Electronic Supplementary Material

Article Title: Codon Optimization Significantly Enhanced the Expression of Human 37 kDa iLRP in *Escherichia coli*

Journal Name: 3 Biotech

Authors: Bainan Liu¶, Qianqian Kong¶, Dong Zhang, Lingli Yan.

Affiliations: Department of Immunology, Zunyi Medical University, Zunyi, Guizhou Province 563000, China

Corresponding Author: Bainan Liu, Ph.D.

E-mail: bnliu@hotmail.com

Phone: +86-851-28642716

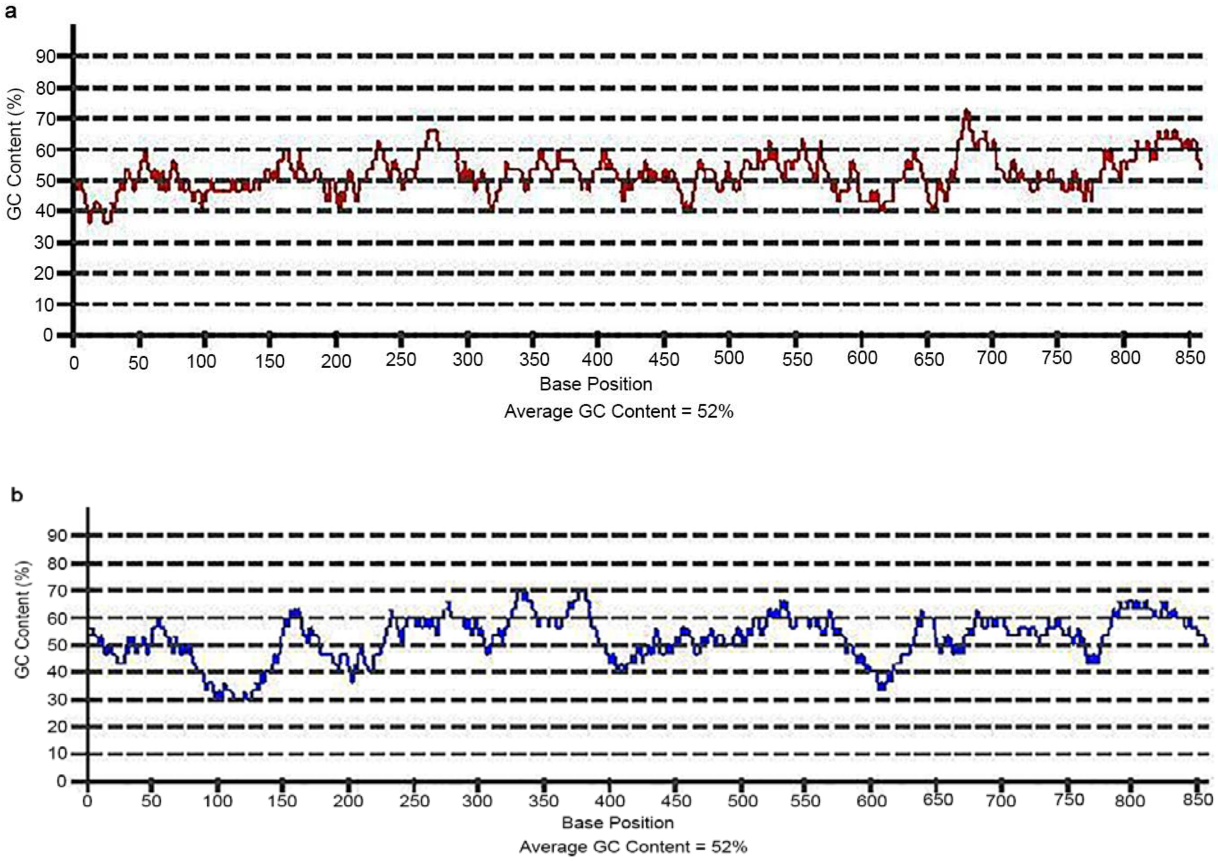


Fig. S1 The GC content of optimized gene and wild-type gene. Y axis represents local GC content calculated using the window of 30 base pairs, and X axis represents base position. Although the average value of GC content for both genes is 52%, a smoother variation curve of GC content was obtained after optimization, with significant changes at base positions 50 to 100 and 300 to 400. The figure part **a** denotes the variation curve for the optimized *iLRP* gene, and the part **b** for the wild gene

