

[51] Int. Cl.
G01N 33/50 (2006.01)

1. 一种提高所关注蛋白质的稳定性的方法，所述方法包含以下步骤：

鉴别所关注蛋白质的表面残基在与所关注的所述蛋白质相关的其它蛋白质中是非高度保守的；和

用在生理 pH 值下带正电的氨基酸残基置换多个非保守表面残基。
2. 一种提高所关注蛋白质的稳定性的方法，所述方法包含以下步骤：

鉴别所关注蛋白质的表面残基在与所关注的所述蛋白质相关的其它蛋白质中是非高度保守的；和

用在生理 pH 值下带负电的氨基酸残基置换多个非保守表面残基。
3. 根据权利要求 1 或 2 所述的方法，其中所述非保守表面残基是疏水性的。
4. 根据权利要求 1 所述的方法，其中所述非保守表面残基是亲水性的或带负电的。
5. 根据权利要求 2 所述的方法，其中所述非保守表面残基是亲水性的或带正电的。
6. 一种提高所关注蛋白质的稳定性的方法，所述方法包含以下步骤：

鉴别所关注蛋白质的表面残基；

鉴别所关注的所述蛋白质的表面残基在与所关注的所述蛋白质相关的其它蛋白质中是非高度保守的；

给每一个所述所鉴别的非保守表面残基指定疏水性值；和

用在生理 pH 值下带电的氨基酸残基置换至少一个表面残基，其中所述被置换的残基全部突变为相同类型的残基，其中所述类型的残基是带正电残基或带负电残基。
7. 根据权利要求 6 所述的方法，其中所述置换步骤包含置换至少一个疏水性表面残基。
8. 根据权利要求 6 所述的方法，其中所述置换步骤包含置换至少一个亲水性表面残基。
9. 根据权利要求 6 所述的方法，其中所述置换步骤包含置换至少一个带电表面残基。
10. 根据权利要求 6 所述的方法，其中所述置换步骤包含用赖氨酸残基置换至少一个表面残基。
11. 根据权利要求 6 所述的方法，其中所述置换步骤包含用天冬氨酸或谷氨酸残基置换至少一个表面残基。
12. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含置换至少两个表面残

基。

13. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含置换至少五个表面残基。
14. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含置换至少十个表面残基。
15. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含置换至少二十个表面残基。
16. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含置换至少三十个表面残基。
17. 根据权利要求 1、2 或 6 所述的方法，其中所述方法产生在生理 pH 值下净电荷比所关注的初始蛋白质高的修饰蛋白质。
18. 根据权利要求 2 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带负电荷比所关注的初始蛋白质多的所关注的修饰蛋白质。
19. 根据权利要求 2 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带负电荷比所关注的初始蛋白质多至少-5 的所关注的修饰蛋白质。
20. 根据权利要求 2 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带负电荷比所关注的初始蛋白质多至少-10 的所关注的修饰蛋白质。
21. 根据权利要求 2 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带负电荷比所关注的初始蛋白质多至少-15 的所关注的修饰蛋白质。
22. 根据权利要求 2 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带负电荷比所关注的初始蛋白质多至少-20 的所关注的修饰蛋白质。
23. 根据权利要求 1 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带正电荷比所关注的初始蛋白质多的所关注的修饰蛋白质。
24. 根据权利要求 1 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带正电荷比所关注的初始蛋白质多至少+5 的所关注的修饰蛋白质。
25. 根据权利要求 1 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带正电荷比所关注的初始蛋白质多至少+10 的所关注的修饰蛋白质。
26. 根据权利要求 1 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带正电荷比所关注的初始蛋白质多至少+15 的所关注的修饰蛋白质。
27. 根据权利要求 1 或 6 所述的方法，其中所述方法产生在生理 pH 值下所带正电荷比所关注的初始蛋白质多至少+20 的所关注的修饰蛋白质。

28. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是易聚集的蛋白质。
29. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是疏水性蛋白质。
30. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是膜蛋白。
31. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是很难过度表达的蛋白质。
32. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是很难纯化的蛋白质。
33. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是受体。
34. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是转录因子。
35. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是酶。
36. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是结构蛋白。
37. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是荧光蛋白。
38. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是绿色荧光蛋白(GFP)。
39. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是细胞外蛋白。
40. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是链霉亲和素。
41. 根据权利要求 1、2 或 6 所述的方法, 其中所述所关注蛋白质是谷胱甘肽-S-转移酶。
42. 根据权利要求 1、2 或 6 所述的方法, 其中所述鉴别所述表面残基的步骤包含计算机模拟所述蛋白质的三维结构。
43. 根据权利要求 1、2 或 6 所述的方法, 其中所述鉴别所述表面残基的步骤包含使用算法预测残基是否可见于蛋白质的表面。
44. 根据权利要求 1、2 或 6 所述的方法, 其中所述鉴别所述表面残基的步骤包含鉴别 AvNAPSA 值小于阈值的残基。
45. 根据权利要求 1、2 或 6 所述的方法, 其中所述鉴别所述非高度保守的表面残基的步骤包含比对所述蛋白质的氨基酸序列与来自同一蛋白质家族的至少一个其它蛋白质。
46. 根据权利要求 1、2 或 6 所述的方法, 其中所述鉴别所述非高度保守的表面残基的步骤包含比对所述蛋白质的氨基酸序列与来自不同物种的所述蛋白质的至少一个其它氨基酸序列。
47. 根据权利要求 45 或 46 所述的方法, 其中所述比对以至少两个其它蛋白质序列进行。
48. 根据权利要求 45 或 46 所述的方法, 其中所述比对以至少三个其它蛋白质序列进行。
49. 根据权利要求 45 或 46 所述的方法, 其中所述比对以至少五个其它蛋白质序列进行。
50. 根据权利要求 6 所述的方法, 其中所述指定疏水性值的步骤包含使用辛醇/水分配

系数值。

51. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含诱变所述蛋白质的序列以用在生理 pH 值下带电的天然氨基酸残基置换所述鉴别的疏水性表面残基。
52. 根据权利要求 51 所述的方法，其中所述天然氨基酸残基选自由赖氨酸、谷氨酸、天冬氨酸、组氨酸以及精氨酸组成的群组。
53. 根据权利要求 51 所述的方法，其中所述天然氨基酸残基选自由赖氨酸、组氨酸以及精氨酸组成的群组。
54. 根据权利要求 51 所述的方法，其中所述天然氨基酸残基是赖氨酸。
55. 根据权利要求 51 所述的方法，其中所述天然氨基酸残基选自由谷氨酸和天冬氨酸组成的群组。
56. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含定点诱变所述鉴别的表面残基。
57. 根据权利要求 1、2 或 6 所述的方法，其中所述置换步骤包含 PCR 诱变所述鉴别的表面残基。
58. 一种绿色荧光蛋白 (+36 GFP)，其具有以下氨基酸序列：

MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLT
LKFICTTGKLPVPWPTLVTTLTYGVCFSRYPKHKRHDFFKSAMPKGYVQERTISF
KKGDKYKTRAEVKFEGRTLVRNRIKLKGRDFKEKGNI LGHKLRYNFNSHKVYITAD
KRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDP
KEKR DHMVLLFVTAAGIKHGRDERYK.

59. 一种多核苷酸，其编码根据权利要求 58 所述的绿色荧光蛋白。
60. 根据权利要求 59 所述的多核苷酸，其具有以下序列：

ATGGGGCATCATCATCATCACACGGCGGGGCGTCTAAGGGAGAGCGCTTGTTTC
GCGGCAAAGTCCCGATTCTTGTGGAGCTCAAAGGTGATGTAAATGGTCATAAATT
TAGTGTGCGCGGGAAAGGGAAAGGAGATGCTACGCGGGGCAAGCTCACCTGAA
ATTTATTTGCACAACCGGCAAACCTGCCAGTGCCGTGGCCTACATTAGTCACTACT
CTGACGTACGGTGTTCAGTGCTTTTCTCGCTATCCCAAACACATGAAACGCCATG
ATTTCTTCAAGAGCGCGATGCCAAAAGGTTATGTGCAGGAACGCACCATCAGCTT
TAAAAAAGACGGCAAATATAAAACCCGTGCAGAAGTTAAATTCGAAGGCCGCAC
CCTGGTCAACCGCATTAAGTGAAGGTCGTGACTTCAAAGAGAAAGGTAATAT
TCTTGGTCACAACTGCGCTATAATTTCAACTCTCACAAAGTTTATATTACGGCG
GATAAACGTAAAAACGGGATTAAAGCGAAATTTAAGATTCGTCATAATGTTAAA
GACGGCAGTGTGCAGTTAGCGGATCATTATCAGCAGAATACCCCAATTGGTCGC
GGTCCAGTGCTGCTGCCGCGTAACCATTATCTGTGCGACCCGCAGCAAACCTCAGCA
AAGACCCGAAAGAAAAACGTGACCACATGGTATTACTGGAATTTGTGACCGCAG
CAGGCATTAAACATGGCCGCGATGAACGTTACAAATAG.

61. 一种绿色荧光蛋白 (+49 GFP), 其具有以下氨基酸序列:

MGHHHHHHGGRSKGKRLFRGKVPILVKLKGDVNGHKFSVRGKGKDATRGLTLK
FICTTGKLPVPWPTLVTTLTYGVCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKK
DGKYKTRAEVKFKGRTL VNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRK
NGIKAKFKIRHNVKDGSVQLAKHYQQNTPIGRGPVLLPRKHLYLSTRSKLSKDPKEKR
DHMVLKEFVTAAGIKHGRKERYK.

62. 一种多核苷酸, 其编码根据权利要求 61 所述的绿色荧光蛋白。

63. 根据权利要求 62 所述的多核苷酸, 其具有以下序列:

ATGGGCCACCATCATCATCACCGGGGACGCTCTAAAGGTAAACGTCTGTTTC
GTGGAAGGTGCCATTCTGGTTAACTCAAAGGTGATGTCAACGGCCATAAGTT
TTCGGTTCGTGGCAAAGGTAAAGGTGATGCGACGCGGGAAATTAACACTGAA
ATTTATTTGCACAACCGGAAAACCTCCCTGTGCCGTGGCCGACTTTGGTGACCACA
TTAACCTATGGTGTTCATGCTTCTCACGTTATCCGAAGCATATGAAACGTCATG
ATTTTTTCAAATCGGCTATGCCGAAAGGTTACGTCCAGGAGCGCACCATCTCATT
TAAGAAAGACGGTAAGTATAAAACCCGTGCTGAAGTAAAATTCAAAGGACGCAC
CCTGGTGAATCGCATTAACTGAAAGGTCTGATTTCAAAGAAAAGGGAAATAT
TTAGGGCATAAGCTCCGTTATAATTTTAACAGTCATAAGGTGTATATTACCGCT
GATAACGCAAAAACGGAATCAAAGCGAAATTTAAGATCCGTCATAATGTAAAA
GATGGCTCAGTCCAACCTGGCAAAACATTACCAGCAGAATACCCCGATCGGCCGC
GGTCCTGTGCTTCTGCCGCGTAAACACTACTTGTGACCCGGTCAAATTGAGTA
AAGATCCGAAGGAAAAGCGTGATCACATGGTCTTGAAGGAATTTGTAACCTGCAG
CAGGTATTAAACACGGGCGCAAAGAACGTTACAAATAG.

64. 一种绿色荧光蛋白 (-29 GFP), 其具有以下氨基酸序列:

MGHHHHHHGGASKGEELFDGEVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKF
ICTTGELPVPWPTLVTTLTYGVCFSRYPDHMDQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHDVYITADKQENGI
KAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLSTESALS KDPNEDRDHM
VLLEFVTAAGIDHGMDELYK.

65. 一种多核苷酸, 其编码根据权利要求 64 所述的绿色荧光蛋白。

66. 根据权利要求 65 所述的多核苷酸, 其具有以下序列:

ATGGGGCATCACCATCATCATCATGGCGGTGCGTCTAAGGGGGAGGAGTTATTTG
ATGGTGAAGTGCCGATCCTGGTGGAGCTTGATGGCGATGTTAACGGCCATGAATT
TTCTGTCCGCGGTGAAGGGGAGGGTGATGCCACGGAAGGGGAGCTGACACTTAA
ATTTATTTGCACCACCGGTGAACTCCCGGTCCCGTGGCCGACCCTGGTGACCACC
CTGACCTACGGCGTTCAATGCTTTTACGTTATCCGGATCACATGGACCAACACG
ACTTCTTTAAAAGCGCGATGCCTGAAGGCTATGTTCAAGAACGTACAATTAGTTT
TAAAGATGACGGCACCTACAAGACCCGTGCGGAAGTAAAATTTGAAGGGGACAC
TTAGTGAAACCGCATCGAGCTGAAAGGGATCGATTTTAAAGAAGATGGGAATAT
CCTGGGACACAACTTGAATACAACCTTAAATAGTCATGACGTCTATATCACGGCG
GACAAACAGGAAAACGGAATTAAGGCAGAATTTGAGATTCGGCATAATGTGCA
GATGGCTCGGTACAGTTGGCTGATCACTATCAGCAGAATACGCCGATTGGAGAT
GGTCCGTTTTTATTACCAGACGATCACTATCTGTCCACCGAATCCGCCCTGAGCA
AAGATCCGAATGAAGACCGGGACCATATGGTTCTGCTGGAATTTGTTACGGCGG
CTGGTATTGACCATGGCATGGATGAGCTGTATAAGTAG.

67. 一种链霉亲和素 (-40 SAV), 其具有以下氨基酸序列:

MGHHHHHHGGAEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGDAESEYVLT
GRYDSAPATDGS GTALGWTVAWKNDYENAHSA TTWSGQYVGGAEARINTQWLLT
SGTTEADAWKSTLVGHDTFTKVEPSAAS.

68. 一种多核苷酸, 其编码根据权利要求 67 所述的链霉亲和素。

69. 根据权利要求 68 所述的多核苷酸, 其具有以下序列:

GGTTCAGCCATGGGTCATCACCACCACCATCACGGTGGCGCCGAAGCAGGTATT
ACCGGTACCTGGTATAACCAGTTAGGCTCAACCTTTATTGTGACCGCGGGAGCGG
ACGGCGCCTTAACCGGTACCTACGAATCAGCTGTAGGTGACGCGGAATCAGAGT
ACGTATTAACCGGTCGTTATGATAGCGCGCCGGCGACTGACGGTAGCGGTACTGC
TTTAGGTTGGACCGTAGCGTGGAAGAATGATTATGAAAACGCACATAGCGCAAC
AACGTGGTCAGGGCAGTACGTTGGCGGAGCTGAGGCGCGCATTAAACACGCAGTG
GTTATTAAGTAGCGGCACCACTGAAGCTGATGCCTGGAAGAGCACGTTAGTGGG

TCATGATACCTTCACTAAAGTGGAACCTTCAGCTGCGTCATAATAATGACTCGAG
ACCTGCA.

70. 一种链霉亲和素 (+52 SAV), 其具有以下氨基酸序列:

MGHHHHHHGGAKAGITGTWYNQLGSTFIVTAGAKGALTGTYESAVGNAKSRYVLT
GRYDSAPATKSGGTALGWTVAWKNKYRNAHSA TTWSGQYVGGAKARINTQWLLT
SGTTKAKAWKSTLVGHDTFTKVKPSAAS.

71. 一种多核苷酸, 其编码根据权利要求 70 所述的链霉亲和素。

72. 根据权利要求 71 所述的多核苷酸, 其具有以下序列:

GGTTCAGCCATGGGTCATCACCACCACCATCACGGTGGCGCCAAAGCAGGTATT
ACCGGTACCTGGTATAACCAGTTAGGCTCAACCTTTATTGTGACCGCGGGAGCGA
AAGGCGCCTTAACCGGTACCTACGAATCAGCTGTAGGAAACGCAAAATCACGCT
ACGTATTAACCGGTCGTTATGATAGCGCGCCGGCGACTAAAGGTAGCGGTACTG
CTTTAGGTTGGACCGTAGCGTGGAAGAATAAGTATCGTAATGCGCACAGTGCTAC
CACTTGGTCAGGGCAGTACGTAGGGGGAGCCAAAGCACGTATCAACACGCAGTG
GTTATTAACATCAGGTACCACCAAAGCGAAAGCCTGGAAGAGCACGTTAGTGGG
TCATGATACCTTCACTAAAGTGAAACCTTCAGCTGCGTCATAATAATGACTCGAG
ACCTGCA.

73. 一种谷胱甘肽-S-转移酶 (-40 GST), 其具有以下氨基酸序列:

MGHHHHHHGGPPYTITYFPVRGRCEAMRMLLADQDQSWEEEVVTMETWPPLKPSC
LFRQLPKFQDGDLTLYQSNAILRHLGRSFGLYGEDEEEAALVDMVNDGVEDLRCKY
ATLIYTDYEAGKEEYVEELPEHLKPFETLLSENEGGEAFVVGSEISFADYNLLDLLRIH
QVLNPSCLDAFPLLSAYVARLSARPEIEAFLASPEHVDRPINGNGKQ.

74. 一种多核苷酸，其编码根据权利要求 73 所述的链霉亲和素。

75. 根据权利要求 74 所述的多核苷酸，其具有以下序列：

GGTTCAGCCATGGGTCATCACCACCACCATCACGGTGGCCCGCCGTACACCATTA
CATACTTTCCGGTACGTGGTCGTTGTGAAGCGATGCGTATGTTATTAGCGGACCA
GGACCAATCATGGGAAGAAGAAGTAGTGACAATGGAAACCTGGCCGCCGTAAA
GCCTAGCTGTTTATTCCGTCAATTACCGAAGTTTCAGGATGGTGATTTAACCTTAT
ACCAGTCTAACGCGATCTTACGTCATTTAGGTCGCTCATTTGGTTTATACGGTGA
AGATGAAGAAGAAGCAGCCTTAGTGGAATATGGTGAATGATGGCGTGGAAGACTT
ACGTTGTAAATACGCGACGTTAATTTACACTGATTATGAAGCCGGTAAAGAGGA
GTACGTGGAAGAATTACCTGAACACCTGAAGCCGTTTGAACATTACTGAGCGA
AAATGAAGGAGGTGAGGCGTTTCGTAGTTGGTAGCGAAATTAGCTTCGCTGATTAT
AAGTTATTAGACTTATTACGCATTCACCAGGTTTTAAATCCTAGCTGTTTAGACGC
TTTCCCGTTACTGAGCGCATATGTAGCGCGCCTGAGCGCCCGTCCGGAAATTGAA
GCTTTCTTAGCGTCACCTGAACACGTAGACCGCCCGATTAACGGAAACGGCAAG
CAGTAATAATGAGGTACCACCTGCA.

