



HAL
open science

Perspective on the SAMPL and D3R Blind Prediction Challenges for Physics-Based Free Energy Methods

Nicolas Tielker, Lukas Eberlein, Oliver Beckstein, Stefan Güssregen, Bogdan Iorga, Stefan M. Kast, Shuai Liu

► **To cite this version:**

Nicolas Tielker, Lukas Eberlein, Oliver Beckstein, Stefan Güssregen, Bogdan Iorga, et al.. Perspective on the SAMPL and D3R Blind Prediction Challenges for Physics-Based Free Energy Methods. Free Energy Methods in Drug Discovery: Current State and Future Directions, Chapter 3, pp. 67-107, 2021, ACS Symposium Series, Vol. 1397, 9780841298064. 10.1021/bk-2021-1397.ch003 . hal-03666161

HAL Id: hal-03666161

<https://hal.science/hal-03666161>

Submitted on 12 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Perspective on the SAMPL and D3R Blind Prediction Challenges for Physics-based Free Energy Methods

Nicolas Tielker,¹ Lukas Eberlein,¹ Oliver Beckstein,² Stefan Güssregen,³
Bogdan I. Iorga,⁴ Stefan M. Kast,¹ and Shuai Liu^{5,*}

¹Physikalische Chemie III, Technische Universität Dortmund, 44227 Dortmund, Germany

²Department of Physics and Center for Biological Physics, Arizona State University, P.O. Box 871504, Tempe, AZ 85287-1504, United States

³R&D Integrated Drug Discovery, Sanofi-Aventis Deutschland GmbH, 65926 Frankfurt am Main, Germany

⁴Université Paris-Saclay, CNRS, Institut de Chimie des Substances Naturelles, 91198 Gif-sur-Yvette, France

⁵XtalPi Inc., 245 Main Street, Cambridge, MA 02142, United States

*E-mail: shuailiu25@gmail.com

Abstract

Solvation and binding thermodynamics of a drug-like molecule is quantified by the respective free energy (FE) change that governs physical properties like $\log P/\log D$ and binding affinities as well as more complex features such as solubility or permeability. The drug discovery process benefits significantly from reliable predictions of FEs, which are hence a key area for the theoretical and modeling community. Despite the clear physical background rooted in statistical mechanics, the desired accuracy goal is hard to achieve. Current modeling methods still need to be improved in various areas related to the FE problem, such as the quality of force fields and quantum-mechanical approximations, the efficiency of sampling algorithms as well as the robustness of computational workflows. In this context, blind prediction challenges, where participants are tasked with testing their computational methods and workflows on compound property predictions without knowing the experimental data, are excellent testbeds to evaluate and improve the modeling methodology. SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) and Drug Design Data Resource-Grand Challenges (D3R-GCs) represent widely known initiatives demonstrating how the “blind

prediction” concept boosts the development of FE predictions. In this chapter, we summarize the status of recent SAMPL and D3R-GCs from the point of view of long-time participants, with the aim of providing the community with a collection of datasets and references.

Introduction

Reliable predictions of thermodynamic properties of solvated molecules are an important requirement when it comes to applying methods of computational chemistry in a drug discovery project. Owing to the complexity of the problem, calculating ensemble properties in a dense, interacting, many-body system, molecular-level approaches are in practice restricted to only a few tractable properties with pharmaceutical and physiological relevance. These encompass, for instance, partition and distribution coefficients of drug-like molecules between water and a non-aqueous phase, $\log P$ and $\log D$ at specified pH, which are related to bioavailability and measure the concentration ratios of purely neutral ($\log P$) or neutral and charged species ($\log D$) between the phases. Another example is the binding affinity of a ligand interacting with a protein target under idealized, e.g., purely aqueous environmental conditions, ignoring the complex interplay with other cellular components. Even under these simplified conditions that merely mimic narrow aspects of complete biological systems, computational chemistry methods are pushed to their limits in terms of speed and accuracy because the target quantity, the (standard) free energy change (i.e., Gibbs or Helmholtz energy, depending on external conditions), which is directly linked to the equilibrium constant of the phenomena mentioned above, is challenging to compute.

From a statistical-mechanical point of view, free energy (FE) calculations are difficult as the FE is directly related to the system's partition function and can be expressed as an ensemble average over the Boltzmann factor of the Hamiltonian, leading to large uncertainty when it is obtained directly from molecular dynamics (MD) over typically accessible time scales. The simulation time to sample all states sufficiently is correlated with the size of the system. For example, modeling ligand-biomolecule binding involves sampling of all possible states of protein side chains, water in the binding site, etc., which requires more simulation time compared to small systems where, for instance the solvation free energy (SFE) shall be computed. Consequently, special FE simulations methods have been developed to allow for sufficiently accurate predictions. These methods form the basis of atomistic calculations of host-guest and protein-ligand binding constants or partition coefficients, commonly employing empirical molecular potential energy functions ("force fields", FF) for modeling interactions, which represent another layer of uncertainty besides sampling errors originating from finite simulation times. The FF quality is strongly correlated with the accuracy of FE predictions. For FF development, generalizability and quality need to be balanced, for instance with respect to combining small molecule FFs with water and protein interaction models.

Quantum mechanics (QM) plays an important role for all aspects of modeling those processes with biological relevance, not only for the determination of FF parameters, but also for phenomena where covalent bonding patterns change. This might happen with targets that covalently bind to a ligand molecule, an extremely challenging area of computational biophysics and chemistry, but is of immediate importance already for small molecule chemistry itself. For instance, electronic polarization induced by the solvent changes its energetics. Furthermore, knowledge of the protonation state is an important ingredient as the pH-dependent ensemble determines the population of different forms and modulates partitioning properties or binding affinities. Protonation patterns in a given ionization state might also vary, a phenomenon known as tautomerization that grows exponentially in complexity with the number of titratable sites. Within physics-based modeling approaches, pro-

tonation and tautomerization state calculations require QM-based FE methods for the determination of acidity constants, pK_a , which add another layer of uncertainty, as the electronic structure depends on levels of theory, basis sets, and solvation models that polarize the wave function of the species. Of course, each protonation state is additionally characterized by a conformational ensemble to be sampled adequately, further complicating the computational workflow.

Obviously, the challenges and uncertainties mentioned above have a clear methodical interconnection when it comes to modeling protein-ligand binding and characterization of small molecule physical properties as indirect measures for bioavailability. Given a detailed protein target model, a realistic atomistic modeling workflow would start with the determination of a ligand's most probable protonation state at a specified pH, followed by assignment of an appropriate FF to all components and sampling by FE simulation methods to determine (relative or absolute) affinities, i.e., binding constants and the corresponding standard binding FE. In parallel, the small molecule protonation features would be required for calculating $\log P$ and $\log D$ between water and an organic solvent, the latter case usually under the assumption that charged species do not partition into the nonaqueous phase.

From the point of view of practitioners and method developers in the respective fields of computational biophysics and (medicinal) chemistry the assessment of the application domains and quantitative model uncertainties is of utmost importance. In this context the SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) series of blinded prediction challenges provided a timely platform for unbiased evaluation of computational methods for liquid phase FE calculations, addressing all key quantities discussed in the preceding paragraphs. Blind challenges provide an excellent format to examine the quality of computational modeling methods. When participants submit predictions without knowing the answer before they perform the calculations the community is able to examine the robustness of specific methods and to resolve confusions in the field, in which methods usually perform better with datasets from within a specific compound domain while being far less predictive for novel datasets. In other words, blind challenges serve to identify and, if possible, to prevent selection bias and model overfitting.

There are several examples of blind challenges in the field of computational chemistry and biophysics, such as the "Critical Assessment of the Structure or Proteins" (CASP) and the Cambridge Crystallographic Data Centre's blind tests of small molecule crystal structure prediction. In the context of FE predictions, besides SAMPL, the D3R (Drug Design Data Resource) organized the Grand Challenges (GCs), continuing on an earlier predecessor, the Community Structure-Affinity Resource (CSAR) [1–4]. In this chapter, we focus on discussing SAMPL and D3R-GCs with the hope that readers will get an overview on the current status of FE predictions within blind challenges in recent years and that this opens a door for future discussions on how to improve the accuracy of FE predictions.

Beginning with SAMPL, currently funded by NIH and organized by David L. Mobley (UC Irvine) with co-investigators John D. Chodera (MSKCC), Bruce C. Gibb (Tulane), and Lyle Isaacs (Maryland), the first challenge was organized by a research group at Stanford University and scientists at OpenEye Scientific Software. Previous organizers also included several other academic researchers [5]. Especially J. Peter Guthrie (Western University) played a key role in curating SFE data from the literature, which made the first few SAMPL challenges possible. As SAMPL7 recently completed and SAMPL8 is ongoing,

we will focus our discussions here on the fully disclosed challenges SAMPL0-6 while including currently available results for SAMPL7.

Table 1. Summary of SAMPL and D3R-GC Challenges

<i>Challenge</i>	<i>Years</i>	<i>Overview paper</i>	<i>Participant papers</i>	<i>Challenge data</i>
SAMPL0	2007	[6]	[6]	[7]
SAMPL1	2008	[8]	[9-14]	[15]
SAMPL2	2009	[16,17]	[13,18-26]	[27]
SAMPL3	2011-2012	[28,29]	[30-45]	[46]
SAMPL4	2013-2014	[47-51]	[52-73]	[74]
SAMPL5	2015-2016	[75-78]	[79-103]	[104]
SAMPL6	2017-2018	[105-107]	[108-144]	[145]
SAMPL7	2020-2021	[146,147]	[148-154]	[155]
SAMPL8	2021-2022			
D3R-GC1	2015-2016	[156]	[157-169]	[170]
D3R-GC2	2016-2017	[171]	[172-194]	[195]
D3R-GC3	2017-2018	[196]	[197-205]	[206]
D3R-GC4	2018-2019	[207]	[208-222]	[223]

D3R was funded by NIH through the years 2014 to 2019 and the project was located at the University of California San Diego (UCSD) where it was co-directed by Rommie Amaro and Michael Gilson. For D3R, we will concentrate on GCs, and only briefly mention the Continuous Evaluation of Ligand Protein Prediction (CELPP), a weekly organized pose prediction challenge, and other tasks that were solely focusing on pose predictions within GCs. The first GCs was named GC2015, where 2015 indicates the year of the challenge. Subsequent GCs were named sequentially. In this chapter, for convenience, we denote GC1 to mean GC2015. A summary of SAMPL and D3R challenges is collected in Table 1, and we will break down the discussions based on the type of properties listed in Table 2. With the help of the organizers of the previous challenges we collected input data from past challenges in publicly accessible GitHub repositories as indicated in the column *Challenge date* in Table 1; these data are made available under a permissive license.

Table 2. Type of Properties Predicted in Different Challenges

<i>Property</i>	<i>SAMPL</i>							<i>D3R</i>			
	0	1	2	3	4	5	6	7	GC1-4	CSAR	CELPP
Solvation free energies	■	■	■	■	■						
$\log P / \log D$						■	■	■			
pK_a							■	■			
Tautomer ratio			■								
Host-guest				■	■	■	■	■			
Protein-ligand affinities		■		■	■			■	■	■	■

Aqueous Solvation Free Energy (SAMPL1-SAMPL4)

The aqueous solvation free energy (SFE) ΔG_s , also known as gas to liquid water transfer energy or hydration free energy, is the free energy to move one solute molecule from the gas phase to the liquid water phase at a given standard state. The calculation of solvation free energies was an early focus of the SAMPL challenges because prediction of a complex thermodynamic property for which good experimental data were available, or could be generated, was considered a good metric [28]. For the more recent log P and log D challenges (see below), SFE prediction remained a fundamental theme because a common approach to predicting log P consists of computing the difference between the SFE in a solvent such as 1-octanol or cyclohexane and the SFE for water (i.e., the hydration free energy). The experimental data sets that were generated for the SAMPL SFE challenges have become useful community resources such as the Minnesota Solvation Database [224] and FreeSolv [225], which provide test and training data for novel SFE methods.

The **SAMPL0** challenge was organized as an informal blind challenge to predict the hydration free energy of 17 compounds that contained multiple interacting polar groups together with a few monofunctional molecules for comparison with their more complicated analogues [6]. The dataset was meant to challenge current SFE prediction methods and spanned -11.01 ± 0.2 kcal/mol to $+1.07 \pm 0.2$ kcal/mol; the SFE here and in the following is quoted for the Ben-Naim standard state of 1 M gas to 1 M aqueous solution and all errors were estimated [6]. A Poisson-Boltzmann (PB) implicit solvent, single conformer approach achieved a prediction accuracy of 1.87 ± 0.03 kcal/mol and 2.57 ± 0.03 kcal/mol for different Born radii sets. An explicit solvent alchemical free energy MD approach with the *AMBER GAFF* force field and different charge models yielded predictions with RMSEs ranging from 1.33 ± 0.05 kcal/mol to 2.05 ± 0.05 kcal/mol. For both methods, weaknesses with particular chemical groups (benzamides for PB, esters for MD) were uncovered. Although limited in scope, the first SAMPL exercise demonstrated the usefulness of true predictions for well-defined physicochemical properties.

The **SAMPL1** data set contained 63 molecules biased towards a drug-like character (including an important number of nitro derivatives and pesticide compounds) for which aqueous solvation free energies could be extracted from the literature [8] although due to problems in data preparation only 56 could be used by all participants. The experimental ΔG_s ranged from -20.2 ± 1.9 kcal/mol to -1.5 ± 0.1 kcal/mol. In this SAMPL challenge, only a handful of groups were invited to participate. Approaches included quantum chemical methods with various approximations for the solvent [9,14], Poisson-Boltzmann calculations [11,12], and molecular dynamics (MD) all-atom explicit solvent alchemical FE calculations [10]. The overall accuracy of the predictions remained modest with RMSEs between 2.4 and 3.6 kcal/mol [8], even though some of the tested methods had previously achieved accuracies of 0.5 kcal/mol in non-blind calculations. SAMPL1 also highlighted the challenge to obtain reliable experimental SFEs, given that new data are rarely published, and errors are often not or only insufficiently included. It also showed that errors in the experimental data may be uncovered by comparison to the computational predictions [14].

Given that SAMPL1 had demonstrated the need to test established methods in blind predictions and motivated the field to improve existing approaches, **SAMPL2** was launched

as an open challenge for all interested participants. In the SFE prediction challenge, organizers made available 23 blinded compounds (the “obscure” dataset for which predictions had to be made) as well as 17 compounds for which SFE were either communicated to the participants (the “explanatory” dataset) or not known (the “investigatory” dataset) [16]. These compounds formed a relatively diverse group, with fragment-like structures in the “explanatory” and “investigatory” datasets, and more drug-like structures in the “obscure” dataset, the latter having uracil and paraben derivatives significantly over-represented. The experimental SFEs of the obscure dataset ranged from -25.4 ± 0.2 kcal/mol to -7.0 ± 0.1 kcal/mol. Implicit solvent methods with single conformers predicted with RMSEs in the range 2.0–2.2 kcal/mol, implicit solvent methods with multiple conformers showed overall best agreement with experiment with RMSEs within the range 1.5–1.7 kcal/mol, and all-atom explicit solvent MD methods obtained RMSEs in the range 2.4–2.5 kcal/mol. Flexible molecules such as sugars proved to be particularly challenging, possibly due to the difficulty of sufficiently sampling all relevant conformers. Prediction of the rank order of multiple halogenated uracil compounds was also difficult for many methods, possibly due to a lack of inclusion of polarization effects, which might have led to underprediction of SFEs for halogenated compounds, including hexachlorobenzene in the explanatory dataset. In general, methods performed better for smaller, more rigid compounds with less negative SFEs [16].

The **SAMPL3** challenge [28] focused on chlorinated and especially polychlorinated compounds because of the difficulty to obtain new datasets for drug-like compounds, the possibility to explore series of substitutions on a common scaffold, the importance of polychlorinated compounds in atmospheric and environmental chemistry, and lastly, the surprisingly poor performance of many approaches in SAMPL2 for two chlorinated compounds, hexachloroethane and hexachlorobenzene. The dataset of 36 chloro-organic compounds was organized in three distinct chemical subsets with varying numbers of chlorine substituents. Subset 1 contained polychlorinated derivatives of ethane (up to six chlorine atoms) with all possible substitution combinations. Polychlorinated derivatives of biphenyl and of dibenzo-*p*-dioxin were present in subsets 2 and 3, respectively, with aromatic ring systems bearing different substitution patterns for the chlorine substituents. The experimental SFEs in this dataset covered the fairly narrow range from -4.61 ± 0.25 kcal/mol to $+1.87 \pm 0.1$ kcal/mol, with a standard deviation of only 1.42 kcal/mol about the mean, which made it difficult for any prediction to distinguish itself from the hypothetical null model of just choosing the (experimental) mean. More than half of the predictions consisted of explicit solvent MD approaches [31,37,39], with the remainder being implicit [37,38,43] and semi-explicit [43]. Overall, the challenge confirmed that the accuracy of most methods decreased with increasing number of chlorine substituents; an exception was an MD-based method with an RMSE of 1.2 kcal/mol for which the authors explicitly parameterized chlorine for compounds with four or more Cl substituents on an aromatic ring [39]. The accuracy of the experimental SFE for two polychlorinated biphenyls was questioned [24,39] but could not be independently checked due to the difficulty of obtaining new data. Due to its focus on chloroderivatives and narrow dynamic range, the dataset was not well suited to assess the overall ability of methods to correctly predict SFEs but pointed out specific problems with chlorines, namely transfer energies that were generally too positive, possibly due to under-estimation of polarizability [28].

