CHAPTER

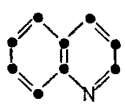
VI

Cluster Analysis and the Design of Congener Sets

VI-1 POSSIBILITIES OF MOLECULAR MODIFICATION

As the study of structure-activity relationships develops and especially as work in the biomedicinal chemical field gains momentum, the need to deal with many more variables becomes pressing. From an economic standpoint, the probelm of ruling out irrelevant parameters and focusing on the relevant ones early in a structure-activity study is one of extreme importance. This problem is both crucial and complex and deserves much more systematic attention than has been customarily allotted to it in the past.

It has been pointed out[1] that the number of derivatives that can be made from a set of N substituents where m is the number of nonsymmetric positions on the parent compound is N^m; for example, if one were making derivatives of quinoline using the 166 well-char-



acterized substituents of Table VI-1 in all possible combinations, this would amount to 166^7 or approximately 3.5×10^{15} molecules. Of course, 166 is a small fraction of the almost 2000 substituents in Appendix I. What constitutes a reasonable sample of 10^{15} congeners? Even if we made only one-billionth of the possibilities, it would amount to one million molecules; yet relatively few drug modification programs make as many as a thousand derivatives.

A general formula for calculating the possibilities is

$$X^k \cdot \frac{n!}{k!(n-k)!}$$

In this expression, X is the number of substituents to be considered, n is the total number of nonsymmetric positions on the parent molecule, and k is the number of substituents to be placed on the parent compound at one time. With this formula we can consider simpler and more varied cases. For example, using only 100 substituents from the 2000 of Appendix I and considering only three out of the seven possible positions on quinoline leads to 35,000,000 analogs. With 100 substituents and only two positions, one still has to face 210,000 possibilities. Even considering only 20 substituents, two at a time, means 8400 possibilities. No wonder that "me too" drugs are always being developed.

VI-2 THE COLLINEARITY PROBLEM

Since the cost of modifying a parent structure is so great and since the possibilities are so enormous, one wants to

Table VI-1 Well-Characterized[a] Aromatic Substituents[b]

