A Computational Physics-based Approach to Predict Unbound Brain-to-Plasma Partition Coefficient, Kp,uu

Morgan Lawrenz,*[a] Mats Svensson,[b] Mitsunori Kato,[b] Karen Dingley,[b] Jackson Chief Elk,[c] Zhe Nie,[a] Yefen Zou,[a] Zachary Kaplan,[b] H. Rachel Lagiakos,[b] Hideyuki Igawa,[b] and Eric Therrien*[b]

[a]Schrödinger Inc., San Diego, California, 92122, United States
[b]Schrödinger Inc., New York, New York, 10036, United States
[c]Schrödinger Inc., Portland, Oregon, 97204, United States

ABSTRACT

The blood-brain barrier (BBB) plays a critical role in preventing harmful endogenous and exogenous substances from penetrating the brain. Optimal brain penetration of small molecule CNS drugs is characterized by a high unbound brain/plasma ratio (Kp,uu). While various medicinal chemistry strategies and *in silico* models have been reported to improve BBB penetration, none were developed to predict Kp,uu directly. We describe a physics-based computational approach, solvation free energy calculations (energy of solvation or E-sol), to predict Kp,uu. Prospective application of this method to internal CNS drug discovery programs highlighted the utility and accuracy of this new method, which showed a categorical accuracy of 79% and a R^2 of 0.61 from a linear regression model.
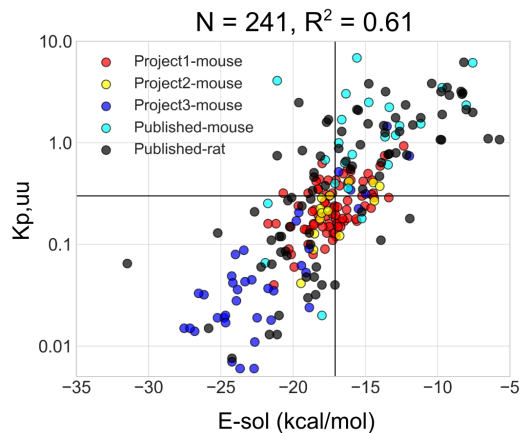
INTRODUCTION

Unbound brain-to-plasma drug partition coefficient (Kp,uu) has been a key parameter to efficiently and reliably support decision making in CNS preclinical projects for many years.[1] As first reported by Prof. Margareta Hammarlund-Udenaes, Kp,uu provides a direct quantitative description of the blood-brain barrier (BBB) permeability represented by passive transport and active influx/efflux.[2] The BBB is enriched with efflux transporters which, as a protective mechanism, reduce the brain penetration of potentially harmful endogenous and exogenous substances.[3–5] Thus, *in vitro* assays are routinely used to identify efflux substrates in projects requiring good brain penetration. The Madin-Darby canine kidney (MDCK) cell line has been used to evaluate the apparent permeability (Papp), estimating the time required to reach distribution equilibrium between brain and plasma. The multidrug resistance protein 1 (MDR1 or P-glycoprotein, P-gp) and breast cancer resistance

protein (BCRP) are two major efflux transporters at the BBB, and their respective assays have been validated to significantly improve the prediction of brain penetration of substrates across species.[6] **Scheme 1** illustrates the equilibrium of drug concentration between the blood and brain, and defines the Kp,uu metric. Fraction unbound (Fu) is a critical parameter for Kp,uu determination that needs to be measured accurately from *in vitro* plasma protein binding and brain tissue binding experiments.

Scheme 1. Schematic representation of drug concentrations between the blood and brain, with definitions for the partition coefficient of unbound drug (Kp,uu). BBB is the blood-brain barrier.

$$K_{p,uu} = C_{unbound,brain} / C_{unbound,plasma}$$
$$C_{unbound} = C_{total} * Fu$$
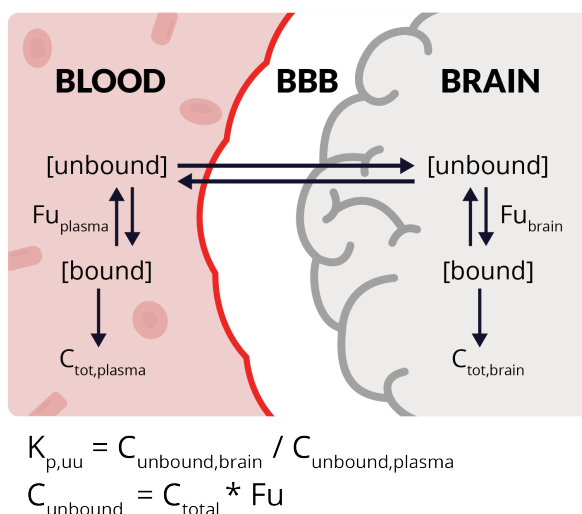
In silico methods have been reported to predict BBB penetration of small molecules. These computational approaches include multi-parameter optimization (MPO) scoring functions,[7–9] Bayesian models (QSAR and machine learning)[10–14] and simple alignment of physicochemical properties (Lipinski rules, RoCNS and others).[15–18] These existing computational protocols proved to be good classifiers for identifying molecules that were more probable to cross the BBB, but to

our knowledge, they lack prospective accuracy and robustness. In addition, none of these methods were intended to predict Kp,uu directly.

Solvation free energy is a fundamental molecular property that connects to several medicinal chemistry end-points, such as solubility, protein-ligand interactions, and any process driven by partitioning, including permeability. It is defined as the energy of hydration of a compound going from the gas phase to water. Solvation free energy can be viewed as one component of logP, the other one being the partition between gas and octanol (Scheme 2).

Scheme 2. Schematic depiction of the partition coefficient (logP) and solvation free energy (E-sol) in water. Figures adapted from the SAMPL6 challenge to predict logP. [19]
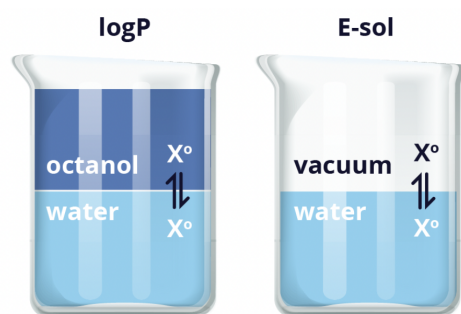
Solvation free energy (E-sol) can be accurately calculated by quantum mechanics (QM) using implicit solvent models.[20–24] It can also be calculated using free energy perturbation calculations (FEP), which is more computationally expensive, but includes effects of explicit waters and enhanced conformational sampling.[25] Herein, we demonstrate that E-sol, a ligand-based and physics-based computational approach, has predictive power for brain Kp,uu and P-glycoprotein mediated efflux. We also show that this approach has a significantly stronger predictive power compared to other calculated metrics with previously demonstrated enrichment of CNS penetrating molecules, such as polar surface area, hydrogen bond donors, molecular weight, logP/D, and

combinatorial metric MPO scores. Prospective application of E-sol in our internal CNS drug discovery programs shows promising performance as a new way of predicting Kp,uu.

RESULTS AND DISCUSSION

Predictive power of E-sol for Kp,uu. To improve permeability and brain exposure, medicinal chemists have used calculated properties such as polar surface area, hydrogen bond donors, molecular weight, and logP/D, also combined as MPO scores or in QSAR models. However, to the best of our knowledge, the accuracy and prospective impact have been limited with respect to direct optimization of key brain penetration measured endpoint Kp,uu. As part of our ongoing drug discovery efforts, we attempted to identify the best possible correlations between a measured Kp,uu endpoint and physicochemical properties of a given molecule. In our work, a strong relationship emerged between E-sol, a physics-based calculated property, and Kp,uu. Predictions of the Kp,uu endpoint by E-sol are shown in **Figure 1** for published Kp,uu datasets,[26–34] as well as data from three Schrödinger CNS projects. Kp,uu is plotted in log scale, to enable better visualization of a linear relationship between the kinetic rate to units of energy (in kcal/mol). Where available, data was filtered by brain unbound fraction (Fu) > 1% and kinetic solubility > 1 μM due to the potential errors introduced in calculating Kp,uu using these measurements (see **Methods** for details on filtering). We found significant linear correlation for E-sol and Kp,uu, with R^2=0.61 for a linear regression model to the data. To characterize predictive power another way, we also showed that E-sol has 79% categorical accuracy for Kp,uu prediction, determined by confusion matrix analysis detailed in **Methods**. In **Figure 1**, a representative favorable Kp,uu of

0.3 is plotted with a horizontal line, and the E-sol threshold which maximizes categorical accuracy for achieving this target Kp,uu value is denoted with a vertical line (-17.1 kcal/mol).

Figure 1. Kp,uu data-points are plotted in log-scale (y-axis) with E-sol (kcal/mol) for compounds from internal projects and published datasets (x-axis). The total number of endpoints was 241 and the R^2 from a linear regression model of the data was 0.61. We denote a favorable Kp,uu of 0.3 with a horizontal line and the E-sol threshold which maximizes categorical accuracy for achieving this target Kp,uu with a vertical line (-17.1 kcal/mol).

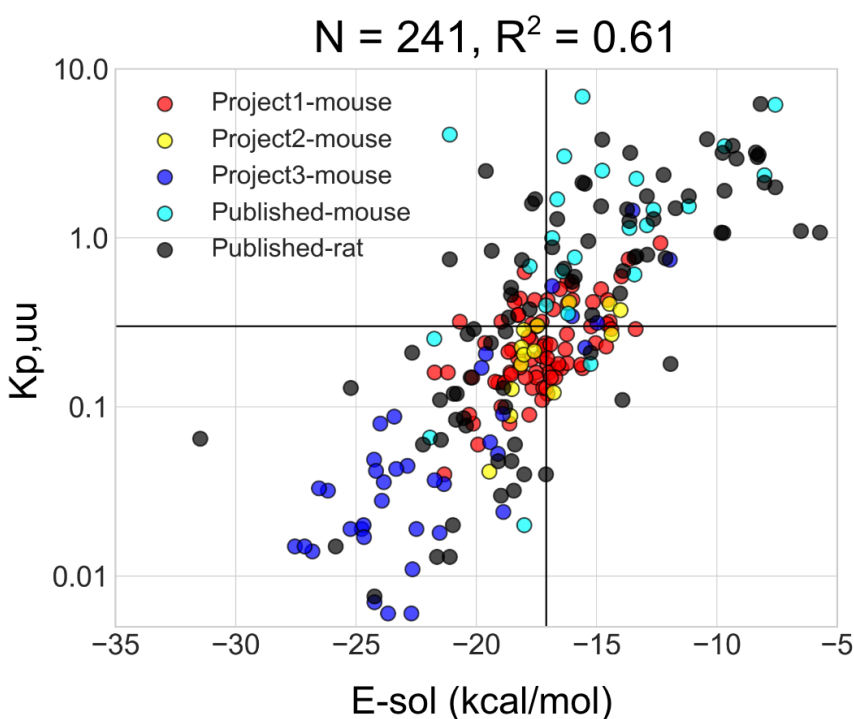**Figure 2** plots the relationship of Kp,uu with other popular calculated ligand descriptors, none of which showed predictive power. By comparison, the trend illustrated for E-sol in **Figure 1** stands out as exceptional for Kp,uu predictions and represents for the first time, to our knowledge, that a calculated ligand property has significant predictive power for Kp,uu. Furthermore, the E-sol

dataset showed no clear systematic shifts in the linear regression model for the compounds grouped

by project to represent distinct chemotypes. This indicated that the trend is robust and transferable.

Figure 2. Kp,uu data-points are plotted in log-scale (y-axis) with calculated metrics for compounds from internal projects and published datasets (x-axis). Calculated endpoints (A) Pfizer MPO,[7] (B) Merck pMPO,[8] (C) molecular weight (MW), (D) polar surface area (PSA), and (E) Schrödinger RRCK predictions[35] are shown. Sample size for the data and R^2 fit from a linear regression model are reported for each plot.
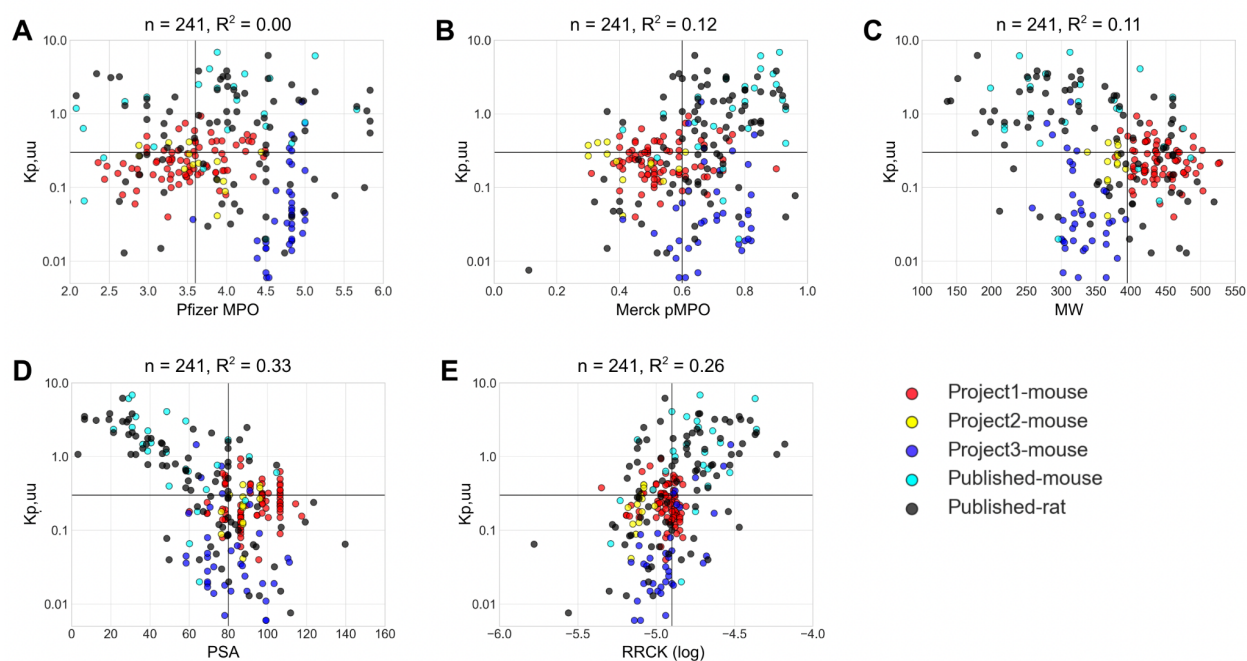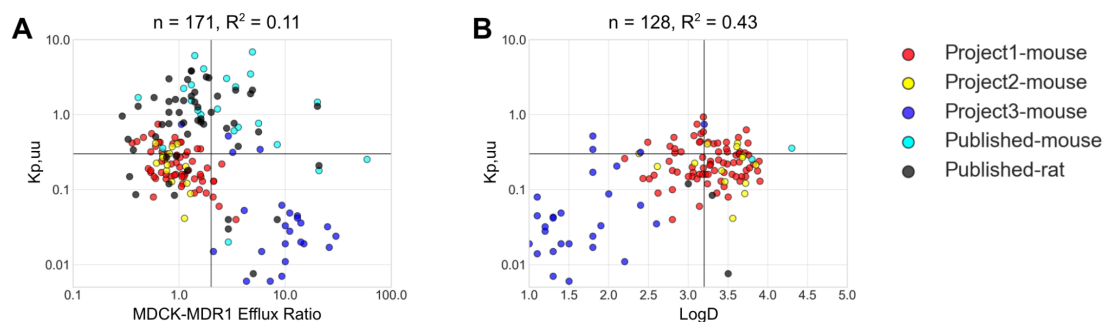

Figure 3. Kp,uu data-points are plotted in log-scale (y-axis) with measured endpoints for compounds from internal projects and published datasets (x-axis). Endpoints (A) MDCK-MDR1 Efflux Ratio and (B) measured logD are shown. Sample size for the data and R^2 fit from linear regression model are reported for each plot.

Comparison of Predictive Power with Other Calculated and Measured Metrics for Kp,uu Prediction. Guidelines for desirable property space to achieve brain penetration have been elaborated and quantified in several CNS multiparameter optimization (MPO) scores,[7,8] which are combinations of calculated properties commonly used in medicinal chemistry, such as logP, polar surface area (PSA), hydrogen bond donor (HBD), molecular weight (MW), and empirical or predicted pKa. Sets of marketed and proprietary compounds were used to establish cutoffs or train these models. The demonstrated utility of the MPO metrics was in retrospective enrichment of brain penetrant compounds and in enrichment of favorable overall ADME (Absorption, Distribution, Metabolism and Excretion) profiles. However, we found the MPO scores not to be predictive of Kp,uu for driving lead optimization stages of internal CNS programs. This is shown in **Figure 2A** and **2B**, in which the MPO scores have no predictive power for Kp,uu. Similar results are shown for single descriptors such as MW and PSA in **Figure 2C** and **2D**.

A low P-gp transporter efflux ratio (ER) is a desirable profile for small molecule drug candidates to achieve adequate oral exposure to reach target tissues. Several brain penetration modeling approaches incorporate ER data into complex trained models to improve performance.[10] Our analysis for Schrödinger drug discovery programs confirmed common knowledge that P-gp substrates are poorly brain penetrant, as most of the molecules with ER > 2 did not achieve a Kp,uu > 0.1. However, in the regime of non-P-gp substrates (ER < 2), results from these *in vitro* experiments alone did not predict Kp,uu, which ranged from less than 0.1 to greater than 1 (**Figure 3A**). In addition, in our analysis of published datasets, there were at least 10 examples of high Kp,uu (Kp,uu > 0.3) compounds with high efflux, up to ER = 20 for Citalopram with Kp,uu (rat) = 1.3. This indicates that even strong P-gp substrates can have good brain penetration, particularly

if they are substrates of uptake transporters.[26,36] Importantly, the E-sol predictive power for Kp,uu held for compounds with Kp,uu > 1 (e.g. Citalopram E-sol = -12.65 kcal/mol).

Passive membrane permeability also has been shown to play a role in BBB function[37,38] and our in-house tool for prediction of Ralph Russ canine kidney cells (RRCK) passive permeability[35] was also compared for Kp,uu prediction (**Figure 2E**). Similar to the trend for ER, the passive permeability model identified compounds with poor brain penetration, but it did not have overall predictive power for Kp,uu. Similarly, experimentally measured logD also provided a necessary, but not sufficient, condition for brain penetration in our dataset. **Figure 3B** shows that the majority of polar (measured logD < 2) compounds were not brain penetrant, but there was no predictive power for compounds in a more drug-like logD range (2-4). This analysis was limited by the scarce availability of measured logD from public Kp,uu datasets. Overall, these results show the utility of efflux ratio and logD measurements for biasing away from poorly brain penetrant molecules, but underscore the inability of these experiments to guide optimization of Kp,uu. E-sol gives a much stronger signal for predicting Kp,uu directly.

