

[19]中华人民共和国国家知识产权局

[51]Int. Cl<sup>6</sup>

C07K 14/585

C12N 15/16 A61K 38/23

# [12] 发明专利申请公开说明书

[21] 申请号 97193123.2

[43]公开日 1999年5月26日

[11]公开号 CN 1217724A

[22]申请日 97.2.4 [21]申请号 97193123.2

[30]优先权

[32]96.2.6 [33]US[31]08/595,868

[86]国际申请 PCT/US97/01652 97.2.4

[87]国际公布 WO97/29127 英 97.8.14

[85]进入国家阶段日期 98.9.16

[71]申请人 柏内布拉斯加公司

地址 美国内布拉斯加州

[72]发明人 F·W·瓦格内 J·S·斯多特

D·B·亨里克森 B·E·帕特里德格

B·霍姆斯特

J·A·弗兰克

[74]专利代理机构 中国国际贸易促进委员会专利商标事  
务所

代理人 陈文平

权利要求书 5 页 说明书 63 页 附图页数 12 页

[54]发明名称 降钙素片段的重组制备方法和其在制备降钙素和相关类似物中的用途

[57]摘要

本发明提供了一种降钙素片段的重组制备方法和该片段在制备降钙素和相关类似物中的用途。该方法包括重组形成融合蛋白,该融合蛋白包括连接到碳酸酐酶上的降钙素片段。其后切割重组形成的融合蛋白以产生包括降钙素片段的多肽。本发明也提供了一种产生降钙素类似物的方法,该方法包括使重组形成的降钙素片段与脱氨九肽缩合。

ISSN 1000-84274

# 权 利 要 求 书

1. 一种合成降钙素片段的重组方法，该方法包括：

(a) 重组形成一种融合蛋白，该融合蛋白包括通过一个切割位点连接到碳酸酐酶上的靶序列；以及

(b) 用一种切割试剂切割上述融合蛋白以产生包括靶序列的第一多肽，其中所说的靶序列包括具有下式的氨基酸序列：

A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>-A<sub>13</sub>-A<sub>14</sub>-A<sub>15</sub>-A<sub>16</sub>-A<sub>17</sub>-A<sub>18</sub>-A<sub>19</sub>-A<sub>20</sub>-A<sub>21</sub>-A<sub>22</sub>-A<sub>23</sub>-A<sub>24</sub>-A<sub>25</sub>-A<sub>26</sub>-A<sub>27</sub>-  
Gly-A<sub>29</sub>-A<sub>30</sub>-A<sub>31</sub>-Pro-Xxx (SEQ ID NO:1)

其中 A<sub>10</sub> 是 Gly 或者 Ser，A<sub>11</sub> 是 Lys、Thr 或者 Ala，A<sub>12</sub> 是 Leu 或者 Tyr，A<sub>13</sub> 是 Ser、Thr 或者 Trp，A<sub>14</sub> 是 Gln、Lys 或者 Arg，A<sub>15</sub> 是 Glu、Asp 或者 Asn，A<sub>16</sub> 是 Leu 或者 Phe，A<sub>17</sub> 是 His 或者 Asn，A<sub>18</sub> 是 Lys 或者 Asn，A<sub>19</sub> 是 Leu、Tyr 或者 Phe，A<sub>20</sub> 是 Gln 或者 His，A<sub>21</sub> 是 Thr 或者 Arg，A<sub>22</sub> 是 Tyr 或者 Phe，A<sub>23</sub> 是 Pro 或者 Ser，A<sub>24</sub> 是 Arg、Gly 或者 Gln，A<sub>25</sub> 是 Thr 或者 Met，A<sub>26</sub> 是 Asp、Ala、Gly 或者 Asn，A<sub>27</sub> 是 Val、Leu、Ile、Phe 或者 Thr，A<sub>29</sub> 是 Ala、Val、Pro 或者 Ser，A<sub>30</sub> 是 Gly、Val 或者 Glu，A<sub>31</sub> 是 Thr、Val 或者 Ala，以及 -Xxx 是 -OH、-NH<sub>2</sub>、氨基酸残基或者多肽基团。

2. 权利要求 1 的方法，其中 -Xxx 是具有  $\alpha$ -C(O)NH<sub>2</sub> 基团的氨基酸残基。

3. 权利要求 1 的方法，其中 A<sub>10</sub> 是 Gly；切割位点包括 Asn-A<sub>10</sub>；切割步骤包括使融合蛋白与羟胺相接触。

4. 权利要求 1 的方法，其中 -Xxx 是 -OH；该方法还包括使第一多肽与一种酰胺化剂进行反应以形成第二肽，其中 -Xxx 是 -NH<sub>2</sub>。

5. 权利要求 1 的方法，其中 -Xxx 包括氨基酸残基；该方法还包括把 Pro-XXX 序列转化成为 C 末端 Pro-NH<sub>2</sub> 残基。

6. 权利要求 5 的方法，其中转化步骤包括用邻硝基苄胺使 -Xxx 转酰胺化以形成第三肽；以及光解第三肽以形成包括具有下式的氨基酸序列的第二肽：

**A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>-A<sub>13</sub>-A<sub>14</sub>-A<sub>15</sub>-A<sub>16</sub>-A<sub>17</sub>-A<sub>18</sub>-A<sub>19</sub>-A<sub>20</sub>-A<sub>21</sub>-A<sub>22</sub>-A<sub>23</sub>-A<sub>24</sub>-A<sub>25</sub>-A<sub>26</sub>-A<sub>27</sub>-Gly-A<sub>29</sub>-A<sub>30</sub>-A<sub>31</sub>-Pro-NH<sub>2</sub> (SEQ ID NO:2)**

7. 权利要求 1 的方法，其中 A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>-A<sub>13</sub>-A<sub>14</sub>-A<sub>15</sub>-A<sub>16</sub>-A<sub>17</sub>-A<sub>18</sub>-A<sub>19</sub>-A<sub>20</sub>-A<sub>21</sub>-A<sub>22</sub>-A<sub>23</sub>-A<sub>24</sub>-A<sub>25</sub>-A<sub>26</sub>-A<sub>27</sub>-Gly-A<sub>29</sub>-A<sub>30</sub>-A<sub>31</sub>-Pro-Xxx (SEQ ID NO:1)序列是：

**Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-Xxx (SEQ ID NO:6)**

8. 权利要求 1 的方法，该方法包括使所述融合蛋白与一种保护试剂进行反应以形成一种保护的融合蛋白，所说的保护的融合蛋白包括具有被保护的氨基、羟基或者羧基活性侧链基团的氨基酸残基。

9. 权利要求 8 的方法，该方法包括使所述融合蛋白与一种保护试剂进行反应以形成一种保护的融合蛋白，所说的保护的融合蛋白包括具有被保护侧链氨基的 Lys 残基。

10. 权利要求 1 的方法，其中切割步骤包括使所述融合蛋白与切割试剂相接触以形成一种微型融合蛋白质，所说的微型融合蛋白质包括具有下式的氨基酸序列：

**接头肽 -A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>-A<sub>13</sub>-A<sub>14</sub>-A<sub>15</sub>-A<sub>16</sub>-A<sub>17</sub>-A<sub>18</sub>-A<sub>19</sub>-A<sub>20</sub>-A<sub>21</sub>-A<sub>22</sub>-A<sub>23</sub>-A<sub>24</sub>-A<sub>25</sub>-A<sub>26</sub>-A<sub>27</sub>-Gly-A<sub>29</sub>-A<sub>30</sub>-A<sub>31</sub>-Pro-Xxx**

11. 权利要求 10 的方法，其中微型融合蛋白质具有下式：

**Val-Asp-Asn-Trp-Arg-Pro-Ala-Gln-Pro-Leu-Lys-Asn-Arg-Glu-Ile-Lys-Ala-Phe-Val-Asp-Asp-Asp-Asn-Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-Ala-Pro-Ala (SEQ ID NO:24).**

12. 权利要求 11 的方法，其中所说的碳酸酐酶是人碳酸酐酶 II，所说的切割试剂包括溴化氰。

13. 权利要求 11 的方法，该方法还包括用羟胺切割上述微型融合蛋白质以形成一种肽，该肽包括具有下式的氨基酸序列：

**Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-Xxx (SEQ ID NO:6)**

14. 权利要求 10 的方法，该方法还包括使所述微型融合蛋白质与一种保护试剂进行反应以形成一种保护的微型融合蛋白质，所说的保护性微型融合蛋白质包括具有被保护的氨基、羟基或者羧基活性侧链基团的氨基酸残基。

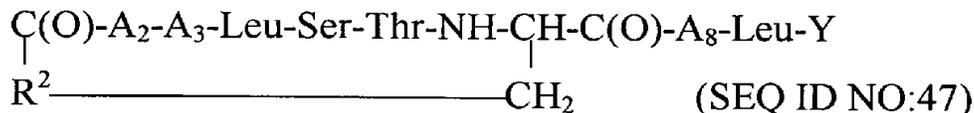
15. 权利要求 1 的方法，其中所说的碳酸酐酶是人碳酸酐酶 II；所说的靶序列包括具有下式的氨基酸序列： Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-Ala<sub>C</sub> (SEQ ID NO:23)，其中 Ala<sub>C</sub> 是 C 末端残基；并且该方法还包括用邻硝基苯基甘氨酸对-Ala<sub>C</sub> 进行转肽基作用以形成一种第三肽；以及光解所说的第三肽以形成一种第二肽，所说的第二肽包括具有下式的氨基酸序列：

Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-NH<sub>2</sub> (SEQ ID NO:6)。

16. 权利要求 15 的方法，其中所说的切割位点包括 Asn-A<sub>10</sub>；以及切割步骤包括使上述融合蛋白与羧胺相接触。

17. 一种核酸序列，该核酸序列包含一个具有下式的序列：  
GGT AAA CTG TCT CAG GAG CTC CAT AAA CTG CAG ACT TAC  
CCG CGT ACT GAC GTT GGT ACC CCG (SEQ ID NO:5)

18. 一种用于产生降钙素碳类似物的方法，该方法包括将下式的 N 末端片段：



其中 A<sub>2</sub> 是 Gly、Ser 或者 Ala，A<sub>3</sub> 是 Asn 或者 Ser，A<sub>8</sub> 是 Val 或者 Met，R<sup>2</sup> 是(CH<sub>2</sub>)<sub>4</sub> 或者-CH(NH<sub>2</sub>)CH<sub>2</sub>S-S-，以及 Y 是 OH、OR<sup>1</sup>，其中-R<sup>1</sup> 是低级烷基；

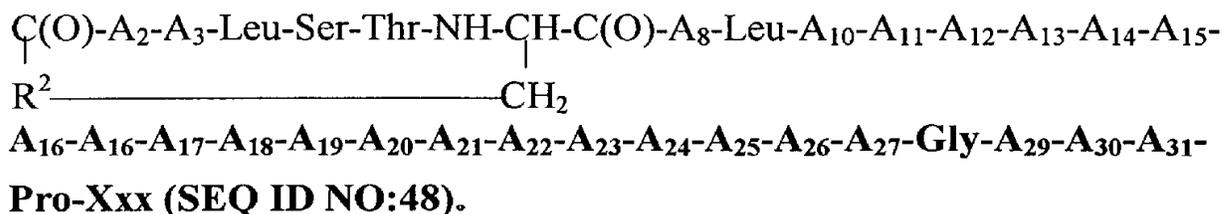
在非酶促偶联试剂存在下与重组形成的具有下式的多肽缩合：

A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>-A<sub>13</sub>-A<sub>14</sub>-A<sub>15</sub>-A<sub>16</sub>-A<sub>17</sub>-A<sub>18</sub>-A<sub>19</sub>-A<sub>20</sub>-A<sub>21</sub>-A<sub>22</sub>-A<sub>23</sub>-A<sub>24</sub>-A<sub>25</sub>-A<sub>26</sub>-A<sub>27</sub>-  
Gly-A<sub>29</sub>-A<sub>30</sub>-A<sub>31</sub>-Pro-Xxx (SEQ ID NO:1)

其中 A<sub>10</sub> 是 Gly 或者 Ser，A<sub>11</sub> 是 Lys、Thr 或者 Ala，A<sub>12</sub> 是 Leu

或者 Tyr, A<sub>13</sub> 是 Ser、Thr 或者 Trp, A<sub>14</sub> 是 Gln、Lys 或者 Arg, A<sub>15</sub> 是 Glu、Asp 或者 Asn, A<sub>16</sub> 是 Leu 或者 Phe, A<sub>17</sub> 是 His 或者 Asn, A<sub>18</sub> 是 Lys 或者 Asn, A<sub>19</sub> 是 Leu、Tyr 或者 Phe, A<sub>20</sub> 是 Gln 或者 His, A<sub>21</sub> 是 Thr 或者 Arg, A<sub>22</sub> 是 Tyr 或者 Phe, A<sub>23</sub> 是 Pro 或者 Ser, A<sub>24</sub> 是 Arg、Gly 或者 Gln, A<sub>25</sub> 是 Thr 或者 Met, A<sub>26</sub> 是 Asp、Ala、Gly 或者 Asn, A<sub>27</sub> 是 Val、Leu、Ile、Phe 或者 Thr, A<sub>29</sub> 是 Ala、Val、Pro 或者 Ser, A<sub>30</sub> 是 Gly、Val 或者 Glu, A<sub>31</sub> 是 Thr、Val 或者 Ala, 以及 -Xxx 是 -OH、-NH<sub>2</sub>、氨基酸残基或者多肽基团;

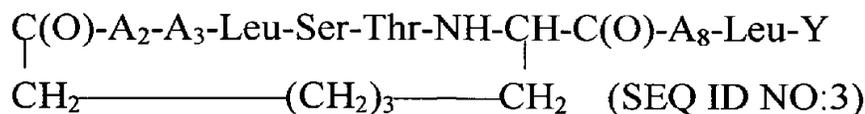
形成具有下式的降钙素衍生物:



19. 权利要求 18 的方法, 其中所说的非酶促偶联试剂包括 N-羟基琥珀酰亚胺、N-羟基苯并三唑或者碳二亚胺。

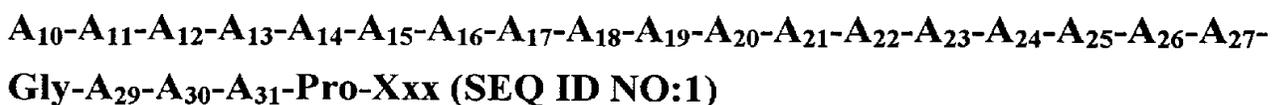
20. 权利要求 18 的方法, 其中所说的非酶促偶联试剂包括(i) N-羟基琥珀酰亚胺和二环己基碳二亚胺; (ii) N-羟基琥珀酰亚胺和 N-乙基-N'-二甲基氨基丙基碳二亚胺或者(iii) N-羟基苯并三唑和二环己基碳二亚胺。

21. 一种用于产生降钙素或者相关类似物的方法, 该方法包括将一种具有下式的脱氧九肽:



其中 A<sub>2</sub> 是 Gly、Ser 或者 Ala, A<sub>3</sub> 是 Asn 或者 Ser, A<sub>8</sub> 是 Val 或者 Met, 以及 Y 是 OH、OR<sup>1</sup>, 其中 -R<sup>1</sup> 是低级烷基;

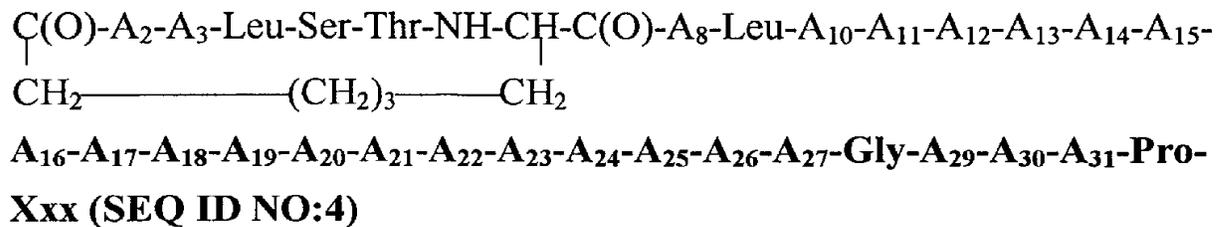
在非酶促偶联剂存在下与重组形成的具有下式的多肽缩合:



其中 A<sub>10</sub> 是 Gly 或者 Ser, A<sub>11</sub> 是 Lys、Thr 或者 Ala, A<sub>12</sub> 是 Leu 或者 Tyr, A<sub>13</sub> 是 Ser、Thr 或者 Trp, A<sub>14</sub> 是 Gln、Lys 或者 Arg,

A<sub>15</sub>是Glu、Asp或者Asn，A<sub>16</sub>是Leu或者Phe，A<sub>17</sub>是His或者Asn，A<sub>18</sub>是Lys或者Asn，A<sub>19</sub>是Leu、Tyr或者Phe，A<sub>20</sub>是Gln或者His，A<sub>21</sub>是Thr或者Arg，A<sub>22</sub>是Tyr或者Phe，A<sub>23</sub>是Pro或者Ser，A<sub>24</sub>是Arg、Gly或者Gln，A<sub>25</sub>是Thr或者Met，A<sub>26</sub>是Asp、Ala、Gly或者Asn，A<sub>27</sub>是Val、Leu、Ile、Phe或者Thr，A<sub>29</sub>是Ala、Val、Pro或者Ser，A<sub>30</sub>是Gly、Val或者Glu，A<sub>31</sub>是Thr、Val或者Ala，以及-Xxx是-OH、-NH<sub>2</sub>、氨基酸残基或者多肽基团；

形成具有下式的降钙素衍生物：



降钙素片段的重组制备方法和  
其在制备降钙素和相关类似物中的用途

人们已知降钙素和相关类似物(如 Elcatonin)是可以用于治疗骨萎缩(参见, 例如美国专利 4,086,221)的多肽。天然存在的降钙素(如鳗鱼、鲑鱼或者人体降钙素)是 C 末端酰胺化的多肽, 该多肽由 32 个氨基酸组成, 在每一种情况中其第一和第七个氨基酸是 L-半胱氨酸, 两个半胱氨酸的巯基通过二硫键的形成来相互连接。可以例如通过从哺乳动物甲状腺中提取获得天然降钙素(参见, 例如美国专利 5,428,129)。

Elcatonin 是一种改进的合成降钙素“碳”类似物, 其活性可与鳗鱼降钙素相比(Morikawa 等, *Experienta*, 32, 1004, 1976)。与鳗鱼降钙素相比, Elcatonin 缺乏氨基末端, 并且鳗鱼降钙素的二硫键为 $-(CH_2)_5-$ “碳桥”所取代。

当前, 利用纯化学过程制备 Elcatonin 的各种方法是已知的。这些化学方法包括相应的氨基酸或者肽的缩合(参见, 例如美国专利 4,086,221 和美国专利 5,428,129)。然而纯化学方法都有其缺点, 即由于所要求的精细纯化方法, 只能以低产率获得 Elcatonin, 随之而来的是它的制备十分昂贵。

因此, 在通过利用某一方法制备 Elcatonin 的过程中能够避免纯化学方法的不利之处将是有益的, 这些方法包括分子部分的重组制备。这一目的是可以实现的, 例如, 如果有一种用于重组制备 C 末端多肽片段的简单方法。然后将重组合成的 C 末端片段用作为制备降钙素或者碳类似物(如 Elcatonin)的起始肽。部分重组方法也将有利于降钙素的肽类/肽类似物、Elcatonin 和相关类似物或者衍生物的合成, 这些衍生物可能包括非天然的氨基酸。

