# GeneCrafter v1.1 Report for COL1A2

## Result Summary

| Metric | Before | After |
|---|---|---|
| CAI Score | N/A | **0.8513** |
| GC Content (%) | 76.4 | **63.0** ▼ |
| Complex DNA Repeats | 1395 | **21** ▼ |
| Immunogenicity Score | 0.09 | **0.08** |

### Interpretation

- ℹ️ **GC Content Adjusted:** The GC content moved from 76.4% to 63.0%, getting closer to the target of 63.0%.
- ℹ️ **Improvement:** The 'Complex DNA Repeats' score was reduced from 1395 to 21.
- ℹ️ **Improvement:** The 'Immunogenicity Score' score was reduced from 0.09 to 0.08.

---

## Optimization Options

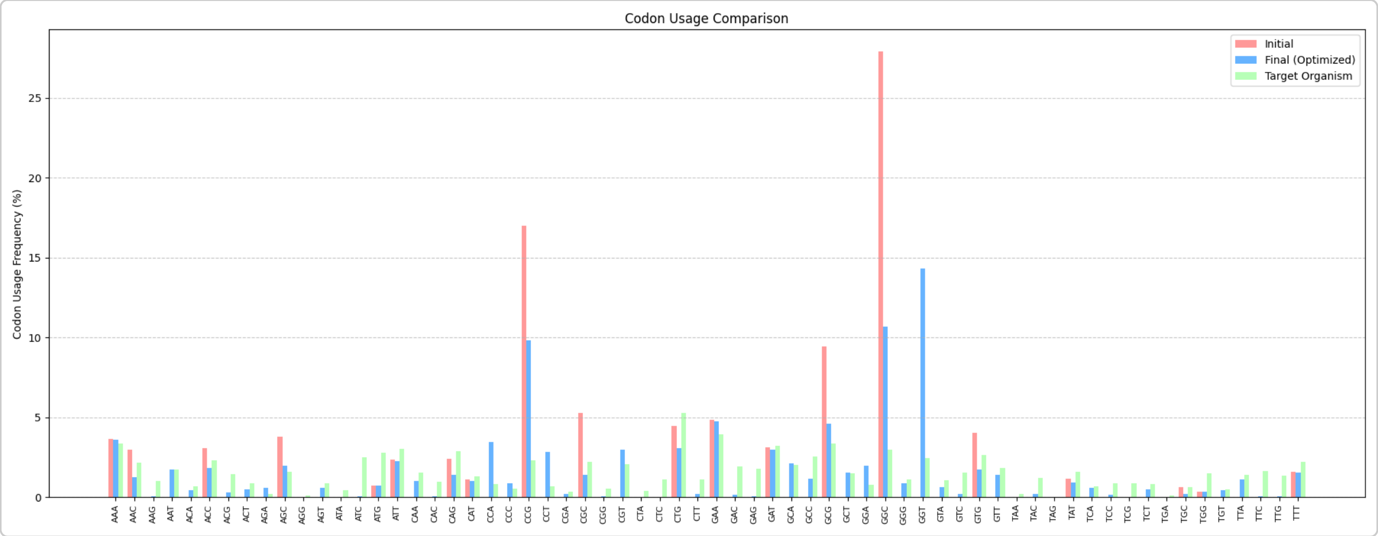
| Option | Value |
|---|---|
| **Organism** | Escherichia coli str. K-12 substr. MG1655 |
| **Optimizer** | GA |
| **Goal** | Maximize |
| **Targets** | cai, gc, cpb, tfbs |
| **Target GC %** | 63.0 (User Input) |
| **Avoided Enzymes** | NdeI(CATATG), XhoI(CTCGAG) |

