

疟蚊基因组中新硒蛋白的计算机识别^{*}

刘琼^{①**} 姜亮^{①②} 陈平^③ 高中洪^③ 徐辉碧^③

(① 深圳大学生命科学学院, 深圳 518060; ② 中国科学院长春应用化学研究所, 长春 130022; ③ 华中科技大学化学系, 武汉 430074)

摘要 硒蛋白的生物合成取决于硒代半胱氨酸插入蛋白质的过程. 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疟蚊基因组预测信息、采用 PERL 语言编程, 对疟蚊基因组中的硒蛋白进行了计算机检索与分析. 结果表明: 以 TGA 码终止的基因有 11365 条, 其中具有 SECIS 结构的基因有 918 条, 同时具有半胱氨酸同源类似物的基因 58 条. 再经 Sec 侧翼序列比对, 最终检索到具有硒蛋白全部特点的基因 7 条. 从硒蛋白的基本生物功能推测, 冈比亚疟蚊基因组中存在的新硒蛋白可能与疟蚊的氧化耐受特性及其相关蛋白的调控相关联. 因此, 研究疟蚊硒蛋白将为干涉疟蚊带菌能力、从蚊媒传播途径防止疟疾提供理论基础.

关键词 硒蛋白 硒代半胱氨酸 冈比亚疟蚊 基因组 生物信息学

硒是人和动物必需的微量元素^[1]. 它在生物体内的生理功能主要是抗氧化作用、调节甲状腺激素的分泌和参与细胞增殖与凋亡. 硒蛋白是硒存在于动物体内的主要形式, 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, 它由基因中编码区内的 TGA 码翻译而成, 在蛋白质生物合成过程中掺入分子中^[2-5]. 因此, 硒蛋白的预测决定于对 Sec 的识别与鉴定.

硒蛋白基因具有明显的结构特征, 可用以下标准来识别它们. 第一, 硒蛋白基因的 3'-非翻译区(3'-UTR)存在硒代半胱氨酸插入序列, 即 SECIS 结构;

第二, 硒蛋白基因的开放阅读框(ORF)中包含有 TGA 密码子; 第三, 对已知硒蛋白的分析表明, 它们均具有以半胱氨酸取代硒代半胱氨酸的同源类似物. 在硒蛋白与其同源蛋白的序列比对中, 对应于硒代半胱氨酸位置的氨基酸残基为半胱氨酸, 即同源蛋白之间存在 Sec/Cys 对应关系.

从基因组序列预测硒蛋白是一项很困难的工作, 因为 TGA 密码子同时具有终止码和翻译为 Sec 的双重功能^[6]. 目前通用的基因注释软件中, TAA, TAG 和 TGA 都只作为开放阅读框(ORF)的终止码来识别^[7,8], 因此, 在用这些程序进行基因预测时, TGA 只作为终

收稿日期: 2006-04-22; 接受日期: 2006-06-09

* 国家自然科学基金资助项目(批准号: 30370352, 30570420)

** E-mail: liuqiong@szu.edu.cn

止码, 而不翻译成Sec, 因此, 这些程序均不能正确识别硒蛋白^[2]. 原核生物和真核生物中的绝大多数硒蛋白都是在基因组注释完成后, 通过其它软件和方法发现的.

冈比亚疟蚊基因组的测序工作于 2002 年完成, Li Jun 等人在 2005 年联合使用从头预测法和比较分析预测法对疟蚊基因组做了一次重新预测, 由此得到了准确率更高的全新疟蚊 cDNA 数据库, 含有 31254 条单独基因^[9,10], 但这些序列的预测和注释过程同样没有考虑到硒蛋白合成机制的特殊性, 以至于不包含硒蛋白. 硫氧还蛋白还原酶(TrxR)在哺乳动物中是一种硒蛋白, 而在原核生物中则不是. 疟蚊基因组注释中的TrxR类似于原核生物TrxR, 也不是硒蛋白, 其中Sec为Cys所取代. 由此我们推断疟蚊基因组中可能含有其它硒蛋白^[11].

