

# Βασικές αρχές της γονιδιακής δομής, της γονιδιακής έκφρασης και της οργάνωσης του ανθρώπινου γονιδιώματος

## 2

### ΠΕΡΙΕΧΟΜΕΝΑ

2.1	ΓΟΝΙΔΙΑ ΠΟΥ ΚΩΔΙΚΕΥΟΥΝ ΠΡΩΤΕΪΝΕΣ: ΔΟΜΗ ΚΑΙ ΕΚΦΡΑΣΗ	23
2.2	ΓΟΝΙΔΙΑ ΤΟΥ RNA ΚΑΙ ΜΗ ΚΩΔΙΚΟ RNA	32
2.3	ΚΑΤΑΝΟΩΝΤΑΣ ΤΙΣ ΛΕΠΤΟΜΕΡΕΙΕΣ ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ ΜΑΣ ΚΑΙ ΤΙ ΣΗΜΑΙΝΟΥΝ	37
2.4	ΓΡΗΓΟΡΗ ΠΕΡΙΓΗΓΗΣΗ ΣΕ ΗΛΕΚΤΡΟΝΙΚΕΣ ΠΗΓΕΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΑΛΛΗΛΟΥΧΙΑΣ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ ΚΑΙ ΤΩΝ ΓΟΝΙΔΙΑΚΩΝ ΠΡΟΪΟΝΤΩΝ	41
2.5	Η ΟΡΓΑΝΩΣΗ ΚΑΙ Η ΕΞΕΛΙΞΗ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ	44
	ΠΕΡΙΛΗΨΗ	58
	ΕΡΩΤΗΣΕΙΣ	59
	ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΜΕΛΕΤΗ	60

Το γονιδιώμα μας είναι περίπλοκο, περιλαμβάνοντας περίπου 3,2 Gb ( $3,2 \times 10^9$  ζεύγη βάσεων, Gb = giga base = 1 δισεκατομμύριο βάσεις) DNA. Μία από τις κύριες εργασίες του ανθρώπινου γονιδιώματος είναι να παράγει μια τεράστια ποικιλία διαφορετικών πρωτεϊνών οι οποίες υπαγορεύουν τον τρόπο λειτουργίας των κυττάρων μας. Παραδόξως, το **κωδικό DNA (coding RNA)**, δηλαδή οι αλληλουχίες του DNA που καθορίζουν τα πολυπεπίδια των πρωτεϊνών μας, αντιστοιχούν σχεδόν μόνο στο 1,5% του DNA μας.

Το υπόλοιπο του γονιδιώματός μας είναι μη κωδικό DNA (non coding DNA) το οποίο δεν παράγει πρωτεΐνες. Ένα μεγάλο τμήμα του μη κωδικού DNA είναι λειτουργικά σημαντικό, και σε αυτό συμπεριλαμβάνονται πολλές διαφορετικές κατηγορίες ρυθμιστικών αλληλουχιών του DNA που ελέγχουν τον τρόπο με τον οποίο λειτουργούν τα γονίδια μας (όπως είναι οι υποκινητές και οι ενισχυτές) και αλληλουχίες του DNA που καθορίζουν βραχέος μήκους στοιχεία ρυθμιστικής αλληλουχίας που λειτουργούν σε επίπεδο RNA.

Επιπλέον, έχουμε πολλές χιλιάδες γονίδια που δεν παράγουν πολυπεπτιδικές αλυσίδες, αλλά δημιουργούν διαφορετικές κατηγορίες λειτουργικού μη κωδικού RNA. Μερικά από αυτά τα **γονίδια RNA (RNA genes)** –όπως γονίδια που κωδικεύουν το ριβοσωμικό RNA και το μεταφορικό RNA που χρειάζονται για τη σύνθεση πρωτεϊνών– είναι γνωστά εδώ και δεκαετίες, αλλά μία από τις μεγάλες εκπλήξεις τα τελευταία χρόνια ήταν ο τεράστιος αριθμός και η ποικιλία μη κωδικών RNA στα κύτταρά μας. Εκτός από τα γονίδια RNA, τα γονίδια μας που κωδικεύουν πρωτεΐνες δημιουργούν συχνά μετάγραφα μη κωδικών RNA καθώς και αγγελιαφόρα RNA (mRNAs).

Όπως άλλα πολύπλοκα γονιδιώματα, το γονιδιώμα μας έχει ένα μεγάλο ποσοστό μέτριου ως υψηλού βαθμού επαναλαμβανόμενες αλληλουχίες DNA. Μερικές από αυτές είναι σημαντικές για τη λειτουργία του κεντρομερούς και των τελομερών, ενώ άλλες είναι σημαντικές στην εξέλιξη του γονιδιώματος.

Μέχρι το 2003, το Πρόγραμμα Ανθρώπινου Γονιδιώματος (Human Genome Project – HGP) διαμόρφωσε τις πρώτες ολοκληρωμένες γνώσεις για το γονιδιώμα μας, παρέχοντας μια ουσιαστικά πλήρη νουκλεοτιδική αλληλουχία του πλούσιου σε γονίδια ευχρωματινικού συστατικού του γονιδιώματος. Οι επόμενες μελέτες συνέκριναν το γονιδιώμα μας με άλλα γονιδιώματα, βοηθώντας μας να κατανοήσουμε πώς εξελίχθηκε το γονιδιώμα μας. Οι συγκριτικές μελέτες των γονιδιωμάτων, μαζί με λειτουργικές και βιοπληροφορικές αναλύσεις σε όλο το γονιδίωμα, παρέχουν σημαντικές πληροφορίες για το πώς λειτουργεί το γονιδιώμα μας.

