文章编号:1000-7423(2011)-04-0279-06

【论著】

cDNA Cloning and Sequence Analysis of *Musca domestica* Antifungal Peptide-1 (MAF-1)

FU Ping, WU Jianwei*, GUO Guo

【**Abstract**】 **Objective** To clone the cDNA sequence of *Musca domestica* antifungal peptide-1 (MAF-1) and analyze the amino acid sequence of MAF-1 by bioinformatics method. **Methods** Based on the primer designed according to the N-terminal amino acid sequence of MAF-1, the cDNA and amino sequence of MAF-1 were obtained by the methods of RACE and NestPCR. The accuracy of the experiment was confirmed by RT-PCR. The characteristic of the sequence was analyzed by bioinformatics software. **Results** The length of the cDNA sequence of MAF-1 was 568 bp by 3′RACE, including an open reading frame (ORF) of 441 bp length and 3′UTR of 127 bp. It was a novel sequence with the submission number of HM178948 in GenBank since none homology was found when compared with other sequences by Blast. Added with the 9 amino acids that were not used to design primer, the whole sequence of MAF-1 was 156 amino acids conferred from its cDNA. 139 bp cDNA sequence was obtained by 5′RACE and the result was consistent to 3′RACE. The result of RT-PCR showed the cDNA of MAF-1 mature peptide was accurate. The bioinformatics analysis deduced that the theoretic molecular weight and isoelectric point of the whole protein sequence of MAF-1 gene were similar to those detected. The ExPASy illustrated that the MAF-1 gene had a signal peptide. There were abundant α-helix in it, the domain located between the 128 and 153 amino acid residuals. Subcellular analysis showed MAF-1 was almost in the nucleus. PredictProtein found two protein kinase C phosphalation sites and one N-myristoylation site, and predicted that it was not a globular protein. In the end, the three dimension image of MAF-1 was set up with 3D-pssm of ExPASy. **Conclusion** The cDNA sequence and the amino acid sequence of MAF-1 have been obtained and analyzed successfully.

【**Key Words**】 *Musca domestica*; Antifungal peptide; RACE; cDNA; Bioinformatics

CLC No: R384.21 Document code: A

家蝇抗真菌肽-1 的 cDNA 克隆及其编码蛋白序列的生物信息学分析

付萍, 吴建伟*, 国果

(贵阳医学院寄生虫学教研室, 贵阳, 550004)

【**摘要**】 **目的** 克隆家蝇抗真菌肽-1 (MAF-1) 的 cDNA 序列并对其编码的蛋白序列进行生物信息学分析。 **方法** 根据 MAF-1 的 N 端 30 个氨基酸序列设计简并引物, 采用 RACE 法和巢式 PCR 技术克隆 MAF-1 的 3′端和 5′端 cDNA 序列, 获得 MAF-1 的 cDNA 序列和氨基酸序列。根据所得 MAF-1 成熟肽部分的 cDNA 序列设计引物进行 RT-PCR, 对序列进行验证并运用生物信息学软件进行分析。 **结果** MAF-1 的 3′端 cDNA 序列长度为 568 bp, 开放阅读框长度为 441 bp, 编码蛋白序列共 147 个氨基酸, 3′端非编码区 127 bp。经 NCBI 中 Blast 比对未找到同源序列, 提示该基因为一全新序列, 登录到 GenBank, 获得登录号 HM178948。该基因编码的 147 个氨基酸加上 MAF-1 的 N 端未用来设计引物的 9 个氨基酸, 全长共 156 个氨基酸。由 5′RACE 获得 139 bp cDNA 序列, 分析所得 MAF-1 成熟肽的氨基酸序列与前述一致。RT-PCR 结果证实了 RACE 所得 MAF-1 序列的正确性。根据生物信息学分析方法对所推导的 MAF-1 蛋白全长序列进行分析, 其理论相对分子质量和等电点均与其实际检测值相近。利用 ExPASy 的各种分析工具对 MAF-1 分析得知, 该蛋白具有信号肽, 富含 α 螺旋, 有 3 个 α 螺旋区, 亚细胞定位分析其主要分布于细胞核内。利用 PredictProtein 分析发现,

Supported by the National Natural Science Foundation of China (No. 30960343), the Doctoral Fund of Ministry of Education [No. (2008)220], the Breeding Project of the Education Department, Guizhou Province[No. (2009) 0137], the Fund of the Education Department, Guizhou Province (No. F2008-6), the Science and Technical Fund of the Health Department, Guizhou Province (No. 2010-1-031)

