






研究报告

Research Report

几种植物类伐昔洛韦水解酶基因的生物信息学分析

巴德仁贵 , 李爽 , 王绍萌 , 马勇 , 哈斯阿古拉 

内蒙古大学生命科学学院/内蒙古自治区牧草与特色作物生物技术重点实验室, 呼和浩特, 010021

 通讯作者, hasind@sina.com;  作者

计算分子生物学, 2015 年, 第 4 卷, 第 4 篇 doi: 10.5376/cmb.cn.2015.04.0004

收稿日期: 2015 年 04 月 28 日 接受日期: 2015 年 04 月 28 日 发表日期: 2015 年 04 月 28 日




© 2015 BioPublisher 生命科学中文期刊出版平台

本文首次发表在《基因组学与应用生物学》印刷版 2015 年, 第 34 卷, 第 02 期上。现依据版权所有人授权的许可协议, 采用 [Creative Commons Attribution License](#) 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。



摘要 以拟南芥和蓖麻为重点, 利用生物信息学对不同植物的类伐昔洛韦水解酶(valacyclovir hydrolase-like, VACVase-Like)蛋白的理化性质、保守结构域、信号肽、跨膜结构域、导肽、亲疏水性、二级结构及三级结构等进行预测和分析, 并构建了该家族系统发生树。结果表明, 除了可可树属于稳定蛋白外, 其余均属于不稳定蛋白。该基因家族均含有明显的信号肽、导肽、跨膜结构和亲疏水结构, 且结构组成上相似, 属于分泌蛋白, 但是对亚细胞定位分析发现, 在不同植物中存在和发挥功能部位均不相同。分析结果可为植物 VACVase-like 的研究提供有价值的信息, 为进一步研究其在植物体内的功能和作用机理提供依据。

关键词 类伐昔洛韦水解酶; 生物信息学; α/β 水解酶

Bioinformatical Analysis Of VACVase-Like Gene From Several Plants

Bade Rengui , Li shuang , Wang Shaomeng , Ma Yong , Hasi Agula 

College of Life Sciences, Inner Mongolia University, Inner Mongolia Key Laboratory of Herbage & Endemic Crop Biotechnology, Hohhot, 010021

 Corresponding author, hasind@sina.com;  Authors

Abstract Base on the *Arabidopsis thaliana* and *Ricinus communis*, The physical/chemical properties, conserved domain, signal peptide, trans-membrane domain, leader peptide, hydrophobicity/hydrophilicity, the secondary and tertiary structure of the proteins were predicted and analyzed for valacyclovir hydrolase-like (VACVase-Like) of several plants by bioinformatics tools, phylogenetic trees were constructed. The results showed that most of these VACVase-Like proteins are structural stable except for *Theobroma cacao*, In all these VACVase-Like proteins, have signal peptide, leader peptide, trans-membrane structure hydrophobicity /hydrophilicity domain, and similar to composition and structure, belong to the secretory protein, but the subcellular localization analysis found that exist and functional are not the same position in the different plants. The result of the plant VACVase-Like studies may provide very valuable information for experimental research and application development and theoretical basis for further study of the relationships between mechanism and function.

Keywords VACVase-like, Bioinformatical, α/β hydrolase

伐昔洛韦(Valaciclovir, VACV)是鸟嘌呤类似物类抗病毒药物, 它是阿昔洛韦(嘌呤核苷类似物)的前体药物, 是一种特异性疱疹病毒抑制剂, 为阿昔洛韦的 L-缬氨酸酯。伐昔洛韦在人(Stephen et al., 1993)、猴子(de Miranda and Burnette.,1994)、小鼠(Burnette and de Miranda., 1994)等体内可通过伐昔洛韦水解酶(valacyclovir hydrolase, VACVase)迅速

几乎完全转化为阿昔洛韦和 L-缬氨酸并被快速的吸收。研究发现 VACVase 以氨基酸酯前体药物的核苷类似物和 α -氨基酸为底物进行水解(Lai et al.,2008), 已研究表明 VACVase 能够广泛的水解如伐昔洛韦 (Valaciclovir) 和缬更昔洛韦 (valganciclovir)、齐多夫定(zidovudine)、氟尿苷(floxuridine)、吉西他滨(gemcitabine)等抗病毒、抗

癌核苷类似物(Kim et al., 2003; 2004)。

VACVase 是 α -氨基酸酯水解酶,属于 α/β 折叠水解酶家族,在水存在的条件下能够水解代谢 α -氨基酸的酯类。 α/β 折叠水解酶家族包括蛋白酶、脂肪酶、酯酶、去卤化酶、过氧化物酶以及环氧化物水解酶类等酶,其家族最显著的结构特点是含有 8 条 β 平行的折叠片层并 α 螺旋包围在两边,只有第二条 β 链是反向平行的。 α/β 水解酶能够为各类酶提供一个稳定的活性位点,催化残留由三个高度保守的三联体组成:一个亲核试剂(丝氨酸、半胱氨酸或天冬氨酸)定位于 $\beta 5$ 链,酸性残基几乎总是在定位于 $\beta 7$ 链上,绝对保守的组氨酸残基则位于最后一个 β 链上,证明在折叠结构大大增加变异底物专一性(Nardini et al.,1999)。

在人体中, VACVase 是水解伐昔洛韦和缬更昔洛韦的重要的酶,由 291 个氨基酸组成, VACVase 在几个癌细胞系中的高效表达,已从人体结肠癌细胞系(Caco-2 cells)内分离、克隆并纯化,在肝脏、肾脏中表达量极高,而在心脏、肠道、及骨骼肌中表达量极弱,证明该酶具有重要的解毒功能(Puente et al.,1995)。目前, VACVase 基因只有在动物体中有相关研究,在植物体内有何功能,作用机理尚不清楚,本文利用生物信息学法以拟南芥和蓖麻为重点,对不同植物的 VACVase-like 基因的核苷酸序列以及所对应的氨基酸序列的理化性质、亲疏水性、保守结构域、信号肽、跨膜结构域、导肽、二级结构及三级结构等进行预测和分析,为下一步研究该基因家族在植物体内的具体功能和作用机理提供重要的线索。

