High Accuracy Assessment in CASP14

Andrei N. Lupas
Joana Pereira
Marcus D. Hartmann

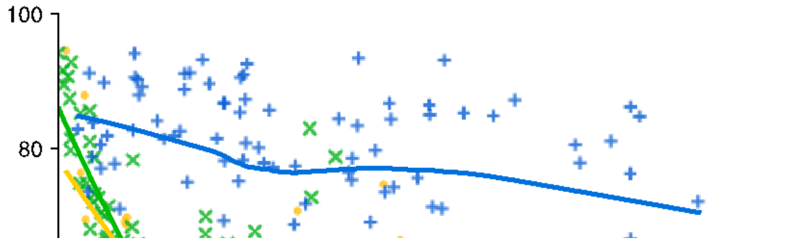
Max Planck Institute for Developmental Biology

High accuracy in CASP14

In 2020, “high accuracy” gained a new meaning…

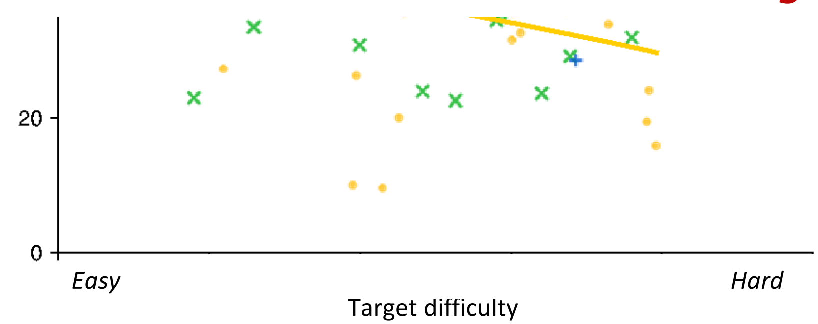
… mostly due to the performance of a single group

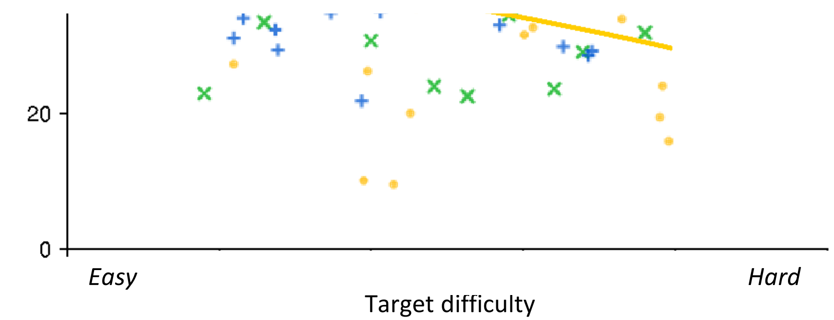
Best model



Excluding the 1st ranking group

ALL targets are now “high accuracy” targets and we analyzed all of them!





Nevertheless there was a significant improvement on model accuracy by all other groups!

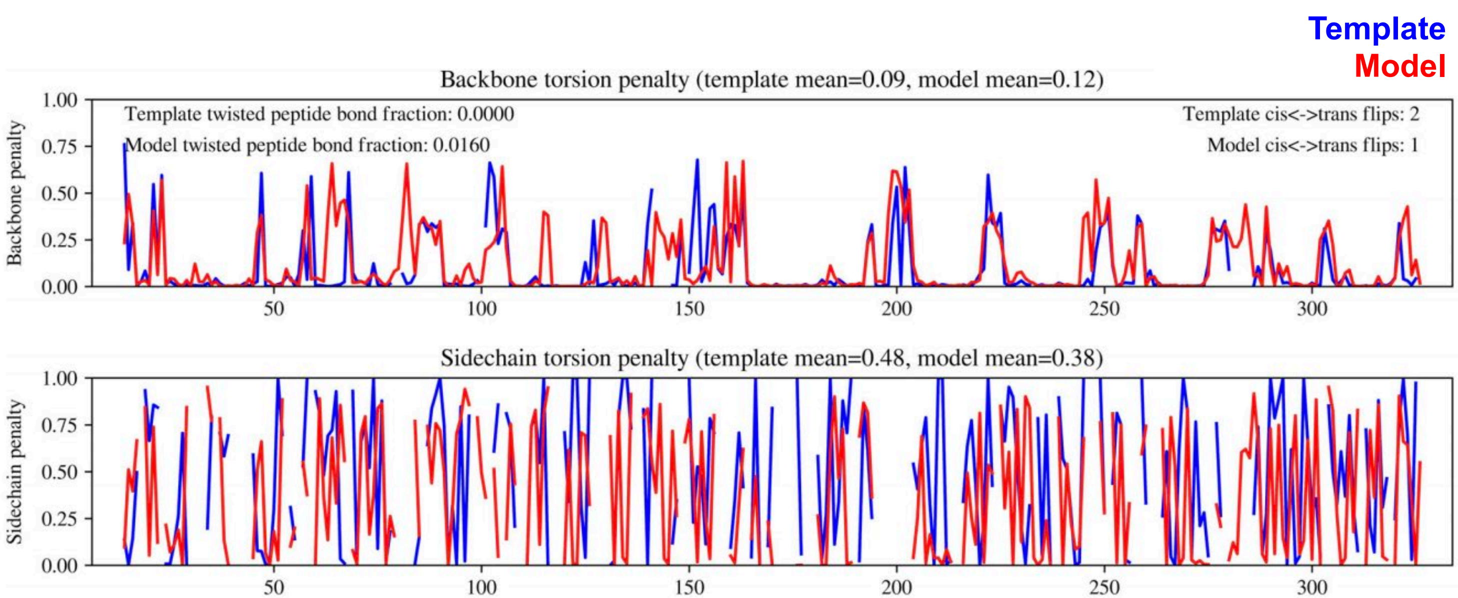
How was it done in CASP13?

Croll *et al.* Proteins (2019)

In CASP13, the high accuracy assessing team used a combination of overall and local metrics:

$$S_{CASP13} = \frac{1}{16}(Z_{IDDT} + Z_{CADaa} + Z_{SG} + Z_{sidechain}) + \frac{1}{8}(Z_{MolPrb-clash} + Z_{backbone}) + \frac{1}{4}(Z_{GDT-HA} + Z_{ASE})$$

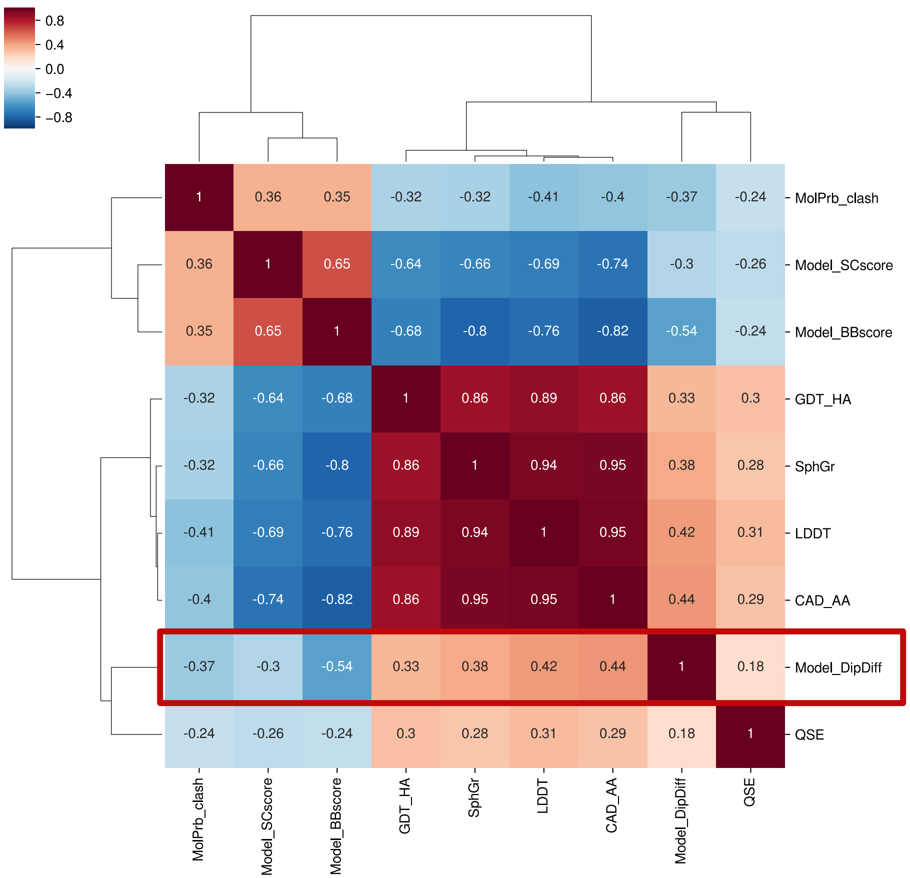
How much does a model's **backbone dihedral and sidechain chi angles** deviate from those in the target?



How did we do it in CASP14?

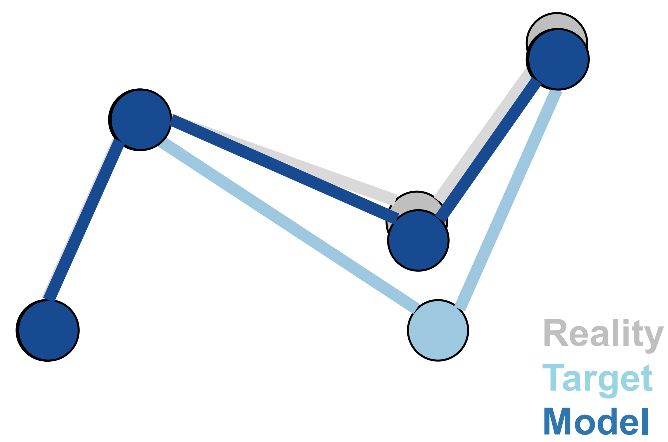
In CASP14, we decided to not “re-invent the wheel” and used the same function, adding only one more metric:

$$S_{CASP14} = \frac{1}{16}(Z_{lDDT} + Z_{CADaa} + Z_{SG} + Z_{SCscore}) + \frac{1}{12}(Z_{MolPrb-clash} + Z_{BBscore} + Z_{DipDiff}) + \frac{1}{4}(Z_{GDT-HA} + Z_{ASE})$$



Are the models actually better than the targets?

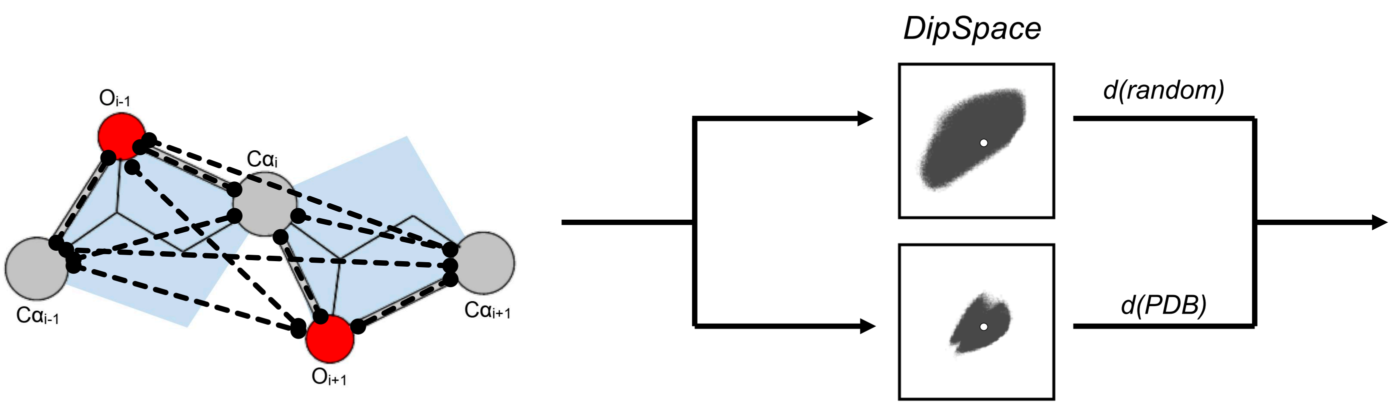
A backbone, superposition-free metric that is orthogonal to any of the other



DipDiff

Pereira and Lamzin, IUCrJ (2017)

DipDiff is based on the **DipScore**, which measures the likelihood of the backbone distances around a given Cα to be correct :



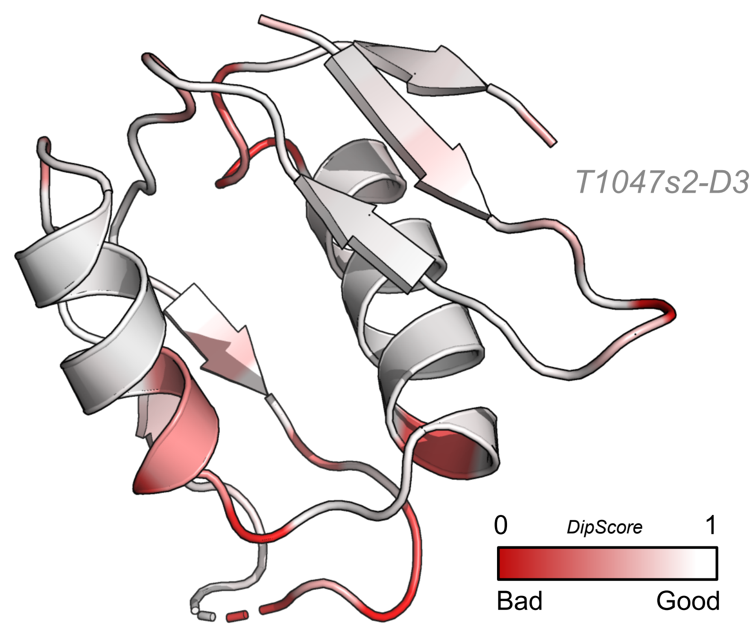
DipScore

$$\frac{d(PDB)}{d(PDB) + d(random)}$$

Given a protein backbone, each Cα can be assigned a DipScore

Useful for:

