

乙酰羟酸合成酶突变体对除草剂阔草清抗性的预测研究

王白帆 何寅武 文欣 牛聪伟* 席真*

(南开大学元素有机化学国家重点实验室 化学生物学系 天津 300071)

摘要 本工作测定了商品化除草剂阔草清对大肠杆菌乙酰羟酸合成酶 II 及其突变体的抑制常数。利用生物大分子的定量构效关系方法(MB-QSAR)建立了定量、可视的突变体性质与结构关系模型。实验数据与计算数据之间显示了良好的相关性: MB-QSAR/CoMFA (comparative molecular field analysis)模型的交叉验证系数 $q^2=0.691$, 非交叉验证相关性系数 $r^2=0.947$, 外部预测能力 $r^2_{\text{pred}}=0.759$; MB-QSAR/CoMSIA (comparative molecular similarity indices analysis)模型的相应值分别为 0.625, 0.960 和 0.619。模型的三维结构-性质关系图为突变体抗性提供了直观阐释, 为抗性规避性农药的设计提供了有价值的信息。同时, 该模型也在一定程度上显示了除草剂种属选择性的预测能力。

关键词 乙酰羟酸合成酶; 阔草清; 除草剂抗性; 抗性预测; MB-QSAR (mutation-dependent biomacromolecular quantitative structure-activity relationship)

Prediction on the Resistance of Acetohydroxyacid Synthase Mutants to Herbicide Flumetsulam

Wang, Baifan He, Yinwu Wen, Xin Niu, Congwei* Xi, Zhen*

(State Key Laboratory of Elemento-Organic Chemistry, Department of Chemical Biology, Nankai University, Tianjin 300071, China)

Abstract Flumetsulam is a widely used herbicide that targets the acetohydroxyacid synthase (AHAS). Mutations in AHAS have caused serious herbicide resistance which threatened the field application of this herbicide. We have recently established a mutation-dependent biomacromolecular quantitative structure-activity relationship method, called MB-QSAR, which could be used to quantitatively predict the mutational drug resistance in molecular level and elucidating the three dimensional structure-resistance relationships for the design of resistance evading inhibitors. In this work, we employ MB-QSAR method to predict the molecular drug resistance of AHAS mutants towards flumetsulam, and to depict the structure resistance relationships in AHAS mutants. A series of AHAS mutants concerned with the herbicide resistance were constructed, and the inhibitory properties of flumetsulam against these mutants were measured. Then the structures of these mutants were constructed, optimized and aligned for the subsequent MB-QSAR modelling. The CoMFA (comparative molecular field analysis) and CoMSIA (comparative molecular similarity indices analysis) molecular field values (steric, electrostatic, hydrophobic, hydrogen-bond donor and hydrogen-bond acceptor descriptors) were calculated in the flumetsulam binding pocket for these mutants. The CoMFA and CoMSIA molecular field values were used as independent variables, while the pK_i values for each mutants were used as dependent variables in the partial least squares (PLS) regression analyses to derive the MB-QSAR models. The built MB-QSAR model showed excellent correlation between experimental and computational data (MB-QSAR/CoMFA model: $q^2=0.691$, $r^2=0.947$, $r^2_{\text{pred}}=0.759$; MB-QSAR/CoMSIA model: $q^2=0.625$, $r^2=0.960$, $r^2_{\text{pred}}=0.619$), indicating the good prediction for the inhibition properties of flumetsulam against AHAS mutants. The comparison of the molecular interaction diagrams from MB-QSAR models provides information about which positions in the polypeptide chain could have a higher propensity to acquire herbicide resistant mutations, which in turn provides guidelines for modifying the existing herbicide as well as for designing new resistance-evading herbicides. The obtained MB-QSAR model also showed reasonable predictive power toward AHAS from various species, indicating the potential application of our MB-QSAR method in the prediction of the selectivity of drugs towards targets from different species.

Keywords acetohydroxyacid synthase; flumetsulam; herbicide resistance; resistance prediction; MB-QSAR (mutation-dependent biomacromolecular quantitative structure-activity relationship)

1 引言

农药抗性及其种属选择性关系到农药的使用效力、

使用周期及物种安全, 而成为农药设计开发过程中必须关注的两个重要问题。研究农药与靶标及其突变体的作用机制, 建立抗性/选择性的预测方法, 是设计兼具高效

* E-mail: niucongwei@nankai.edu.cn; zhenxi@nankai.edu.cn; Tel.: 022-23504782; Fax: 022-23504782

Received November 20, 2021; published December 23, 2021.

Supporting information for this article is available free of charge via the Internet at <http://sioc-journal.cn>.

Project supported by the National Natural Science Foundation of China (Nos. 21837001, 21740002).

项目受国家自然科学基金(Nos. 21837001, 21740002)资助。

性、选择性及抗性规避性农药的可行途径。

当前已报道的利用计算机模拟方法研究药物分子抗性和选择性的方法,可粗略分为基于能量计算方法和统计回归方法两大类。代表性的工作如: Hao 等^[1]发展出计算突变扫描方法来预测 HIV (human immunodeficiency virus)蛋白酶对药物的抗性; Cao 等^[2-3]通过计算小分子与靶标蛋白的结合势能变化,对 HIV 蛋白酶突变抗性进行了预测研究; Hauser 等^[4]利用结合自由能的变化对 Abl 蛋白激酶的突变抗性进行了预测; Goodford 等利用探针原子计算蛋白结合空腔中的分子场,对其进行主成分分析,发展了 GRID 及 GRID/PCA 方法^[5-6]; COMBINE^[7-8]及 SIFT^[9]等方法则以小分子与靶标蛋白复合物结构为基础,计算两者之间的相互作用能量,将能量分解到各个重要位置的氨基酸残基上,以分解能量值为自变量,与小分子生物活性进行统计回归,得到的模型用于预测小分子选择性。

然而,现有的预测方法的突出问题是缺乏直观阐释抗性并为反抗性药物设计提供结构-活性关系指导的功能。在前期工作中,我们将小分子 3D-QSAR (three dimensional quantitative structure-activity relationship)研究中广泛使用的比较分子场分析(Comparative Molecular Field Analysis, CoMFA)^[10]方法移植到生物大分子突变酶系的定量构效关系研究中,建立了 MB-QSAR 方法 (Mutation-dependent biomacromolecular QSAR, 图 1),

