Supporting Information for

Predicting Antimicrobial Properties of Lignin Derivatives

Ryan M. Kalinoski[a] *, Qing Shao[b], Jian Shi[a]*

[a]Biosystems and Agricultural Engineering, 128 C.E. Barnhart Building, University of Kentucky, Lexington, KY, 40506, USA
[b]Department of Chemical and Materials Engineering, 159 FPAT, University of Kentucky, Lexington, KY, 40506, USA

*Corresponding to: Dr. Ryan Kalinoski (rmka232@uky.edu), University of Kentucky; Phone: (859) 218-4330; Fax: (859) 257-5671
and
Dr. Jian Shi (j.shi@uky.edu); University of Kentucky; Phone: (859) 218-4321; Fax: (859) 257-5671

Table S1. Molecular descriptors calculated using RDKit and their descriptions; Selected descriptors used for the ChEMBL database QSAR model development.

Table S2. Selected descriptors used for the ChEMBL database QSAR model development. Descriptor meaning can be found in Supplementary Table S1.

Table S3. QSAR model performance for the B-All (*B. subtilis*) and E-All (*E. coli*) ChEMBL datasets using the different machine learning algorithms.

Table S4. QSAR model performance for the B-Sort (*B. subtilis*) and E-Sort (*E. coli*) ChEMBL datasets using the different machine learning algorithms.

Table S5. QSAR model performance for the B-Phenol (*B. subtilis*) and E-Phenol (*E. coli*) ChEMBL datasets using the different machine learning algorithms.

Table S6. Source articles that reported antimicrobial data (MIC) and converted pMIC values for phenolics that can be derived from lignin against both *B. subtilis* (B-Meta dataset) and *E. coli* (E-Meta dataset).

Table S7. The applicability domain (AD) limit value for each of the *B. subtilis* ChEMBL datasets and the respective Euclidian distances for each compound in the B-Meta dataset.

Table S8. The applicability domain (AD) limit value for each of the *E. coli* ChEMBL datasets and the respective Euclidian distances for each compound in the B-Meta dataset.

Supplementary Table S1: Molecular descriptors calculated using RDKit and their descriptions.

Descriptor Name	Description	Dimension	Extended class
BalabanJ	Balaban's J value for a molecule,Chem. Phys. Lett. 89:399-404 (1982).	2	Topological descriptors
BertzCT	A topological index meant to quantify "complexity" of molecules.J. Am. Chem. Soc. 103:3599-601 (1981).	2	Topological descriptors
Chi0	From equations (1),(9) and (10) of Rev. Comp. Chem. vol 2, 367-422, (1991)	2	Connectivity descriptors
Chi1	From equations (1),(11) and (12) of Rev. Comp. Chem. vol 2, 367-422, (1991)	2	Connectivity descriptors
Chi0v	From equations (5),(9) and (10) of Rev. Comp. Chem. vol 2, 367-422, (1991)	2	Connectivity descriptors
Chi1v	From equations (5),(11) and (12) of Rev. Comp. Chem. vol 2, 367-422, (1991)	2	Connectivity descriptors
Chi2v	From equations (5),(15) and (16) of Rev. Comp. Chem. vol 2, 367-422, (1991)	2	Connectivity descriptors
Chi3v	From equations (5),(15) and (16) of Rev. Comp. Chem. vol 2, 367-422, (1991)	2	Connectivity descriptors
Chi4v	From equations (5),(15) and (16) of Rev. Comp. Chem. vol 2, 367-422, (1991)	2	Connectivity descriptors
Chi0n	Similar to Hall Kier Chi0v, but uses nVal instead of valence This makes a big difference after we get out of the first row.Rev. Comput. Chem. 2:367-422 (1991).	2	Connectivity descriptors
Chi1n	Similar to Hall Kier Chi1v, but uses nVal instead of valence.Rev. Comput. Chem. 2:367-422 (1991).	2	Connectivity descriptors
Chi2n	Similar to Hall Kier Chi2v, but uses nVal instead of valence This makes a big difference after we get out of the first row.Rev. Comput. Chem. 2:367-422 (1991).	2	Connectivity descriptors
Chi3n	Similar to Hall Kier Chi3v, but uses nVal instead of valence This makes a big difference after we get out of the first row.Rev. Comput. Chem. 2:367-422 (1991).	2	Connectivity descriptors
Chi4n	Similar to Hall Kier Chi4v, but uses nVal instead of valence.This makes a big difference after we get out of the first row.Rev. Comput. Chem. 2:367-422 (1991).	2	Connectivity descriptors
EState_VSA1	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA2	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA3	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA4	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA5	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA6	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA7	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG	2	MOE-type descriptors

	paper).		
EState_VSA8	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA9	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA10	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
EState_VSA11	MOE-type descriptors using EState indices and surface area contributions (developed at RD, not described in the CCG paper).	2	MOE-type descriptors
ExactMolWt	The molecule's exact molecular weight.	2	Molecular property descriptors
FractionCSP3	The fraction of C atoms that are SP3 hybridized.	1	Constitutional descriptors
HallKierAlpha	The Hall-Kier alpha value for a molecule.Rev. Comput. Chem. 2:367-422 (1991).	2	Topological descriptors
HeavyAtomCount	Number of heavy atoms of a molecule.	1	Constitutional descriptors
HeavyAtomMolWt	The average molecular weight of the molecule ignoring hydrogens	1	Constitutional descriptors
Ipc	the information content of the coefficients of the characteristic polynomial of the adjacency matrix of a hydrogen-suppressed graph of a molecule.	2	Topological descriptors
Kappa1	Hall-Kier Kappa1 value	2	Topological descriptors
Kappa2	Hall-Kier Kappa2 value	2	Topological descriptors
Kappa3	Hall-Kier Kappa2 value	2	Topological descriptors
LabuteASA	Labute's Approximate Surface Area (ASA from MOE)	2	MOE-type descriptors
MolLogP	Wildman-Crippen LogP value.Wildman and Crippen JCICS 39:868-73 (1999)	2	Molecular property descriptors
MolMR	Wildman-Crippen MR value.Wildman and Crippen JCICS 39:868-73 (1999)	2	Molecular property descriptors
MolWt	The average molecular weight of the molecule	2	Molecular property descriptors
NHOHCount	Number of NHs or OHs	1	Constitutional descriptors
NOCount	Number of Nitrogens and Oxygens	1	Constitutional descriptors
NumAliphaticCarbocycles	The number of aliphatic (containing at least one non-aromatic bond) carbocycles for a molecule	1	Constitutional descriptors
NumAliphaticHeterocycles	The number of aliphatic (containing at least one non-aromatic bond) heterocycles for a molecule	1	Constitutional descriptors
NumAliphaticRings	The number of aliphatic (containing at least one non-aromatic bond) rings for a molecule	1	Constitutional descriptors

