噻唑-吡唑啉衍生物作为 EGFR 抑制剂的 3D-QSAR 和分子对接模型 研究

秦悦, 叶春林*(浙江科技学院生物与化学工程学院,浙江省农产品化学与生物加工技术重点实验室,杭州 310023)

摘要:目的 建立 EGFR 抑制剂结构和活性之间的关系模型,基于对分子活性产生影响的重要结构性因素的信息,设计 新的抑制剂分子并预测其活性,为抑制剂分子的设计提供依据。方法 使用 Discovery Studio 2019 软件进行 3D-QSAR 的 研究以及偏最小二乘的计算;利用 Autodock 进行分子对接;使用 LigPlot 研究二维相互作用。结果 模型具有较高的 q²(0.521),和 r²(r²training=0.993, r²test=0.916, r²blind=0.940),表明模型具有较高的预测能力和拟合能力。结论 预测结果表 明,新设计的化合物活性较高,为 EGFR 抑制剂分子的设计提供了参考。

关键词:抑制剂;模拟筛选;聚类分析;分子对接

中图分类号: R966 文献标志码: B 文章编号: 1007-7693(2023)21-2984-06

DOI: 10.13748/j.cnki.issn1007-7693.20223336

引用本文:秦悦,叶春林. 噻唑-吡唑啉衍生物作为 EGFR 抑制剂的 3D-QSAR 和分子对接模型研究[J]. 中国现代应用药 学, 2023, 40(21): 2984-2989.

3D-QSAR and Molecular Docking Model Study of Thiazole-pyrazoline Derivatives as EGFR Inhibitors

QIN Yue, YE Chunlin^{*}(College of Biological and Chemical Engineering, Zhejiang University of Science and Technology, Zhejiang Provincial Key Laboratory of Chemistry and Bioprocessing Technology for Agricultural Products, Hangzhou 310023, China)

ABSTRACT: OBJECTIVE To establish a model of the relationship between the structure and activity of EGFR inhibitors, and to design new inhibitor molecules and predict their activity based on the information obtained on important structural factors affecting activity, so as to provide a basis for the design of inhibitor molecules. **METHODS** Discovery Studio 2019 software was used for 3D-QSAR research and partial least squares calculation. Autodock was used for molecular docking. Study two-dimensional interactions by using LigPlot. **RESULTS** The model had a high $q^2(0.521)$, and $r^2(r^2_{\text{training}}=0.993, r^2_{\text{test}}=0.916, r^2_{\text{blind}}=0.940)$, indicated that the model had high predictive ability and fitting ability. **CONCLUSION** The prediction results show that the newly designed compounds have higher activity, providing reference for the design of EGFR inhibitor molecules. **KEYWORDS:** inhibitor; simulation screening; cluster analysis; molecular docking

表皮生长因子受体(epidermal growth factor receptor, EGFR)是一种具有酪氨酸激酶活性的膜表面受体,广泛存在于人表皮细胞和基质细胞中。在人体正常细胞中,EGFR 酪氨酸激酶(EGFR tyrosine kinase, EGFR-TK)对酪氨酸激酶活性的控制准确,若人体细胞基因点突变,会导致 EGFR-TK 活性控制持续增强,继而诱发肿瘤。EGFR 抑制剂则是一类通过抑制活化原癌基因表达产物 EGFR 发挥抗肿瘤作用的化合物,可调节 EGFR 抑制肿瘤细胞增殖、迁移和分化,抑制肿瘤血管生成^[1]。

EGFR 在细胞信号通路中的重要作用使其成为肿瘤治疗的主要靶点,其在各种人类肿瘤中可过度表达,如头颈部鳞癌、乳腺癌、肺癌、结直肠癌、前列腺癌、肾癌、胰腺癌、卵巢癌和脑癌。

而人体会对激素治疗、细胞毒剂和放射治疗进行 抵抗,导致治疗结果往往不尽如人意^[2]。因此,国 际指南建议将抗 EGFR 药物作为晚期 EGFR 突变 患者的一线治疗药物^[3]。

