

Structure des Acides Nucléiques

Daniel GAUTHERET

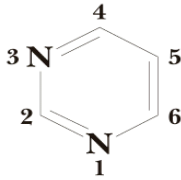
Gautheret@esil.univ-mrs.fr

– Quelques sources:

- Derek Logan (http://falco.biokemi.su.se/~derek/teaching/na_structure/)
- Richard Hallick (http://www.blc.arizona.edu/Molecular_Graphics/DNA_Structure/DNA_Tutorial.HTML)
- Peter de Rijk (<http://www-rrna.uia.ac.be/~peter/doctoraat/struct.html>)
- RNA secondary structures, Robin Gutell (<http://www.rna.icmb.utexas.edu/>)
- The RNA World at IMB Jena (<http://www.imb-jena.de/RNA.html>)
- M. Zuker home page (<http://bioinfo.math.rpi.edu/~zukerm/>)

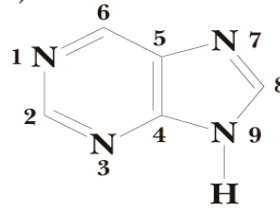
Les bases

(a)

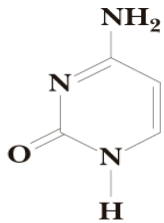


The pyrimidine ring

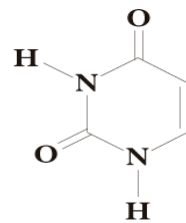
(b)



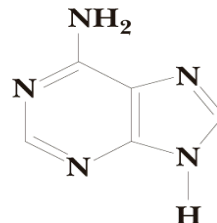
The purine ring system



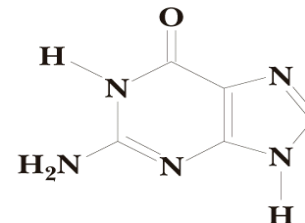
Cytosine
(2-oxy-4-amino pyrimidine)



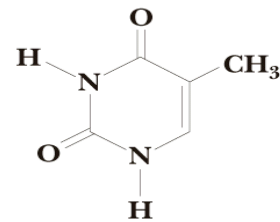
Uracil
(2-oxy-4-oxy pyrimidine)



Adenine
(6-amino purine)

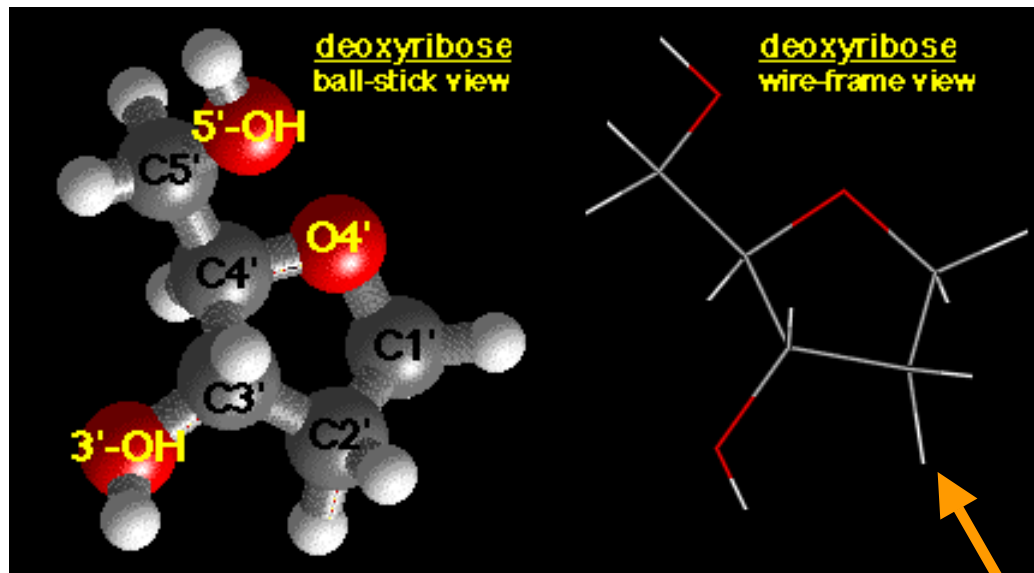


Guanine
(2-amino-6-oxy purine)



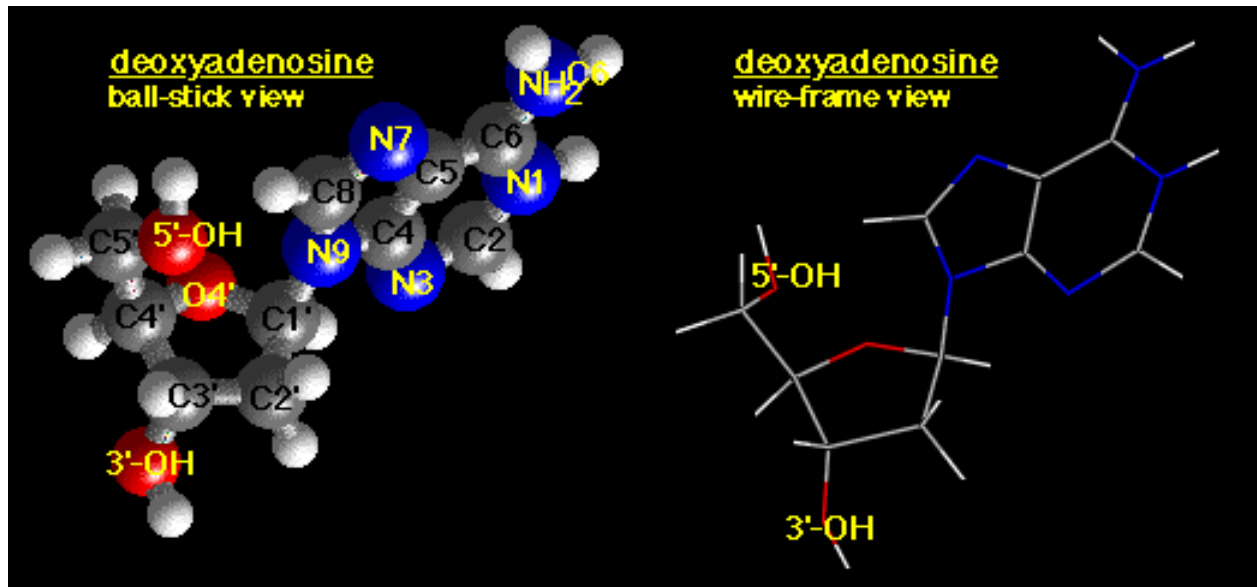
Thymine
(2-oxy-4-oxy
5-methyl pyrimidine)

Le ribose ou déoxyribose

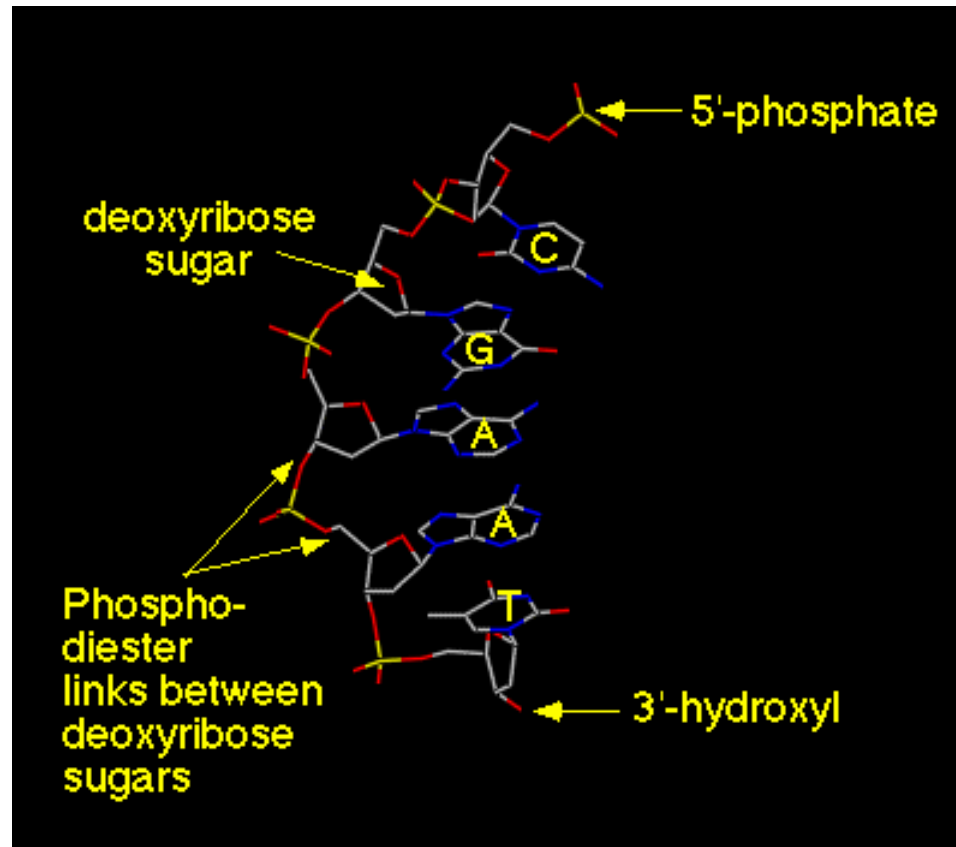


2'OH: ribose

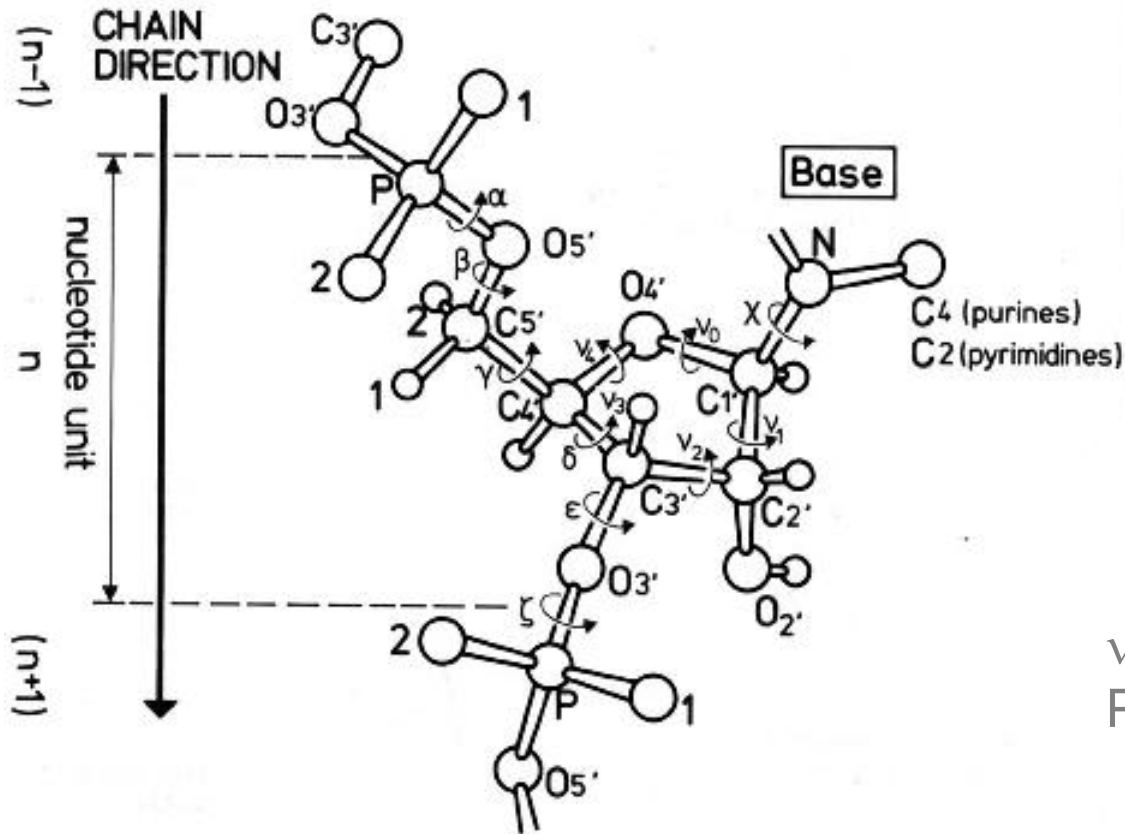
Nucléosides



La chaîne d'ADN/ARN



Les degrés de liberté du nucléotide



ν_0 à ν_4 résumés par:
Phase+amplitude

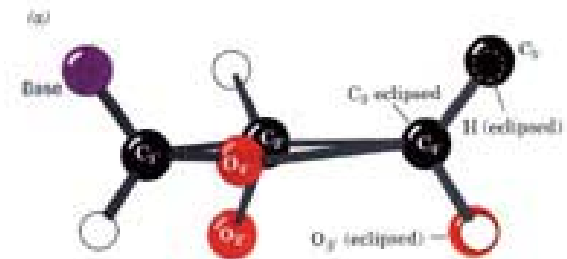
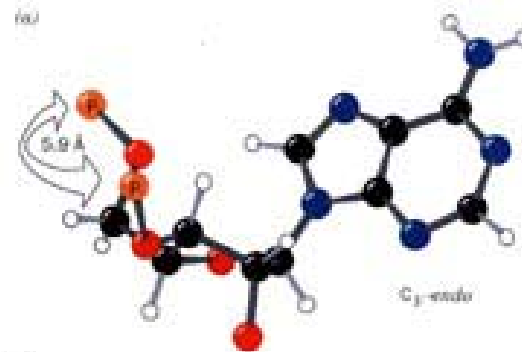
Le plissement du sucre (sugar pucker)

4 des 5 atomes du ribose sont dans un plan, mais le 5ème sort du plan de 0,5 Å environ.

4 conformations majeures: C2': endo, C3'-exo, C3'-endo, C2' -exo.

Endo: coté base, exo: loin base

C3' endo



C2' endo

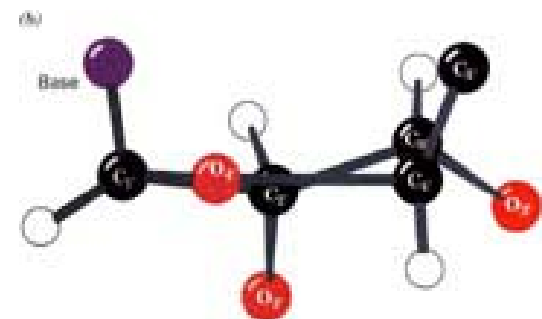
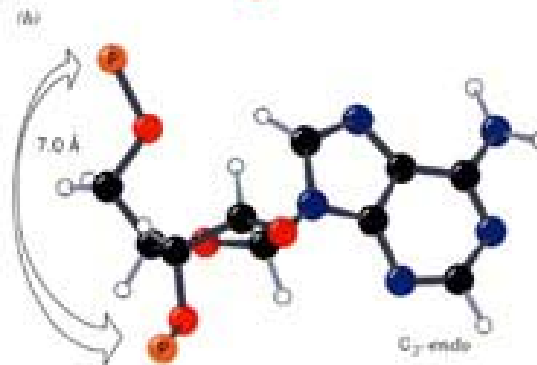
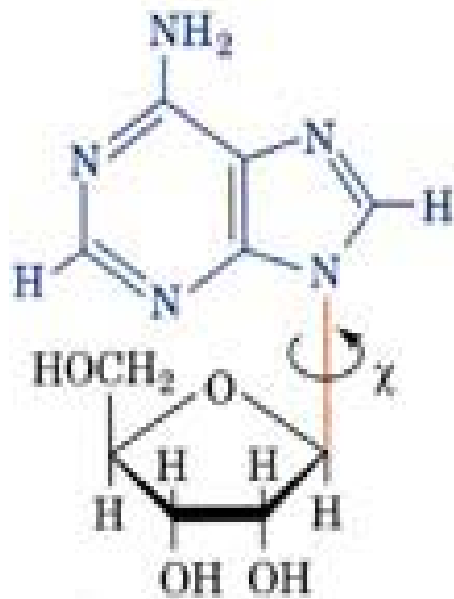


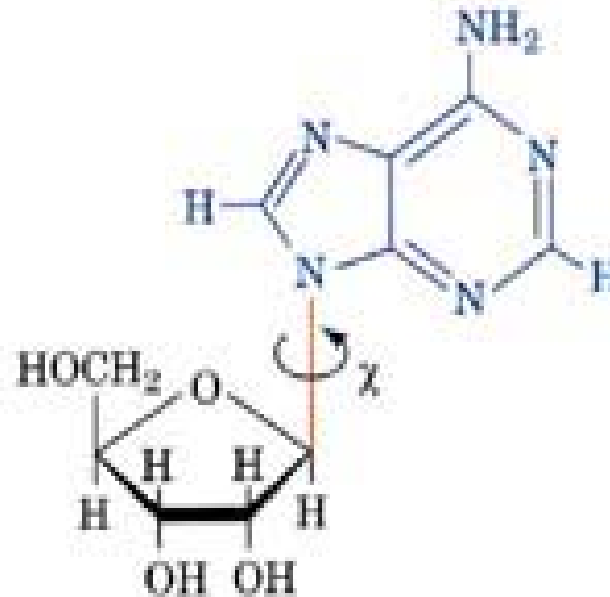
Figure 28-21
Nucleotides in (a) the C(3')-endo conformation [on the same side of the sugar ring as C(5')], and (b) the C(2')-endo conformation which occur, respectively, in A-DNA and B-DNA. The distances between adjacent P atoms in the sugar-phosphate backbone are indicated. [After Saenger, W., *Principles of Nucleic Acid Structure*, p. 237, Springer-Verlag (1983).]