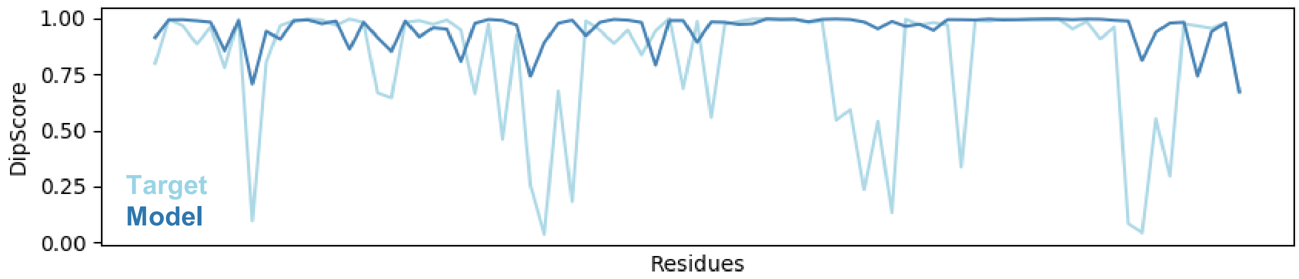
- Automatic backbone tracing of electron density maps
- Detection of strained residues
- **Protein backbone validation**

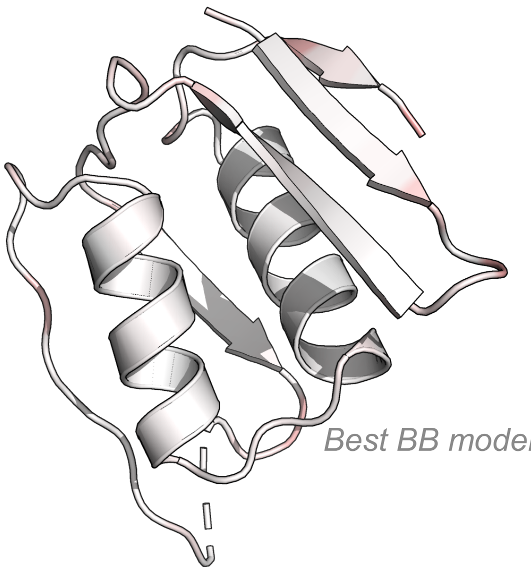


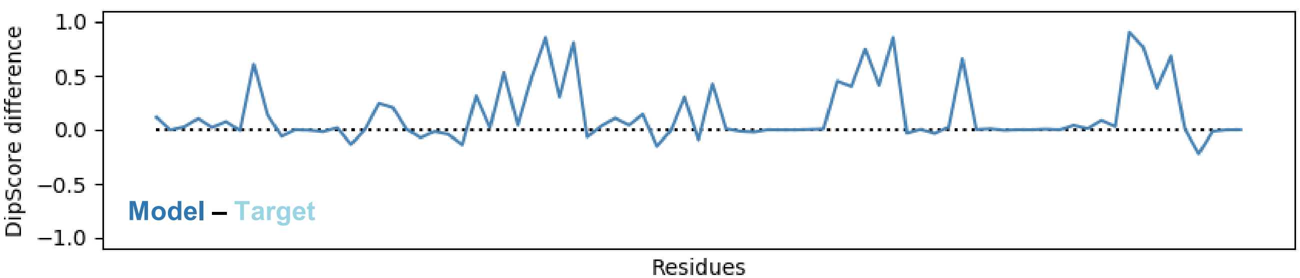
DipDiff

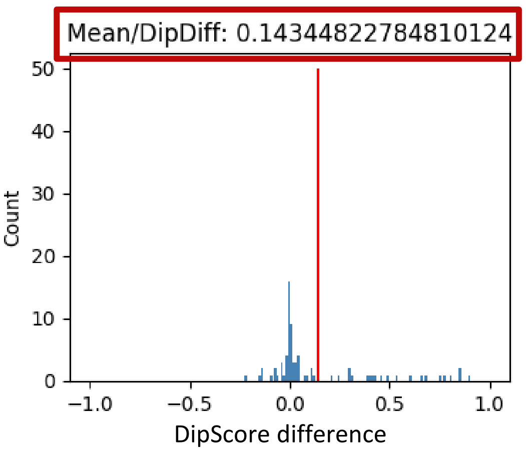
DipDiff measures how well does the backbone distance-based geometry of the models fit into that of the target:

The **DipScore** measures the likelihood of the backbone distances around a given Cα to be correct







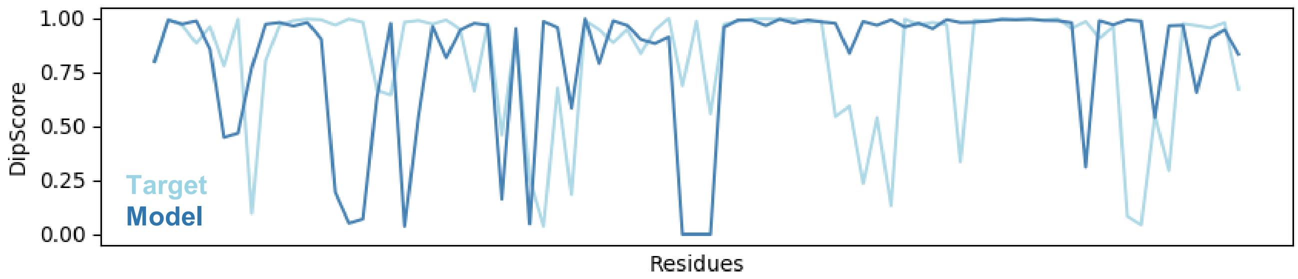


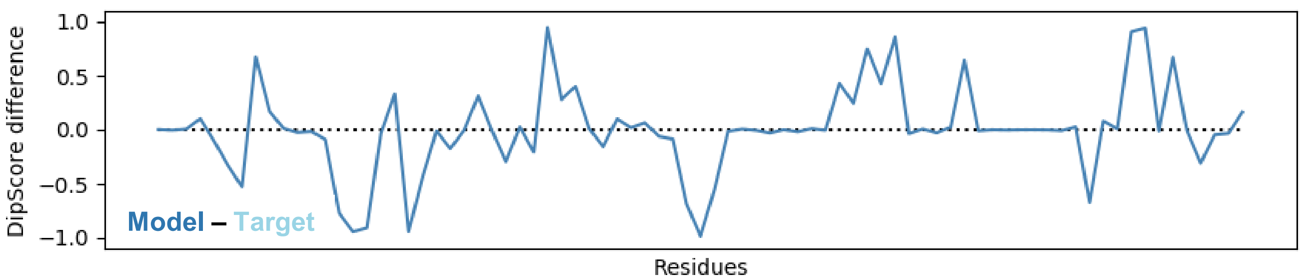
The **DipDiff** is the average difference between the local DipScores of the target and a given model

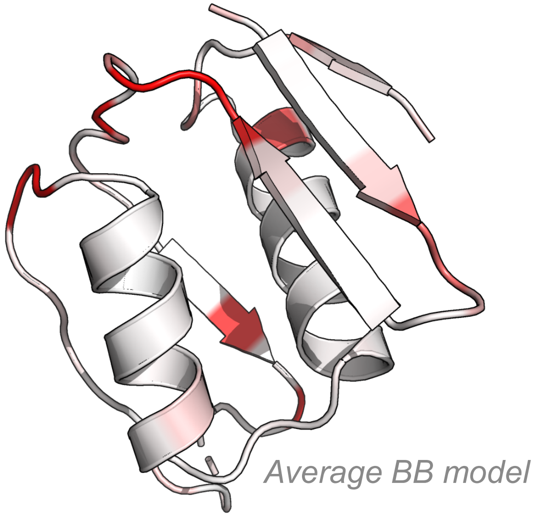
DipDiff

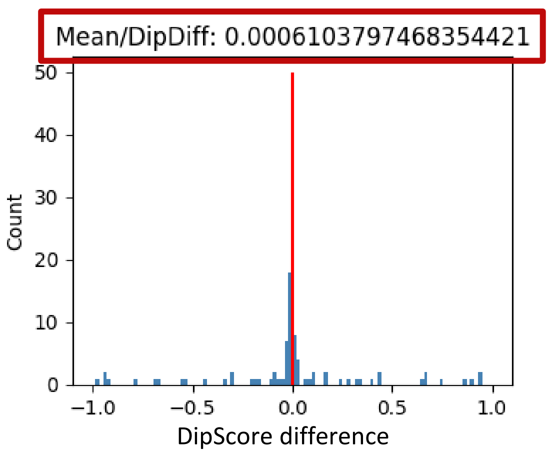
DipDiff measures how well does the backbone distance-based geometry of the models fit into that of the target:

The **DipScore** measures the likelihood of the backbone distances around a given Cα to be correct





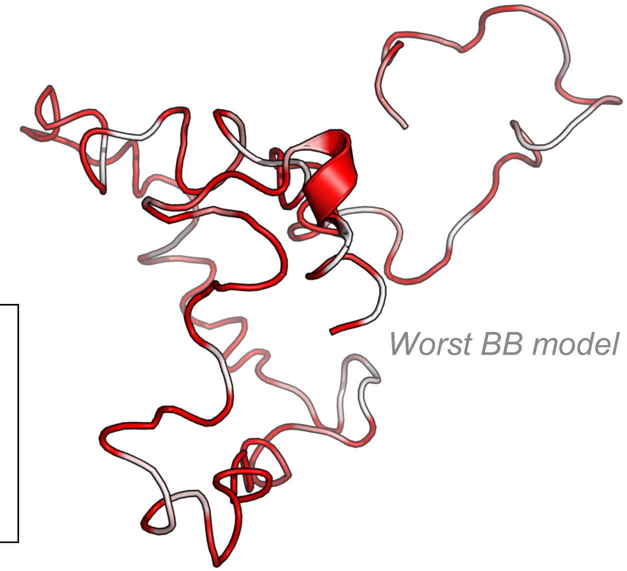




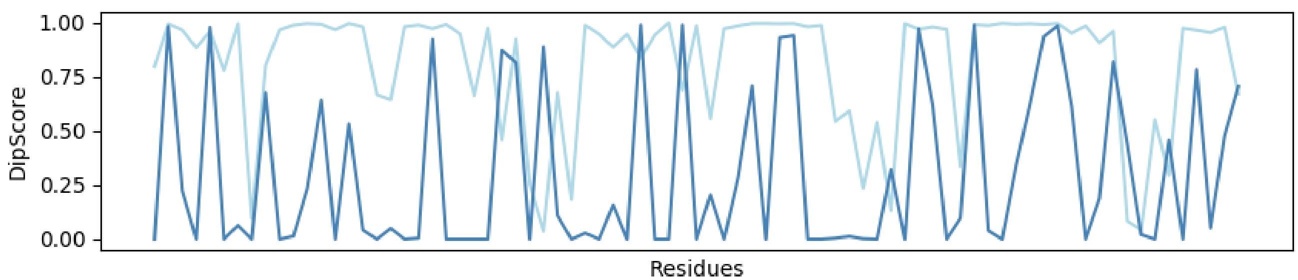
The **DipDiff** is the average difference between the local DipScores of the target and a given model

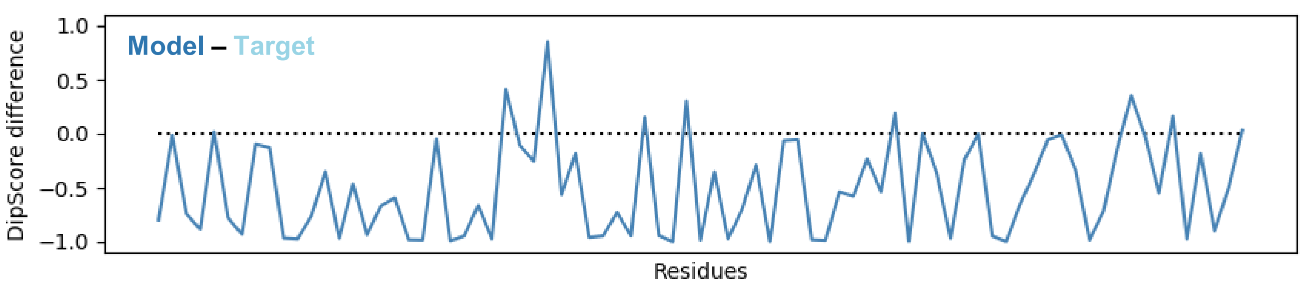
DipDiff

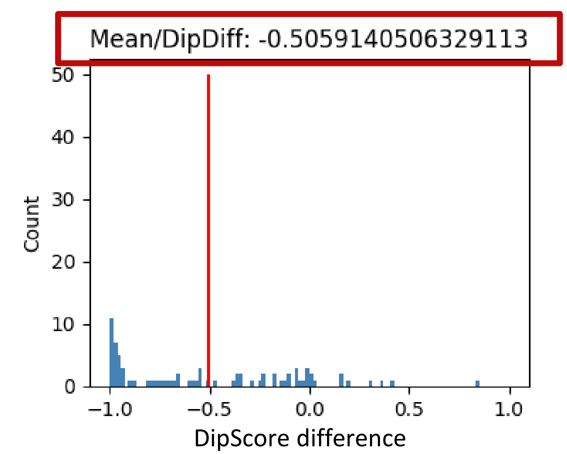
DipDiff measures how well does the backbone distance-based geometry of the models fit into that of the target:



The **DipScore** measures the likelihood of the backbone distances around a given Cα to be correct







