论文

5-羧基苯并咪唑类 HCV NS5B 聚合酶抑制剂的 HQSAR 研究及分子设计

王子恒¹, 陈娴¹, 张玉芳¹, 王月平^{2*}, 何严萍^{1*}

1. 云南大学化学科学与工程学院/药学院, 教育部自然资源药物化学重点实验室, 昆明 650091

2. 昆明理工大学理学院应用化学系, 昆明 650050

*通讯作者, E-mail: wangyp@kmust.edu.cn; yphe@ynu.edu.cn

收稿日期: 2016-10-19; 接受日期: 2016-11-14; 网络发表日期: 2017-02-17

国家自然科学基金(编号: 21262044, 21362017)资助项目

摘要 利用分子全息技术研究了 129 个 5-羧基苯并咪唑类 HCV NS5B 聚合酶抑制剂的结构与活性之间的关系. 讨论了分子碎片大小、碎片区分参数及全息长度对模型质量的影响. 利用偏最小二乘法(partial least square, PLS)建立了一组以 99 个化合物为训练集的最优模型, 该模型的交叉验证相关系数 $q^2=0.820$, 非交叉验证相关系数 $r^2=0.963$, 标准偏差 $SEE=0.213$; 用最优模型对由 30 个化合物组成的测试集进行预测, 得到其相关系数 $r^2_{pred}=0.98$, 表明了该模型具有良好的预测能力及拟合能力. 利用色码图对模型中不同原子及不同结构的贡献进行了解释, 在此基础上根据最优 HQSAR 模型设计了几种具有良好抗 HCV 活性的苯并咪唑类 HCV NS5B 聚合酶抑制剂分子, 为新型 HCV NS5B 聚合酶抑制剂的设计和优化提供了参考.

关键词 5-羧基苯并咪唑类衍生物, HCV NS5B 聚合酶抑制剂, 分子全息定量构效关系, 色码, 分子设计

1 引言

丙型肝炎病毒(HCV)所致的一种全球性分布的肝脏炎症性疾病. 目前, 丙型肝炎病毒(HCV)感染者已超过 1.85 亿, 约 80% 的感染者无法彻底清除病毒而演变为慢性肝炎, 成为引起肝硬化、肝细胞癌的主要因素. 我国丙型肝炎患病率高达 3.2%, 约有 3800 万 HCV 感染者^[1-3]. HCV 严重地威胁着人类的健康和生命, 现已成为全球性的健康问题. 随着对 HCV 生命周期的

认识, 直接抗丙肝药物研究取得重要进展, 该类药物通过作用于 HCV 病毒编码蛋白或致病相关宿主细胞因子, 从不同阶段阻断 HCV 在肝细胞中的复制, 从而达到抗病毒疗效. 现已发现的抗 HCV 药物主要作用靶点有: NS2/NS3 蛋白酶、NS3/NS4A 丝氨酸蛋白酶、NS4B、NS5A 及 NS5B 聚合酶等^[4]. 其中 NS5B 聚合酶是 HCV 基因组 RNA 复制的关键酶^[5], 该酶以 HCV 释放出的正链 RNA 为模板催化 RNA 负链的合成, 再通过负链复制出大量的正链, 与多聚蛋白体裂解产生的病毒蛋白共同包装为成熟病毒颗粒^[6], 由于

引用格式: 王子恒, 陈娴, 张玉芳, 王月平, 何严萍. 5-羧基苯并咪唑类 HCV NS5B 聚合酶抑制剂的 HQSAR 研究及分子设计. 中国科学: 化学, 2017, 47: 350-360
Wang Z, Chen X, Zhang Y, Wang Y, He Y. HQSAR study and molecular design of benzimidazole 5-carboxylic derivatives of hepatitis C virus NS5B polymerase. *Sci Sin Chim*, 2017, 47: 350-360, doi: 10.1360/N032016-00191

人类细胞不表达与 NS5B RNA 依赖性 RNA 聚合酶(RNA-dependent RNA polymerase, NS5B RdRp)功能相同的酶, 这使针对 NS5B RdRp 的抑制剂具有良好的选择性, 从而成为抗 HCV 药物研究的重要领域之一^[7].

苯并咪唑类化合物是一类重要的 HCV NS5B 抑制剂, 2001 年 Tobacco 公司和 Boehringer-Ingelheim 公司筛选出具有 NS5B 抑制活性的苯并咪唑类化合物^[8,9], 随后 Hirashima 等^[10-12]在苯并咪唑环 C-5 引入羧基, 结合 N-1 和 C-2 位结构修饰发现了系列较高活性的 5-羧基苯并咪唑类 NS5B 聚合酶抑制剂, 其中化合物 JTK109 已进入一期临床研究^[13]. 目前, 人们已合成结构多样的 5-羧基苯并咪唑类 NS5B 抑制剂并对其构效关系进行了探讨, 但这些研究较多集中在基于复合物晶体或分子模型的构效分析, 关于定量构效关系(QSAR)的研究却报道较少.