76. 一种蛋白质，其由根据权利要求 1-57 中任一权利要求所述的方法修饰。

77. 一种多核苷酸，其编码根据权利要求 76 所述的蛋白质。

78. 一种试剂盒，其用于进行根据权利要求 1-57 中任一权利要求所述的方法。

蛋白质表面重建

相关申请案

本申请案根据 35 U.S.C. § 119(e)主张 2006 年 6 月 2 日申请的美国临时专利申请案 USSN 60/810,364 的优先权，其全部内容以引用的方式并入本文中。

政府支持

本文中所述的研究部分得到美国国立卫生研究院（National Institutes of Health）基金（GM065400）的支持。美国政府可具有本发明的某些权利。

技术领域

无

背景技术

蛋白质是细胞的主要组成成分。蛋白质催化化学反应，在生物系统中转导信号，在细胞和细胞外基质中提供结构元件，充当信使等。蛋白质不良表现的主要原因之一是聚集。这不仅是实验室中的问题，而且是诸如阿尔茨海默病（Alzheimer's disease）等许多疾病的问题。在达到计算设计的蛋白质时，聚集是尤其烦恼的问题。举例来说，TOP7 是具有新颖折叠的计算设计的蛋白质。TOP7 的较长形式（延长的 TOP7（TOP7 extended））极易聚集。TOP7ex 主要表达为不可溶的聚集体。

随着更多蛋白质被设计或修饰以用作研究生物系统的工具或随着更多野生型或修饰蛋白质被用作治疗剂，需要常规上将这些蛋白质修饰为更稳定和/或防止聚集的系统。

发明内容

本发明提供一种修饰蛋白质以使其更稳定的系统。本发明的产生是基于对修饰蛋白质表面上的疏水性区域可提高蛋白质的超热力学性质的认识。本发明系统尤其适用于提高所关注蛋白质的溶解性，提高蛋白质的抗聚集性，和/或提高蛋白质的复性能力。所有这些特性尤其适用于蛋白质生产、蛋白质纯化以及使用蛋白质作为治疗剂和研究工具。

在一方面，本发明提供一种改变蛋白质的一级序列以增加宽范围条件下蛋白质的抗聚集性、溶解性、重折叠能力和/或总体稳定性的方法。修饰蛋白质的活性优选大致或大体上与未修饰的蛋白质相同。在某些实施例中，修饰蛋白质保留至少 50%、75%、90% 或 95% 的野生型蛋白质的活性。在一实施例中，所述方法包括以下步骤：（a）鉴别所关

注蛋白质的表面残基；(b) 鉴别特定表面残基在与所关注蛋白质相关的其它蛋白质中是非高度保守的（即，确定哪些氨基酸对蛋白质活性或功能来说不是不可缺少的）；(c) 确定所鉴别的非保守表面残基的疏水性；以及(e) 用在生理 pH 值下极性更高或带电的氨基酸置换至少一个或一个以上所鉴别的疏水性非保守残基。各以上步骤均可以使用所属领域中已知的任何技术、计算机软件、算法、范式等进行。产生修饰蛋白质后，可测试其活性和/或所寻求的预期性质。在某些实施例中，修饰蛋白质更加稳定。在某些实施例中，修饰蛋白质不易聚集。本发明方法通常增加在生理 pH 值下蛋白质的净电荷（正或负）。

在另一方面，本发明提供一种通过使蛋白质“超带电（supercharging）”来改变蛋白质的一级序列以增加宽范围条件下蛋白质的抗聚集性、溶解性、重折叠能力和/或总体稳定性的方法。也就是说，与野生型蛋白质相比，修饰蛋白质上总的净电荷增加（正电荷或负电荷）。修饰蛋白质的活性优选大致或大体上与未修饰的蛋白质相同。在某些实施例中，所述方法包括以下步骤：(a) 鉴别所关注蛋白质的表面残基；(b) 鉴别特定表面残基在与所关注蛋白质相关的其它蛋白质中是非高度保守的（即，确定哪些氨基酸对蛋白质活性或功能来说不是不可缺少的）；(c) 确定所鉴别的非保守表面残基的亲水性；以及(e) 用在生理 pH 值下带电的带电氨基酸置换至少一个或一个以上所鉴别的带电或极性、溶剂暴露非保守残基。在某些实施例中，为制备带负电的“超带电”蛋白质，使所鉴别用于修饰的残基突变为天冬氨酸（Asp）或谷氨酸（Glu）残基。在某些其它实施例中，为制备带正电的“超带电”蛋白质，使所鉴别用于修饰的残基突变为赖氨酸（Lys）或精氨酸（Arg）残基。各以上步骤都可以使用所属领域中已知的任何技术、计算机软件、算法、范式等进行。产生修饰蛋白质后，可测试其活性和/或所寻求的预期性质。在某些实施例中，修饰蛋白质（“超带电蛋白质”）更加稳定。在某些实施例中，修饰蛋白质不易聚集。本发明方法通常增加在生理 pH 值下蛋白质的净电荷（正或负）。

编入蛋白质数据库（Protein Data Bank, PDB）中的超过 80% 的蛋白质的理论净电荷在 ± 10 范围内。本发明所产生的修饰蛋白质通常具有小于 -10 或大于 +10 的净电荷。在某些实施例中，修饰蛋白质具有小于 -20 或大于 +20 的净电荷。在某些实施例中，修饰蛋白质具有小于 -30 或大于 +30 的净电荷。在某些实施例中，修饰蛋白质具有小于 -40 或大于 +40 的净电荷。在某些实施例中，修饰蛋白质具有小于 -50 或大于 +50 的净电荷。修饰蛋白质能够恰当折叠并保留其生物活性。

可使用本发明系统修饰任何蛋白质，并且本发明系统所产生的蛋白质变异体以及编码变异蛋白质的多核苷酸或载体和表达变异蛋白质的细胞都视为本发明的一部分。已经

使用本发明系统产生了数种新颖的绿色荧光蛋白（green fluorescent protein, GFP）变异体。这些变异体保留其荧光性；然而，其在宽范围环境下比 GFP 的现有形式更稳定。本发明 GFP 即使经过很长时间和在诱导聚集的环境中也不会聚集，并且即使通过煮沸变性后也能够重折叠为荧光蛋白。也已经使用本发明系统产生了新颖的链霉亲和素（streptavidin）和谷胱甘肽-S-转移酶（glutathione-S-transferase, GST）变异体。这些变异体保留其生物活性并且在加热时仍可溶。本发明还包括编码本发明 GFP、链霉亲和素以及 GST 蛋白序列的多核苷酸序列；包括任何这些核苷酸序列的载体；以及包括这样的多核苷酸序列或载体或表达本发明变异体的细胞。在某些实施例中，本发明包括过度表达本发明变异体的细菌或其它细胞。本发明变异体可用于所属领域已知的多种生物测定中。举例来说，超带电 GFP 可用于目前使用 GFP 作为报告蛋白的任何测定中。

在另一方面，本发明提供已经由本发明系统修饰的其它蛋白质。这些修饰蛋白质优选保留其显著比例的初始活性。在某些实施例中，修饰蛋白质保留至少 99%、98%、95% 或 90% 的未修饰形式的活性。修饰蛋白质在多种条件可更可溶，抗聚集，重折叠能力增加和/或稳定性更高。由本发明系统修饰的蛋白质包括疏水性蛋白、重组蛋白、膜蛋白、结构蛋白、酶、细胞外蛋白、治疗蛋白（例如，胰岛素、细胞激素、免疫球蛋白、免疫球蛋白片段等）、受体、细胞信号传导蛋白、胞浆蛋白、核内蛋白、转录因子等。在某些特定实施例中，蛋白质是用于人类或兽医学的治疗蛋白。在某些实施例中，蛋白质是非天然蛋白质，例如计算设计的蛋白质。在其它实施例中，蛋白质是杂合蛋白、融合蛋白、变构蛋白、突变蛋白、基因工程蛋白或已经通过人工改变的任何其它蛋白质。

还提供用于实施本发明的试剂盒。试剂盒可包括修饰所关注蛋白质以使其更抗聚集、增加其复性能力或增加其总体稳定性所需的试剂。这些试剂盒可包括所有或部分以下各物：多核苷酸、计算机软件、核苷酸、引物、载体、细胞系、说明书、培养板、培养基、缓冲液、酶、艾本德管（Eppendorf tube）、定点诱变试剂盒等。试剂盒优选经简便包装以供实验室装置使用。研究员通常提供待使用本发明技术修饰的蛋白质的 DNA 编码序列。

定义

“氨基酸”：术语“氨基酸”是指蛋白质的基本结构亚单元（subunit）。 α -氨基酸由氨基、羧基、氢原子以及侧链（即，R 基团）组成，均键结于中心碳原子。这个中心碳原子因为与羧基相邻而称为 α 碳。存在二十种天然氨基酸，包括甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、苯丙氨酸、酪氨酸、色氨酸、半胱氨酸、蛋氨酸、丝氨酸、苏氨酸、赖氨酸、精氨酸、组氨酸、天冬氨酸、谷氨酸、天冬酰胺、谷氨酸以及脯氨酸。

疏水性氨基酸包括丙氨酸、缬氨酸、亮氨酸、异亮氨酸以及苯丙氨酸。芳香族氨基酸包括苯丙氨酸、酪氨酸、色氨酸以及组氨酸。极性氨基酸包括酪氨酸、半胱氨酸、丝氨酸、苏氨酸、赖氨酸、精氨酸、组氨酸、天冬氨酸、谷氨酸、天冬酰胺以及谷氨酰胺。含硫氨基酸包括半胱氨酸和蛋氨酸。碱性氨基酸包括赖氨酸、精氨酸以及组氨酸。酸性氨基酸包括天冬氨酸和谷氨酸。也有蛋白质中被插入非天然氨基酸。在某些实施例中，当使用术语“氨基酸”时是指二十种天然氨基酸。

“抗体”：术语“抗体”是指天然或者完全或部分合成产生的免疫球蛋白。维持特异性结合能力的所有其衍生物也包括在这一术语内。这一术语也涵盖具有与免疫球蛋白结合域同源或者基本上同源的结合域的任何蛋白质。这些蛋白质可源自天然来源或者部分或完全合成产生。抗体可为单克隆抗体或多克隆抗体。抗体可为任何免疫球蛋白类别（包括人类类别 IgG、IgM、IgA、IgD 以及 IgE 中的任一类别）的成员。

“保守”：术语“保守”是指多核苷酸序列或氨基酸序列的相应核苷酸或氨基酸残基在所比较的两个或两个以上相关序列的相同位置未发生变化。相对保守的核苷酸或氨基酸是在相关序列中比序列中其它地方所出现的核苷酸或氨基酸保守的那些核苷酸或氨基酸。

“同源”：如本文中所使用的术语“同源”是所属领域通晓的术语，是指核酸或蛋白质在核苷酸或氨基酸序列层面高度相关。彼此同源的核酸或蛋白质称为同源物。同源可指两个序列（即，核苷酸序列或氨基酸）之间的序列相似性程度。本文中所提及的同源性%数字反映两个序列之间的最大可能同源性，即，比对两个序列以具有最大数目的匹配（同源）位置时的同源性%。同源性可由已知方法容易地加以计算，诸如以下文献中所述的方法：计算分子生物学（Computational Molecular Biology），莱斯克（Lesk, A. M.）编，牛津大学出版社（Oxford University Press），纽约（New York），1988；生物计算：信息学和基因组计划（Biocomputing: Informatics and Genome Projects），史密斯（Smith, D. W.）编，学术出版社（Academic Press），纽约（New York），1993；分子生物学中的序列分析（Sequence Analysis in Molecular Biology），凡·海耶（von Heinje, G.），学术出版社（Academic Press），1987；序列数据的计算机分析第 I 部分（Computer Analysis of Sequence Data, Part I），格里菲因（Griffin, A. M.）和格里菲因（Griffin, H. G.）编，胡玛纳出版社（Humana Press），新泽西州（New Jersey），1994；和序列分析引物（Sequence Analysis Primer），吉布考夫（Gribskov, M.）和德夫莱克（Devereux, J.）编，米斯托克顿出版社（M Stockton Press），纽约（New York），1991；各自以引用的方式并入本文中。测定序列之间同源性的常用方法包括（但不限于）以下文献中所揭示的方法：卡里

罗 (Carillo, H.) 和里皮曼 (Lipman, D.), SIAM 应用数学 (SIAM J Applied Math.), 48:1073 (1988); 其以引用的方式并入本文中。测定同源性的技术被编成了公开可用的计算机程序。测定两个序列之间同源性的示范性计算机软件包括 (但不限于) GCG 程序包 (德夫莱克 (Devereux, J.) 等人, 核酸研究 (Nucleic Acids Research), 12(1), 387 (1984))、BLASTP、BLASTN 以及 FASTA (阿特舒 (Atschul, S. F.) 等人, 分子生物学杂志 (J Molec. Biol.), 215, 403(1990))。

术语“同源”必定指至少两个序列 (核苷酸序列或氨基酸序列) 之间的比较。根据本发明, 如果两个核苷酸序列所编码的多肽有至少一段的至少 20 个氨基酸至少约 50-60% 一致, 优选约 70% 一致, 那么将其视为同源。同源核苷酸序列的特征优选还在于编码一段至少 4-5 个特别指定的氨基酸的能力。关于待考虑同源性的核苷酸序列, 必须考虑这些氨基酸彼此之间的一致性和大致间隔。对于长度小于 60 个核苷酸的核苷酸序列, 由编码一段至少 4-5 个特别指定的氨基酸的能力确定同源性。

“肽”或“蛋白质”: 根据本发明, “肽”或“蛋白质”包含一串通过肽键连接在一起的至少三个氨基酸。术语“蛋白质”和“肽”可互换使用。本发明的肽优选仅含有天然氨基酸, 但也可使用所属领域已知的非天然氨基酸 (即, 自然界中不存在但可并入多肽链中的化合物) 和/或氨基酸类似物。同时, 可修饰本发明的肽中的一个或一个以上氨基酸, 例如通过添加碳水化合物基团、磷酸根基、法呢基 (farnesyl)、异法呢基、脂肪酸基团、用于连接、官能化或其它修饰 (例如, α 酰胺化) 的连接基团等化学实体来修饰。在优选实施例中, 对肽修饰产生更稳定的肽 (例如, 活体内半衰期更高)。这些修饰可包括使肽环化、并入 D-氨基酸等。这些修饰均不对肽的预期生物活性产生实质性的干扰。在某些实施例中, 对肽修饰产生生物学活性更高的肽。

“多核苷酸”或“寡核苷酸”: 多核苷酸或寡核苷酸是指核苷酸的聚合物。多核苷酸通常包含至少三个核苷酸。聚合物可包括天然核苷 (即, 腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷以及脱氧胞苷)、核苷类似物 (例如, 2-氨基腺苷、2-硫代胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、C5-丙炔基胞苷、C5-丙炔基尿苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-甲基胞苷、7-脱氮腺苷 (7-deazaadenosine)、7-脱氮鸟苷 (7-deazaguanosine)、8-氧代腺苷 (8-oxoadenosine)、8-氧代鸟苷 (8-oxoguanosine)、O(6)-甲基鸟嘌呤以及 2-硫代胞苷)、化学修饰碱基、生物修饰碱基 (例如, 甲基化碱基)、插入碱基、修饰糖 (例如, 2'-氟核糖、核糖、2'-脱氧核糖、阿拉伯糖 (arabinose) 以及己糖) 和/或修饰磷酸根基 (例如, 硫代磷酸根基和 5'-N-亚磷酰胺键联)。

“小分子”: 如本文中所使用的术语“小分子”是指实验室中制备或自然界中可见的

非肽、非寡聚有机化合物。如本文中所使用的小分子可指“天然产物样”化合物，然而术语“小分子”不限于“天然产物样”化合物。更确切地说，小分子的特征通常在于其含有数个碳碳键并且具有小于 1500 的分子量，但出于本发明的目的，这一特征不打算限制本发明。在某些其它优选实施例中，利用天然产物样小分子。

“稳定”：如本文中所使用的术语“稳定”提及蛋白质时是指蛋白质稳定性的任何方面。与初始野生型蛋白质相比，稳定的修饰蛋白质具有一个或一个以上的以下特征：更可溶、更抗聚集、更抗变性、更抗解折叠、更抗不适当或不希望的折叠、复性能力更强、热稳定性增加、在多种环境（例如，pH 值、盐浓度、存在洗涤剂、存在变性剂等）下稳定性增加以及非水性环境下稳定性增加。在某些实施例中，稳定的修饰蛋白质显示至少两个上述特征。在某些实施例中，稳定的修饰蛋白质显示至少三个上述特征。这些特征可使活性蛋白质在更高水平下生产。举例来说，可在比蛋白质未修饰形式高的水平下过度表达修饰蛋白质而无聚集。这些特征也可使蛋白质用作治疗剂或研究工具。

附图说明

图1：超带电绿色荧光蛋白（GFP）。(a) GFP 变异体的蛋白质序列，形成荧光团的残基突出为绿色、带负电残基突出为红色，而带正电残基突出为蓝色；(b) sfGFP（左）、GFP(+36)（中）以及 GFP(-30)（右）的表面静电势，自 -25 kT/e（红色）至 +25 kT/e（蓝色）着色。

图2：GFP 变异体的分子内性质。(a) 纯 GFP 变异体的染色和紫外荧光。各泳道和试管含有 0.2 μ g 蛋白质。(b) GFP 变异体的圆二色性谱。(c) GFP 变异体的热力学稳定性，由胍盐诱导的解折叠测量。

图3：超带电蛋白质的分子间性质。(a) 纯 GFP 变异体的紫外照射样品（“天然”），在 100°C 加热 1 分钟的样品（“煮沸”）和随后在 25°C 冷却 2 小时的样品（“冷却”）。(b) 在 25°C 下用 40% TFE 诱导 GFP 变异体聚集并通过直角光散射监测。(c) 超带电 GFP 可逆地粘附于带相反电荷的大分子。样品 1：于 30 μ l 25 mM Tris (pH 7.0) 和 100 mM NaCl 中的 6 μ g GFP(+36)。样品 2：向样品 1 中添加 6 μ g GFP(-30)。样品 3：向样品 1 中添加 30 μ g 鲑鱼精 DNA。样品 4：向样品 1 中添加 20 μ g 大肠杆菌 (*E. coli*) tRNA。样品 5：向样品 4 中添加 NaCl 达 1 M。样品 6-8：除使用 sfGFP 代替 GFP(+36) 外，分别与样品 1、2 以及 4 相同。使所有样品在微型离心机中短暂旋转并且在紫外光下观察。(d) GST 变异体的酶法测定。反应含有 0.5 mg/mL GST 变异体、20 mM 二硝基氯苯、20 mM 谷胱甘肽以及 100 mM 磷酸钾 (pH 6.5)。在 340 nm 下监测产物的形成，得到反应速率观察值

(k_{obs}), 野生型 GST 为 6 min^{-1} , GST(-40) 为 2.2 min^{-1} 而煮沸和冷却后 GST(-40) 为 0.9 min^{-1} 。

图 4: GFP 变异体的 (a) 激发光谱和 (b) 发射光谱。各样品含有通过在 490 nm 下发色团吸光度所定量, 等量的蛋白质。

图 5: 链霉亲和素变异体的生物素结合活性, 如先前所述 (卡达 (Kada) 等人, 利用荧光猝灭或荧光偏振快速估算抗生物素蛋白和链霉亲和素 (Rapid estimation of avidin and streptavidin by fluorescence quenching or fluorescence polarization). 生物化学与生物物理学报 (Biochim. Biophys. Acta) **1427**, 44-48 (1999), 其以引用的方式并入本文中) 通过监测生物素-4-荧光素 (英杰公司 (Invitrogen)) 依赖性结合而测量。向 $0.3 \mu\text{M}$ 生物素-4-荧光素 (B4F)、 100 mM NaCl、 1 mM EDTA、 0.1 mg/mL 牛血清白蛋白 (bovine serum albumin BSA)、 50 mM 磷酸钾 (pH 7.5) 中滴定蛋白质样品。在 470 nm 下激发的 Perkin-Elmer LS50B 荧光分光光度仪上测量 526 nm 下的荧光猝灭。相对于含有 600 倍过量的无荧光生物素的对照滴定, 将测量结果归一化。包括图例下部的三个蛋白质作为阴性对照。

具体实施方式

本发明提供一种修饰蛋白质以使其更加稳定的系统。所述系统通过将蛋白质表面上的非保守氨基酸变成极性更大或带电的氨基酸残基来运行。待修饰的氨基酸残基可以呈疏水性、亲水性或其组合。可通过使用本发明系统修饰任何蛋白质来产生更稳定的变异体。已发现对表面残基进行这些修饰可以提高蛋白质的超热力学性质。随着蛋白质日益用作治疗剂和随着其继续用作研究工具, 用于改变蛋白质以使其更加稳定的系统也变得重要和适用。由本发明方法修饰的蛋白质通常抗聚集, 重折叠能力增加, 抗不适当折叠, 溶解性提高并且在包括变性条件 (诸如热或存在洗涤剂) 在内的宽范围条件下通常更加稳定。

可通过使用本发明系统修饰任何蛋白质来产生更稳定的变异体。可修饰天然和非天然蛋白质 (例如, 工程蛋白质)。可被修饰的蛋白质实例包括受体、膜结合蛋白、跨膜蛋白、酶、转录因子、细胞外蛋白、治疗蛋白、细胞因子、信使蛋白、DNA 结合蛋白、RNA 结合蛋白、参与信号转导的蛋白质、结构蛋白、胞浆蛋白、核内蛋白、疏水性蛋白、亲水性蛋白等。待修饰的蛋白质可源自植物、动物或微生物的任何物种。在某些实施例中, 蛋白质是哺乳动物蛋白质。在某些实施例中, 蛋白质是人类蛋白质。在某些实施例中, 蛋白质源自研究中通常使用的生物体。举例来说, 待修饰的蛋白质可来自灵长类动物 (例如, 猿、猴)、啮齿动物 (例如, 兔子、仓鼠、沙鼠)、猪、狗、猫、鱼 (例如,

