The Prediction of Oral Pharmacokinetics Using Combination of In Silico Descriptors and In Vitro ADME Properties

Yohei Kosugi, and Natalie Hosea.*

Global DMPK, Takeda California Inc., San Diego, California 92121, United States

Supporting Information

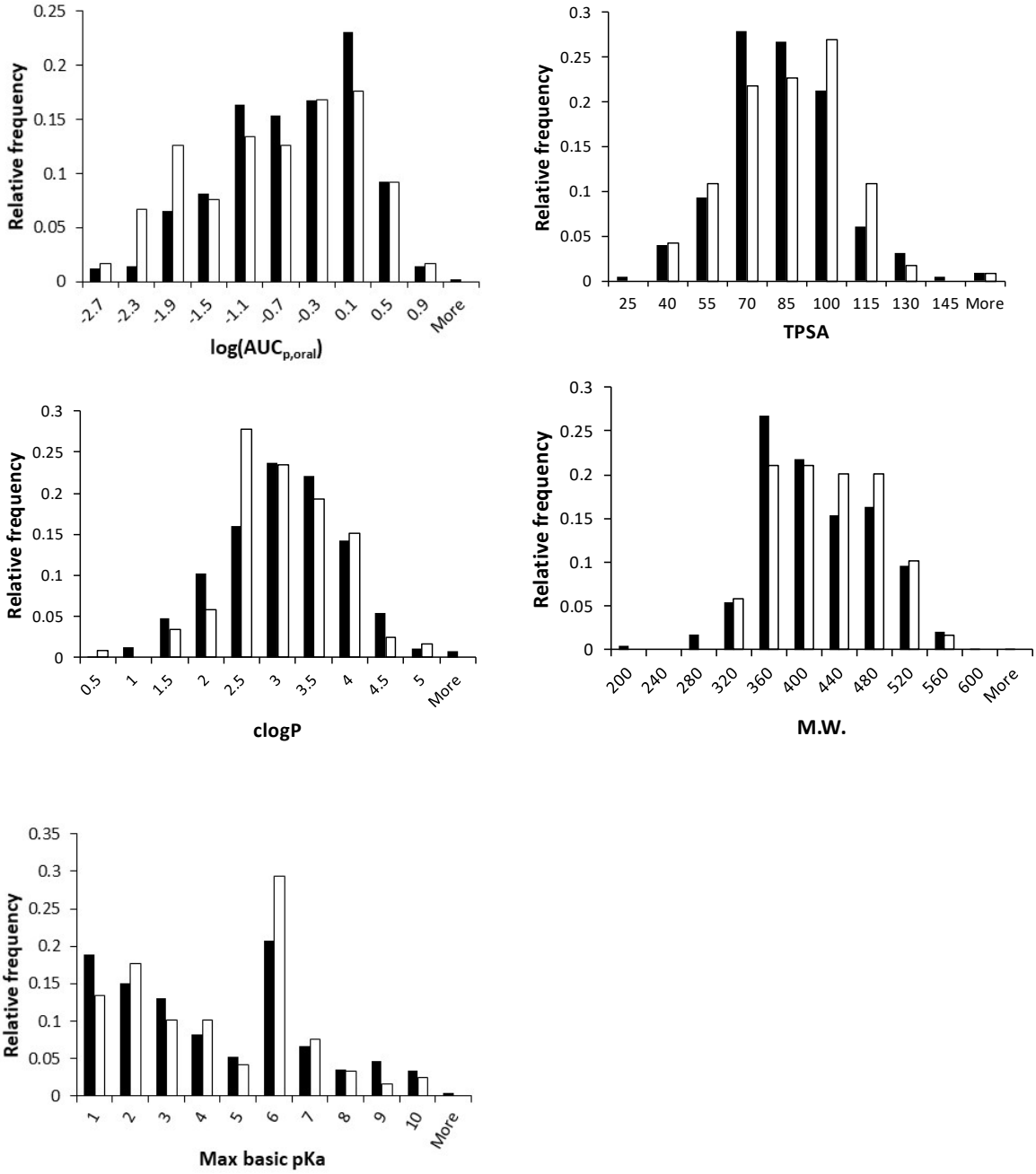

Figure S1. Distribution of log($AUC_{p,oral}$) and physicochemical parameters, TPSA, clogP, MW, and maximum basic pKa between the cluster-split training set (solid column) and the test set (open column). TPSA, clogP, and MW were calculated by StarDrop. Maximum basic pKa was

calculated by JChem for Excel (ver.16.3, ChemAxon, Budapest, Hungary). The relative frequency was calculated by dividing a frequency count by sum of all frequencies. The number of compounds in training and test sets was 476 and 119, respectively.