应用定量构效关系技术建立起化合物分子结构与活性之间的关系, 是指导新化合物合成及分析化合物性质的有效方法. 分子全息定量构效关系(HQSAR)方法于 1997 年由 Hurst 和 Heritage 建立^[14], 该方法是将系列化合物的分子结构转换为各类型的分子碎片, 对这些碎片进行全息编码后作为 QSAR 的描述符, 与生物活性之间经程序化建立相应的结构-活性关系, 并得出生物活性的预测结果. 与其他 QSAR 方法相比, HQSAR 方法既避免了经典 QSAR 方法的共线性和参数计算与选择的问题, 又避免了 3D-QSAR 等技术所必需的分子叠合过程, 其分子碎片和全息的产生完全自动生成, 非常迅速, 通过 HQSAR 方法得到的活性贡献图, 能够明确显示对活性产生影响的取代基及其位置, 从而在分子设计中精确定位和寻找合适取代基. 作为一种简单、快速和具有较高预测能力的 QSAR 新技术, HQSAR 方法已成功应用于指导各类型分子设计^[15-17]. 本文采用分子全息定量构效关系(HQSAR)技术分析了 Hirashima

和 Beaulieu 课题组^[10-13]所合成的一系列 5-羧酸苯并咪唑类 HCV NS5B 聚合酶抑制剂分子对 HCV 的半数抑制浓度 IC₅₀ 的构效关系, 建立了稳定且预测能力良好的 HQSAR 模型, 并以此为指导, 对本课题组^[18]所发现的 2-硫乙酰芳胺苯并咪唑类抗 HCV 活性化合物进行结构修饰, 设计了一系列新型的 5-羧基苯并咪唑类 HCV NS5B 聚合酶抑制剂分子, 采用该模型进行预测, 设计的化合物对 HCV NS5B 聚合酶均有较高的抑制活性, 本研究为新型苯并咪唑类 HCV NS5B 聚合酶抑制剂的分子设计和优化提供了参考.

2 计算方法

2.1 化合物及活性数据

用于 HQSAR 分析的化合物结构及活性数据选自文献^[10-13], 所选择的 129 个化合物具有足够的活性变化梯度和结构多样性, 活性指标用对感染了 HCV 的细胞半数有效抑制浓度 IC₅₀ (M) 的负对数 pIC₅₀ (-lgIC₅₀) 表示. 随机选取了 99 个化合物作为训练集建立模型, 其余 30 个化合物为测试集(表 1 中带*号的分子)用于检验模型的预测能力. 具体结构如图 1 和表 1 所示, 活性数据见表 2.

2.2 分子建模

本文所有计算工作均在 Tripos 公司 Sybyl 7.2 分子模拟建模软件上完成^[19]. 计算中各项参数除特别指明外均采用缺省值, 使用 SYBYL/Base、Minimize、HQSAR 等模块. 首先利用 Sybyl 软件中的 Sketch 选项构建 129 个苯并咪唑类分子的初始三维结构, 随后采用 Tripos 分子力学程序, Minimize 进行能量优化, 负载 Gasteiger Huckel 原子电荷, 能量收敛标准为 0.001 kcal/mol, 迭代 5000 次, 其余均采用 Sybyl 程序的默认值, 优化后得到各个分子的低能构象用于 HQSAR 的分析.

