

00462053003_REVISIED_SEQUENCE_LISTING

SEQUENCE LISTING

<110> Merck Patent GmbH
Zillmann, Martin
Orlando, Joe

<120> Soluble Intein Fusion Proteins And
Methods For Purifying Biomolecules

<130> 0046.2053-003

<140> 15/511,908

<141> 2017-03-16

<150> PCT/US2015/057125

<151> 2015-10-23

<150> US 62/074,494

<151> 2014-11-03

<150> US 62/209,010

<151> 2015-08-24

<160> 63

<170> FastSEQ for windows Version 4.0

<210> 1

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein with flanking non-intein sequences

<400> 1

```

Met Thr Arg Ser Gly Tyr Cys Leu Asp Leu Lys Thr Gln Val Gln Thr
 1          5          10          15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
          20          25          30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
          35          40          45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
          50          55          60
Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65          70          75          80
Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly
          85          90          95

```

<210> 2

<211> 96

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 N-intein variant with flanking non-intein
sequences

<400> 2

```

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

```

00462053003_REVISSED_SEQUENCE_LISTING

Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
 50 55 60
 Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
 65 70 75 80
 Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly
 85 90 95

<210> 3
 <211> 96
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> GP41-1 N-intein variant with flanking non-intein sequences

<400> 3
 Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
 1 5 10 15
 Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
 20 25 30
 Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
 35 40 45
 Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
 50 55 60
 Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
 65 70 75 80
 Ser Gly Gly Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu Gly Gly
 85 90 95

<210> 4
 <211> 96
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> GP41-1 N-intein variant with flanking non-intein sequences

<400> 4
 Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
 1 5 10 15
 Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
 20 25 30
 Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
 35 40 45
 Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
 50 55 60
 Cys Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
 65 70 75 80
 Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly
 85 90 95

<210> 5
 <211> 96
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> GP41-1 N-intein variant with flanking non-intein sequences

<400> 5
 Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
 1 5 10 15

00462053003_REVISIED_SEQUENCE_LISTING

```

Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
      20      25      30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
      35      40      45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
      50      55      60
Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
      65      70      75      80
Ser Gly Gly Leu Lys Glu Gly Met Met Leu Tyr Val Lys Glu Gly Gly
      85      90      95

```

<210> 6
 <211> 96
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> GP41-1 N-intein variant with flanking non-intein sequences

```

<400> 6
Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
  1      5      10      15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
      20      25      30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
      35      40      45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
      50      55      60
Thr Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
      65      70      75      80
Ser Gly Gly Leu Lys Glu Gly Met Met Leu Tyr Val Lys Glu Gly Gly
      85      90      95

```

<210> 7
 <211> 96
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> GP41-1 N-intein variant with flanking non-intein sequences

```

<400> 7
Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
  1      5      10      15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
      20      25      30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
      35      40      45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
      50      55      60
Ala Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
      65      70      75      80
Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly
      85      90      95

```

<210> 8
 <211> 96
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> GP41-1 N-intein variant with flanking non-intein sequences

00462053003_REVISSED_SEQUENCE_LISTING

<400> 8

```

Met Thr Arg Ser Gly Tyr Ala Leu Asp Leu Lys Thr Gln Val Gln Thr
 1          5          10          15
Pro Gln Gly Met Lys Glu Ile Ser Asn Ile Gln Val Gly Asp Leu Val
          20          25          30
Leu Ser Asn Thr Gly Tyr Asn Glu Val Leu Asn Val Phe Pro Lys Ser
          35          40          45
Lys Lys Lys Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile
          50          55          60
Thr Ser Glu Glu His Leu Phe Pro Thr Gln Thr Gly Glu Met Asn Ile
65          70          75          80
Ser Gly Gly Leu Lys Glu Gly Met Lys Leu Tyr Val Lys Glu Gly Gly
          85          90          95

```

<210> 9

<211> 42

<212> PRT

<213> cyanophage

<400> 9

```

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

```

<210> 10

<211> 157

<212> PRT

<213> Artificial Sequence

<220>

<223> GP41-1 C-intein-thioredoxin fusion protein

<400> 10

```

Met Gly Lys Asn Ser Met Met Leu Lys Lys Ile Leu Lys Ile Glu Glu
 1          5          10          15
Leu Asp Glu Arg Glu Leu Ile Asp Ile Glu Val Ser Gly Asn His Leu
          20          25          30
Phe Tyr Ala Asn Asp Ile Leu Thr His Asn Met Ser Asp Lys Ile Ile
          35          40          45
His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
          50          55          60
Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
65          70          75          80
Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
          85          90          95
Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
          100          105          110
Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
          115          120          125
Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
          130          135          140
Phe Leu Asp Ala Asn Leu Ala His His His His His His
145          150          155

```

<210> 11

<211> 51

<212> PRT

<213> E. coli

<400> 11

```

Met Arg Glu Tyr Pro Asn Gly Glu Lys Thr His Leu Thr Val Met Ala

```

00462053003_REVISIED_SEQUENCE_LISTING

```

1           5           10           15
Ala Gly Phe Pro Ser Leu Thr Gly Asp His Lys Val Ile Tyr Val Ala
20
Ala Asp Arg His Val Thr Ser Glu Ile Leu Glu Ala Ala Ile Arg
35
Leu Leu Ser
50

