

doi:10.3969/j.issn.1672-5565.2013.01.11

# 家蝇小热休克蛋白(sHsp20.6)的生物信息学分析

李显航,刘红美\*

(贵阳医学院医学生物技术教研室,贵阳 550004)

**摘要:**研究家蝇小热休克蛋白 sHsp20.6 的生物学功能。方法应用生物信息学的方法和工具对家蝇 sHsp20.6 的理化性质、疏水性、跨膜区和信号肽、膜体分析、二级结构功能域、蛋白质的功能分类预测、多重序列比对与系统发育树构建、三级结构建模进行分析。结果表明:家蝇 sHsp20.6 是一个亲水蛋白,分子量为 20.64kD,等电点为 5.66,不具有跨膜区和信号肽,包含有一个 HSP20 的结构域,主要构成原件为  $\alpha$  螺旋和无规则卷曲,三维结构预测显示该蛋白为棒状结构,C 端结构域具有 7 个片层结构。聚类分析显示,家蝇 sHsp20.6 蛋白与昆虫中的直系同源小热休克蛋白(orthologous small heat shock protein)聚为一类。

**关键词:** 家蝇,小热休克蛋白,生物信息学

**中图分类号:** Q518.2    **文献标识码:** C    **文章编号:** 1672-5565(2013)-01-065-07

## Bioinformatic analysis of small heat shock protein (sHsp20.6) in *Musca Domestica L.*

LI Xian-hang, LIU Hong-mei\*

(Department of Medical Biotechnology, Guiyang Medical College, Guiyang 550004, China)

**Abstract:** To research biological functions of the small heat shock protein (sHsp20.6) in *Musca Domestica L.*, many bioinformatic methods were used to analysis its physical and chemical properties, hydrophobicity, transmembrane region, signal peptide, motif, secondary structure, functional domain and assortment, multiple alignment, phylogenetic tree and three-dimensional structures. The results showed that sHsp20.6 was hydrophobicity protein and the calculated molecular mass was 20.64kD. The theoretical isoelectric point was 5.66. The protein without transmembrane regions and signal peptide contains a sHsp20 functional domain. Main composition of the protein secondary structure were  $\alpha$ -helix and random coil. The three-dimensional structure prediction expressed a rod-like structure which had seven layer structure of the C end. Clustering analysis showed that the protein belong to insects small heat shock protein orthologous class.

**Key words:** *Musca domestica L.*, Small Heat Shock Protein, Bioinformatics

热休克蛋白(heat shock protein, HSP)是由 Ritossa 在对果蝇热诱导时发现的<sup>[1]</sup>。热休克蛋白是生物体在高温(高于生长温度 8~12℃)或者其他类型的胁迫(冷、有机物、重金属、缺氧、微生物感染等)诱导细胞合成或表达量改变的一类蛋白。根据已知各种 Hsps 的分子量大小,普遍将 Hsps 分为 Hsp110、Hsp90、Hsp70、Hsp60 和小热休克蛋白

(sHsp)5 个家族。sHsp 分子量大约为 15~30kD,是一个极具多样的家族,广泛分布于微生物、低等动物和高等动物中。不同的物种含有不同的 sHsp,而且不同的组织,各种 sHsp 表达量也会不一样。大量研究已表明,不论是机体处于应激状态还是处于正常生理状态,sHsp 都有相应的表达,参与机体的生物学过程。sHsp 的主要结构由 N 端和 C 端 2 个部分

收稿日期:2012-05-16;修回日期:2012-06-20.

基金项目:贵州省科技厅联合基金(黔科合[2010]3141;贵州省高层次人才科研条件补助经费(867)。

作者简介:李显航,男,贵州凯里人,细胞生物学硕士研究生,Tel:18786124134,E-mail:372560673@qq.com.

\*通信作者:刘红美,E-mail:hmliu@gmc.edu.cn.

