Supporting Information

Reaction-based Enumeration, Active Learning, and Free Energy Calculations to Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin Dependent Kinase 2 Inhibitors

Kyle D. Konze[‡], Pieter H. Bos[‡], Markus K. Dahlgren, Karl Leswing, Ivan Tubert-Brohman, Andrea Bortolato, Braxton Robbason, Robert Abel, Sathesh Bhat[*]

[‡]These authors contributed equally to this work

[*] sathesh.bhat@schrodinger.com

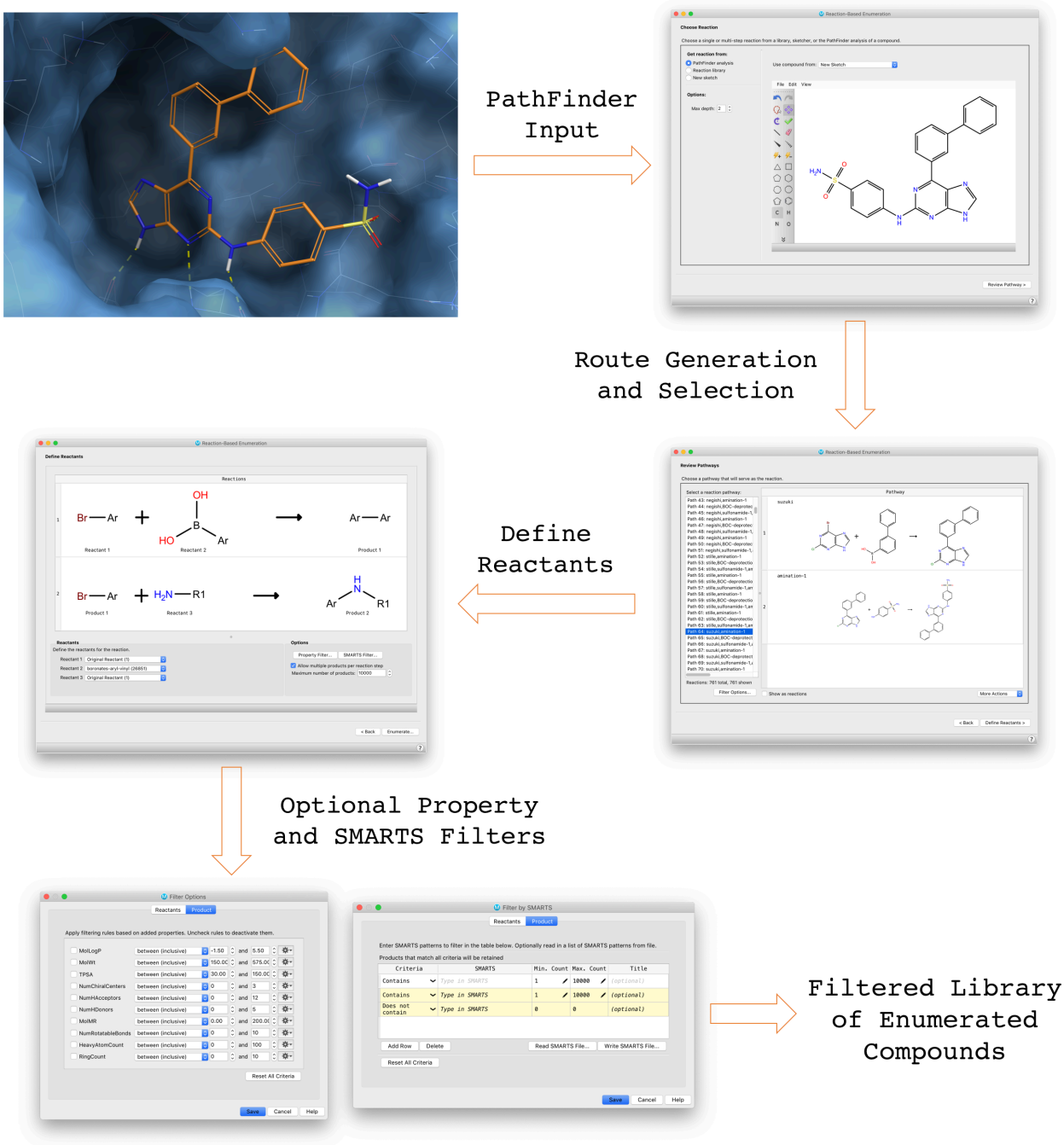


Figure S1. Example of the R-group Enumeration Workflow in the PathFinder GUI.