Department of Parasitology, Guiyang Medical College, Guiyang 550004, China

* Corresponding author, E-mail: wjw@gmc.edu.cn

MAF-1序列中有 2 个蛋白激酶 C 磷酸化位点、1 个 N 端酰基化位点，并预测 MAF-1 为非球形蛋白。最后，利用 ExPASy 中的 3D-pssm (Phyre Version0.2) 模建了 MAF-1 的三维空间结构图。 **结论** 成功克隆了家蝇抗真菌肽-1（MAF-1）的 cDNA 序列，获知其编码的氨基酸序列和生物信息学信息。

【关键词】 家蝇；抗真菌肽；RACE；cDNA；生物信息学

基金项目： 国家自然科学基金（No. 30960343）；国家教育部博士点专项基金 [No.（2008）220]；贵州省教育厅培育项目 [No. (2009) 0137]； 贵州省教育厅基金（No. F2008-6）；贵州省卫生厅科学技术基金（No. 2010-1-031)

* 通讯作者， E-mail: wjw@gmc.edu.cn

Antimicrobial peptides are the key components of the insect immune system. It is also an insect's first defense line in resisting invasion by pathogenic microorganisms. Antimicrobial peptides are widely distributed among different insects and the presence of these functional peptides varies from species to species[1,2]. These peptides are heat-resistant and advantageously cause no damage to the cells of higher animals and hence these peptides have been regarded as new sources for antibacterial and antifungal agents[3]. Consequently the different varieties of antimicrobial peptides have become an increasingly important area of functional peptide research. At present, researchers have isolated antimicrobial and antifungal peptides from all kinds of insects. They cloned these peptide genes, got the expression products or modified the genes in order to get more appropriate products[4-8]. *Musca domestica* is an important medium as it breeds in pathogenic clustered environments and is consequently highly resistant to pathogenic invasion. We found that *Musca domestica* not only shares homologous peptide molecules with other insects, but also possesses new, unique antibacterial or antifungal peptides[9-11]. *Musca domestica* antifungal peptide-1 (MAF-1), isolated from the hemolymph of the third instar larvae of housefly, is a kind of acidic antifungal peptide with high molecular weight[11]. As this study shows it can kill *Candida albicans*. It is different from those alkaline peptides with low molecular weight. Further study of the relationship between the structure and the function, the mechanism of regulation and expression of this antifungal peptide is important to reveal the efficient mechanism of the efficient innate immunity of *Musca domestica*. Such study will also contribute to the efficient screening of new antifungal agents. In this study, according to the MAF-1 N-terminal amino acid sequence, the degenerate primer was designed and the MAF-1′s 3′-cDNA sequence was cloned by rapid amplification of cDNA ends. The MAF-′s complete amino acid sequences were obtained finally. At the same time, bioinformatic analysis methods have been used to forecast the MAF-1 functional domains and the secondary structure. It was expected useful to establish the expression syst-em, to prove control mechanisms and to reveal the relationship between structure and function.

Materials and methods

1 Primer designing and synthesis

According to the N-terminal amino acid sequences of MAF-1, ESAPAPEVSGDAVFSAIQNGLKNL GNAFF[11]. two degenerate primers W1 and W2 were designed for 3′end rapid amplification of cDNA ends. Reverse transcription primers and two downstream primers were provided by 3′RACE kit, 5′-RACE primers were designed based on the results of 3'RACE (table 1). The primers were synthesized by TaKaRa Biotechnology Co.Ltd.

Table 1 Primers for the cDNA Cloning of MAF-1

Primer name	Primer sequences(5′-3′)
W1(3′RACE)	GCICCIGCICCIGARGTIAGYGGIGAYGCIGTITT
W2(3′RACE)	GGIGAYGCIGTITTYAGYGCIATICARAAYGG
R1(5′RACE)	GGTGGGCGGTCTTGAAGTCATT
R2(5′RACE)	GGCACCCTCAATCCAGACCTTC
YW1(RT-PCR)	GGAATTCGAATCTGCC CCCGCCCCT GAGGT
YW2(RT-PCR)	CCCAAGCTTCTAGGCATGGGGCTTCATTTCCTTGGC

2 Extraction total RNA from housefly larvae

TRIzol Reagent (invitrogen, USA) was used to extract the total RNA from housefly larvae. The integrity of extracted RNA was determined by 1% agarose gel electrophoresis. Spontaneously, took appropriate

total RNA to determine its optical density in UV 260 and 280 nm.