NumAromaticCarbocycles	The number of aromatic carbocycles for a molecule	1	Constitutional descriptors
NumAromaticHeterocycles	The number of aromatic heterocycles for a molecule	1	Constitutional descriptors
NumAromaticRings	The number of aromatic rings for a molecule	1	Constitutional descriptors
NumHAcceptors	Number of Hydrogen Bond Acceptors	1	Constitutional descriptors
NumHDonors	Number of Hydrogen Bond Donors	1	Constitutional descriptors
NumHeteroatoms	Number of Heteroatoms	1	Constitutional descriptors
NumRadicalElectrons	The number of radical electrons the molecule has (says nothing about spin state)	1	Constitutional descriptors
NumRotatableBonds	Number of Rotatable Bonds	1	Constitutional descriptors
NumSaturatedCarbocycles	The number of saturated carbocycles for a molecule	1	Constitutional descriptors
NumSaturatedHeterocycles	The number of saturated heterocycles for a molecule	1	Constitutional descriptors
NumSaturatedRings	The number of saturated rings for a molecule	1	Constitutional descriptors
NumValenceElectrons	The number of valence electrons the molecule has	1	Constitutional descriptors
PEOE_VSA1	MOE Charge VSA Descriptor 1 (-inf < x < -0.30)	2	MOE-type descriptors
PEOE_VSA2	MOE Charge VSA Descriptor 2 (-0.30 <= x < -0.25)	2	MOE-type descriptors
PEOE_VSA3	MOE Charge VSA Descriptor 3 (-0.25 <= x < -0.20)	2	MOE-type descriptors
PEOE_VSA4	MOE Charge VSA Descriptor 4 (-0.20 <= x < -0.15)	2	MOE-type descriptors
PEOE_VSA5	MOE Charge VSA Descriptor 5 (-0.15 <= x < -0.10)	2	MOE-type descriptors
PEOE_VSA6	MOE Charge VSA Descriptor 6 (-0.10 <= x < -0.05)	2	MOE-type descriptors
PEOE_VSA7	MOE Charge VSA Descriptor 7 (-0.05 <= x < 0.00)	2	MOE-type descriptors
PEOE_VSA8	MOE Charge VSA Descriptor 8 (0.00 <= x < 0.05)	2	MOE-type descriptors
PEOE_VSA9	MOE Charge VSA Descriptor 9 (0.05 <= x < 0.10)	2	MOE-type descriptors
PEOE_VSA10	MOE Charge VSA Descriptor 10 (0.10 <= x < 0.15)	2	MOE-type descriptors
PEOE_VSA11	MOE Charge VSA Descriptor 11 (0.15 <= x < 0.20)	2	MOE-type descriptors
PEOE_VSA12	MOE Charge VSA Descriptor 12 (0.20 <= x < 0.25)	2	MOE-type descriptors
PEOE_VSA13	MOE Charge VSA Descriptor 13 (0.25 <= x < 0.30)	2	MOE-type descriptors
PEOE_VSA14	MOE Charge VSA Descriptor 14 (0.30 <= x < inf)	2	MOE-type descriptors
RingCount	The number of rings for a molecule	1	Constitutional descriptors

SMR_VSA1	MOE MR VSA Descriptor 1 (-inf < x < 1.29)	2	MOE-type descriptors
SMR_VSA2	MOE MR VSA Descriptor 2 (1.29 <= x < 1.82)	2	MOE-type descriptors
SMR_VSA3	MOE MR VSA Descriptor 3 (1.82 <= x < 2.24)	2	MOE-type descriptors
SMR_VSA4	MOE MR VSA Descriptor 4 (2.24 <= x < 2.45)	2	MOE-type descriptors
SMR_VSA5	MOE MR VSA Descriptor 5 (2.45 <= x < 2.75)	2	MOE-type descriptors
SMR_VSA6	MOE MR VSA Descriptor 6 (2.75 <= x < 3.05)	2	MOE-type descriptors
SMR_VSA7	MOE MR VSA Descriptor 7 (3.05 <= x < 3.63)	2	MOE-type descriptors
SMR_VSA8	MOE MR VSA Descriptor 8 (3.63 <= x < 3.80)	2	MOE-type descriptors
SMR_VSA9	MOE MR VSA Descriptor 9 (3.80 <= x < 4.00)	2	MOE-type descriptors
SMR_VSA10	MOE MR VSA Descriptor 10 (4.00 <= x < inf)	2	MOE-type descriptors
SlogP_VSA1	MOE logP VSA Descriptor 1 (-inf < x < -0.40)	2	MOE-type descriptors
SlogP_VSA2	MOE logP VSA Descriptor 2 (-0.40 <= x < -0.20)	2	MOE-type descriptors
SlogP_VSA3	MOE logP VSA Descriptor 3 (-0.20 <= x < 0.00)	2	MOE-type descriptors
SlogP_VSA4	MOE logP VSA Descriptor 4 (0.00 <= x < 0.10)	2	MOE-type descriptors
SlogP_VSA5	MOE logP VSA Descriptor 5 (0.10 <= x < 0.15)	2	MOE-type descriptors
SlogP_VSA6	MOE logP VSA Descriptor 6 (0.15 <= x < 0.20)	2	MOE-type descriptors
SlogP_VSA7	MOE logP VSA Descriptor 7 (0.20 <= x < 0.25)	2	MOE-type descriptors
SlogP_VSA8	MOE logP VSA Descriptor 8 (0.25 <= x < 0.30)	2	MOE-type descriptors
SlogP_VSA9	MOE logP VSA Descriptor 9 (0.30 <= x < 0.40)	2	MOE-type descriptors
SlogP_VSA10	MOE logP VSA Descriptor 10 (0.40 <= x < 0.50)	2	MOE-type descriptors
SlogP_VSA11	MOE logP VSA Descriptor 11 (0.50 <= x < 0.60)	2	MOE-type descriptors
SlogP_VSA12	MOE logP VSA Descriptor 12 (0.60 <= x < inf)	2	MOE-type descriptors
TPSA	The polar surface area of a molecule based upon fragments	2	Molecular property descriptors
VSA_EState1	VSA EState Descriptor 1 (-inf < x < 4.78)	2	MOE-type descriptors
VSA_EState2	VSA EState Descriptor 2 (4.78 <= x < 5.00)	2	MOE-type descriptors
VSA_EState3	VSA EState Descriptor 3 (5.00 <= x < 5.41)	2	MOE-type descriptors
VSA_EState4	VSA EState Descriptor 4 (5.41 <= x < 5.74)	2	MOE-type descriptors