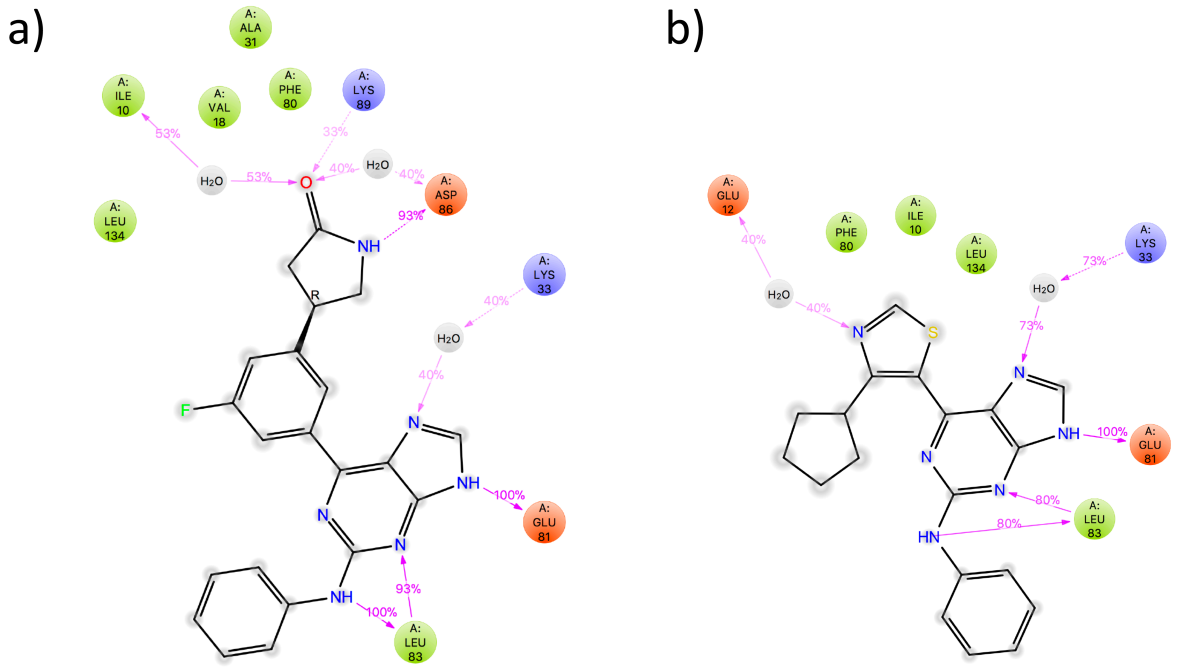
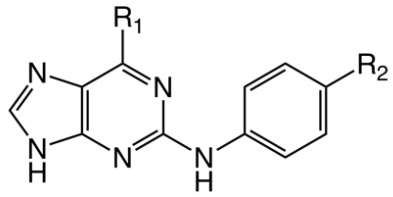
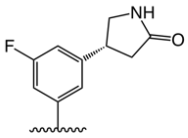
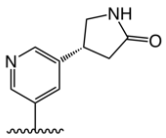
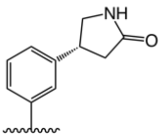
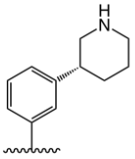
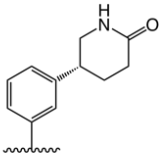
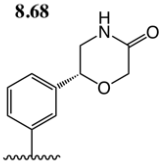
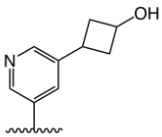
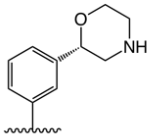
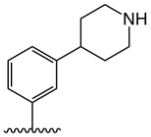
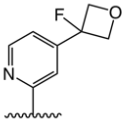
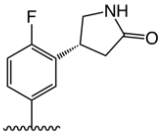
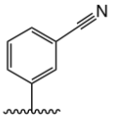
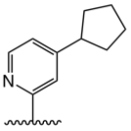
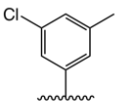
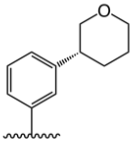
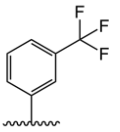
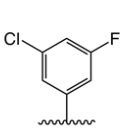
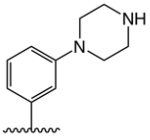
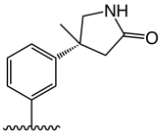
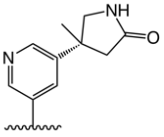
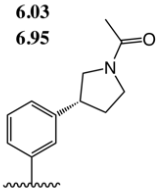
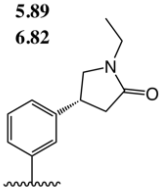
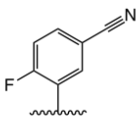
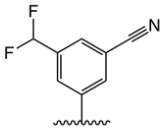
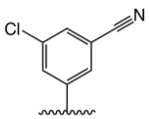


Figure S2. FEP Simulation Interaction Diagrams. a) SID of **2** with CDK2 during the FEP simulation indicating key features contributing to binding affinity: hydrogen bonding to the hinge, and direct/water-mediated hydrogen bonds to I10, D86, and K89. **b**) SID of **6** with CDK2 during the FEP simulation indicating key features contributing to binding affinity: hydrogen bonding to the hinge, water-mediated hydrogen bond of the thiazole nitrogen, and hydrophobic interactions of the cyclopentyl moiety.

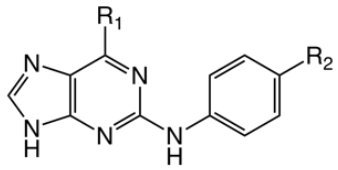
Table S1. Phenyl and Pyridine R-groups from PathFinder Enumeration.[a]

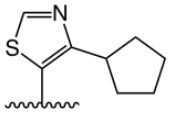
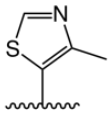
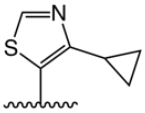
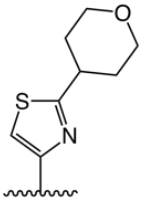
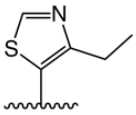
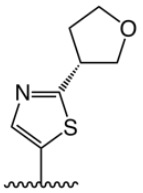
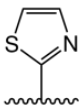
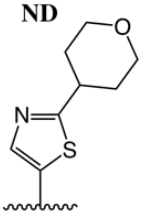


