

文章编号: 1674 - 5566(2011)01 - 0008 - 07

## 大珠母贝珍珠质相关基因 *N36* 和 *N45* 的克隆与序列特征分析

王玉梅<sup>1,2</sup>, 夏建红<sup>1</sup>, 唐仁生<sup>1</sup>, 喻达辉<sup>1</sup>

(1. 中国水产科学研究院南海水产研究所, 农业部海水养殖生态与质量控制重点开放实验室, 广东广州 510300;  
2. 上海海洋大学水产与生命学院, 上海 201306)

**摘要:** 采用同源克隆从大珠母贝外套膜中扩增到 2 种编码 N66 类似蛋白的基因, 分别命名为 *N36* 和 *N45*。*N36* 的开放阅读框为 987 bp, 编码 329 个氨基酸, 推测分子量为 35.7 ku。*N45* 的开放阅读框为 1 164 bp, 编码 387 个氨基酸, 推测分子量为 44.6 ku。二者都富含甘氨酸、天冬酰胺和天冬氨酸。*N36* 蛋白不含信号肽, *N45* 蛋白在 N-端有一段 22 个氨基酸的信号肽。*N36* 蛋白和 *N45* 蛋白分别有 2 个和 1 个糖基化位点, 分别有 13 个和 16 个潜在的磷酸化位点, 它们的二级结构都由  $\alpha$  螺旋、 $\beta$  折叠和无规卷曲组成。系统进化分析显示 *N36* 和 *N45* 与 N66 进化关系最近, 而与其他 nacrein 蛋白则距离较远。与合浦珠母贝的 nacrein 和大珠母贝的 N66 类似, *N36* 和 *N45* 蛋白都包含 2 种功能结构域: 1 个  $\alpha$ -碳酸酐酶结构域和 1 个 Gly-Xaa-Asn (Xaa = Asp, Asn or Glu) 重复结构域。推测 *N36* 和 *N45* 与珍珠层碳酸钙晶体的形成有关。

**研究亮点:** 珍珠形成是一种特殊的生物矿化作用, 所形成的珍珠层具有高硬度的机械性能, 且具有特殊的结构和良好的医药价值, 但其形成机制以及组成成份并不完全清楚。本研究克隆了 2 个与珍珠形成相关的新基因, 并对其序列特征进行了分析, 为进一步研究其结构和功能奠定了重要基础。

**关键词:** 大珠母贝; 珍珠质相关基因;  $\alpha$ -碳酸酐酶结构域; Gly-Xaa-Asn 重复结构域

**中图分类号:** S 917

**文献标识码:** A

珍珠质(nacre)是珍珠贝外套膜分泌的有机质(主要是蛋白质)调控  $\text{CaCO}_3$  形成的文石晶体复合物。最早报道的与珍珠质形成相关的基因是合浦珠母贝(*Pinctada fucata*)的 nacrein 基因<sup>[1]</sup>, nacrein 蛋白是一种水溶性基质蛋白, 有 2 个功能结构域——碳酸酐酶结构域和 Gly-Xaa-Asn 重复结构域。碳酸酐酶结构域能够催化  $\text{CO}_2$  的水合反应, Gly-Xaa-Asn 重复结构域则具有  $\text{Ca}^{2+}$  结合活性, 这类蛋白可能参与珍珠层碳酸钙晶体的形成, 对碳酸钙晶体的沉淀具有负调节作用<sup>[2]</sup>。后来又报道了一些与 nacrein 类似的基质蛋白, 包括大珠母贝(*Pinctada maxima*)的 N66<sup>[3-5]</sup> 和大珠母贝、合浦珠母贝、虾夷扇贝(*Mizuhopecten yessoensis*)、岩牡蛎(*Crassostrea*

*nippona*)的几种 nacrein-like proteins<sup>[3]</sup>, 它们都具有碳酸酐酶结构域和 Gly-Xaa-Asn 重复结构域。本研究以大珠母贝外套膜组织为材料, 对 nacrein 类基因进行同源克隆和序列特征分析, 以期探讨珍珠质形成的分子生物学机制提供基础资料。

### 1 材料与方法

#### 1.1 材料

大珠母贝采集于南海水产研究所海南三亚热带水产研究开发中心, 解剖后取外套膜, 用液氮速冻后保存于  $-80\text{ }^\circ\text{C}$  低温冰箱。

#### 1.2 总 RNA 的提取与 cDNA 的合成

取大珠母贝外套膜组织, 加液氮浸泡并充分研磨, 按照 SV Total RNA Isolation System

收稿日期: 2010-01-12 修回日期: 2010-05-02

基金项目: 国家“八六三”高技术研究发展计划项目(2006AA10A415); 广东省科技兴海项目(A200701C02, A200899C04); 中央级公益性科研院所基本科研业务费专项资金项目(2007TS07)

