

doi:10.3969/j.issn.1672-5565.2014.01.04

双峰驼凝乳酶原基因的生物信息学分析

杨艺, 普燕, 张富春*

(新疆大学生命科学与技术学院, 新疆生物资源与基因工程重点实验室, 新疆, 乌鲁木齐, 830046)

摘要:通过生物信息学的方法对双峰驼凝乳酶原基因及相应的氨基酸序列的同源性、理化性质、保守结构域、亚细胞定位、信号肽、跨膜结构域、亲水性/疏水性、二级结构进行预测分析。结果表明,双峰驼凝乳酶原基因开放阅读框全长1 146 bp, 编码381个氨基酸,属于胃蛋白酶A超家族,预测定位于内质网(膜)的稳定亲水性蛋白,具有一个16个氨基酸的信号肽,其不含跨膜结构域。无规卷曲是其二级结构中最大量的结构元件, α 螺旋和延伸链分散于整个蛋白质中,活性位点的分析表明,编码蛋白有6类活性位点。分析双峰驼凝乳酶原基因及其编码蛋白质的特征,能够为深入开展双峰驼凝乳酶的表达和凝乳特性研究提供理论依据。

关键词:双峰驼;凝乳酶原;克隆;生物信息学

中图分类号:Q557 **文献标志码:**A **文章编号:**1672-5565(2014)-01-018-09

Bioinformatics analysis of prochymosin gene in *camelus bactrianus*

YANG Yi, PU Yan, ZHANG Fuchun*

(Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi, Xinjiang 830046, China)

Abstract: In this paper the research on *camelus bactrianus* prochymosin gene sequence was illustrated using Bioinformatics methods to analyze amino acid sequence homology, physical and chemical characters, conserved domains, subcellular localization, signal peptide, transmembrane domain, hydrophilic/hydrophobic, and secondary structure. The results showed that the length of *camelus bactrianus* prochymosin is 1 146bp, encoding a protein with 381 amino acids. This protein belongs to pepsin A superfamily, and is a stable hydrophilic protein most probably located in endoplasmic reticulum (membrane). It has a signal peptide of 16 amino acids and no transmembrane domain. In the secondary structure, random coil is the majority of structural elements, while α -helix and extended stretch distribute throughout the protein chain. The active site analysis showed there are six kinds of active sites in the prochymosin. Analysis of *camelus bactrianus* prochymosin genes and the characteristics of their encoded proteins provide a theoretical reference to study curd characteristics and expression conditions.

Keywords: *camelus bactrianus*; Prochymosin; Cloning; Bioinformatics

随着乳品行业的高速发展,人们对发酵乳制品的需求也日益增加。近年来国内外干酪市场发展迅猛,不仅对干酪的研究日益深入,而且制作干酪所用的凝乳酶需求量大增。干酪加工中添加凝乳酶的主要目的是促使牛乳凝结,为排除乳清提供条件^[1]。凝乳酶是一种从未断奶的小牛皱胃中发现

的天门冬氨酸蛋白酶,可以专一地切割乳中 κ -酪蛋白的Phe105-Met106之间的肽键,破坏乳中的酪蛋白胶束引起蛋白质凝聚,因此在乳制品尤其是在干酪加工制作中应用广泛^[2],理想的凝乳酶应能迅速、专一地打开 κ -酪蛋白Phe105-Met106键,但水解蛋白其他肽键的能力较低,即两者的活力比值要高^[3]。

收稿日期:2013-11-19;修回日期:2013-12-25.

基金项目:新疆动物学重点学科-乳品工程资助(2011001)。

作者简介:杨艺,女,硕士研究生,研究方向:生物化学与分子生物学; E-mail: 563226049@qq.com.

*通信作者:张富春,男,教授,研究方向:分子生物学; E-mail: zfcxju@gmail.com.

凝乳酶凝乳后制成的干酪相当于将原料奶中的蛋白质和脂肪浓缩了十倍左右,另外干酪中还含有糖类、有机酸、常量矿物质元素等多种成分,具有很高的营养价值。凝乳酶主要为动物来源的凝乳酶,约占市场的 70%,同时还有植物来源的凝乳酶、微生物来源的凝乳酶和利用基因工程技术生产的重组凝乳酶。牛凝乳酶有着强大的凝乳活性和较低的蛋白水解活性,是凝乳制作干酪的优质酶,目前对牛凝乳酶的研究已经较为成熟,市场上销售的凝乳酶主要是牛凝乳酶,而相应骆驼凝乳酶研究较少,因此骆驼凝乳酶开发利用前景广阔。Kappeler 等^[4]研究发现单峰驼凝乳酶与牛凝乳酶有不同的特性,单峰驼凝乳酶表现出比牛凝乳酶高 70% 的凝乳活性,却只有牛凝乳酶的 20% 非特异性蛋白酶活性。动力学分析表明,达到半饱和时,其所需的底物仅是牛凝乳酶的 50% 还少,并且牛凝乳酶不能凝结骆驼乳,说明骆驼凝乳酶有较高的凝乳活性。Jensen 等^[5]研究表明单峰驼凝乳酶的稳定性高于牛凝乳酶。

本研究主要运用生物信息学的方法对双峰驼凝乳酶基因及相应的氨基酸序列的同源性、理化性质、保守结构域、亚细胞定位、信号肽、跨膜结构域、亲水性/疏水性、二级结构以及活性位点进行预测分析。旨在为研究双峰驼凝乳酶的凝乳特性及其条件提供理论参考。

1 材料和方法

1.1 双峰驼凝乳酶原基因序列的获得

首先从 Broe^[6] 等人在 2011 申请的专利中获得双峰驼凝乳酶原基因序列,由上海生工生物工程有 限公司合成双峰驼凝乳酶原基因开放阅读框序列,在大肠杆菌进行原核表达,发现双峰驼凝乳酶原融合蛋白以包涵体形式表达,分子大小与理论值相符。将双峰驼凝乳酶原融合蛋白包涵体溶于尿素,但透析复性的双峰驼凝乳酶原融合蛋白经过酸化/中和处理并没有得到有活性的凝乳酶,说明该基因序列所产生的双峰驼凝乳酶原有误。随后通过检索 GeneBank 数据库,发现 Kappeler 等人正式注册的双峰驼凝乳酶原基因序列,与从专利获得的序列比对后,发现两者双峰驼凝乳酶原有六个氨基酸的差异,推测可能是这六个氨基酸的差异导致合成的凝乳酶无活性。本研究采用从 GeneBank 数据库中获得的 双峰驼凝乳酶原基因序列进行分析。