图 1 NS5B 聚合酶抑制剂分子的通用结构

表 1 129 个化合物的结构^{a)}

序号	通用结构	R ₁	R ₂	序号	通用结构	R ₁	R ₂	R ₃
001	a	cyclohexyl	2-pyridine	065	b	cyclopentyl	Ph-CH ₂ -	-
002	a	H	2-pyridine	066*	b	cyclopentyl	4-Me-Ph-CH ₂ -	-
003*	a	cycloheptyl	2-pyridine	067	b	cyclopentyl	4-CF ₃ -Ph-CH ₂ -	-
004	a	cyclopentyl	2-pyridine	068*	b	cyclopentyl	4-t-Bu-Ph-CH ₂ -	-
005*	a	cyclobutyl	2-pyridine	069*	b	cyclopentyl	4-(COOH)-Ph-CH ₂ -	-
006*	a	isoamyl	2-pyridine	070*	b	cyclopentyl	4-Cl-Ph-CH ₂ -	-
007	a	4-CH ₃ -cyclohexyl	2-pyridine	071*	b	cyclopentyl	4-OMe-Ph-CH ₂ -	-
008*	a	2-CH ₃ -cyclohexyl	2-pyridine	072	b	cyclopentyl	4-Pyridyl-CH ₂ -	-
009*	a	-Ph	2-pyridine	073	b	cyclopentyl	2-Cl-Ph-CH ₂ -	-
010	a	cyclohexyl		074	b	cyclopentyl	3-Cl-Ph-CH ₂ -	-
011	a	cyclohexyl	-Ph	075	b	cyclopentyl	3,5-Cl-Ph-CH ₂ -	-
012	a	cyclohexyl		076	b	cyclopentyl	3,4-Cl-Ph-CH ₂ -	-
013	a	cyclohexyl		077	b	cyclohexyl	Bn	-
014*	a	cyclohexyl	-C ₆ H ₁₁	078	b	cyclohexyl	Ph(CH ₂) ₃ -	-
015	a	cyclohexyl	-CH=CH(CH ₃) ₂	079	b	cyclohexyl	Ph(CH ₂) ₅ -	-
016	a	cyclohexyl	-COOCH ₂ CH ₃	080	b	cyclohexyl	PhO(CH ₂) ₂ -	-
017	a	cyclohexyl	-COPh	081	b	cyclohexyl	Ph ₂ CHCH ₂ -	-
018	a	cyclohexyl	2-Pyridine	082	b	cyclohexyl	(2-Ph)PhCH ₂ -	-
019	a	cyclohexyl	3-Pyridine	083	b	cyclohexyl	(2-Ph)Ph-	-
020	a	cyclohexyl	2-oxole	084	b	cyclohexyl	(3-Ph)PhCH ₂ -	-
021	a	cyclohexyl	3-oxole	085	b	cyclohexyl	(2-Ph)PhCH ₂ CH ₂ -	-
022*	a	cyclohexyl	2-Pyrrole	086	b	cyclohexyl	Ph ₂ CH-	-
023*	a	cyclohexyl	2-midazole	087	b	cyclohexyl	PhCH ₂ -	-
024	a	cyclohexyl	1-CH ₃ -2-Pyrrole	088	b	cyclohexyl	4-Me-Ph-CH ₂ -	-
025	a	cyclohexyl	3-thiophene	089	b	cyclohexyl	4-Cl-Ph-CH ₂ -	-
026*	a	cyclohexyl	2-thiophene	090*	b	cyclohexyl	4-F-Ph-CH ₂ -	-
027	a	cyclohexyl	2-thiazole	091	b	cyclohexyl	3-F-Ph-CH ₂ -	-
028	b	cyclohexyl		092	b	cyclohexyl	Ph-CH ₂ -	2-F
029	b	cyclohexyl		093*	b	cyclohexyl	4-F-Ph-CH ₂ -	2-F
030	b	cyclohexyl		094*	b	cyclohexyl	4-CONMe ₂ -Ph-CH ₂ -	2-F
031	b	cyclohexyl		095	b	cyclohexyl	3-Pyridyl-CH ₂ -	2-F
032	b	cyclohexyl		096	b	cyclohexyl	(3-Pyridyl)PhCH-	2-F
033	b	cyclohexyl		097	c		4-Cl-Ph-	-

续表 1

序号	通用结构	R ₁	R ₂	序号	通用结构	R ₁	R ₂	R ₃
034	d	-Cl	Ph	098	c		4-Cl-OPh	-
035*	d	-Cl	4-Cl-Ph	099	c		-OMe	-
036*	d	-Cl	4-Me-Ph	100	c		1-Piperidyl	-
037	d	-Cl	4-OMe-Ph	101*	c		4-OH-Piperidyl	-
038*	d	-Cl	4-COOH-Ph	102*	c		4-COOH-Piperidyl	-
039	d	-Cl	4-CONH ₂ -Ph	103	c		4-CH ₃ -Piperidyl	-
040	d	-Cl	4-CN-Ph	104	c		-N(CH ₂) ₂ N-Ac	-
041	d	-Cl	3-Cl-Ph	105	c		-N(CH ₂) ₂ O	-
042	d	-Cl	3-Pyridine	106	c		-N(CH ₂) ₂ O	-
043	d	H	4-Cl-Ph	107	c		-N(CH ₂) ₂ O	-
044	d	4-OMe	4-Cl-Ph	108	c		O(CH ₂) ₂ N-	-
045	d	4-SO ₂ NH ₂	4-Cl-Ph	109	c	H	O(CH ₂) ₂ N-	-
046	d	4-SO ₂ NMe ₂	4-Cl-Ph	110	c	Cl	O(CH ₂) ₂ N-	-
047*	d	4-Ac	4-Cl-Ph	111	e	-	-	2-F
048	d	4-NMe ₂	4-Cl-Ph	112	e	-	-	2-Cl
049*	d	4-NHAc	4-Cl-Ph	113	e	-	-	2-OMe
050	d	4-COOH	4-Cl-Ph	114	e	-	-	2-CF ₃
051	d	4-CN	4-Cl-Ph	115	e	-	-	3-F
052	d	4-CONH ₂	4-Cl-Ph	116	e	4-Cl	CONHMe	2-F
053	d	4-CONHMe	4-Cl-Ph	117	e	4-Cl	CONH <i>n</i> -Pr	2-F
054*	d	4-CONMe ₂	4-Cl-Ph	118	e	4-Cl	CONH <i>i</i> -Pr	2-F
055	d	4-CONHBn	4-Cl-Ph	119	e	4-Cl	-CONHC ₂ H ₅ OH	2-F
056*	d	5-CONMe ₂	4-Cl-Ph	120*	e	4-Cl	CONMe ₂	2-F
057	c		O(CH ₂) ₂ N-	121	e	4-Cl	-CO-Piperidyl	2-F
058	c		O(CH ₂) ₂ N-	122	e	4-Cl		2-F
059	c	NMeAc	O(CH ₂) ₂ N-	123	e	4-Cl	4-OH-CO-Piperidyl	2-F
060*	c	NHAc	O(CH ₂) ₂ N-	124	e	4-Cl	NHAc	2-F
061	c		O(CH ₂) ₂ N-	125	e	4-Cl	NMeAc	2-F
062	c	Ni-PrAc	O(CH ₂) ₂ N-	126*	e	4-Cl	NEtAc	2-F
063*	c	CONMe ₂	O(CH ₂) ₂ N-	127	e	4-Cl	NMeCO <i>i</i> -Pr	2-F
064	c	CONH <i>i</i> -Pr	O(CH ₂) ₂ N-	128	e	4-Cl	Ni-PrAc	2-F
				129	e	4-Cl	NHCONMe ₂	2-F