## Detailed Analysis

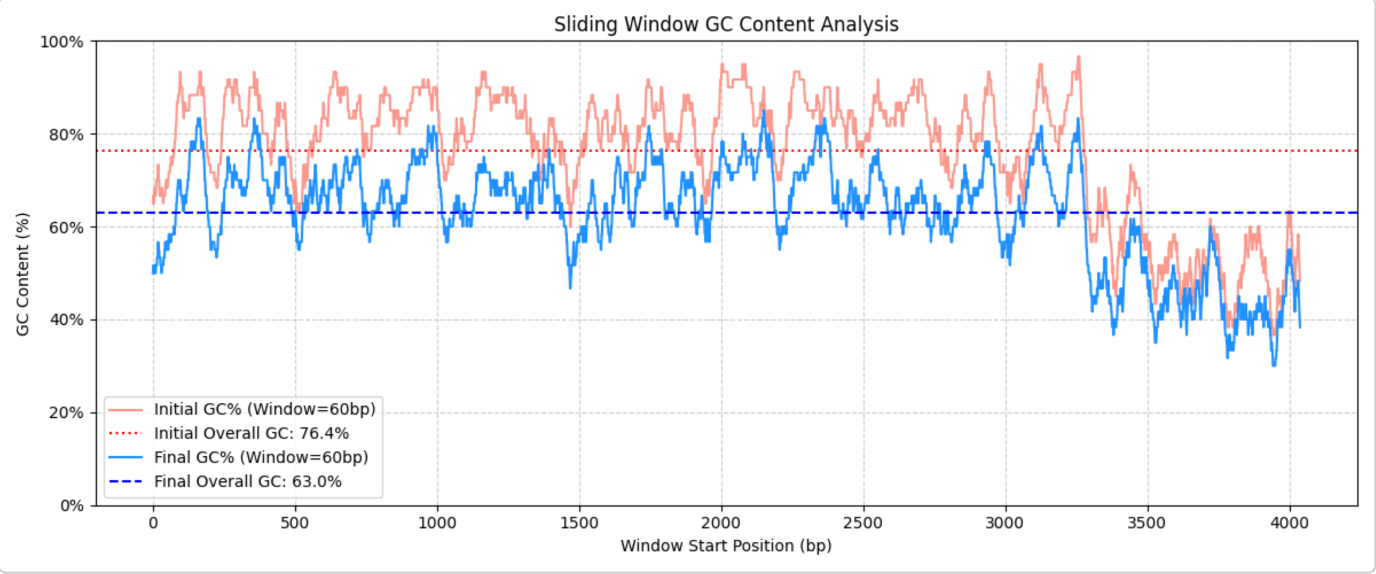
| Metric | Initial | Final (Optimized) |
|---|---|---|
| Sequence Length | 4098 | 4098 |
| Gc Content Percent | 76.4 | 63.0 |
| Cai Score | N/A | 0.8513 |
| Cpb Score | 0.0000 | 0.0000 |
| Tfbs Count | 1 | 0 |
| Restricted Enzyme Sites | 0 | 0 |
| Five Prime Mfe Kcal Mol | -15.30 | -6.65 |
| Cpg Dinucleotide Count | 677 | 375 |
| Upa Dinucleotide Count | 40 | 117 |
| Immunogenicity Score | 0.09 | 0.08 |
| Immunogenicity Risk | Low | Low |
| Complex Dna Repeats Count | 1395 | 21 |

## Visual Analysis

### Codon Usage Analysis

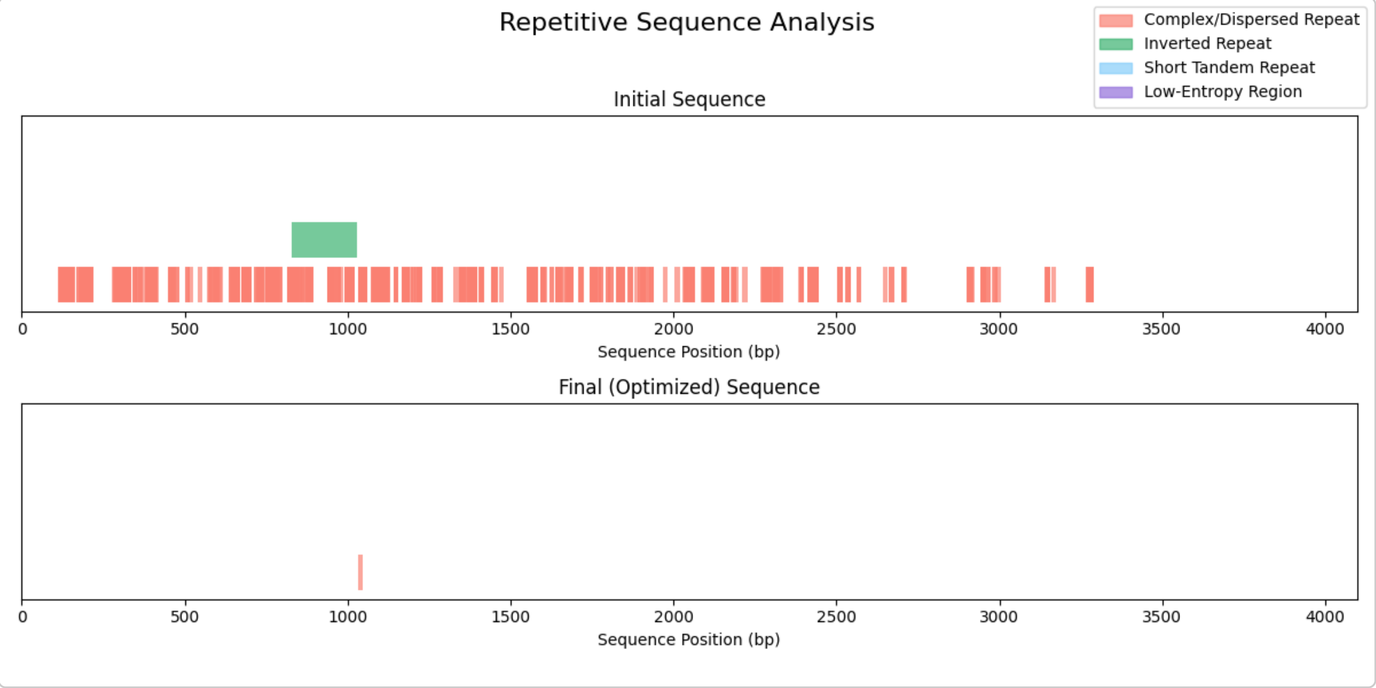


## GC Content Distribution



## Repetitive Sequence Analysis

This plot shows locations of complex repeats. Effective optimization should eliminate these regions.