1.2 生物信息学分析

利用 NCBI 及 ProtParam、ProtScale、SignalP 4.0 Server、TMHMM2.0、等在线工具进行核酸及氨基酸序列的组成、理化性质、信号肽和亲水性/疏水性分析;利用 Blast 和 DNAMAN 在线工具以及 MEGA4.0 软件,进行氨基酸序列的序列比对和同源性分析。所用软件或在线工具有关信息见表 1。

表 1 预测基因结构和功能应用的网站及软件

Table 1 The websites and softwares for predicting the gene structure and function

编号	软件名称或网站网址	主要功能	参考文献
1	http://www.ncbi.nlm.nih.gov/gorf/gorf.html	分析并找到序列的 ORF 区	[7]
2	DNAMAN	核酸及氨基酸多重序列的比对	[7]
3	NCBI 中 Conserved Domain Database	分析蛋白保守结构域	[7]
4	http://web.expasy.org/protparam/	对蛋白质基本理化性质的分析	[8]
5	http://www.expasy.org/protscale/	对蛋白进行亲/疏水性分析	[8]
6	http://www.cbs.dtu.dk/services/SignalP/	分析序列是否含有信号肽	[9]
7	http://psort.hgc.jp/form.html	亚细胞定位	[10]
8	http://www.cbs.dtu.dk/services/TMHMM/	跨膜结构预测	[11]
9	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl/page=/NPSA/npsa_gor4.html	蛋白二级结构分析	[12]
10	MEGA4.0	绘制系统进化树	[13]
11	http://npsa-pbil.ibcp.fr/	对活性位点的分析	[14]

2 结果与分析

2.1 双峰驼凝乳酶原基因序列

双峰驼凝乳酶原基因序列是一个完整的开放阅读框,为 1 146 bp,编码 381 个氨基酸(见图 1)。

```

1  ATGCGGTGCCTCGTGGTGCTACTTGCAGCCCTCGCTCTCTCCCAGGCCAGTGGGATCACC
1  M R C L V V L L A A L A L S Q A S G I T
61  AGGATCCCTCTGCACAAAGGCAAGACTCTGAGAAAAGCGCTGAAGGAGCGTGGGGCTCCTG
21  R I P L H K G K T L R K A L K E R G L L
121  GAGGACTTTCTGCAGAGACAACAGTATGCCGTCAGCAGCAAGTACTCCAGCTTGGGGAAG
41  E D F L Q R Q Q Y A V S S K Y S S L G K
181  GTGGCCAGGGAACCCCTGACCAGCTACCTGGATAGTCAGTACTTTGGGAAGATCTACATC
61  V A R E P L T S Y L D S Q Y F G K I Y I
241  GGGACCCACCCAGGAGTTCACCGTGGTGTGTTGACACTGGCTCCTCTGACCTGTGGGTG
81  G T P P Q E F T V V F D T G S S D L W V
301  CCCTCTATCTACTGCAAGAGCAATGTCTGCAAAAACCACCACCGCTTTGACCCGAGAAAG
101  P S I Y C K S N V C K N H H R F D P R K
361  TCGTCCACCTTCCGGAACCTGGGCAAGCCCCTGTCCATCCATTACGGCACGGGCAGCATG
121  S S T F R N L G K P L S I H Y G T G S M
421  GAGGGCTTTCTGGGCTACGACACCGTCACCGTCTCCAACATTGTGGACCCCAACCAGACT
141  E G F L G Y D T V T V S N I V D P N Q T
481  GTGGGCCTGAGCACCGAGCAACCTGGCGAGGTCTTCACCTACTCCGAGTTTGACGGGGATC
161  V G L S T E Q P G E V F T Y S E F D G I
541  CTGGGGCTGGCCTACCCCTCGCTTGCCTCCGAGTACTCGGTGCCCGTGTGTTGACAATATG
181  L G L A Y P S L A S E Y S V P V F D N M
601  ATGGACAGACACCTGGTGGCCCGAGACCTGTTCTCGGTTTACATGGACAGGAATGGCCAG
201  M D R H L V A R D L F S V Y M D R N G Q
661  GGGAGCATGCTTACACTGGGGGCCATTGACCCGTCCTACTACACCGGCTCCCTGCACTGG
221  G S M L T L G A I D P S Y Y T G S L H W
721  GTGCCCGTGACCTTGCAGCAGTACTGGCAGTTCACCGTGGACAGTGTACCATCAACGGG
241  V P V T L Q Q Y W Q F T V D S V T I N G
781  GTGGCAGTGGCCTGTGTGGTGGCTGTCAGGCCATCCTGGACACGGGTACCTCCGTGCTG
261  V A V A C V G G C Q A I L D T G T S V L
841  TTCGGGCCAGCAGCGACATCCTCAAAATTGAGATGGCTATTGGAGCCACAGAGAACCGA
281  F G P S S D I L K I Q M A I G A T E N R
901  TATGGTGAGTTTGACGTCAACTGTGGGAACCTGAGGAGCATGCCACCGTGGTCTTCGAG
301  Y G E F D V N C G N L R S M P T V V F E
961  ATCAATGGCAGAGACTACCCACTGTCCCCCTCCGCTACACAAGCAAGGACCAGGGCTTC
321  I N G R D Y P L S P S A Y T S K D Q G F
1021  TGCACCAGTGGCTTTCAAGGTGACAACAATTCCGAGCTGTGGATCCTGGGGGATGTCTTC
341  C T S G F Q G D N N S E L W I L G D V F
1081  ATCCGGGAGTATTACAGTGTCTTTGACAGGGCCAACAATCGCGTGGGGCTGGCCAAGGCC
361  I R E Y Y S V F D R A N N R V G L A K A
1141  ATCTGA
381  I *

