



(12)发明专利



(10)授权公告号 CN 102883734 B

(45)授权公告日 2018.01.02

(21)申请号 201180022849.3

(22)申请日 2011.05.20

(65)同一申请的已公布的文献号
申请公布号 CN 102883734 A

(43)申请公布日 2013.01.16

(30)优先权数据
10163564.7 2010.05.21 EP
61/428,016 2010.12.29 US

(85)PCT国际申请进入国家阶段日
2012.11.06

(86)PCT国际申请的申请数据
PCT/EP2011/058307 2011.05.20

(87)PCT国际申请的公布数据
W02011/144756 EN 2011.11.24

(73)专利权人 XL-蛋白有限责任公司

地址 德国弗赖辛

(72)发明人 阿恩·斯科拉 尤里·宾德尔
马丁·斯拉珀斯其

(74)专利代理机构 北京英赛嘉华知识产权代理
有限责任公司 11204

代理人 王达佐 洪欣

(51)Int.Cl.
A61K 38/00(2006.01)
C07K 7/00(2006.01)

(56)对比文件
WO 0178503 A2,2001.10.25,
审查员 周英

权利要求书3页 说明书70页
序列表53页 附图25页

(54)发明名称

生物合成的脯氨酸/丙氨酸无规卷曲多肽及其用途

(57)摘要

本发明涉及生物合成的无规卷曲多肽(random coil polypeptide)或生物合成的无规卷曲多肽区段或生物合成的缀合物,其中所述生物合成的无规卷曲多肽、所述生物合成的无规卷曲多肽区段或所述生物合成的缀合物包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少约50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。所述至少约50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基可以是异源多肽或异源多肽构建体的组成部分。还描述了这些生物合成的无规卷曲多肽或多肽区段或所述缀合物的用途和使用方法。特别地,所述用途可以包括医药用途,诊断用途或食品工业用途以及其它工业应用中的用途,例如造纸工业、采油等。具体而言,提供了药物缀合物,其包含(i)生物合成的无规卷曲多肽或多肽区段,所述多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基

酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成,和(ii)药物,所述药物选自(a)包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽,和(b)小分子药物。此外,公开了编码所述生物合成的无规卷曲多肽或多肽区段和/或所述生物活性异源蛋白的核酸分子,以及包含所述核酸分子的载体和细胞。此外,本发明提供包含本发明化合物的组合物,以及本发明的无规卷曲多肽或多肽区段、生物活性蛋白、药物缀合物或核酸分子、载体和细胞的具体用途。还提供了产生和/或获得本发明生物合成的无规卷曲多肽或多肽区段的方法,以及产生和/或获得本发明的生物活性异源蛋白和/或多肽构建体或缀合物(例如药物缀合物)的方法。

1. 药物缀合物, 包含

(i) 生物合成的无规卷曲多肽或多肽区段, 所述多肽或多肽区段由仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列组成, 其中所述氨基酸序列由至少50个脯氨酸(Pro) 和丙氨酸(Ala) 氨基酸残基组成, 其中所述无规卷曲多肽或多肽区段的构成部分由选自以下的氨基酸序列组成:

AAPAAPAAPAAPAAPAAP (SEQ ID NO:1), 和

AAAPAAAPAAAPAAAPAAAP (SEQ ID NO:3), 以及

(ii) 药物, 所述药物选自 (a) 包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽, 和 (b) 小分子药物。

2. 如权利要求1所述的药物缀合物, 其中所述无规卷曲多肽或多肽区段包含由50至3000个氨基酸残基组成的氨基酸序列。

3. 如权利要求1或2所述的药物缀合物, 其中所述脯氨酸残基构成大于10%且小于75%的所述氨基酸序列。

4. 如权利要求1所述的药物缀合物, 其中所述无规卷曲多肽或多肽区段包含多个氨基酸重复, 其中所述重复由脯氨酸和丙氨酸残基组成, 并且其中不多于6个连续氨基酸残基是相同的。

5. 如权利要求1所述的药物缀合物, 其中所述具有生物活性的多肽、所述包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽选自结合蛋白、抗体片段、细胞因子、生长因子、激素或酶。

6. 如权利要求5所述的药物缀合物, 其中所述具有生物活性的多肽是结合蛋白, 并且其中所述结合分子选自抗体、Fab片段、F(ab')₂片段、CDR衍生的肽模拟物、单链可变片段(scFv)、结构域抗体、凝集素、免疫球蛋白结构域、纤连蛋白结构域、蛋白A结构域、SH3结构域、锚蛋白重复结构域和脂质运载蛋白。

7. 如权利要求1所述的药物缀合物, 其中所述生物活性蛋白选自粒细胞集落刺激因子、人生长激素、 α -干扰素、 β -干扰素、 γ -干扰素、肿瘤坏死因子、红细胞生成素、凝血因子VIII、gp120/gp160、可溶性肿瘤坏死因子I和II受体、瑞替普酶、艾塞那肽-4、阿那白滞素、白介素-2、中性粒细胞明胶酶相关脂质运载蛋白、卵泡刺激素、葡糖脑苷脂酶、胸腺素 α 1、胰高血糖素、生长激素抑制素、腺苷脱氨酶、白介素11、凝血因子VIIa、凝血因子IX、hematide、 λ -干扰素、瘦素、白介素-22受体 α 亚基(IL-22ra)、白介素-22、透明质酸酶、成纤维细胞生长因子18、成纤维细胞生长因子21、胰高血糖素样肽1、骨保护素、IL-18结合蛋白、生长激素释放因子、可溶性TACI受体、血小板反应蛋白-1、可溶性VEGF受体Flt-1和IL-4突变蛋白。

8. 如权利要求1所述的药物缀合物, 其中所述生物合成的无规卷曲多肽或多肽介导所述药物缀合物的增加的体内和/或体外稳定性。

9. 如权利要求8所述的药物缀合物, 其中所述增加的体内稳定性是与缺少所述无规卷曲多肽或多肽区段的对照多肽或对照缀合物的稳定性相比, 所述药物缀合物的延长的血浆半衰期, 所述药物缀合物包含所述生物合成的无规卷曲多肽或多肽区段, 所述多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列, 其中所述氨基酸序列由至少50个脯氨酸(Pro) 和丙氨酸(Ala) 氨基酸残基组成。

10. 如权利要求1所述的药物缀合物, 其中所述小分子选自地高辛配基、荧光素、多柔比

星、卡里奇霉素、喜树碱、烟曲霉素、地塞米松、格尔德霉素、紫杉醇、多西他赛、伊立替康、环孢霉素、丁丙诺啡、纳曲酮、纳洛酮、长春地辛、万古霉素、利培酮、阿立哌唑、帕洛诺司琼、格拉司琼、阿糖胞苷NX1838、亮丙瑞林、戈舍瑞林、布舍瑞林、奥曲肽、替度鲁肽、西仑吉肽、阿巴瑞克、恩夫韦地、葛瑞林、 $\alpha 4$ 整合素抑制剂、反义核酸、小干扰RNA、微RNA、类固醇、DNA或RNA适体和肽和/或肽模拟物。

11. 组合物, 包含权利要求1-10中任一项所述的药物缀合物。

12. 如权利要求11所述组合物, 其是药物组合物或诊断组合物, 所述药物组合物任选地还包含药学可接受的载体, 所述诊断组合物任选地还包含诊断组合物中可接受的载体。

13. 编码权利要求1-10中任一项所述的药物缀合物中所包含的无规卷曲多肽或多肽区段的核酸分子, 或编码蛋白缀合物的核酸分子, 所述蛋白缀合物包含权利要求5-7中任一项所述的生物活性蛋白并包含生物合成的无规卷曲多肽或多肽区段, 所述多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列, 其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。

14. 编码权利要求1-9中任一项所述的药物缀合物的核酸分子, 所述核酸分子包含

(i) 编码翻译的氨基酸和/或前导序列的核酸序列;

(ii) 编码生物合成的无规卷曲多肽或多肽区段的核酸序列, 所述多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列, 其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成;

(iii) 编码包含具有或介导生物和/或治疗活性的氨基酸序列或是具有或介导生物和/或治疗活性的氨基酸序列的生物活性蛋白或多肽的核酸序列; 和

(iv) 代表翻译终止密码子或是翻译终止密码子的核酸序列。

15. 如权利要求14所述的核酸分子, 其中(ii)和(iii)中所述的所述核酸分子部分或区段在所述编码药物缀合物的核酸分子上的位置互换。

16. 如权利要求14或15所述的核酸分子, 任选地在(i)和(ii)所述的部分或区段之间和/或(ii)和(iii)所述的部分或区段之间包含蛋白酶和/或化学切割位点和/或识别位点。

17. 载体, 包含权利要求13-16中任一项所述的核酸分子。

18. 宿主细胞, 包含权利要求13-16中任一项所述的核酸分子或权利要求17所述的载体。

19. 如权利要求18所述的宿主细胞, 其中所述宿主细胞是真核宿主细胞。

20. 如权利要求19所述的宿主细胞, 其中所述真核宿主细胞是真菌或动物细胞。

21. 制备权利要求1-10中任一项所述的药物缀合物中包含的无规卷曲多肽或无规卷曲多肽区段、制备包含所述无规卷曲多肽或所述无规卷曲多肽区段的生物活性蛋白或药物或食品缀合物和/或制备权利要求1-7中任一项所述的包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列并且还包含所述无规卷曲多肽或无规卷曲多肽区段的多肽的方法,

所述方法包括培养权利要求18-20中任一项所述的细胞, 并从培养物或所述细胞分离所述无规卷曲多肽或生物活性蛋白和/或所述生物活性蛋白或所述多肽。

22. 权利要求1、2和4-10中任一项所述的药物缀合物、权利要求11或12所述的组合物、权利要求13-15中任一项所述的核酸分子、权利要求17所述的载体、权利要求18-20中任一

项所述的细胞或通过权利要求21所述的方法制备的包含所述无规卷曲多肽或无规卷曲多肽区段的生物活性蛋白或多肽,其用于治疗激素缺陷相关病症、自身免疫疾病、增殖性病症、贫血症、新生血管性疾病、感染性/炎症性疾病、变应性病症、血栓形成、心肌梗塞、视网膜变性、糖尿病、不孕症、高球氏症、慢性乙型肝炎、丙型肝炎、低血糖症、肢端肥大症、腺苷脱氨酶缺陷、血小板减少症、血友病、贫血症、肥胖症、阿尔茨海默病、脂肪代谢障碍、银屑病、骨关节炎、血脂异常、风湿性关节炎、系统性红斑狼疮、多发性硬化、哮喘、骨质疏松症和再灌注损伤或其它肾病。

23. 权利要求1、2和4-10中任一项所述的药物缀合物、权利要求11或12所述的组合物、权利要求13-15中任一项所述的核酸分子、权利要求17所述的载体、权利要求18-20中任一项所述的细胞或通过权利要求21所述的方法制备的包含所述无规卷曲多肽或无规卷曲多肽区段的生物活性蛋白或多肽,其用于治疗癌症。

24. 权利要求1、2和4-10中任一项所述的药物缀合物、权利要求11或12所述的组合物、权利要求13-15中任一项所述的核酸分子、权利要求17所述的载体、权利要求18-20中任一项所述的细胞或通过权利要求21所述的方法制备的包含所述无规卷曲多肽或无规卷曲多肽区段的生物活性蛋白或多肽,其用于治疗转移性黑素瘤。

25. 权利要求1、2和4-10中任一项所述的药物缀合物、权利要求11或12所述的组合物、权利要求13-15中任一项所述的核酸分子、权利要求17所述的载体、权利要求18-20中任一项所述的细胞,其用作药物,所述药物具有所述无规卷曲多肽或多肽区段、生物活性蛋白或药物缀合物的增加的体内和/或体外稳定性。

生物合成的脯氨酸/丙氨酸无规卷曲多肽及其用途

[0001] 本发明涉及生物合成的无规卷曲多肽(random coil polypeptide)或生物合成的无规卷曲多肽区段或缀合物,所述生物合成的无规卷曲多肽或生物合成的无规卷曲多肽区段或缀合物包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少约50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。所述至少约50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基可以是异源多肽或异源多肽构建体的组成部分。还描述了这些生物合成的无规卷曲多肽、所述多肽区段或所述缀合物的用途和使用方法。所述用途可特别包括医药用途、诊断用途或食品工业用途以及在其它工业应用中的用途,例如造纸工业、采油等。本发明还涉及本文提供的生物合成的无规卷曲多肽或生物合成的无规卷曲多肽区段或缀合物的具体用途,所述生物合成的无规卷曲多肽或生物合成的无规卷曲多肽区段或缀合物包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。本文提供的生物合成的无规卷曲多肽或生物合成的无规卷曲多肽区段的氨基酸序列由至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350或至少约400个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。所述至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350或至少约400个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基优选为(a)异源多肽或异源多肽构建体的组成部分或优选为(b)缀合物的组成部分,所述缀合物例如药物缀合物、具有食品或美容用品成分或添加剂的缀合物、具有生物活性化合物的缀合物或具有光谱学活性化合物的缀合物。具体而言,本文提供异源蛋白,其中这些蛋白包含至少两个结构域,其中所述至少两个结构域中的第一结构域包含具有和/或介导诸如生物活性的活性的氨基酸序列,所述至少两个结构域中的第二结构域包含本发明的生物合成的无规卷曲脯氨酸/丙氨酸多肽或脯氨酸/丙氨酸多肽区段。本发明特别涉及药物缀合物,其包含(i)生物合成的无规卷曲多肽或多肽区段,所述多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成,和(ii)药物,所述药物选自(a)包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽,和(b)小分子药物。本发明的另一主题是药物缀合物,其包含本文提供的生物合成的无规卷曲脯氨酸/丙氨酸多肽或脯氨酸/丙氨酸多肽区段,以及药学或医学有用的分子,例如小分子、肽或生物大分子(例如蛋白、核酸、碳水化合物、脂质囊泡)等,所述药学或医学有用的分子与所述生物合成的无规卷曲脯氨酸/丙氨酸多肽或脯氨酸/丙氨酸多肽区段连接和/或偶联。此外,公开了编码生物合成的无规卷曲多肽或多肽区段和/或生物活性异源蛋白的核酸分子以及包含所述核酸分子的载体和细胞。此外,公开了产生本文所述的本发明生物合成的无规卷曲多肽或多肽区段和相应药物或食品缀合物(即,包含本文所述的生物合成的无规卷曲多肽或多肽区段和食品成分或食品添加剂的缀合物)的方法。还公开了相应的缀合物(作为一种组成部分包含本文公开的生物合成的无规卷曲多肽或多肽区段),其特别包含美容用品成分或添加剂或生物学或光谱学活性化合物。此外,本发明提供包含本发明的化合物的组合物(即,本文公开的无规卷曲多肽或无规卷曲多肽区段和编码其的核酸分子,所述无规卷曲多肽或无规卷曲多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列),以及本发明的所述无规卷

曲多肽或多肽区段、包含所述无规卷曲多肽或无规卷曲多肽区段的生物活性蛋白、药物缀合物、食品缀合物或核酸分子、载体和细胞的具体用途。还提供了产生和/或获得本发明生物合成的无规卷曲多肽或多肽区段的方法,以及产生和/或获得本发明生物活性异源蛋白和/或多肽构建体或药物缀合物的方法。此外,本文提供了生物合成的无规卷曲多肽或多肽区段(或包含所述无规卷曲多肽或多肽区段的分子和缀合物)的医学用途、药学用途以及诊断用途,所述无规卷曲多肽或多肽区段包含本文所述的仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。所述医学用途或药学用途可以包括所述生物合成的无规卷曲多肽或多肽区段作为血浆扩容剂等用途。然而,本文提供的方式和方法不限于药学、医学和生物学用途,并且还可以用于其它工业领域,例如造纸工业、采油等。

[0002] 由肾过滤导致的从血液循环快速清除是小分子(包括小蛋白和肽)的通常性质。然而,通过使表观分子维度增大至超过肾小球的孔径,可以将治疗蛋白的血浆半衰期延长到医学上有用的数天的范围。实现该效果的一个策略是将生物制剂与合成聚合物聚乙二醇(PEG)进行化学偶联。这种策略已经获得了数种获批准的药物,例如PEG-干扰素 $\alpha 2a$ (Pegasys®)、PEG-G-CSF(Neulasta®)以及近期的PEG化的 α TNF-Fab片段(Cimzia®)。尽管如此,“PEG化”技术具有几个缺点:临床级PEG衍生物很昂贵,并且它们与重组蛋白的共价偶联需要额外的下游加工和纯化步骤,从而会降低产率并提高成本。此外,PEG不是生物可降解的,这能带来副作用,例如连续治疗后的肾上皮空泡形成;参见,例如,Gaberc-Porekar (2008) Curr Opin Drug Discov Devel 11:242-50;Knop (2010) Angew Chem Int Ed Engl 49:6288-308或Armstrong in:Veronese (Ed.), “PEGylated Protein drugs: Basic Science and Clinical Applications”; Birkhäuser Verlag, Basel 2009。

[0003] 为了克服PEG技术的一些缺点,本领域中已提供了某些重组多肽模拟物,其中的一些是基于天然存在的氨基酸序列或合成的氨基酸段(stretch)。

[0004] 大部分天然氨基酸序列在生理溶液中的表现不像理想的随机链,因为它们倾向于采取折叠的构象(二级结构),或者如果它们不折叠,它们通常是不溶的并形成聚集体。事实上,研究多肽的随机链行为的大部分经典实验是在变性条件下进行的,即,在化学变性剂例如尿素或盐酸胍的存在下(参见,例如,Cantor (1980) Biophysical Chemistry. W.H. Freeman and Company, New York)。因此,这些技术通常依赖于,即使与折叠的治疗蛋白结构域基因融合的情况下,也能在生理缓冲液条件和温度下抵抗折叠、聚集以及非特异性吸附并从而能提供稳定的随机链的特殊氨基酸序列。在这些情况下,这类重组PEG模拟物会使尺寸的增加大大超过仅基于它们的分子量所能通常预期的尺寸,最终阻碍肾过滤并有效地延长重要因子所连接的生物制剂的血浆半衰期。

[0005] 然而,这些技术中的很多具有其它注意事项和缺点。

[0006] 例如,已经测试了天然存在的重复性氨基酸序列在医学和药学方法中的可用性。这些方法中的一种涉及克氏锥虫(*Trypanosoma cruzi*)的反式唾液酸酶。其含有680个氨基酸残基的催化结构域,催化结构域之后是C末端重复性结构域,被称为“流急性期抗原(shed acute phase antigen)”(SAPA),其包含数量可变的12聚体氨基酸重复。含有13个亲水性和(在生理pH下)带负电的对应氨基酸重复(具有天然序列DSSAHSTPSTPA)的反式唾液酸酶的小鼠药代动力学(PK)研究显示出血浆半衰期比缺失了C末端重复性序列的重组酶长5倍(Buscaglia (1999) Blood 93:2025-32)。在相同的反式唾液酸酶融合之后,即,它的76kDa催

化结构域与在克氏锥虫蛋白抗原13中发现的13个带电的序列EPKSA氨基酸重复,也观察到相似的半衰期延长效应。在来自SAPA的重复和来自抗原13的重复与来自日本血吸虫(*Schistosoma japonicum*)的异源蛋白谷胱甘肽S转移酶(GST)的两个C末端基因融合之后,两种重复都能将该同二聚化酶的血浆半衰期延长7-8倍(参见Buscaglia,同上)。然而,尽管这些来自人病原体的天然存在的重复性氨基酸序列原则上似乎很适合于优化治疗蛋白的药代动力学,但是发现它们的免疫原性很高(参见Affranchino(1989),*MolBiochem Parasitol* 34:221-8或Buscaglia(1998),*J Infect Dis* 1998;177:431-6)。

[0007] 另一方法涉及使用明胶。明胶是水解并变性了的动物胶原蛋白,其含有长段的Gly-Xaa-Yaa重复,其中Xaa和Yaa分别主要由脯氨酸和4-羟基脯氨酸构成。明胶的琥珀酰化主要是通过天然散在的赖氨酸侧链的 ϵ -氨基,这能增加该生物聚合物的亲水性并降低其等电点(pI)。据推测,经修饰的侧链的带负电的羧化物基团之间的分子内静电排斥能使分子展开为多少伸展的构象。所实现的增大的体积使琥珀酰化的明胶能作为用于人体血浆扩容剂的大分子,特别是在市场上以**Volplex®**(Beacon Pharmaceuticals Ltd)或**Gelofusine®**(B.Braun Melsungen AG)销售。此外,粒细胞集落刺激因子(G-CSF)与人工明胶样多肽的基因融合实现了半衰期延长效应(Huang(2010)*Eur J Pharm Biopharm* 74:435-41)。为此,天然明胶的所有疏水性侧链被亲水性残基交换,这产生了包含不同顺序的氨基酸G、P、E、Q、N、S和K的116个氨基酸的明胶样蛋白(GLK)。G-CSF在其N末端与4拷贝的该GLK序列融合,并在巴斯德毕赤酵母(*Pichia pastoris*)中分泌。巴斯德毕赤酵母看起来可作为GLK融合蛋白的良好制备生物体;然而,其它生物体中是否也能产生GLK还是未知的,因为已知的是重组明胶片段在大肠杆菌中仅低量表达,如Olsen(2003),*Adv Drug Deliv Rev* 55:1547-67所示。

[0008] 弹性蛋白是很多组织中细胞外基质的组分。其由可溶性前体弹性蛋白原形成,弹性蛋白原由亲水性的富含Lys/Ala的结构域和具有重复性序列的疏水性弹性体结构域组成。亲水性结构域中赖氨酸侧链的酶学交联实现不可溶性弹性蛋白的形成。弹性蛋白样多肽(ELP)是人工设计的、源于弹性蛋白原的疏水性结构域的重复性氨基酸序列。ELP的最常见的重复序列基序是V-P-G-X-G,其中“X”可以是除Pro之外的任何氨基酸(MacEwan(2010)*Biopolymers* 94:60-77;Kim(2010)*Adv Drug Deliv Rev* 62:1468-78)。合适的ELP可以与治疗蛋白和融合,并在大肠杆菌中产生。因此,ELP在注射后形成凝胶样贮库的能力能显著延长所连接的生物制剂的体内半衰期,但是是通过的机制不同于其它无组织的多肽。然而,ELP连接能阻碍融合伴侣的生物活性,正如在IL-1诱导的淋巴细胞增殖生物学检测中对于白介素-1受体拮抗剂所显示的(Shamji(2007)*Arthritis Rheum*.11:3650-3661)。此外,ELP易于被诸如胶原蛋白酶的内源性蛋白酶降解。同样,聚集的蛋白通常对免疫原性更敏感。

[0009] 其它方法涉及使用聚阴离子聚合物。例如,聚谷氨酸(PG)已与用于癌症治疗的可溶性较差的细胞毒性小分子药物化学偶联。相应的产品是Opaxio™,其是目前处于临床III期研究中的紫杉醇药物缀合物。与未修饰的化合物相比,紫杉醇PG缀合物的半衰期延长3-14倍(Singer(2005)*JControl Release* 109:120-6)。其它融合蛋白,例如在N末端与一段175个连续Glu残基融合的G-CSF,或在C末端携带84个残基PG尾部的IFN- α 2,在大肠杆菌细胞质中以可溶性状态产生(参见W02002/077036)。对于有效翻译,N末端融合需要前导肽,其随后由烟草蚀纹病毒(TEV)蛋白酶切割来去除。G-CSF和INF- α 2的聚谷氨酸融合物在细胞培

养测定中表现出生物活性。然而,至今这些PG融合物的药代动力学数据还没有报道。另外,由于人为静电吸引或排斥效应,PG融合物的大量负电荷是相对于生物分子相互作用(例如,靶受体或可溶性因子的结合)而言的普遍缺点。

[0010] WO 2006/081249描述了具有约2-500个重复单元的多肽序列,所述重复单元为3-6个氨基酸,其中G、N或Q代表主要组成部分,而次要组成部分可以是A、S、T、D或E。该氨基酸组成允许糖基化序列子Asn-Xaa-Ser/Thr(其中Xaa是除Pro之外的任何氨基酸)的整合,用于真核表达体系中Asn侧链的N连接糖基化。所得融合蛋白增加的大分子尺寸,包括大的溶剂化碳水化合物结构的翻译后修饰在内,能延长基因缀合蛋白的药代动力学。这些寡糖连接物(“糖工程”)通常能减小对蛋白水解的易感性并增加流体动力学体积(Sinclair(2005) J Pharm Sci 94:1626-35)。缺点是糖基化生物大分子的内在分子异质性,这使得在生物技术生产和质量控制过程中需要付出更多劳动。

[0011] WO 2010/091122(以及WO 2007/103515)和Schellenberger(2009) Nat Biotechnol 27:1186-90公开了无组织的非重复性氨基酸聚合物,所述聚合物包括和包含残基P、E、S、T、A和G。这组氨基酸显示的组成不是不同于上文进一步描述的PSTAD重复,这组氨基酸经过序列的系统性筛选,从而产生具有大分子尺寸的溶剂化多肽,由于避免能引起聚集和可以导致HLA/MHC-II介导的免疫应答的疏水性侧链(特别是F、I、L、M、V和W),所述溶剂化多肽适合于生物药物研发。另外,排除了可能交联的Cys残基、可以与带负电的细胞膜相互作用的阳离子氨基酸K、R和H、以及可能易于水解的N和Q的酰胺侧链(参见Schellenberger(2009)同上)。针对大肠杆菌中的可溶性表达水平,筛选编码包含PESTAG残基组的非重复性序列(其与绿色荧光蛋白(GFP)融合)的合成基因文库,以及进一步研究所得到的亚组的基因稳定性、蛋白可溶性、热稳定性、聚集倾向和杂质属性。最终,进一步测试了含有216个Ser残基(25.0摩尔%)、72个Ala残基(8.3摩尔%)和Pro、Thr、Glu和Gly中每种144个氨基酸(16.7摩尔%)的864个氨基酸的序列与GLP-1受体激动剂艾塞那肽-4(E-XTEN)和几种其它生物制剂的融合。在大肠杆菌细胞质中产生可溶状态的融合蛋白,并将其分离,所述融合蛋白通常携带纤维素结合结构域,该纤维素结合结构域随后被切掉。通过圆二色(CD)光谱对E-XTEN的研究显示二级结构的缺乏,同时在尺寸排阻色(SEC)过程中,融合蛋白表明驻留明显少于预期的84kDa蛋白,从而证实增加的流体动力学体积(Schellenberg(2009)同上)。分布在XTEN序列之间的携带高的净负电荷的氨基酸之间的静电排斥,可以有利于PESTAG多肽的无序结构和流体动力学半径的相关增加(参见WO 2010/091122)。然而,另一项研究即Geething(2010) PLoS One 2010;5:e10175表明,XTEN降低其治疗融合伴侣的效力。在细胞培养测定中,胰高血糖素XTEN融合物仅表现出未修饰的肽的生物活性的15%。人生长激素(hGH)的XTEN融合显示出受体亲和力的更大丢失(EC₅₀增加17倍);参见WO 2010/144502。

[0012] 此外,甘氨酸作为最小且结构最简单的氨基酸,在理论层面上,其已被认为是在构象上柔性最好的氨基酸;参见,例如Schulz GE, Schirmer RH. Principles of Protein Structure. Springer, New York 1979。此外,计算机模拟已经表明,Gly聚合物缺少二级结构,并且在溶液中易于形成无规卷曲;参见Shental-Bechor(2005) Biophys J 88:2391-402。从化学的观点来看,聚甘氨酸是线性无支链的聚酰胺,其与聚醚PEG在某些方面表现出一定相似性,例如它们本质上都是一维的大分子,沿着链具有很多转动自由度,由被氢键键

合且高度溶剂化的极性基团规律性中断的重复短碳氢化合物单元组成。因此,聚甘氨酸应当构成最简单的基因可编码的PEG模拟物,其预期能延长治疗蛋白的血浆半衰期。分别以“同氨基酸聚合物(HAP)”的形式或富含甘氨酸的序列(GRS)采用该观点;参见,Schlapschy (2007) *Protein Eng Des Sel* 20:273-84;WO 2007/103515。然而,早就已知的是,Gly的化学合成的纯聚合物在水中表现出差的可溶性;特别可参见Bamford CH et al. *Synthetic Polypeptides-Preparation, Structure, and Properties*. Academic Press, New York 1956。因此,进行了不同的尝试来增加亲水性,其中通过引入氢键键合的丝氨酸醇侧链(WO 2007/103515以及Schlapschy (2007) 同上)或者另外引入带负电的谷氨酸残基(WO2007/103515)。已注意到,为了以柔性方式连接融合蛋白中的结构域,具有(Gly₄Ser)_n组成的肽间隔物在本领域中已有描述。在分析SEC中,对于这些融合蛋白检测到显著增加的流体动力学体积。对于200个残基的HAP形式,与未融合的Fab片段相比,表观尺寸增加为120%,而真实质量仅增大29%,因此表现出流体动力学体积增加的效果,这是由于聚甘氨酸标签的溶剂化无规卷曲结构所致。此外,CD差光谱对于HAP部分的无序二级结构是特征性的。最后,携带200个残基HAP的Fab片段在鼠中的最终血浆半衰期延长约3倍。尽管这仅是中等程度的延长,但该效果可以适合于某些(专门)诊断应用,例如体内成像;参见Schlapschy (2007);同上。不幸的是,具有较长的(Gly₄Ser)_n重复序列的融合蛋白的制备似乎不太可行,这是由于形成聚集体的倾向增加,从而造成对有一定纯度的甘氨酸聚合物作为PEG模拟物的应用天然限制。

[0013] WO 2008/155134公开了具有Pro、Ala和Ser(即,PAS)残基的合适混合的序列导致它们独特二级结构偏好的共同取消,并因此导致稳定无序的多肽。然而,WO 2008/155134还记载了,具有仅由丝氨酸和丙氨酸(SA)残基组成的结构域(即,结构域仅包含两种氨基酸)的融合蛋白不会形成无规卷曲,而是会形成 β 折叠结构。

[0014] 多肽的化学合成是公知的,并且在本领域中已有描述。Izuka公开了含有脯氨酸的多肽的化学合成(参见Izuka (1993), *Bull. Chem. Soc. Jpn* 66, 1269-1272)。这些共聚多肽(copolyptide)含有脯氨酸分别和甘氨酸、L-丙氨酸、L- α -氨基丁酸(Abu)、L-正缬氨酸(Nva)或L-亮氨酸的随机序列,并且通过化学共聚作用合成。Izuka公开了,这些共聚多肽通常具有限定的胶原蛋白样构象。此外,该出版物还描述了脯氨酸和丙氨酸(或脯氨酸和L- α -氨基丁酸)的共聚多肽在水中是部分可溶的,而其它共聚多肽完全不可溶。Izuka推测,脯氨酸/丙氨酸共聚多肽可以具有局部无序构象。Izuka强调,具有随机脯氨酸/丙氨酸序列的化学合成多肽主要出现在胶原蛋白样构象中,即,在有组织的构象中。

[0015] 因此,本发明的基本技术问题是提供具有真实无规卷曲构象的大的多肽。通过提供权利要求中描述的和本文提供的实施方案能解决该技术问题。

[0016] 因此,本发明涉及提供和应用生物合成的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段包含由至少约50、特别是至少约100、特别是至少约150、特别是至少约200、特别是至少约250、特别是至少约300、特别是至少约350、特别是至少约400个脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。因此,本发明涉及提供生物合成的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段包含至少50个氨基酸残基的氨基酸序列,所述氨基酸序列仅由脯氨酸和丙氨酸氨基酸残基组成且包含至少一个脯氨酸和至少一个丙氨酸。本发明还提供了药物缀合物,其包含(i)生物合成的无规卷曲多肽或多肽区段,所述无规卷

曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成,和(ii)药物,所述药物选自(a)包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽和(b)小分子药物。本文提供的具有真实无规卷曲构象的多肽和具有真实无规卷曲构象的多肽区段还可用于以下环境:美容用品应用以及食品工业应用和饮料生产。本文提供的表现出真实无规卷曲构象的大的多肽仅由脯氨酸(P,Pro)和丙氨酸(A,Ala)残基组成,并且包含大于至少50个氨基酸、特别是至少约100、特别是至少约150、特别是至少约200、特别是至少约250、特别是至少约300、特别是至少约350、特别是至少约400个脯氨酸和丙氨酸氨基酸残基。两种氨基酸P和A都需要存在于本文提供的具有真实无规卷曲构象的大的多肽和具有真实无规卷曲构象的多肽区段中。本文还提供了编码本文公开的生物合成的无规卷曲多肽或多肽区段的核酸分子,以及编码药物或食品缀合物的核酸分子,所述药物或食品缀合物包含所述生物合成的无规卷曲多肽或多肽区段和(共价连接的)感兴趣的蛋白,例如生物活性蛋白。

[0017] 本文所述的和可用于本文提供的药物或食品缀合物中的且包含由至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400个脯氨酸(P)和丙氨酸(A)氨基酸残基组成的氨基酸序列的生物合成的无规卷曲多肽或生物合成的无规卷曲多肽区段,特别可用于异源环境中,即,用于生物活性异源蛋白、蛋白构建体和/或药物缀合物中,所述药物缀合物包含所述生物合成的无规卷曲多肽或多肽区段和药学上或医学上有用的分子,例如小分子、肽或生物大分子,例如蛋白、核酸、碳水化合物、脂质囊泡等。如后文实施例所示,本发明人成功提供了由本文所述的真实无规卷曲多肽和生物活性蛋白或蛋白段组成的药物缀合物,以及由包含和/或连接于本文所述的无规卷曲多肽的小分子或小分子药物组成的药物缀合物,所述无规卷曲多肽仅由脯氨酸和丙氨酸氨基酸残基组成(即,由氨基酸P和A组成)。

[0018] 因此,本发明特别提供了生物活性异源蛋白,其包含至少两个结构域,其中(a)所述至少两个结构域的第一结构域包含具有和/或介导所述生物活性的氨基酸序列;和(b)所述至少两个结构域的第二结构域包含生物合成的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段由氨基酸序列组成,所述氨基酸序列由至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400个脯氨酸和丙氨酸氨基酸残基组成。按照本发明,所述“第一结构域和所述“第二结构域”不包含在天然(即,天然存在)蛋白或来源于天然存在的编码核酸序列(例如开放读码框等)的假定蛋白中。

[0019] 此外,本发明提供了药物缀合物,其由生物合成的无规卷曲多肽或多肽区段和药学上、治疗上和/或医学上有用的分子组成,所述无规卷曲多肽或多肽区段包含由至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400个脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,所述药学上、治疗上和/或医学上有用的分子例如小分子、肽或生物大分子,例如蛋白、核酸、碳水化合物、脂质囊泡等,所述药学上、治疗上和/或医学上有用的分子与所述生物合成的无规卷曲多肽或多肽区段缀合。此外,应当注意到,对于本文公开的缀合物,术语“生物活性”不限于纯生物分子,还包括具有医学活性的、治疗活性的、药学活性的分子等。对于本领域技术人员显而易见的是,本文提供的方式和方法不限于药学和医学用途,还可以用于很多技术中,包括但不限于美容用品、食品、饮

料和营养学技术、石油工业、造纸工业等。

[0020] 相比于化学合成的共聚多肽(例如Izuka,同上),本文提供的无规卷曲多肽是通过生物合成方式产生的。本文所用的术语“生物合成”是指借助于生物技术方法合成(与化学合成不同)。这些生物技术方法是本领域公知的,并且在下文有进一步的描述。本发明的无规卷曲多肽的生物合成允许产生具有规定的脯氨酸和丙氨酸残基序列、规定的长度和/或规定的脯氨酸和丙氨酸残基比率的多肽。此外,本发明所提供的多肽是基本上纯的,即,所产生的多肽是基本上均一的,并共有以上特征(即,规定的序列、规定的长度和/或规定的氨基酸比率)。按照本发明,由至少约50、特别是至少约100、特别是至少约150、特别是至少约200、特别是至少约250、特别是至少约300、特别是至少约350、特别是至少约400个脯氨酸和丙氨酸氨基酸残基组成的无规卷曲多肽例如包含在生物活性异源多肽/多肽构建体和/或药物或食品缀合物以及在其它工业领域有用的其它缀合物中,所述其它工业领域例如,但不限于,造纸工业、石油工业等。

[0021] 总体上,本发明多肽的上述特征允许形成稳定的无规卷曲多肽,并且这些无规卷曲多肽具有令人意想不到的和有利的性质。例如,本发明的多肽在水性溶液中是完全可溶的,并且具有增加的流体动力学体积。令人意想不到的是,本文所述的无规卷曲多肽还能提供增加的体内/体外稳定性。这对于医学应用是特别重要的,例如,对于包含本发明无规卷曲多肽的生物活性蛋白或药物缀合物。然而,本发明无规卷曲多肽的许多有利性质不仅允许它们可用于医学领域,而且还可用于其它领域,例如美容用品/美容治疗或营养学和食品技术领域,例如乳品业或肉品加工。可用于食品工业等的缀合物的实例,是包含本文公开的无规卷曲多肽或多肽区段和可用于这些技术的化合物(例如用作乳化剂的非离子型表面活性剂的聚氧化丙烯或聚氧化乙烯聚合物)的缀合物,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。本文还考虑到在生物化学方法和技术过程中使用本文所述的生物合成的无规卷曲多肽,例如造纸、采油等。本文提供的仅由脯氨酸和丙氨酸残基组成的生物合成的无规卷曲多肽(以及本文公开的包含所述生物合成的真实无规卷曲多肽的缀合物和构建体,例如药物或食品缀合物/构建体)的令人意想不到的和有利的特征在下文更详细地描述。此外,下文提供了利用这些本发明生物合成的无规卷曲多肽的示例性的应用和方式和方法。本文还提供了产生所述生物合成的无规卷曲多肽以及生物活性异源多肽或多肽构建体的方式和方法,以及制备本文公开的包含所述无规卷曲多肽的缀合物和构建体例如药物构建体的方式和方法。

[0022] 在本发明的背景下,惊奇地发现,具有均匀组成的脯氨酸-丙氨酸聚合物/多肽形成稳定的无规卷曲构象。这在后文的实施例中也得到证明,其中通过圆二色(CD)光谱验证生物合成的脯氨酸/丙氨酸(共)聚合物/多肽的无规卷曲结构。获得和利用所述生物合成的真实无规卷曲多肽/聚合物是令人意想不到的,因为已建立的Chou-Fasman法(Chou and Fasman (1974), *Biochemistry* 13, 223-245)预测到由脯氨酸和丙氨酸组成的聚合物/多肽(或其区段)为100% α -螺旋二级结构,如图7所示。然而,本文惊奇地发现并通过实验证实,具有均匀组成的脯氨酸-丙氨酸聚合物/多肽形成稳定的无规卷曲构象。这在后文的实施例中也得到证明,其中通过诸如圆二色(CD)光谱和尺寸排阻色谱(SEC)的实验技术验证脯氨酸/丙氨酸(共)聚合物/多肽的无规卷曲结构。

[0023] 相比于本发明的多肽/聚合物,例如Izuka (1993) (同上)中所述的化学合成多肽具

有任意/无规定的且随机的序列以及不同的长度。因此,化学合成多肽包含具有不同脯氨酸/丙氨酸比率、长度等的完全不同的肽的混合物。如Izuka所述,这样的混合物中的化学合成多肽不会(或仅会部分地)形成无规卷曲,并因此不会具有本文下文提供和描述的生物合成的多肽的任何有利性质。因此,本发明包括和涉及包含本文公开的本发明生物合成的无规卷曲多肽/聚合物的组合物,其中所述生物合成的无规卷曲多肽/聚合物特别由它们的仅包含脯氨酸和丙氨酸残基的序列定义。在一个具体实施方案中,本发明涉及包含作为一个组成部分的本文公开的这些无规卷曲多肽/聚合物的缀合物,例如药物或食品缀合物。在一个实施方案中,所述组合物中包含的这些本发明生物合成的无规卷曲多肽/聚合物具有均匀长度。

[0024] 如上文所述,本发明的仅由脯氨酸和丙氨酸残基组成的生物合成的无规卷曲多肽(或无规卷曲多肽区段)令人意想不到地形成稳定的无规卷曲构象。本文所用的术语“无规卷曲”通常涉及包括氨基酸聚合物/氨基酸序列/多肽在内的聚合分子的任何构象,其中形成所述聚合结构的各个单体元素基本上随机地定向朝向相邻的单体元素,同时还与所述相邻的单体元素化学结合。具体而言,采取/具有形成“无规卷曲构象”的多肽、氨基酸序列或氨基酸聚合物基本上缺少规定的二级结构和三级结构。对于本发明的多肽,形成聚合结构(即,多肽/氨基酸序列)的单体元素是诸如脯氨酸和丙氨酸的单个氨基酸本身或诸如下文进一步描述和限定的“氨基酸重复”/“氨基酸盒”/“盒重复”/“构成部分(building block)”/“模块(module)”(或其片段)的肽段。

[0025] 多肽无规卷曲的性质和它们的实验鉴定方法是本领域技术人员已知的,并且在科学文献中已有描述(Cantor(1980)Biophysical Chemistry,2nd ed.,W.H.Freeman and Company,New York;Creighton(1993)Proteins-Structures and Molecular Properties,2nd ed.,W.H.Freeman and Company,New York;Smith(1996)Fold Des 1:R95-R106)。本文所用的术语“区段”是指本文所限定的生物合成的无规卷曲多肽的一部分,其中所述的一部分可以本文所述的生物合成的无规卷曲多肽的内部一部分。所述“区段”可以是,例如,一个(或多个)氨基酸被缺失的本文所限定的生物合成的无规卷曲多肽,例如从本发明多肽的起点和/或终点缺失。此外,所述“区段”可以被用作或可以形成较大的蛋白或多肽的一部分,例如,具有生物活性蛋白的融合蛋白的一部分。所述“融合蛋白”还是本发明的异源生物活性多肽/蛋白/多肽构建体的实例。本文所用的术语“异源”在本文的下文中定义。

[0026] 本发明提供的和可用于本发明的无规卷曲多肽(或其无规卷曲区段)例如在水性溶液或在生理条件下采取/形成无规卷曲构象。术语“生理条件”是本领域已知的,并且涉及蛋白通常采取其天然的、折叠构象的条件。更具体而言,术语“生理条件”涉及对于高等生命形式、特别是哺乳动物、最优选人类通常很重要的生物物理参数。术语“生理条件”可以涉及在哺乳动物并特别是人类的机体(特别是体液)中通常存在的生物化学和生物物理参数。所述“生理条件”可以涉及健康机体中存在的相应参数以及疾病条件下或人类患者中存在的参数。例如,当所述哺乳动物或所述人类发热时,患病哺乳动物或人类患者可以具有较高、但任然是“生理”的温度条件。对于蛋白采取其天然构象/状态的“生理条件”,最重要的参数是温度(对人体是37°C)、pH(对于人类血液为7.35-7.45)、同渗容摩(280-300mmol/kg H₂O)、以及如果必要的话,蛋白含量(66-85g/l血清)。然而,本领域技术人员知道,在生理条件下,这些参数可以变化,例如在指定的体液或组织液例如血液、脑脊液、腹膜液和淋巴液

中,温度、pH、同渗容摩和蛋白含量可以不同(Klinke (2005) Physiologie, 5th ed., Georg Thieme Verlag, Stuttgart)。例如,在脑脊液中,同渗容摩可以为约290mmol/kg H₂O,蛋白浓度可以为0.15g/l至0.45g/l,而在淋巴液中,pH可以为约7.4,蛋白含量可以为3g/l至5g/l。当使用下文所述的方法确定多肽(或其区段)/氨基酸序列在实验条件下是否形成/采取无规卷曲构象时,诸如温度、pH、同渗容摩和蛋白含量的生物物理参数可以不同于体内通常存在的生理条件。1°C至42°C或优选4°C至25°C的温度可以被认为可用于在体外测试和/或验证蛋白在生理条件下的生物物理性质和生物活性。

[0027] 数种缓冲液,特别是在实验环境中(例如在确定蛋白结构,特别是在CD测量和允许本领域技术人员测定蛋白/氨基酸段的结构性质的其它方法中)或在药物组合物的缓冲剂、溶剂和/或赋形剂中,被认为能代表体外“生理溶液”/“生理条件”。这些缓冲液的实例是,例如磷酸缓冲盐溶液(PBS:115mM NaCl、4mM KH₂PO₄、16mM Na₂HPO₄pH 7.4)、Tris缓冲液、乙酸盐缓冲液,柠檬酸盐缓冲液或类似缓冲液,例如后文实施例中所用的那些缓冲液。通常,代表“生理溶液条件”的缓冲液的pH应当位于6.5至8.5的范围内,优选位于7.0至8.0的范围内,最优选位于7.2至7.7的范围内,同渗容摩应当位于10至1000mmol/kg H₂O的范围内,更优选位于50至500mmol/kg H₂O的范围内,最优选位于200至350mmol/kg H₂O的范围内。可选地,代表生理溶液条件的缓冲液的蛋白含量可以位于0至100g/l的范围内,忽略具有生物活性的蛋白本身,其中可以使用典型稳定化蛋白,例如人或牛血清白蛋白。

[0028] 本文已发现,多肽(或其区段)不仅在生理条件下形成无规卷曲构象,而且更通常在水性溶液中形成无规卷曲构象。术语“水性溶液”是本领域公知的。“水性溶液”可以是水(H₂O)含量为至少约20%、至少约30%、至少约40%、至少约50%、至少约60%、至少约70%、至少约80%或至少约90%H₂O(重量/重量)的溶液。因此,本发明的多肽(或其区段)可以在可能含有其它可混溶剂的水性溶液中形成无规卷曲构象,或在具有更宽范围的温度、pH值、同渗容摩或蛋白含量的水性分散液中形成无规卷曲构象。这对于无规卷曲多肽(或其区段)在医药治疗或体内诊断之外的应用是特别相关的,例如在美容用品、营养学或食品技术中。

[0029] 因此,本发明中还考虑到,可以将本发明的脯氨酸/丙氨酸生物合成多肽(或其区段)的无规卷曲构象得到保持和/或利用于药物组合物例如液体药品/生物制品或冻干的药物组合物环境中。这对于本文提供的特别包含本发明无规卷曲多肽(或多肽区段)的生物活性异源蛋白或药物缀合物是特别重要的。优选地,“生理条件”会被用于相应的缓冲液体系、溶剂和/或赋形剂中。然而,例如在冻干的或干燥的组合物(例如,药物组合物/生物制品)中,考虑到本文提供的无规卷曲多肽(或多肽区段)的无规卷曲构象暂时不存在和/或不能被检测到。然而,在相应的缓冲液/溶液/赋形剂/溶剂复原后或在给予机体后,所述无规卷曲多肽(或多肽区段)会再次采取/形成其无规卷曲。确定多肽(或其区段)是否形成/采取无规卷曲构象的方法是本领域已知的(Cantor (1980) 同上;Creighton (1993) 同上;Smith (1996) 同上)。这些方法包括本文后文所列举的圆二色(CD)光谱。CD光谱代表光吸收光谱方法,其中测量某种物质对右旋圆偏振光和左旋圆偏振光的吸光度的差异。可以利用波长为约190-250nm的远紫外光谱通过CD光谱确定蛋白的二级结构。在这些波长下,可以分析多肽中常见的不同二级结构,因为 α 螺旋、平行和反向平行 β 折叠以及无规卷曲构象中的每一种都能产生特征性的CD谱形状和幅度。因此,通过使用CD光谱法,本领域技术人员能容易地确定多肽(或其区段)在水性溶液中或在生理条件下是否形成/采取无规卷曲构象。其它建立

的生物物理学方法包括核磁共振(NMR)光谱、吸收光谱法、红外线和拉曼光谱、通过尺寸排阻色谱测量流体动力学体积、分析超离心或动态/静态光散射,以及测量摩擦系数或固有粘度(Cantor(1980)同上;Creighton(1993)同上;Smith(1996)同上)。

[0030] 除了上述实验方法之外,还描述了预测蛋白的二级结构的理论方法。这种理论方法的一个实例是Chou-Fasman法(Chou and Fasman,同上),其是基于根据例如用X射线晶体学解析的已知的蛋白结构来分析 α 螺旋、 β 折叠和转角中每种氨基酸的相对频率。然而,已知蛋白二级结构的理论预测是不可靠的。如下文所示例,根据Chou-Fasman法预期采取 α 螺旋二级结构的氨基酸序列通过实验发现形成无规卷曲。因此,诸如Chou-Fasman算法的理论方法对于指定多肽是否采取无规卷曲构象仅可以具有有限的预测价值,这在后文的实施例和图中也有证实。尽管如此,上述理论预测通常是评估给定多肽/氨基酸序列的推定二级结构中的第一方法。无规卷曲结构的理论预测还表明,通过以上实验手段验证给定多肽/氨基酸序列是否确实具有无规卷曲构象可能是非常值得的。

[0031] 大部分氨基酸、特别是疏水性氨基酸的同聚物在水性溶液中通常是不溶的(Bamford(1956) *Synthetic Polypeptides-Preparation, Structure, and Properties*, 2nd ed., Academic Press, New York)。数种亲水性氨基酸的同聚物已知能形成二级结构,例如,对于Ala为 α 螺旋(Shental-Bechor (2005) *Biophys J* 88:2391-2402),对于Ser为 β 折叠(Quadrifoglio(1968) *J Am Chem Soc* 90:2760-2765),而聚脯氨酸是硬度最强的同聚寡肽(Schimmel(1967) *Proc Natl Acad Sci USA* 58:52-59),其在水性溶液中形成II型反式螺旋(Cowan(1955) *Nature* 176:501-503)。

[0032] 例如,对于聚甘氨酸,利用聚合物生物物理学的理论原理,200个氨基酸残基的链的无规卷曲直径相当于约75Å,其计算为均方根端到端距离 $\sqrt{\langle r^2 \rangle_0} = l \cdot \sqrt{n \cdot C_\infty}$ 其中 $n=200$,对于每个Ca-Ca距离,长度 l 的可转动的键=3.8 Å,对于聚(Gly)的“特征比” $C_\infty \approx 2.0$ (Brant(1967) *J Mol Biol* 23:47-65;Creighton, (1993) 同上)。这种关系表明,本领域技术人员会预期到:通过(a)利用较长的链长度 l 或通过(b)利用表现出较大的特征比 C_∞ 的氨基酸可以增大随机链氨基酸聚合物的流体动力学体积。 C_∞ 是分子随机链的固有硬度的衡量标准,并且对于大多数氨基酸的一般值为9(Brant(1967) 同上)。只有缺少侧链的Gly以及亚氨基酸Pro表现出显著较小的值。因此,Gly和Pro(在变性条件下)预期能有助于减小无规卷曲蛋白的维度(Miller(1968) *Biochemistry* 7:3925-3935)。因此,包含脯氨酸残基的氨基酸序列预期具有相对紧凑的流体动力学体积。然而,与该教导相反,本文表明:与预期的流体动力学体积相比,如通过分析凝胶渗透/尺寸排阻色谱所测定的,包含脯氨酸和丙氨酸残基混合物的本发明氨基酸聚合物/多肽的流体动力学体积具有明显增加的流体动力学体积。事实上,令人意想不到的是,包含这两种氨基酸(脯氨酸和丙氨酸)的混合物的多肽在生理条件下采取无规卷曲构象,这两种氨基酸中的每一种单独倾向于形成具有规定的二级结构的同聚寡肽。这些本发明的脯氨酸/丙氨酸多肽具有的流体动力学半径大于包含相同数量的Gly残基的同聚物的半径,例如,并且它们为本发明的生物活性蛋白或构建体(即,生物活性异源蛋白或药物缀合物)提供更好的可溶性。

[0033] 如上文所述,本发明的生物合成的无规卷曲脯氨酸/丙氨酸多肽与化学合成多肽的不同之处在于,通过简单的方式和方法,它们能采取规定的均匀长度。然而,现有技术提

供的多肽的混合物/组合物在肽的长度方面具有明显变化,而本发明能提供具有规定的长度的生物合成的无规卷曲多肽的混合物/组合物。优选地,这样的混合物/组合物中包含的基本上所有本发明的多肽具有相同的规定的长度,并因此共有相同的生物化学性质。这样的均匀组合物在可应用生物合成的无规卷曲多肽的各种医学、美容用品、营养学应用中更有优势。此外,特别是在医学或药学背景下,包含由至少约50、特别是至少约100、特别是至少约150、特别是至少约200、特别是至少约250、特别是至少约300、特别是至少约350、特别是至少约400个脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列的本文所限定的生物合成的无规卷曲多肽或多肽区段,还可用于预防、改善和/或治疗与受损的血浆状况有关和/或相关的病症,例如损伤、烧伤、手术后等。因此,所述生物合成的无规卷曲多肽或多肽区段的一个医学应用是用作血浆扩容剂。然而,应当注意到,按照本发明,本文所述的药物缀合物和异源多肽或异源多肽构建体可以可用于与受损的血浆量或血浆内容物相关的病症或与受损的血容量相关的病症的医学或药物干预。

[0034] 因此,在一个实施方案中,本发明涉及生物合成的随机多肽(或其区段),其包含仅由至少约50个脯氨酸和丙氨酸氨基酸残基、至少约100个脯氨酸和丙氨酸氨基酸残基、至少约150个脯氨酸和丙氨酸氨基酸残基或至少约200个脯氨酸和丙氨酸残基组成的氨基酸序列,特别是当所述生物合成的随机多肽(或其区段)包含在异源蛋白/多肽/多肽构建体或药物缀合物中时。本发明还涉及生物合成的无规卷曲多肽,其包含仅由至少约200个脯氨酸和丙氨酸氨基酸残基、甚至更优选至少约300个脯氨酸和丙氨酸氨基酸残基、特别优选至少约400个脯氨酸和丙氨酸氨基酸残基、更特别优选至少约500个脯氨酸和丙氨酸氨基酸残基和最优选至少约600个脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。形成无规卷曲构象的氨基酸序列可以由最多约3000个脯氨酸和丙氨酸氨基酸残基、最多约2000个脯氨酸和丙氨酸氨基酸残基、最多约1500个脯氨酸和丙氨酸氨基酸残基、最多约1200个脯氨酸和丙氨酸氨基酸残基、最多约800个脯氨酸和丙氨酸氨基酸残基组成。因此,脯氨酸/丙氨酸氨基酸序列段可以由约50、约100、约150、约200、约250、约300、约350、约400、约500、约600、约700、约800、约900至约3000个脯氨酸和丙氨酸氨基酸残基组成。在某些实施方案中,本发明生物合成的氨基酸序列包含约200至约3000个脯氨酸和丙氨酸残基、约200至约2500个脯氨酸和丙氨酸残基、约200至约2000个脯氨酸和丙氨酸残基、约200至约1500个脯氨酸和丙氨酸残基、约200至约1000个脯氨酸和丙氨酸残基、约300至约3000个脯氨酸和丙氨酸残基、约300至约2500个脯氨酸和丙氨酸残基、约300至约2000个脯氨酸和丙氨酸残基、约300至约1500个脯氨酸和丙氨酸残基、约300至约1000个脯氨酸和丙氨酸残基、约400至约3000个脯氨酸和丙氨酸残基、约400至约2500个脯氨酸和丙氨酸残基、约400至约2000个脯氨酸和丙氨酸残基、约400至约1500个脯氨酸和丙氨酸残基、约400至约1000个脯氨酸和丙氨酸残基、约500至约3000个脯氨酸和丙氨酸残基、约500至约2500个脯氨酸和丙氨酸残基、约500至约2000个脯氨酸和丙氨酸残基、约500至约1500个脯氨酸和丙氨酸残基、约500至约1000个脯氨酸和丙氨酸残基、约600至约3000个脯氨酸和丙氨酸残基、约600至约2500个脯氨酸和丙氨酸残基、约600至约2000个脯氨酸和丙氨酸残基、约600至约1500个脯氨酸和丙氨酸残基、约600至约1000个脯氨酸和丙氨酸残基、约700至约3000个脯氨酸和丙氨酸残基、约700至约2500个脯氨酸和丙氨酸残基、约700至约2000个脯氨酸和丙氨酸残基、约700至约1500个脯氨酸和丙氨酸残基、约700至约1000个脯氨酸和丙氨酸残基、约800至约3000个脯氨酸和丙氨酸残基、

