Natural Products Dereplication by Diffusion Ordered NMR Spectroscopy (DOSY)

Guy Kleks,[ab] Darren C. Holland,[ab] Joshua Porter[ab] and Anthony R. Carroll*[ab]

[a]School of Environment and Science, Griffith University, Gold Coast, QLD 4222, Australia

[b]Griffith Institute for Drug Discovery, Griffith University, Brisbane, QLD 4111, Australia

Supporting Information

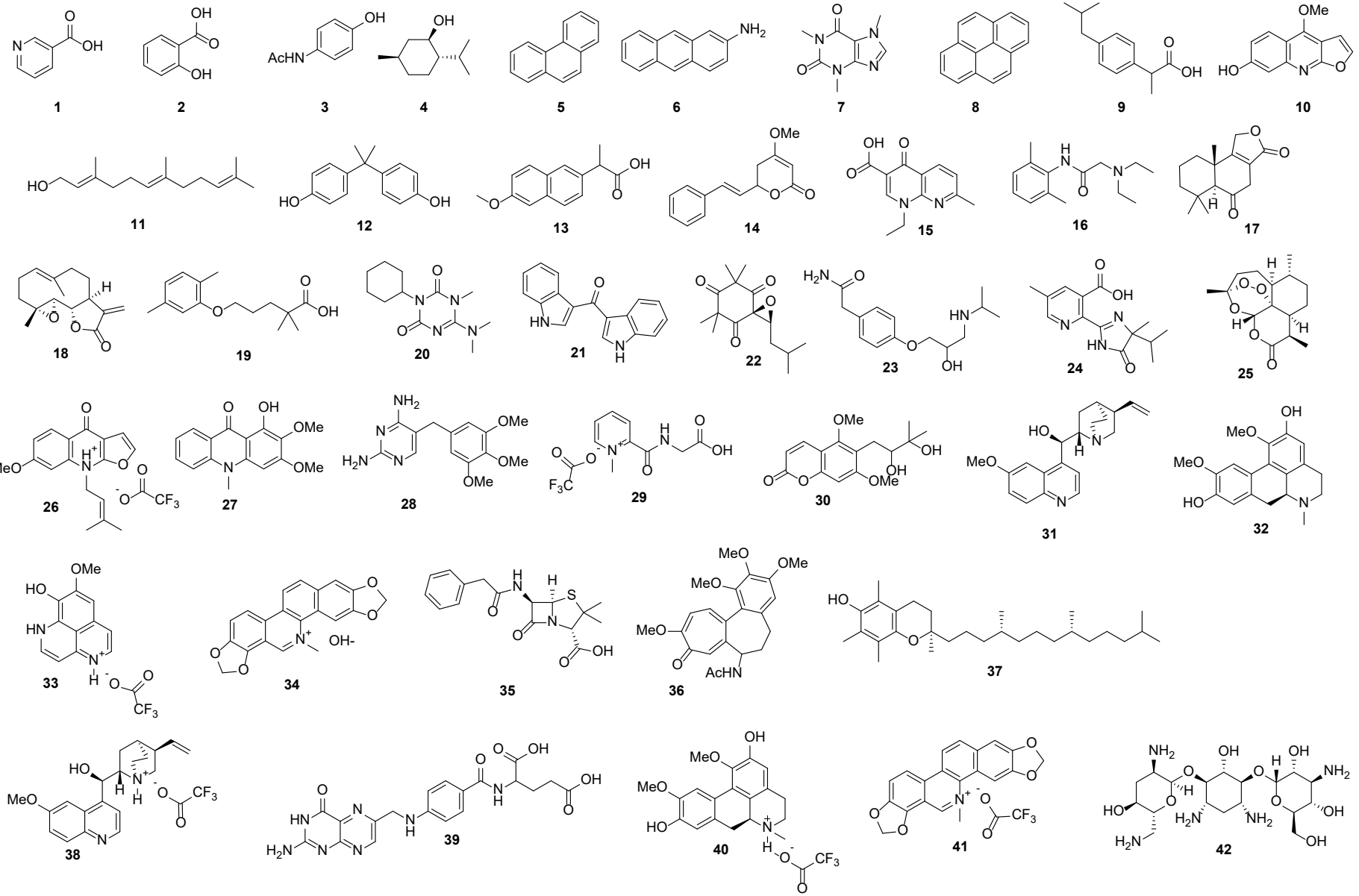
Table S1 List of compounds used to generate models 1-4

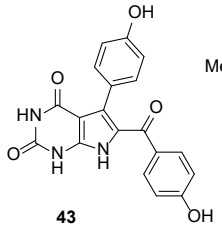
#	Compound	Code	MW	Class	Group	MeOH%	clogP	δ_H protic	$^{eff}\delta_H$ protic	A ($\Delta\delta_H$)[c]	D	MW_{pre} Model 1	MW_{pre} Model 2	MW_{pre} Model 3
1	niacin	nia	123.11	NP[a]	HBD	6%	0.144	13.43	13.43		3.817E-10	153	116	112
2	salicyclic acid	sal	138.12	NP	HBD	22%	0.799	11.35	11.35		3.714E-10	160	132	130
3	acetaminophen	ace	151.16	drug	HBD	24%	1.018	9.64	9.64		3.013E-10	226	198	183
4	menthol	men	156.27	NP	HBD		2.411	4.28	4.28		4.071E-10	137	148	
5	phenanthrene	phe	178.23	NP	disc	86%	4.048		0		4.489E-10	117	148	150
6	2-aminoanthracene	ant	193.24	Syn	HBD	75%	3.371	5.55	5.55	0.22 (1.63)	3.497E-10	177	181	205
7	caffeine	caf	194.19	NP	HBA	37%	-0.178		0		3.822E-10	153	193	138
8	Pyrene	pyr	202.25	NP	disc	90%	4.690		0		4.155E-10	133	168	174
9	ibuprofen	ibu	206.28	drug	HBD	68%	3.003	12.24	12.24		2.924E-10	237	189	257
10	7-hydroxydictamnine	hyd	215.20	NP	HBD	62%	2.294	10.16	10.16		2.959E-10	233	200	242
11	farnesol	far	222.37	NP	HBD		5.657	4.42	4.42		3.130E-10	212	227	
12	bisphenol A	bisA	228.29	Syn	HBD	67%	2.866	9.15	9.15	0.48 (3.56)	2.460E-10	316	282	333
13	naproxen	nap	230.26	drug	HBD	57%	2.695	12.28	12.28		2.742E-10	264	209	281
14	Kavain	kav	230.26	NP	HBA	66%	2.072		0		3.177E-10	207	262	210
15	nalidixic acid	nal	232.24	drug	HBA	55%	0.536	**14.89**	0		3.349E-10	190	240	191
16	lignocaine	lig	234.34	drug	HBD	70%	2.157	9.15	9.15	0.04 (0.22)	3.239E-10	200	179	222
17	fragrolide	fra	248.32	NP	HBA	64%	2.129		0		3.253E-10	199	252	212
18	parthenolide	par	248.32	NP	HBA	68%	2.753		0		3.256E-10	199	251	217
19	gemfibrozil	gem	250.33	drug	HBD	78%	3.555	12.03	12.03		2.712E-10	269	215	309
20	hexazinone	hex	252.31	Syn	HBA	60%	1.384		0		3.002E-10	227	287	234
21	bis-indole-methanone	bis-ind	260.29	NPd	HBD	63%	2.868	11.81	11.81		2.607E-10	287	232	297
22	woodsianone B	woo	266.33	NP	HBA	79%	1.746		0		2.971E-10	231	292	269
23	atenolol	ate	266.34	drug	HBD	26%	0.307	7.38	7.38		2.364E-10	337	322	269
24	imazapic	ima	275.30	Syn	HBD	24%	0.642	11.8	11.8		2.357E-10	339	273	266

25	artemisinin	art	282.33	NP	HBA	74%	2.258		0		3.256E-10	199	251	227
26	acrophylline*	acr*	397.35	NP	HBA	75%	3.498		0		2.673E-10	275	348	310
27	arborinine	arb	285.29	NP	HBA	69%	2.635	**14.86**	0		3.092E-10	216	274	238
28	trimethoprim	tri2	290.32	drug	HBD	43%	1.069	6.07	6.07	0.18 (1.30)	2.299E-10	353	354	314
29	lamourouic acid*	Lam*	308.21	NP	HBD		-4.642	12.7	12.7		2.219E-10	374	292	
30	toddalolactone	todd	308.33	NP	HBD	59%	1.363	4.14	4.14		2.520E-10	304	328	303
31	quinine	qui	324.42	NP	HBD	72%	2.610	5.61	5.61		2.520E-10	303	310	332
32	boldine	bol	327.37	NP	HBD	54%	2.788	9.09	9.09	0.52 (3.89)	2.153E-10	394	352	375
33	9-demethylaaptamine*	aap*	328.24	NP	HBD		1.059	10.14	10.14		2.255E-10	365	313	
34	Sanguinarine (OH^-)	san	349.10	NP	HBA	85%	0.664		0		2.585E-10	291	367	348
35	penicillin G	pen	334.39	NP	HBD		1.538	8.68	8.68		2.121E-10	403	366	
36	colchicine	col	399.44	NP	HBD	58%	1.863	8.57	8.57		2.058E-10	424	386	412
37	α-tocopherol	toc	430.71	NP	HBD	100%	9.644	7.35	7.35	0.43 (3.17)	2.126E-10	402	383	521
38	quinine*****	qui*	438.44	NP	HBD		0.501	5.61	5.61		2.147E-10	395	403	
39	folic acid	fol	441.40	NP	HBD	17%	-1.646	12.1	12.1		1.523E-10	697	555	498
40	boldine*****	bol*	441.40	NP	HBD		0.679	9.55	9.55		1.855E-10	503	442	
41	sanguinarine*****	san*	445.34	NP	HBA		0.664		0		2.279E-10	358	452	
42	tobramycin	tob	467.51	NP	HBD		-7.020	4.88	4.88		1.693E-10	585	612	
43	rigidin	rig	363.32	NP	HBD		1.738	11.86	11.86		1.686E-10	589	473	
44	emetine	eme	480.64	NP	HBD	72%	4.910	2.00[d]	2.00[d]		2.019E-10	438	511	467
45	doxycycline hyclate	doxy	480.90	NPd[b]	HBD		-2.466	**11.48**	10.19		1.707E-10	577	494	
46	fusidic acid	fus	516.71	NP	HBD		5.823	4.03	4.03		1.672E-10	597	646	
47	prunolide C	prunC	574.53	NP	HBD	64%	2.860	10.4	10.4		*1.499E-10*	715	607	700
48	rutin	rut	610.52	NP	HBD	51%	-1.257	**12.60**	10.72		1.420E-10	782	656	696
49	oleandomycin	oleery	687.86	NP	HBD		1.473	4.23	4.23		1.676E-10	595	638	
50	erythromycin	ery	733.93	NP	HBD	67%	1.672	4.28	4.28		1.606E-10	639	684	643
51	clarithromycin	clar	747.95	NPd	HBD	73%	2.100	5.00	5.00		1.707E-10	577	601	609