构成, sHsp 的 C 端具有一个与  $\alpha$  晶体蛋白的 C 端结构相似的保守区域, 编码 110 个左右的氨基酸残基。sHsp 在生物体内以一个低聚复合体 (200 ~ 800kD) 的形式存在, 在细胞受到胁迫时起到分子伴侣的功能, 对恢复和稳定细胞内环境有重要的作用。研究显示, 小热休克蛋白在昆虫中广泛存在, 并与昆虫的生长发育密切相关<sup>[2]</sup>。

前期研究中, 课题组在热胁迫后的家蝇幼虫 cDNA 文库中, 获得一个小热休克蛋白的 EST 序列<sup>[3]</sup>。经过进一步测序和序列比对, 获得其全长 (GenBank: ADT92004.1), 并依据其编码的小热休克蛋白分子量大小命名为家蝇 sHsp20.6。这是首次报道家蝇中的小热休克蛋白。为了进一步研究其生物学功能, 本文采用生物信息学的方法对家蝇 sHsp20.6 蛋白序列进行系统的预测与研究。

## 1 材料与方法

### 1.1 序列来源

家蝇 sHsp20.6 的氨基酸序列 (GenBank: ADT92004.1)。

### 1.2 理化性质、疏水性与跨膜区和信号肽

利用 ProtParam tool (<http://web.expasy.org/protparam/>) 在线软件对 sHsp 进行基本理化性质的分析。使用蛋白质分析软件 ProtScale (<http://web.expasy.org/protscale/>) 氨基酸分值参数选用 Hphob. / Ky t e & Dool i t t l e) 在线进行分析 sHsp 蛋白的疏水域; 用丹麦技术大学的 TMHMM2.0 程序 (<http://www.cbs.dtu.dk/services/TMHMM/>) 分析跨膜区, 用于跨膜区的定位。利用 Signalp 4.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) 软件对信号肽进行预测。

### 1.3 膜体分析、二级结构预测与分析、结构功能域的预测、蛋白质的功能分类预测

用 Prosite 程序 (<http://prosite.expasy.org/>) 对 motif 进行分析, 用于功能和结构域的划分; 使用 Predictprotein 工具 (<http://www.predictprotein.org/>) 对 sHsp 蛋白进行二级结构的预测。使用 SMART (<http://smart.embl-heidelberg.de/>) 预测家蝇 sHsp20.6 的结构功能域。家蝇 sHsp20.6 蛋白的功能分类采用 ProtFun 2.2 Server (<http://www.cbs.dtu.dk/services/ProtFun/>) 进行预测。

### 1.4 序列比对与构建系统发育树

通过 NCBI 提供的 BLAST 在线搜索服务软件 (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) 获取果蝇 (*Drosophila melanogaster*)、蜜蜂 (*Apis mellifera*)、

拟谷盗 (*Tribolium castaneum*)、冈比亚按蚊 (*Anopheles gambiae*) 以及酿酒酵母 (*Saccharomyces cerevisiae*) 的小热休克蛋白的氨基酸序列, 基因库登录号见表 1 所示。使用 MEGA4.1 软件自带的系统发育树构建功能对这些氨基酸进行分析, 选择邻接法 (neighbor-joining method, NJ) 构建系统发育树。挑选系统发育树中家蝇 sHsp20.6 所在聚类组包含的氨基酸序列采用 ClustalX2 软件进行多重序列比对, 获得他们之间蛋白质一级结构的区别。通过分析从一个方面推断不同昆虫小热休克蛋白基因之间的进化关系。

表 1 系统发育分析所用的蛋白质  
Table 1 Phylogenetic analysis of the protein

蛋白质 Protein	登录号 Accession number	来源 Origin
DmelHsp20.8	NM_134482	Drosophila melanogaster
DmelHsp20.6	NM_079275	Drosophila melanogaster
DmelHsp23.0	NM_079273	Drosophila melanogaster
DmelHsp23.6	NM_079276	Drosophila melanogaster
DmelHsp22.2	NM_079270	Drosophila melanogaster
DmelHsp18.0	NM_139898	Drosophila melanogaster
DmelHsp23.8	NM_140047	Drosophila melanogaster
DmelHsp46.9	NM_079274	Drosophila melanogaster
DmelHsp21.3	NM_079103	Drosophila melanogaster
DmelHsp24.5	NM_135499	Drosophila melanogaster
AmelHsp25.6	XM_392405	Apis mellifera
AmelHsp23.0	XM_001120194	Apis mellifera
AmelHsp22.0	XM_394333	Apis mellifera
AmelHsp21.3	XM_393575	Apis mellifera
AmelHsp24.2a	XM_001119830	Apis mellifera
AmelHsp27.7	XM_001120137	Apis mellifera
AmelHsp24.2b	XM_001120006	Apis mellifera
AmelHsp20.4	XM_395659	Apis mellifera
TcasHsp29.8	XM_969297	Tribolium castaneum
TcasHsp20.7a	XM_968349	Tribolium castaneum
TcasHsp20.7b	XM_961687	Tribolium castaneum
TcasHsp18.3	XM_969274	Tribolium castaneum
TcasHsp19.7	XM_968251	Tribolium castaneum
TcasHsp16.0	XM_968285	Tribolium castaneum
TcasHsp13.2	XM_968317	Tribolium castaneum
TcasHsp21.8a	XM_963667	Tribolium castaneum
TcasHsp22.2	XM_970284	Tribolium castaneum
TcasHsp21.8b	XM_968592	Tribolium castaneum
AgamHsp20.9	XM_560153	Anopheles gambiae
AgamHsp21.6	XM_308606	Anopheles gambiae
AgamHsp23.7	XM_308609	Anopheles gambiae
AgamHsp23.5a	XM_315549	Anopheles gambiae
AgamHsp23.4	XM_315550	Anopheles gambiae
AgamHsp23.5b	XM_308607	Anopheles gambiae
AgamHsp21.7	XM_308608	Anopheles gambiae
Scer23.9	NP_009628	Saccharomyces cerevisiae
MdomHsp20.6	ADT92004	Musca domestica L.

