

# Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge

Ariën S. Rustenburg,<sup>1,2,\*</sup> Justin Dancer,<sup>3,†</sup> Baiwei Lin,<sup>3,‡</sup> Jianwen A. Feng,<sup>4,§</sup> Daniel F. Ortwine,<sup>3,¶</sup> David L. Mobley,<sup>5,\*\*</sup> and John D. Chodera<sup>2,††</sup>

<sup>1</sup>Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065

<sup>2</sup>Computational Biology Program, Sloan Kettering Institute,

Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States

<sup>3</sup>Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, United States

<sup>4</sup>Denali Therapeutics, 201 Gateway Blvd, South San Francisco, CA 94080, United States

<sup>5</sup>Department of Pharmaceutical Sciences and Department of Chemistry, University of California, Irvine, Irvine, California 92697, United States

(Dated: Received: date / Accepted: date)

Small molecule distribution coefficients between immiscible nonaqueous and aqueous phases—such as cyclohexane and water—measure the degree to which small molecules prefer one phase over another at a given pH. As distribution coefficients capture both thermodynamic effects (the free energy of transfer between phases) and chemical effects (protonation state and tautomer effects in aqueous solution), they provide an exacting test of the thermodynamic and chemical accuracy of physical models without the long correlation times inherent to the prediction of more complex properties of relevance to drug discovery, such as protein-ligand binding affinities. For the SAMPL5 challenge, we carried out a blind prediction exercise in which participants were tasked with the prediction of distribution coefficients to assess its potential as a new route for the evaluation and systematic improvement of predictive physical models. These measurements are typically performed for octanol-water, but we opted to utilize cyclohexane for the nonpolar phase. Cyclohexane was suggested to avoid issues with the high water content and persistent heterogeneous structure of water-saturated octanol phases, since it has which has greatly reduced water content and a homogeneous liquid structure. Using a modified shake-flask LC-MS/MS protocol, we collected cyclohexane/water distribution coefficients for a set of 53 druglike compounds at pH 7.4. These measurements were used as the basis for the SAMPL5 Distribution Coefficient Challenge, where 18 research groups predicted these measurements before the experimental values reported here were released. In this work, we describe the experimental protocol we utilized for measurement of cyclohexane-water distribution coefficients, report the measured data, propose a new bootstrap-based data analysis procedure to incorporate multiple sources of experimental error, and provide insights to help guide future iterations of this valuable exercise in predictive modeling.

*Keywords:* partition coefficients; distribution coefficients; blind challenge; predictive modeling; SAMPL

## I. INTRODUCTION

Rigorous assessment of the predictive performance of physical models is critical in evaluating the current state of physical modeling for drug discovery, assessing the potential impact of current models in active drug discovery projects, and identifying limits of the domain of applicability that require new models or improved algorithms. Past iterations of the SAMPL (**S**tatistical **A**ssessment of the **M**odeling of **P**roteins and **L**igands) experiment have demonstrated that blind predictive challenges can expose weaknesses in computational methods for predicting protein-ligand binding affinities and poses, hydration free energies, and host-guest binding affinities [1–4]. In addition, these blind challenges have contributed new, high-quality datasets to the community that have enabled retrospective validation studies and

data-based parameterization efforts to further advance the current state of physical modeling.

By focusing community effort on the prediction of hydration free energies in the first few iterations of this challenge, the SAMPL experiments have now brought physical modeling approaches to the point where they can reliably identify erroneous experimental data [5]. While hydration free energy exercises have shown their utility in improving the state of physical modeling, they are laborious, require specialized equipment no longer found in modern laboratories, are (at least using traditional protocols) limited in dynamic range, and are of questionable applicability in their ability to mimic protein-to-solvent transfer. As a result, no experimental laboratory has emerged to provide new hydration free energy measurements to sustain this aspect of the SAMPL challenge. We sought to replace this component of the SAMPL challenge portfolio with a new physical property that was easy to measure, accessible to multiple laboratories, had a wide dynamic range (in a free energy scale), and better mimicked physical and chemical effects relevant to protein-to-solvent transfer free energies, but was still free of the conformational sampling challenges protein-ligand binding affinities present. As the measurement of partition and distribution coefficients is now widespread in pharma (due to its relevance in optimizing lipophilicity of small molecules), we posited that a blind challenge centered around the prediction of distribu-

\* [arr2011@med.cornell.edu](mailto:arr2011@med.cornell.edu)

† Current address: Theravance Biopharma, South San Francisco, CA 94080, United States

‡ [lin.baiwei@gene.com](mailto:lin.baiwei@gene.com)

§ [feng@dnli.com](mailto:feng@dnli.com)

¶ [ortwine.daniel@gene.com](mailto:ortwine.daniel@gene.com)

\*\* [dmobley@uci.edu](mailto:dmobley@uci.edu)

†† [john.chodera@choderalab.org](mailto:john.chodera@choderalab.org); Corresponding author

tion coefficients—which face many of the same physical and chemical effects (such as protonation state [6, 7] and tautomer issues [8]) observed in protein-ligand binding—might provide such a challenge.

While the measurement of octanol/water distribution coefficients is commonplace (a 2008 benchmark of structure- and property-based log P prediction methods used 96,000 experimental measurements [9]), a number of previously-reported complications in the physical simulation of 1-octanol suggested that this might be too complex for an initial distribution coefficient challenge [10–13], despite some recent reports of success [14]. In particular, water-saturated octanol is very wet, containing  $47 \pm 1$  mg water/g solution [15], and forms complex microclusters or inverse-micelles that create a heterogeneous environment that persist for long simulation times [10–13]. For the inaugural distribution coefficient challenge in SAMPL5, we therefore chose to measure cyclohexane/water distribution coefficients. The water content of water-saturated cyclohexane is much lower than water-saturated octanol—0.12 mg water/g solution, approximately 400 times smaller [16–18], and possesses no long-lived heterogeneous structure [19].

The number of freely available sources of cyclohexane-water partition is very limited, and for the purpose of the SAMPL5 distribution coefficient challenge [20], blind data was required. As part of an internship program at Genentech arranged by the coauthors, the lead author was dispatched to work out modifications of a high-throughput shake-flask protocol [21] currently in use for octanol/water distribution coefficient measurements. In particular, the low dielectric constant of cyclohexane (2.0243) compared to 1-octanol (10.30) [22] and cyclohexane’s surprising ability to dissolve laboratory consumables presented some unexpected challenges. In this report, we describe the modified protocol that resulted, and provide suggestions on how it can further be refined for future iterations of the distribution coefficient challenge. Of 95 lead-like molecules with diverse functional groups selected for measurement, we report 53 log D measurements that passed quality controls that were used in the SAMPL5 challenge.

To ensure the reported experimental dataset is useful in assessing, falsifying, and improving computational physical models of physical properties, we require a robust approach to estimating the experimental error (uncertainty in experimental measurements). We explored several procedures for propagating known sources of error in the measurement process into the final reported log distribution coefficients, and report those efforts here. Our primary approach features a parametric bootstrap, which allows the use of a physical model of the data generating process to sample additional realizations of the data, using distributions specified in the model. These additional realizations are new data points, over which estimates can be calculated. We compared this to a nonparametric bootstrap, which can be useful if a physical model can not be constructed. This method generates new data points as well, but it constructs them from selection with replacement from the existing data. We also calculated the arithmetic mean and standard error of the measured

data. We hope that future efforts to measure cyclohexane-water distribution coefficients can benefit from the model we have developed, so that this work will also be useful for future challenges.

All code used in the analysis, as well as raw and processed data, can be found at <https://github.com/choderalab/sampl5-experimental-logd-data>.

## Theory of distribution coefficients

The *distribution coefficient*,  $D$ , is a measure of preferential distribution of a given compound (solute) between two immiscible solvents at a specified pH, usually specified as log  $D$  in its base-10 logarithmic form,

$$\log D_{\text{solvent1/solvent2}}^{\text{pH}} = \log_{10} \frac{[\text{Solute}]_{\text{solvent1, pH}}}{[\text{Solute}]_{\text{solvent2, pH}}} \quad (1)$$

Typically, one solvent is aqueous and buffered at the specified pH (e.g. Tris pH 7.4), while the other is apolar (e.g. 1-octanol). At the given pH, the solute may populate multiple protonation or tautomeric states, but the total concentration summed over all states is used in the calculation of concentrations in Equation (1). The total salt concentration of the aqueous phase can also play a role, in case salts can provide stabilization of an ionic state of the ligand in the aqueous phase [23]. Because of this, care must be exercised when comparing distribution coefficients obtained under different experimental conditions.