1 结果与分析

1.1 VACVase-like 理化性质的分析

利用 ProtParam 在线软件分析植物 VACVase-like 基因的蛋白序列(见表 1),植物 14 个氨基酸残基数目除了甜瓜 MELO3C009679P1 的 398 个氨基酸外,均在 440 以上,拟南芥的分子量最大,为 56.2111 kDa,甜瓜 MELO3C009679P1 分子量最小,为 44.7561kDa,分析其稳定性发现除了可可树稳定以外,其余均为不稳定的蛋白,黄瓜的 VACVase-like 蛋白最不稳定为 51.30%,稳定指数 $>40\%$ 不稳定。对所有的 VACVase-like 蛋白的总疏水性进行分析,发现除了葡萄、可可树、草莓正值外。其余均为负值,然而他们的值都在 $-0.2\sim+0.12$

之间均为两性氨基酸,研究报道认为数值介于 $-0.5\sim+0.5$ 主要为两性氨基酸(付海辉等, 2010),说明这三种植物的 VACVase-like 蛋白是疏水性蛋白。

1.2 植物 VACVase-like 氨基酸序列系统发生树的构建与分析

为了研究基因之间的进化关系,植物 14 个 VACVase-like 基因的系统进化树的构建采用距离依靠法(distance methods)中的邻位相接法(Neighbor-Joining, NJ)方法,对 VACVase-like 蛋白进行序列的比对分析,通过自举法分析(Bootstrap Analysis),当 Bootstrap 的经验值越大,越接近于初始设定的重复抽样次数,即可信度就越高。分析结果表明,自举法分析的期望值在每个节点上都较高,这很可能是 VACVase-like 家族成员之间的氨基酸序列在进化过程中高度保守所导致的,而进化树的可靠性通过独立元素法(discrete character methods)中的另外一种构建法,即最大简约法(Maximum Parsimony methods, MP)构建的进化树进行验证。并通过 MEGA 6.0 软件构建系统发生树(图 1),通过综合分析可进一步获知,在物种进化的过程中,在功能分化上的 VACVase-like 基因家族是高度保守的,在进化分析过程中,拟南芥(AT1G64670)、甜瓜(MELO3C009679P1)与其他植物的进化上的亲缘关系较远,进化初期便发生了分歧,但是各分支内仍然有物种内独特的旁系同源基因。桃和草莓、大豆和鹰嘴豆、番茄的两个蛋白以及黄瓜和甜瓜的蛋白在进化过程中高度保守,在进化上的亲缘关系最近。

1.3 VACVase-like 亲/疏水性的分析

通过蛋白质的亲水性和疏水性值的分析,可预测蛋白质表面的氨基酸残基的分布。利用 ProtScale 在线工具对植物 14 个 VACVase-like 氨基酸序列的亲水性和疏水性进行预测,结果如图 2 所示。蓖麻和拟南芥蛋白的异亮氨酸有最高值为 4.5,表明其疏水性最强,精氨酸的最低分值为-4.5,表明其亲水性最高,亲疏水性氨基酸分布均匀。经整体分析,蓖麻和拟南芥蛋白存在明显的疏水区和亲水区,用同样的方法分别对黄瓜、葡萄、可可树、毛果杨、桃、草莓、大豆、鹰嘴豆、番茄和甜瓜等植物的亲疏水性进行分析,结果表明这几种植物均存在明显的疏水区和亲水区(其它几种植物的图片分析结果未显示)。

表 1 植物 VACVase-like 的氨基酸组成及的理化性质

Table 1 Analysis of amino acid composition and physical/chemical properties VACVase-like in plants

物种 Species	登录号 Accession No.	氨基酸数目(aa) No. of amino acid (aa)	分子量(kD) Molecular weight (kD)	等电点 PI	脂肪系数 Aliphatic index	不稳定指数(%) Instability index (%)	总平均 疏水性 GRAVY
蓖麻 <i>Ricinus communis</i>	XP_002518915	476	53.631 5	9.24	90.88	42.58	-0.056
黄瓜 <i>Cucumis sativus</i>	XP_004143839	476	53.142 1	8.80	89.68	51.30	-0.005
葡萄 <i>Vitis vinifera</i>	XP_002279898	476	53.658 3	8.95	94.39	45.24	0.003
可可树 <i>Theobroma cacao</i>	EOY23882	477	53.979 1	8.92	97.42	38.81	0.012
毛果杨 <i>Populus trichocarpa</i>	XP_002304581	476	53.633 1	8.95	93.95	42.86	-0.039
桃 <i>Prunus persica</i>	EMJ10313	463	52.394 0	9.18	93.26	40.18	-0.033
草莓 <i>Fragaria vesca</i>	XP_004309537	460	52.115 4	8.75	89.39	44.84	0.003
大豆 <i>Glycine max</i>	XP_003535395	490	54.928 6	8.92	93.08	44.69	-0.020
鹰嘴豆 <i>Cicer arietinum</i>	XP_004495509	489	55.631 5	9.14	89.08	45.23	-0.130
番茄 <i>Solanum lycopersicum</i>	XP_004244657	470	53.628 3	8.91	97.21	48.04	-0.020
拟南芥 <i>Arabidopsis thaliana</i>	XP_004245454	449	50.574 7	8.75	91.40	43.26	-0.020
甜瓜 <i>Cucumis melo</i>	NP_194145	498	56.211 1	9.10	92.37	45.05	-0.128
人 <i>Homo sapiens</i>	AT1G64670	469	53.427 8	9.30	88.04	41.52	-0.149
	MELO3C005787P1	458	51.442 2	8.65	90.85	48.59	-0.021
	MELO3C009679P1	398	44.756 1	6.86	84.52	43.87	-0.181
	NP_004323	291	32.542 5	9.20	88.83	25.06	-0.242

1.4 VACVase-like 信号肽及导肽的预测和分析

信号肽是新合成多肽链中用于指导蛋白质的跨膜转移的 N-末端的氨基酸序列, 由 15~30 个氨基酸组成。一个为带有正电的碱性氨基末端及带有负电荷的 C 较长末端, 其含小分子氨基酸, 是信号序列切割位点, 而信号肽的主要功能区一段能够形成 α 螺旋结构疏水序列。在信号肽的引导下, 新合成的蛋白质进入内质网腔。而信号肽序列则在信号肽酶的作用下被切除(王汉屏, 2008)。利用 SignalP 4.1 Server 在线软件, 使用神经网络(NN)和隐马尔可夫模型(HMM)两种分析蓖麻和拟南芥氨基酸序列的信号肽所在位置, 结果见图 2。拟南芥和蓖麻的 max.C、Y 值都是 YES, 并且预测存在信号肽的可能性分别为 0.671、0.468 和 0.510, 并且存在明

显的酶切位点, 预测可能属于分泌蛋白, 其它植物的 VACVase-like 氨基酸序列采用同样的方法对进行分析, 结果显示均含有信号肽结构, 推测可能属于分泌蛋白。

