Introduction to Bioengineering
BIOE/ENGR.80
Stanford University

Spring 2020 Class Slides

Day 8
22 April 2020

Analysis & design of biomolecules

mRNA loaded with ribosomes translating ORFs to make protein

RNA polymerase initiating transcription

DNA

Transcription terminator? (somewhere around here)

Promoter

0.5 μm

Visualization of Bacterial Genes in Action

Abstract. *The morphology of active structural and putative ribosomal RNA genes was observed by electron microscopy after lysis of fragile* Escherichia coli *cells. Conclusions drawn are: most of the chromosome is not genetically active at any one instant; translation is completely coupled with transcription; the 16S and 23S ribosomal RNA cistrons occur in tandem, in regions which are widely spaced on the chromosome.*

```
GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTTTCGGATATTTCTGATGAGTCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTTGTTTACGAATTAAATCGAAGTGGAC
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGCGCAGCTCGAGAAGCTCTTACTTTGCGACCTTTCGCCATCAACTAACGATTCTGTCAAAAACTGACGCGTTGGATGAGGAGAAG
TGGCTTAATATGCTTGGCACGTTCGTCAAGGACTGGTTTAGATATGAGTCACATTTTGTTCATGGTAGAGATTCTCTTGTTGACATTTTAAAAGAGCGTGGATTACTATCTGAGTCCGAT
GCTGTTCAACCACTAATAGGTAAGAAATCATGAGTCAAGTTACTGAACAATCCGTACGTTTCCAGACCGCTTTGGCCTCTATTAAGCTCATTCAGGCTTCTGCCGTTTTGGATTTAACCG
AAGATGATTTCGATTTTCTGACGAGTAACAAAGTTTGGATTGCTACTGACCGCTCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCTGGACTTTGTGGGATACCCTCGCT
TTCCTGCTCCTGTTGAGTTTATTGCTGCCGTCATTGCTTATTATGTTCATCCCGTCAACATTCAAACGGCCTGTCTCATCATGGAAGGCGCTGAATTTACGGAAAACATTATTAATGGCG
TCGAGCGTCCGGTTAAAGCCGCTGAATTGTTCGCGTTTACCTTGCGTGTACGCGCAGGAAACACTGACGTTCTTACTGACGCAGAAGAAAACGTGCGTCAAAAATTACGTGCGGAAGGAG
TGATGTAATGTCTAAAGGTAAAAAACGTTCTGGCGCTCGCCCTGGTCGTCCGCAGCCGTTGCGAGGTACTAAAGGCAAGCGTAAAGGCGCTCGTCTTTGGTATGTAGGTGGTCAACAATT
TTAATTGCAGGGGCTTCGGCCCCTTACTTGAGGATAAATTATGTCTAATATTCAAACTGGCGCCGAGCGTATGCCGCATGACCTTTCCCATCTTGGCTTCCTTGCTGGTCAGATTGGTCG
TCTTATTACCATTTCAACTACTCCGGTTATCGCTGGCGACTCCTTCGAGATGGACGCCGTTGGCGCTCTCCGTCTTTCTCCATTGCGTCGTGGCCTTGCTATTGACTCTACTGTAGACAT
TTTTACTTTTTATGTCCCTCATCGTCACGTTTATGGTGAACAGTGGATTAAGTTCATGAAGGATGGTGTTAATGCCACTCCTCTCCCGACTGTTAACACTACTGGTTATATTGACCATGC
CGCTTTTCTTGGCACGATTAACCCTGATACCAATAAAATCCCTAAGCATTTGTTTCAGGGTTATTTGAATATCTATAACAACTATTTTAAAGCGCCGTGGATGCCTGACCGTACCGAGGC
TAACCCTAATGAGCTTAATCAAGATGATGCTCGTTATGGTTTCCGTTGCTGCCATCTCAAAAACATTTGGACTGCTCCGCTTCCTCCTGAGACTGAGCTTTCTCGCCAAATGACGACTTC
TACCACATCTATTGACATTATGGGTCTGCAAGCTGCTTATGCTAATTTGCATACTGACCAAGAACGTGATTACTTCATGCAGCGTTACCATGATGTTATTTCTTCATTTGGAGGTAAAAC
CTCTTATGACGCTGACAACCGTCCTTTACTTGTCATGCGCTCTAATCTCTGGGCATCTGGCTATGATGTTGATGGAACTGACCAAACGTCGTTAGGCCAGTTTTCTGGTCGTGTTCAACA
GACCTATAAACATTCTGTGCCGCGTTTCT████████████████████████████████████AAGAGATTCAGTACCTTAACGCTAAGG
TGCTTTGACTTATACCGATATTGCTGGCG██                                ██GTGATTCGTCTAAGAAGTTTAAGATTGC
TGAGGGTCAGTGGTATCGTTATGCGCCTT██                                ██GTGATTTGCAAGAACGCGTACTTATTCG
CCACCATGATTATGACCAGTGTTTCCAGT██       "Genome: bought the      ██CGACCACTCGCGATTCAATCATGACTTC
GTGATAAAAGATTGAGTGTGAGGTTATAA██                                ██TTTTCTGCTTAGGAGTTTAATCATGTTT
CAGACTTTTATTTCTCGCCATAATTCAAA██        book; hard to read."     ██GACACCTAAAGCTACATCGTCAACGTTA
TATTTTGATAGTTTGACGGTTAATGCTGG██                                ██TTCTGTTGGTGCTGATATTGCTTTTGAT
GCCGACCCTAAATTTTTTGCCTGTTTGGT██              — Eric Lander      ██TGGTCGCCATGATGGTGGTTATTATACC
GTCAAGGACTGTGTGACTATTGACGTCCT██                                ██TAAATGCCGCGGATTGGTTTCGCTGAAT
CAGGTTATTAAAGAGATTATTTGTCTCCA██                                ██GGTGGCGCCATGCTAAATTGTTTGGAG
GCGGTCAAAAAGCCGCCTCCGGTGGCATT██                                ██GCCATTCAAGGCTCTAATGTTCCTAACC
CTGATGAGGCCGCCCCTAGTTTTGTTTCT██                                ██GTTTCTGATAAGTTGCTTGATTTGGTTG
GACTTGGTGGCAAGTCTGCCGCTGATAAA██                                ██GGTGCTGATGCTTCCTCTGCTGGTATGG
TTGACGCCGGATTTGAGAATCAAAAAGAG██                                ██GAGATTGCTGGCATTCAGTCGGCGACTT
CACGCCAGAATACGAAAGACCAGGTATAT████████████████████████████████████GAAAACACCAATCTTTCCAAGCAACAGC
AGGTTTCCGAGATTATGCGCCAAATGCTTACTCAAGCTCAAACGGCTGGTCAGTATTTTACCAATGACCAAATCAAAGAAATGACTCGCAAGGTTAGTGCTGAGGTTGACTTAGTTCATC
AGCAAACGCAGAATCAGCGGTATGGCTCTTCTCATATTGGCGCTACTGCAAAGGATATTTCTAATGTCGTCACTGATGCTGCTTCTGGTGTGGTTGATATTTTTCATGGTATTGATAAAG
CTGTTGCCGATACTTGGAACAATTTCTGGAAAGACGGTAAAGCTGATGGTATTGGCTCTAATTTGTCTAGGAAATAACCGTCAGGATTGACACCCTCCCAATTGTATGTTTTCATGCCTC
CAAATCTTGGAGGCTTTTTTATGGTTCGTTCTTATTACCCTTCTGAATGTCACGCTGATTATTTTGACTTTGAGCGTATCGAGGCTCTTAAACCTGCTATTGAGGCTTGTGGCATTTCTA
CTCTTTCTCAATCCCCAATGCTTGGCTTCCATAAGCAGATGGATAACCGCATCAAGCTCTTGGAAGAGATTCTGTCTTTTCGTATGCAGGGCGTTGAGTTCGATAATGGTGATATGTATG
TTGACGGCCATAAGGCTGCTTCTGACGTTCGTGATGAGTTTGTATCTGTTACTGAGAAGTTAATGGATGAATTGGCACAATGCTACAATGTGCTCCCCCAACTTGATATTAATAACACTA
TAGACCACCGCCCCGAAGGGGACGAAAAATGGTTTTTAGAGAACGAGAAGACGGTTACGCAGTTTTGCCGCAAGCTGGCTGCTGAACGCCCTCTTAAGGATATTCGCGATGAGTATAATT
ACCCCAAAAAGAAAGGTATTAAGGATGAGTGTTCAAGATTGCTGGAGGCCTCCACTATGAAATCGCGTAGAGGCTTTGCTATTCAGCGTTTGATGAATGCAATGCGACAGGCTCATGCTG
ATGGTTGGTTTATCGTTTTTGACACTCTCACGTTGGCTGACGACCGATTAGAGGCGTTTTATGATAATCCCAATGCTTTGCGTGACTATTTTCGTGATATTGGTCGTATGGTTCTTGCTG
CCGAGGGTCGCAAGGCTAATGATTCACACGCCGACTGCTATCAGTATTTTTGTGTGCCTGAGTATGGTACAGCTAATGGCCGTCTTCATTTCCATGCGGTGCACTTTATGCGGACACTTC
CTACAGGTAGCGTTGACCCTAATTTTGGTCGTCGGGTACGCAATCGCCGCCAGTTAAATAGCTTGCAAAATACGTGGCCTTATGGTTACAGTATGCCCATCGCAGTTCGCTACACGCAGG
ACGCTTTTTCACGTTCTGGTTGGTTGTGGCCTGTTGATGCTAAAGGTGAGCCGCTTAAAGCTACCAGTTATATGGCTGTTGGTTTCTATGTGGCTAAATACGTTAACAAAAAGTCAGATA
TGGACCTTGCTGCTAAAGGTCTAGGAGCTAAAGAATGGAACAACTCACTAAAAACCAAGCTGTCGCTACTTCCCAAGAAGCTGTTCAGAATCAGAATGAGCCGCAACTTCGGGATGAAAA
TGCTCACAATGACAAATCTGTCCACGGAGTGCTTAATCCAACTTACCAAGCTGGGTTACGACGCGACGCCGTTCAACCAGATATTGAAGCAGAACGCAAAAAGAGAGATGAGATTGAGGC
TGGGAAAAGTTACTGTAGCCGACGTTTTGGCGGCGCAACCTGTGACGACAAATCTGCTCAAATTTATGCGCGCTTCGATAAAAATGATTGGCGTATCCAACCTGCA (wraparound)
```

https://www.ncbi.nlm.nih.gov/nuccore/9626372

Analysis of biomolecules often starts as pattern recognition, where patterns are set by what we already know for how biology works…
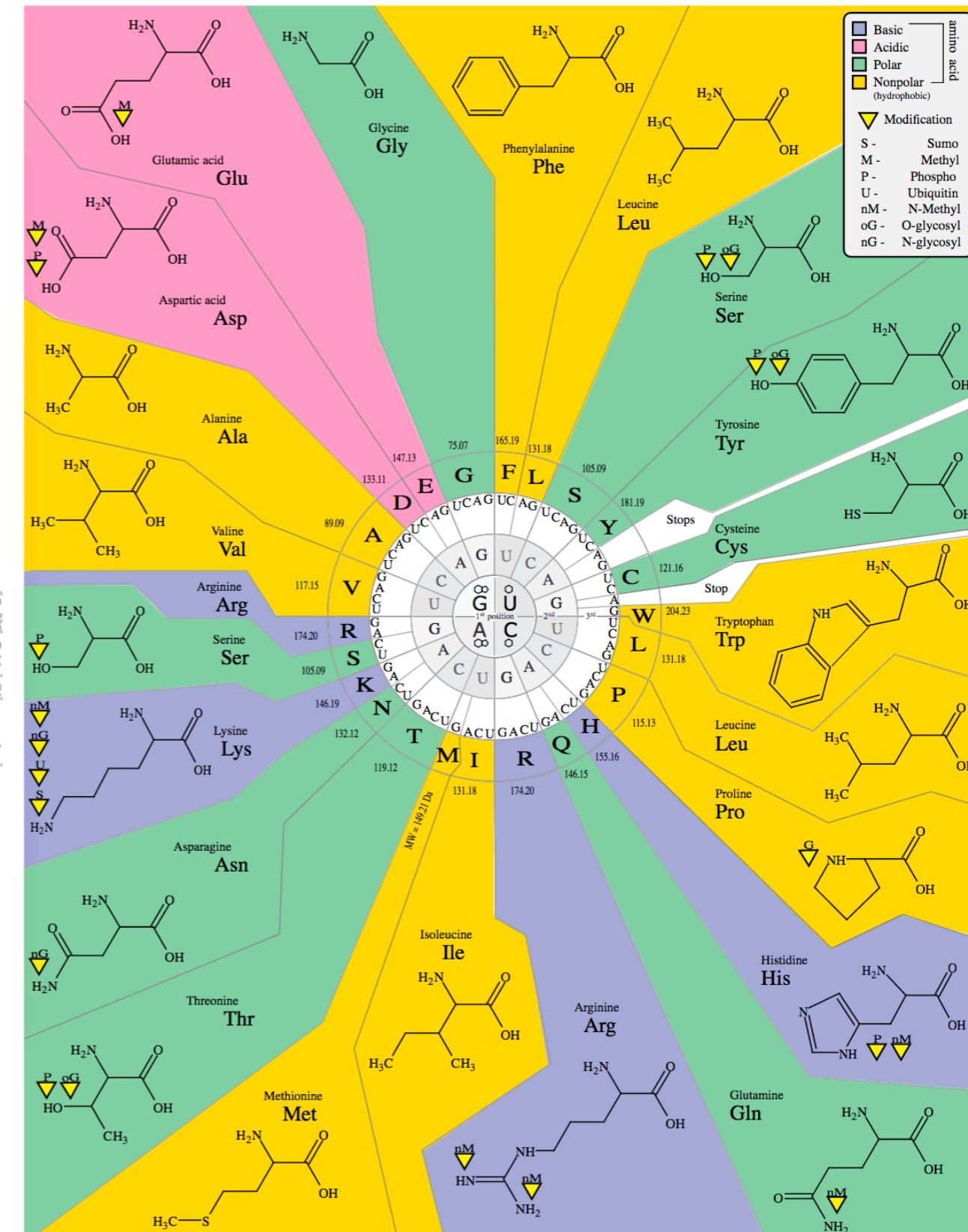
Open Reading Frame (ORF)
Decoding scheme defined by genetic code

NOTE: mRNA is translated, not DNA. U (mRNA) = T (DNA). Hence coding tables use U.