斑马鱼 (*zebrafish*))、线虫 (例如, 秀丽线虫 (*C. elegans*))、酵母 (例如, 酿酒酵母 (*Saccharomyces cerevisiae*)) 或细菌 (例如, 大肠杆菌 (*E. coli*))。

本发明系统尤其适用于修饰易聚集或具有稳定性问题的蛋白质。本系统也可用于修饰被过度表达的蛋白质。举例来说, 重组产生的治疗蛋白可以从本发明系统的修饰中受益。这些经过修饰的治疗蛋白不仅更易生产和纯化, 而且就蛋白质的保存和使用来说可能更加稳定。

本发明系统包括鉴别所关注蛋白质的非保守表面残基和用在生理 pH 值下为亲水性、极性 or 带电的残基置换这些残基中的一些。本发明系统不仅包括修饰蛋白质的方法, 而且包括适用于修饰蛋白质以使其更稳定的试剂和试剂盒。

使用所属领域中已知的任何方法鉴别待修饰的蛋白质的表面残基。在某些实施例中, 通过计算机模拟蛋白质鉴别表面残基。在某些实施例中, 已知和/或测定蛋白质的三维结构, 并通过观察蛋白质的结构鉴别表面残基。在其它实施例中, 使用计算机软件预测表面残基。在某些特定实施例中, 使用每个侧链原子的平均相邻原子 (Average Neighbor Atoms per Sidechain Atom, AvNAPSA) 预测表面暴露。AvNAPSA 是一种已作为计算机程序实施的自动化表面暴露法。参见附录 A。低 AvNAPSA 值表示表面暴露的残基, 而高值表示位于蛋白质内部的残基。在某些实施例中, 使用所述软件预测蛋白质的二级结构和/或三级结构并基于这一预测鉴别表面残基。在其它实施例中, 表面残基的预测是基于残基的疏水性和亲水性以及其于蛋白质一级序列中的聚类分析。除计算机模拟 (*in silico*) 方法外, 也可使用多种生物化学技术 (例如蛋白酶裂解、表面修饰等) 鉴别蛋白质的表面残基。

然后确定表面残基中哪些是保守的或对蛋白质的功能来说是重要的。可使用所属领域中已知的任何方法确定保守残基的鉴别。在某些实施例中, 通过比对所关注蛋白质的一级序列与相关蛋白质来鉴别保守残基。这些相关蛋白质可来自同一蛋白质家族。举例来说, 如果蛋白质是免疫球蛋白, 那么可使用其它免疫球蛋白序列。相关蛋白质也可以是来自不同物种的相同蛋白质。举例来说, 可通过比对来自不同物种的相同蛋白质的序列来鉴别保守残基。作为另一实例, 可比对具有类似功能或生物活性的蛋白质。优选使用 2、3、4、5、6、7、8、9 或 10 个不同序列来确定蛋白质的保守氨基酸。在某些实施例中, 如果超过 50%、60%、70%、75%、80% 或 90% 的序列在特定位置具有相同氨基酸, 那么残基就被视为是保守的。在其它实施例中, 如果超过 50%、60%、70%、75%、80% 或 90% 的序列在特定位置具有相同或类似 (例如, 缬氨酸、亮氨酸以及异亮氨酸; 甘氨酸和丙氨酸; 谷氨酰胺和天冬酰胺; 或天冬氨酸和谷氨酸) 氨基酸, 那么残基就被

视为是保守的。许多软件包可用于比对和比较如本文中所述的蛋白质序列。所属领域的技术人员应了解可首先确定保守残基或者可首先确定表面残基。顺序并不重要。在某些实施例中，计算机软件包可同时确定表面残基和保守残基。也可通过诱变蛋白质来鉴别蛋白质中的重要残基。举例来说，可使用蛋白质的丙氨酸扫描来确定蛋白质中的重要氨基酸残基。在其它实施例中，可使用定点诱变。

一旦鉴别出蛋白质的非保守表面残基，就将各残基鉴别为呈疏水性或亲水性。在某些实施例中，给残基指定疏水性评分。举例来说，可给各非保守表面残基指定辛醇/水分配系数的对数值 ($\log P$)。也可使用其它疏水性参数。这些氨基酸标度已论述于以下文献中：简因 (Janin)，球状蛋白中的表面和内部体积 ("Surface and Inside Volumes in Globular Proteins")，自然 (*Nature*) 277:491-92, 1979；沃尔芬顿 (Wolfenden) 等人，氨基酸侧链对溶剂水的亲和力 ("Affinities of Amino Acid Side Chains for Solvent Water")，生物化学 (*Biochemistry*) 20:849-855, 1981；卡特 (Kyte) 等人，呈现蛋白质水疗特性的简单方法 ("A Simple Method for Displaying the Hydropathic Character of a Protein")，分子生物学杂志 (*J. Mol. Biol.*) 157:105-132, 1982；罗丝 (Rose) 等人，球状蛋白中的氨基酸残基的疏水性 ("Hydrophobicity of Amino Acid Residues in Globular Proteins")，科学 (*Science*) 229:834-838, 1985；考纳特 (Cornette) 等人，检测蛋白质中的两亲结构的疏水性标度和计算技术 ("Hydrophobicity Scales and Computational Techniques for Detecting Amphipathic Structures in Proteins")，分子生物学杂志 (*J. Mol. Biol.*) 195:659-685, 1987；查顿 (Charton) 和查顿 (Charton)，氨基酸疏水性参数的结构依赖性 ("The Structure Dependence of Amino Acid Hydrophobicity Parameters")，理论生物学杂志 (*J. Theor. Biol.*) 99:629-644, 1982；各自以引用的方式并入本文中。本发明方法中可使用任何这些疏水性参数来确定修饰哪些非保守残基。在某些实施例中，鉴别亲水性或带电残基来修饰。

然后选择至少一个所鉴别的非保守或非重要表面残基加以修饰。在某些实施例中，选择疏水性残基加以修饰。在其它实施例中，选择亲水性和/或带电残基加以修饰。在某些实施例中，选择不止一个残基加以修饰。在某些实施例中，选择 1、2、3、4、5、6、7、8、9 或 10 个所鉴别的残基加以修饰。在某些实施例中，选择超过 10 个、超过 15 个或超过 20 个残基加以修饰。所属领域的技术人员应了解，蛋白质越大，则需要修饰的残基将越多。同时，蛋白质疏水性越高或者越易聚集或沉淀，则需要修饰的残基将越多。在某些实施例中，产生多个各具有不同修饰的蛋白质变异体并加以测试以确定就生物活性和稳定性来说最佳的变异体。

在某些实施例中，使选择加以修饰的残基突变为更亲水的残基（包括带电残基）。通常使残基突变为更亲水的天然氨基酸。在某些实施例中，使残基突变为在生理 pH 值下带电的氨基酸。举例来说，残基可变为精氨酸、天冬氨酸、谷氨酸、组氨酸或赖氨酸。在某些实施例中，使待修饰的所有残基变为同一种不同残基。举例来说，使所选择的残基全部变为谷氨酸残基。在其它实施例中，使所选择的残基变为不同残基；然而，可使所有最终残基在生理 pH 值下带正电或带负电。在某些实施例中，为产生带负电的蛋白质，使待突变的所有残基转化为谷氨酸和/或天冬氨酸残基。在某些实施例中，为产生带正电的蛋白质，使待突变的所有残基转化为赖氨酸残基。举例来说，所有所选择的待修饰残基是天冬酰胺、谷氨酰胺、赖氨酸和/或精氨酸，并且使这些残基突变为天冬氨酸或谷氨酸残基。作为另一实例，所有所选择的待修饰残基是天冬氨酸、谷氨酸、天冬酰胺和/或谷氨酰胺，并且使这些残基突变为赖氨酸。这种方法允许在最大程度上修饰蛋白质上的净电荷。

在其它实施例中，蛋白质可在修饰后在修饰蛋白质上保持与未修饰蛋白质上相同的净电荷。在其它实施例中，蛋白质可在修饰后减少蛋白质上的总净电荷，而增加表面上带电残基的总数。在某些实施例中，理论净电荷增加至少+1、+2、+3、+4、+5、+10、+15、+20、+25、+30 或+35。在某些实施例中，理论净电荷减少至少-1、-2、-3、-4、-5、-10、-15、-20、-25、-30 或-35。在某些实施例中，使所选择的氨基酸变为非离子极性残基（例如，半胱氨酸、丝氨酸、苏氨酸、酪氨酸、谷氨酰胺、天冬酰胺）。

蛋白质中的这些修饰或突变可使用所属领域中已知的任何技术来实现。所属领域中熟知在蛋白质序列中引入这些变化的重组 DNA 技术。在某些实施例中，通过定点诱变编码所述蛋白质的多核苷酸进行修饰。引入突变的其它技术论述于以下文献中：分子克隆：实验室手册（*Molecular Cloning: A Laboratory Manual*），第 2 版，山姆布鲁克（Sambrook），福里茨（Fritsch）以及曼尼亚提斯（Maniatis）编，（冷泉港实验室出版社（Cold Spring Harbor Laboratory Press）：1989）；论文酶学中的方法（*Methods in Enzymology*）（学术出版社（Academic Press, Inc.），纽约（N.Y.））；奥苏贝尔（Ausubel）等人，最新分子生物学实验方法汇编（*Current Protocols in Molecular Biology*）（约翰威利父子出版社（John Wiley & Sons, Inc.），纽约（New York），1999）；各自以引用的方式并入本文中。表达并测试所述修饰蛋白质。在某些实施例中，制备一系列变异体并测试各变异体确定其生物活性和其稳定性。选择用于后续使用的变异体可为最稳定的变异体、最具活性的变异体或者活性和稳定性的组合总体最大的变异体。制备第一组变异体后，可以基于从第一组变异体学到的知识制备另一组变异体。通常使用所属领域中已知

的重组技术产生和过度表达变异体。

已使用本发明系统产生了 GFP 变异体。已显示这些变异体更加稳定并且保留其荧光性。来自维多利亚水母 (*Aequorea victoria*) 的 GFP 描述于基因库寄存编号 P42212 中, 其以引用的方式并入本文中。这一野生型 GFP 的氨基酸序列如下:

```
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPW
PTLVTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVK
FEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNI
EDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDHMLLEFVTAAG
GITHGMDELYK (SEQ ID NO: 1)
```

野生型 GFP 的理论净电荷为-7。使用本发明系统已产生理论净电荷为-29、-30、-25、+36、+48 以及+49 的变异体。即使加热+36 GFP 到 95℃后, 100%的变异蛋白质也是可溶的并且蛋白质保留 $\geq 70\%$ 的荧光性。

已产生的 GFP 变异体的氨基酸序列包括:

GFP-NEG25

```
MGHHHHHHGGASKGEELFTGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKF
ICTTGELPVPWPVLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHDVYITADKQENGI
KAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLSTESALSCKDPNEDRDHM
VLEFVTAAGIDHGMDLYK (SEQ ID NO: 2)
```

GFP-NEG29

```
MGHHHHHHGGASKGEELFDGEVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKF
ICTTGELPVPWPVLVTTLTYGVQCFSRYPDHMDQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHDVYITADKQENGI
KAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLSTESALSCKDPNEDRDHM
VLEFVTAAGIDHGMDLYK (SEQ ID NO: 3)
```

GFP-NEG30

```
MGHHHHHHGGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKF
ICTTGELPVPWPVLVTTLTYGVQCFSRYPDHMDQHDFFKSAMPEGYVQERTISFKDD
GTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHDVYITADKQENGI
KAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLSTESALSCKDPNEDRDHM
VLEFVTAAGIDHGMDLYK (SEQ ID NO: 4)
```

GFP-POS36

```
MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKKGKGDATRGKLTCLK
FICTTGKLPVPWPVLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKK
DGKYKTRAEVKFEGRTLNRILKGRDFKEKGNILGHKLRYNNSHKVYITADKRK
NGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKR
DHMLLEFVTAAGIKHGRDERYK (SEQ ID NO: 5)
```

GFP-POS42

MGHHHHHHHGGRSKGKRLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGLTLK
FICTTGKLPVPWPTLVTTLTYGVCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKK
DGKYKTRAENVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRR
NGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRKHLYLSTRSKLSKDPKEKR
DHMVLLFVTAAGIKHGRKERYK (SEQ ID NO: 6)

GFP-POS49

MGHHHHHHHGGRSKGKRLFRGKVPILVKLKGDVNGHKFSVRGKGKGDATRGLTLK
FICTTGKLPVPWPTLVTTLTYGVCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKK
DGKYKTRAENVKFKGRTL VNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRR
NGIKAKFKIRHNVKDGSVQLAKHYQQNTPIGRGPVLLPRKHLYLSTRSKLSKDPKEKR
DHMVLEFVTAAGIKHGRKERYK (SEQ ID NO: 7)

所属领域的技术人员应了解，同源蛋白质同样视为在本发明范围内。举例来说，任何包括一段 20、30、40、50 或 100 个氨基酸与任何上述序列 60%、70%、80%、90%、95%或 100%同源的蛋白质视为本发明的一部分。另外，本发明还涵盖添加和缺失变异体。在某些实施例中，具有如任何上述序列中所示的突变残基的任何 GFP 视为本发明的一部分。在某些实施例中，序列包括 2、3、4、5、6、7、8、9、10 个或 10 个以上如任何以上序列中所示的突变。

编码上述 GFP 变异体的任何 DNA 序列也包括在本发明范围内。编码以上各变异体的示范性 DNA 序列如下：

GFP-NEG25

ATGGGGCATCACCATCATCATCATGGCGGTGCGTCTAAGGGGGAGGAGTTATTTA
CGGGTGTGGTGCCGATCCTGGTGGAGCTTGATGGCGATGTTAACGGCCATGAATT
TTCTGTCCGCGGTGAAGGGGAGGGTGATGCCACGGAAGGGGAGCTGACACTTAA
ATTTATTTGCACCACCGGTGAACTCCCGGTCCCGTGGCCGACCCTGGTGACCACC
CTGACCTACGGCGTTCAATGCTTTTCACGTTATCCGGATCACATGAAGCAACACG
ACTTCTTTAAAAGCGCGATGCCTGAAGGCTATGTTCAAGAACGTACAATTAGTTT
TAAAGATGACGGCACCTACAAGACCCGTGCGGAAGTAAAATTTGAAGGGGACAC
TTTAGTGAAACCGCATCGAGCTGAAAGGGATCGATTTTAAAGAAGATGGGAATAT
CCTGGGACACAACTTGAATACAACTTTAATAGTCATGACGTCTATATCACGGCG
GACAAACAGGAAAACGGAATTAAGGCAGAATTTGAGATTCGGCATAATGTGCGAA
GATGGCTCGGTACAGTTGGCTGATCACTATCAGCAGAATACGCCGATTGGAGAT
GGTCCGGTTTTATTACCAGACGATCACTATCTGTCCACCGAATCCGCCCTGAGCA
AAGATCCGAATGAAGACCGGGACCATATGGTTCTGCTGGAATTTGTTACGGCGG
CTGGTATTGACCATGGCATGGATGAGCTGTATAAGTAG (SEQ ID NO: 8)

GFP-NEG29

ATGGGGCATCACCATCATCATCATGGCGGTGCGTCTAAGGGGGAGGAGTTATTTG
ATGGTGAAGTGCCGATCCTGGTGGAGCTTGATGGCGATGTAAACGGCCATGAATT
TTCTGTCCGCGGTGAAGGGGAGGGTGATGCCACGGAAGGGGAGCTGACACTTAA
ATTTATTTGCACCACCGGTGAACTCCCGGTCCCGTGGCCGACCCTGGTGACCACC
CTGACCTACGGCGTTCAATGCTTTTCACGTTATCCGGATCACATGGACCAACACG
ACTTCTTTAAAAGCGCGATGCCTGAAGGCTATGTTCAAGAACGTACAATTAGTTT
TAAAGATGACGGCACCTACAAGACCCGTGCGGAAGTAAAATTTGAAGGGGACAC
TTTAGTGAACCGCATCGAGCTGAAAGGGATCGATTTTAAAGAAGATGGGAATAT
CCTGGGACACAACTTGAATACAACCTTTAATAGTCATGACGTCTATATCACGGCG
GACAAACAGGAAAACGGAATTAAGGCAGAAATTTGAGATTTCGGCATAATGTCGAA
GATGGCTCGGTACAGTTGGCTGATCACTATCAGCAGAATACGCCGATTGGAGAT
GGTCCGGTTTTATTACCAGACGATCACTATCTGTCCACCGAATCCGCCCTGAGCA
AAGATCCGAATGAAGACCGGGACCATATGGTTCTGCTGGAATTTGTTACGGCGG
CTGGTATTGACCATGGCATGGATGAGCTGTATAAGTAG (SEQ ID NO: 9)

GFP-NEG30

ATGGGGCATCACCATCATCATCATGGCGGTGCGTCTAAGGGGGAGGAGTTATTTG
ATGGTGTGGTGCCGATCCTGGTGGAGCTTGATGGCGATGTAAACGGCCATGAATT
TTCTGTCCGCGGTGAAGGGGAGGGTGATGCCACGGAAGGGGAGCTGACACTTAA
ATTTATTTGCACCACCGGTGAACTCCCGGTCCCGTGGCCGACCCTGGTGACCACC
CTGACCTACGGCGTTCAATGCTTTTCAGATTATCCGGATCACATGGACCAACACG
ACTTCTTTAAAAGCGCGATGCCTGAAGGCTATGTTCAAGAACGTACAATTAGTTT
TAAAGATGACGGCACCTACAAGACCCGTGCGGAAGTAAAATTTGAAGGGGACAC
TTTAGTGAACCGCATCGAGCTGAAAGGGATCGATTTTAAAGAAGATGGGAATAT
CCTGGGACACAACTTGAATACAACCTTTAATAGTCATGACGTCTATATCACGGCG
GACAAACAGGAAAACGGAATTAAGGCAGAAATTTGAGATTTCGGCATAATGTCGAA
GATGGCTCGGTACAGTTGGCTGATCACTATCAGCAGAATACGCCGATTGGAGAT
GGTCCGGTTTTATTACCAGACGATCACTATCTGTCCACCGAATCCGCCCTGAGCA
AAGATCCGAATGAAGACCGGGACCATATGGTTCTGCTGGAATTTGTTACGGCGG
CTGGTATTGACCATGGCATGGATGAGCTGTATAAGTAG (SEQ ID NO: 10)

GFP-POS36

ATGGGGCATCATCATCATCACCACGGCGGGGCGTCTAAGGGAGAGCGCTTGTTTC
GCGGCAAAGTCCCGATTCTTGTGGAGCTCAAAGGTGATGTAAATGGTCATAAATT
TAGTGTGCGCGGGAAAGGGAAAGGAGATGCTACGCGGGGCAAGCTCACCCTGAA
ATTTATTTGCACAACCGGCAAACCTGCCAGTGCCGTGGCCTACATTAGTCACTACT
CTGACGTACGGTGTTCAAGTCTTTCTCGCTATCCCAAACACATGAAACGCCATG
ATTTCTTCAAGAGCGCGATGCCAAAAGGTTATGTGCAGGAACGCACCATCAGCTT
TAAAAAAGACGGCAAATATAAAACCCGTGCAGAAAGTTAAATTCGAAGGCCGCAC
CCTGGTCAACCGCATTAACCTGAAAGGTTCGTGACTTCAAAGAGAAAGGTAATAT
TCTTGCTCACAACCTGCGCTATAATTTCAACTCTCACAAGTTTATATTACGGCG
GATAAACGTAAAAACGGGATTAAGCGAAATTTAAGATTTCGTCATAATGTTAAA
GACGGCAGTGTGCAGTTAGCGGATCATTATCAGCAGAATACCCCAATTGGTCGC
GGTCCAGTGCTGCTGCCGCGTAACCATTAATCTGTCGACCCGCAGCAAACCTCAGCA
AAGACCCGAAAGAAAAACGTGACCACATGGTATTACTGGAATTTGTGACCGCAG
CAGGCATTAAACATGGCCGCGATGAACGTTACAAATAG (SEQ ID NO: 11)