### 1.5 三维结构建模

蛋白质的生物学活性很大程度上取决于蛋白质

的空间构象,同源建模是目前最为成功且实用的蛋白质三维结构预测方法。利用 SWISS - MODEL(<http://swissmodel.expasy.org/>)软件对家蝇 sHsp20.6 蛋白进行同源建模,预测其三维结构。

## 2 结果与分析

### 2.1 理化性质、疏水性、跨膜区和信号肽分析

家蝇 sHsp20.6 基因包括 546bp 的开放阅读框,编码 181 个氨基酸。推测该蛋白的分子式为: C<sub>903</sub>H<sub>1449</sub>N<sub>251</sub>O<sub>289</sub>S<sub>6</sub>,相对分子量为 20.64kD,等电点为 5.66,不稳定参数为 52.89。根据不稳定参数的数值在 40 以下才是稳定蛋白的标准,推测 sHsp20.6 为不稳定的蛋白。该基因编码的氨基酸组成如图 1 所示。从图中发现 Glu 和 Ser 的含量较多,其中 Ser 的含量高达 11%。研究发现:小热休克蛋白丝氨酸的磷酸化作用会导致小热休克蛋白低聚

状态的改变,从而使小热休克蛋白的生物学功能得以发挥<sup>[4]</sup>。相对含量较少的氨基酸有 His 和 Tyr 等。带正电荷的残基(Arg + Lys)总数为 26 个,带负电荷的残基(Asp + Glu)总数为 30 个。家蝇 sHsp20.6 蛋白疏水性分析结果如图 2 所示,亲水氨基酸占 68.33%,而疏水氨基酸占 27.78%,为亲水蛋白质,在该蛋白质 116 - 140 区域的亲水性较强,在 C 端的 164 - 176 处有一个疏水性区域。跨膜分析的结果(图 3)显示该蛋白不含跨膜区域,这正好与热休克蛋白合成以后直接在细胞内发挥其分子伴侣的功能相关。使用 SignalP 4.0 Setver 软件对家蝇 sHsp20.6 蛋白预测结果(图 4)显示 mean S - score 值为 0.133,依据 mean S - score > 0.5 才能判断为分泌蛋白的标准,推测家蝇 sHsp 不具有信号肽,同样说明家蝇 sHsp20.6 蛋白是一种在细胞内表达以后直接在胞内起生理作用的蛋白。