Figure 28-20
The substituents to (a) a planar ribose ring [here viewed down the C(3')—C(4') bond] are all eclipsed. The resulting steric strain is partially relieved by ring puckering such as in (b), a half-chair conformation in which C(3') is the out-of-plane atom.

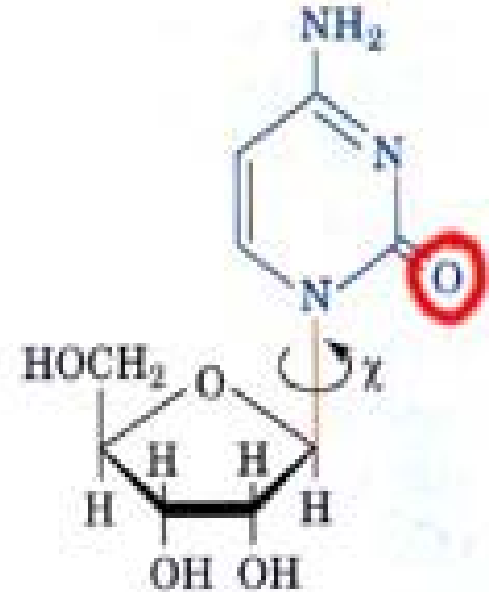
L'orientation de la base



syn Adenosine

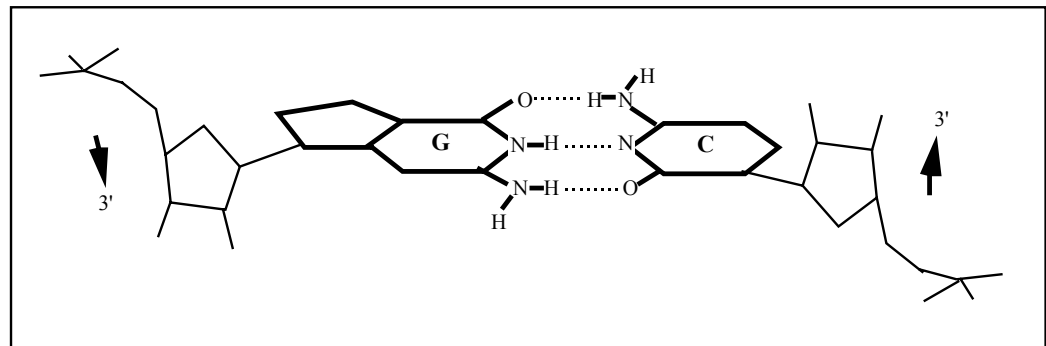


anti Adenosine



anti Cytidine

- Effet sur l'orientation des brins (parallèle ou antiparallèle):



Contraintes sur les angles de torsion

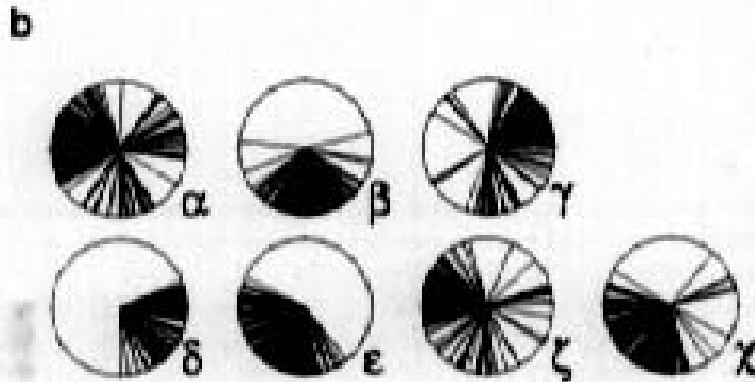
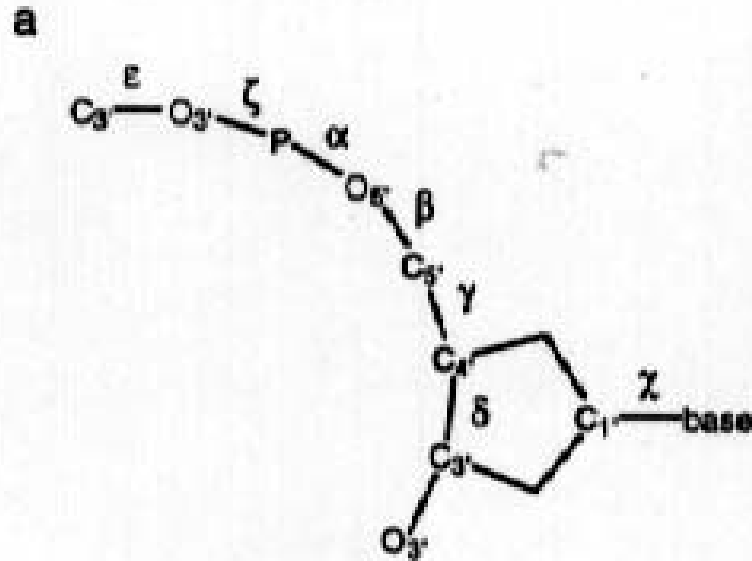
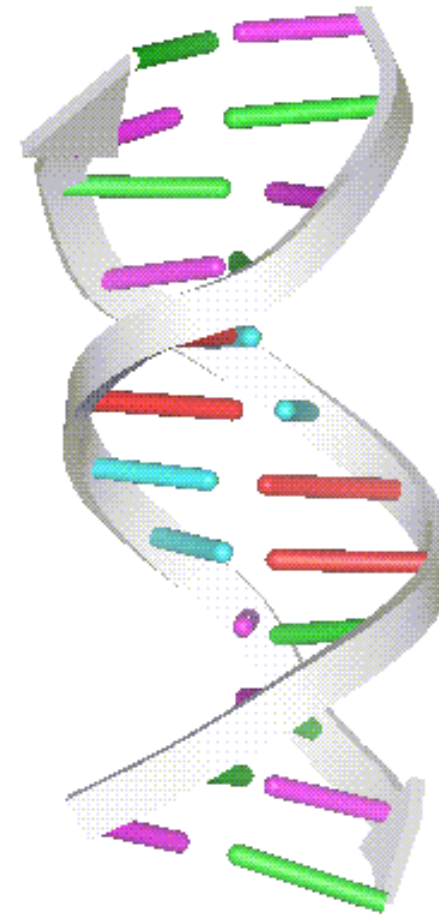
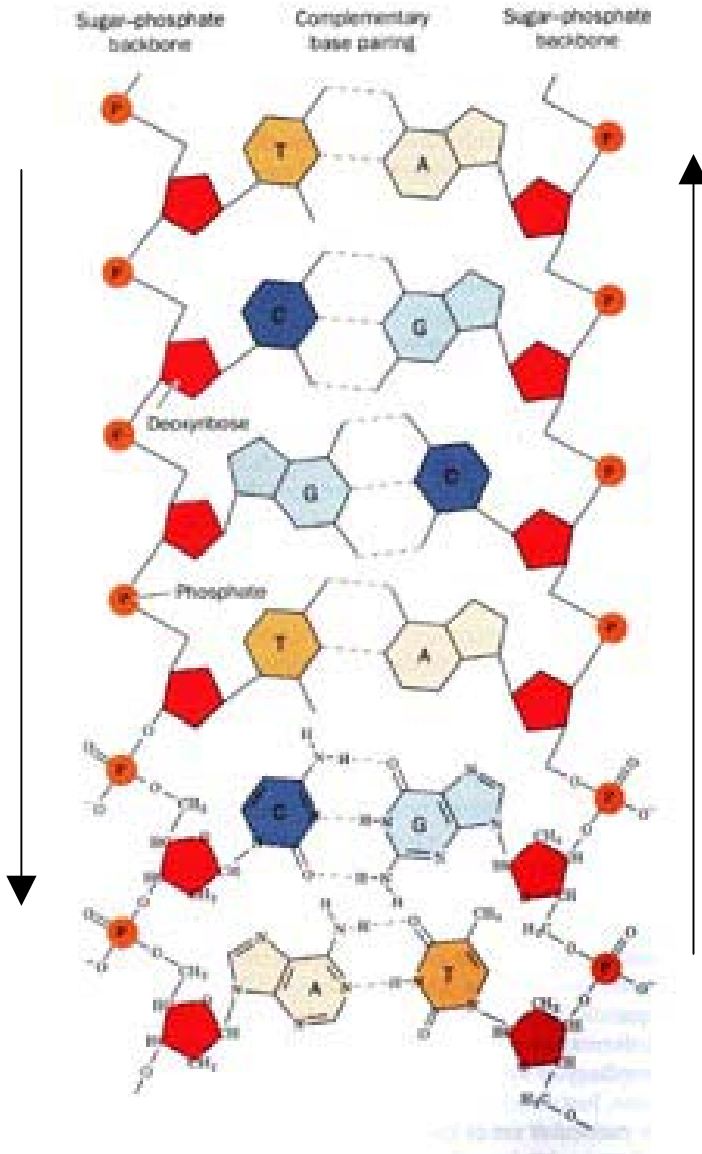


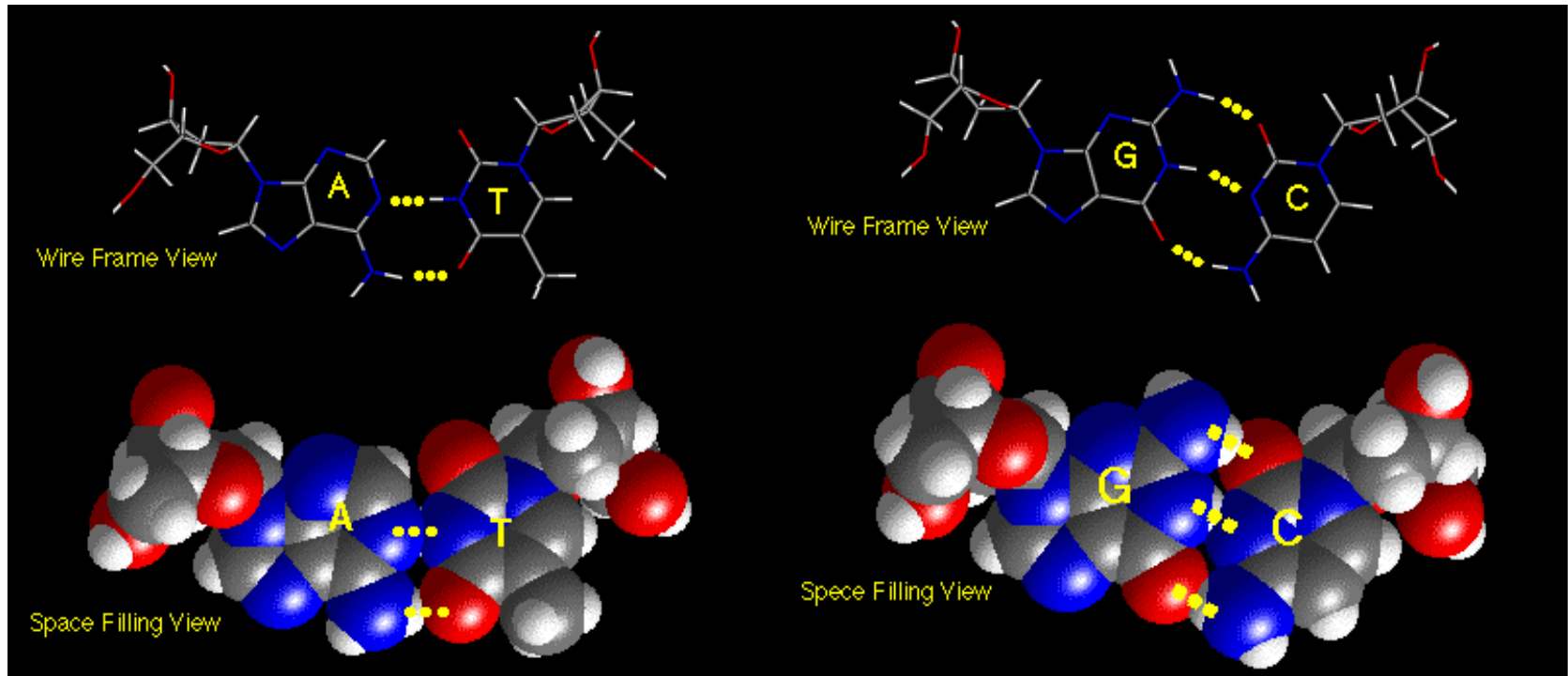
Figure 1. a. Nucleotide torsion angle nomenclature. b. Conformational wheels representing the distribution of each nucleotide torsion angle in the nucleic acid database used here.

– Roues conformationnelles

La double-hélice

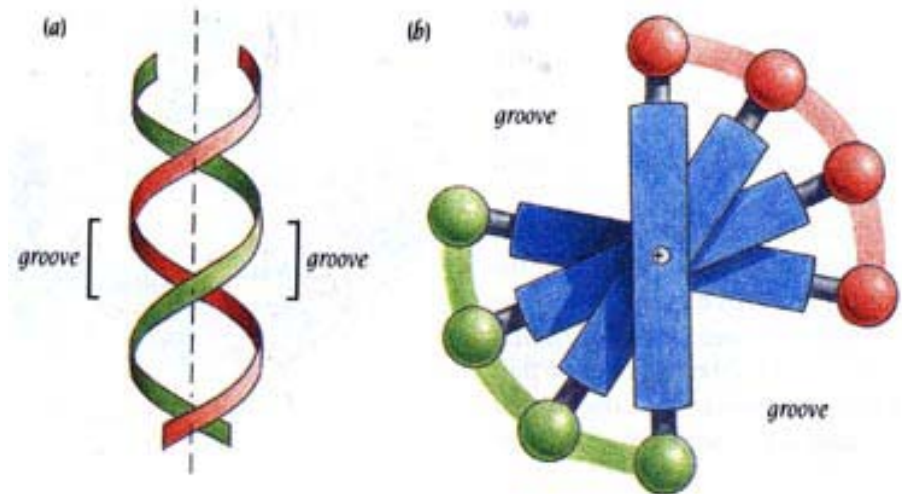
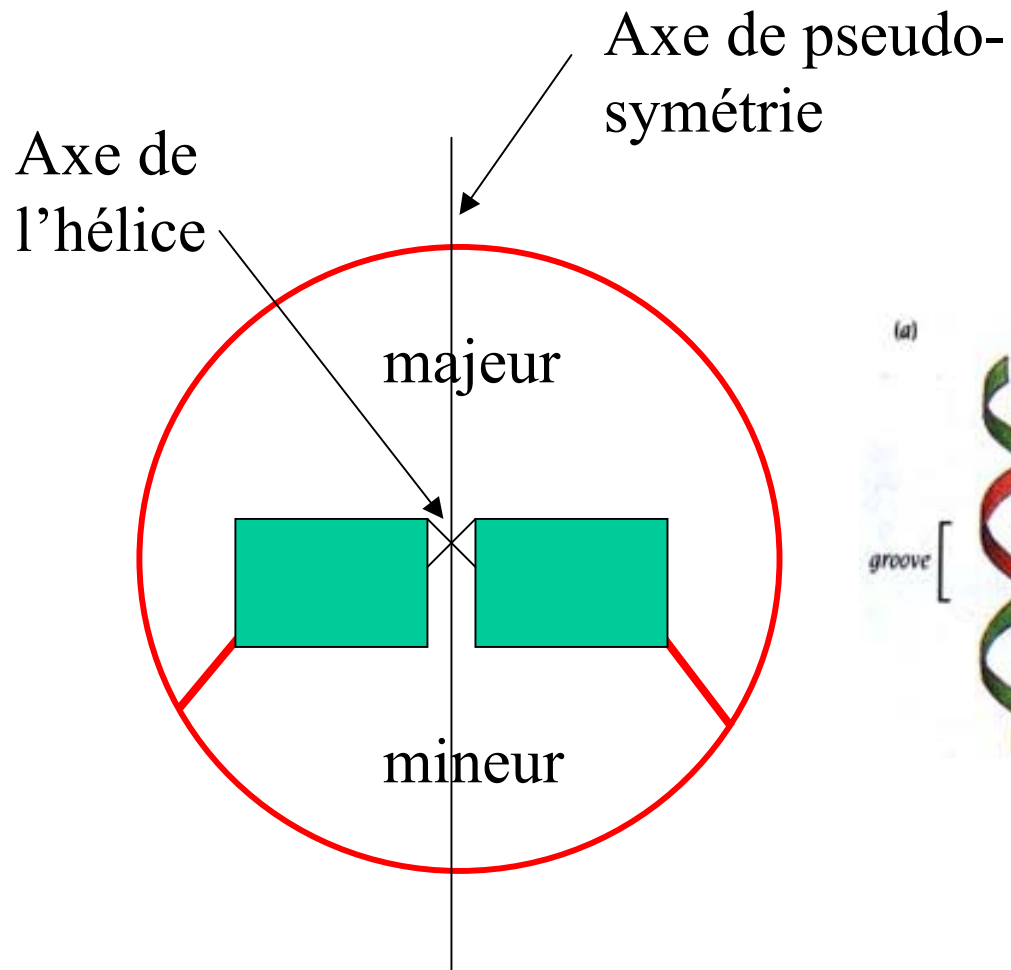


Les paires Watson-Crick



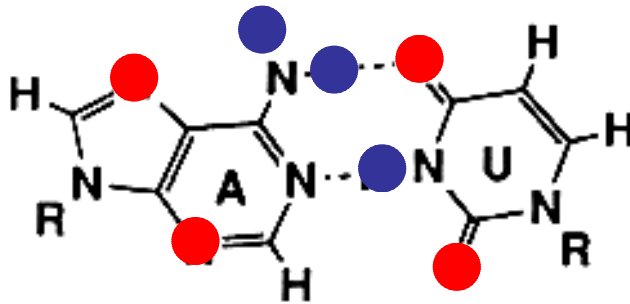
- La distance entre les deux points d'attachement des paires A - T et G - C sont identiques.
- La même géométrie est compatible avec des paires
- A - T et T - A ou G - C et C - G
- Pour les structures hélicoïdales:
- Les quatre bases peuvent être présentes à n'importe quel niveau de la double hélice.