SAMPL4 [50] was the last challenge with a hydration free energy prediction because no new experimental measurements of SFEs had become available and even obtaining sufficiently obscure data from the literature had become difficult. The data set provided for the

SAMPL4 challenge consisted of a diverse set of 52 small molecules divided into two subsets: a blind set of 24 compounds, whose hydration free energies had not been published but were known to the SAMPL4 organizers, and a supplementary set of 28 compounds, for which hydration free energies were available in the literature or could be calculated from literature data. After reassessment of the reliability of the experimental data and other problems, 47 compounds were selected to be retained for assessment of the predictions [50,51]. The dataset contained a varied selection of small molecules such as branched alkanes and alkenes, methoxyphenol/guaiacol and their chlorinated derivatives, cyclohexane derivatives, anthracene derivatives, and a number of polyfunctional and flexible compounds. Many compounds contained multiple oxygen or hydroxyl substituents, capable of hydrogen bonding, resulting in a large SFE dynamic range from -23.62 ± 0.31 kcal/mol (mannitol) to $+0.14 \pm 0.10$ kcal/mol (cyclohexene) [51]. Almost half of all predictions were made with alchemical free energy MD or Monte Carlo methods with explicit solvent [50,63,65,68,69,71,73], followed by single conformation or multi-conformer implicit methods (about one third) [61,62,64,65,67,72]. Hybrid solvent [72] and other methods, including a quantum mechanics (QM)+implicit solvent approach [66] provided a diverse set of strategies. The top performing methods with $RMSE \leq 1.23$ kcal/mol included the QM approach with implicit solvent and functional group corrections [66], a single conformation PB approach [61], and alchemical explicit solvent MD calculations with the *AMBER GAFF* force field and updated hydroxyl parameters [50]. MD with a polarizable force field was less successful than many of the classical force field MD approaches [68]. The very low SFE of mannitol was predicted as too positive by several approaches with a high average unsigned error across all submissions of 3.94 kcal/mol [50]. For example, the *COSMO-RS* implicit approach, one of the consistently well performing approaches on previous SAMPL datasets [14,21,38], predicted mannitol as its sole outlier with an error of +6 kcal/mol, resulting in an RMSE of 1.46 kcal/mol, which would have been 1.18 kcal/mol if mannitol had been omitted [67]. SAMPL4 showed a diverse set of methods converging on robust, predictive protocols with SFE RMS errors below 1.5 kcal/mol even though difficulties with larger and polyfunctional molecules remain [50].

Overall, the SFE prediction challenges indicated that, over the years, prediction accuracy improved moderately and by the time of SAMPL4 (2013), the state of the art for small molecule hydration free energy predictions was in the 1.2–1.5 kcal/mol RMSE range, using a variety of implicit and explicit approaches. Of course, it is not surprising that classical MD simulations in particular showed limited accuracy when predicting SFEs because of the large difference in polarization between the gas phase and the aqueous environment that is not captured in classical force fields. The lack of suitable experimental data has made it difficult to assess if any improvements have been made since SAMPL4. SAMPL3 with its focus on a small and specific subset of chemical space showed that all methods will have likely blind spots that, if not accounted for, will diminish their accuracy. The general difficulty to provide a representative breadth of molecular targets (for instance, charged species have been absent) cautions to not too literally take the SAMPL SFE results as the ultimate assessment of method accuracy. It is, however, encouraging that the overall improvement in accuracy in SFE predictions from SAMPL0 to SAMPL4 was achieved for increasingly more challenging molecules, suggesting that some lessons were learned in previous challenges and applied to improve the computational methods.

Tautomers, $\log P$, $\text{p}K_{\text{a}}$, $\log D$ (SAMPL2, SAMPL5-SAMPL7)

Starting with SAMPL5, the SFE challenges were replaced with challenges to predict partition coefficients ($\log P$) or distribution coefficients ($\log D$) between water and an organic phase such as octanol or cyclohexane. Partly this shift reflected the difficulty in acquiring new unknown SFE datasets, but at least as important was the insight that SFEs do not necessarily make a good test case for typical applications such as drug binding to a receptor. As mentioned above, SFEs include a transfer between two environments (a gaseous and liquid phase) with a strong shift in polarization whereas this shift is less pronounced in a binding process or in moving between two condensed phase environments. Therefore, partition coefficient predictions ought to make a more suitable test for many methods such as ones based on classical force fields.

We begin this section with the $\log D$ perspective for the distribution of species between water and a nonaqueous phase first introduced during SAMPL5, where participants were asked to predict the coupled problem of protonation equilibria, $\text{p}K_{\text{a}}$ (assumed to be relevant for the aqueous phase only), and partitioning of neutral states, $\log P$. Following the historical context, we then focus on the separate parts of the SAMPL6 challenge, here designated “SAMPL6” and “SAMPL6.2” targeting $\text{p}K_{\text{a}}$ and $\log P(\text{octanol-water})$, respectively. SAMPL7 revisited the problem of determining $\log D$ by challenging participants with $\text{p}K_{\text{a}}$ and $\log P(\text{octanol-water})$ predictions of the same set of compounds, allowing for calculation of $\log D$ to be compared to experimental data. The organizers of the SAMPL5-SAMPL7 challenges provided participants with a set of protonation microstates, i.e. tautomers, allowing for treating the macroscopic problem on the basis of statistical populations of individual microstates. Participants were allowed to add their own microstates to the list. This approach culminated in the requirement to explicitly submit relative free energies between all microstate pairs during SAMPL7, although we only briefly elaborate on this challenge as not all relevant publications are accessible at the time of writing. During these later challenges, it was implicitly assumed that tautomerization free energies are computable and necessary for inclusion in the expressions for macroscopic quantities. Indeed, the quality of these individual microstate free energies was only tested once, during SAMPL2. We therefore review these results at the end of this section.

$\log D$

The SAMPL5 challenge tasked the participants with predicting the cyclohexane-water distribution coefficient of 53 compounds. In their analysis of the SAMPL5 challenge results, the organizers characterized methods that yielded good accuracy as measured by RMSE and MAE (mean absolute error) as well as correlation measured by R^2 and Kendall’s Tau [75]. The best-performing submission of this challenge achieved an RMSE of 2.1 $\text{p}K$ units by combining $\log P$ and $\text{p}K_{\text{a}}$ predictions under the common assumption of negligible presence of charged species in the organic phase. Due to the difficulty of predicting aqueous $\text{p}K_{\text{a}}$ values, several groups decided to submit $\log P$ predictions only [91]. The RMSE of 2.1 is still not close to so-called chemical accuracy of about 1 $\text{p}K$ unit, especially considering the small dynamic range of the measured distribution coefficients. Only two other models could also be considered consistently well-performing, i.e., belonging to the top 10 for at least three out of the four metrics mentioned above [81,88]. One of these models only submitted $\log P$ values while the other included the ionization behavior. Additionally, only

five models in total achieved RMSEs equal to or smaller than 2.5, which emphasizes the peculiar difficulty of predicting distribution coefficients.

An additional surprising result of the SAMPL5 challenge was that even a “null hypothesis” which assumes a $\log D$ of every compound of exactly zero would have performed better than any submitted model [75]. These disenchanting results were a major reason for the design of the subsequent SAMPL6 and SAMPL7 challenges, where $\text{p}K_{\text{a}}$ and $\log P$ predictions were investigated separately to identify the major sources of predictive uncertainty.

$\text{p}K_{\text{a}}$

During the SAMPL6 challenge a broad range of conceptually different empirical and physics-based computational methods had been used to predict $\text{p}K$ values, as presented and discussed with respect to their performance in detail in the original overview paper [226]. To provide some context for the results of the SAMPL7 challenge, the main results are summarized here. The empirical approaches used during SAMPL6 can be divided into three categories, Database Lookup (DL), Linear Free Energy Relationship (LFER), and Quantitative Structure-Property / Machine Learning (QSPR/ML) approaches. The physical approaches can be divided into pure quantum-mechanical (QM) methods, QM with a linear empirical correction (QM+LEC) to account for the Gibbs Energy of the proton in solution or potential systematic errors caused by the chosen method, and QM in combination with molecular mechanics (QM+MM). Generally speaking, the empirical methods require significantly less computational effort than their physics-based counterparts once they are properly parametrized.

The best-performing models included four empirical and one QM-based model. These five methods were able to predict the acidity constants of the SAMPL6 challenge compounds to within 1 $\text{p}K$ unit upon submission. In fact, while most empirical models – except for the DL and two of the five QSPR/ML approaches – were able to predict the acidity constants to within about 1.5 $\text{p}K$ units, the range of predictions was much wider for the QM-based models. Unlike in the SAMPL7 challenge, the number of submissions was not limited here, and many groups in this category submitted multiple predictions to test the performance of different variations using the same basic methodology, encompassing, e.g., different levels of theory, model parameters, or conformational ensembles.

The high-performing empirical models included both LFER methods, such as *ACD/pKa Classic* (submission ID *xmyhm*) and *Epik Scan (nb007)*, and QSPR/ML methods such as *MoKa (nb017)* and *S+pKa (gyuhx)*, all performing with root mean square errors (RMSE) between 0.73 and 0.95 $\text{p}K$ units [227–230]. All these well-established tools for the prediction of physicochemical properties therefore demonstrated their reliability and high quality.

Among the physics-based models, the most straightforward approach involved calculation of the acidity constants without any empirical corrections and using the experimental value for the Gibbs energy of solvation of the proton [231]. One group applied different calculation schemes to the compounds of the SAMPL6 challenge that differed in the use of gas phase and/or solution phase geometries as well as additional high-level single point gas phase calculations [108]. While the results achieved by this method were quite promising, with an initial RMSE of 1.77 $\text{p}K$ units (*ryzue*) that could be improved to 1.40 by including

a standard state correction and a different value for the free energy of the proton, the authors also showed the effectiveness of a simple linear regression scheme to correct the raw acidity constants. In this case the RMSE of the best-performing model decreased further from 1.40 to 0.73 pK units after regression.

This type of empirical correction was in fact used by most QM-based approaches, including the best-performing method of the SAMPL6 challenge, improving some systematic deficiencies of the QM level of theory and basis sets, while the proton's Gibbs energy of solvation was effectively an adjustable parameter [226]. The best-performing QM+LEC method, *xvxzd*, achieved an RMSE of 0.68 pK units during the challenge using the *COSMO-RS* solvation model. This also made it the best-performing model overall, and the two methods *ykgga* and *8xt50* that used the same solvation model were just slightly worse, with RMSEs of 1.01 and 1.07 pK units, respectively [116,226,232].

A QM+LEC method using a different, integral equation-based solvation approach, *EC-RISM*, only achieved an RMSE of 1.70 pK units for the submitted model (*nb001*), but a post-submission optimization of the conformer generation workflow and use of exact QM-derived electrostatic interactions improved the RMSE to 1.13, which is more in line with the other high-performing QM+LEC methods [124]. Another solvation model, performing not quite as well as *COSMO-RS* during the SAMPL6 challenge, is the CPCM implicit solvation model used by one group [123]. For these two models, differing only by training either a single LEC for all compounds (*35bdm*) or two separate LECs for deprotonation of neutral compounds to anions and deprotonation of cations to neutral compounds (*p0jba*), the RMSEs for the full set of compounds were 2.04 and 1.95 pK units, respectively. Nevertheless, the results obtained by a pure QM model using the same method were significantly worse, with an RMSE of 3.74. All these results show that accurate values can be predicted when using the QM+LEC approach with different solvation models.

A slightly different approach was used by one participant (*0w/z0*) where QM calculations of the Gibbs energy of deprotonation and thermodynamic integration, an MM method, were combined to calculate the difference of the solvation free energies between the acid and its conjugate base [110]. This approach yielded only average results during the SAMPL6 challenge, with an RMSE of 2.89 for the macroscopic acidity constants calculated from the submitted microscopic acidity constants, excluding two compounds, SM14 and SM18 from the analysis as they exhibited multiple values too close to each other.

In the **SAMPL7** challenge a consistent set of molecules was investigated for their log *P* and p*K*_a except for two compounds, SM28 and SM33, for which no ionization could be detected within the experimental pH range. Probably due to the more complex challenge design, which required the submission of microstate free energies relative to some reference state from which macroscopic acidity constants were computed, only 9 blind predictions were submitted. Of the blind predictions only one method had a performance similar to the best models of the SAMPL6 challenge (*EC-RISM*, a QM+LEC method) with an RMSE of 0.72 while the next best model had one of 1.82 (*IEFPCM/MST*) [147]. Most other methods, surprisingly, had RMSEs of 2.90 or higher, a stark contrast to the various methods with good performance in the SAMPL6 challenge.

It is possible that the challenge design also led to lower diversity of submitted methods. While in the SAMPL6 challenge different empirical and physics-based models were used,

here almost all submissions were pure QM-based models without an empirical correction. These were also not among the best-performing models in the SAMPL6 challenge, where empirical models and QM+LEC approaches had lower average errors.

log *P*

The first calculations of partition coefficients were already necessary during the **SAMPL5** challenge to determine cyclohexane-water distribution coefficients, but octanol-water partition coefficients only came into focus in the second part of the SAMPL6 challenge. In their analysis of the **SAMPL6** challenge results, this time the organizers compared the 20 best methods, as measured by their RMSE, MAE, and correlation measures R^2 and Kendall's Tau [105]. The seven consistently well-performing methods included three different empirical and four QM-based approaches, three of which employed different methods. With RMSEs between 0.38 (*hmz0n*) and 0.54 (*dqxx4*) these models showed excellent performance on the kinase inhibitor fragments of the SAMPL6 challenge. The QM models differed by use of their solvation model, using either *COSMO-RS*, *EC-RISM* or *SMD* to represent the solvent [128,129,142]. The consistently well-performing empirical models on the other hand were three different QSPR models [105,126].

While some of the empirical and QM-based methods appeared to be able to predict water-octanol partition coefficients with good accuracy, no MM-based or "mixed" model showed up in the list of consistently well-performing methods. Although the best MM model achieved a very good RMSE of 0.74 by using the modified Expanded Ensembles method (*nh6c0*), this submission reached only rank 28 when compared with all submitted methods [131]. This absolute rank should not be used to judge model performance as participants were allowed to submit more than one prediction using only slightly varying models, thus distorting a fair comparison. In later challenges, submitting more than one "ranked" model was no longer allowed, while it was still possible to submit several "unranked" predictions for internal comparison. The mixed models are less consistent, with some breaking into the top 20 as measured by performance, but failing to achieve the necessary correlation, while for others the opposite is the case. For example, the third-best performing model with an RMSE of 0.41 (*3vqbi*) ranked only at 25 for the Kendall's Tau and was thus excluded from this group [128]. On the other hand, mixed methods with good correlation did not achieve the required performance, such as (*5krdi*) [125].

In the **SAMPL7** challenge 31 blind submissions were made to predict the partition coefficients of 22 different *N*-acylsulfonamides and structurally similar compounds [147,233]. Performance on these compounds was generally worse, with the five consistently well-performing models achieving RMSEs of 0.58 (*TFE-MLR*) to 1.55 (*TFE-NHLBI-TZVP-QM*). Due to the lower number of submissions, the consistently well-performing models were chosen from the 10 models performing the best in every category instead of 20. A possible cause for the worse performance on average is the different makeup of the two data sets: while in the SAMPL6 challenge various kinase inhibitor fragments were investigated, here an *N*-acylsulfonamide motif was present in all molecules.

While not an official part of the SAMPL7 challenge, the compounds' octanol-water distribution coefficients were measured, and for groups that had submitted both partition coef-

ficients and acidity constants it was possible to calculate distribution coefficient predictions. Here, the main problem was a lack of submissions as only six submissions fit these criteria. Still, the performance was significantly better than during the SAMPL5 challenge with the best model (*TFE IEFPCM MST + IEFPCM/MST*) predicting the log *D* to within 1.27 [147]. Furthermore, five of the six models achieved RMSEs of less than 2.30 which was significantly better than what was achieved during the SAMPL5 challenge. It is important to note that, in this case, the organic phase was represented by octanol, instead of cyclohexane, which may be part of the reason for the significantly better performance, as octanol is more commonly used for investigating partitioning behavior. Because of this, there is more data for octanol to train, e.g., empirical models, but also force fields and empirical corrections for other methods. A larger number of acidity constant predictions could lead to more insight into the distribution behavior because unfortunately many of the best-performing log *P* submissions had no corresponding *pK_a* model.