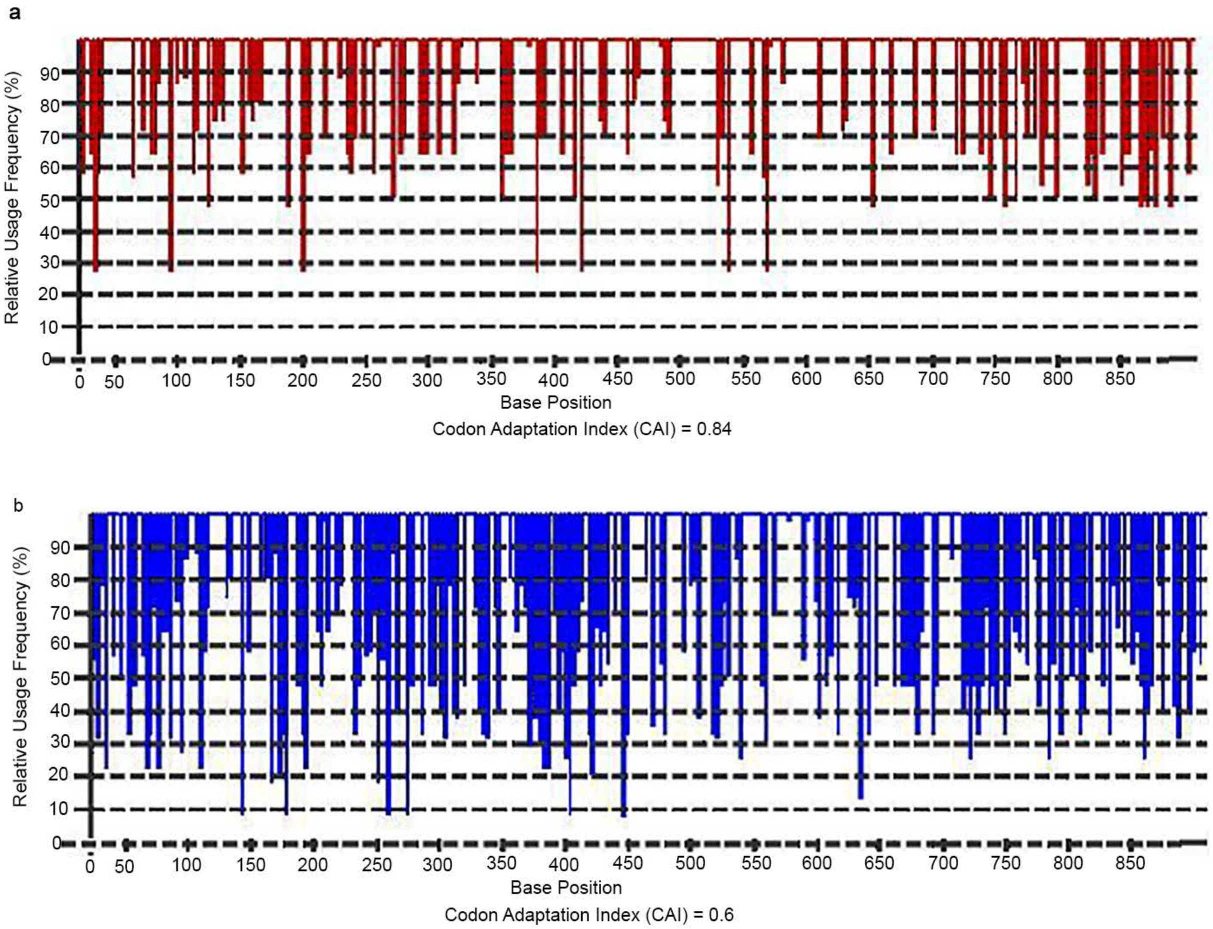
