

Biologie Moléculaire et Organismes Modèles

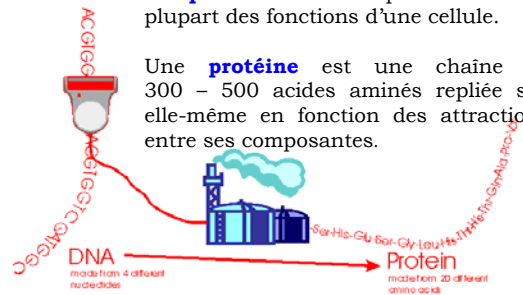
Sami Khuri
Department of Computer Science
San José State University
khuri@cs.sjsu.edu

©2002-08 Sami Khuri

Usine de Protéines

Les **protéines** sont responsables de la plupart des fonctions d'une cellule.

Une **protéine** est une chaîne de 300 - 500 acides aminés repliée sur elle-même en fonction des attractions entre ses composantes.



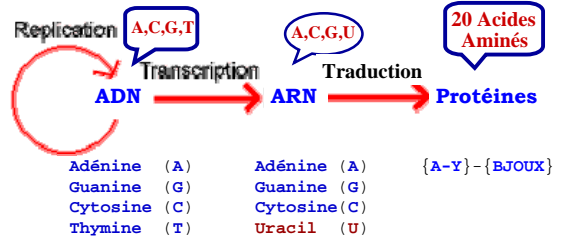
©2002-08 Sami Khuri

Plan du Cours

- Biologie Moléculaire et Organismes Modèles
- Qu'est-ce que la Bioinformatique
- Alignement de deux séquences
- Bases de données
- Alignement multiple de séquences
- Arbres de Phylogénie

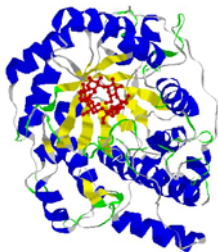
©2002-08 Sami Khuri

Dogme Central: Biologie Moléculaire



©2002-08 Sami Khuri

Biologie Moléculaire



- ADN
- ARN
- Protéines
- Dogme Central
- Transcription
- Traduction

Mis à jour: Le 30 Juin 2008

©2002-08 Sami Khuri

Deux Grandes Catégories de Cellules

Les procaryotes

- Pas de compartiment cellulaire
- L'ADN est dans le cytoplasme de la cellule
- Organismes unicellulaires
- Eubactéries
- Archéobactéries

Les eucaryotes

- Compartiments cellulaires
- Noyau : une membrane de l'ADN
- Organismes uni/pluricellulaires
- Unicellulaires : les levures
- Pluricellulaires : Animaux et végétaux

©2002-08 Sami Khuri

Les Acides Nucléiques

- Deux types d'**acides nucléiques**:
 - **ADN** (acide désoxyribonucléique), support du génome
 - **ARN** (acide ribonucléique)
- Molécule intermédiaire qui permet le passage du gène à la protéine (ARN messager: **ARNm**)
- Molécules impliquées dans la synthèse des protéines (ARN de transfert: **ARNt** et ARN ribosomal: **ARNr**)
- Polymères de nucléotides
 - 4 nucléotides différents: A, T, G, C
 - la séquence en nucléotides représente l'information génétique

©2002-08 Sami Khuri

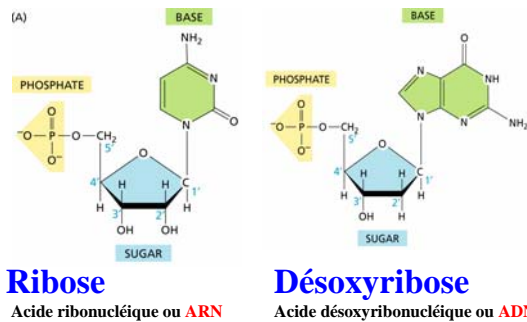
L'ADN

- L'**ADN** est une molécule que l'on trouve au cœur de chacune de nos cellules.
- L'**ADN** est formée de deux chaînes entrelacées; en forme de double hélice.
- L'**ADN** est composée de 4 bases chimiques différentes:

Adénine (A) **Cytosine** (C)
Guanine (G) **Thymine** (T)

©2002-08 Sami Khuri

Les Sucres dans l'ADN et l'ARN



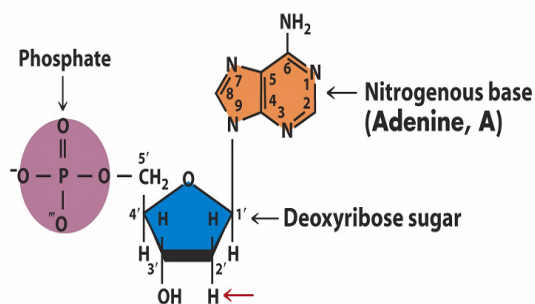
©2002-08 Sami Khuri

La Double Hélice d'ADN

- Les 2 brins s'enroulent en **double hélice**.
- La double hélice est stabilisée par des liaisons hydrogène qui s'établissent entre les bases : les **bases A** forment des paires avec les **bases T**, et les **bases C** avec les **bases G**.
- Les 2 brins d'ADN sont des séquences **inverses complémentaires**.

©2002-08 Sami Khuri

Structure Chimique de l'Adénine

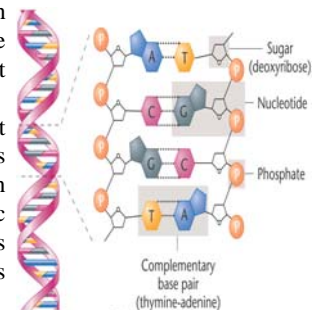


©2002-08 Sami Khuri

Colonnes Vertébrales de l'ADN

La colonne vertébrale d'un **acide nucléique** est une suite alternée de **sucres** et de **phosphates**.

Une molécule d'ADN est formée par deux brins d'ADN qui s'associent en orientation opposée, avec les bases orientées vers l'intérieur et les squelettes sucre-phosphate en surface.



©2002-08 Sami Khuri

Les Gènes



C'est un fragment d'ADN portant les informations nécessaires à la fabrication d'une protéine.

Un **gène** comprend:

- 1) la séquence en nucléotides qui sera transcrite puis traduite en acides aminés
- 2) des séquences permettant de réguler la fabrication de protéines en fonction des conditions cellulaires.

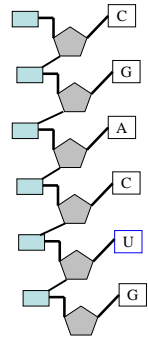
©2002-08 Sami Khuri

Acide Ribonucléique - ARN

- Egalement composé de 4 nucléotides:

- Adénine = Uracile
- Guanine = Cytosine

- Sous forme simple: un brin
- Peut se replier sur lui-même pour acquérir une structure 3D
- Nombreuses fonctions :
 - Synthèse des protéines et leur régulation
 - Génome de certains virus



©2002-08 Sami Khuri

Le Génome

- Le **génom**e est l'ensemble de l'information génétique d'un organisme.
- Une copie du **génom**e est présente dans chacune de ses cellules.
- Le **génom**e est transmis de génération en génération.
- Le **génom**e humain a environ 30,000 gènes.

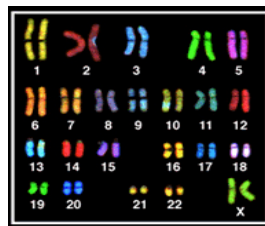
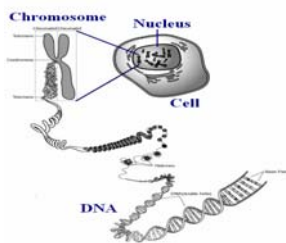
©2002-08 Sami Khuri

Acides Aminés et Protéines

- Il existe 20 **acides aminés** pour la synthèse des **protéines**.
- La structure tridimensionnelle et la composition de la surface déterminent la fonction d'une protéine.
- La formation d'un complexe moléculaire (assemblage de plusieurs molécules) peut être nécessaire à l'acquisition d'une fonction.