R_1 :					
Compound ID :	**2**	**S1**	**S2**	**5**	**S3**
R_2 = H :	**7.81**	**7.61**	**7.42**	**7.36**	**7.31**
R_2 = SO_2NH_2 :	**8.68**	**8.47**	**8.74**	**ND**	**8.89**
R_1 :					
Compound ID :	**S4**	**S5**	**S6**	**S7**	**S8**
R_2 = H :	**7.19**	**6.84**	**6.71**	**6.70**	**6.64**
R_2 = SO_2NH_2 :	**8.22**	**7.63**	**ND**	**7.03**	**7.22**
R_1 :					
Compound ID :	**S9**	**7**	**S10**	**S11**	**S12**
R_2 = H :	**6.54**	**6.41**	**6.26**	**6.12**	**6.10**
R_2 = SO_2NH_2 :	**8.48**	**7.48**	**7.15**	**6.80**	**7.42**
R_1 :					
Compound ID :	**S13**	**S14**	**S15**	**S16**	**S17**
R_2 = H :	**6.03**	**5.89**	**6.88**	**6.60**	**6.38**
R_2 = SO_2NH_2 :	**6.95**	**6.82**	**ND**	**ND**	**ND**
R_1 :					
Compound ID :	**S18**	**S19**	**S20**	**S21**	**S22**
R_2 = H :	**5.40**	**5.26**	**6.20**	**6.13**	**5.87**
R_2 = SO_2NH_2 :	**ND**	**ND**	**ND**	**ND**	**ND**

[a]The predicted dG and error estimates in kcal/mol can be found in the Supporting Information (Table S6).

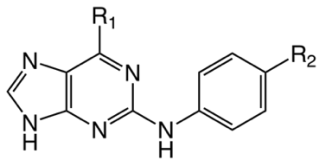
Table S2. Thiazole R-groups from PathFinder Enumeration.[a]

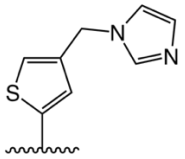
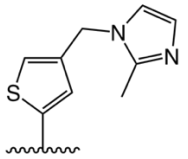
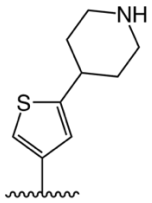
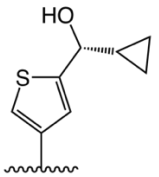
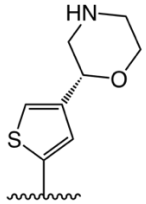
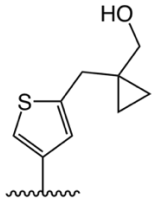
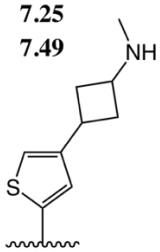
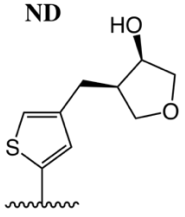


R_1 :		
Compound ID :	**6**	**S26**
R_2 = H :	**7.11**	**6.43**
R_2 = SO_2NH_2 :	**8.36**	**7.61**
R_1 :		
Compound ID :	**S23**	**S27**
R_2 = H :	**6.90**	**6.05**
R_2 = SO_2NH_2 :	**8.07**	**ND**
R_1 :		
Compound ID :	**S24**	**S28**
R_2 = H :	**6.76**	**5.22**
R_2 = SO_2NH_2 :	**8.22**	**ND**
R_1 :		
Compound ID :	**S25**	**S29**
R_2 = H :	**6.48**	**4.77**
R_2 = SO_2NH_2 :	**7.55**	**ND**

[a]The predicted dG and error estimates in kcal/mol can be found in the Supporting Information (Table S6).

Table S3. Thiophene R-groups from PathFinder Enumeration.[a]



R_1 :		
Compound ID :	**S30**	**S33**
R_2 = H :	**7.66**	**6.83**
R_2 = SO_2NH_2 :	**7.79**	**ND**
R_1 :		
Compound ID :	**4**	**S34**
R_2 = H :	**7.39**	**6.75**
R_2 = SO_2NH_2 :	**7.86**	**ND**
R_1 :		
Compound ID :	**S31**	**S35**
R_2 = H :	**7.25**	**6.62**
R_2 = SO_2NH_2 :	**7.49**	**ND**
R_1 :		
Compound ID :	**S32**	**S36**
R_2 = H :	**6.87**	**6.40**
R_2 = SO_2NH_2 :	**7.64**	**ND**

[a]The predicted dG and error estimates in kcal/mol can be found in the Supporting Information (Table S6).

Table S4. PathFinder Reactions - Schrödinger Suite 2019-1

Name	Description	Reactant Library 1	Reactant Library 2	Reactant Library 3
123-triazole-1	Copper-catalyzed Huisgen cycloaddition of an alkyne and an azide to form a 123-triazole (implicit azide formation from primary and secondary halides).	Monosubstituted alkynes	Primary and secondary halides	
123-triazole-2	Copper-catalyzed Huisgen cycloaddition of an alkyne and an azide to form a 123-triazole.	Monosubstituted alkynes	Azides	
123-triazole-3	Ruthenium-catalyzed Huisgen cycloaddition of an alkyne and an azide to form a 123-triazole (implicit azide formation from primary and secondary halides).	Monosubstituted alkynes	Primary and secondary halides	
123-triazole-4	Ruthenium-catalyzed Huisgen cycloaddition of an alkyne and an azide to form a 123-triazole.	Monosubstituted alkynes	Azides	
124-triazole-1	124-triazole from an aryl nitrile and an amidine.	Aryl nitriles	Amidines	
124-triazole-2	124-triazole from a nitrile and hydrazide.	Nitriles	Hydrazides	
124-triazole-3	124-triazole from a nitrile and a carboxylate using hydrazine.	Nitriles	Carboxylic acids	
5-het-amination	Amination of 5-member heterocycles (n-alkyl-imidazole, thiazole, oxazole, benzothiazole, benzimidazole, and benzoxazole).	2-Halo-heterocycles	Primary and secondary amines	
5-het-arylation	Arylation of 5-member heterocycles, including: thiophene, n-methyl-imidazole, 4-methyl-thiazole, 2-methyl-furan, benzothiophene, and benzothiazole.	2-Unsubstituted heterocycles	Aryl bromides	
BOC-deprotection	Deprotection of a Boc-protected amine			