本发明涉及降钙素片段的重组制备方法和该片段在制备降钙素和相关类似物中的用途, 该相关类似物包括如 Elcatonin 的碳类似物(在下文称作为“降钙素碳类似物”)。本发明包括重组形成融合蛋白, 该融合蛋白包括通过切割位点连接到碳酸酐酶上的靶序列。该靶序列包括至少大约 15 个氨基酸残基, 该残基对应于接近降钙素 C-末端的片段或者对应于这种片段的密切相关类似物。典型地, 靶序列包括对应于降钙素或者密切相关类似物的氨基酸残基 10 直至 32(在下文共同称作为“10-32 片段”)的氨基酸序列。其后用切割试剂切割重组形成的融合蛋白以产生包括靶序列的多肽。通过使融合蛋白与化学切割试剂或者酶促切割试剂接触来实现切割反应。合适切割试剂和掺入融合蛋白中的相应切割位点的选择都将取决于存在于融合蛋白中的具体的靶序列和碳酸酐酶序列。典型地, 可这样选择切割试剂以及切割位点, 其中构成切割位点的氨基酸序列并不出现在靶序列或者碳酸酐酶的氨基酸序列之中。例如, 不能对包括猪、牛或者羊降钙素 10-32 片段的融合蛋白在蛋氨酸上进行溴化氰切割。

切割位点典型地存在于连接碳酸酐酶和靶序列的接头序列中。此外, 融合蛋白可以包括一种构建体, 其中碳酸酐酶的 C-末端直接与靶序列的 N-末端连接。这可以发生在碳酸酐酶的 C 末端残基和靶序列的 N 末端残基构成切割位点(其使得可以进行上述两片段之间的肽键的切割)的情况下。除存在于接头序列中的切割位点之外, 融合蛋白的碳酸酐酶部分也可以包括不同的切割位点, 该切割位点使得融合蛋白可以被切割从而形成“微型融合蛋白质”, 即具有仍然与靶序列连接的碳酸酐酶的 C 末端部分的一种多肽。

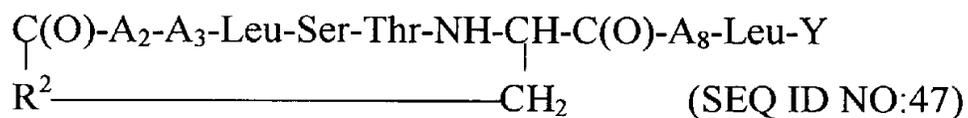
本发明的一种实施方案包括一种用于多肽的重组制备的方法, 该多肽对应于降钙素或者相关类似物的氨基酸 10-32(“10-32 个片段”)。该方法典型地包括下式的多肽片段(“10-32 片段-Xxx”)的重组制备:

A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>-A<sub>13</sub>-A<sub>14</sub>-A<sub>15</sub>-A<sub>16</sub>-A<sub>17</sub>-A<sub>18</sub>-A<sub>19</sub>-A<sub>20</sub>-A<sub>21</sub>-A<sub>22</sub>-A<sub>23</sub>-A<sub>24</sub>-A<sub>25</sub>-A<sub>26</sub>-A<sub>27</sub>-  
Gly-A<sub>29</sub>-A<sub>30</sub>-A<sub>31</sub>-Pro-Xxx (SEQ ID NO:1)

其中 A<sub>10</sub> 是 Gly 或者 Ser, A<sub>11</sub> 是 Lys、Thr 或者 Ala, A<sub>12</sub> 是 Leu 或者 Tyr, A<sub>13</sub> 是 Ser、Thr 或者 Trp, A<sub>14</sub> 是 Gln、Lys 或者 Arg,

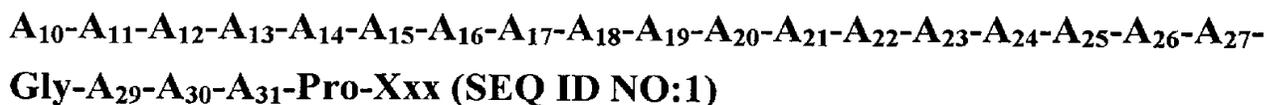
A<sub>15</sub>是Glu、Asp或者Asn，A<sub>16</sub>是Leu或者Phe，A<sub>17</sub>是His或者Asn，A<sub>18</sub>是Lys或者Asn，A<sub>19</sub>是Leu、Tyr或者Phe，A<sub>20</sub>是Gln或者His，A<sub>21</sub>是Thr或者Arg，A<sub>22</sub>是Tyr或者Phe，A<sub>23</sub>是Pro或者Ser，A<sub>24</sub>是Arg、Gly或者Gln，A<sub>25</sub>是Thr或者Met，A<sub>26</sub>是Asp、Ala、Gly或者Asn，A<sub>27</sub>是Val、Leu、Ile、Phe或者Thr，A<sub>29</sub>是Ala、Val、Pro或者Ser，A<sub>30</sub>是Gly、Val或者Glu，A<sub>31</sub>是Thr、Val或者Ala。C-末端-Xxx基团典型地是C末端羧酸(“-OH”)，C末端甲酰胺(“-NH<sub>2</sub>”)，或者能转化成C末端甲酰胺的基团，如氨基酸残基或者多肽基团(典型地，具有2至大约10个氨基酸残基)。由残基A<sub>10</sub>至A<sub>32</sub>(SEQ ID NO:2)表示的10-32片段相应于得自鳗鱼(SEQ ID NO:37)，鲑鱼I(SEQ ID NO:38)、鲑鱼II(SEQ ID NO:39)、鲑鱼III(SEQ ID NO:40)，小鸡(SEQ ID NO:41)、人(SEQ ID NO:42)、兔(SEQ ID NO:43)、猪(SEQ ID NO:44)、牛(SEQ ID NO:45)和羊(SEQ ID NO:46)降钙素或者密切相关类似物(参见显示在图9中的降钙素序列)的氨基酸序列的残基10直至32。本方法也用来重组产生与修饰的降钙素序列相应的10-32片段。修饰的降钙素序列可以包括一个或多个在天然氨基酸序列中的保守氨基酸取代。

10-32片段可以用于降钙素和相关类似物的制备。这种制备典型地包括下式的N末端片段：



其中A<sub>2</sub>是Gly、Ser或者Ala，A<sub>3</sub>是Asn或者Ser，A<sub>8</sub>是Val或者Met，R<sup>2</sup>是(CH<sub>2</sub>)<sub>4</sub>或者-CH(NH<sub>2</sub>)CH<sub>2</sub>S-S-，以及Y是OH、OR<sup>1</sup>，其中-R<sup>1</sup>是低级烷基；

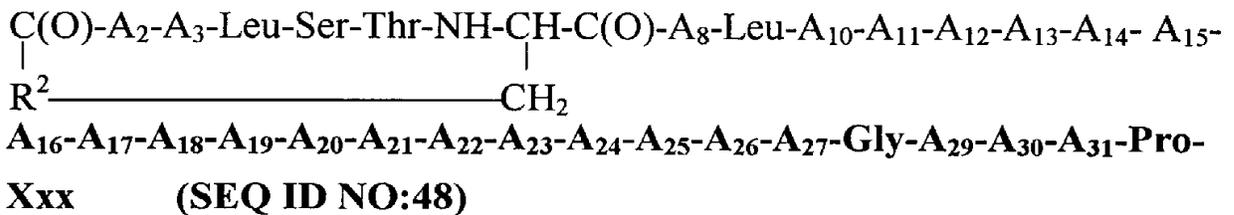
与重组形成的具有下式的多肽在非酶促试剂的存在下缩合：



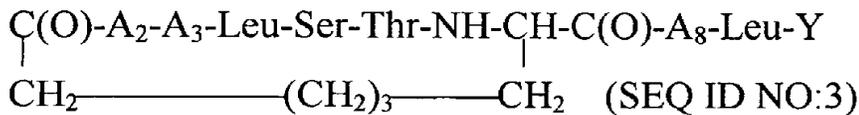
其中A<sub>10</sub>是Gly或者Ser，A<sub>11</sub>是Lys、Thr或者Ala，A<sub>12</sub>是Leu或者Tyr，A<sub>13</sub>是Ser、Thr或者Trp，A<sub>14</sub>是Gln、Lys或者Arg，A<sub>15</sub>是Glu、Asp或者Asn，A<sub>16</sub>是Leu或者Phe，A<sub>17</sub>是His或者Asn，

A<sub>18</sub>是Lys或者Asn，A<sub>19</sub>是Leu、Tyr或者Phe，A<sub>20</sub>是Gln或者His，A<sub>21</sub>是Thr或者Arg，A<sub>22</sub>是Tyr或者Phe，A<sub>23</sub>是Pro或者Ser，A<sub>24</sub>是Arg、Gly或者Gln，A<sub>25</sub>是Thr或者Met，A<sub>26</sub>是Asp、Ala、Gly或者Asn，A<sub>27</sub>是Val、Leu、Ile、Phe或者Thr，A<sub>29</sub>是Ala、Val、Pro或者Ser，A<sub>30</sub>是Gly、Val或者Glu，A<sub>31</sub>是Thr、Val或者Ala，以及-Xxx是-OH、-NH<sub>2</sub>、氨基酸残基或者多肽基团；

形成具有下式的降钙素衍生物：



在本发明的一个实施方案中，用脱氨九肽缩合重组形成的 10-32 片段。脱氨九肽是 N 末端降钙素片段的碳类似物并且典型地具有式：



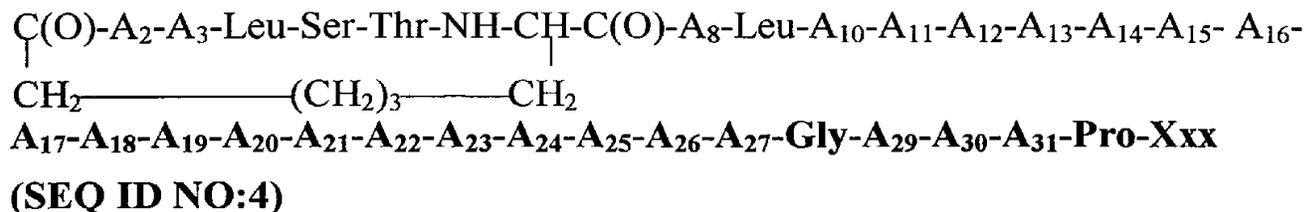
其中 A<sub>2</sub>是 Gly、Ser 或者 Ala，A<sub>3</sub>是 Asn 或者 Ser，A<sub>8</sub>是 Val 或者 Met，以及 Y 是 OH、OR<sup>1</sup>，其中 -R<sup>1</sup> 是低级烷基(即 C<sub>1</sub>-C<sub>6</sub> 烷基)。利用如在美国专利 4,086,221 和美国专利 5,428,129 所描述的那些化学偶联反应可以实现缩合反应，这些专利的公开内容并入本文作为参考。化学偶联剂是本领域技术人员所熟知的。合适的化学偶联剂包括碳二亚胺和各种其它非酶促试剂，这些非酶促试剂能够与肽的 α-羧酸基团进行反应从而形成活化的羧酸衍生的 α-羧酸衍生物，该 α-羧酸衍生物与另一个氨基酸或者多肽的 N 末端 α-氨基反应。其中使脱氨九肽的 C 末端 α-羧酸转化成为酰基叠氮、混合酸酐、咪唑酰胺或者活性酯的化学偶联反应可以用于本发明中。在存在碳二亚胺和能够形成活性酯的试剂的情况下，可以实现一种特别有效的偶联两个肽片段的方法，所说试剂例如为二环己基碳二亚胺(“DCC”)和 N-羟基琥珀酰亚胺(“HOSu”)或者 1-羟基苯并三唑(“HOBT”)的混合物。

用于缩合的重组形成的 10-32 片段优选地具有下式：

**A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>-A<sub>13</sub>-A<sub>14</sub>-A<sub>15</sub>-A<sub>16</sub>-A<sub>17</sub>-A<sub>18</sub>-A<sub>19</sub>-A<sub>20</sub>-A<sub>21</sub>-A<sub>22</sub>-A<sub>23</sub>-A<sub>24</sub>-A<sub>25</sub>-A<sub>26</sub>-A<sub>27</sub>-  
Gly-A<sub>29</sub>-A<sub>30</sub>-A<sub>31</sub>-Pro-Xxx (SEQ ID NO:1)**

其中，A<sub>10</sub> 直至 A<sub>31</sub> 和-Xxx 如本文所定义。

偶联反应的产物典型地是具有下式的降钙素碳类似物：

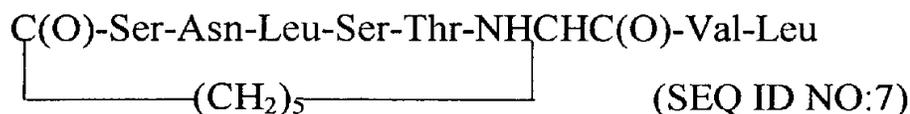


其中，A<sub>10</sub> 直至 A<sub>31</sub> 和-Xxx 如本文脱氨九肽(SEQ ID NO:3)或者 10-32 片段-Xxx (SEQ ID NO:1)下所定义的一样。典型地，在存在非酶促偶联试剂的情况下实现脱氨九肽和重组形成的肽的偶联。

本发明的一个优选实施方案提供了一种方法，该方法用于具有下式的多肽片段(这里称作为“ECF2-酰胺”)的重组制备和酰胺化：

**Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-  
Asp-Val-Gly-Ala-Gly-Thr-Pro-NH<sub>2</sub> (SEQ ID NO:6)**

以及用于偶联 ECF2-酰胺到 Elcatonin 的氨基末端片段(下文称作为“ECH1”), 该氨基末端片段具有式：



本发明也提供了一个核酸序列，该核酸序列包括编码降钙素或者相关类似物的氨基酸 10-32 的序列。该核酸序列典型地编码融合蛋白，该融合蛋白包括通过切割位点连接到碳酸酐酶上的 10-32 片段。优选利用用于目标宿主细胞(如大肠杆菌、酿酒酵母或者巴斯德毕赤酵母)的最佳密码子使用，设计编码 10-32 片段的基因部分。

#### 附图简要描述

图 1 显示质粒 pET31F1mhCAII 的图。该质粒包括与 pSP65 复制起点的相反方向上引入的来自质粒 pEMBL8 的复制起点 F1。

图 2 显示质粒 pABN 的图。质粒 pABN 包括编码接头肽片段(SEQ ID

NO:49)的核酸序列(SEQ ID NO:47 和 SEQ ID NO:48), 该接头肽片段插入到质粒中邻近人碳酸酐酶 II (“hCAII”)的基因序列的羧基末端处。

图 3 显示质粒 pTBN 的图。质粒 pTBN 是四环素抗性表达载体, 该载体得自 pABN 的 1.5 kb 片段向质粒 pBR322 中的插入。该 1.5 kb 片段包括 pABN 的 T7 启动子、hCAII、接头和 T7 终止子序列。

图 4A 是用 PCR 方法制备编码 ECF2-Ala 和附加的限制酶切位点(使得将片段克隆进入质粒得以实现)的核苷酸序列的第一步的示意说明。通过利用 Taq DNA 聚合酶在 PCR MIX 1 上实施 PCR 反应实现这一步骤。

图 4B 是编码 ECF2-Ala 的 DNA 序列的制备的第二步的示意性说明。得自在 PCR MIX 1 中的寡核苷酸 2 和 3 的延伸(分别为 2<sup>Ext</sup>(SEQ ID NO:8) 和 3<sup>Ext</sup>(SEQ ID NO:50))的 PCR 产物在它们的 3' 末端上具有 20 bp 互补序列。利用 Taq DNA 聚合酶的第二个 PCR 反应产生了双链核苷酸片段, 该核苷酸片段包括编码 ECF2-Ala (SEQ ID NO:9)的全长无间断基因序列。

图 5 显示质粒 pTBN26 的图。质粒 pTBN26 包括总 hCAII-接头-ECF2-Ala 融合蛋白(“hCA-ECF2-Ala”)构建体的基因序列。

图 6A 显示 hCAII-接头-ECF2-Ala 融合蛋白构建体的 N 末端蛋氨酸和 hCAII 残基 1-162 的核苷酸(SEQ ID NO:11)和氨基酸(SEQ ID NO:12)序列。

图 6B 显示 hCAII-接头-ECF2-Ala 融合蛋白构建体的 hCAII 残基 162-257 以及接头和 ECF2-Ala 片段的核苷酸(SEQ ID NO:11)和氨基酸(SEQ ID NO:12)序列。

图 7 显示 Elcatonin 和鳗鱼降钙素的 23 氨基酸 C 末端片段(“ECF2”)的核苷酸序列(SEQ ID NO:5)和氨基酸(SEQ ID NO:6)序列。所显示的核苷酸序列是编码 ECF2 的序列, ECF2 存在于掺入到质粒 pTBN26 中的 hCA-ECF2-Ala 融合蛋白的基因中。

图 8 显示包含 ECF2-酰胺(SEQ ID NO:6)的样品的代表性制备性 HPLC 图(利用聚磺乙基天冬酰胺(polysulfolethylaspartamide)柱)。ECF2-酰胺的峰值出现在 14.5 和 16.3 分钟之间。

图 9A 显示来源于一些物种的降钙素的残基 1-15 的氨基酸序列。

图 9B 显示来源于一些物种的降钙素的残基 16-32 的氨基酸序列。

图 10 显示经由 PCR 方法从 5 寡核苷酸(描述在实施例 1.1 中)合成的双链 DNA 片段(SEQ ID NO:9)以及由该双链 DNA 片段(SEQ ID NO:10)编码的肽序列。

图 11 描述在编码 ECF2 的基因序列的 PCR 合成期间产生的寡核苷酸 2<sup>Ext</sup> 的核苷酸序列(SEQ ID NO:8)。

图 12 描述在编码 ECF2 的基因序列的 PCR 合成期间产生的寡核苷酸 3<sup>Ext</sup> 的核苷酸序列(SEQ ID NO:50)。

## 发明详述

按照本发明的融合蛋白的重组制备包括编码靶序列的核酸序列(“FP 核酸序列”)的构建, 该靶序列通过切割位点连接到碳酸酐酶上。在图 6 中所显示的 FP 核酸序列是用于本方法的合适核酸序列的一个例子。在图 6 中所显示的核酸序列包括编码人碳酸酐酶 II (“hCAII”)的残基 1-257 的密码子。

这一氨基酸序列或者碳酸酐酶的其它功能活性片段在融合蛋白中的包含极大地有利于融合蛋白从细胞碎片中分离出来。如本文所使用的, 术语“碳酸酐酶”包括天然存在的碳酸酐酶(和任一等位变体)、功能活性碳酸酐酶片段或者其修饰体(“突变体”)。这里的“功能活性碳酸酐酶片段和突变体”意指碳酸酐酶的片段或者修饰体, 其仍显示出 hCAII 的酶抑制剂结合性质。具有这样性质的片段能够极紧密地结合碳酸酐酶抑制剂(如磺胺)。具有这样结合性质的功能活性片段显示出高度选择性的与低分子量配体(例如磺胺或者其合成衍生物)的亲合结合。一般来说, 这种结合是如此强烈以致碳酸酐酶/配体缀合物将显示出不超过大约  $10^{-7}$  M 的溶液解离常量(结合常数的倒数)。一般地, 这种配体是碳酸酐酶的可逆抑制剂。合适的碳酸酐酶片段可以缺乏酶的 N 末端或者 C 末端部分, 只要该片段保持酶的功能性抑制剂结合活性。同样地, 修饰碳酸酐酶(在本发明融合蛋白中可以用作为结合蛋白)可以通过一个或多个氨基酸的加入、缺失或者