Comparison of Categorical Accuracy for E-sol and Other Calculated and Measured Metrics.
For a more coarse evaluation of the metric performance at Kp,uu prediction, we applied a confusion matrix analysis (see **Methods**) to compare compound classification with respect to brain penetration and to calculate categorical accuracy. Shown in **Figure S1**, E-sol categorical accuracy was the highest of the metrics at 79%, and this accuracy was achieved using a threshold of E-sol of -17.1 kcal/mol for our datasets (vertical line in **Figure 1**). Of all metrics, the E-sol method also had the lowest false negative (FN) rate and lowest false positive (FP) rate, at 10% and 11% respectively. This is consistent with the linear correlation shown in **Figure 1**. By comparison, the MPO scores enriched Kp,uu threshold with a more modest, but statistically significant categorical

accuracy of 55% to 63% for our dataset. This performance was slightly inferior to the results reported in the respective publications for these metrics, possibly due to differences in the compound datasets and the challenge in directly predicting Kp,uu endpoint as described here. Notably, MW had similar categorical accuracy at 61%, and the single descriptor PSA had higher categorical accuracy close to 70%. These results reinforce the common practice to design low molecular weight compounds with lower PSA on CNS programs,[39] and they validate use of these properties in the CNS MPO scores evaluated here. Likely due to the smaller size of datasets available, statistically significant categorical accuracy could only be evaluated for measured ER and experimental logD at Kp,uu threshold of 0.2, at 63% and 62%, respectively. Accuracy for Schrödinger RRCK membrane permeability metric[35,40] was 68%. The performance of these last three endpoints further support the strategy of reducing P-gp efflux, controlling lipophilicity, and maintaining good passive permeability to help with CNS penetration. However, E-sol again gives superior performance for prediction of Kp,uu.

Computational Cost and Machine Learning Approximation. The total computational cost of calculating E-sol was dependent on the molecule conformational complexity and the number of conformations sampled (see **Methods**). Considering ligands of variable size and flexibility, our workflow required approximately one CPU-hour per conformation and typically used 8 to 40 sampled conformations (water and gas-phases combined). This process is pleasantly parallel and, with adequate compute resources, can be optimized to concurrently compute E-sol estimates for large sets of molecules within a few hours.

Although the calculations are relatively inexpensive, the overhead of multiple CPU-hours per ligand makes screening ultra-large ligand libraries impractical. Therefore, we have also explored

machine learning models as an alternative method to triage large libraries with modest computing resources. To do so, we trained a supervised machine learning (ML) model[41] on a training set of ~24,000 QM-computed E-sol values for ligands collected from internal Schrödinger projects and public literature sets.[42] The resulting ML model enables an estimation of E-sol in milliseconds per ligand, but can suffer from inaccuracy due to a lack of generalization to novel chemical matter. We show a comparison of the ML model predictions versus the QM-computed E-sol values in **Figure S2**. The E-sol values from internal projects and published datasets were predicted with an MUE of 1.80 kcal/mol by the ML model and the predictions were largely underpredicted. This error increased 2-fold compared with the uncertainty estimated for the QM E-sol method from sampling analysis QM (MUE = 0.84 kcal/mol, see **Methods)**. To demonstrate the utility of the ML model in predicting Kp,uu, **Figure 4** shows the correlation between prospective ML E-sol predictions and log(Kp,uu) measurements for our dataset, with an R^2 for linear regression of 0.5. Categorical accuracy of the ML E-sol model was 72%.

Although the ML model was less accurate than the physics-based calculation, we can leverage the ML E-sol model by first screening large compound libraries, and then verifying compounds with predicted favorable Kp,uu with full-fidelity QM E-sol calculations. This workflow allows the E-sol method to be extended to profile modern and ultra-large virtual libraries or enumerations.[43] When combined with an enterprise informatics platform,[44] the low latency predictions also enable rapid feedback for medicinal chemists during compound ideation, as drawn candidate ligands can be screened in real time.

Figure 4. Correlation of machine learning model E-sol (kcal/mol) predictions, plotted on the x-axis, against log(Kp,uu) measurements, on the y-axis.

n = 241, R^2 = 0.50
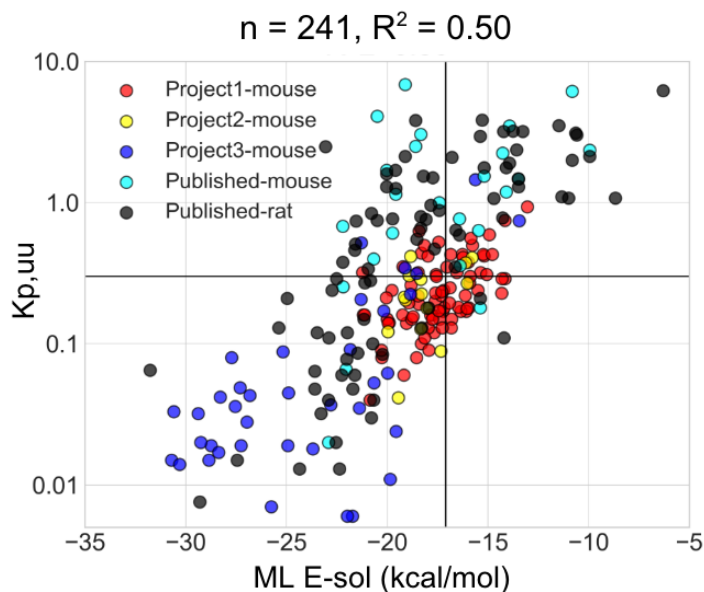
Application to Efflux Prediction. With the known contribution of P-gp mediated efflux to the BBB, we also examined the predictive power of E-sol calculations for ER, shown for 310 compounds across five internal projects (including two non-CNS projects) in **Figure 5**. As expected, we see a good linear relationship with R^2 of 0.64, and 90% categorical accuracy using an E-sol of -19.5 kcal/mol threshold to predict ER < 2. Thus, E-sol calculations can be more broadly applicable to predict efflux ratios.[45] Another QM approach has been previously described to have predictive power on a large efflux dataset from Amgen.[46]

Figure 5. E-sol (kcal/mol) plotted with measured efflux ratios for 310 compounds across five internal projects. $R^2 = 0.64$. Categorical accuracy for prediction of Efflux Ratio < 2 was maximized at 90% with E-sol = -19.5 kcal/mol. P-values << 0.01 as evaluated using Fischer's exact test. All data were from MDCK-MDR1 cells except Project 5, from Caco-2.

$n = 310, R^2 = 0.64$
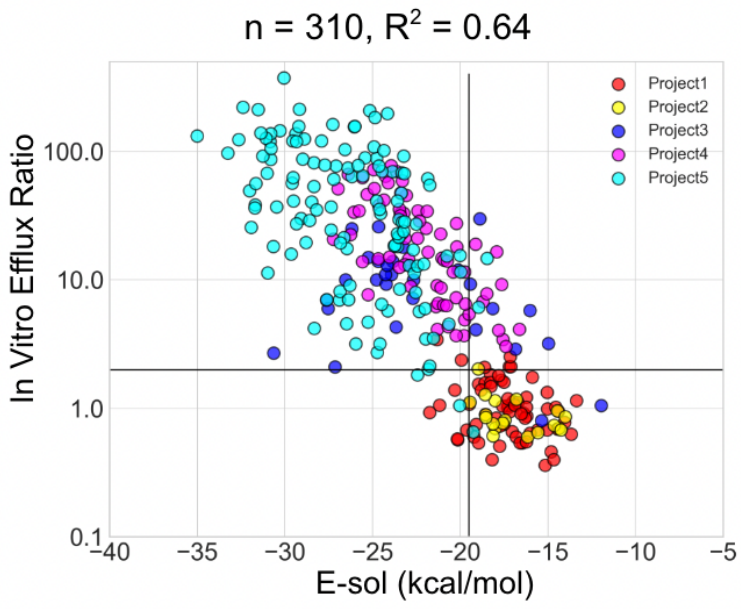
Prospective application of E-sol predictions to internal drug discovery projects. During the lead optimization stage of a drug discovery project, medicinal chemists often struggle to improve brain penetration of lead molecules while maintaining overall favorable druglike properties. At Schrödinger, we have successfully incorporated E-sol predictions into a project multiple parameter optimization (Project MPO) scoring function that includes potency, selectivity, and ADME endpoints predictions, so that design ideas with high probability of achieving brain penetration as predicted by E-sol are prioritized for synthesis. **Tables 1-3** show select matched molecular pairs from an internal CNS project, comparing Kp,uu (determined from *in vivo* mouse PK), the E-sol value, measured logD value, PSA and Merck CNS pMPO score. R group changes from various series are represented. As shown in **Table 1**, compound **1** featuring a basic piperidine (pKa of 8.2) showed reasonable brain penetration as indicated by Kp,uu (0.38). However, safety liabilities such as hERG inhibition with lipophilic basic compounds prompted us to evaluate pKa modulated or neutral molecules (compound **2** and **3**). The E-sol value of **2** (pKa of 5.5) was least favorable among the three, which indeed agreed with the Kp,uu (0.15). Compound **3,** with a neutral tetrahydropyran, showed good brain penetration (Kp,uu = 0.43) and was correctly predicted by E-sol. Merck's CNS pMPO scores were not able to differentiate these molecules in terms of predicted brain penetration. Compounds **4-6** are examples with various ether substitutions off a tetrahydropyran ring to achieve better target selectivity. E-sol values correctly predicted that the OCH_2CF_2 substituted analog **6** would have the worst brain penetration (Kp,uu = 0.09) and that the OEt analog **5** would have favorable brain penetration (Kp,uu = 0.3).

Another example of excellent correlation of E-sol to Kp,uu is shown in **Table 2** for a different chemical series, where various cyclic substitutions were introduced to the imidazopyridine core to adjust physicochemical properties of the lead series, including solubility, permeability and brain

penetration. Interestingly, E-sol correctly identified compound **7** as the least-brain penetrant molecule, while the Merck CNS pMPO predicted favorable scores for all of the molecules (pMPO > 0.6). Additional matched molecular pair examples shown in **Table 3** demonstrated the impact of fluorine atom addition on brain penetration. Incorporation of fluorine atoms into potential drug candidates has been widely used by medicinal chemists to improve metabolic stability, membrane permeability, neighboring group conformation, electronic property of the aromatic rings, and oral bioavailability.[47] E-sol accurately predicted the 7-F indazole analog **13** to be more brain penetrant than 5-F indazole analog **12**, whereas again, the Merck CNS pMPO scores, logD and permeability (MDCK-MDR1; Papp A-B and ER) were similar among all three analogs.

Table 1. Matched molecular pair analysis from a Schrödinger project. R group changes in pyrazolopyridinone series. Kp,uu was obtained either from *in vivo* mouse discrete PK or cassette PK (see Experimental section for details).
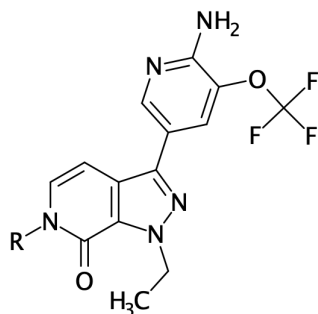
compound	R =	Kp,uu (mouse)	E-sol (kcal/mol)	LogD	MDCK-MDR1 (Papp A-B) (10^{-6} cm/s) / ER	PSA	Merck pMPO	exp pKa
1		0.38	-16.81	2.77	18.6 / 0.6	91	0.67	8.2
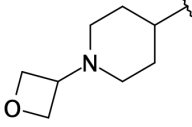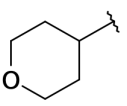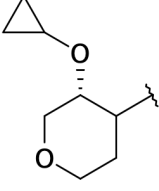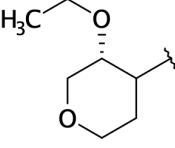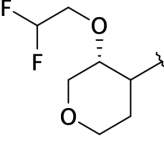
#	Structure							
2		0.15	-20.17	3.05	15.8/ 0.6	100	0.59	5.5
3		0.43	-14.55	3.48	7.6 / 1.0	97	0.51	--
4		0.22	-16.30	3.82	8.0 / 0.8	106	0.41	--
5		0.30	-15.24	3.84	6.13 / 1.5	106	0.43	--
6		0.09	-17.81	3.36	6.8 / 1.8	106	0.40	--

Table 2. Matched molecular pair analysis from a Schrödinger project. R group changes in imidazopyridine series. $K_{p,uu}$ was obtained either from *in vivo* mouse discrete PK or cassette PK (see Experimental section for details).
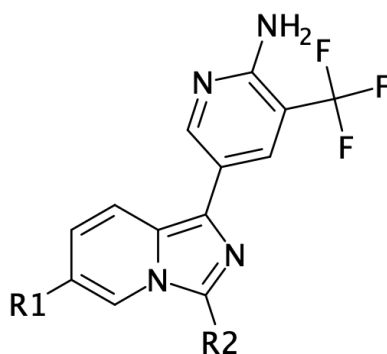
compound	R1 =	R2 =	Kp,uu (mouse)	E-sol (kcal/mol)	LogD	MDCK-MDR1 (Papp A-B) (10^{-6} cm/s) / ER	PSA	Merck pMPO
7		Et	0.08	-20.15	3.14	23.6 / 0.6	77	0.68
8		Et	0.18	-17.07	2.80	10.6 / 1.6	59	0.90
9		Et	0.24	-17.13	3.28	15.6 / 1.0	69	0.73
10		H	1.2	-15.37	3.04	45.5 / 0.4	59	0.94

Table 3. Matched molecular pair analysis from a Schrödinger project showing R group changes in an indazole series. Kp,uu was obtained either from *in vivo* mouse discrete PK or cassette PK (see Experimental section for details).
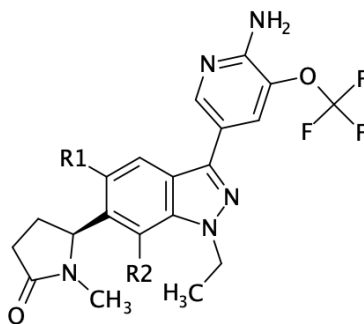
compound	R1 =	R2 =	Kp,uu (mouse)	E-sol (kcal/mol)	LogD	MDCK-MDR1 (Papp A-B) (10^{-6} cm/s) / ER	PSA	Merck pMPO

11	H	H	0.12	-17.05	3.24	9.8 / 1.1	86	0.53
12	F	H	0.13	-17.12	3.65	4.56 / 2.1	86	0.51
13	H	F	0.31	-15.89	3.53	4.8 / 1.8	86	0.51

Application of E-sol to examples from public datasets. Toledo-Sherman et al[32] recently reported on potent Ataxia Telangiectasia-Mutated (ATM) kinase inhibitors. Optimization of permeability and reduction of P-gp mediated efflux led to the discovery of brain-penetrant thioxanthene analogs (**Table 4**). Modifications of compound **14** by the introduction of solubilizing groups to modulate physicochemical properties such as compounds **15-17** led to significant changes in the brain Kp,uu. Although a relationship between Kp,uu and *in vitro* MDCK-MDR1 efflux ratio was observed, the differences in Kp,uu were also correctly captured by the E-sol calculations.
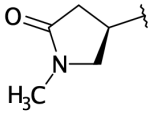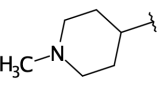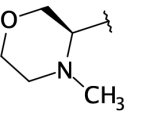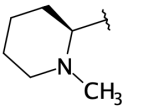
For another example, in an effort to improve brain penetration of clinically proven EGFR TKIs for the treatment of NSCLC with brain metastases, scientists from AstraZeneca worked on a series of 7-methoxyquinazoline carbamates[31] (**Table 5**). Compound **18** demonstrated promising brain penetration (Kp,uu = 0.34). However, it showed high clearance in rats due to oxidative metabolism on the piperazine ring. Compounds **19-25** are examples either with the introduction of methyl substitutions or bridging substitutions to the piperazine ring to add steric hindrance, or examples that changed the piperazine to another moiety (such as pyrrolidine). E-sol predictions accurately rank-ordered brain penetration of these analogs and agreed very well with Kp,uu. Due to the nature of these small changes, it would not be possible to use any other parameters, including Merck CNS pMPO scores, to differentiate the molecules for brain penetration prediction. An important caveat to note is that accuracy in measuring plasma protein binding and brain tissue binding could influence brain Kp,uu. For example, the differences of Kp,uu between enantiomers **22** and **23**

could be due to variability in measured unbound rat plasma protein binding (3.5% for **22** and 6.4%

for **23**),[48] which is beyond the scope of the E-sol calculations.

Table 4. Matched molecular pair analysis from literature paper.[32] Representative thioxanthene analogs with various solubilizing R groups, as potent and selective Ataxia Telangiectasia-Mutated kinase inhibitors for the treatment of Huntington's disease.
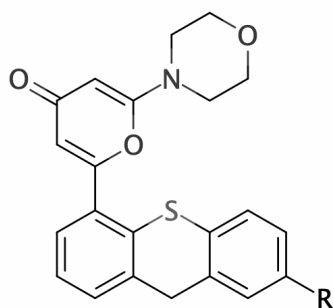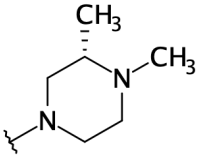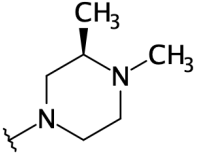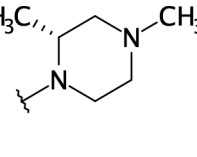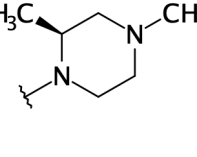
compound	R =	Kp,uu (mouse)	E-sol (kcal/mol)	MDCK (Papp A-B) (10^{-6} cm/s) / ER	PSA	Merck pMPO
14		0.44	-16.92	21.8 / 1.5	52	0.58
15		0.14	-18.77	39.6 / 1.6	61	0.50
16		0.07	-20.76	57.4 / 4.7	61	0.53
17		0.05	-27.88	14.5 / 4.8	75	0.67

Table 5. Matched molecular pair analysis from literature paper.[31] Representative 7-methoxyquinazoline analogs with various R groups, as potent, orally available and CNS- penetrant EGFR tyrosine kinase inhibitors for the treatment of brain metastases.
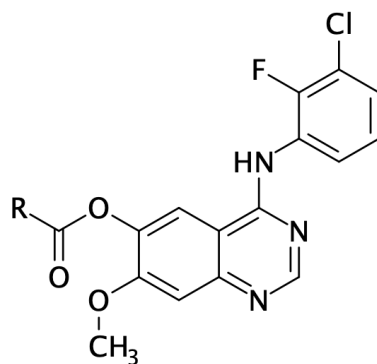
compound	R =	Kp,uu (rat)	E-sol (kcal/mol)	MDCK-MDR1 (Papp A-B) (10^{-6} cm/s) / ER	PSA	Merck pMPO
18		0.34	-18.66	28 / 0.4	80	0.67
19		0.15	-20.31	17 / 0.3	80	0.72
20		0.27	-20.31	47 / 0.8	80	0.72
21		0.086	-20.54	14 / 0.4	80	0.67

22		0.84	-19.01	59 / 1.5	80	0.65
23		0.28	-19.01	54 / 0.9	80	0.65
24 (AZD3759 in PhI)		1.3	-16.50	36 / 0.4	80	0.65
25		1.6	-16.50	51 / 0.9	80	0.65

CONCLUSIONS

With highly curated Kp,uu datasets, we have demonstrated a method with significant predictive power for guiding brain penetration for medicinal chemistry projects. Our method significantly outperformed other reported calculated metrics as well as experimentally measured descriptors such as efflux ratio (ER) and logD. We prospectively use similar E-sol thresholds reported here to predict brain penetration on internal projects, with the specific threshold value refined as data for individual projects is collected. With the categorical accuracies evaluated for other metrics in this work, we also confirm the rationale behind setting molecular property thresholds to enrich brain

penetration, and support common knowledge around improving ADME profiles with high passive permeability and low efflux ratios. The free energy of solvation in water (E-sol) captures an essential character of molecules, including 3D molecular shape and polarity related to permeability and efflux. By computing the partition between water and vacuum, the method gives a clear signal related to the complex *in vivo* endpoint of brain penetration (Kp,uu). The E-sol method is also amenable to machine learning approximation, which permits the screening of ultra-large scale ligand libraries at minimal cost. By combining the machine-learned model with full-fidelity QM calculations, we can enrich for promising brain penetrant compounds from large chemical libraries. Successful prospective applications of this method in our internal projects demonstrated that E-sol provides a promising approach to help develop better brain penetrant drugs in the future.

