

# 家蚕硒蛋白组的计算机识别

陈平 段军 姜亮 刘琼 赵萍 夏庆友\* 徐辉碧

( 西南大学农业部蚕桑重点实验室, 重庆 400716; 深圳大学生命科学学院, 深圳 518060; 华中科技大学化学系, 武汉 430074. \* 联系人, E-mail: xiaqy@swau.cq.cn)

**摘要** 硒是一种生物必需微量元素, 主要以硒蛋白形式发挥生物学功能. 硒蛋白的生物合成取决于硒代半胱氨酸插入蛋白质的合成过程. TGA 码既是终止码, 又可翻译成硒代半胱氨酸, 这使普通基因注释软件无法正确预测硒蛋白, 导致现有数据库中许多物种的硒蛋白被错误注释或丢失. 本研究基于已公布的家蚕基因组预测信息, 采用 PERL 语言编程, 对家蚕基因组中硒蛋白进行了计算机检索与分析. 结果表明, 在家蚕数据库 18510 个已注释基因中, 以 TGA 码终止的基因有 6348 条, 其中含有 SECIS 结构的基因 249 条, 兼含半胱氨酸同源类似物的基因 52 条. 再经硒代半胱氨酸(Sec)侧翼序列比对, 最终检索到完全具备硒蛋白特点的基因 5 条, 其中谷胱甘肽硫转移酶(GST)是一种已知微生物硒蛋白, 而其他 4 种则是新硒蛋白, 分别在已有基因表中被注释为 CG6024 蛋白, CG5195 蛋白, ATP-结合盒转运蛋白 A 型(ABCA)和核 VCP 相似蛋白. 通过对 GST, ABCA 和 VCP 主要性质的分析, 推测家蚕硒蛋白在氧化调节、硒储存运输和细胞凋亡等过程中起重要作用.

**关键词** 蚕 硒 硒蛋白 基因组 硒代半胱氨酸

硒是人和动物必需的微量元素 [1]. 硒蛋白是硒在生物体内的主要存在形式, 也是硒发挥其生物功能的重要分子. 而其中作为活性中心的硒代半胱氨酸(Sec)是第 21 种氨基酸, 由基因编码区内 TGA 码翻译而成, 在蛋白质生物合成过程中掺入分子中 [2-6]. 而通常情况下, 密码子 TGA 作为终止码, 起到终止蛋白合成过程的作用. TGA 码的这种特殊解码方式需要一种特殊的 RNA 茎环结构——硒代半胱氨酸插入元件(SECIS)的指导. 真核生物和古细菌类硒蛋白 SECIS 元件位于其 mRNA 的 3'-非翻译区(3'-UTR), 而细菌类硒蛋白 SECIS 元件则紧邻于其 mRNA 中 UGA 密码子下游 [7-11]. 根据硒蛋白的特点可设计非常规基因预测方法以识别硒蛋白. 硒蛋白预测的难点在于 TGA 密码子即可能翻译成 Sec 又可能发挥终止码的作用, 而常规基因预测程序只将 TGA 解码为终止码, 造成目前众多基因组的注释工作均遗漏了对硒蛋白的注释 [12-19]. 所以现有真核生物和原核生物硒蛋白都是在基因组注释工作完成后独立进行的 [2,12,20-22].

家蚕被认为是继果蝇后第 2 种适用于遗传学研究的模式昆虫 [23-25]. 家蚕也可用作蛋白药物的生物反应器和生物材料来源 [24]. 已有研究表明, 从富硒蚕蛹中提取的氨基酸在抑制人肝癌细胞方面明显强于亚硒酸盐和硒代蛋氨酸, 由此推断家蚕体内可

能存在新硒蛋白 [26]. 随着家蚕基因组测序工作的完成和家蚕数据库的建立, 基因组范围内对家蚕基因的注释工作也已完成 [24,25,27-29]. 但由于 TGA 密码子的双重功能, 注释基因中目前没有硒蛋白. 本研究采用生物信息学方法对家蚕基因组进行重新分析预测, 找到 5 条家蚕硒蛋白基因.

## 1 材料和方法

( ) 序列数据库资源. 家蚕完整序列数据从日本和欧洲分子生物学实验室和 GenBank 数据库获取, 其序列号为 AADK00000000, 版本 AADK01-000000. 此外家蚕基因组数据还可从 SilkDB 获取, 网址为 <http://silkworm.genomics.org.cn> [25]. 基因组数据包括我们最新获得的 23155 个家蚕基因组支架(scaffolds)和 428.7 Mb ~6X 序列草图. 2005 年 8 月从 NCBI dbEST 数据库下载 115360 条家蚕 EST 序列, 并从该数据库 ftp 服务器下载 NCBI 非冗余蛋白数据库(NR). NR 数据库包含 1990024 条蛋白序列(667623348 个氨基酸残基). BLAST 程序(2.2.13 版本)同样从 NCBI ftp 服务器获取.