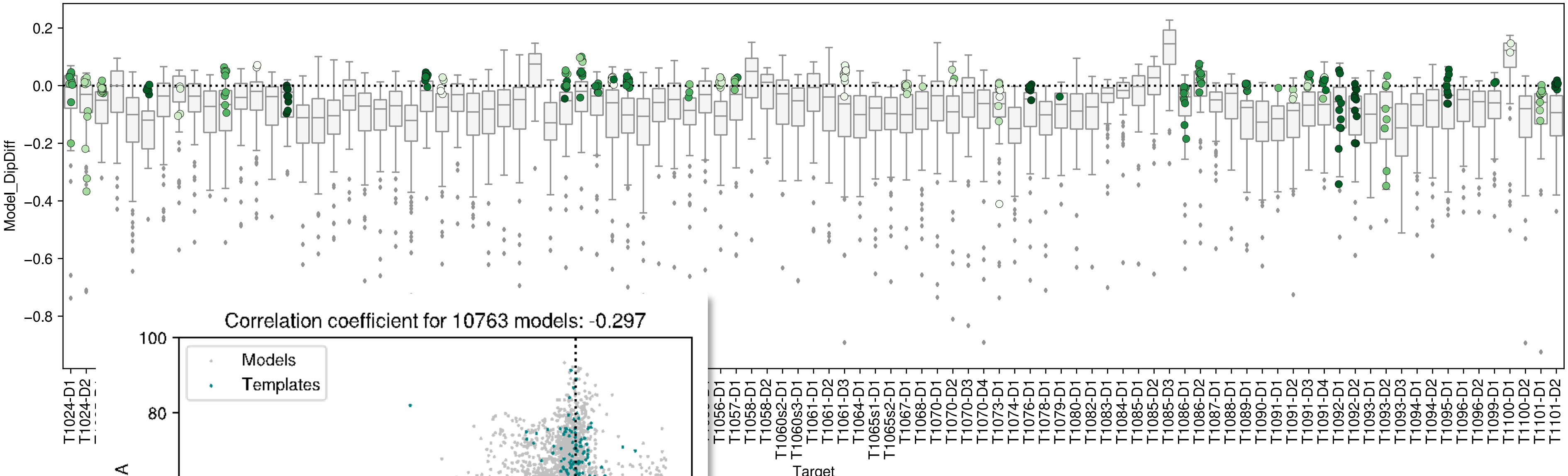
The **DipDiff** is the average difference between the local DipScores of the target and a given model

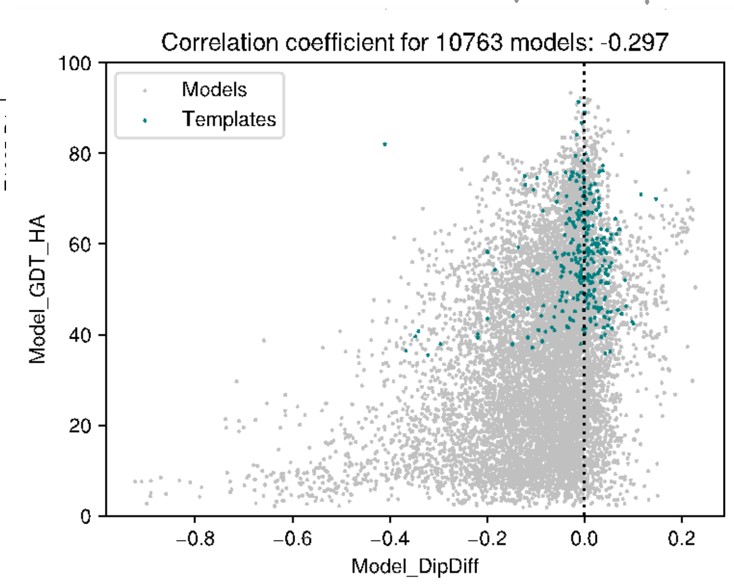
Backbone geometry in CASP14

Most models have a **worst** backbone geometry than the target…

… but some targets seem to always be modelled with a better backbone geometry!

*DipDiff distribution of groups 1st model per target, compared to **selected templates***





A decrease in GDT_HA may be due to an improvement in protein backbone quality…

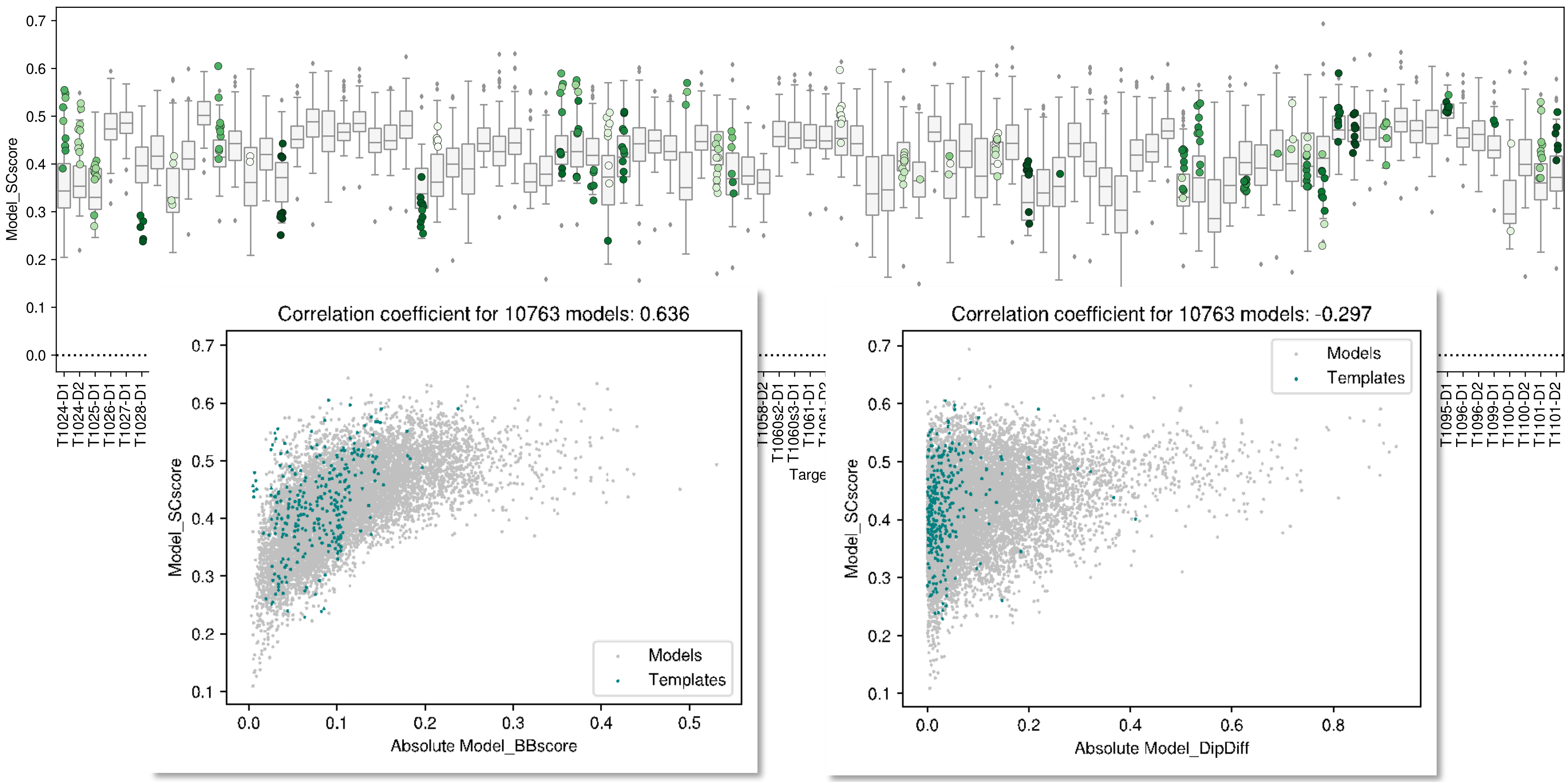
But in CASP14 this is most often not the case!

Sidechain geometry in CASP14

As expected, sidechains are **harder to build correctly**…

… even when there is a good template available.

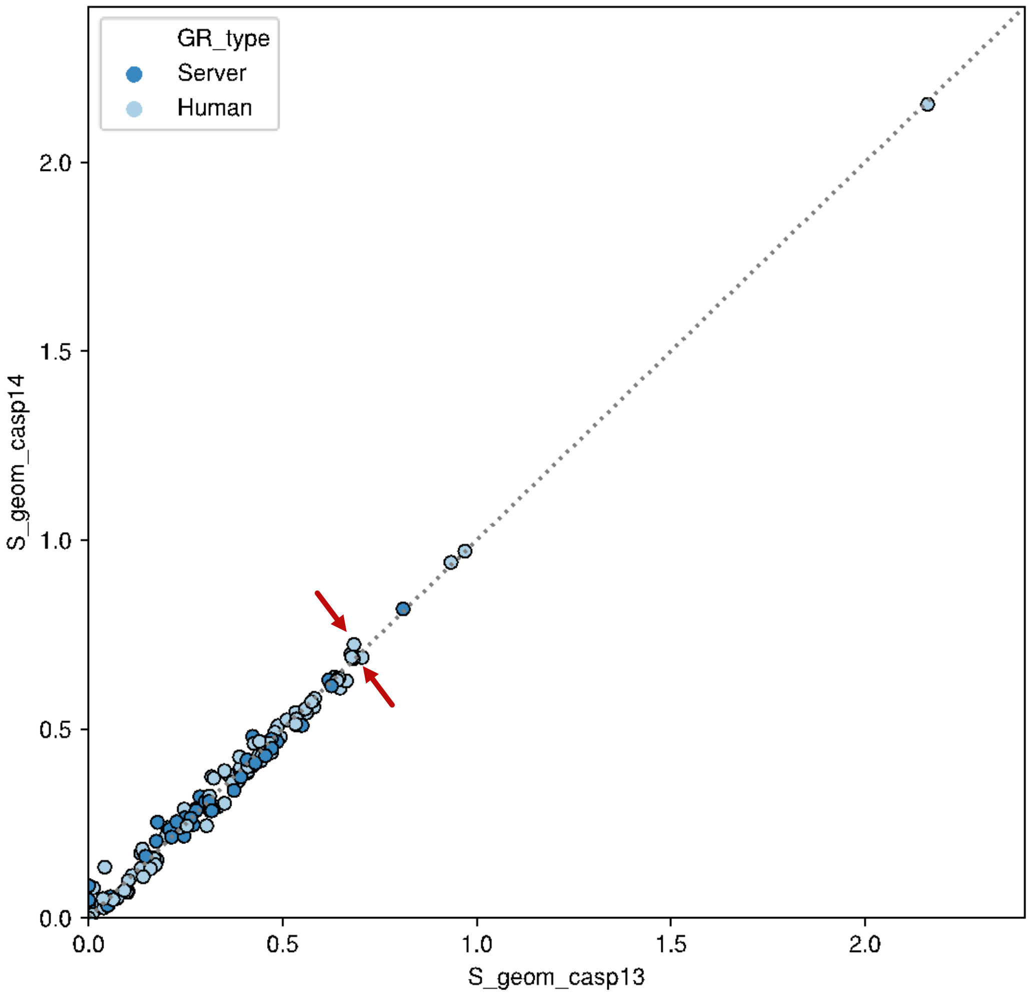
*SCscore distribution of groups 1st model per target, compared to **selected templates***

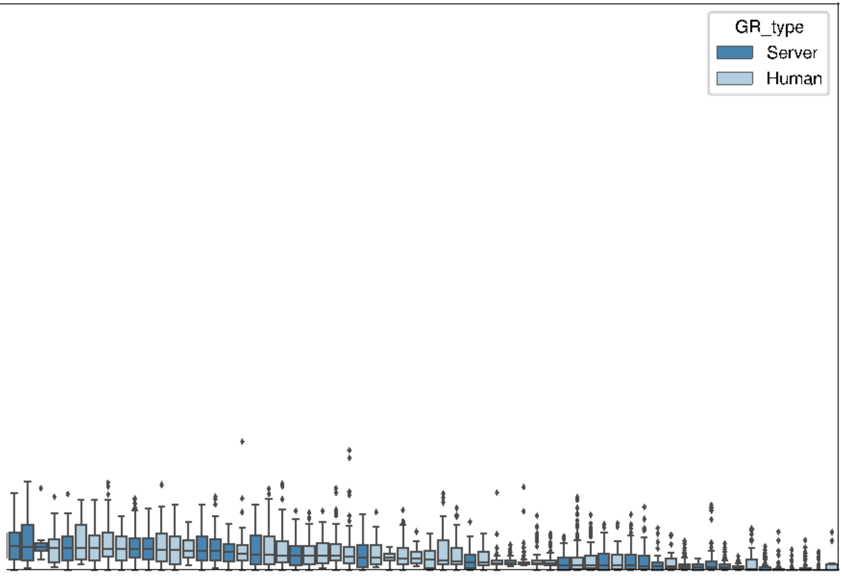


Overall group ranking

When we rank the groups based on the median of our CASP14 scoring function for their 1st models, there is a clear leader:

*Group ranking based on the **median CASP14 score of the 1st model** (only for groups that submitted models for at least 10 targets)*

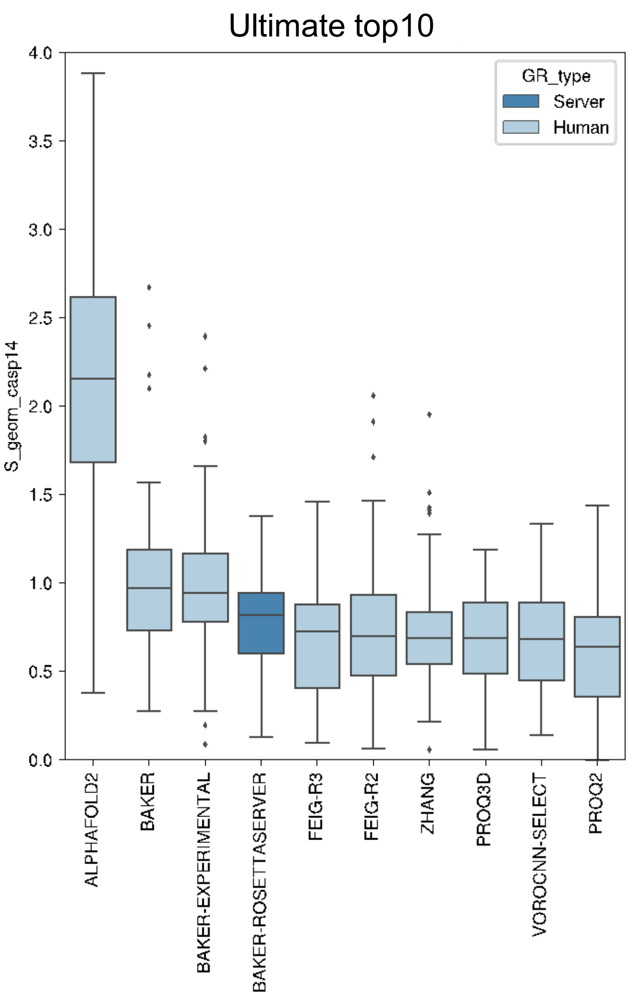




The CASP14 scoring function allowed for a **better resolution** in the middle of the ranking