GFP-POS44

ATGGGCCATCATCATCACCACCACGGCGGCCGCTCAAAAGGTAAACGCTTGTTCC
GTGGTAAAGTACCGATCTTAGTGGAGCTCAAAGGGGATGTGAATGGCCATAAGT
TCTCGGTTTCGTGGCAAAGGTAAGGGAGATGCGACGCGCGGCAAATTAACGCTGA
AATTCATTTGTACTACAGGTAAACTGCCGGTGCCATGGCCTACTCTCGTCACCAC
GTTGACCTATGGGGTTCAATGCTTCAGCCGGTACCCTAAACACATGAAGCGCCAC
GATTTCTTCAAATCGGCGATGCCAAAGGGGTATGTCCAGGAACGCACTATCAGCT
TCAAAAAAGACGGTAAGTATAAACTCGTGCTGAAGTTAAATTGGAAGGACGCA
CACTGGTAAATCGCATTAAATTGAAGGGGCGGACTTTAAGGAAAAAGGTAATA
TCTTAGGTCACAAATTGCGCTACAACTTCAACTCTCATAAAGTTTACATTACAGC
AGATAAGCGTAAAAATGGCATCAAAGCGAAATTCAAAATTCGTCACAATGTGAA
AGATGGTAGCGTGCAATTAGCCGATCATTACCAGCAGAATACGCCGATCGGTGCG
CGGCCAGTACTGTTGCCGCGCAAACATTACTTATCTACCCGGAGTAAACTGTCT
AAAGACCCAAAAGAGAAGCGCGACCATATGGTTCTCCTGGAGTTTGTCAACCGCC
GCCGGAATTAAACACGGCCGCAAAGAGCGCTATAAATAG (SEQ ID NO: 12)

GFP-POS49

ATGGGCCACCATCATCATCACCACGGGGGACGCTCTAAAGGTAAACGTCTGTTTC
GTGGAAAGGTGCCCATTTCTGGTTAAACTCAAAGGTGATGTCAACGGCCATAAGTT
TTCGGTTTCGTGGCAAAGGTAAGGTGATGCGACGCGCGGGAAATTAACACTGAA
ATTTATTTGCACAACCGGAAAACTCCCTGTGCCGTGGCCGACTTTGGTGACCACA
TTAACCTATGGTGTTCAATGCTTCTCACGTTATCCGAAGCATATGAAACGTCATG
ATTTTTTCAAATCGGCTATGCCGAAAGGTTACGTCCAGGAGCGCACCATCTCATT
TAAGAAAGACGGTAAGTATAAAACCCGTGCTGAAGTAAAATTCAAAGGACGCAC
CCTGGTGAATCGCATTAAACTGAAAGGTCGTGATTTCAAAGAAAAGGGAAATAT
TTTAGGGCATAAGCTCCGTTATAATTTTAACAGTCATAAGGTGTATATTACCGCT
GATAAACGCAAAAACGGAATCAAAGCGAAATTTAAGATCCGTCATAATGTAAAA
GATGGCTCAGTCCAACCTGGCAAACATTACCAGCAGAATACCCCGATCGGCCGC
GGTCCTGTGCTTCTGCCGCGTAAACACTACTTGTGACCCGGTCAAATTTGAGTA

AAGATCCGAAGGAAAAGCGTGATCACATGGTCTTGAAGGAATTTGTAACGTGCAG
CAGGTATTAAACACGGGCGCAAAGAACGTTACAAATAG (SEQ ID NO: 13)

与上述序列同源的多核苷酸序列也在本发明范围内。在某些实施例中，多核苷酸序列包括与任何一个上述序列 60%、70%、80%、90%、95%、98%、99%或 100%同源的一段 50、100 或 150 个核苷酸。本发明也包括一个上述序列中插入或缺失一个或一个以上核苷酸的序列。任何具有任何上述序列中所示的突变的多核苷酸序列都视为本发明的一部分。在某些实施例中，序列包括 2、3、4、5、6、7、8、9、10 个或 10 个以上如任何以上序列中所示的突变。

本发明也提供包含使用本发明系统修饰的本文中的任何本发明序列或任何其它序列（DNA 或蛋白质）的载体（例如，质体、粘粒、病毒等）。在某些实施例中，载体包括适用于在细胞中过度表达本发明 GFP 变异体的元件（诸如启动子、增强子、核糖体结合位点等）序列。本发明也包括包含本发明序列或载体的细胞。在某些实施例中，细胞过度表达变异 GFP。细胞可为细菌细胞（例如，大肠杆菌）、真菌细胞（例如，巴斯德毕赤酵母（*P. pastoris*））、酵母细胞（例如，酿酒酵母）、哺乳动物细胞（例如，CHO 细胞）或人类细胞。

已使用本发明系统产生了链霉亲和素变异体。已显示这些变异体形成结合生物素的可溶性四聚体。这一野生型链霉亲和素的氨基酸序列如下：

AAEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATD
GSGTALGWTVAWKNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWK
STLVGHDTFTKVKPSAAS (SEQ ID NO: XX)

野生型链霉亲和素的理论净电荷为-4。使用本发明系统已产生理论净电荷为-40 和 +52 的变异体。即使加热变异体到 100℃后，蛋白质仍可溶。

已产生的链霉亲和素变异体的氨基酸序列包括：

SAV-NEG40

MGHHHHHHGGAEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGDAESEYVLT
GRYDSAPATDGSGTALGWTVAWKNKYRNAHSATTWSGQYVGGAEARINTQWLLT
SGTTEADAWKSTLVGHDTFTKVEPSAAS (SEQ ID NO: XX)

SAV-POS52

MGHHHHHHGGAKAGITGTWYNQLGSTFIVTAGAKGALTGTYESAVGNAKSRYVLT
GRYDSAPATKSGTALGWTVAWKNKYRNAHSATTWSGQYVGGAKARINTQWLLT
SGTTKAKAWKSTLVGHDTFTKVKPSAAS (SEQ ID NO: XX)

所属领域的技术人员应了解，同源蛋白质同样视为在本发明范围内。举例来说，任何包括一段 20、30、40、50 或 100 个氨基酸与任何上述序列 60%、70%、80%、90%、95%或 100%同源的蛋白质视为本发明的一部分。另外，本发明还涵盖添加和缺失变异体。在某些实施例中，具有如任何上述序列中所示的突变残基的任何链霉亲和素视为本发明的一部分。在某些实施例中，序列包括 2、3、4、5、6、7、8、9、10 个或 10 个以上如任何以上序列中所示的突变。

编码上述链霉亲和素变异体的任何 DNA 序列也包括在本发明范围内。编码上述各变异体的示范性 DNA 序列如下：

SAV-NEG40

GGTTCAGCCATGGGTCATCACCACCACCATCACGGTGGCGCCGAAGCAGGTATT
ACCGGTACCTGGTATAACCAGTTAGGCTCAACCTTTATTGTGACCGCGGGAGCGG
ACGGCGCCTTAACCGGTACCTACGAATCAGCTGTAGGTGACGCGGAATCAGAGT
ACGTATTAACCGGTCGTTATGATAGCGCGCCGGCGACTGACGGTAGCGGTACTGC
TTTAGGTTGGACCGTAGCGTGGAAGAATGATTATGAAAACGCACATAGCGCAAC
AACGTGGTCAGGGCAGTACGTTGGCGGAGCTGAGGCGCGCATTAACACGCAGTG
GTTATTAAGTACGGCACCCTGAAGCTGATGCCTGGAAGAGCACGTTAGTGGG
TCATGATACCTTCACTAAAGTGGAACCTTCAGCTGCGTCATAATAATGACTCGAG
ACCTGCA (SEQ ID NO: XX)

SAV-POS52

GGTTCAGCCATGGGTCATCACCACCACCATCACGGTGGCGCCAAAGCAGGTATT
ACCGGTACCTGGTATAACCAGTTAGGCTCAACCTTTATTGTGACCGCGGGAGCGA
AAGGCGCCTTAACCGGTACCTACGAATCAGCTGTAGGAAACGCAAAATCACGCT
ACGTATTAACCGGTCGTTATGATAGCGCGCCGGCGACTAAAGGTAGCGGTACTG
CTTTAGGTTGGACCGTAGCGTGGAAGAATAAGTATCGTAATGCGCACAGTGCTAC
CACTTGGTCAGGGCAGTACGTAGGGGGAGCCAAAGCACGTATCAACACGCAGTG

GTTATTAACATCAGGTACCACCAAAGCGAAAGCCTGGAAGAGCACGTTAGTGGG
TCATGATACCTTCACTAAAGTGAAACCTTCAGCTGCGTCATAATAATGACTCGAG
ACCTGCA (SEQ ID NO: XX)

与上述序列同源的多核苷酸序列也在本发明范围内。在某些实施例中，多核苷酸序列包括与任何一个上述序列 60%、70%、80%、90%、95%、98%、99%或 100%同源的一段 50、100 或 150 个核苷酸。本发明也包括一个上述序列中插入或缺失一个或一个以上核苷酸的序列。任何具有任何上述序列中所示的突变的多核苷酸序列都视为本发明的一部分。在某些实施例中，序列包括 2、3、4、5、6、7、8、9、10 个或 10 个以上如任何以上序列中所示的突变。

本发明也提供包含使用本发明系统修饰的本文中的任何本发明序列或任何其它序列（DNA 或蛋白质）的载体（例如，质体、粘粒、病毒等）。在某些实施例中，载体包括适用于在细胞中过度表达本发明链霉亲和素变异体的元件（诸如启动子、增强子、核糖体结合位点等）序列。本发明也包括包含本发明序列或载体的细胞。在某些实施例中，细胞过度表达变异链霉亲和素。细胞可为细菌细胞（例如，大肠杆菌）、真菌细胞（例如，巴斯德毕赤酵母）、酵母细胞（例如，酿酒酵母）、哺乳动物细胞（例如，CHO 细胞）或人类细胞。

已使用本发明系统产生了谷胱甘肽-S-转移酶（glutathione-S-transferase, GST）的变异体。已显示这些变异体保留野生型 GST 的催化活性。这一野生型 GST 的氨基酸序列如下：

MGHHHHHHGGPPYTITYFPVGRCEAMRMLLADQDQSWKEEVVTMETWPPLKPSC
LFRQLPKFQDGLTLYQSNAILRHLGRSFGLYGKDQKEAALVDMVNDGVEDLRCKY
ATLIYTNYEAGKEKYVKELPEHLKPFETLLSQNQGGQAFVVGSIQSFADYNLLDLLRI
HQVLNPSCLDAFPLLSAYVARLSARPKIKAFSLASPEHVNRPINGNGKQ (SEQ ID NO:
XX)

野生型 GST 的理论净电荷为+2。使用本发明系统已产生理论净电荷为-40 的变异体。这一变异体催化谷胱甘肽添加至二硝基氯苯，其比活性仅为野生型 GST 的 1/2.7。即使加热变异体到 100℃后，蛋白质仍可溶，并且经冷却时蛋白质恢复其 40%的催化活性。

GST 变异体的氨基酸序列包括：

GST-NEG40

MGHHHHHHGGPPYTITYFPVRGRCEAMRMLLADQDQSWEEEVVTMETWPPLKPSC
LFRQLPKFQDGLTLYQSNAILRHLGRSFGLYGEDEEEAALVDMVNDGVEDLRCKY
ATLIYTDYEAGKEEYVEELPEHLKPFETLLSENEGGEAFVVGSEISFADYNLLDLLRIH
QVLNPSCLDAFPLLSAYVARLSARPEIEAFLASPEHVDRPINGNGKQ (SEQ ID NO:
XX)

GST-POS50

MGHHHHHHGGPPYTITYFPVRGRCEAMRMLLADQKQSWKEEVVTMKTWPPLKPSC
LFRQLPKFQDGKLTLYQSNAILRHLGRSFGLYGKKQKEAALVDMVNDGVEDLRCKY
ATLIYTKYKAGKKKYVKKLPKHLKPFETLLSKNKGKAFVVGSKISFADYNLLDLLR
IHQVLNPSCLKAFPLLSAYVARLSARPKIKAFASPEHVKRPINGNGKQ (SEQ ID NO:
XX)

所属领域的技术人员应了解，同源蛋白质同样视为在本发明范围内。举例来说，任何包括一段 20、30、40、50 或 100 个氨基酸与任何上述序列 60%、70%、80%、90%、95% 或 100% 同源的蛋白质视为本发明的一部分。另外，本发明还涵盖添加和缺失变异体。在某些实施例中，具有如任何上述序列中所示的突变残基的任何链霉亲和素视为本发明的一部分。在某些实施例中，序列包括 2、3、4、5、6、7、8、9、10 个或 10 个以上如任何以上序列中所示的突变。

编码上述 GST 变异体的任何 DNA 序列也包括在本发明范围内。编码上述各变异体的示范性 DNA 序列如下：

GST-NEG40

GGTTCAGCCATGGGTCATCACCACCACCATCACGGTGGCCCGCCGTACACCATTA
CATACTTTCCGGTACGTGGTCGTTGTGAAGCGATGCGTATGTTATTAGCGGACCA
GGACCAATCATGGGAAGAAGAAGTAGTGACAATGGAAACCTGGCCGCGCTTAAA
GCCTAGCTGTTTATTCCGTCAATTACCGAAGTTTCAGGATGGTGATTTAACCTTAT
ACCAGTCTAACGCGATCTTACGTCATTTAGGTCGCTCATTTGGTTTATACGGTGA
AGATGAAGAAGAAGCAGCCTTAGTGGATATGGTGAATGATGGCGTGGAAGACTT
ACGTTGTAAATACGCGACGTTAATTTACACTGATTATGAAGCCGGTAAAGAGGA
GTACGTGGAAGAATTACCTGAACACCTGAAGCCGTTTGAAACATTACTGAGCGA
AAATGAAGGAGGTGAGGCGTTCGTAGTTGGTAGCGAAATTAGCTTCGCTGATTAT

AACTTATTAGACTTATTACGCATTCACCAGGTTTTAAATCCTAGCTGTTTAGACGC
TTTCCCGTTACTGAGCGCATATGTAGCGCGCCTGAGCGCCCGTCCGGAATTGAA
GCTTTCTTAGCGTCACCTGAACACGTAGACCGCCCGATTAAACGGAACGGCAAG
CAGTAATAATGAGGTACCACCTGCA (SEQ ID NO: XX)

GST-POS50

GGTTCAGCCATGGGTCATCACCACCACCATCACGGTGGCCCGCCGTACACCATTA
CATACTTTCCGGTACGTGGTTCGTTGTGAAGCGATGCGTATGTTATTAGCGGACCA
GAAACAATCATGGAAAGAAGAAGTAGTGACAATGAAGACCTGGCCGCCGTAA
GCCTAGCTGTTTATTCCGTCAATTACCGAAGTTTCAGGATGGTAAATTAACCTTAT
ACCACTCTAACGCGATCTTACGTCATTTAGGTCGCTCATTTGGTTTATACGGTAA
GAAGCAGAAAGAAGCAGCCTTAGTGGATATGGTGAATGATGGCGTGGAAGACTT
ACGTTGTAAATACGCGACGTTAATTTACACTAAATATAAAGCCGGTAAAAAGAA
GTACGTGAAAAAATTACCTAAACACCTGAAGCCGTTTGAAACATTACTGAGCAA
AAATAAAGGAGGTAAGGCGTTCGTAGTTGGTAGCAAGATTAGCTTCGCTGATTAT
AACTTATTAGACTTATTACGCATTCACCAGGTTTTAAATCCTAGCTGTTTAAAGGC
TTTCCCGTTACTGAGCGCATATGTAGCGCGCCTGAGCGCCCGTCCGAAGATCAAA
GCTTTCTTAGCGTCACCTGAACACGTGAAGCGCCCGATTAACGGAAACGGCAAG
CAGTAATAATGAGGTACCACCTGCA (SEQ ID NO: XX)

本发明也提供包含使用本发明系统修饰的本文中的任何本发明序列或任何其它序列 (DNA 或蛋白质) 的载体 (例如, 质体、粘粒、病毒等)。在某些实施例中, 载体包括适用于在细胞中过度表达本发明 GST 变异体的元件 (诸如启动子、增强子、核糖体结合位点等) 序列。本发明也包括包含本发明序列或载体的细胞。在某些实施例中, 细胞过度表达变异 GST。细胞可为细菌细胞 (例如, 大肠杆菌)、真菌细胞 (例如, 巴斯德毕赤酵母)、酵母细胞 (例如, 酿酒酵母)、哺乳动物细胞 (例如, CHO 细胞) 或人类细胞。

本发明也包括用于修饰所关注蛋白质以产生更稳定的蛋白质变异体的试剂盒。这些试剂盒通常包括产生蛋白质的更稳定变异体所需的所有或大部分试剂。在某些实施例中, 试剂盒包括辅助研究员基于本发明方法设计更稳定变异蛋白质的计算机软件。试剂盒还可包括以下所有或部分: 试剂、引物、寡核苷酸、核苷酸、酶、缓冲液、细胞、培养基、培养板、试管、说明书、载体等。使用本试剂盒的研究通常提供用于突变以产生更稳定变异体的 DNA 序列。通常包装这些内含物以便于在实验室中使用。

在考虑下文**实例**部分后, 将会进一步了解本发明的这些和其它方面, 这些实例旨在说明本发明的某些特定实施例, 但不打算限制如权利要求书所定义的本发明的范围。

实例

实例 1-使蛋白质超带电可赋予极大的恢复力 (Resilience)

蛋白质聚集, 是人类疾病一个众所周知的原因 (考汉 (Cohen, F. E.); 凯莉 (Kelly, J. W.), 自然 (*Nature*) 2003, 426, (6968), 905-9; 池蒂 (Chiti, F.); 多伯森 (Dobson, C. M.), 生物化学年鉴 (*Annu Rev Biochem*) 2006, 75, 333-66; 各自以引用的方式并入本文中), 也是使用蛋白质作为治疗或诊断剂所面临的主要问题 (福罗克杰 (Frokjaer, S.); 奥特珍 (Otzen, D. E.), 自然·评论: 药物发现 (*Nat Rev Drug Discov*) 2005, 4, (4), 298-306; 法沃勒 (Fowler, S. B.); 普恩 (Poon, S.); 墨佛 (Muff, R.); 池蒂 (Chiti, F.); 多伯

森 (Dobson, C. M.); 佐多 (Zurdo, J.), 美国国家科学院院刊 (*Proc Natl Acad Sci USA*) 2005, 102, (29), 10105-10; 各自以引用的方式并入本文中)。已经从对天然蛋白质的研究中获得对蛋白质聚集问题的了解。我们已经知道, 蛋白质在携带零净电荷的等电点时可溶性最小 (李伯 (Loeb, J.), 普通生理学杂志 (*J Gen Physiol*) 1921, 4, 547-555, 以引用的方式并入本文中)。新近, 已显示净电荷的微小差异 (± 3 电荷单位) 可以预测球状蛋白变异体间 (池蒂 (Chiti, F.); 斯泰法尼 (Stefani, M.); 泰戴 (Taddei, N.); 雷姆普尼 (Ramponi, G.); 多伯森 (Dobson, C. M.), 自然 (*Nature*) 2003, 424, (6950), 805-8, 以引用的方式并入本文中) 以及固有无序肽间 (帕瓦 (Pawar, A. P.); 督拜 (Dubay, K. F.); 佐多 (Zurdo, J.); 池蒂 (Chiti, F.); 文德斯可罗 (Vendruscolo, M.); 多伯森 (Dobson, C. M.), 分子生物学杂志 (*J Mol Biol*) 2005, 350, (2), 379-92, 以引用的方式并入本文中) 的聚集倾向。这些观察结果连同一些蛋白质可忍受净电荷显著变化的最近证据 (例如, 发现碳酸酐酶在其表面赖氨酸彻底化学乙酰化后保留催化活性 (古迪克森 (Gudiksen) 等人, 美国化学会志 (*J Am Chem Soc*) 2005, 127, (13), 4707-14, 以引用的方式并入本文中)) 一起, 使我们推断通过使表面大范围突变以显著增加净电荷 (我们在本文中称之为“超带电”的过程) 可能会明显增强一些蛋白质的溶解性和抗聚集性而不消除其折叠或功能。