Tautomers

Returning to the problem of microstate thermodynamics mentioned above, a brief review of the SAMPL2 tautomer challenge results is in order [16,17]. The dataset contained 68 tautomer pairs for which 7 participants submitted 20 models with predicted tautomerization free energies. Analogously to the SFE part of the challenge, it was divided into three components: “obscure”, “explanatory” and “investigatory”. These consisted of provided, undisclosed, and unknown aqueous phase experimental equilibrium constants respectively. All models were QM-based, mostly taking only a single conformation per molecule into account and differing in the solvation models used. Two different approaches were applied by the participants: direct calculation of the FE differences in solution, e.g., with the *COSMO-RS* [234] and *EC-RISM* [20] solvation models, and calculation of solvation free energies, e.g. via *COSMO-RS* [21], *IEF-MST* [26] and *SM_x* (*x*=8, 8AD, D) [25], together with tautomerization energies in the gas-phase. The latter allows for the calculation of SFE and gas-phase reaction free energies at different levels of theory, which reduced the RMSE of *COSMO-RS* based submissions from 3.2 to 2.9 kcal/mol. The best performing model with an overall RMSE of 2.0 kcal/mol was *EC-RISM*, which was able to predict the correct sign for all except one tautomer transition. However, performance for different components of the dataset differed widely, indicating large uncertainty for either computational predictions, experimental data, or both. The performance of the individual models for the respective tautomer pairs is described in detail in the overview paper [16]. As an example for imbalanced predictions, several submissions e.g. *IEF-MST* and *SM_x* showed discrepancies in predictive power between the lactim-lactam tautomerism of six-membered rings and 1,2-diketones (pairs 1-8), for which they yielded good results, and five-membered pyrazolones and isoxazolones (pairs 10-16) that performed less satisfactorily [25,26]. Other discrepancies were revealed for *COSMO-RS* and *EC-RISM* in post-submission analyses [20,234,235]. In fact, the unknown experimental uncertainty of tautomerization free energies in water makes an analysis of computational predictive power very difficult. Similarly, due to the experimental difficulties in obtaining reliable numbers, later challenges only

implicitly accounted for the tautomerization problem by challenging participants with highly tautomerizable species.

General remarks on tautomers, $\log P$, pK_a , $\log D$ (SAMPL2, SAMPL5-7)

Viewing the problem sets of this section in context, one recognizes considerable progress in the field on one hand, but persistent issues on the other. There is clear improvement in physics-based acidity predictions that require QM calculations, almost on par with empirical approaches. Predicting partitioning thermodynamics, though, suffers from apparent inconsistencies. These inconsistencies could be related to diverse chemistries among different datasets that could result in worse-performing force field parameters in one context and high quality in another. At the same time, the tautomer problem has so far not been adequately explored due to limited high-quality data availability. This lack of reference data could hamper predictions that heavily rely on the adequate calculation of populations underlying macroscopic properties. Taken together, successful calculations of derived properties such as distribution coefficients must be based on well-performing models in different areas. This means that further challenges addressing diverse chemistries in a variety of solvents, including ionization processes, will guide the way towards improvement. This is explored in the SAMPL8 challenge, which involves the calculation of acidity constants in water and distribution coefficients $\log D$ between water and a broad variety of polar and apolar organic solvents.

Host-guest (SAMPL3-SAMPL7)

SAMPL3-SAMPL7 contained host-guest challenges. In each challenge, there were several supramolecular hosts and a group of guest molecules associated with each host. The goal was to predict the binding free energies of the guest bound to the host. The host-guest system had its own advantages, compared to the solvation free energy challenges: The model could mimic the ligand-receptor interaction as well as the water dynamics in the binding pocket and compared to the protein-ligand challenges, the system was relatively small, so it would have less errors introduced by inadequate sampling of the conformational states of proteins. This meant that participants could focus more on reducing the errors introduced by other components of the simulation protocols. There were several methods to collect the experimental binding affinity including NMR, UV/Vis spectroscopy and ITC, with the experimental conditions such as buffer concentrations varying across challenges. Different computational methods were applied in predictions such as pathway methods like double decoupling methods [236], binding energy distribution analysis method (BEDAM) [237], orthogonal space random walk (OSRW) [238], attach-pull-release (APR) [239] method and metadynamics [240]. Thermodynamic integration (TI) [241], Bennett acceptance ratio, (BAR) [242] methods are often included in those pathway methods to calculate the energy; end point methods like MM/PBSA [243], MM/GBSA [243], solvent interaction energy (SIE) [244], Mining Minima (M2) [245] and Movable Type [246] were also employed. From the perspective of energy evaluation, participants used force fields like *OPLS*, *GAFF*, *MMFF94*, *CHARMM*, *Openff* or *CGenFF*, fixed-charge models like *RESP/AMI-BCC* as well as polarizable force fields like *AMOEBA*. QM methods like DFT have also been used to evaluate the energy. For solvent models, pathway methods and end

point methods were usually coupled with either explicit solvents or implicit solvents respectively. For conformational sampling, MD was commonly used although other methods like docking and MC methods were also used to generate conformations.

Highlights and Lessons Learned for Each Challenge

SAMPL3 [29] was the first time that the host-guest system was introduced. There were 19 submissions provided by 10 groups. For the datasets, there were three hosts, H1, H2, and H3 with 7, 2, and 2 guests respectively. The hosts were all cucurbit[n]uril (CB), H1 was acyclic cucurbit[n]uril; H2 and H3 were macrocyclic cucurbit[n]uril. The overall performances were relatively poor with high RMSE and low correlations or both, and none of the predictions could yield a low RMSE with high correlations. The SIE approach [244] achieved some of the most accurate results. In this approach the receptor was kept rigid after the minimization step which indicates that the more elaborate sampling methods like MD may not help improve the accuracy. This could be related to the coupling between sampling and energy evaluation, as improving sampling efficiency may sometimes amplify the effect of inaccurate energy models. Another interesting observed trend was that conceptually similar free energy estimators like TI, MBAR, and FEP provided quite different results suggesting that the details of how to set up those simulations as well as the choices of force field may strongly affect the predictions. Furthermore, the different protonation states of bound/unbound ligands may be related to the source of errors of H1 systems.

In **SAMPL4** [47] two hosts, cucurbit[7]uril (CB7) and Octa-acid (OA), with 14 and 9 guests were provided, respectively. There were 22 submissions for CB7 and 13 submissions for the OA system. Free energy simulations with classical force fields were particularly popular in this challenge and 19 out of 35 used an explicit water model, which was in contrast to SAMPL3 where the implicit solvent model was the most popular water representation. In general, the overall performance of this challenge was similar to SAMPL3; none of the methods ranked best for all error metrics while few alchemical free energy methods were among the top of all metrics in both hosts. Alchemical free energy methods calculate the free energies by employing some unphysical intermediates. In this challenge, two null models were introduced, one assigned a constant binding affinity to all guests and the other assigned -1.5 kcal/mol per heavy atom. For CB7, 18 of 22 submissions outperformed both null models in terms of correlation but fewer than half performed better in RMS. For OA, many of the methods performed equally well or better compared to null models in both the correlation and the RMS error.

For **SAMPL5** [77], there were three hosts CBClip (10 guests), Octa-acid (OA, 6 guests), Tetra-endo-methyl octa-acid (TEMOA, 6 guests) receiving 12, 21, and 21 submissions respectively. SOMD-double decoupling [98] and attach-pull-release (APR) [103] were two new methods that had not been tested in previous SAMPL challenges. They both performed relatively well and were among the top performers in all host systems. Although these both were pathway methods, they differed in the construction of the pathway and the force field employed.

In **SAMPL6** [106] three hosts OA (7 guests), TEMOA (7 guests), CB8 (13 guests) were presented, receiving 42, 43, and 34 submissions respectively. In general, the CB8 system

was more challenging compared to the OA systems. Compared to the CB systems in previous SAMPL challenges, this round was more challenging mainly because of the complexity of the guest sets. In general, simulation-based free energy methods yielded above average R^2 and classical force fields tended to overestimate the binding affinity.

SAMPL7 [146] contained three types of host systems (1) cucurbituril-derivatives, (2) Gibb deep cavity cavitands (GDCCs) and (3) modified cyclodextrins. The cucurbituril-derivatives system consisted of one host TrimerTrip with 16 guests, which received 7 submissions. The GDCCs system consisted of two hosts (OA and exo-OA) each with 8 guests and 16 submissions. The modified cyclodextrins system had 2 guests bound to 9 different hosts with 7 submissions. In this challenge, submissions with the *AMOEBA* polarizable force field had outstanding performances, indicating that the choice of force field may be the dominant factor of the accuracy for this challenge. Considering different host conformations could improve the accuracy for the TrimerTrip system.

Conclusions of which methods or parameter sets outperform others were hard to draw since (1) methods and parameters varied along challenges and, (2) more challenges were needed for statistically validated answers. Other factors associated with specific systems like (1) protonation state, (2) experimental ranges and, (3) potential host conformations were also important, as treating them differently will significantly alter the performance for given methods/parameters. Across SAMPL challenges, some trends were still worth noting such as 1. Predictions using explicit solvent free energy simulations seemed to have more consistent performance. 2. In early challenges, the QM methods without sampling were not successful, so the percentage of those submissions reduced in later challenges. 3. The overall RMSE through SAMPL3-6 did not significantly improve, suggesting the force fields still need to improve. 4. The *AMOEBA* force field performed well in the latest SAMPL7 challenge, which emphasized the importance of a polarizable model. 5. In most recent challenges, some physical methods have begun employing empirical corrections based on the specific host being considered.

Protein-ligand (SAMPL1,3,4,7, D3R-GC1-4)

The “protein-ligand” challenges related to protein-ligand interactions aimed at evaluating the performance of existing methods for reproducing crystallographic complexes (pose prediction), for classifying docking conformations according to their relative affinities (ranking) and for prediction of relative free energies in smaller, structurally homogeneous datasets (affinity). These challenges were initiated in 2008 with SAMPL1, and were continuously present in SAMPL3 (2011), SAMPL4 (2013), D3R-GC1 (2015), D3R-GC2 (2016), D3R-GC3 (2017), D3R-GC4 (2018) and SAMPL7 (2019). Several other similar challenges (e.g. CSAR [1–4], <http://www.csardock.org/>) took place in parallel during this period but will not be discussed in detail here.

SAMPL1

The **SAMPL1** challenge included two retrospective data sets focused on compounds targeting the JNK3 kinase and the urokinase. The first data set was designed by Vertex, with 52 co-crystal structures belonging to ~12 classes and K_i values spanning 3.55 pK units

(5.15-8.70). The second dataset was proposed by Abbott and included 27 co-crystal structures from ~9 classes with K_i values spanning 4.9 pK (4.3-9.2) [247,248].

Overall, 174 submissions from 54 groups were received for the three sections of the challenge: pose prediction (cross- and self-docking), virtual screening and affinity prediction. In the **pose prediction** section, the self-docking on urokinase provided predictions with median RMSD of 1.0-2.0 Å whereas on JNK3 the results were more varied, with RMSD for some predictions in the range 1.0-1.5 Å and for others >2.0 Å. As expected, the cross-docking was more challenging for both proteins, with RMSDs ranging from < 0.5 Å for a few predictions to >2.0 Å for the majority of them. It is interesting to note that in this subchallenge a human expert manually posing compounds (Marti Head) outperformed the automated methods. In the **virtual screening** section, the participants used *FRED*, *GLIDE* (with *SP* and *HTVS* parameters), *GOLD* (with the *ASP* scoring function) and *ROCS* (with and without *EON*). No significant overall differences were observed between different tools, but a smaller intra-method variance was observed for *ROCS* and *FRED* compared with the other methods. In the affinity prediction section, the results were system-dependent, with median errors of 0.7-1.9 and 0.6-3.7 kcal/mol and Kendall tau values ranging from 0.35 to 0.55 and from -0.2 to 0.43 for urokinase and JNK3, respectively [249].

SAMPL3

In **SAMPL3**, the participants of the protein-ligand binding section were required to rank 500 fragment-like potential binders of bovine trypsin according to their predicted relative affinities [250]. The main difficulty in this prospective SAMPL3 challenge was that the binding affinities were separated by less than the apparent predictive limit of the methods employed, which in most cases can be evaluated at 1 kcal/mol for intermolecular interactions in aqueous environment [250].

There were 11 submissions using different approaches: free energy decomposition scheme based on a thermodynamic cycle and empirical scores (*LISA*) [42], exhaustive search and solvated interaction energy (SIE) [35], *RosettaLigand* with protein flexibility [33], *GOLD* with increased search efficiency [34] and annealing of chemical potential in a Grand Canonical Monte Carlo (GC/MC) simulation [30]. The best predictions [34] were obtained by selecting in an initial benchmark the most adapted docking software (*GOLD*) for the target protein, with increased search efficiency (by contrast with the reduced conformational search that is a standard approach in virtual screening campaigns involving datasets of this size). This protocol provided an enrichment factor of ~10 for Top 20 compounds, with bootstrapped ROC AUC of 0.77 ± 0.07 [34].

SAMPL4

The dataset provided in 2013 for the virtual screening **SAMPL4** challenge [49] consisted of 321 drug-like compounds (305 after removal of some problematic or duplicate structures), potential binders in the micromolar range to the HIV-1 integrase. The challenge was focused on three binding sites known for this protein: LEDGF/p75, fragment, and Y3 sites. The challenge was focused on three aspects: (1) virtual screening, to identify the binders from the proposed dataset; (2) pose prediction, to predict binding modes; and (3) affinity

prediction, to predict absolute or relative affinities for 8 compounds with known binding modes. The main difficulties encountered during this challenge were the high structural similarity between active and inactive compounds and the presence of three different binding sites, with some compounds exhibiting multiple site binding.

For the **virtual screening** section there were 26 submissions from 9 research groups. The best prediction involved docking with *GOLD*, then pharmacophore/electrostatic similarity search with *MOE* for filtering, but also substantial manual intervention and expert knowledge from a person with more than 10 years of experience working on this specific target [53]. Another interesting submission used *AutoDock/Vina* docking calculations followed by BEDAM alchemical binding free energy calculations to score predictions [58]. To avoid inconsistencies, the dataset and all submissions were re-analyzed after the challenge by dropping all alternate isomers of active compounds, i.e., excluding the “inactive or very weak active” category and retaining only true actives and inactives. This reduced the number of compounds analyzed from 305 to 189, while retaining the same 56 active compounds. In these conditions, another submission that used *GOLD/ChemScore* was ranked second, with an overall ROC AUC of 0.73 and enrichment factor at 10% of 2.89 [59]. The **pose prediction** section received 12 submissions from 5 research groups. The best predictions used *XP Glide* with rescoring via *DrugScore* and MMPB/SA, *AutoDock Vina*, *Wilma* docking with SIE re-scoring or *DOCK 3.7*. The **affinity** section received 15 submissions from 4 groups. Most submissions used docking to predict affinities and submitted docking scores as “affinity” predictions, with one notable exception which used an MM-PB/SA approach [49].

D3R-GC1

The D3R Grand Challenge 2015 (**D3R-GC1**) [156] was focused on two protein targets: Heat Shock Protein 90 (HSP90) and Mitogen-Activated Protein Kinase Kinase Kinase Kinase 4 (MAP4K4). In Phase 1 the participants were asked to provide affinity predictions for 180 HSP90 ligands and pose prediction for 6 of them, as well as pose prediction for 30 MAP4K4 ligands and affinity predictions for 18 of them. In Phase 2 the participants were required to provide the same affinity predictions as in Phase 1, taking into account the additional structural data released at the end of Phase 1. The chemical structures of compounds from the HSP90 dataset belonged to three distinctive classes (benzimidazolones, aminopyrimidines and benzophenone-like), whereas the MAP4K4 dataset was more diverse.

In the **affinity ranking** section, the best-performing submissions used *DockBench*, *GOLD*, *PlantsPLP*, *rDock*, *Surflex-GRIM*, and *RosettaLigand-Omega-ROCS*. Surprisingly, information about ligand poses from phase 2 did not lead to more accurate affinity rankings. In the **pose prediction** section, 39 and 30 submissions were received for HSP90 and MAP4K4, respectively. For HSP90, half of the submissions provided rank 1 poses with median RMSD < 2 Å, but their success can be attributed more to the human expertise than to the choice of docking software, as submissions using the same software packages (e.g., *AutoDock Vina* or *Glide*) yielded differing levels of accuracy. Along the same lines, 4 of the most successful 11 methods involved visual inspection of computationally generated poses, while it seems that none of the 28 less successful methods involved human inter-

vention. The **binding free energy** subchallenge consisted of three sets of 5, 4, and 10 chemically similar HSP90 ligands, whose binding free energies were spanning 2.6, 3.8 and 2.1 kcal/mol, respectively. Among the 32 predictions submitted, 29 used docking-based methods, with free energy estimates based on scoring functions, force fields with implicit solvent, and electronic structure calculations with implicit solvent, and 3 submissions used alchemical free energy methods with explicit solvent. These latter simulations were performed with *FESetup* and *Sire/OpenMM* [251] and with *AMBER/MBAR* [158]. It was observed that more rigorous methods did not necessarily yield improved accuracies, compared with the results of simpler approaches based on scoring functions [156].