硒蛋白的基本生物功能是参与调节体内氧化还原水平. 因此, 研究冈比亚疟蚊硒蛋白有助于认识疟蚊的氧化耐受特性及其调控机理, 为防治疟蚊等蚊虫传染性疾​​病提供一定的理论依据. 本文在已注释基因的基础上, 使用 PERL 语言自编程序, 结合使用现有的基因分析软件, 对疟蚊数据库进行处理筛选, 以期识别疟蚊基因组中的硒蛋白基因并推断其生物功能, 探索从传染源途径阻断疟疾传播的可能性.

1 材料和方法

1.1 序列数据库和信息资源

为了从疟蚊现有基因数据中检索可能存在的硒蛋白, 我们从NCBI数据库和疟蚊基因组注释者数据库下载了疟蚊现有的最新全基因数据. 最新cDNA序列包含 31254 条单独基因, 全部由从头基因预测法结合比较基因预测法得到. DNA数据包括疟蚊染色体的全长DNA数据, 数据容量 407 M^[9]. 数据更新日期: 2006 年 1 月 5~6 日. 网址: <ftp://ftp.ncbi.nih.gov/pub/omes/ReAnoXcel>, <http://vlab.umn.edu/download/GenomeAnnotation/>.

NCBI 无冗余蛋白值数据库(NR 库)从 NCBI ftp 服务器下载, 网址: <ftp://ftp.ncbi.nih.gov/blast/db/>, 该蛋白质数据库包含 2868050 条无重复蛋白序列(下载日期 2006 年 4 月).

Blast 程序从数据库 NCBI ftp 服务器下载, 网址: <ftp://ftp.ncbi.nih.gov/blast/db/>. 版本号: BLASTP 2.2.13 [Nov-27-2005].

1.2 编程检索以 TGA 码终止的基因

疟蚊基因组 cDNA 数据库包含 31254 条单独基因, 这些基因都是使用现有基因预测软件预测获得, 使用的预测算法规则没有考虑到 TGA 翻译成 Sec 的可能性, 以至于没有发现硒蛋白. 为了从疟蚊现有的数据库中挖掘硒蛋白, 我们采用 PERL 语言编辑程序, 用以检索上述数据库中以 TGA 终止的全部基因. 然后从相应 DNA 染色体数据库中抽提出 TGA 终止基因下游紧邻的 999 个核苷酸序列, 并分析是否含有 SECIS 元件.

1.3 SECIS 结构的检索与序列分析

为了分析SECIS元件的存在与否, 通过以下网址 <http://genome.unl.edu/SECISearch.html>^[12] 访问, 并在线使用SECISearch (version 2.19)程序检索以 TGA终止的疟蚊基因的 3'-非翻译区是否存在SECIS结构, 这个程序由两个模块组成, 第一个模块以SECIS元件的一、二级结构的保守区域作为筛选标准. 第二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元件中螺旋 I 加上内环结构的自由能和螺旋 II 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^[2].

绝大多数哺乳动物基因组中, 硒蛋白SECIS结构与终止码之间的距离在 500 bp以内^[6], 由此我们将疟蚊所有以TGA作为终止码结尾的cDNA序列后 999 个碱基作为输入数据, 使用SECISearch程序, 采取最宽松的限制条件检索SECIS元件, 筛选出同时具有TGA码和 3'-SECIS结构的基因. 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. 列出SECIS结构的一级保守序列, 保留含有类似于已知硒蛋白SECIS结构的一级保守序列类型的基因, 舍弃含其余SECIS结构序列的基因.