```

```

<210> 12
<211> 77
<212> PRT
<213> E. coli

```

```

<400> 12
Met Ser His Leu Asp Glu Val Ile Ala Arg Val Asp Ala Ala Ile Glu
1           5           10           15
Glu Ser Val Ile Ala His Met Asn Glu Leu Leu Ile Ala Leu Ser Asp
20
Asp Ala Glu Leu Ser Arg Glu Asp Arg Tyr Thr Gln Gln Gln Arg Leu
35
Arg Thr Ala Ile Ala His His Gly Arg Lys His Lys Glu Asp Met Glu
50
Ala Arg His Glu Gln Leu Thr Lys Gly Gly Thr Ile Leu
65           70           75

```

```

<210> 13
<211> 83
<212> PRT
<213> E. coli

```

```

<400> 13
Met Asn Lys Glu Thr Gln Pro Ile Asp Arg Glu Thr Leu Leu Lys Glu
1           5           10           15
Ala Asn Lys Ile Ile Arg Glu His Glu Asp Thr Leu Ala Gly Ile Glu
20
Ala Thr Gly Val Thr Gln Arg Asn Gly Val Leu Val Phe Thr Gly Asp
35
Tyr Phe Leu Asp Glu Gln Gly Leu Pro Thr Ala Lys Ser Thr Ala Val
50
Phe Asn Met Phe Lys His Leu Ala His Val Leu Ser Glu Lys Tyr His
65           70           75           80
Leu Val Asp

```

```

<210> 14
<211> 53
<212> PRT
<213> E. coli

```

```

<400> 14
Met Ser Leu Glu Asn Ala Pro Asp Asp Val Lys Leu Ala Val Asp Leu
1           5           10           15
Ile Val Leu Leu Glu Glu Asn Gln Ile Pro Ala Ser Thr Val Leu Arg
20
Ala Leu Asp Ile Val Lys Arg Asp Tyr Glu Lys Lys Leu Thr Arg Asp
35
Asp Glu Ala Glu Lys
50

```

```

<210> 15
<211> 69
<212> PRT
<213> E. coli

```

00462053003_REVISIED_SEQUENCE_LISTING

<400> 15

```

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1      5      10      15
Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile
      20      25      30
Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly
      35      40      45
Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr Arg
      50      55      60
Asn Glu Tyr Arg Trp
65

```

<210> 16

<211> 70

<212> PRT

<213> E. coli

<400> 16

```

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1      5      10      15
Val Lys Glu Gln Trp Gly Cys Lys Leu Thr Asp Asp Asp Met Thr Ile
      20      25      30
Ile Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr
      35      40      45
Gly Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr
      50      55      60
Arg Asn Glu Tyr Arg Trp
65      70

```

<210> 17

<211> 70

<212> PRT

<213> E. coli

<400> 17

```

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1      5      10      15
Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile
      20      25      30
Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly
      35      40      45
Cys Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu Thr
      50      55      60
Arg Asn Glu Tyr Arg Trp
65      70

```

<210> 18

<211> 71

<212> PRT

<213> E. coli

<400> 18

```

Met Asn Lys Asp Glu Ala Gly Gly Asn Trp Lys Gln Phe Lys Gly Lys
 1      5      10      15
Val Lys Glu Gln Trp Gly Lys Leu Thr Asp Asp Asp Met Thr Ile Ile
      20      25      30
Glu Gly Lys Arg Asp Gln Leu Val Gly Lys Ile Gln Glu Arg Tyr Gly
      35      40      45
Cys Gly Tyr Gln Lys Asp Gln Ala Glu Lys Glu Val Val Asp Trp Glu
      50      55      60
Thr Arg Asn Glu Tyr Arg Trp
65      70

```

<210> 19

00462053003_REVISSED_SEQUENCE_LISTING

<211> 92
 <212> PRT
 <213> E. coli

<400> 19
 Met Ile Ala Glu Phe Glu Ser Arg Ile Leu Ala Leu Ile Asp Gly Met
 1 5 10 15
 Val Asp His Ala Ser Asp Asp Glu Leu Phe Ala Ser Gly Tyr Leu Arg
 20 25 30
 Gly His Leu Thr Leu Ala Ile Ala Glu Leu Glu Ser Gly Asp Asp His
 35 40 45
 Ser Ala Gln Ala Val His Thr Thr Val Ser Gln Ser Leu Glu Lys Ala
 50 55 60
 Ile Gly Ala Gly Glu Leu Ser Pro Arg Asp Gln Ala Leu Val Thr Asp
 65 70 75 80
 Met Trp Glu Asn Leu Phe Gln Gln Ala Ser Gln Gln
 85 90

