

Supplementary Materials

Sequence-based prediction of the solubility of peptides containing non-natural amino acids

Marc Oeller¹, Ryan Kang¹, Hannah Bolt², Ana Gomes dos Santos³, Annika Langborg
Weinmann⁴, Antonios Nikitidis⁵, Pavol Zlatoidsky⁵, Wu Su⁵, Werngard Czechtizky⁵,
Leonardo De Maria⁵, Pietro Sormanni^{1*}, Michele Vendruscolo^{1*}

¹*Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry,
University of Cambridge, Cambridge, UK*

²*Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK*

³*Pharmaceutical Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK*

⁴*Early Chemical Development, Pharmaceutical Sciences,
BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*

⁵*Medicinal Chemistry, Research and Early Development, Respiratory and Immunology,
BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*

*Corresponding Authors: mv245@cam.ac.uk, ps589@cam.ac.uk

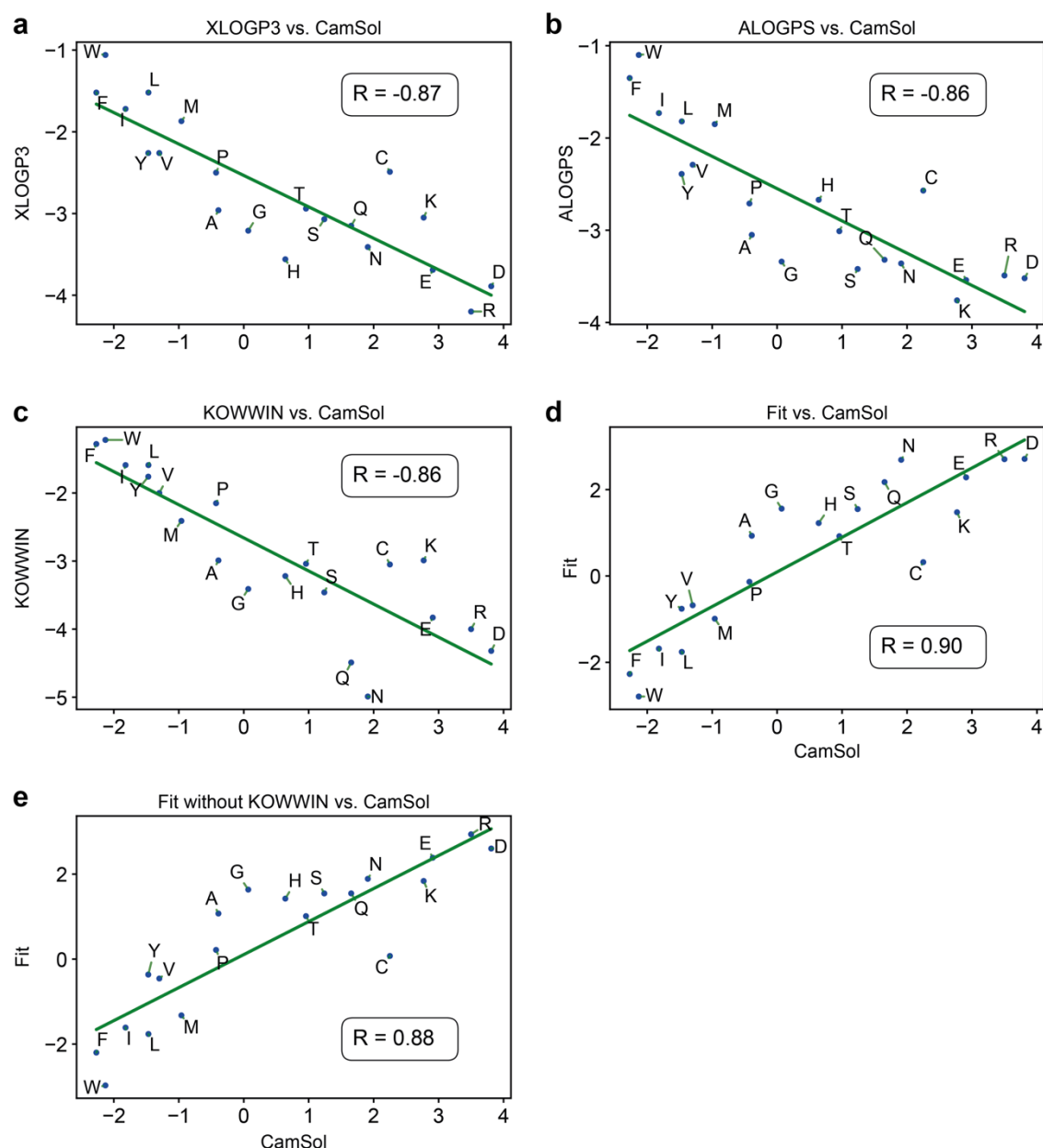