Les sillons

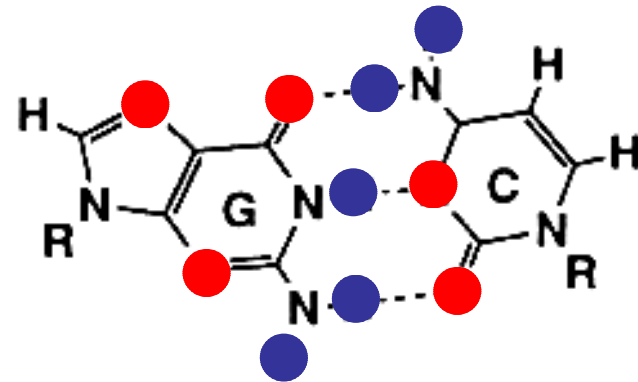


Donneurs et accepteurs de ponts-H: l'identité des paires de bases

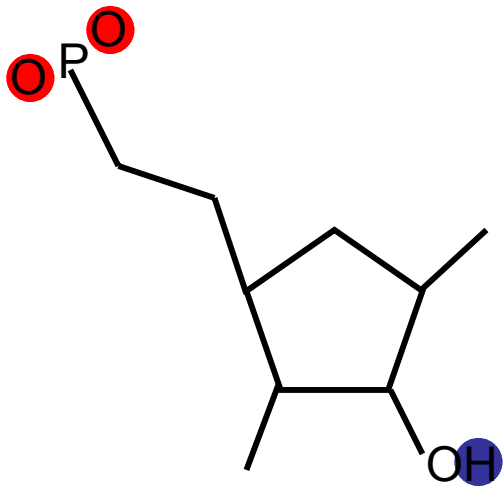
majeur



majeur



mineur



(RNA)

mineur

- donneur
- accepteur

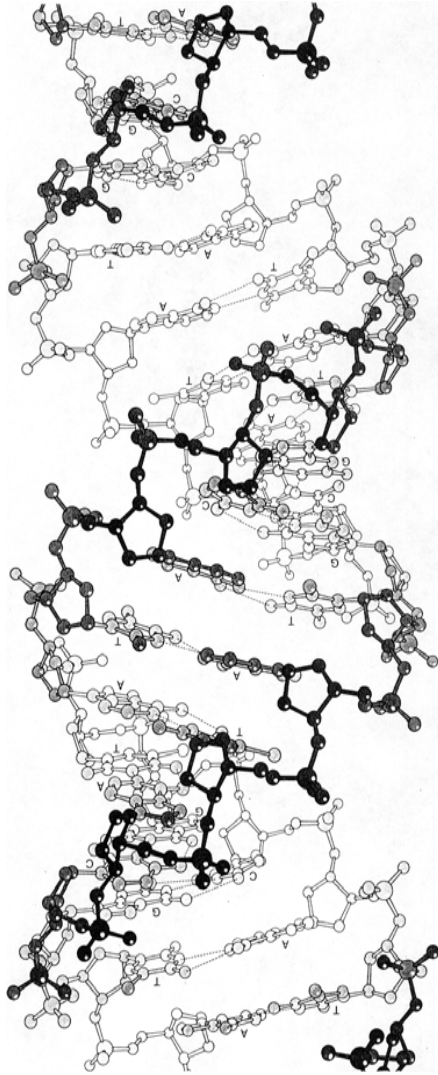
Les types d'hélices

Selon: force ionique,
solvents, degré
d'hydratation.

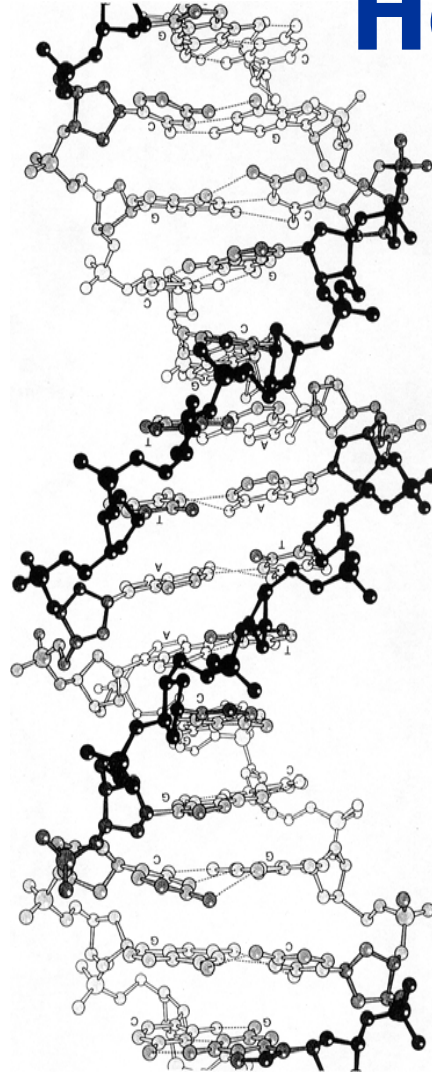
A	B	Z
ADN/ARN	ADN	ADN/ARN

Nt/tour		11	10	12
Sens hélice		droit	droit	gauche
Confo sucre		3' endo	2' endo	2' endo (py) 3' endo (pu)
Diamètre		26 Å	20 Å	18 Å
Liaison glycosidique		anti	anti	Anti (py) Syn (pu)
Déplacement bp/axe		4 Å	aucun	
Sillon majeur	largeur	3 Å	12 Å	plat
	prof	13,5 Å	9 Å	
Sillon mineur	largeur	11 Å	6 Å	étroit
	prof	3 Å	7,5 Å	profond

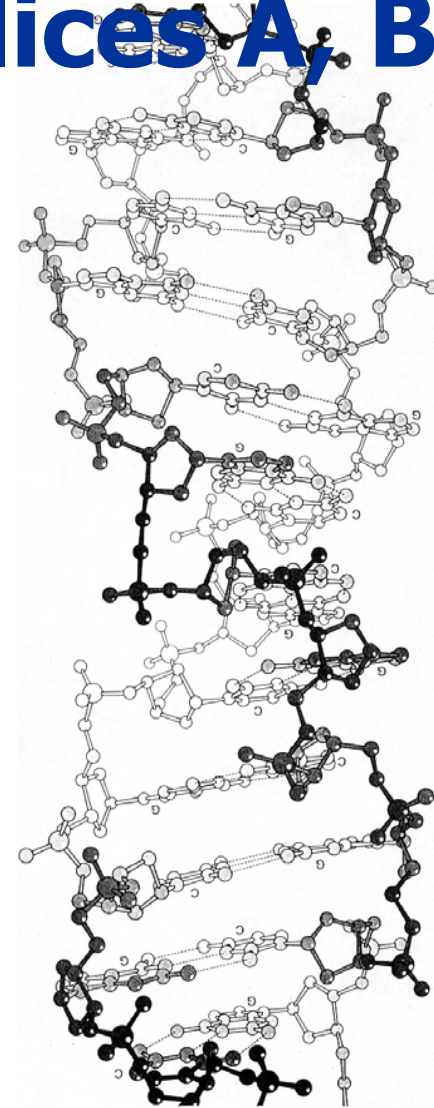
Hélices A, B et Z



A

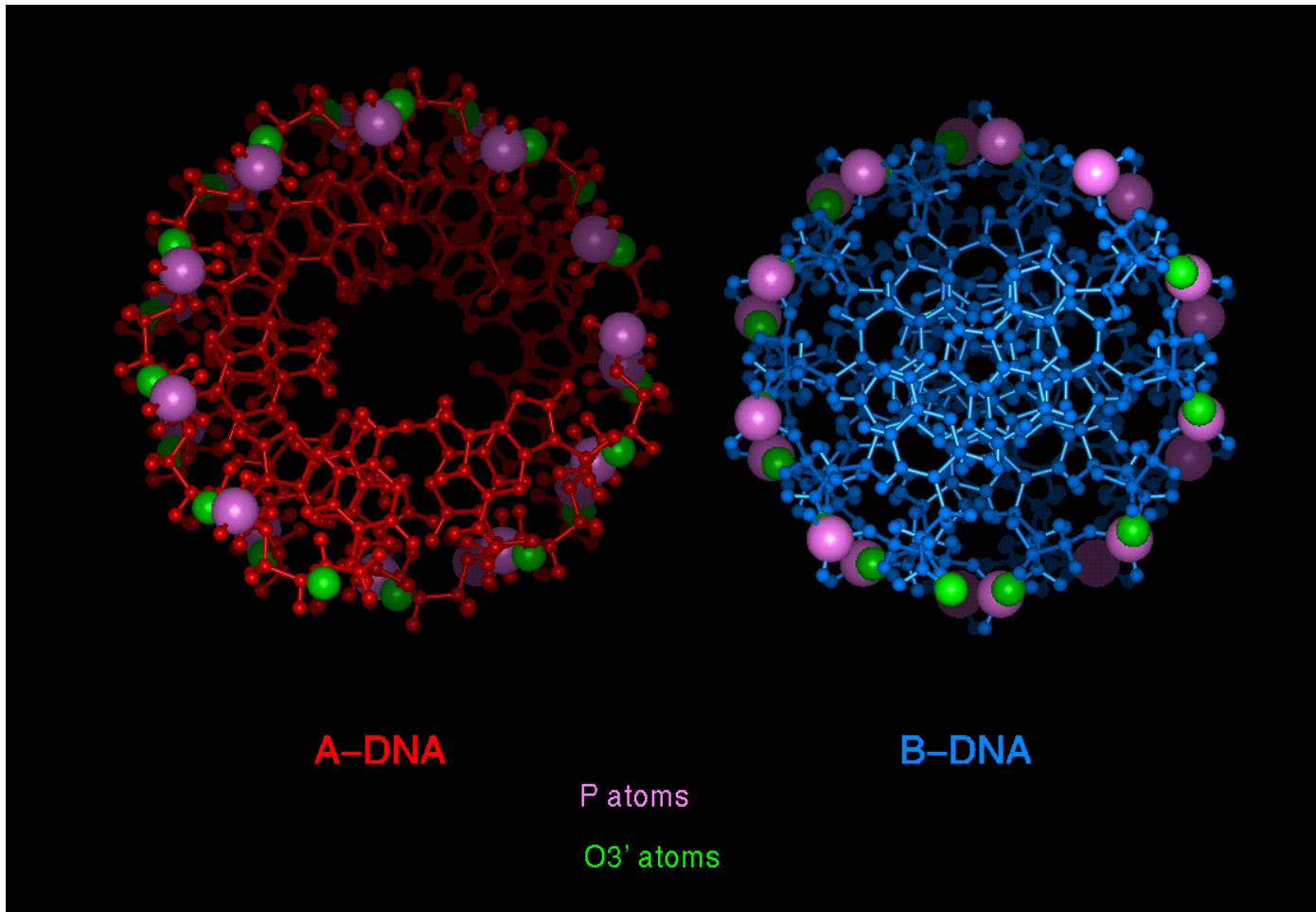


B

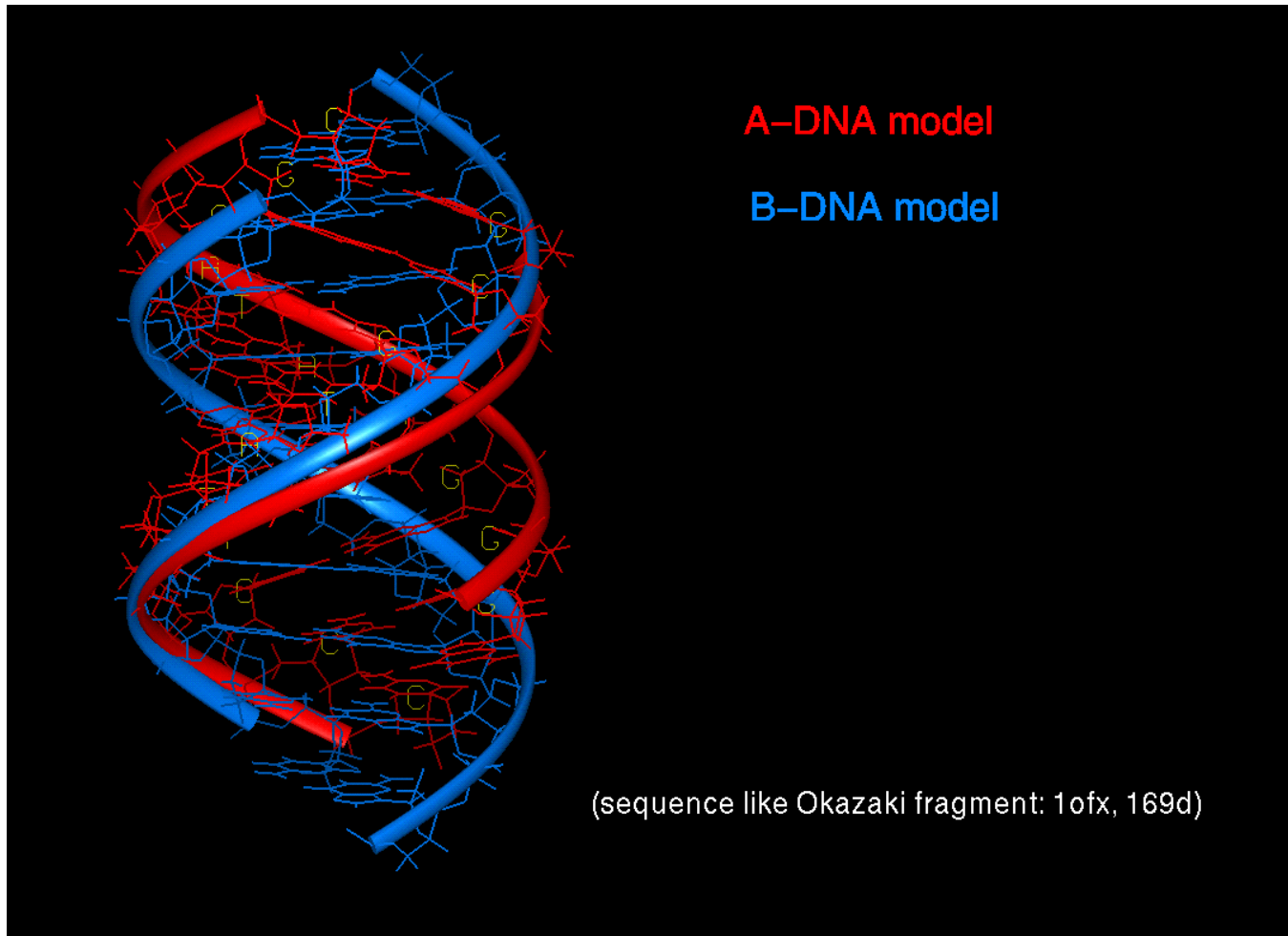


Z

Vues axiales des hélices A et B

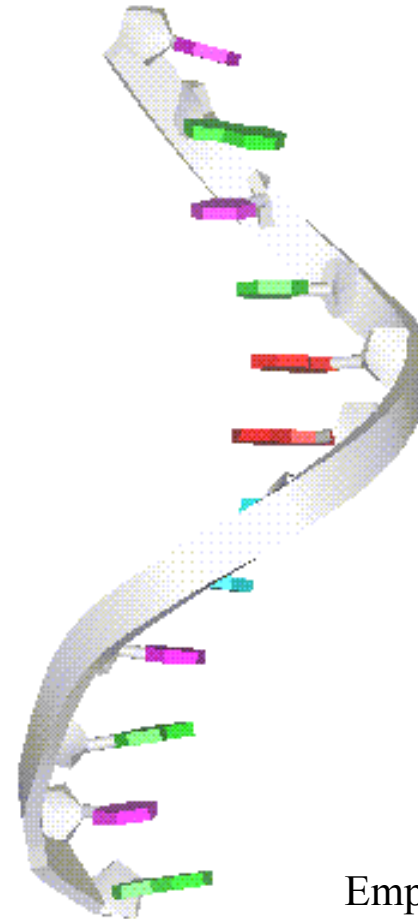


largeur des sillons



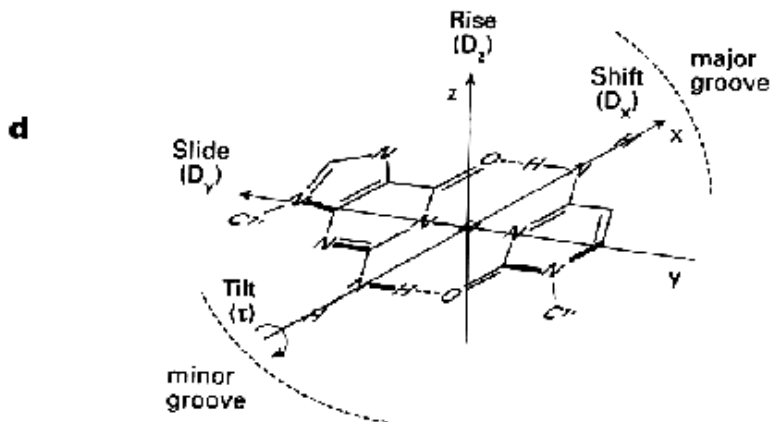
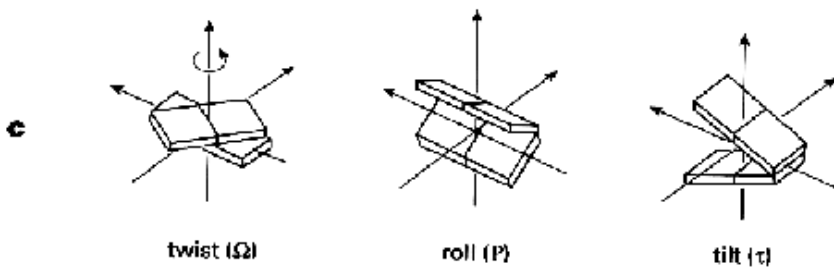
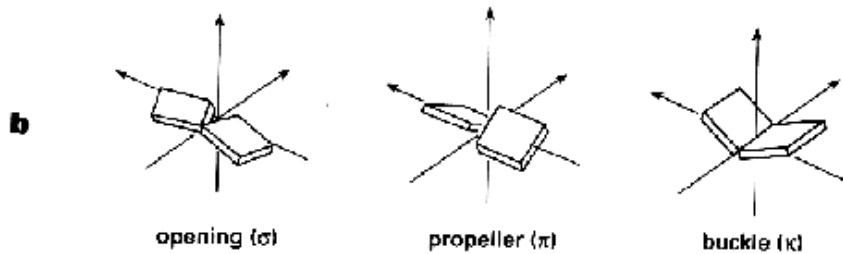
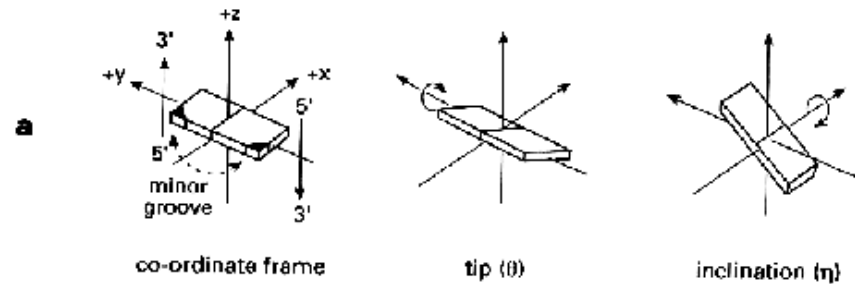
L'empilement des bases (stacking)