<210> 20
 <211> 95
 <212> PRT
 <213> E. coli

<400> 20
 Met Gln Leu Asn Ile Thr Gly Asn Asn Val Glu Ile Thr Glu Ala Leu
 1 5 10 15
 Arg Glu Phe Val Thr Ala Lys Phe Ala Lys Leu Glu Gln Tyr Phe Asp
 20 25 30
 Arg Ile Asn Gln Val Tyr Val Val Leu Lys Val Glu Lys Val Thr His
 35 40 45
 Thr Ser Asp Ala Thr Leu His Val Asn Gly Gly Glu Ile His Ala Ser
 50 55 60
 Ala Glu Gly Gln Asp Met Tyr Ala Ala Ile Asp Gly Leu Ile Asp Lys
 65 70 75 80
 Leu Ala Arg Gln Leu Thr Lys His Lys Asp Lys Leu Lys Gln His
 85 90 95

<210> 21
 <211> 192
 <212> PRT
 <213> E. coli

<400> 21
 Met Asp Thr Ser Asn Ala Thr Ser Val Val Asn Val Ser Ala Ser Ser
 1 5 10 15
 Ser Thr Ser Thr Ile Tyr Asp Leu Gly Asn Met Ser Lys Asp Glu Val
 20 25 30
 Val Lys Leu Phe Glu Glu Leu Gly Val Phe Gln Ala Ala Ile Leu Met
 35 40 45
 Phe Ser Tyr Met Tyr Gln Ala Gln Ser Asn Leu Ser Ile Ala Lys Phe
 50 55 60
 Ala Asp Met Asn Glu Ala Ser Lys Ala Ser Thr Thr Ala Gln Lys Met
 65 70 75 80
 Ala Asn Leu Val Asp Ala Lys Ile Ala Asp Val Gln Ser Ser Thr Asp
 85 90 95
 Lys Asn Ala Lys Ala Lys Leu Pro Gln Asp Val Ile Asp Tyr Ile Asn
 100 105 110
 Asp Pro Arg Asn Asp Ile Ser Val Thr Gly Ile Ser Asp Leu Ser Gly
 115 120 125
 Asp Leu Ser Ala Gly Asp Leu Gln Thr Val Lys Ala Ala Ile Ser Ala
 130 135 140
 Lys Ala Asn Asn Leu Thr Thr Val Val Asn Asn Ser Gln Leu Glu Ile
 145 150 155 160
 Gln Gln Met Ser Asn Thr Leu Asn Leu Leu Thr Ser Ala Arg Ser Asp
 165 170 175
 Val Gln Ser Leu Gln Tyr Arg Thr Ile Ser Ala Ile Ser Leu Gly Lys

180

185

190

<210> 22
 <211> 68
 <212> PRT
 <213> Fasciola hepatica

<400> 22
 Met Pro Ser Val Glu Val Glu Lys Leu Leu His Val Leu Asp Arg Asn
 1 5 10 15
 Gly Asp Gly Lys Val Ser Ala Glu Glu Leu Lys Ala Phe Ala Asp Asp
 20 25 30
 Ser Lys Tyr Pro Leu Asp Ser Asn Lys Ile Lys Ala Phe Ile Lys Glu
 35 40 45
 His Asp Lys Asn Lys Asp Gly Lys Leu Asp Leu Lys Glu Leu Val Ser
 50 55 60
 Ile Leu Ser Ser
 65

<210> 23
 <211> 11
 <212> PRT
 <213> Fasciola hepatica

<400> 23
 Met Pro Ser Val Glu Val Glu Lys Leu Leu His
 1 5 10

<210> 24
 <211> 328
 <212> PRT
 <213> E. coli

<400> 24
 Met Gly Gln Leu Ile Asp Gly Val Trp His Asp Thr Trp Tyr Asp Thr
 1 5 10 15
 Lys Ser Thr Gly Lys Phe Gln Arg Ser Ala Ser Ala Phe Arg Asn
 20 25 30
 Trp Leu Thr Ala Asp Gly Ala Pro Gly Pro Thr Gly Lys Gly Gly Phe
 35 40 45
 Ala Ala Glu Lys Asp Arg Tyr His Leu Tyr Val Ser Leu Ala Cys Pro
 50 55 60
 Trp Ala His Arg Thr Leu Ile Met Arg Lys Leu Lys Gly Leu Glu Pro
 65 70 75 80
 Phe Ile Ser Val Ser Val Val Asn Pro Leu Met Leu Glu Asn Gly Trp
 85 90 95
 Thr Phe Asp Asp Ser Phe Pro Gly Ala Thr Gly Asp Thr Leu Tyr Gln
 100 105 110
 His Glu Phe Leu Tyr Gln Leu Tyr Leu His Ala Asp Pro His Tyr Ser
 115 120 125
 Gly Arg Val Thr Val Pro Val Leu Trp Asp Lys Lys Asn His Thr Ile
 130 135 140
 Val Ser Asn Glu Ser Ala Glu Ile Ile Arg Met Phe Asn Thr Ala Phe
 145 150 155 160
 Asp Ala Leu Gly Ala Lys Ala Gly Asp Tyr Tyr Pro Pro Ala Leu Gln
 165 170 175
 Pro Lys Ile Asp Glu Leu Asn Gly Trp Ile Tyr Asp Thr Val Asn Asn
 180 185 190
 Gly Val Tyr Lys Ala Gly Phe Ala Thr Ser Gln Gln Ala Tyr Asp Glu
 195 200 205
 Ala Val Ala Lys Val Phe Glu Ser Leu Ala Arg Leu Glu Gln Ile Leu
 210 215 220
 Gly Gln His Arg Tyr Leu Thr Gly Asn Gln Leu Thr Glu Ala Asp Ile
 225 230 235 240
 Arg Leu Trp Thr Thr Leu Val Arg Phe Asp Pro Val Tyr Val Thr His