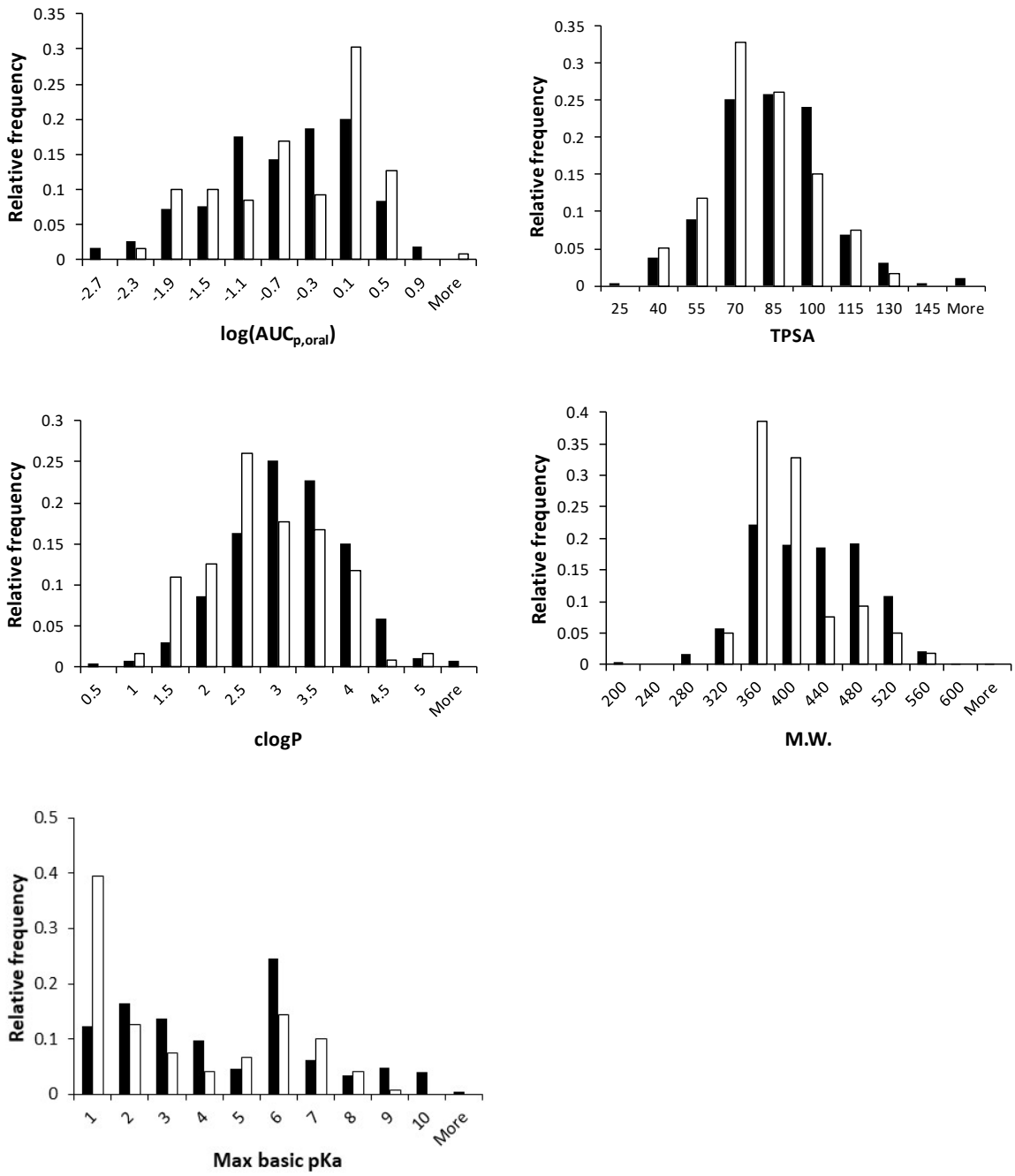

Figure S2. Distribution of log($AUC_{p,oral}$) and physicochemical parameters, TPSA, clogP, MW, and maximum basic pKa between the time-split training set (solid column) and the test set (open column). TPSA, clogP, and MW were calculated by StarDrop. Maximum basic pKa was

calculated by JChem for Excel (ver.16.3, ChemAxon, Budapest, Hungary). The relative frequency was calculated by dividing a frequency count by sum of all frequencies. The number of compounds in training and test sets was 476 and 119, respectively.