©2002-08 Sami Khuri

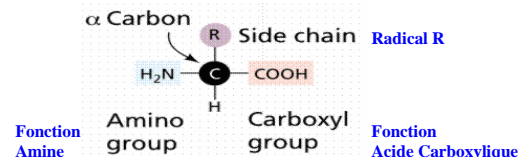
Le Génome et les Chromosomes



Le **génom**e est réparti sur plusieurs molécules d'ADN. Chaque molécule s'appelle un **chromosome**.

©2002-08 Sami Khuri

Structure d'Acides Aminés



Les **acides aminés** ont tous en commun une structure formée par un groupe d'atomes, porté par le même carbone, dans lequel on distingue:

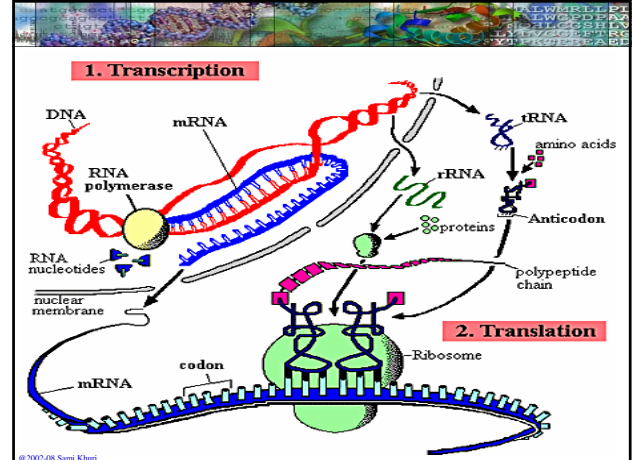
- Une fonction aminée
- Une fonction carboxylique
- Un atome hydrogène
- Un radical R

©2002-08 Sami Khuri

Propriétés Chimiques et Radical R

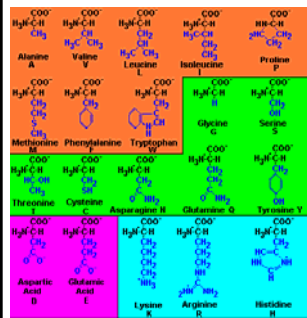
- C'est le radical R qui détermine les propriétés chimiques de chaque acide aminé.
- Les protéines, appelées également polypeptides, sont généralement constituées de 50 à 1000 acides aminés.
- La diversité des protéines résulte des différentes combinaisons linéaires des 20 acides aminés.

©2002-08 Sami Khuri



©2002-08 Sami Khuri

Les Vingt Acides Aminés



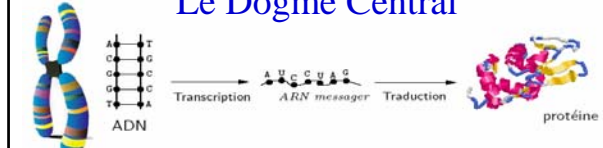
Orange: hydrophobes, c'est-à-dire qu'ils n'absorbent pas l'eau.

Les autres acides aminés sont : hydrophiles.

Magenta: Les radicaux sont acides.
Bleu ciel: Les radicaux sont basiques.

©2002-08 Sami Khuri

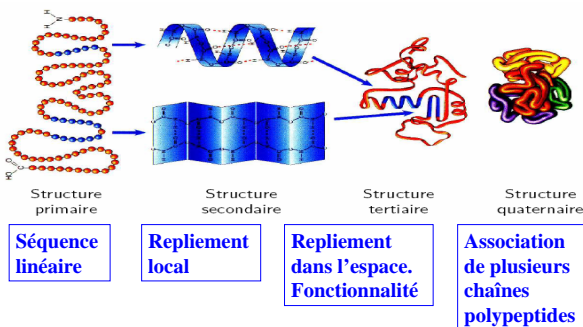
Le Dogme Central



- La **transcription** est la synthèse d'une molécule d'ARN complémentaire (**ARN messenger**) à une séquence d'ADN.
- La **transcription** est initiée par une polymérase ARN.
- Chez les eucaryotes, l'**ARNm** subit une maturation avant la traduction.
- Les **ARNm** sont traduits tels quels chez les procaryotes.