## 2.1 ΓΟΝΙΔΙΑ ΠΟΥ ΚΩΔΙΚΕΥΟΥΝ ΠΡΩΤΕΪΝΕΣ: ΔΟΜΗ ΚΑΙ ΕΚΦΡΑΣΗ

Οι πρωτεΐνες είναι τα κύρια λειτουργικά τελικά προϊόντα της γονιδιακής έκφρασης και επιτελούν μια τεράστια ποικιλία ρόλων που αφορούν τον τρόπο λειτουργίας των κυττάρων (ενεργώντας ως δομικά συστατικά, ένζυμα, πρωτεΐνες φορείς, διάλυτοι ιόντων, μόρια σηματοδότησης, ρυθμιστές γονιδίων κ.λπ.). Κάθε μία πρωτεΐνη αποτελείται από μία ή περισσότερες πολυπεπτιδικές αλυσίδες, μακριές αλληλουχίες αμινοξέων που κωδικεύονται από ένα κωδικό DNA. Σε πολλές περιπτώσεις μια πρωτεΐνη περιέχει επίσης υδατάνθρακες ή λιπιδικά συστατικά (τα οποία δεν καθορίζονται γενετικά).

Τα γονίδια που κωδικεύουν πρωτεΐνες έχουν μια τεράστια ποικιλία οργάνωσης, όπως περιγράφεται παρακάτω, και συνθέτουν μία ή περισσότερες πολυπεπτιδικές αλυσίδες. Ωστόσο, η σύνθεση του πολυπεπτιδίου δεν είναι το τελικό προϊόν. Μία νεοσυντιθέμενη πολυπεπτιδική αλυσίδα πρέπει να υποβληθεί σε πολλαπλά διαφορετικά στάδια ωρίμασης, που συνήθως περιλαμβάνουν γεγονότα χημικής τροποποίησης και διάσπασης (cleavage), και συχνά στη συνέχεια συνδέεται με άλλες πολυπεπτιδικές αλυσίδες για να σχηματίσουν μια λειτουργική πρωτεΐνη.

### Οργάνωση γονιδίων: εξόνια και ιντρόνια

Τα γονίδια που κωδικεύουν τις πρωτεΐνες των βακτηρίων είναι μικρά (κατά μέσο όρο έχουν μήκος περίπου 1000 ζεύγη βάσεων) και απλά. Το γονίδιο μεταγράφεται για να προκύψει ένα mRNA με μια συνεχή κωδική αλληλουχία η οποία στη συνέχεια μεταφράζεται για να δώσει μια γραμμική αλληλουχία περίπου 300 αμινοξέων κατά μέσο όρο. Απροσδόκητα, τα γονίδια των ευκαρυωτικών κυττάρων αποδείχθηκαν πολύ μεγαλύτερα και πολύ πιο περίπλοκα από ό,τι αναμενόταν. Και, όπως θα δούμε, τα γονίδια μας που κωδικεύουν πρωτεΐνες περιέχουν συχνά μια μάλλον μικρή ποσότητα κωδικού DNA.

Για τα περισσότερα ευκαρυωτικά γονίδια που κωδικεύουν πρωτεΐνες, το κωδικό DNA χωρίζεται σε τμήματα (**εξόνια**<sup>1</sup> – **exons**) που διαχωρίζονται από μη κωδικές αλληλουχίες DNA (**ιντρόνια**<sup>2</sup> – **introns**). Ο αριθμός των εξονίων και των ιντρονίων σε ένα γονίδιο ποικίλλει σημαντικά (δεν φαίνεται να υπάρχει μια συγκεκριμένη λογική για το που ακριβώς εισάγονται τα ιντρόνια εντός των γονιδίων).

Εξαιρετικών των γονιδίων που φέρουν ένα εξόνιο (ορισμένα γονίδια στερούνται ιντρονίων), το μέσο μήκος των εξονίων παρουσιάζει μέτρια διακύμανση από γονίδιο σε γονίδιο, αλλά τα ιντρόνια μπορεί να εμφανίζουν εξαιρετικά μεγάλες διαφορές μεγέθους. Τα γονίδια μας είναι επομένως συχνά μεγάλα, έχοντας μερικές φορές μήκος μεγαλύτερο από ένα εκατομμύριο βάσεις DNA (**Πίνακα 2.1**).

### Συναρμογή RNA: συρραφή των γενετικών πληροφοριών σε εξόνια

Όπως όλα τα γονίδια, τα γονίδια που χωρίζονται σε εξόνια μεταγράφονται αρχικά από μια RNA πολυμεράση για να δώσουν ένα μακρύ μετάγραφο RNA. Αυτό το πρωτογενές μετάγραφο (ή πρόδρομο mRNA, pre-mRNA) είναι πανομοιότυπο σε αλληλουχία βάσεων με τη μεταγραφόμενη περιοχή της νοηματικής αλυσίδας του DNA, εκτός από το ότι η U αντικαθιστά τη T (η μεταγραφόμενη περιοχή του DNA ονομάζεται **μονάδα μεταγραφής** – **transcription unit**). Στη συνέχεια, το πρωτογενές μετάγραφο

1. Προέρχεται από τον αγγλικό όρο “expressed sequence” (εκφραζόμενη αλληλουχία), ωστόσο παγιώθηκε ως “exon”. Στην ελληνική βιβλιογραφία αποδόθηκε ως εξόνιο και στο παρόν χρησιμοποιείται στην απλοποιημένη του γραφή ως εξόνιο (Σ.τ.Ε.).

2. Προέρχεται από τον αγγλικό όρο “intra-genic region” (ενδογονιδιακή περιοχή) (κατά W. Gilbert, 1978) και κατά άλλους από τον όρο “intervening sequence” (ενδιάμεση ή παρεμβαλλόμενη αλληλουχία), όροι οι οποίοι σε τελική ανάλυση μπορούν να θεωρηθούν συνώνυμοι. Στην ελληνική βιβλιογραφία αποδόθηκε αρχικά ως εσώνιο (Φ. Καφάτος), αλλά στο παρόν έχει υιοθετηθεί η νέα απόδοση του όρου ως ιντρόνιο (Σ.τ.Ε.).