我们以最近报导的在折叠效率和变性剂抗性方面已高度优化的绿色荧光蛋白 (GFP) 的现有技术变异体 (称为“超折叠 GFP” (sfGFP)) (派德拉克 (Pedelacq) 等人, 国家生物技术 (*Nat Biotechnol*) 2006, 24, (1), 79-88, 以引用的方式并入本文中) 开始。超折叠 GFP 的净电荷为 -7, 与野生型 GFP 类似。根据计算氨基酸的溶剂暴露的简单算法 (参见“材料和方法”部分), 我们通过使 29 个最溶剂暴露的 (most solvent-exposed) 残基突变为带正电的氨基酸而设计出理论净电荷为 +36 的 GFP 超带电变异体 (图 1)。编码 sfGFP 或 GFP(+36) 的基因的表达浓密地产生绿色荧光菌。蛋白质纯化后, 测量 GFP(+36) 的荧光性质并且发现与 sfGFP 的荧光性质非常类似。受这一发现鼓励, 我们设计和纯化了净电荷为 +48、-25 以及 -30 的其它超带电 GFP, 同样发现全部显示类似 sfGFP 的荧光性 (图 2a)。所有超带电 GFP 变异体显示类似于 sfGFP 的圆二色性谱, 表明蛋白质具有类似的二级结构内容 (图 2b)。超带电 GFP 变异体虽然存在多达 36 处突变, 但其热力学稳定性仅比 sfGFP 稍低 (1.0-4.1 kcal/mol, 图 2c 和表 1)。

虽然 sfGFP 是 GFP 优化的历史悠久的产物 (吉普曼斯 (Giepmans) 等人, 科学 (*Science*) 2006, 312, (5771), 217-24; 以引用的方式并入本文中), 但其仍容易受到热或化学解折叠诱导而聚集。加热 sfGFP 到 100°C 诱导其定量沉淀和荧光性不可逆丧失 (图

3a)。相反,超带电 GFP(+36)和 GFP(-30)当加热到 100℃时仍可溶,而冷却后恢复明显的荧光性(图 3a)。重要的是,40% 2,2,2-三氟乙醇(TFE)在 25℃下几分钟内就诱导 sfGFP 完全聚集,而+36 和-30 超带电 GFP 变异体在相同条件下数小时都未经历明显的聚集或荧光性丧失(图 3b)。

除这一显著的抗聚集性之外,超带电 GFP 变异体显示对具有相反电荷的高度带电的大分子具有强大的可逆亲合力(图 3c)。当 GFP(+36)和 GFP(-30)以 1:1 化学计量混合在一起时立即形成绿色荧光共沉淀,表明折叠蛋白质缔合。GFP(+36)类似地与高浓度 RNA 或 DNA 共沉淀。添加 NaCl 足以溶解这些复合物,这一点与其形成的静电基础一致。相反, sfGFP 不受添加 GFP (-30)、RNA 或 DNA 影响(图 3c)。

我们接着需要确定超带电原理是否可能适用于除 GFP 以外的蛋白质,GFP 是单体并且具有屏蔽良好的荧光团。为此目的,我们对两种与 GFP 无关的蛋白质应用超带电过程。链霉亲和素是总净电荷为-4 的四聚体。使用溶剂暴露算法,我们设计出两种净电荷为-40 或+52 的超带电链霉亲和素变异体。两种超带电链霉亲和素变异体都能够形成结合生物素的可溶性四聚体,但亲和力减小。

使谷胱甘肽-S-转移酶(GST)(总净电荷为+2 的二聚体)超带电来产生净电荷为-40 的二聚体,其催化谷胱甘肽至二硝基氯苯的加成,比活性仅为野生型 GST 的 1/2.7(图 3d)。此外,超带电链霉亲和素和超带电 GST 在加热到 100℃时仍可溶,这与其野生型对应物相反,后者类似于 sfGFP 发生定量且不可逆的沉淀(表 1)。另外, GST(-40)冷却后恢复其催化活性的 40%(图 3d)。

总之,我们已经证明可通过简单地用带类似电荷的氨基酸置换大多数溶剂暴露的残基来使具有不同结构和功能的单体和多聚蛋白质“超带电”。超带电可大大改变蛋白质的分子间性质,赋予明显的抗聚集性和以折叠形式与带相反电荷的大分子(如“分子维可牢(molecular Velcro)”)缔合的能力。我们注意到这些不寻常的分子间性质是源于高净电荷,而不是源于带电氨基酸的总数,带电氨基酸的总数并不因超带电过程明显改变(表 1)。

与这些显著的分子间效应相反,此处研究的 7 种超带电蛋白质的分子内性质(包括折叠、荧光性、配体结合以及酶促催化)仍基本上完好。因此,超带电可能是一种降低聚集倾向和提高蛋白质溶解性而不消除其功能的有效方法。这些原理可能尤其适用于新颖(*de novo*)蛋白质的设计工作,其中包括聚集在内的不可预测的蛋白质处理性质仍是一项重大挑战。根据使天然蛋白质超带电的上述结果,人们不禁推测所设计的蛋白质的抗聚集性亦可能通过使设计过程偏向于增加带类似电荷的氨基酸在预测位于折叠蛋白

质外部的位置上的频率来提高。

蛋白质超带电说明蛋白质表面的显著可塑性并突出源于溶剂暴露残基的突变忍受性 (mutational tolerance) 的机会。举例来说, 最近显示一些蛋白质的热力学稳定性可通过合理地工程改造电荷-电荷相互作用来增强 (斯曲克勒 (Strickler) 等人, 生物化学 (Biochemistry) 2006, 45, (9), 2761-6, 以引用的方式并入本文中)。蛋白质超带电证明这种可塑性如何能以不同的方式加以利用从而赋予极强的蛋白质聚集抗性。我们的发现与补充研究的结果一致, 在补充研究中从泛素去除所有电荷, 其折叠性保持完好而其溶解性明显受损 (劳拉兹 (Loladze) 等人, 蛋白质科学 (Protein Sci) 2002, 11,(1), 174-7, 以引用的方式并入本文中)。

这些观察结果也可说明天然蛋白质的适度 (modest) 净电荷分布 (卡耐特 (Knight) 等人, 美国国家科学院院刊 (Proc Natl Acad Sci USA) 2004, 101, (22), 8390-5; 吉特林 (Gitlin) 等人, 德国应用化学杂志 (Angew Chem Int Ed Engl) 2006, 45, (19), 3022-60, 各自以引用的方式并入本文中): 例如, 84%的蛋白质数据库 (PDB) 多肽的净电荷在 ± 10 范围内。我们的结果与高净电荷产生足够的静电排斥从而迫使解折叠的假设相冲突。的确, GFP(+48)具有高于目前 PDB 中的任何多肽的正净电荷, 但仍保留折叠和发荧光的能力。实际上, 我们的发现提示非特异性分子间粘附可能不赞成太多高电荷天然蛋白质的演化。几乎所有具有极高净电荷的天然蛋白质 (诸如结合 RNA 的核糖体蛋白 L3(+36) 和 L15(+44)或结合钙阳离子的肌集钙蛋白(-80)) 都与带相反电荷的物质缔合, 作为其必需细胞功能的一部分。

材料和方法

设计程序和超带电的蛋白质序列: 从公开的结构数据鉴别 $AvNAPSA < 150$ (其中 $AvNAPSA$ 是每个侧链原子的平均相邻原子 (在 $10 \text{ \AA}$ 范围内)) 的溶剂暴露残基 (以下以灰色显示) (韦伯 (Weber, P.C.), 奥兰多夫 (Ohlendorf, D.H.), 文多罗斯奇 (Wendoloski, J.J.) 以及莎勒麦 (Salemme, F.R.), 与链霉亲和素结合的高亲和力生物素的结构起因 (Structural origins of high-affinity biotin binding to streptavidin). 科学 (Science) 243, 85-88 (1989); 迪尔 (Dirr, H.), 瑞纳默 (Reinemer, P.) 以及胡伯 (Huber, R.) $2.1 \text{ \AA}$ 分辨率下猪类 π 谷胱甘肽 S-转移酶 (pGST PI-1) 的细化晶体结构 (Refined crystal structure of porcine class Pi glutathione S-transferase (pGST PI-1) at $2.1 \text{ \AA}$ resolution), 分子生物学杂志 (J Mol Biol) 243, 72-92 (1994); 普德拉克 (Pedelacq, J.D.), 卡班特斯 (Cabantous, S.), 曲恩 (Tran, T.), 特威尔格 (Terwilliger, T.C.) 以及瓦尔多 (Waldo, G.S.) 超折叠绿色荧光蛋白的工程改造和表征 (Engineering and characterization of a superfolder

green fluorescent protein). 自然生物技术 (*Nat Biotechnol*) **24**, 79-88 (2006), 各自以引用的方式并入本文中)。突变带电或高极性的溶剂暴露残基 (DERKNQ): 对于超带负电 (红色), 突变为 Asp 或 Glu; 或对于超带正电 (蓝色), 突变为 Lys 或 Arg。绿色荧光蛋白 (GFP) 变异体中待突变的其它表面暴露位置根据 GFP 同源物中这些位置的序列变异性来选择。链霉亲和素 (SAV) 和谷胱甘肽-S-转移酶 (GST) 的超带电设计过程是全自动的: 首先通过溶剂暴露分选残基, 然后突变最溶剂暴露的带电或高极性残基: 对于超带正电, 突变为 Lys; 或对于超带负电, 突变为 Glu (除非开始残基是 Asn, 在此状况下突变为 Asp)。

```

1
SAV(-40)  MGHHHHHGGAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGDAESRYVLTGRYDSAPATDGSSTA
wtSAV      AARAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSSTA
SAV(+52)   MGHHHHHGGAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSSTA

71
SAV(-40)  LGWTVAWKNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVPKPSAAS
wtSAV     LGWTVAWKNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVPKPSAAS
SAV(+52)  LGWTVAWKNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVPKPSAAS

1
GST(-40)  MGHHHHHGGPPYITITYFVRGRCEAMRMLLADQDSWKEEVVTMTWPELKPSCLFROLPKFODGDLTYQSNA
wtGST     MGHHHHHGGPPYITITYFVRGRCEAMRMLLADQDSWKEEVVTMTWPELKPSCLFROLPKFODGDLTYQSNA
GST(+50)  MGHHHHHGGPPYITITYFVRGRCEAMRMLLADQDSWKEEVVTMTWPELKPSCLFROLPKFODGDLTYQSNA

75
GST(-40)  ILRHLGRSFGLYGDQKEAALVDMVNDGVEDLRCKYATLIYTDYEAGKEKYVDELPEHLKPFETLLSQNGGQAF
wtGST     ILRHLGRSFGLYGDQKEAALVDMVNDGVEDLRCKYATLIYTDYEAGKEKYVDELPEHLKPFETLLSQNGGQAF
GST(+50)  ILRHLGRSFGLYGDQKEAALVDMVNDGVEDLRCKYATLIYTDYEAGKEKYVDELPEHLKPFETLLSQNGGQAF

151
GST(-40)  VVGSQISFADYNLLDLRIHOVLNPSCLDAFPLLSAYVARLSARPKIKAFASPETHVNRPEINGNGKO
wtGST     VVGSQISFADYNLLDLRIHOVLNPSCLDAFPLLSAYVARLSARPKIKAFASPETHVNRPEINGNGKO
GST(+50)  VVGSQISFADYNLLDLRIHOVLNPSCLDAFPLLSAYVARLSARPKIKAFASPETHVNRPEINGNGKO

```

蛋白质的表达和纯化: 从 DNA 2.0 购得针对大肠杆菌密码子使用进行优化的合成基因, 克隆入 pET 表达载体 (诺瓦杰公司 (Novagen)) 并在 15°C 下在大肠杆菌 BL21 (DE3) pLysS 中过度表达 5-10 小时。通过离心收集细胞并通过超声波作用溶解。通过 Ni-NTA 琼脂糖色谱法 (凯杰公司 (Qiagen)) 纯化蛋白质, 将缓冲液换成 100 mM NaCl、50 mM 磷酸钾 (pH 7.5), 并通过超滤 (密理博公司 (Millipore)) 浓缩。在天然条件下纯化所有 GFP 变异体。从普洛麦格公司 (Promega) 购买野生型链霉亲和素。在变性条件下纯化超带电的链霉亲和素变异体并且如先前关于野生型链霉亲和素所报导进行重折叠 (赛姆普森 (Thompson) 等人, 合成链霉亲和素编码基因在大肠杆菌中的构建和表达 (Construction and expression of a synthetic streptavidin-encoding gene in *Escherichia coli*), 基因 (*Gene*) **136**, 243-246 (1993), 以引用的方式并入本文中), 超带电 GST 也是如此。

在天然或变性条件下纯化野生型 GST，得到具有可比活性的蛋白质。

表面静电势计算（图 1b）：-30 和+48 超带电 GFP 变异体的模型是基于超折叠 GFP 的晶体结构（普德拉克（Pedelacq）等人，超折叠绿色荧光蛋白的工程改造和表征（Engineering and characterization of a superfolder green fluorescent protein）. 自然生物技术（*Nat Biotechnol*） **24**, 79-88 (2006)，以引用的方式并入本文中）。使用-25 kT/e (红色)至+25kT/e (蓝色)的标度，使用 APBS（贝克（Baker）等人，纳米系统的静电学：对微管蛋白和核糖体的应用（Electrostatics of nanosystems: application to microtubules and the ribosome）. 美国国家科学院院刊（*Proc Natl Acad Sci USA*） **98**, 10037-10041 (2001)，以引用的方式并入本文中）计算静电势并用 PyMol（德拉诺（Delano, W.L.），PyMOL 分子制图系统（The PyMOL Molecular Graphics System），www.pymol.org (2002)，以引用的方式并入本文中）表现。

蛋白质染色和紫外光诱导的荧光性（图 2a）：在 10%变性聚丙烯酰胺凝胶中通过电泳分析 0.2 微克的各 GFP 变异体并用考马斯亮蓝（Coomassie brilliant blue）染料染色。在 0.2 mL 艾本德管（Eppendorf tube）中放入 0.2 微克于具有 100 mM NaCl 的 25 mM Tris（pH 8.0）中的相同蛋白质样品并在紫外光（360 nm）下拍照。

热变性和聚集（图 3a）：将纯 GFP 变异体在 25 mM Tris（pH 8.0）、100 mM NaCl 以及 10 mM β -巯基乙醇（BME）中稀释为 2 mg/mL，然后在紫外光照射下拍照（“天然”）。加热样品到 100℃持续 1 分钟，然后再次在紫外光照射下拍照（“煮沸”）。最后，在室温下冷却样品 2 小时并再次在紫外光照射下拍照（“冷却”）。

化学诱导的聚集（图 3b）：添加 2,2,2-三氟乙醇（TFE）以产生具有 1.5 mg/mL 蛋白质、25 mM Tris（pH 7.0）、10 mM BME 以及 40% TFE 的溶液。在 25℃下通过直角光散射监测聚集。

尺寸排阻色谱法（表 1）：通过在休珀代克斯 75（Superdex 75）凝胶过滤柱上分析 20-50 微克蛋白质来确定 SAV 和 GST 变异体的多聚状态。缓冲液为 100 mM NaCl、50 mM 磷酸钾（pH 7.5）。通过与一组在相同条件下单独分析的具有已知分子量的单体蛋白质标准物作比较确定分子量。

表 1：计算的和实验确定的蛋白质性质

名称	MW (kD)	长度 (aa)	n _正	n _负	n _{带电}	Q _净	pI	ΔG (kcal/mol) ^a	天然 MW(kD) ^b	煮沸后的可溶百分率 ^c
GFP(-30)	27.8	248	19	49	68	-30	4.8	10.2	n.d.	98

GFP(-25)	27.8	248	21	46	67	-25	5.0	n.d.	n.d.	n.d.
sfGFP	27.8	248	27	34	61	-7	6.6	11.2	n.d.	4
GFP(+36)	28.5	248	56	20	76	+36	10.4	8.8	n.d.	97
GFP(+48)	28.6	248	63	15	78	+48	10.8	7.1	n.d.	n.d.
SAV(-40)	14.3	137	5	15	20	-10	5.1	n.d.	55±5 (四聚体)	99
wtSAV	13.3	128	8	9	17	-1	6.5	n.d.	50±5 (四聚体)	7
SAV(+52)	14.5	137	16	3	19	+13	10.3	n.d.	55±5 (四聚体)	97
GST(-40)	24.7	217	17	37	54	-20	4.8	n.d.	50±5 (二聚体)	96
wtGST	24.6	217	24	23	47	+1	7.9	n.d.	50±5 (二聚体)	3
GST(+50) ^d	24.7	217	39	14	53	+25	10.0	n.d.	n.d.	n.d.

n_正为带正电的氨基酸的数目（每个单体）

n_负为带负电的氨基酸的数目

n_{带电}为带电氨基酸的总数

Q_净为中性 pH 值下的理论净电荷

pI 为计算等电点

n.d.表示未测定到

^a通过胍盐变性测量（图 2c）。

^b通过尺寸排阻色谱法测量。

^c在 100℃下 5 分钟，冷却到 25℃并短暂离心后上清液中所保留的蛋白质百分率。

^d未能在大肠杆菌中表达的蛋白质。

其它实施例

所属领域的一般技术人员应容易地了解前述内容仅代表本发明的某些优选实施例。

在不脱离随附权利要求书中所陈述的本发明精神或范围的情况下，可对上述程序和组合物进行多种改变和修正。

附录 A

```
#!/usr/local/bin/perl

#####
#
# avnapsa
#
# prints list of AvNAPSA values for the specified PDB
#
# Mike Lawrence/Kevin Phillips 3/17/2006
#
#####

sub show_usage
{
print "\n",

"Usage: avnapsa <start_pdb> [params]\n",
"  -3  use 3-letter aa abbreviations (default)\n",
"  -1  use 1-letter aa abbreviations\n",
"  -oncol  print one column only (i.e. only the AvNAPSA results)\n\n";
}

##### global variables #####

@atoms;
# fields loaded from PDB:
# type
# atomNum
# atomName
# resName
# chain
# resNum
# x, y, z
# computed fields
# neighborCount

@distances;

@residues;
# fields copied from PDB
# resNum (PDB numbering)
# resName
# computed fields
# avNapsa

#####

## parse command line

$use3or1 = 3;
$oncol_flag = 0;

$start_pdb = $ARGV[0];

for (my $a = 1; $a < @ARGV; ++$a)
{
    if ($ARGV[$a] eq "-1") { $use3or1 = 1; }
}
```

```

    elif ($ARGV[$a] eq "-3") { $use3or1 = 3; }
    elif ($ARGV[$a] eq "-onecol") { $onecol_flag = 1; }
    else { show_usage(); die "Invalid argument $ARGV[$a]\n"; }
}

unless (lc $start_pdb =~ /\.pdb/) { show_usage(); die "No starting pdb
specified.\n"; }

## read PDB and compute molecular parameters

read_PDB($start_pdb);
tabulate_residues();
$res = @residues;

compute_distances();
compute_neighbor_counts();
compute_residue_avNapsa();

print_residues();
exit;

#
# print_residues
#
#
sub print_residues
{
    for (my $r = 0; $r < @residues; $r++)
    {
        my $name = $residues[$r]{resName};
        $name = toggle31($name) if ($use3or1 == 1);
        printf "%d %s AvNAPSA ", $residues[$r]{resNum}, $name unless
$onecol_flag;
        printf "%.0f\n", $residues[$r]{avNapsa};
    }
    print "\nNum residues = ", $#residues+1, "\n\n" unless $onecol_flag;
}

#
# tabulate_residues
#
# goes through list of atoms and makes a list of amino acid residues
# and stores it in global variable @residues
#
sub tabulate_residues
{
    for ($a = 0; $a < @atoms; $a++)
    {
        $resNum = $atoms[$a]{resNum};
        if ( ! resNum_exists($resNum) )
        {
            push @residues,
            {
                resNum => $resNum,
            }
        }
    }
}

