The T4 rII genes and the history of molecular genetics

Sean Eddy
Harvard University

The classic experiments of Crick, Brenner, Benzer, and others depended on elegant logic and genetics – but not on what the rII genes actually do.

What *do* the rII genes do?

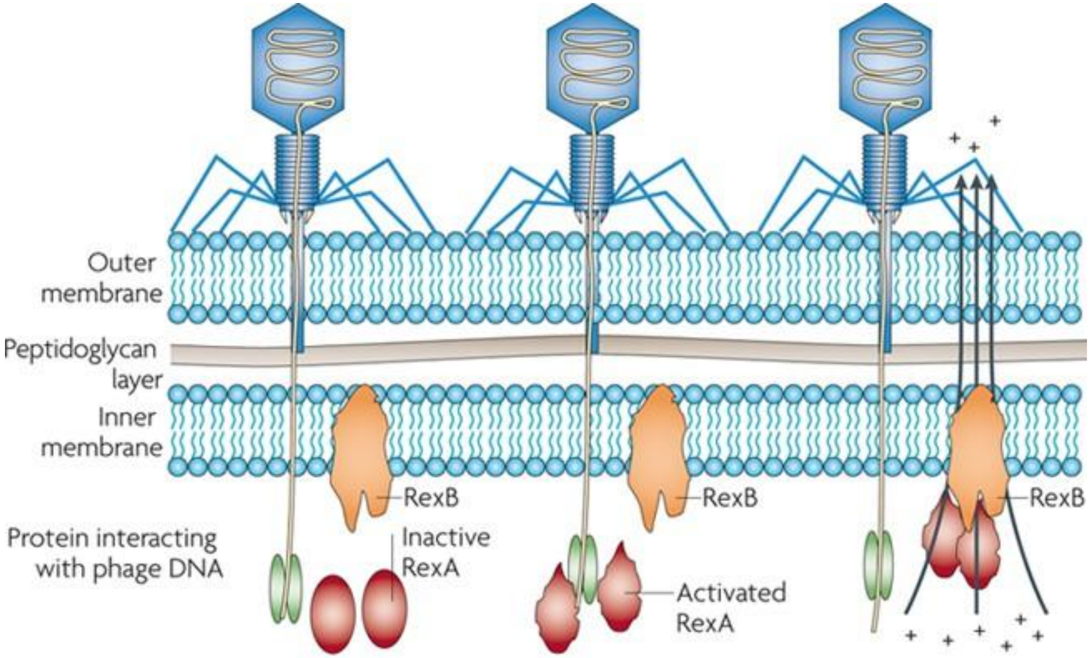
> *"...the status of the rII genes of T4 in the history of genetics compares to that of the gene causing Mendel's smooth and rough peas..."*
>
> Snyder and Kauffman (1994)

> *"... the r-controlled lysis phenomenon remains shrouded in mystery, all the more vexing in light of how long this event has been available for genetic analysis, and also, perhaps more embarrassingly, how intimately related it is to some of the definitive and founding experiments of molecular genetics..."*
>
> Young (1992)

Phage lambda rexA/rexB are an antiphage defense system

rexB, the boobytrap: an activatable proton pore
rexA, the trigger
rIIA/rIIB somehow counter the rexA/rexB system

lambda,
integrated in
host bacterial genome



Labrie, Samson, Moineau Nat Rev Micro (2010)