METHODS

Kp,uu Data Collection. We have collected rodent brain penetration (Kp,uu) data from available published articles[26–34] and curated data from internal Schrödinger drug discovery programs. Internal datasets were obtained from *in vivo* PK studies in mice, while published datasets were from studies in either rats or both mice and rats. We only analyzed brain penetration data that was supported by brain and plasma unbound fraction measurements, and measured efflux ratios were collected when available. Compounds with high brain tissue binding (brain unbound fraction < 1%) or very low kinetic solubility ($< 1 \mu$M) were not included in the analysis due to the potential errors introduced in calculating Kp,uu using these measurements. These and other contributions to errors in Kp,uu measurements are discussed in a subsequent section. Plasma unbound fraction correlated well with brain unbound fraction (**Figure S3**) and was not used as an independent filter for data curation. Altogether, the brain penetration data analyzed in this work originated from 241

diverse endpoints from compounds at various stages of drug development, as well as approved drugs. General properties and a 2D view of the diverse chemical space explored is shown in **Figure S4**.

Computational Protocol. Our E-sol protocol **(Scheme 3)** began with a 2D representation of a molecule as an input. An initial 3D structure of a ligand was generated using the LigPrep[49] module from Maestro[50] in a neutral form. The OPLS4 force field[51] was used for molecular mechanics. Conformations were generated by MacroModel[52] using mixed Monte Carlo Multiple Minimum (MCMM) and Low-Mode Conformational Search (LMCS) methods. GB/SA was used as an implicit water model for MacroModel. Up to 20 low energy conformers were kept for each ligand in water and from gas-phase within 5 kcal/mol from the energy minimum and with the threshold of 0.7Å RMSD to eliminate duplicate conformers. QM geometry optimization of all conformers was performed with B3LYP/LACVP* using Jaguar,[53] and single point energy was calculated with M06-2X/LACVP** with and without PBF water solvation model. The energy difference between lowest energy conformations in gas and water phases was used to calculate the solvation free energy (E-sol, **equation 1**). The state penalty (SP, **equation 2**) was defined as the neutralization energy for a charged molecule at pH 7.4 (there were no acidic molecules in the current set). Measured pKa was used when available, otherwise predicted pKa by Jaguar[54] was used for the state penalty. The Jaguar pKa was further corrected with experimental pKa values of related molecules where appropriate to improve the linear regression model in the Jaguar pKa method. No training on brain penetration or efflux data was involved in the generation of this model.

$$\text{E-sol} = \Delta G_{\text{hydration, neut}} - \text{SP} \qquad\qquad (1)$$

Where $\Delta G_{\text{hydration, neut}}$ is hydration energy of a neutral species, and SP in kcal/mol is the state penalty energy to neutralize a molecule from an ionized state defined as

$$SP = log(1 + 10^{pKa-pH}) * 1.36 \text{ for a base} \tag{2}$$

Scheme 3. E-sol workflow, progressing from 2D molecular representation to a neutral 3D form that was conformationally sampled, geometry optimized in vacuum, then evaluated with single point energy in a PBF water implicit solvent model. The final E-sol value was the dE between lowest energy conformations in gas and water phases, with a state penalty value added for titratable groups.
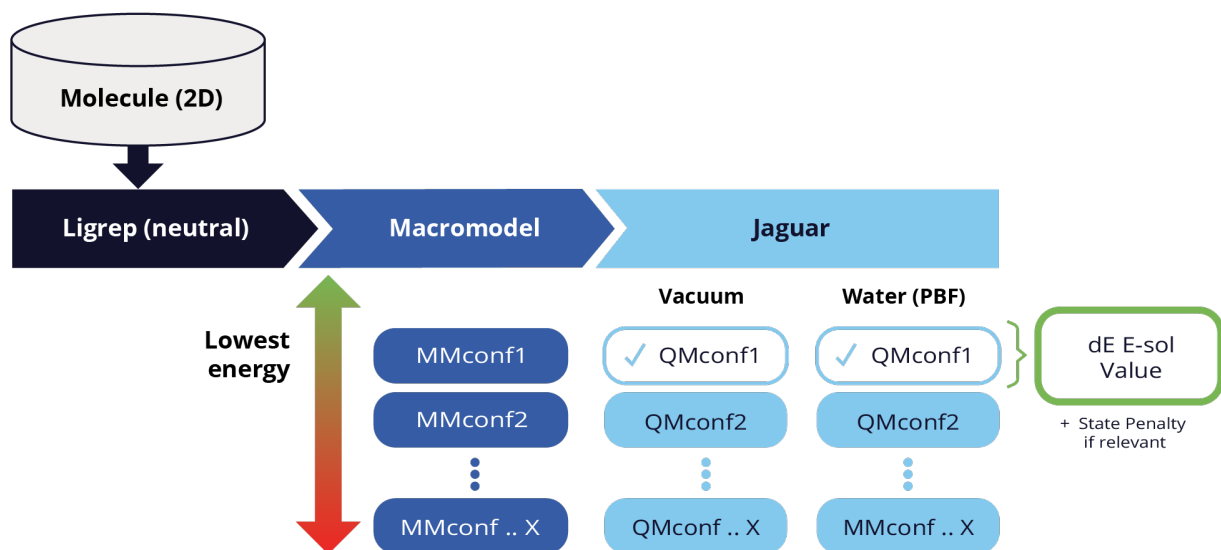
CNS MPO calculations. Pfizer CNS desirability MPO[7] was implemented in LiveDesign.[44] HBD, MW, PSA, and logP were calculated by RDKit,[55] and Epik pKa[56] was used to convert from logP to logD at pH 7.4. Merck CNS pMPO[8] was implemented in LiveDesign[44] using a published protocol.[57]

Confusion matrix analysis. For confusion matrix analysis, thresholds for each metric in **Figures 1, 2 and 3** were evaluated for prediction of Kp,uu of 0.3 and 0.2, maximizing true positive rates and minimizing false positive rates using Receiving Operator Characteristic (ROC) plots. **Table S1** lists the area under the curve (AUC) scores for this analysis, where high values indicated robust and accurate model performance. The E-sol calculation predicted Kp,uu with the highest AUC (AUC = 0.88) compared with other metrics (all AUC ≤ 0.7). Categorical accuracy was calculated from the confusion matrices as (TP+TN)/N, and statistics were evaluated with Fishers' exact test and only reported if p=value < 0.05. We note that a limitation of this matrix approach was the sensitivity of the reported accuracy to the chosen thresholds for classification, particularly for the low to moderate sample sizes here.

Errors in measurements and calculations. The data curation described above removed identifiable uncertainties to allow a clear relationship between brain penetration measurements and our predictions of solvation free energy. The curation also exemplified data that would be used to drive a CNS drug discovery project. We eliminated highly insoluble compounds (kinetic solubility < 1 uM at pH 7.4) and highly brain tissue bound compounds (brain unbound fraction < 1%) from our analysis, because these characteristics can cause a variety of problems in experiments that lead to large variations and uncertainties in key endpoints.

The measured Kp,uu endpoint comes from complex *in vivo* experiments that can have errors from various sources. We have combined rat and mouse data for final analysis in **Figures 1**, **2**, and **3** to increase statistical powering of our study, but differences in P-gp expression in these species could contribute to variance in Kp,uu and affect accuracy of model predictions.[58] Differences in exposures due to cassette vs. discrete PK experiments, and the particular time point of the measurement, contribute to variation in Kp,uu, with errors up to 26% observed in our internal

Project 1 data (n = 16). From the same Project 1, we found that repeated measurements of *in vitro* measured plasma and brain tissue binding could give variance that impacts Kp,uu by up to 20%.

In addition to errors from experiments, model performance was affected by uncertainties in our calculation. These uncertainties come from limitations of the employed sampling algorithms in the water and gas phase, as well as errors in scoring that most likely come from the implicit solvent model. To better understand the sensitivity of the conformational sampling, we have compared the E-sol values using quick (4 conformations in each gas and water phases) and extensive (20 conformations) sampling methods. **Figure S5A** shows the correlation between the 2 methods with an error estimate of 0.84 kcal/mol. Although there can be significant differences in results from the two sampling methods, the impact on the R^2 for model prediction of Kp,uu was negligible (**Figure S5B** and **S5C**). From our analysis, we confirmed that compounds with higher flexibility (number of rotatable bonds > 5) usually required the extensive sampling method to converge the energy value, as illustrated in **Figure S6**. We have ongoing work to further optimize the workflow by comparing implicit *vs* explicit solvent models, using QM *vs* free energy perturbation to compute the energies, and evaluating various conformational sampling techniques to increase overall calculation speed. However, the method described in **Scheme 3** was used for all calculations in this work.

EXPERIMENTAL SECTION

General

All compounds used in this study originating from our internal projects were >95% pure by HPLC analysis.

***In vitro* ADME profiling**. Data on the following *in vitro* ADME properties for in house CNS projects were generated at Pharmaron *in vitro* ADME laboratory (Beijing, China), utilizing the corresponding HT assays: (a) P-gp efflux (ER) was measured utilizing the MDCKII-MDR1 cell line, an MDCK line stably transfected with the MDR1 gene, which expresses a functionally active human P-gp. The cell line was obtained from the Netherlands Cancer Institute (Amsterdam), used at passages between 10 and 20. Passive apparent permeability was determined in both directions, from apical to basolateral (A to B) and basolateral to apical (B to A) directions. Efflux ratios were calculated using the resultant Papp values (ER = Papp(B to A) / Papp (A to B)). 3% BSA (bovine serum albumin) was added into the incubation system to reduce nonspecific binding of the compounds to the plates thus to increase recovery. (B) Plasma protein binding was measured in pooled CD-1 mouse plasma using equilibrium dialysis method using 5 uM of the compound with 6 hr incubation; (C) Brain tissue binding was measured in CD-1 mouse brain homogenate using equilibrium dialysis method and 1 uM of the compound with 6 hr incubation.

***In vivo* brain penetration (Kp,uu)**. Mouse Kp,uu for in house CNS projects was obtained either from a CD-1 mouse IV cassette PK (each study includes 4 compounds plus 1 positive control, 0.5 mpk IV dose, CD-1 mouse, N=3) at 30 min and/or 2 hr post-dose, or from a single dose discrete oral PK study (5 mpk PO dose, CD-1 mouse, N=3 for each time point) using AUC over 24 hr period. Multiple PK studies have confirmed that Kp,uu determined from IV cassette PK data at 2 h post-dose was generally consistent with discrete PK data.

CORRESPONDING AUTHORS

Morgan Lawrenz - Email: morgan.lawrenz@schrodinger.com

Eric Therrien - Email: eric.therrien@schrodinger.com

ACKNOWLEDGMENTS

ABBREVIATIONS USED

Kp,uu, unbound brain-to-plasma drug partition coefficient; CNS, central nervous system; MPO, multiparameter optimization; MW, molecular weight; BBB, blood-brain barrier; MDCK, Madin-Darby canine kidney; Papp, apparent permeability; MDR1, multidrug resistance protein 1; P-gp, P-glycoprotein 1; BCRP, breast cancer resistance protein; QSAR, quantitative structure–activity relationship; RoCNS, Lipinski's rule for central nervous system drugs; PSA, polar surface area; PK, pharmacokinetics; logP, partition coefficient; logD, distribution coefficient; FEP, free energy perturbation; E-sol, Energy of solvation; Fu, unbound fraction; HBD, hydrogen bond donor; ADME, Absorption, Distribution, Metabolism and Excretion; ER, efflux ratio; RRCK, Ralph Russ canine kidney; ROC, Receiving Operator Characteristic; AUC, area under the curve; TP, true positive; TN, true negative; FP, false positive; FN, false negative; CPU, central processing unit; ML, machine learning; Caco-2, human epithelial cell line; hERG, human ether-a-go-go-related gene; ATM, Ataxia Telangiectasia-Mutated; EGFR TKI, EGFR tyrosine kinase inhibitor; NSCLC, non-small cell lung cancer; QM, quantum mechanics; t-SNE, t-distributed stochastic neighbor embedding; MCMM, Monte Carlo Multiple Minimum; LMCS, Low-Mode Conformational

Search; GB/SA, Generalized Born/Solvent Accessible Surface; RMSD, root mean square deviation; PBF, Poisson Boltzmann Finite; SP, state penalty; MUE, mean unsigned error.

REFERENCES

(1) Loryan, I.; Reichel, A.; Feng, B.; Bundgaard, C.; Shaffer, C.; Kalvass, C.; Bednarczyk, D.; Morrison, D.; Lesuisse, D.; Hoppe, E.; Terstappen, G. C.; Fischer, H.; Di, L.; Colclough, N.; Summerfield, S.; Buckley, S. T.; Maurer, T. S.; Fridén, M. Unbound Brain-to-Plasma Partition Coefficient, Kp,uu,brain—a Game Changing Parameter for CNS Drug Discovery and Development. *Pharm. Res.* **2022**, *39* (7), 1321–1341.

(2) Hammarlund-Udenaes, M.; Fridén, M.; Syvänen, S.; Gupta, A. On the Rate and Extent of Drug Delivery to the Brain. *Pharm. Res.* **2008**, *25* (8), 1737–1750.

(3) Pardridge, W. M. The Blood-Brain Barrier: Bottleneck in Brain Drug Development. *NeuroRx* **2005**, *2* (1), 3–14.

(4) Begley, D. J. ABC Transporters and the Blood-Brain Barrier. *Curr. Pharm. Des.* **2004**, *10* (12), 1295–1312.

(5) Löscher, W.; Potschka, H. Blood-Brain Barrier Active Efflux Transporters: ATP-Binding Cassette Gene Family. *NeuroRx* **2005**, *2* (1), 86–98.

(6) Feng, B.; West, M.; Patel, N. C.; Wager, T.; Hou, X.; Johnson, J.; Tremaine, L.; Liras, J. Validation of Human MDR1-MDCK and BCRP-MDCK Cell Lines to Improve the Prediction of Brain Penetration. *J. Pharm. Sci.* **2019**, *108* (7), 2476–2483.

(7) Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach to Enable Alignment of Druglike Properties. *ACS Chem. Neurosci.* **2010**, *1* (6), 435–449.

(8) Gunaydin, H. Probabilistic Approach to Generating MPOs and Its Application as a Scoring Function for CNS Drugs. *ACS Med. Chem. Lett.* **2016**, *7* (1), 89–93.

(9) Gupta, M.; Lee, H. J.; Barden, C. J.; Weaver, D. F. The Blood–Brain Barrier (BBB) Score. *J. Med. Chem.* **2019**, *62* (21), 9824–9836.

(10) Watanabe, R.; Esaki, T.; Ohashi, R.; Kuroda, M.; Kawashima, H.; Komura, H.; Natsume-Kitatani, Y.; Mizuguchi, K. Development of an In Silico Prediction Model for P-Glycoprotein Efflux Potential in Brain Capillary Endothelial Cells toward the Prediction of Brain Penetration. *J. Med. Chem.* **2021**, *64* (5), 2725–2738.

(11) Zhang, L.; Zhu, H.; Oprea, T. I.; Golbraikh, A.; Tropsha, A. QSAR Modeling of the Blood–Brain Barrier Permeability for Diverse Organic Compounds. *Pharm. Res.* **2008**, *25* (8), 1902–1914.

(12) Toropov, A. A.; Toropova, A. P.; Beeg, M.; Gobbi, M.; Salmona, M. QSAR Model for Blood-Brain Barrier Permeation. *J. Pharmacol. Toxicol. Methods* **2017**, *88* (Pt 1), 7–18.

(13) Kumar, R.; Sharma, A.; Alexiou, A.; Bilgrami, A. L.; Kamal, M. A.; Ashraf, G. M. DeePred-BBB: A Blood Brain Barrier Permeability Prediction Model With Improved Accuracy. *Front. Neurosci.* **2022**, *16*, 858126.

(14) Liu, L.; Zhang, L.; Feng, H.; Li, S.; Liu, M.; Zhao, J.; Liu, H. Prediction of the Blood–Brain Barrier (BBB) Permeability of Chemicals Based on Machine-Learning and Ensemble Methods. *Chem. Res.*

Toxicol. **2021**, *34* (6), 1456–1467.

(15) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv*. *Drug Deliv*. *Rev*. **1997**, *23* (1), 3–25.

(16) Rankovic, Z. CNS Drug Design: Balancing Physicochemical Properties for Optimal Brain Exposure. *J*. *Med*. *Chem*. **2015**, *58* (6), 2584–2608.

(17) Xiong, B.; Wang, Y.; Chen, Y.; Xing, S.; Liao, Q.; Chen, Y.; Li, Q.; Li, W.; Sun, H. Strategies for Structural Modification of Small Molecules to Improve Blood–Brain Barrier Penetration: A Recent Perspective. *J*. *Med*. *Chem*. **2021**, *64* (18), 13152–13173.

(18) Fernandes, T. B.; Segretti, M. C. F.; Polli, M. C.; Parise-Filho, R. Analysis of the Applicability and Use of Lipinski's Rule for Central Nervous System Drugs. *Letters in Drug Design & Discovery*. 2016, pp 999–1006. https://doi.org/10.2174/1570180813666160622092839.