导肽是新生蛋白 N-端一段大约 20~80 个氨基酸的肽链, 又称转运肽(transit peptide)或导向序列(targeting sequence), 它是游离核糖体上合成的蛋白质的 N-端信号。主要是带正电荷的碱性氨基酸(特别是 Arg 和 Lys)含量较为丰富, 具有较强的穿透组织和细胞膜的能力, 将肽链准确的定位在某一个组织上或者细胞的细胞器上是其主要功能(金元昌等, 2008)。因此, 通过导肽的预测, 对蛋白质的功能分析、作用机制和亚细胞定位具有一定的意义。利用 targetP1.1 server(Emanuelsson et al., 2000; 2007;

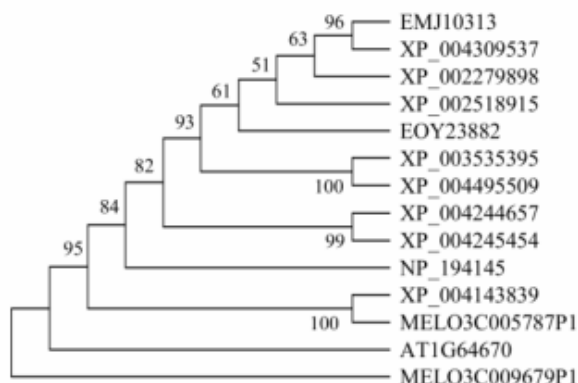


图1 VACVase-like 基因家族的系统发生分析, 自举法检验为 1 000 次重复

注: 黄瓜(XP_004143839); 葡萄(XP_002279898); 可可树(EOY-23882); 毛果杨(XP_002304581); 桃(EMJ10313); 草莓(XP_004-309537); 大豆(XP_003535395); 鹰嘴豆(XP_004495509); 番茄(XP_004244657, XP_004245454); 拟南芥(NP_194145, AT1G6-4670); 甜瓜(MELO3C005787P1, MELO3C009679P1); 蓖麻(XP_002518915)

Figure 1 Phylogenetic analysis of VACVase-like genes family in plants, and bootstrap analysis was conducted with 1000 replicates
Note: *Cucumis sativus* (XP_004143839); *Vitis vinifera* (XP_002-279898); *Theobroma cacao* (EOY23882); *Populus trichocarpa* (XP_002304581); *Prunus persica* (EMJ10313); *Fragaria vesca* (XP_004309537); *Glycine max* (XP_003535395); *Cicer arietinum* (XP_004495509); *Solanum lycopersicum* (XP_004244657, XP_0-04245454); *Arabidopsis thaliana* (NP_194145, AT1G64670); *Cucumis melo* (MELO3C005787P1, MELO3C009679P1); *Ricinus communis* (XP_002518915)

Nielsen et al.,1997;)在线软件对蓖麻和拟南芥氨基酸序列进行预测, 结果如图 3。结果表明, 蓖麻 VACVase-like 存在导肽结构, 定位于分泌途径其预测值为 0.520, 且可信度为 5 级, 拟南芥 VACVase-like 存在导肽结构, 定位于分泌途径其预测值为 0.901, 且可信度为 3 级(可信度区分 1:diff > 0.800, 2: 0.800 > diff > 0.600, 3: 0.600 > diff > 0.400, 4: 0.400 > diff > 0.200、5: 0.200 > diff), CmVACVL02 为分泌途径信号肽的可能最大。

1.5 VACVase-like 跨膜结构域的预测和分析

跨膜结构域(transmembrane domain, TM)是由 20 个左右的疏水氨基酸残基形成 α 螺旋, 与膜的脂双分子层的疏水区域相互作用, 将跨膜蛋白锚定在细胞膜中形成跨膜通道(翟中和等, 2007)。通过 TMpred 和 TMHMM Server v. 2.0 在线软件同时预测拟南芥和蓖麻 VACVase-like 和人 VACVase 氨基

酸序列的跨膜结构域, 再利用 TMHMM Server v. 2.0 在线软件进行证明结果的准确性。结果显示(图 4), 拟南芥(NP_194145)含有四个跨膜螺旋, 第一个跨膜螺旋从 16-40 位氨基酸, 由膜外到膜内; 第二个跨膜螺旋从 219-242 位氨基酸, 由膜内到膜外; 第三个跨膜螺旋从 291-308 位氨基酸, 相由膜外到膜内; 第四个跨膜螺旋从 347-365 位氨基酸, 由膜内到膜外。蓖麻 VACVase 蛋白同样含有四个跨膜螺旋, 第一个跨膜螺旋从 22-40 位氨基酸, 由膜外到膜内; 第二个跨膜螺旋从 199-221 位氨基酸, 由膜内到膜外; 第三个跨膜螺旋从 326-344 位氨基酸, 由膜外到膜内; 第四个跨膜螺旋从 347-365 位氨基酸, 由膜内到膜外。对其它植物 VACVase-like 蛋白序列进行分析, 结果如表 2。通过证明跨膜结构域存在也证实了信号肽存在的可能性。

1.6 亚细胞定位预测和分析

亚细胞定位是指某种蛋白或表达产物在细胞内的具体存在和发挥功能的部位。利用 WoLF PSORT(Horton et al., 2007)在线分析软件对植物 VACVase-like 蛋白的亚细胞定位进行预测和分析, 结果表明, 拟南芥(AT1G64670)的质膜得分 8.0, 拟南芥(NP_194145)在各组织中的分值较低, 蓖麻的胞质和叶绿体的分值分别为 5.0 和 4.0。番茄(XP_004244657)的在质膜得分 10.0, 桃的在细胞核的得分 9.0, 草莓的在胞质的得分 9.0, 其它在各组织中的分布预测结果如图 5。分析结果表明, VACVase-like 蛋白在不同植物中存在和发挥功能部位均不相同。

1.7 蛋白质二级结构的预测和分析

蛋白质的多肽链的主链骨架依靠羰基和酰胺基团之间形成的氢键作用, 盘曲折叠, 形成有规律的空间排布, 形成比较稳定的二级结构, 进一步加工形成有活性的特定构象。常见的二级结构有 α -螺旋和 β -折叠, 另外还有 β 转角、延伸链和无规则卷曲。VACVase-like 蛋白的二级结构本文利用 CFSSP(chou&fasman secondary structure prediction server) (Peter and Gerald, 1974a ;1974b)在线软件进行预测, 通过预测结果的分析, α -螺旋、 β -折叠、无规则卷曲和 β 转角是蓖麻和拟南芥 VACVase-like 蛋白的组成元件, 而 α -螺旋和 β -折叠是最主要的构成元件。对其它植物的 VACVase-like 蛋白的二级结构采用同样的方法进行预测, 通过预

