Supporting Information

MolFind: A Software Package Enabling HPLC/MS Based Identification of Chemical Structures

Lochana C. Menikarachchi[†], Shannon Cawley[†], Dennis W. Hill[†], L. Mark Hall[‡], Lowell Hall[§], Steven Lai[¬], Janine Wilder[†], and David F. Grant[*,†]

[†]Department of Pharmaceutical Sciences, University of Connecticut, Storrs, Connecticut, United States

[‡] Hall Associates Consulting, Quincy, Massachusetts, United States

[§]Department of Chemistry, Eastern Nazarene College, Quincy, Massachusetts, United States

[¬]Waters Corporation, Beverly, Massachusetts, United States

[*]Corresponding author: Dr. David F. Grant, University of Connecticut, 69 N Eagleville Rd, Storrs, CT 06269. Phone (860)486-4265, Fax (860)486-5792, email: david.grant@uconn.edu

Experimental Methods

Chemicals

Trifluoroacetic acid, LC/MS grade methanol (Optima®), mefenamic acid, and 2,3,3a,4,5,7,8,9,9a,9b-decahydro-3-hydroxy-3a-methyl-1h-benz(e)inden-7-one were purchased from Sigma-Aldrich (St. Louis, MO). HPLC grade acetonitrile was purchased from Fisher Scientific (Hampton, NH) and HPLC grade heptafluorobuteric acid was purchased from Thermo Scientific (Rockford, IL). Nitromethane, nitroethane, *n*-nitropropane, *n*-nitrobutane, *n*-nitropentane, and *n*-nitrohexane were purchased from Aldrich (St. Louis, MO). A previously described method[1] was used to synthesize the remaining *n*-nitroalkanes; *n*-nitroheptane, *n*-nitrooctane, *n*-nitrononane, and *n*-nitrodecane. Reagent grade water was generated by a Barnstead Diamond® reverse-osmosis water purification system. Detailed information describing the compounds examined in this study, including their sources, can be found in table S1.

Measurement of Retention Indices

The retention times of test compounds used in this study were determined on an Agilent 1100 HPLC system using a Zorbax Stable Bond, C-18, 1.0 mm x 150 mm (3.5 µm particle size) column at a flow rate of 75 µL/min with a linear solvent gradient from 0.76 mM heptofluorobutyric acid (HFBA) to 0.76 mM HFBA in 10% (v/v) water/acetonitrile. A Zorbax 1.0 mm x 17 mm (5 µm particle size) pre-column (Agilent, Santa Clara, CA) was used prior to the analytical column. The homologous series of n-nitroalkane retention index reference compounds were detected on an Agilent diode array detector at 210 nm and the test compounds were detected on a Qtof-2 (Waters, Beverly, MA) mass spectrometer using electrospray ionization in the positive ion mode and a mass range of 50 to 1000 Da. Ethisterone or uracil was

analyzed on both the uv detector and the mass spectrometer with each batch of analysis to determine the retention offset between the nitroalkane reference standards and the test compounds. Test mixture A (2-aminopyrimidine, aminoquinoline, p-anisidine, benzylamine, brucine, di-sec-butylamine, N-carbobenzyloxy-L-aspartic acid, 2-chloropyrazine, 2, 4, 6-collidine, cyclohexylamine, indole, 1-phenyl-3-pyrazolidinone, probenecid, quinine, and yohimbine) or B (2-aminopyrimidine, corticosterone, N, N-dimethylbenzamide, indomethacin, nicotinic acid, phenylbutazone, probenecid, tetraethylthiuram disulfide, thonzonium bromide) was analyzed with each batch of analysis to monitor retention index reproducibility. As previously described[2], the retention time of each n-nitroalkane was assigned a retention index value 100 times the number of carbons in the respective reference compound. The retention indices of test compounds were assigned a value proportional to their retention times relative to the retention times of the n-nitroalkanes. The retention index of test compounds that eluted during the isocratic portion of the mobile phase were calculated by the interpolation of the log of the retention time of the test compounds relative to the log of the retention times of the n-nitroalkanes that eluted before and after the test compound. The relationships between the retention indices and retention times of the n-nitroalkanes that eluted during the gradient portion of the mobile phase were fitted to a 4^{th} order polynomial equation. This equation was then used to calculate the retention indices of test compounds eluting during the linear gradient.