EGFR 抑制剂是迄今为止研究最为广泛的口服小分子抑制剂,其作用机制为通过与 ATP 竞争性结合胞外的配体结合位点,阻断分子内酪氨酸的自身磷酸化,阻断酪氨酸激酶活化,抑制 EGFR激活,从而抑制细胞周期进程、加速细胞凋亡、抑制血管生成、抑制浸润和转移^[4-6]。这类药物耐受性好、不良反应小,最为常见的不良反应为痤疮样皮疹^[7]。

QSAR 模型被广泛用作实验研究的替代方法,可以根据分子的结构预测分子的活性^[8]。在药物设

*通信作者:叶春林,男,博士,教授 E-mail: chlye2005@126. com

中国现代应用药学 2023 年 11 月第 40 卷第 21 期

基金项目:浙江省自然科学基金项目(LY17C020003)

作者简介:秦悦,女,硕士 E-mail: yueqin5587@foxmail.com

 $[\]cdot\,2984\cdot$ Chin J Mod Appl Pharm, 2023 November, Vol.40 No.21

计中广泛使用的是 3D-QSAR 建模。本研究通过 2 种用于测量静电势和立体效应的探针计算得到了 所需的能量格点,并采用该能量格点作为描述符 构建了一个偏最小二乘模型。

与分子对接结果并行的 3D-QSAR 研究可能 能够解释不同官能团在分子周围的场及其与受体 相互作用方面的贡献。本研究将对 EGFR 抑制剂 进行 3D-QSAR 和分子对接的研究,以为其分子设 计提供一些参考。

1 材料和方法

1.1 数据集的选取

EGFR 抑制剂的结构及其生物活性均取自 Lyu 等^[9]的研究工作。使用分层聚类的方法将数据集划 分为训练集和测试集。从 Alkamaly 等^[10]的研究中 获取的化合物作为盲集,进一步考察所建立的模 型的外部预测精度。这些化合物的预测结果不是 用于建立模型,而是证明所提议的模型在药物设 计中外用的可靠性。各化合物的化学结构及抑制 活性见表 1。

1.2 分子建模与分子叠合

使用 ChemDraw Professional 17.0(PerkinElmer USA)构建了 46 个分子的 3D 化学结构。将构建的 所有分子在 Discovery Studio 2019(BIOVIA Corp, San Diego, CA, USA)中进行能量最小化, Input Ligands 选择键选择对所有分子进行处理,其余参数保持默认。

通常来说分子叠合的好坏决定了模型可信度, 本研究选取了基于立体场和静电场通过场匹配的 方式来进行分子叠合。化合物 11 拥有最高的抑制 活性,因此,作为模板分子来进行叠合,其结构式 见图 1。



