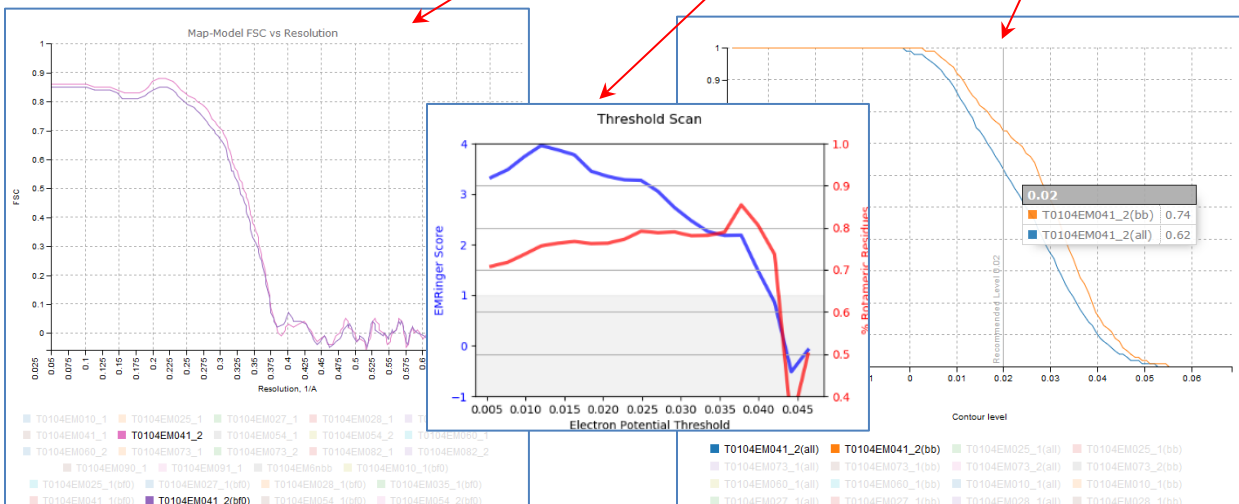
Target: Horse Liver Alcohol Dehydrogenase: T0104 (emd_0406)

Filter by method: ab-initio optimized cryoEM model fully automated partially automated, with some manual steps filter

download csv

#	Model	Group	Method	EM Map	Tempy										Phenix										EMRinger		Q-score		Atom Inclusion	
					MapID	Res. ok	CCC	LAP	CCC (ov)	LAP (ov)	MI	MI (ov)	SMDC	box CC	CC(mask)	CC(vol)	CC(peak)	FSC@0.5	EMRinger score	Q-score	Atom Inclusion All	Atom Inclusion BB								
1	T0104EM073_1	singharoy	optimized	auto	emd_0406_2.9	0.32	0.48	0.19	0.98	0.98	0.04	0.23	0.72	0.72	0.76	0.76	0.82	0.82	0.82	0.81	0.73	3.01	3.01	5.58	0.64	0.64	0.63	0.63	0.76	
2	T0104EM035_1	phenix	ab-initio	man	emd_0406_2.9	0.32	0.49	0.20	0.98	0.98	0.04	0.24	0.74	0.74	0.64	0.78	0.85	0.85	0.84	0.56	0.76	2.98	3.01	5.15	0.65	0.65	0.65	0.65	0.77	
3	T0104EM052_1	rosetta	ab-initio	man	emd_0406_2.9	0.31	0.49	0.19	0.98	0.98	0.04	0.24	0.66	0.73	0.73	0.75	0.88	0.82	0.86	0.81	0.70	0.72	2.98	3.04	5.11	0.50	0.63	0.33	0.64	0.75
4	T0104EM025_2	rosetta	ab-initio	man	emd_0406_2.9	0.31	0.49	0.19	0.98	0.98	0.04	0.24	0.66	0.73	0.76	0.75	0.88	0.82	0.86	0.81	0.73	0.72	2.97	3.03	5.05	0.51	0.64	0.32	0.64	0.75
5	T0104EM091_1	chlu	optimized	man	emd_0406_2.9	0.32	0.48	0.20	0.98	0.98	0.04	0.23	0.73	0.73	0.74	0.77	0.89	0.83	0.88	0.82	0.70	0.75	2.95	2.98	4.92	0.65	0.65	0.64	0.64	0.76
6	T0104EM028_1	copem	ab-initio	man	emd_0406_2.9	0.31	0.48	0.19	0.98	0.98	0.04	0.23	0.71	0.71	0.73	0.75	0.90	0.82	0.89	0.81	0.69	0.73	2.95	3.03	4.84	0.63	0.63	0.62	0.62	0.74
7	T0104EM027_1	kuamar	ab-initio	man	emd_0406_2.9	0.32	0.49	0.20	0.98	0.98	0.04	0.24	0.74	0.74	0.65	0.78	0.85	0.84	0.85	0.84	0.57	0.76	3.00	3.04	4.78	0.65	0.65	0.64	0.64	0.76
8	T0104EM010_1	yu	ab-initio	man	emd_0406_2.9	0.32	0.48	0.20	0.98	0.98	0.04	0.24	0.74	0.74	0.77	0.89	0.83	0.87	0.83	0.70	0.75	2.95	2.97	4.69	0.66	0.66	0.64	0.64	0.75	
9	T0104EM025_1	cdmd	optimized	auto	emd_0406_2.9	0.31	0.46	0.18	0.98	0.98	0.04	0.21	0.71	0.71	0.73	0.73	0.79	0.79	0.79	0.79	0.70	0.70	3.12	3.12	4.25	0.60	0.60	0.61	0.61	0.75
10	T0104EM041_2	aswarup	ab-initio	auto	emd_0406_2.9	0.32	0.47	0.19	0.98	0.98	0.04	0.22	0.73	0.73	0.74	0.75	0.89	0.82	0.86	0.86	0.76	0.71	2.99	3.05	4.96	0.65	0.65	0.67	0.62	0.74
11	T0104EM073_2	singharoy	optimized	auto	emd_0406_2.9	0.31	0.45	0.18	0.98	0.98	0.04	0.19	0.69	0.69	0.73	0.73	0.78	0.78	0.78	0.69	0.69	3.11	3.11	3.92	0.60	0.60	0.61	0.60	0.75	
12	T0104EM066b	-	-	-	emd_0406_2.9	0.31	0.45	0.18	0.98	0.98	0.04	0.20	0.70	0.70	0.69	0.73	0.84	0.78	0.83	0.78	0.83	3.09	3.11	3.75	0.59	0.58	0.60	0.60	0.75	
13	T0104EM054_4	khara	ab-initio	man	emd_0406_2.9	0.29	0.44	0.16	0.98	0.98	0.04	0.18	0.59	0.65	0.65	0.66	0.75	0.70	0.73	0.70	0.58	0.59	3.31	3.33	3.39	0.41	0.53	0.30	0.56	0.69
14	T0104EM011_1	aswarup	ab-initio	man	emd_0406_2.9	0.32	0.48	0.19	0.98	0.98	0.04	0.23	0.73	0.73	0.78	0.77	0.91	0.83	0.88	0.82	0.74	2.97	3.01	3.03	0.64	0.64	0.63	0.63	0.74	
15	T0104EM060_2	deeptreacher	ab-initio	man	emd_0406_2.9	0.28	0.40	0.14	0.98	0.98	0.03	0.15	0.61	0.62	0.62	0.62	0.66	0.66	0.66	0.55	0.55	3.42	3.42	2.78	0.45	0.45	0.45	0.54	0.70	
16	T0104EM060_1	mbaker	ab-initio	man	emd_0406_2.9	0.31	0.46	0.18	0.98	0.98	0.04	0.21	0.71	0.71	0.61	0.73	0.80	0.80	0.80	0.79	0.51	0.70	3.03	3.11	2.62	0.61	0.61	0.61	0.61	0.75
17	T0104EM060_1	deeptreacher	ab-initio	man	emd_0406_2.9	0.29	0.41	0.15	0.98	0.98	0.03	0.15	0.62	0.63	0.64	0.64	0.67	0.67	0.68	0.68	0.56	0.42	3.42	3.42	2.18	0.46	0.50	0.45	0.54	0.71
18	T0104EM054_1	khara	ab-initio	man	emd_0406_2.9	0.27	0.41	0.14	0.98	0.98	0.03	0.15	0.56	0.61	0.61	0.61	0.67	0.63	0.65	0.64	0.53	0.52	3.57	3.57	1.61	0.35	0.47	0.29	0.51	0.65