VSA_EState5	VSA EState Descriptor 5 (5.74 <= x < 6.00)	2	MOE-type descriptors
VSA_EState6	VSA EState Descriptor 6 (6.00 <= x < 6.07)	2	MOE-type descriptors
VSA_EState7	VSA EState Descriptor 7 (6.07 <= x < 6.45)	2	MOE-type descriptors
VSA_EState8	VSA EState Descriptor 8 (6.45 <= x < 7.00)	2	MOE-type descriptors
VSA_EState9	VSA EState Descriptor 9 (7.00 <= x < 11.00)	2	MOE-type descriptors
VSA_EState10	VSA EState Descriptor 10 (11.00 <= x < inf)	2	MOE-type descriptors
fr_Al_COO	Number of aliphatic carboxylic acids	1	Constitutional descriptors
fr_Al_OH	Number of aliphatic hydroxyl groups	1	Constitutional descriptors
fr_Al_OH_noTert	Number of aliphatic hydroxyl groups excluding tert-OH	1	Constitutional descriptors
fr_ArN	Number of N functional groups attached to aromatics	1	Constitutional descriptors
fr_Ar_COO	Number of Aromatic carboxylic acide	1	Constitutional descriptors
fr_Ar_N	Number of aromatic nitrogens	1	Constitutional descriptors
fr_Ar_NH	Number of aromatic amines	1	Constitutional descriptors
fr_Ar_OH	Number of aromatic hydroxyl groups	1	Constitutional descriptors
fr_COO	Number of carboxylic acids	1	Constitutional descriptors
fr_COO2	Number of carboxylic acids	1	Constitutional descriptors
fr_C_O	Number of carbonyl O	1	Constitutional descriptors
fr_C_O_noCOO	Number of carbonyl O, excluding COOH	1	Constitutional descriptors
fr_C_S	Number of thiocarbonyl	1	Constitutional descriptors
fr_HOCCN	Number of C(OH)CCN-Ctert-alkyl or C(OH)CCNcyclic	1	Constitutional descriptors
fr_Imine	Number of Imines	1	Constitutional descriptors
fr_NH0	Number of Tertiary amines	1	Constitutional descriptors
fr_NH1	Number of Secondary amines	1	Constitutional descriptors
fr_NH2	Number of Primary amines	1	Constitutional descriptors
fr_N_O	Number of hydroxylamine groups	1	Constitutional descriptors
fr_Ndealkylation1	Number of XCCNR groups	1	Constitutional descriptors
fr_Ndealkylation2	Number of tert-alicyclic amines (no heteroatoms, not quinine-like bridged N)	1	Constitutional descriptors

fr_Nhpyrrole	Number of H-pyrrole nitrogens	1	Constitutional descriptors
fr_SH	Number of thiol groups	1	Constitutional descriptors
fr_aldehyde	Number of aldehydes	1	Constitutional descriptors
fr_alkyl_carbamate	Number of alkyl carbamates (subject to hydrolysis)	1	Constitutional descriptors
fr_alkyl_halide	Number of alkyl halides	1	Constitutional descriptors
fr_allylic_oxid	Number of allylic oxidation sites excluding steroid dienone	1	Constitutional descriptors
fr_amide	Number of amides	1	Constitutional descriptors
fr_amidine	Number of amidine groups	1	Constitutional descriptors
fr_aniline	Number of anilines	1	Constitutional descriptors
fr_aryl_methyl	Number of aryl methyl sites for hydroxylation	1	Constitutional descriptors
fr_azide	Number of azide groups	1	Constitutional descriptors
fr_azo	Number of azo groups	1	Constitutional descriptors
fr_barbitur	Number of barbiturate groups	1	Constitutional descriptors
fr_benzene	Number of benzene rings	1	Constitutional descriptors
fr_benzodiazepine	Number of benzodiazepines with no additional fused rings	1	Constitutional descriptors
fr_bicyclic	Bicyclic	1	Constitutional descriptors
fr_diazo	Number of diazo groups	1	Constitutional descriptors
fr_dihydropyridine	Number of dihydropyridines	1	Constitutional descriptors
fr_epoxide	Number of epoxide rings	1	Constitutional descriptors
fr_ester	Number of esters	1	Constitutional descriptors
fr_ether	Number of ether oxygens (including phenoxy)	1	Constitutional descriptors
fr_furan	Number of furan rings	1	Constitutional descriptors
fr_guanido	Number of guanidine groups	1	Constitutional descriptors
fr_halogen	Number of halogens	1	Constitutional descriptors
fr_hdrzine	Number of hydrazine groups	1	Constitutional descriptors
fr_hdrzone	Number of hydrazone groups	1	Constitutional descriptors
fr_imidazole	Number of imidazole rings	1	Constitutional descriptors