David Parma

```
ATGAAATTTGTAAAAATTGATTCTTCTAGCGTTGATATGAAAAAATATAAATTGCAGAACAATGTTCGTCGTTCTATTAAATCCTCTTCAATGAACTATGCGAATGTCGCTATTATGACAGACGCAGATCACGATGGATAGCCTTCGGGC
TATCTATAGAAATACCTCATAATTAAGAGATTATTGGATTAGGTTCTATTTATCCTTCTCTGCTCGGATTTTTTAGTAATTGGCCAGAATTGTTTGAGCAAGGACGAATTCGCTTTGTCAAAACTCCTGTAATCATCGCTCAGGTCGGTA
AAAAACAAGAATGGTTTTATACAGTCGCTGAATATGAGAGTGCCAAAGATGCTCTACCTAAACATAGCATCCGTTATATTAAGGGACTTGGCTCTTTGGAAAAATCTGAATATCGTGAGATGATTCAAAACCCAGTATATGATGTTGTTA
AACTTCCTGAGAACTGGAAAGAGCTTTTTGAAATGCTCATGGGAGATAATGCTGACCTTCGTAAAGAATGGATGAGCCAGTAGTTTACTTTACCACAAGGATGTGGTATAATTAATTGGGCAAATGAGGATATTGAAATGAAATCATATA
AAGTAAATTTAGAACTTTTTGATAAAGCAGTTCATCGAGAATATAGAATCATTCAACGCTTTTTCGATATGGGAGAAGCCGAAGAATTTAAAACCCGCTTTAAAGATATTAGAGATAAAATTCAATCCGACACCGCAACTAAAGATGAAC
TACTAGAAGTTGCTGAAGTTATTAAGCGTAATATGAATTAATGAGGAAATTATGATTATCACCACTGAAAAAGAAACAATTCTTGGTAATGGTTCTAAATCAAAAGCATTTAGCATCACAGCATCTCCTAAAGTATTTAAAATTCTGTCA
TCTGATTTGTATACAAACAAGATTCGCGCAGTAGTCCGTGAATTGATTACTAACATGATTGATGCTCATGCACTTAATGGAAATCCTGAAAAATTTATCATACAAGTTCCTGGACGTTTAGACCCACGATTTGTTTGTCGAGATTTTGGT
CCGGGTATGAGTGATTTTGATATTCAAGGTGATGATAATTCTCCTGGGTTGTATAATTCATACTTCAGTTCATCTAAAGCTGAATCTAATGACTTTATTGGCGGATTTGGTTTAGGTTCTAAATCTCCGTTTAGTTATACTGATACGTTT
AGTATTACTTCGTATCATAAAGGTGAAATTCGTGGTTATGTAGCTTACATGGATGGTGATGGTCCACAGATTAAACCTACATTCGTAAAAGAAATGGGTCCAGATGATAAAACTGGTATTGAAATCGTAGTTCCAGTTGAAGAAAAAGAC
TTTAGAAACTTTGCTTATGAAGTTTCTTATATCATGCGACCGTTCAAAGATTTGGCTATCATTAATGGTCTTGACCGCGAAATTGATTATTTTCCGGATTTTGATGACTATTACGGTGTAAATCCAGAAAGATACTGGCCTGATCGTGGT
GGATTATATGCTATCTACGGTGGTATTGTTTATCCTATCGATGGTGTTATTAGAGACCGTAACTGGCTAAGCATTCGCAATGAAGTGAATTACATTAAGTTTCCAATGGGTTCACTTGATATTGCTCCATCTCGCGAGGCTCTTTCACTG
GATGATCGCACTCGTAAAAATATTATTGAACGAGTTAAAGAACTCAGTGAGAAAGCATTTAATGAAGATGTAAAACGATTTAAAGAATCTACATCTCCTCGTCACACATATCGTGAATTGATGAAGATGGGGTATTCTGCTCGAGATTAT
ATGATTAGTAATTCAGTCAAATTCACGACTAAAAATCTGTCATATAAAAAGATGCAGAGCATGTTTGAACCTGACAGTAAGTTATGCAACGCGGGAGTTGTGTATGAAGTAAATCTTGACCCTCGACTGAAGCGCATTAAGCAAAGTCAT
GAAACTTCAGCCGTTGCATCAAGTTATCGTCTGTTTGGTATTAATACAACAAAAATTAATATCGTTATTGATAATATTAAAAATCGTGTTAATATTGTCCGTGGATTAGCACGTGCGTTAGATGATAGTGAATTTAATAACACTTTGAAT
ATTCATCATAACGAACGTCTTCTGTTTATTAATCCAGAAGTAGAATCGCAGATTGATTTGCTTCCTGATATTATGGCGATGTTTGAAAGTGATGAAGTTAACATTCATTATTTGTCAGAAATTGAAGCTTTAGTAAAAAGTTATATTCCA
AAGGTAGTTAAAAGTAAAGCTCCTCGTCCTAAAGCTGCTACAGCGTTTAAGTTTGAAATTAAAGACGGGCGCTGGGAAAAGAGGAATTATTTACGCCTCACATCAGAAGCAGATGAAATTACTGGTTATGTAGCGTATATGCATCGTTCT
GATATTTTCTCTATGGATGGTACTACATCTCTTTGTCATCCATCTATGAATATTTTGATTCGTATGGCTAATCTTATTGGCATTAATGAATTTTATGTTATTCGTCCGCTTTTACAGAAAAAGGTAAAAGAACTCGGTCAGTGCCAATGT
ATTTTTGAAGCTTTGCGTGATTTATATGTAGATGCTTTTGATGATGTAGATTATGATAAGTATGTAGGTTATTCAAGTTCAGCTAAACGATATATTGATAAAATTATCAAGTATCCTGAGTTAGATTTTATGATGAAGTACTTCAGTATA
GATGAAGTTTCTGAAGAATATACACGACTCGCTAATATGGTTAGTTCATTACAGGGTGTATATTTTAATGGTGGAAAAGATACCATCGGTCATGACATTTGGACAGTAACTAATCTTTTTGATGTATTATCAAATAATGCTTCAAAAAAC
AGTGATAAAATGGTTGCTGAGTTTACCAAGAAATTCCGTATTGTTTCCGACTTCATCGGTTATCGCAACTCTTTAAGTGATGATGAAGTTTCCCAAATCGCTAAAACTATGAAGGCCCTTGCGGCCTAATAAGGAAAATTATGTACAATA
TTAAATGCCTGACCAAAAACGAACAAGCTGAAATTGTTAAACTGTATTCAAGTGGTAATTACACCCAACAGGAATTGGCTGATTGGCAAGGTGTATCGGTTGACACAATCCGTCGTGTTTTGAAAAATGCTGAAGAAGCTAAACGCCCTA
AAGTTACTATTAGCGGTGATATTACAGTTAAAGTTAATAGCGATGCAGTTATTGCTCCAGTTGCTAAATCTGACATTATTTGGAATGCATCTAAAAAAATTCATTTCAATTACTGTTGATGGCGTAACTTATAACGCAACTCCTAATACTC
ATTCAAACTTCCAGGAAATTCTTAATCTGCTTGTAGCGGATAAGTTGGAAGAAGCGGCACAAAAAAATTAATGTTCGTCGTGCTGTTGAAAAAATATATTTCCGGCGATGTTCGAATTGAAGGTGGAAGCTTGTTCTATCAAAATATTGAAT
TGCGGTCTGGTTTGGTTGATCGTATTCTTGACTCGATGGAAAAAGGCGAAAACTTTGAATTTTATTTTCCGTTCTTGGAAAATCTGTTGGAAAACCCAAGCCAAAAAGCGGTATCTCGACTCTTTGATTTCTTGGTAGCAAACGATATTG
AAATTACAGAAGATGGTTACTTCTATGCTTGGAAAGTAGTTCGCAGCAATTACTTTGACTGTCACTCAAACACCTTTGATAACAGTCCGGGTAAAGTAGTTAAAATGCCACGTACTCGTGTGAATGACGATGATACACAAACTTGTTCTC
GTGGTTTGCATGTGTGTTCTAAATCTTATATTCGTCACTTTGGTAGTTCAACCAGCCGAGTTGTAAAAGTTAAAGTTCATCCTCGCGATGTAGTATCAATTCCGATTGATTACAACGATGCTAAAATGCGTACCTGCCAATATGAAGTAG
TTGAAGACGTTACTGAACAATTTAAATAAGGCTTCGGCCCTTAACTAAGGAAAATTATGTTAGCTTATCAAGCACGAGTAAAAGAAGAATACGATCAATTAATGCTCAAAATTAATGCACTGAGCAAATTTTTAGAAAGCGCAAAGTTTC
TAACAGTTAGCGCAGTTGAGCAAGAACTGCTACTTTCGCAGTTCATCTCAATGAAATCTTATGCTGAATGTCTAGAGAAAAGAATTGCACAATTCAAGTAAAATAAAGGGCTTCGGCCTTTTGTTTTAAGAGAAATTATGATTTATATGA
ATATCGGTGATTCCGACATAAAGGAAAGTTTAAATGCAGAAAACGAATCCTGGGTTACAGAGACTATTTCAGATTCCGACATTTACCCTATCGAACAGTGACTTGACCTGTGAAATGAAGGTCAAAATTGCTGATACTGCAAGATACTCT
TTAAAACAAAACCCAAATCAAGATAAGGCAGAAGTTATCGAAAGATGCCGTATCGCTGTGTACGCAGAGTTTTTTGTGGCAGATTGGTTGAGTGGATATGTCAATAAAGGCCAAGAGGATGTCGATGATCCGTACACATATGCATGGGAT
GTATTGGCTCATCCAAAATACTGCGGGCTTCGTGTAGAAGTTAAGACACATCAAACTGACTCACGTTGGATTTCGGTAACAACGGGATGCAGCGGAGAGTATCCATATGGTTCTGGAATAAATCTAGGGCCCATTCTAAATCATCAAGTC
GCTGACTGTATAATTATATTCAACACTAAAGAAATTCATCCAGGTGTCATCCAGTACACTCCGAAGTTCATCGGTGACAGAGAAGACCTTCGTAAAGTTGTAAGAAAAAGCAACTACAACGGATGGTATCTTTCCATTTAAAAAATTTCA
```

```
ATGAAATTTGTAAAAATTGATTCTTCTAGCGTTGATATGAAAAAATATAAATTGCAGAACAATGTTCGTCGTTCTATTAAATCCTCTTCAATGAACTATGCGAATGTCGCTATTATGACAGACGCAGATCACGATGGATAGCCTTCGGGC
TATCTATAGAAATACCTCATAATTAAGAGATTATTGGATTAGGTTCTATTTATCCTTCTCTGCTCGGATTTTTAGTAATTGGCCAGAATTGTTTGAGCAAGGACGAATTCGCTTTGTCAAAACTCCTGTAATCATCGCTCAGGTCGGTA  g60
AAAAACAAGAATGGTTTTATACAGTCGCTGAATATGAGAGTGCCAAAGATGCTCTACCTAAACATAGCATCCGTTATATTAAGGGACTTGGCTCTTTGGAAAAATCTGAATATCGTGAGATGATTCAAAACCCAGTATATGATGTTGTTA
AACTTCCTGAGAACTGGAAAGAGCTTTTTGAAATGCTCATGGGAGATAATGCTGACCTTCGTAAAGAATGGATGAGCCAGTAGTTTACTTTACCACAAGGATGTGGTATAATTAATTGGGCAAATGAGGATATTGAAATGAAATCATATA
AAGTAAATTTAGAACTTTTTGATAAAGCAGTTCATCGAGAATATAGAATCATTCAACGCTTTTTCGATATGGGAGAAGCCGAAGAATTTAAAACCCGCTTTAAAGATATTAGAGATAAAATTCAATCCGACACCGCAACTAAAGATGAAC
TACTAGAAGTTGCTGAAGTTATTAAGCGTAATATGAATTAATGAGGAAATTATGATTATCACCACTGAAAAAGAAACAATTCTTGGTAATGGTTCTAAATCAAAAGCATTTAGCATCACAGCATCTCCTAAAGTATTTAAAATTCTGTCA
TCTGATTTGTATACAAACAAGATTCGCGCAGTAGTCCGTGAATTGATTACTAACATGATTGATGCTCATGCACTTAATGGAAATCCTGAAAAATTTATCATACAAGTTCCTGGACGTTTAGACCCACGATTTGTTTGTCGAGATTTTGGT
CCGGGTATGAGTGATTTTGATATTCAAGGTGATGATAATTCTCCTGGGTTGTATAATTCATACTTCAGTTCATCTAAAGCTGAATCTAATGACTTTATTGGCGGATTTGGTTTAGGTTCTAAATCTCCGTTTAGTTATACTGATACGTTT
AGTATTACTTCGTATCATAAAGGTGAAATTCGTGGTTATGTAGCTTACATGGATGGTGATGGTCCACAGATTAAACCTACATTCGTAAAAGAAATGGGTCCAGATGATAAAACTGGTATTGAAATCGTAGTTCCAGTTGAAGAAAAAGAC  rIIA
TTTAGAAACTTTGCTTATGAAGTTTCTTATATCATGCGACCGTTCAAAGATTTGGCTATCATTAATGGTCTTGACCGCGAAATTGATTATTTTCCGGATTTTGATGACTATTACGGTGTAAATCCAGAAAGATACTGGCCTGATCGTGGT
GGATTATATGCTATCTACGGTGGTATTGTTTATCCTATCGATGGTGTTATTAGAGACCGTAACTGGCTAAGCATTCGCAATGAAGTGAATTACATTAAGTTTCCAATGGGTTCACTTGATATTGCTCCATCTCGCGAGGCTCTTTCACTG
GATGATCGCACTCGTAAAAATATTATTGAACGAGTTAAAGAACTCAGTGAGAAAGCATTTAATGAAGATGTAAAACGATTTAAAGAATCTACATCTCCTCGTCACACATATCGTGAATTGATGAAGATGGGGTATTCTGCTCGAGATTAT
ATGATTAGTAATTCAGTCAAATTCACGACTAAAAATCTGTCATATAAAAAGATGCAGAGCATGTTTGAACCTGACAGTAAGTTATGCAACGCGGGAGTTGTGTATGAAGTAAATCTTGACCCTCGACTGAAGCGCATTAAGCAAAGTCAT
GAAACTTCAGCCGTTGCATCAAGTTATCGTCTGTTTGGTATTAATACAACAAAAATTAATATCGTTATTGATAATATTAAAAATCGTGTTAATATTGTCCGTGGATTAGCACGTGCGTTAGATGATAGTGAATTTAATAACACTTTGAAT
ATTCATCATAACGAACGTCTTCTGTTTATTAATCCAGAAGTAGAATCGCAGATTGATTTGCTTCCTGATATTATGGCGATGTTTGAAAGTGATGAAGTTAACATTCATTATTTGTCAGAAATTGAAGCTTTAGTAAAAAGTTATATTCCA
AAGGTAGTTAAAAGTAAAGCTCCTCGTCCTAAAGCTGCTACAGCGTTTAAGTTTGAAATTAAAGACGGGCGCTGGGAAAAGAGGAATTATTTACGCCTCACATCAGAAGCAGATGAAATTACTGGTTATGTAGCGTATATGCATCGTTCT
GATATTTTCTCTATGGATGGTACTACATCTCTTTGTCATCCATCTATGAATATTTTGATTCGTATGGCTAATCTTATTGGCATTAATGAATTTTATGTTATTCGTCCGCTTTTACAGAAAAAGGTAAAAGAACTCGGTCAGTGCCAATGT
ATTTTTGAAGCTTTGCGTGATTTATATGTAGATGCTTTTGATGATGTAGATTATGATAAGTATGTAGGTTATTCAAGTTCAGCTAAACGATATATTGATAAAATTATCAAGTATCCTGAGTTAGATTTTATGATGAAGTACTTCAGTATA
GATGAAGTTTCTGAAGAATATACACGACTCGCTAATATGGTTAGTTCATTACAGGGTGTATATTTTAATGGTGGAAAAGATACCATCGGTCATGACATTTGGACAGTAACTAATCTTTTTGATGTATTATCAAATAATGCTTCAAAAAAC
AGTGATAAAATGGTTGCTGAGTTTACCAAGAAATTCCGTATTGTTTCCGACTTCATCGGTTATCGCAACTCTTTAAGTGATGATGAAGTTTCCCAAATCGCTAAAACTATGAAGGCCCTTGCGGCCTAATAAGGAAAATTATGTACAATA
TTAAATGCCTGACCAAAAACGAACAAGCTGAAATTGTTAAACTGTATTCAAGTGGTAATTACACCCAACAGGAATTGGCTGATTGGCAAGGTGTATCGGTTGACACAATCCGTCGTGTTTTGAAAAATGCTGAAGAAGCTAAACGCCCTA
AAGTTACTATTAGCGGTGATATTACAGTTAAAGTTAATAGCGATGCAGTTATTGCTCCAGTTGCTAAATCTGACATTATTTGGAATGCATCTAAAAAATTCATTTCAATTACTGTTGATGGCGTAACTTATAACGCAACTCCTAATACTC
ATTCAAACTTCCAGGAAATTCTTAATCTGCTTGTAGCGGATAAGTTGGAAGAAGCGGCACAAAAAATTAATGTTCGTCGTGCTGTTGAAAAATATATTTCCGGCGATGTTCGAATTGAAGGTGGAAGCTTGTTCTATCAAAATATTGAAT  rIIB
TGCGGTCTGGTTTGGTTGATCGTATTCTTGACTCGATGGAAAAAGGCGAAAACTTTGAATTTTATTTTCCGTTCTTGGAAAATCTGTTGGAAAACCCAAGCCAAAAAGCGGTATCTCGACTCTTTGATTTCTTGGTAGCAAACGATATTG
AAATTACAGAAGATGGTTACTTCTATGCTTGGAAAGTAGTTCGCAGCAATTACTTTGACTGTCACTCAAACACCTTTGATAACAGTCCGGGTAAAGTAGTTAAAATGCCACGTACTCGTGTGAATGACGATGATACACAAACTTGTTCTC
GTGGTTTGCATGTGTGTTCTAAATCTTATATTCGTCACTTTGGTAGTTCAACCAGCCGAGTTGTAAAAGTTAAAGTTCATCCTCGCGATGTAGTATCAATTCCGATTGATTACAACGATGCTAAAATGCGTACCTGCCAATATGAAGTAG
TTGAAGACGTTACTGAACAATTTAAATAAGGCTTCGGCCCTTAACTAAGGAAAATTATGTTAGCTTATCAAGCACGAGTAAAAGAAGAATACGATCAATTAATGCTCAAAATTAATGCACTGAGCAAATTTTTAGAAAGCGCAAAGTTTC
TAACAGTTAGCGCAGTTGAGCAAGAACTGCTACTTTCGCAGTTCATCTCAATGAAATCTTATGCTGAATGTCTAGAGAAAAGAATTGCACAATTCAAGTAAAATAAAGGGCTTCGGCCTTTTGTTTTAAGAGAAATTATGATTTATATGA
ATATCGGTGATTCCGACATAAAGGAAAGTTTAAATGCAGAAAACGAATCCTGGGTTACAGAGACTATTTCAGATTCCGACATTTACCCTATCGAACAGTGACTTGACCTGTGAAATGAAGGTCAAAATTGCTGATACTGCAAGATACTCT
TTAAAACAAAACCCAAATCAAGATAAGGCAGAAGTTATCGAAAGATGCCGTATCGCTGTGTACGCAGAGTTTTTTGTGGCAGATTGGTTGAGTGGATATGTCAATAAAGGCCAAGAGGATGTCGATGATCCGTACACATATGCATGGGAT  denV
GTATTGGCTCATCCAAAATACTGCGGGCTTCGTGTAGAAGTTAAGACACATCAAACTGACTCACGTTGGATTTCGGTAACAACGGGATGCAGCGGAGAGTATCCATATGGTTCTGGAATAAATCTAGGGCCCATTCTAAATCATCAAGTC
GCTGACTGTATAATTATATTCAACACTAAAGAAATTCATCCAGGTGTCATCCAGTACACTCCGAAGTTCATCGGTGACAGAGAAGACCTTCGTAAAGTTGTAAGAAAAAGCAACTACAACGGATGGTATCTTTCCATTTAAAAAATTTCA
```

T4 rIIB (939 base pairs)