## RNA Secondary Structure (2D Visualization)

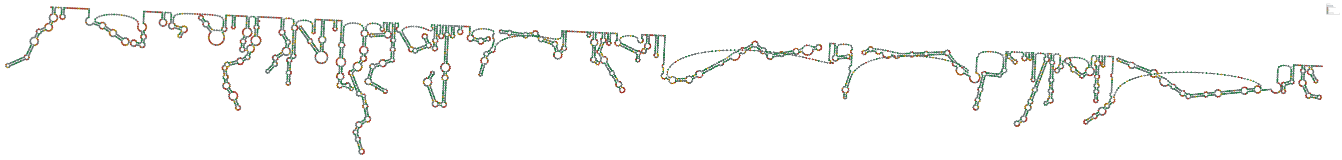
DMS-reactivity-style coloring: Green=base-paired A/C, Orange-Red=unpaired A/C, Grey=U/G (not DMS-reactive). Top: **Before Optimization** | Bottom: **After Optimization**.

### Before Optimization (COL1A2)

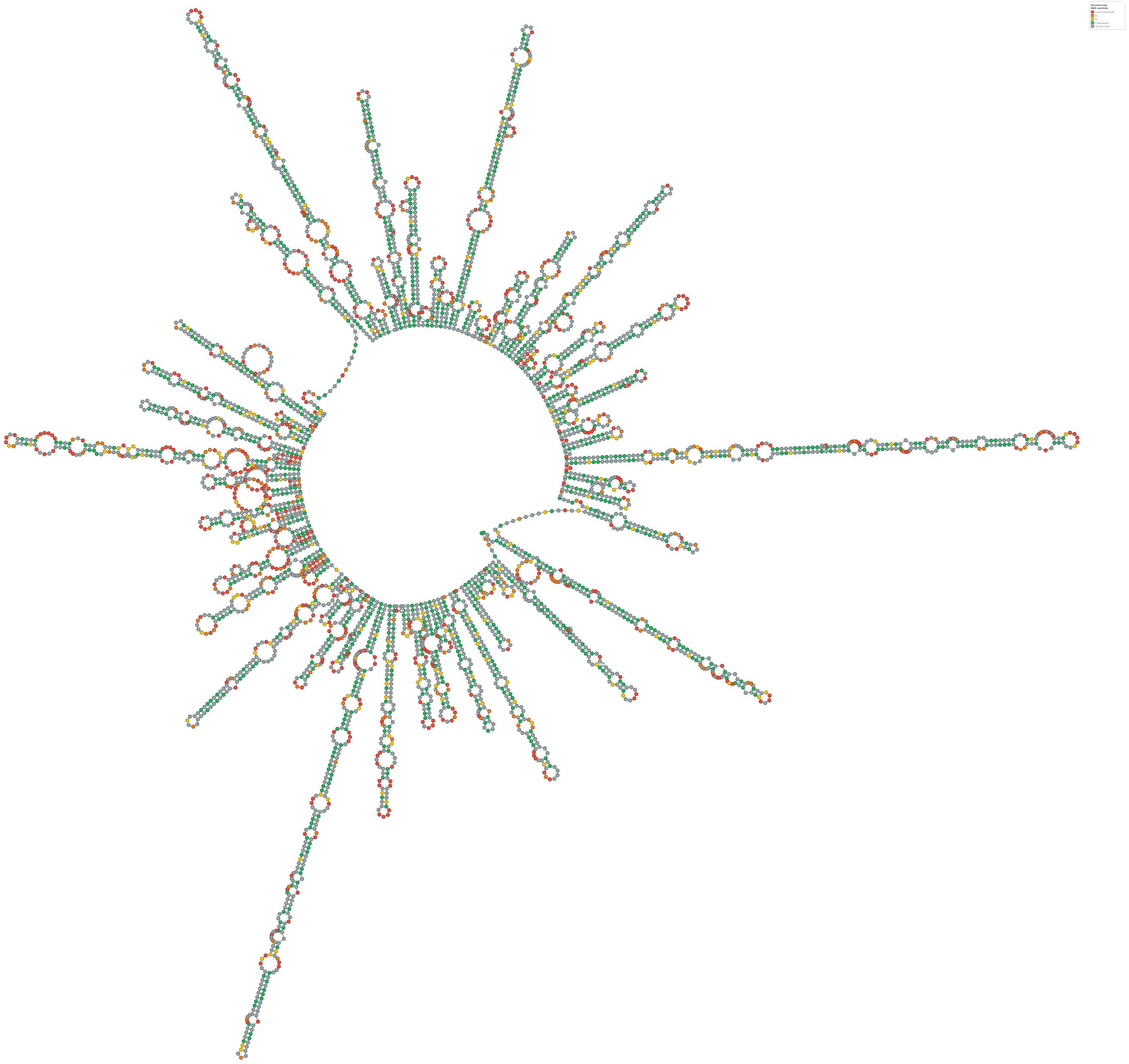
**Seeded LinearFold** **Layout: RNAplot Puzzler + Refine**