**ΠΙΝΑΚΑΣ 2.1 ΠΑΡΑΔΕΙΓΜΑΤΑ ΔΙΑΦΟΡΙΚΗΣ ΓΟΝΙΔΙΑΚΗΣ ΟΡΓΑΝΩΣΗΣ ΓΙΑ ΑΝΘΡΩΠΙΝΑ ΓΟΝΙΔΙΑ ΠΟΥ ΚΩΔΙΚΕΥΟΥΝ ΠΡΩΤΕΪΝΕΣ**

Ανθρώπινο γονίδιο	Μέγεθος γονιδιώματος (kb)	Αριθμός εξονίων	Μέσο μέγεθος εξονίου (bp)*	Μέσο μέγεθος εξονίου (bp)**
<i>SRY</i> (καθοριστής φύλου στους άντρες)	0,9	1	850	–
<i>HBB</i> (β-σφαιρίνη)	1,6	3	150	490
<i>TP53</i> (p53)	39	10	236	3076
<i>F8</i> (παράγοντας VIII)	186	26	375	7100
<i>CFTR</i> (διαμεμβρανικός ρυθμιστής κυστικής ίνωσης)	250	27	227	9100
<i>DMD</i> (δυστροφίνη)	2400	79	180	30.770

Τα στοιχεία σε παρένθεση υποδεικνύουν το όνομα της πρωτεΐνης. Kb, χιλιοβάσεις (=1000 ζεύγη βάσεων). bp, base pairs (ζεύγη βάσεων)

\* Ας σημειωθεί ότι το βραχύτερο ανθρώπινο εξόνιο έχει μήκος μόνο δύο νουκλεοτίδια, ενώ τα τελικά εξόνια μπορούν να είναι αρκετά συχνά μεγάλα σε μήκος, με το μακρύτερο που έχει καταγραφεί να έχει μήκος 27.303 bp.

\*\* Το βραχύτερο ανθρώπινο ιντρόνιο έχει μήκος 26 bp, ενώ το μακρύτερο έχει μήκος 1.160.411 bp – βλ. PMID 31164174.

RNA υφίσταται μια μορφή επεξεργασίας που ονομάζεται **συναρμογή του RNA<sup>3</sup> (RNA splicing)**.

Η συναρμογή του RNA περιλαμβάνει πρώτα την τομή του μετάγραφου RNA στις συνδέσεις μεταξύ των εξονίων και των ιντρονίων που μεταγράφηκαν. Οι μεμονωμένες αλληλουχίες των ιντρονίων που έχουν μεταγραφεί συχνά αποικοδομούνται, αλλά οι αλληλουχίες των εξονίων που έχουν μεταγραφεί στη συνέχεια συνδέονται ομοιοπολικά (συναρμολογούνται) με τη σειρά για να δημιουργήσουν ένα ώριμο RNA (**Εικόνα 2.1**). Η συναρμογή του RNA πραγματοποιείται εντός του πυρήνα από σωματίδια συναρμογής (spliceosomes), που είναι σύνθετα σύμπλοκα πρωτεϊνικών παραγόντων και μορίων μικρού πυρηνικού RNA (small nuclear RNA – snRNA).

Δεν κατανοούμε πλήρως πώς το σωματίδιο συναρμογής μπορεί να αναγνωρίσει και να κόψει το πρωτογενές μετάγραφο σε ακριβείς θέσεις που σηματοδοτούν την αρχή και το τέλος των ιντρονίων. Ωστόσο, αυτό που γνωρίζουμε είναι ότι ορισμένες αλληλουχίες είναι σημαντικές για τη σηματοδότηση των θέσεων συναρμογής που ορίζουν τα όρια εξονίου-ιντρονίου. Για παράδειγμα, σχεδόν όλα τα ιντρόνια ξεκινούν με ένα δινουκλεοτίδιο GT στην κωδική αλυσίδα του DNA και τελειώνουν με ένα AG, έτσι ώστε η μεταγραφόμενη αλληλουχία των ιντρονίων να αρχίζει με ένα GU (που σηματοδοτεί τη **θέση δότη της συναρμογής – splice donor site**) και να τελειώνει σε ένα AG (που σηματοδοτεί τη **θέση**

**δέκτη της συναρμογής – splice acceptor site**). Οι ακραίες αλληλουχίες GT (GU) και AG πρέπει να ενσωματωθούν σε ευρύτερες συναινετικές αλληλουχίες (consensus sequences) στις θέσεις συναρμογής που θα περιγράψουμε στην Ενότητα 6.1 όταν εξετάσουμε πώς ρυθμίζεται η γονιδιακή έκφραση. Όπως θα δούμε στο Κεφάλαιο 7, οι μεταλλαγές στις θέσεις συναρμογής είναι σημαντικές αιτίες νόσου.

Η Εικόνα 2.1 μπορεί να δώσει τη λανθασμένη εντύπωση ότι όλα τα γονίδια που κωδικεύουν πρωτεΐνες υφίστανται έναν συγκεκριμένο, μοναδικό τύπο συναρμογής του RNA. Ωστόσο, σχεδόν το 10% των γονιδίων μας που κωδικεύουν τις πρωτεΐνες έχουν ένα μόνο, μη διακοπτόμενο εξόνιο και δεν υφίστανται καθόλου συναρμογή του RNA – αξιοσημείωτα παραδείγματα περιλαμβάνουν τα γονίδια των ιστονών. Και τα περισσότερα από τα γονίδια που περνούν από τη συναρμογή του RNA υφίστανται εναλλακτικά μοτίβα συναρμογής του RNA. Ένα μόνο γονίδιο μπορεί επομένως να παράγει διαφορετικά γονιδιακά προϊόντα που μπορεί να είναι λειτουργικά διαφορετικά. Εξετάζουμε την έννοια της εναλλακτικής συναρμογής με περισσότερες λεπτομέρειες στο Κεφάλαιο 6, στο πλαίσιο της γονιδιακής ρύθμισης.

### Η εξελικτική αξία της συναρμογής του RNA

Όπως θα δούμε στην Ενότητα 2.2, πολλά γονίδια RNA υφίστανται επίσης συναρμογή του RNA.