并将这一方法应用于除草剂重要靶标乙酰羟酸合成酶 (Acetohydroxyacid Synthases, AHAS, EC 2.2.1.6)及突变体对除草剂氯嘧磺隆、氯磺隆和双草醚的抗性研究^[11-13]。我们所建立的 MB-QSAR 模型均表现出较强的抗性预测能力,同时也可直观地为反抗性小分子设计提供指导。这一方法的另一重要应用是除草剂对源于不同种属靶标的选择性预测。本质上,不同种属酶可以看作是酶进化中的不同突变体。药物对靶标的选择性和靶标对药物的抗性的共同本质是药物与靶标的相互作用,并由此构成了评价药物选择性和抗性的共同标准。因此 MB-QSAR 方法不仅是靶标抗性预测方法,也是药物选择性的预测方法。

AHAS 是微生物和植物中支链氨基酸(缬氨酸、亮氨酸和异亮氨酸)合成途径中的关键性酶^[14-18]。AHAS 催化两分子的丙酮酸缩合生成 2-乙酰乳酸,或催化丙酮酸与 2-酮丁酸形成 2-乙酰-2-羟基丁酸。在植物体内,如果支链氨基酸的合成受到阻碍,会使得蛋白质合成的停止^[19],从而导致植物生长受损直至死亡。由于哺乳动物中并没有 AHAS,因而 AHAS 成为了当前最重要的绿色除草剂作用靶标之一。以 AHAS 为靶标的商品化除草剂包括磺酰脲类、咪唑啉酮类、嘧啶水杨酸类和三唑嘧啶磺酰胺类等类型,58 个品种。然而,由于除草剂的广泛使用,造成大田抗性严重。据统计,截至到 2021 年 11 月 8 日,全世界已有 509 种对除草剂具有抗性的杂草,

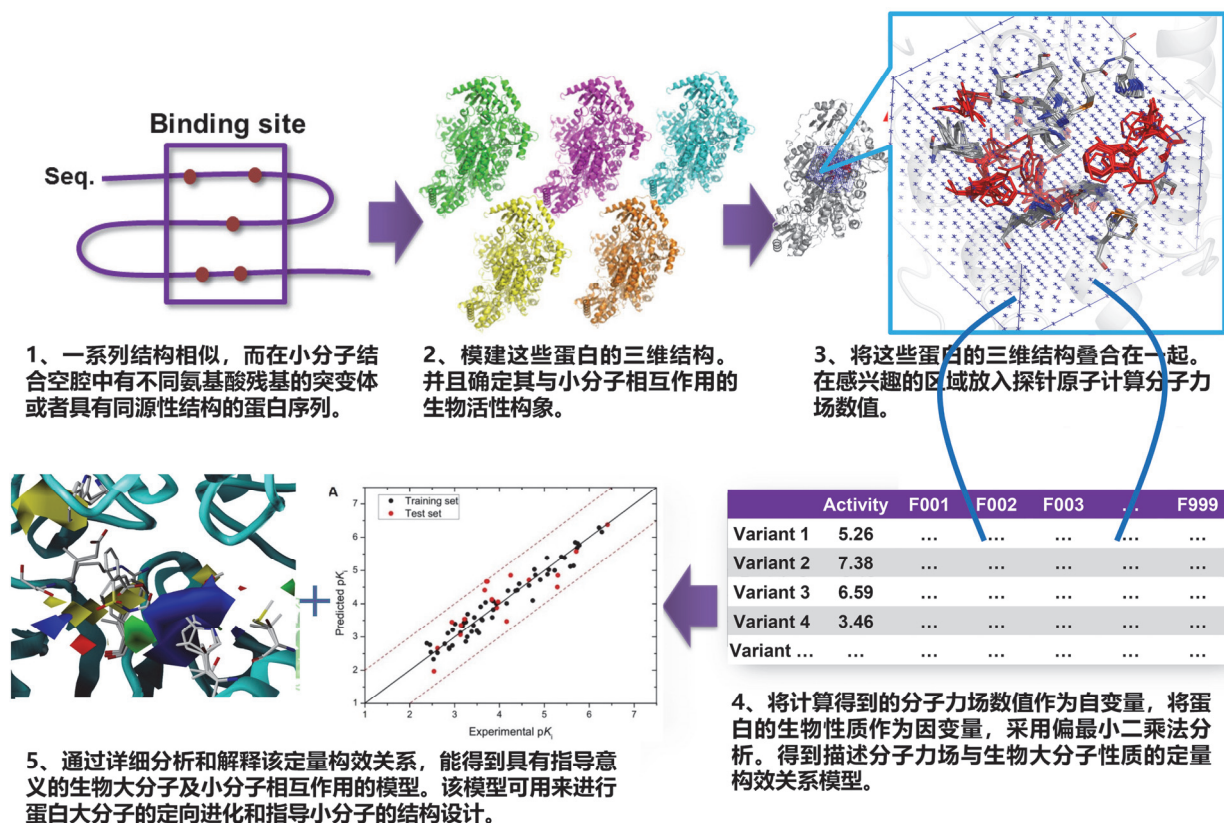
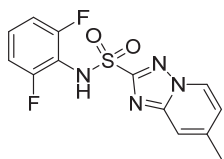


图 1 MB-QSAR 方法流程图

Figure 1 Flow chart of MB-QSAR method

其中对 AHAS 除草剂有抗性的杂草有 169 种 (<http://www.weedscience.org/>). 大部分 AHAS 除草剂抗性所依赖的分子基础源于 AHAS 几个位点, 即 A122、P197、A205、W574、S653、G654(以拟南芥 AHAS 氨基酸序列为准)等位点的氨基酸残基的突变^[20-25].

前期工作中, 我们在大肠杆菌 AHAS II (*E. coli* AHAS II) 的等位位点(A26、V99、S100、A108、M460、W464)构建了一系列突变体^[11-12,26-27]. 在本文中, 我们考察了三唑嘧啶磺酰胺类除草剂阔草清(flumetsulam, 图式 1)对 AHAS 突变体的抑制作用(抑制常数 K_i), 利用 MB-QSAR 方法建立了 AHAS 对阔草清抗性的预测模型, 并就除草剂对不同种属 AHAS 酶的选择性进行了预测.



图式 1 阔草清的分子结构

Scheme 1 The chemical structure of flumetsulam

2 结果与讨论

2.1 野生型 *E. coli* AHAS II 及其突变体与阔草清的相互作用

野生型 *E. coli* AHAS II 及其突变体与阔草清相互作用的 pK_i 值以及抗性倍数 RF 值列于表 S1(见支持信息). 图 2 直观显示了六个突变位点不同突变体的抗性差异.