fr_imide	Number of imide groups	1	Constitutional descriptors
fr_isocyan	Number of isocyanates	1	Constitutional descriptors
fr_isothiocyan	Number of isothiocyanates	1	Constitutional descriptors
fr_ketone	Number of ketones	1	Constitutional descriptors
fr_ketone_Topliss	Number of ketones excluding diaryl, a,b-unsat.	1	Constitutional descriptors
fr_lactam	Number of beta lactams	1	Constitutional descriptors
fr_lactone	Number of cyclic esters (lactones)	1	Constitutional descriptors
fr_methoxy	Number of methoxy groups -OCH3	1	Constitutional descriptors
fr_morpholine	Number of morpholine rings	1	Constitutional descriptors
fr_nitrile	Number of nitriles	1	Constitutional descriptors
fr_nitro	Number of nitro groups	1	Constitutional descriptors
fr_nitro_arom	Number of nitro benzene ring substituents	1	Constitutional descriptors
fr_nitro_arom_nonortho	Number of non-ortho nitro benzene ring substituents	1	Constitutional descriptors
fr_nitroso	Number of nitroso groups, excluding NO2	1	Constitutional descriptors
fr_oxazole	Number of oxazole rings	1	Constitutional descriptors
fr_oxime	Number of oxime groups	1	Constitutional descriptors
fr_para_hydroxylation	Number of para-hydroxylation sites	1	Constitutional descriptors
fr_phenol	Number of phenols	1	Constitutional descriptors
fr_phenol_noOrthoHbond	Number of phenolic OH excluding ortho intramolecular Hbond substituents	1	Constitutional descriptors
fr_phos_acid	Number of phosphoric acid groups	1	Constitutional descriptors
fr_phos_ester	Number of phosphoric ester groups	1	Constitutional descriptors
fr_piperdine	Number of piperdine rings	1	Constitutional descriptors
fr_piperzine	Number of piperzine rings	1	Constitutional descriptors
fr_priamide	Number of primary amides	1	Constitutional descriptors
fr_prisulfonamd	Number of primary sulfonamides	1	Constitutional descriptors
fr_pyridine	Number of pyridine rings	1	Constitutional descriptors
fr_quatN	Number of quarternary nitrogens	1	Constitutional descriptors

fr_sulfide	Number of thioether	1	Constitutional descriptors
fr_sulfonamd	Number of sulfonamides	1	Constitutional descriptors
fr_sulfone	Number of sulfone groups	1	Constitutional descriptors
fr_term_acetylene	Number of terminal acetylenes	1	Constitutional descriptors
fr_tetrazole	Number of tetrazole rings	1	Constitutional descriptors
fr_thiazole	Number of thiazole rings	1	Constitutional descriptors
fr_thiocyan	Number of thiocyanates	1	Constitutional descriptors
fr_thiophene	Number of thiophene rings	1	Constitutional descriptors
fr_unbrch_alkane	Number of unbranched alkanes of at least 4 members (excludes halogenated alkanes)	1	Constitutional descriptors
fr_urea	Number of urea groups	1	Constitutional descriptors
MaxAbsEStateIndex	Returns a tuple of EState indices for the molecule, Reference: Hall, Mohney and Kier. JCICS _31_ 76-81 (1991)	2	Topological descriptors
MaxAbsPartialCharge	Returns molecular charge descriptors	2	Topological descriptors
MaxEStateIndex	Returns a tuple of EState indices for the molecule, Reference: Hall, Mohney and Kier. JCICS _31_ 76-81 (1991)	2	Topological descriptors
MaxPartialCharge	Returns molecular charge descriptors	2	Topological descriptors
MinAbsEStateIndex	Returns a tuple of EState indices for the molecule, Reference: Hall, Mohney and Kier. JCICS _31_ 76-81 (1991)	2	Topological descriptors
MinAbsPartialCharge	Returns molecular charge descriptors	2	Topological descriptors
MinEStateIndex	Returns a tuple of EState indices for the molecule, Reference: Hall, Mohney and Kier. JCICS _31_ 76-81 (1991)	2	Topological descriptors
MinPartialCharge	Returns molecular charge descriptors	2	Topological descripto

Supplementary Table S2: Selected descriptors used for the ChEMBL database QSAR model development. Descriptor meaning can be found in Supplementary Table S1.