**MFE:** -2223.60 kcal/mol | **Shape:** Multi-domain (Dumbbell) | **Paired:** 2,576 / 4,098 nt (62.9%) | **Max depth:** 87 | (19.23s)

**RNAplot Puzzler + Refine**



**RNAplot Naview + Refine**

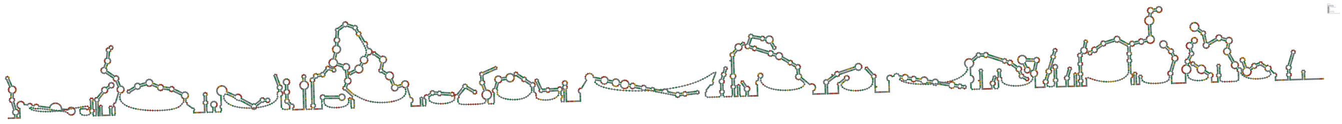


### After Optimization (COL1A2)

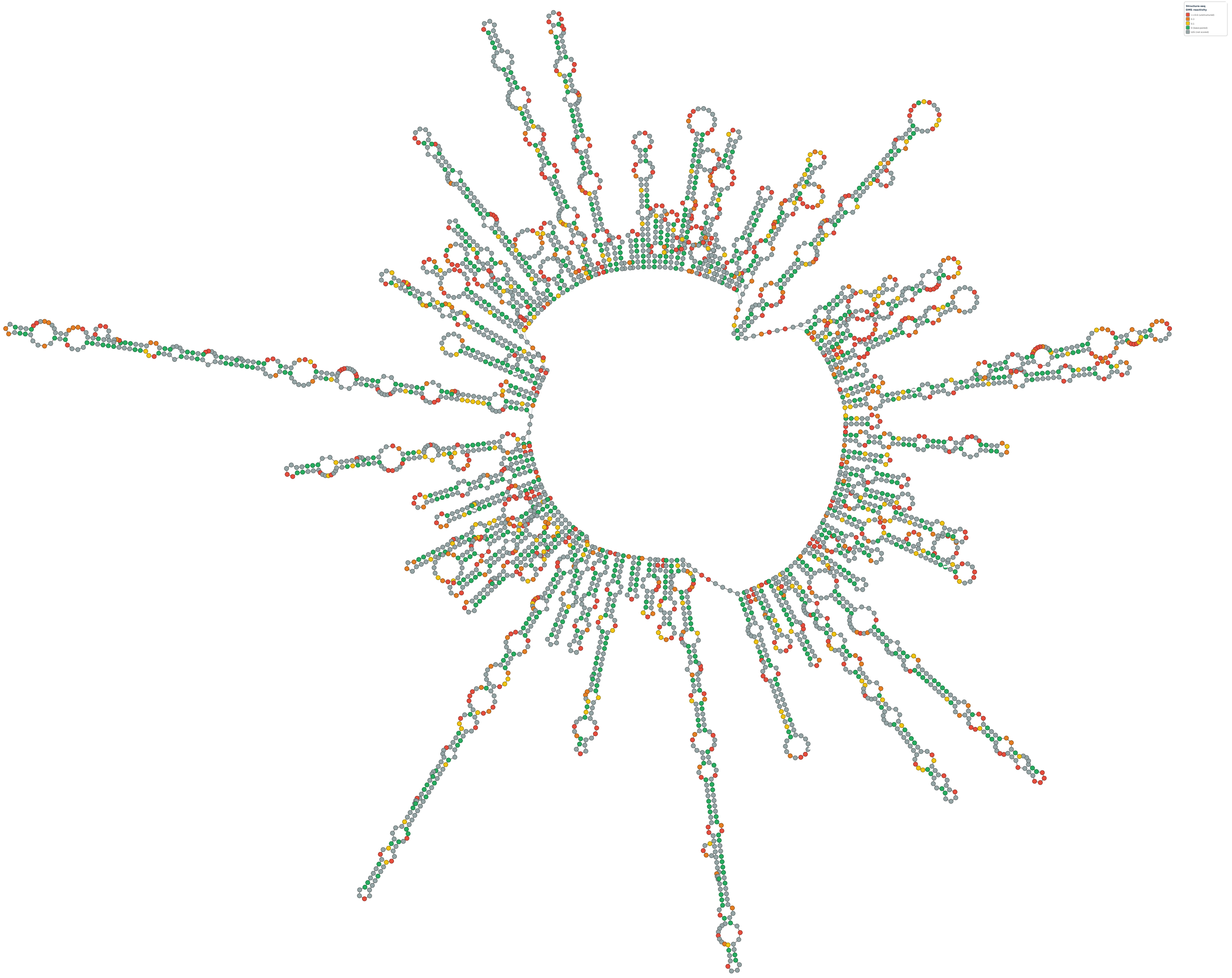
**Seeded LinearFold** **Layout: RNAplot Puzzler + Refine**