3. Ο όρος “splicing” δεν μπορεί να αποδοθεί μονολεκτικά στην ελληνική γλώσσα. Με τον αγγλικό όρο περιγράφεται η αφαίρεση των ιντρονίων και η σύνδεση των εξονίων. Οι χρησιμοποιούμενοι όροι απόδοσης στην ελληνική βιβλιογραφία “μάτισμα”, “συρραφή”, “συναρμογή” μπορούν να θεωρηθούν συνώνυμοι της “σύνδεσης” κι ως εκ τούτου μπορούν να χρησιμοποιηθούν εάν προσδιορίζεται το “τι ματίζεται”, “τι συρράπτεται” ή το “τι συναρμολογείται”, δηλαδή εν προκειμένω τα εξόνια, μετά την αποκοπή και αφαίρεση των ιντρονίων. Στον παρόν έχει επιλεγεί η χρήση του όρου συναρμογή (συναρμολόζω: συνδέω δύο ή περισσότερα ομοειδή στοιχεία μεταξύ τους) (Σ.τ.Ε.).

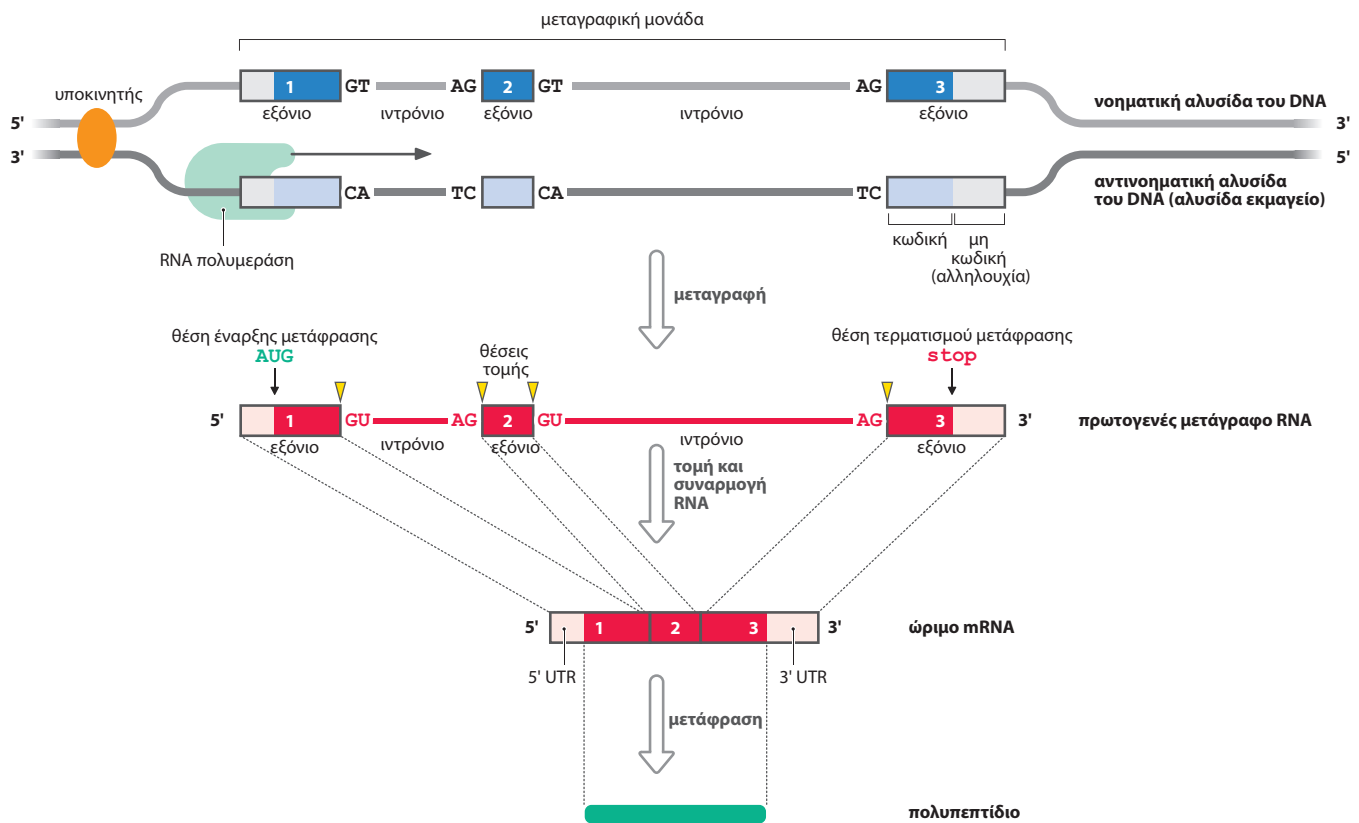
Σε αυτό το στάδιο, θα μπορούσε κανείς εύλογα να αναρωτηθεί γιατί η συναρμογή του RNA είναι τόσο σημαντική στα ευκαρυωτικά κύτταρα και τόσο ιδιαίτερα διαδεδομένη σε σύνθετους πολυκύτταρους οργανισμούς. Γιατί χρειάζεται να χωριστούν οι γενετικές πληροφορίες στα γονίδια σε τόσα πολλά διαφορετικά μικρά εξόνια; Η απάντηση είναι για να συμβάλουν στο έναυσμα σχηματισμού νέων γονιδίων και νέων γονιδιακών προϊόντων που μπορούν να επιτρέψουν μεγαλύτερη λειτουργική πολυπλοκότητα κατά τη διάρκεια της εξέλιξης.

Η τεράστια πολυπλοκότητα των ανθρώπων και άλλων πολυκύτταρων οργανισμών οφείλεται στην εξέλιξη του γονιδιώματος. Εκτός από τον περιοδικό διπλασιασμό των γονιδίων, διάφοροι γενετικοί μηχανισμοί επιτρέπουν μεμονωμένα εξόνια να αντι-

γράφονται ή να ανταλλάσσονται από το ένα γονίδιο στο άλλο σε μια εξελικτική χρονική κλίμακα. Αυτό επιτρέπει διαφορετικούς τρόπους συνδυασμού εξονίων για την παραγωγή νέων υβριδικών γονιδίων. Μια επιπλέον πηγή πολυπλοκότητας προέρχεται από τη χρήση διαφορετικών συνδυασμών εξονίων για τη δημιουργία εναλλακτικών μεταγράφων από το ίδιο γονίδιο (εναλλακτική συναρμογή).