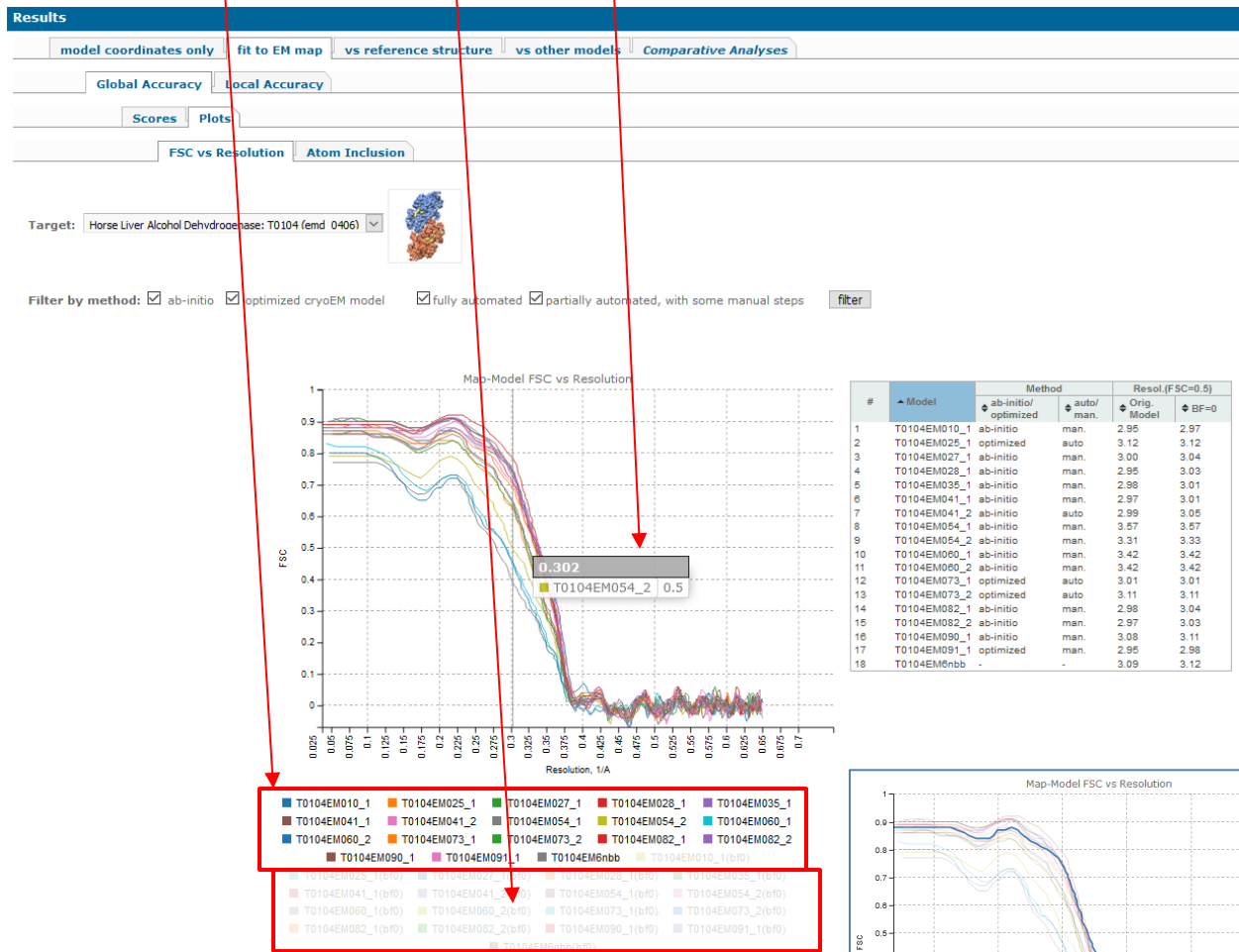
Clicking on the scores in one of the five clickable columns brings up associated plots for these models, e.g. for T0104EM041_2 (highlighted in yellow):



The **Fit to EM map: Global Accuracy: Plots** tab includes results of the **FSC versus Resolution** calculations and **Atom Inclusion** plots.

fit to EM map → Global Accuracy → Plots → FSC versus Resolution

In the **FSC versus Resolution** plots, the lines for all submitted models are shown by default with **model legends provided beneath the graph**. Lines for post-processed models with B-factors set to 0 (bf0) are not shown but **listed in light grey beneath the original models**. The visibility of a line in the graph can be changed by clicking on the model name in the model list. The table to the right of the graph shows resolution values corresponding to FSC=0.5. Moving the mouse along the selected curve shows corresponding **coordinates for this model**.



Hovering the mouse over the model name highlights the line for the selected model with other lines being greyed out.

With mouse over the plot area, rolling the scroll wheel zooms the plot in and out along the x-axis.

fit to EM map → Global Accuracy → Plots → Atom Inclusion

In the **Atom Inclusion** plots, the lines for all submitted models are shown by default with model legends provided beneath the graph. **Model names and atom inclusion ratios** for different contour levels are shown with hovering mouse over the curve(s). For each model, two curves can be shown: one considering all atoms in the model (all) and the other backbone atoms only (bb). Status of a line in the graph (visible /invisible) can be changed by clicking on the model name in the model list (model names in black /grey correspond to visible /invisible lines, correspondingly).



The *Fit to EM map: Local Accuracy* results are presented through seven tabs:

Per-chain Summary, TemPy, Phenix, EMRinger, Q-score, Atom Inclusion and All scores.

fit to EM map → Local Accuracy → Per-chain Summary

While the *Global Accuracy → Scores* tab shows scores for the whole multimeric model, the *Local Accuracy → Per-chain Summary* tab provides model-to-map fit scores for separate chains. Scores in TemPy (SMOC) and PHENIX (box CC) columns are clickable. The links take users to score-specific pages for the selected model (see below).

Results

model coordinates only | fit to EM map | vs reference structure | vs other models | Comparative Analyses | help

Global Accuracy | Local Accuracy

Per-chain Summary | TemPy | Phenix | EMRinger | Q-score | Atom Inclusion | All scores

Target: Horse Liver Alcohol Dehydrogenase: T0104 (emd 0406)

Filter by method: ab-initio optimized cryoEM model fully automated partially automated, with some manual steps

[download csv](#)

#	Model	Gr.Name	Method		Chain	EM Map	TemPy	Phenix
			ab-initio/optimized	auto/man.		MapID	SMOC	box CC
1	T0104EM041_2	arpwarp	ab-initio	auto	A	emd_0406	0.73	0.84
2	T0104EM041_2	arpwarp	ab-initio	auto	B	emd_0406	0.73	0.85
3	T0104EM025_1	cdmd	optimized	auto	A	emd_0406	0.71	0.79
4	T0104EM025_1	cdmd	optimized	auto	B	emd_0406	0.70	0.78
5	T0104EM073_1	singharoy	optimized	auto	A	emd_0406	0.72	0.81
6	T0104EM073_1	singharoy	optimized	auto	B	emd_0406	0.72	0.81
7	T0104EM073_2	singharoy	optimized	auto	A	emd_0406	0.69	0.78
8	T0104EM073_2	singharoy	optimized	auto	B	emd_0406	0.70	0.78
9	T0104EM060_1	deeptrocer	ab-initio	man.	A	emd_0406	0.62	0.67
10	T0104EM060_1	deeptrocer	ab-initio	man.	B	emd_0406	0.62	0.67
11	T0104EM010_1	yu	ab-initio	man.	A	emd_0406	0.74	0.84
12	T0104EM010_1	yu	ab-initio	man.	B	emd_0406	0.74	0.84
13	T0104EM027_1	kumar	ab-initio	man.	A	emd_0406	0.74	0.78
14	T0104EM027_1	kumar	ab-initio	man.	B	emd_0406	0.74	0.78
15	T0104EM028_1	copem	ab-initio	man.	A	emd_0406	0.71	0.86
16	T0104EM028_1	copem	ab-initio	man.	B	emd_0406	0.72	0.86
17	T0104EM035_1	phenix	ab-initio	man.	A	emd_0406	0.74	0.77
18	T0104EM035_1	phenix	ab-initio	man.	B	emd_0406	0.74	0.77
19	T0104EM041_1	arpwarp	ab-initio	man.	A	emd_0406	0.73	0.87
20	T0104EM041_1	arpwarp	ab-initio	man.	B	emd_0406	0.73	0.88
21	T0104EM054_1	kihara	ab-initio	man.	A	emd_0406	0.66	0.83
22	T0104EM054_1	kihara	ab-initio	man.	B	emd_0406	0.66	0.85
23	T0104EM054_2	kihara	ab-initio	man.	A	emd_0406	0.60	0.71
24	T0104EM054_2	kihara	ab-initio	man.	B	emd_0406	0.69	0.70
25	T0104EM060_2	deeptrocer	ab-initio	man.	A	emd_0406	0.60	0.65
26	T0104EM060_2	deeptrocer	ab-initio	man.	B	emd_0406	0.61	0.65
27	T0104EM082_1	rosetta	ab-initio	man.	A	emd_0406	0.66	0.83
28	T0104EM082_1	rosetta	ab-initio	man.	B	emd_0406	0.66	0.83
29	T0104EM082_2	rosetta	ab-initio	man.	A	emd_0406	0.66	0.84
30	T0104EM082_2	rosetta	ab-initio	man.	B	emd_0406	0.66	0.84
31	T0104EM090_1	mbaker	ab-initio	man.	A	emd_0406	0.71	0.73
32	T0104EM090_1	mbaker	ab-initio	man.	B	emd_0406	0.71	0.73
33	T0104EM091_1	chiu	optimized	man.	A	emd_0406	0.73	0.85
34	T0104EM091_1	chiu	optimized	man.	B	emd_0406	0.73	0.85
35	T0104EM6nbb	-	-	-	A	emd_0406	0.69	0.79
36	T0104EM6nbb	-	-	-	B	emd_0406	0.70	0.79

fit to EM map → Local Accuracy → TemPy / Phenix / EMRinger / Q-score / Atom Inclusion

The **five score-specific tabs** show different local (per-residue) scores for models.