(19) Işık, M.; Bergazin, T. D.; Fox, T.; Rizzi, A.; Chodera, J. D. Assessing the Accuracy of Octanol–water Partition Coefficient Predictions in the SAMPL6 Part II Log P Challenge. *Journal of computer* **2020**.

(20) Cortis, C. M.; Friesner, R. A. An Automatic Three-Dimensional Finite Element Mesh Generation System for the Poisson-Boltzmann Equation. *J*. *Comput*. *Chem*. **1997**, *18* (13), 1570–1590.

(21) Cortis, C. M.; Friesner, R. A. Numerical Solution of the Poisson-Boltzmann Equation Using Tetrahedral Finite-Element Meshes. *J*. *Comput*. *Chem*. **1997**, *18* (13), 1591–1608.

(22) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. New Model for Calculation of Solvation Free Energies: Correction of Self-Consistent Reaction Field Continuum Dielectric Theory for Short-Range Hydrogen-Bonding Effects. *J*. *Phys*. *Chem*. **1996**, *100* (28), 11775–11788.

(23) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 4 and Intramolecular Charge Polarization. *J*. *Chem*. *Theory Comput*. **2007**, *3* (6), 2046–2054.

(24) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *J*. *Chem*. *Theory Comput*. **2007**, *3* (6), 2011–2033.

(25) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J*. *Am*. *Chem*. *Soc*. **2015**, *137* (7), 2695–2703.

(26) Summerfield, S. G.; Zhang, Y.; Liu, H. Examining the Uptake of Central Nervous System Drugs and Candidates across the Blood-Brain Barrier. *J*. *Pharmacol*. *Exp*. *Ther*. **2016**, *358* (2), 294–305.

(27) Kim, M.; Laramy, J. K.; Mohammad, A. S.; Talele, S.; Fisher, J.; Sarkaria, J. N.; Elmquist, W. F. Brain Distribution of a Panel of Epidermal Growth Factor Receptor Inhibitors Using Cassette Dosing in Wild-Type and Abcb1/Abcg2-Deficient Mice. *Drug Metab*. *Dispos*. **2019**, *47* (4), 393–404.

(28) Orozco, C. C.; Atkinson, K.; Ryu, S.; Chang, G.; Keefer, C.; Lin, J.; Riccardi, K.; Mongillo, R. K.; Tess, D.; Filipski, K. J.; Kalgutkar, A. S.; Litchfield, J.; Scott, D.; Di, L. Structural Attributes Influencing Unbound Tissue Distribution. *Eur*. *J*. *Med*. *Chem*. **2020**, *185*, 111813.

(29) Bewley, B. R.; Spearing, P. K.; Weiner, R. L.; Luscombe, V. B.; Zhan, X.; Chang, S.; Cho, H. P.; Rodriguez, A. L.; Niswender, C. M.; Conn, P. J.; Bridges, T. M.; Engers, D. W.; Lindsley, C. W. Discovery of a Novel, CNS Penetrant M4 PAM Chemotype Based on a 6-Fluoro-4-(piperidin-1-Yl)quinoline-3-Carbonitrile Core. *Bioorg*. *Med*. *Chem*. *Lett*. **2017**, *27* (18), 4274–4279.

(30) Temple, K. J.; Engers, J. L.; Long, M. F.; Gregro, A. R.; Watson, K. J.; Chang, S.; Jenkins, M. T.; Luscombe, V. B.; Rodriguez, A. L.; Niswender, C. M.; Bridges, T. M.; Conn, P. J.; Engers, D. W.; Lindsley, C. W. Discovery of a Novel 3,4-Dimethylcinnoline Carboxamide M4 Positive Allosteric Modulator (PAM) Chemotype via Scaffold Hopping. *Bioorg*. *Med*. *Chem*. *Lett*. **2019**, *29* (21), 126678.

(31) Zeng, Q.; Wang, J.; Cheng, Z.; Chen, K.; Johnström, P.; Varnäs, K.; Li, D. Y.; Yang, Z. F.; Zhang, X. Discovery and Evaluation of Clinical Candidate AZD3759, a Potent, Oral Active, Central Nervous System-Penetrant, Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor. *J. Med. Chem.* **2015**, *58* (20), 8200–8215.

(32) Toledo-Sherman, L.; Breccia, P.; Cachope, R.; Bate, J. R.; Angulo-Herrera, I.; Wishart, G.; Matthews, K. L.; Martin, S. L.; Cox, H. C.; McAllister, G.; Penrose, S. D.; Vater, H.; Esmieu, W.; Van de Poël, A.; Van de Bospoort, R.; Strijbosch, A.; Lamers, M.; Leonard, P.; Jarvis, R. E.; Blackaby, W.; Barnes, K.; Eznarriaga, M.; Dowler, S.; Smith, G. D.; Fischer, D. F.; Lazari, O.; Yates, D.; Rose, M.; Jang, S.-W.; Muñoz-Sanjuan, I.; Dominguez, C. Optimization of Potent and Selective Ataxia Telangiectasia-Mutated Inhibitors Suitable for a Proof-of-Concept Study in Huntington's Disease Models. *J. Med. Chem.* **2019**, *62* (6), 2988–3008.

(33) Colclough, N.; Chen, K.; Johnström, P.; Strittmatter, N.; Yan, Y.; Wrigley, G. L.; Schou, M.; Goodwin, R.; Varnäs, K.; Adua, S. J.; Zhao, M.; Nguyen, D. X.; Maglennon, G.; Barton, P.; Atkinson, J.; Zhang, L.; Janefeldt, A.; Wilson, J.; Smith, A.; Takano, A.; Arakawa, R.; Kondrashov, M.; Malmquist, J.; Revunov, E.; Vazquez-Romero, A.; Moein, M. M.; Windhorst, A. D.; Karp, N. A.; Finlay, M. R. V.; Ward, R. A.; Yates, J. W. T.; Smith, P. D.; Farde, L.; Cheng, Z.; Cross, D. A. E. Preclinical Comparison of the Blood-Brain Barrier Permeability of Osimertinib with Other EGFR TKIs. *Clin. Cancer Res.* **2021**, *27* (1), 189–201.

(34) Fagerberg, J. H.; Karlsson, E.; Ulander, J.; Hanisch, G.; Bergström, C. A. S. Computational Prediction of Drug Solubility in Fasted Simulated and Aspirated Human Intestinal Fluid. *Pharm. Res.* **2015**, *32* (2), 578–589.

(35) Leung, S. S. F.; Mijalkovic, J.; Borrelli, K.; Jacobson, M. P. Testing Physical Models of Passive Membrane Permeation. *J. Chem. Inf. Model.* **2012**, *52* (6), 1621–1636.

(36) Rochat, B.; Baumann, P.; Audus, K. L. Transport Mechanisms for the Antidepressant Citalopram in Brain Microvessel Endothelium. *Brain Res.* **1999**, *831* (1-2), 229–236.

(37) Fischer, H.; Gottschlich, R.; Seelig, A. Blood-Brain Barrier Permeation: Molecular Parameters Governing Passive Diffusion. *J. Membr. Biol.* **1998**, *165* (3), 201–211.

(38) Fong, C. W. Permeability of the Blood–Brain Barrier: Molecular Mechanism of Transport of Drugs and Physiologically Important Compounds. *J. Membr. Biol.* **2015**, *248* (4), 651–669.

(39) Wager, T. T.; Chandrasekaran, R. Y.; Hou, X.; Troutman, M. D.; Verhoest, P. R.; Villalobos, A.; Will, Y. Defining Desirable Central Nervous System Drug Space through the Alignment of Molecular Properties, in Vitro ADME, and Safety Attributes. *ACS Chem. Neurosci.* **2010**, *1* (6), 420–434.

(40) Leung, S. S. F.; Sindhikara, D.; Jacobson, M. P. Simple Predictive Models of Passive Membrane Permeability Incorporating Size-Dependent Membrane-Water Partition. *J. Chem. Inf. Model.* **2016**, *56* (5), 924–929.

(41) Leswing, Z. K. S. E. *White paper: Benchmark study of DeepAutoQSAR, ChemProp, and DeepPurpose on the ADMET subset of the Therapeutic Data Commons*. Schrodinger. https://www.schrodinger.com/science-articles/benchmark-study-deepautoqsar-chemprop-and-deeppurpose-admet-subset-therapeutic-data.

(42) All compounds with Kp,uu were omitted from the training set.

(43) Bos, P. H.; Houang, E. M.; Ranalli, F.; Leffler, A. E.; Boyles, N. A.; Eyrich, V. A.; Luria, Y.; Katz, D.; Tang, H.; Abel, R.; Bhat, S. AutoDesigner, a DE Novo Design Algorithm for Rapidly Exploring Large Chemical Space for Lead Optimization: Application to the Design and Synthesis of D-Amino Acid Oxidase Inhibitors. *J. Chem. Inf. Model.* **2022**, *62* (8), 1905–1915.

(44) *Schrödinger Release 2022-3: LiveDesign, Schrödinger, LLC, New York, NY, 2022*. Schrodinger. https://www.schrodinger.com/products/livedesign/drug-discovery (accessed 2022-10-24).

(45) We have ongoing efforts that will be reported in due time.

(46) Gunaydin, H.; Weiss, M. M.; Sun, Y. De Novo Prediction of P-Glycoprotein-Mediated Efflux Liability for Druglike Compounds. *ACS Med. Chem. Lett.* **2013**, *4* (1), 108–112.

(47) Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A. Applications of Fluorine

in Medicinal Chemistry. *J. Med. Chem.* **2015**, *58* (21), 8315–8359.

(48) 1.6% and 1.8% for Fu of BTB for 22 and 23.

(49) *Schrödinger Release 2022-3: LigPrep, Schrödinger, LLC, New York, NY, 2021.*
https://www.schrodinger.com/products/ligprep (accessed 2022-10-24).

(50) *Schrödinger Release 2022-3: Maestro, Schrödinger, LLC, New York, NY, 2022.*
https://www.schrodinger.com/products/maestro (accessed 2022-10-24).

(51) 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**, *17* (7),
4291–4300.

(52) *Schrödinger Release 2022-3: MacroModel, Schrödinger, LLC, New York, NY, 2021.*
https://www.schrodinger.com/products/macromodel (accessed 2022-10-24).

(53) *Schrödinger Release 2022-3: Jaguar, Schrödinger, LLC, New York, NY, 2021.*
https://www.schrodinger.com/products/jaguar (accessed 2022-10-24).

(54) Tang, H.; Jensen, K.; Houang, E.; McRobb, F. M.; Bhat, S.; Svensson, M.; Bochevarov, A.; Day, T.;
Dahlgren, M. K.; Bell, J. A.; Frye, L.; Skene, R. J.; Lewis, J. H.; Osborne, J. D.; Tierney, J. P.;
Gordon, J. A.; Palomero, M. A.; Gallati, C.; Chapman, R. S. L.; Jones, D. R.; Hirst, K. L.; Sephton,
M.; Chauhan, A.; Sharpe, A.; Tardia, P.; Dechaux, E. A.; Taylor, A.; Waddell, R. D.; Valentine, A.;
Janssens, H. B.; Aziz, O.; Bloomfield, D. E.; Ladha, S.; Fraser, I. J.; Ellard, J. M. Discovery of a
Novel Class of D-Amino Acid Oxidase Inhibitors Using the Schrödinger Computational Platform. *J.
Med. Chem.* **2022**, *65* (9), 6775–6802.

(55) *RDKit.* http://www.rdkit.org/.

(56) *Schrödinger Release 2021-3: Epik, Schrödinger, LLC, New York, NY, 2021.*
https://www.schrodinger.com/products/epik (accessed 2022-10-24).

(57) Johnson, S. A. *Probabilistic Multi-Parameter Optimization (pMPO).*
https://github.com/Merck/pmpo (accessed 2022-10-24).

(58) Sato, S.; Matsumiya, K.; Tohyama, K.; Kosugi, Y. Translational CNS Steady-State Drug Disposition
Model in Rats, Monkeys, and Humans for Quantitative Prediction of Brain-to-Plasma and
Cerebrospinal Fluid-to-Plasma Unbound Concentration Ratios. *AAPS J.* **2021**, *23* (4), 81.

SUPPORTING INFORMATION

A Computational Physics-based Approach to Predict Unbound Brain-to-Plasma Partition Coefficient, Kp,uu

Morgan Lawrenz,*[a] Mats Svensson,[b] Mitsunori Kato,[b] Karen Dingley,[b] Jackson Chief Elk,[c] Zhe Nie,[a] Yefen Zou,[a] Zachary Kaplan,[b] H. Rachel Lagiakos,[b] Hideyuki Igawa,[b] and Eric Therrien*[b]

[a]Schrödinger Inc., San Diego, California, 92122, United States
[b]Schrödinger Inc., New York, New York, 10036, United States
[c]Schrödinger Inc., Portland, Oregon, 97204, United States

Table of contents

1. Supporting Figures and Tables

Figure S1. Categorical accuracy and confusion matrices for E-sol and other calculated and measured properties for prediction of Kp,uu. Categorical accuracy was calculated from the confusion matrices as (TP+TN)/N and only reported if p=value < 0.05, Two example matrices at the bottom of the figure label quadrants for positively and negatively correlated matrices. See **Table S1** for AUC scores from threshold analysis.

A Accuracy = 79%

	E-sol	
Kp,uu =0.3	0.10	0.33
	0.46	0.11

B Accuracy = 55%

	Pfizer MPO	
Kp,uu =0.3	0.14	0.29
	0.26	0.31

C Accuracy = 63%

	Merck MPO	
Kp,uu =0.3	0.14	0.29
	0.33	0.24

D Accuracy = 63%

	Efflux	
Kp,uu =0.2	0.45	0.15
	0.22	0.18

E Accuracy = 68%

	PSA	
Kp,uu =0.3	0.31	0.12
	0.20	0.37

F Accuracy = 68%

	RRCK	
Kp,uu =0.3	0.16	0.27
	0.41	0.16

G Accuracy = 61%

	MW	
Kp,uu =0.3	0.27	0.16
	0.24	0.33

H Accuracy = 62%

	eLogD	
Kp,uu =0.2	0.19	0.27
	0.35	0.20

	model1	
exp	FN	TP
	TN	FP

	model2	
exp	TP	FN
	FP	TN

Figure S2. Correlation plot showing machine learning model fit on calculated E-sol (kcal/mol) on the y-axis and the full fidelity QM calculations for compounds from internal projects and published datasets (n = 241) on the x-axis (MUE = 1.80 kcal/mol).
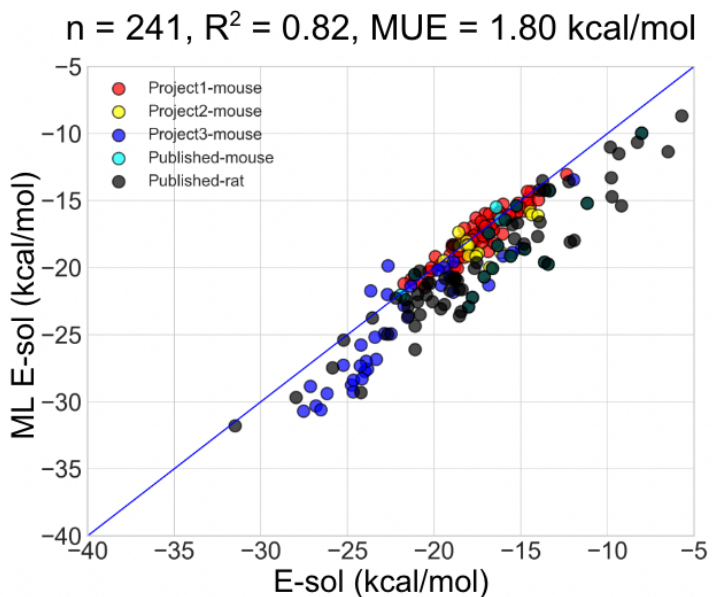
Figure S3. Correlation of plasma protein binding and brain tissue binding for analyzed data.
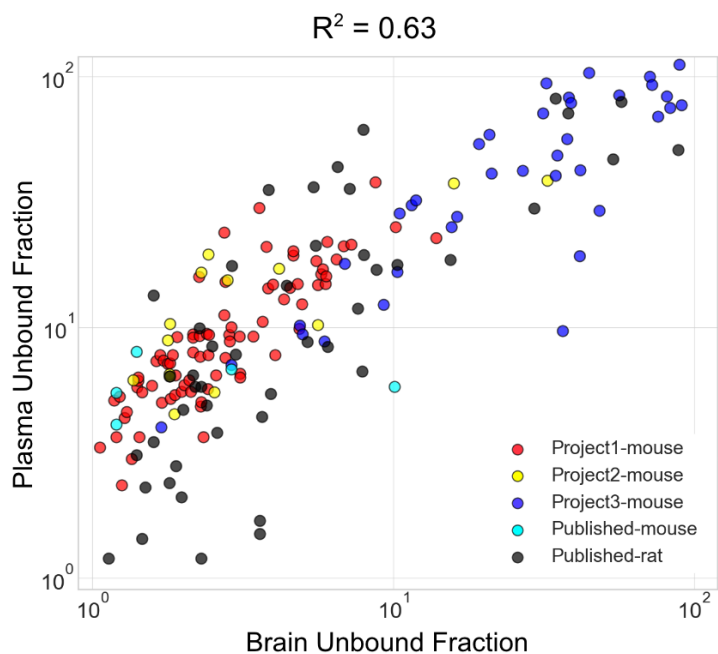
Figure S4. Basic calculated properties compared for Schrödinger project compounds (A) and published compounds (B) analyzed in this work. These two groups of compounds are also shown in t-distributed stochastic neighbor embedding (t-SNE) projections to create a graphical representation of the chemical space (C).
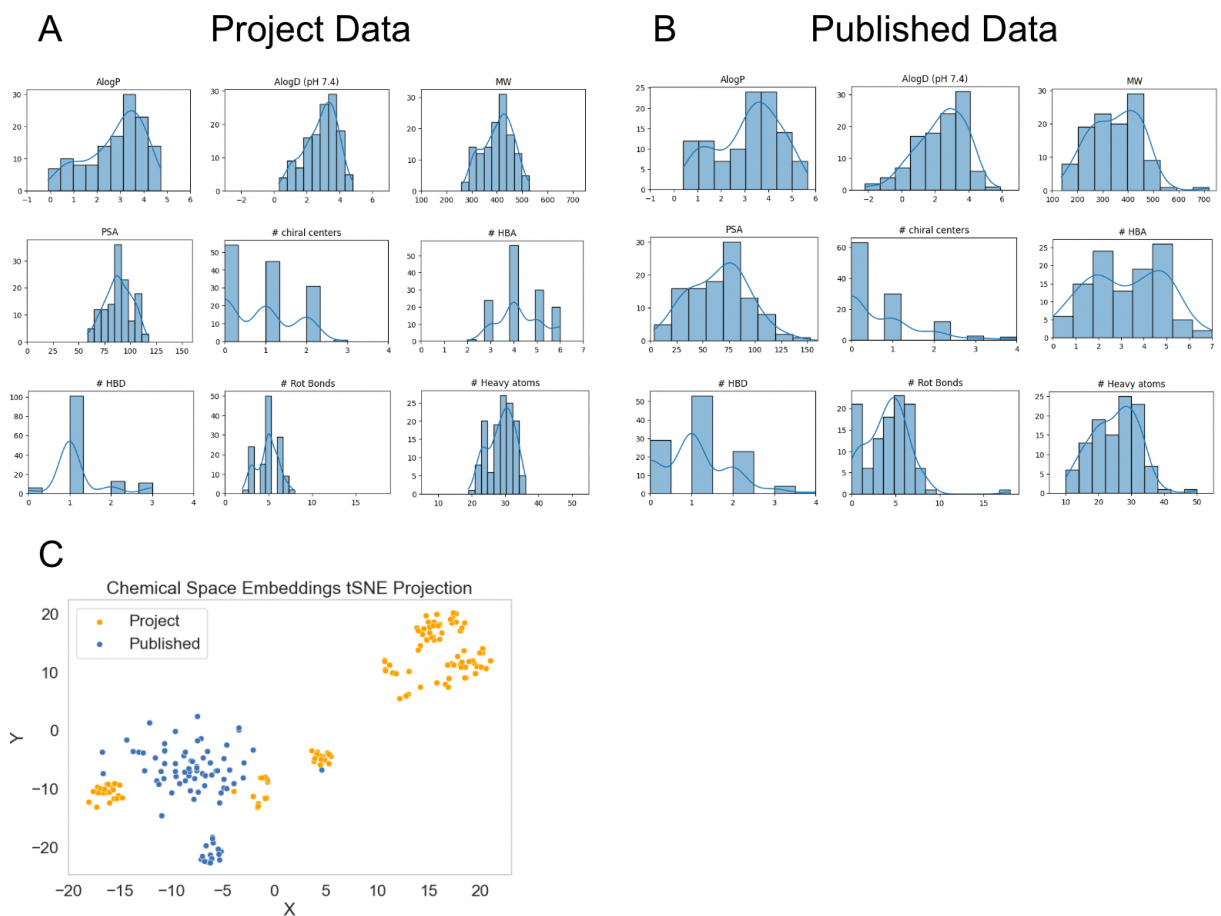
Figure S5. Correlation between quick and extensive sampling methods for E-sol (A) Kp,uu data-points plotted in log-scale (y-axis) with E-sol (kcal/mol) for compounds from internal projects and published datasets (n = 241) on the x-axis for quick (B) and extensive (B) sampling methods.
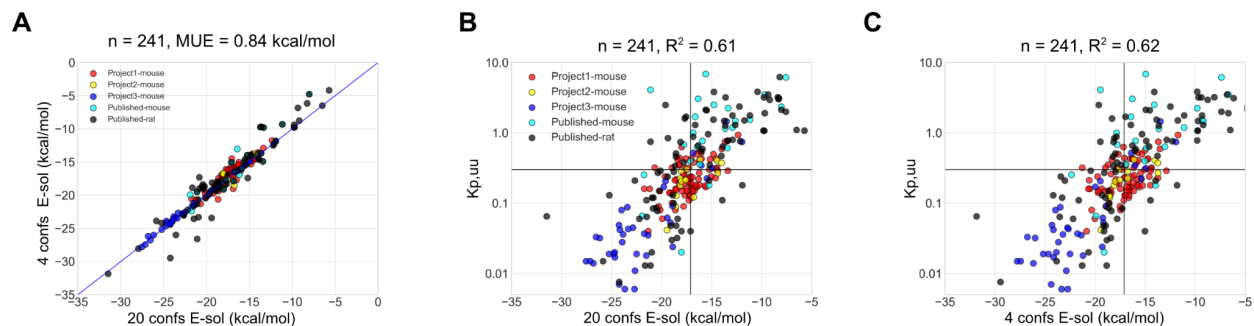
Figure S6. Plot showing the difference of E-sol (kcal/mol) calculated between the quick and extensive sampling methods for the public and internal data sets sorted by the number of rotatable bonds (compounds with 0 rotatable bond not shown).