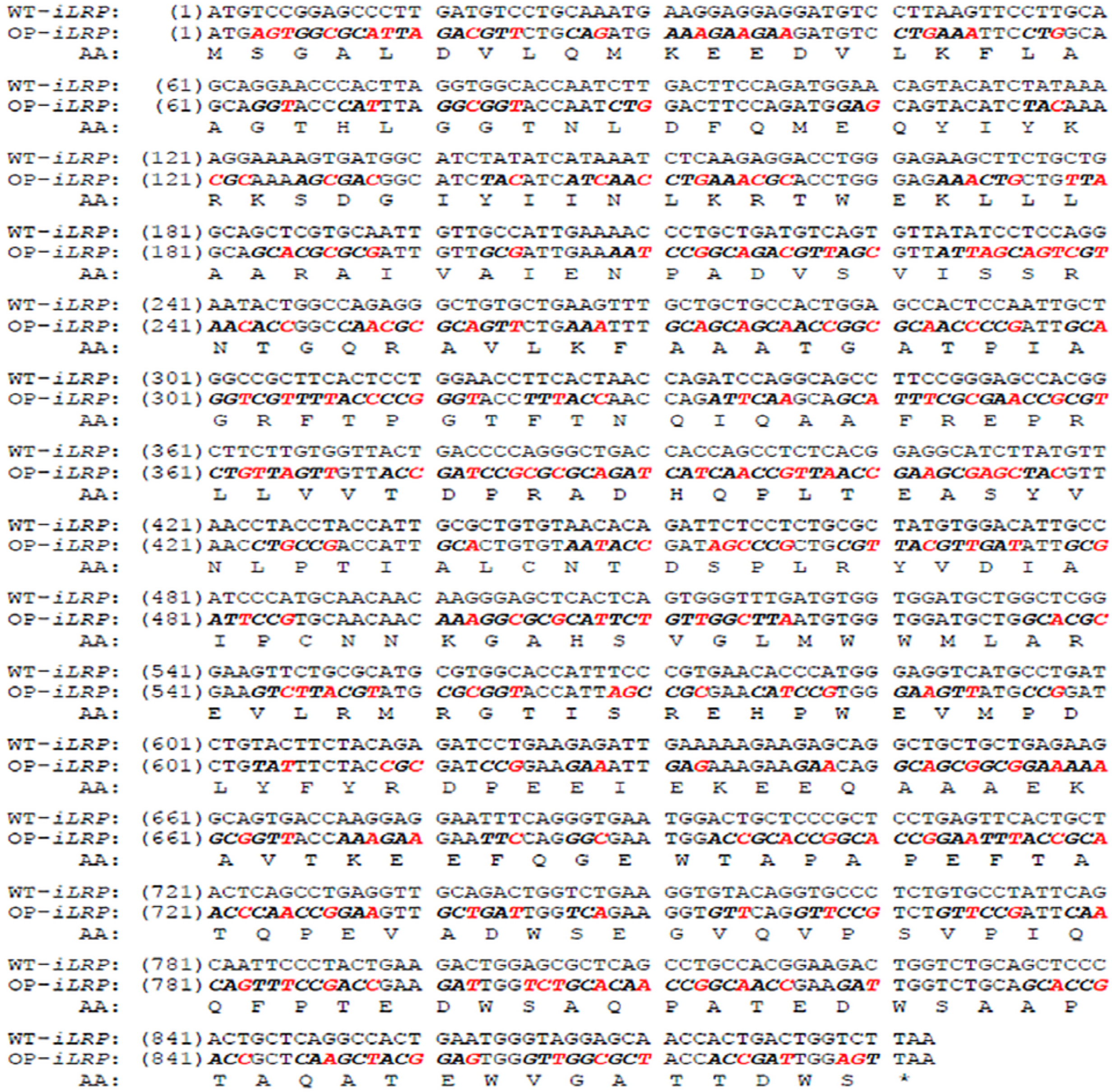