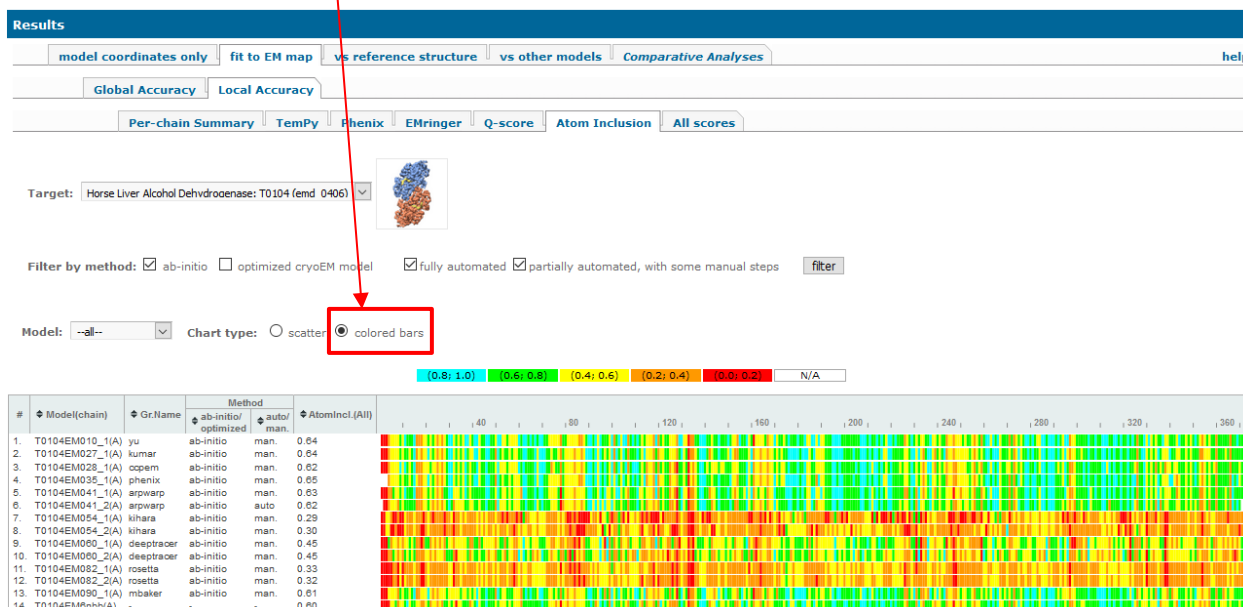
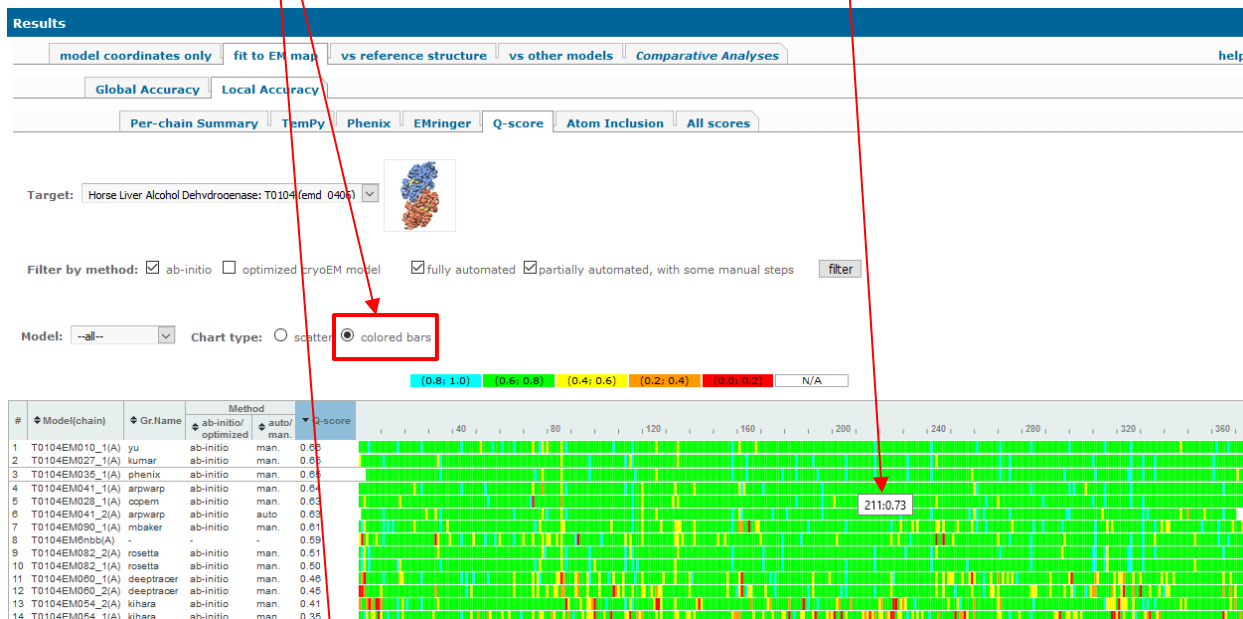
If a specific model is selected in the dropdown menu then the plot shows data for one chain of this model by default. Clicking on the **chain name in the chain list** beneath the plot hides /unhides lines for other chains. The cumulative per-chain scores are **shown to the right of the plot.**



If **all models** are selected, the plot shows evaluation data for one chain in each of the models (the representative chain name is listed after the model name in parenthesis). Lines for models can be hidden/unhidden by clicking on the **model name in the model list**.

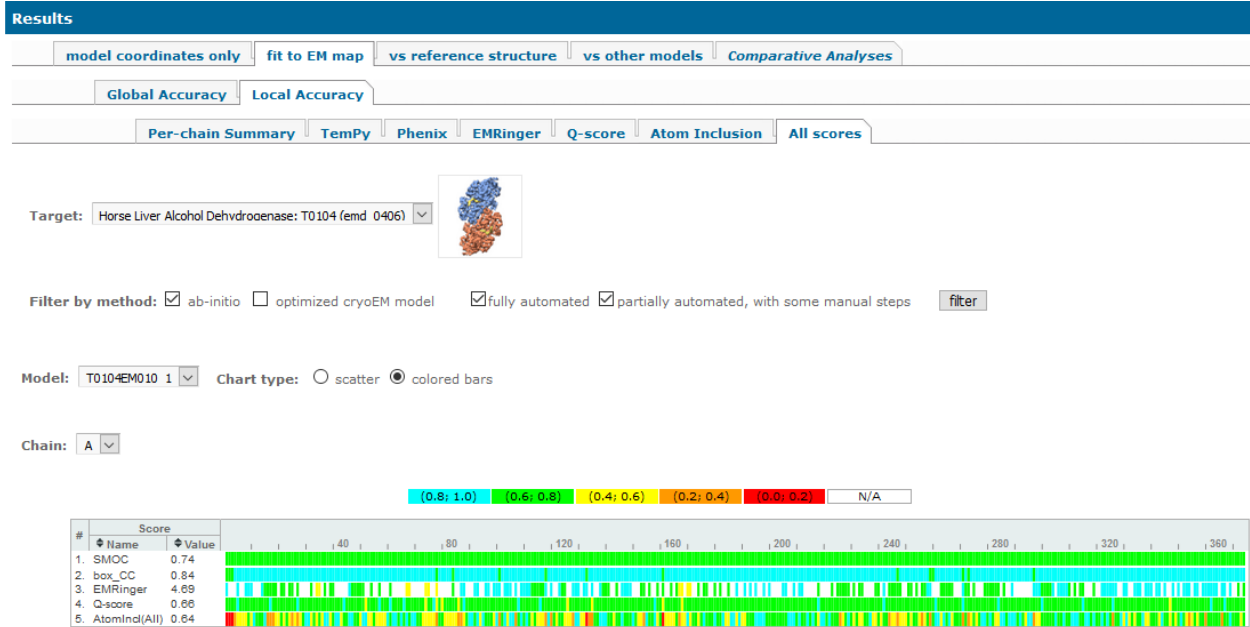


All graphs are shown as scatter plots by default. The view can be switched to colored bars by clicking on the **'colored bars'** radio button (see Q-score and Atom Inclusion pages as examples below). Hovering over a bar shows value of the score for the selected residue.



fit to EM map → *Local Accuracy* → *All Scores*

This tab shows all five local scores (see above) for a selected model in one web page.



Assessment of models versus reference structure(s) is carried out at two levels:

- Monomeric: accuracy of separate subunits (protein chains) are evaluated;
- Multimeric: model-reference agreement is evaluated for the whole multimeric structure.

Evaluation versus reference structure (monomeric mode)

Overview

Submitted multimeric models are first split into chains. The separate chains are checked for similarity to each other, and all differing chain-based models are evaluated separately.

Comparisons to reference structures are made using rigid-body superposition-based measures (RMSD, GDT_TS, GDT_HA and GDC-SC), and local-based superposition-free measures (LDDT and CAD).