- L'empilement n'est pas imposé par la double hélice: il est l'un des principaux facteurs de stabilité des A.N.
- Cycles aromatiques séparés d'environ 4 Å
- Causes: hydrophobicité et interactions VdW



Empilement dans un brin de type B

Paramètres des hélices



On peut définir une double hélice avec les paramètres suivants:

Tilt (θ -t): autour de l'axe pseudo-sym

Twist (t): tour/residu autour axe hélice

propeller Twist (θ -p): entre une base et son appariée

axial rise (h): élévation/residu

Dislocation (D): distance entre centre bp et axe hélice

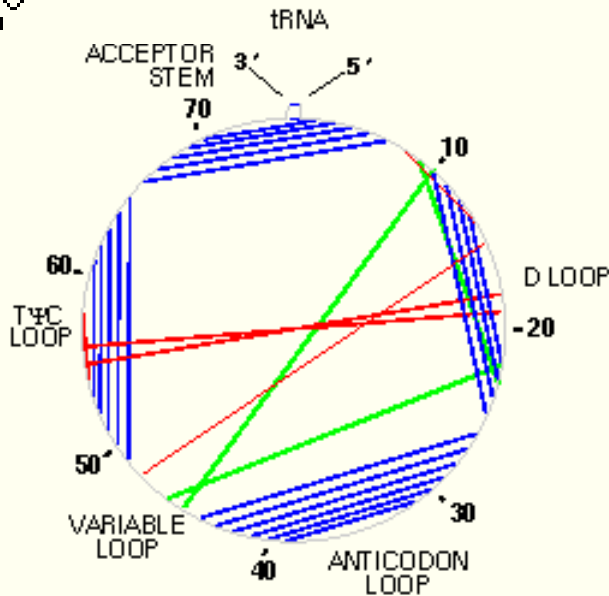
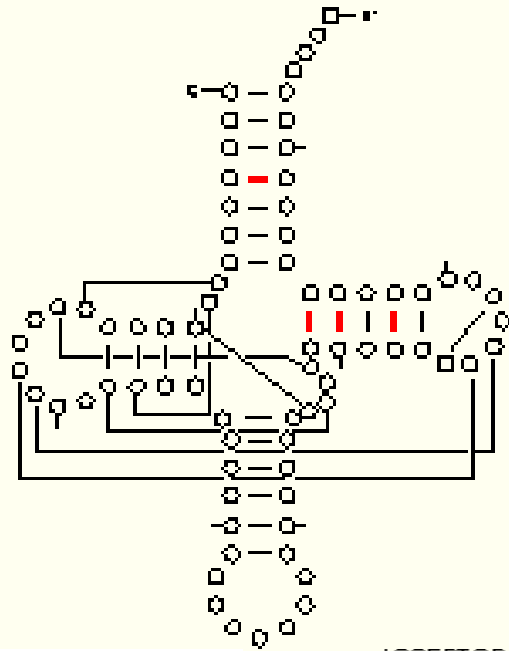
Roll (θ -r): angle autour du 3ème axe (axe C6-C8).

L'ARN

- L'ARNm: très hétérogène
- Les ARNt. Potentiellement 64 différents dans un génomes, en pratique une quarantaine dans les génomes microbiens. Qq centaines dans les génomes de mammifères.
- Les ARNr: 5S, 16S, 23S pour les procaryotes, 5.5S, 18S et 28S pour les eucaryotes. observées en quelques exemplaires chez les procaryotes (7 opérons chez coli), alors que les génomes vertébrés peuvent accueillir plusieurs centaines de copies identiques. (clusters de gènes)
- Les ARNsn (small nuclear RNA) éléments du spliceosome.
- Les ARNsno , guides de méthylation.
- L'ARN viral: longs simple ou double brin
- Les micro-ARN: ARN régulateurs se liant par ex. aux ARNm
- Autres ARN 4.5S, 10Sa, Spot42, DicF, MicF, OxyS, DsrA, 6S (procaryotes) produits des gènes XIST, H19, IPW, 7H4, His-1, NTT (mammifères).

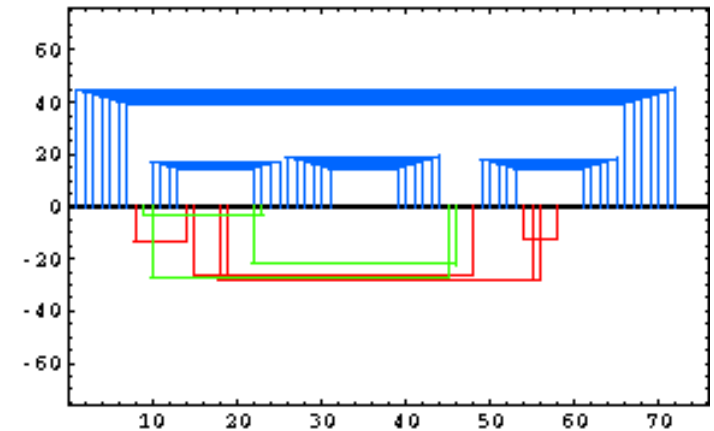
ARNr 18S et 28S: promoteurs pol-I. ARNt et
ARNr 5S: promoteurs pol-III

tRNA: Reddot-Greendot - 85,3% Similarity

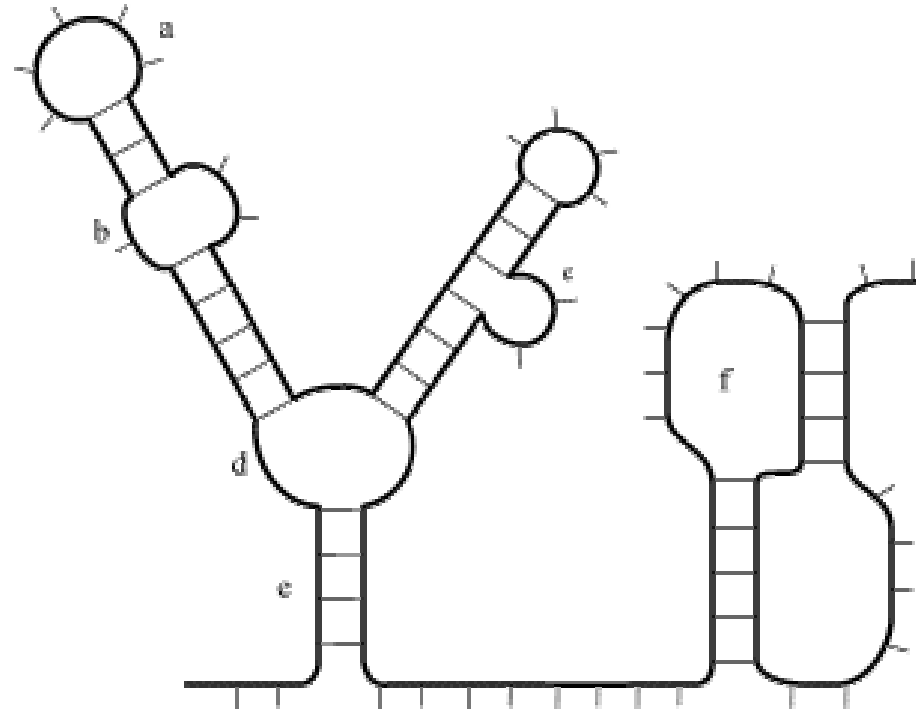


Interactions secondaires et tertiaires

The Structure of tRNA



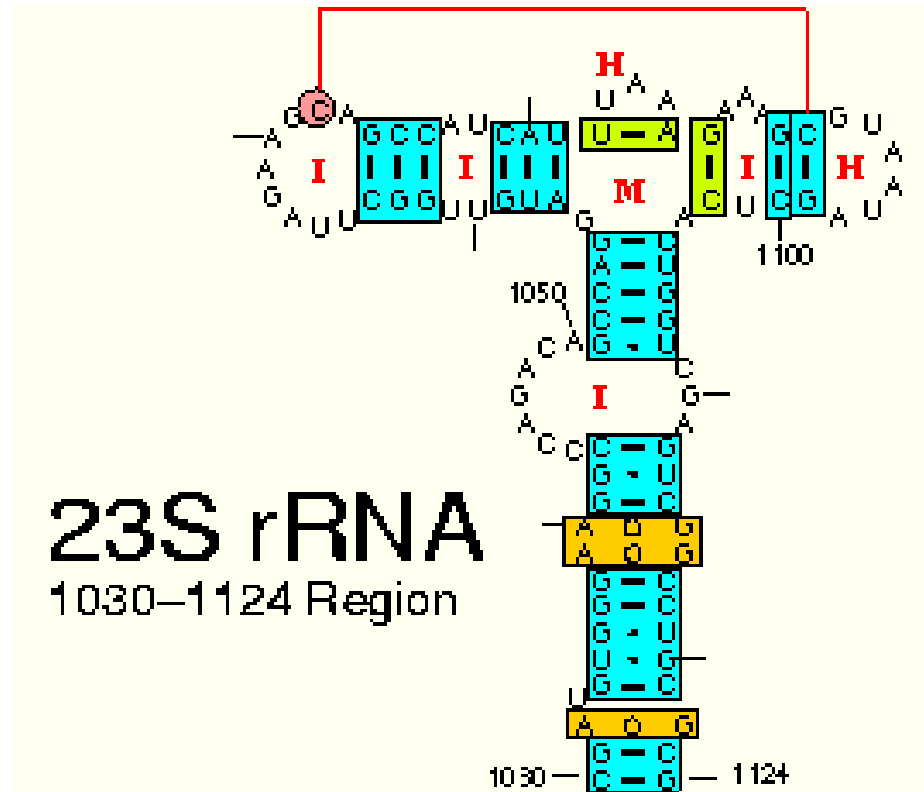
Éléments de structure secondaire



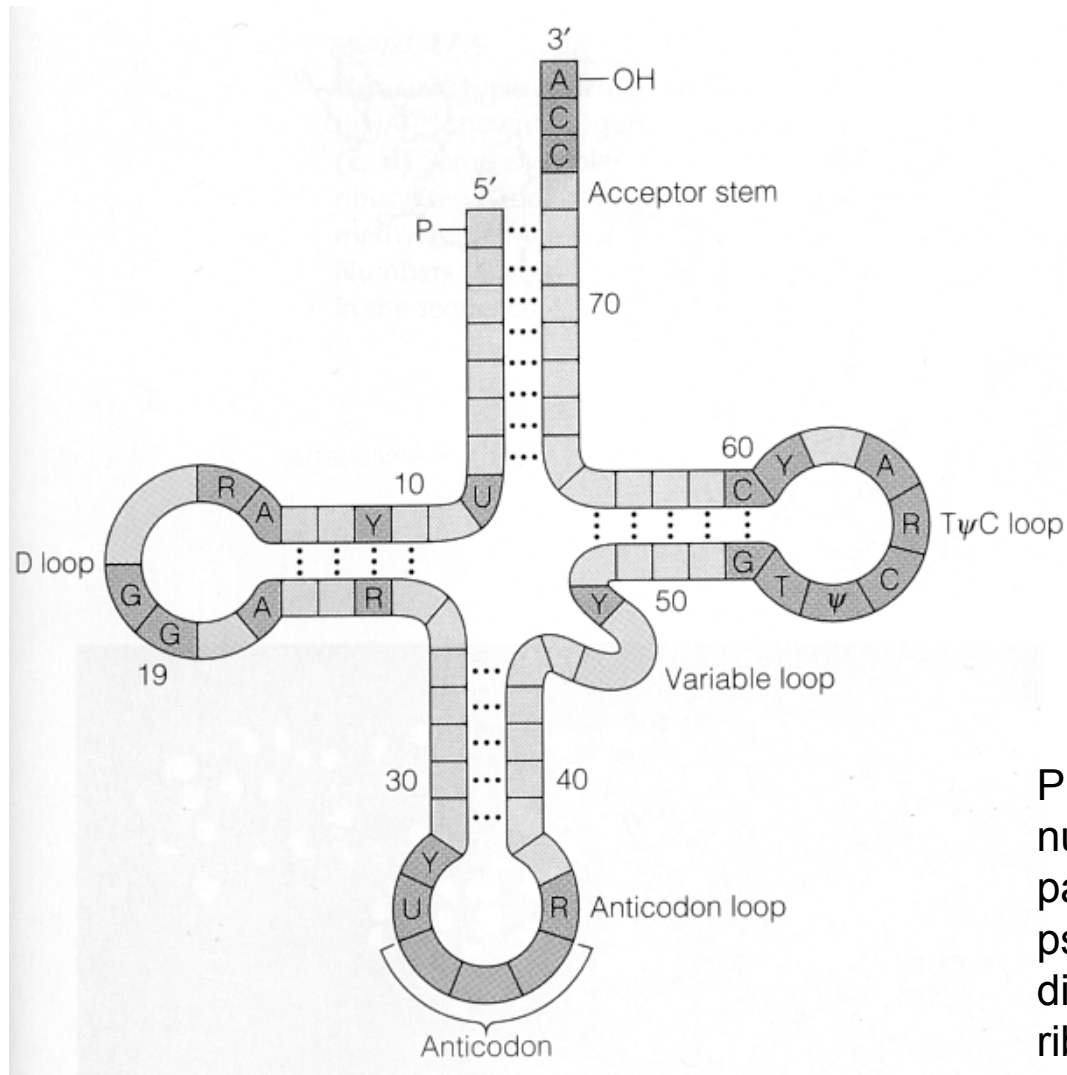
- a. épingle à cheveux (hairpin)
- b. interne
- c. bulge
- d. multi-branche
- e. duplex (longue-distance)
- f. pseudonoeud

Secondaire et tertiaire dans la vie réelle...

