

# Applications of Systematic Molecular Scaffold Enumeration to Enrich SAR Information

*N. Yi Mok\* and Nathan Brown*

Cancer Research UK Cancer Therapeutics Unit, Division of Cancer Therapeutics, The Institute of Cancer Research, London, SM2 5NG, UK.

\* To whom correspondence should be addressed.

E-mail: [Yi.Mok@icr.ac.uk](mailto:Yi.Mok@icr.ac.uk)

**Table S1.** The list of 43 literature medicinal chemistry compound series containing a minimum of 100 compounds and their top Level 1 scaffolds used to exemplify the *EnCore* molecular scaffold enumeration approach. These publications were retrieved from the *Journal of Medicinal Chemistry* using the ChEMBL database.

Year	Volume	First Page	PubMed ID	Protein Target	Number of Compounds in Publication	Top Level 1 Scaffold	Number of Unique Scaffolds after Two <i>EnCore</i> Generations
1980	23	830	7401110	Thrombin	109	c1ccc2ccccc2c1	13
1986	29	2298	3783591	Androgen receptor	130	O=C1CCC2CCCCC2N1	164
1988	31	230	3121857	Aldose reductase	127	O=C1NC(=O)C2(CCOC=C2)N1	130
1988	31	2235	2848124	Cholecystokinin receptor	135	O=C1CN=Cc2ccccc2N1	76
1990	33	992	2308149	Arachidonate 5-lipoxygenase	110	c1ccc2ccccc2c1	13
1990	33	2240	2374150	Adenosine A1 receptor	103	C1Cn2cnnc2C=N1	60
1991	34	2231	1712395	HIV-1 reverse transcriptase	125	O=C1NC=CNc2ncccc12	76
1991	34	2248	2066997	Voltage-gated calcium channel	100	C=C1C=NC=CC1c2ccccc2	58
1992	35	1650	1578493	Platelet activating factor receptor	103	C(c1ccccc1)[n+] <sub>2</sub> ccsc2	1
1992	35	3919	1331449	Oxytocin receptor	100	C1CC2(CCN1)C=CC=C2	47
1992	35	4344	1360026	Sigma receptor	121	C(Oc1ccccc1)C2CCNCC2	114
1993	36	4276	8277510	Cholecystokinin B receptor	121	O=C1CN=Cc2ccccc2N1	76
1993	36	3993	8258821	Oxytocin receptor	159	C1CC2(CCN1)CCC=C2	84
1993	36	2279	8360872	Thymidylate synthase	102	C1Cc2cnnc2C=C1	54
1994	37	2808	8064808	Angiotensin II receptor	101	O=C1NN=CN1Cc2ccccc2	79
1995	38	2748	7629813	Potassium-transporting ATPase	125	c1ccc2ncccc2c1	23
1996	39	1243	8632431	Acyl coenzyme A: cholesterol acyltransferase	167	O=C(NS(=O)(=O)Oc1ccccc1)Oc2ccccc2	141
1997	40	1347	9135032	Cyclooxygenase-2	109	c1ccc(cc1)c2ccn[nH]2	71
1997	40	1955	9207936	Arachidonate 5-lipoxygenase	127	c1ccc2sccc2c1	22
1997	40	2017	9207943	Sodium/hydrogen exchanger	113	C1CCN(CC1)c2ccccc2	66
1998	41	877	9526562	Protein farnesyltransferase	157	C1Cc2cccnc2CC=C1	76
2000	43	4288	11063624	Neuropeptide Y receptor	100	c1ncc2[nH]ccc2n1	97
2000	43	4135	11063610	Phosphoglycerate kinase	200	c1ncc2nc[nH]c2n1	100
2000	43	1293	10753467	Protein tyrosine phosphatase 1B	116	o1ccc2ccccc12	37
2001	44	1777	11356112	Protein tyrosine phosphatase CD45	100	O=C1C=Cc2ccccc2C1=O	36
2002	45	3794	12166951	C-C chemokine receptor 3	109	C(C1CCNCC1)c2ccccc2	88
2002	45	5260	12431053	Monoamine oxidase A	118	C1Cc2ccccc2C1	30
2004	47	2776	15139756	Endothelin receptor	104	O=C1CNCc2ccccc2N1	101
2004	47	5021	15456246	Estrogen receptor	108	o1cnc2ccccc12	46
2004	47	5894	15537345	Cyclin-dependent kinase	100	O=C1NC=Nc2[nH]ncc12	90
2005	48	7781	16302817	Dynamin-1	133	O=C(NCCCNC(=O)\C=C\c1ccccc1)\C=C\c2ccccc2	157