a) *表示化合物为测试集

表 2 129 个化合物的实验值、计算值和误差^{a)}

No.	pIC ₅₀	Pred. ^{b)}	Resi. ^{c)}	No.	pIC ₅₀	Pred. ^{b)}	Resi. ^{c)}	No.	pIC ₅₀	Pred. ^{b)}	Resi. ^{c)}
001	5.367	5.230	0.137	044	7.432	7.296	0.136	087	6.229	6.343	-0.114
002	3.301	3.258	0.043	045	7.796	7.900	-0.104	088	5.921	5.965	-0.044
003*	4.310	5.017	-0.707	046	7.638	7.628	0.010	089	6.032	6.224	-0.192
004	4.921	4.937	-0.016	047*	7.745	7.292	0.453	090	6.097	6.107	-0.010
005*	4.119	4.966	-0.847	048	7.201	7.129	0.072	091	6.119	6.139	-0.020
006*	3.857	5.071	-1.214	049*	7.658	7.173	0.485	092	7.018	6.321	0.697
007	4.886	4.405	-0.104	050	7.620	7.441	0.179	093	6.523	6.487	0.036
008*	3.444	4.777	0.109	051	7.387	7.361	0.026	094	6.194	6.896	-0.702
009*	5.367	5.757	-0.390	052	7.886	7.666	0.220	095	5.959	6.214	-0.255
010	4.886	4.817	0.069	053	7.745	7.880	-0.135	096	6.569	6.526	0.043
011	5.066	5.239	-0.173	054*	7.824	7.807	0.017	097	7.770	7.768	0.002
012	3.921	3.983	-0.062	055	7.538	7.476	0.062	098	6.638	6.282	-0.190
013	4.237	4.304	-0.067	056*	7.495	7.469	0.026	099	6.387	6.368	0.019
014*	4.699	4.740	-0.041	057	7.149	7.233	-0.084	100	6.824	6.788	0.036
015	4.721	4.839	-0.118	058	7.046	6.913	0.133	101*	6.824	6.471	0.353
016	4.208	4.171	0.037	059	7.292	7.252	0.040	102*	6.745	6.584	0.161
017	4.357	4.345	0.012	060	7.155	6.755	0.400	103	6.886	6.876	0.100
018	4.553	4.909	-0.356	061	7.004	7.077	-0.073	104	6.959	6.974	0.015
019	4.770	4.656	0.114	062	7.377	7.418	-0.041	105	7.180	7.268	-0.088
020	5.432	5.395	0.037	063*	7.167	7.513	-0.346	106	6.000	5.976	0.024
021	5.796	5.293	0.503	064	7.319	7.454	-0.135	107	6.585	6.573	-0.012
022*	5.229	4.840	0.389	065	5.495	5.600	-0.105	108	7.180	7.268	-0.088
023*	5.000	5.147	-0.147	066*	5.796	5.537	0.259	109	6.886	6.885	-0.001
024	5.745	5.763	-0.018	067	5.745	5.774	-0.029	110	7.000	6.877	0.123
025	5.201	5.460	-0.259	068*	5.824	5.586	0.238	111	7.000	6.803	0.197
026*	5.469	5.239	0.230	069*	5.174	5.762	-0.588	112	6.456	6.642	-0.186
027	5.337	5.381	-0.044	070*	5.921	5.997	-0.076	113	5.959	5.962	-0.003
028	6.237	6.168	0.069	071*	5.678	5.735	-0.057	114	6.032	6.143	-0.111
029	6.770	6.880	-0.110	072	5.377	5.519	-0.142	115	6.420	6.441	-0.021
030	7.046	7.000	0.046	073	5.638	5.727	-0.089	116	7.921	7.995	-0.074
031	7.056	6.986	0.070	074	5.796	5.685	0.111	117	7.678	7.636	0.042
032	7.276	7.225	0.051	075	5.770	5.496	0.274	118	7.886	7.818	0.068
033	6.824	6.960	-0.136	076	5.886	5.873	0.013	119	7.678	7.657	0.021
034	6.699	6.858	-0.159	077	5.854	5.893	-0.039	120*	7.921	7.910	0.011
035*	7.301	7.121	0.180	078	6.097	6.139	-0.042	121	7.745	7.652	0.093
036*	7.009	6.706	0.303	079	6.000	5.943	0.057	122	7.854	7.847	0.007
037	7.000	7.157	-0.157	080	5.921	5.755	0.166	123	7.854	7.850	-0.004
038*	7.237	7.107	0.130	081	6.009	6.075	-0.066	124	7.854	7.909	-0.055
039	7.161	7.347	-0.186	082	6.523	6.375	0.148	125	7.796	7.858	-0.062
040	7.260	7.172	0.088	083	6.237	6.098	0.139	126*	7.721	7.636	0.085
041	6.319	6.781	-0.462	084	6.409	6.322	0.087	127	7.495	7.420	0.075
042	6.538	6.372	0.166	085	6.469	6.356	0.113	128	7.678	7.671	0.007
043	7.108	7.052	0.056	086	6.229	6.343	-0.114	129	7.658	7.820	-0.162

a) *表示化合物为测试集; b) Pred 为计算值; c) Resi 为误差

2.3 HQSAR 模型的构建

采用 SYBYL/HQSAR 模块产生分子全息, 过程如下^[20]: 首先, 将训练集中每个分子结构划分为包含一定原子个数的分子碎片, 碎片中所包含的结构类型称之为碎片区分参数. 一般所使用的碎片区分