**MFE:** -1672.75 kcal/mol | **Shape:** Multi-domain (Dumbbell) | **Paired:** 2,450 / 4,098 nt (59.8%) | **Max depth:** 72 | (19.05s)

## RNAplot Puzzler + Refine



## RNAplot Naview + Refine



## Input Sequence (FASTA)

```
>Initial_Sequence
ATGCTGAGCTTTGTGGATACCCGCACCCTGCTGCTGCTGGCGGTGACCCTGTGCCTGGCGACCTGCCAGAGCCTGCAGGAAGAAACCGTGCGCAA
AGGCCCGGCGGGCGATCGCGGCCCGCGCGGCGAACGCGGCCCGCCGGGCCCGCCGGGCCGCGATGGCGAAGATGGCCCGACCGGCCCGCCGGGCC
CGCCGGGCCCGCCGGGCCCGCCGGGCCTGGGCGGCAACTTTGCGGCGCAGTATGATGGCAAAGGCGTGGGCCTGGGCCCGGGCCCGATGGGCCTG
ATGGGCCCGCGCGGCCCGCCGGGCGCGGCGGGCGCGCCGGGCCCGCAGGGCTTTCAGGGCCCGGCGGGCGAACCGGGCGAACCGGGCCAGACCGG
CCCGGCGGGCGCGCGCGGCCCGGCGGGCCCGCCGGGCAAAGCGGGCGAAGATGGCCATCCGGGCAAACCGGGCCGCCCGGGCGAACGCGGCGTGG
TGGGCCCGCAGGGCGCGCGCGGCTTTCCGGGCACCCCGGGCCTGCCGGGCTTTAAAGGCATTCGCGGCCATAACGGCCTGGATGGCCTGAAAGGC
CAGCCGGGCGCGCCGGGCGTGAAAGGCGAACCGGGCGCGCCGGGCGAAAACGGCACCCCGGGCCAGACCGGCGCGCGCGGCCTGCCGGGCGAACG
CGGCCGCGTGGGCGCGCCGGGCCCGGCGGGCGCGCGCGGCAGCGATGGCAGCGTGGGCCCGGTGGGCCCGGCGGGCCCGATTGGCAGCGCGGGCC
CGCCGGGCTTTCCGGGCGCGCCGGGCCCGAAAGGCGAAATTGGCGCGGTGGGCAACGCGGGCCCGGCGGGCCCGGCGGGCCCGCGCGGCGAAGTG
GGCCTGCCGGGCCTGAGCGGCCCGGTGGGCCCGCCGGGCAACCCGGGCGCGAACGGCCTGACCGGCGCGAAAGGCGCGGCGGGCCTGCCGGGCGT
GGCGGGCGCGCCGGGCCTGCCGGGCCCGCGCGGCATTCCGGGCCCGGTGGGCGCGGCGGGCGCGACCGGCGCGCGCGGCCTGGTGGGCGAACCGG
GCCCGGCGGGCAGCAAAGGCGAAAGCGGCAACAAAGGCGAACCGGGCAGCGCGGGCCCGCAGGGCCCGCCGGGCCCGAGCGGCGAAGAAGGCAAA
CGCGGCCCGAACGGCGAAGCGGGCAGCGCGGGCCCGCCGGGCCCGCCGGGCCTGCGCGGCAGCCCGGGCAGCCGCGGCCTGCCGGGCGCGGATGG
CCGCGCGGGCGTGATGGGCCCGCCGGGCAGCCGCGGCGCGAGCGGCCCGGCGGGCGTGCGCGGCCCGAACGGCGATGCGGGCCGCCCGGGCGAAC
CGGGCCTGATGGGCCCGCGCGGCCTGCCGGGCAGCCCGGGCAACATTGGCCCGGCGGGCAAAGAAGGCCCGGTGGGCCTGCCGGGCATTGATGGC
CGCCCGGGCCCGATTGGCCCGGCGGGCGCGCGCGGCGAACCGGGCAACATTGGCTTTCCGGGCCCGAAAGGCCCGACCGGCGATCCGGGCAAAAA
CGGCGATAAAGGCCATGCGGGCCTGGCGGGCGCGCGCGGCGCGCCGGGCCCGGATGGCAACAACGGCGCGCAGGGCCCGCCGGGCCCGCAGGGCG
TGCAGGGCGGCAAAGGCGAACAGGGCCCGCCGGGCCCGCCGGGCTTTCAGGGCCTGCCGGGCCCGAGCGGCCCGGCGGGCGAAGTGGGCAAACCG
GGCGAACGCGGCCTGCATGGCGAATTTGGCCTGCCGGGCCCGGCGGGCCCGCGCGGCGAACGCGGCCCGCCGGGCGAAAGCGGCGCGGCGGGCCC
GACCGGCCCGATTGGCAGCCGCGGCCCGAGCGGCCCGCCGGGCCCGGATGGCAACAAAGGCGAACCGGGCGTGGTGGGCGCGGTGGGCACCGCGG