### Μετάφραση: η αποκωδικοποίηση του αγγελιαφόρου RNA για την παραγωγή μιας πολυπεπτιδικής αλυσίδας

Τα μόρια του αγγελιαφόρου RNA (mRNA) που παράγονται από την συναρμογή του RNA στον πυ-



**Εικόνα 2.1 Η συναρμογή του RNA συνδέει τις αλληλουχίες των εξονίων που έχουν μεταγραφεί.** Τα περισσότερα από τα γονίδια που κωδικεύουν πρωτεΐνες (και πολλά γονίδια RNA) υφίστανται συναρμογή του RNA. Σε αυτό το γενικευμένο παράδειγμα απεικονίζεται ένα γονίδιο που κωδικεύει μια πρωτεΐνη με έναν ανοδικό υποκινητή και τρία εξόνια που διαχωρίζονται από δύο ιντρόνια όπου το καθένα ξεκινά με το διουκλεοτίδιο GT και τελειώνει στο διουκλεοτίδιο AG. Το κεντρικό εξόνιο (εξόνιο 2) αποτελείται εξ' ολοκλήρου από κωδικό DNA (βαθύ κόκκινο), αλλά τα εξόνια 1 και 3 έχουν τόσο κωδικό DNA όσο και μη κωδικές αλληλουχίες DNA (εμφανίζονται με ροζ χρώμα· τελικά θα χρησιμοποιηθούν για τη δημιουργία αμετάφραστων αλληλουχιών στο mRNA). Τα τρία εξόνια και τα δύο διαχωρισμένα ιντρόνια μεταγράφονται μαζί για να δώσουν ένα μεγάλο πρωτογενές μετάγραφο RNA. Το μετάγραφο RNA τέμνεται σε θέσεις που αντιστοιχούν στα όρια εξονίου-ιντρόνιου. Η καθεμία από τις δύο αλληλουχίες των ιντρόνιων που μεταγράφηκαν και στη συνέχεια αποκόπηκαν, αποικοδομείται, ενώ οι αλληλουχίες εξονίων που έχουν μεταγραφεί θα ενωθούν (θα συναρμοστούν) μαζί για να σχηματίσουν ένα συνεχόμενο ώριμο RNA το οποίο έχει μη κωδικές αλληλουχίες τόσο στο 5' άκρο, όσο και στο 3' άκρο. Στο ώριμο mRNA αυτές οι τελικές αλληλουχίες στα 5' και 3' άκρα του δεν θα μεταφραστούν κι έτσι είναι γνωστές ως αμετάφραστες περιοχές (UTRs). Η κεντρική κωδική αλληλουχία του mRNA ορίζεται από μια θέση έναρξης της μετάφρασης (η οποία είναι σχεδόν πάντα είναι το τριουκλεοτίδιο AUG) και μια θέση τερματισμού της μετάφρασης και διαβάζεται (μεταφράζεται) για να παραχθεί μια πολυπεπτιδική αλυσίδα.

ρήνα εξάγονται στο κυτταρόπλασμα. Εδώ δεσμεύονται από ριβοσώματα, πολύ μεγάλα σύμπλοκα απαρτιζόμενα από τέσσερις τύπους ριβοσωμικού RNA (ribosomal RNA – rRNA) και πολλές διαφορετικές πρωτεΐνες.

Αν και ένα mRNA σχηματίζεται μόνο από εξόνια, έχει αλληλουχίες στα άκρα 5' και 3' που δεν είναι κωδικές. Έχοντας δεσμευτεί στο mRNA, το έργο των ριβοσωμάτων είναι να σαρώσουν την αλληλουχία του mRNA για να βρουν και να ερμηνεύσουν μια κεντρική κωδική αλληλουχία που θα μεταφραστεί για να συντεθεί μια πολυπεπτιδική αλυσίδα. Οι μη κωδικές αλληλουχίες στα άκρα του mRNA είναι γνωστές ως **αμετάφραστες περιοχές (untranslated regions – UTRs)**, όπως φαίνεται στην Εικόνα 2.1) και περιέχουν αλληλουχίες που είναι σημαντικές για τη ρύθμιση της γονιδιακής έκφρασης.

Μια πολυπεπτιδική αλυσίδα είναι ένα πολυμερές αποτελούμενο από μια γραμμική αλληλουχία **αμινοξέων (amino acids)** (Εικόνα 2.2A). Τα αμινοξέα έχουν τον γενικό τύπο  $\text{NH}_2\text{-CH(R)-COOH}$ , όπου το R είναι μια μεταβλητή πλευρική αλυσίδα που καθορίζει τη χημική ταυτότητα του αμινοξέος και συνδέεται με το κεντρικό (άλφα) άτομο άνθρακα της αλληλουχίας πλαισίου  $\text{NH-CH-CO}$ . Υπάρχουν 20 κοινά αμινοξέα (Εικόνα 2.2Γ). Οι πολυπεπτιδικές αλυσίδες παράγονται με μια αντίδραση συμπύκνωσης μεταξύ της καρβοξυλομάδας (COOH) ενός αμινοξέος και της αμινομάδας ( $\text{NH}_2$ ) ενός άλλου αμινοξέος, σχηματίζοντας έναν πεπτιδικό δεσμό (βλέπε Εικόνα 2.2B).