NOTE: AUG is the (best) start codon.

https://www.khanacademy.org/science/biology/gene-expression-central-dogma/central-dogma-transcription/a/the-genetic-code-discovery-and-properties
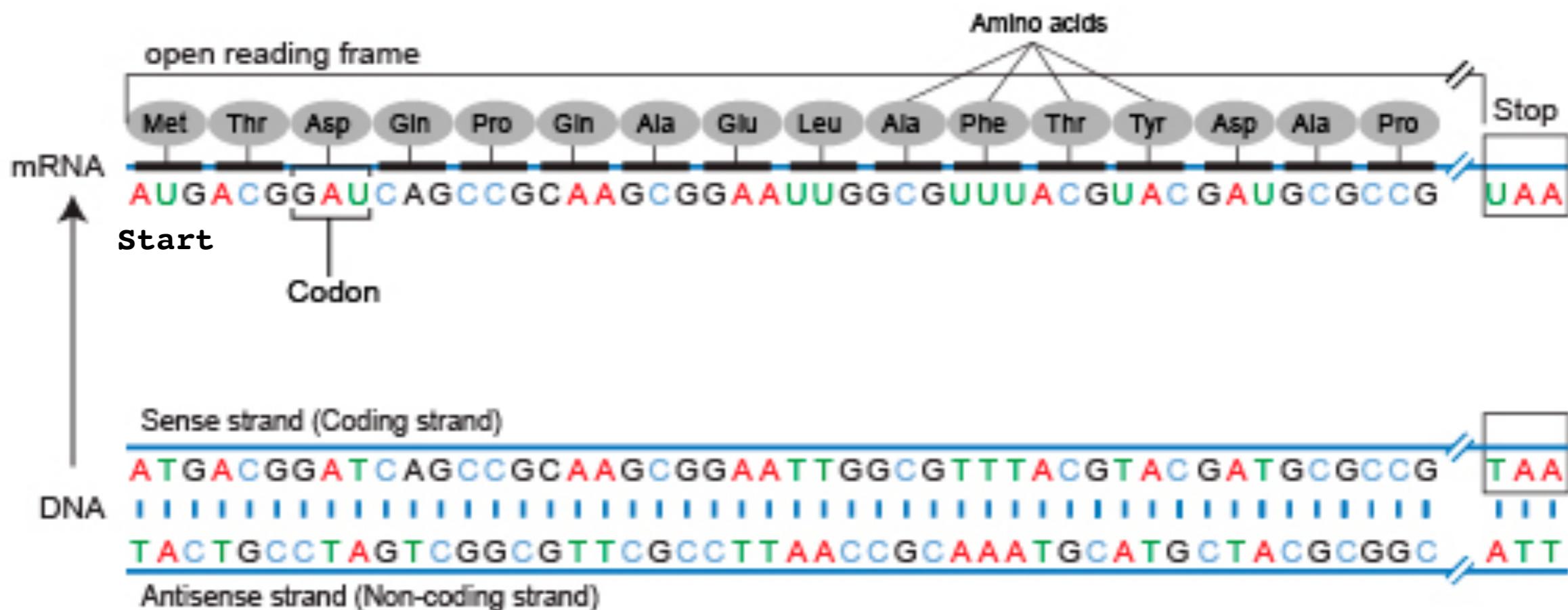
https://en.wikipedia.org/wiki/Genetic_code

Open Reading Frame (ORF)
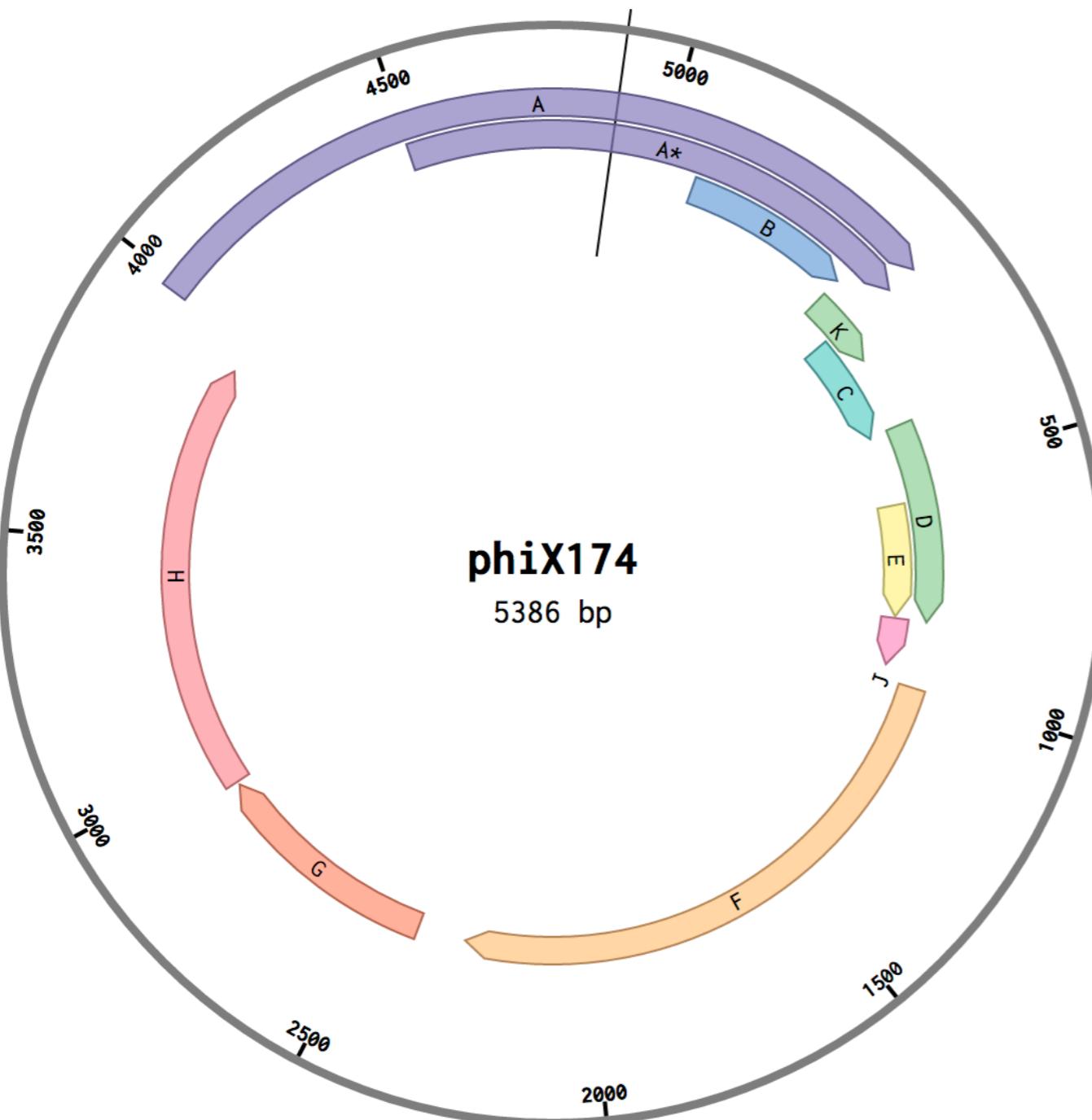
DNA-encoded sequence that, as mRNA, is decoded by the ribosome to make the so-specified protein

GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTTTCGGATATTTCTGATGAGTCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTTGTTTACGAATTAAATCGAAGTGGAC
TGCTGGCGGAAAA**TGA**GAAAATTCGACCTATCCTTGCGCAGCTCGAGAAGCTCTTACTTTGCGACCTTTCGCCATCAACTAACGATTCTGTCAAAAACTGACGCGTTGGATGAGGAGAAG
TGGCTTAATATGCTTGGCACGTTCGTCAAGGACTGGTTTAGATATGAGTCACATTTTGTTCATGGTAGAGATTCTCTTGTTGACATTTTAAAAGAGCGTGGATTACTATCTGAGTCCGAT
GCTGTTCAACCACTAATAGGTAAGAAATCATGAGTCAAGTTACTGAACAATCCGTACGTTTCCAGACCGCTTTGGCCTCTATTAAGCTCATTCAGGCTTCTGCCGTTTTGGATTTAACCG
AAGATGATTTCGATTTTCTGACGAGTAACAAAGTTTGGATTGCTACTGACCGCTCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCTGGACTTTGTGGGATACCCTCGCT
TTCCTGCTCCTGTTGAGTTTATTGCTGCCGTCATTGCTTATTATGTTCATCCCGTCAACATTCAAACGGCCTGTCTCATCATGGAAGGCGCTGAATTTACGGAAAACATTATTAATGGCG
TCGAGCGTCCGGTTAAAGCCGCTGAATTGTTCGCGTTTACCTTGCGTGTACGCGCAGGAAACACTGACGTTCTTACTGACGCAGAAGAAAACGTGCGTCAAAAATTACGTGCGGAAGGAG
TGATGTAATGTCTAAAGGTAAAAAACGTTCTGGCGCTCGCCCTGGTCGTCCGCAGCCGTTGCGAGGTACTAAAGGCAAGCGTAAAGGCGCTCGTCTTTGGTATGTAGGTGGTCAACAATT
TTAATTGCAGGGGCTTCGGCCCCTTACTTGAGGATAAATTATGTCTAATATTCAAACTGGCGCCGAGCGTATGCCGCATGACCTTTCCCATCTTGGCTTCCTTGCTGGTCAGATTGGTCG
TCTTATTACCATTTCAACTACTCCGGTTATCGCTGGCGACTCCTTCGAGATGGACGCCGTTGGCGCTCTCCGTCTTTCTCCATTGCGTCGTGGCCTTGCTATTGACTCTACTGTAGACAT
TTTTACTTTTTATGTCCCTCATCGTCACGTTTATGGTGAACAGTGGATTAAGTTCATGAAGGATGGTGTTAATGCCACTCCTCTCCCGACTGTTAACACTACTGGTTATATTGACCATGC
CGCTTTTCTTGGCACGATTAACCCTGATACCAATAAAATCCCTAAGCATTTGTTTCAGGGTTATTTGAATATCTATAACAACTATTTTAAAGCGCCGTGGATGCCTGACCGTACCGAGGC
TAACCCTAATGAGCTTAATCAAGATGATGCTCGTTATGGTTTCCGTTGCTGCCATCTCAAAAACATTTGGACTGCTCCGCTTCCTCCTGAGACTGAGCTTTCTCGCCAAATGACGACTTC
TACCACATCTATTGACATTATGGGTCTGCAAGCTGCTTATGCTAATTTGCATACTGACCAAGAACGTGATTACTTCATGCAGCGTTACCATGATGTTATTTCTTCATTTGGAGGTAAAAC
CTCTTATGACGCTGACAACCGTCCTTTACTTGTCATGCGCTCTAATCTCTGGGCATCTGGCTATGATGTTGATGGAACTGACCAAACGTCGTTAGGCCAGTTTTCTGGTCGTGTTCAACA
GACCTATAAACATTCTGTGCCGCGTTTCTTTGTTCCTGAGCATGGCACTATGTTTACTCTTGCGCTTGTTCGTTTTCCGCCTACTGCGACTAAAGAGATTCAGTACCTTAACGCTAAAGG
TGCTTTGACTTATACCGATATTGCTGGCGACCCTGTTTTGTATGGCAACTTGCCGCCGCGTGAAATTTCTATGAAGGATGTTTTCCGTTCTGGTGATTCGTCTAAGAAGTTTAAGATTGC
TGAGGGTCAGTGGTATCGTTATGCGCCTTCGTATGTTTCTCCTGCTTATCACCTTCTTGAAGGCTTCCCATTCATTCAGGAACCGCCTTCTGGTGATTTGCAAGAACGCGTACTTATTCG
CCACCATGATTATGACCAGTGTTTCCAGTCCGTTCAGTTGTTGCAGTGGAATAGTCAGGTTAAATTTAATGTGACCGTTTATCGCAATCTGCCGACCACTCGCGATTCAATCATGACTTC
GTGATAAAGATTGAGTGTGAGGTTATAACGCCGAAGCGGTAAAAATTTTAATTTTTGCCGCTGAGGGGTTGACCAAGCGAAGCGCGGTAGGTTTTCTGCTTAGGAGTTTAATCATGTTT
CAGACTTTTATTTCTCGCCATAATTCAAACTTTTTTTCTGATAAGCTGGTTCTCACTTCTGTTACTCCAGCTTCTTCGGCACCTGTTTTACAGACACCTAAAGCTACATCGTCAACGTTA
TATTTTGATAGTTTGACGGTTAATGCTGGTAATGGTGGTTTTCTTCATTGCATTCAGATGGATACATCTGTCAACGCCGCTAATCAGGTTGTTTCTGTTGGTGCTGATATTGCTTTTGAT
GCCGACCCTAAATTTTTTGCCTGTTTGGTTCGCTTTGAGTCTTCTTCGGTTCCGACTACCCTCCCGACTGCCTATGATGTTTATCCTTTGAATGGTCGCCATGATGGTGGTTATTATACC
GTCAAGGACTGTGTGACTATTGACGTCCTTCCCCGTACGCCGGGCAATAACGTTTATGTTGGTTTCATGGTTTGGTCTAACTTTACCGCTACTAAATGCCGCGGATTGGTTTCGCTGAAT
CAGGTTATTAAAGAGATTATTTGTCTCCAGCCACTTAAGTGAGGTGATTTATGTTTGGTGCTATTGCTGGCGGTATTGCTTCTGCTCTTGCTGGTGGCGCCATGTCTAAATTGTTTGGAG
GCGGTCAAAAAGCCGCCTCCGGTGGCATTCAAGGTGATGTGCTTGCTACCGATAACAATACTGTAGGCATGGGTGATGCTGGTATTAAATCTGCCATTCAAGGCTCTAATGTTCCTAACC
CTGATGAGGCCGCCCCTAGTTTTGTTTCTGGTGCTATGGCTAAAGCTGGTAAAGGACTTCTTGAAGGTACGTTGCAGGCTGGCACTTCTGCCGTTTCTGATAAGTTGCTTGATTTGGTTG
GACTTGGTGGCAAGTCTGCCGCTGATAAAGGAAAGGATACTCGTGATTATCTTGCTGCTGCATTTCCTGAGCTTAATGCTTGGGAGCGTGCTGGTGCTGATGCTTCCTCTGCTGGTATGG
TTGACGCCGGATTTGAGAATCAAAAAGAGCTTACTAAAATGCAACTGGACAATCAGAAAGAGATTGCCGAGATGCAAAATGAGACTCAAAAAGAGATTGCTGGCATTCAGTCGGCGACTT
CACGCCAGAATACGAAAGACCAGGTATATGCACAAAATGAGATGCTTGCTTATCAACAGAAGGAGTCTACTGCTCGCGTTGCGTCTATTATGGAAAACACCAATCTTTCCAAGCAACAGC
AGGTTTCCGAGATTATGCGCCAAATGCTTACTCAAGCTCAAACGGCTGGTCAGTATTTTACCAATGACCAAATCAAAGAAATGACTCGCAAGGTTAGTGCTGAGGTTGACTTAGTTCATC
AGCAAACGCAGAATCAGCGGTATGGCTCTTCTCATATTGGCGCTACTGCAAAGGATATTTCTAATGTCGTCACTGATGCTGCTTCTGGTGTGGTTGATATTTTTCATGGTATTGATAAAG
CTGTTGCCGATACTTGGAACAATTTCTGGAAAGACGGTAAAGCTGATGGTATTGGCTCTAATTTGTCTAGGAAATAACCGTCAGGATTGACACCCTCCCAATTGTATGTTTTCATGCCTC
CAAATCTTGGAGGCTTTTTT**ATG**GTTCGTTCTTATTACCCTTCTGAATGTCACGCTGATTATTTTGACTTTGAGCGTATCGAGGCTCTTAAACCTGCTATTGAGGCTTGTGGCATTTCTA
CTCTTTCTCAATCCCCAATGCTTGGCTTCCATAAGCAGATGGATAACCGCATCAAGCTCTTGGAAGAGATTCTGTCTTTTCGTATGCAGGGCGTTGAGTTCGATAATGGTGATATGTATG
TTGACGGCCATAAGGCTGCTTCTGACGTTCGTGATGAGTTTGTATCTGTTACTGAGAAGTTAATGGATGAATTGGCACAATGCTACAATGTGCTCCCCCAACTTGATATTAATAACACTA
TAGACCACCGCCCCGAAGGGGACGAAAAATGGTTTTTAGAGAACGAGAAGACGGTTACGCAGTTTTGCCGCAAGCTGGCTGCTGAACGCCCTCTTAAGGATATTCGCGATGAGTATAATT
ACCCCAAAAAGAAAGGTATTAAGGATGAGTGTTCAAGATTGCTGGAGGCCTCCACTATGAAATCGCGTAGAGGCTTTGCTATTCAGCGTTTGATGAATGCAATGCGACAGGCTCATGCTG
ATGTTGGTTTATCGTTTTTGACACTCTCACGTTGGCTGACGACCGATTAGAGGCGTTTTATGATAATCCCAATGCTTTGCGTGACTATTTTCGTGATATTGGTCGTATGGTTCTTGCTG
CCGAGGGTCGCAAGGCTAATGATTCACACGCCGACTGCTATCAGTATTTTTGTGTGCCTGAGTATGGTACAGCTAATGGCCGTCTTCATTTCCATGCGGTGCACTTTATGCGGACACTTC
CTACAGGTAGCGTTGACCCTAATTTTGGTCGTCGGGTACGCAATCGCCGCCAGTTAAATAGCTTGCAAAATACGTGGCCTTATGGTTACAGTATGCCCATCGCAGTTCGCTACACGCAGG
ACGCTTTTTCACGTTCTGTTGGTTGTGGCCTGTTGATGCTAAAGGTGAGCCGCTTAAAGCTACCAGTTATATGGCTGTTGGTTTCTATGTGGCTAAATACGTTAACAAAAAGTCAGATA
TGGACCTTGCTGCTAAAGGTCTAGGAGCTAAAGAATGGAACAACTCACTAAAAACCAAGCTGTCGCTACTTCCCAAGAAGCTGTTCAGAATCAGAATGAGCCGCAACTTCGGGATGAAAA
TGCTCACAATGACAAATCTGTCCACGGAGTGCTTAATCCAACTTACCAAGCTGGGTTACGACGCGACGCCGTTCAACCAGATATTGAAGCAGAACGCAAAAAGAGAGATGAGATTGAGGC
TGGGAAAAGTTACTGTAGCCGACGTTTTGGCGGCGCAACCTGTGACGACAAATCTGCTCAAATTTATGCGCGCTTCGATAAAAATGATTGGCGTATCCAACCTGCA (wraparound)