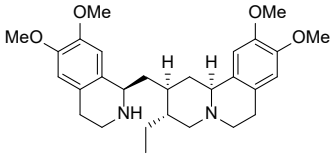
52	oleandomycin triacetate	ole-tri	813.97	NPd	HBA	83%	2.927		0	1.660E-10	604	761	681
53	rifampicin	rifa	822.94	NPd	HBD	73%	4.713	**12.44**	9.52	1.456E-10	751	658	780
54	digitonin	dig	1229.31	NP	HBD	81%	-5.302	5.00[d]	5.00[d]	1.078E-10	1234	1282	1304
55	vancomycin hydrochloride	vanco	1485.71	NP	HBD		-4.551	9.46	9.46	8.942E-11	1679	1471	
56	botryllamide C	botC	404.265	NP	Br[e]					2.048E-10	427		
57	methyl 2-(5-bromo-1*H*-indol-3-yl)-2-oxoacetate	5-BIO	282.09	NP	Br					3.234E-10	199		
58	methyl 2-(6-bromo-1*H*-indol-3-yl)-2-oxoacetate	6-BIO	282.09	NP	Br					3.333E-10	190		
59	prunolide B	prunB	1047.91	NP	Br					1.410E-10	792		
60	prunolide A	prunA	1205.70	NP	Br					1.402E-10	799		
61	procerolide B	procB	563.03	NP	Br					2.092E-10	414		
62	echinosulfone A	echA	498.15	NP	Br					2.338E-10	343		
63	bis(6-bromo-1*H*-indol-3-yl)methanone	Br-bis-ind	418.08	NPd	Br					2.448E-10	318		

[*]TFA salt. [a]NP = Natural Product. [b]NPd = Natural Product-derived. [c]Determined H-bond acidity ($\Delta\delta_H$ DMSO/$CDCl_3$).
[d]Not observed, estimated chemical shift. [e]Br = Brominated.