```

```

        resName => $atoms[$a]{resName}
    };
}
}

#
# resNum_exists
#
# returns 1 if resNum is contained in @residues
#

sub resNum_exists($)
{
    my ($resNum) = @_;

    for ($r = 0; $r < @residues; $r++)
    {
        return 1 if ($residues[$r]{resNum} == $resNum);
    }
    return 0;
}

#
# resNum_to_resindex
#
# converts PDB numbering to index in @residues
#

sub resNum_to_resindex($)
{
    my ($resNum) = @_;

    for ($r = 0; $r < @residues; $r++)
    {
        return $r if ($residues[$r]{resNum} == $resNum);
    }
    return "none";
}

#
# readPDB(filename)
#
# reads the atoms from a PDB and returns them as an array of hashes
#

sub read_PDB($)
{
    my ($filename) = @_;

    open (PDB, $filename) or die("Could not open $filename\n");

    $#atoms = -1;          # clear atoms storage

```

```

# read the file

foreach (<PDB>) {
    my $type = trim(substr($_, 0,6));          # RTyp field is columns 1-6
    next unless ($type eq "ATOM" || $type eq "HETATM");

    my $resName = trim(substr($_, 17, 3));      # Res field is columns 18-
20
    my $atomName = trim(substr($_, 12, 4));      # Atm field is columns
13-16

    next if uc $resName eq "HOH";                # omit waters
    next if uc $atomName =~ /^[0-9]*H/;          # omit protons

    # add a hash to the array, containing data from this record of the PDB

    push @atoms, {
        type => $type,
        resName => $resName,
        atomName => $atomName,
        atomNum => trim(substr($_, 6,5)),        # Num field is columns 7-11
        chain => trim(substr($_, 21,1)),         # Chain field is column 22
        resNum => trim(substr($_, 22,4)),         # ResNo field is columns 23-26
        x => trim(substr($_, 30,8)),             # X field is columns 31-38
        y => trim(substr($_, 38,8)),             # Y field is columns 39-46
        z => trim(substr($_, 46,8))              # Z field is columns 37-54
    };
}

close(PDB);
}

#
# trim
#
# removes whitespace from start and end of string
#

sub trim($)
{
    my ($string) = @_;          # retrieve the passed argument
    $string =~ s/^\s+//;        # remove leading whitespace
    $string =~ s/\s+$//;        # remove trailing whitespace
    return $string;
}

#
# is_number
#
# returns 1 if passed argument is a number (allows whitespace, negative, and
decimal point)
# returns 0 if passed argument is blank or not a number
#

```

```

sub is_number($)
{
    $_ = shift;
    s/^\\s+//;
    s/\\s+$//;
    return 1 if /^-?[0-9]+$ / || /^-?[0-9]*\\. [0-9]+$ / || /^-?[0-9]+\\. [0-9]*$/;
    return 0;
}

```

```

#
# inter_residue_distance
#
# returns the minimum distance between any atoms of the specified residues
# (residues are specified according to index in @residues)
#

```

```

sub inter_residue_distance($, $)
{
    my ($r1, $r2) = @_;

    ## convert to PDB numbering

    my $resNum1 = $residues[$r1]{resNum};
    my $resNum2 = $residues[$r2]{resNum};

    my $min_dist = 1000000;

    for ($a1 = 0; $a1 < @atoms; ++$a1)
    {
        next unless ( $atoms[$a1]{resNum} == $resNum1 );
        for ($a2 = 0; $a2 < @atoms; ++$a2)
        {
            next unless ( $atoms[$a2]{resNum} == $resNum2 );
            my $dist = $distances[$a1][$a2];
            $min_dist = $dist if ($dist < $min_dist);
        }
    }
    return $min_dist;
}

```

```

#
# compute_distances
#
# computes the distances between all atoms
#

```

```

sub compute_distances
{
    for(my $atom1=0; $atom1 < @atoms; $atom1++)
    {
        for(my $atom2=$atom1; $atom2 < @atoms; $atom2++)

```

```

    {
        my ($x1,$y1,$z1) = ($atoms[$atom1]->{x}, $atoms[$atom1]->{y},
$atoms[$atom1]->{z});
        my ($x2,$y2,$z2) = ($atoms[$atom2]->{x}, $atoms[$atom2]->{y},
$atoms[$atom2]->{z});

        my $distance = sqrt(($x1-$x2)**2 + ($y1-$y2)**2 + ($z1-$z2)**2);

        $distances[$atom1][$atom2] = $distance;
        $distances[$atom2][$atom1] = $distance;
    }
}

```

```

#
# compute_neighbor_counts
#
# computes the number of neighbors that each atom has.
# paramter is the cutoff, in Angstroms, for atomic neighborhood
#

sub compute_neighbor_counts
{
    $DISTANCE_CUTOFF = 10;      # criterion for neighborhood, in Angstroms

    for ($atom1=0; $atom1 < @atoms; $atom1++)
    {
        my $count = 0;
        for ($atom2=0; $atom2 < @atoms; $atom2++)
        {
            $count++ if ($distances[$atom1][$atom2] <= $DISTANCE_CUTOFF
&& $atom1 != $atom2);
        }
        $atoms[$atom1]{neighborCount} = $count;
    }
}

```

```

#
# compute_residue_avNapsa
#
# for each residue, compute
# Average Neighbor Atoms Per Sidechain Atom (AvNAPSA)
# (sidechain atoms are all those except N, C, O, CA)
# for glycines, just use CA
#

sub compute_residue_avNapsa
{
    for (my $r = 0; $r < @residues; $r++)
    {
        my $numSideChainAtoms = 0;
        my $totalNeighbors = 0;
        my $resName = $residues[$r]{resName};
        my $resNum = $residues[$r]{resNum};
    }
}

```

```

for (my $a = 0; $a < @atoms; $a++)
{
    if ($atoms[$a]{resNum} == $resNum)
    {
        my $atomName = $atoms[$a]{atomName};
        if (
            (
                $atomName ne "C"
                && $atomName ne "O"
                && $atomName ne "N"
                && $atomName ne "CA"
            )
            || ( $atomName eq "CA" && $resName eq "GLY")
        )
        {
            $numSideChainAtoms++;
            $totalNeighbors += $atoms[$a]{neighborCount};
        }
    }
}
my $avNapsa = $totalNeighbors / $numSideChainAtoms;
$residues[$r]{avNapsa} = $avNapsa;
}
}

#
# toggle31
#
# converts 3-letter abbrev to 1-letter
# or 1-letter abbrev to 3-letter
#

sub toggle31($)
{
    %conv3tol = ( "ALA" => "A", "CYS" => "C", "SER" => "S", "LEU" => "L",
        "ILE" => "I", "PHE" => "F", "ARG" => "R", "ASN" => "N", "GLN" => "Q",
        "TYR" => "Y", "LYS" => "K", "ASP" => "D", "GLU" => "E", "VAL" => "V",
        "TRP" => "W", "MET" => "M", "HIS" => "H", "GLY" => "G", "PRO" => "P",
        "THR" => "T" );

    %convlto3 = reverse %conv3tol;

    my ($abbrev) = @_ ;

    $abbrev = uc $abbrev;
    return $convlto3{$abbrev} if length ($abbrev) == 1;
    return $conv3tol{$abbrev} if length ($abbrev) == 3;
    die "in toggle31(): invalid amino acid abbreviation $abbrev\n";
}

#
# is_aa
#

```



```
# returns 1 if passed argument is a 1-letter amino acid
#

sub is_aa($)
{
    my ($string) = @_;

    return 1 if (length toggle31($string) == 3);
    return 0;
}
```

<110> 哈佛大学校长及研究员协会

<120> 蛋白质表面重建

<130> 034291-0223 (HU 2772)

<140> PCT/US2007/070254

<141> 2007-06-01

<160> 23

<170> PatentIn version 3.5

<210> 1

<211> 238

<212> PRT

<213> 维多利亚水母

<400> 1

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
1 5 10 15

Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
20 25 30

Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
35 40 45

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe
50 55 60

Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln
65 70 75 80

His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
85 90 95

Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val
100 105 110

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile
115 120 125

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn
130 135 140

Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly

145	150	155	160
Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val	165	170	175
Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro	180	185	190
Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser	195	200	205
Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val	210	215	220
Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys	225	230	235
<210> 2			
<211> 248			
<212> PRT			
<213> 人工序列			
<220>			
<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)			
<400> 2			
Met Gly His His His His His His Gly Gly Ala Ser Lys Gly Glu Glu	1	5	10
Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val	20	25	30
Asn Gly His Glu Phe Ser Val Arg Gly Glu Gly Glu Gly Asp Ala Thr	35	40	45
Glu Gly Glu Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Glu Leu Pro	50	55	60
Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys	65	70	75
Phe Ser Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser	85	90	95
Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Ser Phe Lys Asp	100	105	110

Asp Gly Thr Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr
115 120 125

Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly
130 135 140

Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Phe Asn Ser His Asp Val
145 150 155 160

Tyr Ile Thr Ala Asp Lys Gln Glu Asn Gly Ile Lys Ala Glu Phe Glu
165 170 175

Ile Arg His Asn Val Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr
180 185 190

Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asp
195 200 205

His Tyr Leu Ser Thr Glu Ser Ala Leu Ser Lys Asp Pro Asn Glu Asp
210 215 220

Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Asp
225 230 235 240

His Gly Met Asp Glu Leu Tyr Lys
245

<210> 3
<211> 248
<212> PRT
<213> 人工序列

<220>
<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 3

Met Gly His His His His His His Gly Gly Ala Ser Lys Gly Glu Glu
1 5 10 15

Leu Phe Asp Gly Glu Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val
20 25 30

Asn Gly His Glu Phe Ser Val Arg Gly Glu Gly Glu Gly Asp Ala Thr
35 40 45

Glu Gly Glu Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Glu Leu Pro
50 55 60

Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys
65 70 75 80

Phe Ser Arg Tyr Pro Asp His Met Asp Gln His Asp Phe Phe Lys Ser
85 90 95

Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Ser Phe Lys Asp
100 105 110

Asp Gly Thr Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr
115 120 125

Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly
130 135 140

Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Phe Asn Ser His Asp Val
145 150 155 160

Tyr Ile Thr Ala Asp Lys Gln Glu Asn Gly Ile Lys Ala Glu Phe Glu
165 170 175

Ile Arg His Asn Val Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr
180 185 190

Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asp
195 200 205

His Tyr Leu Ser Thr Glu Ser Ala Leu Ser Lys Asp Pro Asn Glu Asp
210 215 220

Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Asp
225 230 235 240

His Gly Met Asp Glu Leu Tyr Lys
245

<210> 4
<211> 248
<212> PRT
<213> 人工序列

<220>

<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 4

Met Gly His His His His His His Gly Gly Ala Ser Lys Gly Glu Glu
1 5 10 15

Leu Phe Asp Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val
20 25 30

Asn Gly His Glu Phe Ser Val Arg Gly Glu Gly Glu Gly Asp Ala Thr
35 40 45

Glu Gly Glu Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Glu Leu Pro
50 55 60

Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys
65 70 75 80

Phe Ser Asp Tyr Pro Asp His Met Asp Gln His Asp Phe Phe Lys Ser
85 90 95

Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Ser Phe Lys Asp
100 105 110

Asp Gly Thr Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr
115 120 125

Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly
130 135 140

Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Phe Asn Ser His Asp Val
145 150 155 160

Tyr Ile Thr Ala Asp Lys Gln Glu Asn Gly Ile Lys Ala Glu Phe Glu
165 170 175

Ile Arg His Asn Val Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr
180 185 190

Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asp
195 200 205

His Tyr Leu Ser Thr Glu Ser Ala Leu Ser Lys Asp Pro Asn Glu Asp
210 215 220

Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Asp
225 230 235 240

His Gly Met Asp Glu Leu Tyr Lys
245

<210> 5

<211> 248

<212> PRT

<213> 人工序列

<220>

<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 5

Met Gly His His His His His His Gly Gly Ala Ser Lys Gly Glu Arg
1 5 10 15

Leu Phe Arg Gly Lys Val Pro Ile Leu Val Glu Leu Lys Gly Asp Val
20 25 30

Asn Gly His Lys Phe Ser Val Arg Gly Lys Gly Lys Gly Asp Ala Thr
35 40 45

Arg Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro
50 55 60

Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys
65 70 75 80

Phe Ser Arg Tyr Pro Lys His Met Lys Arg His Asp Phe Phe Lys Ser
85 90 95

Ala Met Pro Lys Gly Tyr Val Gln Glu Arg Thr Ile Ser Phe Lys Lys
100 105 110

Asp Gly Lys Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Arg Thr
115 120 125

Leu Val Asn Arg Ile Lys Leu Lys Gly Arg Asp Phe Lys Glu Lys Gly
130 135 140

Asn Ile Leu Gly His Lys Leu Arg Tyr Asn Phe Asn Ser His Lys Val
145 150 155 160

Tyr Ile Thr Ala Asp Lys Arg Lys Asn Gly Ile Lys Ala Lys Phe Lys
165 170 175

Ile Arg His Asn Val Lys Asp Gly Ser Val Gln Leu Ala Asp His Tyr
180 185 190

Gln Gln Asn Thr Pro Ile Gly Arg Gly Pro Val Leu Leu Pro Arg Asn
195 200 205

His Tyr Leu Ser Thr Arg Ser Lys Leu Ser Lys Asp Pro Lys Glu Lys
210 215 220

Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Lys
225 230 235 240

His Gly Arg Asp Glu Arg Tyr Lys
245

<210> 6
<211> 248
<212> PRT
<213> 人工序列

<220>
<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 6

Met Gly His His His His His Gly Gly Arg Ser Lys Gly Lys Arg
1 5 10 15

Leu Phe Arg Gly Lys Val Pro Ile Leu Val Glu Leu Lys Gly Asp Val
20 25 30

Asn Gly His Lys Phe Ser Val Arg Gly Lys Gly Lys Gly Asp Ala Thr
35 40 45

Arg Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro
50 55 60

Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys
65 70 75 80

Phe Ser Arg Tyr Pro Lys His Met Lys Arg His Asp Phe Phe Lys Ser
85 90 95

Ala Met Pro Lys Gly Tyr Val Gln Glu Arg Thr Ile Ser Phe Lys Lys
100 105 110

Asp Gly Lys Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Arg Thr
115 120 125

Leu Val Asn Arg Ile Lys Leu Lys Gly Arg Asp Phe Lys Glu Lys Gly
130 135 140

Asn Ile Leu Gly His Lys Leu Arg Tyr Asn Phe Asn Ser His Lys Val
145 150 155 160

Tyr Ile Thr Ala Asp Lys Arg Lys Asn Gly Ile Lys Ala Lys Phe Lys
165 170 175

Ile Arg His Asn Val Lys Asp Gly Ser Val Gln Leu Ala Asp His Tyr
180 185 190

Gln Gln Asn Thr Pro Ile Gly Arg Gly Pro Val Leu Leu Pro Arg Lys
195 200 205

His Tyr Leu Ser Thr Arg Ser Lys Leu Ser Lys Asp Pro Lys Glu Lys
210 215 220

Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Lys
225 230 235 240

His Gly Arg Lys Glu Arg Tyr Lys
245

<210> 7
<211> 248
<212> PRT
<213> 人工序列

<220>
<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 7

Met Gly His His His His His His Gly Gly Arg Ser Lys Gly Lys Arg

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

His Gly Arg Lys Glu Arg Tyr Lys
245

<210> 8
<211> 747
<212> DNA
<213> 人工序列

<220>
<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 8
atggggcatc accatcatca tcatggcggt gcgtctaagg gggaggagtt atttacgggt 60
gtgggtgccga tcctggtgga gcttgatggc gatgttaacg gccatgaatt ttctgtccgc 120
ggatgaagggg agggatgatgc cacggaaggg gagctgacac ttaaatttat ttgcaccacc 180
ggatgaactcc cgggtcccggt gccgacctg gtgaccaccc tgacctacgg cgttcaatgc 240
ttttcacggt atccggatca catgaagcaa cacgacttct taaaagcgc gatgcctgaa 300
ggctatgttc aagaacgtac aattagtttt aaagatgacg gcacctacaa gacccggtgcg 360
gaagtaaaat ttgaagggga cacttttagtg aaccgcatcg agctgaaagg gatcgatttt 420
aaagaagatg ggaatatacct gggacacaaa cttgaataca actttaatag tcatgacgtc 480
tatatcacgg cggacaaaca ggaaaacgga attaaggcag aatttgagat tcggcataat 540
gtcgaagatg gctcggtaca gttggctgat cactatcagc agaatacgcc gatttgagat 600
ggatccggttt tattaccaga cgatcactat ctgtccaccg aatccgccct gagcaaagat 660
ccgaatgaag accgggacca tatggttctg ctggaatttg ttacggcggc tggatttgac 720
catggcatgg atgagctgta taagtag 747

<210> 9
<211> 747
<212> DNA
<213> 人工序列

<220>
<223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 9
atggggcatc accatcatca tcatggcggt gcgtctaagg gggaggagtt atttgatggt 60
gaagtgccga tcctggtgga gcttgatggc gatgttaacg gccatgaatt ttctgtccgc 120
ggatgaagggg agggatgatgc cacggaaggg gagctgacac ttaaatttat ttgcaccacc 180

ggtgaactcc cgggtcccggtg gccgaccctg gtgaccaccc tgacctacgg cgttcaatgc	240
ttttcacgtt atccggatca catggaccaa cacgacttct ttaaaagcgc gatgcctgaa	300
ggctatgttc aagaacgtac aattagtttt aaagatgacg gcacctacaa gacctgtgcg	360
gaagtaaaat ttgaagggga cacttttagtg aaccgcatcg agctgaaagg gatcgatttt	420
aaagaagatg ggaatatacct gggacacaaa cttgaataca actttaatag tcatgacgtc	480
tatatcacgg cggacaaaca ggaaaacgga attaaggcag aatttgagat tcggcataat	540
gtcgaagatg gctcggtaca gttggctgat cactatcagc agaatacgcc gattggagat	600
ggtccggttt tattaccaga cgatcactat ctgtccaccg aatccgccct gagcaaagat	660
ccgaatgaag accgggacca tatggttctg ctggaatttg ttacggcggc tggatttgac	720
catggcatgg atgagctgta taagtag	747

<210> 10
 <211> 747
 <212> DNA
 <213> 人工序列

<220>
 <223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 10	
atggggcatc accatcatca tcatggcggg gcgtctaagg gggaggagtt atttgatggt	60
gtgggtgccga tcctggtgga gcttgatggc gatgttaacg gccatgaatt ttctgtccgc	120
ggtgaagggg agggatgatgc cacggaaggg gagctgacac ttaaatttat ttgcaccacc	180
ggtgaactcc cgggtcccggtg gccgaccctg gtgaccaccc tgacctacgg cgttcaatgc	240
ttttcagatt atccggatca catggaccaa cacgacttct ttaaaagcgc gatgcctgaa	300
ggctatgttc aagaacgtac aattagtttt aaagatgacg gcacctacaa gacctgtgcg	360
gaagtaaaat ttgaagggga cacttttagtg aaccgcatcg agctgaaagg gatcgatttt	420
aaagaagatg ggaatatacct gggacacaaa cttgaataca actttaatag tcatgacgtc	480
tatatcacgg cggacaaaca ggaaaacgga attaaggcag aatttgagat tcggcataat	540
gtcgaagatg gctcggtaca gttggctgat cactatcagc agaatacgcc gattggagat	600
ggtccggttt tattaccaga cgatcactat ctgtccaccg aatccgccct gagcaaagat	660
ccgaatgaag accgggacca tatggttctg ctggaatttg ttacggcggc tggatttgac	720
catggcatgg atgagctgta taagtag	747

<210> 11
 <211> 747
 <212> DNA
 <213> 人工序列

<220>
 <223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 11
 atggggcatc atcatcatca ccacggcggg gcgtctaagg gagagcgctt gtttcgcggc 60
 aaagtcccca ttcttgtgga gctcaaaggt gatgtaaatg gtcataaatt tagtgtgcgc 120
 gggaaaggga aaggagatgc tacgcggggc aagctcacc tgaatttat ttgcacaacc 180
 ggcaaactgc cagtgccgtg gcctacatta gtcactactc tgacgtacgg tgttcagtgc 240
 ttttctcgct atcccaaaca catgaaacgc catgatttct tcaagagcgc gatgccaaaa 300
 ggttatgtgc aggaacgcac catcagcttt aaaaaagacg gcaaatataa aaccctgca 360
 gaagttaaatt tcgaaggccg caccctggtc aaccgcatta aactgaaagg tcgtgacttc 420
 aaagagaaag gtaatatctt tggtcacaaa ctgcgctata atttcaactc tcacaaagtt 480
 tatattacgg cggataaacg taaaaacggg attaaagcga aatttaagat tcgtcataat 540
 gttaaagacg gcagtgtgca gttagcggat cattatcagc agaatacccc aattggtcgc 600
 ggtccagtgc tgctgccgcg taaccattat ctgtcgacct gcagcaaact cagcaaagac 660
 ccgaaagaaa aacgtgacca catggtatta ctggaatttg tgaccgcagc aggcattaaa 720
 catggccgcg atgaacgtta caaatag 747