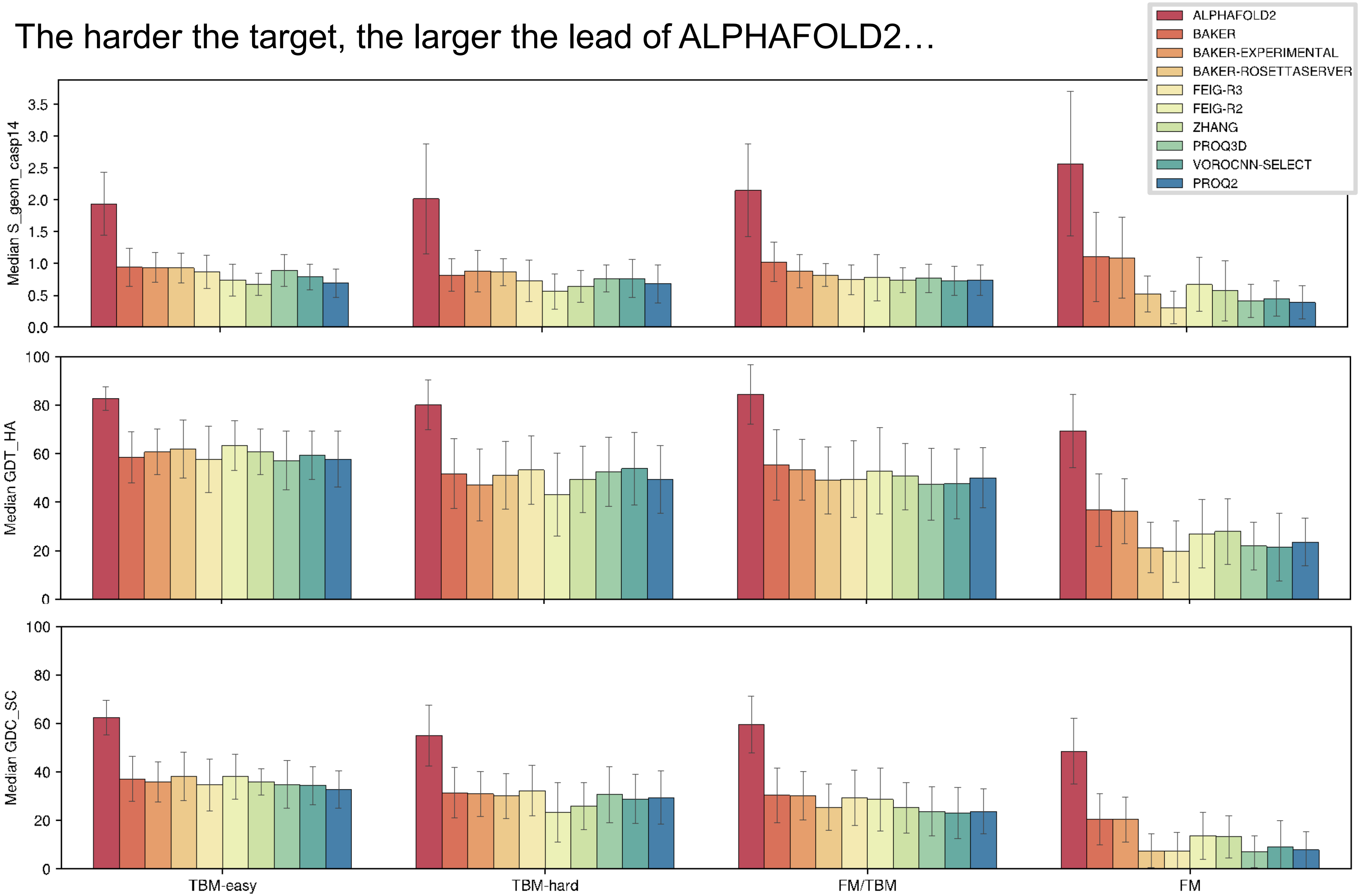
Overall group ranking

When we rank the groups based on the median of our CASP14 scoring function for their 1st models, there is a clear leader:



Overall group ranking

The harder the target, the larger the lead of ALPHAFOLD2…



How useful are the models?

How useful are the models?

Evaluating the usability of models for Molecular Replacement for

49 evaluation units (T????-D?) from
30 targets (T????)

- Using **phaser LLG scripts** provided by **Randy Read / Gabor Bunkoczi**

  - for all models submitted by all groups, done for
    - all **49 evaluation units**
    - all **30 full targets**

- Using **AMPLE pipeline**, performed by the **Dan Rigden** group: **Adam Simpkin, Ronan Keegan**

  - for Alphafold2 submissions for
    - all **30 full targets**

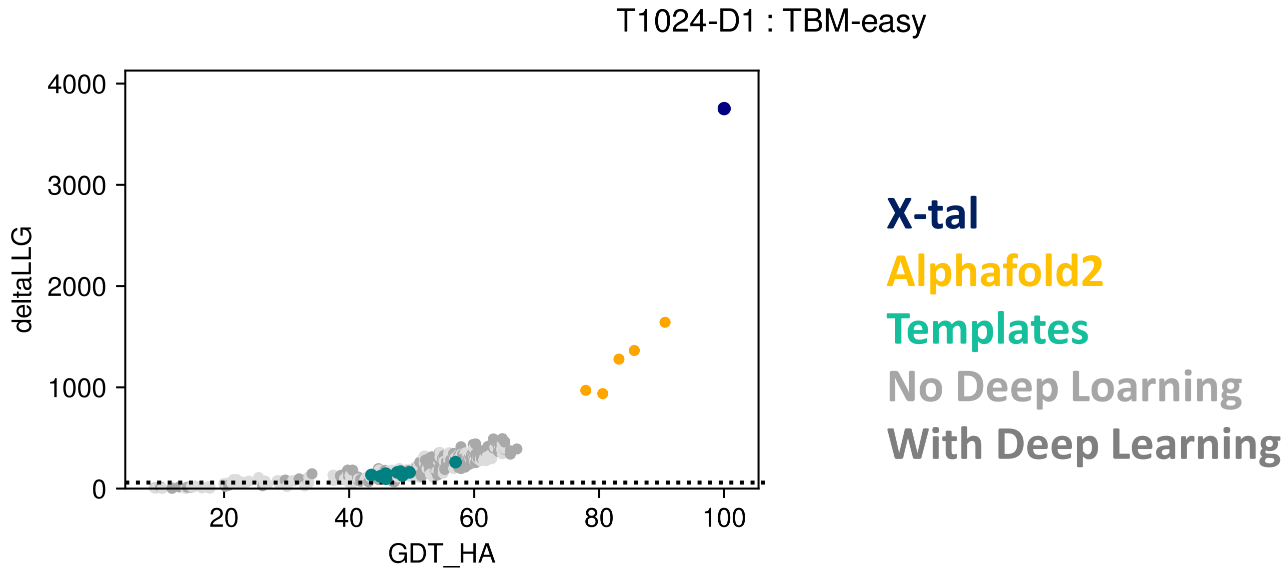
  - for test best 20 non-Alphafold2 models in the LLG analysis for
    - **T1058**, **T1089**, **T1100** (full targets)

How useful are the models?

Evaluating the usability of models for Molecular Replacement for

49 evaluation units (T????-D?)

- Using **phaser LLG scripts** for **T1024** (TBM-easy)

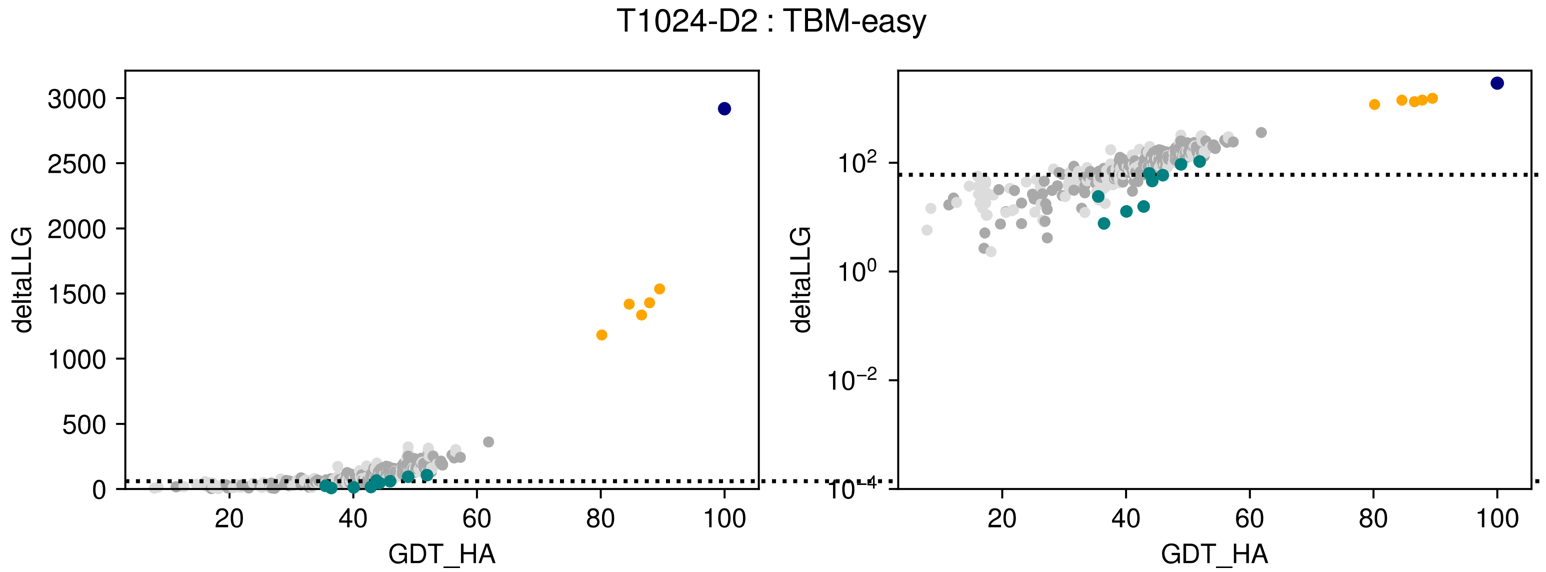


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

49 evaluation units (T????-D?)

- Using **phaser LLG scripts** for **T1024** (TBM-easy)

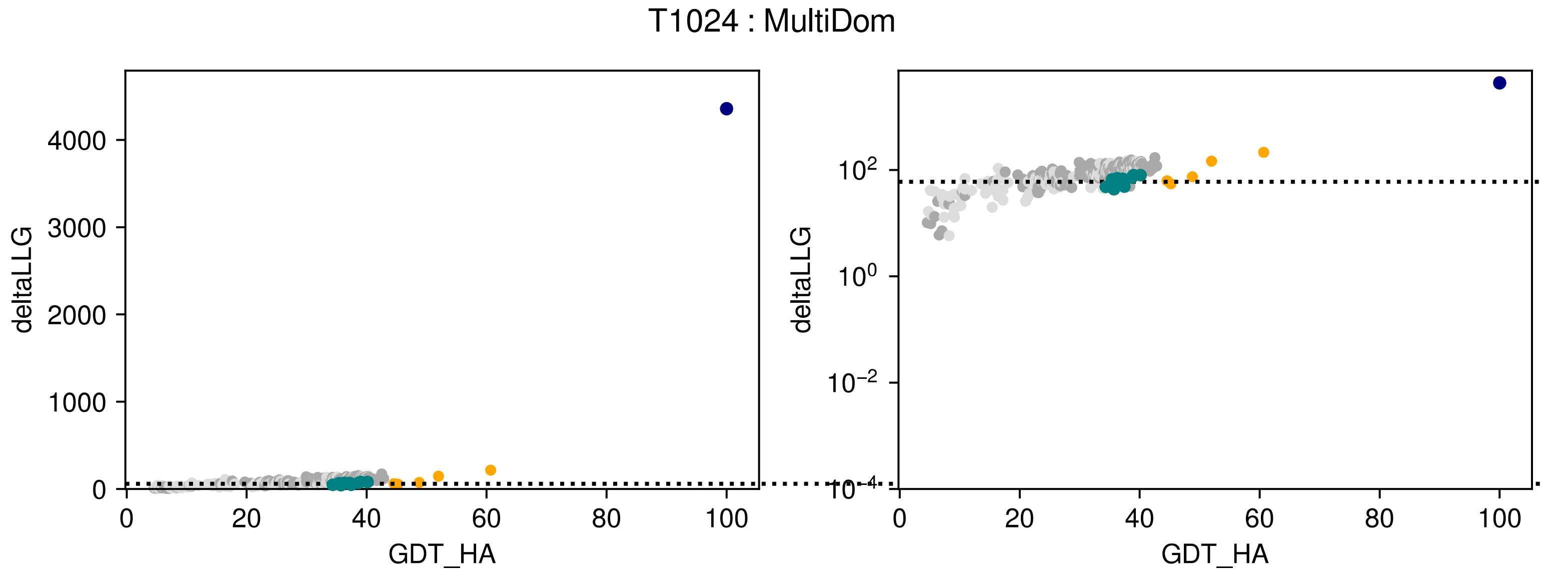


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

49 evaluation units (T????-D?)