1.4 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: (i) 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; (ii) 将 TGA 翻译成 Sec (简称为 U); (iii) 找到每条蛋白序列中第一个出现的 U, 以此位置为中心, 分别保留上下游各 166 个氨基酸, 去掉其余氨基酸, 获得 U 两侧的氨基酸序列片段; (iv) 将此氨基酸序列与 NR 库中的蛋白质序列采用 BLASTP 软件进行比对, 阈值设为 1; (v) 使用程序 Clustal W 分析和展示比对结果; (vi) 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, 取同源性比对优异者做进一步分析。

1.5 硒蛋白开放阅读框的分析

取上步筛选得到的基因, 逐条分析其起始码、阅读框中的 TGA 码、终止码和 SECIS 结构。将基因的 TGA 翻译为 Sec 后, 编码区的终止码将顺延为 TGA 后出现的第一个框内终止码。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 Cys 时, 该 TGA 即为终止码。SECIS 结构出现在 ORF 框内, 或跨 ORF 与 3'-UTR 时, 该基因舍弃。如果终止码出现在 SECIS 结构内部或前面, 该序列将被剔除。作图标记所获得的硒蛋白基因的 ORF 框、框内 TGA 和 3'-UTR 中的 SECIS 元件。

2 结果

2.1 疟蚊硒蛋白基因的识别

疟蚊基因由 GeneFinder, GeneScan 等软件预测得到, 共有 31254 条。由于软件算法的限制, 上述基因中没有符合硒蛋白特点的基因。为了弥补这个缺陷我们使用 PERL 语言编辑了多个程序用于从已有数据库中检索硒蛋白。表 1 展示了利用生物信息学方法检索识别硒蛋白的步骤以及每步筛选结果。在已预测基因中, 有 11365 条基因是以 TGA 终止的, 其中 5710 条位于正义链, 5655 条位于反义链。抽提出这些基因, 对于正义链, 截取其 DNA 中 TGA 下游 999 个核苷酸, 连接在完整 cDNA 开放阅读框后。很明显所

截取部分处于该基因的 3'-UTR 区域; 而对于反义链则截取并连接 TGA 上游 999 个核苷酸。由前面的介绍我们知道, 硒蛋白基因 3'-UTR 区存在 SECIS 结构, 故对上一步截取得到的部分进行 SECIS 结构检索。结果显示, 918 条含 TGA 基因的 3'-UTR 存在 SECIS 结构, 其中 445 条来自正义链, 473 条来自反义链。

表 1 硒蛋白识别的主要步骤

硒蛋白识别步骤	基因数
基因预测	31254
检索以 TGA 码终止的基因	11365
检索 SECIS 结构	918
检索 Sec/Cys 配对同源序列	58
分析硒蛋白开放阅读框	42
分析 SECIS 结构	7

为了识别硒蛋白, 我们把同时具有编码区内 TGA 和下游区域 SECIS 结构的基因翻译成氨基酸序列, 将 ORF 中的 TGA 解码为 U, 得到 918 条包含 U 残基的蛋白序列。以序列中第一个 U 为中点, 截取与 U 相邻的上游和下游序列中各 166 个氨基酸残基, 得到 918 条蛋白质局部序列片段。采用 BLASTP 程序将这些片段与 NCBI 无冗余蛋白质 NR 库进行同源性比对, 寻找所截取的蛋白片段在其它生物中的同源蛋白, 生成序列对。再用 Clustal W 程序对这些序列对进行比对处理, 并用软件 GeneDoc 观察筛选结果。由于大部分已知硒蛋白都存在序列中 Sec 与其同源序列中 Cys 相对应的现象, 即同源蛋白间存在 U/C 配对, 故我们使用以下规则来筛选疟蚊硒蛋白: 第一, 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。第二, 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。如此筛选后得到 58 条蛋白序列。图 1 示出 7 条预测的冈比亚疟蚊硒蛋白与其同源蛋白序列的比对情况。图中左侧文字为基因获取号, 符号 ▲ 指示 U 的位置。结果显示同源蛋白在 U 位置为 C, 即疟蚊蛋白与同源蛋白之间具有 U/C 配对关系, 同时 U/C 对两侧的氨基酸序列同源性较好, 符合硒蛋白基因的特征。在上述检索结果基础上, 对所得硒蛋