Measures

Superposition-based LGA family of scores (GDT_TS^{13,23}, GDT_HA, GDC_SC²⁴)

GDT_TS, GDT_HA and GDC_SC scores are all calculated with the LGA package.

GDT_TS (*Global Distance Test – Total Score*) reports the average percentage of model C α atoms that can be superimposed with the reference structure under 1, 2, 4, and 8 Å distance cutoffs. Only well-modeled regions contribute to the GDT_TS score, in contrast to RMSD, where all residues contribute, including superposition outliers. The GDT_TS score is in the range [0-100] with higher scores corresponding to better fit. GDT_TS scores over 50 indicate structures with significant similarity, while scores below 25 indicate unrelated structures.

GDT_HA (*Global Distance Test – High Accuracy*) is a modification of the GDT_TS score that uses tighter distance cut-offs (0.5, 1, 2 and 4 Å) and thus is better suited for the evaluation of high accuracy models. GDT_HA scores are in the range 0-100; they are highly correlated with GDT_TS scores, usually 10-20 points lower for the same models.

GDC_SC (*Global Distance Calculation for Side Chains*) measure is calculated similarly to GDT_TS but using a characteristic atom near the end of each side chain as a residue representative (instead of C α). This measure implicitly brings the accuracy of side chain modeling into the scoring formula. GDC_SC is scaled in the range 0-100.

Superposition-independent structure-based family of scores (LDDT, CADaa)

LDDT and CAD scores are superposition-free measures of local structure; they are more effective in assessing quality of multi-domain models. While rigid body superposition-based scores (e.g., GDT_TS) are very sensitive to relative domain orientation (as superposition of two multi-domain structures is usually dominated by one of the domains), these local measures are

practically insensitive to spatial inter-domain arrangements and therefore are well suited for evaluation of model quality in such cases.

LDDT¹⁴ is based on comparison of all-atom distance maps between model and target structures. The algorithm determines the percentage of preserved distances between all pairs of atoms in the target structure that are closer in space than a predefined cutoff. The final score is the average of the percentages of the preserved distances under four distance tolerance cutoffs (0.5, 1, 2 and 4 Å). The LDDT score range is [0-1].

CAD-score¹⁵ estimates similarity of two structures based on differences in inter-residue-residue contact areas. The inter-residue contact areas can be defined for any subset of atoms in a residue (e.g., backbone or side-chain-only). In our system we report a variant of the CAD-score that is based on comparison of contact areas for all atoms in a residue (**CADaa**). The contact areas are calculated using the Voronyi tessellation approach in the target and the model separately, and then per-residue differences are summed and normalized to the [0-1] interval. Based on CASP evaluation data, the CAD-score has a bell-shaped distribution with around 90% of scores falling in the range [0.3; 0.7]. CAD score has a desired feature of favoring models with better stereo-chemical arrangements²⁵.

Guide to web infrastructure

The *vs reference structure* → *Monomers* tab provides evaluation results for overall similarity of the monomeric subunits of the model to the reference structures (under *Global Accuracy* tab) and their similarity on a per-residue basis (*Local Accuracy* tab).

vs reference structure → *Monomers* → *Global Accuracy* → *Scores*

The *Scores* tab reports all scores calculated in the '*vs reference structure* → *Monomers*' evaluation track. Values in the **GDT_TS** column are clickable. Clicking on a GDT_TS value for the selected model brings up a GDT plot showing the percentage of fit residues for distance cutoffs from 0 to 10 Å (see comments in the *Monomers* → *Global Accuracy* → *Plots* tab below).

Results

model coordinates only
fit to EM map
vs reference structure
vs other models
Comparative Analyses

Multimer
Monomers

Global Accuracy
Local Accuracy

Scores
Plots

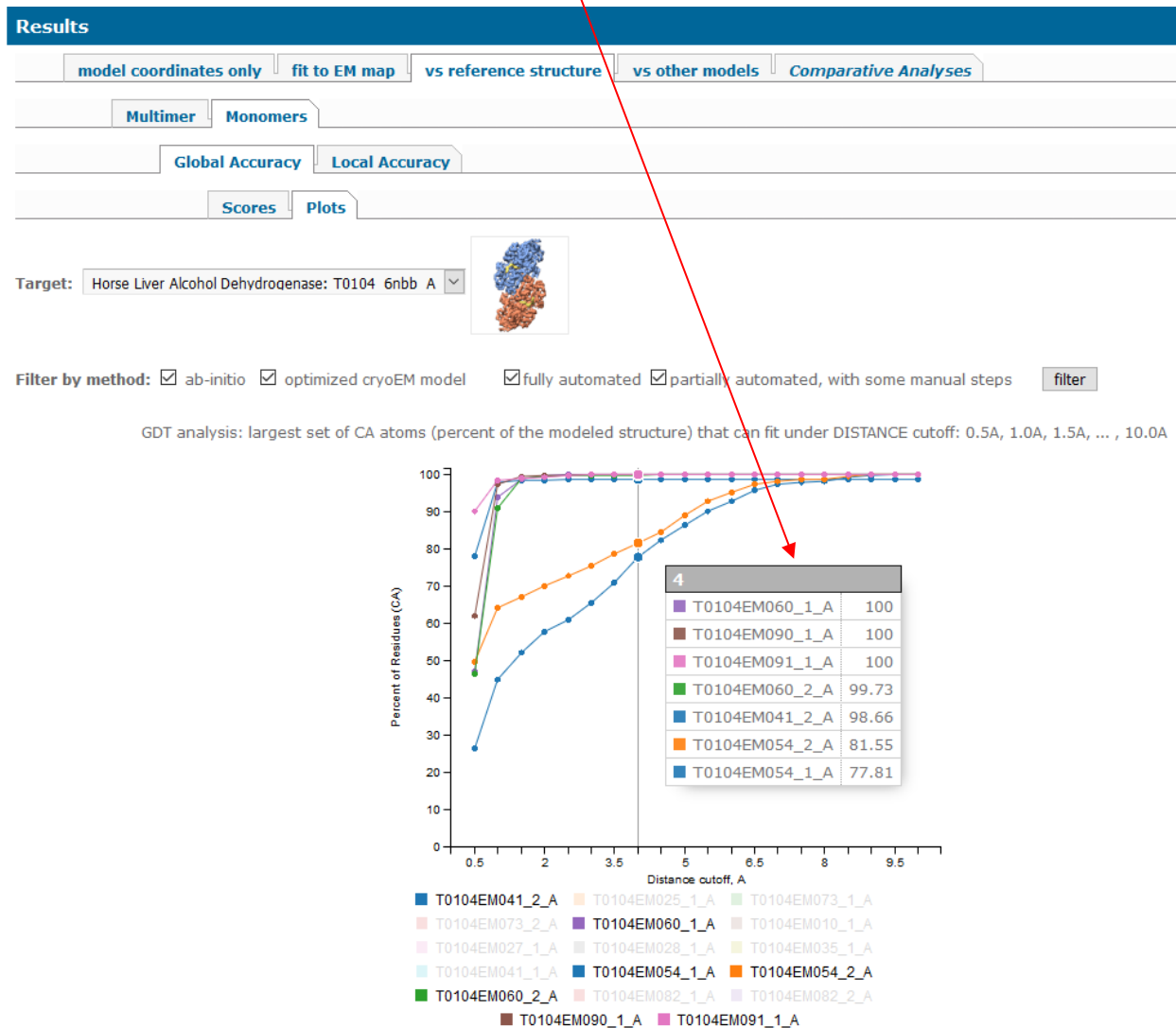
Target: Horse Liver Alcohol Dehydrogenase: T0104_6nbb_A

Filter by method: ab-initio optimized cryoEM model fully automated partially automated, with some manual steps filter

#	Model	Gr.Name	Method		Superposition-free measures (similarity of local geometric features)			Superposition-based measures (rigid body similarity)			
			ab-initio/ optimized	auto/ man.	LDDT	CAD	RMSD	GDT_TS	GDT_HA	GDC_ALL	GDC_SC
1	T0104EM041_2_A	arpwarp	ab-initio	auto	0.80	0.79	0.45	98.40	93.25	86.49	71.49
2	T0104EM025_1_A	cdmd	optimized	auto	0.91	0.83	0.45	99.67	94.92	89.09	72.90
3	T0104EM073_1_A	singharoy	optimized	auto	0.93	0.85	0.42	99.53	96.12	91.70	77.94
4	T0104EM073_2_A	singharoy	optimized	auto	0.95	0.89	0.37	99.60	97.59	94.12	84.81
5	T0104EM060_1_A	deeptracer	ab-initio	man.	0.83	0.75	0.66	98.40	85.16	79.00	58.96
6	T0104EM010_1_A	yu	ab-initio	man.	0.92	0.83	0.42	99.53	95.86	90.13	75.76
7	T0104EM027_1_A	kumar	ab-initio	man.	0.93	0.84	0.42	99.60	96.06	91.12	76.85
8	T0104EM028_1_A	ccpem	ab-initio	man.	0.90	0.81	0.48	99.33	94.78	87.85	69.11
9	T0104EM035_1_A	phenix	ab-initio	man.	0.92	0.84	0.38	99.00	95.59	89.99	74.83
10	T0104EM041_1_A	arpwarp	ab-initio	man.	0.90	0.82	0.46	99.60	95.32	88.22	70.56
11	T0104EM054_1_A	kihara	ab-initio	man.	0.54	0.53	3.34	69.65	51.74	45.22	33.11
12	T0104EM054_2_A	kihara	ab-initio	man.	0.63	0.61	2.93	78.61	66.38	59.85	46.84
13	T0104EM060_2_A	deeptracer	ab-initio	man.	0.83	0.75	0.74	97.53	84.16	78.13	58.24
14	T0104EM082_1_A	rosetta	ab-initio	man.	0.92	0.83	0.44	99.60	95.12	90.17	75.69
15	T0104EM082_2_A	rosetta	ab-initio	man.	0.92	0.83	0.43	99.53	95.45	90.01	74.22
16	T0104EM090_1_A	mbaker	ab-initio	man.	0.86	0.78	0.57	99.27	89.77	84.29	68.06
17	T0104EM091_1_A	chiu	optimized	man.	0.94	0.86	0.42	99.40	96.93	92.41	80.45