- Using **phaser LLG scripts** for **T1024** (TBM-easy)

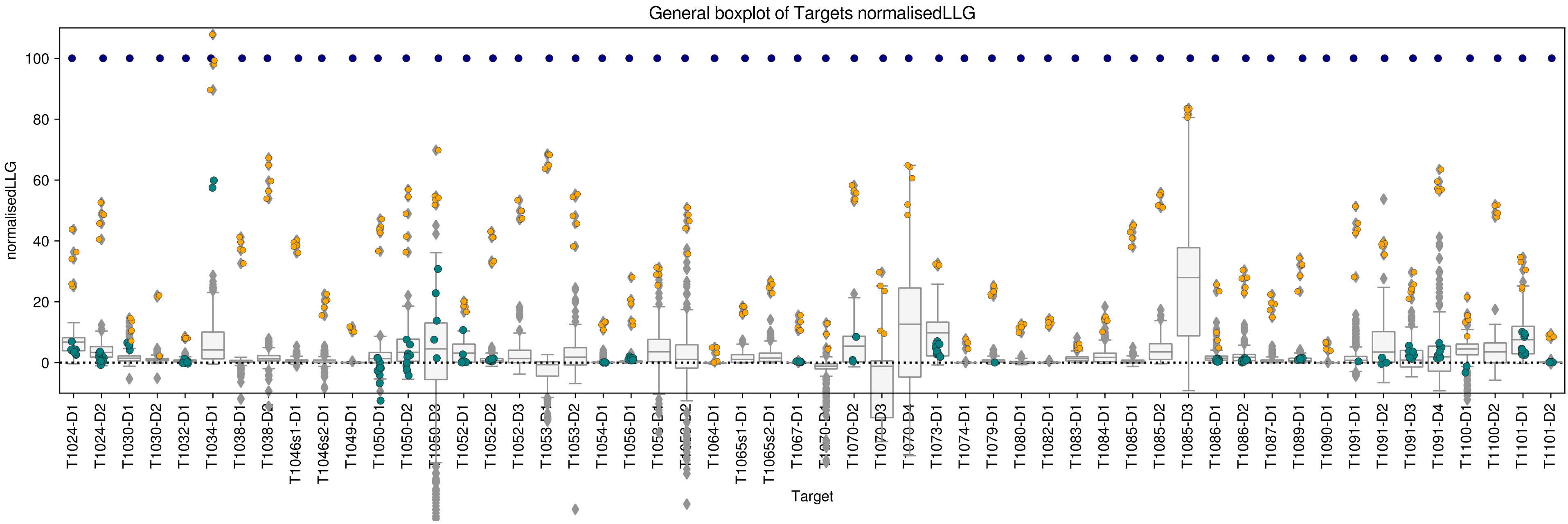


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

49 evaluation units (T????-D?)

- Overview of 49 evaluation units based on normalized LLG

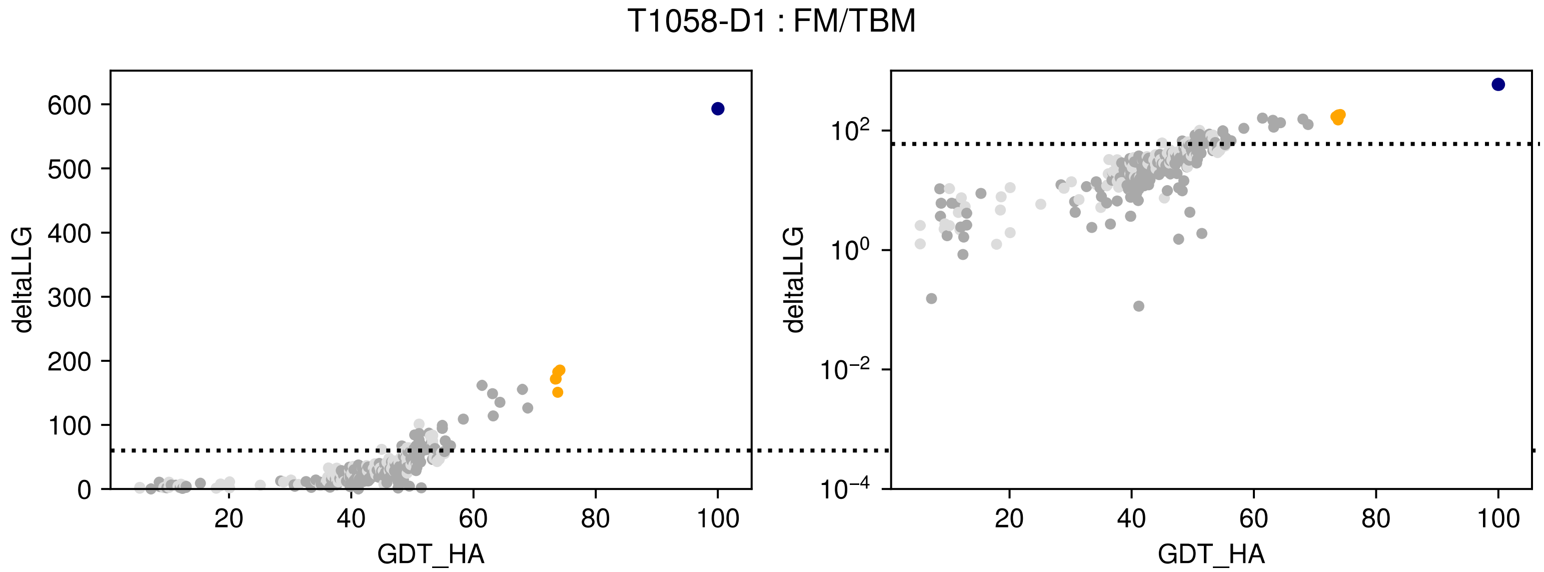


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

Three structures that were solved using Aplhafold2 models

- **T1058** (FM/TBM), two evaluation units

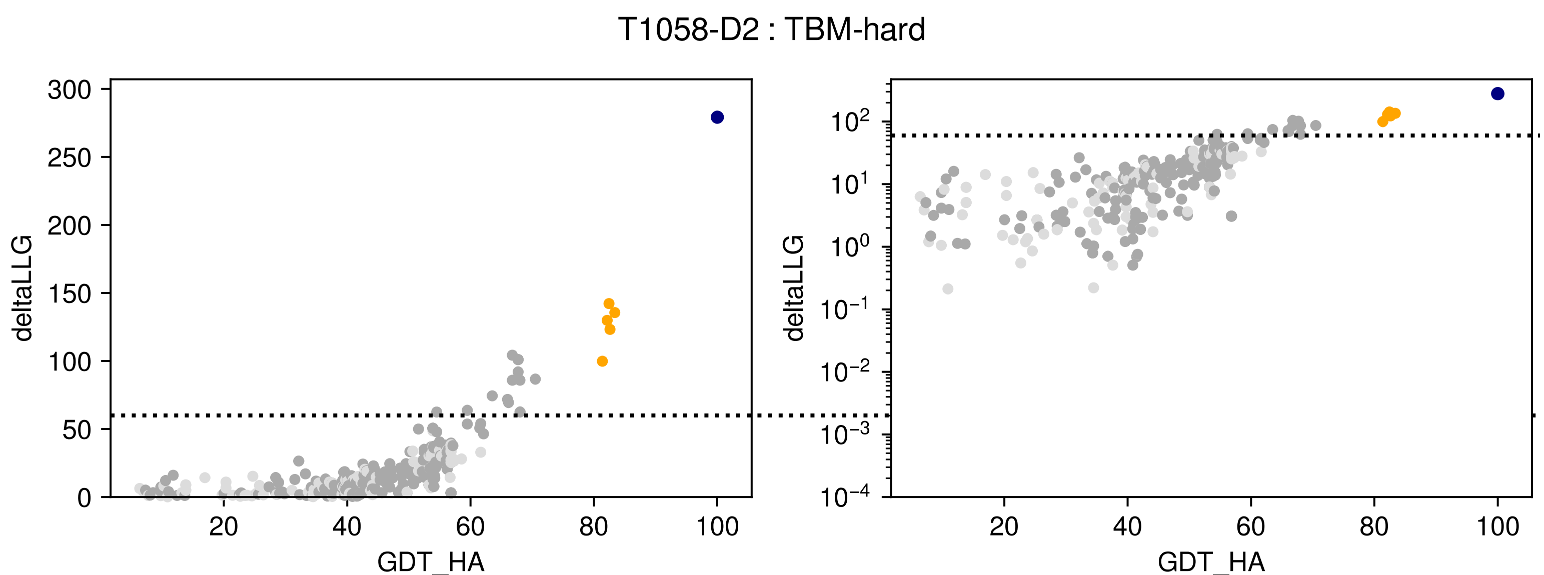


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

Three structures that were solved using Aplhafold2 models

- **T1058** (FM/TBM), two evaluation units

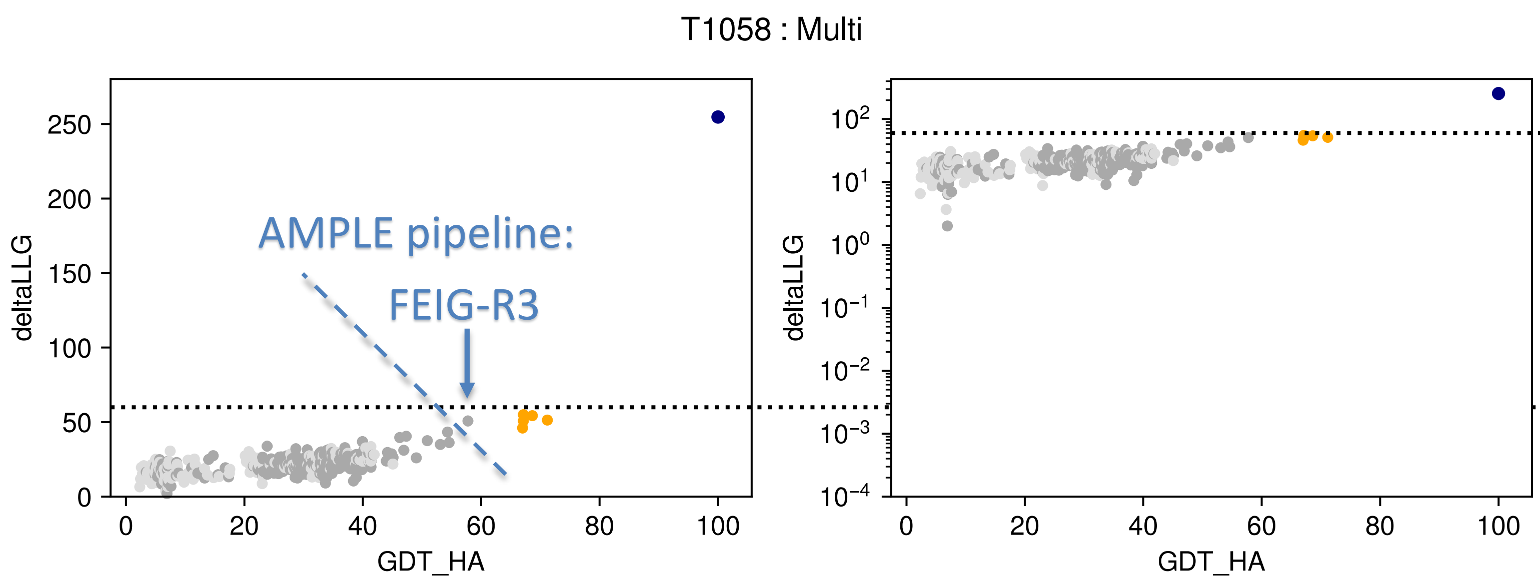


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

Three structures that were solved using Aplhafold2 models

- **T1058** (FM/TBM), two evaluation units

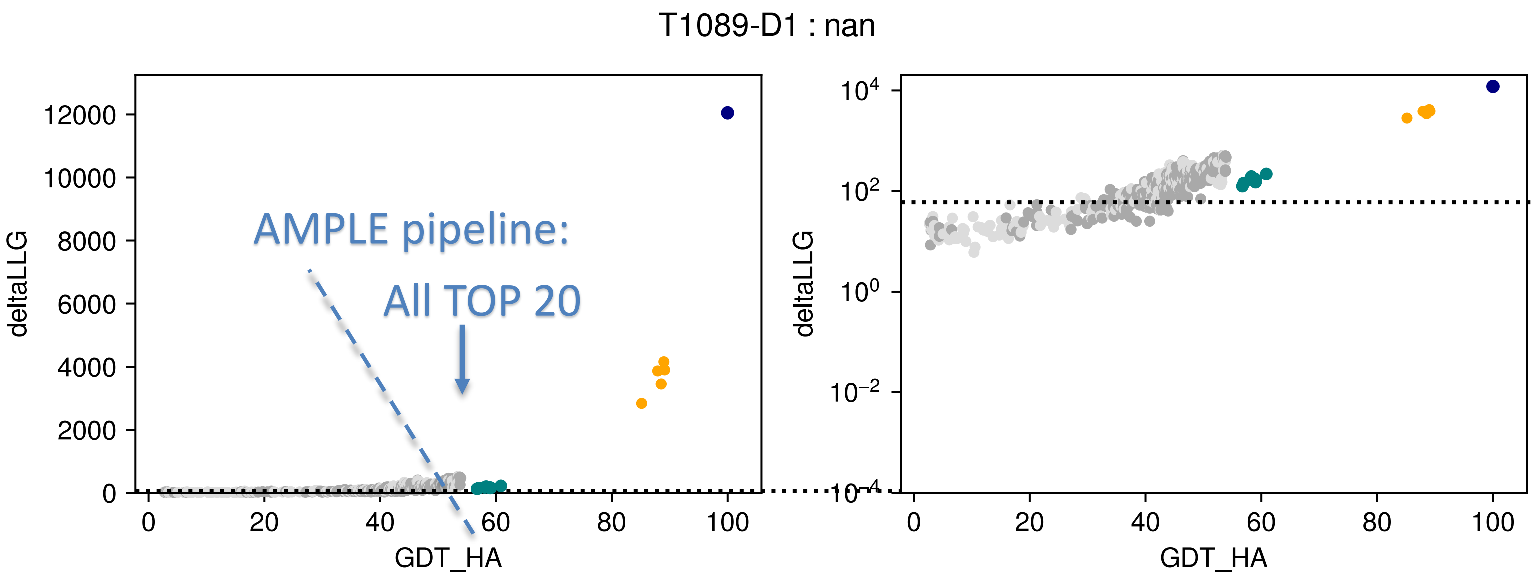


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

Three structures that were solved using Aplhafold2 models

- **T1089** (FM/TBM)

