

Titre

merci aux organisateurs de m'avoir confié la tâche...

titre pompeux

cours accéléré de biologie moléculaire

aspects biomathématiques et bioinformatiques laissés aux collègues français

Plan de l'Exposé

hétéropolymères biologiques

polypeptides (protéines) : polymères d'acides aminés

acides nucléiques (ADN et ARN) : polymères de nucléotides

expression génique

transcription : passage de l'ADN à l'ARN messenger

traduction : passage de l'ARN messenger aux protéines

réplication de l'information génétique

cycle et division cellulaires

réplication de l'ADN

organisation des génomes eucaryotes

eucaryotes vs procaryotes

eucaryotes

matériel génétique dans un noyau

plusieurs chromosomes linéaire

animaux, plantes, champignons et protistes

procaryotes

pas de noyau; matériel génétique dans le cytoplasme

un seul chromosome circulaire

bactéries

tailles des génomes

éléments répétés

compactage de l'ADN

Hétéropolymères biologiques

Polypeptides

Structure d'un acide aminé

élément de base (monomère) des polypeptides (protéines)

carbone alpha lié à 4 groupes différents

un groupe amine primaire (NH₂)

un groupe acide carboxylique (COOH)

un proton (H)

une chaîne latérale variable (R)

carbone alpha chiral (asymétrique)

paire de stéréoisomères actifs optiquement (énantiomères)

D- ou L- (dans les protéines)

ion dipolaire (zwitterion) en solution aqueuse : NH₃⁺ et COO⁻

amphotérique (à la fois base et acide)

Liaison peptidique

réaction de condensation

entre groupe carboxyle et groupe amine

élimination d'eau
formation d'un groupe amide

Structure d'un polypeptide

dipeptide, tripeptide et polypeptide (protéines)
acides aminés deviennent des résidus
alternance carbone alpha et groupe amide forme le squelette
liaison plane : cis et trans (favorisée)
chaînes latérales déterminent la fonction de la protéine
typiquement de 100 à 1500 résidus dans une chaîne
chaîne peptidique directionnelle
extrémité libre NH₂ (N-terminale) et COOH (C-terminale)
conventionnellement, le sens d'écriture est de N vers C

Diversité des acides aminés

20 acides aminés différents
existe des acides aminés supplémentaires (modifiés)
plusieurs classements possibles; par exemple :
chaînes aliphatiques simples
hydrophobes (sauf glycine ?)
acides aminés aromatiques
hydrophobes
encombrants
absorbent dans les UV (280 nm)
tyrosine : groupe phénol (fonction alcool)
chaînes basiques
hydrophiles
positives
chaînes acides et dérivés amides
hydrophiles
négatives (acides)
chaînes aliphatiques à fonction alcool
hydrophiles
chaînes soufrées
cystéine (groupe sulfhydryle) : hydrophile?
méthionine (liaison thioether) : hydrophobe
proline (amine secondaire)
hydrophobe ?

Interactions entre résidus

liaisons hydrogène (faibles)
entre atomes des groupes amides des liaisons peptidiques
hélice alpha
feuillet bêta (parallèle ou anti-parallèle)
entre atomes des chaînes latérales des résidus
ponts salins
entre chaînes latérales basiques (+) et acides (-)
liaison ionique (électrostatique)

ponts disulfure

entre les groupes sulfhydryles de deux cystéines

réaction d'oxydation

liaison covalente

coordination d'un atome métallique

effet de certaines chaînes latérales

liaison covalente

interactions hydrophobes

entre chaînes latérales hydrophobes

incapables de former des liaisons hydrogène avec l'eau

sinon tendance à diminuer l'entropie du système

agglutination des molécules hydrophobes

protéines-protéines

protéines-lipides

acides nucléiques

liaisons très importantes en biochimie

moins fortes que les précédentes

mais concernant une large surface

donc coopératives et stabilisatrices

Conformation tridimensionnelle

structure primaire : chaîne polypeptidique

structure secondaire

motifs répétés

hélice alpha

feuillet bêta

triple hélice de collagène

liaisons hydrogène entre atomes des liaisons peptidiques

structure tertiaire

interactions entre atomes des chaînes latérales (cf. ci-dessus)

complètement déterminée par la structure primaire

conformation de moindre énergie

mais parfois besoin de chaperonnes

structure quaternaire

combinaison de plusieurs chaînes polypeptidiques

mêmes interactions que structure tertiaire

groupes prosthétiques

fonctionnalités chimiques non fournies par les résidus

exemples : groupes redox, ions métalliques...