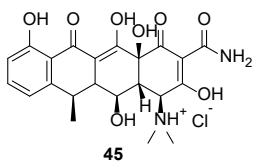

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42



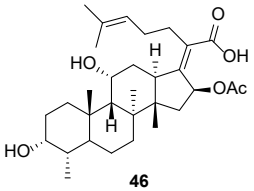
43



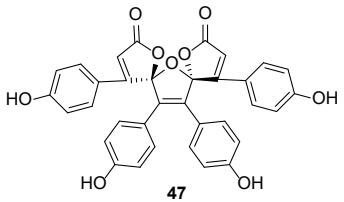
44



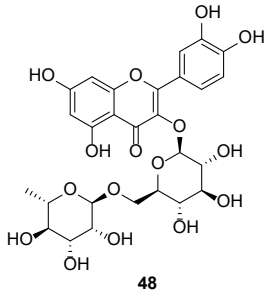
45



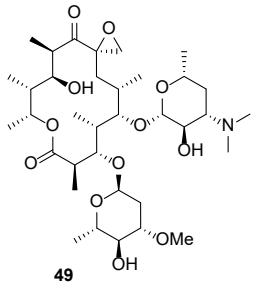
46



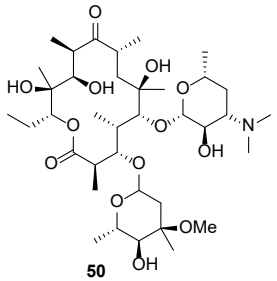
47



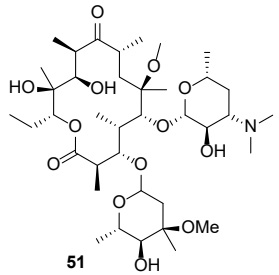
48



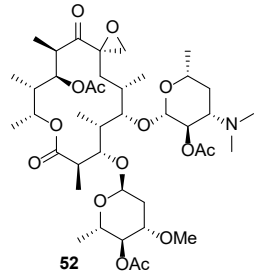
49



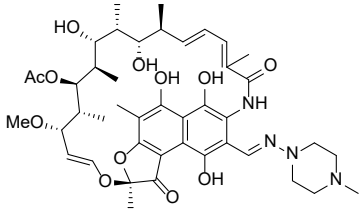
50



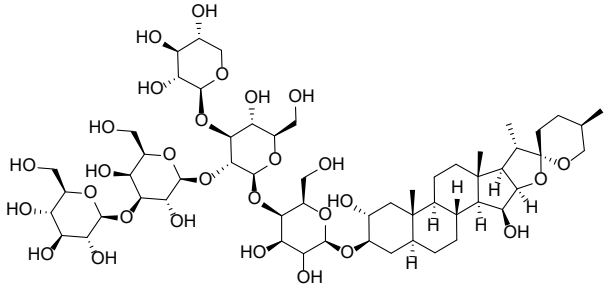
51



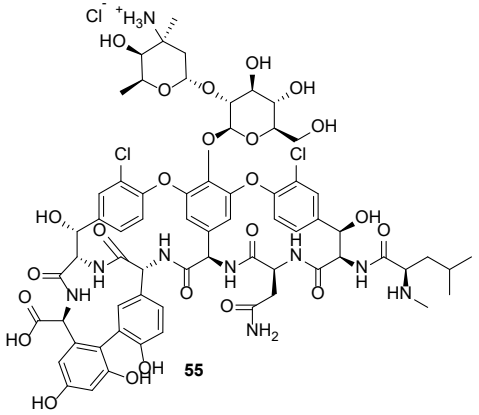
52



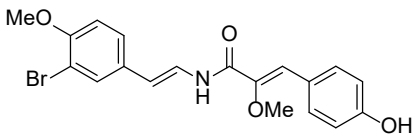
53



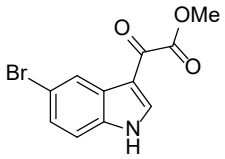
54



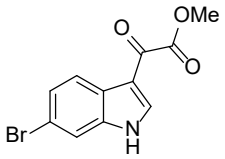
55



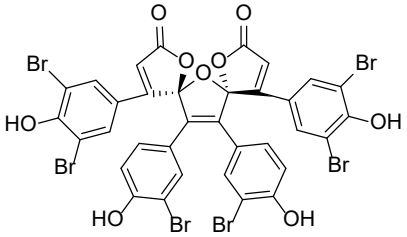
56



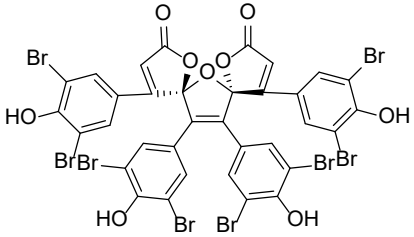
57



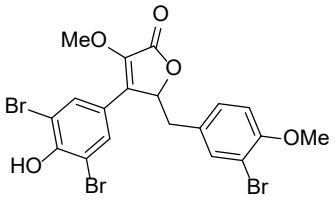
58



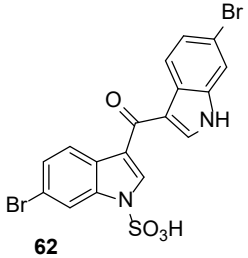
59



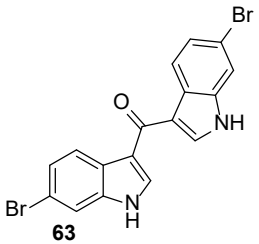
60



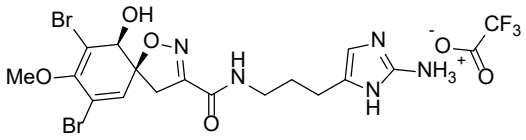
61



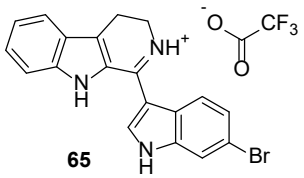
62



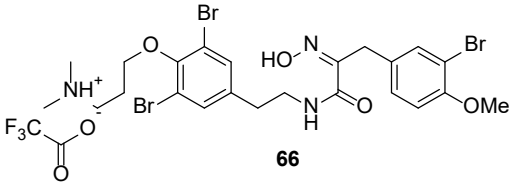
63



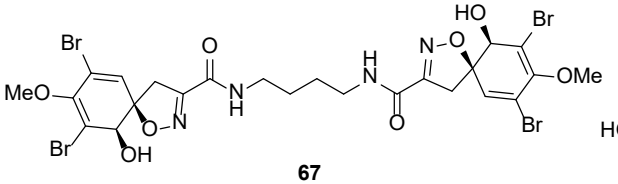
64



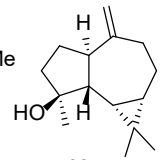
65



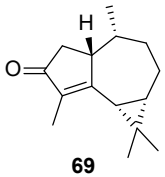
66



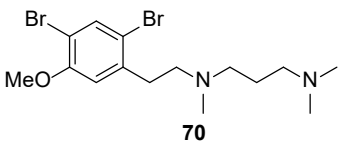
67



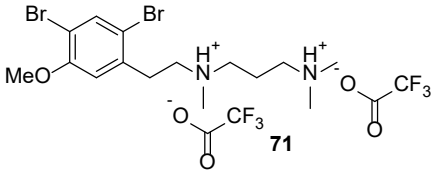
68



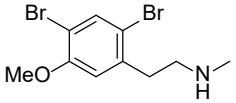
69



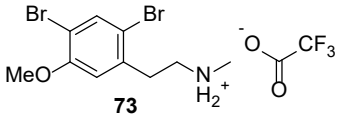
70



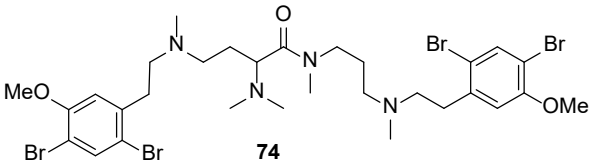
71



72



73



74

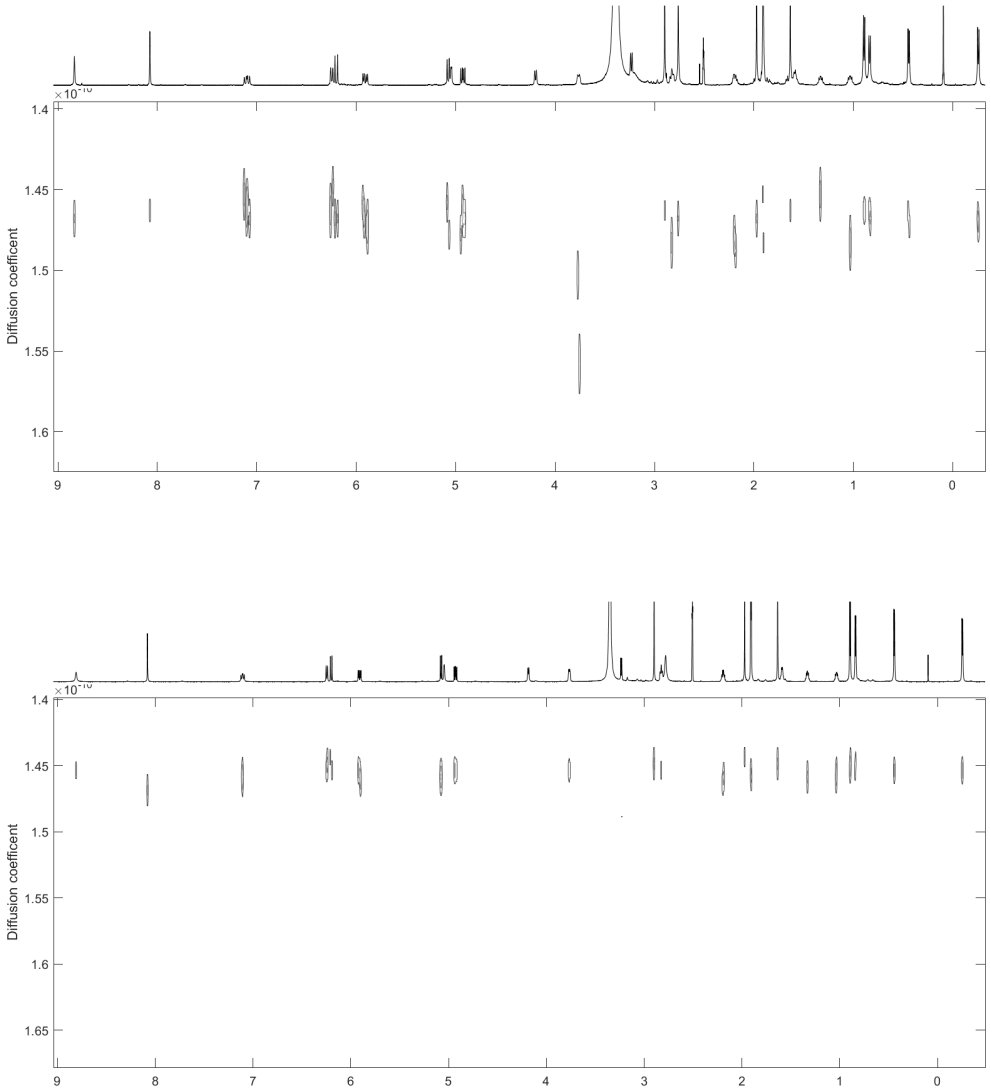


Fig. S1 ^{1}H DOSY spectrum of rifampicin (**53**), acquired on Bruker 500 MHz spectrometer (upper) and 800 MHz (bottom). The data was acquired for different samples at different concentrations (20.6 mM and 10.4 mM), prepared on different days (08/03/2020 and 20/11/2019 respectively). Both spectra were referenced to the TTMS signal at $3.157 \cdot 10^{-10}$.

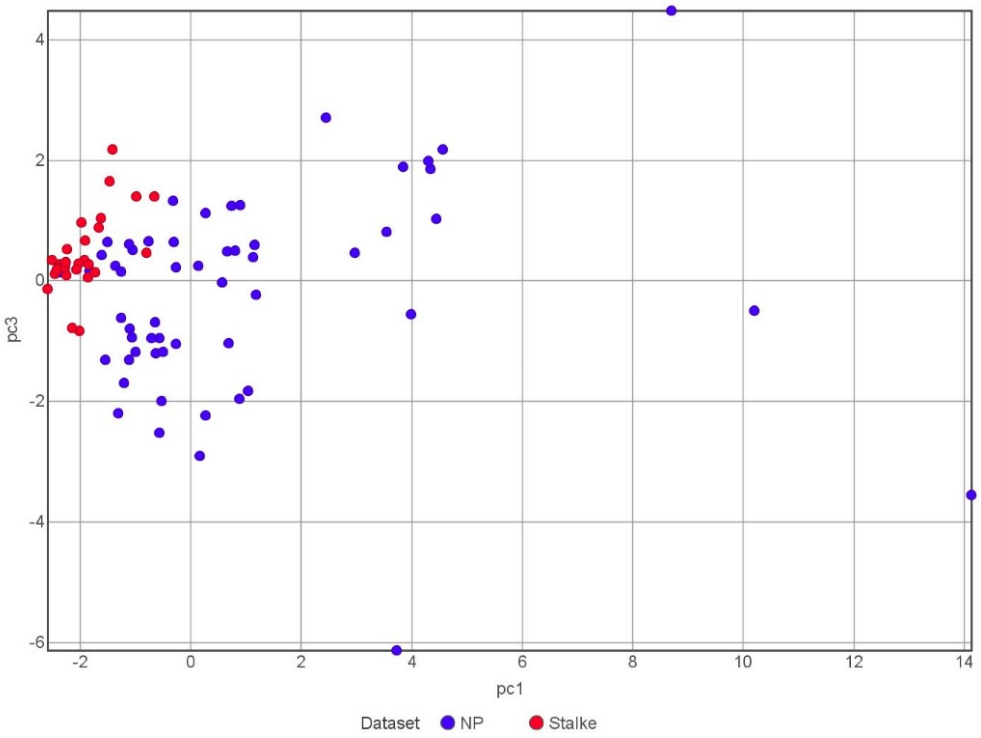


Fig. S2 Principal component analysis (PCA) plot generated for the following physicochemical properties: Total Molweight, cLogP, H-Acceptors, H-Donors, Polar Surface Area, Shape Index, Molecular Flexibility, Electronegative Atoms, Rotatable Bonds, Small Rings, Aromatic Rings, Aromatic Atoms, sp3-Atoms, Symmetric atoms, Amides, Amines, Alkyl-Amines, Aromatic Amines, Aromatic Nitrogens, Basic Nitrogens, Acidic Oxygens. The figure contains the NP data set (n = 55) and those compounds used by Stalke and co-workers (n = 27) to create MW prediction models for expanded disc-like (ED), compact spherical molecules (CS) and dissipated spheres and ellipsoids (DSE).[16]