( ) 编程检索以 TGA 码终止的基因. 下载的 18518 条家蚕注释基因是由现有的基因预测软件 BGF 分析所得, 所使用的算法规则是经典基因预测算法 [24], 无法注释硒蛋白. 为此, 我们采用 PERL 语言编辑程序, 计算机检索上述数据库中的所有基因编码

区,寻找符合标准基因规则的以ATG作为起始码的基因编码区。随后对这些基因的终止码进行分析,显然其中以TGA作为终止信号的基因,可能是被错误识别的硒蛋白基因。于是将这些以TGA作为终止码的基因抽提出来,并对筛选出的基因从相应scaffold数据库中抽提出这些基因下游紧邻的1000个核苷酸,用以分析其中是否含有SECIS元件。

( ) SECIS结构的检索与序列分析。为了分析SECIS元件,用SECISearch (2.19 版本)程序在线检索以TGA码终止的家蚕基因的3'-UTR是否存在SECIS结构(图2),可通过网址<http://genome.unl.edu/SECISearch.html>访问并使用该程序的网络版本 [2,30]。这个程序由两个模块组成:第1个模块以SECIS元件的一、二级结构的保守区域作为筛选标准;第2个模块估算候选基因SECIS元件的自由能,分别估算假定硒蛋白SECIS元件中DNA双螺旋加上内环结构的自由能和DNA双螺旋加上顶环结构的自由能,并与标准自由能限制参数比较,筛选出符合条件的候选序列。其自由能限制参数是基于目前已知哺乳动物硒蛋白SECIS结构的最低自由能数据统计估算的。SECIS结构的一级保守序列采用NTGA\_AA\_GA, NTGA\_CC\_GA, 和NTGA\_AG\_GA (N代表任一核苷酸)3种模式。在使用SECISearch检索家蚕基因组硒蛋白的过程中,参数设置如下:核心结构能量参数为-5;整体结构能量参数为-11。

( ) 硒蛋白可读框的分析。取上步筛选得到的基因,逐条分析其起始码、可读框中的TGA码、终止码和SECIS结构。在将基因的TGA翻译为Sec后,编码区的终止码将顺延为TGA之后可读框内出现的第一个终止码。对含有多个框内TGA的基因,当TGA所对应的同源序列氨基酸不再是半胱氨酸(Cys)时,该TGA即为终止码。

( ) 编程检索具有Sec/Cys配对的同源序列。采用PERL语言,按如下步骤编写程序:(1)将上步得到的具有TGA码和SECIS结构的基因序列翻译成蛋白质的氨基酸序列;(2)将TGA翻译成Sec(简称为U);(3)将此氨基酸序列与NR库中的蛋白质序列采用BLASTx程序进行比对,阈值设为 $e < 1$ ;(4)用程序Clustal W分析和展示比对结果;(5)将比对结果中出现Cys/Sec(即U/C)配对的同源序列抽提出来,分析U两侧的氨基酸序列同源性,同源性比对优异者为硒蛋白。

## 2 结果

### 2.1 使用TGA作为终止码的家蚕基因

用BGI预测并注释的基因数为18510条。然而该程序所使用的预测算法依然是经典基因预测方法,密码子TGA只解码为终止信号,所预测结果中不包含任何硒蛋白基因 [24]。图1列出家蚕硒蛋白的识别步骤和各步结果。在BGI所预测的基因中6348条以TGA码终止,其中3315条基因来自正义链,3303条来自负义链。从家蚕数据库中抽提出这些以TGA终止的基因及可读框内TGA的下游含1000 bp核苷酸的序列。

### 2.2 家蚕硒蛋白SECIS的结构特征

用SECISearch在线检索家蚕基因可读框内TGA的下游序列。如图1所示,249条以TGA码终止的基因的下游序列中含有SECIS元件。其中122条基因来自正义链,127条来自负义链。最终检索到5个SECIS元件能完全满足硒蛋白基因的各项特征。图2列出这5个SECIS元件的保守序列和结构特征,显示了家蚕SECIS元件与哺乳动物硒蛋白SECIS元件之间的相似性。在这5条基因中,2条的一级结构模式为ATGA\_AA\_GA,另2条为ATGA\_CC\_GA,剩余1条为TTGA\_AG\_GA。

图3展示了这些SECIS元件的高级结构。可以看到,这些SECIS结构都含有2个DNA双螺旋、2个环(内环和顶环)以及一个SECIS核心结构(非Watson-Crick四碱基聚合物)。其中两个SECIS元件(即Bmb007600和Bmb018303)在顶环上还分别附加微型茎环结构。图4给出SECIS元件的3种模式:模式a与绝大多数哺乳动物SECIS元件相同;模式b与两种哺乳动物硒蛋白SeIM和SeIO相似;模式c与硫氧还原蛋白还原酶2(TR2)相似。家蚕硒蛋白结构内TGA密码子与SECIS元件之间的距离与大多数哺乳动物硒蛋白类似,大于204 nt。