图 1 模板分子 Fig. 1 Template molecules

1.3 3D-QSAR 模型的建立

采用 Discovery Studio 2019 建立偏最小二乘 模型,2种用于测量静电势和立体效应的探针进行

中国现代应用药学 2023 年 11 月第 40 卷第 21 期

表1 化学结构、抑制活性及预测活性

Tab. 1 Chemical structure, inhibitory activity and predictive activity

predicti	ive activity					
序号	R_1	R_2	R ₃	EXP. ^a	3D-C	QSAR
					Pred.	Kes.
			F	R2		
			\checkmark			
			_N			
	2.4.2011	4.5	ŝ.	R ₀		0.015
4	3,4-2CH ₃	4-F	4-H	5.762	5.///	0.015
5	3,4-2CH ₃	4-CI 4 Dr	4-H 4 H	5.027	5.672	0.040
07	3,4-2CH ₃	4-DI	4-n 1 U	5.735	5.515	-0.002
2 2	3,4-2CH3	4-CH3	4-11 1/11	5.575	5.515	-0.039
0	3,4-2CH ₃	4-0CH3	4-11 4-H	5 717	5 703	_0.024
10	3 4-2CH	4-NO2	4-H	5 415	5 376	-0.039
11	3.4-2CH ₂	4-F	4-Cl	7.222	7.237	0.015
12	3.4-2CH ₃	4-C1	4-Cl	6.721	6.720	-0.001
13	3,4-2CH ₃	4-Br	4-Cl	6.432	6.369	-0.063
14^{*}	3,4-2CH ₃	$4-CH_3$	4-Cl	5.983	5.633	-0.350
15	3,4-2CH ₃	4-OCH ₃	4-Cl	6.086	6.112	0.026
16	3,4-2CH ₃	4-OH	4-Cl	6.013	6.037	0.024
17	3,4-2CH ₃	$4-NO_2$	4-Cl	6.201	6.265	0.065
18 [*]	4-Br	4-F	4-H	4.949	5.046	0.097
19	4-Br	4-Cl	4-H	5.028	4.974	-0.054
20	4-Br	4-Br	4-H	5.048	5.090	0.042
2 1 [*]	4-Br	$4-CH_3$	4-H	4.707	4.789	0.082
22	4-Br	$4-OCH_3$	4-H	4.993	5.031	0.039
23*	4-Br	4-OH	4-H	4.651	4.855	0.204
24*	4-Br	$4-NO_2$	4-H	4.804	4.840	0.036
25	4-Br	4-F	4-Cl	4.617	4.620	0.003
26	4-Br	4-C1	4-Cl	4.749	4.771	0.022
27	4-Br	4-Br	4-Cl	4.589	4.633	0.044
28 [*]	4-Br	$4-CH_3$	4-Cl	4.550	4.796	0.246
29	4-Br	$4-OCH_3$	4-Cl	4.703	4.732	0.029
30	4-Br	4-OH	4-Cl	4.667	4.707	0.040
31	4-Br	$4-NO_2$	4-Cl	4.613	5.265	0.652
32	4-OCH ₃	4-F	4-H	5.243	5.305	0.061
33	$4-OCH_3$	4-Cl	4-H	5.306	5.151	-0.155
34	4-OCH ₃	4-Br	4-H	5.255	5.226	-0.029
35	4-OCH ₃	4-CH ₃	4-H	5.126	5.114	-0.012
36	4-0CH ₃	4-0CH ₃	4-H	5.160	5.192	0.032
3/	4-0CH ₃	4-0H	4-FI 4 LI	5.272	5.027	0.041
30	4-0CH	4-NO ₂	4-n 4 Cl	5 205	5.027	-0.037
39	4-0CH	4-1 4-C1	4-C1	5 100	5 201	-0.090
40	4-0CH	4-01 4-Br	4-C1	5 200	5.436	0.092
42	4-0CH ₃	4-CH	4-C1	5 330	5 313	_0.017
43	4-0CH	4-0CH	4-C1	5 203	5 142	-0.060
43	4-0CH	4-0H	4-C1	5 116	5 125	0.000
45	4-OCH ₃	4-NO ₂	4-Cl	5.074	5.098	0.023
	2	Bia	\sim \square	~ / ^R 2		
		. —	\rightarrow	Ť		
			/	NH		
40				6.076	5 804	0 272
4a	,	2	-	0.070	5.804	-0.272
	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	CI CI		6 107	6 001	0.100
4b		2	_	6.18/	6.001	-0.186
		_	~	/R ₂		
				_		
		R1-		<b>'</b> ↓ [©]		
		~ ~				
5b	₹ North		_	6.678	5.852	-0.826
	, <b>~</b>	ž		0.070	2.902	
				H2		
Ry Ry IS						
N° Y						
	~ ~ ~	<u></u>		H₂Ń		
7c		,[]	-	6.638	5.973	-0.665
<b>F</b> 1	F~ ~	~~		7.500		
Eriotinit	) —	_	_	1.523	_	_

注: ^a-实验所得 pIC₅₀; ^b-模型预测值; ^c-误差值; ^{*}-测试集。 Note: ^a-experimental result of pIC₅₀; ^b-model predicted values; ^c-error value; ^{*}-test set.