```

图 1 双峰驼凝乳酶原序列及其氨基酸序列

Fig.1 Nucleotide acid sequence and amino acid sequence of prochymosin in *camelus bactrianus*

2.2 双峰驼凝乳酶原的同源性分析

通过 GeneBank 数据库中的 Blastp 程序,对双峰驼凝乳酶的氨基酸序列进行同源性分析,数据显示其与单峰驼、牦牛、牛、山羊、印度水牛、绵羊凝乳酶氨基酸序列相似性分别为 99%、84%、84%、84%、

83%和 84%。利用 DNAMAN 软件对双峰驼凝乳酶原的氨基酸序列和其它物种凝乳酶原的氨基酸序列进行多重比对(见图 2)。运用 MEGA4.0 软件,根据上述动物的凝乳酶原的氨基酸序列构建的系统进化树(见图 3),进一步的验证了其同源性。

双峰驼凝乳酶原	MRCLVLLRALALSQASCITRIPLHKGKILRKALKKRGLEDFLQRCQYAVSSKYSGLGVVREPLTSLYLDQYFGKIYL	80
单峰驼凝乳酶原	MRCLVLLRALALSQASCITRIPLHKGKILRKALKKRGLEDFLQRCQYAVSSKYSGLGVVREPLTSLYLDQYFGKIYL	80
牦牛凝乳酶原	MRCLVLLRVFALSQGABITRIPLKYGKSLRKALKKRGLEDFLQRCQYAVSSKYSGLGVVREPLTSLYLDQYFGKIYL	80
牛凝乳酶原	MRCLVLLRVFALSQGABITRIPLKYGKSLRKALKKRGLEDFLQRCQYAVSSKYSGLGVVREPLTSLYLDQYFGKIYL	80
山羊凝乳酶原	MRCLVLLRVFALSQGABITRIPLKYGKSLRKALKKRGLEDFLQRCQYAVSSKYSGLGVVREPLTSLYLDQYFGKIYL	80
印度水牛凝乳酶原	MRCLVLLRVFALSQGABITRIPLKYGKSLRKALKKRGLEDFLQRCQYAVSSKYSGLGVVREPLTSLYLDQYFGKIYL	80
绵羊凝乳酶原	MRCLVLLRVFALSQGABITRIPLKYGKSLRKALKKRGLEDFLQRCQYAVSSKYSGLGVVREPLTSLYLDQYFGKIYL	80
consensus	Mrclvllra lals qasc itripl hkgk lrkalkke gll edflq rcqy av ss kys glgv vrep lts l yldsq yfgki y	
双峰驼凝乳酶原	GTPPQEFVTFD TGSSDLWVPSIYCKSNVCKNHFRFDPKRSSTFRNLGKPLSIRHYGTGSMGGFLGYDVTVTSNIVDINQCT	160
单峰驼凝乳酶原	GTPPQEFVTFD TGSSDLWVPSIYCKSNVCKNHFRFDPKRSSTFRNLGKPLSIRHYGTGSMGGFLGYDVTVTSNIVDINQCT	160
牦牛凝乳酶原	GTPPQEFVTFD TGSSDLWVPSIYCKSNVCKNHFRFDPKRSSTFRNLGKPLSIRHYGTGSMGGFLGYDVTVTSNIVDINQCT	160
牛凝乳酶原	GTPPQEFVTFD TGSSDLWVPSIYCKSNVCKNHFRFDPKRSSTFRNLGKPLSIRHYGTGSMGGFLGYDVTVTSNIVDINQCT	160
山羊凝乳酶原	GTPPQEFVTFD TGSSDLWVPSIYCKSNVCKNHFRFDPKRSSTFRNLGKPLSIRHYGTGSMGGFLGYDVTVTSNIVDINQCT	160
印度水牛凝乳酶原	GTPPQEFVTFD TGSSDLWVPSIYCKSNVCKNHFRFDPKRSSTFRNLGKPLSIRHYGTGSMGGFLGYDVTVTSNIVDINQCT	160
绵羊凝乳酶原	GTPPQEFVTFD TGSSDLWVPSIYCKSNVCKNHFRFDPKRSSTFRNLGKPLSIRHYGTGSMGGFLGYDVTVTSNIVDINQCT	160
consensus	gtpppqeftv fdtgssdl wvpsiycksn cknh rfdprkstf rnlgkplsi rhygtgsm g flgydvtvtsnivd inqct	
双峰驼凝乳酶原	VGLSTQPPGVFTVTFEFDGILEYAYPSLASEYSVPVFDNMMDRFLVRCDLFSVYMDRNGGCSMLTLGAIDPSYYTGLSLHW	240
单峰驼凝乳酶原	VGLSTQPPGVFTVTFEFDGILEYAYPSLASEYSVPVFDNMMDRFLVRCDLFSVYMDRNGGCSMLTLGAIDPSYYTGLSLHW	240
牦牛凝乳酶原	VGLSTQPPGVFTVTFEFDGILEYAYPSLASEYSVPVFDNMMDRFLVRCDLFSVYMDRNGGCSMLTLGAIDPSYYTGLSLHW	240
牛凝乳酶原	VGLSTQPPGVFTVTFEFDGILEYAYPSLASEYSVPVFDNMMDRFLVRCDLFSVYMDRNGGCSMLTLGAIDPSYYTGLSLHW	240
山羊凝乳酶原	VGLSTQPPGVFTVTFEFDGILEYAYPSLASEYSVPVFDNMMDRFLVRCDLFSVYMDRNGGCSMLTLGAIDPSYYTGLSLHW	240
印度水牛凝乳酶原	VGLSTQPPGVFTVTFEFDGILEYAYPSLASEYSVPVFDNMMDRFLVRCDLFSVYMDRNGGCSMLTLGAIDPSYYTGLSLHW	240
绵羊凝乳酶原	VGLSTQPPGVFTVTFEFDGILEYAYPSLASEYSVPVFDNMMDRFLVRCDLFSVYMDRNGGCSMLTLGAIDPSYYTGLSLHW	240
consensus	vglst qppgvftv tfe fdgile y ay ps la se ys vpvfdnmm dr flvrc dl fs v y mdr ngg cs ml t l ga id ps yy t gl sl hw	
双峰驼凝乳酶原	VEVTLQCYWQFTVDSVTHSGAVVACGGCCQAILDGTSTRLVGPSSDIINICQAI GATONCYGEFVIDIDCNLSYMPTVVFE	320
单峰驼凝乳酶原	VEVTLQCYWQFTVDSVTHSGAVVACGGCCQAILDGTSTRLVGPSSDIINICQAI GATONCYGEFVIDIDCNLSYMPTVVFE	320
牦牛凝乳酶原	VEVTLQCYWQFTVDSVTHSGAVVACGGCCQAILDGTSTRLVGPSSDIINICQAI GATONCYGEFVIDIDCNLSYMPTVVFE	320
牛凝乳酶原	VEVTLQCYWQFTVDSVTHSGAVVACGGCCQAILDGTSTRLVGPSSDIINICQAI GATONCYGEFVIDIDCNLSYMPTVVFE	320
山羊凝乳酶原	VEVTLQCYWQFTVDSVTHSGAVVACGGCCQAILDGTSTRLVGPSSDIINICQAI GATONCYGEFVIDIDCNLSYMPTVVFE	320
印度水牛凝乳酶原	VEVTLQCYWQFTVDSVTHSGAVVACGGCCQAILDGTSTRLVGPSSDIINICQAI GATONCYGEFVIDIDCNLSYMPTVVFE	320
绵羊凝乳酶原	VEVTLQCYWQFTVDSVTHSGAVVACGGCCQAILDGTSTRLVGPSSDIINICQAI GATONCYGEFVIDIDCNLSYMPTVVFE	320
consensus	vev t l q c y w q f t v d s v t h s g a v v a c g g c c q a i l d t g t s r l v g p s s d i i n i c q a i g a t o n c y g e f v i d i d c n l s y m p t v v f e	
双峰驼凝乳酶原	INGRDIYPLSISAYTSKDCGFCTSGFCGDNSELWILGDVFIREYYSVDFRANNVGLAKAI	381
单峰驼凝乳酶原	INGRDIYPLSISAYTSKDCGFCTSGFCGDNSELWILGDVFIREYYSVDFRANNVGLAKAI	381
牦牛凝乳酶原	INGKMYPLIISAYTSQDCGFCTSGFCGENRHCQWILGDVFIREYYSVDFRANNVGLAKAI	381
牛凝乳酶原	INGKMYPLIISAYTSQDCGFCTSGFCGENRHCQWILGDVFIREYYSVDFRANNVGLAKAI	381
山羊凝乳酶原	INGKMYPLIISAYTSQDCGFCTSGFCGENRHCQWILGDVFIREYYSVDFRANNVGLAKAI	381
印度水牛凝乳酶原	INGKMYPLIISAYTSQDCGFCTSGFCGENRHCQWILGDVFIREYYSVDFRANNVGLAKAI	381
绵羊凝乳酶原	INGKMYPLIISAYTSQDCGFCTSGFCGENRHCQWILGDVFIREYYSVDFRANNVGLAKAI	381
consensus	ing r d i y p l s i s a y t s k d c g f c t s g f c g d n s e l w i l g d v f i r e y y s v d f r a n n v g l a k a i	