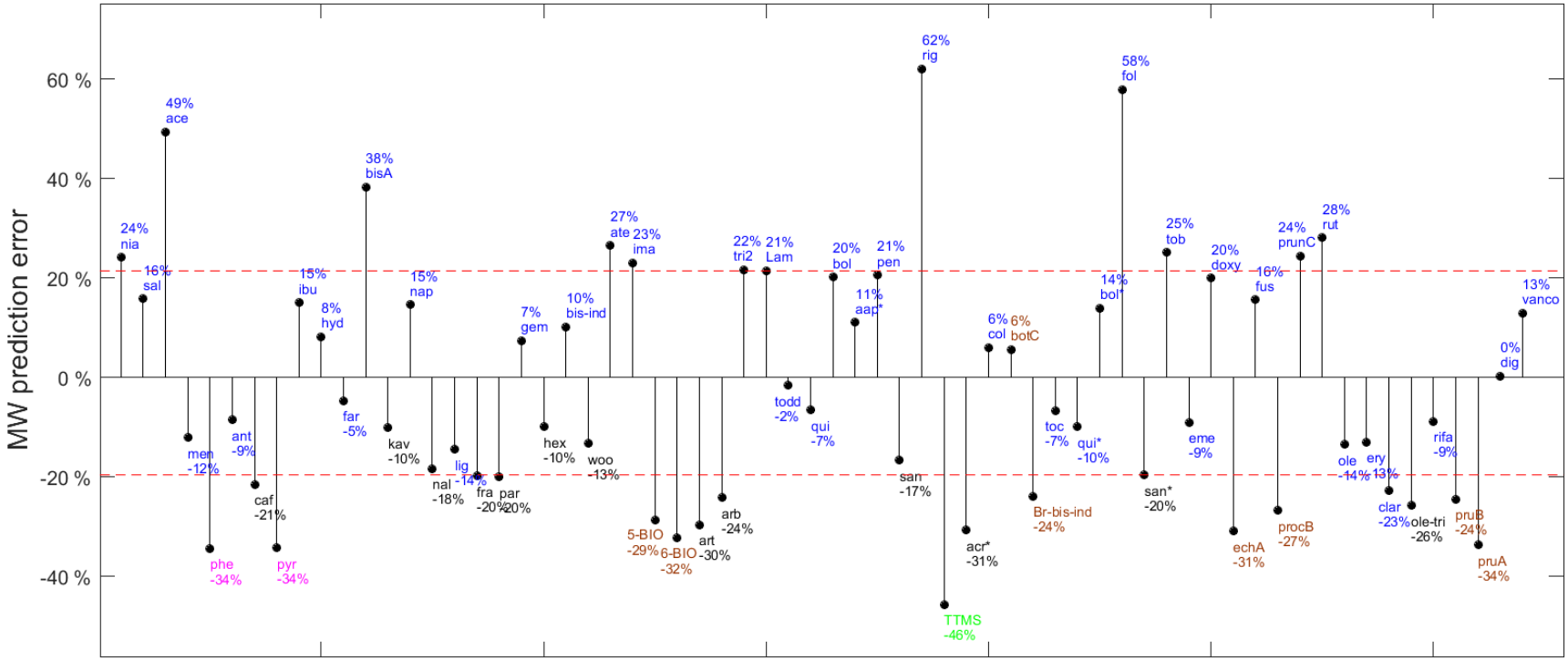


Fig. S3 Residual plot of MW_{pre} errors for model 1 generated for HBDs, non-HBDs and discs (*n*=55). HBDs are in blue, non-HBDs in black, discs in magenta, spheres in green, brominated in brown. Average MW_{pre} error range displayed by red dashed lines.

Table S2. Model 1 (*n*=55, R^2=0.852, adjusted R^2=0.849)

Coefficients	Estimate	SE	tStat	pValue
log(*A*)	-8.0952	0.087971	-92.021	3.75E-60
α	-0.60572	0.034735	-17.439	1.30E-23

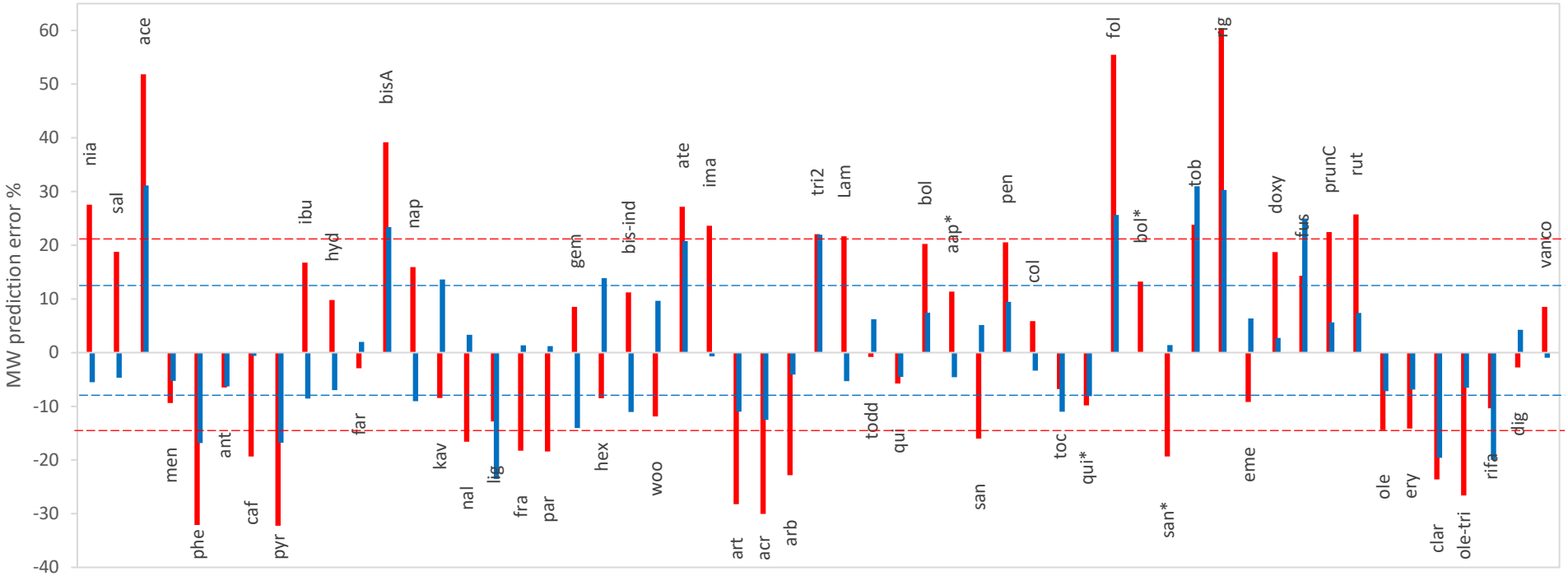


Fig. S4 Residual plot of MW_{pre} errors using Model 1 (red) and Model 2 (blue), with average MW_{pre} error range displayed by dashed lines.

Table S3. Model 2 (n=55, R^2=0.949, adjusted R^2=0.947)

Coefficients	Estimate	SE	tStat	pValue
Intercept	-8.02802	0.052556	-152.752	1.16E-70
logMW	-0.6077	0.02058	-29.529	3.66E-34
$^{eff}\delta_H$	-0.01019	0.001024	-9.95027	1.23E-13

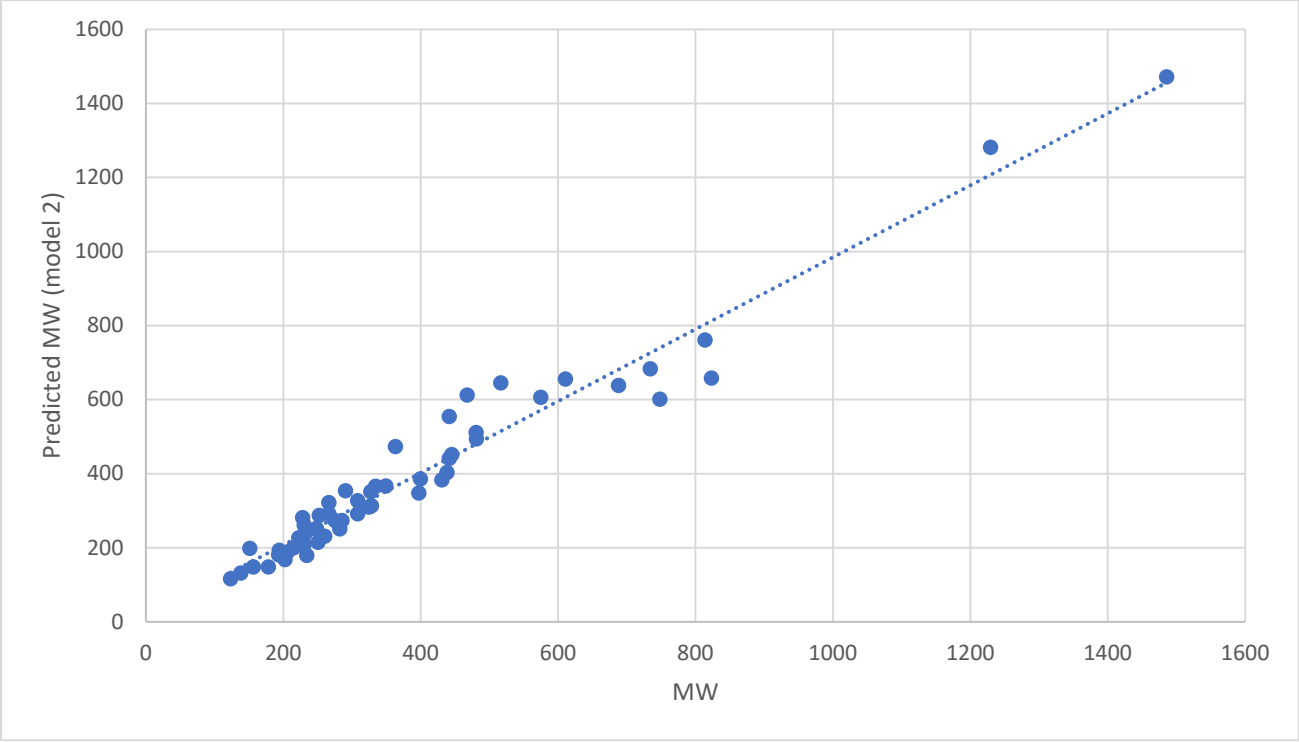


Fig. S5 MW vs. Predicted MW for natural products, drug and synthetic compounds (n= 55) derived from the diffusion model 2.