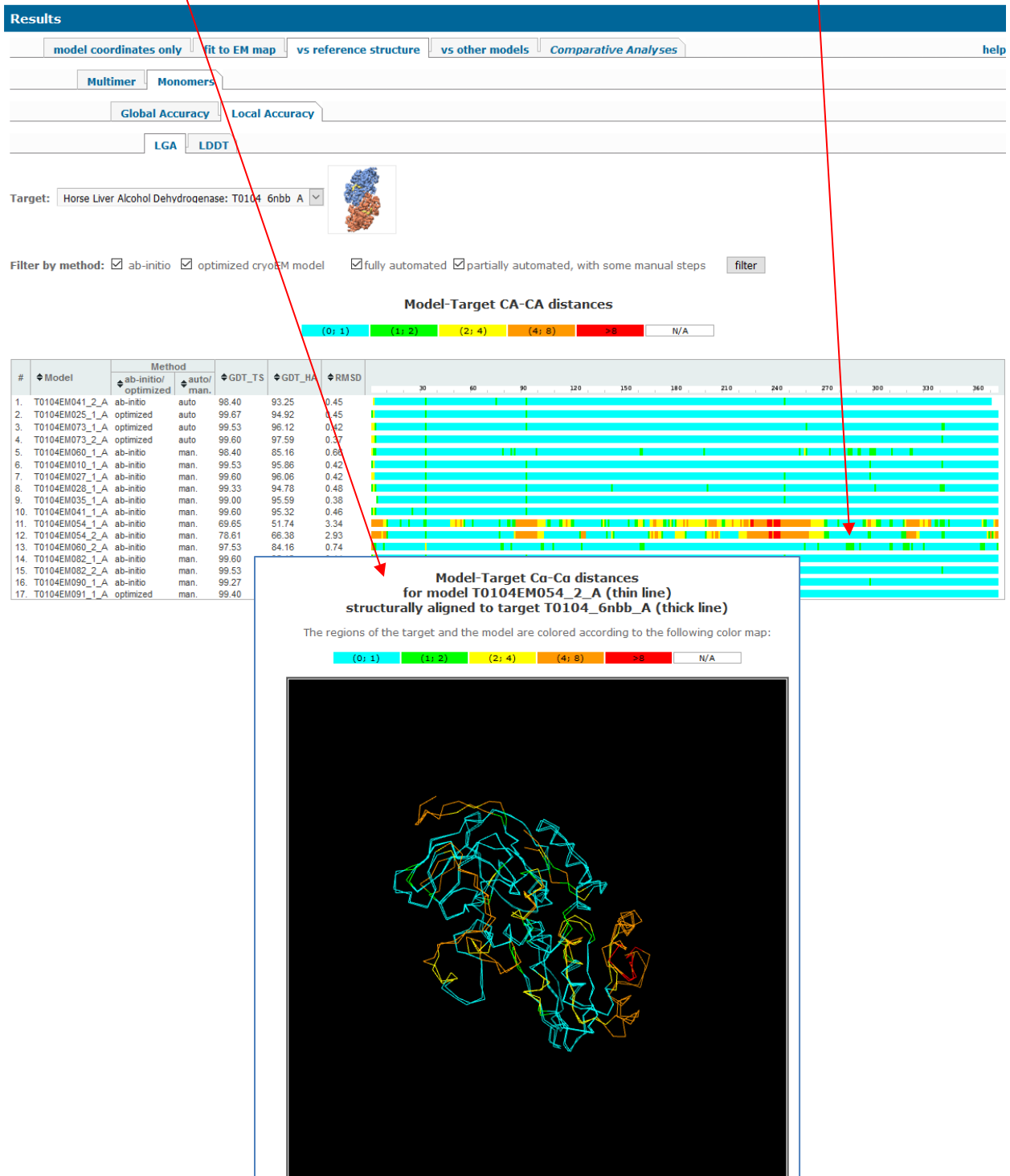
vs reference structure → *Monomers* → *Global Accuracy* → *Plots*

Accuracy of the model versus the reference structure is visually summarized by GDT plots showing percentage of residues in the model that can be superimposed into the target under the specified residue-residue distance cutoff. A larger area under the curve indicates a more accurate model. An ideal model would be represented by a line going straight up and then staying horizontally across the whole range of distance cutoffs. Graphs are interactive: lines can be switched on and off and the **underlying scores** can be shown using the techniques described in the General Information section.



vs reference structure → Monomers → Local Accuracy → LGA /LDDT

There are two tabs under the Local Accuracy– LGA and LDDT. Clicking on either tab shows per-residue accuracy of models as color-coded bars. The LGA tab displays Cα-Cα distances between corresponding residues in model and target after their optimal LGA superposition, while the LDDT tab shows per-residue LDDT score. **Clicking on a data bar** shows structural superposition of the model and the target colored the same way as the underlying bar.



Evaluation versus reference structure (multimeric mode)

Overview

Submitted multimeric models are evaluated as single units.

Comparison to the reference structures is made with rigid-body superposition-based measures (GDT_TS, GDT_HA, GDC_ALL and GDC_SC, RMSD), local-based superposition-free measures (QS-best, QS-global, LDDTo and LDDTw), descriptive statistical measures (N_close, N_far, CA_score and Seq.match) and hydrogen bonding similarity measures (precision and Jaccard_coefficient).

Measures

QS scores

Evaluation of multimers vs the reference structure is carried out using the **QS**¹² suite of measures (QS stands for Quaternary Structure). The score quantifies the similarity between quaternary structures in terms of shared interfacial contacts of their subunits. The package first finds the best mapping between the target and model chains using the structure symmetry, and then reports five scores:

QS_best: fraction of interchain contacts ($C\beta-C\beta < 12\text{\AA}$) shared between two structures for best fitting interface;

QS_global: fraction of interchain contacts shared between two structures for all interfaces;

RMSD calculated on the whole aligned structure (C α 's of all common chains);

LDDTo (*Local Distance Difference Test, oligomeric*): the LDDT score (see monomeric section) calculated on whole oligomeric structure;

LDDTw (*Local Distance Difference Test, weighted*): the LDDT score calculated first on each chain separately and then length-weighted for the whole multimeric structure.

The LDDT scores described above were adapted for multimeric structures in a way that does not penalize for over-prediction, e.g. a tetrameric model (containing a perfect dimeric model) vs the dimeric target is giving a perfect score. QS-scores are ranked in [0-1] interval. Scores above 0.7 indicate highly similar quaternary structures, while scores below 0.3 indicate low assembly similarity.

Superposition-based LGA family of scores (GDT_TS^{13,23}, GDT_HA, GDC_SC²⁴, GDC_ALL)

LGA scores for multimeric structures are calculated similarly to monomeric ones (see monomeric section), using the chain correspondence established with the QS-tool (above).

phenix.chain_comparison module

PHENIX's chain_comparison module calculates the proximity of model and target structures, once coordinates of both are optimally fit to the density. This is important when analyzing *ab initio* models, which may be incomplete, have sequence errors, or have regions of unassigned sequence. The method reports

N(close) – number of C α atoms within 3Å of corresponding atoms in the target;

N(far) – number of C α s further than 3Å;

CA_score – number of C α s within 3Å of the target divided by the rmsd;

Seq.match % - percentage of C α atoms with correct residue name.

Hydrogen bonds scores

Hydrogen bond scores report accuracy of reproducing the target's hydrogen bonds. Hydrogen bonds are first identified with HBPLUS¹⁶ and then compared with precision and Jaccard coefficient statistics.

Precision (>0) = TP/(TP+FP) is the fraction of correctly reproduced hydrogen bonds in a model. TP is the number of correctly reproduced hydrogen bonds in the model; FP is the number of hydrogen bonds in the model that are absent in the target;

Precision (>6): the precision measure (above) calculated on a subset of non-local hydrogen bonds (minimal sequence separation of six residues);

Jaccard coefficient (>0) = TP/(TP+FP+FN) is a statistical measure of similarity of hydrogen bonds in a model and the target. The measure is stricter than the precision as it additionally penalizes models for not reproducing hydrogen bonds present in the target (FN is the number of hydrogen bonds in the target that were not reproduced in the model).