### 2.3 家蚕硒蛋白的含Cys同源序列

对于同时含有框内TGA和SECIS结构的家蚕基因,采用BLASTx进行同源序列比对,以检索它们是否具有含Cys的同源序列及是否含TGA/Cys对。从图1看到,在249条同时具有框内TGA和SECIS元件的基因中,98条含TGA/Cys对,同时产生1374对硒蛋白候选基因与Cys同源序列的一一对应序列。对于上述比对结果,我们根据以下规则筛选家蚕硒蛋

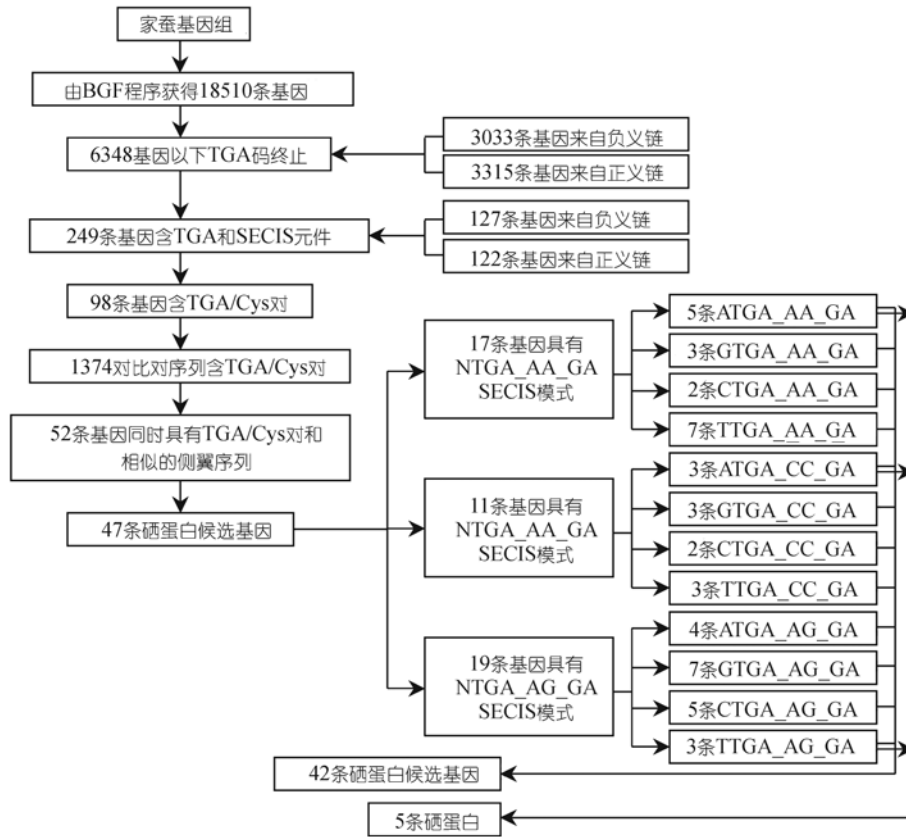


图 1 家蚕基因组中硒蛋白组织识别过程图

	螺旋 I	四聚体	螺旋 II	顶 环		螺旋 II	四聚体	螺旋 I	
>Bmb007600	TTTAAACA	ATACGAAGTGC	<b>ATGAA</b>	TGATGACGATC	AA	ATATGGTGTAGTTTAAACAACAAC	GTCGGTCTCA	CGAG AATCGAGC	GCTTACA
>Bmb001751	TAAATTAT	ACACAATCCGAGA	<b>ATGAG</b>	TTCTTGGAAAGCG	AA	CCCGTTA	ATATCCGAGGC	CGAT AGITTAACC	GGATGT
>Bmb016238	TACCGTG	GTCCACAGGCATC	<b>ATGAA</b>	GTGAGCCTCT	CC	GCCCTAATGAAGTC	GGAGTCTCGA	TCAC CTCGAATCTC	CCAAGTG
>Bmb018303	GACCTCA	TGTCTCAAGGCGT	<b>ATGAT</b>	GCTGTGATGT	CC	ATGGGCTCCCGTAGCCCT	TAACACCAGG	TCAC TAA	TGGAGAT
>Bmb014121	CAGGACT	ATTTAATGCTGGG	<b>TTGAC</b>	AAGCTTACAGC	AG	CTCGAACACATCACTAAGT	CTATGTGAGTT	GGAA TTCCTAAC	AGTAATT