The model developed to predict RI values is specific for the HPLC system described in this study. The normalization of compound retention relative to a set of reference standards will correct retention times due to minor variability in mobile phase composition, column packing and chemistry, and solvent gradient profile. Since a reverse phase mode of retention was used for this system, very hydrophilic compounds will have very little retention and may elute in the

void volume. The predicted retention indices for these compounds will not differentiate between compounds of this type, but will differentiate between these compounds and other more hydrophobic candidate compounds with the same monoisotopic mass. An appropriate HPLC system, such as the use of a HILIC column in the normal phase mode, would need to be developed for the separation of very polar compounds and a new predictive retention model would have to be developed for this system.

Measurement of $Ecom_{50}$ Values

Experimental $Ecom_{50}$ values were obtained on the Qtof-2 mass spectrometer. The cone potential was set at 20 V and the collision energy was varied to determine experimental $Ecom_{50}$ values. All other conditions remained identical to those described for the RI study. A Hamilton syringe pump (Reno, NV) was used to introduce samples into the positive ion electrospray source at a flow rate of 5 µL/min. Stock solutions for each of the compounds were prepared at a concentration of 10 µmol/mL in an appropriate solvent based on solubility determined by predicted LogP values provided in PubChem. Test solutions for this study were prepared using the stock solutions of four compounds, each separated in mass by at least 6 Da, and diluted to a final concentration of 10 nmol/mL, in 0.01% (v/v) trifluoroacetic acid in acetonitrile/water (1:1, v/v). If a sample appeared to saturate the micro-channel plate (MCP) detector on the mass spectrometer, it was diluted accordingly. A few samples were prepared and run individually due to their instability or potential interaction with other compounds in their respective mixtures. The molecular ion of each protonated compound was isolated and subjected to CID in the collision cell using argon as the collision gas. The CID spectra of each compound were determined at odd collision energies from 1 eV to 39 eV. The spectra for each test compound at each collision energy were co-added and processed by smoothing, centering and eliminating noise below a 1%

threshold using Masslynx® software. The collision energy at 50% survival yield (CE_{50}) was calculated and converted to the center of mass energy at 50% survival yield ($Ecom_{50}$) as described[3]. If early parent ion fragmentation resulted in a fast-decaying survival yield curve, the sample was analyzed again at unit collision energies from 1 eV to 20 eV and the data was processed in the same manner.

Drift Time Measurements

A Synapt G2 HDMS (Waters, Beverly, MA) mass spectrometer was operated in ion mobility mode, using positive electrospray ionization (ESI). Leucine-enkephalin was used as the single point lock mass calibrant. The capillary voltage was adjusted as necessary to obtain the desired sample peak intensity and ranged from 0.5 kV to 4.2 kV. The system was operated with a sample cone potential of 40 V and an extraction cone potential of 60 V. The source block temperature was set at 80°C and the desolvation temperature at 200°C. The trap T-wave cell was operated with a potential of 4.0 eV and an argon pressure of 1.52×10^{-2} mbar. The ion mobility cell was maintained at a pressure of 2.23 mbar (N_2), with a traveling wave speed of 1800 m/s, and pulse amplitude of 40 V. Test compounds were prepared individually at concentrations ranging from 0.1 to 100 nmol/mL in 0.01% TFA (v/v) in methanol, water, or methanol/water (1:1, v/v), based on their LogP value. These samples were introduced into the source by flow injection using a Hamilton syringe pump (Reno, NV) operated at a flow rate of 5 µL/min. Data was acquired in the continuum mode over a mass range of 50-1200 Da, for a period of 10 milliseconds. The reconstructed precursor ion drift peak was smoothed and integrated to determine the drift time for each test compound.

(1) Aderjan, R.; Bogusz, M. *J. Chromatogr. A*. **1988**, *454*, 345–351.

(2) Hall, L. M.; Hall, L. H.; Kertesz, T. M.; Hill, D. W.; Sharp, T. R.; Oblak, E. Z.; Dong, Y. W.; Wishart, D. S.; Chen, M.-H.; Grant, D. F. *J. Chem. Inf. Model.* **2012**, *52*, 1222–1237.