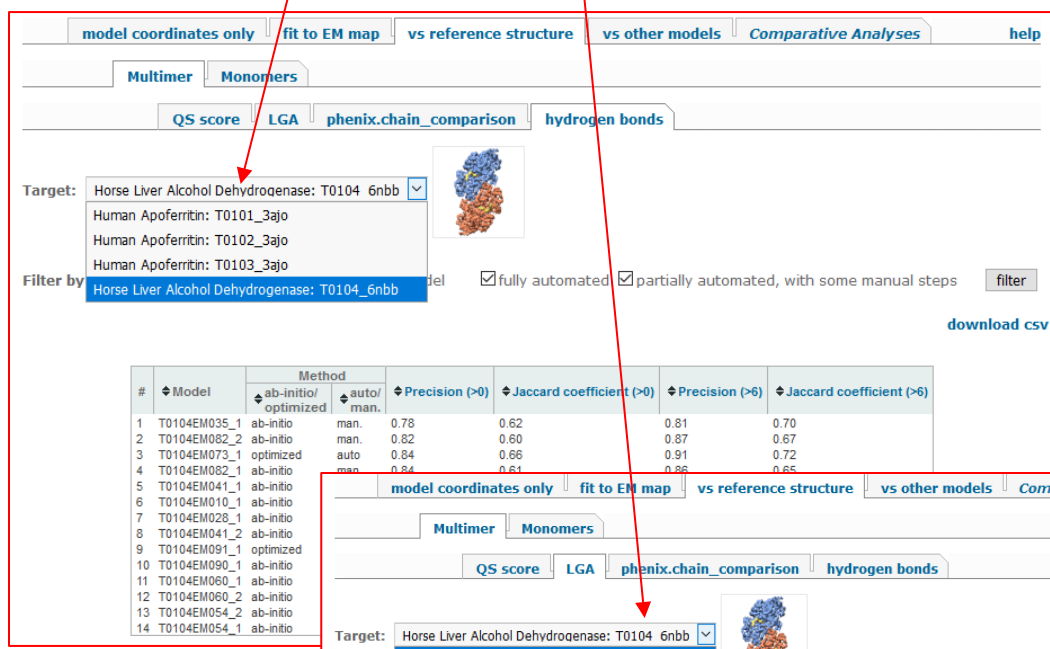
Jaccard coefficient (>6): the Jaccard coefficient measure (above) calculated on a subset of non-local hydrogen bonds (minimal sequence separation of six residues).

Guide to web infrastructure

The *vs reference structure* → *Multimers* tab provides evaluation results on similarity of whole submitted models to reference structures.

Please note that the term ‘multimers’ is used here in a wider sense meaning whole models and whole targets. If a target is monomeric and the evaluation results make sense for both multimeric and monomeric structures, then results would appear under the *Multimers* tab. This pertains to the *phenix.chain_comparison* and *hydrogen bonds* calculations. The ‘Target’ dropdown menu for these two analyses allows switching between all targets including multimers and monomers. However, if the results of multimeric calculations do not make sense in the monomeric context (e.g. QS score) or the results of multimeric calculations are the same as monomeric (LGA scores for monomers) then these results are not shown under the *Multimers* tab and the ‘Target’ dropdown menu contains only the list of multimeric targets.

vs reference structure → *Multimers* → *QS score* / *LGA* / *chain_comparison* / *Hbonds*



model coordinates only | fit to EM map | vs reference structure | vs other models | Comparative Analyses | help

Multimer | Monomers

QS score | LGA | phenix.chain_comparison | hydrogen bonds

Target: Horse Liver Alcohol Dehydrogenase: T0104_6nbb

Human Apoferritin: T0101_3ajo

Human Apoferritin: T0102_3ajo

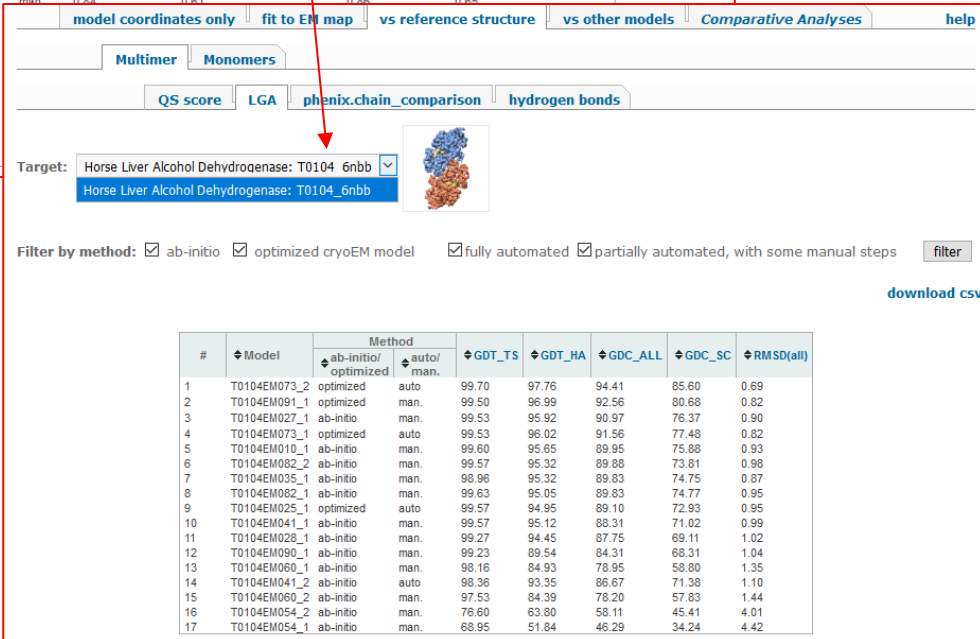
Human Apoferritin: T0103_3ajo

Filter by: Horse Liver Alcohol Dehydrogenase: T0104_6nbb

fully automated partially automated, with some manual steps

download csv

#	Model	Method		Precision (>0)	Jaccard coefficient (>0)	Precision (>6)	Jaccard coefficient (>6)
		ab-initio/optimized	auto/man.				
1	T0104EM035_1	ab-initio	man.	0.78	0.62	0.81	0.70
2	T0104EM082_2	ab-initio	man.	0.82	0.60	0.87	0.67
3	T0104EM073_1	optimized	auto	0.84	0.66	0.91	0.72
4	T0104EM082_1	ab-initio	man.	0.84	0.61	0.86	0.65
5	T0104EM041_1	ab-initio					
6	T0104EM010_1	ab-initio					
7	T0104EM028_1	ab-initio					
8	T0104EM041_2	ab-initio					
9	T0104EM091_1	optimized					
10	T0104EM090_1	ab-initio					
11	T0104EM060_1	ab-initio					
12	T0104EM060_2	ab-initio					
13	T0104EM054_2	ab-initio					
14	T0104EM054_1	ab-initio					



model coordinates only | fit to EM map | vs reference structure | vs other models | Comparative Analyses | help

Multimer | Monomers

QS score | LGA | phenix.chain_comparison | hydrogen bonds

Target: Horse Liver Alcohol Dehydrogenase: T0104_6nbb

Horse Liver Alcohol Dehydrogenase: T0104_6nbb

Filter by method: ab-initio optimized cryoEM model fully automated partially automated, with some manual steps

download csv

#	Model	Method		GDT_TS	GDT_HA	GDC_ALL	GDC_SC	RMSD(all)
		ab-initio/optimized	auto/man.					
1	T0104EM073_2	optimized	auto	99.70	97.76	94.41	85.60	0.69
2	T0104EM091_1	optimized	man.	99.50	96.99	92.56	80.68	0.82
3	T0104EM027_1	ab-initio	man.	99.53	95.92	90.97	76.37	0.90
4	T0104EM073_1	optimized	auto	99.53	96.02	91.56	77.48	0.82
5	T0104EM010_1	ab-initio	man.	99.60	95.65	89.95	75.88	0.93
6	T0104EM082_2	ab-initio	man.	99.57	95.32	89.88	73.81	0.98
7	T0104EM035_1	ab-initio	man.	98.96	95.32	89.83	74.75	0.87
8	T0104EM082_1	ab-initio	man.	99.63	95.05	89.83	74.77	0.95
9	T0104EM025_1	optimized	auto	99.57	94.95	89.10	72.93	0.95
10	T0104EM041_1	ab-initio	man.	99.57	95.12	88.31	71.02	0.99
11	T0104EM028_1	ab-initio	man.	99.27	94.45	87.75	69.11	1.02
12	T0104EM090_1	ab-initio	man.	99.23	89.54	84.31	68.31	1.04
13	T0104EM060_1	ab-initio	man.	98.16	84.93	78.95	58.80	1.35
14	T0104EM041_2	ab-initio	auto	98.36	93.35	86.67	71.38	1.10
15	T0104EM060_2	ab-initio	man.	97.53	84.39	78.20	57.83	1.44
16	T0104EM054_2	ab-initio	man.	76.60	63.80	58.11	45.41	4.01
17	T0104EM054_1	ab-initio	man.	68.95	51.84	46.29	34.24	4.42

Comparison versus other models

Overview

If several models are submitted on the same target, higher level of model conservancy (global and local) can be an indicator of higher model reliability (overall or per-residue). Evaluation in this regime includes calculation of pairwise GDT_TS scores (all-against-all models) and inter-residue distances (as reported in the LGA's optimal superposition).