Chin J Mod Appl Pharm, 2023 November, Vol.40 No.21 · 2985 ·

能量格点计算,将叠合后的分子集表单中添加实验 所得 pIC₅₀,随后建立 3D QSAR 模型,设置 Input Ligands 为筛选的训练集, Input test Ligands 为测试 集, Activity Property 为实验测定的 pIC₅₀,其余保 持默认数值。盲集作为外部测试集使用相同操作。

1.4 分子对接研究

从 RCSB PDB 数据库(https://www.rcsb.org/ pdb)中检索并获取受体大分子,晶体结构(PDB code: 1M17),在进行对接之前,使用 AutoDockTool-1.5.7^[11]软件删除晶体结构中含有的 水及配体小分子,然后将极性氢和 Gasteiger 电荷 添加到蛋白质结构中。

使用 Open Babel-3.1.1^[12]将配体小分子(化合物 **11**)的 mol2 文件格式转换为 pdb 格式,利用 AutoDockTool-1.5.7 软件添加极性氢和 Gasteiger 电荷。

#### 2 结果

2.1 数据集的聚类分析

层次聚类是解决聚类问题使用最广泛的方法 之一。使用层次聚类将数据集划分为训练集和测 试集,目的是在化学结构的多样性和生物活性之 间取得相对公平,以便获得具有代表性的测试集 化合物。基于层次聚类得到的信息,将数据集分 为含有 34 个化合物的训练集和含有 8 个化合物的 测试集,见图 2。



Fig. 2 Results of cluster analysis

### 2.2 3D-QSAR 模型的统计结果

基于含有 35 个化合物的训练集的模型利用偏 最小二乘回归分析,将化学成分进行关联建模。 五重交叉验证的结果见表 2,其中主成分值为 5 时 q²最大,均方根误差值为 0.424,平均绝对误差值 为 0.310,均为最小值。

· 2986 · Chin J Mod Appl Pharm, 2023 November, Vol.40 No.21

表 2 训练集 3D-QSAR 模型的五重交叉验证结果

Tab. 2Five-fold cross validation results of the training set3D-QSAR model

主成分值	交叉验证相关系 数(q ² )	均方根误差	平均绝对误差
1	0.407	0.467	0.364
2	0.470	0.443	0.339
3	0.503	0.431	0.321
4	0.520	0.423	0.311
5	0.521	0.424	0.310
6	0.511	0.428	0.313

选取主成分值为 5 时所建立的模型与训练集 之间进行比对,结果显示,模型的拟合能力较强, 基本在趋势线周围均匀分布,训练集 r²=0.993,测 试集 r²=0.916 均表明模型可靠,具有一定的预测 能力。结果见图 3。



图 3 训练集的预测值与实际值拟合曲线(A)和测试集的 预测值与实际值的拟合曲线(B)

Fig. 3 Predicted value of the training set fits the actual value to the curve(A) and curve of the test set's prediction and actual values(B)

#### 2.3 盲集测试结果分析

盲集是一种外部预测能力的度量,从 Alkamaly等^[10]研究中选取了4个对EGFR有抑制 作用的化合物。这些化合物的预测结果可以为建 议的外用模型在药物设计中的可靠性提供合理的 证据。建立的3D-QSAR模型与外部测试盲集的统 计结果见表1。这组数据的实验和预测 pIC₅₀值均 可以在表1中找到。 同时进行五重交叉验证,当主成分值为 4 时 q²最大,均方根误差值为 0.400,平均绝对误差值 为 0.286,均为最小值,结果见表 3。生成的拟合 曲线 r²=0.940,盲集拟合曲线表示所建立的模型可 靠,具有一定的预测能力,见图 4。

表3 盲集 3D-QSAR 模型的五重交叉验证结果

**Tab. 3** Five-fold cross validation results of blind set 3D-OSAR model



**Fig. 4** Fitted curve of the predicted and actual values of the blind set

#### 2.4 力场结果分析

为了更直观地说明静电场和立体场对于化合物的活性贡献,展示了模型静电场等势图和立体场等势图,见图 5。在等势图中定义积极影响区域系数为 80%,消极影响区域系数为 20%。以静电场 50%,立体场 50%进行模板的叠合。其中图 5A中红色和蓝色区域代表着静电场对化合物活性的影响,在蓝色区域内引入正电荷基团可提高化合物活性,红色区域则相反。图 5B中黄色和绿色区域代表着立体场对于化合物活性的影响,绿色区域中若引入大基团取代基则会使化合物活性增高,黄色区域则提示引入小体积基团对化合物活性有积极影响。

