Printed in the Republic of Korea
ANALYTICAL SCIENCE & TECHNOLOGY
Vol. 24, No. 6, 533-543, 2011
http://dx.doi.org/10.5806/AST.2011.24.6.533

hERG 이온채널 저해제에 대한 2D-QSAR 분석

전을혜 · 박지현 · 정진희 · 이성광★
한남대학교 화학과
(2011. 12. 6. 접수, 2011. 11. 30. 수정, 2011. 12. 8. 승인)

2D-QSAR analysis for hERG ion channel inhibitors

Eul Hye Jeon, Ji Hyeon Park, Jin Hee Jeong and Sung Kwang Lee★
Department of Chemistry, Hannam University, Daejeon 305-811, Korea
(Received December 6, 2011; Revised November 30, 2011; Accepted December 8, 2011)

요 약: hERG (human ether-a-go-go related gene) 이온채널은 심장 재분극의 중요 요소이며 이 채널의 저해제는 부정맥과 돌연사를 유발할 수 있다. 따라서, 신약개발과정에서 후보물질이 hERG 이온채널의 잠재적인 저해제일 경우에는 심장독성 부작용을 유발하므로, 이를 최소화하고자 많은 노력이 집중되고 있다. 본 연구는 HEK(인간 배아 신장)세포에서 얻은 202개 유기화합물의 IC_{50} 데이터를 이용하여 2차원 구조-활성의 정량적 관계(2D-QSAR)방법으로 예측하는 모델을 개발하였다. hERG이온채널 저해제의 기계학습방법으로는 다중선형회귀(Multiple Linear Regression), 서포트 벡터 머신(Support Vector Machine: SVM)방법과 인공신경망(Artificial Neural Network)방법이며, 교차검증을 적용한 모집단 기반 전진선택(forward selection)방법과 결합하여 각 학습모델에 적합한 최적의 표현자들을 결정하였다. 가장 우수한 방법은 14종의 표현자를 사용한 인공신경망방법(R^2_{CV}=0.617, RMSECV=0.762, MAECV=0.583)이었고, 다중선형회귀방법을 통해서 hERG이온채널 저해물질의 구조적 특징과 수용체와의 상호작용을 설명할 수 있다. QSAR모델의 검증은 교차검증과 Y-scrambling test방법으로 수행하였다.

Abstract: The hERG (human ether-a-go-go related gene) ion channel is a main factor for cardiac repolarization, and the blockade of this channel could induce arrhythmia and sudden death. Therefore, potential hERG ion channel inhibitors are now a primary concern in the drug discovery process, and lots of efforts are focused on the minimizing the cardiotoxic side effect. In this study, IC_{50} data of 202 organic compounds in HEK (human embryonic kidney) cell from literatures were used to develop predictive 2D-QSAR model. Multiple linear regression (MLR), Support Vector Machine (SVM), and artificial neural network (ANN) were utilized to predict inhibition concentration of hERG ion channel as machine learning methods. Population based-forward selection method with cross-validation procedure was combined with each learning method and used to select best subset descriptors for each learning algorithm. The best model was ANN model based on 14 descriptors (R^2_{CV}=0.617, RMSECV=0.762, MAECV=0.583) and the MLR model could describe the structural characteristics of inhibitors and interaction with hERG receptors. The validation of QSAR models was evaluated through the 5-fold cross-validation and Y-scrambling test.

Key words: 2D-QSAR, hERG ion channel inhibitor, machine learning, MLR, SVM, ANN, cross-validation

★ Corresponding author
Phone : +82-(0)42-629-8874 Fax : +82-(0)42-629-8811
E-mail : leesk@hnu.kr

1. 서 론

신약 탐색 과정에서 개발 비용과 시간을 절감하면서 신약 후보 물질을 개발하기 위해서는 개발 초기에 흡수, 분포, 대사, 배설, 독성(Absorption, Distribution, Metabolism, Excretion, Toxicity:ADME/Tox)과 같이 약물이 기본적으로 지녀야 하는 특성을 고려하여 후보물질을 설계하는 것이 매우 중요하다. 특히 화합물 라이브러리에 대하여 ADME/Tox 특성 결과를 얻을 수 있는 고속 대량 검색법(high throughput screening)이 현재도 개발 되고 있으며, 이와 더불어 컴퓨터를 이용한 ADME/Tox물성을 예측하는 방법도 함께 개발되면서 신약 개발 초기 단계에 부적합한 후보 물질을 선별하여 실패 요인을 줄이려는 노력들이 많이 진행되고 있다.

hERG (human ether-a-go-go related gene) 이온채널은 심장 박동수에 영향을 미치는 것으로 알려 져 있다. 특히 hERG 이온채널의 저해(inhibition)은 심전도상의 QT 연장 증후군(long QT syndrome)과 관련이 있으며, 심혈 관계의 이상으로 돌연사를 일으킬 수 있다.[1] 최근에 일부 의약품(terfenadine, cisapride, astemizole, grepafloxacin)들이 이러한 원인으로 말미암아 의약품 시장에서 퇴출되기도 했다.[2] 따라서 QT 연장 증후군을 일으키는 hERG이온채널의 저해와 관련하여 후보 물질들을 테스트할 수 있는 방법을 개발하므로써, 심장 부정맥의 부작용을 예방하는 것이 중요하다. 그러나 생체 내에서hERG이온채널과 관련된 QT 연장 증후군 임상 실험은 고비용과 많은 시간을 요하므로 신약 개발의 지연을 초래하고 있다. 이 부분을 극복하기 위하여 고속 대량 검색법의 개발과 더불어 hERG이온채널 저해 관련 예측 모델을 실험 대체 도구로 사용하고 있다.

기존에 진행된 연구들을 살펴보면, Gunturi[3]등은 2차원 및 3차원 표현자를 이용하여 k-Nearest Neighbor (k-NN)과 Local Lazy Regression(LLR)방법을 학습방법으로 유전자 알고리즘(Genetic Algorithm, GA)과 결합하여 최적의 표현자들을 선택하였다. 연구 결과에서는 3차원 표현자를 이용하여 LLR방법을 적용한 방법이 가장 좋은 결과(Q^2_{LOO}=0.818)를 얻었으나, 모델 구현을 위해서는 항상 3차원 화학 구조를 제공해야 하는 단점을 지니고 있다. Yoshida[4] 등은 104종의 유기화합물에 대하여 2차원 표현자를 이용하여 다중 선형 회귀(MLR) 방법을 적용하였고, 옥탄올/물 분배 계수(octanol/water partition coefficient), 위상학적 극성 표면적(topological polar surface area), 직경(diameter), 부분 전하를 가진 원자의 표면적 합계(summed surface area of atoms with partial charges) 표현자 만을 사용하여 물리 화학적인 해석을 중점으로 이용해 MLR 예측 모델(Q^2_{LOO}=0.671)을 구현하였다. 두 연구에서 모두 모델 검증을 위해서 Leave-one-out method (LOO, 훈련 데이터 중에서 한종의 데이터를 제거하여 모델을 구현한 뒤에 제거한 데이터로 모델의 예측 정도를 평가하는 방법)를 사용하였으나, 최근에 발표한 연구[5]에서는 모델 검증에 있어서 LOO방법은 모델의 예측 능력을 평가하는데 부적합하다는 의견을 제시하고 있다.