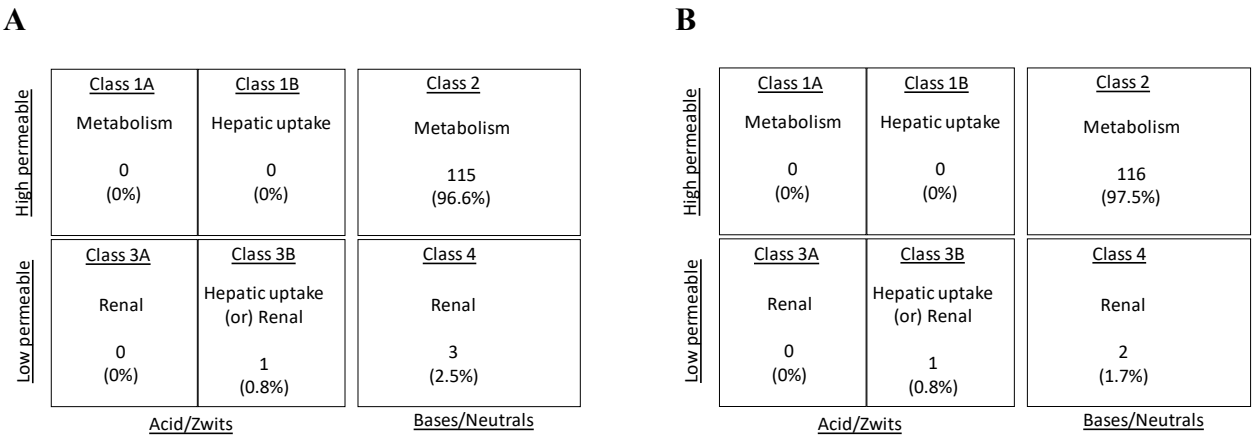

Figure S3. ECCS classification in (A) cluster-split, and (B) time-split test sets. ECCS class was assessed for total 119 compounds in the cluster-split and time-split test sets. The Papp value of 50 nm/s was applied as the low/high permeability class boundary.

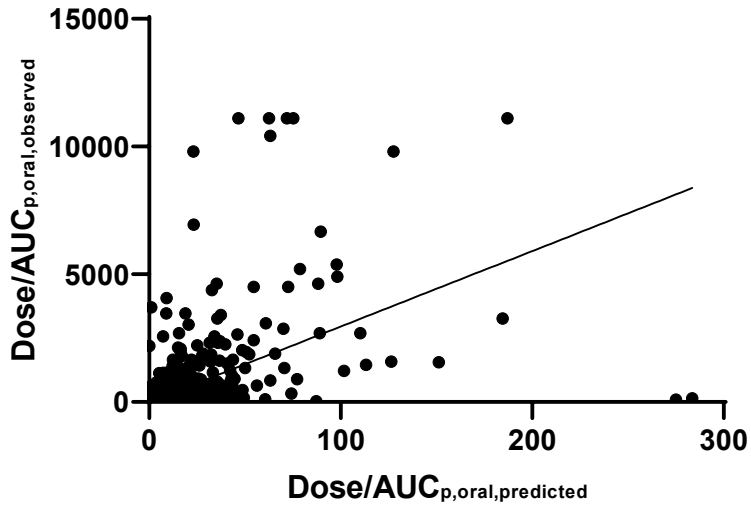

Figure S4. Relationship between mean observed dose/$AUC_{p,oral}$ and values predicted by IVIVE in all data (n=595). Solid line represents the regression.