Table 2 Transmembrane domain prediction of the VACVase-like protein from plants

物种 Species	登录号 Accession No.	跨膜螺旋数 No. of trans- membrane helix	总分值 Score
蓖麻 <i>Ricinus communis</i>	XP_002518915	4	4 157
黄瓜 <i>Cucumis sativus</i>	XP_004143839	4	4 308
葡萄 <i>Vitis vinifera</i>	XP_002279898	4	4 605
可可树 <i>Theobroma cacao</i>	EOY23882	4	4 963
毛果杨 <i>Populus trichocarpa</i>	XP_002304581	4	4 551
桃 <i>Prunus persica</i>	EMJ10313	4	4 743
草莓 <i>Fragaria vesca</i>	XP_004309537	4	4 533
大豆 <i>Glycine max</i>	XP_003535395	4	4 141
鹰嘴豆 <i>Cicer arietinum</i>	XP_004495509	5	4 879
番茄 <i>Solanum lycopersicum</i>	XP_004244657	3	3 732
拟南芥 <i>Arabidopsis thaliana</i>	XP_004245454	4	4 270
甜瓜 <i>Cucumis melo</i>	NP_194145	4	4 021
	MELO3C005787P1	3	3 603
	MELO3C009679P1	4	4 301
		3	4 120

测结果可以看出, 该蛋白家族在二级结构的构成元件上比例大致相似, 说明 VACVase-like 基因编码的蛋白质家族的二级结构和功能上具有一定的相似性。

1.8 植物 VACVase-like 蛋白的三级结构的预测和分析

蛋白质三级结构是指蛋白质分子处于它的天然折叠状态的三维构象, 是在蛋白质二级结构的基础上进一步盘绕, 折叠形成紧密的立体空间结构。随着蛋白质结构数据库中数据量的增多, 利用信息学的方法对蛋白质的高级结构进行预测和分析, 对理解的结构及功能具有一定意义, 一般采用从头算法 (*ab initio*)、线串法 (threading) 和同源建模 (homology modeling) 3 个计算方法, 对蛋白质分子

的三维结构进行初步模拟。利用 Swiss-Model Workspace 在线分析工具 (Johansson et al., 2012; Guex et al., 2009; Schwede et al., 2003), 对拟南芥和蓖麻的 VACVase-like 蛋白的三级结构进行同源建模, 并应用 Swiss-PdbViewer (DeepView v4.1) 软件显示, 结果表明, 与人 VACVase 蛋白三级结构比较发现, 蓖麻和拟南芥 VACVase-like 蛋白预测 C-末端是一个典型的 α/β 水解酶的折叠结构, 其中心的 β 片层是由一组 8 个折叠股构成的, β 片层的两侧则分布有五个和六个的 α 螺旋, N-端部分的氨基酸残基同样也参与 α/β 水解酶折叠结构域的形成, 折叠结构是 α/β 水解酶超家族的典型的结构特征, α/β 水解酶折叠的三个结构域区域都可以插入整个结构域或者一些片段, 也可以相应的进行延长。从同源建模的结果得知, 蓖麻和拟南芥 VACVase-like 蛋白氨基酸序列与参考蛋白的氨基酸序列的同源性较高均在 50% 以上, 所得到的建模结果的准确性均较高。拉氏构象图 (The Ramachandran Diagram) 主要是鉴定蛋白质的构象的合理性。对于一个三肽来说, 当中心肽基处在所有 ϕ 角和 ψ 角值时计算非键合原子间的允许距离, 通过在线软件 Swiss-PdbViewer 分析预测的结果, 通过分析 Psi (ψ) 角和 Phi (ϕ) 角的分布, 可大致的评估模拟的结构域与自然结构的相同程度及合理性。图 7 中的黄色区域是最理想的 Phi (ϕ) 角和 Psi (ψ) 角分布区域, 而蓝色区域外部则表示不合理区域 (α 表示 α -螺旋, $L\alpha$ 表示左旋 α -螺旋, β 表示 β -折叠, ppII 表示聚脯氨酸 II)。预测的蛋白质的空间结构较是否稳定取决于蛋白质残基的二面角 (>90%) 是否都位于黄色区域, 如果预测的二面角均在黄色区域内, 则同源建模的结果相对可靠。

1.9 植物 VACVase-like 蛋白的保守序列分析

利用在线工具 MEME Suite version 4.7.0 (Bailey and Elkan., 1995) 分别对植物 VACVase-like 蛋白进行了基序 (motif) 分析, 在每个 VACVase-like 蛋白质中分别得到 5 个保守基序 (见表 3), N-端在不同植物中进化变异性较高, 但 N-端的氨基酸残基也参与形成 α/β 水解酶折叠结构域, C-端则高度保守, 是一个 α/β 折叠水解酶典型结构, 且每个植物的保守区的位置基本相同 (图 8), 表明该基因家族在进化过程中高度保守, 说明 VACVase-like 基因家族在生物体中具有十分重要的作用。甜瓜

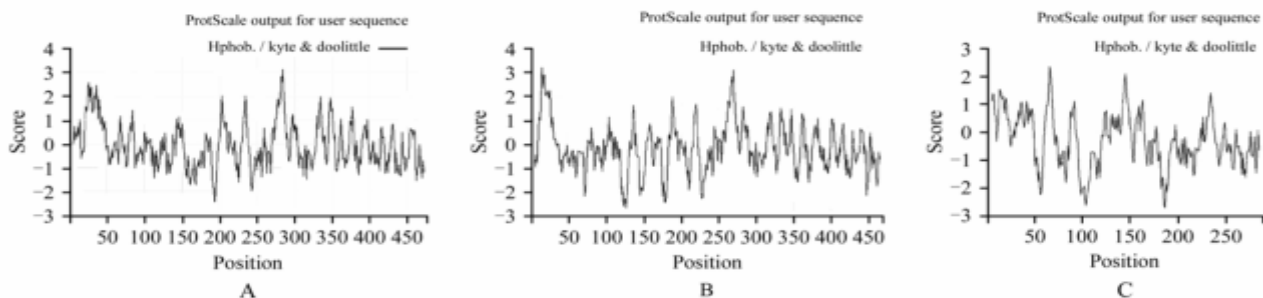


图2 蓖麻, 拟南芥 VACVase-like 和人 VACVase 氨基酸序列的亲水性和疏水性进行预测

注: A: 蓖麻(XP_002518915); B: 拟南芥(NP_194145); C: 人(NP_004323)