Figure S1. Comparison of different hydrophobicity predictors with CamSol hydrophilicity values. Correlation between the hydrophobicity values of XLOGPS (a), ALOGPS (b) and KOWWIN (c) and the tabulated hydrophilicity values used in CamSol. (d) Values from all three predictors were used to fit a linear regression model to the CamSol hydrophilicity values and then plotted to show the correlation between the fit and the tabulated CamSol hydrophilicity values (Pearson's coefficient of correlation = 0.9). (e) Fit without KOWWIN against CamSol hydrophilicity values (Pearson's coefficient of correlation = 0.88).

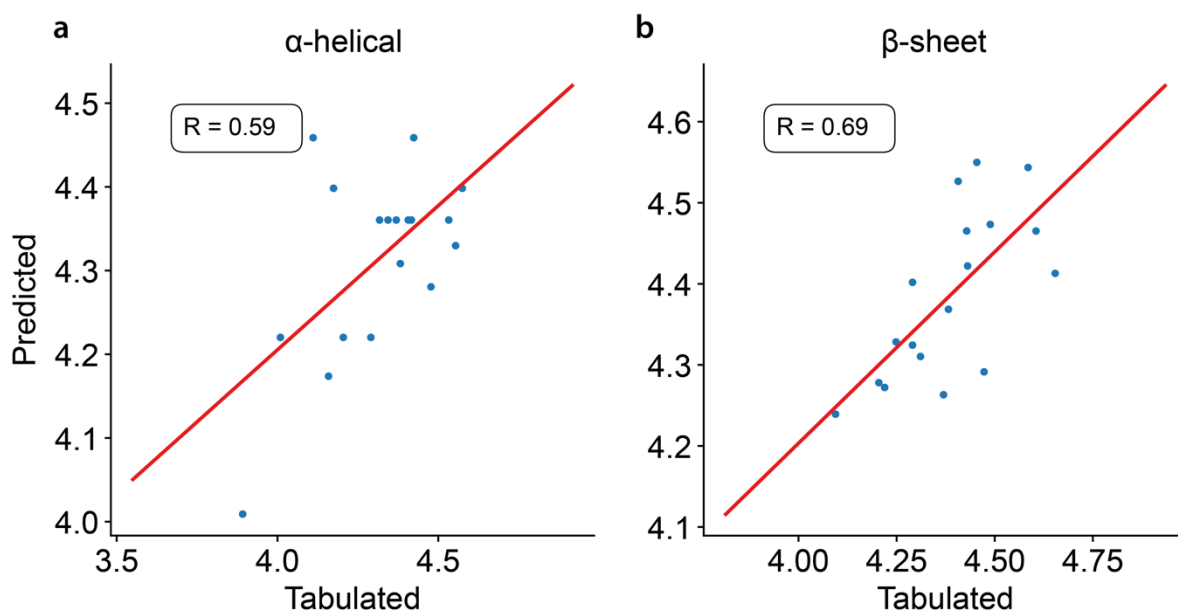


Figure S2. Correlation between predicted and tabulated secondary structure propensities. The secondary structure propensity predictor predicts the α -helical propensity **(a)** and β -sheet propensity **(b)** for the 20 natural amino acids with Pearson's coefficients of correlation of 0.59 and 0.69, respectively.