BOC-protection	Protection of a primary or secondary amine with a Boc-group	Primary and secondary amines	
N-azole-arylation	N-arylation of azoles, including: indole, pyrazole, imidazole, triazole, tetrazole, benzimidazole, and indazoles.	Aryl boronates	Azoles
acylation-1	Acylation of an alkyne with an acid chloride.	Monosubstituted alkynes	Acid chlorides
alkylation-1	Alkylation of a primary or secondary amine using a primary or secondary alkyl halide.	Primary and secondary halides	Primary and secondary amines
alkylation-2	N-Alkylation of heterocycles using a primary or secondary alkyl halide.	Primary and secondary halides	Heterocycles
amide_coupling-1	Traditional amide coupling using a carboxylic acid and a primary or secondary amine.	Carboxylic acids	Primary and secondary amines
amide_coupling-2	Schotten-Baumann: Amide coupling using an acid chloride and a primary or secondary amine.	Acid chlorides	Primary and secondary amines
amide_coupling-3	Intramolecular amide coupling		
amide_coupling-4	Intramolecular amide coupling		
amination-1	Buchwald-Hartwig Amination: N-arylation of primary and secondary amines using aryl halides.	Aryl halides	Primary and secondary amines
amination-2	Reductive Amination: Alkylation of primary or secondary amines with a ketone or an aldehyde.	Aldehydes and ketones	Primary and secondary amines
aminothiophene-1	Gewald Aminothiophene synthesis.	2-Unsubstituted ketones	2-Cyano esters
benzimidazole-1	Benzimidazole synthesis from an ortho-amino aniline and a carboxylic acid or ester.	Ortho-amino anilines	Carboxylic acids

benzimidazole-2	Benzimidazole synthesis from an ortho-amino aniline and an aldehyde.	Ortho-amino anilines	Aldehydes
benzimidazole-3	2-Chloro-benzimidazole formation from 2-amino anilines.	Ortho-amino anilines	
benzimidazole-4	Benzimidazole formation from 2-amino anilines.	Ortho-amino anilines	
benzofuran-1	Benzofuran synthesis from a 2-iodo phenol and an alkyne.	2-Iodo phenols	Monosubstituted alkynes
benzothiazole-1	Benzothiazole synthesis from a 2-thiol aniline and an aldehyde.	2-Mercapto anilines	Aldehydes
benzothiophene	Benzothiophene synthesis.	2-Iodo thioanisoles	Monosubstituted alkynes
benzoxazole-1	Benzoxazole synthesis from a 2-amino phenol and a carboxylic acid.	2-Amino phenols	Carboxylic acids
benzoxazole-2	Benzoxazole synthesis from a 2-amino phenol and an aryl aldehyde.	2-Amino phenols	Aryl aldehydes
carbamate	Carbamate sythesis from isocyanates and primary or secondary alcohols.	Isocyanates	Primary and secondary alcohols
carbamate-2	Carbamate sythesis from primary or secondary amines and primary or secondary alcohols.	Primary and secondary amines	Primary and secondary alcohols
castro-stephens-1	Castro-Stephens coupling: synthesis of indoles.	2-Iodo anilines	Monosubstituted alkynes
castro-stephens-2	Castro-Stephens coupling: synthesis of disubstituted acetylenes.	Aryl and vinyl halides	Monosubstituted alkynes
claisen-1	Crossed Claisen condensation: 1,3-diketone synthesis from an enolizable ketone and a non-enolizable ester.	Enolizable ketones	Non-enolizable esters

claisen-2	Claisen condensation: beta-keto ester synthesis from an enolizable ester and a non-enolizable ester.	Enolizable esters	Non-enolizable esters
conrad-limpach	Conrad-Limpach synthesis of substituted 4-hydroxyquinolines.	3-Ketoesters	Anilines
corey-chaykovsky-1	Corey-Chaykovsky synthesis of an epoxide from an aldehyde or a ketone.	Aldehydes and ketones	
corey-chaykovsky-2	Corey-Chaykovsky cyclopropane synthesis.	Unsaturated ketones	
coumarin	Pechmann condensation: Coumarin synthesis from a phenol and a beta-keto ester.	3-Ketoesters	Phenols
cyanohydrin	Cyanohydrin formation from an aldehyde.	Aldehydes	
decarboxylative_coupling-1	Aryl C-C bond formation from an aryl carboxylic acid and an aryl halide.	Aryl carboxylic acids	Aryl halides
diyne	Heterocoupling of two alkynes to form a 1,3-diyne (implicit formation of a haloalkyne).	Monosubstituted alkynes	Monosubstituted alkynes
epoxidation	Epoxidation of alkenes.		
erlenmeyer-plochl	Erlenmeyer-Plochl azlactone synthesis from aldehydes or ketones and N-acyl glycine.	Aldehydes and ketones	*N*-acyl glycines
ester-hydrolysis-1	Hydrolysis of an ester to the parent carboxylic acid.		
ester-hydrolysis-2	Hydrolysis of an ester to the parent alcohol.		
esterification-1	Esterification of an alcohol using an acid chloride.	Acid chlorides	Alcohols