본 연구에서는 이러한 부분을 보강하기 위하여 선형 및 비선형 QSAR모델을 구현한 뒤에 묶음(fold)으로 구분된 데이터들을 이용하여 검증하는 leave-many-out (LMO)방법을 적용하여 모델을 최적하고, 모델의 신뢰성을 검증하도록 하였다. 특히 모집단 기반 전진 선택(forward selection)방법을 학습방법과 함께 연동하여 LMO방법의 검증 결과를 통해 최적의 표현자를 설정하도록 하였다. 이를 통해 최종적으로 hERG 이온채널의 저해 농도를 검증된 모델로 예측하므로서 심혈관계 부작용을 신약 설계 단계에서 고려할 수 있도록 하고자 한다.

2. 방 법

2.1. 사용 데이터

이 연구에 사용된 데이터는 의약품 주성분과 약유사 성분으로 구성된 202종의 유기 화합물에 대하여 실험적으로 측정한 hERG 이온채널의 저해 농도(IC_{50})를 이용하였고, 다양한 문헌으로부터 실험 데이터를 수집하였다.[4,6,7] 실험값 중에서 인체 배아 신장(Human Embryonic Kidney, HEK)에서 얻은 hERG 이온채널의 IC_{50}(μM)값만 사용하였으며, -log로 나타낸 pIC_{50}으로 전환하여 모델의 종속변수(Y)로 사용하였다. 수집된 화합물 데이터 중에서 표현자 계산을 하기 곤란한 비단일 화합물(주로 염이나 혼합물)이나 무기원소를 포함하는 화합물은 데이터에서 제외하였다. 사용된 화합물의 pIC_{50} 값과 다양한 예측모델의 예측값을 *Table* 1에 나타내었다.

2.2. 분자 표현자(molecular descriptor) 계산

hERG 저해제 예측 모델을 개발하기 위하여 사용된 이론적인 분자 표현자는 PreADMET v2.0프로그램[8]으로 계산하였다. 사용되는 표현자들을 특성에 따라 구

Table 1. Results for QSAR models on 202 drugs and drug-like compounds for hERG pIC_{50}

No	Compounds	Exp. pIC_{50} (μM)[a]	Pre. pIC_{50}(μM)[b]			No	Compounds	Exp. pIC_{50} (μM)	Pre. pIC_{50}(μM)		
			MLR	SVM	ANN				MLR	SVM	ANN
1	2-hydroxymethylolanzapine	-0.477	-0.785	-0.889	-0.904	35	Clozapine N-oxide	2.510	1.854	1.878	2.053
2	Acetylmethadone	-0.505	-0.993	-0.316	-1.052	36	Clozapine	-0.652	-0.955	-0.760	-0.894
3	Almokalant	-0.558	-0.181	-0.138	-0.243	37	Cocaethylene	-1.152	-0.958	-0.743	-0.893
4	Alosetron	0.679	0.322	-0.062	0.688	38	Cocaethyline	-1.435	-1.044	-0.284	-0.658
5	Ambasilide	-2.041	-2.236	-1.760	-2.498	39	Cocaine	-1.117	-0.360	-0.704	-0.293
6	Amitriptyline	2.408	1.960	1.908	2.070	40	Cyamemazine	1.854	1.616	-0.392	2.003
7	Amsacrine	-0.010	-0.107	-0.815	0.233	41	D-703	0.233	-0.575	-0.228	0.100
8	Artemisin	-0.093	0.156	0.140	0.247	42	Desbutylhalofantrine	1.398	0.552	0.813	0.786
9	Astemizole	-1.322	-0.999	-0.839	-1.010	43	Desipramine	1.454	0.438	0.108	0.947
10	Azimilide	1.702	0.411	-0.117	0.445	44	Desloratidine	-0.820	-0.097	-0.407	-1.090
11	Benzoylecgonine	-3.601	-1.138	-1.629	-2.487	45	Desmethylastemizole	-0.820	-0.990	-0.407	-0.745
12	Bepridil	1.161	1.388	0.716	1.365	46	Desmethylclozapine	-0.120	-0.967	-0.378	-0.758
13	Berberine	0.000	0.180	0.413	-0.136	47	Desmethylolanzapine	0.050	0.029	-0.533	-0.093
14	Bisindolylmaleimide I	-1.404	-1.356	-0.946	-1.449	48	Dextro-propoxyphene	-0.050	-0.982	-0.310	-0.714
15	BMCL20051225_3hS	-0.875	-0.698	-0.410	-1.145	49	Diltiazem	1.390	0.858	1.696	0.417
16	BMCL20065859_08i	-0.934	-0.422	-1.048	-0.558	50	Diphenhydramine	-2.290	-2.317	-1.877	-1.997
17	BRL_32872	-0.398	0.458	-0.272	-0.127	51	Disopyramide	-2.438	-2.100	-2.605	-2.583
18	Bupivacaine	-0.936	-0.465	-0.643	-0.408	52	Dofetilide	-0.255	0.105	0.361	0.459
19	Buprenorphine	-0.820	-0.986	-1.179	-1.386	53	Dolasetron	-0.592	-0.067	-1.004	-0.142
20	Caffeine	1.634	0.400	-0.024	0.403	54	Dolasetron	0.337	0.105	-0.280	0.394
21	Carvedilol	-0.599	-0.473	0.442	0.106	55	Domperidone	-1.870	-1.854	-2.282	-2.322
22	Cetirizine	-1.517	-2.311	-1.713	-1.511	56	Doxazosin	-0.571	0.120	0.376	-0.213
23	Chloroquine	-0.462	-0.314	-0.654	-0.018	57	Droperidol	1.588	1.508	1.318	1.088
24	Chlorpheniramine	1.900	1.018	1.487	1.067	58	E-4031	1.381	-0.101	0.548	0.814
25	Chlorpromazine	0.745	1.056	0.648	1.213	59	Ebastine	-0.531	-0.433	-0.540	-0.864
26	Chromanol 293B	-0.479	-0.811	-0.777	-0.626	60	Ecgognine-methyl-ester	-1.301	-1.394	-0.888	-1.449
27	Ciprofloxacin	0.543	-0.462	-0.155	-0.279	61	EGIS-7229	1.553	2.042	1.447	1.350
28	Cisapride	-2.125	-0.473	-0.272	-0.354	62	EMD-60263	-1.204	-1.004	-0.792	-0.484
29	Citalopram	-0.079	-1.262	-0.873	-0.913	63	EMD-60417	-1.220	-0.116	-0.795	-0.409
30	Clarithromycin	-0.785	-1.413	-0.886	-1.001	64	EMD-66398	-0.146	-0.871	-0.810	-1.002
31	Clobutinol	0.060	-0.680	-1.188	-1.316	65	EMD-66430	-1.061	-0.792	-0.822	-0.915
32	Clofilium	0.740	0.287	-0.143	0.043	66	Epinastine	-1.804	-0.883	-0.830	-0.938
33	Clomiphene	1.144	1.027	0.732	0.735	67	ER-118585	-0.556	-0.524	-0.858	-0.662
34	Clotrimazol	-0.143	-0.640	-0.555	-0.968	68	Erythromycin	-0.869	-0.783	-0.745	-0.927