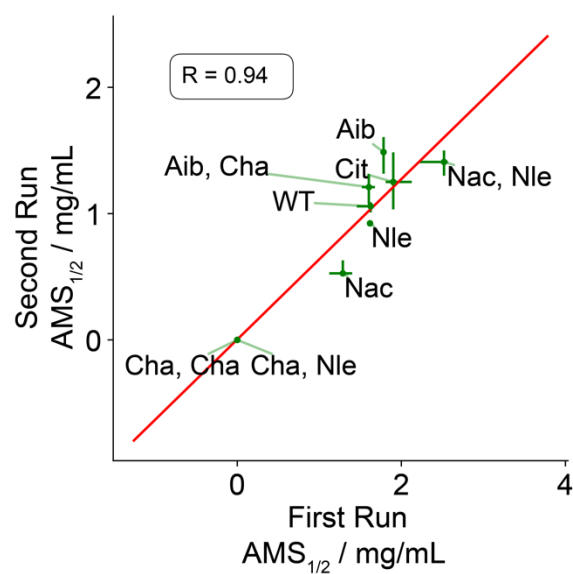


Figure S3. Comparison of independent AMS runs. To verify that replacing PEG with AMS continues to yield reliable and replicable results we performed two independent solubility experiments with AMS on two separate days. The correlation indicates that replacing PEG with AMS is feasible.

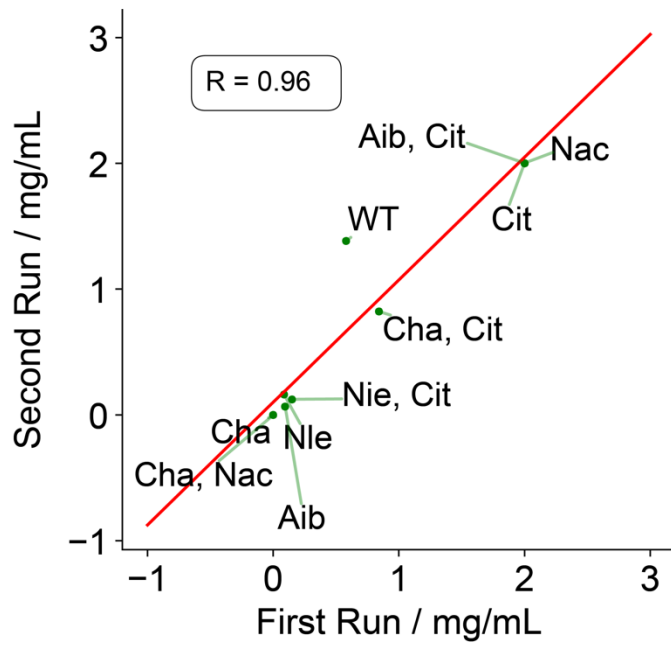
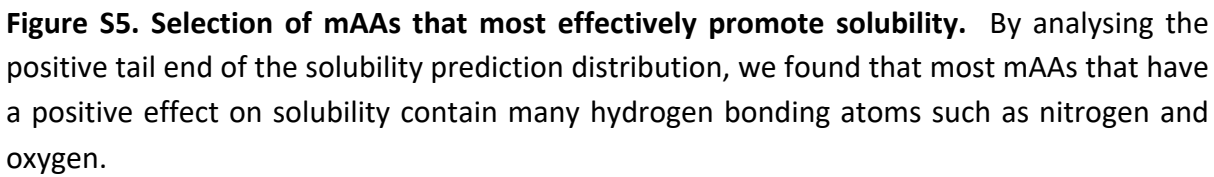


Figure S4. Comparison of independent ultracentrifugation runs. To verify the reproducibility of the ultracentrifugation method we performed two independent experiments on different days. The high correlation shows that the results from the ultracentrifugation assay are reliable.