esterification-2	Esterification of an alcohol and a carboxylic acid.	Carboxylic acids	Alcohols	
ether-1	Ullmann Aryl Ether Synthesis: Gives a biaryl ether from an aryl halide and a phenol.	Aryl halides	Phenols	
ether-2	Williamson Ether Synthesis: Gives an ether from a primary or secondary halide and a primary, secondary or (hetero)aryl alcohol.	Primary and secondary halides	Alcohols	
flavone	Flavone Synthesis: Gives a 2-aryl flavone from a 2-acyl phenol and an aryl acid chloride.	2-Keto phenols	Aryl acid chlorides	
friedel-crafts-1	Friedel-Crafts acylation.	Friedel-Crafts substrates	Acid chlorides	
furan-1	Feist-Benary furan synthesis from a beta-keto-ester and a 2-halo ketone.	3-Ketoesters	2-Halo ketones	
heck	Heck Reaction: coupling of activated alkene and aryl or vinyl halides.	Aryl and vinyl halides	Activated alkenes	
imidazole-1	Imidazole synthesis from an alpha-chloro aldehyde and an amidine.	2-Halo aldehydes	Amidines	
imidazole-2	Imidazole synthesis from an aryl aldehyde, 2-halo ketone and a primary amine.	Aryl aldehydes	2-Halo aryl ketones	Primary and secondary amines
imidazole-3	Synthesis of triaryl-imidazoles.	Diaryl diketones	Aryl aldehydes	
indole-1	Bischler-Mohlau Indole: Gives a 2-aryl Indole from an aniline and a 2-halo aryl ketone.	2-Halo aryl ketones	Anilines	
indole-2	Fischer Indole: Gives an indole from an aryl hydrazine and a ketone.	Aryl hydrazines	2-Unsubstituted ketones	

isocyanate-1	Isocyanate formation from primary amines.	Primary amines	
isoxazole-1	Claisen isoxazole synthesis from a beta-ketoester and hydroxylamine.	3-Ketoesters	
mitsunobu-1	Mitsunobu: Gives an aryl ether from an alcohol and a phenol.	Primary and secondary alcohols	Phenols
mitsunobu-2	Mitsunobu: Alkylation of a sulfonamide with a primary or secondary alcohol.	Primary and secondary alcohols	Primary and secondary sulfonamides
mitsunobu-3	Mitsunobu: N-alkylation of a tetrazole with a primary or secondary alcohol.	Primary and secondary alcohols	Tetrazoles-2H
mitsunobu-4	Mitsunobu: N-alkylation of a tetrazole with a primary or secondary alcohol.	Primary and secondary alcohols	Tetrazoles-1H
mitsunobu-5	Mitsunobu: Esterification from an alcohol and a carboxylic acid.	Primary and secondary alcohols	Carboxylic acids
miyaura-borylation	Miyaura borylation.	Aryl and vinyl halides	
negishi	Negishi: Aryl C-C coupling from two halides (one will be prepared as an organozinc).	Aryl, alkynyl and vinyl halides	Aryl and vinyl halides
oxadiazole-1	Oxadiazole synthesis from a nitrile (converted to an amidoxime) and a carboxylic acid.	Nitriles	Carboxylic acids
oxazole-1	Fischer Oxazole Synthesis: Generates an oxazole from an aldehyde and a cyanohydrin.	Aldehydes	Cyanohydrins
oxazole-2	2,4-disubstituted oxazole synthesis from monosubstituted alkynes and acetonitrile or proprionitrile (PhIO as oxygen source).	Monosubstituted alkynes	Acetonitrile or Proprionitrile

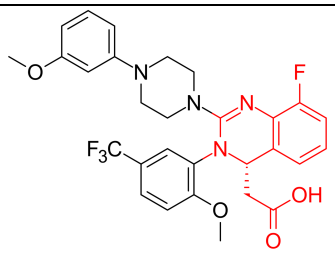
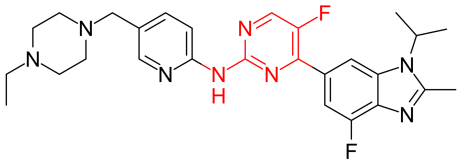
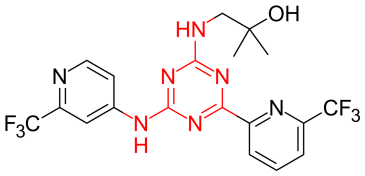
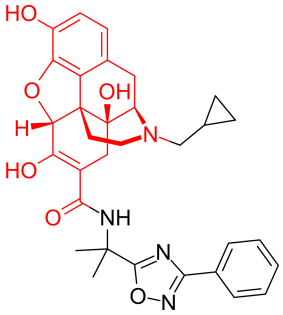
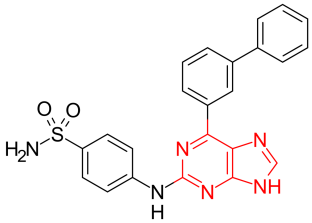
phthalazinone	Phthalazinone synthesis from 2-acylbenzoic acid and a hydrazine.	2-Acyl benzoic acids	Hydrazines	
pictet-spengler	Pictet-Spengler reaction.	Arylethyl amines	Aldehydes	
pomeranz-fritsch	Pomeranz-Fritsch synthesis of isoquinolines.	Aryl ketones and aldehydes		
pyrazole-1	Pyrazole synthesis from 1,3-diketones and hydrazines.	3-Keto ketones	Hydrazines	
pyridine-1	Hantzsch pyridine synthesis from two beta-ketoesters, an aldehyde, and ammonia (stepwise).	3-Keto esters	3-Keto esters	Aldehydes
pyridine-2	Krohnke pyridine synthesis from a 2-halo ketone, and an alfa,beta-unsaturated ketone.	2-Halo ketones	Unsaturated ketones	
pyridine-3	Guareshi-Thorpe pyridine synthesis from a beta-ketoester, and a cyanoacetic ester.	3-Keto esters	2-Cyanoacetic ester	
pyridone-1	Pyridone synthesis from cyanoacetamide and 1,3-diketones.	cyanoacetamides	3-Keto ketones	
pyrimidine-1	Pinner pyrimidine synthesis.	3-Keto ketones	Amidines	
pyrimidinedione-1	Pyrimidinedione synthesis from amino-acrylates and isocyanates.	Amino-acrylates	Isocyanates	
pyrimidinedione-2	Bicyclic pyrimidinedione synthesis from 2-aminobenzoates and isocyanates.	2-Aminobenzoates	Isocyanates	
pyrimidone-1	Biginelli 2-pyrimidone synthesis.	3-Keto esters	Aldehydes	
pyrimidone-2	Biginelli 2-mercaptopyrimidine synthesis.	3-Keto esters	Aldehydes	