Bacillus subtilis			***Escherichia coli***		
B-All	**B-Sort**	**B-Phenol**	**E-All**	**E-Sort**	**E-Phenol**
MaxEStateIndex	MaxEStateIndex	MaxEStateIndex	MaxEStateIndex	MaxEStateIndex	MaxEStateIndex
MinEStateIndex	MinEStateIndex	MinEStateIndex	MinEStateIndex	MinEStateIndex	MinEStateIndex
MinAbsEStateIndex	MinAbsEStateIndex	MinAbsEStateIndex	MinAbsEStateIndex	MinAbsEStateIndex	MinAbsEStateIndex
qed	qed	qed	qed	qed	qed
MolWt	MolWt	MolWt	MolWt	MolWt	MolWt
MaxPartialCharge	MaxPartialCharge	MaxPartialCharge	MaxPartialCharge	MaxPartialCharge	MaxPartialCharge
MinPartialCharge	MinPartialCharge	MinPartialCharge	MinPartialCharge	MinPartialCharge	MinPartialCharge
FpDensityMorgan1	FpDensityMorgan1	FpDensityMorgan1	FpDensityMorgan1	FpDensityMorgan1	FpDensityMorgan1
BalabanJ	BalabanJ	BalabanJ	BalabanJ	BalabanJ	BalabanJ
Ipc	Ipc	Ipc	Ipc	HallKierAlpha	HallKierAlpha
Kappa3	PEOE_VSA10	PEOE_VSA10	Kappa3	Ipc	Ipc
PEOE_VSA11	PEOE_VSA11	PEOE_VSA12	PEOE_VSA11	PEOE_VSA1	PEOE_VSA10
PEOE_VSA13	PEOE_VSA12	PEOE_VSA13	PEOE_VSA13	PEOE_VSA10	PEOE_VSA11
PEOE_VSA14	PEOE_VSA13	PEOE_VSA14	PEOE_VSA3	PEOE_VSA11	PEOE_VSA12
PEOE_VSA3	PEOE_VSA14	PEOE_VSA2	PEOE_VSA4	PEOE_VSA12	PEOE_VSA13
PEOE_VSA4	PEOE_VSA2	PEOE_VSA5	PEOE_VSA5	PEOE_VSA13	PEOE_VSA14
PEOE_VSA5	PEOE_VSA3	PEOE_VSA6	PEOE_VSA6	PEOE_VSA14	PEOE_VSA2
PEOE_VSA6	PEOE_VSA5	PEOE_VSA7	PEOE_VSA8	PEOE_VSA2	PEOE_VSA3
PEOE_VSA8	PEOE_VSA6	PEOE_VSA8	PEOE_VSA9	PEOE_VSA3	PEOE_VSA5
PEOE_VSA9	PEOE_VSA7	PEOE_VSA9	SMR_VSA10	PEOE_VSA5	PEOE_VSA6
SMR_VSA10	PEOE_VSA8	SMR_VSA10	SMR_VSA2	PEOE_VSA6	PEOE_VSA7
SMR_VSA2	PEOE_VSA9	SMR_VSA4	SMR_VSA4	PEOE_VSA7	PEOE_VSA8
SMR_VSA4	SMR_VSA10	SMR_VSA5	SMR_VSA6	PEOE_VSA8	PEOE_VSA9
SMR_VSA6	SMR_VSA4	SMR_VSA6	SMR_VSA7	PEOE_VSA9	SMR_VSA10
SMR_VSA7	SMR_VSA6	SMR_VSA7	SMR_VSA9	SMR_VSA10	SMR_VSA6
SMR_VSA9	SMR_VSA7	SlogP_VSA1	SlogP_VSA10	SMR_VSA6	SMR_VSA7
SlogP_VSA10	SlogP_VSA1	SlogP_VSA3	SlogP_VSA11	SMR_VSA7	SMR_VSA9
SlogP_VSA11	SlogP_VSA3	SlogP_VSA4	SlogP_VSA12	SMR_VSA9	SlogP_VSA1
SlogP_VSA12	SlogP_VSA8	SlogP_VSA8	SlogP_VSA4	SlogP_VSA1	SlogP_VSA3
SlogP_VSA4	EState_VSA2	EState_VSA2	SlogP_VSA7	SlogP_VSA3	SlogP_VSA8
SlogP_VSA7	EState_VSA3	EState_VSA3	SlogP_VSA8	SlogP_VSA8	EState_VSA11
SlogP_VSA8	EState_VSA4	EState_VSA4	EState_VSA11	EState_VSA11	EState_VSA2
EState_VSA11	EState_VSA5	EState_VSA5	EState_VSA3	EState_VSA2	EState_VSA3
EState_VSA3	EState_VSA6	EState_VSA6	EState_VSA4	EState_VSA3	EState_VSA4
EState_VSA4	EState_VSA7	EState_VSA7	EState_VSA5	EState_VSA4	EState_VSA5

EState_VSA5	EState_VSA8	EState_VSA8	EState_VSA6	EState_VSA5	EState_VSA6
EState_VSA6	EState_VSA9	EState_VSA9	EState_VSA7	EState_VSA6	EState_VSA7
EState_VSA7	VSA_EState2	VSA_EState6	EState_VSA8	EState_VSA7	EState_VSA8
EState_VSA8	VSA_EState4	VSA_EState7	EState_VSA9	EState_VSA8	EState_VSA9
EState_VSA9	VSA_EState5	VSA_EState8	VSA_EState1	EState_VSA9	VSA_EState2
VSA_EState1	VSA_EState6	NumAliphaticCarbocycles	VSA_EState4	VSA_EState2	VSA_EState4
VSA_EState3	VSA_EState7	NumAliphaticHeterocycles	VSA_EState7	VSA_EState4	VSA_EState5
VSA_EState4	VSA_EState8	NumAliphaticRings	VSA_EState8	VSA_EState5	VSA_EState6
VSA_EState5	VSA_EState9	NumAromaticCarbocycles	VSA_EState9	VSA_EState6	VSA_EState7
VSA_EState7	FractionCSP3	NumAromaticHeterocycles	FractionCSP3	VSA_EState7	VSA_EState8
VSA_EState8	NumAliphaticHeterocycles	NumSaturatedHeterocycles	NumAliphaticCarbocycles	VSA_EState8	NumAliphaticHeterocycles
VSA_EState9	NumAromaticHeterocycles	NumSaturatedRings	NumAliphaticHeterocycles	VSA_EState9	NumAromaticHeterocycles
FractionCSP3	RingCount	MolLogP	NumAromaticHeterocycles	NumAliphaticHeterocycles	NumRotatableBonds
NumAliphaticCarbocycles	MolLogP	fr_Al_COO	RingCount	NumAromaticHeterocycles	RingCount
NumAliphaticHeterocycles	fr_Al_COO	fr_Al_OH	MolLogP	RingCount	MolLogP
NumAromaticHeterocycles	fr_Ar_COO	fr_Ar_COO	fr_Al_COO	MolLogP	fr_Al_COO
RingCount	fr_aldehyde	fr_aldehyde	fr_Al_OH	fr_Al_COO	fr_Ar_COO
MolLogP	fr_allylic_oxid	fr_allylic_oxid	fr_ArN	fr_Ar_COO	fr_Ar_OH
fr_Al_COO	fr_aryl_methyl	fr_aryl_methyl	fr_Ar_COO	fr_Ar_OH	fr_aldehyde
fr_Al_OH	fr_bicyclic	fr_bicyclic	fr_Ar_NH	fr_aldehyde	fr_allylic_oxid
fr_ArN	fr_epoxide	fr_epoxide	fr_Ar_OH	fr_allylic_oxid	fr_aryl_methyl
fr_Ar_COO	fr_furan	fr_furan	fr_COO	fr_aryl_methyl	fr_bicyclic
fr_Ar_NH	fr_ketone	fr_ketone	fr_C_S	fr_bicyclic	fr_epoxide
fr_Ar_OH	fr_ketone_Topliss	fr_lactone	fr_HOCCN	fr_epoxide	fr_ester
fr_COO	fr_lactone	fr_para_hydroxylation	fr_Imine	fr_ester	fr_furan
fr_C_S	fr_para_hydroxylation	fr_unbrch_alkane	fr_NH0	fr_furan	fr_ketone
fr_HOCCN	fr_term_acetylene		fr_N_O	fr_ketone	fr_ketone_Topliss
fr_Imine			fr_Ndealkylation1	fr_ketone_Topliss	fr_lactone
fr_NH0			fr_Ndealkylation2	fr_lactone	fr_para_hydroxylation
fr_N_O			fr_SH	fr_para_hydroxylation	fr_term_acetylene
fr_Ndealkylation1			fr_aldehyde	fr_term_acetylene	fr_unbrch_alkane
fr_Ndealkylatio			fr_alkyl_carbam	fr_unbrch_alkan	