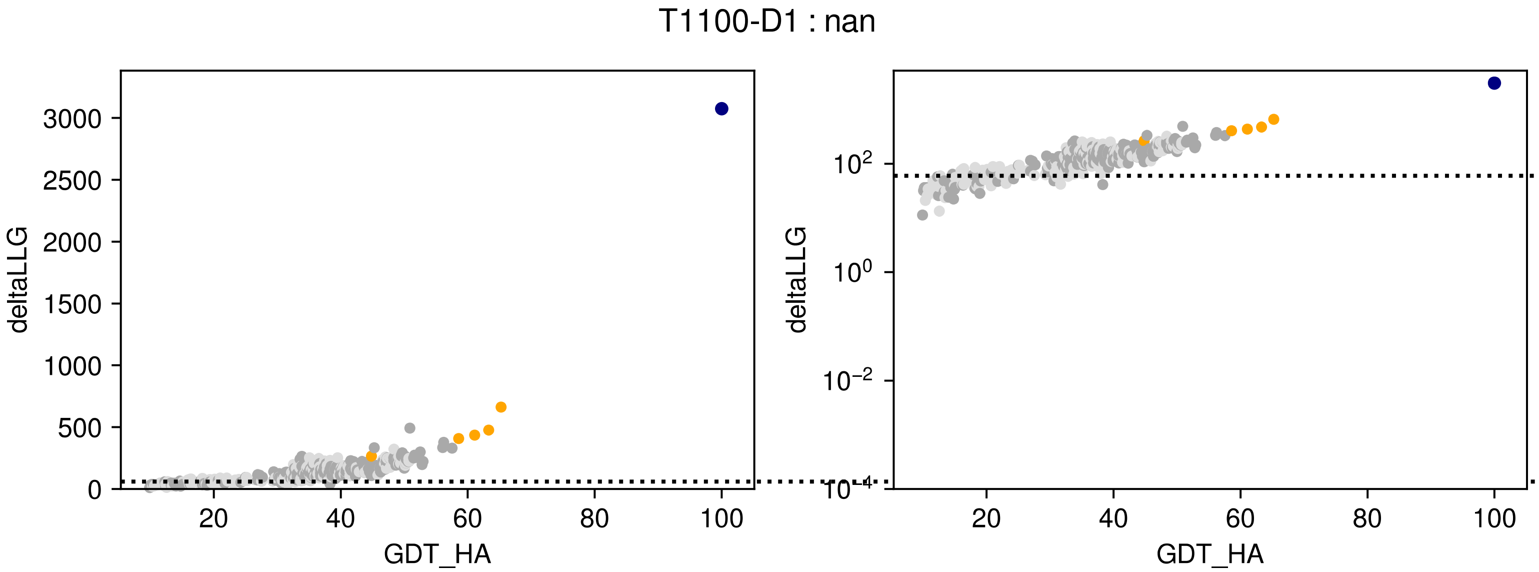


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

Three structures that were solved using Aplhafold2 models

- **T1100** (FM/TBM), two evaluation units

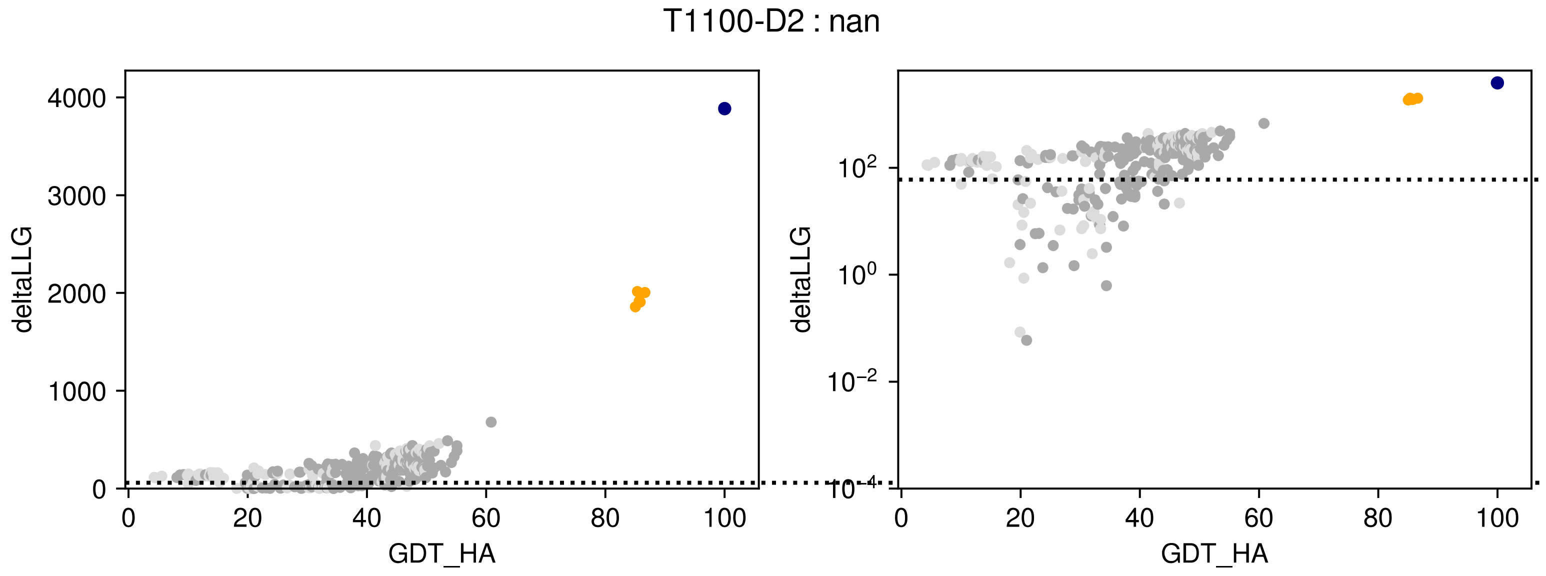


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

Three structures that were solved using Aplhafold2 models

- **T1100** (FM/TBM), two evaluation units

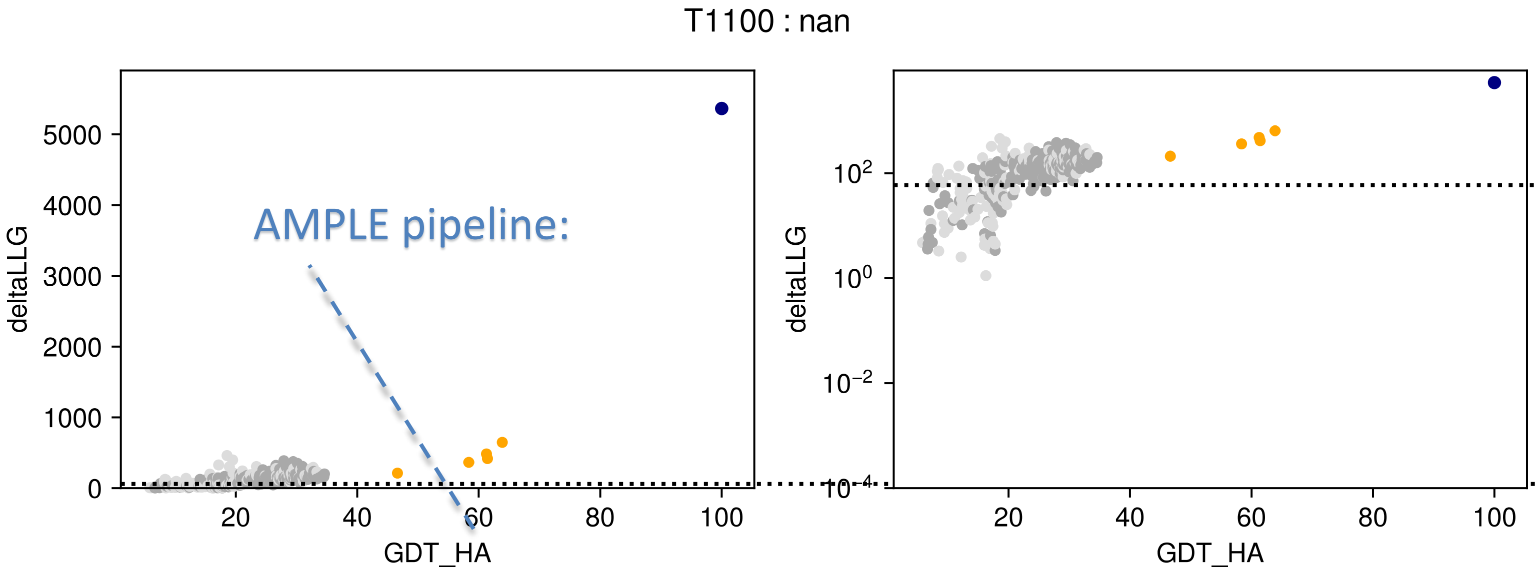


How useful are the models?

Evaluating the usability of models for Molecular Replacement for

Three structures that were solved using Aplhafold2 models

- **T1100** (FM/TBM), two evaluation units

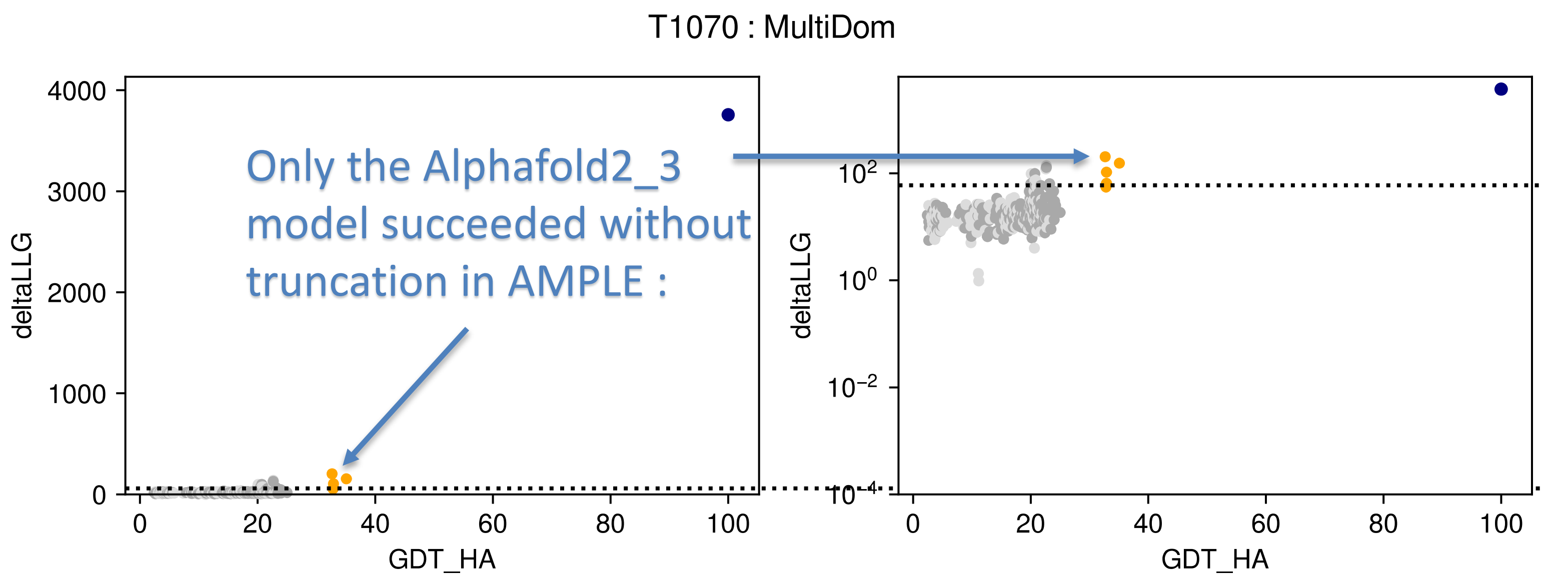


How useful are the models?

Evaluating the usability of models for Molecular Replacement

One of three structures for which truncation of Alphafold2 models were needed

- **T1070** (FM/TBM), four evaluation units

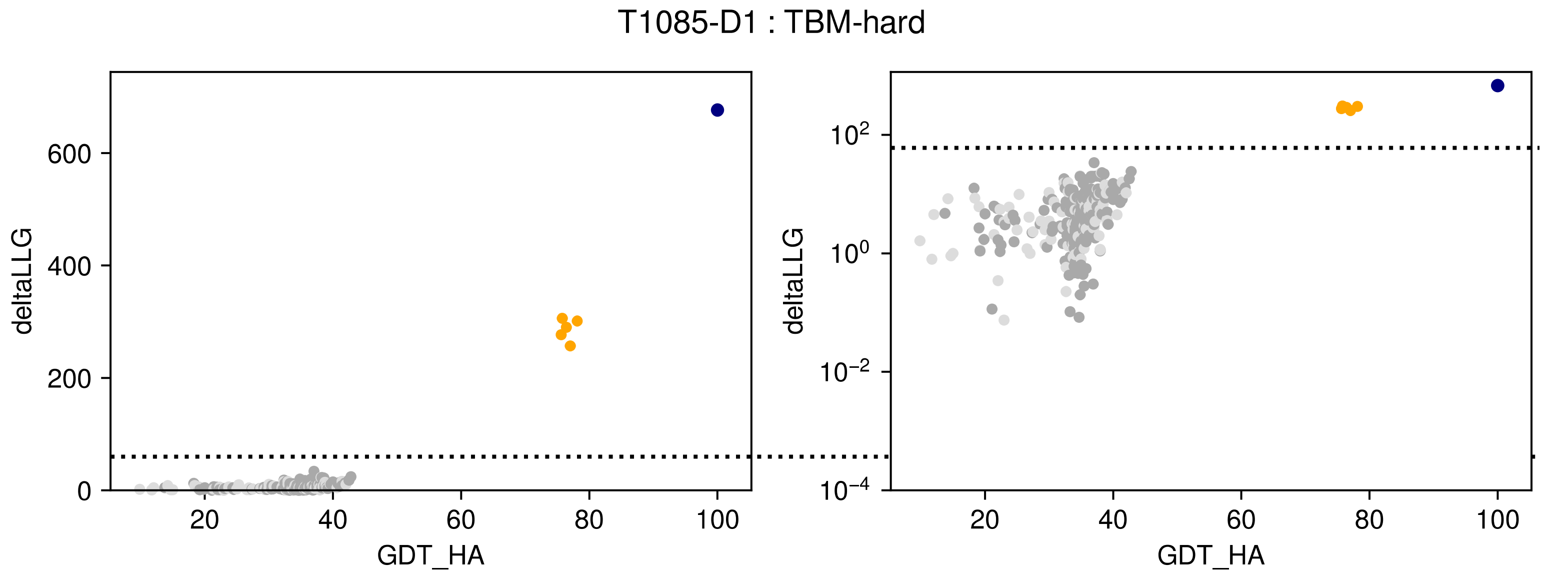


How useful are the models?