For the SAMPL5 challenge, we concern ourselves with the cyclohexane-water distribution coefficient, where phosphate-buffered saline (PBS) at pH 7.4 is used for the aqueous phase:

$$\log D_{\text{chx/wat}}^{\text{pH 7.4}} = \log_{10} \frac{[\text{Solute}]_{\text{cyclohexane}}}{[\text{Solute}]_{\text{PBS, pH 7.4}}} \quad (2)$$

Another commonly reported value is the *partition coefficient*  $P$ , which quantifies the relative concentration of the neutral species in each phase, again usually specified in log<sub>10</sub> form,

$$\log P_{\text{chx/wat}} = \log_{10} \frac{[\text{Solute}]_{\text{cyclohexane}}^{\text{neutral}}}{[\text{Solute}]_{\text{PBS, pH 7.4}}^{\text{neutral}}} \quad (3)$$

For ligands with a single titratable site and known  $pK_a$ , one can readily convert between log  $P$  and log  $D$  for a given pH (see, e.g. [23]), but ligands with more complex protonation state effects or tautomeric state effects make accounting for the transfer free energies of all species significantly more challenging.

## II. EXPERIMENTAL METHODS

In the following sections we describe how we measured cyclohexane/water distribution coefficients for the 53 com-

150 pounds displayed in Figure 1. The compound selection pro-  
151 cedure is described in Section II A.

152 Distribution coefficient measurements utilized a shake-  
153 flask approach based on a LC-MS/MS technique previously  
154 developed for 1-octanol/water distribution coefficient mea-  
155 surements [21]. The approach is described in Section II B,  
156 and the procedure is schematically summarized in Figure 2.

157 The measured data was subjected to a quality control pro-  
158 cedure that eliminated measurements thought to be too  
159 unreliable for use in the SAMPL5 challenge (Section II C). Re-  
160 maining data were analyzed using a physical model of the  
161 experiment by means of a parametric bootstrap procedure.  
162 We compared this approach to a nonparametric bootstrap  
163 approach, and the arithmetic mean and standard error of the  
164 data without bootstrap analysis. In Section II D, we describe  
165 each approach. The results for each approach can be found  
166 in Table I.

### 167 A. Compound selection

168 Compounds were initially selected from a database of  
169 9115 lead-like molecules available in eMolecules that were  
170 present in the Genentech chemical stores in quantities of  
171 over 2 mg, with molecular weights between 150-350 Da. The  
172 lower bound on molecular weight was chosen to increase  
173 the likelihood of detectability by mass spectrometry, and the  
174 upper bound to limit molecular complexity.

175 We initially chose approximately 88 compounds based on  
176 several criteria:

- 177 • First, we selected 8 carboxylic acid compounds. These  
178 were of potential interest for the purpose of the chal-  
179 lenge, since it was suspected these could potentially  
180 partition along into the cyclohexane phase together  
181 with water or cations [23].
- 182 • MoKa 2.5 was used to calculate cLogP, cLogD, and  
183 pKa values [24, 25]. This version of MoKa was trained  
184 with Roche internal data to improve accuracy. We se-  
185 lected 20 compounds with predicted pKa values that  
186 would potentially be measurable with a Sirius T3 in-  
187 strument (Sirius Analytical) so validation with an or-  
188 thogonal technique (electrochemical titration) could  
189 be performed in the future.
- 190 • The remaining compounds were divided into 10 equal-  
191 size bins that spanned the predicted dynamic range  
192 of log P values (-3.0 to 6.6), and 6 compounds were  
193 drawn from each bin, to a total of 60.

194 This set of 88 total was later reduced to 64 due to the un-  
195 availability of some compounds or the inability to detect  
196 molecular fragments by mass spectrometry at the time of  
197 measurement. This selection was expanded to include 31  
198 compounds used as internal standards for the previously de-  
199 veloped octanol/water assay protocol [21], bringing the total  
200 number of compounds for which measurements were per-  
201 formed to 95. These compounds were randomly assigned num-  
202 erical SAMPL\_XXX designations for the SAMPL5 blind chal-

203 lenge. After the quality control filtering phase (Section II C),  
204 the resulting data set contained 53 compounds, which are  
205 displayed in Figure 1. Canonical isomeric SMILES represen-  
206 tations for the compounds can also be found in Table S1.  
207 These were generated using OpenEye Toolkits v2015.06 by  
208 converting 3D SDF files, after manually verifying the correct  
209 stereochemistry.

### 210 B. Shake-flask measurement protocol for 211 cyclohexane/water distribution coefficients

212 We adapted a shake-flask assay method from an original  
213 octanol/water LC-MS/MS protocol [21] to accommodate the  
214 use of cyclohexane for the nonaqueous phase. Our modified  
215 protocol is described here, and the procedure is explained  
216 schematically in Figure 2.

217 The log D is estimated by quantifying the concentration  
218 of a solute directly from two immiscible layers, present as an  
219 emulsion in a single vial. Capped glass 1.5 mL auto-injector  
220 vials with PTFE-coated silicone septa<sup>1</sup> were used for parti-  
221 tioning, as cyclohexane was found to dissolve polystyrene  
222 96-well plates used in the original protocol.

223 For each individual experiment, 10  $\mu$ L of 10 mM compound  
224 in dimethyl sulfoxide (DMSO)<sup>2</sup> and 5  $\mu$ L of 200  $\mu$ M propanolol  
225 in acetonitrile (an internal standard) were added to 500  $\mu$ L cy-  
226 clohexane<sup>3</sup>, followed by the addition of 500  $\mu$ L of phosphate  
227 buffered saline (PBS) solution<sup>4</sup>. The ionic components of the  
228 buffer were chosen to replicate the buffer conditions used  
229 in other in-vitro assays at Genentech. Unlike the original  
230 protocol, neither phase was presaturated prior to pipetting.

231 The solute was allowed to partition between solvents  
232 while the mixture was shaken for 50 minutes using a plate  
233 shaker<sup>5</sup> at 800 RPM, while the vials were mounted in a vial  
234 holder and taped down to the sides of the vial holder<sup>6</sup>. The  
235 two solvents were then separated by centrifugation for 5 min-  
236 utes at 3700 RPM in a plate centrifuge, using the plate rotor<sup>7</sup>,  
237 with the vials seated in the same vial holder.

238 Aliquots were extracted from each separated phase using  
239 a standard adjustable micropipette, and transferred into a  
240 384-well glass-coated polypropylene plate for subsequent  
241 quantification<sup>8</sup>. Cyclohexane wells were first prepared with  
242 45  $\mu$ L of 1-octanol<sup>9</sup> per well. 5  $\mu$ L of cyclohexane was ex-

<sup>1</sup> Shimadzu cat. no. 228-45450-91

<sup>2</sup> DMSO stocks from Genentech compound library

<sup>3</sup> ACS grade  $\geq 99\%$ , Sigma-Aldrich cat. no 179191-2L, batch #00555ME

<sup>4</sup> 136 mM NaCl, 2.6 mM KCl, 7.96 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.46 mM KH<sub>2</sub>PO<sub>4</sub>, with pH adjusted to 7.4, prepared by the Genentech Media lab

<sup>5</sup> Thermo Fisher Scientific, Titer Plate Shaker, model: 4625, Waltham, MA, USA

<sup>6</sup> Agilent Technologies, Vial plate for holding 54 x 2 mL vials part no. G2255-68700

<sup>7</sup> Eppendorf, Centrifuge 5804, Hamburg, Germany

<sup>8</sup> 384-well glass coat plate: Thermo Scientific, Microplate, 384-Well; Web-seal Plate; Glass-coated Polypropylene; Square well shape; U-Shape well bottom; 384 wells; 90  $\mu$ L sample volume; catalog number: 3252187

<sup>9</sup> ACROS Organics, 1-octanol 99% pure, catalog number: AC150630010, Geel, Belgium

243 tracted from the top phase by micropipette and mixed with  
244 45  $\mu\text{L}$  of octanol in the 384 well plate. 50  $\mu\text{L}$  of aqueous so-  
245 lution was subsequently extracted from the bottom phase.  
246 The octanol dilution was performed mainly to prevent ac-  
247 cumulation of cyclohexane on the C18 HPLC columns<sup>10</sup> that  
248 were used. For the aqueous (bottom) phase, the aliquot of  
249 50  $\mu\text{L}$  was transferred directly into the 384-well plate, into  
250 wells that did not contain octanol. The 384-well plates were  
251 sealed with using glueless aluminum foil seals<sup>11</sup>, and frag-  
252 ment concentrations assayed using quantitative LC-MS/MS.  
253 Measuring solute distribution into the two phases depends  
254 on two separate mass spectrometry measurements<sup>12</sup>:

- 255 • The solute is analyzed to identify and select parent  
256 and daughter ions, and optimize ion fragment param-  
257 eters<sup>13</sup>.

