The structure predictions for the mAbs dataset were not successful for all the structures.

To identify if the structure prediction for the sequence in each row was successfully predicted → structure\_status column: “Yes” for structure was predicted and “No” for not predicted.

Thus, the dataset “AbStruc\_mAbs\_developability\_final.csv.zip” has rows with empty (NA) values. ONLY in case for structure\_status = “No”

Hence, this dataset has the same number of rows as the original dataset used to generate it (aligned\_mAbs\_imgt.csv).

If you want to examine the rows that failed you can do something like

AbStruc\_mAbs\_developability\_final %>%

filter (structure\_status == "No")

dataset nrows = 1024

nrows for successful structures = 1015

Failed structures = 9

**Update: 21/07/2022**

The whole\_mAbs\_only dataset:

This dataset is a subset of the mAbs dataset. It includes only the therapeutics of the format (“Whole mAb”, and “Whole mAb (Mouse)”) as defined by TheraSabDAB.

This dataset can be found in the daughter folder “whole\_mabs\_only” which includes:

* “whole\_mAbs\_only.csv”: raw sequence and alignment data, contains 782 rows for 391 unique (non-redundant mAbs).
* “AbChain\_whole\_mAbs\_developability\_final.csv”: single-chain (unpaired) sequence developability parameters.
* “AbStruc\_whole\_mAbs\_developability\_final.csv”: single-chain (unpaired) structure developability parameters.
* “AbStruc\_whole\_mAbs\_developability\_paired.csv”: waiting for an update from Eva: 02/08/2022 → done 04/08/2022  
  This dataset has only 384 mAbs → 7 mAbs we were unable to build PAIRED structure for.