(3) Kertesz, T. M.; Hall, L. H.; Hill, D. W.; Grant, D. F. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1759–1767.

Table S1. Experimental Data

Test Compound	PubChem ID	MIMW	Exp RI	Exp $ECOM_{50}$ (eV)	Exp Drift Time (msec)	Compound Source
6a-Methylprednisolone	4159	374.2093	397	0.77	5.89	Sigma
Adiphenine	2031	311.1885	428	1.80	5.18	Sigma (as HCl)
Albuterol (Salbutamol)	2083	239.1521	270	1.25	4.16	Sigma
Amfenac	2136	255.0895	453	1.46	4.10	Sigma
Aminophylline (Theophylline)	2153	180.0647	252	3.01	2.97	Sigma
Ampicillin	2174	349.1096	289	1.06	5.67	Sigma
Antipyrine	2206	188.0950	306	3.37	3.19	Sigma
Antipyrine-4-amino	2151	203.1059	280	1.98	3.40	Sigma
Bumetanide	2471	364.1093	501	1.24	5.08	Alltech Associates
Cromolyn	2882	468.0693	255	0.86	7.18	MP Biomedicals (as Na)
Cymarin	539061	548.2985	369	0.54	9.29	MP Biomedicals
Daunorubicin	2958	527.1791	383	0.48	8.37	Sigma (as HCl)
Dihydroergotamine	3066	583.2795	388	1.40	8.91	Sigma (as Tartrate)
Diphenoxylate	13505	452.2464	549	2.29	7.72	Alltech Associates
Doxorubicin (Adriamycin)	1691	543.1741	350	0.53	8.75	Sigma (as HCl)
Enalapril	3222	376.1998	368	1.46	5.89	Sigma
Enalaprilat	5362033	348.1685	289	1.47	5.35	Sigma
Ephedrine	5032	165.1154	289	1.67	3.08	West-Ward Pharmaceutical
Ergocristine	98255	609.2951	437	0.69	9.72	Sigma (as HCl)
Fenoterol	3343	303.1471	290	1.56	4.91	Sigma
Ketorolac	3826	255.0895	418	1.82	4.10	Sigma
Leucine Enkephalin	3903	555.2693	332	1.08	8.75	Sigma
Methionine Enkephalin	42785	573.2257	317	1.03	8.86	Sigma (as Acetate)
Methotrexate	4112	454.1713	287	1.12	6.53	Sigma
Methylergonovine	4140	339.1947	308	1.93	5.62	Sigma (as Maleate)
Nimesulide	4495	308.0467	549	1.67	4.75	Sigma
Noscapine	4544	413.1474	357	1.56	6.32	Sigma
Oxaprozin	4614	293.1052	554	1.69	4.86	Sigma