图 2 家蚕硒蛋白基因 SECIS 元件的序列比对严格保守核苷酸在图中用黑体标出

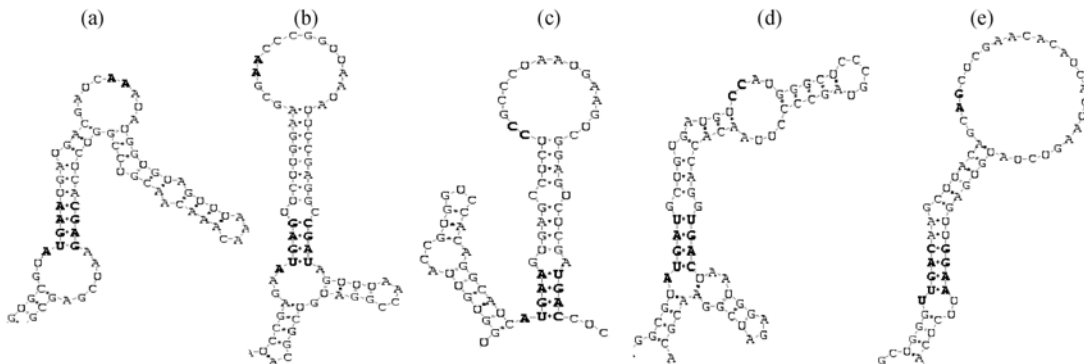


图 3 家蚕硒蛋白基因中 SECIS 元件的保守核苷酸及其高级结构

保守序列中的保守结构用黑体标出. (a) Bmb007600; (b) Bmb001751; (c) Bmb016238; (d) Bmb018303; (e) Bmb014121

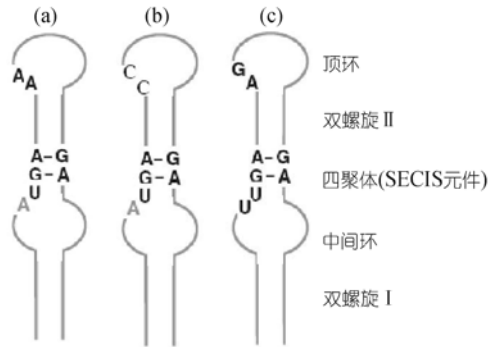


图 4 家蚕硒蛋白基因 SECIS 元件的保守核苷酸和结构特点  
(a) AUGA\_AA\_GA; (b) AUGA\_CC\_GA; (c) UUGA\_AG\_GA

白基因、剔除假阳性结果: ( ) 只有第一个 U 与 C 成对的序列被保留作为硒蛋白的候选基因, 第一个 U 如没有与 C 成对, 则该序列将被舍弃; ( ) 在 U/C 对的两侧序列一定要有较好的同源相似性, 否则也将被舍弃. 如此, 从这 98 条基因中筛选出满足以上两条规则的 52 条基因, 在这 52 条基因中不存在可读框内多于一个编码 Sec 的 TGA 的基因.

### 2.4 家蚕硒蛋白基因可读框的分析

对上述 52 条基因进行可读框的分析, 发现其中 5 条基因的 SECIS 结构的位置是错误的: SECIS 结构出现在 ORF 框内或跨 ORF 与 3'-UTR 区域, 这些基因被舍弃. 终止码出现在 SECIS 结构内部或前面的该序列也被剔除. 分析余下基因的 SECIS 结构, 只有 14 条基因的保守模式与已知硒蛋白基因的保守模式相同或相似. 其中, 5 条为 ATGA\_AA\_GA, 3 条为 G-TGA\_AA\_GA, 3 条为 ATGA\_CC\_GA, 3 条为 TTGA\_A\_AG\_GA. 而这 14 条基因中, 只有 5 条具有多于 2 个的含 Cys 的同源序列. 这 5 条基因具有已知硒蛋白的所有结构特征. 图 5 显示这 5 条基因的含 Cys 的同源基因比对结果. 表 1 中列出这 5 条基因的注释信息, 其中只有谷胱甘肽硫转移酶(GST)已报道为一种微生物硒蛋白 [21], 其他 4 种蛋白在其他物种中均未发现

表 1 家蚕硒蛋白基因的注释

基因序号	SECIS 模式	注释
Bmb007600	ATGA_AA_GA	谷胱甘肽硫转移酶(已知硒蛋白)
Bmb001751	ATGA_AA_GA	CG6024 蛋白
Bmb016238	ATGA_CC_GA	CG5195 蛋白
Bmb018303 ATP	ATGA_CC_GA (ABCA)	ATP-结合盒转运蛋白 A 型 (ABCA)
Bmb014121	TTGA_AG_GA	核 VCP 相似蛋白

为硒蛋白.