3 3′ RACE

Following the 3′ Full RACE Core set Ver2.0 instruction (Takara, Japan), with the 3′ RACE Adaptor, 1 μg total RNA was reversely transcripted into cDNA as templates for nested PCR. The first PCR system had the volume in 50 μl containing 3.5 μl cDNA templates, 6.5 μl 1×cDNA Dilution Buffer Ⅱ, 0.25 μl TaKaRa LA Taq (5 U/μl), 2 μl Primer W1 (10 μmol/L), 2 μl 3′ RACE Outer Primer (10 μmol/L), 4 μl 10×LA PCR Buffer Ⅱ (Mg^{2+} Free), 3 μl $MgCl_2$ (25 mmol/L), 28.75 μl dH_2O. The amplication procedure had predenaturation in 94 ℃ 3 minutes then started the cycle 94 ℃ denaturation 30 sec, 55 ℃ annealing 30 sec, 72 ℃ extension 1 min. After 20 cycles, continue the 72 ℃ extension 10 min. The second PCR was carried out with the previous products as templates, containing W2 (10 μmol/L) 2 μl, 3′ RACE Inner Primer (10 μmol/L) 2 μl, 10×LA PCR Buffer Ⅱ (Mg^{2+} Free) 5 μl, $MgCl_2$ (25 mmol/L) 5 μl, LA *Taq* (5 U/μl) 0.5 μl, dNTP Mixture (2.5 mmol/L each) 8 μl, dH_2O_2 6.5 μl. The reaction condition followed the first one in 30 cycles. After the amplification, the product was checked in 5 μl PCR solution by 1% agarose electrophoresis. The special strand was recycled by kits. The objective DNA was linked to the pMD-218T vector and then transformed into *E. coli* DH5α. The positive clone was chosen out to inoculate Luria-Bertani medium through blue-white selection. In a night incubation, the plasmid DNA was extracted by kits and transmitted to TaKaRa for DNA sequencing.

4 5′ RACE

5′ RACE primers were designed based on the results of 3'RACE (table 1). First-strand cDNA synthesis was performed using DNase-treated RNA according to the manual of the SMART RACE cDNA amplification kit (Clontech). A series of reactions were conducted with Advantage 2 polymerase mix (Clon tech) using the UPM primer (Clontech manual) in combination with 5′ RACE primers R1 and R2. The outer and inner PCR program with the following settings was used: 94 ℃ for 3 min, and 30 cycles of 94 ℃ for 30 s, 55 ℃ for 30 s and 72 ℃ for 2 min, followed by 72 ℃ for 10 min. After the amplification, the product was checked in 5 μl PCR solution by 1% agarose electrophoresis. The special strand was recycled by kits. The objective DNA was linked to the pMD-18T vector and then transformed into *E. coli* DH5α. The positive clone was chosen out to inoculate Luria-Bertani medium through blue-white selection. In a night incubation, the plasmid DNA was extracted by kits and transmitted to TaKaRa for DNA sequencing.

5 RT-PCR

Reactions were conducted following the instruction manual of the One Step RT-PCR kit (TaKaRa). Each reaction was performed in 50 μl total volume with 20 μl RNase Free dH_2O, 1 μl total RNA, 25 μl 2×1 step buffer, 1 μl 20 μmol/L of each forward and re-verse gene-specific primer (table 1). The program used was: 50 ℃ for 30 min, 94 ℃ for 2 min, then 30 cycles of 94 ℃ for 30 s, 60 ℃ for 30 s, and 72 ℃ for 1 min. The product was checked in 5 μl PCR solution by 1% agarose electrophoresis. The strand was recycled by kits and was linked to the pMD-18T vector and then transformed into *E. coli* DH5α. The po-sitive clone was chosen out to inoculate Luria-Bertani medium through blue-white selection. In a night incubation, the plasmid DNA was extracted by kits and transmitted to TaKaRa for DNA sequencing.

6 Characterization of the cloned sequence

We used the Blast software in NCBI website to analyze the amplificated cDNA and calculated the amino acid sequence by DNAMAN, then the complete sequence of the peptide and its molecular weight and isoelectric point were obtained. Put the sequence into Blastp in NCBI for the comparative analysis.