Oxybutynin	4634	357.2304	469	1.99	5.83	Sigma
Oxytetracycline	54675779	460.1482	306	0.91	6.53	Sigma
Perindopril	107807	368.2311	374	1.53	5.99	Sigma
Piperacetazine	19675	410.2028	412	1.90	6.64	Sigma
Poldine*	11018	340.1913	366	2.29	5.45	Theta (as Methylsulfate)
Prazosin	4893	383.1594	335	2.40	6.26	Sigma
Prednisolone	4894	360.1937	366	0.78	5.62	Sigma
Prednisolone Tebutate	4898	458.2668	723	0.38	7.61	United States Pharmacopeia
Prednisone	4900	358.1780	368	1.02	5.56	Sigma
Prolintane	14592	217.1830	369	2.65	3.83	Boehringer Ingelheim & Rhein (as HCl)
Sulfadimethoxine	5323	310.0736	358	1.90	4.75	Sigma
Tenoxicam	54677971	337.0191	334	1.49	4.91	Sigma
Terbutaline	5403	225.1365	272	1.73	3.94	MP Biomedicals (as Hemisulfate)
Terfenadine	5405	471.3137	572	1.50	8.64	Sigma
Thioridazine	5452	370.1537	516	1.95	5.78	Sigma
Thonzonium Bromide	5456	511.4376	1217	0.99	9.67	Sigma
Tyramine (Hydroxyphenethylamine)	5610	137.0841	257	1.25	2.48	Sigma
Methimazole	1349907	114.0252	N/A	N/A	2.30	Chembridge
1-methylpiperidin-4-amine	2737531	114.1157	N/A	N/A	2.32	Chembridge
2-methyl-4(3H)-quinazolinone	15674	160.0637	N/A	N/A	2.81	Chembridge
(3-methylpiperidin-1-yl)acetic acid	3163244	157.1103	N/A	N/A	2.89	Chembridge
1,2,3-trimethyl-4-phenyl-piperidin-4-ol	2838298	219.1623	N/A	N/A	3.68	Chembridge
1,3-dimethyl-N-(5-methyl-3-isoxazolyl)-1H-pyrazole-4-sulfonamide	6466738	256.0630	N/A	N/A	3.97	Chembridge
3-{[(cyclopentylmethyl)amino]methyl}-N,N-diethyl-2-pyridinamine	28389621	261.2205	N/A	N/A	4.48	Chembridge
N-(1-butyl-4-methyl-1H-pyrazol-5-yl)-2-phenyl-1H-imidazole-4-carboxamide	46995060	323.1746	N/A	N/A	5.59	Chembridge
N-(3-acetylphenyl)-3-methyl-2-phenyl-quinoline-4-carboxamide	1313097	380.1525	N/A	N/A	5.94	Chembridge
N-{4-[(mesitylamino)sulfonyl]phenyl}-2-methylbenzamide	2910668	408.1508	N/A	N/A	6.89	Chembridge

N-(2-ethoxyphenyl)-3-methyl-2-(4-methylphenyl)quinoline-4-carboxamide	2168581	396.1838	N/A	N/A	6.78	Chembridge
N-{[2-[(cyclohexylmethyl)sulfonyl]-1-(2-methoxyethyl)-1H-imidazol-5-yl]methyl}-N,3-dimethyl-1-butanamine	42514440	399.255563	N/A	N/A	6.97	Chembridge
4-{[(3,4-dimethylphenyl)(methylsulfonyl)amino]methyl}-N-(2-methylbenzyl)benzamide	1350894	436.1821	N/A	N/A	7.83	Chembridge
3-propoxy-N-{4-[4-(2-thienylcarbonyl)-1-piperazinyl]phenyl}benzamide	2987700	449.1773	N/A	N/A	8.26	Chembridge
N-[2-(benzoylamino)benzoyl]-N-(1,5-dimethyl-3-oxo-2-phenyl-2,3-dihydro-1H-pyrazol-4-yl)phenylalaninamide	2902191	573.2376	N/A	N/A	9.29	Chembridge
3,3'-{2,2-propanediylbis[4,1-phenyleneoxy(2-hydroxy-3,1-propanediyl)]}bis(5,5-dimethyl-2,4-imidazolidinedione)	2884368	596.2846	N/A	N/A	9.13	Chembridge

Table S2. Effective Drift Times for Poly-Alanine Standards

No of Ala	m/z	He Drift Tube Derived Cross Sectional Area ($Å^2$)	Mass Independent Cross Sectional Area ($Å^2$)	Average Effective Drift Time (milliseconds)
3	232.1297	89.0	176.5	3.90
4	303.1668	102.9	204.5	4.64
5	374.2039	115.0	228.9	5.52
6	445.2411	127.0	252.9	6.43
7	516.2782	140.5	280.0	7.42
8	587.3153	155.5	310.0	8.64
9	658.3524	167.9	334.9	9.78