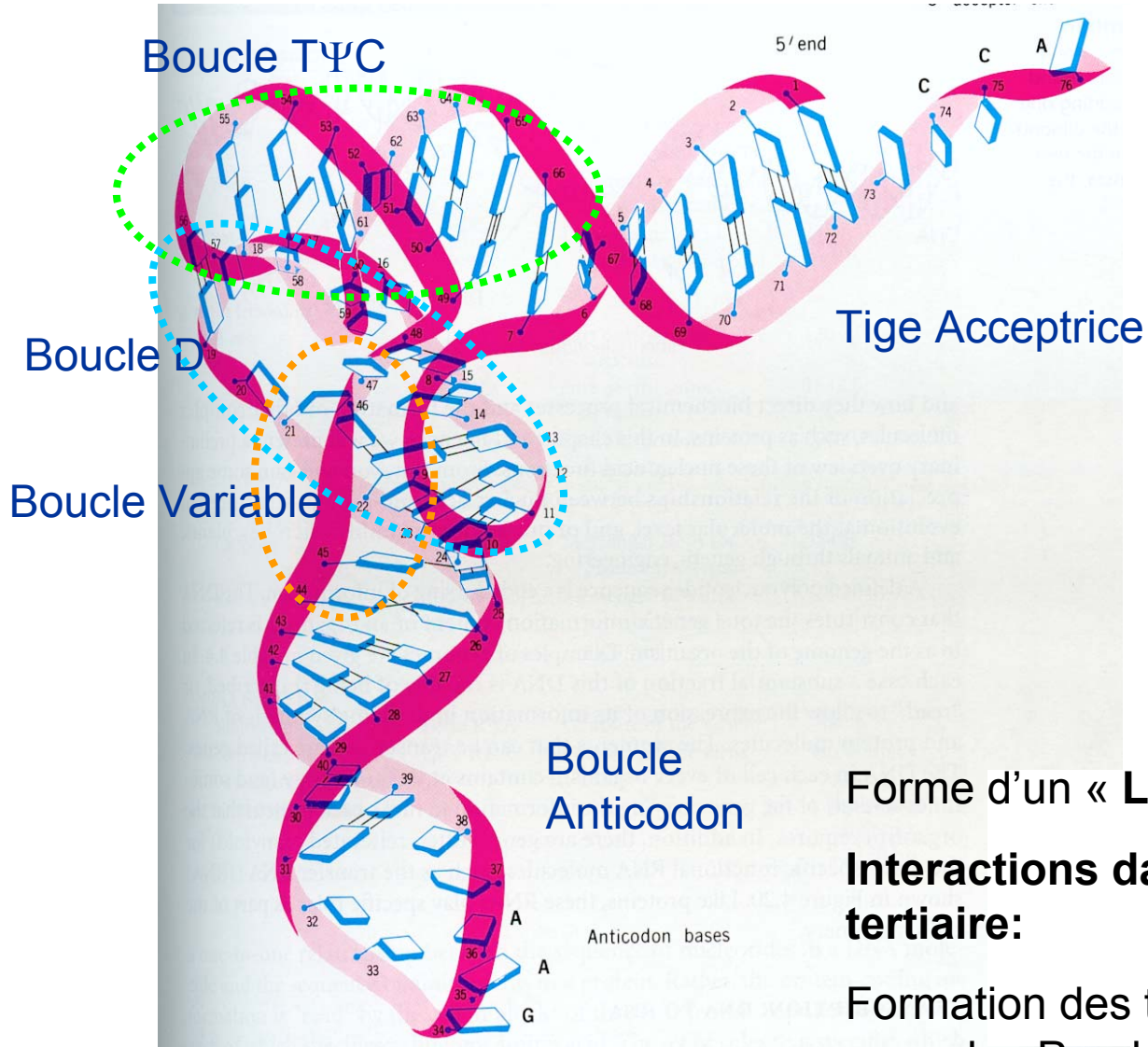
- The 1030-1124 Region of 23S rRNA has several tertiary interactions: two canonical lone base pairs (1082:1086 and 1087:1102), a base triple ((1092:1099):1072) and three non-canonical tertiary base pairs (1032:1122, 1039:1116, and 1040:1115)



L'ARNt



Présence de nombreux nucléosides inhabituels, par ex: inosine (I), pseudouridine (Ψ), dihydrouridine (D), ribothymidine (T), base Y, etc.



Forme d'un « L »

Interactions dans la structure tertiaire:

Formation des triplets; interactions avec des P – du backbone et avec le 2'OH des riboses

ARNt: interactions tertiaires

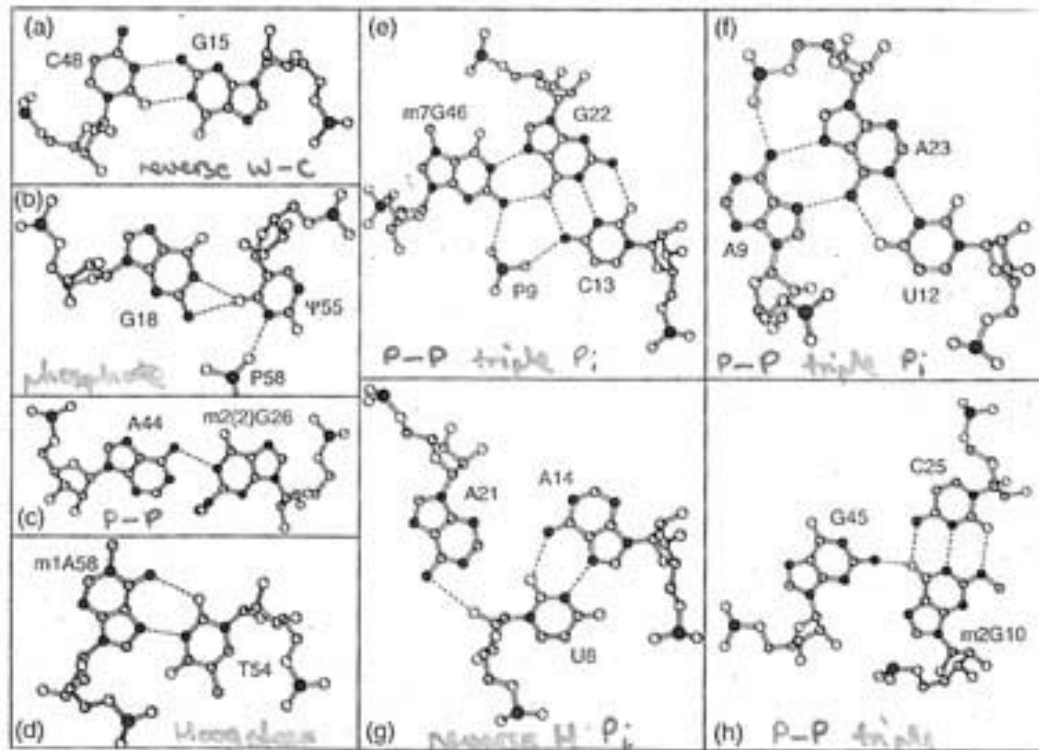
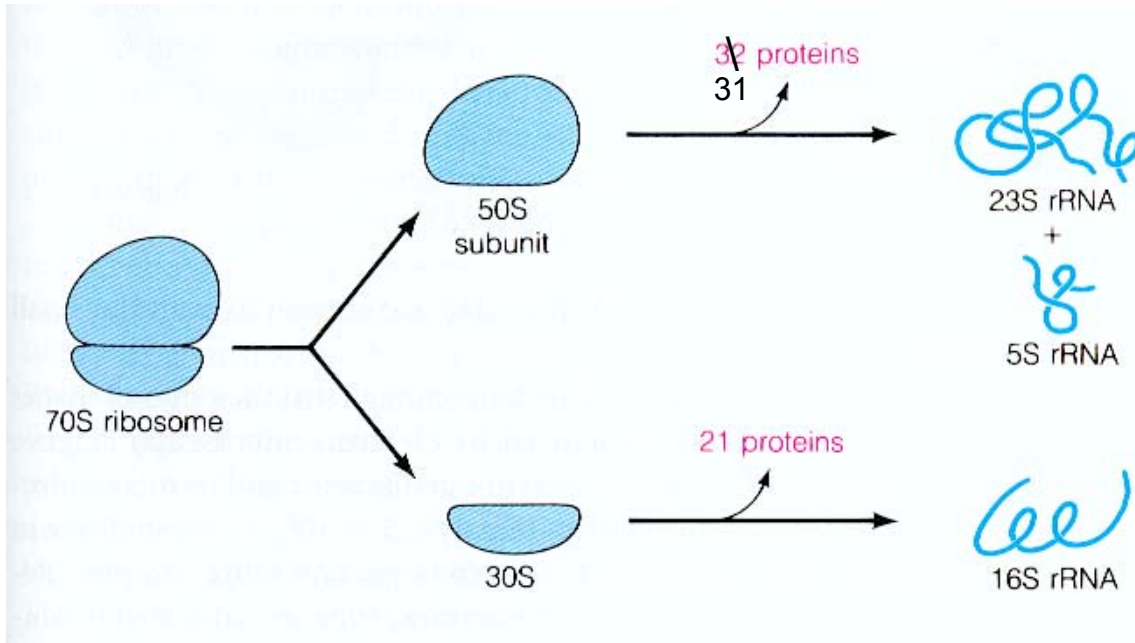


Fig. 2 Tertiary interactions in tRNA^{Phe} (11, 12). The pair G19–C56, a standard Watson–Crick base pair, is not shown. Atoms P–N–C–O are grey-coded in this and subsequent figures with P being the darkest and O the lightest grey.

La structure du ribosome :

Le ribosome



Masse totale: 2.6×10^3 kDa (100 fois plus que le Lysosyme)

Composition: 1/3 protéine 2/3 nucléotides

Sous-unité 30S : Interaction avec les codons du mRNA et les anticodons du tRNA

Sous-unité 50S : activité peptidyl-transferase et interaction avec le GTP-binding protéine.

Meilleures structures aujourd'hui :

Ribosome 70S de *T. thermophilus* à 7.8 Å (1999)

Sous-unité 30S de *T. thermophilus* à 4.5 Å (1999)

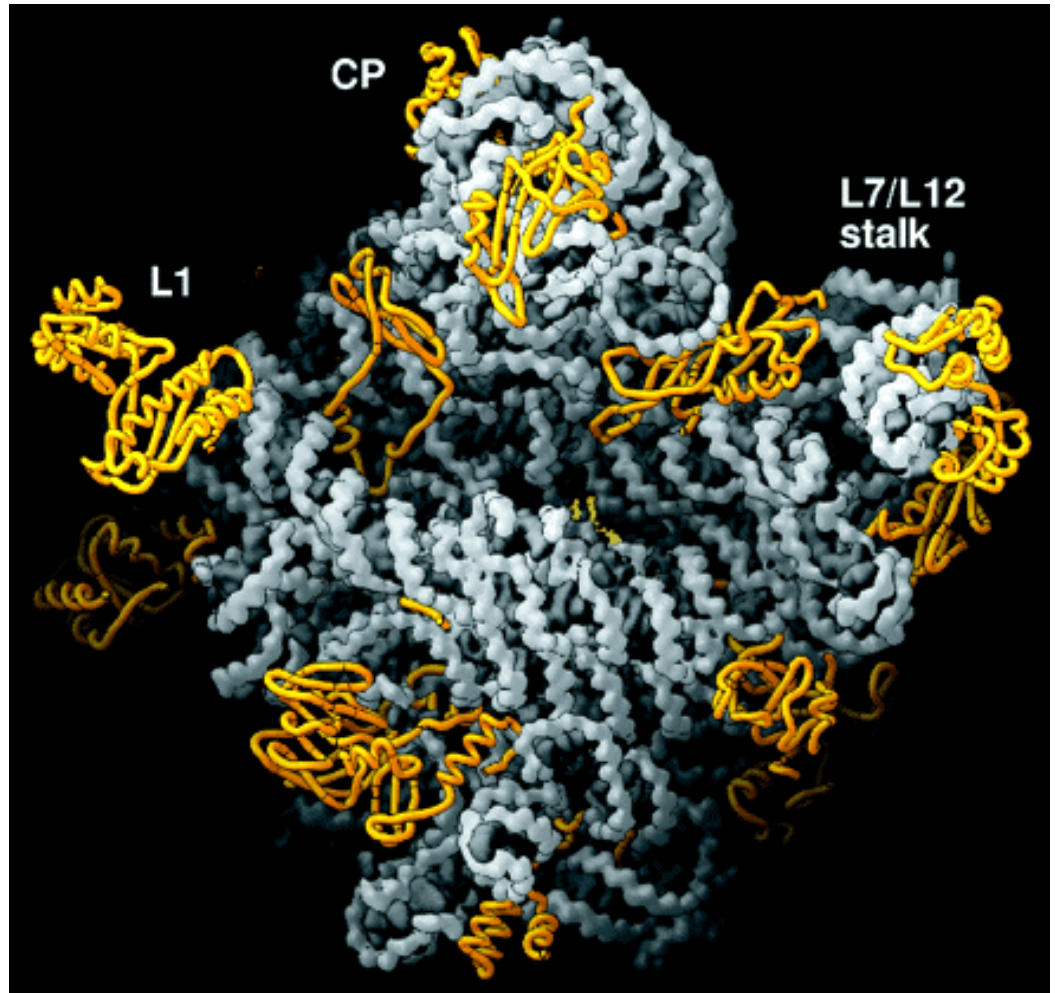
Sous-unité 50S de *H. marsimortui* à 2.4 Å (2000)

La grande sous-unité (ARN 23S)

crystal structure of the
large ribosomal
subunit from
Haloarcula
marismortui at 2.4
angstrom resolution

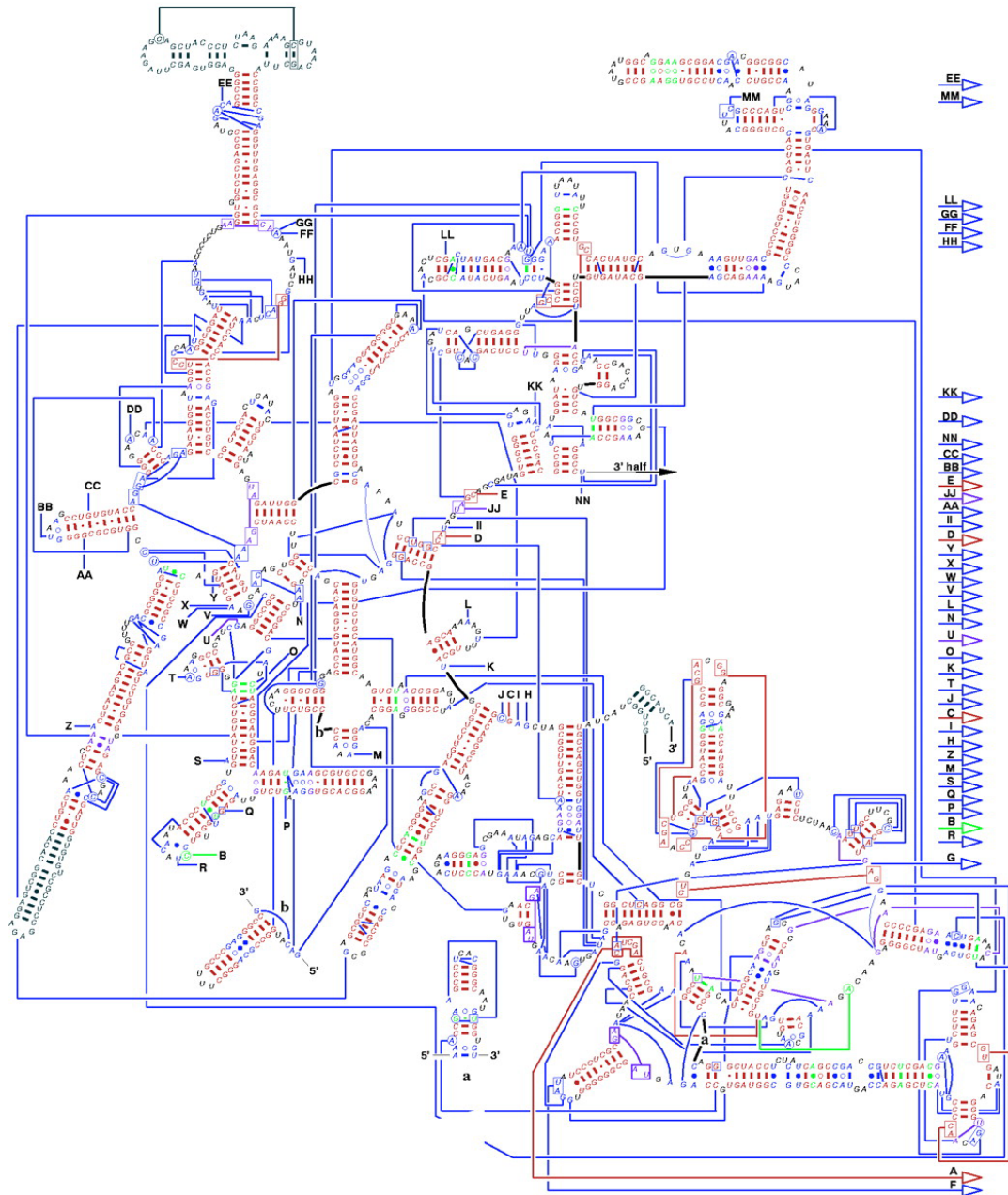
Grande sous Unité: 35
prot + 2 ARN (23S,
5S)

Nenad Ban, Poul Nissen,
Jeffrey Hansen, **Peter B.
Moore, Thomas A. Steitz**
Science. 289:878-9, 2000