https://www.ncbi.nlm.nih.gov/nuccore/9626372

GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTTTCGGATATTTCTGATGAGTCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTTGTTTACGAATTAAATCGAAGTGGAC
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGCGCAGCTCGAGAAGCTCTTACTTTGCGACCTTTCGCCATCAACTAACGATTCTGTCAAAAACTGACGCGTTGGATGAGGAGAAG
TGGCTTAATATGCTTGGCACGTTCGTCAAGGACTGGTTTAGATATGAGTCACATTTTGTTCATGGTAGAGATTCTCTTGTTGACATTTTAAAAGAGCGTGGATTACTATCTGAGTCCGAT
GCTGTTCAACCACTAATAGGTAAGAAATCATGAGTCAAGTTACTGAACAATCCGTACGTTTCCAGACCGCTTTGGCCTCTATTAAGCTCATTCAGGCTTCTGCCGTTTTGGATTTAACCG
AAGATGATTTCGATTTTCTGACGAGTAACAAAGTTTGGATTGCTACTGACCGCTCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCTGGACTTTGTGGGATACCCTCGCT
TTCCTGCTCCTGTTGAGTTTATTGCTGCCGTCATTGCTTATTATGTTCATCCCGTCAACATTCAAACGGCCTGTCTCATCATGGAAGGCGCTGAATTTACGGAAAACATTATTAATGGCG
TCGAGCGTCCGGTTAAAGCCGCTGAATTGTTCGCGTTTACCTTGCGTGTACGCGCAGGAAACACTGACGTTCTTACTGACGCAGAAGAAAACGTGCGTCAAAAATTACGTGCGGAAGGAG
TGATGTAATGTCTAAAGGTAAAAAACGTTCTGGCGCTCGCCCTGGTCGTCCGCAGCCGTTGCGAGGTACTAAAGGCAAGCGTAAAGGCGCTCGTCTTTGGTATGTAGGTGGTCAACAATT
TTAATTGCAGGGGCTTCGGCCCCTTACTTGAGGATAAATTATGTCTAATATTCAAACTGGCGCCGAGCGTATGCCGCATGACCTTTCCCATCTTGGCTTCCTTGCTGGTCAGATTGGTCG
TCTTATTACCATTTCAACTACTCCGGTTATCGCTGGCGACTCCTTCGAGATGGACGCCGTTGGCGCTCTCCGTCTTTCTCCATTGCGTCGTGGCCTTGCTATTGACTCTACTGTAGACAT
TTTTACTTTTTATGTCCCTCATCGTCACGTTTATGGTGAACAGTGGATTAAGTTCATGAAGGATGGTGTTAATGCCACTCCTCTCCCGACTGTTAACACTACTGGTTATATTGACCATGC
CGCTTTTCTTGGCACGATTAACCCTGATACCAATAAAATCCCTAAGCATTTGTTTCAGGGTTATTTGAATATCTATAACAACTATTTTAAAGCGCCGTGGATGCCTGACCGTACCGAGGC
TAACCCTAATGAGCTTAATCAAGATGATGCTCGTTATGGTTTCCGTTGCTGCCATCTCAAAAACATTTGGACTGCTCCGCTTCCTCCTGAGACTGAGCTTTCTCGCCAAATGACGACTTC
TACCACATCTATTGACATTATGGGTCTGCAAGCTGCTTATGCTAATTTGCATACTGACCAAGAACGTGATTACTTCATGCAGCGTTACCATGATGTTATTTCTTCATTTGGAGGTAAAAC
CTCTTATGACGCTGACAACCGTCCTTTACTTGTCATGCGCTCTAATCTCTGGGCATCTGGCTATGATGTTGATGGAACTGACCAAACGTCGTTAGGCCAGTTTTCTGGTCGTGTTCAACA
GACCTATAAACATTCTGTGCCGCGTTTCTTTGTTCCTGAGCATGGCACTATGTTTACTCTTGCGCTTGTTCGTTTTCCGCCTACTGCGACTAAAGAGATTCAGTACCTTAACGCTAAAGG
TGCTTTGACTTATACCGATATTGCTGGCGACCCTGTTTTGTATGGCAACTTGCCGCCGCGTGAAATTTCTATGAAGGATGTTTTCCGTTCTGGTGATTCGTCTAAGAAGTTTAAGATTGC
TGAGGGTCAGTGGTATCGTTATGCGCCTTCGTATGTTTCTCCTGCTTATCACCTTCTTGAAGGCTTCCCATTCATTCAGGAACCGCCTTCTGGTGATTTGCAAGAACGCGTACTTATTCG
CCACCATGATTATGACCAGTGTTTCCAGTCCGTTCAGTTGTTGCAGTGGAATAGTCAGGTTAAATTTAATGTGACCGTTTATCGCAATCTGCCGACCACTCGCGATTCAATCATGACTTC
GTGATAAAGATTGAGTGTGAGGTTATAACGCCGAAGCGGTAAAAATTTTAATTTTTGCCGCTGAGGGGTTGACCAAGCGAAGCGCGGTAGGTTTTCTGCTTAGGAGTTTAATCATGTTT
CAGACTTTTATTTCTCGCCATAATTCAAACTTTTTTTCTGATAAGCTGGTTCTCACTTCTGTTACTCCAGCTTCTTCGGCACCTGTTTTACAGACACCTAAAGCTACATCGTCAACGTTA
TATTTTGATAGTTTGACGGTTAATGCTGGTAATGGTGGTTTTCTTCATTGCATTCAGATGGATACATCTGTCAACGCCGCTAATCAGGTTGTTTCTGTTGGTGCTGATATTGCTTTTGAT
GCCGACCCTAAATTTTTTGCCTGTTTGGTTCGCTTTGAGTCTTCTTCGGTTCCGACTACCCTCCCGACTGCCTATGATGTTTATCCTTTGAATGGTCGCCATGATGGTGGTTATTATACC
GTCAAGGACTGTGTGACTATTGACGTCCTTCCCCGTACGCCGGGCAATAACGTTTATGTTGGTTTCATGGTTTGGTCTAACTTTACCGCTACTAAATGCCGCGGATTGGTTTCGCTGAAT
CAGGTTATTAAAGAGATTATTTGTCTCCAGCCACTTAAGTGAGGTGATTTATGTTTGGTGCTATTGCTGGCGGTATTGCTTCTGCTCTTGCTGGTGGCGCCATGTCTAAATTGTTTGGAG
GCGGTCAAAAAGCCGCCTCCGGTGGCATTCAAGGTGATGTGCTTGCTACCGATAACAATACTGTAGGCATGGGTGATGCTGGTATTAAATCTGCCATTCAAGGCTCTAATGTTCCTAACC
CTGATGAGGCCGCCCCTAGTTTTGTTTCTGGTGCTATGGCTAAAGCTGGTAAAGGACTTCTTGAAGGTACGTTGCAGGCTGGCACTTCTGCCGTTTCTGATAAGTTGCTTGATTTGGTTG
GACTTGGTGGCAAGTCTGCCGCTGATAAAGGAAAGGATACTCGTGATTATCTTGCTGCTGCATTTCCTGAGCTTAATGCTTGGGAGCGTGCTGGTGCTGATGCTTCCTCTGCTGGTATGG
TTGACGCCGGATTTGAGAATCAAAAAGAGCTTACTAAAATGCAACTGGACAATCAGAAAGAGATTGCCGAGATGCAAAATGAGACTCAAAAAGAGATTGCTGGCATTCAGTCGGCGACTT
CACGCCAGAATACGAAAGACCAGGTATATGCACAAAATGAGATGCTTGCTTATCAACAGAAGGAGTCTACTGCTCGCGTTGCGTCTATTATGGAAAACACCAATCTTTCCAAGCAACAGC
AGGTTTCCGAGATTATGCGCCAAATGCTTACTCAAGCTCAAACGGCTGGTCAGTATTTTACCAATGACCAAATCAAAGAAATGACTCGCAAGGTTAGTGCTGAGGTTGACTTAGTTCATC
AGCAAACGCAGAATCAGCGGTATGGCTCTTCTCATATTGGCGCTACTGCAAAGGATATTTCTAATGTCGTCACTGATGCTGCTTCTGGTGTGGTTGATATTTTTCATGGTATTGATAAAG
CTGTTGCCGATACTTGGAACAATTTCTGGAAAGACGGTAAAGCTGATGGTATTGGCTCTAATTTGTCTAGGAAATAACCGTCAGGATTGACACCCTCCCAATTGTATGTTTTCATGCCTC
CAAATCTTGGAGGCTTTTTTATGGTTCGTTCTTATTACCCTTCTGAATGTCACGCTGATTATTTTGACTTTGAGCGTATCGAGGCTCTTAAACCTGCTATTGAGGCTTGTGGCATTTCTA
CTCTTTCTCAATCCCCAATGCTTGGCTTCCATAAGCAGATGGATAACCGCATCAAGCTCTTGGAAGAGATTCTGTCTTTTCGTATGCAGGGCGTTGAGTTCGATAATGGTGATATGTATG
TTGACGGCCATAAGGCTGCTTCTGACGTTCGTGATGAGTTTGTATCTGTTACTGAGAAGTTAATGGATGAATTGGCACAATGCTACAATGTGCTCCCCCAACTTGATATTAATAACACTA
TAGACCACCGCCCCGAAGGGGACGAAAAATGGTTTTTAGAGAACGAGAAGACGGTTACGCAGTTTTGCCGCAAGCTGGCTGCTGAACGCCCTCTTAAGGATATTCGCGATGAGTATAATT
ACCCCAAAAAGAAAGGTATTAAGGATGAGTGTTCAAGATTGCTGGAGGCCTCCACTATGAAATCGCGTAGAGGCTTTGCTATTCAGCGTTTGATGAATGCAATGCGACAGGCTCATGCTG
ATGTTGGTTTATCGTTTTTGACACTCTCACGTTGGCTGACGACCGATTAGAGGCGTTTTATGATAATCCCAATGCTTTGCGTGACTATTTTCGTGATATTGGTCGTATGGTTCTTGCTG
CCGAGGGTCGCAAGGCTAATGATTCACACGCCGACTGCTATCAGTATTTTTGTGTGCCTGAGTATGGTACAGCTAATGGCCGTCTTCATTTCCATGCGGTGCACTTTATGCGGACACTTC
CTACAGGTAGCGTTGACCCTAATTTTGGTCGTCGGGTACGCAATCGCCGCCAGTTAAATAGCTTGCAAAATACGTGGCCTTATGGTTACAGTATGCCCATCGCAGTTCGCTACACGCAGG
ACGCTTTTTCACGTTCTGGTTGGTTGTGGCCTGTTGATGCTAAAGGTGAGCCGCTTAAAGCTACCAGTTATATGGCTGTTGGTTTCTATGTGGCTAAATACGTTAACAAAAAGTCAGATA
TGGACCTTGCTGCTAAAGGTCTAGGAGCTAAAGAATGGAACAACTCACTAAAAACCAAGCTGTCGCTACTTCCCAAGAAGCTGTTCAGAATCAGAATGAGCCGCAACTTCGGGATGAAAA
TGCTCACAATGACAAATCTGTCCACGGAGTGCTTAATCCAACTTACCAAGCTGGGTTACGACGCGACGCCGTTCAACCAGATATTGAAGCAGAACGCAAAAAGAGAGATGAGATTGAGGC
TGGGAAAAGTTACTGTAGCCGACGTTTTGGCGGCGCAACCTGTGACGACAAATCTGCTCAAATTTATGCGCGCTTCGATAAAAATGATTGGCGTATCCAACCTGCA (wraparound)