7 Bioinformatics analysis for MAF-1

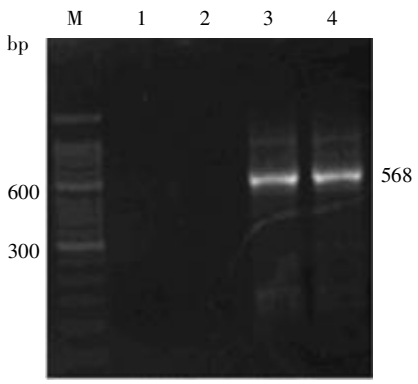
The MAF-1 amino acid sequence was analyzed through Tmpred in the bioinformatics website ExPASy (http://www.expasy.org/tools/) for transmembrane region, subcellular location in SignalP3.0 Server for signal peptide, domain function predicting in SMART, secondary structure check in SOPMA, functional site in PredictProtein. Finally, the three dimension model of

MAF-1 was set up with 3D-pssm of ExPASy.

Results

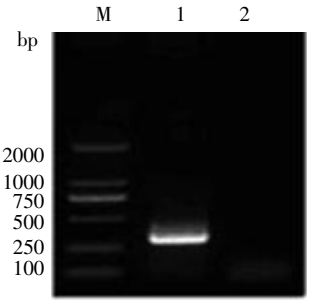
1 cDNA cloning of *Musca domestica* Antifungal Peptide-1 (MAF-1)

The total RNA of the third instar larvae of housefly was reversely transcripted into cDNA by RT-PCR. After the first PCR, the products were checked with 1% agarose gel electrophoresis, the result showed no specific bands. But after the second PCR with the primer W2, the specific band appeared near the 600 bp (Figure 1). 568 bp nucleotide sequence was obtained by sequencing. It was a novel sequence with the submission number of HM178948 in GenBank since none homology was found when compared with other sequences by Blast. With DNAMAN software translation of 147 amino acids, which GDA-VFSAMQNGLKNL-GNAFF was consistent to the N-terminal amino acid sequence of MAF-1. Added with the 9 amino acids that were not used to design primer, the whole sequence of MAF-1 was 156 amino acids conferred from its cDNA. After 5′RACE, the products of PCR were checked with 1% agarose gel electrophoresis, the result showed a specific band appeared near the 250 bp (Figure 2). 139 bp nucleotide sequence was obtained by sequencing. With DNAMAN software translation of 46 amino acids, which ESAPAPEVS was consistent to the N-terminal amino acid sequence of MAF-1. After the RT-PCR, the products were checked with 1% agarose gel electrophoresis, the result showed a specific band near the 500 bp (Figure 3). The sequencing result was consistent with the RACE.



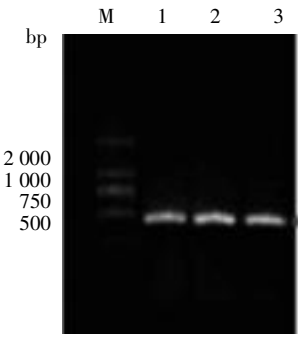
M: 50 bp DNA ladder marker; 1, 2: the product of first PCR; 3, 4: the product of second PCR.

Fig. 1 The result of 3′RACE for MAF-1



M: DL 2000 DNA Marker; 1: the product of inner PCR; 2: the product of outer PCR.

Fig. 2 The result of 5′RACE for MAF-1



M: DL 2000 DNA Marker; 1,2,3: the product of RT- PCR for MAF-1.

Fig. 3 The result of RT-PCR for MAF-1

2 Characterization of the cloned sequence

It included 13.5% lysine which was the highest content, and 12.8% alanine, 11.5% glutamic. The sequence of MAF-1 did not contain cysteine, suggesting no disulfide in it. After analysis, the theoretical molecular weight of MAF-1 was 17 181.2, it was similar to the accurate weight by mass spectrometry. The theoretical isoelectric point was 5.13, which was close to the accurate point.

3 The transmenbrane, signal peptide and subcellular location for MAF-1

The Tmpred suggested that MAF-1 had no transmembrane helix, it was a non-transmembrane protein. The SignalP 3.0 Server found that the most likely cleavage site was between pos. 33 and 34: ASA-RV. The analysis by the PSORT Ⅱ server showed that MAF-1 existed mainly in nucleus, little in cyroplasm.