258 We used a flow rate of 0.2 mL/min, mobile phase of  
259 water/acetonitrile/formic acid (50/50/0.1 v/v/v) and  
260 1.5 minutes run time. All parameters were automati-  
261 cally stored for further MRM analyses. For several com-  
262 pounds, the fragment identification LC-MS/MS proce-  
263 dure did not yield high intensity fragments, and these  
264 could therefore not be measured using the MRM ap-  
265 proach.

- 266 • A separate mass spectrometer is employed using  
267 multiple-reaction monitoring (MRM) to select for par-  
268 ent ions and daughter ions of the solute identified in  
269 the previous step. The mass/charge ( $m/z$ ) intensity  
270 (proportional to the absolute number of molecules) is  
271 quantified as a function of the retention time<sup>14</sup>. Infor-  
272 mation on the gradient can be found in Supplementary  
273 Table 1 of Lin and Pease 2013 [21].

274 Highest  $m/z$  intensity fragments were selected using 5 mM  
275 solutions consisting of 50% DMSO, 50% acetonitrile.

276 From each solvent phase in the partitioning experiment,  
277 one aliquot was prepared, and replicate MRM measurements  
278 were performed 3 times per aliquot. The  $\log D$  can be cal-  
279 culated from the relative MRM-signals, obtained by integrat-  
280 ing the single peak in the MRM-chromatogram, using Equa-  
281 tion (4).

$$\log D_{\text{chx/wat}}^{\text{pH } 7.4} = \log_{10} \frac{\text{MRM signal}_{\text{cyclohexane}} / [d_{\text{chx}} v_{\text{inj, chx}}]}{\text{MRM signal}_{\text{PBS, pH } 7.4} / v_{\text{inj, PBS}}} \quad (4)$$

<sup>10</sup> Waters Xbridge C18 2.130 mm with 2.5  $\mu\text{m}$  particles

<sup>11</sup> Agilent cat no 24214-001

<sup>12</sup> All LC solvents were HPLC-grade and purchased from OmniSolv (Charlotte, NC, USA)

<sup>13</sup> This was done using a Shimadzu NexeraX2 consisting of an LC-30AD(pump), SIL-30AC (auto-injector), and SPD-20AC(UV/VIS detector) with Sciex API4000QTRP (MS)

<sup>14</sup> This was done using a Shimadzu NexeraX2 consisting of an LC-30AD(pump), SIL-30AC (auto-injector), and SPD-20AC(UV/VIS detector) with Sciex API4000 (MS)

282 The cyclohexane signal is normalized by the dilution fac-  
283 tor of our cyclohexane aliquots,  $d_{\text{chx}} = 0.1$ , and the injec-  
284 tion volume  $v_{\text{inj, chx}}$ . As the PBS aliquots were not diluted,  
285 this is only normalized by the injection volume  $v_{\text{inj, PBS}}$ . Ex-  
286 periments were carried out independently at least in dupli-  
287 cate, repeated from the same DMSO stock solutions. In-  
288 jection volumes of the MRM procedure were 1  $\mu\text{L}$  for cy-  
289 clohexane (diluted in octanol), and 2  $\mu\text{L}$  for PBS samples.  
290 For one set of measurements, we carried out 2 additional  
291 repeat experiments with 2  $\mu\text{L}$  injections for cyclohexane  
292 (diluted in octanol), and 1  $\mu\text{L}$  for PBS. This set included  
293 SAMPL5\_003, SAMPL5\_005, SAMPL5\_006, SAMPL5\_011,  
294 SAMPL5\_027, SAMPL5\_049, SAMPL5\_050, SAMPL5\_055 ,  
295 SAMPL5\_058, SAMPL5\_060, and SAMPL5\_061.

### 296 C. Quality control

297 In order to eliminate measurements thought to be too  
298 unreliable for the SAMPL5 challenge, we utilized a simple  
299 quality control filter after MRM quantification. Compounds  
300 where the integrated MRM signal within either phase var-  
301 ied between replicates or repeats by more than a factor of  
302 10 were excluded from further analysis. We additionally re-  
303 moved compounds that exceeded the dynamic range of the  
304 assay because they did not produce a detectable MRM sig-  
305 nal in either the cyclohexane or buffer phases during the  
306 quantification.

### 307 D. Bootstrap analysis

308 Since our ultimate goal is to compare predicted distribu-  
309 tion coefficients to experiment to evaluate the accuracy of  
310 current-generation physical modeling approaches, it is criti-  
311 cal to have an accurate assessment of the uncertainty in  
312 the experimental measurement. Good approaches to uncer-  
313 tainty analysis propagate all known sources of experimental  
314 error into the final estimates of uncertainty. To accomplish  
315 this, we developed a parametric bootstrap model [26] of  
316 the experiment based on earlier work [27], with the goal of  
317 propagating pipetting volume and technical replicate errors  
318 through the complex analysis procedure to estimate their  
319 impact on the overall estimated  $\log D$  measurements.

320 Bootstrap approaches provide new synthetic data sets,  
321 denoted as realizations, sampled using some function of  
322 the observed data that approximates the distribution that  
323 the observed data was drawn from. For each compound  
324 that was measured, suppose our data set provides  $N$  inde-  
325 pendent repeats (from the same stock solution, typically 2  
326 or 4), and 3 technical replicates for each repeat (quantita-  
327 tion experiments from each repeat, typically 3). Each real-  
328 ization of the bootstrap process leads to a new synthetic  
329 data set, of the same size, from which a set of synthetic dis-  
330 tribution coefficients can be computed for the realization.  
331 We applied two additional approaches for comparison to  
332 assess the performance of our parametric bootstrap method  
333 (Section II D 1). One features a nonparametric bootstrap ap-

334 proach (Section II D 2), which does not include any physical  
335 details. The other is a calculation of the arithmetic mean  
336 and standard error that is limited to the observed data (Sec-  
337 tion II D 3).

### 338 1. Parametric bootstrap

339 We used a parametric bootstrap [28] method to introduce  
340 a random bias and variance into the data, based on known  
341 experimental sources. This procedure allows us to use a  
342 model to propagate known uncertainty throughout the pro-  
343 cedure [28]. This allows us to better estimate the distribution  
344 that the observed data was drawn from, so that more accu-  
345 rate estimates of the means and sample variance can be  
346 obtained.

347 Uncertainties in pipetting operations were modeled based  
348 on manufacturer descriptions [29, 30], following the work of  
349 Hanson, Ekins and Chodera [27]. Technical replicate varia-  
350 tion was modeled by calculating the coefficient of variation  
351 (CV) between individual experimental replicates. We then  
352 took the mean CV of the entire data set, which was found  
353 to be  $\sim 0.3$ . As a control, we verified that the CV did not de-  
354 pend on the solvent phase that was measured. We included  
355 this in the parametric model by adding a signal imprecision,  
356 modeled by a normal distribution with zero mean, and a stan-  
357 dard deviation of 0.3. We perform a total of 5 000 realizations  
358 of this process, and calculate statistics over all realizations,  
359 such as the mean (expectation) and standard deviation (esti-  
360 mate of standard error) for each measurement.

### 361 2. Nonparameteric bootstrap

362 A traditional nonparametric Monte Carlo procedure was  
363 applied to resample data points[26]. This approach can es-  
364 timate the distribution that the observed data was drawn  
365 from by resampling from the observed data with replace-  
366 ment, to generate a new set of data points with size equal  
367 to the observed data set. Nonparametric bootstrap can be a  
368 useful approach if larger amounts of data are available, and  
369 a detailed physical model of the experiment is absent. We  
370 implemented the procedure in two stages:

- 371 1. A set of  $N$  repeats is drawn with replacement from the  
372 original set of measured repeats.
- 373 2. For each of the repeats, we similarly draw a set of 3  
374 technical replicates from the original set of technical  
375 replicates.

376 This yields a sample data set with the same size as the origi-  
377 nally observed data ( $N$  repeats, with 3 replicates each). We  
378 perform a total of 5 000 realizations of this process, and calcu-  
379 late statistics over all realizations, such as the mean (expec-  
380 tation) and standard deviation (estimate of standard error)  
381 for each measurement.