apoprotéine (sans son groupe prosthétique)

Formes et fonctions des polypeptides

Formes

globulaires (sphériques)

fibreuse (beaucoup plus longues que larges)

Fonctions

catalyse

- enzymes responsables des réactions biochimiques
- communication
 - hormones et récepteurs hormonaux
- transport et stockage
 - hémoglobine
 - lipoprotéines sanguines
- structure et mouvement
 - collagène (peau, os et cartilage)
 - kératine (cheveux, poils)
 - cytosquelette (actine, myosine)
- nutrition
 - caséine (lait)
 - ovalbumine (oeufs)
 - globulines (graines)
- immunité
 - anticorps
- régulation
 - facteurs de transcription

Acides nucléiques

Structure d'un nucléotide

élément de base (monomère) des acides nucléiques

ADN et ARN

composé de 3 éléments

un sucre à 5 carbones (pentose)

ribose (2 groupes OH libres) : ARN

désoxyribose (1 groupe OH libre) : ADN

didéoxyribose (plus aucun groupe OH libre)

une base azotée sur le carbone 1' (liaison glycosidique)

un, deux ou trois groupes phosphates sur le carbone 5' (ester)

élimination de deux phosphates lors de la polymérisation

Liaison phosphodiester

réaction de condensation

entre les groupes OH du phosphate en 5' et du pentose en 3'

élimination d'eau

formation d'un diester de phosphate

Structure d'un oligonucléotide

dinucléotide, trinucleotides, oligonucléotides...

alternance pentose et phosphate forme le squelette

bases azotées déterminent l'information stockée

jusqu'à des centaines de millions de nucléotides polymérisés

chaîne nucléotidique directionnelle

extrémité libre 5'-OH (avec ou sans phosphate) et 3'-OH

conventionnellement, le sens d'écriture est de 5' vers 3'

à pH neutre, chaque groupe phosphate a une charge négative

anions d'acides forts très chargés négativement

Diversité des bases azotées

hétérocycles aromatiques azotés

2 familles

purines

bicycliques

adénine, guanine

pyrimidines

monocycliques

thymine, uracile, cytosine

thymine étant un méthyl-uracile

distribution des bases

ADN : adénine, guanine, thymine et cytosine

ARN : adénine, guanine, uracile et cytosine

Complémentarité des bases azotées

toujours une purine avec une pyrimidine

thymine (ou uracile) avec adénine

2 liaisons hydrogène

appariement relativement faible

cytosine avec guanine

3 liaisons hydrogène

appariement plus solide

chaque paire a une géométrie plane et des dimensions similaires

hétéropolymère a structure régulière

Conformation 3D de l'ADN

double hélice d'ADN

découverte en 1953 par James Watson et Francis Crick

données cristallographiques de M. Wilkins et R. Franklin

conformation la plus courante

double hélice dextrogyre

2 brins antiparallèles (pentose-phosphate)

bases appariées vers l'intérieur

paires complémentaires

liaisons hydrogène et interactions hydrophobes

dimensions

diamètre : 20 Å

un tour complet : 34 Å (10 pb)

2 sillons

majeur : 22 Å

mineur : 12 Å

Comparaison de l'ADN et de l'ARN

3 différences principales

ARN ribose vs ADN desoxyribose

ARN uracile vs ADN thymine

ARN simple brin (globulaire) vs ADN double brin (hélice)

liaisons intramoléculaires vs intermoléculaires

fonctions

ADN

stockage de l'information génétique

ARN

fonctions beaucoup plus variées

depuis transport de l'information génétique

jusqu'à catalyse (ribozymes)