约800至约2500个脯氨酸和丙氨酸残基、约800至约2000个脯氨酸和丙氨酸残基、约800至约1500个脯氨酸和丙氨酸残基、约800至约1000个脯氨酸和丙氨酸残基。从本发明的内容可以看出,较大的生物合成的氨基酸序列(基本上由脯氨酸和丙氨酸组成)也在本发明的范围内,并且能容易地用于本文所限定的生物活性蛋白或蛋白构建体,所述生物活性蛋白或蛋白构建体包含作为至少两个结构域中的一个结构域的具有和/或介导所述生物活性的氨基酸序列和作为至少两个结构域中的另一个结构域的生物合成的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段由至少约50个脯氨酸和丙氨酸氨基酸残基、至少约100个脯氨酸和丙氨酸氨基酸残基、至少约150个脯氨酸和丙氨酸氨基酸残基、至少约200、至少约250、至少约300、至少约350、至少约400个脯氨酸和丙氨酸氨基酸残基组成。这类生物合成的无规卷曲多肽或多肽区段相当于异源蛋白/蛋白构建体的生物合成的无规卷曲部分。这些生物合成的脯氨酸/丙氨酸段由最多约3000个脯氨酸和丙氨酸氨基酸残基组成。这些氨基酸序列(脯氨酸/丙氨酸段)包含作为主要或独特残基的脯氨酸和丙氨酸,这在下文中将进一步解释。

[0035] 考虑到,本文所限定的生物合成的氨基酸序列仅由脯氨酸(P)和丙氨酸(A)氨基酸残基组成,其形成/采取/具有无规卷曲构象。在最简单的情况下,生物合成的多肽或多肽区段由具有本文所限定的无规卷曲构象的氨基酸序列组成。然而,除了本文所述的形成/采取/具有无规卷曲构象的氨基酸序列之外,生物合成的多肽(或其区段)还可以包含不能有助于无规卷曲构象的形成或本身不能形成/采取/具有无规卷曲构象的氨基酸序列/氨基酸残基。在不脱离本发明主旨的情况下,这样的生物合成的多肽(或其区段)也是生物合成的“无规卷曲”多肽或多肽区段。例如,其它氨基酸序列/氨基酸残基可以被用作连接子。特别是,本发明中还考虑到生物合成的无规卷曲多肽的二聚体、三聚体,即,一般多聚体,并且这些多聚体可以通过不形成无规卷曲构象的氨基酸序列/残基连接。可以包含这样的无规卷曲多肽的蛋白的实例是本文提供的生物活性蛋白,除了本文所限定的由脯氨酸和丙氨酸氨基酸残基组成的无规卷曲多肽之外,其还可以包含具有/介导生物活性的另一多肽。此外,这样的构建体可以是本文所限定的异源生物活性蛋白或多肽构建体。

[0036] 术语“至少约50/100/150/200/300/400/500/600/700/800/等个氨基酸残基”不限于所述明确数量的氨基酸残基,还包括这样的氨基酸段,所述氨基酸段包含增加约1-20%(例如10%至20%)的残基或减少约1-20%(例如约10%至20%)的残基。例如“至少约100个氨基酸残基”还可以包括约80至100和约100至120个氨基酸残基,而不脱离本发明的主旨。例如“至少约200个氨基酸残基”还可以包括约160至200和约200至240个氨基酸残基,而不脱离本发明的主旨。在必要的变通下,上文给出的定义和解释还适用于术语“最多约3000/2000/1500/1200/800个氨基酸残基”等。因此,对于较长的氨基酸序列(例如包含最多3000个氨基酸残基或由最多3000个氨基酸残基组成的氨基酸序列),术语“约”不局限于或不受限于所述明确数量的氨基酸残基。因此,术语“最多约3000/2000/1500/1200/800个氨基酸残基”还可以包括这样的氨基酸段,所述氨基酸段包含增加10%至20%或减少10%至20%的残基,这不脱离本发明。

[0037] 此外,生物合成的无规卷曲多肽(或其区段)的特征在于氨基酸残基、特别是主要组成部分脯氨酸和丙氨酸的规定的含量或比率。如上文所述,本发明涉及生物合成的无规卷曲多肽或多肽区段,其包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所

述氨基酸序列由至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成,特别是当所述生物合成的无规卷曲多肽或多肽区段包含在异源生物活性蛋白/蛋白构建体/多肽或药物缀合物中时。本文所用的术语“仅”表示,优选至少约90%或至少约95%的氨基酸是脯氨酸和丙氨酸,其中脯氨酸和丙氨酸构成主要部分,但可以不是仅有的氨基酸残基,即,本发明的这些氨基酸序列未必是100%的脯氨酸和丙氨酸氨基酸段。因此,本发明的生物合成的多肽/氨基酸序列还可以包含作为次要组成部分的除脯氨酸和丙氨酸之外的其它氨基酸,只要氨基酸序列形成/采取/具有无规卷曲构象。通过本文提供的方式和方法能容易地确定这类无规卷曲构象。因此,对于术语“仅”,可以包含少量(小于约10%或小于约5%)的其它氨基酸残基。所述“其它”的次要氨基酸残基在下文中有定义。

[0038] 因此,在一个实施方案中,本发明涉及生物合成的无规卷曲多肽(或其区段),其中氨基酸序列主要由脯氨酸和丙氨酸组成,并且其中脯氨酸残基构成大于约10%且小于75%的氨基酸序列。丙氨酸残基包含所述氨基酸序列(或无规卷曲多肽或多肽区段,如果其由氨基酸序列组成)的剩余至少25%至90%。

[0039] 优选地,氨基酸序列包含大于约10%、优选大于约12%、甚至更优选大于约14%、特别优选大于约18%、更特别优选大于约20%、甚至更特别优选大于约22%、23%或24%和最优选大于约25%的脯氨酸残基。氨基酸序列优选包含小于约75%、更优选小于70%、65%、60%、55%或50%的脯氨酸残基,其中较低的值是优选的。甚至更优选地,氨基酸序列包含小于约48%、46%、44%、42%的脯氨酸残基。特别优选的是,氨基酸序列包含小于约41%、40%、39%、38%、37%或36%的脯氨酸残基,其中较低的值是优选的。最优选地,氨基酸序列包含小于约35%的脯氨酸残基;还可参见下文提供的PA构建体。

[0040] 反之亦然,氨基酸序列优选包含小于约90%、更优选小于88%、86%、84%、82%或80%的丙氨酸残基,其中较低的值是优选的。甚至更优选地,氨基酸序列包含小于约79%、78%、77%、76%的丙氨酸残基,其中较低的值是优选的。最优选地,氨基酸序列包含小于约75%的丙氨酸残基。

[0041] 本文还优选的是,氨基酸序列包含大于约25%、优选大于约30%、甚至更优选大于约35%、特别优选大于约40%、更特别优选大于约45%或50%、甚至更特别优选大于约52%、54%、56%、58%或59%的丙氨酸残基,其中较高的值是优选的。甚至更优选地,氨基酸序列包含大于约60%、61%、62%、63%或64%的丙氨酸残基,最优选大于约65%的丙氨酸残基。

[0042] 因此,无规卷曲多肽(或其区段)可以包含由约25%的脯氨酸残基和约75%的丙氨酸残基组成的氨基酸序列。或者,无规卷曲多肽(或其区段)可以包含由约35%的脯氨酸残基和约65%的丙氨酸残基组成的氨基酸序列。上文所用的术语“约X%”不限于所述明确数量的百分比,还包括增加10%至20%或减少10%至20%的残基的值。例如,术语10%还可以分别涉及11%或12%和9%和8%。

[0043] 然而,如上文所提及的和下文进一步的详细描述,所述无规卷曲多肽(或多肽区段)并且特别是氨基酸序列还可以包含不同于脯氨酸和丙氨酸的作为次要组成部分的其它氨基酸。如上文已经讨论的那样,所述次要组成部分,即,除脯氨酸或丙氨酸之外的氨基酸,可以构成本发明生物合成的无规卷曲多肽/聚合物的小于约10%、小于约9%、小于约8%、小于约7%、小于约6%、小于约5%、小于约4%、小于约4%、小于约3%或小于约2%。

[0044] 本领域技术人员明白,当除脯氨酸或丙氨酸之外的其他残基作为次要组成部分包含在氨基酸序列/多肽(多肽区段)中时,所述氨基酸序列/多肽(或其区段)还可以形成无规卷曲构象。本文所用的术语“次要组成部分”表示,在本发明的生物合成的无规卷曲多肽/聚合物中,最多5%或最多10%的氨基酸残基不同于脯氨酸或丙氨酸。这表示,在100个氨基酸中,最多10个可以不同于脯氨酸和丙氨酸;优选最多8%,即,在100个氨基酸中,最多8个可以不同于脯氨酸和丙氨酸;更优选最多6%,即,在100个氨基酸中,最多6个可以不同于脯氨酸和丙氨酸;甚至更优选最多5%,即,在100个氨基酸中,最多5个可以不同于脯氨酸和丙氨酸;特别优选最多4%,即,在100个氨基酸中,最多4个可以不同于脯氨酸和丙氨酸;更特别优选最多3%,即,在100个氨基酸中,最多3个可以不同于脯氨酸和丙氨酸;甚至更特别优选最多2%,即,在100个氨基酸中,最多2个可以不同于脯氨酸和丙氨酸;最优选最多1%,即,在无规卷曲多肽(或其区段)包含的100个氨基酸中,最多1个可以不同于脯氨酸和丙氨酸。所述不同于脯氨酸和丙氨酸的氨基酸可以选自Arg、Asn、Asp、Cys、Gln、Glu、Gly、His、Ile、Leu、Lys、Met、Phe、Thr、Trp、Tyr和Val,包括翻译后修饰的氨基酸或非天然氨基酸(参见,例如, Budisa (2004) *Angew Chem Int Ed Engl* 43:6426-6463或Young (2010) *J Biol Chem* 285: 11039-11044)。当生物合成的无规卷曲多肽/构建体/聚合物(或其片段)的“次要组成部分”(即,除脯氨酸或丙氨酸之外的氨基酸)包含Ser作为“其它氨基酸”/“不同的氨基酸”时,所述Ser氨基酸优选构成这些(次要)氨基酸残基的小于50%、更优选小于40%、小于30%、小于20%或小于10%。在最优选的实施方案中,本文所述的生物合成的无规卷曲多肽/构建体/聚合物或本文所述的(例如)融合蛋白的无规卷曲多肽部分不包含丝氨酸残基。通常,本文优选的是,这些“次要”氨基酸(除脯氨酸和丙氨酸之外)不存在于本文提供的生物合成的无规卷曲多肽/构建体/聚合物或(例如)融合蛋白的无规卷曲多肽部分。按照本发明,生物合成的无规卷曲多肽(或其区段)/氨基酸序列可以特别地仅仅由脯氨酸和丙氨酸氨基酸残基构成(即,无规卷曲多肽或氨基酸序列中不存在其它氨基酸残基)。

[0045] 虽然上文涉及无规卷曲多肽(或其区段)中包含的氨基酸序列的总长度和脯氨酸/丙氨酸含量,但下文更详细地涉及具体的示例性氨基酸序列(或其片段)。

[0046] 在一个实施方案中,例如在水性溶液中或在生理条件下采取无规卷曲构象的氨基酸序列/多肽(本文所限定的无规卷曲多肽或其区段)可以包含多个“氨基酸重复”/“氨基酸盒”/“盒重复”,其中所述“氨基酸重复”/“氨基酸盒”/“盒重复”/“构成部分”/“模块”(这些术语在本文中可交换使用)主要或仅仅由脯氨酸(Pro, P)和丙氨酸(Ala, A)氨基酸残基组成(本文描述为“PA”或“AP”),其中不多于6个连续氨基酸残基是相同的。示例性的“构成部分”是例如“AP”,并且在后文示例性的实例中将其提供为本发明的功能性生物合成的无规卷曲结构域。该示例性的实例是序列“P1A1”,其也以APAPAPAPAPAPAPAPAPAP (SEQ ID NO:51)的形式提供,即,“聚PA”“氨基酸重复”/“氨基酸盒”/“盒重复”。在优选的实施方案中,包含上文所限定的“氨基酸重复”/“氨基酸盒”/“盒重复”等的氨基酸序列/多肽包含不多于5个相同的连续氨基酸残基。对于示例性的个体构成部分,下文提供了其它可选实施方案。

[0047] 在本发明的无规卷曲多肽(或其区段)中,氨基酸重复可以是相同的或不相同的。由脯氨酸和丙氨酸残基组成的“氨基酸重复”、“构成部分”、“模块”、“重复”、“氨基酸盒”等的非限制性实例在下文提供;参见,例如SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6和SEQ ID NO:51(所附的序列表还包含编码这些“重复”/

“模块”等的示例性核酸序列。与本文共同提交的所述序列列表中的附加序列构成本说明书的一部分)。此外,本文考虑了这些序列(相同的和/或不相同的)片段的用途,其中“片段”包含至少2个氨基酸并包含至少一个脯氨酸和/或丙氨酸,优选至少一个脯氨酸和一个丙氨酸。本发明中用于产生无规卷曲多肽(或其区段)的这些序列的“片段”可以由选自所述SEQ ID No:1、2、3、4、5、6和51的氨基酸序列的至少3个、优选至少4个、更优选至少5个、甚至更优选至少6个、仍然更优选至少8个、特别优选至少10个、更特别优选至少12个、甚至更特别优选至少14个、仍然更特别优选至少16个、并且最优选至少18个连续氨基酸组成(此处应当注意到,SEQ ID No:51由示例性“AP”或“PA”重复组成)。

[0048] 基于本文给出的教导,本领域技术人员能够容易地制备能例如在水性或生理条件下形成无规卷曲构象并且如本文所限定主要由脯氨酸和丙氨酸组成的其它氨基酸序列/多肽。可用作本文所限定的无规卷曲多肽(或其区段)的构成部分或模块的形成无规卷曲构象的氨基酸序列/多肽的其它实例可以特别包含上文所示的具体“构成部分”、“聚合物盒”或“聚合物重复”的组合和/或片段或环状排列形式。因此,无规卷曲多肽/氨基酸序列的示例性模块/序列单元/聚合物重复/聚合物盒还可以提供可以按照本发明重新组合而形成其它模块/序列单元/聚合物重复/聚合物盒的个体片段。

[0049] 术语“模块”、“序列单元”、“聚合物重复”、“聚合物盒”和“构成部分”在本文作为同义词使用,并涉及可以用于形成本文所限定的无规卷曲多肽(或其区段)/氨基酸序列的个体氨基酸段。

[0050] 氨基酸重复(用作本发明生物合成的无规卷曲多肽的“构成部分”等)可以由至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30或更多个氨基酸残基组成,其中每个重复都包含脯氨酸和丙氨酸残基。然而,如后文的SEQ ID No:51所示,所述“构成部分”还可以仅由2个本文提供的氨基酸残基P和A组成,即“PA”或“AP”的形式。在一个实施方案中,本发明的氨基酸重复不包含大于50个氨基酸残基。然而,对于本领域技术人员显而易见的是,所述“重复”可以包含甚至大于50个氨基酸残基,例如当所述本发明生物合成的无规卷曲多肽/聚合物包含大于约例如100个氨基酸、大于约150个氨基酸、大于约200个氨基酸等的情况下。因此,所述“重复”中包含的氨基酸残基的最大量以本文提供的生物合成的多肽(或其区段)/聚合物的总长度为前提条件。

[0051] 然而,应当注意到,包含上述重复等的生物合成的无规卷曲多肽/氨基酸序列应当优选具有上文限定和解释的总长度和/或脯氨酸/丙氨酸含量,即,由约50、约100、约150、约200、约250、约300、约350、约400至约3000个氨基酸组成和/或包含大于约10%且小于约75%的脯氨酸残基。在必要的变通下,上文中关于此方面给出的所有限定也适用于此处。

[0052] 正如本文详细讨论和上文所提供的那样,本发明提供了在药学、医学和/或治疗领域中特别有用的生物活性异源蛋白或蛋白构建体。这些生物活性异源蛋白/蛋白构建体包含作为所述至少两个结构域中的至少一个结构域的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸残基组成的氨基酸序列,其中所述氨基酸序列由约50、约100、约150、约200、约250、约300、约350、约400至约3000个脯氨酸(Pro)和丙氨酸(Ala)残基组成。

[0053] 对于本文公开的生物活性异源蛋白、多肽或蛋白构建体,术语“异源”涉及所述蛋白、多肽或蛋白构建体中的至少两个结构域,其中所述至少两个结构域的第一结构域提供、

具有和/或介导规定的生物活性,并且其中所述至少两个结构域的第二结构域包含仅由脯氨酸和丙氨酸氨基酸残基组成的生物合成的无规卷曲多肽,并且其中所述至少两个结构域在自然界不是彼此有效连接的,或不是由自然界存在的单个编码核酸序列(例如开放读码框)所编码的。本文提供的和用于本发明生物活性异源蛋白/蛋白构建体的仅由脯氨酸和丙氨酸氨基酸残基组成的生物合成的无规卷曲多肽/多肽区段优选不进行其它(化学)修饰,例如它们优选不是糖基化或羟基化的。

[0054] 应当注意到,某些天然存在的蛋白或从测序的天然存在核酸段推定的假定蛋白被描述为包含相对高(即,高于平均值)含量的脯氨酸和丙氨酸。例如,已对于利什曼虫(*Leishmania*)主要株Friedlin描述了同源假定蛋白(Ivens (2005) *Science* 309, 436-442.)。所公开的包含1514个三联密码子的读码框包括一段412个三联密码子,其由240个Ala、132个Pro、34个Lys和4个Val密码子组成。Lys残基在生理缓冲液条件下是带正电的,它们几乎均匀地分布在该序列中,这提示增溶效果。然而,从本文公开内容能够明显判断出,这种源自天然存在的核酸分子或开放读码框、包含高于平均值的高脯氨酸和丙氨酸含量的同源假定蛋白不是本发明的一部分。本发明是基于下述事实:提供了在医学/药物领域特别有用的相当大的无规卷曲多肽或多肽区段,其不在自然中以分离的形式存在,并且其包含仅由脯氨酸和丙氨酸残基组成的氨基酸序列,其中所述氨基酸序列由约50、约100、约150、约200、约250、约300、约350、约400至约3000个脯氨酸(Pro)和丙氨酸(Ala)残基组成。本文所述的不在自然中以分离的形式存在的分离的生物合成无规卷曲多肽或多肽区段也包含在本文公开的在药学、医学和/或医疗领域中特别有用的生物活性异源蛋白或蛋白构建体中。这些生物活性异源蛋白/蛋白构建体包含作为所述至少两个结构域中的至少一个结构域的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸残基组成的氨基酸序列,其中所述氨基酸序列由约50、约100、约150、约200、约250、约300、约350、约400至约3000个脯氨酸(Pro)和丙氨酸(Ala)残基组成。

[0055] 另外,阿拉伯半乳聚糖蛋白(AGP)(富含Pro蛋白)和伸展蛋白属于一大组称为富含羟基脯氨酸(Hyp)的糖蛋白(HRGPs)的糖蛋白,其在植物界广泛表达。包含Ala-Pro重复(AP)₅₁的一种所述AGP基序在转基因拟南芥(*Arabidopsis thaliana*)中被表达为具有N末端信号序列和C末端绿色荧光蛋白的合成糖模块肽,并被作为脯氨酰羟化酶和后续的羟基脯氨酸残基O-糖基化的底物得到研究(Estévez (2006) *Plant Physiol.* 142, 458-470)。此外,所公开的能与水分子形成氢键的羟化和/或糖基化Pro侧链看起来具有增溶效果。

[0056] 应当注意到,本文所述的“包含作为(至少)一个结构域的生物合成的无规卷曲多肽或多肽区段的生物活性蛋白或蛋白构建体,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸残基组成的氨基酸序列”涉及在自然界中通常不存在的并因此是“异源”的蛋白或蛋白构建体。此外,相比于植物界中所述的富含脯氨酸的序列,本文所述的生物合成的无规卷曲多肽/多肽区段优选不是化学修饰的,即,它们优选不是糖基化或羟基化的。

[0057] 本发明的生物合成的无规卷曲多肽或多肽区段的特别优势在于它们固有亲水、但不带电的性质。因此,作为本文所述的生物合成的无规卷曲多肽或多肽段中的“次要”氨基酸(除脯氨酸和丙氨酸之外),这些氨基酸优选不具有疏水性侧链(例如Val、Ile、Leu、Met、Phe、Tyr或Trp)和/或不具有带电侧链(例如Lys、Arg、Asp或Glu)。本发明考虑到(当这些个体氨基酸包含于本发明生物合成的无规卷曲多肽/多肽区段中时),本文所述的生物合成的

无规卷曲多肽(或其区段)中具有疏水性侧链(例如Val,Ile,Leu,Met,Phe,Tyr或Trp)和/或具有带电侧链(例如Lys,Arg,Asp或Glu)的每种个体氨基酸的总含量不超过8%、7%、6%、5%、4%、3%、2%或1%。

[0058] 本发明的生物合成的无规卷曲多肽/氨基酸序列可以包含个体构成部分的多连体,所述个体模块包含序列(Pro)_x-(Ala)_y的组合的脯氨酸/丙氨酸段,其中x可具有1至优选15、更优选1至10、甚至更优选1至5的整数值,y可具有1至优选15、更优选1至10、甚至更优选1至5的整数值,并且x和y在相继的构成部分之间可以不同。所述x和y也可以是以下整数:1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。

[0059] 能在水性溶液中或在生理条件下形成无规卷曲构象的氨基酸序列/多肽可以具有式(I):

[0060] $[\text{Pro}_x\text{Ala}_y]_n$

[0061] 其中x独立选自整数1至5。此外,对于每个n,y独立选自整数1至5。最后,n是任何整数,条件是无规卷曲多肽(或其区段)/氨基酸序列优选由至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400个氨基酸残基和最高至约3000个氨基酸残基组成。在该方面还应当注意到,包含上述多连体或具有上述式(I)的多肽/氨基酸序列应当优选具有上文所限定和解释的总长度和/或脯氨酸/丙氨酸含量,即,由约50、约100、约150、约200、约250、约300、约350、约400至约3000个氨基酸组成和/或包含大于约10%且小于约75%的脯氨酸残基。此外,在必要的变通下,上文对于该方面给出的所有限定也适用于此处。

[0062] 本发明还涉及包含选自以下的氨基酸段的无规卷曲多肽(多肽区段)/氨基酸序列:AAPAAPAPAAPAAPAPAPA (SEQ ID NO:1);AAPAAPAPAAPAAPAPAAP (SEQ ID NO:2);AAAPAAAPAAAPAAAPAAAP (SEQ ID NO:3是 $[\text{Pro}_1\text{Ala}_3]_5$ 的一个实例);AAPAAPAPAAPAAPAPAAP (SEQ ID NO:4);APAAPAPAAAPAPAAAPAPAAAP (SEQ ID NO:5);AAAPAAPAPAPAPAPAPAPAP (SEQ ID NO:6)和APAPAPAPAPAPAPAPAP (SEQ ID NO:51是 $[\text{Pro}_1\text{Ala}_1]_{10}$ 的一个实例)或作为这些序列整体或这些序列的一部分的这些序列的环状排列形式或多聚体。因此,无规卷曲多肽(其多肽区段)/氨基酸序列可以包含氨基酸段AAPAAPAPAAPAAPAPAPA (SEQ ID NO:1),AAPAAPAPAAPAAPAPAPA (SEQ ID NO:1);AAPAAPAPAAPAAPAPAAP (SEQ ID NO:2);AAAPAAAPAAAPAAAPAAAP (SEQ ID NO:3);AAPAAPAPAAPAAPAPAAP (SEQ ID NO:4);APAAPAPAAAPAPAAAPAPAAAP (SEQ ID NO:5);AAAPAAPAPAPAPAPAPAPAP (SEQ ID NO:6)和APAPAPAPAPAPAPAPAP (SEQ ID NO:51),以及这些基序的组合或这些基序的片段和部分的组合,只要所产生的生物合成的无规卷曲多肽仅由脯氨酸和丙氨酸氨基酸残基组成,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。

[0063] 此外,根据本发明,可以使用上述氨基酸序列的环状排列形式。可以容易地产生例如AAPAAPAPAAPAAPAPAPA (SEQ ID NO:1)的示例性的环状排列形式,例如通过去除第一丙氨酸并在上述序列的末端添加另一丙氨酸。SEQ ID NO:1的这样的环状排列形式会是APAAPAPAAPAAPAPAA (SEQ ID NO:7)。此外,SEQ ID NO:1的环状排列形式的非限制性实例是:

[0064] PAAPAPAPAPAPAPAPAAA (SEQ ID NO:8),

- [0065] AAPAAPAAPAAPAAPAAAP (SEQ ID NO:9) ,
[0066] APAPAAPAAPAAPAAAPA (SEQ ID NO:10) ,
[0067] PAPAAPAAPAAPAAAPAA (SEQ ID NO:11) ,
[0068] APAAPAAPAAPAAAPAAP (SEQ ID NO:12) ,
[0069] PAAPAAPAAPAAAPAAPA (SEQ ID NO:13) ,
[0070] AAPAAPAAPAAAPAAPAP (SEQ ID NO:14) ,
[0071] APAAPAAPAAAPAAPAPA (SEQ ID NO:15) ,
[0072] PAAPAAPAAAPAAPAPAA (SEQ ID NO:16) 等。

[0073] 基于本发明的教导,本领域技术人员能容易地制备SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6和SEQ ID NO:51(所述SEQ ID No:51完全基于“AP”重复,和环状排列形式可以完全基于“PA”或“AP”重复/构成部分)所示的氨基酸段的相应环状排列形式。

[0074] 这类环状排列形式还可以被认为是本文提供的多肽/氨基酸序列的其它“模块”/“构成部分”等的实例,因而可用于本发明。

[0075] 对于本领域技术人员显而易见的是,本文提供的氨基酸段的“模块”和(较短的)片段或环状排列形式可以被用作本文所限定的无规卷曲多肽(或其区段)/氨基酸序列的“模块”、“重复”和/或构成部分。

[0076] 按照上文,形成无规卷曲构象的无规卷曲多肽/氨基酸序列可以包含任何上述氨基酸段(或其环状排列形式或片段)的多聚体,优选SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6和SEQ ID NO:51所示的氨基酸段的多聚体。应当注意到,这些序列并非意图限制本发明。

[0077] 此外,包含上述氨基酸段(或其片段)、环状突变形式(或其片段)的多肽/氨基酸序列应当优选具有上文所限定和解释的总长度和/或脯氨酸/丙氨酸含量,即,由约50、约100、约150、约200、约250、约300、约350、约400至约3000个氨基酸组成和/或包含大于约10%且小于约75%的脯氨酸残基。在必要的变通下,上文对于该方面给出的所有限定也适用于此处。此外,术语“片段”在上文已有定义。

[0078] 如上文所述,对于本发明,意外发现本文提供的生物合成的无规卷曲多肽(或多肽区段)/聚合物的特征为相对较大的流体动力学体积。通过分析凝胶过滤(也称为尺寸排阻色谱,SEC)可以容易地测定该流体动力学体积(也称为表观尺寸)。优选地,无规卷曲多肽(或其区段)具有的表观尺寸为至少10kDa、优选至少25kDa、更优选至少50kDa、甚至更优选至少100kDa、特别优选至少200kDa,最优选至少400kDa。本领域技术人员能容易地测定具体蛋白的流体动力学体积。这些方法可以包括下文所示例的动态/静态光散射、分析超离心或分析凝胶过滤。分析凝胶过滤代表本领域已知的用于测量大分子的流体动力学体积的方法。或者,球形多肽的流体动力学体积可以通过其分子量来估计(Creighton(1993)同上)。如本文所述,相对于基于分子量对于相应的折叠球形蛋白所估计出的流体动力学体积,本发明的优选由至少约50、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400至约3000个脯氨酸和丙氨酸氨基酸残基组成并具有无规卷曲构象的多肽的流体动力学体积意想不到地表现出高值。下文涉及生物活性异源蛋白或蛋白构建体,其特别包含上文描述和限定的、代表本发明优选实施方案的生物合成的无规卷曲多肽(或其区

段)。在不被理论束缚的情况下,在本发明中意外地发现,本文提供的仅由脯氨酸和丙氨酸组成的生物合成的无规卷曲多肽段能提供的流体动力学体积甚至高于具有相同氨基酸残基总数、但仅由脯氨酸、丙氨酸和丝氨酸组成的相应的生物合成的无规卷曲段(如WO 2008/155134中所提供的)的流体动力学体积。

[0079] 诸如血清白蛋白(HSA)和免疫球蛋白(Ig)(包括人源化抗体)的常见人血浆蛋白表现出长的半衰期,通常为2至3周,这归因于它们与新生Fc受体(FcRn)的特异性相互作用,从而导致胞内体再循环(Ghetie (2002) *Immunol Res*, 25:97-113)。相比之下,大部分其它药理学上感兴趣的蛋白,特别是重组抗体片段、激素、干扰素等,会遭受快速(血液)清除。这对于大小小于约70kDa的肾过滤的阈值的蛋白是特别明显的(Caliceti (2003) *Adv Drug Deliv Rev* 55:1261-1277)。在这些情况下,未修饰的药理学蛋白的血浆半衰期可以显著小于1小时,因此使其基本无法用于大部分治疗应用。为了实现持续的药理学作用以及改善的患者依从性(使所需的给药间隔延长至数天或数周),为了生物药理学药物开发目的之前已经建立了数种策略。

[0080] 第一,已经通过下述方式利用了天然血浆蛋白的再循环机制:产生具有Ig的Fc部分的融合蛋白(例如**Enbrel®**,其是TNF α 受体的细胞外结构域和人IgG1的杂合体(Goldenberg (1999) *Clin Ther* 21:75-87))或具有血清白蛋白的融合蛋白(例如**Albuferon®**(albinterferon alfa-2b, ZALBIN™, **JOULFERON®**),其是IFN α 与HAS的对应融合物(Osborn (2002) *J Pharmacol Exp Ther* 303:540-548))。还也以间接的方式使用具有600 μ M的高血浆浓度的白蛋白,其作为配置了白蛋白结合功能的生物药物的载体,例如,通过与来自链球菌蛋白G的细菌白蛋白结合结构域(ABD)进行融合(Makrides (1996) *J Pharmacol Exp Ther* 277:534-542)或与从噬菌体展示文库针对HAS选择的肽进行融合(Dennis (2002) *J Biol Chem*, 277:35035-35043; Nguyen (2006) *Protein Eng Des Sel* 19:291-297)。

[0081] 第二,延长生物药物的血浆半衰期的本质上不同的方法是与高度溶剂化的和生理上惰性的化学聚合物缀合,从而有效增大治疗蛋白的流体动力学半径,使其超过约3-5nm的肾小球孔径(Caliceti (2003) 同上)。在生物化学上的温和条件下与聚乙二醇(PEG)的活化衍生物的共价偶联已取得一定成功,并且目前正用于数种获批准的药物,所述共价偶联是通过Lys侧链随机地偶联(Clark (1996) *J Biol Chem* 271:21969-21977)或借助于特别引入的Cys残基(Rosendahl (2005) *BioProcess International*:52-60)。已经取得了相应的优势,特别是在具有具体药理学活性的小蛋白中,例如**Pegasys®**,其是化学PEG化的重组IFN α -2a(Harris (2003) *Nat Rev Drug Discov*, 2:214-221; Walsh (2003) *Nat Biotechnol* 21:865-870)。

[0082] 然而,生物活性蛋白与合成聚合物的化学偶联在生物药物研发和生产方面具有缺点。合适的PEG衍生物是昂贵的,特别是需要高纯度时,并且它们与重组蛋白的缀合需要额外的体外加工和纯化步骤,这会降低产率并提高生产成本。事实上,PEG经常被醛类和过氧化物污染(Ray (1985) *Anal Biochem* 146:307-312),并且其在氧的存在下贮存时本质上易于化学降解。另外,如果治疗蛋白的生物化学活性位点附近的氨基酸侧链在PEG化过程中被修饰,则该治疗蛋白的药物功能可能受到影响。此外,与合成聚合物的化学偶联通常产生分子的异质混合物,这可能会表现出明显的体内活性变化。

[0083] 第三,已经提议使用生物活性蛋白的糖基化类似物来延长血浆半衰期,其中引入新的N连接糖基化共有序列;参见WO 02/02597;Perlman(2003)J Clin Endocrinol Metab 88:2327-2335;或Elliott(2003)Nat Biotechnol 21:414-420)。然而,所述糖工程化蛋白表现出改变的体内活性,这表明新的碳水化合物侧链影响工程化蛋白的生物活性。此外,额外的碳水化合物侧链可能会增加产生的生物活性分子的抗原性,这会明显增加安全性问题。此外,据报道,包含克氏锥虫来源的人工重复性序列PSTAD的融合蛋白能诱导反式唾液酸酶的延长的血浆半衰期(Alvarez(2004)JBC 279:3375-3381)。然而,据报道,这样的克氏锥虫来源的重复会诱导体液免疫应答(Alvarez(2004)同上)。因此,需要延长生物活性蛋白的作用的备选策略。

[0084] 意外地发现,本文公开的和按照本发明仅由脯氨酸和丙氨酸组成的生物合成的氨基酸序列/多肽能采取无规卷曲构象,特别是在生理条件下。因此,它们是提供下文限定的生物活性蛋白/多肽的“第二结构域”的有利的分子,即,所述生物活性蛋白/多肽包含在生理条件下形成无规卷曲构象并从而能为生物活性(“功能”)蛋白或多肽介导增加的体内和/或体外稳定性,特别是增加的血浆半衰期的多肽段。与所述无规卷曲结构域融合的功能蛋白的流体动力学体积显著增加,这可以通过使用本文所述的标准方法进行估计。因为无规卷曲结构域被认为不会干扰生物活性蛋白的第一结构域的生物活性,所以与其融合的功能蛋白所介导的生物活性基本上能被保留。此外,形成本文公开的无规卷曲结构域的氨基酸聚合物/多肽被认为是生物学上惰性很大的,特别是在血浆中的蛋白水解、免疫原性、等电点/静电行为、与细胞表面受体的结合以及内化的方面,但是其仍是生物可降解的,这会提供超过诸如PEG的合成聚合物的清楚优势。

[0085] 按照上文,本发明涉及包含本文所述的生物合成的无规卷曲多肽的生物活性蛋白。包含本文所述的生物合成的无规卷曲多肽的生物活性蛋白/蛋白构建体是异源生物活性蛋白/蛋白构建体。具体而言,本文还公开了包含至少两个结构域或由至少两个结构域组成的生物活性异源蛋白,其中

[0086] (a)所述至少两个结构域的第一结构域包含具有和/或介导所述生物活性的氨基酸序列或由具有和/或介导所述生物活性的氨基酸序列组成;和

[0087] (b)所述至少两个结构域的第二结构域包含本文所述的和限定的无规卷曲多肽或多肽区段或由本文所述的和限定的无规卷曲多肽或多肽区段组成。

[0088] 应当注意到,按照本发明,所述“第一结构域”和所述“第二结构域”涉及这样的蛋白段,所述蛋白段在同一蛋白中不是天然存在的或预期不是自然界存在的编码核酸序列(例如,开放读码框)所编码的同一假定蛋白的一部分。

[0089] 在必要的变通下,上文对于无规卷曲多肽或其多肽区段的给出的限定和解释也适用于包含所述无规卷曲多肽(或其多肽区段)的生物活性蛋白。

[0090] 优选地,所述无规卷曲构象介导所述生物活性蛋白的增加的体内和/或体外稳定性,例如生物样品中或生理环境中的体内和/或体外稳定性。

[0091] 例如,本文考虑到,包含本文所限定的能在水性溶液中或在生理条件下采取无规卷曲构象的其它“第二结构域”(例如,由约200或约400或约600个氨基酸残基组成并包含作为“构成部分”的PA#1/SEQ ID NO:1、PA#2/SEQ ID NO:2、PA#3/SEQ ID NO:3、PA#4/SEQ ID NO:4、PA#5/SEQ ID NO:5、PA#6/SEQ ID NO:6和/或P1A1/SEQ ID NO:51的聚合物)的蛋

白与缺少所述无规卷曲构象的对照相比,具有有利的血清稳定性或血浆半衰期,甚至在体内,(特别是在静脉内给药时)。

[0092] 在WO 2008/155134(如上文所讨论)中证实,包含具有能采取无规卷曲构象的氨基酸序列的结构域的生物活性蛋白具有增加的体内和/或体外稳定性。具体而言,WO 2008/155134所公开的无规卷曲结构域由脯氨酸、丙氨酸和丝氨酸(PAS)残基组成。该现有技术文件中描述了这三种残基的存在,作为在水性溶液中形成稳定和可溶性无规卷曲的基本要求。

[0093] 如上文的背景部分所讨论的,WO 2007/103515描述了无组织的重组聚合物,其包含多种氨基酸作为主要组成部分,特别是甘氨酸、天冬氨酸、丙氨酸、丝氨酸、苏氨酸、谷氨酸和脯氨酸。然而,相比于术语“生物合成”和“无规卷曲”,术语“无组织的重组聚合物”没有公认清楚含义。

[0094] 上文还描述了WO 2006/081249。该文件描述了蛋白缀合物,其包含与包含2至500个氨基酸重复单元的多肽偶联的生物活性蛋白,所述氨基酸重复具有作为主要组成部分的Gly、Asn和Gln和作为次要组成部分的Ser、Thr、Asp、Gln、Glu、His和Asn。与未缀合的生物活性蛋白相比,所述蛋白缀合物被描述为具有增加的或减少的血浆半衰期。然而,WO 2006/081249未提供任何教导来预测具体的氨基酸重复能减少还是增加缀合物的血浆半衰期。此外,WO 2006/081249未教导或暗示,当缀合的蛋白包含本发明所示的形成无规卷曲构象的氨基酸重复时,蛋白的血浆半衰期能增加。此外,WO 2006/081249所公开的氨基酸重复包含至少两种选自Gly、Asn和Gln的残基,这与本发明的生物合成的无规卷曲多肽明显不同,本发明的生物合成的无规卷曲多肽包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。

[0095] 令人意想不到的是,本申请发现,本文提供的生物合成的无规卷曲氨基酸序列,与现有技术相比,仅包含脯氨酸和丙氨酸残基(即,其优选不包含大量的任何其它氨基酸,也不包含大量的丝氨酸或根本不包含丝氨酸),并且也形成有效的无规卷曲结构。考虑到WO 2008/155134的公开内容,即,具有仅由丝氨酸和丙氨酸(SA)残基组成的结构域(即,缺少脯氨酸残基)的融合蛋白,本申请的发现是特别意想不到的,从而表明这样仅包含两种氨基酸的结构域不能形成无规卷曲,而是形成 β 折叠结构。这些丝氨酸-丙氨酸结构域也未表现出对于“PAS”所观察到的增加的流体动力学体积,或特别是未表现出对于本文提供的“P/A”序列所观察到的增加的流体动力学体积。

[0096] 本文所用的术语“生物活性”描述了某一物质对生命体的生物效应。因此,本文所用的术语“生物活性蛋白”涉及能在暴露于所述蛋白或多肽的活的细胞/生物体中诱导生物效应的蛋白。然而,应当注意到,对于本发明,术语“生物活性蛋白”涉及包含具有和/或介导所述生物活性的氨基酸序列(所述第一结构域)和能采取/形成无规卷曲构象并仅由脯氨酸和丙氨酸组成的本发明氨基酸序列(所述第二结构域)的本发明完整蛋白。

[0097] 因此,本文所用的术语“具有和/或介导生物活性的氨基酸序列”或“具有生物活性的氨基酸序列”涉及上文所限定的本发明的生物活性蛋白的“第一结构域”,其介导或具有或能介导或能具有上文所限定的“生物活性”。术语“具有和/或介导生物活性的氨基酸序列”或“具有生物活性的氨基酸序列”还包括在体内或体外半衰期需要被延长的任何感兴趣的蛋白(及其功能片段,例如抗体片段、包含膜受体的细胞外或细胞内结构域的片段、生长因子或细胞因子等的截短形式)。在本发明的一个实施方案中,本发明的具有和/或介导生

物活性的氨基酸序列可以来自任何“感兴趣的蛋白”，即，药学或生物学所关注的任何蛋白或可用作治疗剂/诊断剂的任何蛋白。

[0098] 因此，生物活性蛋白可以包含第一结构域，所述第一结构域包含来源于天然产生的多肽或通过重组DNA技术产生的多肽的生物活性氨基酸序列。在优选的实施方案中，感兴趣的蛋白可以选自结合蛋白/结合分子、免疫球蛋白、抗体片段、转运蛋白、膜受体、诸如细胞因子、生长因子、激素或酶等的信号蛋白/肽。

[0099] 如上文所解释的，包含于生物活性蛋白的第二结构域中的无规卷曲多肽(或多肽区段)形成无规卷曲构象，特别是在生理条件下。这对于可以形成待给予个体或患者的药物组合物的一部分的生物活性蛋白是特别重要的。

[0100] 应当注意到，本发明的生物活性蛋白的生物合成的无规卷曲结构域(所述“第二结构域”)天然地(即，在生理条件下)采取/形成/具有无规卷曲构象，特别是在体内以及给予需要医学干预的哺乳动物或人类患者时。相比之下，本领域已知的是，具有非随机二级结构和/或三级结构作为天然构象的蛋白倾向于在非生理条件下(即，在变性条件下)采取无规卷曲构象。然而，与包含本发明的无规卷曲多肽的生物活性蛋白相比，这样的变性蛋白具有完全不同的性质。因此，本发明的主旨是，本文提供的“生物活性蛋白”和融合蛋白/融合构建体的生物活性部分，当与本发明生物合成的无规卷曲多肽(或多肽区段)结合和/或连接时，也保持它们的生物功能。

[0101] 此外，无规卷曲多肽(或多肽区段)在生理条件下仍然保持可溶性。因此，还考虑到，本发明的蛋白构建体(包含上文所限定的“第一结构域”和“第二结构域”)可以包含暂时或暂且不处于无规卷曲构象的“第二”无规卷曲形成/采取结构域，例如，当处于诸如冻干的或干燥的组合物的某些组合物形式时。然而，重要的是，本发明蛋白构建体的这类“第二结构域”，当例如在相应的缓冲液(优选“生理”缓冲液/赋形剂和/或溶剂)中复原之后，能再次采取本文所限定的无规卷曲构象。所述“第二结构域”能(如果需要，则在相应的复原之后)介导本发明生物活性蛋白的增加的体内和/或体外稳定性。本文优选的是，本文所限定的“第二结构域”由本发明的无规卷曲多肽(或多肽区段)组成。

[0102] 本文所用的术语“结构域”涉及能自主采取具体结构和/或功能的任何氨基酸序列区域/部分。因此，对于本发明，“结构域”可以代表功能结构域或结构域。如本文所述，本发明的蛋白包含至少一个具有和/或介导生物活性的结构域/部分和至少一个形成无规卷曲构象的结构域/部分。然而，本发明的蛋白也可以由大于两个结构域组成，并且可以包含例如本文所限定的两个结构域/部分之间的额外的连接子或间隔子结构或另一结构域/部分，例如蛋白酶敏感切割位点、亲和标签(例如His₆-标签或Strep标签)、信号肽、停留肽、靶向肽(例如膜转位肽)或其它效应结构域，例如用于与抗肿瘤毒素相关的肿瘤靶向的抗体片段或用于前药活化的酶等。

[0103] 在另一个实施方案中，通过分析凝胶过滤(也称为尺寸排阻色谱，SEC)测定，本发明的生物活性蛋白具有的流体动力学体积为至少50kDa、优选至少70kDa、更优选至少80kDa、甚至更优选至少100kDa、特别优选至少125kDa，最优选至少150kDa。本领域技术人员能容易地测定具体蛋白的流体动力学体积。示例性的方法在上文关于无规卷曲多肽中已有描述。本领域技术人员还能根据本发明的生物活性蛋白容易地调整这些方法。如下文所述，相比于基于氨基酸残基的分子量或分子数/组成，对相应的折叠球形蛋白所估计出的流体

动力学体积,本发明的包含上文所限定的第二结构域(即,包含本文提供的无规卷曲多肽(或其区段)或由本文提供的无规卷曲多肽(或其区段)组成的结构域)的生物活性蛋白的流体动力学体积被证实具有意想不到的大流体动力学体积。

[0104] 应当注意到,包含“具有和/或介导生物活性的氨基酸序列”的第一结构域在另一多肽或氨基酸序列的情况下或与另一多肽或氨基酸序列结合之后还可以保留其生物活性。例如,抗体的Fab片段,例如抗肿瘤抗体赫塞汀(Herceptin)的一个Fab片段(Eigenbrot (1993) J.Mol.Biol.229:969-995),由两个不同多肽链组成,即免疫球蛋白轻链和免疫球蛋白重链的片段,它们还可以通过链间二硫键被连接起来。按照本发明,将那些链中的一个(例如通过基因融合)与无规卷曲多肽(或多肽区段)连接,同时通过与其它链结合来重建完整的生物活性蛋白就可以是足够的。例如,可以通过以下实现所述重建,如后文实施例所述的不同多肽(一方面是一条链与无规卷曲多肽的融合蛋白,另一方便是其它链)在相同宿主细胞的共表达,或体外重建,例如,作为再折叠方案的一部分。

[0105] 因此,这样的蛋白(包含两个单独的多肽链)也被视为本发明的生物活性蛋白。在这样的情况下,本文所限定的第一结构域可以包含仅非共价连接的两个单独的多肽链。此外,生物活性蛋白/结构域的单链可以各自都与无规卷曲多肽(或多肽区段)连接。除抗体片段之外,还有很多其它由数个相关的多肽链组成并适合于本发明的感兴趣的同寡聚蛋白或异寡聚蛋白(例如,胰岛素、血红蛋白等)。

[0106] 本文所用的术语“结合蛋白”涉及这样的分子,其能特异性地与可能的结合伴侣相互作用,从而能将所述可能的结合伴侣和作为所述可能结合伴侣的多个不同分子区别至这样的程度,即在所述作为可能的结合伴侣的多个不同分子的库中,仅有所述可能的结合伴侣是结合的或显著结合的。测定结合蛋白和可能的结合伴侣之间的结合的方法是本领域已知的,并且能常规地实施,例如,通过使用ELISA、等温滴定量热法、平衡透析法、拉下实验(pull down assay)、表面等离子体共振或Biacore装置。可用于本发明的示例性的结合蛋白/结合分子包括但不限于,抗体、抗体片段,例如Fab片段、F(ab')₂片段、单链可变片段(scFv)、抗体的分离的可变区(VL和/或VH区)、CDR、单结构域抗体/免疫球蛋白、CDR衍生的肽模拟物、凝集素、免疫球蛋白结构域、纤连蛋白结构域、蛋白A结构域、SH3结构域、锚蛋白重复结构域、脂质运载蛋白(lipocalin)或各种支架衍生的结合蛋白,例如,如Skerra (2000) J Mol Recognit 13:167-187, Gebauer (2009) Curr Opin Chem Biol 13:245-255, Binz (2005) Nat Biotechnol 23:1257-1268或Nelson (2009) Nat Biotechnol 27:331-337中所述。

[0107] 可用于本发明的其它示例性的感兴趣的生物活性蛋白(特别是生物活性蛋白的第一结构域中包含的或组成/作为生物活性蛋白的第一结构域的蛋白)包括但不限于,粒细胞集落刺激因子、人生长激素、 α -干扰素、 β -干扰素、 γ -干扰素、 λ -干扰素、肿瘤坏死因子、红细胞生成素、凝血因子(例如凝血因子VIII、凝血因子VIIa、凝血因子IX)、gp120/gp160、可溶性肿瘤坏死因子I和II受体、溶栓剂(例如瑞替普酶)、具有代谢效应的肽(例如GLP-1或艾塞那肽-4)、免疫抑制/免疫调节蛋白(例如白介素-1受体拮抗剂或阿那白滞素、白介素-2和中性粒细胞明胶酶相关脂质运载蛋白)或其它天然的或工程化的脂质运载蛋白或例如, Walsh (2003) Nat Biotechnol 21:865-870或Walsh (2004) Eur J Pharm Biopharm 58:185-196列出的蛋白或化合物,或诸如<http://www.biopharma.com/approvals.html>或

<http://www.drugbank.ca>的在线数据库列出的蛋白或化合物。可用于本发明的其它生物活性蛋白(特别是生物活性蛋白的第一结构域中包含的或组成/作为生物活性蛋白的第一结构域的蛋白)特别是卵泡刺激素、葡糖脑苷脂酶、胸腺素 $\alpha 1$ 、胰高血糖素、生长激素抑制素、腺苷脱氨酶、白介素11、hematide、瘦素、白介素-20、白介素-22受体 α 亚基(IL-22ra)、白介素-22、透明质酸酶、成纤维细胞生长因子18、成纤维细胞生长因子21、胰高血糖素样肽1、骨保护素、IL-18结合蛋白、生长激素释放因子、可溶性TACI受体、血小板反应蛋白-1、可溶性VEGF受体Flt-1、 α -半乳糖苷酶A、肌骨素拮抗剂、抑胃多肽、 α -1抗胰蛋白酶、IL-4突变蛋白等。从本文公开的内容可以显而易见的是,本发明还涉及包含生物合成的无规卷曲脯氨酸/丙氨酸多肽或脯氨酸/丙氨酸多肽区段和药学上或医学上有用的分子,例如小分子、肽或生物大分子,例如蛋白、核酸、碳水化合物、脂质囊泡等,特别是药学上或医学上有用的蛋白,例如(但不限于)结合蛋白/结合分子、免疫球蛋白、抗体片段、转运蛋白、膜受体、信号蛋白/肽、细胞因子、生长因子、激素或酶等,并且它们可以被包含在本文所限定的药物构建体中,但它们还可以是本文所限定的包含所限定的至少两个结构域或由所限定的至少两个结构域组成的生物活性异源蛋白的一部分。在这样的情况下,所述具体的药学上或医学上有用的蛋白(或其功能片段)可以是包含具有和/或介导所述生物活性的氨基酸序列或由具有和/或介导所述生物活性的氨基酸序列组成的所述至少两个结构域的“第一结构域”。在这样的情况下,功能片段是所述药学上或医学上有用的蛋白的片段,所述片段仍能在体内和/或体外引发所需的生物学或药学应答和/或仍具有或介导所需的生物活性。

[0108] 在所述第一和所述第二结构域之间插入的上述多肽连接子/间隔子优选包含多个亲水性、肽键连接的氨基酸,其与两个结构域共价连接。在另一个实施方案中,所述多肽连接子/间隔子包含血浆蛋白酶切割位点,其允许包含具有和/或介导生物活性的多肽的所述第一结构域的受控释放。可以容易地鉴定出不同类型或长度的连接子,而获得具体蛋白的最佳生物活性。

[0109] 在优选的实施方案中,本发明的生物活性蛋白是融合蛋白。本文所述的融合蛋白可以在单个多结构域多肽中包含至少一个能介导生物活性的结构域和至少一个包含本文所述的生物合成的无规卷曲多肽(或多肽区段)的其它结构域。此外,应当注意到,本发明不限于其中一个结构域介导生物活性的融合蛋白。本文还提供了其它“融合蛋白”/“融合构建体”,其中一个部分/结构域是本发明的脯氨酸/丙氨酸无规卷曲多肽/聚合物或包含本发明的脯氨酸/丙氨酸无规卷曲多肽/聚合物,并且其它部分/结构域包含另一蛋白段/结构。

[0110] 特别是,对于融合蛋白,本发明的无规卷曲多肽(或多肽区段)不一定在其氨基或羧基末端携带Pro或Ala残基。在可选实施方案中,本发明的生物活性蛋白可以代表蛋白缀合物,其中感兴趣的蛋白或具有和/或介导生物活性的多肽/多肽段/肽/氨基酸序列通过非肽键与形成/采取无规卷曲构象的氨基酸序列、特别是本文提供的且仅由脯氨酸和丙氨酸残基组成的无规卷曲多肽(或多肽区段)缀合。可用于将蛋白与包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列的生物合成的无规卷曲多肽或多肽区段进行交联的非肽键是本领域已知的,其中所述氨基酸序列由本文提供的至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。所述非肽键可以包括例如Cys侧链之间的二硫键、硫醚键或由化学交联子产生的非肽共价键,例如双琥珀酰亚胺辛二酸酯(DSS)或磺基琥珀酰亚胺4-[p-马来酰亚胺苯基]丁酸酯(硫代SMPB)、金属螯合/复合基团,以及非共价蛋白-蛋白相互作用。

[0111] 应当注意到,本发明的“生物活性蛋白”还可以包含多于一个“具有和/或介导生物活性的氨基酸序列”。此外,生物活性蛋白还可以包含大于生物合成的无规卷曲多肽(或其区段)。在最简单的情况下,生物活性蛋白由两个结构域组成,即,包含具有和/或介导生物活性的氨基酸序列的第一结构域和包含生物合成的多肽(或其区段)的第二结构域。应当注意到,本发明不限于与本文公开的生物合成的无规卷曲多肽或多肽区段连接的“生物学上或治疗上的活性蛋白”,所述生物合成的无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。此外,在例如食品或饮料工业、美容用品工业等的其它工业中有重要作用其它感兴趣的蛋白或分子可以通过本文提供的方式和方法来制备。

[0112] 本领域技术人员知晓,可以将“包含具有和/或介导生物活性的氨基酸序列的结构域”和“包含在本发明的生物活性蛋白中包含的无规卷曲多肽(或其区段)的第二结构域”以具体的顺序组织在一起。

[0113] 因此,在本发明中,本发明生物活性多肽的所述“第一”和“第二”结构域的顺序可以按顺序安排,由此所述“第一结构域”(即,感兴趣的蛋白;“具有和/或介导所述生物活性的氨基酸序列”)位于生物活性蛋白的氨基(N-)末端,且所述“第二结构域”(即,包含本文提供的无规卷曲多肽(或其区段)的结构域)位于生物活性蛋白的羧基(C-)末端。然而,该顺序还可以颠倒,例如所述“第一结构域”(即,感兴趣的蛋白;“具有和/或介导所述生物活性的氨基酸序列”)位于生物活性蛋白的羧基(C-)末端,所述“第二结构域”(即,包含本文提供的无规卷曲多肽(或其区段)的结构域)位于生物活性蛋白的氨基(N-)末端。如果生物活性蛋白仅由一个第一结构域和一个第二结构域组成,则结构域顺序因此可以为(从N末端至C末端):第一结构域(具有和/或介导生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段))。反之亦然,结构域顺序可以为(从N末端至C末端):第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导生物活性的氨基酸序列)。

[0114] 还考虑到,多于一个包含具有和/或介导所述生物活性的氨基酸序列或由具有和/或介导所述生物活性的氨基酸序列组成的结构域可用于本发明的蛋白构建体。例如,生物活性蛋白可以包含两个“第一结构域”,即,两个具有和/或介导生物活性的具体氨基酸序列,由此这种生物活性可以是相同的或不同的活性。如果生物活性蛋白由两个这样的“第一结构域”即两个具有和/或介导生物活性的具体氨基酸序列和一个“第二结构域”(包含生物合成的无规卷曲多肽(或其区段)组成,则结构域顺序可以为(从N末端至C末端):第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体(任选地不同)生物活性的氨基酸序列)。