38
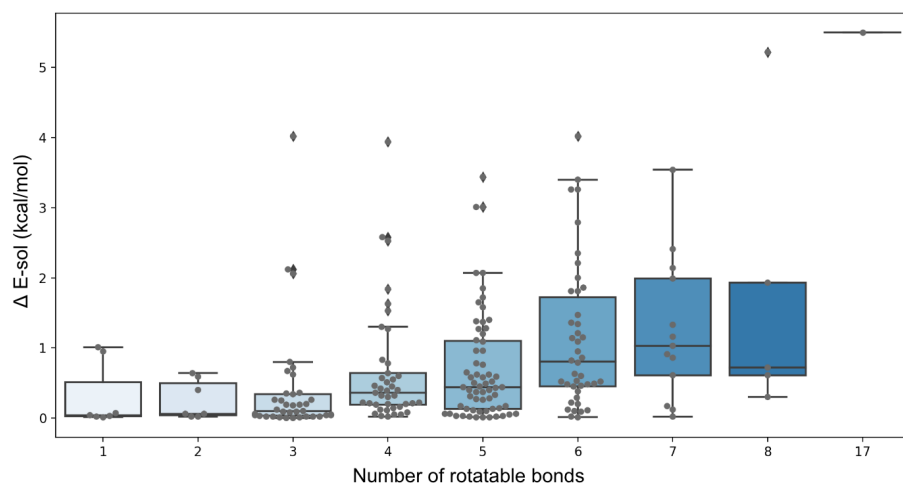
Table S1. AUC of ROC plots evaluated for the metrics in this study, with two Kp,uu thresholds used for classification.

Metric	Kp,uu = 0.3 AUC	Kp,uu = 0.2 AUC
E-sol Hydration	0.88	0.87
Merck pMPO	0.67	0.58
Pfizer MPO	0.55	0.48
RRCK	0.72	0.67
PSA	0.27	0.38
MW	0.33	0.42
LogD	0.60	0.69
MDCK-MDR1 ER	0.43	0.37

2. HPLC chromatograms

Empower 3 SOFTWARE

Compound 1

SAMPLE INFORMATION

Sample Name:	NBK0147-113-P1-A	Acq. Method Set:	202007
Vial:	1:A,7	Sample Set Name:	10min 20 Normal
Injection Volume:	0.70 ul	Channel Name:	280.0nm, 254.0nm, 214.0nm
Run Time:	10.0 Minutes	Date Acquired:	7/15/2020 11:24:27 AM CST

Stacked Chromatograms

Processed Channel Descr. PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm
Processed Channel Descr. PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm
Processed Channel Descr. PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm
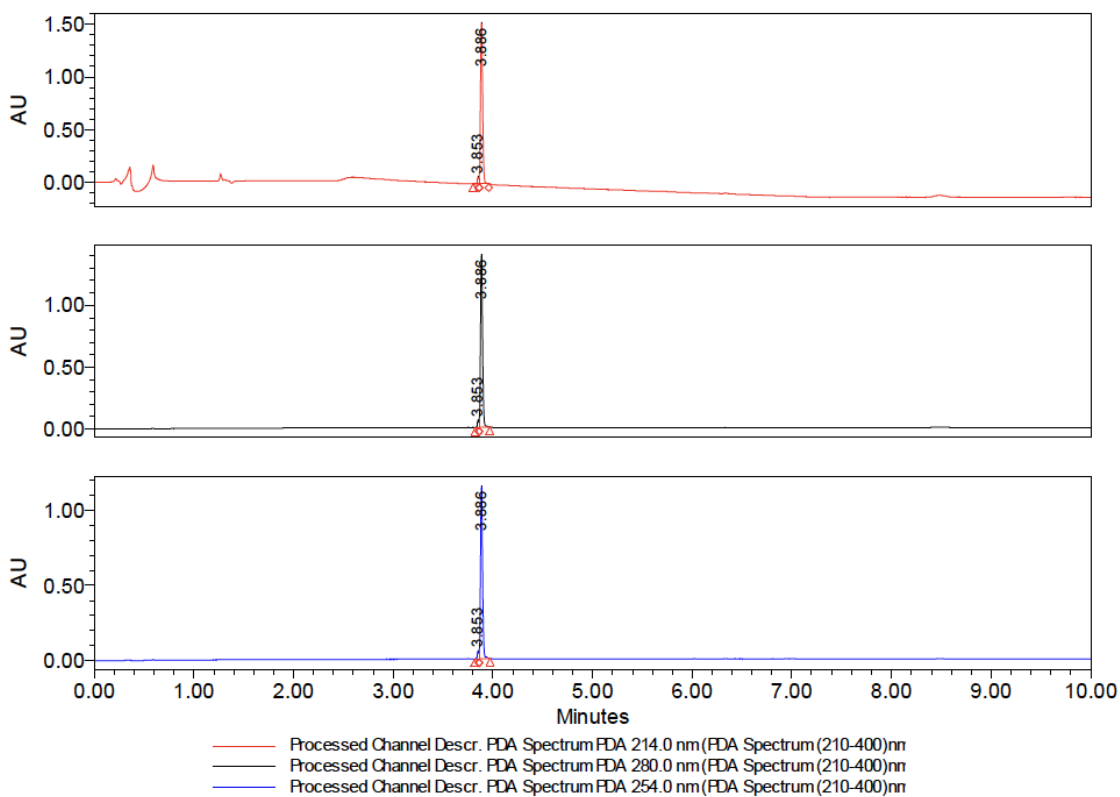
Reported by User: System
Report Method: 1
Report Method ID: 1889
Page: 1 of 2

Project Name: Installation\2020-05\2020-07
Date Printed:
7/21/2020
9:39:59 AM PRC

Processed Channel Descr.: PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)	3.853	72892	3.16	67982
2	PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)	3.886	2236076	96.84	1534123

Processed Channel Descr.: PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)	3.853	54089	3.27	52775
2	PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)	3.886	1601087	96.73	1155069

Processed Channel Descr.: PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)	3.853	60448	3.00	58927
2	PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)	3.886	1952641	97.00	1404384

Reported by User: System
Report Method: 1
Report Method ID: 1889
Page: 2 of 2

Project Name: Installation\2020-05\2020-07
Date Printed:
7/21/2020
9:39:59 AM PRC

41

S A M P L E   I N F O R M A T I O N

Sample Name:	NBK0147-139-P1-A	Acq. Method Set:	202007
Vial:	1:C,5	Sample Set Name:	10min 20 Normal
Injection Volume:	0.70 ul	Channel Name:	280.0nm, 254.0nm, 214.0nm
Run Time:	10.0 Minutes	Date Acquired:	8/11/2020 1:39:32 PM CST

Stacked Chromatograms

Processed Channel Descr. PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm
Processed Channel Descr. PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm
Processed Channel Descr. PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm

Reported by User: System
Report Method: 1
Report Method ID: 2256
Page: 1 of 2

Project Name: Installation\2020-05\2020-08
Date Printed:
8/12/2020
2:14:36 PM PRC

42

Processed Channel Descr.: PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)	3.866	1616543	100.00	1184959

Processed Channel Descr.: PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)	3.730	20456	1.83	17493
2	PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)	3.866	1096868	98.17	818662

Processed Channel Descr.: PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)	3.731	8584	0.64	8518
2	PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)	3.866	1331147	99.36	993103

Reported by User: System
Report Method: 1
Report Method ID: 2256
Page: 2 of 2

Project Name: Installation\2020-05\2020-08
Date Printed:
8/12/2020
2:14:36 PM PRC

43

S A M P L E   I N F O R M A T I O N

Sample Name:	NBK0149-296-P1-A	Acq. Method Set:	202011
Vial:	1:A,1	Sample Set Name:	10min 20 Normal
Injection Volume:	0.60 ul	Channel Name:	280.0nm, 254.0nm@1, 214.0nm
Run Time:	10.0 Minutes	Date Acquired:	11/11/2020 9:07:51 AM CST

Stacked Chromatograms

Processed Channel Descr. PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm
Processed Channel Descr. PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm
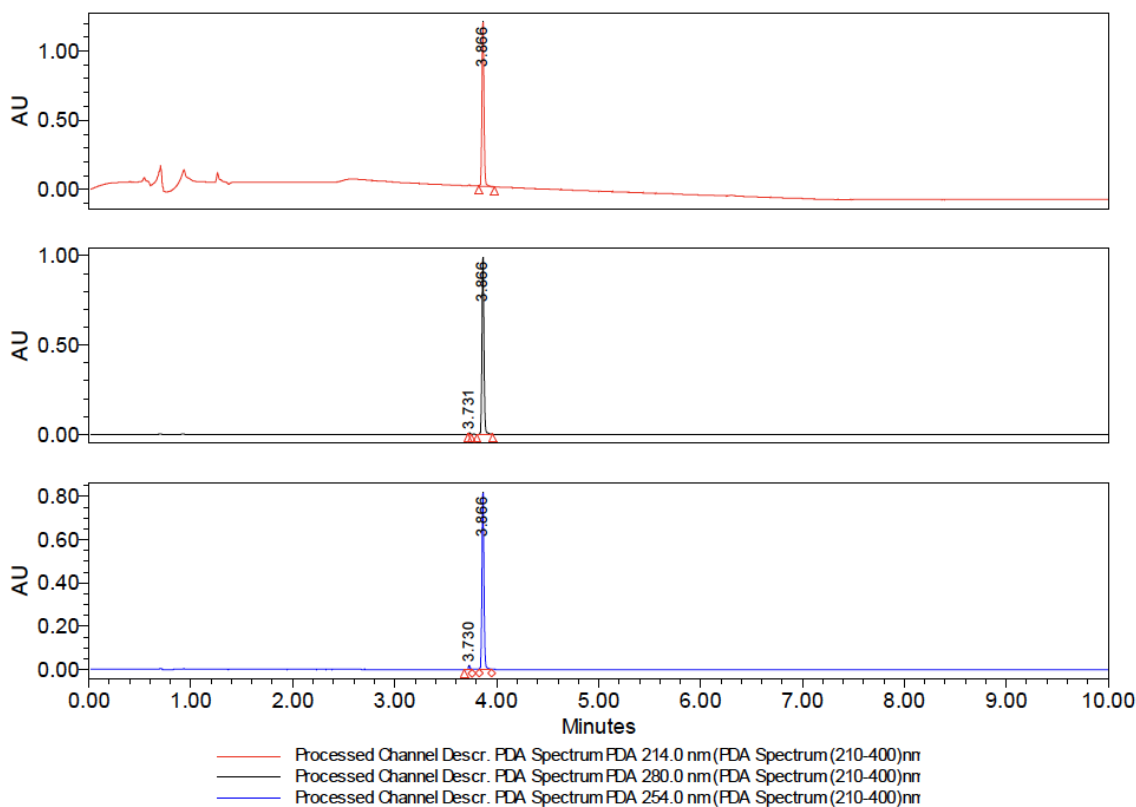Processed Channel Descr. PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm
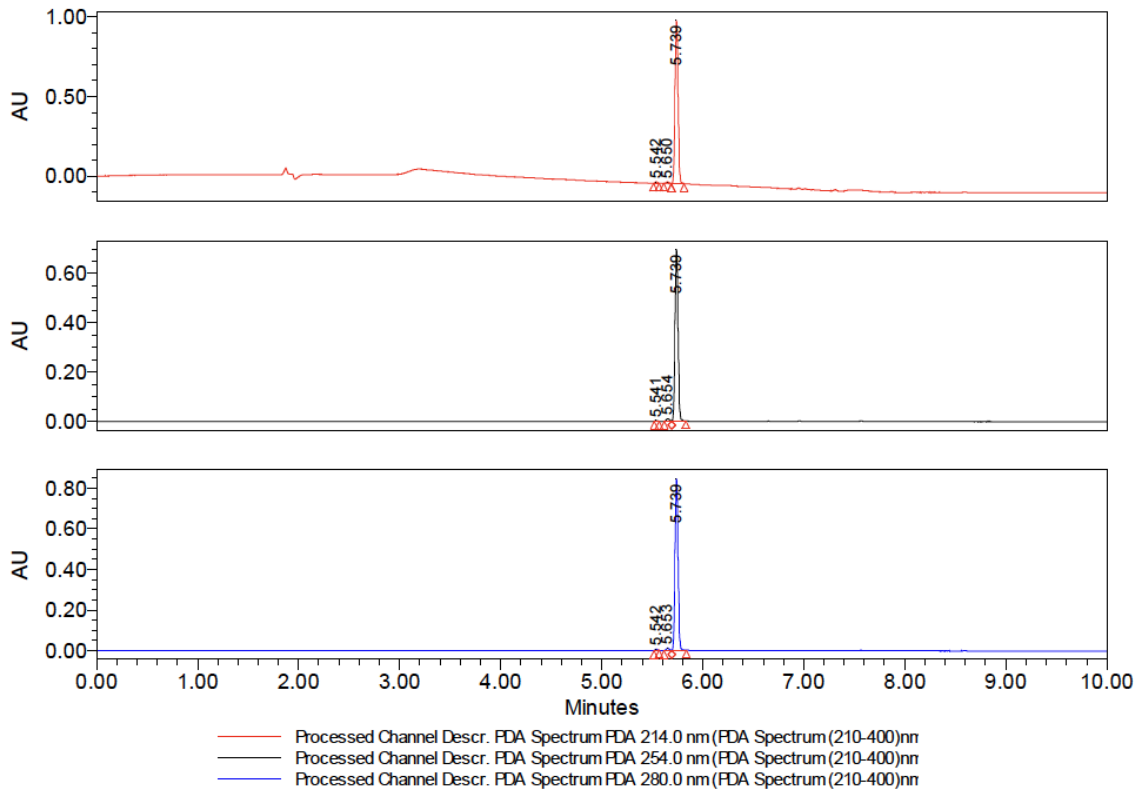
Reported by User: System
Report Method: 1
Report Method ID: 2211
Page: 1 of 2

Project Name: Installation\2020-05\2020-11
Date Printed:
11/11/2020
9:28:00 AM PRC

44

Processed Channel Descr.: PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)	5.542	14339	0.65	8386
2	PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)	5.650	23025	1.04	11169
3	PDA Spectrum PDA 214.0 nm (PDA Spectrum (210-400)nm)	5.739	2166855	98.30	1026214

Processed Channel Descr.: PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)	5.541	4246	0.29	2726
2	PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)	5.654	23079	1.55	10428
3	PDA Spectrum PDA 254.0 nm (PDA Spectrum (210-400)nm)	5.739	1461965	98.17	696726

Processed Channel Descr.: PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)