Figure 2 Hydrophobicity/phydrophilicity prediction of the VACVase-like proteins and VACVase from *Ricinus communis*, *Arabidopsis thaliana* and VACVase protein from *Homo sapiens*

Note: A: *Ricinus communis* (XP_002518915); B: *Arabidopsis thaliana* (NP_194145); C: *Homo sapiens* (NP_004323)

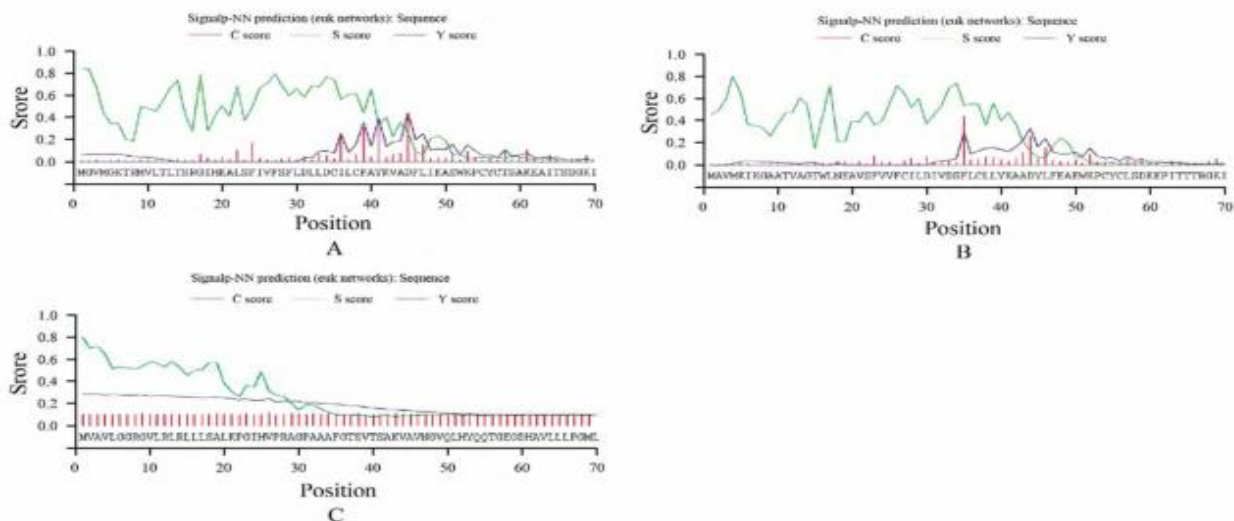


图3 拟南芥和蓖麻 VACVase-like 蛋白和人 VACVase 蛋白的信号肽的预测和分析

注: C 值: 原始剪切位点的分值; S 值: 信号肽的分值; Y 值: 综合剪切位点的分值; A: 蓖麻 (XP_002518915); B: 拟南芥(NP_194145); C: 人(NP_004323)

Figure 3 Signal peptide prediction of the deduced VACVase-like protein from *Ricinus communis* and *Arabidopsis thaliana* and VACVase protein from *Homo sapiens*

Note: C score: Primitive cleavage site score; S score: Signal peptide score; Y score: Synthetic cleavage site score; A: *Ricinus communis* (XP_002518915); B: *Arabidopsis thaliana* (NP_194145); C: *Homo sapiens* (NP_004323)

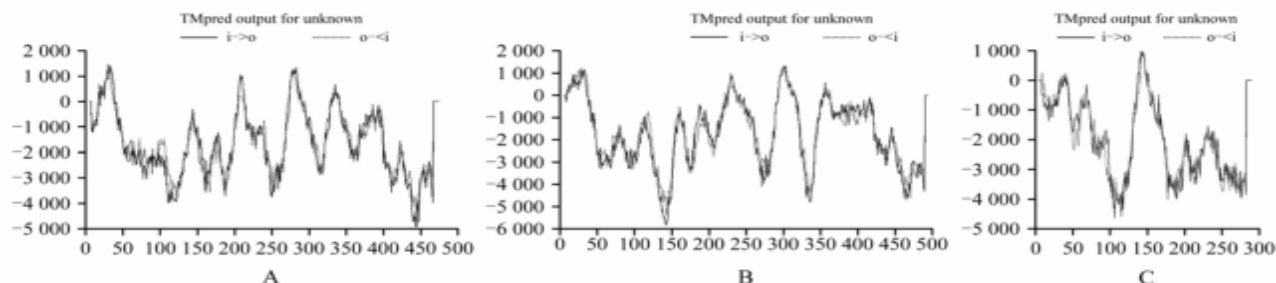


图4 拟南芥和蓖麻 VACVase-like 和人 VACVase 蛋白的跨膜结构预测

注: A: 蓖麻(XP_002518915); B: 拟南芥(NP_194145); C: 人(NP_004323)

Figure 4 Transmembrane domain prediction of the VACVase-like protein from *Ricinus communis* and *Arabidopsis thaliana* and VACVase protein from *Homo sapiens*

Note: A: *Ricinus communis* (XP_002518915); B: *Arabidopsis thaliana* (NP_194145); C: *Homo sapiens* (NP_004323)

(MELO3C009679P1)基序 motif 2 和 motif 4 在进化过程中缺失, 可能在功能上有冗余。随着选择压力的减小, 每个基因积累各自不同的遗传和变异, 从而形成结构和功能不同的新基因。