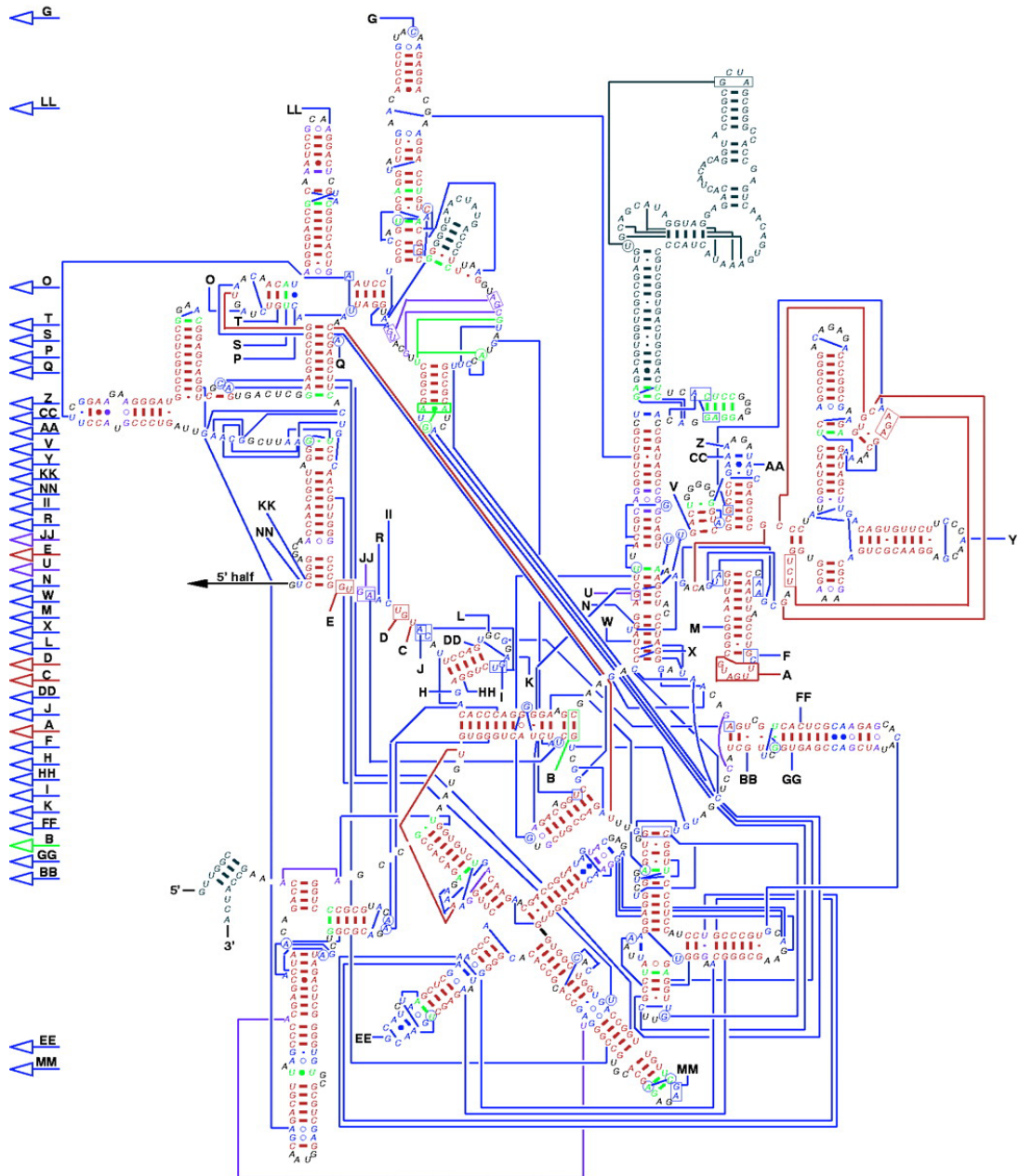
23S: interactions secondaires et tertiaires

Secondary Structure: large subunit ribosomal RNA - 5' half



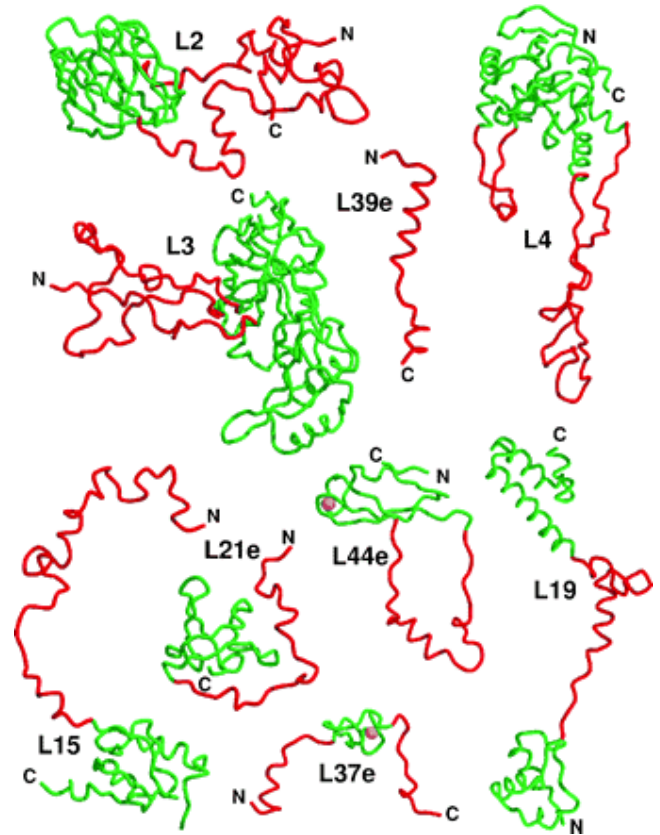
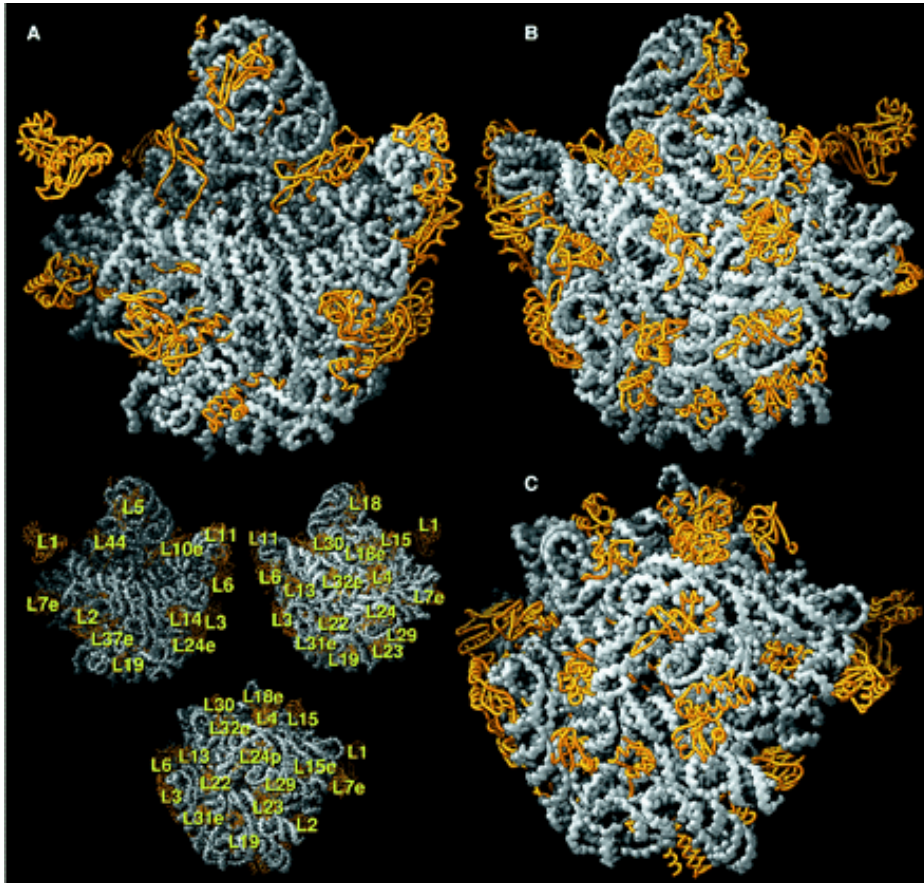
(moitié 5')

23S: interactions secondaires et tertiaires

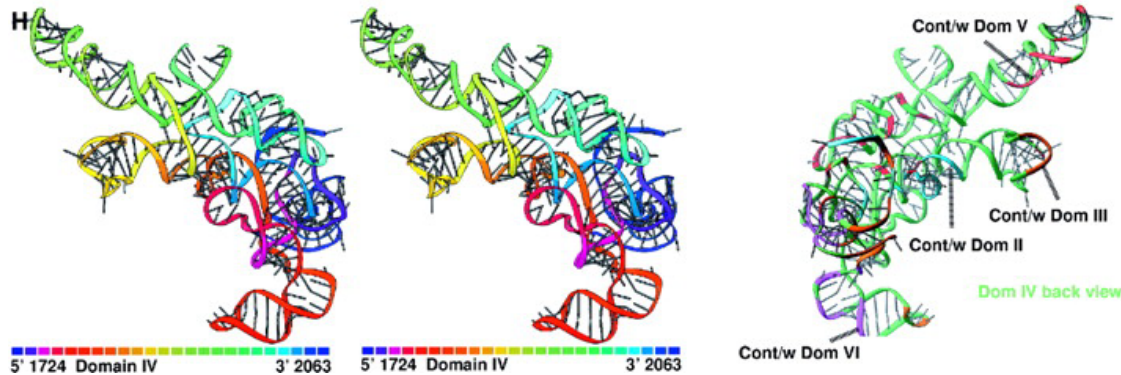
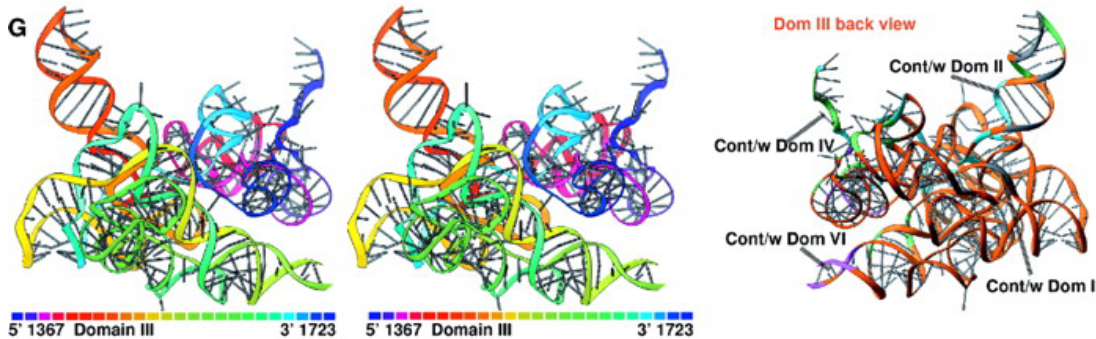
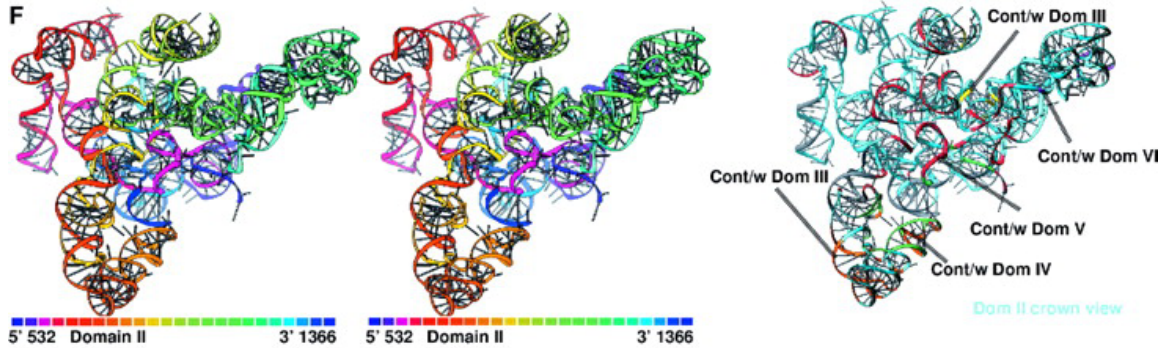
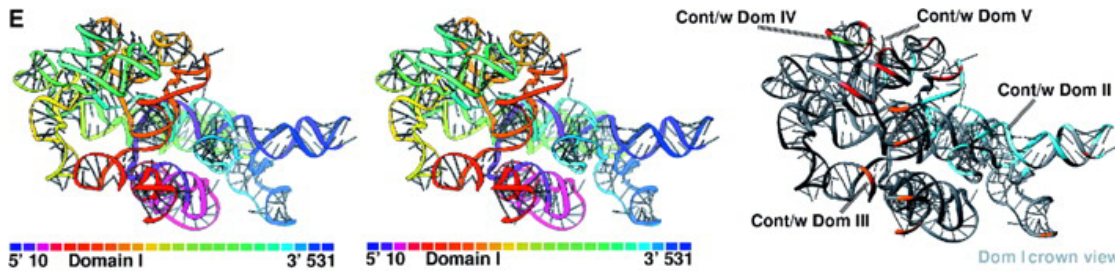


(moitié 3')

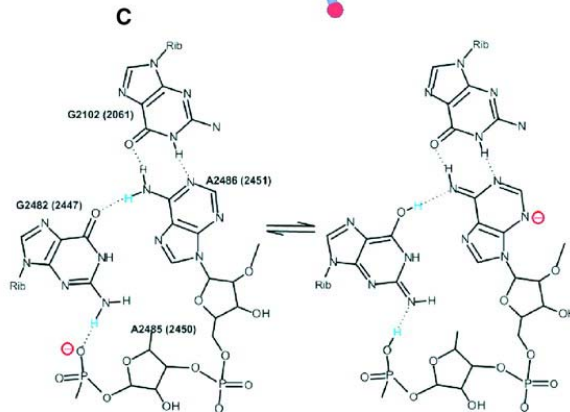
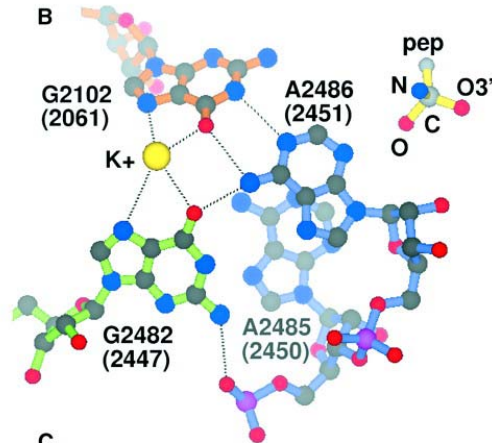
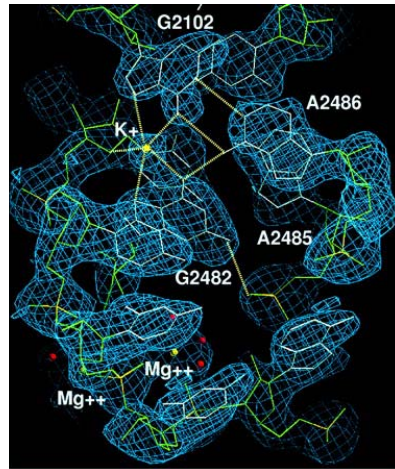
Protéines de la grande sous-Unité



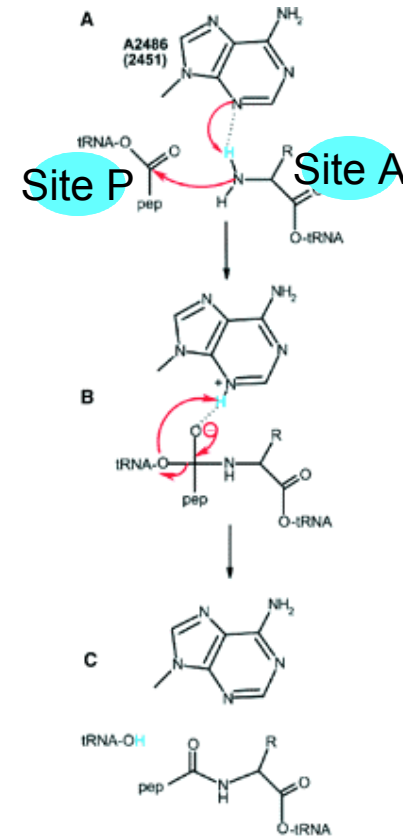
Replieement des domaines dominés par les interactions inter-hélices