4 Struture analysis for MAF-1

Using the Smart in ExPASy, the domain which had 26 amino acids was found between 128 and 153 amino acid (Figure 4), it was KKFKETADKLIESAKQ-

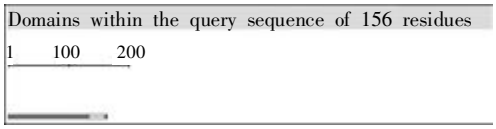


Fig. 4 The domains prediction for MAF-1

QLESLAKEMK, its' theoretical molecular weight was M_r 3 022 and the theoretical isoelectric point was 10.05. The analysis by SOPMA in ExPASy showed there was abundant α-helix in MAF-1. The PredictProtein found two protein kinase C phosphorylation site, one N-myristoylation site and multiple protein interaction sites, and MAF-1 was predicted not to be globular. Finally, the three dimension model of MAF-1 was set up with 3D-pssm of ExPASy. It showed that MAF-1 was a linear protein (Figure 5).

Fig. 5 The tertiary structure prediction for MAF-1

Discussion

Musca domestica (housefly) breeds in harsh environments where bacteria are ever present and yet are rarely infected. As a result their immune defense mechanism has aroused the interest of many researchers. Studies have found a variety of potent antimicrobial peptides in their immune defense mechanisms[3,12]. Researchers, using bioinformatics resources, have cloned cecropin, attacin, defensin, diptericin and other homologous molecules from housefly[4-8]. *Musca domestica* shares many homologous peptide molecules with other insects and in addition has a rich variety of unique antifungal, antivirus, antitumor peptide. Some peptides can also kill *Toxoplasma gondii*[12]. It is important to discover and clarify these new antimicrobial peptides and research the specific immune defense mechanism of housefly. In this study, according to the result of Fu and Wu[11], the antifungal peptide MAF-1 (including 568 bp 3′ end of cDNA sequence and a 156 amino acid residue of the mature peptide) was obtained from the third instar larvae of housefly by using RACE. It confirms that MAF-1 is a new antifungal insect peptide that has no homology with other antimicrobial peptides such as cecropins.

Insect antimicrobial peptides, based on structure and composition, are currently divided into several categories, namely linear-helix peptides, with disulfide bonds or the end of the open circular peptide, proline-rich and glycine-rich antimicrobial peptides, etc. Most linear-helical peptides have an alkaline isoelectric point, are constitutively expressed and are of low molecular weight, which basic amino acids were significantly higher than that of acidic amino acids [13]. MAF-1, a linear constitutively expressed antifungal peptide of high molecular weight, is rich in α-helix, which acidic amino acids are about 20% of the whole sequence, basic amino acids too. But it has an acidic isoelectric point so that we speculate it may be structurally modified. MAF-1 does not fully comply with the standards of a classification and it should be a new class of antimicrobial peptide.

In the MAF-1 protein sequence analysis, we found that the amino acid residue 128 to 153 may exist as a functional domain. The domain has 26 amino acids residues and its' pI is alkaline. The entire sequence has two protein kinase C phosphorylation sites and multiple protein binding sites. Protein kinase C phosphorylation sites and protein function are closely related to the activation or inactivation of the protein binding sites suggesting that MAF-1 may function with other immune molecules to play an effective role in coordinating and supporting other antimicrobial peptides to boost the innate immunity of *Musca domestica* through antimicrobial peptide synergy. Therefore we have reason to infer that MAF-1 works with other peptides or proteins, which enhance or reduce the biological activity of MAF-1 by binding with the MAF-1. It is yet to be determined which type of proteins or peptides combine with MAF-1 and how these proteins or peptides regulate the activity of MAF-1. The answer to these problems will lay the foundation for the further study of the innate immune system of *Mu-*

sca domestica.

Analysis of the functional domains of MAF-1 may identify the major functional areas of the peptide. If the MAF-1 gene expression in the function domain can be understood and its biological activity confirmed, new antifungal peptides of low toxicity could be provided for clinical use. In the future work we will therefore establish the recombinant expression system of the new antifungal peptide MAF-1 through cDNA sequence and bioinformatics analysis. An indepth study of the relationship between structure and function as well as specific regulatory mechanisms and so on, will also be performed to uncover the mechanisms behind *Musca domestic*′s efficient immune defense mechanisms. It is hoped this knowledge will lead to the synthesis of novel peptide drugs.

References

[1] Eleanor RH, Yannick M, Michael TSJ, *et al*. Antimicrobial defense and persistent infection in insects[J]. Science, 2008, 11(322): 1257-1259.

[2] Rahuma N, Ghenghesh KS, Ben Aissa R, *et al*. Carriage by the housefly (*Musca domestica*) of multiple-antibiotic-resistant bacteria that are potentially pathogenic to humans, in hospital and other urban environments in Misurata, Libya [J]. Ann Trop Med Parasitol, 2005, 99(8): 795-802.