ATG TAC AAT ATT AAA TGC CTG ACC AAA AAC GAA CAA GCT GAA ATT GTT AAA CTG TAT TCA AGT GGT AAT TAC ACC CAA CAG GAA TTG GCT

GAT TGG CAA GGT GTA TCG GTT GAC ACA ATC CGT CGT GTT TTG AAA AAT GCT GAA GAA GCT AAA CGC CCT AAA GTT ACT ATT AGC GGT GAT

ATT ACA GTT AAA GTT AAT AGC GAT GCA GTT ATT GCT CCA GTT GCT AAA TCT GAC ATT ATT TGG AAT GCA TCT AAA AAA TTC ATT TCA ATT

ACT GTT GAT GGC GTA ACT TAT AAC GCA ACT CCT AAT ACT CAT TCA AAC TTC CAG GAA ATT CTT AAT CTG CTT GTA GCG GAT AAG TTG GAA

GAA GCG GCA CAA AAA ATT AAT GTT CGT CGT GCT GTT GAA AAA TAT ATT TCC GGC GAT GTT CGA ATT GAA GGT GGA AGC TTG TTC TAT CAA

AAT ATT GAA TTG CGG TCT GGT TTG GTT GAT CGT ATT CTT GAC TCG ATG GAA AAA GGC GAA AAC TTT GAA TTT TAT TTT CCG TTC TTG GAA

AAT CTG TTG GAA AAC CCA AGC CAA AAA GCG GTA TCT CGA CTC TTT GAT TTC TTG GTA GCA AAC GAT ATT GAA ATT ACA GAA GAT GGT TAC

TTC TAT GCT TGG AAA GTA GTT CGC AGC AAT TAC TTT GAC TGT CAC TCA AAC ACC TTT GAT AAC AGT CCG GGT AAA GTA GTT AAA ATG CCA

CGT ACT CGT GTG AAT GAC GAT GAT ACA CAA ACT TGT TCT CGT GGT TTG CAT GTG TGT TCT AAA TCT TAT ATT CGT CAC TTT GGT AGT TCA

ACC AGC CGA GTT GTA AAA GTT AAA GTT CAT CCT CGC GAT GTA GTA TCA ATT CCG ATT GAT TAC AAC GAT GCT AAA ATG CGT ACC TGC CAA

TAT GAA GTA GTT GAA GAC GTT ACT GAA CAA TTT AAA **TAA**

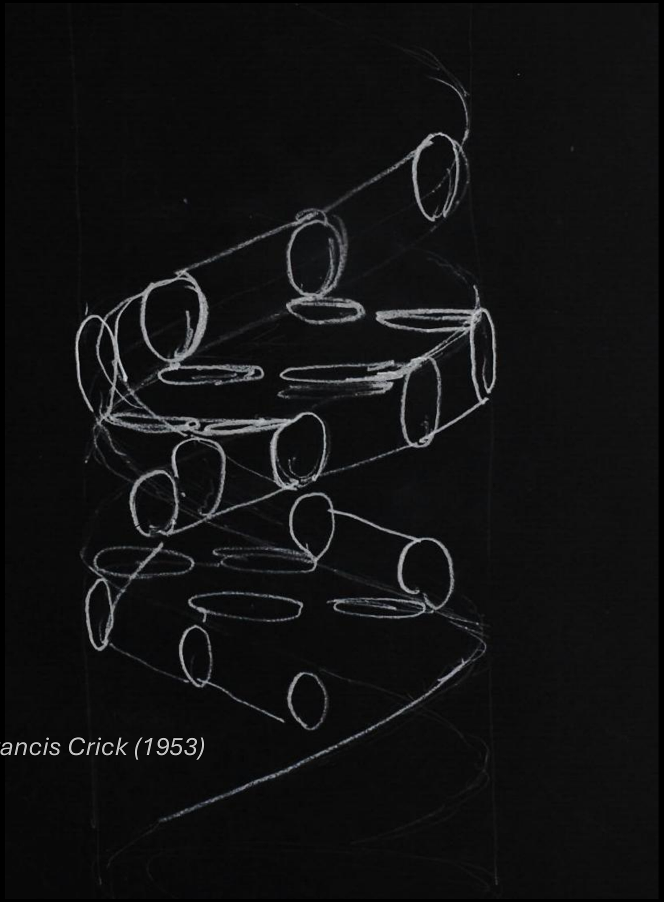
T4 rIIB (312 amino acids)

```
ATG TAC AAT ATT AAA TGC CTG ACC AAA AAC GAA CAA GCT GAA ATT GTT AAA CTG TAT TCA AGT GGT AAT TAC ACC CAA CAG GAA TTG GCT
Met Tyr Asn Ile Lys Cys Leu Thr Lys Asn Glu Gln Ala Glu Ile Val Lys Leu Tyr Ser Ser Gly Asn Tyr Thr Gln Gln Glu Leu Ala

GAT TGG CAA GGT GTA TCG GTT GAC ACA ATC CGT CGT GTT TTG AAA AAT GCT GAA GAA GCT AAA CGC CCT AAA GTT ACT ATT AGC GGT GAT
Asp Trp Gln Gly Val Ser Val Asp Thr Ile Arg Arg Val Leu Lys Asn Ala Glu Glu Ala Lys Arg Pro Lys Val Thr Ile Ser Gly Asp

ATT ACA GTT AAA GTT AAT AGC GAT GCA GTT ATT GCT CCA GTT GCT AAA TCT GAC ATT ATT TGG AAT GCA TCT AAA AAA TTC ATT TCA ATT
Ile Thr Val Lys Val Asn Ser Asp Ala Val Ile Ala Pro Val Ala Lys Ser Asp Ile Ile Trp Asn Ala Ser Lys Lys Phe Ile Ser Ile

ACT GTT GAT GGC GTA ACT TAT AAC GCA ACT CCT AAT ACT CAT TCA AAC TTC CAG GAA ATT CTT AAT CTG CTT GTA GCG GAT AAG TTG GAA
Thr Val Asp Gly Val Thr Tyr Asn Ala Thr Pro Asn Thr His Ser Asn Phe Gln Glu Ile Leu Asn Leu Leu Val Ala Asp Lys Leu Glu

GAA GCG GCA CAA AAA ATT AAT GTT CGT CGT GCT GTT GAA AAA TAT ATT TCC GGC GAT GTT CGA ATT GAA GGT GGA AGC TTG TTC TAT CAA
Glu Ala Ala Gln Lys Ile Asn Val Arg Arg Ala Val Glu Lys Tyr Ile Ser Gly Asp Val Arg Ile Glu Gly Gly Ser Leu Phe Tyr Gln

AAT ATT GAA TTG CGG TCT GGT TTG GTT GAT CGT ATT CTT GAC TCG ATG GAA AAA GGC GAA AAC TTT GAA TTT TAT TTT CCG TTC TTG GAA
Asn Ile Glu Leu Arg Ser Gly Leu Val Asp Arg Ile Leu Asp Ser Met Glu Lys Gly Glu Asn Phe Glu Phe Tyr Phe Pro Phe Leu Glu

AAT CTG TTG GAA AAC CCA AGC CAA AAA GCG GTA TCT CGA CTC TTT GAT TTC TTG GTA GCA AAC GAT ATT GAA ATT ACA GAA GAT GGT TAC
Asn Leu Leu Glu Asn Pro Ser Gln Lys Ala Val Ser Arg Leu Phe Asp Phe Leu Val Ala Asn Asp Ile Glu Ile Thr Glu Asp Gly Tyr

TTC TAT GCT TGG AAA GTA GTT CGC AGC AAT TAC TTT GAC TGT CAC TCA AAC ACC TTT GAT AAC AGT CCG GGT AAA GTA GTT AAA ATG CCA
Phe Tyr Ala Trp Lys Val Val Arg Ser Asn Tyr Phe Asp Cys His Ser Asn Thr Phe Asp Asn Ser Pro Gly Lys Val Val Lys Met Pro

CGT ACT CGT GTG AAT GAC GAT GAT ACA CAA ACT TGT TCT CGT GGT TTG CAT GTG TGT TCT AAA TCT TAT ATT CGT CAC TTT GGT AGT TCA
Arg Thr Arg Val Asn Asp Asp Asp Thr Gln Thr Cys Ser Arg Gly Leu His Val Cys Ser Lys Ser Tyr Ile Arg His Phe Gly Ser Ser

ACC AGC CGA GTT GTA AAA GTT AAA GTT CAT CCT CGC GAT GTA GTA TCA ATT CCG ATT GAT TAC AAC GAT GCT AAA ATG CGT ACC TGC CAA
Thr Ser Arg Val Val Lys Val Lys Val His Pro Arg Asp Val Val Ser Ile Pro Ile Asp Tyr Asn Asp Ala Lys Met Arg Thr Cys Gln

TAT GAA GTA GTT GAA GAC GTT ACT GAA CAA TTT AAA TAA
Tyr Glu Val Val Glu Asp Val Thr Glu Gln Phe Lys
```

T4 rIIB N-terminus (50 amino acids)

1 MYNIKCLTKNEQAEIVKLYSSGNYTQQELADWQGVSVDTIRRVLKNAEEA... 50

The genome is digital information

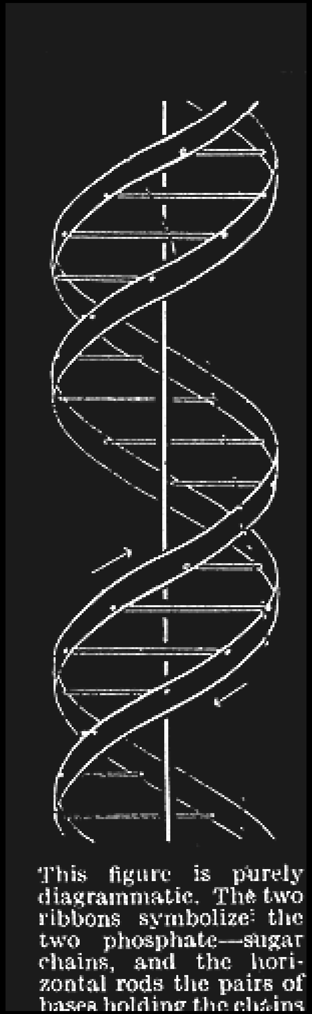


Francis Crick (1953)

1892: Friedrich Miescher postulates that genetic information may be encoded in a linear sequence of a few different chemical units...

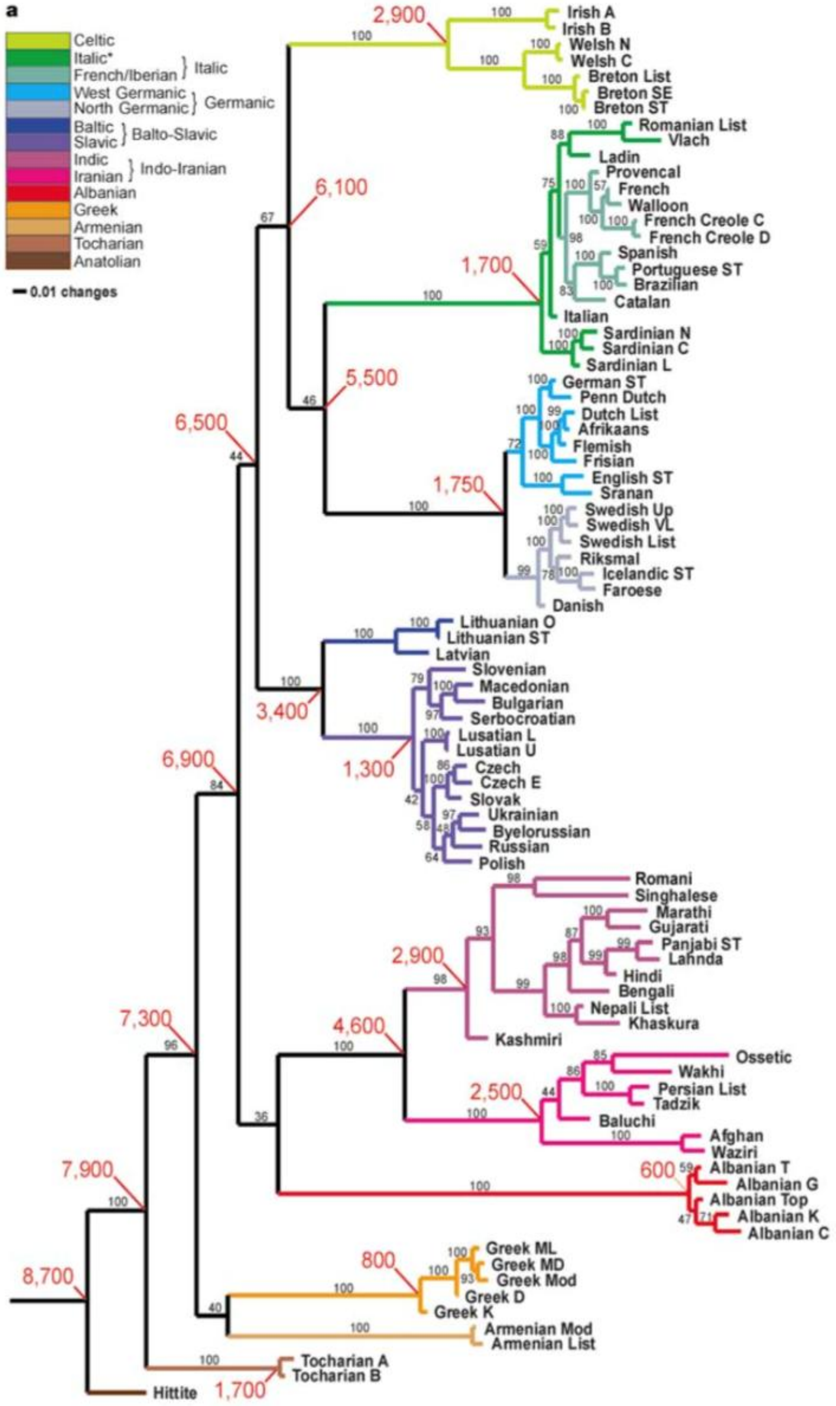
> *"...just as all the words and concepts in all languages can find expression in twenty-four to thirty letters of the alphabet."*