### 382 3. Arithmetic mean and sample variance

383 We calculated the arithmetic mean over all replicates and  
384 repeats, and estimated the standard error from the total of 6  
385 or 12 data points, to compare to our bootstrap estimates.<sup>15</sup>

### 386 E. Kernel densities

387 As a visual guide, in Figure 3 data are plotted on top of an  
388 estimated density of points. This density was calculated us-  
389 ing kernel density estimation [31], which is a nonparametric  
390 way to estimate a distribution of points using kernel func-  
391 tions. Kernel functions assign density to individual points in  
392 a data set, so that the combined set of data points reflects  
393 a distribution of of the data. We used the implementation  
394 available in seaborn 0.7.0 [32]. We used a product of Gaus-  
395 sian kernels, with a bandwidth of 0.4 for log D and 0.3 for the  
396 standard error. To prevent artifacts such as negative density  
397 estimates for the standard errors, they were first transformed  
398 by the natural logarithm  $\ln$ , and the results were then con-  
399 verted back into standard errors by exponentiation.

### 400 III. DISTRIBUTION COEFFICIENTS

401 The log D values and their uncertainties for the 53 small  
402 molecules that passed quality controls are presented in Ta-  
403 ble I. In the following two sections, we describe the differ-  
404 ences between the analysis results in more detail.

### 405 A. Mean and standard errors in log D

406 The results from the arithmetic mean and sample variance  
407 calculation ( Section II D 3) are plotted in Figure 3c.

408 Despite the compound selection effort, the distribution  
409 of data along the log D-axis is less dense in the region -1 to  
410 0 log units. The data outside this region seems to be cen-  
411 tered around -2 log units, or around 1 log unit. We could  
412 attribute this distribution of data to coincidence, though this  
413 way warrant future investigations into systematic errors. Us-  
414 ing the arithmetic mean of the combined repeat and replicate  
415 measurements (Section II D 3) the distribution coefficients  
416 measured spanned from -3.9 to 2.5 log units.

417 The log D measurements distribution appears bimodal  
418 along the uncertainty axis. A subset of mostly negative log D  
419 values (Figure 3c) has a smaller estimated standard deviation,  
420 though this is not the case for the majority of negative log D  
421 values. The average standard error, rounded to 1 significant  
422 figure, is 0.2 log units for the arithmetic mean calculation.

<sup>15</sup> For the purpose of the D3R/SAMPL5 workshop, we originally erroneously reported the standard deviation  $\cdot\sqrt{3}$  instead of the standard error  $\cdot\sqrt{3}$ . The factor of  $\sqrt{3}$  corrects the sample standard deviation across all MRM measurements for the correlation between the 3 replicate measurements belonging to a single independent experimental repeat.

## B. Bootstrap results

Estimates of the log D span the range between -3.9 to 2.6 log units, using either of the two bootstrap approaches (Section II D 1 and Section II D 2). The log D estimates do not differ significantly from the arithmetic mean calculations. The difference between the results is seen when we compare the estimated standard errors. When applying our bootstrap procedures (Section II D 1 and Section II D 2), we see an upwards shift in the uncertainties, compared to the sample variance calculations. The nonparametric approach yields an average uncertainty of 0.3 log units. The parametric approach yields an average uncertainty of 0.4 log units. The parametric bootstrap suggests that by propagating errors such as the cyclohexane dilution, and the replicate variability into the model, some of the observed low uncertainties might be an artifact of the low number of measurements. This suggests that simply calculating the arithmetic mean, and standard error of all measured data might not reliably capture the error in the experiments. We also note that for certain compounds, bootstrap distributions exhibit multimodal character and as such, standard errors might not accurately capture the full extent of the experimental uncertainty. We provide the bootstrap sample distributions of the parametric model in the supplementary information.

Using the parametric scheme, we see an average shift of uncertainties to larger values compared to the nonparametric bootstrap. The density estimate suggests we should expect a lower bound to the error that we have now incorporated into the analysis. Not every compound shows the same increase in uncertainty, though if we compare the two bootstrap approaches, results are similar above this empirically observed lower bound. The nonparametric approach returns higher uncertainties for some data on average, but estimates lower uncertainties for some as well. It can be concluded that the error would typically be underestimated without the use of a bootstrap approach. Without a physical model, a nonparametric approach might still underestimate errors due to the limited sample size for each measurement (either 2 or 4 fully independent repeats, and a total of 3 replicates per data point).

## IV. DISCUSSION

### 1. Solvent conditions

It is important to consider the fact that different cosolvents may have on the measured values. The solutions contained approximately 1% DMSO, as well as approximately 0.5% acetonitrile. Further work would benefit from a comparison with experiments starting from dry stocks, and thereby not adding extra solvents. This would eliminate DMSO and acetonitrile, by dispensing compound directly into either cyclohexane, or the mixture of cyclohexane and PBS. In this case, care should be taken that all compound is dissolved. If found to be necessary, we could then consider starting all experiments from dry compound stocks, to entirely eliminate effects from co-

solvents such as DMSO and acetonitrile. This would make experiments much more laborious, and would therefore reduce the bandwidth of the method.

One of the things we could not completely account for in the model was the exact ratio of cyclohexane and PBS solution. We attempted to account for volumetric errors from pipetting by performing independent repeat experiments, although this may still leave some systematic error. Cyclohexane is also a volatile compound, especially compared to water. For comparison, the vapor pressure of cyclohexane is 97.81 torr [33], versus 23.8 torr for water [34]. It is possible that evaporation may occur, which could lead to a systematic overestimation of the cyclohexane volume. For future investigations, it may be fruitful to study the evaporation so that this can be accounted for in the model of the experiment, and to take note of systematic bias in the pipette models used.

### 2. Compound detection limits

Calculations using COSMO-RS software [35] suggested a systematic underestimation of log D values in the negative log unit range, in particularly past a log D of -2. Without further experimental investigation, we can not draw definite conclusions as to whether this is the case, or if so, where the source of the systematic error lies.

One possibility that may cause an artificial reduction of the dynamic range—especially at high log D values—is the potential for MS/MS detector saturation at high ligand concentrations. Previous work (Figure 2 from [21]) examined detector saturation effects, finding it possible to reach sufficiently high compound concentrations (generally  $\geq 10 \mu\text{M}$ ) that MRM is no longer linear in compound concentration for that phase. This work also found that different compounds reach detector saturation at different concentrations [21], in principle requiring an assessment of detector saturation to be performed for each compound. While we could not deduce obvious signs of detector saturation in our LC-MS/MS chromatograms, these effects could be mitigated by performing a dilution series of the aliquots sampled from each phase of the partitioning experiment to ensure detector response is linear in the range of dilutions measured. This may also reveal whether compound dimerization may be a complicating factor in quantitation.

### 3. Experimental design considerations

In order to adjust our experimental setup, we had to switch away from using polystyrene 96 well plates, as these were dissolved by cyclohexane. We attempted the use of glass inserts, and glass tubes but these were too narrow and provided insufficient mixing when shaken. We switched to glass vials because their larger diameter provides improved mixing when shaken. For future work, we would recommend the use of glass coated plates, which have the automation advantages of the plates used in the original protocol [21].

528 Plate seals need to be selected carefully. We experimented  
529 with silicone sealing mats, but these absorbed significant  
530 quantities of cyclohexane. We also had to discontinue use of  
531 aluminum seals that contained glue, since the glue is soluble  
532 in cyclohexane and would contaminate LC-MS/MS measure-  
533 ments. In the end, we used aluminum PlateLoc heat seals  
534 and glass coated 384 well plates to circumvent these issues.

535 Sensitivity also suffered due to the need to dilute cyclohex-  
536 ane in octanol to prevent its accumulation on C18 columns  
537 used in the LC-MS/MS phase of the experiment. Trial injec-  
538 tions on a separate system and chromatograms showed ac-  
539 cumulation of unknown origin at the end of each UV chro-  
540 matogram. Accumulation was reduced by injecting less cy-  
541 clohexane. As a result, we diluted the cyclohexane with 1-  
542 octanol for the experiments described here, and ran blank  
543 injections containing ethanol between batches of 64 mea-  
544 surements to ensure the column was clean.

