

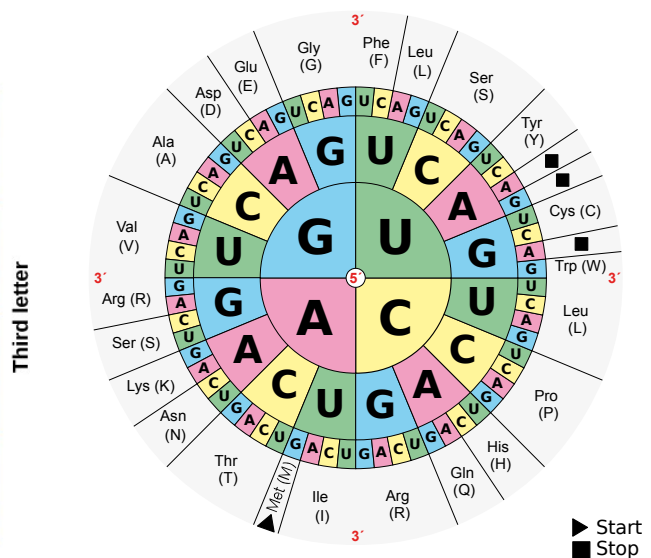
# ΥΠΟΛΟΓΙΣΤΙΚΗ ΓΛΩΣΣΟΛΟΓΙΑ

## ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ

### "Ο ΓΕΝΕΤΙΚΟΣ ΚΩΔΙΚΑΣ"

Ο Γενετικός Κώδικας γνωστός και ως "The Triplet Code" είναι ο κώδικας μετατροπής της πληροφορίας από το mRNA σε ακολουθίες αμινοξέων που δημιουργούν πρωτεΐνες. Ο κώδικας αυτός είναι κοινός για όλα τα έμβια όντα της Γης και για την ανακάλυψή του οι ερευνητές Marshall Nirenberg και Har Gobind Khorana τιμήθηκαν με βραβείο Nobel το 1968. Ο κώδικας φαίνεται στα ακόλουθα σχήματα, σε δύο μορφές, ως πίνακας (αριστερά) και ως δίσκος (δεξιά). Μπορείτε να χρησιμοποιήσετε όποιο σας είναι πιο εύκολο στην ανάγνωση.

		Second letter			
		U	C	A	G
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }
	A	AUU } Ile AUC } AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }
	G	GUU } Val GUC } GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }



Το mRNA κωδικοποιείται ως μια συμβολοσειρά με αλφάβητο  $\Sigma = \{U, C, A, G\}$ . Για παράδειγμα:

UGGCCAGUAUGACUGAUUCCAUGAGAUGAUCAC

Σε κάθε τρία συνεχόμενα σύμβολα αντιστοιχεί ένα αμινοξύ σύμφωνα με τον γενετικό κώδικα. Όλοι οι συνδυασμοί τριών συνεχόμενων συμβόλων είναι 64. Ο πίνακας δίνει την συντομογραφία του ονόματος του αμινοξέος για κάθε συνδυασμό, πχ. UGG => Trp (Tryptophan). Στον πίνακα η πληροφορία διαβάζεται γραμμές-στήλες (αριστερή γραμμή U, στήλη G, δεξιά γραμμή G => Trp), στον δίσκο από το κέντρο προς την περιφέρεια. Στον δίσκο φαίνεται επιπλέον και ένας μονοψήφιος κωδικός δίπλα στη συντομογραφία (πχ. W για το Trp), που χρησιμοποιείται αν θέλουμε να καταγράψουμε τις ακολουθίες αμινοξέων με ακόμα πιο συνοπτικό τρόπο.