[3] Zhao F. Advances of antimicrobial active substance from *Musca domestica* (houseflies)[J]. J Shanxi Agr Sci, 2007, 35(9): 23-26. (in Chinese)
(赵飞. 家蝇抗菌活性物质研究进展[J]. 山西农业科学, 2007, 35(9):23-26.)

[4] Liu FS, Wang LN, Tang T, *et al*. Cloning and characterisation of the diptericin gene in *Musca domestica*[J]. Acta Entomol Sin, 2009, 52(10): 1078-1082. (in Chinese)
(柳峰松, 王丽娜, 唐婷, 等. 家蝇抗菌肽 Diptericin 基因的克隆与分析[J]. 昆虫学报, 2009, 52(10): 1078-1082.)

[5] Xu QY, Zhu JY. Cloning and expression of defensin gene of *Musca domestica* larvae[J]. Chin Trop Med, 2005, 5(1): 4-6. (in Chinese)
(许琴英, 朱家勇. 家蝇幼虫 Defensin 成熟蛋白序列的克隆及在原核中的表达[J]. 中国热带医学, 2005, 5(1): 4-6.)

[6] Yang XR, Jing XB, Zeng AH, *et al*. Clone of full-length diptericin from *Musca domestica* and its bioinformatics analysis[J]. Chin J Vector Biol Control, 2009, 20(5): 393-396. (in Chinese)
(杨小蓉, 金小宝, 曾爱华, 等. 全长家蝇抗菌肽基因 diptericin的克隆及生物信息学分析[J]. 中国媒介生物学及控制杂志, 2009, 20(5): 393-396.)

[7] Geng H, An CJ, Hao YJ, *et al*. Molecular cloning and expression of attacin from housefly (*Musca domestica*)[J]. Acta Gen Sin, 2004, 31(12): 1344-1350. (in Chinese)
(耿华, 安春菊, 郝友进, 等. 家蝇攻击素 (Attacin) 基因的克隆与表达[J]. 遗传学报, 2004, 31(12): 1344-1350.)

[8] Wang W, Zheng XL. The prokaryotic expression of cecropin gene in *Musca domestica* and immunological identification of its exprressed products[J]. Chin J Zoonoses, 2009, 25(5): 439-445. (in Chinese)
(王炜, 郑学礼. 家蝇 Cecropin 基因的原核表达及表达产物的免疫学鉴定[J]. 中国人兽共患病学报, 2009, 25(5): 439-445.)

[9] Guo G, Wu JW, Fu P, *et al*. Biochemical characteristics of the antibacterial peptides in secretion of *Musca domestica* larvae[J]. Acta Entomol Sin, 2006, 49(6): 918-923. (in Chinese)
(国果, 吴建伟, 付萍, 等. 家蝇幼虫分泌物抗菌肽的生化特性初步研究[J]. 昆虫学报, 2006, 49(6): 918-923.)

[10] Guo G, Wu JW, Fu P, *et al*. Isolation and purification of antibacterial peptides from the larvae secretion of housefly and the characteristics[J]. Chin J Parasitol Parasit Dis, 2007, 25(2): 87-92. (in Chinese)
(国果, 吴建伟, 付萍, 等. 家蝇幼虫分泌物抗菌肽的分离及其部分特性[J]. 中国寄生虫学与寄生虫病杂志, 2007, 25(2): 87-92.)

[11] Fu P, Wu JW, Guo G. Purification and molecular identification of an antifungal peptide from the hemolymph of *Musca domestica* (housefly)[J]. Cel Mol Immun, 2009, 6(4): 245-251.

[12] Ma HX, Sun N, Pei ZH, *et al*. Research progress on antibiotic peptide of *Musca domestica*[J] Chin J Vet Drug, 2007, 4(11): 45-49. (in Chinese)
(马红霞, 孙娜, 裴志花, 等. 家蝇抗菌肽的研究进展[J]. 中国兽药杂志, 2007, 4(11): 45-49.)

[13] Wang YP, Lai R. Insect antimicrobial peptides: structures, properties and gene regulation[J]. Zool Res, 2010, 31(1): 27-34. (in Chinese)
(王义鹏, 赖仞. 昆虫抗菌肽结构、性质和基因调控[J]. 动物学研究, 2010, 31(1): 27-34.)

(**Received**: 2011-03-11 Editors: YU Sen-hai, YI Feng-yun)