[0115] 相同的解释适用于生物活性蛋白包含多于一个“第二结构域”(即,生物活性蛋白包含多于一个无规卷曲多肽(或其区段)的情况。如果生物活性蛋白由两个这样的“第二结构域”即两个包含生物合成的无规卷曲多肽(或其区段)的结构域和一个“第一结构域”(包含具有和/或介导生物活性的氨基酸序列)组成,则结构域顺序可以为(从N末端至C末端):第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段))。如果生物活性蛋白包含多于一个“第二结构域”,本文考虑到,这些“第二结构域”可以是相同的或可以是不同的。

[0116] 如上文所述,生物活性蛋白可以包含多于一个“第一结构域”,即,多于一个具有

和/或介导生物活性的具体氨基酸序列,和多于一个“第二结构域”(生物合成的无规卷曲多肽(或其区段)),其中这些“第一结构域”可以是相同的或不同,和/或其中所述“第二结构域”可以是相同的或不同。在这些情况下,下述示例性的结构域顺序是可能的(从N末端至C末端):

[0117] -第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段));

[0118] -第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段));

[0119] -第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段))-第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体生物活性的氨基酸序列);

[0120] -第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体生物活性的氨基酸序列);

[0121] -第二结构域(无规卷曲多肽(或其区段))-第二结构域(无规卷曲多肽(或其区段))-第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第一结构域(具有和/或介导具体生物活性的氨基酸序列);或

[0122] -第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第一结构域(具有和/或介导具体生物活性的氨基酸序列)-第二结构域(无规卷曲多肽(或其区段))-第二结构域(无规卷曲多肽(或其区段))。

[0123] 对于本领域技术人员,其它相应的结构域顺序(特别是生物活性蛋白中包含大于两个“第一结构域”或“大于两个“第二结构域”的情况下)是容易想到的。

[0124] 对于本发明多肽/生物活性蛋白的所有实施方案,包含具有和/或介导所述生物活性的氨基酸序列的所述结构域,还可以是具有所需生物功能的给定蛋白的生物活性片段。因此,本文所限定的“第二结构域”(优选包含本文提供的无规卷曲多肽(或其区段))还可以位于感兴趣的蛋白的两个生物活性片段之间或两个感兴趣的蛋白的生物活性片段之间。在必要的变通下,上文对于“全长”感兴趣的蛋白/多肽(即,当氨基酸序列自身具有/介导某一生物活性时)给出的所有解释和限定适用于这样的片段。

[0125] 此外,上述发明不限于包含具有“生物活性功能”的“结构域”的构建体。本发明的构建体还可以包含具有其它功能的结构域,并且不限于生物活性。这些仅是本发明的实施方案,并且对于本领域技术人员显而易见的是,在不脱离本发明的主旨的情况下,能容易地制备和使用其它构建体。因此,在必要的变通下,本文关于“具有和/或介导具体生物活性的氨基酸序列”的所述适用于其它构建体,例如可用于其它技术领域的构建体,例如美容用品、食品加工、乳制品、造纸工业等。如上文所述,本发明的生物合成的多肽/聚合物也可以用于与例如小分子等连接。

[0126] 此外,应当指出,术语“具有和/或介导第一生物活性的氨基酸序列”不限于具有和/或介导所述生物活性或功能的全长多肽,还包括它的生物学和/或药理学活性片段。特

别是(但不仅如此),在本发明的“生物活性蛋白”中包含两个或更多个本文所限定的“第一结构域”时,还考虑到,这些“第一结构域”是或代表蛋白复合物的不同的部分或蛋白复合物的该部分的片段。

[0127] 如下文所例示,与缺少无规卷曲结构域的未修饰的生物活性蛋白相比,经过修饰而包含所述无规卷曲多肽的本发明的生物活性蛋白意外地表现出增加的体内和/或体外稳定性。本文所用的术语“体内稳定性”涉及被给予活体的具体物质保持生物可利用性和生物活性的能力。在体内,由于排泄、肾过滤、肝摄取、聚集、降解和/或其它代谢过程,物质可以被移除和/或失活。因此,对于本发明,具有增加的体内稳定性的生物活性蛋白通过肾(尿)或通过排泄物排泄的速度可以较慢,和/或可以更稳定对抗蛋白水解,特别是对抗生物液体(例如血液、脑脊液、腹膜液和淋巴液)中的体内蛋白水解。在一个实施方案中,生物活性蛋白的增加的体内稳定性表现为所述生物活性蛋白的延长的血浆半衰期。具体而言,与缺少第二结构域的生物活性蛋白相比,生物活性蛋白的增加的体内稳定性是包含所述第二结构域的所述生物活性蛋白的延长的血浆半衰期。

[0128] 测量生物活性蛋白的体内稳定性的方法是本领域已知的。如下文所例示,利用Western印迹技术或酶连接免疫吸附测定(ELISA),可以在血浆中特异性地检测生物活性蛋白。然而,本领域技术人员知晓,其它方法可以用于特异性地测定感兴趣的蛋白的血浆半衰期。这些方法包括但不限于放射性标记的感兴趣的蛋白的物理学检测。放射性标记蛋白方法(例如通过放射性碘化)是本领域已知的。

[0129] 本文所用的术语“增加的体外稳定性”涉及生物活性蛋白在体外环境中抵抗降解和/或聚集并保持其原始生物活性的能力。测量生物活性蛋白的生物活性的方法是本领域公知的。

[0130] 此外,提供了药物缀合物,其包含本文所述和限定的无规卷曲多肽或多肽区段和小分子药物,所述小分子药物与所述无规卷曲多肽或多肽区段缀合。小分子的非限制性实例是地高辛配基、荧光素、多柔比星、卡里奇霉素、喜树碱、烟曲霉素、地塞米松、格尔德霉素、紫杉醇、多西他赛、伊立替康、环孢霉素、丁丙诺啡、纳曲酮、纳洛酮、长春地辛、万古霉素、利培酮、阿立哌唑、帕洛诺司琼、格拉司琼、阿糖胞苷、NX1838、亮丙瑞林、戈舍瑞林、布舍瑞林、奥曲肽、替度鲁肽、西仑吉肽、阿巴瑞克、恩夫韦地、葛瑞林和衍生物、tubulysins、铂衍生物、 $\alpha 4$ 整合素抑制剂、反义核酸、小干扰RNA、微RNA、类固醇、DNA或RNA适体、肽、肽模拟物。通常,本发明还涉及药物构建体,所述药物构建体包含本文所限定的无规卷曲多肽或多肽区段和特别是药学上或医学上有用的分子,例如小分子、肽或生物大分子,例如蛋白、核酸、碳水化合物、脂质囊泡等。在后文的示例性实验部分(参见,例如实施例22)中,成功地制备了本发明的构建体/缀合物,在构建体中,“小化学分子”与本文公开的无规卷曲多肽缀合。因此,本发明的附图和相应的图例中的实验信息提供了说明性的示例,其中本文公开的药物缀合物包含(i)生物合成的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成,和(ii)小分子,作为说明,所述小分子选自地高辛配基和荧光素。应当注意到,这些不是仅有的理论实例。荧光素或荧光素衍生物常用作诊断制剂,并且医学荧光素溶液以商品名Fluorescein®、AK-FLUOR®或Fluorescein®销售。这些化合物显然等够受益于本文提供的方式和方法。地高辛配基形成地高辛的类固醇部

分,地高辛是公知的具有心脏活性功能的植物次生代谢产物,其还含有三个洋地黄毒糖。地高辛,以及在较小的程度上说,密切相关的化合物洋地黄毒苷,被广泛用于治疗室性快速型心律失常和充血性心力衰竭(Hauptman (1999) *Circulation* 99:1265-1270)。所有具有心脏活性的类固醇是位于细胞质膜中的 Na^+/K^+ -ATP酶的强力且高度特异性的抑制剂,因此发挥交感神经阻滞作用或正性肌力作用。

[0131] 在必要的变通下,上文关于无规卷曲多肽或其多肽区段给出的限定和解释也适用于这样的药物缀合物,其包含无规卷曲多肽(或其多肽区段)和选自以下的药物:(a)包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽和(b)小分子药物。

[0132] 本文所限定和提供的形成无规卷曲构象的氨基酸聚合物/无规卷曲多肽(或其区段)可以与小分子/小分子药物缀合。由此,小分子/小分子药物的血浆半衰期和/或可溶性可以增加,非特异性毒性可以降低,并且体内的目标细胞或结构暴露于活性药物的时间延长可以实现药效增强。

[0133] 无规卷曲多肽的N末端与活化的药物衍生物的位点特异性缀合,例如N-羟基琥珀酰亚胺(NHS)酯衍生物(Hermanson (1996) *Bioconjugate Techniques*, Academic Press, San Diego, CA),是可能的。通常,N末端氨基可以与多种官能团化学偶联,例如醛和酮(从而形成席夫碱,例如,利用硼氢化钠或氰基硼氢化钠可以将其还原为胺)或活化的碳酸衍生物(酸酐、氯化物、酯等,从而形成酰胺)或其它反应性化学物,例如异氰酸酯、异硫氰酸酯、磺酰基氯化物等。另外,可以首先用合适的保护基修饰氨基酸聚合物/多肽的N末端,例如乙酰基、BOC基团或FMOC基团(Jakubke (1996) *Peptide. Spektrum Akademischer Verlag*, Heidelberg, Germany)。此外,可以用焦谷氨酰基保护氨基末端,焦谷氨酰基能形成自Pro/Ala多肽或多肽区段之前的编码的Gln氨基酸残基。在例如利用常用试剂EDC(N-(3-二甲基氨基丙基)-N'-乙基碳化二亚胺)和NHS使C末端羧基团活化后,如果药物携带例如自由氨基,则可以实现与受保护的无规卷曲多肽的C末端的位点特异性偶联。

[0134] 或者,可以用能提供马来酰亚胺基团的可商购的连接试剂来修饰形成无规卷曲构象的氨基酸聚合物/无规卷曲多肽的N末端或C末端,从而允许与作为药物分子的一部分的硫醇基化学偶联。按照该方式,可容易地获得均一的药物缀合物。本领域公知的类似技术(Hermanson (1996) 同上)可用于将无规卷曲多肽与肽或甚至与蛋白药物偶联。可以容易地制备携带Lys或Cys侧链的这些肽或蛋白,这允许它们通过NHS酯或马来酰亚胺活性基团与形成无规卷曲构象的氨基酸聚合物进行体外偶联。通常,用包含无规卷曲多肽(或其区段)的融合蛋白可以制备相似的药物缀合物。然而,如后文实施例和附图所示,本发明还提供了本发明的创新缀合物中包含的无规卷曲多肽或无规卷曲多肽区段的制备方式。

[0135] 作为单一位点特异性缀合的可选方案,可以使无规卷曲多肽在N末端或C末端或内部装配适于化学修饰的额外的侧链,例如,具有 ϵ -氨基的赖氨酸残基、具有硫醇基团的半胱氨酸残基或甚至是非天然氨基酸,从而允许利用例如NHS酯或马来酰亚胺活性基团实现多个小分子的缀合。

[0136] 除稳定的缀合之外,可以使前药与无规卷曲多肽暂时连接。可以将连接设计为能以可预见的方式在体内被切割,所述切割可以通过酶学机制或通过生理pH下开始的缓慢水解,这与例如低可溶性抗肿瘤剂喜树碱与PEG聚合物缀合是相似的,从而实现生物分布增

加、毒性减小、功效和肿瘤累积增强 (Conover (1998) Cancer Chemother Pharmacol, 42: 407-414)。其它前药的实例是化疗剂, 例如多西他赛 (Liu (2008) J Pharm Sci. 97: 3274-3290)、多柔比星 (Veronese (2005) Bioconjugate Chem. 16: 775-784) 或紫杉醇 (Greenwald (2001) J Control Release 74: 159-171)。

[0137] 此外, 小分子可以与融合蛋白偶联, 所述融合蛋白包含与诸如抗体片段的靶向结构域基因融合的氨基酸聚合物/多肽, 从而实现小分子药物的特异性递送。对于后一种情况, 例如, 如果靶向结构域是针对经历内化的细胞表面受体, 可以通过与细胞毒性小分子缀合容易地制备免疫毒素。

[0138] 按照上文, 本发明因此还涉及提供本文公开的生物合成的无规卷曲多肽或多肽区段用于与其它所选化合物进行进一步的和其它的偶联, 所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。所述进一步的和其它的偶联可以是和/或可以包含所述生物合成的无规卷曲多肽或生物合成的无规卷曲多肽区段与另一化合物的首先偶联。

[0139] 在另一个实施方案中, 本发明涉及编码本文所述的无规卷曲多肽(或其区段)或生物活性蛋白的核酸分子。因此, 所述核酸分子可以包含编码具有生物活性的多肽的核酸序列和编码无规卷曲多肽(或其区段)的核酸序列。本文所用的术语“核酸分子”意图包括诸如DNA分子和RNA分子的核酸分子。所述核酸分子可以是单链的或双链的, 但优选双链DNA。优选地, 所述核酸分子可以被包含在载体中。

[0140] 因此, 本发明还涉及编码本文提供的缀合物(例如本文所限定的药物缀合物)中包含的无规卷曲多肽或多肽区段的核酸分子, 或编码包含上文所限定的生物活性蛋白并且还包含生物合成的无规卷曲多肽或多肽区段的蛋白缀合物的核酸分子, 所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列, 其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。

[0141] 在一个实施方案中, 提供了编码缀合物(例如, 上文所限定的药物缀合物或食品缀合物)的核酸分子, 所述核酸分子包含:

[0142] (i) 编码翻译的氨基酸和/或前导序列的核酸序列;

[0143] (ii) 编码生物合成的无规卷曲多肽或多肽区段的核酸序列, 所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列, 其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成;

[0144] (iii) 编码包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的所述生物活性蛋白或多肽或感兴趣的蛋白的核酸序列, 所述感兴趣的蛋白例如是可用于诸如食品工业的其它工业领域的蛋白; 和

[0145] (iv) 提供翻译终止密码子或是翻译终止密码子的核酸序列。

[0146] 例如, (i) 中的上述“翻译的氨基酸和/或前导序列”可以是起始“M”, 即, 来自相应的起始密码子的蛋氨酸, 其还可以包含mRNA的非翻译序列, 例如5'序列直至起始密码子, 非翻译序列例如包含核糖体结合位点。然而, 这样的序列还可以包含典型前导和/或信号序列, 例如用于使表达的蛋白分泌到周质或培养基中。原核信号肽例如是OmpA、MalE、PhoA、DsbA、pelB、Afa、npr、STII。真核信号肽例如是蜂素信号序列、酸性糖蛋白gp67信号序列、小鼠IgM信号序列、hGH信号序列。

[0147] 包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽以及其它感兴趣的蛋白(例如可用于其它工业领域的蛋白)已在上文提供。在必要的变通下,所述实施方案适用于上文所述的核酸分子(部分/区段(iii))。

[0148] 可用于本文提供的核酸分子中的翻译终止密码子是本领域公知的,例如是密码子UAA、UAG或UGA。

[0149] 在上文提供的核酸分子的一个实施方案中,上述核酸分子部分/区段(ii)和(iii)在编码缀合物(例如药物缀合物或食品缀合物)的所述核酸分子中的位置互换。这样的核酸分子会包含下述顺序的部分/区段:

[0150] (i) 编码翻译的氨基酸和/或前导序列的核酸序列;

[0151] (ii) 编码包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列的生物活性蛋白或多肽或感兴趣的蛋白的核酸序列,所述感兴趣的蛋白例如是可用于诸如食品工业的其它工业领域的蛋白;

[0152] (iii) 编码生物合成的无规卷曲多肽或多肽区段的核酸序列,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成;以及

[0153] (iv) 代表翻译终止密码子或是翻译终止密码子的核酸序列。

[0154] 上文提供的核酸分子还可以任选地包含在部分/区段(i)和(ii)之间和/或在部分/区段(ii)和(iii)之间包含蛋白酶和/或化学切割位点和/或识别位点。这些化学切割位点是本领域公知的,并且包括例如特定的个体氨基酸序(参见,例如Lottspeich and Engels(Hrsg.) (2006) Bioanalytik. 2. Auflage. Spektrum Akademischer Verlag, Elsevier, München, Germany)。例如,溴化氰或氯化氰切割Met残基后的肽键;羟基野芝麻碱切割天冬酰胺-甘氨酸键;甲酸切割Asp-Pro;2-(2'-硝基苯基亚氧硫基)-3-甲基-3-溴吡啶,2-碘苯甲酸或N-氯琥珀酰亚胺切割Trp之后的键;2-硝基-5-硫代羧酰苯甲酸切割Cys之后的键。本文还考虑到并且也可能的是,Pro/Ala多肽或多肽区段之前的残基可以通过定点诱变被取代为Met,并且然后所得的融合蛋白可以被BrCN切割。相似地,可通过定点诱变将包含切割位点的其它氨基酸序列引入重组融合蛋白或其编码核酸。

[0155] 其它有用的蛋白酶识别/切割位点是本领域已知的。这些包括但不限于:胰蛋白酶、糜蛋白酶、肠激酶、烟草蚀纹病毒(TEV)蛋白酶、PreScission蛋白酶、HRV 3C蛋白酶、SUMO蛋白酶、分拣酶A、颗粒酶B、弗林蛋白酶、凝血酶、因子Xa或自我可切割的内含肽。因子Xa在氨基酸序列IleGluGlyArg的C末端水解肽键,氨基酸序列IleGluGlyArg可以被插在N末端融合伴侣和Pro/Ala多肽或多肽区段之间。实现蛋白水解切割的特别简单的方法是,通过在Pro/Ala多肽或多肽区段的N末端插入或取代Lys或Arg侧链,然后用胰蛋白酶进行消化,胰蛋白酶不会切割Pro/Ala多肽或多肽区段的内部,只要在内部避免Lys或Arg侧链。示例性的识别位点可以是,但不限于,D-D-D-D-K(肠激酶)、ENLYFQ(G/S)(TEV蛋白酶),I-(E/D)-G-R(因子Xa)、L-E-V-L-F-Q-G-P(HRV 3C)、R-X-(K/R)-R(弗林蛋白酶)、LPXTG(分拣酶A)、L-V-P-R-G(凝血酶)或I-E-X-D-X-G(颗粒酶B)。

[0156] 从上文的公开内容可以明显看出,本发明提供了重组产生的生物合成的无规卷曲多肽和多肽区段,其可以与所选的分子缀合,所述所选的分子例如是有用的蛋白、药学活性多肽或小分子、诊断上有用的多肽或小分子,或者其它工业领域(例如食品或造纸工业或采

油)的其它有用的蛋白或小分子。因此,本发明还提供了编码生物合成的无规卷曲多肽或多肽区段的核酸,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成,所述核酸分子包含:

[0157] (i) 编码翻译的氨基酸和/或前导序列的核酸序列;

[0158] (ii) 编码所述生物合成的无规卷曲多肽或多肽区段的核酸序列,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列;和

[0159] (iii) 代表翻译终止密码子或是翻译终止密码子的核酸序列。

[0160] 上述核酸分子可以任选地在(i)和(ii)之间包含蛋白酶和/或化学切割位点和/或识别位点)。

[0161] 此外,对于上述核酸分子,上文关于前两种核酸分子(即,蛋白酶和/或化学切割位点和/或识别)提供的实施方案在必要的变通下适用于此处。

[0162] 可用于本发明的有用的和示例性的信号序列包括,但不限于,原核序列,例如OmpA、MalE、PhoA、DsbA、pelB、Afa、npr、STII,或真核序列,例如蜂素信号序列、酸性糖蛋白gp67信号序列、小鼠IgM信号序列、hGH信号序列

[0163] 具体而言,编码本发明的生物合成的无规卷曲多肽或多肽区段的核酸分子可用于下文提供的以及后文实施例和附图所示的方法,所述生物合成的无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列。这样的表达的无规卷曲多肽或无规卷曲多肽区段可以从例如表达该无规卷曲多肽或无规卷曲多肽区段的宿主细胞分离。这类宿主细胞可以是例如用本文提供的载体转染的细胞。

[0164] 因此,考虑到用本文所述的核酸分子或载体转染细胞。在另一实施方案中,本发明涉及这样的核酸分子,所述核酸分子在表达时能编码本发明的无规卷曲多肽(或其区段)或生物活性蛋白。然而,在另一实施方案中,本发明涉及这样的核酸分子,所述核酸分子在表达时能编码本文公开的在水性溶液中或在生理条件下完全地或部分地形成/采取无规卷曲构象的多肽。所述核酸分子可以与本领域已知的合适的表达控制序列融合,从而确保多肽以及信号序列的合适的转录和翻译,进而确保细胞分泌或靶向细胞器。这类载体还可以包含其它基因,例如允许所述载体在合适的宿主细胞中和在合适的条件下的选择的标记基因。

[0165] 优选地,本发明的核酸分子被包含在重组载体中,在所述重组载体中,编码本文所述的生物活性蛋白的核酸分子与允许在原核或真核细胞中表达的表达控制序列可操作地连接。所述核酸分子的表达包括核酸分子转录为可翻译的mRNA。允许原在核宿主细胞中表达的调节元件包括,例如,大肠杆菌中的 λ PL、lac、trp、tac、ara、phoA、tet或T7启动子。确保在真核细胞、优选哺乳动物细胞或酵母中表达的可能的调节元件是本领域技术人员公知的。它们通常包括确保启动转录的调节序列以及任选的影响转录终止和转录物稳定化的多聚腺苷酸(polyA)信号。其它调节元件可以包括转录增强子以及翻译增强子,和/或天然相关的启动子区或异源启动子区。允许在真核宿主细胞中表达的调节元件的实例是酵母中的AOX1或GAL1启动子或CMV、SV40、RSV(劳斯氏肉瘤病毒)启动子、CMV增强子、SV40增强子或哺乳动物和其它动物细胞中的球蛋白内含子。除了负责启动转录的元件,所述调节元件还可以包括编码区下游的转录终止信号,例如SV40多聚腺苷酸位点或tk多聚腺苷酸位点。

[0166] 可使用本领域技术人员公知的方法来构建重组载体(参见,例如,Sambrook (1989),Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory NY 和Ausubel (1989),Current Protocols in Molecular Biology,Green Publishing Associates and Wiley Interscience,NY中描述的技术)。在本发明中,合适的表达载体是本领域已知的,例如Okayama-Berg cDNA表达载体pcDV 1(Pharmacia)、pCDM8、pRc/CMV、pcDNA1、pcDNA3、pPICZαA (Invitrogen)或pSPORT 1(GIBCO BRL)。此外,根据所用的表达体系,可以将能将多肽引导至细胞区室或将多肽分泌到培养基中的前导序列添加到本发明的核酸分子的编码序列。

[0167] 本发明还涉及常规用于基因工程的载体,特别是质粒、粘粒、病毒和噬菌体,所述载体包含编码本发明的无规卷曲多肽(或其区段)或生物活性蛋白的核酸分子。优选地,所述载体是表达载体和/或基因转移或靶向载体。来自诸如逆转录病毒、痘病毒、腺相关病毒、疱疹病毒或牛乳头状瘤病毒等病毒的表达载体可以用于将本发明的多核苷酸或载体递送到靶细胞群中。

[0168] 通过公知的方法可以将含有本发明的核酸分子的载体转染到宿主细胞中,所述方法根据细胞类型会有不同。因此,本发明还涉及包含所述核酸分子或所述载体的细胞。这类方法,例如,包括Sambrook (1989),同上和Ausubel (1989),同上,中描述的技术。因此,氯化钙转染或电穿孔常用于原核细胞,而磷酸钙处理或电穿孔可以用于其它细胞宿主(Sambrook (1989),同上)。作为另一可选方案,可以将本发明的核酸分子和载体包装在脂质体中,用于递送至靶细胞。存在于宿主细胞中的本发明的核酸分子或载体可以被整合到宿主细胞的基因组中或可以保持在染色体外。因此,本发明还涉及包含本发明的核酸分子和/或载体的宿主细胞。用于表达多肽的宿主细胞是本领域公知的,并包括原核细胞以及真核细胞,例如大肠杆菌细胞、酵母细胞、无脊椎动物细胞、CHO细胞、CHO-K1细胞、HEK 293细胞、HeLa细胞、COS-1猴细胞、黑素瘤细胞(例如Bowes细胞)、小鼠L-929细胞、来自瑞士、Balb-c或NIH小鼠的3T3细胞系、BHK或HaK仓鼠细胞系等。

[0169] 另一方面,本发明包含制备本发明的缀合物以及本文提供的生物合成的无规卷曲多肽(或其区段)或生物活性蛋白的方法,并包括培养本发明的(宿主)细胞以及从本文所述的培养物分离所述无规卷曲多肽(或其区段)或缀合物或生物活性蛋白。通常,可以通过重组DNA技术产生本发明的无规卷曲多肽(或其区段)、包含无规卷曲结构域的缀合物或生物活性蛋白,例如,通过培养包含编码本发明的生物活性蛋白或无规卷曲多肽(或其区段)的所述核酸分子或载体的细胞,以及从培养物分离所述蛋白/多肽。可以在任何合适的细胞培养体系中产生本发明的生物活性蛋白或无规卷曲多肽(或其区段),包括原核细胞,例如大肠杆菌BL21、KS272或JM83,或真核细胞,例如巴斯德毕赤酵母、酵母株X-33或CHO细胞。本领域已知的其它合适的细胞系可获自细胞系保藏单位,例如美国典型培养物保藏中心(ATCC)。

[0170] 术语“原核”意图包括细菌细胞,而术语“真核”意图包括酵母、高等植物、昆虫和哺乳动物细胞。可以使转化的宿主在发酵罐中生长并按照本领域已知的技术进行培养,从而实现最佳的细胞生长。在另一实施方案中,本发明涉及制备上文所述的无规卷曲多肽(或其区段)或生物活性蛋白的方法,包括在适于生物活性蛋白或无规卷曲多肽(或其区段)的表达的条件下培养本发明的细胞,以及从细胞或培养基分离所述蛋白/多肽。

[0171] 本发明的无规卷曲多肽(或其区段)本身优选不包含任何化学反应基团,然而,也可能有一个N末端伯胺基团(或,对于脯氨酸为仲胺)和位于聚合物的C末端的一个羧酸基团。然而,对于本领域技术人员显而易见的是,本文提供的生物合成的无规卷曲多肽/聚合物可以包含化学反应基团,例如,当所述无规卷曲多肽/聚合物是“融合蛋白”/“融合构建体”的一部分时。还如上文所述,可以按照本领域技术人员公知的方法,以数种方式通过转化的细胞中的重组表达来制备生物合成的无规卷曲多肽(或其区段),例如:(i)借助于N末端Met残基/起始密码子在细胞质中直接表达;(ii)通过N末端信号肽(例如OmpA、PhoA (Monteilhet (1993) Gene. 1993125:223-228)、蜂毒肽 (Tessier (1991) Gene 98:177-183)、白介素2 (Zhang (2005) J Gene Med 7:354365)、hGH (Pecceu (1991) Gene 97 (2):253-258)等)分泌,然后是细胞内切割,从而得到成熟的N末端,例如Ala或Pro;(iii)表达为与另一可溶蛋白的融合蛋白,例如,麦芽糖结合蛋白位于N末端并具有散布的蛋白酶切割位点 (Kapust and Waugh (2000) Protein Expr. Purif. 19:312-318),然后是体外或体内特异性蛋白酶切割,从而释放具有成熟N末端(例如Ala或Pro)的氨基酸聚合物/多肽。另一合适的融合伴侣是SUMO蛋白,其可以被SUMO蛋白酶切割,正如实施例20和21所述。其它融合伴侣包括在不限于,谷胱甘肽-S-转移酶、硫氧还蛋白、纤维素结合结构域、白蛋白结合结构域、荧光蛋白(例如GFP)、蛋白A、蛋白G、内含肽等 (Malhotra (2009) Methods Enzymol. 463:239-258)。

[0172] 如上文所解释的,所述的无规卷曲多肽(或多肽区段)/聚合物主要由丙氨酸和脯氨酸残基组成,而O-糖基化或N-糖基化所需的丝氨酸、苏氨酸或天冬酰胺优选不存在。因此,产生多肽本身或产生包含无规卷曲多肽(或其多肽区段)的生物活性蛋白或通常产生包含无规卷曲多肽(或其多肽区段)的融合蛋白,令人意想不到地能够产生单分散的产物,所述产物优选在Pro-Ala序列中没有翻译后修饰。这对于真核细胞中的重组蛋白产生是特别有利的,所述真核细胞例如是常被选择用于复合蛋白的生物合成的中国仓鼠卵巢细胞(CHO)或酵母。例如,酵母已被用于产生获批准的治疗蛋白,例如胰岛素、粒细胞-巨噬细胞集落刺激因子、血小板衍生的生长因子或水蛭素 (Gerngross (2004) Nat. Biotechnol. 22:1409-1414)。CHO细胞已被用于产生治疗蛋白,例如凝血因子IX、干扰素 β -1a、替奈普酶 (Chu (2001) Curr. Opin. Biotechnol. 12:180-187)或促性腺激素,其中糖组分可以正性地影响数方面,例如功能活性、折叠、二聚化、分泌以及受体相互作用、信号转导和代谢清除 (Walsh (2006) Nat. Biotechnol. 24:-1241-1252)。因此,本文还公开了本发明的构建体、无规卷曲多肽和缀合物在真核表达体系中的制备。

[0173] 利用本文提供的方式和方法,现已能够制备和提供本文公开的缀合物和分子,所述缀合物和分子包含(i)生物合成的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列和(ii)另一感兴趣的分子,例如有用的蛋白、蛋白区段或小分子。因此,本发明还提供了制备或生产所述缀合物以及生物合成的无规卷曲多肽和/或分子或包含所述生物合成的无规卷曲多肽和/或分子的缀合物的方法。因此,本发明还提供了制备和/或生产缀合物(例如药物缀合物、食品缀合物、诊断缀合物等)中包含的无规卷曲多肽或无规卷曲多肽区段的方法。此外,提供了制备和/或生产包含无规卷曲多肽或无规卷曲多肽区段的生物活性蛋白或缀合物的方法。此外,提供了制备和/或生产多肽的方法,所述多肽包含具有或介导生物活性的氨基酸序列或是具有或介导生物活性的氨基酸序列并且还包含所述无规卷曲多肽或无规卷曲多肽区段。这些方法

特别包括(作为一个步骤)培养上文提供的(宿主)细胞,和从培养物或所述细胞分离(作为另一个步骤)所述无规卷曲多肽或生物活性蛋白和/或所述生物活性蛋白和/或所述多肽缀合物。这种分离的无规卷曲、包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列的生物合成的无规卷曲多肽或多肽区段以及分离的缀合物可以被进一步加工。例如,所述包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列的生物合成的无规卷曲多肽或多肽区段可以与感兴趣的分子化学连接或偶联,如后文实施例所示。此外并作为备选方案,感兴趣的分子可以例如通过转谷氨酰胺酶(Besheer (2009) J Pharm Sci. 98:4420-8)或其它酶(Subul (2009) Org. Biomol. Chem. 7:3361-3371)与所述包含由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列的生物合成的无规卷曲多肽或多肽区段酶学缀合。

[0174] 无规卷曲多肽(或其区段)和/或包含无规卷曲多肽(或其区段)和感兴趣的蛋白(例如,生物活性蛋白或治疗活性蛋白,或可用于例如诊断方法中的蛋白)的蛋白缀合物可以(特别是)从生长介质、细胞裂解物、周质或细胞膜组分分离。(此外,本发明不限于可用于医学或药物领域的(蛋白)缀合物。本文提供的方式和方法还可用于其它工业领域,例如但不限于食品和饮料工业、营养品工业、造纸工业、生物制剂工业、研究工具和试剂工业、需要利用酶的工业、美容用品工业、石油加工和采油等)。可以通过任何常规方式进行本发明的表达的多肽的分离和纯化(Scopes (1982) "Protein Purification", Springer, New York, NY),包括硫酸铵沉淀、亲和纯化、柱层析、凝胶电泳等,并且可以涉及利用针对例如与本发明的生物活性蛋白融合的标签的单克隆或多克隆抗体。例如,可以利用链霉亲和素亲和层析、通过Strep标签II纯化蛋白(Skerra (2000) Methods Enzymol. 326:271-304),如后文实施例所述。具有至少约90至95%同质性(在蛋白水平上)的基本纯的多肽是优选的,并且98至99%或更高的同质性是最优选的,特别是对于药物用途/应用。根据制备过程中所用的宿主细胞/生物体,本发明的多肽可以是糖基化的或可以是非糖基化的。

[0175] 本发明还涉及本发明的生物活性蛋白、无规卷曲多肽(或其区段)或缀合物(例如药物缀合物)、本发明的核酸分子、本发明的载体或本发明的(宿主)细胞在制备药物中的用途,其中与不包含或不连接于生物合成的无规卷曲多肽或多肽区段的对照分子相比,所述生物活性蛋白或药物(或任何其它感兴趣的小分子或蛋白)具有增加的体内和/或体外稳定性,所述生物合成的无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。

[0176] 在另一实施方案中,本发明涉及治疗能受益于所述生物活性蛋白或药物的提高的稳定性的疾病和/或病症的方法,所述方法包括将本文所述的生物活性蛋白或药物缀合物给予需要这种治疗的哺乳动物。根据本发明蛋白或药物缀合物的生物活性,本领域技术人员能容易地确定利用本发明的具体生物活性蛋白或药物缀合物能治疗哪种疾病/病症。一些非限制性实例提供在下表中:

[0177]

生物活性蛋白 (或其生物活性组分/片段)或药物	待治疗的病症/疾病
粒细胞集落刺激因子	癌症和/或化疗相关的嗜中性白血球减少症
人生长激素	生长激素缺陷相关的低血糖症和/或生长不足
干扰素 α	癌症、病毒感染、丙型肝炎
干扰素 β	自身免疫疾病, 多发性硬化
干扰素 γ	病毒感染
肿瘤坏死因子	癌症
白介素-20	银屑病
α -半乳糖苷酶 A	法布瑞氏症
肌肉抑制素拮抗剂	肌肉减少症
胃抑制多肽	II 型糖尿病
α -1 抗胰蛋白酶	酶替换疗法, 囊性纤维化、慢性阻塞性肺病、急性呼吸综合征、严重哮喘
红细胞生成素	贫血症
凝血因子 VIII	血友病

[0178]

gp120/gp160	HIV
可溶性肿瘤坏死因子 I 和 II 受体	炎性疾病
瑞替普酶	血栓形成, 心肌梗塞
艾塞那肽-4	糖尿病
白介素-1 受体拮抗剂 (IL-1ra; 阿那白滞素)	自身免疫疾病, 风湿性关节炎
白介素-2	癌症
胰岛素	糖尿病
天冬酰胺酶	急性成淋巴细胞性白血病, 非霍奇金氏淋巴瘤
onconase	恶性间皮瘤和其它类型癌症
链激酶	血栓性疾病
中性粒细胞明胶酶相关脂质运载蛋白	微生物感染, 肾再灌注损伤
抗体及其片段, 包括单结构域抗体, 单链和其它工程化片段, 包括 CDR 模拟肽和 CDR	免疫学疾病, 肿瘤学疾病, 新生血管性疾病和感染性疾病等
粒细胞-巨噬细胞集落刺激因子	化疗相关嗜中性白血球减少症
卵泡刺激素	不孕症
葡糖脑苷脂酶	高歇氏病
胸腺素 $\alpha 1$	慢性乙型肝炎, 癌症
胰高血糖素	低血糖症
生长激素抑制素	肢端肥大症
腺苷脱氨酶	腺苷脱氨酶缺陷
白介素-11	血小板减少症
凝血因子 VIIa	血友病
凝血因子 IX	血友病
hematide	贫血症

[0179]

干扰素 λ	丙型肝炎
瘦素	脂肪营养不良、肥胖症、阿尔茨海默病、I 型糖尿病
白介素-22 受体 α 亚基 (IL-22ra)	银屑病
白介素 22	转移性黑素瘤
透明质酸酶	实体瘤
成纤维细胞生长因子 18	骨关节炎
成纤维细胞生长因子 21	II 型糖尿病、肥胖症、血脂异常、代谢病症
胰高血糖素样肽 1	糖尿病
骨保护素	癌症、骨质疏松症、风湿性关节炎
IL-18 结合蛋白	风湿性关节炎
生长激素释放因子	HIV 相关脂肪营养不良
可溶性 TACI 受体	系统性红斑狼疮、多发性硬化、风湿性关节炎
血小板反应蛋白-1	癌症
可溶性 VEGF 受体 Flt-1	癌症
IL-4 突变蛋白 (IL-4 受体拮抗剂)	哮喘
环孢霉素	器官排斥
烟曲霉素	癌症
纳曲酮	酒精依赖
奥曲肽	肢端肥大症、类癌肿瘤
替度鲁肽	短肠综合征、克罗恩氏病
戈舍瑞林	晚期前列腺癌症、乳腺癌
喜树碱	癌症
万古霉素	革兰氏阳性菌肺炎

[0180] 按照上文,生物活性蛋白、无规卷曲多肽(或其区段)、药物缀合物、核酸、载体或细胞可以用于制备药物,所述药物优选具有或提供增加的体内和/或体外稳定性,特别是对于生物活性蛋白和/或药物组分,所述药物用于治疗例如激素缺乏或相关病症、自身免疫疾病、癌症、贫血症、新生血管性疾病、感染性/炎症性疾病、血栓形成、心肌梗塞、糖尿病、不孕症、高球氏症、肝炎、低血糖症、肢端肥大症、腺苷脱氨酶缺陷、血小板减少症、血友病、贫血症、肥胖症、阿尔茨海默病、脂肪代谢障碍、银屑病、转移性黑素瘤、骨关节炎、血脂异常、风湿性关节炎、系统性红斑狼疮、多发性硬化、哮喘、骨质疏松症和再灌注损伤或其它肾病。在

一个实施方案中,生物活性蛋白、药物缀合物、核酸、载体或细胞被用作药物,所述药物具有所述生物活性蛋白/药物缀合物的增加的体内和/或体外稳定性。相似地,生物活性蛋白、无规卷曲多肽(或其区段)、药物缀合物、核酸、载体或细胞被用于治疗例如激素缺乏或相关病症、自身免疫疾病、增殖性病症(例如癌症、贫血症、新生血管性疾病)、感染性和/或炎性疾病、血栓形成、心肌梗塞、中风、糖尿病、不孕症、阴茎功能障碍、高球氏症、法布瑞氏症、肌肉减少症、囊性纤维化、阻塞性肺病、急性呼吸综合征、肝炎、低血糖症、肢端肥大症、腺苷脱氨酶缺陷、血小板减少症、血友病、贫血症、肥胖症、阿尔茨海默病、脂肪代谢障碍、银屑病、转移性黑素瘤、骨关节炎、血脂异常、风湿性关节炎、系统性红斑狼疮、多发性硬化、哮喘、骨质疏松症和再灌注损伤或其它肾病。

[0181] 本发明还涉及核酸分子、载体以及本文提供的并且包含本发明核酸分子或载体的转染的细胞在医学方法中的应用,例如基于细胞的基因治疗方法或基于核酸的基因治疗方法。

[0182] 在另一实施方案中,本文提供的无规卷曲多肽(或其多肽区段)、包含生物合成的无规卷曲多肽(或其多肽区段)的生物活性异源蛋白/蛋白构建体或药物或食品缀合物或其它缀合物和/或本发明的核酸分子或载体或宿主细胞是组合物的一部分。所述组合物可以包含下述中的一种或多种:本发明的无规卷曲多肽(或其多肽区段)、生物活性蛋白、食品缀合物、感兴趣的缀合物、药物缀合物或编码和/或表达上述物质的核酸分子、载体和/或宿主细胞。所述组合物可以是药物组合物,任选地还包含药学可接受的载体和/或稀释剂。在另一实施方案中,本发明涉及本文所述的生物活性蛋白、无规卷曲多肽(或其区段)或药物缀合物在制备药物组合物中的用途,所述药物组合物用于预防、治疗或改善需要摄取该药物组合物的疾病。

[0183] 如上文所述,本文公开的诸如药物缀合物或诊断缀合物的缀合物和/或生物活性异源蛋白/蛋白构建体(包含本发明无规卷曲多肽或其多肽区段)特别具有医学或药学用途。此外,所述无规卷曲多肽或多肽区段本身可以用于医学领域,例如作为“血浆扩容剂”或作为血液代替品,用于改善、预防和/或治疗与受损的血浆量或血浆内容物相关的病症,或用于改善、预防和/或治疗与受损的血容量相关的病症。需用血浆扩容剂治疗的病症包括但不限于,与失血相关的病症,例如损伤、手术、烧伤、创伤或急腹症、感染、脱水等。然而,这样的医学用途不限于本发明的无规卷曲多肽或多肽区段,还可以扩展到本文公开的某些药物缀合物或甚至扩展到某些生物活性异源蛋白/蛋白构建体。

[0184] 在一个实施方案中,本文所述的组合物可以是诸如成像剂的诊断组合物,其任选还包含适于检测的部分,其中所述诊断组合物具有增加的体内和/或体外稳定性。

[0185] 本发明的组合物可以是固体或液体形式,并且特别可以是粉末、片剂、溶液或气溶胶的形式。此外考虑到,本发明的药物可以包含其它生物活性剂,这取决于药物组合物的预期用途。

[0186] 可以通过不同方式实现合适的(药学)组合物的给药,例如,通过肠胃外给药、皮下给药、静脉内给药、动脉内给药、腹膜内给药、局部给药、支气管内给药、肺内给药和鼻内给药,并且如果需要定位治疗的话,可以通过病灶内给药。肠胃外给药包括腹膜内、肌肉内、真皮内、皮下、静脉内或动脉内给药。还可以将本发明的组合物直接给药至目标位点,例如,通过基因枪递送至外部或内部目标位点,例如具体受累的器官。

[0187] 合适的药物载体、赋形剂和/或稀释剂的实例是本领域公知的,并包括磷酸缓冲盐溶液或其它缓冲溶液、水、乳剂(例如油/水乳剂)、各种润湿剂、无菌溶液等。可以通过公知的常规方法配制包含所述载体的组合物。合适的载体可以包括任何材料,只要在与本发明的生物活性蛋白/药物缀合物组合时能保留生物和/或药物活性(参见Remington's Pharmaceutical Sciences(1980) 16th edition, Osol, A. Ed, Mack Publishing Company, Easton, PA)。用于肠胃外给药的制剂可以包括无菌的水性或非水性溶液、悬液和乳剂。药物组合物中所用的缓冲液、溶剂和/或赋形剂优选为上文所限定的“生理的”。非水性溶剂的实例是丙二醇、聚乙二醇、植物油(例如橄榄油)和可注射的有机酯(例如油酸乙酯)。水性载体包括水、醇溶液/水性溶液、乳剂或悬液,包括盐水和缓冲介质。肠胃外介质可以包括氯化钠溶液、林格氏葡萄糖、葡萄糖和氯化钠、乳酸林格氏液或固定油。静脉内介质可以包括液体和营养补充剂、电解质补充剂(例如基于林格氏葡萄糖的电解质补充剂)等。也可以存在防腐剂和/或其它添加剂,包括抗菌剂、抗氧化剂、螯合剂和/或惰性气体等。此外,本发明药物组合物可以包含蛋白性载体,例如,血清白蛋白或免疫球蛋白,优选来源于人。

[0188] 可以将这些药物组合物以合适的剂量给予个体。剂量方案由主治医师和临床因素确定。医学领域中公知的是,给予任一患者的剂量取决于很多因素,包括患者体型、体表面积、年龄、待给予的具体化合物、性别、给药时间和途径、一般健康状况和同时给予的其它药物。药学活性物质存在的量可以为每次剂量 $1\mu\text{g}$ 至 20mg/kg 体重,例如 0.1mg 至 10mg/kg 体重,例如 0.5mg 至 5mg/kg 体重。如果治疗方案是连续输注,则所述量应当为 $1\mu\text{g}$ 至 10mg/kg 体重/分钟。然而,也考虑到了低于或高于上述示例性的范围的剂量,特别是考虑到上述因素时。

[0189] 此外还考虑到,本发明的药物组合物可以包含其它生物或药学活性剂,这取决于药物组合物的预期用途。这些其它生物或药学活性剂可以是,例如抗体、抗体片段、激素、生长因子、酶、结合分子、细胞因子、趋化因子、核酸分子和药物。

[0190] 应当注意到,本发明不限于药物组合物。还考虑到用于研究的组合物或作为诊断制剂的组合物。例如,考虑到了本文所限定的包含无规卷曲结构域或组分的生物活性蛋白或药物缀合物可用于诊断领域。为此,可以对本发明的生物活性蛋白或药物缀合物进行标记以允许进行检测。所述标记包括但不限于,放射性标记(例如 ^3H 氢、 ^{125}I 碘或 ^{123}I 碘)、荧光标记(包括荧光蛋白,例如绿色荧光蛋白(GFP)或荧光团,例如异硫氰酸荧光素(FITC))或NMR标记(例如钆螯合物)。本文限定的标记或标记物并不是限制性的,而仅是提供示例性的实例。本发明的诊断组合物特别可用于追踪或成像实验或诊断医学领域。在后文的实施例和附图中,提供了相应的构建体的制备,其包括缀合物,所述缀合物包含(i)生物合成的无规卷曲多肽或多肽区段,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成,和(ii)荧光素或地高辛配基;参见附图13和14和相应的图例以及示例性的实施例22。

[0191] 但是,本发明的主旨不仅包括本文提供的方式和方法的药学或诊断用途。本文提供的化合物/缀合物还可用于某些其它工业领域,例如食品工业、饮料工业、美容用品工业、石油工业、造纸工业等。因此,本发明还提供了本文提供的生物合成的无规卷曲多肽在这些工业领域中的用途。因此,本发明的另一部分涉及产生美容用品、用于美容治疗的化合物、食品或饮料的方法,所述方法包括培养包含核酸分子(或载体)的细胞,所述核酸分子(或载

体)编码本文所限定的无规卷曲多肽或编码生物活性蛋白和/或包含具有或介导活性的氨基酸序列或是具有或介导活性的氨基酸序列的生物活性蛋白和/或多肽。上述方法还包括从培养物或所述细胞分离所述无规卷曲多肽、所述生物活性蛋白和/或包含具有或介导活性(例如生物活性)的氨基酸序列或是具有或介导活性的氨基酸序列并且还包含所述无规卷曲多肽或无规卷曲多肽区段的所述生物活性蛋白或所述多肽。在相同情况下,可以产生其它感兴趣的缀合物,例如可用于不同工业领域(例如石油或造纸工业)的缀合物。本领域技术人员能够容易地调整本文提供的方式和方法,从而制备出相应的分子/重组构建体以及制备包含生物合成的无规卷曲多肽或多肽区段和感兴趣的小分子或多肽的缀合物,所述无规卷曲多肽或多肽区段包含仅由脯氨酸和丙氨酸氨基酸残基组成的氨基酸序列,其中所述氨基酸序列由至少50个脯氨酸(Pro)和丙氨酸(Ala)氨基酸残基组成。

[0192] 在另一实施方案中,本发明提供了试剂盒,所述试剂盒包含无规卷曲多肽(或其多肽区段)、生物活性蛋白、药物缀合物、编码所述生物活性蛋白、编码所述生物活性蛋白和/或编码所述生物活性蛋白和/或编码包含具有或介导活性(例如生物活性)的氨基酸序列或是具有或介导活性的氨基酸序列的多肽的核酸分子、包含所述核酸分子的载体或包含本文所述的核酸或载体的细胞。本发明的试剂盒可以还包含缓冲液、贮存溶液和/或进行医学、科学或诊断测定和应用所需的其它试剂或材料。此外,本发明试剂盒的各部分可以在小瓶或瓶子中单独包装,或在容器或多容器单元中组合包装。

[0193] 本发明的试剂盒可以特别有利地用于实施本发明的方法,并且可以在本文所述的多种应用中使用,例如,作为诊断试剂盒、研究工具或医学工具。此外,本发明的试剂盒可以含有适于科学、医学和/或诊断用途的检测装置。试剂盒的制造优选依照本领域技术人员已知的标准操作。

[0194] 以下非限制性附图和实施例进一步说明了本发明。

附图说明

[0195] 图1.PA#1Pro/Ala聚合物/多肽序列的基因设计。

[0196] PA#1的构成部分的核苷酸和编码的氨基酸序列(SEQ ID NO:1),其是通过两个互补寡脱氧核苷酸(上方/编码链寡脱氧核苷酸SEQ ID NO:17,下方/非编码链寡脱氧核苷酸SEQ ID NO:18)的杂交而获得的。所得到的核酸具有两个分别对应于Ala密码子和反密码子的粘性末端(如小写字母所示),并且能相互配合。通过这样的构成部分的重复连接,可以获得编码不同长度的Pro-Ala多肽的多连体,并且然后可以将其克隆,例如,通过SapI限制性位点。

[0197] 图2.Pro/Ala聚合物/多肽序列与Fab片段或人IFN α 2b的融合物的克隆策略。

[0198] (A) 位于由pASK88-Fab-2xSapI(SEQ ID NO:22)(pASK75的衍生物)编码的抗体Fab片段的免疫球蛋白轻链的C末端附近的核苷酸和编码的氨基酸序列段(上方/编码链SEQ ID NO:19,下方/非编码链SEQ ID NO:20,编码的氨基酸序列SEQ ID NO:21),用于亚克隆Pro/Ala聚合物/多肽序列和表达相应的生物活性蛋白。核苷酸序列携带两个互逆方向的SapI识别位点,通过消化,其可以导致能与图1所示的合成基因盒配合的DNA末端。轻链的识别序列和C末端氨基酸用下划线标出。

[0199] (B) 将图1所示的单个盒插入pASK88-Fab-2xSapI质粒后的具有20个残基的PA#1聚

合物/多肽的核苷酸序列和编码的氨基酸序列(上方/编码链SEQ ID NO:23,下方/非编码链SEQ ID NO:24,编码的氨基酸序列SEQ ID NO:25)。相似地连接/插入10个这样的重复的盒能得到质粒载体pFab-PA#1(200)(Seq ID NO:28),其编码具有200个残基的聚合物/多肽(SEQ ID NO:26和27)。位于Pro/Ala聚合物编码序列侧翼的SapI限制性位点已标出(识别序列用下划线标出)。

[0200] (C) pFab-PA#1(200)(SEQ ID NO:28)的质粒图谱。Fab-PA#1(200)的重链(HC)和轻链(LC)的结构基因受到四环素启动子/操纵基因($tet^{p/o}$)的转录控制,并且操纵子的末端为脂蛋白终止子(t_{1pp})。HC包含细菌OmpA信号肽、可变结构域(VH)和第一人IgG1重链恒定C结构域(CH)以及His₆标签。LC包含细菌PhoA信号肽、可变结构域(VL)和人轻链恒定(CL)结构域、具有200个残基的PA#1聚合物/多肽。侧翼为XbaI和HindIII限制性位点的表达盒之外的pFab-PA#1(200)质粒骨架与基因克隆和表达载体pASK75的质粒骨架是相同的(Skerra(1994)Gene 151:131-135)。标出了单限制性位点。

[0201] (D) 克隆在pASK-IFNa2b(SEQ ID NO:32)上的人IFNa2b的N末端附近的核苷酸和氨基酸序列段(上方/编码链SEQ ID NO:29,下方/非编码链SEQ ID NO:30,编码的氨基酸序列SEQ ID NO:31)。可用于插入Pro/Ala聚合物编码序列的单限制性位点SapI已标出(识别序列用下划线标出)。Strep标签II的两个C末端氨基酸用下划线标出。成熟IFNa2b的第一氨基酸用+1标记。

[0202] (E) 在插入一个图1所示的PA#1聚合物序列盒之后,IFNa2b的N末端的核苷酸和编码的氨基酸序列段(上方/编码链SEQ ID NO:33,下方/非编码链SEQ ID NO:34,编码的氨基酸序列SEQ ID NO:35)。在插入Pro/Ala聚合物编码序列后保留的单限制性位点SapI被标出(识别序列用下划线标出)。作为融合蛋白的一部分的IFNa2b的第一氨基酸被标出(1),并且Strep标签II的两个C末端氨基酸用下划线标出。相似地连接/插入10个重复的PA#1聚合物序列盒得到质粒载体pPA#1(200)-IFNa2b,其编码具有200个残基的聚合物/多肽(SEQ ID NO:36)

[0203] (F) pPA#1(200)-IFNa2b的质粒图谱(SEQ ID NO:37)。生物活性蛋白PA#1(200)-IFNa2b的结构基因(包含细菌OmpA信号肽、Strep标签II、具有200个残基的PA#1聚合物/多肽区段和人IFNa2b)受四环素启动子/操纵基因($tet^{p/o}$)的转录控制并且末端为脂蛋白终止子(t_{1pp})。侧翼为XbaI和HindIII限制性位点的表达盒之外的质粒骨架与基因克隆和表达载体pASK75的质粒骨架是相同的(Skerra(1994)同上)。标出了单限制性位点。

[0204] 图3. 纯化的重组Fab片段和纯化的重组IFNa2b以及它们的Pro/Ala多肽/聚合物融合物的SDS-PAGE分析。

[0205] 通过周质分泌在大肠杆菌KS272(Strauch(1988)Proc.Natl.Acad.Sci.USA 85:1576-80)中产生重组蛋白,并分别借助于His₆标签(Fab)或Strep标签II(IFNa2b)、利用固定的金属亲和层析或链霉亲和素亲和层析进行纯化。

[0206] (A) 利用12%SDS-PAGE,分析纯化的重组Fab以及它的具有200个残基的PA#1融合蛋白。凝胶显示出Fab和Fab-PA#1(200)中每一种的2μg蛋白样品。左侧样品用2-巯基乙醇还原,而右侧重复的样品未还原。蛋白标记物(在还原条件下施加)的大小显示在左侧边缘。通过链间二硫键的还原,Fab片段及其200个残基的PA#1融合物显示为两个均匀条带。对于还原的Fab片段,两个条带的分子大小分别为约24和26kDa,对应于单独的LC和HC。对于还原的

Fab-PA#1 (200) 融合蛋白, 24kDa处的条带对应于HC, 而约90kDa处的条带对应于与PA#1 (200) 多肽区段融合的LC。在非还原条件下, Fab片段及其PA#1 (200) 融合物显示为单个均匀的条带, 表观分子大小分别为约45kDa和100kDa。Fab-PA#1 (200) 融合蛋白的两个表观尺寸值显著大于计算质量, 对于非还原Fab-PA#1 (200), 计算质量为64.3kDa, 对于分离的LC-PA#1 (200), 计算质量为39.1kDa。这种效果是由于添加Pro/Ala聚合物/多肽区段所致, 因为计算质量为48.0kDa的Fab片段本身或其未融合的轻链显示出基本正常的电泳迁移率。

[0207] (B) 利用12%SDS-PAGE, 分析纯化的重组IFNa2b以及它的具有200个残基的PA#1融合蛋白。凝胶显示出IFNa2b和PA#1 (200)-IFNa2b中每一种的2 μ g蛋白样品。左侧样品用2-巯基乙醇还原, 而对应的右侧样品未还原。蛋白标记物(在还原条件下施加)的大小显示在左侧边缘。两种蛋白显示为单个均匀条带, 还原形式的表观分子大小为约20kDa和约80kDa。后一值显著大于PA#1 (200)-IFNa2b的计算质量37.0kDa。这种效果是由于添加Pro/Ala聚合物/多肽区段所致, 因为计算质量为20.9kDa的IFNa2b本身显示出基本正常的电泳迁移率。非还原状态的IFNa2b具有稍高的电泳迁移率, 这表明由于其两个分子内二硫键而产生更紧凑的形式。