	π	H-Accpt	H-Donor	MR	$\mathcal{F}$	$\mathcal{R}$	σ_m	σ_p
Br	0.86	0	0	8.88	0.44	−0.17	0.39	0.23
Cl	0.71	0	0	6.03	0.41	−0.15	0.37	0.23
F	0.14	0	0	0.92	0.43	−0.34	0.34	0.06
SO_2F	0.05	1	0	8.65	0.75	0.22	0.80	0.91
SF_5	1.23	0	0	9.89	0.57	0.15	0.61	0.68
I	1.12	0	0	13.94	0.40	−0.19	0.35	0.18
IO_2	−3.46	1	0	63.51	0.63	0.20	0.68	0.78
NO	−1.20	1	0	5.20	0.50	0.45	0.62	0.91
NO_2	−0.28	1	0	7.36	0.67	0.16	0.71	0.78
NNN	0.46	0	0	10.20	0.30	−0.13	0.27	0.15
H	0.00	0	0	1.03	0.00	0.00	0.00	0.00
OH	−0.67	1	1	2.85	0.29	−0.64	0.12	−0.37
SH	0.39	0	1	9.22	0.28	−0.11	0.25	0.15
$B(OH)_2$	−0.55	1	1	11.04	−0.07	0.18	−0.01	0.12
NH_2	−1.23	1	1	5.42	0.02	−0.68	−0.16	−0.66
NHOH	−1.34	1	1	7.22	0.06	−0.40	−0.04	−0.34
SO_2NH_2	−1.82	1	1	12.28	0.41	0.19	0.46	0.57
$NHNH_2$	−0.88	1	1	8.44	0.17	−0.71	−0.02	−0.55
5-Cl-1-Tetrazolyl	−0.65	1	0	23.16	0.58	0.07	0.60	0.61
$N{=}CCl_2$	0.41	0	0	18.35	0.23	−0.08	0.21	0.13
CF_3	0.88	0	0	5.02	0.38	0.19	0.43	0.54
OCF_3	1.04	1	0	7.86	0.38	0.00	0.38	0.35
SO_2CF_3	0.55	1	0	12.86	0.73	0.26	0.79	0.93
SCF_3	1.44	0	0	13.81	0.35	0.18	0.40	0.50
CN	−0.57	1	0	6.33	0.51	0.19	0.56	0.66
NCS	1.15	1	0	17.24	0.51	−0.09	0.48	0.38
SCN	0.41	1	0	13.40	0.36	0.19	0.41	0.52
CO_2^-	−4.36	1	0	6.05	−0.15	0.13	−0.10	0.00
1-Tetrazolyl	−1.04	1	0	18.33	0.52	0.02	0.52	0.50
NHCN	−0.26	1	1	10.14	0.26	−0.18	0.21	0.06
CHO	−0.65	1	0	6.88	0.31	0.13	0.35	0.42
CO_2H	−0.32	1	1	6.93	0.33	0.15	0.37	0.45
CH_2Br	0.79	0	0	13.39	0.10	0.05	0.12	0.14
CH_2Cl	0.17	0	0	10.49	0.10	0.03	0.11	0.12
CH_2I	1.50	0	0	18.60	0.09	0.03	0.10	0.11
NHCHO	−0.98	1	1	10.31	0.25	−0.23	0.19	0.00
$CONH_2$	−1.49	1	1	9.81	0.24	0.14	0.28	0.36
CH=NOH	−0.38	1	1	10.28	0.25	−0.13	0.22	0.10
CH_3	0.56	0	0	5.65	−0.04	−0.13	−0.07	−0.17
$NHCONH_2$	−1.30	1	1	13.72	0.04	−0.28	−0.03	−0.24
$NHC{=}S(NH_2)$	−1.40	1	1	22.19	0.23	−0.05	0.22	0.16
OCH_3	−0.02	1	0	7.87	0.26	−0.51	0.12	−0.27
CH_2OH	−1.03	1	1	7.19	0.00	0.00	0.00	0.00
$SOCH_3$	−1.58	1	0	13.70	0.52	0.01	0.52	0.49
SO_2CH_3	−1.63	1	0	13.49	0.54	0.22	0.60	0.72
OSO_2CH_3	−0.88	1	0	16.99	0.39	0.00	0.39	0.36
SCH_3	0.61	0	0	13.82	0.20	−0.18	0.15	0.00
$SeCH_3$	0.74	0	0	17.03	0.13	−0.12	0.10	0.00
$NHCH_3$	−0.47	1	1	10.33	−0.11	−0.74	−0.30	−0.84
$NHSO_2CH_3$	−1.18	1	1	18.17	0.25	−0.20	0.20	0.03
CF_2CF_3	1.68	0	0	9.23	0.44	0.11	0.47	0.52
C≡CH	0.40	0	0	9.55	0.19	0.05	0.21	0.23
$NHCOCF_3$	0.08	1	1	14.30	0.36	−0.21	0.30	0.12
CH_2CN	−0.57	1	0	10.11	0.21	−0.18	0.16	0.01

Table VI-1 Well-Characterized[a] Aromatic Substituents[b] (Continued)