### 3 讨论

家蚕硒蛋白基因识别过程中采用了 3 条基本规则: ( ) 在 3'-UTR 一定含有 SECIS 结构; ( ) 可读框内部含有 TGA 密码子; ( ) 绝大多数已知硒蛋白基因具有含 Cys 的同源序列, 在该同源序列中硒蛋白的 Sec 所对应氨基酸残基为 Cys. 硒蛋白基因的这些特征为其基因注释工作带来了困难, 但也为硒蛋白的识别提供了可能性. 到目前为止, 已经有数种新的硒蛋白计算机识别方法. 一种方法是基于对 SECIS 元件的检索和结构可读框内 TGA 密码子的分析 [2,22]. 该方法在应用时需要针对不同物种进行计算机程序培训. 例如, 在使用该程序预测果蝇硒蛋白的工作前, 必须通过对程序进行培训以得到果蝇的相应工作参数, 这些参数只适应于果蝇, 对其他物种包括家蚕在内均不适用. 而程序培训过程含有对大量数据的收集和分析工作、以及对众多程序的修改编辑工作, 费时且困难. 因此这类方法无法广泛适用于所有物种. 另外一种硒蛋白预测方法基于以下共识, 即绝大多数硒蛋白具有含 Sec 残基或 Cys 残基的同源序列, 对这些同源序列比对结果进行分析时可知, 硒蛋白 Sec 残基在同源蛋白的相应位置被 Cys 所取代 [21]. 根据硒蛋白的这种性质, 可以通过同源性比对检索候选蛋白基因 TGA 密码子对应于同源蛋白相应位置上的氨基酸残基, 假如是 Sec 或 Cys 残基, 则候选蛋白很可能是硒蛋白, 所产生的 Sec/Sec (TGA/TGA) 或 Sec/Cys (TGA/Cys) 配对表明, 该 TGA 密码子有编码为 Sec 的可能 [10]. 这类方法不考虑所预测物种的特异性, 但在预测过程中需处理极大量数据, 所以实施过程需要超级计算机硬件支持.

本研究采用了一种新的从基因组中识别硒蛋白的方法. 它既具有物种特异性也具有广泛使用的可能性. 此外, 它不需要过多的数据处理. 但该方法是基于基因组数据库中已注释基因的正确性的基础上的一种硒蛋白识别方法. 对于家蚕, 现有注释基因 18510 条, 其中没有硒蛋白 [24]. 为了弥补这种不足, 我们拓展了上述方法以检索全部家蚕注释基因. 首先, 找出所有以 TGA 作为终止密码子的基因, 这个 TGA 密码子可能解码为 Sec 残基或终止密码子; 其次, 检索可读框内 TGA 后和 ORF 下游 3'-UTR 中的 SECIS 元件; 最后, 进行含 TGA/Cys 对同源序列比对分析, 以确定硒蛋白的存在. 采用这种方法, 本研