作者简介: 王玉梅(1982-), 女, 硕士研究生, 专业方向为海洋生物技术。E-mail: yumei20081982@163.com

通讯作者: 喻达辉, E-mail: pearlydh@163.com

(Promega)操作手册提取总 RNA。按照 SMART™ PCR cDNA Synthesis Kit (Clontech) 操作手册,合成大珠母贝外套膜 cDNA。

### 1.3 N66L 基因的克隆

根据已知的大珠母贝的 N66 基因<sup>[4]</sup> (GenBank 登录号 AB032612)、nacrein-like protein 基因<sup>[3]</sup> (AB252480, AB252479) 和合浦珠母贝的 nacrein 基因<sup>[1]</sup> (D83523) 序列设计合成 2 对引物 N66L-F1 (ATG GAG GCT CAT TTG GTK TTC C)/N66L-R1 (TGT GTK GGA CGT CTR GTT C) 和 N66L-F2 (ATG TGG AGA ATG ACG ACG CT)/N66L-R2 (TAT GGA GTT TTT GTA CAC), 对上述 cDNA 进行 PCR 扩增。PCR 反应总体积为 20  $\mu$ L, 包括 0.2 mmol/L dNTP, 1  $\mu$ mol/L 的引物, 4  $\mu$ mol/L 的 MgCl<sub>2</sub>, 2  $\mu$ L 10  $\times$  buffer, 0.5 U 的 Taq 酶。PCR 程序如下: 95  $^{\circ}$ C 5 min; 95  $^{\circ}$ C 30 s, 57  $^{\circ}$ C 30 s, 72  $^{\circ}$ C 2 min, 30 个循环; 72  $^{\circ}$ C 10 min。

PCR 产物经 1.0% 琼脂糖凝胶电泳后, 用 AxyPrep™ DNA Gel Extraction Kit (AXYGEN) 对目的片段进行纯化, 纯化产物与 pMD18-T 载体 (TaKaRa) 连接, 转化到感受态大肠杆菌 (*Escherichia coli*) Top10, 经 PCR 鉴定后将阳性克隆送往上海生工公司测序。

### 1.4 序列特征与结构分析

用 BLASTX 程序搜索同源基因<sup>[5]</sup>, 用 BLASTP 分析蛋白氨基酸序列的保守域。用 SignalP 3.0 预测信号肽序列<sup>[6]</sup>, 用 Tmpred 预测穿膜结构域<sup>[7]</sup>。用 Prosite 分析蛋白质的结构域<sup>[8]</sup>, 用 Predator 预测其二级结构<sup>[9]</sup>。通过 NetNGlyc 1.0 预测糖基化位点<sup>[10]</sup>, 通过 NetPhos 2.0 预测磷酸化位点<sup>[11]</sup>。用 Clustal X 进行多序列比对<sup>[12]</sup>, 并用 MEGA 3.1 构建进化树<sup>[13]</sup>。

## 2 结果与分析

### 2.1 序列分析

以 N66L-F1/N66L-R1 和 N66L-F2/N66L-R2 为引物, 分别扩增得到 987 bp 和 1 232 bp 的 cDNA 片段。BLAST 分析表明获得的 2 个片段都与 GenBank 中大珠母贝基因 N66 的同源性最高, 因而参照 N66 以蛋白质分子量命名的方式, 分别将它们命名为 N36 (GenBank 登录号 FJ913471) 和 N45 (GenBank 登录号 FJ913472)。N36 (图 1) 包含 987 bp 的开放阅读框, 编码 329 个氨基酸,

推测分子量为 35.7 ku, 理论等电点为 6.35。N45 (图 2) 包含 1 164 bp 的开放阅读框, 编码 387 个氨基酸, 推测分子量为 44.6 ku, 理论等电点为 8.50。氨基酸组成分析表明, 组成 N36 蛋白的主要氨基酸残基为 Asn (22.8%)、Gly (18.2%) 和 Asp (6.7%), 组成 N45 蛋白的主要氨基酸残基为 Gly (8.5%)、Asn (6.5%) 和 Asp (6.5%)。

### 2.2 结构预测

用 PROSITE 搜索蛋白结构域表明, N36 蛋白不含信号肽, N45 蛋白第 1-22 个氨基酸残基为信号肽 (图 2), 表明 N45 蛋白是 1 种分泌型或膜结合型蛋白。但是 N45 蛋白除信号肽外, 不含穿膜结构域, 如疏水性  $\alpha$ -螺旋或  $\beta$ -折叠, 因而 N45 蛋白可能是 1 种分泌型蛋白。蛋白二级结构预测分析表明, N36 包含 3 个  $\alpha$  螺旋 (19.1%)、7 个  $\beta$  折叠 (12.2%) 和无规卷曲 (random coil, 68.7%); N45 包含 5 个  $\alpha$  螺旋 (17.1%)、12 个  $\beta$  折叠 (16.5%) 和无规卷曲 (66.4%)。N36 蛋白和 N45 蛋白各有 1 个  $\alpha$ -碳酸酐酶结构域 ( $\alpha$ -carbonic anhydrases domain), 在 H4 和 H154、H156、H179 分别有 1 个和 3 个锌结合位点 (图 2)。N36 蛋白和 N45 蛋白都有 1 段 Gly-Xaa-Asn (Xaa = Asp, Asn 或 Glu) 重复结构域 (Gly-Xaa-Asn repeat domain) (图 2)。由于 Gly-Xaa-Asn 重复结构域的插入,  $\alpha$ -碳酸酐酶结构域被间隔成 2 个亚结构域。