https://www.ncbi.nlm.nih.gov/nuccore/9626372

"Accepted" ORF annotation

phiX174
5386 bp

"Unfiltered" ORF annotation

unfiltered annotation

https://www.google.com/search?q=phix174+benchling

https://www.pnas.org/content/116/48/24206

https://en.wikipedia.org/wiki/Phi_X_174
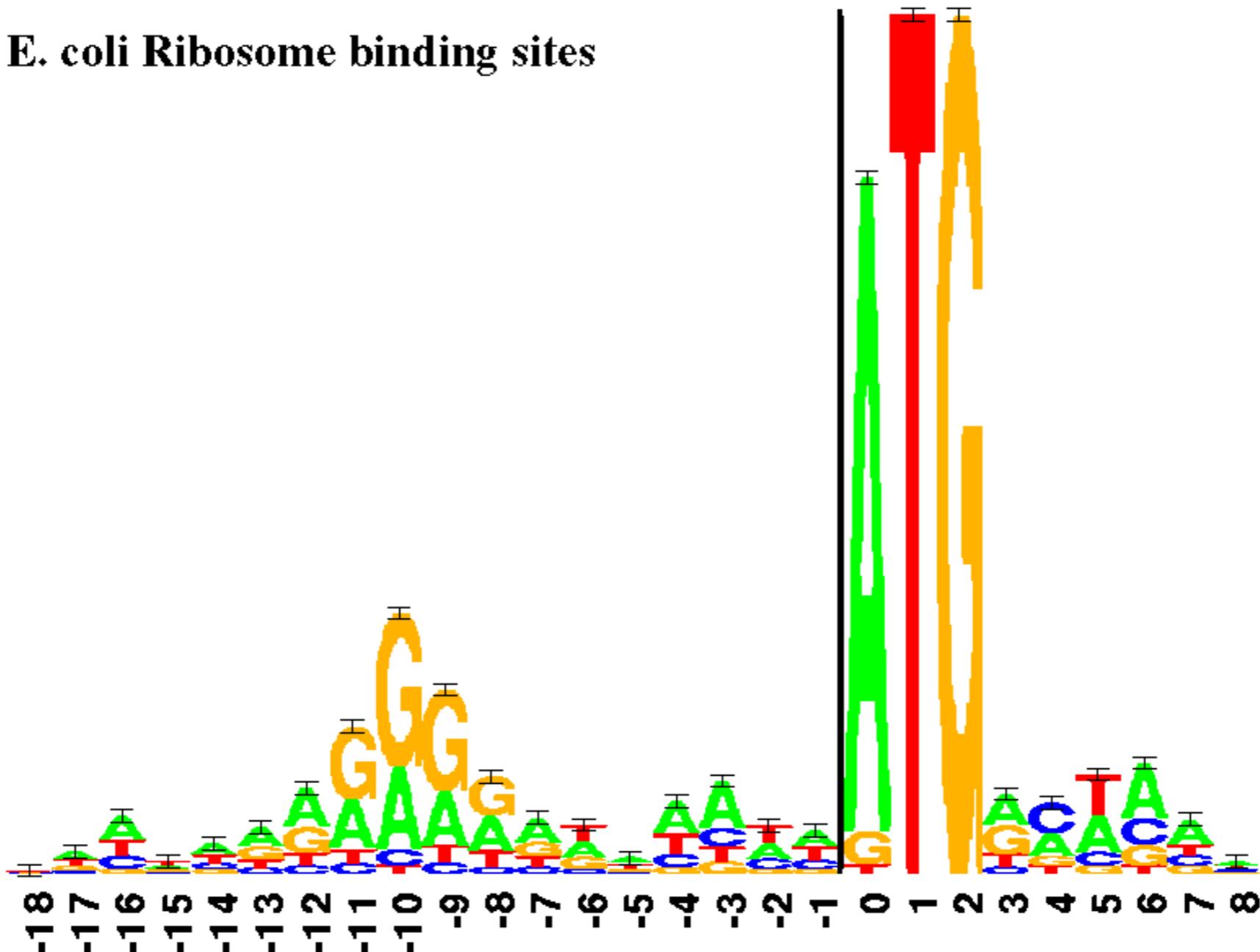
Ribosome loading on mRNA involves RNA:RNA interaction (base pairing)

Ribosome Binding Site (RBS)
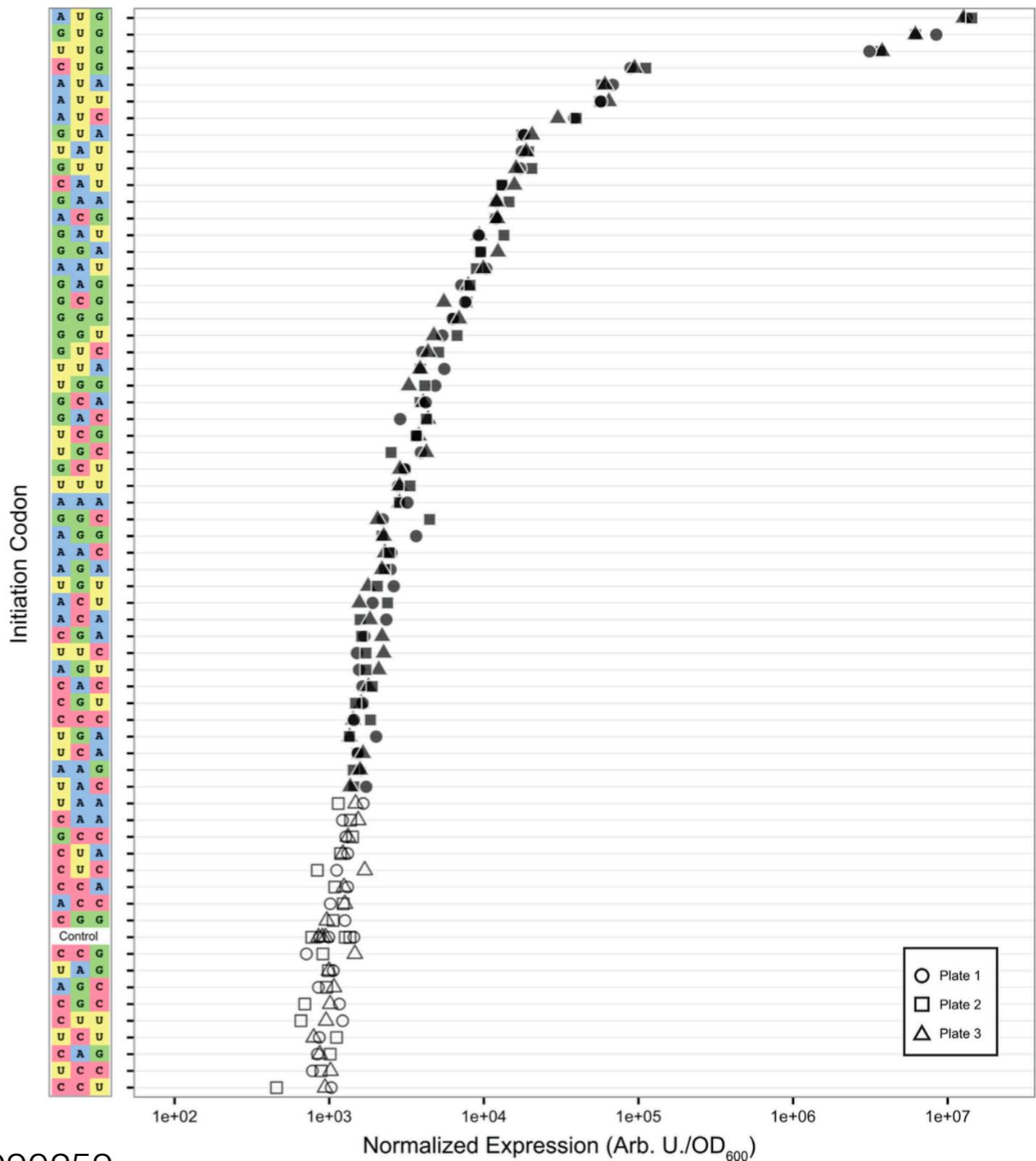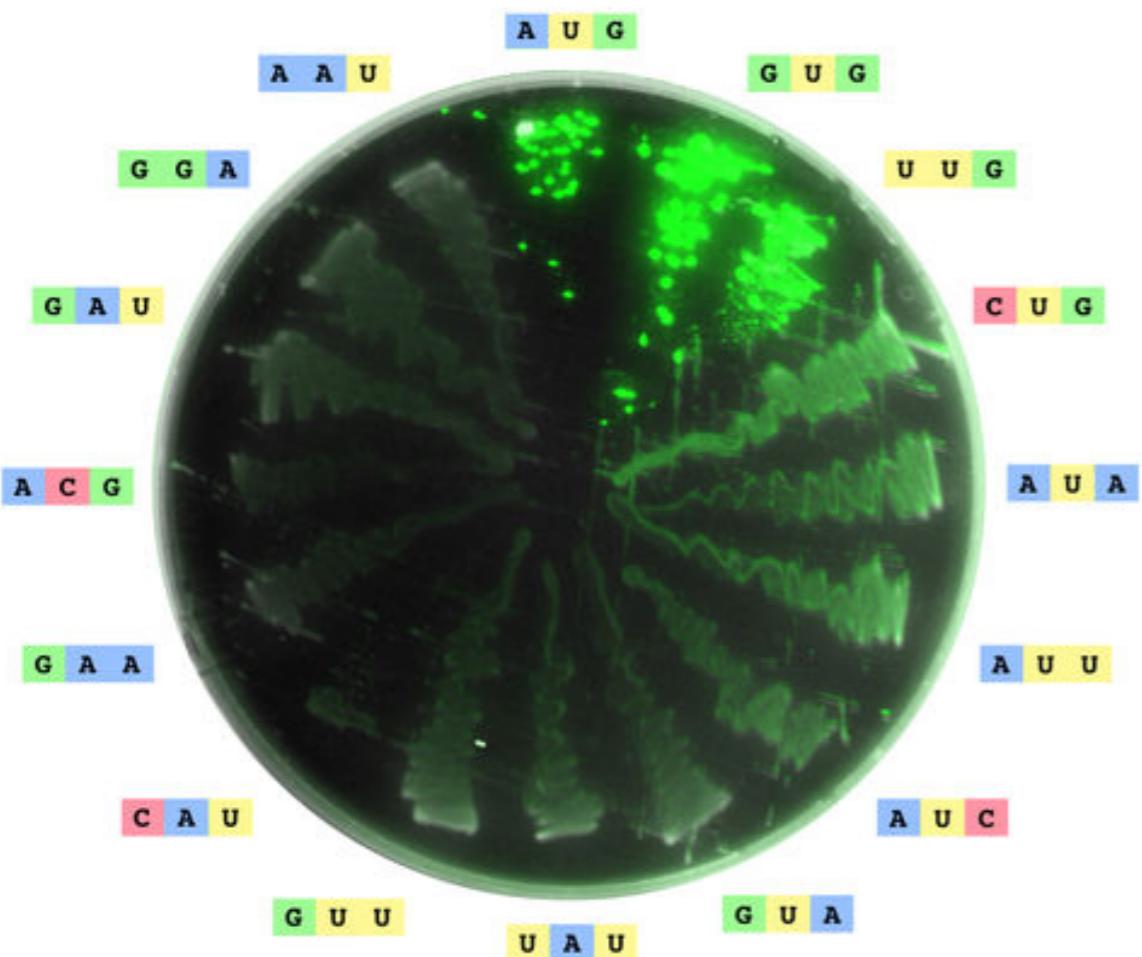
mRNA sequence recognized by ribosomes to initiate translation