n2			ate	e	
fr_SH			fr_alkyl_halide		
fr_aldehyde			fr_allylic_oxid		
fr_alkyl_carbamate			fr_amidine		
fr_alkyl_halide			fr_aniline		
fr_allylic_oxid			fr_aryl_methyl		
fr_amidine			fr_azide		
fr_aniline			fr_azo		
fr_aryl_methyl			fr_barbitur		
fr_azide			fr_bicyclic		
fr_azo			fr_dihydropyridine		
fr_barbitur			fr_epoxide		
fr_bicyclic			fr_ester		
fr_dihydropyridine			fr_ether		
fr_epoxide			fr_furan		
fr_ester			fr_halogen		
fr_ether			fr_hdrzine		
fr_furan			fr_hdrzone		
fr_guanido			fr_imidazole		
fr_halogen			fr_imide		
fr_hdrzine			fr_isothiocyan		
fr_hdrzone			fr_ketone		
fr_imidazole			fr_lactam		
fr_imide			fr_lactone		
fr_isothiocyan			fr_methoxy		
fr_ketone			fr_morpholine		
fr_lactam			fr_nitrile		
fr_lactone			fr_nitro		
fr_methoxy			fr_nitroso		
fr_morpholine			fr_oxazole		
fr_nitrile			fr_oxime		
fr_nitro			fr_para_hydroxylation		
fr_nitroso			fr_phos_acid		
fr_oxazole			fr_piperdine		
fr_oxime			fr_piperzine		
fr_para_hydroxylation			fr_priamide		
fr_phos_acid			fr_pyridine		
fr_piperdine			fr_quatN		
fr_piperzine			fr_sulfide		
fr_priamide			fr_sulfonamd		

fr_pyridine			fr_sulfone		
fr_quatN			fr_term_acetylene		
fr_sulfide			fr_tetrazole		
fr_sulfonamd			fr_thiazole		
fr_sulfone			fr_thiocyan		
fr_term_acetylene			fr_thiophene		
fr_tetrazole			fr_unbrch_alkane		
fr_thiazole			fr_urea		
fr_thiocyan					
fr_thiophene					
fr_unbrch_alkane					
fr_urea					

Table S3: QSAR model performance for the B-All (*B. subtilis*) and E-All (*E. coli*) ChEMBL datasets using the different machine learning algorithms. Measured by average coefficient of determination (R^2) and root mean square error (RMSE) for both the training and test sets, where values are mean±SE (n=3). Each dataset was split into random test and train sets three different times to obtain the average performance score. The number of compounds, selected descriptors, and number of principle components used to develop models can be found in Table 1 and Supplementary Table S2 (Additional File 2).

	B-All				**E-All**			
	Test		**Train**		**Test**		**Train**	
Model	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
SVM	0.64 ± 0.005	0.66 ± 0.005	0.68 ± 0.001	0.50 ± 0.001	0.63 ± 0.004	0.65 ± 0.005	0.72 ± 0.002	0.52 ± 0.002
RF	0.65 ± 0.002	0.61 ± 0.006	0.95 ± 0.000	0.24 ± 0.001	**0.69 ± 0.004**	**0.62 ± 0.002**	**0.95 ± 0.000**	**0.24 ± 0.001**
KNN	**0.69 ± 0.008**	**0.58 ± 0.009**	**0.86 ± 0.001**	**0.39 ± 0.002**	0.68 ± 0.005	0.68 ± 0.006	0.71 ± 0.000	0.41 ± 0.000
DT	0.39 ± 0.027	0.81 ± 0.028	0.47 ± 0.008	0.75 ± 0.008	0.46 ± 0.004	0.82 ± 0.004	0.65 ± 0.004	0.65 ± 0.003
NN	0.63 ± 0.009	0.63 ± 0.004	0.83 ± 0.002	0.42 ± 0.002	0.63 ± 0.008	0.68 ± 0.003	0.87 ± 0.014	0.40 ± 0.021

Table S4: QSAR model performance for the B-Sort (*B. subtilis*) and E-Sort (*E. coli*) ChEMBL datasets using the different machine learning algorithms. Measured by average coefficient of determination (R^2) and root mean square error (RMSE) for both the training and test sets, where values are mean±SE (n=3). Each dataset was split into random test and train sets three different times to obtain the average performance score. The number of compounds, selected descriptors, and number of principle components used to develop models can be found in Table 1 and Supplementary Table S2 (Additional File 2).