Table S3. Molecular Descriptors used in random forest models

Descriptor	Definition	Class*
VP.0	Chi path - valence path of order 0	2D
VABC	Van der Waals volume	2D
ECCEN	Eccentric connectivity index - A topological descriptor combining distance and adjacency information	2D
apol	Sum of atomic polarizabilities	2D
MOMI.Y	Moment of inertia along y-axiz	3D
MobCSA	Mobcal cross sectional area	3D
WPATH	Wiener path number	2D
VP.1	Chi path - valence path of order 1	2D
SP.1	Chi path - simple path of order 1	2D
GRAVH.2	Square root of hydrogen-included gravitational index	3D
GRAVH.1	Gravitational index - hydrogens included	3D
WA.unity	Unit weighted WHIM size descriptor – Quadratic contribution to the total dimension of the molecule	3D
WTPT.1	Weighted path of order 1	2D
SP.0	Chi path -simple path of order 0	2D
MOMI.X	Moment of inertia along x-axis	3D
WV.unity	Unit weighted WHIM size descriptor – Linear contribution to the total dimension of the molecule	3D
VAdjMat	Vertex adjacency information	2D
nB	Bond count - number of bonds of a certain bond order	2D
WT.unity	Unit weighted WHIM size descriptor – Total contribution to the size	3D
ATSp1	Autocorellation polarizability of lag 1	2D
GRAVH.3	Cube root of hydrogen-included gravitational index	3D
nAtom	Atom count	2D
Kier1	Kappa shape indices	2D
WPSA.1	Charged partial surface area	3D
Zagreb	Zagreb index - sum of the squared atom degree of all heavy atoms	2D

MOMI.Z	Moment of inertia along z-axis	3D
ATSm1	Autocorellation mass of lag 1	2D
SP.2	Chi path - simple path of order 2	2D
ATSp2	Autocorellation polarizability of lag 2	2D
GRAV.6	Grav3 for all pairs of atoms (not just bonded pairs)	3D
MOMI.R	Radius of gyration	3D
GRAV.5	Grav2 for all pairs of atoms (not just bonded pairs)	3D
KIer2	Kappa shape indices	2D
VP.2	Chi path - valence path of order 2	2D
ATSm2	Autocorellation mass of lag 2	2D
SP.4	Chi path - simple path of order 4	2D
VP.7	Chi path - valence path of order 7	2D
fragC	Fragment complexity	2D
SP.3	Chi path - simple path of order 3	2D
GRAV.4	Grav1 for all pairs of atoms (not just bonded pairs)	3D
VP.3	Chi path - valence path of order 3	2D

* 3D – requires a 3D (xyz coordinates) representation of the structure. 2D – only molecular graph (connectivity between atoms) is needed

Table S4. Reproducibility of RI

Date	Nicotinic acid	N,N-dimethylbenzamide	Corticosterone	Probenecid	Indomethacin	Phenylbutazone	Tetraethylthiuram disulfide	Thonzonium
3/2/12	105	318	433	523	610	662	756	1222
3/5/12	104	317	432	524	610	663	758	1222
3/6/12	105	318	432	524	610	663	757	1220
3/7/12	106	318	433	524	609	663	757	1228
3/8/12	106	318	434	524	611	664	757	1222
3/9/12	104	318	434	526	612	665	760	1219
3/12/12	107	319	434	525	610	663	760	1223
3/13/12	104	318	434	526	613	667	761	1222
3/13/12	106	318	433	526	612	665	760	1224
3/13/12	105	318	434	526	612	666	760	1222
3/13/12	106	319	433	525	611	664	759	1220
3/13/12	107	318	433	526	612	666	760	1224
3/14/12	102	318	433	523	610	663	757	1216
3/19/12	105	318	434	525	612	664	759	1226
3/21/12	101	318	433	523	609	663	757	1218
3/23/12	105	319	434	525	610	664	756	1223
STD	1.74	0.47	0.60	1.19	1.19	1.31	1.61	2.93

Table S5. Reproducibility of $ECOM_{50}$ (eV)

Date	Ampicillin	Isoxsuprine	Drofenine	Tetramisole
5/18/12	1.09	1.27	1.97	3.26
5/21/12	1.09	1.26	1.98	3.27
5/22/12	1.10	1.27	1.98	3.26
5/23/12	1.09	1.26	1.98	3.28
5/29/12	1.08	1.26	1.98	3.25
5/30/12	1.09	1.26	1.99	3.29
5/31/12	1.08	1.26	1.98	3.27
6/1/12	1.09	1.26	1.98	3.26
6/2/12	1.10	1.27	1.98	3.26
6/4/12	1.09	1.26	1.98	3.26
STD	0.01	<0.01	<0.01	0.01

