Systematic evaluation of computational biophysical measurements of missense mutations in the absence of experimental structures

Qisheng Pan[1,2], **Stephanie Portelli**[1,2], **Thanh Binh Nguyen**[1,2], **Douglas E.V. Pires**[2,3], **David B. Ascher**[1,2]

[1]*School of Chemistry and Molecular Bioscience, University of Queensland, Brisbane Queensland 4072, Australia*
[2]*Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne Victoria 3004, Australia*
[3]*School of Computing and Information Systems, University of Melbourne, Melbourne, VIC 3053, Australia*

Qisheng Pan
QishengPan
qisheng.pan@uq.net.au

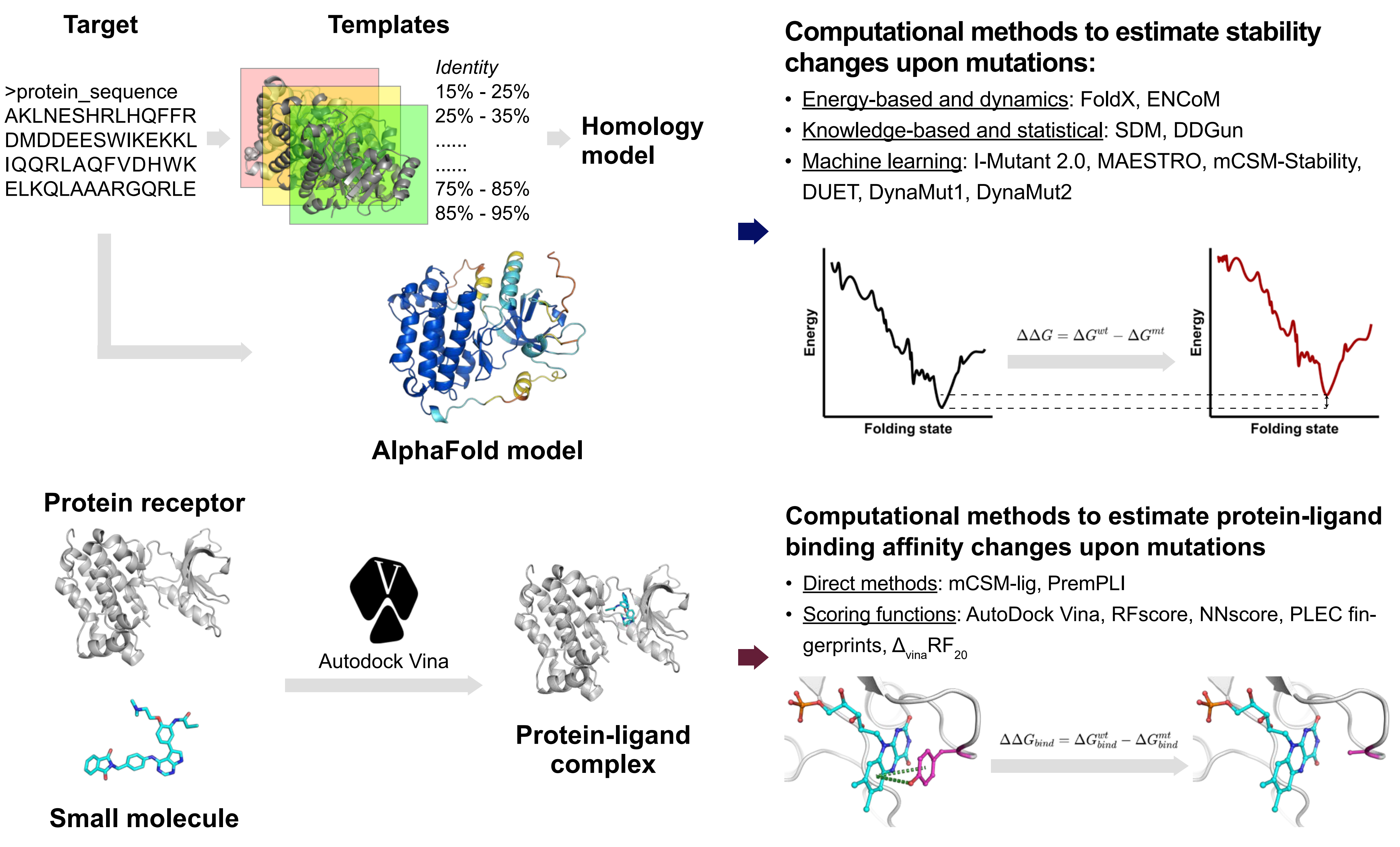
Highlights

- We present the <u>first systematic study</u> assessing how methods to predict <u>effect of mutations on stability changes and protein-ligand binding affinity</u> cope in the absence of high-resolution experimental protein structures.
- This work provides <u>a detailed guideline for *in silico* mutation analysis</u>, which will assist users in appropriately using non-experimental models, such as AlphaFold2 models, <u>on protein engineering and drug development</u>.