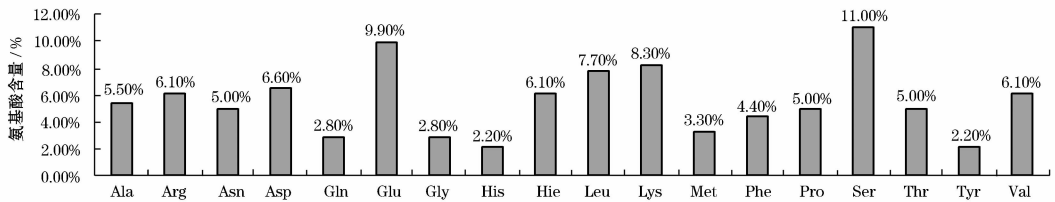


图1 家蝇 sHsp20.6 基因编码氨基酸组成

Fig. 1 Amino acid composition of sHsp20.6

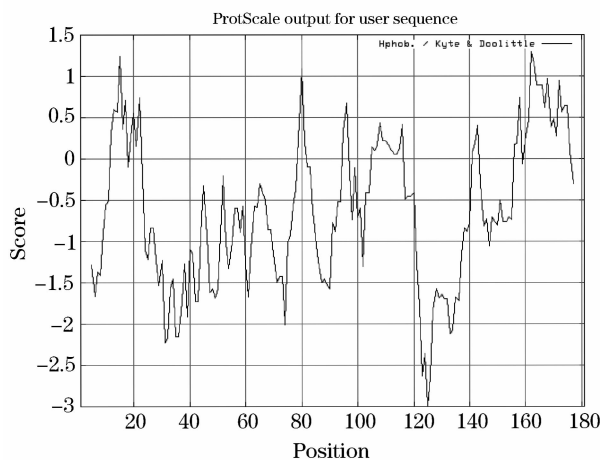


图2 家蝇 sHsp20.6 蛋白亲疏水值分析  
(参数选用 Hphob Kyte & Doolittle)

Fig. 2 Hydrophobicity analysis of sHsp20.6  
(parameter: Hphob Kyte & Doolittle)

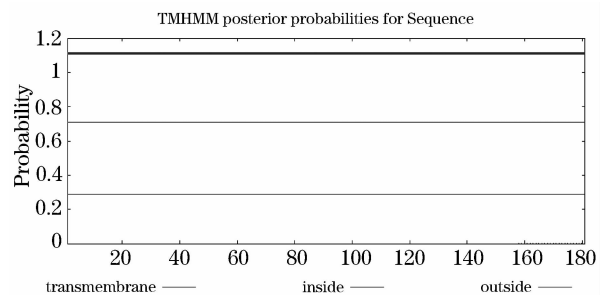


图3 家蝇 sHsp20.6 蛋白跨膜域分析

Fig. 3 Transmembrane region prediction of sHsp20.6

### 2.2 膜体分析、二级结构预测与分析

通过 PROSITE 软件的检索结果(图 5)显示该蛋白质具有一个 HSP20 家族的 motif,位于第 82 - 166 处。使用 SOPMA 软件对家蝇 sHsp 蛋白二级结构分析结果(图 6)显示,该蛋白质是由 33.70%的  $\alpha$  螺

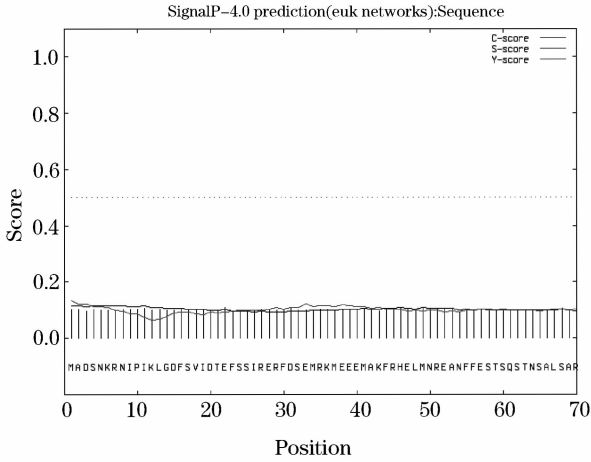


图 4 家蝇 sHsp20.6 蛋白信号肽分析  
Fig. 4 Signal peptide prediction of sHsp20.6

旋,16.57%的 $\beta$ 折叠,5.52% $\beta$ 转角和44.20%的无规则卷曲,其中 $\alpha$ 螺旋和无规则卷曲是其主要的构成原件,在20-57处构成了一个 $\alpha$ 螺旋的区域,其余的 $\alpha$ 螺旋和 $\beta$ 转角散布于蛋白质中。

结构功能域是蛋白质亚基结构中介于二级与三级结构之间的一种独立结构和功能单位。通过SMART预测sHsp20.6的结构功能域。结果如图7所示,通过图7可以看到家蝇sHsp氨基酸在85-179的区域有含一个HSP20家族的结构域,它是一个分子量在20kD左右的蛋白家族,作为伴侣它可以保护其他蛋白质在胁迫时尽可能的不发生变性;在22-55的区域含有一个卷曲螺旋区域,这与二级结构预测出的 $\alpha$ 螺旋区域所在的位置几乎一致。