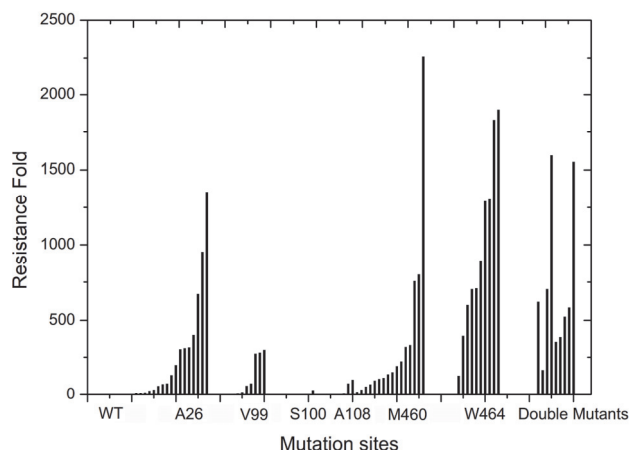


图 2 大肠杆菌 AHAS II 突变体对于阔草清抗性倍数的分布

Figure 2 The distribution of the resistance fold values of *E. coli* AHAS II mutants against flumetsulam

由表 S1 中数据及图 2 所示, 突变后抗性倍数最大的突变体为 M460S, 其抗性倍数 RF 值为 2265. 突变影响最明显的位点为 464 位点, 将该位点的色氨酸突变为其它氨基酸残基后, 普遍会对阔草清产生较大程度的分子抗性. 影响其次的为 26 位点和 460 位点. 将 26 位点的

丙氨酸突变为带电性残基如天冬氨酸后, 会对阔草清产生严重的分子抗性. 99 位点的带电性残基的突变体 V99R、V99D 以及将该位点的缬氨酸突变为色氨酸也会引起一定程度的分子抗性, 而该位点的其他氨基酸残基突变体对阔草清的影响不大. 在 108 位点, 丙氨酸突变为带电性残基如赖氨酸(K)、谷氨酸(E)后会引起对阔草清一定程度的分子抗性, 该位点的其它氨基酸残基突变体对阔草清的影响不大. 100 位点带负电荷的突变体 S100E 会对阔草清有影响, 其它类型氨基酸残基突变体的影响则较小.

2.2 阔草清在野生型 *E. coli* AHAS II 催化亚基中的结合模式

我们通过分子对接, 确定了阔草清与野生型 *E. coli* AHAS II 的作用模式. 如图 3 所示, 阔草清的杂环部分与 464 位点的色氨酸吡啶环部分有堆积作用, 其磺酰胺部分与 276 位点的精氨酸及 159 位点的赖氨酸有极性相互作用. 这一结合模式与磺酰胺类除草剂分子如氯嘧磺隆和氯磺隆等与 *E. coli* AHAS II 的结合模式非常类似(见支持信息, 图 S1)^[11-12]. 通过比较可以发现, 阔草清的杂环部分与磺酰胺分子的杂环部分重合, 两类分子都与 159 位点的赖氨酸和 276 位点的精氨酸有着强的极性相互作用.

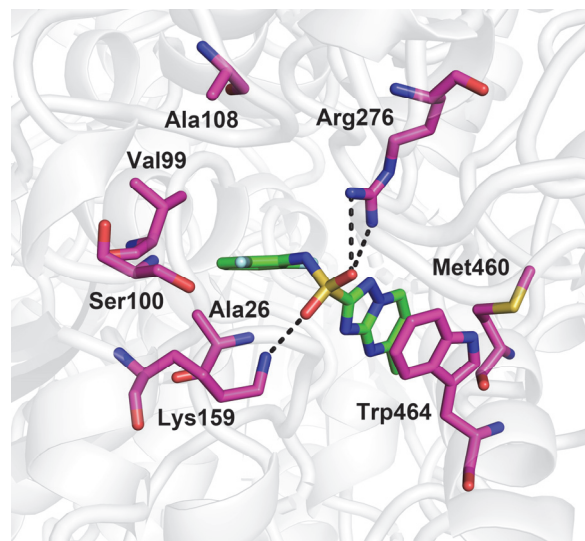


图 3 阔草清分子与大肠杆菌 AHAS II 的结合模式. 其中抑制剂分子用绿色表示, AHAS 残基用洋红色表示

Figure 3 The binding mode of flumetsulam in *E. coli* AHAS II. Flumetsulam is showing in green and the residues of AHAS are showing in magenta

2.3 MB-QSAR 模型

我们以计算获得的分子场数值(CoMFA 值)或分子场相似性指数(CoMSIA 值)作为自变量, 以阔草清抑制常数 K_i 的负对数为因变量, 采用偏最小二乘法进行了统计回归分析. 考虑到研究体系的叠合方式对 CoMFA 或 CoMSIA 模型的影响, 本文以蛋白结构的三种叠合方式

计算分子场数值(见实验部分), 突显生物活性差异的结构特征, 避开可能存在的干扰模型的无关差异. 通过相应的回归分析, 一些干扰模型相关性的突变体数据被排除, 包括 A26D、A26S、A26W、M460K、S100W、V99E、V99H、W464K 和 A26T-W464F. 最终以野生型及 58 个突变体构成训练集, 19 个突变体为测试集, 建立了 MB-QSAR/CoMFA 及 MB-QSAR/CoMSIA 模型, 回归分析结果列于表 1 及图 4.

如表 1 所示, 在叠合结合空腔中残基骨架原子的方式下(A-II), 均得到较好的回归模型. 最高的交叉验证系数 $q^2(0.691)$ 是在叠合空腔中抑制剂周围 5 Å 距离内残基骨架原子时得到(A-II-5), 该模型同样具有较好的非交叉验证相关性系数 $r^2(0.947)$ 及 F 检验值(190.608), 且该模型也具有最好的外部预测能力, 即最高的 r^2_{pred} 值(0.759). 因此该模型被作为最终 MB-QSAR/CoMFA 模型, 用来进行后续的突变体结构与抗性关系分析.