	B-Sort				**E-Sort**			
	Test		**Train**		**Test**		**Train**	
Model	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
SVM	0.49 ± 0.015	0.63 ± 0.015	0.79 ± 0.013	0.42 ± 0.005	0.41 ± 0.022	0.59 ± 0.024	0.75 ± 0.004	0.53 ± 0.001
RF	0.45 ± 0.038	0.65 ± 0.032	0.92 ± 0.002	0.26 ± 0.003	0.38 ± 0.023	0.73 ± 0.025	0.92 ± 0.000	0.28 ± 0.001
KNN	0.35 ± 0.028	0.71 ± 0.014	0.78 ± 0.006	0.44 ± 0.006	**0.49 ± 0.067**	**0.79 ± 0.017**	**0.69 ± 0.007**	**0.42 ± 0.007**
DT	0.21 ± 0.026	0.78 ± 0.008	0.29 ± 0.004	0.78 ± 0.001	0.00 ± 0.073	0.94 ± 0.040	0.40 ± 0.009	0.78 ± 0.011
NN	**0.71 ± 0.014**	**0.49 ± 0.005**	**0.79 ± 0.032**	**0.41 ± 0.036**	0.18 ± 0.041	0.85 ± 0.043	0.88 ± 0.051	0.32 ± 0.095

Table S5: QSAR model performance for the B-Phenol (*B. subtilis*) and E-Phenol (*E. coli*) ChEMBL datasets using the different machine learning algorithms. Measured by average coefficient of determination (R^2) and root mean square error (RMSE) for both the training and test sets, where values are mean±SE (n=3). Each dataset was split into random test and train sets three different times to obtain the average performance score. The number of compounds, selected descriptors, and number of principle components used to develop models can be found in Table 1 and Supplementary Table S2 (Additional File 2).

	B-Phenol				**E-Phenol**			
	Test		**Train**		**Test**		**Train**	
Model	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
SVM	0.55 ± 0.023	0.65 ± 0.113	0.67 ± 0.187	0.31 ± 0.001	0.22 ± 0.011	0.87 ± 0.040	0.69 ± 0.005	0.56 ± 0.003
RF	0.50 ± 0.023	0.57 ± 0.033	0.93 ± 0.001	0.22 ± 0.002	0.17 ± 0.049	0.89 ± 0.064	0.89 ± 0.002	0.33 ± 0.007
KNN	**0.57 ± 0.007**	**0.59 ± 0.036**	**0.63 ± 0.005**	**0.42 ± 0.007**	**0.38 ± 0.019**	**0.75 ± 0.041**	**0.53 ± 0.015**	**0.76 ± 0.002**
DT	0.49 ± 0.025	0.63 ± 0.015	0.43 ± 0.010	0.62 ± 0.004	0.02 ± 0.065	1.02 ± 0.053	0.17 ± 0.016	0.90 ± 0.015
NN	0.51 ± 0.052	0.58 ± 0.072	0.80 ± 0.046	0.37 ± 0.049	0.00 ± 0.085	1.04 ± 0.094	0.88 ± 0.061	0.30 ± 0.117

Table S6: Source articles that reported antimicrobial data (MIC) and converted pMIC values for phenolics that can be derived from lignin against both *B. subtilis* (B-Meta dataset) and *E. coli* (E-Meta dataset).

Compound	Organism	MIC (ug/ml)	pMIC	Source
caffeic acid	B. subtilis	720.64	3.6	Barber, McConnell (8)
caffeic acid	B. subtilis	600	3.5	Taguri, Tanaka (33)
coniferaldehyde	B. subtilis	712.72	3.6	Barber, McConnell (8)
coniferyl alcohol	B. subtilis	1441	3.9	Barber, McConnell (8)
eugenol	B. subtilis	656.8	3.6	Barber, McConnell (8)
ferulic acid	B. subtilis	388.36	3.3	Barber, McConnell (8)
p-coumaraldehyde	B. subtilis	296.32	3.3	Barber, McConnell (8)
p-coumaric acid	B. subtilis	328.0875	3.3	Barber, McConnell (8)
p-coumaryl alcohol	B. subtilis	1201.396	3.9	Barber, McConnell (8)
pyrocatechol	B. subtilis	533	3.7	Taguri, Tanaka (33)
sinapaldehyde	B. subtilis	1664.589	3.9	Barber, McConnell (8)
sinapic acid	B. subtilis	448.42	3.3	Barber, McConnell (8)
sinapyl alcohol	B. subtilis	1682	3.9	Barber, McConnell (8)
gallic acid	B. subtilis	1600	3.97	Taguri, Tanaka (33)
protocatechuic acid	B. subtilis	2667	4.2	Taguri, Tanaka (33)
pyrogallol	B. subtilis	267	3.3	Taguri, Tanaka (33)
resveratrol	B. subtilis	2667	4.07	Taguri, Tanaka (33)
1-(4-hydroxy-3-methoxyphenyl)-2-propanone	E. coli	375	3.3	Zemek, Košíková (12)
2,3-bis(a-hydroxyvanillyl)-l,4-butanediol	E. coli	375	2.99	Zemek, Košíková (12)
4-hydroxy-3-methoxy-b-hydroxy-propiophenonoe	E. coli	375	3.3	Zemek, Košíková (12)
caffeic acid	E. coli	1441.28	3.9	Barber, McConnell (8)
caffeic acid	E. coli	2667	4.	Taguri, Tanaka (33)
coniferaldehyde	E. coli	356.36	3.3	Barber, McConnell (8)
coniferylalcohol	E. coli	375	3.3	Zemek, Košíková (12)
dehydrodiferulic acid	E. coli	150	2.6	Zemek, Košíková (12)
dehydrodiisoeugenol	E. coli	180	2.7	Zemek, Košíková (12)
di-o-acetylpinoresinol	E. coli	150	2.6	Zemek, Košíková (12)
eugenol	E. coli	656.8	3.6	Barber, McConnell (8)
eugenol	E. coli	3000	4.26	Zemek, Košíková (12)
ferulic acid	E. coli	388.36	3.30	Barber, McConnell (8)
ferulic acid	E. coli	375	3.3	Zemek, Košíková (12)
isoeugenol	E. coli	100	2.8	Zemek, Košíková (12)
p-coumaraldehyde	E. coli	296.32	3.3	Barber, McConnell (8)
p-coumaric acid	E. coli	328.0875	3.3	Barber, McConnell (8)
p-coumaryl alcohol	E. coli	1201	3.9	Barber, McConnell (8)
pyrocatechol	E. coli	533	3.7	Taguri, Tanaka (33)
sinapaldehyde	E. coli	832.2944	3.6	Barber, McConnell (8)
sinapic acid	E. coli	448.42	3.3	Barber, McConnell (8)