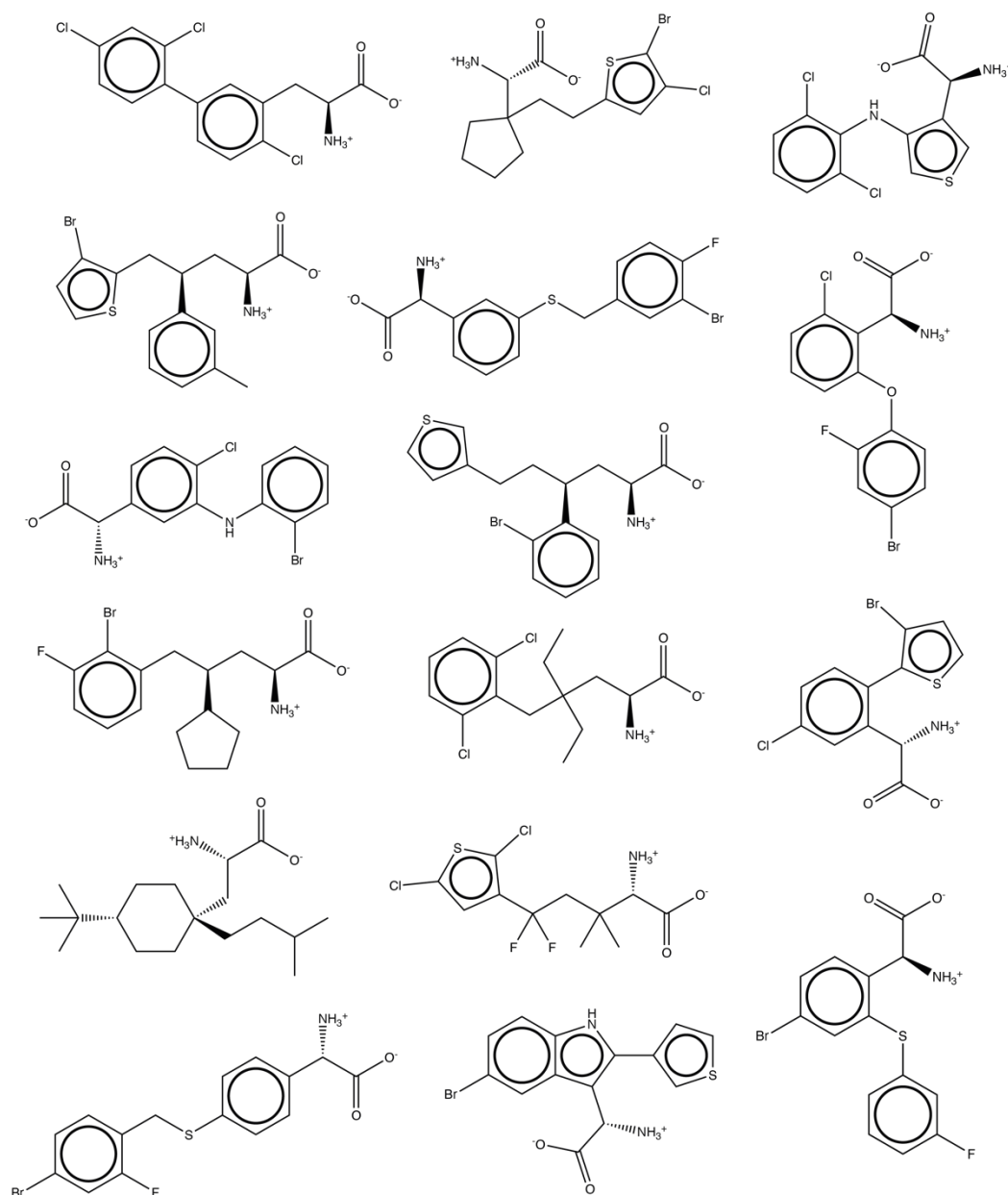


Figure S6. Selection of mAAs that most effectively decrease solubility. By analysing the negative tail end of the solubility prediction distribution, we found that most mAAs that have a negative effect on solubility contain several aromatic rings and often halogens such as chlorine or bromine.