00462053003_REVISSED_SEQUENCE_LISTING

245 250 255
 Phe Lys Cys Asp Lys His Arg Ile Ser Asp Tyr Leu Asn Leu Tyr Gly
 260 265 270
 Phe Leu Arg Asp Ile Tyr Gln Met Pro Gly Ile Ala Glu Thr Val Asn
 275 280 285
 Phe Asp His Ile Arg Asn His Tyr Phe Arg Ser His Lys Thr Ile Asn
 290 295 300
 Pro Thr Gly Ile Ile Ser Ile Gly Pro Trp Gln Asp Leu Asp Glu Pro
 305 310 315 320
 His Gly Arg Asp Val Arg Phe Gly
 325

<210> 25
 <211> 120
 <212> PRT
 <213> Enterobacteria phage lambda

<400> 25
 Met Ala Ser Trp Ser His Pro Gln Phe Glu Lys Ala Ser Lys Glu Thr
 1 5 10 15
 Phe Thr His Tyr Gln Pro Gln Gly Asn Ser Asp Pro Ala His Thr Ala
 20 25 30
 Thr Ala Pro Gly Gly Leu Ser Ala Lys Ala Pro Ala Met Thr Pro Leu
 35 40 45
 Met Leu Asp Thr Ser Ser Arg Lys Leu Val Ala Trp Asp Gly Thr Thr
 50 55 60
 Asp Gly Ala Ala Val Gly Ile Leu Ala Val Ala Asp Gln Thr Ser
 65 70 75 80
 Thr Thr Leu Thr Phe Tyr Lys Ser Gly Thr Phe Arg Tyr Glu Asp Val
 85 90 95
 Leu Trp Pro Glu Ala Ala Ser Asp Glu Thr Lys Lys Arg Thr Ala Phe
 100 105 110
 Ala Gly Thr Ala Ile Ser Ile Val
 115 120

<210> 26
 <211> 396
 <212> PRT
 <213> E. coli

<400> 26
 Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
 1 5 10 15
 Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
 20 25 30
 Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
 35 40 45
 Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
 50 55 60
 His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
 65 70 75 80
 Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
 85 90 95
 Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
 100 105 110
 Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
 115 120 125
 Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
 130 135 140
 Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
 145 150 155 160
 Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
 165 170 175
 Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
 180 185 190
 Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly

00462053003_REVISSED_SEQUENCE_LISTING

```

195      200      205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210      215      220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225      230      235
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245      250      255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260      265      270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275      280      285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290      295      300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305      310      315
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325      330      335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys
340      345      350
Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala
355      360      365
Val Arg Thr Ala Val Ile Asn Ala Ala Ser Gly Arg Gln Thr Val Asp
370      375      380
Glu Ala Leu Lys Asp Ala Gln Thr Arg Ile Thr Lys
385      390      395

```

<210> 27
 <211> 109
 <212> PRT
 <213> E. coli

```

<400> 27
Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp
1      5      10      15
Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
20      25      30
Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
35      40      45
Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn
50      55      60
Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu
65      70      75      80
Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser
85      90      95
Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala
100      105

```

<210> 28
 <211> 495
 <212> PRT
 <213> E. coli

```

<400> 28
Met Asn Lys Glu Ile Leu Ala Val Val Glu Ala Val Ser Asn Glu Lys
1      5      10      15
Ala Leu Pro Arg Glu Lys Ile Phe Glu Ala Leu Glu Ser Ala Leu Ala
20      25      30
Thr Ala Thr Lys Lys Lys Tyr Glu Gln Glu Ile Asp Val Arg Val Gln
35      40      45
Ile Asp Arg Lys Ser Gly Asp Phe Asp Thr Phe Arg Arg Trp Leu Val
50      55      60
Val Asp Glu Val Thr Gln Pro Thr Lys Glu Ile Thr Leu Glu Ala Ala
65      70      75      80
Arg Tyr Glu Asp Glu Ser Leu Asn Leu Gly Asp Tyr Val Glu Asp Gln
85      90      95
Ile Glu Ser Val Thr Phe Asp Arg Ile Thr Thr Gln Thr Ala Lys Gln