Table 1. Continued

No	Compounds	Exp. pIC_{50} (μM)[a]	Pre. pIC_{50}(μM)[b]			No	Compounds	Exp. pIC_{50} (μM)	Pre. pIC_{50}(μM)		
			MLR	SVM	ANN				MLR	SVM	ANN
69	Erythromycylamine	-0.398	-0.840	-0.810	-0.974	102	JMC20066569_34	-0.820	0.453	-0.088	0.368
70	Fentanyl	-0.556	-0.859	-0.812	-0.974	103	JMC20066569_35	-2.230	-2.456	-2.137	-2.575
71	Flecainide	-1.179	-0.880	-0.838	-0.974	104	JMC20066569_36	0.454	0.287	0.676	0.325
72	Fluoxetine	-0.699	-0.804	-0.373	-0.843	105	JMC20066569_40	2.124	1.427	1.316	1.314
73	Gatifloxacin	-0.602	-0.880	-0.835	-0.974	106	JMC20066569_42	-2.047	-1.624	-1.922	-2.075
74	Glibenclamide	-0.826	-0.929	-0.814	-1.016	107	JMC20066569_43	-0.890	-0.182	-0.247	0.218
75	Glimepiride	-1.021	-0.885	-0.832	-1.002	108	JMC20066569_44	-1.562	-1.667	-1.440	-2.085
76	Glyceryl-nonivamide	-1.201	-0.823	-0.843	-0.947	109	JMC20066569_46	-2.168	-2.366	-1.829	-2.040
77	Granisetron	-0.792	-0.905	-0.860	-1.002	110	JMC20066569_47	-1.643	-1.371	-1.130	-1.540
78	Grepafloxacin	-1.299	-1.412	-1.117	-1.125	111	JMC20066569_49	0.721	0.954	0.309	0.686
79	Halofantrine	-0.748	-0.769	-0.944	-0.884	112	JMC20066569_51	-1.061	-1.050	-0.648	-1.204
80	Haloperidol	-1.130	-0.885	-0.827	-1.002	113	JMC20066569_52	-1.403	-1.840	-1.610	-1.651
81	Hydrodolasetron	-1.086	-0.871	-0.764	-1.002	114	JMC20066569_53	-2.789	-2.044	-2.009	-2.038
82	Imipramine	-1.017	-1.252	-0.854	-1.028	115	JMC20066569_54	-2.060	-1.857	-2.005	-1.903
83	IQB-9302	-0.732	-0.869	-0.849	-1.028	116	Josamycin	-1.079	-0.907	-0.892	-0.955
84	Isobutylmethylxanthine	-1.326	-1.361	-1.213	-1.085	117	Ketanserin	1.560	1.029	0.853	0.815
85	JMC_20042405_1	-1.167	-0.899	-0.942	-1.062	118	Ketoconazole	-0.362	-0.137	-0.267	0.268
86	JMC_20042405_30	-0.929	-1.283	-0.914	-1.062	119	LAAM	-0.499	0.380	-0.087	0.275
87	JMC_20055888-7	-1.307	-1.406	-0.881	-1.068	120	Levofloxacin	0.019	-0.692	-0.536	-0.298
88	JMC20051725_8f	-1.679	-0.575	-1.570	-1.725	121	Lidoflazine	0.689	-0.481	-0.131	-0.273
89	JMC20063614_32	-2.010	-2.385	-2.422	-2.430	122	Lignocaine	-2.531	-2.276	-2.272	-2.217
90	JMC20063766_14	-0.340	-0.376	0.072	0.392	123	Lopinavir	0.091	-0.535	0.099	-0.703
91	JMC20066569_01	-0.196	0.505	0.156	0.444	124	Loratadine	-1.399	-0.415	-1.004	-0.819
92	JMC20066569_02	-0.342	-0.785	-0.889	-0.892	125	Losartan	-0.114	0.178	0.534	0.396
93	JMC20066569_17	1.796	1.707	0.446	1.371	126	Lumefantrine	-2.117	-2.173	-2.529	-2.426
94	JMC20066569_18	-0.934	-0.638	-0.593	-0.663	127	LY-97241	-3.477	-2.361	-2.316	-2.072
95	JMC20066569_19	0.027	-1.146	-0.605	-1.308	128	Maprotiline	-2.380	-2.475	-2.046	-1.969
96	JMC20066569_20	-0.890	-0.169	-0.301	-0.835	129	MCI-154	-1.310	-0.673	-0.904	-1.100
97	JMC20066569_21	-0.908	0.382	0.535	0.252	130	MDL-74156	1.985	1.140	1.828	1.259
98	JMC20066569_22	-0.716	-1.211	-0.462	-1.109	131	Mefloquine	-0.196	-1.088	-0.298	-0.474
99	JMC20066569_23	-0.574	-1.371	-0.474	-0.612	132	Mesoridazine	-2.143	-1.360	-1.731	-1.814
100	JMC20066569_30	0.428	-0.141	0.424	-0.606	133	Methadone	-0.440	-0.667	-0.867	-0.714
101	JMC20066569_31	-0.991	-0.750	-0.808	-0.836	134	Methoxyverapamil	-0.756	-0.247	-0.344	-0.668