2 讨论

VACVase 在动物体内的肝脏、肾脏中表达量极高, 而在心脏、肠道及骨骼肌中表达量极弱, 证明该酶具有重要的解毒功能, 对其特异性分析表明 VACVase 是一个高度特定的氨基酸酯水解酶, 但在植物体内的功能尚不清楚。本文通过生物信息学软件和在线分析工具, 对蓖麻、黄瓜、葡萄、可可树、毛果杨、桃、草莓、大豆、鹰嘴豆、番茄、拟南芥和甜瓜等植物的 VACVase-like 蛋白进行分析。结果显示, 其稳定性发现除了可可树稳定以外, 其余均为不稳定的蛋白, 其中黄瓜的 VACVase-like 蛋白最不稳定为 51.30%, 研究认为, 不稳定指数>40%, 说明该蛋白为不稳定。该蛋白家族除了甜瓜和人均含有信号肽结构, 而 VACVase-like 基因编码的蛋白存在导肽和跨膜结构, 跨膜螺旋数目在 3-5 之间, 也证明了信号肽结构的存在。在系统发生分析过程中, 多数蛋白在进化上高度保守, 且各分支上都存在直系同源关系, 但是在各自的进化分支上仍然发现有物种内的独特的旁系同源基因, 可能在进化演变过程发生了独立的基因重复事件, 基因重复在演化过程中扮演重要角色, 它与物种的分化、生物体基因组大小的进化、新基因的产生等都密切相关。从而产生了旁系同源基因, 并且在功能上, 由于进化提供了较合适且更多的选择位点, 当基因重复事件发生后, 只要有一个基因是正常发挥功能, 这些基因的复制品通常可幸免于选择压力或选择压力就会大大减少, 当重复基因积累到一定程度便获得新的功能, 而且也是新基因产生的重要途径之一 (Force et al., 1999)。桃和草莓、大豆和鹰嘴豆、番茄的两个蛋白以及黄瓜和甜瓜的蛋白在不同物种间而进化过程中则高度保守, 在进化上的亲缘关系最近。可能是由同一祖先进化而来, 但在物种分化后, 在不同的选择压力下序列发生了不同方向的进化。虽然 VACVase-like 蛋白都属于 $\alpha\beta$ 折叠水解酶, 导致在亚细胞中的定位存在一定的差异主要与其所处的空间位置和功能相关。高级结构分析表明, 所有植物 VACVase-like 蛋白都由 α -螺旋、 β -折叠、 β 转角和无规则卷曲所组成且 α -螺旋和 β -折叠是主

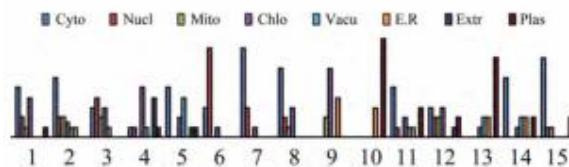


图 5 植物 VACVase-like 蛋白的亚细胞定预测

注: 1: 蓖麻(XP_002518915); 2: 黄瓜(XP_004143839); 3: 葡萄(XP_002279898); 4: 可可树(EOY23882); 5: 毛果杨(XP_002304581); 6: 桃(EMJ10313); 7: 草莓(XP_004309537); 8: 大豆(XP_003535395); 9: 鹰嘴豆(XP_004495509); 番茄(10: XP_004244657; 11: XP_004245454); 拟南芥(12: NP_194145; 13: AT1G64670); 甜瓜(14: MELO3C005787P1; 15: MELO3C009679P1)

Figure 5 The subcellular localization prediction of the VACVase-like protein from plants

Note: 1: *Ricinus communis* (XP_002518915); 2: *Cucumis sativus* (XP_004143839); 3: *Vitis vinifera* (XP_002279898); 4: *Theobroma cacao* (EOY23882); 5: *Populus trichocarpa* (XP_002304581); 6: *Prunus persica* (EMJ10313); 7: *Fragaria vesca* (XP_004309537); 8: *Glycine max* (XP_003535395); 9: *Cicer arietinum* (XP_004495509); *Solanum lycopersicum* (10: XP_004244657; 11: XP_004245454); *Arabidopsis thaliana* (12: NP_194145; 13: AT1G64670); *Cucumis melo* (14: MELO3C005787P1; 15: MELO3C009679P1)

要的构成元件, 三级结构中 VACVase-like 蛋白 C-末端是一个典型的 $\alpha\beta$ 水解酶的折叠结构, 其中心的 β 片层是由一组 8 个折叠股构成的, β 片层的两侧则分布有五个和六个的 α 螺旋, N-端是一个 β 推进器结构, 大约 10-30 个氨基酸残基组成, 该 β 推进器结构也参与形成 $\alpha\beta$ 水解酶折叠结构域 (Polgár, 1987), 折叠结构是 $\alpha\beta$ 水解酶折叠超家族的典型的结构特征。

3 材料与方法

3.1 数据来源

选择一个含有典型 DUF1234/ABHYDROLASE-1 结构域的蓖麻 (ACCESSION: XP_002518915) 和拟南芥 (NP_194145) 蛋白序列作为探针序列, 在 NCBI (National Center For Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>) 通过 blastp 检索蛋白数据库, 选择 $E \leq 10^{-10}$ 的序列为候选蛋白, 再利用 SMART (<http://smart.embl-heidelberg.de/>) 在线软件预测这些候选蛋白质的 DUF1234/ABHYDROLASE-1 结构域, 若存在 DUF1234/ABHYDROLASE-1 结构

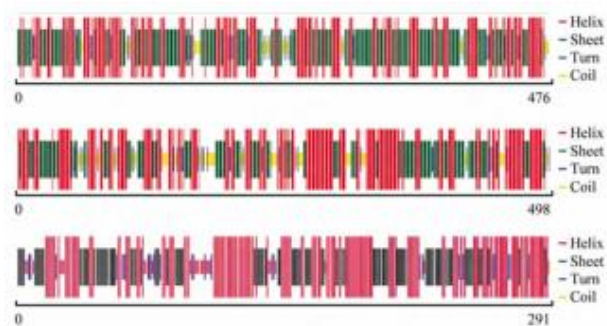


图 6 蓖麻(XP_002518915)拟南芥(NP_194145) VACVase-like 和人(NP_004323) VACVase 蛋白的二级结构的预测

Figure 6 Predicted the secondary structure of the VACVase-like protein from *Ricinus communis* and *Arabidopsis thaliana* and VACVase protein from *Homo sapiens*

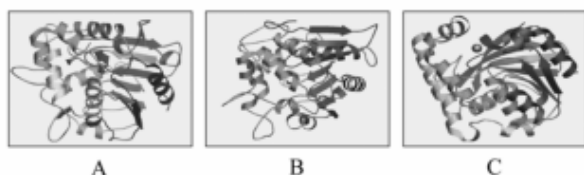


图 7 蓖麻(XP_002518915)拟南芥(NP_194145) VACVase-like 和人(NP_004323) VACVase 蛋白的三维高级结构的预测

注: A: 蓖麻(XP_002518915); B: 拟南芥(NP_194145); C: 人(NP_004323)

Figure 7 Predicted the tertiary structure proteins of the VACVase-like protein from *Ricinus communis* and *Arabidopsis thaliana* and VACVase protein from *Homo sapiens*

Note: A: *Ricinus communis* (XP_002518915); B: *Arabidopsis thaliana* (NP_194145); C: *Homo sapiens* (NP_004323)