参数包括: 原子(atoms, A)、键(bonds, B)、连接性(connections, Co)、氢原子(hydrogen atoms, H)、手性特征(chirality, Ch)以及氢键供体或受体(donor & acceptor, DA)等. 其中参数 A 用来区分原子的种类; B 用来区分原子间所形成的化学键差异; Co 可以区分

碎片内部原子杂化状态; Ch 可以显示碎片中的原子手性以及化学键的立体化学信息; DA 用来显示碎片的氢键给体或受体情况. 碎片中所包含的原子数目称为碎片大小参数, 一般由最小原子数 N 和最大原子数 M 指定. 其中, 较小分子碎片为 1~3, 中等分子碎片为 4~7, 较大分子碎片为 3~10^[21]. 然后, 采用循环冗余校验算法(cyclic redundancy check algorithm, CLC)将各个分子碎片映射为 0~231 的伪随机整数, 使得每个分子可被表示为一定长度的整数串. 进而采用哈希法(Hashing)将它们转换为具有相同长度的整数串, 即为分子全息. HQSAR 模块提供了 12 个缺省质数(53~401)为全息长度^[22]. 通过调整碎片大小、碎片结构类型、全息长度等参数, 可以得到不同的分子全息. 随后以各个分子全息为自变量, 活性为因变量, 采用偏最小二乘法(PLS)进行线性回归分析, 采用留一法(LOO)进行交叉验证, 计算交叉验证相关系数 q^2 , 选择 q^2 最大时的主成分数为最佳主成分数(NOC), 应用最佳主成分数进行非交叉验证 PLS 分析进行模型的构建, 综合考虑模型的 q^2 、非交叉验证相关系数(r^2), 非交叉验证标准误差(SEE)等参数来衡量

模型的预测能力、稳健程度和拟合能力, 得到最终的 HQSAR 模型, 并通过 HQSAR 模型色码图直接反映原子或基团对化合物活性的贡献, 得到对分子进行修饰的有利信息^[22].

3 结果与讨论

3.1 碎片区分参数的确定

HQSAR 模型的统计能力与 3 个参数有关: 分子碎片区分参数、碎片大小参数以及分子全息长度, 通过调整它们之间的组合, 可获得最佳 QSAR 模型^[23]. 首先, 在默认分子碎片(4~7)及较小的分子碎片(1~3)下, 考察不同碎片区分参数以及分子全息长度对模型的影响, 结果见表 3, 因本项研究中的分子均无手性, 因此在 HQSAR 分析中不考虑此参数. 由表 3 中数据可知, 当使用的碎片区分参数为 B/Co/H/D 时, 可得到最佳的模型结果. 此时, 留一法交叉验证相关系数 q^2 为 0.815, 具有显著的统计意义, 非交叉验证相关系数 r^2 为 0.963, 最佳主成分数(NOC)为 8, 最佳分子全息长度(BL)为 307.

表 3 不同碎片区分参数以及分子全息长度对模型的影响

碎片区分参数	1~3					4~7				
	q^2 ^{a)}	BL ^{b)}	NOC ^{c)}	r^2 ^{d)}	SEE ^{e)}	q^2 ^{a)}	BL ^{b)}	NOC ^{c)}	r^2 ^{d)}	SEE ^{e)}
A	0.530	59	1	0.625	0.654	0.613	59	4	0.839	0.435
B	0.547	71	3	0.694	0.596	0.676	61	4	0.832	0.445
Co	0.535	71	1	0.637	0.643	0.565	257	1	0.646	0.635
A/B	0.551	257	1	0.631	0.648	0.689	199	10	0.974	0.179
A/H	0.549	53	3	0.686	0.604	0.598	257	3	0.736	0.554
A/D	0.542	61	1	0.631	0.648	0.740	97	12	0.964	0.215
B/Co	0.607	199	2	0.694	0.594	0.678	257	4	0.886	0.336
B/H	0.547	71	3	0.694	0.596	0.676	61	4	0.832	0.445
Co/H	0.535	71	1	0.637	0.643	0.565	257	1	0.646	0.635
H/D	0.526	61	1	0.645	0.636	0.703	61	12	0.931	0.299
A/B/Co	0.610	307	2	0.687	0.600	0.691	257	10	0.978	0.166
A/B/H	0.539	53	1	0.624	0.654	0.654	353	14	0.984	0.143
A/B/D	0.578	257	1	0.652	0.630	0.798	353	7	0.963	0.211
A/Co/D	0.577	53	1	0.657	0.625	0.700	83	3	0.826	0.450
B/Co/H	0.698	199	7	0.897	0.354	0.678	257	4	0.886	0.366
Co/H/D	0.562	53	1	0.655	0.627	0.700	83	4	0.842	0.431
A/B/Co/H	0.630	61	14	0.919	0.327	0.587	71	2	0.704	0.584
A/B/Co/D	0.695	199	6	0.893	0.359	0.812	199	8	0.968	0.198
A/Co/H/D	0.569	59	2	0.651	0.634	0.648	61	2	0.715	0.573
B/Co/H/D	0.698	199	7	0.897	0.354	0.815	307	8	0.963	0.212
A/B/Co/H/D	0.668	97	8	0.903	0.345	0.692	401	12	0.980	0.160