GCCCGAGCGGCCCGAGCGGCCTGCCGGGCGAACGCGGCGCGGCGGGCATTCCGGGCGGCAAAGGCGAAAAAGGCGAACCGGGCCTGCGCGGCGAA
ATTGGCAACCCGGGCCGCGATGGCGCGCGCGGCGCGCCGGGCGCGGTGGGCGCGCCGGGCCCGGCGGGCGCGACCGGCGATCGCGGCGAAGCGGG
CGCGGCGGGCCCGGCGGGCCCGGCGGGCCCGCGCGGCAGCCCGGGCGAACGCGGCGAAGTGGGCCCGGCGGGCCCGAACGGCTTTGCGGGCCCGG
CGGGCGCGGCGGGCCAGCCGGGCGCGAAAGGCGAACGCGGCGCGAAAGGCCCGAAAGGCGAAAACGGCGTGGTGGGCCCGACCGGCCCGGTGGGC
GCGGCGGGCCCGGCGGGCCCGAACGGCCCGCCGGGCCCGGCGGGCAGCCGCGGCGATGGCGGCCCGCCGGGCATGACCGGCTTTCCGGGCGCGGC
GGGCCGCACCGGCCCGCCGGGCCCGAGCGGCATTAGCGGCCCGCCGGGCCCGCCGGGCCCGGCGGGCAAAGAAGGCCTGCGCGGCCCGCGCGGCG
ATCAGGGCCCGGTGGGCCGCACCGGCGAAGTGGGCGCGGTGGGCCCGCCGGGCTTTGCGGGCGAAAAAGGCCCGAGCGGCGAAGCGGGCACCGCG
GGCCCGCCGGGCACCCCGGGCCCGCAGGGCCTGCTGGGCGCGCCGGGCATTCTGGGCCTGCCGGGCAGCCGCGGCGAACGCGGCCTGCCGGGCGT
GGCGGGCGCGGTGGGCGAACCGGGCCCGCTGGGCATTGCGGGCCCGCCGGGCGCGCGCGGCCCGCCGGGCGCGGTGGGCAGCCCGGGCGTGAACG
GCGCGCCGGGCGAAGCGGGCCGCGATGGCAACCCGGGCAACGATGGCCCGCCGGGCCGCGATGGCCAGCCGGGCCATAAAGGCGAACGCGGCTAT
CCGGGCAACATTGGCCCGGTGGGCGCGGCGGGCGCGCCGGGCCCGCATGGCCCGGTGGGCCCGGCGGGCAAACATGGCAACCGCGGCGAAACCGG
CCCGAGCGGCCCGGTGGGCCCGGCGGGCGCGGTGGGCCCGCGCGGCCCGAGCGGCCCGCAGGGCATTCGCGGCGATAAAGGCGAACCGGGCGAAA
AAGGCCCGCGCGGCCTGCCGGGCCTGAAAGGCCATAACGGCCTGCAGGGCCTGCCGGGCATTGCGGGCCATCATGGCGATCAGGGCGCGCCGGGC
AGCGTGGGCCCGGCGGGCCCGCGCGGCCCGGCGGGCCCGAGCGGCCCGGCGGGCAAAGATGGCCGCACCGGCCATCCGGGCACCGTGGGCCCGGC
GGGCATTCGCGGCCCGCAGGGCCATCAGGGCCCGGCGGGCCCGCCGGGCCCGCCGGGCCCGCCGGGCCCGCCGGGCGTGAGCGGCGGCGGCTATG
ATTTTGGCTATGATGGCGATTTTTATCGCGCGGATCAGCCGCGCAGCGCGCCGAGCCTGCGCCCGAAAGATTATGAAGTGGATGCGACCCTGAAA
AGCCTGAACAACCAGATTGAAACCCTGCTGACCCCGGAAGGCAGCCGCAAAAACCCGGCGCGCACCTGCCGCGATCTGCGCCTGAGCCATCCGGA
ATGGAGCAGCGGCTATTATTGGATTGATCCGAACCAGGGCTGCACCATGGATGCGATTAAAGTGTATTGCGATTTTAGCACCGGCGAAACCTGCA
TTCGCGCGCAGCCGGAAAACATTCCGGCGAAAAACTGGTATCGCAGCAGCAAAGATAAAAAACATGTGTGGCTGGGCGAAACCATTAACGCGGGC
AGCCAGTTTGAATATAACGTGGAAGGCGTGACCAGCAAAGAAATGGCGACCCAGCTGGCGTTTATGCGCCTGCTGGCGAACTATGCGAGCCAGAA
CATTACCTATCATTGCAAAAACAGCATTGCGTATATGGATGAAGAAACCGGCAACCTGAAAAAAGCGGTGATTCTGCAGGGCAGCAACGATGTGG
AACTGGTGGCGGAAGGCAACAGCCGCTTTACCTATACCGTGCTGGTGGATGGCTGCAGCAAAAAAACCAACGAATGGGGCAAAACCATTATTGAA
TATAAAACCAACAAACCGAGCCGCCTGCCGTTTCTGGATATTGCGCCGCTGGATATTGGCGGCGCGGATCAGGAATTTTTTGTGGATATTGGCCC
GGTGTGCTTTAAA
```