82 - 166: score = 16.285  
 PLIQEDgdnKVLKLRFDVSQY APEEI VVK TVDQKLLVHAKHEEKSD---TKSVYRE---YNRE  
 FLLPKGVNPESIRSSLSkDGVLTVDAPL

图 5 家蝇 sHsp20.6 蛋白 motif 分析  
Fig. 5 Motif prediction of sHsp20.6

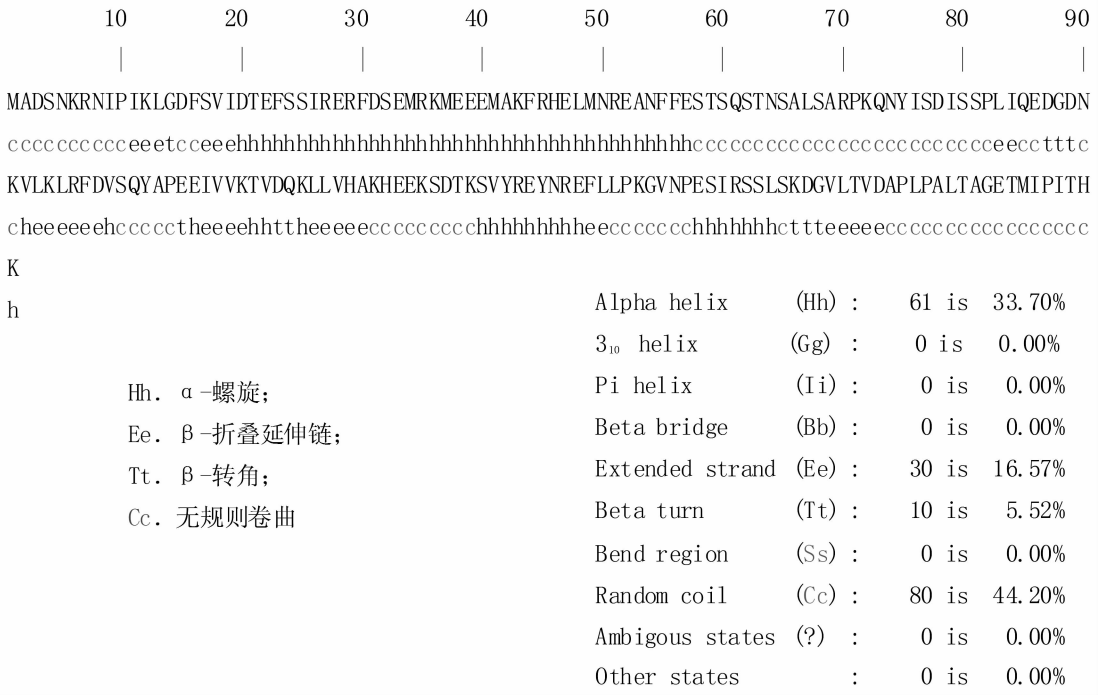


图 6 家蝇 sHsp20.6 氨基酸序列二级结构预测  
Fig. 6 Secondary structure prediction of sHsp20.6

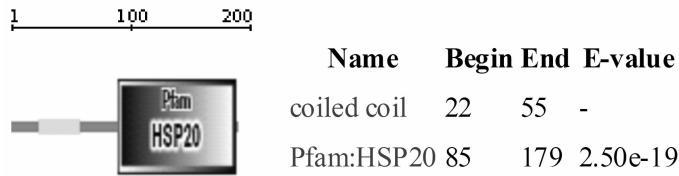


图 7 家蝇 sHsp20.6 氨基酸序列功能域分析

Fig. 7 Functional domain of sHsp20.6

通过 ProtFun 软件对家蝇 sHsp 蛋白功能分类进行预测,结果如图 8 所示,小热休克蛋白不具备酶的活性。功能类别(Functional category)显示该蛋白可能在翻译的时候发挥生物学功能,比较符合小热休克蛋白作为分子伴侣时所发挥生物学功能的作用。基因本体分类(Gene Ontology category)提示小热休克蛋白可能是一种生长因子,说明其在家蝇生长、发育和进化过程中起到一定的作用。

### 2.3 多重序列比对与系统发育树的构建

通过 MAGE4.1 软件构建的 NJ 树如图 9 所示。分析序列主要分为 5 个类群,其中用于构建系统发育树的 4 种昆虫物种的大部分小热休克蛋白各自分为一个类群,显示小热休克蛋白有物种特异性;而其中一个类群则包含了 4 种昆虫的其中一个小热休克蛋白,家蝇 sHsp20.6 蛋白也属于这一类。这与 Zi - Wen Li 等的研究结论相同,即昆虫中的小热休克蛋