```

00462053003_REVISSED_SEQUENCE_LISTING

```

100      105      110
Val Ile Val Gln Lys Val Arg Glu Ala Glu Arg Ala Met Val Val Asp
115
Gln Phe Arg Glu His Glu Gly Glu Ile Ile Thr Gly Val Val Lys Lys
130
Val Asn Arg Asp Asn Ile Ser Leu Asp Leu Gly Asn Asn Ala Glu Ala
145      150      155      160
Val Ile Leu Arg Glu Asp Met Leu Pro Arg Glu Asn Phe Arg Pro Gly
165
Asp Arg Val Arg Gly Val Leu Tyr Ser Val Arg Pro Glu Ala Arg Gly
180      185      190
Ala Gln Leu Phe Val Thr Arg Ser Lys Pro Glu Met Leu Ile Glu Leu
195
Phe Arg Ile Glu Val Pro Glu Ile Gly Glu Glu Val Ile Glu Ile Lys
210
Ala Ala Ala Arg Asp Pro Gly Ser Arg Ala Lys Ile Ala Val Lys Thr
225      230      235      240
Asn Asp Lys Arg Ile Asp Pro Val Gly Ala Cys Val Gly Met Arg Gly
245      250      255
Ala Arg Val Gln Ala Val Ser Thr Glu Leu Gly Gly Glu Arg Ile Asp
260      265      270
Ile Val Leu Trp Asp Asp Asn Pro Ala Gln Phe Val Ile Asn Ala Met
275      280      285
Ala Pro Ala Asp Val Ala Ser Ile Val Val Asp Glu Asp Lys His Thr
290      295      300
Met Asp Ile Ala Val Glu Ala Gly Asn Leu Ala Gln Ala Ile Gly Arg
305      310      315      320
Asn Gly Gln Asn Val Arg Leu Ala Ser Gln Leu Ser Gly Trp Glu Leu
325      330      335
Asn Val Met Thr Val Asp Asp Leu Gln Ala Lys His Gln Ala Glu Ala
340      345      350
His Ala Ala Ile Asp Thr Phe Thr Lys Tyr Leu Asp Ile Asp Glu Asp
355      360      365
Phe Ala Thr Val Leu Val Glu Gly Phe Ser Thr Leu Glu Glu Leu
370      375      380
Ala Tyr Val Pro Met Lys Glu Leu Leu Glu Ile Glu Gly Leu Asp Glu
385      390      395      400
Pro Thr Val Glu Ala Leu Arg Glu Arg Ala Lys Asn Ala Leu Ala Thr
405      410      415
Ile Ala Gln Ala Gln Glu Glu Ser Leu Gly Asp Asn Lys Pro Ala Asp
420      425      430
Asp Leu Leu Asn Leu Glu Gly Val Asp Arg Asp Leu Ala Phe Lys Leu
435      440      445
Ala Ala Arg Gly Val Cys Thr Leu Glu Asp Leu Ala Glu Gln Gly Ile
450      455      460
Asp Asp Leu Ala Asp Ile Glu Gly Leu Thr Asp Glu Lys Ala Gly Ala
465      470      475      480
Leu Ile Met Ala Ala Arg Asn Ile Cys Trp Phe Gly Asp Glu Ala
485      490      495

```

<210> 29

<211> 88

<212> PRT

<213> GP41-1 N-intein (cyanophage)

<400> 29

```

Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu
1      5      10
Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr
20      25      30
Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Ser Tyr Lys
35      40      45
Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu
50      55      60
Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu
65      70      75      80
Gly Met Cys Leu Tyr Val Lys Glu

```

<210> 30
 <211> 88
 <212> PRT
 <213> Unknown

<220>
 <223> N-terminal domain of GP41.8

<400> 30
 Cys Leu Ser Leu Asp Thr Met Val Val Thr Asn Gly Lys Ala Ile Glu
 1 5 10 15
 Ile Arg Asp Val Lys Val Gly Asp Trp Leu Glu Ser Glu Cys Gly Pro
 20 25 30
 Val Gln Val Thr Glu Val Leu Pro Ile Ile Lys Gln Pro Val Phe Glu
 35 40 45
 Ile Val Leu Lys Ser Gly Lys Lys Ile Arg Val Ser Ala Asn His Lys
 50 55 60
 Phe Pro Thr Lys Asp Gly Leu Lys Thr Ile Asn Ser Gly Leu Lys Val
 65 70 75 80
 Gly Asp Phe Leu Arg Ser Arg Ala
 85

<210> 31
 <211> 105
 <212> PRT
 <213> Unknown

<220>
 <223> N-terminal domain of NrdJ1

<400> 31
 Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr
 1 5 10 15
 Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln
 20 25 30
 Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
 35 40 45
 Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu
 50 55 60
 Ile Asn Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His
 65 70 75 80
 Pro Val Tyr Thr Lys Asn Arg Asp Tyr Val Arg Ala Asp Glu Leu Thr
 85 90 95
 Asp Asp Asp Glu Leu Val Val Ala Ile
 100 105

<210> 32
 <211> 101
 <212> PRT
 <213> Unknown

<220>
 <223> N-terminal domain of IMPDH1

<400> 32
 Cys Phe Val Pro Gly Thr Leu Val Asn Thr Glu Asn Gly Leu Lys Lys
 1 5 10 15
 Ile Glu Glu Ile Lys Val Gly Asp Lys Val Phe Ser His Thr Gly Lys
 20 25 30
 Leu Gln Glu Val Val Asp Thr Leu Ile Phe Asp Arg Asp Glu Glu Ile
 35 40 45
 Ile Ser Ile Asn Gly Ile Asp Cys Thr Lys Asn His Glu Phe Tyr Val
 50 55 60