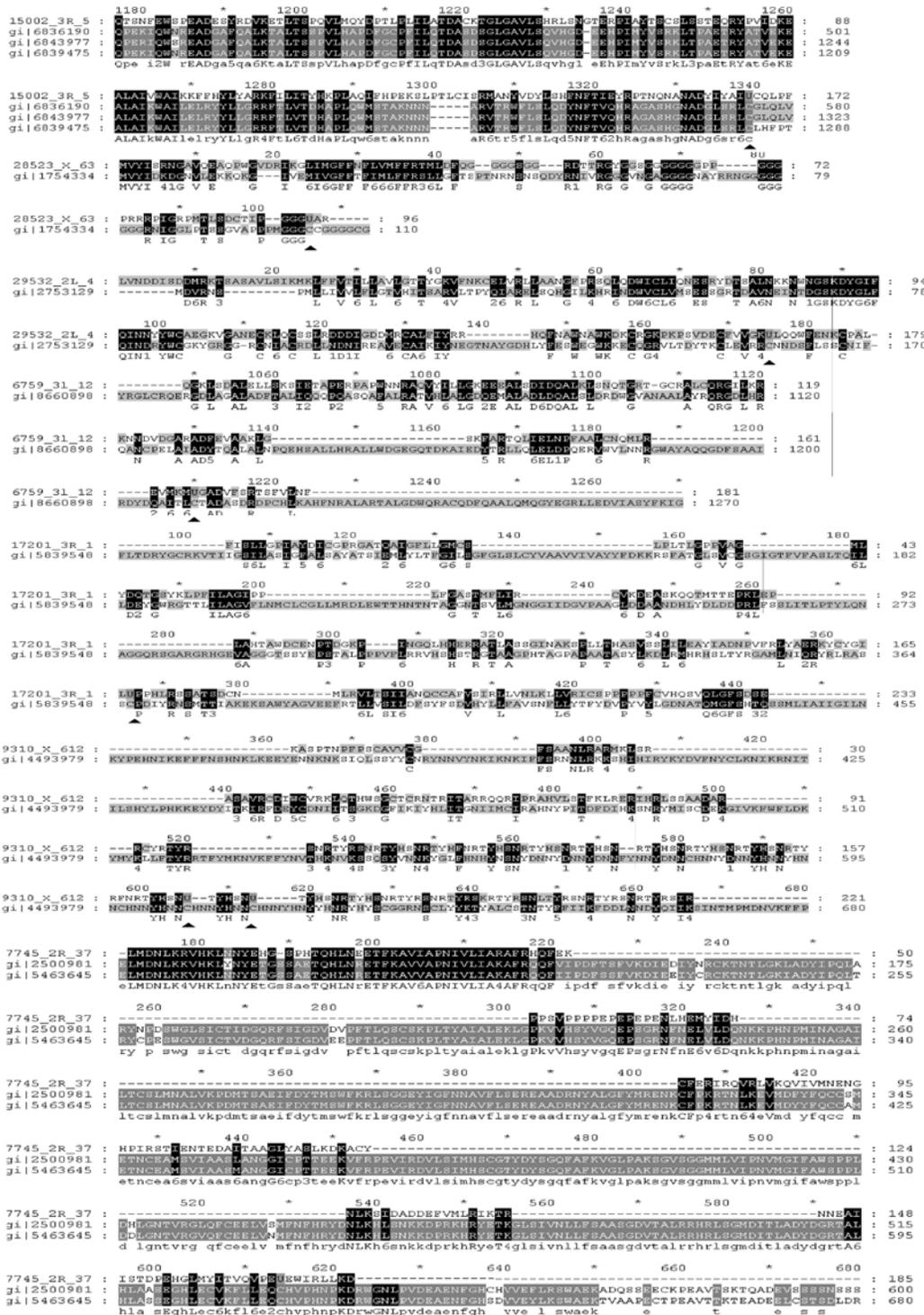


图 1 疟蚊涎蛋白与同源蛋白基因的多重序列比对

白候选基因的编码区进行重新分析，以 TGA 后出现的第一个框内终止码来结束编码区。同时分析 SECIS 的具体位置，当 SECIS 结构出现在新的开放阅读框中或者跨越了开放阅读框和 3'-UTR，该基因将被舍弃。只有 SECIS 结构位于 3'-UTR 的基因才有可能为硒蛋白。如表 1 所示，我们得到了 42 条基因，它们均含有编码区内的 TGA，3'-UTR 内的 SECIS 结构以及同源类似物具有 U/C 配对和相似的双侧翼序列，符合硒蛋白基因的主要特征要求。然而，在对上述 42 条基因序列的 SECIS 一级保守序列进行分析时，我们发现仅有 7 条基因的 SECIS 序列与已知硒蛋白基因的 SECIS 保守区相似，即为 AUGA_AA_GA/AUGA_CC_GA/UUGA_AG_GA，其余 35 条序列的 SECIS 保守区目前尚未在已知蛋白中发现，无法证实它们是硒蛋白基因中的 SECIS 结构，为保证结果不出现假阳性，舍弃这 35 条基因，最终获得 7 条完全符合硒蛋白基因特点的疟蚊硒蛋白。

2.2 疟蚊硒蛋白 SECIS 结构特点

(1) 疟蚊 SECIS 的一级保守序列：不同硒蛋白的 SECIS 都有保守的一级序列，从 5' 端到 3' 端依次含有 3 段特征保守碱基 AUGA-AA-GA，5' 端有一个腺嘌呤 A，极少数硒蛋白该位置出现的是鸟嘌呤 G。中间的两个腺嘌呤 AA 位于 SECIS 茎环结构的顶环，少数硒蛋白在该位置由两个胞嘧啶 CC 替代，如硒蛋白 M (SelM)^[15-17]。所以现有的真核生物硒蛋白基因 SECIS 一级结构保守序列主要包含 4 种形式 AUGA_AA_GA, GUGA_AA_GA, AUGA_CC_GA 和 UUGA_AG_GA。图 2 列出所得 7 条基因的 SECIS 一级结构，其