糖基化位点预测发现, N36 蛋白有 1 个 Asn-Gly-Ser (aa 164-166) 和 1 个 Asn-Glu-Thr (aa 286-288) N-连接的糖基化序列, N45 蛋白有 1 个 Asn-Glu-Thr (aa 354-356) N-连接的糖基化序列。磷酸化位点预测分析表明, N36 蛋白含有 13 个潜在的磷酸化位点, 包括 4 个 Ser 位点 (Ser196, Ser231, Ser301, Ser321)、3 个 Thr 位点 (Thr260, Thr325, Thr329) 和 6 个 Tyr 位点 (Tyr46, Tyr97, Tyr116, Tyr122, Tyr216, Tyr314), N45 蛋白含有 16 个潜在的磷酸化位点, 包括 8 个 Ser 位点 (Ser48, Ser52, Ser87, Ser160, Ser165, Ser168, Ser264, Ser299)、4 个 Thr 位点 (Thr6, Thr174, Thr328, Thr378) 和 4 个 Tyr 位点 (Tyr31, Tyr148, Tyr221, Tyr284)。

### 2.3 同源性比较和系统发生分析

用 BLASTP 在 GenBank 中进行同源搜索, 显示 N36 和 N45 蛋白和大珠母贝的 N66 matrix

protein<sup>[4]</sup>、nacrein-like protein<sup>[3]</sup>, 合浦珠母贝的 nacrein<sup>[1]</sup> 以及 2 种 nacrein-like protein<sup>[3]</sup>, 虾夷扇贝的 nacrein-like protein P1、P2<sup>[3]</sup> 和岩牡蛎的 nacrein-like protein C1、C2<sup>[3]</sup> 具有较高的相似性。

尤其是  $\alpha$ -碳酸酐酶结构域的 36 个活性位点, N36 和 N45 蛋白与其他蛋白显示高度一致性(图 3)<sup>[14]</sup>。除 N36 中有 16 个位点缺失外, N45 只有 3 个氨基酸位点不保守。

```

1  ATGGAGGCTCATTGGTTTTCCATCATGATGATAAAAAGGAAATCAAACCTCCAAGGGTT
1  M E A (H) L V F H H D D K K E I K P P R V
61  AAGTTAGGGGAGTGTACGCTGGTCGTAACAAATTTGTTGCGTTGGAGTCTTTCTAGAG
21  K L G G V Y A G R N K F V V V G V F L E
121  GTGGGTGATGAAGGATACGGTGTGAACCGGACGACGATGAATGTAAGCGCATATTAAG
41  V G D E G Y G D E P D D D E C K R I L K
181  GGTCAATGCGACAACGATGGGACAATGGTAACAATTGTGATAACGGCGACAATGGTAAC
61  G H C D N D G D N G N N C D N G D N G N
241  AACGGCAACAATGGTAACAACGGAAACAATGGTAATGGTAACAACGGTTATAACGGTAAT
81  N G N N G N N G N N G N G N N G Y N G N
301  AACGGTGACAATGGAAACAATGGCAATGGTAATGGTAACAACGGTTATAACGGTAATAAC
101  N G D N G N N G N G N G N N G Y N G N N
361  GGTTACAATGGCAACAACGGAAACAATGGTAATGGTAACAATGGCAATAATGGTAACGGT
121  G Y N G N N G N N G N G N N G N N G N G
421  AACACGGAAATAACGGTGGCAATGGTAACAACGGAAACAATGGTAATGGTAACAATGAA
141  N N G N N G G N G N N G N N G N G N N E
481  AATAATGGTAACGGTAGTAATGGTAACAACGGTGGAAACGAGAACAATGGTAATAACGGT
161  N N G N G S N G N N G G N E N N G N N G
541  GATAACGGTAATGGCGACAATGGTTATAACGGTGATAATGGTAACAGTGACGAGCGACTC
181  D N G N G D N G Y N G D N G N S D E R L
601  AGACGCTGGGATTTGGCAAATGTCCGACGCATGCATGCCGAGCGATATCACTTTAGCGGA
201  R R W D L A N V R R M H A E R Y H F S G
661  AGATGTATGGTCAAAAAGCTAAACGCCTCAGCAGGATTCTCGAATGCGCATATAGACAC
221  R C M V K K A K R L S R I L E C A Y R H
721  AAAAAAGTCAGAGAATTCAAAAGGAATGGAGAAGAAAAAGGTCTTGATGTTGATATTACA
241  K K V R E F K R N G E E K G L D V D I T
781  CCGGAAATGGTTTTACCGCAATGAAATACAGACAATACTATACCTATGAAGGATCTTTG
261  P E M V L P P M K Y R Q Y Y T Y E G S L
841  ACAACCCCTCCTTGCAATGAGACAGTCCTTTGGGTTGTTGAAAAATGCCACGTGCAAGTA
281  T T P P C N E T V L W V V E K C H V Q V
901  TCCAGAAGGGTGCTTGATGCATTGCGGAACGTGGAAGGATATGAGGATGGTACCACGCTG
301  S R R V L D A L R N V E G Y E D G T T L
961  AGCAAGTATGGAAC TAGACGTCCAACACA
321  S K Y G T R R P T