域则认为该蛋白属于 α/β 折叠水解酶家族的类伐昔洛韦水解酶, 最终确定, 黄瓜(XP_004143839)、葡萄(XP_002279898)、可可树(EOY23882)、毛果杨(XP_002304581)、桃(EMJ10313)、草莓(XP_004309537)、大豆(XP_003535395)、鹰嘴豆(XP_004495509)、番茄(XP_004244657, XP_004245454)、拟南芥(NP_194145, AT1G64670)、甜瓜(MELO3C005787P1、MELO3C009679P1)和人(NP_004323)。

3.2 类伐昔洛韦水解酶的生物信息学分析

应用 ExPasy 数据库中的 ProtParam (<http://www.expasy.org/tools/protparam.html>) 软件分析预测候选蛋白的理化性质包括: 氨基酸数目, 分子量, 理论等电点, 分子式, 消光系数, 脂肪系数等。氨基酸的同源比对及构建进化树应用 MEGA 5.10 软件分析。ProtScale(<http://www.expasy.org/protscale/>) 软件预测

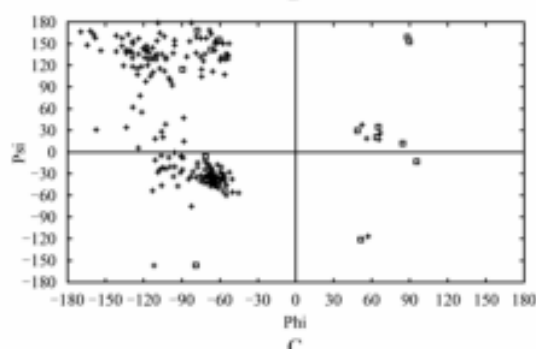
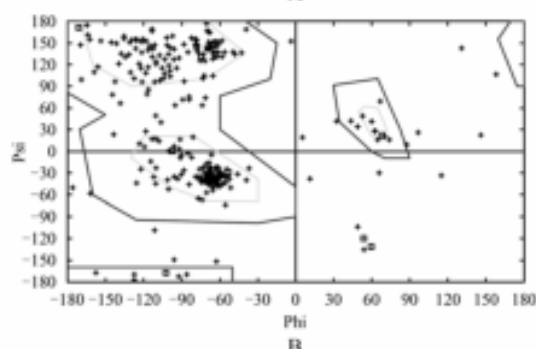
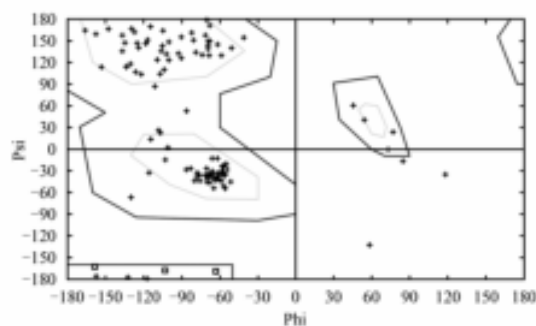


图 8 蓖麻(XP_002518915)拟南芥(NP_194145) VACVase-like 和人(NP_004323) VACVase 蛋白三维建模的拉式构象图

注: A: 蓖麻(XP_002518915); B: 拟南芥(NP_194145); C: 人(NP_004323)

Figure 8 The tertiary structure of ramachandran diagram of the VACVase-like protein from *Ricinus communis* and *Arabidopsis thaliana* and VACVase protein from *Homo sapiens*

Note: A: *Ricinus communis* (XP_002518915); B: *Arabidopsis thaliana* (NP_194145); C: *Homo sapiens* (NP_004323)

其亲水性, 采用 SMART(<http://smart.embl-heidelberg.de/>)在线软件预测蛋白质的保守结构域, SignalP V2.0.b2 (<http://www.cbs.dtu.dk/services/SignalP-2.0/>) 预测可能的信号肽, TargetP 1.1 Server(<http://www.cbs.dtu.dk/services/TargetP/>)预测可能的前导肽, 跨膜结构域的预测应用 TMHMM Server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>)和

表 3 植物 VACVase-like 蛋白保守基序分析

Table 3 Analysis of conserved motifs of VACVase-like protein in plants

模体	序列标识
Motif	Sequence logo
模体 1	DL LGFGRSPKP[ATS][DE]SLYTL[RK]EH[LV][ED]MIE[KR]SV[LI]E[APQ][YH]KVKSFHIVAHSGLGCILA
Motif 1	
模体 2	EGFFCHTHNAAWHTLHNIICGT[AG][GS]K[IL][DE][GR]YL[DE][AT]V[RK][DE][RHN][LV][KN]C[DE]V[NT][IV]
Motif 2	FHGK[DN]DE
模体 3	[QE][AP]TQ[YH]VMR[KR][VI]APRRVWP[PV][IM]A[FL]GAS[IL]ACWYEHI[ST]RT[IV]CL[LV]ICKNHR[LF]WEFL
Motif 3	
模体 4	CSYNV[QK][SK][RK][IV]PRA[RQ]V[KN]V[IV][EDQ][NK]KDHTIV[VI]GRQKAFARELEEIW
Motif 4	
模体 5	[KD]DK[AP][REKQ]E[DN]VLFIHGFISSAFWTET[LV]FPN[FL]S[SA][AST]AK
Motif 5	

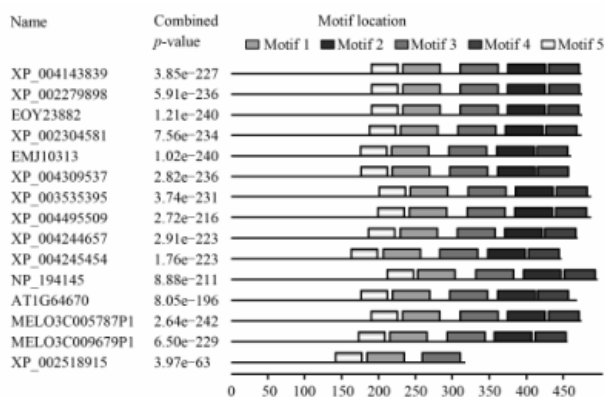


图 9 植物 VACVase-like 蛋白保守基序分布

Figure 9 Distribution of conserved motifs of VACVase-like protein in plants

Tmpred(http://embnet.vital-it.ch/software/TMPRED_form.html) 在线工具进行。二级结构预测由 BioGem.Org 数据库中的 CFSSP(chou&fasman secondary structure prediction server)(<http://www.biogem.org/tool/chou-fasman/>) 进行, 三级结构由 Swiss-Model(http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1) 进行同源模块的构建, 再用 swiss-pdb viewer 计算得出拉氏构象图 (Ramachandran), 以检验其合理性。利用在线工具 MEME Suite version 4.7.0 (<http://meme.sdsc.edu/meme/meme-intro.html>) 分别对植物 VACVase-like 蛋白进行了基序分析。