中 5 条硒蛋白一级序列保守核苷酸模式为经典的 AUGA_AA_GA 模式，余下 2 条分别为 GUGA_AA_GA 和 UUGA_AG_GA，这 3 种疟蚊硒蛋白保守核苷酸模式与已知真核生物硒蛋白的 SECIS 保守核苷酸模式相同。

(2) 疟蚊 SECIS 的二级结构特点：图 3 示出上步所得 7 条疟蚊硒蛋白 SECIS 元件的高级结构图。疟蚊 SECIS 的二级结构与已知硒蛋白的 SECIS 元件相同，是一种茎环结构。这种二级结构中含有 2 个或 3 个环，下部茎端是由 A(或 U)UGA-GA 通过非 Watson-Crick 碱基配对构成的四碱基单元，上端环中含有保守的 AA 或 GA 核苷酸。由图可见，只有疟蚊 17201 基因的 SECIS 结构属于一型 SECIS 结构，其它疟蚊硒蛋白 SECIS 结构的顶环上均有附加小环出现，属于二型 SECIS 结构。

2.3 疟蚊硒蛋白分子的功能预测

上述硒蛋白基因的预测是采用 GenScan 或 SNAP 等基因预测程序，通过密码子解码概率算法或者基因同源性比较算法获得原始序列，并根据硒蛋白基因特点筛选出 7 条疟蚊硒蛋白基因，其功能仍是未知的。为此我们采用 BLASTP 在线程序进行疟蚊硒蛋白序列相对于 NCBI 蛋白库的同源性比对，分析推测疟蚊硒蛋白的功能，结果如表 2 所示。疟蚊基因 28253 为已知硒蛋白，它与伊蚊硒蛋白 G 具有同源性。余下 6 条蛋白皆为未知硒蛋白，分别与溶菌酶、CG2 蛋白、谷氨酰胺酶、TRP 重复序列蛋白、单羧酸盐转运蛋白和转座酶的序列具有高同源性，由此推断它们的功能与上述蛋白相似。