D3R-GC2

The D3R Grand Challenge 2 (**D3R-GC2**) [171] involved a single protein, farnesoid X receptor (FXR). In Phase 1 the participants were asked to provide affinity predictions for 102 FXR ligands and pose predictions for 36 of them. In Phase 2, the same affinity predictions as in Phase 1 were required, taking into account the additional structural data (36 new protein-ligand complexes) released at the end of Phase 1. Most of the ligands could be organized in four homogeneous classes based on their chemical structures (benzimidazoles, isoxazoles, sulfonamides, spiro compounds), along with a few diverse structures. The challenge also included the prediction of relative affinities for two homogeneous subsets of 15 and 18 compounds that are suited for free energy calculations.

In the **pose prediction** section, the predictions used one or several docking programs, sometimes combined with molecular dynamics simulations or machine learning. A few methods superposed the challenge ligands onto the structures of similar ligands in available co-crystal structures, instead of using docking. About half of submissions achieved a median RMSD of <2.0 Å for their top-ranked pose, despite the flexibility of the binding site, and its relatively featureless and hydrophobic character. However, the most accurate results were obtained when the predictions could be guided by available co-crystal structures of similar ligands with FXR. Unlike the previous challenges, manual intervention was not involved in the top ranked submissions of this challenge. In the **ranking** section, 59 and 82 submissions were received in Phases 1 and 2, respectively. Most of them used structure-based approaches, with a wide range of methods based on force fields with implicit solvent models, electronic structure methods with implicit solvent models, and methods that combined physical models and machine learning. The top performing submissions used docking software such as *SMINA*, *Vina*, and *Glide*, the *IChem-GRIM* and *HYDE* scoring methods, and *IDOCK*. A few ligand-based approaches were also used, making use of Quantitative Structure-Activity Relationships (QSAR) models or combining ligand binding pose data. Interestingly, the availability of accurate poses in Phase 2 did not improve the ranking accuracy. In the **free energy** section, 30 submissions used alchemical approaches in explicit solvent to provide relative binding free energies between pairs of ligands, whereas 39 submissions applied other approaches, including methods based on scoring functions, force fields combined with implicit solvent, and electronic structure calculations with implicit solvent. As in the case of ligand ranking, no clear improvement was observed with additional structural information available in Phase 2. The explicit-solvent free energy methods did not provide better accuracy compared with the other methods. The submissions that performed well on one or another subset involved the *Autodock Vina* energy score, a trained random forest model, and MMGB/SA calculations trained on available FXR binding data,

but only one method performed well across both subsets, a knowledge-based scoring function (*ITScore_v2_TF*) developed with a statistical mechanics-based iterative method using available FXR binding data. Interestingly, similar alchemical methods provided consistent results between participants, as two independent groups using Schrodinger’s *FEP+* obtained centered root-mean-square error (RMSE_c) values of 1.48 and 1.52 kcal/mol for subset 1, and 1.31 and 1.49 kcal/mol for subset 2. It is also noteworthy that an absolute binding free energy method, combining Jarzynski non-equilibrium pulling and umbrella sampling, performed well in Phase 2 for the subset 2, leading to an RMSE_c of 0.94 kcal/mol and tau of 0.62 [171].

D3R-GC3

The dataset proposed for the D3R Grand Challenge 3 (**D3R-GC3**) [196] was focused on cathepsin S (CatS) and five kinases. In the **Subchallenge 1**, for Phase 1A the participants were asked to predict the affinity ranking for 136 CatS ligands, the crystallographic poses for 24 of them, and the relative binding affinities for a subset of 33 compounds. In phase 1B, the organizers released the crystallographic structures of the complexes without the coordinates of the ligands and asked to predict again the protein-ligand complexes using self-docking instead of cross-docking. In Phase 2 the participants repeated the ranking predictions taking into account the additional structural data released at the end of Phase 1B. Subchallenges 2 to 5 included only ranking predictions for kinases, as follows. **Subchallenge 2** proposed three subsets with 85, 89, and 72 diverse ligands for kinases vascular endothelial growth factor receptor 2 (VEGFR2), Janus Kinase 2 (JAK2), and p38- α (also called mitogen-activating protein kinase 14 or MAPK14), respectively; 54 of these ligands were common for all three kinases. **Subchallenge 3** included 17 congeneric JAK2 ligands (that were different from those of Subchallenge 2), whereas **Subchallenge 4** contained 18 congeneric ligands of the kinase Angiopoietin-1 receptor (TIE2). Finally, **Subchallenge 5** required the prediction of affinities for two compounds in the complex with the wild type and five mutants of the nonphosphorylated ABL1 protein: ABL1(F317I), ABL1(F317L), ABL1(H396P), ABL1(Q252H), and ABL1(T315I).

The **pose prediction** challenge (involving CatS ligands only) proved to be quite difficult, with few submissions achieving a mean or median RMSD for the top-ranking pose below 2.5 Å. However, the best predictions did well, with lowest median RMSD values of 1.87 Å and 1.01 Å, in Phases 1A and 1B, respectively. Interestingly, all but two of the top-performing submissions in Phase 1A used visual inspection to help with their pose predictions and, in contrast, only one of the ten lowest-performing methods, based on pose 1 median RMSD, used visual inspection. In the **ranking** section, a notable submission for the CatS dataset in Phase 1 clearly outperformed the others, using a physics-based energy function (*ICM VLS* score) and the knowledge-based *Atomic Property Field 3D* QSAR approach, in conjunction with poses produced by *ICM-Dock* with ligand bias and 4D receptor conformational ensembles (*LigBEnD*) [201]. For ABL1, the top-performing methods include the *Rhodium* docking and scoring algorithm developed by Southwest Research Institute and a topology-based machine-learning method [202]. For JAK2 SC2, the best prediction used *GNINA* docking and a convolutional neural network scoring model [200]. For JAK2 SC3, the three best predictions involved a knowledge-based scoring function, *ITSCORE2*, and two variations of a convolutional neural network docking and scoring

method [200]. For TIE2, the top-performing methods included two topology-based machine-learning methods [202] and a convolutional neural network docking and scoring method [200]. Although in this challenge there was an increased use of machine- and deep-learning methods, it is not clear that such methods performed better than alternative methods. Both types of approaches provided similar overall performance for all targets, with the exception of TIE2, for which all but three submissions used machine-learning. In the **free energy** section, a single submission was received, which used explicit solvent alchemical free energy methods [196].

D3R-GC4

The D3R Grand Challenge 4 (**D3R-GC4**) [207] was based on two protein targets: cathepsin S (CatS), which was already present in the previous D3R-GC3, and beta-secretase 1 (BACE). The BACE subchallenge included three stages. In Phase 1A, the participants were required to predict the affinity ranking for 154 ligands, the crystallographic poses of a subset of 20 ligands, and the binding affinities for a designated free energy subset of 34 structurally homogeneous ligands. In Phase 1B, the organizers released the corresponding receptor structures (without ligand coordinates) for the 20 BACE ligands composing the pose prediction subset, and the participants could repeat the pose prediction using this additional structural information. In Phase 2, the organizers released the complexes from the pose prediction subset and the participants were required to repeat the affinity predictions for the 154 ligands and the relative free energy for the subset of 34 compounds, taking into account the 20 released protein-ligand complexes. The CatS subchallenge included a single stage, Phase 2, with ranking prediction for two datasets of 459 ligands and 39 ligands, the latter being designed for relative free energy calculations.

For the **ranking** section, 10 submissions were received for CatS, using a custom *ICM-docking* procedure and iterative *3D Atomic Property Field* quantitative structure–activity relationships (QSAR) model [217], topology-based deep learning methods with features generated by algebraic graphs, differential geometry, and algebraic topology scores [218] and the *DeepScaffOpt* method with an ensemble of deep neural networks trained on CatS data from ChEMBL. In BACE1 phase 1 there was an approach, using the *GOLD* docking software and *Goldscore* scoring function, that clearly outperformed all other methods [252]. In BACE1 phase 2, the top submissions included the same *GOLD/GoldScore* method [252] and a second method using *SkeleDock* with the *Kdeep* scoring function. In the **affinity prediction** section for BACE1 ligands, 60% of all phase 1A submissions achieved cross-docking with a median pose 1 RMSD < 2.5 Å, whereas in self-docking (phase 1B) 59 out of 71 submissions (83%) predicted a median pose 1 with RMSD < 2.5 Å. Interestingly, the best submissions performed well in both phases 1A and 1B. In the **free energy** section, the BACE1 dataset involved scaffold hopping, while the CatS dataset included only one chemical scaffold. Among all submissions, only one and five methods used alchemical free energy methods in explicit solvent for BACE1 phase 2 and CatS, respectively. All other submissions were structure-based and ligand-based scoring methods providing relative binding free energies between pairs of ligands. For CatS, the top-performing methods included four submissions that used explicit solvent alchemical free energy methods [208], whereas for BACE1 the one submission using an alchemical free energy approach exhibited poorer performance.

SAMPL7

The docking section of **SAMPL7** involved the second bromodomain of the Pleckstrin homology domain interacting protein (PHIP2). This challenge was organized in three phases: (1) identification of binders from a dataset of 799 unique fragments that were screened by X-ray crystallography; (2) prediction of fragment binding modes of the active compounds; and (3) selection of new compounds for screening from an experimental database containing more than 41 million compounds [253]. A total of 44 submissions were received (21, 19, and 4 for phases 1, 2, and 3 respectively). In phase 3, the participants suggested new compounds for screening, but none were actually screened because the experimental group concluded the suggested compounds were too un-drug-like (e.g., too hydrophobic) and also because of logistical difficulties due to the COVID-19 pandemic.

The main difficulty of this challenge was the extremely low affinity of the fragments and the presence of 4 different binding sites. Therefore, most participants made predictions only for the main binding site, which was well characterized in the literature. The methods used for these predictions were ligand-based, generally involving machine learning (ML) approaches, or structure-based, with docking with or without constraints, and in some cases with post-processing based on molecular dynamics (MD) simulations or ML models.

CELPP

The Continuous Evaluation of Ligand Protein Predictions (**CELPP**) [254] was a special kind of blinded prediction challenge that took place during a 66-week period in 2017 and 2018. It was designed to address reproducibility issues in the prediction of protein-ligand complexes using docking calculations that are related to human intervention (e.g., preparation of the protein and ligand, choice of the protein structure(s), or docking parameters). Participants build a workflow for the prediction of protein-ligand complexes, which was then challenged with the prediction of 10–100 new protein-ligand crystal structures each week. After the publication of the pre-release notifications by the PDB that takes place every Friday at 20:00, the complexes to be predicted were selected and prepared in a standardized format (including the identity of protein and ligand, crystallization conditions, etc.), then the challenge was open for submissions from Sunday at 0:00 to Tuesday at 14:59. The submissions were evaluated from Tuesday to Friday, until the beginning of the next round. This is a typical example of a cross-docking challenge, where the protein used for docking was crystallized with another ligand or in the apo form. In total, 1,989 targets were evaluated during this period. The results obtained during this continuous challenge on a very diverse dataset of proteins and ligands can be further used to identify the strengths and weaknesses of existing approaches, to improve the docking algorithms, and more generally to improve the efficiency of structure-guided drug design.

General Remarks on Protein-ligand Challenges

Although it is difficult to evaluate how the prediction accuracy of protein-ligand interactions evolved over time due to the diversity of datasets proposed and the limited number of independent benchmarks, these challenges acted as incentives for the development of existing methodologies. In the vast majority of cases, the use of existing information to

guide the predictions led to significantly improved results. The manual intervention at different stages of the whole process (choice of the receptor structure, of the docking software and the scoring function, post-processing and selection of best poses, etc.) generally proved to be beneficial, although there were some exceptions. Over time, the submissions involving alchemical free energy calculations showed good performance in many cases, but these methods are computationally demanding and hampered by limitations such as force field accuracy and incomplete sampling. These approaches are now challenged by the machine- and deep-learning methods that became more and more present in the recent challenges, showing performances similar or even superior to the traditional approaches.

Conclusion and Perspectives

The SAMPL and D3R grand challenges were initiated to provide an unbiased comparison between different computational methods based on curated sets of unpublished experimental data in order to help practitioners in the field to assess their applicability and reliability. We summarized the historical evolution of these competitions, the scientific questions they sought to explore, and the progress of the community in reliably doing so. In this concluding section we look to evaluate the impact of these challenges on the field.

While gaining new theoretical insights and improvements have been a strong incentive for all participants, one would have hoped that predictions from physics-based methods would have generally improved over the years. Surprisingly, results obtained by participants in the SAMPL challenges make it clear that this is not the case. More progress has been seen for host-guest/protein-ligand interactions compared to QM-based predictions of solution properties. Overall, empirical methods that can be tuned quite quickly are still in the lead over physics-based methods, which evolve more slowly.

Identifying reasons for this finding remains a difficult but important task for future blind challenges. Sources of errors will have to be investigated from a theoretical but also experimental point of view. Limitations in comparing theoretical results with experimental data often arise from the fact that experimental errors from donated data are hardly ever known, and therefore, it is difficult to assess whether predictions achieved an accuracy within the boundaries of the experimental errors. This is important because while certain properties like binding constants to soluble proteins like proteases can be measured with high accuracy, experimental determination of other properties, like tautomer ratios in different solvents, proves to be very challenging. Also, meaningful statistical correlations between predicted and measured data are often hampered by small dynamic ranges of measured properties, which might be a consequence of limited structural diversity in ligand sets. Nevertheless, some experimental data points have been identified as likely experimental outliers by consistent mispredictions from different participants.

Of course, there are theoretical factors that constitute sources of errors. Most molecular properties of relevance are not governed by a single conformation in the gas phase but by conformational ensembles in complex condensed phases. While prediction of properties in solution is at the heart of the SAMPL challenges, and significant progress has been made here, sampling and ranking of conformations is still a weak link and likely source of error propagating into the final result. This problem is understood in principle, but sampling at a high level of theory would be needed for proper treatment, which is still simply out of

reach for practical reasons, and most initiatives focused on improvements to force fields [255,256] or fast semiempirical or DFT techniques [116,257]. Approaches making use of neural network potentials such as *ASE-ANI* [258] augment physics-based methods with state-of-the-art AI-based machine learning making high quality results available at a fraction of the time. The drawback that each property to be predicted requires separate training, is overcome by techniques like *SchNOorb* [259] that predict the molecular wave function directly, from which in turn different properties can be derived. Alternatively, using quantum computers would be an obvious solution for this task, but those with a quantum volume that is high enough for the required calculations are still a mere promise for the foreseeable future.

After more than a decade since the first SAMPL challenge began, all reasons to conduct those challenges are still valid. New computational techniques and approaches will continue to be developed and SAMPL challenges are well positioned for continuing to critically monitor the advances. Still, the long-term perspective is unclear and needs to be secured by broadening the base for funding not only for the hosts, but also for participants as well, as they are investing considerable computational and human resources. Therefore, the SAMPL roadmap suggests *containerization* [260] by introducing a new category of challenge formats where participants submit methods in software containers rather than predictions. Additionally, a steady stream of high-quality experimental data is key to future SAMPL challenges, and would allow reviving challenges like the SFE challenge, which was stopped due to lack of new experimental data. Establishing a crowd sourcing initiative for experimental data could be an effective way to achieve this and a way of broadening the funding base at the same time. Capacity for testing does exist in academia, but is much higher in industry, where there is also much broader access to chemical matter available. Here, industry could be of tremendous help in multilateral partnerships with academia. While intellectual property and data privacy are of high importance, industry has entered into innovative pre-competitive collaborations already. Recent examples for this are the EU-funded MELLODDY project on data exchange and distributed federated learning [261] or the Open Force Field initiative [256]. To make the blind challenges shiny, this is the way!

Acknowledgments

We thank David Mobley (University of California, Irvine), Rommie Amaro (University of California, San Diego) and Michael Gilson (University of California, San Diego) for great suggestions to gather and distribute previous SAMPL and D3R structures and experimental values. We thank all previous organizers for organizing the challenges and thank experimental data contributors for making the blind dataset available to the community.

OB was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM118772. BII was supported by the Laboratory of Excellence in Research on Medication and Innovative Therapeutics (LERMIT, grant ANR-10-LABX-33) and by the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR, grant ANR-14-JAMR-002). SMK was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2033 – Projektnummer 390677874, and under the Research Unit FOR 1979.