[0208] 图4. 纯化的重组Fab和IFNa2b以及它们的PA#1 (200) 融合物的流体动力学体积的定量分析。

[0209] (A) Fab和Fab-PA#1 (200) 的分析尺寸排阻色谱(SEC)。将浓度为0.25mg/ml的250 μ l 纯化蛋白施加到用PBS缓冲液平衡的Superdex S20010/300GL柱。监测280nm处的吸收, 每个层析运行的峰被标准化为值等于1。箭头指示柱的孔隙体积(7.8ml)。

[0210] (B) 利用Superdex S20010/300GL柱, 来自(A)的色谱图的校准曲线。将标记蛋白的分子量(MW)(细胞色素c: 12.4kDa; 碳酸酐酶, 29.0kDa; 卵白蛋白: 43.0kDa; 牛血清白蛋白: 66.3kDa; 醇脱氢酶: 150kDa, β -淀粉酶: 200kDa, 去铁铁蛋白: 440kDa)的对数与它们的洗脱体积(黑色圆圈)作图, 并进行直线拟合。根据所观察到的Fab片段及其PA#1融合蛋白(黑色方块)的洗脱体积, 它们的表观分子大小被确定如下。Fab: 31kDa(真实质量: 48.0kDa); Fab-PA#1 (200): 237kDa(真实质量: 64.3kDa)。这些数据表明: 与PA#1多肽的融合赋予了明显增大的流体动力学体积。

[0211] (C) IFNa2b和PA#1 (200)-IFNa2b的分析尺寸排阻色谱。将浓度为0.25mg/ml的250 μ l 每种纯化蛋白施加到用磷酸缓冲盐溶液(PBS)平衡的Superdex S20010/300GL柱。监测280nm处的吸收, 每个层析运行的峰被标准化为值等于1。箭头指示柱的孔隙体积(7.8ml)。

[0212] (D) 利用Superdex S20010/300GL柱, 来自(C)的色谱图的校准曲线。将标记蛋白的分子量(MW)(参见B)的对数与它们的洗脱体积(黑色圆圈)作图, 并进行直线拟合。根据所观察到的IFNa2b及其PA#1融合蛋白(黑色方块)的洗脱体积, 它们的表观分子大小被确定如下。IFNa2b: 22.5kDa(真实质量: 20.9kDa); PA#1 (200)-IFNa2b: 229.0kDa(真实质量: 37.0kDa)。这些数据表明: 与PA#1多肽的融合赋予了明显增大的流体动力学体积。

[0213] 图5. 通过圆二色(CD)光谱对重组蛋白及其PA#1聚合物/多肽融合物的实验性二级结构进行分析。

[0214] 对于每种蛋白, 记录室温下、50mM K₂SO₄、20mM磷酸钾、pH 7.5的溶液中的光谱, 并将其标准化为摩尔椭圆率(Θ M)。

[0215] (A) 纯化的重组Fab和Fab-PA#1 (200) 的CD光谱。Fab片段的CD光谱显示了典型的主

要为 β 折叠的蛋白的特征,具有位于216nm附近的宽的负最大值(negative maximum)(Sreerama in: Circular Dichroism-Principles and Applications (2000) Berova, Nakanishi and Woody (Eds.) Wiley, New York, NY, pp. 601-620),这指示细菌产生的Fab片段的正确折叠。Fab片段与Pro/Ala聚合物/多肽的融合蛋白的光谱显示了低于200nm的优势负带,这指示无规卷曲构象。此外,在220nm附近具有肩部,这来自Fab片段的 β 折叠贡献,并指示即使Fab片段作为融合蛋白的一部分也能正确折叠。

[0216] (B) 通过减去Fab片段的光谱而获得的Fab-PA#1 (200) 的摩尔差异CD光谱。差异CD光谱表示200个残基的PA#1聚合物/多肽区段的二级结构,并且显示出200nm附近的强的最小值,这明确表明缓冲水性溶液中的无规卷曲构象 (Greenfield (1969) Biochemistry 8: 4108-4116; Sreerama (2000) 同上; Fändrich (2002) EMBO J. 21: 5682-5690)。

[0217] (C) 纯化的重组IFNa2b和PA#1 (200) -IFNa2b的CD光谱。IFNa2b的CD光谱显示了典型的主要为 α 螺旋蛋白的特征,具有位于208nm和220nm附近的两个负带 (Sreerama (2000) 同上),这指示细菌产生的人IFNa2b的正确折叠。IFNa2b与Pro/Ala聚合物/多肽的融合蛋白的光谱显示出特征性偏差,其具有位于200nm附近的优势最小值,这指示无规卷曲构象。此外,在220nm附近具有肩部,这来自IFNa2b的 α 螺旋贡献,并指示即使IFNa2b作为融合蛋白的一部分也能正确折叠。

[0218] (D) 通过减去IFNa2b的光谱而获得的PA#1 (200) -IFNa2b的摩尔差异CD光谱。差异CD光谱表示200个残基的PA#1聚合物/多肽区段的二级结构,并且显示出200nm附近的强的最小值,这与(B)中所示基本相同。这再次明确表明本发明的包含Pro和Ala残基的生物聚合物在缓冲水性溶液中的无规卷曲构象。

[0219] 图6.人生长激素(hGH)和基因编码的PA#1聚合物的融合蛋白在CHO细胞中的分泌产生。

[0220] (A) 克隆在pASK75-His₆-hGH (SEQ ID NO: 41) 上的hGH的N末端附近的核苷酸和氨基酸序列段(上方/编码链SEQ ID NO: 38, 下方/非编码链SEQ ID NO: 39, 编码的氨基酸序列SEQ ID NO: 40)。可与HindIII(未示出)一起用于亚克隆的单限制性位点NheI和可用于插入Pro/Ala聚合物编码序列的SapI被标出(识别序列用下划线标出)。His₆标签的6个氨基酸用下划线标出。hGH的第一氨基酸用+1标记。

[0221] (B) 在插入一个图1所示的PA#1聚合物序列盒后,hGH的N末端的核苷酸和编码的氨基酸序列(上方/编码链SEQ ID NO: 42, 下方/非编码链SEQ ID NO: 43, 编码的氨基酸序列SEQ ID NO: 44)。可用于亚克隆的单限制性位点NheI和在插入Pro/Ala聚合物编码序列后仍保留的SapI被标出(识别序列用下划线标出)。作为融合蛋白的一部分的hGH的第一氨基酸被标出(1), His₆标签的氨基酸用下划线标出。相似地连接/插入10个重复的PA#1聚合物序列盒能得到质粒载体pASK75-His₆-PA#1 (200) -hGH,其编码成熟融合蛋白SEQ ID NO: 45。

[0222] (C) pASK75-His₆-PA#1 (200) -hGH (SEQ ID NO: 46) 的质粒图谱。生物活性蛋白His₆-PA#1 (200) -hGH的结构基因(包括细菌OmpA信号肽、His₆标签、具有200个残基的PA#1聚合物/多肽区段和人GH)受四环素启动子/操纵基因(tet^{r/o})的转录控制,并且末端为脂蛋白终止子(t_{1pp})。侧翼为XbaI和HindIII限制性位点的表达盒之外的质粒骨架与基因克隆和表达载体pASK75的质粒骨架与相同 (Skerra (1994) 同上)。标出了单限制性位点。

[0223] (D) 编码His₆-PA#1 (200) -hGH融合蛋白 (SEQ ID NO: 47) 的pCHO-PA#1 (200) -hGH的

质粒图谱。结构基因(包含人生长激素信号肽(Sp)、His₆标签、具有200个残基的PA#1聚合物/多肽序列(PA#1(200))、人生长激素(hGH)并含有牛生长激素聚腺苷酸化信号(bGH pA))受巨细胞病毒启动子(CMV^p)的转录控制。标出了单限制性位点NheI和HindIII。新霉素磷酸转移酶(neo)的抗性基因受SV40启动子(SV40^p)的控制,并且后面是SV40聚腺苷酸化信号(SV40pA)。此外,质粒含有细菌ColE1复制起点(ColE1-ori)、噬菌体f1复制起点(f1-ori)和β-内酰胺酶基因(bla),从而允许在大肠杆菌中增殖和选择。

[0224] (E) CHO细胞产生的hGH和基因编码的200个残基的PA#1聚合物的融合蛋白与重组hGH的Western印迹分析比较。用pCHO-PA#1(200)-hGH(SEQ ID NO:48)或pCHO-hGH(SEQ ID NO:49)转染CHO-K1细胞,pCHO-hGH是编码hGH但不具有PA#1(200)序列的质粒(但是也携带His₆标签)。转染后两天,将细胞培养物上清的样品进行SDS-PAGE并利用与辣根过氧化物酶缀合的抗hGH抗体进行Western印迹。两种蛋白表现为箭头所指的单条带,表观分子大小为约23kDa(His₆-hGH)和约90kDa(His₆-PA#1-hGH)。在60kDa附近还有弱条带,其来自培养基中的血清蛋白。而具有His₆标签的hGH以23.5kDa的计算质量出现,His₆-PA#1-hGH的表观分子大小显著大于其计算质量39.5kDa。这种效果是由于Pro-Ala聚合物的亲水性无规卷曲性质所致。

[0225] 图7.PA#1Pro/Ala多肽/聚合物序列的二级结构的理论预测。

[0226] 该图显示了按照Chou-Fasman法(Chou and Fasman (1974) Biochemistry 13: 222-245)的CHOFAS计算机算法的结果,该计算是在弗吉尼亚大学的序列比较和二级结构预测服务器上执行的(URL:http://fasta.bioch.virginia.edu/fasta_www2)。为了避免氨基和羧基末端的边界效应,粘贴了3个连续拷贝的图1的20聚体氨基酸重复(得到了与重复连接/插入合成基因盒之后所编码的相似的多连体),并且仅考虑了中央的20聚体序列构成部分(用框圈出的)的结果。对于PA#1多肽序列/区段(SEQ ID NO:1),Chou-Fasman算法预测为100%α螺旋二级结构。这与实验观察到的作为融合蛋白的一部分的PA#1多肽/多肽区段主要无规卷曲构象不符(参见图5B/D)。

[0227] 图8.BALB/c小鼠中纯化的重组Fab片段以及其具有200个和600个残基的PA#1聚合物融合物的药代动力学的定量分析。

[0228] 利用夹心ELISA,测定来自实施例16的血浆样品的Fab、Fab-PA#1(200)和Fab-PA#1(600)浓度。为了估计Fab、Fab-PA#1(200)和Fab-PA#1(600)的血浆半衰期,将测量的浓度值对静脉内注射后的时间作图,并假定双指数衰减进行数字拟合。未融合的Fab片段表现出非常快的清除,清除半衰期为 1.3 ± 0.1 h(小时)。相比之下,对于Fab-PA#1(200)和Fab-PA#1(600)测定的清除期显著较慢,终末半衰期分别为 4.1 ± 1.8 h和 38.8 ± 11.2 h,从而与未融合的Fab片段相比,由于具有200或600个残基Pro/Ala聚合物的融合而表现出约3倍和约30倍的延长的循环。

[0229] 图9:作为与具有200个残基的P1A1或P1A3聚合物的融合物的纯化的重组Fab片段的分析。

[0230] 通过周质分泌在大肠杆菌KS272中产生重组蛋白,并借助于His₆标签、利用固定的金属亲和层析进行纯化。用12%SDS-PAGE分析纯化的蛋白。凝胶显示出Fab-P1A1(200)和Fab-P1A3(200)中的每一种以及为了比较用的未融合的Fab片段的2μg蛋白样品(参见图3A)。左侧样品用2-巯基乙醇还原,而右侧的相似样品未还原。蛋白标记物(在还原条件下施

加)的大小显示在左侧边缘。链间二硫键还原后,Fab片段及其200个残基的Pro/Ala融合物表现为两个均匀条带。对于还原的Fab片段,两个条带的分子大小分别为约24和26kDa,其对应于单独的轻链(LC)和重链片段(HC)。对于还原的Fab-P1A1(200)融合蛋白,24kDa处的条带对应于HC,而约90kDa处的条带对应于与P1A1(200)多肽融合的LC。对于还原的Fab-P1A3(200)融合蛋白,24kDa处的条带对应于HC,而约75kDa处的条带对应于与P1A5(200)多肽融合的LC。在非还原条件下,Fab片段、其P1A1(200)和其P1A3(200)融合蛋白表现为单一优势条带,表观分子大小分别为约45kDa、110kDa和90kDa。Fab-P1A1(200)和Fab-P1A3(200)融合蛋白的表观尺寸显著大于计算质量,其中对于非还原的Fab-P1A1(200),计算质量为65.3kD,对于非还原的Fab-P1A3(200),计算质量为64.0kDa。另外,相应的还原的轻链的表观尺寸显著大于计算质量,其中对于P1A1(200)LC,计算质量为40.7kDa,对于P1A3(200)LC,计算质量为39.4kDa。这种效果是由于添加Pro/Ala聚合物/多肽区段所致,因为计算质量为48.0kDa的Fab片段自身或它的计算质量为23.4kDa的未融合的轻链表现出基本正常的电泳迁移率。

[0231] 图10.纯化的重组Fab-P1A1(200)和Fab-P1A3(200)融合蛋白的流体动力学体积的定量分析。

[0232] Fab-P1A1(200)和Fab-P1A3(200)的分析尺寸排阻色谱(SEC)。将浓度为0.25mg/ml的250 μ l纯化蛋白施加到用PBS平衡的Superdex S20010/300GL柱。监测280nm处的吸收,每个层析运行的峰被标准化为值等于1。箭头指示柱的孔隙体积(7.8ml)。根据所观察到的融合蛋白的洗脱体积,利用与图4B所示相似的校准曲线,将它们的表观分子大小确定如下。Fab-P1A1(200):180.7kDa(真实质量:65.3kDa);Fab-P1A3(200):160.2kDa(真实质量:64.0kDa)。这些数据表明:蛋白与P1A1和/或P1A5多肽的融合赋予了明显增大的流体动力学体积。

[0233] 图11.通过圆二色(CD)光谱对Fab-P1A1(200)和Fab-P1A3(200)融合蛋白的实验性二级结构进行分析。

[0234] 对于每种蛋白,记录室温下、50mM K₂SO₄、20mM磷酸钾、pH 7.5的溶液中的光谱,并将其标准化为摩尔椭圆率(Θ M)。

[0235] (A)纯化的重组Fab-P1A1(200)和Fab-P1A3(200)的CD光谱。具有Pro/Ala聚合物/多肽的Fab融合蛋白的CD光谱都表现出明显的低于200nm的负带,这指示无规卷曲构象。此外,在220nm附近具有肩部,这来自Fab片段的 β 折叠贡献,并指示即使Fab片段作为融合蛋白的一部分也能正确折叠。

[0236] (B)通过减去未融合的Fab片段的光谱而获得的Fab-P1A1(200)和Fab-P1A3(200)的摩尔差异CD光谱(参见图5A)。差异CD光谱分别表示200个残基的P1A1(SEQ ID NO:51)和P1A3(SEQ ID NO:3)聚合物/多肽区段的二级结构,并且显示出200nm附近的强的最小值,这明确表明它们在缓冲水性溶液中的无规卷曲构象(Greenfield(1969) Biochemistry 8: 4108-4116;Sreerama(2000)同上;Fändrich(2002)EMBO J.21:5682-5690)。

[0237] 图12:分离的生物合成的Pro/Ala聚合物/多肽的制备。

[0238] (A) pSUMO-PA#1(200)(SEQ ID NO:60)的质粒图谱。融合蛋白MK-His(6)-SUMO-PA#1(200)的结构基因(包含起始蛋氨酸密码子、然后是赖氨酸密码子、6个连续His残基的N末端亲和和标签、可切割的小泛素样修饰蛋白(SUMO)蛋白Smt3p(Panavas(2009)Methods Mol

Biol.497:303-17) 和具有200个残基的PA#1聚合物/多肽区段(SEQ ID NO:60))受噬菌体T7的基因10启动子的转录控制,并且末端为t ϕ 终止子。其它质粒元件包括复制起点(ori)、氨苄青霉素抗性基因(bla)和f1复制起点。除了通过沉默突变消除的SapI限制性位点之外,侧翼为NdeI和HindIII限制性位点的表达盒之外的质粒骨架与一般克隆和表达载体pRSET5a(Schoepfer(1993)124:83-85)的质粒骨架是相同的。

[0239] SEQ ID NO:60提供在附带的序列表中(其也是本说明书的一部分),并且也可参见下文。

```

[0240] gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa 60
[0241] atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat tgaaaaagga 120
[0242] agagtatgag tattcaacat ttccgtgtcg cccttattcc cttttttgcg gcattttgcc 180
[0243] ttccgtgttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg 240
[0244] gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt gagagttttc 300
[0245] gccccgaaga acgtttttcca atgatgagca cttttaaagt tctgctatgt ggcgcggtat 360
[0246] tatcccgat tgacgccggg caagagcaac tcggtcgccg catacactat tctcagaatg 420
[0247] acttggttga gtactcacca gtcacagaaa agcatcttac ggatggcatg acagtaagag 480
[0248] aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta cttctgacaa 540
[0249] cgatcggagg accgaaggag ctaaccgctt ttttgacaa catgggggat catgtaactc 600
[0250] gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca 660
[0251] cgatgcctgt agcaatggca acaacgttgc gaaactatt aactggcgaa ctacttactc 720
[0252] tagcttcccg gcaacaatta atagactgga tggaggcgga taaagttgca ggaccacttc 780
[0253] tgcgctcggc ccttcgggt ggctggttta ttgctgataa atctggagcc ggtgagcgtg 840
[0254] ggtctcgcg tatcattgca gcactggggc cagatggtaa gccctcccg atcgtagtta 900
[0255] tctacagac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag 960
[0256] gtgcctcact gattaagcat tggttaactgt cagaccaagt ttactcatat atactttaga1020
[0257] ttgatttaaa acttcatttt taatttaaaa ggatctaggt gaagatcctt tttgataatc 1080
[0258] tcatgaccaa aatcccttaa cgtgagtttt cgttcactg agcgtcagac cccgtagaaa 1140
[0259] agatcaaagg atcttcttga gatccttttt ttctgcgcgt aatctgctgc ttgcaaacaa 1200
[0260] aaaaaccacc gctaccagcg gtggtttgtt tgccggatca agagctacca actctttttc 1260
[0261] cgaaggtaac tggcttcagc agagcgcaga taccaaatac tgtccttcta gtgtagccgt 1320
[0262] agttaggcca ccacttcaag aactctgtag caccgctac atacctcgct ctgctaatac 1380
[0263] tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg gactcaagac 1440
[0264] gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca 1500
[0265] gcttggagcg aacgacctac accgaactga gatactaca gcgtgagcta tgagaaagcg 1560
[0266] ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg gtcggaacag 1620
[0267] gagagcgcac gagggagctt ccagggggaa acgcttggt tctttatagt cctgtcgggt 1680
[0268] ttgccacct ctgacttgag cgtegatttt tgtgatgtc gtcagggggg cggagcctat 1740
[0269] ggaaaaacgc cagcaacgcg gcctttttac gtttcttggc cttttgctgg ccttttgctc 1800
[0270] acatgttctt tctcggtta tccctgatt ctgtggataa ccgtattacc gcctttgagt 1860
[0271] gagctgatac cgctcgccgc agccgaacga ccgagcgag cgagtcagtg agcgaggaag 1920

```

[0272] cggagaagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt cattaatgca 1980
 [0273] ggatctcgat cccgcgaaat taatacgact cactataggg agaccacaac ggtttccctc 2040
 [0274] tagaaataat tttgtttaac ttttaagaagg agatatacat atgaaacatc accaccatca 2100
 [0275] ccattcggac tcagaagtca atcaagaagc taagccagag gtcaagccag aagtcaagcc 2160
 [0276] tgagactcac atcaatttaa aggtgtccga tggatcttca gaaatcttct ttaagatcaa 2220
 [0277] aaagaccact cctttaagaa ggctgatgga agcgttcgct aaaagacagg gtaaggaaat 2280
 [0278] ggactcctta agattcttgt acgacggtat tagaattcaa gctgatcaga cccctgaaga 2340
 [0279] tttggacatg gaggataacg atattattga ggctcacaga gaacagattg gtggcgccgc 2400
 [0280] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2460
 [0281] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2520
 [0282] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2580
 [0283] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2640
 [0284] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2700
 [0285] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2760
 [0286] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2820
 [0287] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2880
 [0288] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgccgc 2940
 [0289] tccagctgca cctgctccag cagcacctgc tgcaccagct ccggtgctc ctgctgcctg 3000
 [0290] aagagcaagc ttgatccggc tgctaacaa gcccgaagg aagctgagtt ggctgctgcc 3060
 [0291] accgtgagc aataactagc ataaccctt ggggcctcta aacgggtctt gaggggtttt 3120
 [0292] ttgtgaaag gaggaactat atccgatct ggctgaatag cgaagaggcc cgcaccgatc 3180
 [0293] gcccttccca acagttgcgc agcctgaatg gcgaatggga cgcgccctgt agcggcgcat 3240
 [0294] taagcgcggc ggggtgtgtg gttacgcgca gcgtgaccgc tacacttgcc agcgccttag 3300
 [0295] cgcccgtcc tttcgcttcc ttccttccct ttctcgccac gtctcgccgc tttccccgtc 3360
 [0296] aagctctaaa tcgggggctc cctttagggt tccgatttag tgctttacgg cacctcgacc 3420
 [0297] ccaaaaaact tgattagggt gatggttcac gtagtgggcc atcgccctga tagacggttt 3480
 [0298] ttcgcccttt gacgttggag tccacgttct ttaatagtgg actcttgctc caaactggaa 3540
 [0299] caacactcaa ccctatctcg gtctattctt ttgatttata agggattttg ccgatttcgg 3600
 [0300] cctatttggt aaaaaatgag ctgattttaac aaaaatttaa cgcgaaattt aacaaaatat 3660
 [0301] taacgettac aatttaggtg 3680

[0302] (B) 利用12%SDS-PAGE,对细菌产生的His (6)-SUMO-PA#1 (200) 融合蛋白及其切割产物的分析。凝胶显示了在用Ub1特异性蛋白酶1 (SUMO蛋白酶) 进行蛋白水解切割之前(泳道1) 和之后(泳道2) 的SUMO-PAS#1 (200) 融合蛋白,正如实施例21所述,所述SUMO-PAS#1 (200) 融合蛋白是从大肠杆菌中提取并通过固定的金属亲和层析(IMAC) 和尺寸排阻色谱(SEC) 纯化的。所有样品都用2-巯基乙醇还原。蛋白标记物(M) (在还原条件下施加) 的大小显示在左侧边缘。His (6)-SUMO-PA#1 (200) 融合蛋白表现为单一均匀条带,表观分子大小约为100kDa。因此,在SDS-PAGE中观察到的His (6)-SUMO-PA#1 (200) 融合蛋白的表观尺寸显著大于计算质量28.3kDa,这是由于存在Pro/Ala聚合物/多肽所致。切割后,无法检测到亲水性PA#1 (200) 多肽的考马斯亮蓝染色;因此,仅有小的融合蛋白残余部分和切割过的His

(6)-SUMO蛋白能在SDS聚丙烯酰胺凝胶上观察到(泳道2)。His (6)-SUMO蛋白显示了均匀条带,表观分子大小为约16kDa(泳道2),这很好地符合其计算分子量12.2kDa。

[0303] 图13:生物合成的Pro/Ala聚合物/多肽与化学化合物和/或药物的缀合。

[0304] (A-D)通过分析尺寸排阻色谱(SEC)监测具有生物合成的PA#1 (200) 聚合物/多肽(SEQ ID NO:61)的荧光素缀合物的产生。图(从上至下)显示了以下的SEC运行:纯化的His (6)-SUMO-PA#1 (200) (A),在SUMO蛋白酶的存在下进行切割反应后的His (6)-SUMO-PA#1 (200) (B)、与荧光素NHS酯化学偶联后的一批切割的His (6)-SUMO-PA#1 (200) (C)和IMAC纯化后的荧光素-PA#1 (200) 缀合物(D)。将浓度为约0.5mg/ml的250 μ l蛋白/多肽被施加到Äkta 纯化仪系统上的用PBS平衡的Superdex S20010/300GL柱。利用UV-900UV/VIS检测仪(GE Healthcare)监测225nm、280nm和494nm处的吸收,并且将每个色谱图主峰标准化为值等于1。箭头指示柱的孔隙体积(7.3ml)。

[0305] (E-K)通过SEC和UV/VIS光谱对游离荧光素、生物合成的PA#1 (200) 聚合物/多肽及其荧光素缀合物的表征。3幅色谱图(从上至下)显示了纯化的PA#1 (200) (E)、化学化合物荧光素(F)和纯化的荧光素-PA#1 (200) 缀合物(G)。4个UV/VIS谱显示了纯化的His (6)-SUMO-PA#1 (200) 融合蛋白(H)、纯化的PA#1 (200) 聚合物/多肽(I)、游离荧光素(J)和纯化的荧光素-PA#1 (200) 缀合物(K)(所有都位于PBS中)。箭头指示SUMO(280nm)、PA#1 (200) (225nm)、和荧光素(494nm)的特征性吸收带/肩部。

[0306] (L)利用Superdex S20010/300GL柱,来自(A-G)的色谱图的校准曲线。将标记蛋白的分子量(MW)(抑肽酶:6.5kDa;细胞色素C:12.4kDa;碳酸酐酶:29.0kDa;牛血清白蛋白:66.3kDa;醇脱氢酶:150kDa; β -淀粉酶:200kDa;去铁铁蛋白:440kDa)的对数与它们的洗脱体积(x)作图,并进行直线拟合。根据所观察到的His (6)-SUMO-PA#1 (200) (10.81ml)、PA#1 (200) (11.51ml)、荧光素-PA#1 (200) (11.49ml)和荧光素(27.57ml)的洗脱体积,它们的表观分子大小被确定如下。His (6)-SUMO-PA#1 (200):215.6kDa,PA#1 (200):154.1kDa(真实质量:16.1kDa),荧光素-PA#1 (200):155.6kDa(真实质量:16.6kDa);SUMO:25.7kDa(真实质量:12.2kDa);荧光素:0.09kDa(真实质量:0.33kDa)。这些数据表明:与未修饰的化合物相比,与Pro/Ala多肽/聚合物的融合物为缀合的药物赋予了明显增大的流体动力学体积。

[0307] (M)通过电喷雾电离质谱(ESI-MS)对生物合成的PA#1 (200) 多肽/聚合物和类固醇化合物地高辛配基的化学缀合物进行表征。地高辛配基-PA#1 (200)的去卷积ESI-MS光谱显示出质量为16671.4Da,这基本符合地高辛配基-PA#1 (200) 缀合物的计算质量(16670.6Da)。

[0308] 图14:生物合成的PA#1 (200) 多肽/聚合物和小分子药物的化学缀合物的说明。

[0309] (A)与生物合成的PA#1 (200)的N末端偶联的荧光素。

[0310] (B)与生物合成的PA#1 (200)的N末端偶联的地高辛配基。

实施例

[0311] 通过以下示例性的非限制性实施例进一步描述本发明,这些实施例为本发明及其诸多优势提供了更好的理解。

[0312] 除非另外指明,按照例如Sambrook (2001) (同上)所述,使用已有的重组基因技术方法。

[0313] 实施例1:Pro/Ala氨基酸聚合物/多肽的基因合成。

[0314] 如上文所述,由Pro和Ala残基组成的氨基酸重复在本文中被称为Pro/Ala或“PA”。通过图1中所示的两个互补寡脱氧核苷酸(SEQ ID NO:17和SEQ ID NO:18)的杂交来获得编码包含Pro和Ala的重复聚合物序列(PA#1,其对应于SEQ ID NO:1)的基因片段,然后通过它们的互相配合、但非回文的粘性末端的DNA连接以定向方式形成多连体。寡脱氧核苷酸购自ThermoScientific(Ulm,Germany)并通过制备型尿素聚丙烯酰胺凝胶电泳进行纯化。寡脱氧核苷酸的核酸序列如图1所示(SEQ ID NO:17和18,其包含额外的丙氨酸GCC密码子,通过与相应的粘性末端连接,该密码子成为下一个PA#1序列重复的一部分。通过以下进行酶促磷酸化:将200pmol的两种寡脱氧核苷酸在100μl 50mM Tris/HCl(pH 7.6)、10mM MgCl₂、5mM DTT、1mM ATP中混合,并在37°C下、在10u多核苷酸激酶(MBI Fermentas,St.Leon-Rot,Germany)的存在下孵育30分钟。80°C下变性10分钟后,将混合物冷却至室温过夜以实现杂交。然后,通过以下使50μl的该溶液进行连接反应:添加1u T4DNA连接酶(MBI Fermentas)和10μl 100mM Tris/HCl(pH 7.4)、50mM MgCl₂、20mMDTT、10mMATP并且在某些情况下添加各5mM的dATP、dCTP、dGTP和dTTP,总体积100μl,并在冰上孵育55分钟。70°C下热灭活10分钟后,在TAE缓冲液(40mM Tris、20mM乙酸、1mM EDTA)的存在下通过1.5%(w/v)琼脂糖凝胶电泳使连接产物分开。溴化乙锭染色后,对应于长度为300bp的组装基因区段的条带被切掉并分离。

[0315] 实施例2:作为Fab-PA#1融合蛋白的表达载体的pFab-PA#1(200)的构建。

[0316] 为了克隆编码来自实施例1的PA#1的20个氨基酸的序列的合成基因片段的10聚体重复,使用质粒载体pASK88-Fab-2xSapI(SEQ ID NO:22),其是Fab片段的表达质粒(Schlapschy(2007)Protein Eng.Des.Sel.20:273-284),携带在轻链的3'端具有处于反向互补方向的两个SapI限制性位点的核苷酸序列(图2A)。用SapI切割该载体(其是pASK75(Skerra,A.(1994)Gene 151:131-135)的衍生体),用虾碱性磷酸酶(USB,Cleveland,OH)去磷酸化,并与获自实施例1的300bp的合成DNA片段盒连接。用SapI再次切割得到的中间体质粒pFab-PA#1(100),用虾碱性磷酸酶去磷酸化,并与获自实施例1的300bp的合成DNA片段盒连接(如图2B的举例说明,但仅用了一个PA#1(20)聚合物/多肽盒)。得到的质粒被指定为pFab-PA#1(200)(SEQ ID NO:28)(图2C)。应当注意到,在该质粒上,200个残基的PA#1序列重复的编码区的侧翼为两个SapI限制性位点,这使得能够准确切除并进一步亚克隆携带5'-GCC核苷酸突出端的整个序列盒。

[0317] 大肠杆菌XL 1-Blue(Bullock(1987)Biotechniques 5:376-378)转化后,制备出质粒,并利用BigDye™终止试剂盒以及能从两侧进行测序的寡脱氧核苷酸引物,通过限制性分析和双链DNA测序(ABI-Prism™310Genetic analyzer,Perkin-Elmer Applied Biosystems,Weiterstadt,Germany)确认克隆的合成核酸插入物的序列。

[0318] 实施例3:作为PA#1(200)-IFNa2b融合蛋白的表达载体的pASK-PA#1(200)-IFNa2b的构建。

[0319] 为了构建编码与200个残基的PA#1序列重复PA#1(200)的融合物的IFNa2b的表达质粒,用SapI切割pASK-IFNa2b(SEQ ID NO:32)(图2D),用虾碱性磷酸酶去磷酸化,并与编码200个残基的PA#1多肽的基因片段连接切下来的(如图2E的举例说明,但仅用了一个PA#1(20)聚合物/多肽盒),所述基因片段是通过SapI的限制性消化从之前构建的质粒pFab-PA#

1 (200) (实施例2)。大肠杆菌JM83转化后 (Yanisch-Perron. (1985) Gene 33:103-119), 制备出质粒, 并通过限制性分析验证是否存在正确的插入物。得到的质粒被指定为pPA#1 (200)-IFNa2b (SEQ ID NO:37) (图2F)。

[0320] 实施例4: Fab片段和基因编码的PA#1聚合物/多肽的融合蛋白的细菌产生和纯化。

[0321] 利用含有100mg/l氨苄青霉素和30mg/l氯霉素的2L LB培养基和振荡培养、在22°C下、在大肠杆菌KS272中产生Fab片段 (计算质量: 48.0kDa) 和Fab-PA#1 (200) 融合蛋白 (计算质量: 64.3kDa), 所述大肠杆菌KS272携带来自实施例3的相应的表达质粒, 以及折叠辅助质粒pTUM4 (Schlupschy (2006) Protein Eng. Des. Sel. 19:385-390)。通过添加0.4mg无水四环素在OD₅₅₀=0.5的情况下过夜诱导重组基因表达 (通常在收获时的OD₅₅₀为约1.0)。按照其它人的描述 (Breustedt (2005) Biochim. Biophys. Acta 1764:161-173), 在500mM蔗糖、1mM EDTA、含50μg/ml溶解酵素的100mM Tris/HCl (pH 8.0) 的存在下进行周质提取, 然后借助于His6标签, 利用固定的金属亲和层析 (Skerra (1994) Gene 141:79-84), 进行纯化, 咪唑梯度为0至200mM, 配制于500mM甜菜碱、50mM磷酸钠 (pH 7.5) 中。

[0322] 获得了两种重组Fab片段的均匀蛋白制备物 (图3A), 未融合的Fab的产率为0.2mg L⁻¹OD⁻¹, Fab-PA#1 (200) 的产率为0.1mg L⁻¹OD⁻¹。利用高摩尔浓度Tris缓冲体系 (Fling (1986) Anal. Biochem. 155:83-88) 进行SDS-PAGE。对于未融合的Fab及其PA#1聚合物融合蛋白, 利用68290M⁻¹cm⁻¹的计算的消光系数 (Gill (1989) Anal. Biochem. 182:319-326), 按照280nm下的吸收测定蛋白浓度, 因为Pro/Ala聚合物由于缺少芳香氨基酸而不会对UV吸收做出贡献。

[0323] 实施例5: IFNa2b和基因编码的PA#1聚合物/多肽的融合蛋白的细菌产生和纯化。

[0324] 利用含有100mg/l氨苄青霉素和30mg/l氯霉素的2L LB培养基和振荡培养、在22°C下、在大肠杆菌KS272中产生IFNa2b (计算质量: 20.9kDa) 和PA#1 (200)-IFNa2b (计算质量: 37.0kDa), 所述大肠杆菌KS272携带来自实施例3的相应的表达质粒, 以及折叠辅助质粒pTUM4 (Schlupschy (2006) 同上)。通过添加0.4mg无水四环素在OD₅₅₀=0.5的情况下过夜诱导重组基因表达 (通常在收获时的OD₅₅₀为约1.0)。按照其它人的描述 (Breustedt (2005) 同上), 在500mM蔗糖、1mM EDTA、含50μg/ml溶解酵素的100mM Tris/HCl (pH 8.0) 的存在下进行周质提取, 然后通过Strep标签II、利用链霉亲和素亲和层析 (Schmidt (2007) Nat. Protoc. 2:1528-1535), 在150mM NaCl、1mM EDTA、100mM Tris/HCl (pH 8.0) 的存在下, 进行纯化。

[0325] 获得了两种重组IFNa2b蛋白的均匀蛋白制备物 (图3B), IFNa2b的产率为0.15mg L⁻¹OD⁻¹, PA#1 (200)-IFNa2b的产率为0.1mg L⁻¹OD⁻¹。利用高摩尔浓度Tris缓冲体系 (Fling (1986) 同上) 进行SDS-PAGE。对于未融合的IFNa2b及其PA#1聚合物融合蛋白, 利用23590M⁻¹cm⁻¹的计算的消光系数 (Gill (1989) 同上), 按照280nm下的吸收测定蛋白浓度。

[0326] 实施例6: 通过分析凝胶过滤测量Fab片段和基因编码的200个残基的PA#1聚合物的重组融合蛋白的流体动力学体积。

[0327] 在Superdex S200HR 10/300GL柱 (GE Healthcare Europe, Freiburg, Germany) 上进行尺寸排阻色谱 (SEC), 流速为ml/min, 利用Äkta Purifier 10系统 (GE Healthcare), 利用PBS (115mM NaCl、4mM KH₂PO₄、16mM Na₂HPO₄; pH 7.4) 作为工作缓冲液。分别施加250μl的纯化的Fab片段及其200个残基的PA#1融合物样品, 所述样品的浓度为0.25mg/ml, 配制于

PBS中,所述纯化的Fab片段及其200个残基的PA#1融合蛋白获自实施例4中所述的金属亲和层析。两种蛋白洗脱为图4A所示的单个同质峰。

[0328] 为了柱校准(图4B),施加下述球形蛋白(Sigma,Deisenhofen,Germany)的250 μ l合适的混合物,所述蛋白配制于PBS中,蛋白浓度为0.2mg/ml至0.5mg/ml:细胞色素C:12.4kDa;碳酸酐酶:29.0kDa;卵白蛋白:43.0kDa;牛血清白蛋白:66.3kDa;醇脱氢酶:150kDa; β -淀粉酶:200kDa;去铁铁蛋白:440kDa。

[0329] 作为结果,具有200个残基的PA#1聚合物/多肽的融合蛋白表现出的大小显著大于具有相同分子量的对应球形蛋白。相比于未融合的Fab片段,Fab-PA#1(200)的表观尺寸增加7.4倍,而真实质量仅增大1.3倍。这样的结果明确指示本发明的Pro/Ala多肽区段赋予生物活性Fab片段明显增加的流体动力学体积。

[0330] 实施例7:通过分析凝胶过滤测量IFNa2b和基因编码的具有200个残基的PA#1聚合物的重组融合蛋白的流体动力学体积。

[0331] 在Superdex S200HR 10/300GL柱(GE Healthcare)上对IFNa2b和PA#1(200)-IFNa2b进行尺寸排阻色谱,流速为1ml/min,利用ÄktaPurifier10系统(GE Healthcare),这与实施例6所述相似。两种蛋白洗脱为图4C所示的单个同质峰。

[0332] 结果,具有200个残基的PA#1聚合物/多肽的融合蛋白表现出的大小显著大于具有相同分子量的对应球形蛋白(图4D)。相比于未融合的IFNa2b蛋白,PA#1(200)-IFNa2b的表观尺寸增加10.2倍,而真实质量仅增大1.8倍。这样的结果明确指示本发明的Pro/-Ala聚合物/多肽赋予生物活性干扰素明显增加的流体动力学体积。

[0333] 实施例8:通过圆二色光谱检测与Fab片段融合的生物合成的PA#1聚合物的无规卷曲构象。

[0334] 利用配备了石英比色杯106-QS(0.1mm光程长;Hellma,Müllheim,Germany)的J-810分光偏振计(Jasco,Groß-Umstadt,Germany)分析二级结构。利用获自实施例4的配制于50mM K₂SO₄,20mM磷酸钾(pH 7.5)的3.12至15.4 μ M的蛋白溶液,在室温下通过累积16次运行(带宽1nm,扫描速度100nm/min,响应时间4s)记录190至250nm的光谱。进行溶液空白校正后,利用仪器软件将光谱平滑化,并且按照以下方程计算摩尔椭圆率 Θ_M :

$$[0335] \quad \Theta_M = \frac{\Theta_{obs}}{c \cdot d}$$

[0336] 其中 Θ_{obs} 表示测得的椭圆率,c表示蛋白浓度[mol/l],d表示石英比色杯的光程长[cm]。利用Kaleidagraph(Synergy Software,Reading,PA)将 Θ_M 值对波长作图。

[0337] 对于重组Fab,测得的圆二色(CD)光谱符合 β 折叠优势免疫球蛋白折叠,而Fab-PA#1(200)融合蛋白的光谱显示出无规卷曲构象的重要贡献(图5A)。为了更详细地分析Pro/Ala多肽区段的光谱贡献,通过将前一光谱从Fab-PA#1(200)的光谱中减除,针对未融合的Fab片段计算了摩尔差异CD光谱(图5B)。结果,观察到200nm附近的强的最小值,这是无规卷曲构象的特征。因此,作为重组融合蛋白的一部分的Pro/Ala序列似乎在生理缓冲液条件下以无规卷曲聚合物存在。

[0338] 实施例9:通过圆二色光谱检测与IFNa2b融合的基因编码的PA#1聚合物的无规卷曲构象。

[0339] 按照实施例8所述,使用3.6至38.7 μ M蛋白溶液,通过CD测量结果分析IFNa2b和PA#

1 (200)-IFNa2b (获自实施例5) 的二级结构。PA#1 (200)-IFNa2b的光谱显示出 α 螺旋二级结构的重要贡献,这表明干扰素的已知的 α 螺旋成束折叠,并且还显示出无规卷曲构象的重要贡献(图5C)。为了更详细地分析Pro/Ala聚合物融合伴侣的光谱贡献,通过减去两个分别光谱,针对未融合的IFNa2b,计算摩尔差异CD光谱(图5D)。结果,观察到200nm附近的强的最小值,这是无规卷曲构象的特征。因此,作为重组融合蛋白的一部分的Pro/Ala多肽区段似乎在水性缓冲液条件下以无规卷曲聚合物存在。

[0340] 实施例10: Fab片段、IFNa2b以及它们的200个残基的PA#1聚合物融合蛋白的二级结构的定量分析。

[0341] 利用二级结构去卷积程序CDNN ver. 2.1 (Böhm (1992) Protein Eng. 5:191-195), 根据在实施例8和9测得的相应的CD光谱, 分别定量Fab片段、Fab-PA#1 (200)、IFNa2b和PA#1 (200)-IFNa2b的二级结构含量, 其中使用一套33个基础光谱用于复合CD光谱的去卷积。该分析的结果提供在下表中:

[0342]

	Fab	Fab-PA#1(200)	差异: PA#1(200)	IFNa2b	PA#1(100) -IFNa2b	差异: PA#1(200)
α 螺旋	9.5%	7.5%	2.1%	38.2%	31.0%	0.7%
反平行 β 折叠	40.4%	3.1%	0%	1.8%	0.2%	4.6%
平行 β 折叠	6.9%	1.3%	0.3%	8.4%	0.7%	0.6%
β 转角	6.2%	50.4%	78.6%	19.2%	75.2%	69.7%
无规卷曲	37.2%	63.4%	94.8%	35.9%	64.4%	97.5%
Σ 全部	100.2%	125.8%	175.8%	103.5%	171.4%	170.0%
Σ β 转角和无	43.4%	113.8%	173.4%	55.1%	139.6%	169.1%

[0343]

规卷曲						
-----	--	--	--	--	--	--

[0344] 与重组Fab片段的主要 β 折叠二级结构含量(这符合其已知的免疫球蛋白折叠(参见Eigenbrot (1993) J. Mol. Biol. 229:969-995))相比, 如果PA#1聚合物与Fab片段融合, 则无组织构象(包括无规卷曲和 β 转角)的比例明显增加。Pro/Ala多肽区段的差异CD光谱显示出明确的无规卷曲构象。二级结构的分析显示出存在高比例的无组织构象(包括无规卷曲和 β 转角), 这几乎占了全部二级结构的100%。相似地, 与重组IFNa2b的主要 α 螺旋二级结构含量(这符合其已知的 α 螺旋成束蛋白三维结构(Radhakrishnan (1996) Structure 4:1453-1463))相比, 如果PA#1聚合物与IFNa2b融合, 则完整蛋白的无组织构象的比例明显增加。Pro/Ala多肽区段的差异CD光谱显示出明确的无规卷曲构象。二级结构分析显示出存在高比例的无组织构象(包括无规卷曲和 β 转角), 这几乎占了全部二级结构的100%。

[0345] 当使用Chou-Fasman算法(Chou and Fasman (1974) Biochemistry 13:222-245)对PA#1聚合物序列进行理论分析时, 获得的结果不同。该分析的结果示于图7中。该算法预

测100%的 α 螺旋二级结构,这与实验数据明显相反。因此,该算法不能用于可靠地预测本发明氨基酸聚合物的无组织构象。

[0346] 实施例11:构建作为His₆-PA#1 (200)-hGH融合蛋白的表达载体的pASK75-His₆-PA#1 (200)-hGH。

[0347] 为了构建编码作为与200个残基的PA#1序列重复PA#1 (200)的融合物的hGH的表达质粒,用SapI切割pASK75-His₆-hGH (SEQ ID NO:41) (图6A),用虾碱性磷酸酶去磷酸化,并与编码200个残基的PA#1多肽的基因片段连接,所述基因片段是通过SapI限制性消化从之前构建的质粒pFab-PA#1 (200) (实施例2)中切下来的(如图6B的举例说明,但仅用了一个PA#1 (20) 聚合物/多肽盒)。大肠杆菌JM83转化后 (Yanisch-Perron. (1985) 同上),制备出质粒,通过限制性分析验证是否存在正确的插入物。得到的质粒被指定为pASK75-His₆-PA#1 (200)-hGH (SEQ ID NO:46) (图6C)。

[0348] 实施例12:构建用于在中国仓鼠卵巢细胞中分泌产生与200个残基的PA#1聚合物/多肽融合的人生长激素的表达载体。

[0349] 用NheI和HindIII切割载体pASK75-His₆-PA#1 (200)-hGH (SEQ ID NO:46),该载体是pASK75 (Skerra (1994) 同上)的衍生物,允许hGH PA#1融合蛋白的原核产生。通过琼脂糖凝胶电泳纯化该片段,并与相应切割的载体pCHO (SEQ ID NO:50)连接。大肠杆菌XL1-Blue (Bullock (1987) 同上)转化后,制备出质粒,并通过限制性分析证实片段的正确插入。得到的质粒被指定为pCHO-PA#1 (200)-hGH (SEQ ID NO:48)并显示在图6D中,其编码与His₆标签、PA#1 (200)多肽区段和人生长激素(hGH)融合的hGH信号肽。

[0350] 实施例13:人生长激素(hGH)和基因编码的PA#1聚合物的融合蛋白在CHO细胞中的分泌产生。

[0351] 在100mm塑料培养皿中,用Quantum 263培养基(PAA Laboratories, Cölbe, Germany)培养CHO-K1细胞(ATCC号CCL-61),直至达到50%汇合。利用Nanofectin试剂盒(PAA Laboratories, Cölbe, Germany),用8 μ g pCHO-PA#1 (200)-hGH (SEQ ID NO:48)或作为对照的pCHO-hGH (SEQ ID NO:49)转染细胞,pCHO-hGH是编码hGH的相似的质粒,但不具有PA#1 (200)序列。6h后,用7ml Opti-MEM®-I还原的血清培养基(Invitrogen, Darmstadt, Germany)更换细胞培养基,并在37°C下、含5%CO₂的湿润环境中孵育细胞。两天后,获取20 μ l的细胞培养物上清,并用5 μ l含 β -巯基乙醇的SDS-PAGE上样缓冲液稀释。在95°C下加热5分钟后,将15 μ l的每种样品进行12%SDS-PAGE。利用半干式印迹装置电转移到硝酸纤维素膜(Schleicher & Schuell, Dassel, Germany)之后,用10ml PBST (含0.1%v/v吐温20的PBS)将膜洗涤3次,持续15分钟。用10ml稀释度为1:1000的与辣根过氧化物酶缀合的抗人生长激素抗体ab1956 (Abcam, Cambridge, UK)孵育膜。孵育1h、用20ml PBST将膜洗涤2次持续5分钟并用PBS将膜洗涤2次持续5分钟之后,在15ml SIGMAFAST™ 3,3'-二氨基联苯胺溶液(Sigma-Aldrich Chemie, Munich, Germany)的存在下进行显色反应。通过水洗和将膜风干来终止反应。对于两种重组蛋白样品,印记都显示出信号(图6E),从而证实了具有PA#1多肽的hGH融合蛋白在CHO细胞中分泌产生。

[0352] 实施例14:hGH和基因编码的PA#1聚合物/多肽的融合蛋白的细菌产生和纯化。

[0353] 在大肠杆菌KS272中产生人生长激素(hGH) (计算质量:23.4kDa)、PA#1 (200)-hGH (计算质量:39.6kDa)、PA#1 (400)-hGH (计算质量:55.8kDa)和PA#1 (600)-hGH (计算质量:

72.0kDa), 大肠杆菌KS272携带来自实施例11的相应的表达质粒或它们的分别具有双倍(编码400个残基)或三倍(600个残基)PA#1序列盒的衍生物。利用含2.5g/L葡萄糖、0.5g/L脯氨酸和100mg/l氨苄青霉素的2L LB培养基和振荡培养,在22°C下进行细菌产生。通过添加0.4mg无水四环素,在 $OD_{550}=0.5$ 的情况下诱导重组基因表达3h。按照其它人的描述(Breustedt (2005) 同上),在500mM蔗糖、1mM EDTA、含50 μ g/ml溶解酵素的100mM Tris/HCl (pH 8.0)的存在下进行周质提取,然后通过His₆标签、利用HisTrap高效亲和柱(GEHealthcare)进行纯化,用40mM磷酸钠(pH 7.5)、0.5M NaCl作为缓冲液。利用0至150mM的咪唑浓度梯度(溶解于工作缓冲液并用HCl调节至pH7.5)洗脱蛋白,并通过尺寸排阻色谱、利用用PBS (115mM NaCl, 4mMKH₂PO₄, 16mM Na₂HPO₄, pH 7.4)平衡的Superdex 200-HR10/30柱(GEHealthcare)进一步纯化蛋白。

[0354] 尺寸排阻色谱后,对于所有重组hGH融合蛋白均获得均匀的蛋白制备物,而没有聚集的迹象,hGH的产率为1mg L⁻¹OD⁻¹,PA#1 (200)-hGH的产率为0.3mg L⁻¹OD⁻¹,PA#1 (400)-hGH的产率为0.3mg L⁻¹OD⁻¹,PA#1 (600)-hGH的产率0.2mg L⁻¹OD⁻¹。利用高摩尔浓度Tris缓冲体系(Fling (1986) 同上)进行SDS-PAGE。对于未融合的hGH和其所有PA#1多肽融合物,利用16050M⁻¹cm⁻¹的计算的消光系数(Gill (1989) 同上),按照280nm下的吸收测定蛋白浓度。

[0355] 实施例15:利用表面等离子体共振测定人生长激素及其PA#1聚合物融合蛋白对人生长激素受体的细胞外结构域的结合亲和力。

[0356] 通过在Biacore 2000系统(GE Healthcare)上的表面等离子体共振(SPR)实时测量来测定hGH及其PA#1多肽融合蛋白对人生长激素受体Fc融合蛋白(hGHR-Fc; R&D Systems)的亲和力。首先,利用胺偶联试剂盒(GEHealthcare),将15 μ l配制于10mM醋酸钠(pH 5.0)中浓度为100 μ g/ml的小鼠抗人IgG-Fc捕获抗体(Jackson Immuno Research)固定于CMDP芯片(XanTec bioanalytics)的两个流动通道的表面。这样获得了约2700个响应单元(RU)。用作为流动缓冲液的PBS/T(含0.05%(v/v)吐温20的PBS)进行平衡后,将2 μ g/ml hGHR-Fc以5 μ l/min的流速充入芯片的一个通道,直至达到约300RU的额外信号。然后,以不同浓度注射75 μ l配制于PBS/T中的hGH或其PA#1多肽融合蛋白,并且在20 μ l/min的连续缓冲液流下测量缔合相和解离相。为了再生,施加3次6 μ l的10mM甘氨酸/HCl (pH 2.7)脉冲。通过将针对无固定的受体的通道测量的相应信号和由几个缓冲液空白注射测定的平均基线减去,校正传感图(Myszka (1999) Mol. Recognit. 12:279-284)。利用BIA评估软件3.1版(GE Healthcare),按照1:1朗格缪尔结合模型,通过至少7个不同的样品注射的痕迹的整体拟合进行动力学数据评估。从hGH或其PA#1融合蛋白和人生长激素受体之间的复合体的动力学及衍生平衡常数的SPR测量所获得的值总结于下表中:

[0357]

hGH变体	$k_{on}[10^5 M^{-1} s^{-1}]$	$k_{off}[10^{-6} s^{-1}]$	$K_D[pM]$
hGH	10.2	10.6	10.4
PA#1 (200)-hGH	4.75	9.18	19.3
PA#1 (400)-hGH	3.26	14.0	42.9
PA#1 (600)-hGH	3.29	12.5	38.0

[0358] 这些数据表明:hGH与不同长度的PA#1多肽的融合不会显著干扰受体结合。与缺少PA#1多肽的重组hGH相比,所有hGH PA#1多肽融合蛋白保留5倍之内的受体结合活性。

[0359] 实施例16:检测Fab片段和基因编码的PA#1聚合物的重组融合蛋白的延长的血体内浆半衰期。

[0360] 按照下表对成年BALB/c小鼠 (SPF stock breeding; TU München, Freising, Germany) 进行静脉内注射:

[0361]	组别	A	B	D
	测试物	Fab	Fab-PA#1(200)	Fab-PA#1(600)
	给药途径	静脉内		
	剂量[mg/kg 体重]	5.0	5.0	5.0
	浓度[mg/ml]	1.0	1.0	1.0
[0362]	施用体积[ml/kg 体重]	5.0		
	每组的动物数	9	9	9
	采血时间点数	12	12	12
	动物数/采样时间点	3	3	3
	采血次数/动物	4/1	4/1	4/1

[0363] 按照给药当天记录的个体体重 (b.w.) 计算静脉内给予的测试物的总体积 (例如 20g 体重的动物接受 100 μ l 的 1mg/ml 测试物)。按照下表进行采血:

		注射后的采血时间点											
测试物	亚组	10 分钟	30 分钟	1 h	2 h	3 h	4 h	6 h	8 h	12 h	24 h	36 h	48 h
Fab Fab-PA#1(200) Fab-PA#1(600)	1	x			x			x			x		
		x			x			x			x		
		x			x			x			x		
	2		x			x			x			x	
			x			x			x			x	
			x			x			x			x	
	3			x			x			x			x
				x			x			x			x
				x			x			x			x

[0364]

[0365] 对于每种物质(测试物),总共注射9只动物(分为3个亚组1-3,每个亚组3只动物),每只动物提供不同时间点的4个样品。从尾静脉采集血液样品(约50μl),并在4℃下保存30

分钟。10000g和4°C下离心10分钟后,立即冷冻上清(血浆)并保存在-20°C。

[0366] 为了在ELISA中定量检测Fab融合蛋白,用50μl配制于50mMNaHCO₃ (pH 9.6)中的10 μg/ml的重组Her2/ErbB2胞外域抗原溶液将96孔微量滴定板 (Maxisorb,NUNC,Denmark)的孔在4°C下包被过夜。然后,用200μl配制于PBS中的3% (w/v) BSA将孔封闭1h,并用PBS/T (含0.1% (v/v) 吐温20的PBS) 洗涤3次。将血浆样品在含0.5% (v/v) 来自未处理的动物的小鼠血浆的PBS/T中连续稀释,并孵育1h。然后,用PBS/T将孔洗涤3次,并用50μl以1:1000稀释于PBS/T的抗人Cκ抗体碱性磷酸酶缀合物溶液孵育1h。用PBS/T洗涤2次并用PBS洗涤2次后,通过添加50μl配制于100mM Tris/HCl (pH 8.8)、100mM NaCl、5mMMgCl₂中的0.5μg/ml的磷酸对硝基苯酯作为底物,开始显色反应,25°C下保持15分钟后,测量405nm处的吸光度。通过比较测得的信号和标准曲线来定量血浆样品中Fab、Fab-PA#1 (200) 和Fab-PA#1 (600) 的浓度,针对相应的纯化蛋白以规定的浓度在PBS/T中连续稀释来确定标准曲线,其中PBS/T含0.5% (v/v) 来自未处理的动物的小鼠血浆。

[0367] 为了估计Fab、Fab-PA#1 (200) 和Fab-PA#1 (600) 的血浆半衰期,从ELISA测量结果测定每个时间点的浓度值c(t),并将浓度值c(t)对静脉内注射后的时间t作图。利用KaleidaGraph软件、假定双指数衰减、按照以下方程将这些数据进行数字拟合:

$$[0368] \quad c(t) = c_{\alpha} e^{-\ln 2 \frac{t}{\tau_{1/2}^{\alpha}}} + (c_0 - c_{\alpha}) e^{-\ln 2 \frac{t}{\tau_{1/2}^{\beta}}}$$