Evaluating the usability of models for Molecular Replacement

Outlook: Contribution of refinement

- **T1085-D1** (TBM-hard)

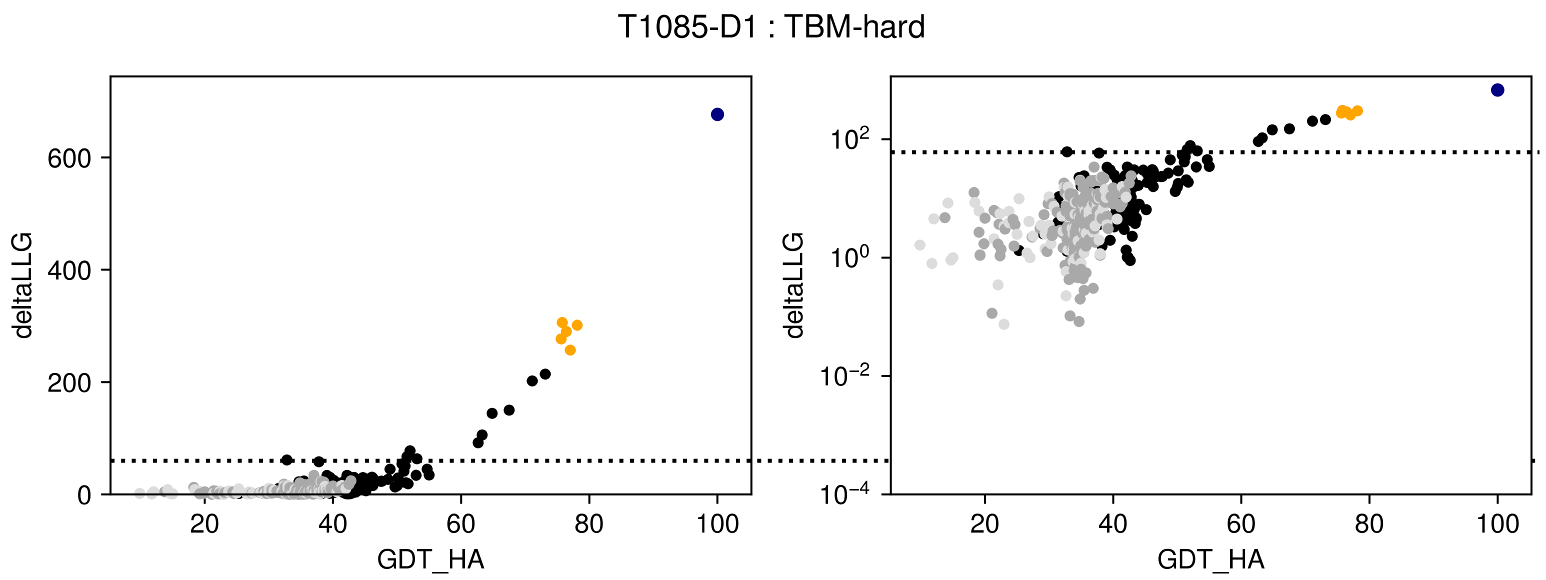


How useful are the models?

Evaluating the usability of models for Molecular Replacement

Outlook: Contribution of refinement

- **T1085-D1** (TBM-hard), **with refinement**

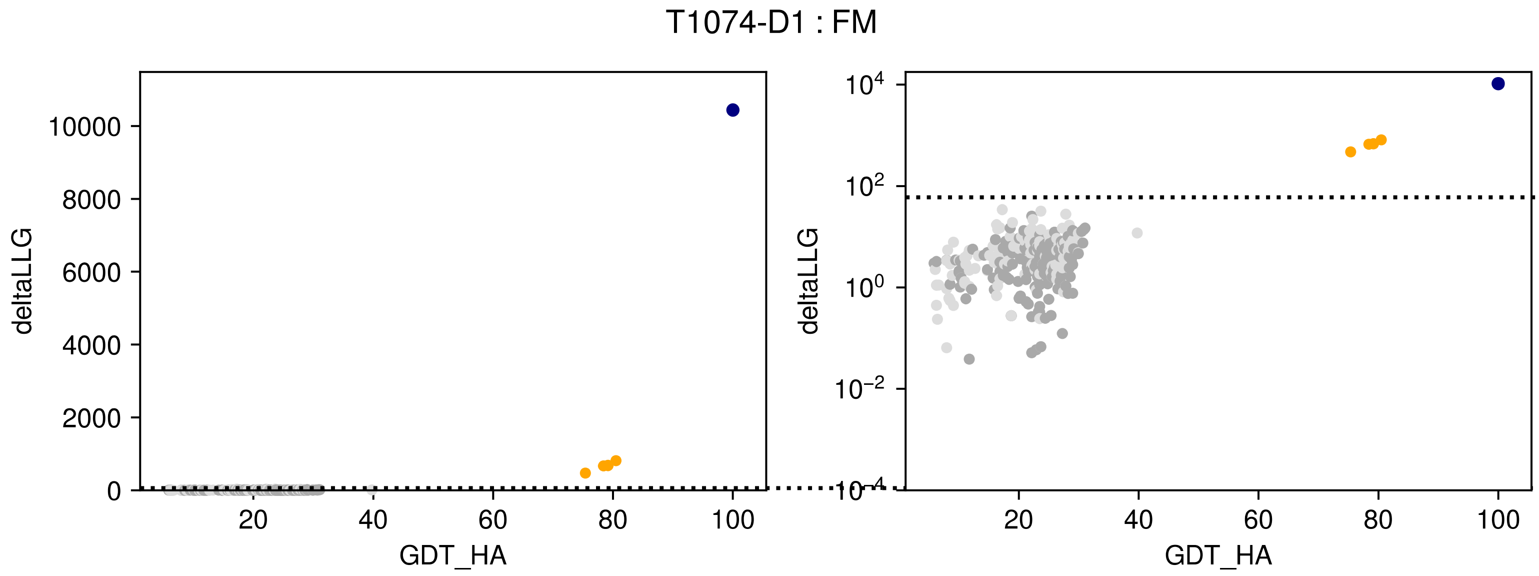


How useful are the models?

Evaluating the usability of models for Molecular Replacement

Outlook: Contribution of refinement

- **T1074** (FM)

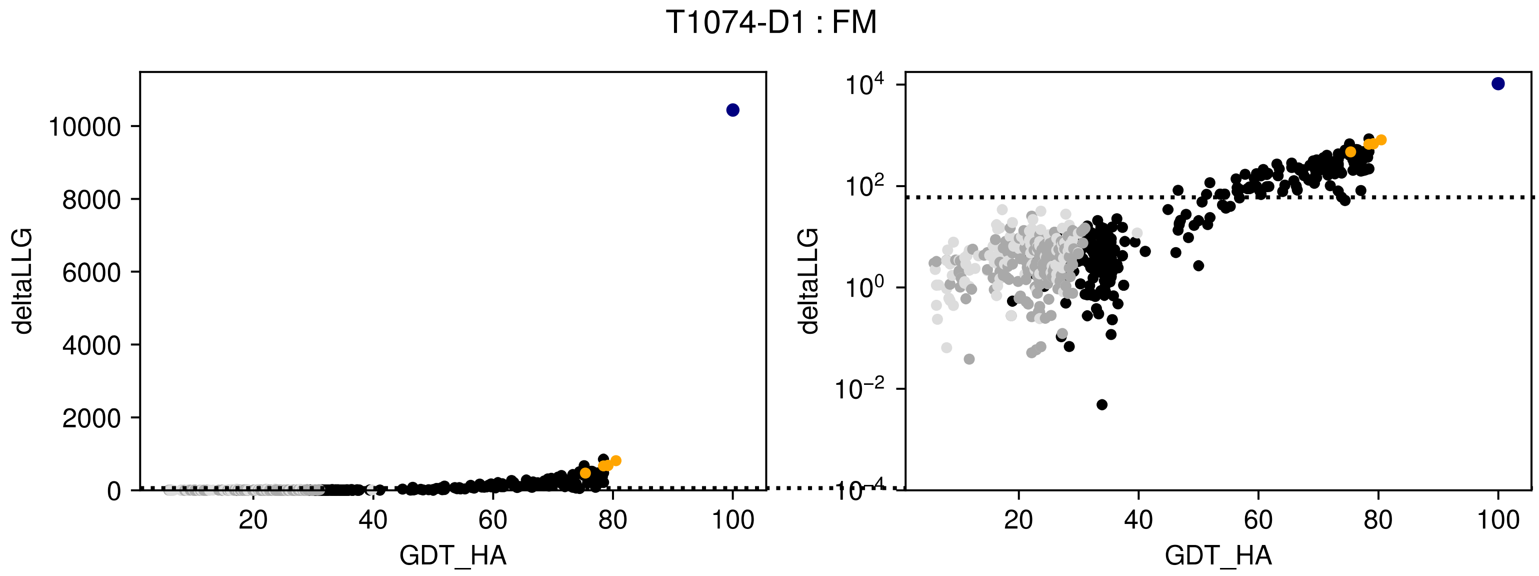


How useful are the models?