图 2 双峰驼凝乳酶原以及其它动物种凝乳酶原的氨基酸序列比对

Fig.2 Homology analysis of camelus bactrianus prochymosin amino acid sequences and those from other animal species

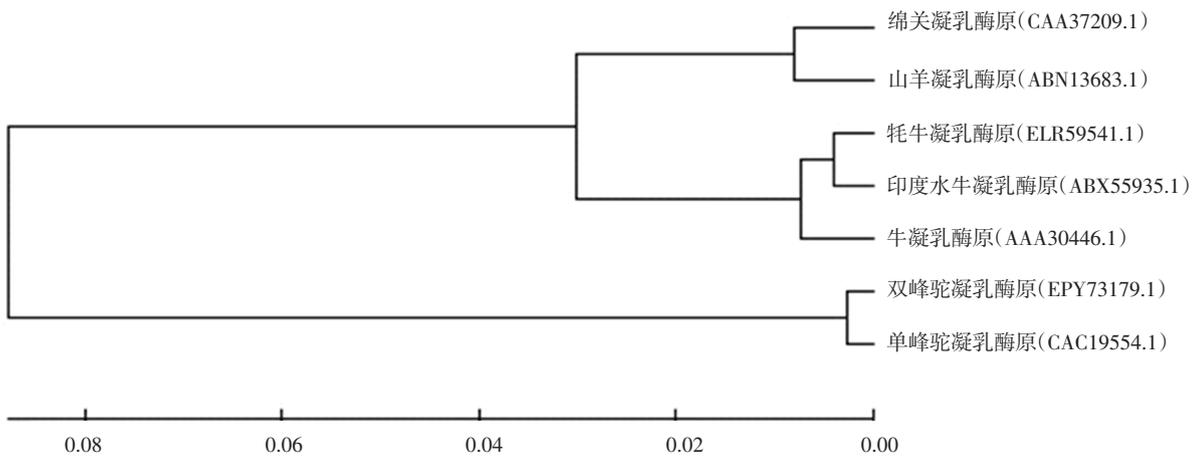


图 3 不同物种基于凝乳酶原氨基酸序列的系统进化树

Fig.3 Phylogenetic tree of different species based on amino acid of prochymosin in camelus bactrianus