[0369] 其中 $\tau_{1/2}^{\alpha}$ 和 $\tau_{1/2}^{\beta}$ 分别是分布相α和清除相β的半衰期值。 c_0 是在0时间点时的总血液浓度,而 c_{α} 是分布相的浓度幅度。

[0370] 图8描述了BALB/c小鼠中3个测试物的药代动力学。尽管重组Fab显示了快速血液清除,清除半衰期仅为约1.3h,而Fab-PA#1 (200) 和Fab-PA#1 (600) 融合蛋白具有的半衰期延长了大于3倍和29倍,相应的值分别为约4.1h和38.8h。这些数据证实,由于与Pro/Ala聚合物/多肽的融合,Fab片段的体内血浆半衰期显著延长,从而随着氨基酸聚合物的长度增加,半衰期变得更长。

[0371] 实施例17:P1A1和P1A3氨基酸聚合物/多肽的基因合成和作为Fab-P1A1 (200) 和Fab-P1A3 (200) 融合蛋白的表达载体的pFab-P1A1 (200) 和pFab-P1A3 (200) 的构建。

[0372] 通过互补寡脱氧核苷酸对的杂交获得编码包含Pro/Ala多肽/聚合物P1A1 (SEQ ID NO:51) 和指定为PA#3的P1A3 (SEQ ID NO:3) 的重复聚合物序列的基因片段,其中对于P1A1,互补寡脱氧核苷酸对为SEQ ID NO:52和SEQ ID NO:53,对于P1A3,互补寡脱氧核苷酸对为SEQ ID NO:54和SEQ ID NO:55,正如实施例1所述。按照与实施例2中所述的pFab-PA#1 (200) 相似的方式来构建pFab-P1A1 (200) (Seq ID NO:58) 和pFab-P1A3 (200) (Seq ID NO:59),它们编码Fab片段,Fab片段在轻链(LC) (LC Fab-P1A1 (200) 的氨基酸序列:SEQ ID NO:56;LC Fab-P1A3 (200) 的氨基酸序列:SEQ ID NO:57) 的C末端处具有相应的200个残基的Pro/Ala聚合物/多肽区段。

[0373] 下文给出了SEQ ID NO:56、57、58和59。然而,这些序列还包含在后附序列表中,所述序列表是本公开和本发明描述的特定部分。

[0374] SEQ ID NO:56

[0375] Asp Ile Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly

[0376]	1	5	10	15
[0377]	Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala			
[0378]	20	25	30	
[0379]	Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile			
[0380]	35	40	45	
[0381]	Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg Phe Ser Gly			
[0382]	50	55	60	
[0383]	Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro			
[0384]	65	70	75	80
[0385]	Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr Thr Pro Pro			
[0386]	85	90	95	
[0387]	Thr Phe Gly Gln Gly Thr Lys Leu Glu Ile Lys Arg Thr Val Ala Ala			
[0388]	100	105	110	
[0389]	Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly			
[0390]	115	120	125	
[0391]	Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala			
[0392]	130	135	140	
[0393]	Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln			
[0394]	145	150	155	160
[0395]	Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser			
[0396]	165	170	175	
[0397]	Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr			
[0398]	180	185	190	
[0399]	Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser			
[0400]	195	200	205	
[0401]	Phe Asn Arg Gly Glu Cys Ser Ser Ala Pro Ala Pro Ala Pro Ala Pro			
[0402]	210	215	220	
[0403]	Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro			
[0404]	225	230	235	240
[0405]	Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro			
[0406]	245	250	255	
[0407]	Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro			
[0408]	260	265	270	
[0409]	Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro			
[0410]	275	280	285	
[0411]	Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro			
[0412]	290	295	300	
[0413]	Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro			
[0414]	305	310	315	320

[0415] Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
 [0416] 325 330 335
 [0417] Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
 [0418] 340 345 350
 [0419] Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
 [0420] 355 360 365
 [0421] Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
 [0422] 370 375 380
 [0423] Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
 [0424] 385 390 395 400
 [0425] Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
 [0426] 405 410 415
 [0427] Ala
 [0428] SEQ ID NO:57
 [0429] Asp Ile Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
 [0430] 1 5 10 15
 [0431] Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala
 [0432] 20 25 30
 [0433] Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
 [0434] 35 40 45
 [0435] Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg Phe Ser Gly
 [0436] 50 55 60
 [0437] Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
 [0438] 65 70 75 80
 [0439] Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr Thr Pro Pro
 [0440] 85 90 95
 [0441] Thr Phe Gly Gln Gly Thr Lys Leu Glu Ile Lys Arg Thr Val Ala Ala
 [0442] 100 105 110
 [0443] Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
 [0444] 115 120 125
 [0445] Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
 [0446] 130 135 140
 [0447] Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
 [0448] 145 150 155 160
 [0449] Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
 [0450] 165 170 175
 [0451] Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
 [0452] 180 185 190
 [0453] Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser

[0454]	195	200	205
[0455]	Phe Asn Arg Gly Glu Cys Ser Ser Ala Ala Ala Pro Ala Ala Ala Pro		
[0456]	210	215	220
[0457]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0458]	225	230	235
[0459]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0460]	245	250	255
[0461]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0462]	260	265	270
[0463]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0464]	275	280	285
[0465]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0466]	290	295	300
[0467]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0468]	305	310	315
[0469]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0470]	325	330	335
[0471]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0472]	340	345	350
[0473]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0474]	355	360	365
[0475]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0476]	370	375	380
[0477]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0478]	385	390	395
[0479]	Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro		
[0480]	405	410	415
[0481]	Ala		
[0482]	SEQ ID NO:58		
[0483]	accgcacacc atcgaaatggc cagatgatta attcctaatt tttgttgaca ctctatcatt	60	
[0484]	gatagagtta ttttaccact ccctatcagt gatagagaaa agtgaaatga atagttcgac	120	
[0485]	aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac	180	
[0486]	tggctggttt cgctaccgta gcgcaggccg aagttaaact gcaggaatcc ggtggtggtc	240	
[0487]	tggttcagcc aggtggttcc ctgcggctct cgtgtgctgc ttccggtttc aacatcaaag	300	
[0488]	acacctacat ccactgggtt cgtcaggctc cgggtaaagg cctggaatgg gttgctcgta	360	
[0489]	tctaccgcac caacggttac accaggatat cggattcagt taaaggctgt ttcaccatct	420	
[0490]	cggccgacac ttccaaaaac accgcttacc tccagatgaa ctccctgcgt gctgaagaca	480	
[0491]	cagctgttta ttattgctcc cgttgggggtg gtgacggttt ctacgctatg gactactggg	540	
[0492]	gtcagggtac cctggtcacc gtctctcag cctccaccaa gggcccatcg gtcttcccc	600	

[0493] tggcaccctc ctccaagagc acctctgggg gcacageggc cctgggctgc ctgggtcaagg 660
 [0494] actacttccc cgaaccggtg acgggtgtcgt ggaactcagg cgccctgacc agcggcgtgc 720
 [0495] acaccttccc ggctgtccta cagtctcag gactctactc cctcagcagc gtgggtgactg 780
 [0496] tgccctccag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca 840
 [0497] acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccacat caccattaat 900
 [0498] aaccatggag aaaataaagt gaaacaaagc actattgcac tggcactctt accgttactg 960
 [0499] tttaccctg tgacaaaagc cgacatcgag ctacaccaat ccccgctcctc cctgtccgct1020
 [0500] tccgttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgccgtal080
 [0501] gcgtggtatc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgcttccttcl140
 [0502] ctgtattccg gagttccgag caggttcagt gggtcccggt ccggtaccga cttcaccctgl200
 [0503] acgatactct cctccagcc ggaagacttc gctacctact actgtcaaca gcactacacc1260
 [0504] accccgccga ccttcgggtc gggtagcaaa ctcgagatca aacggactgt ggctgcacal320
 [0505] tctgtcttca tcttcccgcc atctgatgag cagttgaaat ctggaactgc ctctgtgtg1380
 [0506] tgctgtctga ataacttcta tcccagagag gccaaagtac agtggaaggt ggataacgcc1440
 [0507] ctccaatcgg gtaactccca ggagagtgtc acagagcagg acagcaagga cagcacctac1500
 [0508] agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacacaa agtctacgcc1560
 [0509] tgccaagtca cccatcaggg cctgagttcg cccgtcaca agagcttcaa ccgcggagag1620
 [0510] tgctcttctg cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct1680
 [0511] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct1740
 [0512] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct1800
 [0513] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct1860
 [0514] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct1920
 [0515] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct1980
 [0516] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct2040
 [0517] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct2100
 [0518] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct2160
 [0519] cctgcaccag cccctgctcc tgctccagca cctgcaccag cacctgctcc agcaccagct2220
 [0520] cctgcaccag cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacattgtg2280
 [0521] cgacatTTTT tttgtctgcc gtttaccgct actgcgtcac ggatctccac gcgccctgta2340
 [0522] gcggcgcatt aagcgcggcg ggtgtggtgg ttacgcgcag cgtgaccgct acacttgcca2400
 [0523] gcgccctagc gcccgctcct ttcgctttct tcccttcctt tctcgccacg ttcgccggct2460
 [0524] tccccgtca agctctaaat cgggggctcc ctttaggggt cggatttagt gctttacggc2520
 [0525] acctgaccc caaaaaactt gattaggggt atggttcacg tagtgggcca tcgccctgat2580
 [0526] agacggTTTT tgcaccttg acgttggagt ccacgttctt taatagtgga ctcttgttcc2640
 [0527] aaactggaac aacactcaac cctatctcgg tctattcttt tgatttataa gggattttgc2700
 [0528] cgatttcggc ctatttggtta aaaaatgagc tgatttaaca aaaatttaac gcgaatttta2760
 [0529] aaaaaatatt aacgtttaca atttcagggt gcacttttcg gggaaatgtg cgcggaaccc2820
 [0530] ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataaccct2880
 [0531] gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg2940

[0532] cccttattcc cttttttgcg gcattttgcc ttctgtttt tgctcaccca gaaacgctgg3000
[0533] tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc3060
[0534] tcaacagcgg taagatcctt gagagttttc gccccgaaga acgtttttcca atgatgagca3120
[0535] cttttaaagt tctgctatgt ggcgcggtat tateccgtat tgacgccggg caagagcaac3180
[0536] tcggtcgccg catacactat tctcagaatg acttggttga gtactcacca gtcacagaaa3240
[0537] agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata accatgagtg3300
[0538] ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt3360
[0539] ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg3420
[0540] aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc3480
[0541] gcaaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga3540
[0542] tggaggcgga taaagttgca ggaccacttc tgcgctcggc ctttccggct ggctggttta3600
[0543] ttgctgataa atctggagcc ggtgagcgtg gctctcgcg tatcattgca gcactggggc3660
[0544] cagatggtaa gccctccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg3720
[0545] atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat3780
[0546] taatgatgtc tcgttttagat aaaagtaaag tgattaacag cgcattagag ctgcttaatg3840
[0547] aggtcggaat cgaaggttta acaaccgta aactcgcca gaagctaggt gtagagcagc3900
[0548] ctacattgta ttggcatgta aaaaataagc gggctttgct cgacgcctta gccattgaga3960
[0549] tgttagatag gcaccatact cacttttgcc ctttagaagg ggaaagctgg caagattttt4020
[0550] tacgtaataa cgctaaaagt tttagatgtg ctttactaag tcatcgcgat ggagcaaaag4080
[0551] tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct4140
[0552] ttttatgcca acaaggtttt tccactagaga atgcattata tgcactcagc gcagtggggc4200
[0553] attttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaaggg4260
[0554] aaacacctac tactgatagt atgccgccat tattacgaca agctatcgaa ttatttgatc4320
[0555] accaaggtgc agagccagcc ttcttattcg gccttgaatt gatcatatgc ggattagaaa4380
[0556] aacaacttaa atgtgaaagt gggctcttaa agcagcataa cttttttccg tgatggtaac4440
[0557] ttcactagtt taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc4500
[0558] cttaacgtga gttttcgttc cactgagcgt cagaccccg agaaaagatc aaaggatctt4560
[0559] cttgagatcc tttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac4620
[0560] cagcgggtgt ttgtttgccc gatcaagagc taccaactct tttccgaag gtaactggct4680
[0561] tcagcagagc gcagatacca aatactgtcc ttctagtgtg gccgtagtta ggccaccact4740
[0562] tcaagaactc tgtagcaccc cctacatacc tcgctctgct aatcctgtta ccagtggctg4800
[0563] ctgccagtgg cgataagtcg tgtcttaccg ggttggaactc aagacgatag ttaccggata4860
[0564] aggcgcagcg gtcgggctga acgggggggt cgtgcacaca gccagcttg gagcgaacga4920
[0565] cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag4980
[0566] ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg5040
[0567] agcttccagg gggaaacgcc tggatatctt atagtctgt cgggtttcgc cacctctgac5100
[0568] ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca5160
[0569] acgcggcctt ttacggttc ctggcctttt gctggccttt tgctcacatg 5210
[0570] SEQ ID NO:59

[0571] acccgacacc atcgaatggc cagatgatta attcctaatt tttgttgaca ctctatcatt60
 [0572] gatagagtta ttttaccact ccctatcagt gatagagaaa agtgaaatga atagttcgac120
 [0573] aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac180
 [0574] tggctggttt cgctaccgta gcgcaggcgc aagttaaact gcaggaatcc ggtgggtggc240
 [0575] tggttcagcc aggtgggttc ctgcggctct cgtgtgctgc ttccggtttc aacatcaaag300
 [0576] acacctacat ccactgggtt cgtcaggctc cgggtaaagg cctggaatgg gttgctcgta360
 [0577] tctacccgac caacggttac accaggtatg ccgattcagt taaaggctgt ttcaccatct420
 [0578] cggccgacac ttccaaaaac accgcttacc tccagatgaa ctccctgcgt gctgaagaca480
 [0579] cagctgttta ttattgctcc cgttgggggtg gtgacggttt ctacgctatg gactactggg540
 [0580] gtcagggtac cctggtcacc gtctcctcag cctccacca gggcccatcg gtcttcccc600
 [0581] tggcaccctc ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggtcaagg660
 [0582] actacttccc cgaaccgggtg acgggtgtgt ggaactcagg cgccctgacc agcggcgtgc720
 [0583] acacttccc ggtgttecta cagtctcag gactctactc cctcagcagc gtggtgactg780
 [0584] tgccctccag cagcttgggc acccagacct acatctgcaa cgtaatcac aaaccagca840
 [0585] acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccaccat caccattaat900
 [0586] aaccatggag aaaataaagt gaaacaaagc actattgcac tggcactctt accgttactg960
 [0587] tttacccttg tgacaaaagc cgacatcgag ctacccaat ccccgctc cctgtccgt1020
 [0588] tccgttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgccgtal080
 [0589] gcgtggtatc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgcttcctcl140
 [0590] ctgtattccg gagttccgag caggttcagt ggttcccggt ccggtaccga cttaccctgl200
 [0591] acgatatect ccctccagcc ggaagacttc gctacctact actgtcaaca gcactacacc1260
 [0592] accccgccga ccttcgggtc ggggtaccaa ctcgagatca aacggactgt ggctgcaccal320
 [0593] tctgtcttca tcttcccgcc atctgatgag cagttgaaat ctggaactgc ctctgttgtl380
 [0594] tgctgtctga ataacttcta tcccagagag gccaaagtac agtgggaagg ggataacgcc1440
 [0595] ctccaatcgg gtaactcca ggagagtgtc acagagcagg acagcaagga cagcacctac1500
 [0596] agcctcagca gcaccctgac gctgagcaaa gcagactacg agaaacacaa agtctacgcc1560
 [0597] tgcgaagtca cccatcaggg cctgagttcg cccgtcaca agagcttcaa ccgcggagag1620
 [0598] tgctcttctg ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgcal680
 [0599] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgcal740
 [0600] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgcal800
 [0601] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgcal860
 [0602] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgcal920
 [0603] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgcal980
 [0604] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca2040
 [0605] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca2100
 [0606] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca2160
 [0607] gcagctccag ccgctgcacc tgctgcagca cctgctgcag ctccagcagc tgctcctgca2220
 [0608] gcagctccag cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacattgtg2280
 [0609] cgacattttt tttgtctgcc gtttaccgt actgcgtcac ggatctccac gcgccctgta2340

[0610] gcggcgcatt aagcgcggcg ggtgtggtgg ttacgcgcag cgtgaccgct acacttgcca2400
 [0611] gcgccttagc gcccgcctct ttcgctttct tcccttctt tctcgccacg ttcgccggct2460
 [0612] ttccccgtca agctctaaat cgggggctcc ctttaggggtt ccgatttagt gctttacggc2520
 [0613] acctcgaccc caaaaaactt gattaggggtg atggttcacg tagtgggcca tcgccctgat2580
 [0614] agacggtttt tcgccctttg acgttggagt ccacgttctt taatagtgga ctcttgttcc2640
 [0615] aaactggaac aacactcaac cctatctcgg tctattcttt tgatttataa gggattttgc2700
 [0616] cgatttcggc ctatttggtt aaaaatgagc tgatttaaca aaaatttaac gcgaatttta2760
 [0617] acaaaatatt aacgtttaca atttcagggtg gcacttttcg gggaaatgtg cgcggaaccc2820
 [0618] ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataaccct2880
 [0619] gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg2940
 [0620] cctttattcc cttttttgcg gcattttgcc ttctgtttt tgctcaccca gaaacgctgg3000
 [0621] tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc3060
 [0622] tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca3120
 [0623] cttttaaaagt tctgctatgt ggcgcggtat tatcccgat tgacgccggg caagagcaac3180
 [0624] tcggtcgccg catacactat tctcagaatg acttggttga gtactacca gtcacagaaa3240
 [0625] agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata accatgagt3300
 [0626] ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt3360
 [0627] ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg3420
 [0628] aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc3480
 [0629] gcaaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga3540
 [0630] tggagcgga taaagttgca ggaccacttc tgcgctcgcc ctttccggct ggctggttta3600
 [0631] ttgctgataa atctggagcc ggtgagcgtg gctctcgcg tatcattgca gcactggggc3660
 [0632] cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg3720
 [0633] atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat3780
 [0634] taatgatgtc tcgttttagat aaaagtaaag tgattaacag cgcattagag ctgcttaatg3840
 [0635] aggtcggaat cgaaggttta acaaccgta aactcgccca gaagctaggt gtagagcagc3900
 [0636] ctacattgta ttggcatgta aaaaataagc gggctttgct cgacgcctta gccattgaga3960
 [0637] tgttagatag gcaccatact cacttttgcc ctttagaagg ggaaagctgg caagattttt4020
 [0638] tacgtaataa cgctaaaagt ttttagatgtg ctttactaag tcatcgcgat ggagcaaaag4080
 [0639] tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct4140
 [0640] ttttatgcca acaaggtttt tcaactagaga atgcattata tgcactcagc gcagtggggc4200
 [0641] attttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaaggg4260
 [0642] aaacacctac tactgatagt atgcgcctat tattacgaca agctatcgaa ttatttgatc4320
 [0643] accaaggtgc agagccagcc ttcttattcg gccttgaatt gatcatatgc ggattagaaa4380
 [0644] aacaacttaa atgtgaaagt gggctcttaa agcagcataa ctttttccg tgatggtaac4440
 [0645] ttactagtt taaaaggatc taggtgaaga tccttttga taatctcatg accaaaatcc4500
 [0646] cttaacgtga gttttcgttc cactgagcgt cagaccccg agaaaagatc aaaggatctt4560
 [0647] cttgagatcc tttttttctg cgcgtaatct gctgcttgc aacaaaaaaa ccaccgctac4620
 [0648] cagcgggtgt ttgtttgcg gatcaagagc taccaactct ttttccgaag gtaactggct4680

[0649] tcagcagagc gcagatacca aatactgtcc ttctagtgtgta gccgtagtta ggccaccact4740
[0650] tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg4800
[0651] ctgccagtgg cgataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata4860
[0652] aggcgagcgc gtcgggctga acggggggtt cgtgcacaca gccagcttg gagcgaacga4920
[0653] cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag4980
[0654] ggagaaaggc ggacaggtat ccggttaagcg gcagggtcgg aacaggagag cgcacgaggg5040
[0655] agcttccagg gggaaacgcc tggatatcttt atagtcctgt cgggtttcgc cacctctgac5100
[0656] ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca5160
[0657] acgcggcctt tttacggttc ctggcctttt gctggccttt tgctcacatg 5210

[0658] 实施例18:通过分析凝胶过滤测量Fab片段和基因编码的P1A1或P1A3多肽/聚合物的重组融合蛋白的流体动力学体积。

[0659] 利用Äkta Purifier 10系统(GE Healthcare),用PBS作为工作缓冲液,在Superdex S200HR 10/300GL柱(GE Healthcare Europe, Freiburg, Germany)上进行SEC,流速为1ml/min。分别施加250 μ l的Fab-P1A1 (200)和Fab-P1A3 (200)融合蛋白样品,样品浓度为0.25mg/ml,配制于PBS中,所述样品的产生和纯化(图9)与实施例4中关于Fab-PA#1 (200)所述相似。两种蛋白洗脱为图10所示的单个同质峰。

[0660] 结果,具有200个残基的P1A1或P1A3聚合物/多肽的融合蛋白表现出的大小显著大于相应的未融合Fab片段。与Fab片段相比,Fab-P1A1 (200)和Fab-P1A3 (200)的表观尺寸分别增加5.8倍和5.2倍(参见图4B),而真实质量仅增大1.4倍和1.3倍。该结果明确表明,本发明的生物合成的P1A1和P1A3多肽区段赋予生物活性Fab片段明显增加的流体动力学体积。

[0661] 实施例19:通过圆二色(CD)光谱检测与Fab片段融合的生物合成的P1A1和P1A3聚合物/多肽的无规卷曲构象。

[0662] 按照实施例8中所述记录Fab-P1A1 (200)和Fab-P1A3 (200)的CD光谱,分别利用4.2和6.5 μ M的蛋白溶液,所述蛋白溶液的制备与实施例4所述相似,利用50mM K₂SO₄、20mM磷酸钾(pH 7.5)作为缓冲液。

[0663] Fab-P1A1 (200)和Fab-P1A3 (200)融合蛋白的光谱显示了大比例的无规卷曲构象(图11A)。为了更详细地分析Pro/Ala多肽区段的光谱贡献,通过在标准化至相同摩尔浓度之后将前一光谱分别从Fab-P1A1 (200)和Fab-P1A3 (200)的光谱减去来针对未融合的Fab片段计算摩尔差异CD光谱(参见实施例8)(图11B)。作为结果,观察到波长约200nm处的强的最小值,这是无规卷曲构象的特征。因此,作为重组融合蛋白的一部分,P1A1和P1A3序列似乎在生理缓冲液条件下以无规卷曲构象存在。

[0664] 实施例20:作为His (6)-SUMO-PA#1 (200)融合蛋白的表达载体的pSUMO-PA#1 (200)的构建。

[0665] 为了构建编码6残基的His标签和与200个残基的PA#1序列重复融合的小泛素样修饰蛋白(SUMO)(Panavas (2009) Methods Mol. Biol. 497:303-17)的表达质粒,通过聚合酶链式反应(PCR)从克隆的cDNA扩增来自酿酒酵母(*Saccharomyces cerevisiae*)的SUMO蛋白[也称为Smt3p; Uniprot: Q12306]。5'引物在PCR产物中引入NdeI限制性位点,其含有Met起始密码子(ATG)和另一Lys密码子,以及His₆标签编码序列,而3'引物在PCR产物中引入HindIII和SapI限制性位点。用NdeI和HindIII消化所得的DNA片段,并与相应消化的质粒

pSA1衍生物 (Schmidt (1994) J.Chromatogr. 676:337-345) 连接,在所述质粒pSA1衍生物中, SapI限制性位点通过沉默突变已被去除。用SapI切割得到的质粒,用虾碱性磷酸酶去磷酸化,并与编码200个残基的PA#1多肽区段的基因片段连接,所述基因片段是通过SapI限制性消化(已与图2E中举例说明的方式相似)从质粒pFab-PA#1 (200) (如实施例2所述)切下来的。得到的质粒被指定为pSUM0-PA#1 (200) (SEQ ID NO:60),并且描述在图12A中。

[0666] 实施例21:基因编码的PA#1 (200) 聚合物/多肽的细菌表达和分离。

[0667] PA#1 (200) 多肽(计算质量:16.1kDa)在大肠杆菌BLR (DE3) (NEB, Ipswich, MA, USA) 的细胞质中初始产生为与小泛素样修饰蛋白 (SUMO) (计算质量:12.2kDa) 的融合蛋白,所述大肠杆菌BLR (DE3) 携带表达质粒pSUM0-PA#1 (200) (如实施例21中所述) 以及抑制T7启动子的质粒pLysE (Studier (1991) J.Mol.Biol. 219:37-44)。利用2L含2.5g/L D-葡萄糖、0.5g/L L-脯氨酸、100mg/l氨苄青霉素和30mg/l氯霉素的LB培养基和振荡培养,在30°C下进行细菌产生。通过添加异丙基-β-D-硫代半乳糖苷 (IPTG) 至0.5mM的终浓度来诱导重组基因表达。诱导后3h收获细菌,将其重悬在100mM NaCl、40mM磷酸钠 (pH 7.5) 中,并使用弗氏压碎器 (Thermo Scientific, Waltham, MA, USA) 来裂解。裂解物离心 (15分钟, 15000g) 后,未观察到内含体。

[0668] 将含有可溶性融合蛋白的上清在70°C下孵育15分钟,并进行离心 (15分钟, 15000g) 来去除热不稳定的宿主细胞蛋白。利用与Äkta纯化系统 (GE Healthcare) 连接的12ml Ni²⁺充电的HisTrap高效柱 (GE Healthcare)、通过IMAC (Skerra (1994) Gene 141:79-84) 从上清纯化His (6) -SUM0-PA#1 (200) 融合蛋白,并用配制于500mM NaCl、40mM磷酸钠 (pH 7.5) 中的0至150mM的咪唑梯度进行洗脱。在随后的制备型SEC步骤之后,获得了His (6) -SUM0-PA#1 (200) 融合蛋白的同质制备物 (图12B),产率为约5mg/L细菌培养物 (OD550=1)。对于His (6) -SUM0-PA#1 (200) 多肽融合蛋白,利用1280M⁻¹cm⁻¹的计算的消光系数 (Gill (1989), 同上),按照280nm下的吸收测定蛋白浓度。注意到,PA#1 (200) 多肽区段对于280nm下的吸收没有贡献,这是由于其缺少芳香氨基酸侧链或含硫氨基酸侧链。

[0669] 通过位点特异性蛋白水解切割 (Pro/Ala多肽区段之前的Gly-Gly基序的下游) 从融合蛋白释放生物合成的PA#1 (200) 多肽,所述位点特异性蛋白水解切割是利用2U/mg的来自酿酒酵母的Ubl特异性蛋白酶1 (Invitrogen, Carlsbad, CA, USA),在切割缓冲液 (0.2w/v% 胰加漂 (Igepal)、1mM DTT、150mM NaCl、50mM Tris-HCl (pH 8.0)) 中、在30°C下进行1h。利用高摩尔浓度Tris缓冲体系 (Fling (1986) Anal. Biochem. 155:83-88),通过SDS-PAGE (图12B) 检查切割过程。为了去除携带His₆标签的切割的His (6) -SUM0蛋白、残余的未切割的融合蛋白以及SUM0蛋白酶,其中利用5ml Ni²⁺充电的HisTrap高效柱 (GE Healthcare) 和作为工作缓冲液的500M NaCl、20mM磷酸盐 (pH 7.5),将反应混合物进行另一次IMAC。此时,通过流含有纯的生物合成的PA#1 (200) 多肽 (图13E)。注意到,按此方式制备的生物合成的PA#1 (200) 多肽/聚合物 (SEQ ID NO:61) 包含总共201个氨基酸残基,这是来自于10个连接的双链寡脱氧核苷酸构成部分 (每个编码20个氨基酸残基,如图1所示) 的编码组合基因产物和另外的一个Ala残基,所述Ala残基由用于克隆的下游SapI限制性位点的三联DNA突出端编码。

[0670] 实施例22:具有PA#1 (200) 的小分子/药物缀合物的制备和表征。

[0671] 将来自实施例21的His (6) -SUM0-PA#1 (200) 融合蛋白的未纯化的蛋白水解切割反

应混合物在4°C下针对50mM NaHCO₃ (pH 8.3) 进行两次透析,并且在与10倍摩尔过量的配制于无水二甲基甲酰胺(DMF)中的6-[荧光素-5(6)-酰胺基(carboxamido)]己酸N-羟基琥珀酰亚胺酯(荧光素-NHS酯;Sigma-Aldrich)溶液混合后,在室温下孵育1h。为此,将200μl 2.5mg/ml His(6)-SUMO-PA#1(200)切割混合物溶液添加至17.6μl溶解于DMF中的10mM荧光素-NHS酯溶液。将得到的混合物在室温下孵育1h,并如实施例21所述施加到IMAC,从而去除切割的His(6)-SUMO蛋白、残余的未切割的融合蛋白和SUMO蛋白酶,并利用制备型SEC、在用PBS平衡的Superdex S20010/300GL柱上、以0.5ml/min的流速进一步纯化。

[0672] 然后,通过分析SEC、在用PBS平衡的Superdex S20010/300GL柱上、以0.5ml/min的流速分析来自不同的步骤的样品。通过SUMO蛋白的芳香侧链、在280nm下检测SUMO蛋白;在225nm下检测肽键,包括those of the Pro/Ala多肽或多肽区段的肽键;而在494nm下检测荧光素(图13A-G)。为了比较,利用λ9仪器(Perkin Elmer, Waltham, MA, USA)测量游离荧光素溶液(Sigma-Aldrich)和来自SEC中检测到的每个不同峰的级分的UV/VIS谱(图13H-K)。为了层析柱的大小校准(图13L),将250μl下列球形蛋白(Sigma-Aldrich)的合适的混合物添加在PBS中,浓度为0.2至0.5mg/ml:抑肽酶:6.5kDa;细胞色素C:12.4kDa;碳酸酐酶:29.0kDa;牛血清白蛋白:66.3kDa;醇脱氢酶:150kDa;β-淀粉酶:200kDa;去铁铁蛋白:440kDa。

[0673] 结果,在生物合成的PA#1(200)多肽/聚合物与荧光素-NHS酯偶联后,通过IMAC和SEC分离到大分子缀合物,其基本表现出PA#1(200)多肽/聚合物的大小性质和小分子即荧光素基团的光谱学特征。这表明,小分子与生物合成的Pro/Ala多肽/聚合物成功地偶联,按照本发明,这能够明显增加缀合的小分子药物或化合物的流体动力学体积。

[0674] 为了制备生物合成的Pro/Ala多肽/聚合物和植物类固醇地高辛配基的相似缀合物,将0.1mg来自实施例21的纯化PA#1(200)多肽如上文所述针对50mM NaHCO₃ (pH 8.3)进行透析。按照205nm下的吸收测定纯化的PA#1(200)多肽的浓度(Gill(1989)同上)。将PA#1(200)多肽与10倍摩尔过量的地高辛配基-3-O-甲基羰基-ε-氨基己酸NHS酯(DIG-NHS酯;Roche Diagnostics, Mannheim, Germany)偶联。为此,将100μl配制于50mM NaHCO₃ (pH 8.3)的1mg/ml的纯化PA#1(200)多肽溶液添加至2μl溶解于无水DMF中的30mM的DIG-NHS酯溶液,并且将反应混合物在室温下孵育1h。利用截留值为7kDa的Zeba™离心脱盐柱(Thermo Scientific)纯化得到的缀合物溶液,针对10mM醋酸铵缓冲液(pH 6.8)进行两次透析,并通过ESI质谱、在Q-ToF Ultima设备(Waters, Eschbronn, Germany)上、利用阳离子模式进行分析。结果,地高辛配基-PA#1(200)缀合物的光谱显示出16671.4Da的质量,这基本符合16670.6Da的计算质量(图13M)。这明确表明,生物合成的Pro/Ala多肽/聚合物,特别是PA#1(200),能与小分子药物有效缀合。

[0675] 本发明涉及并引用以下示例性的序列,从而后附序列列表是作为描述的一部分来提供,并因此作为说明书的一部分。

[0676] SEQ ID NO:1显示了PA#1的氨基酸序列。

[0677] SEQ ID NO:2显示了PA#2的氨基酸序列。

[0678] SEQ ID NO:3显示了PA#3的氨基酸序列。

[0679] SEQ ID NO:4显示了PA#4的氨基酸序列。

[0680] SEQ ID NO:5显示了PA#5的氨基酸序列。

- [0681] SEQ ID NO:6显示了PA#6的氨基酸序列。
- [0682] SEQ ID NO:7显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0683] SEQ ID NO:8显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0684] SEQ ID NO:9显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0685] SEQ ID NO:10显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0686] SEQ ID NO:11显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0687] SEQ ID NO:12显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0688] SEQ ID NO:13显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0689] SEQ ID NO:14显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0690] SEQ ID NO:15显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0691] SEQ ID NO:16显示了SEQ ID NO:1的环状排列形式的氨基酸序列。
- [0692] SEQ ID NO:17显示了用于产生PA#1构成部分的上方/编码链寡脱氧核苷酸的核酸序列。
- [0693] SEQ ID NO:18显示了用于产生PA#1构成部分的下方/非编码链寡脱氧核苷酸的核酸序列。
- [0694] SEQ ID NO:19显示了pASK88-Fab-2xSapI上编码的抗体Fab片段的免疫球蛋白轻链的C末端附近的核酸序列段(上方/编码链)。
- [0695] SEQ ID NO:20显示了pASK88-Fab-2xSapI上编码的抗体Fab片段的免疫球蛋白轻链的C末端附近的核酸序列段(下方/非编码链)。
- [0696] SEQ ID NO:21显示了pASK88-Fab-2xSapI上编码的Fab片段的轻链的C末端的氨基酸序列。
- [0697] SEQ ID NO:22显示了pASK88-Fab-2xSapI的核酸序列。
- [0698] SEQ ID NO:23显示了插入一个PA#1 (20) 聚合物之后,编码Fab轻链的C末端的氨基酸序列的核酸序列段(上方/编码链)。
- [0699] SEQ ID NO:24显示了插入一个PA#1 (20) 聚合物之后,Fab轻链的C末端的氨基酸段的核酸序列(下方/非编码链)。
- [0700] SEQ ID NO:25显示了插入一个PA#1 (20) 聚合物之后,Fab轻链的C末端的氨基酸序列段。
- [0701] SEQ ID NO:26显示了pFab-PA#1 (200) 上编码的Fab重链的氨基酸序列。
- [0702] SEQ ID NO:27显示了pFab-PA#1 (200) 上编码的与PA#1 (200) 聚合物融合的Fab轻链的氨基酸序列。
- [0703] SEQ ID NO:28显示了pFab-PA#1 (200) 的核酸序列。
- [0704] SEQ ID NO:29显示了编码INFa2b的N末端和Strep标签II (仅最后两个氨基酸)的氨基酸序列的核酸序列(上方/编码链)。
- [0705] SEQ ID NO:30显示了编码INFa2b的N末端和Strep标签II (仅最后两个氨基酸)的氨基酸序列的核酸序列(下方/非编码链)。
- [0706] SEQ ID NO:31显示了Strep标签II的C末端和INFa2b的N末端的氨基酸序列。
- [0707] SEQ ID NO:32显示了pASK-IFNa2b的核酸序列。
- [0708] SEQ ID NO:33显示了插入一个PA#1聚合物序列盒之后,编码Strep标签II的C末端

和IFNa2b的N末端的核酸序列段(上方/编码链)。

[0709] SEQ ID NO:34显示了插入一个PA#1聚合物序列盒之后,Strep标签II的C末端和IFNa2b的N末端的核酸序列段(下方/非编码链)。

[0710] SEQ ID NO:35显示了与一个PA#1聚合物盒融合后,Strep标签II的C末端和IFNa2b的N末端的氨基酸序列段。

[0711] SEQ ID NO:36显示了pPA#1(200)-IFNa2b上编码的与PA#1(200)聚合物融合的IFNa2b和Strep标签II的氨基酸序列。

[0712] SEQ ID NO:37显示了pPA#1(200)-IFNa2b的核酸序列。

[0713] SEQ ID NO:38显示了pASK75-His₆-hGH上编码His₆-hGH的N末端附近的氨基酸序列的核酸序列段(上方/编码链)。

[0714] SEQ ID NO:39显示了pASK75-His₆-hGH上编码hGH的N末端附近的氨基酸序列的核酸序列段(下方/非编码链)。

[0715] SEQ ID NO:40显示了pASK75-His₆-hGH上编码的His₆-hGH的N末端的氨基酸序列段。

[0716] SEQ ID NO:41显示了pASK75-His₆-hGH的核酸序列。

[0717] SEQ ID NO:42显示了插入PA#1(20)聚合物之后,编码His₆-hGH的N末端的氨基酸序列的核酸序列段(上方/编码链)。

[0718] SEQ ID NO:43显示了插入一个PA#1聚合物序列盒之后,编码hGH的N末端的核酸序列(下方/非编码链)。

[0719] SEQ ID NO:44显示了插入PA#1(20)聚合物之后,His₆-hGH的N末端的氨基酸序列。

[0720] SEQ ID NO:45显示了pASK75-His₆-PA#1(200)-hGH上编码的成熟His₆-PA#1(200)-hGH的氨基酸序列。

[0721] SEQ ID NO:46显示了pASK75-His₆-PA#1(200)-hGH的核酸序列。

[0722] SEQ ID NO:47显示了pCHO-PA#1(200)-hGH上编码的His₆-PA#1(200)-hGH的氨基酸序列。

[0723] SEQ ID NO:48显示了pCHO-PA#1(200)-hGH的核酸序列。

[0724] SEQ ID NO:49显示了pCHO-hGH的核酸序列。

[0725] SEQ ID NO:50显示了pCHO的核酸序列。

[0726] SEQ ID NO:51显示了P1A1的氨基酸序列。

[0727] SEQ ID NO:52显示了用于产生P1A1的构成部分的上方/编码链寡脱氧核苷酸的核酸序列。

[0728] SEQ ID NO:53显示了用于产生P1A1的构成部分的下方/非编码链寡脱氧核苷酸的核酸序列。

[0729] SEQ ID NO:54显示了用于产生P1A3的构成部分的上方/编码链寡脱氧核苷酸的核酸序列。

[0730] SEQ ID NO:55显示了用于产生P1A3的构成部分的下方/非编码链寡脱氧核苷酸的核酸序列。

[0731] SEQ ID NO:56显示了pFab-P1A1(200)上编码的与P1A1(200)聚合物融合的Fab轻链的氨基酸序列。

[0732] SEQ ID NO:57显示了pFab-P1A3 (200) 上编码的与P1A3 (200) 聚合物融合的Fab轻链的氨基酸序列。

[0733] SEQ ID NO:58显示了pFab-P1A1 (200) 的核酸序列。

[0734] SEQ ID NO:59显示了pFab-P1A3 (200) 的核酸序列。

[0735] SEQ ID NO:60显示了pSUMO-PA#1 (200) 的核酸序列。

[0736] SEQ ID NO:61显示了用于制备药物缀合物的PA#1 (200) 多肽/聚合物(其是通过将10个20聚体的编码基因盒连接起来获得的,并包括来自下游连接位点的一个额外的C末端Ala残基)。

序列表

<110> XL-蛋白有限责任公司
慕尼黑科技大学
阿恩·斯科拉
尤里·宾德尔
马丁·斯拉珀斯其

<120> 生物合成的脯氨酸/丙氨酸无规卷曲多肽及其用途

<130> 12C80943CN

<160> 61

<170> PatentIn version 3.3

<210> 1

<211> 20

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: PA#1的氨基酸序列"

<400> 1

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
1 5 10 15

Ala Ala Pro Ala
20

[0001]

<210> 2

<211> 20

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: PA#2的氨基酸序列"

<400> 2

Ala Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala
1 5 10 15

Pro Ala Ala Pro
20

<210> 3

<211> 20

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: PA#3的氨基酸序列"

<400> 3

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro
1 5 10 15

[0002]

Ala Ala Ala Pro
20

<210> 4
<211> 24
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: PA#4的氨基酸序列"

<400> 4

Ala Ala Pro Ala Ala Pro Ala Ala Pro Ala Ala Pro Ala Ala Pro Ala
1 5 10 15

Ala Pro Ala Ala Pro Ala Ala Pro
20

<210> 5
<211> 24
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: PA#5的氨基酸序列"

<400> 5

Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala
1 5 10 15

Ala Pro Ala Pro Ala Ala Ala Pro
20

<210> 6
<211> 24
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: PA#6的氨基酸序列"

<400> 6

Ala Ala Ala Pro Ala Ala Pro Ala Ala Pro Pro Ala Ala Ala Ala Pro
1 5 10 15

Ala Ala Pro Ala Ala Pro Pro Ala
20

<210> 7
<211> 20
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 7

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala
1 5 10 15

Ala Pro Ala Ala
20

<210> 8

<211> 20

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 8

Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala
1 5 10 15

Pro Ala Ala Ala
20

<210> 9

<211> 20

<212> PRT

<213> 人工序列

[0003]

<220>

<221> 来源

<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 9

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro
1 5 10 15

Ala Ala Ala Pro
20

<210> 10

<211> 20

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 10

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
1 5 10 15

Ala Ala Pro Ala
20

<210> 11

<211> 20
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 11

Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala
1 5 10 15

Ala Pro Ala Ala
20

<210> 12
<211> 20
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 12

Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala
1 5 10 15

[0004]

Pro Ala Ala Pro
20

<210> 13
<211> 20
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 13

Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro
1 5 10 15

Ala Ala Pro Ala
20

<210> 14
<211> 20
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 14

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
1 5 10 15

[0005]

Ala Pro Ala Pro
20<210> 15
<211> 20
<212> PRT
<213> 人工序列<220>
<221> 来源
<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 15

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala
1 5 10 15Pro Ala Pro Ala
20<210> 16
<211> 20
<212> PRT
<213> 人工序列<220>
<221> 来源
<223> 注释="人工序列的描述: SEQ ID NO: 1的轮转排列形式的氨基酸序列"

<400> 16

Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro
1 5 10 15Ala Pro Ala Ala
20<210> 17
<211> 60
<212> DNA
<213> 人工序列<220>
<221> 来源
<223> 注释="人工序列的描述: 用于产生PA#1构成部分的上方/编码链寡脱氧核苷酸的核酸序列"<400> 17
gccgctccag ctgcacctgc tccagcagca cctgtgtcac cagctccggc tgcctctgct 60<210> 18
<211> 60
<212> DNA
<213> 人工序列<220>
<221> 来源
<223> 注释="人工序列的描述: 用于产生PA#1构成部分的下方/非编码链寡脱氧核苷酸的核酸序列"<400> 18
ggcagcagga gcagccggag ctggtgcagc aggtgtgtgt ggagcaggtg cagctggagc 60

[0006]

<210> 19
 <211> 49
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK88-Fab-2xSapI上编码的抗体Fab片段的免疫球蛋白轻链的C末端附近的核酸序列段(上方/编码链)"

<400> 19
 aagagettca accgeggaga gtgetettct gccfgaagag ctttaagett 49

<210> 20
 <211> 49
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK88-Fab-2xSapI上编码的抗体Fab片段的免疫球蛋白轻链的C末端附近的核酸序列段(下方/非编码链)"

<400> 20
 aagettaage tcttcaggca gaagageact etccgcggtt gaagetctt 49

<210> 21
 <211> 11
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK88-Fab-2xSapI上编码的Fab片段的轻链的C末端的氨基酸序列"

<400> 21
 Lys Ser Phe Asn Arg Gly Glu Cys Ser Ser Ala
 1 5 10

<210> 22
 <211> 4610
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK88-Fab-2xSapI的核酸序列"

<400> 22
 acccgacacc atcgaaiggc cagatgatta attcctaatt ttigtigaca ctctatcatt 60
 gatagagtta tttaccact cctatcagt gatagagaaa agtgaaatga atagttcgac 120
 aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatecgatt gcagtggcac 180
 tggetgggtt cgtaccgta ggcagggcgg aagttaaact gcaggaatcc ggtggtggtc 240
 tggttcagcc aggtggttcc ctgggctct cgtgtgtgc ttccggttcc aacatcaaag 300
 acaactacat ccactgggtt cgtcaggtc cgggtaaagg ccgggaatgg gttgetcgt 360
 tetacccgac caacggttac accaggtatg ccgattcagt taaggtcgt ttaccatct 420
 cggcggacac ttccaaaaac accgettacc tccagatgaa ctccctcgt gctgaagaca 480

	cagctgttta ttattgtctc cgttgggggtg gtgacgggtt ctacgtatg gactactggg	540
	gtcagggtac cctggtaacc gtctctcag cctccacca gggcccatcg gtcttcccc	600
	tggcaccctc ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggtcaagg	660
	actacttccc cgaaccgggtg acgggtgtgt ggaactcagg cgccttgacc agcggcgtgc	720
	acaccttccc ggtgttctta cagtcctcag gactctact cctcagcagc gtggtgaactg	780
	tgcctctcag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca	840
	acaccaaggi cgacaagaaa gttagagcca aatcttgcca tcaccacct caccatlaat	900
	aaccatggag aaaataaagt gaaacaaagc actattgcac tggcactctt accgttactg	960
	tttacccttg tgacaaaagc cgacatcgag ctacccaat cccgtctct cctgtccgt	1020
	tccgttggcg accgtgttac catcaegtgt agggcctcgc aagacgtaaa caccgcctga	1080
	gcgtggatc agcagaaacc cgggaaagct cggaaactgc tgatctatag cgttctctc	1140
	ctgtattcgg gattccgag caggttcagt ggttccgtt ccgttaccga cttcaccctg	1200
	acgatatact ccttcagcc ggaagacttc gtacctact actgtcaaca gcactacacc	1260
	acccgcctga ccttcggta gggtaacaaa ctgagatca aacggactgt ggttgcacca	1320
	tctgtcttca tcttccgcc atctgatgag cagttgaaat ctggaactgc ctctgttgtg	1380
	tgcctgtga ataacttcta tcccagagag gccaaagtac agtgggaagg ggataacgcc	1440
[0007]	ctccaatcgg gtaactccca ggagagtgtc acagagcagg acagcaagga cagcacctac	1500
	agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacaana agtctacgcc	1560
	tgcgaagta cccatcaggg cctgagttcg cccgtcaca agagcttcaa ccgcggagag	1620
	tgtcttctg cctgaagagc itaagcttga cctgigaagt gaaaaatggc gcacattgtg	1680
	cgacattttt tttgtctgcc gtttaccgt actgcgtcac ggatctcac ggcctctga	1740
	gcggcgcat aagcgcgcg ggtgtggtgg ttacgcgcag cgtgaccgt acacttgcca	1800
	gcgcctagc gccgcctct ttgcctttct tcccttctt tctcgcacg ttgcctggct	1860
	ttcccgta agctctaaat cgggggctcc ctttagggtt ccgatttagt gctttacgce	1920
	acctgaccc caaaaaactt gattaggggt atggttcacg tagtgggcca tgcctctgat	1980
	agacggtttt tgcctcttg acgttggagt ccaegtctt taatagtga ctctgttcc	2040
	aaactggaac aacctcaac cctatctcgg tetattcttt tgatttataa gggattttgc	2100
	cgatttcggc ctatttggtt aaaaatgagc tgatttaaca aaaatttaac ggaatttta	2160
	acaaaatatt aacgtttaca atttcaggtg gcacttttcg gggaaatgtg cgcggaacc	2220
	ctatttgttt atttttctaa atacattcaa atagtatcc gctcatgaga caataacct	2280
	gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg	2340
	cccttatcc cttttttgg gcattttgcc ttctgtttt tctcaccga gaaacgttg	2400
	tgaagtaaa agatgtgaa gatcagttgg gtgcacaggt gggttacac gaactggatc	2460
	tcaacagcgg taagatctt gagagtttc gccccaaga acgttttcca atgatgagca	2520

	citttaaagt tctgetaigt ggcgcggtat tatcccgat tgacgcggg caagagcaac	2580
	tcggtcgcgc catacaatat tctcagaatg acttgggtga gtactcacca gtcacagaaa	2640
	agcatcttac ggaigggcatg acagtaagag aattatgcag tgetgccata acctgagtg	2700
	ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt	2760
	ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg	2820
	aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc	2880
	gcnaactatt aactggcgaa ctacttaact tagcttcccg gcaacaattg atagactgga	2940
	tggaggcgga taaagttgca ggaccacttc tgcctcggc ccttcggct ggctggtta	3000
	ttgtgataa atctggagcc ggtgagcgtg gctctcggg tatcattgca gcactggggc	3060
	cagatggtaa gccctcccg atcgtagtta tctacacgac ggggagtcag gcaactatgg	3120
	atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat	3180
	taatgatgtc tcgttttagat aaaagtaaa tgattaacag cgcattagag ctgcttaatg	3240
	aggctggaat cgaaggttta acaaccgta aactcgcga gaagctaggt gtagagcagc	3300
	ctacattgta ttgcctgta aaaaataagc ggcctttgct cgacgcctta gccattgaga	3360
	tgttagatag gcaccatact cactttlccc ctllagaagg ggaaagctgg caagattttt	3420
	tacgtaataa cgctaaaagt tttagatgig ctttactaag tcctcgcgat ggagcaaaag	3480
	tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct	3540
[0008]	ttttatgcca acaaggtttt tctactagaga atgcattata tgcactcagc gcagtggggc	3600
	attttacttt aggttgcgta ttggaagatc aagagcaca agtcgctaaa gaagaaaggg	3660
	aaacacctac tactgatagt atgcggccat tattacgaca agctatcgaa ttatttgatc	3720
	accaagggtc agagccagcc ttcttatctg gccttgaatt gatcatatgc ggattagaaa	3780
	aacaacttaa atgtgaaagi gggcttlaaa agcagcataa ccttttccg tcatggtaac	3840
	ttactagtt taaaaggatc taggtgaaga tctttttga taatctcatg accaaaaacc	3900
	cttaacgtga gttttcgtc cactgagcgt cagaccctgt agaaaagatc aaaggatctt	3960
	cttgagatcc tttttttctg cgcgtaatct gctgcttgc aacaaaaaaa ccaccgtac	4020
	cagcgggtgt ttgtttccg gatcaagagc taaccaactct ttttcgaag gtaactggct	4080
	tcagcagagc gcagatacca aatctgtcc ttctagtgt gccgtagtta ggcaccact	4140
	tcaagaacte ttagcaacc cctacatacc tcgtctctgt aatctgtta ccagtggctg	4200
	ctgccagtgg cgataagtcg tgtcttaccg ggttgactc aagacgatag ttaccggata	4260
	agggcagcgc gtcgggctga acggggggtt cgtgcacaca gccagcttg gacgaacga	4320
	cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg ctccccgaag	4380
	ggagaaagcc ggacaggtat ccggtgaagc gcagggtcgg aacaggagag cgcacgaggg	4440
	agcttcacag gggaaacgcc tggatcttt atagtcctgt ccggtttccg cactctgac	4500
	ttgagcgtcg atttttga tgcctcagc gggggcggag cctatggaaa aacgccagca	4560
	acgggccttt tttaaggctc ctggccttt gctggccttt tgctcaatg	4610

[0009]

<210> 23
 <211> 85
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 插入一个PA#1(20)聚合物之后, 编码Fab轻链的C末端的氨基酸序列的核酸序列段(上方/编码链)"

<400> 23
 gggtgtcttt ctgccgcicc agctgcacct gctccagcag caccigtctc accagctcgc 60
 gctgctcctg ctgctgaag agctt 85

<210> 24
 <211> 85
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 插入一个PA#1(20)聚合物之后, Fab轻链的C末端的氨基酸段的核酸序列(下方/非编码链)"

<400> 24
 aagctcttca ggcagcagga gcagccggag ctggtgcage aggtgctgct ggagcaggtg 60
 cagctggage ggcagaagag cactc 85

<210> 25
 <211> 25
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 插入一个PA#1(20)聚合物之后, Fab轻链的C末端的氨基酸序列段"

<400> 25
 Glu Cys Ser Ser Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
 1 5 10 15
 Ala Pro Ala Pro Ala Ala Pro Ala Ala
 20 25

<210> 26
 <211> 229
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pFab-PA#1(200)上编码的Fab重链的氨基酸序列"

<400> 26
 Glu Val Lys Leu Gln Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
 1 5 10 15
 Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Asn Ile Lys Asp Thr

[0010]

[illegible]

<210>	27
<211>	417
<212>	PRT
<213>	人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: pFab PA#1(200)上编码的与PA#1(200)聚合物融合的Fab轻链的氨基酸序列"
<400> 27

Asp Ile Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala
 20 25 30
 Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
 35 40 45
 Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg Phe Ser Gly
 50 55 60
 Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
 65 70 75 80
 Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr Thr Pro Pro
 85 90 95
 Thr Phe Gly Gln Gly Thr Lys Leu Glu Ile Lys Arg Thr Val Ala Ala
 100 105 110
 Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
 115 120 125
 Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
 130 135 140
 [0011] Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
 145 150 155 160
 Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
 165 170 175
 Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
 180 185 190
 Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
 195 200 205
 Phe Asn Arg Gly Glu Cys Ser Ser Ala Ala Pro Ala Ala Pro Ala Pro
 210 215 220
 Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
 225 230 235 240
 Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
 245 250 255
 Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
 260 265 270
 Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
 275 280 285

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro
290 295 300

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
305 310 315 320

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
325 330 335

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
340 345 350

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
355 360 365

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro
370 375 380

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
385 390 395 400

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
405 410 415

Ala

[0012]

<210> 28

<211> 5210

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: pFab-PA#1(200)的核酸序列"

<400> 28

```

accgcacacc atcgaatgga cagatgatia attcctaatt ttgttgaca ctctatcatt      60
gatagagtta tttaccact cctatcagt gatagagaaa agtgaatga atagttcgac      120
aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac      180
tggttggttt cgtaccgta ggcagggcgg aagttaaaact gcaggaatec ggtgggtgct      240
tggttcagcc aggtgggttc ctgcggctct cgtgltcgtc ttcgggttcc aacatcaaag      300
acacetacat ccactgggtt cgtcaggctc cgggttaaagg cctggaatgg gtgtctcgta      360
tctaccggac caacggttac accaggtatg ccgattcagt taanggtcgt ttcaccatct      420
cggcggacac ttccaasaac accgcttacc tccagatgaa ctccctgctg gctgaagaca      480
cagctgttta ttattgtccc cgttgggggtg gtgacgggtt ctacgctatg gactactggg      540
gtcagggttac cctggtcacc gtctctcag cctccacca gggcccatcg gtcttcccc      600
tggcacccct ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggtcaagg      660
actacttccc cgaaccggtg acggtgtcgt ggaactcagg cgcctgacc agcggcgtgc      720