Fig. S2 Analysis of codon usage bias by Codon Adaptation Index (CAI). CAI measures the deviation of a given protein coding sequence according to a reference set of genes. It provides an indication of gene expression level with a value range from 0 to 1. A higher value predicts a higher expression level. The figure part **a** represents the plot of optimized *iLRP* gene with CAI value of 0.84, and the part **b** represents the plot of wild *iLRP* gene with CAI value of 0.6



```
WT-iLRP:    (1)ATGTCCGGAGCCCTT GATGTCCTGCAAATG AAGGAGGAGGATGTC CTTAAGTTCCTTGCA
OP-iLRP:    (1)ATGAGTGGCGCATTA GACGTTCTGCAGATG AAAGAAGAAGATGTC CTGAAATTCCTGGCA
     AA:        M  S  G  A  L   D  V  L  Q  M   K  E  E  D  V   L  K  F  L  A

WT-iLRP:   (61)GCAGGAACCCACTTA GGTGGCACCAATCTT GACTTCCAGATGGAA CAGTACATCTATAAA
OP-iLRP:   (61)GCAGGTACCCATTTA GGCGGTACCAATCTG GACTTCCAGATGGAG CAGTACATCTACAAA
     AA:        A  G  T  H  L   G  G  T  N  L   D  F  Q  M  E   Q  Y  I  Y  K

WT-iLRP:  (121)AGGAAAAGTGATGGC ATCTATATCATAAAT CTCAAGAGGACCTGG GAGAAGCTTCTGCTG
OP-iLRP:  (121)CGCAAAAGCGACGGC ATCTACATCATCAAC CTGAAACGCACCTGG GAGAAACTGCTGTTA
     AA:        R  K  S  D  G   I  Y  I  I  N   L  K  R  T  W   E  K  L  L  L

WT-iLRP:  (181)GCAGCTCGTGCAATT GTTGCCATTGAAAAC CCTGCTGATGTCAGT GTTATATCCTCCAGG
OP-iLRP:  (181)GCAGCACGCGCGATT GTTGCGATTGAAAAT CCGGCAGACGTTAGC GTTATTAGCAGTCGT
     AA:        A  A  R  A  I   V  A  I  E  N   P  A  D  V  S   V  I  S  S  R

WT-iLRP:  (241)AATACTGGCCAGAGG GCTGTGCTGAAGTTT GCTGCTGCCACTGGA GCCACTCCAATTGCT
OP-iLRP:  (241)AACACCGGCCAACGC GCAGTTCTGAAATTT GCAGCAGCAACCGGC GCAACCCCGATTGCA
     AA:        N  T  G  Q  R   A  V  L  K  F   A  A  A  T  G   A  T  P  I  A

WT-iLRP:  (301)GGCCGCTTCACTCCT GGAACCTTCACTAAC CAGATCCAGGCAGCC TTCCGGGAGCCACGG
OP-iLRP:  (301)GGTCGTTTTACCCCG GGTACCTTTACCAAC CAGATTCAAGCAGCA TTTCGCGAACCGCGT
     AA:        G  R  F  T  P   G  T  F  T  N   Q  I  Q  A  A   F  R  E  P  R

WT-iLRP:  (361)CTTCTTGTGGTTACT GACCCCAGGGCTGAC CACCAGCCTCTCACG GAGGCATCTTATGTT
OP-iLRP:  (361)CTGTTAGTTGTTACC GATCCGCGCGCAGAT CATCAACCGTTAACC GAAGCGAGCTACGTT
     AA:        L  L  V  V  T   D  P  R  A  D   H  Q  P  L  T   E  A  S  Y  V

WT-iLRP:  (421)AACCTACCTACCATT GCGCTGTGTAACACA GATTCTCCTCTGCGC TATGTGGACATTGCC