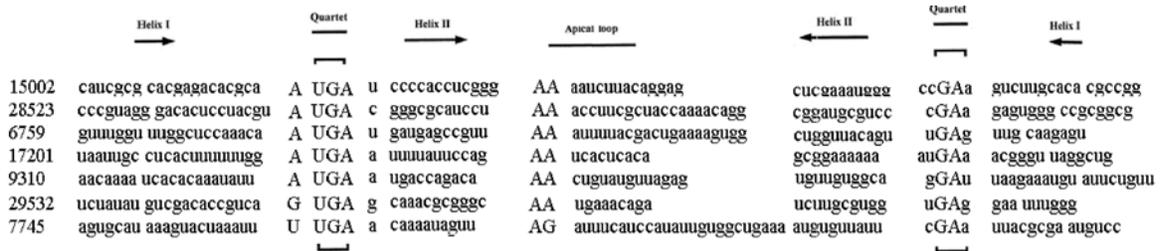


图 2 疟蚊硒蛋白 SECIS 结构

表 2 疟蚊硒蛋白基因注释表

基因	染色体	SECIS 模式	功能注释
29532	2L	TGA_AA_GA	溶菌酶同源蛋白
28523	X	ATGA_AA_GA	硒蛋白 G 同源蛋白
9310	X	ATGA_AA_GA	CG2 蛋白同源蛋白
7745	2R	TTGA_AG_GA	谷氨酰胺酶同源蛋白
6759	3L	ATGA_AA_GA	TRP 重复序列同源蛋白
17201	3R	ATGA_AA_GA	单羧酸盐转运同源蛋白
15002	3R	ATGA_AA_GA	转座酶同源蛋白

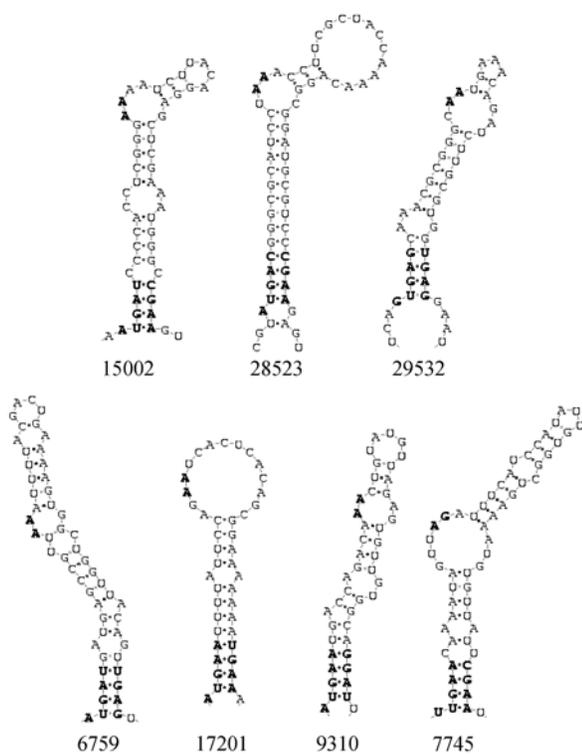


图 3 疟蚊硒蛋白 SECIS 二级结构

3 讨论和结论

硒蛋白生物合成过程涉及到硒元素的特殊掺入机制, 位于ORF内部的密码子UGA翻译为Sec进入蛋白质分子, 而UGA在通常情况下作为终止码介导蛋白质的合成. 硒蛋白基因合成过程中UGA的这种特殊解码方式是由mRNA中的特定茎环结构——SECIS元件所决定 [13~16]. 不同硒蛋白的SECIS元件都有保守的一级序列和特定的高级结构. 真核生物中SECIS元件位于mRNA的3'-非翻译区(UTR). 硒蛋白基因序列中TGA 密码子和SECIS元件之间的距离最

大可达 217 kb, 最小可为 0~111 bp, 而原核生物SECIS元件在紧邻TGA密码子的下游 [20]. 此外, 大部分已知的硒蛋白序列都具有一种特性, 在其同源蛋白序列中, 对应于Sec的氨基酸残基为Cys, 即同源蛋白之间具有Sec/Cys配对 [21]. 研究表明: 疟蚊硒蛋白基因同样含有编码区内TGA并解码为Sec, 由此介导硒元素掺入蛋白质分子. 7 条疟蚊硒蛋白基因的SECIS结构与编码区内TGA密码子的距离在 50~1000 之间, SECIS保守碱基的模式与已知硒蛋白的SECIS结构相同. 7 种疟蚊硒蛋白的序列分别与伊蚊硒蛋白 G, 溶菌酶, CG2 蛋白, 谷氨酰胺酶, TPR重复序列蛋白, 单羧酸盐转运蛋白和转座酶具有高同源性, 并与之形成U/C配对关系.