Table S6. Rankings for Old Bins

Test Compound	22% prec CID energy	After pre filters	After RI	After $ECOM_{50}$	Mass Frontier rank after pre filters	Mass Frontier rank after RI and $ECOM_{50}$	MetFrag peaks rank (22% prec CID spectra)	MetFrag score rank (22% prec CID spectra)	MetFrag peaks rank (composite spectra)	MetFrag score rank (composite spectra)	MetFrag score rank (intensity corrected composite spectra)
6a-Methylprednisolone	20	186	156	129	11	9	6	2	10	3	3
Adiphenine	20	599	588	585	6	6	6	6	15	5	4
Albuterol	20	137	117	99	15	14	9	33	3	3	7
Amfenac	30	334	252	246	11	10	14	24	28	55	57
Aminophylline (Theophylline)	10	84	63	58	21	19	27	18	31	8	13
Ampicillin	20	607	536	510	1	1	1	1	1	1	1
Antipyrine	30	302	301	294	97	95	68	167	107	161	64
Antipyrine-4-amino	10	222	218	169	16	13	29	18	17	17	13
Bumetanide	20	599	UFO	N/A	10	N/A	N/A	N/A	N/A	N/A	N/A
Cromolyn	20	31	UFO	N/A	2	N/A	N/A	N/A	N/A	N/A	N/A
Cymarin	20	59	39	UFO	8	N/A	N/A	N/A	N/A	N/A	N/A
Daunorubicin	20	105	UFO	N/A	12	N/A	N/A	N/A	N/A	N/A	N/A
Dihydroergotamine	30	34	13	13	1	1	1	1	1	2	2
Doxorubicin	20	59	31	31	3	3	4	7	3	6	6
Enalapril	20	188	162	153	1	1	57	28	2	7	7
Enalaprilat	20	333	224	222	2	2	7	4	4	11	11
Ephedrine	30	242	236	177	5	5	32	8	38	16	16
Ergocristine	30	14	11	7	1	1	1	1	1	4	4
Fenoterol	20	322	243	228	5	5	2	5	4	4	3
Tyramine (Hydroxyphenethylamine)	10	163	161	77	166	77	77	77	14	27	22
Ketorolac	20	334	309	306	37	35	39	14	19	7	4
Leucine Enkephalin	20	51	12	11	2	2	2	3	2	2	2

Methionine Enkephalin	20	64	16	15	1	1	1	1	1	1	1
Methotrexate	20	636	242	237	116	42	49	8	12	9	9
Methylergonovine	30	500	373	371	1	1	2	2	3	2	2
Norpropoxyphene	10	377	368	317	15	12	30	8	20	6	6
Noscapine	30	246	205	196	3	3	82	39	156	38	39
Oxybutynin	30	84	67	59	6	5	3	1	3	1	1
Oxytetracycline	20	451	53	53	4	4	1	1	1	1	1
Perindopril	20	51	42	30	2	1	2	2	12	3	3
Piperacetazine	30	465	261	260	1	1	1	1	1	1	1
Poldine	30	680	651	649	19	19	22	21	71	13	14
Prazosin	40	149	114	112	4	4	3	36	2	43	42
Prednisolone	20	222	192	161	13	11	4	1	7	2	2
Prednisone	20	315	267	234	6	5	7	4	14	4	4
Prolintane	20	105	104	67	9	5	13	11	11	17	12
Sulfadimethoxine	30	90	69	66	18	15	4	3	4	3	3
Taurocholate	20	54	42	UFO	4	N/A	N/A	N/A	N/A	N/A	N/A
Tenoxicam	20	28	24	24	1	1	1	1	1	1	1
Terbutaline	20	170	155	145	31	25	23	23	15	24	28
Median					6	5	6	6	7	5	4
Average		253	195	180	19	13	18	17	18	15	12
STD					36	21	23	31	32	29	16