```

[0013]	acaccttccc ggctgtecta cagtcctcag gactctactc cctcagcagc gtggtgactg	780
	tgccttcacag cagcttgggc acccagacct acatctgcaa cgtaateac aaacccagea	840
	acaccaaggt cgacaagaaa gttgagccca aatcttgcca tcaccacat caccatlaat	900
	aaccatggag aaaataaagt gaaacaaagc actatigcac tggcactctt accgttactg	960
	tttacecctg tgacaaaage cgacatcgag ctacaccaat ccccgctctc cctgtccctg	1020
	tccgttggcg accgtgttac catcacgtgt agggcctcgc aagacgtaaa caccgcgcta	1080
	gggtggtatc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgttccctc	1140
	ctgtattccg gagttccgag caggttcagt ggttcccggt ccggtaccga cttcaccctg	1200
	aegatatact ccttcacgcc ggaagacttc gctacctact actgtcaaca gcactacacc	1260
	accccgccga ccttcgggtc gggtagcaaa ctcgagatca aacggactgt ggtgcacca	1320
	tctgtcttca tcttcccgcc atctgatgag cagttgaaat ctggaactgc ctctgttgtg	1380
	tgcctgtcga ataacttcta tcccagagag gccaaagtac agtgggaaggt ggataacgcc	1440
	ctccaatcgg gtaactccca ggagagtgtc acagagcagg acagcaagga cagcacctac	1500
	agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacacaa agtctacgcc	1560
	tgcgaagtea cccatcaggg cctgagtctg cccgtcaca agagcttcaa ccgcggagag	1620
	tgtctttctg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	1680
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	1740
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	1800
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	1860
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	1920
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	1980
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	2040
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	2100
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	2160
	gtctctgtg ccgtccacgc tgcacctgct ccagcagcac ctgtgcacc agctccggt	2220
	gtctctgtg cctgaagage ttaagcttga cctgigaagt gaaaaatggc gcacattgtg	2280
	cgacattttt ttgtctgcc gtttaccgt actgcgtcac ggatctccac gcgcctgta	2340
	gcggcgcaat aagcgcgcg ggtgtgtgg ttacgcgcag cgtgaccgt acacttgcca	2400
	gcgcctagc gccgcctct ttcgtttct tcccttctt tctcgccacg ttgcgggt	2460
	ttcccgtea agctctaaat cgggggctcc ctttagggt cggatttagt gttttacgc	2520
	acctgaccc caaaaaactt gattaggggt atggttcacg tagtgggcca tgcacctgat	2580
	agacggtttt tgcaccttg acgttggagt ccacgttctt taatagtga ctctgttcc	2640
	aaactggaac aacactcaac cctatctcgg tctattctt tgalitataa gggattttgc	2700
	cgatttcggc ctatttggtt aaaaatgagc tgatttaaca aaaatttaac ggaatttta	2760

	acaaaatatt aacgtttaca atticaggtg gcacttttgc gggaaatgtg cgeggaaccc	2820
	ctatttggtt atttttctaa atacattcaa atatgtatcc gctcaigaga caataaccct	2880
	gataaatgct tcaataatat tgaaaaagga agagtatgag tatcaacat ttcctgtctg	2940
	cccttatccc cttttttgcg gcattttgcc ttectgtttt tgctcaccca gaaacgctgg	3000
	tgaagtaaa agatgctgaa gatcagttgg gtcacagagt gggttacatc gaactggatc	3060
	tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca	3120
	cttttaagt tctgctatgt ggcgggtat taccctgat tgacgcggg caagagcaac	3180
	tggctgcgcg catacactat tctcagaatg acttggttga gtactacca gtcacagaaa	3240
	agcatcttac ggatggcatg acagtaagag aattatgcag tcttgccata accatgagt	3300
	ataacacigc ggccaacila ctctgacaa cgaicggagg accgaaggag ctaccgcctt	3360
	ttttgcacaa caiggggat catgtaactc gcttgatcg ttgggaaccg gagctgaatg	3420
	aagccatacc aaacgacgag cgtgacacca cgaatgcctg agcaatggca acaacgttgc	3480
	gcaaaactatt aactggcgaa ctacttactc tagcttccc gcaacaattg atagactgga	3540
	tggagcgga taaagttgca ggaccattc tgcctcgcc ccttcggct ggctggttta	3600
	ttctgataa atctggagcc ggtgagcgtg gctctcgcc tatcattgca gcactgggce	3660
	cagatggtaa gccctccct atcgtagtta tctacacgac ggggagtcag gcaactatgg	3720
	atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat	3780
[0014]	taatgatgct tegttagat aaaagtaag tgattaacag cgcattagag ctgcttaatg	3840
	aggctggaat cgaaggttta acaacccgta aactcgccca gaagctaggt gtagagcagc	3900
	ctacattgta ttggeatgta aaaaaaagc gggtttgct cgacgcctta gccattgaga	3960
	tgtagatag gcaccatact cacttttgc ctttagaagg ggaaagctgg caagattttt	4020
	tacgtaataa cgtataaagt tttagatgtg ctttactaag tcctcgcat ggagcaaaag	4080
	tacatttagg tacacgcct acagaaaaac agtatgaac tctcgaaat caattagcct	4140
	ttttatgcca acaaggtttt tcaatagaga atgcattata tgcactcagc gcagtgggce	4200
	attttacitt aggttgcgt ttggaagatc aagagcatca agtcctaaa gaagaaagg	4260
	aaacacctac tactgatagt atgcgccat tattacgaca agctatcgaa ttattigate	4320
	accaaggtgc agagccagcc ttcttattcg gccctgaatt gatcatatgc ggattagaaa	4380
	aacaacttaa atgtgaaagt ggtctttaa agcagcataa ccttttccg tgatggtaac	4440
	ttactagtt taaaaggatc taggtgaaga tctttttga taatctcatg accaaaatcc	4500
	cttaacgtga gttttcttc cactgagcgt cagaccocgt agaaaagatc aaaggatctt	4560
	cttgagatcc tttttttctg cgcgtaatct gctgcttgc aacaaaaaaa ccaccgtac	4620
	cagcgtgggt ttgtttgcg gatcaagagc taccaactct ttttcgaag gtaactggct	4680
	tcagcagagc gcagatacca aatactgtcc ttctagtgta gccgtagtta ggccaccact	4740
	tcaagaacte tgtagcaccg ctacataacc tgcctctgct aatcctgtta ccagtgcctg	4800
	ctgccagtgg cgataagtcg tgtcttaccg ggttggaact aagacgatag ttaccggata	4860

[0015]

```

aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca gccagcttg gagcgaacga 4920
cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg ctcccggaag 4980
ggagaaaggc ggacaggatc ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg 5040
agcttccagg gggaaacgac tggatcttt atagtcctgt cgggtttcgc caectctgac 5100
ttgagcgtcg atttttctga tgcctgctag gggggcggag cctatggaan aacgccagca 5160
acgcggcctt ttacgggtc ctggccttt gctggccttt tgcacatg 5210

```

<210> 29
 <211> 57
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 编码INFa2b的N末端和Strep标签II (仅最后两个氨基酸)的氨基酸序列的核酸序列(上方/编码链)"

```

<400> 29
gaaaaaggcg ccagctcttc tgcctgtgat ctgcctcaaa ccacagcct ggtagc 57

```

<210> 30
 <211> 57
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 编码INFa2b的N末端和Strep标签II (仅最后两个氨基酸)的氨基酸序列的核酸序列(下方/非编码链)"

```

<400> 30
gtatcccagg ctatgggttt gaggcagatc acaggcagaa gagctggegc cttttc 57

```

<210> 31
 <211> 19
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: Strep标签II的C末端和INFa2b的N末端的氨基酸序列"

```

<400> 31
Glu Lys Gly Ala Ser Ser Ser Ala Cys Asp Leu Pro Gln Thr His Ser
1           5           10           15

```

Leu Gly Ser

<210> 32
 <211> 3721
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK-IFNa2b的核酸序列"

[0016]

```

<400> 32
ccatcgaatg gccagatgat taaticctaa tttttgtga cactctatca ttgatagagt 60
tattttacca etccctatca gtgatagaga aaagtgaat gaatagtctg acaaaaatct 120
agataacgag ggcaaaaaat gaaaaagaca gctatcgga ttgcagtggc actggctggt 180
ttcgctaccg tagcgagcg cgttagctgg agccaccgc agttcgaaa aggcgccagc 240
tettctgctt gtagcttgc tcaaacccac agcttgggtg gcaggaggac cttagctctc 300
ctggcacaga tgaggagaat ctctctttt tcttgcctga aggacagaca tgactttgga 360
tttcccagg aggagtttgg caaccagttc caaaaggctg aaaccatccc tgctctccat 420
gagatgatec agcagatctt caatctctt agcacaagg actcatctgc tgettgggat 480
gagaccctec tagacaaatt ctacactgaa ctctaccagc agctgaatga cctggaagcc 540
tgttgtatac aggggggtgg ggtagacagag actccctga tgaaggagga ctccattctg 600
gctgtgagga aatacttcca aagaatcact ctctatctga aagagaagaa atacagccct 660
tgtgctggg aggttgtcag agcagaaatc atgagatctt ttctttgtc aacaaacttg 720
caagaaagtt taagaaagta ggaataagct tgacctgta agtgaaaaat ggccacatt 780
gigcgacatt tttttgtct gcgctttacc gctactgct caeggatctc caegccct 840
gtagcggcgc attaagcgc ggggtgttgg tggttacgc cagcgtgacc gctacacttg 900
ccagcgcct agcgcgcct ctcttgcct tctccctc ctctctgc acgttcgcg 960
gcttccctc tcaagctct aatcgggggt tcccttagg gtccgattt agtgccttac 1020
ggcacctcga cccccaaaaa cttagtagg gtgatgttc agtagtggg ccategcct 1080
gatagacggt ttttgcctt ttgacgttgg agtcacggt cttaaatagt ggactctgt 1140
tccaaactgg aacnacact aacctatct cgtctctt ttttgattt taagggattt 1200
tgccgattt ggctatttgg ttaaaaaatg agctgattt acaaaaattt aacgcgaatt 1260
ttaacaaat attaacgctt acaattcag gtggcactt tggggaaat gtgcgcgaa 1320
ccctatttt tttattttt taatacatt caaatatgt tccgtctatg agacaataac 1380
cctgataaat gcttcaataa tatgaaaaa ggaagagtat gagtattca catttccgtg 1440
tcgcccttat tccctttttt gggcatttt gccttctgt tttgtctac ccagaaacgc 1500
tggtagaagt aaagatgct gaagatcagt tgggtgcac agtgggttac atcgaaactg 1560
aictcaacag cggtaagatc ettgagagt ttcgccccga agaaegtiti ccaatgatga 1620
gcacttttaa agttctgcta tgtggcgcg tattatccg tattgacgc gggcaagagc 1680
aactcgctc ccgcataac tatctcaga atgacttggg tgagtactca ccagtcacag 1740
aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgcctgc ataaccatga 1800
gtgataaac tcggccaac ttacttctga caacgatcgg aggaccgaag gagctaaccg 1860
cttttttga caacatggg gatcatgtaa ctgccttga tegtiggaa ccggagctga 1920
atgaagccat accaaacgac gacgtgaca ccaagatgc ttagcaatg gcaacaact 1980
tgccaaaact attaaactgc gaactacta ctctagctt ccggcaaca ttgatagact 2040

```

[0017]

```

ggatggagge ggataaagtt gcaggaccac ttctgogctc ggccttcag gctggetggt 2100
ttatigtctga taaatctgga gccggtagcg gtggctctcg cggtaacatt gcagcactgg 2160
ggccagaagg laagccctcc cglalegiag llateiacac gacggggagc caggcaacta 2220
tggatgaacg aatatagacag atcgtctgaga taggtgccic actgattaag cattggtagg 2280
aattaatgat gicctgttta gataaaagta aagtattaa cagegcatta gagctgctta 2340
atgaggtegg aatcgaaggt ttaacaacc gtaactcgc ccagaagcta ggtgtagagc 2400
agcciacall gtaattggcat gtaaaaaata agcgggctll gctcgacgc itagccallg 2460
agatgttaga taggcaccat actcactttt gccctttaga aggggaaagc tggcaagatt 2520
ttttacgtaa taacgtaaa agtttttagat gtgctttact aagtcacgc gatggagcaa 2580
aagtacattt aggtacacgg cctacagaaa aacagtatga aactctcgaa aatcaattag 2640
ccttttlaig ccaacnaggi lllcactag agaatgcalt alalgcactc agcgcagigg 2700
ggcattttac tttaggttgc gtattggaag atcaagagca tcaagtcgt aaagaagaaa 2760
gggaacacc tactactgat agtatgcgc cattattaag acaagctate gaattatttg 2820
atcaceaaag tgcagagcca gccttcttat teggccttga attgatcata tgcggattag 2880
aaaaacaact taaatgtgaa agtgggtctt aaaagcagca taacctttt ccgtgatggt 2940
aacttcacta gtttaaaagg atctaggta agatcctttt tgataatctc atgaccaaaa 3000
tcccttaacg tgagtttctg ttccactgag cgtcagacc cgtagaaaag atcaaaggat 3060
cttcttgaga tcttttttt ctgcgcgtaa tctgtctgtt gcaaaacaaa aaaccacgc 3120
taccagcggg ggtttgtttg cggatcaag agctaccaac tcttttccg aaggtaactg 3180
gcttcagcag agcgcagata ccaataacig tccctctagt gtacccgtag ttaggcacc 3240
acttcaagaa ctctgtagca ccgcctacat acctcgtct gctaatectg ttaccagigg 3300
ctgcgcagcagg tggcgataag tctgtcttla ccgggttga ctaagacga tagttacgg 3360
ataaggcgca ggggtcgggc tgaacggggg gtctgtgcac acagcccagc ttggagcgaa 3420
cgacctacac cgaactgaga tactacagc gtgagctatg agaaagcgc acgttcccg 3480
aaggggagaa ggcggacagg tatccggtaa gcggcagggt cggaaacagga gagcgacga 3540
gggagcttcc agggggaac gcctggatc tttatagtc tctcgggttt cgcacctct 3600
gacttgagcg tcgatttttg tgatctctgt caggggggcg gacctatgg aaaaacgca 3660
gcaacgcggc ctttllaagg ttcttgccct ttgttgcc tttgtctac atgacccgac 3720
a 3721

```

<210> 33
 <211> 90
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 插入一个PA#1聚合物序列盒之后, 编码Strep标签II的C末端和IPNa2b的N末端的核酸序列段(上方/编码链)"

[0018]

```

<400> 33
gaaaaaggcg ccagctcttc tgcgcgtcca gctgcacctg ctccagcagc acctgctgca      60
ccagctccgg ctgctcctgc tgcctgtgat                                         90

<210> 34
<211> 90
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: 插入一个PA#1聚合物序列盒之后, Strep标签II的C末端和IFNa2b的N末端的核
酸序列段(下方/非编码链)"

<400> 34
atcacaggca gcaggagcag ccggagctgg tgcagcaggt gcigctggag caggtgcage      60
tggagcggca gaagagctgg cgcctttttc                                         90

<210> 35
<211> 30
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: 与一个PA#1聚合物盒融合后, Strep标签II的C末端和IFNa2b的N末端的氨基酸
序列段"

<400> 35
Glu Lys Gly Ala Ser Ser Ser Ala Ala Pro Ala Ala Pro Ala Pro Ala
1          5          10          15

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Cys Asp
          20          25          30

<210> 36
<211> 381
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: pPA#1(200)-IFNa2b上编码的与PA#1(200)聚合物融合的IFNa2b和Strep标签II
的氨基酸序列"

<400> 36
Ala Ser Trp Ser His Pro Gln Phe Glu Lys Gly Ala Ser Ser Ser Ala
1          5          10          15

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala
          20          25          30

Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala
          35          40          45

Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala
          50          55          60

```

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala
65 70 75 80

Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala
85 90 95

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala
100 105 110

Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala
115 120 125

Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala
130 135 140

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala
145 150 155 160

Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala
165 170 175

Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala
180 185 190

[0019] Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala
195 200 205

Pro Ala Pro Ala Ala Pro Ala Ala Cys Asp Leu Pro Gln Thr His Ser
210 215 220

Leu Gly Ser Arg Arg Thr Leu Met Leu Leu Ala Gln Met Arg Arg Ile
225 230 235 240

Ser Leu Phe Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro Gln
245 250 255

Glu Glu Phe Gly Asn Gln Phe Gln Lys Ala Glu Thr Ile Pro Val Leu
260 265 270

His Glu Met Ile Gln Gln Ile Phe Asn Leu Phe Ser Thr Lys Asp Ser
275 280 285

Ser Ala Ala Trp Asp Glu Thr Leu Leu Asp Lys Phe Tyr Thr Glu Leu
290 295 300

Tyr Gln Gln Leu Asn Asp Leu Glu Ala Cys Val Ile Gln Gly Val Gly
305 310 315 320

Val Thr Glu Thr Pro Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg
325 330 335

Lys Tyr Phe Gln Arg Ile Thr Leu Tyr Leu Lys Glu Lys Lys Tyr Ser
340 345 350

Pro Cys Ala Trp Glu Val Val Arg Ala Glu Ile Met Arg Ser Phe Ser
355 360 365

Leu Ser Thr Asn Leu Gln Glu Ser Leu Arg Ser Lys Glu
370 375 380

<210> 37

<211> 4321

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: pPA#1(200)-IFNa2b的核酸序列"

<400> 37

[0020]

```
ccatcgaaig gccagatgat taattcctaa ttttgttga cactctatca ttgatagagt 60
tattttacca ctccctatca gtgatagaga aaagtgaat gaatagttcg acaaaaatct 120
agataacgag ggcaaaaaat gaaaaagaca getatcgga ttgcagtggc actggctggt 180
ttcgtaceg tagcgcaggc cgctagctgg agccaccgc agttcgaaaa aggcgccage 240
tettctgeeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 300
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 360
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 420
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 480
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 540
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 600
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 660
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 720
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 780
ctgctgceeg ctccagctgc acctgctcca gcagcaactg ctgcaccage tccggtgct 840
ctgctgceet gtgatctgcc tcaaacccac agcctgggta gcaggaggac cttgatgctc 900
ctggcacaga tgaggagaat ctctcttttc tctgcttga aggacagaca tgactttgga 960
tttccccagg aggagtttgg caaccagttc caaaaggctg aaaccatccc tgtctctcat 1020
gagatgatec agcagatctt caatctcttc agcacaagg actcatctgc tgettgggat 1080
gagacctec tagacaaatt ctacactgaa ctctaccage agctgaatga cctggaagcc 1140
tgtgtgatac aggggggtggg ggtgacagag actccccga tgaaggagga ctccattctg 1200
gctgtgagga aatacttcca aagaatcaat ctctatctga aagagaagaa atacagccct 1260
tgtgccfagg aggttgtcag agcagaaatc atgagatctt tttctttgtc aacaaacttg 1320
caagaaagtt taagaagtaa ggaataaget tgacctgtga agtgaaaaat ggcgcacatt 1380
```

[0021]

```

gtgcgacatt tttttgtct gccgtttacc gctactgcgt caccgatctc caccgccect 1440
gtagcggcgc attaagcgcg gcgggtgtgg tggttacgcg cagcgtgacc gctacacttg 1500
ccagcgccect agcgcgcgct cctttcgcct tcttcccttc ctttctcgcc acgttcgcgc 1560
gccttccccc tcaagctcta aatcgggggc tccctttagg gtcccgattt agtgccttac 1620
ggcacctcga ccccaaaaaa cttgattagg gtgatgggtc acgtagtggg ccatgccect 1680
gatagacggt ttttcgccect ttgacgttgg agtccacgtt ctttaatagt ggactcttgt 1740
tccaaactgg aacaacactc aacctatctc cggctctatt ttttgattta taagggattt 1800
tgccgatttc ggccatttgg ttaaaaaatg agctgattta acaaaaattt aacgcgaatt 1860
ttaacaaaat attaacgctt acnatttcag gtggcacttt tcggggaaat gtgcgcgaa 1920
ccccatttg tttatttttc taaatacatt caaatatgta tccgcctatg agacaataac 1980
cctgataaat gcttcaataa tatigaaaaa ggaagagtat gagtattcaa catttccttg 2040
tcgcccttat tccctttttt gcggcatttt gccttccgtt ttttgcctac ccagaaacgc 2100
tggtgaaagt acaagatgct gaagatcagt tgggtgcacg agtgggttac atcgaactgg 2160
atctcaacag cggtaagatc cttagagatt ttcccccga agaacgtttt ccaatgatga 2220
gcacttttaa agttctgcta tgtggcgcgg tattateccg tattgacgcc gggaagage 2280
aactcggctc ccgcatacac tattctcaga atgacttggg tgagtactca ccagtcacag 2340
aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgcctcc ataaccatga 2400
gtgataacac tgcggccaac ttacttctga caacgatcgg aggaaccgaag gagctaaccg 2460
cttttttgca caacatgggg gatcatgtaa ctgccttga tegtgggaa ccggagctga 2520
atgaagccat accaaacgac gagegtgaca ccacgatgcc tgtagcaatg gcaacaacgt 2580
tgcgcaaaat attaaactgc gaactactta ctctagcttc ccgcaacaa ttgatagact 2640
ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttcgc gctggctggg 2700
ttatigtcta taaatctgga gccggtagc gtgctctcg cggatcatt gcagcactgg 2760
ggccagatgg taagccctcc cgtatcgtag ttatctacac gaaggggagt caggcaacta 2820
tggatgaacg aaatagacag atcgtgaga taggtgcctc actgattaaag cattggtagg 2880
aattaatgat gtctcgttta gataaaagta aagtgattaa cagcgcatta gagctgetta 2940
atgaggtegg aatcgaaggt ttaacaaccc gtaaacctgc ccagaagcta ggtgtagagc 3000
agcctacatt gtatiggcat gtaaaaaata agcgggcctt gctcgacgcc ttagecattg 3060
agatgttaga taggcacctt actacatttt gccctttaga aggggaaagc tggcaagatt 3120
ttttacgtaa taacgctaaa agttttagat gtgccttact aagtcctgc gatggagcaa 3180
aagtacattt aggtacacgg cctacagaaa aacagtaiga aactctcga aatcaattag 3240
cctttttatg ccaacaaggt ttttcaactag agaatgcatt atatgcactc agcgcagtgg 3300
ggcattttac tttaggttgc gtattggaag atcaagagca tcaagtcgct aaagaagaaa 3360
gggaacaccc tactactgat agtatgcgc cattattacg acaagctatc gaattatttg 3420
atcaccaagg tgcagagcca gcctcttat tcggccttga attgatcata tgcggattag 3480

```

[0022]

```

aaaaacaact taaatgtgaa agtgggtctt aaaagcagca taaccttttt cegtgatggt 3540
aactticacia gtttaaaagg atctaggfga agatcccttt tgataatctc atgaccaaaa 3600
tcccttaacg tgagtitttg ttccactgag cgtcagaccc cgtagaaaag atcaanngat 3660
cttcttgaga tctttttttt ctgcgcgtaa tctgtctgct gcaaacaaaa aaaccacgcg 3720
taccagcggt ggttttgttg cgggatacag agctaccac cttttttcgg aaggttaactg 3780
gcttcagcag agcgcagata ccaataactg tctttctagt gtacgcgtag ttaggccacc 3840
acttaagaa ctcgttagca ccgcatacat acctcgcctt gctaactctg ttaccagtgg 3900
ctgcctccag tggcgataag tctgtcttta cgggtttgga ctaagacga tagttaccgg 3960
ataaggcgca gcggtcgggc tgaacggggg gtctgtgca acagcccagc ttggagcgaa 4020
cgacctacac cgaactgaga tactacagc gtgagctatg agaaagcgcc acgcttcccg 4080
aaggggaaaa ggcggacagg tatccggtaa gcggcagggt cggaacagga gacgcacga 4140
gggagcttcc agggggaaac gcctggctac ttatagtc tgctgggttt cgcacctct 4200
gacttgagcg tcgatttttg tgatgctcgt cagggggggc gagcctatgg aaaaacgca 4260
gcaacgcggc ctttttacgg ttcttggeet tttgtggcc ttttgctcac atgacccgac 4320
a 4321

```

<210> 38
 <211> 54
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK75-His6-hGH上编码His6-hGH的N末端附近的氨基酸序列的核酸序列段(上方/编码链)"

```

<400> 38
gccgtagcc atcaccacca tcaccatgga gccagctctt ctgccttccc aacc 54

```

<210> 39
 <211> 54
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK75- His6-hGH上编码hGH的N末端附近的氨基酸序列的核酸序列段(下方/非编码链)"

```

<400> 39
gcttgggaag gcagaagaga tggcgccaig gtgatggtgg tgatggctag cggc 54

```

<210> 40
 <211> 18
 <212> PRF
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK75- His6-hGH上编码的His6-hGH的N末端的氨基酸序列段"

[0023]

<400> 40

Ala Ala Ser His His His His His His Gly Ala Ser Ser Ser Ala Phe
1 5 10 15

Pro Thr

<210> 41

<211> 3793

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: pASK75- His6-hGH的核酸序列"

<400> 41

```

acccgacacc atcgaatggc cagatgatta attcctaatt ttgttgaca ctctatcatt    60
gatagagtta ttttaccact cctatcagt gatagagaaa agtgaaatga atagttcgac    120
aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gcagtggcac    180
tggttggttt cgtaccgta ggcagggcgc ctagecatca ccaccatcac catggcgcca    240
gctcttctgc ctteccaacc attccttat ccaggctttt tgacaacgct atgetecgcg    300
ccatcgctct gcaccagctg gcctttgaca cctaccagga gtttgaagaa gctatatacc    360
caaaggaaca gaagtattca ttctgcaga acccccagac ctccctctgt ttctcagagt    420
ctaticegac accctccac agggaggaaa caaacagaa atccaaacta gagtgcctc    480
gcctctccct gctgctcctc cagtctgggc tggagcccggt gcagttcttc aggagtgtct    540
tcgccaacag cctggtgtac ggcgcctctg acagcaacgt ctatgacctc ctaaaggacc    600
tagaggaagg catccaaacg ctgatgggga ggetggaaga tggcagcccc cggactgggc    660
agatcttcaa gcagacctac agcaagtgc acacaaactc acacaacgat gacgcactac    720
tcasagaacta cgggtctctc tactgcttca ggaaggacat ggacaaggtc gagacattcc    780
tgcgcatcgt gcagtgcgc tctgtggagg gcagctgtgg cttctaagct tgacctgtga    840
agtgaaaaat ggcgacatt gtgcgacatt tttttgtct gccgtttacc gctactgcgt    900
cacggatctc cagcgccct gtacggcgc attaagcgg ggggtgttg tggttacgg    960
cagctgacc gctacaattg ccagcgccct agcgcccgct ctttcgctt tcttccctc    1020
cttctcgcg acgttcgcg gcttccccc tcaagctcta aatcgggggc tccctttagg    1080
gttccgattt agtgccttac ggcacctga cccccaaaaa cttgattagg gtgatggttc    1140
acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg agtccacgtt    1200
ctttaatagt ggaactttgt tccaaactgg aacaacactc aaacctatct cggctatctc    1260
ttttgattta taagggtatt tgcgatttc ggcctattgg tlaaaaaa agctgattta    1320
acaaaaattt aacgcgaatt ttaacaaat attaacgttt acaatttcag gtggcacttt    1380
tcggggaaat gtgcgggaa cccctatttg tttatttttc taataacatt caaatatgta    1440
tcctctcag agacaataac ctgataaat gtttcaataa tattgaaaaa ggaagagtat    1500

```

[0024]

gagtattcaa catttccgtg tgcacctat tccctttttt ggcgcatttt gccctccgtg 1560
 ttttgcacac ccagaaacgc tggtagaagt aaaagatgct gaagatcagt tgggtgcacg 1620
 agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcccccga 1680
 agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgagg tattatcccg 1740
 tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggt 1800
 tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg 1860
 cagtgtgcc ataaccatga gtgataacac tgcggccaac ttactctga caacgatcgg 1920
 aggaccgaag gagctaacgc cttttttgca caacatgggg gatcatgtaa ctgccttga 1980
 tcgttgggaa ccggagctga atgaagccat accaaacgac gacgctgaca ccacgatgcc 2040
 tgtagcaatg gcaacaacgt tgcgcaaac attaactggc gaactactta ctctagette 2100
 ccggcaacaa ttgatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgtc 2160
 ggcccttccg gctggctggg ttattgctga taaatctgga gccggigagc gtggctctcg 2220
 cggtatcatt gcagcactgg gccagatgg taagccctcc cgtatcgtag ttatctacac 2280
 gacggggagt caggcaacta tggatgaacg aaatagacag atcgtcgaga taggtgctc 2340
 actgattaag cattggtagg aattaatgat gtctcgttta gataaaagta aagtattaa 2400
 cagegcatta gagctgctta atgaggtcgg aatcgaaggt ttaacaaccc gtaaacctgc 2460
 ccagaagcta ggtgtagagc agcctacatt gtattggcat gtaaaaaata agcgggcttt 2520
 gctcgacgcc ttagccattg agatgtaga taggcacct actcactttt gccctttaga 2580
 aggggaaagc tggcaagatt ttttacgtaa taacgctaaa agttttagat gtgtttact 2640
 aagtcacgc gatggagcaa aagtacattt aggtacacgg cctacagaaa aacagtatga 2700
 aactctcgaa aatcaattag cttttttatg ccaacaaggt tttcactag agaatgcatt 2760
 atatgcactc agcgcagtgg ggcattttac tttaggttgc gtattggaag atcaagagca 2820
 tcaagtcgt aaagaagaaa gggaaacacc tactactgat agtatgccgc cattattacg 2880
 acaagctatc gaattatttg atcaccaagg tgcagagcca gccctcttat tggccttga 2940
 attgatcata tgcggattag aaaaacaact taaatgtgaa agtgggtctt aaagcagca 3000
 taaccttttt ccgtgatggg aacttcaacta gttttaaagg atctaggtga agatcctttt 3060
 tgataatctc atgacaaaaa tcccttaacg tgagttttcg ttccactgag cgtcagaccc 3120
 cgtagaaaag atcaaaggat cttcttgaga tctttttttt ctgcgctgaa tctgtgctt 3180
 gcaaacaaaa aaaccacgcg taccagcggt ggtttgtttg ccgatcaag agctaccaac 3240
 tctttttccg aaggtaactg gcttcagcag agcgcagata ccaataactg tcttctagt 3300
 gtagecgtag ttaggcacc acttcaagaa ctctgtagca ccgcctacat acctgcctc 3360
 gctaactctg ttaccagtgg ctgctgccag tggcgataag tctgtcttta ccgggttga 3420
 ctcaagaaga tagttaccg ataagcgca gggctgggc tgaacggggg gtctgtgcac 3480
 acagcccagc ttggagcgaa cgacctacac cgaactgaga taactacagc gtgagctatg 3540

[0025]

```

agaaagcgcc acgtttcccg aaggagagaa gccggacagg tatccggtaa gccgcagggt 3600
cggaacagga gagcgacaga gggagcttcc aggggggaaac gcctgggtalc ttatagtec 3660
tgtegggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt caggggggcg 3720
gagcctatgg aaaaacgeca gcaargcggc ctttttacgg ttcttgccct ttgctggcc 3780
tttctctac atg 3793

```

<210> 42
 <211> 114
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 插入PA#1(20)聚合物之后, 编码His6-hGH的N末端的氨基酸序列的核酸序列(上方/编码链)段"

```

<400> 42
gacgtatgcc atcaccacca tcacctggc gccagctctt ctgcctctcc agctgcacct 60
gtctcagcag cactctctgc accagctcgg gtctctcttg ctgccttccc aacc 114

```

<210> 43
 <211> 114
 <212> DNA
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 插入一个PA#1聚合物序列盒之后, 编码hGH的N末端的核酸序列(下方/非编码链)"

```

<400> 43
ggttgggaag gcagcaggag cagcaggagc tggatgcaga ggtctctctg gacagggcgc 60
agctggagcg gcagaagagc tggcgccatg gtgatggtag tgatggctag cggc 114

```

<210> 44
 <211> 38
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释="人工序列的描述: 插入PA#1(20)聚合物之后, His6-hGH的N末端的氨基酸序列"

```

<400> 44
Ala Ala Ser His His His His His Gly Ala Ser Ser Ser Ala Ala
1          5          10          15

```

```

Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala
20          25          30

```

```

Pro Ala Ala Phe Pro Thr
35

```

<210> 45
 <211> 405
 <212> PRT

[0026]

<213> 人工序列

<220>

<221> 来源

<223> 注释-“人工序列的描述: pASK75- His6-PA#1(200)-hGH上编码的成熟His6-PA#1(200)-hGH的氨基酸序列”

<400> 45

Ala Ser His His His His His His Gly Ala Ser Ser Ser Ala Ala Pro
1 5 10 15

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro
20 25 30

Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala
35 40 45

Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro
50 55 60

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala
65 70 75 80

Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro
85 90 95

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro
100 105 110

Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala
115 120 125

Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro
130 135 140

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala
145 150 155 160

Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro
165 170 175

Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro
180 185 190

Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala
195 200 205

Pro Ala Ala Pro Ala Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe
210 215 220

Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp
225 230 235 240

Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr
 245 250 255
 Ser Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile
 260 265 270
 Pro Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu
 275 280 285
 Leu Leu Arg Ile Ser Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val
 290 295 300
 Gln Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser
 305 310 315 320
 Asp Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln
 325 330 335
 Thr Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile
 340 345 350
 Phe Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp
 355 360 365
 Ala Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met
 370 375 380
 Asp Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu
 385 390 395 400
 Gly Ser Cys Gly Phe
 405

[0027]

<210> 46
 <211> 4393
 <212> DNA
 <213> 人工序列
 <220>
 <221> 来源
 <223> 注释="人工序列的描述: pASK75 His6 PA#1(200) hGH的核酸序列"

<400> 46
 gaccgagac catcgaaatgg ccagatgatt aattctaat ttctgttag acctatcat 60
 tgatagagtt attttaaccac tccctatcag tgatagagaa aagtgaatg aatagttga 120
 caaaaatcta gataacgagg gcaaaaaatg aaaaagacag ctatcgcat tgcagtggca 180
 ctggettggtt tgetaccgt agcgcaggcc gctagccatc accaccatca ceatggegcc 240
 agctctttctg ccgtccagc tgcacctgt ccagcagcac ctgtgcacc agctccggt 300
 gctctgtgt ccgtccagc tgcacctgt ccagcagcac ctgtgcacc agctccggt 360
 gctctgtgt ccgtccagc tgcacctgt ccagcagcac ctgtgcacc agctccggt 420
 gctctgtgt ccgtccagc tgcacctgt ccagcagcac ctgtgcacc agctccggt 480

	gctcctgctg ccgtccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt	540
	gctcctgctg ccgtccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt	600
	gctcctgctg ccgtccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt	660
	gctcctgctg ccgtccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt	720
	gctcctgctg ccgtccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt	780
	gctcctgctg ccgtccagc tgcacctgct ccagcagcac ctgctgcacc agctccggt	840
	gctcctgctg ccttcccaac cattccctta tccaggttt ttgacaacgc tatgtccgc	900
	gcccatcgtc tgcaccagct ggctttgac acctaccagg agttigaaga agcctatata	960
	ccaaaggaac agaagtatc attcctgcag aacccccaga cctccctctg tttctcagag	1020
	tctatccga caccctccaa caggaggaa acacaacaga aatccaacct agagctgctc	1080
	cgcctctccc tgcctgctat ccagtcgtgg ctggagcccg tgcagttcct caggagtgtc	1140
	ttcgccaaca gccctggta cggcgctct gacagcaacg tctatgacct cctaaaggac	1200
	ctagaggaag gcctccaaac gctgatggg aggcctggaag atggcagccc ccggactggg	1260
	cagatcttca agcagacct cagcaagttc gacacaaact cacacaacga tgacgacta	1320
	ctcaagaact aegggtgct ctactgttc aggaaggaca tggacaaggt cgagacatc	1380
	ctggcatcg tgcagtccg ctctgtggag ggcagctgtg gcttctaagc ttgacctgtg	1440
[0028]	aagtgtaaaa tggcgacat tgtgcgacat ttttttgc tgcctttac cgtactgcg	1500
	tcacggatct ccacgcgcc tgtagcgcg cattaagcgc ggcggtgtg gtggttacgc	1560
	gcagcgtgac cgtacactt gccagcgccc tagcgccgc tcttctgct ttcttccctt	1620
	cctttctgc cagttcgcc ggtttccccc gtaagctct aaatggggg ctccctttag	1680
	ggttccgatt tagtcttta cggcacctcg accccaaaaa acttgattag ggtgatggt	1740
	cacgtagtgg gccatcgccc tgatagcgg ttttgcgcc tttagcgttg gattccacgt	1800
	tctttaatag tggactcttg ttccaaactg gaacaacact caacctata tggctatatt	1860
	cttttgattt ataagggtt ttgcgattt cggcctattg gttaaaaaat gactgattt	1920
	aacaaaaatt taacgcgaat ttttaaaaaa tattaacgtt tacaatttca ggiggcactt	1980
	ttcggggaaa tgtgcgcgga acctatatt gttaattttt ctaatacat tcaaatatgt	2040
	atccgctcat gagacaataa cctgataaa tgcctcaata atattgaaa aggaagagta	2100
	tgagtattca acatttccgt gtcgacctta ttcccttttt tgcggcattt tgccttctg	2160
	ttttgtca cccagaaaac ctggigaaag taaaagatgc tgaagatcag ttgggtgcac	2220
	gagtgggtta catcgaaact gatctcaaca gggtaagat ccttgagagt ttteggcccg	2280
	aagaacgttt tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc	2340
	gtattgacgc cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg	2400
	ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat	2460
	gcagtgtgc cataacctg agtgataaca ctgcggccaa cttacttctg acaacgatcg	2520

[0029]

```

gaggaccgaa ggagetaacc gcttttttgc acaacatggg ggateatgta actegcettg 2580
atcgttggga acggagcctg aatgaagcca taacaaaga cgagcglgac accacgatgc 2640
ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt 2700
ccccgaaca attgatagac tggatggagg cggataaagt tgcaggacca cttctgeget 2760
cgcccttcc ggctggcctg tttattgctg ataaatcctg agcgggtgag cgtggetctc 2820
gggttactat tgcagcactg ggcccagatg gtaagcctc ccgtatcgta gttatctaca 2880
cgacggggag teaggcaact atggatgaac gaaatagaca gatcgtgag ataggtcct 2940
cactgattaa gcattggtag gaattaatga tgtctcgttt agataaaagt aaagtgatta 3000
acagcgcatt agagetgett aatgaggctg gaatcgaagg tttacaacc cgtaaactcg 3060
cccagaagct aggtgtagag cagctacat tgtattggca tgtaaaaat aagcgggctt 3120
tgctcgacgc cttagccatt gagatgttag ataggcacca tactcacttt tgccttttag 3180
aaggggaaag ctggcaagat tttttacgta ataacgctaa aagtittaga tgtgctttac 3240
taagtcacgc cgtggagca aaagtacatt taggtacacg gctacagaa aaacagiatg 3300
aaactctega aaatcaatta gcttttttat gccaaacagg ttttacta gagaatgcatt 3360
tataatcact cagcgcagtg ggccatttta ctttaggttg cgtatiggaa gatcaagagc 3420
atcaagtcgc taagaagaa agggaaacac ctactactga tagtatgcgc ccattattac 3480
gacaagctat cgaattattt gatcaccaag gtgcagagcc agccttctta ttgcgecttg 3540
aatlgatcat atggggatta gaaaacacac taaatgiga aatgggctt taaaagcagc 3600
ataacctttt tccgtgatgg taacttcaat agttttaaag gatctaggtg aagatccttt 3660
ttgataatct catgaccana atccctaac gtgagtttct gttccactga gcgtcagacc 3720
ccgtagaaaa gatcaaaagg tcttcttgag atcccttttt tctgcgcgta atctgctgct 3780
tgcaaacaaa aaaaccaccg ctaccagcgg tggtttggtt gcgggatcaa gactaccaa 3840
ctctttttcc gaaggtaact ggcctcagca gagegcagat accaataact gtccttctag 3900
tgtagccgta gttagccac cacttcaaga actctgtagc accgcctaca taactcctc 3960
tgctaactct gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg 4020
actcaagacg atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca 4080
cacagcccag ctggagcga acgaactaca ccgaactag atacctacag cgtgagctat 4140
gagaagcgc cacgcttccc gaaggagaa agcgggacag gtatccgta agcggcaggg 4200
tcggaacagg agagcgcacg agggagcttc cagggggaaa cgcctggtat ctttatagtc 4260
ctgtcggggt tgcacactc tgacttgagc gtcgttttt gtgaltcctg tcaggggggt 4320
ggagcctatg gaaaaacgc agcaacgcgg cctttttacg gttccttgcc ttttgcgtgc 4380
cttttgcctc cat 4393

```

<210> 47
<211> 404
<212> PRT
<213> 人工序列

[0030]

```

<220>
<221> 来源
<223> 注释="人工序列的描述: pCHO-PA#1 (200)-hGH上编码的His6-PA#1 (200)-hGH的氨基酸序列"
<400> 47
Ser His His His His His His Gly Ala Ser Ser Ser Ala Ala Pro Ala
1          5          10          15
Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
20          25          30
Ala Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
35          40          45
Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
50          55          60
Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro
65          70          75          80
Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
85          90          95
Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
100         105         110
Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
115         120         125
Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
130         135         140
Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro
145         150         155         160
Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
165         170         175
Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
180         185         190
Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
195         200         205
Ala Ala Pro Ala Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp
210         215         220
Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr
225         230         235         240
Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser
245         250         255

```

[0031]

Phe Leu Gln Asn Pro Gln Thr Ser Leu Cys Phe Ser Glu Ser Ile Pro
260 265 270

Thr Pro Ser Asn Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu
275 280 285

Leu Arg Ile Ser Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln
290 295 300

Phe Leu Arg Ser Val Phe Ala Asn Ser Leu Val Tyr Gly Ala Ser Asp
305 310 315 320

Ser Asn Val Tyr Asp Leu Leu Lys Asp Leu Glu Glu Gly Ile Gln Thr
325 330 335

Leu Met Gly Arg Leu Glu Asp Gly Ser Pro Arg Thr Gly Gln Ile Phe
340 345 350

Lys Gln Thr Tyr Ser Lys Phe Asp Thr Asn Ser His Asn Asp Asp Ala
355 360 365

Leu Leu Lys Asn Tyr Gly Leu Leu Tyr Cys Phe Arg Lys Asp Met Asp
370 375 380

Lys Val Glu Thr Phe Leu Arg Ile Val Gln Cys Arg Ser Val Glu Gly
385 390 395 400

Ser Cys Gly Phe

<210> 48

<211> 6653

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: pCHO-PA#1(200)-hGH的核酸序列"

<400> 48

ctattctata gtgtcaccta aatgctagag ctgctgac agcctegact gtgccttcta 60

gttgcagcc atctgttgtt tgcctctccc ccgtgccttc cttgacctg gaagggtcca 120

ctcccactgt cctttcctaa taaaatgagg aaattgcac gcattgtctg agtaggtgtc 180

attctattct ggggggtggg gtggggcagg acagcaagg ggaggattgg gaagacaata 240

gcaggcatgc tgggatgcg gtgggtctta tggttctga ggaggaaaga accagctggg 300

gctctagggg gtatccccac gcgcctgta gcgggcatt aagcgcggcg ggtgtgtgg 360

ttaacgcag cgtaaccgt acacttgca gcgcctagc gccgctcct ttgccttct 420

tcccttctt tctgcacg ttgcgggt ttcctgtca agctctaaat eggggcatcc 480

ctttagggtt ccgatttagt gtttaaggc acctegacc caaaaaactt gattagggtg 540

	atggttcaeg tagtgggcca tgccectgat agacggtttt tgccectttg acgttggagt	600
	ccacgttctt taatagtga ctcttgctcc aaactiggaac aacactcaac cctatctcgg	660
	tctattcttt tgatttataa gggattttgg ggatttcggc ctatttggtta aaaaatgagc	720
	tgatttaaca aaaatttaac gcgaattaat tcigtggaat gtgtgtcagt taggggtgtg	780
	aaagtcceca ggctceccag gcaggcagaa gtatgcaaag catgcatctc aattagtcag	840
	caaccaggtg tggaaagtcc ccaggctccc caggcaggcag aagtatgcaa agcatgcac	900
	tcaattagtc agcaaccata gtcccgcgcc taactcgcgc catccgcgcc ctaactcgc	960
	ccagttccgc ccattctcgc ccccatggct gactaatitt ttttatttat gcagaggccg	1020
	aggccgcctc tgcctctgag ctattccaga agtagtgagg aggccttttt ggaggcctag	1080
	gcttttgcaa aaagctccc ggagcttgta tatccatttt cggatctgat caagagacag	1140
	gatgaggatc gtttcgcagc atigaacaag atggattgca cgcaggttct ccggccgctt	1200
	gggtgagag gctattcggc tatgaatggg cacaacagac aatcgctgc tctgaigccg	1260
	ccgtgttccg gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg	1320
	gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc acgacgggcg	1380
	ttccttgcgc agctgtgctc gacgttgta ctgaagcggg aagggaactgg ctgetattgg	1440
	gcgaagtgc ggggcaggat ctctgtcat ctcaacttgc tctgcgcag aaagtatcca	1500
[0032]	tcatggctga tgaatgcgg cggctgcata cgtttgatcc ggctacctgc ccattcgacc	1560
	accaagcgaa acatgcacg gagcagagac gtactcggat ggaagccggt cttgtcgatc	1620
	aggatgatct ggacgaagag catcaggggc tgcgcgcagc cgaactgttc gccagctca	1680
	aggcgcctat gccgcagcg gaggatctcg tcgtgaccca tggcgatgcc tgcctgcga	1740
	atatcatggt ggaaaatggc cgtttttctg gattcatcga ctgtggcccg ctgggtgtgg	1800
	cggaccgcta tcaggacata gcgttgcta cccgtgatat tgcgtgaagag cttggcggcg	1860
	aatgggctga ccgttctcgc gtgctttacg gtatgcgcgc tcccgattcg cagcgcacgc	1920
	ctttctatcg ctttcttgac gattctctct gagcgggact ctggggttcg aaatgaccga	1980
	ccaagcgagc cccaacctgc catcacgaga ttctgattcc acgcgcgcct tctatgaaag	2040
	gttgggcctc ggaatcgttt tccgggacgc cggctggatg atcctccagc gcggggatct	2100
	catgctggag ttcttcgcgc accccaactt gtttattgca gattataatg gttacaaata	2160
	aagcaatagc atcacaatc tcacaaataa agcatttttt tcaactgcatt ctagtgtgtg	2220
	tttgtccaaa ctcatcaatg tatcttatca tgtctgata ccgtcgacct ctagetagag	2280
	cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttaicgc tcacaattcc	2340
	acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaat gagtgcgcta	2400
	actcacatta attgcgttgc gctcactgcc cgttttcag tgggaaacc tgtcgtgcca	2460
	gtcgcattaa tgaatcgccc aacgcgcggg gagaggcggg ttgcgtattg ggcgccttc	2520
	cgttctctcg ctcaactgact cgtctgcctc ggctgttcgg ctgcggcgag cggatcagc	2580

[0033]

tcactcaaag gcggtataac ggttatecac agaatacagg gataacgcag gaaagaacat	2640
gtagacaaaa ggcacagaaa aggcchggaa ccglanaaag gccgcgllgc tggcglllil	2700
ccataggetc cgcceccctg acgagcatca caaaaatcga cgetcaagtc agaggtggcg	2760
aaacccgaca ggaactataa gataccagge gtttcccctt ggaagctccc tegtgcgtc	2820
tcctgttccg accctgcgcg ttaccggata cctgtccgce ttctccctt cgggaagcgt	2880
ggcgttttct caatgcctac gcgttaggta tctcagttcg gtgtaggteg ttcgtccaa	2940
gcgtggctgt gtgcacgaac ccccgcttca gcccgaccgc tgcgccttat ccgtaacta	3000
tcgtcttgag tccaaccccg taagacacga ctatccgcca ctggcagcag ccactggtaa	3060
caggattage agagcgaggt atgtaggcgg tctacagag ttcttgaagt ggtggcctaa	3120
ctaaggctac actagaagga cagtatttgg tatctgcgtt ctgtcgaagc cagttaacct	3180
cggaaaaaga gtgttagctt ctgacccgg caaacaaacc accgctggta ggggtgttt	3240
tttgtttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag atcctttgat	3300
cttttctacg gggtctgacg ctcagtggaa cgaaaactca cgttaaggga tttgtgcat	3360
gagattatca aaaaggatct tcaactagat ctttttaa ataaaaatga gttttaaate	3420
aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc	3480
acctatctca gcatctgtc tatttcttcc atccatagtt gcctgaacct ccgtctgta	3540
gataactacg atacgggagg gcttaccatc tggcccagc gctgcaatga taccgcgaga	3600
cccacgctca cggctccag atllalcage aataaaccag ccagccggaa gggcgcagcg	3660
cagaagtggc ccgcacatt tatccgctc catccagctt attaatgtt gccgggaagc	3720
tagagtaagt agttcccgag ttaafagttt gcgcaacgtt gttgccattg ctacagcat	3780
cgtgggtgca cgtctgtctt ttggtatggc tteattcagc tccgttccc aacgatcaag	3840
gcgagttaca tgatccccc tgtgtgcaa aaaagcgggt agctccctcg gtcctccgat	3900
cgttgctaga agtaagtgg ccgcagtggt atcactcatg gttatggcag caatgcataa	3960
ttctcttact gtcatgccat ccgtaagaat cttttctgt actggtgagt actcaacca	4020
gtcattctga gaatagtga tgcggcgacc gagggtctt tgcggcgct caatacggga	4080
taataccgag ccacatagca gaactitaaa agtgcctatc attggaanaa gttcttcggg	4140
gcgaaaaact tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccactcgtc	4200
acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg	4260
aaggcaaaat gccgcaaaa aggaataaag ggcgacacgg aaatgttgaa tactcatact	4320
cttcttttt caataatatt gaagcaitla tcagggttat tgtctcatga gggatatac	4380
atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt	4440
gccacctgac gtcagcgat cgggagatct cccgacccc tatggtgac tctcagtaaa	4500
atctgctctg atgcgcata gtttaagccag tatctgtccc ctgcttgtgt gttggaggtc	4560
gtgagtagt gccgagcaa aatlaagct acaacaagga aagcttgaac cgaatitgc	4620
atgaagaate tgcctagggt taggcgtttt gcgtctcttc gcgatgtac gccagatat	4680

acgcgttgac attgattatt gactagttaa taatagtaat caattacggg gtcattagtt	4740
catagcccat atattggagt ccgcgttaca taacttacgg taaatggccc gcctggctga	4800
ccgcccaacg acccccgcgc attgacgtca ataatacgt atgttcccat agtaacgcca	4860
atagggaactt tccattgacg tcaatgggtg gactatttac ggtaaactgc ccacttgcca	4920
gtacatcaag tgtatcatat gccaaagtag cccctatig acgtcaatga cggtaaattg	4980
cccgcctggc attatgccc gtacatgacc ttatgggaact ttctacttg gcagtacac	5040
taegtattag tcatgctat taccatggtg atgggtttt ggcagtacat caatgggct	5100
ggatagcggg ttgactcagc gggatttcca agtctccacc ccattgacgt caatgggagt	5160
ttgttttggc accaaaatca acgggaactt ccaaatgtc gtaacaactc cgcctcattg	5220
acgcaaatgg gcggtaggcg tgtacgggtg gaggtctata taagcagagc tctctggeta	5280
actagagaac ccactgctta ctggcttacc gaaattaata cgactcacta tagggagacc	5340
caagetgcca ccattgctac aggtcccg gcgtccctgc tctggtctt ttgctgctc	5400
tgcctgcctt ggcttcaaga gggcagcgt agccatcacc accatcacca tggcgcagc	5460
tctctgcctt ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5520
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5580
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5640
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5700
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5760
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5820
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5880
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	5940
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	6000
cctgctgcct ctccagctgc acctgctcca gcagcaccctg ctgcaccagc tccggtctct	6060
cctgctgcct tcccaaccat tcccttacc aggtttttt acaacgtat gctccgcgc	6120
catgctctgc accagctgac cttgacacc taccaggagt tgaagaagc ctatatccca	6180
aaggaaacga agtattcatt cctgcagaa cccagacct cctctgttt ctacaggtct	6240
attccgacac cctccaacag ggaggaaaca caacagaat ccaacctaga gctgctcgc	6300
atctccctgc tgcctcaca gtcgtggtg gacccctgc agtctctcag gactgtcttc	6360
gccaaacagc tgggtacgg cgcctctgac agcaacgtct atgaactcct aaaggaccta	6420
gaggaaggca tccaaacgt gatggggagg ctggaagatg gcagccccc gactgggcag	6480
atcttcaagc agaactacag caagttagc acaaaactac acaacgaatg cgcactactc	6540
aagaactacg ggtgctctca ctgtctcagg aaggacatgg acaaggctga gacattctg	6600
cgcactgtgc agtgcctgc tgtggagggc agctgtggtc tctaagcttg gcc	6653

[0034]

<211>	6041	
<212>	DNA	
<213>	人工序列	
<220>		
<221>	来源	
<223>	注释="人工序列的描述: pCHO-hGH的核酸序列"	
<400>	49	
	ctattctata ggttcaccta aatgetagag ctgctgate agcctcgact gtgccttcta	60
	gttgccagcc atctgttgtt tgcctctccc cgtgcttcc ctgacccctg gaaggtagcca	120
	ctcccactgt cctttcctaa taaaatgagg aaattgcac gcattgtctg agtaggtgtc	180
	attctattct ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata	240
	gcaggcatgc tggggatgcg gtgggctcta tggcttctga ggcggaaaga accagctggg	300
	gctctagggg gtatccccc gcgcccctga gcggcgcat aagcggcgcg ggtgtggtgg	360
	ttacgcgcag cgigaccgtt acacttgcca gcgcccctag gcccgctctt ttcgctttct	420
	tcccttctct tctgcccacg ttgcggcgct ttcccctga agctctaaat cggggcatcc	480
	cttttaggggt cegatttagt gctttacggc acctcgacc caaaaaactt gattagggtg	540
	atggttcacg tagtgggcca tgcctctgat agacggtttt tgcgctttg acgttggagt	600
	ccacgttctt taatagtggg ctcttgttcc aaactggaac aacactcaac cctatctcgg	660
	tctattcttt tgatttataa gggattttgg ggatttcggc ctatttggtt aaaaatgagc	720
[0035]	tgatttaaca aaaatttaac ggaattaat tctgtggaat gtgtgtcagt tagggtgtgg	780
	aaagtcccca ggtcccccag gcaggcagaa gtatgcaaag catgcctctc aattagtcat	840
	caaccagggtg tggaaagtc ccaggctccc caggcggcag aagtatgcaa agcatgcac	900
	tcaattagtc agcaaccata gtcccgccc taactcggc catcccgccc ctaactcgc	960
	ccagttcgc ccattctccg ccccatggtt gactaatttt ttttatttat gcagagggcg	1020
	aggccgccc tgcctctgag ctattccaga agtagtgagg aggtttttt ggaggccatg	1080
	gcttttgcaa aaagctccc ggagcttgta tatccattt cggatctgat caagagacag	1140
	gatgaggatc gtttcgatg attgaacaag atggattgca cgcaggttct ccggccgctt	1200
	gggtggagag gctattcggc tatgaactggg cacaacagac aatcggctgc tctgatccg	1260
	cctgttccg getgtcagc caggggcgcc cggttctttt tgtaagacc gacctgtccg	1320
	gtgcctgaa tgaactgcag gacgaggcag cgcggctatc gtggttgccc acgacggcg	1380
	ttccttgcgc agctgtgctc gacgttgta ctgaagcggg aagggaactg ctgtatttgg	1440
	gcgaagtgc ggggcaggat ctctgtcat ctcaecttgc tctgcccag aaagtatcca	1500
	tcatggtga tgcaatgcg cggtgcata cgttgatcc ggctacctgc cattedgacc	1560
	accaagcgaa acatgcacg gacgagcac gtaactggat ggaagccggt ctgtctgac	1620
	aggatgatct ggacgaagag catcaggggc tgcgccagc cgaactgttc gccaggtca	1680
	aggcgcgcac gccgcaggc gaggatctg tctgaacca tggcgatgcc tcttgcgga	1740
	atatcatggt ggaaaatgac cgtttttctg gattcatcga ctgtggccgg ctgggtgtgg	1800