E. coli Ribosome binding sites

Advanced Topic

What is a start codon, exactly?

https://academic.oup.com/nar/article/45/7/3615/2990259
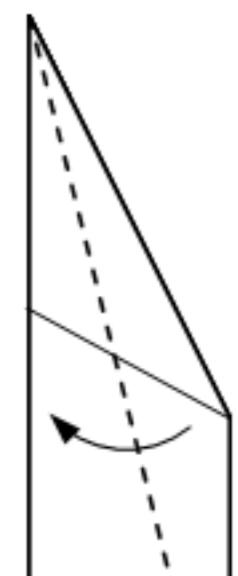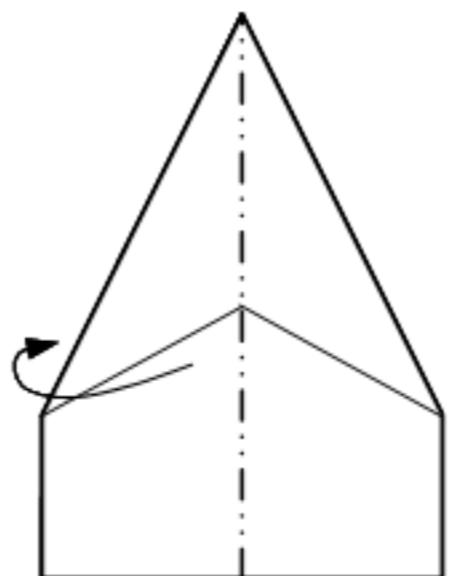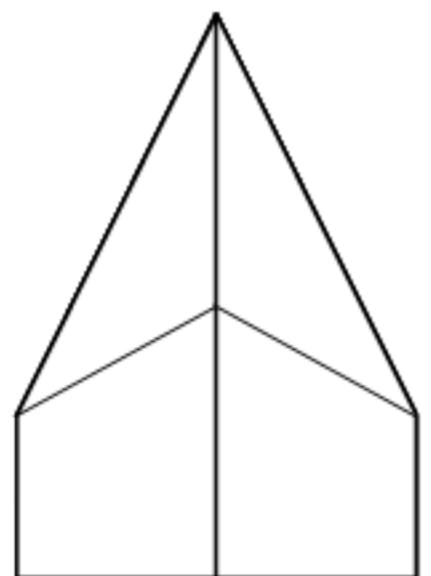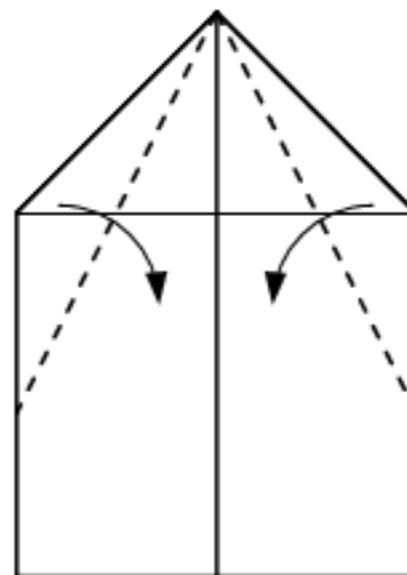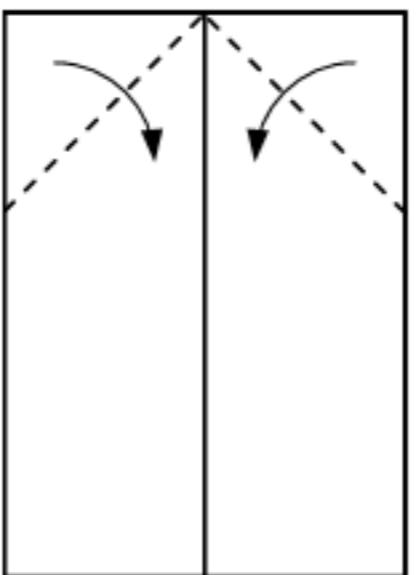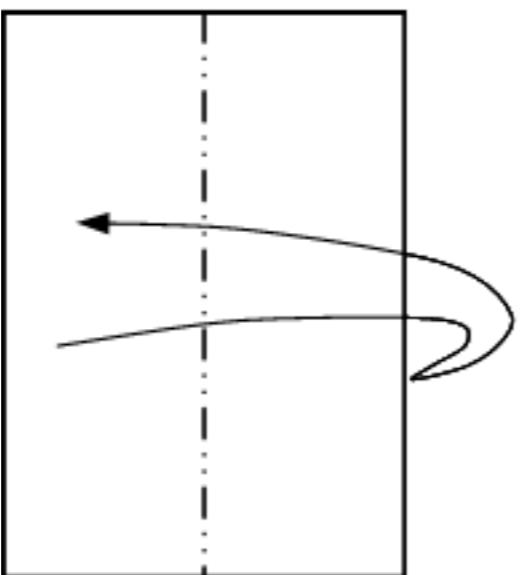
```
main( )
{
    printf("hello, world\n");
}
```
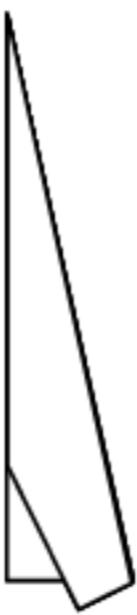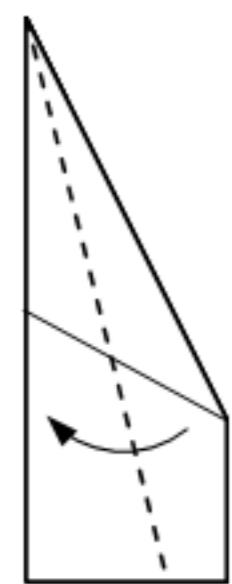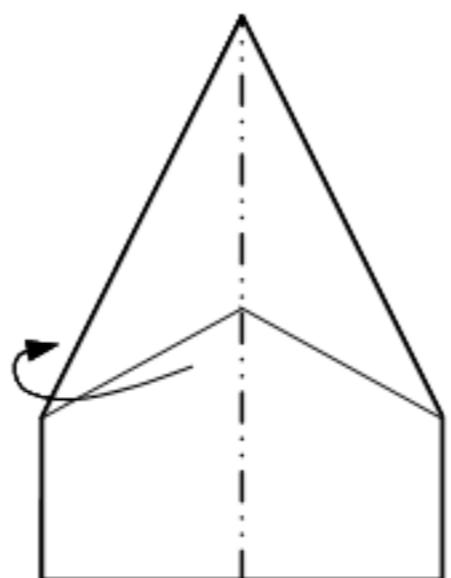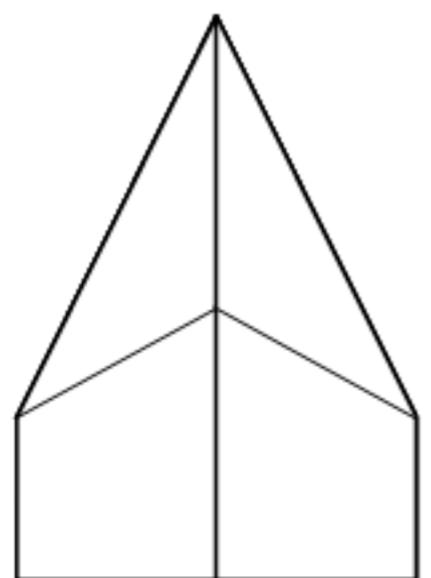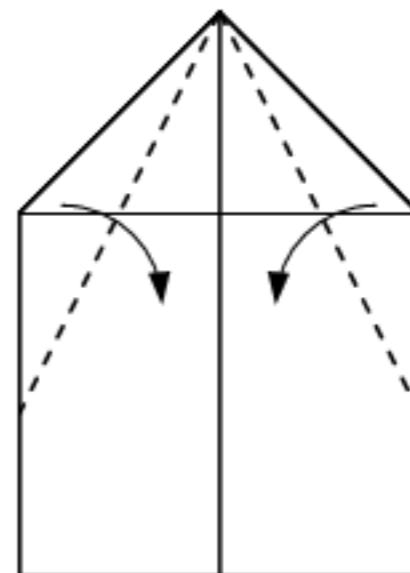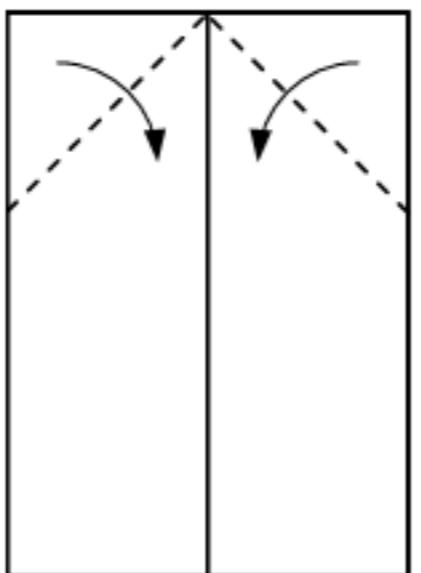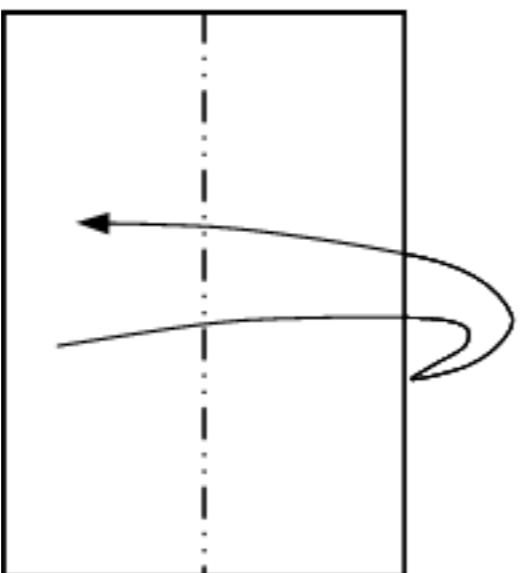
Brian Kernighan
Programming in C: A Tutorial
Bell Labs ~1970s

```
cell( )
{
    express("your favorite protein");
}
```

# What happens when you can't see and feel, etc?

UW iGEM
MAKE IT OR BREAK IT

| Home | About Us | Diesel Production | Gluten Destruction | iGEM Toolkits | Data Page | Protocols | Outreach | Safety |
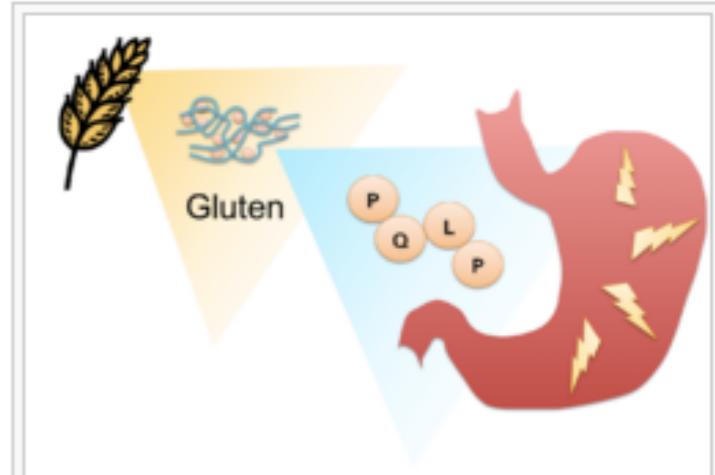
Gluten Destruction: Background

What is Gluten Intolerance?

People who suffer from gluten intolerance have an adverse reaction to gluten proteins found in wheat, barley, and rye products. The glutens invoke an immune response in the digestive tract of genetically predisposed individuals resulting in inflammation of the gut, impeding the absorption of nutrients. Symptoms can appear in early childhood or later in life, and range widely in severity, from diarrhea, fatigue and weight loss to abdominal distension, anemia, and neurological symptoms. There are currently no effective therapies for this lifelong disease except the total elimination of glutens from the diet. Although celiac sprue remains largely underdiagnosed, its prevalence in the US and Europe is estimated at 0.5-1.0% of the population. With this in mind, we set out to design an enzyme therapeutic for gluten intolerance that could be taken in pill form.

Proline (P)- and glutamine (Q)-rich components of gluten known as 'gliadins' appear to be responsible for the bulk of the immune response in most patients. Their high PQ content protects gliadin oligopeptides from degradation by gastrointenstinal endoproteases, but also presents a target for drug design. Any peptidase capable of cleaving at or near the P-Q bond while remaining active at the temperature and harsh pH of the stomach would have pharmacological potential as a therapy for celiac sprue.