Measures

Davis_QA score¹⁷ estimates accuracy of a model based on its similarity to other models submitted on the target. The method superimposes all models pairwise by running LGA in the sequence dependent mode. For each model, the quality score is calculated by averaging the GDT_TS scores from all pairwise comparisons. In the local mode, per-residue scores are obtained by averaging the S-function-transformed distances²¹ between the corresponding residues in pairwise LGA superpositions of the selected model with the other models submitted on the target.

Guide to web infrastructure

The *vs other models* tab provides global **DAVIS QA score** and per-residue similarity scores in the form of color-coded bars. Clicking on a data bar shows structural superposition of the model and the target colored according to the underlying bar.



Comparative analyses

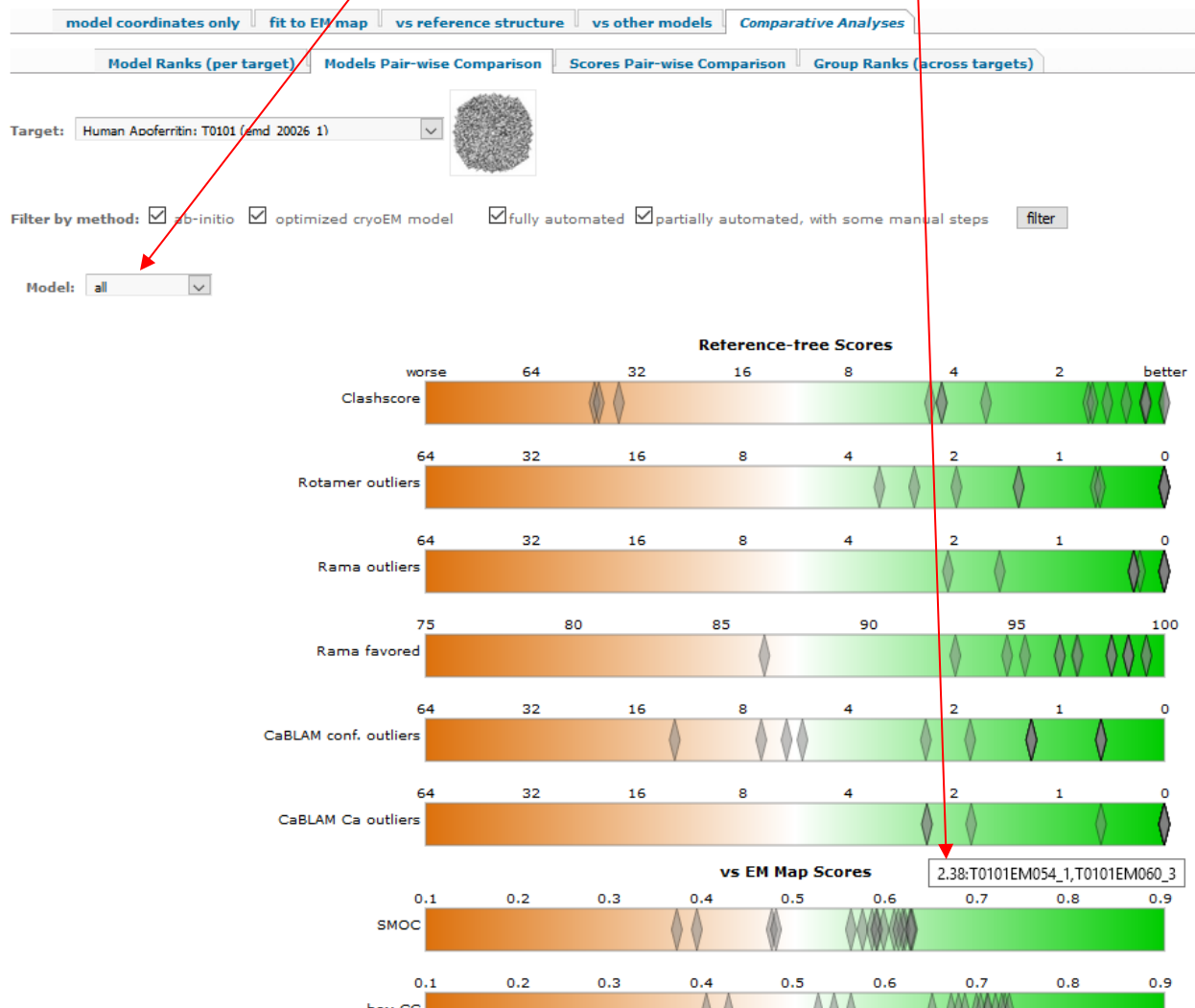
Overview

This tab provides a web resource for comparing methods, models and scores from different evaluation tracks.

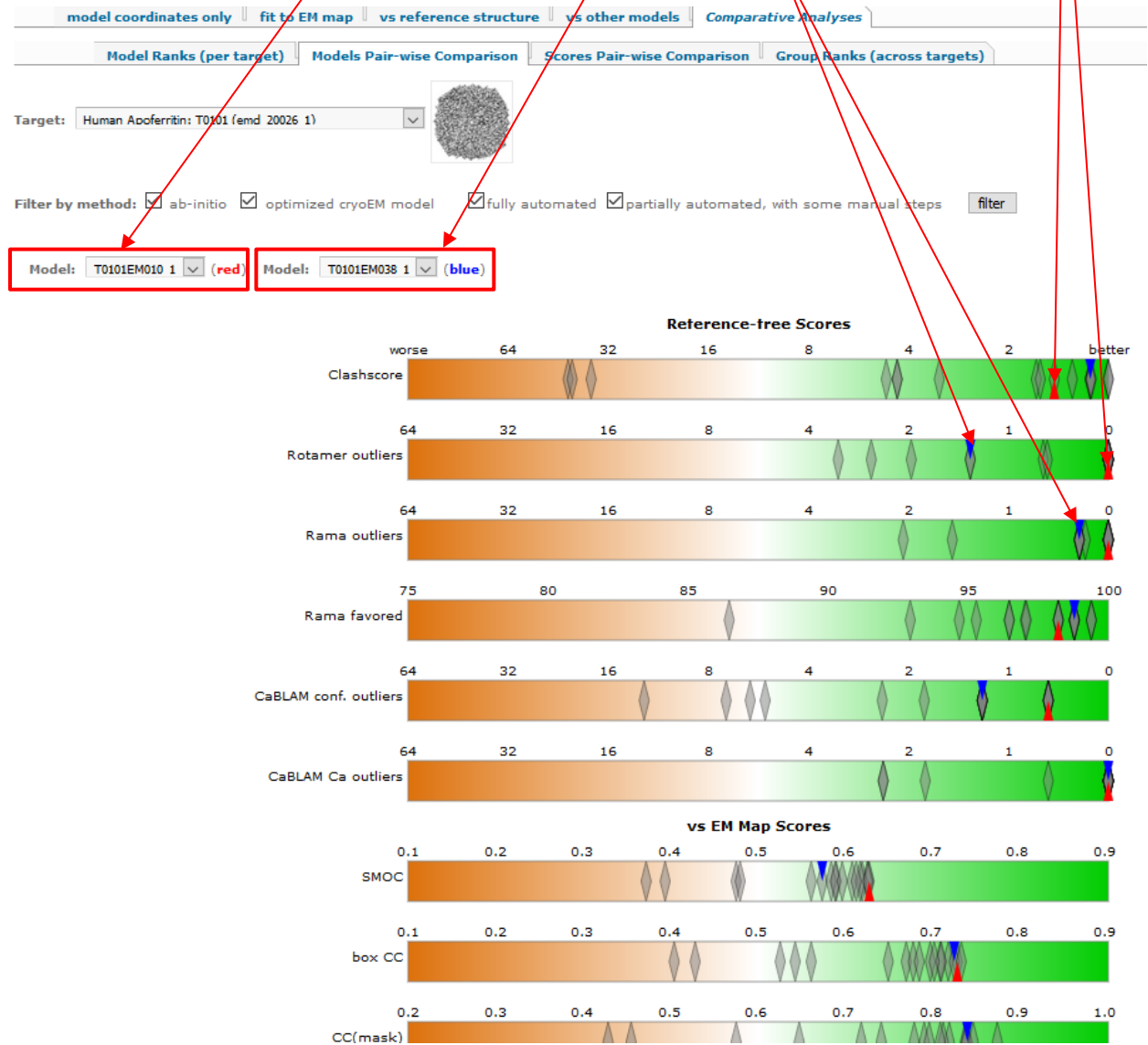
Guide to web infrastructure

Comparative analyses → Models Pair-wise Comparison

Visual comparison of a models' accuracy according to different evaluation scores can be performed using the slider tool. The sliders enable visualization of all major scores for all models submitted to a single target in one page. The default view has 'all' models selected in the 'Model' dropdown menu. Scores are shown as semi-transparent grey diamonds. Overlapping diamonds make grey color more intense. Hovering over a diamond identifies model(s) and shows the score value (if scores are identical, all models are listed). Scores in different evaluation tracks are grouped together (e.g., 'Reference-free scores' or 'vs EM Map Scores').