Introduction

- AlphaFold2 and other methods for protein modelling provide substantial resources of protein structures.
- However, there is **no systematic evaluation** of the reliability of current computational biophysical measurements **in the absence of experiment-determined structure**.
- We have, therefore, systematically investigated the performance and robustness of widely used structural methods to predict the effect of mutations on **protein stability** and **protein-ligand binding affinity** when presented with these non-experimental models.

Protein modelling and mutation analysis



Computational methods to estimate stability changes upon mutations:

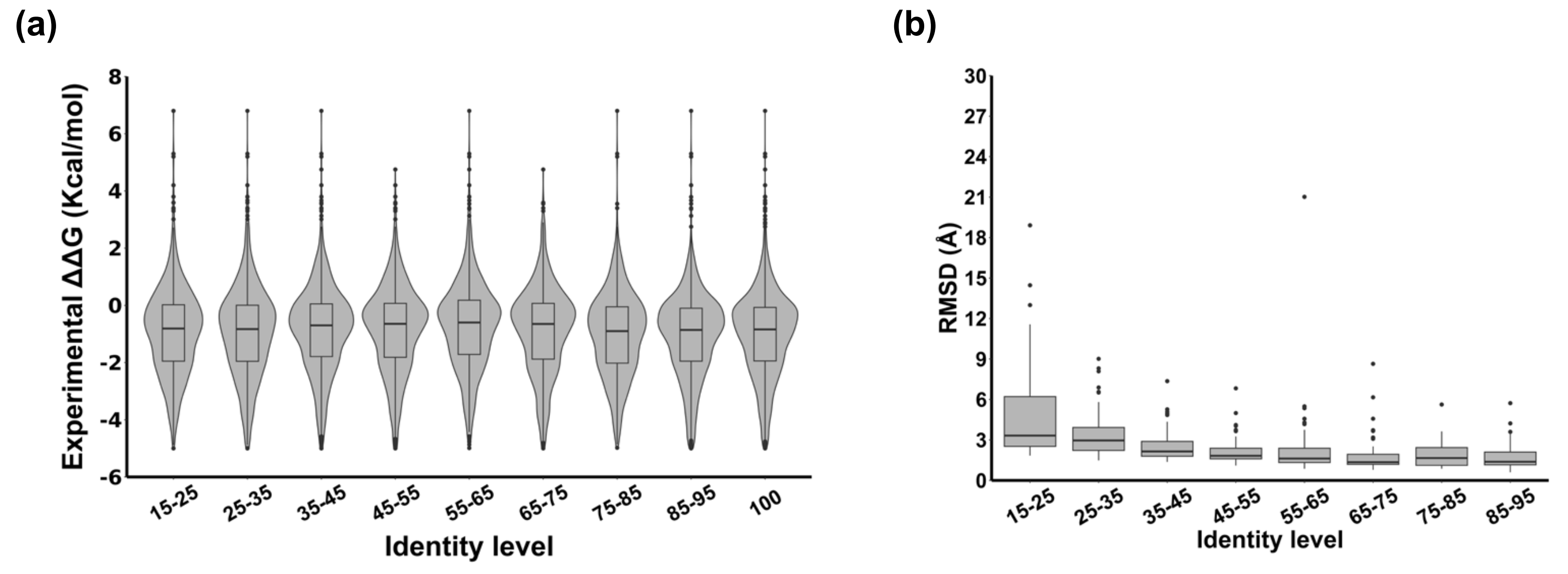
- <u>Energy-based and dynamics</u>: FoldX, ENCoM
- <u>Knowledge-based and statistical</u>: SDM, DDGun
- <u>Machine learning</u>: I-Mutant 2.0, MAESTRO, mCSM-Stability, DUET, DynaMut1, DynaMut2