Καθώς όλα τα αμινοξέα είναι 20 ενώ οι συνδυασμοί 64, κάποια αμινοξέα κωδικοποιούνται με περισσότερους από έναν συνδυασμούς, από 2 μέχρι και 6, όπως φαίνεται στον ακόλουθο κατάλογο που δείχνει όλους τους εναλλακτικούς συνδυασμούς για κάθε αμινοξύ:

Ala	:	GCU	GCC	GCA	GCG		
Arg	:	CGU	CGC	CGA	CGG	AGA	AGG
Asn	:	AAU	AAC				
Asp	:	GAU	GAC				
Cys	:	UGU	UGC				
Gln	:	CAA	CAG				
Glu	:	GAA	GAG				
Gly	:	GGU	GGC	GGA	GGG		
His	:	CAU	CAC				
Ile	:	AUU	AUC	AUA			
Leu	:	UUA	UUG	CUU	CUC	CUA	CUG
Lys	:	AAA	AAG				
Phe	:	UUU	UUC				
Pro	:	CCU	CCC	CCA	CCG		
Ser	:	UCU	UCC	UCA	UCG	AGU	AGC
Thr	:	ACU	ACC	ACA	ACG		
Trp	:	UGG					
Tyr	:	UAU	UAC				
Val	:	GUU	GUC	GUA	GUG		
Start & Met	:	AUG					
Stop	:	UAA	UAG	UGA			

Επιπλέον, υπάρχουν 4 συνδυασμοί ελέγχου (control) που καθορίζουν πού θα ξεκινήσει και πού θα σταματήσει η διαδικασία επεξεργασίας του mRNA. Ο συνδυασμός AUG σηματοδοτεί το σημείο έναρξης της επεξεργασίας και οι συνδυασμοί UAA, UAG και UGA τερματίζουν τη διαδικασία (οποιοσδήποτε από τους τρεις κι αν εμφανιστεί). Προφανώς αν δεν βρεθεί συνδυασμός τερματισμού η διαδικασία θα σταματήσει όταν τελειώσουν τα σύμβολα του mRNA.

Έτσι, το string του mRNA το διαβάζουμε από αριστερά προς τα δεξιά, χαρακτήρα προς χαρακτήρα, μέχρι να εντοπίσουμε την τριάδα AUG. Από το σημείο εκείνο ξεκινά η αποκωδικοποίηση. Όλοι οι προηγούμενοι χαρακτήρες αγνοούνται:

UGGCCAGU**AUG**ACUGAUUCCAUGAGAUGAUCAC

Στη συνέχεια, οι χαρακτήρες διαβάζονται όχι ένας-ένας, αλλά ανά τρεις:

UGGCCAGU**AUG**.ACU.GAU.UCC.AUG.AGA.UGA.UCA.C

και χρησιμοποιούμε τον κώδικα για να κάνουμε τη μετράτοπη κάθε τριάδας σε αμινοξύ:

UGGCCAGU	<b>AUG</b>	.ACU	.GAU	.UCC	.AUG	.AGA	<b>UGA</b>	.UCA	.C
	Met	Thr	Asp	Ser	Met	Arg			

Και εκεί σταματάμε, επειδή βρήκαμε τον συνδυασμό τερματισμού UGA. Όλη η υπόλοιπη συμβολοσειρά αγνοείται. Η πρωτεΐνη που έχει συντεθεί είναι η:

MetThrAspSerMetArg

Προσέξτε ότι οι συνδυασμοί τερματισμού δεν παράγουν αμινοξέα, ενώ ο συνδυασμός έναρξης AUG παράγει (το Met). Μάλιστα, ενδέχεται το AUG να βρεθεί και εσωτερικά μέσα στην ακολουθία, οπότε θα ξαναπαράξει Met χωρίς να θεωρηθεί εκεί ως συνδυασμός

έναρξης:

```
UGGCCAGUAUG . ACU . GAU . UCC . AUG . AGA . UGA . UCA . C
      |      |      |      |      |      |
      Met Thr Asp Ser Met Arg
```

Προσέξτε επίσης ότι η ίδια ακολουθία περιείχε και νωρίτερα τον συνδυασμό UGA, που όμως δεν θεωρήθηκε συνδυασμός τερματισμού καθώς οι 2 πρώτοι χαρακτήρες του ήταν στην προηγούμενη τριάδα και ο τρίτος στην επόμενη:

```
UGGCCAGUAUG . ACU . GAU . UCC . AUG . AGA . UGA . UCA . C
```