pyrrole-1	Knorr pyrrole synthesis.	3-Keto esters	3-Keto esters	
pyrrole-2	Paal-Knorr pyrrole synthesis.	4-Keto ketones	Primary amines	
quinazoline-1	Quinazoline synthesis from a quinazolinone and a primary or secondary amine.	Quinazolinones	Primary and secondary amines	
quinazolinone-1	Quinazolinone synthesis from a 2-carboxy aniline, a primary amine, and an aldehyde.	2-Carboxy anilines	Primary and secondary amines	Aldehydes
quinazolinone-2	Niementowski quinazolinone synthesis from a 2-carboxy aniline, and a primary amide.	2-Carboxy anilines	Primary amides	
quinoline-1	Quinoline synthesis from a 2-amino benzaldehyde and a ketone.	2-Amino benzaldehydes	2-Unsubstituted ketones	
quinoline-2	Niementowski quinoline synthesis from a 2-carboxy aniline and a ketone.	2-Carboxy anilines	2-Unsubstituted ketones	
rubottom	Rubottom oxidation: alfa hydroxylation of an enolizable ketone *via* the corresponding silyl enol ether.	Enolizable ketones	-	
sonogashira	Sonogashira Coupling: Arylation of an alkyne with an aryl/vinyl halide.	Aryl and vinyl halides	Monosubstituted alkynes	
spiro-chromanone	Spiro-chromanone synthesis from an o-hydroxyacetophenone and a cyclohexanone.	*o*-Hydroxyacetophenones	Cyclohexanones	
stille	Stille Coupling: Aryl C-C coupling from an aryl or vinyl halide and an organostannane.	Aryl and vinyl halides	Organostannanes	
stille-2	Stille Carbonylative Coupling: Aryl C-C coupling with CO insertion from an aryl or vinyl halide and an organostannane.	Aryl and vinyl halides	Organostannanes	
sulfonamide-1	Sulfonamide synthesis from a sulfonyl chloride and a primary or secondary amine.	Sulfonyl chlorides	Primary and secondary amines	

sulfone-1	Oxidation of a thioether to a sulfone.			
sulfone-2	Oxidation of a sulfoxide to a sulfone.			
sulfoxide	Oxidation of a thioether to a sulfoxide.			
suzuki	Suzuki Coupling: Aryl C-C coupling from an aryl halide and an aryl/vinyl boronate.	Aryl and vinyl halides	Aryl and vinyl boronates	
tetrahydroindole-1	Tetrahydroindole synthesis from a cyclohexanone, 2-OH ketone and a primary amine.	Cyclohexanones	2-Hydroxy ketones	Primary amines
tetrazole-1	Tetrazole from nitrile and alkyl halide; first regioisomer.	Nitriles	Primary and secondary halides	
tetrazole-2	Tetrazole from nitrile and alkyl halide; second regioisomer.	Nitriles	Primary and secondary halides	
tetrazole-3	Synthesis of a terminal tetrazole from a nitrile.	Nitriles		
thiazole-1	Hantzsch thiazole synthesis from a thioamide and alpha-halo ketone.	Thioamides	2-Halo ketones	
thiazole-2	Hantzsch thiazole synthesis from thioformamide and an alpha-halo ketone.	2-Halo ketones		
thioether-1	Thioether synthesis from styrenes and thiols.	Styrenes	Thiols	
thioether-2	Thioether synthesis from halides and thiols.	Primary and secondary arylhalides	Thiols	
thiourea-1	Thiourea synthesis from thioisocyanates and primary or secondary amines.	Thioisocyanates	Primary and secondary amines	

urea-1	Urea synthesis from isocyanates and primary or secondary amines.	Isocyanates	Primary and secondary amines
urea-2	Urea synthesis from primary or secondary amines and primary or secondary amines.	Primary and secondary amines	Primary and secondary amines
vinyl-reduction	Reduction of vinyl groups.		
yamaguchi	Yamaguchi macrolactonization.		

Table S5. PathFinder Reaction-Based Enumeration of a Set of Drug Molecules Approved by the FDA in 2017 and Compound C73.[a]

Compound	Structure[b]	Target	# Enumerated Compounds[c]
Letermovir		Antiviral	1,058,608
Abemaciclib		Dual CDK4/6 inhibitor	9,220,785
Enasidenib		IDH2 allosteric inhibitor	4,077,689
Naldemedine		μ-, ∂-, κ-opioid receptor antagonist	1,390,474
Compound **C73**[1]		CDK2 inhibitor	7,430,158

[a]PathFinder was run with a maximum depth of 2 and the core of each compound was kept constant. If a reaction contained multiple reactants that did not contain the core, only one of the components was varied (*i.e.* Route-1-enumeration-1: Reactant A (core; kept constant) + Reactant B (kept constant) -> Product 1, then Product 1 + Reactant C (varied) -> Final Product; Route-1-enumeration-2: Reactant A (core; kept constant) + Reactant B (varied) -> Product 1, then Product 1 + Reactant C (kept constant) -> Final Product. [b] The cores that were kept constant are highlighted in red. [c]Total number of unique ligands generated using all available chemistry in the PathFinder tool while keeping the highlighted core constant.

FEP+ Supplemental Methods

FEP simulations were performed using FEP+.[2,3] All FEP simulations used 16 lambda windows, hydrogen mass repartitioning (enabling 4 fs timestep), SPC waters48,[4] and the OPLS3 force field.[5] Single edge FEP was run for 1 ns and full cycle closure FEP was run for 20 ns simulation length.

