The structure predictions for the mutated sampled 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\_mutated\_sampled\_developability.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 (mutated\_dataset\_sampled.csv).

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

AbStruc\_mutated\_sampled\_developability %>%

filter (structure\_status == "No")

dataset nrows = 234798

nrows for successful structures = 232722

Failed structures = 2076

**#update - 20220713**

We decided

* to re-predict structures of mutated antibodies using IgFold (ABodyBuilder is not enough to detect structure changes with single mutations).
* Include only the mutations in the CDRs → 63,211 sequences
* Predicted successfully with struc developability measures → 36,210

**#Datafrme produced on 20220803 by Habib from the files provided by Eva in (“/developability\_project/from\_eva/sensitivity\_structure\_parameters\_20220713”) - cyverse**

→ AbStruc\_mutated\_cdrs\_developability\_igfold.csv.zip has the rowid and the mutation type plus all the 48 structure parameters for the sampled mutated antibodies with mutations occurring only in the CDRs.

We are awaiting Eva’s attempt to build the full 63,211 dataset though.

**#update - 20220830**

Eva has re-predicted the structures **for all the 63K mutated antibodies**. She gave them to Habib on 29 August 2022 and Habib curated the data and uploaded it to the [shared folder.](https://drive.google.com/drive/folders/1MRyTvaR4hVDEo2aRjkOlUAOycG0JaOXe) File name **AbStruc\_mutated\_cdrs\_developability\_igfold\_63K.csv**