	Processed Channel Descr.	Retention Time (min)	Area	% Area	Height
1	PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)	5.542	10619	0.59	6424
2	PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)	5.653	23584	1.30	10973
3	PDA Spectrum PDA 280.0 nm (PDA Spectrum (210-400)nm)	5.739	1775624	98.11	846303

Reported by User: System
Report Method: 1
Report Method ID: 2211
Page: 2 of 2

Project Name: Installation\2020-05\2020-11
Date Printed:
11/11/2020
9:28:00 AM PRC

45

Compound 4

Empower™ 3
SOFTWARE

SAMPLE INFORMATION

Sample Name:	NBK0337-105-P1-A	Acquired By:	System
Sample Type:	Unknown	Sample Set Name:	20210621
Vial:	2:D,4	Acq. Method Set:	viva_Acid waters BEH_15min
Injection #:	1	Processing Method:	PM
Injection Volume:	0.20 ul	Channel Name:	PDA Ch1 214nm@4.8nm, PDA Ch2
Run Time:	15.0 Minutes	Proc. Chnl. Descr.:	PDA Ch1 214nm@4.8nm, PDA Ch2

Method Info:
Mobile Phase: A:Water(0.05%FA)   B: ACN(0.05%FA)
Gradient: 5%B for 0.5min,10%B increase to 95%B within 8.5min,95%B for 3.0min
back to 5%B within 1.5min,5%B for 2.0min
Flow Rate: 0.3mL/min   Column Temperature: 30 C
Column: ACQUITY UPLC BEH C18 1.7um,2.1*150mm

Date Acquired: 6/22/2021 1:39:27 PM CST
Date Processed: 6/22/2021 1:56:32 PM CST, 6/22/2021 1:56:36 PM CST, 6/22/2021 1:56:45 PM CST
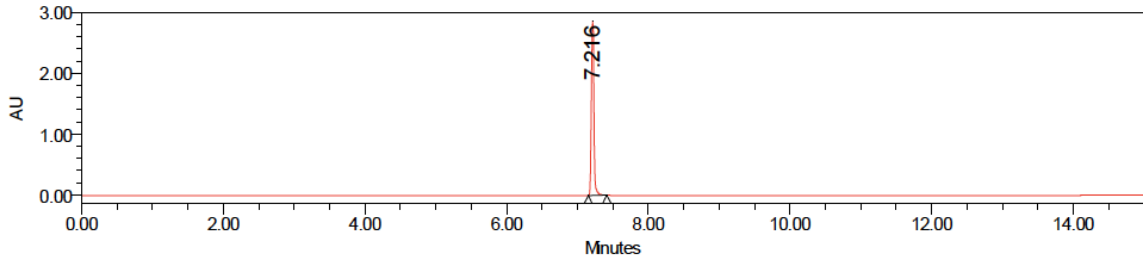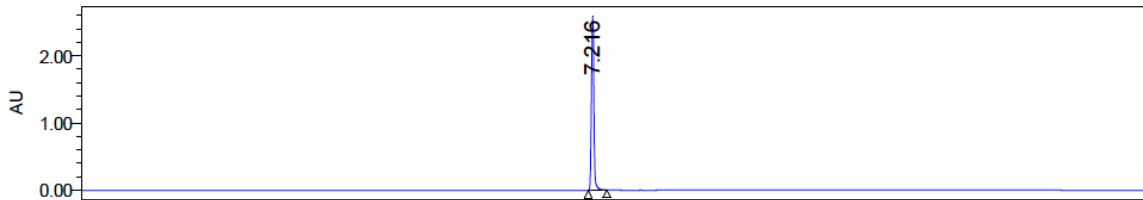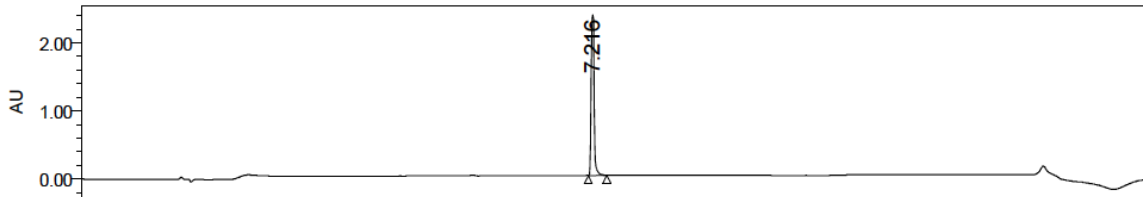
_____ Channel: PDA Ch1 214nm@4.8nm; Processed Channel: PDA Ch1 214nm@4.8nm; Result Id: 4165;
Processing Method: PM
_____ Channel: PDA Ch2 254nm@4.8nm; Processed Channel: PDA Ch2 254nm@4.8nm; Result Id: 4164;
Processing Method: PM
_____ Channel: PDA Ch3 280nm@4.8nm; Processed Channel: PDA Ch3 280nm@4.8nm; Result Id: 4163;
Processing Method: PM

Reported by User: System   Project Name: 2021-06
Report Method: VIva Title update report   Date Printed:
Report Method ID:2795   6/22/2021
Page: 1 of 2   1:56:57 PM PRC

46

Processed Channel Descr.: PDA Ch1 214nm@4.8nm

	Processed Channel Descr.	RT	Height	Area	% Area
1	PDA Ch1 214nm@4.8nm	7.216	2358333	6925580	100.00

Processed Channel Descr.: PDA Ch2 254nm@4.8nm

	Processed Channel Descr.	RT	Height	Area	% Area
1	PDA Ch2 254nm@4.8nm	7.216	2595639	6253541	100.00

Processed Channel Descr.: PDA Ch3 280nm@4.8nm

	Processed Channel Descr.	RT	Height	Area	% Area
1	PDA Ch3 280nm@4.8nm	7.216	2861569	7249959	100.00

Reported by User: System
Report Method: Vlva Title update report
Report Method ID: 2795
Page: 2 of 2

Project Name: 2021-06
Date Printed:
6/22/2021
1:56:57 PM PRC

47

Compound 5

Empower™ 3 SOFTWARE

S A M P L E   I N F O R M A T I O N

Sample Name:	NBK0265-137-P1-A	Acquired By:	System
Sample Type:	Unknown	Sample Set Name:	20210714
Vial:	2:C,7	Acq. Method Set:	viva_Acid waters BEH_15min
Injection #:	1	Processing Method:	PM
Injection Volume:	0.50 ul	Channel Name:	PDA Ch1 214nm@4.8nm, PDA Ch2
Run Time:	15.0 Minutes	Proc. Chnl. Descr.:	PDA Ch1 214nm@4.8nm, PDA Ch2

Method Info: Mobile Phase: A:Water(0.02%FA) B: ACN(0.02%FA)
Gradient: 5%B for 0.5min,10%B increase to 95%B within 8.5min,95%B for 2.5min
back to 5%B within 1.5min,5%B for 2.0min
Flow Rate: 0.3mL/min Column Temperature: 30 C
Column: ACQUITY UPLC BEH C18 1.7um,2.1*150mm

Date Acquired: 7/14/2021 12:51:20 PM CST
Date Processed: 7/14/2021 1:18:02 PM CST, 7/14/2021 1:18:13 PM CST, 7/14/2021 1:18:37 PM CST
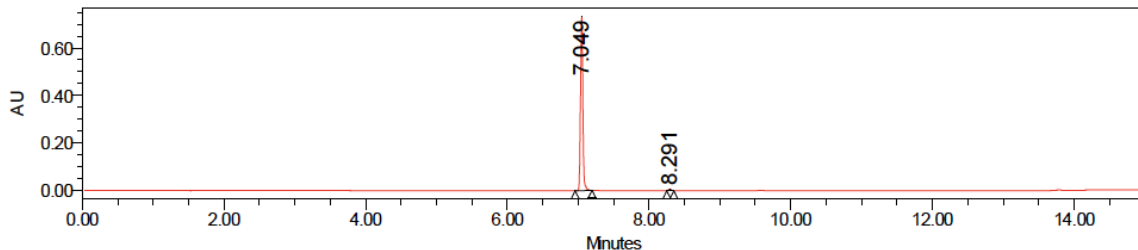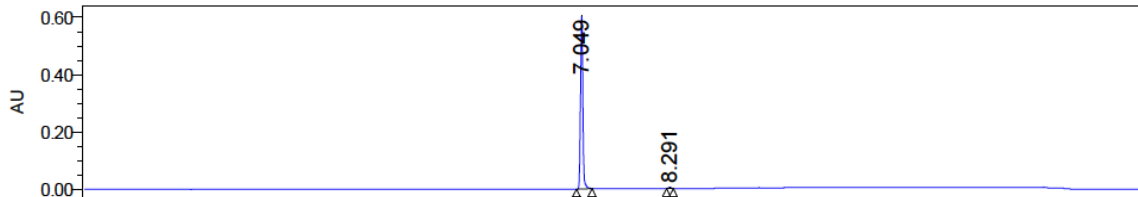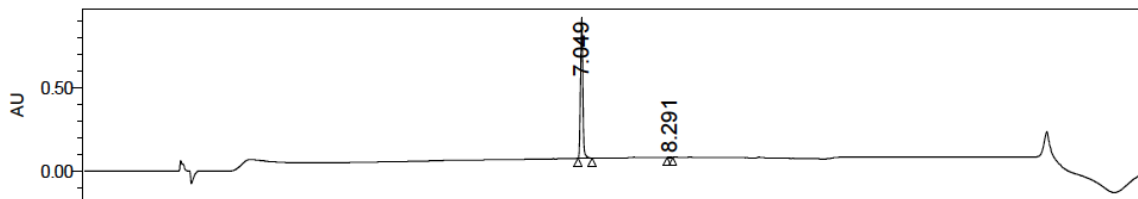
Channel: PDA Ch1 214nm@4.8nm; Processed Channel: PDA Ch1 214nm@4.8nm; Result Id: 3059;
Processing Method: PM
Channel: PDA Ch2 254nm@4.8nm; Processed Channel: PDA Ch2 254nm@4.8nm; Result Id: 3058;
Processing Method: PM
Channel: PDA Ch3 280nm@4.8nm; Processed Channel: PDA Ch3 280nm@4.8nm; Result Id: 3057;
Processing Method: PM

Reported by User: System
Report Method: Title update report
Report Method ID:1014
Page: 1 of 2

Project Name: 2021-07
Date Printed:
7/14/2021
1:19:02 PM PRC

48

Processed Channel Descr.: PDA Ch1 214nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch1 214nm@4.8nm	7.049	1942433	99.54	843061
2	PDA Ch1 214nm@4.8nm	8.291	8908	0.46	4105

Processed Channel Descr.: PDA Ch2 254nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch2 254nm@4.8nm	7.049	1400103	99.34	607281
2	PDA Ch2 254nm@4.8nm	8.291	9335	0.66	4067

Processed Channel Descr.: PDA Ch3 280nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch3 280nm@4.8nm	7.049	1698622	99.20	736584
2	PDA Ch3 280nm@4.8nm	8.291	13721	0.80	5900

Error Log

SI Plot with Legend group contains information that doesn't match the data being reported.

Reported by User: System

Report Method: Title update report

Report Method ID: 1014

Page: 2 of 2

Project Name: 2021-07

Date Printed:

7/14/2021

1:19:02 PM PRC

49

Compound 6

```
=====================================================================
Acq. Operator   : admin
Sample Operator : admin
Acq. Instrument : 1260                      Location :   P1-A-03
Injection Date  : 8/2/2021 10:19:24 AM           Inj :   1
                                       Inj Volume : 5.000 µl
Different Inj Volume from Sample Entry!  Actual Inj Volume : 0.700 µl
Acq. Method     : C:\Users\Public\Documents\ChemStation\1\Methods\Base-15min.M
Last changed    : 8/2/2021 10:18:42 AM by admin
                  (modified after loading)
Analysis Method : C:\USERS\PUBLIC\DOCUMENTS\CHEMSTATION\1\METHODS\Base-15min.M
Last changed    : 7/22/2021 9:09:34 AM by admin
                  (modified after loading)
Method Info     : Mobile:  A: Water(10mM NH4HCO3)  B: ACN      Flow Rate:1.0mL/min
                  Column:XBridge BEH C18  2.5um  4.6mm*150mm   Column Temperature:30 oC
                  Gradient:10%B for 0.5min, increase to 95%B within 9.0min, 95%B for 2.0min,
                  back to 5%B within 1.0min, 10%B for 2.5min

Sample Info     : Walkup method: 'Base-15min.M'
                  Target:

Additional Info : Peak(s) manually integrated
```
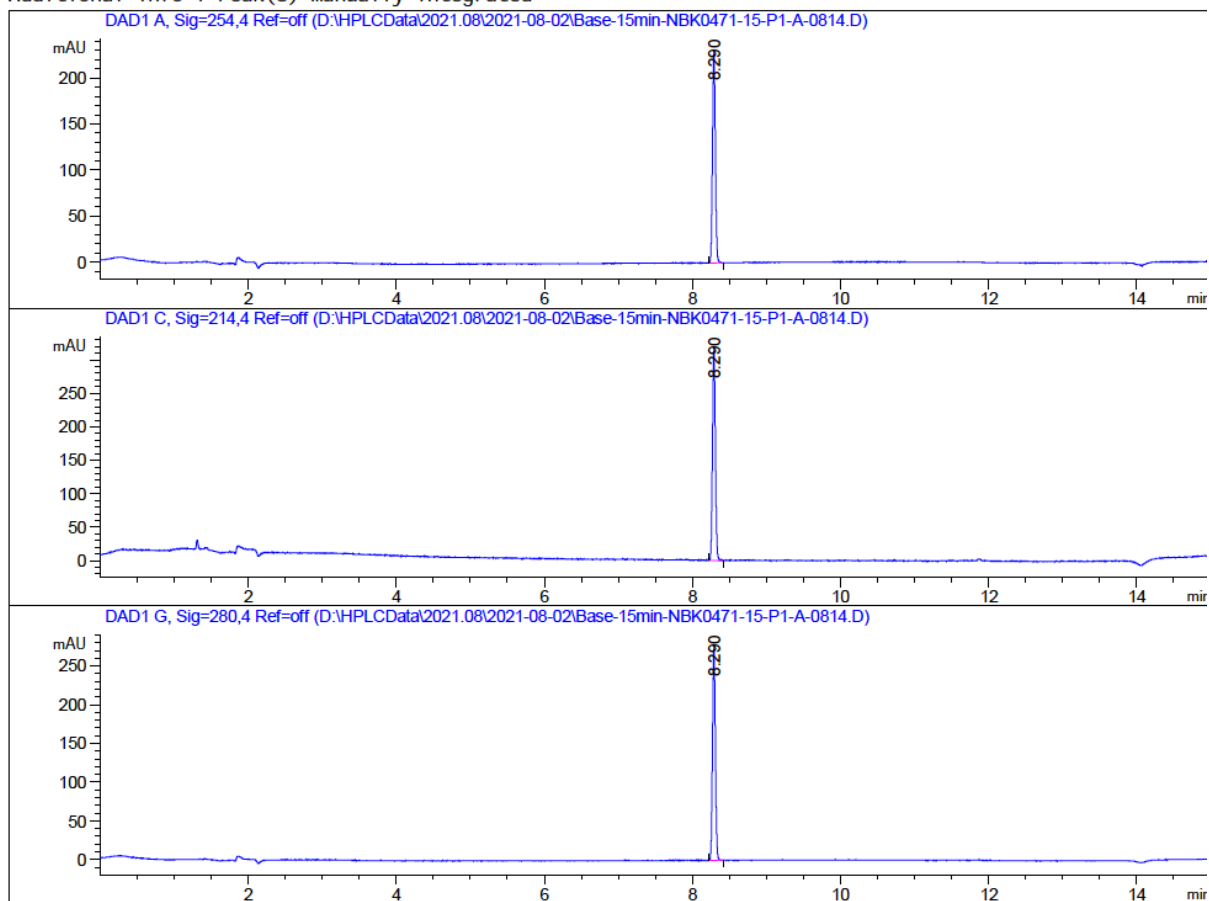
```
========================================================================
                        Area Percent Report
========================================================================

Sorted By           :        Signal
Multiplier          :        1.0000
Dilution            :        1.0000
Do not use Multiplier & Dilution Factor with ISTDs


Signal 1: DAD1 A, Sig=254,4 Ref=off

Peak RetTime Type  Width     Area        Height      Area
  #   [min]        [min]   [mAU*s]       [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   8.290 BB   0.0415  624.73181  231.05818 100.0000

Totals :                 624.73181  231.05818




Signal 2: DAD1 C, Sig=214,4 Ref=off

Peak RetTime Type  Width     Area        Height      Area
  #   [min]        [min]   [mAU*s]       [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   8.290 BB   0.0419  869.93011  318.03067 100.0000

Totals :                 869.93011  318.03067




Signal 3: DAD1 G, Sig=280,4 Ref=off

Peak RetTime Type  Width     Area        Height      Area
  #   [min]        [min]   [mAU*s]       [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   8.290 BB   0.0421  749.70026  277.09290 100.0000

Totals :                 749.70026  277.09290



========================================================================
                        *** End of Report ***
```

Empower™ 3
SOFTWARE

SAMPLE INFORMATION

Sample Name:	NBK0022-196-P1-A	Acquired By:	System
Sample Type:	Standard	Sample Set Name:	20200318
Vial:	1:B,1	Acq. Method Set:	viva_Acid waters BEH_15min
Injection #:	1	Processing Method:	VIVA standards Method_214nm,
Injection Volume:	1.00 ul	Channel Name:	PDA Ch3 214nm@4.8nm, PDA Ch2
Run Time:	15.0 Minutes	Proc. Chnl. Descr.:	PDA Ch3 214nm@4.8nm, PDA Ch2