Friedrich Miescher to his uncle, private letter, 1892



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains

Odile Crick (1953)



English: give us this day our daily bread

Swedish: giv oss i dag vårt dagliga bröd

Dutch: geef ons heden ons dagelijks brood

German: gib uns heute unser tägliches brot

Polish: chleba naszego powszedniego daj nam dzisiaj

Gray and Atkinson, Nature 426:435 (2003)

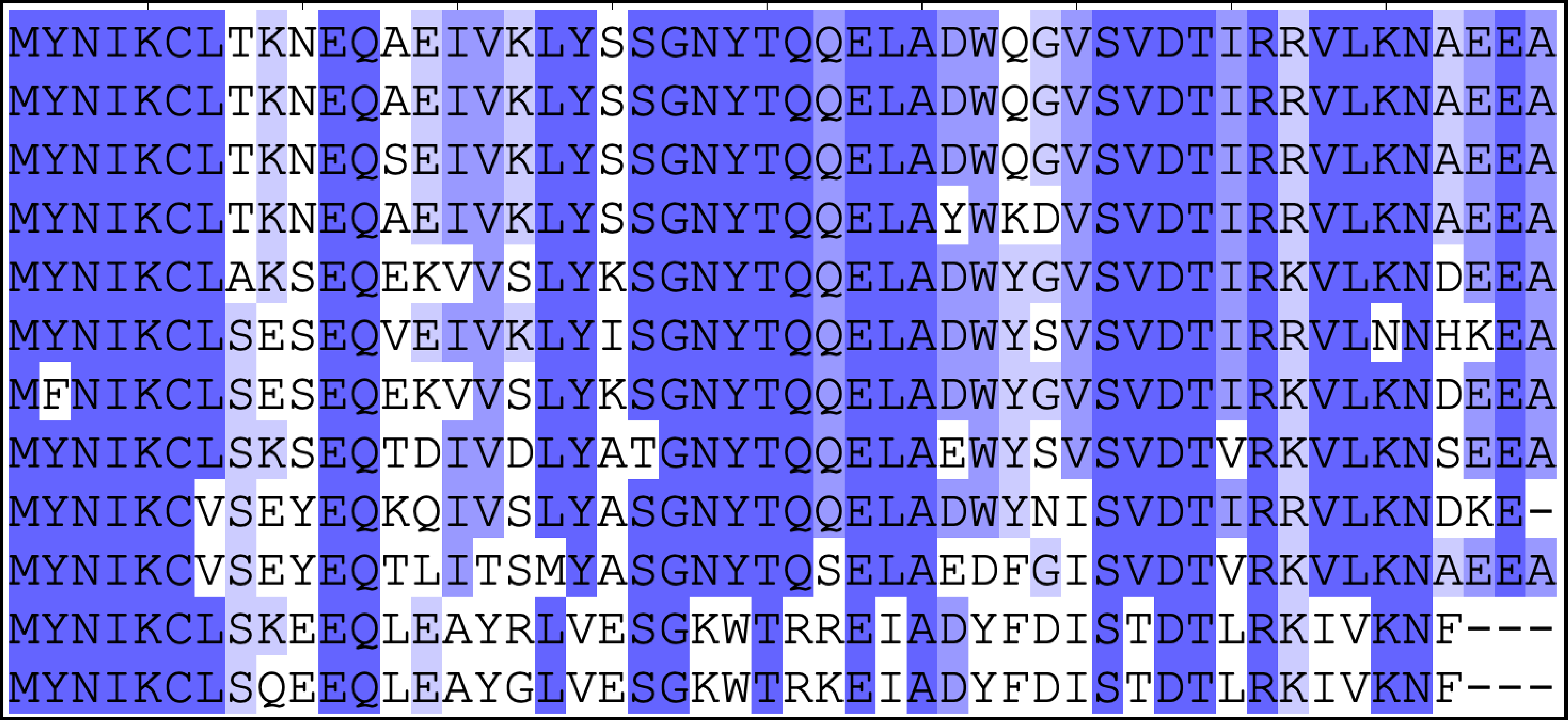
Guy Deutscher, *The Unfolding of Language* (2005)

Linear B
Mycenae, ca. 1500-1200 BC
deciphered by Michael Ventris, John Chadwick, and Alice Kober, 1953

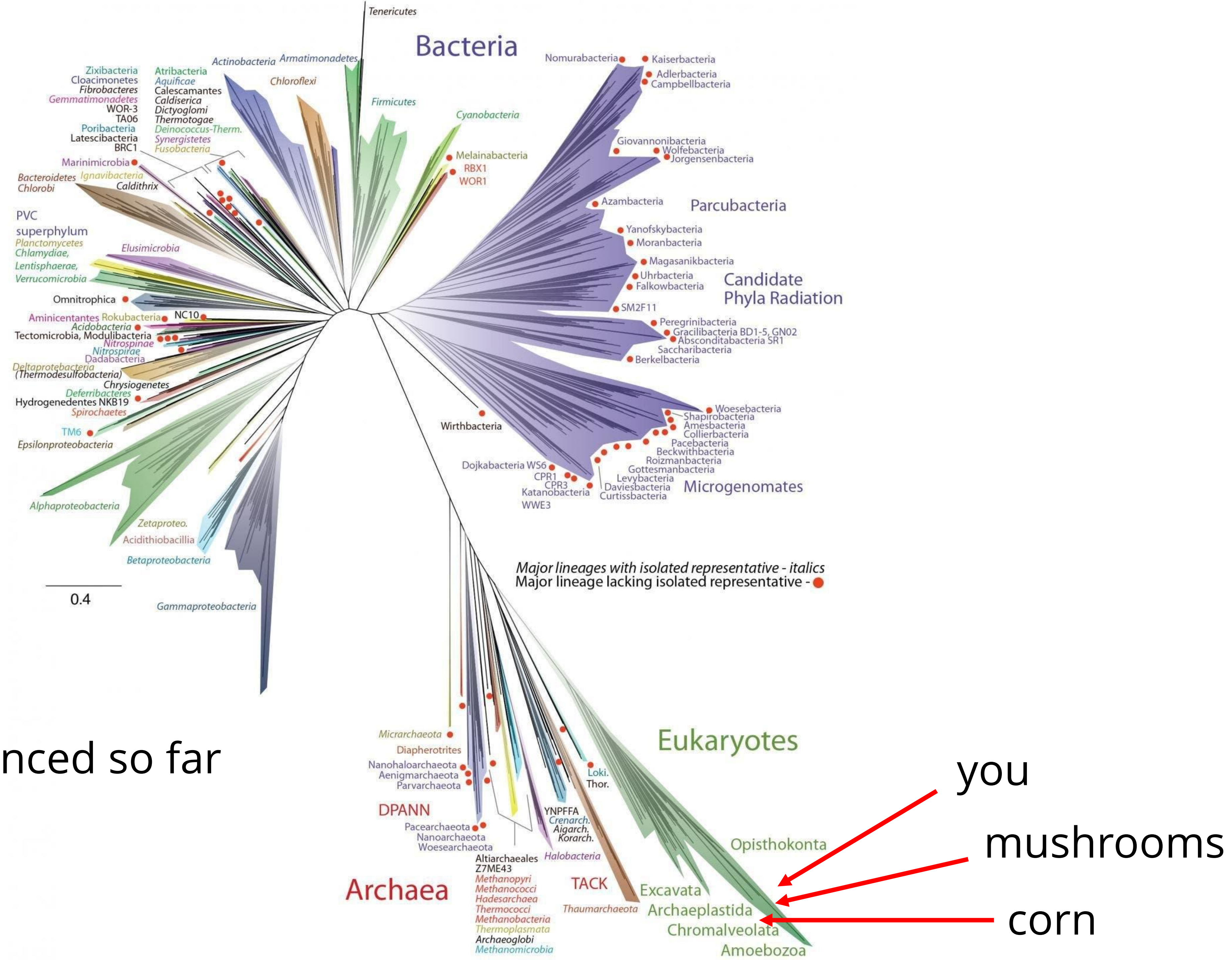
“Cryptography has contributed a new weapon to the student of unknown scripts.... the basic principle is the analysis and indexing of coded texts, so that underlying patterns and regularities can be discovered. *If a number of instances can be collected,* it may appear that a certain group of signs in the coded text has a particular function....”

John Chadwick,
The Decipherment of Linear B
Cambridge University Press, 1958

a multiple sequence alignment (MSA)
of homologs of the rIIB N-terminus from several other phage genomes



```
MYNIKCLTKNEQAEIVKLYSSGNYTQQELADWQGVSVDTIRRVLKNAEEA
MYNIKCLTKNEQAEIVKLYSSGNYTQQELADWQGVSVDTIRRVLKNAEEA
MYNIKCLTKNEQSEIVKLYSSGNYTQQELADWQGVSVDTIRRVLKNAEEA
MYNIKCLTKNEQAEIVKLYSSGNYTQQELAYWKDVSVDTIRRVLKNAEEA
MYNIKCLAKSEQEKVVSLYKSGNYTQQELADWYGVSVDTIRKVLKNDEEA
MYNIKCLSESEQVEIVKLYISGNYTQQELADWYSVSVDTIRRVLNNHKEA
MFNIKCLSESEQEKVVSLYKSGNYTQQELADWYGVSVDTIRKVLKNDEEA
MYNIKCLSKSEQTDIVDLYATGNYTQQELAEWYSVSVDTVRKVLKNSEEA
MYNIKCVSEYEQKQIVSLYASGNYTQQELADWYNISVDTIRRVLKNDKE-
MYNIKCVSEYEQTLITSMYASGNYTQSELAEDFGISVDTVRKVLKNAEEA
MYNIKCLSKEEQLEAYRLVESGKWTRREIADYFDISTDTLRKIVKNF---
MYNIKCLSQEEQLEAYGLVESGKWTRKEIADYFDISTDTLRKIVKNF---
```



you

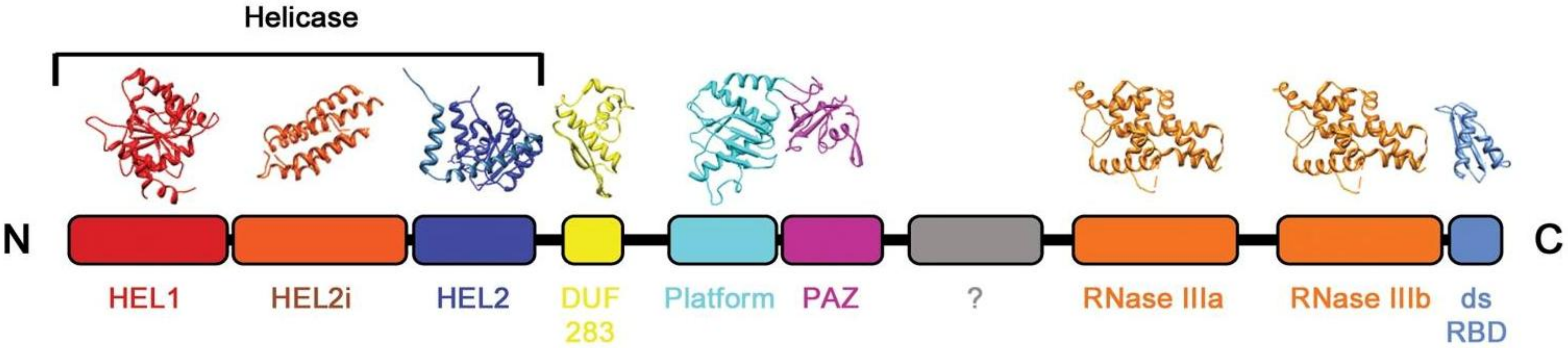
mushrooms

corn

~10^6 species sequenced so far

~10^7-10^8 to go

About 30,000 different conserved protein domains account for most protein sequence