表 1 MB-QSAR/CoMFA 模型结果

Table 1 Summary of the MB-QSAR/CoMFA results

	MB-QSAR/CoMFA					
	A-I	A-II				A-III
		A-II-3	A-II-4	A-II-5	A-II-6	
ONC ^a	5	5	5	5	5	5
$q^2(\text{LOO})^b$	0.580	0.666	0.673	0.691	0.684	0.372
SEE ^c	0.315	0.261	0.256	0.259	0.264	0.401
r^{2d}	0.922	0.947	0.948	0.947	0.945	0.874
$F\text{-value}^e$	125.318	187.64	195.036	190.608	182.745	73.208
	(5, 53)	(5, 53)	(5, 53)	(5, 53)	(5, 53)	(5, 53)
$r^2_{\text{pred}}^f$	0.725	0.746	0.747	0.759	0.754	0.621
Contributions ^g						
S	0.657	0.709	0.706	0.695	0.701	0.809
E	0.343	0.291	0.294	0.305	0.299	0.191

^a 最佳主成分数; ^b 在交叉验证中得到的相关性系数; ^c 在非交叉验证中得到的标准差; ^d 在非交叉验证中得到的相关性系数; ^e F -test 检验; ^f 通过预测测试集中的突变体得到的相关性系数; ^g 各分子场在模型中所占的比重, S 代表立体场, E 代表静电场.

我们在构建 MB-QSAR/CoMSIA 模型时, 首先是对蛋白结构不同叠合情况下 CoMSIA 场组合方式进行了评价. 对于由五种基本 CoMSIA 场构成的全部 31 种场组合方式, 计算了其交叉验证相关性系数 q^2 (见支持信息, 表 S2). 其中在 A-I 叠合方式下, 立体场(S), 疏水场(H)及氢键受体场(A)的组合, 具有较高 q^2 值及较好的预测能力. 该模型被用作最终的 MB-QSAR/CoMSIA 模型进行后续的突变体结构与抗性关系分析. 其 MB-QSAR/CoMSIA 模型的相关参数列于表 2.

表 S1 中数据及图 4 表明, 在 MB-QSAR/CoMFA 模型的训练集和 MB-QSAR/CoMSIA 模型的训练集中, 所有突变体的预测误差都不超过一个对数单位. 这个结果说明建立的 MB-QSAR 模型具有很好的内部预测能力. 对 19 个突变体构成的测试集中, MB-QSAR/CoMSIA 模

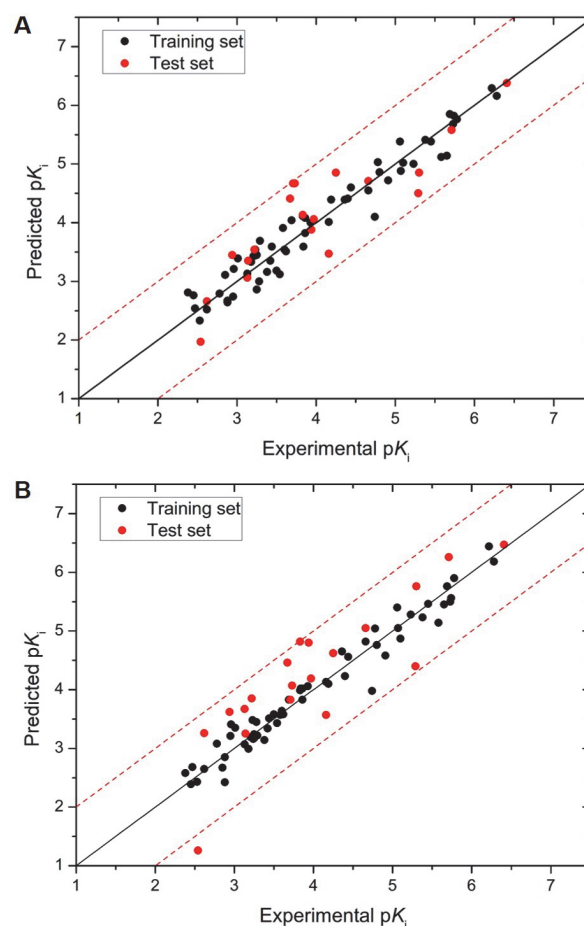


图 4 由 MB-QSAR 模型中得到 AHAS 突变体针对阔草清的 pK_i 值的预测值与实验值的相关图. (A) MB-QSAR/CoMFA 模型; (B) MB-QSAR/CoMSIA 模型

Figure 4 Plots of the experimental and predicted pK_i values for AHAS mutants against flumetsulam from MB-QSAR/CoMFA (A) and MB-QSAR/CoMSIA (B) models

型对 V99M-W464Q 的预测误差为 1.28. 除去该突变体外, MB-QSAR 模型对其余所有突变体的预测误差都小于一个对数单位, 说明模型同样具有很好的外部预测能力.

2.4 构效关系

由模型 MB-QSAR/CoMFA 及模型 MB-QSAR/CoMSIA 得到的 *E. coli* AHAS II 结构与阔草清抗性的三维构效关系如图 5 所示.

在立体效应图中(图 5A, 5C), 绿色区块表示增强该区域的立体相互作用有利于增强阔草清的抑制作用, 而黄色区域表示增强该区域的立体相互作用对阔草清的抑制作用是不利的, 在该区域引入强的立体相互作用可能引起一定程度的分子抗性. 如图 5 所示, 在 99 位点及 100 位点与阔草清之间有大块的黄色区块出现, 表示在该位点发生的大体积的氨基酸残基突变对阔草清的抑制作用是不利的, 如突变体 V99W 的抗性倍数 RF 值为 302. 在 MB-QSAR/CoMSIA 模型的立体效应图中, 黄

表2 MB-QSAR/CoMSIA 模型在各种叠合方式下的结果

Table 2 Statistics of MB-QSAR/CoMSIA with different alignment rules

力场组合	SHA	SA	SH	SEA	SEHA	SEHDA	SDA	EHA	S
ONC	6	6	5	6	6	6	5	6	4
$q^2(\text{LOO})^a$	0.625	0.611	0.565	0.594	0.590	0.552	0.532	0.547	0.536
SEE	0.228	0.249	0.329	0.241	0.237	0.232	0.317	0.288	0.431
r^2	0.960	0.952	0.915	0.955	0.957	0.959	0.921	0.936	0.851
F -value	208.110	172.876	113.771	184.961	192.019	200.976	123.593	127.218	77.340
r^2_{pred}	0.619	0.536	0.468	0.608	0.581	0.417	0.361	0.556	0.708
Contributions ^b									
S	0.270	0.340	0.571	0.204	0.176	0.118	0.177		1.000
E				0.364	0.319	0.217		0.397	
H	0.178		0.429		0.119	0.084		0.159	
D						0.321	0.444		
A	0.552	0.660		0.432	0.386	0.259	0.379	0.444	

^a 交叉验证相关性系数, 采用的交叉验证方法为 SAMPLS; ^b 各个分子场所占的比率, S 为立体场, E 为静电场, H 为疏水场, D 为氢键给体场, A 为氢键受体场.