```

图 1 大珠母贝 N36 基因的核苷酸序列和推测的氨基酸序列

Fig. 1 Nucleotide and deduced amino acid sequences of N36 gene from *P. maxima*

注:阴影表示 Gly-Xaa-Asn 重复结构域;粗体表示  $\alpha$ -碳酸酐酶结构域;圆圈表示锌结合位点;ATC 表示起始密码子。

```

1   ATGTGGAGAATGACGACGCTTCTTCACTTGACTGCTCTGCTTGTCTGATTCCATTATGT
1   M W R M T T L L H L T A L L V L I P L C
61  CATTGCGCCTCCATGCACAGGCATGACCATTATATGGACATGGATCAAACCTACCCTAAT
21  H C A S M H R H D H Y M D M D Q T Y P N
121 GGATTGGGATACTGTGAACCTTCAGGTGAAAGCAGCTGTAAAGCCGGATTTAGCTACAAT
41  G L G Y C E P S G E S S C K A G F S Y N
181 AGAGACATATGCCAAGGTCGGTATCATTGGCACACTATATCTAGTTGCTATAAGGCATGT
61  R D I C Q G P Y H W H T I S S C Y K A C
241 GGACATAAAAGGAGACAATCACCAATCAACATTTGGTCACATAAAGCTGTATTCTTACCT
81  G H K R R Q S P I N I W S H K A V F L P
301 TATCTGCCAAGACTGAAATTCAGCCACATATGAAGTCATTGGATACGGATGTGACAAAT
101 Y L P R L K F K P H M K S L D T D V T N
361 CACCAAAATCGTGCCCTGAATTCGAGCCGGAGGACGGAGATAAGCTTCATGTGAAACTA
121 H Q N R A P E F E P E D G D K L H V K L
421 AAGAATCTTGTGATGGACATTATAAATTTACAATCTCCATATTCACAACGGCAAAAGT
141 K N L V D G H Y K F H N L (H) I (H) N G K S
481 AGACGAAAGGGCTCGGAACACAGCGTGAACAGACATTTTACGCCCATGGAGGCTCATTGT
161 R R K G S E H S V N R H F T P M E A (H) L
541 GTGTTCCATCATGATGATAAAAGGAAATCAAACCTCCAAGGGTTAAGTTAGGGGGAGTG
181 V F H H D D K K E I K P P R V K L G G V
601 TACGCTGGTCGTAACAAATTTGTTGTCGTTGGAGTCTTTCTAGAGGTGGGTGATGAAGGA
201 Y A G R N K F V V V G V F L E V G D E G
661 TACGGTGATGAACCGGACGACGATGAATGTAAGCGCATATTAAGGGTCATTGCGGAGAAC
221 Y G D E P D D D E C K R I L K G H C E N
721 AATGGGGACAATGGTAACAACTGTGATAACGGTAATGGCGACAATGGTTATAACGGTGAT
241 N G D N G N N C D N G N G D N G Y N G D
781 AATGGTAACAGTGACGGGCGACTCAGACGCTGGGATTTGGCAAATGTCCGACGCATGCAT
261 N G N S D G R L R R W D L A N V R R M H
841 GCCGAGCGATATCACTTTAGCGGAGGATGTATCGTCAAAAAAGCTAAACGCCTCAGCAGG
281 A E R Y H F S G G C I V K K A K R L S R
901 ATTCTTGAATGCGCATATAGACACAAAAAGTCAGAGAATTCAGAAGGAATGGAGAAGAA
301 I L E C A Y R H K K V R E F R R N G E E
961 AAAGGCTTGTGATGTTGATATTACACCGGAAATGGTTTTACGCCAATGAAATACAGACAT
321 K G L D V D I T P E M V L P P M K Y R H
1021 TACTATACTTATGAAGGATCTTTGTCAACCCCTCCTTGCAATGAGACCGTCTTTGGGTT
341 Y Y T Y E G S L S T P P C N E T V L W V
1081 GTTGA AAATGCCACGTGCAAGTATCCAGAAGGGTGCTTGATGCGTTGCGGAACGTGCAAG
361 V E N A T C K Y P E G C L M R C G T S K
1141 GATATGAAGATGGTACCACGCTGAGCAAGTATGGAACCAGACGTCCACACGAAGAAACA
381 D M K M V P R *
1201 AGCATCCTCTACGTGTGTACAAAACTCCATA