插入得到修饰，只要修饰酶本质上显示出以上所描述的抑制剂结合活性。典型地，通过缺失或者改变氨基酸残基修饰碳酸酐酶，结果是修饰酶不包含将要在融合蛋白中用作为切割位点的特异氨基酸。例如，当溴化氰将要用作为切割试剂时，可以修饰碳酸酐酶以除去或者取代正常存在的任一蛋氨酸(“Met”)残基。

FP 核酸序列包括编码 10-32 片段的序列，即降钙素或者密切相关类似物的氨基酸残基 10-32。密切相关类似物可以得自在降钙素 10-32 片段中的保守氨基酸取代。可以设计这部分核酸序列以便优化融合蛋白在特异宿主细胞中的表达。例如，在将要利用肠杆菌(如大肠杆菌)作为宿主有机体进行融合蛋白的表达的情况下，典型地基于在目标宿主有机体中的密码子使用频率(参见，例如，Gribskov 等，核酸研究，12, 539-549, 1984 中的密码子使用频率的讨论)构建编码 10-32 片段的核酸序列和切割位点。

切割位点可以是化学切割位点或者酶促切割位点。在 PCT 专利申请号 WO 92/01707 中详尽地描述了化学和酶促切割位点和用来进行与这些位点之一连接的肽键切割的相应试剂，该专利申请的公开内容引入本文作为参考。适合于在本发明中用作为切割位点的肽序列(以及编码它的 DNA 基因序列)和相应的切割酶或者化学切割条件的例子显示在表 1 中。所表明的基因序列是编码相应肽序列的一种可能序列。可以构建编码相同肽序列的其它 DNA 序列。优选地，基于将要用于融合蛋白表达的宿主细胞的每一氨基酸的更通常使用的密码子，设计编码切割位点的核酸序列。例如，核苷酸序列可以基于在如大肠杆菌的肠杆菌中的最佳密码子使用(参见，例如 Gribskov 等，1984，核酸研究，12, 539-549)。

切割位点可以存在作为接头肽的一部分，该接头肽连接碳酸酐酶和靶序列。例如，在图 6 中描述的 hCAII-接头-ECF2-Ala(“hCA-ECF2-Ala”)的氨基酸序列(SEQ ID NO:12)包括一个七氨基酸接头序列(-Phe-Val-Asp-Asp-Asp-Asn-; (SEQ ID NO:14))，该接头序列建立了可由羟胺(“Asn-Gly”)切割的化学切割位点，该切割位点位于接头的 C 末端天冬酰胺残基和靶序列的 N 末端甘氨酸残基之间。

表 1

<u>用于切割的酶</u>	<u>肽序列</u>	<u>DNA 序列</u>
肠激酶	(Asp) <sub>4</sub> Lys (SEQ ID NO:16)	GACGACGACGATAAA (SEQ ID NO:15)
因子 Xa	IleGluGlyArg (SEQ ID NO:18)	ATTGAAGGAAGA (SEQ ID NO:17)
凝血酶	GlyProArg 或者 GlyAlaArg	GGACCAAGA 或者 GGAGCGAGA
遍在蛋白切割酶	ArgGlyGly	AGAGGAGGA
肾素	HisProPheHisLeu- LeuValTyr (SEQ ID NO:20)	CATCCTTTTCATC- TGCTGGTTTAT (SEQ ID NO:19)
胰蛋白酶	Lys 或者 Arg	AAA 或者 CGT
胰凝乳蛋白酶	Phe 或 Tyr 或 Trp	TTT 或 TAT 或 TGG
梭菌蛋白酶	Arg	CGT
S. aureus V8	Glu	GAA
<u>化学切割</u>	<u>肽序列</u>	<u>DNA 序列</u>
(在 pH3)	AspGly 或 AspPro	GATGGA 或 GATCCA
(羟胺)	AsnGly	AATCCA
(CNBr)	蛋氨酸	ATG
BNPS-粪臭素	Trp	TGG
2-硝基-5- 氟硫基苯甲酸	Cys	TGT

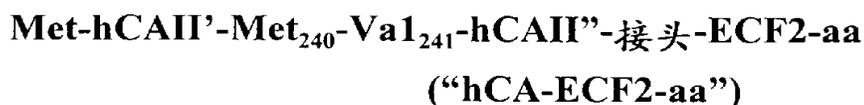
融合蛋白包括至少一个拷贝，并且可以包括多个靶序列拷贝。在存在多个靶序列拷贝的情况下，这些拷贝可以串联地连接在一起。此外，靶序列的拷贝可以由内部接头肽连接在一起。上述内部接头肽可以与外部接头肽相同或者不相同，上述外部接头肽把碳酸酐酶与靶肽的第一个拷贝相连接。如果上述内部和上述外部接头肽是相同的或者包括相同的切割位点，

那么可以直接切割融合蛋白从而产生一些片段, 这些片段的每一个包含靶肽的单拷贝。然而, 如果上述内部和上述外部接头肽包含不同的切割位点, 那么就有可能首先把碳酸酐酶片段切割下来从而形成具有一个以上靶肽拷贝的中间多肽。

利用本方法可以从多拷贝构建体(具有包括蛋氨酸残基的内部连接肽)中产生无蛋氨酸残基的靶序列。在蛋氨酸残基与靶序列的 C-末端直接连接的情况下, 可以用溴化氰切割上述多拷贝构建体。可以利用例如丝氨酸羧肽酶(如羧肽酶 Y)的羧肽酶对所产生的片段进行转肽基作用从而用 $\alpha$ -酰胺化的氨基酸取代 C 末端高丝氨酸残基。该片段也可以利用羧肽酶转酰胺化从而用 2-硝基苄胺化合物取代 C 末端高丝氨酸残基。这一过程产生具有 C 末端(2-硝基苄基)酰氨基的片段, 所说的酰氨基可以以光化学方式分解从而产生无高丝氨酸残基的 $\alpha$ -酰胺化的肽片段。

在构建体中包括多拷贝靶序列的融合蛋白的一个例子中, 所说的构建体包括 hCA-(MetValAspAspAspAspAsn-ECF2)<sub>n</sub>-Xxx, 其中 hCA、ECF2 和 Xxx 如本文所定义, n 是一个整数(典型地是 2-20)。可以利用 CNBr 处理这种构建体从而形成 ValAspAspAspAspAsn-ECF2-Hse (SEQ ID NO:49)肽片段(其中 Hse 是由 CNBr 与 Met 残基的反应产生的高丝氨酸残基)。然后当如羧肽酶 Y 的肽酶存在时可以使该肽片段与如邻硝基苯基甘氨酸酰胺(“ONPGA”)的亲核体进行反应, 结果导致 Hse 残基为 ONPGA 所取代。利用光解, 将转肽基作用产物转化成为 C 末端甲酰胺。通过用羟胺处理可以从所述片段上切除 N 末端尾序列 ValAspAspAspAspAsn (SEQ ID NO:50)。

本发明的一个优选实施方案涉及 C 末端鳗鱼降钙素多肽片段即 ECF2-酰胺(SEQ ID NO:6)的制备。该制备包括下列蛋白质构建体的起始遗传表达:



其中, aa 是氨基酸残基, Met 是任一大肠杆菌蛋白质所需要的 N 末端残基。将 Met 残基加入至 hCAII'(人碳酸酐酶 II 的 239 氨基酸 N 末端多肽

片段)的第一个残基的 N-末端上, Met<sub>240</sub>-Val<sub>241</sub> 是在人碳酸酐酶 II (“hCAII”)中的唯一的溴化氰不稳定的肽键, hCAII”是来源于邻近 hCAII (残基 242-257)的 C 末端的 16 氨基酸片段(SEQ ID NO:21)。 ECF2-aa 的 C 末端氨基酸残基(SEQ ID NO:22)(即 aa)提供了酰胺化信号从而使之能够转变成为 Pro-酰胺, 该 Pro-酰胺构成 Elcatonin (SEQ ID NO:13)的 C-末端。 aa 残基典型地是如丙氨酸的氨基酸残基, 该氨基酸残基能够经由转酰胺反应与亲核体的交换。

所需要的产物 ECF2-aa (SEQ ID NO:22)可以用至少两种方法得自 hCA-ECF2-aa 蛋白质构建体。第一种方法使用利用羟胺在 Asn-Gly 键上对 hCA-ECF2-aa 的切割, 从而直接产生 ECF2-aa (SEQ ID NO:22)。此外, 利用溴化氰(CNBr)的切割产生微型融合蛋白质, 该融合蛋白质包括连接到 ECF2-aa 多肽(SEQ ID NO:22)上的 hCAII 的 hCAII”C 末端片段 (SEQ ID NO:21)。其后可以在微型融合蛋白质侧链残基的衍生化作用之前或者之后用羟胺切割该微型融合蛋白质从而提供 ECF2-aa (SEQ ID NO:22), 所说的衍生化作用是指例如, 其中 ECF2-aa (SEQ ID NO:22)的 Lys 残基已得到衍生而形成具有 Z-保护侧链氨基的 Lys 残基。在其中衍生化作用试剂修饰 Lys 的侧链氨基功能的情况下, 在衍生的微型融合蛋白质切割之后, 所产生的带保护基的 ECF2-aa (SEQ ID NO:22)在 N 末端甘氨酸残基的 $\alpha$ 位上将仅仅具有单自由氨基功能团。自由 N 末端 $\alpha$ -氨基可以用于尔后的特异性化学反应, 例如偶联到 ECF1 (SEQ ID NO:7)的 C 末端残基上从而提供 Elcatonin (SEQ ID NO:13)。这种类型的偶联反应可以在 ECF2-aa 多肽(SEQ ID NO:22)已转化成为具有 C 末端残基 Pro-酰胺 (“ECF2-酰胺”, (SEQ ID NO:6))的 ECF2 衍生物之后进行, 或者可以用来产生衍生形式的 Elcatonin(例如“Elcatonin-aa”, (SEQ ID NO:30)), 该衍生形式的 Elcatonin 用于尔后的向 Elcatonin (SEQ ID NO:13)的转变。

ECF2-aa 的优选序列是:

Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-Ala (“ECF2-Ala”, SEQ ID NO:23)

ECF2-Ala 肽片段可以得自 hCA-ECF2-Ala 蛋白质构建体的切割, 例

如通过利用羧胺的处理。hCA-ECF2-Ala 蛋白质构建体可以包括序列：  
hCAII'-Met-Val-Asp-Asn-Trp-Arg-Pro-Ala-Gln-Pro-Leu-Lys-Asn-Arg-  
Gln-Ile-Lys-Ala-Phe-Val-Asp-Asp-Asp-Asn-Gly-Lys-Leu-Ser-Gln-  
Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-  
Thr-Pro-Ala (SEQ ID NO:12)

此外，可以在不同键上切割 hCA-ECF2-Ala 蛋白质构建体从而产生前 ECF2-Ala 肽，即微型融合蛋白质。例如，可以切割 hCA-ECF2-Ala 蛋白质构建体从而产生前 ECF2-Ala 肽，这些肽包括 hCAII 的 C 末端片段或者其修饰体连接在 ECF2-Ala (SEQ ID NO:23) 的 N-末端上。在本发明的一个优选实施方案中，可以用溴化氰(CNBr)切割 hCA-ECF2-Ala 蛋白质构建体 (SEQ ID NO:12) 从而产生具有下列氨基酸序列的微型融合蛋白质 (“MFP”)：

Val-Asp-Asn-Trp-Arg-Pro-Ala-Gln-Pro-Leu-Lys-Asn-Arg-Glu-Ile-Lys-  
Ala-Phe-Val-Asp-Asp-Asp-Asp-Asn-Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-  
Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-Ala  
(SEQ ID NO:24)

MFP 包括 hCAII 的 C 末端的 Val<sub>241</sub>-hCAII” 片段：

Val-Asp-Asn-Trp-Arg-Pro-Ala-Gln-Pro-Leu-Lys-Asn-Arg-Glu-Ile-Lys-  
Ala (SEQ ID NO:25)

以及具有下式的接头序列：

Phe-Val-Asp-Asp-Asp-Asp-Asn (SEQ ID NO:14).

人碳酸酐酶 II (hCAII) 的多肽片段 hCAII'-Met<sub>240</sub>-Val<sub>241</sub>-hCAII” 已在 生物化学, 9, 2638 (1970) 中有报道，并且可以用作为结合蛋白片段，用来利用如在 (WO 92/01707) 中所描述的那些方法纯化碳酸酐酶-ECF2-aa 蛋白质构建体。按照 WO 92/01707，碳酸酐酶片段或者修饰的碳酸酐酶也可以用作作为结合蛋白片段。

用于一些其它碳酸酐酶的氨基酸序列也有报道 (参见，例如 Hewett-Emmett 等，在 碳酸酐酶 中，Dodgson 等编著，第 2 章，pp.15-32, 1991)。如果特异碳酸酐酶的氨基酸序列是已知的但没有基因可供使用，那么可以利

用本领域技术人员所熟知的方法(参见, 例如 Sambrook 等, 分子克隆, 实验室手册, 冷泉港实验室, 冷泉港, N.Y, 1989; 以及 Tanhauser 等, 基因, 117, 113-117, 1992)分离编码碳酸酐酶的 cDNA。这些方法典型地包括构建编码所讨论的碳酸酐酶片段的核酸探针(例如长为大约 20-30 个碱基对)。基于已知氨基酸序列或者相关 DNA 序列的简并性核苷酸探针可以用来筛选 cDNA 文库(参见, 例如 Wallace 等, 核酸研究, 6, 3543, 1979 和 Wallace 等, 核酸研究, 9, 879, 1981)。

在本发明的另一个实施方案中, 靶序列可以包括其中 10-32 片段的至少一个 C 末端残基已为一个或多个氨基酸残基(“离去单位”)所取代的氨基酸序列。可以在除去碳酸酐酶部分的融合蛋白切割之前或者之后经由转肽基反应修饰靶序列的 C 末端。转肽基反应典型地导致离去单位从靶序列的 C 末端上除去以及它被“10-32 片段”的丢失的 C 末端残基(或者多个残基)所取代。例如, 可以利用 *S. aureus* V8 对具有 C 末端离去单位的重组产生的肽(如鳗鱼降钙素的氨基酸残基 10-30 偶联到 C 末端丙氨酸上的鳗鱼(10-30)-Ala)进行转肽基作用从而在残基 Glu<sub>30</sub> 取代 Ala C 末端残基之后引入 Thr-Pro-NH<sub>2</sub> 二肽。

除用于降钙素类似物(如 Elcatonin)的制备外, 本发明的重组合成的 10-32 片段可以用于降钙素的合成。例如, 可以利用非酶促偶联试剂将鳗鱼降钙素的环氧化形式的残基 1-9 与 ECF2 酰胺的侧链保护的衍生物偶联, 例如, 其中 Ser、Thr 和 Glu 残基被保护作为苄基醚或者酯并且 Lys 残基为 CBZ 基团所保护。可以用来实现偶联反应的合适的非酶促偶联试剂的例子包括 DCC、N-乙基-N'-二甲基氨丙基-碳二亚胺(“EDAPC”)、DCC/HOSu、DCC/HOBt 以及 EDAPC/HOSu(参见, 例如美国专利 5,429,129, 其公开内容并入本文作为参考)。

## I. 融合蛋白的重组形成

### a. 载体

选择掺入融合蛋白基因的表达载体以便使之可与宿主细胞相容。一般地, 所利用的载体具有许多共同的特性。这些特性包括可与宿主细胞相容

的复制起点、用于转录和转录调节(对于可诱导系统)的调节 DNA 序列、有效的核糖体结合位点(对于原核宿主)、聚 A 信号(对于真核宿主)。此外,可以包括表型基因、调节区以及前导序列。

原核载体(如那些用于在大肠杆菌中表达的载体)的特点在于复制起点、用于转化细菌的选择的遗传标记(表型)以及将指导感兴趣的基因的表达的 DNA 调节序列。典型地,所说的调节序列将包括驱动转录的启动子、控制转录(开/关开关)的操纵基因、起始翻译的有效的核糖体结合位点以及转录终止信号。起始和终止密码子由插入(融合蛋白)的基因提供。

原核载体尤其包含合适的“表达盒”,该表达盒基于一些可供使用的启动子/操纵基因系统之任一个。包含在原核载体中的典型的启动子包括乳糖、色氨酸、T7、脂蛋白、碱性磷酸酶、 $\lambda$  左或右启动子或者它们的混合物(杂合启动子)。乳糖和色氨酸操纵基因以及温敏的 $\lambda$  启动子是典型的开/关开关,这种开关可以包括在原核载体中。包含在原核载体中的典型的表型标记包括用于产生对氯苄青霉素、四环素、卡那霉素以及氯霉素的抗性的基因。

真核载体(如那些用于在酵母、酿酒酵母中表达的载体)典型地是穿梭载体,该穿梭载体包含用于大肠杆菌的复制起点和用于酿酒酵母的复制起点、两种细胞类型的遗传标记以及将指导在酵母中表达的 DNA 调节序列。所说的调节序列典型地包括启动子、调节序列以及转录终止信号(包括聚腺苷酸化信号)。用于指导细胞分泌的可有可无的信号序列也可以插入到真核载体中。将要掺入的典型的标记对基因中的突变的互补提供阳性选择,该基因对于尿嘧啶、亮氨酸、组氨酸、腺嘌呤、色氨酸等的产生是必要的。可以优选地掺入真核载体中的启动子序列包括醇脱氢酶 I 或者 II、磷酸甘油醛脱氢酶、磷酸甘油酸激酶(phosphoglycerokinase)、半乳糖、色氨酸、交配因子 $\alpha$  等等。

依据所选择的载体的性质,可以在各种有机体中表达 ECF2-aa 基因片段。具有特异宿主范围的载体和具有广泛宿主范围的载体都适合用于本发明中。具有特异宿主范围的载体的例子,例如对于大肠杆菌来说,是 pBR322 (Bolivar 等, 基因, 2, 95-113, 1977)、PUC18/19 (Yanisch Perron

等, 基因, 23, 103-119, 1985)、 pK18/19 (Pridmore, 基因, 56, 309-312, 1987)、 pRK290X (Alvarez-Morales 等, 核酸研究, 14, 4207-4227, 1986) 和 pRA95(可得自 Nycomed Pharm a AS, Huidove, 丹麦)。

可以使用的其它载体是适合用于革兰氏阴性细菌中的“广泛宿主范围”载体。这样的“广泛宿主范围”载体的例子是 pRK290 (Ditta 等, Proc. Nat. Acad. Sci., 77, 7347-7351, 1980)、 pKT240 (Bagdasarian 等, 基因, 26, 273-282, 1983)、 如 pLAFR1 (Long 等, 自然, 289, 485-488, 1982)的 pRK290 衍生物、 如 pMMB66EH (Furste 等, 基因, 48, 119-131, 1986)或者 pGSS33 (Sharpe, 基因, 29, 93-102, 1984)的 pKT240 衍生物。

#### b. 质粒 pTBN 的构建

包含 hCAII 基因的表达载体 pET31F1mhCAII 获自佛罗里达大学的 P.J Laipis 博士。氨苄青霉素抗性的 T7 表达载体由在 Tanhauser 等, 基因, 117, 113-117 (1992)中所描述的方法从质粒 pET-3c (Studier 等, Methods Enzymol., 185, 60-89, 1990)中构建。将得自 pET-3c 载体的 T7 表达盒转移到截短的 pSP65 质粒上。为了使单链质粒 DNA 能够产生, 将得自 pEMBL8+ (Dente 等, 核酸研究, 11, 1645-1655, 1983)的 F1 起点连接到阳性克隆的 Bgl II 限制酶切位点中。指定所产生的质粒为 pET31F1m, 其中 F1m 表示 F1 起点具有 pSP65 复制起点的相反方向。最后, 将人碳酸酐酶 II cDNA (hCAII cDNA)克隆在 pET31F1m 质粒中 Nde<sup>±</sup>和 BamH I 位点之间, 从而给出指定为 pET31F1mhCAII(参见图 1)的质粒。