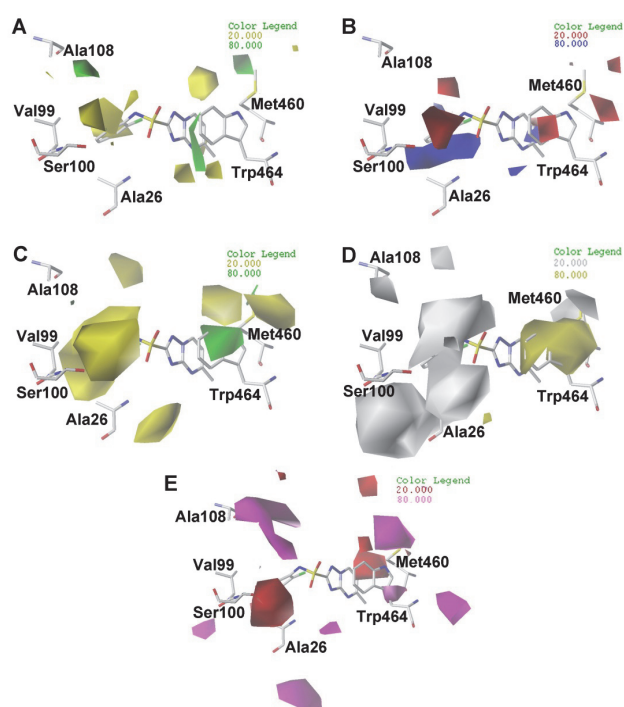


图5 MB-QSAR/CoMFA 模型的立体效应图(A)、静电效应图(B)和 MB-QSAR/CoMSIA 模型的立体效应图(C)、疏水效应图(D)和氢键受体效应图(E)

Figure 5 The MB-QSAR/CoMFA contour maps of steric (A) and electrostatic (B) effects and the MB-QSAR/CoMSIA contour maps of steric (C), hydrophobic (D) and hydrogen bond acceptor (E) effects (alignment rule A-I) were shown

色区域也出现在 26 位点的丙氨酸和阔草清之间, 表明在该位点发生的大体积的氨基酸残基突变对阔草清的抑制作用也是不利的, 如突变体 A26W 的抗性倍数 RF 值为 957, 突变体 A26F 的抗性倍数 RF 值为 402. 从结构上看, 将 Val99 和 Ala26 突变成侧链具有较大体积的色氨酸和苯丙氨酸残基, 其侧链会和阔草清分子的苯环和杂环取代基发生冲突, 从而影响了阔草清与 AHAS 的结合. 在 464 位点的色氨酸周围, 与之前研究的磺酰脲类除草剂分子类似^[11-12], 有绿色区块的出现, 说明该位点的立体相互作用对阔草清的抑制作用是有利的, 当发生

突变变为较小体积的氨基酸残基会引起针对阔草清严重的分子抗性, 如突变体 W464I 的抗性倍数 RF 值为 1297. 从我们的分子对接模型中可以看出, Trp464 的吡咯环与阔草清的杂环存在芳香堆积作用, 将其突变成异亮氨酸以及其他类型的氨基酸则无法与阔草清的杂环形成堆积作用, 且其侧链有可能与阔草清的杂环发生空间上的冲突, 进而引起阔草清与 AHAS 的结合下降. 与此同时, 将 Trp464 的侧链突变成同样带有芳环的苯丙氨酸(W464F)和酪氨酸(W464Y), 能够在一定程度上维持与阔草清杂环部分的芳香堆积作用, 故而能够在很大程度上维持阔草清与 AHAS 的结合, 而不引起抗性.

在静电效应图中(图 5B), 蓝色区块出现在 100 位点丝氨酸及 26 位点丙氨酸与阔草清之间, 蓝色区块表示在该区域引入带负电性基团的突变则可能引起一定程度的分子抗性. 如突变体 S100E 的抗性倍数 RF 值为 35, 突变体 A26D 的抗性倍数 RF 值为 1358. 从阔草清与 AHAS 结合的模型上看, 阔草清的磺胺部分与 Lys159 和 Arg276 形成了氢键相互作用. 将 Ser100 和 Ala26 分别突变成侧链带有负电的谷氨酸和天冬氨酸残基, 有可能与侧链带有正电的 Lys159 和 Arg276 形成静电作用, 从而破坏了这两个残基与阔草清之间氢键作用, 削弱了阔草清与 AHAS 的结合. 在 460 位点的甲硫氨酸周围, 散落着红色区块. 红色区块表示在该区域引入带正电性基团的突变可能引起一定程度的分子抗性, 如突变体 M460K 的抗性倍数 RF 值为 805.

在疏水效应图中(图 5D), 在抑制剂结合空腔外侧, 即 99、100 及 26 位点的残基周围, 有大片的灰色区块出现, 灰色区块表示该区域发生疏水性氨基酸残基的突变可能引起一定程度的分子抗性. 在上述三个位点上发生的疏水性氨基酸残基的突变有可能产生针对阔草清的抗性. 在结合空腔内侧, 有黄色区块包裹着 464 位点的色氨酸, 黄色区块表示在该区域发生亲水性氨基酸残基的突变可能产生一定程度的分子抗性. 即将 464 位点的色氨酸突变为亲水性的氨基酸残基也可能是造成该位点突变体抗性的一个重要原因.

在氢键受体场效应图中(图 5E), 粉色区块表示在该区域引入具有氢键受体的氨基酸残基突变对抑制剂的抑制作用是有利的, 红色区块表示在该区域引入具有氢键受体的氨基酸残基突变则对抑制剂的抑制作用是不利的, 有可能产生一定程度的分子抗性. 26 位点的丙氨酸周围出现红色的区块, 则在该位点引入具有氢键受体的氨基酸残基突变对阔草清的抑制作用是不利的. 另外, 粉红色区块散落在各个突变位点的周围, 难以找到对应的结构抗性关系来描述其特征.

2.5 MB-QSAR 模型对除草剂种属选择性的预测

本质上, 不同种属的酶可以看作是酶进化中的不同突变体. 药物对靶标的选择性和靶标对药物的抗性的共同本质是药物与靶标的相互作用, 由此构成了评价药物选择性和抗性的共同标准. 因此 MB-QSAR 方法不仅是靶标抗性预测方法, 也是药物选择性的预测方法.