从图 5B 立体场的等势图中可以看出 R₁的 3 位附近集中有绿色区域,表示体积较大的基团有 利于提高活性,但进行建模的数据集 R₁的 3 位 并未有除了-CH₃的取代基。R₁的 4 位的取代基

中国现代应用药学 2023 年 11 月第 40 卷第 21 期

分别为-OCH₃、-Br、-CH₃,取代基由-Br取代变 为-OCH₃取代时,活性顺序为 39>25,40>26, 41>27,42>28,43>29,44>30,45>31。所以在 R₁的4位进行较大体积取代基的替换可能会有 利于化合物的活性值的增加。大面积红色区域包 裹R₂取代基,表示在该位增加电负性取代基有 利于增加活性。





图 5 静电场等势图(A)和立体场等势图(B)

A-红色和蓝色区域分别表示在该区域引入吸电子基能够增强或减弱 活性;B-绿色和黄色区域分别表示在该区域引入大体积基团能够增强 或减弱化合物活性。

**Fig. 5** Electrostatic field equipotential plot(A) and stereoscopic field equipotential plot(B)

A-Red and blue regions indicated that the introduction of an electron-absorbing group in this region could enhance or attenuate activity, respectively. B-green and yellow regions indicated that the introduction of large volumes of groups in this area could enhance or attenuate compound activity, respectively.

但是在模型中需要综合 4 种颜色的区域范围 确定引入基团,进行综合考虑。

#### 2.5 对接结果分析

选取训练集中活性最高的化合物 11 与 EGFR 晶体结构(pdb code:1M17)进行对接模拟,以探究 抑制剂分子与蛋白可能的结合模式。对接产生 10 个对接结果,利用 Autodock 进行对接的打分结果 见表 4。通常最佳结合能为数据最小值。化合物 11 与酶活性口袋 Asp985 和 Val987 残基之间形成 2 个氢键,且观察到与 Met983、Tyr992、Asp988、 Val986 有关的疏水相互作用,R₁ 的 2 位周围的 Asp988 和 R₂ 的 3,4 位周围的 Met983 表明引入疏 水基团会提高活性。使用 LigPlot^[13]软件简化了相 互作用的结果,见图 6。

Chin J Mod Appl Pharm, 2023 November, Vol.40 No.21  $\cdot$  2987  $\cdot$ 

表 4	灭	接打分结果	
Tah	4	Results of docking	scoring

Tab. 4 Result	<b>b. 4</b> Results of docking scoring			
对接编号	结合能	分子间能量	内能	
1	-27.41	-32.38	-5.69	
2	-26.36	-31.34	-5.27	
3	-26.15	-31.13	-5.82	
4	-29.16	-30.00	-5.82	
5	-24.10	-29.08	-5.44	
6	-25.90	-30.88	-5.02	
7	-25.44	-30.42	-5.86	
8	-23.47	-28.49	-5.40	
9	-22.93	-27.91	-5.69	
10	-19.54	-24.52	-5.61	



**图 6** 化合物 **11** 与 EGFR 激酶(PDB code: 1M17)的二维相 互作用

Fig. 6 Two-dimensional interaction of compound 11 with EGFR kinase(PDB code: 1M17)

## 2.6 设计新的化合物

基于 3D-QSAR 和分子对接的研究,获得了 影响抑制剂活性的结构特性,模板分子为化合物 11(图 1),在 R₁的 3 位引入一个新基团,结果表 明预测活性较好,见表 5。新引入的取代基增加 了化合物与 EGFR 激酶之间的静电相互作用。与 化合物 11 相比,结构修饰后设计的分子预测活 性有所提高。此外,还选择了新设计的小分子与 受体蛋白对接,研究了其与酶的结合方法,见图 7。结果表明,新的小分子与受体蛋白所形成的 氢键作用更强,但实际的抑制活性仍需通过实 验验证。