```

图2 大珠母贝 *N45* 基因的核苷酸序列和推测的氨基酸序列

Fig. 2 Nucleotide and deduced amino acid sequences of *N45* gene from *P. maxima*

注:下划线表示信号肽;阴影表示 Gly-Xaa-Asn 重复结构域;粗体表示  $\alpha$ -碳酸酐酶结构域;圆圈表示锌结合位点;ATG 表示起始密码子;TGA 表示终止密码子。

Pf-nacrein	: AGFSYDRSICEGPRYWTISKCFIACGIGQROSPINIVSYDAKFRQRLPKLKEKPHMEKLTTEVTNH	: 67
Pf-nacrein	: -----	: -
Cn-nacrein	: AGFSYDRSICEGPHYWHTISKCFIACGIGQROSPINIVSYDAKFRQRLPKLKEKPHMEKLTTEVTNH	: 67
My-nacrein	: -----QSPINIVSYDAKFRQRLPKLKEKPHMEKLTTEVTNH	: 36
My-nacrein	: AGFSYDRSICEGPHYWHTISKCFIACGIGQROSPINIVSYDAKFRQRLPKLKEKPHMEKLTTEVTNH	: 67
Pm-nacrein	: AGFSYDRSICEGPHYWHTISKWFIACGIGQROSPINIVSYDAKFRQRLPKLKEKPHMEKLTTEVTNH	: 67
Pf-nacrein	: AGFSYDRSICEGPHYWHTISKCFIACGIGQROSPINIVSYDAKFRQRLPKLKEKPHMEKLTTEVTNH	: 67
Cn-nacrein	: -----QSPINIVSYDAKFLRRLPKLKEKPHMEKLTTEVTNH	: 36
Pm-N36	: -----	: -
Pm-N66	: AGFSYNRDQCQGFYHWHHTISSCYKACGHKRRQSPINIWSHKAVFLPYLPRLKEKPHMKSLLDVTNHN	: 67
Pm-N45	: AGFSYNRDQCQGFYHWHHTISSCYKACGHKRRQSPINIWSHKAVFLPYLPRLKEKPHMKSLLDVTNHN	: 67
Pf-nacrein	: QNRAPEPEPEDGENLYVKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDEQT	: 134
Pf-nacrein	: -----VKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDEQT	: 51
Cn-nacrein	: QNRAPEPEPEDGENLYVKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDEQT	: 134
My-nacrein	: QNRAPEPEPEDGENLYVKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDDQT	: 103
My-nacrein	: QNRAPEPEPEDGENLYVKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDDQT	: 134
Pm-nacrein	: QNRAPEPEPEDGENLYVKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDDQT	: 134
Pf-nacrein	: QNRAPEPEPEDGENLYVKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDDQT	: 134
Cn-nacrein	: QNRAPEPEPEDGENLYVKLNNLVDGHYKFNHLELVNNGRTRRRKGEHSHVNGRFTPEEAALVFFHHDEQT	: 103
Pm-N36	: -----MEEAALVFFHHDDKK	: 13
Pm-N66	: QNRAPEPEPEDGDKLVKLNLDGHYKFNHLELVNNGKSRKGEHSHVNRHFTPEEAALVFFHHDDKK	: 134
Pm-N45	: QNRAPEPEPEDGDKLVKLNLDGHYKFNHLELVNNGKSRKGEHSHVNRHFTPEEAALVFFHHDDKK	: 134
Pf-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKHILKGHP-----DNNE	: 189
Pf-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKHILKGHP-----DNNE	: 106
Cn-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKHILKGHP-----DNNE	: 189
My-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKRILKGHP-----DNNE	: 158
My-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKRILKGHP-----DNNE	: 189
Pm-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKRILKGHP-----DNNE	: 189
Pf-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKHILKGHP-----DNNE	: 189
Cn-nacrein	: HFEPTRTKLGGAAPGHNDPVLVGVFLEVGDGDFGDEPDDEECKHILKGHP-----DNNE	: 158
Pm-N36	: EIKPFRVKLGGVYAGRNKEFVVGVFLEVGDGDFGDEPDDEECKRILKGHCNDSDERLRRWDLANVR	: 80
Pm-N66	: EIKPFRVKLGGVYAGRNKEFVVGVFLEVGDGDFGDEPDDEECKRILKGHC--SDGLRRWDLANVR	: 199
Pm-N45	: EIKPFRVKLGGVYAGRNKEFVVGVFLEVGDGDFGDEPDDEECKRILKGHC--SDGLRRWDLANVR	: 199
Pf-nacrein	: NGNGHKH----GCRVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMALPPLKYRHYTTY	: 252
Pf-nacrein	: NGNGHKH----GCLVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMALPPLK-----	: 162
Cn-nacrein	: NGNGHKH----GCRVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMPLPPLKYRHYTTY	: 252
My-nacrein	: NGNGHKH----GCRVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMPLPPLKNRHYTTY	: 221
My-nacrein	: NGNGHKH----GCRVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMPLPPLKNRHYTTY	: 252
Pm-nacrein	: NGNGHKH----GCRVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMALPPLKYRHYTTY	: 252
Pf-nacrein	: NGNGHKH----GCRVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMALPPLKYRHYTTY	: 252
Cn-nacrein	: NGNGHKH----GCRVKKAKHLSRILECAYRNDKVRREFKKVGEEEGLDVHLTPEMPLPPLKYRHYTTY	: 221
Pm-N36	: RMHAERYHFSGRCMVKKAKRLSRILECAYRHKKVREFKRNGBEKGLDVDITPEMVLPPMKYRHYTTY	: 147
Pm-N66	: RMHAERYHFSGGCIVKAKRLSRILECAYRHKKVREFKRNGBEKGLDVDITPEMVLPPMKYRHYTTY	: 266
Pm-N45	: RMHAERYHFSGGCIVKAKRLSRILECAYRHKKVREFKRNGBEKGLDVDITPEMVLPPMKYRHYTTY	: 266
Pf-nacrein	: EGSLTTPPECTESVLMVYQKCHVQVSRRLVHALRNVGEYKDGTTLRKYGTRRPTQKNKVTVYKSF--	: 316
Pf-nacrein	: -----	: -
Cn-nacrein	: EGSLTTPPECTESVLMVYQKCHVQVSRRLVHALRNVGEYKDGTTLRKYGTRRPTQKNKVTVYKSF--	: 316
My-nacrein	: EGSLTTPPECTESVLMVYQKCHVQVSRRLVHALRNVGEY-----	: 259
My-nacrein	: EGSLTTPPECTESVLMVYQKCHVQVSRRLVHALRNVGEYKDGTTLRKYGTRRPTQKNKVTVYKSF--	: 316
Pm-nacrein	: EGSLTTPPECTESVLMVYQKCHVQVSRRLVHALRNVGEYKDGTTLRKYGTRRPTQKNKVTVYKSF--	: 316
Pf-nacrein	: EGSLTTPPECTESVLMVYQKCHVQVSRRLVHALRNVGEYKDGTTLRKYGTRRPTQKNKVTVYKSF--	: 316
Cn-nacrein	: EGSLTTPPECTESVLMVYQKCHVQVSRRLVHALRNVGEY-----	: 259
Pm-N36	: EGSLTTPCNETVLMVYQKCHVQVSRRLDALRNVGEYEDGTTLSKYGTRRPT-----	: 200
Pm-N66	: EGSLTTPCNETVLMVYQKCHVQVSRRLDALRNVGEYEDGTTLSKYGTRRPTQRNKHPLRVYKNS	: 332
Pm-N45	: EGSLTTPCNETVLMVYENATCKYP-----EGCLMRCGTSKDMEMVPR-----	: 309