Date Acquired: 3/18/2020 11:45:36 AM CST
Date Processed: 3/18/2020 12:53:34 PM CST, 3/18/2020 12:53:45 PM CST, 3/18/2020 12:54:00 PM CST
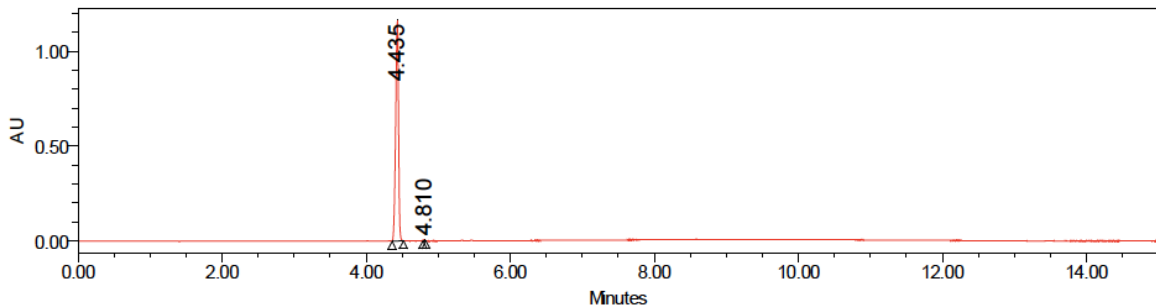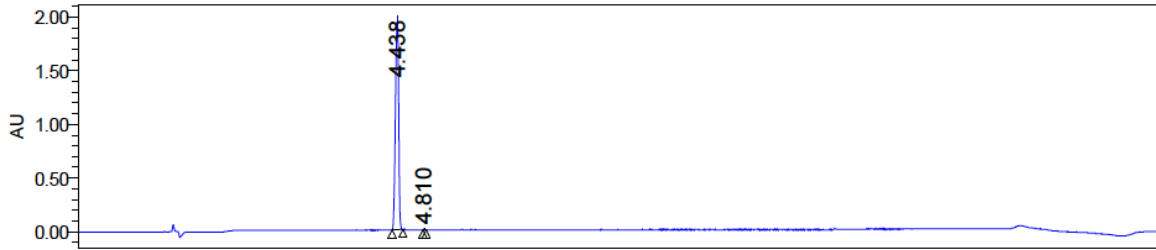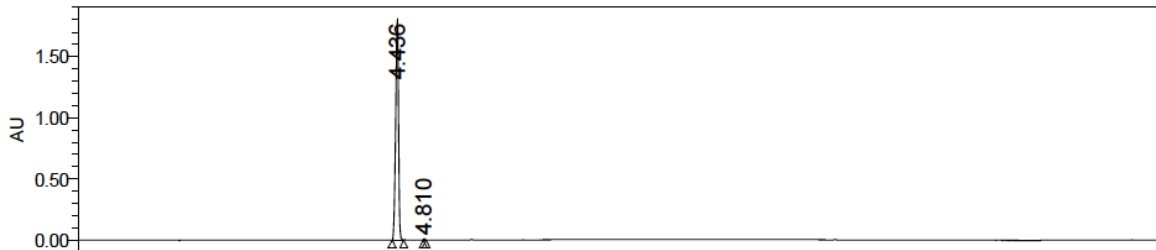
_____ Channel: PDA Ch2 280nm@4.8nm; Processed Channel: PDA Ch2 280nm@4.8nm; Result Id: 12241;
Processing Method: VIVA standards Method 280
_____ Channel: PDA Ch3 214nm@4.8nm; Processed Channel: PDA Ch3 214nm@4.8nm; Result Id: 12243;
Processing Method: VIVA standards Method_214nm
_____ Channel: PDA Ch1 254nm@4.8nm; Processed Channel: PDA Ch1 254nm@4.8nm; Result Id: 12242;
Processing Method: VIVA standards Method 254

Reported by User: System
Report Method: Injection Summary Report1
Report Method ID: 5922
Page: 1 of 2

Project Name: VIVA_2020.01month
Date Printed:
3/18/2020
12:54:18 PM PRC

52

Processed Channel Descr.: PDA Ch1 254nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch1 254nm@4.8nm	4.435	3199835	99.74	1161456
2	PDA Ch1 254nm@4.8nm	4.810	8277	0.26	6387

Processed Channel Descr.: PDA Ch2 280nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch2 280nm@4.8nm	4.436	5000624	99.76	1805959
2	PDA Ch2 280nm@4.8nm	4.810	11846	0.24	9262

Processed Channel Descr.: PDA Ch3 214nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch3 214nm@4.8nm	4.438	6134056	99.78	1987673
2	PDA Ch3 214nm@4.8nm	4.810	13309	0.22	10281

Reported by User: System
Report Method: Injection Summary Report1
Report Method ID: 5922
Page: 2 of 2

Project Name: VIVA_2020.01month
Date Printed:
3/18/2020
12:54:18 PM PRC

53

Compound 8

Injection Summary Report

S A M P L E I N F O R M A T I O N	

Sample Name:	V1503-059-P1-A	Acquired By:	System
Sample Type:	Standard	Sample Set Name	20190515
Vial:	1:B,5	Acq. Method Set:	VIVA QC_WATERS BEH C18
Injection #:	1	Processing Method	VIVA01
Injection Volume:	0.50 ul	Channel Name:	PDA Ch3 214nm@4.8nm, PDA
Run Time:	15.0 Minutes	Proc. Chnl. Descr.:	PDA Ch3 214nm@4.8nm, PDA
Date Acquired:	5/15/2019 1:26:59 PM CST		
Date Processed:	5/15/2019 1:42:55 PM CST, 5/15/2019 1:43:06 PM CST, 5/15/2019 1:43:16 PM CST		
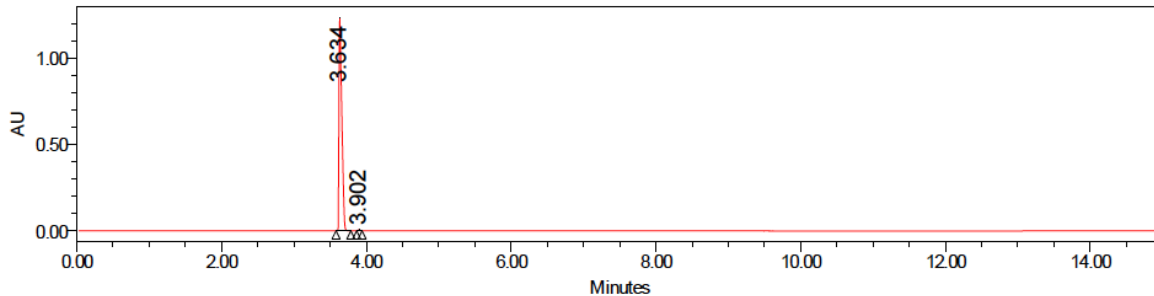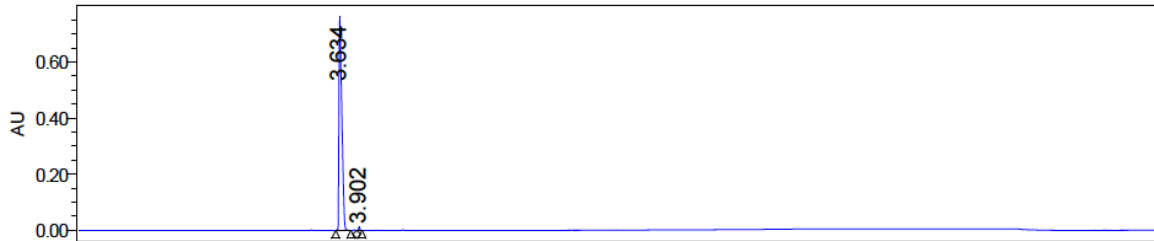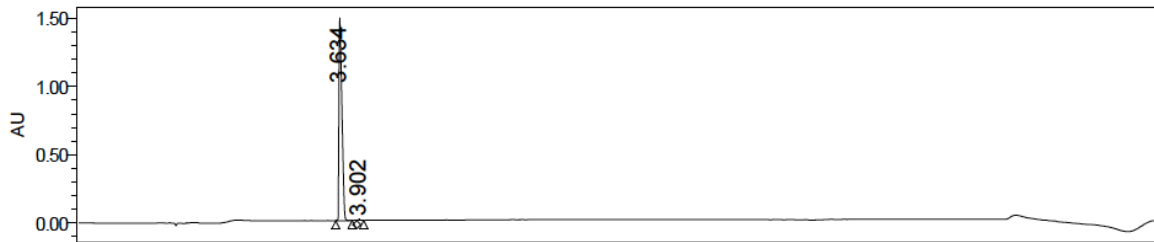
Channel: PDA Ch3 214nm@4.8nm;   Processed Channel: PDA Ch3 214nm@4.8nm;   Result Id: 6258;
Processing Method: VIVA01
Channel: PDA Ch1 254nm@4.8nm;   Processed Channel: PDA Ch1 254nm@4.8nm;   Result Id: 6260;
Processing Method: VIVA01
Channel: PDA Ch2 280nm@4.8nm;   Processed Channel: PDA Ch2 280nm@4.8nm;   Result Id: 6259;
Processing Method: VIVA01

Reported by User: System
Report Method: Injection Summary Repoi
Report Method II 1368
Page: 1 of 2

Project Name: 2019.04month
Date Printed:
5/15/2019
1:44:27 PM PRC

54

Processed Channel Descr.: PDA Ch1 254nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch1 254nm@4.8nm	3.634	2107527	99.35	761652
2	PDA Ch1 254nm@4.8nm	3.902	13737	0.65	11616

Processed Channel Descr.: PDA Ch2 280nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch2 280nm@4.8nm	3.634	3409105	99.67	1234923
2	PDA Ch2 280nm@4.8nm	3.902	11219	0.33	9054

Processed Channel Descr.: PDA Ch3 214nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch3 214nm@4.8nm	3.634	4264655	99.70	1485576
2	PDA Ch3 214nm@4.8nm	3.902	12663	0.30	8727

Reported by User: System
Report Method: Injection Summary Repor
Report Method II 1368
Page: 2 of 2

Project Name: 2019.04month
Date Printed:
5/15/2019
1:44:27 PM PRC

55

Compound 9

SAMPLE INFORMATION		
Sample Name:	NBK0012-31-P2-A	
Sample Type:	Standard	
Vial:	2:F,3	
Injection #:	1	
Injection Volume:	0.50 ul	
Run Time:	15.0 Minutes	

Acquired By:	System
Sample Set Name	2090716
Acq. Method Set:	VIVA QC_WATERS BEH C18
Processing Method	VIVA Process standrads method,
Channel Name:	PDA Ch3 214nm@4.8nm, PDA
Proc. Chnl. Descr.:	PDA Ch3 214nm@4.8nm, PDA

Date Acquired: 7/16/2019 12:42:51 PM CST
Date Processed: 7/16/2019 12:58:21 PM CST, 7/16/2019 12:58:31 PM CST, 7/16/2019 12:58:51 PM
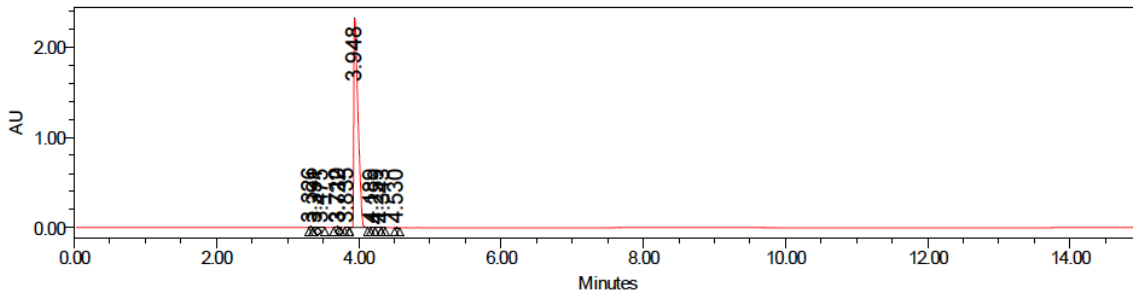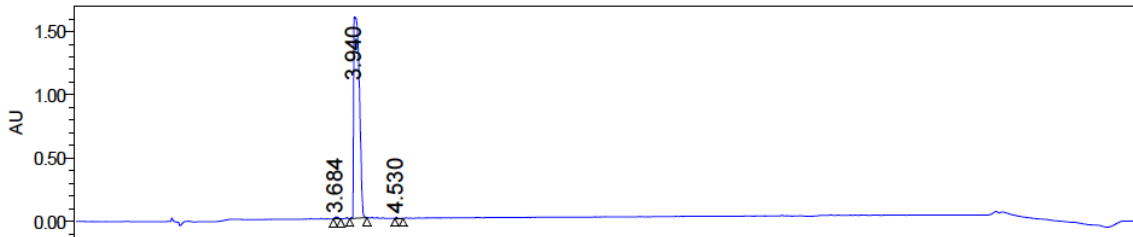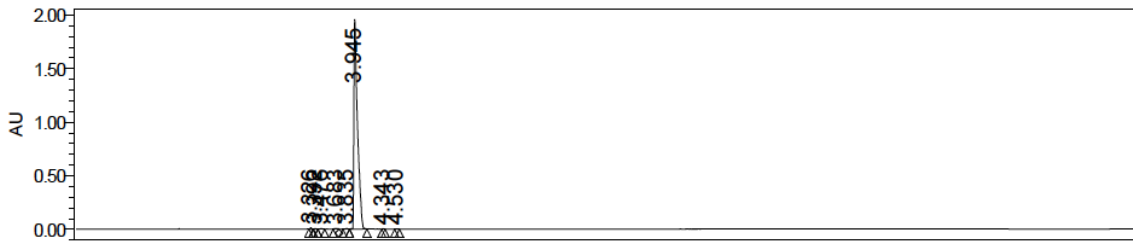
Channel: PDA Ch1 254nm@4.8nm; Processed Channel: PDA Ch1 254nm@4.8nm; Result Id: 5608;
Processing Method: VIVA Process standrads method
Channel: PDA Ch3 214nm@4.8nm; Processed Channel: PDA Ch3 214nm@4.8nm; Result Id: 5610;
Processing Method: 214 nm standrads method
Channel: PDA Ch2 280nm@4.8nm; Processed Channel: PDA Ch2 280nm@4.8nm; Result Id: 5609;
Processing Method: VIVA Process standrads method

Reported by User: System
Report Method: Injection Summary Repor
Report Method I 2412
Page: 1 of 2

Project Name: 2019.07month
Date Printed:
7/16/2019
1:02:21 PM PRC

56

Processed Channel Descr.: PDA Ch1 254nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch1 254nm@4.8nm	3.326	16418	0.24	13087
2	PDA Ch1 254nm@4.8nm	3.392	5996	0.09	5222
3	PDA Ch1 254nm@4.8nm	3.476	11804	0.17	6500
4	PDA Ch1 254nm@4.8nm	3.683	20525	0.30	6490
5	PDA Ch1 254nm@4.8nm	3.835	12117	0.18	8051
6	PDA Ch1 254nm@4.8nm	3.945	6840152	98.84	1959098
7	PDA Ch1 254nm@4.8nm	4.343	5180	0.07	4361
8	PDA Ch1 254nm@4.8nm	4.530	8081	0.12	5306

Processed Channel Descr.: PDA Ch2 280nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch2 280nm@4.8nm	3.326	18750	0.19	14776
2	PDA Ch2 280nm@4.8nm	3.391	7101	0.07	5943
3	PDA Ch2 280nm@4.8nm	3.475	10339	0.11	5810
4	PDA Ch2 280nm@4.8nm	3.710	27036	0.28	11765
5	PDA Ch2 280nm@4.8nm	3.722	2125	0.02	5043
6	PDA Ch2 280nm@4.8nm	3.835	13519	0.14	9829
7	PDA Ch2 280nm@4.8nm	3.948	9514828	98.84	2320903
8	PDA Ch2 280nm@4.8nm	4.189	5784	0.06	4072
9	PDA Ch2 280nm@4.8nm	4.259	6109	0.06	3591
10	PDA Ch2 280nm@4.8nm	4.343	7140	0.07	5658
11	PDA Ch2 280nm@4.8nm	4.530	13850	0.14	9843

Processed Channel Descr.: PDA Ch3 214nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch3 214nm@4.8nm	3.684	47052	0.53	15850
2	PDA Ch3 214nm@4.8nm	3.940	8810595	99.29	1587598
3	PDA Ch3 214nm@4.8nm	4.530	16009	0.18	9399

Reported by User: System
Report Method: Injection Summary Repor
Report Method I[2412
Page: 2 of 2

Project Name: 2019.07month
Date Printed:
7/16/2019
1:02:21 PM PRC

57

Compound 10

SAMPLE INFORMATION

Sample Name:	V1516-024-P1-A	Acquired By:	System
Sample Type:	Standard	Sample Set Name	20190515
Vial:	1:A,7	Acq. Method Set:	VIVA QC_WATERS BEH C18
Injection #:	1	Processing Method	VIVA01
Injection Volume:	0.10 ul	Channel Name:	PDA Ch3 214nm@4.8nm, PDA
Run Time:	15.0 Minutes	Proc. Chnl. Descr.:	PDA Ch3 214nm@4.8nm, PDA

Date Acquired: 5/15/2019 12:23:22 PM CST
Date Processed: 5/15/2019 12:39:08 PM CST, 5/15/2019 12:39:31 PM CST, 5/15/2019 12:39:44 PM
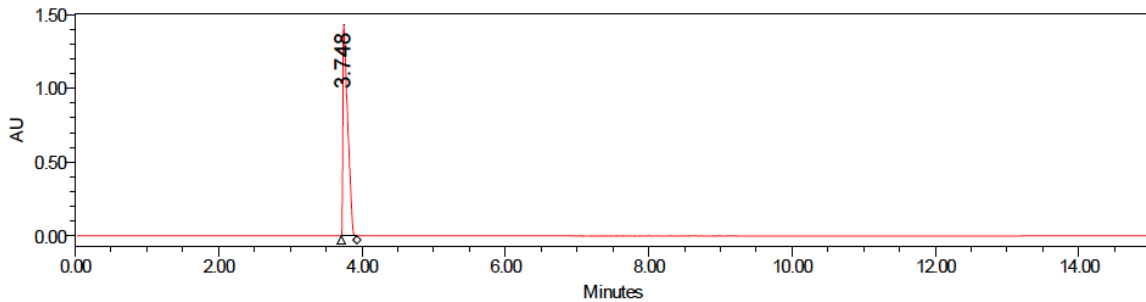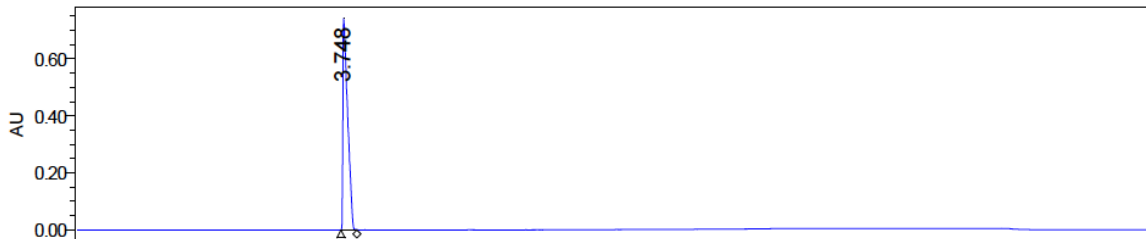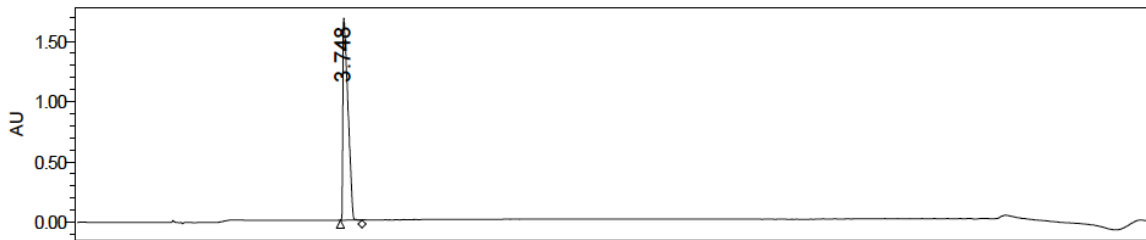
Channel: PDA Ch3 214nm@4.8nm; Processed Channel: PDA Ch3 214nm@4.8nm; Result Id: 6210;
Processing Method: VIVA01
Channel: PDA Ch1 254nm@4.8nm; Processed Channel: PDA Ch1 254nm@4.8nm; Result Id: 6221;
Processing Method: VIVA01
Channel: PDA Ch2 280nm@4.8nm; Processed Channel: PDA Ch2 280nm@4.8nm; Result Id: 6220;
Processing Method: VIVA01