<210> 12
 <211> 747
 <212> DNA
 <213> 人工序列

<220>
 <223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 12
 atgggccatc atcatcacca ccacggcggc cgctcaaaag gtaaagcgtt gttccgtggt 60
 aaagtaccga tcttagtgga gctcaaaggg gatgtgaatg gccataagtt ctcggttcgt 120
 ggcaaaggta agggagatgc gacgcgcggc aaattaacgc tgaattcat ttgtactaca 180
 ggtaaactgc cggtgccatg gcctactctc gtcaccacgt tgacctatgg ggttcaatgc 240
 ttcagccggt accctaaaca catgaagcgc cagatttct tcaaatcggc gatgccaaag 300

gggtatgtcc aggaacgcac tatcagcttc aaaaaagacg gtaagtataa aactcgtgct	360
gaagttaa at tcgaaggacg cacactggta aatcgcat ta aattgaagg ggcgcacttt	420
aaggaaaaag gtaatatctt aggtcacaaa ttgcgctaca acttcaactc tcataaagtt	480
tacattacag cagataagcg taaaaatggc atcaaagcga aattcaaa at tcgtcacaa t	540
gtgaaagatg gtagcgtgca attagccgat cattaccagc agaatacgcc gatcggtcgc	600
ggcccagtac tgttgccgcg caaacattac ttatctaccc ggagtaaact gtctaaagac	660
ccaaaagaga agcgcgacca tatggttctc ctggagt ttg tcaccgccgc cggaattaaa	720
cacggccgca aagagcgcta taaatag	747

<210> 13
 <211> 747
 <212> DNA
 <213> 人工序列

<220>
 <223> 来自维多利亚水母的表面修饰绿色荧光蛋白 (GFP)

<400> 13	
atgggccacc atcatcatca ccacggggga cgctctaaag gtaaactct gtttcgtgga	60
aaggtgcca ttctggttaa actcaaaggt gatgtcaacg gccataagtt ttcggttcgt	120
ggcaaaggta aaggtgatgc gacgcgcggg aaattaacac tgaaatttat ttgcacaacc	180
ggaaaactcc ctgtgccgtg gccgactttg gtgaccacat taacctatgg tgttcaatgc	240
ttctcacgtt atccgaagca tatgaaacgt catgattttt tcaaactcgc tatgccgaaa	300
ggttacgtcc aggagcgcac catctcat tt aagaaagacg gtaagtataa aaccctgct	360
gaagtaaa at tcaaaggacg caccctgggtg aatcgcat ta aactgaaagg tcgtgatttc	420
aaagaaaagg gaaatatttt agggcataag ctccgttata attttaacag tcataaggtg	480
tatattaccg ctgataaacg caaaaacgga atcaaagcga aatttaagat ccgtcataat	540
gtaaaagatg gtcagtcca actggcaaaa cattaccagc agaatacccc gatcggccgc	600
ggtcctgtgc ttctgccgcg taaacactac ttgtcgacce ggtcaaaatt gagtaaagat	660
ccgaaggaaa agcgtgatca catggctttg aaggaatttg taactgcagc aggtattaaa	720
cacgggcgca aagaacgtta caaatag	747

<210> 14
 <211> 128
 <212> PRT
 <213> 人工序列

<220>

<223> 野生型链霉亲和素

<400> 14

Ala Ala Glu Ala Gly Ile Thr Gly Thr Trp Tyr Asn Gln Leu Gly Ser
1 5 10 15

Thr Phe Ile Val Thr Ala Gly Ala Asp Gly Ala Leu Thr Gly Thr Tyr
20 25 30

Glu Ser Ala Val Gly Asn Ala Glu Ser Arg Tyr Val Leu Thr Gly Arg
35 40 45

Tyr Asp Ser Ala Pro Ala Thr Asp Gly Ser Gly Thr Ala Leu Gly Trp
50 55 60

Thr Val Ala Trp Lys Asn Asn Tyr Arg Asn Ala His Ser Ala Thr Thr
65 70 75 80

Trp Ser Gly Gln Tyr Val Gly Gly Ala Glu Ala Arg Ile Asn Thr Gln
85 90 95

Trp Leu Leu Thr Ser Gly Thr Thr Glu Ala Asn Ala Trp Lys Ser Thr
100 105 110

Leu Val Gly His Asp Thr Phe Thr Lys Val Lys Pro Ser Ala Ala Ser
115 120 125

<210> 15

<211> 137

<212> PRT

<213> 人工序列

<220>

<223> 表面修饰链霉亲和素 (SAV-NEG40)

<400> 15

Met Gly His His His His His His Gly Gly Ala Glu Ala Gly Ile Thr
1 5 10 15

Gly Thr Trp Tyr Asn Gln Leu Gly Ser Thr Phe Ile Val Thr Ala Gly
20 25 30

Ala Asp Gly Ala Leu Thr Gly Thr Tyr Glu Ser Ala Val Gly Asp Ala
35 40 45

Glu Ser Glu Tyr Val Leu Thr Gly Arg Tyr Asp Ser Ala Pro Ala Thr
50 55 60

Asp Gly Ser Gly Thr Ala Leu Gly Trp Thr Val Ala Trp Lys Asn Asp
65 70 75 80

Tyr Glu Asn Ala His Ser Ala Thr Thr Trp Ser Gly Gln Tyr Val Gly
85 90 95

Gly Ala Glu Ala Arg Ile Asn Thr Gln Trp Leu Leu Thr Ser Gly Thr
100 105 110

Thr Glu Ala Asp Ala Trp Lys Ser Thr Leu Val Gly His Asp Thr Phe
115 120 125

Thr Lys Val Glu Pro Ser Ala Ala Ser
130 135

<210> 16
<211> 137
<212> PRT
<213> 人工序列

<220>
<223> 表面修饰链霉亲和素 (SAV-POS52)

<400> 16

Met Gly His His His His His His Gly Gly Ala Lys Ala Gly Ile Thr
1 5 10 15

Gly Thr Trp Tyr Asn Gln Leu Gly Ser Thr Phe Ile Val Thr Ala Gly
20 25 30

Ala Lys Gly Ala Leu Thr Gly Thr Tyr Glu Ser Ala Val Gly Asn Ala
35 40 45

Lys Ser Arg Tyr Val Leu Thr Gly Arg Tyr Asp Ser Ala Pro Ala Thr
50 55 60

Lys Gly Ser Gly Thr Ala Leu Gly Trp Thr Val Ala Trp Lys Asn Lys
65 70 75 80

Tyr Arg Asn Ala His Ser Ala Thr Thr Trp Ser Gly Gln Tyr Val Gly
85 90 95

Gly Ala Lys Ala Arg Ile Asn Thr Gln Trp Leu Leu Thr Ser Gly Thr
 100 105 110

Thr Lys Ala Lys Ala Trp Lys Ser Thr Leu Val Gly His Asp Thr Phe
 115 120 125

Thr Lys Val Lys Pro Ser Ala Ala Ser
 130 135

<210> 17
 <211> 442
 <212> DNA
 <213> 人工序列

<220>
 <223> 表面修饰链霉亲和素 (SAV-NEG40)

<400> 17
 gggtcagcca tgggtcatca ccaccacat cacggtggcg cgaagcagg tattaccggt 60
 acctggtata accagttagg ctcaaccttt attgtgaccg cgggagcgga cggcgcctta 120
 accggtacct acgaatcagc tgtaggtgac gcggaatcag agtacgtatt aaccggtcgt 180
 tatgatagcg cgccggcgac tgacggtagc ggtactgctt taggttggac cgtagcgtgg 240
 aagaatgatt atgaaaacgc acatagcgca acaacgtggt cagggcagta cgttggcgga 300
 gctgaggcgc gcattaacac gcagtggta ttaactagcg gcaccactga agctgatgcc 360
 tggaagagca cgtagtgagg tcatgatacc ttactaaag tggaaccttc agctgcgtca 420
 taataatgac tcgagacctg ca 442

<210> 18
 <211> 442
 <212> DNA
 <213> 人工序列

<220>
 <223> 表面修饰链霉亲和素 (SAV-POS52)

<400> 18
 gggtcagcca tgggtcatca ccaccacat cacggtggcg ccaaagcagg tattaccggt 60
 acctggtata accagttagg ctcaaccttt attgtgaccg cgggagcgaa aggcgcctta 120
 accggtacct acgaatcagc tgtaggaaac gcaaaatcac gctacgtatt aaccggtcgt 180
 tatgatagcg cgccggcgac taaaggtagc ggtactgctt taggttggac cgtagcgtgg 240
 aagaataagt atcgtaatgc gcacagtgt accacttggt cagggcagta cgtaggggga 300
 gccaaagcac gtatcaacac gcagtggta ttaacatcag gtaccaccaa agcgaaagcc 360

tggaagagca cgtagtggg tcatgatacc ttactaaag tgaaaccttc agctgcgtca	420
taataatgac tcgagacctg ca	442

<210> 19
 <211> 217
 <212> PRT
 <213> 人工序列

<220>
 <223> 野生型谷胱甘肽-S-转移酶 (GST)

<400> 19

Met	Gly	His	His	His	His	His	His	Gly	Gly	Pro	Pro	Tyr	Thr	Ile	Thr
1				5					10					15	

Tyr	Phe	Pro	Val	Arg	Gly	Arg	Cys	Glu	Ala	Met	Arg	Met	Leu	Leu	Ala
			20					25					30		

Asp	Gln	Asp	Gln	Ser	Trp	Lys	Glu	Glu	Val	Val	Thr	Met	Glu	Thr	Trp
		35					40					45			

Pro	Pro	Leu	Lys	Pro	Ser	Cys	Leu	Phe	Arg	Gln	Leu	Pro	Lys	Phe	Gln
	50					55					60				

Asp	Gly	Asp	Leu	Thr	Leu	Tyr	Gln	Ser	Asn	Ala	Ile	Leu	Arg	His	Leu
65					70					75					80

Gly	Arg	Ser	Phe	Gly	Leu	Tyr	Gly	Lys	Asp	Gln	Lys	Glu	Ala	Ala	Leu
			85						90					95	

Val	Asp	Met	Val	Asn	Asp	Gly	Val	Glu	Asp	Leu	Arg	Cys	Lys	Tyr	Ala
		100						105					110		

Thr	Leu	Ile	Tyr	Thr	Asn	Tyr	Glu	Ala	Gly	Lys	Glu	Lys	Tyr	Val	Lys
		115					120					125			

Glu	Leu	Pro	Glu	His	Leu	Lys	Pro	Phe	Glu	Thr	Leu	Leu	Ser	Gln	Asn
	130					135					140				

Gln	Gly	Gly	Gln	Ala	Phe	Val	Val	Gly	Ser	Gln	Ile	Ser	Phe	Ala	Asp
145				150						155					160

Tyr	Asn	Leu	Leu	Asp	Leu	Leu	Arg	Ile	His	Gln	Val	Leu	Asn	Pro	Ser
				165					170					175	

Cys Leu Asp Ala Phe Pro Leu Leu Ser Ala Tyr Val Ala Arg Leu Ser
180 185 190

Ala Arg Pro Lys Ile Lys Ala Phe Leu Ala Ser Pro Glu His Val Asn
195 200 205

Arg Pro Ile Asn Gly Asn Gly Lys Gln
210 215

<210> 20
<211> 217
<212> PRT
<213> 人工序列

<220>
<223> 表面修饰谷胱甘肽-S-转移酶 (GST-NEG40)

<400> 20

Met Gly His His His His His His Gly Gly Pro Pro Tyr Thr Ile Thr
1 5 10 15

Tyr Phe Pro Val Arg Gly Arg Cys Glu Ala Met Arg Met Leu Leu Ala
20 25 30

Asp Gln Asp Gln Ser Trp Glu Glu Glu Val Val Thr Met Glu Thr Trp
35 40 45

Pro Pro Leu Lys Pro Ser Cys Leu Phe Arg Gln Leu Pro Lys Phe Gln
50 55 60

Asp Gly Asp Leu Thr Leu Tyr Gln Ser Asn Ala Ile Leu Arg His Leu
65 70 75 80

Gly Arg Ser Phe Gly Leu Tyr Gly Glu Asp Glu Glu Glu Ala Ala Leu
85 90 95

Val Asp Met Val Asn Asp Gly Val Glu Asp Leu Arg Cys Lys Tyr Ala
100 105 110

Thr Leu Ile Tyr Thr Asp Tyr Glu Ala Gly Lys Glu Glu Tyr Val Glu
115 120 125

Glu Leu Pro Glu His Leu Lys Pro Phe Glu Thr Leu Leu Ser Glu Asn
130 135 140

Glu Gly Gly Glu Ala Phe Val Val Gly Ser Glu Ile Ser Phe Ala Asp

145	150	155	160
Tyr Asn Leu Leu Asp 165	Leu Leu Arg Ile His 170	Gln Val Leu Asn Pro 175	Ser
Cys Leu Asp Ala Phe Pro Leu Leu Ser 180	Ala Tyr Val Ala Arg 185	Leu Ser	
Ala Arg Pro Glu Ile Glu Ala Phe Leu Ala Ser Pro Glu His Val Asp 195		200	205
Arg Pro Ile Asn Gly Asn Gly Lys Gln 210	215		
<210> 21			
<211> 217			
<212> PRT			
<213> 人工序列			
<220>			
<223> 表面修饰谷胱甘肽-S-转移酶 (GST-POS50)			
<400> 21			
Met Gly His His His His His His Gly Gly Pro Pro Tyr Thr Ile Thr 1	5	10	15
Tyr Phe Pro Val Arg Gly Arg Cys Glu Ala Met Arg Met Leu Leu Ala 20	25	30	
Asp Gln Lys Gln Ser Trp Lys Glu Glu Val Val Thr Met Lys Thr Trp 35	40	45	
Pro Pro Leu Lys Pro Ser Cys Leu Phe Arg Gln Leu Pro Lys Phe Gln 50	55	60	
Asp Gly Lys Leu Thr Leu Tyr Gln Ser Asn Ala Ile Leu Arg His Leu 65	70	75	80
Gly Arg Ser Phe Gly Leu Tyr Gly Lys Lys Gln Lys Glu Ala Ala Leu 85	90	95	
Val Asp Met Val Asn Asp Gly Val Glu Asp Leu Arg Cys Lys Tyr Ala 100	105	110	
Thr Leu Ile Tyr Thr Lys Tyr Lys Ala Gly Lys Lys Lys Tyr Val Lys 115	120	125	

Lys Leu Pro Lys His Leu Lys Pro Phe Glu Thr Leu Leu Ser Lys Asn
130 135 140

Lys Gly Gly Lys Ala Phe Val Val Gly Ser Lys Ile Ser Phe Ala Asp
145 150 155 160

Tyr Asn Leu Leu Asp Leu Leu Arg Ile His Gln Val Leu Asn Pro Ser
165 170 175

Cys Leu Lys Ala Phe Pro Leu Leu Ser Ala Tyr Val Ala Arg Leu Ser
180 185 190

Ala Arg Pro Lys Ile Lys Ala Phe Leu Ala Ser Pro Glu His Val Lys
195 200 205

Arg Pro Ile Asn Gly Asn Gly Lys Gln
210 215

<210> 22
<211> 682
<212> DNA
<213> 人工序列

<220>
<223> 表面修饰谷胱甘肽-S-转移酶 (GST-NEG40)

<400> 22
ggttcagcca tgggtcatca ccaccacat cacggtggcc cgccgtacac cattacatac 60
tttccggtac gtggtcgttg tgaagcgatg cgtatgttat tagcggacca ggaccaatca 120
tgggaagaag aagtagtgac aatggaaacc tggccgccgt taaagcctag ctgtttattc 180
cgtcaattac cgaagtttca ggatggatgat ttaaccttat accagtctaa cgcgatctta 240
cgtcatttag gtcgctcatt tggtttatac ggtgaagatg aagaagaagc agccttagtg 300
gatatggtga atgatggcgt ggaagactta cgttgtaaata acgcgacgtt aatttacact 360
gattatgaag ccggtaaaga ggagtacgtg gaagaattac ctgaacacct gaagccgttt 420
gaaacattac tgagcgaaaa tgaaggaggt gaggcgttcg tagttggtag cgaaattagc 480
ttcgttgatt ataacttatt agacttatta cgcattcacc aggtttttaa tcctagctgt 540
ttagacgctt tcccgttact gagegcataat gtagcgcgcc tgagcgcccg tccggaaatt 600
gaagctttct tagcgtcacc tgaacacgta gaccgcccga ttaacggaaa cggcaagcag 660
taataatgag gtaccacctg ca 682

<210>	23	
<211>	682	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	表面修饰谷胱甘肽-S-转移酶 (GST-POS50)	
<400>	23	
ggttcagcca tgggtcatca ccaccacat cacggtggcc cgccgtacac cattacatac		60
tttccggtac gtggtcgttg tgaagcgatg cgtatgttat tagcggacca gaaacaatca		120
tggaaagaag aagtagtgac aatgaagacc tggccgccgt taaagcctag ctgtttattc		180
cgtcaattac cgaagtttca ggatggtaaa ttaaccttat accagtctaa cgcgatctta		240
cgtcatttag gtcgctcatt tggtttatac ggtaagaagc agaaagaagc agccttagtg		300
gatatgggtga atgatggcgt ggaagactta cgttgtaa at acgcgacgtt aatttacact		360
aaatataaag ccggtaaaaa gaagtacgtg aaaaaattac ctaaacacct gaagccgttt		420
gaaacattac tgagcaaaaa taaaggaggt aaggcgttcg tagttggtag caagattagc		480
ttcgctgatt ataacttatt agacttatta cgcattcacc aggtttttaa tcttagctgt		540
ttaaaggcitt tcccgttact gagcgcatat gtagcgcgcc tgagcgcccg tccgaagatc		600
aaagcitttct tagcgtcacc tgaacacgtg aagcgcccga ttaacggaaa cggcaagcag		660
taataatgag gtaccacctg ca		682

图1a-1
图1a-2

图1a

GFP (-30) MGH HHH HHH G G A S K G E E L F D G V V P I L V E L D G D V N G H E F S V R G E G E G D A T E G
GFP (-25) MGH HHH HHH G G A S K G E E L F T G V V P I L V E L D G D V N G H E F S V R G E G E G D A T E G
sfGFP MGH HHH HHH G G A S K G E E L F T G V V P I L V E L D G D V N G H K F S V R G E G E G D A T N G
GFP (+36) MGH HHH HHH G G A S K G E R L F R G K V P I L V E L K G D V N G H K F S V R G K G D A T R G
GFP (+48) MGH HHH HHH G G R S K G K R L F R C K V P I L V K L K G D V N G H K F S V R C K G D A T R G

GFP (-30) E L T L K F I C T T G E L P V P W P T L V T T L T Y G V Q C F S D Y P D H M D Q H D F F K S A M P E
GFP (-25) E L T L K F I C T T G E L P V P W P T L V T T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E
sfGFP K L T L K F I C T T G K L P V P W P T L V T T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E
GFP (+36) K L T L K F I C T T G K L P V P W P T L V T T L T Y G V Q C F S R Y P K H M K R H D F F K S A M P K
GFP (+48) K L T L K F I C T T G K L P V P W P T L V T T L T Y G V Q C F S R Y P K H M K R H D F F K S A M P K

GFP (-30) G Y V Q E R T I S F K D D G T Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G H K
GFP (-25) G Y V Q E R T I S F K D D G T Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G H K
sfGFP G Y V Q E R T I S F K D D G T Y K T R A E V K F E G D T L V N R I E L K G I D F K E D G N I L G H K
GFP (+36) G Y V Q E R T I S F K K D G K Y K T R A E V K F E G R T L V N R I K L K G R D F K E K G N I L G H K
GFP (+48) G Y V Q E R T I S F K K D G K Y K T R A E V K F K G R T L V N R I K L K G R D F K E K G N I L G H K