545 Another change to the protocol that we would like to con-  
546 sider for future measurements is to optimize the time spent  
547 equilibrating the mixture. In this work, we separated phases  
548 via centrifugation and sampled aliquots for concentration  
549 measurement within minutes. The post-centrifugation time  
550 prior to sampling aliquots could be extend to 24 hours to  
551 allow for more equilibration for the solute between phases.  
552 This may have a downside, since we would have to consider  
553 the effects that may follow if compounds prefer to be in the  
554 interface-region between cyclohexane and water, or water  
555 and air. These could cause high local concentrations, intro-  
556 ducing a dependency of the results on exactly which part of  
557 the solution aliquots are taken from. We can get around this  
558 by only taking samples from the pure cyclohexane and aque-  
559 ous regions, avoiding the interfaces. This way, we still get the  
560 right distribution coefficients for partitioning between bulk  
561 phases even if some compound is lost to the interfaces.

562 It may be worthwhile to consider other effects of pipetting  
563 operations on the procedure. Some compounds could poten-  
564 tially stick to the surface of pipettes, or glass surfaces. This  
565 could adversely affect our measurements by changing local  
566 concentrations.

567 We also consider that assay results might be less variable if  
568 we presaturated water and cyclohexane before mixing them.  
569 While cyclohexane and water have much lower mutual sol-  
570 ubility than octanol, it is still possible that this affects the  
571 measurement.

572 The computational end of the challenge featured some  
573 difficulties with respect to interpretation of the data. This  
574 was mainly because we could not separate out many effects  
575 from the data that could affect the interpretation. We ob-  
576 served that for the many effects used to explain discrepan-  
577 cies between model and experiment, none of those could  
578 easily be tested with the current state of the data set. It may  
579 be possible to discombobulate matters such as compound  
580 dimerization, the transfer of water or ions into the cyclohex-  
581 ane phase, from changes in the log D values based on the  
582 molecular properties itself.

583 For future challenges, we would recommend that these  
584 assays are carried out at multiple final concentrations of  
585 the ligand in the assay. This could be achieved using differ-

586 ent volumes of 10 mM ligand stocks. This would help detect  
587 dimerization issues, and may help account for issues with  
588 detector oversaturation. Note that the absolute errors in  
589 these stock volumes will not be critical, since the measure-  
590 ments rely on the relative measurement between the two  
591 phases. We could build models that allow for extrapolation  
592 to the infinite dilution limit, which should then provide sim-  
593 pler test cases for challenge participants to reproduce. On  
594 the opposite end, it may be useful to even investigate ways  
595 to design an experimental set that represents these type of  
596 issues, such as compound dimerization, so that we can focus  
597 more on these.

#### 598 4. Uncertainty analysis

599 We hope the experience from this challenges will lay the  
600 groundwork for improving the reliability of data sets regard-  
601 ing the physical properties that we as a modeling commu-  
602 nity rely on. Many computational studies are limited in the  
603 amount of high-quality experimental data that they have ac-  
604 cess to. Unfortunately, most data is taken straight from liter-  
605 ature tables, without much thought being spent on the data  
606 collection process. By performing the experimental part of  
607 the SAMPL5 challenge we were in the position to provide new  
608 data to the modeling community, with an opportunity to de-  
609 cide on an analysis strategy that suits modeling applications.  
610 This not only allows for blind validation of physics-based  
611 models, but also a re-evaluation of the exact properties a  
612 data set should have to provide utility to the modeling com-  
613 munity. An important fact that we feel needs reemphasizing  
614 is that experimental data is limited in utility by the method  
615 that was used to analyze it.

616 Among the lessons learned from this challenge, we would  
617 recommend that future challenges would also feature a rig-  
618 orous statistical treatment of the experimental analysis pro-  
619 cedure, ideally going beyond these initial efforts. One crucial  
620 part of the analysis procedure is obtaining not only accu-  
621 rate estimates of the observable, but also its uncertainty.  
622 As indicated in our data set, standard error estimates from  
623 small populations may underestimate the error. Several ap-  
624 proaches can be taken to resolve part of this issue. Among the  
625 options are the use of statistical tests, such as the bootstrap-  
626 ping methods we applied in this work. These can help us both  
627 propagate information on uncertainty into the model (such  
628 as a parametric bootstrap) or extract uncertainty already  
629 available in the data (such as nonparametric bootstrap). The  
630 parametric approaches can be improved in terms of the phys-  
631 ical models that are used to analyze the data. These models  
632 should ideally include all known sources of error, such as  
633 pipetting errors, evaporation of solvent, errors in integration  
634 software, fluctuations in temperature, pressure and likewise  
635 many other conditions that could affect the results.

636 Another approach would be to perform statistical infer-  
637 ence on the data set, to provide uncertainty estimates from  
638 the data itself. The model structure can provide ways to  
639 incorporate data and propagate uncertainty from multiple  
640 experiments. Common parameters, such as variance in mea-

641 surements between experiments could be inferred from combin- 693  
642 ing the entire data set into one model. When prior knowl- 694  
643 edge on the experimental parameters is available, a Bayesian 695  
644 model can be used to effectively infer this type of uncertainty 696  
645 from the data, and use it to propagate the error into log D 697  
646 estimates. Distinctions could be made between an objec- 698  
647 tive treatment of the problem, or an empirical Bayesian ap- 699  
648 proach, where prior parameters are derived from the data. 700  
649 One could use a maximum *a posteriori* (MAP) probability ap- 701  
650 proach to obtain an estimate of one of the modes of the 702  
651 parameter distribution. This has obvious downsides when 703  
652 posterior densities are multimodal, and in such a case, one 704  
653 may wish to estimate the shape of the entire posterior distri- 705  
654 bution instead. An approach like Markov chain Monte Carlo 706  
655 could provide such estimates, and will allow for calculation 707  
656 of credible intervals. MCMC methods can be computationally 708  
657 intensive compared to MAP, though if the resulting posterior 709  
658 is complicated, a MAP estimate can give poor results. Un- 710  
659 fortunately, we were unable to construct a Bayesian model 711  
660 of the experiments within our time constraints. We would 712  
661 encourage future challenges to make an attempt at creating 713  
662 a Bayesian model, since this would allow for robust inference 714  
663 of all experimental parameters. 715

#### 664 5. Funding future challenges

665 The execution of this work would not have been possi- 720  
666 ble without the resources provided by Genentech. Access 721  
667 to a rich library of compounds onsite allowed us to select 722  
668 a dataset that was both challenging and useful for the pur- 723  
669 poses of the SAMPL challenge. At the same time, the instru- 724  
670 mentation provided us with the bandwidth to perform many 725  
671 measurements. Rapid redesign of experiments by trial and 726  
672 error, as a result of the difficulties with cyclohexane compati- 727  
673 bility of laboratory consumables and equipment, would not 728  
674 have been possible without the expertise shared by Genen- 729  
675 tech scientists and the opportunities to do many measure- 730  
676 ments. 731

677 Future iterations of this challenge would benefit from con- 732  
678 tinued collaboration between industry and academia. Aca- 733  
679 demic groups can partner with industry groups to pair avail- 734  
680 able skilled academic labor (graduate students and postdoc- 735  
681 toral researchers) with specialized measurement equipment 736  
682 and compound libraries. The graduate student industry in- 737  
683 ternship model proved to be a particularly successful ap- 738  
684 proach, with measurements for a blind challenge providing 739  
685 a well-defined, limited-scope project with clear high value to 740  
686 the modeling community. 741

#### 687 V. CONCLUSION

688 The experimental data provided by this study was very 743  
689 useful for hosting the first small-molecule distribution coeffi- 744  
690 cient challenge in the context of SAMPL. It revealed that log D 745  
691 prediction, as well as measurement, is not always straightfor- 746  
692 ward. We showed that it was possible to perform cyclohex- 747

ane/water log D measurements in the same manner as the 743  
original octanol/water assays, though further optimizations 744  
are needed to reach the same level of throughput. Cyclo- 745  
hexane did pose several challenges for experimental design, 746  
such as the need for different container types, and the po- 747  
tential accumulation of substrate on reversed phase HPLC 748  
columns. 749

750 Many details, such as protonation states, tautomer states, 751  
and dimerization might need to be accounted for in order 752  
to reproduce experiments. This challenge taught us con- 753  
siderations that should be made on the experimental side. 754  
Cases where dimerization were pointed out as possible rea- 755  
son for discrepancy between experiment and model, could 756  
only be hypothesized from the modeling end and not tested 757  
experimentally. Issues with detector saturation could also 758  
be affecting the overall quality of the data set. Future experi- 759  
ments would benefit from more rigorous protocols, such as 760  
measurements at multiple concentrations, and models of all 761  
experimental components. 762