Centre peptidyl-transférase



La stabilisation de l'imino tautomère du A2486 augmente le pKa de A2486 N3 à 7.6



Abstraction proton par N3 de 2486

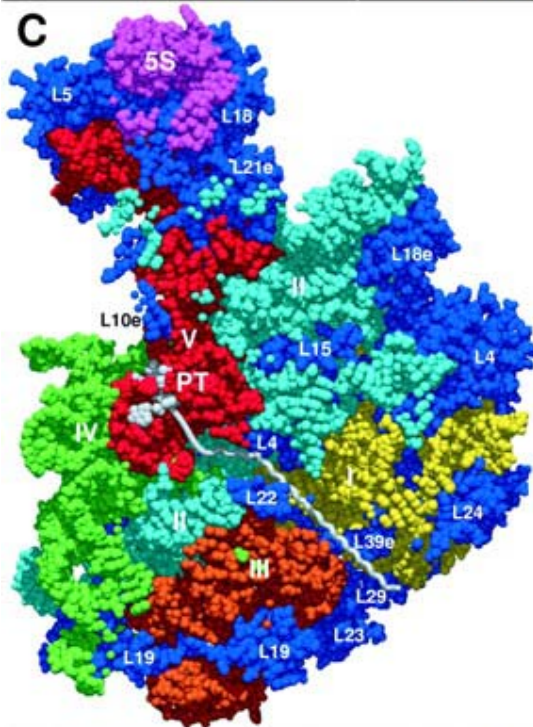
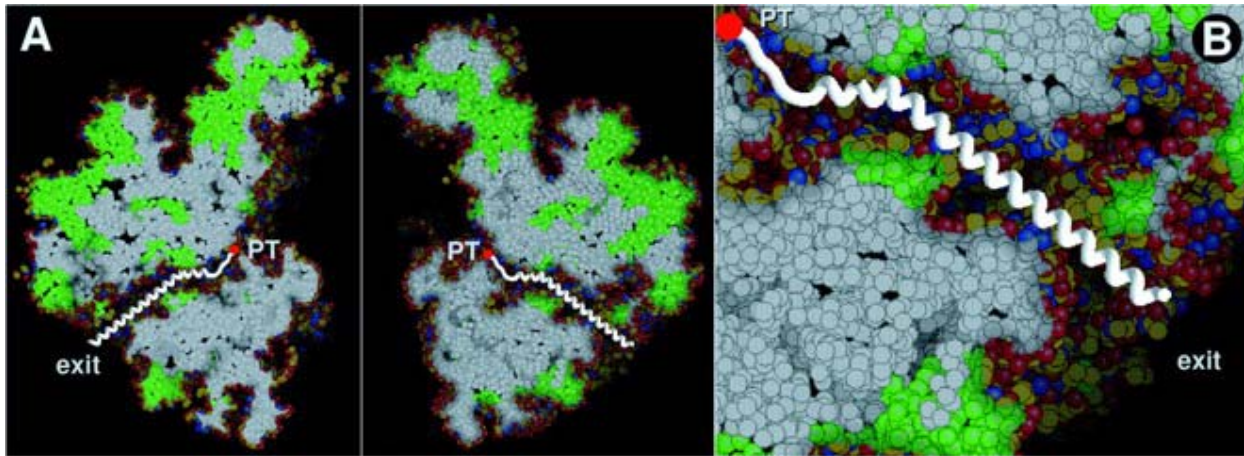
Attaque nucléophile

stabilisation du carbone tetrahedrique par N+

Site P libéré par le déplacement du peptide sur le tRNA du site A

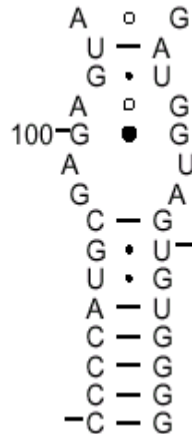
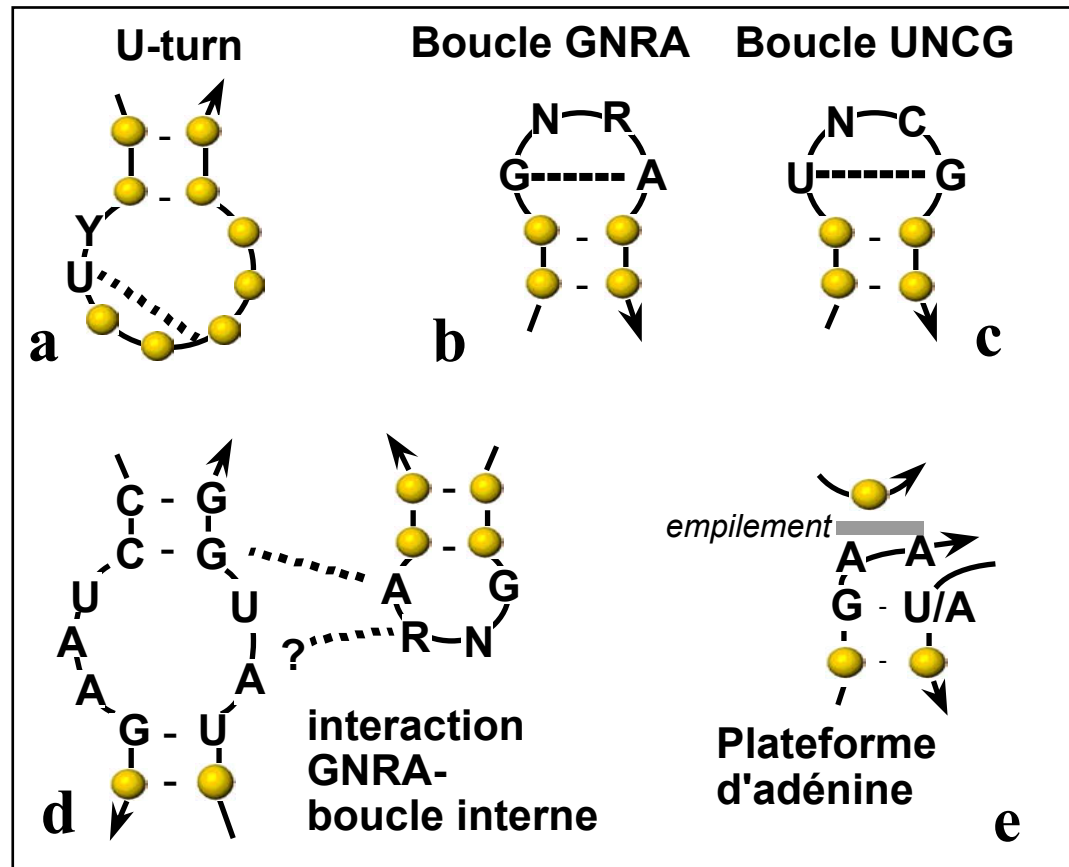
A proposed mechanism of peptide synthesis catalyzed by the ribosome. A2486 (A2451) is shown as the standard tautomer in all steps, but could be represented as the imino tautomer, which would have a negative unprotonated N3 and a neutral protonated N3. We expect that the electronic distribution is actually between these two extremes. (A) The N3 of A2486 abstracts a proton from the NH₂ group as the latter attacks the carbonyl carbon of the peptidyl-tRNA. (B) A protonated N3 stabilizes the tetrahedral carbon intermediate by hydrogen bonding to the oxyanion. (C) The proton is transferred from the N3 to the peptidyl tRNA 3' OH as the newly formed peptide deacylates. Among the variations on this mechanism that should be considered would be a protonated A2486 stabilizing the intermediate, as in (B), with less contribution on acid-base catalysis, as shown in (A) and (C).

Tunnel de sortie du polypeptide



The polypeptide exit tunnel. (A) The subunit has been cut in half, roughly bisecting its central protuberance and its peptide tunnel along the entire length. The two halves have been opened like the pages of a book. All ribosome atoms are shown in space-filling representation, with all RNA atoms that do not contact solvent shown in white and all protein atoms that do not contact solvent shown in green. Surface atoms of both protein and RNA are color-coded with carbon yellow, oxygen red, and nitrogen blue. A possible trajectory for a polypeptide passing through the tunnel is shown as a white ribbon. PT, peptidyl transferase site. (B) Detail of the polypeptide exit tunnel showing distribution of polar and nonpolar groups, with atoms colored as in (A), the constriction and bend in the tunnel formed by proteins L4 and L22 (green patches close to PT), and the relatively wide exit of the tunnel. A modeled polypeptide is in white. (C) The tunnel surface is shown with backbone atoms of the RNA color coded by domain. Domains I (yellow), II (light blue), III (orange), IV (green), V (light red), 5S (pink), and proteins are blue.

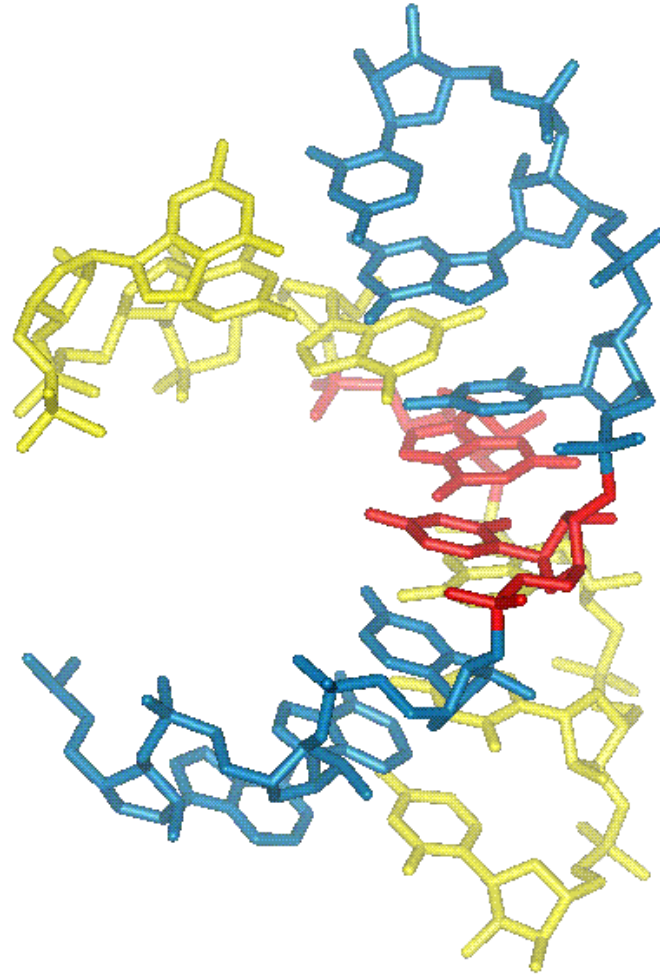
Les motifs ARN



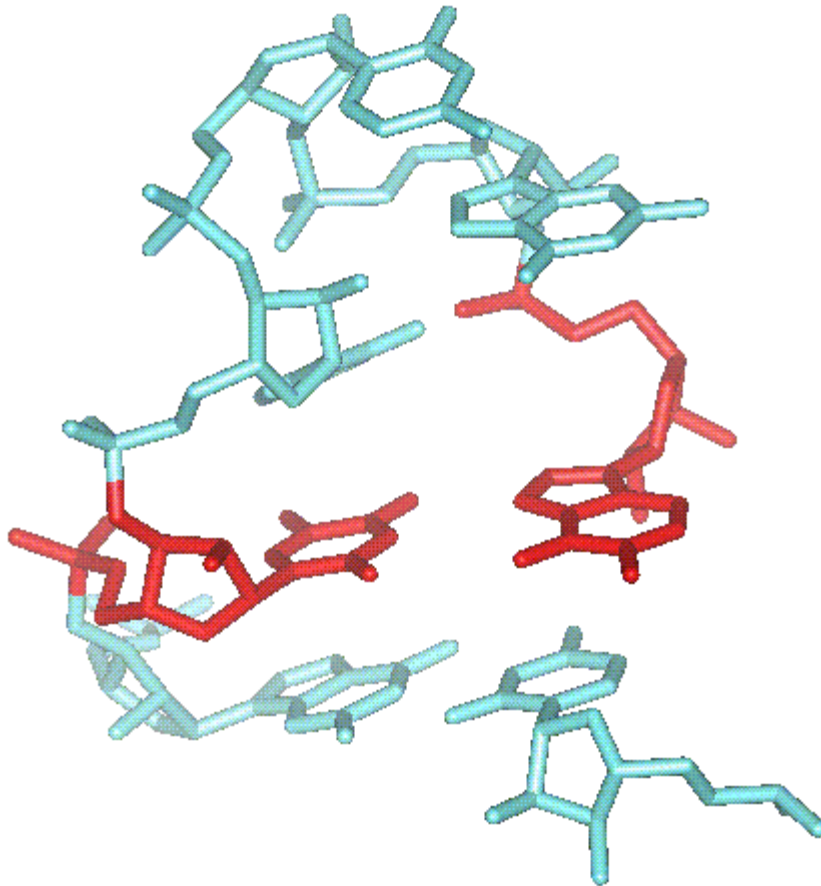
Boucle E

Quelques motifs ARN. Les traits pointillés désignent des interactions tertiaires. Les flèches indiquent le sens 5'->3'. Le point d'interrogation dans le motif *d* indique que le partenaire de cette interaction tertiaire n'est pas connu.