Pick-Wei Lau et al (2012)
The molecular architecture of human Dicer

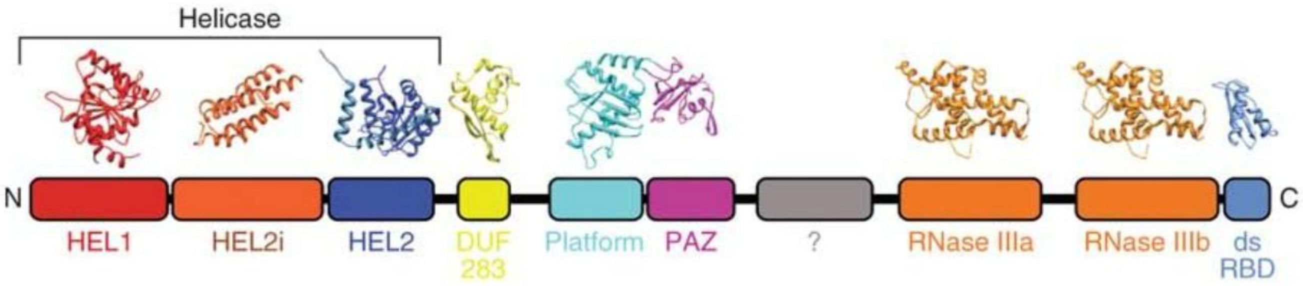
Algorithms and software tools for homology search and alignment

HMMER

protein homology search: profile HMMs

hmmer.org

55K lines source code
>40,000 downloads/yr



human Dicer | Lau *et al* (2012)

Pfam and Interpro

30,000 protein domain families

www.ebi.ac.uk/interpro/

Xfam Consortium and HMMER server team: Alex Bateman, Rob Finn
European Bioinformatics Institute, Cambridge UK

rIIA

: found in many ATP-driven motors: gyrases, DNA mismatch repair, others

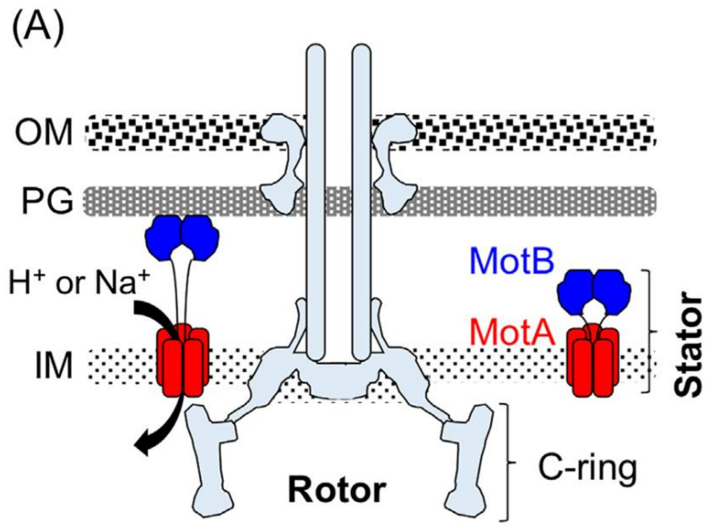
rIIB

: helix-turn-helix DNA binding proteins

rexA

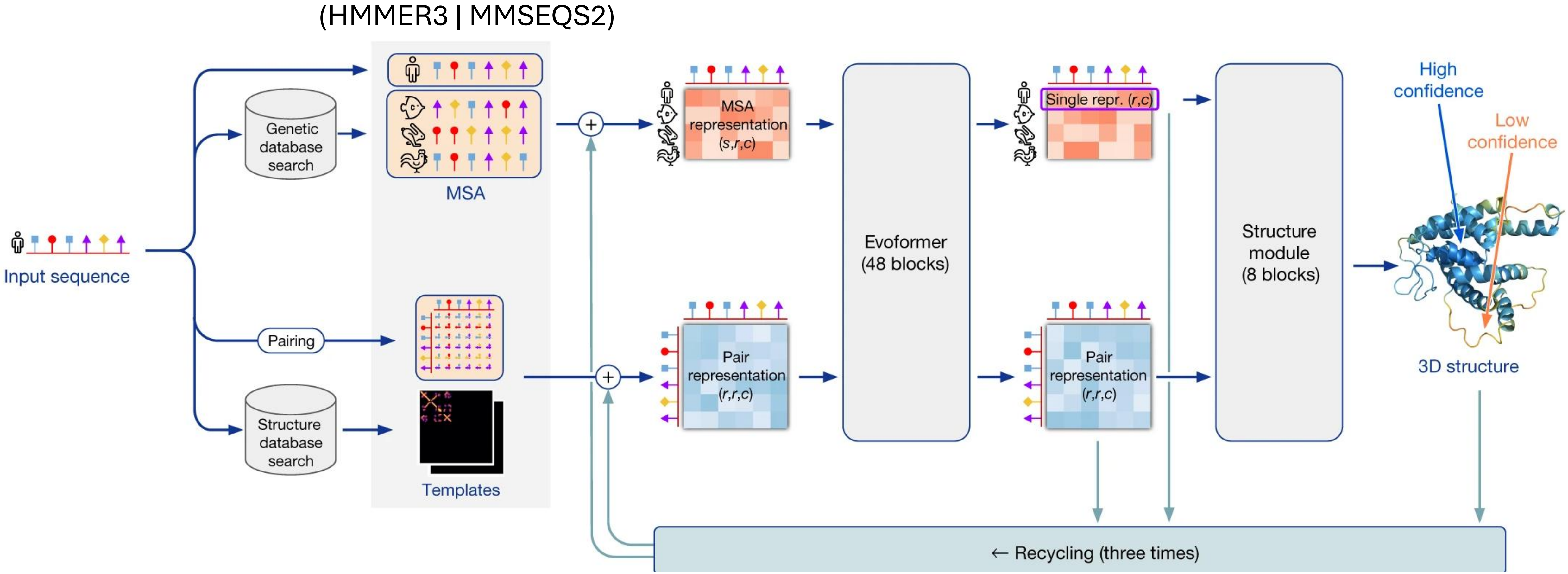
rexB

: motA is the proton-driven stator motor
of E. coli chemotaxis

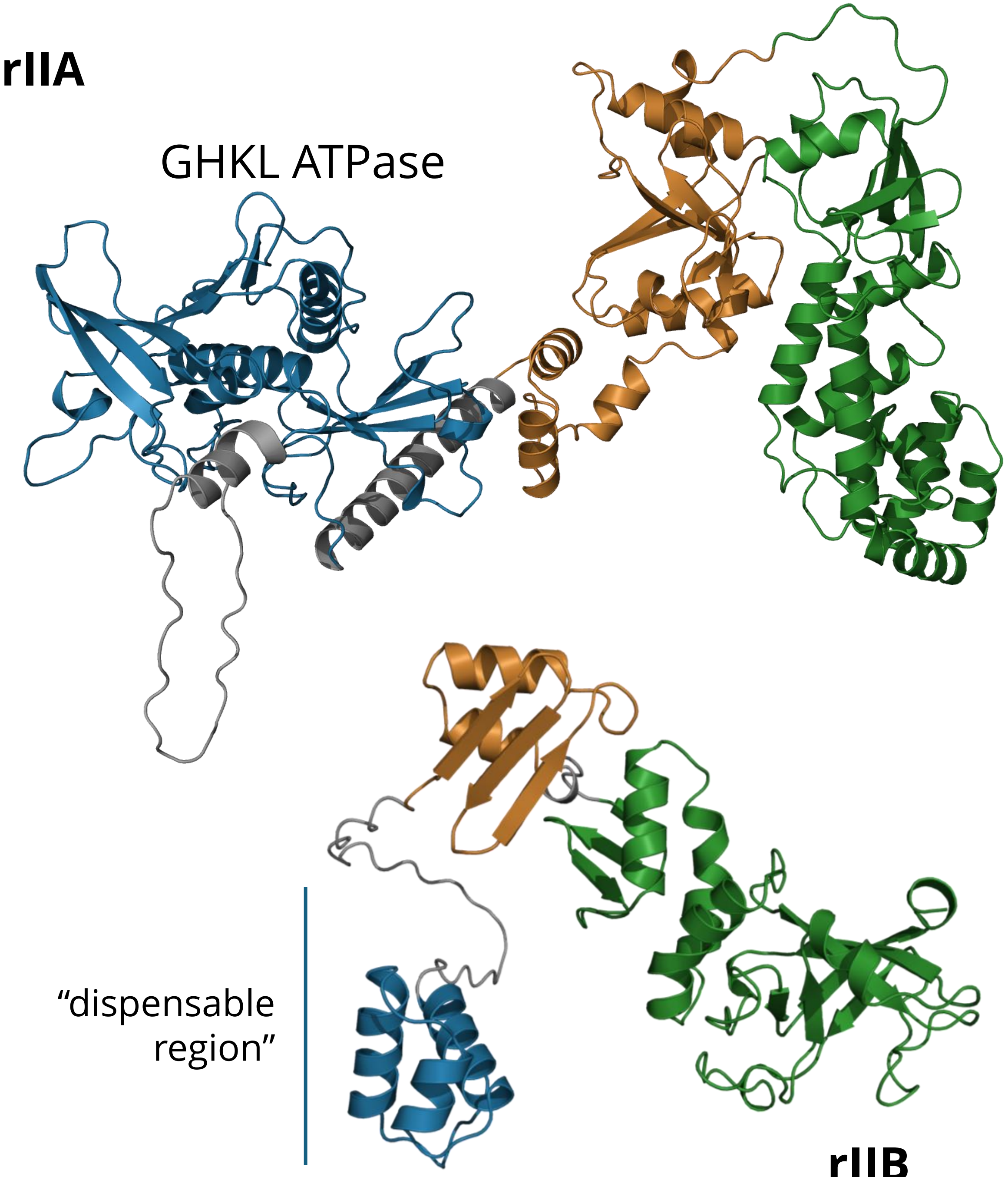


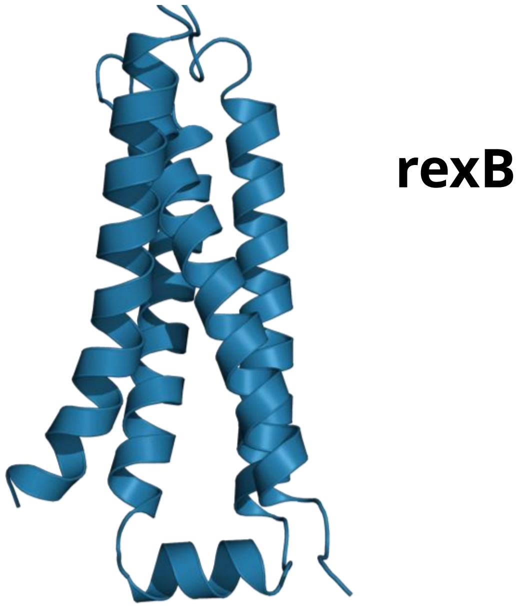
Nishikino et al, BBRC (2022)

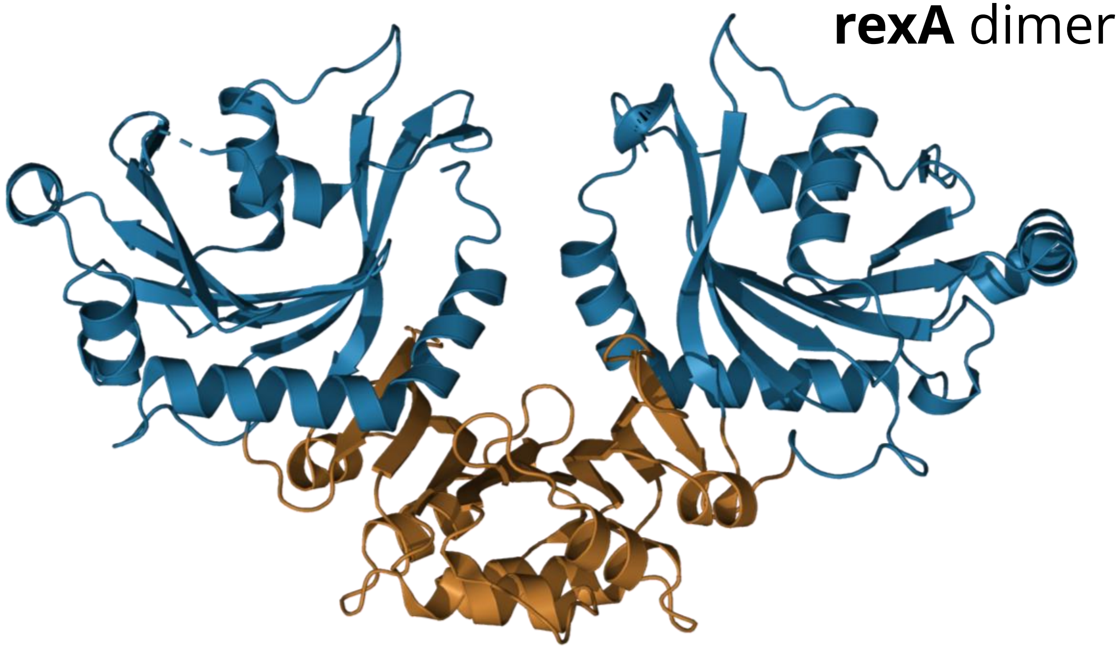
AlphaFold: from MSA to 3D protein structure (at last!)