Expression génique

Dogme central de la génétique

énoncé par Francis Crick en 1958

idées

l'ADN guide la synthèse d'ARN (transcription)

exploitation de la complémentarité des bases

l'ARN guide la synthèse des protéines (traduction)

notions de codons et de code génétique

les protéines fabriquent les cellules vivantes

[critiqué vers 1970]

l'ADN seul ne peut pas fabriquer une cellule

interdit la reverse transcription de l'ARN en ADN (rétrovirus)

interdit la répllication directe des protéines (prions)

Transcription

Bulle de transcription

action de l'ARN polymérase

déroule et dissocie localement la double hélice d'ADN

brin sens et brin antisens (matrice)

synthétise un ARN messager

complémentaire au brin matrice (copie de l'autre brin)

par incorporation de ribonucléotides triphosphates

progresses sur le brin matrice dans le sens 3'-5'

l'élongation de l'ARN se faisant dans le sens 5'-3'

à la vitesse de 40 nt/sec (E. coli)

nécessité de signaux de démarrage et d'arrêt codés dans l'ADN

Expression d'un gène eucaryote

éléments de contrôle

promoteur

deux (ou plus) motifs consensus attirant la RNA polymérase

éléments amplificateurs (distants, boucles d'ADN)

terminateur

structure secondaire de l'ARN (effet direct ou indirect)

région transcrite (CDS)

éléments codants et non-codants (ORF)

exons

fragments codants du gène

destinés à être traduits en protéines

introns

- éliminés par épissage
- autres éléments non-codants
 - 5'-UTR (leader)
 - portion d'ARN messenger avant le premier codon
 - 3'-UTR (trailer)
 - portion d'ARN messenger après le dernier codon
- autres modifications de l'ARN messenger
 - ajout d'une coiffe de méthyl-guanosine triphosphate
 - action de la guanyltransférase
 - fonction dans la protection, le transport et la traduction
 - clivage 3' et synthèse d'une queue poly-A
 - jusqu'à 250 résidus adénine
 - motif consensus signal
 - action de la poly(A) polymérase
 - fonction de protection de l'ARN messenger
 - édition de l'ARN messenger

Epissage d'un intron

- formation du spliceosome
 - composé de l'ARN messenger et de snRNPs
 - motifs consensus aux deux extrémités de l'intron
 - rapprochement des deux exons à rabouter
- réaction en deux étapes
 - clivage de l'intron en 5' et formation d'un lasso
 - clivage de l'intron en 3' et raboutage des exons
 - dégradation de l'intron
 - transport de l'ARN messenger mature dans le cytoplasme
- épissage alternatif
 - plusieurs protéines à partir d'un seul gène

Traduction

Code génétique

- associe un codon (triplet de nucléotide) à un acide aminé
- 64 codons différents pour 20 acides aminés
- redondance (code dégénéré)
 - détails
 - 2 acides aminés n'ont qu'un seul codon
 - 9 acides aminés ont deux codons
 - 1 acide aminé a trois codons
 - 5 acides aminés ont quatre codons
 - 3 acides aminées ont six codons
 - redondance organisée précisément
 - meilleure résistance aux mutations géniques
 - 3e base : souvent pas d'effet sur l'acide aminé spécifié
 - 1ere ou 2e base : souvent un acide aminé similaire
- codons spéciaux
 - le codon AUG (methionine) est aussi le codon start

3 codons stop arrêtent la traduction (UAA, UAG et UGA)
code universel pour sa plus grande partie
origine unique de la vie
biais dans l'usage des codons selon les organismes

Structure d'un ARN de transfert

molécule adaptatrice associant un codon à un acide aminé
petit ARN (moins de 100 nt)
structure secondaire en forme de trèfle
conformation en forme de L
contient de nombreuses bases modifiées
site d'attachement de l'acide aminé (3')
action d'une enzyme spécifique au couple ARNt-AA
boucle de l'anticodon

Mécanisme de la traduction

se déroule généralement dans le cytoplasme
assurée par le ribosome

complexe ribonucléoprotéique

2 sites

site P (polypeptide)

site A (acide aminé)

processus en 3 étapes

initiation

le ribosome s'associe au codon start d'un ARN messager
un ARNt chargé de méthionine se place au site P