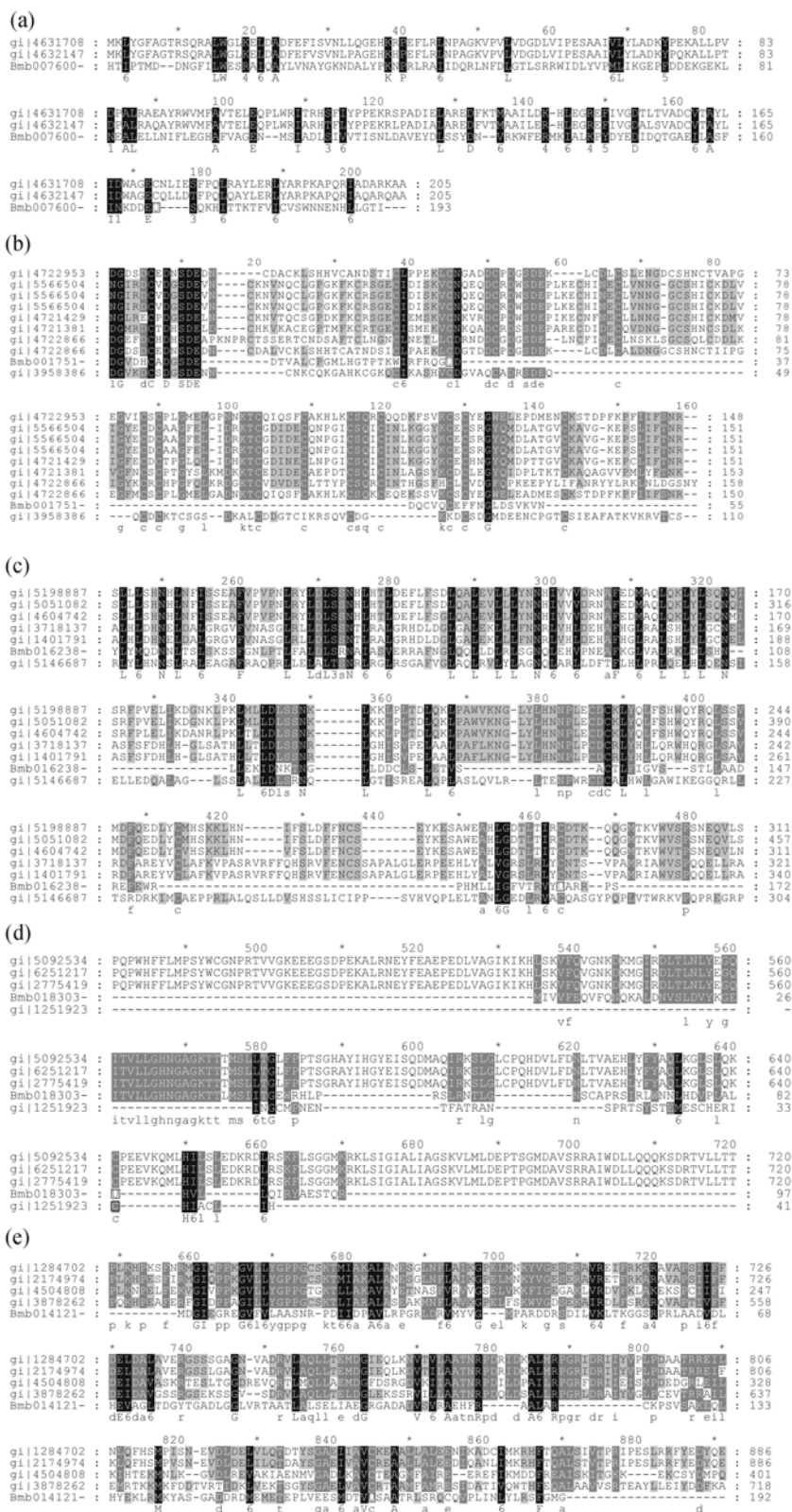


图 5 家蚕硒蛋白与其含 Cys 的同源蛋白的多重序列比对

保守氨基酸残基用黑体标出。方框残基为预测的硒代半胱氨酸 Sec (U) 残基, 它与其同源蛋白中的半胱氨酸 Cys (C) 残基型匹配。各家蚕硒蛋白分别由基因组中以下基因翻译而得: (a) Bmb007600; (b) Bmb001751; (c) Bmb016238; (d) Bmb018303; (e) Bmb014121

究预测到5条家蚕硒蛋白基因。其中GST已被报道为一种微生物蛋白<sup>[21]</sup>。其他4条基因在现有的基因库中被注释为CG6024蛋白、CG5195蛋白、ATP-结合盒转运蛋白A型(ABCA)和核VCP相似蛋白。到目前为止,这4种蛋白还没有在其他物种中被报道为硒蛋白。CG6024和CG5195蛋白的生物学功能目前尚不清楚,而ABCA和VCP相似蛋白的功能正在研究之中<sup>[31-37]</sup>。

ABCA位于质膜中,在肝、巨噬细胞、肺、胎盘、肠、胃和单核细胞中都有表达,它的另外一个名字是胆固醇外流调节蛋白,与家族性脂蛋白缺乏病有关。这是一种高保守性蛋白家族,它们参与许多酶作用底物如离子、氨基酸、肽链、糖、维生素和类固醇激素等的跨膜运输<sup>[33-34]</sup>。VCP是与细胞活性相关的ATP酶(AAA)超级家族的一个成员。它与核因子- $\kappa$ B(NF $\kappa$ B)的抑制物——抑制因子 $\kappa$ B $\alpha$ (I $\kappa$ B $\alpha$ )的蛋白酶降解途径有关<sup>[35,36]</sup>。转染VCP的细胞株出现稳定的NF $\kappa$ B激活和磷酸化I $\kappa$ B $\alpha$ 的快速降解<sup>[37]</sup>。尽管ABCA和VCP在其他物种中并不是硒蛋白,但是他们的生物学功能与几种已知硒蛋白密切相关。硒蛋白P(SeI $P$ )是哺乳动物中的一种硒储存和转载蛋白,在家蚕硒蛋白基因组中我们没有发现这种蛋白,因此,ABCA在家蚕基因组中可能承担SeI $P$ 的功能,作为一种硒储存和转载蛋白。长期以来,大量研究表明,硒通过调节氧化应激下而作用于NF $\kappa$ B。家蚕VCP被发现为硒蛋白是对这种观点的有力支持。其他物种中的VCP尽管不是硒蛋白,但它们是含Cys的蛋白,其Cys可通过氧化还原反应与硒化合物作用,以此调节NF $\kappa$ B的活性。

我们以前的研究表明,从富硒蚕蛹中提取的氨基酸能够抑制人肝癌细胞SMC-7721的生长,其抑制作用强于亚硒酸钠和硒代蛋氨酸,而普通蚕蛹氨基酸却没有这种功能,这暗示富硒蚕蛹中存在新硒蛋白<sup>[26]</sup>。本研究发现的5种家蚕硒蛋白不但是对前面工作的支持。ABCA已被报道为家族性高密度脂蛋白HDL缺乏症<sup>[33]</sup>和丹吉尔病(一种家族性疾病,表现为血清中缺乏高密度脂蛋白、扁桃体及其他组织中有胆固醇酯沉积)<sup>[38,39]</sup>的一种致病基因,在细胞中具有输出含胆固醇和胆碱的磷脂的功能<sup>[40]</sup>,它们也对Stargardt病以及多种退行性和炎症型视网膜疾病的发病机制起关键作用<sup>[41-43]</sup>。VCP表达水平和肝癌<sup>[44]</sup>、胃癌<sup>[45]</sup>和前列腺癌<sup>[46]</sup>的发病率和存活时间有着