Table S4. Model: logD = lowMW + clogP + intercept (*n*=55, R^2=0.87, adjusted R^2=0.865)

	Estimated coefficients	**SE**	**tStat**	**pValue**
Intercept	-8.1686	0.087434	-93.425	1.34E-59
clogP	0.007662	0.002832	2.7051	0.00921
logMW	-0.58146	0.034025	-17.089	5.50E-23

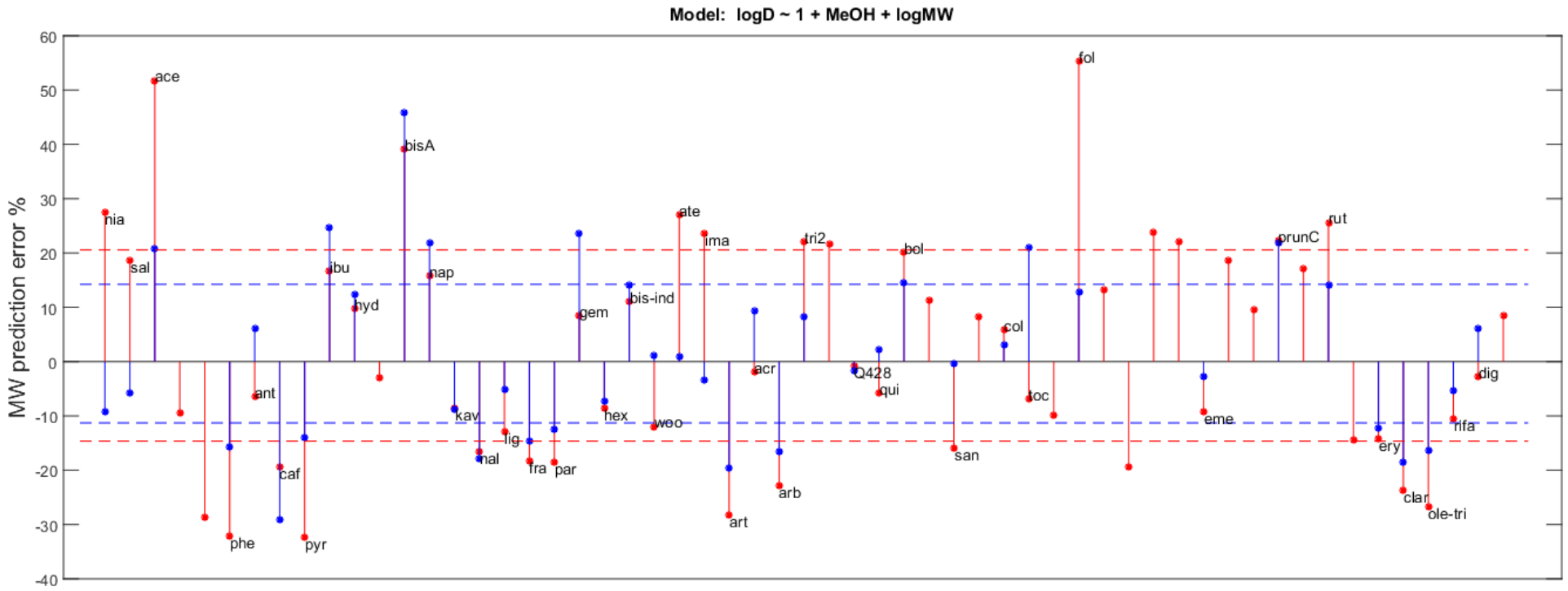


Fig. S6 Residual plot of MW_{pre} errors using Model 1 (red) and Model 3 (blue), with average MW_{pre} error range displayed by dashed lines.

Table S5. Model 3 (n=41, R^2=0.908, adjusted R^2=0.903)

Coefficients	Estimate	SE	tStat	pValue
Intercept	-8.0979	0.0789	-102.5700	4.75E-48
MeOH%	0.1906	0.0354	5.3831	3.99E-06
logMW	-0.6497	0.0337	-19.2890	3.17E-21

Table S6. Model 4 Data and NPs tested using Model 4

#	Compound	code	MW	$D\text{ref}_{TTMS}$	Polarity	Shape	%acid/ ArOH	%NHBD	%Br	%AHBD	HeavyO	$D_{model\ 4}$	% error
1	Niacin	nia	123.1	3.817E-10	0.386	0.667	0.138	0.000	0.000	0.000	1	3.659E-10	-4.1
2	Salicyclic acid	sal	138.1	3.714E-10	0.378	0.600	0.123	0.000	0.000	0.000	2	3.600E-10	-3.1
3	Acetaminophen	ace	151.2	3.013E-10	0.311	0.727	0.112	0.099	0.000	0.000	1	3.024E-10	0.4
4	Menthol	men	156.3	4.071E-10	0.097	0.636	0.000	0.000	0.000	0.109	0	4.247E-10	4.3
5	phenanthrene	phe	178.2	4.489E-10	0.000	0.571	0.000	0.000	0.000	0.000	0	4.146E-10	-7.6
6	2-aminoanthracene	ant	193.2	3.497E-10	0.101	0.600	0.000	0.078	0.000	0.000	0	3.496E-10	0.0
7	Caffeine	caf	194.2	3.822E-10	0.363	0.500	0.000	0.000	0.000	0.000	2	3.692E-10	-3.4
8	Pyrene	pyr	202.3	4.155E-10	0.000	0.500	0.000	0.000	0.000	0.000	0	3.863E-10	-7.0
9	ibuprofen	ibu	206.3	2.924E-10	0.151	0.667	0.082	0.000	0.000	0.000	1	2.984E-10	2.1
10	7-hydroxydictamnine	hyd	215.2	2.959E-10	0.302	0.500	0.079	0.000	0.000	0.000	2	2.985E-10	0.9
11	Farnesol	far	222.4	3.130E-10	0.062	0.813	0.000	0.000	0.000	0.076	0	3.147E-10	0.5
12	bisphenol A	bisA	228.3	2.460E-10	0.145	0.647	0.149	0.000	0.000	0.000	0	2.412E-10	-2.0
13	naproxen	nap	230.3	2.742E-10	0.202	0.647	0.074	0.000	0.000	0.000	2	2.818E-10	2.8
14	Kavain	kav	230.3	3.177E-10	0.176	0.647	0.000	0.000	0.000	0.000	3	3.333E-10	4.9
15	nalidixic acid	nal	232.2	3.349E-10	0.309	0.529	0.000	0.000	0.000	0.000	3	3.307E-10	-1.2
16	lignocaine	lig	234.3	3.239E-10	0.138	0.588	0.000	0.064	0.000	0.000	1	3.096E-10	-4.4
17	fragrolide	fra	248.3	3.253E-10	0.200	0.444	0.000	0.000	0.000	0.000	3	3.340E-10	2.7
18	Parthenolide	par	248.3	3.256E-10	0.210	0.444	0.000	0.000	0.000	0.000	3	3.330E-10	2.3
19	gemfibrozil	gem	250.3	2.712E-10	0.174	0.611	0.068	0.000	0.000	0.000	2	2.733E-10	0.8
20	hexazinone	hex	252.3	3.002E-10	0.228	0.556	0.000	0.000	0.000	0.000	2	3.110E-10	3.6
21	bis-indole-methanone	bis-ind	260.3	2.607E-10	0.206	0.550	0.000	0.115	0.000	0.000	1	2.711E-10	4.0
22	woodsianone B	woo	266.3	2.971E-10	0.283	0.474	0.000	0.000	0.000	0.000	4	3.115E-10	4.8
23	Atenolol	ate	266.3	2.364E-10	0.288	0.737	0.000	0.113	0.000	0.064	2	2.461E-10	4.1
24	imazapic	ima	275.3	2.357E-10	0.354	0.500	0.062	0.054	0.000	0.000	2	2.409E-10	2.2
25	Artemisinin	art	282.3	3.256E-10	0.276	0.400	0.000	0.000	0.000	0.000	5	3.110E-10	-4.5
26	acrophylline [TFA]	acr*	397.3	2.673E-10	0.181	0.476	0.000	0.000	0.000	0.000	5	2.424E-10	-9.3