采用本文及文献[12]中建立的 MB-QSAR 模型, 我们初步预测了四个商品化除草剂氯磺隆、氯嘧磺隆、双草醚和阔草清对不同种属 AHAS 酶的抑制选择性, 其预测的 pK_i 值如表 3. 为验证这些数据, 本实验室构建表达了拟南芥、酵母、枯草芽孢杆菌及大肠杆菌的 AHAS

酶, 并实验测定了四个除草剂对上述酶的抑制常数(表 3 括号内数据). 数据表明, 以 *E. coli* AHAS II 野生型及突变体构建的 MB-QSAR 模型, 可以较好地预测其它细菌类 AHAS 酶的抑制活性, 即四个商品化除草剂对大肠杆菌 AHAS 三个同功酶、枯草芽孢杆菌 AHAS 酶抑制常数的实测数据与预测数据具有较好的一致性; 对酵母 AHAS 抑制常数的预测要差一些; 对于拟南芥 AHAS 酶, 抑制常数的实验和预测数据偏差较大, 并且实验数据值高于预测值. 分析这一结果的原因, 可能有三:

(1) 小分子与酶作用方式的差异. MB-QSAR 的基本假设是研究体系中生物大分子与药物小分子具有相同或类似的作用方式. 这样, 除去数据及模型质量的因素外, 预测差的结果通常暗示了被预测体系与模型体系中分子作用方式的差异. 事实上, 报道拟南芥 AHAS 与咪唑啉酮复合物晶体结构的文章中特别提到, 咪唑啉酮结合在拟南芥 AHAS 的两个区域^[28]. 除出现在本文阔草清分子结合的对应该区域外, 还结合在距此大约 2 nm 的区域. 如果阔草清类似于咪唑啉酮结合在拟南芥 AHAS 两个区域, 那么以单区域构建的 MB-QSAR 模型预测的抑制常数低是有可能的.

表 3 利用 MB-QSAR 模型对四种除草剂分子作用种属选择性的预测结果

Table 3 Prediction of the pK_i of four herbicides against AHAS from different species using MB-QSAR model

序列编号	种属	Pred. pK_i (氯磺隆)	Pred. pK_i (氯嘧磺隆)	Pred. pK_i (双草醚)	Pred. pK_i (阔草清)
AAO53551.1	小麦	3.18	4.28	4.71	1.12
ACD74789.1	水稻	3.96	4.60	3.57	1.22
NP_001151761.1	玉米	2.95	3.84	4.44	0.75
CAA87083.1	棉花	3.51	6.03	4.86	0.25
AD156521.1	马铃薯	4.25	5.04	3.65	0.60
P09342	烟草	3.70	4.43	4.64	1.14
P09114	烟草	3.56	4.79	4.42	1.34
ACU30048.1	大豆	3.38	4.50	4.90	0.63
ACV84152.1	菜豆	2.48	4.18	3.77	1.40
P27818	油菜	3.69	4.25	4.82	1.43
P14874	油菜	3.35	4.11	4.43	1.61
P27819	油菜	3.61	4.26	4.82	1.36
XP_002511176.1	蓖麻	3.58	5.84	5.15	0.35
EEF51778.1	蓖麻	3.58	5.84	5.15	0.35
ACR55067.1	燕麦	3.74	5.39	4.89	1.52
AAT07322.1	向日葵	3.51	5.17	4.75	1.24
AAT07327.1	向日葵	3.89	5.10	4.74	0.67
AAT07329.1	向日葵	3.72	3.91	4.27	2.30
AEL89170.1	臭甘菊	4.75	5.11	4.82	1.28
AEL89171.1	臭甘菊	4.48	4.89	5.15	1.06
ABY57315.1	苜蓿	3.58	5.07	3.62	0.58
ABY57318.1	苜蓿	3.70	5.02	3.34	0.51
AAK50821.1	藜	4.14	4.63	5.27	0.97
AAK50820.1	藜	4.12	4.76	5.31	1.04
ABM92357.2	莎草	3.42	4.90	4.01	0.80
ADQ54157.1	猪毛菜	4.15	5.24	4.84	0.90
ACF47582.1	苣荬菜	3.58	4.65	4.83	1.33
P17597	拟南芥	3.91 (7.04)	4.66 (7.50)	3.96 (6.09)	0.37 (6.30)

续表

序列编号	种属	Pred. pK_i (氯磺隆)	Pred. pK_i (氯嘧磺隆)	Pred. pK_i (双草醚)	Pred. pK_i (阔草清)
P36620	粟酒裂殖酵母	4.22 (6.11)	4.34 (8.15)	5.51 (5.12)	2.02 (3.95)
P00893	大肠杆菌 I	3.45 (3.32)	2.90 (3.30)	3.05	0.86
P00892	大肠杆菌 II	6.29 (6.49)	7.15 (7.66)	4.98 (4.62)	5.68 (5.73)
P08142	大肠杆菌 III	4.38 (3.81)	4.02 (3.73)	3.45	2.23 (3.22)
P37251	枯草芽孢杆菌	4.81 (4.78)	4.25 (2.94)	4.06 (2.87)	2.33 (3.94)
P42463	谷氨酸棒杆菌	5.23	4.09	3.16	4.35
POA622	结核分支杆菌	3.53	3.80	3.74	3.99
POA623	牛分支杆菌	3.53	3.80	3.74	3.99
P45261	流感嗜血杆菌	5.33	5.77	3.71	2.52
O33112	麻风病分支杆菌	3.67	4.31	3.41	3.09
Q59498	副结核杆菌亚种	4.54	4.16	3.86	3.50
P57321	宿主豆长管蚜内共生菌	4.34	4.77	4.69	2.37
O85293	宿主麦二叉蚜内共生菌	4.06	4.67	4.93	1.85
Q89AP7	宿主木瘦蚜属内共生菌	4.01	4.12	4.39	2.50
Q02137	乳酸乳球菌	5.27	4.49	2.79	3.55
P40811	鼠伤寒沙门菌	4.38	4.23	3.75	2.47
Q9RQ65	宿主角倍蚜内共生菌	3.64	4.08	4.66	2.59

^a 括号内数据为实验测定值。

(2) 计算分子场值所涉及的氨基酸残基(场值氨基酸残基)的差异。我们分析了预测种属 AHAS 酶与模型种属 AHAS 酶(*E. coli* AHAS II)场值氨基酸残基的差异性(见支持信息, 图 S2 及表 4)。分析表明拟南芥 AHAS 酶的场值氨基酸残基中, 与 *E. coli* AHAS II 比较, 非保守残基较多。但是与 *E. coli* AHAS III 相比, 该值差异并不显著, 而 *E. coli* AHAS III 的 pK_i 值预测结果却与实验测试值更为接近。因而, 该值是否是造成拟南芥和酵母 AHAS 对于阔草清 pK_i 值预测结果差的主要原因, 还值得进一步分析, 比如非保守氨基酸残基侧链的构象, 及其对于相邻残基构象的影响, 都有可能最终影响到抑制剂分子的结合。