Selecting a model from the *Model* dropdown menu marks scores for the model with red triangles in the lower halves of all slider bars. At the same time, a second *Model* menu appears allowing selection of a different model and comparison of the two sets of scores. Scores for the second model are marked in the upper halves of the slider bars by blue triangles.



Comparative analyses → Scores Pair-wise Comparison → Scatter Plots

This page allows pairwise comparisons of different per-target scores, so that one can evaluate correlation (or lack thereof) for any pair of scores. Six panels are shown. By default, these panels show relation between six different scores (y axis) and the PHENIX's Box_CC score (x-axis). Scatter plot for the desired pair of scores can be drawn by selecting the scores from the dropdown menu beneath each panel and clicking on the "Redraw all plots with updated XY" button. Each point in the plot represents a model. Model scores and names can be identified by hovering the mouse over the point of interest. Each graph has a separate menu that appears after placing the mouse in the plot area. Meaning of symbols in the menu are explained with the hovering mouse. Graphs also can be zoomed in and out by selecting a rectangular area in the plot (click on the desired corner of the area and drag the mouse to the opposite corner).



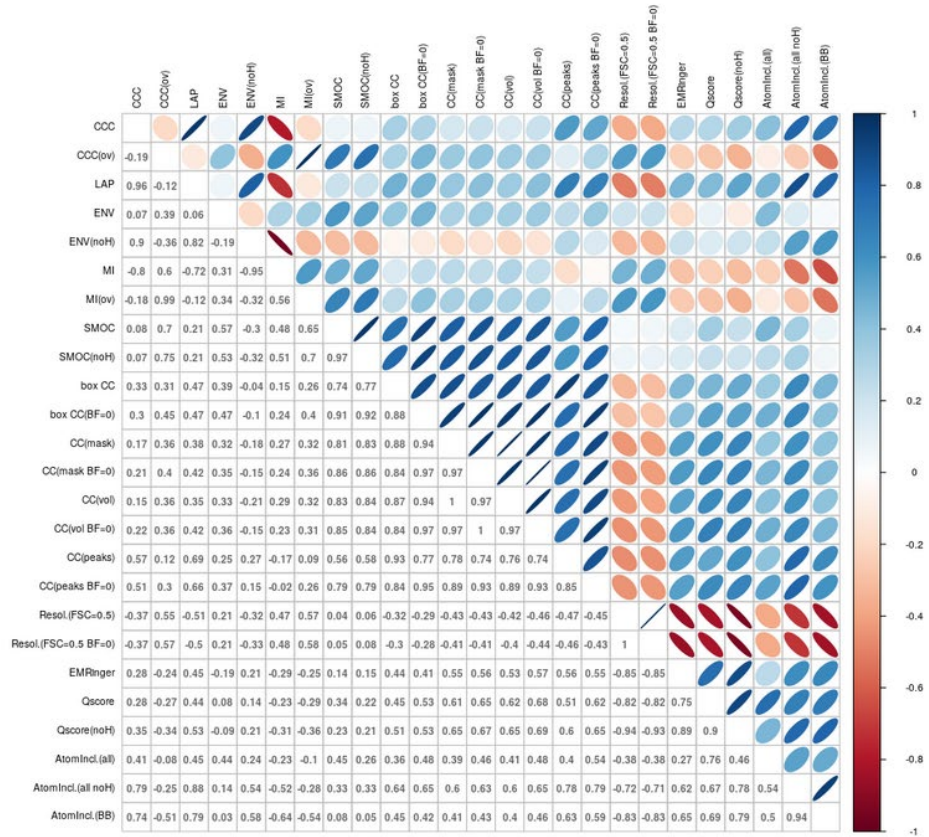
Comparative analyses → Scores Pair-wise Comparison → Correlation Summary

The *Correlation Summary* page shows Pearson's correlation coefficients between different evaluation scores calculated on all submitted models for all targets.

The *Comparative analyses* → *Scores Pair-wise Comparison* → *Correlation Summary* → *Selected scores* tab shows correlation coefficients between preselected scores in four different evaluation tracks. Scores within each evaluation track are marked with the black squares. The **DAVIS QA score** (a *vs other models* score) is placed in the 'vs reference structure – Monomers' block.



Four evaluation-track specific subtabs of the *Comparative analyses* → *Scores Pair-wise Comparison* → *Correlation Summary* page (i.e., *Geometry Scores*, *fit to EM map*, *vs reference structure (multi)*, and *vs reference structure (mono)*) show correlation coefficients for all evaluation scores within the selected evaluation track. The correlation table for the ‘*fit to EM map*’ track is shown as an example below.



Comparative analyses → Model ranks (per target)

Page under development.

This page enables ranking of models on each target according to the user-selected combination of measures.

For each target, the original (raw) scores are transferred into the distribution-normalized z-scores (standard scores). Values of the z-scores depend on the raw score and the mean and standard deviation of the target's score population:

$$z_score(model) = \frac{raw_score(model) - Mean}{StandardDeviation}.$$

A z-score shows relative accuracy of a model with respect to other models submitted on the target. Z-scores can take any values and are dimensionless so that they can be combined with desired weights. This way each model can be assigned a cumulative ranking score in separate assessment tracks (e.g. for model-to-map fit) or for a combination of the assessment tracks.

Comparative analyses → Group ranks (across targets)

Page under development.

This page enables ranking of prediction methods on all attempted targets according to the user-selected combination of measures. Per-target z-scores (see above) from a group (modeling method) are summed or averaged and ranked accordingly.

References:

1. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Echols N, Headd JJ, Hung LW, Jain S, Kapral GJ, Grosse Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner RD, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. The Phenix software for automated determination of macromolecular structures. *Methods* 2011;55(1):94-106.
2. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 2010;66(Pt 1):12-21.
3. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, 3rd, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. MolProbity: More and better reference data for improved all-atom structure validation. *Protein science : a publication of the Protein Society* 2018;27(1):293-315.
4. Ray A, Lindahl E, Wallner B. Improved model quality assessment using ProQ2. *BMC Bioinformatics* 2012;13:224.
5. Farabella I, Vasishtan D, Joseph AP, Pandurangan AP, Sahota H, Topf M. : a Python library for assessment of three-dimensional electron microscopy density fits. *J Appl Crystallogr* 2015;48(Pt 4):1314-1323.
6. Vasishtan D, Topf M. Scoring functions for cryoEM density fitting. *J Struct Biol* 2011;174(2):333-343.
7. Joseph AP, Malhotra S, Burnley T, Wood C, Clare DK, Winn M, Topf M. Refinement of atomic models in high resolution EM reconstructions using Flex-EM and local assessment. *Methods* 2016;100:42-49.
8. Joseph AP, Lagerstedt I, Patwardhan A, Topf M, Winn M. Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. *Journal of structural biology* 2017;199(1):12-26.
9. Barad BA, Echols N, Wang RY, Cheng Y, DiMaio F, Adams PD, Fraser JS. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* 2015;12(10):943-946.
10. Pintilie G, Chiu W. Assessment of structural features in Cryo-EM density maps using SSE and side chain Z-scores. *Journal of structural biology* 2018;204(3):564-571.
11. Lagerstedt I, Moore WJ, Patwardhan A, Sanz-Garcia E, Best C, Swedlow JR, Kleywegt GJ. Web-based visualisation and analysis of 3D electron-microscopy data from EMDB and PDB. *Journal of structural biology* 2013;184(2):173-181.
12. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep* 2017;7(1):10480.
13. Zemla A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res* 2003;31(13):3370-3374.
14. Mariani V, Biasini M, Barbato A, Schwede T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 2013;29(21):2722-2728.
15. Olechnovic K, Kulberkyte E, Venclovas C. CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins* 2013;81(1):149-162.
16. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. *Journal of molecular biology* 1994;238(5):777-793.
17. Kryshtafovych A, Monastyrskyy B, Fidelis K. CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL. *Proteins* 2014;82 Suppl 2:7-13.

18. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 2004;60(Pt 12 Pt 1):2184-2195.
19. Uziela K, Shu N, Wallner B, Elofsson A. ProQ3: Improved model quality assessments using Rosetta energy terms. *Sci Rep* 2016;6:33509.
20. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, Labonte JW, Pacella MS, Bonneau R, Bradley P, Dunbrack RL, Jr., Das R, Baker D, Kuhlman B, Kortemme T, Gray JJ. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* 2017;13(6):3031-3048.
21. Gerstein M, Levitt M. Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci* 1998;7(2):445-456.
22. Afonine PV, Klaholz BP, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, Adams PD, Urzhumtsev A. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta crystallographica Section D, Structural biology* 2018;74(Pt 9):814-840.
23. Zemla A, Venclovas, Moutl J, Fidelis K. Processing and evaluation of predictions in CASP4. *Proteins* 2001;Suppl 5:13-21.
24. Keedy DA, Williams CJ, Headd JJ, Arendall WB, 3rd, Chen VB, Kapral GJ, Gillespie RA, Block JN, Zemla A, Richardson DC, Richardson JS. The other 90% of the protein: assessment beyond the Calphas for CASP8 template-based and high-accuracy models. *Proteins* 2009;77 Suppl 9:29-49.
25. Olechnovic K, Monastyrskyy B, Kryshchak A, Venclovas C. Comparative analysis of methods for evaluation of protein models against native structures. *Bioinformatics* 2018.