## Final Optimized Sequence (FASTA) (Identity: 84.1%)

```
>Optimized_COL1A2
ATGCTGAGTTTTGTTGATACACGTACCCTGCTGTTACTGGCAGTCACCCTGTGCTTAGCGACCTGTCAATCTCTGCAGGAAGAAACTGTACGGAA
AGGCCCGGCTGGCGATCGTGGTCCACGTGGTGAAAGAGGACCGCCGGGTCCCCCAGGTAGAGATGGCGAAGACGGCCCAACCGGCCCGCCGGGCC
CACCGGGTCCGCCCGGTCCGCCGGGTCTGGGTGGAAACTTTGCTGCTCAGTATGATGGGAAAGGAGTTGGCTTAGGCCCGGGCCCAATGGGTCTG
ATGGGACCTCGTGGTCCGCCTGGAGCGGCTGGTGCGCCGGGTCCGCAGGGTTTTCAGGGTCCAGCTGGTGAACCAGGCGAACCGGGTCAGACCGG
CCCTGCCGGCGCCCGTGGCCCCGCCGGTCCGCCAGGCAAAGCGGGTGAAGATGGCCATCCTGGTAAACCGGGCCGTCCGGGCGAACGCGGTGTTG
TGGGTCCGCAGGGTGCGCGTGGTTTTCCCGGCACCCCGGGCCTGCCGGGTTTTAAAGGTATTCGCGGCCACAATGGTCTGGATGGTCTGAAAGGT
CAGCCGGGCGCGCCGGGTGTGAAAGGCGAACCGGGTGCACCGGGTGAAAATGGCACGCCGGGTCAAACTGGTGCGCGCGGTCTGCCGGGTGAACG
TGGTCGTGTAGGTGCGCCAGGTCCGGCGGGTGCACGTGGTAGTGATGGTTCCGTGGGCCCGGTGGGCCCTGCGGGCCCAATTGGCAGCGCGGGAC
CGCCAGGTTTTCCAGGCGCACCGGGTCCAAAAGGTGAAATTGGCGCTGTAGGCAATGCAGGTCCGGCAGGCCCTGCAGGCCCCCGTGGTGAAGTT
GGTCTGCCGGGCTTATCAGGTCCGGTGGGTCCGCCAGGTAACCCTGGTGCAAATGGCTTGACCGGCGCGAAAGGTGCAGCGGGTCTGCCGGGCGT
GGCCGGTGCGCCGGGATTACCGGGTCCGCGTGGTATTCCGGGTCCGGTCGGCGCGGCAGGTGCCACCGGCGCCCGCGGCCTGGTGGGTGAACCAG
GCCCTGCAGGCAGTAAAGGTGAAAGTGGTAACAAGGGCGAACCGGGTTCTGCTGGCCCGCAGGGTCCTCCGGGACCAAGTGGTGAAGAAGGTAAA
CGTGGTCCGAACGGCGAAGCCGGTAGCGCAGGTCCTCCAGGTCCCCCAGGTCTGCGTGGCAGCCCGGGTAGCCGTGGTCTGCCGGGTGCTGATGG
TCGTGCAGGCGTAATGGGCCCACCTGGTAGCCGCGGTGCAAGTGGCCCAGCGGGTGTTCGTGGCCCGAATGGCGATGCGGGTCGTCCAGGTGAAC
CTGGCTTAATGGGCCCTCGCGGACTTCCGGGCAGCCCAGGTAACATTGGGCCGGCGGGCAAAGAAGGCCCGGTGGGCCTGCCTGGCATTGATGGT
CGTCCGGGCCCAATCGGCCCGGCTGGCGCGAGAGGCGAACCGGGAAATATTGGTTTTCCGGGTCCGAAAGGTCCGACCGGTGATCCGGGTAAAAA
TGGTGATAAAGGTCATGCGGGATTAGCGGGCGCACGCGGTGCGCCTGGTCCAGATGGTAATAATGGCGCTCAGGGTCCGCCTGGGCCGCAAGGAG
TTCAAGGTGGAAAAGGTGAACAAGGTCCTCCGGGCCCTCCGGGTTTTCAGGGTCTGCCGGGCCCGTCCGGCCCAGCGGGTGAAGTTGGTAAACCG
GGTGAACGCGGGCTGCATGGTGAATTTGGCCTGCCAGGTCCGGCCGGCCCCAGAGGCGAACGTGGTCCGCCGGGCGAATCAGGCGCGGCGGGCCC
GACAGGTCCGATTGGCAGCCGTGGTCCAAGCGGCCCGCCAGGTCCTGATGGTAATAAAGGCGAACCGGGCGTTGTCGGCGCAGTAGGTACTGCGG
GTCCGAGCGGGCCGTCTGGTCTGCCTGGAGAACGTGGTGCTGCGGGCATTCCGGGAGGCAAAGGTGAAAAAGGAGAGCCGGGCCTTCGCGGCGAA