(3) 本文中的不同种属 AHAS 均是在大肠杆菌表达系统中表达的。通常, 真核生物的蛋白质有更多的可修饰氨基酸, 不能排除在表达过程中氨基酸未预期的修饰或变异所带来的微结构的变化及由此引起的功能上的改变。这样, 基于通常结构的预测, 有可能出现偏差。

因此, 更深入地研究不同除草剂与不同种属 AHAS 酶的作用方式, 获取其复合物结构, 同时系统分析场值氨基酸残基的权重是非常必要的。目前, 我们针对这四个除草剂与不同种属 AHAS 的结合模式也在进行进一步的优化和验证。

表 4 模型种属 *E. coli* AHAS II 与预测种属 AHAS 场值氨基酸残基差异性比较

Table 4 Comparison of the residues of AHAS in herbicide binding site from model species and AHAS from prediction set

模式种属	不同场值氨基酸残基数	非保守氨基酸残基数
拟南芥	19	13
酵母	18	11
枯草芽孢杆菌	17	9
大肠杆菌 I	14	10
大肠杆菌 III	26	12

3 结论

本文测定了三唑嘧啶磺酰胺类除草剂阔草清对 *E. coli* AHAS II 及其突变体的抑制常数。利用生物大分子的定量构效关系方法(MB-QSAR)建立了定量、可视的突变体性质与结构关系模型。实验数据与计算数据之间显示了良好的相关性: MB-QSAR/CoMFA 模型的交叉验证系数 $q^2=0.691$, 非交叉验证相关性系数 $r^2=0.947$, 外部预测能力值 $r^2_{\text{pred}}=0.759$; MB-QSAR/CoMSIA 模型的相应值分别为 0.625, 0.960 和 0.619。该模型具有较好的外部预测能力, 使预警除草剂抗性突变成为可能。模型的三维结构-性质关系图直观阐释突变体抗性结构基础, 为抗性规避型除草剂分子设计提供了有价值的指导信息。同时, 该模型也在一定程度上显示了除草剂种属选择性的预测能力。

4 实验部分

4.1 试剂与仪器

阔草清由本课题组合成, 其纯度在 95%以上。其结构通过 ^1H NMR (nuclear magnetic resonance)、 ^{13}C NMR、LC-MS (liquid chromatography-mass spectrometry)等方法进行了验证。其他试剂和仪器参见文献^[27]。

4.2 实验方法

4.2.1 *E. coli* AHAS II 突变体的构建、表达及活性测定

E. coli AHAS II 突变体的构建、表达及酶活性以及化合物的筛选参照文献^[11-12,27]进行。突变体在质粒 pQE-GMwt (由以色列 Ben-Gurion University of the Negev 的 David M. Chipman 教授提供)的基础上构建的。野生型 AHAS 酶由质粒 pQE-GMwt 表达得到。阔草清用二甲基亚砜(DMSO)溶解, 控制 DMSO 在反应体系中的

体积分数不超过 4%。反应液中含有 50 mmol/L 磷酸钾缓冲液(pH 7.5), 10 mmol/L MgCl_2 , 1 mmol/L 硫胺素焦磷酸(ThDP), 10 $\mu\text{mol/L}$ 黄素腺嘌呤二核苷酸(FAD), 100 mmol/L 丙酮酸, 以及各种浓度的化合物。阔草清针对每种突变体的抑制常数 K_i 至少重复测定 2~4 次。抑制常数 K_i 的处理方法参考文献^[11-12,27,29]。

我们定义 K_i (突变体)与 K_i (野生型)的比值为抗性因子(resistance factors, RF , $RF = K_i(\text{突变体})/K_i(\text{野生型})$) (表 S1)。 RF 值越大, 说明酶的点突变对酶抑制剂的抑制活性影响越大, 与野生型酶比较, 该突变酶对酶抑制剂越不敏感, 或者说, 抑制剂对酶渐渐失去抑制活性。

4.2.2 野生型 *E. coli* AHAS II 催化亚基及其突变体活性构象的构建

前期工作中, 我们已经获得了 *E. coli* AHAS II 催化亚基及系列突变体的三维结构^[11,30]。本文中, 我们利用 AutoDock4.0^[30] 软件将阔草清对接到野生型 *E. coli* AHAS II 催化亚基结构中, 并通过结构优化获得野生型 *E. coli* AHAS II 催化亚基与阔草清相互作用的活性构象。按阔草清在野生型 *E. coli* AHAS II 催化亚基中的结合模式, 将其分别置于系列突变体结构中, 通过 SYBYL 软件中(Tripos Inc.)的 MINIMIZE SUBSET 模块优化, 获得 *E. coli* AHAS II 催化亚基突变体与阔草清相互作用的活性构象。相应计算参数参照文献^[11-13,31]。

4.2.3 野生型 *E. coli* AHAS II 催化亚基及其突变体活性构象的叠合

在 3D-QSAR 研究中, 叠合方式的选取应突显研究体系生物活性差异的结构特征, 避开可能存在的干扰模型的无关差异。参照前期工作^[11-13,30], 在本研究中我们尝试了三种叠合方式:

(1)叠合 *E. coli* AHAS II 催化亚基上所有氨基酸残基的骨架原子(A-I)。

(2)叠合阔草清结合空腔中的重要氨基酸残基骨架原子(A-II-3, A-II-4, A-II-5, A-II-6, 其中阿拉伯数字代表复合物结构中阔草清与周围氨基酸残基的最短距离, 单位为 Å)。

(3)叠合复合物结构中的阔草清中所有非氢原子(A-III)。

4.2.4 训练集和测试集的划分

依据训练集(Training set)与测试集(Test set)中的突变体氨基酸位置、类型相匹配, 生物活性数据分布均衡的原则, 将野生型 *E. coli* AHAS II 及其 77 个突变体划分训练集与测试集(表 S1)。

4.2.5 分子场计算及回归分析

对已叠合的一系列蛋白质大分子, 参照我们前期工作, 以如下参数, 计算比较分子场分析(CoMFA)中的分子场数值。格子空间范围: 1.2 nm $\times$ 1.2 nm $\times$ 1.2 nm, 空间格点步长为 0.2 nm, 探针原子类型: sp^3 , +1, 能量阈