Για να δημιουργηθεί μια πολυπεπτιδική αλυσίδα, η κωδική αλληλουχία σε ένα mRNA μεταφράζεται σε ομάδες τριών νουκλεοτιδίων κάθε φορά, που ονομάζονται **κωδικόνια (codons)**. Υπάρχουν 64 πιθανά κωδικόνια (τέσσερις πιθανές βάσεις σε καθεμία από τις τρεις θέσεις νουκλεοτιδίων δίνουν  $4 \times 4 \times 4$  συνδυασμούς). Από αυτά, 61 χρησιμοποιούνται για τον προσδιορισμό ενός αμινοξέος, τα άλλα τρία σηματοδοτούν το τέλος της πρωτεϊνοσύνθεσης. Ο καθολικός **γενετικός κώδικας (genetic code)**, το σύνολο των κανόνων που υπαγορεύουν τον τρόπο ερμηνείας των κωδικονίων, έχει συνεπώς ενσωματωμένο κάποιον πλεονασμό (redundancy). Για παράδειγμα, το αμινοξύ σερίνη μπορεί να προσδιοριστεί με οποιοδήποτε από τα έξι κωδικόνια (UCA, UCC, UCG, UCU, AGU και AGG) και, κατά μέσο όρο, ένα αμινοξύ προσδιορίζεται από οποιοδήποτε από τρία κωδικόνια. Ως αποτέλεσμα, οι αντικαταστάσεις νουκλεοτιδίων εντός του κωδικού DNA αρκετά συχνά δεν προκαλούν αλλαγή του αμινοξέος.

Εξετάζουμε τον γενετικό κώδικα λεπτομερέστερα στην Ενότητα 7.2 όταν θα μελετήσουμε τα αποτελέσματα των αντικαταστάσεων ενός νουκλεοτιδίου.

## Η διαδικασία της μετάφρασης

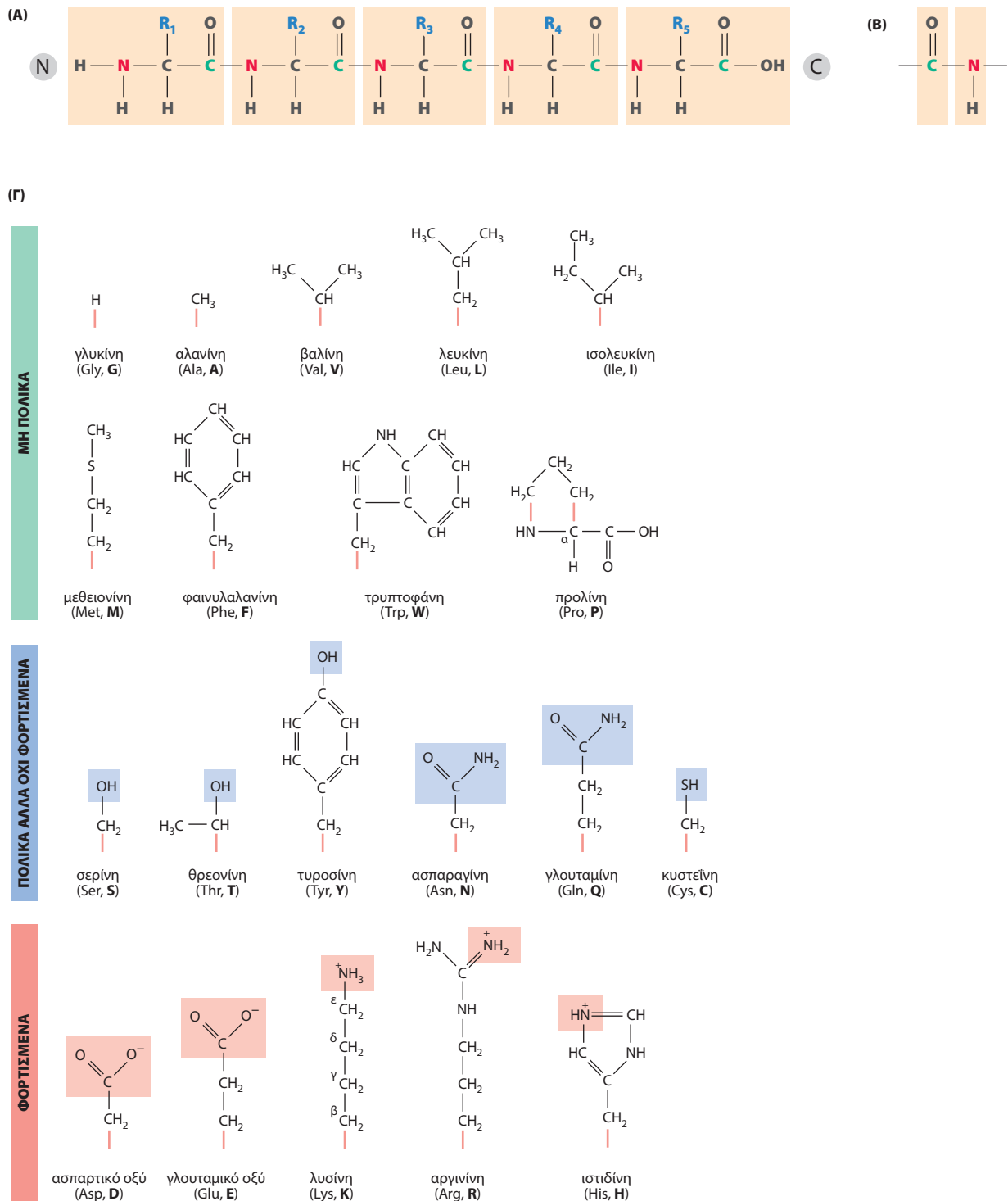
Η μετάφραση αρχίζει όταν τα ριβοσώματα συνδέονται στο 5' άκρο ενός mRNA και στη συνέχεια κινούνται κατά μήκος του RNA για να βρουν μια θέση έναρξης μετάφρασης, το κωδικόνιο έναρξης – ένα τρινουκλεοτίδιο AUG ενσωματωμένο στην ευρύτερη, λιγότερο καλά καθορισμένη συναινετική αλληλουχία Kozak (GCCPuCCAUGG• οι πιο συντηρημένες βάσεις εμφανίζονται με έντονους χαρακτήρες και το Pu αντιπροσωπεύει μια πουρίνη).