a) q^2 为交叉验证相关系数; b) BL 为最佳分子全息长度; c) NOC 为最佳主成分数; d) r^2 为非交叉验证相关系数; e) SEE 为非交叉验证标准误差

3.2 碎片大小的确定

在给出分子碎片时, 碎片的大小是可以改变的, 因为碎片的大小对分析结果也有一定的影响. 碎片长度的不同所代表的分子结构信息亦有所不同, 一般碎片长度为 1 可以区分不同的原子类型. 较小碎片长度范围(1~3)可区分原子类型、化学键及官能团性质; 中等长度的碎片范围(4~7)可区分脂肪环和芳香环的性质、取代基的类型和位置等特征^[24]. 为了得到最佳的预测模型, 以 B/Co/H/D 为最优碎片区分参数, 采用不同的碎片长度建立 HQSAR 模型, 表 4 反映了碎片长度对 HQSAR 分析结果的影响. 表 4 数据表明, 当采用分子碎片为(2~7)时, 统计模型为最优, 其中 $q^2=0.820$, $r^2=0.963$.

3.3 最佳模型的评价与检验

在 HQSAR 分析过程中, r^2 和 q^2 是检验模型质量最重要的标准. 其中 $r^2 \geq 0.9$ 说明建立的模型具有良好的相关性, q^2 可检验模型的预测能力. 一般认为 $q^2 > 0.5$ 的模型具有较好的预测能力. 本文所得的最佳模型的 $r^2=0.963$, $q^2=0.820$, 这表明所建立的模型具有良好的线性相关性和较强的预测能力. 另一个评价模型好坏的指标是非交叉验证标准误差(SEE). 当 SEE 与活性数值范围的比例小于 10%时, 可以认为模型良好. 本文所得的最佳模型的 SEE 为 0.213, 活性数据的数值范围约为 4.62, 比值约为 4.59%, 小于 10%的标准, 因此所建立的模型良好^[25]. 为进一步检验 HQSAR 模型的预测能力和稳定性, 采用最优模型对测试集分子的活性进行了预测, 结果如图 2 和表 2 所示.

可见实际值与预测值很接近, 误差很小. 此外经线性回归分析计算, 测试集中的化合物的实际值和预测值的相关系数 r^2_{pred} 达到 0.98, 进一步表明此模

型具有良好的稳定性及预测能力. 综上所述, 使用碎片区分参数为 B/Co/H/D, 碎片长度为 2~7 时, 得到的模型最佳. 在分子全息长度为 307 时产生最佳分子全息模型, 此时预测相关系数 $r^2_{\text{pred}}=0.98$, 交叉验证相关系数 $q^2=0.820$, SEE=0.213, 非交叉验证相关系数平方 $r^2=0.963$, 最佳主成分数为 8.

3.4 HQSAR 模型图解释

HQSAR 在计算结束时会给出每个分子结构的每个部分对其相关联的某个性质的贡献图示. 用不同的颜色来代表贡献程度的大小. 如图 3 色谱带所示, 从红色到绿色部分表示基团或原子对分子活性由提供负贡献到正贡献的变化. 红色、橙红色表示该部位为分子活性提供负贡献; 白色代表该部位对分子活性贡献中等; 黄色、绿色表示该部位原子或基团对整个分子活性提供正贡献. 由活性最差、属 a 类结构的 002 号化合物($\text{pIC}_{50}=3.301$)的色谱图可以看出, 位于咪唑环、吡啶环、苯并咪唑 C-5 位上的羧基、苯并咪

图 2 训练集和测试集样本的实验值与计算值相关关系(网络版彩图)

表 4 不同碎片大小对 HQSAR 分析结果影响

碎片大小	q^2 ^{a)}	r^2 ^{b)}	SEE ^{c)}	BL ^{d)}	NOC ^{e)}
1~4	0.712	0.908	0.334	199	7
2~7	0.820	0.963	0.213	307	8
3~8	0.786	0.959	0.224	401	7
4~5	0.771	0.919	0.315	61	7
5~5	0.815	0.923	0.305	61	7
6~7	0.790	0.954	0.237	307	7
9~10	0.759	0.923	0.303	151	5

a) q^2 是交叉验证相关系数; b) r^2 为非交叉验证相关系数; c) SEE 为非交叉验证标准误差; d) BL 为最佳分子全息长度; e) NOC 为最佳主成分数