References

- (1) Smith, R. D.; Dunbar, J. B.; Ung, P. M.-U.; Esposito, E. X.; Yang, C.-Y.; Wang, S.; Carlson, H. A. CSAR Benchmark Exercise of 2010: Combined Evaluation Across All Submitted Scoring Functions. *J. Chem. Inf. Model.* **2011**, *51* (9), 2115–2131, doi: 10.1021/ci200269q.
- (2) Damm-Ganamet, K. L.; Smith, R. D.; Dunbar, J. B.; Stuckey, J. A.; Carlson, H. A. CSAR Benchmark Exercise 2011–2012: Evaluation of Results from Docking and Relative Ranking of Blinded Congeneric Series. *J. Chem. Inf. Model.* **2013**, *53* (8), 1853–1870, doi: 10.1021/ci400025f.
- (3) Smith, R. D.; Damm-Ganamet, K. L.; Dunbar, J. B.; Ahmed, A.; Chinnaswamy, K.; Delproposto, J. E.; Kubish, G. M.; Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Stuckey, J. A.; Baker, D.; Carlson, H. A. CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. *J. Chem. Inf. Model.* **2016**, *56* (6), 1022–1031, doi: 10.1021/acs.jcim.5b00387.
- (4) Carlson, H. A.; Smith, R. D.; Damm-Ganamet, K. L.; Stuckey, J. A.; Ahmed, A.; Convery, M. A.; Somers, D. O.; Kranz, M.; Elkins, P. A.; Cui, G.; Peishoff, C. E.; Lambert, M. H.; Dunbar, J. B. CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J. Chem. Inf. Model.* **2016**, *56* (6), 1063–1077, doi: 10.1021/acs.jcim.5b00523.
- (5) SAMPL Challenge. *Wikipedia*, 2021. Available: https://en.wikipedia.org/wiki/SAMPL_Challenge. Accessed: 2021-09-12.
- (6) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *J. Med. Chem.* **2008**, *51* (4), 769–779, doi: 10.1021/jm070549+.
- (7) *Samplchallenges/SAMPL0*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL0>. Accessed: 2021-09-11, doi:10.5281/zenodo.5508276.
- (8) Guthrie, J. P. A Blind Challenge for Computational Solvation Free Energies: Introduction and Overview. *J. Phys. Chem. B* **2009**, *113* (14), 4501–4507, doi: 10.1021/jp806724u.
- (9) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113* (14), 4538–4543, doi: 10.1021/jp809094y.
- (10) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113* (14), 4533–4537, doi: 10.1021/jp806838b.
- (11) Nicholls, A.; Wlodek, S.; Grant, J. A. The SAMPL1 Solvation Challenge: Further Lessons Regarding the Pitfalls of Parametrization. *J. Phys. Chem. B* **2009**, *113* (14), 4521–4532, doi: 10.1021/jp806855q.
- (12) Sulea, T.; Wanapun, D.; Dennis, S.; Purisima, E. O. Prediction of SAMPL-1 Hydration Free Energies Using a Continuum Electrostatics-Dispersion Model. *J. Phys. Chem. B* **2009**, *113* (14), 4511–4520, doi: 10.1021/jp8061477.
- (13) Purisima, E. O.; Corbeil, C. R.; Sulea, T. Rapid Prediction of Solvation Free Energy. 3. Application to the SAMPL2 Challenge. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 373–383, doi: 10.1007/s10822-010-9341-9.

- (14) Klamt, A.; Eckert, F.; Diedenhofen, M. Prediction of the Free Energy of Hydration of a Challenging Set of Pesticide-Like Compounds. *J. Phys. Chem. B* **2009**, *113* (14), 4508–4510, doi: 10.1021/jp805853y.
- (15) *Samplchallenges/SAMPL1*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL1>. Accessed: 2021-09-11, doi:10.5281/zenodo.5510231.
- (16) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 Blind Prediction Challenge: Introduction and Overview. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 259–279, doi: 10.1007/s10822-010-9350-8.
- (17) Skillman, A. G.; Geballe, M. T.; Nicholls, A. SAMPL2 Challenge: Prediction of Solvation Energies and Tautomer Ratios. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 257–258, doi: 10.1007/s10822-010-9358-0.
- (18) Ellingson, B. A.; Skillman, A. G.; Nicholls, A. Analysis of SM8 and Zap TK Calculations and Their Geometric Sensitivity. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 335–342, doi: 10.1007/s10822-010-9355-3.
- (19) Kast, S. M.; Heil, J.; Güssregen, S.; Schmidt, K. F. Erratum to: Prediction of Tautomer Ratios by Embedded-Cluster Integral Equation Theory. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 355–355, doi: 10.1007/s10822-010-9360-6.
- (20) Kast, S. M.; Heil, J.; Güssregen, S.; Schmidt, K. F. Prediction of Tautomer Ratios by Embedded-Cluster Integral Equation Theory. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 343–353, doi: 10.1007/s10822-010-9340-x.
- (21) Klamt, A.; Diedenhofen, M. Blind Prediction Test of Free Energies of Hydration with COSMO-RS. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 357–360, doi: 10.1007/s10822-010-9354-4.
- (22) Klimovich, P. V.; Mobley, D. L. Predicting Hydration Free Energies Using All-Atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 307–316, doi: 10.1007/s10822-010-9343-7.
- (23) Meunier, A.; Truchon, J.-F. Predictions of Hydration Free Energies from Continuum Solvent with Solute Polarizable Models: The SAMPL2 Blind Challenge. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 361–372, doi: 10.1007/s10822-010-9339-3.
- (24) Nicholls, A.; Wlodek, S.; Grant, J. A. SAMPL2 and Continuum Modeling. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 293–306, doi: 10.1007/s10822-010-9334-8.
- (25) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Prediction of SAMPL2 Aqueous Solvation Free Energies and Tautomeric Ratios Using the SM8, SM8AD, and SMD Solvation Models. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 317–333, doi: 10.1007/s10822-010-9333-9.
- (26) Soteras, I.; Orozco, M.; Luque, F. J. Performance of the IEF-MST Solvation Continuum Model in the SAMPL2 Blind Test Prediction of Hydration and Tautomerization Free Energies. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 281–291, doi: 10.1007/s10822-010-9331-y.
- (27) *Samplchallenges/SAMPL2*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL2>. Accessed: 2021-09-11, doi:10.5281/zenodo.5508280.
- (28) Geballe, M. T.; Guthrie, J. P. The SAMPL3 Blind Prediction Challenge: Transfer Energy Overview. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 489–496, doi: 10.1007/s10822-012-9568-8.
- (29) Muddana, H. S.; Daniel Varnado, C.; Bielawski, C. W.; Urbach, A. R.; Isaacs,

- L.; Geballe, M. T.; Gilson, M. K. Blind Prediction of Host–Guest Binding Affinities: A New SAMPL3 Challenge. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 475–487, doi: 10.1007/s10822-012-9554-1.
- (30) Kulp, J. L.; Blumenthal, S. N.; Wang, Q.; Bryan, R. L.; Guarnieri, F. A Fragment-Based Approach to the SAMPL3 Challenge. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 583–594, doi: 10.1007/s10822-012-9546-1.
- (31) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical Prediction of Hydration Free Energies for SAMPL. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 551–562, doi: 10.1007/s10822-011-9528-8.
- (32) Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding Affinities in the SAMPL3 Trypsin and Host–Guest Blind Tests Estimated with the MM/PBSA and LIE Methods. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 527–541, doi: 10.1007/s10822-011-9524-z.
- (33) Kumar, A.; Zhang, K. Y. J. Computational Fragment-Based Screening Using RosettaLigand: The SAMPL3 Challenge. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 603–616, doi: 10.1007/s10822-011-9523-0.
- (34) Surpateanu, G.; Iorga, B. I. Evaluation of Docking Performance in a Blinded Virtual Screening of Fragment-like Trypsin Inhibitors. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 595–601, doi: 10.1007/s10822-011-9526-x.
- (35) Sulea, T.; Hogues, H.; Purisima, E. O. Exhaustive Search and Solvated Interaction Energy (SIE) for Virtual Screening and Affinity Prediction. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 617–633, doi: 10.1007/s10822-011-9529-7.
- (36) König, G.; Brooks, B. R. Predicting Binding Affinities of Host-Guest Systems in the SAMPL3 Blind Challenge: The Performance of Relative Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 543–550, doi: 10.1007/s10822-011-9525-y.
- (37) Sulea, T.; Purisima, E. O. Predicting Hydration Free Energies of Polychlorinated Aromatic Compounds from the SAMPL-3 Data Set with FiSH and LIE Models. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 661–667, doi: 10.1007/s10822-011-9522-1.
- (38) Reinisch, J.; Klamt, A.; Diedenhofen, M. Prediction of Free Energies of Hydration with COSMO-RS on the SAMPL3 Data Set. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 669–673, doi: 10.1007/s10822-012-9576-8.
- (39) Beckstein, O.; Iorga, B. I. Prediction of Hydration Free Energies for Aliphatic and Aromatic Chloro Derivatives Using Molecular Dynamics Simulations with the OPLS-AA Force Field. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 635–645, doi: 10.1007/s10822-011-9527-9.
- (40) Gallicchio, E.; Levy, R. M. Prediction of SAMPL3 Host-Guest Affinities with the Binding Energy Distribution Analysis Method (BEDAM). *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 505–516, doi: 10.1007/s10822-012-9552-3.
- (41) Muddana, H. S.; Gilson, M. K. Prediction of SAMPL3 Host–Guest Binding Affinities: Evaluating the Accuracy of Generalized Force-Fields. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 517–525, doi: 10.1007/s10822-012-9544-3.
- (42) Benson, M. L.; Faver, J. C.; Ucisik, M. N.; Dashti, D. S.; Zheng, Z.; Merz, K. M. Prediction of Trypsin/Molecular Fragment Binding Affinities by Free Energy Decomposition and Empirical Scores. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 647–659, doi: 10.1007/s10822-012-9567-9.
- (43) Kehoe, C. W.; Fennell, C. J.; Dill, K. A. Testing the Semi-Explicit Assembly Solvation Model in the SAMPL3 Community Blind Test. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 563–568, doi: 10.1007/s10822-011-9536-8.

- (44) Newman, J.; Dolezal, O.; Fazio, V.; Caradoc-Davies, T.; Peat, T. S. The DINGO Dataset: A Comprehensive Set of Data for the SAMPL Challenge. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 497–503, doi: 10.1007/s10822-011-9521-2.
- (45) Lawrenz, M.; Wereszczynski, J.; Ortiz-Sánchez, J. M.; Nichols, S. E.; McCammon, J. A. Thermodynamic Integration to Predict Host-Guest Binding Affinities. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 569–576, doi: 10.1007/s10822-012-9542-5.
- (46) *Samplchallenges/SAMPL3*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL3>. Accessed: 2021-09-11.
- (47) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 Host–Guest Blind Prediction Challenge: An Overview. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 305–317, doi: 10.1007/s10822-014-9735-1.
- (48) Gibb, C. L. D.; Gibb, B. C. Binding of Cyclic Carboxylates to Octa-Acid Deep-Cavity Cavitand. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 319–325, doi: 10.1007/s10822-013-9690-2.
- (49) Mobley, D. L.; Liu, S.; Lim, N. M.; Wymer, K. L.; Perryman, A. L.; Forli, S.; Deng, N.; Su, J.; Branson, K.; Olson, A. J. Blind Prediction of HIV Integrase Binding from the SAMPL4 Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 327–345, doi: 10.1007/s10822-014-9723-5.
- (50) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. Blind Prediction of Solvation Free Energies from the SAMPL4 Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 135–150, doi: 10.1007/s10822-014-9718-2.
- (51) Guthrie, J. P. SAMPL4, a Blind Challenge for Computational Solvation Free Energies: The Compounds Considered. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 151–168, doi: 10.1007/s10822-014-9738-y.
- (52) Muddana, H. S.; Yin, J.; Sapra, N. V.; Fenley, A. T.; Gilson, M. K. Blind Prediction of SAMPL4 Cucurbit[7]Uril Binding Affinities with the Mining Minima Method. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 463–474, doi: 10.1007/s10822-014-9726-2.
- (53) Voet, A. R. D.; Kumar, A.; Berenger, F.; Zhang, K. Y. J. Combining in Silico and in Cerebro Approaches for Virtual Screening and Pose Prediction in SAMPL4. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 363–373, doi: 10.1007/s10822-013-9702-2.
- (54) Monroe, J. I.; Shirts, M. R. Converging Free Energies of Binding in Cucurbit[7]Uril and Octa-Acid Host–Guest Systems from SAMPL4 Using Expanded Ensemble Simulations. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 401–415, doi: 10.1007/s10822-014-9716-4.
- (55) Hogues, H.; Sulea, T.; Purisima, E. O. Exhaustive Docking and Solvated Interaction Energy Scoring: Lessons Learned from the SAMPL4 Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 417–427, doi: 10.1007/s10822-014-9715-5.
- (56) Mikulskis, P.; Cioloboc, D.; Andrejić, M.; Khare, S.; Brorsson, J.; Genheden, S.; Mata, R. A.; Söderhjelm, P.; Ryde, U. Free-Energy Perturbation and Quantum Mechanical Study of SAMPL4 Octa-Acid Host–Guest Binding Energies. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 375–400, doi: 10.1007/s10822-014-9739-x.
- (57) Hsiao, Y.-W.; Söderhjelm, P. Prediction of SAMPL4 Host–Guest Binding Affinities Using Funnel Metadynamics. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 443–454, doi: 10.1007/s10822-014-9724-4.
- (58) Gallicchio, E.; Deng, N.; He, P.; Wickstrom, L.; Perryman, A. L.; Santiago, D.

- N.; Forli, S.; Olson, A. J.; Levy, R. M. Virtual Screening of Integrase Inhibitors by Large Scale Binding Free Energy Calculations: The SAMPL4 Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 475–490, doi: 10.1007/s10822-014-9711-9.
- (59) Colas, C.; Iorga, B. I. Virtual Screening of the SAMPL4 Blinded HIV Integrase Inhibitors Dataset. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 455–462, doi: 10.1007/s10822-014-9707-5.
- (60) Perryman, A. L.; Santiago, D. N.; Forli, S.; Santos-Martins, D.; Olson, A. J. Virtual Screening with AutoDock Vina and the Common Pharmacophore Engine of a Low Diversity Library of Fragments and Hits against the Three Allosteric Sites of HIV Integrase: Participation in the SAMPL4 Protein–Ligand Binding Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 429–441, doi: 10.1007/s10822-014-9709-3.
- (61) Ellingson, B. A.; Geballe, M. T.; Wlodek, S.; Bayly, C. I.; Skillman, A. G.; Nicholls, A. Efficient Calculation of SAMPL4 Hydration Free Energies Using OMEGA, SZYBKI, QUACPAC, and Zap TK. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 289–298, doi: 10.1007/s10822-014-9720-8.
- (62) Park, H. Extended Solvent-Contact Model Approach to SAMPL4 Blind Prediction Challenge for Hydration Free Energies. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 175–186, doi: 10.1007/s10822-014-9729-z.
- (63) Genheden, S.; Cabedo Martinez, A. I.; Criddle, M. P.; Essex, J. W. Extensive All-Atom Monte Carlo Sampling and QM/MM Corrections in the SAMPL4 Hydration Free Energy Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 187–200, doi: 10.1007/s10822-014-9717-3.
- (64) Fu, J.; Liu, Y.; Wu, J. Fast Prediction of Hydration Free Energies for SAMPL4 Blind Test from a Classical Density Functional Theory. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 299–304, doi: 10.1007/s10822-014-9730-6.
- (65) König, G.; Pickard, F. C.; Mei, Y.; Brooks, B. R. Predicting Hydration Free Energies with a Hybrid QM/MM Approach: An Evaluation of Implicit and Explicit Solvation Models in SAMPL4. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 245–257, doi: 10.1007/s10822-014-9708-4.
- (66) Sandberg, L. Predicting Hydration Free Energies with Chemical Accuracy: The SAMPL4 Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 211–219, doi: 10.1007/s10822-014-9725-3.
- (67) Reinisch, J.; Klamt, A. Prediction of Free Energies of Hydration with COSMO-RS on the SAMPL4 Data Set. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 169–173, doi: 10.1007/s10822-013-9701-3.
- (68) Manzoni, F.; Söderhjelm, P. Prediction of Hydration Free Energies for the SAMPL4 Data Set with the AMOEBA Polarizable Force Field. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 235–244, doi: 10.1007/s10822-014-9733-3.
- (69) Beckstein, O.; Fourrier, A.; Iorga, B. I. Prediction of Hydration Free Energies for the SAMPL4 Diverse Set of Compounds Using Molecular Dynamics Simulations with the OPLS-AA Force Field. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 265–276, doi: 10.1007/s10822-014-9727-1.
- (70) Coleman, R. G.; Sterling, T.; Weiss, D. R. SAMPL4 & DOCK3.7: Lessons for Automated Docking Procedures. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 201–209, doi: 10.1007/s10822-014-9722-6.
- (71) Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E. Testing and Validation of the Automated Topology Builder (ATB) Version 2.0: Prediction of Hydration Free Enthalpies. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 221–233, doi:

- 10.1007/s10822-014-9713-7.
- (72) Li, L.; Dill, K. A.; Fennell, C. J. Testing the Semi-Explicit Assembly Model of Aqueous Solvation in the SAMPL4 Challenge. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 259–264, doi: 10.1007/s10822-014-9712-8.
- (73) Muddana, H. S.; Sapra, N. V.; Fenley, A. T.; Gilson, M. K. The SAMPL4 Hydration Challenge: Evaluation of Partial Charge Sets with Explicit-Water Molecular Dynamics Simulations. *J. Comput. Aided Mol. Des.* **2014**, *28* (3), 277–287, doi: 10.1007/s10822-014-9714-6.
- (74) *Samplchallenges/SAMPL4*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL4>. Accessed: 2021-09-11, doi:10.5281/zenodo.5508284.
- (75) Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. Blind Prediction of Cyclohexane–Water Distribution Coefficients from the SAMPL5 Challenge. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 927–944, doi: 10.1007/s10822-016-9954-8.
- (76) Rustenburg, A. S.; Dancer, J.; Lin, B.; Feng, J. A.; Ortwine, D. F.; Mobley, D. L.; Chodera, J. D. Measuring Experimental Cyclohexane–Water Distribution Coefficients for the SAMPL5 Challenge. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 945–958, doi: 10.1007/s10822-016-9971-7.
- (77) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 Host–Guest Challenge: Are We Doing Better? *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 1–19, doi: 10.1007/s10822-016-9974-4.
- (78) Sullivan, M. R.; Sokkalingam, P.; Nguyen, T.; Donahue, J. P.; Gibb, B. C. Binding of Carboxylate and Trimethylammonium Salts to Octa-Acid and TEMOA Deep-Cavity Cavitands. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 21–28, doi: 10.1007/s10822-016-9925-0.
- (79) Brini, E.; Paranahegawa, S. S.; Fennell, C. J.; Dill, K. A. Adapting the Semi-Explicit Solvation Model for Estimating Water-Cyclohexane Partitioning with the SAMPL5 Molecules. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1067–1077, doi: 10.1007/s10822-016-9961-9.
- (80) Genheden, S.; Essex, J. W. All-Atom/Coarse-Grained Hybrid Predictions of Distribution Coefficients in SAMPL5. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 969–976, doi: 10.1007/s10822-016-9926-z.
- (81) Pickard, F. C.; König, G.; Tofoleanu, F.; Lee, J.; Simmonett, A. C.; Shao, Y.; Ponder, J. W.; Brooks, B. R. Blind Prediction of Distribution in the SAMPL5 Challenge with QM Based Protomer and PKa Corrections. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1087–1100, doi: 10.1007/s10822-016-9955-7.
- (82) Bosisio, S.; Mey, A. S. J. S.; Michel, J. Blinded Predictions of Distribution Coefficients in the SAMPL5 Challenge. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1101–1114, doi: 10.1007/s10822-016-9969-1.
- (83) König, G.; Pickard, F. C.; Huang, J.; Simmonett, A. C.; Tofoleanu, F.; Lee, J.; Dral, P. O.; Prasad, S.; Jones, M.; Shao, Y.; Thiel, W.; Brooks, B. R. Calculating Distribution Coefficients Based on Multi-Scale Free Energy Simulations: An Evaluation of MM and QM/MM Explicit Solvent Simulations of Water-Cyclohexane Transfer in the SAMPL5 Challenge. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 989–1006, doi: 10.1007/s10822-016-9936-x.
- (84) Santos-Martins, D.; Fernandes, P. A.; Ramos, M. J. Calculation of Distribution Coefficients in the SAMPL5 Challenge from Atomic Solvation Parameters and Surface Areas. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1079–1086, doi:

- 10.1007/s10822-016-9951-y.
- (85) Chung, K.-C.; Park, H. Extended Solvent-Contact Model Approach to Blind SAMPL5 Prediction Challenge for the Distribution Coefficients of Drug-like Molecules. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1019–1033, doi: 10.1007/s10822-016-9928-x.
- (86) Jones, M. R.; Brooks, B. R.; Wilson, A. K. Partition Coefficients for the SAMPL5 Challenge Using Transfer Free Energies. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1129–1138, doi: 10.1007/s10822-016-9964-6.
- (87) Diaz-Rodriguez, S.; Bozada, S. M.; Phifer, J. R.; Paluch, A. S. Predicting Cyclohexane/Water Distribution Coefficients for the SAMPL5 Challenge Using MOSCED and the SMD Solvation Model. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1007–1017, doi: 10.1007/s10822-016-9945-9.
- (88) Paranahe wage, S. S.; Gierhart, C. S.; Fennell, C. J. Predicting Water-to-Cyclohexane Partitioning of the SAMPL5 Molecules Using Dielectric Balancing of Force Fields. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1059–1065, doi: 10.1007/s10822-016-9950-z.
- (89) Kamath, G.; Kurnikov, I.; Fain, B.; Leontyev, I.; Illarionov, A.; Butin, O.; Olevanov, M.; Pereyaslavets, L. Prediction of Cyclohexane-Water Distribution Coefficient for SAMPL5 Drug-like Compounds with the QMPFF3 and ARROW Polarizable Force Fields. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 977–988, doi: 10.1007/s10822-016-9958-4.
- (90) Kenney, I. M.; Beckstein, O.; Iorga, B. I. Prediction of Cyclohexane-Water Distribution Coefficients for the SAMPL5 Data Set Using Molecular Dynamics Simulations with the OPLS-AA Force Field. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1045–1058, doi: 10.1007/s10822-016-9949-5.
- (91) Klamt, A.; Eckert, F.; Reinisch, J.; Wichmann, K. Prediction of Cyclohexane-Water Distribution Coefficients with COSMO-RS on the SAMPL5 Data Set. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 959–967, doi: 10.1007/s10822-016-9927-y.
- (92) Luchko, T.; Blinov, N.; Limon, G. C.; Joyce, K. P.; Kovalenko, A. SAMPL5: 3D-RISM Partition Coefficient Calculations with Partial Molar Volume Corrections and Solute Conformational Sampling. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1115–1127, doi: 10.1007/s10822-016-9947-7.
- (93) Tielker, N.; Tomazic, D.; Heil, J.; Kloss, T.; Ehrhart, S.; Güssregen, S.; Schmidt, K. F.; Kast, S. M. The SAMPL5 Challenge for Embedded-Cluster Integral Equation Theory: Solvation Free Energies, Aqueous pK_a , and Cyclohexane–Water $\log D$. *J. Comput. Aided Mol. Des.* **2016**, *30* (11), 1035–1044, doi: 10.1007/s10822-016-9939-7.
- (94) Pal, R. K.; Haider, K.; Kaur, D.; Flynn, W.; Xia, J.; Levy, R. M.; Taran, T.; Wickstrom, L.; Kurtzman, T.; Gallicchio, E. A Combined Treatment of Hydration and Dynamical Effects for the Modeling of Host–Guest Binding Thermodynamics: The SAMPL5 Blinded Challenge. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 29–44, doi: 10.1007/s10822-016-9956-6.
- (95) Tofoleanu, F.; Lee, J.; Pickard IV, F. C.; König, G.; Huang, J.; Baek, M.; Seok, C.; Brooks, B. R. Absolute Binding Free Energies for Octa-Acids and Guests in SAMPL5. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 107–118, doi: 10.1007/s10822-016-9965-5.
- (96) Lee, J.; Tofoleanu, F.; Pickard, F. C.; König, G.; Huang, J.; Damjanović, A.; Baek, M.; Seok, C.; Brooks, B. R. Absolute Binding Free Energy Calculations of CBClip Host–Guest Systems in the SAMPL5 Blind Challenge. *J. Comput.*

- Aided Mol. Des.* **2017**, *31* (1), 71–85, doi: 10.1007/s10822-016-9968-2.
- (97) Caldarraru, O.; Olsson, M. A.; Riplinger, C.; Neese, F.; Ryde, U. Binding Free Energies in the SAMPL5 Octa-Acid Host–Guest Challenge Calculated with DFT-D3 and CCSD(T). *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 87–106, doi: 10.1007/s10822-016-9957-5.
- (98) Bosisio, S.; Mey, A. S. J. S.; Michel, J. Blinded Predictions of Host–Guest Standard Free Energies of Binding in the SAMPL5 Challenge. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 61–70, doi: 10.1007/s10822-016-9933-0.
- (99) Pal, R. K.; Haider, K.; Kaur, D.; Flynn, W.; Xia, J.; Levy, R. M.; Taran, T.; Wickstrom, L.; Kurtzman, T.; Gallicchio, E. Erratum to: A Combined Treatment of Hydration and Dynamical Effects for the Modeling of Host–Guest Binding Thermodynamics: The SAMPL5 Blinded Challenge. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 45–45, doi: 10.1007/s10822-016-9987-z.
- (100) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 147–161, doi: 10.1007/s10822-016-9977-1.
- (101) Bansal, N.; Zheng, Z.; Cerutti, D. S.; Merz, K. M. On the Fly Estimation of Host–Guest Binding Free Energies Using the Movable Type Method: Participation in the SAMPL5 Blind Challenge. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 47–60, doi: 10.1007/s10822-016-9980-6.
- (102) Bhakat, S.; Söderhjelm, P. Resolving the Problem of Trapped Water in Binding Cavities: Prediction of Host–Guest Binding Free Energies in the SAMPL5 Challenge by Funnel Metadynamics. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 119–132, doi: 10.1007/s10822-016-9948-6.
- (103) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Gilson, M. K. The SAMPL5 Host–Guest Challenge: Computing Binding Free Energies and Enthalpies from Explicit Solvent Simulations by the Attach–Pull–Release (APR) Method. *J. Comput. Aided Mol. Des.* **2017**, *31* (1), 133–145, doi: 10.1007/s10822-016-9970-8.
- (104) *Samplchallenges/SAMPL5*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL5>. Accessed: 2021-09-11, doi:10.5281/zenodo.5508286.
- (105) Işık, M.; Bergazin, T. D.; Fox, T.; Rizzi, A.; Chodera, J. D.; Mobley, D. L. Assessing the Accuracy of Octanol–Water Partition Coefficient Predictions in the SAMPL6 Part II Log P Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 335–370, doi: 10.1007/s10822-020-00295-0.
- (106) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 Host–Guest Binding Affinity Prediction Challenge. *J. Comput. Aided Mol. Des.* **2018**, *32* (10), 937–963, doi: 10.1007/s10822-018-0170-6.
- (107) Işık, M.; Levorse, D.; Mobley, D. L.; Rhodes, T.; Chodera, J. D. Octanol–Water Partition Coefficient Measurements for the SAMPL6 Blind Prediction Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 405–420, doi: 10.1007/s10822-019-00271-3.
- (108) Zeng, Q.; Jones, M. R.; Brooks, B. R. Absolute and Relative pK_a Predictions via a DFT Approach Applied to the SAMPL6 Blind Challenge. *J. Comput. Aided Mol. Des.* **2018**, *32* (10), 1179–1189, doi: 10.1007/s10822-018-0150-x.
- (109) Laury, M. L.; Wang, Z.; Gordon, A. S.; Ponder, J. W. Absolute Binding Free Energies for the SAMPL6 Cucurbit[8]Uril Host–Guest Challenge via the AMOEBA Polarizable Force Field. *J. Comput. Aided Mol. Des.* **2018**, *32* (10),

- 1087–1095, doi: 10.1007/s10822-018-0147-5.
- (110) Prasad, S.; Huang, J.; Zeng, Q.; Brooks, B. R. An Explicit-Solvent Hybrid QM and MM Approach for Predicting PKa of Small Molecules in SAMPL6 Challenge. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1191–1201, doi: 10.1007/s10822-018-0167-1.
- (111) Caldararu, O.; Olsson, M. A.; Misini Ignjatović, M.; Wang, M.; Ryde, U. Binding Free Energies in the SAMPL6 Octa-Acid Host–Guest Challenge Calculated with MM and QM Methods. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1027–1046, doi: 10.1007/s10822-018-0158-2.
- (112) Papadourakis, M.; Bosisio, S.; Michel, J. Blinded Predictions of Standard Binding Free Energies: Lessons Learned from the SAMPL6 Challenge. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1047–1058, doi: 10.1007/s10822-018-0154-6.
- (113) Nishikawa, N.; Han, K.; Wu, X.; Tofoleanu, F.; Brooks, B. R. Comparison of the Umbrella Sampling and the Double Decoupling Method in Binding Free Energy Predictions for SAMPL6 Octa-Acid Host–Guest Challenges. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1075–1086, doi: 10.1007/s10822-018-0166-2.
- (114) Song, L. F.; Bansal, N.; Zheng, Z.; Merz, K. M. Detailed Potential of Mean Force Studies on Host–Guest Systems from the SAMPL6 Challenge. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1013–1026, doi: 10.1007/s10822-018-0153-7.
- (115) Hudson, P. S.; Han, K.; Woodcock, H. L.; Brooks, B. R. Force Matching as a Stepping Stone to QM/MM CB[8] Host/Guest Binding Free Energies: A SAMPL6 Cautionary Tale. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 983–999, doi: 10.1007/s10822-018-0165-3.
- (116) Pracht, P.; Wilcken, R.; Udvarhelyi, A.; Rodde, S.; Grimme, S. High Accuracy Quantum-Chemistry-Based Calculation and Blind Prediction of Macroscopic PKa Values in the Context of the SAMPL6 Challenge. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1139–1149, doi: 10.1007/s10822-018-0145-7.
- (117) Işık, M.; Levorse, D.; Rustenburg, A. S.; Ndukwe, I. E.; Wang, H.; Wang, X.; Reibarkh, M.; Martin, G. E.; Makarov, A. A.; Mobley, D. L.; Rhodes, T.; Chodera, J. D. pKa Measurements for the SAMPL6 Prediction Challenge for a Set of Kinase Inhibitor-like Fragments. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1117–1138, doi: 10.1007/s10822-018-0168-0.
- (118) Dixon, T.; Lotz, S. D.; Dickson, A. Predicting Ligand Binding Affinity Using On- and off-Rates for the SAMPL6 SAMPLing Challenge. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1001–1012, doi: 10.1007/s10822-018-0149-3.
- (119) Han, K.; Hudson, P. S.; Jones, M. R.; Nishikawa, N.; Tofoleanu, F.; Brooks, B. R. Prediction of CB[8] Host–Guest Binding Free Energies in SAMPL6 Using the Double-Decoupling Method. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1059–1073, doi: 10.1007/s10822-018-0144-8.
- (120) Bannan, C. C.; Mobley, D. L.; Skillman, A. G. SAMPL6 Challenge Results from pKa predictions Based on a General Gaussian Process Model. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1165–1177, doi: 10.1007/s10822-018-0169-z.
- (121) Procacci, P.; Guarrasi, M.; Guarnieri, G. SAMPL6 Host–Guest Blind Predictions Using a Non Equilibrium Alchemical Approach. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 965–982, doi: 10.1007/s10822-018-0151-9.
- (122) Eken, Y.; Patel, P.; Díaz, T.; Jones, M. R.; Wilson, A. K. SAMPL6 Host–Guest Challenge: Binding Free Energies via a Multistep Approach. *J. Comput. Aided Mol. Des.* **2018**, 32 (10), 1097–1115, doi: 10.1007/s10822-018-0159-1.
- (123) Selwa, E.; Kenney, I. M.; Beckstein, O.; Iorga, B. I. SAMPL6: Calculation of Macroscopic PKa Values from Ab Initio Quantum Mechanical Free Energies. *J.*

- Comput. Aided Mol. Des.* **2018**, *32* (10), 1203–1216, doi: 10.1007/s10822-018-0138-6.
- (124) Tielker, N.; Eberlein, L.; Güssregen, S.; Kast, S. M. The SAMPL6 Challenge on Predicting Aqueous pK_a Values from EC-RISM Theory. *J. Comput. Aided Mol. Des.* **2018**, *32* (10), 1151–1163, doi: 10.1007/s10822-018-0140-z.
- (125) Arslan, E.; Findik, B. K.; Aviyente, V. A Blind SAMPL6 Challenge: Insight into the Octanol-Water Partition Coefficients of Drug-like Molecules via a DFT Approach. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 463–470, doi: 10.1007/s10822-020-00284-3.
- (126) Lui, R.; Guan, D.; Matthews, S. A Comparison of Molecular Representations for Lipophilicity Quantitative Structure–Property Relationships with Results from the SAMPL6 LogP Prediction Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 523–534, doi: 10.1007/s10822-020-00279-0.
- (127) Prasad, S.; Brooks, B. R. A Deep Learning Approach for the Blind LogP Prediction in SAMPL6 Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 535–542, doi: 10.1007/s10822-020-00292-3.
- (128) Loschen, C.; Reinisch, J.; Klamt, A. COSMO-RS Based Predictions for the SAMPL6 LogP Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 385–392, doi: 10.1007/s10822-019-00259-z.
- (129) Guan, D.; Lui, R.; Matthews, S. Log P Prediction Performance with the SMD Solvation Model and the M06 Density Functional Family for SAMPL6 Blind Prediction Challenge Molecules. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 511–522, doi: 10.1007/s10822-020-00278-1.
- (130) Krämer, A.; Hudson, P. S.; Jones, M. R.; Brooks, B. R. Multi-Phase Boltzmann Weighting: Accounting for Local Inhomogeneity in Molecular Simulations of Water–Octanol Partition Coefficients in the SAMPL6 Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 471–483, doi: 10.1007/s10822-020-00285-2.
- (131) Nikitin, A. Non-Zero Lennard-Jones Parameters for the Toukan–Rahman Water Model: More Accurate Calculations of the Solvation Free Energy of Organic Substances. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 437–441, doi: 10.1007/s10822-019-00256-2.
- (132) Ouimet, J. A.; Paluch, A. S. Predicting Octanol/Water Partition Coefficients for the SAMPL6 Challenge Using the SM12, SM8, and SMD Solvation Models. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 575–588, doi: 10.1007/s10822-020-00293-2.
- (133) Ding, Y.; Xu, Y.; Qian, C.; Chen, J.; Zhu, J.; Huang, H.; Shi, Y.; Huang, J. Predicting Partition Coefficients of Drug-like Molecules in the SAMPL6 Challenge with Drude Polarizable Force Fields. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 421–435, doi: 10.1007/s10822-020-00282-5.
- (134) Fan, S.; Iorga, B. I.; Beckstein, O. Prediction of Octanol-Water Partition Coefficients for the SAMPL6-log P Molecules Using Molecular Dynamics Simulations with OPLS-AA, AMBER and CHARMM Force Fields. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 543–560, doi: 10.1007/s10822-019-00267-z.
- (135) Zamora, W. J.; Pinheiro, S.; German, K.; Ràfols, C.; Curutchet, C.; Luque, F. J. Prediction of the N-Octanol/Water Partition Coefficients in the SAMPL6 Blind Challenge from MST Continuum Solvation Calculations. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 443–451, doi: 10.1007/s10822-019-00262-4.
- (136) Jones, M. R.; Brooks, B. R. Quantum Chemical Predictions of Water–Octanol Partition Coefficients Applied to the SAMPL6 LogP Blind Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 485–493, doi: 10.1007/s10822-020-00286-1.