©2002-08 Sami Khuri

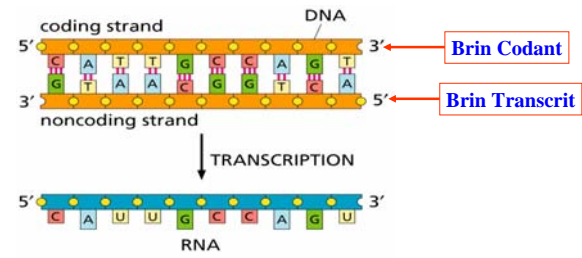
Structures des Protéines



©2002-08 Sami Khuri

Brin Transcrit et Brin Codant

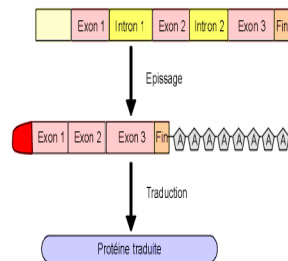
- Un seul brin d'ADN est **transcrit**, c'est à dire sert de modèle à la polymérisation des ribonucléotides.



©2002-08 Sami Khuri

L'épissage des Introns

Les gènes sont d'abord transcrits sous forme d'un ARN pré-messager, lequel sera ensuite capé, polyadénylé, puis les séquences introniques seront éliminées par un mécanisme appelé **épissage** pour générer l'ARNm mature.



©2002-08 Sami Khuri

Le Code Génétique

Première lettre	Deuxième lettre			Troisième lettre							
	U	C	A	G							
U	UUU	Phe	UCU	UUA	Tyr	UGU	Cys	U	Alanine	Ala	A
	UUC	Phe	UCC	UAC	Tyr	UGC	Cys	C	Arginine	Arg	R
	UUA	Leu	UCA	UAA	Stop	UGA	Stop	A	Asparagine	Asn	N
	UUG	Leu	UCG	UAG	Stop	UGG	Trp	G	Acide aspartique	Asp	D
C	CUU	Leu	CCU	CAU	His	CGU	Arg	U	Cystéine	Cys	C
	CUC	Leu	CCC	CAC	His	CGC	Arg	C	Glutamine	Gln	Q
	CUA	Leu	CCA	CAA	Gln	CGA	Arg	A	Acide glutamique	Glu	E
	CUG	Leu	CCG	CAG	Gln	CGG	Arg	G	Glycine	Gly	G
A	AUU	Ile	ACU	AAU	Asn	AGU	Ser	U	Histidine	His	H
	AUC	Ile	ACC	AAC	Asn	AGC	Ser	C	Isoléucine	Ile	I
	AUA	Ile	ACA	AAA	Stop	AGA	Arg	A	Leucine	Leu	L
	AUG	Met	ACG	AAG	Lys	AGG	Arg	G	Lysine	Lys	K
G	GUU	Val	GCU	GAU	Asp	GGU	Gly	U	Méthionine	Met	M
	GUC	Val	GCC	GAC	Asp	GGC	Gly	C	Phénylalanine	Phe	F
	GUA	Val	GCA	GAA	Glu	GGA	Gly	A	Proline	Pro	P
	GUG	Val	GCG	GAG	Glu	GGG	Gly	G	Sérine	Ser	S

©2002-08 Sami Khuri

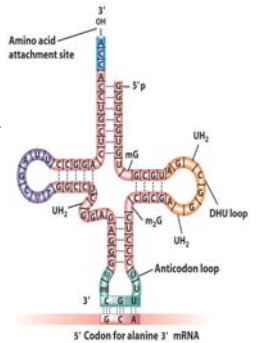
Epissage Alternatif

- Pour un même **gène**, et dans un même organisme, l'élimination des introns peut être différente selon la cellule concernée. Ainsi, pour un même gène, l'**ARNm** sera différent et donnera naissance à une **protéine** différente.
- C'est un phénomène d'économie pour la cellule, omniprésent chez tous les eucaryotes pluricellulaires.