Evaluating the usability of models for Molecular Replacement

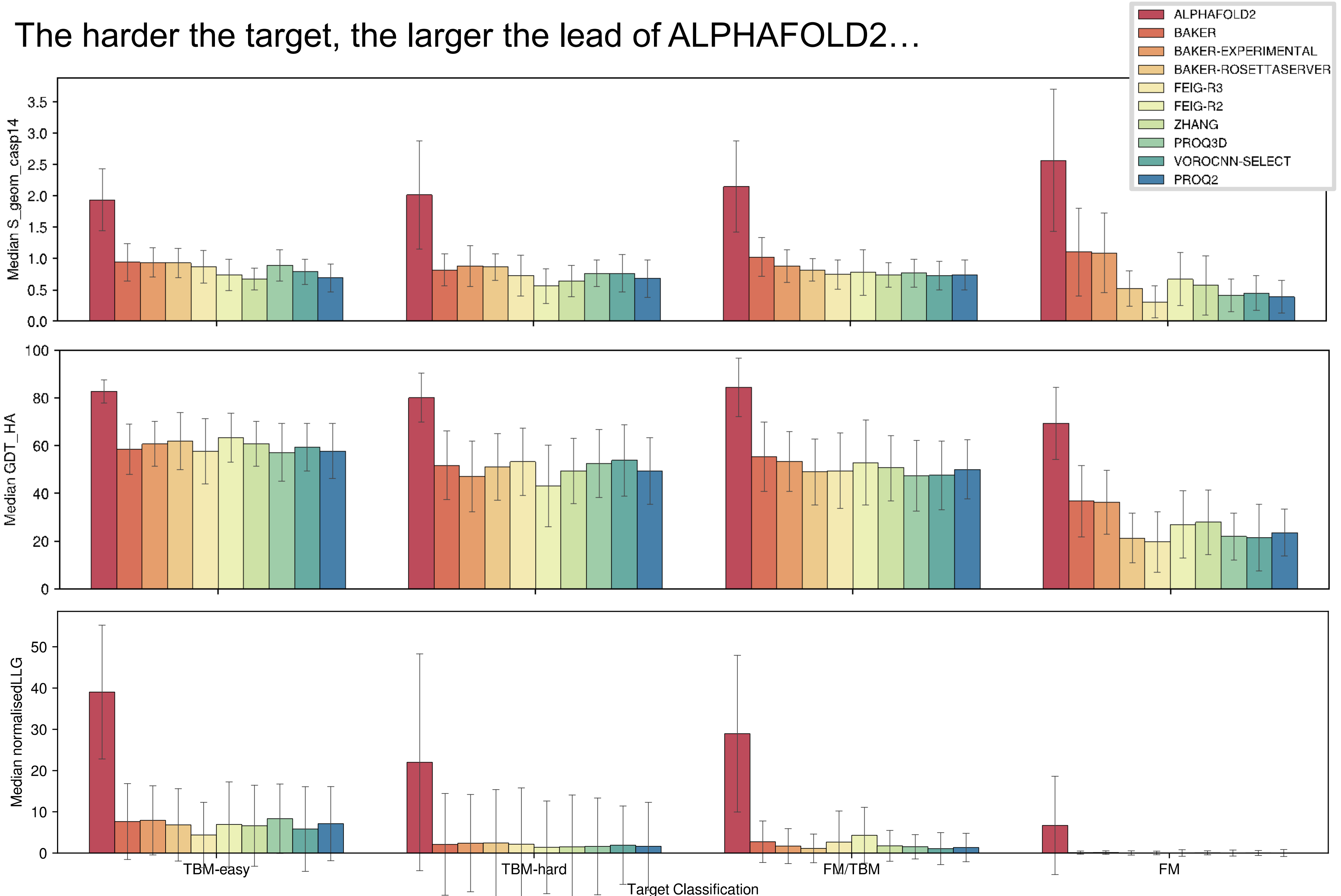
Outlook: Contribution of refinement

- **T1074** (FM), **with refinement**



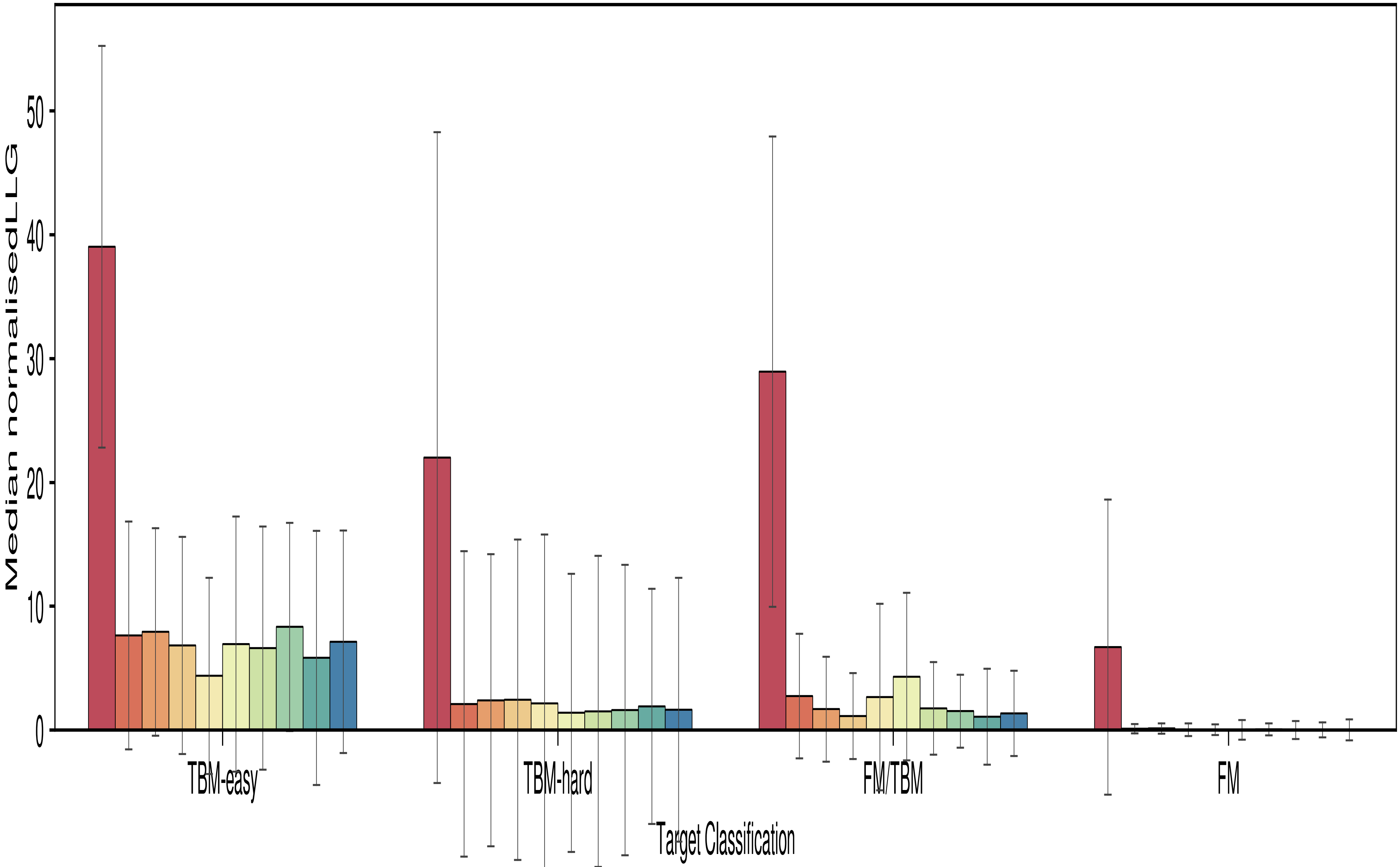
How useful are the models?

The harder the target, the larger the lead of ALPHAFOLD2…



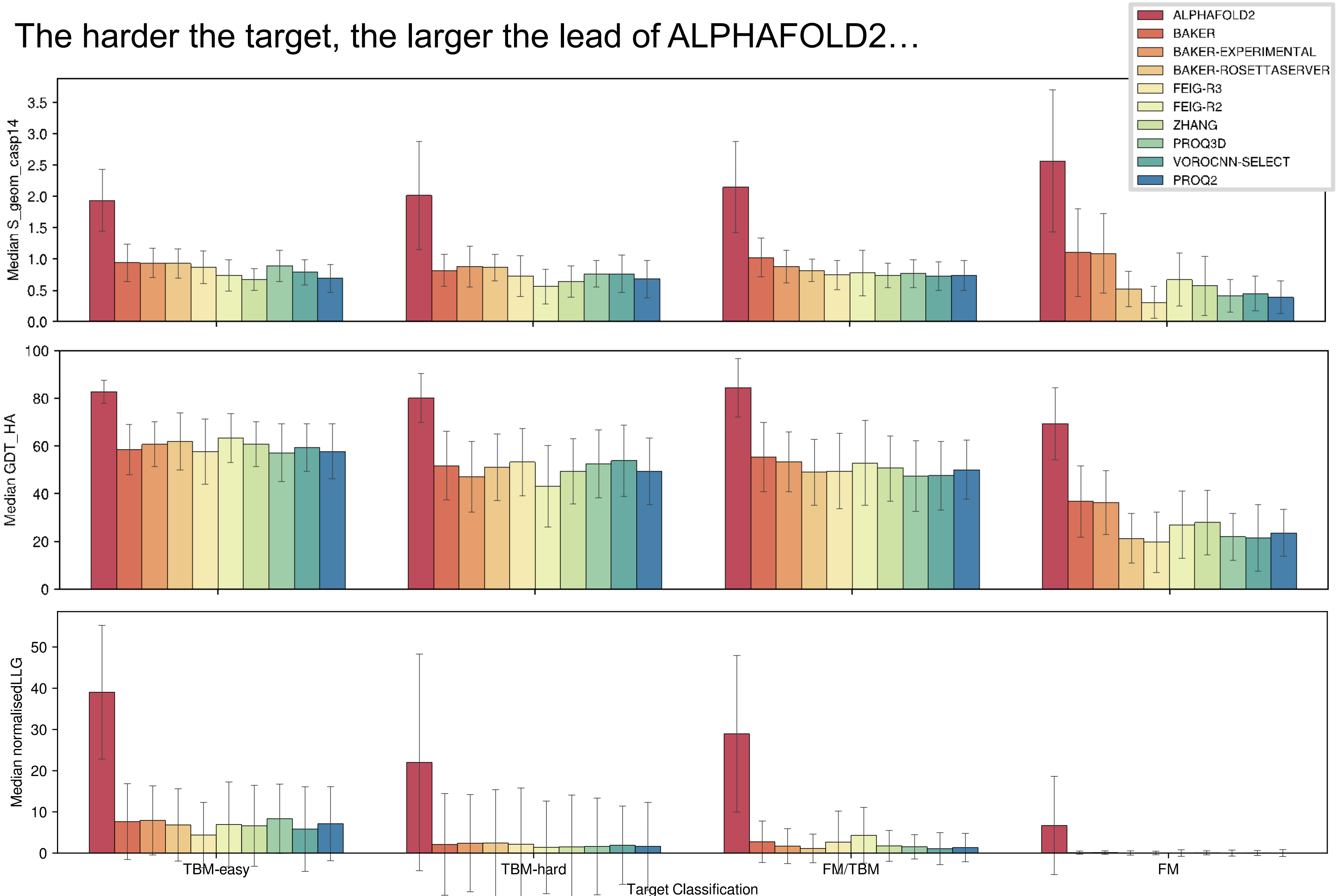
How useful are the models?

The harder the target, the larger the lead of ALPHAFOLD2…



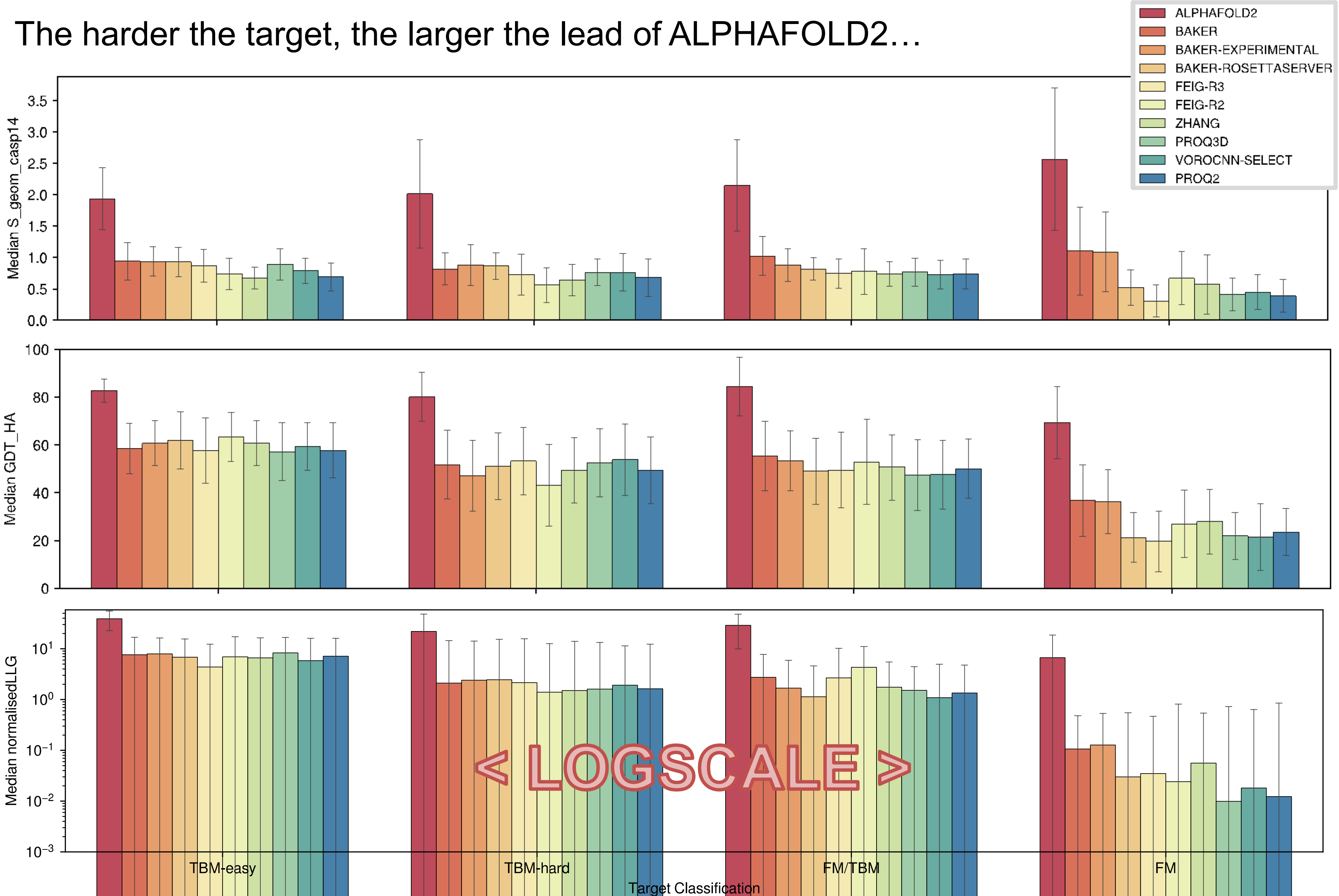
How useful are the models?

The harder the target, the larger the lead of ALPHAFOLD2…



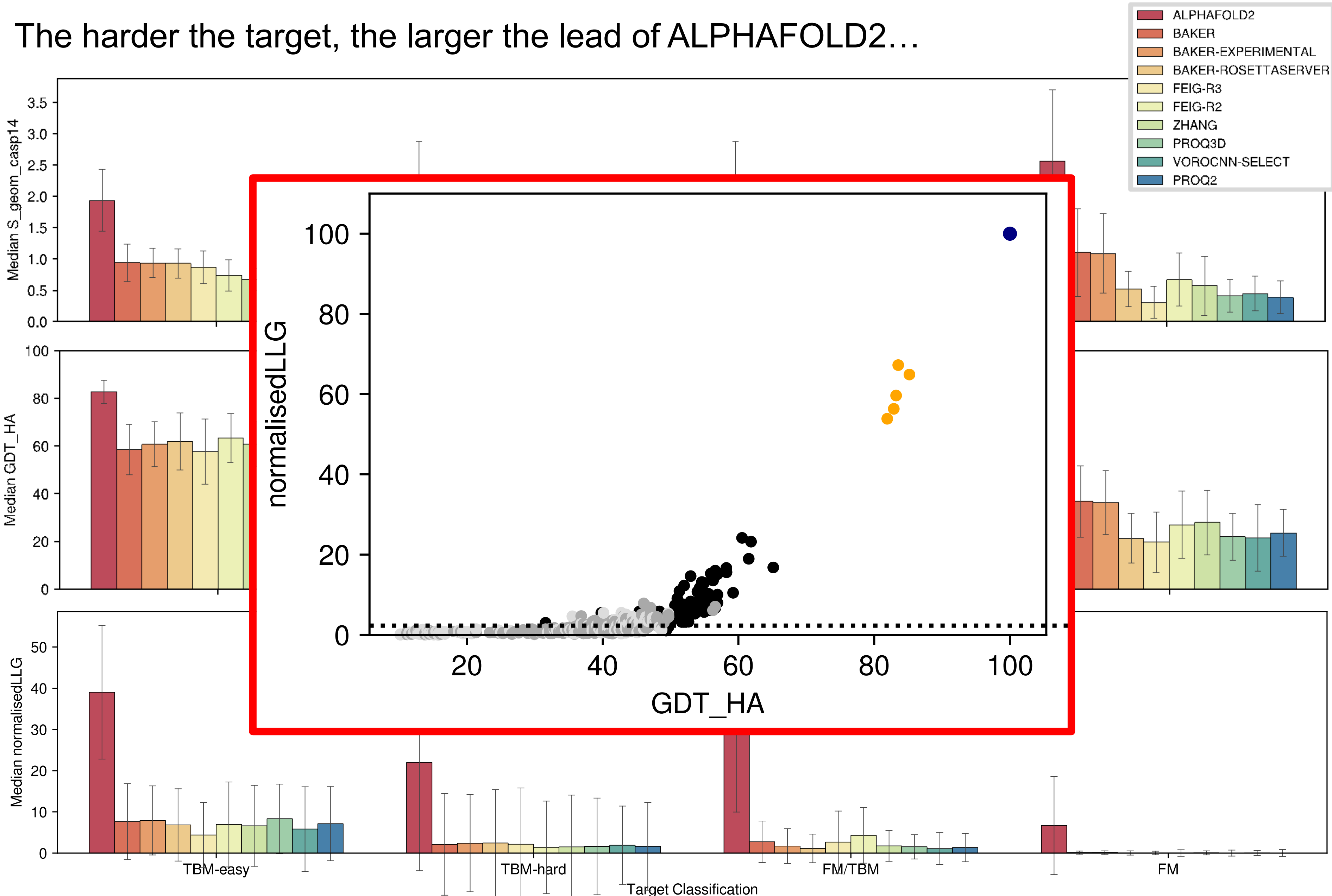
How useful are the models?

The harder the target, the larger the lead of ALPHAFOLD2…



How useful are the models?

The harder the target, the larger the lead of ALPHAFOLD2…



Acknowledgements

@ University of Cambridge

Gabor Bunkoczi

Randy Read

@ University of Liverpool

Adam Simpkin

Ronan Keegan

Dan Rigden

@ MPI Tübingen

Felipe Merino

Vikram Alva