OP-iLRP:  (421)AACCTGCCGACCATT GCACTGTGTAATACC GATAGCCCGCTGCGT TACGTTGATATTGCG
     AA:        N  L  P  T  I   A  L  C  N  T   D  S  P  L  R   Y  V  D  I  A

WT-iLRP:  (481)ATCCCATGCAACAAC AAGGGAGCTCACTCA GTGGGTTTGATGTGG TGGATGCTGGCTCGG
OP-iLRP:  (481)ATTCCGTGCAACAAC AAAGGCGCGCATTCT GTTGGCTTAATGTGG TGGATGCTGGCACGC
     AA:        I  P  C  N  N   K  G  A  H  S   V  G  L  M  W   W  M  L  A  R

WT-iLRP:  (541)GAAGTTCTGCGCATG CGTGGCACCATTTCC CGTGAACACCCATGG GAGGTCATGCCTGAT
OP-iLRP:  (541)GAAGTCTTACGTATG CGCGGTACCATTAGC CGCGAACATCCGTGG GAAGTTATGCCGGAT
     AA:        E  V  L  R  M   R  G  T  I  S   R  E  H  P  W   E  V  M  P  D

WT-iLRP:  (601)CTGTACTTCTACAGA GATCCTGAAGAGATT GAAAAAGAAGAGCAG GCTGCTGCTGAGAAG
OP-iLRP:  (601)CTGTATTTCTACCGC GATCCGGAAGAAATT GAGAAAGAAGAACAG GCAGCGGCGGAAAAA
     AA:        L  Y  F  Y  R   D  P  E  E  I   E  K  E  E  Q   A  A  A  E  K

WT-iLRP:  (661)GCAGTGACCAAGGAG GAATTTCAGGGTGAA TGGACTGCTCCCGCT CCTGAGTTCACTGCT
OP-iLRP:  (661)GCGGTTACCAAAGAA GAATTCCAGGGCGAA TGGACCGCACCGGCA CCGGAATTTACCGCA
     AA:        A  V  T  K  E   E  F  Q  G  E   W  T  A  P  A   P  E  F  T  A

WT-iLRP:  (721)ACTCAGCCTGAGGTT GCAGACTGGTCTGAA GGTGTACAGGTGCCC TCTGTGCCTATTCAG
OP-iLRP:  (721)ACCCAACCGGAAGTT GCTGATTGGTCAGAA GGTGTTCAGGTTCCG TCTGTTCCGATTCAA
     AA:        T  Q  P  E  V   A  D  W  S  E   G  V  Q  V  P   S  V  P  I  Q

WT-iLRP:  (781)CAATTCCCTACTGAA GACTGGAGCGCTCAG CCTGCCACGGAAGAC TGGTCTGCAGCTCCC
OP-iLRP:  (781)CAGTTTCCGACCGAA GATTGGTCTGCACAA CCGGCAACCGAAGAT TGGTCTGCAGCACCG
     AA:        Q  F  P  T  E   D  W  S  A  Q   P  A  T  E  D   W  S  A  A  P

WT-iLRP:  (841)ACTGCTCAGGCCACT GAATGGGTAGGAGCA ACCACTGACTGGTCT TAA
OP-iLRP:  (841)ACCGCTCAAGCTACG GAGTGGGTTGGCGCT ACCACCGATTGGAGT TAA
     AA:        T  A  Q  A  T   E  W  V  G  A   T  T  D  W  S   *
```

Fig. S3 Sequence comparison between the optimized *iLRP* gene and wild *iLRP* gene. OP: codon optimized; WT: wild type; AA: amino acid. Differences in synonymous codons between the wild and optimized gene were indicated by bold and italic letters. Individual bases in red represent the base replacement with bacterial preferred ones. Translated amino acids are shown in standard single-letter. There is no difference in amino acid sequences between two DNA sequences after translation