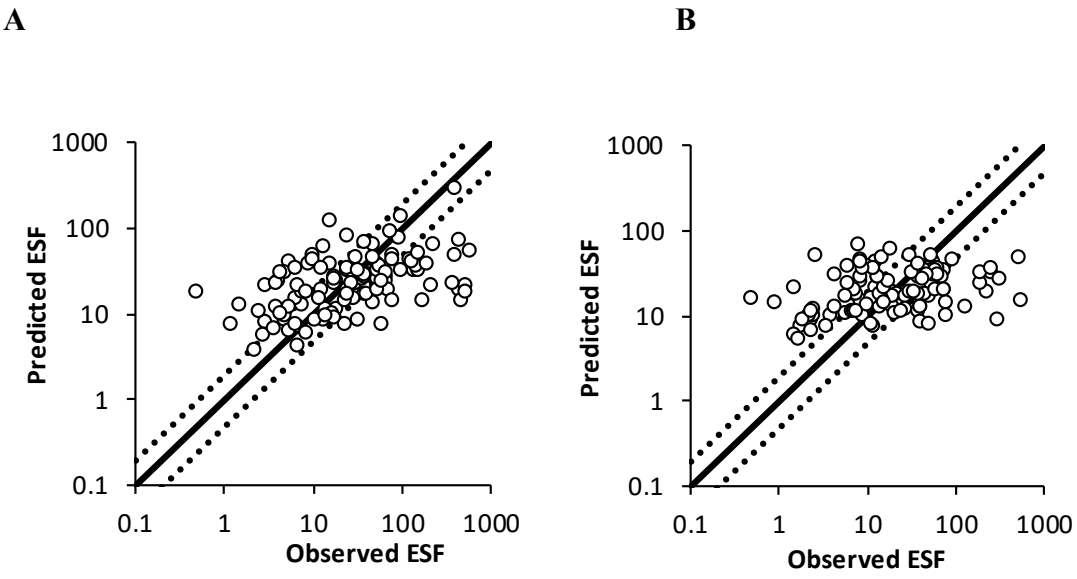

Figure S5. Relationship between observed ESF and values predicted by RF in (A) cluster-split, and (B) time-split test sets. The solid line indicates unity. Dashed lines represent ± 2-fold of unity.