UFO – Unknown Filtered Out

Table S7. Rankings for New Bins

Test Compound	22% prec CID energy	Bin Size	After pre filters	After RI	After RI and $ECOM_{50}$	MetFrag peaks rank after pre filters (22% prec CID spectra)	MetFrag peaks rank after RI and $ECOM_{50}$ (22% prec CID spectra)	MetFrag score rank (intensity corrected composite spectra)
6a-Methylprednisolone	20	4099	2389	2131	1819	51	43	16
Adiphenine	20	3657	2642	2533	2452	7	7	7
Albuterol	20	1679	1356	1093	943	82	73	35
Amfenac	30	2162	1334	958	927	32	22	152
Aminophylline (Theophylline)	10	718	295	218	200	89	70	40
Ampicillin	20	6828	3249	2874	2734	1	1	1
Antipyrine	30	1868	1490	1476	1423	388	373	347
Antipyrine-4-amino	10	1881	1447	1410	1182	289	262	133
Dihydroergotamine	30	493	302	203	152	5	3	3
Doxorubicin	20	795	223	124	122	6	6	7
Enalapril	20	3854	2111	1885	1785	706	644	30
Enalaprilat	20	3976	2458	1924	1804	58	52	36
Ephedrine	30	1803	1410	1371	966	231	225	101
Ergocristine	30	328	189	100	81	2	1	10
Fenoterol	20	3272	2071	1631	1534	3	3	20
Tyramine (Hydroxyphenethylamine)	10	922	788	770	218	788	218	39
Ketorolac	20	2162	1334	1145	1136	146	136	54
Leucine Enkephalin	20	879	462	173	138	2	2	14
Methionine Enkephalin	20	659	282	98	88	1	1	10
Methotrexate	20	6332	4114	1678	1655	697	301	11
Methylergonovine	30	6286	4780	3597	3526	8	8	7
Norpropoxyphene	10	2471	1909	1819	1549	192	154	51

Noscapine	30	5801	1662	1470	1439	619	553	135
Oxybutynin	30	1819	871	658	608	15	14	4
Oxytetracycline	20	3679	1702	394	386	8	8	6
Perindopril	20	2256	676	561	491	22	19	7
Piperacetazine	30	5115	2899	1825	1803	1	1	1
Poldine*	30	5102	4084	3802	3723	14	14	14
Prazosin	40	2509	1223	1022	1019	32	30	407
Prednisolone	20	4315	2480	2252	1941	25	21	18
Prednisone	20	4604	2364	1823	1636	69	64	40
Prolintane	20	1006	878	869	579	157	120	54
Sulfadimethoxine	30	1754	481	325	307	10	10	8
Tenoxicam	20	871	124	112	106	4	4	4
Terbutaline	20	1477	1148	1031	966	113	104	121
MEDIAN						28.5	22	18
AVERAGE			1635	1296	1184	142	102	56
STD						228	157	91

Table S8. Estimated MetFrag rankings for smallest RI and $ECOM_{50}$ windows

Test Compound	% compounds remaining at $ECOM_{50}$ = ± 0.5 and RI = ± 40	MetFrag score rank (intensity corrected composite spectra)	MetFrag peaks rank (intensity corrected composite spectra)
6a-Methylprednisolone	16.6	5	35
Adiphenine	12.3	5	17
Albuterol	5.2	8	7
Amfenac	12.0	28	11
Aminophylline (Theophylline)	12.5	8	17
Ampicillin	7.6	1	1
Antipyrine	21.1	93	160
Antipyrine-4-amino	3.7	7	5
Dihydroergotamine	24.5	3	3
Doxorubicin	11.2	6	5
Enalapril	10.4	9	10
Enalaprilat	10.7	6	5
Ephedrine	21.4	24	45
Ergocristine	12.2	3	1
Fenoterol	2.8	13	6
Tyramine (Hydroxyphenethylamine)	3.9	14	10
Ketorolac	28.8	20	24
Leucine Enkephalin	7.6	4	2
Methionine Enkephalin	2.8	2	1
Methotrexate	31.6	10	29
Methylergonovine	5.0	2	2
Norpropoxyphene	7.9	10	29
Noscapine	4.7	5	42
Oxybutynin	17.5	4	11
Oxytetracycline	1.9	2	6
Perindopril	9.8	3	19
Piperacetazine	10.0	1	1
Poldine*	14.6	12	148
Prazosin	16.6	170	6
Prednisolone	14.2	4	15
Prednisone	10.4	11	32
Prolintane	29.4	31	32
Sulfadimethoxine	16.0	5	5
Tenoxicam	25.8	4	4
Terbutaline	7.1	11	17