Senior et al. Nature 2020
Jumper et al. Nature 2021
Abramson et al. Nature 2024

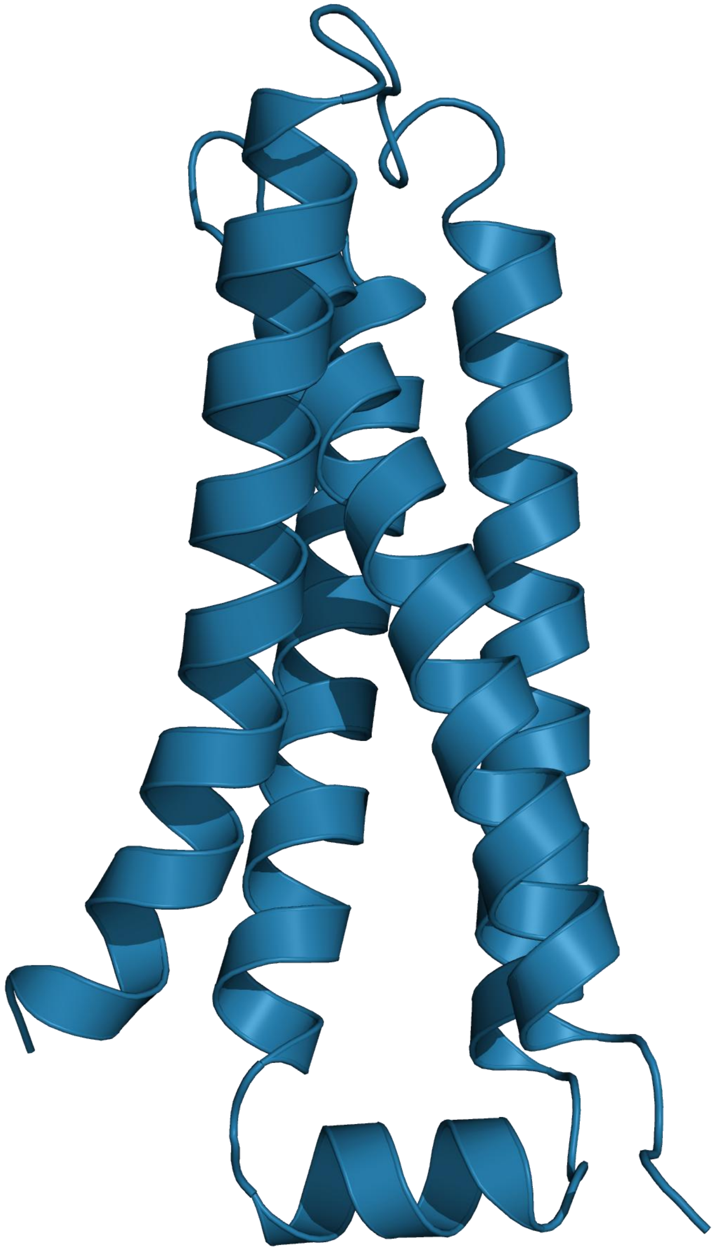






PDB:8TWQ, Adams et al NAR (2024)

rexB appears to be a four-pass transmembrane protein, just as Parma inferred



Parma et al.

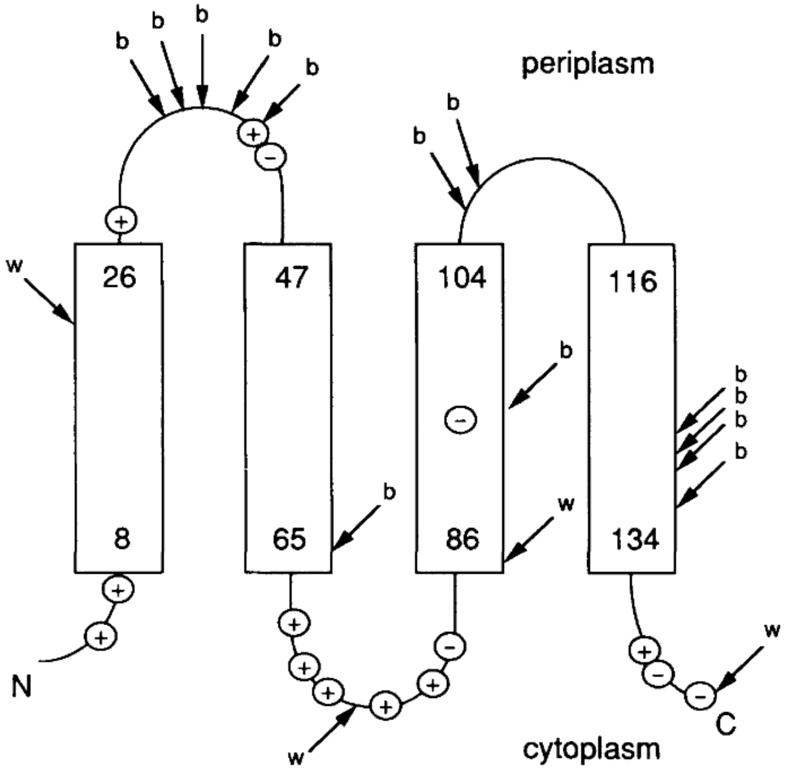
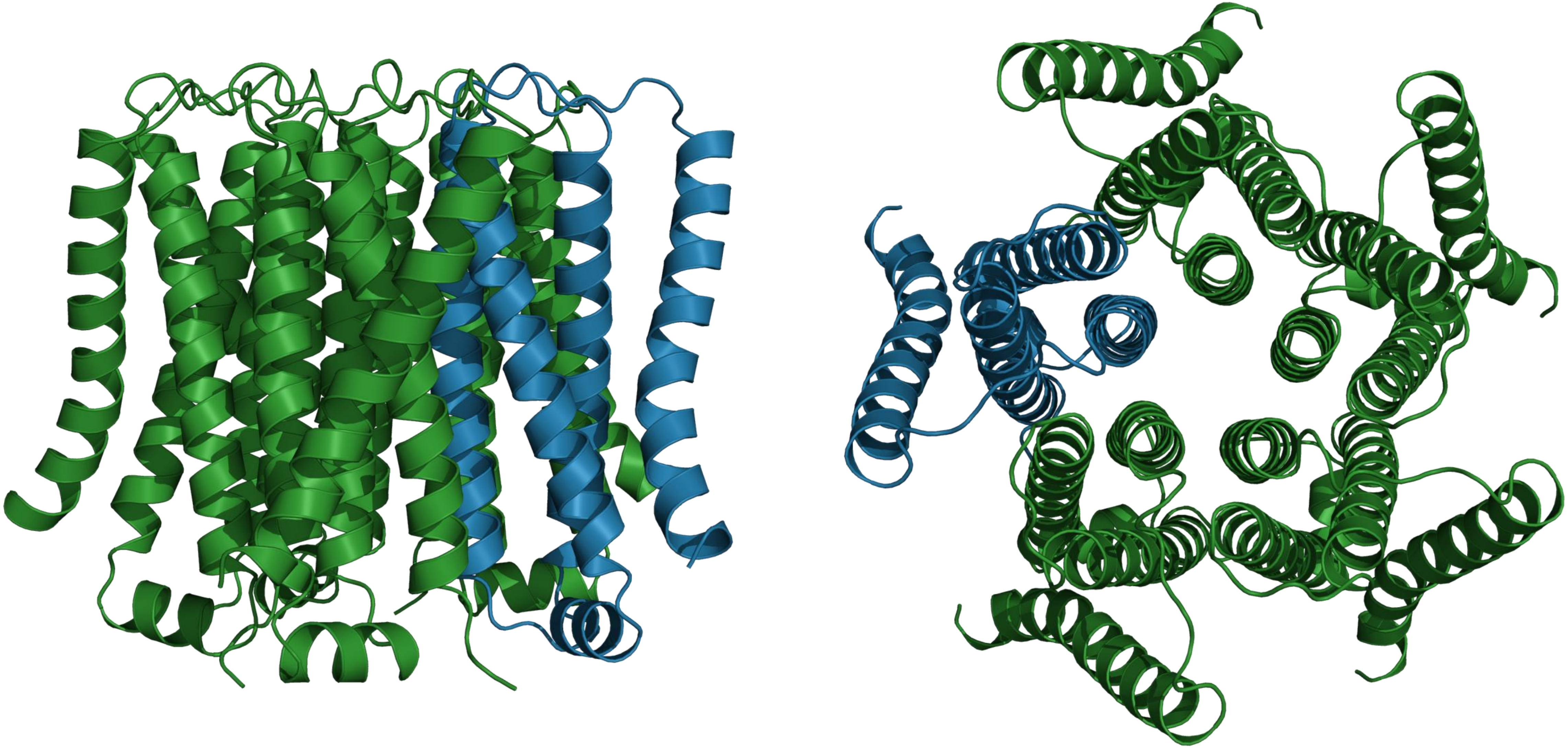


Figure 4. Predicted topology of the RexB protein. Transmembrane domains are shown as rectangles; cytoplasmic and periplasmic domains are lines. Junctions of in-frame RexB–PhoA′ fusions are marked with arrows; their color on X-phosphate plates is indicated: (b) Blue, active; (w) white, inactive. Charged amino acids are depicted by + and − signs. The amino and carboxyl termini are designated by N and C, respectively.