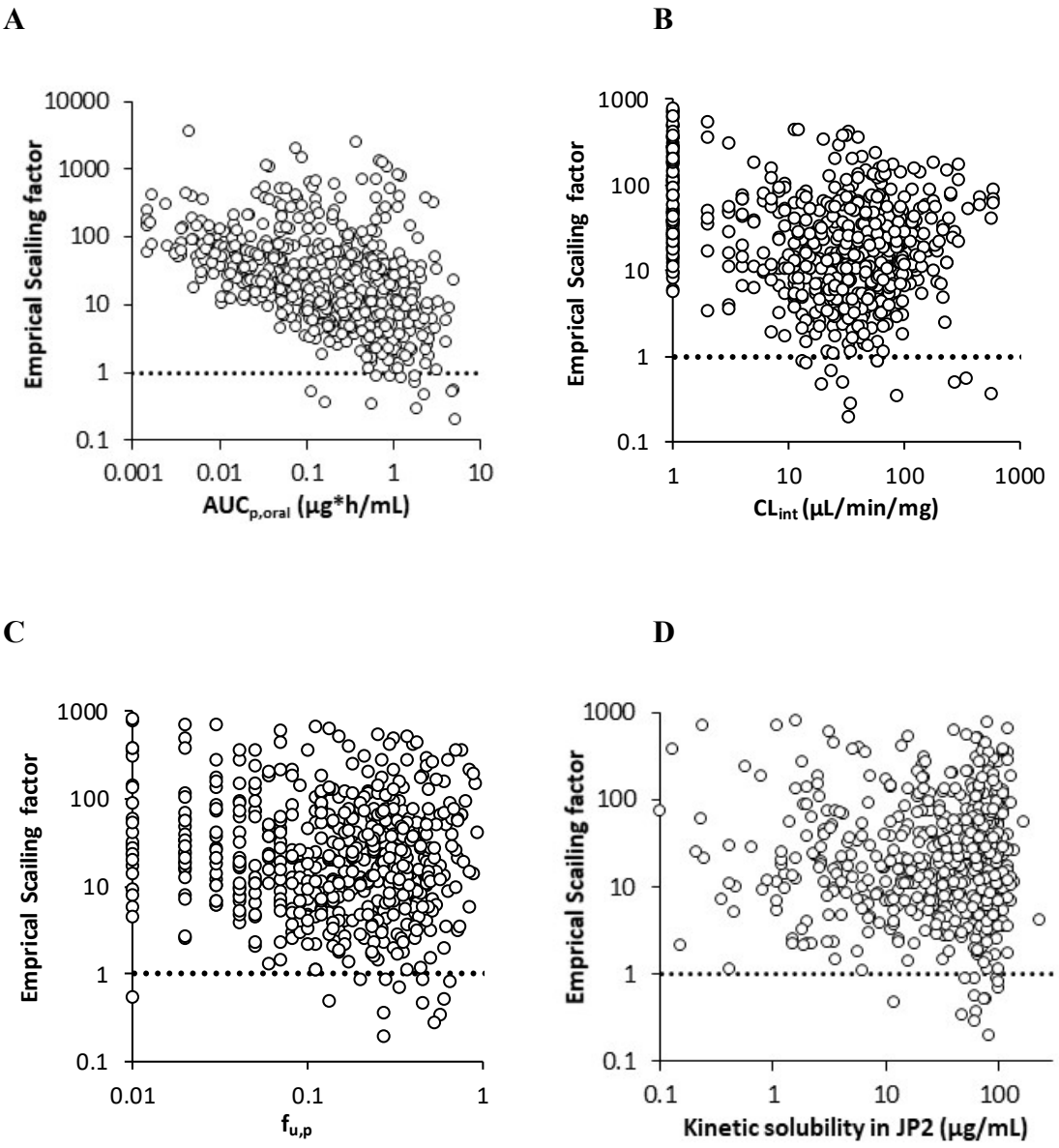

Figure S6. Relationship between ESF and (A) mean observed $AUC_{p,oral}$, (B) CL_{int}, (C) $f_{u,p}$, and (D) kinetic solubility in JP2. Each figure shows the results of 595 compounds. Dashed line represents the line of unity.