[0036]

```

eggaccgeta tcaggacata gcgttggeta cecgtgatat tgctgaagag cttagcgggcg 1860
aatgggetga cegcttcctc gtcctttacg gtatcgccgc tcccgattcg cagcgcacg 1920
ccttcctatcg ccttccttac gagttcttct gagegggact cgggggttcg aatgaccga 1980
ccaagcgacg cccaacctgc catcaagaga tttegattee accgcgcctt tctatgaaag 2040
gltaggcctc ggaacgctt tccgggacgc cggctggatg atccctccgc ggggggact 2100
catgctggag ttcttcgcgc accccaactt gttatttga gttatattg gttacaata 2160
aagcaatagc atcacaatt tcacaaataa agcatttttt tcaatgcatt ctagtgttg 2220
tttgctcaaa ctcactaatg tatcttata tctctgata ccttcgacct ctactagag 2280
cttagcgtaa tcatgtcat agctgttccc tgtgtgaat tgttatccg tcacaattcc 2340
acacaacata cgagccggaa gcataaagt taaagcctgg ggtgcctaat gattgagca 2400
actcacatta attagcgtgc gctcactgcc cgtctccag tggggaacc tctctgcca 2460
gtctcattaa tgaatcgcc aacgcgcggg gagagcggt ttgcgtattg gcgcctctc 2520
cgttctctcg ctactgact cgtgcgcgc ggtcgttcgg ctgcggcgag cggatcagc 2580
tcactcaaac gggtaaac ggtaaacac agaacaggg gataacgcg gaaagacat 2640
gtgagcaaaa ggccagcaaa agccagga cgttaaaag gcgcgttgc tggcgtttt 2700
ccatagctc cgcctccctg acgagcatca caaaaatga cgtcaagtc agaggtggcg 2760
aaaccgcaca ggaclataaa gataccaggc gttccccc tgaagctcc tctgcgcct 2820
tctgttcg accctgcgc ttaccggata cctgtccgc ttctccct cgggaagct 2880
ggcgtttct caatgctac gctgtagga tctcgttcg gctgtaggc ttgcctcaa 2940
gtcggcgtg ggcacgaac ccccgctca gccgcgcgc tgcgcctat ccggttaata 3000
tcgtctgag tcaaccgcg taagacaga ctatcgcca ctgcgcagc cactggtaa 3060
caggattagc agagcgaggt atgtagcgg tctacagag ttctgaagt ggtgcctaa 3120
ctacggctac actagaagga cagtattgg tatctgcgt ctgcgaagc cagttaact 3180
cggaaaaaga gttggtagct ctgatacgc caaacaacc accgtggtg gcggtggtt 3240
ttttgttgc aagcgcaga ttacgcgcg aaaaaagga tctcaagaag atcctttgat 3300
ctttctacg ggtctgacg ctacgtgga cgaacctca cgttaaggga ttgtgctat 3360
gagattatca aaaggtatc tcacctagat ctttttaat taataatga gttttaata 3420
aatctaaagt atataagat aaacttggtc tgacagtta caatgctaa tcaatgagc 3480
acclactca ggaactgtc tatctgtc atccatagt gccgacct cctctgta 3540
gataactacg atagggagg gcttaccatc tggcccgat gctcaatga taccgcaga 3600
cccacgctca cggctccag attatcagc aataaccag ccagccgaa ggccgagcg 3660
cagagtggt cctgaactt tatccgctc catccagct attaatgtt gccgggaag 3720
tagagtaagt agtgcgcag ttaatatgt ggcacagtt gttgcattg ctacagcat 3780
cgtggtgca cgtctcgt ttgtatgc ttactcagc tccgttccc aacgatcaag 3840
gcgattaca tgatcccca tgtgtgcaa aaaagcggt agctcttcg gctctcgt 3900

```

[0037]

cgttgctcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag cactgcataa 3960
 ttctcttaact gtcattgccat ccgtaagatg cttttctgtg actgggtgagt actcaaccaa 4020
 gtcattctga gaatagtgta tgcgcgcacc gagtggctct tgcctggcgt caatacggga 4080
 taataccgcg ccacatagca gaactttaaa agtgcctatc attggaaaaa gttcttcggg 4140
 gegaaaacte tcaaggatct taccgtgttt gagatccagt tcgatgtaac ccactcgtgc 4200
 acccaactga tcttcagcat cttttacttt caccagcgtt tetgggtgag caaaaacagg 4260
 aaggcaaaat gccgcaaaaa agggaaataag ggcgacacgg aaatgttgaa tactcatact 4320
 ctctcttttt caatattatt gaagcattta tcagggttat tgtctcaiga ggggatacat 4380
 atttgaaigt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt 4440
 gccacctgac gtgcacggat cgggagatct cccgateccc tatggtegac tctcagtaca 4500
 atctgctctg atgcgcata gtttaagcag tatctgctcc ctgcttgtgt gttggaggtc 4560
 gctgagtagt gcgcgagcaa aatttaagct acacaagge aagcttgac cgacaattgc 4620
 atgaagaatc tgcctagggt taggcgtttt gcctgcttc gcgatgtac ggccagatat 4680
 acgcgttgac attgattatt gactagtatt taatagtaat caattacggg gtcattagtt 4740
 catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gccggctga 4800
 ccgcaccaag acccccgcgc attgacgtca ataatgacgt atgttcccat agtaacgcca 4860
 atagggaactt tccattgacg tcaatgggtg gactatttac ggtaaaactgc ccacttgcca 4920
 gtacatcaag tgtatcatat gccaaagtag cccctattg acgtcaatga cggtaaattgg 4980
 cccgcctggc attatgccc gtaactgacc ttatgggaact ttctacttg gcagtacatc 5040
 tacgtattag tcatcctat taccatggtg atgcggtttt ggcagtacat caatgggcgt 5100
 ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt 5160
 ttgttttggc accaaaaatca acgggaacttt ccaaaatgtc gtaacaacte cgccecatg 5220
 acgcaaatgg gcggtaggcg tglacgggg gaggtctata taagcagagc tctctggcta 5280
 actagagaac ccactgctta ctggttatac gaaattaata cgactcaata tagggagacc 5340
 caagctgcca ccattgctac aggtcccggt aegtcctgc tcttggttt ttgctctgc 5400
 tgcctgcctt ggttcaaga gggcagcgt agccatcac accatcacca tggcgccttc 5460
 ccaaccattc cttatccag gctttttgac aacgtatgc tccgcgcca tgcctgcac 5520
 cagctggcct ttgacaacta ccaggagttt gaagaagcct atatccaaa ggaacagaag 5580
 tattcattec tgcagaacct ccagacctcc ctctgtttct cagagtctat tccgacacc 5640
 tccaacaggg aggaacaca acagaaatcc aaactagagc tgcctgcct ctccctgctg 5700
 ctcatccagt cgtggtgga gcccgtcag tctctcagga gtgtcttcgc caacagcctg 5760
 gtgtacggcg cctctgacag caacgtctat gacctctaa aggacctaga ggaaggcatc 5820
 caaacgtga tggggaggct ggaagatgc agccccgga ctgggcagat ctcaangcag 5880
 acctacagca agttcgacac aaactcacac aacgatgac cactactcaa gaactacggg 5940

	ctgetctact gcttcaggaa ggacatggac aaggctcgaga cattcctgcg catcgtgcag	6000
	tgccgctctg tggagggcag ctgtggettc taagettggc c	6041
	<210> 50	
	<211> 5449	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<221> 来源	
	<223> 注释="人工序列的描述: pCHO的核酸序列"	
	<400> 50	
	ctattctata gtgtcaccta aatgctagag ctegetgac agcctcgact gtgccttcta	60
	gttgccagcc atctgttgtt tgcctctccc ccgtgcttc ctgacctg gaagggtcca	120
	ctcccacigt cctttcctaa taaaatgagg aaattgcac gcatgtctg agtaggtgtc	180
	attctattct ggggggtggg gtggggcagg acagcaaggz ggaggattgg gaagacaata	240
	gcaggcatgc tggggatgag gtgggctcta tgccttctga ggcggaaaga accagctggg	300
	gtctagggg gtaacccac gcgccctgta ggcgcgcat aagcgcggcg ggtgtggtgg	360
	ttacgcgcag cgtgacctg acacttgcca gcgccctag gccgctct ttcgtttct	420
	tccttctctt tctgccacg ttgcgcggt tccccgtca agctctaaat cggggcctcc	480
	ctttagggtt ccgatttagt gcttaagcc acctcgaccc caaaaaactt gattagggtg	540
[0038]	atggttcacg tagtgggcca tcgccctgat agacggtttt tcgcccttg acgttggagt	600
	ccacgttctt taatagtga ctcttgctcc aaaciggaa acacactcaac cctatctcg	660
	tctattcttt tgatttataa gggattttgg ggatttcgga ctatttggtt aaaaatgagc	720
	tgatttaaca aaaatttaac gcgaattaat tctgtggaat gtgtgtcagt taggtgtgg	780
	aaagtcccca ggtcccccag gcaggcagaa gtatgcaaag catgccttc aattagtcag	840
	caaccagggtg tggaaagtcc ccaggtccc cagcaggcag aagtatgeaa agcatgcac	900
	tcaattagtc agcaaccata gtcccgccc taactccgc cctccgccc ctaactccg	960
	ccagttccgc ccattctccg ccccatggt gactaatttt tttatttat gcagagccg	1020
	aggccgctc tgcctctgag ctatccaga agtagtgagg agctttttt ggaggcctag	1080
	gcttttgcaa aaagctccg ggagcttgta tatccatttt cggatctgat caagagacag	1140
	gatgaggatc gtttcgatg attgaacaag atggattgca cgcaggttct ccggccgctt	1200
	gggtggagag gctattcgga tatgaatggg cacaacagac aatcgctgc tctgatgcg	1260
	ccgtgtccg gctgtcagc caggggcgcc cggttctttt tctcaagacc gacctgtcc	1320
	gtgcctgaa tgaactgcag gacgaggcag cgcggctatc gtggtgccc acgacggcg	1380
	ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg aagggaactg ctgctattgg	1440
	gcgaagtgc ggggcaggat ctctgtcat ctacattgc tctgcccag aaagtatcca	1500
	tcatggtga tgaatgcg cggctgcata cgttgatcc ggtacctgc ccattcgacc	1560
	accaagegaa acatgcacg gacgagcac gtactcggat ggaagccggt ctgtctgac	1620

[0039]

```

aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggetca 1680
aggcgcgcct gcccgacggc gaggatctcg tcgtgacca tggcgatgcc tgcttgccga 1740
atatcatggt ggaaatggc cgtttttctg gattcatcga ctgtggcccg ctgggtgtgg 1800
eggaccgcta tcaggacata gegttagcta cccgtgatai tgcgaagag ctggcgggcg 1860
aatgggetga cegcttctc gtgttttaag giatcgccgc tcccgattcg cagcgcctcg 1920
ctttctatcg cctttctgac gagttcttct gagcgggact ctgggggttcg aatgaccga 1980
ccaagcgacg cccaacctgc catcacgaga ttctgattcc accgccgct tctatgaaag 2040
gttgggettgc ggnatcgitt tccgggaecg cggttgatg atctccagc ggggggatct 2100
catgctggag ttcttcgcc accccaactt gttatttgcg gattataatg gttacaaata 2160
aagcaatage atcacaattt tcacaaataa agcatttttt tcaatgcatt ctagtgtgg 2220
tttgccaaa ctcatcaatg tatcttata tgctgtata ccgtcgacct ctagctagag 2280
cttggcgtaa tcatggtcat agctgttcc tgtgtgaat tgttatccgc tcacaattcc 2340
acacaacata cgaacggaa gcataaagtg taaagccgtg ggtgcctaat gagtgaacta 2400
actcacatta attgcttgc gctacgtcc cgctttccag tccggaaacc tgcgtgcca 2460
gctgcattaa tgaatcgcc aacgcgcggg gagaggcgtt ttgcgtatg ggcgtcttc 2520
cgcttctctg ctcaetgact cgcgcgcgc ggctgttcgg ctgcggcgag cggtatcagc 2580
tcactcaag gggtaatac gttatccac agaatcaggg gataacgcag gaaagacat 2640
gtgagcaaaa ggccagcaaa agccagga cegtaaaag gccgcgttgc tggcgttttt 2700
ccataggetc cgccectg acgagcata caaaaatga cgtcaagtc agagggtggc 2760
aaaccgcaca ggaactataa gataccagc gtttccctt ggaagctcc tcgtgcctc 2820
tcctgttcg acctgcgcg ttaccgata cctgtccgc tttctccct cgggaagcgt 2880
ggcgctttct caatgctac gctgtaggta tctcagttcg gtgtaggtcg ttgcctcaa 2940
gttgggetgt gtgcacgaac ccccgctca gcccgaccgc tgcgcttat ccggttaacta 3000
tcgttttag tcaaccgg taagacaga ctatcgcca ctggcagcag ccaatggtaa 3060
caggattagc agagcgaggt atgtaggcg tctacagag tcttgaagt ggtggcctaa 3120
ctacgctac actagaagga cagtatttg tatctgcgt ctgtgaagc cagttacct 3180
cgaaaaaga gttgtagct ctgatccg caacaaacc accgtgcta gcgttggtt 3240
ttttgttgc aagcagcaga ttacgcgcg aaaaaagga tctcaagaag atccttgat 3300
ctttctacg ggtctgacg ctcatggaa cgaaaactca cgttaaggga ttttgctat 3360
gagattatca aaaaggatct tcacctagat ctttttaat taanaatga gttttaate 3420
aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagttaggc 3480
acctatctca gcatctgtc tatttcgtc atccatagtt gcctgactcc cgtctgtta 3540
gataactacg atacgggagg gctaacatc tggcccgat gctgaatga taccgcgaga 3600
cccacgctca cggctccag attatcagc aataaaccag ccagccgga ggccgagcg 3660
cagaagtgtt cctgaactt tategcctc catcagttct attaatgtt gccgggaage 3720

```

[0040]

```

tagagtaagt agttegccag ttaatagttt ggcgaacgtt gttgccattg ctacaggcat 3780
cgtagtgtea cgtctgtgtt ttggtatggt ttcattcagc tccggttccc aacgatcaag 3840
gcgagttaca tgatecccca tgttgtgcaa aaaagcgggt agtccttcg gtctccgat 3900
cgttgtcaga agtaagttag ccgcagtggt atcactcatg gttatggcag cactgcataa 3960
ttctcttaet gtcatgccat ccgtaagatg cttttctgtg actggtgagt actcaacca 4020
gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgcctggcgt caatacggga 4080
taataccgcg ccacatagca gaactttaaa agtctcctc attggaaaaa gtctctcggg 4140
gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccactcgtgc 4200
acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg 4260
aaggcaaaat gcgcgaaaaa agggataaag ggcgacacgg aaatgttgaa tactcatact 4320
cttccititt caatattati gaagcattta tcagggttat tgtctcatga gcggatacat 4380
atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt 4440
gccacctgac gtcgacggat cgggagatct cccgatcccc tatggctgac tctcagtaca 4500
atctgctctg atgcgcata gtttaagcag tatctgctcc ctgcttgtgt gttggaggtc 4560
gcgagtagt gcgcgagcaa aatttaaget acaacaaggc aaggttgac cgacaattgc 4620
atgaagaatc tcttiagggt tagcggtttt gcctgcttc gcgatgtacg ggccagatat 4680
acgcgttgac attgattatt gactagtatt taatagtaat caattacggg gtcattagtt 4740
catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga 4800
ccgcceaacg acccccgcgc attgaegtea ataatgaagt atgtcccat agtaacgcca 4860
aiagggaatt tccattgacg tcaatgggtg gactatitac ggtaaaactgc ccactlgcca 4920
gtacatcaag tgtaicatat gccaaagtacg cccctattg acgtcaatga cggtaaatgg 4980
cccgcctggc attatgccc gtacatgacc ttaigggact ttctacttg gcagtacac 5040
tacgtattag tcatgctat taccatggtg atgcggtttt ggcagtacat caatggcggt 5100
ggatagcggg ttgaetcaag gggatttcca agtctccacc ccattgacgt caatgggagt 5160
ttgttttggc accaaaaatea acgggacttt ccaaaaatgt gtaacaacte cgccecatg 5220
acgcaaatgg gcggtaggcg tgtacgggtg gaggctctata taagcagagc tctctggcta 5280
actagagaac ccactgctta ctgcttctc gaaatttaata cgactcacta tagggagacc 5340
caagetgcca ccattgctac aggtcccggt acgtccctgc tctggtttt tggectgctc 5400
tgctgcctt ggttcaaga gggcagcgt agctctagaa agcttgccc 5449

```

<210> 51

<211> 20

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: P1A1的氨基酸序列"

[0041]

<400> 51

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
 1 5 10 15

Ala Pro Ala Pro
 20

<210> 52

<211> 60

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: 用于产生P1A1的构成部分的上方/编码链寡脱氧核苷酸的核酸序列"

<400> 52

gccccgtgctc ctgctccagc acctgcacca gcacctgctc cagcaccagc tctgcacca 60

<210> 53

<211> 60

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: 用于产生P1A1的构成部分的下方/非编码链寡脱氧核苷酸的核酸序列"

<400> 53

ggctgggtgca ggagctgggtg ctggagcagg tctgggtgca ggtgctggag caggagcagg 60

<210> 54

<211> 60

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: 用于产生P1A3的构成部分的上方/编码链寡脱氧核苷酸的核酸序列"

<400> 54

ggcgtgtcac ctgctgcagc acctgctgca gctccagcag ctgctcctgc agcagctcca 60

<210> 55

<211> 60

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: 用于产生P1A3的构成部分的下方/非编码链寡脱氧核苷酸的核酸序列"

<400> 55

ggctggagct gctgcaggag cagctgctgg agctgcagca ggtgctgcag caggtcagc 60

<210> 56

<211> 417

<212> PRT

<213> 人工序列

<220>

<221> 来源

[0042]

<223> 注释="人工序列的描述:pFab-P1A1(200)上编码的与P1A1(200)聚合物融合的Fab轻链的氨基酸序列"

<400> 56

Asp Ile Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala
20 25 30

Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr Thr Thr Pro Pro
85 90 95

Thr Phe Gly Gln Gly Thr Lys Leu Glu Ile Lys Arg Thr Val Ala Ala
100 105 110

Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly
115 120 125

Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala
130 135 140

Lys Val Gln Trp Lys Val Asp Asn Ala Leu Gln Ser Gly Asn Ser Gln
145 150 155 160

Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr Tyr Ser Leu Ser
165 170 175

Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val Tyr
180 185 190

Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser
195 200 205

Phe Asn Arg Gly Glu Cys Ser Ser Ala Pro Ala Pro Ala Pro Ala Pro
210 215 220

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
225 230 235 240

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
245 250 255

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

[0043]

[illegible]

Ala

<210>	57
<211>	417
<212>	PRT
<213>	人工序列

<220>	
<221>	来源
<223>	注释="人工序列的描述: pFab-P1A3(200)上编码的与P1A3(200)聚合物融合的Fab轻链的氨基酸序列"
<400>	57

Asp Ile Glu Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp Val Asn Thr Ala
20 25 30

Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

	Ser	Arg	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr	Ile	Ser	Ser	Leu	Gln	Pro	
	65					70					75					80	
	Glu	Asp	Phe	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	His	Tyr	Thr	Thr	Pro	Pro	
				85						90					95		
	Thr	Phe	Gly	Gln	Gly	Thr	Lys	Leu	Glu	Ile	Lys	Arg	Thr	Val	Ala	Ala	
			100						105					110			
	Pro	Ser	Val	Phe	Ile	Phe	Pro	Pro	Ser	Asp	Glu	Gln	Leu	Lys	Ser	Gly	
			115					120					125				
	Thr	Ala	Ser	Val	Val	Cys	Leu	Leu	Asn	Asn	Phe	Tyr	Pro	Arg	Glu	Ala	
		130					135					140					
	Lys	Val	Gln	Trp	Lys	Val	Asp	Asn	Ala	Leu	Gln	Ser	Gly	Asn	Ser	Gln	
	145					150					155					160	
	Glu	Ser	Val	Thr	Glu	Gln	Asp	Ser	Lys	Asp	Ser	Thr	Tyr	Ser	Leu	Ser	
					165					170					175		
	Ser	Thr	Leu	Thr	Leu	Ser	Lys	Ala	Asp	Tyr	Glu	Lys	His	Lys	Val	Tyr	
			180						185					190			
[0044]	Ala	Cys	Glu	Val	Thr	His	Gln	Gly	Leu	Ser	Ser	Pro	Val	Thr	Lys	Ser	
			195					200					205				
	Phe	Asn	Arg	Gly	Glu	Cys	Ser	Ser	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
		210					215					220					
	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
	225					230					235					240	
	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
				245						250					255		
	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
				260				265					270				
	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
			275					280				285					
	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
		290					295					300					
	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
	305					310					315					320	
	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	Ala	Ala	Ala	Pro	
				325				330							335		

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro
340 345 350

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro
355 360 365

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro
370 375 380

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro
385 390 395 400

Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro Ala Ala Ala Pro
405 410 415

Ala

<210> 58
<211> 5210
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: pFab-P1A1(200)的核酸序列"

[0045]

<400> 58
accgcacacc atcgaatggc cagatgatta attcctaatt ttgttgaca ctctatcatt 60
gatagagtta tttaaccact ccctatcagt gatagagaaa agtgaatga atagttcgac 120
aaaaatctag ataacgaggg caaaaaatga aaaagacagc tatcgcgatt gaagtggaac 180
tggtcgtgtt cgtaccgta gcgcaggccg aagttaaact gcaggaatcc ggtggtggtc 240
tggttcagcc aggtggttcc ctgcgctct cgtgtgctgc ttccggttcc aacatcaaag 300
acacctacat ccactgggtt cgtcaggctc cgggtaaagg cctggaatgg gttgctcgta 360
tetaccgac caacggttac accaggtatg cagattcagt taaaggctgt ttaccatct 420
cgcccgacac ttccaaaaac accgcttacc tccagatgaa ctccctgcgt gcgaagaca 480
cagctgttta ttattgctcc cgttgggggtg gtagcgggtt ctacgctatg gaactactgg 540
gtcagggtac cctggtcacc gtctctcag cctccacca gggcccatcg gtcttcccc 600
tggcacccct ctccaagagc acctctgggg gcacagcggc cctgggctgc ctggtcaagg 660
actacttccc cgaaccgggt acggtgtcgt ggaactcagg cgcctgacc agcggcgtgc 720
acacttccc ggetgtctta cagtcctcag gactctactc cctcagcagc gttgtgactg 780
tgccctccag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca 840
acaccaaggt cgacaagaaa gttgagccca aatcttgcca taccacat caccattaat 900
aaccatggag aaaataaagt gaaacaaagc actattgcac tggcactctt accgttactg 960
tttaccctg tgacaaagc cgacatcgag ctaccaccaat cccgctctc ctgtccgt 1020
tcgttggcg accgtgttac catcacgtgt agggcctcgc aagaagtaaa caccgcgta 1080

	gcgtggatc agcagaaacc cgggaaagct ccgaaactgc tgatctatag cgttccttc	1140
	ctgtattccg gagttccgag caggttcagt gggtcccggt ccggtaccga ettcaccctg	1200
	acgatatact cctccagcc ggaagacitc gctacctact actgtcaaca gcactacacc	1260
	accccgcga ccttcgggtc gggtaacaaa ctcgagatca aacggactgt ggctgcacca	1320
	tctgtcttca tcttcctgcc atctgatgag cagttgaaat ctggaaactgc ctctgttgtg	1380
	tgcctgtga ataacttcta tcccagagag gccaaagtac agtgaaggt ggataacgcc	1440
	ctccaatcgg gtaactccca ggagagigt acagagcagg acagcaagga cagcacctac	1500
	agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacacaa agtctacgcc	1560
	tgcgaagtca cccatcaggg cctgagttcg cccgtcaca agagcttcaa ccgcggagag	1620
	tgtctttctg cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	1680
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	1740
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	1800
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	1860
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	1920
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	1980
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	2040
[0046]	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	2100
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	2160
	cctgcaccag cccctgctc tctccagca cctgcaccag cacctgctc agcaccagct	2220
	cctgcaccag cctgaagagc ttaagcttga ccgtgaagt gaaaaaiggc gcacatigig	2280
	cgacattttt ttgtctgcc gtttaccgt actgcgtcac ggatctccac gcgcctgta	2340
	gcggcgcatt aagcgcggcg ggtgtggtgg ttacgcgcag cgtgaccgt acacttgcga	2400
	gcgcctagc gccgctctt ttcgtttct tcccttcct tctgcacg ttcgcggct	2460
	tccccgtca agctctaat cgggggctc cttaggggt ccgatttagt gctttacgce	2520
	acctgaccc caaaaaactt gattagggtg atggttcacg tagtgggcca tgcctgat	2580
	agacgtttt tgccttttg acgttggagt ccacgttctt taatagtga ctctgttcc	2640
	aaactggaac aacctcaac cctatctcg tctattctt tgattataa gggattttgc	2700
	cgatttcgce ctatttgta aaaaatgagc tgatttaaca aaaatttaac gcgaatttta	2760
	acaaaatatt aacgtttaca atttcaggtg gcacttttcg gggaaatgtg cgcggaacce	2820
	ctatttggtt attttctaa atacattcaa ataigtatcc gctcatgaga caataacct	2880
	gataaatgt tcaataatat tgaanaagga agagtatgag tatteaacat ttccgtgtcg	2940
	cccttatcc cttttttgcg gcattttgc ttcctgttt tctcaccga gaaacgtcg	3000
	tgaagtaaa agatgtgaa gatcagttgg gtgcacgagt gggttacac gaactggate	3060
	tcaacagcgg taagatctt gagagtttc gccccaaga acgttttcca atgatgagca	3120

[0047]

```

cttttaaagt tctgetatgt ggcgcggtat tatcccgat tgacgcggg caagagcaac 3180
tgggtgcgcg catacaatat tctcagaatg acttgggttg gtactcaaca gtcacagaa 3240
agcatcttac ggaatggcat acagtaagag aattatgcag tgetgcata accatgagt 3300
ataacactgc ggecaactta ctctgacaa egatcggagg accgaaggag ctaacegett 3360
ttttgcacaa catgggggat catgtaact gccttgatcg ttgggaaccg gagetgaatg 3420
aagccatacc aaacgaagag cgtgacacca cgatgcctgt agcaatggca acaacgttc 3480
gcaaaattat aactggcgaa ctacttact tagcttcccg gcaacaattg atagactgga 3540
tggaggcgga taaagttgca ggaccacttc tgcctcggc ccttcggct ggctggtta 3600
ttgctgataa atctggagcc ggtgagcgtg gctctcgcg tatcatgca gcactgggc 3660
cagatggtaa gccctccgt atcgtagtta tctacagac ggggagtcag gcaactatgg 3720
atgaacgaaa tagacagatc gctgagatag gtgcctcaet gattaagcat tggtagaat 3780
taatgatgtc tcgttagat aaaagtaag tgattaacag cgcattagag ctgcttaatg 3840
aggctcggaat cgaagggtta acaaccgta aactcgcca gaagctaggt gtagagcagc 3900
ctacattgta ttggcatgta aaaaataagc ggccttgcg cgacgcctta gccattgaga 3960
tgtagatag gcaccatact cacttttgc ctttagaagg ggaaagctgg caagatttt 4020
tacgtaataa cgctaaaagt tttagatgtg ctttactaag tcctcgcat ggagcaaaag 4080
tacatttagg tacacggcct acagaaaaac agtatgaaac tctcgaaaat caattagcct 4140
ttttatgcca acaaggtttt tcactagaga atgcattata tgcactcagc gcagtgggc 4200
attttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaagg 4260
aaacacctac tacgatagt atgcggcat tattacgaa agctategaa ttatttgatc 4320
accaaggtgc agagccagcc ttctatttc gccttgaatt gatcatatgc ggattagaaa 4380
aacaacttaa atgigaaagt gggtctttaa agcagcataa cctttttcgc tgaiggtaac 4440
ttcactagtt taaaaggatc taggtgaaga tctttttga taatctcatg accaaaatcc 4500
cttaacgtga gtttctgtc caetgagcgt cagaccctg agaaaagatc aaaggatctt 4560
cttgagatcc tttttttcgc cgcgtaatct gctgcttgc aacaaaaaaa ccaccgtac 4620
cagcgggtgt ttgtttgcg gateaagagc taccactct ttttcgaag gtaactggt 4680
tcagcagagc gcagatacca aatactgtc ttctagtga gcgtagtta ggccaccact 4740
tcaagaactc tgtagcaccg cctacatacc tcgtctgct aatcctgtta ccagtggctg 4800
ctgccagtgg cgataagtcg tgccttaccg ggttgactc aagacgatag ttaccggata 4860
aggcgcagcg gtcgggtga acggggggtt cgtgcacaca gccagcttg gacggaacga 4920
cctacaccga actgagatac ctacagcgtg agctatgaga aagcgcacg ctteccgaag 4980
ggagaaagcg ggacaggtat ccgtaagcg gcagggtcgg aacaggagag cgcacgagg 5040
agcttcaggc gggaaacgct tggatcttt atagtctgt cgggtttcgc cactctgac 5100
ttgagcgtc atttttga tgcctcagc gggggcggag cctatgaaa aacgccagca 5160
acgggcctt tttacggtc ctggccttt gctggcctt tgcacatg 5210

```

[0048]

```

<210> 59
<211> 5210
<212> DNA
<213> 人工序列

<220>
<221> 来源
<223> 注释="人工序列的描述: pFab-P1A3(200)的核酸序列"

<400> 59
accgcacacc ategaatggc cagatgatta attcctaatt ttigtigaca ctctateatt 60
gatagagtta ttttaccact cccatcagtg gatagagaaa agtgaaatga atagtgcgac 120
aaaaatctag ataacgaggg caaaaaatga aaaaagacgc tatecgcgatt gcagtgccac 180
tggttggttt cgtatccgta gcgcagggcg aagttaaact gcaggaatcc ggtgtgtgtc 240
tggttcagcc aggttggttc ctgcggctct cgtgtgtgtc ttccggtttc aacatcaaag 300
acacctacat ccactgggtt cgtcaggttc cgggtaaagg cctggaatgg gttgtctgta 360
tctaccgac caacggttac accagtaig ccgattcagt taaaggtctt ttcacctct 420
cggcgcacac ttccaaaac accgcttacc tccagatgaa ctccctgcgt gctgaagaca 480
cagctgttta ttattgtctc cgttggggtg gtgacggttt ctacgctatg gactactggg 540
gtcagggtac cctgttcacc gtctctcag cctccaccaa ggcccctcg gtcttcccc 600
tggcaccctc ctccaagagc acctctgggg gcacagcgge cctgggctgc ctggtcaagg 660
actacttccc cgaacgggtg acggtgtctt ggaactcagg cgccctgacc agcggcgtgc 720
acaccttccc ggetgttcta cagtctcag gactctactc cctcagcagc gtgggtaactg 780
tgccctcag cagcttgggc acccagacct acatctgcaa cgttaatcac aaaccagca 840
acaccaaggt cgacaagaaa gttgagccca aalettgcca tcaaccacct caccattant 900
aaccatggag aaataaagt gaaacaaagc actattgac tggcactctt accgttactg 960
tttacccttg tgacaaaagc cgacatcgag ctcaaccaat ccccgctctc cctgtccgt 1020
tcggttggeg accgtgttac cctcagctgt agggcctcgc aagacgtaaa caccgcgta 1080
gcgttggtatc agcagaaccc cgggaaagct ccgaaactgc tgatctatag cgttctctc 1140
ctgtattccg gagttccgag caggttcagt ggttccggtt ccggtaccca cttaccctg 1200
acgatatctt cctccagcc ggaagacttc gctacctact actgtcaaca gcactacacc 1260
accccgccga ccttgggtca gggtaccaaa ctgagatca aacggactgt ggtgcacca 1320
tctgtcttea tcttcgcgc atctgatgag cagttgaaat ctggaactgc ctctgttgtg 1380
tgctgtctga ataacttcta tcccagagag gccaaagtac agtggaaggt ggataacgcc 1440
ctccaatcgg gtaactccca ggagagtgtc acagagcagg acagcaagga cagcacctac 1500
agcctcagca gcacctgac gctgagcaaa gcagactacg agaaacacaa agtctacgcc 1560
tgcaagtc aacctcagg cctgagttcg cccgtcaca agagcttcaa ccgcggagag 1620
tgctctcttg ccgtgcacc tgtgcagca cctgtgcag ctccagcagc tgctctgca 1680
gcagctccag ccgtgcacc tgtgcagca cctgtgcag ctccagcagc tgctctgca 1740

```

[0049]	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	1800
	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	1860
	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	1920
	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	1980
	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	2040
	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	2100
	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	2160
	gcagctccag ccgctgcacc tgetgcagca cctgetgcag ctccagcage tgcctctgca	2220
	gcagctccag cctgaagagc ttaagcttga cctgtgaagt gaaaaatggc gcacattgtg	2280
	cgacattttt tttgtctgcc gtttaacct aetgcgtcac ggatctccac gegccctgta	2340
	ggggcgcat aagcgggcg ggtgtgggtg ttacgcgcag cgtgacctgt acatttgcca	2400
	gegccctage gcccgctcct ttccctttct tcccttcctt tctgcacag ttcccggtct	2460
	ttcccgctca agctctaaat cgggggctcc ctttaggggt cggatttagt gctttacgce	2520
	acctcgacct caaaaaactt gatttaggtg aiggttcac tagtgggcca tgcctctgat	2580
	agacggtttt tgcctcttg acgttggagt ccacgttctt taatagtga ctctgttcc	2640
	aaactggaac aacactcaac cctatctcgg tctattcttt tgatttataa gggattttgc	2700
	cgttttcggc ctatttggtt aaaaatgagc tgatttaaca aaaatttaac gcgaatttta	2760
	acaaaaatatt aacgtttaca atttcaggtg gcacttttcg gggaaatgtg cgcggaacct	2820
	ctatttggtt attttctaa atacattcaa atatgtatcc gctcatgaga caataacctt	2880
	gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaaat ttccgtgtcg	2940
	ccctatttcc cttttttgcg gcattttgcc ttctgtttt tgcacacca gaaacgttg	3000
	tgaangtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc	3060
	tcaacagcgg taagatcctt gagagtttcc gcccgagaag acgttttcca atgatgagca	3120
	cttttaangt tctgctatgt ggcgcggtat tatccgtat tgacgcggg caagagcaac	3180
	tgggtgcgg catacactat tctcagaatg acttggtiga gtactacca gtcacagaaa	3240
	agcatcttac ggatggcatg acagtaagag aattatgcag tgcgccata acctagagt	3300
	ataacactgc ggcaactta cttctgacaa cgtcggagg accgaaggag ctaacctgtt	3360
	ttttgcacaa catgggggat catgtaactc gccctgatcg ttgggaaccg gagctgaatg	3420
	aagccatacc aaacgacgag cgtgacacca cgtgcctgt agcaatggca acaacgttgc	3480
	gcaaaatatt aactggcgaa ctacttactc tagcttcccg gcaacaattg atagactgga	3540
	tggaggcgga taaagttgca ggaccacttc tgcgtcggc ctttcgggt ggtgtgttta	3600
	ttgtgataa atctggagcc ggtgagcgtg gctctcggg tatcattgca gcactgggce	3660
	cagatggtaa gccctccgt atcgtagtta tetacaagac ggggagtcag gcaactatgg	3720
	atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaggaat	3780

```

taatgatgtc tcgtttagat aaaagtaaag tgattaacag cgcattagag ctgcttaatg 3840
aggtcggaat cgaagggtta acaacccgta aactcgccca gaagctaggt gtagagcagc 3900
ctacattgta ttggcatgta aaaaataagc gggttttgcg cgacgcctta gccattigaga 3960
tgtttagatag gcaccatact cacttttgcg ctttagaagg ggaagctgg caagattttt 4020
tacgtaataa cgtataaagt tttagatgtg ctttactaag tcatecgcat ggagcaaaag 4080
tacatttagg tacacggcct acagaaaaac agtatgaac tctcgaaat caattagcct 4140
ttttatgcca acaagggttt tctactagaga atgcattata tgcactcage gcagtggggc 4200
attttacttt aggttgcgta ttggaagatc aagagcatca agtcgctaaa gaagaaaggg 4260
aaacacctac tacgatagt atgcgcccat tattacgaca agctatcgaa ttattigatc 4320
accaagggtc agagccagcc ttcttattcg gccctgaatt gatcataatc ggattagaaa 4380
aacaacttaa atgtgaaagt ggtctttaa agcagcataa cttttttccg tgatggtaac 4440
ttcactagtt taaaaggatc taggtgaaga tctttttga taatctcatg accaaaatcc 4500
cttaacgtga gttttcgtc cactgagcgt cagaccccg agaaaagatc aaaggatctt 4560
cttgagatcc tttttttctg cgcgtaatct gctgcttgca aacaaaaaa ccaccgctac 4620
cagcgggtgt ttgtttgcg gatcaagagc taccaactct tttccgaag gtaactggct 4680
tcagcagagc gcagatacca aatactgtcc ttctagtgtt gccgtagtta ggccaccact 4740
tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatctgttta ccagtggctg 4800
[0050] ctgccagttg cgataagtcg tgccttaccg ggttggactc aagacgatag ttaccggata 4860
agggcgagcg gtcgggtcga acgggggggt cgtgcacaca gcccagcttg gagcgaacga 4920
cctacaccga actgagatcc ctacagcgtg agctatgaga aagcgccaag ctcccggaag 4980
ggagaaaggc ggacaggtat ccggttaagc gcagggtcgg aacaggagag cgcacgaggg 5040
agcttccagg gggaaacgcc tggatatctt atagtcctgt cgggtttcgc caccctgac 5100
ttgagcgtcg atttttgtga tgcctgtcag gggggcggag cctatggaaa aacgccagca 5160
acggcgccct tttacggttc ctggcccttt gctggccctt tgcctacatg 5210

```

<210> 60

<211> 3680

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> 注释="人工序列的描述: pSUM0-PA#1(200)的核酸序列"

<400> 60

```

gcacttttcg gggaaatgtg cgcggaacc cttattgttt atttttctaa atacattcaa 60
atatgtatcc gctcatgaga caataaccct gataaatget tcaataatat tgaanaagga 120
agagtatgag tattcaacat ttccgtgtcg ccttatctcc ctttttttgcg gcattttgce 180
ttctgttttt tgcctaccca gaaacgctgg tgaaagtaaa agatgetgaa gatcagttgg 240
gtgcacgagt gggttacatc gaactggatc tcaacagegg taagatccct gagagttttc 300

```

[0051]

```

gccccgaaga acgitttcca atgaatgaca cttttaaagt tctgctatgt ggccgggtat 360
tatcccgat tgaagccggg caagagcanc tggctcgccg catacactat tctcagaatg 420
acttgggtga gtactcacca gtcacagaaa agcatcttac ggatggcatg acagtaagag 480
aattatgcag tgcigccata accatgagtg ataacactgc ggccaaetta cttctgacaa 540
cgatcgaggg accgaaggag ctacccgctt ttttgcacaa catgggggat catgtaacte 600
gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca 660
cgatgcctgt agcaatggca acaacgttgc gcaaaactatt aactggcgaa ctacttacte 720
tagcttcccg gcaacaatta atagactgga tggaggcgga taaagttgca ggaccacttc 780
tgcctcggcg ccttcgget ggctgggtta ttgtgataa atctggagcc ggtgagcggtg 840
ggctcgcggt tatcattgca gcactggggc cagatggtaa gccctcccgat atcgtagtta 900
tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag 960
gtgcctcact gattaaagcat tggtaactgt cagaccaagt ttactcatal atactttaga 1020
ttgattttaa acttcatttt taatttataa gcatctaggt gaagatcctt ttgataatc 1080
tcattgacaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac ccgtagaaa 1140
agatcaaagg atcttcttga gactctttt ttctgcgcgt aacttgcctc ttgcaacaa 1200
aaaaaccacc gtaaccagcg gtggtttgtt tgcgggatac agagctacca actcttttct 1260
cgaaggiaac tggcttcagc agagcgagc taccataaac tgccttctta gtgtagccgt 1320
agttaggcca caacttaag aactctgtag caccgcctac atacctcgt ctgctaatec 1380
tgtaaccagt ggtctctgcc agtggcgata agtcgtgtct taccgggttg gactcaagac 1440
gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca 1500
gcttggagcg aacgacctac accgaactga gatacttaca gcgtgagcta tgagaaagcg 1560
ccacgcttcc cgaagggaga aaggcgagca ggtatccggt aagcggcagg gtcggaacag 1620
gagagcgcac gaggagctt ccagggggaa acgcctggta tctttatagt cctgtcgggt 1680
ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc gtcagggggg cggagcctat 1740
ggaaaaacgc cagcaacgcg gcccttttac ggttctcggc cttttgctgg ccttttgcct 1800
acatgttctt tctgcgtta tccccgatt ctgtggataa ccgtattacc gcccttgagt 1860
gagctgatac cgtcgcgcgc agccgaacga ccgagcgag ccagtcagtg agcgaggag 1920
cggagaagcg cccaatacgc aaaccgcctc tccccgcgcg ttgcccatt cattaatgca 1980
ggatctcgat ccccggaat taatacgact cactataggg agaccacaa gcccttccctc 2040
tagaataat ttgttttac tttaagaagg agataatcat atgaacatc accaccatca 2100
ccattcggac tcagaagta atcaagaagc taagccagag gtcaagccag aagtaagcc 2160
tgagactcac atcaatttaa aggtgtcga tggatcttca gaaatcttct ttaagatcaa 2220
aaagaccact ctttaagaa ggtgatgga agcgttcgct aaaagacagg gtaaggaaat 2280
ggactcctta agattctgt acgacggtat tagaattcaa gctgacaga cccctgaaga 2340
tttgacatg gaggataacg atattattga ggctcacaga gaacagattg gtggcgccgc 2400

```

[0052]

```

tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2460
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2520
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2580
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2640
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2700
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2760
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2820
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2880
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 2940
tccagctgca cctgtccag cagcaccctgc tgcaccagct ccggctgctc ctgttgcgc 3000
aagagcaagc ttgatccgc tgcatacaaa gcccgaaagg aagctgagtt ggcctgtgc 3060
accctgagc aataactagc ataaccctt ggggcctcta aacgggtctt gaggggttt 3120
ttgtgaang gaggaactat atccggatct ggcgtaatag cgaagaggcc cgcaccgatc 3180
gcccttccca acagttgcgc agcctgaatg gcgaatggg cgcgcctgt agcgggcgat 3240
taagcgcgc ggggtgtgtg gttacgcgc gcgtgaccgc tacacttgc agcgcctag 3300
cgccgcctcc ttctcttct ttcccttct ttctgcac gttgcgcgc ttccctctc 3360
aagctctaaa tgggggctc ccttaggtt tccgatttag tgccttacgg caccctgacc 3420
ccaaaaaact tgattagggt gatggttcac gtagtgggc atcgccctga tagacggtt 3480
ttcgcccttt gacgttgag tccagttct ttaatagtgg actcttctc caaactggaa 3540
caacacfcac cctatctcg gctattctt ttgatttata agggattttg ccgatttcgg 3600
cctatttggt aaaaaatgag ctgatitac aaaaaattaa cgcgaatttt aacaaaatat 3660
taacgcttac aatttaggtg 3680

```

<210> 61
 <211> 201
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> 注释-“人工序列的描述: 用于制备药物缀合物的PA#1(200)多肽/聚合物(其实通过将10个20聚体的编码基因盒连接起来获得的, 并包括来自下游连接位点的一个额外的C末端Ala残基)”

<400> 61

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
1 5 10 15

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
20 25 30

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro
35 40 45

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
50 55 60

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
65 70 75 80

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
85 90 95

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
100 105 110

Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro
115 120 125

[0053]

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Ala Pro Ala
130 135 140

Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
145 150 155 160

Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Pro
165 170 175

Ala Ala Pro Ala Ala Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala
180 185 190

Ala Pro Ala Pro Ala Ala Pro Ala Ala
195 200

```
gccgctccagctgcacctgctccagcagcacctgctgcaccagctccggctgctcctgct  
|||||  
CGAGGTCGACGTGGACGAGGTCGTCGTGGACGACGTGTCGAGGCCGACGAGGACGAcg  
AlaAlaProAlaAlaProAlaAlaProAlaAlaProAlaProAlaProAlaProAla
```

图1

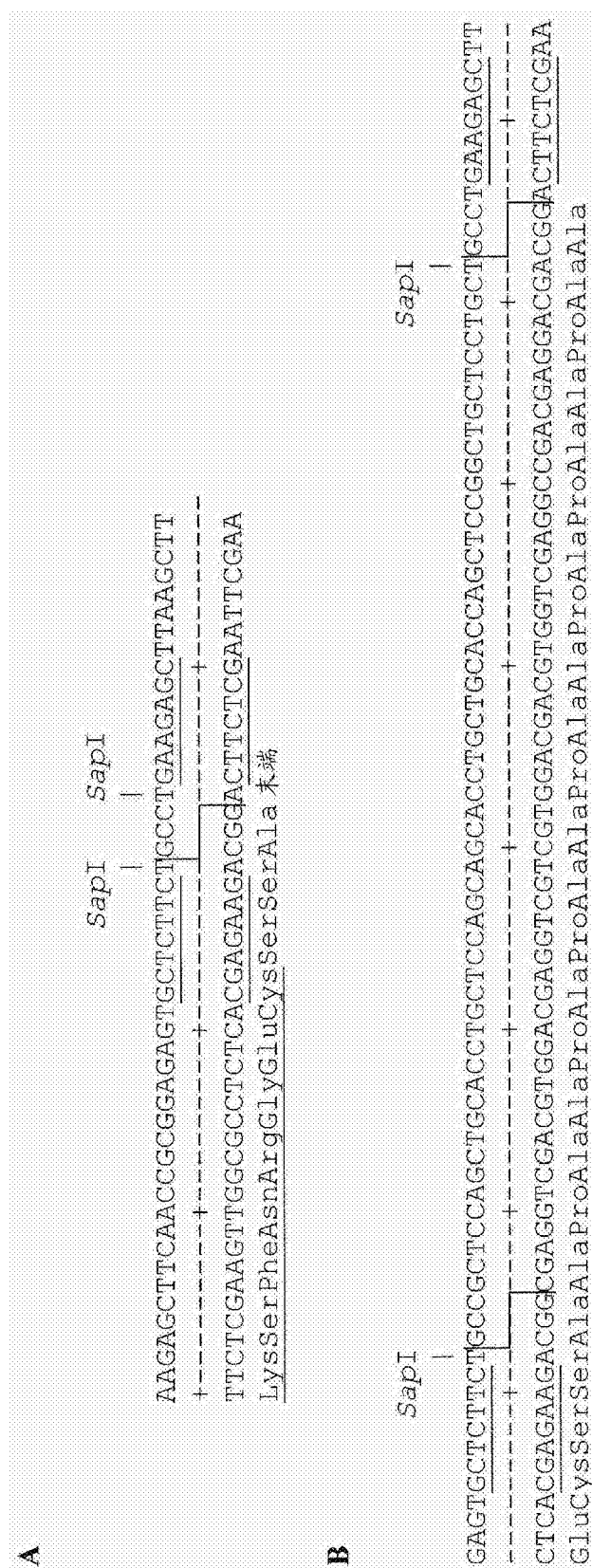


图 2

C

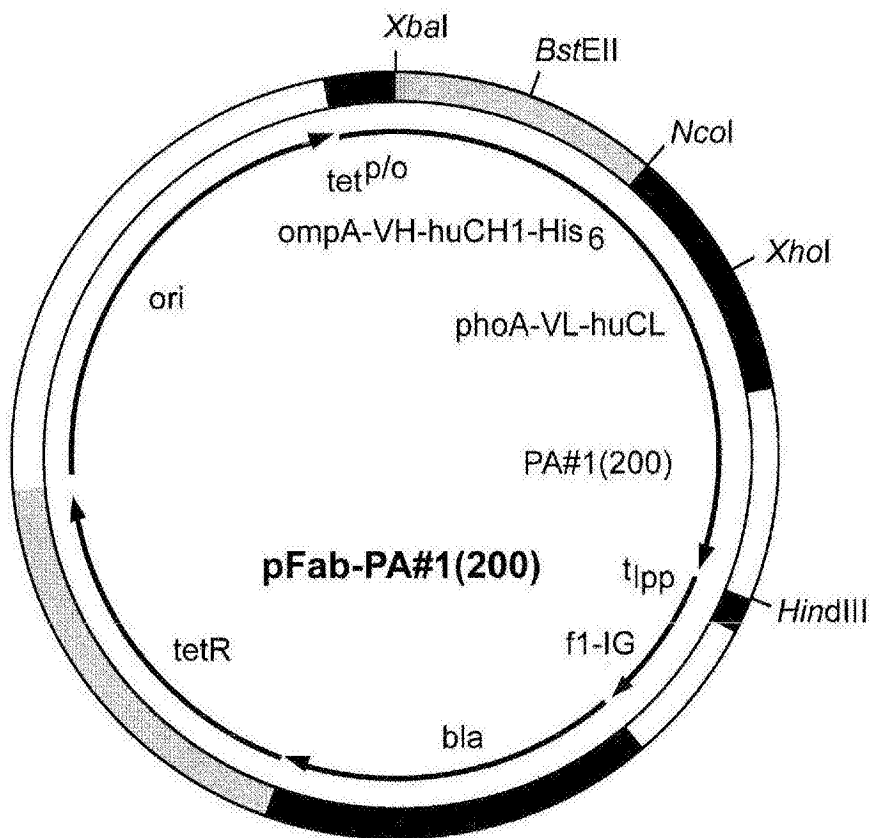


图2续

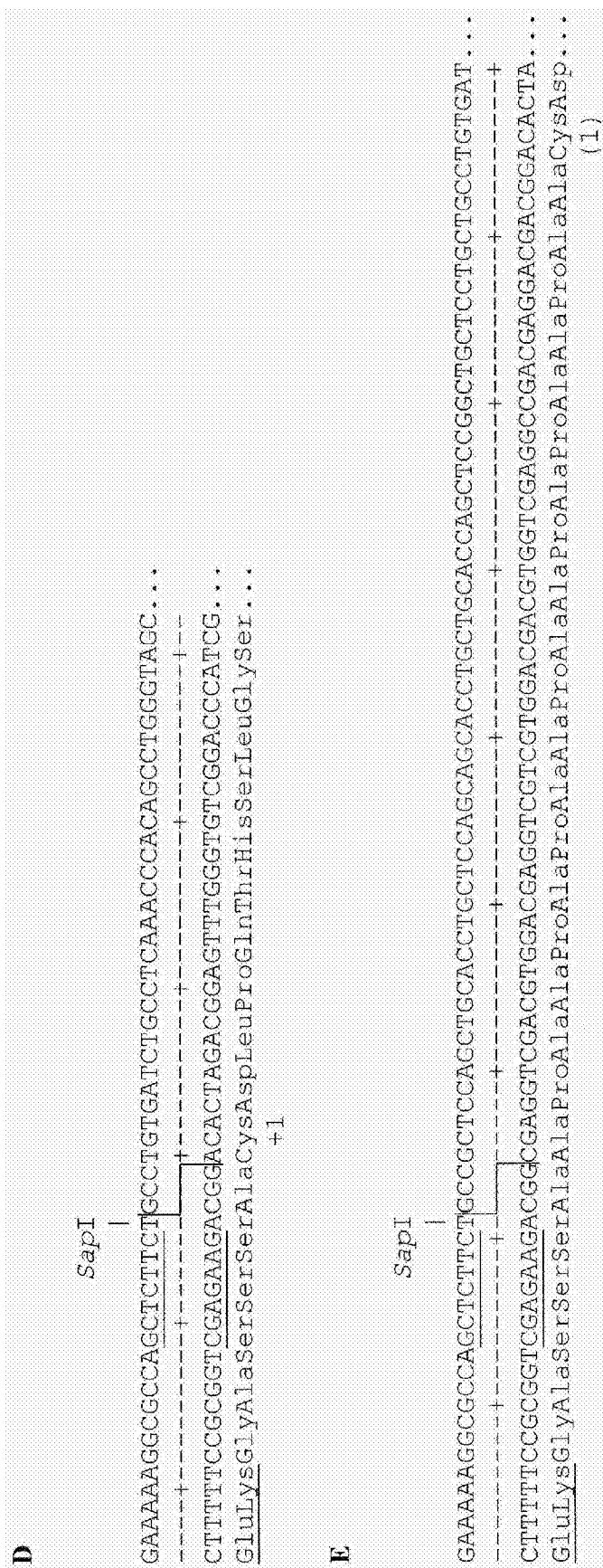


图2续

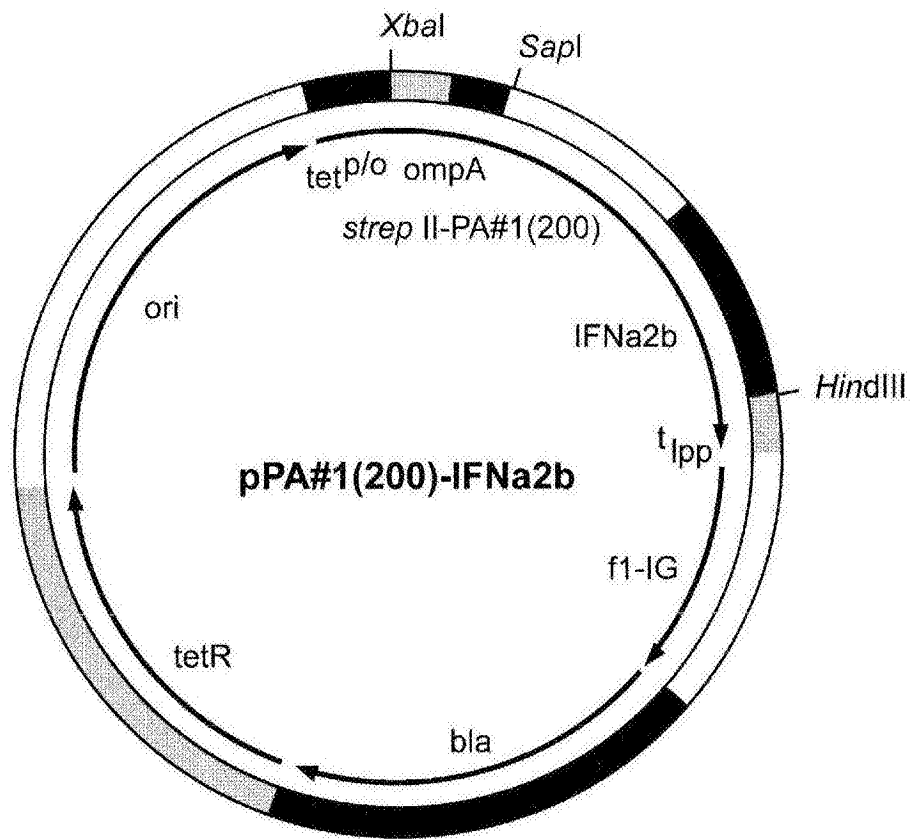
F

图2续

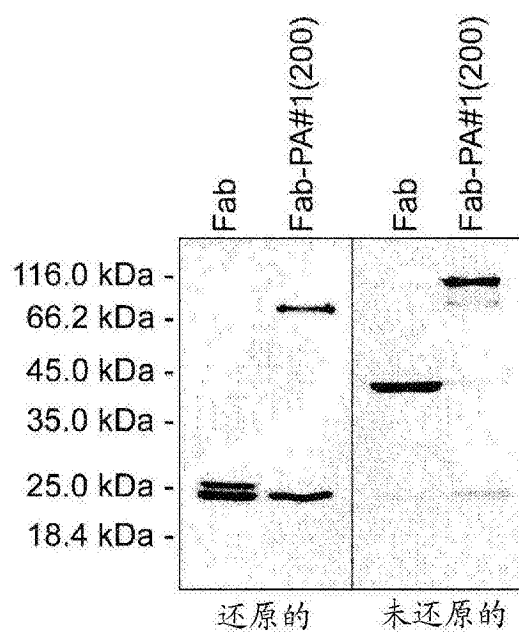
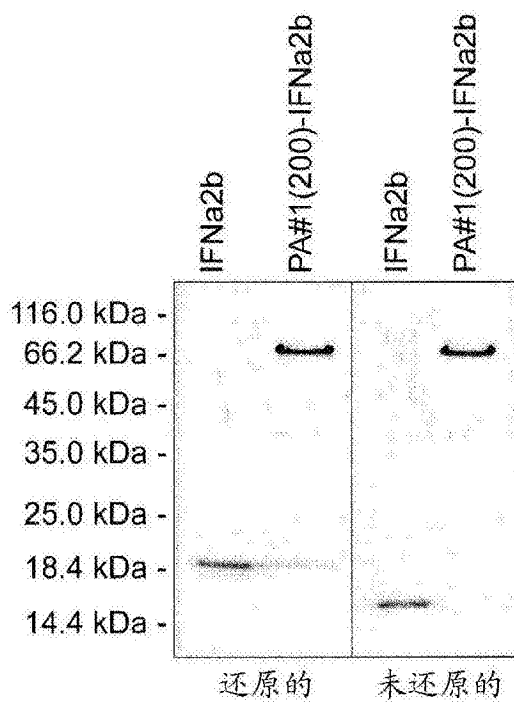
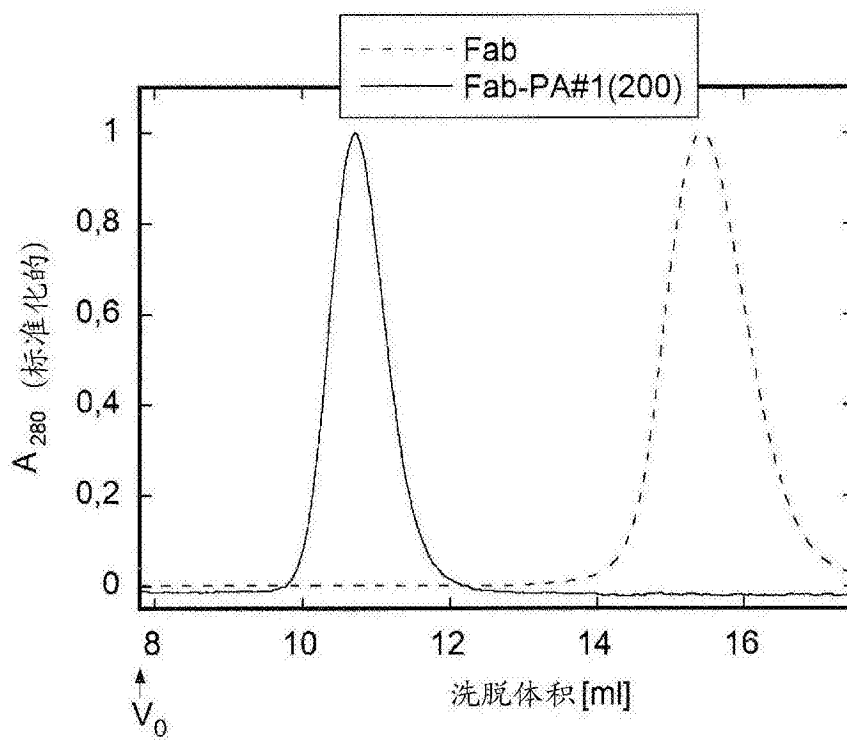
A**B**