00462053003_REVISSED_SEQUENCE_LISTING

Ile Asp Lys Glu Asn Ala Asn Arg Val Asn Glu Asp Asn Ile His Leu
 65 70 75 80
 Phe Ala Arg Trp Val His Ala Glu Glu Leu Asp Met Lys Lys His Leu
 85 90 95
 Leu Ile Glu Leu Glu
 100

<210> 33
 <211> 106
 <212> PRT
 <213> Unknown

<220>
 <223> N-terminal domain of NrdA-2

<400> 33
 Cys Leu Thr Gly Asp Ala Lys Ile Asp Val Leu Ile Asp Asn Ile Pro
 1 5 10 15
 Ile Ser Gln Ile Ser Leu Glu Glu Val Val Asn Leu Phe Asn Glu Gly
 20 25 30
 Lys Glu Ile Tyr Val Leu Ser Tyr Asn Ile Asp Thr Lys Glu Val Glu
 35 40 45
 Tyr Lys Glu Ile Ser Asp Ala Gly Leu Ile Ser Glu Ser Ala Glu Val
 50 55 60
 Leu Glu Ile Ile Asp Glu Glu Thr Gly Gln Lys Ile Val Cys Thr Pro
 65 70 75 80
 Asp His Lys Val Tyr Thr Leu Asn Arg Gly Tyr Val Ser Ala Lys Asp
 85 90 95
 Leu Lys Glu Asp Asp Glu Leu Val Phe Ser
 100 105

<210> 34
 <211> 102
 <212> PRT
 <213> Nostoc punctiforme

<400> 34
 Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu
 1 5 10 15
 Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser
 20 25 30
 Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His
 35 40 45
 Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
 50 55 60
 Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln
 65 70 75 80
 Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg
 85 90 95
 Val Asp Asn Leu Pro Asn
 100

<210> 35
 <211> 105
 <212> PRT
 <213> Synechocystis

<400> 35
 Cys Ile Ser Gly Asp Ser Leu Ile Ser Leu Ala Ser Thr Gly Lys Arg
 1 5 10 15
 Val Ser Ile Lys Asp Leu Leu Asp Glu Lys Asp Phe Glu Ile Trp Ala
 20 25 30
 Ile Asn Glu Gln Thr Met Lys Leu Glu Ser Ala Lys Val Ser Arg Val
 35 40 45
 Phe Cys Thr Gly Lys Lys Leu Val Tyr Ile Leu Lys Thr Arg Leu Gly

00462053003_REVISSED_SEQUENCE_LISTING

```

      50      55      60
Arg Thr Ile Lys Ala Thr Ala Asn His Arg Phe Leu Thr Ile Asp Gly
65      70      75      80
Trp Lys Arg Leu Asp Glu Leu Ser Leu Lys Glu His Ile Ala Leu Pro
      85      90      95
Arg Lys Leu Glu Ser Ser Ser Leu Gln
      100      105

```

<210> 36
 <211> 45
 <212> PRT
 <213> Unknown

<220>
 <223> C-terminal domain of GP41.8

```

<400> 36
Met Cys Glu Ile Phe Glu Asn Glu Ile Asp Trp Asp Glu Ile Ala Ser
 1      5      10      15
Ile Glu Tyr Val Gly Val Glu Glu Thr Ile Asp Ile Asn Val Thr Asn
      20      25      30
Asp Arg Leu Phe Phe Ala Asn Gly Ile Leu Thr His Asn
      35      40      45

```

<210> 37
 <211> 40
 <212> PRT
 <213> Unknown

<220>
 <223> C-terminal domain of NrdJ1

```

<400> 37
Met Glu Ala Lys Thr Tyr Ile Gly Lys Leu Lys Ser Arg Lys Ile Val
 1      5      10      15
Ser Asn Glu Asp Thr Tyr Asp Ile Gln Thr Ser Thr His Asn Phe Phe
      20      25      30
Ala Asn Asp Ile Leu Val His Asn
      35      40

```

<210> 38
 <211> 40
 <212> PRT
 <213> Unknown

<220>
 <223> C-terminal domain of IMPDH1

```

<400> 38
Met Lys Phe Lys Leu Lys Glu Ile Thr Ser Ile Glu Thr Lys His Tyr
 1      5      10      15
Lys Gly Lys Val His Asp Leu Thr Val Asn Gln Asp His Ser Tyr Asn
      20      25      30
Val Arg Gly Thr Val Val His Asn
      35      40

```

<210> 39
 <211> 34
 <212> PRT
 <213> Unknown

<220>
 <223> C-terminal domain of NrdA-2

00462053003_REVISIED_SEQUENCE_LISTING

<400> 39

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

<210> 40

<211> 36

<212> PRT

<213> Nostoc punctiforme

<400> 40

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

<210> 41

<211> 48

<212> PRT

<213> Synechocystis

<400> 41

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

<210> 42

<211> 27

<212> PRT

<213> Unknown

<220>

<223> GP41-2 N-intein sequence

<400> 42

Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Gln Gln Gly Leu Lys Asp
 1 5 10 15
 Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu
 20 25