Reported by User: System
Report Method: Injection Summary Repor
Report Method IC 1368
Page: 1 of 2

Project Name: 2019.04month
Date Printed:
5/15/2019
12:39:54 PM PRC

58

Processed Channel Descr.: PDA Ch1 254nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch1 254nm@4.8nm	3.748	3411073	100.00	742888

Processed Channel Descr.: PDA Ch2 280nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch2 280nm@4.8nm	3.748	6603674	100.00	1433549

Processed Channel Descr.: PDA Ch3 214nm@4.8nm

	Processed Channel Descr.	RT	Area	% Area	Height
1	PDA Ch3 214nm@4.8nm	3.748	8257088	100.00	1675705

Reported by User: System
Report Method: Injection Summary Repor
Report Method II 1368
Page: 2 of 2

Project Name: 2019.04month
Date Printed:
5/15/2019
12:39:54 PM PRC

59

Compound 11

```
========================================================================
 Acq. Operator   : admin
 Sample Operator : admin
 Acq. Instrument : 1260                      Location :   P1-A-05
 Injection Date  : 7/27/2020 10:23:28 AM         Inj :   1
                                           Inj Volume : 5.000 µl
 Different Inj Volume from Sample Entry!  Actual Inj Volume : 0.500 µl
 Acq. Method     : C:\Users\Public\Documents\ChemStation\1\Methods\Base-15min.M
 Last changed    : 7/27/2020 10:22:46 AM by admin
                   (modified after loading)
 Analysis Method : C:\USERS\PUBLIC\DOCUMENTS\CHEMSTATION\1\METHODS\Base-15min.M
 Last changed    : 7/27/2020 10:04:56 AM by admin
                   (modified after loading)
 Method Info     : Mobile:  A: Water(10mM NH4HCO3)  B: ACN      Flow Rate:1.0mL/min
                   Column:XBridge BEH C18  2.5um  4.6mm*150mm   Column Temperature:40 C
                   Gradient:5%B for 0.5min, increase to 95%B within 9.0min, 95%B for 2.0min,
                   back to 5%B within 1.0min, 5%B for 2.5min

 Sample Info     : Walkup method: 'Base-15min.M'
                   Target:

 Additional Info : Peak(s) manually integrated
```

60
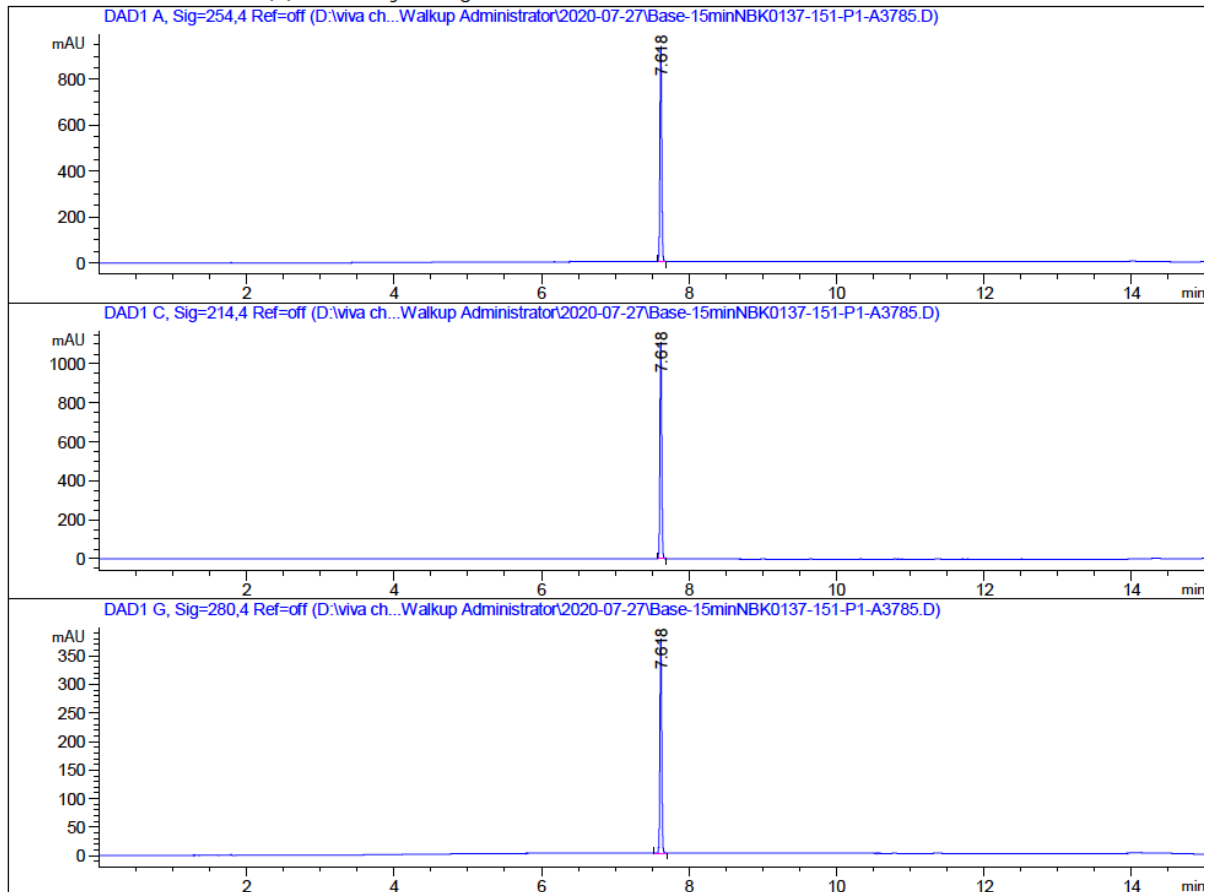
```
=================================================================
                        Area Percent Report
=================================================================

Sorted By              :        Signal
Multiplier             :        1.0000
Dilution               :        1.0000
Do not use Multiplier & Dilution Factor with ISTDs


Signal 1: DAD1 A, Sig=254,4 Ref=off

Peak RetTime Type  Width     Area        Height      Area
  #   [min]        [min]    [mAU*s]       [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   7.618 BB    0.0258 1553.19482  937.31970 100.0000

Totals :                  1553.19482  937.31970




Signal 2: DAD1 C, Sig=214,4 Ref=off

Peak RetTime Type  Width     Area        Height      Area
  #   [min]        [min]    [mAU*s]       [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   7.618 BB    0.0258 1839.49609 1109.71033 100.0000

Totals :                  1839.49609 1109.71033




Signal 3: DAD1 G, Sig=280,4 Ref=off

Peak RetTime Type  Width     Area        Height      Area
  #   [min]        [min]    [mAU*s]       [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   7.618 BB    0.0259  623.37848  375.74017 100.0000

Totals :                   623.37848  375.74017


=================================================================
                      *** End of Report ***
```

Empower™ 3
SOFTWARE

S A M P L E   I N F O R M A T I O N

Sample Name:	NBK0137-244-P1-A	Acquired By:	System
Sample Type:	Unknown	Sample Set Name:	20201110
Vial:	2:E,8	Acq. Method Set:	viva_Acid waters BEH_15min
Injection #:	1	Processing Method:	PM
Injection Volume:	2.00 ul	Channel Name:	PDA Ch1 214nm@4.8nm, PDA Ch2
Run Time:	15.0 Minutes	Proc. Chnl. Descr.:	PDA Ch1 214nm@4.8nm, PDA Ch2

Method Info:
Mobile Phase: A:Water(0.02%FA)   B: ACN(0.02%FA)
Gradient: 5%B for 0.5min,10%B increase to 95%B within 8.5min,95%B for 2.5min
            back to 5%B within 1.5min,5%B for 2.0min
Flow Rate: 0.3mL/min   Column Temperature: 35 C
Column: ACQUITY UPLC BEH C18 1.7um,2.1*150mm

Date Acquired: 11/10/2020 10:34:38 AM CST
Date Processed: 11/10/2020 12:40:46 PM CST, 11/10/2020 12:40:52 PM CST, 11/10/2020 12:40:57 PM CST
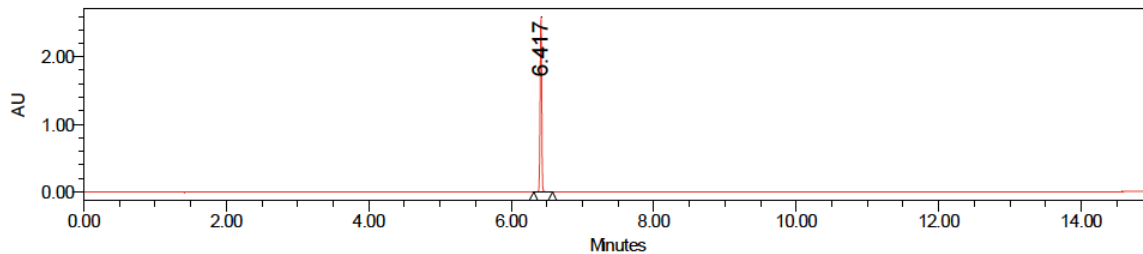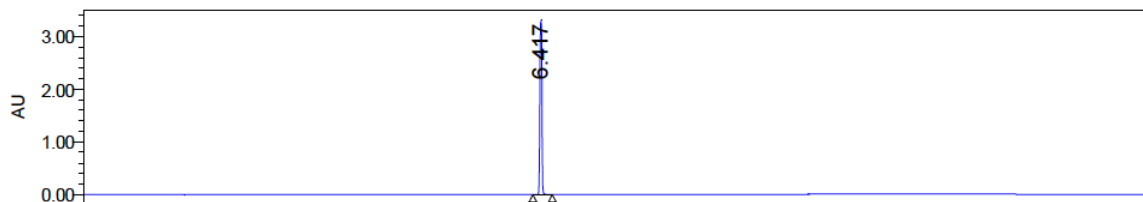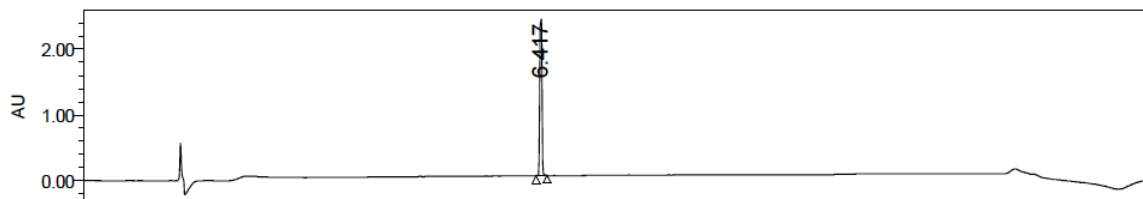
_____ Channel: PDA Ch1 214nm@4.8nm; Processed Channel: PDA Ch1 214nm@4.8nm; Result Id: 7482;
Processing Method: PM
_____ Channel: PDA Ch2 254nm@4.8nm; Processed Channel: PDA Ch2 254nm@4.8nm; Result Id: 7483;
Processing Method: PM
_____ Channel: PDA Ch3 280nm@4.8nm; Processed Channel: PDA Ch3 280nm@4.8nm; Result Id: 7484;
Processing Method: PM

Reported by User: System
Report Method: VIva Title update report
Report Method ID: 1361
Page: 1 of 2

Project Name: VIVA-2020.11month
Date Printed:
11/10/2020
12:41:11 PM PRC

62

Processed Channel Descr.: PDA Ch1 214nm@4.8nm

	Processed Channel Descr.	RT	Height	Area	% Area
1	PDA Ch1 214nm@4.8nm	6.417	2377061	5438631	100.00

Processed Channel Descr.: PDA Ch2 254nm@4.8nm

	Processed Channel Descr.	RT	Height	Area	% Area
1	PDA Ch2 254nm@4.8nm	6.417	3328551	6651686	100.00

Processed Channel Descr.: PDA Ch3 280nm@4.8nm

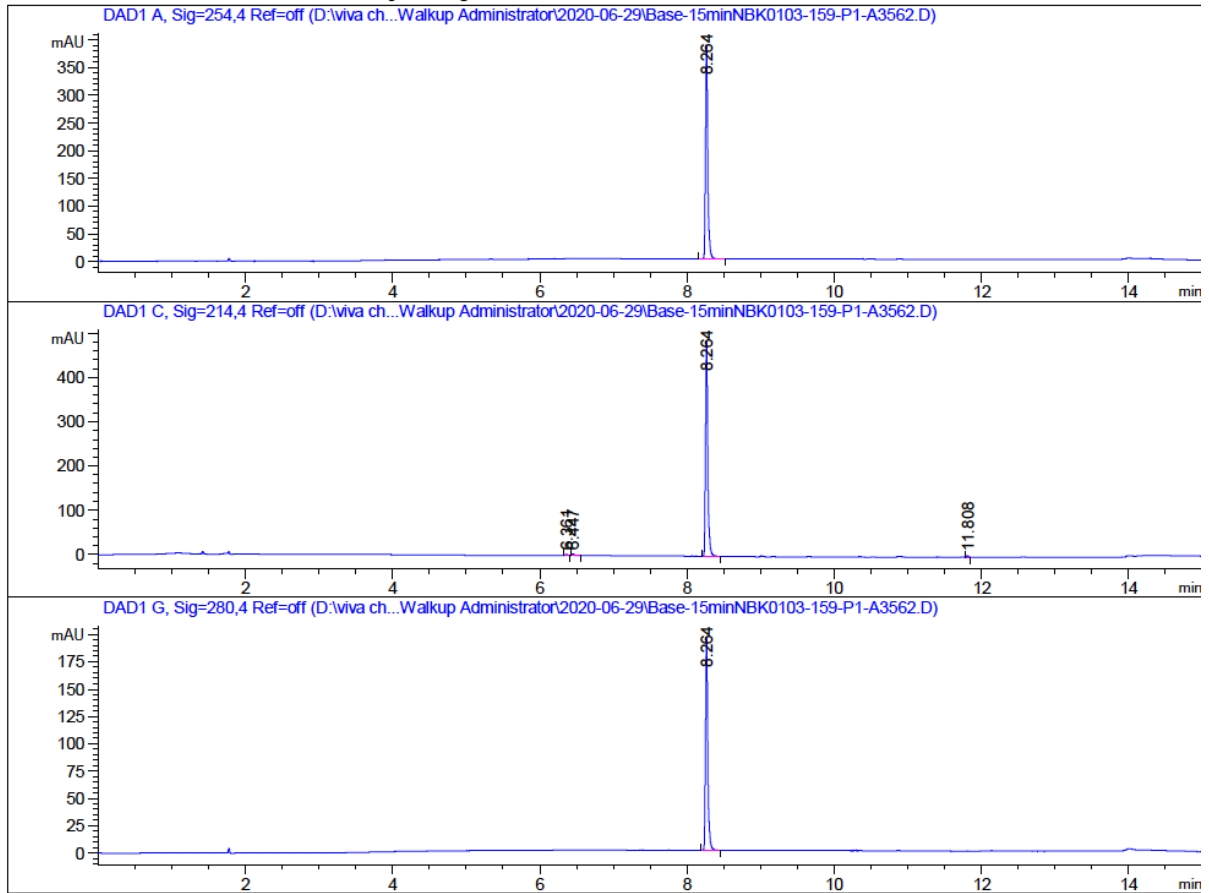|	Processed Channel Descr.	RT	Height	Area	% Area
1	PDA Ch3 280nm@4.8nm	6.417	2593967	4144660	100.00

Reported by User: System
Report Method: VIva Title update report
Report Method ID: 1361
Page: 2 of 2

Project Name: VIVA-2020.11month
Date Printed:
11/10/2020
12:41:11 PM PRC

63

Sample Name: NBK0103-159-P1-A

Compound 13

===

Acq. Operator   : admin
Sample Operator : admin
Acq. Instrument : 1260   Location :   P1-B-06
Injection Date : 6/29/2020 9:28:21 AM   Inj :   1
                                            Inj Volume : 5.000 µl
Different Inj Volume from Sample Entry! Actual Inj Volume : 2.000 µl
Acq. Method   : C:\Users\Public\Documents\ChemStation\1\Methods\Base-15min.M
Last changed   : 6/29/2020 9:27:39 AM by admin
                  (modified after loading)
Analysis Method : C:\Users\Public\Documents\ChemStation\1\Methods\standby.M
Last changed   : 6/29/2020 1:33:50 PM by admin
                  (modified after loading)
Sample Info   : Walkup method: 'Base-15min.M'
                  Target:

Additional Info : Peak(s) manually integrated

```
===================================================================
                       Area Percent Report
===================================================================

Sorted By           :        Signal
Multiplier          :        1.0000
Dilution            :        1.0000
Do not use Multiplier & Dilution Factor with ISTDs


Signal 1: DAD1 A, Sig=254,4 Ref=off

Peak RetTime Type  Width     Area       Height      Area
  #   [min]        [min]    [mAU*s]      [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   8.264 BB    0.0318  823.43774  387.68594 100.0000

Totals :                   823.43774  387.68594




Signal 2: DAD1 C, Sig=214,4 Ref=off

Peak RetTime Type  Width     Area       Height      Area
  #   [min]        [min]    [mAU*s]      [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   6.361 BB    0.0303    3.89124    2.04023   0.3701
  2   6.447 BB    0.0294    7.61238    3.87572   0.7240
  3   8.264 BB    0.0317 1033.83057  487.38068  98.3295
  4  11.808 MM R  0.0335    6.05956    3.01099   0.5763

Totals :                  1051.39375  496.30761




Signal 3: DAD1 G, Sig=280,4 Ref=off

Peak RetTime Type  Width     Area       Height      Area
  #   [min]        [min]    [mAU*s]      [mAU]        %
----|-------|----|-------|----------|----------|--------|
  1   8.264 BB    0.0318  414.52899  195.34364 100.0000

Totals :                   414.52899  195.34364


===================================================================
                      *** End of Report ***
```