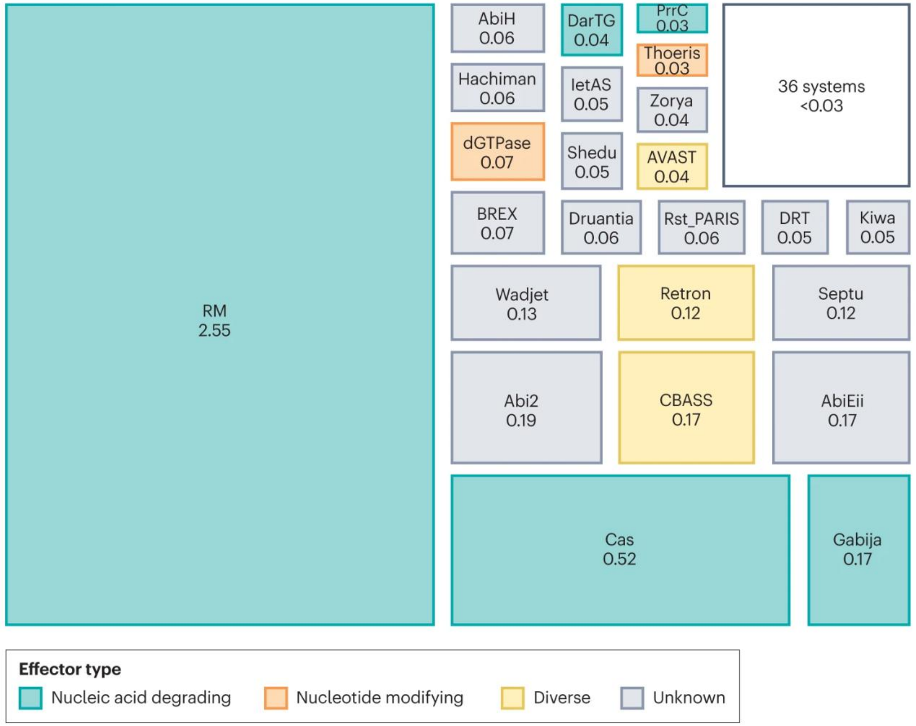
Parma et al. Genes Dev (1992)

if homologous to motA: rexB might make a pentamer, forming a proton channel

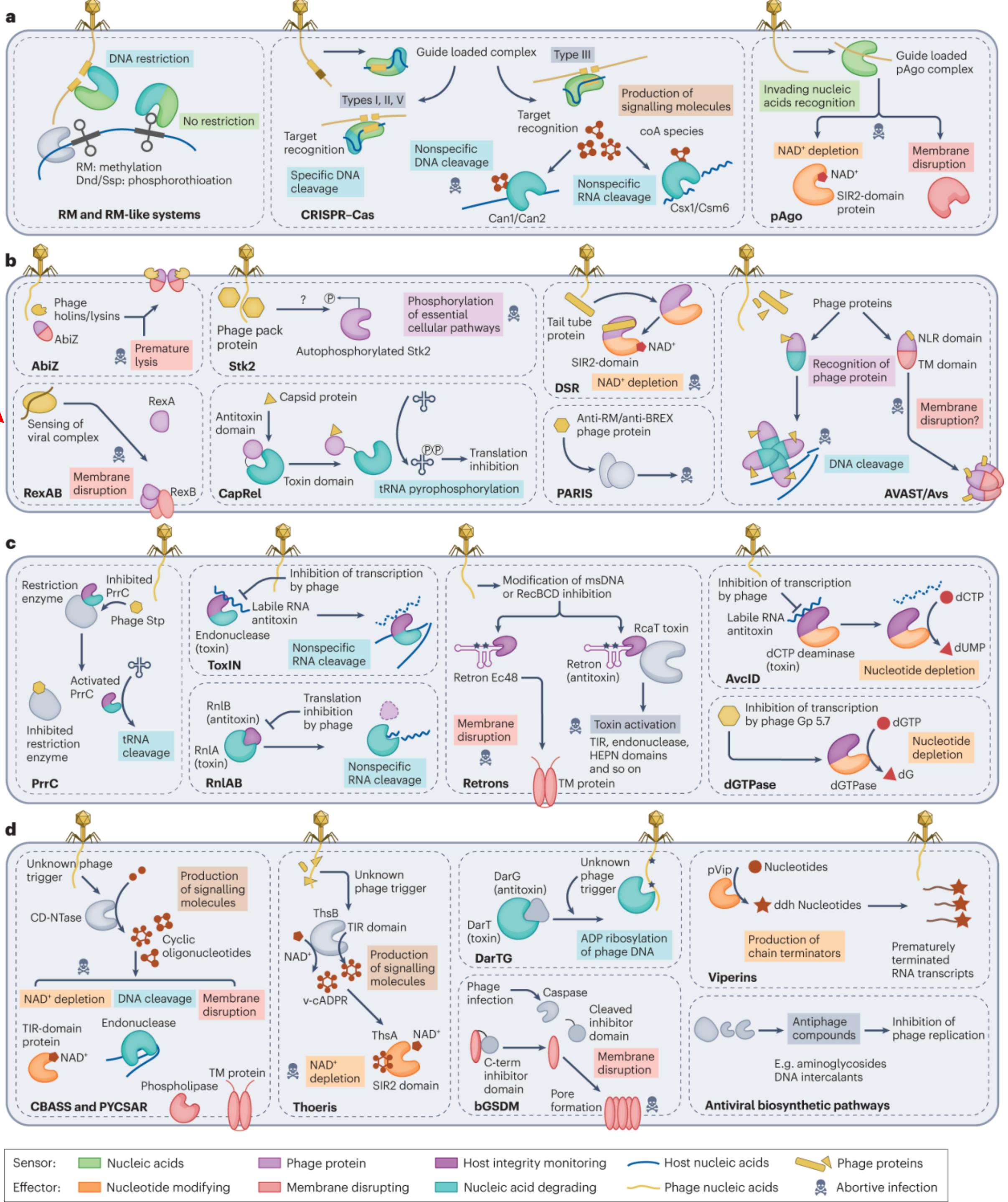


Antiphage defense systems (and counter-systems)

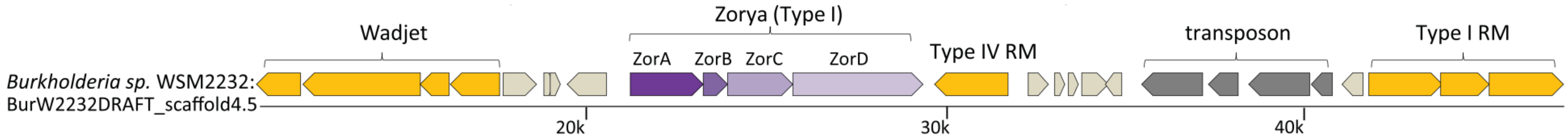
restriction enzymes,
CRISPR,
and many many more

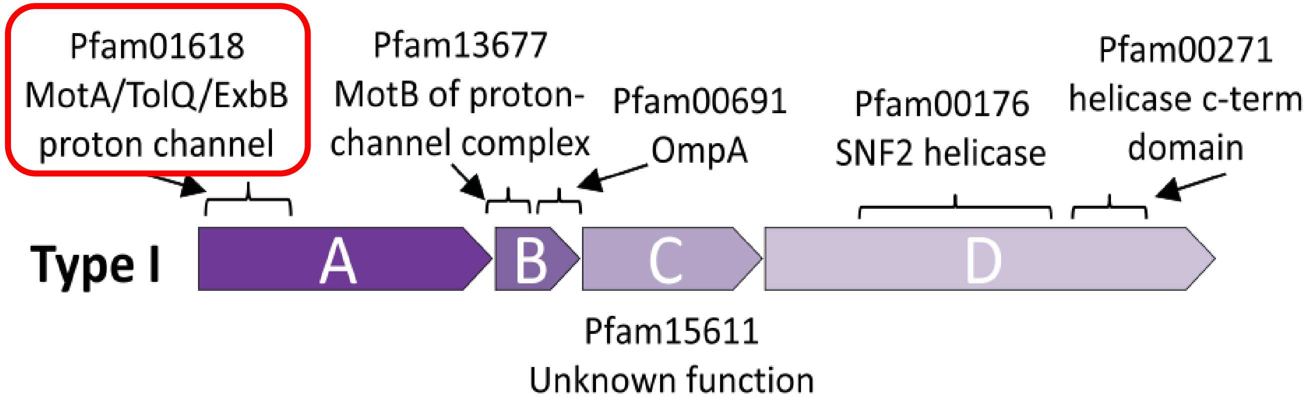


rexAB



Gerojon & Bernheim, Nat Rev Micro (2023)