2.3 双峰驼凝乳酶原氨基酸一级结构及理化性质预测分析

通过 ExPASyProtParam 预测双峰驼凝乳酶原基因编码的氨基酸序列的组成和理化性质,并与单峰驼、牛等动物凝乳酶原基因的一级结构进行比较,结

果见表 2。结果显示,凝乳酶原基因均编码 381 个氨基酸,并且其带正、负电荷氨基酸数目相差甚微,分子量差异很小,等电点在 4.8 到 5.9 之间,其半衰期均大于 10 小时,表明其有较高稳定性,根据定义当不稳定系数分值小于 40 时,预测蛋白质比较稳定;

当不稳定系数分值大于 40 时,则该蛋白质不稳定^[7],因此七种凝乳酶原均为较稳定的蛋白。

表 2 不同物种凝乳酶原的氨基酸理化性质的比较

Table 2 The physical and chemical characteristics of amino acid sequences of different prochymosin

名称	氨基酸残基数(个)	正负电荷氨基酸数		分子量(Da)	等电点(PI)	半衰期(h)	蛋白质不稳定指数
		+	-				
双峰驼	381	32	36	42 041.6	5.90	>10	34.47
单峰驼	381	32	36	42 082.6	5.90	>10	35.10
牦牛	381	25	36	42 184.6	4.90	>10	39.22
牛	381	24	36	42 195.6	4.90	>10	36.80
山羊	381	24	38	42 092.5	4.86	>10	33.35
印度水牛	381	25	36	42 217.7	4.83	>10	38.27
绵羊	381	26	38	42 074.5	4.80	>10	37.33

2.3 双峰驼凝乳酶原的保守结构域分析

通过 NCBI 保守结构域数据库 (Conserved Domain Database, CDD), 分析双峰驼凝乳酶原中的

保守结构域(见图 4), 结果显示双峰驼凝乳酶原属于胃蛋白酶 A 超家族, 含有多个胃蛋白酶保守结构域。

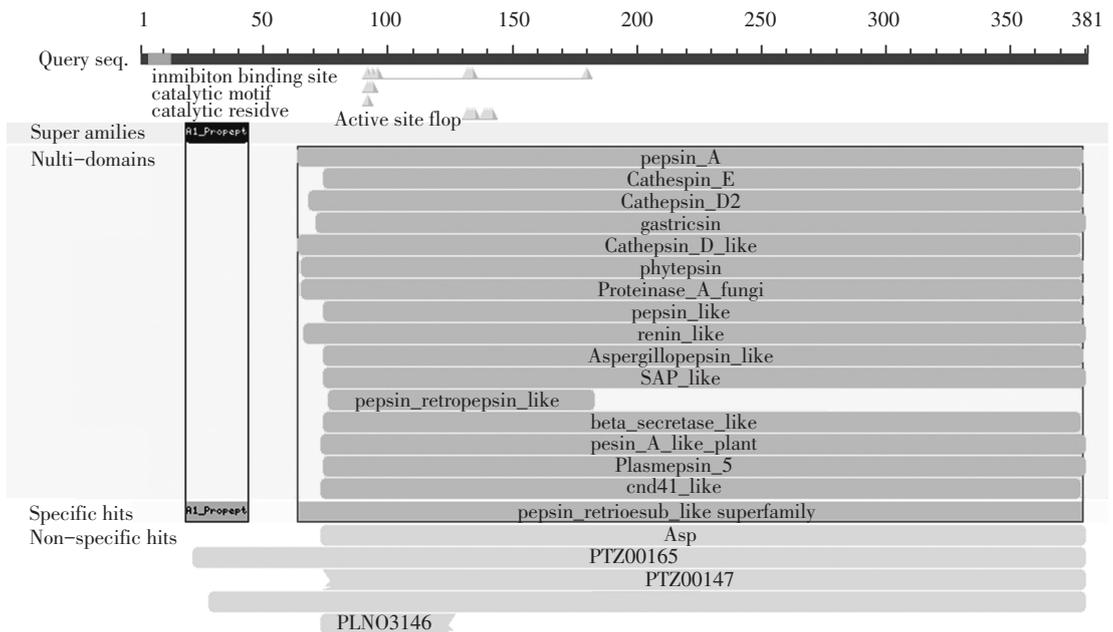


图 4 预测的双峰驼凝乳酶原保守结构域

Fig.4 Conserved domain prediction of prochymosin in camelus bactrianus

2.4 双峰驼凝乳酶的亚细胞定位

应用 Psort 工具预测双峰驼凝乳酶的亚细胞定位, 结果见表 2, 其可能定位于内质网(膜)、过氧化物酶体、细胞质膜和内质网(腔)。表 3 显示双峰驼凝乳酶定位于内质网(膜)的概率较高, 为 0.820, 定位于过氧化物酶体、细胞质膜和内质网(腔)的概率甚微。

表 3 双峰驼凝乳酶原的亚细胞定位

Table 3 Subcellular location of prochymosin of camelus bactrianus

亚细胞定位	概率
内质网(膜) endoplasmic reticulum (membrane)	0.820
过氧化物酶体(peroxisome)	0.219
细胞质膜(plasma membrane)	0.190
内质网(腔) endoplasmic reticulum (lumen)	0.100