	π	H-Accpt	H-Donor	MR	$\mathcal{F}$	$\mathcal{R}$	σ_m	σ_p
$CH{=}CHNO_2$-(trans)	0.11	1	0	16.42	0.33	−0.05	0.32	0.26
$CH{=}CH_2$	0.82	0	0	10.99	0.07	−0.08	0.05	−0.02
$NHC{=}O(CH_2Cl)$	−0.50	1	1	19.77	0.23	−0.25	0.17	−0.03
$COCH_3$	−0.55	1	0	11.18	0.32	0.20	0.38	0.50
$SCOCH_3$	0.10	1	0	18.42	0.36	0.11	0.39	0.44
$OCOCH_3$	−0.64	1	0	12.47	0.41	−0.07	0.39	0.31
CO_2CH_3	−0.01	1	0	12.87	0.33	0.15	0.37	0.45
$NHCOCH_3$	−0.97	1	1	14.93	0.28	−0.26	0.21	0.00
$NHCO_2CH_3$*	−0.37	1	1	16.53	0.14	−0.28	0.07	−0.15
$C{=}O(NHCH_3)$	−1.27	1	1	14.57	0.34	0.05	0.35	0.36
$CH{=}NOCH_3$	0.40	1	0	14.93	0.39	−0.06	0.37	0.30
$NHC{=}S(CH_3)$	−0.42	1	1	23.40	0.27	−0.13	0.24	0.12
$CH{=}NNHC{=}S(NH_2)$	−0.27	1	1	29.92	0.46	−0.02	0.45	0.40
CH_2CH_3	1.02	0	0	10.30	−0.05	−0.10	−0.07	−0.15
$CH{=}NNHCONHNH_2$	−1.32	1	1	24.86	0.23	−0.05	0.22	0.16
CH_2OCH_3	−0.78	1	0	12.07	0.01	0.02	0.02	0.03
OCH_2CH_3	0.38	1	0	12.47	0.22	−0.44	0.10	−0.24
SOC_2H_5*	−1.04	1	0	18.35	0.52	0.01	0.52	0.49
SC_2H_5	1.07	0	0	18.42	0.23	−0.18	0.18	0.03
SeC_2H_5*	1.28	0	0	21.68	0.13	−0.12	0.10	0.00
NHC_2H_5	0.08	1	1	14.98	−0.11	−0.51	−0.24	−0.61
$SO_2C_2H_5$*	−1.09	1	0	18.14	0.54	0.22	0.60	0.72
$N(CH_3)_2$	0.18	1	0	15.55	0.10	−0.92	−0.15	−0.83
$NHSO_2C_2H_5$*	−0.64	1	1	22.82	0.25	−0.20	0.20	0.03
$P(CH_3)_2$	0.44	0	0	21.19	−0.08	0.39	0.03	0.31
$PO(OCH_3)_2$	−1.18	1	0	21.87	0.37	0.19	0.42	0.53
$C(OH)(CF_3)_2$	1.28	1	1	15.18	0.28	0.05	0.29	0.30
$CH{=}CHCN$	−0.17	1	0	16.23	0.26	−0.07	0.24	0.17
Cyclopropyl	1.14	0	0	13.53	−0.03	−0.19	−0.07	−0.21
COC_2H_5*	0.06	1	0	15.83	0.32	0.20	0.38	0.50
$SCOC_2H_5$*	0.64	1	0	23.07	0.36	0.11	0.39	0.44
$CO_2C_2H_5$	0.51	1	0	17.47	0.33	0.15	0.37	0.45
$OCOC_2H_5$*	−0.10	1	0	17.12	0.41	−0.07	0.39	0.31
$CH_2CH_2CO_2H$	−0.29	1	1	16.52	−0.02	−0.05	−0.03	−0.07
$NHCO_2C_2H_5$	0.17	1	1	21.18	0.14	−0.28	0.07	−0.15
$CONHC_2H_5$*	−0.73	1	1	19.22	0.34	0.05	0.35	0.36
$NHCOC_2H_5$*	−0.43	1	1	19.58	0.28	−0.26	0.21	0.00
$CH{=}NOC_2H_5$*	0.94	1	0	19.58	0.39	−0.06	0.37	0.30
$NHC{=}S(C_2H_5)$*	0.12	1	1	28.05	0.27	−0.13	0.24	0.12
$CH(CH_3)_2$	1.53	0	0	14.96	−0.05	−0.10	−0.07	−0.15
C_3H_7	1.55	0	0	14.96	−0.06	−0.08	−0.07	−0.13
$NHC{=}S(NHC_2H_5)$	−0.71	1	1	31.66	0.38	−0.28	0.30	0.07
$OCH(CH_3)_2$	0.85	1	0	17.06	0.30	−0.72	0.10	−0.45
OC_3H_7	1.05	1	0	17.06	0.22	−0.45	0.10	−0.25
$CH_2OC_2H_5$*	−0.24	1	0	16.72	0.01	0.02	0.02	0.03
SOC_3H_7*	−0.50	1	0	23.00	0.52	0.01	0.52	0.49
$SO_2C_3H_7$*	−0.55	1	0	22.79	0.54	0.22	0.60	0.72
SC_3H_7*	1.61	0	0	23.07	0 23	−0.18	0.15	0.00
SeC_3H_7*	1.82	0	0	26.33	0.13	−0.12	0.10	0.00
NHC_3H_7*	0.62	1	1	19.63	−0.11	−0.51	−0.24	−0.61
$NHSO_2C_3H_7$*	−0.10	1	1	27.47	0.25	−0.20	0.20	0.03
$\overset{+}{N}(CH_3)_3$	−5.96	0	0	21.20	0.89	0.00	0.88	0.82
$Si(CH_3)_3$	2.59	0	0	24.96	−0.04	−0.04	−0.04	−0.07
$CH{=}C(CN)_2$	0.05	1	0	21.53	0.58	0.30	0.66	0.84

Table VI-1 Well-Characterized[a] Aromatic Substituents[b] (Continued)