Table 1. Continued

No	Compounds	Exp. pIC_{50} (μM)[a]	Pre. pIC_{50}(μM)[b]			No	Compounds	Exp. pIC_{50} (μM)	Pre. pIC_{50}(μM)		
			MLR	SVM	ANN				MLR	SVM	ANN
135	Methylecgonidine	-0.216	-0.231	-0.735	-0.321	169	Prazosin	-0.280	-0.868	-0.671	-0.864
136	Mibefradil	-0.761	0.805	0.404	0.237	170	Procainamide	-1.820	-0.661	-0.304	-0.607
137	Mizolastine	0.206	-0.423	-0.210	-0.588	171	Propafenone	-0.690	-2.274	-2.048	-2.281
138	MK-499	0.539	0.191	0.795	0.370	172	Propoxyphene	0.060	0.542	0.424	0.675
139	Morin	-1.562	-1.456	-2.191	-1.283	173	Protopine	-0.900	-0.280	-0.839	-0.730
140	Mosapride	-1.185	-0.713	-0.837	-1.058	174	Prucalopride	-2.980	-1.958	-2.364	-2.056
141	Moxafloxacin	1.635	0.712	1.272	0.663	175	Pyrilamine	-0.080	-1.262	-0.873	-0.889
142	Naringenin	-1.506	-0.948	0.030	-0.371	176	Quetiapine	-0.650	-1.046	-0.238	-0.944
143	N-Demethylerythromycin	-2.150	-2.003	-1.812	-1.831	177	Quinidine	-0.440	-0.667	-0.867	-0.714
144	N-desbutylhalofantrine	-4.000	-1.706	-1.945	-1.083	178	Risperidone	-1.210	-0.973	-0.739	-0.961
145	N-desmethylclozapine	-2.000	-1.389	-0.625	-1.313	179	Roxithromycin	-0.770	-0.610	-0.357	-0.605
146	Nelfinavir	-0.100	0.955	0.402	1.174	180	RP-58866	0.790	0.442	1.806	0.707
147	Nicotine	-1.248	-0.853	-0.414	-0.464	181	Saquinavir	1.600	1.554	1.346	1.702
148	Nifedipine	1.758	1.283	0.636	0.719	182	Sertindole	-3.780	-2.354	-2.250	-2.551
149	NIP-142	-0.470	-0.034	-1.205	-0.118	183	Sildenafil	-2.220	-1.536	-1.107	-1.175
150	Nitrendipine	0.716	0.329	0.759	0.191	184	Sotalol	-1.750	-1.830	-2.159	-2.002
151	Noracetylmethadol	-0.462	0.034	-0.291	-0.043	185	Sparfloxacin	-1.870	-1.886	-1.457	-1.042
152	Norastemizole	-1.952	-1.076	-1.132	-1.120	186	Sulfamethoxazole	1.000	-1.116	-0.622	-1.044
153	Norfluoxetine	-2.380	-1.244	-1.521	-1.341	187	Tadalafil	-1.820	-1.895	-2.411	-2.052
154	Norpropoxyphene	-1.107	-0.885	-0.418	-0.642	188	Tamoxifen	-1.000	-1.993	-1.903	-2.083
155	Norverapamil	0.658	0.452	-0.063	0.323	189	Terazosin	-3.055	-1.971	-1.677	-1.660
156	NPS-2143	-0.079	-1.442	-0.500	-1.343	190	Terfenadine	2.800	1.281	0.906	1.127
157	Olanzapine	0.903	0.739	1.207	0.882	191	Terikalant	-0.019	-1.622	-0.496	-0.751
158	Oleandomycin	-0.041	-0.152	-0.061	-0.090	192	Terodiline	-1.080	-1.311	-0.655	-1.148
159	Ondansetron	-0.763	-0.154	-1.004	-0.120	193	Thioridazine	-0.160	0.727	0.606	1.449
160	OPC-18790	-0.352	-0.062	0.005	0.582	194	Trazodone	-0.320	-1.722	-1.423	-1.440
161	Oxybutynin	-1.616	-0.977	-0.540	-0.993	195	Trimebutine	-0.560	-0.870	-0.874	-0.839
162	Paliperidone	-0.775	-0.147	0.149	-0.217	196	Trimethoprim	-0.890	-0.779	-0.478	-1.505
163	Pentobarbital	-2.420	-1.933	-1.854	-2.073	197	Vardenafil	-2.220	-1.133	-0.603	-1.257
164	Perhexiline	-1.083	-1.311	-0.655	-1.148	198	Verapamil	0.680	0.644	1.083	0.789
165	Phenobarbital	1.144	1.027	0.732	0.735	199	Vesnarinone	0.120	-1.863	-2.334	-1.988
166	Phenytoin	-0.652	-0.955	-0.753	-0.894	200	Vitamin K	1.160	0.635	0.788	0.803
167	Pilsicainide	0.357	0.255	-0.083	-0.111	201	Way123398	-0.760	0.799	-0.348	0.293
168	Pimozide	0.350	-0.068	0.104	0.414	202	Ziprasidone	0.690	0.788	-0.280	0.668

[a] Experimental pIC50 (μM) obtained from reference [3,5,6].

[b] Calculated Exp pIC50 (μM) obtained from MLR, SVM and ANN model.

분하면, 구조적(constitutional), 물리화학적(physico-chemical), 위상학적(topological) 특성을 수치로 나타낸 표현자와 MPEOE 전하를 이용한 정전기적(electrostatic) 표현자[9,10] 그리고 화학적 상호 작용점(chemical feature) 들간의 결합간 거리 정보를 포함한 CATS(chemical advanced template search)표현자[11]가 사용되었다.

2.3. 데이터 전처리(data-processing)

본 연구에서는 화학 구조적 특징을 수치로 표현한 표현자들의 수가 매우 많으므로, 이 중에서 모델에 부적합한 표현자를 미리 제거하는 과정이 필요하다. 다음과 같은 기준에 해당하는 표현자들은 다음 단계의 모델 구현에서 사용하지 않았다. 1) 전체 화합물에 대하여 분자 표현자가 모두 계산이 되지 않는 표현자, 2) 모든 분자에 대한 분자 표현자의 표준편차가 0.01보다 작은 표현자, 3) pIC_{50} 값과의 상관계수(R^2)값이 0.01보다 작은 표현자, 4) 두 분자 표현자간의 상관계수(R^2)값이 0.9이상인 경우, 다른 표현자와의 상관관계가 큰 표현자는 이 과정에서 제거하였다. PreADMET 프로그램에서 계산한 559종의 2차원 표현자들 중에서 최종적으로 200종의 분자 표현자만 다음 단계의 모델 구현에 사용하게 되었다. *Fig.* 1는 QSARs 모델링의 모든 과정을 보여 준다.

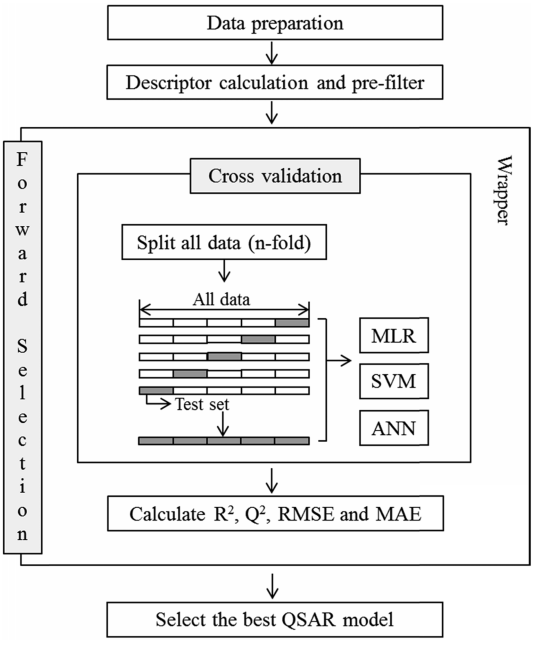

Fig. 1. Process of QSAR modeling for prediction of hERG inhibitor.