AVERAGE	15.5	21.8
MEDIAN	6	10
STD	31.3	35.3

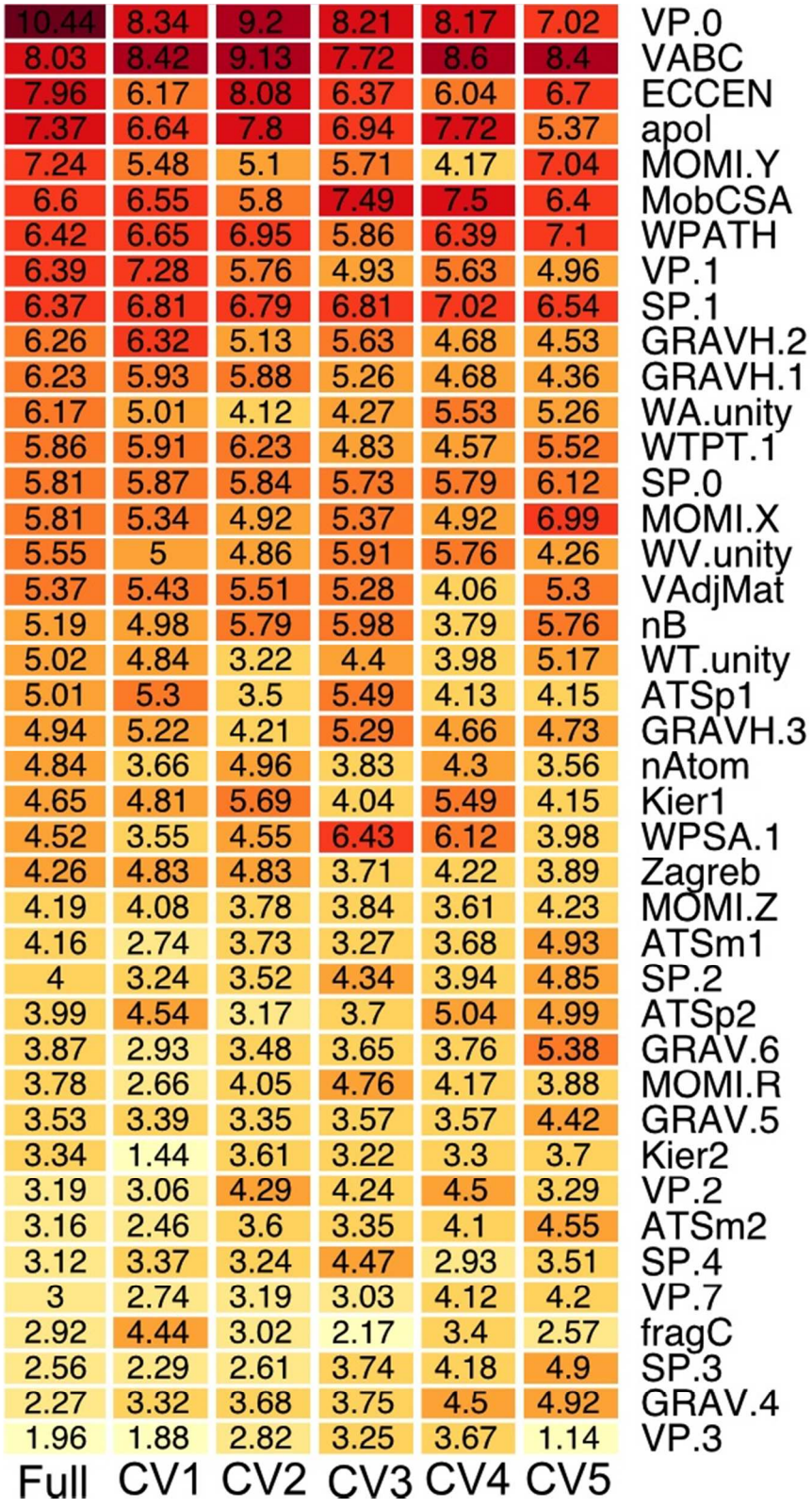

Fig S1. Relative descriptor importance for 5 cross validation replicates