27	arborinine	arb	285.3	3.092E-10	0.236	0.476	0.000	0.000	0.000	0.000	4	2.999E-10	-3.0
28	Trimethoprim	tri2	290.3	2.299E-10	0.362	0.571	0.000	0.103	0.000	0.000	3	2.475E-10	7.7
29	Lamourouic acid [TFA]	Lam*	308.2	2.219E-10	0.356	0.643	0.055	0.049	0.000	0.000	4	2.224E-10	0.2
30	toddalolactone	todd	308.3	2.520E-10	0.298	0.500	0.000	0.000	0.000	0.110	4	2.694E-10	6.9
31	Quinine	qui	324.4	2.520E-10	0.152	0.542	0.000	0.000	0.000	0.052	1	2.581E-10	2.4
32	Boldine	bol	327.4	2.153E-10	0.208	0.417	0.104	0.000	0.000	0.000	2	2.201E-10	2.2
33	Aaptamine [TFA]	aap*	328.2	2.255E-10	0.232	0.500	0.052	0.046	0.000	0.000	3	2.290E-10	1.6
34	Sanguinarine	san	349.3	2.585E-10	0.188	0.480	0.000	0.000	0.000	0.000	4	2.615E-10	1.2
35	Penicillin G	pen	334.4	2.121E-10	0.366	0.565	0.051	0.045	0.000	0.000	3	2.132E-10	0.5
36	Colchicine	col	399.4	2.058E-10	0.251	0.414	0.000	0.038	0.000	0.000	5	2.330E-10	13.2
37	α-tocopherol	toc	430.7	2.126E-10	0.062	0.645	0.039	0.000	0.000	0.000	1	1.946E-10	-8.5
38	quinine [TFA]	qui*	438.4	2.147E-10	0.197	0.542	0.000	0.000	0.000	0.039	3	2.109E-10	-1.8
39	Folic acid	fol	441.4	1.523E-10	0.504	0.625	0.077	0.136	0.000	0.000	4	1.451E-10	-4.7
40	boldine [TFA]	bol*	441.4	1.855E-10	0.252	0.417	0.077	0.000	0.000	0.000	4	1.894E-10	2.1
41	Sanguinarine [TFA]	san*	445.3	2.279E-10	0.188	0.480	0.000	0.000	0.000	0.000	6	2.255E-10	-1.1
42	Tobramycin	tob	467.5	1.693E-10	0.573	0.469	0.000	0.160	0.000	0.182	4	1.565E-10	-7.6
43	rigidin	rig	363.3	1.686E-10	0.397	0.481	0.093	0.124	0.000	0.000	3	1.746E-10	3.6
44	Emetine	eme	480.6	2.019E-10	0.146	0.486	0.000	0.031	0.000	0.000	4	2.027E-10	0.4
45	Doxycycline	doxy	480.9	1.707E-10	0.449	0.375	0.035	0.031	0.000	0.071	5	1.793E-10	5.0
46	Fusidic acid	fus	516.7	1.672E-10	0.195	0.432	0.033	0.000	0.000	0.066	3	1.804E-10	7.9
47	Prunolide C	prunC	574.5	1.499E-10	0.274	0.349	0.118	0.000	0.000	0.000	5	1.483E-10	-1.1
48	Rutin	rut	610.5	1.420E-10	0.488	0.419	0.084	0.000	0.000	0.167	7	1.387E-10	-2.4
49	Oleandomycin	ole	687.9	1.676E-10	0.285	0.313	0.000	0.000	0.000	0.074	9	1.730E-10	3.2
50	erythromycin	ery	733.9	1.606E-10	0.285	0.294	0.000	0.000	0.000	0.116	8	1.617E-10	0.7
51	clarithromycin	clar	748.0	1.707E-10	0.272	0.288	0.000	0.000	0.000	0.091	9	1.639E-10	-4.0
52	Oleandomycin triacetate	ole-tri	814.0	1.660E-10	0.286	0.316	0.000	0.000	0.000	0.000	15	1.705E-10	2.7
53	Rifampicin	rifa	823.0	1.456E-10	0.284	0.373	0.021	0.018	0.000	0.041	9	1.414E-10	-2.9
54	Digitonin	dig	1229.3	1.078E-10	0.427	0.388	0.000	0.000	0.000	0.235	12	1.067E-10	-1.0
55	Vancomycin [HCl]	vanco	1485.7	8.942E-11	0.410	0.248	0.046	0.081	0.000	0.069	14	8.854E-11	-1.0
56	botrylamide	botC	404.3	2.048E-10	0.203	0.680	0.042	0.037	0.197	0.000	3	2.160E-10	5.5

57	5-bromo-indole-oxoacetate	5-BIO	282.1	3.250E-10	0.288	0.563	0.000	0.053	0.284	0.000	3	3.214E-10	-1.1
58	6-bromo-indole-oxoacetate	6-BIO	282.1	3.344E-10	0.288	0.625	0.000	0.053	0.284	0.000	3	3.150E-10	-5.8
59	Prunolide B	prunB	1047.9	1.410E-10	0.214	0.306	0.065	0.000	0.458	0.000	5	1.430E-10	1.5
60	Prunolide A	prunA	1205.7	1.402E-10	0.199	0.294	0.056	0.000	0.531	0.000	5	1.380E-10	-1.6
61	Procerolide B	procB	563.0	2.087E-10	0.185	0.500	0.031	0.000	0.426	0.000	4	2.216E-10	6.2
62	ehinosulfone A	echA	498.1	2.338E-10	0.282	0.500	0.000	0.029	0.308	0.000	4	2.273E-10	-2.8
63	bis(6-bromoindole)-methanone	Br-bis-ind	418.1	2.449E-10	0.174	0.591	0.000	0.072	0.383	0.000	1	2.487E-10	1.5
64	Aerophobin-2 [TFA]	AERAO	619.2	1.699E-10	0.368	0.630	0.000	0.048	0.258	0.000	4	1.726E-10	1.5
65	19-bromoisoeudistomin	PL01O	478.3	2.021E-10	0.167	0.522	0.000	0.063	0.167	0.000	2	2.078E-10	2.8
66	aplysamine 2 [TFA]	APL 2	764.2	1.661E-10	0.183	0.667	0.022	0.020	0.314	0.000	5	1.592E-10	-4.1
67	Aerothionin	AERNO	818.1	1.520E-10	0.315	0.600	0.000	0.037	0.391	0.007	6	1.624E-10	6.8
68	spathulenol	spa	220.4	3.397E-10	0.080	0.438	0.000	0.000	0.000	0.077	0	3.553E-10	4.6
69	cyclocolorenone	cyc	218.3	3.720E-10	0.079	0.438	0.000	0.000	0.000	0.000	1	3.696E-10	-0.6
70	convolutamine K	CONK	408.2	2.646E-10	0.066	0.650	0.000	0.000	0.392	0.000	1	2.755E-10	4.1
71	Convolutamine K [TFA]	CONK*	636.2	1.959E-10	0.066	0.650	0.000	0.000	0.251	0.000	5	1.944E-10	-0.8
72	Convolutamine L	CONL	323.0	3.264E-10	0.117	0.643	0.000	0.000	0.495	0.000	1	3.440E-10	5.4
73	Convolutamine L [TFA]	CONL*	437.0	2.508E-10	0.117	0.643	0.000	0.000	0.366	0.000	3	2.629E-10	4.8
74	volutamine K	volK	828.3	1.773E-10	0.093	0.585	0.000	0.000	0.386	0.000	3	1.700E-10	-4.1

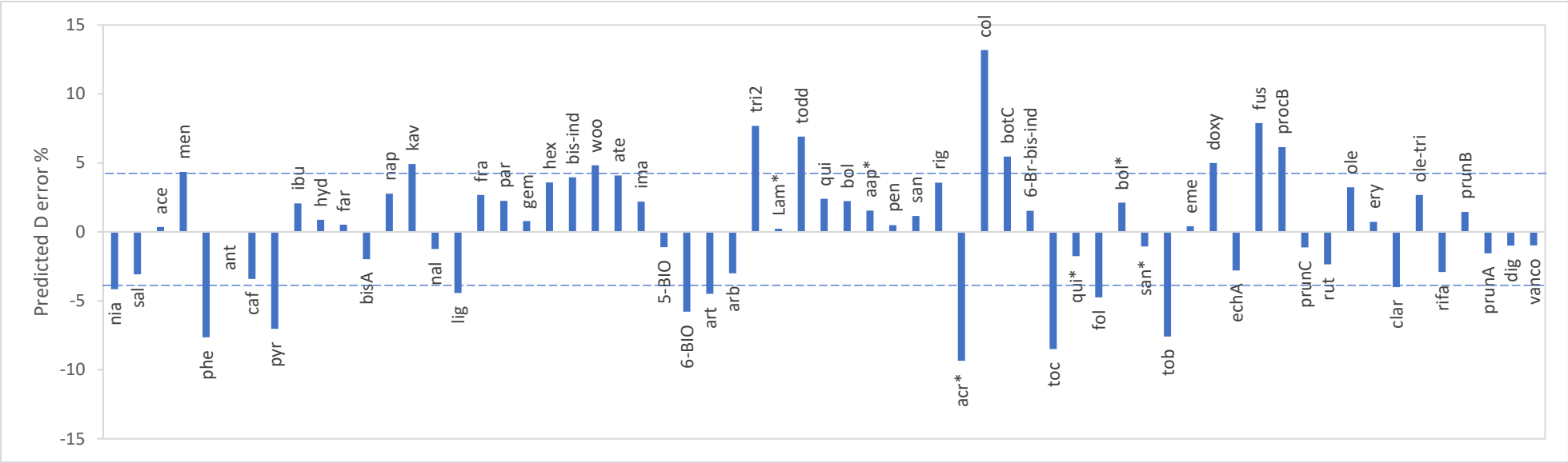
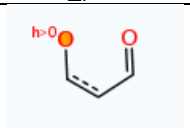
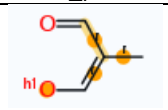
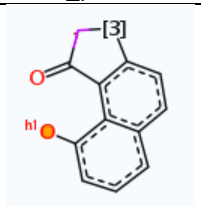
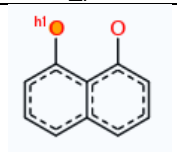
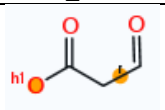


Figure S7. Residual plot of predicted diffusion coefficient (D_{pre}) error % from experimental *D*, with dashed lines showing the average error range.