- (137) Procacci, P.; Guarnieri, G. SAMPL6 Blind Predictions of Water-Octanol Partition Coefficients Using Nonequilibrium Alchemical Approaches. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 371–384, doi: 10.1007/s10822-019-00233-9.
- (138) Sun, Z.; He, Q.; Li, X.; Zhu, Z. SAMPL6 Host–Guest Binding Affinities and Binding Poses from Spherical-Coordinates-Biased Simulations. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 589–600, doi: 10.1007/s10822-020-00294-1.
- (139) Patel, P.; Kuntz, D. M.; Jones, M. R.; Brooks, B. R.; Wilson, A. K. SAMPL6 log P Challenge: Machine Learning and Quantum Mechanical Approaches. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 495–510, doi: 10.1007/s10822-020-00287-0.
- (140) Riquelme, M.; Vöhringer-Martinez, E. SAMPL6 Octanol–Water Partition Coefficients from Alchemical Free Energy Calculations with MBIS Atomic Charges. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 327–334, doi: 10.1007/s10822-020-00281-6.
- (141) Gunner, M. R.; Murakami, T.; Rustenburg, A. S.; Işık, M.; Chodera, J. D. Standard State Free Energies, Not pK_a s, Are Ideal for Describing Small Molecule Protonation and Tautomeric States. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 561–573, doi: 10.1007/s10822-020-00280-7.
- (142) Tielker, N.; Tomazic, D.; Eberlein, L.; Güssregen, S.; Kast, S. M. The SAMPL6 Challenge on Predicting Octanol–Water Partition Coefficients from EC-RISM Theory. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 453–461, doi: 10.1007/s10822-020-00283-4.
- (143) Rizzi, A.; Jensen, T.; Slochow, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoimes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing Challenge: Assessing the Reliability and Efficiency of Binding Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2020**, *34* (5), 601–633, doi: 10.1007/s10822-020-00290-5.
- (144) Wang, S.; Riniker, S. Use of Molecular Dynamics Fingerprints (MDFPs) in SAMPL6 Octanol–Water log P Blind Challenge. *J. Comput. Aided Mol. Des.* **2020**, *34* (4), 393–403, doi: 10.1007/s10822-019-00252-6.
- (145) *Samplchallenges/SAMPL6*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL6>. Accessed: 2021-09-11, doi:10.5281/zenodo.1422854.
- (146) Amezcua, M.; El Houry, L.; Mobley, D. L. SAMPL7 Host–Guest Challenge Overview: Assessing the Reliability of Polarizable and Non-Polarizable Methods for Binding Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 1–35, doi: 10.1007/s10822-020-00363-5.
- (147) Bergazin, T. D.; Tielker, N.; Zhang, Y.; Mao, J.; Gunner, M. R.; Francisco, K.; Ballatore, C.; Kast, S.; Mobley, D. Evaluation of log P , pK_a and log D Predictions from the SAMPL7 Blind Challenge. *J. Comput. Aided Mol. Des.* **2021**, *35*, 771–802, doi: 10.1007/s10822-021-00397-3.
- (148) Eken, Y.; Almeida, N. M. S.; Wang, C.; Wilson, A. K. SAMPL7: Host–Guest Binding Prediction by Molecular Dynamics and Quantum Mechanics. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 63–77, doi: 10.1007/s10822-020-00357-3.
- (149) Huai, Z.; Yang, H.; Li, X.; Sun, Z. SAMPL7 TrimerTrip Host–Guest Binding Affinities from Extensive Alchemical and End-Point Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 117–129, doi: 10.1007/s10822-020-00351-9.

- (150) Kellett, K.; Slochower, D. R.; Schauerperl, M.; Duggan, B. M.; Gilson, M. K. Experimental Characterization of the Association of β -Cyclodextrin and Eight Novel Cyclodextrin Derivatives with Two Guest Compounds. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 95–104, doi: 10.1007/s10822-020-00350-w.
- (151) Khalak, Y.; Tresadern, G.; de Groot, B. L.; Gapsys, V. Non-Equilibrium Approach for Binding Free Energies in Cyclodextrins in SAMPL7: Force Fields and Software. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 49–61, doi: 10.1007/s10822-020-00359-1.
- (152) Procacci, P.; Guarnieri, G. SAMPL7 Blind Predictions Using Nonequilibrium Alchemical Approaches. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 37–47, doi: 10.1007/s10822-020-00365-3.
- (153) Shi, Y.; Laury, M. L.; Wang, Z.; Ponder, J. W. AMOEBA Binding Free Energies for the SAMPL7 TrimerTrip Host–Guest Challenge. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 79–93, doi: 10.1007/s10822-020-00358-2.
- (154) Sun, Z. SAMPL7 TrimerTrip Host–Guest Binding Poses and Binding Affinities from Spherical-Coordinates-Biased Simulations. *J. Comput. Aided Mol. Des.* **2021**, *35* (1), 105–115, doi: 10.1007/s10822-020-00335-9.
- (155) *Samplchallenges/SAMPL7*; The SAMPL Challenges, 2021. Available: <https://github.com/samplchallenges/SAMPL7>. Accessed: 2021-09-11, doi:10.5281/zenodo.3346021.
- (156) Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K. D3R Grand Challenge 2015: Evaluation of Protein–Ligand Pose and Affinity Predictions. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 651–668, doi: 10.1007/s10822-016-9946-8.
- (157) Sunseri, J.; Ragoza, M.; Collins, J.; Koes, D. R. A D3R Prospective Evaluation of Machine Learning for Protein–Ligand Scoring. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 761–771, doi: 10.1007/s10822-016-9960-x.
- (158) Misini Ignjatović, M.; Caldararu, O.; Dong, G.; Muñoz-Gutierrez, C.; Adasme-Carreño, F.; Ryde, U. Binding-Affinity Predictions of HSP90 in the D3R Grand Challenge 2015 with Docking, MM/GBSA, QM/MM, and Free-Energy Simulations. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 707–730, doi: 10.1007/s10822-016-9942-z.
- (159) Salmaso, V.; Sturlese, M.; Cuzzolin, A.; Moro, S. DockBench as Docking Selector Tool: The Lesson Learned from D3R Grand Challenge 2015. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 773–789, doi: 10.1007/s10822-016-9966-4.
- (160) Slynko, I.; Da Silva, F.; Bret, G.; Rognan, D. Docking Pose Selection by Interaction Pattern Graph Similarity: Application to the D3R Grand Challenge 2015. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 669–683, doi: 10.1007/s10822-016-9930-3.
- (161) Ruiz-Carmona, S.; Barril, X. Docking–Undocking Combination Applied to the D3R Grand Challenge 2015. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 805–815, doi: 10.1007/s10822-016-9979-z.
- (162) Prathipati, P.; Nagao, C.; Ahmad, S.; Mizuguchi, K. Improved Pose and Affinity Predictions Using Different Protocols Tailored on the Basis of Data Availability. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 817–828, doi: 10.1007/s10822-016-9982-4.
- (163) Santos-Martins, D. Interaction with Specific HSP90 Residues as a Scoring Function: Validation in the D3R Grand Challenge 2015. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 731–742, doi: 10.1007/s10822-016-9943-y.

- (164) Deng, N.; Flynn, W. F.; Xia, J.; Vijayan, R. S. K.; Zhang, B.; He, P.; Menten, A.; Gallicchio, E.; Levy, R. M. Large Scale Free Energy Calculations for Blind Predictions of Protein–Ligand Binding: The D3R Grand Challenge 2015. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 743–751, doi: 10.1007/s10822-016-9952-x.
- (165) Selwa, E.; Martiny, V. Y.; Iorga, B. I. Molecular Docking Performance Evaluated on the D3R Grand Challenge 2015 Drug-like Ligand Datasets. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 829–839, doi: 10.1007/s10822-016-9983-3.
- (166) Ye, Z.; Baumgartner, M. P.; Wingert, B. M.; Camacho, C. J. Optimal Strategies for Virtual Screening of Induced-Fit and Flexible Target in the 2015 D3R Grand Challenge. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 695–706, doi: 10.1007/s10822-016-9941-0.
- (167) Grudin, S.; Kadukova, M.; Eisenbarth, A.; Marillet, S.; Cazals, F. Predicting Binding Poses and Affinities for Protein - Ligand Complexes in the 2015 D3R Grand Challenge Using a Physical Model with a Statistical Parameter Estimation. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 791–804, doi: 10.1007/s10822-016-9976-2.
- (168) Kumar, A.; Zhang, K. Y. J. Prospective Evaluation of Shape Similarity Based Pose Prediction Method in D3R Grand Challenge 2015. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 685–693, doi: 10.1007/s10822-016-9931-2.
- (169) Piotta, S.; Di Biasi, L.; Fino, R.; Parisi, R.; Sessa, L.; Concilio, S. Yada: A Novel Tool for Molecular Docking Calculations. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 753–759, doi: 10.1007/s10822-016-9953-9.
- (170) *Drugdata/D3R_GC2015*; Drug Design Data Resource, 2021. Available: https://github.com/drugdata/D3R_GC2015. Accessed: 2021-09-11, doi: 10.5281/zenodo.5501709.
- (171) Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V. A.; Walters, W. P.; Kuhn, B.; Rudolph, M. G.; Burley, S. K.; Gilson, M. K.; Amaro, R. E. D3R Grand Challenge 2: Blind Prediction of Protein–Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. *J. Comput. Aided Mol. Des.* **2018**, *32* (1), 1–20, doi: 10.1007/s10822-017-0088-4.
- (172) Kumar, A.; Zhang, K. Y. J. A Cross Docking Pipeline for Improving Pose Prediction and Virtual Screening Performance. *J. Comput. Aided Mol. Des.* **2018**, *32* (1), 163–173, doi: 10.1007/s10822-017-0048-z.
- (173) Olsson, M. A.; García-Sosa, A. T.; Ryde, U. Binding Affinities of the Farnesoid X Receptor in the D3R Grand Challenge 2 Estimated by Free-Energy Perturbation and Docking. *J. Comput. Aided Mol. Des.* **2018**, *32* (1), 211–224, doi: 10.1007/s10822-017-0056-z.
- (174) Rifai, E. A.; van Dijk, M.; Vermeulen, N. P. E.; Geerke, D. P. Binding Free Energy Predictions of Farnesoid X Receptor (FXR) Agonists Using a Linear Interaction Energy (LIE) Approach with Reliability Estimation: Application to the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, *32* (1), 239–249, doi: 10.1007/s10822-017-0055-0.
- (175) Hogues, H.; Sulea, T.; Gaudreault, F.; Corbeil, C. R.; Purisima, E. O. Binding Pose and Affinity Prediction in the 2016 D3R Grand Challenge 2 Using the Wilma-SIE Method. *J. Comput. Aided Mol. Des.* **2018**, *32* (1), 143–150, doi: 10.1007/s10822-017-0071-0.
- (176) Selwa, E.; Elisée, E.; Zavala, A.; Iorga, B. I. Blinded Evaluation of Farnesoid X Receptor (FXR) Ligands Binding Using Molecular Docking and Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2018**, *32* (1), 273–286, doi:

- 10.1007/s10822-017-0054-1.
- (177) Ding, X.; Hayes, R. L.; Vilseck, J. Z.; Charles, M. K.; Brooks, C. L. CDOCKER and λ -Dynamics for Prospective Prediction in D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 89–102, doi: 10.1007/s10822-017-0050-5.
- (178) Salmaso, V.; Sturlese, M.; Cuzzolin, A.; Moro, S. Combining Self- and Cross-Docking as Benchmark Tools: The Performance of DockBench in the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 251–264, doi: 10.1007/s10822-017-0051-4.
- (179) Kadukova, M.; Grudinin, S. Docking of Small Molecules to Farnesoid X Receptors Using AutoDock Vina with the Convex-PL Potential: Lessons Learned from D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 151–162, doi: 10.1007/s10822-017-0062-1.
- (180) Mey, A. S. J. S.; Jiménez, J. J.; Michel, J. Impact of Domain Knowledge on Blinded Predictions of Binding Energies by Alchemical Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 199–210, doi: 10.1007/s10822-017-0083-9.
- (181) Duan, R.; Xu, X.; Zou, X. Lessons Learned from Participating in D3R 2016 Grand Challenge 2: Compounds Targeting the Farnesoid X Receptor. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 103–111, doi: 10.1007/s10822-017-0082-x.
- (182) Baumgartner, M. P.; Evans, D. A. Lessons Learned in Induced Fit Docking and Metadynamics in the Drug Design Data Resource Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 45–58, doi: 10.1007/s10822-017-0081-y.
- (183) Lam, P. C.-H.; Abagyan, R.; Totrov, M. Ligand-Biased Ensemble Receptor Docking (LigBEnD): A Hybrid Ligand/Receptor Structure-Based Approach. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 187–198, doi: 10.1007/s10822-017-0058-x.
- (184) Yakovenko, O.; Jones, S. J. M. Modern Drug Design: The Implication of Using Artificial Neuronal Networks and Multiple Molecular Dynamic Simulations. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 299–311, doi: 10.1007/s10822-017-0085-7.
- (185) Wingert, B. M.; Oerlemans, R.; Camacho, C. J. Optimal Affinity Ranking for Automated Virtual Screening Validated in Prospective D3R Grand Challenges. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 287–297, doi: 10.1007/s10822-017-0065-y.
- (186) Kurkcuoglu, Z.; Koukos, P. I.; Citro, N.; Trellet, M. E.; Rodrigues, J. P. G. L. M.; Moreira, I. S.; Roel-Touris, J.; Melquiond, A. S. J.; Geng, C.; Schaarschmidt, J.; Xue, L. C.; Vangone, A.; Bonvin, A. M. J. J. Performance of HADDOCK and a Simple Contact-Based Protein–Ligand Binding Affinity Predictor in the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 175–185, doi: 10.1007/s10822-017-0049-y.
- (187) Fradera, X.; Verras, A.; Hu, Y.; Wang, D.; Wang, H.; Fells, J. I.; Armacost, K. A.; Crespo, A.; Sherborne, B.; Wang, H.; Peng, Z.; Gao, Y.-D. Performance of Multiple Docking and Refinement Methods in the Pose Prediction D3R Prospective Grand Challenge 2016. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 113–127, doi: 10.1007/s10822-017-0053-2.
- (188) Réau, M.; Langenfeld, F.; Zagury, J.-F.; Montes, M. Predicting the Affinity of Farnesoid X Receptor Ligands through a Hierarchical Ranking Protocol: A D3R Grand Challenge 2 Case Study. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 231–238, doi: 10.1007/s10822-017-0063-0.
- (189) Bhakat, S.; Åberg, E.; Söderhjelm, P. Prediction of Binding Poses to FXR Using

- Multi-Targeted Docking Combined with Molecular Dynamics and Enhanced Sampling. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 59–73, doi: 10.1007/s10822-017-0074-x.
- (190) Padhorny, D.; Hall, D. R.; Mirzaei, H.; Mamonov, A. B.; Moghadasi, M.; Alekseenko, A.; Beglov, D.; Kozakov, D. Protein–Ligand Docking Using FFT Based Sampling: D3R Case Study. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 225–230, doi: 10.1007/s10822-017-0069-7.
- (191) da Silva Figueiredo Celestino Gomes, P.; Da Silva, F.; Bret, G.; Rognan, D. Ranking Docking Poses by Graph Matching of Protein–Ligand Interactions: Lessons Learned from the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 75–87, doi: 10.1007/s10822-017-0046-1.
- (192) Schindler, C.; Rippmann, F.; Kuhn, D. Relative Binding Affinity Prediction of Farnesoid X Receptor in the D3R Grand Challenge 2 Using FEP+. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 265–272, doi: 10.1007/s10822-017-0064-z.
- (193) Athanasiou, C.; Vasilakaki, S.; Dellis, D.; Cournia, Z. Using Physics-Based Pose Predictions and Free Energy Perturbation Calculations to Predict Binding Poses and Relative Binding Affinities for FXR Ligands in the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 21–44, doi: 10.1007/s10822-017-0075-9.
- (194) Gao, Y.-D.; Hu, Y.; Crespo, A.; Wang, D.; Armacost, K. A.; Fells, J. I.; Fradera, X.; Wang, H.; Wang, H.; Sherborne, B.; Verras, A.; Peng, Z. Workflows and Performances in the Ranking Prediction of 2016 D3R Grand Challenge 2: Lessons Learned from a Collaborative Effort. *J. Comput. Aided Mol. Des.* **2018**, 32 (1), 129–142, doi: 10.1007/s10822-017-0072-z.
- (195) *Drugdata/D3R_GC2*; Drug Design Data Resource, 2021. Available: https://github.com/drugdata/D3R_GC2. Accessed: 2021-09-11.
- (196) Gaieb, Z.; Parks, C. D.; Chiu, M.; Yang, H.; Shao, C.; Walters, W. P.; Lambert, M. H.; Nevins, N.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K. D3R Grand Challenge 3: Blind Prediction of Protein–Ligand Poses and Affinity Rankings. *J. Comput. Aided Mol. Des.* **2019**, 33 (1), 1–18, doi: 10.1007/s10822-018-0180-4.
- (197) Xie, B.; Minh, D. D. L. Alchemical Grid Dock (AIGDock) Calculations in the D3R Grand Challenge 3. *J. Comput. Aided Mol. Des.* **2019**, 33 (1), 61–69, doi: 10.1007/s10822-018-0143-9.
- (198) Chaput, L.; Selwa, E.; Elisée, E.; Iorga, B. I. Blinded Evaluation of Cathepsin S Inhibitors from the D3RGC3 Dataset Using Molecular Docking and Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2019**, 33 (1), 93–103, doi: 10.1007/s10822-018-0161-7.
- (199) He, X.; Man, V. H.; Ji, B.; Xie, X.-Q.; Wang, J. Calculate Protein–Ligand Binding Affinities with the Extended Linear Interaction Energy Method: Application on the Cathepsin S Set in the D3R Grand Challenge 3. *J. Comput. Aided Mol. Des.* **2019**, 33 (1), 105–117, doi: 10.1007/s10822-018-0162-6.
- (200) Sunseri, J.; King, J. E.; Francoeur, P. G.; Koes, D. R. Convolutional Neural Network Scoring and Minimization in the D3R 2017 Community Challenge. *J. Comput. Aided Mol. Des.* **2019**, 33 (1), 19–34, doi: 10.1007/s10822-018-0133-y.
- (201) Lam, P. C.-H.; Abagyan, R.; Totrov, M. Hybrid Receptor Structure/Ligand-Based Docking and Activity Prediction in ICM: Development and Evaluation in D3R Grand Challenge 3. *J. Comput. Aided Mol. Des.* **2019**, 33 (1), 35–46, doi: 10.1007/s10822-018-0139-5.