一定关系。Se被报道在预防动脉硬化和癌症中起重要作用,家蚕硒蛋白ABCA及VCP类似蛋白的发现不仅显示硒在调节胆固醇通量和细胞凋亡中起到重要作用,而且也暗示在人类ABCA和VCP蛋白中,可以通过用Se取代其中的S来加强人类对这些疾病的预防能力。

致谢 本工作为国家自然科学基金(批准号:30370352和30570420)以及国家重点基础研究发展计划(批准号:2005CB121000)资助项目。

### 参 考 文 献

- 1 Hatfield D L. Selenium: Its molecular biology and role in human health. Boston: Kluwer Academic Publishers, 2001. 313—319
- 2 Kryukov G V, Castellano S, Novoselov S V, et al. Characterization of mammalian selenoproteomes. *Science*, 2003, 300(30): 1439—1443[DOI]
- 3 Atkins J F, Gesteland R F. The twenty-first amino acid. *Nature*, 2000, 407(6803): 463—465[DOI]
- 4 Bock A. Biosynthesis of selenoproteins: An overview. *Biofactors*, 2000, 11(1-2): 77—78
- 5 Hatfield D L, Gladyshev V N. How selenium has altered our understanding of the genetic code. *Mol Cell Biol*, 2002, 22(11): 3565—3576[DOI]
- 6 Driscoll D M, Copeland P R. Mechanism and regulation of selenoprotein synthesis. *Annu Rev Nutr*, 2003, 23: 17—40[DOI]
- 7 Berry M J, Banu L, Chen Y Y, et al. Recognition of UGA as a selenocysteine codon in type deiodinase requires sequences in the 3' untranslated region. *Nature*, 1991, 353(6341): 273—276[DOI]
- 8 Walczak R, Westhof E, Carbon P, et al. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA*, 1996, 2(4): 367—379
- 9 Tujebajeva R M, Copeland P R, Xu X M, et al. Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep*, 2000, 1(2): 158—163[DOI]
- 10 Zhang Y, Gladyshev V N. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. *Bioinformatics*, 2005, 21(11): 2580—2589[DOI]
- 11 徐辉碧, 黄开勤, 瞿祥虎, 等. 采用 RNA Draw 程序识别真核生物硒蛋白基因的 SECIS 结构. *科学通报*, 2001, 46(7): 556—558
- 12 Castellano S, Morozova N, Morey M, et al. In silico identification of novel selenoproteins in the *Drosophila melanogaster* genome. *EMBO Rep*, 2001, 2(8): 697—702[DOI]
- 13 Burge C B, Karlin S. Finding the genes in genomic DNA. *Curr Opin Struct Biol*, 1998, 8(3): 346—353 [DOI]
- 14 Haussler D. Computational gene finding. *Trends Biochem Sci*, Supplementary Guide to Bioinformatics, 1998. 12—15
- 15 Cataldo L, Baig K, Oko R, et al. Developmental expression, intracellular localization, and selenium content of the cysteine-rich protein associated with the mitochondrial capsules of mouse sperm. *Mol Reprod Dev*, 1996, 45(3): 320—331 [DOI]
- 16 Gladyshev V N, Jeang K T, Stadtman T C. Selenocysteine, identified as the penultimate C-terminal residue in human T-cell thiore-