Computational methods to estimate protein-ligand binding affinity changes upon mutations

- <u>Direct methods</u>: mCSM-lig, PremPLI
- <u>Scoring functions</u>: AutoDock Vina, RFscore, NNscore, PLEC fingerprints, $\Delta_{vina}RF_{20}$

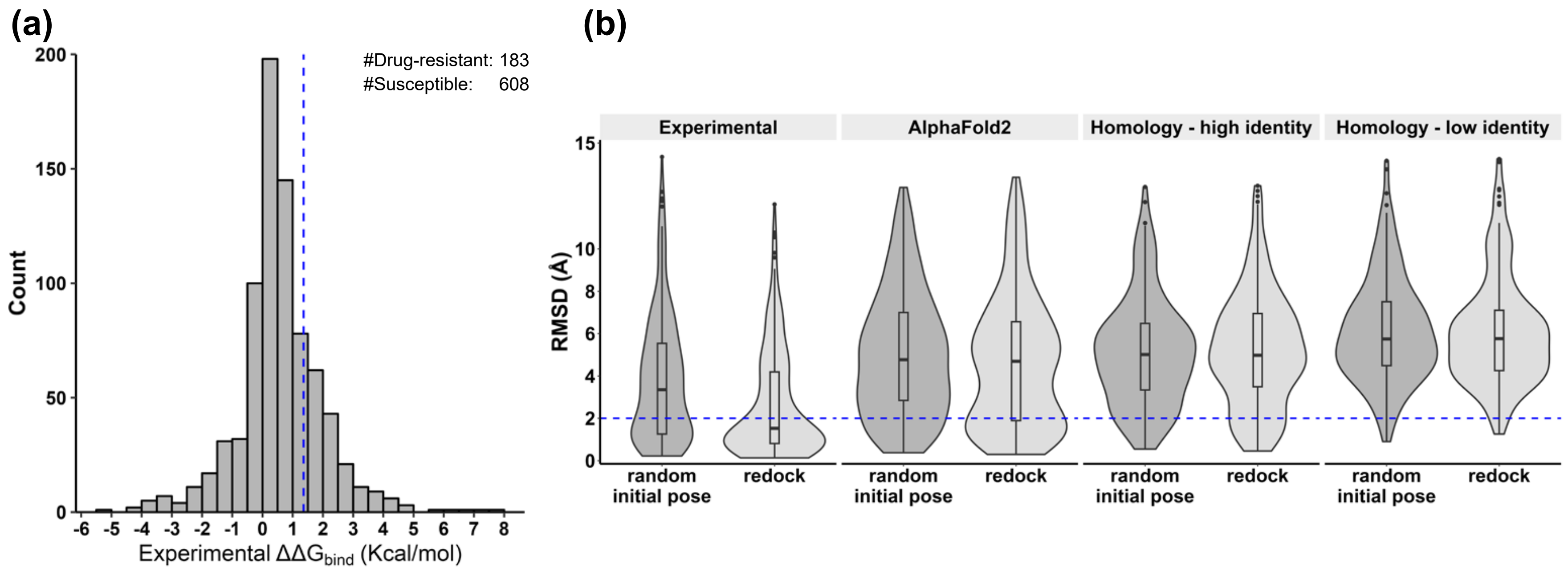
- We modelled the protein structures and protein-ligand complexes, and put these non-experimental models to evaluate the performance of computational biophysical measurements.