- (202) Nguyen, D. D.; Cang, Z.; Wu, K.; Wang, M.; Cao, Y.; Wei, G.-W. Mathematical Deep Learning for Pose and Binding Affinity Prediction and Ranking in D3R Grand Challenges. *J. Comput. Aided Mol. Des.* **2019**, *33* (1), 71–82, doi: 10.1007/s10822-018-0146-6.
- (203) Ignatov, M.; Liu, C.; Alekseenko, A.; Sun, Z.; Padhorny, D.; Kotelnikov, S.; Kazennov, A.; Grebenkin, I.; Kholodov, Y.; Kolosvari, I.; Perez, A.; Dill, K.; Kozakov, D. Monte Carlo on the Manifold and MD Refinement for Binding Pose Prediction of Protein–Ligand Complexes: 2017 D3R Grand Challenge. *J. Comput. Aided Mol. Des.* **2019**, *33* (1), 119–127, doi: 10.1007/s10822-018-0176-0.
- (204) Koukos, P. I.; Xue, L. C.; Bonvin, A. M. J. J. Protein–Ligand Pose and Affinity Prediction: Lessons from D3R Grand Challenge 3. *J. Comput. Aided Mol. Des.* **2019**, *33* (1), 83–91, doi: 10.1007/s10822-018-0148-4.
- (205) Kumar, A.; Zhang, K. Y. J. Shape Similarity Guided Pose Prediction: Lessons from D3R Grand Challenge 3. *J. Comput. Aided Mol. Des.* **2019**, *33* (1), 47–59, doi: 10.1007/s10822-018-0142-x.
- (206) *Drugdata/D3R_GC3*; Drug Design Data Resource, 2021. Available: https://github.com/drugdata/D3R_GC3. Accessed: 2021-09-11, doi: 10.5281/zenodo.5501712.
- (207) Parks, C. D.; Gaieb, Z.; Chiu, M.; Yang, H.; Shao, C.; Walters, W. P.; Jansen, J. M.; McGaughey, G.; Lewis, R. A.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K. D3R Grand Challenge 4: Blind Prediction of Protein–Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 99–119, doi: 10.1007/s10822-020-00289-y.
- (208) Zou, J.; Tian, C.; Simmerling, C. Blinded Prediction of Protein–Ligand Binding Affinity Using Amber Thermodynamic Integration for the 2018 D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1021–1029, doi: 10.1007/s10822-019-00223-x.
- (209) El Khoury, L.; Santos-Martins, D.; Sasmal, S.; Eberhardt, J.; Bianco, G.; Ambrosio, F. A.; Solis-Vasquez, L.; Koch, A.; Forli, S.; Mobley, D. L. Comparison of Affinity Ranking Using AutoDock-GPU and MM-GBSA Scores for BACE-1 Inhibitors in the D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1011–1020, doi: 10.1007/s10822-019-00240-w.
- (210) Basciu, A.; Koukos, P. I.; Mallocci, G.; Bonvin, A. M. J. J.; Vargiu, A. V. Coupling Enhanced Sampling of the Apo-Receptor with Template-Based Ligand Conformers Selection: Performance in Pose Prediction in the D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 149–162, doi: 10.1007/s10822-019-00244-6.
- (211) Sasmal, S.; El Khoury, L.; Mobley, D. L. D3R Grand Challenge 4: Ligand Similarity and MM-GBSA-Based Pose Prediction and Affinity Ranking for BACE-1 Inhibitors. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 163–177, doi: 10.1007/s10822-019-00249-1.
- (212) Santos-Martins, D.; Eberhardt, J.; Bianco, G.; Solis-Vasquez, L.; Ambrosio, F. A.; Koch, A.; Forli, S. D3R Grand Challenge 4: Prospective Pose Prediction of BACE1 Ligands with AutoDock-GPU. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1071–1081, doi: 10.1007/s10822-019-00241-9.
- (213) Wang, B.; Ng, H.-L. Deep Neural Network Affinity Model for BACE Inhibitors in D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 201–217, doi: 10.1007/s10822-019-00275-z.

- (214) Kadukova, M.; Chupin, V.; Grudinin, S. Docking Rigid Macrocycles Using Convex-PL, AutoDock Vina, and RDKit in the D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 191–200, doi: 10.1007/s10822-019-00263-3.
- (215) Yang, Y.; Lu, J.; Yang, C.; Zhang, Y. Exploring Fragment-Based Target-Specific Ranking Protocol with Machine Learning on Cathepsin S. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1095–1105, doi: 10.1007/s10822-019-00247-3.
- (216) Kumar, A.; Zhang, K. Y. J. Improving Ligand 3D Shape Similarity-Based Pose Prediction with a Continuum Solvent Model. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1045–1055, doi: 10.1007/s10822-019-00220-0.
- (217) Lam, P. C.-H.; Abagyan, R.; Totrov, M. Macrocycle Modeling in ICM: Benchmarking and Evaluation in D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1057–1069, doi: 10.1007/s10822-019-00225-9.
- (218) Nguyen, D. D.; Gao, K.; Wang, M.; Wei, G.-W. MathDL: Mathematical Deep Learning for D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 131–147, doi: 10.1007/s10822-019-00237-5.
- (219) Elisée, E.; Gapsys, V.; Mele, N.; Chaput, L.; Selwa, E.; de Groot, B. L.; Iorga, B. I. Performance Evaluation of Molecular Docking and Free Energy Calculations Protocols Using the D3R Grand Challenge 4 Dataset. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1031–1043, doi: 10.1007/s10822-019-00232-w.
- (220) Shin, W.-H.; Kihara, D. Predicting Binding Poses and Affinity Ranking in D3R Grand Challenge Using PL-PatchSurfer2.0. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1083–1094, doi: 10.1007/s10822-019-00222-y.
- (221) Kotelnikov, S.; Alekseenko, A.; Liu, C.; Ignatov, M.; Padhorny, D.; Brini, E.; Lukin, M.; Coutsiyas, E.; Dill, K. A.; Kozakov, D. Sampling and Refinement Protocols for Template-Based Macrocycle Docking: 2018 D3R Grand Challenge 4. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 179–189, doi: 10.1007/s10822-019-00257-1.
- (222) Stroganov, O. V.; Novikov, F. N.; Medvedev, M. G.; Dmitrienko, A. O.; Gerasimov, I.; Svitanko, I. V.; Chilov, G. G. The Role of Human in the Loop: Lessons from D3R Challenge 4. *J. Comput. Aided Mol. Des.* **2020**, *34* (2), 121–130, doi: 10.1007/s10822-020-00291-4.
- (223) *Drugdata/D3R_GC4*; Drug Design Data Resource, 2021. Available: https://github.com/drugdata/D3R_GC4. Accessed: 2021-09-11, doi: 10.5281/zenodo.5501714.
- (224) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database — Version 2009*; University of Minnesota: Minneapolis, MN 55455, 2009.
- (225) Mobley, D. L.; Guthrie, J. P. FreeSolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput. Aided Mol. Des.* **2014**, 1–10, doi: 10.1007/s10822-014-9747-x.
- (226) Işık, M.; Rustenburg, A.; Rizzi, A.; Gunner, M.; Mobley, D.; Chodera, J. Overview of the SAMPL6 pK_a Challenge: Evaluating Small Molecule Microscopic and Macroscopic pK_a Predictions. *J. Comput. Aided Mol. Des.* **2020**, *35* (2), 131–166, doi: 10.1007/s10822-020-00362-6.
- (227) Shelley, J.; Cholleti, A.; Frye, L.; Greenwood, J.; Timlin, M.; Uchimaya, M. Epik: A Software Program for pK_a Prediction and Protonation State Generation for Drug-like Molecules. *J. Comput. Aided Mol. Des.* **2007**, *21* (12), 681–691, doi: 10.1007/s10822-007-9133-z.

- (228) *MoKa*, Molecular Discovery.; Hertfordshire: UK, 2018.
- (229) *Simulations Plus ADMET Predictor v8.5*, Simulations Plus.; Lancaster: CA, 2018.
- (230) *ACD/PKa Classic (ACD/Percepta Kernel v1.6)*, Advanced Chemistry Development Inc.; Toronto: Canada, 2018.
- (231) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Grundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *J Phys Chem A* **1998**, *102* (40), 7787–7794, doi: 10.1021/jp982638r.
- (232) Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. First Principles Calculations of Aqueous pK_a Values for Organic and Inorganic Acids Using COSMO–RS Reveal an Inconsistency in the Slope of the pK_a Scale. *J Phys Chem A* **2003**, *107* (44), 9380–9386, doi: 10.1021/jp034688o.
- (233) Francisco, K. R.; Varricchio, C.; Paniak, T. J.; Kozlowski, M. C.; Brancale, A.; Ballatore, C. Structure Property Relationships of N-Acylsulfonamides and Related Bioisosteres. *Eur. J. Med. Chem.* **2021**, *218*, 113399, doi: 10.1016/j.ejmech.2021.113399.
- (234) Klamt, A.; Diedenhofen, M. Some Conclusions Regarding the Predictions of Tautomeric Equilibria in Solution Based on the SAMPL2 Challenge. *J. Comput. Aided Mol. Des.* **2010**, *24* (4), 621–625, doi: 10.1007/s10822-010-9332-x.
- (235) Tielker, N.; Eberlein, L.; Hessler, G.; Schmidt, K. F.; Güssregen, S.; Kast, S. M. Quantum–Mechanical Property Prediction of Solvated Drug Molecules: What Have We Learned from a Decade of SAMPL Blind Prediction Challenges? *J. Comput. Aided Mol. Des.* **2020**, doi: 10.1007/s10822-020-00347-5.
- (236) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72* (3), 1047–1069.
- (237) Gallicchio, E.; Lapelosa, M.; Levy, R. M. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein–Ligand Binding Affinities. *J. Chem. Theory Comput.* **2010**, *6* (9), 2961–2977, doi: 10.1021/ct1002913.
- (238) Zheng, L.; Chen, M.; Yang, W. Random Walk in Orthogonal Space to Achieve Efficient Free-Energy Simulation of Complex Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (51), 20227–20232, doi: 10.1073/pnas.0810631106.
- (239) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational Calorimetry: High-Precision Calculation of Host–Guest Binding Thermodynamics. *J. Chem. Theory Comput.* **2015**, *11* (9), 4377–4394, doi: 10.1021/acs.jctc.5b00405.
- (240) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (16), 6358–6363, doi: 10.1073/pnas.1303186110.
- (241) Straatsma, T. P.; McCammon, J. A. Multiconfiguration Thermodynamic Integration. *J. Chem. Phys.* **1991**, *95* (2), 1175–1188, doi: 10.1063/1.461148.
- (242) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22* (2), 245–268, doi: 10.1016/0021-9991(76)90078-4.
- (243) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.* **1998**, *120* (37), 9401–9409, doi: 10.1021/ja981844+.
- (244) Naïm, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.;

- Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. Solvated Interaction Energy (SIE) for Scoring Protein–Ligand Binding Affinities. 1. Exploring the Parameter Space. *J. Chem. Inf. Model.* **2007**, *47* (1), 122–133, doi: 10.1021/ci600406v.
- (245) Chang, C.-E.; Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host–Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126* (40), 13156–13164, doi: 10.1021/ja047115d.
- (246) Zheng, Z.; Ucisik, M. N.; Merz, K. M. The Movable Type Method Applied to Protein–Ligand Binding. *J. Chem. Theory Comput.* **2013**, *9* (12), 5526–5538, doi: 10.1021/ct4005992.
- (247) OpenEye Scientific 2008-03 | CUP IX | Santa Fe, NM. Available: <https://www.eyesopen.com/events/2008/03/cup-09>. Accessed: 2021-09-08.
- (248) SAMPL 1 at First Glance.
- (249) Skillman, G. SAMPL 1 at First Glance; Santa Fe, NM, 2008.
- (250) Skillman, A. G. SAMPL3: Blinded Prediction of Host–Guest Binding Affinities, Hydration Free Energies, and Trypsin Inhibitors. *J. Comput. Aided Mol. Des.* **2012**, *26* (5), 473–474, doi: 10.1007/s10822-012-9580-z.
- (251) Mey, A. S. J. S.; Juárez-Jiménez, J.; Hennessy, A.; Michel, J. Blinded Predictions of Binding Modes and Energies of HSP90- α Ligands for the 2015 D3R Grand Challenge. *Bioorg. Med. Chem.* **2016**, *24* (20), 4890–4899, doi: 10.1016/j.bmc.2016.07.044.
- (252) Elisée, E.; Gapsys, V.; Mele, N.; Chaput, L.; Selwa, E.; de Groot, B. L.; Iorga, B. I. Performance Evaluation of Molecular Docking and Free Energy Calculations Protocols Using the D3R Grand Challenge 4 Dataset. *J. Comput. Aided Mol. Des.* **2019**, *33* (12), 1031–1043, doi: 10.1007/s10822-019-00232-w.
- (253) Dudgeon, T.; Grosjean, H.; Mobley, D.; Delft, F. von; Chodera, J. Compound Database and Subsets Generated by the Fragment Network for Stage 3 of the PHIP2 SAMPL7 Challenge, 2019, doi: 10.5281/zenodo.3576140.
- (254) Wagner, J. R.; Churas, C. P.; Liu, S.; Swift, R. V.; Chiu, M.; Shao, C.; Feher, V. A.; Burley, S. K.; Gilson, M. K.; Amaro, R. E. Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking. *Structure* **2019**, *27* (8), 1326–1335.e4, doi: 10.1016/j.str.2019.05.012.
- (255) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, doi: 10.1021/acs.jctc.1c00302.
- (256) Open Force Field Initiative. Available: <https://openforcefield.org/>. Accessed 2021-09-12.
- (257) Ajani, H.; Pecina, A.; Eyrilmez, S. M.; Fanfrlík, J.; Haldar, S.; Řezáč, J.; Hobza, P.; Lepšík, M. Superior Performance of the SQM/COSMO Scoring Functions in Native Pose Recognition of Diverse Protein–Ligand Complexes in Cognate Docking. *ACS Omega* **2017**, *2* (7), 4022–4029, doi: 10.1021/acsomega.7b00503.
- (258) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203, doi: 10.1039/C6SC05720A.
- (259) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying Machine Learning and Quantum Chemistry with a Deep Neural Network

for Molecular Wavefunctions. *Nat. Commun.* **2019**, *10* (1), 5024, doi:
10.1038/s41467-019-12875-2.

- (260) The SAMPL Roadmap. Available: <https://www.sAMPLchallenges.org/roadmap/>. Accessed 2021-09-11.
- (261) MELLODDY. Available: <https://www.melloddy.eu/>. Accessed 2021-09-12.