élongation

un second ARNt chargé (complémentaire) se place au site A
catalyse d'une liaison peptidique entre les 2 acides aminés
libération de l'ARNt du site P

glissement 3' de trois nucléotides

le polypeptide en cours d'élongation est au site P

le site A est disponible pour le prochain ARNt

terminaison

un facteur d'arrêt reconnaît le codon stop au site A

dissociation du ribosome

libération du polypeptide

Appariement imparfait

théoriquement, autant d'ARNt que de codons

en pratique, ce n'est pas le cas

wobble hypothesis

suggérée par Francis Crick en 1966

globalement vérifiée depuis lors

première base de l'anticodon tolère des appariements imparfaits

C ne reconnaît que G

A n'existe pas en tant que telle (convertie en I)

U n'existe pas dans les anticodons

G reconnaît effectivement U et C

I reconnaît U, C et A

inosine

adénine déaminée en cétone

2 liaisons hydrogène avec A, C et U

un seul ARNt reconnaît 3 (sur 4) codons Gly (GGU, GGC et GGA)

Structure d'un ribosome

machinerie assurant la traduction de l'ARN messager en protéines

fonctionne généralement dans le cytoplasme

complexe ribonucléoprotéique composé de deux sous-unités

grande sous-unité 60S

3 ARN ribosomiques (28S, 5.8S et 5S)

environ 45 protéines

petite sous-unité 30S

1 ARN ribosomique (18S)

environ 30 protéines

capacité d'auto-assemblage

ARN ribosomiques

molécules universelles

très conservées

structure secondaire complexe

activité enzymatique (ribozymes)

quelques chiffres

25% du poids sec de la cellule (E. coli)

10% des protéines

80% des ARN

collectivement : 10^6 aa/sec (cellules eucaryotes)

Réplication de l'information génétique

Cycle cellulaire

toute cellule eucaryote a un cycle cellulaire bien défini

dure quelques heures et se termine par la division cellulaire

4 phases

G1 : (gap) phase la plus longue durant laquelle la cellule grandit

S : phase durant laquelle l'information génétique est répliquée

G2 : courte pause précédant la division cellulaire

M : division cellulaire (mitose)

ensemble, G1, S et G2 constituent l'interphase

Division cellulaire

interphase

matériel génétique (chromatine) est décondensé

synthèses actives

chromatine invisible au microscope optique

4 phases

prophase

la chromatine se condense pour former les chromosomes

chaque chromosome possède deux chromatides identiques
accrochées par le centromère
effet de la réplication (phase S)
un fuseau de fibres se forme pour guider le mouvement
les chromosomes s'y accrochent
par le biais du kinétochore, voisin du centromère
la membrane nucléaire disparaît

métaphase

les chromosomes se placent au centre de la cellule
plaque équatoriale
effet de la tension du fuseau

anaphase

les chromatides de chaque chromosome se séparent
il y a maintenant 2 jeux de chromosomes identiques

télophase

le fuseau régresse pour disparaître
la membrane nucléaire se reforme
les chromosomes retournent à l'état décondensé
le cytoplasme se divise à son tour (cytokinèse)

Chromosome métaphasique

chromosome

2 chromatides soeurs (identiques, cf. bandes de couleurs)

centromère

fuseau

kinétochore

fibre du fuseau

Fourche de réplication

mécanisme semi-conservatif

chaque brin encode la même information (brins complémentaires)

chaque brin guide la synthèse d'un brin complémentaire

par incorporation de désoxyribonucléotides triphosphates

un brin parental et un brin nouvellement synthétisé

chaque nouvelle hélice sera l'une des deux chromatides soeurs

la double hélice est déroulée et dissociée

action d'une hélicase

les brins monocaténaire sont stabilisés par des protéines

évite la dégradation

évite la renaturation

synthèse du brin leader

similaire à la transcription

l'ADN polymérase III progresse dans le sens 3'-5'

l'élongation du brin complémentaire se faisant dans le sens 5'-3'

synthèse du brin antiparallèle

il est impossible de synthétiser dans le sens 3'-5'

l'astuce consiste à attendre quelque temps (brin retard)

une ARN primase fabrique un petit ARN complémentaire
l'ADN polymérase III fabrique un fragment dans le sens 5'-3'
fragment d'Okazaki (100 à 200 nt pour les eucaryotes)
l'ADN polymérase I remplace l'amorce ARN par de l'ADN
une DNA ligase raboute les fragments d'Okazaki
en fait, le brin leader est aussi initié par une amorce ARN
nécessité d'un signal de départ (origine) codé dans l'ADN