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] "DNA replication, mRNA transcription, and protein translation walk-through", Dr. Jennifer Louten (2022), YouTube, <https://www.youtube.com/watch?v=ORPtbsJD6kU>
- [2] "Transcription and Translation Practice Problems", Dr. Jennifer Louten (2022), YouTube, [https://www.youtube.com/watch?v=8xamWBYhv\\_w](https://www.youtube.com/watch?v=8xamWBYhv_w)
- [3] "The Genetic Code", Khan Academy, Course AP/College Biology, Unit 6, Lesson 4, <https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/translation/a/the-genetic-code-discovery-and-properties>
- [4] "The Genetic Code", OpenStax, Biology Unit 3, Chapter 15, <https://openstax.org/books/biology/pages/15-1-the-genetic-code>
- [5] "Genetic code", Wikipedia, [https://en.wikipedia.org/wiki/Genetic\\_code](https://en.wikipedia.org/wiki/Genetic_code)
- [6] "DNA and RNA codon tables", Wikipedia, [https://en.wikipedia.org/wiki/DNA\\_and\\_RNA\\_codon\\_tables](https://en.wikipedia.org/wiki/DNA_and_RNA_codon_tables)
- [7] "Nucleic Acids to Amino Acids: DNA Specifies Protein", Ann P. Smith, Ph.D. (2008), Nature Education 1(1):126, <https://www.nature.com/scitable/topicpage/nucleic-acids-to-amino-acids-dna-specifies-935/>
- [8] "Essential Amino Acids: Chart, Abbreviations and Structure", Karen Steward Ph.D. (2023), Technology Networks Applied Sciences, <https://www.technologynetworks.com/applied-sciences/articles/essential-amino-acids-chart-abbreviations-and-structure-324357>

## ΠΡΟΕΤΟΙΜΑΣΙΑ

Για την άσκηση αυτή θα χρειαστείτε το πρόγραμμα JFLAP (<https://www.jflap.org/>), το οποίο μπορείτε να κατεβάσετε απευθείας από εδώ: <https://www.jflap.org/jflaptmp/>  
Προτιμήστε την έκδοση 7.x. Αποφύγετε την 8.x. Δείτε τα εισαγωγικά videos εδώ:

[https://www.youtube.com/playlist?list=PLaAjeNjt7tTAH3LvvMVeR\\_rOVOgLLx6D](https://www.youtube.com/playlist?list=PLaAjeNjt7tTAH3LvvMVeR_rOVOgLLx6D)  
και το tutorial εδώ: <https://www.jflap.org/tutorial/> Υπάρχει επίσης και αυτό το πλήρες  
εγχειρίδιο: <https://www.jflap.org/jflapbook/jflapbook2006.pdf>

## ΔΙΑΔΙΚΑΣΙΑ

Σχεδιάστε στο JFLAP κατάλληλη μηχανή (Mealy, Moore, ό,τι θέλετε) που θα δέχεται ως είσοδο ένα string με σύμβολα από το αλφάβητο  $\Sigma = \{U, C, A, G\}$  και θα εξάγει το string με την πρωτεΐνη, σύμφωνα με τον γενετικό κώδικα, όπως περιγράφηκε.

Παραδείγματα:

ΕΙΣΟΔΟΣ	UGGCCAGUAUGA CUGAUUCCAUGA GAUGAUCAC	AGUAGCCGU AGGAUA	UGAAUGAUG AUGUC	CUGAACUGAAUGCUAGGAUC UUCUCAUAGAUAACUACGG UUACUAGUA
ΕΞΟΔΟΣ	MetThrAspSer MetArg		MetMetMet	MetLeuGlySerSerHisAr gSerThrThrValThrSer

Ελέγξτε τη μηχανή που σχεδιάσατε με τα παραπάνω παραδείγματα και άλλα 4 δικά σας, με Multiple Run και δείξτε screenshot με τη μηχανή και το test.

## ΠΑΡΑΔΟΤΕΟ

Παραδίδετε ένα συμπιεσμένο αρχείο (zip/rar) που περιέχει:

- ένα αρχείο κειμένου (κατά προτίμηση PDF) στο οποίο θα έχετε γράψει τα στοιχεία σας (Ονοματεπώνυμο, Τμήμα, Τομέας ή Πρόγραμμα Σπουδών, Έτος Φοίτησης και Αριθμός Μητρώου) και τις απαντήσεις (screenshots),
- το αρχείο (.jff) από το πρόγραμμα JFLAP με τη λύση της άσκησης.