A Bayesian model averaging approach for solvation energy prediction

Christopher C. Overall,[†] Luke J. Gosink,[‡] Paul D. Whitney,[¶] Elizabeth R. Jurrus,[‡] David L. Mobley,[§] and Nathan A. Baker[*,‡,‖]

Biological Sciences Division, Pacific Northwest National Laboratory, Computational and Statistical Analytics Division, Pacific Northwest National Laboratory, Advanced Computing, Mathematics, and Data Division, Pacific Northwest National Laboratory, Pharmaceutical Sciences, Pacific Northwest National Laboratory, and Division of Applied Mathematics, Brown University

E-mail: nathan.baker@pnnl.gov

⋆⋆ Author list order is arbitrary ⋆⋆

⋆⋆ Title is tentative ⋆⋆

Abstract

⋆⋆ We need an abstract. ⋆⋆

1 Introduction

⋆⋆ Background on SAMPL results and accuracy limits ⋆⋆

[*]To whom correspondence should be addressed
[†]Biological Sciences Division, Pacific Northwest National Laboratory
[‡]Computational and Statistical Analytics Division, Pacific Northwest National Laboratory
[¶]Advanced Computing, Mathematics, and Data Division, Pacific Northwest National Laboratory
[§]Pharmaceutical Sciences, Pacific Northwest National Laboratory
[‖]Division of Applied Mathematics, Brown University

⋆⋆ Background on solvation calculations and need for better accuracy ⋆⋆

⋆⋆ Background on BMA and past applications[1] ⋆⋆

2 Methods

2.1 SAMPL dataset

⋆⋆ Overview of SAMPL dataset[2,3] ⋆⋆

2.2 Molecular test cases

⋆⋆ Groupings of molecules into subsets (if feasible) or scoring based on $\log P$ and related methods ⋆⋆

2.3 Solvation models used in ensemble

⋆⋆ Groupings of input methods by type (per Mobley paper), cost, etc. ⋆⋆

2.4 Bayesian model averaging approach

⋆⋆ Overview of BMA method ⋆⋆

3 Results

⋆⋆ Testing/training results ⋆⋆

⋆⋆ Subgroup performance ⋆⋆

4 Discussion

5 Conclusions

⋆⋆ **Need for external validation** ⋆⋆

⋆⋆ **Prospects for use in future SAMPL events** ⋆⋆

Acknowledgement

The authors thank . . . ⋆⋆ **include supporting organizations.** ⋆⋆

Supporting Information Available

⋆⋆ **In case we have supporting information.** ⋆⋆

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Gosink, L. J.; Hogan, E. A.; Pulsipher, T. C.; Baker, N. A. *Proteins: Structure, Function and Bioinformatics* **2014**, *82*, 354–363.

(2) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. *Journal of computer-aided molecular design* **2014**, *28*, 135–50.

(3) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *Journal of medicinal chemistry* **2008**, *51*, 769–79.

Graphical TOC Entry