Distribution of datasets to assess stability changes



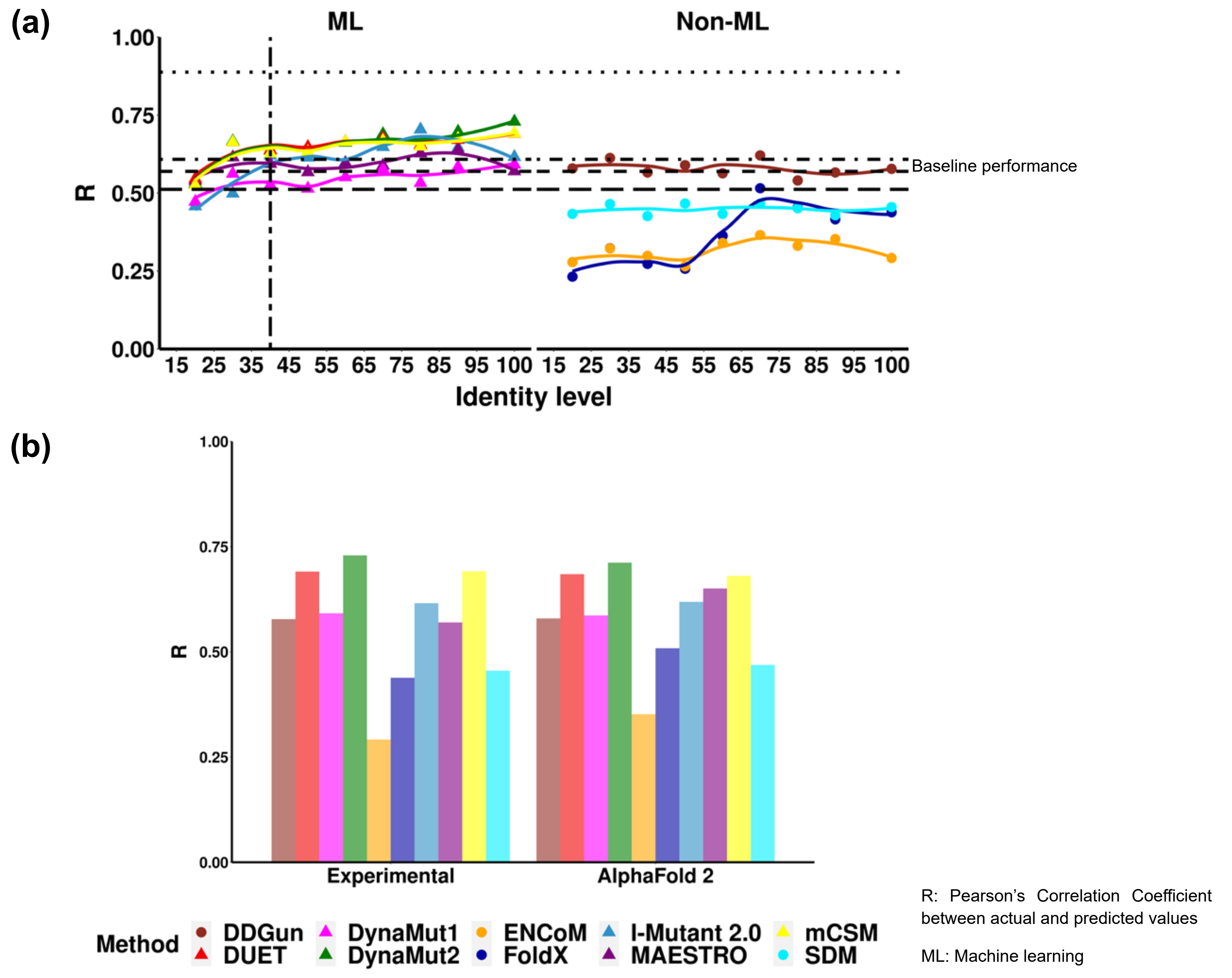
- Datasets share similar distribution of ΔΔG values (a).
- The higher the target-template identity is, the more similar the homology models with the experimental structures will be (b).

Distribution of datasets to assess binding affinity changes



- An imbalanced distribution of the phenotypes of drug-resistance was observed on this dataset (a).
- Our docking pipeline showed an expected successful rate, with around 20% docked ligand which has an RMSD lower than 2Å (b).