作者贡献

本论文构思、数据分析和写作由巴德仁贵完成; 论文的修改和定稿由哈斯阿古拉完成; 国家自

然科学基金由哈斯阿古拉主持;

致谢

感谢内蒙古自治区高等学校创新团队发展计划 (No. NMGIRT1401)

参考文献

- Bailey T.L., Elkan C., 1995, The value of prior knowledge in discovering motifs with MEME, Proc Int Conf Intell Syst Mol Biol, 3:21-29
- Burnette T.C., and De Miranda P., 1994, Metabolic disposition of the acyclovir prodrug valaciclovir in the rat, DMD January, 22 (1): 60-64
- De Miranda P., and Burnette T.C., 1994, Metabolic fate and pharmacokinetics of the acyclovir prodrug valaciclovir in cynomolgus monkeys, Drug Metab Dispos January, 22: 55-59
- Emanuelsson O., Nielsen H., Brunak S., and von Heijne G., 2000, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, Journal of molecular biology, 300(4): 1005-1016
- Emanuelsson O., Brunak S., von Heijne G., Nielsen H., 2007, Locating proteins in the cell using TargetP, SignalP, and related tools, Nature Protocols, 2(4): 953-971
- Force A., Lynch M., Pickett F.B., Amores A., Yan Y.L., and Postlethwait J., 1999, Preservation of duplicate genes by complementary, degenerative mutations, Genetics, 151(4):1531-1545
- Fu H.H., Xin P.Y., Xu Y.L., Liu Y., Wei Y.J., Dong J., Cao Y.L.,

- and Zhou J., 2010, Bioinformatics Analysis of UFGT Gene from Several Economic Plants, *jiyinzuxue yu yingyong shengwuxue (Genomics and Applied Biology)*, 30(1) : 92-102(付海辉, 辛培尧, 许玉兰, 刘岩, 韦援教, 董娇, 曹有龙, 周军, 2010, 几种经济植物 UFGT 基因的生物信息学分析, 基因组学与应用生物学, 30(1) : 92-102
- Guex N., Peitsch, M.C., and Schwede T., 2009, Automated comparative protein structure modeling with SWISS-MODEL and Swiss-Pdb Viewer: a historical perspective, *Electrophoresis*, 30, S162-S173
- Horton P., Park K.J., Obayashi T., Fujita N., Harada H., Adams-Collier C.J., and Nakai K., 2007, WoLF PSORT: protein localization predictor, *Nucleic Acids Research*, 35(2): W585-W587.
- Jin Y.C., Xiang Y.J., Li J.P., 2008, Purification and Activity Analysis of Recombinant Human GnRH and Leader Peptide, *shengwu jishu tongbao(Biotechnology Bulletin)*, (3) : 87-91(金元昌, 向育君, 李景鹏, 2008, 重组人促性腺激素释放激素及导肽的分离纯化与活性分析[J]. 生物技术通报 (3) : 87-91
- Johansson M.U., Zoete V., Michielin O., and Guex N., 2012, Defining and searching for structural motifs using Deep View/ Swiss-Pdb Viewer, *BMC Bioinformatics*, 13:173
- Kim I., Song X.Q., Vig B.S., Mittal S., Shin H.C., Lorenzi P.J., Amidon G.L., Nardini M., and Dijkstra B.W., 1999, α/β Hydrolase fold enzymes: the family keeps growing, *Current Opinion in Structural Biology*, 9(6): 732-737.
- Kim I., Chu X.Y., Kim S., Provoda C.J., Lee K.D., and Amidon G.L., 2003, Identification of a Human Valacyclovirase biphenyl hydrolase-like protein as valacyclovir hydrolase, *The Journal Of Biological Chemistry*, 278(28): 25348-25356.
- Kim I., Song X.Q., Vig B.S., Mittal S., Shin H.C., Lorenzi P.J., and Amidon G.L., 2004, A novel nucleoside prodrug-activating enzyme: substrate specificity of biphenyl hydrolase-like protein, *Mol Pharm*, 1(2): 117-127.
- Lai L.S., Xu Z.H., Zhou J.H., Lee K.D., and Amidon G.L., 2008, Molecular Basis of Prodrug Activation by Human Valacyclovirase, an α -Amino Acid Ester Hydrolase, *The Journal of Biological Chemistry*, 283(14): 9318-9327.
- Nielsen H., Engelbrecht J., Brunak S. and von Heijne G., 1997, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, *Protein Engineering*, 10(1): 1-6.
- Peter Y.C., and Gerald D., 1974, Prediction of protein conformation, *Biochemistry*, 13(2): 222-245
- Peter Y.C., and Gerald D.F., 1974, Conformational parameters for amino acids in helical, β -sheet, and random coil regions calculated from proteins, *Biochemistry*, 13(2): 211-222
- Puente X.S., and López-Otín C., 1995, Cloning and expression analysis of a novel human serine hydrolase with sequence similarity to prokaryotic enzymes involved in the degradation of aromatic compounds, *Journal of Biological Chemistry*, 270(21): 12926-12932.
- Polgár L., 1987, Structure and function of serine proteases, *New Comprehensive Biochemistry*, 16:159-200.
- Schwede T., Kopp J., Guex N., and Peitsch M.C., 2003, SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Research*, 31(13):3381-3385.
- Wang H.P., 2008, Bioinformatics Analysis of Defensins in Different Plant, *zhiwu shenglixue tongxun (Plant Physiology Journal)*, 44(1):25-32(王汉屏., 2008, 不同植物防御素的生物信息学分析, 植物生理学通讯, 44(1):25-32
- Weller S., Blum M.R., Doucette M., Burnette T., Cederberg D.M., de Miranda P., and Smiley M.L., 1993, Pharma-cokinetics of the acyclovir pro-drug valaciclovir after escalating single-and multiple-dose administration to normal volunteers, *Clinical Pharmacology and Therapeutics*, 54(6): 595-605
- Zhai Z.H., Wang X.Z., and Ding X.M., 2007, *Cell Biology*, Higher Education Press, Beijing, China, pp.91(翟中和, 王喜忠, 丁明孝, 主编, 2007: 细胞生物学, 高等教育出版社, 中国, 北京, pp. 91)