- doxin reductase, corresponds to TGA in the human placental gene. Proc Natl Acad Sci USA, 1996, 93(12): 6146—6151[DOI]
- 17 Guimaraes M J, Peterson D, Vicari A, et al. Identification of a novel selD homolog from eukaryotes, bacteria, and archaea: Is there an autoregulatory mechanism in selenocysteine metabolism? Proc Natl Acad Sci USA, 1996, 93(26): 15086—15091[DOI]
- 18 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. Nature, 2001, 409(6822): 860—921[DOI]
- 19 Venter J C, Adams M D, Myers E W, et al. The sequence of the human genome. Science, 2001, 291(5507): 1304—1351[DOI]
- 20 Taskov K, Chapple C, Kryukov G V, et al. Nematode selenoproteome: The use of the selenocysteine insertion system to decode one codon in an animal genome? Nucleic Acids Res, 2005, 33(7): 2227—2238[DOI]
- 21 Zhang Y, Fomenko D E, Gladyshev V N. The microbial selenoproteome of the Sargasso Sea, Genome Biol, 2005, 6(4): R37[DOI]
- 22 Kryukov G V, Gladyshev V N. The prokaryotic selenoproteome. EMBO Rep, 2004, 5(5): 538—543[DOI]
- 23 Adams M D, Celniker S E, Holt R A, et al. Finishing a whole-genome shotgun: Release 3 of the *Drosophila* euchromatic genome sequence. Science, 2000, 287(5461): 2185—2195[DOI]
- 24 Xia Q Y, Zhou Z Y, Lu C, et al. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science, 2004, 306(10): 1937—1940
- 25 Wang J, Xia Q Y, He X M, et al. SilkDB: A knowledgebase for silkworm biology and genomics. Nucleic Acids Res, 2005, 33(Database issue): D399—D402[DOI]
- 26 Hu D C, Liu Q, Cui H R, et al. Effects of amino acids from selenium-rich silkworm pupas on human hepatoma cells. Life Sci, 2005, 77(17): 2098—2110[DOI]
- 27 Zha X F, Xia Q Y, Zhao P, et al. Detection and analysis of alternative splicing in the silkworm by aligning expressed sequence tags with the genomic sequence. Insect Mol Biol, 2005, 14 (2): 113—119[DOI]
- 28 Wang Z L, Ji J, Xia Q Y, et al. Identification and expression pattern of Bmlark, a homolog of the *Drosophila* gene lark in *Bombyx mori*. DNA Seq, 2005, 16(3): 224—229
- 29 Cheng T C, Xia Q Y, Qian J F, et al. Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain Dazao. Insect Biochem Mol Biol, 2004, 34(6): 523—530[DOI]
- 30 Kryukov G V, Kryukov V M, Gladyshev V N. New mammalian selenocysteine-containing proteins identified with an algorithm that searches for selenocysteine insertion sequence elements. J Biol Chem, 1999, 274(48): 33888—33897[DOI]
- 31 Higgins C F. ABC transporters: From microorganisms to man. Annu Rev Cell Biol, 1992, 8: 67—113[DOI]
- 32 Kaminski W E, Orso E, Diederich W, et al. Identification of a novel human sterol-sensitive atp-binding cassette transporter (ABCA7). Biochem Biophys Res Commun, 2000, 273(2): 532—538[DOI]
- 33 Langmann T, Klucken J, Reil M, et al. Molecular cloning of the human ATP-binding cassette transporter 1 (hABC1): Evidence for sterol-dependent regulation in macrophages. Biochem Biophys Res Commun, 1999, 257(1): 29—33[DOI]
- 34 Klucken J, Buchler C, Orso E, et al. ABCG1 (ABC8), the human homolog of the *Drosophila* white gene, is a regulator of macrophage cholesterol and phospholipid transport. Proc Natl Acad Sci USA, 2000, 97(2): 817—822[DOI]
- 35 Dai R M, Chen E, Longo D L, et al. Involvement of valosin-containing protein, an ATPase co-purified with Ikappa-Balpha and 26 S proteasome, in ubiquitin-proteasome-mediated degradation of Ikappa-Balpha. J Biol Chem, 1998, 273(6): 3562—3573[DOI]
- 36 Yamamoto S, Tomita Y, Hoshida Y, et al. Expression level of valosin-containing protein (VCP) as a prognostic marker for gingival squamous cell carcinoma. Ann Oncol, 2004, 15(9): 1432—1438 [DOI]
- 37 Asai T, Tomita Y, Nakatsuka S et al. VCP (p97) regulates NfkappaB signaling pathway, which is important for metastasis of osteosarcoma cell line. Jpn J Cancer Res, 2002, 93(3): 296—304
- 38 Brooks-Wilson A, Marcil M, Clee S M, et al. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. Nat Genet, 1999, 22(4): 336—345[DOI]
- 39 Rust S, Rosier M, Funke H, et al. Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. Nat Genet, 1999, 22(4): 352—355[DOI]
- 40 Orso E, Broccardo C, Kaminski W E, et al. Transport of lipids from Golgi to plasma membrane is defective in Tangier disease patients and Abc1 deficient mice. Nat Genet, 2000, 24(2): 192—196[DOI]
- 41 Allikmets R, Singh N, Sun H, et al. A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. Nat Genet, 1997, 15(3): 236—246[DOI]
- 42 Allikmets R, Shroyer N F, Singh N, et al. Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. Science, 1997, 277(5333): 1805—1807[DOI]
- 43 Martinez-Mir A, Paloma E, Allikmets R, et al. Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. Nat Genet, 1998, 18(1): 11—12[DOI]
- 44 Yamamoto S, Tomita Y, Nakamori S et al. Elevated expression of valosin-containing protein (p97) in hepatocellular carcinoma is correlated with increased incidence of tumor recurrence. J Clin Oncol, 2003, 21(3): 447—452[DOI]
- 45 Yamamoto S, Tomita Y, Hoshida Y, et al. Expression level of valosin-containing protein is strongly associated with progression and prognosis of gastric carcinoma. J Clin Oncol, 2003, 21(13): 2537—2544[DOI]
- 46 Tsujimoto Y, Tomita Y, Hoshida Y et al. Elevated expression of Valosin-containing protein (p97) is associated with poor prognosis of prostate cancer. Clin Cancer Res, 2004, 10(9): 3007—3012[DOI]

(2006-05-09 收稿, 2006-09-06 接受)