表5 新抑制剂分子的预测活性值

 Tab. 5
 Predicted activity values of new inhibitor molecules



**图 7** 化合物 N1 与 EGFR 激酶(PDB code: 1M17)的二维相 互作用

^{· 2988 ·} Chin J Mod Appl Pharm, 2023 November, Vol.40 No.21

Fig. 7 Two-dimensional interaction of compound N1 with EGFR kinase(PDB code: 1M17)

中国现代应用药学 2023 年 11 月第 40 卷第 21 期

#### 3 讨论

本研究建立了一系列 EGFR 抑制剂的 3D-QSAR 模型。用偏最小二乘法研究了生物学数据与模型之间的相关性。统计结果证实了所建模型的预测性和可靠性。对建立的模型进行了分析,以显示与分子图中每个基团的位置相关的影响,从而为设计新的 EGFR 抑制剂提供了更好的思路。之后进行了分子对接,发现目标受体引入疏水基团会提高活性。根据得到的数据,设计的新化合物的预测结果表明它们都具有较高的活性。本研究为 EGFR 激酶抑制剂分子的设计提供了一定的参考。

#### REFERENCES

- ZHENG C Y, GAO C Y, SHI E, et al. Drug research progress of EGFR20 exon 20 insertion mutation in non-small cell lung cancer[J]. Chin J Clin Ration Drug Use(临床合理用药杂志), 2021, 14(10): 178-181.
- [2] REN X L, WANG X, LI Z Y, et al. Research progresses of epidermal growth factor receptor tyrosine kinase inhibitors[J]. Chin New J Drugs(中国新药杂志), 2005, 14(7): 821-826.
- [3] HUANG L, HU X X, HONG Y. Research on the third-generation epidermal growth factor receptor trysine kinase inhibitor drugs and their patents[J]. Chin J New Drugs(中国新药杂志), 2022, 31(16): 1553-1559.
- [4] ZHANG X T, LI L Y. Clinical progress of epidermal growth factor receptor tyrosine kinase inhibitor in anti-tumor therapy[J]. Chin J Lung Cancer(中国肺癌杂志), 2002, 5(4): 310-312.
- [5] SUN L, MCMAHON G. Inhibition of tumor angiogenesis by synthetic receptor tyrosine kinase inhibitors[J]. Drug Discov

Today, 2000, 5(8): 344-353.

- [6] LI Y, TIAN C, WANG L, et al. Research progress of small-molecule epidermal growth factor receptor-tyrosine kinase inhibitors[J]. Anti Tumor Pharm(肿瘤药学), 2016, 6(2): 81-88.
- [7] BANERJEE R, RACHID Z, MCNAMEE J, et al. Synthesis of a prodrug designed to release multiple inhibitors of the epidermal growth factor receptor tyrosine kinase and an alkylating agent: A novel tumor targeting concept[J]. J Med Chem, 2003, 46(25): 5546-5551.
- [8] POURBASHEER E, AALIZADEH R, GANJALI M R, et al. QSAR study of Nav1.7 antagonists by multiple linear regression method based on genetic algorithm (GA-MLR)[J]. Med Chem Res, 2014, 23(5): 2264-2276.
- [9] LYU P C, LI D D, LI Q S, et al. Synthesis, molecular docking and evaluation of thiazolyl-pyrazoline derivatives as EGFR TK inhibitors and potential anticancer agents[J]. Bioorg Med Chem Lett, 2011, 21(18): 5374-5377.
- [10] ALKAMALY O M, ALTWAIJRY N, SABOUR R, et al. Dual EGFR/VEGFR2 inhibitors and apoptosis inducers: Synthesis and antitumor activity of novel pyrazoline derivatives[J]. Arch Pharm (Weinheim), 2021, 354(4): e2000351.
- [11] GOODSELL D S, OLSON A J. Automated docking of substrates to proteins by simulated annealing[J]. Proteins, 1990, 8(3): 195-202.
- [12] O'BOYLE N M, BANCK M, JAMES C A, et al. Open Babel: An open chemical toolbox[J]. J Cheminform, 2011(3): 33.
- [13] WALLACE A C, LASKOWSKI R A, THORNTON J M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions[J]. Protein Eng, 1995, 8(2): 127-134.

收稿日期: 2022-09-25 (本文责编:沈倩)