通过首先在佛罗里达大学的 DNA Synthesis Core of the Interdisciplinary Center for Biotechnology 合成两种互补寡核苷酸来建立质粒的接头区:

5' A GCT TTC GTT GAC GAC GAC GAT ATC TT 3'

(SEQ ID NO:26)

5' AGC TAA GAT ATC GTC GTC GTC AAC GAA 3'

(SEQ ID NO:27)

将上述寡核苷酸磷酸化、退火并连接到邻近 hCAII 基因序列的羧基末端的 Hind III 位点中。包含插入片段的质粒具有唯一的 EcoR V 位点,

该位点紧随多肽片段即接头(SEQ ID NO:14)中的第四个天冬氨酸残基。所产生的质粒称作为 pABN(参见图 2)。

利用质粒 pABN 和 pBR322 (Bolivar 等, 基因, 2, 95-113, 1977)构建四环素抗性的表达载体, 该表达载体用于 ECF2-aa (SEQ ID NO:27)的产生, 例如其中 aa 是 Ala。将得自 pABN 的 1.5-kb Ssp I/BspE I 片段插入到 pBR322 的 ScaI 位点中, 结果产生 pTBN (参见图 3)。

### c. 宿主菌株和转化

用于按照本发明进行限制性切割、连接、转化、选择、培养和裂解的方法中的过程一般采用本领域所知的标准方法。提供这些方法的细节的文献包括 Sambrook 等, 分子克隆, 实验室手册, 冷泉港实验室, 冷泉港, N.Y. (1989), 其公开内容并入本文作为参考。

为了制备用于发酵的生产菌株, 把编码 hCA-ECF2-aa 的 DNA 片段引入适合于 hCA-ECF2 融合蛋白表达的宿主菌株。适合于表达这一基因的微生物(典型地包括具有底物和原材料的高耐受性的菌株)的例子是肠杆菌, 如埃希氏菌属。大肠杆菌种的微生物是特别优选的。该微生物可以包含在载体分子上的或者在整合到它们的染色体中的 hCA-ECF2-aa DNA 片段。用包含 hCA-ECF2-aa DNA 片段的载体利用本领域技术人员所熟知的方法(参见, 例如 Sambrook 等, 上述引用的)转化所选择的微生物。合适的生产菌株的例子是大肠杆菌 JM 109 (DE3)和大肠杆菌 BL21 (DE3), 在所说的每一种生产菌株的情况下都包含编码融合蛋白的质粒, 例如 pTBN26。

真核细胞可以包括单细胞生物(如酵母细胞)以及得自高等生物的无限增殖细胞(如植物、昆虫或者哺乳动物细胞)。合适的真核宿主细胞包括酿酒酵母、巴斯德毕赤酵母、黑曲霉、草地夜蛾以及玉米、烟草或者大豆植物细胞。可作为宿主的高等生物包括具有可进行转化的生殖细胞的高等植物和动物。包括在其中的是如烟草、玉米、大豆和有果植物的植物和鱼、鸟的无脊椎动物和脊椎动物以及如绵羊、山羊、母牛, 马和猪的哺乳动物。

典型地, 将所转化的宿主菌株从加有抗生素的选择性营养培养基分离出来, 其中由于标记基因定位在载体上面因而宿主菌株具有对抗生素的抗

性。

#### d. 发酵

利用微生物实现 hCA-ECF2-aa 融合蛋白的重组制备，该微生物包含 hCA-ECF2-aa DNA 片段和/或载体质粒(包含 hCA-ECF2-aa 片段)。利用本质上已知的方法可以实现上述过程，例如利用在 WO 92/01707 中所描述的方法。据此，可以利用市售的生长培养基，如 Luria 肉汤。分批向发酵液中补加氧和氨基酸补料以提供高细胞密度。在发酵完成之后，可以如同在 WO 92/01707 中所描述的一样破坏微生物，并且纯化 hCA-ECF2-aa 融合蛋白，例如利用如 WO 92/01707 中所描述的磺胺亲和层析。

#### II. 产生 ECF2-Ala 的融合蛋白的切割

可以在氨基酸 Asn 7 (接头序列的氨基酸 7)和 Gly A<sub>10</sub> (ECF2-Ala 的氨基酸 1)之间(由下列的“\*”指出)切割具有以下所显示的序列的融合蛋白：

hCAII'-Met-Val-Asp-Asn-Trp-Arg-Pro-Ala-Gln-Pro-Leu-Lys-Asn-Arg-Gln-Ile-Lys-Ala-Ser-Phe-Val-Asp-Asp-Asp-Asp-Asn\*-Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-Ala

(SEQ ID NO:12)

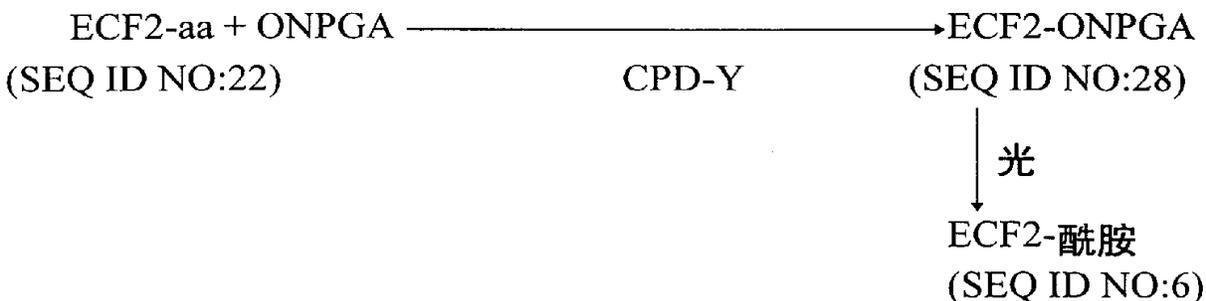
通过用羟胺处理可以实现在该位点的特异性切割。所产生的 ECF2-Ala 片段(SEQ ID NO:23)可以由常规方法纯化。

#### III. 形成 ECF2-酰胺(SEQ ID NO:6)的 ECF2-aa (SEQ ID NO:22)的酰胺化

通过在有羧肽酶 Y(参见下述的方案 1)的情况下利用亲核氨基化合物转酰胺化如 ECF2-Ala (SEQ ID NO:23)的 ECF2-aa 肽，按照 WO 92/05271 可以实现 ECF2-aa (SEQ ID NO:22)的酰胺化，从而形成能够被反应或者分解而形成 ECF2-NH<sub>2</sub>(SEQ ID NO:6)的肽中间体。合适的亲核体的例子包括邻硝基苄胺，如邻硝基苄基甘氨酸酰胺(“ONPGA”)。转酰胺化导致 ECF2-aa (SEQ ID NO:22)的 C 末端-aa 残基为亲核胺所取代。所产生的光不稳定中间体(例如 ECF2-ONPGA (SEQ ID NO:28))的光解导致亲核体的切割和在羧基末端具有酰胺化的脯氨酸的 ECF2 片段(“ECF2-酰胺”，

SEQ ID NO:6)的产生。 ECF2-酰胺可以由常规方法纯化。

### 方案 1

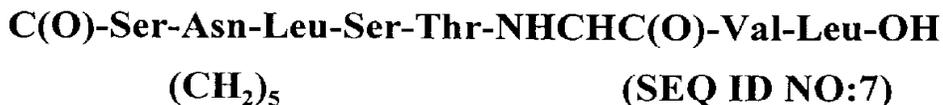


### IV. Elcatonin 片段 1 (ECF1)和 ECF2-酰胺的缩合

利用如本文所描述的标准的肽偶联反应，例如在有 DCC/HOSu、EDAPC/HOSu 或者 DCC/HOBt 的情况下，可以偶联 N 末端 Elcatonin 片段 ECF1 (SEQ ID NO:7)和 Elcatonin 相关片段 ECF2-Xxx (SEQ ID NO:1)。如果利用无保护形式的 ECF1 和 ECF2-酰胺进行偶联反应，那么 ECF1 片段的 C 末端 $\alpha$ -羧酸在 ECF2 片段加入之前就典型地转化成为活化酯。通过偶联反应产生的 Elcatonin (SEQ ID NO:13)可以由常规技术(如制备性 HPLC 方法)纯化。

利用 ECF1 (SEQ ID NO:7)和其它 10-32 片段，例如鲑鱼 I 降钙素 10-32 片段(SEQ ID NO:29)，用类似的方法可以制备其它降钙素的 Carba 类似物(例如鲑鱼 I 降钙素, (SEQ ID NO:38))和相关降钙素碳类似物。通过得自一物种的降钙素的 C 末端片段和与得自另一不同物种的降钙素相应的 N 末端碳片段的结合，本发明也可以制备碳类似物。

### 方案 2



+

Gly-Lys-Leu-Ser-Gln-Glu-Leu-His-Lys-Leu-Gln-Thr-  
Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-Thr-Pro-NH<sub>2</sub>

(SEQ ID NO:6)

在非酶促偶联试剂存  
在下缩合

Elcatonin (SEQ ID NO:13)

#### V. 给出微型融合蛋白质(MFP)的融合蛋白 hCA-ECF2-aa 的切割

在本发明的另一个实施方案(其在 Elcatonin (SEQ ID NO:13)的形成中使用化学偶联反应)中, 经由 CNBr 处理可以在 Met<sub>240</sub> 和 Val<sub>241</sub> 残基之间切割 Met-hCAII'-Met<sub>240</sub>-Val<sub>241</sub>-hCAII''-接头-ECF2-Ala 融合蛋白 (SEQ ID NO:12), 从而产生 48 氨基酸微型融合蛋白质(Val<sub>241</sub>-hCAII''-接头-ECF2-Ala, 下文的“MFP” (SEQ ID NO:24)). MFP 由 3 种组分片段组成。微型融合蛋白质的氨基末端部分基本上由在 ECF2-Ala 片段(SEQ ID NO:23)的 N-末端上的 24 残基“生物保护基团”组成。可以由常规方法纯化 MFP (SEQ ID NO:24), 其后使之衍生从而保护氨基酸侧链残基。

#### VI. 向微型融合蛋白质引入保护基团

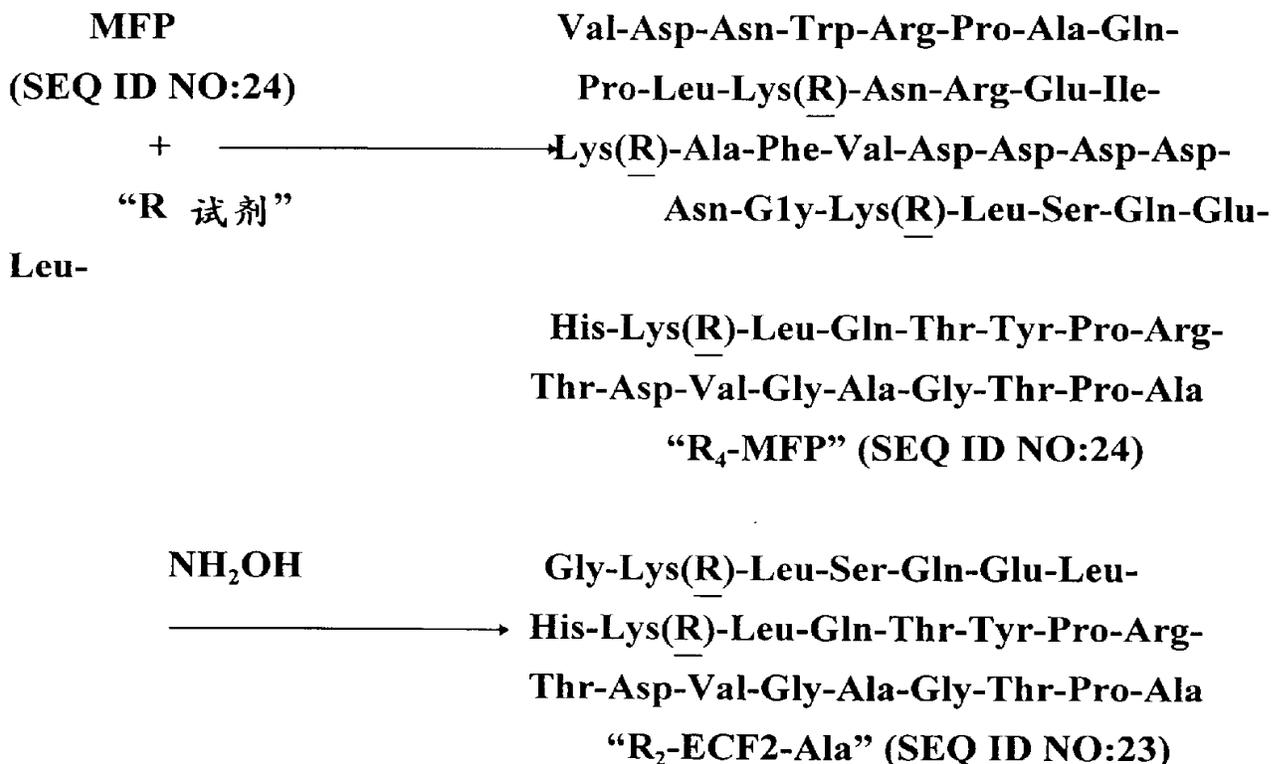
当要以化学保护形式分离 ECF3-aa (SEQ ID NO:22)时, 可以从切割之后的融合蛋白中分离出微型融合蛋白质, 然后使之接受化学衍生化作用从而用各种保护基团(“R”)封阻活性侧链基团, 如赖氨酸残基的ε氨基功能团。N 末端肽片段起保护基团的作用, 保护 ECF2-aa 的 N 末端 Gly 的α-氨基不受化学保护试剂衍生化。

在多肽化学中常用的氨基保护基团 R, 例如那些在 Houben-Weyl (Methoden der organischen Chemie 15/1 和 15/2, Thieme Verlag Stuttgart, 1974)中描述的保护基团, 都适合于用作为保护基团。优选的氨基保护基团包括苄氧基羰基- (“Z”), 叔丁氧基羰基- (“BOC”), 苄基甲氧基羰基- (“FMOC”)或者金刚烷氧基羰基- (“ADOC”)。其它活性侧链基团也可以由保护基团保护, 例如可以保护羟基和羧基分别作为苄基醚和苄基酯。

#### VII. 保护的 ECF2-Ala 片段的制备

通过首先用适当的衍生化试剂(“R”试剂)处理微型融合蛋白质可以形成保护的 10-32 片段, 其中 R 是氨基保护基团(其中 n 与微型融合蛋白质中的赖氨酸残基的数量相对应)。然后切割所产生的保护的微型融合蛋白质(“R<sub>n</sub>-MFP” (SEQ ID NO:24))从而形成保护的 10-32 片段(“R<sub>n</sub>-ECF2-Ala” (SEQ ID NO:23)), 其中只有 Lys 残基的自由ε-氨基为“R”基团所保护而α-氨基没有得到保护(参见以下的方案 3)。可以以化学方式或者酶促方式进行切割。对于化学切割, 羟胺可以用来切割 Asn-Gly 键(用箭头指出), 例如按照由 Bomstein, 生物化学, 12, 2408-2421, 1979 所描述的方法, 从而产生 R<sub>n</sub>-ECF2-Ala 的片段(SEQ ID NO:23)。其后可以由常规化学方法(如果需要的话)纯化所产生的 R<sub>n</sub>-ECF2-Ala (SEQ ID NO:23)。在本发明的这一实施方案的一个例子中, 通过使 MFP 与 BOC-酸酐进行反应可以形成 BOC 保护的 Val<sub>241</sub>-hCAII”-接头-ECF2-Ala (“BOC<sub>4</sub>-MFP” (SEQ ID NO:24))。可以用羟胺在 Asn 和 Gly 氨基酸残基之间切割所产生的 BOC<sub>4</sub>-MFP (SEQ ID NO:24)从而产生 BOC<sub>2</sub>-ECF2-Ala (SEQ ID NO:23)。

### 方案 3



## VIII. 产生非酰胺化的 Elcatonin 的保护性 R<sub>n</sub>-ECF2-aa 与 ECF1 片段的偶联

本发明还涉及 ECF2-aa (SEQ ID NO:22) 或者其修饰形式(如 R<sub>n</sub>-ECF2-aa)在如 Elcatonin (SEQ ID NO:13)的降钙素类似物的制备中的用途。在这一实施方案中,用本领域技术人员所熟知的方法,利用非酶促偶联试剂可以首先使 R<sub>n</sub>ECF2-aa (SEQ ID NO:22)的自由α-氨基与肽片段(如 ECF1 (SEQ ID NO:7))或者保护的 ECF1 的自由羧基缩合,从而产生 Elcatonin 前体,如 Elcatonin-Ala。



### “Elcatonin-Ala” (SEQ ID NO:31)

例如,用其中 Ser 和 Thr 残基的羟基和 Lys 残基的侧链氨基如同以上所描述的一样得到保护的起始物可以实现用来产生 Elcatonin 前体的缩合。该缩合导致形成侧链保护的、氨基酸延伸的、非酰胺化形式的 Elcatonin (“R<sub>n</sub>-Elcatonin-Ala (SEQ ID NO:31))。

由碳二亚胺或者叠氮化物方法,利用已知的方式可以实现缩合。在两个片段已经以化学方式缩合之后,例如通过如(Houben Weyl, Meth. der Org. Chemie, 15, 1-2 Thieme Verlag, Stuttgart, 1974)所描述的氢解、水解、酸化、还原或者肼解,利用适合于具体保护基的常规技术可以使产物脱保护。

## IX. Elcatonin-Ala 向 Elcatonin 的转化

按照 WO 92/05271, 通过在有羧肽酶的情况下使前体多肽 Elcatonin-aa (SEQ ID NO:30)与亲核化合物(例如 ONPGA)进行反应以给出可切割的中间体,然后该中间体可以由光解切割从而给出 Elcatonin(其作为 C 末端α-甲酰胺存在),如此可以实现氨基酸延伸的 Elcatonin 前体向具有下式的 Elcatonin 的转化:

(CH<sub>2</sub>)<sub>5</sub>

C(O)-Ser-Asn-Leu-Ser-Thr-NHCHC(O)-Val-Leu-Gly-Lys-Leu-Ser-Gln-  
Glu-Leu-His-Lys-Leu-Gln-Thr-Tyr-Pro-Arg-Thr-Asp-Val-Gly-Ala-Gly-  
Thr-Pro-NH<sub>2</sub> (SEQ ID NO:13)

本文用于氨基酸的缩写是: Gly, 甘氨酸; Lys, L-赖氨酸; Leu, L-亮氨酸; Ser, L-丝氨酸; Gln, L-谷氨酰胺; Glu, L-谷氨酸; His, L-组氨酸; Thr, L-苏氨酸; Tyr, L-酪氨酸; Pro, L-脯氨酸; Arg, L-精氨酸; Asp, L-天门冬氨酸; Val, L-缬氨酸; Ala, L-丙氨酸; Met, L-蛋氨酸; Asn, L-天冬酰胺; Trp, L-色氨酸; Ile, L-异亮氨酸; 以及 Phe, L-苯丙氨酸。

### 实施例

#### 1.1 包含编码 hCA-ECF2-Ala 融合蛋白的 DNA 的质粒 pTBN26 的制备

利用 PCR 方法构建编码 ECF2 的氨基酸 10-32 的基因。在佛罗里达大学合成下列五种寡核苷酸:

1. 5' GGA TCC AAG CTT GTT A 3' (SEQ ID NO:32)
2. 5' GTC GAC GAA TTC GAT A 3' (SEQ ID NO:33)
3. 5' GGA TCC AAG CTT GTT AAC GGT AAA CTG TCT CAG  
GAG CTC CAT AAA CTG 3' (SEQ ID NO:34)
4. 5' CTG ACG TTG GTG CTG GTA CCC CGG CTT AAG ATA  
TCG AAT TCG TCG AC 3' (SEQ ID NO:35)
5. 5' GGT ACC AGC ACC AAC GTC AGT ACG CGG GTA AGT  
CTG CAG TTT ATG GAG CTC CTG AGA 3' (SEQ ID NO:36)

将寡核苷酸 2-5 混合在 PCR 反应混合物 PCR MIX 1 中。寡核苷酸 3 的 3' 末端互补于寡核苷酸 5 的 3' 末端, 而寡核苷酸 2 互补于寡核苷酸 4 的 3' 末端。将上述四种核苷酸一起退火, 如图 4A 中所示。如图 A 的虚线所指出的一样, 在 PCR 期间, Taq DNA 聚合酶用来伸展寡核苷酸 3 到寡核苷酸 5 的 5' 末端, 寡核苷酸 5 则伸展到寡核苷酸 3 的 5' 末端, 寡核苷酸 2 则伸展到寡核苷酸 4 的 5' 末端。

第二个 PCR 反应用来连接得自 PCR MIX 1 中的 PCR 产物, 结果产生双链核苷酸片段(SEQ ID NO:9), 该片段包含完全的、无间断的 Asn-ECF2-Ala 基因序列以及有利于克隆的限制酶切位点。分别得自寡核苷酸 2(2<sup>Ext</sup>(SEQ ID NO:8))和 3(3<sup>Ext</sup>(SEQ ID NO:50))的 PCR 延伸产物在它们的 3'末端(参见图 4B 中的示意说明)都具有 20 bp 的互补序列。在与 PCR MIX 2 的最初几个循环的反应期间, 寡核苷酸 2<sup>Ext</sup>(SEQ ID NO:8)和 3<sup>Ext</sup>(SEQ ID NO:50)(其中它们的序列是互补的)得到退火并延伸从而建立双链片段, 该双链片段包括 ECF2(SEQ ID NO:10)的全长无间断基因序列。最后, 寡核苷酸 1 和 2 用来扩增上述全长基因序列。图 11 和 12 显示完全的、无间断的 ECF2 基因序列(SEQ ID NO:10)的核酸序列以及分别为寡核苷酸 2<sup>Ext</sup>(SEQ ID NO:8)和 3<sup>Ext</sup>(SEQ ID NO:50)的核酸序列。

用 Hpa I 和 EcoR V 消化所扩增的 PCR 产物。在唯一的 EcoR V 位点上将这一平端 DNA 片段(编码多肽片段接头的 C 末端氨基酸(Asn)以及整个多肽片段 ECF2)插入到 pABN 质粒中。筛查该质粒以确保 Asn-ECF2 基因以正确的取向插入。所产生的质粒指定为 pABN26。最后, 用 Xba I 和 BspE I 消化 pABN26, 并将整个 Met-hCAII'-Met<sub>240</sub>-Val<sub>241</sub>-hCAII''-接头-ECF2-aa 序列转移到 pTBN 质粒(已用相同的酶消化过)中从而产生质粒 pTBN26(图 5)。

最后构建体的 DNA 序列显示在图 6 中, 该构建体具有加下划线的 Met 溴化氰切割位点和 Asn-Gly (SEQ ID NO:11)以及用箭头指出的相应的氨基酸序列(SEQ ID NO:12)切割位点。

### 1.2 质粒 pTBN26(包含 hCA-ECF2-aa DNA 片段的载体)的转化

按照在 WO 92/01707 中所描述的方法实现转化。所利用的宿主微生物是大肠杆菌 HB101 和大肠杆菌 BL21 (DE3), 两者都描述在 van Heeke 等, 蛋白质表达和纯化, 4, 265-275, 1993 中。

### 1.3 包含 hCA-ECF2-AlaDNA 的载体的选择

利用如那些在 WO 92/01707 中所描述的标准方法实现选择。选择基于四环素抗性向宿主生物中的引入。

## 2. 包含质粒 pTBN26 的大肠杆菌的发酵

### 2.1 用于接种发酵罐的预培养物的生长

在 37 °C 下在摇瓶中使包含质粒 pTBN26 的大肠杆菌菌株 BL21 (DE3) 生长在 Luria 肉汤(LB)培养基(包含四环素(15 mg/L)和葡萄糖溶液(50 mg/L))中直到在 550 nm (OD550)达到大约 4 的光密度, 该过程历时大约 14 小时。Luria 肉汤培养基(LB 培养基)的组合物是: 1.0 g/L 胰蛋白酶、1.0 g/L NaCl 和 0.5 g/L 酵母提取物。

### 2.2 发酵

在 New Brunswick MPP-80 发酵罐中完成发酵。在发酵罐中加入包含 300 g 酵母提取物、30 g NaCl 和溶于 15 L H<sub>2</sub>O 的 1200 g 酪蛋白氨基酸的发酵培养基, 其后加入 30 L 蒸馏水。在 121 °C 下将发酵罐灭菌 25 分钟。在发酵罐的内含物已冷却至 37 °C 之后, 依次加入下列无菌溶液: 在 800 mL H<sub>2</sub>O 中的 480 g 葡萄糖、在 250 mL H<sub>2</sub>O 中的 120 g MgSO<sub>4</sub>·H<sub>2</sub>O、在 3.0 L H<sub>2</sub>O 中的 495 g K<sub>2</sub>PO<sub>4</sub> 和 465 g KH<sub>2</sub>PO<sub>4</sub>、以及在 30 mL 95% EtOH 和 20 mL H<sub>2</sub>O 的混合物中的 0.9 g 盐酸四环素。此外, 还加入无菌无机混合物, 该无菌无机混合物包含 3.6 g FeSO<sub>4</sub>·7H<sub>2</sub>O, 3.6 g CaCl<sub>2</sub>·2H<sub>2</sub>O, 0.90 g MnSO<sub>4</sub>, 0.90 g AlCl<sub>3</sub>·6H<sub>2</sub>O, 0.09 g CuCl<sub>2</sub>·2H<sub>2</sub>O, 0.18 g 钼酸, 以及溶于 490.0 mL H<sub>2</sub>O 和 10.0 mL 浓 HCl 的 0.36 g CoCl<sub>2</sub>·6H<sub>2</sub>O。

将试剂级别的 NH<sub>4</sub>OH (28%) 加入到发酵罐的自动 pH 进料泵中并将 pH 调整至 6.8。利用校准电极连续地监控发酵液, 并且通过间歇地加入 NH<sub>4</sub>OH 使 pH 保持在 6.8。另外, 连续地监控溶氧, 利用校准氧探测器。通过调整搅拌速率保持氧浓度。通气速率保持在 40 L/min。当所有系统都正常运行时, 通过无菌方法加入 600 mL 接种物进行接种。

在整个发酵过程中监控浊度、溶氧以及葡萄糖水平。当浊度达到 15-20 OD(550)时, 向培养基补加 300 g 酵母提取物和 1,200 g 酪蛋白氨基酸。当搅拌速率达到 500 rpm 时, 向以 5 L/min 速率进给的空气中补加纯氧并使通气速率减少至 20 L/min 以保持 40% 的溶氧。

当浊度达到 30 OD<sub>550</sub> 单位时, 向发酵培养基提供异丙基硫代半乳糖

甚至 2 mM 以诱导质粒的表达。此外，加入足够的  $\text{ZnCl}_2$  以提供  $100\mu\text{M}$  终浓度，同时在培养基中加入无菌氨基酸溶液，该氨基酸溶液包含 225 g L-Ser 和各为 75 g 的 L-Tyr、L-Trp、L-Phe、L-Pro 和 L-His。在诱导之后第 2 小时，取细胞样品用于分析。对样品进行干重和 SDS-PAGE 蛋白质分析，其中样品在诱导时采集并且其后每 30 分钟采集一次。

在发酵结束时，在无菌条件下将培养基转移到无菌存贮槽中并冷却至  $8\text{ }^\circ\text{C}$ 。利用装备有 200 kD 分子量截断膜的 Millipore Pro Stack 通过错流过滤从废培养基中分离出细胞。将细胞浓缩至 5 L，然后立即处理或者以 1 L 等分试样包装在塑料冷藏袋中并存储在  $-20\text{ }^\circ\text{C}$  下用于以后处理。

用 32 L 裂解缓冲液(50 mM Tris, 1, mM EDTA, 0.5% Triton-X100, pH7.8, 包含 0.05 mM 基甲磺酰氟)稀释浓缩的细胞(5 L)，然后使之两次通过在  $4-15\text{ }^\circ\text{C}$  下以 12,000 psi 运行的 Gaulin APV 中试规模匀浆器。然后制备溶菌酶的  $1.3\text{ }\mu\text{M}$  溶液，培养 15 分钟(在  $4-15\text{ }^\circ\text{C}$  下)，然后使之迅速通过匀浆器。

可溶性大肠杆菌蛋白质的大约 50% 为 hCA-ECF2-aa，通过按照 Verpoorte 等，生物学化学杂志，242，4221-4229，1967 测量融合蛋白的人碳酸酐酶部分的酶促活性来评估。

### 2.3 hCA-ECF2-Ala 融合蛋白的纯化

用裂解缓冲液(没有加入甲苯磺酰氟)以 1:1 稀释获自实施例 2.2 的裂解液(32 L)，并加入聚乙烯亚胺至 0.35% 的终浓度。将上述整个溶液搅拌 20 分钟，然后以  $10,000\text{ } \times\text{ g}$  离心以除去沉淀的 DNA、RNA、非必需的蛋白质以及膜泡，然后利用 Pall Profile Filter ( $1.0\text{ }\mu\text{m}$ ) 过滤。

其后由亲和层析纯化可溶性融合蛋白。通过加入 Tris 碱将所过滤的蛋白质溶液的 pH 调整至 8.7，并以 200 ml/min 流速将该溶液装载在 1 L 的对氨基苯磺酰胺亲和树脂(van Heeke 等, *Methods. Molec. Biol.*, 36, 245-260, 1994) 柱上。在装载之后，用 5 倍柱量的 0.1 M pH9.0 Tris-硫酸盐缓冲液(包含 0.2 M  $\text{K}_2\text{SO}_4$  和 0.5 mM EDTA) 洗涤上述柱。然后用 5 倍柱量的 0.1 M pH7.0 Tris-硫酸盐缓冲液(包含 1 M NaCl) 洗涤上述树脂。用 5 倍柱量的 0.1 M pH6.8 Tris-硫酸盐缓冲液(包含 0.4 mM 硫氰酸钾和

0.5 mM EDTA)从亲和材料上洗脱重组产生的 hCA-ECF2-Ala 融合蛋白。混合包含 hCA-ECF2-Ala 融合蛋白的级分，并用乙酸处理该混合物至 pH4.0 以沉淀产物，其后离心收集该产物。冻结并冷冻干燥所产生的糊状物。这一步骤的产率是 85%。

### 3. 给出 ECF2-Gla 的 hCA-ECF2-Gla 融合蛋白的切割

#### 3.1 利用羟胺对 hCA-ECF2-Ala 融合蛋白的切割

用羟胺在 Asn 和 Gly 之间切割融合蛋白而成为包含 hCAII 的片段以及多肽片段 ECF2-Ala (SEQ ID NO:23)。通过将溶解于羟胺缓冲液(2 M 盐酸羟胺, 5 M 盐酸胍, 50 mM 3-(环己基氨基)-2-羟基-1-丙烷磺酸, 用氢氧化锂调整 pH 至 10)中的 40 g hCA-ECF2-Ala 培养 4 小时来实现切割。每小时取出一等分试样, 融合蛋白切割成为 ECF2-Ala 的程度则由 HPLC 分析(C18 Vydac, 4.6 x 300 mM, 缓冲液 A: 0.1%三氟乙酸, 5%体积的乙腈, 95%体积水; 缓冲液 B: 0.1%三氟乙酸, 5%体积水, 95%体积乙腈; 5%缓冲液 A 至 68%缓冲液 B 的线性梯度; 流速: 1 ml/min)测定。然后用 15%乙酸将反应产物稀释至 4 L, 所沉淀的包含 hCAII 的片段由 10000 x G 的离心除去。

离心所产生上清液包含 ECF2-Ala (SEQ ID NO:23), 然后通过将其以 50 ml/min 流速装载到用在 5%乙腈中的 10 mM 乙酸平衡的制备性 C8 柱(5x5.1 cm)上来除去上清液的盐分。在装载之后, 用在 10%乙腈中的 10 mM 乙酸洗涤上述柱, 并用在 45%乙腈中的 10 mM 乙酸洗脱 ECF2-Ala。包含 ECF2-Ala 的级分由如上所述的分析型 HPLC 确定, 集中并冻干这些级分。获得总计 1.14 g 的 ECF2-Ala (SEQ ID NO:23), 相应的产率为 82%。

#### 3.2 ECF2-Ala 的纯化

为了进一步纯化 ECF2-Ala (SEQ ID NO:23), 利用了如 5.3 部分所详尽描述的半制备性聚磺乙基天冬酰胺 HPLC, 其产率为 85%。

### 4. ECF2-Ala 向 ECF2-酰胺的转化

按照在 WO 92/05271 中所描述的方法实现 ECF2-Ala (SEQ ID NO:23)的酰胺化。将 ECF2-Ala (12.8 mg)溶解在 1 mL 的 5 mM EDTA、25 mM 吗啉代乙烷-磺酸、pH7.0 中,并在其中加入 72 mg 邻硝基苯基甘氨酸酰胺(ONPGA)。用 5 M NaOH 将 pH 调整至 6.0 并加入羧肽酶-Y (120 $\mu$ g)。在暗处搅拌大约 48 小时之后,加入乙腈(1 mL),并由如 5.2 部分的 C18 反相 HPLC 纯化 ECF2-ONPGA (SEQ ID NO:28)产物,冻干所得产物。接着酰胺化反应过程的步骤是在 10、60、120 和 180 分钟之后提取样品,提取的样品用于分析型 C18 HPLC (如 WO 92/05271 中所描述的)的分析。

为了进行尔后的光解,将冻干的 ECF2-ONPGA (12mg)溶解在 5mL 的 50%乙醇中。在该溶液中加入 NaHSO<sub>3</sub> (26 mg)和苯甲酸钠(7.2 mg)并用 5 M NaOH 将 pH 调整至 9.5。然后使氮气通过反应混合物 15 分钟。按照在 WO 92/05271 中所描述的方法进行尔后的光解、反应过程的分析以及所产生的 ECF2-酰胺(SEQ ID NO:6)的鉴定。

## 5. 产生微型融合蛋白质的 hCA~ECF2-Ala 融合蛋白的切割

### 5.1 利用溴化氰的 hCA~ECF2-Ala 融合蛋白的化学切割

当使用化学而非酶促片段偶联时, MFP (SEQ ID NO:24)首先从融合蛋白(SEQ ID NO:12)上切割下来。为了切割 Met<sub>240</sub>-Val<sub>241</sub> 键(在 ECF2-Ala 的开始之前的 24 个氨基酸),在暗处在氩气环境下在室温下利用 0.02 M 溴化氰(在 70%甲酸中)处理 40 mg/mL 的 hCA-ECF2-Ala 融合蛋白(SEQ ID NO:12)6 小时。加入蛋氨酸(0.03 M)至 0.03M 的终浓度以终止切割反应,并且搅拌所产生的溶液 30 分钟。在反应混合物中两次加入反应量的包含 10%乙酸和 112g/L 硫酸钠的溶液以沉淀包含 hCAII”的片段,同时搅拌混合物 20 分钟。离心(10 分钟, 10,000 x G)除去所沉淀的材料。微型融合蛋白质(SEQ ID NO:24)仍然溶解在上清液中。获得相对于所使用的 hCA-ECF2-Ala 融合蛋白(SEQ ID NO:12)的 58% MFP(SEQ ID NO:24)回收率。

### 5.2 使微型融合蛋白质脱盐

将得自酸沉淀的上清液(4.5 g 微型融合蛋白质)装载在 C8 Vydac 半制备柱(22x250 mM)上, 该柱已用 0.1%三氟乙酸和 5%乙腈平衡。利用 0.1%三氟乙酸(在 25%乙腈中)洗脱上述蛋白质, 尔后冻干。获得总计 4.05 g 的 85%纯微型融合蛋白质, 相应的产率为 90%。

### 5.3 利用聚磺乙基天冬酰胺 HPLC 的微型融合蛋白质的纯化

将聚磺乙基天冬酰胺层析用于纯化微型融合蛋白质(SEQ ID NO:24)、 ECF2-Ala (SEQ ID NO:23)和 ECF2-酰胺(SEQ ID NO:6)。将肽溶解在缓冲液 A (25 mM 乙酸, 35%乙腈)中从而使浓度为 5 mg/mL, 然后以 20 ml/min 流速将其装载在聚磺乙基天冬酰胺 HPLC 柱(2.2x25cm)上。其后就可能利用 10%缓冲液 B(25 mM 乙酸, 400 mM 乙酸钠, 35%乙腈)至 47%缓冲液 B 的线性梯度洗脱蛋白质, 该过程历时 30 分钟。产率典型地是 50%或者更大。图 8 显示包含 ECF2-酰胺(SEQ ID NO:6)的样品的代表性 HPLC 图。 ECF2-酰胺的峰值出现在 14.5 和 16.3 分钟之间。

### 5.4 聚磺乙基天冬酰胺 HPLC 后的脱盐

冻干之后, 在 C8 柱(5x20 cm, 用 95%乙醇平衡)上装载蛋白质用于脱盐。然后用 4 倍柱量的水和用 4 倍柱量的 5%乙醇(包含 1%乙酸)洗涤该柱。然后用 2 倍柱量的 90%乙醇洗脱肽。由 HPLC 分析各组分。产率典型地是 80%或者更大。

## 6. 保护基团向微型融合蛋白质中的掺入

### 一般方法:

在重组 MFP (SEQ ID NO:24)之内的 ECF2 片段(SEQ ID NO:6)的 $\alpha$ -氨基是被生物保护的, 即通过 N 末端 hCAII<sup>9</sup>片段的的存在得到保护。MFP 中的 Lys 基团的 $\epsilon$ -氨基被酰基供体(如 Z、 BOC、 ADOC 或者 FMOC)所保护。精氨酸-胍基功能团与 Z-OSU 的反应通过加入如 N - 羟基琥珀酰亚胺的试剂被封阻。

### 实验方法:

将 MFP (SEQ ID NO:24) (270 mg)溶解在水(20 mL)中, 同时加入二

噁烷(4 mL)和 680 $\mu$ L 的 0.1N HCl(以确保两个精氨酸-胍基功能团上形成盐)。在反应之前, 加入 78 mg N-羟基琥珀酰亚胺和 104.7 $\mu$ L 三乙胺以保护 His 氮, 然后在整个溶液中加入 169 mg Z-OSU、146 mg BOC-OSU、135 mg ADOC 氟化物或者 229 mg FMOC-OSU, 并伴随冷却。

然后在室温下搅拌反应混合物 24 小时, 其后在真空中干燥。用二氯甲烷(两次)然后用乙腈(两次)细心地研磨所产生的残渣。通过过滤收集产物并在真空中干燥。Z 保护的微型融合蛋白质的产量是 300 mg, BOC 保护的蛋白质的产量是 295 mg, ADOC 保护的和 FMOC 保护的蛋白质的产量是 310 mg。