2.4. 표현자 선택(descriptor selection)

QSAR모델을 구현하고자 할 때, 적용이 가능한 많은 표현자들 중에서 모델 구현에 적합한 최적의 표현자 집합을 선택하는 과정이 매우 중요하다. 본 연구에서는 모집단 형태로 적용하는 전진 선택(forward selection)방법[12]을 사용하여 각 학습방법에 적합한 표현자 집합을 선정하였다. 이 방법을 간단히 소개하면, 처음에 n개의 표현자를 각각 1개씩 포함하는 n개의 개별 수식 형태의 모집단(population)을 구성하고, 모델의 적합도(goodness of fit)을 가장 잘 향상시킬 수 있는 표현자를 각 개별 수식에 추가하도록 한다. 이 때 추가되는 표현자는 기존의 개별 수식에 포함되지 않은 것이어야 한다. 이 과정을 통해 개별 수식에 추가되는 k개의 가장 뛰어난 수식을 유지시키도록 함으로서, 최종적으로 k×n개의 개별 수식이 모집단에 남아있게 된다. 본 과정에서 k를 5로 설정하여, 모집단 안에 5×202=1010개의 개별 수식이 존재하도록 하였다. 표현자를 추가할 때 기준으로 사용한 적합도는 제곱근 오차(root mean square error, RMSE)를 사용하였으며, 추가적으로 결정계수(R^2), 평균절대오차(mean absolute error, MAE)를 계산하였다. 이 적합도의 계산식은 다음과 같다.

$$R^2 = 1 - \frac{\Sigma(y_{obs} - y_{pre})^2}{\Sigma(y_{obs} - y_{mean})^2} \quad RMSE = \sqrt{\frac{\Sigma(y_{obs} - y_{pre})^2}{N}}$$

$$MAE = \frac{\Sigma|y_{obs} - y_{pre}|}{N}$$

여기서 y_{obs}는 실험값, y_{pre}는 예측된 값, y_{mean}의 모든 실험값의 평균값, 그리고 N은 모델에 사용된 화합물의 수를 나타낸다.

앞에서 언급한 모델의 적합도들은 모델에 포함되는 표현자의 수가 증가할수록 항상 개선이 되므로, 실제 모델의 예측 능력을 표현할 수 있도록 교차 검증(cross-validation)방법을 적용하였다. 교차 검증 법은 전체 훈련 데이터를 정해진 수의 묶음(fold)으로 나눈 뒤에, 예측에 사용될 한 묶음을 제외한 나머지 묶음들로 모델을 구현하고, 제외시킨 한 묶음으로 모델이 잘 맞는지는 예측하도록 한다. 이 과정을 묶음의 수만큼 반복하여 예측에만 사용한 묶음의 데이터를 모두 모으면 전체 훈련 데이터의 수와 같아 지며, 이때 예측된 값을 이용하여 이들의 모델 적합도를 계산하므로써, 실제 모델의 예측 능력을 평가하는 방법이다. 본 연구에서는 5개의 그룹으로 구성하여 적합도를 계산하였고, 교차 검증으로 얻은 적합도를 훈련 데이터에

서만 얻은 적합도와 구분하여, R^2_{CV}, RMSECV(root mean square error of cross-validation), MAECV(mean absolute error of cross-validation)로 표기하였다. 각 학습 방법에서 최적의 표현자 수는 교차 검증을 통해 얻은 Jp값이 가장 낮은 것으로 결정하였고, 표현자 수가 증가할 때 Jp값이 증가되도록 하여 최적의 표현자 수를 계산할 수 있다. 그 수식은 다음과 같다.

$$J_p = \frac{(n+p)\cdot s^2}{n} = \frac{RSS}{n}\cdot\frac{(n+p)}{(n-p)} = \frac{RMSECV^2\cdot(n+p)}{(n-p)}$$

여기서 n은 전체 화합물의 수, p는 모델에 사용된 표현자 수를 나타낸다.

2.5. 모델 구현에 사용된 학습 방법

다중 선형 회귀(Multiple Linear Regression)모델은 일반적으로 널리 사용되는 매우 유용한 방법으로, 화합물의 분자 표현자와 화합물의 저해 농도 사이에 선형적인 관계를 다음과 같은 선형 관계식으로 나타낸다.

$$Y = a_0 + a_1X_1 + a_2X_2 + \ldots. + a_nX_n$$

여기서 $[X_1, X_2, \ldots, X_n]$은 분자 표현자를, a_0는 선형 모델의 절편, $[a_1, a_2, a_n]$은 표현자들의 회귀계수를 나타낸다. 이 방법은 수식 내에서 각 표현자의 회귀계수의 부호를 확인하므로서 종속변수(Y)와의 관계를 설명할 수 있다.

서포트 벡터 머신(Support Vector Machine, SVM)방법은 최근에 각광을 받는 새로운 기계학습 방법으로, 원래는 Cortes와 Vapnik[13]에 의해서 구조적 위험 최소화(structural risk minimization)원리에 기반하여 분류(classification)를 하기 위하여 개발된 학습 방법이다. 추후에 SVM방법은 ε-insensitive loss함수를 도입함으로써 회귀 예측에도 적용이 가능하게 되었다. 이 방법의 기본적인 개념은 입력 데이터들을 비선형 변환 함수를 이용하여 더 높은 차원의 공간으로 변환하고, 이 어서 그 공간상에서 선형 회귀식을 수행하는 방법이다. SVM방법의 성능은 주로 변환 함수로 사용되는 kernel함수의 형태와 최적화를 수행해야 하는 C, γ값 그리고 kernel변수에 좌우된다. 본 연구에서는 SVM 방법내의 추가적인 매개변수를 제거할 수 있는 γ-SVM방법[14]을 사용하였고, kernel함수로 radial basis function (RBF)을 사용하였으며, RBF함수내 최적화를 수행해야 하는 변수인 ν와 C, γ값은 표현자 선택 과정이 모두 수행된 다음에 grid search방법으로 최적화 하였다.

인공 신경망(Artificial Neural Network, ANN)방법은 사람의 두뇌를 모델로 하여 여러 정보를 처리하는데 있어서 신경세포들간의 연결선을 통해 병렬 분산 형태로 정보를 처리하는 알고리즘이다. ANN방법 중에서 입력 데이터의 의한 학습을 통해 신경세포를 연결하는 가중치를 재조정하여 실험값과의 오차를 줄여나가는 역전파(back-propagation) 신경망법[15]이 주로 회귀모델에 사용되며, 본 연구에서도 이 방법을 사용하였다. 본 연구에서는 입력층, 은닉층, 출력층으로 구성된 신경망 구조를 사용하였고, 각 층에서는 활성 함수(activation function)는 sigmoid함수를 사용하였다. ANN방법의 매개변수인 은닉층의 뉴론 수, learning rate, momentum값은 모델의 성능 평가를 통하여 최적화하였다.

3. 결과 및 고찰

다양한 유기 화합물의 hERG 이온채널 저해 농도를 예측하기 위하여 3가지 다른 학습방법으로 QSAR모델을 구현하였다. 선형적인 관계는 다중 선형 회귀 모델(MLR)으로, 비선형적인 관계는 인공 신경망(ANN)방법과 서포트 벡터 머신(SVM)방법을 이용하여 확인하였다. 많은 QSAR분석들은 훈련 데이터의 적합도에 중점을 두어 화합물 수에 비하여 많은 표현자를 사용하기도 하지만(일반적으로 모델에 사용된 표현자의 수는 화합물 수의 최대 1/5이하로 하며, 일반적으로는 1/10이하로 진행함), 훈련 데이터의 적합도가 뛰어난 모델보다 검증 수준이 뛰어난 모델을 사용하는 것이 모델의 예측도를 높일 수 있는 방법이다. 그러므로 본 연구에서는 훈련 데이터에 대하여 모집단을 기반으로 한 전진 선택(forward selection)방법을 통해 모델 내의 표현자들을 선택하도록 하였고, 5 묶음 교차-검증(5 fold cross-validation)법을 통해서 가장 낮은 RMSECV 값을 지닌 모델을 최적의 표현자 모델로 결정하였다. *Fig.* 2는 각 학습모델에 사용된 표현자의 수에서 대한 교차 검증으로 얻은 Jp값을 그래프로 나타낸 것이다. 이 결과를 통해서 Jp값이 가장 낮은 모델을 각 방법의 최적 모델로 결정하였는데, MLR방법에서는 15종, SVM방법은 13종, ANN방법은 14종의 표현자로 구성된 모델로 결정되었다. 모델 구현을 위해 사용한 방법들은 모두 공개 프로그램인 RapidMiner v5.0[16]을 이용하여 결과를 얻었다.