763 We recommend that future challenges, and experiments in 764  
general, use physical models of experiments in the analysis 765  
of experimental uncertainty. These should be part of the 766  
analysis procedure, but also in experimental design. These 767  
will reveal abnormalities in data more clearly. 768

769 We recommend that future challenges look into the use 770  
of bootstrap models such as those considered here. Addi- 771  
tionally, the use of Bayesian inference methods, that allow 772  
the incorporation of prior information should lead to a more 773  
robust estimate of experimental uncertainty. They will allow 774  
for joint inference on multiple experiments, thereby increas- 775  
ing the information gain by using the model. 776

777 Lastly, the sponsoring of this internship by Genentech was 778  
fundamental to generating this data. Access to compound 779  
libraries, and the equipment to perform the experiments is 780  
crucial to the design and execution of a study. Close collabo- 781  
rations with Genentech scientists were important in solving 782  
many technical challenges. The collaboration between indus- 783  
try and academics was not only fruitful, but fundamental in 784  
establishing standardized challenges for the modeling field. 785  
The amount of data we were able to gather would have been 786  
hard to come by without industry resources. At the same 787  
time, the need and expertise in investigating these challeng- 788  
ing physical chemical problems provided by the community, 789  
and the forum provided by the SAMPL challenge was essen- 790  
tial in turning this challenge into a success. We welcome 791  
such future efforts and collaborations, as it is apparent that 792  
both experimental and computational approaches for ob- 793  
taining log D estimates for small molecules, would benefit 794  
from further optimization. 795

#### 796 VI. SUPPLEMENTARY INFORMATION

797 Canonical isomeric smiles for each of the measured com- 798  
pound are available in Table S1. An sdf file containing all com- 799  
pounds, including the measured distribution coefficients is 800  
available as part of the supplementary information. Inte- 801  
grated MRM data including excluded data points are avail- 802



748 able as part of the supplementary information. Bootstrap  
749 distributions from the parametric bootstrap samples for each  
750 compound are provided. We also include a csv file containing  
751 a full list of SAMPL5\_XXX identifiers and canonical isomeric  
752 smiles, including unmeasured compounds.

## 753 VII. FINANCIAL SUPPORT

754 This work was performed as part of an internship by  
755 ASR sponsored by Genentech, Inc., 1 DNA Way, South San  
756 Francisco, CA 94080, United States. JDC acknowledges sup-  
757 port from the Sloan Kettering Institute and NIH grant P30  
758 CA008748. DLM appreciates financial support from National  
759 Science Foundation (CHE 1352608).

## 760 VIII. CONFLICT OF INTEREST STATEMENT

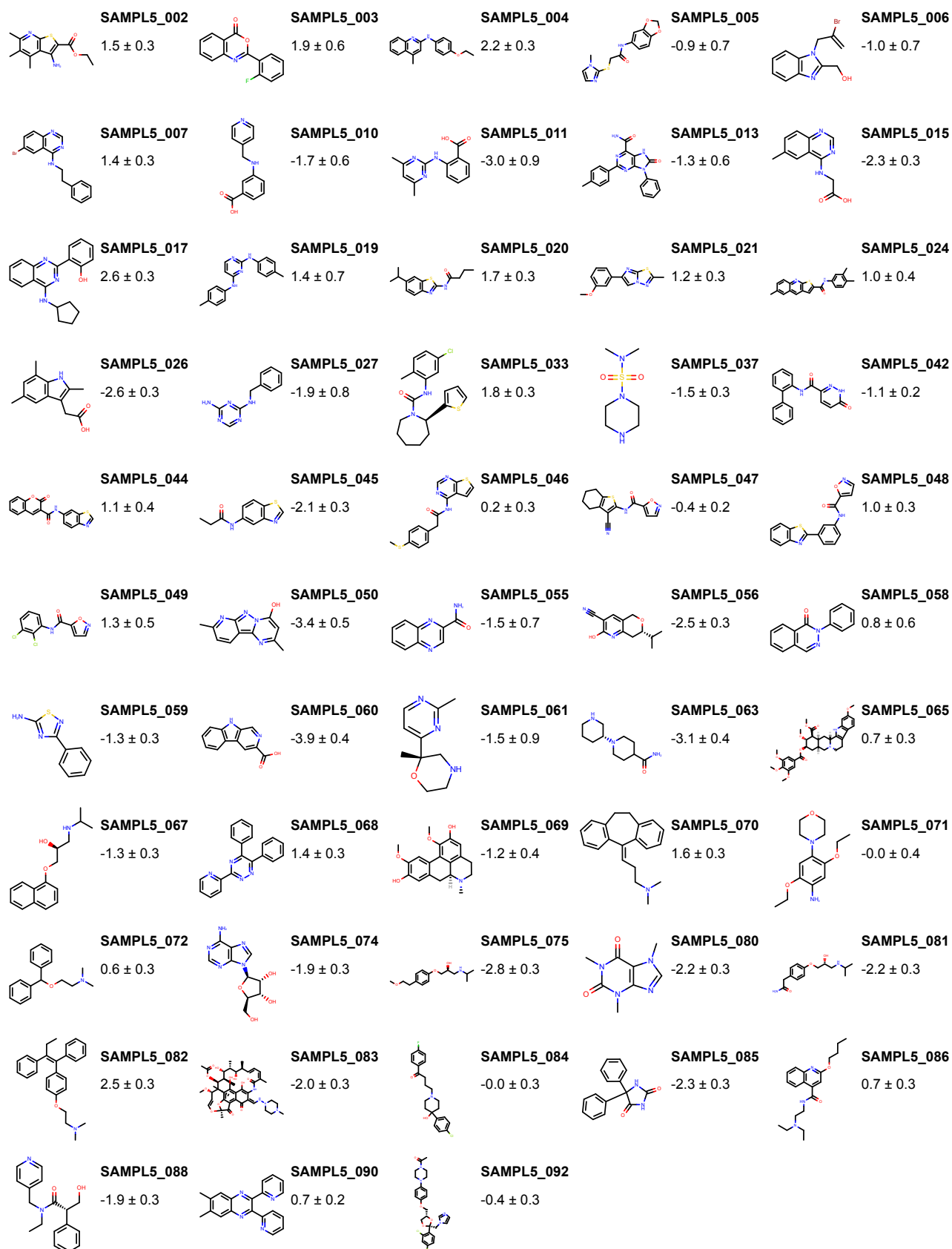
761 DLM and JDC are members of the Scientific Advisory Board  
762 for Schrödinger, LLC.

## IX. ACKNOWLEDGMENTS

764 The authors acknowledge Christopher Bayly (OpenEye Sci-  
765 entific) and Robert Abel (Schrödinger) for their contributions  
766 to discussions on compound selection; Joseph Pease (Genen-  
767 tech) for discussions of the experimental approach and aid in  
768 compound selection; Delia Li (Genentech) for her assistance  
769 in performing experimental work; Alberto Gobbi (Genentech),  
770 Man-Ling Lee (Genentech), and Ignacio Aliagas (Genentech)  
771 for helpful feedback on experimental issues; Andreas Klamt  
772 (Cosmologic) and Jens Reinisch (Cosmologic) for invigorating  
773 discussions regarding experimental data; Patrick Grinaway  
774 (MSKCC) for helpful discussions on analysis procedures; and  
775 Anthony Nicholls (OpenEye) for originating and supporting  
776 earlier iterations of SAMPL challenges.