白分为两类,一类是直系同源蛋白(即序列在进化的过程中没有发生太大的改变,较完整的保留了祖先蛋白质的序列),序列间具有高度同源性;另一类则具有物种特异性<sup>[5]</sup>。选取这些直系同源蛋白质的氨基酸序列通过 ClustalX2 软件进行多重序列比对,结果如图 10 所示。这些氨基酸序列具有高度的相似性,并且其中的某些区域的氨基酸序列完全一致。通过同源序列比对发现家蝇的小热休克蛋白与果蝇小热休克蛋白 20.8 之间只有 14 个氨基酸不同,并且在系统发育树上它们之间的进化距离也最近,同时也符合在昆虫分类学上的亲缘关系,都属于昆虫中小热休克蛋白的直系同源蛋白。当前,对昆虫中的直系同源小热休克蛋白的生物学研究较少,但推测这些蛋白应该具有重要的生物学功能,值得深入研究。

# Functional category	Prob	Odds	# Gene Ontology category	Prob	Odds
Amino_acid_biosynthesis	0.011	0.484	Signal_transducer	0.081	0.378
Biosynthesis_of_cofactors	0.041	0.574	Receptor	0.005	0.028
Cell_envelope	0.044	0.721	Hormone	0.002	0.329
Cellular_processes	0.037	0.507	Structural_protein	0.003	0.106
Central_intermediary_metabolism	0.054	0.857	Transporter	0.025	0.230
Energy_metabolism	0.027	0.296	Ion_channel	0.013	0.225
Fatty_acid_metabolism	0.016	1.265	Voltage-gated_ion_channel	0.005	0.239
Purines_and_pyrimidines	0.080	0.328	Cation_channel	0.010	0.220
Regulatory_functions	0.158	0.982	Transcription	0.124	0.967
Replication_and_transcription	0.247	0.921	Transcription_regulation	0.081	0.652
<b>Translation</b>	<b>=&gt; 0.122</b>	<b>2.781</b>	Stress_response	0.027	0.311
Transport_and_binding	0.023	0.055	Immune_response	0.012	0.142
# Enzyme/nonenzyme	Prob	Odds	<b>Growth_factor</b>	<b>=&gt; 0.048</b>	<b>3.424</b>
Enzyme	0.178	0.623	Metal_ion_transport	0.012	0.026
<b>Nonenzyme</b>	<b>=&gt; 0.822</b>	<b>1.151</b>			

图 8 家蝇 sHsp20.6 蛋白的功能分类

Fig. 8 Functional assortment of sHsp20.6



## 2.4 三级结构建模

通过同源建模的方式得到了 sHsp20.6 的三维预测模型(图 11),经过 RasMol 软件统计,该蛋白具有 9 个片层结构,4 个螺旋结构,17 个转角。可以明显的看到蛋白质成一个棒状结构,并能够清楚的区分出 N 端结构域和 C 端结构域。家蝇 sHsp20.6 有 7 个片层结构集中在 C 端结构域,它是所有小热休克蛋白家族成员都具有的保守结构,并且同  $\alpha$  晶体蛋白之间也具有很高的同源性,在同种物种之间该结构有很高的同源性,即使在不同的物种之间也有较高的同源性。小热休克蛋白以及  $\alpha$  晶体蛋白 C 端结构域具有大量片层结构也在其他文献中有过报道<sup>[4]</sup>。家蝇 sHsp20.6 的螺旋结构主要集中在 N 端结构域,N 端结构域与小热休克蛋白形成多聚复合物的调节有关,在同种物种之间具有一定的保守性,但在不同的物种之间没有保守性。