Overall performance on assessing stability changes

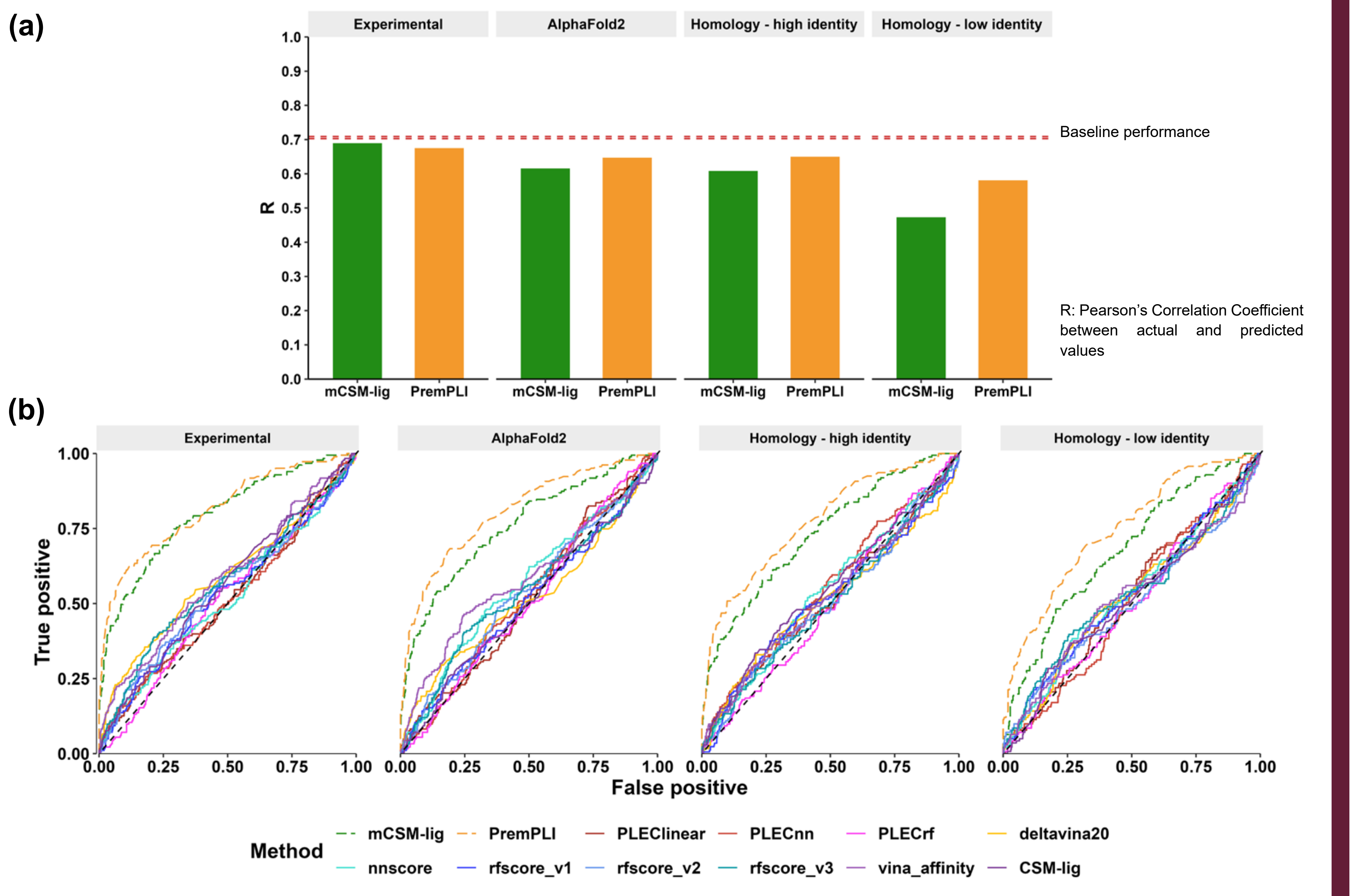


R: Pearson's Correlation Coefficient between actual and predicted values

ML: Machine learning

- In general, the predictive performance of the evaluated methods increases with target-template identity (a).
- Alternatively, we observed **a consistent performance deterioration** for **all structure-based methods**, particularly in **machine learning based methods** and FoldX, when the sequence identity of the homology modelling template dropped.
- Performance of most methods on AlphaFold2 models is close to those obtained on experimental structures (b).

Overall performance on assessing binding affinity changes



R: Pearson's Correlation Coefficient between actual and predicted values

- We observed **a small performance deterioration** when the direct methods were presented with **AlphaFold2 models** and **homology models with high sequence identity**, compared with the baseline (a).
- There is **a large performance drop** when using **homology models with low sequence identity** as inputs.
- The **scoring functions** may **not be suitable** to estimate the effect of mutation on protein-ligand binding affinity changes (b).