利用保护的和无保护的肽由薄层层析对具有各自保护基团的微型融合蛋白质的完全反应进行检验。在薄层二氧化硅平板(洗脱液: 苯酚: H<sub>2</sub>O = 775: 225)上检验完全反应。当利用 540 mg MFP (SEQ ID NO:24)时 Z-OSU 反应的产率是 75%, 并且该产率相对于利用其它封阻剂所发现的产率来说是典型的。

## 7. 产生保护性 ECF2-Ala 片段的保护性微型融合蛋白质的切割

### 7.1 利用羟胺的切割

为了释放为氨基酸序列所生物保护的 $\alpha$ -氨基, 用羟胺在 Asn 和 Gly 之间切割微型融合蛋白质成为 Z-保护的 ECF2-Ala 片段(SEQ ID NO:23)和 24 个氨基酸长的包含 Trp 的片段。羟胺缓冲液的组成与 3.1 部分所描述的相同。

在 260 mg Z-保护的微型融合蛋白质(SEQ ID NO:24)中加入羟胺缓冲液(26 mL), 然后超声处理(20 秒)全部溶液并用 4M 氢氧化锂调整 pH 至 pH10.0。在 30  $^{\circ}$ C 和 pH10.0 下培养所产生的混合液多达 5 小时。然后利用浓乙酸调整 pH 至 6.0。

每小时取出一等分试样, Z-保护性微型融合蛋白质(SEQ ID NO:24)切割成为 Z-保护性 ECF2-Ala 片段(SEQ ID NO:23)的程度则由 HPLC 分析(C18 Vydac, 5 x 300 mM, 缓冲液 A: 0.1%三氟乙酸; 缓冲液 B: 0.1%三氟乙酸, 5%体积水, 95%乙腈; 5%缓冲液 A 至 68%缓冲液 B 的线性梯

度; 1 ml/min)测定。羟胺切割的终止如在 3.1 部分中所描述的一样。

利用 Vydac C18 柱由 HPLC 分析得自 C8 柱的级分。用于 HPLC 的缓冲液是那些以前所描述的缓冲液。流速是 1 mL/min。用 41%缓冲液 A 至 71%缓冲液 B 的线性梯度洗脱蛋白质。检测在 210 nm 进行。获得总计 156 mg 的 Z-保护性 ECF2-Ala (“Z-ECF2-Ala” (SEQ ID NO:23)), 相应的产率为 60%。

### 7.2 Z-ECF2-Ala 的纯化

利用半制备 C8 HPLC 纯化 Z-ECF2-Ala (SEQ ID NO:23)。将 Z-ECF2-Ala (160 mg)溶解在 5 M 乙酸(5 mL)中, 其后加入缓冲液 A(100 mM 乙酸, 5%乙腈) (5 mL)。然后将样品装载在半制备 HPLC C8 柱上(细节如 3.2 部分一样)。在羟胺切割和纯化之后, 获得 120 mg Z-ECF2-Ala (SEQ ID NO:23), 对于 C8 纯化步骤来说相应的产率为 75%。

为了进一步纯化 Z-ECF2-Ala (SEQ ID NO:23), 如以上 5.3 部分所描述的一样利用半制备聚磺乙基天冬酰胺 HPLC。利用这一方法, 可获得 59 mg 95-98%的纯 Z-ECF2-Ala, 相应的产率为 85.5%。

## 8. ECF2-酰胺和 ECF1-OMe 的缩合

在有非酶促偶联试剂的情况下, 可以偶联按照实施例 4 产生的酰胺化的 ECF2 (SEQ ID NO:6)和环状 Elcatonin 片段 ECF1 (SEQ ID NO:7)。例如, 在 ECF1(SEQ ID NO:3) (Thr 和 Ser 的侧链羟基作为苄醚被保护)和 ECF2-酰胺(SEQ ID NO:2) (具有 Lys、Ser、Thr 和/或 Glu 残基的活性侧链基团被保护)的溶液中, 可以加入二环己基碳二亚胺(DCC)和 N-羧基琥珀酰亚胺(HOSu)。在大约 0 °C 下搅拌反应 8 小时。在这一期间的片段缩合之后的是 HPLC 分析。在反应终止之后, 可以通过加入 1%三氟乙酸来稀释上述混合物。产物 Elcatonin (SEQ ID NO:13)可以由制备 HPLC 纯化。可以由 HPLC 分离和纯化未反应的 ECF2-酰胺(SEQ ID NO:2), 并使之再循环。

## 9. Z-ECF2 与 ECF1 的化学偶联

利用碳二亚胺或者叠氮化物方法(参见, 例如 Greenstein 等, 氨基酸的化学, 卷 2, John Wiley, 纽约, pp 804ff, 1016ff, 1961), 可以实现 Z-ECF2-Ala 片段(即具有为苄氧羰基所保护的侧链氨基的 EFC2-Ala)的自由 $\alpha$ -氨基与肽片段 ECF1 的自由 C 末端 $\alpha$ -羧基基团的偶联。典型地, 通过加入 Z-ECF2-Ala 片段至活化的酯(从 ECF1 片段的 C 末端 $\alpha$ -羧酸形成)中实现偶联反应。经由氢解可以实现从偶联产物中除去 Cbz 保护基团, 所产生的产物则由制备性 HPLC 纯化从而产生 Elcatonin-Ala (即“ECF1-ECF2-Ala”; SEQ ID NO:31)。

#### 10. Elcatonin-Ala 向 Elcatonin-酰胺的转变

利用在第 4 部分中所描述的方法, 可以使按照第 9 部分产生的 Elcatonin-Ala (SEQ ID NO:31)肽酰胺化。可以纯化酰胺化的 Elcatonin (SEQ ID NO:13), 其后利用在 5.3 和 5.4 部分中所描述的方法可以使之脱盐。

虽然已按照各种具体的和优选的实施方案和技术描述了本发明。然而, 应该明白的是可以进行许多变化和修饰而保持在本发明的精神与范围之内。

本说明书中所涉及的出版物表明了本发明所属领域的一般技术水平, 并且这些出版物相同程度地引入本文作为参考, 如同每一种出版物被具体地和单独地引述。

## 序列表

### (1) 一般信息

#### (i) 申请人:

(A) 名称: **BioNebraska, Inc.**

(B) 街道: **西北第 46 街 3820 号**

(C) 城市: **Lincoln**

(D) 州: **Nebraska**

(E) 邮政编码(ZIP): **68524**

(F) 国家: **美国**

(ii) 发明名称: **降钙素的重组制备方法和其在制备降钙素和相关类似物中的用途**

(iii) 序列数: **50**

#### (iv) 通讯地址:

(A) 地址: **Merchant & Gould**

(B) 街道: **3100 Norwest Center, 90 S. 7th Street**

(C) 城市: **Minneapolis**

(D) 州: **明尼苏达**

(E) 国家: **美国**

(F) ZIP: **55402**

#### (v) 计算机可读形式:

(A) 介质类型: **软盘**

(B) 计算机: **IBM 兼容机**

(C) 操作系统: **DOS**

(D) 软件: **FastSEQ 版本 1.5**

#### (vi) 当前的申请数据:

(A) 申请号: **未知**

(B) 申请日期: **1997 年 2 月 4 日**

(C) 分类号:

#### (vii) 在先申请的数据:





(A) 名称/关键词: 其它

(B) 位置: 1...6

(D) 其它信息: 注 = 在位置 1 和 6 之间的 L-氨基辛二酸键

(xi) 序列描述: SEQ ID NO:3:

Xaa Xaa Leu Ser Thr Xaa Xaa Leu

1

5

(2) SEQ ID NO:4 的信息:

(i) 序列特征:

(A) 长度: 31 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: N 末端

(vi) 原始来源:

(ix) 特征:

(A) 名称/关键词: 其它

(B) 位置: 1...6

(D) 其它信息: 注 = 在 1 和 6 位置之间的 L-氨基辛二酸键

(A) 名称/关键词: 其它

(B) 位置: 31...31

(D) 其它信息: 注 = 具有-OH、-NH<sub>2</sub>、氨基酸或者多肽的 $\alpha$ -羧基

(xi) 序列描述: SEQ ID NO:4:

Xaa Xaa Leu Ser Thr Xaa Xaa Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
1 5 10 15  
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Pro  
20 25 30

(2) SEQ ID NO:5 的信息:



(v) 片段类型: 内部

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:6:

```
Gly Lys Leu Ser Gln Glu Leu His Lys Leu Gln Thr Tyr Pro Arg
 1           5           10           15
Thr Asp Val Gly Ala Gly Thr Pro
                20
```

(2) SEQ ID 的 NO:7 信息:

(i) 序列特征:

(A) 长度: 8 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: N 末端

(vi) 来源:

(ix) 特征:

(A) 名称/关键词: 其它

(B) 位置: 1...6

(D) 其它信息: 注 = 在 1 和 6 位置之间的 L-氨基辛二酸键

(xi) 序列描述: SEQ ID NO:7:

```
Ser Asn Leu Ser Thr XaaVal Leu
```

1

5

(2) SEQ ID NO:8 的信息:

(i) 序列特征:

(A) 长度: 47 个碱基对

(B) 类型: 核酸

- (C) 链型: 单链
- (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:8:

GACGACGAAT TCGATATCTT AAGCCGGGGT ACCAGCACCA ACGTCAG

47

**(2) SEQ ID NO:9 的信息:**

- (i) 序列特征:
  - (A) 长度: 111 个碱基对
  - (B) 类型: 核酸
  - (C) 链型: 双链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (ix) 特征:
  - (A) 名称/关键词: 编码序列
  - (B) 位置: 16...90
  - (D) 其它信息:
- (xi) 序列描述: SEQ ID NO:9:

GGATCCAAGC TTGTT AAC GGT AAA CTG TCT CAG GAG CTC CAT AAA CTG CAG  
 Asn Gly Lys Leu Ser Gln Glu Leu His Lys Leu Gln  
 1 5 10

51

ACT TAC CCG CGT ACT GAC GTT GGT GCT GGT ACC CCG GCT TAAGATATCG AAT 103  
 Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro Ala  
           15                                  20                                  25

TCGTCGAC 111

**(2) SEQ ID NO:10 的信息:**

**(i) 序列特征:**

**(A) 长度: 25 个氨基酸**

**(B) 类型: 氨基酸**

**(C)链型: 单链**

**(D)拓扑结构: 线性**

**(ii) 分子类型: 蛋白质**

**(iii) 假设: 否**

**(iv) 反义: 否**

**(v) 片段类型: 内部**

**(vi) 原始来源:**

**(xi) 序列描述: SEQ ID NO:10 :**

Asn Gly Lys Leu Ser Gln Glu Leu His Lys Leu Gln Thr Tyr Pro Arg  
 1                                  5                                  10                                  15  
 Thr Asp Val Gly Ala Gly Thr Pro Ala  
                                   20                                  25

**(2) SEQ ID NO:11 的信息:**

**(i) 序列特征:**

**(A) 长度: 864 个碱基对**

**(B) 类型: 核酸**

**(C)链型: 单链**

**(D)拓扑结构: 线性**

**(ii) 分子类型: 基因组 DNA**

**(iii) 假设: 否**

**(iv) 反义: 否**

**(v) 片段类型:**

(vi) 原始来源:

(ix) 特征:

(A) 名称/关键词: 编码序列

(B) 位置: 1..864

(D) 其它信息:

(xi) 序列描述: SEQ ID NO:11:

ATG TCC CAT CAC TGG GGG TAC GGC AAA CAC AAC GGA CCT GAG CAC TGG	48
Met Ser His His Trp Gly Tyr Gly Lys His Asn Gly Pro Glu His Trp	
1 5 10 15	
CAT AAG GAC TTC CCC ATT GCC AAG GGA GAG CGC CAG TCC CCT GTT GAC	96
His Lys Asp Phe Pro Ile Ala Lys Gly Glu Arg Gln Ser Pro Val Asp	
20 25 30	
ATC GAC ACT CAT ACA GCC AAG TAT GAC CCT TCC CTG AAG CCC CTG TCT	144
Ile Asp Thr His Thr Ala Lys Tyr Asp Pro Ser Leu Lys Pro Leu Ser	
35 40 45	
GTT TCC TAT GAT CAA GCA ACT TCC CTG AGG ATC CTC AAC AAT GGT CAT	192
Val Ser Tyr Asp Gln Ala Thr Ser Leu Arg Ile Leu Asn Asn Gly His	
50 55 60	
GCT TTC AAC GTG GAG TTT GAT GAC TCT CAG GAC AAA GCA GTG CTC AAG	240
Ala Phe Asn Val Glu Phe Asp Asp Ser Gln Asp Lys Ala Val Leu Lys	
65 70 75 80	
GGA GGA CCC CTG GAT GGC ACT TAC AGA TTG ATT CAG TTT CAC TTT CAC	288
Gly Gly Pro Leu Asp Gly Thr Tyr Arg Leu Ile Gln Phe His Phe His	
85 90 95	
TGG GGT TCA CTT GAT GGA CAA GGT TCA GAG CAT ACT GTG GAT AAA AAG	336
Trp Gly Ser Leu Asp Gly Gln Gly Ser Glu His Thr Val Asp Lys Lys	
100 105 110	
AAA TAT GCT GCA GAA CTT CAC TTG GTT CAC TGG AAC ACC AAA TAT GGG	384
Lys Tyr Ala Ala Glu Leu His Leu Val His Trp Asn Thr Lys Tyr Gly	
115 120 125	
GAT TTT GGG AAA GCT GTG CAG CAA CCT GAT GGA CTG GCC GTT CTA GGT	432
Asp Phe Gly Lys Ala Val Gln Gln Pro Asp Gly Leu Ala Val Leu Gly	
130 135 140	
ATT TTT TTG AAG GTT GGC AGC GCT AAA CCG GGC CTT CAG AAA GTT GTT	480
Ile Phe Leu Lys Val Gly Ser Ala Lys Pro Gly Leu Gln Lys Val Val	
145 150 155 160	
GAT GTG CTG GAT TCC ATT AAA ACA AAG GGC AAG AGT GCT GAC TTC ACT	528
Asp Val Leu Asp Ser Ile Lys Thr Lys Gly Lys Ser Ala Asp Phe Thr	
165 170 175	
AAC TTC GAT CCT CGT GGC CTC CTT CCT GAA TCC TTG GAT TAC TGG ACC	576
Asn Phe Asp Pro Arg Gly Leu Leu Pro Glu Ser Leu Asp Tyr Trp Thr	
180 185 190	

TAC CCA GGC TCA CTG ACC ACC CCT CCT CTT CTG GAA TGT GTG ACC TGG	624
Tyr Pro Gly Ser Leu Thr Thr Pro Pro Leu Leu Glu Cys Val Thr Trp	
195 200 205	
ATT GTG CTC AAG GAA CCC ATC AGC GTC AGC AGC GAG CAG GTG TTG AAA	672
Ile Val Leu Lys Glu Pro Ile Ser Val Ser Ser Glu Gln Val Leu Lys	
210 215 220	
TTC CGT AAA CTT AAC TTC AAT GGG GAG GGT GAA CCC GAA GAA CTG ATG	720
Phe Arg Lys Leu Asn Phe Asn Gly Glu Gly Glu Pro Glu Glu Leu Met	
225 230 235 240	
GTG GAC AAC TGG CGC CCA GCT CAG CCA CTG AAG AAC AGG CAA ATC AAA	768
Val Asp Asn Trp Arg Pro Ala Gln Pro Leu Lys Asn Arg Gln Ile Lys	
245 250 255	
GCT TTC GTT GAC GAC GAC GAC AAC GGT AAA CTG TCT CAG GAG CTC CAT	816
Ala Phe Val Asp Asp Asp Asp Asn Gly Lys Leu Ser Gln Glu Leu His	
260 265 270	
AAA CTG CAG ACT TAC CCG CGT ACT GAC GTT GGT GCT GGT ACC CCG GCT	864
Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro Ala	
275 280 285	

**(2) SEQ ID NO:12 的信息:**

**(i) 序列特征:**

**(A) 长度: 288 个氨基酸**

**(B) 类型: 氨基酸**

**(C) 链型: 单链**

**(D) 拓扑结构: 线性**

**(ii) 分子类型: 蛋白质**

**(iii) 假设: 否**

**(iv) 反义: 否**

**(v) 片段类型: C 末端**

**(vi) 原始来源:**

**(xi) 序列描述: SEQ ID NO:12:**

Met	Ser	His	His	Trp	Gly	Tyr	Gly	Lys	His	Asn	Gly	Pro	Glu	His	Trp
1				5					10					15	
His	Lys	Asp	Phe	Pro	Ile	Ala	Lys	Gly	Glu	Arg	Gln	Ser	Pro	Val	Asp
			20					25					30		
Ile	Asp	Thr	His	Thr	Ala	Lys	Tyr	Asp	Pro	Ser	Leu	Lys	Pro	Leu	Ser
		35					40					45			
Val	Ser	Tyr	Asp	Gln	Ala	Thr	Ser	Leu	Arg	Ile	Leu	Asn	Asn	Gly	His
	50					55					60				
Ala	Phe	Asn	Val	Glu	Phe	Asp	Asp	Ser	Gln	Asp	Lys	Ala	Val	Leu	Lys
65				70					75						80
Gly	Gly	Pro	Leu	Asp	Gly	Thr	Tyr	Arg	Leu	Ile	Gln	Phe	His	Phe	His
				85					90					95	
Trp	Gly	Ser	Leu	Asp	Gly	Gln	Gly	Ser	Glu	His	Thr	Val	Asp	Lys	Lys
			100					105						110	
Lys	Tyr	Ala	Ala	Glu	Leu	His	Leu	Val	His	Trp	Asn	Thr	Lys	Tyr	Gly
		115					120					125			
Asp	Phe	Gly	Lys	Ala	Val	Gln	Gln	Pro	Asp	Gly	Leu	Ala	Val	Leu	Gly
	130					135					140				
Ile	Phe	Leu	Lys	Val	Gly	Ser	Ala	Lys	Pro	Gly	Leu	Gln	Lys	Val	Val
145					150					155					160
Asp	Val	Leu	Asp	Ser	Ile	Lys	Thr	Lys	Gly	Lys	Ser	Ala	Asp	Phe	Thr
				165					170					175	
Asn	Phe	Asp	Pro	Arg	Gly	Leu	Leu	Pro	Glu	Ser	Leu	Asp	Tyr	Trp	Thr
			180					185						190	
Tyr	Pro	Gly	Ser	Leu	Thr	Thr	Pro	Pro	Leu	Leu	Glu	Cys	Val	Thr	Trp
		195					200						205		
Ile	Val	Leu	Lys	Glu	Pro	Ile	Ser	Val	Ser	Ser	Glu	Gln	Val	Leu	Lys
	210					215					220				
Phe	Arg	Lys	Leu	Asn	Phe	Asn	Gly	Glu	Gly	Glu	Pro	Glu	Glu	Leu	Met
225					230					235					240
Val	Asp	Asn	Trp	Arg	Pro	Ala	Gln	Pro	Leu	Lys	Asn	Arg	Gln	Ile	Lys
				245					250					255	
Ala	Phe	Val	Asp	Asp	Asp	Asp	Asn	Gly	Lys	Leu	Ser	Gln	Glu	Leu	His
			260					265					270		
Lys	Leu	Gln	Thr	Tyr	Pro	Arg	Thr	Asp	Val	Gly	Ala	Gly	Thr	Pro	Ala
		275					280						285		

(2) SEQ ID NO:13 的信息:

(i) 序列特征:

(A) 长度: 31 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: C 末端

(vi) 原始来源:

(ix) 特征:

(A) 名称/关键词: 其它

(B) 位置: 1...6

(D) 其它信息: 注 = 在 1 和 6 位置之间的 L-氨基辛二酸键

(xi) 序列描述: SEQ ID NO:13:

```
Ser Asn Leu Ser Thr Xaa Val Leu Gly Lys Leu Ser Gln Glu Leu His
 1                5                10                15
Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro
      20                25                30
```

(2) SEQ ID NO:14 的信息:

(i) 序列特征:

(A) 长度: 7 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: 内部

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:14:

Phe Val Asp Asp Asp Asp Asn

1

5

(2) SEQ ID NO:15 的信息:

(i) 序列特征:

(A) 长度: 15 个碱基对

(B) 类型: 核酸

- (C) 链型: 单链
- (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (ix) 特征:
  - (A) 名称/关键词: 编码序列
  - (B) 位置: 1...15
  - (D) 其它信息:
- (xi) 序列描述: SEQ ID NO:15:

GAC GAC GAC GAT AAA  
 Asp Asp Asp Asp Lys  
 1 5

15

**(2) SEQ ID NO:16 的信息:**

- (i) 序列特征:
  - (A) 长度: 5 个氨基酸
  - (B) 类型: 氨基酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 蛋白质
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: 内部
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:16:

Asp Asp Asp Asp Lys  
 1 5

**(2) SEQ ID NO:17 的信息:**

- (i) 序列特征:
  - (A) 长度: 12 个碱基对
  - (B) 类型: 核酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (ix) 特征:
  - (A) 名称/关键词: 编码序列
  - (B) 位置: 1...12
  - (D) 其它信息:
- (xi) 序列描述: SEQ ID NO:17:

ATT GAA GGA AGA  
Ile Glu Gly Arg  
1

12

**(2) SEQ ID NO:18 的信息:**

- (i) 序列特征:
  - (A) 长度: 4 个氨基酸
  - (B) 类型: 氨基酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 蛋白质
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: 内部
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:18:

Ile Glu Gly Arg

1

**(2) SEQ ID NO:19 的信息:**

**(i) 序列特征:**

(A) 长度: 24 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

**(ii) 分子类型: 基因组 DNA**

**(iii) 假设: 否**

**(iv) 反义: 否**

**(v) 片段类型:**

**(vi) 原始来源:**

**(ix) 特征:**

(A) 名称/关键词: 编码序列

(B) 位置: 1...24

(D) 其它信息:

**(xi) 序列描述: SEQ ID NO:19:**

CAT CCT TTT CAT CTG CTG GTT TAT  
His Pro Phe His Leu Leu Val Tyr  
1 5

24

**(2) SEQ ID NO:20 的信息:**

**(i) 序列特征:**

(A) 长度: 8 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

**(ii) 分子类型: 蛋白质**

**(iii) 假设: 否**

**(iv) 反义: 否**



(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:22:

```
Gly Lys Leu Ser Gln Glu Leu His Lys Leu Gln Thr Tyr Pro Arg Thr
 1           5           10           15
Asp Val Gly Ala Gly Thr Pro Xaa
                20
```

(2) SEQ ID NO:23 的信息:

(i) 序列特征:

(A) 长度: 24 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: C 末端

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:23:

```
Gly Lys Leu Ser Gln Glu Leu His Lys Leu Gln Thr Tyr Pro Arg Thr
 1           5           10           15
Asp Val Gly Ala Gly Thr Pro Ala
                20
```

(2) SEQ ID NO:24 的信息:

(i) 序列特征:

(A) 长度: 48 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: C 末端
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:24:

```

Val Asp Asn Trp Arg Pro Ala Gln Pro Leu Lys Asn Arg Glu Ile Lys
 1           5           10           15
Ala Phe Val Asp Asp Asp Asp Asn Gly Lys Leu Ser Gln Glu Leu His
          20           25           30
Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro Ala
          35           40           45

```

(2) SEQ ID NO:25 的信息:

- (i) 序列特征:
  - (A) 长度: 17 个氨基酸
  - (B) 类型: 氨基酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 肽
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: 内部
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:25:

```

Val Asp Asn Trp Arg Pro Ala Gln Pro Leu Lys Asn Arg Glu Ile Lys
 1           5           10           15
Ala

```

(2) SEQ ID NO:26 的信息:

- (i) 序列特征:
  - (A) 长度: 27 个碱基对
  - (B) 类型: 核酸
  - (C) 链型: 单链

- (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:26:

AGCTTTCGTT GACGACGACG ATATCTT

27

**(2)SEQ ID NO:27 的信息:**

- (i) 序列特征:
  - (A) 长度: 27 个碱基对
  - (B) 类型: 核酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:27:

AGCTAAGATA TCGTCGTCGT CAACGAA

27

**(2) SEQ ID NO:28 的信息:**

- (i) 序列特征:
  - (A) 长度: 24 个氨基酸
  - (B) 类型: 氨基酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 肽
- (iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: N 末端

(vi) 原始来源:

(ix) 特征:

(A) 名称/关键词: 其它

(B) 位置: 24...24

(D) 其它信息: /注 = 邻-NO<sub>2</sub>-苯基甘氨酸酰胺

(xi) 序列描述: SEQ ID NO:28:

```
Gly Lys Leu Ser Gln Glu Leu His Lys Leu Gln Thr Tyr Pro Arg Thr
 1           5           10           15
Asp Val Gly Ala Gly Thr Pro Xaa
                20
```

(2) SEQ ID NO:29 的信息:

(i) 序列特征:

(A) 长度: 23 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: C 末端

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:29:

```
Gly Lys Leu Ser Gln Glu Leu His Lys Leu Gln Thr Tyr Pro Arg Thr
 1           5           10           15
Asn Thr Gly Ser Gly Thr Pro
                20
```

(2) SEQ ID NO:30 的信息:

(i) 序列特征:

(A) 长度: 32 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: N 末端

(vi) 原始来源:

(ix) 特征:

(A) 名称/关键词: 其它

(B) 位置: 1...6

(D) 其它信息: /注 = 在 1 和 6 位置之间的 L-氨基辛二酸键

(xi) 序列描述: SEQ ID NO:30:

```
Ser Asn Leu Ser Thr Xaa Val Leu Gly Lys Leu Ser Gln Glu Leu His
 1                5                10                15
Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro Xaa
                20                25                30
```

(2)SEQ ID NO:31 的信息:

(i) 序列特征:

(A) 长度: 32 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: N 末端

(vi) 原始来源:

(ix) 特征:

(A) 名称/关键词: 其它

(B) 位置: 1...6

(D) 其它信息: /注 = 在 1 和 6 位置之间的 L-氨基辛二酸键

(xi) 序列描述: SEQ ID NO:31:

```
Ser Asn Leu Ser Thr Xaa Val Leu Gly Lys Leu Ser Gln Glu Leu His
 1           5           10           15
Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro Ala
 20           25           30
```

(2) SEQ ID NO:32 的信息:

(i) 序列特征:

(A) 长度: 16 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 基因组 DNA

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型:

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:32:

GGATCCAAGC TTGTTA

16

(2) SEQ ID NO:33 的信息:

(i) 序列特征:

(A) 长度: 16 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:33:

GTCGACGAAT TCGATA

16

**(2) SEQ ID NO:34 的信息:**

- (i) 序列特征:
  - (A) 长度: 48 个碱基对
  - (B) 类型: 核酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:34:

GGATCCAAGC TTGTTAACGG TAAACTGTCT CAGGAGCTCC ATAAACTG

48

**(2) SEQ ID NO:35 的信息:**

- (i) 序列特征:
  - (A) 长度: 47 个碱基对
  - (B) 类型: 核酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否

- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (xi) 序列描述: **SEQ ID NO:35:**

CTGACGTTGG TGCTGGTACC CCGGCTTAAG ATATCGAATT CGTCGAC

47

## (2) SEQ ID NO:36 的信息:

- (i) 序列特征:
  - (A) 长度: 57 个碱基对
  - (B) 类型: 核酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 基因组 DNA
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型:
- (vi) 原始来源:
- (xi) 序列描述: **SEQ ID NO:36:**

GGTACCAGCA CCAACGTCAG TACGCGGGTA AGTCTGCAGT TTATGGAGCT CCTGAGA

57

## (2) SEQ ID NO:37 的信息:

- (i) 序列特征:
  - (A) 长度: 32 个氨基酸
  - (B) 类型: 氨基酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 肽
- (iii) 假设: 否
- (iv) 反义: 否

(v) 片段类型: N 末端

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:37:

```
Cys Ser Asn Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Glu Leu
 1           5           10           15
His Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro
          20           25           30
```

**(2) SEQ ID NO:38 的信息:**

(i) 序列特征:

(A) 长度: 32 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 肽

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型: N 末端

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:38:

```
Cys Ser Asn Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Glu Leu
 1           5           10           15
His Lys Leu Gln Thr Tyr Pro Arg Thr Asn Thr Gly Ser Gly Thr Pro
          20           25           30
```

**(2) SEQ ID NO:39 的信息:**

(i) 序列特征:

(A) 长度: 32 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑结构: 线性

- (ii) 分子类型: 肽
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: N 末端
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:39:

```

Cys Ser Asn Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Asp Leu
 1           5           10
His Lys Leu Gln Thr Phe Pro Arg Thr Asn Thr Gly Ala Gly Val Pro
          20           25           30

```

**(2) SEQ ID NO:40 的信息:**

- (i) 序列特征:
  - (A) 长度: 32 个氨基酸
  - (B) 类型: 氨基酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 肽
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: N 末端
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:40:

```

Cys Ser Asn Leu Ser Thr Cys Met Leu Gly Lys Leu Ser Gln Asp Leu
 1           5           10
His Lys Leu Gln Thr Phe Pro Arg Thr Asn Thr Gly Ala Gly Val Pro
          20           25           30

```

**(2) SEQ ID NO:41 的信息:**

- (i) 序列特征:
  - (A) 长度: 32 个氨基酸

- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑结构: 线性
- (ii) 分子类型: 肽
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: N 末端
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:41 :

```

Cys Ala Ser Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Glu Leu
 1           5           10           15
His Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro
          20           25           30

```

**(2) SEQ ID NO:42 的信息:**

- (i) 序列特征:
  - (A) 长度: 32 个氨基酸
  - (B) 类型: 氨基酸
  - (C) 链型: 单链
  - (D) 拓扑结构: 线性
- (ii) 分子类型: 肽
- (iii) 假设: 否
- (iv) 反义: 否
- (v) 片段类型: N 末端
- (vi) 原始来源:
- (xi) 序列描述: SEQ ID NO:42 :

```

Cys Gly Asn Leu Ser Thr Cys Met Leu Gly Thr Tyr Thr Gln Asp Phe
 1           5           10           15
Asn Lys Phe His Thr Phe Pro Gln Thr Ala Ile Gly Val Gly Ala Pro
          20           25           30

```

**(2) SEQ ID NO:43 的信息:**

**(i) 序列特征:**

**(A) 长度: 32 个氨基酸**

**(B) 类型: 氨基酸**

**(C) 链型: 单链**

**(D) 拓扑结构: 线性**

**(ii) 分子类型: 肽**

**(iii) 假设: 否**

**(iv) 反义: 否**

**(v) 片段类型: N 末端**

**(vi) 原始来源:**

**(xi) 序列描述: SEQ ID NO:43:**

```
Cys Gly Asn Leu Ser Thr Cys Met Leu Gly Thr Tyr Thr Gln Asp Leu
 1           5           10
Asn Lys Phe His Thr Phe Pro Gln Thr Asp Ile Gly Val Val Ala Pro
          20           25           30
```

**(2) SEQ ID NO:44 的信息:**

**(i) 序列特征:**

**(A) 长度: 32 个氨基酸**

**(B) 类型: 氨基酸**

**(C) 链型: 单链**

**(D) 拓扑结构: 线性**

**(ii) 分子类型: 肽**

**(iii) 假设: 否**

**(iv) 反义: 否**

**(v) 片段类型: N 末端**

**(vi) 原始来源:**

**(xi) 序列描述: SEQ ID NO:44:**

Cys Ser Asn Leu Ser Thr Cys Val Leu Ser Ala Tyr Trp Arg Asn Leu  
 1 5 10 15  
 Asn Asn Phe His Arg Phe Ser Gly Met Gly Phe Gly Pro Glu Thr Pro  
 20 25 30

**(2) SEQ ID NO:45 的信息:**

**(i) 序列特征:**

**(A) 长度: 32 个氨基酸**

**(B) 类型: 氨基酸**

**(C) 链型: 单链**

**(D) 拓扑结构: 线性**

**(ii) 分子类型: 肽**

**(iii) 假设: 否**

**(iv) 反义: 否**

**(v) 片段类型: N 末端**

**(vi) 原始来源:**

**(xi) 序列描述: SEQ ID NO:45 :**

Cys Ser Asn Leu Ser Thr Cys Val Leu Ser Ala Tyr Trp Lys Asp Leu  
 1 5 10 15  
 Asn Asn Tyr His Arg Phe Ser Gly Met Gly Phe Gly Pro Glu Thr Pro  
 20 25 30

**(2) SEQ ID NO:46 的信息:**

**(i) 序列特征:**

**(A) 长度: 32 个氨基酸**

**(B) 类型: 氨基酸**

**(C) 链型: 单链**

**(D) 拓扑结构: 线性**

**(ii) 分子类型: 肽**

**(iii) 假设: 否**

**(iv) 反义: 否**

(v) 片段类型: N 末端

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:46:

Cys	Ser	Asn	Leu	Ser	Thr	Cys	Val	Leu	Ser	Ala	Tyr	Trp	Lys	Asp	Leu
1				5					10					15	
Asn	Asn	Tyr	His	Arg	Tyr	Ser	Gly	Met	Gly	Phe	Gly	Pro	Glu	Thr	Pro
			20					25					30		

(2) SEQ ID NO:47 的信息:

(i) 序列特征:

(A) 长度: 28 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 基因组 DNA

(iii) 假设: 否

(iv) 反义: 否

(v) 片段类型:

(vi) 原始来源:

(xi) 序列描述: SEQ ID NO:47:

AGCTTTCGTT GACGACGACG ATATCTTA

28

(2) SEQ ID NO:48 的信息:

(i) 序列特征:

(A) 长度: 27 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 基因组 DNA

(iii) 假设: 否



(v) 片段类型:

(vi) 原始来源:

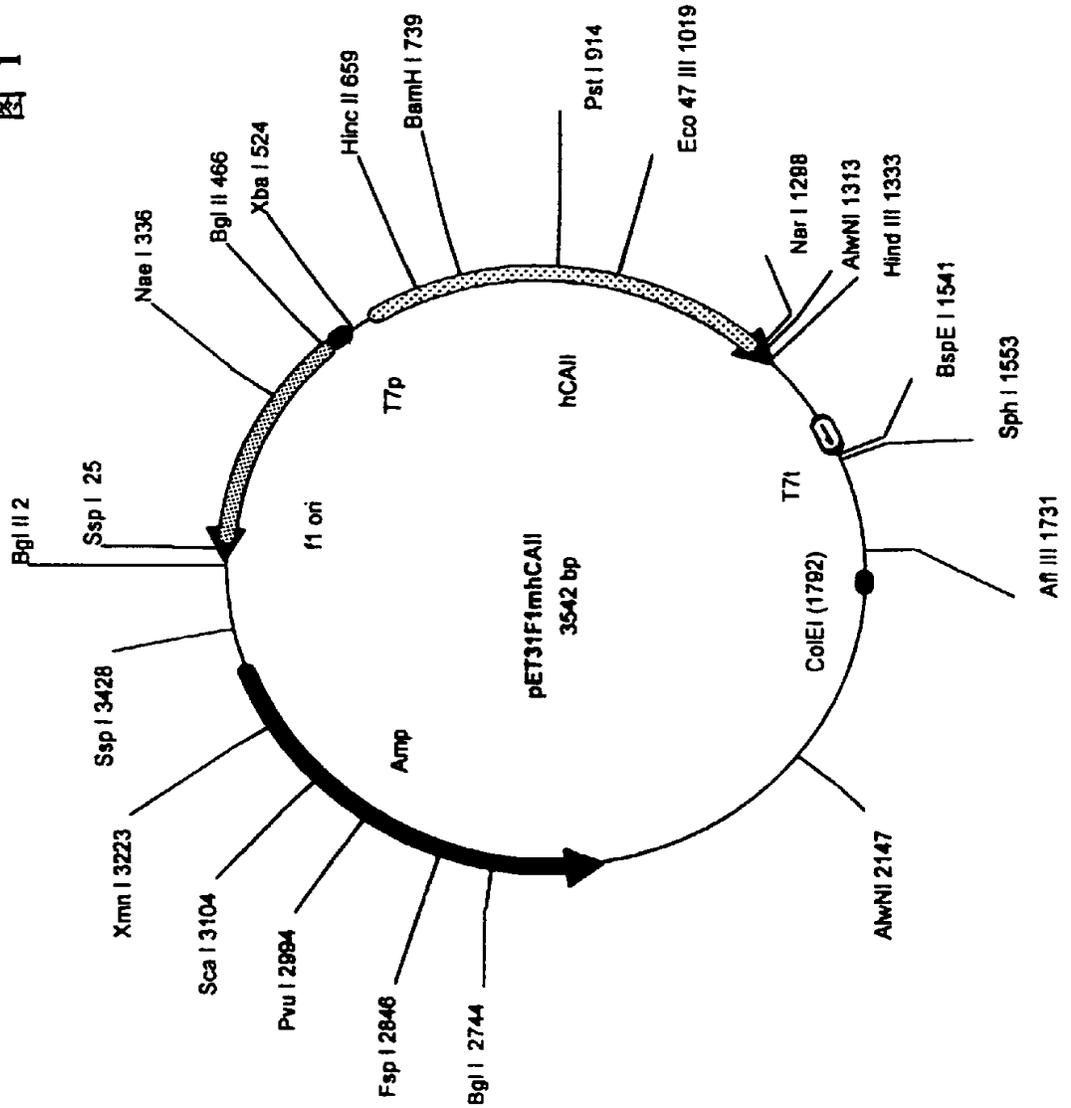
(xi) 序列描述: SEQ ID NO:50:

GGCTCCAAGC TTGTTAACGG TAAACTGTCT CAGGAGCTCC ATAAACTGCA GACTTACCCG  
CGTACTGACG TTGGTGCTGG TACC

60  
84

说明书附图

图 1



<u>HindIII</u>							<u>EcoRV</u>			
A	GCT	TTC	GTT	GAC	GAC	GAC	GAT	ATC	TTA	
		AAG	CAA	CTG	CTG	CTG	CTA	TAG	AAT	CGA
K	A	F	V	D	D	D	D	I	L	