La paire G:U Wobble

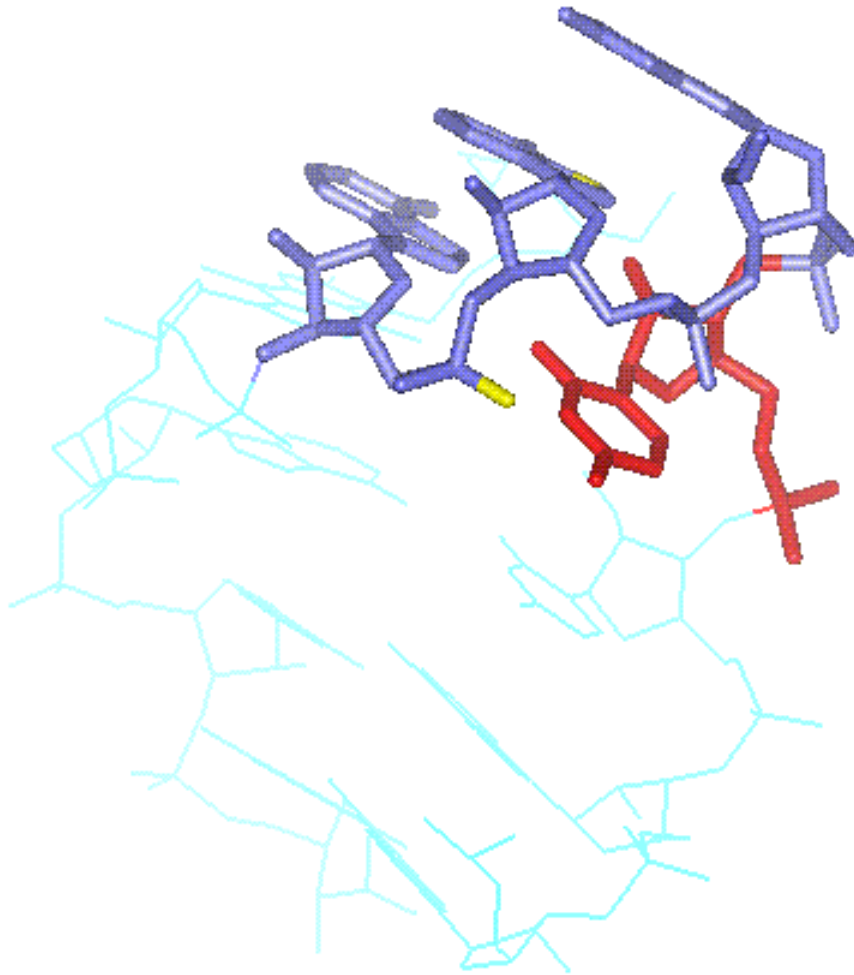


La paire Hoogsteen



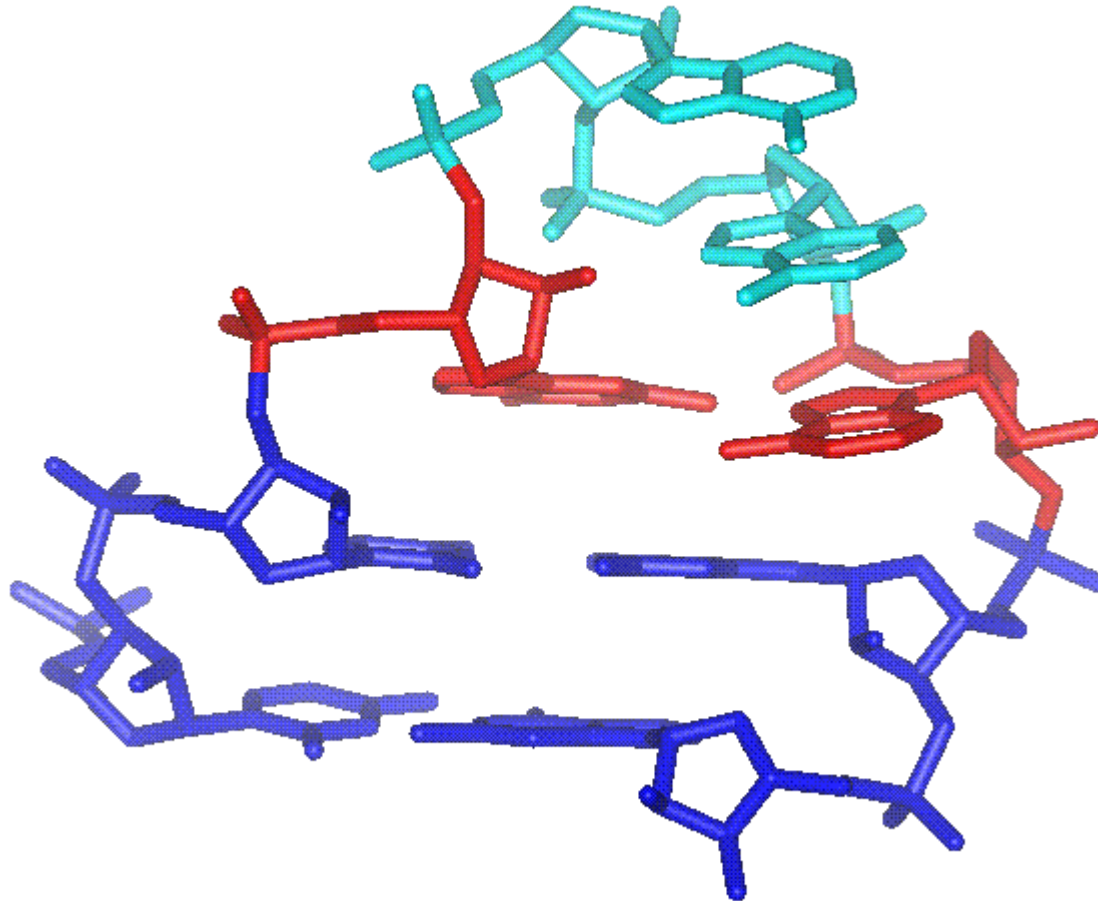
Boucle TTC de tRNA
Phe levure

Le U-turn

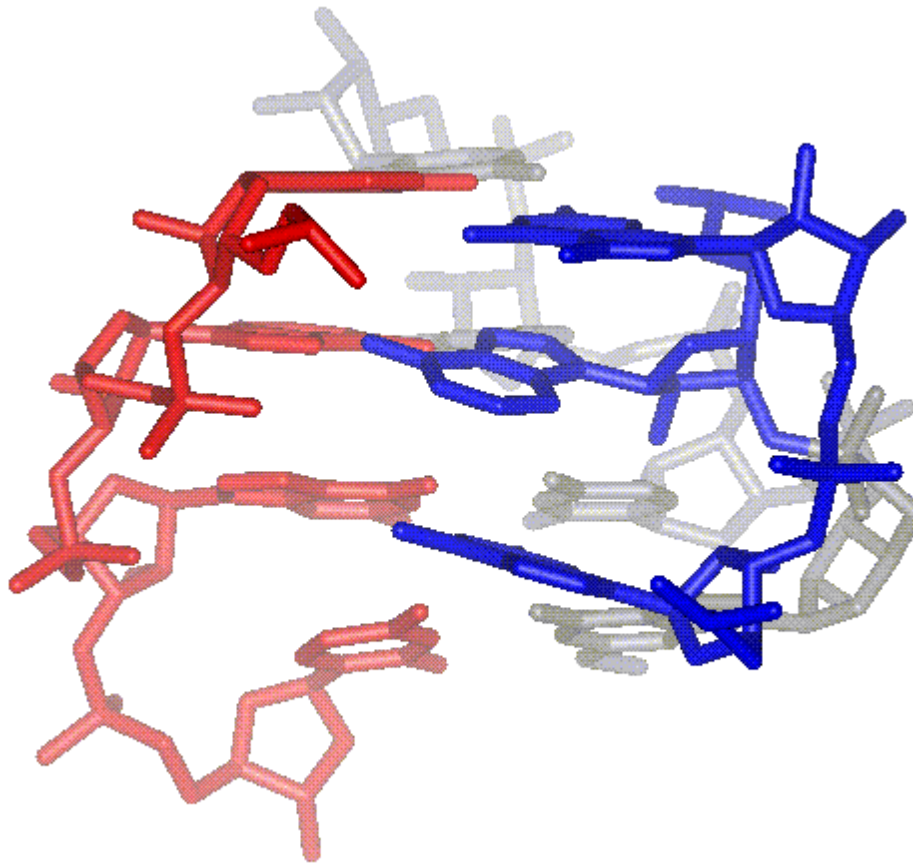


Boucle anticodon de
tRNA Phe levure

La boucle GNRA



Les triplets de bases



Triple hélice 10-13 + 22-
25+ 9+45+46 de tRNA
Phe levure

Algorithme de Zuker et Stiegler

- L'algorithme de programmation dynamique appliqué aux structures secondaires recherche les structures de plus basse énergie pour tous les sous-fragments d'une séquence (Zuker and Stiegler, 1981). Il commence par rechercher dans tous les sous-fragments la structure de plus basse énergie comportant une paire de bases au moins. Les premiers sous fragments considérés sont ceux pouvant former une hairpin fermée par une bp. Les hairpins ayant au minimum une taille 3, la première passe va identifier toutes les structures de plus basse énergie pour tous les pentanucléotides de la séquence. Les énergies de chaque structure sont calculées avec et sans la contrainte d'appariement des nucléotides terminaux. 2 matrices V et W sont ainsi remplies. L'étape suivante recommence avec les fragments de taille 6. A chaque étape, un nt est ajouté au fragment et la meilleure structure est recherchée pour ce fragment, en utilisant les valeurs calculées pour le fragment plus petit. On continue jusqu'à trouver l'énergie la plus basse pour la structure entière. La structure est ensuite reconstituée en retraçant à l'envers toutes les étapes ayant mené à cette solution.
- Cette méthode garantit une solution optimale à l'intérieur du modèle énergétique choisi. N'autorise pas les pseudo-nœuds.

Prédire les structures secondaires: énergies libres d'empilement (stacking free energy) (Turner et al.)

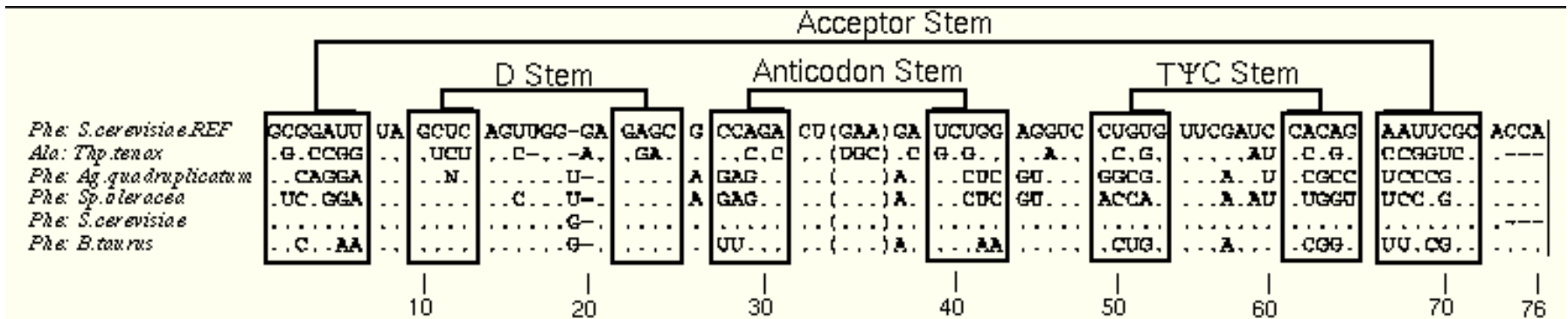
STACKING ENERGIES (exemples):

	Y				Y				Y				Y			
	A	C	G	U	A	C	G	U	A	C	G	U	A	C	G	U
	5' --> 3'				5' --> 3'				5' --> 3'				5' --> 3'			
	UX				UX				UX				UX			
	AY				CY				GY				UY			
	3' <-- 5'				3' <-- 5'				3' <-- 5'				3' <-- 5'			
A	-0.50	-0.30	-0.60	-0.50	-0.50	-0.30	-0.60	-0.50
C	-0.20	-0.10	-1.20	-0.00	-0.20	-0.10	-1.70	0.00
G	-1.40	-1.20	-0.70	-0.20	-0.80	-1.20	-0.30	-0.70
U	-0.30	-0.10	-0.50	-0.80	-0.60	-0.10	-0.60	-0.80

– Attention:



Analyse comparative



- L'analyse comparative est la façon la plus fiable de déterminer les structures secondaires
- Elle repose sur la détection de covariations
- Elle nécessite plusieurs séquences homologues alignées

Mesures de la covariation

- Table de contingence:

```
table p normal 53 61

      a  c  g  u  -
53, 61 | 0 97  0  3  0 61
-----+-----
a  3   | 0  0  0  3  0
c  0   | 0  0  0  0  0
g 97   | 0 97  0  0  0
u  0   | 0  0  0  0  0
-  0   | 0  0  0  0  0
53

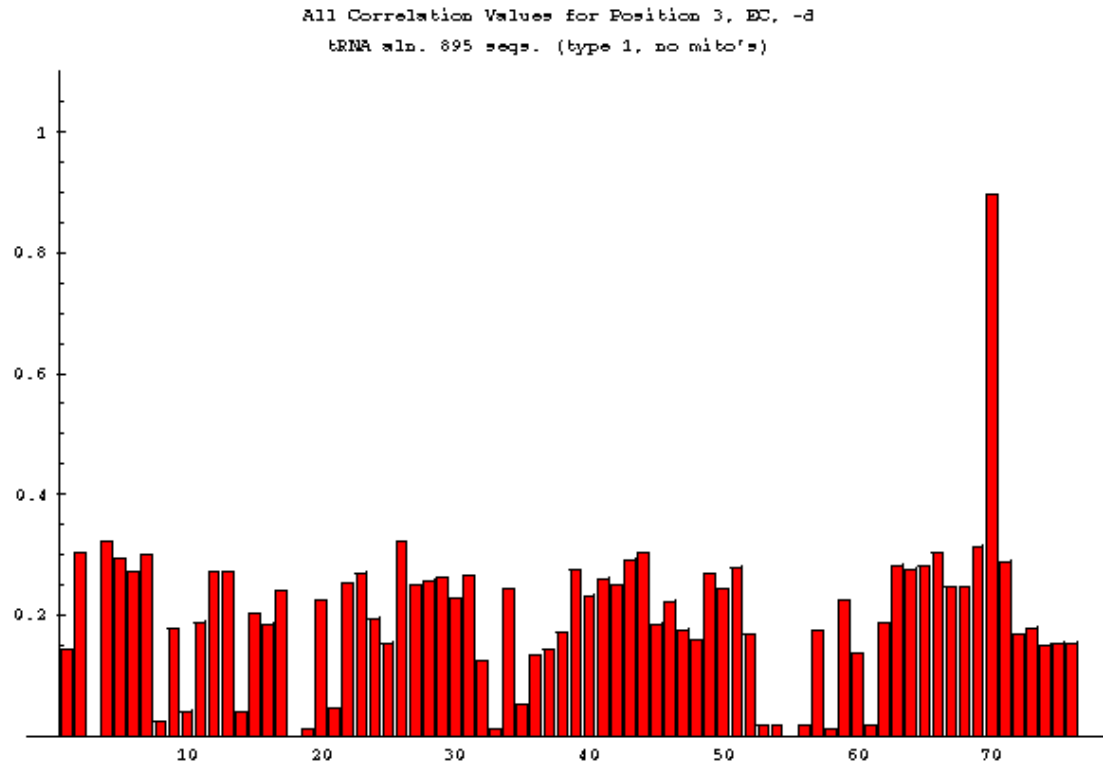
gc=( 29, 28.03, 96.7%)
au=(  1,  0.03,  3.3%)
```

- Tests:
 - Chi 2
 - Information Mutuelle
 - évènements phylogénétiques

$$\chi^2 = \sum_{M,N} \frac{[no(M_i, N_j) - ne(M_i, N_j)]^2}{ne(M_i, N_j)}$$

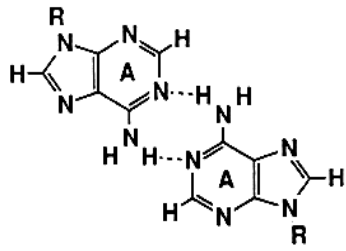
Détection systématique des covariations

- Position 1 du tRNA contre toutes les autres positions:

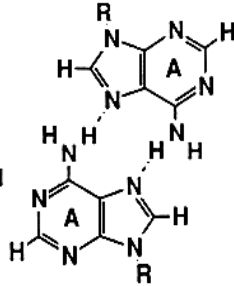


Interactions tertiaires

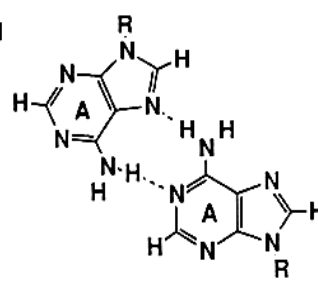
Paaires à 2 Ponts-H



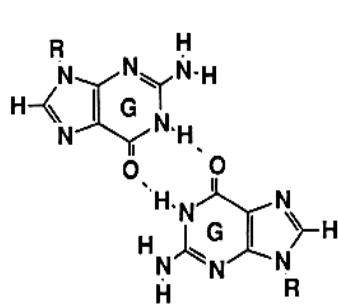
A-A N1-amino,
symmetric



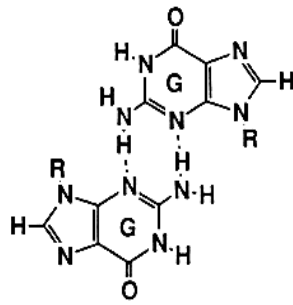
A-A N7-amino,
symmetric



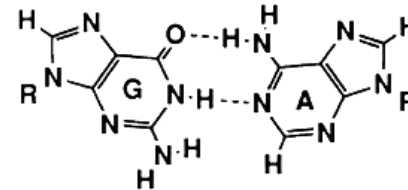
A-A N1-amino,
N7-amino



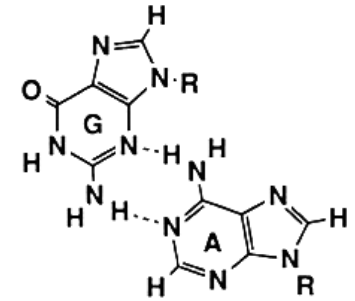
G-G N1-carbonyl,
symmetric



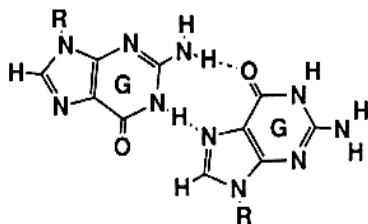
G-G N3-amino,
symmetric



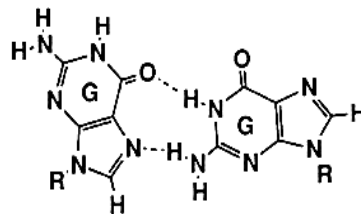
G-A N1-N1,
carbonyl-amino



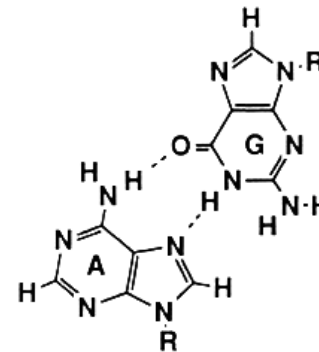
G-A N3-amino,
amino-N1



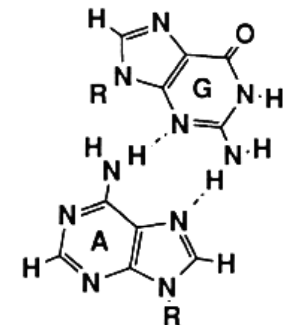
G-G N7-N1,
carbonyl-amino



G-G N1-carbonyl,
N7-amino

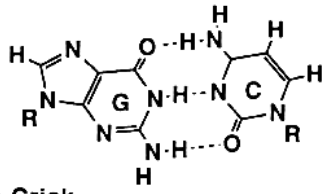
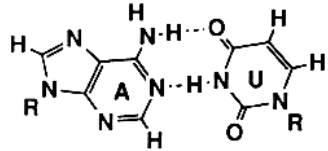


A-G N7-N1,
amino-carbonyl

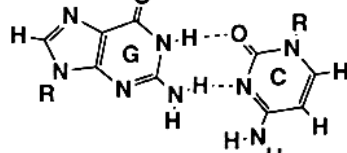
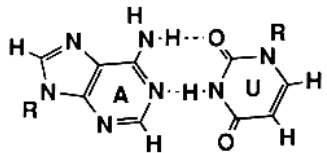


A-G N7-amino,
amino-N3

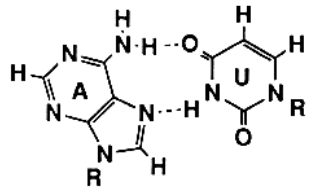
Suite...



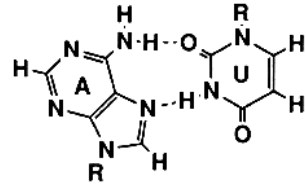
Watson-Crick



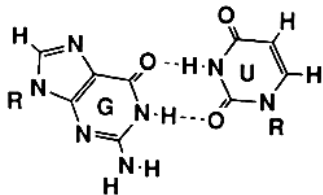
Reverse Watson-Crick



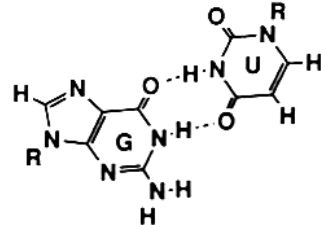
A-U Hoogsteen



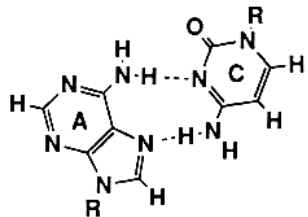
A-U Reverse Hoogsteen



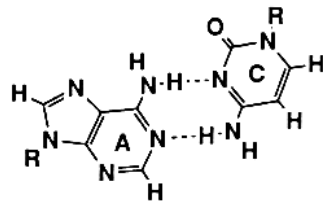
G-U Wobble



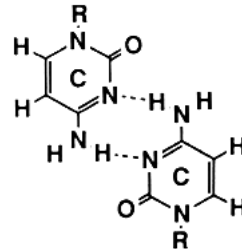
G-U Reverse Wobble



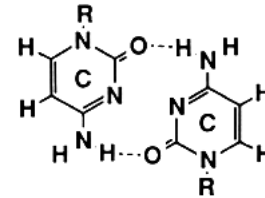
A-C Reverse Hoogsteen



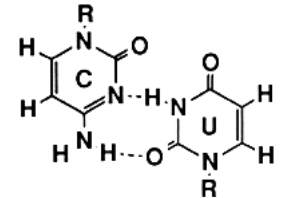
A-C Reverse Wobble



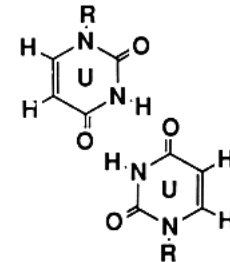
C-C N3-amino,
symmetric



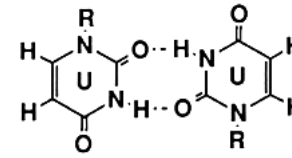
C-C carbonyl-amino,
symmetric



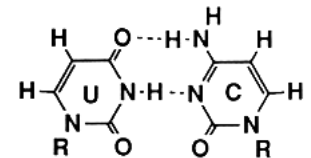
C-U N3-N3,
2-carbonyl-amino



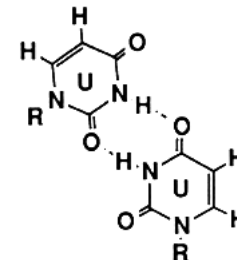
U-U 4-carbonyl-N3,
symmetric



U-U 2-carbonyl-N3,
symmetric



U-C N3-N3,
4-carbonyl-amino



U-U 2-carbonyl-N3,
4-carbonyl-N3