Το κωδικόνιο έναρξης είναι η αρχή ενός **ανοιχτού πλαισίου ανάγνωσης (open reading frame, ORF)** κωδικονίων που προσδιορίζουν διαδοχικά αμινοξέα στην πολυπεπτιδική αλυσίδα (βλέπε Πλαίσιο 2.1 για την έννοια των μεταφραστικών πλαισίων ανάγνωσης). Όπως περιγράφεται παρακάτω, μια οικογένεια **μεταφορικών RNA (transfer RNAs – tRNAs)** είναι υπεύθυνη για τη μεταφορά των σωστών αμινοξέων που πρέπει να εισαχθούν στην απαιτούμενη θέση της αναπτυσσόμενης πολυπεπτιδικής αλυσίδας. Μεμονωμένοι τύποι tRNA μεταφέρουν ένα συγκεκριμένο αμινοξύ και μπορούν να αναγνωρίσουν και να δεσμευτούν σε ένα συγκεκριμένο κωδικόνιο και όταν το κάνουν ξεφορτώνουν το φορτίο του αμινοξέος τους.

Καθώς κάθε νέο αμινοξύ εκφορτώνεται, συνδέεται με το προηγούμενο αμινοξύ έτσι ώστε να σχηματίζεται μια πολυπεπτιδική αλυσίδα (Εικόνα 2.3). Το πρώτο αμινοξύ έχει μια ελεύθερη ομάδα  $\text{NH}_2$  (αμινομάδα) και σηματοδοτεί το N-τερματικό άκρο (N) του πολυπεπτιδίου. Η σύνθεση της πολυπεπτιδικής αλυσίδας τερματίζεται όταν το ριβόσωμα συναντήσει ένα **κωδικόνιο λήξης (stop codon)** (το οποίο σημαίνει ότι το ριβόσωμα πρέπει να αποδεσμευτεί από το mRNA, απελευθερώνοντας το πολυπεπτίδιο· για τη μετάφραση σε κυτταροπλασματικά ριβοσώματα, υπάρχουν τρεις επιλογές κωδικονίου λήξης: UAA, UAG ή UGA). Το τελευταίο αμινοξύ που ενσωματώθηκε στην πολυπεπτιδική αλυσίδα έχει ελεύθερη ομάδα COOH (καρβοξυλομάδα) και σηματοδοτεί το C-τελικό άκρο (C) του πολυπεπτιδίου.

## Το μεταφορικό RNA ως RNA προσαρμοστής

Τα μεταφορικά RNA έχουν μια κλασική δομή τριφυλλίου που προκύπτει από τους ενδομοριακούς δεσμούς υδρογόνου (Εικόνα 2.4A). Λειτουργούν ως



**Εικόνα 2.2 Δομή πολυπεπτιδικής αλυσίδας και αμινοξέων.** (Α) Πρωτοταγής δομή πολυπεπτιδικής αλυσίδας. Απεικονίζεται ένα πενταπεπτιδίο με επισήμανση των πέντε αμινοξέων του. Εδώ, το αριστερό άκρο ονομάζεται N-τελικό άκρο (αμινοτελικό άκρο) επειδή το αμινοξύ έχει μια ελεύθερη αμινομάδα (NH<sub>2</sub>). Το δεξί άκρο είναι το C-τελικό άκρο (καρβοξυτελικό άκρο) επειδή το τελευταίο αμινοξύ έχει μια ελεύθερη καρβοξυλομάδα (COOH). Οι πλευρικές αλυσίδες (R<sub>1</sub> έως R<sub>5</sub>) είναι μεταβλητές και καθορίζουν την ταυτότητα του αμινοξέος. Συνδέονται με το κεντρικό άτομο άνθρακα της επαναλαμβανόμενης αλληλουχίας πλαισίου: -NH-CH-CO-. Ας σημειωθεί ότι σε φυσιολογικό pH οι ελεύθερες αμινομάδες και οι καρβοξυλομάδες θα φορτιστούν: NH<sub>3</sub><sup>+</sup> και COO<sup>-</sup>, αντίστοιχα. (Β) Τα γειτονικά αμινοξέα συνδέονται με έναν πεπτιδικό δεσμό. Ένας πεπτιδικός δεσμός σχηματίζεται από μια αντίδραση συμπύκνωσης μεταξύ της τελικής καρβοξυλομάδας ενός αμινοξέος και της τελικής αμινομάδας ενός άλλου: -COOH + NH<sub>2</sub> -> -CONH- + H<sub>2</sub>O. (Γ) Οι πλευρικές αλυσίδες των 20 κύριων αμινοξέων. Οι κόκκινες γραμμές αναπαριστούν τον ομοιοπολικό δεσμό που συνδέει την πλευρική αλυσίδα στη δομή του πλαισίου της πρωτεΐνης. Παρατηρήστε ότι η δομή της προλίνης είναι ασυνήθιστη και η πλήρης δομή της δίνεται εδώ επειδή η πλευρική της αλυσίδα συνδέεται με το άτομο αζώτου της αμινομάδας πλαισίου καθώς και με τον άλφα άνθρακα, σχηματίζοντας έτσι έναν πενταμελή δακτύλιο.

RNA προσαρμοστές (adaptor RNAs) επειδή η δουλειά τους είναι να ζευγαρώνουν βάσεις με τα mRNA και να συμβάλλουν στην αποκωδικοποίηση των μηνυμάτων της κωδικής αλληλουχίας που μεταφέρονται από τα mRNA. Το ζευγάρι των βάσεων περιορίζεται σε μια αλληλουχία τριών νουκλεοτιδίων στο tRNA που ονομάζεται *αντικωδικόνιο*, η οποία είναι συμπληρωματική προς την αλληλουχία ενός κωδικονίου. Σύμφωνα με την ταυτότητα των αντικωδικονίων τους, διαφορετικά tRNA φέρουν διαφορετικά αμινοξέα ομοιοπολικά συνδεδεμένα στα 3' άκρα τους. Μέσω του ζευγαρώματος βάσεων μεταξύ του κωδικονίου και του αντικωδικονίου, τα μεμονωμένα αμινοξέα μπορούν να διαταχθούν διαδοχικά σύμφωνα με την αλληλουχία των κωδικονίων σε ένα mRNA και να συνδεθούν διαδοχικά

μεταξύ τους για να σχηματίσουν μια πολυπεπτιδική αλυσίδα (βλέπε **Εικόνα 2.4B**).