Proline(P) and glutamine(Q) -rich peptide fragments of gluten provoke an immune response which causes painful inflammation in the digestive tract of gluten intolerant individuals.

http://2011.igem.org/Team:Washington/Celiacs/Background

Registry of Standard Biological Parts

Part:BBa_K590087

Designed by: Sydney Gordon, Daniel Hadidi, Elizabeth Stanley, Sarah Wolf, Angus Toland, Sean Wu   Group: iGEM11_Washington
(2011-09-22)

Coding
KumaMax

Not Released	
Sample It's complicated	
Experience: Works	
1 Uses	
1 Twin	

Get This Part

KumaMax: Kumamolisin-As_N291D, G319S, D358G, D368H
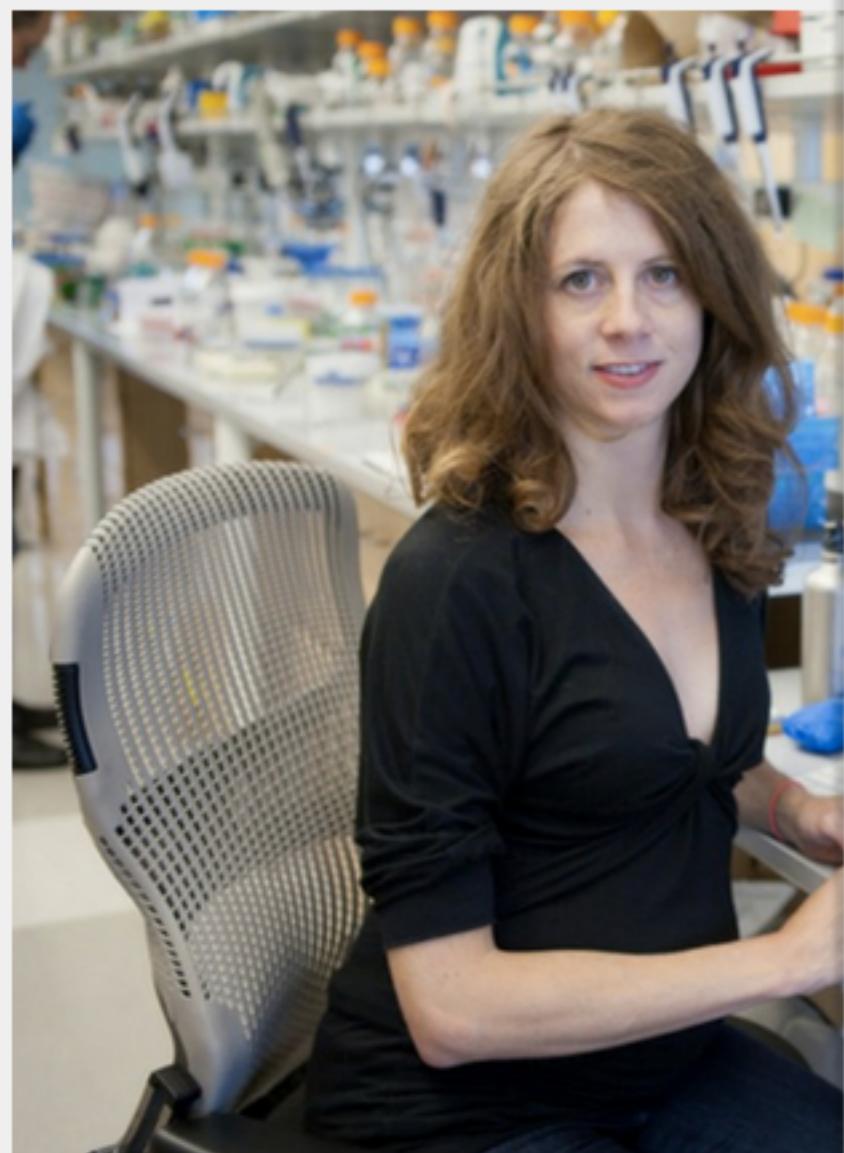
A mutated Kumamolisin-As ⧉ enzyme aimed to break down gluten by increased activity with the PQLP peptide, an antigenic epitope in gliadin.

Usage and Biology

This part (KumaMax) was constructed by the 2011 University of Washington ⧉ iGEM team to break down gluten, the primary cause of Celiac's disease. KumaMax was generated by making rational mutations to the active site of the enzyme, as detailed on our our wiki ⧉. To test BBa _K590021, it was inserted into a protein expression vector, pET29b+. KumaMax (Kumamolisin-As_N219D, S354N, D358G, D368H) was then produced and purified as described in the 2011 iGEM Team's Small Scale Protein Expression and Purification Protocol ⧉. The purified protein was then tested for activity. For a detailed description of the assay, please see the 2011 UW iGEM Purified Enzyme Assay Protocol ⧉. The resulting data is shown below. We achieved an over 100-fold increase in activity on breaking down PQLP from the wild-type enzyme. This variant enzyme is ultimately 784 times better at breaking down PQLP than SC PEP, the enzyme currently in clinical trials for treating gluten intolerance!

http://partsregistry.org/wiki/index.php?title=Part:BBa_K590087

iGEM Startup PvP Biologics Closes a $35 Million Agreement with Pharmaceutical Company Takeda

Ingrid Swanson Pultz, PhD, Translational Investigator
http://www.ipd.uw.edu/2016/01/dr-ingrid-swanson-p...

Barely two months ago, synbio company PvP Biologics, spun out of Washington University. Last week, the company closed a $35 million deal with medical company Takeda to cover phase 1 clinical trials of KumaMax, a new engineered protein capable of breaking down the immune-reactive parts of gluten in the stomach, potentially bettering the life of around 74 million people worldwide estimated to suffer from gluten intolerance.

From iGEM to Company

The work behind the development, however, began five years ago. In 2011, a group of University of Washington undergraduates presented a computer-generated protein design to combat the symptoms of celiac disease as their entry for that year's iGEM competition. The logic behind their idea was that instead of focusing on substrate specificity when choosing their base candidate enzyme, they would first identify an enzyme that was already capable of working properly in the necessary conditions – that is, in the highly acidic pH level of the human stomach. Once said that base enzyme was selected, the team worked to reengineer its substrate specificity through in silico tools, making it now gluten-specific. They went on to become the first US team to win iGEM's global prize.

The project could have ended there, with a nice gold medal and a new BioBrick to UW's name in the Registry of Standard Biological Parts. Instead, recent pHD graduate Ingrid Swanson Pultz decided to take up the further development of KumaMax as their post-doctoral project in 2012 in David Baker's Institute for Protein Design. "The enzyme that the students had generated was still a prototype at that point," says Swanson, "so my goal was to further engineer it and then use more sophisticated methods to test it."

Swanson's interest in engineering life from its base translated nicely from her childhood to her pHD and eventually to the company she founded. "Having been obsessed with Legos as a child, I've always loved the idea of building something that becomes more than the sum of its parts [...] "I liked the idea of studying life on its most basic level, as a simple closed system." When she applied to graduate school, her goal was to learn as much as she could about how bacteria work on a molecular level, since the better she knew their inner working, the better she could reengineer them. "This was in 2005," notes Swanson, "and I had never heard of the term "synthetic biology", but that's what I was interested in".

After reading about iGEM in Nature and finding out the University of Washington, where Swanson was getting her pHD, did not have a team for the competition, she decided to take matters into her own hands and fund the first

After buying Seattle startup PvP Biologics for $330M, Takeda to advance celiac disease treatment
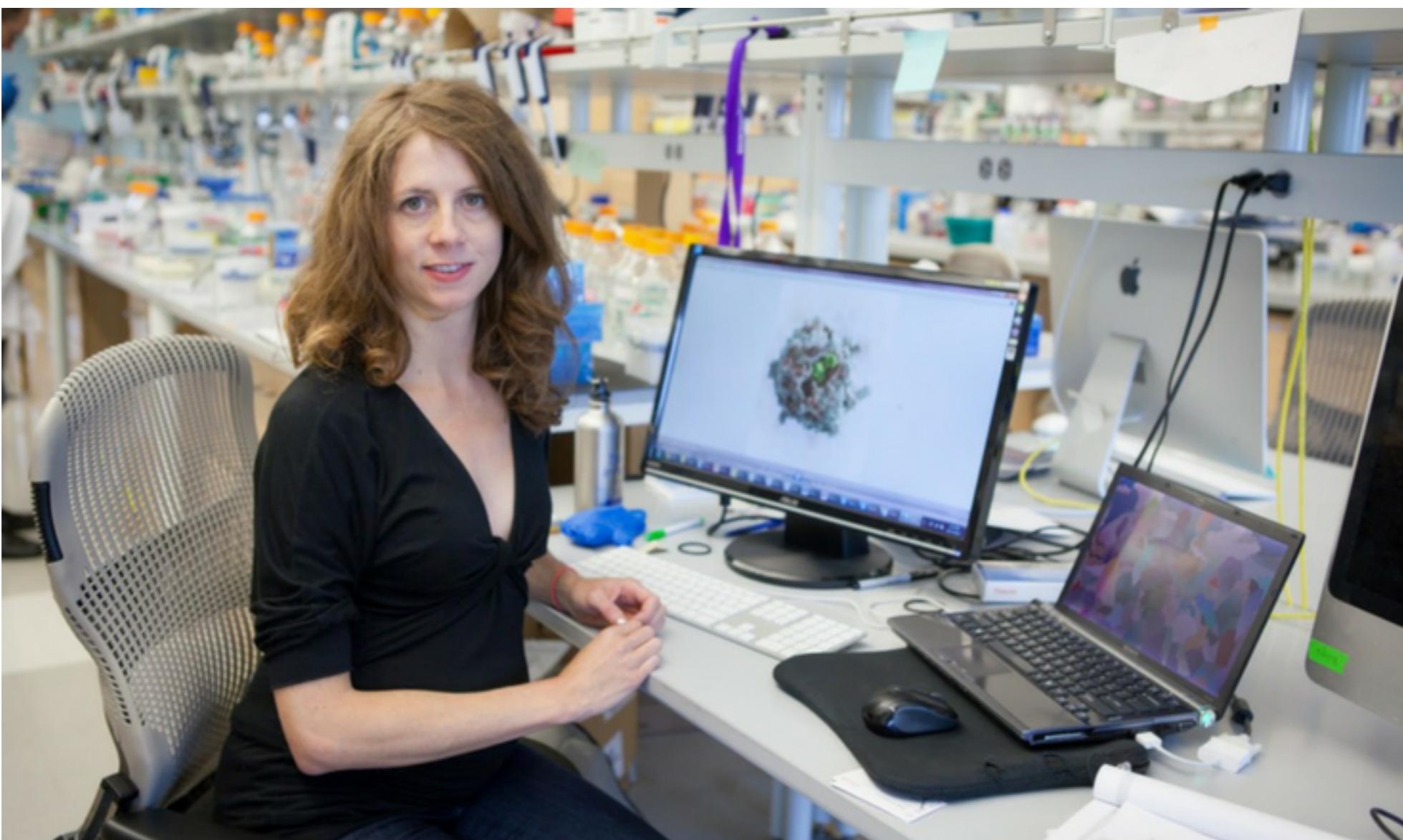
BY **KARINA MAZHUKHINA** on April 8, 2020 at 5:41 pm

Members of the PvP Biologics team at a company gathering prior to its acquisition. (PvP Photo)

Seattle biotech startup PvP Biologics, which developed a promising treatment for people who can't digest gluten, was acquired by Japanese pharmaceutical Takeda for up to $330 million in late February.

Since the acquisition, Takeda has taken over all clinical work, as well as chemistry, manufacturing and control activities. It also laid off PVP's entire staff, said Ingrid Swanson Pultz, co-founder and chief scientific officer of PvP, who was one of the few retained part-time contractors helping Takeda with the transition.
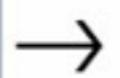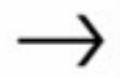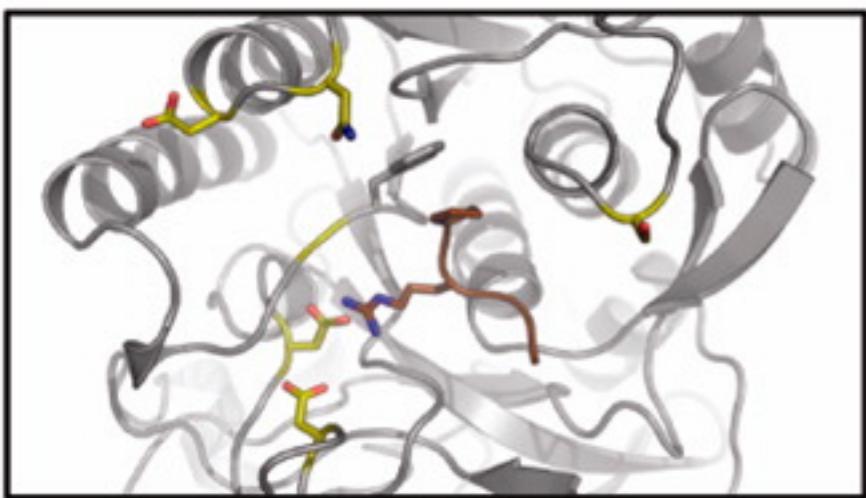
"An ideal oral enzyme therapeutic (OET) for celiac disease would have the following traits:

(1) optimal activity at the pH of the stomach after a meal (in the range of 2–4);([7])
(2) resistance to common digestive proteases;
(3) facile recombinant production in a soluble form; and
(4) specificity for the common proline–glutamine (PQ) motif found in immunogenic α-gliadin oligopeptides.([6], [8])"
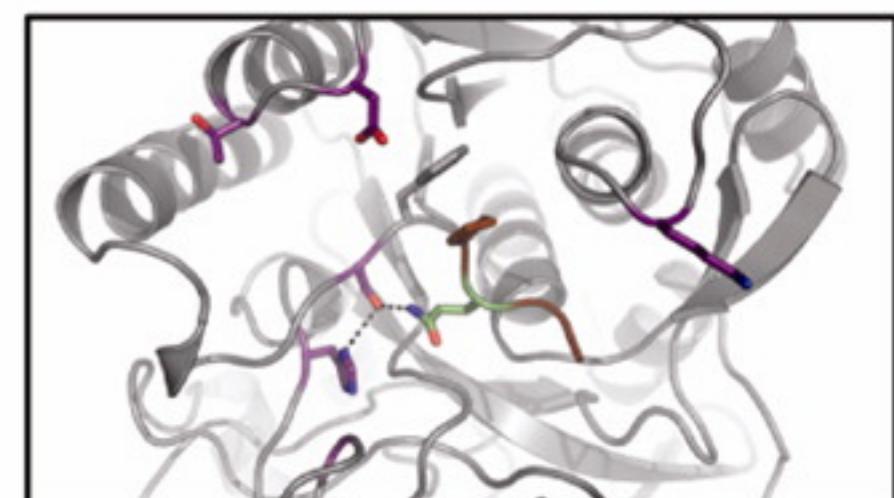
Gordon, Sydney R., et al. "Computational design of an α-gliadin peptidase." *Journal of the American Chemical Society* **134.50 (2012): 20513-20520.**