	π	H-Accpt	H-Donor	MR	$\mathcal{F}$	$\mathcal{R}$	σ_m	σ_p
1-Pyrryl	0.95	1	0	21.85	0.50	−0.09	0.47	0.37
2-Thienyl	1.61	0	0	24.04	0.10	0.04	0.09	0.05
3-Thienyl	1.81	0	0	24.04	0.04	−0.06	0.03	−0.02
$CH{=}CHCOCH_3$	−0.06	1	0	21.10	0.28	−0.27	0.21	−0.01
$CH{=}CHCO_2CH_3$*	0.32	1	0	22.56	0.24	−0.19	0.19	0.03
COC_3H_7*	0.53	1	0	20.48	0.32	0.20	0.38	0.50
$SCOC_3H_7$*	1.18	1	0	27.72	0.36	0.11	0.39	0.44
$OCOC_3H_7$*	0.44	1	0	21.77	0.41	−0.07	0.39	0.31
$CO_2C_3H_7$*	1.07	1	0	22.17	0.33	0.15	0.37	0.45
$(CH_2)_3CO_2H$*	0.25	1	1	21.17	−0.02	−0.05	−0.03	−0.07
$CONHC_3H_7$*	−0.19	1	1	23.87	0.34	−0.05	0.35	0.36
$NHCOC_3H_7$*	0.11	1	1	24.23	0.28	−0.26	0.21	0.00
$NHC{=}OCH(CH_3)_2$	−0.18	1	1	24.23	0.18	−0.26	0.11	−0.10
$NHCO_2C_3H_7$*	0.71	1	1	25.83	0.14	−0.28	0.07	−0.15
$CH{=}NOC_3H_7$*	1.48	1	0	24.23	0.39	−0.06	0.37	0.30
$NHC{=}S(C_3H_7)$*	0.66	1	1	32.70	0.27	−0.13	0.24	0.12
C_4H_9	2.13	0	0	19.61	−0.06	−0.11	−0.08	−0.16
$C(CH_3)_3$	1.98	0	0	19.62	−0.07	−0.13	−0.10	−0.20
OC_4H_9	1.55	1	0	21.66	0.25	−0.55	0.10	−0.32
$CH_2OC_3H_7$*	0.30	1	0	21.37	0.01	0.02	0.02	0.03
$N(C_2H_5)_2$	1.18	1	0	24.85	0.01	−0.91	−0.23	−0.90
NHC_4H_9*	1.16	1	1	24.26	−0.28	−0.25	−0.34	−0.51
$P(C_2H_5)_2$*	1.52	0	0	30.49	−0.08	0.39	0.03	0.31
$PO(OC_2H_5)_2$*	−0.10	1	0	31.16	0.37	0.19	0.42	0.53
$CH_2Si(CH_3)_3$	2.00	0	0	29.61	−0.15	−0.07	−0.16	−0.21
$CH{=}CHCOC_2H_5$*	0.48	1	0	25.75	0.28	−0.27	0.21	−0.01
$CH{=}CHCO_2C_2H_5$	0.86	1	0	27.21	0.24	−0.19	0.19	0.03
$CH{=}NOC_4H_9$*	2.02	1	0	28.88	0.39	−0.06	0.37	0.30
C_5H_{11}*	2.67	0	0	24.26	−0.06	−0.08	−0.08	−0.16
$CH_2OC_4H_9$*	0.84	1	0	26.02	0.01	0.02	0.02	0.03
C_6H_5	1.96	0	0	25.36	0.08	−0.08	0.06	−0.01
$N{=}NC_6H_5$	1.69	0	0	31.31	0.28	0.13	0.32	0.39
OC_6H_5	2.08	1	0	27.68	0.34	−0.35	0.25	−0.03
$SO_2C_6H_5$	0.27	1	0	33.20	0.56	0.18	0.61	0.70
$OSO_2C_6H_5$	0.93	1	0	36.70	0.36	0.00	0.36	0.33
NHC_6H_5	1.37	1	1	30.04	−0.02	−0.38	−0.12	−0.40
$NHSO_2C_6H_5$	0.45	1	1	37.88	0.21	−0.18	0.16	0.01
2,5-di-Me-1-pyrryl	1.95	1	0	31.15	0.52	−0.10	0.49	0.38
$CH{=}CHCOC_3H_7$*	1.02	1	0	30.40	0.28	−0.27	0.21	−0.01
$CH{=}CHCO_2C_3H_7$*	1.40	1	0	31.86	0.24	−0.19	0.19	0.03
Cyclohexyl	2.51	0	0	26.69	−0.13	−0.10	−0.15	−0.22
2-Benzthiazolyl	2.13	1	0	38.88	0.25	0.06	0.27	0.29
COC_6H_5	1.05	1	0	30.33	0.30	0.16	0.34	0.43
$CO_2C_6H_5$	1.46	1	0	32.31	0.33	0.13	0.37	0.44
$OCOC_6H_5$	1.46	1	0	32.33	0.23	−0.08	0.21	0.13
$N{=}CHC_6H_5$	−0.29	1	0	33.01	0.09	−0.63	−0.08	−0.55
$CH{=}NC_6H_5$	−0.29	1	0	33.01	0.31	0.13	0.35	0.42
$NHCOC_6H_5$	0.49	1	1	34.64	0.09	−0.27	0.02	−0.19
$CH_2C_6H_5$	2.01	0	0	30.01	−0.08	−0.01	−0.08	−0.09
$CH_2OC_6H_5$	1.66	1	0	32.19	0.02	0.02	0.03	0.04
$C{\equiv}CC_6H_5$	2.65	0	0	33.21	0.12	0.05	0.14	0.16
$CH{=}NNHCOC_6H_5$	0.43	1	1	42.37	0.33	0.20	0.39	0.51
$CH_2Si(C_2H_5)_3$*	3.26	0	0	43.56	−0 15	−0.07	−0.16	−0.21
$CH{=}CHC_6H_5$-(trans)	2.68	0	0	34.17	0.06	−0.12	0.03	−0.07