Origines de réplication

les fourches de réplication fonctionnent par paires bidirectionnelles
réplicon (bulle de réplication)

la région de départ est l'origine de réplication

motifs consensus

1 seule dans le chromosome bactérien

chromosome de 4000 kbp

réplication à 900 bp/sec

50000 à 100000 dans une cellule de mammifère

chromosome typique de 100000 kbp

réplication à 50 bp/sec

durant la phase S (eucaryotes)

activation successive de clusters de 20 à 50 réplicons

au début, euchromatine

ADN transcriptionnellement actif

partiellement décondensé

ensuite, hétérochromatine

ADN inactif

très condensé

enfin, centromères et télomères

Organisation des génomes eucaryotes

Paradoxe de la C-value

C-value est la quantité d'ADN par génome haploïde

information génétique contenue dans les cellules reproductrices

la C-value varie énormément d'un organisme à l'autre

même au sein du même groupe

exemples : plantes, insectes, amphibiens

sans relation avec la complexité des organismes

les mammifères ne sont pas ceux qui ont les plus gros génomes

facteurs influençant la C-value

nombre et taille des gènes

à nouveau, nombre de gènes pas vraiment lié à la complexité

homme : 20000 à 25000 gènes (données super-récentes)

similaire à la plante *Arabidopsis thaliana*

25000 gènes

115 Mbp

taille des gènes déterminée par les introns

fréquence et taille des introns

fréquence et taille des éléments répétés

degré de polyploïdie

strictement concerne pas la C-value (dépend le point de vue)

nombre de jeux de chromosomes par cellule

haploïde = 1

diploïde = 2

polyploïde = plus

deux types de polyploïdie

auto-polyploïdie (accident de mitose)

allo-polyploïdie (croisement interspécifique)

éléments répétés

séquences uniques

de une à quelques-unes

essentiellement des gènes

mais pas seulement

séquences modérément répétées

clusters de gènes en tandem

précurseur rRNA : 10 à 10000 en tandem

homme : 5 tandems de 40 copies

tandem des 5 histones : quelques centaines

éléments transposables

quelques centaines à quelques milliers de bp

dispersés

se multiplient et se réintègrent par eux-mêmes (transposition)

exemples

L1 et Alu (300000 à 500000 copies)

jusqu'à 10% du génome humain

séquences hautement répétées

ADN satellite

comportement différent du reste du génome

de 2 à 20-30 bp

répétées des milliers de fois en tandem

rôle probable dans le centromère/kinétochore

application : empreinte génétique

Taille des chromosomes

influencée par

la taille du génome haploïde

le nombre de chromosomes

se calcule à partir de la taille de la double hélice d'ADN

un tour d'hélice = 3,4 nm (34 Å) pour 10 bp

10 bp = 0,34 nm

in vivo, chromosomes métaphasiques beaucoup plus compacts

de 0,002 à 0,02 mm

nécessairement un mécanisme de condensation

Condensation de la chromatine

mécanismes de compactage

double hélice d'ADN : 2 nm

nucléosomes

octamère d'histones

protéines très basiques

4 différentes

146 bp d'ADN

histone H1

+ 20 bp d'ADN

stabilise le tout

chromatosome

linker

55 bp en moyenne (10 à 100 bp)

en tout, environ 200 bp

solénoïde

30 nm

interaction entre les histones H1

fibre de chromatine

200 nm

super-enroulement de solénoïde

protéines linker

chromosome métaphasique

700 nm / chromatide

matrice de protéines