다중 선형 회귀 분석(MLR)식에서 결정한 최종 모델은 15종의 표현자를 포함하였고, 그 수식은 다음과 같으며 표현자의 설명은 *Table* 3에 나타내었다.

Table 2. List of descriptors for MLR, SVM and ANN models

Model	Descriptor name	Description
MLR model	NoCdO	The number of double bonds between C atoms and O atoms
	No_{imide}	The number of imide groups
	No_{ketone}	The number of ketone groups
	$VSA_{hydunsat}$	2D van der Waals partial surface area of hydrophobic unsaturated groups
	$FraVSA_{hydunsat}$	Fraction of 2D van der Waals hydrophobic unsaturated surface area
	$E_{state}S_{aaaC}$	Sum of E-state for aaaC type(a: aromatic ring)
	$Kier_{shape}3$	Kier shape index order 3
	$CATS^{a}_{HydAro}10$	Indicator for atom pair in 10 distance between hydrophobic atom and aromatic ring
	$CATS_{Hydpos}5$	Indicator for atom pair in 5 distance between hydrophobic atom and positive charge atom
	$CATS_{AroAro}6$	Indicator for atom pair in 6 distance between aromatic ring and aromatic ring
	$CATS_{AroPos}7$	Indicator for atom pair in 7 distance between aromatic ring and positive charge atom
	$CATS_{DonDon}3$	Indicator for atom pair in 3 distance between H-bond donor atom and H-bond donor atom
	$CATS_{DonAcc}9$	Indicator for atom pair in 9 distance between H-bond donor atom and H-bond acceptor atom
	$CATS_{DonPos}6$	Indicator for atom pair in 6 distance between H-bond donor atom and positive charge atom
	$CATS_{DonNeg}8$	Indicator for atom pair in 8 distance between H-bond donor atom and negative charge atom
SVM model	$No_{AroRing}$	The number of aromatic rings
	No_{cyano}	The number of cyano groups
	No_{ester}	The number of ester groups
	No_{ketone}	The number of ketone groups
	$E_{state}S_{ssCH2}$	Sum of E-state for $-CH_2-$ type
	QminC	The minimum partial charge for C atoms in a molecule (MPEOE[b] charge)
	$CATS_{HydHyd}10$	Indicator for atom pair in 10 distance between hydrophobic atom and hydrophobic atom
	$CATS_{HydAcc}1$	Indicator for atom pair in 1 distance between hydrophobic atom and H-bond acceptor atom
	$CATS_{HydNeg}6$	Indicator for atom pair in 6 distance between hydrophobic atom and negative charge atom
	$CATS_{AroPos}7$	Indicator for atom pair in 7 distance between aromatic ring and positive charge atom
	$CATS_{DonDon}10$	Indicator for atom pair in 10 distance between H-bond donor atom and H-bond donor atom
	$CATS_{DonPos}3$	Indicator for atom pair in 3 distance between H-bond donor atom and positive charge atom
	$CATS_{AccPos}9$	Indicator for atom pair in 9 distance between H-bond acceptor atom and positive charge atom
ANN model	NoCdO	The number of double bonds between C atoms and O atoms
	$No_{AroRing}$	The number of aromatic rings
	No_{imide}	The number of imide groups
	No_{ketone}	The number of ketone groups
	$Kier_{shape}3$	Kier shape index order 3
	$(-)PSA_{MPEOE}$	The sum of VDW surface area sith values of MPEOE charges less than -0.2
	$CATS_{HydHyd}10$	Indicator for atom pair in 10 distance between hydrophobic atom and hydrophobic atom
	$CATS_{HydNeg}2$	Indicator for atom pair in 2 distance between hydrophobic atom and negative charge atom
	$CATS_{HydNeg}10$	Indicator for atom pair in 10 distance between hydrophobic atom and negative charge atom
	$CATS_{AroAro}6$	Indicator for atom pair in 6 distance between aromatic ring and aromatic ring
	$CATS_{AroPos}7$	Indicator for atom pair in 7 distance between aromatic ring and positive charge atom
	$CATS_{DonDon}3$	Indicator for atom pair in 3 distance between H-bond donor atom and positive H-bond donor atom
	$CATS_{DonDon}10$	Indicator for atom pair in 10 distance between H-bond donor atom and positive H-bond donor atom
	$CATS_{AccPos}9$	Indicator for atom pair in 9 distance between H-bond acceptor atom and positive charge atom

[a]Chemical Advanced Template Search descriptor, [b]The Modified Partial Equalization of Orbital Electronegativities charge (ref. 9-10)

$$PIC_{50} = -1.904 - 0.665NoCdO + 0.560No_{inide} + 0.447No_{ketone} + 0.011VSA_{hydunsat} - 3.340FraVSA_{hydunsat} + 0.249E_{state}S_{aaaC} + 0.200Kier_{shape}3 + 0.355 - CATS_{HydAro}10 - 0.384CATS_{HydPos}5 - 0.300CATS_{AroAro}6 + 0.718CATS_{AroPos}7 - 0.835CATS_{DonDon}3 - 1.434CATS_{DonNeg}8$$

$$n = 202, R^2 = 0.662, R^2_{CV} = 0.608, F = 25.04$$

전체 데이터에 대한 적합도의 경우, R^2값이 0.662로 다소 저조하나, 5묶음 교차 검증에 의한 Q^2값이 0.608

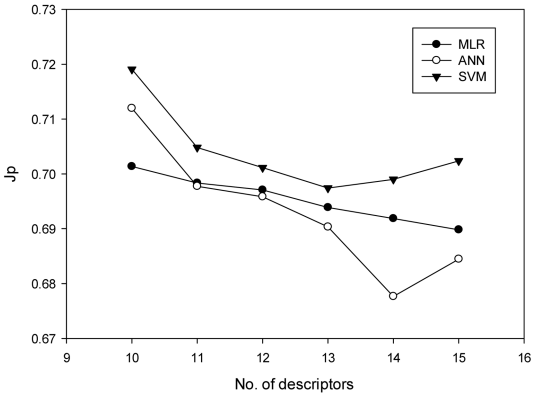

Fig. 2. Cross-validated Jp vs. number of descriptor in MLR, ANN and SVM model.