7 mutations provide solution…

Native Enzyme → Engineered Enzyme

|	Native Motif	
..xxPRxx.. k_{cat}/K_M=132 M^{-1}s^{-1} ←		→ ..xxPRxx.. no activity
..xxPQxx.. k_{cat}/K_M=5 M^{-1}s^{-1} ←	α-gliadin Motif	→ ..xxPQxx.. k_{cat}/K_M=569 M^{-1}s^{-1}
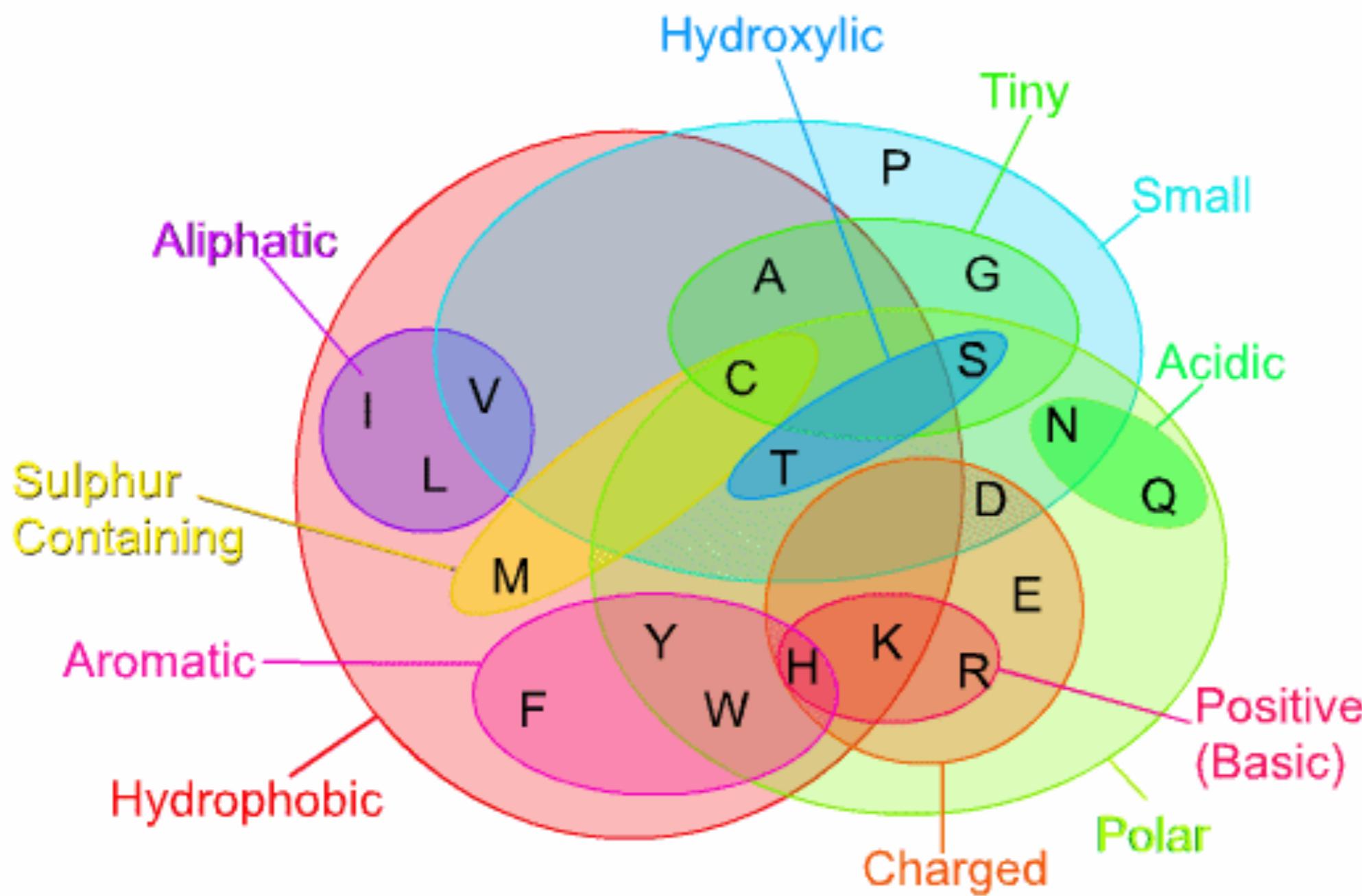
V119D, S262K, N291D, D293T, G319S, D358G, D368H
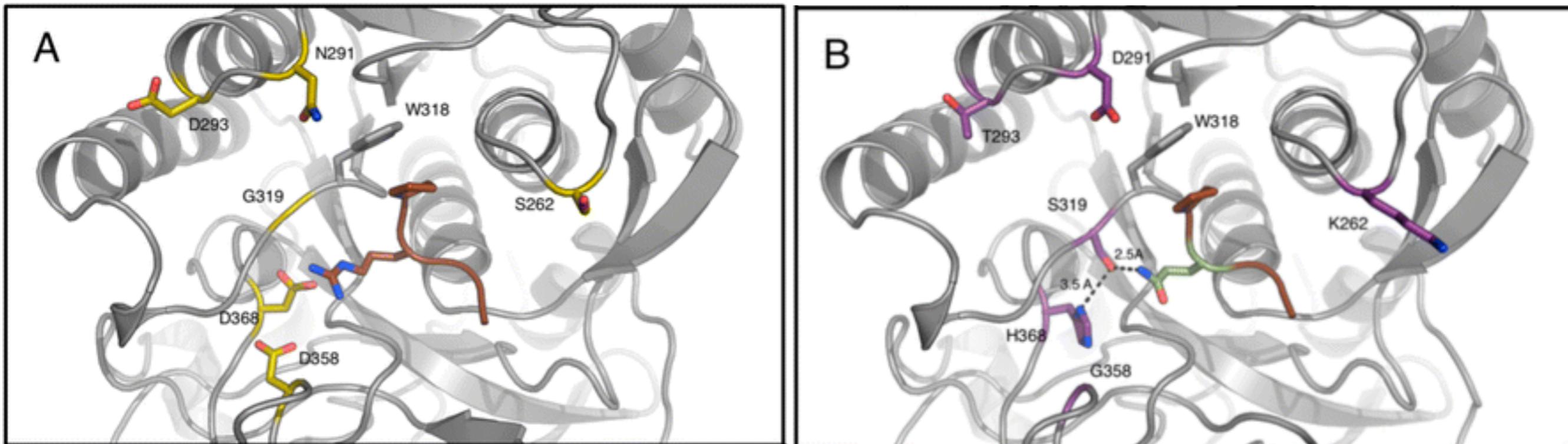
Gordon, Sydney R., et al. "Computational design of an α-gliadin peptidase."
Journal of the American Chemical Society 134.50 (2012): 20513-20520.

"Eats" collagen (PR) v. "Eats" gluten (PQ)

V119D, S262K, N291D, D293T, G319S, D358G, D368H

- ○ V119D (not shown) l in propeptide domain, does not affect catalytic activity
- ○ S262K (S73K) l Likely introduces interaction other residues outside catalytic site
- ○ N291D (N102D)l Likely introduces interaction other residues outside catalytic site
- ○ D293T (D104T) l Likely introduces interaction other residues outside catalytic site
- ○ G319S (G130S) l New H bond w/ Q at P1 position
- ○ D358G (D169G) l New H bond w/ Q at P1 position
- ○ D368H (D179H) l New H bond w/ Q at P1 position

Gordon, Sydney R., et al. "Computational design of an α-gliadin peptidase."
*Journal of the American Chemical Society* 134.50 (2012): 20513-20520.
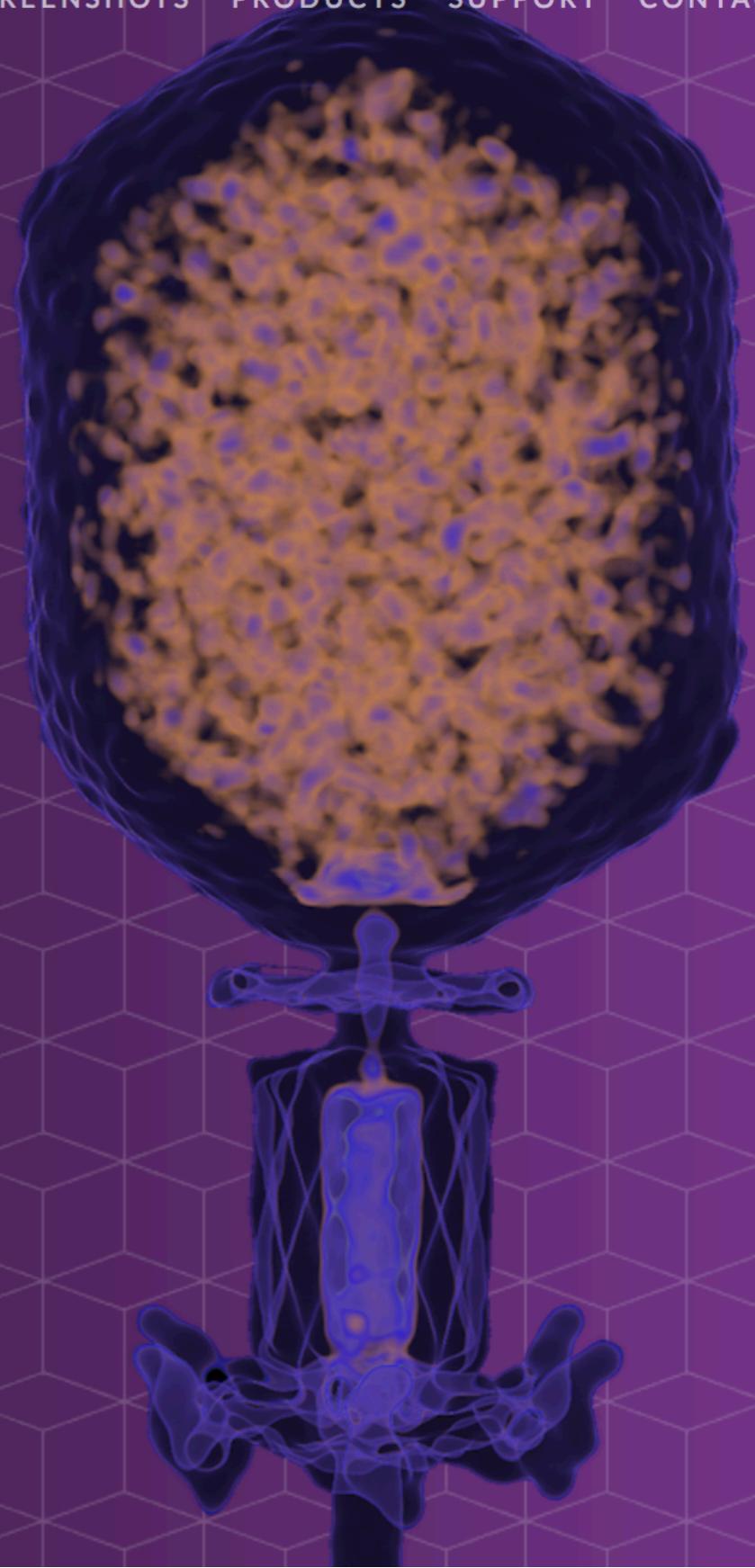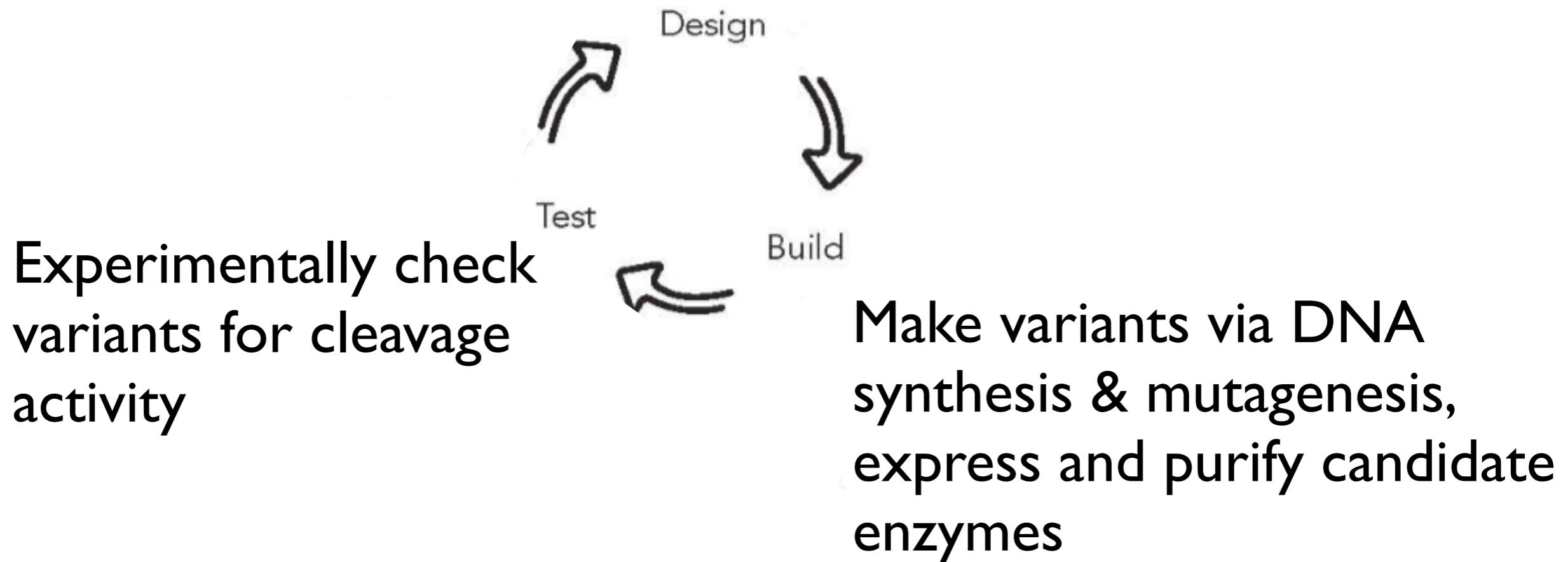
To summarize Ingrid's workflow…

Using modeling & simulation tools (PyMol & Rosetta) to suggest mutants that better bind desired substrate motif (PQLP)

Experimentally check variants for cleavage activity

Make variants via DNA synthesis & mutagenesis, express and purify candidate enzymes