ATTGGTAACCCTGGCCGTGATGGTGCGCGTGGTGCGCCGGGCGCGGTGGGTGCTCCTGGCCCTGCTGGTGCGACTGGTGATAGAGGCGAAGCGGG
TGCAGCCGGCCCGGCAGGTCCAGCGGGTCCGCGCGGAAGCCCAGGCGAACGAGGTGAAGTGGGACCGGCGGGCCCGAATGGCTTTGCGGGCCCGG
CCGGCGCGGCCGGCCAGCCGGGTGCCAAAGGTGAACGCGGTGCGAAAGGACCTAAAGGTGAAAACGGCGTAGTGGGTCCTACCGGCCCAGTTGGA
GCGGCAGGTCCTGCAGGTCCGAATGGTCCTCCGGGCCCGGCAGGTAGCCGTGGTGATGGCGGCCCTCCGGGTATGACCGGCTTTCCGGGCGCGGC
GGGTCGCACCGGCCCCCCGGGTCCGAGCGGCATTAGCGGGCCACCGGGTCCTCCTGGCCCAGCTGGCAAAGAAGGTTTACGTGGCCCGCGTGGTG
ATCAAGGTCCGGTTGGCCGCACGGGAGAAGTGGGTGCGGTTGGCCCTCCTGGTTTTGCGGGCGAAAAAGGTCCGAGCGGTGAAGCAGGTACCGCG
GGCCCACCGGGTACCCCGGGCCCGCAGGGCTTACTGGGTGCACCGGGTATTTTAGGGCTGCCGGGTTCACGTGGCGAACGTGGCTTACCAGGCGT
TGCGGGTGCGGTGGGTGAACCAGGTCCTTTAGGTATTGCAGGTCCGCCGGGTGCACGTGGTCCCCCAGGCGCAGTAGGCAGTCCGGGTGTTAATG
GTGCTCCGGGTGAAGCGGGTCGTGATGGTAACCCGGGTAATGATGGCCCGCCTGGCCGTGATGGACAACCGGGTCATAAAGGCGAACGAGGTTAC
CCGGGTAACATTGGCCCAGTGGGGGCAGCTGGTGCACCAGGTCCGCATGGTCCGGTGGGACCGGCGGGTAAACATGGGAATCGTGGCGAAACCGG
CCCGTCAGGCCCTGTTGGCCCGGCGGGCGCGGTAGGTCCGCGCGGTCCTTCTGGTCCGCAAGGCATTCGCGGTGATAAAGGTGAACCAGGTGAAA
AAGGGCCAAGAGGCCTGCCAGGGCTGAAAGGCCATAACGGTCTGCAAGGCCTGCCGGGTATTGCGGGCCATCATGGCGATCAGGGTGCGCCGGGC
AGCGTGGGCCCGGCTGGCCCTCGTGGTCCGGCGGGTCCATCTGGACCGGCGGGCAAAGATGGTCGCACCGGCCATCCGGGTACTGTAGGACCGGC
TGGTATTCGTGGTCCGCAAGGACATCAGGGCCCAGCAGGTCCGCCGGGCCCCCCGGGTCCTCCCGGCCCACCGGGCGTTTCAGGCGGAGGTTATG
ATTTCGGTTATGATGGTGATTTTTATCGTGCGGATCAACCTAGAAGCGCGCCATCTCTGAGACCGAAAGATTACGAAGTGGATGCTACCTTAAAA
AGCTTAAACAACCAGATTGAAACCCTGCTGACACCGGAAGGTAGCCGTAAAAACCCGGCGCGTACATGTCGAGATCTGCGCCTGAGCCATCCGGA
ATGGTCAAGCGGCTATTATTGGATTGATCCGAATCAGGGTTGTACAATGGATGCCATTAAAGTGTATTGTGATTTTAGCACCGGTGAAACCTGCA
TTCGTGCGCAACCGGAAAATATTCCCGCCAAAAACTGGTATCGTAGCTCAAAAGACAAAAAACATGTGTGGCTGGGCGAAACAATTAATGCGGGG
AGCCAATTTGAATATAACGTTGAAGGCGTGACGAGCAAAGAAATGGCCACCCAGCTGGCGTTTATGCGCCTGCTGGCGAATTATGCGTCACAGAA
CATTACTTATCATTGCAAAAATAGCATTGCCTATATGGATGAAGAAACCGGTAACCTGAAAAAAGCGGTGATTCTTCAGGGCAGTAATGATGTTG
AATTAGTTGCTGAAGGTAATAGCCGTTTTACGTATACCGTGCTGGTGGATGGTTGTAGCAAAAAAACTAATGAATGGGGCAAAACCATTATTGAA
TACAAAACCAATAAACCATCTCGCCTGCCGTTTCTGGATATTGCGCCGCTGGATATTGGCGGTGCTGATCAAGAATTTTTTGTGGATATTGGTCC
GGTTTGTTTTAAA
```