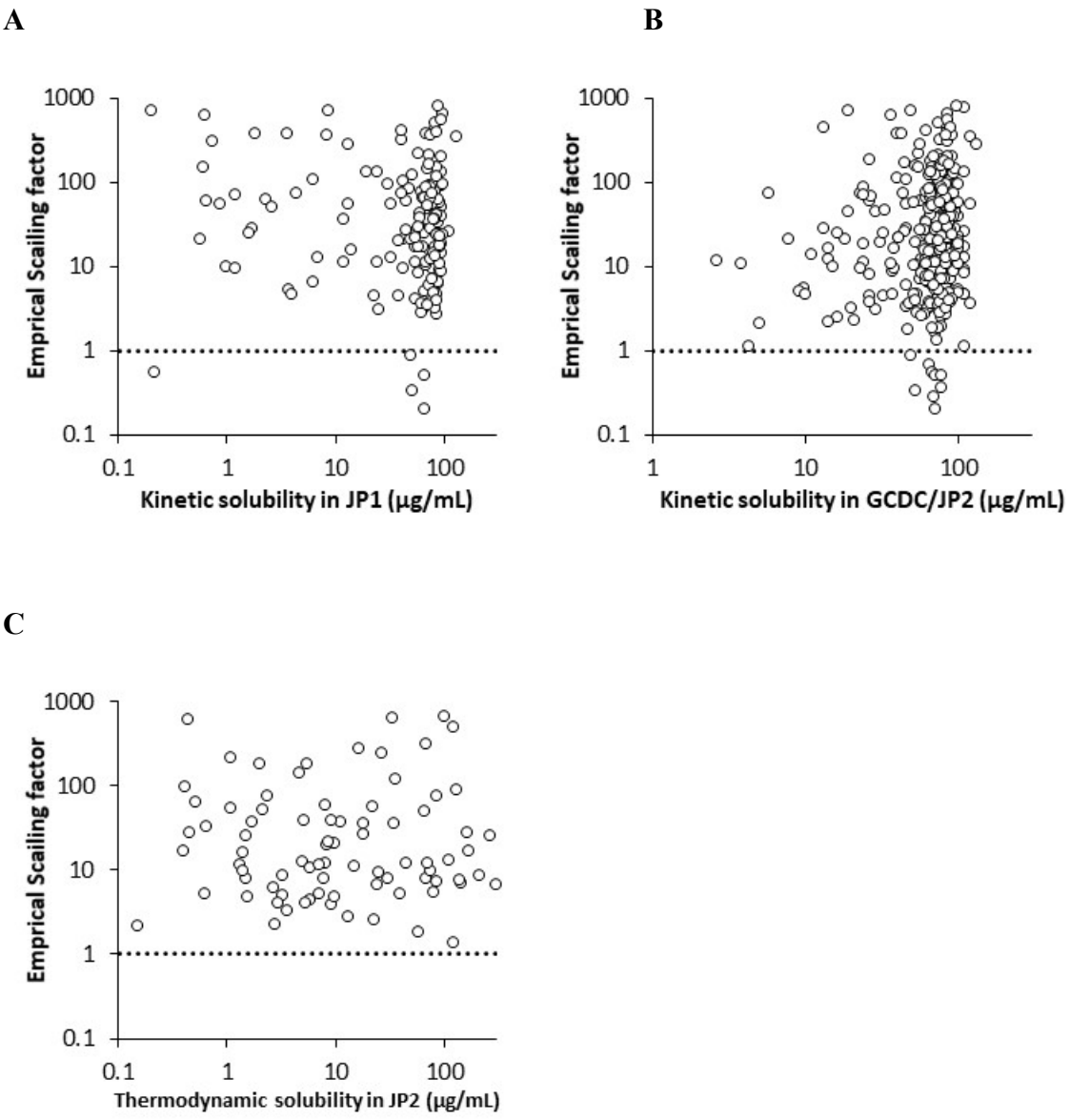

Figure S7. Relationship between ESF and (A) kinetic solubility in JP1 (n=172), (B) kinetic solubility in GCDC/JP2 (n=356), and (C) thermodynamic solubility in JP2 (n=106). Dashed line represents the line of unity.