图3 HQSAR 活性贡献图

唑 N-1 位上的 H 均分布有对活性不利的红色. 将其 N-1 位氢以烷烃或环烷烃取代后, 活性增加了 1~2 个数量级, 如 N-1 环己基取代的 a 类结构化合物 **001** 活性较 **002** 提高了两个数量级($pIC_{50}=5.367$), 说明在苯并咪唑 N-1 位引入一定体积的取代基对活性是有利的. 对化合物 **001** 进行后续修饰, 在保留 N-1 位上的环己基及苯并咪唑 C-5 位羧基结构的基础上, 在苯并咪唑 C-2 位先引入一个苯环, 继而在苯环的 2'位引入 F 原子、在 4'位以 O 原子连接长度、体积、电性不同的取代基, 得到 b~e 结构类型的化合物, 以 b 类化合物 **092** ($pIC_{50}=7.018$)、c 类化合物 **097** ($pIC_{50}=7.770$)、d 类化合物 **052** ($pIC_{50}=7.886$)和 e 类化合物 **116** ($pIC_{50}=7.921$)为例. 化合物 **097**、**052**、**116** 三者的活性均大于化合物 **092**, 比较四者的化学结构, 差别主要为化合物 **097**、**052**、**116** 三者以 O 原子连接的取代基的体积大于化合物 **092**. N-1 位上的环己基在化合物 **092** 的色码图中显示出对活性不利的橙色, 只有苯并咪唑 C-2 位的苯环及该苯环 2'位的 F 原子显示出对活性有利的黄绿色, 而在化合物 **097**、**052**、**116** 三者的色码图中可看出对活性有利的黄绿色相对于化合物 **092** 明显增多, 因此在本课题组后续的设计中, 可以借鉴拟用在 C-2 位苯环的 4'位引入较大体积的取

代基以提高化合物的活性. 从活性最好的化合物 **116** 的色码图可以看出, 在化合物 **092**、**052** 色码图中显示为白色或橙色的苯并咪唑环, 苯并咪唑 C-5 位上的羧基及 N-1 位的环己基部分在化合物 **097**、**116** 的色码图中均显示为绿色, 而在化合物 **092**、**097**、**052** 色码图中显示为白色的 4'位的 O 原子在化合物 **116** 的色码图中却显示为不利于化合物活性的橙色, 说明苯环 4'位的 O 连接子对化合物活性的贡献需进一步探讨. 结合本课题组所合成的苯并咪唑类抗 HCV 活性化合物(图 4(a))结构, 拟用电子等排体 S 原子替代 O 原子并引入酰胺键以观察其对化合物的活性影响.

3.5 基于 HQSAR 最佳模型分析结果的分子设计

基于以上 HQSAR 色码图分析, 在苯并咪唑环 N-1 位上选择环己基或链状取代基, C-2 位选择带有邻 F 取代的苯环, 将醚键的 O 原子更换成 S 原子或用酰胺替换醚键为提高化合物活性的有利因素. 据此, 对本课题组发现的 2-硫乙酰芳胺苯并咪唑类抗 HCV 活性化合物(图 4(a))进行结构改造, 设计了 11 个化合物(均为尚未见文献报道的新颖化合物), 并利用所建立的最佳 HQSAR 模型(碎片区分参数为 B/Co/H/D, 碎片长度为 2~7)预测所设计化合物的活性. 其最优公共结构如图 4(b)所示, 分子结构及预测值见表 5. 由表 5 可以看出, 基于所建模型设计的 11 个化合物都具有较优的活性预测值, 其中在苯并咪唑环 N-1 位上连接链状取代基的化合物活性要优于 N-1 环己基的化合物. 醚键的 O 连接子换为硫乙酰氨或酰胺键均得到较好的活性预测值, 且酰胺连接子的化合物活性略高于硫乙酰氨连接的化合物, 其可靠性有待实验进一步验证.

4 结论

本文采用 HQSAR 方法对系列 5-羧基苯并咪唑类

图4 硫乙酰芳胺苯并咪唑类 HCV 活性化合物(a)及最优模型通用结构(b)

表 5 基于最优 HQSAR 结果所设计分子的结构及活性预测值

化合物	R ₁	R ₂	预测值
1			8.005
2			8.129
3			8.198
4			8.002
5			7.906
6			8.072
7			8.246
8			8.150
9			8.314
10			7.550
11			7.708

抗 HCV 活性化合物的分子全息定量构效关系进行了系统的研究, 得到最优 HQSAR 模型的交叉验证相关系数 q^2 为 0.820, 非交叉验证相关系数 r^2 为 0.963, 标准偏差 SEE 为 0.213. 对由 30 个化合物构成的测试集进行了预测, 其测试集的预测相关系数 r^2_{pred} 达到 0.98,

表明该模型具有较高的预测能力和拟合能力. HQSAR 色码图能直观地将对化合物活性提供正贡献或负贡献的取代基团用不同颜色表示出来, 为 5-羧基酰胺苯并咪唑类 HCV NS5B 聚合酶抑制剂分子的设计及优化提供了一条可探讨的途径.