- 
- 777 [1] J. P. Guthrie, *J Phys Chem B* **113**, 4501 (2009).  
778 [2] M. T. Geballe, A. G. Skillman, A. Nicholls, J. P. Guthrie, and P. J.  
779 Taylor, *J Comput Aided Mol Des* **24**, 259 (2010).  
780 [3] A. G. Skillman, *J Comput Aided Mol Des* **26**, 473 (2012).  
781 [4] H. S. Muddana, A. T. Fenley, D. L. Mobley, and M. K. Gilson, *J*  
782 *Comput Aided Mol Des* **28**, 305 (2014).  
783 [5] D. L. Mobley, K. L. Wymer, N. M. Lim, and J. P. Guthrie, *J Comput*  
784 *Aided Mol Des* **28**, 135 (2014).  
785 [6] P. Czodrowski, C. A. Sotriffer, and G. Klebe, *J. Mol. Biol.* **367**,  
786 1347 (2007).  
787 [7] H. Steuber, P. Czodrowski, C. A. Sotriffer, and G. Klebe, *J. Mol.*  
788 *Biol.* **373**, 1305 (2007).  
789 [8] Y. C. Martin, *J. Comput. Aid. Mol. Des.* **23**, 693 (2009).  
790 [9] R. Mannhold, G. I. Poda, C. Ostermann, and I. V. Tetko, *J. Pharm.*  
791 *Sci.* **98**, 861 (2009).  
792 [10] P. A. Kollman, *Accounts of Chemical Research* **29**, 461 (1996).  
793 [11] S. A. Best, K. M. Merz, and C. H. Reynolds, *The Journal of Phys-*  
794 *ical Chemistry B* **103**, 714 (1999).  
795 [12] B. Chen and J. I. Siepmann, *The Journal of Physical Chemistry*  
796 *B* **110**, 3555 (2006).  
797 [13] A. P. Lyubartsev, S. P. Jacobsson, G. Sundholm, and A. Laakso-  
798 nen, *J. Phys. Chem. B* **105**, 7775 (2001).  
799 [14] N. Bhatnagar, G. Kamath, I. Chelst, and J. J. Potoff, *J. Pharm.*  
800 *Sci.* **137**, 014502 (2012).  
801 [15] S. A. Margolis and M. Levenson, *Fresenius' J. Anal. Chem.* **367**,  
802 1 (2000).  
803 [16] R. Stephenson, J. Stuart, and M. Tabak, *Journal of Chemical &*  
804 *Engineering Data* **29**, 287 (1984).  
805 [17] C. Black, G. G. Joris, and H. S. Taylor, *The Journal of Chemical*  
806 *Physics* **16**, 537 (1948).  
807 [18] S. H. Yalkowsky, Y. He, and P. Jain, *Handbook of aqueous solu-*  
808 *bility data* (CRC press, ADDRESS, 2010).  
809 [19] J. G. Harris and F. H. Stillinger, *J. Chem. Phys.* **95**, 5953 (1991).  
810 [20] C. C. Bannan, K. H. Burley, M. Chiu, M. K. Gilson, and D. L. Mob-  
811 ley, *Journal of Computer-Aided Molecular Design* (2016), un-  
812 published, part of this special issue.
- 813 [21] B. Lin and J. H. Pease, *Comb Chem High Throughput Screen*  
814 **16**, 817 (2013).  
815 [22] W. M. Haynes, *CRC handbook of chemistry and physics* (CRC  
816 press, ADDRESS, 2014).  
817 [23] A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.* **71**, 525 (1971).  
818 [24] F. Milletti, L. Storchi, G. Sforna, and G. Cruciani, *Journal of*  
819 *chemical information and modeling* **47**, 2172 (2007).  
820 [25] F. Milletti, L. Storchi, L. Goracci, S. Bendels, B. Wagner, M. Kansy,  
821 and G. Cruciani, *European journal of medicinal chemistry* **45**,  
822 4270 (2010).  
823 [26] B. Efron, *Ann. Statist.* **7**, 1 (1979).  
824 [27] S. M. Hanson, S. Ekins, and J. D. Chodera, *Journal of computer-*  
825 *aided molecular design* **29**, 1073 (2015).  
826 [28] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*  
827 (CRC press, ADDRESS, 1994).  
828 [29] Rainin Pipet-Lite Multi Pipette L8-200XLS+, <https://www.shoprainin.com/Pipettes/Multichannel-Manual-Pipettes/Pipet-Lite-XLS%2B/Pipet-Lite-Multi-Pipette-L8-200XLS%2B/p/17013805>,  
829 accessed: 2016-06-06.  
830  
831  
832 [30] Rainin Classic Pipette PR-10, <https://www.shoprainin.com/Pipettes/Single-Channel-Manual-Pipettes/RAININ-Classic/Rainin-Classic-Pipette-PR-10/p/17008649>,  
833 accessed: 2016-06-06.  
834  
835 [31] M. Rosenblatt, *Ann. Math. Statist.* **27**, 832 (1956).  
836  
837 [32] M. W. O. B. drewokane; Paul Hobson; Yaroslav Halchenko;  
838 Saulius Lukauskas; Jordi Warmenhoven; John B. Cole; Stephan  
839 Hoyer; Jake Vanderplas; gkunter; Santi Villalba; Eric Quin-  
840 tero; Marcel Martin; Alistair Miles; Kyle Meyer; Tom Augspurger;  
841 Tal Yarkoni; Pete Bachant; Constantine Evans; Clark Fitzgerald;  
842 Tamas Nagy; Erik Ziegler; Tobias Megies; Daniel Wehner;  
843 Samuel St-Jean; Luis Pedro Coelho; Gregory Hitz; Antony Lee;  
844 Luc Rocher; seaborn: v0.7.0 (January 2016), 2016.  
845 [33] *Journal of Chemical and Engineering Data* **12**, 326 (1967).  
846 [34] J. G. Speight *et al.*, *Lange's handbook of chemistry* (McGraw-Hill  
847 New York, ADDRESS, 2005), Vol. 1.  
848

<sup>849</sup> [35] A. Klamt, F. Eckert, J. Reinisch, and K. Wichmann, *Journal of*  
<sup>850</sup> *computer-aided molecular design* (2016), unpublished, part of  
<sup>851</sup> *this special issue.*



**FIG. 1: Molecules and corresponding measured log distribution coefficients for measurements that passed quality controls.** Log D measurements are reported as expectation ± standard errors, calculated using our parametric bootstrap method (Section II D).

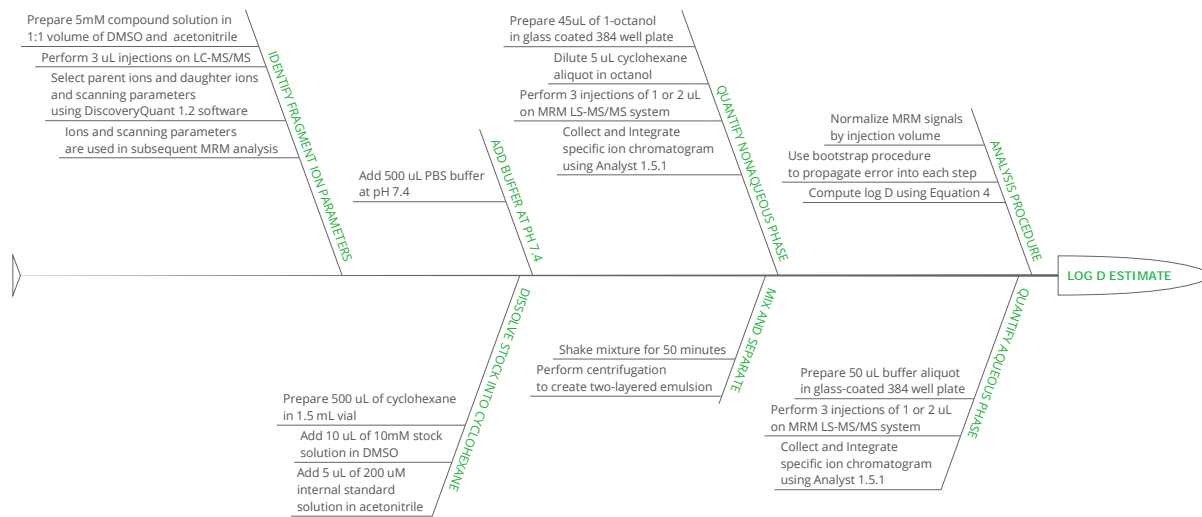
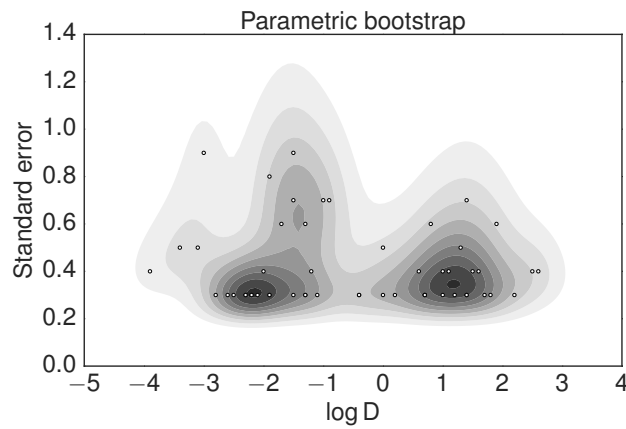


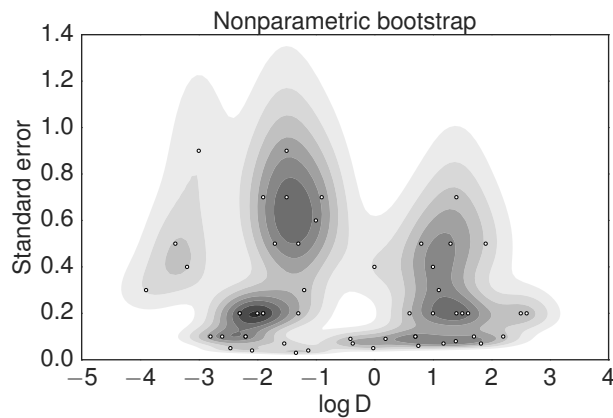
FIG. 2: Illustration of the shake-flask procedure used for cyclohexane-water distribution coefficient measurements.