<210> 43

<211> 46

<212> PRT

<213> Unknown

<220>

<223> GP41-3 N-intein sequence

<400> 43

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

00462053003_REVISIED_SEQUENCE_LISTING

<210> 44
 <211> 88
 <212> PRT
 <213> Unknown

<220>
 <223> GP41-4 N-intein sequence

<400> 44
 Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu
 1 5 10 15
 Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr
 20 25 30
 Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Ser Tyr Lys
 35 40 45
 Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu
 50 55 60
 Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu
 65 70 75 80
 Gly Met Cys Leu Tyr Val Lys Glu
 85

<210> 45
 <211> 88
 <212> PRT
 <213> Unknown

<220>
 <223> GP41-5 N-intein sequence

<400> 45
 Cys Leu Asp Leu Lys Thr Gln Val Gln Thr Pro Gln Gly Met Lys Glu
 1 5 10 15
 Ile Ser Asn Ile Gln Val Gly Asp Leu Val Leu Ser Asn Thr Gly Tyr
 20 25 30
 Asn Glu Val Leu Asn Val Phe Pro Lys Ser Lys Lys Ser Tyr Lys
 35 40 45
 Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu Glu His Leu
 50 55 60
 Phe Pro Thr Gln Thr Gly Glu Met Asn Ile Ser Gly Gly Leu Lys Glu
 65 70 75 80
 Gly Met Cys Leu Tyr Val Lys Glu
 85

<210> 46
 <211> 43
 <212> PRT
 <213> Unknown

<220>
 <223> GP41-6 N-intein sequence

<400> 46
 Ser Tyr Lys Ile Thr Leu Glu Asp Gly Lys Glu Ile Ile Cys Ser Glu
 1 5 10 15
 Glu His Leu Phe Pro Thr Gln Asn Gly Glu Val Asn Ile Lys Gly Gly
 20 25 30
 Leu Lys Glu Gly Met Cys Leu Tyr Val Lys Glu
 35 40

<210> 47
 <211> 26
 <212> PRT
 <213> Unknown

00462053003_REVISIED_SEQUENCE_LISTING

<220>

<223> GP41-7 N-intein sequence

<400> 47

Met Met Leu Lys Lys Ile Leu Lys Ile Glu Glu Leu Asp Glu Arg Glu
 1 5 10 15
 Leu Ile Asp Ile Glu Val Ser Gly Asn His
 20 25

<210> 48

<211> 133

<212> PRT

<213> Unknown

<220>

<223> NrdA-1 N-intein sequence

<400> 48

Cys Val Ala Gly Asp Thr Lys Ile Lys Ile Lys Tyr Pro Glu Ser Val
 1 5 10 15
 Gly Asp Gln Tyr Gly Thr Trp Tyr Trp Asn Val Leu Glu Lys Glu Ile
 20 25 30
 Gln Ile Glu Asp Leu Glu Asp Tyr Ile Ile Met Arg Glu Cys Glu Ile
 35 40 45
 Tyr Asp Ser Asn Ala Pro Gln Ile Glu Val Leu Ser Tyr Asn Ile Glu
 50 55 60
 Thr Gly Glu Gln Glu Trp Lys Pro Ile Thr Ala Phe Ala Gln Thr Ser
 65 70 75 80
 Pro Lys Ala Lys Val Met Lys Ile Thr Asp Glu Glu Ser Gly Lys Ser
 85 90 95
 Ile Val Val Thr Pro Glu His Gln Val Phe Thr Lys Asn Arg Gly Tyr
 100 105 110
 Val Met Ala Lys Asp Leu Ile Glu Thr Asp Glu Pro Ile Ile Val Asn
 115 120 125
 Lys Asp Met Asn Phe
 130

<210> 49

<211> 105

<212> PRT

<213> Unknown

<220>

<223> NrdA-4 N-intein sequence

<400> 49

Cys Leu Ala Gly Asp Thr Thr Val Thr Val Leu Glu Gly Asp Ile Val
 1 5 10 15
 Phe Glu Met Thr Leu Glu Asn Leu Val Ser Leu Tyr Lys Asn Val Phe
 20 25 30
 Ser Val Ser Val Leu Ser Phe Asn Pro Glu Thr Gln Lys Gln Glu Phe
 35 40 45
 Lys Pro Val Thr Asn Ala Ala Leu Met Asn Pro Glu Ser Lys Val Leu
 50 55 60
 Lys Ile Thr Asp Ser Asp Thr Gly Lys Ser Ile Val Cys Thr Pro Asp
 65 70 75 80
 His Lys Val Phe Thr Lys Asn Arg Gly Tyr Val Ile Ala Ser Glu Leu
 85 90 95
 Asn Ala Glu Asp Ile Leu Glu Ile Lys
 100 105

<210> 50

<211> 65

<212> PRT

00462053003_REVISIED_SEQUENCE_LISTING

<213> Unknown

<220>

<223> NrdA-5 N-intein sequence

<400> 50

```

His Thr Glu Thr Val Arg Arg Val Gly Thr Ile Thr Ala Phe Ala Gln
 1           5           10           15
Thr Ser Pro Lys Ser Lys Val Met Lys Ile Thr Asp Glu Glu Ser Gly
          20           25           30
Asn Ser Ile Val Val Thr Pro Glu His Lys Val Phe Thr Lys Asn Arg
          35           40           45
Gly Tyr Val Met Ala Lys Asn Leu Val Glu Thr Asp Glu Leu Val Ile
 50           55           60
Asn
65