R-group single edge parameters: 1 ns, 1 reference compound: **C12**,[1] default mapping, 16 lambda windows, SPC, 5,255 compounds, custom core SMARTS (c1ccccc1-N-c2nc3ncnc3cn2)

R-group cycle closure parameter: 20 ns, 2 reference compounds: **C12** and **C48**,[1] default mapping, 16 lambda windows, SPC, 177 compounds, custom core SMARTS (c1ccccc1-N-c2nc3ncnc3cn2)

Core hops (all cycle closure): 20 ns, 2 reference compounds: **C12** and **C48**,[1] default mapping, 16 lambda windows, SPC, 725 compounds, custom core SMARTS (c1ccccc1-N-c2aaaaa2)

R-group plus sulfonamide compounds (all cycle closure): 20 ns, 2 reference compounds: **C13** and **C70**,[1] default mapping, 16 lambda windows, SPC, 80 compounds, custom core SMARTS (c1ccccc1-N-c2nc3ncnc3cn2)

Core hop plus sulfonamide compounds (all cycle closure): 20 ns, 2 reference compounds: **C13** and **C70**,[1] default mapping, 16 lambda windows, SPC, 80 compounds, custom core SMARTS (c1ccccc1-N-c2aaaaa2)

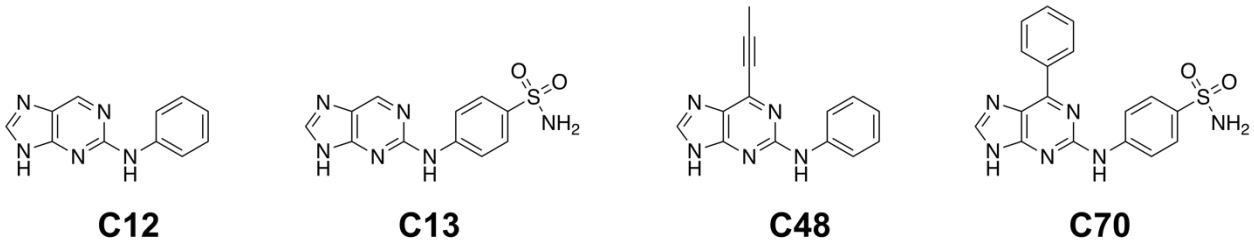


Figure S3. FEP reference compounds.[1]

AutoQSAR/DC details

AutoQSAR/DC runs an ensemble of Graph Convolutional Neural Networks, Dense Neural Networks on ECFP4 Fingerprints, Random Forests on ECFP4 Fingerprints, and XGBoost on ECFP4 Fingerprints. It does a random search of models for 4 hours and picks the top three sets of hyper parameters using 5 fold cross validation. It then ensembles models trained from all three sets of hyper parameters and takes the mean for the final prediction.[6,7]

Round	Model #	Model Type	Fingerprint Length	Dropout	Epochs
1	1	DNN	507 Chiral	0.15	50
1	2	DNN	1019	0.25	100
1	3	DNN	1019 Chiral	0.0	50
2	1	XGBoost	2046 Chiral	N/A	N/A
2	2	RF	2047 Chiral	N/A	N/A
2	3	RF	2048 Chiral	N/A	N/A
3	1	DNN	1029 Chiral	0.5	150
3	2	DNN	1027	0.5	50
3	3	DNN	1023	0.5	150