### Αμετάφραστες περιοχές, 5' καλύπτρα και 3' πολυ(A) άκρο

Όπως απεικονίζεται στην Εικόνα 2.3, κάθε ώριμο mRNA έχει μια μεγάλη κεντρική κωδική αλληλουχία του DNA που πλαισιώνεται από δύο **αμετάφραστες περιοχές (untranslated regions)**, μια μικρή 5' αμετάφραστη περιοχή (5' UTR) και μια μάλλον μακρύτερη 3' αμετάφραστη περιοχή (3' UTR). Οι αμετάφραστες περιοχές ρυθμίζουν τη σταθερότητα του mRNA και περιέχουν ρυθμιστικές αλληλουχίες που είναι σημαντικές για τον προσδιορισμό του τρόπου έκφρασης των γονιδίων.

Εκτός από τις αλληλουχίες που μεταγράφονται

## ΠΛΑΙΣΙΟ 2.1 ΜΕΤΑΦΡΑΣΤΙΚΑ ΠΛΑΙΣΙΑ ΑΝΑΓΝΩΣΗΣ ΚΑΙ ΔΙΑΧΩΡΙΣΜΟΣ ΤΩΝ ΚΩΔΙΚΩΝ ΑΛΛΗΛΟΥΧΙΩΝ ΑΠΟ ΙΝΤΡΟΝΙΑ

### ΜΕΤΑΦΡΑΣΤΙΚΑ ΠΛΑΙΣΙΑ ΑΝΑΓΝΩΣΗΣ

Στα παραδείγματα διαφορετικών μεταφραστικών **πλαisiών ανάγνωσης (reading frames)** που απεικονίζονται παρακάτω, χρησιμοποιούμε αλληλουχίες λέξεων που περιέχουν τρία γράμματα για να αναπαραστήσουμε την τριπλή φύση του γενετικού κώδικα. Ορίζουμε τα πλαίσια ανάγνωσης (RF) ως 1, 2 ή 3 ανάλογα με το αν το πλαίσιο ανάγνωσης αρχίζει πριν από το πρώτο, δεύτερο ή τρίτο νουκλεοτίδιο της αλληλουχίας.

Η ανάγνωση του πλαισίου 1 (RF1) στην **Εικόνα 1** έχει νόημα, αλλά η μετατόπιση σε άλλο πλαίσιο ανάγνωσης δεν παράγει προϊόν με νόημα. Η ίδια αρχή ισχύει γενικά για τις κωδικές αλληλουχίες. Έτσι, για παράδειγμα, εάν ένα ή δύο νουκλεοτίδια απαλειφθούν από μια κωδική αλληλουχία ή υπάρχει μια ένθεση ενός ή δύο νουκλεοτιδίων, το αποτέλεσμα είναι να παραχθεί μια **μετατόπιση πλαισίου (frame-shift)** (αλλαγή πλαισίου ανάγνωσης) που θα οδηγήσει σε άνευ νοήματος αλληλουχία.

### ΔΙΑΧΩΡΙΣΜΟΣ ΤΩΝ ΚΩΔΙΚΩΝ ΑΛΛΗΛΟΥΧΙΩΝ ΑΠΟ ΙΝΤΡΟΝΙΑ

Σε επίπεδο DNA, τα ιντρόνια μπορεί να διακόπτονται μια κωδική αλληλουχία σε έναν από τους τρεις τύπους θέσης: σε ένα σημείο ακριβώς μεταξύ δύο κωδικονίων (ένα *ιντρόνιο φάσης 0*), μετά το πρώτο νουκλεοτίδιο ενός κωδικονίου (ένα *ιντρόνιο φάσης 1*) ή μετά το δεύτερο νουκλεοτίδιο ενός κωδικονίου (ένα *ιντρόνιο φάσης 2*).

Ένα εσωτερικό εξόνιο μπορεί να πλαισιώνεται από ιντρόνια της ίδιας φάσης. Σε ένα εξόνιο όπως αυτό, ο αριθμός των νουκλεοτιδίων διαιρείται πάντα ακριβώς με το τρία. Όταν ένα εξόνιο πλαισιώνεται από δύο ιντρόνια διαφορετικής φάσης, το εξόνιο θα έχει έναν αριθμό νουκλεοτιδίων που δεν διαιρείται ακριβώς με το τρία. Αυτό μπορεί να έχει σημαντικές συνέπειες όταν συμβαίνουν απαλοιφές νουκλεοτιδίων εντός των γονιδίων (βλέπε **Εικόνα 2**).

αλληλουχία: THEOLDMANGOTOFFTHEBUSANDSAWTHEBIGREDDOGANDHERPUP

ΠΑ1: THE OLD MAN GOT OFF THE BUS AND SAW THE BIG RED DOG AND HER PUP

ΠΑ2: T HEO LDM ANG OTO FFT HEB USA NDS AWT HEB IGR EDD OGA NDH ERP UP

ΠΑ3: TH EO L DMA NGO TOF FTH EBU SAN DSA WTH EBI GRE DDO GAN DHE RPU P

**Εικόνα 1** Η σημασία της χρήσης του σωστού μεταφραστικού πλαισίου ανάγνωσης. Η αλληλουχία των γραμμάτων στο επάνω μέρος μπορεί να ομαδοποιηθεί σε σύνολα των τριών (κωδικόνια) που έχουν νόημα στην ανάγνωση του πλαισίου 1 (ΠΑ1), αλλά δεν έχουν νόημα όταν χρησιμοποιείται το πλαίσιο ανάγνωσης 2 (ΠΑ2) ή το πλαίσιο ανάγνωσης 3 (ΠΑ3).