2.5 双峰驼凝乳酶原信号肽预测和分析

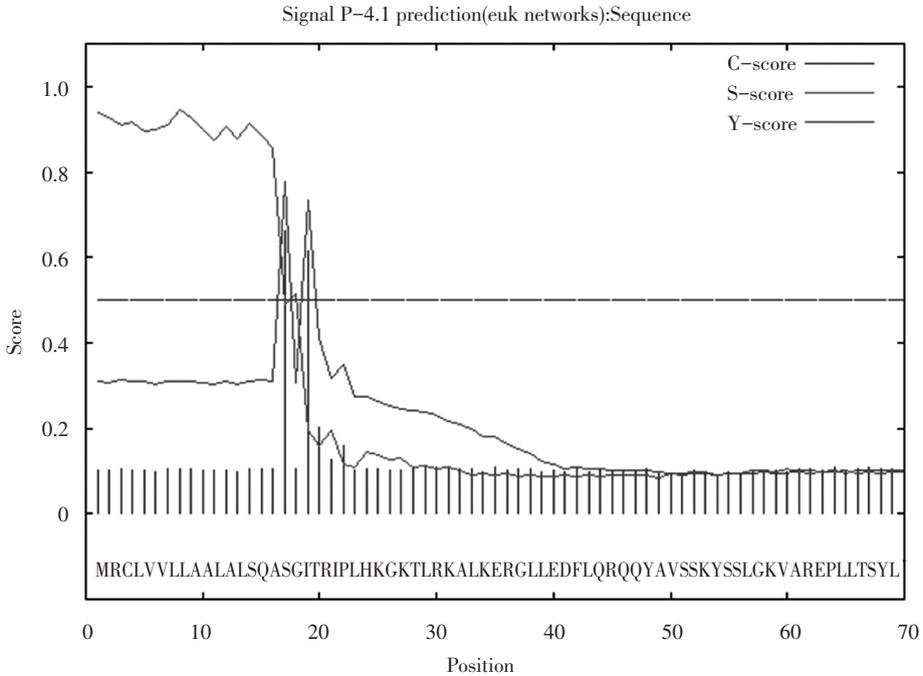


图 5 双峰驼凝乳酶信号肽预测

Fig.5 Signal P prediction of prochymosin in *camelus bactrianus*

注:C score:原始剪切位点分值;S score:信号肽分值;Y score:综合剪切位点分值。

Notes:C score:Scores of Putative cleavage site;S score:Scores of Signal peptide;Y score:Scores of Synthesis cleavage site.

通过 Signal P4.0Server 软件预测双峰驼凝乳酶信号肽(见图 5),结果显示:第 17 位丝氨酸残基具有最高的原始剪切位点分值 0.661,第 8 位亮氨酸残基具有最高的信号肽分值 0.946,第 17 位丝氨酸残基具有最高的综合剪切位点分值 0.778。由于氨基酸残基的原始剪切位点和信号肽的分值均,可推测双峰驼凝乳酶基因所编码的蛋白存在信号肽,属于分泌蛋白,与 Foltman 关于牛凝乳酶的研究一致,该研究表明牛凝乳酶存在一个 16 个氨基酸组成的前

导信号序列,在通过细胞膜分泌中很重要^[15]。

2.6 双峰驼凝乳酶原跨膜结构域预测和分析

跨膜结构域是膜内在蛋白与膜脂相结合的主要部位,它固着于细胞膜上起锚定作用,一般由 20 个左右的疏水氨基酸组成。通过 Expsy 软件中的 TMHMM Server v.2.0 工具预测双峰驼凝乳酶原基因的跨膜螺旋区(见图 6)。预测结果显示,双峰驼凝乳酶原基因氨基酸序列中不存在跨膜螺旋区,蛋白质整体 381 个氨基酸残基全部在膜外,为非跨膜蛋白。

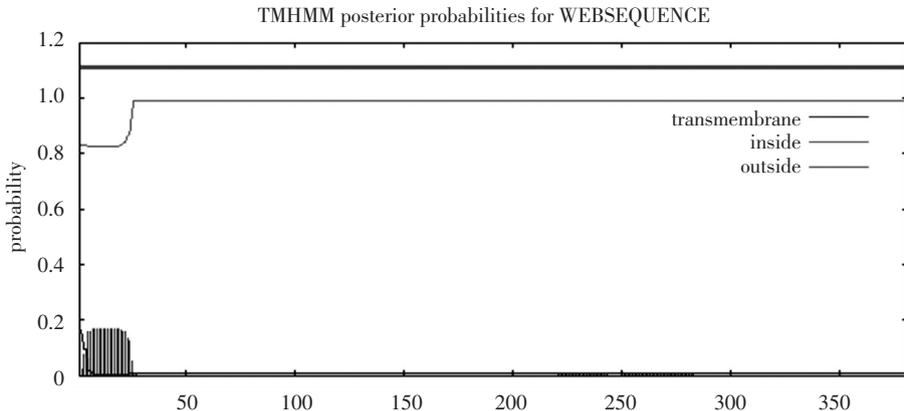


图 6 双峰驼凝乳酶原跨膜结构域预测

Fig.6 Predicted transmembrane domain of prochymosin in *camelus bactrianus*

2.7 双峰驼凝乳酶原氨基酸序列的疏水性\亲水性的预测和分析

通过 ExPASyProtScale 在线软件,对双峰驼凝乳酶原基因所编码的氨基酸序列的疏水性\亲水性进行预测,其结果见图7,分析双峰驼凝乳酶的结果显示,双峰驼凝乳酶,最高的分值3.300,为第7位的亮氨酸,疏水性最强;最低的分值-2.789,为第115和

116位的精氨酸和苯丙氨酸,亲水性最强。GRAVY值是测蛋白质的疏水性标准,通常 GRAVY 值分布在 2 至-2 之间,如为正值,则该蛋白为疏水蛋白;如为负值,则为亲水蛋白。其 GRAVY 值为 -0.096,从整体来看,亲水性氨基酸残基的分布在整条肽链上,且在整体上多于疏水性氨基酸残基,推测双峰驼凝乳酶是一种可溶性蛋白。

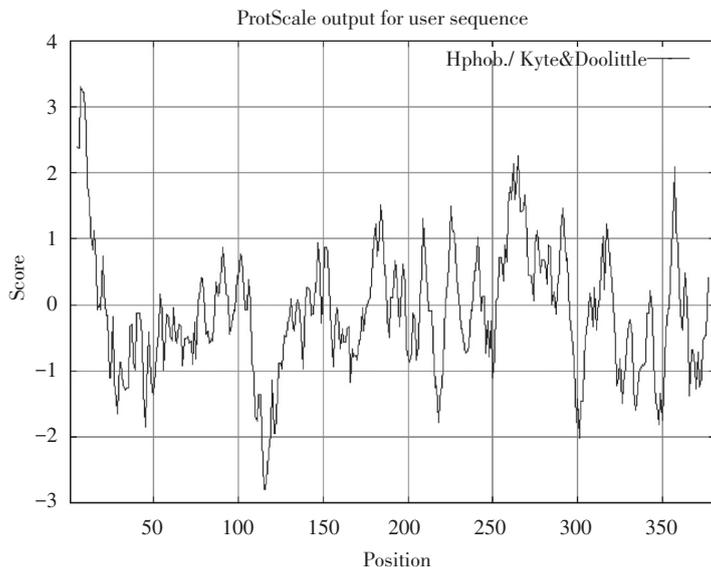


图7 双峰驼凝乳酶原氨基酸序列的疏水性\亲水性预测

Fig.7 Predicted hydrophobicity/hydrophilicity of the amino acidsequence of *camelus bactrianus prochymosin*