©2002-08 Sami Khuri

La Traduction

- La **traduction** est basée sur le code génétique
- Correspondance entre un codon et son acide aminé par les **ARN de transfert (ARNt)**
- La traduction débute avec le **codon d'initiation (ATG)**
- Fin de la traduction : **codons de terminaison**



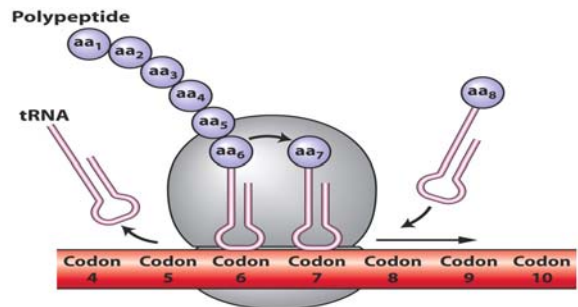
©2002-08 Sami Khuri

Traduction: Synthèse Protéique

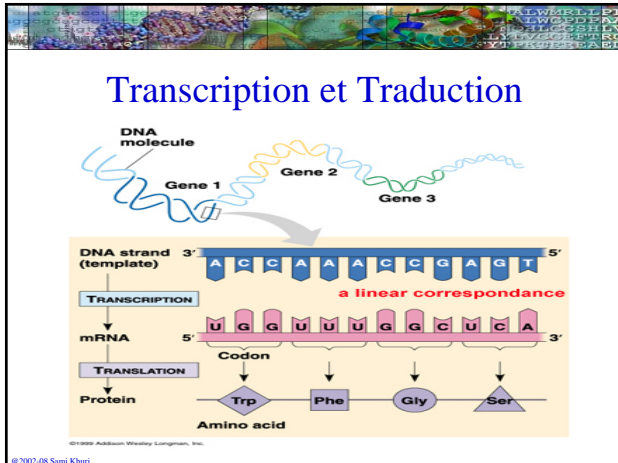
- Processus permettant la synthèse d'une **chaîne polypeptidique (protéine)** à partir d'un brin d'ARN messager.
- L'**ARNm** est traduit en **protéine**
- La **traduction** a lieu au niveau des ribosomes dans le cytoplasme.

©2002-08 Sami Khuri

Une Chaîne d'Acides Aminés

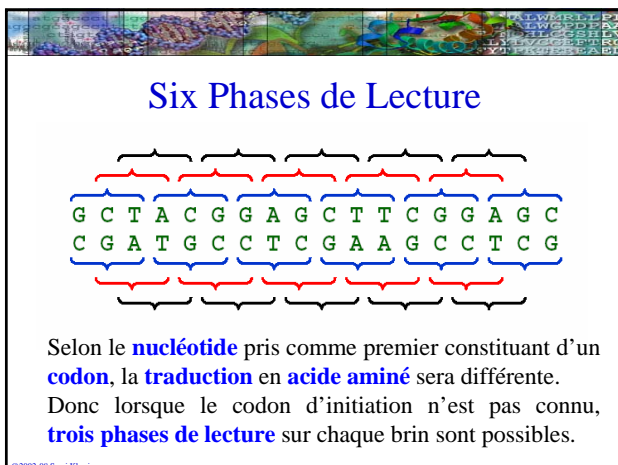


©2002-08 Sami Khuri



- ## Phase Ouverte de Lecture
- **Phase Ouverte de Lecture (Open Reading Frame, ORF):**
 - C'est le **codon d'initiation Met** qui indique le démarrage de la protéine et définit la phase de lecture.
 - Pour que cette phase de lecture soit correcte, il faut aussi trouver à une certaine distance, multiple de trois, un **codon STOP**.
 - Une séquence codante (**CDS**, région traduite en protéine) commence par le **codon initiateur**, et se termine par un des trois **codons STOP**.
- ©2002-08 Sami Khuri