冈比亚疟蚊是疟原虫的载体, 是传播疟疾的主要媒介昆虫. 全世界每年近 5 亿人感染疟疾, 造成 200 万余人的死亡 [22,9]. 冈比亚疟蚊基因组的测序工作, 为从传播途径防治疟疾提供了理论支持, 可以从分子水平解释疟蚊多种生态和生理现象, 如疟蚊对人血的特异性摄食、对多种杀虫剂的强抗药性、疟原虫的强带菌能力及在疟蚊体内的寄生过程等, 以基因调控的方式降低疟蚊的存活率、寿命, 或者降低其抗药性、带菌能力, 或者干涉疟原虫在疟蚊体内的寄生生态, 以期控制疟疾的危害. 目前这一领域的研究已获得很多结果 [23~25]. Kafatos等人 [26]报道蚊子的带菌能力可能与其氧化耐受力相关. 抗氧化作用涉及到谷胱甘肽过氧化物酶、硫氧还蛋白还原酶、过氧化氢酶、超氧化物歧化酶等抗氧化酶的参与, 其中哺乳动物谷胱甘肽过氧化物酶和硫氧还蛋白还原酶就是硒蛋白 [27~29]. 我们通过生物信息学方法, 从疟蚊基因库中发现了疟蚊新硒蛋白, 并推测其生物功能, 希望能从硒蛋白的抗氧化作用角度研究疟蚊, 为干涉其

带菌能力提供线索. 而从同源序列比对结果可见, 疟蚊硒蛋白所具有的生物功能有助于达到上述目的, 如溶菌酶对疟蚊带菌能力的影响, 谷氨酰胺酶蛋白对疟蚊免疫系统和运动系统的影响等, 都可能作为干涉、调控疟蚊生长传播的因素.

目前已知的哺乳动物硒蛋白都有各自独特的结构特点和生物功能, 各硒蛋白之间缺乏共同的氨基酸保守序列. 因此, 通用的计算机生物软件, 如GeneScan, GeneFinder或者BLAST和FASTA等, 都无法单独识别新硒蛋白. 徐辉碧等人^[30]曾报道用RNADraw程序识别硒蛋白的SECIS结构, 但却不能用该程序识别新硒蛋白. 由于硒蛋白基因具有共同点, 即开读框架内部含有TGA密码子、3'-UTR存在SECIS元件、同源蛋白存在U/C配对、以及TGA码两侧序列相似, 因此, 本文探索了一种使用计算机检索基因组序列、识别硒蛋白的方法. 它基于硒蛋白基因的共同特点, 通过使用多种生物学软件并自主编辑多个程序, 在基因组中进行多步检索以识别硒蛋白. 该方法能寻找和发现目前尚未发现的新硒蛋白, 这是通常硒蛋白鉴别手段所不具备的, 因此研究方法优于目前已有的基因注释方法. 通过使用该方法, 我们从疟蚊基因库中发现了7条硒蛋白. 尽管通过生物信息学方法识别的硒蛋白仍需最终由实验证实, 但它从理论计算上预测了物种中可能存在的所有硒蛋白, 为后续实验提供了强有力的理论基础. 该方法同样适用于其它物种中硒蛋白的识别与鉴定.