2.8 双峰驼凝乳酶原二级结构预测和分析

蛋白质的二级结构主要包括 α 螺旋,延伸链, β 转角,无规则卷曲以及模序等蛋白质局部结构组件。通过 GOR4 方法对双峰驼凝乳酶原二级结构进行在线分析(见表4,图8),在双峰驼凝乳酶原二级结构中,无规则卷曲占的比例最高为 53.02%, α 螺旋占 17.32%,延伸链占 29.66%。由此可推测,无规卷曲是双峰驼凝乳酶原二级结构中最大量的结构元件, α 螺旋和延伸链分散于整个蛋白质中。

表4 双峰驼凝乳酶二级结构典型构象统计结果

Table 4 Secondary structure prediction of *camelus bactrianus prochymosin*

二级结构类型	氨基酸残基数目(个)	百分比(%)
α 螺旋	66	17.32%
延伸链	113	29.66%
无规卷曲	202	53.02%

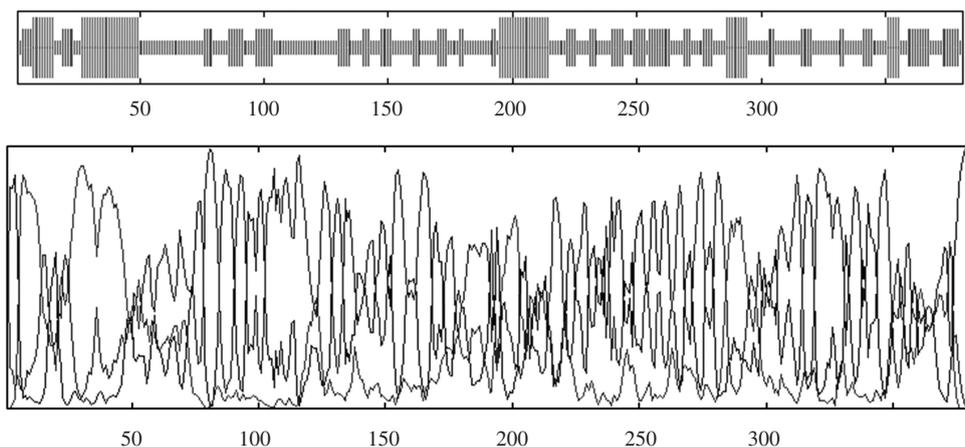


图8 双峰驼凝乳酶原二级结构预测和分析

Fig.8 Predicted secondary structure of *camelus bactrianus prochymosin*

2.9 双峰驼凝乳酶原的活性位点的分析

通过 NPS 的 PROSCAN 对双峰驼凝乳酶原进行活性位点的分析(见表5),结果显示,双峰驼凝乳酶原有6类活性位点,分别是 N-糖基化位点、cAMP 和 cGMP 依

赖性蛋白激酶磷酸化位点、蛋白酶 C 磷酸化位点、酪蛋白激酶 II 磷酸化、N - 十四酰化位点、真核生物和病毒天门冬氨酰蛋白酶活性位点。这可能与凝乳酶的分泌以及自剪切成为有活性的酶等生物学功能相关。

表5 双峰驼凝乳酶原在 NPS 中的活性位点预测分析

Table 5 Scanning of *camelus bactrianus* prochymosin for site/signatures against proscan database in NPS

活性位点名称	序列号	膜体类型	氨基酸序列	概率值
N-糖基化位点	PS00001	$N-\{P\}-[ST]-\{P\}$.	158-161 NQTV. 349-352 NNSE.	5.138e-03
cAMP 和 cGMP 依赖性蛋白激酶磷酸化位点	PS00004	$[RK](2)-x-[ST]$.	119-122 RKSS 29-31 TLR. 52-54 SSK.	1.572e-03
蛋白酶 C 磷酸化位点	PS00005	$[ST]-x-[RK]$.	123-125 TFR. 310-312 SLR. 334-336 TSK. 68-71 SYLD. 173-176 TYSE.	1.423e-02
酪蛋白激酶 II 磷酸化位点	PS00006	$[ST]-x(2)-[DE]$	175-178 SEFD. 334-337 TSKD. 366-369 SVFD. 138-143 GSMEGF 221-226 GSMLTL 260-265 GVAVAC	1.482e-02
N - 十四酰化位点	PS00008	$G-\{EDRKHPFYW\}-x(2)-[STAGCN]-\{P\}$	267-272 GGCQAI 295-300 GATENR 309-314 GSLRSM 376-381 GLAKAI 89-100	1.397e-02
真核生物和病毒天门冬氨酰蛋白酶活性位点	PS00141	$[LIVMFGAC]-[LIVMTADN]-[LIVFSA]-D-[ST]-G-[STAV]-[STAPDENQ]-\{GQ\}-[LIVMFSTN]C-\{EGK\}-[LIVMFGTA]$.	VVFDTGSSDLWV. 271-282 AILDGTGTSVLF.	7.894e-07

3 结 论

骆驼乳具有多种生理活性物质,营养价值远远高于一般的动物乳制品。这些独特成分使其在防病抗病方面起着积极的作用^[16]。Jesper 等研究的结果表明,单峰驼凝乳酶作为一种替代品已经被成功地销售,因为其能提供较高的干酪产量和制备的干酪苦味少,风味很好,并且蛋白水解的活性较低^[17]。Bansal N.等研究的结果表明,单峰驼凝乳酶适合用于契达奶酪的制作,不仅蛋白水解水平较低,并且具有有良好的风味^[18]。目前骆驼凝乳酶研究较少,对于双峰驼凝乳酶原基因的生物信息分析尚未见报道,因此骆驼凝乳酶开发利用前景广阔。