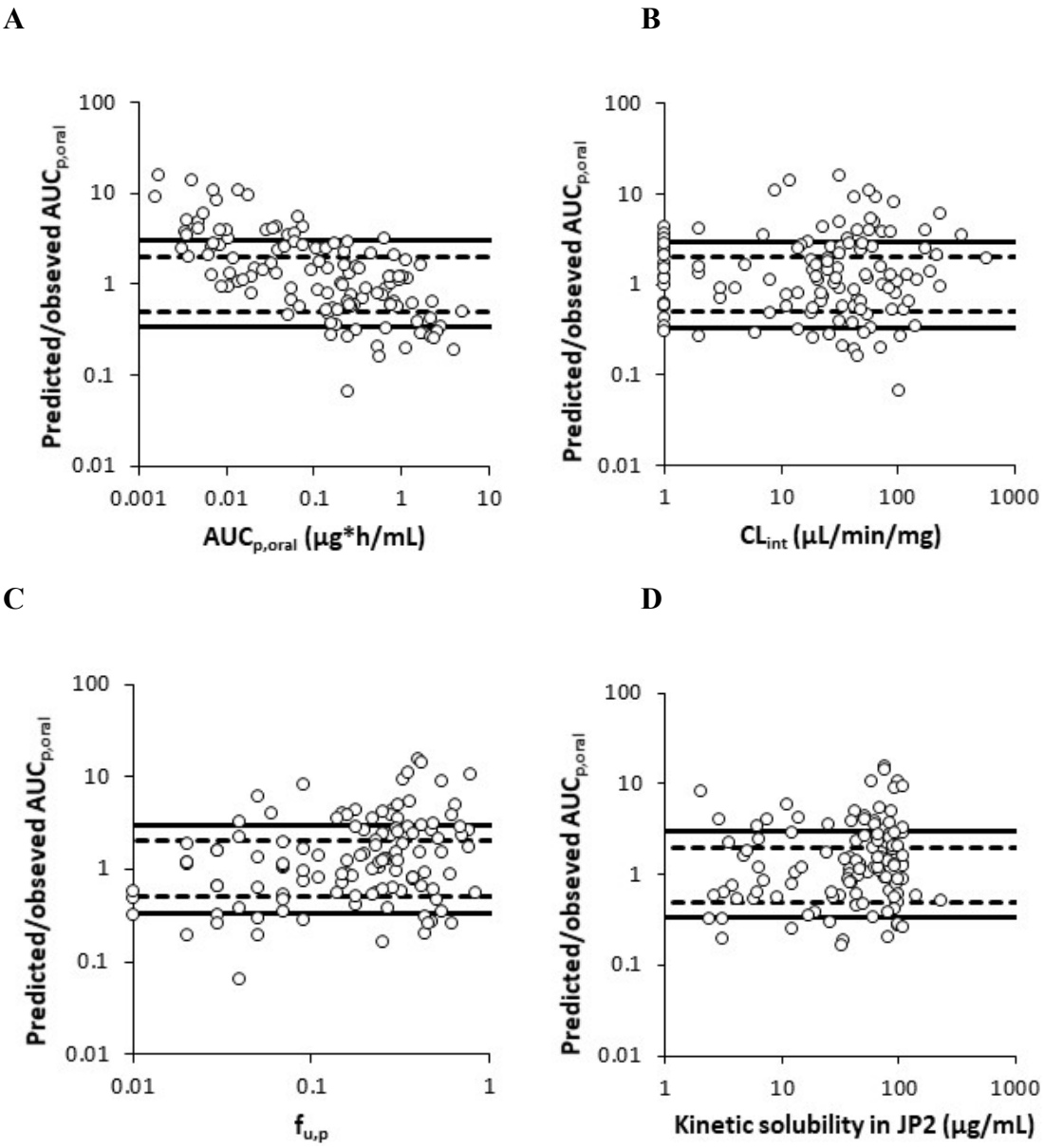

Figure S8. Relationship between the fold error (ratio of predicted to observed $AUC_{p,oral}$) in GPOPT incorporating CL_{int}, $f_{u,p}$, and kinetic solubility in JP2 on the cluster-split test set and (A) mean observed $AUC_{p,oral}$, (B) CL_{int}, (C) $f_{u,p}$, and (D) kinetic solubility in JP2. Dashed lines represent ratios of the predicted $AUC_{p,oral}$ to the observed $AUC_{p,oral}$ of 0.5 and 2, respectively. The solid lines represent ratios of the predicted $AUC_{p,oral}$ to the observed $AUC_{p,oral}$ of 0.33 and 3, respectively.