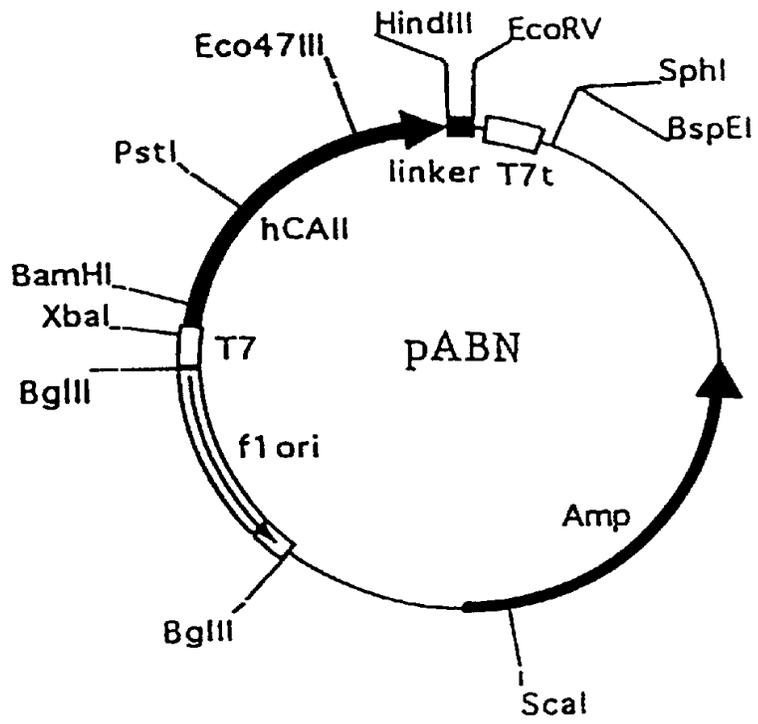


图 2

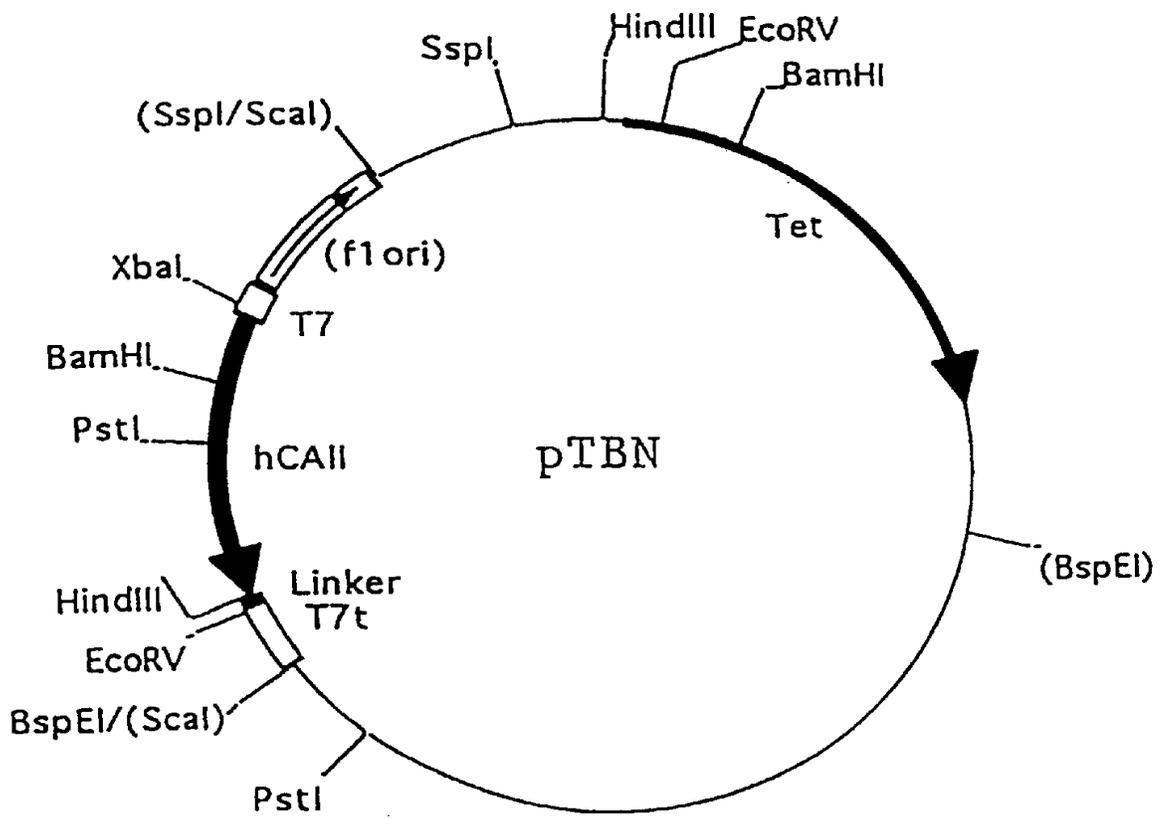


图 3

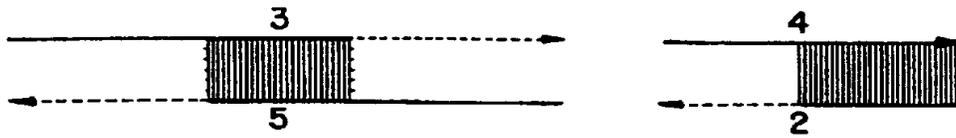


图 4A

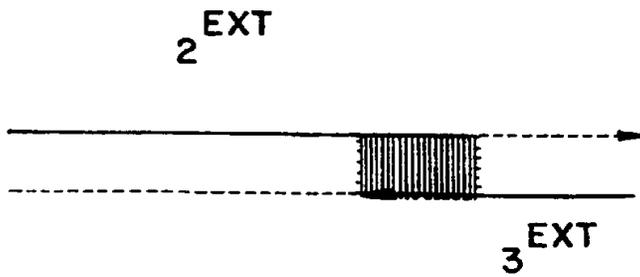


图 4B

5' ATG TCC CAT CAC TGG GGG TAC GGC AAA CAC AAC GGA CCT GAG CAC TGG CAT AAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Met Ser His His Trp Gly Tyr Gly Lys His Asn Gly Pro Glu His Trp His Lys

63 72 81 90 99 108  
 GAC TTC CCC ATT GCC AAG GGA GAG CGC CAG TCC CCT GTT GAC ATC GAC ACT CAT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Asp Phe Pro Ile Ala Lys Gly Glu Arg Gln Ser Pro Val Asp Ile Asp Thr His

117 126 135 144 153 162  
 ACA GCC AAG TAT GAC CCT TCC CTG AAG CCC CTG TCT GTT TCC TAT GAT CAA GCA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Thr Ala Lys Tyr Asp Pro Ser Leu Lys Pro Leu Ser Val Ser Tyr Asp Gln Ala

171 180 189 198 207 216  
 ACT TCC CTG AGG ATC CTC AAC AAT GGT CAT GCT TTC AAC GTG GAG TTT GAT GAC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Thr Ser Leu Arg Ile Leu Asn Asn Gly His Ala Phe Asn Val Glu Phe Asp Asp

225 234 243 252 261 270  
 TCT CAG GAC AAA GCA GTG CTC AAG GGA GGA CCC CTG GAT GGC ACT TAC AGA TTG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Ser Gln Asp Lys Ala Val Leu Lys Gly Gly Pro Leu Asp Gly Thr Tyr Arg Leu

279 288 297 306 315 324  
 ATT CAG TTT CAC TTT CAC TGG GGT TCA CTT GAT GGA CAA GGT TCA GAG CAT ACT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Ile Gln Phe His Phe His Trp Gly Ser Leu Asp Gly Gln Gly Ser Glu His Thr

333 342 351 360 369 378  
 GTG GAT AAA AAG AAA TAT GCT GCA GAA CTT CAC TTG GTT CAC TGG AAC ACC AAA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Val Asp Lys Lys Lys Tyr Ala Ala Glu Leu His Leu Val His Trp Asn Thr Lys

387 396 405 414 423 432  
 TAT GGG GAT TTT GGG AAA GCT GTG CAG CAA CCT GAT GGA CTG GCC GTT CTA GGT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Tyr Gly Asp Phe Gly Lys Ala Val Gln Gln Pro Asp Gly Leu Ala Val Leu Gly

441 450 459 468 477 486  
 ATT TTT TTG AAG GTT GGC AGC GCT AAA CCG GGC CTT CAG AAA GTT GTT GAT GTG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 Ile Phe Leu Lys Val Gly Ser Ala Lys Pro Gly Leu Gln Lys Val Val Asp Val

图 6A

		495		504		513		522		531		540					
CTG	GAT	TCC	ATT	AAA	ACA	AAG	GGC	AAG	AGT	GCT	GAC	TTC	ACT	AAC	TTC	GAT	CCT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Leu	Asp	Ser	Ile	Lys	Thr	Lys	Gly	Lys	Ser	Ala	Asp	Phe	Thr	Asn	Phe	Asp	Pro
		549		558		567		576		585		594					
CGT	GGC	CTC	CTT	CCT	GAA	TCC	TTG	GAT	TAC	TGG	ACC	TAC	CCA	GGC	TCA	CTG	ACC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Arg	Gly	Leu	Leu	Pro	Glu	Ser	Leu	Asp	Tyr	Trp	Thr	Tyr	Pro	Gly	Ser	Leu	Thr
		603		612		621		630		639		648					
ACC	CCT	CCT	CTT	CTG	GAA	TGT	GTG	ACC	TGG	ATT	GTG	CTC	AAG	GAA	CCC	ATC	AGC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Thr	Pro	Pro	Leu	Leu	Glu	Cys	Val	Thr	Trp	Ile	Val	Leu	Lys	Glu	Pro	Ile	Ser
		657		666		675		684		693		702					
GTC	AGC	AGC	GAG	CAG	GTG	TTG	AAA	TTC	CGT	AAA	CTT	AAC	TTC	AAT	GGG	GAG	GGT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Val	Ser	Ser	Glu	Gln	Val	Leu	Lys	Phe	Arg	Lys	Leu	Asn	Phe	Asn	Gly	Glu	Gly
		711		720		729		738		747		756					
GAA	CCC	GAA	GAA	CTG	ATG	GTG	GAC	AAC	TGG	CGC	CCA	GCT	CAG	CCA	CTG	AAG	AAC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Glu	Pro	Glu	Glu	Leu	Met	Val	Asp	Asn	Trp	Arg	Pro	Ala	Gln	Pro	Leu	Lys	Asn
		765		774		783		792		801		810					
AGG	CAA	ATC	AAA	GCT	TTC	GTT	GAC	GAC	GAC	AAC	GGT	AAA	CTG	TCT	CAG	GAG	
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
Arg	Gln	Ile	Lys	Ala	Phe	Val	Asp	Asp	Asp	Asp	Asn	Gly	Lys	Leu	Ser	Gln	Glu
		819		828		837		846		855		864					
CTC	CAT	AAA	CTG	CAG	ACT	TAC	CCG	CGT	ACT	GAC	GTT	GGT	GCT	GGT	ACC	CCG	GCT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Leu	His	Lys	Leu	Gln	Thr	Tyr	Pro	Arg	Thr	Asp	Val	Gly	Ala	Gly	Thr	Pro	Ala

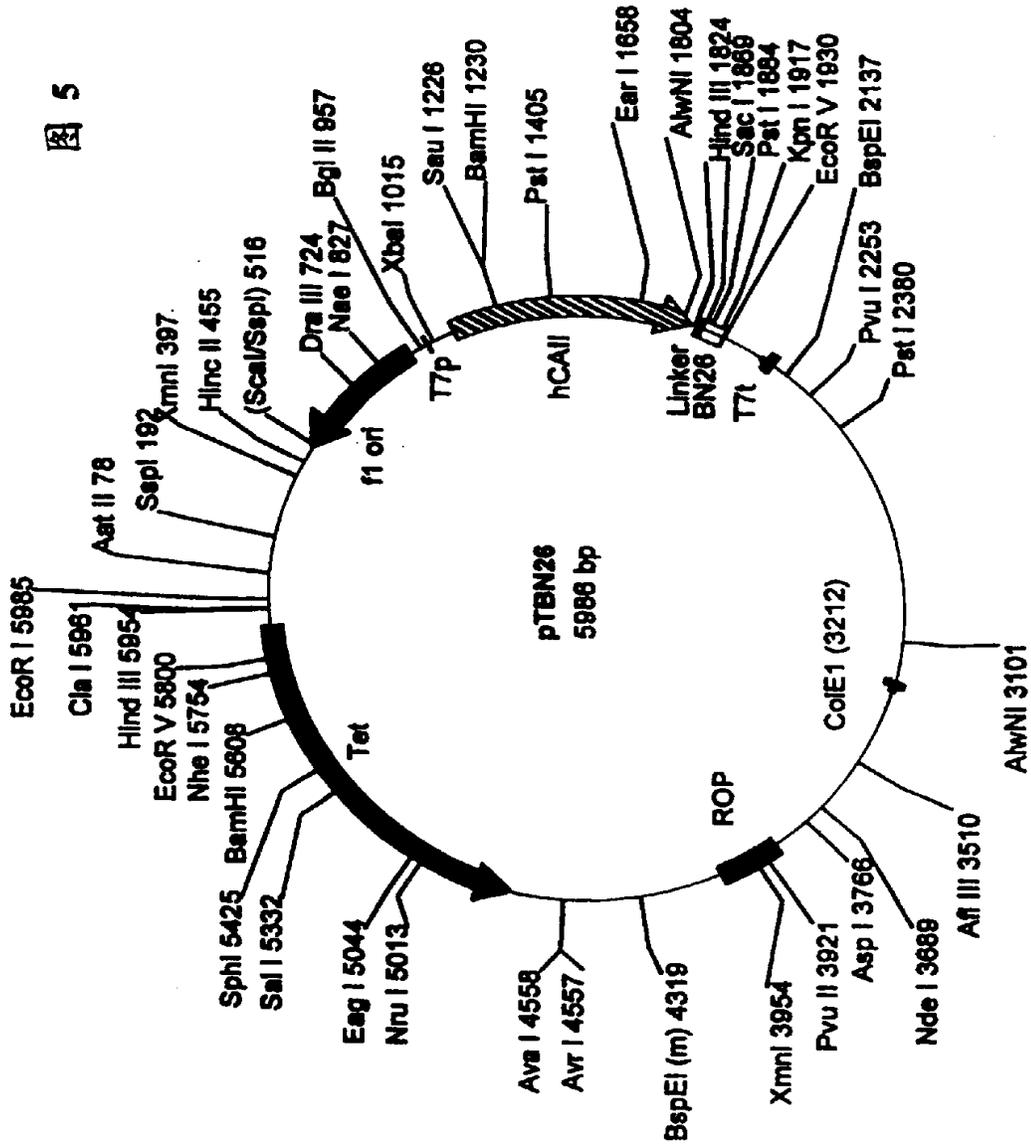
TAA 3'

---

\*\*\*

图 6B

图 5



Gly Lys Leu Ser Gln Glu Leu His  
 GGT AAA CTG TCT CAG GAG CTC CAT

Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr  
 AAA CTG CAG ACT TAC CCG CGT ACT GAC GTT GGT GCT GGT ACC

Pro  
 CCG

图 7

GGA TCC AAG CTT GTT AAC GGT AAA CTG TCT CAG GAG CTC CAT  
 CCT AGG TTC GAA CAA TTG CCA TTT GAC AGA GTC CTC GAG GTA  
 Bam HI HindIII HpaI

AAA CTG CAG ACT TAC CCG CGT ACT GAC GTT GGT GCT GGT ACC  
 TTT GAC GTC TGA ATG GGC GCA TGA CTG CAA CCA CGA CCA TGG

CCG GCT TAA GAT AAT GAA TTC GTC GAC  
 GGC CGA ATT CTA TAG CTT AAG CAG CAG  
 EcoRV EcoRI Sall

图 10

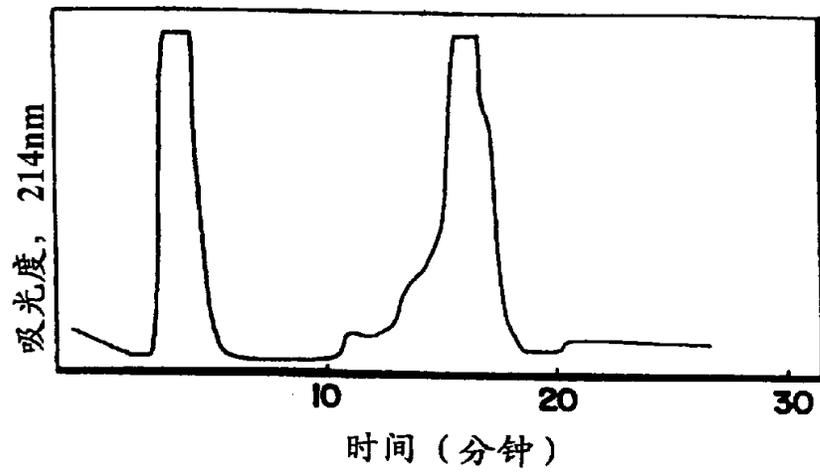


图 8

SEQ

ID

NO:

图 9A

物种

1

5

10

15

37	鳗鱼	Cys Ser Asn Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Glu
38	鲑鱼 I	Cys Ser Asn Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Glu
39	鲑鱼 II	Cys Ser Asn Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Asp
40	鲑鱼 III	Cys Ser Asn Leu Ser Thr Cys Met Leu Gly Lys Leu Ser Gln Asp
41	小鸡	Cys Ala Ser Leu Ser Thr Cys Val Leu Gly Lys Leu Ser Gln Glu
42	人	Cys Gly Asn Leu Ser Thr Cys Met Leu Gly Thr Tyr Thr Gln Asp
43	兔	Cys Gly Asn Leu Ser Thr Cys Met Leu Gly Thr Tyr Thr Gln Asp
44	猪	Cys Ser Asn Leu Ser Thr Cys Val Leu Ser Ala Tyr Trp Arg Asn
45	牛	Cys Ser Asn Leu Ser Thr Cys Val Leu Ser Ala Tyr Trp Lys Asp
46	羊	Cys Ser Asn Leu Ser Thr Cys Val Leu Ser Ala Tyr Trp Lys Asp

图 9B

<u>SEQ</u>	<u>ID</u>	<u>NO:</u>	<u>物种</u>	16	20	25	30
37			鲛鱼	Leu His Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro-NH <sub>2</sub>			
38			鲑鱼 I	Leu His Lys Leu Gln Thr Tyr Pro Arg Thr Asn Thr Gly Ser Gly Thr Pro-NH <sub>2</sub>			
39			鲑鱼 II	Leu His Lys Leu Gln Thr Phe Pro Arg Thr Asn Thr Gly Ala Gly Val Pro-NH <sub>2</sub>			
40			鲑鱼 III	Leu His Lys Leu Gln Thr Phe Pro Arg Thr Asn Thr Gly Ala Gly Val Pro-NH <sub>2</sub>			
41			小鸡	Leu His Lys Leu Gln Thr Tyr Pro Arg Thr Asp Val Gly Ala Gly Thr Pro-NH <sub>2</sub>			
42			人	Phe Asn Lys Phe His Thr Phe Pro Gln Thr Ala Ile Gly Val Gly Ala Pro-NH <sub>2</sub>			
43			兔	Leu Asn Lys Phe His Thr Phe Pro Gln Thr Asp Ile Gly Val Ala Pro-NH <sub>2</sub>			
44			猪	Leu Asn Asn Phe His Arg Phe Ser Gly Met Gly Phe Gly Pro Glu Thr Pro-NH <sub>2</sub>			
45			牛	Leu Asn Asn Tyr His Arg Phe Ser Gly Met Gly Phe Gly Pro Glu Thr Pro-NH <sub>2</sub>			
46			羊	Leu Asn Asn Tyr His Arg Tyr Ser Gly Met Gly Phe Gly Pro Glu Thr Pro-NH <sub>2</sub>			

**2** EXT

5' GAC GAC GAA TTC GAT ATC TTA AGC CCG GGT ACC AGC ACC  
AAC GTC AG

图 11

**3** EXT

5' GGA TCC AAG CTT GTT AAC GGT AAA CTG TCT CAG GAG CTC CAT  
AAA CTG CAG ACT TAC CCG CGT ACT GAC GTT GGT GCT GGT  
ACC

图 12