Table S7. Model 4 (n=63, R^2=0.986, adjusted R^2=0.984)

Coefficients	Estimate	SE	tStat	pValue
Intercept	-7.6365	0.0745	-102.5460	1.45E-63
logMW	-0.7403	0.0265	-27.9582	8.46E-34
Shape	-0.1390	0.0343	-4.0528	1.63E-04
HeavyO	0.00685	0.0018	3.7613	4.18E-04
%phenol/acid	-0.8506	0.0696	-12.2143	3.63E-17
%Br	0.25862	0.0257	10.0676	5.38E-14
%NHBD	-0.4016	0.0806	-4.9855	6.75E-06
%AHBD	-0.0947	0.0647	-1.4638	1.49E-01
Polarity	-0.1282	0.0357	-3.5966	6.99E-04

Table S8. Substructures used to detect intramolecular H-bonding with DataWarrior, then used to calculate ratio of free phenols and carboxylic acids

	Phenol IMHB				Carboxylic acid IMHB	peptide
	IMHB_pA	IMHB_pB	IMHB_pC	IMHB_pD	IMHB_c	Cyclic amide
						
Salicylic acid	1					
Nalidixic acid					1	
Arborinine	1					
Doxycycline hyclate		1				
Rutin	1					
Rifampicin			1	1		

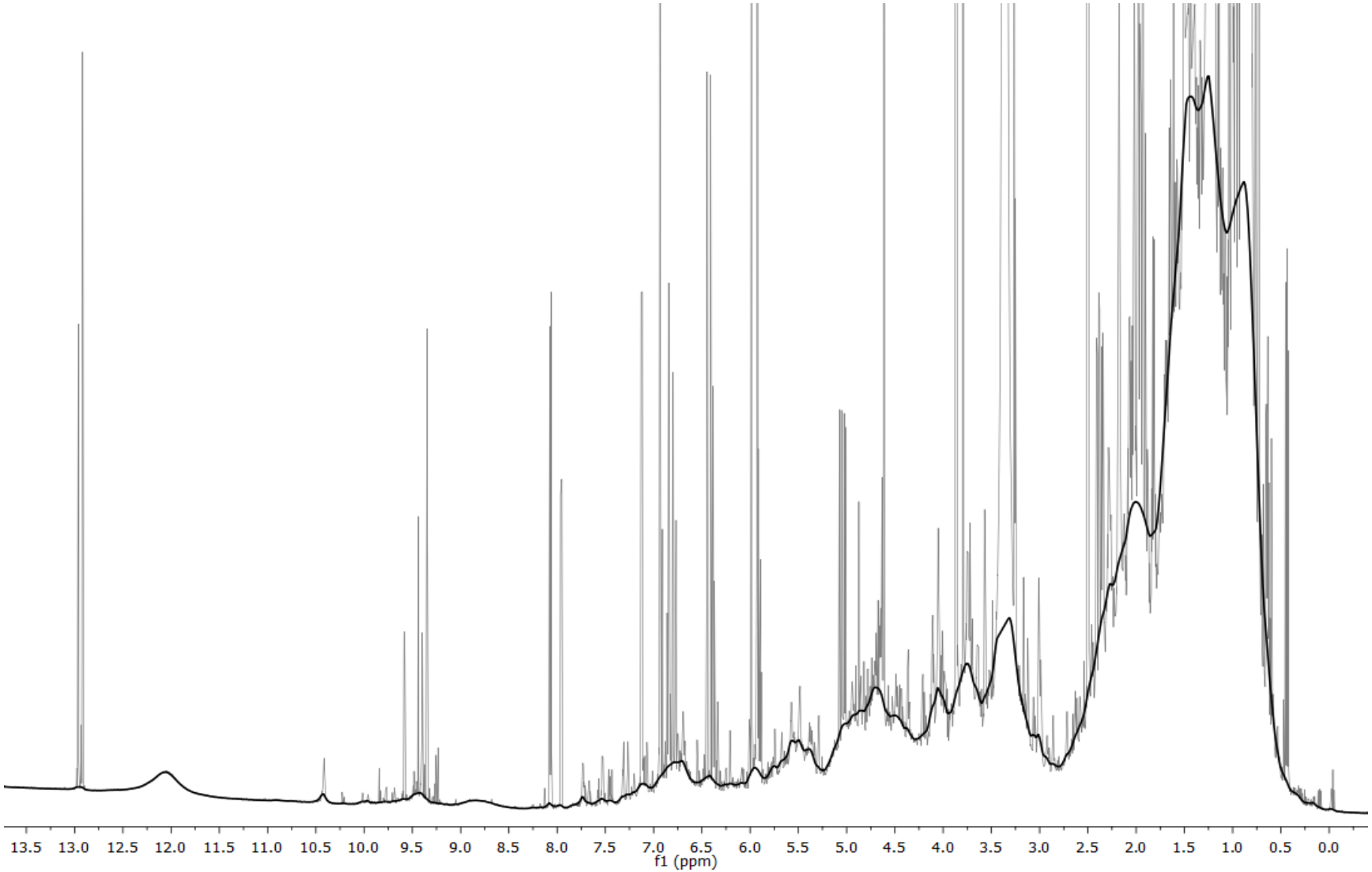


Fig. S8 The baseline shape of RP HPLC fraction 47 at increased intensity.

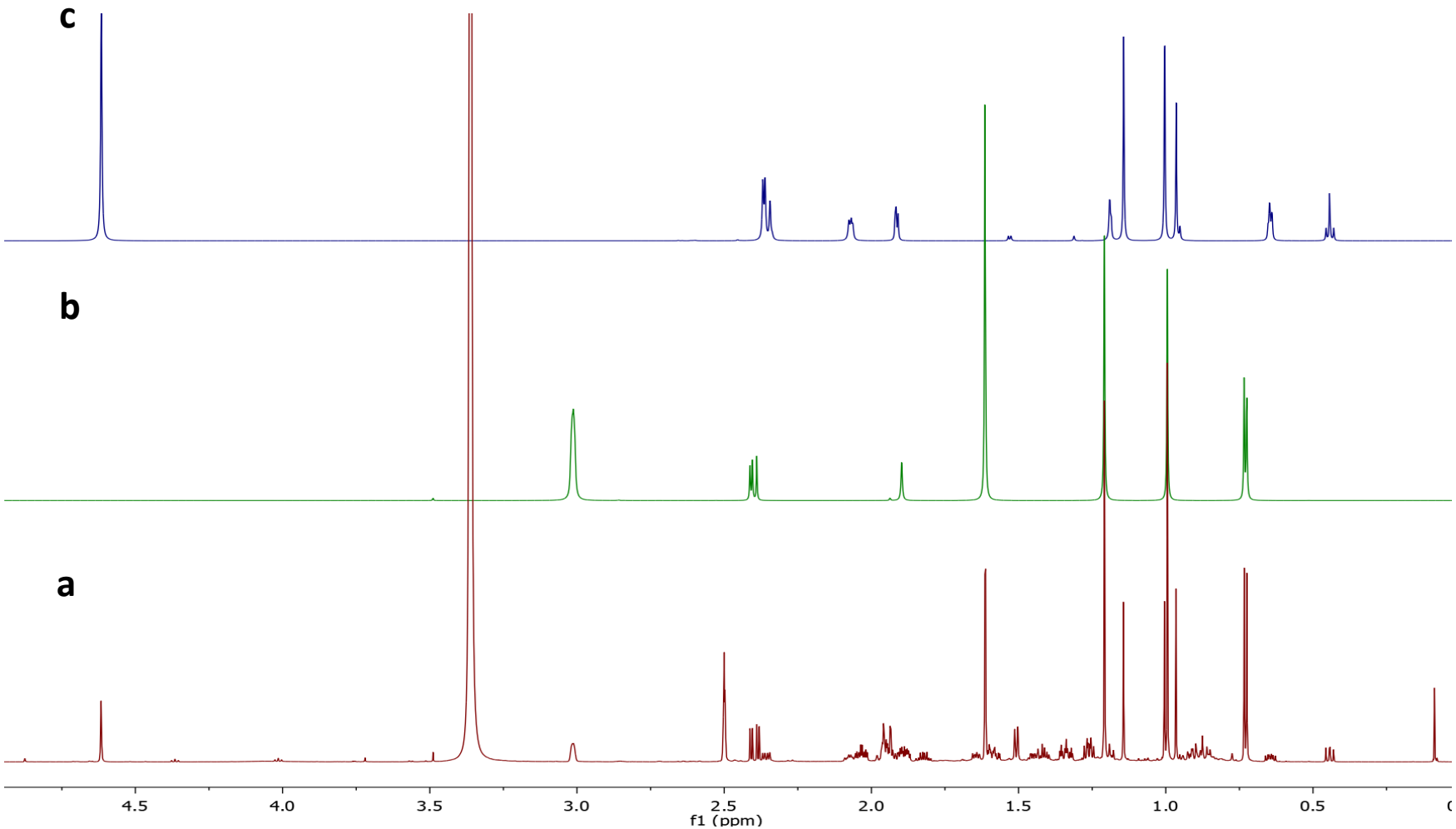


Fig. S9 ^{1}H NMR of F10 from NP separation of *Tasmannia xerophila* extract (a) and projections at D = 3.72014E^{-10} (b) and 3.39743E^{-10} (c) from ^{1}H DOSY spectrum of F10 in DMSO-d_6 at 298 K.