Year	Volume	First Page	PubMed ID	Protein Target	Number of Compounds in Publication	Top Level 1 Scaffold	Number of Unique Scaffolds after Two <i>EnCore</i> Generations
2005	48	7829	16302822	DNA-dependent protein kinase	138	O=C1C=COc2ccccc12	54
2006	49	5093	16913699	Dopamine D4 receptor	106	C1CN(CCN1)c2ccccc2	83
2006	49	4896	16884302	Checkpoint kinase Chk1	103	C1Cc2ccccc2N1	46
2006	49	3832	16789740	Methionine aminopeptidase 2	109	C1CCc2ccccc2C1	41
2006	49	971	16451063	RNA-dependent RNA polymerase	113	O=S1(=O)N=CNc2ccccc12	57
2009	52	3892	19469546	Thyroid hormone receptor	147	O=C(CCN1CCNCC1)c2ccccc2	144
2011	54	6305	21823597	Serotonin 2a (5-HT2a) receptor	112	O=C(NCCCN1CCNCC1)c2c[nH]cn2	356
2012	55	1021	22224594	Transglutaminase 2	106	O=S(=O)(N1CCNCC1)c2ccccc2	91
2012	55	2311	22263837	Aldo-keto reductase	111	N(c1ccccc1)c2ccccc2	24
2012	55	2945	22364528	Hypoxia-inducible factor prolyl hydroxylase 1	100	O=C1NC(=O)C2(CCNCC2)N1	109
2013	56	437	23245311	Valosine containing protein	112	c1cncc(c1)n2cnnc2	40
2014	57	3939	24742150	CDK9/cyclin T1	101	N(c1ccccc1)c2ccnnc2	54

**Table S2.** The list of 93 *Journal of Medicinal Chemistry* publications retrieved from the ChEMBL database that were analyzed in the literature COX-2 inhibitor case study.

Year	Volume	First Page	PubMed ID
1994	37	3878	7966148
1995	38	3895	7562922
1995	38	4570	7473585
1995	38	4897	8523403
1996	39	253	8568815
1996	39	1846	8627608
1997	40	819	9057869
1997	40	980	9083488
1997	40	1347	9135032
1997	40	1619	9171872
1997	40	1634	9171873
1998	41	420	9484493
1998	41	1112	9544211
1998	41	1124	9544212
1998	41	3515	9719605
1998	41	4800	9822550
1999	42	1151	10197959
1999	42	1161	10197960
1999	42	1274	10197970
2000	43	214	10649977
2000	43	775	10715145
2000	43	1661	10794682
2000	43	2280	10841807
2000	43	2860	10956194
2000	43	3168	10956225
2000	43	4582	11101350
2001	44	350	11462976
2001	44	641	11262075
2001	44	2921	11520200
2001	44	3039	11520213
2001	44	3223	11563921
2002	45	1402	11906281
2002	45	1511	11906292

Year	Volume	First Page	PubMed ID
2002	45	4816	12383007
2002	45	4847	12383010
2002	45	5182	12408728
2003	46	3463	12877584
2003	46	3975	12954051
2003	46	4872	14584938
2003	46	5484	14640557
2004	47	792	14761182
2004	47	2180	15084117
2004	47	3874	15239665
2004	47	3972	15267236
2004	47	4881	15369392
2004	47	6108	15537365
2004	47	6195	15566290
2004	47	6749	15615524
2005	48	3428	15857149
2005	48	3613	15887968
2005	48	3930	15916445
2005	48	4061	15943479
2005	48	5705	16134939
2005	48	6516	16190777
2005	48	6523	16220969
2005	48	6997	16250658
2006	49	1668	16509583
2007	50	1425	17341061
2007	50	1449	17335184
2007	50	5403	17915854
2007	50	6367	17994684
2008	51	142	18072726
2008	51	1601	18311898
2008	51	2400	18363350
2008	51	4150	18588282
2008	51	4476	18598017

Year	Volume	First Page	PubMed ID
2008	51	7800	19053765
2009	52	1525	19296694
2009	52	5864	19791801
2010	53	723	19957931
2010	53	3707	20387815
2010	53	4691	20503989
2010	53	5033	20527891
2010	53	6560	20804197
2010	53	7879	20954731
2011	54	1356	21280601
2011	54	2060	21381754
2011	54	2529	21413808
2011	54	3037	21434686
2011	54	3650	21542630
2011	54	7759	21992176
2011	54	8555	22091869
2012	55	688	22148253
2012	55	2287	22263894
2012	55	2311	22263837
2012	55	2932	22236250
2012	55	5143	22404396
2012	55	8152	22916727
2012	55	8807	23043222
2013	56	2429	23432095
2013	56	3191	23534442
2013	56	4277	23651359
2014	57	5638	24920381