本研究运用生物信息学分析软件,系统地研究了双峰驼凝乳酶原的基因序列和氨基酸序列。结果显示,双峰驼凝乳酶原基因全长 1 146 bp,编码 381 个氨基酸,属于胃蛋白酶 A 超家族,可能定位于内质

网(膜)的稳定亲水性蛋白,不含跨膜结构域。无规卷曲是其二级结构中最大量的结构元件, α 螺旋和延伸链分散于整个蛋白质中,编码蛋白有6类活性位点,主要分为 N-糖基化位点,蛋白激酶磷酸化位点,蛋白酶磷酸化位点,N-十四酰化位点和蛋白酶活性位点。通过分析双峰驼凝乳酶原基因及其编码蛋白质的特征为研究其凝乳特性及其条件提供理论参考,对凝乳酶凝乳功能的研究具有重要的意义。

参考文献(References)

- [1] 钟继才.凝乳酶在干酪生产中的应用[J].中国乳品工业,2006,6(1):54-56.
ZHONG Jicai. Application of rennin in cheese making [J]. China Dairy Industry, 2006, 6 (1):54-56.
- [2] WILLIAMS M G, WILSHER J, NUGENT P, et al. Mutagenesis, biochemical characterization and X-ray structural analysis of point mutants of bovine chymosin [J]. Protein Eng, 1997, 10(9): 991-997.

- [3] 周俊清.产高效凝乳酶菌株获得方法的探讨[J].食品科学,2005,26(3):253-256.
ZHOU Junqing. Discussion on method of yielding high effect chymosin-producing strain [J]. Food Science, 2005,26(3):253-256.
- [4] KAPPELER S R, BRINK H M V D, RAHBK-NIELSEN H, et al. Characterization of recombinant camel chymosin reveals superior properties for the coagulation of bovine and camel milk [J]. Biochemical and Biophysical Research Communications, 2006, 342(2): 647-654.
- [5] JANGHOLM JENSEN J, MØLGAARD A, NAVARRO POULSEN J C, et al. Camel and bovine chymosin; the relationship between their structures and cheese-making properties [J]. Biological Crystallography, 2013, D69, 901-913.
- [6] BROE M L, BRINK J V D, HARBOE M R. Marianne Kirsten Harboe. Coagulation of Milk [P]. United States, US2011/0008492A1, 2011-1-13.
- [7] 陈珊珊,郭晋隆,李国印,等. 甘蔗过氧化氢酶基因的电子克隆及生物信息学分析[J]. 生物信息学,2012,10(1):65-70.
CHEN Shanshan, GUO Jinlong, LI Guoyin, et al. Electronic cloning and characterization of CAT gene from *Saccharum officinarum* using bioinformatics tools [J]. Chinese Journal of Bioinformatics, 2012, 10(1):65-70.
- [8] GASTEIGER E, HOOGLAND C, GATTIKER A, et al. Protein identification and analysis tools on the ExPASy server [J]. The Proteomics Protocols Handbook, 2005, 571-607.
- [9] PETERSEN T N, BRUNAK S, HEIJINE G V, et al. Signal P 4. 0: discriminating signal peptides from transmembrane regions [J]. Nature Methods, 2011, 8(10):785-786.
- [10] NAIR R, ROST B. Better prediction of sub-cellular localization by combing evolutionary and structural information [J]. PROTEINS: Structure, Function, and Genetics, 2003,53: 917-930.
- [11] 郝爱平,熊超超.小鼠 Pitx2 基因的生物信息学分析[J]. 生物技术,2012,22(6):13-16.
HAO Aiping, XIONG Chaochao. Bioinformatics analysis of pitx2 gene in mouse [J]. Biotechnology, 2012, 22(6): 13-16.
- [12] MA Tingting, ZHOU Yijun, GAO Fei, et al. Bioinformatical analysis on sequence and functions of peroxidase in *Arabidopsis* [J]. Agricultural Biotechnology, 2012, 13(8): 1645-1650.
- [13] TAMNRA R, DUDLEY J, NEI M, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 [J]. Molecular Biology Evolution, 2007, 24(8): 1596-1599.
- [14] 蔡明,高贝,张道远.齿肋赤藓热激蛋白 60 基因的电子克隆和生物信息学分析[J]. 生物信息学,2013,11(3):216-223.
CAI Ming, GAO Bei, ZHANG Daoyuan. Electronic cloning and characterization of HSP60 gene from *Syntrichia caninervis* using bioinformatics tool [J]. Chinese Journal of Bioinformatics, 2013, 11(3):216-223.
- [15] FOLTMANN B, PEDERSEN V B, JACOBSEN H, et al. The complete amino acid sequence of prochymosin [J]. Proc Natl Acad Sci USA, 1977, 74: 2321-2324.
- [16] 何俊霞,哈斯苏荣,那仁巴图,等.驼乳医疗保健作用的研究进展 II [J].食品科学,2009,30(23):504-507.
HE Junxia, HASISURONG, NARENBATU, et al. Research progress in health care function of camel milk [J]. Food Science, 2009,30(23):504-507.
- [17] SØRENSEN J, PALMER D S, QVIST K B, et al. Initial Stage of Cheese Production: A Molecular Modeling Study of Bovine and Camel Chymosin Complexed with Peptides from the Chymosin-Sensitive Region of K-Casein [J]. Journal of Agricultural and Food Chemistry, 2011, 59: 5636-5647.
- [18] BANSAL N, DRAKE M A D A, PIRAINO P, et al. Suitability of Recombinant Camel (*Camelus Dromedarius*) Chymosin as a Coagulant for Cheddar Cheese [J]. International Dairy Journal, 2009,19:510-517.