- ## Organismes Modèles
- Un **organisme modèle** est un organisme qu'on étudie en détail pour comprendre un phénomène biologique particulier.
 - **Pourquoi organisme modèle?** C'est un point de départ qui permet d'élucider le fonctionnement d'autres modèles.
 - **Pourquoi est-ce possible?** L'évolution utilise à nouveau les mêmes principes biologiques fondamentaux et conserve les voies métaboliques.
- ©2002-08 Sami Khuri



- ## La Levure de Boulangerie
- La **levure** de boulangerie: *Saccharomyces cerevisiae*
 - Environ 38% de ses protéines ont un équivalent chez les mammifères.
 - Grâce à cet organisme, les biologistes ont découvert les mécanismes fondamentaux du contrôle de la division cellulaire et de la réparation de l'ADN.
-
- ©2002-08 Sami Khuri

Le Ver Nématode

- Le ver nématode:
Caenorhabditis elegans
- Un tiers de ses protéines ressemble à celles des Mammifères
- Par exemple, les biologistes utilisent les vers nématodes dans la recherche des médicaments contre le diabète.



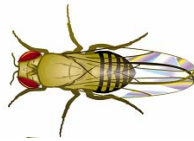
©2002-08 Sami Khuri

Animaux	Plantes
Mammifères	Plantes
Bos taurus (vache)	Arabidopsis thaliana
Canis familiaris (chien)	Avena sativa (avoine)
Felis catus (chat)	Glycine max (soja)
Homo sapiens (homme)	Hordeum vulgare (orge)
Mus musculus (souris)	Lycopersicon esculentum (tomate)
Ovis aries (mouton)	Oryza sativa (riz)
Rattus norvegicus (rat)	Triticum aestivum (blé)
Sus scrofa (porc)	Zea mays (maïs)
Autres Vertébrés	Champignons
Danio rerio (poisson zèbre)	Candida glabrata (responsable des candidoses)
Gallus gallus (poule)	Debaryomyces hansenii
Invertébrés	Encephalitozoon cuniculi
Insectes	Eremothecium gossypii
Anopheles gambiae (moustique, malaria)	Gibberella zeae
Apis mellifera (abeille)	Kluyveromyces lactis
Drosophila melanogaster (mouche du vinaigre)	Magnaporthe oryzae
Nématode	Neurospora crassa
Caenorhabditis elegans	Saccharomyces cerevisiae (levure de boulangerie)
Protistes	Schizosaccharomyces pombe
Plasmodium falciparum (malaria)	Yarrowia lipolytica
Dicryostelium discoideum	

©2002-08 Sami Khuri

La Mouche du Vinaigre

- La mouche du vinaigre
Drosophila melanogaster
- Environ 50% de ses protéines ressemblent à celles des mammifères
- Un grand pourcentage des gènes connus de maladies humaines ont un gène homologue chez cette mouche. Exemple: gène humain p53.



©2002-08 Sami Khuri

Identification d'ADN

- Identifier les suspects dont l'ADN pourrait correspondre à celui trouvé sur les lieux d'un crime.
- Innocenter des personnes accusées à tort.
- Etablir la paternité et autres liens familiaux.
- Trouver un donneur d'organes compatible avec le receveur.

©2002-08 Sami Khuri

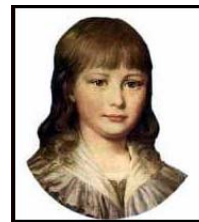
La Souris

- La souris
Mus musculus
- Plus de 90% de ses protéines ressemblent à celles des humains.
- La plupart des médicaments sont testés sur des mammifères: généralement la souris.



©2002-08 Sami Khuri

Louis XVII



Louis XVII: fils de Louis XVI et Marie-Antoinette
Décédé de tuberculose en 1795 à l'âge de 12 ans.

©2002-08 Sami Khuri