sinapyl alcohol	E. coli	1682	3.9	Barber, McConnell (8)
syringaldehyde	E. coli	375	3.3	Zemek, Košíková (12)
gallic acid	E. coli	600	3.55	Taguri, Tanaka (33)
protocatechuic acid	E. coli	2667	4.2	Taguri, Tanaka (33)
pyrogallol	E. coli	256	3.3	Cynthia, Florence (34)
pyrogallol	E. coli	83	2.8	Taguri, Tanaka (33)
resveratrol	E. coli	3200	4.1	Taguri, Tanaka (33)
p-hydroxy benzoic acid	E. coli	400	3.46	Aziz, Farag (35)
vanillic acid	E. coli	400	3.4	Aziz, Farag (35)

Table S7: The applicability domain (AD) limit value for each of the *B. subtilis* ChEMBL datasets and the respective Euclidian distances for each compound in the B-Meta dataset. Where True (T) or False (F) indicates if each compound lies within the AD of the each ChEMBL dataset.

	B-All: Limit 13.95		**B-Sort: Limit 9.73**		**B-Phenol: Limit 9.75**	
Compound	**Distance**	**T/F**	**Distance**	**T/F**	**Distance**	**T/F**
Caffeic acid	37.48	F	20.16	F	17.65	F
Coniferaldehyde	35.31	F	20.22	F	22.17	F
Coniferyl alcohol	30.53	F	16.28	F	16.40	F
Eugenol	35.56	F	18.04	F	18.82	F
Ferulic acid	36.93	F	21.64	F	20.23	F
Gallic Acid	39.74	F	18.74	F	16.02	F
Protocatechuic acid	32.41	F	16.97	F	14.53	F
Pyrocatechol	26.42	F	15.33	F	14.15	F
Pyrogallol	32.13	F	15.33	F	14.08	F
Resveratrol	40.74	F	22.65	F	21.93	F
Sinapaldehyde	41.58	F	22.93	F	21.52	F
Sinapic acid	43.76	F	27.10	F	20.81	F
Sinapyl alcohol	36.09	F	20.85	F	18.25	F
p-Coumaraldehyde	35.08	F	20.85	F	22.74	F
p-Coumaric acid	38.10	F	20.33	F	18.98	F
p-Coumaryl alcohol	28.69	F	16.93	F	15.97	F

Table S8: The applicability domain (AD) limit value for each of the *E. coli* ChEMBL datasets and the respective Euclidian distances for each compound in the B-Meta dataset. Where True (T) or False (F) indicates if each compound lies within the AD of the each ChEMBL dataset.

	E-All: Limit 13.95		**E-Sort: Limit 9.77**		**E-Phenol: Limit 9.71**	
Compound	**Distance**	**T/F**	**Distance**	**T/F**	**Distance**	**T/F**
1-(4-hydroxy-3-methoxyphenyl)-2-propanone	24.47	F	10.50	F	11.11	F
2,3-Bis(a-hydroxyvanillyl)-l,4-butanediol	47.50	F	16.79	F	16.38	F
4-Hydroxy-3-methoxy-B-hydroxy-propiophenonoe	27.64	F	10.76	F	10.77	F
Caffeic acid	34.22	F	17.03	F	14.54	F
Coniferaldehyde	44.11	F	17.57	F	17.47	F
Coniferyl alcohol	30.11	F	13.71	F	12.88	F
Dehydrodiferulic acid	47.47	F	24.42	F	21.26	F
Dehydrodiisoeugenol	47.74	F	22.51	F	20.95	F
Di-O-acetylpinoresinol	52.39	F	18.86	F	18.87	F
Eugenol	37.00	F	13.11	F	13.22	F
Ferulic acid	35.01	F	16.36	F	14.47	F
Ferulic acid	35.01	F	16.36	F	14.47	F
Gallic Acid	33.16	F	18.92	F	16.70	F
Protocatechuic acid	27.79	F	15.74	F	14.01	F
Pyrocatechol	26.27	F	15.69	F	15.04	F
Pyrogallol	28.18	F	14.92	F	14.43	F
Resveratrol	36.45	F	16.71	F	16.77	F
Sinapaldehyde	46.28	F	17.81	F	17.07	F
Sinapic acid	40.47	F	18.58	F	15.58	F
Sinapyl alcohol	32.90	F	14.10	F	12.15	F
Vanillic acid	27.31	F	15.64	F	13.01	F
isoeugenol	34.04	F	13.21	F	12.52	F
p-Coumaraldehyde	44.91	F	17.97	F	18.20	F
p-Coumaric acid	34.51	F	16.83	F	15.33	F
p-Coumaryl alcohol	29.86	F	14.31	F	14.06	F
p-hydroxy benzoic acid	25.91	F	15.22	F	13.49	F
syringaldehyde	40.89	F	15.61	F	14.19	F