图3  $\alpha$ -碳酸酐酶结构域的多序列比对Fig. 3 Multiple sequence alignment of  $\alpha$ -carbonic anhydrase domain

注: 阴影表示  $\alpha$ -碳酸酐酶结构域的活性位点, 方框表示锌结合位点; Pf. 合浦珠母贝; Cn. 岩牡蛎; My. 虾夷扇贝; Pm. 大珠母贝。

使用 MEGA 3.1 软件, 采用 NJ (Neighbor-joining) 法构建  $\alpha$ -碳酸酐酶结构域氨基酸序列 (去除了中间的重叠结构域) 的系统进化树, 结果显示, N36 和 N45 与 N66 进化关系最近, 聚在同一分支, 而其他蛋白则位于另一分支, 而且 2 个分支比较长, 遗传距离较远 (图 4)。

### 3 讨论

碳酸酐酶 (Carbonic Anhydrases, CAs) 是一类分布广泛的含锌金属酶, 它能够可逆性地催化二氧化碳的水合反应:  $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+$  [15-16]。根据进化上的不同, CAs 通常分为 5

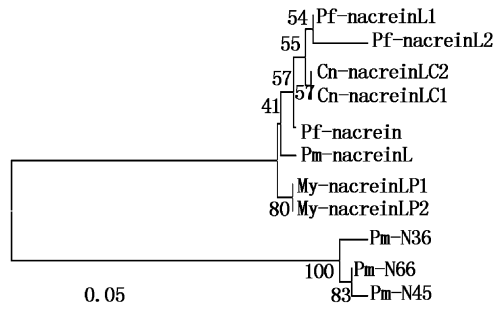


图4 大珠母贝 N36 和 N45 蛋白与其他  $\alpha$ -碳酸酐酶蛋白  $\alpha$ -碳酸酐酶结构域的进化关系

Fig. 4 Phylogenetic analysis of  $\alpha$ -carbonic anhydrase domain of N36, N45 and other  $\alpha$ -carbonic anhydrase proteins