```

<210> 51

<211> 49

<212> PRT

<213> Unknown

<220>

<223> NrdA-6 N-intein sequence

<400> 51

```

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

```

<210> 52

<211> 105

<212> PRT

<213> Unknown

<220>

<223> NrdJ-1 N-intein sequence

<400> 52

```

Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr
 1           5           10           15
Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln
          20           25           30
Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
          35           40           45
Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu
          50           55           60
Ile Asp Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His
          65           70           75           80
Pro Val Tyr Thr Lys Asn Arg Gly Tyr Val Arg Ala Asp Glu Leu Thr
          85           90           95
Asp Asp Asp Glu Leu Val Val Ala Ile
          100          105

```

<210> 53

<211> 105

<212> PRT

<213> Unknown

<220>

<223> NrdJ2 N-intein sequence

<400> 53

```

Cys Leu Val Gly Ser Ser Glu Ile Ile Thr Arg Asn Tyr Gly Lys Thr
 1          5          10          15
Thr Ile Lys Glu Val Val Glu Ile Phe Asp Asn Asp Lys Asn Ile Gln
          20          25          30
Val Leu Ala Phe Asn Thr His Thr Asp Asn Ile Glu Trp Ala Pro Ile
          35          40          45
Lys Ala Ala Gln Leu Thr Arg Pro Asn Ala Glu Leu Val Glu Leu Glu
          50          55          60
Ile Asn Thr Leu His Gly Val Lys Thr Ile Arg Cys Thr Pro Asp His
65          70          75          80
Pro Val Tyr Thr Lys Asn Arg Asp Tyr Val Arg Ala Asp Glu Leu Thr
          85          90          95
Asp Asp Asp Glu Leu Val Val Ala Ile
          100          105

```

<210> 54

<211> 47

<212> PRT

<213> Unknown

<220>

<223> GP41-9 C-intein sequence

<400> 54

```

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

```

<210> 55

<211> 42

<212> PRT

<213> Unknown

<220>

<223> IMPDH-2 C-intein sequence

<400> 55

```

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

```

<210> 56

<211> 41

<212> PRT

<213> Unknown

<220>

<223> IMPDH-3 C-intein sequence

<400> 56

```

Met Lys Phe Lys Leu Val Glu Ile Thr Ser Lys Glu Thr Phe Asn Tyr
 1          5          10          15
Ser Gly Gln Val His Asp Leu Thr Val Glu Asp Asp His Ser Tyr Ser
          20          25          30
Ile Asn Asn Ile Val Val His Asn Ser
          35          40

```

00462053003_REVISSED_SEQUENCE_LISTING

<210> 57
 <211> 34
 <212> PRT
 <213> Unknown

<220>
 <223> NrdA-3 C-intein sequence

<400> 57
 Met Leu Lys Ile Glu Tyr Leu Glu Glu Glu Ile Pro Val Tyr Asp Ile
 1 5 10 15
 Thr Val Glu Glu Thr His Asn Phe Phe Ala Asn Asp Ile Leu Ile His
 20 25 30
 Asn Cys

<210> 58
 <211> 28
 <212> PRT
 <213> Unknown

<220>
 <223> NrdA-5 C-intein sequence

<400> 58
 Met Leu Lys Ile Glu Tyr Leu Glu Glu Glu Ile Pro Val Tyr Asp Ile
 1 5 10 15
 Thr Val Glu Gly Thr His Asn Leu Ala Tyr Ser Leu
 20 25

<210> 59
 <211> 33
 <212> PRT
 <213> Unknown

<220>
 <223> NrdA-6 C-intein sequence

<400> 59
 Met Gly Ile Lys Ile Arg Lys Leu Glu Gln Asn Arg Val Tyr Asp Ile
 1 5 10 15
 Lys Val Glu Lys Ile Ile Ile Phe Cys Asn Asn Ile Leu Val His Asn
 20 25 30
 Cys

<210> 60
 <211> 41
 <212> PRT
 <213> Unknown

<220>
 <223> NrdJ-1 C-intein sequence

<400> 60
 Met Glu Ala Lys Thr Tyr Ile Gly Lys Leu Lys Ser Arg Lys Ile Val
 1 5 10 15
 Ser Asn Glu Asp Thr Tyr Asp Ile Gln Thr Ser Thr His Asn Phe Phe
 20 25 30
 Ala Asn Asp Ile Leu Val His Asn Ser
 35 40

00462053003_REVISIED_SEQUENCE_LISTING

<210> 61
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Loop region of E. coli

<400> 61
 Gly Cys Lys Leu
 1

<210> 62
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Loop region of E. coli

<400> 62
 Gly Cys Tyr Gln
 1

<210> 63
 <211> 5
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Loop region of E. coli

<400> 63
 Gly Cys Gly Tyr Gln
 1 5