参考文献

- Li D, Huang Z, Zhong J. *Sci China Chem*, 2015, 2: 285–295
- Asselah T, Thompson AJ, Flisiak R, Romero-Gomez M, Messinger D, Bakalos G, Shiffman ML. *Plos One*, 2016, 11: 1–20
- 陈娜, 刘叶红, 刘小静, 陈云茹, 郭永红, 刘敏. 中华肝病杂志, 2016, 24: 40–45
- 陈开慧, 谭勇明. 免疫学杂志, 2011, 28: 635–538
- Beaulieu PL, Tsantrizos YS. *Curr Opin Invest Drugs*, 2004, 5: 838–850
- 钱汐晶, 朱勇喆, 戚中田. 中国病毒病杂志, 2013, 3: 149–154
- 聂爱华. 国际药学研究杂志, 2012, 39: 89–103
- Hashimoto H, Mizutani K, Yoshida A. Preparation of heterocyclic compounds as remedies for hepatitis C. World Patent, WO2001047883, 2001-06-05
- Beaulieu PL, Fazal G, Gillard J, Kukolj G, Austel V. Preparation of benzimidazolecarboxylates and related compounds as viral polymerase inhibitors. World Patent, WO2002004425, 2002-01-17
- Beaulieu PL, Bos M, Bousquet Y, Fazal G, Gauthier J, Gillard J, Goulet S, LaPlante S, Poupart MA, Lefebvre S, McKercher G, Pellerin C, Austel V, Kukolj G. *Bioorg Med Chem Lett*, 2004, 14: 119–124
- Hirashima S, Oka T, Ikegashira K, Noji S, Yamanaka H, Hara Y, Goto H, Mizojiri R, Niwa Y, Noguchi T, Ando I, Ikeda S, Hashimoto H. *Bioorg Med Chem Lett*, 2007, 17: 3181–3186
- Ishida T, Suzuki T, Hirashima S, Mizutani K, Yoshida A, Ando I, Ikeda S, Adachi T, Hashimoto H. *Bioorg Med Chem Lett*, 2006, 16: 1859–1863
- Hirashima S, Suzuki T, Ishida T. *Bioorg Med Chem Lett*, 2006, 49: 4721–4736
- HQSAR is the product of Tripos of Inc. St. Louis, MO, 1997
- Michael S, David B. *Quant Struc Act Relat*, 1999, 18: 245–252
- 孙礚礚, 米思奇, 游靖, 胡松青, 刘新泳. 物理化学学报, 2013, 29: 1192–1200
- 王月平, 闫婉露, 郭琼. 科学通报, 2013, 58: 916–921
- 王月平, 闫婉露, 郭琼, 何严萍. *N*-取代苯基-2-((1*H*-苯并咪唑-2-基)巯基)乙酰胺类衍生物及其用途. 中国专利, CN102675212A, 2014-11-05
- Sybyl Molecular Modeling System. Version 7.2. Tripos Inc. St. Louis, MO, 2007
- Doddareddy MR, Lee YJ, Cho YS, Choi K, Koh HY, Pae AN. *Med Chem Res*, 2004, 12: 3815–3820
- 孙礚礚, 米思奇, 游靖, 胡松青, 刘新泳. 物理化学学报, 2013, 29: 1192–1200
- Winkler DA, Burden FR. *Quant Struct Act Relat*, 1998, 17: 224–231
- 崔世海, 刘树深, 王晓栋, 王连生. 科学通报, 2003, 3: 239–242
- Lukaski HC, Bolonchuk WW. *Aviat Space Environ Med*, 1988, 12: 1163–1169
- Averv MA, Alvim GM, Rodrigues CR. *J Math Chem*, 2002, 45: 292–303

HQSAR study and molecular design of benzimidazole 5-carboxylic derivatives of hepatitis C virus NS5B polymerase

Ziheng Wang¹, Xian Chen¹, Yufang Zhang¹, Yueping Wang^{2*}, Yanping He^{1*}

¹ Key Laboratory of Medicinal Chemistry for Natural Resource, Ministry of Education, School of Chemical Science and Technology/College of Pharmacy, Yunnan University, Kunming 650091, China

² Department of Applied Chemistry, Faculty of Science, Kunming University of Science and Technology, Kunming 650050, China

*Corresponding authors (email: wangyp@kmust.edu.cn; yphe@ynu.edu.cn)

Abstract: The structure-activity relationship of 129 benzimidazole 5-carboxylic derivatives of hepatitis C virus NS5B polymerase inhibitors were investigated using hologram quantitative structure-activity relationship (HQSAR) techniques. The model parameters studied include: influences of molecular fragment, fragment distinction and molecular holographic length. In order to establish the optimal model, 99 compounds have been used to composed the training set. Results obtained from this set were analyzed by the partial least square (PLS), with the cross-validated $q^2=0.820$, convention $r^2=0.963$ and standard deviation $SEE=0.213$. The model was then further validated with a test set of 30 compounds. The predicted values were in good agreement with the experimental results ($r^2_{pred}=0.98$), showing that the model has potential ability in structural predicting and property fitting. The contributions of different atoms to activity were able to indicated with color coding figures. Based on this HQSAR model, a series of 11 benzimidazole 5-carboxylic derivatives with anti-HCV inhibition properties were designed. This model could provide additional tool in HCV NS5B polymerase inhibitor designing and optimizing.

Keywords: benzimidazole 5-carboxylic derivatives, HCV NS5B inhibitor, hologram quantitative structure-activity relationship (HQSAR), color coding, molecule design

doi: 10.1360/N032016-00191