值: 125.46 kJ $\cdot\text{mol}^{-1}$ 。

CoMSIA 场的计算采用与 CoMFA 模型相同的格点, 探针原子半径为 0.1 nm。采用高斯方程拟合格点与蛋白质之间的势能值, 衰减因子选用默认的 0.3。CoMSIA 模型里用到的分子场包括立体场(S), 静电场(E), 疏水场(H), 氢键给体场(D)和氢键受体场(A)。我们对所有可能的场组合方式进行了考察。

以计算获得的分子场数值(CoMFA 值)或分子场相似性指数(CoMSIA 值)作为自变量, 以阔草清表观抑制常数 K_i 的负对数($\text{p}K_i$)为因变量, 采用偏最小二乘法进行统计回归分析。通过交叉验证/非交叉验证, 得到最终蛋白质大分子的 MB-QSAR/CoMFA 及 MB-QSAR/CoMSIA 模型。

4.2.6 MB-QSAR 模型对除草剂种属选择性的预测

我们从 UniProt 数据库 (<http://www.uniprot.org/uniprot/>) 下载了部分杂草、作物以及细菌和真菌的 AHAS 酶序列。在 SWISS-MODEL 服务器 (<http://swissmodel.expasy.org/>) 上以我们模建的 *E. coli* AHAS II 的结构为模板同源模建了相应结构。按照 4.2.2 节所述方法得到它们与四个商品化除草剂氯磺隆、氯嘧磺隆、双草醚和阔草清相互作用的活性构象。将这些不同种属的 AHAS 酶视为 *E. coli* AHAS II 酶的突变体, 用文献^[11-13,30]及本文所建立的 MB-QSAR 模型对它们与不同除草剂作用的 $\text{p}K_i$ 值进行预测。

References

- [1] Hao, G. F.; Yang, G. F.; Zhan, C. G. *J. Phys. Chem. B* **2010**, *114*, 9663.
- [2] Chen, Y. Z.; Gu, X. L.; Cao, Z. W. *J. Mol. Graph. Model.* **2001**, *19*, 560.
- [3] Cao, Z. W.; Han, L. Y.; Zheng, C. J.; Ji, Z. L.; Chen, X.; Lin, H. H.; Chen, Y. Z. *Drug Discov. Today* **2005**, *10*, 521.
- [4] Hauser, K.; Negron, C.; Albanese, S. K.; Ray, S.; Steinbrecher, T.; Abel, R.; Chodera, J. D.; Wang, L. *Commun. Biol.* **2018**, *1*, 70.
- [5] Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849.
- [6] Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E.; Fox, T. *J. Med. Chem.* **2000**, *43*, 3033.
- [7] Wang, T.; Wade, R. C. *J. Med. Chem.* **2001**, *44*, 961.
- [8] Murcia, M.; Morreale, A.; Ortiz, A. R. *J. Med. Chem.* **2006**, *49*, 6241.
- [9] Deng, Z.; Chuaqui, C.; Singh, J. *J. Med. Chem.* **2004**, *47*, 337.
- [10] Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
- [11] He, Y. W.; Niu, C. W.; Wen, X.; Xi, Z. *Mol. Inform.* **2013**, *32*, 139.
- [12] He, Y. W.; Niu, C. W.; Wen, X.; Xi, Z. *Chin. J. Chem.* **2013**, *31*, 1171.
- [13] He, Y. W.; Niu, C. W.; Li, H.; Wen, X.; Xi, Z. *Sci. China Chem.* **2013**, *56*, 286.
- [14] Subramanian, M. V.; Gerwick, B. C. *Plant Physiol. Biochem.* **1989**, *277*.
- [15] Hawkes, T. *Monograph-British Crop Protection Council* **1989**, *42*, 131.
- [16] Shaner, D.; Anderson, P. C.; Stidham, M. A. *Plant. Physiol.* **1984**, *76*, 545.
- [17] Ray, T. B. *Plant. Physiol.* **1984**, *75*, 823.
- [18] McCourt, J. A.; Duggleby, R. G. *Amino Acids* **2006**, *31*, 173.
- [19] Rost, T. L. *J. Plant. Growth. Regul.* **1984**, *3*, 51.
- [20] Tranel, P. J.; Wright, T. R. *Weed Sci.* **2002**, *50*, 700.
- [21] Bernasconi, P.; Woodworth, A. R.; Rosen, B. A.; Subramanian, M. V.; Siehl, D. L. *J. Biol. Chem.* **1995**, *270*, 17381.

- [22] Mourad, G.; Williams, D.; King, J. *Planta* **1995**, *196*, 64.
- [23] Sibony, M.; Michel, A.; Haas, H.; Rubin, B.; Hurle, K. *Weed Res.* **2001**, *41*, 509.
- [24] Hattori, J.; Brown, D.; Mourad, G.; Labbé, H.; Ouellet, T.; Sunohara, G.; Rutledge, R.; King, J.; Miki, B. *Mol. Gen. Genet.* **1995**, *246*, 419.
- [25] Mourad, G.; King, J. *Planta* **1992**, *188*, 491.
- [26] Niu, C. W. *Ph.D. Dissertation*, Nankai University, Tianjin, **2005** (in Chinese). (牛聪伟, 博士论文, 南开大学, 天津, **2005**.)
- [27] Xi, Z.; Niu, C. W.; Li, Q. X.; Ouyang, D.; Ban, S. R. *Chin. J. Pestic. Sci.* **2005**, *7*, 215 (in Chinese). (席真, 牛聪伟, 李庆霞, 欧阳砥, 班树荣, 农药学学报, **2005**, *7*, 215.)
- [28] McCourt, J. A.; Pang, S. S.; King-Scott, J.; Guddat, L. W.; Duggleby, R. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 569.
- [29] Xi, Z.; Niu, C. W.; Ban, S. R.; Li, Q. X.; Ouyang, D.; Huang, M. Z. *Chin. J. Pestic. Sci.* **2005**, *7*, 311 (in Chinese). (席真, 牛聪伟, 班树荣, 李庆霞, 欧阳砥, 黄明智, 农药学学报, **2005**, *7*, 311.)
- [30] He, Y. W. *Ph.D. Dissertation*, Nankai University, Tianjin, **2013** (in Chinese). (何寅武, 博士论文, 南开大学, 天津, **2013**.)
- [31] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.

(Cheng, B.)