GFP (-30) L E Y N F N S H D V Y I T A D K Q E N G I K A E F E I R H N V E D G S V Q L A D H Y Q Q N T P I G D
GFP (-25) L E Y N F N S H D V Y I T A D K Q E N G I K A E F E I R H N V E D G S V Q L A D H Y Q Q N T P I G D
sfGFP L E Y N F N S H N V Y I T A D K Q K N G I K A N F K I R H N V E D G S V Q L A D H Y Q Q N T P I G D
GFP (+36) L R Y N F N S H K V Y I T A D K R K N G I K A K F K I R H N V K D G S V Q L A D H Y Q Q N T P I G R
GFP (+48) L R Y N F N S H K V Y I T A D K R K N G I K A K F K I R H N V K D G S V Q L A K H Y Q Q N T P I G R

图1a-1

GFP (-30)	GPVLLP	D	D	H	Y	I	L	S	T	E	S	A	L	S	K	D	P	N	E	D	R	D	H	M	V	L	L	E	F	V	T	A	A	G	I	D	H	G	M	D	E	L	Y	K	
GFP (-25)	GPVLLP	D	D	H	Y	I	L	S	T	E	S	A	L	S	K	D	P	N	E	D	R	D	H	M	V	L	L	L	E	F	V	T	A	A	G	I	D	H	G	M	D	E	L	Y	K
sfGFP	GPVLLP	D	N	H	Y	I	L	S	T	Q	S	A	L	S	K	D	P	N	E	K	R	D	H	M	V	L	L	L	E	F	V	T	A	A	G	I	T	H	G	M	D	E	L	Y	K
GFP (+36)	GPVLLP	R	N	H	Y	I	L	S	T	R	S	K	L	S	K	D	P	K	E	K	R	D	H	M	V	L	L	L	E	F	V	T	A	A	G	I	K	H	G	R	D	E	R	Y	K
GFP (+48)	GPVLLP	R	K	H	Y	I	L	S	T	R	S	K	L	S	K	D	P	K	E	K	R	D	H	M	V	L	L	L	E	F	V	T	A	A	G	I	K	H	G	R	K	E	R	Y	K

图1a-2

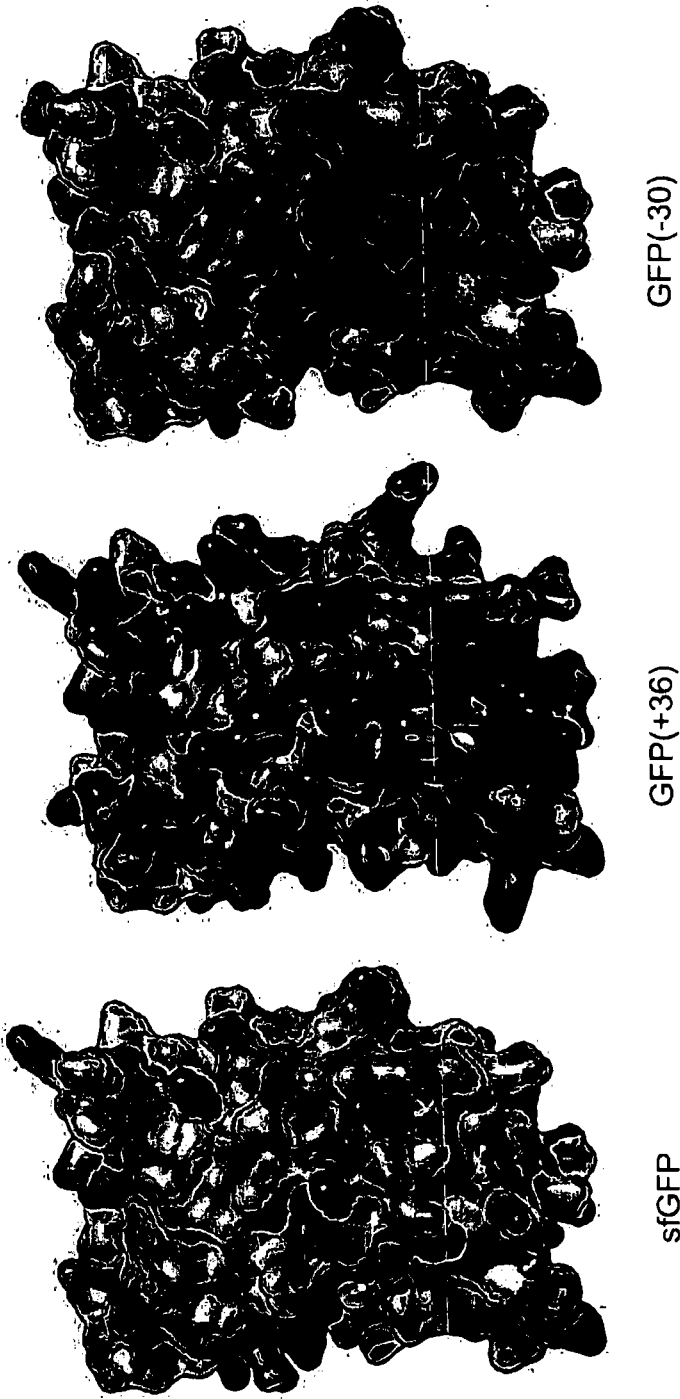


图1b

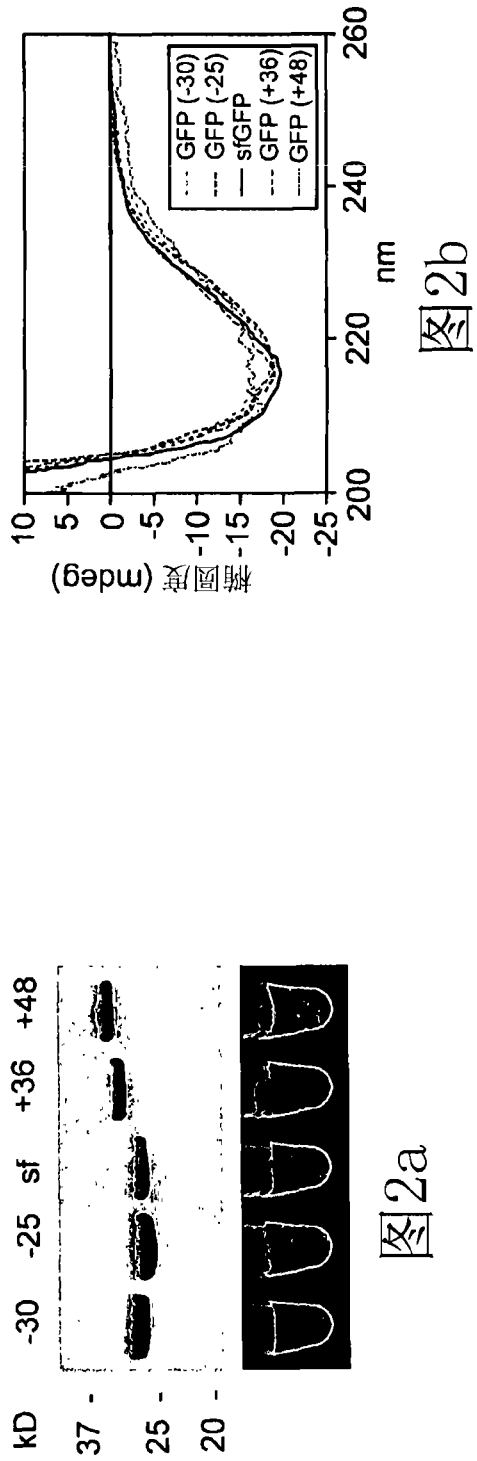


图2a

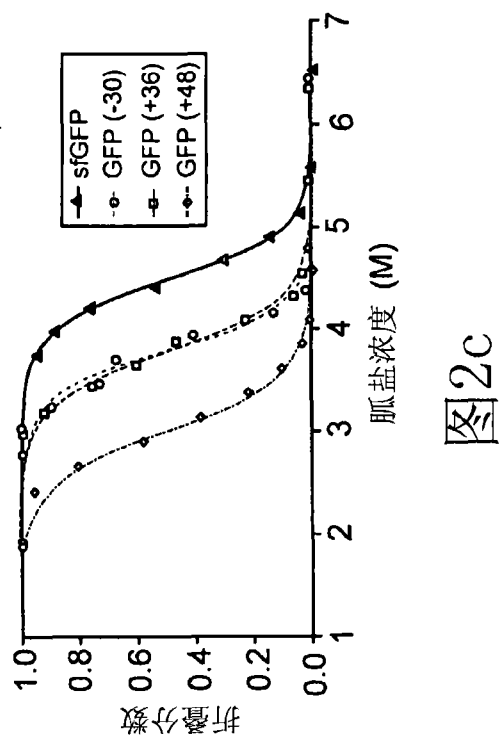


图2c

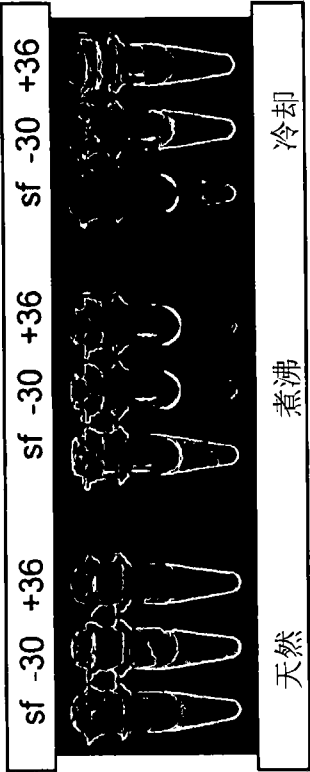


图3a

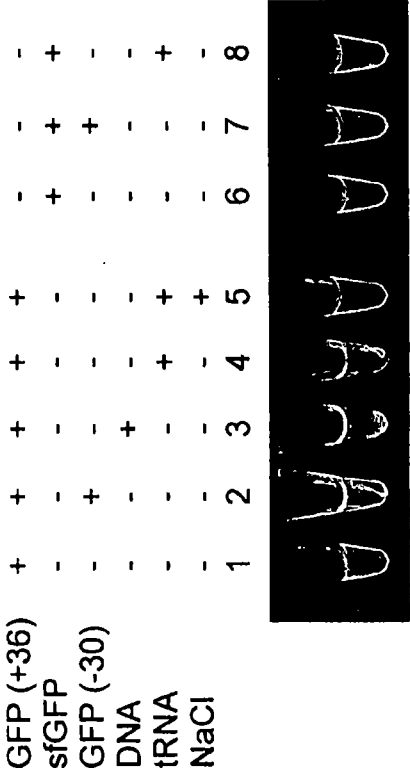


图3c

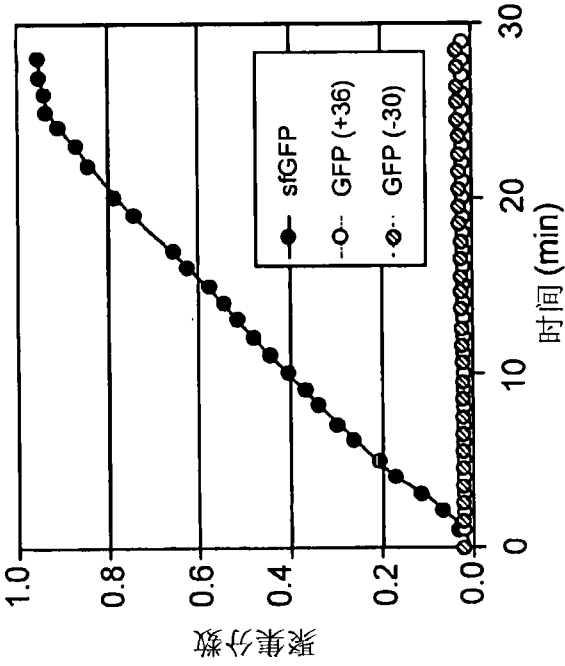


图3b

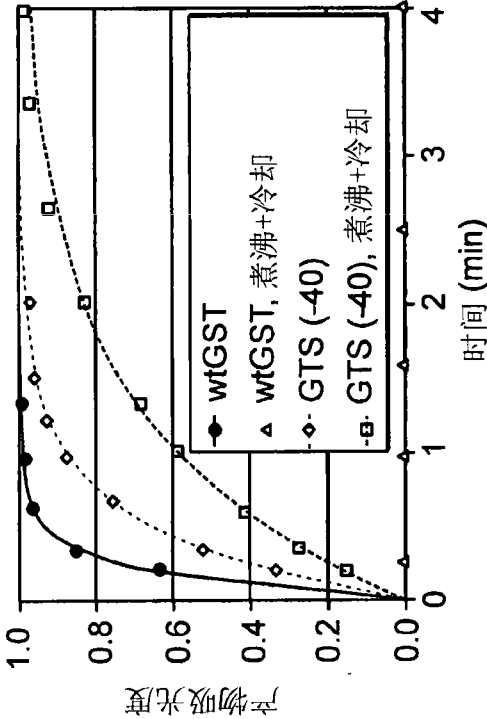


图3d

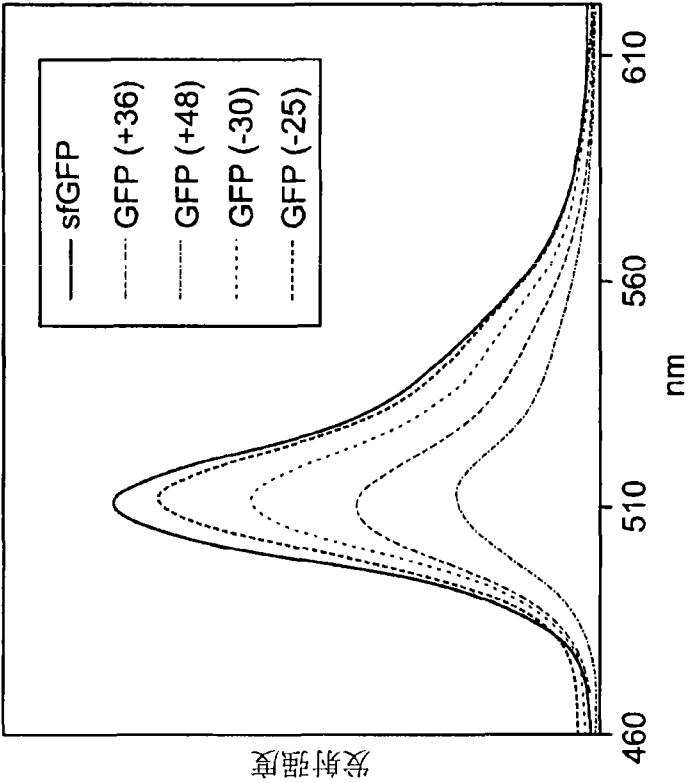


图4a

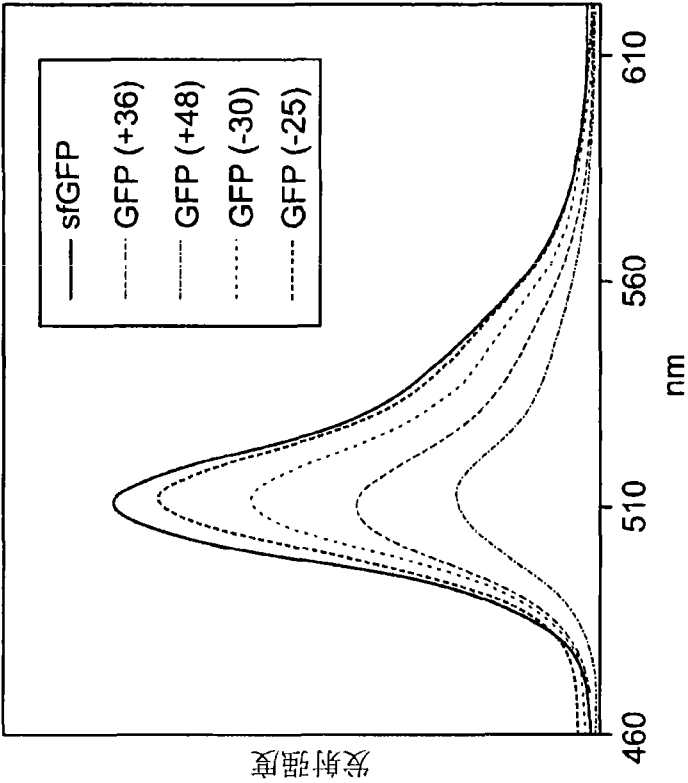


图4b

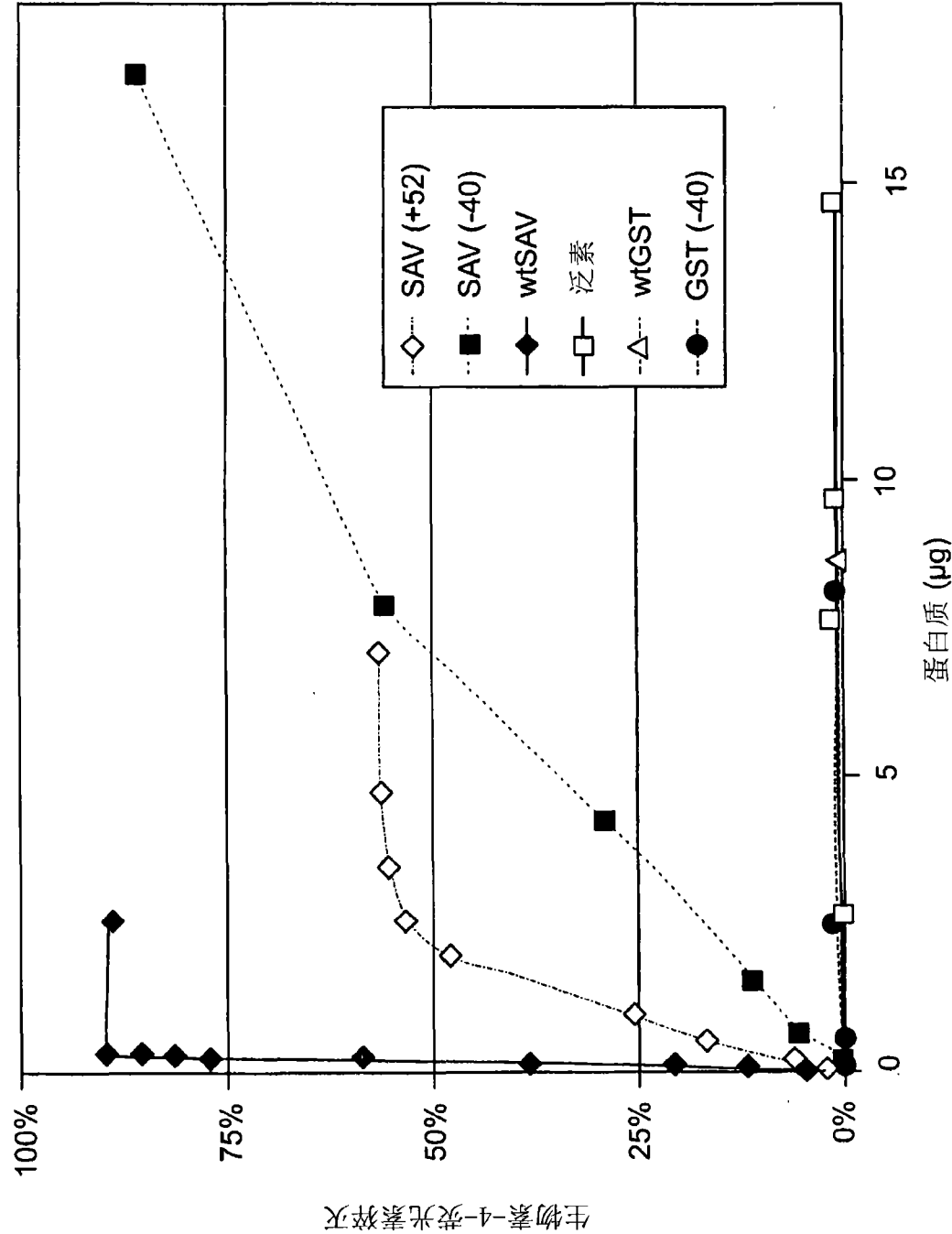


图5