Pfam01618
MotA/TolQ/ExbB
proton channel

Pfam13677
MotB of proton-
channel complex

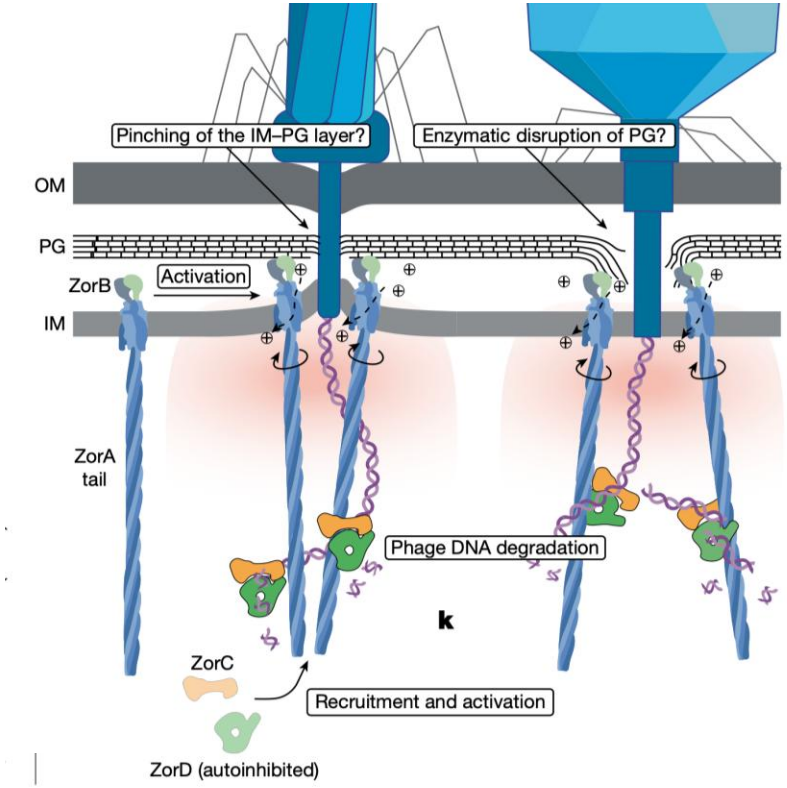
Pfam00691
OmpA

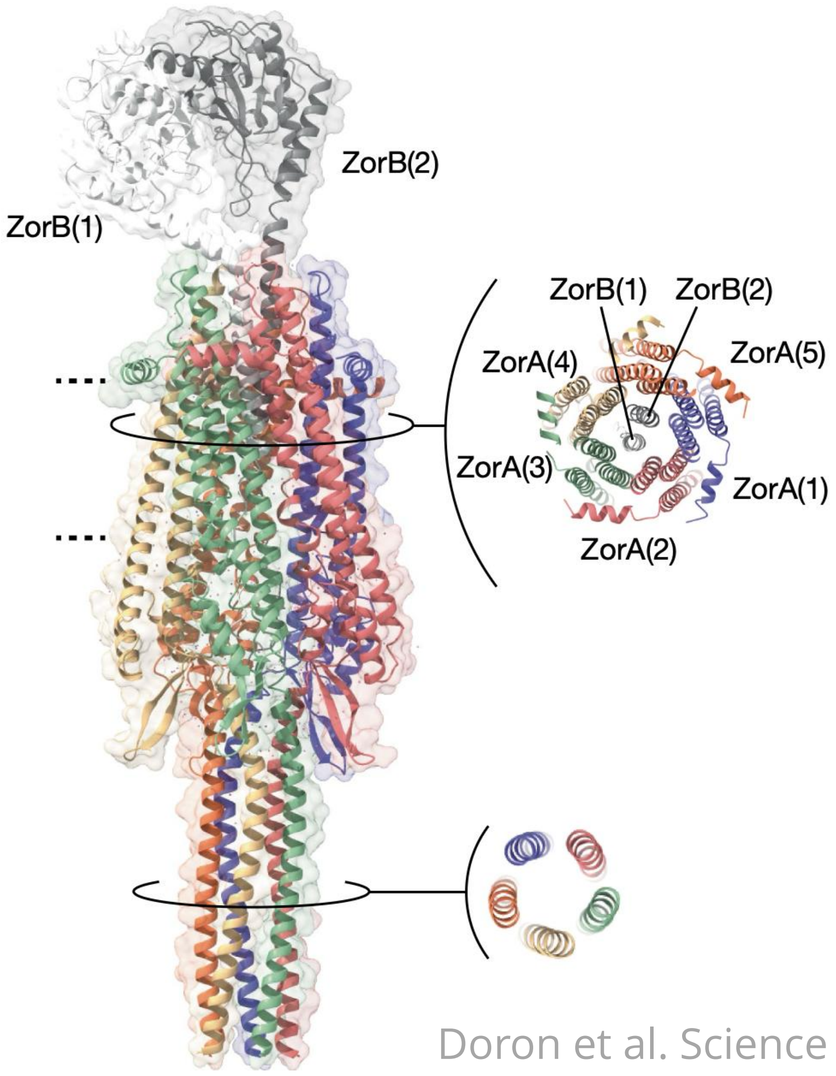
Pfam00176
SNF2 helicase

Pfam00271
helicase c-term
domain

Type I

Pfam15611
Unknown function



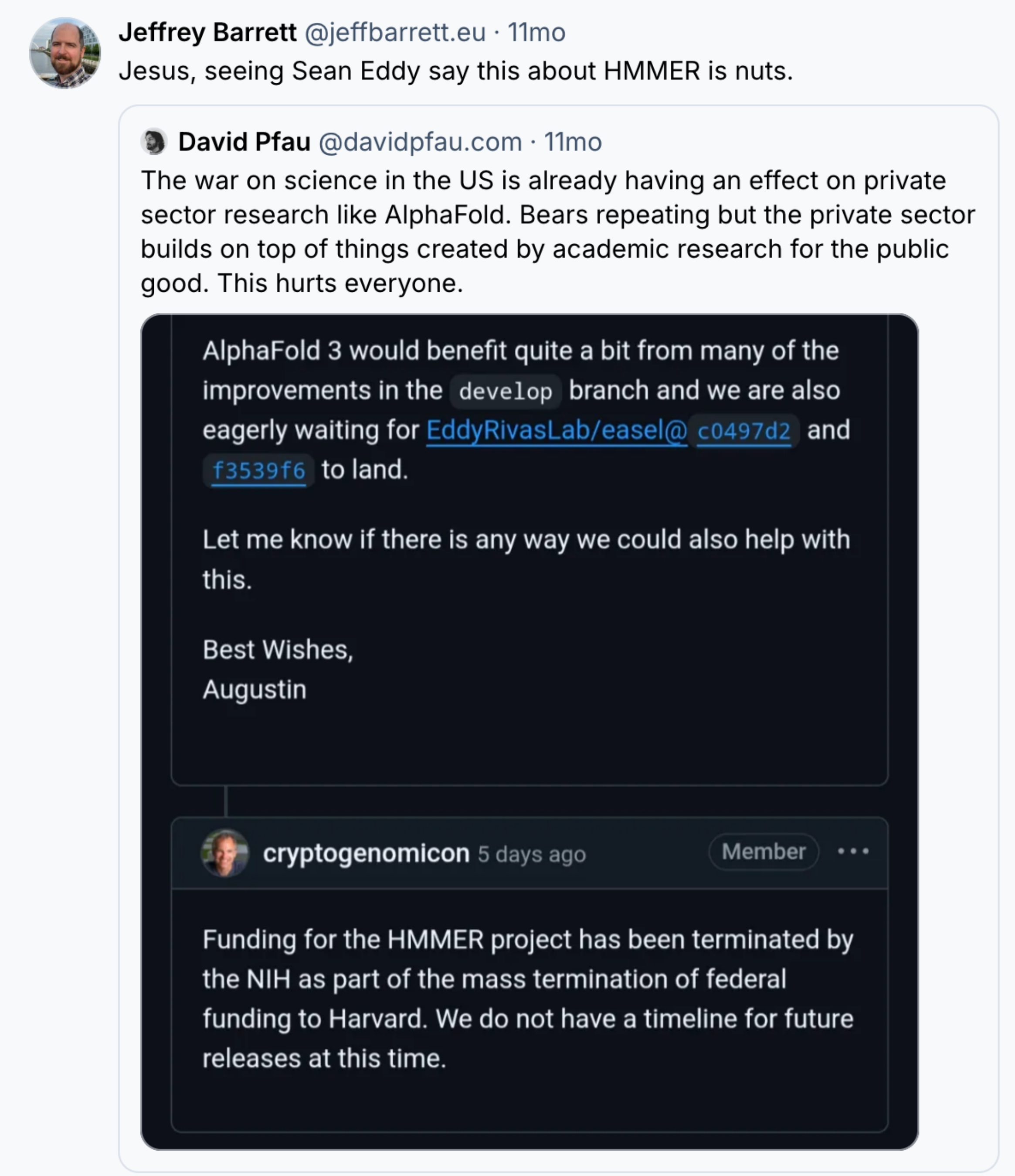


Doron et al. Science (2018)
Hu et al Nature (2025)

Disclaimer:

My work has been determined by the National Institutes of Health to be a misuse of taxpayer dollars and of no benefit to the American people.

Michelle Bulls, Director
NIH Office of Policy for Extramural Research
letter terminating NIH funding to Harvard, May 6, 2025
R.I.P. R01-HG009116, K99-GM146243, NSF GRF



Jeffrey Barrett @jeffbarrett.eu · 11mo
Jesus, seeing Sean Eddy say this about HMMER is nuts.

David Pfau @davidpfau.com · 11mo
The war on science in the US is already having an effect on private sector research like AlphaFold. Bears repeating but the private sector builds on top of things created by academic research for the public good. This hurts everyone.

AlphaFold 3 would benefit quite a bit from many of the improvements in the `develop` branch and we are also eagerly waiting for EddyRivasLab/easel@ `c0497d2` and `f3539f6` to land.

Let me know if there is any way we could also help with this.

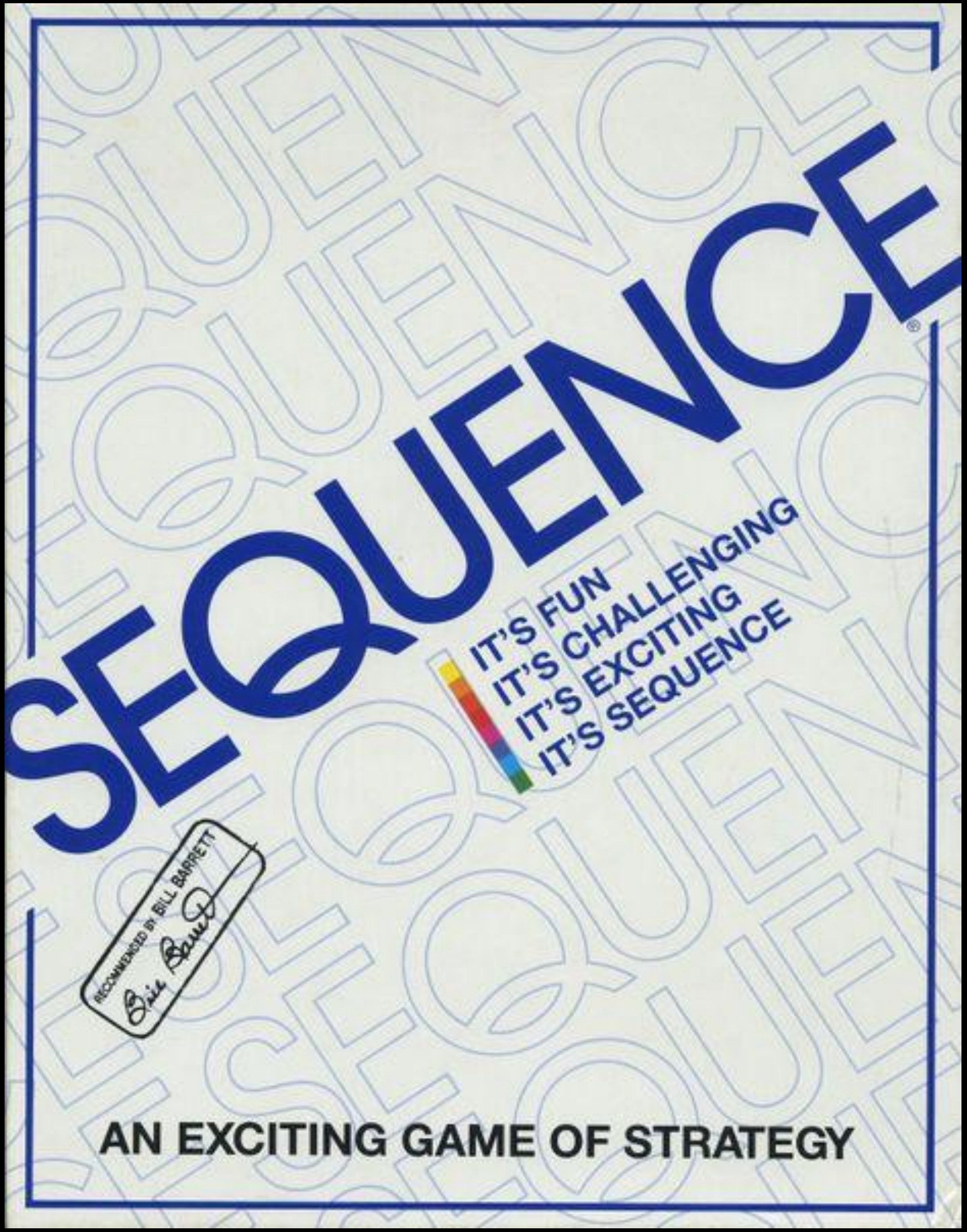
Best Wishes,
Augustin

cryptogenomicon 5 days ago

Funding for the HMMER project has been terminated by the NIH as part of the mass termination of federal funding to Harvard. We do not have a timeline for future releases at this time.

machines made of language

the power of comparative sequence analysis
to deduce structure and function of proteins and RNAs



Harvard University | eddylab.org

Nick Carter
Senior software engineer

Jolene Seme
Lab coordinator

Jenny Chen
Postdoc
(joint w/ Hopi Hoekstra)

Kumaresh Krishnan
Postdoc

Kumar Sarthak
Postdoc

Elena Rivas
Harvard University

Nick Hilgert
PhD student
G5, SSQ Biology
(joint with Kara McKinley)

Arpan Sarkar
PhD student
G5, Biostatistics

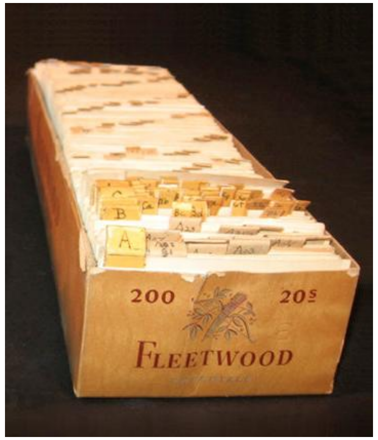
Liana Merk
PhD student
G3, Biophysics

Grace Luo
Harvard College
first year

Avi Burstein
Harvard College
sophomore

Eric Nawrocki
NIH NCBI

Harvard University | eddylab.org

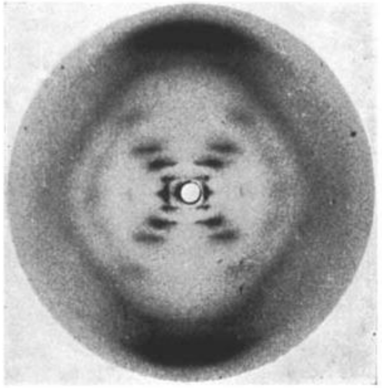


Kober's Linear B files

Alice E. Kober, 43; Lost to History No More

Margalit Fox, New York Times, May 11, 2013.

"…Alice Kober [was] an overworked, underpaid classics professor at Brooklyn College. In the mid-20th century, though hardly anyone knew it, Dr. Kober, working quietly and methodically at her dining table in Flatbush, helped solve one of the most tantalizing mysteries of the modern age."



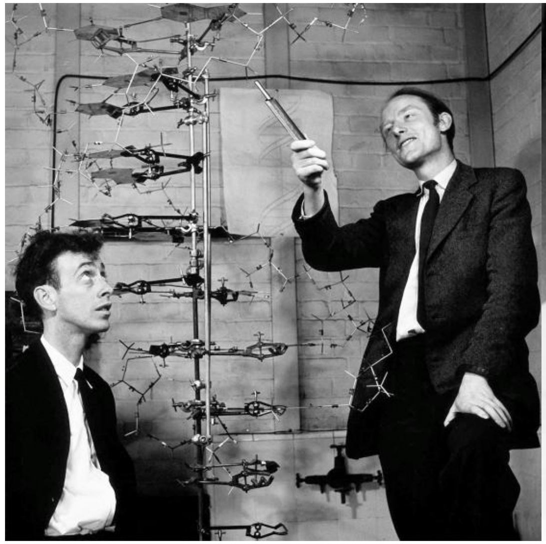
Franklin's Photograph 51

Alice Kober

John Chadwick

Michael Ventris

Jim Watson Francis Crick



Rosalind Franklin