Graphic from "Synthetic Aesthetics" MIT Press (2014)

RCSB PDB

Protein Data Bank

139187 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligand | Go

Advanced Search | Browse by Annotations

PDB-101 | WORLDWIDE PDB PROTEIN DATA BANK | EMDataBank Unified Data Resource for 3DEM | ndb NUCLEIC ACID DATABASE | Worldwide Protein Data Bank Foundation

- Welcome
- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology
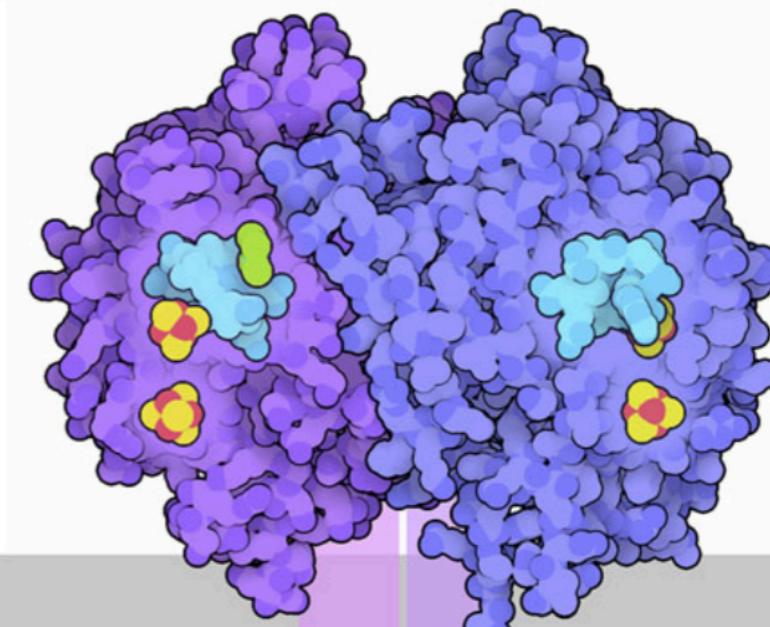
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

### New Video: What is a Protein?
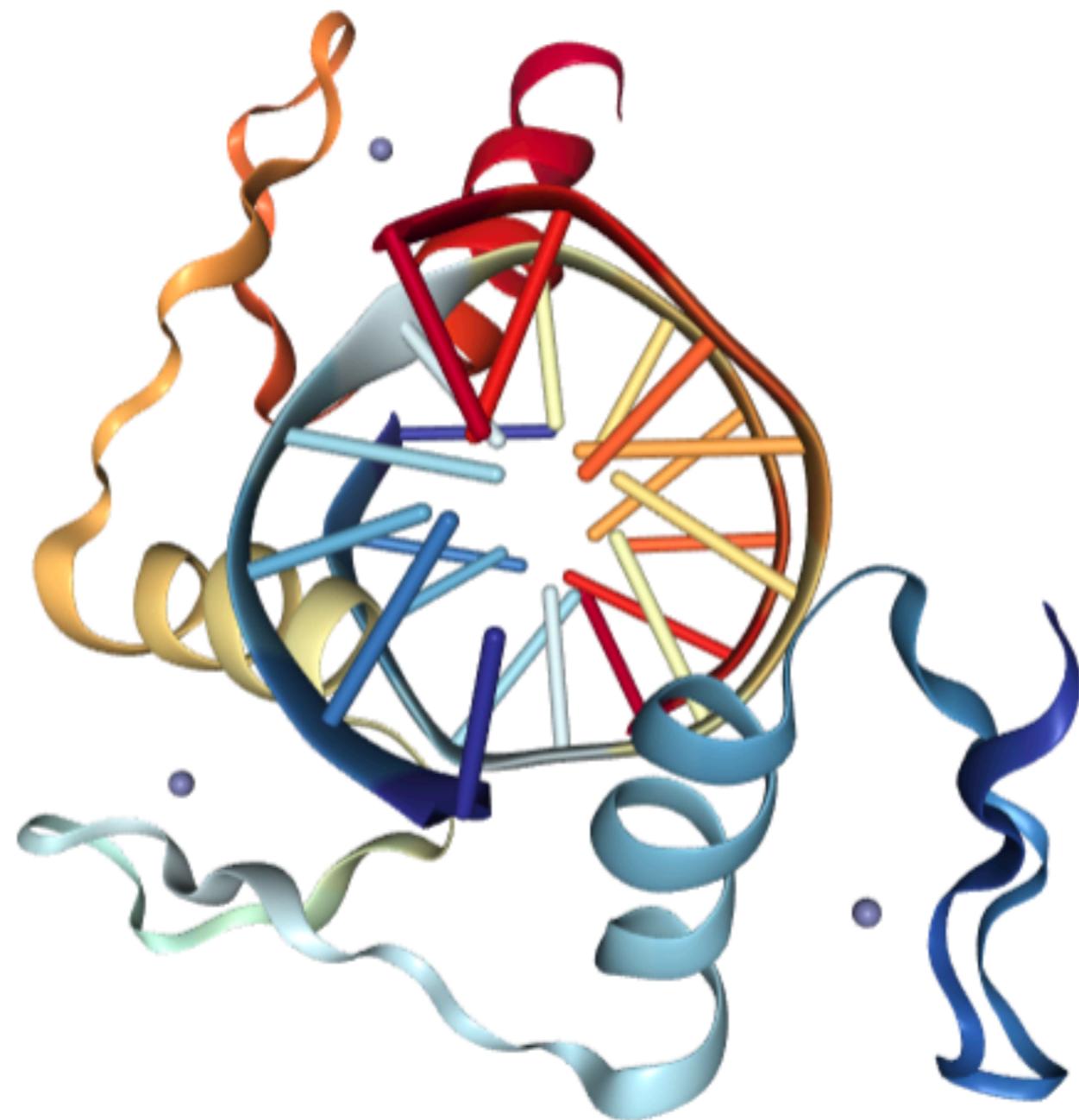
PDB-101 VIDEO WHAT IS A PROTEIN?
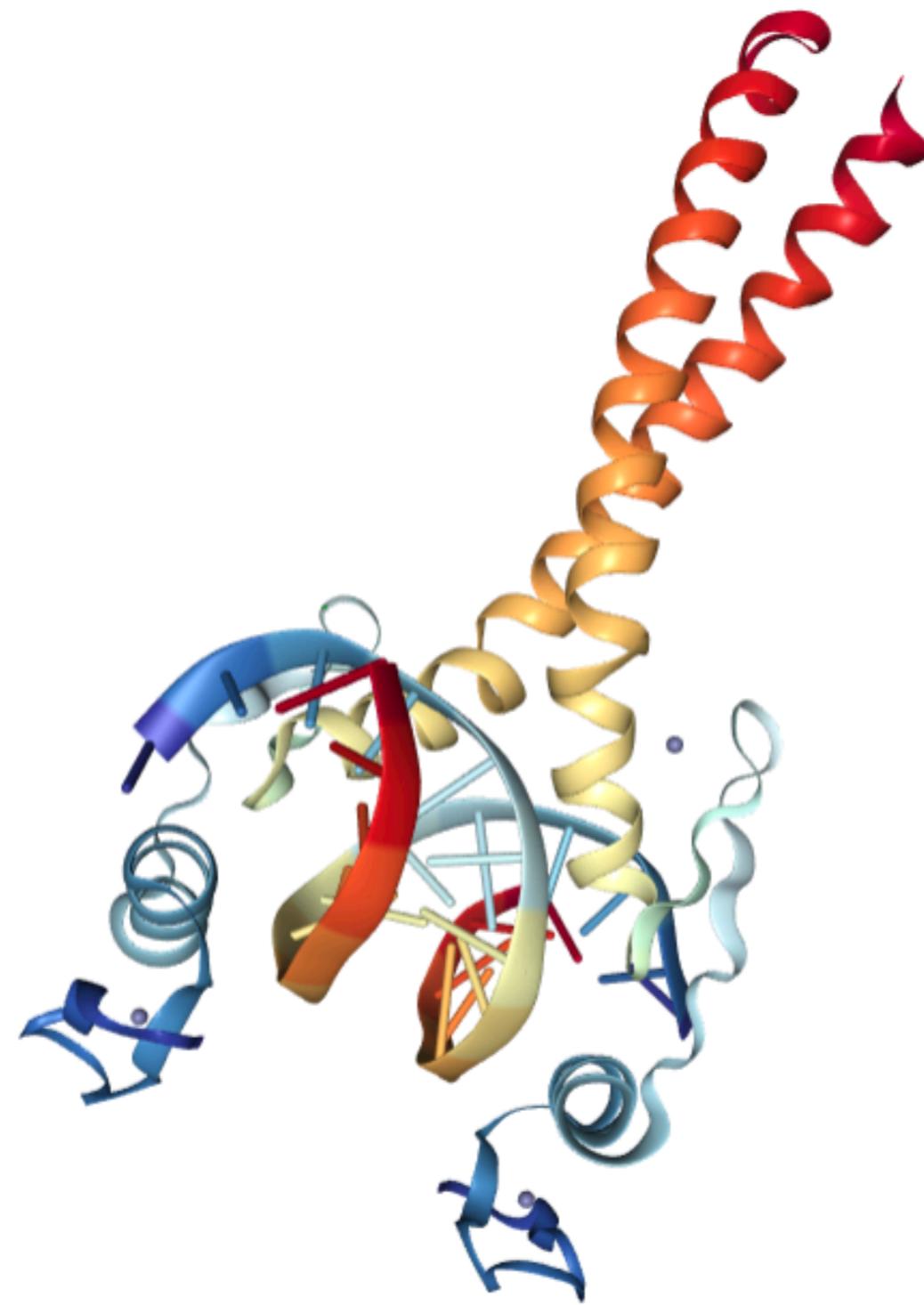
April Molecule of the Month

Dehalogenases

https://www.rcsb.org/

https://www.rcsb.org/3d-view/1AAY/1

https://www.rcsb.org/3d-view/1LLM/1
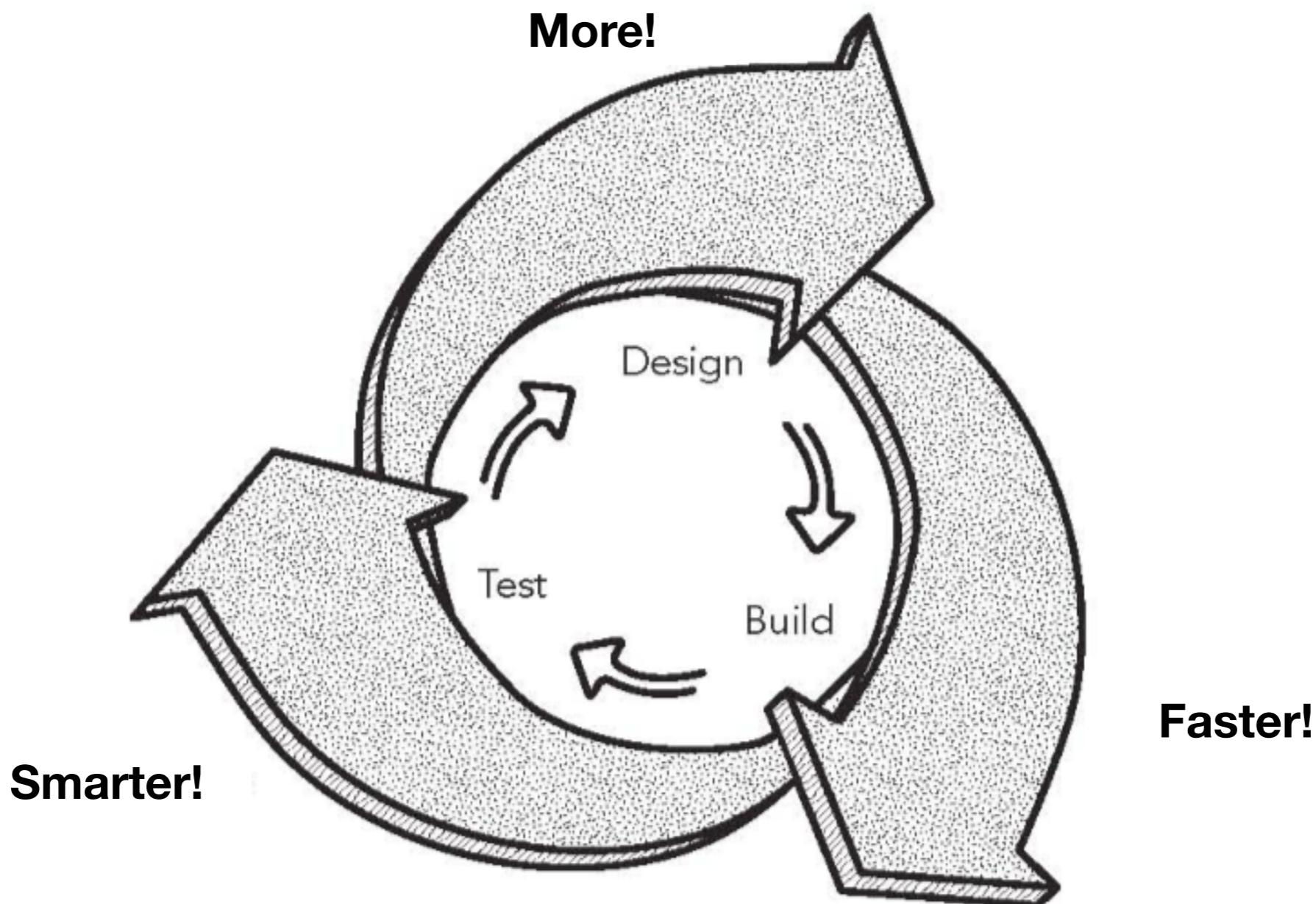
Analysis & design tools
(model) (make) (measure)

More!

Design

Test

Build

Faster!

Smarter!

Improving the tools that enable the core engineering cycle (above) is a big part of bioengineering.

BREAKOUT
Better tools…

Based on what you know and have learned so far, what features or wishes do you have for biomolecular analysis and design tools?

Hint — Try Framestorm & Brainstorm Skills