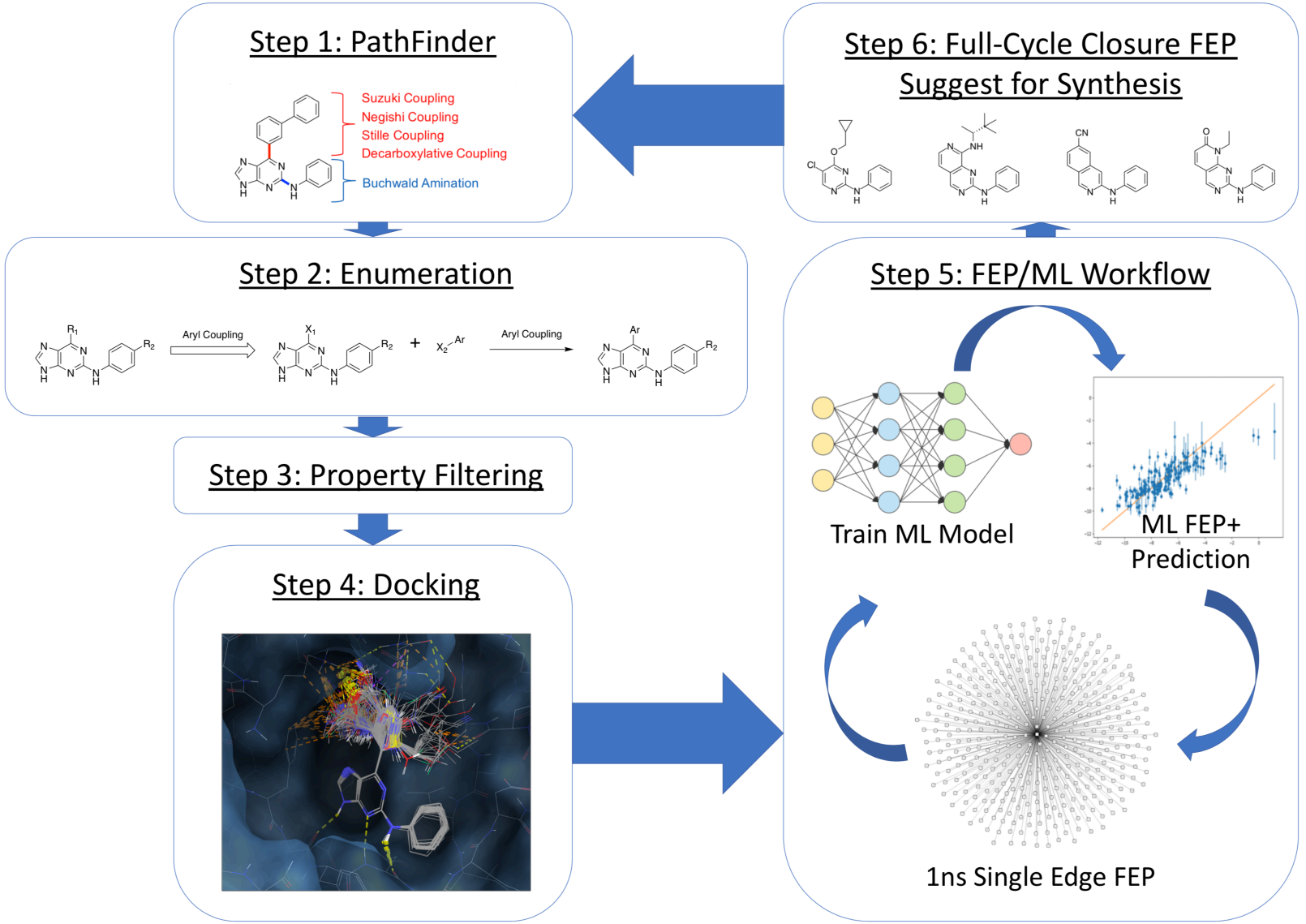


Figure S4. General overview of the workflow combining reaction-based enumeration, active learning, and free energy calculations.

Table S6. Predicted dG and error estimates for compounds described in Table 2, 5, 6, S1, S2, and S3.

ID	R = H		R = SO2NH2	
	Pred dG (kcal/mol)	Pred Err (kcal/mol)	Pred dG (kcal/mol)	Pred Err (kcal/mol)
2	-10.65	0.98	-11.85	1.24
4	-10.08	0.98	-10.73	1.26
5	-10.04	0.94	-10.16	1.26
6	-9.70	0.98	-11.40	1.28
7	-8.75	1.12	-10.21	1.24
22	-9.71	1.22	ND	ND
23	-9.63	1.65	ND	ND
24	-8.72	1.48	ND	ND
25	-10.77	1.51	ND	ND
27	ND	ND	-13.45	0.80
28	ND	ND	-12.66	1.09
29	ND	ND	-10.39	1.25
30	-10.60	1.56	ND	ND
31	ND	ND	-11.24	1.22
C12	-6.03	0.82	-6.61	0.91
C2	-7.69	0.84	-8.15	0.91
S1	-10.39	1.06	-11.55	1.20
S10	-8.54	1.08	-9.76	1.43
S11	-8.35	1.10	-9.28	1.32
S12	-8.32	1.02	-10.12	1.21
S13	-8.23	1.02	-9.48	1.46
S14	-8.04	1.01	-9.31	1.32
S15	-9.39	0.97	ND	ND
S16	-9.01	1.04	ND	ND
S17	-8.71	0.98	ND	ND
S18	-7.37	1.16	ND	ND
S19	-7.17	1.01	ND	ND
S2	-10.12	1.01	-11.92	1.30
S20	-8.46	1.08	ND	ND
S21	-8.36	1.01	ND	ND
S22	-8.01	1.12	ND	ND
S23	-9.42	1.02	-11.00	1.16
S24	-9.22	1.03	-11.21	1.17
S25	-8.84	1.09	-10.29	1.22
S26	-8.78	0.98	-10.38	1.15

S27	-8.25	1.00	ND	ND
S28	-7.13	1.12	ND	ND
S29	-6.51	0.99	ND	ND
S3	-9.97	1.09	-12.14	1.16
S30	-10.45	0.89	-10.62	1.27
S31	-9.89	0.89	-10.22	1.35
S32	-9.37	0.91	-10.42	1.23
S33	-9.32	0.91	ND	ND
S34	-9.21	0.96	ND	ND
S35	-9.03	0.93	ND	ND
S36	-8.73	1.08	ND	ND
S4	-9.81	1.11	-11.22	1.24
S5	-9.33	1.11	-10.41	1.38
S6	-9.16	1.13	-9.41	1.25
S7	-9.13	0.96	-9.59	1.46
S8	-9.06	1.04	-9.84	1.27
S9	-8.92	1.01	-11.58	1.17

References

(1) Coxon, C. R.; Anscombe, E.; Harnor, S. J.; Martin, M. P.; Carbain, B.; Golding, B. T.; Hardcastle, I. R.; Harlow, L. K.; Korolchuk, S.; Matheson, C. J.; et al. Cyclin-Dependent Kinase (CDK) Inhibitors: Structure-Activity Relationships and Insights into the CDK-2 Selectivity of 6-Substituted 2-Arylaminopurines. *J. Med. Chem.* **2017**, *60* (5), 1746–1767. https://doi.org/10.1021/acs.jmedchem.6b01254.

(2) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632. https://doi.org/10.1021/acs.accounts.7b00083.

(3) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; et al. 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. https://doi.org/10.1021/ja512751q.

(4) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271. https://doi.org/10.1021/j100308a038.

(5) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. https://doi.org/10.1021/acs.jctc.5b00864.

(6) Dixon, S. L.; Duan, J.; Smith, E.; Von Bargen, C. D.; Sherman, W.; Repasky, M. P. AutoQSAR: An Automated Machine Learning Tool for Best-Practice Quantitative Structure-Activity Relationship Modeling. *Future Med. Chem.* **2016**, *8* (15), 1825–1839. https://doi.org/10.4155/fmc-2016-0093.

(7) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chem. Sci.* **2018**, *9* (2). https://doi.org/10.1039/c7sc02664a.