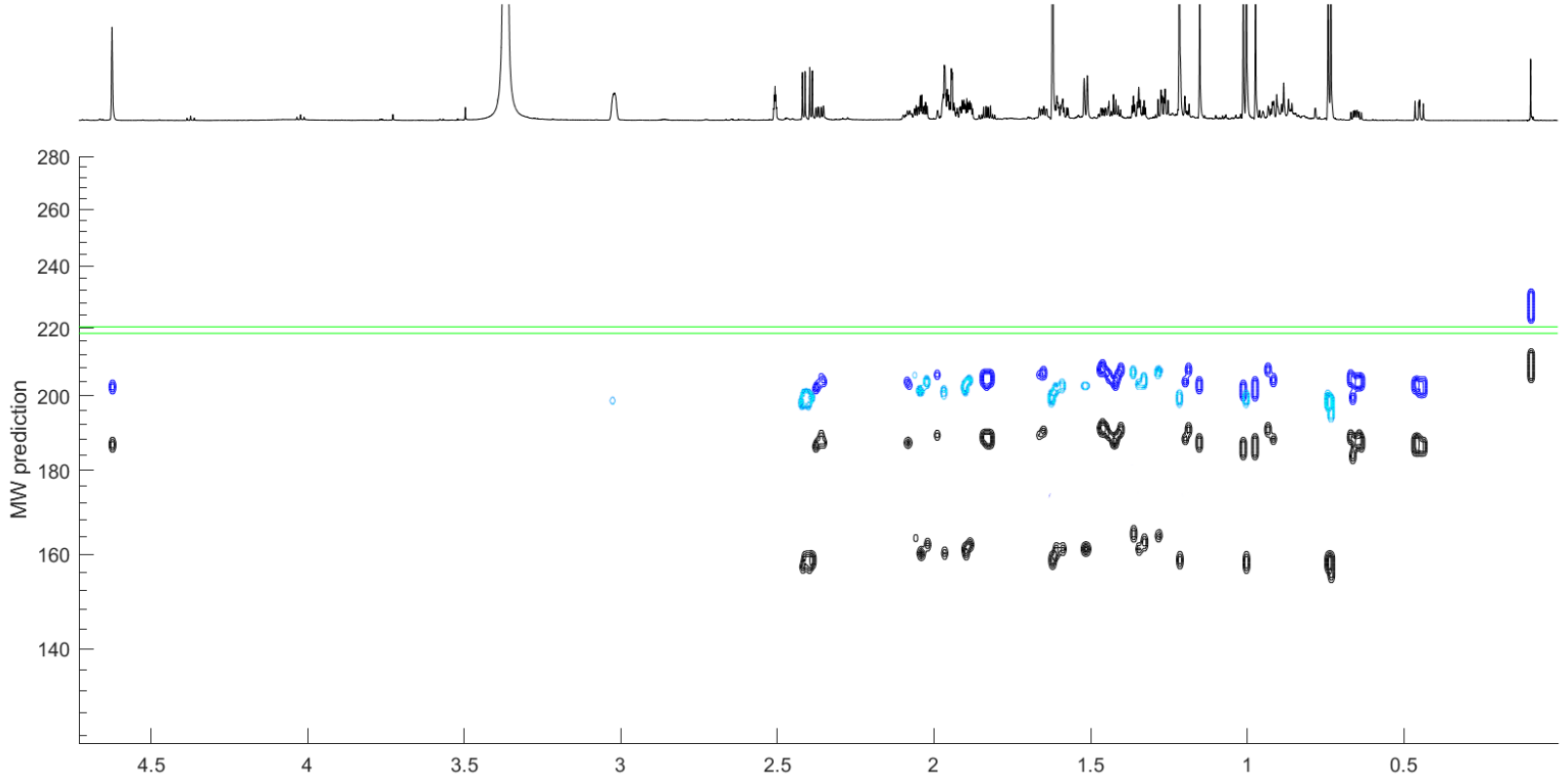


Fig. S10 ^{1}H DOSY spectrum of F10, MW prediction by model 1 (black) and 2 (blue) with $^{eff}\delta_H$ = 3.95 for the slower diffusing compound (spathulenol (**68**) dark blue) and $^{eff}\delta_H$ = 0.00 for the faster diffusing compound (cyclocorolenone (**69**) light blue). True MWs values are represented by green lines.

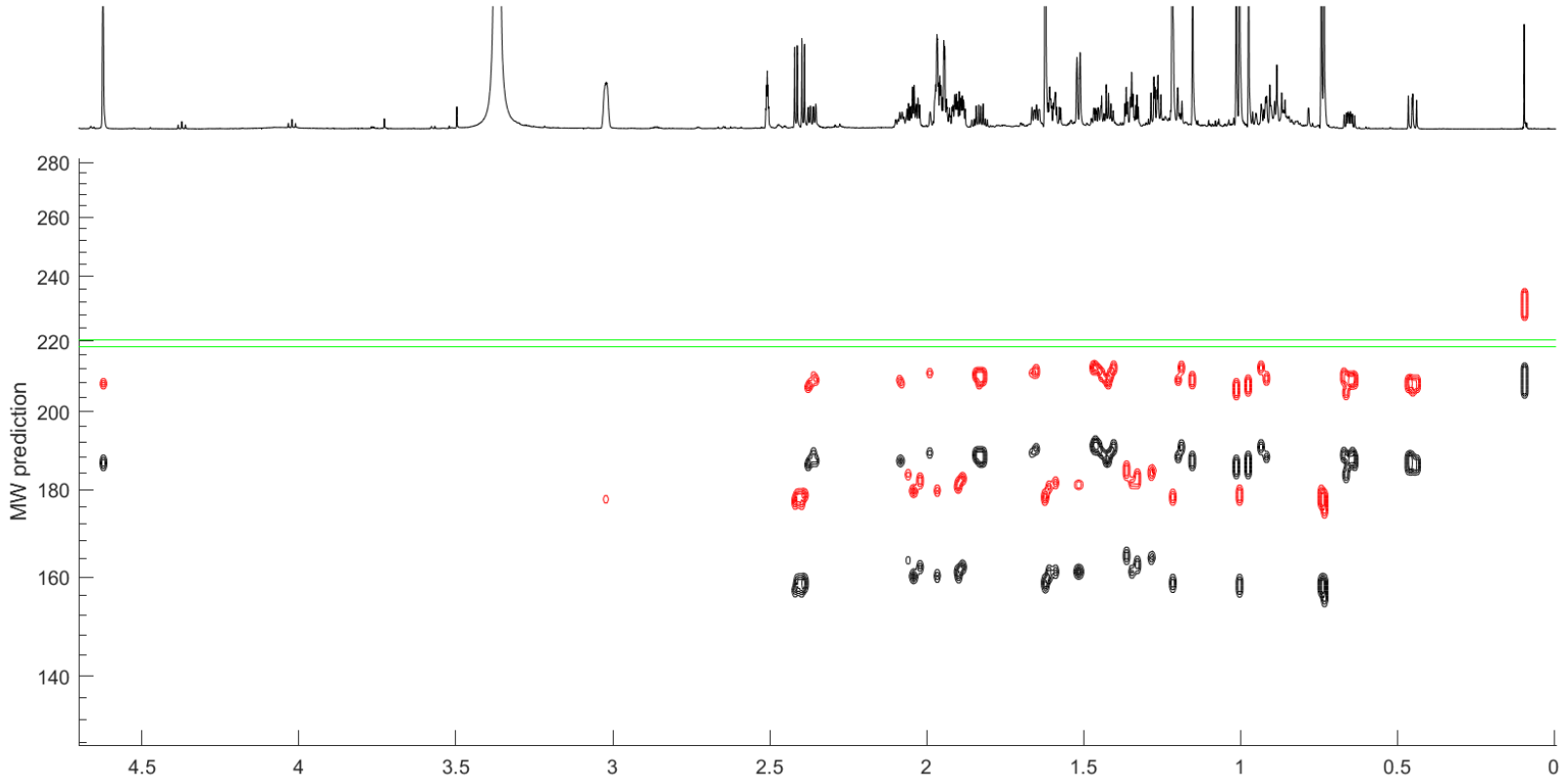
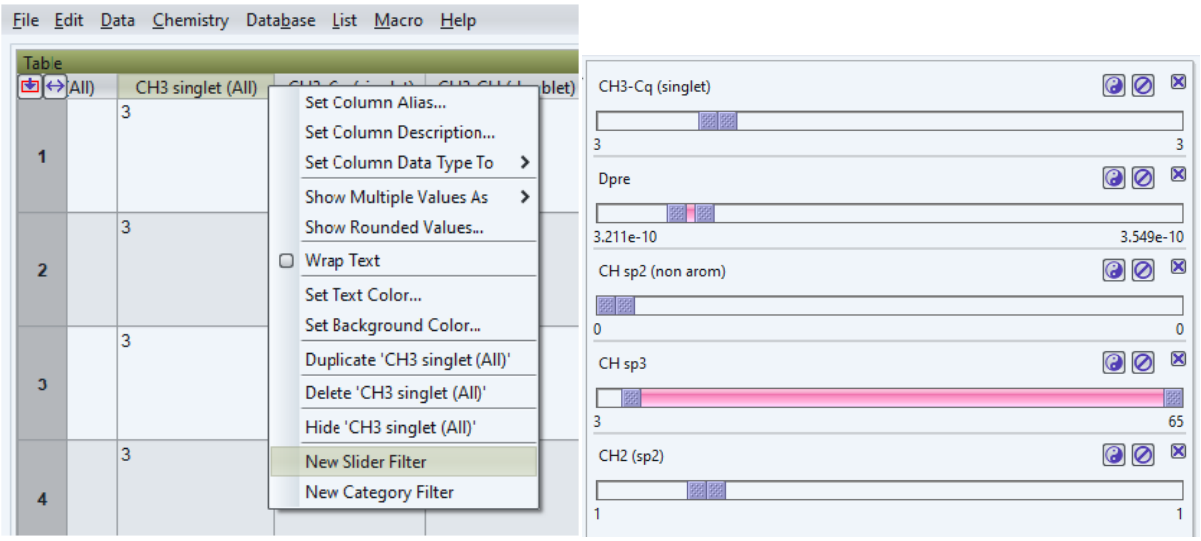


Fig. S11 ^{1}H DOSY spectrum of F10, MW prediction by model 3 with a MeOH% variable of 70% Black resonances (model 1) red resonances (model 3). True MWs values are represented by green lines.

Scheme for DOSY-based dereplication:

* Models 2-4 are can only be used with DOSY data acquired in DMSO-d_6 at 298 °K

1. Acquire DOSY data with an internal standard.
2. Determine average *D* for compound.
3. Reference the DOSY data (for TTMS: D_{stand}=3.157e-10, for DMSO: D_{stand}=6.670e-10):
   $$D = \frac{D_{\text{stand}}}{D_{\text{ref}}} \cdot D_{\text{comp}}$$
4. Using the DEREP-NP database in DataWarrior, add structural filters based on NMR data by selecting a structural feature and adding it as a slider or category filter, and a predicted *D* (D_{pre}) filter (with an error range) based on the average experimental *D* of the compound:



On the right are the structural and D_{pre} filters added for the slower diffusing compound in our NP HPLC fraction.