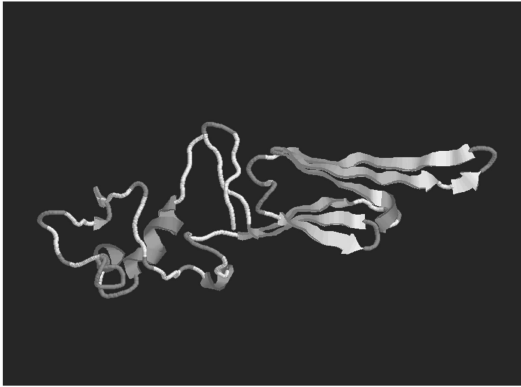


图 11 sHsp20.6 三维结构

Fig. 11 Three-dimensional structures of sHsp20.6

## 3 讨论

小热休克蛋白是热休克蛋白家族中的一个成员,其分子量是热休克蛋白家族中最小的。目前对于小热休克蛋白的研究相对较少,但它作为分子伴侣在生物体的生命活动过程中却发挥了相当大的作用。生物体在生命活动的大部分时期,小热休克蛋白都具有重要生物学功能。它能协助生物体内变性的蛋白质进行重新折叠<sup>[6]</sup>,诱导生物体产生耐热性、稳定细胞骨架<sup>[7]</sup>,在生物体的生长发育以及进化过程中均发挥了重要的作用。

通过分析发现该蛋白是一个分子量为 20.64kD 的亲水性蛋白,它不具备分泌信号肽,推测其合成以后可以直接在细胞内发生生物学功能。通过预测蛋白质的二级结构发现 sHsp20.6 一半的结构是由  $\alpha$  螺旋和  $\beta$  折叠构成。通过 PROSITE 软件预测显示

该蛋白具有一个 HSP20 家族的 motif,同时经过 SMART 预测该蛋白质在 85 - 179 的区域有含一个 HSP20 的结构域,这与 motif 的预测结果一致。对其蛋白功能的预测显示该蛋白质不具备酶的活性,但在其他蛋白质进行翻译时起到一定的生物学功能。

通过家蝇 sHsp20.6 与 4 种昆虫的系统发育树聚类构建分析发现该蛋白属于直系同源蛋白质(orthologous protein)。而直系同源蛋白质是在不同的物种之间来源于共同祖先的蛋白,他能很好的保留其主要序列以及结构,并具有共同或者相似的生物学功能。直系同源蛋白通常是生物体生命活动所必需的酶或关键性调控蛋白,它具有保守的生物学功能,在进化过程中非常缓慢、变化速度可以覆盖整个进化历史、序列变化速度与进化距离相当等特征。它代表了进化上出现分离的物种间保守的单拷贝基因,通常在物种进化过程中保持相同或是相似的功能<sup>[8-9]</sup>。本研究为以后更进一步研究 sHsp20.6 的生物学功能及与其他物种的直系同源小热休克蛋白之间的关系提供奠定基础。

## 参考文献(References)

- [1] F. Ritossa. A new puffing pattern induced by temperature shock and DNP in drosophila[J]. Cellular and Molecular Life Sciences, 1962,18(12):571-573.
- [2] Li-Hua Huang, Chen-Zhu Wang, Le Kang. Cloning and expression of five heat shock protein genes in relation to cold hardening and development in the leafminer, Liriomyza sativa[J]. Journal of Insect Physiology, 2009(55):279-285.
- [3] 刘红美,张洁,王赞,付萍,国果,张勇,吴建伟.热胁迫后家蝇幼虫 cDNA 文库构建与随机 EST 测序分析[J].免疫学杂志,2010,26(9):772-775.
- [4] 夏佳音,张耀洲.小热休克蛋白的结构和功能.中国生物化学与分子生物学报,2007,23(11):911-915.
- [5] Zi-Wen Li, Xue Li, Quan-You Yu, Zhong-Huai Xiang, Hirohisa Kishino, Ze Zhang. The small heat shock protein (sHSP) genes in the silkworm, Bombyx mori, and comparative analysis with other insect sHSP genes[J]. BMC Evolutionary Biology. 2009,9:215.
- [6] Martin Haslbeck. Recombinant expression and in vitro refolding of the yeast small heat shock protein Hsp42[J]. International Journal of Biological Macromolecules,2006,38(2):107-114.
- [7] Roy Quinlan. Cytoskeletal competence requires protein chaperones. Progress in Molecular and Subcellular Biology[J]. 2002,28:219-233.
- [8] Jayashree Balaji, Jonathan H Crouch, Prasad VNS Petite, David A Hoisington. A database of annotated tentative orthologs from crop abiotic stress transcripts. Bioinformatics,2006;1(6):225-227.
- [9] Konstantin V. Krutovsky, Christine G. Elsik, Marta Matvienko, Alex Kozik, David B. Neale. Conserved ortholog sets in forest trees. Tree Genet Genomes, 2006,3(1):61-70.