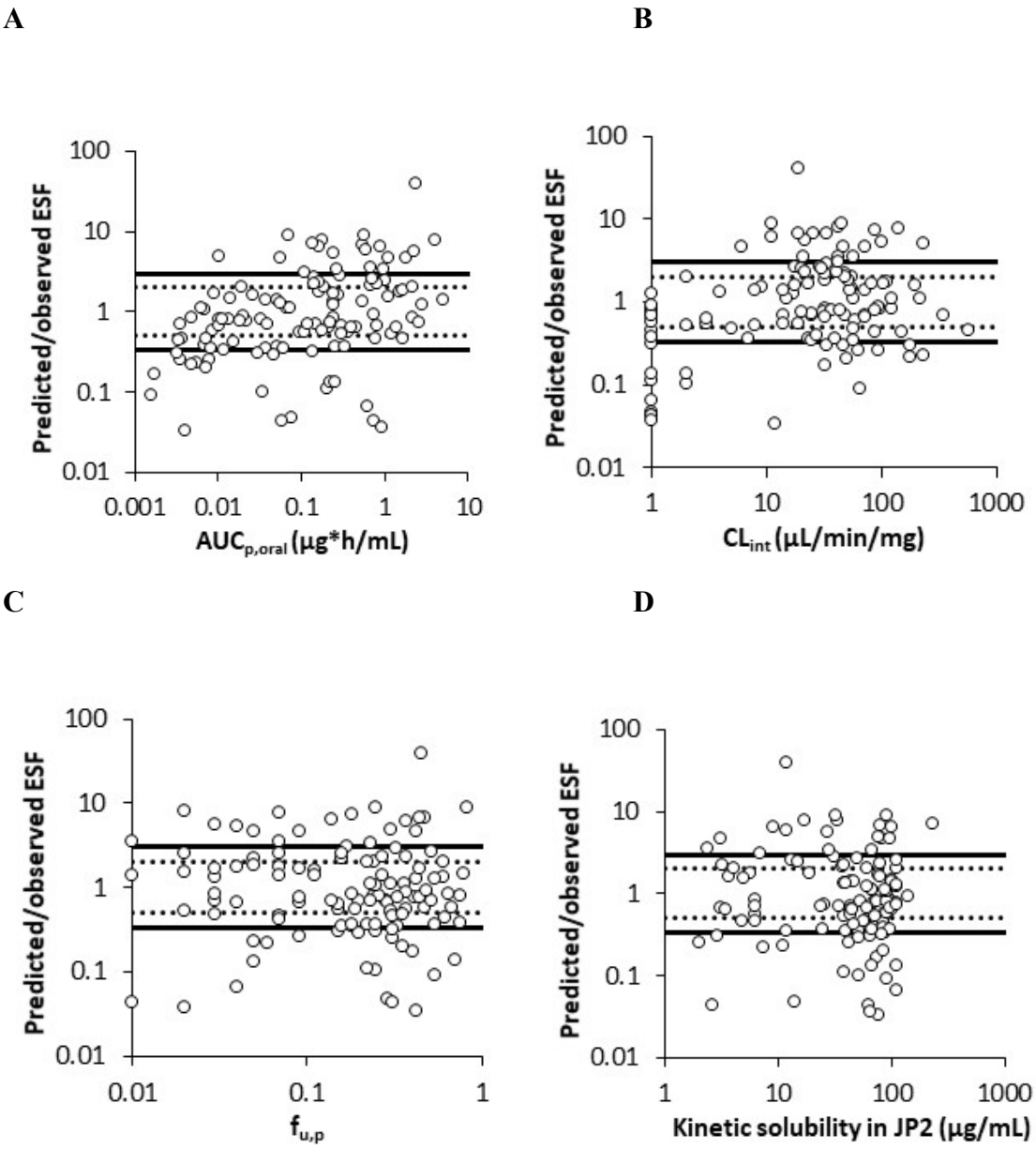

Figure S9. Relationship between the fold error (ratio of predicted to observed ESF) in RF on the cluster-split test set and (A) mean observed $AUC_{p,oral}$, (B) CL_{int}, (C) $f_{u,p}$, and (D) kinetic solubility in JP2. Dashed lines represent ratios of the predicted $AUC_{p,oral}$ to the observed $AUC_{p,oral}$ of 0.5 and 2, respectively. The solid lines represent ratios of the predicted $AUC_{p,oral}$ to the observed $AUC_{p,oral}$ of 0.33 and 3, respectively.

Table S1. $AUC_{p,oral}$ Prediction Using Well-Stirred Models Incorporated with $f_{u,mic}$

Dataset used for IVIVE	Number of test set	Statistics	ESF=1	ESF=29.5	ESF_{pred}
All	595	% < 2-fold	4.9	33	36[a]
		R^2	0.351	0.351	0.461[a]
		RMSE	1.47	0.741	0.666[a]
Time-split test set	119	% < 2-fold	6.7	34	43[b]
		R^2	0.470	0.470	0.538[b]
		RMSE	1.35	0.682	0.608[b]
Cluster-split test set	119	% < 2-fold	2.5	35	40[b]
		R^2	0.468	0.468	0.592[b]
		RMSE	1.49	0.679	0.584[b]

R^2, and RMSE were calculated using $\log(AUC_{p,oral})$. [a]The model for predicting ESF was developed using fivefold cross-validation procedure. [b]The model for predicting ESF was developed using each training set.

Table S2. ESF Prediction Using RF

Validation methods	Number of test set	Statistics	RF
5-fold cross-validation	595	% < 2-fold	37
		Q^2	0.183
		$RMSE_{cv}$	0.656
Time-spliting	119	% < 2-fold	43
		R^2	0.115
		RMSE	0.583
Cluster-spliting	119	% < 2-fold	44
		R^2	0.257
		RMSE	0.566

R^2, Q^2, RMSE, and $RMSE_{CV}$ were calculated using log(ESF). For fivefold cross-validation, the dataset of 595 compounds was randomly split into five different groups, using four of the groups for training and the remaining part for testing. This cross-validation process was repeated five times that all groups were left out once.

Table S3. Summary of the Top 20 Molecular Descriptors of GPOPT on the Cluster-Split Training Set

Molecular descriptor	Description	Molecular descriptor	Description
Flex	The flexibility index	N4	Number of non-aromatic, uncharged, nitrogens with exactly one hydrogen, connected to an aromatic atom
pyridones	Number of sp2 oxygens in a pyridone ring	PRX-time1	Number of amide and sulfonamide side chains
RbasicNH0	Number of cyclic sp3 nitrogens with no hydrogen connected to three sp3 carbons in a molecule with no acidic groups	aaNH	Number of aromatic nitrogens with one hydrogen
aromO	Number of aromatic sp2 oxygens	nH0indole-like	Number of indole nitrogens with no hydrogen
arylNHCO	Number of secondary amides with nitrogen connected to an aromatic atom	tert-amine-t11	Number of tertiary nitrogen non-anilines
ertl-33	Number of sulfur atoms with at least two single bonds	S3	Number of aromatic sulfurs
hetero-halo-di-n-arom	Number of aromatic carbons connected to exactly two aromatic nitrogens and one heteroatom	ew60	Number of trifluoromethyl groups connected to one aromatic atom
frg-26	Number of sp3 nitrogens with no hydrogens connected to two sp3 carbons and one aromatic sp2 carbon	xccn-t12	Number of secondary amines connected to either methyl or ethyl groups on one side and in the beta position of an oxygen or nitrogen atom on the other side
nHindole-	Number of indole nitrogens with one	frg-8	Number of para interactions in a disubstituted

like	hydrogen	benzene ring with no other substituents