을 나타내므로 일반적으로 Q^2값이 0.5이상을 추천하는 기본적인 QSAR 모델 예측도 차원에서는 적합하다고 볼 수 있다.[17] MLR모델에 사용된 표현자와 그 회귀계수를 통해서 hERG 이온채널의 저해 정도를 화학 구조적으로 설명할 수 있다. 즉 회귀계수가 양의 값이면, 해당 표현자 값이 증가할수록 pIC_{50}이 증가함을 나타내며, 음의 값이면 그 반대임을 말한다. MLR 수식에 사용된 화합물의 부분 구조를 나타내는 표현자의 경우, carbonyl group의 수(No.CdO)가 감소할수록, imide group의 수가 증가할수록, ketone group의 수가 증가할수록 pIC_{50} 값이 커지므로 hERG 이온채널의 저해 효과가 커지는 것을 알 수 있으며, 이 관계는 회귀계수의 부호를 통해서 알 수 있다. 또한 분자의 불포화 소수성 표면적($VSA_{hydunsat}$)이 커지거나, fused aromatic ring 주변 원자와의 전기음성도 차($E_{state}S_{aaaC}$)가 커질수록, 화학구조에 탄소의 가지수($Kier_{shape}3$)가 많을수록 hERG 이온채널의 저해 효과가 커짐도 알 수 있다. 또한 6종의 상호 작용점(Aro:방향족, Hyd:소수성, Don:수소결합 주게, Acc:수소결합 받게, Pos:양전하, Neg:음전하 원자)간 결합거리의 유무를 이진수로 표현한 CATS표현자[11]를 통해서는 결합 분자와 수용체(receptor)간의 상호작용을 나타낼 수 있다. 이 모델에서는 방향족 원자와 양전하 원자간의 결합거리가 7개의 결합으로 떨어진 분자($CATS_{AroPos}7$) 들의 저해 효과가 크다고 판단할 수 있으며, 이것은 실제 수용체와의 상호 작용점을 나타낸 것으로 분석할 수 있다. 그 외 음의 회귀계수를 지니는 다른 CATS표현자들은 해당 상호 작용점간의 거리에 원자가 존재할 경우, 입체적인 장애 및 상호작용적으로 부적합하여 hERG이온채널를 저해하지 않음을 나타낸다. 그러므로 모델에 사용된 표현자들을 통해서 hERG이온채널과의 저해 관계를 화학 구조적, 상호작용적으로 설명함으로써, hERG이온채널로 인해 심장독성을 유발할 수 있는 화합물의 구조적 특징과 상호작용에 대하여 이해할 수 있었다.

비선형적인 관계를 알아보기 위하여 사용한 서포트 벡터 머신(SVM)과 인공 신경망(ANN)방법도 앞에서 언급한 forward selection방법과 교차 검증 방법을 결합하여 최적의 표현자 집합을 결정하였다. 결정된 SVM과 ANN방법의 표현자들은 13종, 14종이며, 그 목록은 *Table* 2에 나타내었다. SVM과 ANN방법의 매개변수 최적화 과정은 최적의 표현자 집합이 결정된 뒤에 진행하였다. SVM방법에 사용된 최적화 매개변수는 C, γ, ν이었고, C는 0.1, 1, 10, 100의 범위로, γ는 0.001-0.1 사이로 0.005씩, ν값은 0.1-0.5사이로 0.1씩 변화하여 5묶음 교차 검증 방법에 의해 결정되는 RMSECV 값이 가장 낮은 조건으로 결정하였다. ANN방법의 매개변수는 learning rate, momentum, hidden node수이며, learning rate와 momentum은 0.1-1.0사이로 0.1씩, hidden node의 수는 6, 7, 8개를 적용하여 마찬가지로 모든 조건에서 RMSECV값이 가장 낮은 조건으로 결정하였다. SVM방법에서 최적화된 매개변수의 값은 C=100, γ=0.005, ν=0.5였으며, ANN방법에서 최적화된 매개변수의 값은 learning rate=0.1,

Table 3. Comparative statistical performance of MLR, SVM, and ANN models

Method	No. descriptor	parameter	Training			5-fold Cross-validation			
			R^2	RMSE	MAE	R^2_{cv}	RMSECV[a]	MAECV[b]	Jp[c]
MLR	15		0.662	0.714	0.540	0.608	0.771	0.580	0.690
SVM	13	C=100, gamma=0.005, Nu=.5	0.660	0.720	0.554	0.595	0.783	0.599	0.697
ANN	14	Learning rate=0.1, momentum=0.7, no.hidden nodes=6	0.680	0.696	0.527	0.611	0.768	0.588	0.678

[a] Root mean square error of cross-validation
[b] mean absolute error of cross-validation
[c] Jp statistics

momentum=0.7, hidden node의 수는 6이였고, 최적화된 SVM과 ANN모델의 통계 결과는 *Table* 3에 나타내었다.

교차 검증에 의한 각 학습방법의 적합도를 비교하였을 때, 모델간의 차이는 크지 않았으나, 모든 적합도(R^2_{cv}, RMSECV, MAECV)에서 ANN모델이 가장 우수함을 알 수 있었고, 그 다음 MLR, SVM 순 이었다. 각 모델에 사용된 표현자들을 비교해 보면(*Table* 2), 가장 우수한 결과를 내었던 ANN과 MLR방법의 표현자들이 서로 공통적으로 사용된 것을 확인 할 수 있었고(No.CdO, No_{imide}, No_{ketone}, $Kier_{shape}3$, $CATS_{AroAro}6$, $CATS_{AroPos}7$, $CATS_{DonDon}3$), SVM방법에 사용된 표현자와는 서로 다름을 확인 할 수 있었다. 이를 통해서 ANN모델은 MLR모델에 비선형적으로 개선된 형태임을 알 수 있었고, SVM모델은 ANN과는 다른 비선형적인 모델임을 알 수 있었다. *Fig.* 3은 가장 우수한 결과를 나타낸 ANN방법의 예측된 pIC_{50}와 실험값 pIC_{50} 을 비교하여 도표로 나타내었다.

Bain[7]등은 hERG 이온채널 저해 물질의 기준을 1 μM (pIC_{50}=0)로 정하고, 이보다 낮은 경우를 hERG 저해제로 구분하여 분류(classification)모델을 구현한 바가 있다. 본 연구에서도 이 기준을 적용하여 3가지 학습방법의 분류 정확도(accuracy) 계산하였고, 그 결과

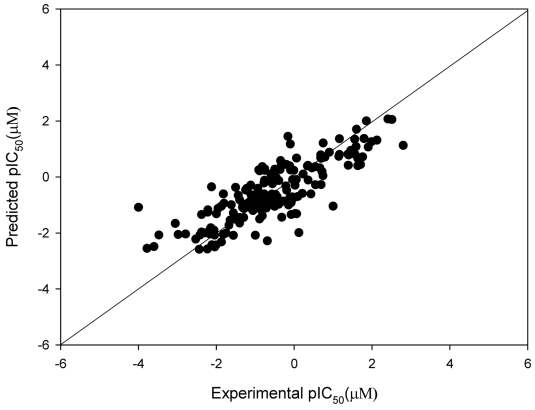

Fig. 3. A plot of experimental vs. predicted pIC_{50}(μM) values by ANN model.