图3

A



B

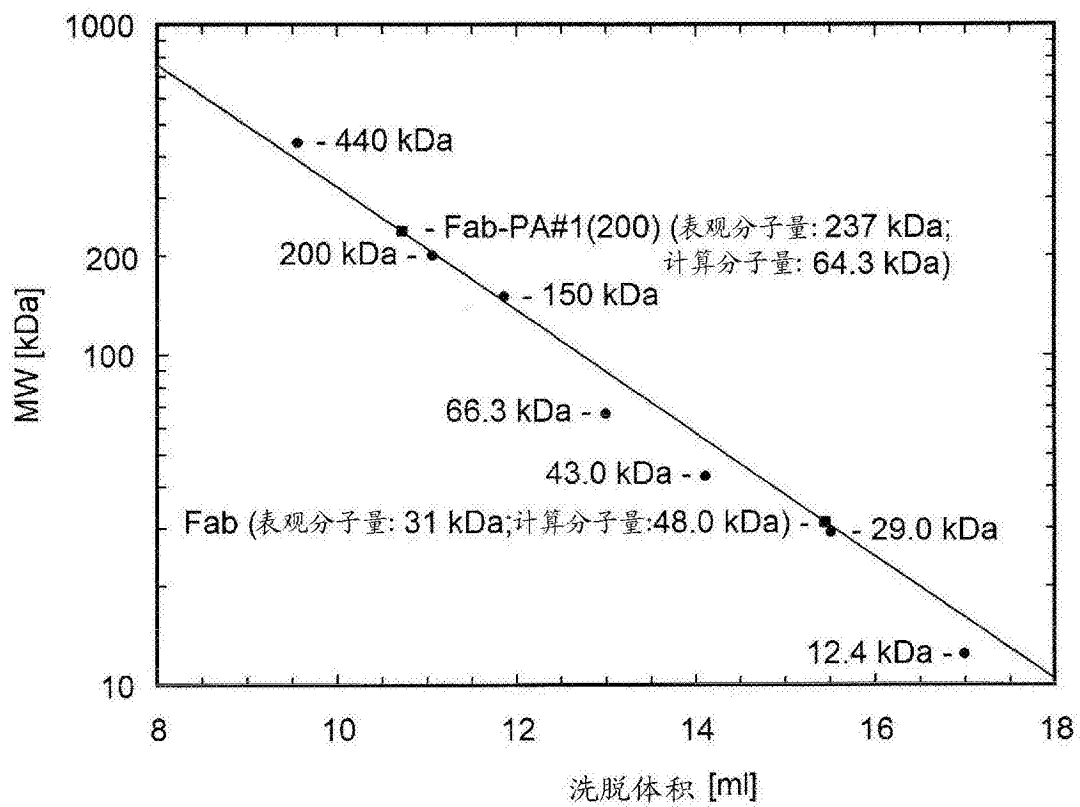
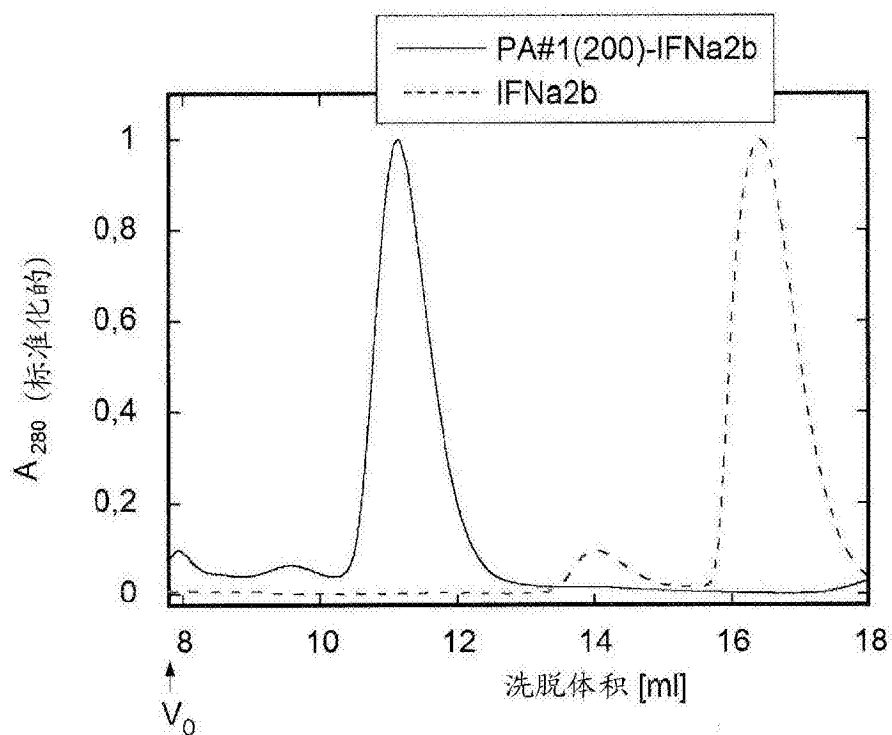


图4

C



D

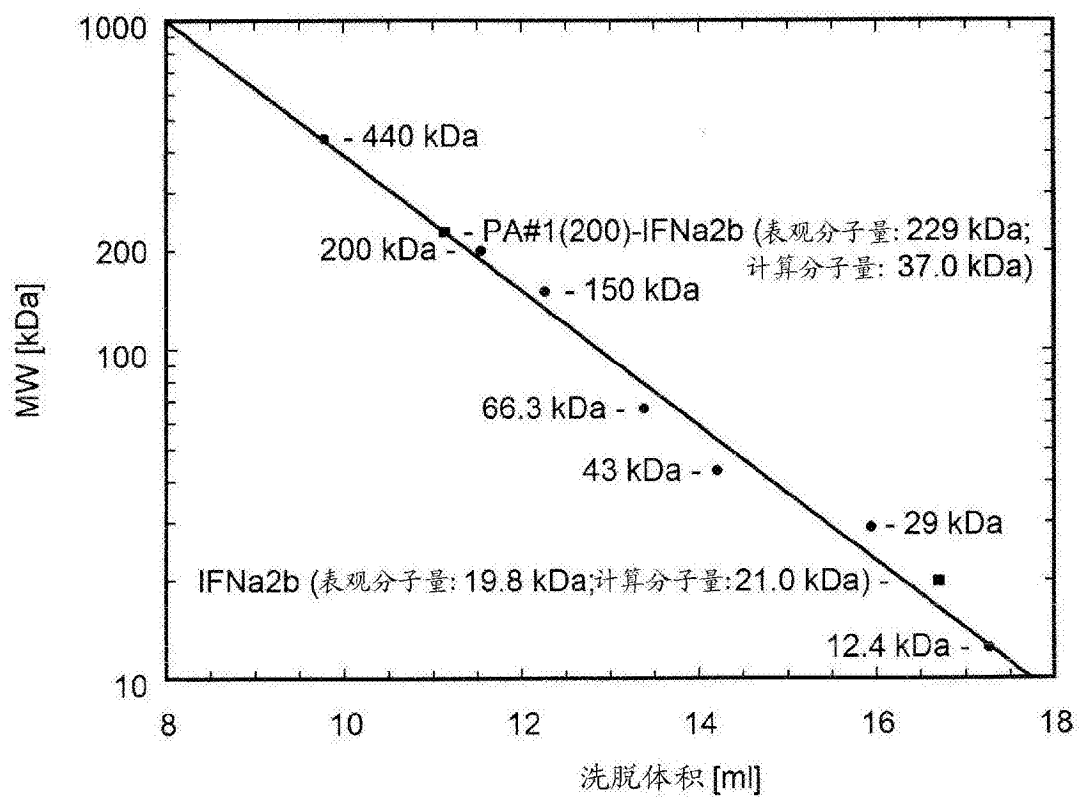


图4续

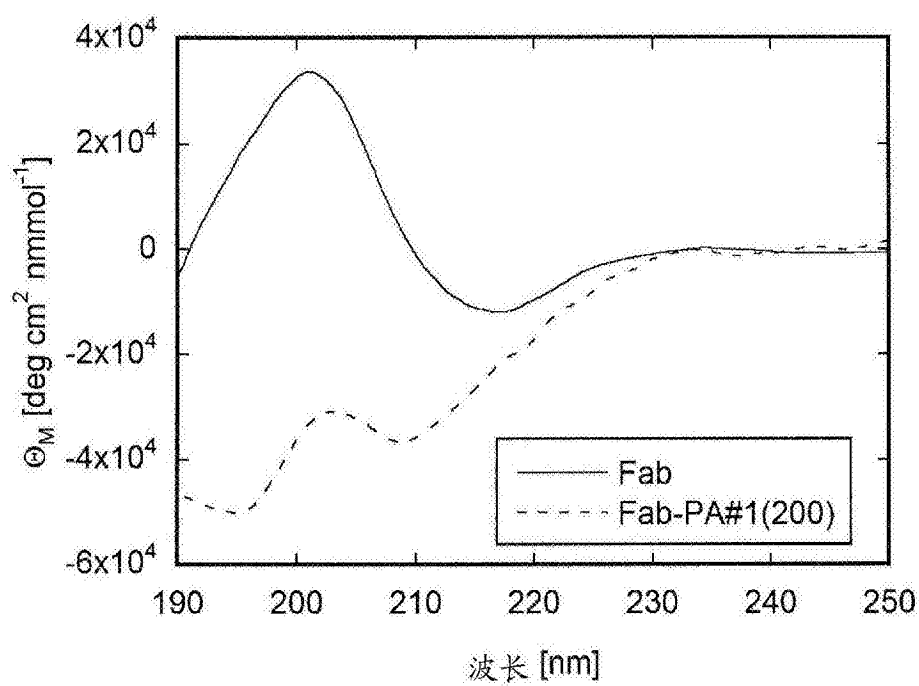
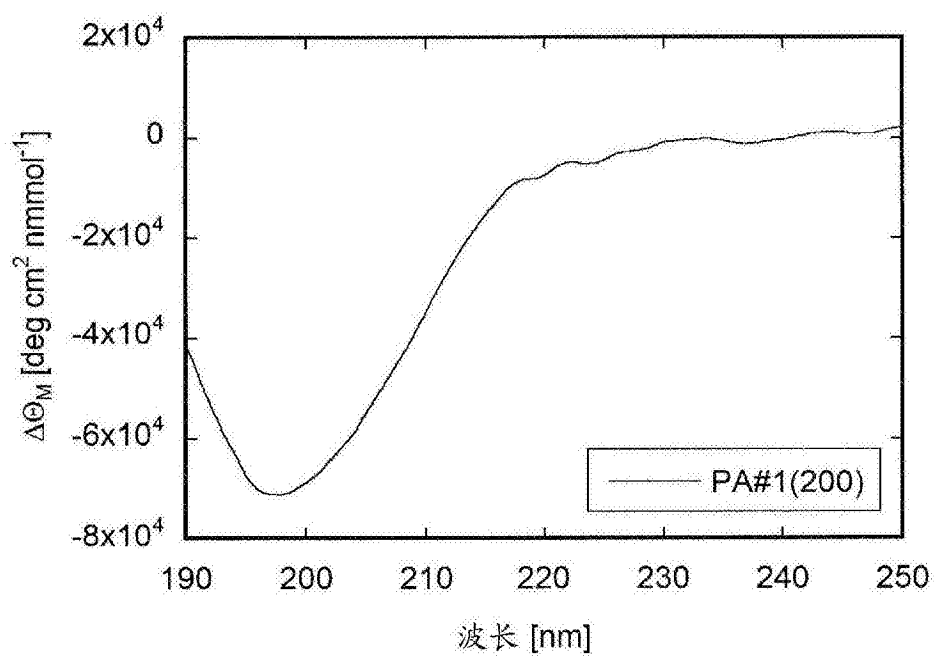
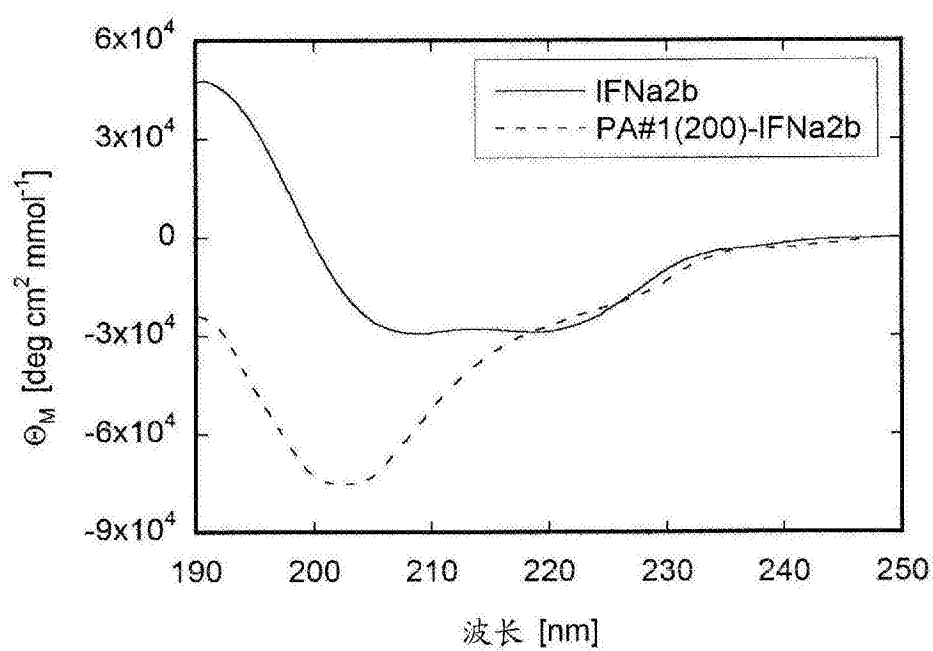
A**B**

图5

C



D

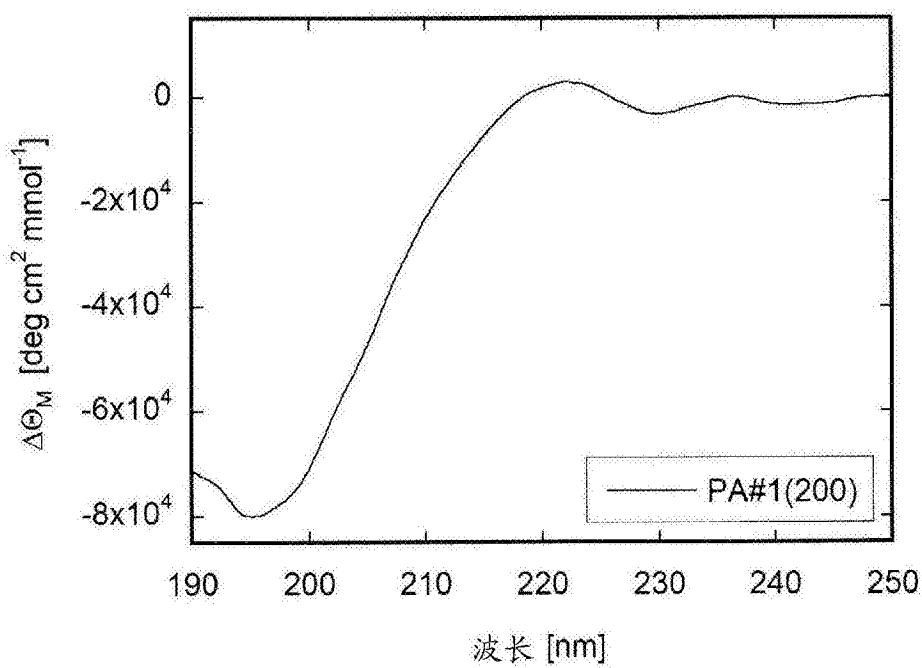


图5续

A

[illegible]

图6

B

<u>NheI</u>		<u>SapI</u>
GC	CGCTAGCCATCACCACCATCACCATGGCGCCAGCTCTCTG	CCGCTCCAGCTGCACCTGCTCCAGCAGCACCTGCTGCACCACTCCG
-----+	-----+	-----+
CG	CGCATCGGTAGTGGTACCGCGGTTCGAGAACGCGCGAGGTCGACGTGGACGAGGTCGTCGTTGGACGACGTGGTCGAGGC	CGTCCAGTGGACGAGGTCGTCGTTGGACGACGTGGTCGAGGC
Ala	AlaSerHisHisHisHisGlyAlaSerSerAlaAlaProAlaAlaProAlaAlaProAlaProAlaPro	AlaAlaProAlaProAlaProAlaProAlaProAlaProAlaProAlaProAlaPro
-----+	-----+	-----+
GCTGCTCCTGCTGCCTTCCCAACC...		
-----+	-----+	-----+
CGACGAGGACGACGGAAGGTTGG...		
AlaAlaProAlaAlaPheProThr...		
(1)		

图6续

C

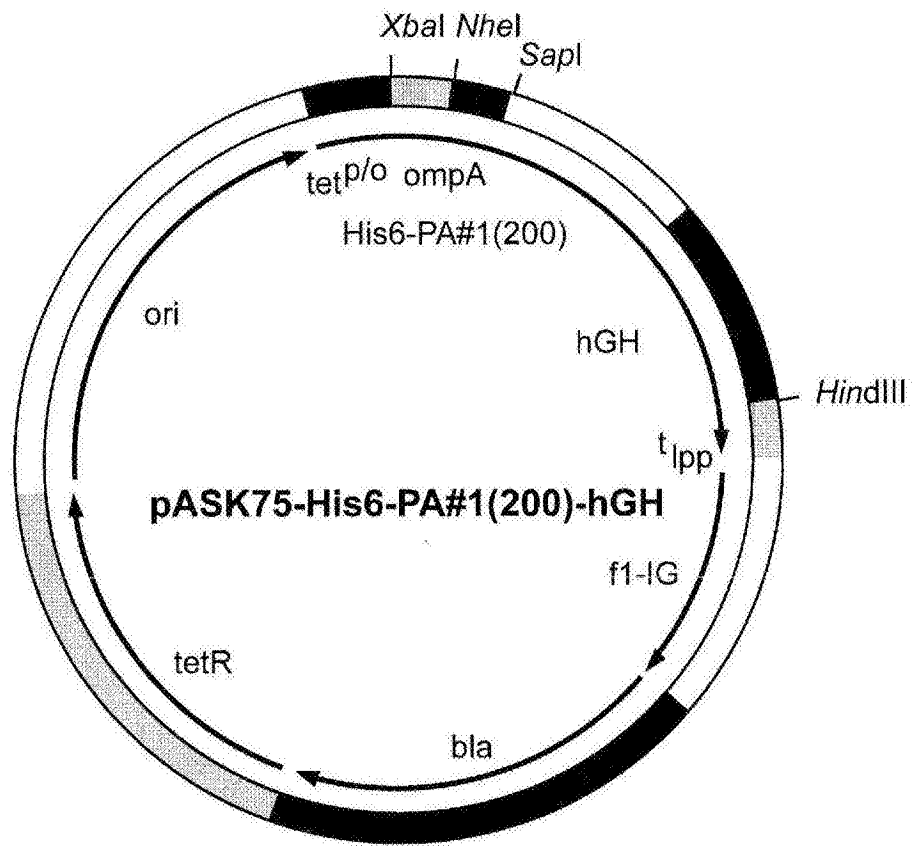


图6续

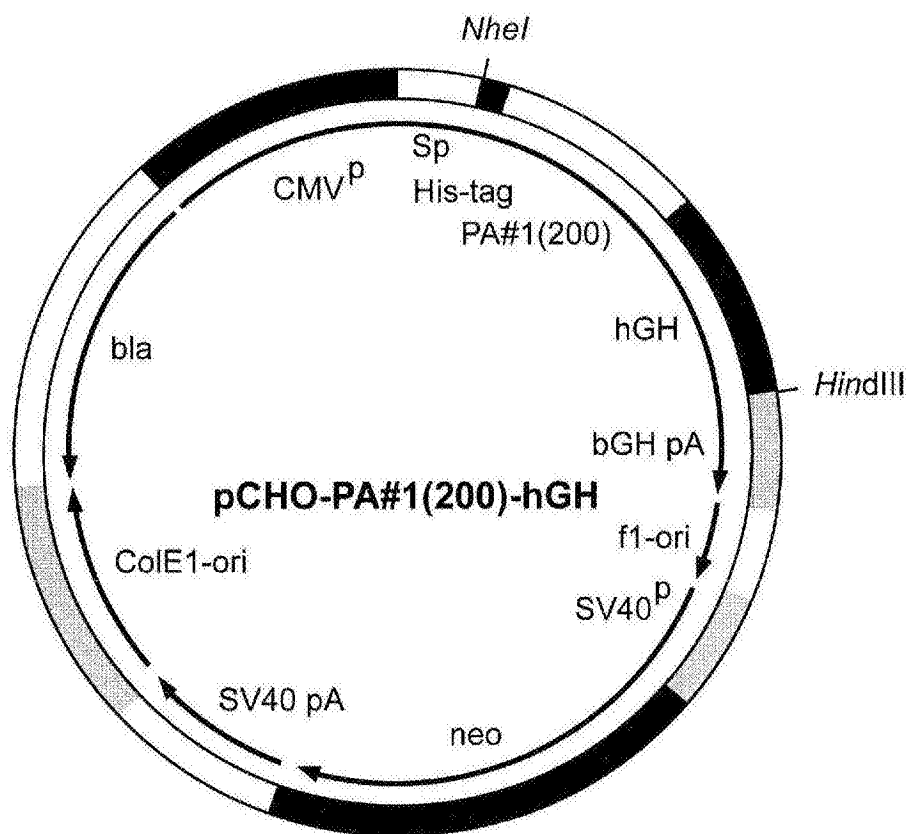
D

图6续

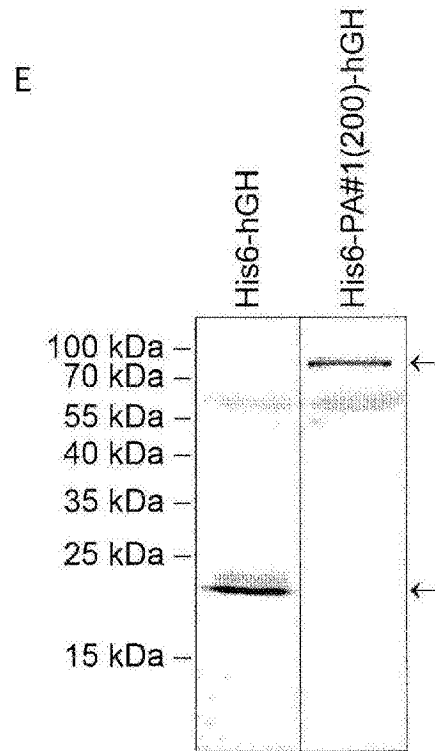


图6续

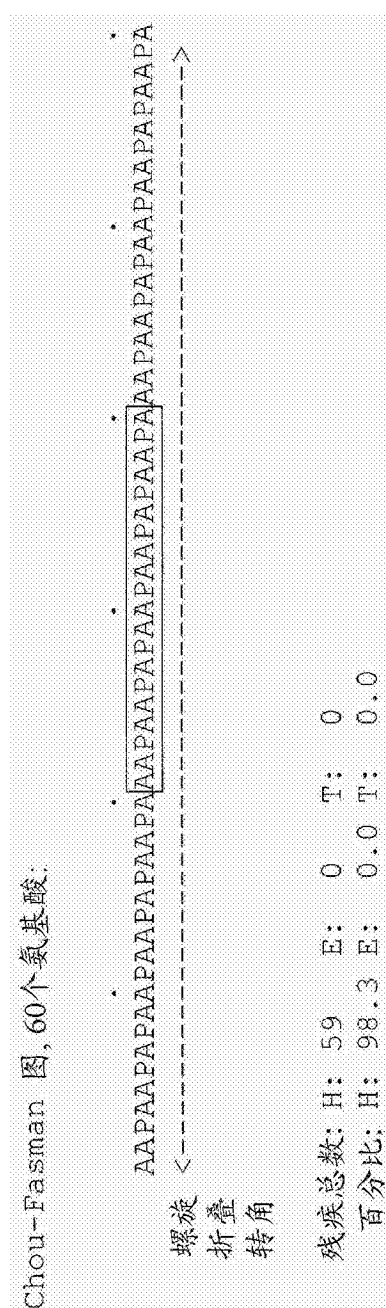


图7

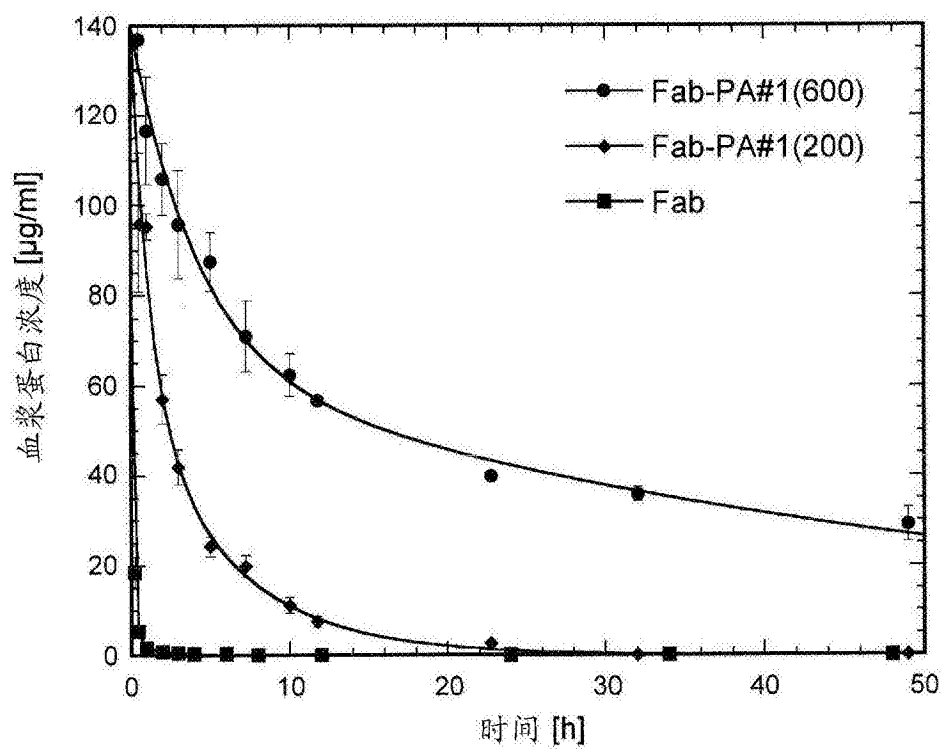


图8

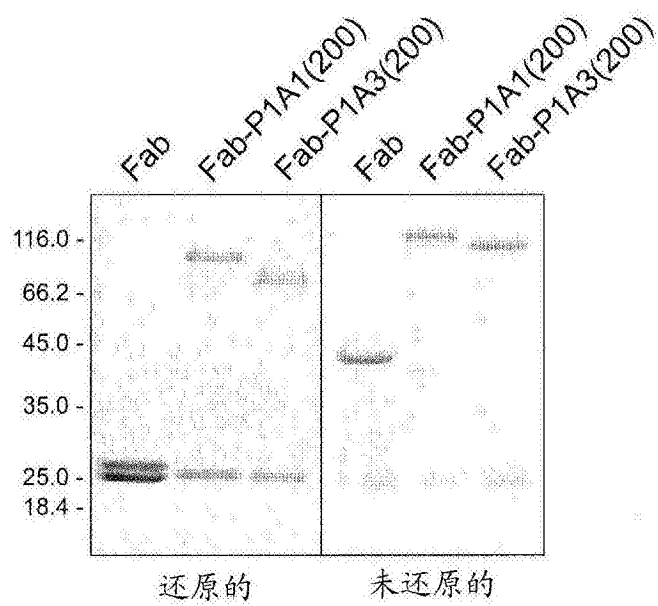


图9

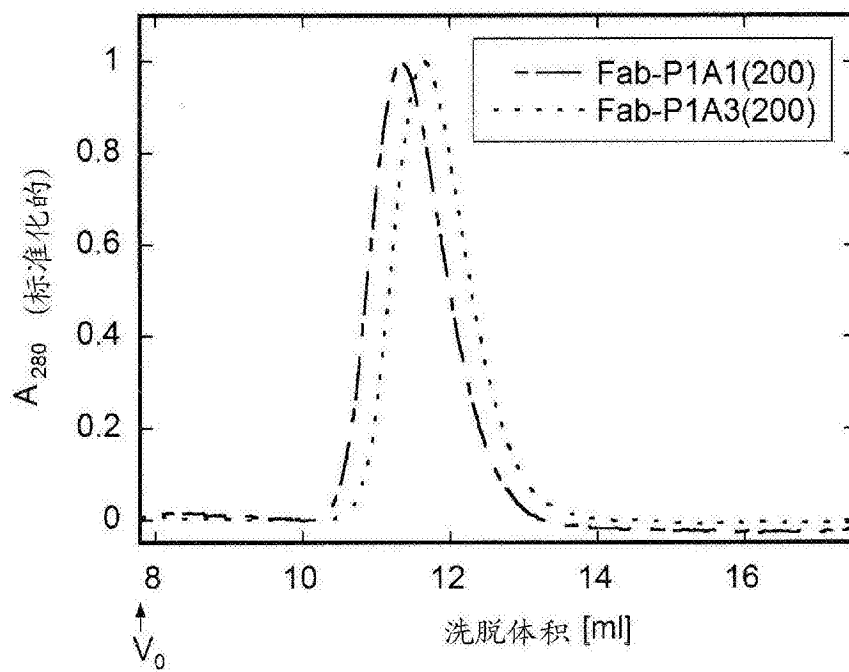


图10

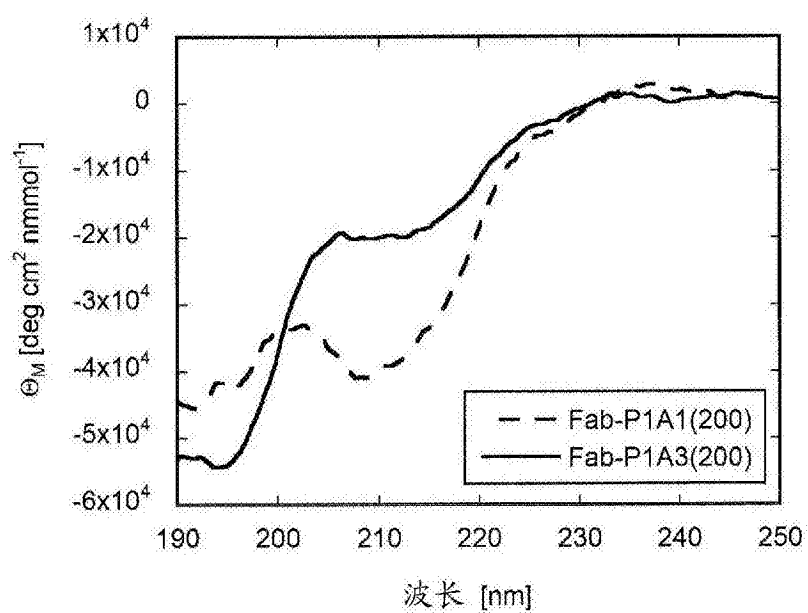
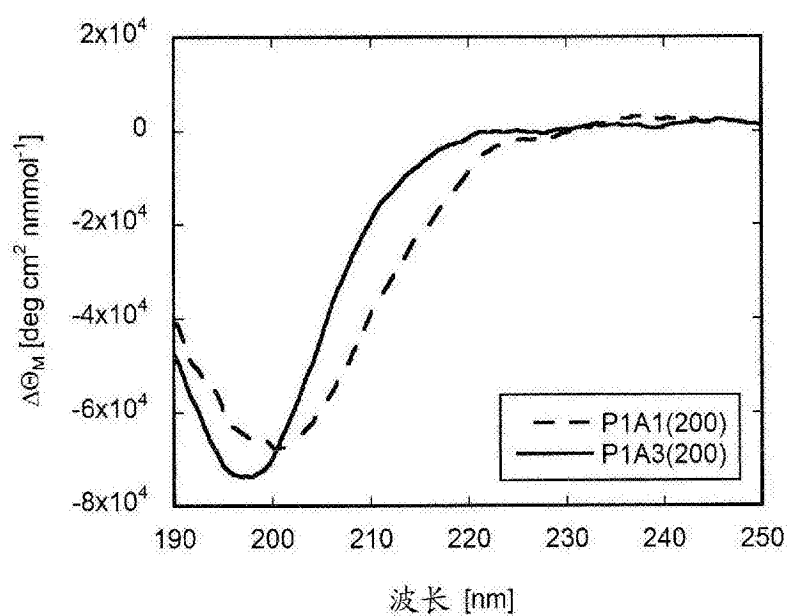
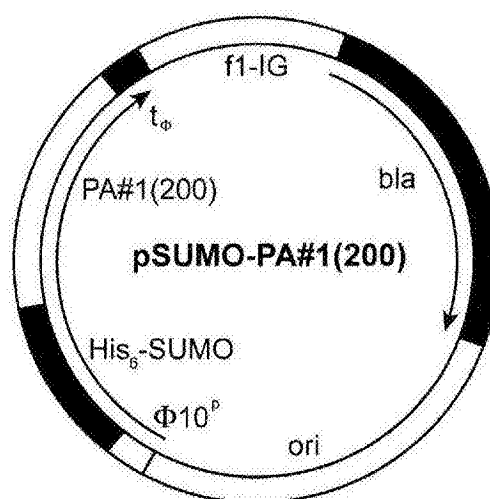
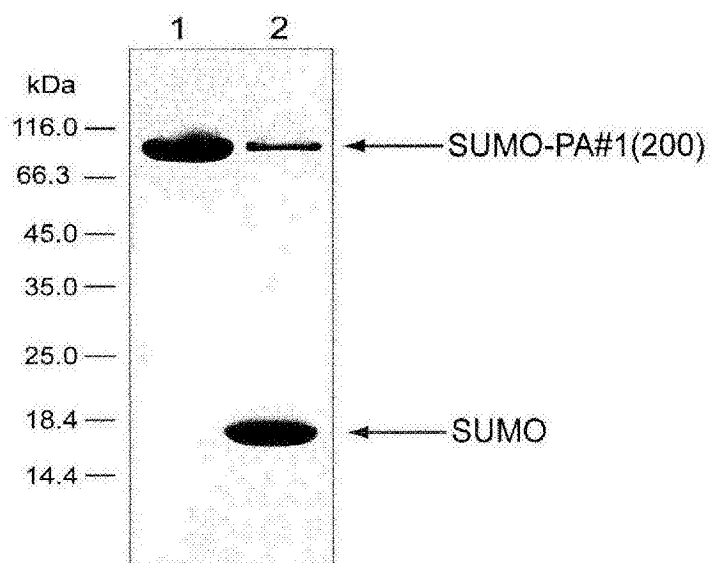
A**B**

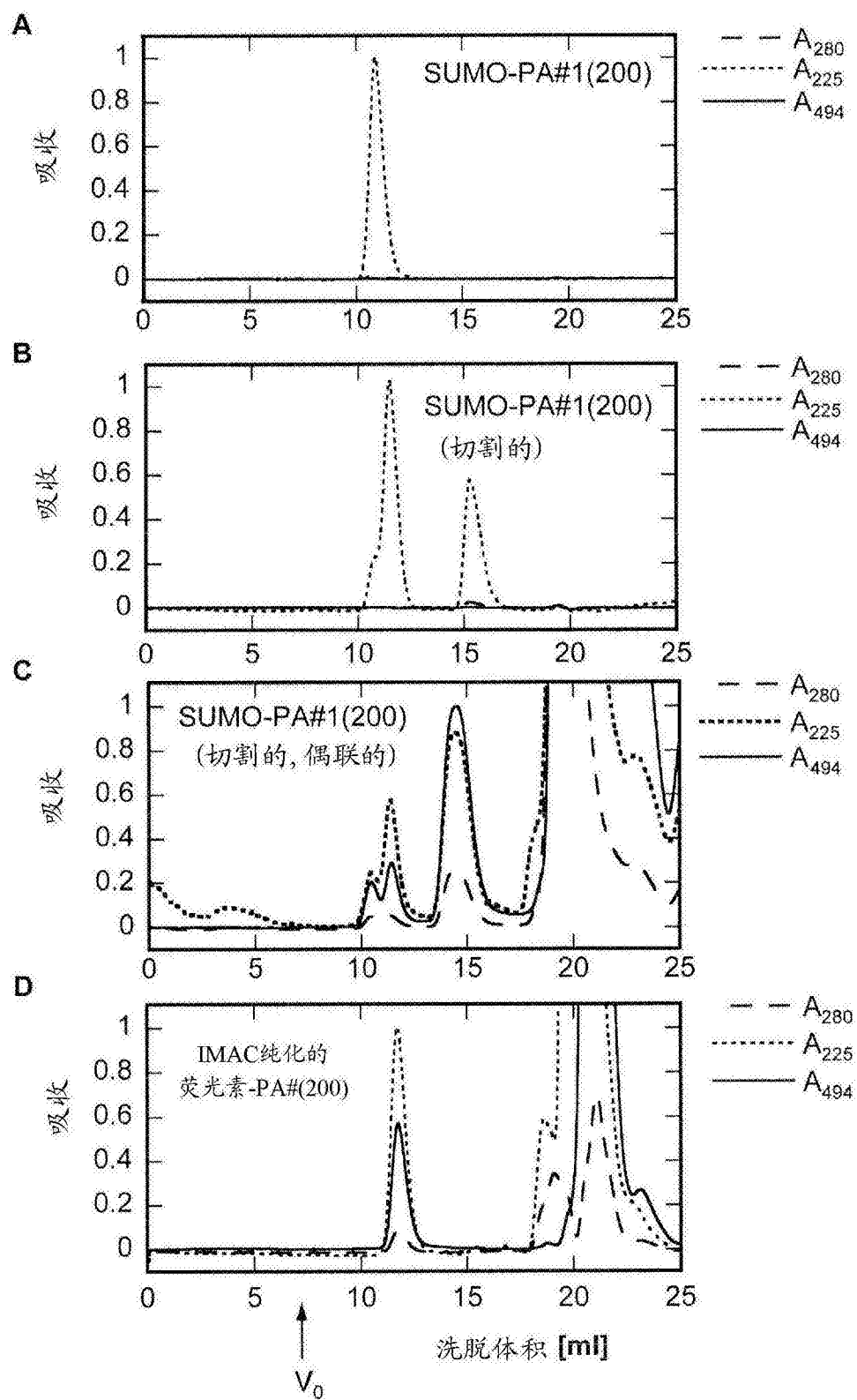
图11

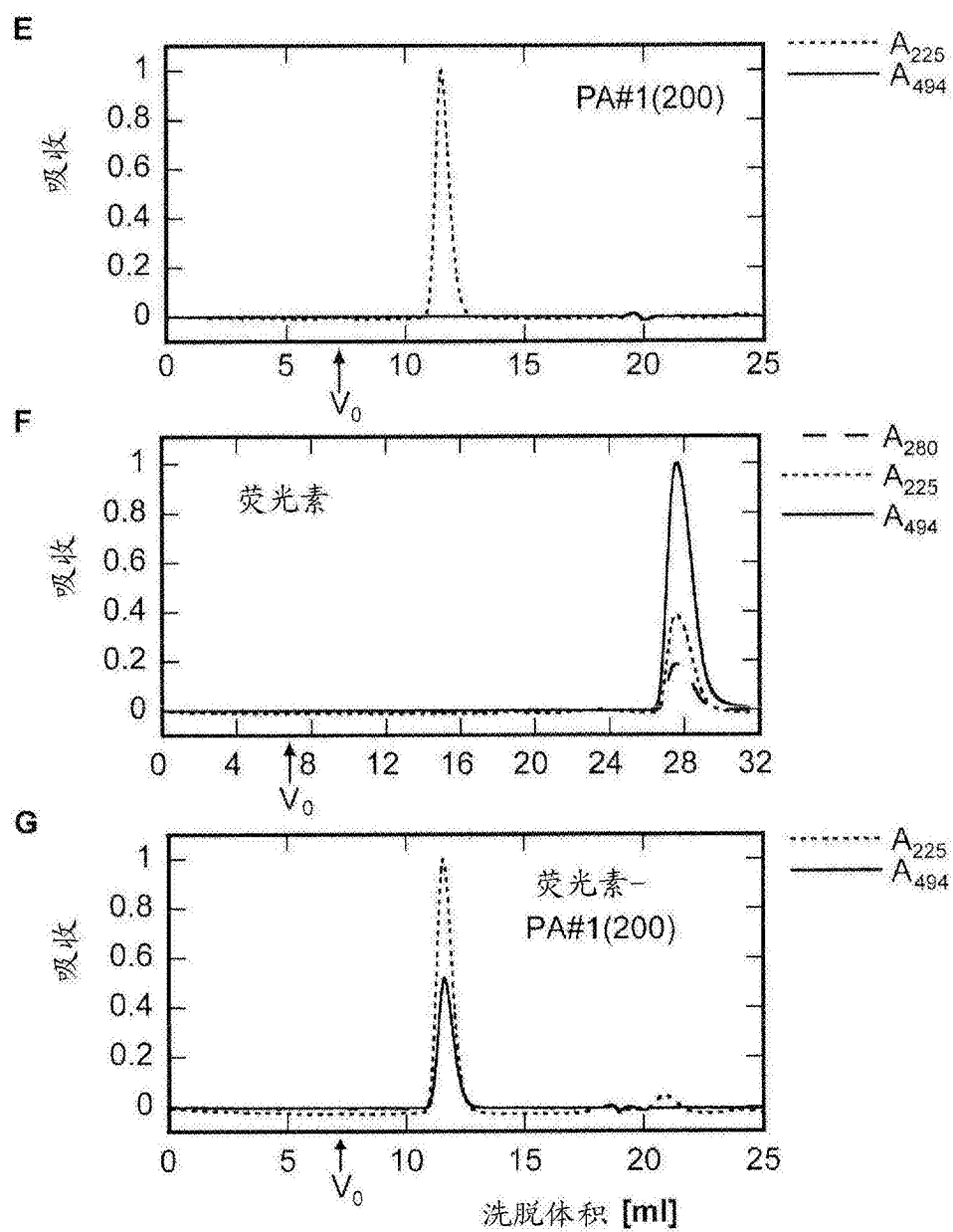
A**B**

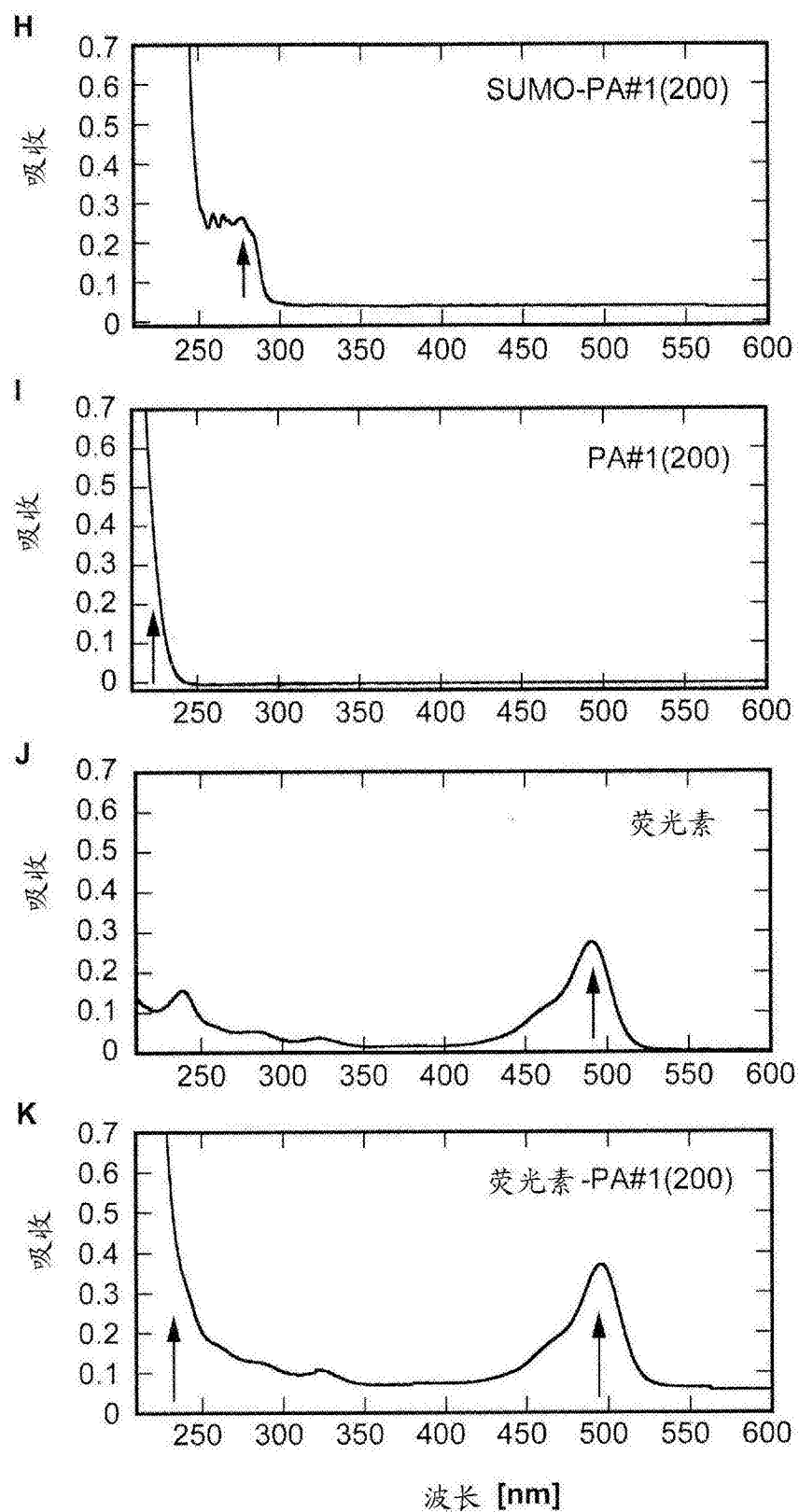
1: 切割前的SUMO-PA#1(200)

2: 用SUMO蛋白酶切割的SUMO-PA#1(200)

图12







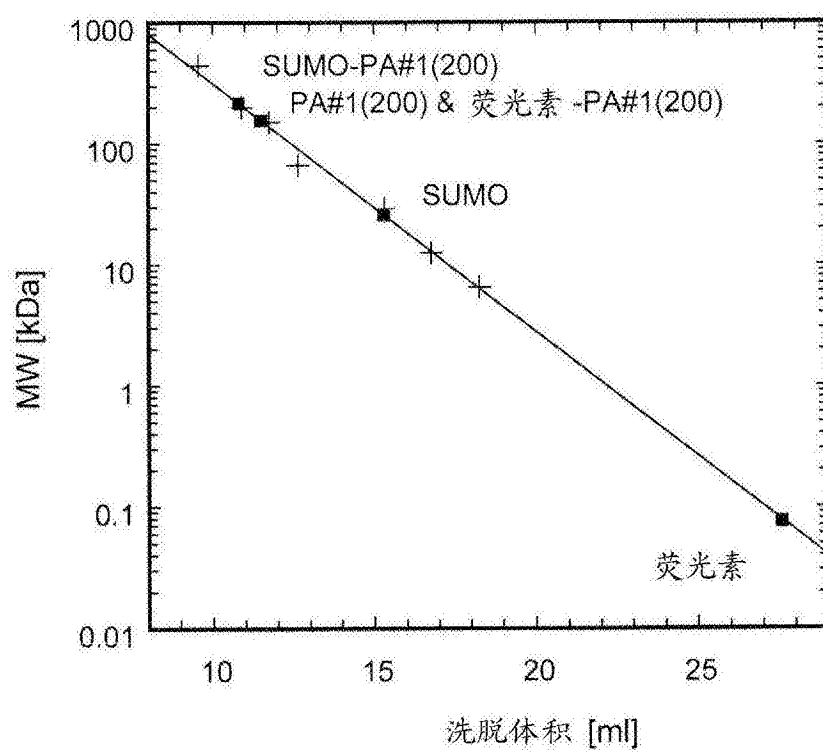
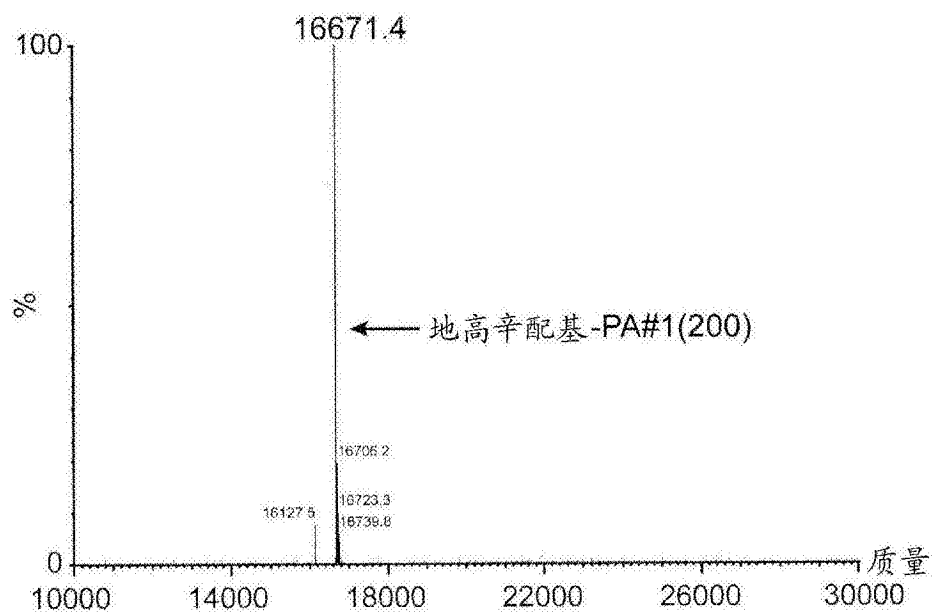
L**M**

图13

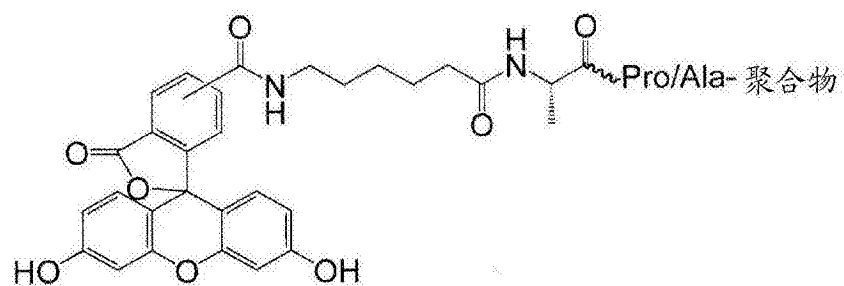
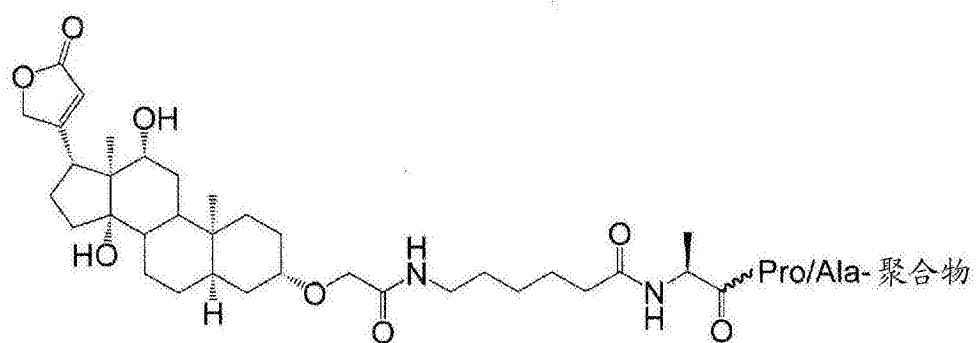
A**B**

图14