TABLE I: **Log distribution coefficient measurements and standard errors.** Estimates of log distribution functions and their associated standard errors are described for parametric bootstrap (Section II D 1), nonparametric bootstrap (Section II D 2), and arithmetic mean and corrected sample variance (Section II D 3).

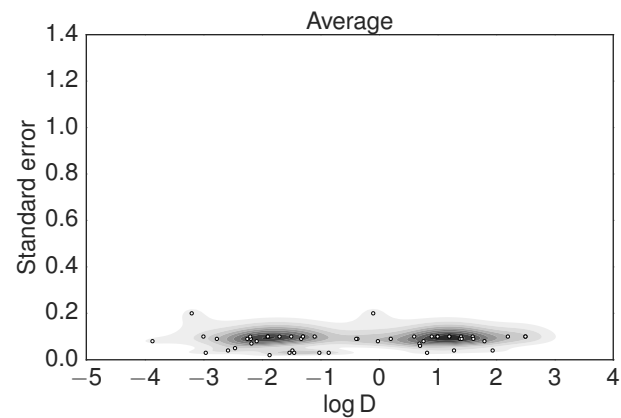
Compound ID	Uncertainty analysis method		
	Bootstrap		Arithmetic mean
	Parametric	Nonparametric	Standard error
SAMPL5_002	1.5 ± 0.3	1.5 ± 0.2	1.4 ± 0.1
SAMPL5_003	1.9 ± 0.6	1.9 ± 0.5	1.94 ± 0.04
SAMPL5_004	2.2 ± 0.3	2.2 ± 0.1	2.2 ± 0.1
SAMPL5_005	-0.9 ± 0.7	-0.9 ± 0.7	-0.86 ± 0.03
SAMPL5_006	-1.0 ± 0.7	-1.0 ± 0.6	-1.02 ± 0.03
SAMPL5_007	1.4 ± 0.3	1.39 ± 0.08	1.38 ± 0.09
SAMPL5_010	-1.7 ± 0.6	-1.7 ± 0.5	-1.7 ± 0.1
SAMPL5_011	-3.0 ± 0.9	-3.0 ± 0.9	-2.96 ± 0.03
SAMPL5_013	-1.3 ± 0.6	-1.3 ± 0.5	-1.5 ± 0.1
SAMPL5_015	-2.3 ± 0.3	-2.3 ± 0.2	-2.25 ± 0.09
SAMPL5_017	2.6 ± 0.3	2.6 ± 0.2	2.5 ± 0.1
SAMPL5_019	1.4 ± 0.7	1.4 ± 0.7	1.2 ± 0.1
SAMPL5_020	1.7 ± 0.3	1.7 ± 0.1	1.6 ± 0.1
SAMPL5_021	1.2 ± 0.3	1.18 ± 0.07	1.2 ± 0.1
SAMPL5_024	1.0 ± 0.4	1.0 ± 0.4	1.0 ± 0.1
SAMPL5_026	-2.6 ± 0.3	-2.6 ± 0.1	-2.58 ± 0.04
SAMPL5_027	-1.9 ± 0.8	-1.9 ± 0.7	-1.87 ± 0.02
SAMPL5_033	1.8 ± 0.3	1.82 ± 0.07	1.80 ± 0.08
SAMPL5_037	-1.5 ± 0.3	-1.54 ± 0.07	-1.53 ± 0.03
SAMPL5_042	-1.1 ± 0.2	-1.13 ± 0.04	-1.1 ± 0.1
SAMPL5_044	1.1 ± 0.4	1.1 ± 0.3	1.0 ± 0.1
SAMPL5_045	-2.1 ± 0.3	-2.09 ± 0.04	-2.09 ± 0.08
SAMPL5_046	0.2 ± 0.3	0.19 ± 0.09	0.20 ± 0.09
SAMPL5_047	-0.4 ± 0.2	-0.37 ± 0.07	-0.37 ± 0.09
SAMPL5_048	1.0 ± 0.3	1.0 ± 0.2	0.9 ± 0.1
SAMPL5_049	1.3 ± 0.5	1.3 ± 0.5	1.28 ± 0.04
SAMPL5_050	-3.4 ± 0.5	-3.4 ± 0.5	-3.2 ± 0.2
SAMPL5_055	-1.5 ± 0.7	-1.5 ± 0.7	-1.48 ± 0.04
SAMPL5_056	-2.5 ± 0.3	-2.46 ± 0.05	-2.46 ± 0.05
SAMPL5_058	0.8 ± 0.6	0.8 ± 0.5	0.82 ± 0.03
SAMPL5_059	-1.3 ± 0.3	-1.34 ± 0.03	-1.33 ± 0.09
SAMPL5_060	-3.9 ± 0.4	-3.9 ± 0.3	-3.87 ± 0.08
SAMPL5_061	-1.5 ± 0.9	-1.5 ± 0.9	-1.45 ± 0.03
SAMPL5_063	-3.1 ± 0.4	-3.2 ± 0.4	-3.0 ± 0.1
SAMPL5_065	0.7 ± 0.3	0.7 ± 0.1	0.69 ± 0.07
SAMPL5_067	-1.3 ± 0.3	-1.3 ± 0.2	-1.3 ± 0.1
SAMPL5_068	1.4 ± 0.3	1.4 ± 0.2	1.41 ± 0.09
SAMPL5_069	-1.3 ± 0.4	-1.2 ± 0.3	-1.3 ± 0.1
SAMPL5_070	1.6 ± 0.3	1.6 ± 0.2	1.61 ± 0.09
SAMPL5_071	-0.0 ± 0.4	-0.0 ± 0.4	-0.1 ± 0.2
SAMPL5_072	0.6 ± 0.3	0.6 ± 0.2	0.6 ± 0.1
SAMPL5_074	-1.9 ± 0.3	-1.9 ± 0.2	-1.9 ± 0.1
SAMPL5_075	-2.8 ± 0.3	-2.8 ± 0.1	-2.77 ± 0.09
SAMPL5_080	-2.2 ± 0.3	-2.2 ± 0.1	-2.18 ± 0.07
SAMPL5_081	-2.2 ± 0.3	-2.2 ± 0.1	-2.19 ± 0.09
SAMPL5_082	2.5 ± 0.3	2.5 ± 0.2	2.5 ± 0.1
SAMPL5_083	-2.0 ± 0.3	-2.0 ± 0.2	-1.9 ± 0.1
SAMPL5_084	-0.0 ± 0.3	-0.02 ± 0.05	-0.02 ± 0.08
SAMPL5_085	-2.3 ± 0.3	-2.3 ± 0.2	-2.2 ± 0.1
SAMPL5_086	0.7 ± 0.3	0.7 ± 0.1	0.70 ± 0.06
SAMPL5_088	-1.9 ± 0.3	-1.9 ± 0.2	-1.9 ± 0.1
SAMPL5_090	0.7 ± 0.2	0.75 ± 0.06	0.76 ± 0.08
SAMPL5_092	-0.4 ± 0.3	-0.41 ± 0.09	-0.39 ± 0.09



(a) **Parametric bootstrap** (Section II D 1). Standard error estimates calculated by using a parametric bootstrap (circles) and a kernel density estimate (contours) of the entire set.



(b) **Nonparametric bootstrap** (Section II D 2). Standard error estimates calculated using a nonparametric bootstrap (circles), and a kernel density estimate (contours) of the entire set.



(c) **Arithmetic mean and sample variance** (Section II D 3). Standard error estimates calculated using corrected sample variance (circles), and a kernel density estimate (contours) of the entire set.

**FIG. 3: Joint kernel density estimates of log distribution coefficient (log D) measurements and measurement error estimates.** log D measurements are plotted with their corresponding estimated standard errors (circles) for the three analysis approaches described in Section II D. A kernel density estimate (contours, described in Section II E) is shown to highlight the differences in error estimates for the different methods.