注: Pm. 大珠母贝; Pf. 合浦珠母贝; My. 虾夷扇贝; Cn. 岩牡蛎; nacreinL. nacrein-like protein。

个家族: $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  和  $\epsilon$ 。 $\alpha$ -CAs 是一个主要由反平行的  $\beta$ -折叠片组成的、由两个半球组成的球形分子,表面还分布有一些相对较短的  $\alpha$ -螺旋结构<sup>[17-18]</sup>。与 N66、nacrein 类似,N36 和 N45 的  $\alpha$ -碳酸酐酶结构域都被 Gly-Xaa-Asn 重复结构域间隔成 2 个亚结构域,这 2 个亚结构域可能作为组成该分子球形结构的两个半球。N36 蛋白的两个  $\alpha$ -碳酸酐酶亚结构域分别含有 4 个和 3 个  $\beta$ -折叠片,N45 分别含有 8 个和 4 个,而 N66(19 个)和 nacrein(18 个)则含有更多,推测 N45 比 N36 形成的三级结构更利于碳酸酐酶结构域发挥催化作用,N66 和 nacrein 则比 N45 更有利。 $\alpha$ -CAs 的活性位点含有 3 个保守的 His 残基, $\alpha$ -CAs 通过 3 个保守的 His 结合 1 个  $Zn^{2+}$ 。N36 蛋白只含 1 个保守的 His 残基,而 N45、N66 和 nacrein 则都含 3 个。除 3 个与  $Zn^{2+}$  结合的 His(作为直接受体通过季铵氮原子与  $Zn^{2+}$  配位结合)外,酸酐酶结构域还有 33 个保守的氨基酸位点<sup>[1,14]</sup>,这些位点有的作为非直接配体形成氢键,有的组成的一个大的疏水袋以结合底物  $CO_2$ ,有的作为质子转运体再生有催化活性的  $EZnOH^-$ 。这 36 个活性位点的高度保守性显示了 N36 和 N45 蛋白与其他 nacrein 样蛋白在碳酸酐酶活性上的高度相似性,而它们保守性的高低可能反映了其碳酸酐酶活性的强弱。无论是三级结构预测还是活性位点分析都显示 N66 在这些蛋白中的碳酸酐酶活性是最强的,nacrein 仅次于 N66 而强于 N45 蛋白,N36 则低于三者。

KONO 等<sup>[4]</sup>认为 Gly-Xaa-Asn 重复结构域很可能参与晶体的生长,而且重复序列 Gly-Xaa-Asn 越长,与钙、基质组份和晶体的结合能力就越强,从而影响珍珠层的形成速度和质量差异。N36 蛋白的 Gly-Xaa-Asn 重复序列长 129 个氨基酸,在 nacrein 样蛋白中仅次于 N66,而 N45 蛋白的仅为 24 个氨基酸,推测 N36 蛋白比 N45 蛋白对  $Ca^{2+}$  有更强的结合能力。

磷酸基和糖链可能和生物矿化过程有密切关系<sup>[19]</sup>。N36 和 N45 蛋白分别有 13 个和 16 个潜在的磷酸化位点,氨基酸残基的磷酸化可能改变蛋白质的三级结构,并且参与蛋白质和钙的相互作用。大多数珍珠质水溶性基质蛋白含有一定数目的糖链,N36 和 N45 蛋白也分别有 2 个和 1 个糖基化位点,推测二者为糖蛋白。TAKAKURA 等<sup>[20]</sup>发现 nacrein 是一种糖蛋白,其糖链中的亚硫酸盐和唾液酸可能为促进  $Ca^{2+}$  的摄取提供必需的负电荷。N36 和 N45 蛋白的糖链可能也有类似的作用。N45 蛋白在 N-端有一段信号肽序列,是一种分泌型蛋白,推测其由大珠母贝外套膜合成后分泌至外套腔,从而参与贝壳或珍珠质的形成。

#### 参考文献:

- [1] MIYAMOTO H, MIYASHITA T, OKUSHIMA M A, *et al.* A carbonic anhydrase from the nacreous layer in oyster pearls [J]. *Proc Natl Acad Sci USA*, 1996, 93(18): 9657-9660.
- [2] MIYAMOTO H, MIYOSHI F, KOHNO J. The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata* [J]. *Zool Sci*, 2005, 22: 311-315.
- [3] NORIZUKI M, SAMATA T. Distribution and function of the nacrein-related proteins inferred from structural analysis [J]. *Mar Biotechnol*, 2008, 10(3): 234-241.
- [4] KONO M, HAYASHI N, SAMATA T. Molecular mechanism of the nacreous layer formation in *Pinctada maxima* [J]. *Biochem and Biophys Res Commun*, 2000, 269(1): 213-218.
- [5] WHEELER D L, CHAPPER C, LASH A E, *et al.* Database resources of the National Center for Biotechnology Information [J]. *Nucl Acids Res*, 2000, 28(1): 10-14.
- [6] EMANUELSSON O, BRUNAK S, von HEIJNE G, *et al.* Locating proteins in the cell using TargetP, SignalP and related tools [J]. *Nature protocols*, 2007, 2(4): 953-971.
- [7] HOFMANN K, STOFFEL W. TMbase-a database of membrane spanning protein segments [J]. *Biol Chem Hoppe-Seyler*,