Table VI-1 Well-Characterized[a] Aromatic Substituents[b] (Continued)

	π	H-Accpt	H-Donor	MR	$\mathcal{F}$	$\mathcal{R}$	σ_m	σ_p
$CH{=}CHCOC_6H_5$	0.95	1	0	40.25	0.22	−0.15	0.18	0.05
Ferrocenyl	2.46	0	0	48.24	−0.15	−0.04	−0.15	−0.18
$N(C_6H_5)_2$	3.61	1	0	54.96	0.07	−0.29	0.00	−0.22
$P{=}O(C_6H_5)_2$	0.70	1	0	59.29	0.31	0.24	0.38	0.53

[a]By well-characterized, we mean that the set of eight constants is known for each substituent; we do not mean to imply that all of the constants are of the highest accuracy.
[b]Substituents are ordered first by number of C, then by number of H, and the remaining elements alphabetically.

Table VI-2 Well-Characterized[a] Aliphatic Substituents

	Fr	H-Accpt	H-Donor	MR	$\mathcal{F}$
Br	0.20	0	0	8.80	0.44
Cl	0.06	0	0	5.93	0.41
F	−0.38	0	0	1.05	0.43
I	0.59	0	0	13.76	0.40
NO_2	−1.16	1	0	6.71	0.67
H	0.23	0	0	1.03	0.00
OH	−1.64	1	1	2.55	0.29
SH	−0.23	0	1	8.76	0.28
NH_2	−1.54	1	1	4.37	0.02
CBr_3	2.03	0	0	28.81	0.27
CCl_3	1.61	0	0	20.12	0.31
CF_3	0.29	0	0	5.02	0.38
CN	−1.27	1	0	5.39	0.51
SCN	−0.48	1	0	13.40	0.36
CO_2^-	−5.19	1	0	5.15	−0.15
CO_2H	−1.11	1	1	6.03	0.33
CH_2Br	0.74	0	0	13.39	0.10
CH_2Cl	0.60	0	0	10.49	0.10
CH_2I	1.13	0	0	18.60	0.09
$CONH_2$	−2.18	1	1	9.81	0.24
$CH{=}NOH$	−1.02	1	1	10.28	0.25
CH_3	0.77	0	0	5.65	−0.04
$NHCONH_2$	−2.90	1	1	13.72	0.04
OCH_3	−1.54	1	0	7.33	0.26
CH_2OH	−1.10	1	1	7.19	0.00
$SOCH_3$	−2.24	1	0	13.70	0.52
OSO_2CH_3	−1.34	1	0	16.99	0.39
SCH_3	−0.02	0	0	13.33	0.20
$NHCH_3$	−1.38	1	1	9.11	−0.11
CF_2CF_3	1.34	0	0	9.23	0.44
$C{\equiv}CH$	0.01	0	1	8.25	0.19
CH_2CN	−0.73	1	0	10.11	0.21
$CH{=}CHNO_2$-trans	−0.63	1	0	16.42	0.33
$CH{=}CH_2$	0.88	0	0	9.79	0.07
$COCH_3$	−1.13	1	0	10.29	0.32
$OCOCH_3$	−0.72	1	0	11.85	0.41
CO_2CH_3	−0.72	1	0	11.85	0.33
$NHCOCH_3$	−1.94	1	1	13.71	0.28
$C{=}O(NHCH_3)$	−1.94	1	1	13.39	0.34

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.