Table 4. Classification results for best models (MLR, SVM, ANN)

Method	Sensitivity[a]	Specificity[b]	Accuracy[c]
MLR	93.29	62.26	85.15
SVM	98.66	79.25	93.56
ANN	94.63	64.15	83.63

[a] Percentage of correctly predicted set of inhibitors,
[b] Percentage of correctly predicted set of non-inhibitors,
[c] Percentage of correctly predicted results (threshold for the separation of inhibitors/non-inhibibitors : 1 μM)

를 *Table* 4에 나타내었다. 3가지 학습방법 중에서 SVM방법의 정확도가 93.5%로 가장 뛰어 났으며, 특히 inhibitor를 분류하는 예측 능력이 다른 방법보다 뛰어난 것을 확인 할 수 있었다. 이를 통해서 SVM방법은 다른 방법과 비교하여 분류 기준이 되는 1 μM 주변에 저해농도를 나타낸 화합물에 대하여 실험값과의 오차가 적은 것임을 말해 준다. 종합적으로 비교해 보면, 실제 저해농도를 예측하기에는 ANN와 MLR방법이 우수하나, 1 μM의 기준으로 한 저해제 분류법으로는 SVM방법이 우수하므로, 목적에 맞게 선별하여 사용할 수 있을 것으로 예상한다.

학습방법을 통해 세워진 QSAR모델은 많은 표현자로부터 도출해 내기 때문에 우연하게 실험 결과를 예측하게 될 수도 있다. 이러한 부분은 우연 상관(chance correlation)이라고 하며, 이를 확인하기 위해서는 주로 Y-scrambling 결과와 비교한다. Y-scrambling test는 종속변수로 사용된 각 화합물의 실험값들을 무작위로 섞어서 새로운 데이터를 생성한 뒤, 본 연구에서 진행한 학습방법으로 그대로 진행하여 모델의 적합도를 평가하는 방법이다. 최적 모델의 적합도와 Y-scrambling을 한 적합도와 비교하여 큰 차이를 나타내면, 본 연구의 모델이 우연 상관으로 얻어진 결과가 아님을 증명하게 된다. 본 연구에서는 Y-scrambling과정을 30번 반복하여 얻은 결과의 적합도를 *Table* 5에 나타내었고, *Table* 3에 나타낸 학습방법의 적합도와 비교하여 모두 큰 차이를 나타내었다. 그러므로, 본 연구에서 개발된 QSAR모델들의 견고성(robustness)을 확인할 수 있었다.

Table 5. Cross-validated results after 30 Y-scrambling test

Method	Cross-validation (Y-scrambling)					
	$R^2_{cv\ ave.}$	$R^2_{cv\ range}$	$RMSECV_{ave.}$	$RMSECV_{range}$	$MAECV_{ave.}$	$MAECV_{range}$
MLR	0.086	0.014-0.114	1.185	1.162-1.257	0.916	0.893-0.953
SVM	0.184	0.136-0.238	1.112	1.075-1.144	0.856	0.828-0.873
ANN	0.124	0.039-0.206	1.154	1.097-1.211	0.895	0.838-0.934

4. 결 론

본 연구에서는 202종의 의약품 주성분과 약유사 성분에 대하여 hERG 이온채널의 저해농도를 예측하는 연구를 학습방법인 MLR, SVM, ANN 방법을 이용하여 QSAR분석 연구를 수행하였다. 분자의 특성을 나타내는 표현자는 모두 2차원 화학 구조에서 빠르게 계산이 될 수 있는 것 만을 사용하였으므로, 신약 설계 단계에서 hERG 관련 심장독성과 관련하여 가상적인 화합물 라이브러리의 독성 예측에 대량으로 적용할 수 있도록 하였다. 본 연구로 구축된 QSAR모델 중에서, MLR방법은 모델에 사용된 표현자를 이용하여 hERG 저해물질과 관련된 화학 구조적 특징과 수용체와의 상호작용을 설명할 수 있었다. ANN방법은 3가지 방법 중에서 저해농도 예측과 관련하여 가장 우수한 예측 능력을 나타내었고, SVM방법은 1 μM를 기준으로 한 저해제 분류 예측에서 가장 우수한 결과를 나타내었다. 본 연구에 사용된 모든 학습방법은 5 묶음 교차 검증 법과 Y-scrambling test에 의해서 얻은 적합도(R^2_{CV}, RMSECV, MAECV)를 통해서 모델의 예측 능력과 견고함을 확인 할 수 있었다.

감사의 글

본 연구는 2011년도 한남대학교 교비연구비의 지원으로 수행되었으며 이에 감사 드립니다.

참고문헌

1. J. I. Vandenberg, B. D. Walker and T. J. Campbell, *Trends Pharmacol. Sci.,* **22**(5), 240-246 (2001).
2. M. Jalaie and D. D. Holsworth, *Mini-Rev. Med. Chem.,* **5**(12), 1083-1091 (2005).
3. S. B. Gunturi, K. Archana, A. Khandelwal and R. Narayanan, *QSAR Combi. Sci.,* **27**(11-12), 1305-1317 (2008).
4. K. Yoshida and T. Niwa, *J. Chem. Inf. Model.,* **46**(3), 1371-1378 (2006).
5. A. M. Doweyko, *J. Comput.-Aided Mol. Des.,* **22**(2), 81-89 (2008).
6. K. M. Thai and G. F. Ecker, *Chem. Biol. Drug Des.,* **72**(4), 279-289 (2008).
7. W. Bains, A. Basman and C. White, *Prog. Biophys. Mol. Biol.,* **86**(2), 205-233 (2004).
8. S. K. Lee, S. H. Park, I. H. Lee and K. T. No, PreADMET Ver.v2.0, BMDRC: Seoul, Korea, 2007.
9. K. T. No, J. A. Grant, M. S. Jhon and H. A. Scheraga, *J. Phys. Chem.,* **94**(11), 4740-4746 (1990).
10. K. T. No, J. A. Grant and H. A. Scheraga, *J. Phys. Chem.,* **94**(11), 4732-4739 (1990).
11. G. Schneider, W. Neidhart, T. Giller and G. Schmid, *Angew. Chem. Int. Ed. Engl.,* **38**(19), 2894-2896 (1999).
12. N. R. Draper and H. Smith, In 'Applied Regression Analysis', 2nd Ed., pp 294-379, John Wiley & Sons Inc., New York, 1981.
13. C. Cortes and V. Vapnik, *Mach. Learn.,* **20**(3), 273-297 (1995).
14. B. Schölkopf, A. J. Smola, R. C. Williamson and P. L. Bartlett, *Neural Comput.,* **12**(5), 1207-1245 (2000).
15. D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature,* **323**(6088), 533-536 (1986).
16. Rapidminer Ver.5.0, Rapid Miner is unquestionable the world-leading open-source system for data mining, Rapid-I: Dortmund, Germany, 2010.
17. B. L. Podlogar, I. Muegge and L. J. Brice, *Curr. Opin. Drug Discovery Dev.,* **4**(1), 102-109 (2001).