- 1993,374: 166 – 170.
- [8] BAIROCH A , BUCHER P, HOFMANN K. The Prosite database, its status in 1997[J]. Nucl Acids Res, 1997, 25(1): 217 – 221.
- [9] JONES D T. Protein secondary structure prediction based on position-specific scoring matrices[J]. J Mol Biol, 1999, 292(2): 195 – 202.
- [10] BLOM N, GAMMELTOFT S, BRUNAK S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites[J]. J Mol Biol, 1999, 294(5): 1351 – 1362.
- [11] BLOM N, GAMMELTOFT S, BRUNAK S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites[J]. J Mol Biol, 1999, 294(5): 1351 – 1362.
- [12] THOMPSON J D, GIBSON T J, PLEWNIAC F, *et al.* The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools[J]. Nucl Acids Res, 1997, 25(24): 4876 – 4882.
- [13] KUMAR S, GADAGKAR S R. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences [J]. Genetics, 2001, 158(3): 1321 – 1327.
- [14] TASHIAN R E. The carbonic anhydrases: widening perspectives on their evolution, expression, and function [J]. BioEssays, 1989, 10: 186 – 192.
- [15] KELLIN D, MANN T. Carbonic anhydrase, purification and nature of the enzyme[J]. Biochem J, 1940, 34: 1163 – 1176.
- [16] HEWETT-EMMETT D, TASHIAN R E. Functional diversity, con-servation, and convergence in the evolution of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -carbonic anhydrase gene families [J]. Mol Phylogenet Evol, 1996, 5: 50 – 77.
- [17] LINDSKOG S. Structure and mechanism of carbonic anhydrase [J]. Pharmacol Ther, 1997, 74: 1 – 20.
- [18] HUNT J A, AHMED M, FIERKE C A. Metal binding specificity in carbonic anhydrase is influenced by conserved hydrophobic core residues [J]. Biochemistry, 1999, 38(28): 9054 – 9062.
- [19] BORBAS J E, WHEELER A P, SIKES C S. Molluscan shell matrix phosphoproteins: correlation of degree of phosphorylation to shell mineral microstructure and to *in vitro* regulation of mineralization[J]. J Exp Zool, 1991, 258(1): 1 – 13.
- [20] TAKAKURA D, NORIZUKI M, ISHIKAWA F, *et al.* Isolation and characterization of the N-linked oligosaccharides in nacrein from *Pinctada fucata*[J]. Mar Biotechnol, 2008, 10(3): 290 – 296.

## Cloning and characterization of nacre-related genes in silver-lip pearl oyster *Pinctada maxima*

WANG Yu-mei<sup>1,2</sup>, XIA Jian-hong<sup>1</sup>, TANG Ren-sheng<sup>1</sup>, YU Da-hui<sup>1</sup>

(1. Key Laboratory of Mariculture, Ecology and Quality Control, Ministry of Agriculture, South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510300, China; 2. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China)

**Abstract:** We have cloned two cDNAs that encode N66-like proteins in *Pinctada maxima* and are named *N36* and *N45*, respectively. The open reading frame (ORF) of *N36* was 987 bp long, encoding 329 amino acids with molecular mass estimated to be 35.7 kD while the ORF of *N45* was 1 164 bp, encoding 387 amino acids with estimated molecular mass of 44.6 kD. Both proteins were rich in Gly, Asn and Asp residues. There was no signal peptide in *N36* protein while the amino acid residues 1 – 22 in *N45* showed a typical signal peptide. *N36* and *N45* contained 2 and 1 putative glycosylation sequences, 9 and 16 putative phosphorylation sites, respectively. Their secondary structure consisted of  $\alpha$ -helix,  $\beta$ -sheet and coil. Phylogenetic analysis indicated that *N36* and *N45* were more closely related to N66 than other nacrein-like proteins. Similar to nacrein in *P. fucata* and N66 in *P. maxima*, *N36* and *N45* proteins contained two functional domains:  $\alpha$ -carbonic anhydrase domain and Gly-Xaa-Asn (Xaa = Asp, Asn or Glu) repeat domain, suggesting that both *N36* and *N45* are likely to participate in calcium carbonate crystal formation in nacreous layer.

**Key words:** *Pinctada maxima*; nacre-related gene;  $\alpha$ -carbonic anhydrase domain; Gly-Xaa-Asn repeat domain