参 考 文 献

- Hatfield D L. Selenium: Its Molecular Biology and Role in Human Health. Boston: Kluwer Academic Publishers, 2001. 313—319
- Kryukov G V, Castellano S, Novoselov S V, et al. Characterization of mammalian selenoproteomes. *Science*, 2003, 300(30): 1439—1443[DOI]
- Atkins J F, Gesteland R F. The twenty-first amino acid. *Nature*, 2000, 407: 463—465[DOI]
- Bock A. Biosynthesis of selenoproteins - an overview. *Biofactors*, 2000, 11: 77—78
- Hatfield D L, Gladyshev V N. How selenium has altered our understanding of the genetic code. *Mol Cell Biol*, 2002, 22: 3565—3576[DOI]
- Castellano S, Morozova N, Morey M, et al. *In silico* identification of novel selenoproteins in the *Drosophila melanogaster* genome. *EMBO Reports*, 2001, 2(8): 697—702[DOI]
- Burge C B, Karlin S. Finding the genes in genomic DNA. *Curr Opin Struct Biol*, 1998, 8: 346—353[DOI]
- Haussler D. Computational gene finding. *Trends Biochem. Sci.*, *Supplementary Guide to Bioinformatics*, 1998, 12—15
- Holt R A, Subramanian G M, Halpern A. The Genome Sequence of the Malaria Mosquito *Anopheles gambiae*. *Science*, 2002, 298(4): 129—149[DOI]
- Li J, Riehle M M, Zhang Y, et al. *Anopheles gambiae* genome re-annotation through synthesis of abinitio and comparative gene prediction algorithms. *Genome Biol*, 2006, 7: R24.1—R24.12
- Bauer H, Gromer S, Urbani A, et al. Thioredoxin reductase from the malaria mosquito *Anopheles gambiae*. *Eur J Biochem*, 2003, 270: 4272—4281[DOI]
- 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]
- Berry M J, Banu L, Chen Y, et al. Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. *Nature*, 1991, 353: 273—276[DOI]
- 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: 367—379
- Tujebajeva R M, Copeland P R, Xu X M, et al. Decoding apparatus for eukaryotic selenocysteine insertion. *EMBO Rep*, 2000, 1: 158—163[DOI]
- 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]
- Lambert A, Lescure A, Gautheret D. A survey of metazoan selenocysteine insertion sequence. *Biochimie*, 2002, 84: 953—959[DOI]
- Fagegaltier D, Lescure A, Walczak R, et al. Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Research*, 2000, 28(14): 2679—2689[DOI]
- Thanbichler M, Bock A. The function of SECIS RNA in translational control of gene expression in *Escherichia coli*. *EMBO J*, 2002, 21(24): 6925—6934[DOI]
- Bock A. In: Burk RF, ed. *Selenium in Biology and Human Health*. New York: Springer-Verlag, 1994. 9—24
- Hatfield D L, Gladyshev V N. How selenium has altered our understanding of the genetic code. *Mol Cell Biol*, 2002, 22: 3565—

- 2576[DOI]
- 22 Nighorn A, Hildebrand J G. Dissecting the molecular mechanisms of olfaction in a malaria-vector mosquito. *PNAS*, 2002, 99(3): 1113—1114[DOI]
- 23 Hill C A, Fox A N, Pitts R J, et al. G protein-coupled receptors in *Anopheles gambiae*. *Science*, 2002, 298: 176—178[DOI]
- 24 Christophides G K, Zdobnov E, Barillas-Mary C, et al. Immunity-Related Genes and Gene Families in *Anopheles gambiae*. *Science*, 2002, 298: 159—165[DOI]
- 25 Zheng L, Conel A J, Wang R, et al. Quantitative trait loci for refractoriness of *Anopheles gambiae* to *Plasmodium cynomolgi* B. *Science*, 1997, 276: 425—428[DOI]
- 26 Kafatos F C, Levashina E, Kokoza E, et al. Man “bytes” mosquito. *EMBL Research Reports*, 2002, 184—189
- 27 Margaret P R. The importance of selenium to human health. *The Lancet*, 2000, 356(9225): 233—241[DOI]
- 28 Burk R F, Hill K E. Orphan selenoproteins. *Bioassays*, 1999, 21(3): 231—237[DOI]
- 29 Glydyshev V N, Jeang K, Wootton J C, et al. A new human selenium-containing protein: Purification, characterization, and cDNA sequence. *J Biol Chem*, 1998, 273(15): 8910—8915[DOI]
- 30 徐辉碧, 黄开勋, 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